

UNIVERSIDAD COMPLUTENSE DE MADRID
FACULTAD DE COMERCIO y TURISMO



TESIS DOCTORAL

Marco conceptual para la toma de decisiones basadas en datos en organizaciones turísticas.
Casos de uso en digitalización de negocios y servicios, experiencia de turista y gestión de
destinos

MEMORIA PARA OPTAR AL GRADO DE DOCTOR

PRESENTADO POR

Juan Vidal Gil

DIRECTORES

Ramón Alberto Carrasco González y María Francisca Blasco López

UNIVERSIDAD COMPLUTENSE DE MADRID
FACULTAD DE COMERCIO y TURISMO
DOCTORADO INTERUNIVERSITARIO EN TURISMO



TESIS DOCTORAL

*Marco conceptual para la toma de decisiones basadas en datos en organizaciones turísticas.
Casos de uso en digitalización de negocios y servicios, experiencia de turista y gestión de destinos*

MEMORIA PARA OPTAR AL GRADO DE DOCTOR

PRESENTADO POR

Juan Vidal Gil

DIRECTORES

Ramón Alberto Carrasco González y María Francisco Blasco López

Dedicatoria

A mi mujer Yolanda, a mis hijos Juan y Ángela que son el motor de mi vida.

A mis padres y hermanos por todo su cariño y por estar siempre conmigo.

Agradecimientos

"Si quieres dominar algo, enséñalo. Cuanto más enseñas, mejor aprendes. La enseñanza es una herramienta poderosa para el aprendizaje" - Richard Feynman.

En primer lugar, deseo expresar mi profundo agradecimiento a mis directores de tesis, Ramón Alberto Carrasco y Francis Blasco. La aceptación que me dieron como doctorando y el constante apoyo que me ofrecieron a lo largo de estos años ha sido fundamentales para culminar esta tesis, y les estoy muy agradecido por ello. Igualmente agradecer su trabajo y apoyo a mi tutor Adolfo Hernández Estrada.

No fue sencillo para mí tomar la decisión de hacer un doctorado, en primer lugar, porque tenía que compaginarlo con el trabajo y en segundo lugar porque ya peinaba algunas canas y pensaba que quizás era algo que tenía que haber empezado antes. Sin embargo, hacer un doctorado y volver a la vida universitaria ha sido una experiencia buena y rejuvenecedora. Me ayudó mucho a tomar esta decisión mi director Ramón Alberto Carrasco, que además de animarme a hacer la tesis, aceptó ser mi director.

Enseguida comprobé que coincidimos en la forma de ver la investigación y nos pusimos manos a la obra. Hoy en día todo el mundo habla de innovación, pero creo que hay pocas personas que realmente sean innovadoras. Una de ellas es Ramón. En nuestras reuniones surgían ideas y siempre planteaba nuevas perspectivas para abordar la investigación que facilitaban el desarrollo de la misma. Creo que nos han sobrado ideas a implementar, los límites y el final lo ha marcado el marco temporal. Cuando compaginas trabajo y tesis, hay que hacer un esfuerzo adicional y una cuestión importante es superar los momentos de desánimo, tengo que decir que te tenido muy pocos y el optimismo de Ramón y Francis ayudaban a superarlo de forma rápida. Quiero destacar también el apoyo de Francis y su capacidad de gestión para facilitar el trabajo y agilizar las cosas.

Agradecer también el apoyo al equipo académico de la Facultad de Comercio y Turismo con el que he tenido ocasión de compartir diversas actividades y eventos en el programa de doctorado: Manuel de la Calle, María García Hernández y Diana Gómez Bruna que realizan un gran trabajo de apoyo al doctorando promoviendo la investigación en Turismo y lanzando muchas iniciativas de formación de gran interés. Ampliar el agradecimiento a todo el equipo de coordinación y a el comité académico del programa de doctorado interuniversitario por su gran trabajo para el desarrollo de la investigación en

Turismo, que ya está dando sus frutos. Agradecer igualmente las valiosas colaboraciones en la investigación de Manuel Cobo Martín de la Universidad de Granada, Eduardo Parra-López de la Universidad de la Laguna y de Gema Fernández-Avilés de la Universidad de Castilla-La Mancha con los que ha sido un honor trabajar y de los que he aprendido mucho.

Por último, y no menos importante, agradecimientos para mi familia y amigos, ya que sin ellos esto hubiera sido imposible. En primer lugar, a Yolanda, mi mujer, que ha vivido de cerca el proceso por su apoyo constante y a mis hijos Juan y Ángela por el tiempo que les haya podido dejar de dedicar, que he tratado de que sea el mínimo. Agradecimientos para mis amigos Darío Chimeno, Ángel Alonso y Alberto Hernández que escucharon pacientemente mis experiencias en la vida de doctorando. Fundamental también el apoyo de mi madre Carmen y de mis hermanos: Loli, Carmen, Rafa, Jesús y Fátima que no dejaron de animarme y apoyarme durante todo el proceso. Finalmente, agradecer a Juan, mi padre, que está en el cielo su ejemplo impagable y todo lo que me ha enseñado, le hubiera encantado ver esta tesis.

Gracias a todos.

Tabla de contenido

Agradecimientos	I
Resumen	12
Abstract	14
Desarrollo de la memoria	16
1. Introducción	17
1.1. Objetivos	22
1.1.1. Objetivo general	22
1.1.2. Objetivos específicos	23
1.2. Estructura de la memoria	24
2. Estado del arte	27
2.1. Trabajos relacionados	27
2.2. Metodología	30
2.3. Resultados	35
2.3.1. Análisis bibliométrico de rendimiento	35
2.3.2. Diagrama estratégico	38
2.3.3. Diagrama de evolución y solapamiento	46
2.3.4. Análisis cualitativo de áreas temáticas	48
2.3.5. Marco de referencia inicial	52
3. Marco teórico de investigación	55
3.1 Análisis de grafos	55
3.1.1 Métodos de agrupamiento para detección de comunidades	56
3.2 Proceso analítico jerárquico (AHP)	57
3.2.1 Definición del problema de decisión	59
3.2.2 Estructuración jerárquica del problema de decisión	59
3.2.3 Comparación por pares y cálculo del vector de ponderaciones	60
3.2.4 Obtención de las ponderaciones de las alternativas finales mediante varios niveles de agregación de ponderaciones	62
3.3 Cuadro de mando integral	62
3.4 Técnicas de agrupamiento de datos geográficos	63
3.4.1 Selección de las variables de agrupamiento y del tipo de distancia	64
3.4.2 Selección de la técnica de agrupamiento	66
3.5 Grandes modelos del lenguaje	70
4. Marco conceptual propuesto	72
4.1. Identificación de gaps	72

	4.2. Marco conceptual para la toma de decisiones basada en datos en una organización turística	79
5.	Caso de uso: detección y análisis de experiencias turísticas a partir de reseñas en línea para la mejora de la gestión de destinos turísticos. Aplicación en la ciudad de Roma.....	83
	5.1. Introducción.....	84
	5.2. Trabajos relacionados	86
	5.3. Metodología	91
	5.3.1. Modelizar experiencias con grafos	91
	5.4. Modelo propuesto	96
	5.4.1. Establecer los objetivos de negocio	97
	5.4.2. Recopilación de los datos.....	98
	5.4.3. Limpieza y depuración	99
	5.4.4. Extracción de la red.....	100
	5.4.5. Normalización.....	102
	5.4.6. Clustering.....	102
	5.4.7. Visualización e interpretación	104
	5.4.8. Toma de decisiones	112
	5.4.8.1. Recomendaciones generales	112
	5.4.8.2. Recomendaciones personalizadas.....	113
	5.5. Conclusiones.....	113
	5.5.1. Implicaciones teóricas.....	114
	5.5.2. Implicaciones en la gestión empresarial	114
	5.6. Trabajo futuro.....	115
6.	Caso de uso: enfoque de criterios múltiples aplicado a la transformación digital de tiendas físicas en el sector turístico. Aplicación en agencias de viajes	116
	6.1. Introducción.....	117
	6.2. Modelo propuesto	120
	6.2.1. Definición del problema de decisión	121
	6.2.2. Estructuración jerárquica del problema de decisión	122
	6.2.3. Comparación por pares y cálculo del vector de ponderaciones.....	123
	6.2.4. Cálculo del vector de ponderaciones para cada matriz y obtención de las ponderaciones de las alternativas finales mediante varios niveles de agregación de ponderaciones.....	127
	6.3. Conclusiones y trabajo futuro.....	128
7.	Caso de uso: plataforma de datos e inteligencia artificial para el posicionamiento de servicios turísticos inteligentes. Aplicación a las	

oficinas de información turística	130
7.1. Introducción.....	131
7.2. Revisión de literatura.....	135
7.3. Modelo propuesto	139
7.3.1. Comprender el dominio de aplicación	139
7.3.2. Creación del conjunto de datos	141
7.3.3. Preprocesamiento de los datos	143
7.3.4. Transformación de los datos	145
7.3.5. Construcción del modelo.....	146
7.3.6. Ejecución del modelo.....	148
7.3.7. Interpretación del conocimiento obtenido....	152
7.4. Resultados	163
7.5. Conclusiones y trabajo futuro.....	168
8. Conclusiones y trabajo futuro	170
8.1. Conclusiones.....	170
8.2. Trabajo futuro.....	172
9. Bibliografía.....	174
Anexo.....	194
a) Artículos publicados.....	195
b) Artículos en revisión	202
c) Menciones y reconocimientos académicos.....	203

Índice de figuras

Figura 1: Esquema de toma de decisiones basadas en datos en el sector del turismo. Fuente: elaboración propia.....	21
Figura 2: Estructura de la memoria	26
Figura 3: Principales pasos de la metodología empleada para la realización del análisis bibliométrico. Fuente: elaboración propia.	31
Figura 4: Diagrama estratégico. Fuente: Cobo et al. (2012).....	34
Figura 5: Evolución temporal del número de documentos por periodo. Fuente: elaborado con SciMAT .	36
Figura 6: Gráfico de barras que contiene las revistas con más de 10 publicaciones. Fuente: elaborado con SciMAT	37
Figura 7: Mapa del mundo con los países con más de 10 publicaciones. Fuente: elaborado con SciMAT	38
Figura 8: Diagrama estratégico para el periodo 1982-2015. Se muestra el número de documentos de cada tema. Fuente: elaborado con SciMAT	39
Figura 9: Diagrama estratégico para el periodo 1982-2015 clasificado por áreas temáticas. Se muestra el número de documentos de cada tema. Fuente: elaborado con SciMAT.....	42
Figura 10: Diagrama estratégico para el periodo 2016-2021. Se muestra el número de documentos de cada tema. Fuente: elaborado con SciMAT.....	43
Figura 11: Diagrama estratégico para el periodo 2016-2021 clasificado por áreas temáticas. Se muestra el número de documentos de cada tema. Fuente: elaborado con SciMAT.....	46
Figura 12: Mapa de evolución conceptual entre periodos. Fuente: elaborado con SciMAT	47
Figura 13: Palabras clave en el tema 'online-reviews'. Fuente: elaborado con SciMAT	50
Figura 14: Palabras clave en el tema 'clustering'. Fuente: elaborado con SciMAT	51
Figura 15: Palabras clave en el tema 'machine learning'. Fuente: elaborado con SciMAT	52
Figura 16: Palabras clave en el tema 'deep learning'. Fuente: elaborado con SciMAT	52
Figura 17: Marco de referencia inicial sobre el uso del dato en la toma de decisiones en organizaciones turísticas. Los temas pertenecientes al segundo periodo se resaltan en gris para diferenciarlos del primer periodo. Fuente: elaboración propia	54
Figura 18: Diagrama de un grafo con 6 vértices y 7 aristas. Fuente: elaboración propia.....	56
Figura 19: Problema de toma de decisiones en una jerarquía H. Fuente: elaboración propia	60
Figura 20: Escala de Saaty. Fuente: Saaty (1980)	61
Figura 21: Distancias euclídea (negro), distancia caminando a pie (azul) entre el Hotel AC Atocha y la oficina de turismo de la estación de Atocha. Fuente: elaborado con Google Maps.	65
Figura 22: Distancia a pie entre la oficina de información turística de la Puerta del Sol de Madrid y el hotel Regina Madrid. Fuente: elaborado con Google Maps	66
Figura 23: Ejemplo de dendograma con distintos de puntos de corte para obtener diferente número de clústeres en un dendograma. Fuente: elaboración propia.	68
Figura 24: Marco conceptual para la toma de decisiones basada en datos en una organización turística. Fuente: elaboración propia.	80
Figura 25: Integración del caso de uso 'Detección y análisis de experiencias turísticas a partir de reseñas en línea para la mejora de la gestión de destinos turísticos. Aplicación en la ciudad de Roma' dentro del marco conceptual propuesto. Fuente: elaboración propia	84
Figura 26: Resumen de los pasos dados para obtener las CDEs de las reseñas. Fuente: elaboración propia	94
Figura 27: CDE relacionada con la seguridad del destino. Fuente: elaboración propia.	94
Figura 28: Principales características de una CDE. Fuente: elaboración propia	95
Figura 29: Exploración personalizada dentro de un ejemplo de CDE relacionada con la seguridad del destino. Fuente: elaboración propia.....	96
Figura 30: Fases del modelo propuesto para identificar y analizar CDEs. Fuente: elaboración propia	97
Figura 31: Ejemplos de palabras agrupadas en el archivo tesauro empleado en Vosviewer. Fuente: elaboración propia	100
Figura 32: Ejemplos de co-ocurrencia de las palabras clave 'arte' y 'arquitectura'. Fuente: elaboración propia.....	101
Figura 33: Red de palabras clave generada por el método de co-ocurrencia. Fuente: elaborado con	

Vosviewer.....	102
Figura 34: Red completa de palabras clave con los 61 clústeres obtenidos. El tamaño de cada palabra clave es proporcional al número de ocurrencias de la misma. Fuente: elaborado con Vosviewer.....	104
Figura 35: Subgrafo del clúster 17. Fuente: elaborado con Vosviewer.....	105
Figura 36: Subgrafo del clúster 29. Fuente: elaborado con Vosviewer.....	106
Figura 37: Subgrafo del clúster 11. Se destacan las relaciones de la palabra clave 'scam'. Fuente: elaborado con Vosviewer.	107
Figura 38: Subgrafo del clúster 3. Se destacan las relaciones de la palabra clave 'pickpocket'. Fuente: elaborado con Vosviewer.	107
Figura 39: Subgrafo del clúster 40. Fuente: elaborado con Vosviewer.....	108
Figura 40: Subgrafo del clúster 46. Fuente: elaborado con Vosviewer.....	109
Figura 41: Subgrafo del clúster 59. Fuente: elaborado con Vosviewer.....	109
Figura 42: Subgrafo de relaciones de la palabra clave 'pickpocket'. Se muestran sólo aquellas relaciones con un número de co-ocurrencias mayor o igual a 10. Los colores muestran los distintos clústeres a los que pertenecen las palabras clave. Fuente: elaborado con Vosviewer.	111
Figura 43: Subgrafo de relaciones de la palabra clave 'crowd'. Se muestran sólo aquellas relaciones con un número de co-ocurrencias mayor o igual a 10. Los colores muestran los distintos clústeres a los que pertenecen las palabras clave. Fuente: elaborado con Vosviewer.	112
Figura 44: Integración del caso de uso 'Enfoque de criterios múltiples aplicado a la transformación digital de tiendas físicas en el sector turístico. Aplicación en agencias de viajes' dentro del marco conceptual propuesto. Fuente: elaboración propia.....	117
Figura 45: Propuesta de modelo MCDM para la digitalización estratégica de las tiendas físicas de una agencia de viajes. Fuente: elaboración propia.....	121
Figura 46: Jerarquía H para la estrategia de digitalización de tiendas físicas de agencias de viajes. Fuente: elaboración propia.....	123
Figura 47: Integración del caso de uso 'Plataforma de datos e inteligencia artificial para el posicionamiento de STS. Aplicación a las oficinas de información turística' dentro del marco conceptual propuesto. Fuente: elaboración propia.....	131
Figura 48: Plataforma de datos e inteligencia artificial para el posicionamiento de STS. Fuente: elaboración propia.....	133
Figura 49: En la imagen de la izquierda se muestra una visualización sobre un mapa de Madrid en el que se resalta la zona geográfica seleccionada (almendra central) con las divisiones territoriales en barrios de los datos seleccionados referentes a oficinas de turismo (puntos negros) y hoteles (puntos grises). En la imagen de la derecha se añade a la anterior información los puntos de interés turístico: teatros (puntos verdes claro), museos (puntos azules) y edificios de carácter monumental (puntos verdes oscuro)......	145
Figura 50: Operaciones realizadas en las fases creación, preprocesamiento y transformación de los conjuntos de datos. Fuente: elaboración propia.....	146
Figura 51: Dendograma obtenido para cada uno de los métodos empleados.....	148
Figura 52: Entorno interactivo para visualizar los resultados del posicionamiento de las nuevas oficinas de turismo propuestas. Se muestra la selección del número de oficinas a proponer y el algoritmo a utilizar.....	149
Figura 53: Entorno interactivo para visualizar los resultados del posicionamiento de las nuevas oficinas de turismo propuestas. Se muestra la selección del número de oficinas a proponer, el algoritmo a utilizar y si se desean visualizar puntos de interés turístico o no. Los puntos de interés turístico se muestran en tonalidades verdes.....	149
Figura 54: Entorno interactivo para visualizar los resultados del posicionamiento de las nuevas oficinas de turismo. Se muestra la selección del mapa y los elementos a visualizar.	150
Figura 55: Recomendaciones generadas en el entorno interactivo para una oficina de turismo (TO1) de las propuestas por el entorno.....	150
Figura 56: Posicionamiento de 4 nuevas oficinas de turismo empleando el algoritmo Ward.D.	151
Figura 57: Cálculos del modelo realizados en batch y cálculos realizados en online en el entorno interactivo. Fuente: elaboración propia.	152
Figura 58: Generación de 4 grupos de hoteles y posicionamiento de oficinas con el método single.....	153
Figura 59: Generación de 4 grupos de hoteles y posicionamiento de oficinas con el método complete..	154
Figura 60: Generación de 4 grupos de hoteles y posicionamiento de oficinas con el método average....	154

Figura 61: Generación de 4 grupos de hoteles y posicionamiento de oficinas con el método mcquitty. .	155
Figura 62: Generación de 4 grupos de hoteles y posicionamiento de oficinas con el método median.....	155
Figura 63: Generación de 4 grupos de hoteles y posicionamiento de oficinas con el método centroid. ..	155
Figura 64: Generación de 4 grupos de hoteles y posicionamiento de oficinas con el método Ward.D....	156
Figura 65: Generación de 4 grupos de hoteles y posicionamiento de oficinas con el método Ward.D2..	156
Figura 66: Generación de 2 grupos de hoteles y posicionamiento de oficinas con el método Ward.D....	157
Figura 67: Generación de 3 grupos de hoteles y posicionamiento de oficinas con el método Ward.D....	158
Figura 68: Generación de 4 grupos de hoteles y posicionamiento de oficinas con el método Ward.D....	158
Figura 69: Generación de 5 grupos de hoteles y posicionamiento de oficinas con el método Ward.D....	158
Figura 70: Generación de 4 grupos de hoteles y posicionamiento de oficinas con el método Ward.D. Se añaden los puntos de interés turística en tonalidades verdes.	159
Figura 71: Visualización en el entorno interactivo de una de las oficinas de turismo TO1 ubicada en la zona Estación de Atocha y Paseo del Prado y recomendaciones obtenidas para la misma.	161
Figura 72: Oficinas de turismo ya existentes.....	164
Figura 73: Oficina de turismo nueva (punto negro) posicionada cerca del eje turístico de Plaza de España y Gran Vía.	165
Figura 74: Oficina de turismo nueva (punto negro) posicionada cerca del eje turístico del Paseo del Prado y la estación de Atocha.	165
Figura 75: Oficina de turismo nueva (punto negro) posicionada en la Castellana en la confluencia con la calle José Ortega y Gasset cerca de la plaza de Colón	166
Figura 76: Oficina de turismo nueva (punto negro) posicionada en la Castellana cerca de la plaza de Cuzco.	166

Índice de tablas

Tabla 1: Resumen de revisiones de literatura o estudios bibliométricos centrados en la investigación basada en datos	30
Tabla 2: Indicadores de rendimiento por temas en el periodo 1982-2015.....	40
Tabla 3: Indicadores de rendimiento por área temática en el periodo 1982-2015.....	42
Tabla 4: Indicadores de rendimiento por temas en el periodo 2016-2021.....	44
Tabla 5: Indicadores de rendimiento por área temática en el periodo 2016-2021	46
Tabla 6: RI según el tamaño de la muestra.....	62
Tabla 7: Métodos de clustering jerárquico	69
Tabla 8: Artículos que analizan datos de tipo UGC para obtener información útil sobre la gestión de los destinos turísticos.	90
Tabla 9: Iniciativas de mejora basadas en la información relevante obtenida de las recomendaciones generales.....	113
Tabla 10: Iniciativas de mejora basadas en la información relevante obtenida de las recomendaciones personalizadas.....	113
Tabla 11: Lista de dispositivos digitales seleccionados.....	121
Tabla 12: Clasificación de las alternativas en el problema MCDM, pesos de los distintos niveles de la jerarquía y CR.....	128
Tabla 13: Trabajos relacionados con nuestra propuesta y diferenciación en base a las diferentes capas de la plataforma propuesta.....	138
Tabla 14: Información relevante identificada clasificada en los distintos tipos de fuentes de datos de la plataforma propuesta.	140
Tabla 15: Conjuntos de datos a utilizar y volumen de datos.	142
Tabla 16: Conjuntos de datos con los datos seleccionados después del filtrado espacial por la zona geográfica de la almendra central y el filtrado de hoteles en base a la proximidad a las actuales oficinas de turismo.	144

Índice de acrónimos

AHP: Analytic Hierarchy Process.

API: Application Programming Interface.

AR: Augmented Reality.

BSC: Balanced Scorecard.

CDE: Common-destination-experience.

CI: Consistency Index.

CR: Consistency Ratio.

DMO: Destination Management Organizations.

eWOM: Electronic Word of Mouth.

GPS: Global Positioning System.

HTML: Hypertext Markup Language.

IoT: Internet of Things.

KDD: Knowledge Discovery in Databases.

LLM: Large Language Model.

MCDM: Multi-criteria Decision Making

NTO: National Tourism Organizations.

NoSQL: Not Only SQL Databases.

OWA: Ordered Weighted Averaging.

RAG: Retrieval Augmented Generation.

RI: Random Consistency Index.

SQL: Structured Query Language.

STS: Smart Tourism Service.

UGC: User Generated Content.

UTM: Universal Transverse Mercator.

VR: Virtual Reality.

wd-5: walkable distance-5 min.

WGS84: World Geodetic System 1984.

WoS: Web of Science.

Resumen

En el actual entorno empresarial, la orientación al mercado se presenta como un requisito estratégico para mantener una posición competitiva sólida y sostenible. Esta orientación implica comprender las necesidades y expectativas de los clientes y analizar las capacidades y estrategias de los competidores. La transformación hacia organizaciones centradas en datos ha cobrado relevancia, y la inversión en estrategias de datos y plataformas de inteligencia artificial se considera fundamental para lograr una auténtica orientación al mercado. Esta estrategia debe generar inteligencia de mercado, difundir la información en toda la organización y convertirla en acciones concretas.

El turismo se ha visto profundamente afectado por estos cambios, convirtiéndose en un ámbito donde el análisis de datos ha revolucionado la forma en que se toman decisiones en la gestión turística. En la última década, el sector turístico ha presenciado un considerable aumento en las fuentes de información disponibles. La proliferación de dispositivos móviles ha llevado a una mayor conectividad de los turistas, generando lo que se conoce como "huellas digitales". Estos datos, manejados con cuidado en términos de privacidad, permiten a las organizaciones tomar decisiones inteligentes para el desarrollo de destinos turísticos. Además de los datos móviles, se dispone de información de redes sociales, opiniones en plataformas turísticas, búsquedas en internet y más, lo que proporciona una visión completa tanto del turista como del mercado. De igual modo, los datos publicados en abierto por instituciones públicas o en espacios de datos compartidos han experimentado un notable crecimiento.

En el sector turístico, la gestión efectiva de las diversas fuentes de datos es esencial para aplicar técnicas analíticas que guíen la toma de decisiones. Es crucial disponer de plataformas de datos que permitan la explotación y análisis de datos para aplicaciones como la inteligencia de negocio o la minería de datos. En el contexto actual, la inteligencia artificial está cobrando mayor relevancia en la toma de decisiones y todo el desarrollo de la misma es de gran interés y aplicación en el sector del turismo. La industria turística se caracteriza por su complejidad inherente debido a la participación de diversos agentes tanto públicos como privados en su cadena de valor. Esta complejidad se manifiesta en la diversidad de datos que provienen de distintas fuentes y ofrecen perspectivas variadas, pero complementarias. En el ámbito del turismo, tomar decisiones efectivas requiere tener en cuenta una variedad de factores y criterios interconectados. En este contexto, el uso de modelos para la toma de decisiones multicriterio (MCDM por sus siglas en inglés Multi-

Criteria Decision Making) puede ser fundamental para abordar los desafíos específicos relacionados con la gestión y planificación en esta industria.

Consideramos, por tanto, que en el complejo contexto actual una orientación a mercado en el sector del turismo precisa de un marco de trabajo sólido que simplifique esta gestión y respalde la toma de decisiones basadas en datos. El objetivo de esta tesis es abordar esta necesidad de establecer un marco de trabajo adecuado.

Como primer paso para conocer la situación del sector del turismo respecto al uso de datos en la toma de decisiones, realizamos un análisis bibliométrico sobre el uso de fuentes de datos en investigaciones turísticas basadas en datos. Utilizaremos los resultados del análisis bibliométrico para identificar los principales temas de investigación turística, las técnicas de análisis de datos utilizadas y las fuentes de información empleadas. Estos conocimientos permitirán obtener un marco de referencia inicial sobre el uso del dato en la toma de decisiones en organizaciones turísticas. A partir de este marco de referencia inicial analizaremos si se cubren adecuadamente las necesidades actuales de una organización turística orientada al dato.

Una vez definida la situación de partida en el marco de referencia inicial e identificadas las necesidades detectadas en el mismo, definiremos el marco conceptual para la toma de decisiones basadas en datos en una organización turística.

Tomando como base este marco conceptual se desarrollarán tres casos de uso que servirán de paradigma de aplicación del mismo:

- Detección y análisis de experiencias turísticas a partir de reseñas en línea para la mejora de la gestión de destinos turísticos. Aplicación en la ciudad de Roma.
- Enfoque de criterios múltiples aplicado a la transformación digital de tiendas físicas en el sector turístico. Aplicación en agencias de viajes.
- Plataforma de datos e inteligencia artificial para el posicionamiento de servicios turísticos inteligentes. Aplicación en oficinas de información turística.

Abstract

In the current business environment, market orientation is a strategic requirement for maintaining a strong and sustainable competitive position. This orientation involves understanding customer needs and expectations and analyzing competitors' capabilities and strategies. The transformation to data-centric organizations has gained prominence, and investment in data strategies and artificial intelligence platforms is seen as critical to achieve true market orientation. This strategy must generate market intelligence, disseminate the information throughout the organization and turn it into concrete actions.

Tourism has been significantly affected by these changes, becoming an area where data analysis has revolutionized the way decisions are made in tourism management. In the last decade, the tourism sector has witnessed a considerable increase in the sources of information available. The proliferation of mobile devices has led to greater connectivity of tourists, generating what are known as "digital footprints". This data, handled with care in terms of privacy, allows organizations to make intelligent decisions for the development of tourism destinations. In addition to mobile data, information is available from social networks, opinions on tourism platforms, internet searches and more, providing a complete view of both the tourist and the market. Similarly, data published openly by public institutions or in shared data spaces has experienced significant growth.

In the tourism sector, effective management of various data sources is essential to apply analytical techniques to guide decision making. It is crucial to have data platforms that allow the exploitation and analysis of data for applications such as business intelligence or data mining. In the current context, artificial intelligence is becoming more relevant in decision making and the whole development of artificial intelligence is of great interest and application in the tourism sector. The tourism industry is characterized by its inherent complexity due to the participation of various public and private agents in its value chain. This complexity is manifested in the diversity of data coming from different sources and offering varied but complementary perspectives. In the field of tourism, making effective decisions requires taking into account a variety of interconnected factors and criteria. In this context, we believe that the use of Multi-Criteria Decision Making (MCDM) models can be instrumental in addressing the specific challenges related to management and planning in this industry.

We believe that in the current complex context, a market orientation in the tourism sector requires a solid framework that simplifies this management and supports data-

driven decision making. The objective of this thesis is to address this need to establish an appropriate framework for market orientation.

As a first step to understand the situation of the tourism sector regarding the use of data in decision making, we conducted a bibliometric analysis on the use of data sources in data-driven tourism research. We will use the results of the bibliometric analysis to identify the main tourism research topics, the data analysis techniques used and the data sources employed. This knowledge will allow us to obtain an initial frame of reference on the use of data in decision-making in tourism organizations. From this initial frame of reference, we will analyze if the current needs of a data-oriented tourism organization are adequately covered.

Once the starting situation has been well defined in the initial frame of reference and the needs detected in it have been identified, we will define the conceptual framework for data-driven decision making in a tourism organization.

Based on this conceptual framework, three use cases will be developed to serve as paradigm of its application:

- Detection and analysis of tourism experiences from online reviews for the improvement of tourism destination management. Application in the city of Rome.
- Multi-criteria approach applied to the digital transformation of physical stores in the tourism sector. Application in travel agencies.
- Data and artificial intelligence platform for the positioning of smart tourism services. Application in tourist information offices.

Desarrollo de la memoria

1. Introducción

En el complejo y cambiante escenario empresarial, la adopción de una orientación a mercado se presenta como un imperativo estratégico para las compañías que aspiran a mantener una posición competitiva sólida y sostenible. La orientación a mercado implica una estrategia empresarial (Kohli & Jaworski, 1990; Narver & Slater, 1990; Rapert & Yarbrough, 1998) centrada en la comprensión profunda de las necesidades y expectativas de los clientes, lo que permite ofrecer productos y servicios alineados con sus deseos. Esta práctica no solo se limita a conocer qué es lo que buscan los consumidores, sino también a anticiparse a sus cambios de comportamiento y adaptar las ofertas en función de esas expectativas. Además, la orientación a mercado se basa en un análisis detallado del entorno competitivo (Sorensen, 2009), observando de cerca las capacidades, fortalezas y estrategias de los competidores. Esto permite identificar oportunidades para diferenciarse, así como amenazas que podrían comprometer la posición de la empresa en el mercado. Al integrar esta visión centrada en el cliente y en los competidores, las organizaciones pueden tomar decisiones más informadas y basadas en datos, lo que facilita la innovación y mejora la adaptación en mercados en constante cambio. De esta forma, la orientación a mercado se convierte en un pilar clave para lograr ventajas competitivas sostenibles, al estar siempre en sintonía con las dinámicas del mercado y la evolución de las preferencias de los consumidores.

En los últimos años, la transformación hacia organizaciones verdaderamente orientadas al dato (Davenport, 2015; Svensson & Taghavianfar, 2020) ha adquirido una gran relevancia, y muchas empresas consideran la inversión en estrategias de datos y plataformas de inteligencia artificial como un pilar fundamental para la consecución de una auténtica orientación a mercado. Obviamente, el objetivo es utilizar estos datos para tener una mejor orientación al mercado (Kohli & Jaworski, 1990, 2023). Esta estrategia de orientación debe basarse en la generación de inteligencia de mercado sobre las necesidades actuales y futuras de los clientes en toda la organización, debe facilitar la difusión de la información en todos los departamentos y debe tener la capacidad de convertir esta información en acciones concretas de mercado (Narver & Slater, 1990; Narver et al., 2004; Borges et al., 2021).

El sector del turismo ha sido parte de esta transformación, donde el uso de datos ha redefinido la manera en que se toman decisiones en la gestión turística. La materia prima de esta transformación es, obviamente, los datos. Las fuentes de información disponibles

en el sector turístico han crecido significativamente en la última década. Los turistas, cada vez más conectados a través de dispositivos móviles, generan constantemente información que conocemos como "huellas digitales". Esta información, aplicada en el contexto de la protección de datos, permite a las organizaciones públicas y privadas tomar decisiones para el desarrollo inteligente de los destinos turísticos, ayudando a obtener una visión 360° tanto del turista con el objetivo de anticiparse a sus necesidades y adaptar los servicios a la realidad del visitante (Camilleri, 2020) como de la competencia, tal y como exige la orientación al mercado. Los datos generados por los dispositivos móviles son solo una parte de los datos generados por los turistas (Li et al., 2018) ya que también contamos con los datos generados en redes sociales, opiniones y comentarios en plataformas de servicios turísticos, búsquedas en internet, reservas online o pagos con tarjeta, entre otros. De igual modo, los datos publicados en abierto (Celdran-Bernabeu et al., 2018) han experimentado un notable crecimiento. Muchas organizaciones publican estadísticas sobre el sector turístico, al igual que diversas fuentes de datos en abierto con información de interés para el sector. En este sentido, cabe destacar la reciente creación de espacios de datos compartidos en el sector turismo entre los que podemos mencionar el ecosistema digital europeo Gaia-X (Gaia-X, 2023; Braud et al., 2021), y los proyectos de espacios de datos turísticos europeos (Tourism Data Space, 2023; DATES, 2023). Así mismo y centrándonos en las iniciativas españolas, hay que destacar el desarrollo de los sistemas de inteligencia turística y la creación de la Plataforma inteligente de destinos (Plataforma inteligente de destinos, 2023).

Estos diferentes tipos de fuentes de datos deben almacenarse y gestionarse con el fin de aplicar técnicas analíticas que permitan obtener información relevante para la toma de decisiones. Las plataformas de gestión de datos han tenido un desarrollo importante en el sector del turismo en la última década. En este sentido, los almacenes de datos han sido ampliamente utilizados en el sector turístico (Navarro & Rubio, 2000; Abdulaziz et al., 2015) y más recientemente también plataformas de datos más actuales como los lagos de datos (Sankaranarayanan & Lalchandani, 2017; Raju et al., 2018) o la evolución de los mismos hacia arquitecturas modernas que combinan ambos enfoques: el de los almacenes de datos y el de los lagos de datos. Este tipo de arquitecturas modernas que tratan de aprovechar lo mejor de los dos enfoques (Harby & Zulkernine, 2022) se suelen referenciar por su término en inglés *data lakehouse*, que utilizaremos en lo sucesivo en esta memoria para referirnos a ellas. El correcto almacenamiento y gestión facilita la explotación y el análisis de los datos. Un tipo de explotación y análisis muy extendido en el sector de turismo es la aplicación de técnicas de inteligencia de negocio con diferentes fines (Bustamante et al., 2020; Fuchs et al., 2013), tales como el seguimiento de indicadores

turísticos, el análisis de percepción del turista, el análisis geoespacial de puntos de interés turístico o el análisis de la sostenibilidad de los destinos turísticos. De igual forma, la aplicación de técnicas de minería de datos (Law et al., 2007; Rodríguez et al., 2018), permite añadir técnicas estadísticas en el análisis de estos datos, posibilitando ir un paso más allá en la analítica de datos, con objeto de obtener patrones, tendencias o relaciones no evidentes. Entre las técnicas de minería de datos es frecuente usar técnicas de agrupamiento (Jain et al., 1999) para detectar grupos de turistas con un comportamiento similar. Estos grupos también se suelen denominar por su término en inglés *cluster*, igualmente las técnicas de agrupamiento se suelen referir comúnmente como técnicas de *clustering*. Usaremos ambos términos en lo sucesivo en esta memoria. La aplicación de estas técnicas permite, entre otras aplicaciones, la predicción de flujos turísticos para anticipar la demanda, la segmentación o agrupamiento de turistas en base a características similares o la optimización de precios o tarifas de servicios turísticos.

Sin embargo, el crecimiento de los datos generados en el sector del turismo en la última década y la necesidad de realizar cada vez análisis más avanzados hacen necesario el uso de técnicas que van más allá de las capacidades convencionales de la inteligencia de negocio y la minería de datos. En este contexto, la aplicación de la inteligencia artificial emerge como una herramienta crítica y transformadora para analizar y aprovechar eficazmente las abundantes fuentes de datos disponibles en el sector turístico. Encontramos en la literatura muchos estudios (Kirtil & Askun, 2021; Li et al., 2018; Lv et al., 2021; Carvalho & Ivanov, 2023) donde quedan patente las múltiples aplicaciones de la inteligencia artificial. Por mencionar algunas de estas múltiples aplicaciones cabe destacar: la creación de asistentes virtuales capaces de interactuar de manera natural con los turistas, técnicas de procesamiento de lenguaje natural para interpretar reseñas y comentarios en línea, proporcionando percepciones sobre la valoración del turista y permitiendo a las empresas adaptar estrategias de marketing y servicios en tiempo real y las técnicas que facilitan la personalización de ofertas en función de las preferencias y comportamientos del turista.

Sin duda la aplicación de estas técnicas ayuda a la toma de decisiones en el sector del turismo, sin embargo, en ocasiones nos encontramos que la multiplicidad de factores que afectan a la actividad turística añade complejidad a estas decisiones. Por su naturaleza, tiene la industria turística una complejidad inherente derivada de los múltiples agentes públicos y privados que intervienen en la cadena de valor de la misma. Esta complejidad se refleja en los datos que provienen de diferentes orígenes y que muestran perspectivas distintas, pero complementarias. En este sentido, la toma de decisiones efectiva en el sector

del turismo implica considerar una multiplicidad de factores y criterios interrelacionados. En este contexto, la aplicación de los modelos MCDM emerge como una técnica esencial para abordar los desafíos particulares asociados con la gestión y la planificación en la industria turística. Estos modelos proporcionan un marco estructurado para evaluar y comparar alternativas en situaciones donde las decisiones no pueden reducirse a un único criterio. Un método específico dentro de los modelos MCDM es el proceso analítico jerárquico (AHP por sus siglas en inglés Analytic Hierarchy Process). EL AHP (Saaty, 1980) ayuda a tomar decisiones complejas al descomponer un problema en una estructura jerárquica de criterios y alternativas. Encontramos interesantes aplicaciones del AHP en el sector del turismo (Hoang et al., 2018; Maracajá et al. 2021). La integración de los modelos MCDM y en particular el AHP, con otras técnicas analíticas, como las de inteligencia artificial, proporciona una sólida estrategia para la toma de decisiones en entornos complejos. El proceso AHP puede recibir como entrada diversas fuentes de datos o los resultados generados por técnicas de inteligencia artificial y otras herramientas analíticas. De este modo, las decisiones derivadas del AHP se fundamentan completamente en datos y técnicas analíticas, siguiendo un proceso riguroso, consistente y estructurado que facilita la gestión de entornos de decisión complejos.

Podemos concluir, en base a lo expuesto, que la orientación al dato en la toma de decisiones implica la gestión de fuentes de datos de diferente naturaleza, así como la aplicación de variadas técnicas de análisis de datos y la integración de las mismas. En la Figura 1 vemos el esquema general del proceso de toma de decisiones basadas en datos en el sector del turismo. Toma como punto de partida el proceso de toma de decisiones basado en el conocimiento, la información y los datos en el contexto urbano de una ciudad inteligente definido por Bellini et al. (2021). Este esquema general es el que vamos a utilizar como punto de partida en esta tesis de cara a generar marcos de trabajo con datos en la toma de decisiones en el sector del turismo.

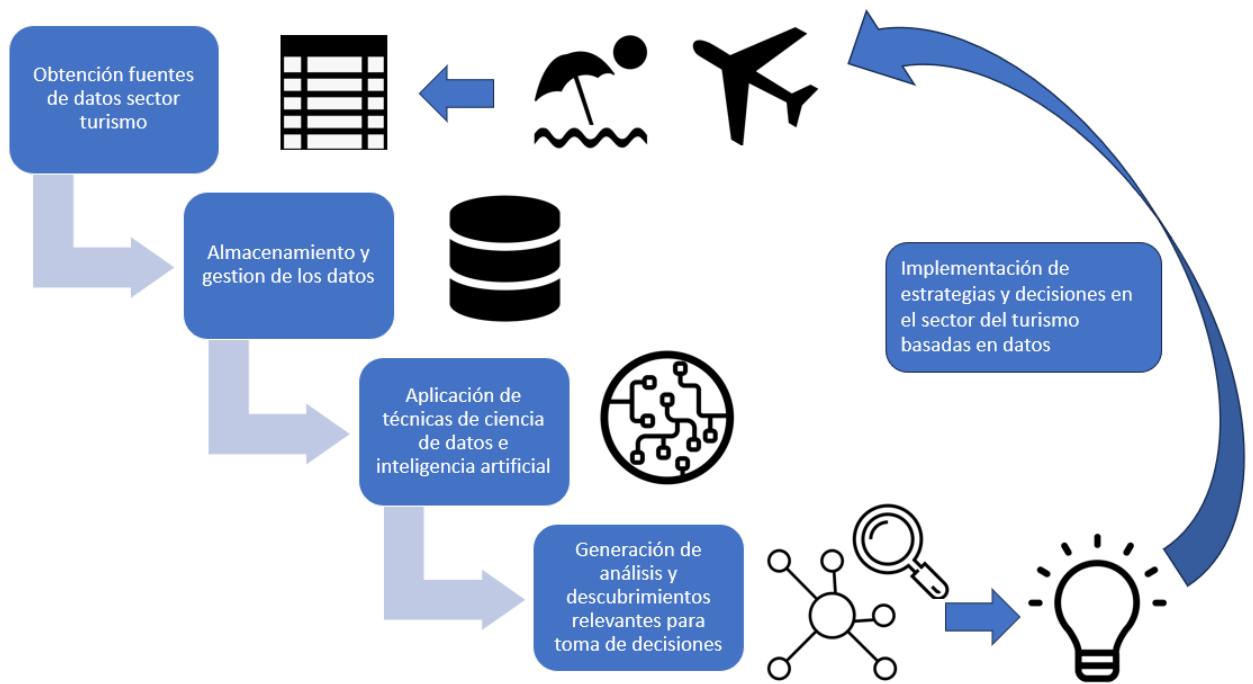


Figura 1: Esquema de toma de decisiones basadas en datos en el sector del turismo. Fuente: elaboración propia

Nos encontramos, por tanto, en un entorno complejo en lo relativo a la gestión y análisis de los datos y precisamos de un marco de trabajo robusto que facilite toda esta gestión y la toma de decisiones basadas en datos. Esta tesis pretende dar respuesta a esta necesidad de disponer de un marco de trabajo. Existen arquitecturas conceptuales genéricas que identifican las piezas necesarias para esta orientación al dato, en las que se integran perfectamente aspectos como la gestión de los datos o la inteligencia artificial (Moreno et al., 2019). Si bien estas arquitecturas conceptuales genéricas, plantean un marco conceptual válido, consideramos necesario adaptarlo a las necesidades concretas del sector del turismo en cuanto a la particularidad de sus fuentes de datos y las aplicaciones propias de este sector. Encontramos en la literatura trabajos en los que la arquitectura se adapta al sector del turismo, tales como el trabajo de Abdulaziz et al. (2015) que se centra en la capa de almacenamiento de datos y Bustamante et al. (2020) que plantea una capa de almacenamiento sobre la que añade una capa de inteligencia de negocio para la explotación de los datos. Ambas son aproximaciones válidas, pero no se pueden considerar un marco de trabajo completo que facilite la orientación al dato en el sector del turismo ya que no contienen todas las capas de la arquitectura que consideramos necesarias, ni todos los tipos de fuentes de datos presentes en el sector del turismo. Igualmente, cabe destacar el trabajo de Jansson et al. (2022) que si bien plantea una plataforma de datos más completa, no considera todas las posibles fuentes de datos y centra su trabajo en su aplicación en el sector público en destinos turísticos inteligentes, concretamente un caso de uso para la ciudad de Gotemburgo. Consideramos necesario plantear un marco conceptual de aplicación general para diferentes tipos de organizaciones turísticas y que contemple todas las posibles

fuentes de datos y contenga en su arquitectura todas las capas necesarias para trabajar en base al esquema planteado en la Figura 1. Este marco conceptual debe facilitar la orientación al dato y por tanto la orientación a mercado de las organizaciones del sector del turismo.

En esta tesis, para la obtención de este marco conceptual realizaremos varios pasos. En primer lugar, conocer el estado del arte relativo al uso de fuentes de datos en investigaciones en turismo basadas en datos, para ello se realizará un análisis bibliométrico que nos permitirá identificar los principales temas de investigación, así como las técnicas de análisis de datos utilizadas y las fuentes de datos empleadas. En segundo lugar, los resultados del estudio bibliométrico nos permitirán tener un marco de referencia inicial sobre el uso del dato en la toma de decisiones en organizaciones turísticas. Este marco de referencia inicial se estructurará en base a las áreas generales identificadas en el esquema de toma de decisiones basadas en datos en el sector del turismo mostrado en la Figura 1.

A partir de este marco analizaremos si se cubren adecuadamente las necesidades actuales de una organización turística orientada al dato. Este análisis nos permitirá identificar brechas, carencias o deficiencias que englobaremos bajo el término en inglés *gap* comúnmente utilizado en este tipo de análisis, término que utilizaremos en lo sucesivo en esta memoria. La identificación de estos gaps se traducirá en una serie de necesidades que se deben cubrir para poder abordar los desafíos actuales del sector del turismo con una estrategia orientada al dato. En tercer lugar, una vez definida la situación inicial y las necesidades detectadas, definiremos el marco conceptual para la toma de decisiones basadas en datos en una organización turística. De igual forma, este marco conceptual también se estructurará en base a las áreas identificadas en el esquema de toma de decisiones basadas mostrado en la Figura 1 para garantizar una correcta orientación al dato en la toma de decisiones. Partiendo de este marco conceptual se realizarán diferentes aplicaciones del mismo a problemas de negocio concretos del sector del turismo para demostrar su utilidad y validez.

1.1. Objetivos

En este apartado se describirá tanto el objetivo general como los objetivos específicos que nos hemos planteado en esta investigación.

1.1.1. Objetivo general

Diseñar un marco conceptual para la toma de decisiones basadas en datos en una organización turística. Esta propuesta propia de marco conceptual se pretende aplicar de forma práctica a diversas problemáticas dentro de este sector.

1.1.2. Objetivos específicos

Además, se han determinado varios objetivos más concretos que quieren complementar el objetivo principal de esta tesis. Estos objetivos son los siguientes:

- **Objetivo 1:** para la consecución del objetivo general, necesitaremos en primer lugar y como objetivo más específico conocer la situación actual de las organizaciones turísticas respecto al uso de los datos en la toma de decisiones, es decir, disponer de un marco de referencia inicial que sirva de punto de partida. Para ello se realizará un análisis bibliométrico sobre el uso de fuentes de datos en investigaciones en turismo basadas en datos que nos permita definir este marco de referencia inicial sobre el uso del dato en la toma de decisiones en organizaciones turísticas.
- **Objetivo 2:** a partir de este marco de referencia inicial, se analizarán las carencias que este marco podría tener de cara a abordar los desafíos actuales del sector del turismo con una estrategia orientada al dato. Estas carencias, brechas o necesidades se traducirán en una serie de gaps que deberán ser tenidos en cuenta para definir un marco conceptual que permita abordar de forma eficiente estrategias orientadas en datos en las organizaciones turísticas. Este marco conceptual será definido en detalle y servirá de base para cualquier estrategia de negocio orientada al dato en una organización turística.
- **Objetivo 3:** una vez definido el marco conceptual se aplicará a diferentes casos de uso en el sector del turismo:
 - **Analizar la experiencia del turista para mejorar la gestión de destinos.** Proponer un modelo para detectar y analizar las experiencias de los turistas en un determinado destino turístico a partir de las reseñas realizadas en plataformas digitales. El modelo pretende ser de gran ayuda para que las organizaciones de gestión de destinos conozcan en profundidad las experiencias de los turistas en un destino turístico y tomen decisiones de mejora basadas en este conocimiento.
 - **Apoyo a la toma de decisiones en la estrategia de digitalización de tienda física en agencias de viajes.** Desarrollar un modelo MCDM para tiendas físicas del sector turístico que quieren digitalizarse, proporcionando un enfoque sistemático para gestionar las prioridades de inversión que involucre a todas las áreas de la organización. Plantear una aplicación en las

tiendas físicas de agencias de viajes en el que se proponga un ranking de prioridades en la adopción de dispositivos digitales.

- **Posicionamiento óptimo de servicios turísticos inteligentes.** Proponer una plataforma de datos e inteligencia artificial, a partir de la cual se desarrolla un modelo de decisión para determinar el posicionamiento óptimo de los servicios turísticos inteligentes que genere recomendaciones personalizadas. Desarrollar una aplicación para el posicionamiento de oficinas de turismo en la ciudad de Madrid, donde se propongan nuevas ubicaciones de oficinas de turismo para la ciudad, con recomendaciones personalizadas sobre su gestión.

1.2. Estructura de la memoria

Esta tesis doctoral está estructurada en dos grandes apartados. Una primera parte en la que se realiza el planteamiento del problema y estudio preliminar y una segunda parte donde se obtienen los resultados y conclusiones de la memoria.

En la primera parte se realiza la introducción, el estado de arte y se presenta el marco teórico de investigación. Se compone de los siguientes capítulos:

El capítulo 1 inicialmente introduce a la toma de decisiones basadas en datos en el sector del turismo. Posteriormente revisa el uso de fuentes de datos en turismo, así como la aplicación de técnicas de ciencia de datos, inteligencia artificial y modelos MCDM. Se propone un esquema de toma de decisiones basadas en datos para el sector del turismo y finalmente se plantea la necesidad de disponer de marco conceptual de trabajo para organizaciones de turísticas basadas en datos.

En el capítulo 2 se realiza un análisis bibliométrico sobre el uso de fuentes de datos en investigaciones en turismo basadas en datos. Este análisis bibliométrico nos permitirá identificar los principales temas de investigación, así como las técnicas de análisis de datos utilizadas y las fuentes de datos empleadas. Finalmente, partiendo de los resultados del estudio bibliométrico se obtiene un marco de referencia inicial sobre el uso del dato en la toma de decisiones en organizaciones turísticas.

En el capítulo 3 se presenta todo el marco teórico necesario para el desarrollo de la investigación. Se presentan todos los fundamentos para el planteamiento del marco conceptual y el desarrollo de los diferentes casos de uso.

La segunda parte de la memoria se centra en los resultados obtenidos, proponiendo un marco conceptual para la toma de decisiones basadas en datos en una organización turística. A partir de este marco conceptual se plantean tres casos de uso en el sector del turismo y finalmente se muestran las conclusiones. Se compone de los siguientes capítulos:

En el capítulo 4 partiendo del marco de referencia inicial obtenido en el capítulo 2 analizaremos si se cubren adecuadamente las necesidades actuales de una organización turística orientada al dato. Una vez definida la situación de partida en el marco de referencia inicial e identificadas las necesidades detectadas en el mismo, definiremos el marco conceptual para la toma de decisiones basadas en datos en una organización turística.

En los capítulos 5, 6 y 7 se presentan diferentes casos de uso en los que se aplica el marco conceptual propuesto. Los casos de uso planteados tienen cada uno de ellos objetivos de negocio diferentes en el ámbito del turismo y son abordados desde una perspectiva orientada al dato. Por otro lado, cada uno de los casos de uso utiliza fuentes de datos o técnicas analíticas diferentes para la toma de decisiones que se integran en el marco conceptual propuesto. Los casos de uso son los siguientes:

- Analizar la experiencia del turista para mejorar la gestión de destinos.
- Apoyo a la toma de decisiones en la estrategia de digitalización de tienda física en el sector del turismo.
- Posicionamiento óptimo de servicios turísticos inteligentes.

El capítulo 8 presenta las conclusiones y trabajo futuro. En primer lugar, se revisa la consecución del objetivo general y los objetivos específicos de la tesis. Se explica la utilidad y los beneficios del marco conceptual propuesto y como el mismo facilita la toma de decisiones basadas en datos en organizaciones turísticas. De forma práctica se revisan los beneficios del uso del marco conceptual propuesto en cada uno de los casos de uso planteados. En segundo lugar, en este capítulo en el ámbito del trabajo futuro, se proponen líneas de actuación directamente relacionadas con la revolución que tenemos actualmente debido a los avances de la inteligencia artificial con los cambios y tendencias actuales en la actividad turística

Se incluye en la memoria un Anexo que contiene una breve referencia de cada uno de los artículos producidos durante la realización de la tesis indicando el estado del artículo (publicado o en revisión) y datos adicionales sobre la publicación o el congreso donde se ha defendido. Igualmente se incluyen en este Anexo menciones y reconocimientos académicos obtenidas en el periodo de realización de esta tesis.

En la Figura 2 se muestra la estructura completa de la memoria, proporcionando una visión general de su contenido:

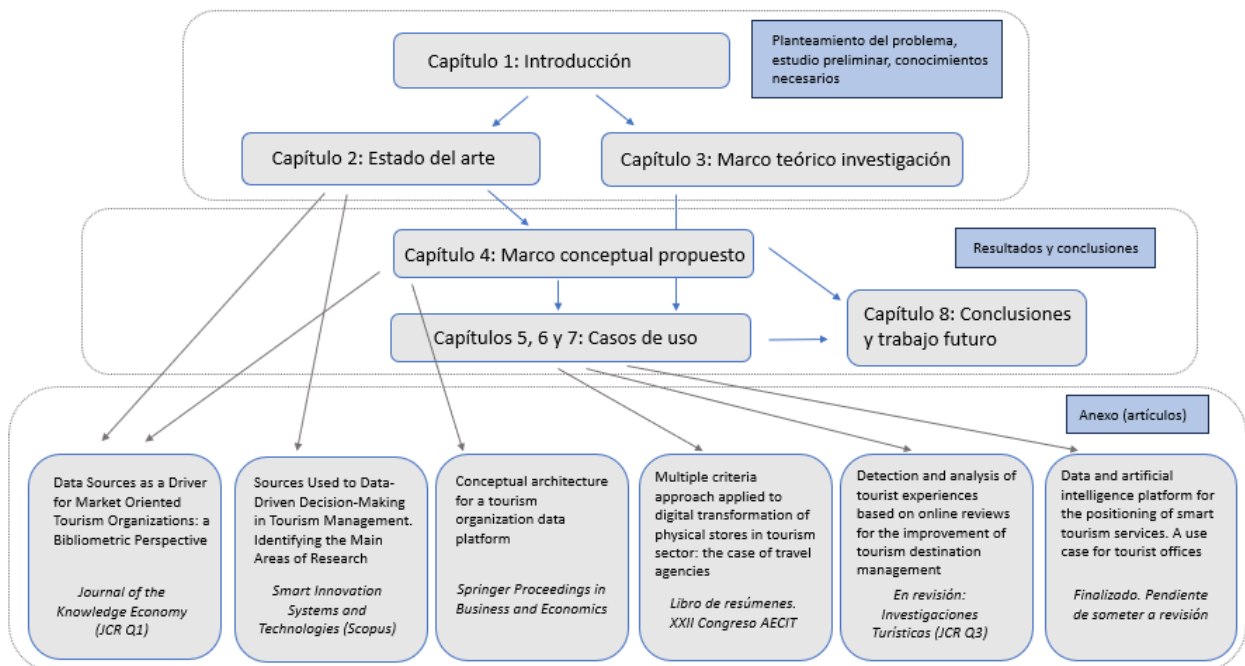


Figura 2: Estructura de la memoria

En la Figura 2 se muestra mediante flechas en azul las conexiones entre los diferentes capítulos desde la primera parte en la que se realiza el planteamiento del problema y estudio preliminar hasta la segunda parte donde se obtienen los resultados y conclusiones de la memoria. Cada uno de los artículos científicos (publicados o en revisión) mostrados en el Anexo son resultados de las investigaciones realizadas en los diferentes capítulos de la tesis tal y como se observa en la Figura 2 con las conexiones mostradas con flechas grises.

2. Estado del arte

Tal y como se ha indicado anteriormente, en este capítulo vamos a realizar un análisis bibliométrico sobre el uso de fuentes de datos en investigaciones turísticas basadas en datos. La bibliometría es una herramienta útil para evaluar y analizar los resultados de la investigación académica. Para explorar un determinado campo de investigación se suelen utilizar dos métodos principales: el análisis del rendimiento y los mapas científicos (Noyons et al., 1999; Van-Raan, 2004). El objetivo del análisis del rendimiento es evaluar el impacto de la producción científica de los distintos agentes científicos. En cambio, los mapas científicos pretenden mostrar la estructura conceptual, social o intelectual de la investigación científica, así como su evolución y sus aspectos dinámicos.

Este análisis bibliométrico va a abarcar un periodo temporal desde 1982 hasta 2021. Utilizaremos los resultados del análisis bibliométrico para identificar los principales temas de investigación de la comunidad científica y obtener percepciones y conclusiones sobre los temas de investigación turística, las técnicas de análisis de datos utilizadas y las fuentes de información empleadas. Estos conocimientos permitirán obtener un marco de referencia inicial sobre el uso del dato en la toma de decisiones en organizaciones turísticas que se estructurará en base a las áreas identificadas en el esquema de toma de decisiones basadas en datos en el sector del turismo mostrado en la Figura 1.

2.1. Trabajos relacionados

Algunos autores han realizado una revisión de los estudios bibliométricos en el ámbito del turismo (Koseoglu et al., 2016) que concluye que existen muchos más estudios de revisión sistemática de la literatura que estudios bibliométricos, señalando también que el número de análisis con técnicas bibliométricas relacionales es relativamente bajo.

Considerando la literatura disponible sobre estudios bibliométricos, encontramos que varios autores han realizado estudios de este tipo en áreas temáticas relacionadas con el turismo desde diferentes puntos de vista.

En primer lugar, encontramos autores que han realizado estudios generales sobre la investigación turística desde diferentes perspectivas, algunos autores se han centrado en analizar la calidad de las revistas (Hall ,2011), otros autores se centran en analizar la contribución de los autores y las relaciones entre autores en base a las citas (Benckendorff

& Zehrer, 2013) y otros autores (Mulet-Forteza et al., 2019) añaden a las perspectivas anteriores un análisis de los principales temas de investigación.

En segundo lugar, encontramos autores que han realizado estudios sobre áreas temáticas específicas de relevancia en el sector turístico como la innovación turística (Omerzel ,2016), el turismo sostenible (Ruhanen, 2015; Martínez-Martínez et al., 2022), la promoción turística (Florido-Benítez, 2022), el turismo médico (de la Hoz-Correa, et al., 2018), impacto económico (Comerio & Strozzi, 2019), turismo rural (Ruiz-Real et al., 2020), turismo de aventura (Cheng et al., 2018), o turismo gastronómico (Okumus, 2018), entre otros.

En tercer lugar, nos centramos en los estudios bibliométricos que se centran en la investigación basada en datos. Es decir, investigaciones en el ámbito del turismo con un enfoque basado en datos y técnicas analíticas. Es en este tipo de estudios en los que se centrará el trabajo que proponemos. Cabe destacar que también existen revisiones de literatura (Li et al., 2018; Mariani et al., 2018; Rahmadian et al., 2022; Samara et al., 2020; Li et al., 2022; Nyanga et al., 2020; Mariani & Baggio, 2020) en este campo. En esta línea encontramos bibliométricos centrados en el uso de la inteligencia artificial en el turismo (Kirtil & Askun, 2021; Sharma et al., 2022). En estos análisis bibliométricos se utiliza la base de datos Scopus e identifican los autores e instituciones más productivos, así como los temas de investigación más relevantes. Se muestra la evolución de las publicaciones a lo largo de los periodos analizados. Proponemos en nuestro estudio una perspectiva diferente que abarca toda la investigación turística basada en fuentes de datos y no sólo la que utiliza técnicas de inteligencia artificial, analizando tanto las técnicas como las fuentes de datos y los temas de investigación, utilizando Web of Science (WoS) como base de datos. También hay autores (Lv et al., 2021) que analizan el uso de la inteligencia artificial añadiendo el uso de arquitecturas y sistemas orientados a la gestión de volúmenes de datos muy elevados. Este tipo de sistemas y arquitecturas se le suelen denominar por su término en inglés *big data* que utilizaremos en lo sucesivo en esta memoria para referirnos a ellas. Este estudio muestra la evolución temporal de las publicaciones, las principales palabras clave y la distribución de las publicaciones por revistas. También identifica y clasifica las principales fuentes de datos y los principales temas de investigación. Así mismo, algunos autores (Gómez & Gil, 2020) realizaron un estudio bibliométrico sobre la gestión del conocimiento y la creación de valor a través del big data en el sector turístico. A diferencia del artículo anterior, proponemos en nuestro trabajo una perspectiva diferente que abarca todas las investigaciones en turismo basadas en fuentes de datos y no sólo aquellas que utilizan técnicas de inteligencia artificial. En cuanto al estudio bibliométrico, incluimos técnicas como los mapas científicos y la evolución conceptual basada en ellos, que son técnicas no

abordadas en el artículo mencionado. Por otro lado, también existen estudios bibliométricos que analizan exclusivamente un tipo específico de fuente de datos, como los estudios centrados en datos procedentes de redes sociales (Nusair et al., 2019), en datos generados por usuarios en plataformas online (Muritala et al., 2020; Mukhopadhyay et al., 2022; Akbari et al., 2022) o en datos generados por dispositivos móviles (Chantre-Astaiza, 2019). Los estudios de este tipo analizan exclusivamente un tipo de fuente de datos; en el estudio que proponemos analizamos los diferentes tipos de fuentes de datos utilizados en la investigación turística basada en datos.

La Tabla 1 muestra un resumen de los artículos mencionados en el tercer párrafo. Se diferencia entre revisión de literatura como resumen narrativo de investigaciones previas sobre un tema para identificar tendencias y brechas, pero sin una metodología estricta, una revisión sistemática de literatura que sigue una metodología rigurosa y estructurada para identificar, evaluar y sintetizar la evidencia relevante sobre una pregunta específica y el análisis bibliométrico que usa métricas cuantitativas para analizar el impacto, las citas y las tendencias en la producción científica.

Referencia	Objetivo	Tipo de revisión
Li et al., 2018	Big data en turismo	Revisión de literatura
Mariani et al., 2018	Inteligencia de negocio y big data en hospitalidad y turismo	Revisión sistemática de literatura
Rahmadian et al., 2022	Uso del big data para turismo sostenible	Revisión sistemática de literatura
Samara et al., 2020	Inteligencia artificial y big data en turismo	Revisión sistemática de literatura
Li et al., 2022	Big data en la investigación del turismo en China	Revisión sistemática de literatura
Nyanga et al., 2020	Mejorar la competitividad del sector turístico mediante el uso de la inteligencia empresarial	Revisión de literatura
Mariani & Baggio, 2020	Big data y analítica en hospitalidad y turismo	Revisión sistemática de literatura
Kirtil & Askun, 2021	Inteligencia artificial en turismo	Bibliométrico
Sharma et al., 2022	Inteligencia artificial en turismo	Bibliométrico
Lv et al., 2021	Literatura sobre big data e inteligencia artificial en hostelería y turismo	Bibliométrico
Gómez & Gil, 2020	Gestión del conocimiento y creación de valor a través del big data en el sector turístico	Bibliométrico
Nusair et al., 2019	Las redes sociales en la investigación sobre hostelería y turismo	Bibliométrico

Muritala et al., 2020	Investigación sobre reseñas en línea en turismo y hostelería	Bibliométrico
Mukhopadhyay et al., 2022;	Investigación basada en eWOM (electronic word of mouth)	Bibliométrico
Akbari et al., 2022	La evolución de la eWOM del pasado al futuro	Bibliométrico
Chantre-Astaiza, 2019	Movilidad turística. Datos sobre dispositivos móviles	Bibliométrico
Nuestra propuesta	Uso de las fuentes de datos en las investigaciones en turismo basadas en datos	Bibliométrico

Tabla 1: Resumen de revisiones de literatura o estudios bibliométricos centrados en la investigación basada en datos

Tras revisar los artículos relacionados, no hemos encontrado ningún trabajo previo que haya intentado ofrecer una visión general del panorama actual de las organizaciones turísticas basadas en datos, definiendo un marco de referencia sobre el uso de datos en la toma de decisiones. Además, el estudio propuesto difiere de los más relacionados, ya mencionados, por varias razones: amplio y actualizado marco temporal (1982-2021) de los artículos WoS. También difiere en el análisis de los temas de investigación y su evolución conceptual a lo largo del periodo, todo ello utilizando las herramientas de mapeo de la ciencia proporcionadas por SciMAT (Science Mapping Analysis software Tool) (Cobo et al., 2012), lo que permite una descripción más profunda de los temas. Se propone un enfoque amplio de la investigación turística basado en las fuentes de datos, analizando no solo los temas de investigación, sino también las fuentes de datos utilizadas y las técnicas de análisis de datos empleadas.

2.2. Metodología

La metodología utilizada en este trabajo está inspirada en Cobo et al. (2011), Carrasco-Aguilar et al. (2022), Galán et al. (2022) y Minhas & Sindakis (2021). La Figura 3 muestra esquemáticamente los pasos a realizar.

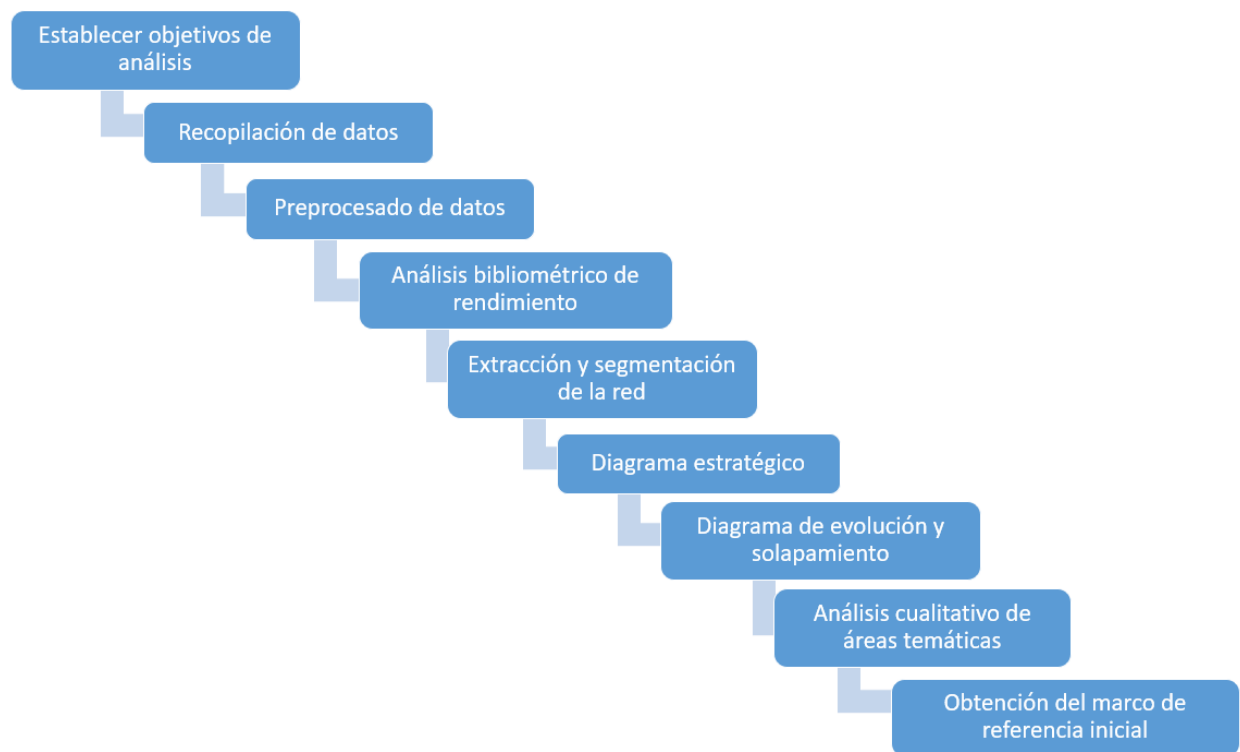


Figura 3: Principales pasos de la metodología empleada para la realización del análisis bibliométrico. Fuente: elaboración propia.

Establecer los objetivos de análisis

Como primer paso de la metodología, es necesario establecer un objetivo. Tal y como se ha mencionado anteriormente, el objetivo es realizar un análisis bibliométrico sobre el uso de fuentes de datos en investigaciones turísticas basadas en datos. Este bibliométrico nos permitirá definir un marco de referencia inicial sobre el uso del dato en la toma de decisiones en organizaciones turísticas.

Recopilación de datos

Los artículos científicos objeto de estudio se descargaron de WoS. Para ello, se utilizó la búsqueda avanzada con la siguiente ecuación de búsqueda:

$$TS = (("tourism" OR "tourist*" OR "hospitality" OR "hotel*") NEAR/3 ("data source*" OR "information source*" OR "dataset*" OR "information system*" OR "database*" OR "data warehouse*" OR "data lake*"))$$

Se obtuvieron un total de 794 documentos en el periodo de tiempo: 1982-2021. Los datos fueron descargados el 21 de septiembre de 2022. En esta investigación para realizar

el análisis bibliométrico hemos utilizado el software SciMAT (Cobo et al., 2012) ya que integra todo lo que requerimos para esta investigación y por su carácter gratuito y de código abierto.

Preprocesado de datos

Los documentos se descargaron de WoS en un fichero de texto sin formato incluyendo el registro completo. Este fichero se cargó en SciMAT para construir la base de datos a partir de la cual se realizará el análisis bibliométrico.

Para mejorar la calidad de los datos, se ha realizado un proceso de depuración. Se han normalizado las palabras clave fusionándolas en sus formas plural y singular, también se han fusionado las palabras con sus correspondientes sinónimos, y se han identificado varias palabras clave que se refieren al mismo concepto utilizando la distancia de Levenshtein en SciMAT.

A continuación, utilizando el gestor de periodos de SciMAT, se establecieron dos periodos de tiempo consecutivos para mostrar la evolución conceptual en el análisis de los mapas científicos. Realizando un análisis previo de la distribución de las publicaciones en el tiempo, se decidió dividir el marco temporal original (1982-2021) en dos periodos asegurando que ambos periodos tuvieran un número similar de documentos. De esta manera, los datos se dividieron en dos periodos temporales consecutivos 1982-2015 con 393 documentos y 2016-2021 con 401 documentos.

Análisis bibliométrico de rendimiento

En esta fase se ha realizado el análisis bibliométrico de rendimiento, en el que se analizan las publicaciones desde múltiples perspectivas: evolución en el tiempo del número de publicaciones, revistas con mayor número de publicaciones, distribución geográfica de las publicaciones y autores más productivos. Los gráficos obtenidos en este análisis bibliométrico se realizan con el lenguaje de programación Python y el paquete Matplotlib (Matplotlib, 2022), utilizando los datos de SciMAT.

Extracción y segmentación de la red

A continuación, se generaron los mapas científicos, que son representaciones espaciales de las relaciones que existen entre documentos, autores, campos o disciplinas. Se han utilizado ampliamente en la literatura científica en campos muy diversos para mostrar estas relaciones de forma gráfica (Huang & Chang, 2014; Liu et al., 2012).

Aplicamos un análisis de co-ocurrencias (Callon et al., 1983) basado en las palabras clave de los documentos de cada periodo. La co-ocurrencia C_{ij} es definida aquí como el número de documentos en las que dos palabras claves i y j aparecen juntas

A continuación, realizamos una agrupación de palabras clave mediante el algoritmo de centro simple (Coulter et al., 1998), que localiza redes de palabras clave estrechamente relacionadas entre sí y que corresponden a centros de interés o problemas de investigación de gran interés entre los investigadores. Esta similitud entre palabras clave se evalúa mediante el índice de equivalencia (Coulter et al., 1998): $e_{ij} = C_{ij}^2 / C_i C_j$ donde C_{ij} es el número de documentos en los que coexisten dos palabras clave i y j , y C_i y C_j representan el número de documentos en los que aparece cada una.

Diagrama estratégico

En esta fase, mediante el diagrama estratégico (He, 1999) y la red temática, se visualizan los temas más destacados. Cada uno de estos temas se caracteriza por dos medidas (Coulter et al., 1998): centralidad y densidad. La centralidad mide el grado de interacción de una red con otras redes, definida como: $c = 10 * \sum e_{kh}$, donde k es una palabra clave perteneciente a ese mismo tema y h una palabra clave perteneciente a otros temas.

Por otro lado, la densidad mide la intensidad interna de la red y puede definirse como $d = 100 * (\sum e_{ij} / w)$, donde i y j son palabras clave pertenecientes al tema, y w el número de palabras clave del tema. Con estas dos medidas, podemos visualizar los temas de investigación mediante un diagrama estratégico bidimensional (Figura 4) en el que los temas de investigación se clasifican en cuatro grupos:

- a) En el cuadrante superior derecho están los temas que tienen una fuerte centralidad y una alta densidad; son, por tanto, los temas impulsores de nuestro campo de investigación. Están bien desarrollados y son importantes para la estructura del campo de investigación.
- b) En el cuadrante superior izquierdo se sitúan los temas altamente especializados pero periféricos. Se trata de temas de importancia marginal para el campo, ya que están bien desarrollados internamente, pero sus vínculos con otros temas externos son muy débiles.
- c) En el cuadrante inferior izquierdo están representados los temas emergentes o en vías de desaparición, ya que tienen una baja densidad y centralidad. Son temas poco desarrollados y marginales.

d) En el cuadrante inferior derecho están representados los temas transversales y generales, es decir, básicos, relevantes para el campo de investigación, pero poco desarrollados.

Diagrama estratégico sobre campos o temas de investigación

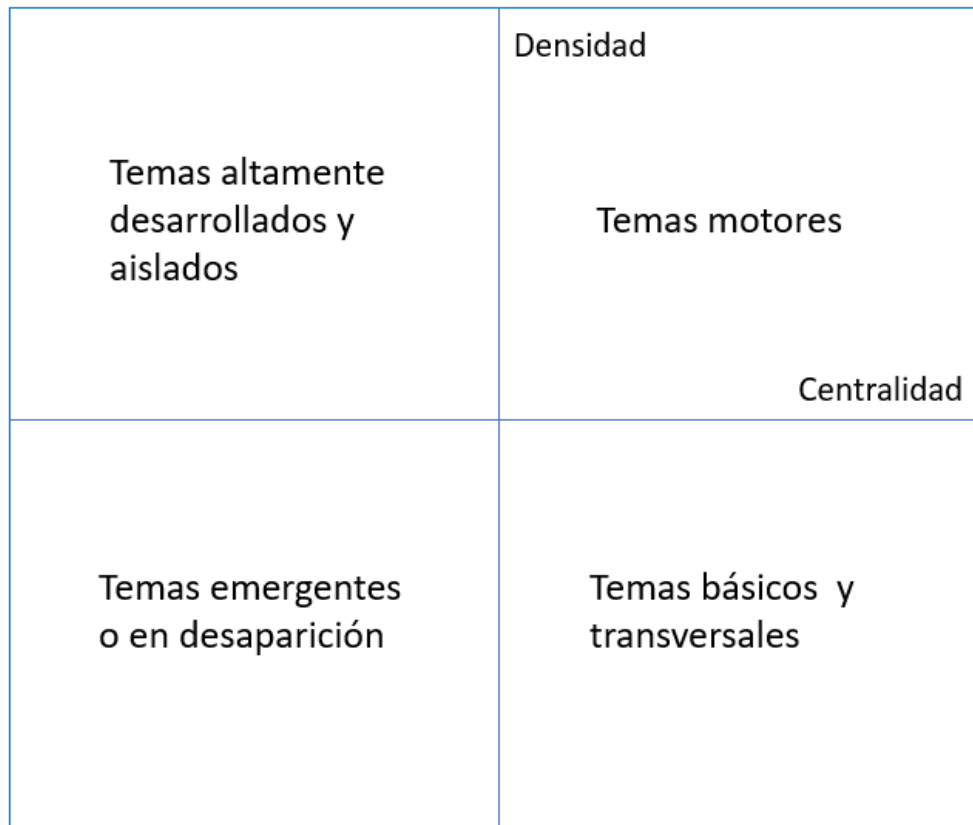


Figura 4: Diagrama estratégico. Fuente: Cobo et al. (2012)

El análisis de rendimiento se realiza como complemento al trabajo de cartografía científica, para lo cual se utilizan los siguientes indicadores bibliométricos: número de artículos publicados, número de citas e índice H (Alonso et al., 2009; Hirsch, 2005). Para evaluar estos indicadores de rendimiento, el programa SciMAT, mediante una función, asigna un conjunto de documentos a cada tema detectado; en concreto, SciMAT devuelve la unión algebraica del conjunto de documentos asociados a las palabras clave del tema. Un documento tiene varias palabras clave, cada una de ellas puede estar asociada a un tema diferente, por lo que un documento podría estar asociado a varios temas.

Una vez obtenido el diagrama estratégico, es interesante identificar las principales áreas temáticas en las que podemos clasificar los temas obtenidos. Esta clasificación en áreas temáticas nos permitirá comprender mejor los resultados bibliométricos.

Diagrama de evolución y solapamiento

En esta fase se detectan los temas de investigación a lo largo de un periodo de tiempo, tras lo cual se analizan las áreas generales de evolución, sus orígenes e interrelaciones. Para ello, se utiliza el índice de inclusión (Sternitzke & Bergman, 2009), que detecta vínculos conceptuales entre temas de investigación de diferentes periodos. Posteriormente, se elabora un mapa bibliométrico con la evolución temática en dos periodos de tiempo, los temas pueden estar desvinculados, vinculados por líneas continuas o líneas discontinuas. Además, dependiendo de si el valor del índice de inclusión es mayor o menor, así será el grosor de las líneas que unen los temas enlazados (Cobo et al., 2015).

Análisis cualitativo de áreas temáticas

Además de las medidas de rendimiento y los mapas de evolución obtenidos, merece la pena realizar un análisis cualitativo de las áreas temáticas identificadas a partir de las principales palabras clave de las mismas.

Obtención del marco de referencia inicial

Una vez realizado el análisis bibliométrico e identificadas y analizadas las principales áreas temáticas, utilizaremos estos conocimientos para obtener el marco de referencia inicial en base a las áreas identificadas en el esquema de toma de decisiones basadas en datos en el sector del turismo mostrado en la Figura 1.

2.3. Resultados

En este apartado, tras seguir los pasos de la metodología expuesta en el apartado anterior, se obtienen los indicadores básicos de rendimiento bibliométrico y el diagrama estratégico. Se identifican los principales temas de interés para la investigación y se obtiene la evolución conceptual. Se identifican tres grandes áreas temáticas y se realiza un análisis cualitativo de las mismas. En base a estas áreas temáticas se obtiene el marco de referencia inicial.

2.3.1. Análisis bibliométrico de rendimiento

En primer lugar, revisamos la evolución temporal de las publicaciones. En la Figura 5 se muestra la distribución de los documentos por años. La gráfica muestra siete periodos de 5 años, excepto el periodo inicial que abarca 8 años debido al escaso número de documentos en esos años y el último periodo que se extiende hasta 2021 y abarca 6 años. Como puede observarse, el número de publicaciones ha ido aumentando a lo largo de los periodos. Los dos primeros periodos tienen un número bajo de documentos, pero se observa un aumento entre el primer y el segundo periodo. El número de documentos se mantiene estable en el tercer y cuarto periodo y vuelve a aumentar considerablemente en los tres últimos periodos. Podemos ver que entre el quinto y el séptimo periodo el número de documentos se cuadruplica.

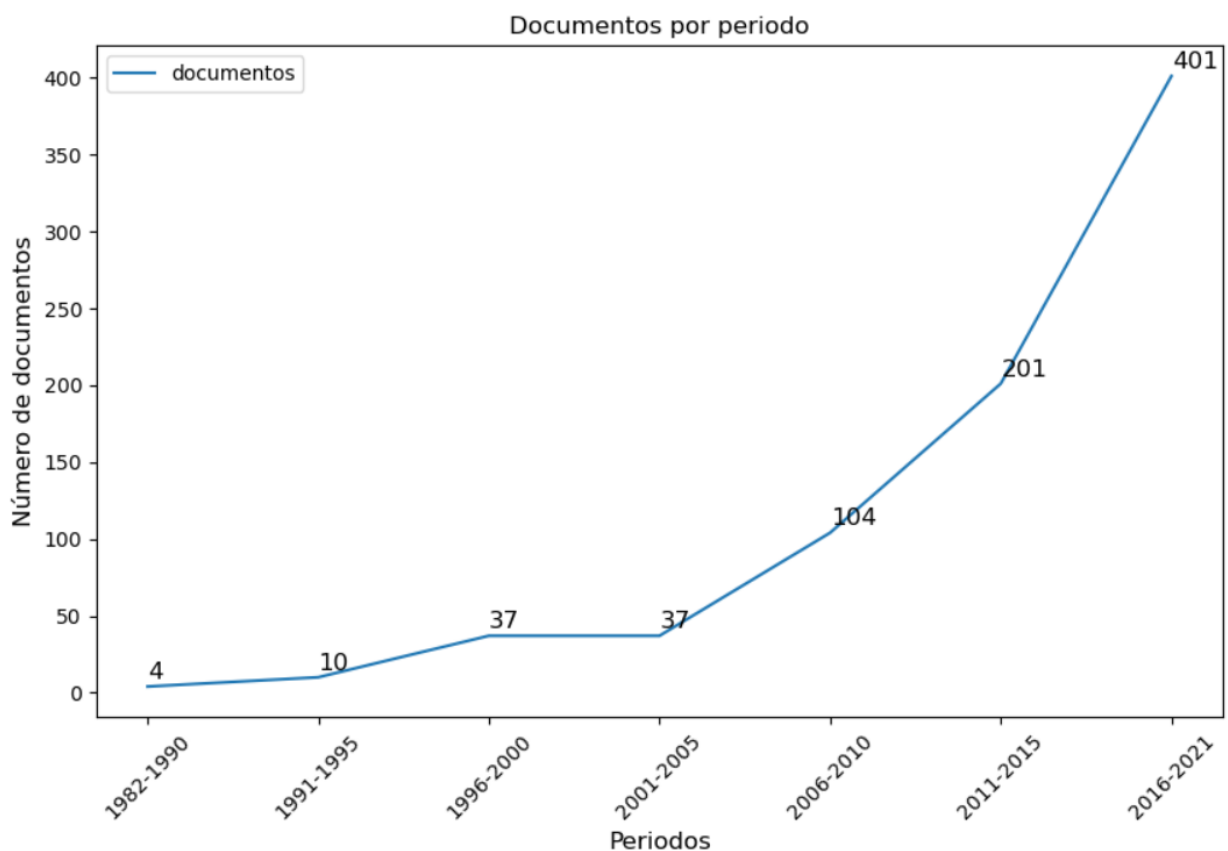


Figura 5: Evolución temporal del número de documentos por periodo. Fuente: elaborado con SciMAT

Si nos centramos en la distribución de publicaciones por revistas, la Figura 6 muestra un gráfico de barras con las revistas con más de 10 publicaciones. Destaca la revista *Tourism Management* con 29 publicaciones.

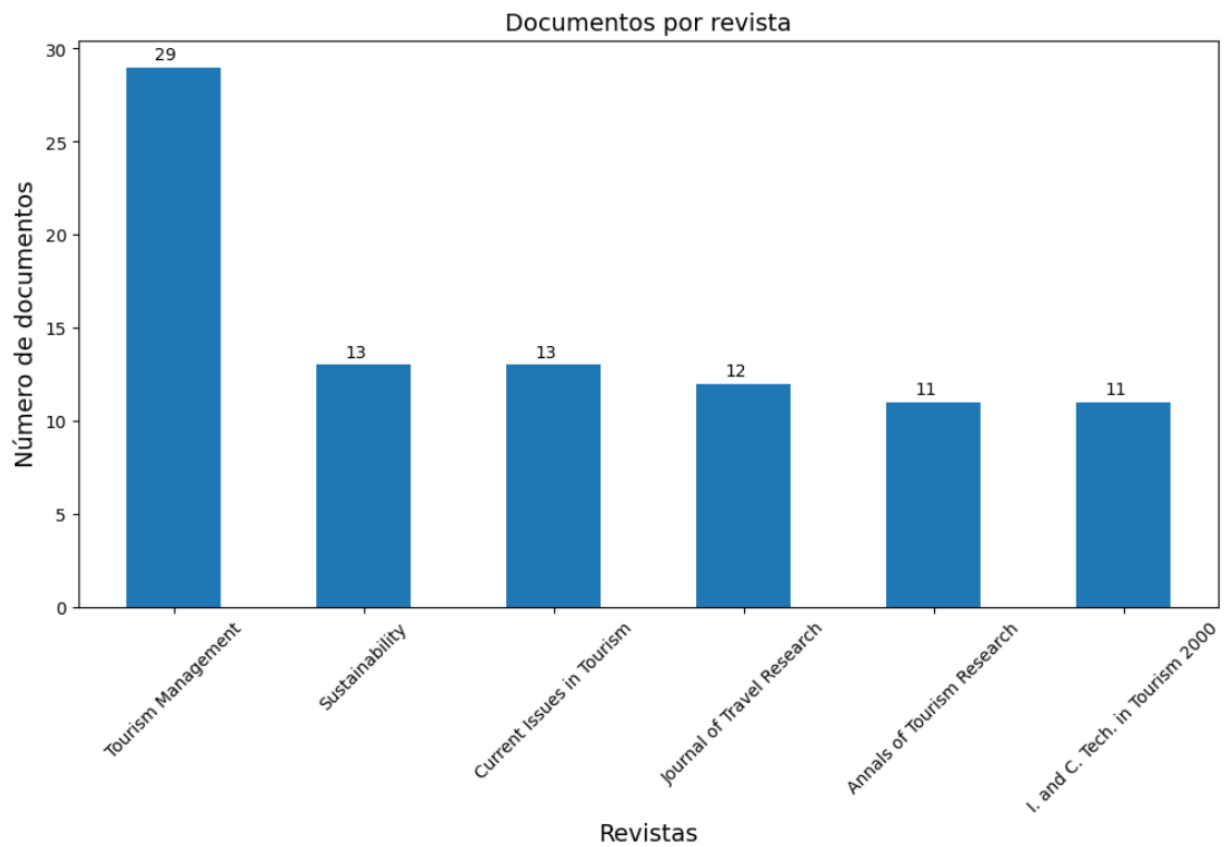


Figura 6: Gráfico de barras que contiene las revistas con más de 10 publicaciones. Fuente: elaborado con SciMAT

En cuanto a la distribución de documentos por países, la Figura 7 muestra un mapa del mundo con los países con más de 10 publicaciones. China es el líder en número de documentos publicados, seguida de España y Estados Unidos



Figura 7: Mapa del mundo con los países con más de 10 publicaciones. Fuente: elaborado con SciMAT

En cuanto a los autores, se observa que la distribución es muy homogénea. No encontramos ningún autor con un número muy elevado de publicaciones. Respecto a los autores más productivos, Proell, Birgit, seguida de Law, Rob y Retschitzegger, Werner. Proell y Retschitzegger tienen publicaciones conjuntas en la Universidad Johannes Kepler de Linz, mientras que Law está afiliado a la Universidad de Macao.

2.3.2. Diagrama estratégico

En este apartado obtenemos las representaciones espaciales de las relaciones que existen entre documentos siguiendo la metodología de Cobo et al. (2011) basada en un análisis de co-ocurrencias descrito en el apartado 2.2.

El análisis se basa en dos periodos (1982-2015 y 2016-2021) y las descripciones se apoyarán en tablas y diagramas estratégicos. En los diagramas estratégicos, el tamaño de la esfera de cada tema de investigación es proporcional al número de artículos publicados sobre ese tema.

Primer periodo (1982-2015)

Como muestra la Figura 8, este primer periodo contiene 13 temas de investigación: 'performance', 'customer-satisfaction', 'tourism-information-system', 'tourism', 'travel-

agencies', 'geosocial-networks', 'perceived-usefulness', 'disability', 'business', 'clustering', 'strategies', 'environment' and 'perceived-risk'.

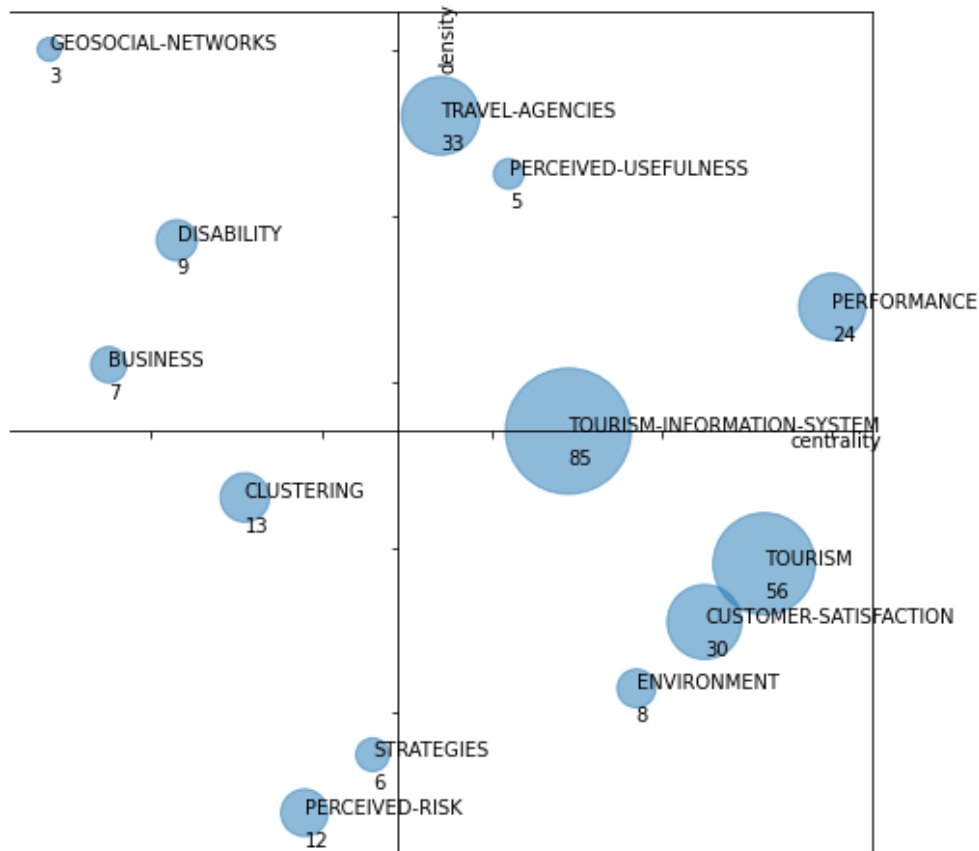


Figura 8: Diagrama estratégico para el periodo 1982-2015. Se muestra el número de documentos de cada tema. Fuente: elaborado con SciMAT

La Tabla 2 muestra los indicadores de rendimiento: número de documentos, citas obtenidas por estos documentos y el índice H. De acuerdo con esos indicadores, pudimos identificar como temas principales los siguientes: 'tourism-information-system', 'tourism', 'travel-agencies', 'customer-satisfaction' and 'performance'. Entre estos 5 temas suman el 74% del número total de citas y el 78% de los documentos. El índice H de estos temas también es superior, aunque no hay una diferencia tan notable como en el número de citas o documentos.

Temas	Número de documentos	Índice H	Citas
Tourism-information-system	85	10	449
Tourism	56	16	891
Travel-agencies	33	11	904
Customer-satisfaction	30	19	1295

Performance	24	16	1012
Clustering	13	6	215
Perceived-risk	12	8	266
Disability	9	6	98
Environment	8	5	142
Business	7	4	263
Strategies	6	6	199
Perceived-usefulness	5	5	362
Geosocial-networks	3	1	28

Tabla 2: Indicadores de rendimiento por temas en el periodo 1982-2015

Estos temas principales se describen a continuación.

El tema 'tourism-information-system' es un tema impulsor; es el tema con mayor número de documentos, aunque tiene menos peso en términos de citas e índice H. Se centra en las fuentes de información utilizadas en turismo, con énfasis en las fuentes de datos de información geográfica y, en menor medida y con mucho menos peso que el anterior, en los sistemas de información estadística. También incluye diversos temas que hacen referencia a cuestiones tecnológicas relacionadas con las fuentes de información, como la web semántica, las ontologías y los procesos de integración.

Otro tema impulsor destacado es 'performance'; se trata del segundo tema con mayor número de citas. Se centra en cuestiones relacionadas con la medición de la competitividad y el impacto desde diferentes perspectivas de la industria turística, destacando temas como la calidad del servicio y, en menor medida, la productividad o las estrategias. La competitividad y la calidad del servicio de los destinos turísticos son uno de los principales objetos de estudio en los análisis basados en datos del sector turístico.

El tercer tema impulsor es 'travel-agencies'; es el tercer tema con mayor número de citas. Incluye temas relacionados con las agencias de viajes, internet o el uso de imágenes de destinos turísticos, lo que refleja la creciente importancia de internet en el turismo, así como el uso de imágenes de destinos. También incluye temas relacionados con las bases de datos relacionales, lo que refleja su uso para el almacenamiento y análisis de información estructurada en el sector turístico.

El tema 'customer-satisfaction' es un tema básico y transversal. Es el tema con mayor número de citas y el segundo índice H más alto. Abarca un amplio abanico de temas, pero destacan los relacionados con el comportamiento del turista, como la fidelidad o las motivaciones. Igualmente, y relacionado con los anteriores, están los temas relacionados

con la búsqueda de información. El comportamiento y la satisfacción del turista es otro de los principales temas de análisis en la investigación turística basada en datos.

El tema 'tourism' es un tema básico y transversal. Es el segundo tema con el índice H más alto, igualado en esta posición con el rendimiento. Como es habitual en los temas de este cuadrante, se trata de un tema bastante general. Incluye cuestiones de gran alcance relacionadas con la gestión o los servicios, como la segmentación del mercado o los canales de distribución.

Considerando todos los temas identificados en este periodo, tiene sentido agruparlos en tres grandes áreas temáticas, a saber:

- **Temas de investigación en turismo:** abarca estos temas de investigación desde una perspectiva empresarial. Los temas 'customer-satisfaction', 'tourism' y 'performance' se centran en temas objeto de estudio en la investigación turística basada en datos y se incluye en esta área temática. Otros temas del periodo, con un peso en número de documentos, índice H y número de citas muy inferior a los anteriores, encontramos una clasificación similar: 'perceived-risk', 'disability', 'environment', 'business', 'strategies' and 'perceived-usefulness'.
- **Fuentes de información:** incluye temas relacionados con las fuentes del sistema de información. Los temas 'travel-agencies' y 'tourism-information-system' pertenecen a esta área. Incluimos además los relacionados con las redes sociales geolocalizadas.
- **Técnicas de análisis de datos:** esta área abarca temas relacionados con las técnicas estadísticas y la ciencia de datos que se aplican en la investigación. El tema 'clustering' pertenece a esta área.

La Figura 9 destaca en diferentes colores los temas de cada una de las áreas temáticas mencionadas anteriormente y la Tabla 3 muestra los indicadores de resultados basados en estas tres áreas temáticas.

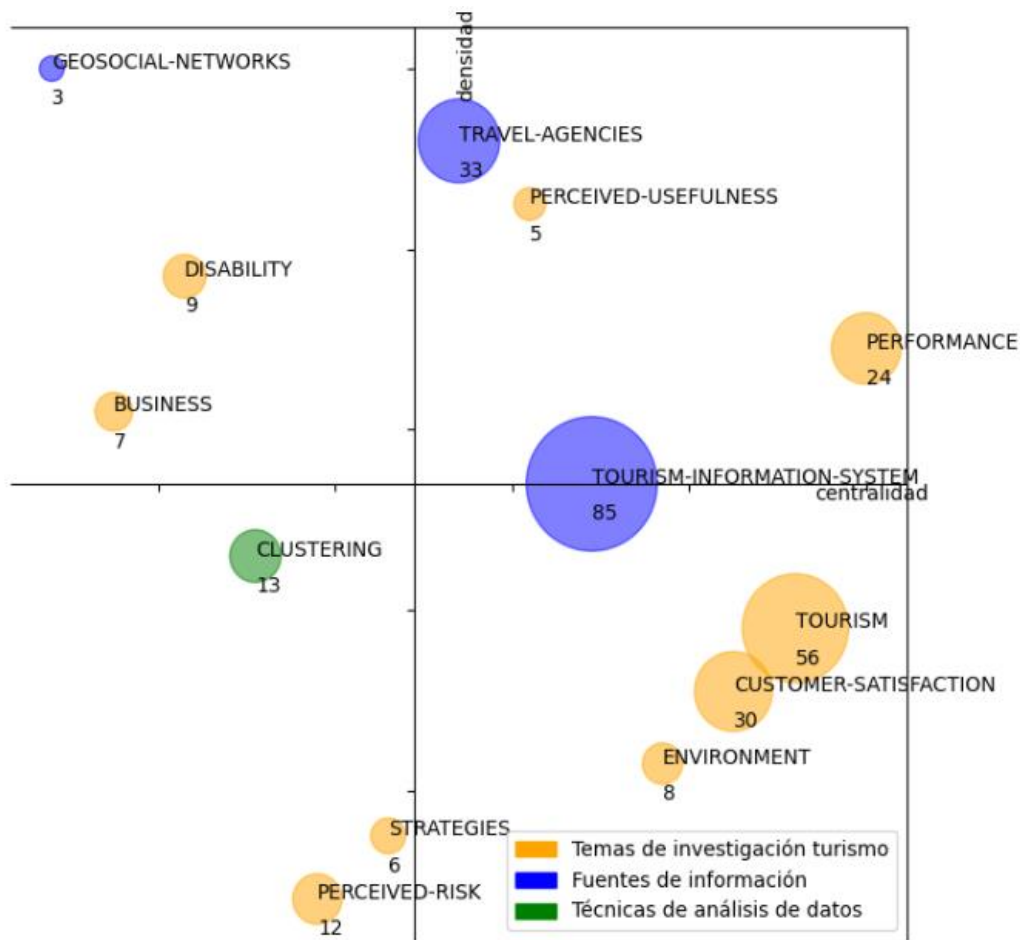


Figura 9: Diagrama estratégico para el periodo 1982-2015 clasificado por áreas temáticas. Se muestra el número de documentos de cada tema. Fuente: elaborado con SciMAT

Áreas temáticas	Número de documentos	Citas
Temas de investigación en turismo	157	4528
Fuentes de información	121	1381
Técnicas de análisis de datos	13	215

Tabla 3: Indicadores de rendimiento por área temática en el periodo 1982-2015

Segundo periodo (2016-2021)

Como puede observarse en la Figura 10, este segundo periodo contiene 15 temas de investigación: 'antecedents', 'deep-learning', 'destination-image', 'economic-growth', 'hotels', 'impact', 'information-sources', 'machine-learning', 'management', 'online-reviews',

'sustainable-tourism', 'tourism', 'tourism-destination', 'tourism-promotion', 'user-generated-content'.

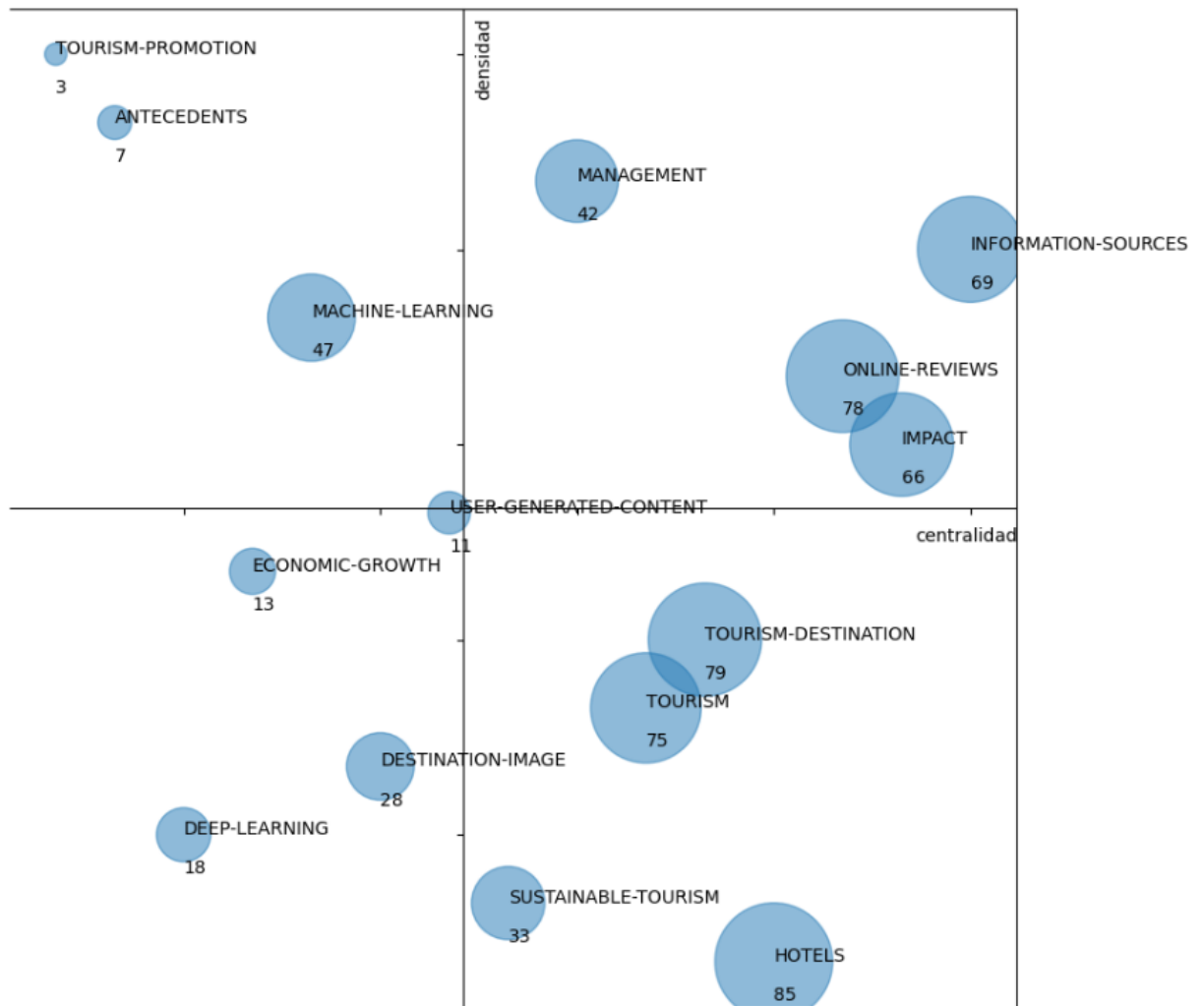


Figura 10: Diagrama estratégico para el periodo 2016-2021. Se muestra el número de documentos de cada tema. Fuente: elaborado con SciMAT

La Tabla 4 muestra los indicadores de rendimiento: número de documentos, citas obtenidas por estos documentos y el índice H. Según estos indicadores, podemos destacar como temas principales los siguientes: 'hotels', 'tourism-destination', 'online-reviews', 'tourism', 'information-sources' y 'impact'. Estos 6 temas representan en conjunto el 73% del número total de citas y el 69% de los documentos. El índice H de estos temas también es superior, aunque la diferencia no es tan grande como en el número de citas o documentos.

Temas	Número de documentos	Índice H	Citas
Hotels	85	10	513

Tourism-destination	79	14	1072
Online-reviews	78	15	979
Tourism	75	12	597
Information-sources	69	14	841
Impact	66	13	655
Machine-learning	47	7	257
Management	42	9	364
Sustainable-tourism	33	7	175
Destination-image	28	8	219
Deep-learning	18	7	236
Economic-growth	13	5	173
User-generated-content	11	6	162
Antecedents	7	5	86
Tourism-promotion	3	2	6

Tabla 4: Indicadores de rendimiento por temas en el periodo 2016-2021.

Estos temas principales se describen a continuación.

El tema 'online-reviews' es un tema impulsor; es el tercer tema con mayor número de documentos y el segundo en cuanto a citas. Se centra en la información generada por el turista, ya sea a través de valoraciones u opiniones en plataformas de intermediación en la contratación de servicios turísticos, encuestas de satisfacción, comentarios en redes sociales, intercambio de información en comunidades virtuales o correos electrónicos. Los datos de este tipo se clasifican de forma más general en datos generados por el usuario (Li et al., 2018). Este tipo de datos se conocen de forma extendida por sus siglas en inglés UGC (*User Generated Content*). Relacionado con lo anterior aparecen cuestiones relacionadas con la analítica de datos. Se refleja la gran relevancia que ha adquirido para el turismo el análisis de la información generada por el turista, tanto porque la tecnología permite capturarla y almacenarla como por el cambio hacia un enfoque más centrado en el análisis del comportamiento y la experiencia del turista.

El tema 'information-sources' es un tema impulsor; es el tercer tema con mayor número de citas y el segundo en términos de índice H igualado con 'tourism-destination'. Contiene temas relacionados con las distintas fuentes de datos diferenciadas por tipo y origen. Abarca datos procedentes de viajes, buscadores de internet, imágenes, comportamiento del turista o reservas hoteleras. Refleja la amplia gama de fuentes de datos disponibles, cuya variedad ha ido creciendo en la última década.

El tema 'impact' es un tema impulsor; contiene cuestiones diferenciadas, como el impacto de las redes sociales en el turismo, cada vez más relevantes, y, por otro lado, un tema clásico como el impacto del turismo en la economía, con énfasis en la economía de los países. Por otro lado, contiene temas relacionados con la hostelería, los canales de distribución y las agencias de turismo online.

El tema 'tourism-destination' es un tema básico y transversal; es el tema con mayor número de citas, el segundo índice H más alto igualado con 'information-sources' y el segundo mayor número de documentos. Se centra en el análisis de los destinos turísticos desde la perspectiva de la demanda, el comportamiento del turista y la experiencia. Aparecen temas de actualidad como el turismo inteligente, que se basa en la tecnología para mejorar los servicios y la experiencia del turista. El análisis de los destinos turísticos desde diferentes perspectivas es uno de los temas centrales de la investigación turística basada en datos.

El tema 'hotels' es básico y transversal; es el tema con mayor número de documentos. Contiene cuestiones relacionadas con la calidad del servicio, la gestión de ingresos, los servicios, la industria, el rendimiento y la productividad. También está relacionado con la plataforma de reservas online Airbnb, que refleja los cambios actuales en las reservas de alojamiento.

El tema 'tourism' es un tema básico y transversal. Al ser un tema muy general, está relacionado con una amplia gama de cuestiones, como los factores determinantes, las diferencias de género, el precio, las redes sociales, la segmentación y otras.

A modo de resumen y utilizando las áreas temáticas identificadas en el primer periodo, vemos que los temas: 'tourism-promotion', 'antecedents', 'tourism-destination', 'impact', 'management', 'tourism', 'economic-growth', 'sustainable-tourism' y 'hotels' pertenecen al área temática 'temas de investigación en turismo'. Los temas 'information-sources', 'online-reviews', 'user-generated-content' y 'destination-image' pertenecen al área temática de 'fuentes de información' y finalmente 'machine-learning' y 'deep-learning' pertenecen al área de 'técnicas de análisis de datos'.

En la Figura 11, los temas de cada una de las áreas temáticas aparecen resaltados en distintos colores, y la Tabla 5 muestra los indicadores de resultados basados en estas tres áreas temáticas.

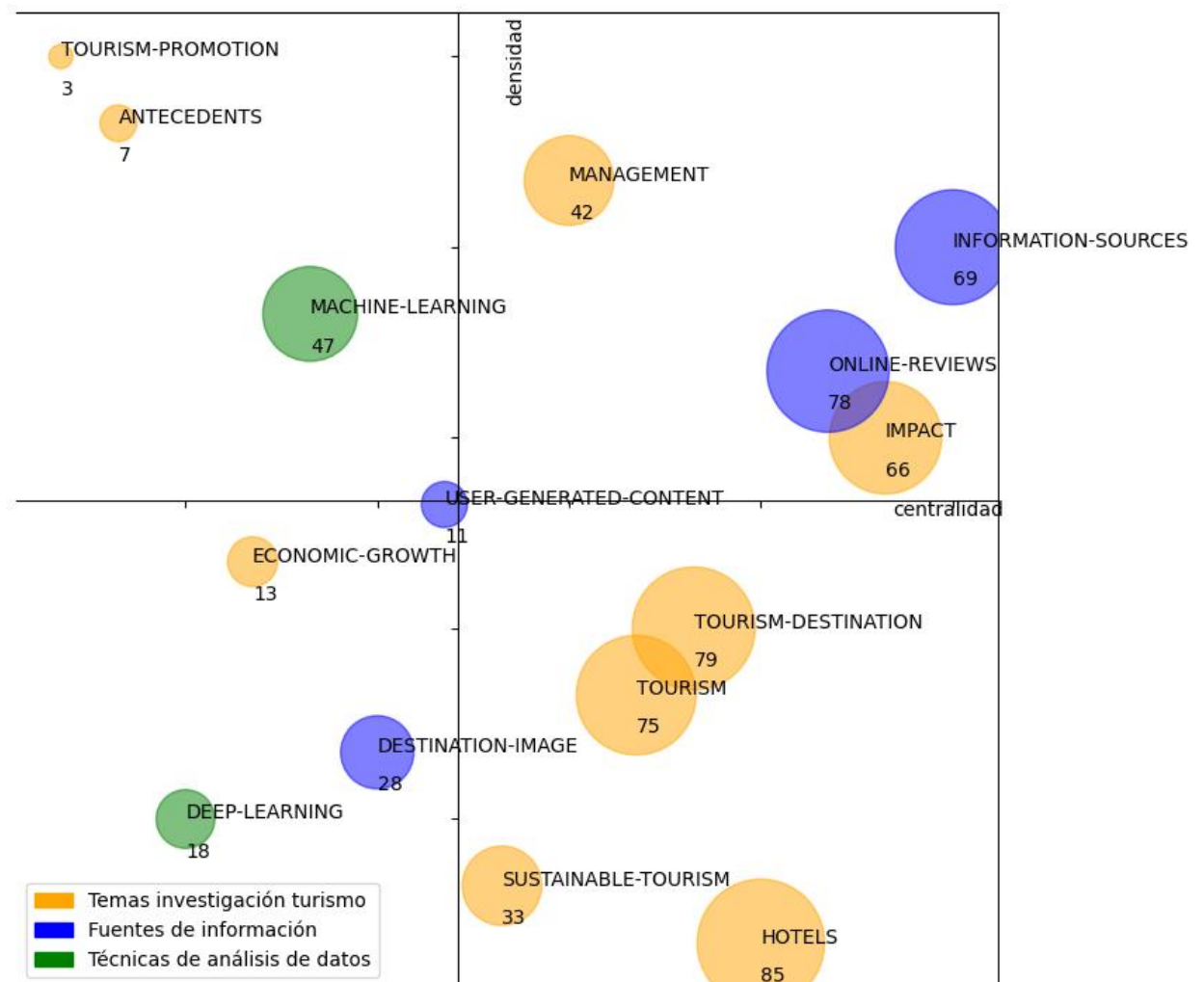


Figura 11: Diagrama estratégico para el periodo 2016-2021 clasificado por áreas temáticas. Se muestra el número de documentos de cada tema. Fuente: elaborado con SciMAT

Áreas temáticas	Número de documentos	Citas
Temas de investigación en turismo	403	3641
Fuentes de información	186	2201
Técnicas de análisis de datos	65	493

Tabla 5: Indicadores de rendimiento por área temática en el periodo 2016-2021

2.3.3. Diagrama de evolución y solapamiento

Se ha desarrollado un análisis de los temas detectados en cada periodo de tiempo teniendo en cuenta las palabras clave y su evolución en el tiempo; para ello se ha vuelto a

utilizar SciMAT. El mapa de solapamiento de SciMAT muestra que el 12% de las palabras clave se comparten entre periodos.

La evolución conceptual se muestra en la Figura 12. En este mapa las líneas continuas significan un nexo temático (comparten palabras claves con el nombre de los temas), una línea discontinua punteada significa que los temas enlazados comparten palabras clave distintas del nombre de los temas, el grosor de la línea es proporcional al índice de inclusión, el tamaño de la esfera es proporcional al número de documentos albergados por cada tema y el color muestra el área temática a la que pertenecen los temas.

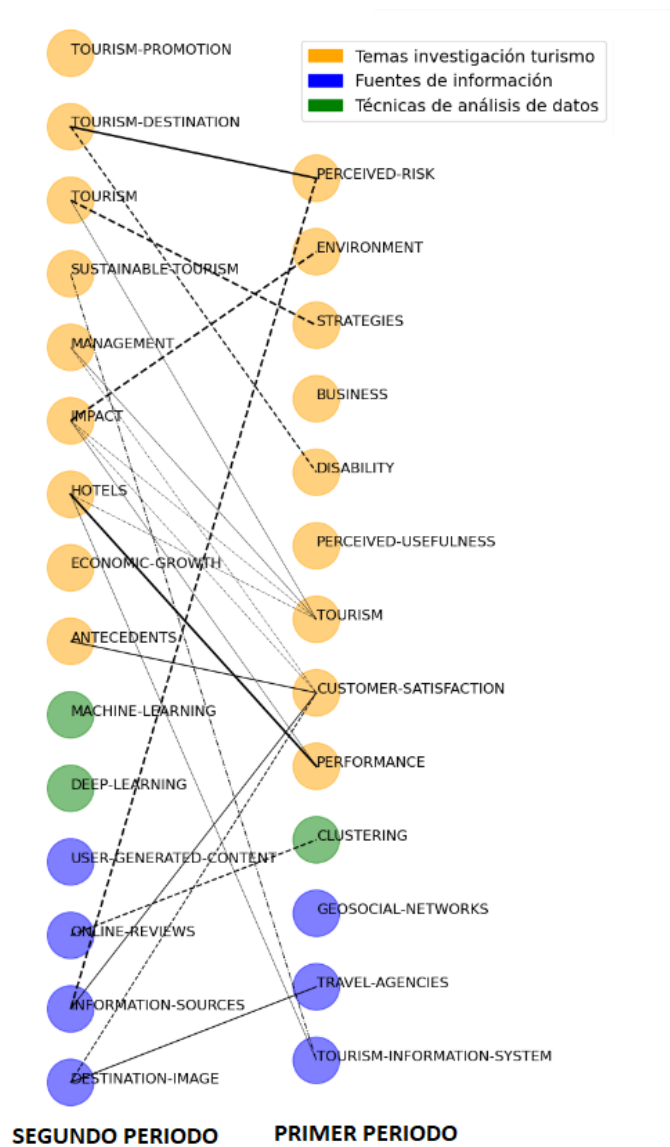


Figura 12: Mapa de evolución conceptual entre periodos. Fuente: elaborado con SciMAT

En el mapa de evolución es interesante repasar la evolución de las áreas temáticas. La mayoría de los temas del área temática 'temas de investigación en turismo' conectan con temas de la misma área. La clara diferencia de temas entre el primer y el segundo periodo refleja el cambio de temas de investigación que se ha producido entre ambos periodos. El tema 'impact', que pertenece al segundo periodo, es un tema destacado ya que conecta con cuatro temas del primer periodo, de los cuales, el vínculo más fuerte es con 'environment', todos ellos del área temática 'temas de investigación en turismo'. También cabe destacar en el área temática de 'temas de investigación en turismo' el fuerte vínculo entre temas como: 'perceived-risk' del primer período con el 'tourism-destination' del segundo período, así como el vínculo entre 'strategies' del primer período con 'tourism' del segundo período y el vínculo entre 'performance' del primer período con 'hotels' del segundo período.

Respecto al área temática de 'técnicas de análisis de datos', vemos que los temas 'machine learning' y 'deep learning' aparecen en el segundo periodo como temas nuevos sin relación con el periodo anterior, lo que refleja la aparición disruptiva de este tipo de técnicas avanzadas de análisis de datos.

Por último, la evolución del tema en el área temática 'fuentes de información' refleja el cambio en el tipo de fuentes de información utilizadas entre el primer y el segundo periodo. En el segundo periodo, aparecen nuevas fuentes de información, como las 'online-reviews', 'destination-image' y 'user-generated-contents'. Estas nuevas fuentes de información son generadas por los propios turistas en diferentes plataformas digitales y es un tipo de información cuyo uso ha crecido sustancialmente.

2.3.4. Análisis cualitativo de áreas temáticas

Además de los indicadores de rendimiento y los mapas de evolución obtenidos, cabe destacar algunos aspectos específicos que detallamos a partir de las tres áreas temáticas mencionadas. Para ello, profundizamos en el análisis revisando las palabras clave contenidas en los distintos temas como resultado del análisis de co-ocurrencias.

Temas de investigación en turismo

Las palabras clave relacionadas con esta área temática incluyen: 'impact', 'tourism-destination', 'customer-satisfaction', 'customer-behaviour', 'travel', 'hospitality', 'management', 'hotels', 'word-of-mouth', 'performance' y 'sustainable-tourism'. Algunos de ellos son bastante genéricos y están asociados al sector turístico, como 'travel', 'hospitality' o 'hotels'. Otros, en cambio, se centran más en temas específicos, como 'customer-

satisfaction' y 'customer-behaviour', que reflejan el creciente interés por analizar la satisfacción y el comportamiento de los turistas. Se observa un notable aumento de la investigación en esta categoría entre el primer y el segundo periodo. Otro grupo de palabras clave puede ser: 'management', 'impact' o 'performance', palabras clave comunes en estudios orientados a medir el rendimiento económico y desde otros puntos de vista de los diferentes actores de la industria turística, medidas de la calidad del servicio turístico y diferentes medidas de impacto, como la medida del impacto del turismo en la economía de los países. Por otro lado, la palabra clave 'sustainable-tourism', que aparece con más frecuencia en el segundo periodo, muestra el creciente interés por las cuestiones de sostenibilidad en la gestión de los servicios turísticos. Algo similar ocurre con la palabra 'tourism-destination', que es más frecuente en el segundo periodo que en el primero, lo que refleja el análisis de los destinos turísticos desde cada vez más perspectivas. Por último, la palabra clave 'word-of-mouth' muestra la creciente importancia en el análisis de los servicios turísticos de las opiniones de los turistas en plataformas on-line de reserva de alojamiento y servicios turísticos, en redes sociales, foros de discusión y otros. Este tipo de datos son comúnmente referidos como e-WOM (siglas en inglés de *Electronic Word of Mouth*) y se trata de un término también más frecuente en el segundo periodo que en el primero.

Fuentes de información

En primer lugar, encontramos 'information-sources' o 'tourist-information-system' entre las palabras clave con mayor número de documentos. También destaca la palabra clave 'GIS', que hace referencia al elevado uso de datos espaciales en la investigación turística, o 'big-data', que refleja el uso de grandes conjuntos de datos.

Revisando las palabras clave en profundidad, encontramos cambios notables entre los periodos en las fuentes de información utilizadas, como ya hemos visto en el mapa de evolución. Por ejemplo, la palabra clave 'online-reviews' y otras con el mismo significado 'online-customer-reviews' o 'online-travel-reviews' aparecen con una alta frecuencia en el segundo periodo, mientras que no aparecen en el primero. Esta palabra clave hace referencia a las reseñas realizadas por los turistas en plataformas online de reserva de alojamiento y servicios turísticos (Booking, Airbnb, Tripadvisor), que se utilizan cada vez más como fuentes de información en la investigación turística basada en datos.

Es interesante observar con más detenimiento el grupo de 'online-reviews'. En la Figura 13 podemos ver algunas de las fuentes donde se generan los datos ('twitter', 'customer-reviews', 'word-of-mouth'), fuentes de información relacionadas ('mobile-positioning-data', 'social-media-data'), técnicas analíticas ('analytics', 'big-data', 'sentiment-

analysis', 'classification') y temas a los que se aplica la investigación basada en estos datos ('reputation', 'sharing-economy').

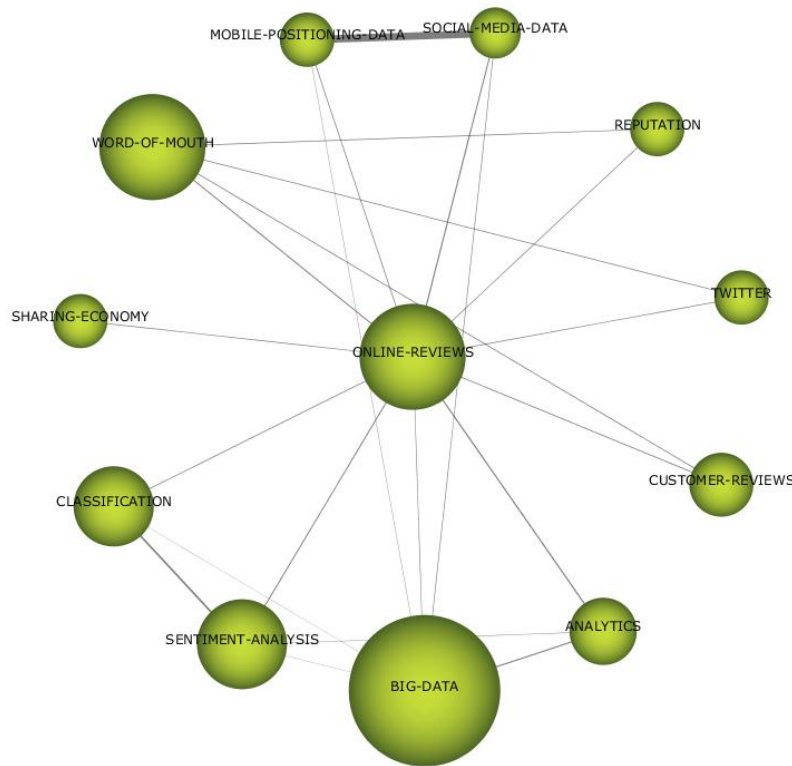


Figura 13: Palabras clave en el tema 'online-reviews'. Fuente: elaborado con SciMAT

Como ya se ha indicado, los datos del tipo 'online-reviews' se clasifican más generalmente en los datos de tipo UGC. De forma similar a lo que sucede con 'online-reviews', la palabra clave 'user-generated-content' aparece en el segundo periodo, pero no en el primero. Algo similar ocurre con las palabras clave referidas a redes sociales o imágenes, con un uso notablemente mayor en el segundo periodo. Otras fuentes de información que también crecen son los datos relacionados con el internet de las cosas, este tipo de datos se conocen de forma extendida por sus siglas en inglés IoT (*Internet of Things*). Es notable también el crecimiento en los datos de dispositivos móviles, en lo referente a las tendencias en buscadores de Internet, también se aprecia un crecimiento, aunque no tan notable como en las anteriores. Del mismo modo, la palabra clave 'big-data' es claramente una palabra clave del segundo periodo, lo que refleja el aumento del uso de grandes conjuntos de datos en la investigación turística en los últimos años.

Tal y como se ha indicado anteriormente, una de las palabras clave con mayor número de documentos es 'tourist-information-system'. En las iniciativas basadas en datos, los repositorios de información integrados son fundamentales. Para detectar este tipo de sistemas que centralizan la información, hemos buscado otras palabras clave como 'data-

warehouse' o 'data-lake' y similares. Revisando los documentos asociados a ellas, hemos encontrado un desarrollo relativamente bajo de este tipo de repositorios en el sector turístico, donde su uso podría ser clave.

Con un objetivo similar buscamos palabras clave que hicieran referencia a los datos estadísticos del turismo, como 'tourism-statistics' o similares. Encontramos un número relativamente bajo de documentos, revisando los documentos encontramos que los datos estadísticos de turismo producidos por diferentes tipos de instituciones tienen un uso relativamente bajo en la investigación del turismo basada en datos.

Encontramos algo similar cuando revisamos las fuentes de datos abiertos y buscamos palabras clave como 'open-data'. Encontramos un bajo número de documentos que denota un escaso uso en investigación de este tipo de fuentes que han tenido un notable crecimiento en la última década.

Técnicas de análisis de datos

En cuanto a esta área temática, se han detectado tres temas: 'clustering' en el primer periodo y 'machine learning' y 'deep learning' en el segundo. Al tratarse de pocos temas y con denominaciones muy genéricas, es interesante explorar la composición por palabras clave de cada tema que podemos obtener con el software SciMAT:



Figura 14: Palabras clave en el tema 'clustering'. Fuente: elaborado con SciMAT

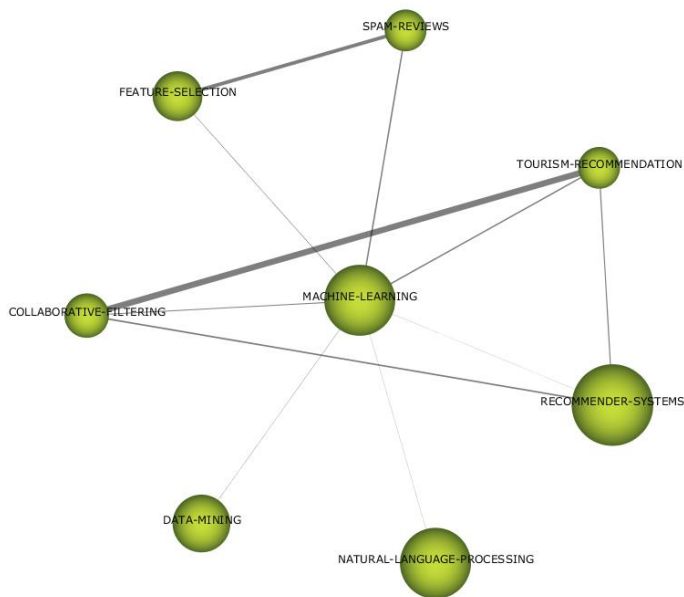


Figura 15: Palabras clave en el tema 'machine learning'. Fuente: elaborado con SciMAT

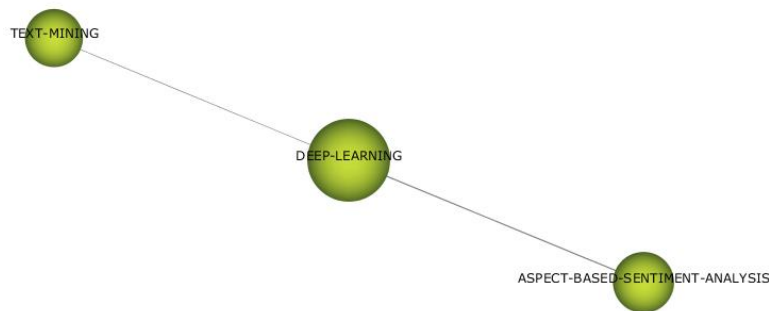


Figura 16: Palabras clave en el tema 'deep learning'. Fuente: elaborado con SciMAT

Las palabras clave relacionadas con técnicas analíticas más avanzadas o innovadoras, como: 'machine learning', 'sentiment-analysis', 'text-mining', 'natural-language-processing', 'recommendation-systems', 'sentiment-analysis' o 'deep learning' aparecen con mucha más frecuencia en el segundo periodo, reflejando el creciente uso de técnicas analíticas avanzadas en la investigación turística, de acuerdo con lo visto en el mapa de evolución.

2.3.5. Marco de referencia inicial

Una vez identificadas y analizadas las principales áreas, vamos a plasmarlas en un marco de referencia inicial sobre el uso del dato en la toma de decisiones en organizaciones

turísticas. Este marco de referencia inicial se estructurará en base a las áreas identificadas en el esquema de toma de decisiones basadas en datos en el sector del turismo mostrado en la Figura 1, por tanto, integramos cada área temática con sus temas en el área correspondiente del esquema (fuentes de datos, almacenamiento y gestión de datos, técnicas de ciencia de datos e inteligencia artificial y generación de análisis y descubrimientos relevantes). Cabe destacar que, en el caso concreto de los temas del área temática 'Técnicas de análisis de datos', dado que se habían detectado pocos temas y muy genéricos, se han incorporado al modelo las palabras clave que componen cada tema y que se han mostrado en el apartado 2.3.4. Los temas pertenecientes al segundo periodo se resaltan en gris para diferenciarlos del primer periodo. El marco propuesto plantea las siguientes áreas:

- **Fuentes de datos:** se identifican las fuentes de datos más utilizadas en la investigación turística. Se muestra la evolución de las fuentes de información de un periodo a otro. Se obtiene un amplio abanico de fuentes de información diferentes, entre las que destacan: sistemas de información turística (Puhretmair et al., 2002), redes geosociales (Yin et al., 2018), imágenes de destinos (Chang & Chiang, 2022), reseñas en línea (Kim & Han, 2022) o UGC (Li et al., 2018; Mukhopadhyay et al., 2022; Akbari et al., 2022). Las fuentes de información: reseñas en línea o UGC son actualmente las más utilizadas, que proporcionan información valiosa sobre la percepción y el comportamiento de los turistas.
- **Almacenamiento y gestión de los datos:** el marco planteado destaca la importancia de contar con una plataforma de datos (Shi, 2020) que facilite el almacenamiento y la organización de los datos de forma centralizada y eficiente. Se han realizado investigaciones en este ámbito (Navarro & Rubio, 2000; Abdulaziz et al., 2015; Ramos et al., 2017), pero en general son escasas y faltan repositorios de datos centralizados en el sector turístico, especialmente repositorios de lagos de datos.
- **Técnicas de ciencia de datos e inteligencia artificial:** se identifican las técnicas de ciencia de datos e inteligencia artificial más utilizadas en la investigación turística. Entre las que destacan: clustering (Feng et al., 2022), personalización (Gupta et al., 2022), filtrado colaborativo (He, 2022), aprendizaje automático (Kayakus, 2022), sistemas de recomendación (Julashokri et al., 2022), minería de datos (Ma, 2022), minería de textos (Loureiro et al., 2022), procesado de lenguaje natural (Ray & Bala, 2021) o aprendizaje profundo (Law et al., 2019). Algunas de las técnicas identificadas como: minería de textos, análisis de sentimientos o procesado de lenguaje natural se aplican a la información de tipo texto no estructurada, que es la que se recoge típicamente en las fuentes como UGC o reseñas en línea identificadas en el área de

fuentes de datos. Destaca la evolución hacia técnicas de aprendizaje automático o aprendizaje profundo.

- **Generación de análisis y descubrimientos relevantes para la toma de decisiones:** se identifican las áreas más importantes de aplicación del conocimiento obtenido. Entre las que destacan: satisfacción del cliente (Padma & Ahn, 2020), promoción del turismo (Wu et al., 2022), turismo sostenible (Nilashi et al., 2019) o destino turístico (Rashad, 2022). Se puede observar la evolución de las áreas de un periodo a otro.

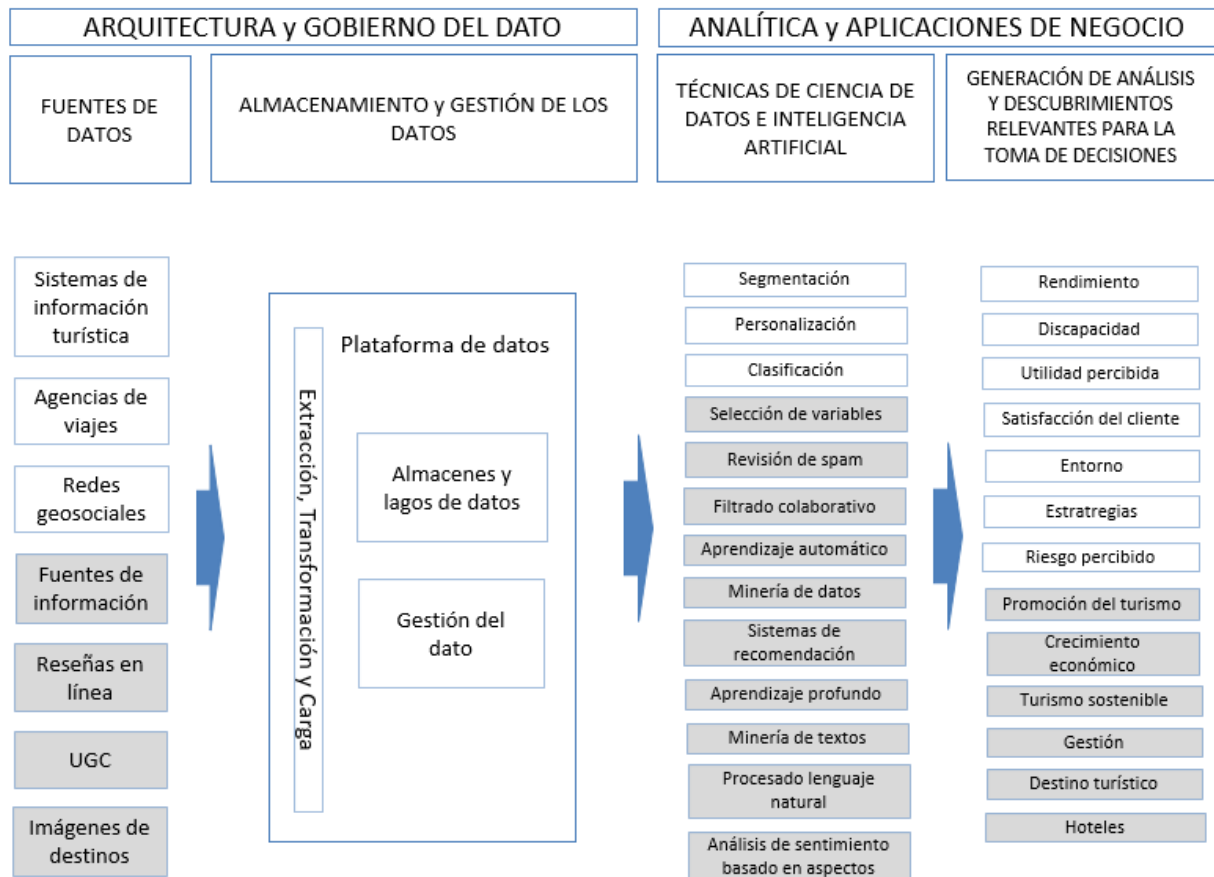


Figura 17: Marco de referencia inicial sobre el uso del dato en la toma de decisiones en organizaciones turísticas. Los temas pertenecientes al segundo periodo se resaltan en gris para diferenciarlos del primer periodo. Fuente: elaboración propia

3. Marco teórico de investigación

En este capítulo, se presentan los conceptos teóricos que sirven de base para las investigaciones realizadas. En esta memoria se va a proponer un marco conceptual para la toma de decisiones basadas en datos en organizaciones turísticas en base al cual se realizarán diferentes casos de uso de aplicación en el sector del turismo. Cada uno de estos casos de uso emplea distintas técnicas de ciencia de datos (véase Figura 17) y tienen, por tanto, un fundamento teórico diferente.

En primer lugar, aplicaremos el marco conceptual propuesto al análisis de experiencias de turistas, se trabajará con comentarios de turistas cuyas palabras se relacionarán mediante grafos y se organizarán con técnicas de agrupamiento para detección de comunidades. En un segundo caso, aplicaremos el marco conceptual propuesto para la digitalización de tiendas físicas en el sector del turismo. Al tratarse de una decisión compleja emplearemos modelos MCDM, concretamente modelos AHP y al tratarse también de una decisión estratégica emplearemos el cuadro de mando integral. En un tercer caso se aplicará el marco conceptual propuesto para el posicionamiento óptimo de oficinas de turismo. En este caso, es importante introducir el tipo de variables geográficas a emplear, el concepto de distancia a utilizar, así como técnicas de agrupamiento jerárquico que nos ayudarán a encontrar la ubicación en la que se posicionarán las oficinas. Por último, y también relacionado con el último caso mencionado se obtendrán recomendaciones utilizando grandes modelos de lenguaje.

3.1 Análisis de grafos

Un grafo es un conjunto de objetos llamados nodos (o vértices) y una selección de pares de nodos, llamados aristas (*edge* en inglés) que pueden ser orientados o no. Típicamente, un grafo se representa mediante una serie de puntos (los nodos) conectados por líneas (las aristas) tal y como vemos en la Figura 18.

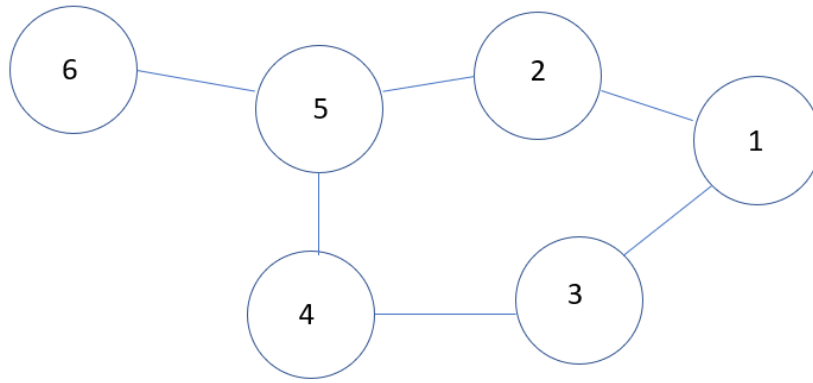


Figura 18: Diagrama de un grafo con 6 vértices y 7 aristas. Fuente: elaboración propia

Una red puede representar un único tipo de relación entre nodos (simple) o más de un tipo de relación (múltiple). En algunos casos, es necesario asignar una dirección a las aristas. Los grafos que contienen aristas con una dirección definida se denominan grafos dirigidos, mientras que un grafo no dirigido es un tipo de grafo en el que las aristas representan relaciones simétricas y no tienen una dirección definida. El tipo de grafos con el que vamos a trabajar son los grafos no dirigidos.

Formalmente un grafo no dirigido se define como un par de conjuntos $G = (V, E)$, donde:

- $V \neq \emptyset$ es el conjunto no vacío de vértices o nodos.
- $E \subseteq \{ (a, b) \in V \times V \}$ es el conjunto de aristas (*edges*), tal que $(a, b) = (b, a)$.

En el caso de los grafos no dirigidos, una arista (a, b) también se puede denotar como $\{a, b\}$.

Dentro de un grafo G podemos encontrar subgrafos que son también grafos cuyos conjuntos de vértices y aristas son subconjuntos de los de G .

3.1.1 Métodos de agrupamiento para detección de comunidades

La teoría de grafos, que representa las relaciones entre nodos y aristas, se combina con las técnicas de agrupamiento, que agrupan nodos similares, para revelar comunidades y patrones intrínsecos en conjuntos de datos de red.

En la teoría de grafos, una comunidad se define como un grupo de nodos densamente conectados entre sí, mientras que hay conexiones más débiles con nodos fuera del grupo o clúster. Los algoritmos de agrupamiento aplicados a grafos buscan identificar estas comunidades, lo cual facilita la comprensión de la estructura subyacente y las relaciones dentro de una red. Existen diferentes métodos de agrupamiento para el análisis y detección de comunidades (Krishna et al., 2018), se pueden resumir en:

- **Métodos jerárquicos**, que buscan particiones naturales en la red, basándose en la idea de que la red tiene una estructura jerárquica.
- **Métodos modulares**, que se centran en tratar de encontrar la partición que maximice el modularidad (una métrica que compara los enlaces internos de una comunidad frente a los enlaces que conectan la comunidad con el resto de la red).

En esta tesis, en lo relativo a detección de comunidades en redes vamos a emplear técnicas de agrupamiento que van en la línea de los métodos modulares. Los métodos modulares ofrecen algunas variaciones, como la técnica de visualización de similitudes (VOS por sus siglas en inglés Visualization of Similarities) utilizada por la herramienta Vosviewer (Van Eck & Waltman, 2007; Van Eck et al., 2010). Esta técnica es interesante ya que permite obtener clústeres pequeños, lo que supone un hándicap en las técnicas modulares que tienen limitaciones en este sentido. La técnica VOS incluye el parámetro de resolución, este parámetro ayuda a tratar el problema del límite de resolución (Fortunato & Barthélemy, 2007) del agrupamiento basado en el modularidad. Consideramos que esta funcionalidad que ofrece la técnica VOS es de gran interés para el caso de uso que se desarrollará en el capítulo 5.

3.2 Proceso analítico jerárquico (AHP)

La toma de decisiones es un proceso que permite seleccionar la mejor solución posible entre una serie de alternativas (que previamente han sido analizadas y ordenadas según la valoración de uno o más criterios) con la finalidad de dar respuesta a un problema dado. Tiene gran importancia, dado que con frecuencia tanto las empresas, como las personas deben participar en procesos de toma de decisiones. Por tanto, se requiere el desarrollo de métodos y modelos que permitan representar fielmente cada problema y analizar las distintas alternativas objetivamente. En los problemas de toma de decisiones se puede utilizar un único criterio para clasificar las alternativas o varios, en cuyo caso se denomina toma de decisiones multicriterio (MCDM). Los problemas de MCDM son más complejos

de resolver que aquéllos en los que solo hay que considerar un criterio, dado que cada criterio puede establecer un orden de preferencia particular y diferente sobre el conjunto de alternativas. Por tanto, a partir de las distintas ordenaciones de preferencia particulares, será necesario establecer algún mecanismo que permita construir un orden global de preferencia.

El ambiente de decisión viene definido por las características y el contexto en el que se lleva a cabo la toma de decisiones. Así, se distinguen tres situaciones, dependiendo de que el ambiente sea de:

- **Certidumbre**, donde se conocen con exactitud todos los elementos y factores que intervienen en el problema, lo que permite asignar valores precisos de utilidad a cada una de las alternativas presentes en el mismo.
- **Riesgo**, donde alguno de los elementos o factores que intervienen responden a las leyes del azar, por lo que estos problemas se resuelven utilizando la Teoría de la Probabilidad.
- **Incertidumbre**, donde la información disponible sobre las distintas alternativas puede ser incompleta, vaga o imprecisa, lo que implica que la utilidad asignada a cada alternativa tenga que ser valorada de forma aproximada. En este caso será necesario emplear enfoques basados en lógica difusa.

En el ámbito de un problema de MCDM, el grupo de expertos desempeña un papel fundamental. Algunas de sus funciones más comunes son: identificación de criterios relevantes, definición de pesos y preferencias, evaluación de alternativas y validación del modelo.

Dentro de los métodos a aplicar para resolver los problemas MCDM uno de los más comúnmente empleados es el AHP. Saaty (1980) desarrolló el AHP como técnica para ayudar a los responsables de la toma de decisiones a resolver problemas complejos en los que intervienen características cualitativas y cuantitativas. Esta técnica permite clasificar y seleccionar posibles alternativas en un conjunto de comparaciones por pares. El responsable de la toma de decisiones evalúa previamente cada criterio en una escala para determinar su importancia. Por último, utilizando la matriz de comparación por pares propuesta por un panel de expertos, se llega a una decisión óptima que tiene en cuenta la importancia de cada criterio. El AHP se ha utilizado ampliamente en la toma de decisiones estratégicas (Zhang et al., 2021; Dong et al., 2018). La técnica permite comprobar si las evaluaciones de los responsables de la toma de decisiones son coherentes, garantizando la reducción del sesgo de decisión. AHP sigue cuatro pasos (Saaty, 1980, 2008; Carrasco et al., 2018) que se detallarán a continuación y son los siguientes: 1) Definir el problema de

decisión, 2) Estructurar jerárquicamente el problema de decisión, 3) Realizar comparaciones por pares y calcular el vector de ponderaciones, y 4) Obtener las ponderaciones de las alternativas finales mediante varios niveles de agregación de ponderaciones.

3.2.1 Definición del problema de decisión

La primera etapa consiste en identificar claramente el problema que hay que resolver, el posible resultado u objetivo del sistema, sus características relevantes y los expertos implicados en la toma de la decisión. Una vez recopilada toda esta información, podemos pasar a la fase de organización del problema de decisión en un modelo jerárquico.

3.2.2 Estructuración jerárquica del problema de decisión

En esta fase, el objetivo, los criterios, los subcriterios y las alternativas previamente identificadas se utilizan para construir la jerarquía que nos permitirá analizar el problema de decisión, como se muestra en la Figura 19. El primer nivel está formado únicamente por el objetivo del problema, el segundo nivel se genera con los distintos criterios considerados y el último nivel contiene las posibles alternativas:

- **Objetivo:** un único elemento, llamado G , que generalmente representa el óptimo resultado para el problema. Este es el primer nivel de la jerarquía, por ejemplo, $k = 0$.
- **Criterios:** características que se tienen en cuenta para la consecución del objetivo. Estos criterios jerárquicos se denominan:

$$C = \bigcup_{k=1}^{\#H-2} c(k),$$

donde $\#H-2$ es el número de niveles en que se han subdividido los criterios (al menos algunos de ellos), k que identifican el nivel del criterio, $c(k) = \{c^k_1, \dots, c^k_{\#c(k)}\}$, donde $\#$ simboliza el número de elementos del conjunto.

- **Subcriterios:** cada uno de estos criterios c^k_i puede, a su vez, ser subdividido en subcriterio en niveles más bajos $k+1$, $c_i(k+1) = \{c^{k+1}_{i1}, \dots, c^{k+1}_{i\#Ci(k+1)}\}$ y así sucesivamente. Las alternativas deben evaluarse para cada uno de los subcriterios de cada criterio. Así, la jerarquía se recorre desde su punto más bajo hasta el primer nivel, donde se encuentran los criterios.
- **Alternativas:** las posibles soluciones del objetivo dada una elección forman una

colección de resultados. $A = \{a_1, \dots, a_{\#A}\}$. Son el último nivel de la jerarquía, $k = \#H-1$.

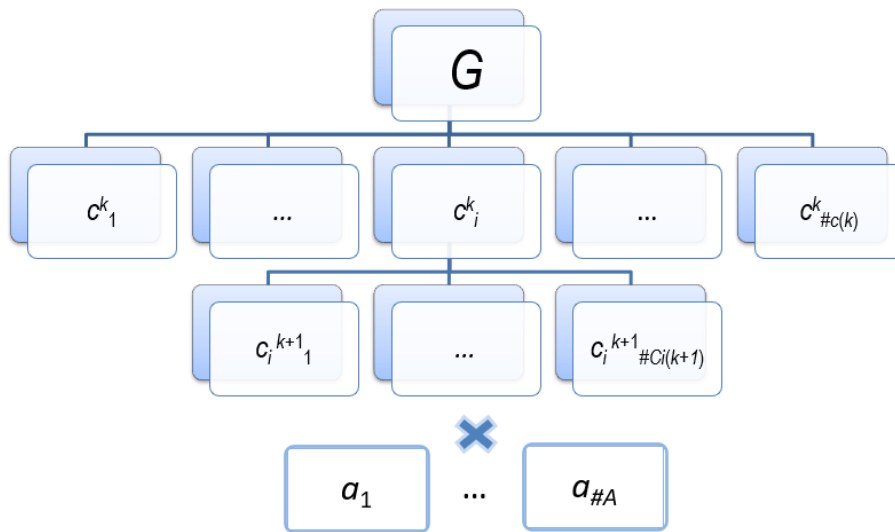


Figura 19: Problema de toma de decisiones en una jerarquía H . Fuente: elaboración propia

3.2.3 Comparación por pares y cálculo del vector de ponderaciones

En este apartado, la opinión de los expertos se utiliza para comparar, por parejas, los elementos de un nivel k con respecto a un elemento padre del nivel superior $k-1$. Se asigna un valor en una escala de 1 a 9 a los distintos elementos, valores que representan la importancia que los expertos encuestados consideran que tiene un elemento en relación con otro. Donde 1 representa que ambos elementos tienen la misma importancia, mientras que 9 significa que uno de los elementos es extremadamente más importante que el otro.

La escala de importancia por pares se ilustra en la Figura 20 siguiendo la escala de Saaty (Saaty, 2008).

Intensidad de la importancia	Definición	Explicación
1	Misma importancia	Ambos elementos contribuyen por igual al objetivo.
2	Débil o leve	
3	Importancia moderada	La experiencia y el juicio favorecen ligeramente un elemento respecto al otro.
4	Importancia moderada-fuerte	
5	Fuerte importancia	La experiencia y el juicio favorecen fuertemente un elemento respecto al otro.
6	Importancia fuerte - muy fuerte	
7	Importancia demostrada o muy fuerte	Se favorece fuertemente un elemento respecto al otro; su dominio está demostrado en la práctica.
8	Importancia muy, muy fuerte	
9	Importancia extrema	La evidencia que favorece un elemento respecto a otro es el más alto orden de afirmación posible.
Recíprocos de arriba		Si la actividad i tiene asignado uno de los números anteriores distintos de cero en comparación con la actividad j , entonces j tiene el valor recíproco en comparación con i .
1.1-1.9	Si las actividades son muy similares	A veces es difícil asignar el mejor valor de los anteriores. Cuando se compara un elemento con respecto a otros, el tamaño de los números pequeños no tiene un efecto destacable, pero aún así pueden utilizarse para indicar la importancia relativa de los elementos.

Figura 20: Escala de Saaty. Fuente: Saaty (1980)

De tal forma que cada elemento de un nivel superior se compara con los elementos del nivel inmediatamente inferior. Sea PW una de estas matrices de comparación por pares construida a partir de los datos generados por las respuestas a los cuestionarios, entonces pw_{ij} es la importancia del i elemento relativo al j y satisfice que: $pw_{ij} > 0$; la propiedad recíproca: $pw_{ij} = 1/pw_{ji}$; y $pw_{ii} = 1$ para todo i . Entonces, el vector de ponderaciones ω se determina mediante el método de vectores propios usando la siguiente ecuación:

$$\sum_{j=1}^n pw_{ij}w_j = \lambda_{max}w_i \quad (1)$$

donde λ_{max} es el valor propio más alto de PW y w es el vector propio normalizado asociado al valor propio principal de PW . Por último, el ratio de consistencia (CR por sus siglas en inglés *Consistency Ratio*) se calcula dividiendo el índice de consistencia (CI por sus siglas en inglés *Consistency Index*) por el índice de consistencia aleatorio (RI por sus siglas en inglés *Random Consistency Index*), que representa la consistencia de una matriz de comparación por pares generada aleatoriamente. El índice de consistencia se define como:

$$\frac{\lambda_{max} - n}{n - 1} \quad (2)$$

El RI fue obtenido por Saaty a partir de una muestra de tamaño 500, de una matriz recíproca generada aleatoriamente utilizando la escala $1/9, 1/8, \dots, 1, \dots, 8, 9$. El RI según el tamaño de la muestra de 1 a 10 se muestra en la Tabla 6.

n	1	2	3	4	5	6	7	8	9	10
RI	0	0	0.58	0.90	1.12	1.24	1.32	1.41	1.45	1.49

Tabla 6: RI según el tamaño de la muestra

En particular, si $CR \leq 0,1$ la inconsistencia es tolerable y puede asumirse, de lo contrario será necesario ajustar los valores de los ítems de comparación por pares, y los decisores tendrán que volver a emitir juicios hasta que sean coherentes. Saaty (1980) explicó por qué el CR no debe ser superior a 0,1. Consideró que la propia inconsistencia es importante, ya que sin ella no podrían admitirse nuevos conocimientos que generasen cambios en el orden de preferencias. Por otra parte, suponer que todos los conocimientos son coherentes es contradictorio con la idea de ajustar las preferencias en función de la comprensión y el conocimiento. Por tanto, el modelo depende de que se admita un cierto nivel de inconsistencia. Nos basamos en este resultado para medir la consistencia, se pueden encontrar ejemplos en la literatura de esta forma de medir la consistencia (Gil-Lafuente et al., 2014; Lin et al., 2020).

3.2.4 Obtención de las ponderaciones de las alternativas finales mediante varios niveles de agregación de ponderaciones

En este paso se obtienen las prioridades globales de las alternativas con el objetivo de calcular su importancia global. Para calcular estas ponderaciones globales, las ponderaciones de los subcriterios de un nivel se multiplican por el criterio correspondiente del nivel superior y se suman para cada alternativa de un nivel según los criterios a los que afecte. Estas prioridades globales se utilizan para ponderar las prioridades locales de las alternativas del nivel inferior y así sucesivamente hasta el nivel más bajo.

3.3 Cuadro de mando integral

El cuadro de mando integral (BSC por sus siglas en inglés Balanced Scorecard) fue desarrollado por Kaplan & Norton (1992), y es un modelo para medir el rendimiento estratégico de una organización. Su objetivo es traducir la misión y la visión de una organización en acciones operativas y de planificación estratégica mediante objetivos e indicadores. En el marco de las estrategias empresariales orientadas a mercado

mencionadas en el capítulo 1, el BSC permite que las organizaciones vinculen su desempeño (Salehzadeh et al., 2017) con las exigencias del entorno competitivo y las expectativas de los clientes. Propone cuatro áreas principales: finanzas, clientes, procesos internos y aprendizaje y crecimiento, que se explican a continuación.

- **Finanzas:** incluye indicadores tradicionales centrados en cuestiones tales como ingresos, márgenes de beneficio, retorno de la inversión, etc. Esta perspectiva se centra en los resultados financieros que son críticos para los accionistas.
- **Clientes:** se concentra en la satisfacción del cliente, la retención, la adquisición de nuevos clientes y la participación del mercado. Evalúa la percepción de la organización por parte de sus clientes y cómo esta puede mejorar para satisfacer sus necesidades.
- **Procesos internos:** evalúa la eficiencia y eficacia de los procesos internos de la organización. Se centra en los procesos clave que afectan directamente a la entrega de productos o servicios a los clientes.
- **Aprendizaje y crecimiento:** involucra el desarrollo y adquisición de habilidades y recursos necesarios para la innovación y la mejora continua. Incluye la capacitación del personal, la tecnología, la cultura organizacional, y otros factores que contribuyen al crecimiento a largo plazo.

Estas cuatro perspectivas proporcionan una visión holística, no meramente financiera, creando un conjunto más amplio de medidas que captan no sólo el rendimiento actual de la empresa, sino también los factores que impulsan su rendimiento futuro. El BSC no solo proporciona medidas e indicadores, sino que también establece una conexión clara entre la estrategia de la organización y las acciones operativas diarias. Al alinear los objetivos en estas cuatro perspectivas, las empresas pueden garantizar que todas las áreas cruciales estén siendo consideradas para lograr sus metas estratégicas. Este enfoque integral ayuda a evitar un peso excesivo de indicadores financieros a corto plazo y promueve una toma de decisiones más equilibrada y alineada con la visión a largo plazo de la organización.

BSC y AHP han sido empleados conjuntamente en aplicaciones en diversas áreas (Yu et al., 2012; Jeong-hwan & Geum, 2016; Yaghoobi & Haddadi 2016; Modak et al., 2017) obteniendo interesantes resultados y metodologías.

3.4 Técnicas de agrupamiento de datos geográficos

Las técnicas de agrupamiento de datos geográficos permiten la agrupación de elementos con ubicaciones similares en clústeres o conjuntos, basándose en la proximidad espacial de los mismos. El objetivo es que elementos dentro de un mismo clúster sean más similares entre sí que con aquellos elementos que estén en otros clústeres, facilitando así la identificación de patrones espaciales y la exploración de la estructura subyacente en los datos. Para realizar un agrupamiento, hay que seguir los siguientes pasos: (i) selección de las variables en base a las cuales se van a agrupar los elementos y del tipo de distancia o medida de similitud que se va a utilizar y (ii) selección de la técnica de agrupamiento más adecuada para el caso de negocio concreto.

Un uso muy extendido del agrupamiento de datos geográficos es el posicionamiento de servicios en base a la distribución espacial del público objetivo.

3.4.1 Selección de las variables de agrupamiento y del tipo de distancia

Las variables utilizadas para realizar el agrupamiento son las coordenadas geográficas de la ubicación de los elementos a agrupar tanto en el sistema de coordenadas universal transversal de Mercator (UTM, por sus siglas en inglés *Universal Transverse Mercator*) como en el sistema geodésico de coordenadas geográficas (WGS84, por sus siglas en inglés *World Geodetic System 1984*).

Con estas coordenadas se lleva a cabo la matriz de distancias empleada en el agrupamiento. De forma general, la distancia $d(x_r; x_s)$, entre dos puntos se representa como:

$$d(x_i; y_j) = \left(\sum_{i=1}^n |x_i - x_j|^p \right)^{\frac{1}{p}}$$

donde, para $p = 1$, la expresión coincide con la distancia Manhattan y para $p = 2$, coincide con la distancia euclídea (Fernández-Avilés & Montero, 2024).

En el caso de uso que se desarrollado en el capítulo 7 vamos a aplicar estas técnicas para el posicionamiento de servicios turísticos en una ciudad, concretamente oficinas de turismo. En este caso, el público objetivo serán los turistas y para tener una referencia de su ubicación emplearemos la ubicación de los hoteles en la ciudad. Por tanto, a la hora de seleccionar la distancia a utilizar en este caso debemos tener en cuenta que nuestra

distancia debe reflejar fielmente la proximidad o cercanía geográfica que hay entre dos puntos de una ciudad. La distancia más corta entre dos puntos sería la euclídea, otra opción sería la distancia Manhattan que será siempre algo mayor. No obstante, debemos considerar que estas distancias serán recorridas por un turista que generalmente realizará el trayecto a pie, por lo que va a afectar el urbanismo de la ciudad, el trazado de sus calles o las barreras arquitectónicas aumentando la distancia recorrida final (véase Figura 21).

En la Figura 21 se puede observar la diferencia entre la distancia euclídea (línea negra) y la distancia recorrida a pie (línea azul), los puntos mostrados son la oficina de turismo ubicada en la estación de tren de Atocha y el hotel AC Atocha y se han obtenido empleando la herramienta Google Maps (Google Maps, 2024) a la que se le solicita la distancia a pie entre estos dos puntos. Se puede observar que la distancia a pie son 450 metros, muy superior a la distancia euclídea que son 255,8 metros.



Figura 21: Distancias euclídea (negro), distancia caminando a pie (azul) entre el Hotel AC Atocha y la oficina de turismo de la estación de Atocha. Fuente: elaborado con Google Maps.

Vamos a utilizar, por tanto, la distancia a pie que representa mejor las distancias que se encuentra un turista en la ciudad (señalar que esta distancia cumple las propiedades de una función distancia de un espacio métrico). Dado que el objetivo del caso de uso es ubicar servicios turísticos próximos geográficamente a los turistas para que proporcionen un servicio de cercanía, resulta crucial definir con exactitud lo que consideramos una distancia cercana o próxima. Según Perry (1929), se considera que la distancia de 5 minutos a pie,

también conocida usualmente por los términos en inglés como: *5-minutes walk*, *pedestrian shed* o *walkable catchment*, es la distancia que la gente está dispuesta a caminar antes de optar por usar un medio de transporte. Basándose en la velocidad media al caminar, un paseo de cinco minutos se representa por un radio de 400 metros (Azmi et al., 2012). Tomamos, por tanto, un radio de 400 metros como distancia que implica cercanía y, en lo sucesivo, nos referiremos a esta distancia con el término: 'wd-5' (del inglés *walkable distance-5 min*).

Por ejemplo, en base a esta definición, podemos decir que la oficina de turismo ubicada en la Puerta del Sol de Madrid da un servicio de proximidad o cercanía a los turistas ubicados en el hotel Regina Madrid, dado que la distancia entre ambos es 350 metros y es inferior a 'wd-5' que son 400 metros (ver Figura 22). La distancia se ha calculado empleando la herramienta Google Maps (Google Maps, 2024) a la que se le solicita la distancia a pie entre estos dos puntos.

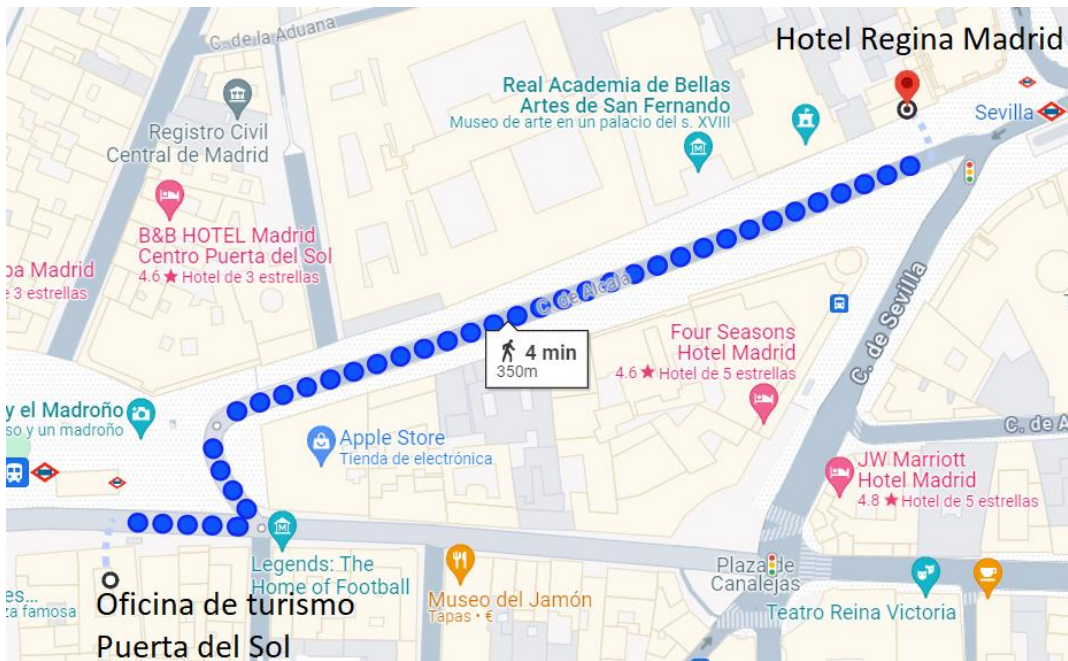


Figura 22: Distancia a pie entre la oficina de información turística de la Puerta del Sol de Madrid y el hotel Regina Madrid. Fuente: elaborado con Google Maps

3.4.2 Selección de la técnica de agrupamiento

Hay distintos tipos de algoritmos de agrupamiento, los podemos clasificar en los siguientes grandes tipos:

- **Jerárquico** (Bridges, 1966), que genera los clústeres de forma secuencial y trabaja sobre una estructura jerárquica. Pueden ser divisivos o aglomerativos.

- **Por particionamiento**, basado en prototipos (Kanungo et al., 2002), que genera clústeres que tienen representantes o prototipos del mismo que determinan el comportamiento de los mismos.
- **Basado en una distribución** (Yu et al., 2017), en el cual los datos forman parte de un clúster según la probabilidad de que un punto pertenezca a un clúster determinado.
- **Basado en densidad** (Khan et al., 2014), que genera agrupamientos por áreas de alta densidad. En esta técnica los clústeres pueden tener formas arbitrarias siempre que correspondan con áreas densas.

Tal y como hemos señalado anteriormente en el capítulo 7 se desarrolla un caso de uso sobre posicionamiento de servicios turísticos en una ciudad, concretamente oficinas de turismo. En este caso de uso vamos a emplear algoritmos de clustering y consideramos que dentro de este tipo de algoritmos los más apropiados para nuestro caso de uso son los algoritmos de agrupamiento jerárquico por aglomeración. Este algoritmo genera un conjunto de clústeres anidados organizados como un árbol jerárquico que se visualizan en forma de dendograma (véase Figura 23). Las razones para seleccionar este tipo de algoritmo son las siguientes:

- Se trata de un algoritmo determinístico. Es una característica que nos interesa especialmente para que el modelo sea reproducible para un conjunto de datos dado.
- Posibilidad de trabajar con una medida de distancia personalizada, en nuestro caso la distancia caminando a pie, calculando para ello previamente la matriz de distancias.
- Facilidad en la interpretación de los resultados. Los dendogramas generados por el agrupamiento jerárquico son sencillos de interpretar, ayudan a identificar la estructura jerárquica del conjunto de datos y a visualizar las relaciones entre los clústeres.
- Determinación del número de clústeres. En este tipo de algoritmos no es necesario especificar el número de clústeres a priori. Esto es especialmente interesante, ya que queremos dejar variable el número de clústeres y una vez creado el dendograma se pueden seleccionar distintos números de clústeres mediante técnicas de corte. Es decir, no necesitamos repetir el proceso de agrupamiento si cambiamos el número de clústeres, toda la información está en los dendogramas.

Igualmente, este algoritmo no hace ninguna asunción sobre el número de clústeres. A partir del dendograma, podemos obtener cualquier número de clústeres 'cortando' por el sitio adecuado. En la Figura 23 observamos el 'corte' necesario para generar tres clústeres (línea roja) y cuatro clústeres (línea azul).

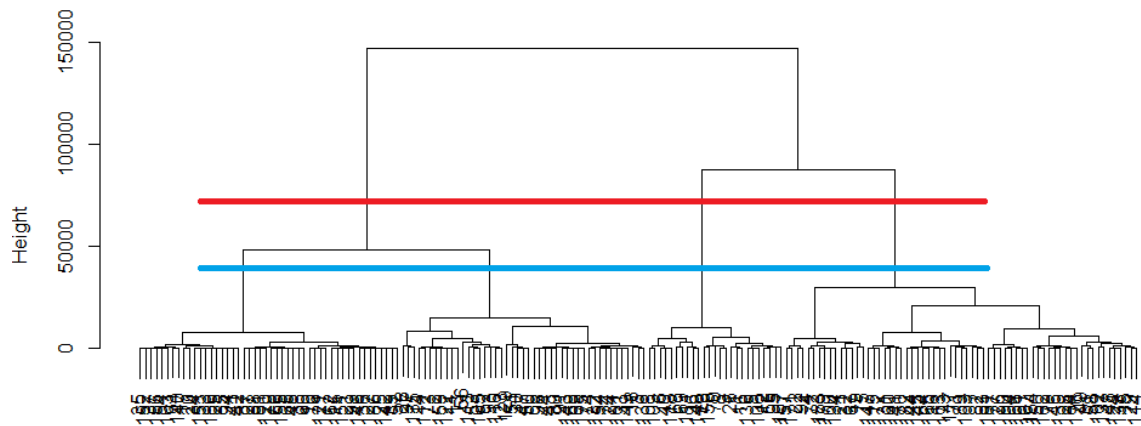


Figura 23: Ejemplo de dendrograma con distintos de puntos de corte para obtener diferente número de clústeres en un dendrograma. Fuente: elaboración propia.

El agrupamiento por aglomeración es la aproximación más frecuente en este tipo de algoritmos. Existen diferentes métodos para realizar el agrupamiento jerárquico aglomerativo (Müllner, 2011) que se basan en las diferentes posibles medidas de distancia que podemos considerar entre dos grupos. En la Tabla 7 resumimos los diferentes métodos.

Método	Definición	Expresión
Single	Dados dos clústeres C_i y C_j , es la distancia mínima entre cualquier punto perteneciente a C_i y cualquier punto perteneciente a C_j .	$D(C_i, C_j) = \min_{x,y} \{ d(x,y) x \in C_i, y \in C_j \}$
Complete	Dados dos clústeres C_i y C_j , es la distancia máxima entre cualquier punto perteneciente a C_i y cualquier punto perteneciente a C_j .	$D(C_i, C_j) = \max_{x,y} \{ d(x,y) x \in C_i, y \in C_j \}$
Average	Dados dos clústeres C_i y C_j , es la distancia media entre los puntos pertenecientes a C_i y los puntos pertenecientes a C_j .	$D(C_i, C_j) = \frac{1}{ C_i C_j } \sum_{x \in C_i, y \in C_j} d(x,y)$
Centroid	Dados dos clústeres C_i y C_j , es la distancia entre el centroide R_i de C_i y el centroide R_j perteneciente a C_j .	$D(C_i, C_j) = d(R_i, R_j)$
Median	Variante del método centroides, trabaja igual, pero difiere en el cálculo de los centroides. El centroide de un clúster C_{ij} se calcula como la media de los centroides del clúster de los que proviene C_i y C_j .	<p>Similar a la fórmula de Centroid, pero el centroide R_{ij} de C_{ij} se calcula partiendo del centroide R_i de C_i y el centroide R_j de C_j como:</p> $R_{ij} = \frac{R_i + R_j}{2}$
Ward	Dados dos clústeres C_i y C_j , es la diferencia entre la suma de cuadrados intracluster para los dos clústeres de manera independiente y la suma de cuadrados intracluster para el clúster resultante de combinar ambos clústeres para dar lugar a C_{ij} .	$D(C_i, C_j) = \sum_{x \in C_{ij}} (x - R_{ij})^2 - \sum_{x \in C_i} (x - R_i)^2 - \sum_{x \in C_j} (x - R_j)^2$ <p>Siendo R_i el centroide de C_i, R_j el centroide de C_j y R_{ij} el centroide de C_{ij}</p>
Mcquitty	Se define recursivamente, donde en cada paso del algoritmo la distancia de los clústeres C_i y C_j fusionados más recientemente C_{ij} con respecto a cualquier otro clúster C_k es la media entre las distancias de C_i a C_k y de C_j a C_k	$D(C_{ij}, C_k) = \frac{D(C_i, C_k) + D(C_j, C_k)}{2}$

Tabla 7: Métodos de clustering jerárquico

3.5 Grandes modelos del lenguaje

Los avances en las técnicas de aprendizaje automático, la gran disponibilidad de datos y la capacidad de cómputo actual han dado lugar a un gran avance en la inteligencia artificial con la aparición de modelos de nueva generación que han alcanzado un rendimiento sin precedentes en tareas como la respuesta a preguntas y la generación de lenguaje natural. Esta nueva generación de modelos llamados grandes modelos de lenguaje (conocidos ampliamente por sus siglas en inglés LLM, *Large Language Models*) son modelos de procesamiento de lenguaje natural (Aditya et al., 2018) basados en arquitecturas neuronales profundas (Patel et al., 2018) que se entrenan con grandes cantidades de texto. Estos modelos son capaces de generar texto, completar frases, responder preguntas, traducir entre idiomas, resumir contenido y realizar muchas otras tareas relacionadas con el lenguaje. La evolución de los LLM en los últimos 5 años ha sido exponencial y su rendimiento en un conjunto notable de tareas diferentes ha llegado a ser muy destacado. Antes de 2017, la mayoría de los modelos de Procesamiento del Lenguaje Natural se entrenaban mediante aprendizaje supervisado para tareas concretas y solo podían utilizarse esas mismas tareas. Estas limitaciones fueron superadas con la arquitectura de red de autoatención, también conocida como Transformer, (Vaswani et al., 2017) que se introdujo a finales de 2017 y se utilizó para desarrollar dos modelos que fueron disruptivos: Bidirectional Encoder Representations from Transformers (BERT) y Generative Pretrained Transformer (GPT) (Devlin et al., 2018; Radford et al., 2023). Ambos modelos consiguieron una capacidad de generalización superior, mediante una combinación de preentrenamiento no supervisado y ajuste fino supervisado, logrando aplicar representaciones lingüísticas preentrenadas a tareas posteriores. Los modelos GPT han evolucionado rápidamente, la tercera versión de GPT (GPT-3) tiene 175.000 millones de parámetros y es 100 veces mayor que GPT-2.

ChatGPT (ChatGPT, 2022) es un gran modelo de lenguaje entrenado por OpenAI, una compañía de investigación y despliegue de inteligencia artificial, lanzado en una versión preliminar de investigación gratuita el 30 de noviembre de 2022. Es un asistente conversacional o *chatbot* (término en inglés ampliamente utilizado) de inteligencia artificial optimizado para el diálogo, para interactuar en una conversación similar a la humana. La interacción con este chatbot se realiza mediante entradas o solicitudes generalmente conocidas por su término en inglés *prompt*. Un *prompt* es una instrucción o entrada inicial que se proporciona a ChatGPT, se trata de la pregunta, frase o fragmento de texto que se introduce para obtener una respuesta o salida del modelo.

Elegir el prompt adecuado es a menudo crucial para obtener la respuesta o salida deseada. Una formulación ligeramente diferente del mismo prompt puede llevar a respuestas diferentes, ya que guía al modelo en una dirección particular o establece un contexto específico para la respuesta. En este sentido, la ingeniería de prompts (Henrickson & Merono-Penuela, 2023) se refiere a la práctica de diseñar y refinar cuidadosamente los prompts o instrucciones iniciales que se dan a un modelo de lenguaje para obtener respuestas específicas o deseables.

En el caso de uso desarrollado en el capítulo 7 trabajaremos con ChatGPT-3.5 (disponible públicamente desde noviembre de 2022). Existen versiones posteriores como ChatGPT-4.0 (disponible comercialmente, desde marzo de 2023) y ChatGPT-4o (disponible comercialmente, desde mayo de 2024). Consideramos que la versión ChatGPT-3.5 ha sido ya extensamente utilizada en múltiples áreas de investigación y genera resultados fiables para el objetivo que pretendemos.

4. Marco conceptual propuesto

En este capítulo abordamos el objetivo general de la tesis, que es la definición de un marco conceptual para la toma de decisiones basadas en datos en una organización turística. Para conseguir este objetivo, partiendo del marco de referencia inicial definido en el capítulo 2 se analizarán las carencias que este marco podría tener de cara a abordar los desafíos actuales del sector del turismo con una estrategia orientada al dato. Estas carencias, brechas o necesidades se traducirán en una serie de gaps que serán tenidos en cuenta para definir un marco conceptual que permita abordar de forma eficiente estrategias basadas en datos en las organizaciones turísticas.

4.1. Identificación de gaps

El marco de referencia inicial obtenido en el capítulo 2 representa el uso de fuentes de datos en turismo bajo un marco temporal de 1982 a 2021. Nos ha permitido conocer la situación de las organizaciones turísticas respecto a las áreas definidas (fuentes de datos, aplicación de técnicas de ciencia de datos e inteligencia artificial y generación de análisis y descubrimientos relevantes) y nos sirve como evaluación inicial. Teniendo como objetivo final el definir un marco conceptual para la toma de decisiones basadas en datos en una organización turística, vamos a identificar brechas, carencias o deficiencias que tiene este marco de referencia inicial que se traducirán en una serie de necesidades que se deben cubrir para poder abordar los desafíos actuales del sector del turismo con una estrategia orientada al dato. Dado que este marco de referencia inicial refleja un marco temporal amplio, pero que termina en 2021 es lógico esperar que muchas de los gaps a identificar tengan que ver con la adopción de técnicas actuales relativas a la gestión y análisis de datos, así como su aplicación a temas de actualidad y tendencias en la industria del turismo. Adicionalmente, es previsible que surjan otros gaps asociados con la situación general del sector turístico en cuanto a su orientación hacia el uso de datos. Estos gaps podrían reflejar áreas en las que el sector turístico aún no ha experimentado un desarrollo completo o no ha alcanzado el nivel de madurez necesario desde la perspectiva de aprovechar plenamente el potencial de los datos. Este segundo tipo de gaps se detectan en la comparativa del sector turístico con otros sectores de actividad en el uso de los datos. Gaps que pueden deberse, entre otras cosas, a la naturaleza heterogénea y multidisciplinar de la industria turística, que involucra diversas disciplinas y, por otra parte, a la naturaleza atómica del sector, configurado por muchas pequeñas empresas, en su mayoría inmersas en un proceso de digitalización. Realizaremos este diagnóstico

centrándonos en las cuatro grandes áreas identificadas en el marco de referencia inicial.

Fuentes de datos

En esta área, nos centramos en las fuentes de datos empleadas en el sector del turismo. En el marco de referencia inicial se identifican fuentes de datos de diferente tipología (UGC, redes geosociales, sistemas de información turística, etc.), pero sin una organización clara. Creemos útil establecer una clasificación que nos facilite su organización y gestión posterior. Una clasificación interesante que podemos utilizar es la planteada por Li et al. (2018), que identifica tres grandes bloques:

- **Operaciones:** datos de transacciones u operaciones como reservas de hotel, pagos, transporte (vuelos, cruceros, transporte ferroviario), visitas a sitios web, etc.
- **Dispositivos:** datos procedentes de dispositivos, principalmente datos móviles generalmente geolocalizados, pero también datos IoT procedentes de sensores u otros dispositivos.
- **Usuarios:** datos de tipo UGC, tales como comentarios en redes sociales, reseñas en línea, motores de búsqueda, comunidades virtuales, cocreación de experiencias turísticas, imágenes, etc.

Tal y como se ha indicado en el apartado 2.3.4 las fuentes de los bloques 'dispositivos' y 'usuarios' son los que más han crecido en los últimos años y son fuentes de datos muy utilizadas actualmente en el sector del turismo y cuya correcto almacenamiento y gestión de crítico y debe ser contemplado en nuestro marco conceptual. De igual forma, hay que tener en cuenta que mientras en el primer bloque (operaciones) los datos son estructurados (formato estándar y estructura bien definida), en los otros dos bloques (dispositivos y usuarios) podemos encontrar datos semiestructurados o no estructurados, lo que debe tenerse en cuenta en su gestión y almacenamiento.

Además de los tres bloques de datos mencionados anteriormente, podemos añadir un cuarto bloque que lo constituirían los datos de tipo estadístico generados y compartidos en abierto, generalmente por instituciones públicas. Aquí encontramos datos estadísticos generados por instituciones especializadas o datos abiertos proporcionados por administraciones locales o estatales u organizaciones internacionales. Por ejemplo, información sobre el grado de ocupación de un destino, procedencia de los turistas, gasto y otros. Este tipo de datos abiertos ha crecido significativamente en la última década y se utiliza en múltiples aplicaciones (Bratucu & Cismaru, 2015). Se trata de información compartida que democratiza el acceso a los datos de todos los agentes públicos y privados

del sector, como señalan Celdrán-Bernabeu et al. (2018). En el sector turístico, en la última década, se han desarrollado diferentes sistemas de inteligencia turística (Gajdosik, 2019), algunos a nivel de la administración estatal y otros a nivel local, que recogen y generan información de gran interés. Tal y como se ha mencionado en el capítulo **¡Error! No se encuentra el origen de la referencia.**, cabe destacar la reciente creación de espacios de datos compartidos en el sector turismo entre los que cabe destacar el ecosistema digital europeo Gaia-X (Gaia-X, 2023; Braud et al., 2021), y los proyectos de espacios de datos turísticos europeos (Tourism Data Space, 2023; DATES, 2023). Así mismo y centrándonos en las iniciativas españolas, hay que destacar el desarrollo de los sistemas de inteligencia turística y la creación de la Plataforma inteligente de destinos (Plataforma inteligente de destinos, 2023). En el análisis bibliométrico y en marco de referencia inicial no se detecta un uso importante de este tipo de datos en abierto, pero es un tipo de información compartida que va a cobrar mayor relevancia en los próximos años y las organizaciones turísticas deben estar preparadas para acceder a este tipo de espacios de datos.

Los diferentes tipos de fuentes de datos identificados provienen de diferentes orígenes que muestran perspectivas complementarias de la industria turística. Esta diversidad refleja una característica que tiene la industria turística como es la complejidad inherente derivada de los múltiples agentes públicos y privados que intervienen en la cadena de valor de la misma. En este sentido, y tal y como se indicó en el capítulo **¡Error! No se encuentra el origen de la referencia.**, la toma de decisiones efectiva en el sector del turismo debe considerar una multiplicidad de factores y criterios interrelacionados. En este contexto, la aplicación de modelos MCDM resulta de gran utilidad para abordar los desafíos particulares asociados con la gestión y la planificación en la industria turística. Estos modelos proporcionan un marco estructurado para evaluar y comparar alternativas en situaciones donde las decisiones no pueden reducirse a un único criterio. Estas alternativas se valoran en base a criterios que son relevantes para la decisión en cuestión. En estos modelos hay un tipo de información que resulta esencial como es la 'información de expertos' que son las evaluaciones proporcionadas por individuos con experiencia y conocimientos en el área relevante. Estos expertos son consultados para evaluar las alternativas en relación con los criterios establecidos para la toma de decisiones. La información de expertos es esencial para asignar pesos a los criterios, ponderar la importancia relativa de cada uno y proporcionar puntuaciones o clasificaciones a las alternativas en función de su desempeño en esos criterios. La combinación de las evaluaciones de expertos y la aplicación de métodos específicos de toma de decisiones multicriterio contribuye a obtener una decisión fundamentada y bien informada. No encontramos en el análisis bibliométrico, ni en el marco de referencia inicial un uso

importante de los modelos MCDM en la toma de decisiones en turismo, pero consideramos esencial su uso y por ello, añadir la 'información de expertos' en un quinto bloque de fuentes de datos.

Se identifican, por tanto 5 bloques de fuentes de datos que representan fielmente las tendencias actuales en uso de fuentes de datos en turismo y que deberán reflejarse en el marco conceptual definitivo.

Almacenamiento y gestión del dato

Una vez definidas las fuentes de datos, en esta área, nos centramos en los aspectos relacionados con su almacenamiento y gestión. En primer lugar, señalar que en el análisis bibliométrico realizado se detecta un uso escaso de almacenes de datos o repositorios de datos centralizados en el sector turístico, siendo este un aspecto crítico en la orientación al dato de una compañía. Por otra parte, tenemos en el sector turístico una amplia diversidad de tipología de fuentes identificada en los 5 grandes bloques del área anterior. Esta diversidad de fuentes muestra perspectivas diferentes y complementarias de la actividad turística cuya integración en repositorios centralizados permitiría tener una visión global.

En lo relativo a los aspectos técnicos es necesario disponer de almacenes de datos que permitan trabajar con diferentes tipos de fuentes de datos, algunas de ellas estructuradas y otras no estructuradas, tal y como se comentó en el área de fuentes de datos. Existen en el mercado diferentes tipos de bases de datos (Sahatqija et al., 2018), para poder almacenar diferentes tipos de datos: bases de datos relacionales, bases de datos clave-valor, bases de datos orientadas a grafos, etc., lo cual permite adaptarse a las diferentes tipologías de fuentes de datos identificadas en el área de fuentes de datos. Estas diferentes bases de datos se pueden agrupar bajo dos grupos: bases de datos relacionales o que emplean el lenguaje de consulta estructurado, comúnmente referido por SQL (por sus siglas en inglés de *Structured Query Language*) y bases de datos NoSQL (abreviatura en inglés de *Not only SQL*). No hemos encontrado en el marco de referencia inicial referencias a estos tipos de bases de datos y tampoco hemos detectado en el análisis bibliométrico un uso extendido de esta diferente tipología de bases de datos en el sector del turismo, sin embargo, consideramos que su uso podría ser de gran utilidad. Mención especial para la información de tipo texto muy común en el sector turístico (opiniones, reseñas, comentarios redes sociales) y que tiene un gran potencial en su análisis. Este tipo de información está siendo muy utilizada actualmente por la inteligencia artificial generativa

con grandes resultados. Lógicamente, dado el marco temporal que comprende nuestro marco de referencia inicial no aparece en el mismo la inteligencia artificial generativa, pero está teniendo múltiples aplicaciones en el sector del turismo (Dwivedi et al., 2023). Para poder emplear esta información de texto en aplicaciones de inteligencia artificial generativa puede ser de gran utilidad disponer de bases de datos vectoriales. Este tipo de bases de datos facilitan el manejo de información de texto y su integración con los LLM (Topsakal & Akinci, 2023), modelos muy utilizados en el ámbito de la inteligencia artificial generativa.

De igual forma, las fuentes de datos se diferencian claramente en su velocidad de generación y necesidades de tratamiento. Por ejemplo, algunas fuentes tienen una velocidad de generación muy alta, como los datos UGC, para los que puede existir necesidad de gestionarlos en tiempo real (Ranganathan et al., 2020), mientras que otras fuentes como las estadísticas disponibles en datos abiertos sobre afluencia de turistas en un destino concreto son datos con una frecuencia de publicación mucho mayor (por ejemplo, mensual) y lejana al tiempo real. Las fuentes con mayor velocidad de generación del dato, generan altos volúmenes de datos que deben ser gestionados con una capacidad de almacenamiento y procesado adecuadas. Es decir, estos almacenes de datos deben disponer de las necesarias capacidades de almacenamiento y procesado de la información y deben estar preparadas para el continuo crecimiento de los datos, esto último hace referencia a sus capacidades de escalabilidad.

Un aspecto relacionado con la gestión de los datos y que no hemos detectado como significativo en el análisis bibliométrico, ni en el marco de referencia inicial es todo lo relativo a la disciplina de gobierno del dato (Yallop et al., 2021). El gobierno del dato se refiere a las políticas, procesos y controles establecidos para garantizar la calidad, integridad, seguridad y uso adecuado de los datos dentro de una organización. Implica la gestión eficiente y responsable de los datos a lo largo de su ciclo de vida, algo crítico para una correcta orientación al dato. Un aspecto fundamental del gobierno del dato es la gestión del catálogo de datos que actúa como un inventario estructurado de conjuntos de datos disponibles, describiendo sus metadatos, esquemas y relaciones. La gestión de los catálogos de datos puede ayudar a las empresas del sector turístico a participar de manera más efectiva en el intercambio de información, por ejemplo, en los espacios de datos compartidos del sector turístico ya mencionados anteriormente. Esto se traduce tanto en la capacidad de acceder y utilizar datos de relevantes para sus operaciones (como consumidores de datos) como en la contribución activa al suministrar datos valiosos a los espacios compartidos (como productores de datos). Igualmente, estos catálogos pueden

facilitar el uso de normas comunes, protocolos, semánticas u ontologías que están siendo definidas e impulsadas por los espacios de datos compartidos para facilitar la interoperabilidad.

En resumen, nos encontramos una tipología de fuentes de datos diversa cuyas necesidades de tratamiento y gestión deben ser contempladas en los almacenes de datos y no encontramos en nuestro marco de referencia inicial, ni en el análisis bibliométrico una respuesta clara a estas necesidades.

Técnicas de ciencia de datos e inteligencia artificial

En lo referente a las técnicas de ciencia de datos e inteligencia artificial, podemos ver en el marco de referencia inicial una variedad de técnicas muy amplia. Lo primero que podemos destacar es la clara evolución temporal desde técnicas más clásicas como puede ser la segmentación o la clasificación hacia técnicas más actuales como algoritmos avanzados de aprendizaje máquina, el aprendizaje profundo, los sistemas de recomendación o el procesado de lenguaje natural, siendo este segundo grupo de técnicas las que más desarrollo están teniendo y que deben ocupar un papel predominante en el marco conceptual definitivo. Dentro de las técnicas ampliamente utilizadas en el sector del turismo cabe mencionar igualmente aquellas relacionadas con la inteligencia de negocio (Bustamante et al., 2020) tales como la generación de informes y cuadros de mando para el seguimiento de indicadores de actividad turística. Encontramos menciones a estas técnicas en el marco de referencia inicial en las referencias a los sistemas de información turística y al rendimiento. Las técnicas de inteligencia de negocio deben quedar bien reflejadas en el marco conceptual definitivo y deben beneficiarse de un almacén de datos integrado (descritos en el área de almacenamiento y gestión del dato) para una explotación óptima de la información.

En el área de fuentes de datos identificamos 5 bloques de fuentes de datos, dentro de la cuales hay dos que están experimentando un crecimiento mayor: 'dispositivos' y 'usuarios'. Ambos tipos de fuentes tienen un gran potencial de cara a aplicar técnicas analíticas. En la información procedente de dispositivos encontramos a menudo información geolocalizada procedente de dispositivos móviles sobre la que se pueden aplicar técnicas de analítica geoespacial (Yang et al., 2012) para conocer el movimiento o afluencia de los turistas, si bien aparecen en el análisis bibliométrico, deben tener una relevancia especial en el marco conceptual definitivo. El segundo bloque mencionado es el que tiene la información generada por los usuarios que puede ser de diferente

naturaleza como, por ejemplo, texto, imágenes o video. Estos tipos de dato son analizados con técnicas que han tenido un gran desarrollo en los últimos años. En lo relativo a imágenes o video, encontramos toda la evolución experimentada por el aprendizaje profundo basado en redes neuronales de diferentes tipologías (Essien & Chukwukelu, 2022). Relativo a la información de texto las técnicas relacionadas con el procesado de lenguaje natural para el análisis de textos, también han experimentado una notable evolución en los últimos años con la aparición de técnicas novedosas como el mecanismo de atención (Vaswani et al., 2017) o los lenguajes preentrenados (Devlin et al., 2018), entre otras técnicas, que han permitido dar un salto significativo en el análisis de textos. En este sentido, y con el objetivo de conocer de forma más profunda el comportamiento del turista, son especialmente interesantes en el ámbito de la inteligencia artificial las técnicas de multimodalidad (Hayashi et al., 2021) que permiten comprender y procesar información proveniente de múltiples modalidades o fuentes. Estas modalidades suelen estar relacionadas con diferentes tipos de datos, como texto, imágenes, audio, video, entre otros. Muchas de estas técnicas mencionadas no aparecen en el marco de referencia inicial debido al marco temporal que abarca, pero son técnicas recientes e innovadoras que deben ser contempladas en el marco conceptual definitivo. Algo similar sucede con la inteligencia artificial generativa que ha tenido un impacto muy considerable en la sociedad y encuentra múltiples aplicaciones en el sector del turismo (Carvalho & Ivanov, 2023).

Por último, en esta área de técnicas de ciencia de datos, mencionar los modelos MCDM, cuya importancia y aplicación en turismo ya han sido analizadas en el capítulo 1. Estos modelos se podrán combinar con el resto de técnicas mencionadas.

Generación de análisis y descubrimientos relevantes para la toma de decisiones

En el marco de referencia inicial encontramos múltiples aplicaciones en el uso de los datos. Estas múltiples aplicaciones o casos de uso reflejan todos los agentes que participan en el sector turístico y las múltiples perspectivas desde las que se puede analizar: turista, destino turístico, servicios, empresas, etc. Dada la variedad de casos y la complejidad en su gestión, consideramos que mejoraría nuestro marco de referencial el disponer de una clasificación en este tipo de aplicaciones que ayude a gestionarla.

Por un lado, detectamos aplicaciones centrados en el turista, como las que analizan su comportamiento, (Miah et al., 2017), sus percepciones (Nave et al., 2018) su satisfacción (Li et al., 2020) o proponiendo recomendaciones y personalizando servicios (Esmaeili et

al., 2020). Todos estos casos de uso intentan cubrir las diferentes fases del ciclo de vida del viaje (antes, durante y después). Por otro lado, encontramos otra perspectiva que serían los casos de uso centrados en las organizaciones, como puede ser un destino turístico o las empresas que ofrecen servicios turísticos. Ejemplos de estos casos de uso, pueden ser: la previsión de la demanda (Li & Jiao, 2020), la gestión de la sostenibilidad (De Marchi et al. 2022) o el análisis de la reputación (Cillo et al., 2019).

Esta doble perspectiva de casos de uso (turista y organizaciones) se refleja también en las iniciativas relacionadas con la creación de ecosistemas digitales para ofrecer servicios integrados de valor en destinos turísticos. Un ejemplo puede ser la, anteriormente mencionada, iniciativa española de creación de una plataforma (Plataforma inteligente de destinos, 2023) que tiene como objetivo integrar, relacionar y combinar datos públicos y privados para generar más inteligencia competitiva en los ecosistemas de los destinos españoles, activando, la innovación continua, interconectando y atendiendo las necesidades de todos los agentes de la cadena de valor turístico: turista, destinos y empresas.

4.2. Marco conceptual para la toma de decisiones basada en datos en una organización turística

En este apartado proponemos el marco conceptual para la toma de decisiones basadas en datos en una organización turística. Este marco conceptual trata de dar respuesta a los gaps identificados en el apartado 4.1 sobre el marco de referencia inicial para poder abordar los desafíos actuales del sector del turismo con una estrategia orientada al dato. En la Figura 24 podemos observar el marco conceptual propuesto.

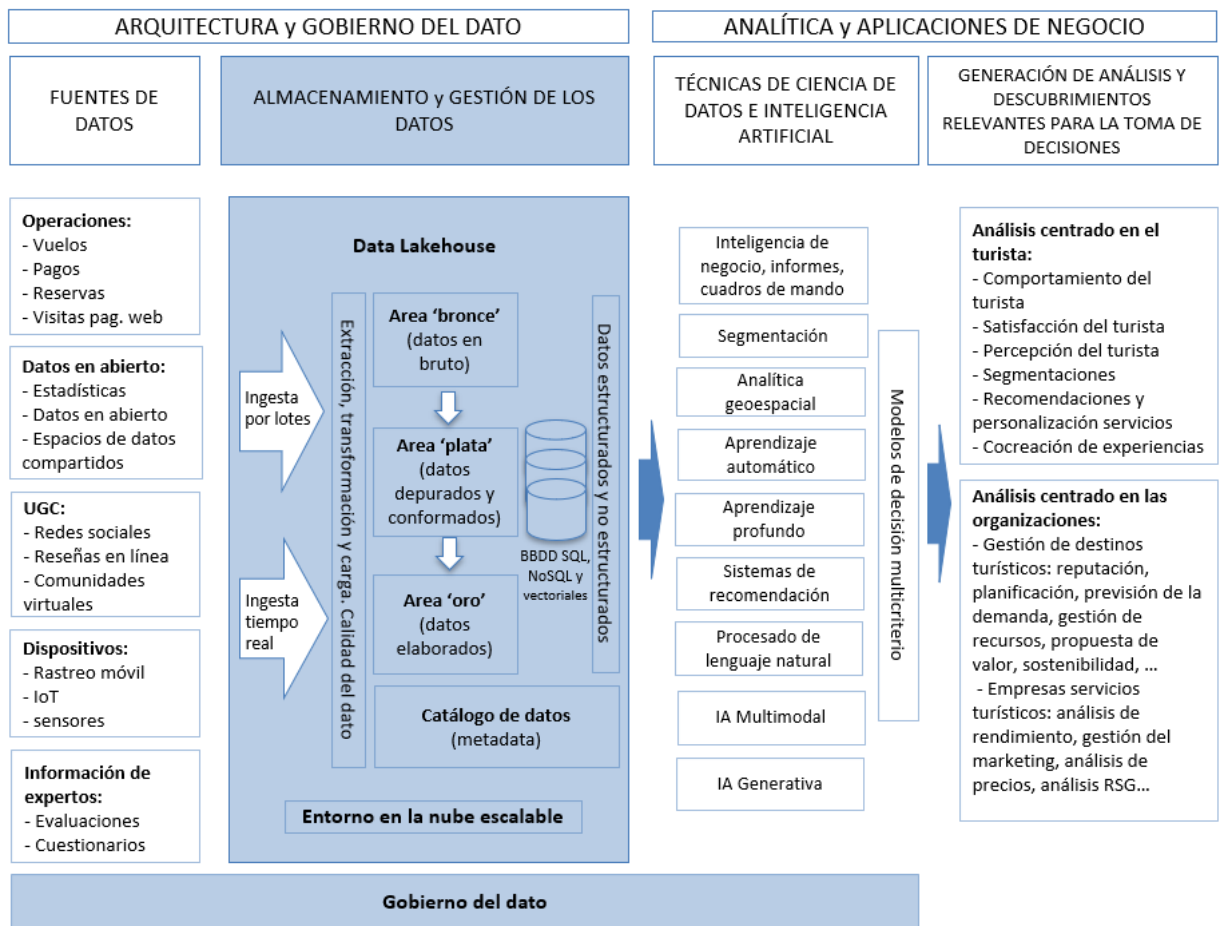


Figura 24: Marco conceptual para la toma de decisiones basada en datos en una organización turística. Fuente: elaboración propia.

A continuación, se explican cada una de las áreas del marco conceptual propuesto.

Fuentes de datos

Agrupamos las fuentes de datos en los **cinco grandes bloques** identificados en el apartado 4.1 (operaciones, datos en abierto, UGC, dispositivos e información de expertos).

Almacenamiento y gestión del dato

Se propone un entorno de almacenamiento y gestión del dato basado en una arquitectura de tipo **data lakehouse** con las siguientes características:

- Organización de los datos en **diferentes áreas** con diferente grado de **elaboración del dato**, lo cual facilita la realización de diferentes tipos de análisis sobre los mismos. De esta manera, los datos están preparados para las diferentes técnicas a aplicar en el apartado de técnicas de ciencia de datos e inteligencia artificial

- Capacidad para **gestionar datos tanto estructurados como no estructurados**, adaptándose a la variada tipología de fuentes indicada en el área de fuentes de datos.
- Capacidad para la **ingesta de datos en procesos periódicos** (diarios, semanales, etc.) como en **tiempo real**, adaptándose a la variada frecuencia de actualización de fuentes indicada en el área de fuentes de datos.
- Trabajo con **diferentes tipos de bases de datos**: relacionales, NoSQL y vectoriales para poder almacenar fuentes de datos de diferentes tipos y para realizar diferentes tipos de análisis de datos
- Se añade una **capa de gobierno del dato** para facilitar las acciones relacionadas con esta disciplina indicadas en el apartado 4.1 en la identificación de gaps
- **Arquitectura en la nube con capacidad para escalar** y desplegar un mayor número de recursos de computación y adaptarse a los cambiantes requisitos de almacenamiento y gestión del dato.

Cabe señalar que, aunque nuestra propuesta se basa en una arquitectura de tipo data lakehouse, existen otras arquitecturas modernas con un importante grado de adopción en otros sectores como las basadas en mallas de datos (Machado et al., 2022) que se suelen referenciar bajo su término en inglés *data mesh*, término que usaremos en lo sucesivo en esta memoria. El data mesh aboga por un enfoque más descentralizado, distribuido y basado en dominios para el almacenamiento y gestión de datos. Nuestro enfoque hacia el data lakehouse implica una apuesta por una **arquitectura centralizada** y aunque la descentralización que aporta el data mesh puede tener ciertas ventajas, creemos que no es adecuada para el sector del turismo en base a los problemas relacionados con el uso de datos detectados en nuestro análisis bibliométrico y marco de referencia inicial. Consideramos que la descentralización del data mesh puede plantear desafíos en términos de coordinación y consistencia en el manejo de datos, lo que podría resultar en silos de información, lo cual ya ha sido un problema habitual en el sector del turismo. Por otro lado, la implementación de un data mesh puede ser compleja y requerir un cambio cultural significativo en una organización y creemos que el sector del turismo no presenta en la actualidad un nivel de madurez de uso de datos adecuado para abordarlo. Consideramos, por tanto, que las arquitecturas a emplear en nuestro marco conceptual deben ser centralizadas y en su evolución deberían ir en línea con los marcos de trabajo que propugnan una gestión unificada de los datos, como por ejemplo las actuales arquitecturas conocidas por su nombre en inglés: *data fabric* (Zaidi et al., 2019).

Técnicas de ciencia de datos e inteligencia artificial

En el apartado 4.1 se detectan un conjunto de técnicas de ciencia de datos e inteligencia artificial que podríamos organizar en tres bloques:

- **Inteligencia de negocio**, basada en informes, cuadros de mando, etc.
- **Técnicas tradicionales ciencia de datos e inteligencia artificial**, tales como segmentación, analítica geoespacial, aprendizaje automático, sistemas de recomendación, procesado de lenguaje natural.
- **Evolución actual inteligencia artificial**, hacia el aprendizaje profundo, inteligencia artificial generativa e inteligencia multimodal.

Todas estas técnicas pueden ser combinadas, si es necesario con los **modelos MCDM** para la toma de decisiones en entornos complejos o con múltiples factores implicados como suele suceder en un buen número de casos de uso en el sector turístico.

Generación de análisis y descubrimientos relevantes para la toma de decisiones

En este apartado, diferenciamos los siguientes grandes bloques en lo relativo a los casos de uso:

- **Análisis centrado en el turista**, comportamiento del turista, satisfacción del turista, percepción del turista, cocreación de experiencias, etc.
- **Análisis centrado en las organizaciones**
 - **Destinos turísticos**, donde se realiza análisis de reputación, planificación, previsión de la demanda, gestión de los recursos, sostenibilidad, propuesta de valor, etc.
 - **Empresas que ofrecen productos o servicios turísticos**, donde se realiza análisis de rendimiento, gestión del marketing, análisis de precios, análisis RSC (Responsabilidad Social Corporativa), etc.

Empleando el marco conceptual planteado se pueden abordar los diferentes casos de uso que podamos identificar en el sector del turismo, tal y como desarrollaremos en los siguientes capítulos 5, 6 y 7. Creemos que los casos de uso seleccionados ofrecen una demostración convincente de la efectividad del marco conceptual propuesto, ya que cada uno de ellos tiene objetivos de negocio diferentes en el ámbito del turismo y cada uno de ellos utiliza fuentes de datos o técnicas analíticas diferentes para la toma de decisiones.

5. Caso de uso: detección y análisis de experiencias turísticas a partir de reseñas en línea para la mejora de la gestión de destinos turísticos. Aplicación en la ciudad de Roma.

A la hora de gestionar destinos turísticos, es fundamental conocer cómo ha sido la experiencia de los turistas para poder mejorarla. Este caso de uso presenta un modelo para detectar y analizar las experiencias de los turistas en un determinado destino turístico a partir de las reseñas realizadas en plataformas digitales. El caso de uso se centra en lo que hemos denominado como 'common-destination-experiences (CDE)', que son experiencias turísticas comunes representativas y relevantes. El modelo genera diferentes tipos de recomendaciones, como recomendaciones generales y recomendaciones personalizadas. Del mapa de las CDEs obtenido por el modelo se pueden derivar recomendaciones generales para mejorar la gestión. El modelo también permite realizar un análisis exploratorio y personalizado de una experiencia concreta para obtener recomendaciones personalizadas. El modelo propuesto podría ser de gran ayuda para que las organizaciones de gestión de destinos conozcan en profundidad las experiencias de los turistas en un destino turístico y tomen decisiones de mejora basadas en este conocimiento.

En la Figura 25 se muestra cómo se integra el caso de uso dentro del marco conceptual propuesto. Como se puede observar en la figura, se resaltan con flechas todas las áreas del marco conceptual que se emplean en el caso de uso, por otro lado, la dirección de las flechas muestra el ciclo de vida del dato desde su ingesta hasta la obtención de análisis con aplicaciones de negocio. Tal y como se explicará en el desarrollo del mismo, en lo relativo a fuentes de datos se trabajará con datos de tipo UGC, concretamente reseñas en línea. Estos datos se ingestarán en el área de almacenamiento y gestión de los datos y se realizará sobre los mismos una limpieza y depuración. Posteriormente se aplicarán diferentes técnicas de análisis de datos, tales como la segmentación y se visualizarán los resultados en entornos que faciliten la exploración y el análisis. Las aplicaciones se centrarán en el ámbito del análisis de la percepción del turista, lo cual facilitará la toma de decisiones en el ámbito de la gestión de destinos turísticos.

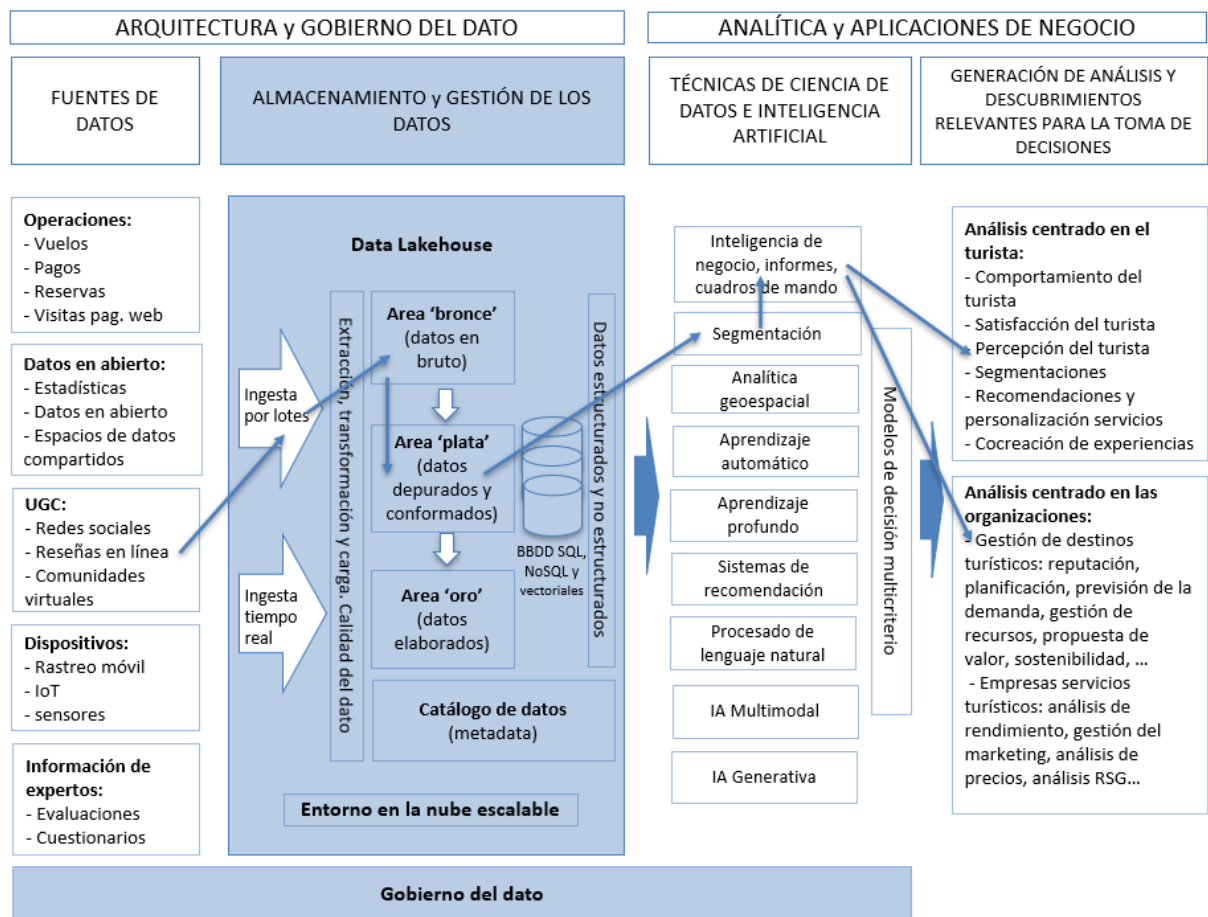


Figura 25: Integración del caso de uso 'Detección y análisis de experiencias turísticas a partir de reseñas en línea para la mejora de la gestión de destinos turísticos. Aplicación en la ciudad de Roma' dentro del marco conceptual propuesto. Fuente: elaboración propia

5.1. Introducción

La competitividad de los destinos turísticos es una cuestión cada vez más relevante. Dwyer & Kim (2003) desarrollaron un Modelo de Competitividad de Destinos que incluye los principales elementos de competitividad, y Crouch (2011) también analizó los principales atributos que afectan a la competitividad de un destino turístico. También son relevantes los estudios que analizan las implicaciones que la competitividad tiene en la comunidad receptora (Crouch & Ritchie, 2012). El tipo de destino turístico y sus características particulares influyen significativamente en su competitividad (Parra-López & Oreja-Rodríguez, 2014). La gestión de los destinos turísticos para mejorar su competitividad implica una renovación continua de sus recursos, infraestructuras, oferta, servicios y experiencia turística. Esta renovación se ha visto acelerada por la pandemia del Covid-19 y ha generado un contexto de incertidumbre que ha requerido una continua

adaptación al cambio. En este contexto de competitividad e incertidumbre, el concepto de destinos inteligentes ha cobrado fuerza en los últimos años (Del Chiappa y Baggio, 2015). Se trata de un concepto en desarrollo, que surge del de ciudad inteligente y está estrechamente vinculado a la transformación digital que están experimentando muchos sectores, entre ellos el turismo. Gretzel et al. (2015) definieron y describieron la idea de un ecosistema turístico inteligente en el que las tecnologías inteligentes, las ciudades inteligentes y el turismo inteligente generan nuevas formas de crear, intercambiar y consumir valor. Buhalis y Amaranggana (2014) consideran que la inteligencia, cuando se refiere a un destino turístico, requiere la capacidad de intercambiar información en tiempo real relacionada con las actividades turísticas, con el objetivo de mejorar la satisfacción de los turistas y la eficiencia en la gestión de los recursos. Estas actividades generan un vasto conjunto de información digital, que se enmarca dentro del concepto de big data, y permite a las organizaciones nacionales de turismo (NTO por sus siglas en inglés *National Tourism Organizations*) y a las organizaciones de gestión de destinos (DMO por sus siglas en inglés *Destination Management Organizations*) extraer valiosos conocimientos. El uso de big data por parte de los destinos turísticos (Wang et al., 2013) puede apoyar la toma de decisiones empresariales y la asignación óptima de recursos, y puede ayudar a comprender mejor la percepción y el comportamiento de los turistas. Buhalis (2020) analizó la naturaleza transformadora y disruptiva de la tecnología utilizada para el turismo.

En este ámbito, como es lógico, las fuentes de datos juegan un papel primordial. En el marco conceptual propuesto en esta tesis se identifican los diferentes tipos de fuentes de datos empleadas en turismo, en su mayoría de aplicación para la gestión de destinos turísticos. Encontramos investigaciones que han empleado datos estructurados (Bratucu & Cismaru, 2015) y geolocalizados (Krsak et al. 2018) para la mejora de la gestión de destinos turísticos, pero son los datos no estructurados los que más interés han despertado en los últimos años. En concreto, los datos de tipo UGC. Este tipo de datos ha sido utilizado en numerosas investigaciones (Barisic et al., 2019) relativas al ámbito de la gestión de destinos turísticos con el objetivo de comprender mejor el comportamiento y la percepción del turista o la creación de nuevos servicios (Buhalis & Sinnarta, 2019) en turismo y hostelería. Dentro de los datos UGC, muchos de los estudios se han centrado en las reseñas realizadas por los turistas en plataformas digitales (Van der Zee & Bertocchi, 2018; Chiu & Leng, 2017) como Tripadvisor, Booking u otras. Estrechamente relacionado con el término UGC encontramos el término eWOM, que hace referencia al intercambio no comercial de información relativa a un producto, servicio u organización, que tiene lugar en un entorno digital.

En este caso de uso, se propone un modelo general en el que utilizaremos las reseñas turísticas para extraer experiencias turísticas, en concreto lo que denominamos CDE, que son experiencias comunes, representativas y relevantes de un destino turístico que servirán para mejorar su gestión. De la recopilación y análisis de estas experiencias comunes de destino obtendremos un mapa general del que extraeremos recomendaciones generales. También se pueden obtener recomendaciones personalizadas a partir de un análisis personalizado de una CDE concreta.

Este caso de uso está organizado como sigue: en el apartado 5.2 revisamos trabajos similares al nuestro y discutimos la novedad e innovación del mismo. El apartado 5.3 trata de la metodología, en el que presentamos los aspectos metodológicos necesarios para nuestro modelo, así como introducimos el concepto de CDE. El apartado 5.4 presenta nuestro modelo de ocho fases que nos permitirá obtener recomendaciones generales y personalizadas para un destino turístico. El modelo propuesto se aplicará a un caso real de gran interés como es la ciudad de Roma. El apartado 5.5 presenta las conclusiones con implicaciones teóricas y de gestión y el apartado 5.6 propone trabajos futuros.

5.2. Trabajos relacionados

Como hemos visto en el apartado anterior, en la literatura podemos encontrar diferentes investigaciones que analizan datos de tipo UGC para obtener conocimiento útil en la gestión de destinos turísticos. La

Tabla 8 muestra un resumen de los artículos que actualmente trabajan en esta dirección. Se incluyen en el resumen una visión general de cada artículo, mostrando las fuentes de datos, las técnicas de análisis de datos utilizadas, el caso de estudio realizado y los conocimientos obtenidos. Nos interesa examinar si estas percepciones son generales o si pueden personalizarse para obtener percepciones más específicas y adaptarlas después a una necesidad de análisis concreta. Como puede observarse, la mayoría de estas investigaciones trabajan con el objetivo de conocer en profundidad la percepción que los turistas tienen del destino turístico. Algunas de ellas se centran en analizar su percepción sobre aspectos concretos como restaurantes u hoteles desde diferentes puntos de vista (Van der Zee & Bertocchi, 2018; Van der Zee et al., 2020; Christodoulou et al., 2020), otras obtienen una visión más general del destino turístico empleando diferentes técnicas (Solazzo et al., 2020; Alrawadieh et al., 2018; Antonio et al., 2020) para comprender los comportamientos de los turistas y la percepción del destino. Asimismo, con una perspectiva general del destino turístico, encontramos autores que se centran en obtener

una percepción global positiva o negativa a partir del análisis de sentimiento (Kim et al. ,2017). También encontramos estudios que obtienen predicciones de la demanda turística en los destinos o anticipan las intenciones de visita de los turistas aplicando diferentes técnicas (Liu et al. ,2021; Gunter et al., 2019; Benslimane & Semaoune ,2021; Sohrabi et al., 2020). De igual forma, encontramos estudios que se centran en los principales atractivos turísticos de estos destinos (Marine-Roig & Clave, 2015; Yu et al. ,2021). Una línea de investigación interesante es aumentar nuestro conocimiento de la experiencia turística como factor crítico de cara a la mejora de la gestión de los destinos turísticos, objetivo de nuestro caso de uso. Desde esta perspectiva de la experiencia, encontramos estudios centrados en conocer la experiencia de los turistas en los principales atractivos turísticos de un destino concreto (Yu et al. ,2021), o durante una actividad turística concreta (Chiu & Leng ,2017) u otro tipo de experiencia específica (Gon ,2021). En nuestra propuesta, tratamos de ampliar estas perspectivas para obtener una imagen completa de la experiencia de un turista en un destino turístico concreto. Para ello, obtenemos lo que denominamos CDE a partir de las reseñas de los turistas, que son experiencias comunes, representativas y relevantes de un destino turístico, y con ellas generamos un mapa global de CDEs. Este mapa nos permite añadir la posibilidad de generar análisis personalizados, realizando exploraciones mediante palabras clave específicas de una determinada CDE que queramos profundizar, de forma que podamos obtener recomendaciones generales o personalizadas que permitan a las organizaciones gestoras del destino turístico tomar medidas concretas de mejora. Por otro lado, como factor diferencial, nuestro caso de uso propone un modelo general de trabajo aplicable a las reseñas obtenidas de cualquier destino turístico.

Referencia del artículo	Fuentes de datos	Técnicas de análisis de datos	Aplicación	Análisis y descubrimientos (generales y personalizados)
Van der Zee & Bertocchi (2018)	UGC de Tripadvisor y ubicación espacial de hoteles y restaurantes	Análisis de redes sociales y geolocalización	Hoteles y restaurantes de Amberes (Bélgica)	El estudio muestra cómo la proximidad geográfica y la proximidad relacional están interrelacionadas en el caso de los evaluadores internacionales, pero no tanto en el de los nacionales. Conclusiones y descubrimientos generales y una serie de recomendaciones para los responsables políticos y los gestores de destinos que intentan lograr una mejor distribución de los turistas en el destino.
Van der Zee et al. (2020)	UGC de Tripadvisor sobre reseñas de restaurantes	Análisis de redes sociales, análisis de geolocalización y clustering	Restaurantes de cinco ciudades artísticas flamencas (Bélgica)	El estudio detecta agrupaciones espaciales de restaurantes con reseñas frecuentes (<i>hot spots</i>) y poco frecuentes (<i>cold spots</i>) en cuatro de las cinco ciudades artísticas. El análisis de agrupaciones espaciales y puntos calientes es capaz de mostrar la huella digital de los turistas y proporciona información general para la gestión de destinos turísticos.
Christodoulou et al. (2020)	UGC de Tripadvisor sobre hoteles, datos sobre el PIB per cápita del Fondo Monetario Mundial, datos sobre los valores culturales de los evaluadores del sitio web de Hofstede.	Modelización de temas, análisis de sentimientos, árboles de decisión	Hoteles de Chipre	Los patrones obtenidos se expresan en términos de reglas que combinan variables exógenas y endógenas de las opiniones de los turistas que conducen a un sentimiento negativo o positivo de éstos. El modelo predictivo permite obtener información general que puede utilizarse para adaptar la estrategia de marketing a los turistas con más probabilidades de quedar satisfechos en el destino.
Solazzo et al. (2020)	Big Social Data: conjunto de datos de la red social en línea Flickr y del microblog Twitter	Marco de trabajo analítico, que implementa: extracción de temas, agrupación de datos geográficos, técnicas de modelización de series temporales, análisis de sentimientos y análisis de imágenes.	Región de Salento (Italia)	Detección de clústeres de puntos de interés, identificación de tendencias y patrones estacionales de demanda turística, seguimiento de eWOM, identificación de lugares atractivos. Información general sobre el comportamiento de los turistas y la percepción de los destinos. Las DMO pueden explotar los conocimientos para satisfacer sus necesidades en términos de apoyo a la toma de decisiones.
Antonio et al. (2020)	UGC de Tripadvisor sobre Salamanca (España) y Coimbra (Portugal)	Minería de textos, análisis de sentimiento y análisis de la cesta de la compra	Salamanca (España) y Coimbra (Portugal)	Patrones según atracción, nacionalidad y visitas repetidas. Comportamientos y tipo de cosas que los visitantes valoran y desean encontrar en su experiencia en este tipo de ciudades. Perspectivas generales y comprensión más profunda de los turistas que visitan los sitios del Patrimonio Mundial útiles para los DMO.

Alrawadieh et al. (2018)	Datos de entradas de blogs de viajes extraídos de TravelPod.com	Análisis de contenidos, inteligencia de negocio	Estambul (Turquía)	Apreciaciones generales sobre la imagen del destino Estambul que, en general, son positivas. En concreto, el destino es apreciado principalmente por su patrimonio cultural e histórico, así como por su infraestructura turística. Se espera que los resultados ofrezcan información relevante a las DMOs turcas.
Kim et al. (2017)	Datos de reseñas recogidos en un sitio de reseñas de viajeros (www.virtualtourist.com)	Minería de textos, análisis de sentimientos	Reseñas de París (Francia) de 14 categorías	Se midió con éxito la percepción de los turistas sobre el servicio en cada categoría. La información sobre la percepción del servicio por parte de los turistas en cada categoría proporciona información general para la gestión de los destinos turísticos.
Gunter et al. (2019)	Datos sobre los 'me gusta' de las publicaciones en las páginas de Facebook de las organizaciones de gestión de destinos de las ciudades. Datos de Google Trends	Retraso autorregresivo distribuido, técnicas de series temporales, minería de textos	Cuatro grandes organizaciones de gestión de destinos de Austria (Graz, Innsbruck, Salzburgo y Viena).	Predicción de llegadas totales de turistas a cuatro ciudades austriacas (Graz, Innsbruck, Salzburgo y Viena) a partir de los 'me gusta' de las publicaciones en las páginas de Facebook de las organizaciones de gestión de destinos de estas ciudades. Predicción de llegadas de turistas a ciudades austriacas que resultan generales y útiles para las DMOs.
Benslimane & Semaoune (2021)	Imágenes de destino, datos eWOM sobre Argelia	Teoría de las Acciones Razonadas	Argelia	Buena capacidad predictiva de la intención de los turistas de visitar un destino emergente. Además, la actitud, las normas subjetivas y la imagen del destino fueron predictores significativos de la intención de visita y del eWOM. Los resultados del estudio son ideas generales que pueden ayudar a desarrollar una estrategia de marketing turístico más eficaz y exitosa.
Sohrabi et al. (2020)	Datos de las redes sociales	Análisis de textos, minería de datos	Irán	Esta investigación estudia el efecto de los intereses en los turistas emisores iraníes, especialmente su selección de destino, y a continuación, utilizando algoritmos de minería de textos y datos, introduce un modelo para predecir los destinos de los turistas en función de sus intereses y antecedentes de viaje.
Liu et al. (2021)	Reseñas en línea de Ctrip (https://www.ctrip.com), UGC geotiquetado	Reglas de asociación, minería de textos	Provincia de Yunnan (China)	La comprensión general de las reglas de asociación de destinos puede ayudar a entender la posibilidad de que los turistas visiten un destino después de haber viajado desde otro. Estos conocimientos son útiles para que las industrias turísticas exploren estrategias y productos de viaje y ofrezcan mejores servicios.
Marine-Roig & Clave (2015)	Más de 100.000 blogs de viajes y reseñas de viajes en línea relevantes (OTR)	Minería de textos, inteligencia de negocio	La Sagrada Familia, principal atracción turística de Barcelona (España)	Obtención de inteligencia de negocio a partir de opiniones de viajeros en línea sobre las visitas al principal monumento/atracción de Barcelona, La Sagrada Familia. Información general útil sobre la imagen de los destinos transmitida por los turistas acerca de la Sagrada Familia que puede ser utilizada por los DMO.

Yu et al. (2021)	Datos de las opiniones de Tripadvisor sobre las atracciones turísticas más populares de Londres.	Minería de textos, análisis de contenidos, análisis de co-ocurrencias, mapeo semántico	Las diez atracciones turísticas más populares de Londres	Identificar una colección de temas y conceptos que reflejen las experiencias memorables de los turistas durante sus visitas a las atracciones. Ofrece ideas y recomendaciones generales para que las DMO y los distintos agentes del turismo urbano planifiquen, comercialicen y gestionen productos y servicios turísticos urbanos.
Gon (2021)	Datos de redes sociales de Instagram con hashtags relacionados con 'experiencia local'.	Minería de textos, Análisis de contenidos	Publicación en Instagram de más de 100 hashtags que incluían la palabra clave 'local'.	Comprensión de qué y quién está asociado a las experiencias locales (por ejemplo, productos locales y actividades locales), lo que sugiere que los medios sociales pueden aportar nuevos conocimientos. Los conocimientos generales sobre la experiencia local pueden ayudar a las DMO en el proceso de diseño de experiencias turísticas.
Chiu & Leng (2017)	Datos de las opiniones de Tripadvisor sobre la experiencia ciclista de los turistas en Singapur	Minería de textos, análisis de contenidos, análisis de co-ocurrencias, mapeo semántico	Experiencia cicloturista en Singapur	Los resultados sugieren que las excursiones en bicicleta pueden proporcionar a los turistas experiencias placenteras. Conclusiones generales útiles para que las DMO diseñen experiencias ciclistas para los turistas
Nuestra propuesta	Datos de Tripadvisor sobre Roma (Italia)	Extracción datos web, análisis de co-ocurrencia, VOS clustering	Roma (Italia)	Análisis y descubrimientos generales proporcionados por un mapa global de experiencias turísticas en un destino turístico que se obtienen desde múltiples perspectivas. Igualmente se pueden obtener análisis personalizados explorando algunas palabras relacionadas con una experiencia específica, lo que permite una exploración personalizada y a medida de esa experiencia. Este mapa global de experiencias y los análisis personalizados permiten a las organizaciones gestoras de destinos tomar decisiones y emprender acciones para mejorar la experiencia turística desde múltiples perspectivas.

Tabla 8: Artículos que analizan datos de tipo UGC para obtener información útil sobre la gestión de los destinos turísticos.

5.3. Metodología

Como hemos indicado anteriormente, nuestro objetivo es identificar experiencias turísticas a partir de reseñas. Supongamos una reseña sobre una visita turística a una iglesia románica, en ella aparecerán palabras clave tales como arquitectura, arte, historia, religión, iglesia románica, etc., que se relacionan entre sí y que pueden representar una experiencia. Teniendo como objetivo la identificación de experiencias, vamos a generar redes de palabras clave a partir de estas reseñas. Trabajando con estas redes de palabras clave vamos a identificar las experiencias y para ello, necesitamos modelar las relaciones que existen entre las palabras clave. Estas relaciones se modelarán mediante grafos. Los grafos proporcionan una representación flexible y sencilla de los objetos y sus relaciones. Una vez que hayamos modelizado nuestro problema con grafos, podremos analizar las relaciones utilizando las matemáticas de la teoría de grafos. Para identificar las experiencias más relevantes, definiremos lo que llamamos CDE, que también se modelarán con grafos y se obtendrán a partir de redes de palabras clave.

5.3.1. Modelizar experiencias con grafos

En el apartado 4.3.1 de esta memoria se introdujo el concepto de grafo, que ahora utilizaremos para modelizar experiencias. Como paso previo, debemos definir el concepto de palabra clave que vamos a utilizar en nuestro caso de uso. Las palabras clave de las reseñas son aquellas palabras de las reseñas que han sido sometidas a un proceso de limpieza y refinamiento y que tienen un número mínimo de apariciones en las reseñas, como se explicará en el apartado 5.4. Dicho de otro modo, son las palabras más relevantes de las reseñas y serán cruciales en nuestro caso de uso.

Dado un conjunto de reseñas $R = \{R_1, \dots, R_n\}$ y un conjunto de palabras claves (*keywords* en inglés) $W = \{W_1, \dots, W_k\}$ una experiencia puede modelarse como un grafo G donde los nodos o vértices V son el conjunto de palabras clave W y el conjunto de aristas (*edges* en inglés) E representa las relaciones o enlaces entre las palabras claves, de tal forma que E_{ij} representa la relación entre W_i y W_j .

A continuación, debemos definir la relación entre las palabras clave, que será la relación entre los nodos de los grafos. Una técnica muy utilizada en las redes de palabras es la de co-ocurrencia de términos (Callon et al., 1983).

La co-ocurrencia C_{ij} es definida aquí como el número de reseñas de un conjunto R en las que dos palabras claves particulares W_i y W_j del conjunto W aparecen juntas.

Una vez definida la relación, podemos fijar un segundo filtro estableciendo un umbral con un valor mínimo de co-ocurrencias para cada palabra clave. Los dos umbrales establecidos sobre la ocurrencia y co-ocurrencia de palabras clave nos permiten obtener grafos cuyos nodos y aristas son representativos de la información contenida en las reseñas. Las experiencias que modelizamos mediante estos grafos son, por tanto, representativas de las reseñas del destino turístico.

En este contexto, podemos considerar una experiencia como un subgrafo dentro del grafo general que contiene todas las palabras clave de las reseñas. Tenemos, por tanto, un conjunto de experiencias que serán un conjunto de subgrafos dentro de un grafo general. Este conjunto de experiencias contendrá experiencias específicas de reseñas específicas, pero nuestro objetivo no es ir a experiencias específicas, sino obtener experiencias más generales que reúnan información de varias reseñas. Una solución para obtener experiencias más generales es agrupar conjuntos de experiencias específicas similares. Esta agrupación reunirá información de varias experiencias específicas similares y, por tanto, de varias reseñas similares. La agrupación de experiencias específicas se denomina *common-destination-experiences* (CDE), un concepto que ya se ha introducido y que se tratará con más detalle a continuación.

Para obtener estos clústeres de experiencias concretas debemos emplear una técnica de clustering adecuada a nuestro caso. Las técnicas de clustering (Krishna et al., 2018) nos permitirán detectar nodos fuertemente conectados entre sí y con un menor grado de conexión con el resto de la red. El enlace entre nodos se ha establecido, como se ha mencionado anteriormente, mediante la técnica de co-ocurrencia. Estos enlaces se utilizarán para realizar el clustering, pero antes de utilizarlos es conveniente realizar un paso de normalización. En nuestra red habrá palabras clave mucho más frecuentes que otras y esto afectará al valor del enlace, por lo que palabras clave muy frecuentes aumentarán el valor de los enlaces. Nos interesa obtener un valor de enlace que no dependa de la frecuencia de cada palabra clave y que represente la fuerza de la relación entre nodos independientemente de la frecuencia de cada palabra clave.

Por lo tanto, es necesario un proceso de normalización (Van Eck & Waltman, 2009). En la literatura se han utilizado diferentes medidas para normalizar una red, entre las cuales las más apropiadas para nuestro caso de uso son las utilizadas en redes bibliométricas como: coseno de Salton (Salton & McGill, 1983), índice de Jaccard (Peters &

Van Raan, 1993), índice de inclusión (Tijssen & Van Raan, 1989) y fuerza de asociación (Van Eck et al. ,2006).

Estudios relacionados (Van Eck & Waltman, 2009) indican que la fuerza de asociación es una medida probabilística, mientras que el coseno de Salton, el índice de inclusión y el índice de Jaccard son medidas teóricas de conjuntos y los datos de co-ocurrencia pueden normalizarse mejor utilizando una medida probabilística. Consideramos, por tanto, que el método de normalización de la fuerza de asociación es muy apropiado cuando se trabaja con redes de co-ocurrencia.

Este método también se conoce como índice de proximidad (Peters & Van Raan, 1993; Rip & Courtial, 1984) o como índice probabilístico de afinidad (Zitt et al., 2000). Utilizando la fuerza de asociación, la similitud S_{ij} entre dos elementos i y j es calculado como:

$$S_{ij} = C_{ij}/W_iW_j$$

donde C_{ij} denota el número de co-ocurrencias de los elementos i y j , i.e., el número de reseñas en las que los artículos i y j aparecen juntos y donde W_i y W_j denotan el número total de apariciones de elementos i y j .

Una vez definida la técnica de normalización, debemos definir la técnica de agrupación que se utilizará para obtener los CDEs, que es el conjunto $CDE = \{CDE_1, \dots, CDE_m\}$. En el apartado 43.1.1 se introdujeron las diferentes técnicas de detección de comunidades en redes. En particular, se mencionaron las técnicas modulares, y se destacó la técnica VOS que será la técnica que emplearemos para realizar las agrupaciones y obtener las CDEs. Consideramos que la técnica VOS es muy apropiada para nuestro caso de uso, tal y como se detallará en la el apartado 5.4 relativo al modelo planteado.

La Figura 26 muestra un resumen de los pasos seguidos para obtener las CDEs:

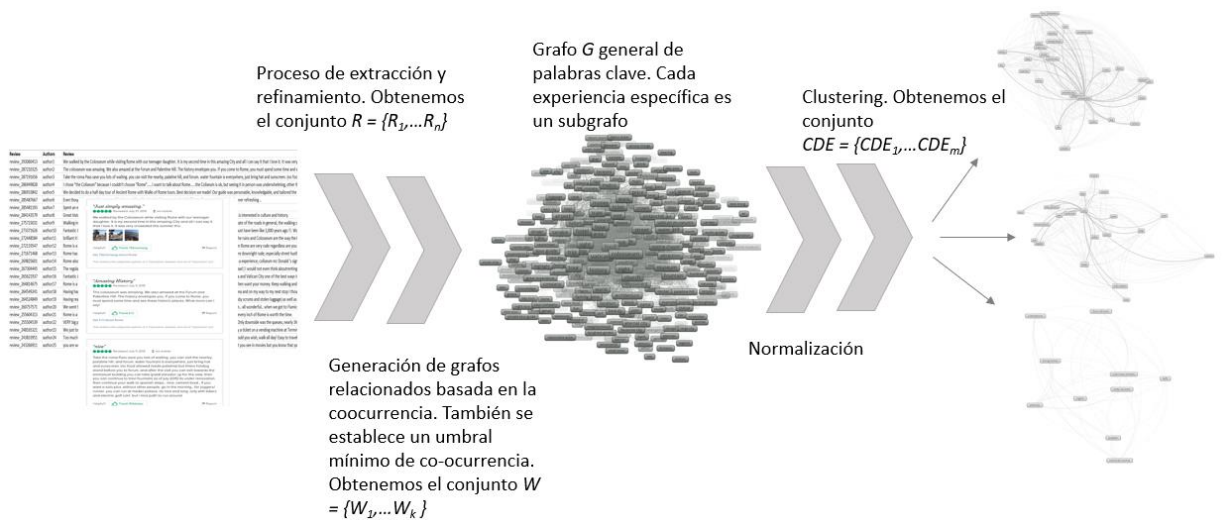


Figura 26: Resumen de los pasos dados para obtener las CDEs de las reseñas. Fuente: elaboración propia

Una vez definidas las técnicas de obtención de las CDE, examinaremos sus principales características.

Como se verá en detalle en el apartado 5.4, hemos obtenido la CDE agrupando experiencias concretas. En la Figura 27 podemos ver un ejemplo de CDE relacionado con la seguridad del turista en un destino turístico. Esta CDE es un clúster que se compone de experiencias específicas con carteristas o ladrones o experiencias particulares en zonas específicas.

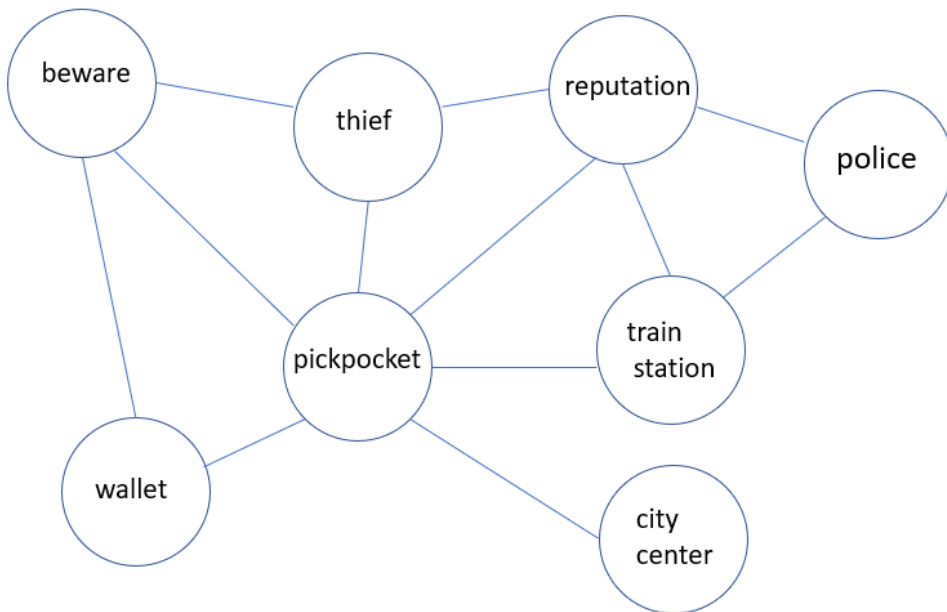


Figura 27: CDE relacionada con la seguridad del destino. Fuente: elaboración propia.

Como ya se ha mencionado, las palabras clave de las redes son representativas de las reseñas de las que proceden. Por tanto, podemos decir que una CDE formado por estas palabras clave es una experiencia representativa de las reseñas. Otra propiedad importante de las CDE es que agrupan experiencias comunes. Además, al consolidar o resumir un conjunto de experiencias específicas similares, podemos considerarlo un patrón de experiencias. Es interesante considerar el impacto que una experiencia de un conjunto de turistas puede tener sobre un destino turístico, desde esta perspectiva podemos decir que una CDE es una experiencia de impacto, ya que recoge información de muchas reseñas y por tanto de muchos turistas. Este impacto puede ser positivo o negativo para el destino turístico. En la Figura 27 podemos deducir que los problemas de seguridad tienen un impacto negativo en la reputación de este destino turístico. En la Figura 28 mostramos un resumen de las principales características de una CDE.



Figura 28: Principales características de una CDE. Fuente: elaboración propia

Como se ha mencionado anteriormente, una de las características de una CDE es que se trata de un patrón de experiencia formado por un conjunto de experiencias específicas. Puede ser interesante, por tanto, profundizar en una CDE realizando análisis personalizados como se detallará en el apartado 5.4. La Figura 29 muestra un análisis

personalizado en el que buscamos palabras clave relacionadas con la palabra clave carterista y que nos puede permitir identificar zonas concretas del destino turístico donde se pueden encontrar carteristas.

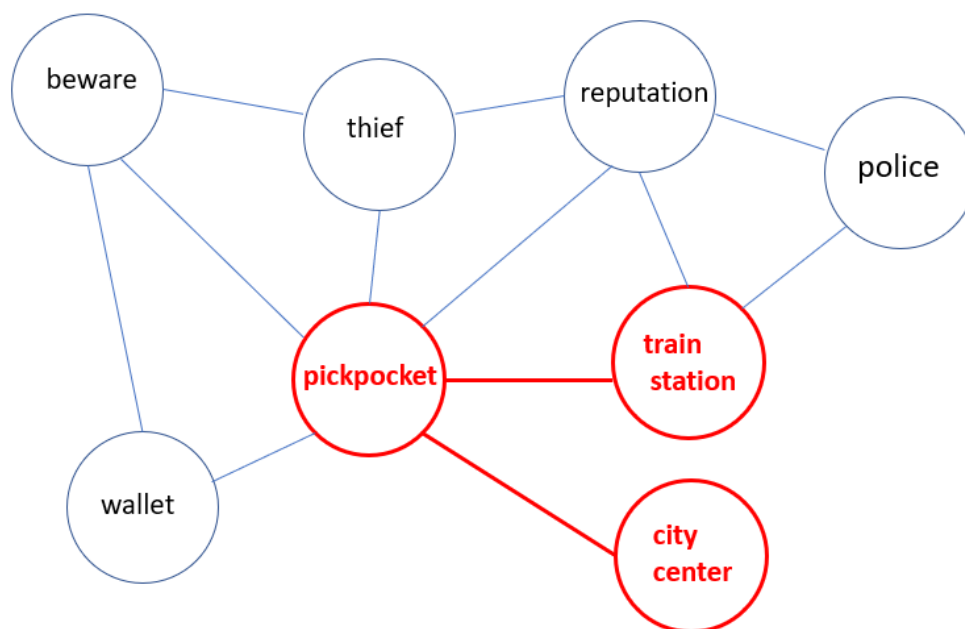


Figura 29: Exploración personalizada dentro de un ejemplo de CDE relacionada con la seguridad del destino. Fuente: elaboración propia

Cabe señalar que investigaciones anteriores (Yu et al., 2021; Chiu & Leng, 2017; Gon, 2021) ya han trabajado el concepto de experiencia de forma similar utilizando agrupaciones dentro del grafo (véanse los 'mapas conceptuales' obtenidos con la herramienta Leximancer) para obtener información relacionada con la experiencia de los turistas. Nuestro caso de uso difiere de estas investigaciones en que modelizamos el concepto de experiencia como un grafo. Otra diferencia es que definimos el concepto de CDE, este concepto amplía la definición de experiencia y lo consideramos más apropiado para su uso en la gestión de destinos turísticos. Por otro lado, nuestro caso de uso propone un modelo de obtención de CDEs, cuestión que aún no ha sido abordada en investigaciones anteriores.

5.4. Modelo propuesto

A continuación, se presenta una propuesta de modelo para identificar y analizar las CDEs a partir de los comentarios de los turistas, que generará recomendaciones para la

gestión de los destinos turísticos. El modelo puede dividirse en ocho fases principales, como se muestra en la Figura 30.

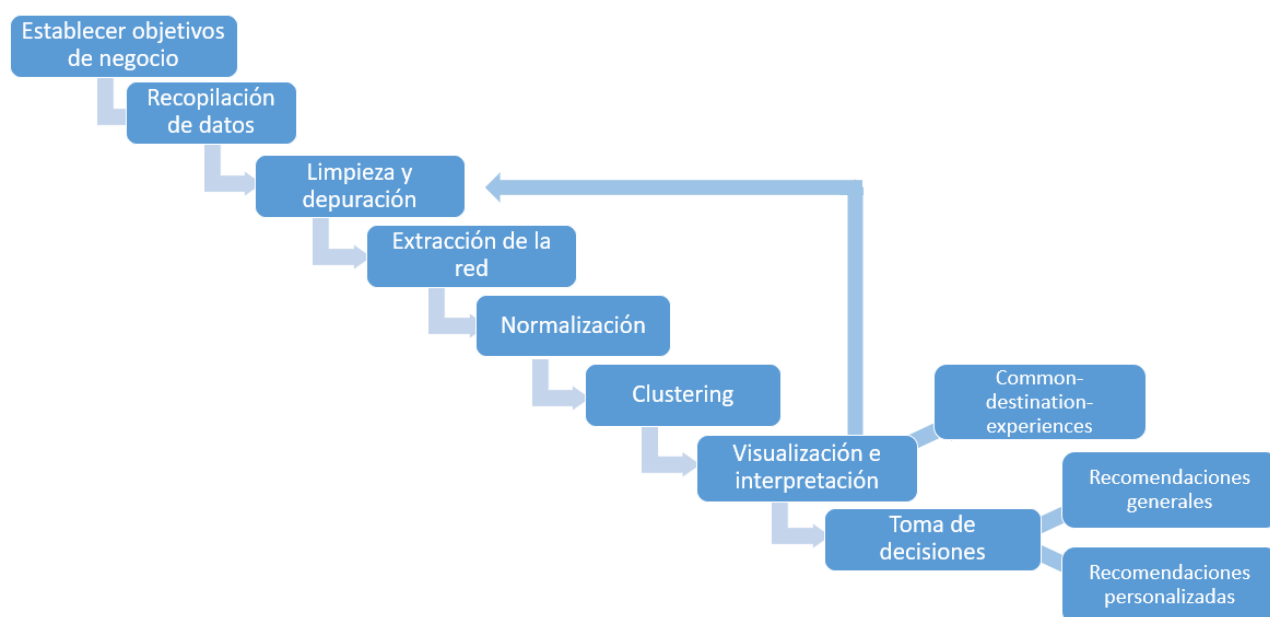


Figura 30: Fases del modelo propuesto para identificar y analizar CDEs. Fuente: elaboración propia

A continuación, detallaremos los pasos llevados a cabo en cada una de las fases del modelo y explicaremos cómo se han aplicado en la ciudad de Roma. El caso de Roma pretende mostrar un ejemplo de uso del modelo, que por supuesto puede aplicarse a otros muchos casos de interés en el sector turístico. Para la aplicación de los pasos, nos hemos basado en la herramienta Vosviewer (Van Eck & Waltman, 2010; Vosviewer, 2022). Se trata de una herramienta de software utilizada para construir y visualizar redes bibliométricas. En nuestra aplicación hemos utilizado la información de las reseñas turísticas en lugar de la de las publicaciones, y hemos emparejado el título de la publicación con el título de la reseña y el resumen de la publicación con la propia reseña.

5.4.1. Establecer los objetivos de negocio

En primer lugar, debemos establecer objetivos específicos desde una perspectiva empresarial. En el modelo propuesto, nuestro objetivo general es detectar y analizar experiencias turísticas a partir de las reseñas realizadas sobre un destino turístico concreto en una plataforma digital, red social u otro tipo de comunidades virtuales. El objetivo será obtener CDEs que ayuden a tomar decisiones y mejorar la gestión de un destino turístico concreto. Tal y como se ha mencionado anteriormente, en el caso de aplicación del modelo, el destino turístico seleccionado para el análisis es Roma, por ser una de las ciudades que

mayor volumen de turistas recibe cada año y por ser uno de los destinos urbanos más importantes de Europa. Otro aspecto interesante es que recibe un alto volumen de opiniones en redes sociales. Nos interesa obtener una visión general de la experiencia de los turistas en Roma. Es decir, no nos centramos en las experiencias de atracciones turísticas u hoteles concretos de Roma, sino en una visión global de las experiencias turísticas en Roma.

5.4.2. Recopilación de los datos

En segundo lugar, debemos recopilar la información que vamos a analizar. En nuestro caso de aplicación a la ciudad de Roma, las reseñas turísticas realizadas en el portal de Tripadvisor (Tripadvisor, 2023) sirven como fuente de información. Seleccionamos, por tanto, las reseñas referidas a Roma. El objetivo es que estas reseñas ofrezcan una visión general de la experiencia de los turistas en un destino turístico, por lo que buscamos reseñas con esta visión global. También es posible obtener reseñas sobre una atracción turística específica de un destino turístico o sobre los hoteles de este destino, sin embargo, las reseñas que necesitamos son aquellas que reflejen la experiencia global de los turistas en el destino turístico.

La ciudad de Roma recibe un alto volumen de opiniones turísticas en el portal de Tripadvisor. En el siguiente enlace, encontramos opiniones generales sobre la experiencia turística en Roma (Enlace Tripadvisor, 2023). Encontramos 2465 reseñas de las cuales las 1004 primeras están escritas en inglés y el resto en otros idiomas. Nos vamos a centrar en estas 1004 reseñas escritas en inglés entre 2003 y 2015. Estos datos serán nuestro punto de partida en nuestro caso de aplicación, aunque, como ya se ha indicado, el modelo propuesto podría funcionar con otros periodos de tiempo o con datos de otro destino turístico, ya que el objetivo del modelo es proponer un método para detectar y analizar experiencias a partir de reseñas turísticas independientemente del destino turístico al que se refieran o del periodo de tiempo de los datos.

Para descargar estas reseñas, utilizamos técnicas de raspado de web, conocidas habitualmente por su término de inglés: *web scraping*. Estas técnicas se basan en descargar páginas HTML (*HyperText Markup Language*) y realizar un análisis sintáctico del código HTML que nos permita recorrer la jerarquía de etiquetas que componen la página para obtener la información buscada. Estas técnicas han sido ampliamente utilizadas para recopilar información de internet (Saurkar et al., 2018) y en particular de redes y plataformas sociales (Pereira & Vanitha 2015).

En el caso de la página web utilizada, es necesario identificar las etiquetas que almacenan la información relevante de la reseña (identificador de la reseña, título y texto) y, por otro lado, es necesario realizar procesos iterativos y de paginación para recorrer todas las reseñas y las diferentes páginas que las contienen. Para ello utilizamos el lenguaje de programación Python y paquetes especializados en técnicas de web scraping como Request (Request, 2022) y BeautifulSoup (BeautifulSoup, 2022). Los resultados se descargan a fecha 2 de abril de 2022 en un fichero cuya información más relevante es el identificador de la reseña, el título y el texto de las 1004 reseñas obtenidas.

5.4.3. Limpieza y depuración

Los datos recuperados de las opiniones turísticas suelen contener errores, por ejemplo, faltas de ortografía en el título o en el propio texto de la reseña. A veces, hay palabras en las reseñas que representan el mismo concepto, pero se escriben de forma diferente; por ejemplo, un lugar de Roma puede escribirse de distintas maneras, por ejemplo: 'Coliseo', 'Colosseum', 'Coliseo'. En otros casos, un concepto puede representarse con distintas palabras (formas léxicas) o siglas, y también representar el mismo concepto. Problemas similares aparecen con las formas singulares o plurales de una palabra. También es importante detectar y eliminar las palabras que no tienen ningún valor desde el punto de vista conceptual o semántico, usualmente denominadas por su término en inglés *stopwords*.

En este paso y partiendo de la información en bruto, debemos realizar un proceso de depuración con la herramienta Vosviewer y para ello nos apoyamos en un archivo tesoro hecho a medida para nuestra aplicación del caso de uso a Roma que nos permitirá descartar o agrupar palabras. Se utiliza el texto del título de la reseña y la propia reseña. En la Figura 31 podemos ver ejemplos de algunas de las palabras más frecuentes que contienen agrupaciones, la columna 'label' contiene el valor original y la columna 'replace by' es el valor en el que se ha agrupado.

label	replace by
colesseum	colosseum
coliseum	colosseum
colisseum	colosseum
colleseum	colosseum
collesium	colosseum
colliseum	colosseum
colloseum	colosseum
collosseum	colosseum
colluseum	colosseum
colluseum	colosseum
coloseum	colosseum
colosium	colosseum
colosseo	colosseum
restaraunt	restaurant
resturant	restaurant
ristorante	restaurant
rome cab	rome cabs
romecab	rome cabs
romecabs	rome cabs

Figura 31: Ejemplos de palabras agrupadas en el archivo tesauro empleado en Vosviewer. Fuente: elaboración propia

El siguiente paso consiste en quedarse con las palabras que tienen un número mínimo de apariciones en el texto. Existen dos técnicas posibles para realizar este cálculo:

- Recuento binario, donde el atributo ocurrencias indica el número de documentos en los que un término aparece al menos una vez.
- Recuento completo, donde el atributo ocurrencias indica el número total de ocurrencias de un término en todos los documentos.

En nuestra aplicación del caso de uso hemos utilizado una frecuencia mínima de 5 apariciones y la técnica de recuento binario. Nos quedan 895 palabras clave.

5.4.4. Extracción de la red

La técnica que vamos a utilizar en nuestro modelo para establecer los vínculos entre palabras clave será la técnica de co-ocurrencia introducida en el apartado 5.3. Obtendremos así la co-ocurrencia para cada combinación de pares de palabras clave.

En la Figura 32 podemos ver ejemplos de co-ocurrencias para las palabras clave 'arte' y 'arquitectura' utilizadas en nuestra aplicación del caso de uso.

Author	Title	Review	Link
jtb1234	Best vacation I have ever had	Would go back to Rome in a heartbeat. Stayed near the train station in a nominally priced hotel and everything was within walking distance - and we walked the whole city! We spent a week there over New Year's and it was awesome. The architecture, art , neighborhoods, Vatican, Villa Borghese were beyond compare with any other European city I have visited. Never had a dull minute there.	https://www.tripadvisor.com/ShowUserReviews-g187791-r46192912-Rome_Lazio.html
timetraveler2000	Inspirational!	Our Vatican tour of Rome, Italy was wonderful! We learned so much history & facts, saw amazing art & architecture & met great people. We sent beautiful postcards home that we bought right there & you can even have them postmarked at the Vatican if you like! St. Peter's Basilica was even more extraordinary & awe-inspiring! Everything worked well, even when dodging all the traffic - we will return soon!	https://www.tripadvisor.com/ShowUserReviews-g187791-r46781087-Rome_Lazio.html
IsabelLoza	great city	I loved Rome, there are so many things to do there and great places to visit. If you're into art, architecture , monuments, historic places, great food and Bernini you have to visit Rome. I wouldn't take children, there aren't activities for them, they would get bored most likely. I suggest walking to get almost everywhere (unless is REALLY far) is the best way, there are so many things to see in every corner.	https://www.tripadvisor.com/ShowUserReviews-g187791-r164807939-Rome_Lazio.html

Figura 32: Ejemplos de co-ocurrencia de las palabras clave 'arte' y 'arquitectura'. Fuente: elaboración propia

Una vez obtenidas las co-ocurrencias, en esta fase del modelo se podría realizar un segundo filtro de palabras clave estableciendo un umbral con un valor mínimo de co-ocurrencias para cada palabra clave. En nuestra aplicación, no hemos establecido dicho umbral para reducir la red, ya que no tenemos un volumen elevado de palabras clave.

En la Figura 33 se muestra la red completa de palabras clave obtenida con Vosviewer:

pequeños nos permitirán identificar experiencias específicamente relacionadas mientras que los clústeres grandes contendrán mezclas de diferentes tipos de experiencias.

El parámetro que permite obtener clústeres pequeños es el parámetro de resolución. Este parámetro determina el nivel de detalle del clustering producido por la técnica de clustering VOS. Cuanto mayor sea el valor del parámetro, mayor será el número de clústeres producidos por la técnica de clustering VOS con el fin de evitar clústeres grandes.

Por otro lado, es conveniente manejar un tamaño mínimo de clúster en términos de número de palabras clave. Un número muy bajo de palabras clave no definiría bien una CDE. El parámetro de tamaño mínimo de clúster nos permite gestionar esta opción. Los clústeres que no tengan el tamaño mínimo definido en el parámetro anterior se fusionarán. Tras realizar diferentes pruebas obtuvimos los siguientes valores óptimos:

- Resolución (*resolution*): 4
- Tamaño mínimo de los clústeres (*minimum cluster size*): 6

Obtuvimos entonces 61 clústeres, cuya visualización es conveniente realizar a nivel individual. La Figura 34 muestra la red completa con todos los clústeres.



Figura 34: Red completa de palabras clave con los 61 clústeres obtenidos. El tamaño de cada palabra clave es proporcional al número de ocurrencias de la misma. Fuente: elaborado con Vosviewer

5.4.7. Visualización e interpretación

En esta fase del modelo se visualizan individualmente algunos de los clústeres más relevantes que se han obtenido y se intenta interpretar el resultado de los mismos. Conviene señalar que, al evaluar los resultados, si los clústeres obtenidos no son satisfactorios, debemos volver a las fases anteriores y realizar cambios (por ejemplo, el valor del número mínimo de ocurrencias de una palabra clave) en la fase de 'limpieza y refinamiento', así como en las siguientes fases: 'extracción de redes', 'normalización' y 'clustering', con el fin de obtener mejores resultados. En otras palabras, planteamos el modelo como un proceso iterativo para refinar y mejorar los clústeres obtenidos.

A continuación, se muestran algunos de los clústeres más relevantes obtenidos en nuestra aplicación del caso de uso. Los clústeres obtenidos representan CDEs como se detalla a continuación. Para obtener estas visualizaciones a nivel de clúster hemos

utilizado Vosviewer y nos hemos apoyado en el lenguaje de programación Python para obtener los clústeres individuales a partir de la red completa generada por Vosviewer.

Clústeres 17 y 29:

En la Figura 35 encontramos los elementos del clúster 17 y la Figura 36 los del clúster 29:

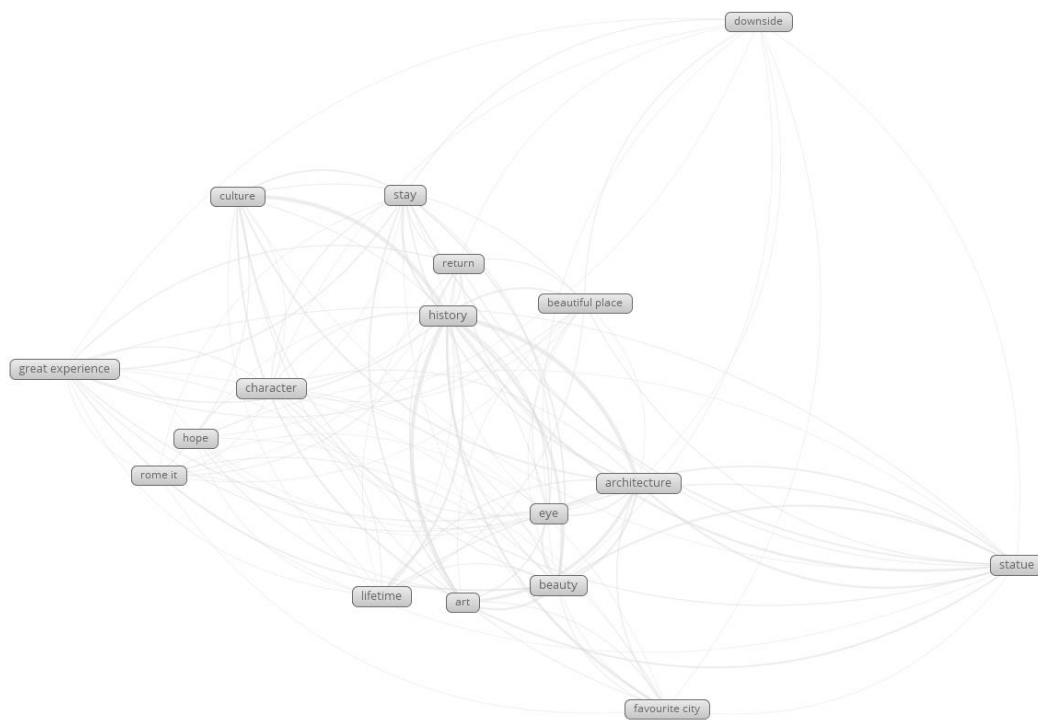


Figura 35: Subgrafo del clúster 17. Fuente: elaborado con Vosviewer.

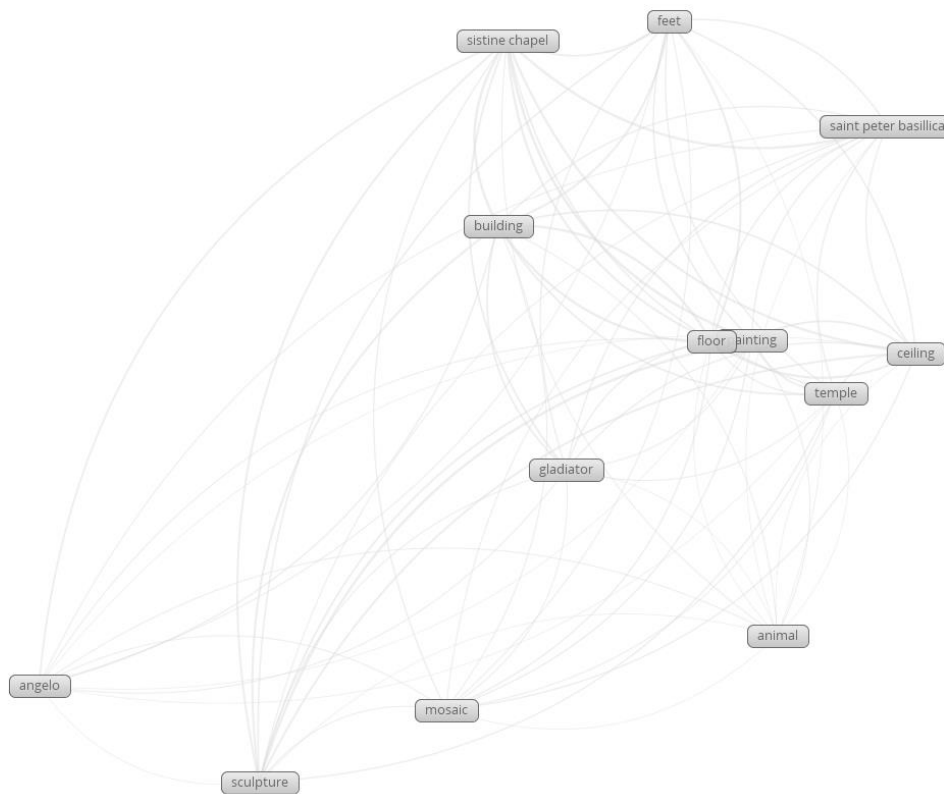


Figura 36: Subgrafo del clúster 29. Fuente: elaborado con Vosviewer

Revisando ambos clústeres podemos intentar extraer una CDE. En estos clústeres encontramos experiencias positivas ('great experience', 'beautiful place', 'favorite city') de los turistas relacionados con el arte, la arquitectura o la historia de Roma, que son sin duda algunos de los grandes atractivos de la ciudad.

Clústeres 3 y 11:

En la Figura 37 encontramos los elementos del clúster 11 y la Figura 38 los del clúster 3:

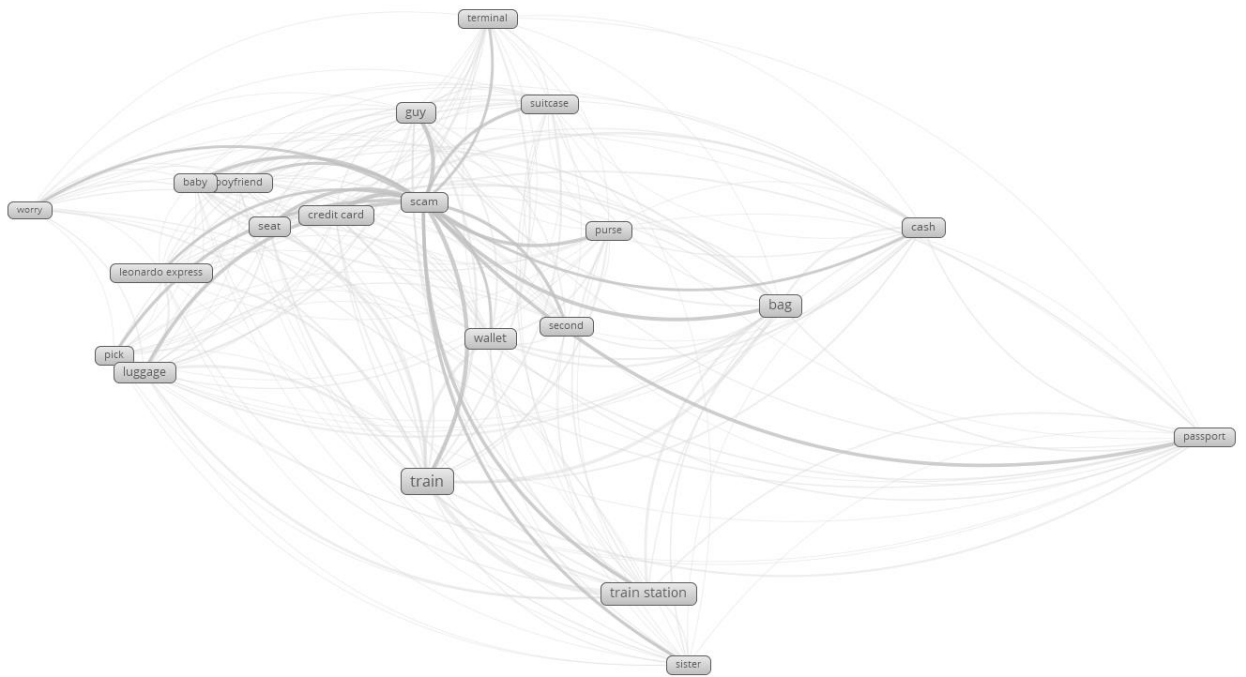


Figura 37: Subgrafo del clúster 11. Se destacan las relaciones de la palabra clave 'scam'. Fuente: elaborado con Vosviewer.

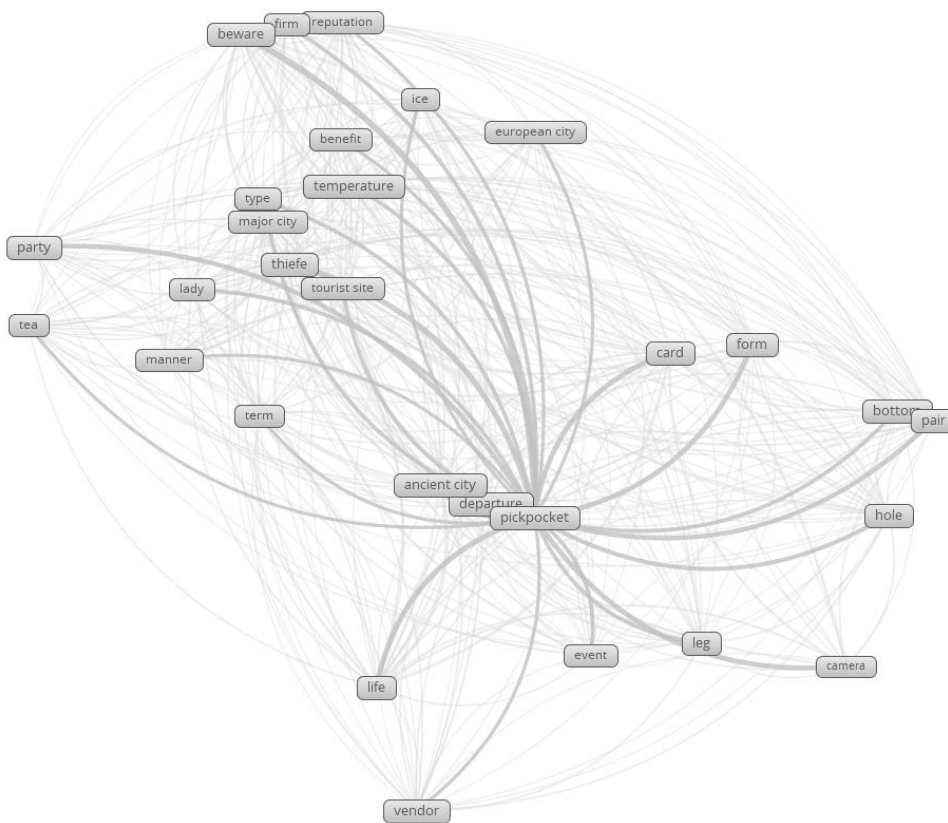


Figura 38: Subgrafo del clúster 3. Se destacan las relaciones de la palabra clave 'pickpocket'. Fuente: elaborado con Vosviewer.

Dentro de estos clústeres encontramos CDEs relacionados con la seguridad turística en Roma, referidos a posibles situaciones de inseguridad ('scam', 'pickpocket', 'thief',

'beware') y elementos relacionados ('wallet', 'suitcase', 'card'), al igual que lugares ('terminal', 'train station'). Hemos destacado las relaciones de las palabras clave que consideramos más relevantes en estos clústeres: 'pickpocket' y 'scam' en ambas figuras.

Clústeres 40, 46 y 59:

En la Figura 39 encontramos los elementos del clúster 40, en la Figura 40 los elementos del clúster 46 y la Figura 41 los del clúster 59:

Encontramos los siguientes elementos en estos clústeres:

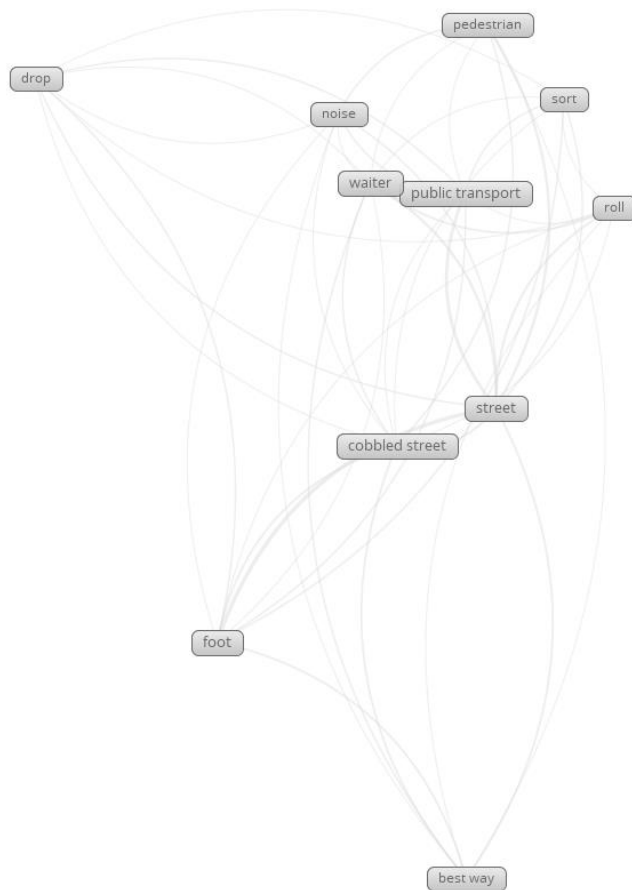


Figura 39: Subgrafo del clúster 40. Fuente: elaborado con Vosviewer

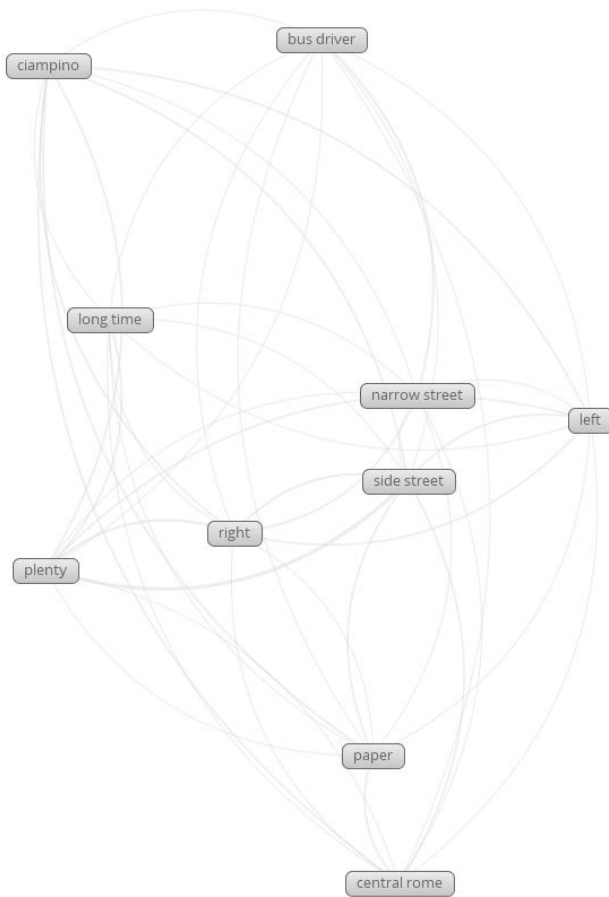


Figura 40: Subgrafo del clúster 46. Fuente: elaborado con Vosviewer

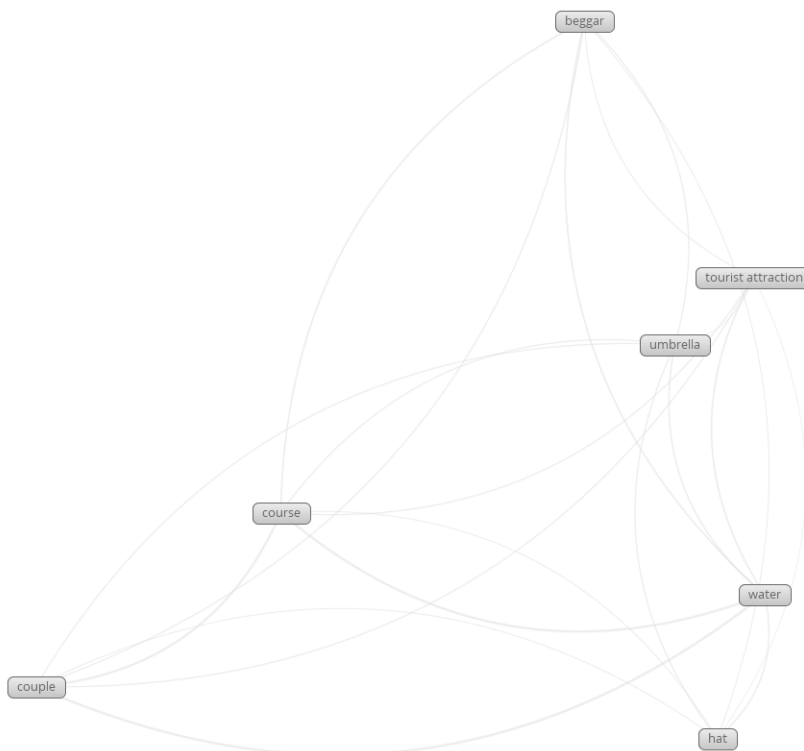


Figura 41: Subgrafo del clúster 59. Fuente: elaborado con Vosviewer.

En estos clústeres podemos deducir CDEs relacionadas con la vida en las calles de Roma: características físicas de las calles ('cobbled street', 'narrow street'), el ambiente ('plenty', 'noise') o cómo moverse por ellas ('pedestrian', 'public transport', 'best way'). Más concretamente, en el clúster 59 encontramos recomendaciones para la vida cotidiana del turista en Roma, como el uso de sombreros, paraguas o llevar agua.

Los clústeres mostrados son una selección del total de clústeres generados. Además de los mostrados, existen otros clústeres que hacen referencia a otros tipos de CDE, como los relacionados con itinerarios turísticos, alojamiento, transporte, alimentación y organización de viajes.

Análisis personalizado. Relaciones relevantes entre palabras clave

Como ya hemos mencionado, una CDE representa un patrón de experiencia sobre el que podemos realizar exploraciones personalizadas. Hemos encontrado algunas palabras clave de especial relevancia en la red en relación con estas CDE, que consideramos interesante analizar con más detalle, ya que podrían proporcionarnos información adicional sobre experiencias concretas. Partiendo de estas palabras clave relevantes, es interesante explorar las palabras clave relacionadas, que también pueden ser palabras clave de otros clústeres.

Seleccionamos las siguientes palabras clave: 'pickpocket' y 'crowd'.

Pickpocket:

Es una palabra clave con un total de 63 apariciones y está directamente relacionada con las CDEs relacionados con la seguridad. En la Figura 42, mostramos las palabras clave más estrechamente relacionadas con ella. Para ello, hemos aplicado un filtro para mostrar sólo las palabras clave en las que el número mínimo de co-ocurrencias es 10.

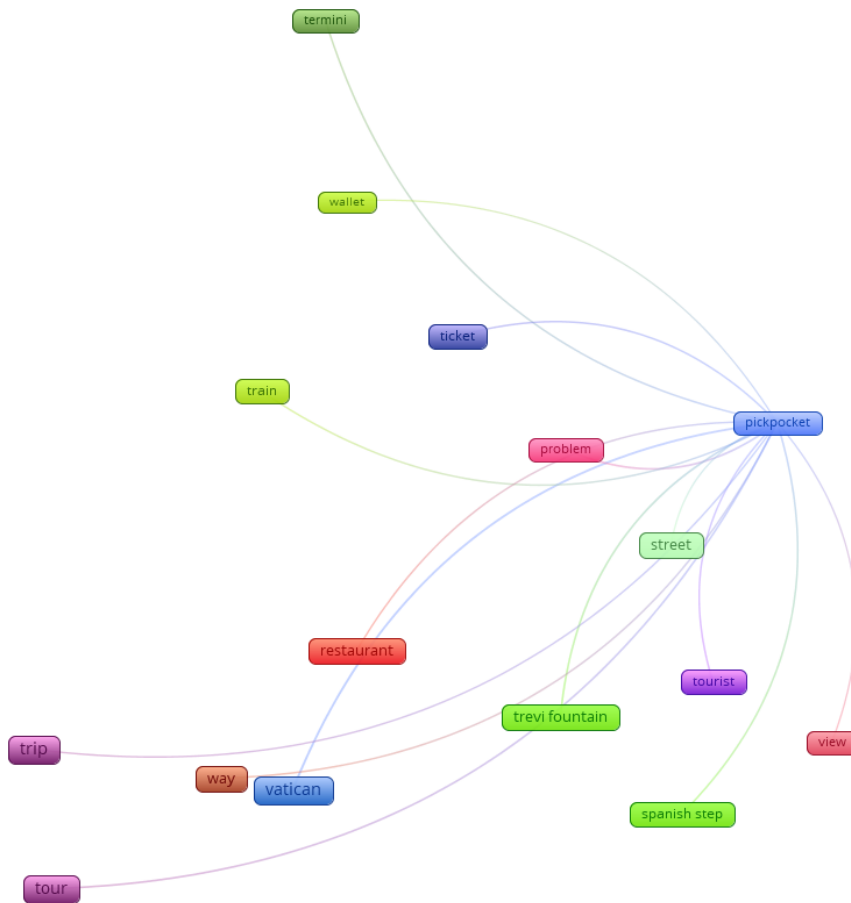


Figura 42: Subgrafo de relaciones de la palabra clave 'pickpocket'. Se muestran sólo aquellas relaciones con un número de co-ocurrencias mayor o igual a 10. Los colores muestran los distintos clústeres a los que pertenecen las palabras clave. Fuente: elaborado con Vosviewer.

Las relaciones obtenidas pueden ser de interés para descubrir cuáles son las zonas con mayor actividad de carteristas en Roma.

Crowd:

Es una palabra clave con un total de 57 ocurrencias y está directamente relacionado con las CDEs relacionadas con itinerarios turísticos (véase Figura 43), ya que puede avisarnos de posibles aglomeraciones en determinadas zonas.

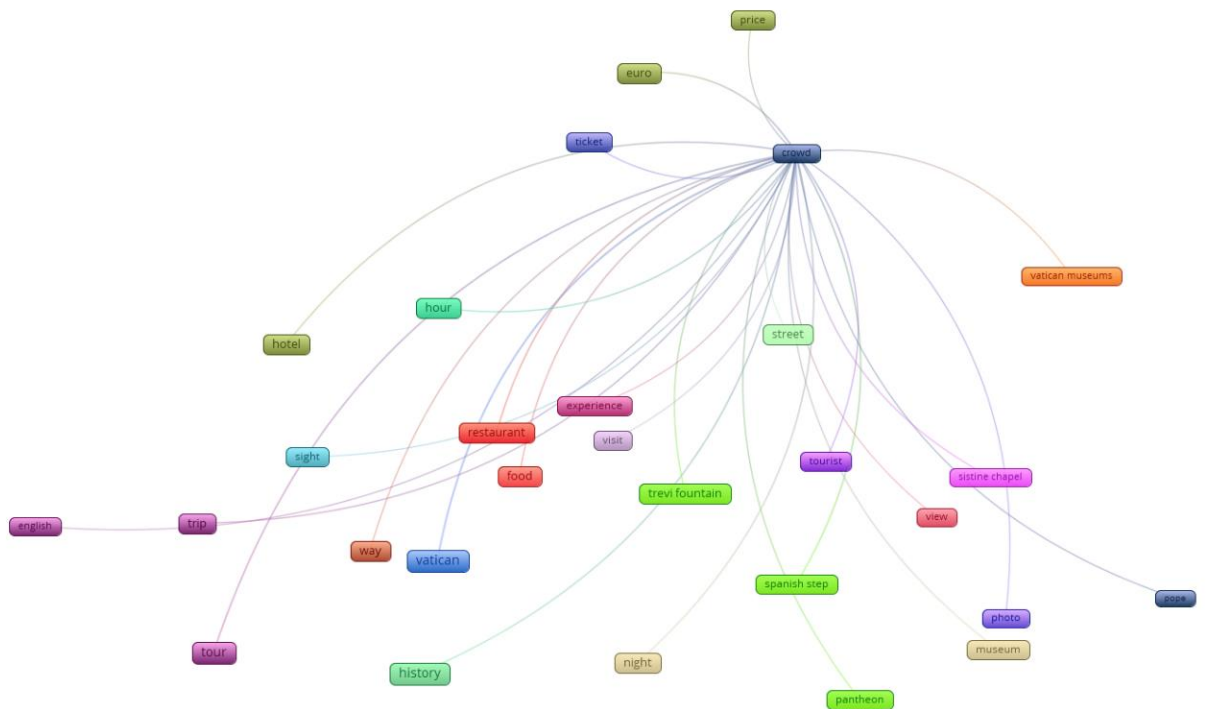


Figura 43: Subgrafo de relaciones de la palabra clave 'crowd'. Se muestran sólo aquellas relaciones con un número de co-ocurrencias mayor o igual a 10. Los colores muestran los distintos clústeres a los que pertenecen las palabras clave. Fuente: elaborado con Vosviewer.

5.4.8. Toma de decisiones

La toma de decisiones se basará en recomendaciones generales y recomendaciones personalizadas. Las primeras se basarán en los clústeres obtenidos que generarán recomendaciones generales o acciones de mejora y las segundas se basarán en recomendaciones personalizadas que dependerán de las exploraciones "a medida" realizadas dentro de los clústeres de palabras clave que consideremos relevantes en nuestro análisis, tal y como se muestra en los análisis personalizados del apartado 5.4.7.

5.4.8.1. Recomendaciones generales

El conjunto de clústeres generados muestra un mapa global de CDEs que podría ser útil a los gestores de un destino turístico para obtener una visión global de las experiencias turísticas más relevantes y llamativas. Esto les permitiría detectar áreas a mejorar y tomar decisiones en función de la experiencia turística. La Tabla 9 muestra algunos ejemplos de iniciativas concretas.

CDE	Información relevante obtenida	Iniciativa de mejora
17 & 29	Se detectan experiencias positivas relacionadas con el arte y la arquitectura.	Promoción de visitas guiadas de arte y arquitectura y actividades relacionadas.
3 & 11	Se detectan experiencias negativas relacionadas con la seguridad.	Mejorar la seguridad pública, especialmente en las zonas de gran afluencia turística. Mejorar la percepción turística de la seguridad.
40, 46 & 59	Los turistas comparten experiencias relacionadas con recomendaciones útiles que facilitan el día a día del turista.	Compartir estas útiles recomendaciones para facilitar el día a día del turista en las oficinas de turismo y puntos de información turística.

Tabla 9: Iniciativas de mejora basadas en la información relevante obtenida de las recomendaciones generales.

5.4.8.2. Recomendaciones personalizadas

Además de estas recomendaciones generales, también podemos obtener recomendaciones personalizadas. El análisis en profundidad de las palabras clave relevantes puede ayudar a comprender mejor una experiencia específica, como se muestra en la Tabla 10.

Palabra clave	Información relevante obtenida	Iniciativa de mejora
pickpocket	Se obtienen áreas específicas en las que se han detectado problemas de seguridad, así como incidentes concretos.	Se definen iniciativas específicas de mejora de la seguridad centradas en determinados tipos de incidentes en puntos concretos.
crowd	Se obtienen zonas de gran afluencia y posibles aglomeraciones.	Mejorar la gestión de accesos, tickets y tiempos de espera para evitar aglomeraciones en momentos puntuales.

Tabla 10: Iniciativas de mejora basadas en la información relevante obtenida de las recomendaciones personalizadas.

5.5. Conclusiones

En este caso de uso se ha presentado un modelo de detección y análisis de CDEs en un destino basado en las reseñas de los turistas. Este modelo se ha aplicado a un caso real

de un destino turístico como es la ciudad de Roma obteniendo los datos de las reseñas realizadas por los turistas sobre esta ciudad en la plataforma Tripadvisor.

5.5.1. Implicaciones teóricas

Se ha definido formalmente el concepto de CDE, y también se han detallado los pasos para obtenerlo a partir de las reseñas de los turistas. Este concepto podría ser útil en el análisis de las experiencias en un destino turístico. Se propone un modelo para detectar y analizar estas CDEs en un destino turístico determinado, que trabaja con las reseñas realizadas en plataformas digitales por los propios turistas. Consideramos que el punto fuerte del modelo es la posibilidad de aplicarlo a cualquier destino turístico del que se disponga de reseñas de turistas. Este modelo puede ser utilizado en investigaciones en el contexto de la competitividad de los destinos turísticos. Desde el punto de vista de la aplicación, podría ser interesante aplicar el modelo para comparar destinos desde la perspectiva de la experiencia turística, una línea de investigación interesante podría ser la comparación de grandes destinos urbanos europeos como Roma, Londres, París o Madrid.

5.5.2. Implicaciones en la gestión empresarial

El modelo propuesto puede ser muy útil para los gestores de destinos turísticos. A partir del modelo se puede obtener un mapa general de CDEs del destino turístico, a partir del cual se pueden hacer recomendaciones generales. En la aplicación del caso de uso, se obtienen recomendaciones sobre cuestiones como la seguridad pública o recomendaciones prácticas sobre cómo mejorar el día a día del turismo. Además, el modelo permite profundizar en una CDE concreta analizando palabras clave específicas relacionadas con la CDE y encontrando palabras clave asociadas que permiten realizar un análisis personalizado del mismo. En la aplicación desarrollada, se revisa la seguridad pública identificando lugares concretos del destino turístico en los que es necesario mejorar la seguridad, así como lugares concretos en los que se producen aglomeraciones o tiempos de espera. Esta información relevante obtenida a partir de recomendaciones generales o personalizadas puede ayudar a implementar medidas concretas para mejorar aspectos de seguridad o acciones operativas para evitar tiempos de espera, evitar aglomeraciones y mejorar los itinerarios turísticos. Así mismo, la información obtenida sobre recomendaciones prácticas para mejorar el día a día de los turistas puede ser utilizada por los destinos turísticos en sus comunicaciones y marketing. El caso de Roma

representa un destino urbano de gran interés. El modelo podría ser útil en la transición de un destino turístico tradicional hacia un destino turístico inteligente que mejore sus productos y servicios basándose en los datos y centrado en el turista.

5.6. Trabajo futuro

Hay varios ámbitos en los que se podría trabajar en el futuro. En primer lugar, el modelo podría proponer pasos para generar una red de palabras clave que admita alternativas. Por ejemplo, se podrían utilizar otras técnicas de normalización de la red u otras técnicas de clustering a la hora de generar los grupos de palabras clave. Por otro lado, se podría añadir información adicional sobre las CDEs obtenidas, como clasificar esta CDE como positiva, negativa o neutro añadiendo análisis de sentimiento (Kim et al., 2017) a las reseñas turísticas. En el caso de obtener estas reseñas de una red social, la información de los 'me gusta' obtenidos por la reseña podría darnos información sobre el impacto de las reseñas y en consecuencia de las CDEs derivadas de las mismas. Del mismo modo, en el caso de datos procedentes de una red social, también podría ser interesante añadir información sobre el autor, ya que cada autor puede tener una cierta relevancia e impacto en la red social. Adoptando un enfoque diferente, también podríamos tratar de analizar una experiencia dada desde diferentes perspectivas como las propuestas por Yu et al. (2021): emocional, cognitiva y conductual, con el fin de obtener una comprensión más profunda de la experiencia turística. Con objeto de mejorar la interpretación de los clústeres de la red generados y obtener las CDEs, se podría utilizar inteligencia artificial generativa. En particular, se podría emplear un modelo LLM al que como entrada (prompt) se le facilitaran las palabras clave que componen un clúster y se le solicitara su interpretación. Previamente este modelo LLM recibiría un entrenamiento en el concepto de CDE, facilitándole ejemplos de clústeres y su CDE correspondiente.

6. Caso de uso: enfoque de criterios múltiples aplicado a la transformación digital de tiendas físicas en el sector turístico. Aplicación en agencias de viajes

No cabe duda de que la transformación digital en el sector turístico ha llegado para quedarse. Para las empresas de este sector, esta digitalización es esencial para seguir siendo competitivas, mejorar la experiencia del cliente y aumentar su compromiso. En este contexto, aquellos actores de la cadena de valor del turismo que operan a través de tiendas físicas han visto cómo los operadores puramente online han ganado cuota de mercado.

Este caso de uso desarrolla un modelo MCDM para tiendas físicas del sector turístico que quieren digitalizarse, proporcionando un enfoque sistemático para gestionar las prioridades de inversión que involucra a todas las áreas de la organización. Para integrar eficientemente todas las áreas de la organización en el proceso de toma de decisiones, nos basamos en las áreas definidas por el Cuadro de Mando Integral. Se desarrolla una aplicación concreta en las tiendas físicas de las agencias de viajes, en el que se utilizan procesos de análisis jerárquico y se obtiene como resultado un ranking de prioridades en la adopción de dispositivos digitales.

En la Figura 44 podemos observar cómo se integra el caso de uso dentro del marco conceptual propuesto en el apartado 4.2. Tal y como se explicará en el desarrollo del mismo, en lo relativo a fuentes de datos se trabajará con información de expertos que contendrá evaluaciones. Estos datos se ingestarán en el área de almacenamiento y gestión de los datos y se realizará sobre los mismos una revisión para asegurar su consistencia. Posteriormente se aplicarán modelos MCDM. Los resultados de los modelos se aplicarán a la estrategia de digitalización de empresas de servicios turísticos, concretamente agencias de viajes.

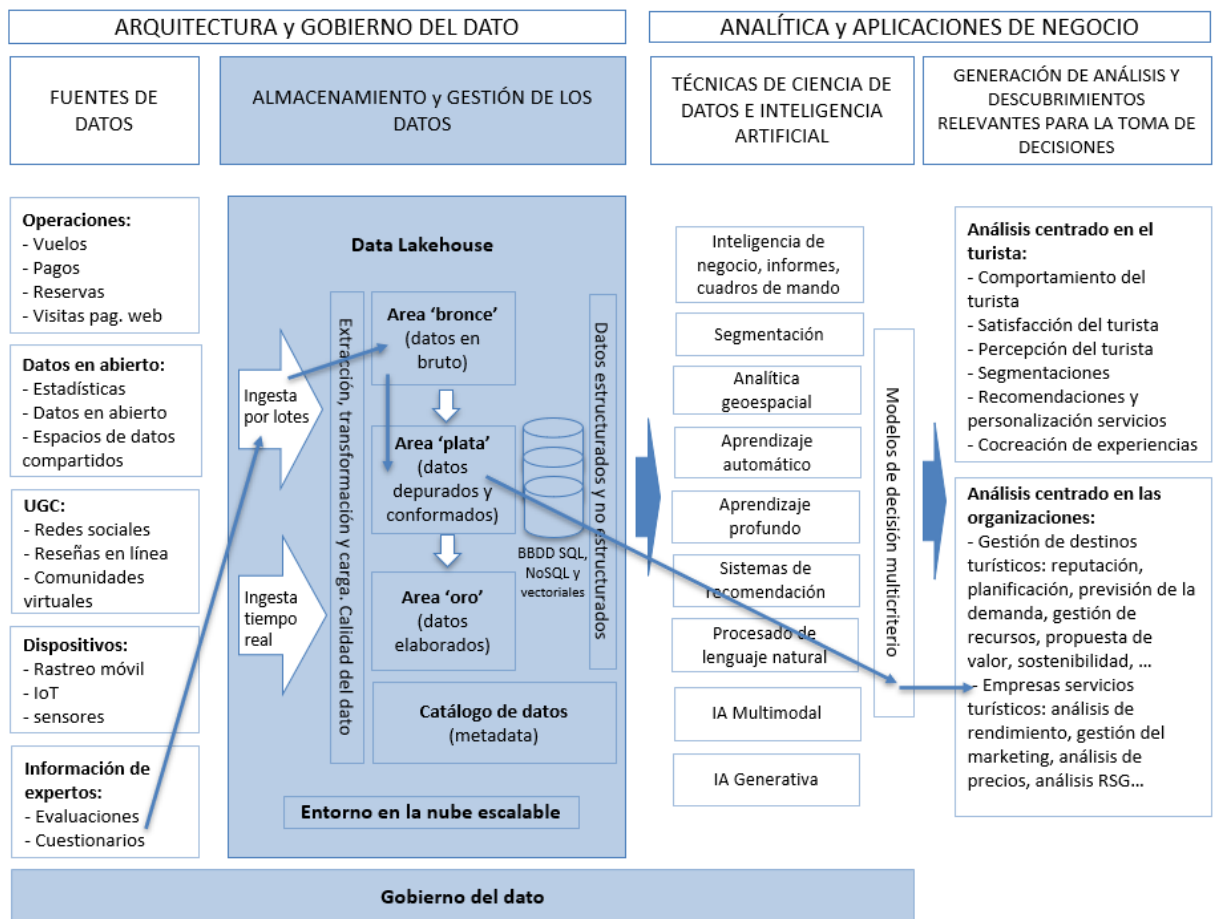


Figura 44: Integración del caso de uso 'Enfoque de criterios múltiples aplicado a la transformación digital de tiendas físicas en el sector turístico. Aplicación en agencias de viajes' dentro del marco conceptual propuesto. Fuente: elaboración propia.

6.1. Introducción

No cabe duda de que la ciencia de datos ha transformado radicalmente la manera en que las empresas interactúan con sus clientes, permitiendo una personalización sin precedentes y una comprensión más profunda de sus necesidades y preferencias. En particular, una de las áreas con mayor impacto es la experiencia del cliente. En el ámbito de la experiencia de cliente, podemos definir el viaje del cliente como el itinerario por el que un consumidor interactúa gradualmente con la empresa a través de puntos de interacción a lo largo del proceso de compra. Este término es usualmente referenciado por su término en inglés *customer journey*, término que usaremos en lo sucesivo en esta memoria. Como explica Richardson (2010), estos puntos de interacción pueden ser tanto físicos como online y comprenden los pasos desde que tienen una necesidad hasta que completan su compra y evalúan y comparten su experiencia como cliente. Situamos el customer journey dentro de una estrategia de orientación al cliente, que implica una gestión centrada en la creación de valor para los clientes (Kumar y Reinartz, 2016). Un

aspecto fundamental de esta orientación de la gestión es la experiencia del cliente, tal y como describen Pine y Gilmore (1999), que abordan específicamente la importancia de las experiencias en la sociedad actual y cómo la creación de experiencias sólidas y duraderas de los clientes puede ser beneficiosa para las empresas.

Esta experiencia del cliente debe gestionarse desde múltiples canales de interacción con el cliente. La gestión omnicanal se ha convertido en la nueva norma (Verhoef et al., 2015). Esta tendencia se ha visto acelerada por la pandemia de Covid19. En el caso de las grandes agencias de viajes con una fuerte presencia física, el comercio online y offline tienen que trabajar de forma coordinada. En este nuevo escenario, el término *phygital* (que es combinación de las palabras *physical* y *digital* y que usaremos en lo sucesivo en esta memoria) parece reflejar esta realidad (Belghiti et al., 2018). La estrategia phygital propone la integración de la experiencia online y offline, de forma que la diferencia entre la tienda física y la tienda digital se diluya para mejorar la experiencia global del usuario.

Dentro de la cadena de valor de la industria turística, encontramos actores que operan puramente en línea, que han supuesto una fuerte competencia para los que operan en tiendas físicas. Esto ha obligado a estos últimos a evolucionar y reinventarse. Un claro ejemplo es el de las agencias de viajes que operan en tiendas físicas, que han tenido que acometer un proceso de digitalización que incluye la incorporación de nuevas tecnologías en sus procesos y en el punto de venta, con el fin de mejorar la experiencia del cliente a través de los entornos físico y digital como medio para competir con los actores puramente online. Los dispositivos y herramientas tecnológicas en las tiendas, juegan un papel decisivo en el aumento de la percepción de calidad y satisfacción del cliente; particularmente en las agencias de viajes estas herramientas mejoran sustancialmente la interacción personal y los resultados (Caro & García, 2008; Viljoen & Roberts-Lombard, 2016). Pencarelli et al. (2020) analizaron la adopción de tecnologías digitales en la tienda por parte de las agencias de viajes tradicionales y las oportunidades de mejora que ofrecen en la relación con los clientes durante el proceso de compra en la tienda. En una línea de investigación similar, Dini et al. (2022) abordaron el papel que desempeñan las nuevas herramientas y dispositivos tecnológicos utilizados en las interacciones con los clientes en las agencias de viajes, analizando las percepciones de los propietarios y gestores de agencias de viajes, antes y después de Covid-19, sobre la eficacia de estas herramientas.

En este contexto, el objetivo de este caso de uso es contribuir al análisis de cómo las tiendas físicas del sector turístico y, en particular, las agencias de viajes pueden mejorar la experiencia del cliente en la tienda mediante el uso de dispositivos digitales. En concreto, se desarrolla un marco para facilitar la toma de decisiones sobre la inversión en

tecnología en la tienda, dentro del amplio abanico de opciones de que disponen. Se analiza la relación entre diferentes tecnologías en tienda y se deriva la mejor opción posible para la agencia de viajes según diferentes criterios. Al tratarse de un proyecto estratégico, se propone el uso del BSC. Por lo tanto, el caso de uso implica la participación de las cuatro áreas clave planteadas por el BSC y descritas en el apartado 3.3: clientes, finanzas, procesos internos y aprendizaje y crecimiento. Estas cuatro áreas deben participar en la clasificación de los dispositivos digitales alternativos que fueron preseleccionados a partir de la revisión de la literatura. Estos dispositivos serán clasificados de acuerdo con los criterios definidos por el CEO de la empresa en relación con elementos tales como el coste de instalación, el coste de mantenimiento, la rentabilidad o el retorno potencial de la inversión, la experiencia del cliente, la salud y la seguridad, la fiabilidad, la gestión de dispositivos, la formación y la innovación.

Los dispositivos incluidos en la aplicación del caso de uso a las agencias de viajes fueron los siguientes: monitor o pantalla LCD, realidad aumentada 3D (AR por sus siglas en inglés *Augmented Reality*) y realidad virtual (VR por sus siglas en inglés *Virtual Reality*), tablet, código QR, sala de espera de clientes, pantalla grande de vídeo, proyección de vídeo en paneles y Wi-Fi para clientes. Los autores consideran apropiado el uso del AHP para modelar este problema de decisión, ya que este modelo permite obtener un resultado final consistente, consensuado entre las diferentes áreas de la organización y además existe una jerarquía implícita en la estructura de la organización.

Encontramos en la literatura casos de aplicación a la digitalización de tiendas físicas (Llorens et al., 2022) en los que se desarrolla una metodología completa que integra BSC y AHP. La novedad de nuestro caso de uso es que aplicamos la metodología indicada a la digitalización de tiendas físicas del sector turístico, en particular a las agencias de viajes.

Las conclusiones generales de este caso de uso suponen una contribución relevante al creciente corpus de investigación en el ámbito de la transformación digital del sector del turismo y en particular proporciona una valiosa ayuda a los profesionales a la hora de decidir su plan de inversión en la digitalización de las agencias de viajes. El resto de este caso de uso está organizado como sigue: en el apartado 6.2 se propone nuestro modelo basado en MCDM utilizando como aplicación real una agencia de viajes en la que se sugieren planes de acción alternativos con la cartera de dispositivos obtenidos en la revisión bibliográfica y su priorización. En el apartado 6.3 se presentan las conclusiones y el trabajo futuro.

6.2. Modelo propuesto

Tal y como se ha comentado anteriormente, en este apartado presentamos un modelo MCDM para dirigir la estrategia de digitalización de una agencia de viajes. Este modelo se basa en el proceso analítico jerárquico (AHP) y en el cuadro de mando integral (BSC) presentados en los apartados 43.2 y 43.3 dentro del marco teórico de investigación de esta memoria.

En el apartado 6.1 se han mencionado diferentes trabajos en los que se analiza el uso de diferentes dispositivos digitales en la digitalización de tiendas físicas y en particular en agencias de viajes. Basándonos en esta revisión de la literatura, hicimos una selección de dispositivos digitales (Pencarelli et al., 2020; Dini M. et al., 2022) para nuestra aplicación del caso de uso a agencias de viajes. La Tabla 11 muestra la lista de dispositivos digitales seleccionados como posibles alternativas para nuestro modelo de decisión.

Dispositivo digital	Descripción
Monitor o pantalla LCD (<i>monitor or LCD screen</i>)	Monitor o pantalla LCD (<i>Liquid Crystal Display</i>) para presentar contenidos multimedia
3D (AR y VR)	La realidad virtual (VR) y la realidad aumentada (AR) permiten vivir experiencias de viaje o visitar destinos en 3D
Tablet	Dispositivo móvil con pantalla táctil que proporciona información sobre productos y servicios
Código QR (<i>QR code</i>)	Más información y ofertas disponibles en el smartphone a través de QR (<i>Quick Response</i>). Esta tecnología garantiza un acceso sin contacto, se utiliza para reducir los puntos de contacto
Sala de espera de clientes (<i>Customer waiting room</i>)	Sala de espera con soluciones de entretenimiento para los clientes
Proyección de vídeo en pantalla grande (<i>Video projection in big screen</i>)	Proyección de vídeo en pantalla grande de destinos que atraen a los turistas
Paneles con proyectores de vídeo (<i>Windows with video projectors</i>)	Paneles con proyectores de vídeo para ofertas
Wi-Fi clientes (<i>Customer Wi-Fi</i>)	Red Wi-Fi para clientes

Tabla 11: Lista de dispositivos digitales seleccionados

El modelo propuesto se estructura en base a los cuatro pasos de los modelos AHP detallados en los apartados 43.2.1 a 43.2.4 del marco teórico de investigación de esta memoria. En la Figura 45 se muestra el modelo propuesto:

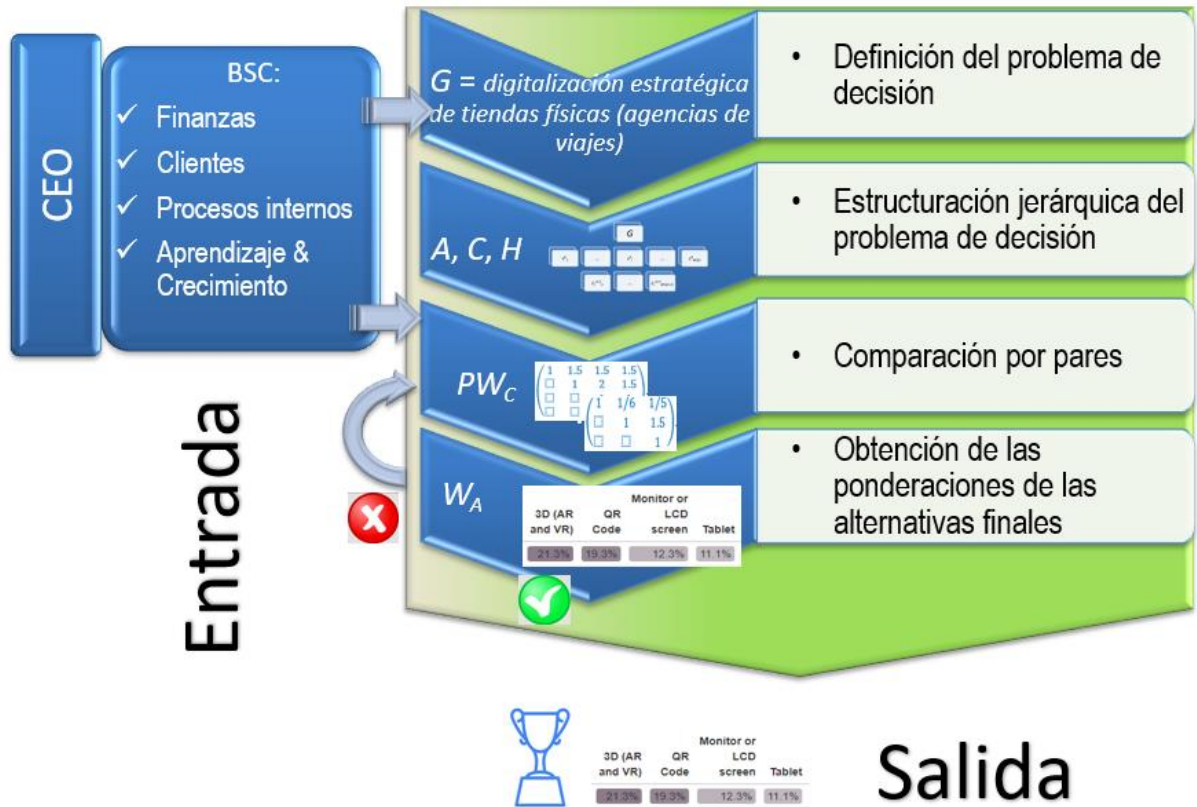


Figura 45: Propuesta de modelo MCDM para la digitalización estratégica de las tiendas físicas de una agencia de viajes. Fuente: elaboración propia

A continuación, en los siguientes subapartados se explican las diferentes etapas del modelo.

6.2.1. Definición del problema de decisión

Para desarrollar el modelo MCDM es necesario identificar el objetivo y los elementos básicos que forman la estructura del AHP (véase apartado 3.2.2). Nuestro objetivo es la selección del dispositivo para implantar en una tienda física de una agencia de viajes. La transformación de las tiendas físicas es una decisión que determinará el éxito a largo plazo de la empresa y, por tanto, es una decisión estratégica. Como se ha mencionado anteriormente, el BSC (Kaplan & Norton, 1992) es una herramienta de gestión del

rendimiento estratégico ampliamente utilizada que se centra en la gestión de la aplicación de una actividad estratégica basada en una perspectiva interempresarial que implica a las siguientes áreas: finanzas, clientes, procesos internos y aprendizaje y crecimiento. Una vez establecido el marco estratégico global de la empresa, este plan debe aplicarse a nivel de cada área. En base a lo argumentado, el conjunto de expertos de nuestro modelo serán los directores de departamento relacionados con estas cuatro áreas del BSC. Este conjunto de expertos establecerá tanto las alternativas a elegir como los criterios (clasificados por áreas) de nuestro modelo. Finalmente, la decisión final será tomada por el CEO, teniendo en cuenta la valoración de estas áreas.

6.2.2. Estructuración jerárquica del problema de decisión

Entre los criterios, C , evaluables dentro de las cuatro áreas del BSC, se han seleccionado aquellos que se consideran relacionados con el objetivo propuesto en nuestra aplicación del caso de uso:

- *Finanzas (Finance)*:
 - *Coste de instalación (Setup Cost)*: es el coste en que se incurre para adquirir el equipamiento del dispositivo.
 - *Coste de mantenimiento (Maintenance Cost)*: es el coste en que se incurre para mantener el equipo en buenas condiciones.
 - *Beneficio (Profit)*: es el potencial de un dispositivo para generar ingresos.
- *Clientes (Clients)*:
 - *Experiencia (Experience)*: es el impacto generado en el cliente a través de múltiples puntos de contacto relacionados con el dispositivo.
 - *Salud & Seguridad (Health & Safety)*: es el impacto sobre la percepción que tiene el cliente de la salud y la seguridad relacionadas con el dispositivo.
- *Procesos internos (Internal Processes)*:
 - *Fiabilidad (Reliability)*: capacidad de un dispositivo para funcionar siempre de acuerdo con sus especificaciones.
 - *Gestión (Management)*: incluye la facilidad de uso y administración del dispositivo desde un punto de vista técnico.
- *Aprendizaje & Crecimiento (Learning & Growth)*:
 - *Formación (Training)*: esta característica se refiere a la facilidad en términos de formación necesaria para que los empleados manejen correctamente estos dispositivos.

- *Innovación (Innovation)*: es el potencial de innovación y mejora de prestaciones que tiene el dispositivo.

Estos factores se evalúan sobre el conjunto de alternativas, A , que representan los dispositivos seleccionados obtenidos en la revisión bibliográfica y descritos en la Tabla 11: monitor o pantalla LCD, 3D (AR y VR), tablet, código QR, sala de espera de clientes, proyección de vídeo en pantalla grande, paneles con videoproyectores y Wi-Fi de clientes.

Una vez definidos los elementos del problema (objetivo, criterios y alternativas), construimos la estructura jerárquica H del problema MCDM, que consta de cuatro niveles, como se muestra en la Figura 46. Como ya se ha mencionado, el objetivo es obtener los dispositivos más adecuados para implantar en una tienda y este objetivo ocupa la posición más alta en la estructura jerárquica. Para tomar esta decisión, incluimos un primer nivel de criterios compuesto por las áreas del BSC. En el segundo nivel de criterios, cada área del BSC incluye sus correspondientes factores mencionados con anterioridad para tener en cuenta en el problema de decisión. En la última etapa, las alternativas se evalúan desde el punto de vista de cada criterio de nivel superior, es decir, desde la perspectiva de los gestores de las correspondientes áreas del BSC.

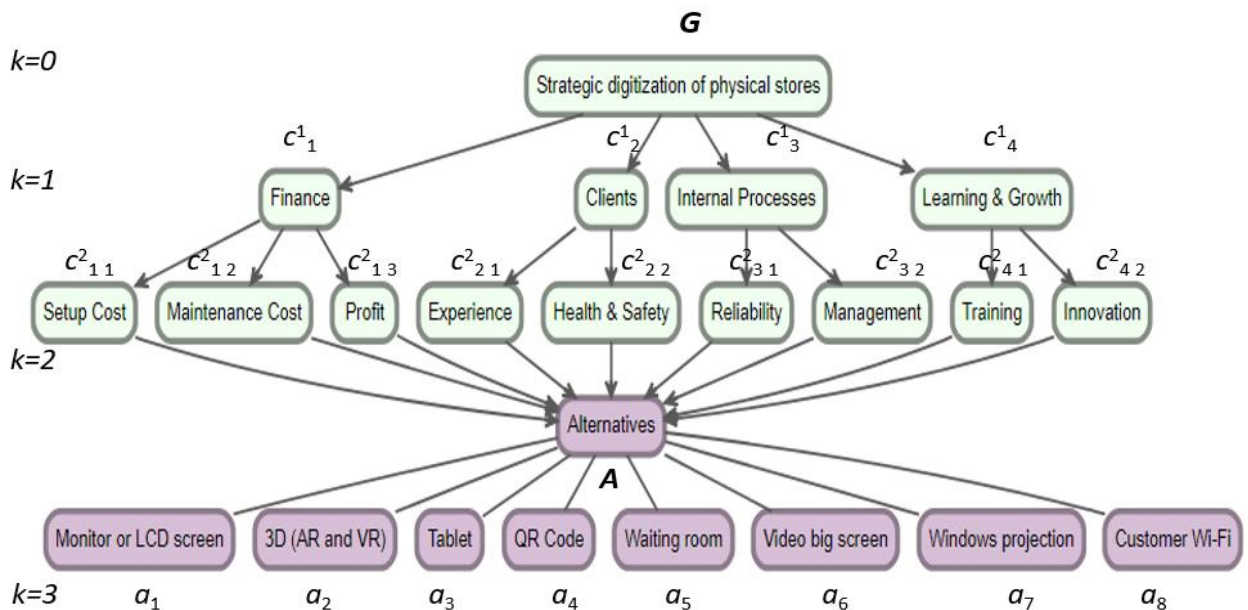


Figura 46: Jerarquía H para la estrategia de digitalización de tiendas físicas de agencias de viajes. Fuente: elaboración propia

6.2.3. Comparación por pares y cálculo del vector de ponderaciones

Como se menciona en el apartado 43.2.3 del marco teórico de investigación de esta memoria, una vez anidados los criterios, es necesario definir las matrices de comparación por pares. Para ello, el área del CEO y cada responsable del área del BSC correspondiente analizan los valores de los elementos de decisión de la jerarquía H e incorporan sus valoraciones realizando una comparación por pares de dichos elementos. En esta jerarquía, cada elemento de decisión del nivel superior, k , sirve para comparar los elementos de un nivel inmediatamente inferior, $k+1$, con respecto al anterior. Así, las alternativas ($k=3$) se comparan con respecto a los subcriterios ($k=2$), los subcriterios se comparan con respecto a los criterios ($k=1$) y los criterios se comparan con respecto al objetivo ($k=0$).

Las matrices definidas por cada responsable de la toma de decisiones se muestran a continuación:

- *CEO*. Las matrices de comparación por pares se construyen a partir de la comparación de la importancia de cada área del BSC en este problema de decisión con respecto al objetivo G , utilizando la escala de Saaty que se muestra en la Figura 20. Sobre los criterios $c(1) = \{c^1_1, \dots, c^1_4\}$, la matriz de pares proporcionada es:

$$PW_{c(1)} = \begin{pmatrix} 1 & 1.5 & 1.5 & 1.5 \\ & 1 & 1.5 & 1.5 \\ & & 1 & 1 \\ & & & 1 \end{pmatrix}.$$

- *Finanzas*. En primer lugar, evalúan los subcriterios dependientes del área, i.e. $c_1(2) = \{c^2_{11}, \dots, c^2_{13}\} = \{\text{Costes de instalación, costes de mantenimiento, Beneficio}\}$:

$$PW_{c_1(2)} = \begin{pmatrix} 1 & 1/6 & 1/5 \\ & 1 & 1 \\ & & 1 \end{pmatrix}.$$

Además, hay que evaluar las alternativas para cada criterio de nivel superior:

- *Costes de instalación*:

$$PW_{c_{11}(2)} = \begin{pmatrix} 1 & 3 & 1 & 1/5 & 5 & 3 & 3 & 1/3 \\ & 1 & 1/3 & 1/7 & 3 & 1 & 1 & 1/5 \\ & & 1 & 1/5 & 5 & 3 & 3 & 1/3 \\ & & & 1 & 9 & 7 & 7 & 3 \\ & & & & 1 & 1/3 & 1/3 & 1/7 \\ & & & & & 1 & 1 & 1/5 \\ & & & & & & 1 & 1/5 \\ & & & & & & & 1 \end{pmatrix}$$

- *Costes de mantenimiento:*

$$PW_{c_{12}(2)} = \begin{pmatrix} 1 & 3 & 1 & 1/3 & 7 & 5 & 7 & 1 \\ & 1 & 1/3 & 1/5 & 5 & 3 & 5 & 1/3 \\ & & 1 & 1/3 & 7 & 5 & 7 & 1 \\ & & & 1 & 9 & 7 & 9 & 3 \\ & & & & 1 & 1/3 & 1 & 1/7 \\ & & & & & 1 & 3 & 1/5 \\ & & & & & & 1 & 1/7 \\ & & & & & & & 1 \end{pmatrix}.$$

- *Beneficio:*

$$PW_{c_{13}(2)} = \begin{pmatrix} 1 & 1/5 & 1 & 5 & 3 & 1 & 1 & 5 \\ & 1 & 5 & 9 & 7 & 5 & 5 & 9 \\ & & 1 & 5 & 3 & 1 & 1 & 5 \\ & & & 1 & 1/3 & 1/5 & 1/5 & 1 \\ & & & & 1 & 1/3 & 1/3 & 3 \\ & & & & & 1 & 1 & 5 \\ & & & & & & 1 & 5 \\ & & & & & & & 1 \end{pmatrix}.$$

- *Clientes.* Con los subcriterios $c_2(2) = \{c^2_{21}, c^2_{22}\} = \{\text{Experiencia, Salud \& Seguridad}\}$ y esta matriz de comparación:

$$PW_{c_2(2)} = \begin{pmatrix} 1 & 1/2 \\ & 1 \end{pmatrix}$$

Las evaluaciones de las alternativas se muestran a continuación:

- *Experiencia:*

$$PW_{c_{21}(2)} = \begin{pmatrix} 1 & 1/5 & 1 & 5 & 3 & 1 & 1 & 3 \\ & 1 & 5 & 9 & 7 & 5 & 5 & 7 \\ & & 1 & 5 & 3 & 1 & 1 & 3 \\ & & & 1 & 1/3 & 1/5 & 1/5 & 1/3 \\ & & & & 1 & 1/3 & 1/3 & 1 \\ & & & & & 1 & 1 & 3 \\ & & & & & & 1 & 3 \\ & & & & & & & 1 \end{pmatrix}.$$

- *Salud \& Seguridad:*

$$PW_{c_2(2)} = \begin{pmatrix} 1 & 3 & 3 & 1/3 & 3 & 1 & 1 & 1 \\ & 1 & 1 & 1/5 & 1 & 1/3 & 1/3 & 1/3 \\ & & 1 & 1/5 & 1 & 1/3 & 1/3 & 1/3 \\ & & & 1 & 5 & 3 & 3 & 3 \\ & & & & 1 & 1/3 & 1/3 & 1/3 \\ & & & & & 1 & 1 & 1 \\ & & & & & & 1 & 1 \\ & & & & & & & 1 \\ & & & & & & & & 1 \end{pmatrix}$$

- *Procesos internos.* Con los subcriterios $c_3(2) = \{c^2_{31}, c^2_{32}\} = \{\text{Fiabilidad}, \text{Gestión}\}$ y esta matriz de comparación:

$$PW_{c_3(2)} = \begin{pmatrix} 1 & 1.5 \\ & 1 \end{pmatrix}.$$

La comparación por pares de las alternativas es la siguiente:

- *Fiabilidad:*

$$PW_{c_{31}(2)} = \begin{pmatrix} 1 & 2 & 1 & 1/2 & 1/2 & 1 & 2 & 1 \\ & 1 & 1/2 & 1/3 & 1/3 & 1/2 & 1 & 1/2 \\ & & 1 & 1/2 & 1/2 & 1 & 2 & 1 \\ & & & 1 & 1 & 2 & 3 & 2 \\ & & & & 1 & 2 & 3 & 2 \\ & & & & & 1 & 2 & 1 \\ & & & & & & 1 & 1/2 \\ & & & & & & & 1 \end{pmatrix}.$$

- *Gestión:*

$$PW_{c_{32}(2)} = \begin{pmatrix} 1 & 2 & 1 & 1/2 & 1 & 1 & 2 & 1 \\ & 1 & 1/2 & 1/3 & 1/2 & 1/2 & 1 & 1/2 \\ & & 1 & 1/2 & 1 & 1 & 2 & 1 \\ & & & 1 & 2 & 2 & 3 & 2 \\ & & & & 1 & 1 & 2 & 1 \\ & & & & & 1 & 2 & 1 \\ & & & & & & 1 & 1/2 \\ & & & & & & & 1 \end{pmatrix}$$

- *Aprendizaje & Crecimiento.* Con los subcriterios $c_4(2) = \{c^2_{41}, c^2_{42}\} = \{\text{Formación}, \text{Innovación}\}$ y esta matriz de comparación:

$$PW_{c_4(2)} = \begin{pmatrix} 1 & 1/4 \\ & 1 \end{pmatrix}.$$

Las matrices por pares de las alternativas son las siguientes:

- *Formación:*

$$PW_{c_{41}(2)} = \begin{pmatrix} 1 & 2 & 1 & 1/2 & 1 & 1 & 2 & 1/2 \\ & 1 & 1/2 & 1/3 & 1/2 & 1/2 & 1 & 1/3 \\ & & 1 & 1/2 & 1 & 1 & 2 & 1/2 \\ & & & 1 & 2 & 2 & 3 & 1 \\ & & & & 1 & 1 & 2 & 1 \\ & & & & & 1 & 2 & 1/2 \\ & & & & & & 1 & 1/3 \\ & & & & & & & 1 \end{pmatrix}.$$

○ *Innovación:*

$$PW_{c_{42}(2)} = \begin{pmatrix} 1 & 1/5 & 1 & 3 & 3 & 1 & 1 & 3 \\ & 1 & 5 & 9 & 9 & 5 & 5 & 9 \\ & & 1 & 3 & 3 & 1 & 1 & 3 \\ & & & 1 & 3 & 1/2 & 1/2 & 3 \\ & & & & 1 & 1/2 & 1/2 & 1 \\ & & & & & 1 & 1 & 1/2 \\ & & & & & & 1 & 1 \\ & & & & & & & 1 \end{pmatrix}$$

6.2.4. Cálculo del vector de ponderaciones para cada matriz y obtención de las ponderaciones de las alternativas finales mediante varios niveles de agregación de ponderaciones

Una vez completadas las matrices de comparación por pares por parte de los decisores, se obtiene el vector de ponderaciones de cada matriz, mediante la ecuación 1, y el correspondiente CI mediante la ecuación 2. A partir del CI obtenemos el CR, tal y como se explica en el apartado 43.2.3. Cualquier matriz cuyo CR supere el 10% es descartada y debe ser modificada por el decisor correspondiente ya que no se considera una decisión suficientemente racional. A partir de los vectores de ponderaciones obtenidos se procede a obtener las ponderaciones de las alternativas finales a través de varios niveles de agregación utilizando la jerarquía H , tal y como hemos explicado en el apartado 43.2.4. Los resultados finales de este paso se muestran en la Tabla 12: pesos de los distintos niveles del problema, ranking de las alternativas y el CR de las matrices de comparación.

	Weight	3D (AR and VR)	QR Code	Monitor or LCD screen	Tablet	Customer Wi-Fi	Video big screen	Windows projection	Waiting room	Inconsistency
Strategic digitization of physical stores	100.0%	21.3%	19.3%	12.3%	11.1%	10.4%	10.0%	8.5%	7.1%	0.8%
Finance	33.2%	7.9%	7.0%	4.3%	4.6%	3.6%	2.5%	2.2%	1.2%	0.4%
Maintenance Cost	15.7%	1.4%	5.5%	2.2%	2.6%	2.6%	0.7%	0.3%	0.3%	3.4%
Profit	14.7%	6.4%	0.4%	1.7%	1.7%	0.4%	1.7%	1.7%	0.8%	2.4%
Setup Cost	2.8%	0.1%	1.1%	0.3%	0.3%	0.6%	0.1%	0.1%	0.1%	2.9%
Clients	27.1%	4.8%	6.1%	3.4%	1.9%	2.8%	3.4%	3.4%	1.3%	0.0%
Health & Safety	18.0%	0.9%	5.9%	2.4%	0.9%	2.4%	2.4%	2.4%	0.9%	0.7%
Experience	9.0%	3.9%	0.2%	1.0%	1.0%	0.4%	1.0%	1.0%	0.4%	2.1%
Internal Processes	19.9%	1.3%	4.4%	2.4%	2.4%	2.4%	2.4%	1.3%	3.5%	0.0%
Reliability	11.9%	0.7%	2.5%	1.4%	1.4%	1.4%	1.4%	0.7%	2.5%	0.2%
Management	8.0%	0.5%	1.9%	1.0%	1.0%	1.0%	1.0%	0.5%	1.0%	0.1%
Learning & Growth	19.9%	7.4%	1.9%	2.2%	2.2%	1.7%	1.8%	1.6%	1.1%	0.0%
Innovation	15.9%	7.1%	1.0%	1.8%	1.8%	0.9%	1.3%	1.4%	0.6%	6.3%
Training	4.0%	0.2%	0.8%	0.5%	0.5%	0.8%	0.5%	0.2%	0.5%	0.5%

Tabla 12: Clasificación de las alternativas en el problema MCDM, pesos de los distintos niveles de la jerarquía y CR.

Si observamos los resultados de la Tabla 12, podemos ver los pesos de las áreas del BSC: finanzas (33,2%), clientes (27,1%), procesos internos (19,9%) y aprendizaje y crecimiento (19,9%). Con estos pesos vemos que la primera alternativa en el ranking es el 3D (AR y VR) (21,3%), seguido del código QR (19,3%) y en tercera posición monitor o pantalla LCD (12,3%). Revisando la columna 3D (AR y VR), podemos ver el detalle de los pesos de los diferentes niveles del problema añadidos a la posición en el ranking. En concreto, vemos como el mayor peso lo tiene el área de finanzas (7,9%) concretamente desde el criterio de beneficio (6,4%) y el área de aprendizaje y crecimiento (7,4%) concretamente desde el criterio de innovación (7,1%). Esto refleja la importancia de mejorar la innovación para aumentar la creación de valor y el beneficio de la tienda. Por otro lado, vemos que este dispositivo ha sido penalizado por el criterio de salud y seguridad (0,9%) probablemente debido a la pandemia de COVID-19, considerando la seguridad del dispositivo.

6.3. Conclusiones y trabajo futuro

En su proceso de digitalización, las organizaciones necesitan tomar decisiones, el modelo propuesto facilita el consenso y la toma de decisiones equilibradas. En la aplicación realizada a agencias de viaje, la elección global tiene en cuenta las elecciones de los expertos individuales de las diferentes áreas del BSC para hacer un ranking global de preferencias. En la Tabla 12 vemos que las preferencias globales son para el 3D (AR y VR) (principalmente debido a las áreas de finanzas y aprendizaje y crecimiento del BSC) y el

código QR (principalmente debido a las áreas de finanzas y clientes del BSC). El modelo MCDM desarrollado puede aplicarse a otra selección de dispositivos sobre los que los expertos realizarán nuevas evaluaciones. En el caso de las agencias de viajes una mejora en la selección de dispositivos podría ser avanzar en experiencias digitales inmersivas, evolucionando desde el 3D (AR y VR) hacia el metaverso (Buhalis et al, 2023).

Se ha realizado una aproximación desde una perspectiva teórica. El valor añadido de este caso de uso reside en que extiende la generalización del proceso de toma de decisiones al contexto digital de los dispositivos en tienda, centrándose en el caso concreto del uso de la digitalización por parte de las agencias de viajes, aunque un marco similar puede aplicarse a otras tiendas físicas de la cadena de valor del sector turístico en las que se requiera una toma de decisiones compleja.

El uso de las áreas del BSC ayuda a tener una visión completa de la empresa y a que las principales perspectivas participen en la toma de decisiones. Una decisión importante son los pesos asignados a cada una de las áreas del BSC, esta decisión la toma el CEO. Una posible mejora y trabajo futuro para el modelo MCDM propuesto podría ser tener la posibilidad de variar estos pesos para trabajar con diferentes escenarios posibles y por lo tanto diferentes resultados finales, es decir, en un escenario podríamos priorizar el área de clientes, pero en otro podríamos priorizar finanzas. Para realizar este trabajo con los pesos de las áreas del BSC proponemos utilizar los operadores OWA (Ordered Weighted Averaging), introducidos por Yager (1988). Se ha propuesto en la literatura, sustituir la media ponderada del proceso de agregación AHP utilizando operadores OWA (Borouhaki & Malczewski, 2008; Zabihi et al., 2019; Llorens et al., 2022). Esta extensión facilita a los responsables de la toma de decisiones la expresión del número de subcriterios que deben satisfacerse mediante el uso de conceptos de alto nivel, como los cuantificadores lingüísticos. El uso de esta extensión implica, por tanto, incluir un cuantificador lingüístico en cada uno de los procesos de agregación de la jerarquía.

7. Caso de uso: plataforma de datos e inteligencia artificial para el posicionamiento de servicios turísticos inteligentes. Aplicación a las oficinas de información turística

Una de las claves de la competitividad de un destino turístico es la calidad de sus servicios turísticos y uno de los aspectos claves de estos servicios es su posicionamiento de cara a ofrecer un servicio de cercanía al turista. En los destinos turísticos inteligentes, la gestión y explotación del dato es un aspecto clave en la toma de decisiones y debe ser utilizado para mejorar los servicios diseñando servicios turísticos inteligentes (STS, por sus siglas en inglés *Smart Tourism Services*).

En este caso de uso se propone una plataforma de datos e inteligencia artificial sobre el que se desarrolla un modelo de decisión basado en una metodología descubrimiento de conocimiento en bases de datos (KDD, por sus siglas en inglés Knowledge Discovery in Databases) para el posicionamiento óptimo de STS. Además de posicionar de forma óptima los servicios turísticos, esta plataforma facilita un entorno interactivo para la interpretación de resultados. De igual modo, este entorno permite personalizar los servicios a su entorno cercano y propone recomendaciones. Se desarrolla una aplicación concreta de posicionamiento de oficinas de información turística para la ciudad de Madrid. Se proponen nuevas ubicaciones de oficinas de turismo para la ciudad de Madrid con recomendaciones personalizadas sobre la gestión de las mismas y se comprueba que los resultados del modelo reflejan correctamente los cambios que se han producido en los últimos años en este destino turístico.

En la Figura 47 podemos observar cómo se integra este caso de uso dentro del marco conceptual propuesto. Tal y como se explicará en el desarrollo del mismo, en lo relativo a fuentes de datos se trabajará con datos en abierto. Estos datos se ingestarán en el área de almacenamiento y gestión de los datos y se realizará sobre los mismos una limpieza, depuración, homogeneización y elaboración. Posteriormente se aplicarán diferentes técnicas de análisis de datos, tales como la segmentación y se generará un entorno interactivo y de recomendación en el que se empleará inteligencia artificial generativa para realizar las recomendaciones. Los resultados de la segmentación se usarán en los modelos de inteligencia generativa y los resultados de ambos se presentarán en un cuadro de mando interactivo. Las aplicaciones se centrarán en el posicionamiento de STS, lo cual tendrá aplicaciones para la gestión de destinos turísticos. En este caso de uso partiendo

del modelo conceptual propuesto se planteará una plataforma de datos e inteligencia artificial adaptada para la problemática concreta del caso.

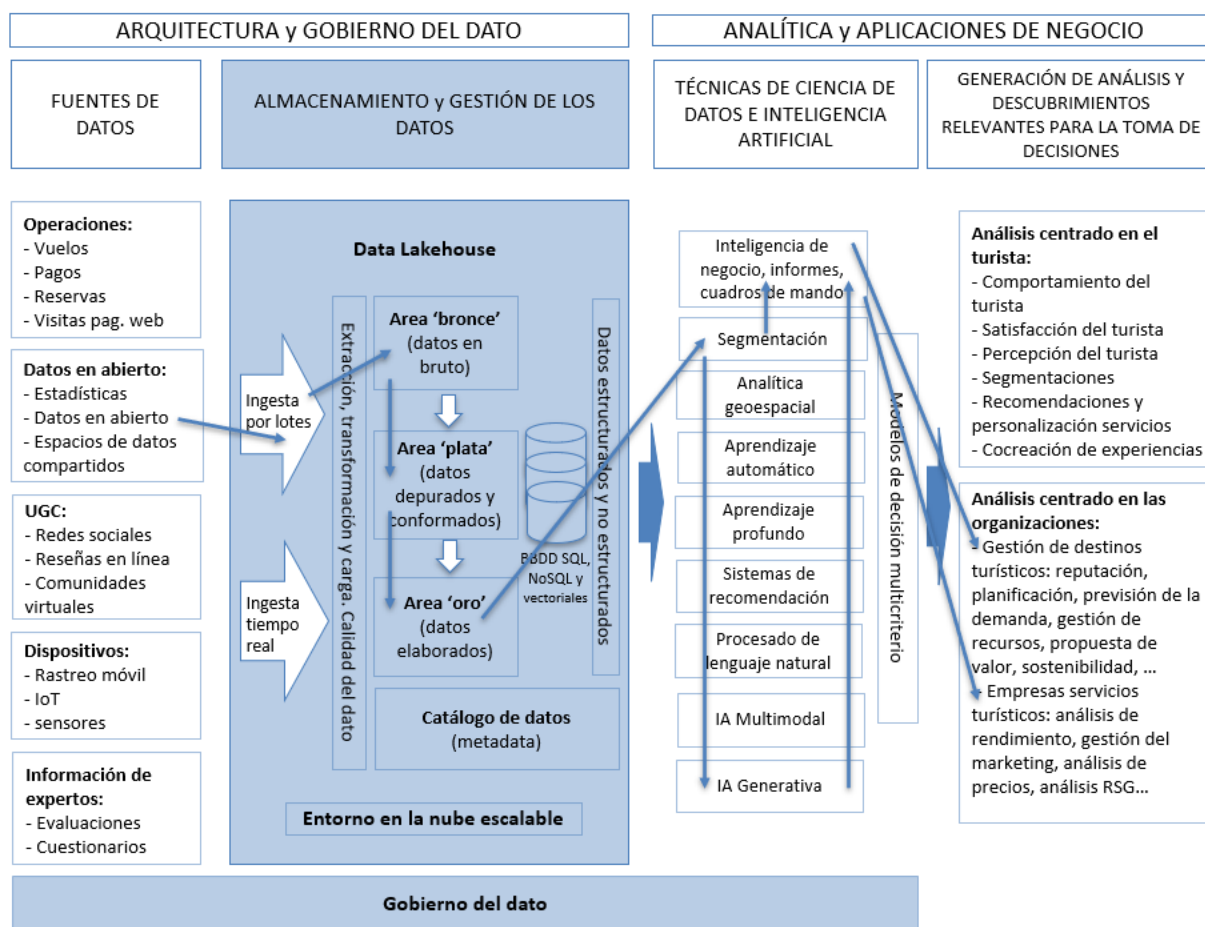


Figura 47: Integración del caso de uso 'Plataforma de datos e inteligencia artificial para el posicionamiento de STS. Aplicación a las oficinas de información turística' dentro del marco conceptual propuesto. Fuente: elaboración propia

7.1. Introducción

Uno de los problemas de negocio comunes en el sector del turismo es la ubicación de servicios turísticos. Desde la perspectiva de un destino turístico, la correcta ubicación de los servicios turísticos es una cuestión clave para su competitividad. Este posicionamiento de servicios afecta a los diferentes agentes de la cadena de valor del turismo tanto del sector privado como público. Tal y como se detallará en el apartado 7.2, el impacto de su posicionamiento ha sido analizado desde las perspectivas tanto de servicios públicos (Inoue & Yamamoto, 2013) como agentes del sector privado tales como los hoteles o restaurantes (Chen & Yan, 2016; Bilen et al., 2018). De igual modo, los destinos turísticos inteligentes (Shafiee et al., 2022) tratan de ofrecer servicios innovadores al turista basándose en datos y tecnología (Ammirato et al., 2018) para construir los STS. El diseño

de los STS debe priorizar la experiencia del turista y debe basarse en el uso de datos e inteligencia artificial (Hu & Li, 2023). Consideramos que el correcto posicionamiento de los STS mejora notablemente la experiencia del turista.

Para abordar la cuestión del posicionamiento de los servicios turísticos vamos a plantear una estrategia orientada al dato en la que la toma de decisión sobre el posicionamiento de los servicios turísticos sea guiada por los datos. Esta estrategia orientada al dato se va a sustentar en una plataforma de datos e inteligencia artificial, sobre la cual vamos a crear un modelo de decisión basado en ciencia de datos para posicionamiento óptimo de los STS. La plataforma propuesta toma como base inicial el marco conceptual para la toma de decisiones basadas en datos de una organización turística propuesto en esta tesis. La plataforma se organiza en tres capas y considera las áreas planteadas en el marco conceptual propuesto en el apartado 4. A continuación, se explican las funcionalidades principales de las tres capas que componen la plataforma, los detalles concretos de los componentes de cada capa serán desarrollados posteriormente.

- **Capa 1: Gestión y almacenamiento de fuentes de datos.** Dentro de la plataforma esta capa es la encargada de recopilar, validar, gestionar, almacenar, organizar y gobernar los datos necesarios para la toma de decisiones, en nuestro caso: cartografía de la zona geográfica, ubicación actual de los servicios turísticos, localización del público objetivo y punto de interés.
- **Capa 2: Técnicas de ciencia de datos e inteligencia artificial.** En esta capa se construyen los modelos de aprendizaje automático necesarios para la generación del conocimiento que permitan la toma de decisiones. En nuestro caso clustering jerárquico basado en distancias. Estos modelos pueden ser costosos computacionalmente y por ello son ejecutados en procesos planificados por lotes (usualmente denominados por su término en inglés *procesos batch*).
- **Capa 3: Generación de análisis y descubrimientos relevantes para la toma de decisiones.** Esta capa se centra en las aplicaciones de negocio y se realiza la interpretación del conocimiento previamente generado empleando para ello un entorno interactivo. Este entorno facilita la visualización de resultados, realizando consultas interactivas y simulaciones. Igualmente permite obtener recomendaciones empleando para ello servicios basados en modelos LLM. Este entorno interactivo se integra en un cuadro de mando.

En la Figura 48 se muestra el detalle de la plataforma propuesta a la que denominaremos Plataforma de datos e inteligencia artificial para el posicionamiento de STS.

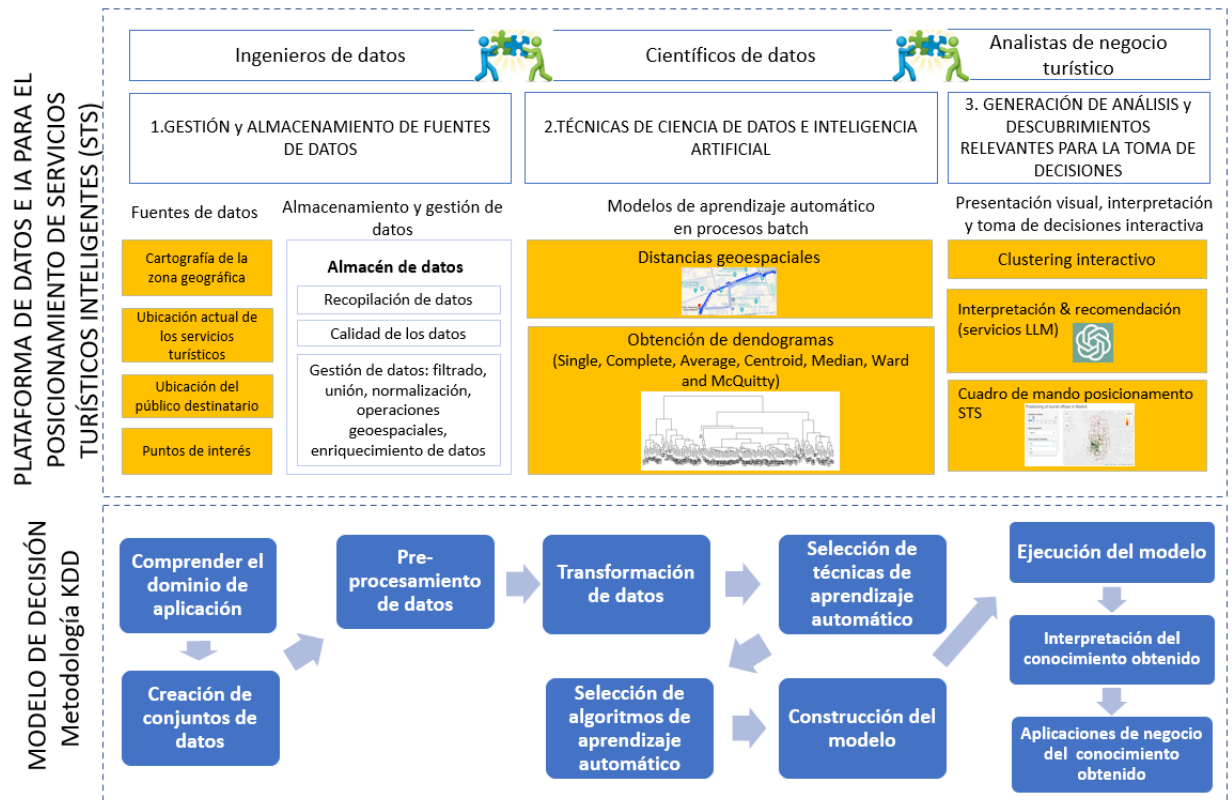


Figura 48: Plataforma de datos e inteligencia artificial para el posicionamiento de STS. Fuente: elaboración propia

Tal y como se ha mencionado anteriormente, las capas de la plataforma se basan en las áreas del marco conceptual propuesto en el apartado 4. Para adaptar dicho marco conceptual a una plataforma específica y centrada en el caso de uso, se consideran las siguientes cuestiones en el diseño de las capas de la plataforma:

- La capa 1 de la plataforma aglutina las áreas de fuentes de datos y almacenamiento y gestión del dato del marco conceptual. Consideramos que no existe una complejidad alta en las actividades de gestión y almacenamiento del dato que justifique el uso de dos áreas.
- En la capa 3 se incorporan técnicas de ciencia de datos, como los modelos LLM y los cuadros de mando interactivos que, considerando la organización del marco conceptual, también podrían haberse ubicado en la capa 2. Sin embargo, dado que son utilizadas por los analistas de negocio para interpretar los resultados y tomar decisiones, consideramos más adecuado integrarlas en esta capa.

Empleando esta plataforma se propone desarrollar un modelo de decisión basado en ciencia de datos para el posicionamiento óptimo de servicios turísticos empleando la

metodología KDD. Tal y como se muestra en la Figura 48, las diferentes fases del modelo quedan integradas en las capas de la plataforma.

En la plataforma se identifican los perfiles claves en cada una de las capas. Las capas 1 ó 2 precisan de perfiles técnicos especializados en datos como los ingenieros de datos o científicos de datos para todo el trabajo de gestión y análisis de datos, mientras que en la capa 3 participan analistas especializados en el negocio del turismo para encontrar las aplicaciones del conocimiento obtenido en la capa 2.

Nuestra propuesta se aplica a la ubicación óptima de oficinas de turismo en la ciudad de Madrid. Las oficinas de turismo son un servicio turístico esencial dentro de un destino turístico y un contacto directo con el turista que precisa proximidad o cercanía geográfica. Consideramos que la proximidad o cercanía geográfica entre las oficinas de turismo y el turista redundará en un mejor servicio, donde la cercanía genere confianza, un servicio más personalizado y en definitiva una mejor experiencia del turista. Se ha seleccionado la ciudad de Madrid ya que consideramos que se trata de un buen ejemplo de destino urbano europeo en el que se están produciendo cambios significativos tanto en la planificación urbanística como en su gestión turística orientándose hacia el concepto de destino turístico inteligente (Curiel et al., 2017). Madrid dispone de un importante patrimonio artístico y cultural concretamente la Comunidad de Madrid dispone de cinco enclaves declarados Patrimonio Mundial por la UNESCO (Martínez & Devesa, 2013). Por otra parte, se trata de un referente mundial en turismo de negocios y congreso (World Travel Awards, 2023) y de una ciudad con un gran dinamismo y vitalidad que dispone de una oferta muy amplia de ocio, cultura y gastronomía en constante renovación y crecimiento. El turismo en la comunidad de Madrid se encuentra en proceso de crecimiento y recibió en 2022 más de 6 millones de turistas extranjeros según el Instituto Nacional de Estadística (INE, 2023). Por otra parte, se trata de un destino con cambios importantes desde el punto de vista de planes urbanísticos y gestión turística. En lo relativo a planes urbanísticos cabe destacar el proyecto 'Madrid nuevo Norte' (Metaxas et al., 2021) y en lo relativo a los cambios en la gestión como destino turístico, cabe destacar la inversión en los diferentes agentes de la oferta turística (ocio, compras, gastronomía, hoteles, etc.) para atraer turismo de alto poder adquisitivo.

El resto del caso de uso se estructura como sigue. En el apartado 7.2 se estudian los trabajos relacionados explicando la originalidad de nuestra propuesta. En el apartado 7.3 se propone el modelo de decisión usando las diferentes etapas de KDD integradas en la plataforma propuesta y se realiza su aplicación al caso de oficinas de turismo. En el

apartado 7.4 se muestran los resultados obtenidos que se corresponde con la última fase del modelo de decisión mostrando las aplicaciones del conocimiento generado. Finalmente, en el apartado 7.5 se establecen las conclusiones y trabajo futuro.

7.2. Revisión de literatura

En este apartado se analizan diversos trabajos relacionados con la problemática planteada en este caso de uso y posteriormente se comparan con la solución planteada en este caso de uso con objeto de demostrar la novedad y los aportes de mismo.

Encontramos en la literatura diferentes investigaciones que han trabajado aspectos similares al objetivo de nuestro caso de uso. En primer lugar, cabe destacar el trabajo de Arana et al. (2016), este artículo evalúa las preferencias de los visitantes por diferentes diseños de servicios en las oficinas de información turística evaluando entre otros atributos la importancia de la ubicación de la oficina de turismo. En lo relativo al posicionamiento de los servicios turísticos encontramos el artículo de Inoue & Yamamoto (2013) que propone un enfoque basado en algoritmos genéticos y sistemas de información geográfica para evaluar la idoneidad de la ubicación de diferentes instalaciones públicas relacionadas con el turismo, entre ellos los centros de información turística. Encontramos diferentes estudios en los que se ha realizado un agrupamiento espacial de hoteles. Las técnicas de clustering y los objetivos de estos estudios son diversos entre ellos, cabe destacar: el análisis de la relación entre el posicionamiento con el precio (Chen & Yan, 2016; Jinwon et al., 2020), la relación entre el posicionamiento y la reputación online (Al-Saad et al., 2023), la relación entre el posicionamiento y la tasa de ocupación (Ghozali & Syafitri, 2020). Todos estos estudios realizan agrupamiento espacial de hoteles, pero con técnicas diferentes a las utilizadas en nuestro caso de uso y con un objetivo claramente diferente. De igual modo, encontramos estudios encaminados a identificar puntos de afluencia turística (Wu et al., 2018; Chang, 2020; Park & Chang, 2021) que podrían estar relacionados con el posicionamiento de servicios turísticos ya que identifican zonas con alta densidad de turistas, pero no lo abordan desde esa perspectiva.

Proponemos un método novedoso para posicionar servicios turísticos basado en técnicas de clustering espacial. En nuestra aplicación concreta del caso de uso, se realizará un agrupamiento espacial de hoteles para posteriormente ubicar las oficinas de turismo en la posición óptima en estos grupos. Los puntos de concentración turística se obtendrán realizando un clustering geográfico de los hoteles y, posteriormente, se obtendrá la situación óptima de las oficinas de turismo nuevas basada en esos grupos. Por otro lado, y a diferencia del resto de trabajos mencionados, nuestro caso de uso añade un entorno

interactivo para la interpretación de los resultados obtenidos y la obtención de recomendaciones. A continuación, en la Tabla 13 mostramos toda la relación de trabajos relacionados y la novedad de nuestra propuesta. Realizamos esta comparativa en base a las diferentes capas de la plataforma de datos e inteligencia artificial para el posicionamiento de STS, analizando como resuelven los diferentes artículos las funcionalidades de cada capa y mostrando al final nuestra propuesta con la novedad aportada. Tras la revisión bibliográfica, podemos decir que, hasta donde sabemos, nuestro caso de uso es el primer estudio que propone una plataforma con el alcance explicado.

Estudio	Principales resultados de la investigación	1) Gestión y almacenamiento de fuentes de datos	2) Técnicas de ciencia de datos e inteligencia artificial	3) Generación de análisis y descubrimientos relevantes para la toma de decisiones
Arana et al. (2016)	Diseño de oficinas de turismo.	Datos basados en encuestas.	Experimento de elección discreta.	Presentación de resultados en tablas.
Inoue & Yamamoto (2013)	Ubicación de instalaciones públicas relacionadas con el turismo.	Mapa digital con datos de infraestructura (distrito central de la ciudad de Yokohama), base de datos de ubicación de instalaciones relacionadas con el turismo, base de datos de excursiones.	Algoritmos genéticos y Sistemas de información geográficos.	Visualización de resultados en mapas geográficos.
Chen & Yan (2013)	Localización óptima de hoteles.	Datos de localización y precios de hoteles en Pekín.	Algoritmo de clustering k-medias.	Gráficos de dispersión.
Jinwon et al. (2020)	Variación espacial de la relación entre el precio y los atributos de un hotel.	Datos de localización, atributos y precios de hoteles en Chicago.	Modelo de precios hedónicos (s-HPM, Hedonic Price Model).	Presentación de resultados en tablas y mapas geográficos.
Al-Saad et al. (2023)	Agrupación espacial de hoteles y restaurantes con mesas al aire libre en base a su reputación on-line.	Datos UGC de Tripadvisor.	Método de agrupación espacial basado en la densidad jerárquica (HDBSCAN, Hierarchical Density-Based Spatial Clustering).	Presentación de resultados en tablas y mapas geográficos.
Ghozali & Syafitri (2020)	Identificar los patrones de distribución de los hoteles y realizar un análisis de agrupación de la ubicación de los hoteles basado en la tasa de ocupación.	Datos espaciales: ubicación de los hoteles, límites de las ciudades y subdistritos y mapas de la red de carreteras de la ciudad de Balikpapan. Datos no espaciales: clasificación de los hoteles, número de	Método de análisis del vecino más próximo (ANN, Average Nearest Neighbor) y de los puntos de alta densidad de turistas (conocidos comúnmente como <i>tourist hotspots</i>).	Presentación de resultados en tablas y mapas geográficos.

		habitaciones y porcentaje de ocupación mensual.		
Wu et al. (2018)	Construcción red de puntos de interés turístico.	Fotos geoetiquetadas tomadas en Pekín (China) desde Flickr.	Agrupación por búsqueda rápida y hallazgo de picos de densidad (CFSFDP, Clustering by Fast Search and Find of Density Peaks).	Visualización de resultados en mapas geográficos.
Chang (2020)	Identificación 'tourist hotspots'.	Datos sobre la ciudad de Jeonju en Corea del Sur: población, transportes, información geográfica, visitantes, atracciones turísticas, negocios locales.	Análisis de componentes principales (PCA, Principal Component Analysis), bosque aleatorio, k-medias clustering, e incrustación de vecinos estocásticos en algoritmos t (t-SNE, T-distributed Stochastic Neighbour Embedding) junto con dos técnicas de optimización: escalada simple Hill y algoritmo genético. (GA, Genetic Algorithm).	Visualización de resultados en mapas geográficos.
Park & Chang (2021)	Identificación 'tourist hotspots'.	Datos sobre la ciudad de Jeonju en Corea del Sur: población, transportes, información geográfica, visitantes, atracciones turísticas, negocios locales.	Análisis de componentes principales, k-medias clustering, bosque aleatorio, algoritmo de aproximación y proyección uniforme de colectores (UMAP, Uniform manifold approximation and projection).	Visualización de resultados en mapas geográficos.
Nuestra propuesta	Plataforma de datos e inteligencia artificial para el posicionamiento de STS que integra un modelo de decisión. Aplicación: oficinas de información turística en Madrid.	Datos geográficos de hoteles, oficinas de información turística, puntos de interés y de delimitación territorial (barrios).	Obtención de distancias geoespaciales reales. Clustering jerárquico en base a diversas técnicas (single, complete, average, centroid, median, Ward and mcquitty).	Entorno interactivo que facilita la Visualización de resultados, realizando consultas interactivas que facilitan su interpretación. Igualmente permite obtener recomendaciones empleando para ello servicios basados LLM.

Tabla 13: Trabajos relacionados con nuestra propuesta y diferenciación en base a las diferentes capas de la plataforma propuesta.

7.3. Modelo propuesto

Como se indicó previamente, en la plataforma de datos e inteligencia artificial para el posicionamiento de STS se implementa un modelo de decisión. Los pasos del modelo que proponemos se van a realizar siguiendo la metodología KDD. La metodología KDD fue presentada por Fayyad et al. (1996) y está inspirada en un trabajo previo de Brachman & Anand (1994). Se trata del primer intento serio de sistematizar el proceso conocido hoy día como ciencia de datos.

Tal y como se ha indicado en el apartado 7.1, el caso de uso se aplica a la ciudad de Madrid. Al tratarse de un destino turístico en proceso de cambio y crecimiento creemos que puede ser de interés analizar la ubicación óptima de las oficinas de turismo que se adapten adecuadamente a estos cambios.

7.3.1. Comprender el dominio de aplicación

El primer paso en la mencionada metodología KDD es el entendimiento del dominio de la aplicación. Para ello debemos identificar claramente los objetivos de nuestro análisis y la información relevante disponible. Tal y como se ha indicado en el apartado 7.1, el objetivo principal de este caso de uso es crear un modelo que busque la ubicación óptima de estos servicios en un destino turístico determinado. Este posicionamiento se va a realizar centrándose en ofrecer un servicio cercano al turista y, por ello, es esencial definir y localizar bien el conjunto de turistas al que nos dirigimos, el cual que será nuestro público objetivo. En nuestra aplicación del caso de uso al posicionamiento de oficinas de turismo, la forma de identificar la localización de este público objetivo la vemos a realizar basándonos en la ubicación de los hoteles de una zona geográfica concreta.

En primer lugar, delimitaremos una zona geográfica que concentre los atractivos turísticos de un destino determinado y en segundo lugar obtendremos los datos de los hoteles ubicados en esa zona. Se posicionarán las oficinas turísticas de forma que estén cerca del mayor número posible de hoteles, lo cual facilite un servicio de proximidad. Será necesario igualmente disponer de la información de las oficinas turísticas ya existentes en el destino de forma que propongamos nuevas oficinas en ubicaciones que complementen a las anteriores. Este posicionamiento, que identificamos como objetivo

principal del modelo, va a depender del número de nuevas oficinas que se deseen ubicar, este número será algo flexible en nuestro modelo de forma que dependiendo del número de nuevas oficinas elegido el resultado podrá cambiar. Dar flexibilidad al modelo es interesante tanto en el número de oficinas como en otras variables que consideremos relevantes en el desarrollo del modelo. Identificamos, por tanto, también como objetivo del modelo el dotarlo de flexibilidad e interactividad para realizar simulaciones.

Otra información relevante en nuestro modelo y que será complementaria a la anterior es la información relativa a las divisiones territoriales a efectos de gestión administrativa de la zona geográfica seleccionada, lo cual nos permite mejor conocer las diferentes áreas locales dentro de la zona geográfica seleccionada. Así mismo, nos permitirá conocer mejor la zona geográfica seleccionada el disponer de la información de los puntos de interés turístico ubicados en la misma. Toda la información relevante identificada puede clasificarse dentro de los distintos tipos de fuentes de datos definidas en la plataforma de datos e inteligencia artificial para el posicionamiento de STS en la capa 1 (véase Figura 48) tal y como se observa en la Tabla 14.

Fuentes de datos (Capa 1)	Información relevante	Caso de aplicación (Madrid)
Cartografía de la zona geográfica	Información cartográfica de la zona geográfica de estudio identificada.	Divisiones territoriales área geográfica de estudio identificada en la ciudad de Madrid.
Ubicación actual de los servicios turísticos	Posición de los servicios turísticos actuales en la zona geográfica de estudio identificada.	Posición de las oficinas de información turística en el área geográfica de estudio identificada en la ciudad de Madrid.
Ubicación del público destinatario	Localización del público objetivo en la zona geográfica de estudio identificada.	Posicionamiento de los hoteles en área geográfica de estudio identificada en la ciudad de Madrid.
Puntos de interés	Puntos de interés de la zona geográfica de estudio identificada.	Puntos de interés en la zona delimitada en el área geográfica de estudio identificada en la ciudad de Madrid.

Tabla 14: Información relevante identificada clasificada en los distintos tipos de fuentes de datos de la plataforma propuesta.

Dentro de la información relevante identificada la ubicación de los hoteles va a jugar un papel fundamental, pues representa el posicionamiento de los turistas y también refleja la dinámica de cambio de un destino turístico debido a planes urbanísticos o a la aparición de nuevas atracciones turísticas en el mismo. Es decir, el

análisis a realizar se basa en la ubicación de los hoteles en un momento determinado, pero si este análisis se realiza después de un periodo de tiempo, lo esperable es que hayan aparecido nuevos hoteles y otros habrán cambiado de ubicación o desaparecido representando la dinámica de cambios de un destino turístico. Identificamos también como objetivo del modelo el que se capaz de reflejar correctamente los cambios que experimenta un destino turístico en el posicionamiento de las nuevas oficinas. Por último, consideramos que toda la información relevante identificada debe también ayudar a personalizar estas oficinas en base a las necesidades de su entorno geográfico próximo, lo cual será también otro objetivo interesante. Para conseguir este objetivo de personalización sería interesante que el propio modelo genere recomendaciones en base a los datos de entorno geográfico próximo a las oficinas identificadas. Es decir, pretendemos un modelo flexible con interactividad para facilitar la interpretación de resultados, realizar simulaciones y que genere recomendaciones.

A modo de resumen, establecemos como objetivo principal del modelo el posicionar de forma óptima las nuevas oficinas de turismo, además hemos identificado otros objetivos adicionales como personalizar las mismas a su entorno cercano y que el modelo refleje correctamente los cambios que se producen en un destino turístico. Así mismo, nos parece interesante poder trabajar en un entorno interactivo para la interpretación de resultados que permita realizar simulaciones y que genere recomendaciones sobre las oficinas. La información relevante identificada incluye la ubicación de los hoteles, las oficinas de turismo ya existentes, los puntos de interés turístico y la división territorial administrativa del destino turístico.

7.3.2. Creación del conjunto de datos

El segundo paso de la metodología KDD es la creación del conjunto de datos. Una vez identificada la información relevante, en esta etapa obtenemos todos los conjuntos de datos necesarios: datos de hoteles con su posición geográfica, oficinas de información turística ya existentes con su posición geográfica, delimitación territorial administrativa de la zona geográfica y puntos de interés turístico también con su posición geográfica. Nos basamos en datos publicados en abierto, concretamente de la Comunidad y el Ayuntamiento de Madrid. Consideramos que el avance de este tipo de iniciativas de datos en abierto permite la aplicación de nuestra plataforma y modelo a otros destinos turísticos. En la Tabla 15 se muestran los conjuntos de datos a utilizar y el volumen de datos de los mismos.

Conjunto de datos	Volumen de datos
Hoteles de Madrid (1 a 5 estrellas) geolocalizados.	5 ficheros con un total de 498 filas
Delimitación territorial	1 fichero con 128 filas
Oficinas de turismo de Madrid geolocalizadas	1 fichero con 56 filas
Puntos de interés turísticos:	
Teatros	1 fichero con 235 filas
Museos	1 fichero con 185 filas
Edificios de carácter monumental	1 fichero con 216 filas

Tabla 15: Conjuntos de datos a utilizar y volumen de datos.

Los datos fueron descargados a fecha 3 de marzo de 2023. Los 5 primeros conjuntos de datos proceden del portal Nomecalles (Nomecalles, 2023) del Sistema de Información Territorial del Instituto de Estadística de la Comunidad de Madrid y el último conjunto de datos procede del portal de datos abierto del Ayuntamiento de Madrid (Portal datos Ayuntamiento de Madrid, 2023).

Los datos geográficos descargados utilizan diferentes sistemas de coordenadas geográficas. Algunos de ellos trabajan con el sistema geodésico de coordenadas geográficas WGS84 y otros con sistema de coordenadas UTM. El sistema WGS84, utiliza la latitud y la longitud para la representación de ubicaciones en superficie terrestre que proporcionan información sobre la posición horizontal de un punto en la Tierra en relación con el ecuador (latitud) y el meridiano de Greenwich (longitud). WGS84 es ampliamente utilizado en sistemas de posicionamiento global (GPS, siglas en inglés de *Global Positioning System*) y en cartografía, y se considera un estándar global para la representación de coordenadas geográficas. El Sistema de Coordenadas UTM está basado en la proyección cartográfica de Mercator. Sus unidades son los metros. Las coordenadas UTM tienen múltiples ventajas; entre ellas, cabe destacar, que simplifica el cálculo de distancias dado que utiliza una proyección cartográfica específica que minimiza las deformaciones de la superficie terrestre, especialmente en áreas pequeñas. Vamos a emplear el sistema WGS84 para visualizar los puntos en aplicaciones que emplean cartografía mundial y el sistema UTM para el cálculo de distancias entre puntos cercanos.

Conviene señalar que existen otros conjuntos de datos que podrían resultar de interés, pero se han descartado al delimitar el ámbito de aplicación del caso de uso. Por ejemplo, complementarios a los datos de hoteles, podrían haber sido los datos de pensiones, albergues o incluso viviendas turísticas, pero nuestra aplicación se centra

únicamente en hoteles. Por otro lado, en lo relativo a los puntos de interés encontramos en Madrid un amplio abanico de opciones (restaurantes, tiendas, parques y jardines, etc.) complementarias a las identificadas, aunque consideramos que los seleccionados son muy representativos.

El lenguaje de programación empleado para realizar la lectura de estos ficheros y que vamos a utilizar en el resto de pasos del modelo es R (R Project, 2023), y los scripts del modelo se han realizado en el entorno de desarrollo RStudio (RStudio, 2023).

7.3.3. Preprocesamiento de los datos

En este paso, partiendo de la información obtenida se realizan todo el procesado necesario para poder trabajar con estos datos en el modelo.

En primer lugar, se realiza una revisión de la calidad del dato con el objeto de detectar posibles errores que pudieran impactar en nuestro análisis tales como valores ausentes o datos duplicados. En esta revisión sólo se detecta un único error en el conjunto de datos edificios de carácter monumental en el que se detecta una fila con nulos en los datos de coordenadas geográficas, esta fila es eliminada, quedando por tanto 215 filas en este conjunto de datos.

En segundo lugar, identificaremos una zona geográfica objetivo de nuestra aplicación del caso de uso. El objetivo es delimitar dentro del destino turístico a estudiar una zona de especial interés para los turistas y contenga los principales atractivos turísticos del destino. Una delimitación geográfica muy empleada en la ciudad de Madrid es la llamada 'almendra central', zona situada dentro de la carretera de circunvalación principal de Madrid (M-30) y que contiene los barrios más céntricos dentro de la ciudad y que mayor patrimonio histórico y cultural disponen. Son estos barrios de la almendra central los que mayor afluencia de turistas reciben. Nos quedamos con 43 barrios y los hoteles, las oficinas de turismo y los puntos de interés turístico que tienen (véase Tabla 16). Todas las operaciones de filtrado geoespacial se han realizado trabajando en el mismo sistema de referencia de coordenadas, concretamente en una proyección UTM, huso 30.

Debido a que el principal objetivo de esta aplicación es proponer nuevas oficinas turísticas cuya ubicación esté cerca del mayor número posible de hoteles y dado que ya existen oficinas turísticas en nuestra zona geográfica de estudio, el primer paso es determinar los hoteles que ya tienen una oficina de turismo cerca y eliminarlos del conjunto de datos de estudio. Tal y como se ha explicado en el apartado 4.3.4.1 del marco

teórico de investigación de esta memoria, en nuestra aplicación la distancia que mejor refleja la proximidad o cercanía geográfica que hay entre dos puntos de una ciudad es la distancia caminando a pie. Tomaremos como referencia el 'wd-5', introducido también en el marco teórico de investigación. Para obtener la distancia a pie entre dos puntos geográficos empleamos la interfaz de programación de aplicaciones (API, por sus siglas en inglés *Application Programming Interface*) de Openrouteservice (Openrouteservice, 2023) que permite obtener la distancia entre dos puntos indicando la forma en la que se recorre, en nuestro caso caminando a pie. Una vez realizado este proceso encontramos que 183 hoteles están dentro de la almendra central y no tienen ninguna oficina de turismo cerca. Este conjunto final de hoteles será el que empleemos para posicionar las nuevas oficinas de turismo de forma óptima buscando la mayor cercanía posible con estos hoteles.

En la Tabla 16 mostramos la composición del conjunto de datos final.

Conjunto de datos	Volumen de datos
Hoteles de Madrid (1 a 5 estrellas) geolocalizados.	183 hoteles
Delimitación territorial	43 barrios
Oficinas de turismo de Madrid geolocalizadas	6 oficinas de turismo
Puntos de interés turísticos:	
Teatros	76 teatros
Museos	63 museos
Edificios de carácter monumental	190 edificios de carácter monumental

Tabla 16: Conjuntos de datos con los datos seleccionados después del filtrado espacial por la zona geográfica de la almendra central y el filtrado de hoteles en base a la proximidad a las actuales oficinas de turismo.

En la Figura 49 se visualizan sobre un mapa de Madrid los datos seleccionados referentes a oficinas de turismo, hoteles y puntos de interés turístico, mostrando las divisiones territoriales en barrios de la zona geográfica seleccionada (almendra central).

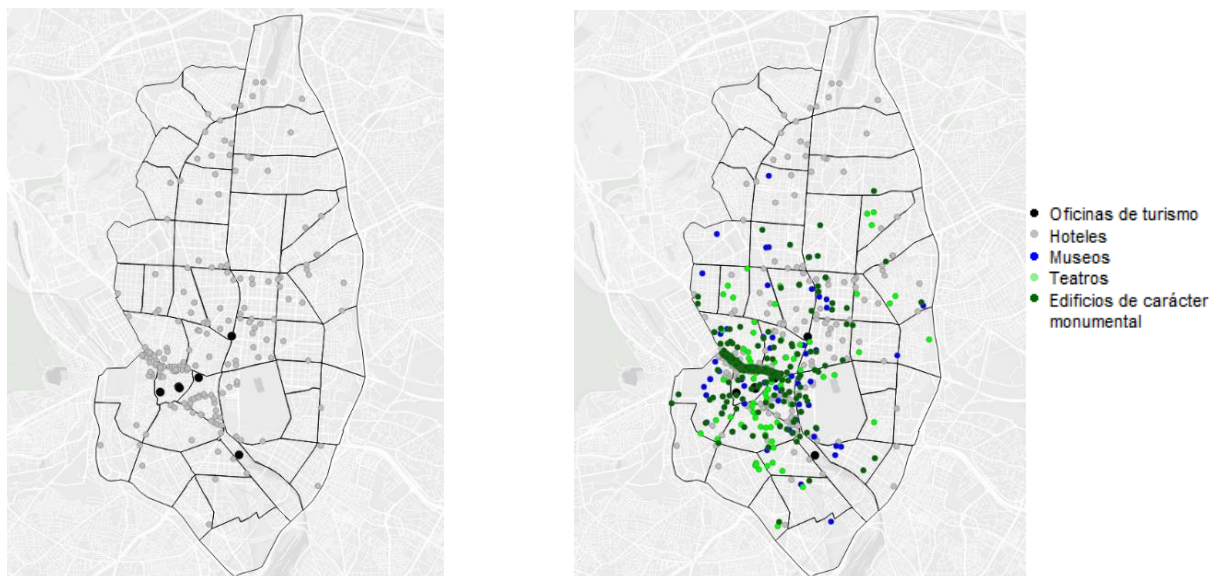


Figura 49: En la imagen de la izquierda se muestra una visualización sobre un mapa de Madrid en el que se resalta la zona geográfica seleccionada (almendra central) con las divisiones territoriales en barrios de los datos seleccionados referentes a oficinas de turismo (puntos negros) y hoteles (puntos grises). En la imagen de la derecha se añade a la anterior información los puntos de interés turístico: teatros (puntos verdes claro), museos (puntos azules) y edificios de carácter monumental (puntos verdes oscuro).

Tal y como se observa en la Figura 48, esta fase y las dos anteriores del modelo quedan integradas en la plataforma en la capa 1.

7.3.4. Transformación de los datos

En el cuarto paso realizamos las transformaciones necesarias de los datos. En nuestro caso al tratar con datos espaciales es necesario trabajar con toda la información espacial bajo el mismo sistema de coordenadas. Transformamos todos los conjuntos de datos al sistema geodésico de coordenadas geográficas WGS84 que será con el que trabajaremos en los siguientes pasos.

A continuación, vamos a proponer nuevas oficinas de turismo posicionándolas en base a la ubicación de los hoteles. Dentro del conjunto de los 183 hoteles, podremos encontrar subconjuntos de hoteles próximos entre sí, la identificación de estos subconjuntos permitirá posicionar una nueva oficina de turismo en una posición central dentro del subconjunto de hoteles, consiguiendo la proximidad de la oficina a los hoteles de ese subconjunto. Necesitamos, por tanto, conocer las distancias entre los hoteles, para posteriormente agruparlos. Calculamos la distancia entre los hoteles, que será una matriz 183×183 , que almacena la información de la distancia de cada hotel con los 182 restantes.

Desde el apartado 7.3.2 hasta el apartado 7.3.4 en las fases creación, preprocesamiento y transformación de los conjuntos de datos se han ido realizando diferentes operaciones con los conjuntos de datos obtenidos, en la Figura 50 mostramos un resumen de todas estas operaciones realizadas con el objeto de tener una visión completa de todo el ciclo de operaciones realizado previo a la construcción del modelo.



Figura 50: Operaciones realizadas en las fases creación, preprocesamiento y transformación de los conjuntos de datos. Fuente: elaboración propia.

7.3.5. Construcción del modelo

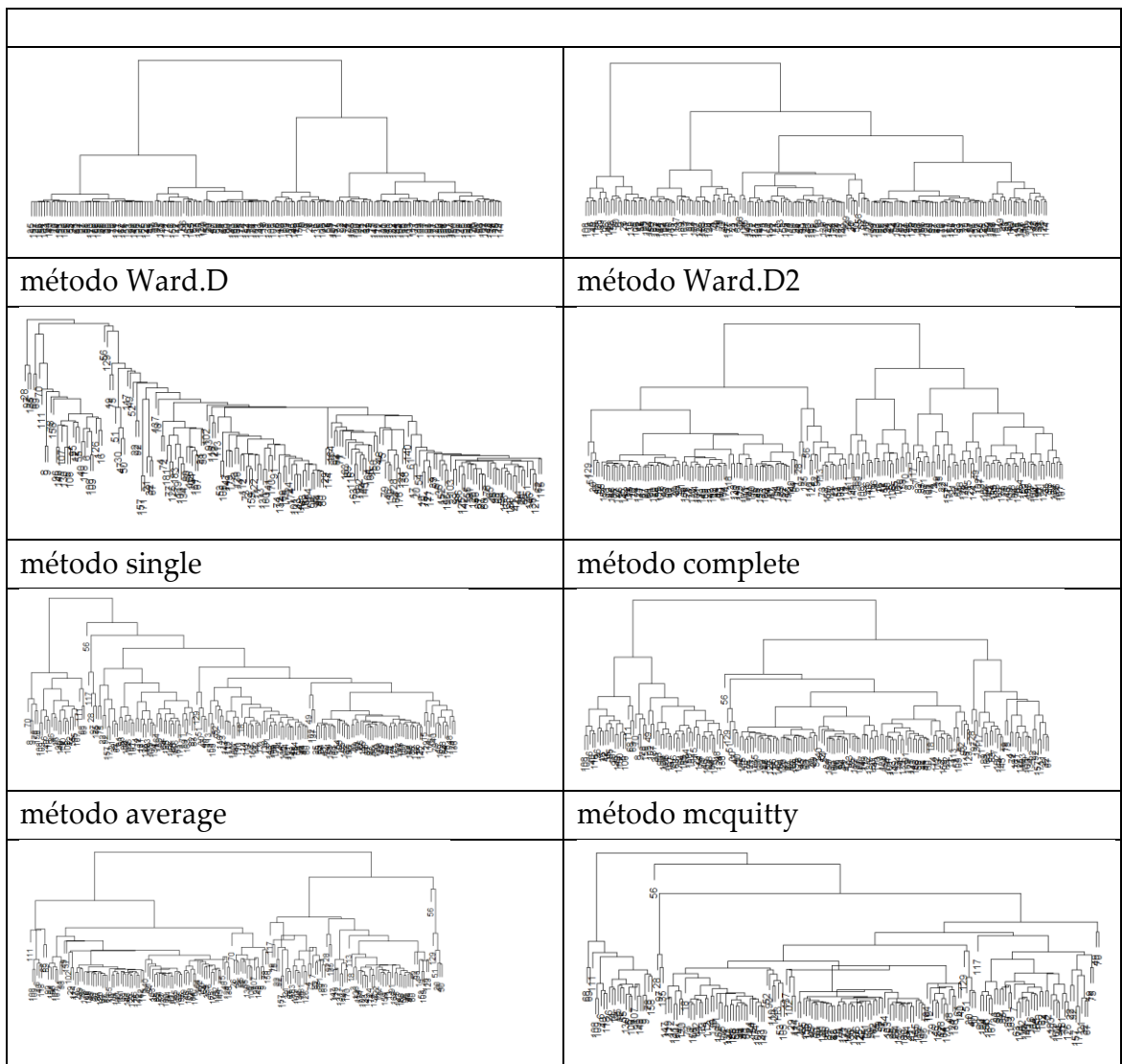
En este paso realizamos la construcción del modelo. Tal y como observamos en Figura 48 previo a la construcción del modelo están los pasos: 'Selección de técnicas de aprendizaje automático' y 'Selección de algoritmo de aprendizaje automático'. Ambos pasos ya han sido justificados en apartados anteriores. En primer lugar, tal y como se ha indicado en el apartado 7.2 vamos a realizar un agrupamiento espacial de hoteles para posteriormente ubicar las oficinas de turismo en la posición óptima en estos grupos. De tal forma que la técnica de aprendizaje automático seleccionada es el clustering. En segundo lugar, y con respecto a la selección del algoritmo de aprendizaje automático en el apartado 4.3.4 dentro del marco teórico de investigación de esta memoria se ha justificado la selección de técnicas de clustering y en particular el uso de algoritmos de clustering jerárquico por aglomeración utilizando diferentes métodos.

Uno de los objetivos, identificado en el apartado 7.3.1, es dotar al modelo de flexibilidad para realizar simulaciones en la fase de interpretación de resultados. Para ello consideramos útil que el método de clustering jerárquico sea un parámetro, así como el número de clústeres a obtener. Por un lado, el número de clústeres determinará el número de nuevas oficinas de turismo que queremos posicionar y por otro el método de clustering jerárquico por aglomeración determinará el modo en el que estos clústeres se obtienen.

Por lo tanto, vamos a realizar el clustering por aglomeración para todos los métodos mencionados anteriormente. Para implementarlo en R vamos a usar la función `hclust`

(del paquete Stats, cargado por defecto en R). Las particularidades de la implementación de estos métodos en la función hclust se pueden consultar en Hclust: Hierarchical Clustering (2023). Esta función nos permite seleccionar el método de aglomeración incluyendo todos los citados: Ward.D, Ward.D2, single, complete, average, mcquitty, median y centroid. El método Ward tiene dos variantes explicadas en el artículo de Murtagh & Legendre (2014).

El siguiente paso es la obtención de los dendogramas para cada uno de los métodos indicados. La obtención de estos dendogramas se basa en la matriz de distancias indicada en el apartado 7.3.4. Conviene señalar que tanto el cálculo de la matriz de distancias, como la obtención de los dendogramas son procesos computacionalmente costosos y que serán ejecutados en procesos batch para su uso posterior en el entorno interactivo. A continuación, en la Figura 51 mostramos el dendograma obtenido para cada uno de los métodos empleados.



método median	método centroid
---------------	-----------------

Figura 51: Dendograma obtenido para cada uno de los métodos empleados.

Los resultados quedan almacenados para su explotación en la siguiente fase. Para obtener un número dado de clústeres del dendograma realizamos un corte a la altura deseada (véase Figura 23). Posteriormente se utilizarán en la ejecución del modelo en el entorno interactivo.

Tal y como se observa en la Figura 48, esta fase y las tres anteriores del modelo quedan integradas en la plataforma propuesta en la capa 2.

7.3.6. Ejecución del modelo

La ejecución del modelo se integra dentro de la plataforma de datos e inteligencia artificial para el posicionamiento de STS en la capa 3. La ejecución del modelo queda determinada por dos parámetros que son el método de clustering jerárquico por aglomeración y el número de clústeres. Para realizar este proceso de la forma más eficiente y para cumplir el objetivo de flexibilidad que nos hemos planteado en el apartado 7.3.1 proponemos el uso de un entorno interactivo. En este entorno interactivo se podrán modificar los parámetros anteriormente indicados y nos permitirá visualizar los resultados obtenidos seleccionando los elementos a mostrar. La visualización mostrará la posición de las oficinas de turismo propuestas en base al número de ellas seleccionado y el método de clustering jerárquico. Esta visualización se realizará sobre un mapa en el que se situarán los hoteles seleccionados y las oficinas de turismo actuales, adicionalmente se podrán añadir los puntos de interés turístico para enriquecer esta visualización. A continuación, vamos a explicar las principales funcionalidades de este entorno interactivo que sobre el que se realiza la ejecución del modelo y posteriormente detallaremos los cálculos de esta ejecución.

Nos centramos primero en las opciones de visualización del entorno interactivo y posteriormente hacemos zoom sobre el mapa para analizar en detalle los distintos elementos mostrados en el mismo. En las Figura 52 y Figura 53 vemos una captura de pantalla del entorno interactivo en el que se pueden observar las selecciones que permite realizar.

Positioning of tourist offices in Madrid

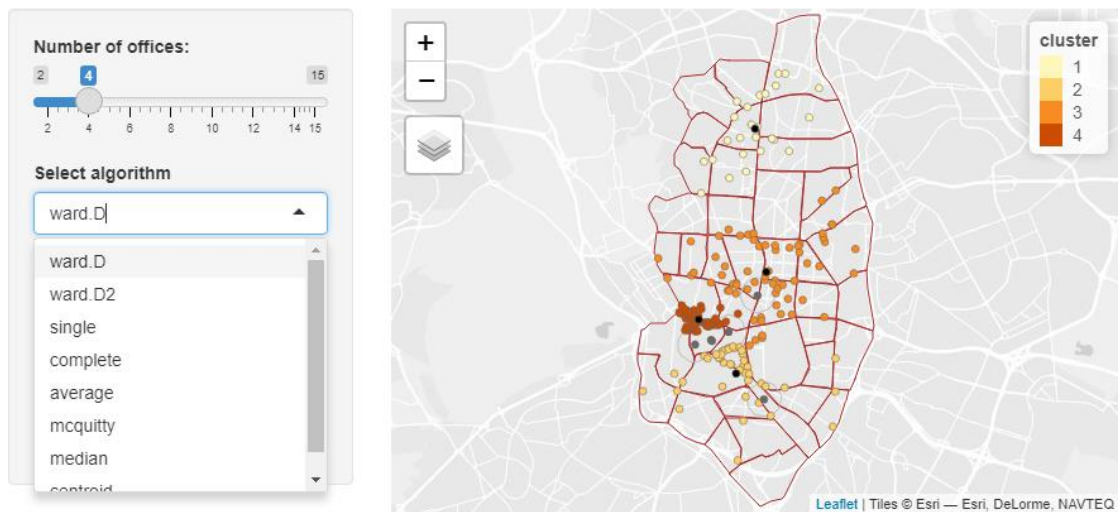


Figura 52: Entorno interactivo para visualizar los resultados del posicionamiento de las nuevas oficinas de turismo propuestas. Se muestra la selección del número de oficinas a proponer y el algoritmo a utilizar.

Positioning of tourist offices in Madrid

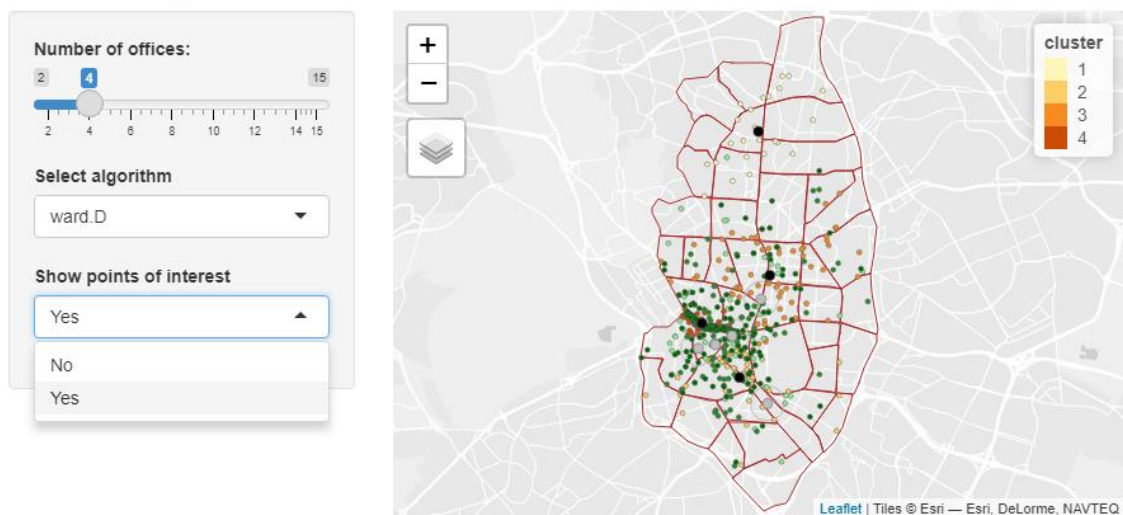


Figura 53: Entorno interactivo para visualizar los resultados del posicionamiento de las nuevas oficinas de turismo propuestas. Se muestra la selección del número de oficinas a proponer, el algoritmo a utilizar y si se desean visualizar puntos de interés turístico o no. Los puntos de interés turístico se muestran en tonalidades verdes.

Dentro del entorno interactivo también se puede seleccionar el mapa sobre el cual se muestran los elementos y se pueden seleccionar los diferentes elementos que se quieren visualizar sobre el mapa (véase Figura 54). Se puede seleccionar para mostrar u ocultar: las divisiones territoriales en barrios, la zona de influencia de cada oficina de turismo existente, los hoteles, las oficinas de turismo propuestas y los puntos de interés turístico.

Positioning of tourist offices in Madrid

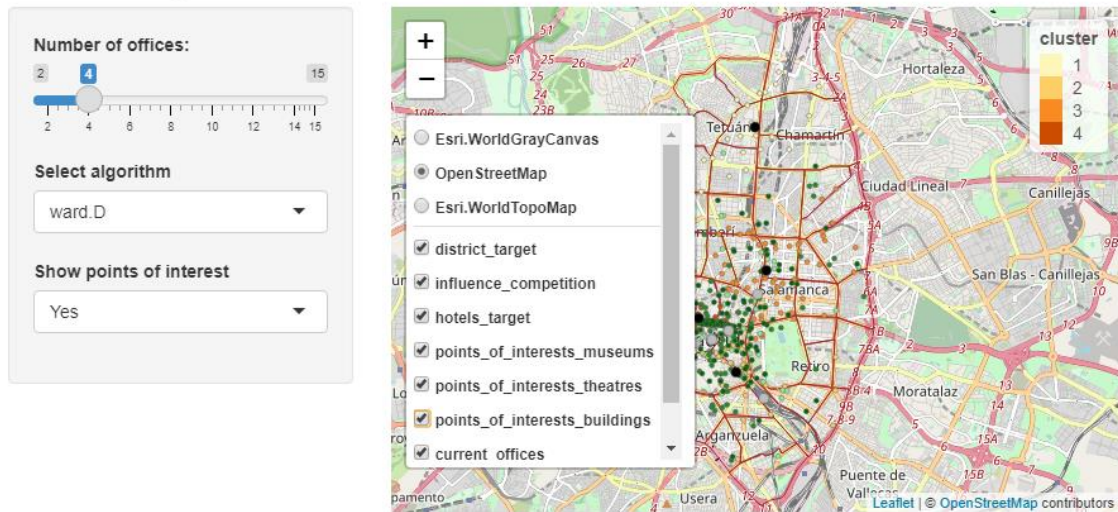


Figura 54: Entorno interactivo para visualizar los resultados del posicionamiento de las nuevas oficinas de turismo. Se muestra la selección de la mapa y los elementos a visualizar.

De igual modo, es posible obtener recomendaciones con ChatGPT3.5 una vez nos centramos en una oficina de turismo de las propuestas por el entorno, tal y como se puede observar en la Figura 55.

Positioning of tourist offices in Madrid

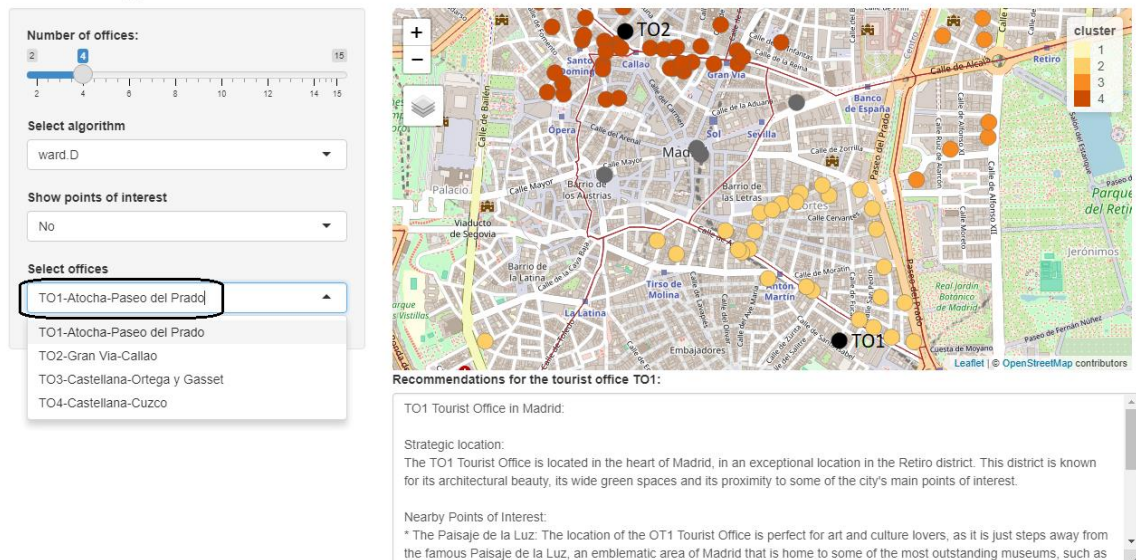


Figura 55: Recomendaciones generadas en el entorno interactivo para una oficina de turismo (TO1) de las propuestas por el entorno.

En la siguiente imagen hacemos zoom sobre la imagen del entorno interactivo para visualizar de forma más detallada los diferentes elementos mostrados. La selección que se realiza son 4 oficinas de turismo a proponer usando el algoritmo Ward.D. En la Figura 56 podemos ver la delimitación de la zona geográfica de la almendra central y su división en barrios en color rojo. Los puntos grises son las oficinas de turismo actuales y el círculo

gris alrededor es la zona de influencia de los mismos con un radio de 400 metros (wd-5). Conviene señalar que esta zona de influencia facilita la visualización, pero no es exacta ya que hemos empleado en el modelo es la distancia caminando a pie, no la euclídea. Los puntos negros muestran el posicionamiento de las 4 nuevas oficinas de turismo propuestas. Los hoteles están agrupados en 4 clústeres que se muestran en diferentes colores. Cada clúster tiene una oficina de turismo propuesta posicionada en un punto central del mismo.

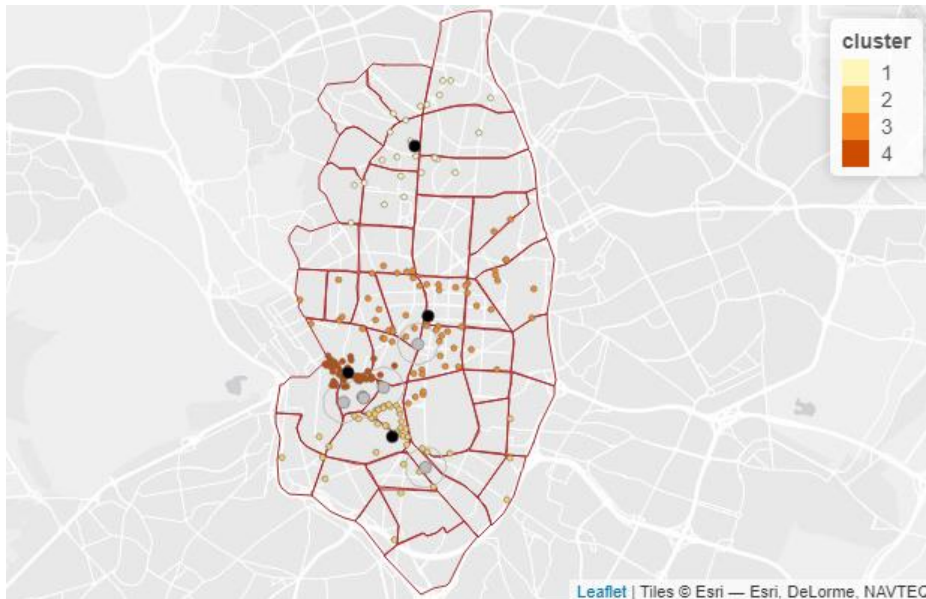


Figura 56: Posicionamiento de 4 nuevas oficinas de turismo empleando el algoritmo Ward.D.

Una vez explicado cómo funciona el entorno interactivo, pasamos a explicar cómo realiza los cálculos para obtener los datos a visualizar. En el paso de construcción del modelo se calculan los dendogramas para todos los métodos de clustering jerárquico considerados, en el entorno interactivo se selecciona un método concreto y un número de oficinas a proponer. El método seleccionado identifica el dendograma concreto con el que se va a trabajar y el número de oficinas el punto de corte a realizar en el dendograma. El resultado son los hoteles agrupados en base al número de clústeres indicado, de forma que podemos obtener el clúster al que pertenece cada hotel. Una vez realizada esta agrupación obtenemos el punto geográfico medio de cada clúster que es dónde se sitúa la oficina de turismo propuesta. Esta operación de obtener el punto medio es similar a la que realiza el algoritmo de clustering k-medias. Tenemos como resultado los hoteles agrupados en clústeres y el posicionamiento de las oficinas de turismo propuestas dentro de cada clúster. Las posiciones de las oficinas se tratan como puntos geográficos y se trabajan bajo la misma proyección del resto de información geográfica que es sistema geodésico de coordenadas geográficas WGS84.

La visualización está creada con el paquete Tmap de R (Tmap, 2023) que permite realizar una visualización interactiva pudiendo realizar acciones tales como el zoom o moverse sobre el mapa. La visualización se puede realizar sobre distintos mapas de base y permite dibujarlo en diferentes capas, de tal forma, que luego se puedan quitar o añadir esas capas de forma interactiva, tal y como se ha mencionado.

El entorno interactivo está realizado empleando el paquete de R Shiny (Shiny, 2023). Shiny es un paquete de R que facilita la creación de aplicaciones web interactivas. El entorno interactivo realiza algunos cálculos para obtener los datos a visualizar, pero conviene señalar que estos cálculos son ligeros y se realizan de forma muy rápida sin generar ninguna espera en la explotación del entorno interactivo. Es decir, podemos realizar cambios en las selecciones y el resultado se muestra en el mapa de forma inmediata. Esto es debido, tal y como se indicó en el apartado 7.3.5, a que los cálculos costosos son ejecutados previamente en un proceso batch. Este proceso batch incluye tanto el cálculo de la matriz de distancias como la obtención de los dendogramas. El entorno interactivo parte de estos dendogramas ya calculados, realizando las selecciones necesarias y generando la visualización tal y como se muestra en la Figura 57.

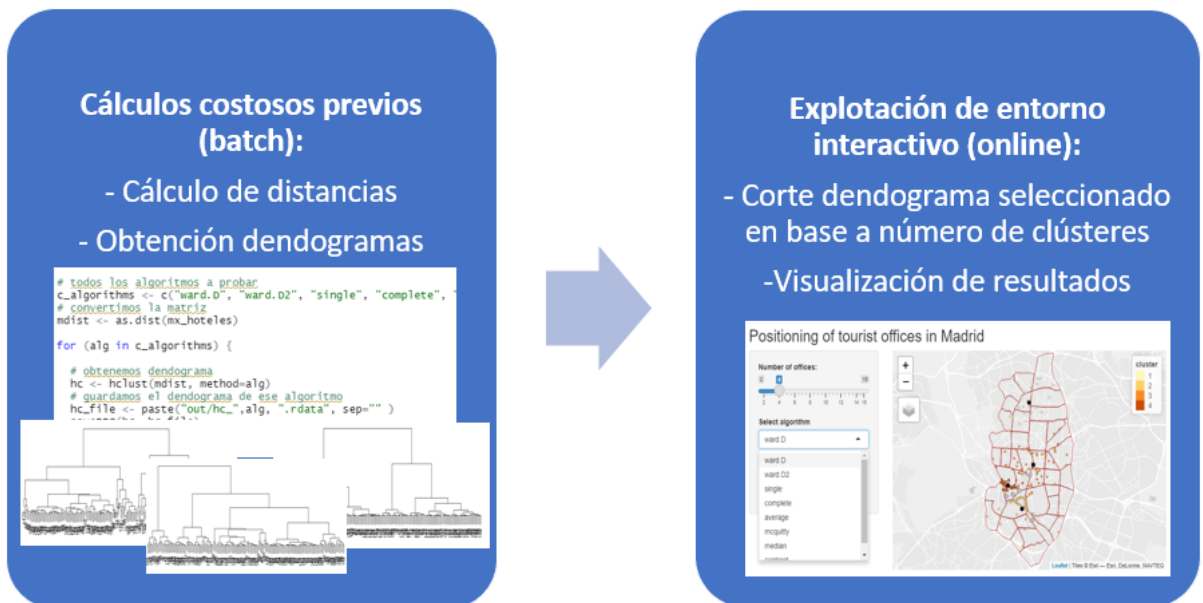


Figura 57: Cálculos del modelo realizados en batch y cálculos realizados en online en el entorno interactivo. Fuente: elaboración propia.

7.3.7. Interpretación del conocimiento obtenido

En este paso el objetivo es interpretar los resultados obtenidos en el paso anterior. En primer lugar, analizamos cómo nos afecta el seleccionar un tipo de método u otro de

clustering jerárquico en la forma en la que se agrupan los hoteles y posicionan las oficinas de turismo.

Los métodos utilizados son: single, complete, average, mcquitty, median, centroid, Ward.D y Ward.D2. Vamos a interpretar los resultados obtenidos con cada uno de ellos.

Empezamos por el método single que tal y como se aprecia en la Figura 58 es muy sensible a los valores atípicos (comúnmente referenciados por su término en inglés: *outlier*) y da lugar a clústeres desbalanceados en tamaño.

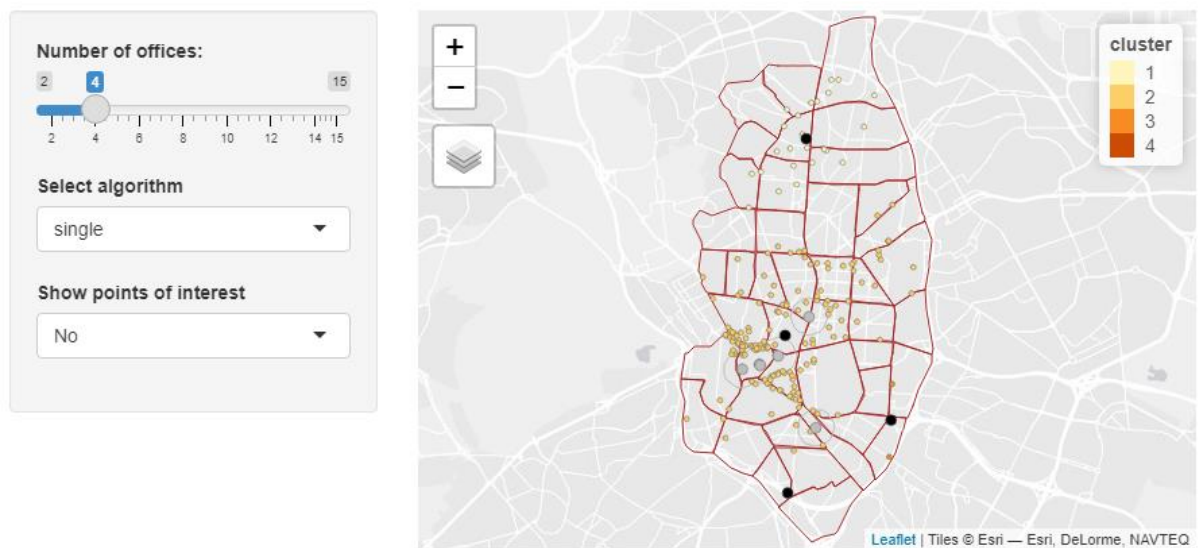


Figura 58: Generación de 4 grupos de hoteles y posicionamiento de oficinas con el método single.

El método complete (véase Figura 59) es menos sensible a los outliers y genera clústeres más balanceados (diámetro similar). En nuestro caso, sitúa nuevas oficinas de turismo muy próximas a las ya existentes.

Number of offices:

Select algorithm:

Show points of interest:

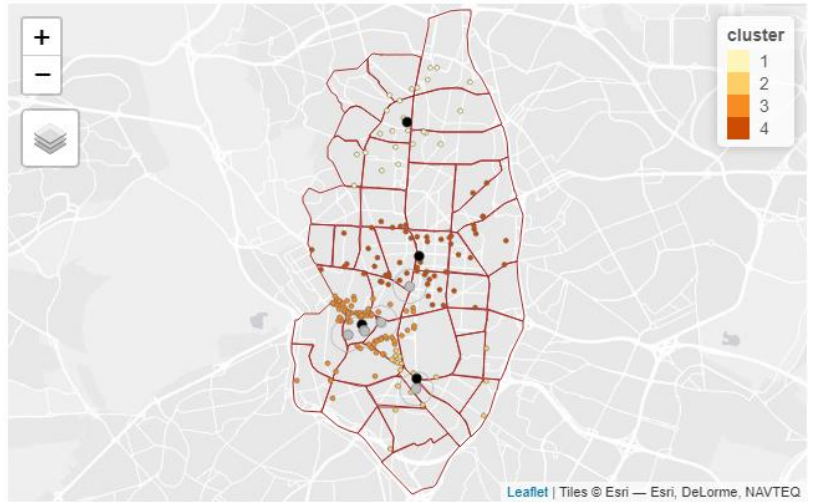


Figura 59: Generación de 4 grupos de hoteles y posicionamiento de oficinas con el método complete.

Los métodos average, mcquitty, median y centroid (véanse Figura 60, Figura 61, Figura 62 y Figura 63) tienden a generar clústeres desbalanceados en tamaño y alguno de los clústeres abarca muchos hoteles. Los resultados también muestran sensibilidad a los outliers.

Number of offices:

Select algorithm:

Show points of interest:

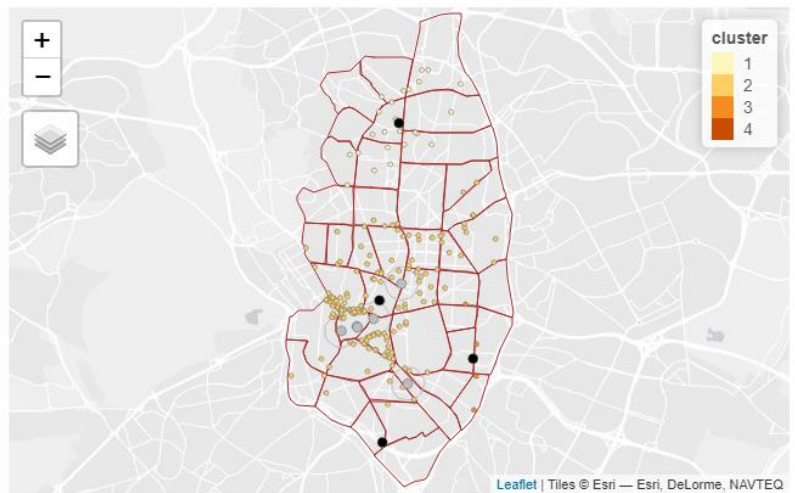


Figura 60: Generación de 4 grupos de hoteles y posicionamiento de oficinas con el método average.

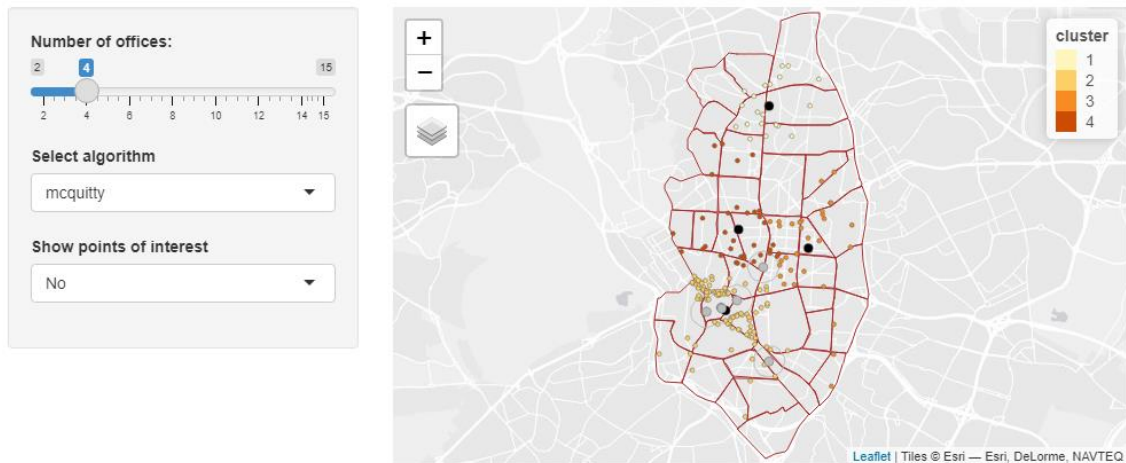


Figura 61: Generación de 4 grupos de hoteles y posicionamiento de oficinas con el método mcquitty.

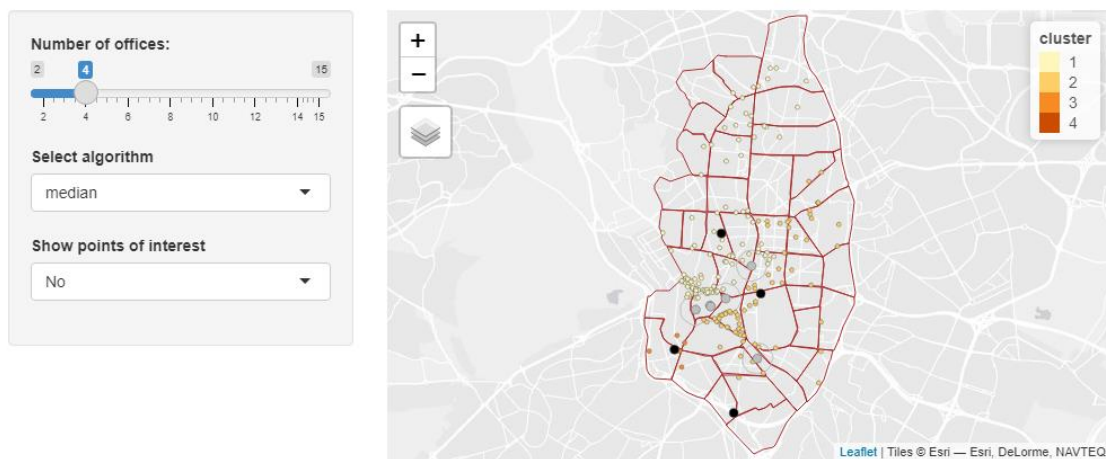


Figura 62: Generación de 4 grupos de hoteles y posicionamiento de oficinas con el método median.

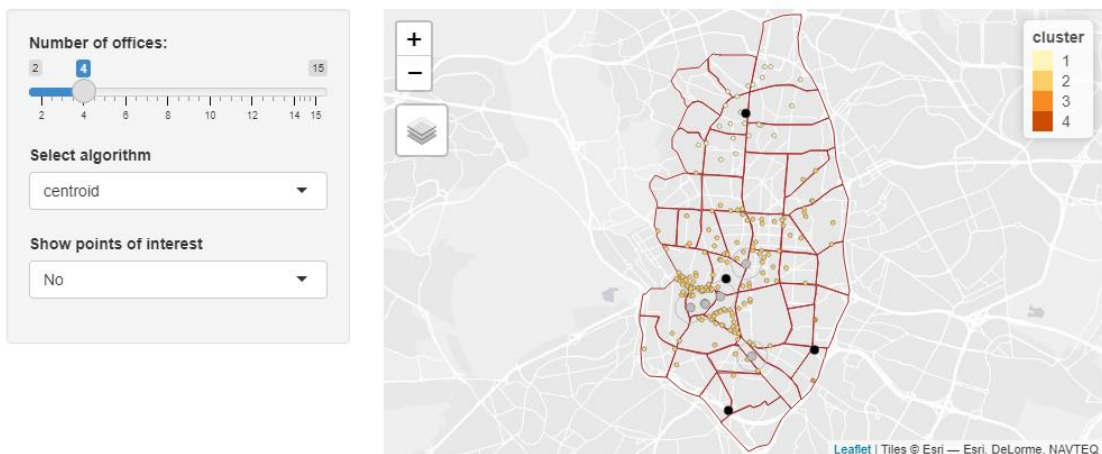


Figura 63: Generación de 4 grupos de hoteles y posicionamiento de oficinas con el método centroid.

Los métodos Ward.D y Ward.D2 (véanse Figura 64 y Figura 65) obtienen clústeres más balanceados en tamaño y no sensibles a outliers. Por otra parte, tienden a posicionar las oficinas en puntos centrados del clúster y con alta densidad de hoteles.

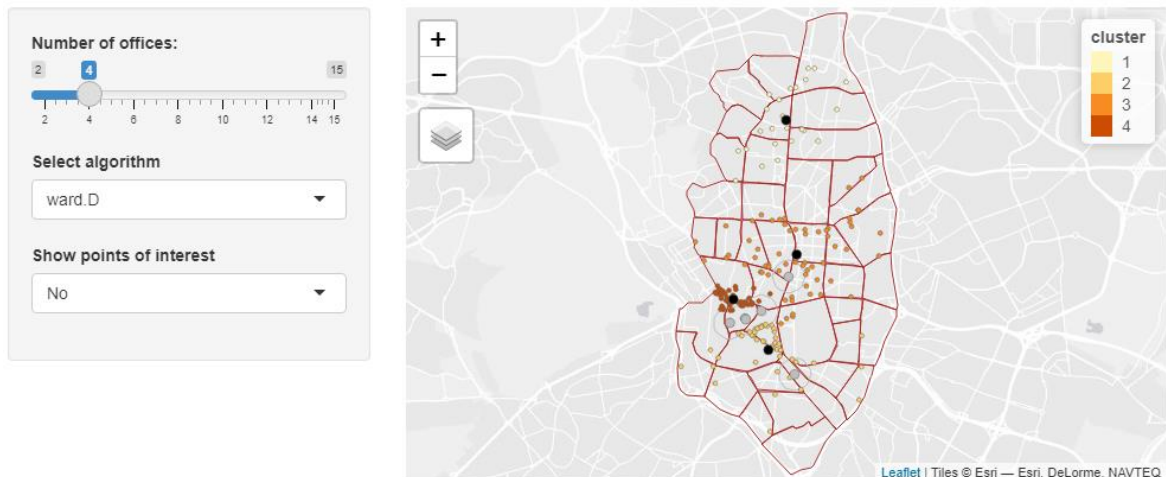


Figura 64: Generación de 4 grupos de hoteles y posicionamiento de oficinas con el método Ward.D.

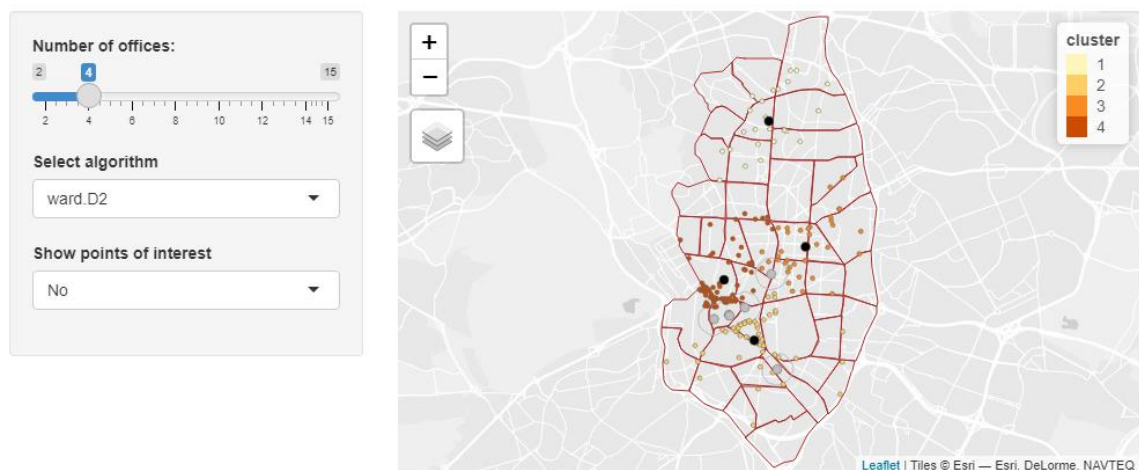


Figura 65: Generación de 4 grupos de hoteles y posicionamiento de oficinas con el método Ward.D2.

Como resumen, cabe señalar que hay muchos métodos afectados por los valores outliers. Los métodos afectados por outliers son: single, average, mcquitty, median y centroid. La elección de un método u otro también afecta al tamaño de los clústeres y hay varios métodos que generan clústeres desbalanceados en tamaño como: single, average, mcquitty, median, centroid y en menor medida complete. Los métodos Ward.D y Ward.D2 obtienen clústeres más balanceados en tamaño y no sensibles a outliers. Por otra parte, estos dos métodos, tienden a posicionar las oficinas en puntos centrados del clúster y con alta densidad de hoteles. Comprobamos, por tanto, que el método seleccionado afecta bastante al resultado y el hecho de poder probar varios en el entorno interactivo permite seleccionar el más adecuado a las necesidades concretas.

De los diferentes métodos probados consideramos que los que dan resultados más adecuados a nuestras necesidades son Ward.D y Ward.D2, siendo más preciso el primero, ya que estos métodos generan clústeres balanceados de forma que todos tengan un número de hoteles similar. Por otra parte, al posicionar las oficinas en puntos centrados en el clúster y con alta densidad de hoteles nos beneficia en el objetivo de proximidad que pretendemos.

Hemos analizado la influencia del método de clustering jerárquico, otro aspecto que podemos modificar es el número de clústeres. Tomamos como método más óptimo Ward.D y probamos diferente número de clústeres. Si seleccionamos números bajos como 2 o 3 vemos clústeres excesivamente grandes y distanciados de los hoteles (véanse Figura 66 y Figura 67) y una de las oficinas nuevas queda posicionada muy cerca de las existentes.

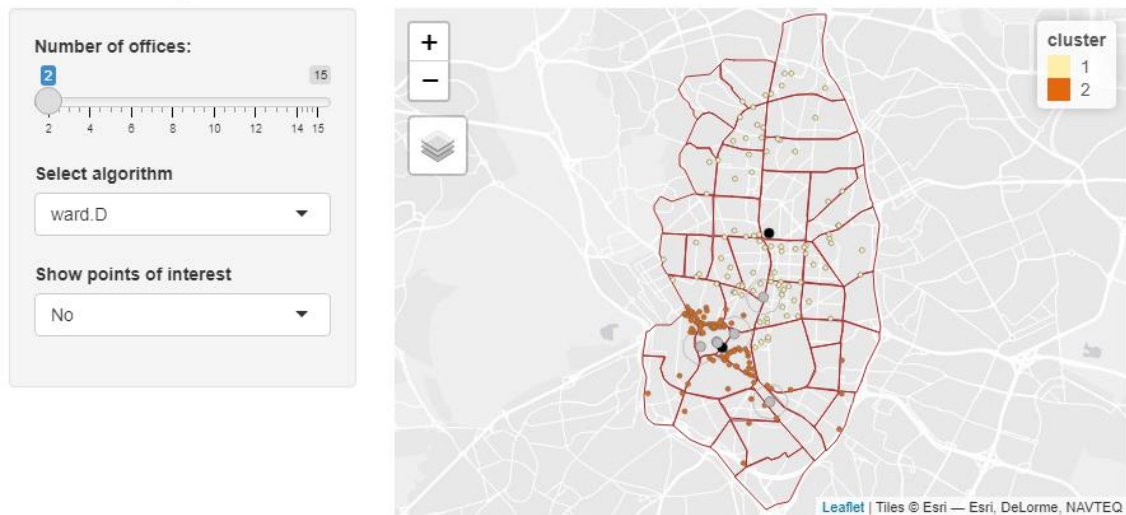


Figura 66: Generación de 2 grupos de hoteles y posicionamiento de oficinas con el método Ward.D.

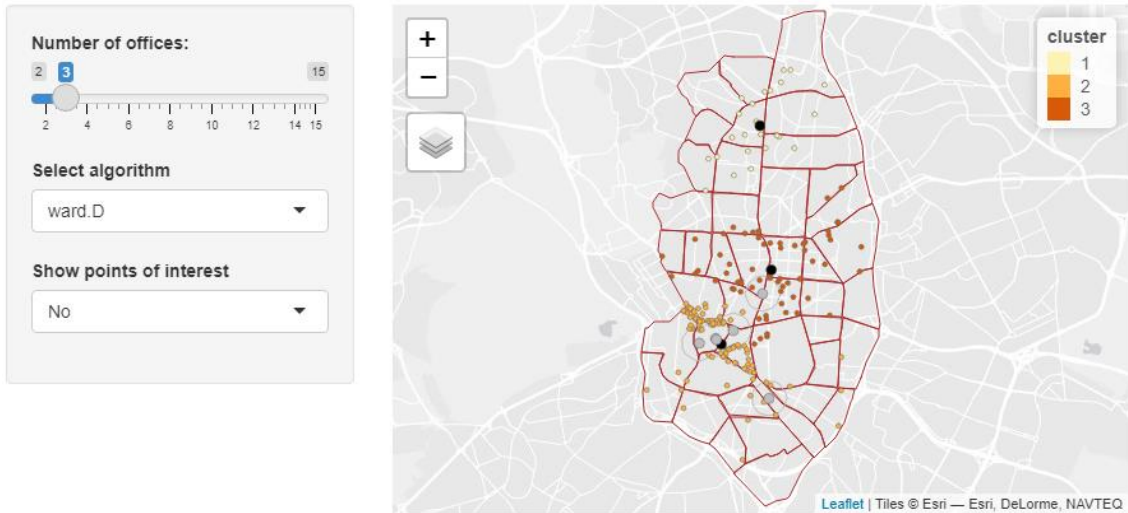


Figura 67: Generación de 3 grupos de hoteles y posicionamiento de oficinas con el método Ward.D.

Consideramos que un número de clústeres de 4 o 5 genera proximidad con los hoteles (véanse Figura 68 y Figura 69), sin tener que abrir un número de oficinas demasiado alto.

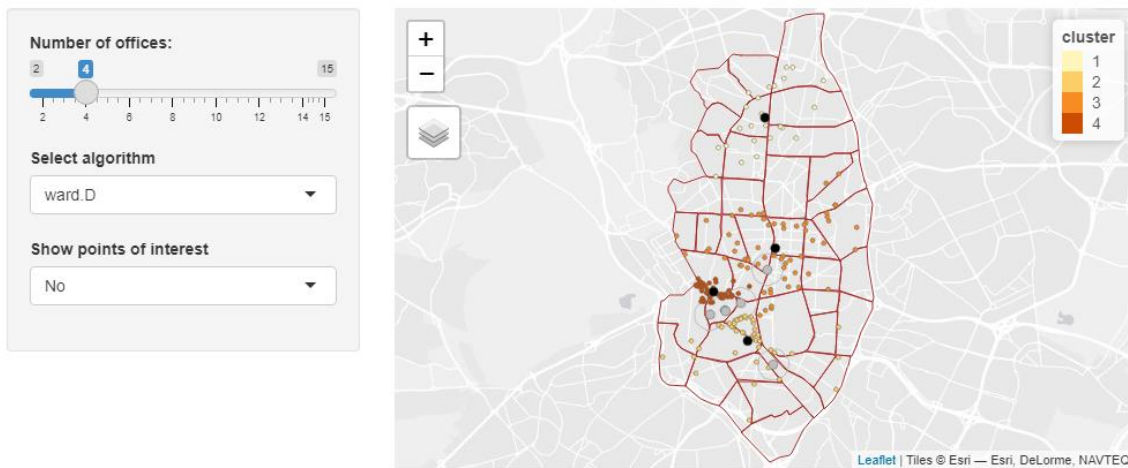


Figura 68: Generación de 4 grupos de hoteles y posicionamiento de oficinas con el método Ward.D.

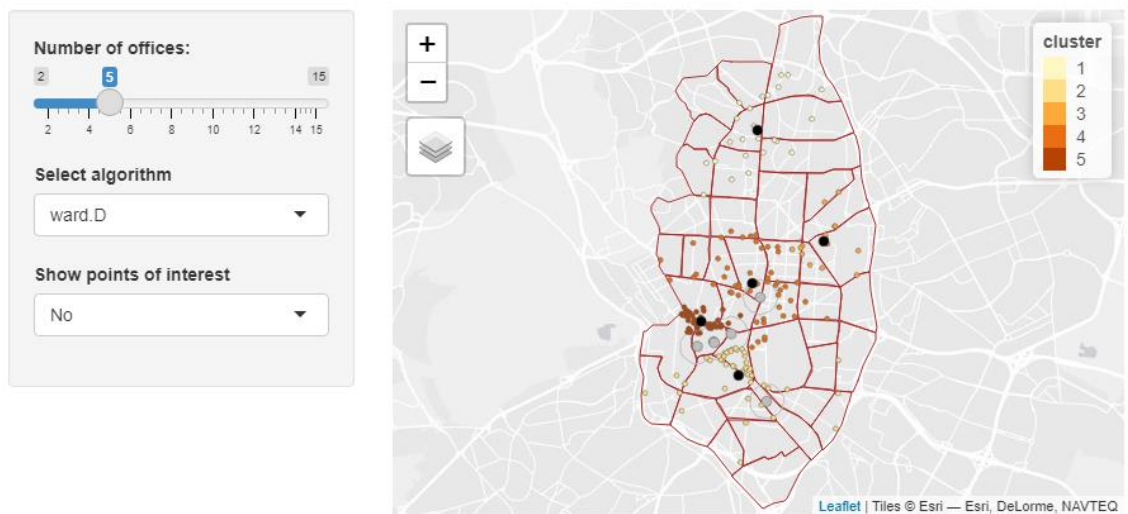


Figura 69: Generación de 5 grupos de hoteles y posicionamiento de oficinas con el método Ward.D.

Por otra parte, vemos que para el caso de Madrid que estamos analizando, existe una importante diferencia en la distribución de las nuevas oficinas de turismo propuestas y las ya existentes. Las últimas están muy próximas entre sí y situadas en la zona central y las nuevas se distribuyen de forma más uniforme por toda la ciudad creciendo hacia la zona norte. Es decir, las oficinas propuestas por el modelo reflejan mejor la distribución de los hoteles que las oficinas ya existentes, como cabría esperar.

De igual forma, en la distribución de las nuevas oficinas propuesta encontramos oficinas que abarcan una zona geográfica mayor y con hoteles más dispersos y otros que abarcan zonas geográficas más pequeñas y con hoteles más concentrados.

Si sobre la visualización anterior añadimos los puntos de interés turístico encontramos que las oficinas con clústeres de hoteles más concentrados son también los que mayor número de puntos de interés turístico tienen cerca tal y como se puede observar en la Figura 70.

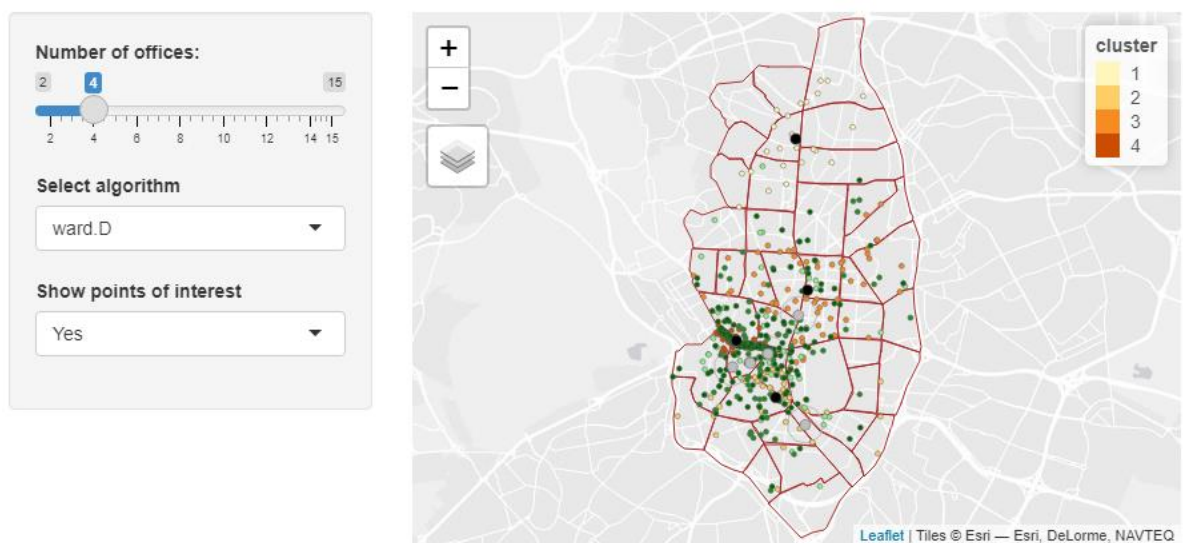


Figura 70: Generación de 4 grupos de hoteles y posicionamiento de oficinas con el método Ward.D. Se añaden los puntos de interés turística en tonalidades verdes.

Finalmente, si tomamos como selección definitiva un número dado de oficinas y un método concreto, podremos obtener recomendaciones sobre estas nuevas oficinas. Vamos a tomar como selección cuatro nuevas oficinas de turismo usando el método Ward.D.

El siguiente paso será obtener recomendaciones para cada una de las cuatro oficinas. Tal y como se explicó anteriormente para obtener recomendaciones vamos a emplear ChatGPT3.5. Para ello generamos un prompt (véase apartado 3.5) con una

pregunta determinada y toda la información de contexto necesaria. La pregunta deberá solicitar una descripción de esta oficina, conocer sus características particulares y recomendaciones de personalización. En lo relativo a la información de contexto vamos a proporcionarle los datos relativos a hoteles y puntos de interés turístico cercanos a la ubicación de la oficina. Vamos a considerar un radio de un kilómetro para obtener información sobre hoteles, teatros, museos y edificios de carácter monumental. Cabe señalar en este punto, que la información de contexto proporcionada está contenida en los conjuntos de datos del modelo, pero podría ser enriquecida con información adicional que consideremos que pueda ser de utilidad.

En nuestra aplicación vamos a añadir información relativa a la situación actual del turismo en la Comunidad de Madrid, para ello vamos a emplear un informe (Anuario de Turismo Madrid 2022 ,2023) generado y publicado con licencia libre por el Centro de Inteligencia Turística del Ayuntamiento de Madrid que detalla la situación del turismo en Madrid en el año 2022. Para el diseño del prompt se han usado como referencia plantillas (Biblioteca de Prompts UCM, 2023) y se han tenido en cuenta las buenas prácticas recomendadas en lo relativo a ingeniería de prompts (Giray, 2023; Korzynski et al., 2023). A continuación, se muestra la plantilla de prompt empleada para las distintas oficinas de turismo obtenidas. Los valores proporcionados entre corchetes serán sustituidos por los valores correspondientes a cada oficina de turismo.

Plantilla de prompt proporcionada a ChatGPT3.5

'Necesito que actúes como experto en turismo. Te voy a dar la siguiente información, en primer lugar, el contexto (apartado 1.CONTEXTO), luego los datos (apartado 2.DATOS) y finalmente la pregunta (apartado 3.PREGUNTA) concreta que quiero que contestes.

1. CONTEXTO: *Queremos situar una nueva oficina de información turística llamada [NOMBRE] en la Comunidad de Madrid. El objetivo es orientar y ayudar a los turistas que visitan la Comunidad de Madrid, así como promocionar y comercializar los productos turísticos en esta Comunidad en colaboración con otros agentes de la cadena de valor del turismo, tales como empresas del sector restauración y ocio, alojamiento, transporte y organismos locales de la Administración y servicios públicos. A continuación, te facilito datos sobre la ubicación de [NOMBRE] y puntos de interés cercanos.*

2. DATOS: *Queremos ubicar [NOMBRE] el punto [COORDENADAS]. Te doy información de puntos de interés cercanos (hoteles, museos teatros y monumentos).*

[DATOS HOTELES: nombre, dirección y coordenadas (latitud y longitud)]

[DATOS MUSEOS: nombre, dirección y coordenadas (latitud y longitud)]

[DATOS TEATROS: nombre, dirección y coordenadas (latitud y longitud)]

[DATOS MONUMENTOS Y EDIFICIOS DE INTERÉS: nombre, dirección y coordenadas (latitud y longitud)]

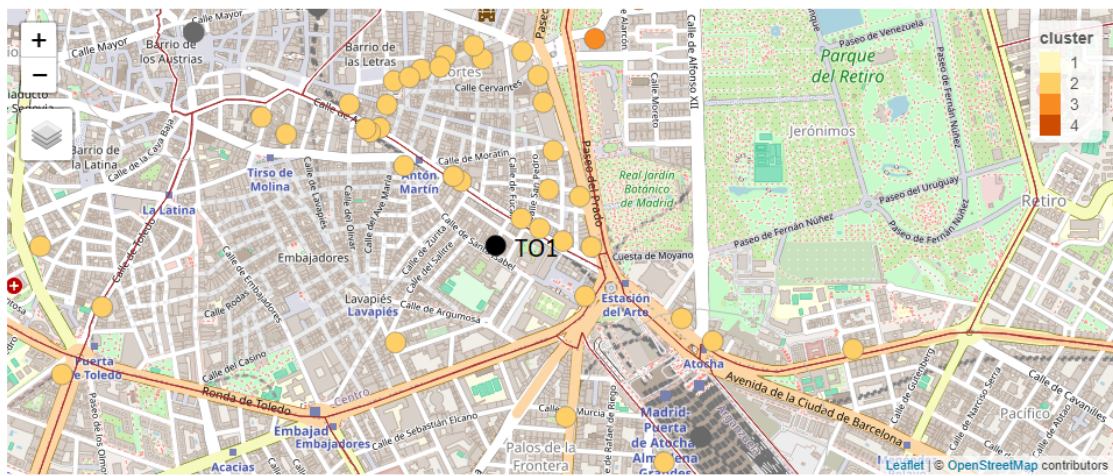
Adicionalmente te proporciono un informe que contiene información relevante sobre el turismo en Madrid durante 2022 que quiero que uses como información de contexto para responder a la pregunta (apartado 3.PREGUNTA).

[DATOS INFORME ANUARIO DE TURISMO MADRID 2022]

Una vez proporcionado todo el contexto y los datos pasamos a la pregunta que quiero que respondas:

3. PREGUNTA: Quiero actúes como experto en turismo y que generes una descripción explicando las características particulares que tiene esta oficina de turismo [OFICINA]. Interesan cuestiones como la ubicación, puntos de interés cercanos y conexiones con otras zonas turísticas de Madrid. Es importante ser lo más específico posible y encontrar características propias que la diferencien de otras posibles ubicaciones y que permitan personalizar una estrategia de marketing concreta. En la generación de la respuesta se deben utilizar los datos proporcionados (apartado 2.DATOS) para contextualizar la respuesta teniendo en cuenta los puntos de interés cercanos y el informe anuario de turismo Madrid 2022. Por favor argumenta tu respuesta paso a paso'

En la Figura 71 se muestra un ejemplo de la salida generada por ChatGPT3.5 para la oficina de turismo propuesta por el modelo en la zona cercana a la Estación de Atocha y el paseo del Prado. En el prompt generado para ChatGPT se ha denominado como TO1 a esta oficina de turismo propuesta.



Recommendations for the tourist office TO1:

TO1 Tourist Office in Madrid:

Strategic location:

The TO1 Tourist Office is located in the heart of Madrid, in an exceptional location in the Retiro district. This district is known for its architectural beauty, its wide green spaces and its proximity to some of the city's main points of interest.

Nearby Points of Interest:

* The Paisaje de la Luz: The location of the OT1 Tourist Office is perfect for art and culture lovers, as it is just steps away from the famous Paisaje de la Luz, an emblematic area of Madrid that is home to some of the most outstanding museums, such as the Prado Museum, the Thyssen-Bornemisza Museum and the Reina Sofia Museum. Visitors can explore the city's rich artistic history and enjoy masterpieces by artists such as Goya, Velázquez and Picasso.

* Retiro Park: Within walking distance of the office, tourists can relax in the Retiro Park, one of Madrid's largest and most beautiful parks. It offers activities such as boat rides on the lake, visits to the Crystal Palace and outdoor cultural events.

* Atocha Station: The OT1 Tourist Office is strategically located near the Atocha train station, which is an important connection point for travel to cities near Madrid. Visitors can easily take high speed trains (AVE) to visit places such as Toledo, Aranjuez or El Escorial, making it a hub for travellers wishing to explore the region.

Figura 71: Visualización en el entorno interactivo de una de las oficinas de turismo TO1 ubicada en la zona Estación de Atocha y Paseo del Prado y recomendaciones obtenidas para la misma.

El texto completo de la recomendación obtenida es el siguiente.

Recomendaciones ChatGPT3.5 oficina de turismo Estación de Atocha – Paseo del Prado:

“Por supuesto, estaré encantado de actuar como experto en turismo y proporcionarte una descripción detallada de la Oficina de Turismo TO1 en Madrid, destacando sus características particulares, ubicación estratégica, puntos de interés cercanos y conexiones con otras zonas turísticas de la ciudad. Utilizaré los datos proporcionados y el informe anual de turismo en Madrid de 2022 para contextualizar mi respuesta.

Oficina de Turismo TO1 en Madrid:

Ubicación Estratégica:

La Oficina de Turismo TO1 se encuentra en el corazón de Madrid, en una ubicación excepcional en el distrito de Retiro. Este distrito es conocido por su belleza arquitectónica, sus amplios espacios verdes y su proximidad a algunos de los principales puntos de interés de la ciudad.

Puntos de Interés Cercanos:

- **El Paisaje de la Luz:** La ubicación de la Oficina de Turismo TO1 es perfecta para los amantes del arte y la cultura, ya que está a pocos pasos del famoso “Paisaje de la Luz”, una zona emblemática de Madrid que alberga algunos de los museos más destacados, como el Museo del Prado, el Museo Thyssen-Bornemisza y el Museo Reina Sofía. Los visitantes pueden explorar la rica historia artística de la ciudad y disfrutar de obras maestras de artistas como Goya, Velázquez y Picasso.
- **Parque del Retiro:** A poca distancia de la oficina, los turistas pueden relajarse en el Parque del Retiro, uno de los parques más grandes y hermosos de Madrid. Ofrece actividades como paseos en bote en el lago, visitas al Palacio de Cristal y eventos culturales al aire libre.
- **Estación de Atocha:** La Oficina de Turismo TO1 está estratégicamente ubicada cerca de la estación de tren de Atocha, que es un importante punto de conexión para viajar a ciudades cercanas a Madrid. Los visitantes pueden fácilmente tomar trenes de alta velocidad (AVE) para visitar lugares como Toledo, Aranjuez o El Escorial, lo que la convierte en un centro neurálgico para los viajeros que desean explorar la región.

Estrategia de Marketing Personalizada:

Basándonos en estas características únicas, la Oficina de Turismo TO1 podría desarrollar una estrategia de marketing personalizada que se centre en los siguientes aspectos:

- **Turismo Cultural y Artístico:** Destacando su proximidad a los museos más prestigiosos de Madrid, la oficina podría promocionar paquetes turísticos culturales que incluyan visitas exclusivas a estas instituciones.
- **Turismo de Lujo:** Dada su ubicación en el distrito de Retiro y su proximidad a hoteles de lujo, la Oficina de Turismo TO1 es ideal para promover el turismo de lujo en Madrid. Puede ofrecer paquetes y recomendaciones exclusivas para aquellos que buscan una experiencia de alta gama, como visitas VIP a museos o reservas en restaurantes gourmet. Colaboración con hoteles de lujo cercanos para ofrecer paquetes de alojamiento y experiencias exclusivas para viajeros que buscan el máximo confort y elegancia.

- **Turismo de Excursión:** Promocionar viajes de un día a las ciudades cercanas desde la estación de Atocha, destacando la comodidad de llegar en tren de alta velocidad y regresar a Madrid por la noche.
- **Gastronomía:** Madrid está experimentando una revolución gastronómica y cuenta con 22 restaurantes con al menos una estrella Michelin. La Oficina de Turismo TO1 está lista para guiar a los amantes de la comida a través de esta emocionante escena culinaria, recomendando restaurantes de renombre y lugares auténticos para probar la rica comida española.
- **Cine y Cultura:** Si eres un amante del cine y la cultura, Madrid es el lugar perfecto para ti. La ciudad se ha convertido en un importante centro de producción de contenido tanto a nivel nacional como internacional.

En resumen, la Oficina de Turismo TO1 en Madrid se destaca por su ubicación estratégica en el distrito de Retiro, su proximidad a los principales puntos de interés, como el Paisaje de la Luz y el Parque del Retiro, y su papel como punto de conexión para explorar ciudades cercanas. Estas características permiten una estrategia de marketing personalizada que puede atraer a diversos tipos de viajeros y mejorar la experiencia turística en Madrid”

La respuesta de ChatGPT3.5 ha sido obtenida el 4 de octubre de 2023 con la actualización del 25 de septiembre. Como se puede observar ChatGPT3.5 nos puede facilitar recomendaciones para personalizar la oficina de turismo. Cabe señalar que las respuestas ofrecidas pueden variar en base a pequeñas modificaciones realizadas en el prompt de entrada.

Tal y como se observa en la Figura 48, esta fase queda integrada en la plataforma propuesta en la capa 3.

7.4. Resultados

En este apartado nos centramos en los resultados obtenidos en la aplicación del modelo propuesto. Este apartado se corresponde con la fase 'Aplicación de negocio del conocimiento obtenido' del modelo propuesto (Figura 48) que se integra en la plataforma propuesta en la capa 3.

Nuestro modelo permite posicionar de forma óptima nuevas oficinas de turismo, pero también nos ofrece un conocimiento que permite distinguir entre diferentes tipos de oficinas y personalizarlas, lo cual era también uno de los objetivos planteados en el apartado 7.3.1. Este conocimiento proviene tanto de las opciones de visualización ofrecidas del entorno interactivo como de las recomendaciones obtenidas de ChatGPT3.5.

Encontramos dos tipos de oficinas en base a su posicionamiento:

- Oficinas tipo 1, que abarcan zonas geográficas amplias, con hoteles dispersos y pocos puntos de interés turístico cerca.
- Oficinas tipo 2, que abarcan zonas geográficas más reducidas, con hoteles más concentrados y muchos puntos de interés turístico cerca.

Las oficinas de tipo 2 es esperable que no estén posicionadas en zonas de alto afluencia de turistas y deberán hacer uso de puntos de información turística móviles que se posicionen en puntos de mayor afluencia de turistas dentro del del área geográfica del clúster de la oficina. Por otro lado, las oficinas de tipo 1 tienen una ubicación más óptima respecto a la proximidad a los turistas, ya que tienen un mayor número de hoteles muy cercanos y puntos de interés turístico próximos.

El modelo está planteado para aplicarlo a cualquier destino turístico del que se dispongan de los datos necesarios, pero a continuación profundizamos en el caso de Madrid para encontrar aplicaciones concretas teniendo en cuenta sus características particulares como destino turístico.

Si nos centramos en el caso anteriormente analizado en la que posicionamos cuatro nuevas oficinas de turismo usando el método Ward.D vemos que las oficinas de turismo ya existentes cubren los ejes turísticos tradicionales como son las zonas de barrio de las letras, barrio de los Austrias, estación de Atocha o plaza Colón tal y como vemos en la Figura 72.

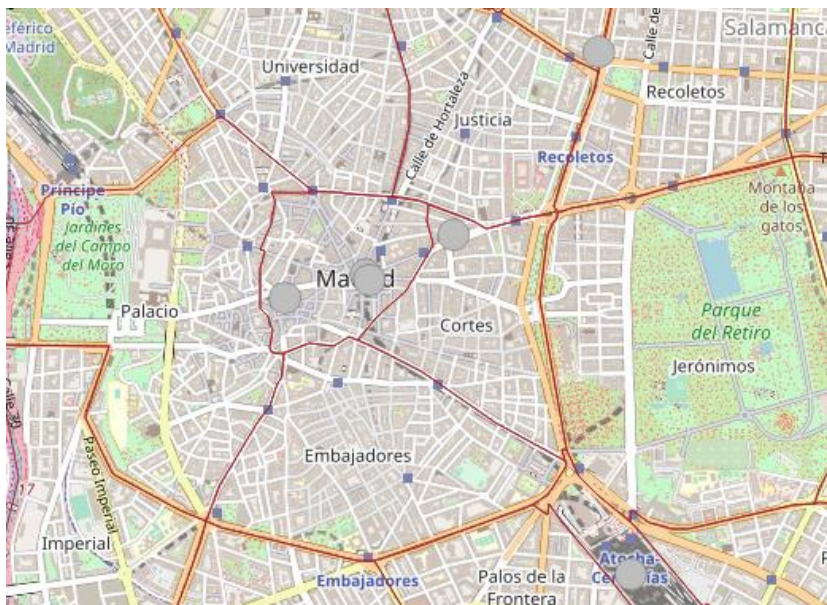


Figura 72: Oficinas de turismo ya existentes

Por otro lado, las nuevas oficinas de turismo propuestas cubren ejes turísticos que han experimentado cambios recientemente. En la Figura 73 vemos una nueva oficina de

eje turístico tradicional que está experimentando un importante crecimiento gracias principalmente al crecimiento del turismo de negocios y congresos, y debido en parte también al crecimiento urbanístico de la ciudad en esa zona impulsado por el proyecto 'Madrid nuevo Norte'. La zona de la Castellana más próxima a la plaza de Colón se ubica cerca del Barrio de Salamanca tradicional eje turístico que ha experimentado un importante crecimiento en hoteles, restaurantes y tiendas de lujo atrayendo turistas de alto poder adquisitivo.



Figura 75: Oficina de turismo nueva (punto negro) posicionada en la Castellana en la confluencia con la calle José Ortega y Gasset cerca de la plaza de Colón

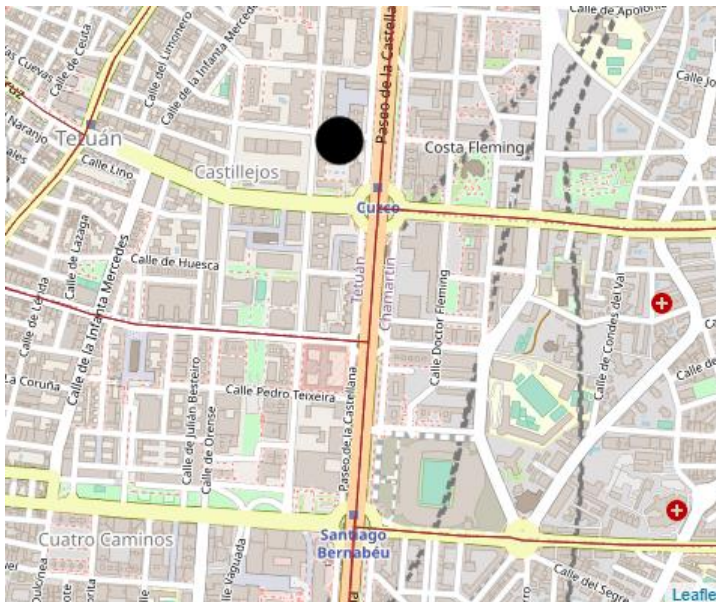


Figura 76: Oficina de turismo nueva (punto negro) posicionada en la Castellana cerca de la plaza de Cuzco.

Conviene señalar que como es lógico el posicionamiento propuesto de las oficinas turísticas es siempre aproximado y será necesario analizar en detalle la localización

propuesta para buscar en el entorno cercano a esta localización un punto con alto tránsito de transeúntes y comprobar la disponibilidad de locales o inmuebles. Es decir, sobre el posicionamiento propuesto habrá ligeras variaciones.

Consideramos que el posicionamiento de las nuevas oficinas de turismo refleja fielmente los cambios que ha experimentado Madrid en los últimos años como destino turístico lo cual era uno de los objetivos del modelo propuesto:

- Nuevos ejes turísticos, como Plaza de España y Paisaje de la Luz.
- Crecimiento turismo de congresos.
- Crecimiento turismo segmento de lujo.

Por otra parte, el conocimiento de la zona donde se ubican las nuevas oficinas de turismo y las recomendaciones de ChatGPT3.5 permiten personalizar y adecuar la oferta al tipo de turista de cada zona: turismo de negocios, turismo de lujo, turismo cultural, etc.

Nótese que hay una diferencia considerable entre la distribución de las oficinas ya existentes y las nuevas. Las primeras están muy próximas entre sí y situadas en la zona central y las segundas se distribuyen de forma más uniforme por toda la ciudad creciendo hacia la zona norte. Esta descentralización de las oficinas también debe ser considerada en la gestión de las nuevas oficinas. En primer lugar, posicionar oficinas turísticas lejos de la zona central de la ciudad puede ayudar no sólo a estar más próximos a turistas ubicados en hoteles lejanos al centro proporcionándoles la información del destino turístico, sino a facilitarles información de movilidad que les permita moverse por la ciudad para desplazarse al centro u otros ejes turísticos de la forma más eficiente y sostenible. Estas nuevas oficinas pueden estar próximas a estaciones de trenes como Atocha u otros puntos de enlace en la infraestructura de transportes, esta ubicación es interesante para ofrecer información de lugares próximos al destino turístico que también sean de interés turístico y dispongan de una buena comunicación desde esta oficina. En el caso de Madrid a una distancia inferior a 50 kilómetros existen tres ciudades que son Patrimonio de la Humanidad como son Aranjuez, Alcalá de Henares y San Lorenzo de El Escorial y una cuarta a 75 kilómetros como es Toledo.

En resumen, podemos decir que los conocimientos obtenidos por el modelo para el caso Madrid ayudan a posicionar de forma más óptima y actualizada las oficinas de turismo, adaptándose a los cambios que experimenta el destino turístico. Así mismo,

ayuda a personalizar las diferentes oficinas en base a su entorno geográfico próximo y las recomendaciones de ChatGPT3.5.

7.5. Conclusiones y trabajo futuro

En este caso de uso se ha planteado una plataforma de datos e inteligencia artificial para el posicionamiento de STS basada en tres capas sobre el que se desarrolla un modelo de decisión basado en la metodología KDD para el posicionamiento óptimo de servicios turísticos. Esta plataforma aparte de posicionar de forma óptima los servicios turísticos, permite personalizar los mismos a su entorno cercano gracias a su entorno interactivo que facilita la interpretación de resultados, realizar simulaciones y genera recomendaciones. Se ha desarrollado una aplicación concreta de posicionamiento de oficinas de información turística para la ciudad de Madrid. En esta aplicación se proponen nuevas ubicaciones de oficinas de turismo para la ciudad de Madrid con recomendaciones personalizadas. Así mismo, se comprueba que los resultados del modelo reflejan correctamente los cambios que se han producido en los últimos años en Madrid como destino turístico. El modelo puede ser aplicado a cualquier otro tipo de servicio turístico sobre el que se disponga de la información relevante identificada en las fuentes de datos de la capa 1 de la plataforma propuesta. Esta plataforma es, por tanto, de aplicación a cualquier destino turísticos inteligente que base sus decisiones en datos y quiera mejorar el posicionamiento de sus servicios turísticos.

En lo relativo a los datos empleados en la plataforma, consideramos que el desarrollo de los espacios de datos compartidos en el sector del turismo (Gaia-X, 2023; Tourism Data Space, 2023) puede ser de gran utilidad para dotar de conjuntos de datos de gran interés a la plataforma y obtener recomendaciones sobre distintos tipos de servicios turísticos.

Si analizamos las técnicas empleadas, una posible línea de trabajo futura es el uso de otras técnicas de clustering diferentes al clustering jerárquico como puede ser el clustering difuso en el cual los elementos tienen un grado de pertenencia a cada clúster y estos no son excluyentes entre sí. El número de clústeres es un parámetro que se selecciona en nuestro modelo, pero sería interesante que el propio modelo propusiera un número óptimo de clústeres en base al conjunto de datos proporcionado. Del mismo modo, la distancia 'wd-5' empleada podría ser un parámetro que pueda ser modificado en el entorno interactivo planteado para poder realizar simulaciones con el mismo. Para implementar este cambio será necesario contar con un entorno de computación adecuado para poder realizar el cálculo de distancias en la última capa de la plataforma.

En esta línea el papel de los procesos batch y online en los cálculos del modelo se podrían distribuir de otra manera. Por ejemplo, si disponemos de una alta capacidad de computación podemos llevar cálculos al proceso online para que se ejecuten en la última capa de la plataforma. Llevar cálculo a la última capa tiene el beneficio de dotar de interactividad a estos cálculos definiendo parámetros que permitan realizar simulaciones, tal y como se ha indicado para el caso de 'wd-5'. Por otra parte, consideramos que el papel de los LLM en nuestra plataforma puede cambiar a futuro, el desarrollo de estos modelos va a permitir obtener mejores recomendaciones y el protagonismo de estos modelos en nuestra plataforma será mayor. Por último, los datos que se incorporan a los LLM se pueden enriquecer y emplear bases de datos vectoriales con conjuntos de datos de altos volumen de información sobre actividad turística en el destino turístico de estudio. La información de estos conjuntos de datos combinada con el uso de las novedosas técnicas de generación aumentada de recuperación (RAG, por sus siglas en inglés Retrieval Augmented Generation) pueden ayudar a mejorar sustancialmente la calidad de las recomendaciones.

8. Conclusiones y trabajo futuro

En este apartado se desarrollan las conclusiones de esta tesis y el posible trabajo futuro a desarrollar.

8.1. Conclusiones

Durante el desarrollo de esta tesis, se ha analizado y evaluado la situación de las organizaciones turísticas respecto al uso de fuentes de datos, conociendo el tipo de fuentes de datos que utilizan, las técnicas de análisis de datos y los principales temas de investigación. Esta revisión se ha realizado mediante un análisis bibliométrico y era el primer objetivo dentro de los objetivos específicos de esta tesis. Durante esta revisión se detecta que no existe actualmente un marco de trabajo robusto que facilite toda esta gestión y la toma de decisiones basadas en datos en el sector del turismo.

Partiendo de los resultados de este análisis bibliométrico se obtiene un marco de referencia inicial sobre el uso de datos en la toma de decisiones en las organizaciones turísticas. A partir de este marco de referencia inicial se identifican una serie de gaps que se traducen en una serie de necesidades que se deben cubrir para poder abordar los desafíos actuales del sector del turismo con una estrategia orientada al dato. Una vez definida la situación inicial y las necesidades detectadas, se ha definido un marco conceptual para la toma de decisiones basadas en datos en una organización turística. Los pasos realizados desde el análisis bibliométrico hasta la obtención del marco conceptual cubren el segundo objetivo definido en los objetivos específicos de esta tesis.

Con la definición de este marco conceptual, se cumple el objetivo general de la tesis definiendo un marco de trabajo robusto que facilita toda esta gestión y la toma de decisiones basadas en datos en el sector del turismo. Para demostrar la eficacia de este marco conceptual, se han desarrollado tres casos de uso, cumpliendo así el tercer objetivo dentro de los objetivos específicos de la tesis. En cada uno de ellos, este marco se ha mostrado como una herramienta eficaz de cara a abordar problemas de negocio desde una perspectiva de datos. En lo relativo al trabajo con los datos en cada caso de uso, el marco conceptual propuesto sirve de guía para la gestión y almacenamiento de las diferentes fuentes de datos, así como para la aplicación de técnicas de ciencia de datos e inteligencia artificial. Así mismo, en el desarrollo del caso del uso el marco conceptual nos muestra todos los ámbitos de negocio en el sector del turismo donde el uso de datos tiene aplicación. Se demuestra que el uso del marco conceptual facilita la toma de

decisiones basadas en datos, obteniendo beneficios para el negocio desde diferentes perspectivas:

- **Analizar la experiencia del turista para mejorar la gestión de destinos.** Para representar la experiencia de los turistas, se ha formalizado el concepto de CDE y se han establecido pasos para su obtención a partir de reseñas de turistas. Se propone un modelo que detecta y analiza estas CDEs, aplicable a cualquier destino turístico. Este modelo obtiene un mapa de experiencias a partir del cual pueden obtenerse recomendaciones generales y personalizadas. Los resultados son de gran ayuda para que las organizaciones de gestión de destinos conozcan en profundidad las experiencias de los turistas y tomen decisiones de mejora basadas en este conocimiento. Este modelo puede ser valioso en investigaciones sobre competitividad turística y permite comparar destinos, como Roma, Londres, París o Madrid, desde la perspectiva de la experiencia. Se ha desarrollado una aplicación para la ciudad de Roma. En esta aplicación se obtienen recomendaciones para la ciudad de Roma sobre seguridad pública, mejoras operativas, optimización itinerarios y gestión de tiempos de espera.
- **Apoyo a la toma de decisiones en la estrategia de digitalización de tienda física en el sector del turismo.** El modelo propuesto facilita la estrategia de digitalización para tiendas físicas del sector turístico, proporcionando un enfoque sistemático basado en AHP para gestionar las prioridades de inversión. El modelo involucra en la toma de decisiones a todas las áreas de la organización mediante el uso del BSC. En el caso concreto de aplicación a agencias de viajes se propone un ranking de prioridades en la adopción de dispositivos digitales. En este ranking destaca la preferencia por tecnologías como 3D (AR y VR) y códigos QR, impulsadas principalmente por las áreas de finanzas y aprendizaje y crecimiento del BSC.
- **Posicionamiento óptimo de servicios turísticos inteligentes.** Se propone una plataforma de datos e inteligencia artificial y un modelo de decisión que permiten determinar el posicionamiento óptimo de los STS. Esta plataforma aparte de posicionar de forma óptima los servicios turísticos, permite personalizar los mismos a su entorno cercano gracias a su entorno interactivo que facilita la interpretación de resultados, realizar simulaciones y genera recomendaciones. Se realiza una aplicación para el posicionamiento de oficinas de turismo en la ciudad de Madrid, donde se proponen nuevas ubicaciones de oficinas de turismo para la ciudad, que reflejan fielmente los cambios recientes de la ciudad como destino turístico y se añaden recomendaciones personalizadas sobre su gestión. Esta plataforma es de aplicación a cualquier destino que desee posicionar de forma óptima los servicios turísticos.

Cada una de las conclusiones concretas referentes a los distintos casos de uso se han desarrollado de forma más detallada en el apartado correspondiente al caso de uso en esta memoria.

8.2. Trabajo futuro

Podemos identificar diferentes líneas de trabajo futuro para el marco conceptual propuesto que están relacionadas con la revolución que tenemos actualmente debido a los avances de la inteligencia artificial y por otro lado con los cambios y tendencias actuales en la actividad turística. Podemos destacar los siguientes:

- Adecuación del marco conceptual para el cumplimiento de las regulaciones relacionadas con el uso de la inteligencia artificial. Los marcos regulatorios entorno al uso de la inteligencia artificial van a tener un importante desarrollo en los próximos años.
- Desde el punto de vista del uso de técnicas de ciencia de datos e inteligencia artificial en el marco conceptual propuesto, es evidente que el desarrollo actual de la inteligencia artificial generativa va a traer nuevas e interesantes aplicaciones al sector del turismo.
- Considerando el auge de las iniciativas de compartición de datos y el desarrollo de los espacios de datos en el sector del turismo, es importante potenciar en el marco conceptual propuesto para la interoperabilidad de datos. Será, por tanto, importante el uso de protocolos de interoperabilidad, normas comunes, semánticas u ontologías que serán definidas en las diferentes iniciativas de compartición de datos en los espacios de datos del sector del turismo.
- El marco conceptual propuesto está preparado para manejar diferentes tipos de fuentes de datos y será necesaria su revisión debido al continuo crecimiento de datos que experimentamos en el sector del turismo, en especial aquellos derivados del uso de IoT, realidad virtual, metaverso, asistentes virtuales y la personalización y automatización de los servicios turísticos.
- Las aplicaciones de negocio del marco conceptual deben irse adaptando a las tendencias a futuro del sector del turismo como puede ser la creciente importancia del turismo de sostenibilidad, el turismo accesible, turismo de salud, turismo de lujo y turismo experiencial, entre otros.

Así mismo, dentro ámbito concreto de aplicación de cada uno de los casos de uso realizados, se proponen líneas de trabajo futuro que se han desarrollado en el apartado correspondiente al caso de uso en esta memoria.

9. Bibliografía

- Abdulaziz, TA., Moawad, IF., Abu-Alam, WM. (2015). Building Data Warehouse System for The Tourism Sector. 2015 IEEE Seventh International Conference on Intelligent Computing and Information Systems (ICICIS), 410-417. DOI: 10.1109/IntelCIS.2015.7397253
- Aditya, J, Gandhar, K , Vraj, S (2018). Natural Language Processing. International Journal of Computer Sciences and Engineering, 6(1), 161-167. DOI:10.26438/ijcse/v6i1.161167
- Akbari, M., Foroudi, P., Fashami, RZ., Mahavarpour, N., Khodayari, M. (2022). Let us talk about something: The evolution of e-WOM from the past to the future. Journal of Business Research, 149, 663-689. DOI:10.1016/j.jbusres.2022.05.061
- Al-Saad, SA, Jawarneh, RN, Aloudat, AS (2023). Spatiotemporal cluster analysis of reputable tourist accommodation in Greater Amman Municipality, Jordan. Journal of Hospitality and Tourism Technology. DOI:10.1108/JHTT-03-2021-0071
- Alonso, S.; Cabrerizo, F. J.; Herrera-Viedma, E.; Herrera, F. (2009). H-Index: A review focused in its variants, computation and standardization for different scientific fields. Journal of Informetrics, 3 (4), 273-289. DOI:10.1016/j.joi.2009.04.001
- Alrawadieh, Z., Dincer, MZ., Dincer, FI., Mammadova, P. (2018). Understanding destination image from the perspective of Western travel bloggers: the case of Istanbul. International Journal of Culture Tourism and Hospitality Research, 12(2), 198-212. DOI:10.1108/IJCTHR-12-2017-0124
- Ammirato, S., Felicetti, AM., Della Gala, M., Raso, C., Cozza, M. (2018). Smart Tourism Destinations: Can the Destination Management Organizations Exploit Benefits of the ICTs? Evidences from a Multiple Case Study. Collaborative Networks of Cognitive Systems, 534, 623-634. DOI:10.1007/978-3-319-99127-6_54
- Antonio, N., Correia, MB. , Ribeiro, FP. (2020). Exploring User-Generated Content for Improving Destination Knowledge: The Case of Two World Heritage Cities. Sustainability, 12 (2). DOI10.3390/su12229654
- Anuario de Turismo Madrid 2022 (2023). https://www.esmadrid.com/mcb/wp-content/uploads/2023/08/anuario_completo_madrid_turismo_2022.pdf. Accessed 29, 09, 2023
- Arana, JE. ,Leon, CJ. , Carballo, MM., Gil, SM. (2016). Designing Tourist Information Offices: The Role of the Human Factor. Journal of Travel Research, 55(6), 764-773. DOI: 10.1177/0047287515587113
- Azmi, D., Karim, H., & Amin, M.Z.M. (2012). Comparing the walking

- behavior between urban and rural residents. *Procedia - Social and Behavioral Sciences*, 68(19), 406-416. DOI: 10.1016/j.sbspro.2012.12.237
- Barisic, P.; Nanic, A., Vrkljan, S. (2019). UGC and FGC as the new challenges for tourist destination management. *Interdisciplinary Management Research XV (IMR)*, 15, 869-890. *International Journal of Mobile Computing and Multimedia Communications*, 11(4), 26-41. DOI:10.4018/IJMCMC.2020100102.
 - BeautifulSoup (2022). <https://pypi.org/project/beautifulsoup4/>. Accessed 01, 03, 2022
 - Belghiti, S., Ochs, A., Lemoine, J. F., & Badot, O. (2018). The phygital shopping experience: An attempt at conceptualization and empirical investigation. In *Academy of Marketing Science World Marketing Congress. Marketing Transformation: Marketing Practice in an Ever Changing World* (pp. 61–74). Springer, Cham. DOI: 10.1007/978-3-319-68750-6_18
 - Bellini, E., Bellini P., Cenni, D., Nesi, P., Pantaleo, G., Paoli, I., Paolucci, M. (2021). An IoE and Big Multimedia Data Approach for Urban Transport System Resilience Management in Smart Cities. *Sensors*, 21(2), DOI:10.3390/s21020435
 - Benckendorff, P., Zehrer, A. (2013). A network analysis of tourism research. *Annals of Tourism Research*, 43 (1), 121-149. DOI:10.1016/j.annals.2013.04.005
 - Benslimane, A. & Semaoune, K. (2021). An Analysis of Tourist's Behavioural Intention in the Digital Era: Using a Modified Model of the Reasoned Action Theory. *International Journal of Marketing Communication and New Media*, 9(17), 4-26. DOI:10.54663/2182-9306.2021.v9n17.4-26
 - Biblioteca de Prompts UCM (2023). <https://ssii.ucm.es/biblioteca-de-prompts>. Accessed 26, 09, 2023.
 - Bilen, T., Erel-Özçevik, Yaslan, Y., Oktug, SF. (2018). A Smart City Application: Business Location Estimator using Machine Learning Techniques. *IEEE 20TH International conference on high performance computing and communications / IEEE 16TH International conference on smart city / IEEE 4TH International conference on data science and systems (HPCC/SMARTCITY/DSS)*, 1314-1321. DOI:10.1109/HPCC/SmartCity/DSS.2018.00219
 - Borges, M., Bernardino, J., Pedrosa, I. (2021). Data-driven decision making strategies applied to marketing. *PROCEEDINGS OF 2021 16TH Iberian Conference on Information Systems and Technologies (CISTI'2021)*. DOI: 10.23919/CISTI52073.2021.9476506
 - Boroushaki, S., & Malczewski, J. (2008). Implementing an extension of the analytical hierarchy process using ordered weighted averaging operators with fuzzy quantifiers in ArcGIS. *Computers & Geosciences*, 34(4), 399–

410. DOI: 10.1016/j.cageo.2007.04.003

- Brachman, R.J., Anand, T. (1994). The process of knowledge discovery in databases: a first sketch. *AAAIWS'94: Proceedings of the 3rd International Conference on Knowledge Discovery and Data Mining*, 1–11
- Bratucu, G. & Cismaru, L. (2015). Developing a business intelligence planning tool for managing ecotourism destinations based on indicators existing at EU level. *International Multidisciplinary Scientific GeoConference-SGEM*, 181-188. DOI:10.5593/SGEM2015/B53/S21.023.
- Braud, A., Fromentoux, G., Radier, B., Le Grand, O. (2021). The Road to European Digital Sovereignty with GAIA-X and IDSA. *IEEE Network*, 35(2), 4-5. DOI: 10.1109/MNET.2021.9387709.
- Bridges, C.C. (1966). Hierarchical Clustering Analysis, *Psychological Reports*, 18(3), 851-&. DOI: 10.2466/pr0.1966.18.3.851
- Buhalis, D., Amaranggana, A. (2014). Smart tourism destinations. *Information and communication technologies in tourism 2014*, 553-564. DOI:10.1007/978-3-319-03973-2_40
- Buhalis, D., Sinarta, Y. (2019). Real-time co-creation and nowness service: lessons from tourism and hospitality. *Journal of Travel & Tourism Marketing*, 36(5), 563-582. DOI:10.1080/10548408.2019.1592059
- Buhalis, D. (2020). Technology in tourism-from information communication technologies to eTourism and smart tourism towards ambient intelligence tourism: a perspective article. *Tourism Review*, 75(1), 267-272. DOI:10.1108/TR-06-2019-0258
- Buhalis, D., Leung, D., Lin, M.C. (2023). Metaverse as a disruptive technology revolutionising tourism management and marketing. *Tourism Management*, 97. DOI: 10.1016/j.tourman.2023.104724.
- Bustamante, A., Sebastia, L., Onaindia, E. (2020). BITOUR: A Business Intelligence Platform for Tourism Analysis. *ISPRS International Journal of Geo-information*, 9(11). DOI:10.3390/ijgi9110671
- Camilleri, M.A. (2020). The use of data-driven technologies in tourism marketing. *Entrepreneurship, Innovation and Inequality: Exploring Territorial Dynamics and Development*, 182-194. DOI:10.4324/9780429292583-11
- Callon, M.; Courtial, J. P.; Turner, W. A.; Bauin, S. (1983). From translations to problematic networks: An introduction to co-word analysis. *Social Science Information*, 22 (2), 191-235. DOI:10.1177/053901883022002003
- Caro, L.M. & García, J.A.M. (2008). Developing a multidimensional and hierarchical service quality model for the travel agency industry. *Tourism Management*, 29 (4), 706-720. DOI: 10.1016/j.tourman.2007.07.014
- Carrasco, R. A., Forero, L. N., López, S. X., Herrera-Viedma, E., & Porcel, C. (2018). Using the AHP model to improve the measurement of satisfaction in the ICT sector. In *Frontiers in artificial intelligence and*

applications: Vol. 303. New trends in intelligent software methodologies, tools and techniques (pp. 299–311). DOI: 10.3233/978-1-61499-900-3-299

- Carrasco-Aguilar, A., Galan, JJ., Carrasco, RA. (2022). Obamacare: A bibliometric perspective. *Frontiers in Public Health*, 10. DOI:10.3389/fpubh.2022.9790
- Carvalho, I., Ivanov, S. (2023). ChatGPT for tourism: applications, benefits and risks. *Tourism Review*. DOI:10.1108/TR-02-2023-0088
- Celdran-Bernabeu, MA, Mazon, JN, Sanchez, DG. (2018). Open Data and tourism. Implications for tourism management in Smart Cities and Smart Tourism Destinations. *Investigaciones Turísticas*, 15, 49-78. DOI:10.14198/INTURI2018.15.03
- Chang, HH., Chiang, CC. (2022). Is virtual reality technology an effective tool for tourism destination marketing? A flow perspective. *Journal of Hospitality and Tourism Technology*, 13(3), 427-440. DOI:10.1108/JHTT-03-2021-0076
- Chang, Y. (2020). Tourist hotspot clustering and explanatory analysis using multidimensional data sources: case study of Jeonju-city, South Korea. *International Journal of Transportation*, 8(2), 23-38. DOI:10.33832/ijt.2020.8.2.03
- Chantre-Astaiza, A., Fuentes-Moraleda, L., Munoz-Mazon, A., Ramírez-González, G. (2019). Science Mapping of Tourist Mobility 1980-2019. Technological Advancements in the Collection of the Data for Tourist Traceability. *Sustainability*, 11 (17). DOI:10.3390/su11174738
- ChatGPT (2022). Introducing ChatGPT. <https://openai.com/blog/chatgpt>. Accessed 28, 09, 2023
- Chen, J., Yan, J. (2016). Location Optimization Model for Hotels with Clustering Algorithm. *Proceedings of the 2016 International Conference on Management Science and Innovative Education*, 60, 69-71. DOI: 10.2991/msie-16.2016.10
- Cheng, MM., Edwards, D., Darcy, S., Redfern, K. (2018). A Tri-Method Approach to a Review of Adventure Tourism Literature: Bibliometric Analysis, Content Analysis, and a Quantitative Systematic Literature Review. *Journal of Hospitality & Tourism Research*, 42 (6), 997-1020. DOI:10.1177/1096348016640588
- Chiu, WS., Leng, HK. (2017). Let's go cycling: an analysis of tourists' experience on online user-generated content. *International Journal of Tourism Cities*, 3(1), 30-42. DOI: 10.1108/IJTC-10-2016-0045
- Christodoulou, E., Gregoriades, A., Pampaka, M., Herodotou, H. (2020). Combination of Topic Modelling and Decision Tree Classification for Tourist Destination Marketing. *Advanced Information Systems Engineering Workshops*, 382, 95-108. DOI: 10.1007/978-3-030-49165-9_9

- Cillo, V., Rialti, R., Del Giudice, M., Usai, A (2019). Niche tourism destinations' online reputation management and competitiveness in big data era: evidence from three Italian cases. *Current Issues in Tourism*, 24(2), 177-191. DOI:10.1080/13683500.2019.1608918
- Cobo, M. J.; López-Herrera, A. G.; Herrera-Viedma, E.; Herrera, F. (2011). An approach for detecting, quantifying, and visualizing the evolution of a research field: A practical application to the Fuzzy Sets Theory field. *Journal of Informetrics*, 5 (1), 146-166. DOI: 10.1016/j.joi.2010.10.002
- Cobo, M. J.; López-Herrera, A. G.; Herrera-Viedma, E.; Herrera, F. (2012). SciMAT: A new science mapping analysis software tool. *Journal of the American Society for Information Science and Technology*, 63 (8), 1609-1630. DOI:10.1002/asi.22688
- Cobo, M. J.; Martínez, M. A.; Gutiérrez-Salcedo, M.; Fujita, H.; Herrera-Viedma, E. (2015). 25 years at Knowledge-Based Systems: A bibliometric analysis. *Knowledge-Based Systems*, 80, 3-13. DOI: 10.1016/j.knosys.2014.12.035
- Comerio, N., Strozzi, F. (2019). Tourism and its economic impact: A literature review using bibliometric tools. *Tourism Economics*, 25 (1), 109-131. DOI:10.1177/1354816618793762
- Coulter, N.; Monarch, I.; Konda, S. (1998). Software engineering as seen through its research literature: A study in co-word analysis. *Journal of the American Society for Information Science*, 49 (13), 1206-1223. DOI:10.1002/(SICI)1097-4571(1998)49:133.3.CO;2-6
- Crouch, G.I. (2011). Destination competitiveness: An analysis of determinant attributes. *Journal of Travel Research*, 50 (1) ,27-45. DOI: 10.1177/0047287510362776
- Crouch, G.I., Ritchie, B.J.R. (2012). Destination competitiveness and its implications for host-community QOL. M. Uysal, R. Perdue, M.J. Sirgy (Eds.), *Handbook of tourism and quality-of-life research: Enhancing the lives of tourists and residents of host communities*, Springer, Dordrecht, Netherlands, 491-514. DOI:10.1007/978-94-007-2288-0_29
- Curiel, JD., Jalón, MLD., Herráez, BR., Antonovica, A.(2017).Smart Tourism Destination in Madrid. *Sustainable Smart Cities: creating spaces for technological, social and business development*, 101-114. DOI:10.1007/978-3-319-40895-8_8
- DATES (2023). <https://www.tourismdataspace-csa.eu/>. Accessed 01, 09, 2023
- Davenport, TH. (2015). How strategists use “big data” to support internal business decisions, discovery and production. *Strategy & Leadership*, 42 (4), pp. 45-50. DOI: 10.1108/SL-05-2014-0034
- De la Hoz-Correa, A., Muñoz-Leiva, F., Bakucz, M (2018). Past themes and

future trends in medical tourism research: A co-word analysis. *Tourism Management*, 65, 200-211. DOI: 10.1016/j.tourman.2017.10.001

- De Marchi, D., Becarelli, R., Di Sarli, L. (2022). Tourism Sustainability Index: Measuring Tourism Sustainability Based on the ETIS Toolkit, by Exploring Tourist Satisfaction via Sentiment Analysis. *Sustainability*, 14(13). DOI:10.3390/su14138049
- Del Chiappa, G., & Baggio, R. (2015). Knowledge transfer in smart tourism destinations: Analyzing the effects of a network structure. *Journal of Destination Marketing & Management*, 4 (3), 145-140. DOI: 10.1016/j.jdmm.2015.02.001
- Devlin J, Chang MW, Lee K, Toutanova K. (2018). BERT: pre-training of deep bidirectional transformers for language understanding. *Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies (NAACL HLT 2019–2019)*,1, 4171–4186. DOI:10.18653/v1/N19-1423
- Dini, M., Splendiani, S., Bravi, L. & Pencarelli, T. (2022). In-store technologies to improve customer experience and interaction: an exploratory investigation in Italian travel agencies (2022). *TQM JOURNAL*, 34(7), 94-114. DOI: 10.1108/TQM-08-2021-0230
- Dong, Y., Liu, Y., Liang, H., Chiclana, F., & Herrera-Viedma, E. (2018). Strategic weight manipulation in multiple attribute decision making. *Omega: The International Journal of Management Science*, 75, 154–164. DOI: 10.1016/j.omega.2017.02.008
- Dwivedi, YK., Pandey, N., Currie, W., Micu, A. (2023). Leveraging ChatGPT and other generative artificial intelligence (AI)-based applications in the hospitality and tourism industry: practices, challenges and research agenda. *International journal of contemporary hospitality management*, 36(1), 1-12. DOI:10.1108/IJCHM-05-2023-0686
- Dwyer, L. & Kim, C. (2003). Destination competitiveness: Determinants and indicators. *Current Issues in Tourism*, 6 (5), 369-414. DOI: 10.1080/13683500308667962
- Enlace Tripadvisor (2023). https://www.tripadvisor.com/ShowUserReviews-g187791-r293060413-Rome_Lazio.html. Accessed 01, 09, 2023
- Esmaili, L., Mardani, S., Golpayegani, SAH., Madar, ZZ. (2020). A novel tourism recommender system in the context of social commerce. *Experts Systems with Applications*, 149. DOI: 10.1016/j.eswa.2020.113301
- Essien, A., Chukwukelu, G. (2022). Deep learning in hospitality and tourism: a research framework agenda for future research. *International Journal of Contemporary Hospitality Management*, 34(12), 4480-4515. DOI: 10.1108/IJCHM-09-2021-1176
- Fayyad, U., PiatetskyShapiro, G., Smyth, P. (1996). The KDD process for extracting useful knowledge from volumes of data. *Communications of*

the ACM, 39(11), 27-34. DOI:10.1145/240455.240464

- Feng, YL., Zhang, HQ., Peng, CT. (2022). Fast Recommendation Method of Personalized Tourism Big Data Information Based on Improved Clustering Algorithm. Book Series Lecture Notes of the Institute for Computer Sciences Social Informatics and Telecommunications Engineering, 417, 284-296. DOI:10.1007/978-3-030-94554-1_23
- Fernández-Avilés, G., Montero Lorenzo, J.M. (2024). Ciencia de datos con R. McGraw Hill.
- Florido-Benitez, L. (2022). The impact of tourism promotion in tourist destinations: a bibliometric study. International Journal of Tourism Cities. DOI:10.1108/IJTC-09-2021-0191
- Fortunato, S. & Barthélemy, M. (2007). Resolution limit in community detection. Proceedings of the National Academy of Sciences of the United States of America, 104 (1), 36-41. DOI: 10.1073/pnas.0605965104
- Fuchs, M., Abadzhiev, A., Svensson, B, Höpken, W. & Lexhagen, M. (2013): Knowledge Destination framework for tourism sustainability – a Business Intelligence application from Sweden. Tourism - An Interdisciplinary Journal, 61(2), 121-148.
- Gaia-X (2023). Gaia-X Hub: <https://www.gaiax.es/>. Accessed 23, 05, 2023
- Gajdosik, T. (2019). Towards a Conceptual Model of Intelligent Information System for Smart Tourism Destinations. Software Engineering and Algorithms in Intelligent Systems, 763, 66-74. DOI: 10.1007/978-3-319-91186-1_8
- Galán JJ, Carrasco RA, Latorre A. (2022). Military applications of machine learning: a bibliometric perspective. Mathematics, 10(9). DOI: 10.3390/math10091397
- Ghozali, A, Syafitri, ED (2020). Spatial patterns of hotel's occupation rate. International Conference on Innovation in Engineering and Vocational Education 2019 (ICIEVE 2019), PTS 1-4. DOI: 10.1088/1757-899X/830/2/022081
- Gil-Lafuente, A. M., Merigó, J. M., Vizueté, E. (2014). Analysis of luxury resort hotels by using the fuzzy analytic hierarchy process and the fuzzy Delphi method. Economic Research-Ekonomska Istraživanja, 27(1), 244-266. DOI: 10.1080/1331677X.2014.952106
- Giray, L (2023). Prompt Engineering with ChatGPT: A Guide for Academic Writers. Annals of Biomedical Engineering. DOI: 10.1007/s10439-023-03272-4
- Gómez, SG., Gil, AR. (2020). Knowledge management and value creation through big data in the tourism sector. Revista internacional de organizaciones, 24, 211-239

- Gon, M. (2021). Local experiences on Instagram: Social media data as source of evidence for experience design. *Journal of Destination Marketing & Management*, 19. DOI: 10.1016/j.jdmm.2020.100435
- Google Maps (2024): <https://www.google.com/maps>. Accessed 01, 03, 2024
- Gretzel, U., Koo, C., Lamsfus, C. & Werthner, H. (2015). Conceptual Foundations for Understanding Smart Tourism Ecosystems. *Computers in Human Behavior*, 50, 558-563. DOI: 10.1016/j.chb.2015.03.043
- Gunter, U., Onder, I., Gindl, S. (2019). Exploring the predictive ability of LIKES of posts on the Facebook pages of four major city DMOs in Austria. *Tourism Economics*, 25(3), 375-401. DOI:10.1177/1354816618793765
- Gupta, S., Modgil, S., Lee, CK., Sivarajah, U. (2022). The future is yesterday: Use of AI-driven facial recognition to enhance value in the travel and tourism industry. *Information Systems Frontiers*. DOI:10.1007/s10796-022-10271-8
- Hall, CM. (2011). Publish and perish? Bibliometric analysis, journal ranking and the assessment of research quality in tourism. *Tourism Management*, 32, 16-27. DOI: 10.1016/j.tourman.2010.07.001
- Harby, A., Zulkernine, F. (2022). From Data Warehouse to Lakehouse: A Comparative Review. *Proceedings - 2022 IEEE International Conference on Big Data, Big Data 2022*, 389 - 395. DOI: 10.1109/BigData55660.2022.10020719
- Hayashi, R., Matsuda, Y., Fujimoto, M., Suwa, H., Yasumoto, K.(2021). Multimodal Tourists' Emotion and Satisfaction Estimation Considering Weather Conditions and Analysis of Feature Importance. 13TH International Conference on Mobile Computing and Ubiquitous Network (ICMU2021). DOI: 10.23919/ICMU50196.2021.9638938
- Hclust: Hierarchical Clustering (2023). <https://www.rdocumentation.org/packages/stats/versions/3.6.2/topics/hclust>. Accessed 12, 07, 2023
- He, Q. (1999). Knowledge discovery through co-word analysis. *Library Trends*, 48 (1), 133-159. DOI:10.2478/cait-2014-0051
- He, S. (2022). Research on Tourism Route Recommendation Strategy Based on Convolutional Neural Network and Collaborative Filtering Algorithm. *Security and Communications Networks*, 2022. DOI:10.1155/2022/4659567
- Henrickson, L, Merono-Penuela, A. (2023) Prompting meaning: a hermeneutic approach to optimising prompt engineering with ChatGPT. *AI & Society*. DOI:10.1007/s00146-023-01752-8
- Hirsch, J. E. (2005). An index to quantify an individuals scientific research output. *Proceedings of the National Academy of Sciences*, 102, 16569-16572. DOI:10.1073/pnas.0507655102
- Hoang, HTT., Truong, QH., Nguyen, AT., Hens, L. (2018). Multicriteria

Evaluation of Tourism Potential in the Central Highlands of Vietnam: Combining Geographic Information System (GIS), Analytic Hierarchy Process (AHP) and Principal Component Analysis (PCA). *Sustainability*, 10(9). DOI: 10.3390/su10093097

- Hu, HM., Li, CH. (2023). Smart tourism products and services design based on user experience under the background of big data. *Soft Computing*, 27(17), 12711-12724. DOI: 10.1007/s00500-023-08851-0
- Huang, M. H.; Chang, C. P. (2014). Detecting research fronts in OLED field using bibliographic coupling with sliding window. *Scientometrics*, 98 (3), 1721-1744. DOI:10.1007/s11192-013-1126-1
- INE (2023). <https://www.ine.es>. Accessed 12, 07, 2023.
- Inoue, M., Yamamoto, K. (2013). Method for Evaluating the Location of Tourist-Related Public Facilities Using Genetic Algorithms and GIS. *Journal of Communication and Computer* 10 (2013) 496-512
- Jain AK., Murty MN., Flynn PJ. (1999). Data clustering: a review. *ACM Comput. Surv.*, 31 (3) (1999), pp. 264-323. DOI:10.1145/331499.331504
- Jansson, J., Johansson, O., Roshan, M. (2022). Initiating a Smart Tourism Ecosystem: A Public Actor Perspective. *Proceedings of the 55th Hawaii International Conference on System Sciences*. DOI:10.24251/HICSS.2022.335
- Jeong-hwan, J., Geum, Y. (2016). BSC-AHP Framework for the Partner Selection in Open Innovation Practice. *Journal of Technology Innovation*, 24(1), 25-48. DOI: 10.14383/SIME.2016.24.1.25
- Jinwon, K., Seongsoo, J., Sanghoon, K., SeungHyun, J. (2020). Why are hotel room prices different? Exploring spatially varying relationships between room price and hotel attributes. *Journal of Business Research*, 107, 118-129. DOI: 10.1016/j.jbusres.2018.09.006
- Julashokri, M., Leonardi, SM., Seabra, P. (2022). Tourism Integrated Recommender System: Setubal Peninsula Case Study. *ICEIS: Proceedings of the 24th International Conference on Enterprise Information Systems*, 1, 623-627. DOI:10.5220/0011102100003179
- Kanungo, T, Mount, DM, Netanyahu, NS, Piatko, CD, Silverman R, Wu, AY. (2002). An Efficient k-Means Clustering Algorithm: Analysis and Implementation. *IEEE Transactions on pattern analysis and machine intelligence*, 24 (7), 881-892. DOI10.1109/TPAMI.2002.1017616
- Kaplan R.S. & Norton D.P. (1992). The balanced scorecard measures that drive performance. *Harvard Bus Rev* 70: 71-79
- Kayakus, M. (2022). Estimating the Changes in the Number of Visitors on the Websites of the Tourism Agencies in the COVID-19 Process by Machine Learning Methods. *Sosyoekonomi*, 30 (53), 11-26. DOI:10.17233/sosyoekonomi.2022.03.01
- Khan, K., Rehman, SU., Aziz, K., Fong, S., Sarasvady, S. (2014). DBSCAN: Past, Present and Future. 2014 FIFTH International Conference on the

Applications of Digital Information and Web Technologies (ICADIWT), 232-238. DOI: 10.1109/ICADIWT.2014.6814687

- Kim, K., Park, OJ., Yun, S., Yun, H. (2017). What makes tourists feel negatively about tourism destinations? Application of hybrid text mining methodology to smart destination management. *Technological Forecasting and Social Change*, 123, 362-369. DOI: 10.1016/j.techfore.2017.01.001
- Kim, JM., Han, J. (2022). Impact of the length of stay at hotels on online reviews. *International Journal of Contemporary Hospitality Management*, 34(4), 1249-1269. DOI:10.1108/IJCHM-05-2021-0659
- Kirtil, IG., Askun, V. (2021). Artificial intelligence in tourism: a review and bibliometrics research. *Advances in Hospitality and Tourism Research-AHTR*, 9 (1), 205-233. DOI:10.30519/ahtr.801690
- Kohli, AK., & Jaworski, BJ. (1990). Market orientation: the construct, research propositions, and managerial implications. *Journal of marketing*, 54(2),1-18, 1990. DOI: 10.2307/1251866
- Kohli, AK., Jaworski, BJ. (2023). Market driving: Some directions for future research. *Industrial Marketing Management*, 113, 348-351. DOI:10.1016/j.indmarman.2023.06.012
- Korzynski, P., Mazurek, G., Krzyrkowska, P., Kurasinski, A. (2023). Artificial intelligence prompt engineering as a new digital competence: Analysis of generative AI technologies such as ChatGPT. *Entrepreneurial Business and Economics Review*, 11(3), 25-37. DOI:10.15678/EBER.2023.110302
- Koseoglu, MA., Rahimi, R., Okumus, F., Liu, J. (2016). Bibliometric studies in tourism. *Annals of Tourism Research*, 61, 180-198. DOI: 10.1016/j.annals.2016.10.006
- Krishna, RJ ; Chaudhry, Y ; Sharma, DP (2018). Analysis of Community Detection Algorithms. *Proceedings of the 2018 second international conference on inventive communication and computational technologies (ICICCT)*, 669-674. DOI: 10.1109/ICICCT.2018.8473246
- Krsak, B., Sidor, C., Strba, L. (2018). Open structured databases' use for spatial econometrics within data-driven governance and business intelligence. *Advances and Trends in Geodesy, Cartography and Geoinformatics*, 183-190. DOI: 10.1201/9780429505645.
- Kumar, V., & Reinartz, W. (2016). Creating enduring customer value. *Journal of Marketing*, 80(6), 36– 68. DOI: 10.1509/jm.15.0414
- Law, R., Mok, H., Goh, C. (2007). Data mining in tourism demand analysis: A retrospective analysis. *Advanced Data Mining and Applications, proceedings*, 4631, 508-515. DOI:10.1007/978-3-540-73871-8_47
- Law, R. , Li, G., Fong, DKC., Han, X. (2019). Tourism demand forecasting:

A deep learning approach. *Annals of Tourism Research*, 75, 410-423. DOI: 10.1016/j.annals.2019.01.014

- Li, G., Jiao, XY. (2020). Tourism forecasting research: a perspective article. *Tourism review*, 75(1), 263-266. DOI: 10.1108/TR-09-2019-0382
- Li, HX.; Liu, Y.; Tan, CW.; Hu, F. (2020). Comprehending customer satisfaction with hotels Data analysis of consumer-generated reviews. *International Journal of Contemporary Hospitality Management*, 32(5), 1713-1735. DOI: 10.1108/IJCHM-06-2019-0581
- Li, HY., Wang, Q., Zhang, LY., Cai, DT. (2022). Big Data in China Tourism Research: A Systematic Review of Publications from English Journals. *Journal of China Tourism Research*, 18(3), 453-471. DOI:10.1080/19388160.2022.2049943
- Li, JJ., Xu, LZ., Tang, L., Wang, SY., Li, L. (2018). Big data in tourism research: A literature review. *Tourism Management*, 68, 301-323. DOI: 10.1016/j.tourman.2018.03.009
- Lin, C., Kou, G., Peng, Y., & Alsaadi, F. E. (2020). Two-stage prioritization procedure for multiplicative AHP-group decision making. *Technological and Economic Development of Economy*, 26(2), 525–545. DOI: 10.3846/tede.2020.12037
- Liu, G.; Hu, J. M.; Wang, H. L. (2012). A co-word analysis of digital library field in China. *Scientometrics*, 91 (1), 203-217. DOI:10.1007/s11192-011-0586-4
- Liu, T., Zhang, Y., Zhang, H., Yang, XP. (2021). A Methodological Workflow for Deriving the Association of Tourist Destinations Based on Online Travel Reviews: A Case Study of Yunnan Province, China. *Sustainability*, 13(9). DOI:10.3390/su13094720
- Llorens, M., Carrasco, R., Bueno, I., Herrera-Viedma, E. & Morente-Molinera, JA (2022). Multiple criteria approach applied to digital transformation in fashion stores: the case of physical retailers in Spain. *Technological and Economic Development of Economy*, 28(2), 500-530. DOI:10.3846/tede.2022.16553
- Loureiro, SMC., Guerreiro, J., Han, H. (2022). Past, present, and future of pro-environmental behavior in tourism and hospitality: a text-mining approach. *Journal of sustainable tourism*, 30(1), 258-278. DOI:10.1080/09669582.2021.1875477
- Lv, H., Shi, S., Gursoy, D. (2021). A look back and a leap forward: a review and synthesis of big data and artificial intelligence literature in hospitality and tourism. *Journal of Hospitality Marketing & Management*, 31 (2). DOI:10.1080/19368623.2021.1937434
- Ma, ZC. (2022). Construction of Tourism Management Engineering Based on Data Mining Technology. *Journal of Electrical and Computer Engineering*, 2022. DOI:10.1155/2022/1982462

- Machado, IA., Costa, C., Santos, MY. (2022). Advancing Data Architectures with Data Mesh Implementations. *Intelligent information systems (Caise forum 2022)*, 453, 10-18. DOI:10.1007/978-3-031-07481-3_2
- Maracajá, KFB., Schramm, VB., Schramm, F., Valduga, V. (2021). A multicriteria model for evaluation of Brazilian wineries from a tourism destination perspective. *International Journal of Wine Business Research*, 34 (1), 52-68. DOI:10.1108/IJWBR-12-2020-0057
- Mariani, M., Baggio, R., Fuchs, M., Hoepken, W. (2018). Business intelligence and big data in hospitality and tourism: a systematic literature review. *International Journal of Contemporary Hospitality Management*, 30 (12), 3514-3554. DOI:10.1108/IJCHM-07-2017-0461
- Mariani, M., Baggio, R. (2022). Big data and analytics in hospitality and tourism: a systematic literature review. *International Journal of Contemporary Hospitality Management*, 34(1), 231-278. DOI:10.1108/IJCHM-03-2021-0301
- Marine-Roig, E., Clave, SA. (2015). Tourism analytics with massive user-generated content: A case study of Barcelona. *Journal of Destination Marketing & Management*, 4(3), 162-172. DOI: 10.1016/j.jdmm.2015.06.004
- Martínez, EO., Devesa, MJS. (2013). Communication and tourists awareness about World Heritage sites in Madrid Regional Area. *Cuadernos de Turismo*, 31, 263-288.
- Martínez-Martínez, A, Cegarra-Navarro, JG, de Valon, T (2022). Impacts and Implications for Advancing in Environmental Knowledge in Hospitality Industry in COVID Society. *Journal of the Knowledge Economy*. DOI:10.1007/s13132-022-00910-5
- Matplotlib (2022). <https://matplotlib.org/>. Accessed 01, 03, 2022.
- Metaxas, T., Juarez, L., Gavriilidis, G. (2021). Planning and Marketing the City for Sustainability: The Madrid Nuevo Norte Project. *Sustainability*, 13(4). DOI:10.3390/su13042094
- Miah, SJ.; Vu, HQ., Gammack, J., McGrath, M (2017). A Big Data Analytics Method for Tourist Behaviour Analysis. *Information & Management*, 54(6), 771-785. DOI:10.1016/j.im.2016.11.011
- Minhas, J., & Sindakis, S. (2021). Implications of Social Cohesion in Entrepreneurial Collaboration: a Systematic Literature Review. *Journal of the Knowledge Economy*, 1-32. DOI:10.1007/s13132-021-00810-0
- Modak, M., Pathak, K., & Ghosh, K. K. (2017). Performance evaluation of outsourcing decision using a BSC and Fuzzy AHP approach: A case of the Indian coal mining organization. *Resources Policy*, 52, 181–191. DOI: 10.1016/j.resourpol.2017.03.002
- Moreno, C.; Carrasco, RA; Herrera-Viedma (2019). Data and Artificial Intelligence Strategy: A Conceptual Enterprise Big Data Cloud Architecture to Enable Market-Oriented Organisations. *International Journal of Interactive Multimedia and Artificial Intelligence*, 5(6), 7-14.

DOI:10.9781/ijimai.2019.06.003

- Mukhopadhyay, S., Pandey, R., Rishi, B. (2022). Electronic word of mouth (eWOM) research - a comparative bibliometric analysis and future research insight. *Journal of Hospitality and Tourism Insights*. DOI:10.1108/JHTI-07-2021-0174
- Mulet-Forteza, C., Genovart-Balaguer, J., Mauleon-Méndez, E., Merigo, J.M. (2019). A bibliometric research in the tourism, leisure and hospitality fields. *Journal of Business Research*, 101, 819-827. DOI: 10.1016/j.jbusres.2018.12.002
- Müllner, D. (2011) Modern Hierarchical, Agglomerative Clustering Algorithms. ArXiv Organization, Technical Report. <https://arxiv.org/abs/1109.2378>
- Muritala, BA., Sánchez-Rebull, MV., Hernández-Lara, AB. (2020). A Bibliometric Analysis of Online Reviews Research in Tourism and Hospitality. *Sustainability*, 12 (23). DOI:10.3390/su12239977
- Murtagh, F., Legendre, P. (2014). Ward's hierarchical agglomerative clustering method: which algorithms implement Ward's criterion? *Journal of Classification*, 31, 274--295. DOI: 10.1007/s00357-014-9161-z.
- Narver J.C., Slater S.F. (1990). The effect of a market orientation on business profitability. *Journal of Marketing*, 54 (4), 20-35. DOI: 10.2307/1251757
- Narver, JC., Slater, SF., MacLachlan, DL. (2004). Responsive and proactive market orientation and new-product success. *Journal of Product Innovation Management*, 21(5), 334-347. DOI:10.1111/j.0737-6782.2004.00086.x
- Navarro, JR., Rubio, JQ (2000). DATATUR: Tourism statistics information system - The experience of Spain. *Information and Communication Technologies in Tourism 2000*, 126-146. DOI:10.1007/978-3-7091-6291-0_12
- Nave, M.; Rita, P., Guerreiro, J. (2018). A decision support system framework to track consumer sentiments in social media. *Journal of Hospitality Marketing & Management*, 27(6), 693-700. DOI:10.1080/19368623.2018.1435327
- Nilashi, M., Ahani, A., Esfahani, MD., Yadegaridehkordi, E., Samad, S., Ibrahim, O., Sharef, NM., Akbari, E. (2019). Preference learning for eco-friendly hotels recommendation: A multi-criteria collaborative filtering approach. *Journal of Cleaner Production*, 215, 767-783. DOI:10.1016/j.jclepro.2019.01.012
- Nomecalles (2023). <https://gestion.comunidad.madrid/nomecalles/DescargaBDTCorte.icm>. Accessed, 12, 07, 2023
- Noyons, E. C. M.; Moed, H. F.; Luwel, M. (1999). Combining mapping and citation analysis for evaluative bibliometric purposes: A bibliometric

- study. *Journal of the American Society for Information Science*, 50 (2), 115-131. DOI:10.1002/(SICI)1097-4571(1999)50:2<115:AID-ASI3>3.0.CO;2-J
- Nusair, K., Butt, I., Nikhashemi, SR. (2019). A bibliometric analysis of social media in hospitality and tourism research. *International Journal of Contemporary Hospitality Management*, 31 (7), 2691-2719. DOI:10.1108/IJCHM-06-2018-0489
 - Nyanga, C., Pansiri, J., Chatibura, D. (2020). Enhancing competitiveness in the tourism industry through the use of business intelligence: a literature review. *Journal of Tourism Futures*, 6(2), 139-151. DOI:10.1108/JTF-11-2018-0069
 - Okumus, B., Koseoglu, MA., Ma, F. (2018). Food and gastronomy research in tourism and hospitality: A bibliometric analysis. *International Journal of Hospitality Management*, 73, 64-74. DOI: 10.1016/j.ijhm.2018.01.020
 - Omerzel, DG. (2016). A systematic review of research on innovation in hospitality and tourism. *International Journal of Contemporary Hospitality Management*, 28 (3), 516-558. DOI:10.1108/IJCHM-10-2014-0510
 - Openrouteservice (2023). <https://openrouteservice.org/>. Accessed 12, 07, 2023
 - Padma, P., Ahn, J. (2020). Guest satisfaction & dissatisfaction in luxury hotels: An application of big data. *International Journal of Hospitality Management*, 84. DOI:10.1016/j.ijhm.2019.102318
 - Park, SH, Chang YH. (2021). Applying an Explanatory Analysis for Hotspot Identification in Tourist Attraction. 2021 IEEE REGION 10 SYMPOSIUM(TENSYP). DOI: 10.1109/TENSYP52854.2021.9550903
 - Parra-López, E. & Oreja-Rodríguez, J. (2014). Evaluation of the competitiveness of tourist zones of an island destination: an application of a Many-Facet Rasch Model (MFRM). *Journal of Destination Marketing & Management*, 3 (2), 114-121
 - Patel, H, Thakkar, A, Pandya, M, Makwana, K. (2018) Neural network with deep learning architectures. *Journal of Information & Optimization Sciences*, 39(1), 31-38. DOI: 10.1080/02522667.2017.1372908
 - Pencarelli, T., Bravi, L., Dini, M. & Splendiani, S. (2020). Enhancing customer experience through technological innovation in traditional travel agencies: evidences from Italy. *Current Issues in Tourism*, 24 (5),590-596. DOI: 10.1080/13683500.2020.173454
 - Pereira, RC & Vanitha T (2015), Web Scraping of Social Networks, *International Journal of Innovative Research in Computer and Communication Engineering*, 3 (7), 237-240. DOI:10.14195/0870-4147_46_13

- Perry, C. (1929). "The neighborhood unit": Regional New York and its environs. New York, United States: Routledge.
- Peters, HPF., & Van Raan, AFJ. (1993). Co-word-based science maps of chemical engineering. Part I: Representations by direct multidimensional scaling. *Research Policy*, 22(1), 23–45. DOI: 10.1016/0048-7333(93)90031-C
- Pine, B. L., & Gilmore, J. H. (1999). *The experience economy: Work is theater & every business a stage*. Harvard Business School Press
- Plataforma inteligente de destinos. <https://www.segittur.es/plataforma-inteligente/proyectos-plataforma-inteligente/plataforma-inteligente-de-destinos/>. Accessed 01, 09, 2023
- Portal datos Ayuntamiento de Madrid (2023). <https://datos.madrid.es/portal/site/egob> Accessed 12, 07, 2023
- Puhretmair, F., Rumetshofer, H., Schaumlechner, E. (2002). Extended decision making in tourism information systems. *Book Series Lecture Notes in Computer Science*, 2455, 57-66. DOI:10.1007/3-540-45705-4_7.
- R Project (2023). <https://www.r-project.org/> Accessed 01, 07, 2023.
- Radford, A, Narasimhan K, Salimans T, Sutskever I. (2023). Improving Language Understanding by Generative Pre-Training. https://cdn.openai.com/research-covers/language-unsupervised/language_understanding_paper.pdf Accessed, 28, 09, 2023.
- Rahmadian, E., Feitosa, D., Zwitter, A (2022). A systematic literature review on the use of big data for sustainable tourism. *Current Issues in Tourism*, 25 (11),1711-1730. DOI:10.1080/13683500.2021.1974358
- Raju, R., Mital, R., Finkelsztein, D. (2018). Data Lake Architecture for Air Traffic Management. 2018 IEEE/AIAA 37TH Digital Avionics Systems Conference (DASC), 604-609. DOI: 10.1109/DASC.2018.8569361
- Ramos, CMQ., Martins, DJ., Serra, F., Lam, R. (2017). Framework for a Hospitality Big Data Warehouse: The Implementation of an Efficient Hospitality Business Intelligence System. *International Journal of Information Systems in the Service Sector*, 9(2), 27-45. DOI:10.4018/IJISS.2017040102
- Ranganathan, I., Thangamuthu, P., Palanimuthu, S., Balusamy, B. (2020). The growing role of integrated and insightful big and real-time data analytics platforms. *Advances in Computers*, 117, 165-186. DOI:10.1016/bs.adcom.2019.09.009
- Rapert, M., Yarbrough, L (1998). Competing strategically through market orientation. *Journal of Hospital Marketing*, 12 (1). DOI:10.1300/J043v12n01_07
- Rashad, AS. (2022). The Power of Travel Search Data in Forecasting the Tourism Demand in Dubai. *Forecasting*, 4(3), 674-684. DOI:10.3390/forecast4030036
- Ray, A., Bala, PK. (2021). User generated content for exploring factors

- affecting intention to use travel and food delivery services. *International Journal of Hospitality Management*, 92. DOI:10.1016/j.ijhm.2020.102730
- Request (2022). <https://pypi.org/project/requests/>. Accessed 01, 03, 2022
 - Richardson, A. (2010). Using customer journey maps to improve customer experience. *Harvard Business Review*, 15(1), 2-5.
 - Rip, A., Courtial, JP. (1984). Co-word maps of biotechnology: An example of cognitive scientometrics. *Scientometrics*, 6(6), 381–400. DOI: 10.1007/BF02025827
 - Rodríguez, J., Semanjski, I., Gautama, S., Van de Weghe, N., Ochoa, D. (2018). Unsupervised Hierarchical Clustering Approach for Tourism Market Segmentation Based on Crowdsourced Mobile Phone Data. *Sensors*, 18(9). DOI:10.3390/s18092972
 - RStudio (2023). <https://posit.co/downloads/> Accessed 12, 07, 2023.
 - Ruhanen, L., Weiler, B., Moyle, BD., McLennan, CLJ (2015). Trends and patterns in sustainable tourism research: a 25-year bibliometric analysis. *Journal of Sustainable Tourism*, 23 (4), 517-535. DOI:10.1080/09669582.2014.978790
 - Ruiz-Real, JL., Uribe-Toril, J., Valenciano, JD., Gazquez-Abad, JC. (2020). Rural tourism and development: Evolution in Scientific Literature and Trends. *Journal of Hospitality & Tourism Research*. DOI:10.1177/1096348020926538
 - Saaty, T. L. (1980). *The analytic hierarchy process*. McGraw Hill.
 - Saaty, T. L. (2008). Decision making with the analytic hierarchy process. *International Journal of Services Sciences*, 1(1), 83–98.
 - Sahatqija, K., Ajdari, J., Zenuni, X., Raufi, B., Ismaili, F. (2018). Comparison between relational and NOSQL databases. 2018 41ST International convention on information and communication technology, electronics and microelectronics (MIPRO), 216-221. DOI:10.23919/MIPRO.2018.8400041
 - Samara, D., Magnisalis, I., Peristeras, V (2020). Artificial intelligence and big data in tourism: a systematic literature review. *Journal of Hospitality and Tourism Technology*, 11(2), 343-367. DOI:10.1108/JHTT-12-2018-0118
 - Sankaranarayanan, HB., Lalchandani, J. (2017). Passenger reviews reference architecture using big data lakes. *Proceedings of the 7th International Conference Confluence 2017 on Cloud Computing, Data Science and Engineering*, 204-209. DOI: 10.1109/CONFLUENCE.2017.7943150
 - Salehzadeh, R., Khazaei, P., Tabaeian, R., Amani, M., Mortazavi, M. (2017). The impact of internal marketing and market orientation on performance: an empirical study in restaurant industry. *Measuring Business Excellence*, 21(4), 273-290. DOI:10.1108/MBE-02-2016-0009.
 - Salton, G., & McGill, MJ. (1983). *Introduction to modern information*

retrieval. New York: McGraw-Hill

- Saurkar, AV., Pathare, KG., Gode, SA. (2018), An Overview On Web Scraping Techniques And Tools, *International Journal on Future Revolution in Computer Science & Communication Engineering*, 4 , 363-367
- Shafiee, S., Ghatari, AR., Hasanzadeh, A., Jahanyan, S. (2022). Developing a model for smart tourism destinations: an interpretive structural modelling approach. *Information Technology & Tourism*, 24(4), 511-546. DOI:10.1007/s40558-022-00236-7
- Sharma, K., Dhir, S., Ongsakul, V. (2022). Artificial intelligence and hospitality industry: systematic review using TCCM and bibliometric analysis. *Journal for International Business and Entrepreneurship Development*, 14(1), 48-71. DOI:10.1504/JIBED.2022.124245
- Shi, FJ. (2020). Constructing a Secured, Reactive and Scalable Data Platform for a Better Exploitation of Rich Data Assets in the Tourism Industry. *Book Series: Lecture Notes in Computer Science*, 12019, 236-244. DOI: 10.1007/978-3-030-45989-5_19
- Shiny (2023). <https://shiny.rstudio.com/>. Accessed 14, 07, 2023
- Sohrabi, B. ;Vanani, IR., Nasiri, N., Rudd, AG (2020), A predictive model of tourist destinations based on tourists' comments and interests using text analytics, *Tourism Management Perspectives*, 35. DOI: 10.1016/j.tmp.2020.100710
- Solazzo, G., Maruccia, Y., Lorenzo, G., Elia, G., Del Vecchio, P., Ndou, V. (2020). Extracting Insights from Big Social Data for Smarter Tourism Destination Management. *15TH INTERNATIONAL FORUM ON KNOWLEDGE ASSET DYNAMICS (IFKAD 2020)*, 1766-1782. DOI:10.1108/MBE-11-2020-0156
- Sorensen, H. (2009). Why competitors matter for market orientation. *European Journal of Marketing*, 43(5/6), 735-761. DOI:10.1108/03090560910947025
- Sternitzke, C.; Bergmann, I. (2009). Similarity measures for document mapping: A comparative study on the level of an individual scientist. *Scientometrics*, 78 (1), 113-130. DOI:10.1007/s11192-007-1961-z
- Svensson, RB., Taghavianfar, M. (2020). Toward Becoming a Data-Driven Organization: Challenges and Benefits. *Research Challenges in Information Science (RCIS 2020)*, 385, 3-19. DOI: 10.1007/978-3-030-50316-1_1
- Tijssen, RJW & Van Raan, AFJ. (1989). Mapping co-word structures: A comparison of multidimensional scaling and LEXIMAPPE. *Scientometrics*, 15(3), 283-295. DOI:10.1007/BF02017203
- Tmap (2023). <https://cran.r-project.org/web/packages/tmap/index.html>. Accessed 12, 07, 2023

- Topsakal, O., Akinci TC.(2023). Creating Large Language Model Applications Utilizing LangChain: A Primer on Developing LLM Apps Fast. 5th International Conference on Applied Engineering and Natural Sciences. DOI: 10.59287/icaens.1127
- Tourism Data Space (2023). <https://dsft.modul.ac.at/tourism-data-inventory/>. Accessed 01, 09, 2023
- Tripadvisor (2023). <https://www.tripadvisor.com>. Accessed 01, 09, 2023
- Van der Zee, E., Bertocchi, D. (2018). Finding patterns in urban tourist behaviour: a social network analysis approach based on TripAdvisor reviews. *Information Technology & Tourism*, 20 (1-4), 153-180. DOI: 10.1007/s40558-018-0128-5
- Van der Zee, E., Bertocchi, D., Vanneste, D. (2020). Distribution of tourists within urban heritage destinations: a hot spot/cold spot analysis of TripAdvisor data as support for destination management. *Current Issues in Tourism*, 23 (2), 175-196. DOI10.1080/13683500.2018.1491955
- Van Eck, NJ., Waltman, L., Van den Berg, J., & Kaymak, U. (2006). Visualizing the computational intelligence field. *IEEE Computational Intelligence Magazine*, 1(4), 6–10. DOI: 10.1109/MCI.2006.329702
- Van Eck, NJ. & Waltman, L. (2007). VOS: A new method for visualizing similarities between objects. In H.-J. Lenz & R. Decker (Eds.), *Advances in data analysis: Proceedings of the 30th Annual Conference of the German Classification Society*, 299–306. Springer. DOI:10.1007/978-3-540-70981-7_34
- Van Eck, NJ., & Waltman, L. (2009). How to normalize cooccurrence data? An analysis of some well-known similarity measures. *Journal of the American Society for Information Science and Technology*, 60(8), 1635–1651. DOI: 10.1002/asi.21075
- Van Eck, NJ., Waltman, L., Dekker, R., Van den Berg, J. (2010). A comparison of two techniques for bibliometric mapping: Multidimensional scaling and VOS. *Journal of the American Society for Information Science and Technology*, 61 (12), 2405-2416. DOI: 10.1002/asi.21421
- Van-Raan, A. F. J. (2004). Measuring Science. *Handbook of Quantitative Science and Technology Research*, pp. 19-50. Dordrecht: Springer Netherlands. DOI:10.1007/1-4020-2755-9_2
- Vaswani A, Shazeer N, Parmar N, Uszkoreit J, Jones L, Gomez AN, Kaiser L, Polosukhin I. (2017). Attention is all you need. *NIPS'17: Proceedings of the 31st International Conference on Neural Information Processing Systems*, 6000–6010. DOI: 10.48550/arXiv.1706.03762
- Verhoef, P. C., Kannan, P., & Inman, J. J. (2015). From multi-channel retailing to omni-channel retailing. *Journal of Retailing*, 91(2), 174–181.

DOI: 10.1016/j.jretai.2015.02.00

- Viljoen, K. & Roberts-Lombard, M. (2016). Customer retention strategies for disintermediated travel agents: how to stop customers from migrating to online booking channels. *Journal of Applied Business Research*, 32 (3), 681-694. DOI: 10.19030/jabr.v32i2.9595
- Vosviewer (2022). <https://www.vosviewer.com/>. Accessed 18, 10, 2022
- Wang, D., Li, X., & Li, Y. (2013). China's 'smart tourism destination' initiative: A taste of the service-dominant logic. *Journal of Destination Marketing & Management*, 2(1), 59–61. DOI: 10.1016/j.jdmm.2013.05.004
- World Travel Awards (2023). <https://www.worldtravelawards.com/profile-1655-madrid-convention-bureau>. Accessed 10, 12, 2023
- Wu, XY, Huang, Z., Peng, X., Chen, YR., Liu, Y. (2018). Building a Spatially-Embedded Network of Tourism Hotspots From Geotagged Social Media Data. *IEEE ACCESS*, 6, 21945-21955. DOI:10.1109/ACCESS.2018.2828032
- Wu, D., Song, ZL., Guo, H. (2022). Artificial Intelligence Algorithms in Ice and Snow Tourism Promotion from Digital Technology. *Wireless Communications & Mobile Computing*, 2022. DOI:10.1155/2022/1806611
- Yager, R. R. (1988). On ordered weighted averaging aggregation operators in multicriteria decision making. *IEEE Transactions on Systems, Man, and Cybernetics*, 18(1), 183–190. DOI: 10.1109/21.87068
- Yaghoobi, T. Haddadi, F. (2016). Organizational performance measurement by a framework integrating BSC and AHP. *International Journal of Productivity and Performance Management*, 65(7), 959-976. DOI: 10.1108/IJPPM-01-2015-0001
- Yallop, AC., Gica, OA., Moisescu, OI., Coros, MM., Séraphin, H.(2021).The digital traveller: implications for data ethics and data governance in tourism and hospitality. *Journal of consumer marketing*, 40(2), 155-170. DOI:10.1108/JCM-12-2020-4278
- Yang, B., Madden, M., Kim, J., Jordan, TR. (2012). Geospatial analysis of barrier island beach availability to tourists. *Tourism Management*, 33(4), 840-854. DOI:10.1016/j.tourman.2011.08.013
- Yin, HZ., Wang, WQ. (2018). Mining Geo-social Networks - Spatial Item Recommendation. *Book Series Lecture Notes in Computer Science*, 10837, XX-XXI
- Yu, J., Zheng, X.B., Zhang, X.H., Lu, C.Z. (2012). The method of combined BSC with AHP and its application in the budget cost allocation. *2011 3rd International Conference on Computer Technology and Development (icctd 2011)*, 3, 549-553. DOI: 10.1115/1.859919.paper355
- Yu, QL., Pickering, S., Geng RQ., Yen, DA. (2021). Thanks for the memories: Exploring city tourism experiences via social media reviews. *Tourism Management Perspectives*, 40. DOI: 10.1016/j.tmp.2021.100851

- Yu, ZW., Zhu, XJ., Wong, HS., You, JE., Zhang, J., Han, GQ. (2017). Distribution-Based Cluster Structure Selection. *IEEE Transactions on Cybernetics*, 47(11), 3554-3567. DOI:10.1109/TCYB.2016.2569529
- Zabihi, H., Alizadeh, M., Kibet Langat, P., Karami, M., Shahabi, H., Ahmad, A., Nor Said, M., & Lee, S. (2019). GIS multi-criteria analysis by Ordered Weighted Averaging (OWA): Toward an integrated citrus management strategy. *Sustainability*, 11(4), 1009. DOI: 10.3390/su11041009
- Zaidi, E., Thoo, E., De Simoni, G., Beyer M. (2019). Data Fabrics Add Augmented Intelligence to Modernize Your Data Integration, Gartner, <https://www.gartner.com/en/documents/3978267>. Accessed 01, 02, 2024
- Zhang, Y., Bouadi, T., Wang, Y., & Martin, A. (2021). A distance for evidential preferences with application to group decision making. *Information Sciences*, 568, 113–132. DOI: 10.1016/j.ins.2021.03.011
- Zitt, M., Bassecoulard, E., Okubo, Y. (2000). Shadows of the past in international cooperation: Collaboration profiles of the top five producers of science. *Scientometrics*, 47(3), 627–657. DOI: 10.1023/A:1005632319799

Anexo

a) Artículos publicados

1. Data Sources as a Driver for Market-Oriented Tourism Organizations: a Bibliometric Perspective:

Título: Data Sources as a Driver for Market-Oriented Tourism Organizations: a Bibliometric Perspective

Autores: Juan Vidal, Ramón A. Carrasco, Manuel J. Cobo, María F. Blasco

Año publicación: 2023

DOI: 10.1007/s13132-023-01334-5

link: <https://hdl.handle.net/20.500.14352/91036>

Publicación: Journal of the Knowledge Economy

Impacto publicación 2023:

JCR Category: Economics

JCR Ranking: 76/597

JCR Quartile: Q1

Impact factor: 4



Data Sources as a Driver for Market-Oriented Tourism Organizations: a Bibliometric Perspective

Juan Vidal¹ · Ramón A. Carrasco² · Manuel J. Cobo³ · María F. Blasco⁴

Received: 16 March 2022 / Accepted: 25 February 2023
© The Author(s) 2023

Abstract

This paper presents a conceptual framework that accurately represents the current and future perspectives of data-driven companies in tourism by means of an analysis of the data sources used in the data-driven tourism research literature, as well as the research topics to which they are applied. For this purpose, a bibliometric analysis of data-driven tourism research is carried out. The framework of the study is all tourism-related publications whose research was based on data sources during the period 1982–2020. The results show some of the basic bibliometric performance indicators and the maps of science. The main themes of research interest are identified, and the conceptual evolution is obtained based on these maps. Three major thematic areas are identified: tourism research topics, information sources, and data analysis techniques. Based on these three thematic areas, the conceptual model of data architecture and processes of a data-driven organization in the tourism sector are obtained. An additional qualitative analysis of the three thematic areas is performed.

Keywords Tourism · Data-driven · Market-oriented · Bibliometric analysis · Data sources · Data analysis

JOURNAL OF THE KNOWLEDGE ECONOMY

Publisher name: SPRINGER

Journal Impact Factor™

4

2023

3.6

Five Year

JCR Category	Category Rank	Category Quartile
ECONOMICS <i>in SSCI edition</i>	76/597	Q1

Source: Journal Citation Reports 2023. [Learn more](#)

Journal Citation Indicator™

1.17

2023

0.91

2022

JCI Category	Category Rank	Category Quartile
ECONOMICS <i>in SSCI edition</i>	104/600	Q1

Citas recibidas:

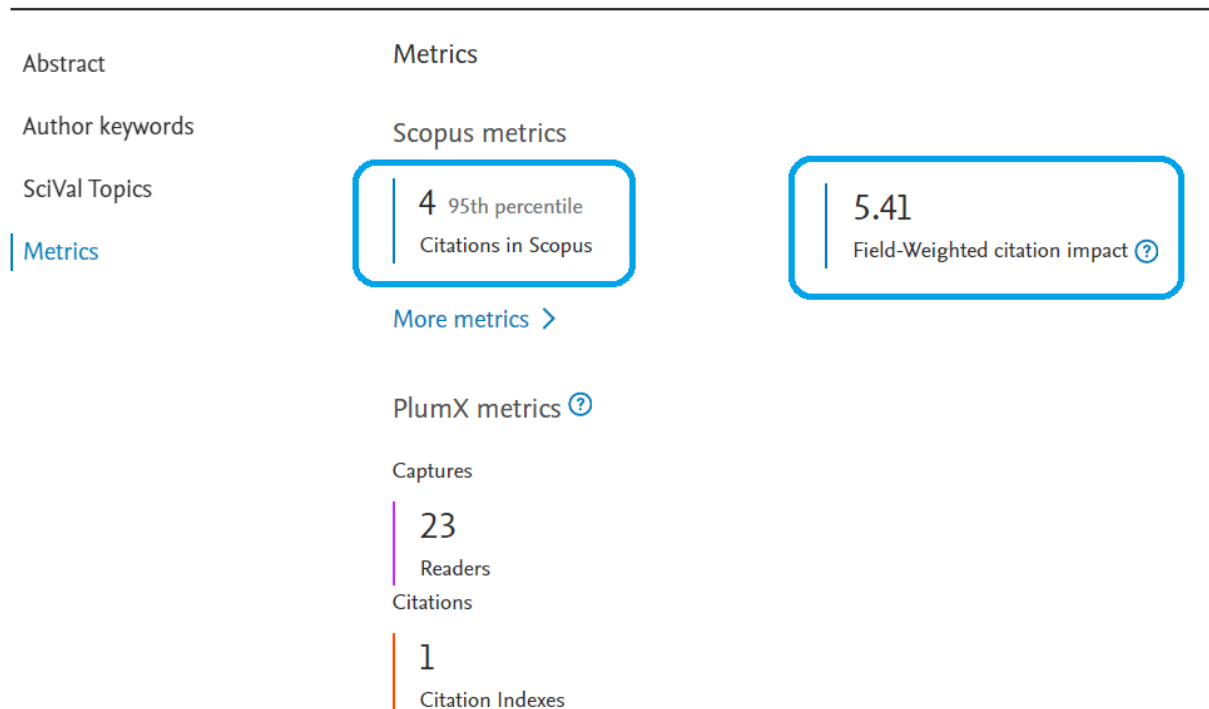
Se adjuntan métricas consultadas a 30 septiembre de 2024:

Citas recibidas: 4 citas Scopus (percentil 95) + 2 citas WoS + 5 citas Google Scholar = **11**

Métrica FWCI (Scopus): 5,41

La métrica FWCI de Scopus es la razón entre las citas del documento y el número promedio de citas recibidas por documentos similares en una ventana de tres años. Cada disciplina contribuye de igual manera a la métrica, lo que elimina las diferencias en los comportamientos de citación entre investigadores. Un valor mayor a 1.00 significa que el documento ha sido citado más de lo esperado según el promedio.

Data Sources as a Driver for Market-Oriented Tourism Organization...



2. Sources Used to Data-Driven Decision-Making in Tourism Management. Identifying the Main Areas of Research:

Título: Sources Used to Data-Driven Decision-Making in Tourism Management. Identifying the Main Areas of Research

Autores: Juan Vidal, Ramón A. Carrasco, Manuel J. Cobo, María F. Blasco

DOI: 10.1007/978-981-16-9268-0_19

Año publicación: 2022

link: https://link.springer.com/chapter/10.1007/978-981-16-9268-0_19

Publicación:

Book Series: Smart Innovation Systems and Technologies

Book title: Marketing and Smart Technologies

Book subtitle: Proceedings of ICMarktech 2021, Vol 1

ISBN: 978-981-16-9267-3

Impacto publicación 2022:


CiteScore: 1,1

SJR: 0,17

SNIP: 0,23

Sources Used to Data-Driven Decision-Making in Tourism Management. Identifying the Main Areas of Research



Juan Vidal, Ramón A. Carrasco , María F. Blasco, and Manuel J. Cobo

Abstract The tourism sector is one of the most affected by the health situation caused by COVID-19. As a result, digital transformation is accelerating in this sector. One of the pillars of this transformation is the management of organizations based on data-driven decision-making. The raw material for such data-driven strategies is obviously the sources of information used. This paper attempts to give a knowledge map of the diverse sources of information used in tourism for this decision-making. To this purpose, we analyse the scientific publications of the last five years in order to identify the main areas of action related to the sources used for data-driven management in tourism. As a result of this bibliometric analysis, we have identified 14 topics that have attracted the interest of the scientific community grouped into three main areas of action.

Keywords Data-driven decision-making · Tourism management · Tourism data sources

3. Conceptual architecture for a tourism organization data platform

Título: Conceptual architecture for a tourism organization data platform

Autores: Juan Vidal, Ramón A. Carrasco, María F. Blasco

Año publicación: 2024

link: <https://link.springer.com/book/9783031526060>

Publicación:

Book series: Springer Proceedings in Business and Economics (SPBE)

Book Title: Tourism and ICTs: Advances in Data Science, Artificial Intelligence and Sustainability

Book Subtitle: Proceedings of the TURITEC 2023 Conference, October 19–20, 2023, Málaga, Spain

ISBN: 978-3-031-52606-0

Impacto publicación 2023:

CiteScore: 0,7

SJR: 0,151

SNIP: 0,140

CONCEPTUAL ARCHITECTURE FOR A TOURISM ORGANIZATION DATA PLATFORM

Juan Vidal Gil, Ramón Alberto Carrasco González, María Francisca Blasco
López

Universidad Complutense de Madrid

juavid01@ucm.es

Abstract

The tourism sector is one of the sectors that has undergone most changes in recent years due to digital transformation. One of the pillars of this transformation is the management of organizations based on data-driven decision making. The raw material for these data-driven strategies is, of course, the sources of information used, which have changed and grown significantly in recent years. This article attempts to provide a conceptual architecture for a modern data platform that effectively manages and analyses these information sources and facilitates data-driven decision-making in tourism organizations.

Keywords: tourism destinations, smart destinations, data-driven organizations, tourism digitalization, tourism data-platform.



[Search](#)
[Sources](#)
[SciVal](#)
[?](#)
[🔔](#)
[🏠](#)

Source details

[Feedback](#) > [Compare sources](#) >

Springer Proceedings in Business and Economics

Years currently covered by Scopus: from 2014 to 2024

ISSN: 2198-7246 E-ISSN: 2198-7254

Subject area: [Economics, Econometrics and Finance: General Economics, Econometrics and Finance](#)

[Business, Management and Accounting: General Business, Management and Accounting](#)

Source type: Conference Proceeding

[View all documents](#) >
 [Set document alert](#)
[Save to source list](#)
[Full Text](#)
[Cite](#)
[1 Cite](#)

CiteScore 2023	0.7	?
SJR 2023	0.151	?
SNIP 2023	0.140	?

4. Multiple criteria approach applied to digital transformation of physical stores in tourism sector: the case of travel agencies

Título: Multiple criteria approach applied to digital transformation of physical stores in tourism sector: the case of travel agencies

Autores: Juan Vidal, Ramón A. Carrasco, María F. Blasco

Año publicación: 2022

link: https://xxiicongresoacit.webs.uvigo.es/wp-content/uploads/2022/10/LIBRO_RESU%CC%81MENES_XXIIAECIT.pdf

Publicación: Libro de resúmenes XXII Congreso Aecit. 6,7,8 Octubre 2022, Universidad de Vigo. El nuevo turismo, innovación, digitalización, sostenibilidad y competitividad.

ISBN: 978-84-09-44041-2

TURISMO Y NUEVAS TECNOLOGÍAS 

71

**MULTIPLE CRITERIA
APPROACH APPLIED TO
DIGITAL
TRANSFORMATION OF
PHYSICAL STORES IN
TOURISM SECTOR: THE
CASE OF TRAVEL
AGENCIES**

Juan Vidal Gil, Ramón Alberto
Carrasco González and María
Francisca Blasco López

There is no doubt that digital transformation in the tourism sector is here to stay. For companies in this sector, this digitalization is essential to remain competitive, improve the customer experience and increase customer engagement. In this context, those players in the tourism value chain that operate through physical stores have seen how purely online operators have gained market share. Investment in digital devices is a constant trend that is penetrating

physical stores as a tool to improve their competitiveness.

These device investments are a key strategic decision where multiple visions in the company have to be taken into account and it is a complex decision problem. This type of problem has been successfully dealt with by means of the so-called multi-criteria decision models, which form a consolidated area of "decision making" allowing to obtain decisions in a structured and validated way, using the appropriate quality criteria.

This work develops a multi-criteria decision making model for physical stores in the tourism sector that want to go digital, providing a systematic approach to manage investment priorities involving all areas of the organization. In order to efficiently integrate all areas of the organization in the decision making process, we rely on the areas defined by the Balanced Scorecard. A use case is developed for its application in the physical stores of travel agencies, in which hierarchical analysis processes are used and a ranking of priorities in the adoption of digital devices is obtained as a result.

Keywords: tourism, digital transformation, store digitalization, multi-criteria decision making, customer experience.

b) Artículos en revisión

5. Detection and analysis of tourist experiences based on online reviews for the improvement of tourism destination management

Título: Detection and analysis of tourist experiences based on online reviews for the improvement of tourism destination management

Autores: Juan Vidal, Ramón A. Carrasco, Eduardo Parra-López, María F. Blasco

Año envío: 2024

Estado: en revisión

Publicación: Investigaciones Turísticas

6. Data and artificial intelligence platform for the positioning of smart tourism services. A use case for tourist offices.

Título: Data and artificial intelligence platform for the positioning of smart tourism services. A use case for tourist offices

Año envío: 2025

Estado: Finalizado, pendiente de someter a revisión.

c) Menciones y reconocimientos académicos

Accesit PhDay 2023:



D. Manuel de la Calle Vaquero, Vicedecano de Investigación y Doctorado de la Facultad de Comercio y Turismo

HACE CONSTAR

Que D. **Juan Vidal Gil**, con DNI 50848571X, ha obtenido un **Accésit** en la I Jornadas de Doctorado «PhDay 2023 Turismo» organizadas por la Facultad de Comercio y Turismo el 27 de septiembre de 2023, con una presentación titulada **Revisión de las fuentes de datos como activo estratégico de las organizaciones turísticas orientadas al dato**. La Jornada estuvo promovida por la Escuela de Doctorado de la Universidad Complutense de Madrid (UCM) en el marco de la VII Edición del «PhDay», con participación de los vicerrectorados de Estudios y de Estudiantes.

DE LA CALLE
VAQUERO
MANUEL -
50083767W

Firmado digitalmente
por DE LA CALLE
VAQUERO MANUEL -
50083767W
Fecha: 2023.10.06
12:01:17 +02'00'

Madrid, 06 de Octubre de 2023