
UNIVERSIDAD
COMPLUTENSE
MADRID

**UNIVERSIDAD COMPLUTENSE DE MADRID**

# MÁSTER EN MINERIA DE DATOS E INTELIGENCIA DE NEGOCIOS

Minería de patrones de movilidad humana y características espaciales-temporales a través de datos geo-etiquetados de redes sociales: Un caso de estudio en Madrid Capital.

Trabajo de fin de máster presentado por: Salvador Alejandro Rueda

Dirigido por: Javier Portela

**Madrid**

Septiembre / 2017

# RESUMEN

La caracterización de los patrones de movilidad humana son esencial para poder comprender el comportamiento humano y su interacción en un ámbito social, económico y ambiental. Además, juega un rol importante en el desarrollo y mejora en la planificación urbana, ingeniería de transporte, salud pública y estrategias de marketing.

Con el incremento en el uso de dispositivos móviles con detección de ubicación (GPS) y el de las redes sociales, ha surgido una nueva forma de estudiar la dinámica de una ciudad a partir del minado de datos georreferenciados generados por los usuarios de estas redes sociales.

El propósito de este estudio es el detectar los patrones de trayectorias y sus características espacio temporales más relevantes dentro de Madrid capital. Para este objetivo, se ha ideado un proceso a través del cual, se han recogido desde la red social Twitter, tweets georreferenciados para posteriormente llevar a cabo un proceso de asignación de categorías las cuales permiten dar una semántica a cada una de las ubicaciones (latitud y longitud) obtenidos en cada uno de los tweets.

El proceso de asignación de las categorías a cada uno de los tweets se realizó a través de la red social Foursquare. Por medio del API disponible por esta red social, es posible consultar las categorías de bajo (nombre de la ubicación), medio y alto nivel asociada a cada posición desde donde se generaron cada uno de los tweets. Adicionalmente, como parte de la caracterización de cada uno de las ubicaciones, se ha asignado a cada una de estas el barrio de Madrid desde el cual se generó cada uno de los tweet.

Luego de caracterización de cada una de las ubicaciones a través de la asignación de bajo, medio, alto nivel y de barrio; se procede a evaluar lo patrones de trayectorias y de asociaciones entre ubicaciones. Para ello, se ha empleado algoritmos de análisis de asociación y el análisis de secuencial para así poder determinar las características de desplazamientos y los comportamientos de movilidad dentro de Madrid Capital.

**Palabras Claves:** Geolocalización, Movilidad, Minado de patrones de trayectorias, Análisis de asociación, Análisis Secuencial, Posicionamiento

# ABSTRACT

The online social networks such as Twitter, allow users to include in their posts the geographical coordinates (latitude and longitude) from the location where the post was created. The time and the geographical coordinates associated to each post, point out the spatial-temporal movements of people in real life. The aim of this investigation is to analyze those movements within Madrid capital. To this end, we define a methodology to collect and process the posts with geographical coordinates from twitter; afterward, we assigned to each geographical coordinates a semantic which define the name of the location and the general characteristics, through the information provided by the online social network Foursquare. Our approach to infers the movements is based on the association and sequential analysis algorithms.

**Keywords**: Geopositioning, Human mobility, association analysis, sequential analysis, Geo-social data.

# TABLA DE CONTENIDO

# ÍNDICES DE FIGURAS

# ÍNDICES DE TABLAS

# CAPÍTULO I - INTRODUCCIÓN

## 1. Introducción

Las redes sociales permiten a sus usuarios asociar contenido geográfico y temporal a cada una de las publicaciones que generan en dichas redes. Además, cada usuario visita diariamente un conjunto de localizaciones, generando desde cada una de estas, una cantidad de datos georreferenciados que oculta información interesante sobre la dinámica humana y comportamiento de movilidad dentro de un contexto urbano.

Los análisis generados a partir de esta información rica en datos georreferenciados puede ser usada en muchas áreas incluyendo la planificación urbana, gestión de tráfico, recomendación de rutas, campañas de marketing, seguridad y monitorización de salud, etc. En este sentido, existe una gran oportunidad dentro del área de minería de datos para desarrollar herramientas que permitan explotar y analizar conjuntos de datos con información espacio-temporal.

Este trabajo presenta una metodología para el estudio de la dinámica de movilidad dentro de la ciudad de Madrid, la cual emplea datos georreferenciados obtenidos a través de la red social Twitter.

A cada una de las posiciones (coordenadas geográficas), desde donde se generaron los tweets, se le asignan varias semánticas las cuales permiten comprender el significado de cada ubicación, ya que por sí solo un par de coordenadas más allá de indicar la ubicación de un tweet no presenta ningún significado. La asignación de las semánticas a cada una de estas ubicaciones se realiza empleado la red social Foursquare, a través de la cual, a cada coordenada geográfica del tweet se le asigna una semántica de bajo nivel que indica el nombre especifico de la ubicación desde donde se generó el tweet. Además, se le asignan semánticas de medio y alto nivel, las cuales hacen referencia a una estructura de categorización más genérica. Así, la semántica de bajo nivel Estación de Madrid-Puerta de Atocha, tiene asociada como semántica de medio y alto nivel Train Station y Travel&Transport respectivamente.

Adicionalmente, aparte de las semánticas de bajo, medio y alto nivel, también se le asigna a cada tweet el barrio correspondiente a la ubicación desde donde se generó, con el objeto de comprender la dinámica de la ciudad a partir de una componente geográfica. De este modo, el barrio asociado a las semánticas indicadas anteriormente corresponde a Atocha.

El punto de partida de este estudio yace en que los individuos tienden a seguir rutas comunes, por ejemplo se desplazan al trabajo cada día usando las mismas vías, estaciones de metro o tren, permanecen en una ubicación durante su hora de trabajo o centros de estudio, se desplazan a otras ubicaciones para comer o realizar compras. Así, si se tiene una

cantidad suficientes datos para modelar estos comportamientos, se pueden extraer patrones que se puedan emplear para predecir y gestionar los futuros desplazamientos de los individuos.

A partir de las categorizaciones de cada uno de los tweets y conociendo la información temporal asociada a estos, se describe por medio del análisis de asociación las relaciones entre las visitas realizadas a cada una de las ubicaciones. Con este análisis se pretende mostrar las implicaciones que existen entre las visitas a diferentes ubicaciones teniendo en cuenta las categorizaciones de medio nivel, bajo nivel y de barrios.

A través del análisis de secuencias, se obtiene las rutas más comunes que siguen los habitantes de la ciudad de Madrid, para ello también se tienen en cuenta la componente temporal asociada a cada ubicación. Este análisis se lleva a cabo por medio de la categorización de bajo nivel en función de los días de la semana. De esta forma, se obtienen los patrones de viajes más frecuentados para cada día de la semana dentro de la ciudad

Tanto para el análisis de asociación como para el análisis temporal, se evalúan los indicadores más importantes resultantes de la aplicaciones de los algoritmos y que proporciona una magnitud de las probabilidades de ocurrencia de dichas trayectorias o asociaciones.

Todas las etapas del desarrollo de este estudio: Recolección de datos, procesamiento, evaluación, aplicaciones de algoritmos y visualización se han realizado por medio de del software de análisis estadístico R[1]. Adicionalmente y para algunas visualizaciones se ha empleado la plataforma de inteligencia de localización CartoDB[2].

El objetivo perseguido en este estudio pretende analizar la información espacial y temporal asociada a los tweets generados dentro de Madrid capital para así detectar trayectorias típicas entre cada uno de los individuos, patrones de desplazamiento y patrones de asociaciones de ubicaciones frecuentemente visitadas y que además pueda servir como base para futuros estudios.

# 2. Motivación.

El rápido incremento del uso de redes sociales ha potenciado el estudio de la movilidad humana y sus implicaciones en la sociedad. A pesar de que existen algunos trabajos que han llevado a cabo la extracción de trayectorias por medio de la explotación de datos georreferenciados generados a través de redes sociales, ninguno de estos se han llevado a cabo en la ciudad de Madrid.

[1] https://www.r-project.org
[2] https://carto.com

Dada la escasa información relacionada con los aspectos de movilidad y la dinámica de la ciudad de Madrid, la motivación para el desarrollo de este trabajo se centra en descubrir todas las principales características que definen el comportamiento y la dinámica de Madrid.

# 3. Definición de objetivos

## 3.1. Objetivo general

Descubrir patrones de movilidad y detección de trayectorias dentro de Madrid capital a partir de análisis secuencial y de asociación por medio de información georreferenciada.

## 3.2. Objetivos específicos

- Recopilar, procesar y depurar tweets con información espacial y temporal generados dentro de Madrid capital.
- Asociar una semántica de barrio, bajo, medio y alto nivel a cada uno de los datos geográficos en función de la ubicación desde donde fueron generado los tweets.
- Determinar la asociación entre ubicaciones visitadas a través de categorización de bajo y medio nivel así como de barrio.
- Obtener patrones de trayectorias entre ubicaciones categorizadas a bajo nivel en función de los días de la semana.

# CAPÍTULO II – ESTADO DEL ARTE

## 1. Introducción

Los datos obtenidos a través de las redes sociales están siendo percibidos como una alternativa a los tradicionales métodos de encuestas debido a la facilidad, volumen de datos disponibles y por la cantidad de información que se puede obtener a partir de estos.

La red social Twitter se han convertido en una fuente de datos de gran interés y que ha recibido una especial atención por parte de la comunidad académica y desde donde se han originado varias líneas de investigación, muchas de estas enfocadas al contenido de los mensajes del tweet (análisis sentimental) y las características de los usuarios de Twitter como por ejemplo los estudiados en [2].

Aparte del minado de opiniones, sentimientos y características de usuarios por medio de Twitter, también existen otras líneas de investigación las cuales explotan los contenidos georreferenciados y características de espacio-temporales que se pueden obtener a través de los datos de esta red social como por ejemplo [3]. Muchas de estas investigaciones están orientadas a determinar los patrones de movilidad y trafico además de trayectorias típicas en diferentes localidades tal y como se ha estudiado en [4] y [5], y entender así la dinámica social de un entorno geográfico determinado tal y como se plantea en [6].

La metodología de recolección de datos y la forma de asignar una semántica a la información geográfica (coordenadas geográficas) llevada a cabo en este trabajo es similar a la ejecutada en [5].Como elemento diferencial, este trabajo introduce una nueva semántica a la hora de categorizar una ubicación geográfica. A parte de una semántica de bajo nivel, la cual corresponde con el nombre especifico de la ubicación, por ejemplo: Museo del Jamón, 100 Montaditos, Primark, etc., también se emplea como semántica asociada a cada ubicación geográfica, el barrio en la cual se encuentra inscrita la ubicación desde la cual se generó el tweet. Es importante considerar la segmentación geográfica de la cuidad como parte de la semántica asociada a la información geográfica, ya que es posible que existan ubicaciones con una misma semántica de bajo nivel, como las indicadas anteriormente y que se encuentren distribuidas en diferentes barrios de la cuidad.

La mayoría de los estudios encontrados en la literatura tienen en cuenta datos georreferenciados generados a partir de redes sociales con el objetivo de determinar patrones de secuencias de trayectorias tal y como se plantea en [5]. Aunque las secuencias de trayectorias proporcionan una visión en cuanto a la movilidad considerando la dimensión temporal; muy pocos estudios se han realizado teniendo en cuenta las asociaciones entre ubicaciones, que al igual que el análisis secuencial, permiten generar una visión de la dinámica de los individuos y así poder aprovecharse en estrategias de marketing o motores de recomendación de desplazamientos.

Las investigaciones llevadas a cabo en referencia a los patrones de movilidad se diferencian principalmente en el entorno geográfico elegido para llevar a cabo el estudio y en los algoritmos aplicados para la obtención de dichos patrones. Otras investigaciones están enfocadas a los algoritmos de la detección de trayectorias tal y como se destalla en [7].

Para la detección de trayectorias de este estudio, llevado a cabo por medio del análisis secuencial, se empleará una adaptación del algoritmo prefix-tree-based search [8], el cual se encuentra implementado dentro del paquete de R TraMineR. De igual forma, para la detección de asociaciones entre ubicaciones se ha empleado el algoritmo Apriori el cual se encuentra implementado dentro del paquete Arules de R.

# CAPITULO III – RECOLECCIÓN DE DATOS

## 1. Recolección de tweets.

Para la recolección de los tweets es necesario establecer una conexión con Twitter con el objetivo de almacenar los tweets para su posterior análisis. La manera ideal de recolectar los tweets de forma automática es a través del uso del API de Twitter. Twitter dispone de dos tipos de APIs[3] para la captura de datos; un API de tipo streaming y un API de tipo REST.

El API de tipo REST proporciona un acceso programático que permite leer y escribir datos en Twitter; crear un nuevo tweet, leer el perfil de un usuario, búsqueda de tweets por palabras específicas, búsqueda de tweets generados en una localización (latitud, longitud) dentro de un determinado radio. En R las interacciones a través del API REST están implementadas a través de paquete twitteR[4].

El API de tipo streaming permite capturar los tweets en tiempo real a medida que se genera. El API streaming requiere abrir una conexión de forma permanente durante el tiempo indicado dentro de la sintaxis de la consulta. Antes de almacenar los resultados, este API permite realizar filtrado y agregación de datos en función de los requerimientos del estudio. En R el API de tipo streaming se encuentra implementado en el paquete streamR[5].

Ambas API permite recolectar los datos y las variables requeridas para este estudio, pero al comparar el funcionamiento de ambas se observa que la recolección de datos en streaming se obtiene mayor cantidad de tweets que los que se obtiene al emplear el API REST. Adicionalmente, el API streaming permite de forma automática obtener, filtrar y almacenar los datos sin necesidad de código adicional. En cambio el API REST requiere que la sentencia de consulta de recolección de datos se repita dentro de un bucle para recopilar los nuevos tweets que se vayan generando y así, dependiendo del intervalo de tiempo de repetición de recolección es posible que se obtengan tweets repetidos, por lo que será necesario un paso adicional para borrar tweets duplicados. Por los explicado anteriormente se ha seleccionado el API de tipo streaming para la recolección de datos y con ello el uso del paquete streamR.

La función filterStream del paquete streamR permite implementar un filtrado a través de la conexión de streaming con Twitter con el objetivo de recolectar los datos necesarios para este trabajo. Dado que este estudio está basado en la ciudad de Madrid, se debe especificar

[3] https://dev.twitter.com/streaming/overview , https://dev.twitter.com/rest/public
[4] https://cran.r-project.org/web/packages/twitteR/twitteR.pdf
[5] https://cran.r-project.org/web/packages/streamR/streamR.pdf

mediante el campo “locations” de la función filterStream, la región sobre la que queremos obtener los tweets.

Para definir la región se debe colocar un par de coordenadas en donde la primera representa la esquina suroeste del filtro, mientras que la otra coordenada representa la esquina noreste. En este estudio se han definido las siguientes coordenadas: -3.922433,40.277213 y -3.495995,40.651585. Con este par de coordenadas el API construirá un rectángulo que funcionará de filtro geográfico y que devolverá todos los tweets que se generen dentro de la frontera definida por esta coordenadas.

Se observa que las coordenadas indicadas en el filtro de la función filterStream abarca toda el área de Madrid capital y algunos municipios aledaños por lo que en la siguiente fase será necesario filtrar únicamente los tweets georreferenciados que se hayan generado dentro de los límites de Madrid capital.

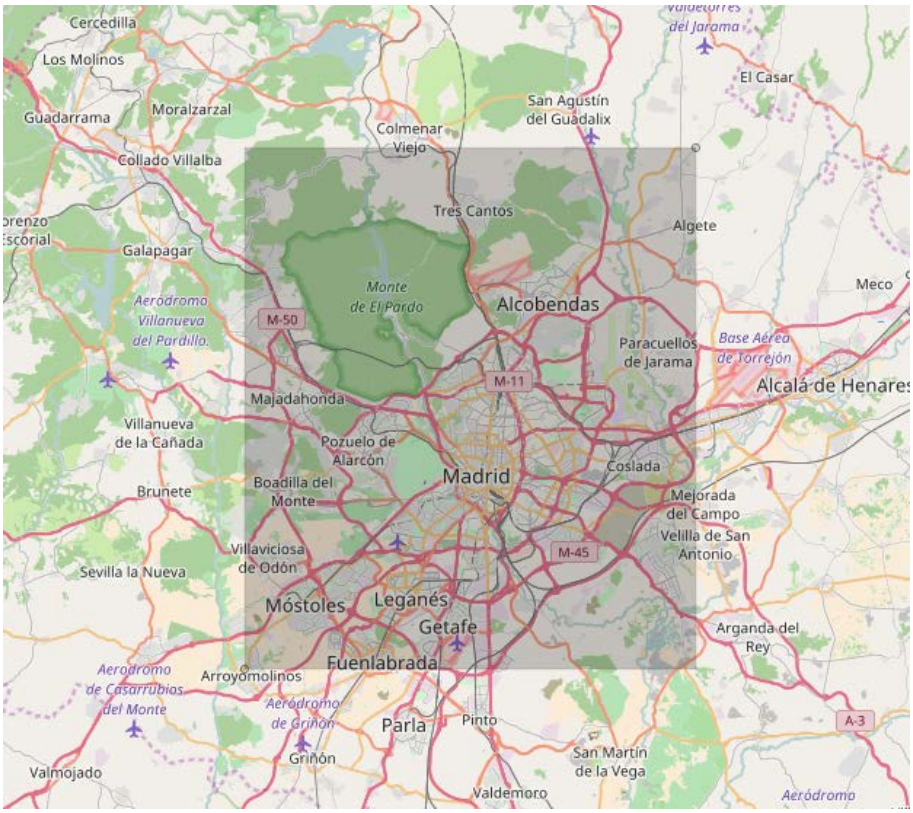



*Ilustración 1:Representación filtro geográfico de la función filterStream.*

Es importante destacar que dentro de esta frontera se recogerán todos los tweets geo-localizados, es decir tweets generados dentro del área definida, pudiendo estar o no georreferenciados, es decir, tweets que incorporen o no la localización (Latitud y Longitud) desde la cual se originó (siempre que el usuario tenga habilitado la opción de geolocalización desde el dispositivo donde genera los tweets). Así, esta frontera no implica

que todos los tweets que se generan dentro de esta sean georreferenciados. En la siguiente fase se filtraran todos aquellos tweets que aporte la información precisa desde donde se originaron (Latitud y Longitud).

La recolección de los datos por medio de API streaming se debe de ejecutar desde una plataforma que garantice la continuidad en la recolección de los datos sin que se vea afectada por fallos tales como eléctricos o de conexión a internet. Con el objetivo de evitar estos fallos y de mantener la integridad en la recolección de los datos se ha decidido ejecutar la recolección de los datos a través de una plataforma en la nube.

Por medio de la plataforma Amazon Web Services (AWS)[6] es posible desplegar de forma rápida, ordenadores virtuales en la nube llamados instancias. Las características de dichas instancias (números de núcleos, RAM, almacenamiento) pueden ser elegidas o ajustadas en función de la complejidad de la tarea que se vaya a llevar a cabo. Para propósito de la recolección de los datos se ha usado el servicio Amazon EC2 (Elastic Compute Cloud). Sobre esta instancia se debe de instalar un sistema operativo y software denominado Amazon Machine Images (AMIs), sobre las que se pueda llevar a cabo las tareas recolección de tweets. Dado que se usará el paquete de R streamR para realizar la recolección de datos, se ha instalado el AMI: RStudio-0.99.447_R-3.2.1_ubuntu-14.04-LTS-64bit(ami-0c13557b), el cual es una imagen del sistema operativo Ubuntu en el que se tiene instalado el software R y el IDE Rstudio.

Configurada la instancia EC2 con el AMI seleccionado, es posible acceder de forma remota a través de una interfaz web a la instancia que ejecuta RStudio para llevar a cabo la recolección de datos. En la ilustración 2, se muestra la consola de RStudio ejecutado la recolección de datos desde la instancia EC2 de AWS.

En la ventana del Script se cargan las librerías streamR y ROAuth. Seguidamente se crean cada uno de los objetos que contienen los parámetros necesarios para realizar la autenticación contra el servidor de Twitter (líneas 3-7, ilustración 2). En la línea 8 del código se crea el objeto my_oauth que se incluirá en la función filterStream y que contiene cada uno de los parámetros definidos anteriormente y necesarios en la autenticación.

La función filterStream del paquete streamR permite abrir un streaming de datos con Twitter y a su vez implementar los filtros indicados en la función sobre los datos que se reciben. En la función se define el nombre y formato del fichero sobre el cual se irán almacenando los datos capturados (rawmadgeotweet.json). El parámetro “locations” tal como se indicó anteriormente, define el área sobre el cual se recolectaran los tweets. En el parámetro “timeout” debe especificar el tiempo en segundos en el que la comunicación del streaming de datos permanecerá abierta. El parámetro “oauth” requiere todos los

[6] https://aws.amazon.com

parámetros de autenticación de usuario para que permita la creación de la conexión contra el servidor de Twitter.

Al ejecutar la recolección de datos desde RStudio en una instancia de AWS, nos garantiza que la comunicación y la transferencia de datos entre los servidores de Twitter y la instancia sea lo más estable posible, además nos asegura un respaldo en caso un fallo de energía, así, la conexión de la parte cliente (instancia EC2) se mantendrá siempre activa.

Existe la posibilidad de que en la parte cliente o servidor de esta conexión se genere algún evento que pueda desconectar el streaming de datos. Por esta razón, la función filterStream se incluye dentro de un bucle que obliga a levantar inmediatamente la conexión en caso de que ocurra algún evento que desconecte streaming de datos. En la ilustración 2 se puede observar en la consola que han ocurrido alguno de estos eventos, pero la conexión se ha levantado automáticamente, además los datos se continua almacenado en el mismo fichero definido inicialmente. El icono rojo en la parte superior izquierda de la ventana de la consola muestra que se mantiene la ejecución de la recolección de datos y almacenamiento de los mismo en el fichero rawmadgeotweet.json (ventana Files de RStudio)

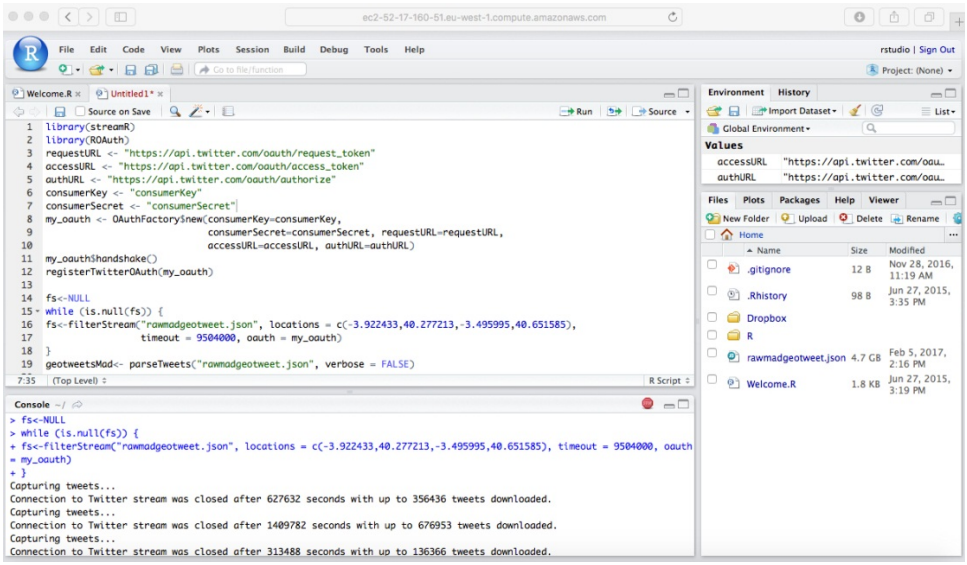



*Ilustración 2:Consola RStudio en EC2 de AWS.*

Se puede acceder al código empleado para la recolección de los tweets a través del siguiente link ConexiónApiTwitter

Es posible llevar a cabo la depuración y el análisis de los datos recolectados desde el RStudio en la instancia de AWS, pero dado que las características configuradas sobre esta instancia son las mínimas posibles para hacer un uso gratuito del servicio; los datos recolectados deben ser transferido desde la instancia al ordenador sobre el cual se continuará con el estudio.

Los datos se han transferido fácilmente a través de una capeta asociada a una cuenta de Dropbox. De esta forma, el archivo con los datos recolectados podrán ser descargados de forma local vía web desde la cuenta de Dropbox que está asociada a la carpeta Dropbox de la instancia de AWS (ventana Files de RStudio).

## 2. Preparación de datos

Tal como se indicó en la fase de la recolección de datos, los tweets son almacenados por la función filterStream en un fichero con formato .json (rawmadgeotweet.json). A diferencia de otros formatos de almacenamiento de datos tales como .csv ó .xlm; json permite almacenar un gran volumen de datos manteniendo un tamaño de fichero relativamente pequeño en comparación con los que se pueden obtener almacenando los mismo datos con los formatos anteriormente mencionados. Además, por ser un formato de texto ligero, permite el intercambio de datos de forma más rápida y sencilla.

A continuación se muestra la estructura de un tweet en formato json y todas las variables que recoge la función filterStream para cada uno de los tweets capturados. Las estructura que se presenta a continuación corresponde al tweet de la ilustración 3.


*Ilustración 3: Ejemplo tweet y su estructura json.*

El campo “created_at” de la estructura .json muestra la hora UTC cuando el tweet fue creado.

Los campos "id" y "id_str" corresponden a un número secuencial que identifica unívocamente a un tweet. El campo "text" muestra el contenido del tweet (texto, hashtags, url, iconos, etc) que el usuario añade para transmitir un mensaje.

El campo "source" indica la fuente de emisión del tweet. Un tweet puede ser originado desde la propia red social Twitter o bien puede estar vinculado a otra red social. La mayoría de las redes sociales como por ejemplo Instagram, Foursquare, Tumblr y entre otras, permiten asociar una cuenta de Twitter, de esta forma toda la actividad que se genere en dichas redes sociales será replicada en Twitter incluyendo todas las propiedades asociadas a la misma tales como la geolocalización, fotografías, texto, etc. En este ejemplo se observa que este tweet fue generado a partir de una publicación de una cuenta de la red social Instagram.

Lo campos "in_reply_to_status_" muestran toda la información relativa a una respuesta de tweet. Este este ejemplo y dado que el tweet se generó sin ser una respuesta a otro tweet, todos los campos tienen un valor Null.

El objeto "user" contiene toda la información relativa al autor del tweet y su cuenta. Los campos "id" y "id_str" representan de forma univoca a un usuario. El campo "name" muestra el nombre del usuario y "screen_name" el nombre de usuario de Twitter. El campo "location" indica la ubicación registrada en la cuenta del usuario. Este campo no representa la ubicación geográfica al momento de generar el tweet.

Otros campos de interés a destacar dentro del objeto "user" son el número de seguidores que tiene el usuario en el momento en que se generó el tweet ("followers_count") o el número de tweets que el usuario ha marcado como favoritos ("favourites_count") y la fecha en la cual la cuenta de Twitter de dicho usuario fue creada ("created_at").

Los objetos "geo" y "coordinates" almacena la información más relevante de los datos contenidos dentro de la estructura json del tweet para nuestro estudio ya que ambos objetos contienen las coordenadas en formato decimal reportados por el usuario o la aplicación cliente (aplicación de red social a la que se asoció la cuenta de Twitter) y desde donde se generó el tweet. En el ejemplo presentado, las coordenadas desde donde se generó el tweet son las siguientes: latitud= 40.44627778 y longitud=-3.69181389. Estas coordenadas ubican a este tweet en la proximidades de la estación de Nuevos Ministerios.

El objeto "place" incluye todos los atributos de las ubicaciones especificadas al generar un tweet. El objeto "place" corresponde con una localización pero no con una coordenada. La localización se corresponde con atributo "name" y lleva asociado un código que se especifica en el atributo "id". Otros campos como "full_name" y "country_code", permiten identificar a nivel de ciudad o país desde donde se generó el tweet. El atributo "bounding_box" enmarca a través de un polígono construido por medio de cuatro coordenadas la ubicación del atributo "id"; así la ubicación con atributo "id" cuyo nombre es "name", se encuentra contenida por los limites definidos en el "bounding_box".

El atributo "bounding_box" del objeto "place" permite junto al filtro "location" definido en la función filterStream capturar todos aquellos tweets cuyo atributo "bounding_box" se encuentre dentro o se solape con el filtro establecido en la función filterStream.

El código json del tweet de la ilustración 3 se presenta a continuación.

```
{
  "created_at":"Fri Feb 10 20:56:26 +0000 2017",
  "id":830158507819270144,
  "id_str":"830158507819270144",
  "text":"Just posted a photo @ Nuevos Ministerios
https:\/\/t.co\/9KkIZdBaAG",
  "source":"\u003ca href=\"http:\/\/instagram.com\"
rel=\"nofollow\"\u003eInstagram\u003c\/a\u003e",
  "truncated":false,
  "in_reply_to_status_id":null,
  "in_reply_to_status_id_str":null,
  "in_reply_to_user_id":null,
  "in_reply_to_user_id_str":null,
  "in_reply_to_screen_name":null,
  "user":{
    "id":1209406033,
    "id_str":"1209406033",
    "name":"Error 404",
    "screen_name":"Alba_forever_",
    "location":"Madrid, Comunidad de Madrid",
    "url":"https:\/\/instagram.com\/alba10_hg",
    "description":"17. | Progressive-Trap-House-Dubstep | @Krewella",
    "protected":false,
    "verified":false,
    "followers_count":279,
    "friends_count":359,
    "listed_count":2,
    "favourites_count":1850,
    "statuses_count":593,
    "created_at":"Fri Feb 22 18:05:35 +0000 2013",
    "utc_offset":null,
    "time_zone":null,
    "geo_enabled":true,
    "lang":"es",
    "contributors_enabled":false,
    "is_translator":false,
    "profile_background_color":"C0DEED",
"profile_background_image_url":"http:\/\/abs.twimg.com\/images\/themes
\/theme1\/bg.png",
"profile_background_image_url_https":"https:\/\/abs.twimg.com\/images\
/themes\/theme1\/bg.png",
    "profile_background_tile":false,
    "profile_link_color":"1DA1F2",
    "profile_sidebar_border_color":"C0DEED",
    "profile_sidebar_fill_color":"DDEEF6",
    "profile_text_color":"333333",
    "profile_use_background_image":true,
"profile_image_url":"http:\/\/pbs.twimg.com\/profile_images\/808031239
462600706\/anT7GDGG_normal.jpg",
"profile_image_url_https":"https:\/\/pbs.twimg.com\/profile_images\/80
803
1239462600706\/anT7GDGG_normal.jpg",
"profile_banner_url":"https:\/\/pbs.twimg.com\/profile_banners\/120940
6033\/1481484633",
```

```
"default_profile":true,
    "default_profile_image":false,
    "following":null,
    "follow_request_sent":null,
    "notifications":null
  },
  "geo":{
    "type":"Point",
    "coordinates":[
      40.44627778,
      -3.69181389
    ]
  },
  "coordinates":{
    "type":"Point",
    "coordinates":[
      -3.69181389,
      40.44627778
    ]
  },
"place":{
    "id":"206c436ce43a43a3",
"url":"https:\/\/api.twitter.com\/1.1\/geo\/id\/206c436ce43a43a3.json"
,
    "place_type":"city",
    "name":"Madrid",
    "full_name":"Madrid, España",
    "country_code":"ES",
    "country":"España",
    "bounding_box":{
      "type":"Polygon",
      "coordinates":[
        [
          [
            -3.889005,
            40.312071
          ],
          [
            -3.889005,
            40.643518
          ],
          [
            -3.518010,
            40.643518
          ],
          [
            -3.518010,
            40.312071
          ]
        ]
      ]
    },
"attributes":{
    }
```

```
  },
  "contributors":null,
  "is_quote_status":false,
  "retweet_count":0,
  "favorite_count":0,
  "entities":{
    "hashtags":[
    ],
    "urls":[
      {
        "url":"https:\/\/t.co\/9KkIZdBaAG",
       "expanded_url":"https:\/\/www.instagram.com\/p\/BPGSJGNB8fB\/",
        "display_url":"instagram.com\/p\/BPGSJGNB8fB\/",
        "indices":[
          41,
          64
        ]
      }
    ],
    "user_mentions":[
    ],
    "symbols":[
    ]
  },
  "favorited":false,
  "retweeted":false,
  "possibly_sensitive":false,
  "filter_level":"low",
  "lang":"es",
  "timestamp_ms":"1486760186520"
}
```


Al importar el fichero rawmadgeotweet.json a R es necesario transformar los datos contenidos dentro de este en un formato relacional sobre el que sea más sencillo llevar a cabo la selección y manipulación de cada una de las variables. El mismo paquete streamR incluye una función que permite analizar sintácticamente el fichero .json y construir así un data frame (marco de datos de estructura relacional) con cada una de las variables incluidas en cada tweet.

```
library(streamR)
rawtweets<- parseTweets("rawmadgeotweet.json", verbose = FALSE)
```

El data frame **rawtweets** contiene un total de **1.130.307** tweets (observaciones), las cuales fueron recogidas desde el Lunes 6 de febrero del 2017 a partir de las 14:12:13 hasta el Sábado 22 de abril del 2017 a las 06:36:44. Este conjunto inicial de datos contiene tanto tweets georreferenciados como no georreferenciados.

Adicionalmente, este conjunto de datos contiene variables que no se emplearan en este estudio, por este motivo se crea una nueva estructura data frame que contienen únicamente las siguientes variables: **lat** (latitud de la ubicación desde donde se generó el tweet), **lon**

(longitud de la ubicación desde donde se generó el tweet), **created_at** (fecha y hora de la creación del tweet), **user_id_str** (identificador univoco del usuario que generó el tweet). Sobre esta misma estructura se seleccionan únicamente los tweets georreferenciados, es decir tweets que incluyan la información de la latitud y longitud desde donde fueron generados. El código empleado para llevar a cabo el filtrado se puede consultar por medio del siguiente link FiltradoTweets.

A partir del conjunto de datos ($TW$), que incluyen únicamente las coordenadas, fecha y hora de la creación del tweet y el identificador del usuario que lo generó, se establece la definición con la que se trabajara a lo largo de esta estudio.

Un tweet georreferenciado $tw$ , que pertenece al conjunto de datos $TW$, esta caracterizado por un individuo $u$ , quien generó el tweet y una ubicación $l$ desde la cual se generó dicho tweet. Dicha ubicación está identificada por un par de coordenadas geográficas $l = (x, y)$, latitud y longitud respectivamente y la fecha y hora $t$ cuando se generó el tweet. De esta forma, un tweet georeferenciado está definido como: $tw = (l, t, u) \; \forall \; tw \in TW$.

Después del filtrado de observaciones que no incluyen las coordenadas geográficas desde donde se generaron los respectivos tweets, el conjunto de datos $TW$ contiene un total de **206.186** tweets georreferenciados.

Dado que el objetivo de este estudio es únicamente la región de Madrid capital, es necesario filtrar el conjunto de datos con los tweets georreferenciados que se hayan generado dentro de los límites geográficos de Madrid capital. A partir de los datos que conforman el polígono de Madrid capital y el paquete splancs[7] de R, es posible desechar los tweets situados fuera de esta región. En el siguiente link de puede acceder al código empleado para el filtrado geográfico de los tweets:filtradoTweetsMadrid

Después de aplicar el filtro geográfico al conjunto de datos geotweets, se tiene un conjunto de datos con un total de **170.854** observaciones. Este cantidad representa un 12% del total de los tweets inicialmente recolectados (conjunto de datos rawtweets).

En la imagen que se muestra a continuación se presentan el polígono que abarca todo Madrid capital y los tweet georreferenciados obtenidos a partir de los pasos previos

[7] https://cran.r-project.org/web/packages/splancs/splancs.pdf

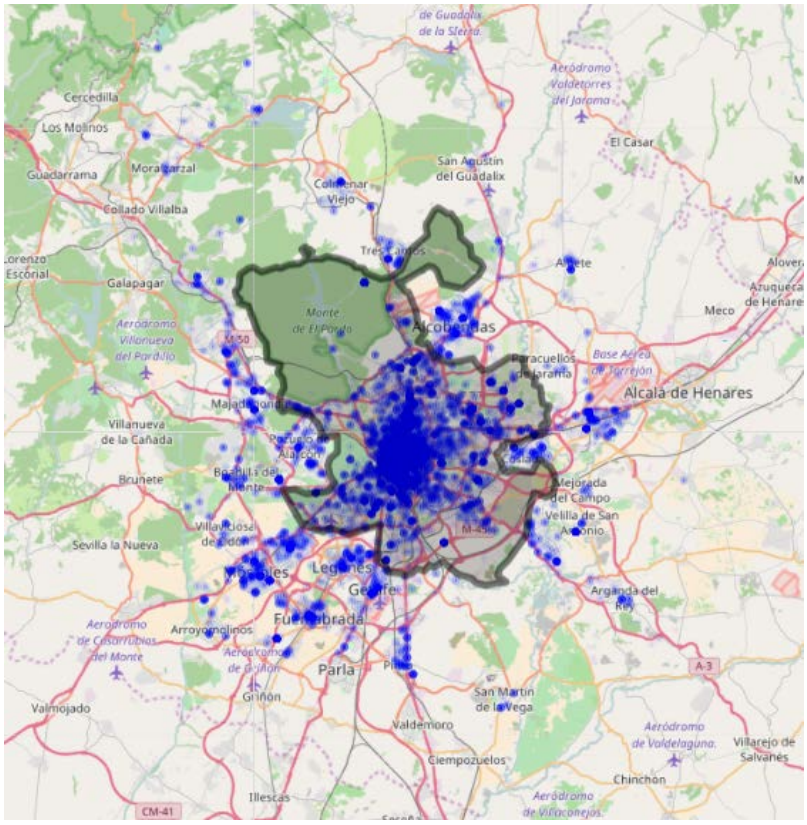



*Ilustración 4:Polígono Madrid capital y tweets recolectados.*

En la siguiente imagen, se muestran los tweets que se generaron dentro de Madrid capital, después de aplicar el filtrado a partir del polígono que define el área de estudio.

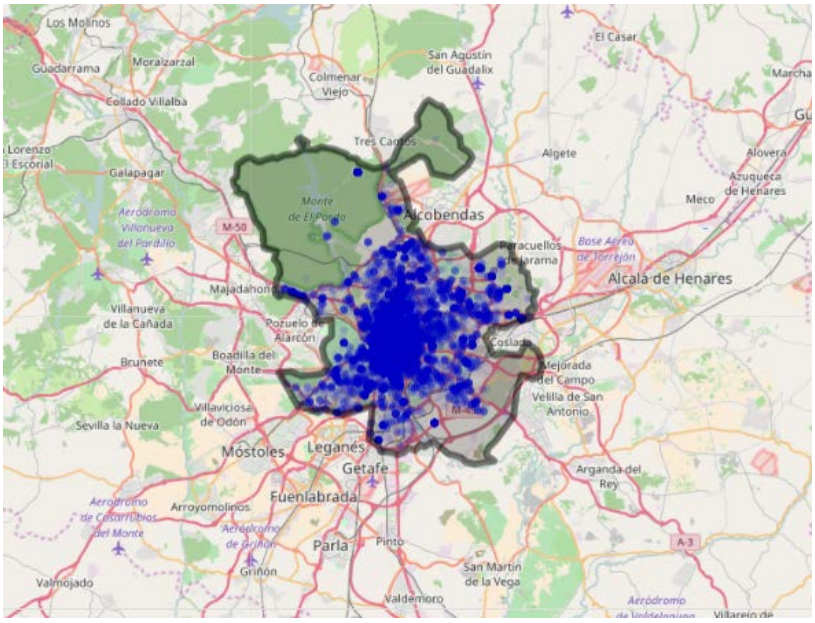



*Ilustración 5:Tweets generados dentro de Madrid capital.*

En este punto, es posible determinar el número de individuos que han generado tweets tanto dentro de Madrid capital como fuera de las fronteras del estudio.

Las **170.854** observaciones obtenidas posteriormente al filtrado geográfico han sido originadas por un total de **29.863** individuos. Ademas, **4.867** de estos individuos (16.3% del total) también generaron tweets fuera de los limites definidos. El código R para la consulta de la cantidad de individuos que generaron los tweets en relación a su ubicación en Madrid se puede consultar por medio del siguiente link .

# 3. Definición de movilidad.

Uno de los objetivos de este trabajo incluye el estudio de los patrones de movilidad de los individuos dentro de Madrid capital. Para determinar la movilidad de cada uno de los individuos, es necesario establecer una definición de movilidad que permita depurar el conjunto de datos y poder trabajar únicamente con aquellos que reflejen claramente patrones de movilidad.

Un individuo $u$ refleja movilidad cuando dos tweets consecutivos ( $tw_1$ , $tw_2 \in u$) las coordenadas desde donde se generaron dichos tweets son diferentes ( $l_1 \neq l_2$ ) independientemente del tiempo $t$ que ha transcurrido entre ambos tweets.

De esta forma, si el individuo $u$ genera en el instante $t_1$ el tweet $tw_1$ desde la ubicación $l_1$ y seguidamente genera en el instante $t_2$ el tweet $tw_2$ también desde la ubicación $l_1$, por la definición anterior, se observa que no existe movilidad para el individuo $u$ entre el instante $t_1$ y el instante $t_2$, por lo que la observación del $tw_2$ del usuario $u$ se eliminan del conjunto de datos ya que no aporta información de valor de acuerdo con los objetivos planteados en este estudio.

```
geotweetsmad <- arrange(geotweetsmad,desc(user_id_str))# se ordena el con
junto de datos por tiempo de forma ascendente y agrupado por user_id_str
head(geotweetsmad, n=23)
```

```
##         lat       lon                       created_at user_id_str
## 1  40.37776 -3.712665 Wed Feb 15 21:00:38 +0000 2017   999885253
## 2  40.40728 -3.709788 Sat Feb 11 20:23:12 +0000 2017    99962353
## 3  40.41356 -3.682056 Wed Feb 08 13:42:32 +0000 2017    99961966
## 4  40.53330 -3.766670 Wed Feb 22 20:01:41 +0000 2017   999512318
## 5  40.42019 -3.703841 Mon Feb 06 23:03:26 +0000 2017   999412620
## 6  40.42803 -3.698280 Tue Feb 07 08:18:40 +0000 2017   999412620
## 7  40.41356 -3.682056 Tue Feb 07 10:48:11 +0000 2017   999412620
## 8  40.42691 -3.699794 Thu Feb 23 00:50:35 +0000 2017   999412620
## 9  40.40954 -3.693172 Sun Feb 12 19:53:40 +0000 2017   999260946
## 10 40.41842 -3.696590 Sun Feb 19 20:37:14 +0000 2017   999181784
## 11 40.43046 -3.714312 Tue Feb 14 13:55:10 +0000 2017    99911876
## 12 40.46690 -3.689200 Sat Feb 18 08:38:13 +0000 2017    99890424
## 13 40.46690 -3.689200 Sat Feb 18 08:44:05 +0000 2017    99890424
```

```
## 14 40.40000 -3.683330 Sat Feb 18 16:37:25 +0000 2017    99890424
## 15 40.41941 -3.702898 Sun Feb 19 18:00:03 +0000 2017    99890424
## 16 40.41941 -3.702898 Sun Feb 19 19:33:36 +0000 2017    99890424
## 17 40.41941 -3.702898 Sun Feb 19 22:08:22 +0000 2017    99890424
## 18 40.41941 -3.702898 Mon Feb 20 12:10:09 +0000 2017    99890424
## 19 40.41941 -3.702898 Mon Feb 20 17:15:19 +0000 2017    99890424
## 20 40.41941 -3.702898 Mon Feb 20 23:04:11 +0000 2017    99890424
## 21 40.41941 -3.702898 Tue Feb 21 11:06:00 +0000 2017    99890424
## 22 40.40000 -3.683330 Wed Feb 08 14:26:08 +0000 2017   998306774
## 23 40.40000 -3.683330 Sat Feb 11 23:58:23 +0000 2017   998306774
```

Después de ordenar el conjunto de datos, se busca las posiciones que ocupan las observaciones que no cumplen con el criterio de movilidad.

```
myvector<-NULL
for (i in 1:length(geotweetsmad$user_id_str)) {
  if(is.na(geotweetsmad$user_id_str[i+1])) {break}
  else {
    if(geotweetsmad$user_id_str[i]==geotweetsmad$user_id_str[i+1] && geot
weetsmad$lat[i]==geotweetsmad$lat[i+1] && geotweetsmad$lon[i]==geotweetsm
ad$lon[i+1])
     { myvector[i] <- i } # Almacena las posiciones que no cumplen con la
 condición de movilidad.
  }
}
myvector <- na.omit(myvector)%>%as.vector()
head(myvector, n=20)

##  [1] 12 15 16 17 18 19 20 22 23 24 25 44 45 46 69 71 81 82 86 87
```

A modo de ejemplo se indican las primeras 20 posiciones que no cumplen el criterio de movilidad. Por ejemplo, la observación 12 y la observación 13 que corresponden a dos tweets consecutivos y que pertenecen al individuo $u = 99890424$, fueron generadas ambas desde la misma ubicación $l = (40.46690, -3.689200)$.

A continuación se eliminan estas observaciones del conjunto de datos.

```
geotweetsmad.mov <- geotweetsmad[-c(myvector),] # Se eliminan las observa
ciones que no cumplen con el criterio de movilidad
head(geotweetsmad.mov,n=20)

##         lat       lon                       created_at user_id_str
## 1  40.37776 -3.712665 Wed Feb 15 21:00:38 +0000 2017   999885253
## 2  40.40728 -3.709788 Sat Feb 11 20:23:12 +0000 2017    99962353
## 3  40.41356 -3.682056 Wed Feb 08 13:42:32 +0000 2017    99961966
## 4  40.53330 -3.766670 Wed Feb 22 20:01:41 +0000 2017   999512318
## 5  40.42019 -3.703841 Mon Feb 06 23:03:26 +0000 2017   999412620
## 6  40.42803 -3.698280 Tue Feb 07 08:18:40 +0000 2017   999412620
## 7  40.41356 -3.682056 Tue Feb 07 10:48:11 +0000 2017   999412620
## 8  40.42691 -3.699794 Thu Feb 23 00:50:35 +0000 2017   999412620
```

```
## 9  40.40954 -3.693172 Sun Feb 12 19:53:40 +0000 2017   999260946
## 10 40.41842 -3.696590 Sun Feb 19 20:37:14 +0000 2017   999181784
## 11 40.43046 -3.714312 Tue Feb 14 13:55:10 +0000 2017    99911876
## 13 40.46690 -3.689200 Sat Feb 18 08:44:05 +0000 2017    99890424
## 14 40.40000 -3.683330 Sat Feb 18 16:37:25 +0000 2017    99890424
## 21 40.41941 -3.702898 Tue Feb 21 11:06:00 +0000 2017    99890424
## 26 40.40000 -3.683330 Thu Feb 16 14:47:03 +0000 2017   998306774
## 27 40.42018 -3.699720 Wed Feb 08 23:57:20 +0000 2017    99819906
## 28 40.39658 -3.709462 Sat Feb 18 12:13:55 +0000 2017    99819906
## 29 40.45664 -3.604500 Sun Feb 19 07:33:14 +0000 2017     9978592
## 30 40.42516 -3.674860 Tue Feb 21 13:58:44 +0000 2017   997780370
## 31 40.39500 -3.772945 Thu Feb 16 11:25:47 +0000 2017    99742396
```

Una vez eliminados aquellos tweets que se generaron consecutivamente por un mismo individuo $u$ desde una misma ubicación ($l_1 \equiv l_2$); es posible que existan dentro del conjunto de datos $TW$, individuos que generaron un tweet una única vez o que después de aplicar la definición de movilidad solo se observa un único tweet para un individuo. Lógicamente, un único tweet asociado a un individuo no aporta información de movilidad por lo que también se procede a eliminar del conjunto de datos todas aquellas observaciones que cumplan con esta condición:$\forall\, u \in TW, count\,(u) = 1$.

```
frecuencia<-table(geotweetsmad.mov$user_id_str)
obs<-as.data.frame(frecuencia[frecuencia >1])$Var1 %>%as.vector()
geotweetsmad.mov <- subset(geotweetsmad.mov,user_id_str %in% obs)
unique(geotweetsmad.mov$user_id_str)%>%length()

## [1] 12382

head(geotweetsmad.mov, n=20)
##         lat       lon          created_at        user_id_str
##       <dbl>     <dbl>              <dttm>              <dbl>
## 1  40.40000 -3.683330 2017-03-23 14:18:27 843865601298915328
## 2  40.54424 -3.692030 2017-03-24 10:10:30 843865601298915328
## 3  40.40000 -3.683330 2017-03-29 14:53:23 843865601298915328
## 4  40.40922 -3.731807 2017-03-30 23:00:27 843598339413327872
## 5  40.40953 -3.731886 2017-04-03 18:45:58 843598339413327872
## 6  40.40568 -3.694497 2017-03-17 19:09:06 842124075006844928
## 7  40.42588 -3.711701 2017-03-18 11:38:37 842124075006844928
## 8  40.43506 -3.695400 2017-03-18 14:19:32 842105837829140480
## 9  40.43944 -3.690692 2017-03-18 14:26:37 842105837829140480
## 10 40.40000 -3.683330 2017-03-30 11:17:21 842105837829140480
## 11 40.42596 -3.687197 2017-03-16 12:58:54 841781181800685568
## 12 40.42599 -3.687086 2017-03-17 13:49:52 841781181800685568
## 13 40.43409 -3.701156 2017-04-01 19:10:12 841781181800685568
## 14 40.39502 -3.646300 2017-03-18 16:48:53 841295983044616192
## 15 40.40000 -3.683330 2017-03-24 13:00:59 841295983044616192
## 16 40.41119 -3.642271 2017-03-26 16:18:36 841295983044616192
## 17 40.39502 -3.646300 2017-03-28 19:38:19 841295983044616192
## 18 40.32444 -3.710556 2017-03-29 00:00:59 841295983044616192
```

```
## 19 40.45000 -3.616670 2017-03-29 15:15:00 841295983044616192
## 20 40.43704 -3.607120 2017-03-29 18:52:19 841295983044616192
```

El conjunto de datos $TW$, una vez depurado aplicando las definiciones de movilidad anteriormente descritas, incluyen un total de **12.382** individuos que han generado **70.684** tweets. En este punto se tiene el conjunto de datos totalmente depurado y preparado para continuar con la siguiente fase de este estudio.

De la ilustración 6 se puede observar que un 38.63% de los individuos generaron únicamente dos tweets. Así mismo, un 19.27% generaron tres tweets y un 10.62% cuatro tweets. La larga cola de esta gráfica indica una gran cantidad de tweets generada por unos pocos individuos. Se observa además como el 20% de los individuos que conforman el conjunto de datos (**2.476**) han generado un total de **41.078** tweets los cuales se corresponden con el 58% de las observaciones del conjunto de datos.

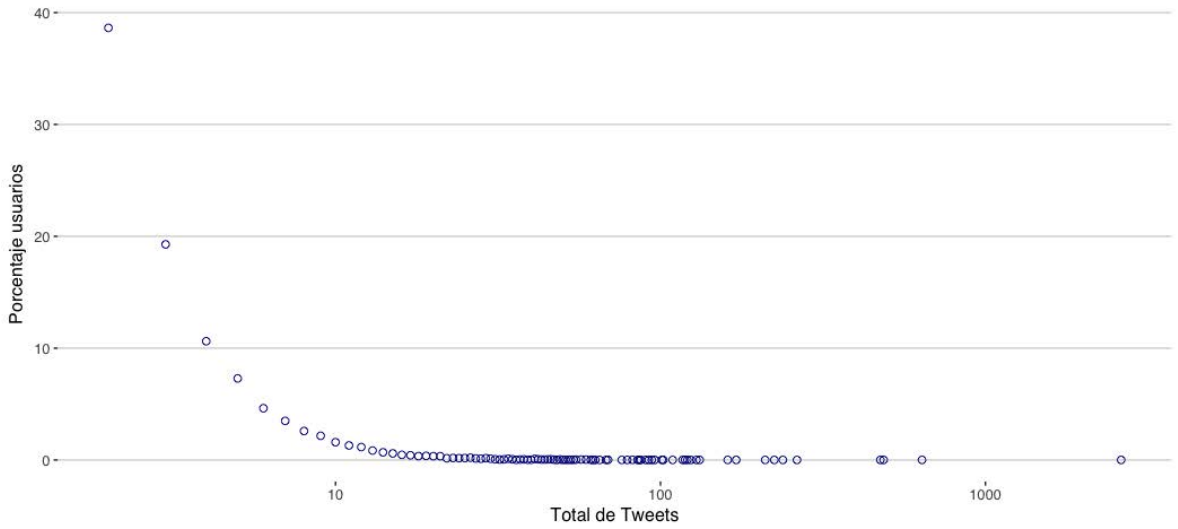



*Ilustración 6: Función de distribución de la cantidad de tweets generados por individuos.*

# 4. Asociación de categorías a tweets georreferenciados.

## 4.1. Categorización de bajo y medio nivel.

Para cada uno de los tweets, se ha obtenido entre otras variables, la ubicación desde la cual fue generado y que está representada por la latitud y longitud ($l$). El objetivo de este estudio requiere categorizar cada coordenada geográfica por medio de una semántica de bajo, medio y alto nivel de forma que permita identificar patrones de movilidad. Por sí solas, un par de coordenadas no representa nada dentro de este estudio, es por ello que es necesario realizar la categorización de estas ubicaciones.

El conocimiento semántico de una ubicación a nivel medio permite identificar cada ubicación en función de las características generales de la misma (museo, restaurante,

hospital, bar, teatro). En cambio, la identificación semántica de bajo nivel busca identificar cada ubicación bajo su nombre especifico (Aeropuerto Adolfo Suárez Madrid-Barajas, Jardines de Sabatini, Puerta del Sol).

Para la detección de las categorías y las identificación semántica de cada una de las ubicaciones se empleará la red social Foursquare[8].

Foursquare es un servicio basado en localización web aplicada a las redes sociales. En esta red social cada usuario marca su ubicación (check-in) desde su dispositivo móvil. Este proceso permite compartir la ubicación con amigos dentro de la misma red social, además de permitir explorar y buscar otras ubicaciones en función de categorías y ubicaciones más frecuentadas.

Foursquare ofrece su servicios a través de una base de datos propia que contiene todas las ubicaciones, categorías y estadísticas que sirve como motor y sobre la cual permite ofrecer sus servicios.

A través del API de Foursquare disponible para los desarrolladores es posible obtener las categorías de bajo, medio y alto nivel a partir de cada una las localizaciones ($l$). De este modo, introduciendo en la consulta API las coordenadas de cada una de las ubicaciones recogidas en los tweets; Foursquare devuelve como respuesta en formato json las ubicaciones más próximas asociadas a las coordenadas consultadas incluyendo la información más importantes para cada una, tales como: Cuenta de check-ins, Nombre de la ubicación, Categorías, dirección, código postal, distancia entre las coordenadas reales y las coordenadas pasadas en la consulta a través del API.

El código en R que permite obtener la información a través del API de la red social Foursquare se puede consultar en el siguiente link

Como resultado de pasar las coordenadas sobre la consulta del API de Foursquare, se obtiene un listado en formato json de las ubicaciones contenidas en la plataforma las cuales se encuentran ordenadas de forma creciente en función de la distancia que existe entre dichas ubicaciones y la ubicación ($l$) introducida en la consulta del API. De esta forma siempre se muestra en primer lugar la ubicación más próxima a las coordenadas pasadas sobre la consulta.

A modo de ejemplo se muestra la respuesta del API de Foursquare para una ubicación cuyo tweet se generó desde las siguientes coordenadas: 40.415556,-3.707222.

```
library(RJSONIO)
c <- paste("https://api.foursquare.com/v2/venues/search?client_id=3FW5WZU
VBHK3LD0LMD5BPTBRD0FXVRKQONJ2Z4N5JDXNN3M1&client_secret= ZWLFTSWT5NY0PDHW
```

[8] https://foursquare.com

```
4NYISXIGQB5LNN1RBWEFFCDMX4WE5PJE&limit=1&v=20170204&ll=",40.415556,",",-3
.707222,sep ="")
url <- fromJSON(c)
url
```

```
## $meta
## $meta$code
## [1] 200
## $meta$requestId
## [1] "58cedc9b9fb6b777813d3832"
## $response
## $response$venues
## $response$venues[[1]]
## $response$venues[[1]]$id
## [1] "4adcda37f964a520193c21e3"
## $response$venues[[1]]$name
## [1] "Plaza Mayor"
## $response$venues[[1]]$contact
## named list()
## $response$venues[[1]]$location
## $response$venues[[1]]$location$address
## [1] "Pl. Mayor"
## $response$venues[[1]]$location$lat
## [1] 40.41555
## $response$venues[[1]]$location$lng
## [1] -3.707353
## $response$venues[[1]]$location$distance
## [1] 11
## $response$venues[[1]]$location$postalCode
## [1] "28012"
## $response$venues[[1]]$location$cc
## [1] "ES"
## $response$venues[[1]]$location$city
## [1] "Madrid"
## $response$venues[[1]]$location$state
## [1] "Madrid"
## $response$venues[[1]]$location$country
NA
## $response$venues[[1]]$location$formattedAddress
NA
## $response$venues[[1]]$categories
## $response$venues[[1]]$categories[[1]]
## $response$venues[[1]]$categories[[1]]$id
## [1] "4bf58dd8d48988d164941735"
```

```
## $response$venues[[1]]$categories[[1]]$name
## [1] "Plaza"
## $response$venues[[1]]$categories[[1]]$pluralName
## [1] "Plazas"
## $response$venues[[1]]$categories[[1]]$shortName
## [1] "Plaza"
## $response$venues[[1]]$categories[[1]]$icon
##                                                      prefix
## "https://ss3.4sqi.net/img/categories_v2/parks_outdoors/plaza_"
##                                                      suffix
##                                                      ".png"
## $response$venues[[1]]$categories[[1]]$primary
## [1] TRUE
## $response$venues[[1]]$verified
## [1] FALSE
## $response$venues[[1]]$stats
## checkinsCount    usersCount      tipCount
##         54884         38881           400
## $response$venues[[1]]$beenHere
## lastCheckinExpiredAt
##                    0
## $response$venues[[1]]$specials
## $response$venues[[1]]$specials$count
## [1] 0
## $response$venues[[1]]$specials$items
## list()
## $response$venues[[1]]$hereNow
## $response$venues[[1]]$hereNow$count
## [1] 1
## $response$venues[[1]]$hereNow$summary
## [1] "One other person is here"
## $response$venues[[1]]$hereNow$groups
## $response$venues[[1]]$hereNow$groups[[1]]
## $response$venues[[1]]$hereNow$groups[[1]]$type
## [1] "others"
## $response$venues[[1]]$hereNow$groups[[1]]$name
## [1] "Other people here"
## $response$venues[[1]]$hereNow$groups[[1]]$count
## [1] 1
## $response$venues[[1]]$hereNow$groups[[1]]$items
## list()
## $response$venues[[1]]$referralId
## [1] "v-1489951899"
## $response$venues[[1]]$venueChains
```

El campo **$response$venues[[1]]$name** permite identificar la semántica a bajo nivel de la ubicación. En este ejemplo el nombre la ubicación corresponde a "Plaza Mayor". El campo **$response$venues[[1]]$location$distance** indica la distancia que existe entre la ubicación desde la cual se generó el tweet y la ubicación real del emplazamiento más próximo desde el cual se generó el tweet. Para este caso, la distancia es de 11 metros.

Otro campo objetivo para este estudio es el que muestra la categoría de nivel medio dentro de la cual se ubica el emplazamiento.

El campo **$response$venues[[1]]$categories[[1]]$name** muestra la categoría de nivel medio de la ubicación. Para este ejemplo, la categoría en la que Foursquare incluye a la "Plaza Mayor" es "Plaza". Este campo nos permitirá determinar la semántica a alto nivel para cada una de las ubicaciones tal y como se mostrará a continuación.

Adicionalmente, en este punto se confirma que no existe ninguna observación del conjunto de datos sin su correspondiente categoría de medio y bajo nivel. El código en R para la verificación de la asignación de cada una de las categorías para cada una de las observaciones se muestra en el siguiente link.

A partir del conjunto de datos construido hasta este punto, se busca determinar la categorización de alto nivel asociada a la categoría de nivel medio y por ende a la categoría de bajo nivel.

## 4.2. Categorización de alto nivel.

Foursquare contiene más de 900 tipos de categorías de nivel medio y que se encuentran agrupadas en 9 grupos definidos por las características intrínsecas que conforman a cada una de estas categorías. Los grupos son: **Arts&Entertainment**, **College&University, Event**, **Food**, **NightlifeSpot**, **Outdoors&Recreation, Professional&OtherPlaces, Shop&Service, Travel&Transport.**

Las categoría de alto nivel **Arts&Entertainment** incluye ubicaciones del tipo: Museum, Stadium, Zoo, Art Gallery, Opera House.

**College&University,** agrupa ubicaciones del ámbito académico o que se encuentren dentro de instalaciones académicas tales como: University, College Library, Medical School, College Residence Hall.

**Event** agrupa a todas aquellas ubicaciones no permanentes y que únicamente permanecerán activas durante un determinado periodo de tiempo. Dentro de esta categoría se incluyen: Christmas Market, Music Festival, Convention.

**Food,** es el grupo que incluye la mayor cantidad de categorías. Dentro de este grupo se incluyen todas aquellas ubicaciones relacionadas con servicios de restauración tales como:

Italian Restaurant, Latin American Restaurant, Restaurant, Mexican Restaurant, Food Truck. Cafeteria, Bakery, Sushi Restaurant y Tapas Restaurant.

El grupo **NightlifeSpot** tiene asociado todas aquellas categorías relacionadas con ubicaciones de la vida nocturna. Dentro de NightlifeSpot se incluyen las siguientes categorías: Nightclub, Brewery, Bar, Karaoke Bar.

Dentro de la categoría de alto nivel **Outdoors&Recreation** se encuentran las siguientes categorías: Soccer Field, Gym, Garden, Park, Castle. En el ejemplo de la localización "Plaza Mayor", categorizado en el plataforma Foursquare como "Plaza"; la categoría de alto nivel para esta ubicación corresponde con Outdoors&Recreation.

La categoría de alto nivel **Professional&OtherPlaces** incluye categorías tales como: Doctor's Office, Office, Coworking Space, Convention Center, Embassy / Consulate, Hospital.

**Shop&Service** incluye categorías tales como: Bank, Supermarket, Laundry Service, Shopping Mall, Electronics Store.

**Travel&Transport** incluye todas categorías relacionadas con transporte, viajes y medios de trasporte. Dentro de esta categoría podremos encontrar: Airport Terminal, Bus Stop, Hotel, Airport, Metro Station y Train Station.

A partir del conjunto de datos construido y dado que para cada observación se tiene asociada una categoría de nivel medio, es posible entonces asociar una observación de alto nivel a cada una de las observaciones a través del listado de relaciones que aporta Foursquare entre categorías de nivel medio y categorías de alto nivel[9].

A continuación se muestra una tabla resumen de las nueve categorías de alto nivel con la cantidad total de categorías de nivel medio de Foursquare y que se agrupan dentro de dicha categoría de alto nivel.

| Categorías Alto nivel | Total Categorías Nivel Medio |
|---|---|
| Arts&Entertainment | 66 |
| College&University | 39 |
| Event | 10 |
| Food | 358 |
| NightlifeSpot | 26 |
| Outdoors&Recreation | 101 |
| Professional&OtherPlaces | 103 |
| Shop&Service | 164 |
| Travel&Transport | 54 |

*Tabla 1: Total de categorías de nivel medio incluidas en categorías de bajo nivel.*

[9] https://developer.foursquare.com/categorytree

El código en R que permite realizar la categorización de alto nivel para cada una de los observaciones se puede consultar por medio del siguiente link.

Una muestra del conjunto de datos que se empleará para continuar con el análisis se presenta a continuación. El conjunto de datos ($TW$) incluye la latitud y longitud de la ubicación ($l$) desde donde se generó el tweet, la hora a la cual se generó ($t$) y el usuario que la genero ($u$), así como la categorización de la ubicación a bajo, medio y alto nivel, representadas respectivamente como:$cb, cm, ca$. Así, $tw = (l, t, u, cb, cm, ca)\ \forall\ tw \in TW$.

| lat | lon | cat_alto | cat_medio | cat_bajo | user_id_str | created_at |
|---|---|---|---|---|---|---|
| 40.5333 | -3.76667 | Outdoors&Recreation | Neighborhood | Real Sitio de El Pardo | 999512318 | 22/02/17 21:01 |
| 40.42392195 | -3.6854465 | Food | Restaurant | Ultramarinos Quintin | 999512318 | 30/03/17 21:58 |
| 40.42019 | -3.70384138 | Shop&Service | Clothing Store | Primark | 999412620 | 07/02/17 00:03 |
| 40.42803 | -3.69828 | Travel&Transport | Hostel | U Hostel Madrid | 999412620 | 07/02/17 09:18 |
| 40.41355556 | -3.68205556 | Professional&OtherPlaces | Monument / Landmark | Palacio de Cristal del Retiro | 999412620 | 07/02/17 11:48 |
| 40.42690565 | -3.6997945 | NightlifeSpot | Nightclub | Teatro Barceló | 999412620 | 23/02/17 01:50 |
| 40.434029 | -3.689458 | Professional&OtherPlaces | Non-Profit | Fundación Rafael del Pino | 9993192 | 16/03/17 10:31 |
| 40.41257894 | -3.71168393 | Food | Vegetarian / Vegan Restaurant | Viva Burger | 9993192 | 02/04/17 15:14 |
| 40.41314084 | -3.72207666 | Arts&Entertainment | Concert Hall | Sala La Riviera | 99931034 | 08/03/17 11:19 |
| 40.42982 | -3.67104 | Food | Indian Restaurant | Tandoori Station | 99931034 | 07/03/17 15:19 |
| 40.47222222 | -3.68222222 | Travel&Transport | Train Station | Estación de Madrid-Chamartín | 99931034 | 09/03/17 18:57 |
| 40.39658214 | -3.7094619 | Outdoors&Recreation | Beach | Playa de Madrid | 99890424 | 21/03/17 17:20 |
| 40.4669 | -3.6892 | Travel&Transport | Bus Station | Intercambiador de Plaza de Castilla | 99890424 | 18/02/17 09:44 |
| 40.435883 | -3.720154 | Travel&Transport | Bus Station | Intercambiador de Moncloa | 99890424 | 12/03/17 21:40 |
| 40.41941341 | -3.70289759 | Travel&Transport | Hotel | Hotel Petit Palace Tres Cruces | 99890424 | 25/02/17 17:08 |
| 40.40126773 | -3.74662035 | Outdoors&Recreation | Neighborhood | Lucero | 99890424 | 02/04/17 00:21 |
| 40.39658214 | -3.7094619 | Outdoors&Recreation | Park | Madrid Río (Sector Central) | 99890424 | 30/03/17 01:41 |
| 40.46583333 | -3.68916667 | Outdoors&Recreation | Plaza | Plaza de Castilla | 99890424 | 25/02/17 20:46 |
| 40.43733611 | -3.72154167 | Outdoors&Recreation | Scenic Lookout | Faro de Moncloa | 99890424 | 12/03/17 23:48 |
| 40.4 | -3.68333 | Travel&Transport | Train Station | Estación de Madrid-Puerta de Atocha | 99890424 | 23/03/17 15:17 |
| 40.4 | -3.68333 | Travel&Transport | Train Station | Estación de Madrid-Puerta de Atocha | 99890424 | 18/02/17 17:37 |
| 40.41997568 | -3.70044972 | Travel&Transport | Hotel | H10 Villa de La Reina | 99835561 | 06/03/17 22:23 |
| 40.41546952 | -3.70647818 | Outdoors&Recreation | Plaza | Plaza Mayor | 99835561 | 09/03/17 00:07 |
| 40.42018 | -3.69972 | Food | French Restaurant | Petit Comité | 99819906 | 09/02/17 00:57 |
| 40.39658214 | -3.7094619 | Outdoors&Recreation | Park | Madrid Río (Sector Central) | 99819906 | 18/02/17 13:13 |
| 40.41546952 | -3.70647818 | Outdoors&Recreation | Plaza | Plaza Mayor | 997828130 | 03/03/17 15:36 |
| 40.4 | -3.68333 | Travel&Transport | Train Station | Estación de Madrid-Puerta de Atocha | 997828130 | 13/03/17 00:33 |
| 40.410905 | -3.692897 | Arts&Entertainment | Art Museum | CaixaForum Madrid | 9977882 | 01/03/17 22:04 |
| 40.4267387 | -3.711 | Food | Italian Restaurant | Jack PERCOCA | 9977882 | 01/03/17 21:52 |
| 40.47222222 | -3.56083333 | Travel&Transport | Airport | Aeropuerto Adolfo Suárez Madrid-Barajas (MAD) (Aeropuerto Adolfo Suárez Madrid-Barajas) | 99657247 | 06/03/17 10:07 |
| 40.46922859 | -3.64853556 | Outdoors&Recreation | Soccer Field | Campo De Futbol Canillas | 99657247 | 27/02/17 03:49 |
| 40.4 | -3.68333 | Travel&Transport | Train Station | Estación de Madrid-Puerta de Atocha | 99657247 | 14/03/17 20:49 |
| 40.4612217 | -3.7531428 | Outdoors&Recreation | Athletics & Sports | Parque Deportivo Puerta de Hierro | 996055734 | 20/03/17 22:37 |
| 40.42 | -3.70361 | Shop&Service | Clothing Store | Primark | 996055734 | 09/03/17 20:28 |
| 40.40641883 | -3.71017925 | Professional&OtherPlaces | Monument / Landmark | Puerta de Toledo | 996055734 | 12/02/17 14:04 |
| 40.4087 | -3.71058 | NightlifeSpot | Nightclub | Shoko Madrid | 996055734 | 25/02/17 15:32 |

*Tabla 2:Muestra del conjunto de datos principal*

Desde el siguiente link se pueden consultar todas las categorizaciones de alto, medio y bajo nivel de cada uno de los tweets del conjunto de datos sobre el mapa de Madrid. A modo de ejemplo se muestra la categorización de un tweet generado desde el la ubicación Real Jardín Botánico Alfonso XIII. (Link al mapa interactivo).

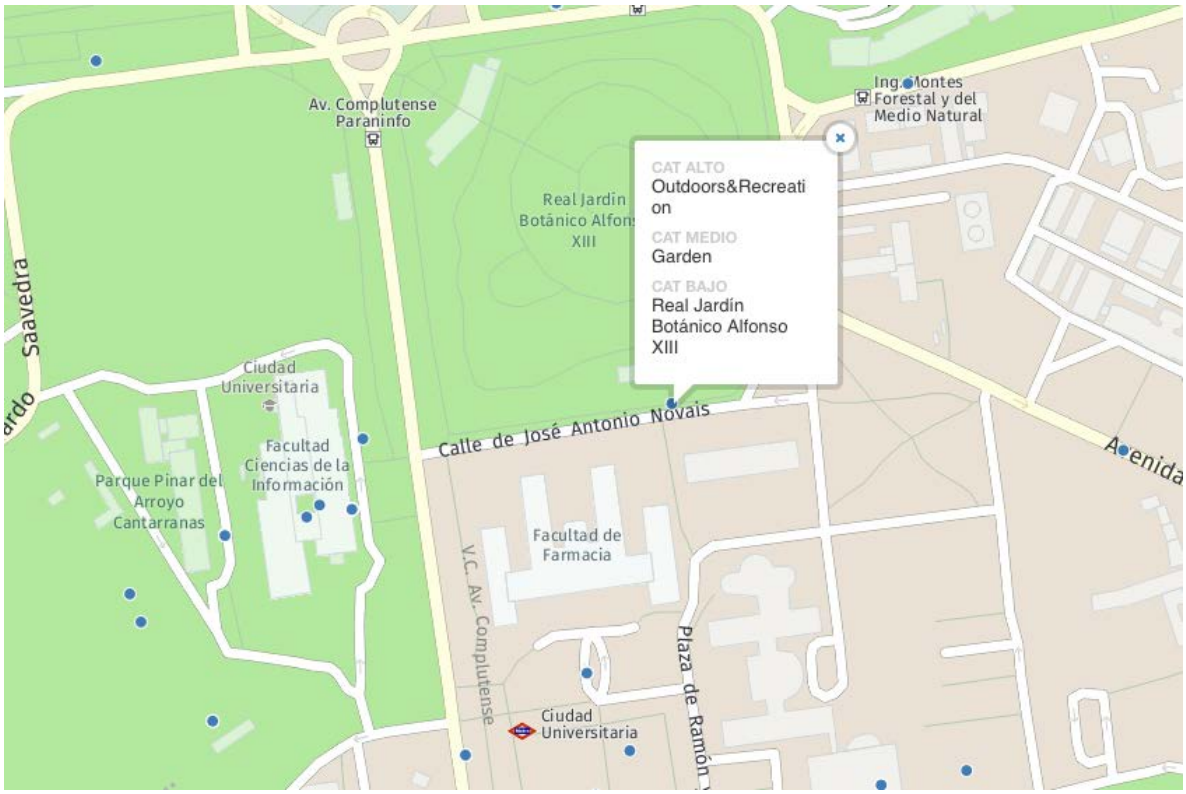



*Ilustración 7:Categorización de tweet en mapa de Madrid (Link al mapa interactivo).*

La ilustración 8 muestra la cantidad de categorizaciones de alto nivel presente en el conjunto de datos. Se observa que el mayor número de tweets han sido clasificados como **Outdoors&Recreation,** seguido de la categoria **Food** con 13.038 tweets y **Travel&Transport** con 12.937 tweets.

En la ilustración 9 se representa, a través de los recuadros interiores la proporción de categorías de nivel medio que componen el conjunto de datos. La amplitud de los recuadros exteriores representan la proporción de categorías de alto nivel del conjunto de datos. De esta forma, de 70.684 observaciones, se tiene que un 23,37% de las observaciones han sido clasificadas dentro de la categoría **Outdoors&Recreation**. Un 18% han sido clasificada tanto **Food** y **Travel&Transport**.

Apenas un 0,026% de las observaciones han sido clasificadas como Event, esto se debe a la características de las categoría de nivel medio que se agrupan dentro de esta categoría de alto nivel; tal como se indicó, dentro de esta categoría se incluye: Parade, Music Festival,

Conference, Covention; es decir evento con características móviles que tienen un presencia esporádica.

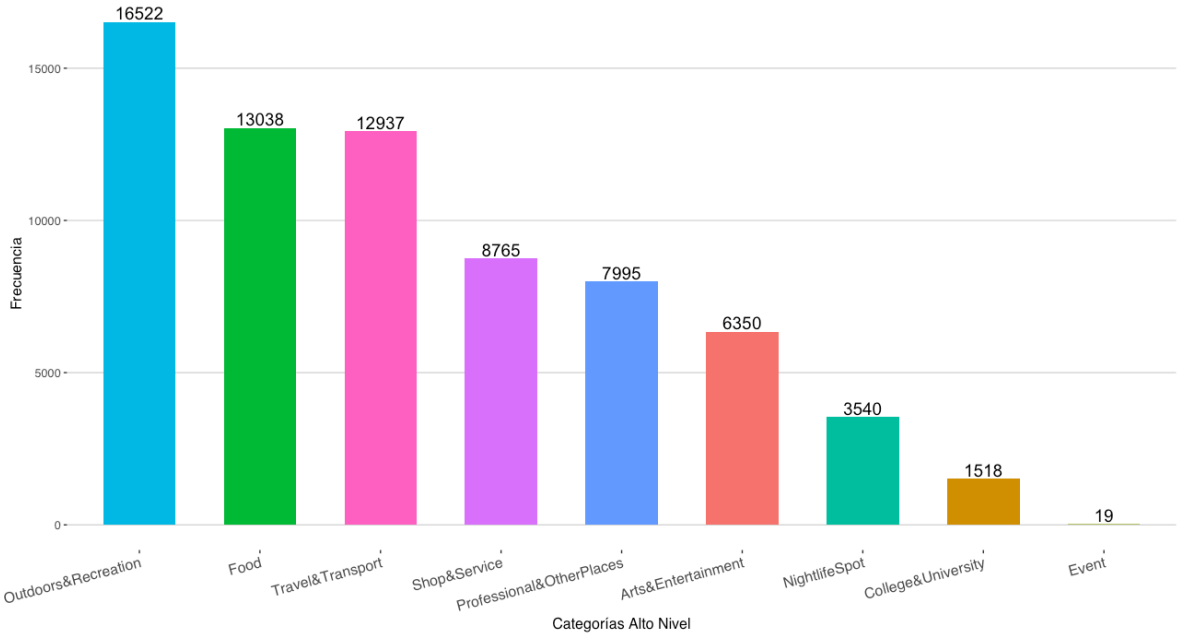



*Ilustración 8: Distribución categorías de alto nivel en el conjunto de datos.*

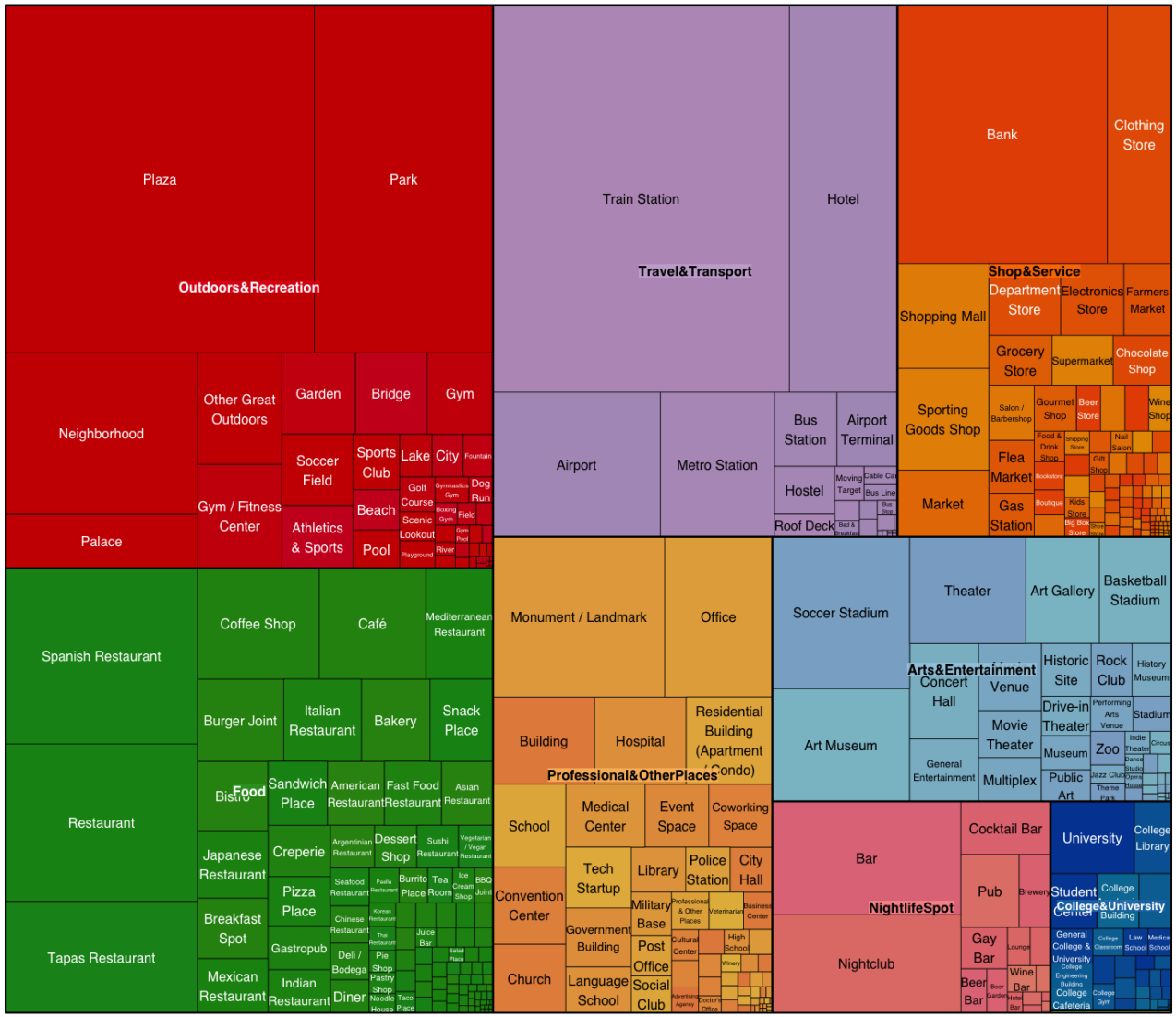



*Ilustración 9: Proporción de categorías de medio y alto nivel en conjunto de datos.*

Las **70.684** observaciones que conforman el conjunto de datos, se corresponden con un total de **5.758** ubicaciones que han sido visitadas por **12.382** individuos. De estas 5.758 ubicaciones se han seleccionados las 15 ubicaciones más frecuentadas las cuales se muestran en el siguiente gráfico.

Se observa que la categoría de bajo nivel más frecuentada corresponde con la ubicación Estación de Madrid Puerta de Atocha.

La mayoría de las ubicaciones de bajo nivel que entran dentro de las primeras 15 categorías, corresponden con ubicaciones principalmente frecuentadas por turistas, tales como el Parque del retiro, Plaza Mayor, Palacio Real de Madrid, Plaza de España y Templo de Debod.

Ubicaciones que agrupan importantes eventos deportivos y culturales tales como el Estadio Santiago Bernabéu y El palacio de Deportes de la Comunidad de Madrid ocupan las posiciones 6 y 15 respectivamente.

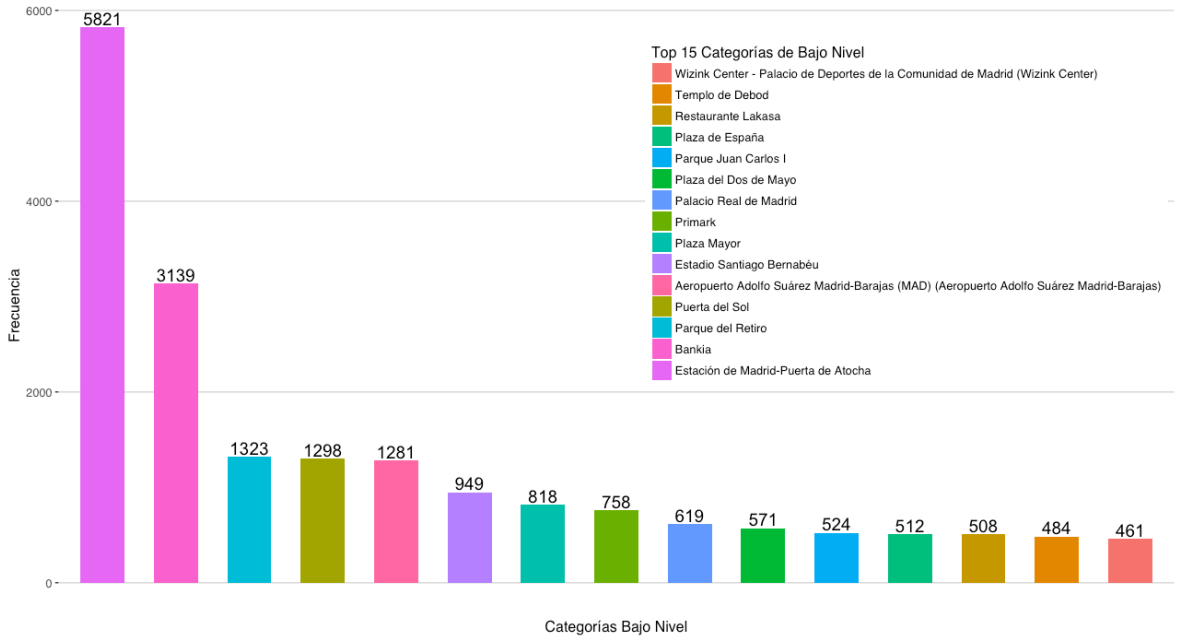



*Ilustración 10: Gráfica de frecuencia para las 15 categorías de bajo nivel más frecuentadas.*

Es interesante destacar como la segunda categoría más frecuentemente visitada con 3.139 observaciones se corresponde con Bankia. Estas observaciones fueron generadas por 1.796 individuos.

Al evaluar geográficamente todos los tweets generados desde las ubicaciones correspondientes a Bankia, se puede observar que 3.138 de estos tweets fueron creados desde una misma ubicación, la cual se corresponde con un área en la cual convergen múltiples servicio de transporte publico de Madrid (especialmente servicios de transporte

ferroviarios), por lo es probable que gran parte de estas observaciones hayan sido generadas por individuos que en determinados intervalos de tiempo estaban usando medios trasporte público de Madrid.

Es impórtate enfatizar a través de los estadísticos descriptivos la distancia existentes entre las coordenadas de cada uno de los tweets y la coordenada real de la ubicación de Bankia con la cual se categorizaron los tweets. Así, la distancia media, mediana y la desviación estándar entre las ubicaciones de los tweets y la ubicación real de Bankia son: 37mts, 36,94mts y 1.26mts respectivamente

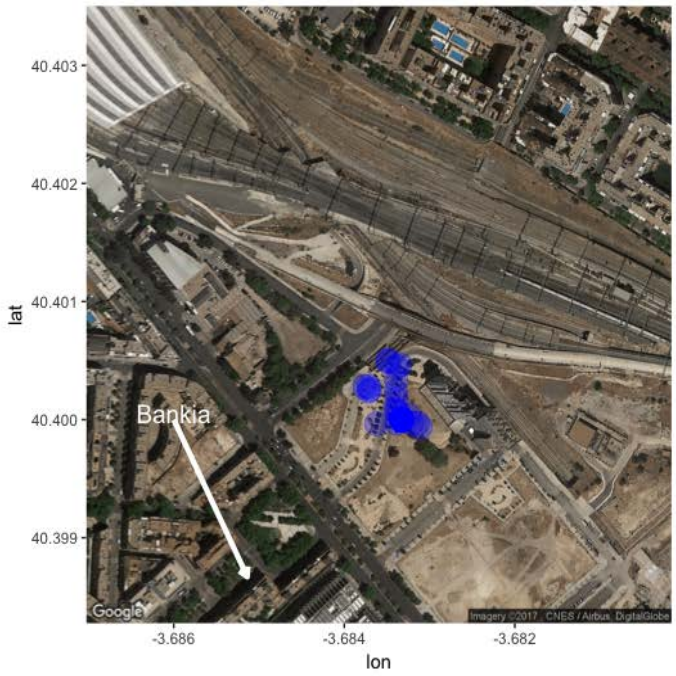



*Ilustración 11:Ubicacion geográfica de tweets con categorización de bajo nivel igual a Bankia.*

En la ilustración 11 se observa la ubicación de los 3.138 tweets cuya categorización de bajo nivel asignada es Bankia y la ubicación real de Bankia. Todas estas observaciones se encuentra muy próximas entre sí, además el entorno sobre el que se originan estas observaciones no muestra ninguna otra ubicación la cual pueda ser asociada como categoría de bajo nivel a alguno de estos tweets.

Tal como se puntualizó anteriormente, es posible que alguna se estas observaciones estén relacionadas a individuos que se encuentra desplazándose por medio del transporte público; tal y como se muestra en la figura anterior, las ubicaciones de estos tweets se encuentran muy próximas a intersecciones de vías férreas. A pesar de la imprecisión intrínseca del GPS de un dispositivo móvil la cual pueda hacer que la ubicación real desde donde se generó el tweet pueda verse desplazada en algunos metros, es lógico pensar que a pesar de la impresión del GPS, los tweets se encuentren distribuidos a lo largo de las vías férrea y no

en un área con todas las observaciones muy próximas entre sí. Es por ello que para efectos de este estudio, todos estos tweets mantendrán la categorización de Bankia.

En la ilustración 12 se muestran las 15 categorías de nivel medio más representadas. Coincidiendo con la categoría de bajo nivel más representada (Estación de Madrid Puerta de Atocha), la categoría de nivel medio más frecuente se corresponde con Estación de tren (Train Station).

Se observa además la alta presencia de categorías frecuentas por turistas tales como: Hotel, Monument/Landmark y Airport. Adicionalmente, se aprecia un alto número de tweets relacionados con alimentación y gastronomía tales como Spanish restaurant, Tapas, Restaurant y Bar.

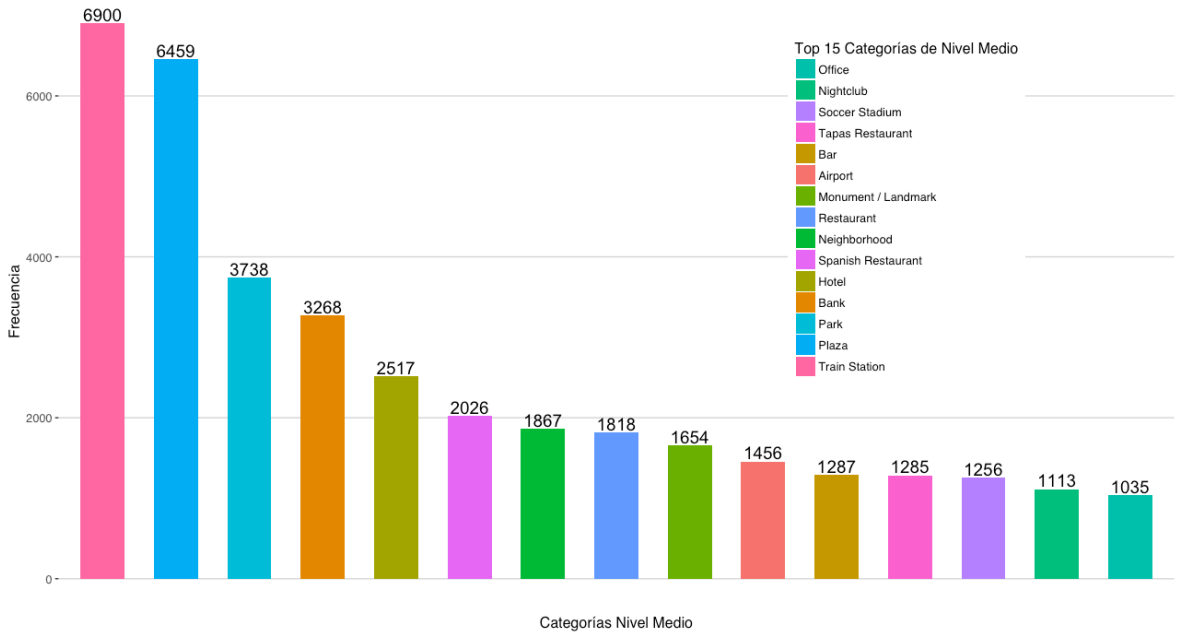



*Ilustración 12: Gráfica de frecuencia para las 15 categorías de nivel medio más frecuentadas.*

En el esquema que se muestra a continuación se resume cada uno de los pasos llevados a cabo para la construcción del conjunto de datos.

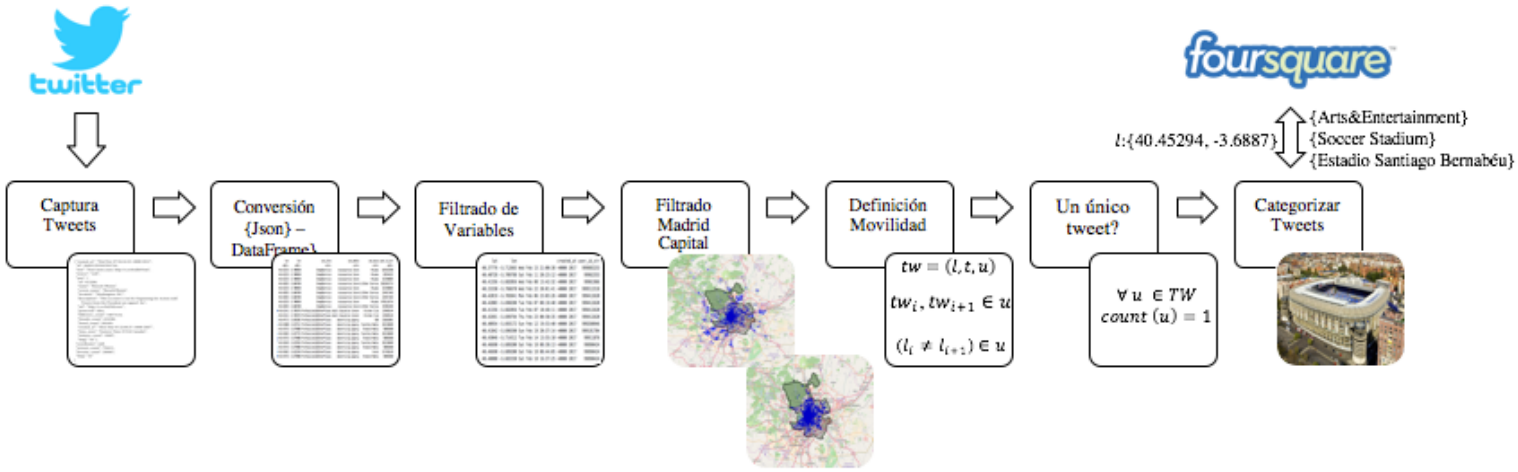



*Ilustración 13: Esquema resumen de recolección, filtrado y categorización de Tweets.*

# CAPÍTULO IV– MINADO DE ASOCIACIONES DE TRAYECTORIA.

## 1. Introducción

A continuación se describe la extracción de los patrones de asociación a partir del conjunto de datos recolectados. El problema de minado de los patrones de trayectorias puede ser modelado como una extensión del análisis de asociación.

El análisis de asociación es un método que nos permite encontrar relaciones de interés dentro de un conjunto de datos transaccional. Por lo general, el análisis de asociación es empleado por empresas tales como supermercados y minoristas para encontrar relaciones ocultas dentro de datos generados a partir de compras, envíos, órdenes y ventas.

En nuestro trabajo, el empleo del análisis de asociación nos permitirá descubrir las asociaciones entre las ubicaciones más importantes y frecuentadas por los individuos de Madrid capital.

Existen varios algoritmos que permiten inspeccionar un conjunto de datos transaccionales y obtener las asociaciones más relevantes. En este estudio se empleará el algoritmo Apriori que se incluye dentro del paquete arules[10] de R.

A continuación se muestra un ejemplo del formato de datos transaccionales sobre la categorización de bajo nivel de las ubicaciones y tomando como identificador de la transacción el identificador que twitter proporciona a cada uno de sus usuarios (user_id_str), de esta forma, cada transacción representa las ubicaciones que fueron visitadas por cada uno los individuos sin tener en cuenta la componte temporal, es decir, cada una de las transacciones incluyen todas las ubicaciones visitadas por cada uno de los individuos sin tener en cuenta el día, hora o intervalo horario en la cual se realizó la visita a cada una de las ubicaciones.

```
$`100044026`

[1] "Estación de Cercanías de Madrid-Atocha" "Gasolinera Repsol"
             "Viking Rock"

$`1001848674`

[1] "Estación de Cercanías de Madrid-Atocha" "Hotel Praga"

$`100257934`
```

[10] https://cran.r-project.org/web/packages/arules/arules.pdf

```
[1] "Goiko Grill"  "Harvard"  "La Maso Sport Center - Padel"

[4] "La Neomudéjar" "Lizarrán"  "Parque Forestal de Valdebebas"

[7] "Planetario de Madrid"

$`100264375`

[1] "Hotel Silken Puerta Madrid" "Parque Juan Carlos I"

$`1003159226`

[1] "C.C. Las Rosas" "Hard Rock Cafe Madrid" "Pueblo Nuevo"

$`1004021640`

[1] "Plaza Mayor"  "Teatro Rialto"

$`1111768654`

[1] "Estación de Cercanías de Madrid-Atocha" "Hotel Praga"
"Goiko Grill"
```

Es necesario en este punto definir algunos conceptos que nos permitirán comprender los parámetros usados por el algoritmo Apriori así como los resultados obtenidos .

El individuo representado por el user_id_str 1004021640 visitó las siguientes ubicaciones: "Plaza Mayor" y "Teatro Rialto". El conjunto de elementos que conformar una transacción se les denomina Itemset. El Itemset se denota como $IS_n$, donde $n$ representa el enésimo número dentro del conjunto de datos transaccionales. Así, teniendo en cuenta el listado de transacciones mostrado anteriormente, se tiene $IS_6 = \{\text{Plaza Mayor}, \text{Teatro Rialto}\}$

Las reglas generadas a partir del algoritmo del análisis de asociaciones se componen por sentencias que incluyen un lado derecho (RHS) y un lado izquierdo (LHS). Estas sentencias indican que al tener un Itemset en el LHS es muy probable que el individuo visite o se desplace a las ubicaciones incluidas en el Itemset del RHS. Las reglas de asociación se denotan como: $IS_x \rightarrow IS_y$, he indican que: si han visitado las ubicaciones del ítemset $x$ es muy probable que se visiten también las ubicaciones del ítemset $y$. Un ejemplo de una regla puede ser: {Plaza Mayor} $\rightarrow$ {Hard Rock Cafe Madrid"}, lo que indica que: un individuo que visita la Plaza Mayor, muy probablemente también visitará el Hard Rock Cafe Madrid.

A continuación se describen las métricas más relevantes usadas para evaluar las asociaciones obtenidas en el análisis del minado de asociaciones sin tomar en cuenta la componente temporal.

## 1.1. Frecuencia

La frecuencia (frecuency) es el número de veces que un particular itemset ocurre en la lista de todas las transacciones. Se puede denotar como $f(IS_n)$, donde $IS_n$ es un particular itemset y $f()$ la función que devuelve la frecuencia de ese itemset dentro de todo el conjunto de datos transaccionales. Teniendo en cuenta el conjunto de datos transaccionales del ejemplo:

$f\left(IS\{Estación\ de\ Cercanías\ de\ Madrid— Atocha, Hotel\ Praga\}\right) = 2$ .Ya que se observa que ambos itemset se repiten dos veces, una vez en cada uno de los siguientes usuarios: user_id_str=1111768654 y user_id_str=1001848674.

## 1.2. Support

La medida Support consiste en el número de veces en el que un particular itemset fue visitado (frecuencia) dividido por el total de transacciones que componen el conjunto de datos, y se denota como: $S(IS_n) = \frac{f(IS_n)}{count(\sum_{i=1}^{n} IS_i)}$, donde $S()$ representa el Support del itemset $IS_n$.Continuando con el ejemplo anterior, el valor del support está dado por:

$S\left(IS\{Estación\ de\ Cercanías\ de\ Madrid— Atocha, Hotel\ Praga\}\right) = \frac{2}{7}$, que es igual a un 28.57%. El support tiene como objetivo medir la calidad de la regla detectando lo que ya sucedió en transacciones anteriores. Cuando se evalúan los resultados de las reglas de asociación se busca que los valores de Support sean lo más elevados posibles.

## 1.3. Confidence

El valor de Confidence de una regla de asociación está definido como la probabilidad de que para una nueva transacción que contiene el itemset en el LHS de la regla, la transacción también contenga el itemset en el lado derecho (RHS) de la regla. Confidence se denota como $C()$ y está dado por $C\left(IS_x \rightarrow IS_y\right) = \frac{S(IS_x \cup IS_y)}{f(IS_x)}$. Valores grandes de confidence representa altas probabilidades de que un individuo que visitó una localización también visite otra localización. Así, el valor de Confidence permite detectar la calidad de la regla predicción basándose en que sucederá en el futuro a partir del datos transaccionales anteriores.

Teniendo en cuenta la regla $\{Estacion\ de\ Cernanias\ de\ Madrid \rightarrow Hotel\ Praga\}$ el valor de Confidence que la describe esta datos por:

$$S(IS\{Estación\ de\ Cercanías\ de\ Madrid— Atocha, Hotel\ Praga\}) = \frac{2}{7}$$

$$S(IS\{Estación\ de\ Cercanías\ de\ Madrid— Atocha\}) = \frac{2}{7}$$

Así, el valor de Confidence para esta regla es

$$C(Estacion\ de\ Cernanias\ de\ Madrid \rightarrow Hotel\ Praga) = 1$$

### 1.4. Lift

El lift es otra métrica que permite evaluar la calidad de una regla. Si el lift es mayor a 1, indica entonces que la presencia del itemset en el LHS es responsable del incremento en la probabilidad de que un individuo vaya a visitar las ubicaciones del itemset del RHS. Si el valor del lift es igual a 1, implica que los ítemsets en el LHS y RHS son independientes, por lo que visitar un itemset no implica visitar las ubicaciones de otro itemset. Si el lift es menor a 1, indica que si un individuo tiene un itemset en el LHS, entonces la probabilidad de visitar el itemset en el RHS es relativamente baja.

El lift de una regla está dado por $L(IS_x \rightarrow IS_y) = \frac{S(IS_x \cup IS_y)}{S(IS_x) * S(IS_y)}$, donde $L()$ representa el lift de una regla.

Para la regla $\{Estacion\ de\ Cernanias\ de\ Madrid \rightarrow Hotel\ Praga\}$, se determina el valor de lift: $L(Estacion\ de\ Cernanias\ de\ Madrid \rightarrow Hotel\ Praga) = \frac{\frac{2}{7}}{\frac{2}{7}*\frac{2}{7}} = 3.5$

A diferencia de otros algoritmos que buscan asociaciones en un conjunto de datos transaccionales tales como por ejemplo Eclat o FP-Growth; Apriori ofrece varias ventajas para el minado de asociaciones dado que permite una forma rápida y sencilla entender los resultados aunque el manejo de memoria computacional no sea el más eficiente [9].

El algoritmo Apriori ejecuta una búsqueda por anchura (breadth-first) para contar los candidatos en el itemset. El proceso del algoritmo Apriori inicia encontrando los itemsets más frecuentes (conjunto de ítems que tiene un valor de Support mínimo el cual es definido como valor de entrada en el algoritmo) level-wisely. El proceso inicia buscando el listado de frecuentes 1-itemsets ( $1 - itemset = \{a\}, \{b\}, \{c\}, \{d\}$ ). Seguidamente, el proceso continua empleando el listado de frecuentes 1-itemsets para encontrar el listado de frecuentes 2-itemsets, $2 - itemset = \{a, b\}, \{b, d\}, \{c, a\}, \{d, a\}$). El proceso continua la ejecución para encontrar nuevos listados de frecuentes k+1 itemset, a partir del listado de frecuente ítem set k-itemset hasta que se encuentren nuevos k-itemset. Finalmente el proceso emplea el listado de frecuentes itemset para generar las reglas de asociaciones.

## 2. Minado de asociaciones entre ubicaciones para categorizaciones de bajo nivel.

A continuación se busca determinar los patrones de asociaciones más importantes para el conjunto de datos teniendo en cuenta las categorizaciones de bajo nivel pero sin considerar la componente temporal asociada a dichas categorizaciones.

A través del código R que se puede acceder por medio del siguiente link, se seleccionan las variables cat_bajo y user_id_str del conjunto de datos principal y se crean un conjunto de datos transaccional el cual contiene un total de 12.382 transacciones.

A continuación se buscan las reglas de asociación que tengan un valor mínimo de Support igual a **0,001** y un valor mínimo Confidence de **0.5**. Con estos parámetros, se obtiene un total de 69 reglas.

Por lo general, entre las reglas generadas se pueden encontrar reglas repetidas o redundantes, por ejemplo una regla puede ser una súper regla o un subconjunto de una regla, es por ello que se procede a eliminar las reglas redundantes.

En la tabla 3, se listan todas las reglas de movilidad obtenidas después de la eliminación de las reglas redundantes y a partir de los parámetros de mínimo Support=0,001 y Confidence=0.05 indicados como parámetros en el algoritmo Apriori. Como resultado se obtienen un total de 58 reglas las cuales se han ordenado de forma decreciente por medio del valor Lift década una de estas.

Cada una de las reglas obtenidas revela mayoritariamente patrones de movimientos entre ubicaciones de gran importancia histórica y cultural dentro de Madrid capital, lo que hace muy evidente que sean patrones de asociaciones entre ubicaciones visitadas mayoritariamente por turistas.

Otra gran parte de las reglas de asociaciones obtenidas implican el paso por la estación de tren Puerta de Atocha y por el Aeropuerto Adolfo Suarez de Madrid. En cambio, en otras, tal y como se muestra en la regla 43 de la tabla 3, se observa la asociación entre ubicaciones que únicamente agrupan estaciones de trenes, autobuses y metros.

La regla 5 (de la tabla 3), muestra la relación que existe entre las ubicaciones Mercado de San Miguel, Templo de Debod y el Palacio Real de Madrid; Así, los individuos que visiten el Templo de Debod y el Marcado de San Miguel, también visitaran el Palacio Real de Madrid. $\{\text{Mercado de San Miguel}, \text{Templo de Debod}\} \longrightarrow \{\text{Palacio Real de Madrid}\}$. La regla esta soportada por un valor de Confidence de 0.619047, lo que significa que un individuo que visite tanto el Mercado de San Miguel como el Templo de Debod, tiene una probabilidad de un 62% de visitar también el Palacio Real de Madrid. El valor del parámetro lift obtenido para esta regla es de 14.627953 ($14.627953 > 1$), por lo que la presencia del ítemset $\{\text{Mercado de San Miguel}, \text{Templo de Debod}\}$ es remposanble por el incremento en la probabilidad del que el individuo también visite $\{\text{Palacio Real de Madrid}\}$.

En el gráfico de dispersión que se muestra a continuación se presentan las 58 reglas obtenidas a partir de este análisis. Se observa que el valor de la métrica lift para cada una de

estas reglas es mayor a 1. De igual forma los valores de Confidence de las reglas obtenidas, tal y como se especificó en el algoritmo es igual o mayor a 0.5.

En este mismo gráfico se puede aprecia que la mayoría de las reglas tiene un valor de Support muy próximo a 0,001 y valores altos de Confidence. A medida que los valores de Support aumentan, los valores de Confidence de las reglas disminuyen. Así, la regla con el valor más alto de Support en este estudio es de 0.00323 y tiene asociada un valor de Confidence de 0.5. Esta regla está dado por $\{\text{Hotel Praga}\} \longrightarrow \{\text{Estacion de Madrid} - \text{Puerta de Atocha}\}$.

El valor de Confidence más elevado se obtiene en la regla 13: $\{\text{Estacion Puerta de Atocha, Plaza Mayor, Puerta de Alcalá }\} \rightarrow \{\text{Puerta del Sol }\}$ con un valor de 76.47%. Esta regla se compone por ubicaciones de tipo de transporte y ubicaciones de tipo turísticas, por lo que es evidente pensar que esta regla es una ruta principalmente frecuentada por turistas.

Existen varias formas de explotar este estudio, por ejemplo, es posible monetizar el análisis de asociación a partir de datos georreferenciados a través de motores de sugerencias el cual ofrezca a los usuarios ofertas o propuestas de desplazamientos a bajo nivel entre ubicaciones en función de los valores de confidence obtenidas en cada una de las reglas de asociación. Es posible además crear productos dirigidos a turistas que integren entradas a sitios turísticos y además abonos de desplazamiento (metros, autobús o cercanías).

| RuleID | rules | support | confidence | lift |
|---|---|---|---|---|
| 1 | {Puerta de Alcalá,Santa Iglesia Catedral de Santa Maria la Real de la Almudena} => {Palacio Real de Madrid} | 0.00104991116136327 | 0.722222222222222 | 17.0659457167091 |
| 2 | {Santa Iglesia Catedral de Santa Maria la Real de la Almudena,Templo de Debod} => {Palacio Real de Madrid} | 0.00113067355839121 | 0.7 | 16.5408396946565 |
| 3 | {Aeropuerto Adolfo Suárez Madrid-Barajas (MAD) (Aeropuerto Adolfo Suárez Madrid-Barajas),Santa Iglesia Catedral de Santa Maria la Real de la Almudena} => {Palacio Real de Madrid} | 0.00121143595541916 | 0.681818181818182 | 16.1112074947953 |
| 4 | {Palacio de Cristal del Retiro,Santa Iglesia Catedral de Santa Maria la Real de la Almudena} => {Palacio Real de Madrid} | 0.00104991116136327 | 0.65 | 15.3593511450382 |
| 5 | {Mercado de San Miguel,Templo de Debod} => {Palacio Real de Madrid} | 0.00104991116136327 | 0.619047619047619 | 14.6279534714649 |
| 6 | {Puerta del Sol,Santa Iglesia Catedral de Santa Maria la Real de la Almudena} => {Palacio Real de Madrid} | 0.00161524794055888 | 0.588235294117647 | 13.8998652896273 |
| 7 | {Plaza Mayor,Santa Iglesia Catedral de Santa Maria la Real de la Almudena} => {Palacio Real de Madrid} | 0.00169601033758682 | 0.583333333333333 | 13.7840330788804 |
| 8 | {Museo Nacional del Prado,Templo de Debod} => {Palacio Real de Madrid} | 0.00113067355839121 | 0.583333333333333 | 13.7840330788804 |
| 9 | {Plaza Mayor,Puerta del Sol,Templo de Debod} => {Palacio Real de Madrid} | 0.00113067355839121 | 0.583333333333333 | 13.7840330788804 |
| 10 | {Estación de Madrid-Puerta de Atocha,Plaza Mayor,Templo de Debod} => {Palacio Real de Madrid} | 0.00104991116136327 | 0.52 | 12.2874809160305 |
| 11 | {Puerta del Sol,Santa Iglesia Catedral de Santa Maria la Real de la Almudena} => {Plaza Mayor} | 0.00161524794055888 | 0.588235294117647 | 10.3606392770479 |
| 12 | {Estación de Madrid-Puerta de Atocha,Puerta del Sol,Templo de Debod} => {Plaza Mayor} | 0.00104991116136327 | 0.541666666666667 | 9.54042200094832 |
| 13 | {Estación de Madrid-Puerta de Atocha,Plaza Mayor,Puerta de Alcalá} => {Puerta del Sol} | 0.00104991116136327 | 0.764705882352941 | 9.31947660954145 |
| 14 | {Puerta del Sol,le cocó} => {Plaza Mayor} | 0.00104991116136327 | 0.5 | 8.80654338549075 |
| 15 | {Aeropuerto Adolfo Suárez Madrid-Barajas (MAD) (Aeropuerto Adolfo Suárez Madrid-Barajas),Templo de Debod} => {Plaza Mayor} | 0.00121143595541916 | 0.5 | 8.80654338549075 |
| 16 | {Estación de Madrid-Puerta de Atocha,Palacio de Cristal del Retiro,Plaza Mayor} => {Parque del Retiro} | 0.00104991116136327 | 0.722222222222222 | 8.54928829403017 |
| 17 | {Terminal 1} => {Aeropuerto Adolfo Suárez Madrid-Barajas (MAD) (Aeropuerto Adolfo Suárez Madrid-Barajas)} | 0.00121143595541916 | 0.555555555555556 | 7.79036114256952 |
| 18 | {Estación de Madrid-Puerta de Atocha,Terminal 4} => {Aeropuerto Adolfo Suárez Madrid-Barajas (MAD) (Aeropuerto Adolfo Suárez Madrid-Barajas)} | 0.00161524794055888 | 0.555555555555556 | 7.79036114256952 |
| 19 | {Palacio Real de Madrid,Parque del Retiro,Plaza Mayor} => {Puerta del Sol} | 0.00121143595541916 | 0.625 | 7.61687992125984 |
| 20 | {Terminal 2} => {Aeropuerto Adolfo Suárez Madrid-Barajas (MAD) (Aeropuerto Adolfo Suárez Madrid-Barajas)} | 0.00258439667048942 | 0.542372881355932 | 7.60550511154577 |
| 21 | {Palacio Real de Madrid,Plaza de Cibeles} => {Puerta del Sol} | 0.00113067355839121 | 0.583333333333333 | 7.10908792650919 |
| 22 | {Museo Nacional del Prado,Templo de Debod} => {Parque del Retiro} | 0.00113067355839121 | 0.583333333333333 | 6.90519439133206 |
| 23 | {Mercado de San Miguel,Plaza Mayor} => {Puerta del Sol} | 0.00185753513164271 | 0.560975609756098 | 6.83661417322835 |
| 24 | {Aeropuerto Adolfo Suárez Madrid-Barajas (MAD) (Aeropuerto Adolfo Suárez Madrid-Barajas),Palacio de Cristal del Retiro} => {Parque del Retiro} | 0.00129219835244471 | 0.551724137931034 | 6.53102129623525 |
| 25 | {Estadio Santiago Bernabéu,Plaza de España} => {Puerta del Sol} | 0.00104991116136327 | 0.52 | 6.33724409448819 |
| 26 | {Estación de Madrid-Puerta de Atocha,Palacio Real de Madrid,Palacio de Cristal del Retiro} => {Parque del Retiro} | 0.00113067355839121 | 0.518518518518518 | 6.13795057007294 |
| 27 | {Mercado de San Miguel,Palacio Real de Madrid} => {Puerta del Sol} | 0.00161524794055888 | 0.5 | 6.09350393700787 |
| 28 | {Estación de Madrid-Puerta de Atocha,Santa Iglesia Catedral de Santa Maria la Real de la Almudena} => {Parque del Retiro} | 0.00145372314650299 | 0.5 | 5.91873804971319 |
| 29 | {Puerto de Cabreira} => {Bankia} | 0.0025843967048942 | 0.507936507936508 | 3.501820624315006 |
| 30 | {Aeropuerto Adolfo Suárez Madrid-Barajas (MAD) (Aeropuerto Adolfo Suárez Madrid-Barajas),Real Madrid Official Store} => {Estación de Madrid-Puerta de Atocha} | 0.00104991116136327 | 0.722222222222222 | 2.69272976680384 |
| 31 | {Museo Nacional del Prado,Santa Iglesia Catedral de Santa Maria la Real de la Almudena} => {Estación de Madrid-Puerta de Atocha} | 0.00121143595541916 | 0.714285714285714 | 2.663139329806 |
| 32 | {El Rincón Secreto} => {Estación de Madrid-Puerta de Atocha} | 0.00113067355839121 | 0.7 | 2.60987654320988 |
| 33 | {Gourmet Experience,Parque del Retiro} => {Estación de Madrid-Puerta de Atocha} | 0.00104991116136327 | 0.619047619047619 | 2.30805408583186 |
| 34 | {Plaza del Dos de Mayo,Real Madrid Official Store} => {Estación de Madrid-Puerta de Atocha} | 0.00129219835244471 | 0.615384615384615 | 2.29439696106363 |
| 35 | {Ana La Santa,Puerta del Sol} => {Estación de Madrid-Puerta de Atocha} | 0.00113067355839121 | 0.608695652173913 | 2.26945786366076 |
| 36 | {Real Madrid Official Store,Starbucks} => {Estación de Madrid-Puerta de Atocha} | 0.00145372314650299 | 0.6 | 2.23703703703704 |
| 37 | {Verbena Bar} => {Estación de Madrid-Puerta de Atocha} | 0.00226134711678243 | 0.583333333333333 | 2.17489711934156 |
| 38 | {Circulo de Bellas Artes,Plaza de España} => {Estación de Madrid-Puerta de Atocha} | 0.00113067355839121 | 0.583333333333333 | 2.17489711934156 |
| 39 | {Palacio Real de Madrid,Parque del Retiro,Plaza Mayor} => {Estación de Madrid-Puerta de Atocha} | 0.00113067355839121 | 0.583333333333333 | 2.17489711934156 |
| 40 | {Parque del Retiro,Real Madrid Official Store} => {Estación de Madrid-Puerta de Atocha} | 0.00177677273461476 | 0.578947368421053 | 2.15854450942217 |
| 41 | {Colegio José Pérez De Ayala} => {Estación de Madrid-Puerta de Atocha} | 0.00104991116136327 | 0.565217391304348 | 2.10735373054214 |
| 42 | {Barrio de Goya,Primark} => {Estación de Madrid-Puerta de Atocha} | 0.00104991116136327 | 0.565217391304348 | 2.10735373054214 |
| 43 | {Plaza de Castilla,Puerta del Sol} => {Estación de Madrid-Puerta de Atocha} | 0.00104991116136327 | 0.565217391304348 | 2.10735373054214 |
| 44 | {Museo Nacional del Prado,Templo de Debod} => {Estación de Madrid-Puerta de Atocha} | 0.00104991116136327 | 0.541666666666667 | 2.01954732510288 |
| 45 | {Palacio Real de Madrid,Plaza de Cibeles} => {Estación de Madrid-Puerta de Atocha} | 0.00104991116136327 | 0.541666666666667 | 2.01954732510288 |
| 46 | {Mercado de Pacífico} => {Estación de Madrid-Puerta de Atocha} | 0.00104991116136327 | 0.52 | 1.93876543209877 |
| 47 | {Aeropuerto Adolfo Suárez Madrid-Barajas (MAD) (Aeropuerto Adolfo Suárez Madrid-Barajas),Parque del Retiro,Plaza Mayor} => {Estación de Madrid-Puerta de Atocha} | 0.00104991116136327 | 0.52 | 1.93876543209877 |
| 48 | {Puerta del Sol,Real Madrid Official Store} => {Estación de Madrid-Puerta de Atocha} | 0.00218058471975448 | 0.519230769230769 | 1.93589743589744 |
| 49 | {Puerta de Alcalá,Puerta del Sol} => {Estación de Madrid-Puerta de Atocha} | 0.00218058471975448 | 0.519230769230769 | 1.93589743589744 |
| 50 | {Palacio Real de Madrid,Palacio de Cristal del Retiro} => {Estación de Madrid-Puerta de Atocha} | 0.00218058471975448 | 0.519230769230769 | 1.93589743589744 |
| 51 | {Terminal 1} => {Estación de Madrid-Puerta de Atocha} | 0.00113067355839121 | 0.518518518518518 | 1.93324188385917 |
| 52 | {El Maestro Churrero} => {Estación de Madrid-Puerta de Atocha} | 0.00129219835244471 | 0.516129032258065 | 1.92433293508562 |
| 53 | {Mercado de San Miguel,Parque del Retiro} => {Estación de Madrid-Puerta de Atocha} | 0.00161524794055888 | 0.512820512820513 | 1.91199746755302 |
| 54 | {Circulo de Bellas Artes,Puerta del Sol} => {Estación de Madrid-Puerta de Atocha} | 0.00169601033758682 | 0.512195121951220 | 1.9096657633243 |
| 55 | {Hotel Praga} => {Estación de Madrid-Puerta de Atocha} | 0.00323049588111775 | 0.5 | 1.8641975308642 |
| 56 | {Circulo de Bellas Artes,Palacio Real de Madrid} => {Estación de Madrid-Puerta de Atocha} | 0.00104991116136327 | 0.5 | 1.8641975308642 |
| 57 | {El Corte Inglés,Parque del Retiro} => {Estación de Madrid-Puerta de Atocha} | 0.00129219835244471 | 0.5 | 1.8641975308642 |
| 58 | {Parque del Retiro,Plaza Mayor,Puerta del Sol} => {Estación de Madrid-Puerta de Atocha} | 0.00153448554353093 | 0.5 | 1.8641975308642 |

*Tabla 3:Reglas espaciales para categorización de ubicaciones a bajo nivel.*

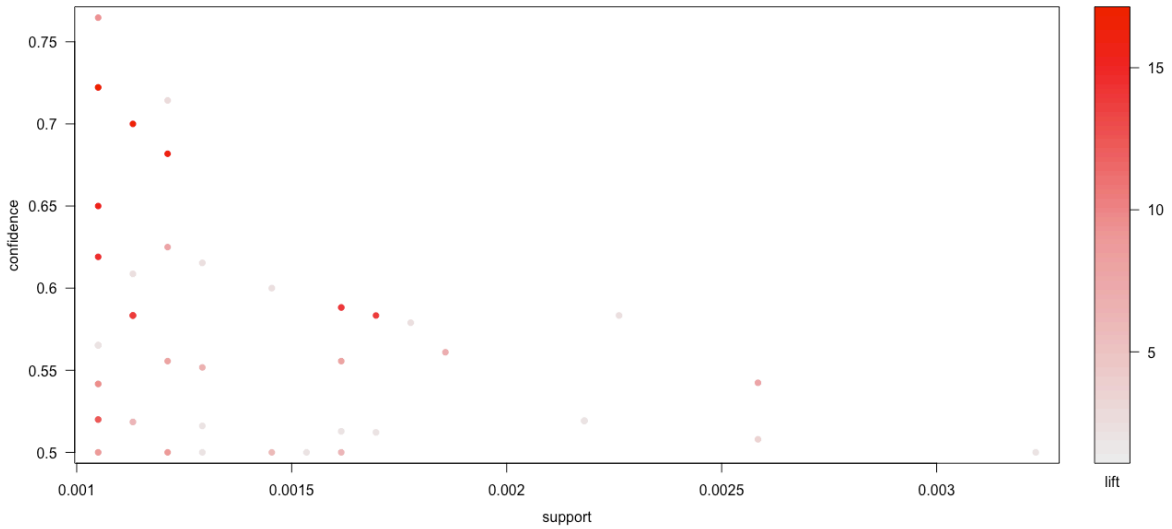



*Ilustración 14:Gráfico de dispersión para 58 reglas a partir de categorización espacial de bajo nivel.*

# 3. Minado de trayectorias para categorizaciones de nivel medio.

A continuación se busca determinar los patrones de asociación para la categorización de nivel medio sin tener en cuenta la componente temporal. A través del estudio de patrones de asociación obtenidos por medio de categorizaciones de nivel medio es posible inferir desplazamientos de usuarios en función de estas categorías.

Para la obtención de las asociaciones por medio de la categorización de nivel medio y tal como se muestra en el código R que se puede acceder por medio de este link, se ha usado como valores de entrada del algoritmo Apriori un valor de Confidence de **0.6** y de Support de **0.001**. Con estos parámetros se ha obtenido un total de 2.287 reglas a partir de un total de 12.382 transacciones.

Tal como se realizó en el estudio de asociación de movilidad a través de categorización de bajo nivel, se procede a la eliminación de la reglas redundantes. Después de la fase de eliminación, se tiene un conjunto de 1.764 reglas. Las 40 reglas más importantes ordenadas por su valor de Confidence se muestran en la tabla 4.

La regla de asociación 38, muestra que un individuo que visite un Aeropuerto, un palacio, una plaza y una estación de tren tiene un 93% de probabilidad de visitar también un museo de arte ({Airport, Palace, Park, Plaza} → {Art Museum}).

La regla 31, muestra lo que suele ser un tipo de asociaciones con características de comportamiento diario: ({Metro Station, Office, Tapas Restaurant} ⟶ {Plaza}). Un individuo

que pase por una estación de metro, por un sitio de oficina y por un restaurante, tiene un 94% de probabilidades de ir a una plaza.

En gran medida, las asociaciones obtenidas en este estudio corresponden a movimientos realizados típicamente por turistas, aunque se aprecia una cantidad importante de rutas que muestran los movimientos cotidianos de la población.

Se observa que la mayoría de las reglas presentan en su RHS la categoría Plaza. Este comportamiento se debe a que tal y como se mostró en la ilustración 11, la categoría Plaza se presenta como la segunda categoría de nivel medio más frecuente en el conjunto de datos, estando presente en 6.459 tweets.

| RulesID | rules | support | confidence | lift |
|---|---|---|---|---|
| 1 | {Bar,Mediterranean Restaurant,Restaurant} => {Spanish Restaurant} | 0.0012921983524471 | 1 | 11.5181395348837 |
| 2 | {Bar,Clothing Store,Neighborhood,Plaza} => {Park} | 0.00104991116136327 | 1 | 5.1228796028134 |
| 3 | {Breakfast Spot,Park,Restaurant} => {Plaza} | 0.00137296074947504 | 1 | 3.30539241857982 |
| 4 | {Bus Station,Hotel,Spanish Restaurant} => {Plaza} | 0.00104991116136327 | 1 | 3.30539241857982 |
| 5 | {Cocktail Bar,Monument / Landmark,Park} => {Plaza} | 0.0012921983524471 | 1 | 3.30539241857982 |
| 6 | {Bar,Department Store,Neighborhood} => {Plaza} | 0.00104991116136327 | 1 | 3.30539241857982 |
| 7 | {Building,CafÈ,Train Station} => {Plaza} | 0.00104991116136327 | 1 | 3.30539241857982 |
| 8 | {Building,Clothing Store,Monument / Landmark} => {Plaza} | 0.00104991116136327 | 1 | 3.30539241857982 |
| 9 | {Building,Hotel,Monument / Landmark} => {Plaza} | 0.00113067355839121 | 1 | 3.30539241857982 |
| 10 | {CafÈ,Coffee Shop,Monument / Landmark} => {Plaza} | 0.00121143595541916 | 1 | 3.30539241857982 |
| 11 | {CafÈ,Clothing Store,Monument / Landmark} => {Plaza} | 0.00121143595541916 | 1 | 3.30539241857982 |
| 12 | {CafÈ,Monument / Landmark,Spanish Restaurant} => {Plaza} | 0.00137296074947504 | 1 | 3.30539241857982 |
| 13 | {CafÈ,Monument / Landmark,Restaurant} => {Plaza} | 0.00137296074947504 | 1 | 3.30539241857982 |
| 14 | {Clothing Store,Neighborhood,Spanish Restaurant} => {Plaza} | 0.00113067355839121 | 1 | 3.30539241857982 |
| 15 | {CafÈ,Hotel,Park,Train Station} => {Plaza} | 0.00153448554353093 | 1 | 3.30539241857982 |
| 16 | {Bar,Coffee Shop,Park,Spanish Restaurant} => {Plaza} | 0.00104991116136327 | 1 | 3.30539241857982 |
| 17 | {Bar,Neighborhood,Spanish Restaurant,Train Station} => {Plaza} | 0.00104991116136327 | 1 | 3.30539241857982 |
| 18 | {Hotel,Park,Restaurant,Spanish Restaurant,Train Station} => {Plaza} | 0.00113067355839121 | 1 | 3.30539241857982 |
| 19 | {Monument / Landmark,Park,Spanish Restaurant,Train Station} => {Plaza} | 0.00177677273461476 | 0.956521739130435 | 3.16167970472852 |
| 20 | {Building,Monument / Landmark,Train Station} => {Plaza} | 0.00153448554353093 | 0.95 | 3.14012279765083 |
| 21 | {Monument / Landmark,Palace,Tapas Restaurant} => {Plaza} | 0.00153448554353093 | 0.95 | 3.14012279765083 |
| 22 | {Monument / Landmark,Palace,Spanish Restaurant} => {Plaza} | 0.00145372314650299 | 0.947368421052632 | 3.1314243965493 |
| 23 | {CafÈ,Hotel,Monument / Landmark} => {Plaza} | 0.00145372314650299 | 0.947368421052632 | 3.1314243965493 |
| 24 | {Bar,Coffee Shop,Neighborhood} => {Plaza} | 0.00145372314650299 | 0.947368421052632 | 3.1314243965493 |
| 25 | {Monument / Landmark,Park,Theater} => {Plaza} | 0.00137296074947504 | 0.944444444444444 | 3.1217595064365 |
| 26 | {Airport,Art Museum,Clothing Store} => {Plaza} | 0.00137296074947504 | 0.944444444444444 | 3.1217595064365 |
| 27 | {Neighborhood,Other Great Outdoors,Plaza} => {Park} | 0.0012921983524471 | 0.941176470588235 | 4.82153374382438 |
| 28 | {Art Museum,Monument / Landmark,Neighborhood} => {Plaza} | 0.0012921983524471 | 0.941176470588235 | 3.11095757042806 |
| 29 | {Art Museum,Hotel,Neighborhood} => {Plaza} | 0.0012921983524471 | 0.941176470588235 | 3.11095757042806 |
| 30 | {Bank,Clothing Store,Monument / Landmark,Park} => {Plaza} | 0.0012921983524471 | 0.941176470588235 | 3.11095757042806 |
| 31 | {Metro Station,Office,Tapas Restaurant} => {Park} | 0.00121143595541916 | 0.9375 | 4.80269962763757 |
| 32 | {Bar,Flea Market} => {Plaza} | 0.00121143595541916 | 0.9375 | 3.09880539241858 |
| 33 | {Coffee Shop,Sandwich Place} => {Plaza} | 0.00121143595541916 | 0.9375 | 3.09880539241858 |
| 34 | {Department Store,Palace} => {Plaza} | 0.00121143595541916 | 0.9375 | 3.09880539241858 |
| 35 | {Metro Station,Office,Tapas Restaurant} => {Plaza} | 0.00121143595541916 | 0.9375 | 3.09880539241858 |
| 36 | {Hotel,Metro Station,Park,Restaurant} => {Plaza} | 0.00121143595541916 | 0.9375 | 3.09880539241858 |
| 37 | {Clothing Store,Park,Restaurant,Train Station} => {Plaza} | 0.00121143595541916 | 0.9375 | 3.09880539241858 |
| 38 | {Airport,Palace,Plaza,Train Station} => {Art Museum} | 0.00113067355839121 | 0.933333333333333 | 16.4155303030303 |
| 39 | {Brewery,Park,Train Station} => {Plaza} | 0.00113067355839121 | 0.933333333333333 | 3.08503292400783 |
| 40 | {Cocktail Bar,Monument / Landmark,Spanish Restaurant} => {Plaza} | 0.00113067355839121 | 0.933333333333333 | 3.08503292400783 |

Tabla 4:Reglas espaciales para categorización de ubicaciones a nivel medio

En la ilustración 15 se muestra el gráfico de dispersión de las asociaciones obtenidas a partir de las categorías de nivel medio. Las mayoría de las reglas obtenidas tiene un valor de Support bajo. Así mismo, a medida que el valor de Support aumenta el valor de Confidence disminuye.

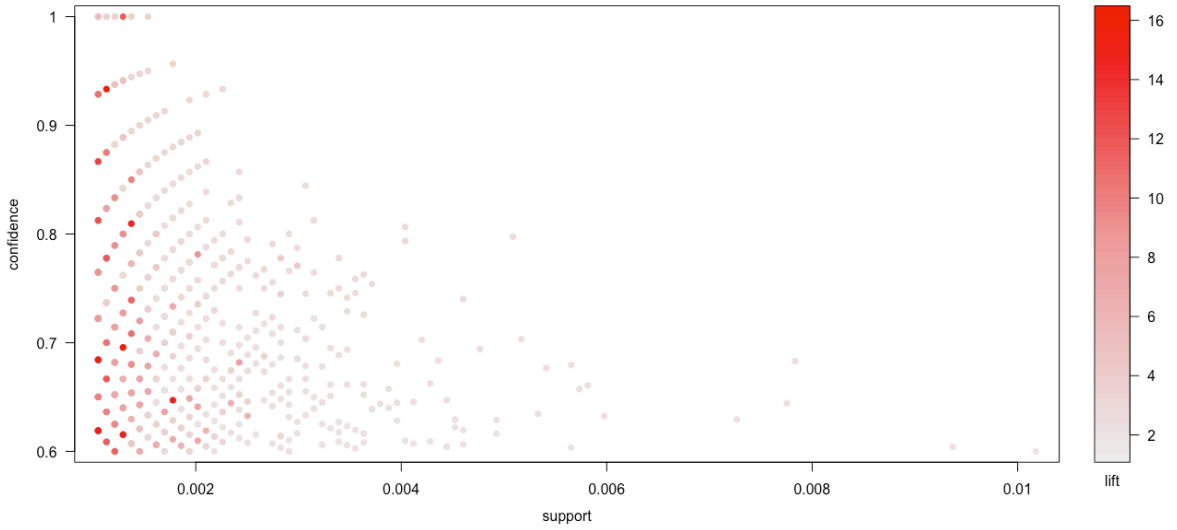



*Ilustración 15: Gráfico de dispersión para 1.764 reglas a partir de categorización de nivel medio.*

# 4. Minado de trayectorias en función de los barrios de Madrid Capital.

En los apartados anteriores se han estudiado las asociaciones referenciadas a categorías de bajo y medio nivel. En estos estudios se le han asignado a cada una de las coordenadas su correspondiente categoría, para así evaluar los patrones de movilidad asociados a las mismas.

En este apartado, se busca estudiar la movilidad de los individuos dentro de un ámbito geográfico y que tenga en cuenta las propias delimitaciones de la ciudad. Así, a partir de este idea se busca evaluar los patrones de asociaciones vinculados a los barrios de Madrid capital.

Madrid capital está dividida en total de 128 barrios. A cada uno de los 70.684 tweets incluidos en el conjunto de datos ($TW$), se le asignará a cada una de esta observaciones el barrio correspondiente, en función del polígono geográfico del barrio desde donde se generó el tweet, de esta forma se asocia a cada tweet el barrio desde donde se generó.

La imprecisión de los teléfonos móviles en el momento de posicionase por medio de su GPS y el error que se puede cometer en el momento de asociar una categoría a la ubicación real del tweet, hace que la asociación de los barrios a cada una de las observaciones sea más precisa en el momento de afirma la presencia de asociaciones de trayectoria. Esto se debe principalmente a que cada uno de los polígonos de los barrios son lo suficientemente

grandes para que en el momento de llevar a cabo la asociación entre la ubicación del tweet y el barrio correspondiente, la imprecisión intrínseca del posicionamiento GPS sea irrelevante y por ende la asociación de cada barrio sea la correcta.

En el mapa que se muestra en la ilustración 16, se presentan los barrios de Madrid, los cuales se han coloreado en función del porcentaje del número de tweets que han sido generados dentro del mismo. De igual forma, en la tabla que se muestra a continuación se presentan los 20 primeros barrios desde donde se han generado la mayor cantidad de tweets.

| Barrios | Total Tweets |
|---|---|
| Atocha | 9146 |
| Sol | 5244 |
| Universidad | 3855 |
| Palacio | 3427 |
| Justicia | 2891 |
| Jerónimos | 2811 |
| Cortes | 2639 |
| Embajadores | 2469 |
| Recoletos | 2149 |
| Casa de Campo | 1532 |
| Hispanoamérica | 1390 |
| Goya | 1359 |
| Aeropuerto | 1259 |
| Cuatro Caminos | 1153 |
| Ciudad Universitaria | 1075 |
| Corralejos | 1029 |
| Almagro | 1026 |
| Ríos Rosas | 919 |
| Argüelles | 918 |

*Tabla 5: Cantidad de tweets en barrios de Madrid.*

Se observa, que el barrio de Atocha es el barrios desde donde se han generado más tweets, con un total de 9146, que representan el 13% del total tweets del conjunto de datos. Le sigue Sol con un 7.42% y Universidad con 5.45%. Se observa, que para el resto de los barrios, el número de tweets ronda el 1 y 2%.

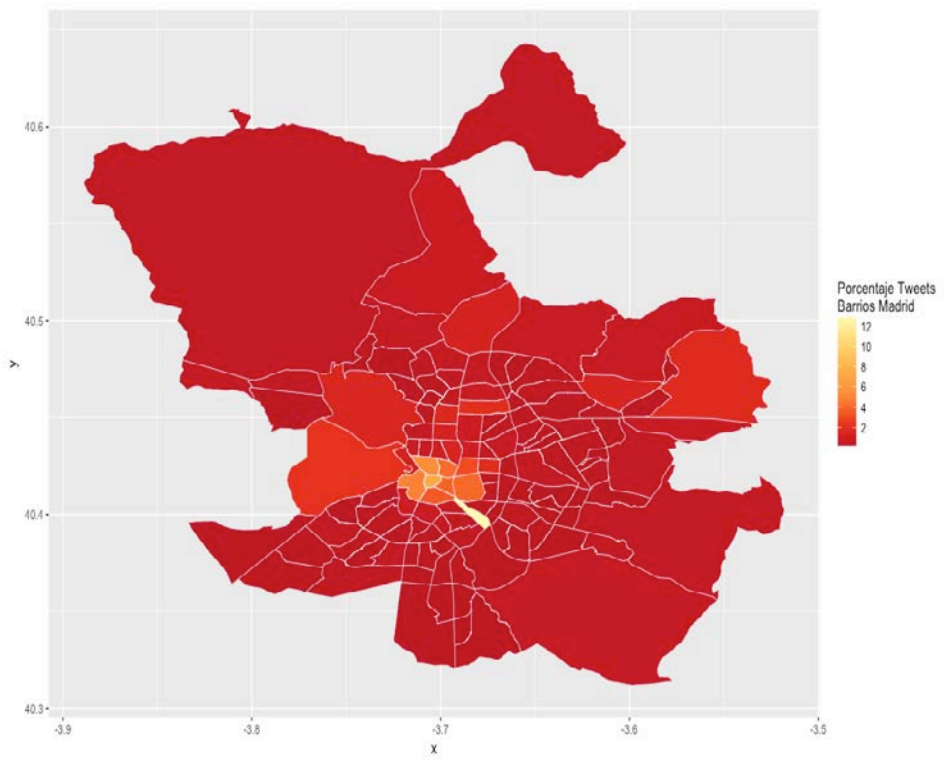
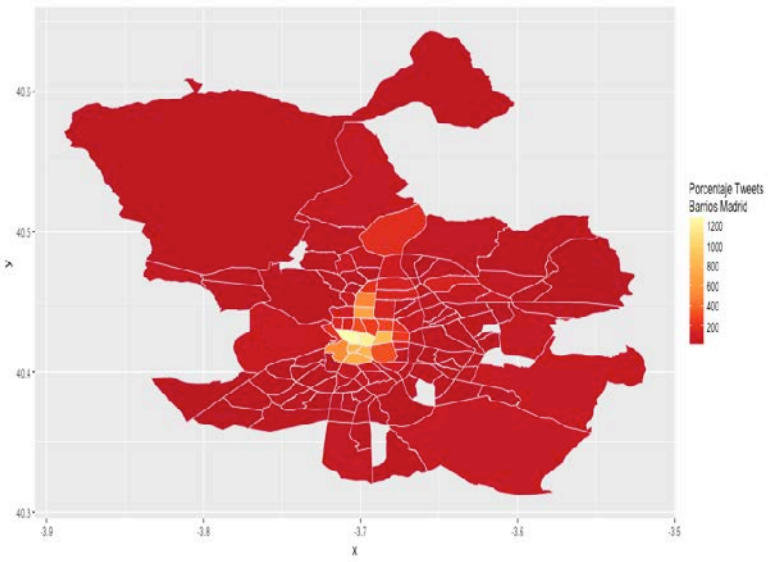



*Ilustración 16:Porcentaje de tweets en barrios de Madrid Capital*

A continuación, se evalúan el número de tweets en cada uno de los barrios, teniendo en cuenta las 9 categorías de alto nivel (Arts&Entertainment, College&University, Event, Food, NightlifeSpot,Outdoors&Recreation,Professional&OtherPlaces,Shop&Service,Travel&Transport*)*.

Con este análisis, se busca identificar cuáles de los barrios tienen más popularidad en cada una de estas 9 categorías. Debajo de cada uno los mapas se muestran los 4 primeros barrios desde los cuales se han generado el mayor número de tweets para dicha categoría. Adicionalmente, aquellos barrios desde los cuales no se hayan generado tweets, su correspondiente polígono no aparecerá trazado sobre el mapa.




| Barrio | Categoría | Total Tweets |
|---|---|---|
| Universidad | Food | 1297 |
| Justicia | Food | 1197 |
| Recoletos | Food | 853 |
| Cortes | Food | 806 |

*Ilustración 17: Total de tweets en los barrios de Madrid para la categoría Food.*

Para la categoría Food, se observa que el barrio Universidad (conocido también como barrio de las Maravilla y que se asocia al área de Malasaña), presenta el mayor número de tweets categorizados como Food. Le sigue el barrio Justicia, dentro del cual se encuentra la zona de Chueca. Tanto el barrio Universidad, como el de Justicia, se localizan dentro del distrito Centro de Madrid.

El tercer barrio con más tweets categorizados con la categoría Food, se corresponde con Recoletos. Este barrio se ubica dentro del distrito de Salamanca y además de contener una importante zona comercial de artículos de lujo (concretamente la calle Serrano), también se ubican algunos restaurantes reconocidos en la ciudad como por ejemplo el Hard Rock Café de Madrid.

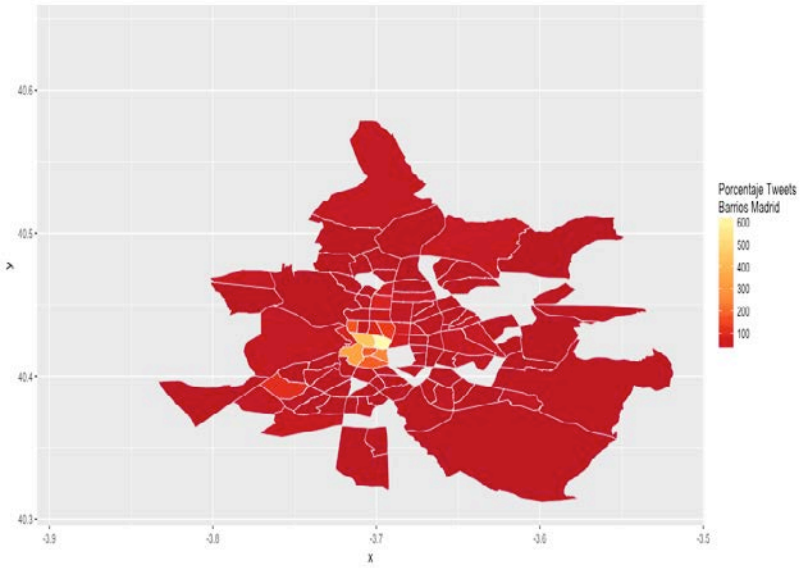
Es evidente, como a partir de este análisis, se aprecia la dinámica de la ciudad en su relación a la categoría Food y sus ubicación geográfica dentro de la misma.

Para la categoría NightlifeSpot; la mayor cantidad de tweets para esta categoría se han generado desde el barrio Justicia (Zona de Chueca), seguido por el barrio Universidad (área de Malasaña).

El tercer barrio con mayor número de tweets con la categoría NightlifeSpot, es Palacio. Este barrio es el de mayor extensión de los 6 barrios que conforman el distrito Centro. Dentro del barrio Palacio se encuentra el Palacio Real de Madrid, El Teatro Real de

Madrid, Plaza de la Villa, la Latina y las cavas (la Cava Alta y la Cava Baja), siendo estas últimas calles famosas por concentrar ubicaciones celebres de la vida nocturna de Madrid.

El cuatro barrio desde donde se generan más tweet categorizados con NightlifeSpot, corresponde a Cortes (popularmente conocido como Huertas). Este barrio, también ubicado dentro del distrito centro, alberga varias opciones de bares y sitios de copas en general.



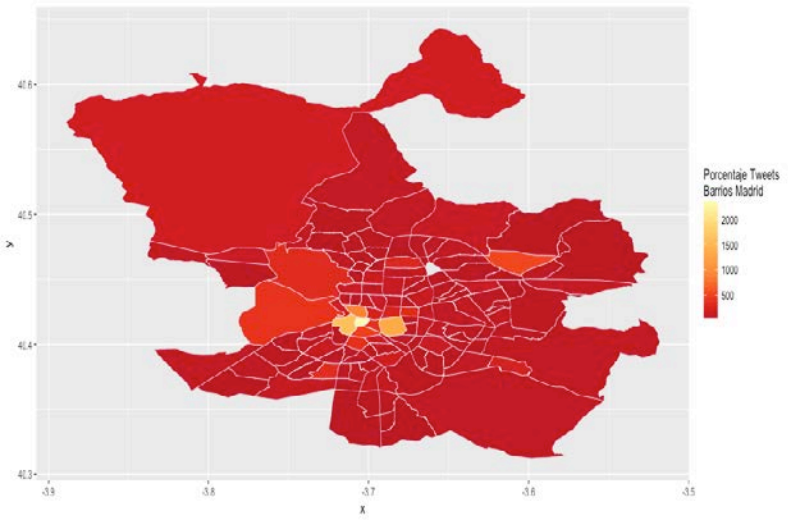

| Barrio | Categoria | Total Tweets |
|---|---|---|
| Justicia | NightlifeSpot | 634 |
| Universidad | NightlifeSpot | 448 |
| Palacio | NightlifeSpot | 331 |
| Cortes | NightlifeSpot | 246 |

*Ilustración 18:Total de tweets en los barrios de Madrid para la categoría NightlifeSpot.*

La categoría Outdoors&Recreation, incluye todas aquellas ubicaciones tales como Plazas, Jardines, Palacios, puentes, etc. El barrio con mayor número de tweets asociados a la categoría Outdoors&Recreation es Sol. Este barrio del distrito centro se caracteriza por contener ubicaciones de relevancia turísticas tales como: Plaza Mayor, Puerta del Sol, Plaza del Callao, Plaza del Carmen.

El barrio Palacio aparece como el segundo barrio con más tweets para esta categoría. Como se indicó, el barrio Palacio es el barrio con mayor extensión de los barrios del distrito centro y al igual que el barrio Sol, existen una cantidad importante de ubicaciones de relevancia turística, tales como el Palacio Real, Plaza de Oriente, Jardines del Campo del Moro, entre otros.

El tercer barrio con más tweets de la categoría Outdoors&Recreation corresponde a Jerónimos. Dentro de este barrio se puede encontrar el Parque del Retiro, El Palacio de Cristal y el Real Jardín botánico.




| Barrio | Categoría | Total Tweets |
|---|---|---|
| Sol | Outdoors&Recreation | 2351 |
| Palacio | Outdoors&Recreation | 1631 |
| JerÃ³nimos | Outdoors&Recreation | 1358 |
| Universidad | Outdoors&Recreation | 897 |

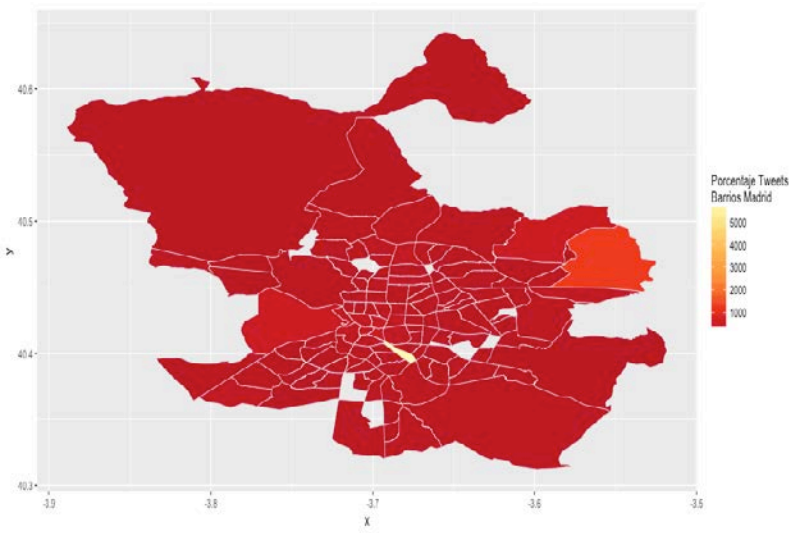
*Ilustración 19:Total de tweets en los barrios de Madrid para la categoría Outdoors&Recreation.*

La ilustración 20 muestra claramente la dinámica de los individuos asociada a la categoría Travel&Transport. El barrio Atocha, ubicado dentro del distrito de Arganzuela, incluye la célebre estación Puerta de Atocha, además dentro de este barrio se ubica el intercambiador de Méndez Álvaro (metro y tren de cercanías) y la estación sur de autobuses.

Como segundo barrio con más tweets dentro de la categoría Travel&Transport, aparece el barrio Aeropuerto, en el cual se ubica el Aeropuerto Adolfo Suarez y alguno hoteles.

El barrio Cortes (conocido también como Huertas), se posiciona como el tercer barrio con mayor número de tweets asociados a la categoría Travel&Transport, aunque en menor medida en comparación con Atocha o Aeropuerto. El barrio Cortes no se caracteriza por incluir varios medios de transporte. Únicamente se podría destacar la presencia de la estación de metro Sevilla. La característica más resaltante de este barrio es la alta concentración de hoteles, hostales y bed and breakfast. Entre los más importante destacan The Westin Palace y el ME Madrid Reina Victoria.

El barrio de Casa de Campo se ubica en la cuarta posición. Dentro de este barrio podemos encontrar la Estación Príncipe Pio la cual incluye un intercambiador de autobuses, Líneas de metros y trenes de cercanía. Además dentro del parque Casa de Campo se encuentra la estación de teleférico, el cual es considerada un medio de transporte.

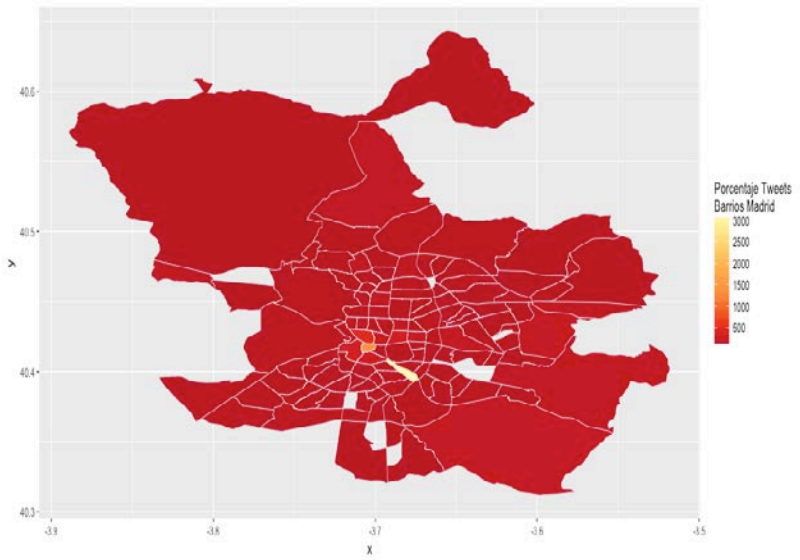



| Barrio | Categoria | Total Tweets |
|---|---|---|
| Atocha | Travel&Transport | 5899 |
| Aeropuerto | Travel&Transport | 1206 |
| Cortes | Travel&Transport | 406 |
| Casa de Campo | Travel&Transport | 388 |

*Ilustración 20:Total de tweets en los barrios de Madrid para la categoría Travel&Transport.*

La categoría Shop&Service está compuesta por ubicaciones tales como Bancos, supermercados, lavandería, centro comerciales y tiendas de equipamiento electrónico. Atocha es el barrio desde donde se generaron la mayor cantidad de tweets de la categoría Shop&Service. Este barrio, por ser uno de los más transitados, se ubican varios bancos y supermercados.

Sol, es el segundo barrio con mayor cantidad de tweets para esta categoría. En este barrio podemos encontrar varias ubicaciones incluidas en esta categoría tales como Supermercado del Corte Ingles en la calle Preciados, Tienda Fnac y el Apple Store de la Puerta del Sol.




| Barrio | Categoria | Total Tweets |
|---|---|---|
| Atocha | Shop&Service | 3140 |
| Sol | Shop&Service | 1390 |
| Universidad | Shop&Service | 440 |
| Palacio | Shop&Service | 281 |

*Ilustración 21:Total de tweets en los barrios de Madrid para la categoría Shop&Service.*

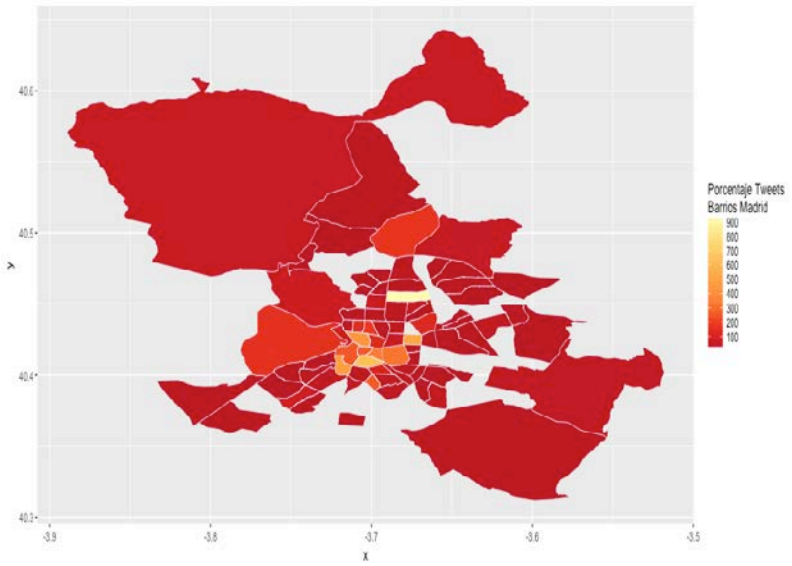
En la categoría Arts&Entertainment podemos encontrar Teatros, Museos, Estadio de futbol y baloncesto, Cines, entre otros.

Hispanoamérica aparece como el barrio con más tweets dentro de esta categoría. Aunque este barrio del distrito de Chamartín no se caracteriza por tener ubicaciones relacionadas con el arte, sí que tiene una de las atracciones principales de ciudad de Madrid; El estadio Santiago Bernabéu.

El barrio Embajadores se caracteriza por ubicar uno de los museo más importante de España como lo es el Museo Nacional de Arte Reina Sofía y muchos teatros como por ejemplo el famoso teatro Calderón.

El tercer barrio con más tweet con la categoría Arts&Entertainment es el barrio Goya. Al igual que el barrio Hispanoamérica, este barrio del distrito de Salamanca no se caracteriza por tener una alta densidad de ubicaciones relacionadas con el arte, en cambio, en este barrio se ubica el Palacio de los deportes de la Comunidad de Madrid, conocido actualmente como el Wizink Center.

En la cuarta posición con 476 tweets aparece el barrio Imperial, en donde se encuentra el estadio Vicente Calderón.




| Barrio | Categoria | Total Tweets |
|---|---|---|
| HispanoamÃ©rica | Arts&Entertainment | 933 |
| Embajadores | Arts&Entertainment | 641 |
| Goya | Arts&Entertainment | 515 |
| Imperial | Arts&Entertainment | 476 |

*Ilustración 22: Total de tweets en los barrios de Madrid para la categoría Arts&Entertainment.*

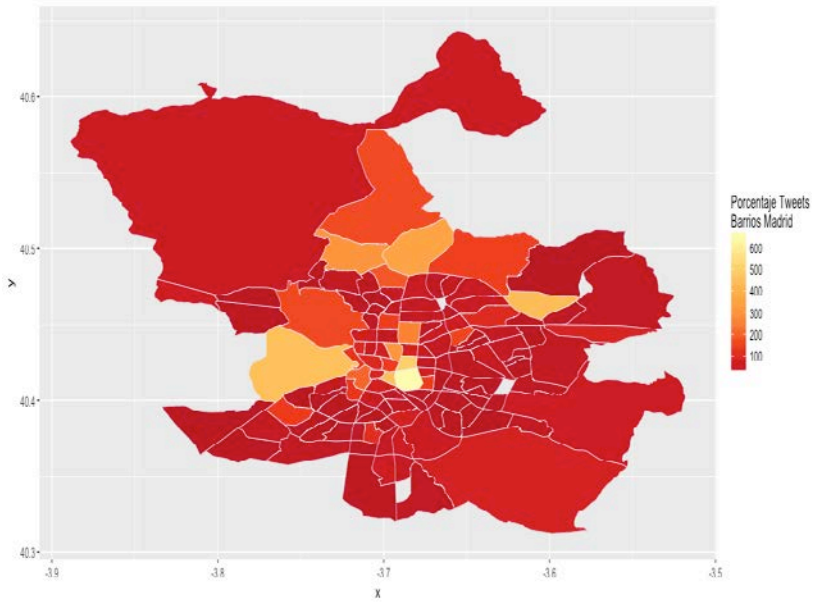
La categoría Professional&OtherPlaces, incluye monumentos, emblemas turísticos, oficinas, edificios, hospitales, etc.

Los Jerónimos aparece como el barrio desde el cual se enviaron más tweet con esta categoría. Este barrio se caracteriza por albergar varios sitios de trabajo tales como el Ayuntamiento (Palacio de Cibeles), Agencia tributaria, Ministerio de Agricultura, además de monumentos como la Puerta de Alcalá y el Palacio de Cristal del Parque del Retiro.

El barrio de Recoletos aparece como el segundo barrio con más tweets dentro de esta categoría. En este barrio se ubica el Colegio de Abogados de Madrid, La biblioteca Nacional de España, Puerta de Alcalá y varias empresas y sitios de trabajo.

Con casi la misma cantidad de tweets generados desde el barrio de Recoletos; el barrio de Casa de Campo aparece como el tercer barrio con más tweets dentro de esta categoría. A diferencia de los Jerónimos y Recoletos, este barrio a pesar de ser uno de los más extensos de Madrid, no se caracteriza por tener una alta densidad de ubicaciones de tipo laborares. En este barrio se encuentra uno de los monumentos o atracción turística más importante de Madrid; Templo de Debod.

En el barrio de Cortes, se ubican varios de los monumentos más importantes de Madrid, como lo son la fuente de Cibeles y la Fuente de Neptuno.



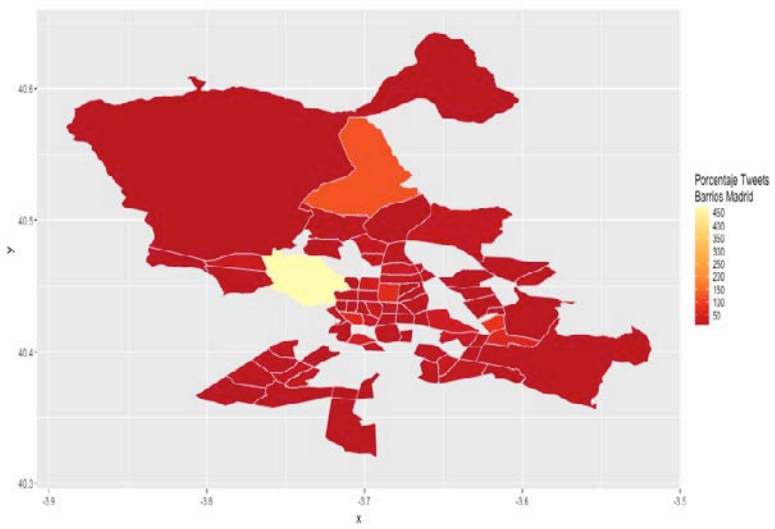

| Barrio | Categoria | Total Tweets |
|---|---|---|
| JerÃ³nimos | Professional&OtherPlaces | 663 |
| Recoletos | Professional&OtherPlaces | 501 |
| Casa de Campo | Professional&OtherPlaces | 456 |
| Cortes | Professional&OtherPlaces | 444 |

*Ilustración 23:Total de tweets en los barrios de Madrid para la categoría Professional&OtherPlaces.*

En cuanto a la categoría College&University, el barrio Ciudad Universitaria ocupa de forma evidente la primera posición. En segunda posición, se ubica el barrio el Goloso en el cual se encuentra la Universidad Autónoma de Madrid.

En tercera lugar se ubica el barrio Arcos. En este barrio se ubica la faculta de óptica y optometría de la Universidad Complutense de Madrid y varios institutos.

El barrio el Viso aunque con poca cantidad de tweets, se posiciona como el cuarto barrio con más tweets dentro de la categoría College&University. En este barrio se ubica el IE Bussiness School y varios institutos, entre ellos el colegio la Salle.

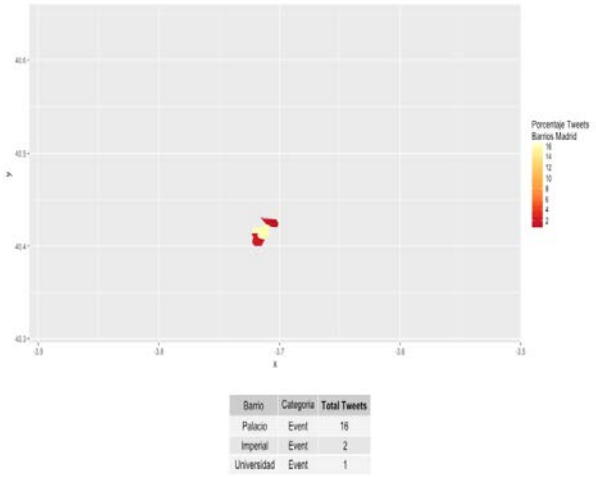



| Barrio | Categoria | Total Tweets |
|---|---|---|
| Ciudad Universitaria | College&University | 474 |
| El Goloso | College&University | 121 |
| Arcos | College&University | 87 |
| El Viso | College&University | 69 |

*Ilustración 24:Total de tweets en los barrios de Madrid para la categoría College&University.*

Como se puede observar, la categoría Event, se origina en únicamente 3 barrios. Esta categoría se asigna a todos aquellas atracciones no permanentes, tales como un mercado de navidad, una atracción itinerante como un circo o un espectáculo.

Los 16 tweets del barrio palacio corresponde con el evento Musical del Rey León, mientras que los 2 tweets originados dentro del barrio Imperial corresponden al evento WordCamp Madrid.




| Barrio | Categoria | Total Tweets |
|---|---|---|
| Palacio | Event | 16 |
| Imperial | Event | 2 |
| Universidad | Event | 1 |

Ilustración 25: Total de tweets en los barrios de Madrid para la categoría Event.

La ilustración 26 presenta una imagen del mapa interactivo en donde se muestran cada uno de los tweets recolectados (link al mapa interactivo). Es posible obtener pulsando sobre cada uno de los tweets sus correspondientes categorías, así como el barrio donde se localizan. Adicionalmente se muestran las divisiones geográficas, según cada uno de los barrios que conforma Madrid Capital.

A continuación se busca determinar los patrones de asociación teniendo en cuenta las categorización de cada uno de los tweets a partir de los barrios de Madrid desde los cuales fueron generados.

Para este estudio se examinan las reglas de asociación entre barrios empleado los siguientes parámetros del algoritmo Apriori; Support igual a **0,003** y un valor de Confidence de **0.7**.Asi, mediante el empleo de estos parámetros se obtiene un total de 117 reglas. Despues de la eliminación de la reglas redundantes se obtiene un total de 88 reglas, que se muestran ordenadas por su valor de Lift de forma decreciente en la tabla 6. El código en R empleado para el análisis de las reglas de asociación para la categorización en función de los barrios se puede consultar en el siguiente link.

En la regla 3, se observa que un individuo que visite el barrio de Cortes (Huertas), Sol y Trafalgar visitará el barrio Universidad. Esta regla esta soportada por un valor de Confidence de un 74,5% y un valor de lift de 4.

Se observa que las 88 reglas obtenidas tienen cuatro RHS que son Palacio, Universidad, Sol y Atocha.

En general todas las reglas cuyo RHS es Sol presentan los valores de Support más elevado en comparación con las reglas cuyo RHS es Palacio, Universidad y Atocha.

| RulesID | rules | support | confidence | lift |
|---|---|---|---|---|
| 1 | {Atocha,Cortes,Jerónimos,Recoletos,Sol} => {Palacio} | 0.00306897108706186 | 0.8444444444444444 | 4.77003244120032 |
| 2 | {Jerónimos,Recoletos,Sol,Universidad} => {Palacio} | 0.00323049588111775 | 0.7407407407407041 | 4.184238983850906 |
| 3 | {Cortes,Sol,Trafalgar} => {Universidad} | 0.00323049588111775 | 0.7547169811320760 | 4.13491400901653 |
| 4 | {Cortes,Embajadores,Justicia,Palacio} => {Universidad} | 0.00355354546922953 | 0.7457627711864407 | 4.085855707214640 |
| 5 | {Aeropuerto,Justicia,Palacio} => {Universidad} | 0.00323049588111775 | 0.7142857142857140 | 3.9134007585335 |
| 6 | {Casa de Campo,Cortes,Jerónimos,Palacio} => {Sol} | 0.00314973348408981 | 0.9285714285714290 | 3.4788415820018590 |
| 7 | {Atocha,Cortes,Embajadores,Jerónimos,Palacio} => {Sol} | 0.00323049588111775 | 0.8695652173913040 | 3.25777807011774 |
| 8 | {Cortes,Embajadores,Jerónimos,Palacio} => {Sol} | 0.00428040704248102 | 0.8548387096774190 | 3.2026060221560700 |
| 9 | {Atocha,Embajadores,Palacio,Recoletos} => {Sol} | 0.0033112582781457 | 0.8367346938775510 | 3.13478032665411 |
| 10 | {Casa de Campo,Cortes,Jerónimos} => {Sol} | 0.00411888224842513 | 0.8360655737704920 | 3.1322735051214 |
| 11 | {Atocha,Casa de Campo,Cortes,Palacio} => {Sol} | 0.00363430786625747 | 0.8333333333333330 | 3.12203731719617 |
| 12 | {Atocha,Casa de Campo,Jerónimos,Palacio} => {Sol} | 0.00347278307220158 | 0.8269230769230770 | 3.0980216455254300 |
| 13 | {Cortes,Jerónimos,Justicia,Palacio} => {Sol} | 0.00363430786625747 | 0.8181818181818180 | 3.06527300233806 |
| 14 | {Cortes,Jerónimos,Palacio,Universidad} => {Sol} | 0.00492650621870457 | 0.8133333333333330 | 3.0471084215834600 |
| 15 | {Casa de Campo,Cortes,Palacio} => {Sol} | 0.0051687934097884 | 0.8101265822784810 | 3.03509450583121 |
| 16 | {Cortes,Justicia,Palacio,Recoletos} => {Sol} | 0.00371507026328541 | 0.8070175438596490 | 3.0234466650741800 |
| 17 | {Atocha,Cortes,Embajadores,Palacio,Universidad} => {Sol} | 0.0033112582781457 | 0.8039215686274510 | 3.01184776482454 |
| 18 | {Atocha,Cortes,Justicia,Palacio,Universidad} => {Sol} | 0.00363430786625747 | 0.8035714285714290 | 3.0105359884439160 |
| 19 | {Cortes,Jerónimos,Palacio,Recoletos} => {Sol} | 0.00419964464545308 | 0.7878787878787880 | 2.95174437262183 |
| 20 | {Atocha,Cortes,Jerónimos,Palacio} => {Sol} | 0.00686480374737522 | 0.7870370370370370 | 2.9485907099574160 |
| 21 | {Aeropuerto,Jerónimos,Universidad} => {Sol} | 0.00323049588111775 | 0.7843137254901960 | 2.93838806324345 |
| 22 | {Atocha,Embajadores,Jerónimos,Palacio} => {Sol} | 0.0051687934097884 | 0.7804878048780490 | 2.9240544629349500 |
| 23 | {Atocha,Embajadores,Justicia,Palacio,Universidad} => {Sol} | 0.00314973348408981 | 0.78 | 2.92222692889561 |
| 24 | {Cortes,Embajadores,Justicia,Palacio} => {Sol} | 0.00371507026328541 | 0.7796610169491530 | 2.9209569476140400 |
| 25 | {Atocha,Cortes,Embajadores,Palacio} => {Sol} | 0.00541108060087223 | 0.7790697674418610 | 2.91874186398339 |
| 26 | {Casa de Campo,Palacio,Recoletos} => {Sol} | 0.00339202067517364 | 0.7777777777777780 | 2.9139014960497600 |
| 27 | {Atocha,Cortes,Justicia,Palacio} => {Sol} | 0.00533031820384429 | 0.7764705882352940 | 2.90900418261102 |
| 28 | {Cortes,Jerónimos,Palacio} => {Sol} | 0.00993377483443709 | 0.7735849056603770 | 2.8981931321896500 |
| 29 | {Embajadores,Jerónimos,Palacio,Universidad} => {Sol} | 0.00371507026328541 | 0.7666666666666670 | 2.87227433182047 |
| 30 | {Casa de Campo,Cortes,Universidad} => {Sol} | 0.00419964464545308 | 0.7647058823529410 | 2.8649283616623700 |
| 31 | {Cortes,Embajadores,Jerónimos,Universidad} => {Sol} | 0.00314973348408981 | 0.7647058823529410 | 2.86492836166237 |
| 32 | {Atocha,Embajadores,Justicia,Palacio} => {Sol} | 0.00492650621870457 | 0.7625 | 2.8566641452344900 |
| 33 | {Jerónimos,Justicia,Palacio,Recoletos} => {Sol} | 0.00306897108706186 | 0.76 | 2.8472980332829 |
| 34 | {Atocha,Cortes,Goya,Palacio} => {Sol} | 0.0033112582781457 | 0.7592592592592590 | 2.8445228890009500 |
| 35 | {Atocha,Cortes,Palacio,Universidad} => {Sol} | 0.00670327895331933 | 0.7545454545454550 | 2.82686287993398 |
| 36 | {Cortes,Palacio,Recoletos,Universidad} => {Sol} | 0.00395735745436925 | 0.7538461538461540 | 2.8242429888847899 |
| 37 | {Cortes,Justicia,Palacio,Universidad} => {Sol} | 0.00508803101276046 | 0.75 | 2.80983358547655 |
| 38 | {Goya,Jerónimos,Justicia} => {Sol} | 0.00347278307220158 | 0.7413793103448280 | 2.7775366477124500 |
| 39 | {Argüelles,Cortes,Palacio} => {Sol} | 0.00363430786625747 | 0.7377049180327870 | 2.763770739813 |
| 40 | {Cortes,Palacio,Recoletos} => {Sol} | 0.00702632854143111 | 0.7372881355932200 | 2.7622092874176300 |
| 41 | {Aeropuerto,Cortes,Jerónimos} => {Sol} | 0.00339202067517364 | 0.7368421052631580 | 2.760538259415560 |
| 42 | {Cortes,Jerónimos,Justicia,Universidad} => {Sol} | 0.00339202067517364 | 0.7368421052631580 | 2.7605382594155600 |
| 43 | {Cortes,Goya,Palacio} => {Sol} | 0.00444193183653691 | 0.7333333333333330 | 2.74739283913263 |
| 44 | {Argüelles,Cortes,Justicia} => {Sol} | 0.00323049588111775 | 0.7272727272727270 | 2.7246871131893800 |
| 45 | {Embajadores,Palacio,Recoletos} => {Sol} | 0.00452269423356485 | 0.7272727272727270 | 2.72468711318938 |
| 46 | {Atocha,Jerónimos,Palacio,Recoletos} => {Sol} | 0.00492650621870457 | 0.7261904761904760 | 2.7206325192709500 |
| 47 | {Atocha,Embajadores,Jerónimos,Universidad} => {Sol} | 0.00363430786625747 | 0.7258064516129030 | 2.71919379239666 |
| 48 | {Cortes,Embajadores,Palacio} => {Sol} | 0.00743014052657083 | 0.7244094488188980 | 2.7138599985705300 |
| 49 | {Cortes,Palacio,Universidad} => {Sol} | 0.0101760620255209 | 0.7241379310344830 | 2.71294277218426 |
| 50 | {Atocha,Cortes,Jerónimos,Justicia} => {Sol} | 0.00379583266031336 | 0.7230769230769230 | 2.7089677644594400 |
| 51 | {Cortes,Justicia,Palacio} => {Sol} | 0.0077531901146826 | 0.7218045112781950 | 2.70420074391728 |
| 52 | {Atocha,Justicia,Palacio,Recoletos} => {Sol} | 0.00395735745436925 | 0.7205882352941180 | 2.6996440331049200 |
| 53 | {Aeropuerto,Cortes,Universidad} => {Sol} | 0.0033112582781457 | 0.7192982456140350 | 2.6948111580009 |
| 54 | {Goya,Jerónimos,Palacio} => {Sol} | 0.00452269423356485 | 0.7179487179487180 | 2.6897552271228500 |
| 55 | {Atocha,Cortes,Jerónimos,Universidad} => {Sol} | 0.00492650621870457 | 0.7176470588235290 | 2.68862507786776 |
| 56 | {Argüelles,Justicia,Palacio} => {Sol} | 0.00347278307220158 | 0.7166666666666670 | 2.6849520927887 |
| 57 | {Embajadores,Jerónimos,Palacio} => {Sol} | 0.00694556614440317 | 0.7166666666666670 | 2.6849520927887 |
| 58 | {Aeropuerto,Jerónimos,Palacio} => {Sol} | 0.00403811985139719 | 0.7142857142857140 | 2.6760319861681400 |
| 59 | {Cortes,Jerónimos,Universidad} => {Sol} | 0.00767242771765466 | 0.7142857142857140 | 2.67603198616814 |
| 60 | {Embajadores,Goya,Palacio} => {Sol} | 0.00371507026328541 | 0.7076923076923080 | 2.6513301524496700 |
| 61 | {Cortes,Jerónimos,Justicia,Recoletos} => {Sol} | 0.00306897108706186 | 0.7037037037037040 | 2.63638706785454 |
| 62 | {Embajadores,Justicia,Recoletos,Sol} => {Atocha} | 0.00306897108706186 | 0.8085106382978720 | 1.9625521904340800 |
| 63 | {Almagro,Justicia,Sol} => {Atocha} | 0.0038765950573413 | 0.7741935483870970 | 1.87925201257186 |
| 64 | {Jerónimos,Palacio,Recoletos,Universidad} => {Atocha} | 0.00355354546922953 | 0.7719298245614030 | 1.8735712364621 |
| 65 | {Jerónimos,Justicia,Palacio,Universidad} => {Atocha} | 0.00371507026328541 | 0.7666666666666670 | 1.86098150689407 |
| 66 | {Embajadores,Jerónimos,Justicia,Sol} => {Atocha} | 0.00314973348408981 | 0.75 | 1.8205253887178900 |
| 67 | {Cortes,Jerónimos,Justicia,Palacio} => {Atocha} | 0.0033112582781457 | 0.7454545454545450 | 1.80949189998396 |
| 68 | {Aeropuerto,Jerónimos,Palacio} => {Atocha} | 0.00419964464545308 | 0.7428571428571430 | 1.8033187051015823 |
| 69 | {Cortes,Jerónimos,Palacio,Recoletos} => {Atocha} | 0.00395735745436925 | 0.7424242424242420 | 1.80213624185394 |
| 70 | {Cortes,Embajadores,Jerónimos,Palacio} => {Atocha} | 0.00371507026328541 | 0.7419354838709680 | 1.8009498453813600 |
| 71 | {Cortes,Sol,Trafalgar} => {Atocha} | 0.00314973348408981 | 0.7358490566037740 | 1.78617585157183 |
| 72 | {Embajadores,Jerónimos,Palacio,Universidad} => {Atocha} | 0.00355354546922953 | 0.7333333333333330 | 1.780069267463900 |
| 73 | {Casa de Campo,Palacio,Sol,Universidad} => {Atocha} | 0.0033112582781457 | 0.7321428571428570 | 1.7771795446271 |
| 74 | {HispanoamÈrica,Jerónimos,Palacio} => {Atocha} | 0.00314973348408981 | 0.7222222222222220 | 1.7530985209871700 |
| 75 | {Embajadores,Justicia,Recoletos} => {Atocha} | 0.0046034566305928 | 0.7215189873417720 | 1.75139151171649 |
| 76 | {Cortes,Goya,Palacio} => {Atocha} | 0.00436116943950896 | 0.72 | 1.747704371691830 |
| 77 | {Cortes,Jerónimos,Palacio,Universidad} => {Atocha} | 0.00436116943950896 | 0.72 | 1.74770437169183 |
| 78 | {Jerónimos,Justicia,Palacio,Sol} => {Atocha} | 0.00452269423356485 | 0.7179487179487180 | 1.7427251569576600 |
| 79 | {Almagro,Palacio,Sol} => {Atocha} | 0.0038765950573413 | 0.7164179104477610 | 1.739009325066490 |
| 80 | {Embajadores,Justicia,Sol,Universidad} => {Atocha} | 0.00541108060087223 | 0.7127659574468080 | 1.7301446941918470 |
| 81 | {Almagro,Cortes,Sol} => {Atocha} | 0.00355354546922953 | 0.7096774193548390 | 1.722647678190870 |
| 82 | {Almagro,Cortes,Palacio} => {Atocha} | 0.00314973348408981 | 0.7090909090909090 | 1.7212240022423770 |
| 83 | {Casa de Campo,Cortes,Jerónimos} => {Atocha} | 0.00347278307220158 | 0.7049180327868850 | 1.7110948994250600 |
| 84 | {Jerónimos,Recoletos,Sol,Universidad} => {Atocha} | 0.00306897108706186 | 0.7037037037037040 | 1.7081472768592900 |
| 85 | {Almagro,Cortes,Justicia} => {Atocha} | 0.00363430786625747 | 0.703125 | 1.706742550480300 |
| 86 | {Palacio,Sol,Trafalgar} => {Atocha} | 0.00379583266031336 | 0.7014925373134330 | 1.7027799641276100 |
| 87 | {Goya,Jerónimos,Universidad} => {Atocha} | 0.00395735745436925 | 0.7 | 1.699157028033720 |
| 88 | {Casa de Campo,Jerónimos,Universidad} => {Atocha} | 0.00395735745436925 | 0.7 | 1.6991570280337200 |

Tabla 6:Reglas espaciales con categorización de ubicaciones con su correspondiente barrio.

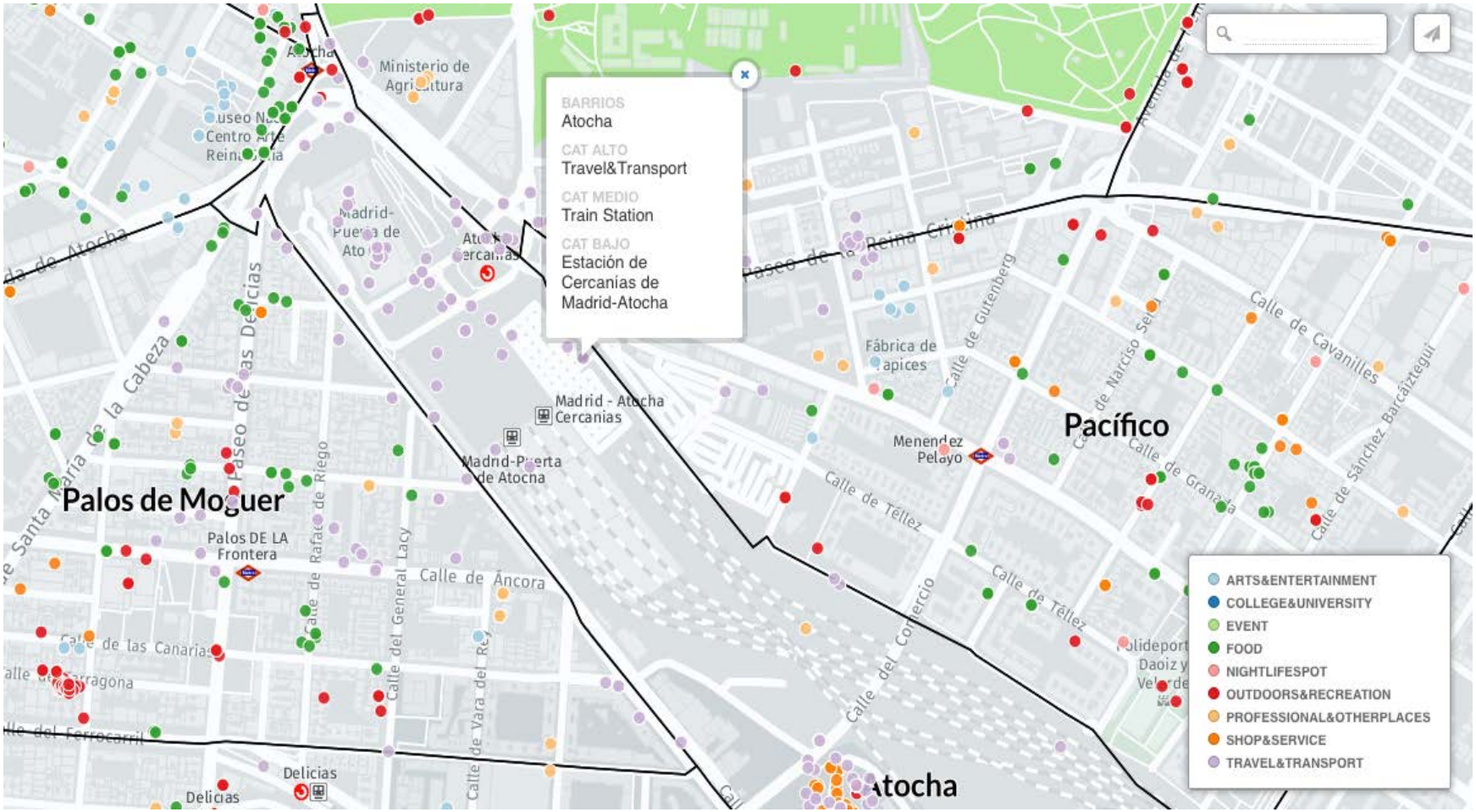



*Ilustración 26: Mapa de barrios de Madrid y categorización de alto nivel de tweets (link al mapa interactivo)*

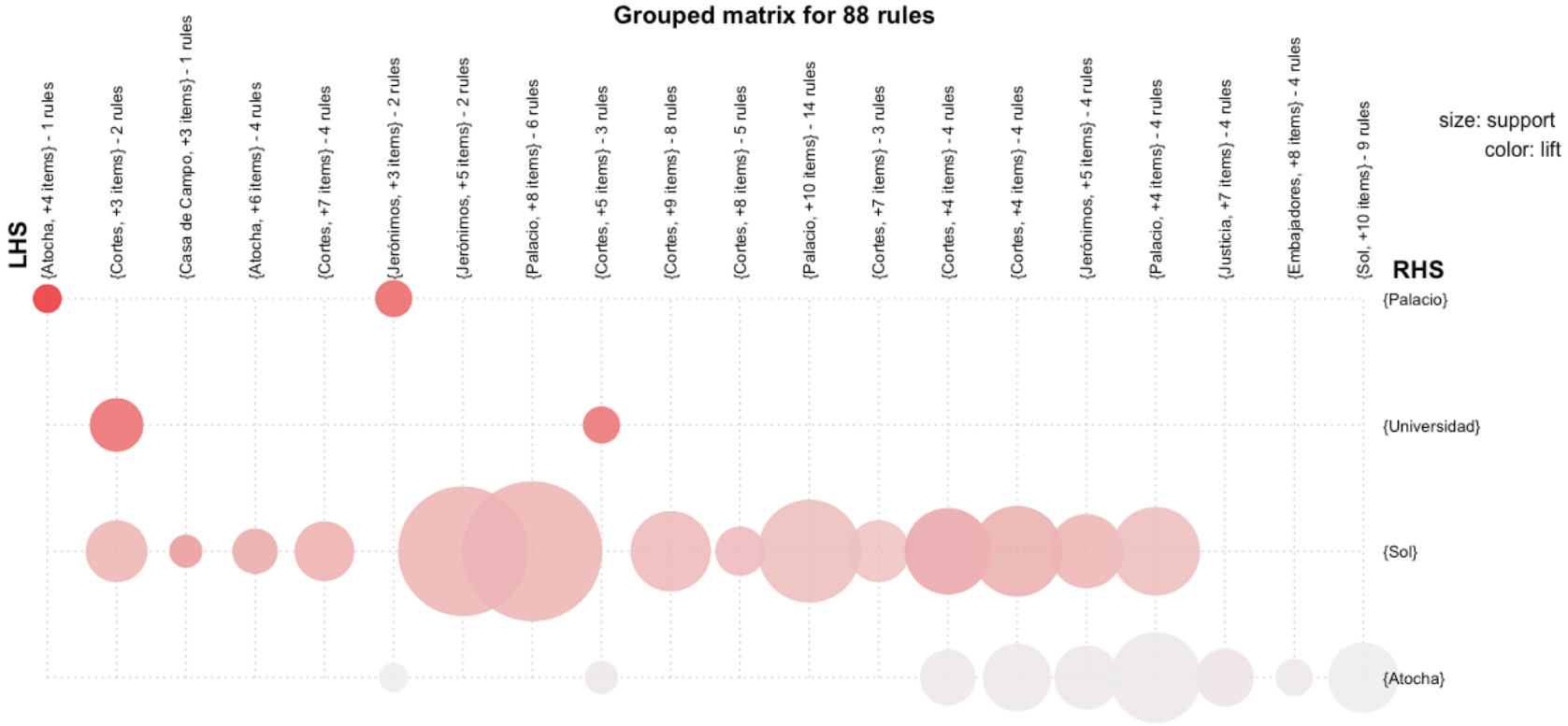



*Ilustración 27:Matriz de las reglas a partir de la categorización de barrios de Madrid.*

# CAPÍTULO V – MINADO DE PATRONES DE TRAYECTORIA CON COMPONENTE TEMPORAL.

## 1. Introducción

En el capítulo anterior se ha estudiado la movilidad de los individuos sin tener en cuenta la componente temporal de sus desplazamientos, ni la secuencia en cómo se desplazan estos individuos entre cada una de las ubicaciones que visita.

Uno de los objetivos planteados en este estudio buscan determinar el comportamiento de los individuos de Madrid capital, teniendo en cuenta la componente temporal en relación a sus desplazamientos.

Tal como se indicó en el capítulo 3, el conjunto de datos $(TW)$ se compone de las siguientes variables: latitud y longitud en la ubicación $(l)$ desde donde se generó el tweet, la hora $(t)$ a la cual se generó y el usuario $(u)$ que generó dicho tweet, así como las categorizaciones de las ubicaciones a bajo, medio y alto nivel.

Dado que cada tweet tiene asociado la componente temporal dentro de la variable **created_at** bajo el siguiente formato: año-mes-día hora:minutos:segundos (ejemplo: 2017-03-21 09:52:25), es posible generar otras variables referidas a la componente temporal de forma que nos facilite el análisis inicial del conjunto de datos. Para ello, se han creado dos nuevas variables a partir de **created_at.** La nueva variable **día**, se corresponde con cada uno de los 7 días de la semana. Con esta variable es posible asociar la recurrencia de visitas a diversas ubicaciones a la variable **día.**

La segunda variable creada a partir de **created_at** es **timeslot.** Esta variable se corresponde con la discretización de la hora a la cual fue generado el tweet. Esta variable está compuesta por 6 categorías, en las cuales se agrupan cada una de las principales franjas horarias de un día.

| timeslot | Nombre Franja horaria | Franja Horaria |
|---|---|---|
| N | Nigth | [00:00:00-05:59:59] |
| EM | EarlyMorning | [06:00:00-09:59:59] |
| M | Morning | [10:00:00-13:59:59] |
| A | Afternoon | [14:00:00-17:59:59] |
| EE | EarlyEvening | [18:00:00-20:59:59] |
| E | Evening | [21:00:00-23:59:59] |

*Tabla 7: Categorías de la variable timeslot.*

La variable **timeslot** nos permitirá identificar cuáles son las franjas horarias en las que un individuo visita de forma recurrente una determinada ubicación de la ciudad.

A continuación se muestra un extracto del conjunto de datos con las dos nuevas variables: **dia** y **timeslot.**

| lat | lon | cat_alto | cat_medio | cat_bajo | user_id_str | created_at | dia | timeslot |
|---|---|---|---|---|---|---|---|---|
| 40.42232714 | -3.70093022 | Shop&Service | Accessories Store | Misako | 625532598 | 21/03/17 09:52 | Tuesday | EM |
| 40.42232714 | -3.70093022 | Shop&Service | Accessories Store | Misako | 85534222 | 30/03/17 16:26 | Thursday | A |
| 40.42232714 | -3.70093022 | Shop&Service | Accessories Store | Misako | 321940034 | 23/03/17 07:18 | Thursday | EM |
| 40.4265289 | -3.6867599 | Shop&Service | Accessories Store | & Other Stories | 3015652713 | 03/03/17 08:07 | Friday | EM |
| 40.42232714 | -3.70093022 | Shop&Service | Accessories Store | Misako | 321940034 | 17/03/17 07:47 | Friday | EM |
| 40.4265289 | -3.6867599 | Shop&Service | Accessories Store | & Other Stories | 224671463 | 19/02/17 05:42 | Sunday | N |
| 40.4265289 | -3.6867599 | Shop&Service | Accessories Store | & Other Stories | 224671463 | 13/03/17 12:46 | Monday | M |
| 40.42232714 | -3.70093022 | Shop&Service | Accessories Store | Misako | 4765116714 | 15/03/17 13:04 | Wednesday | M |
| 40.42653 | -3.68676 | Shop&Service | Accessories Store | & Other Stories | 332501465 | 03/04/17 10:52 | Monday | M |
| 40.42140884 | -3.70357793 | Professional&OtherPlaces | Adult Education Center | Kitchen Club | 135485144 | 24/02/17 16:06 | Friday | A |
| 40.42140884 | -3.70357793 | Professional&OtherPlaces | Adult Education Center | Kitchen Club | 135485144 | 21/03/17 11:54 | Tuesday | M |
| 40.4471256 | -3.6952054 | Professional&OtherPlaces | Advertising Agency | DDB | 218818891 | 17/02/17 14:17 | Friday | A |
| 40.43399923 | -3.63731238 | Professional&OtherPlaces | Advertising Agency | FourCats Media | 262150489 | 12/02/17 07:29 | Sunday | EM |
| 40.4373283 | -3.6798999 | Professional&OtherPlaces | Advertising Agency | Forward Media | 50983030 | 25/03/17 11:31 | Saturday | M |
| 40.43464633 | -3.63773253 | Professional&OtherPlaces | Advertising Agency | FourCats Media | 262150489 | 26/03/17 07:47 | Sunday | EM |
| 40.4373283 | -3.6798999 | Professional&OtherPlaces | Advertising Agency | Forward Media | 50983030 | 24/02/17 03:44 | Friday | N |
| 40.43448856 | -3.63762562 | Professional&OtherPlaces | Advertising Agency | FourCats Media | 262150489 | 14/02/17 09:38 | Tuesday | EM |
| 40.4373283 | -3.6798999 | Professional&OtherPlaces | Advertising Agency | Forward Media | 50983030 | 08/02/17 04:22 | Wednesday | N |
| 40.45462454 | -3.62224776 | Professional&OtherPlaces | Advertising Agency | Carat | 517458149 | 13/02/17 07:38 | Monday | EM |
| 40.4373283 | -3.6798999 | Professional&OtherPlaces | Advertising Agency | Forward Media | 50983030 | 21/02/17 06:09 | Tuesday | EM |
| 40.4288448 | -3.7160627 | Professional&OtherPlaces | Advertising Agency | Mobusi Mobile Advertising | 251795911 | 09/03/17 12:42 | Thursday | M |
| 40.43470811 | -3.63769744 | Professional&OtherPlaces | Advertising Agency | FourCats Media | 262150489 | 10/03/17 23:56 | Friday | E |
| 40.43471531 | -3.63772143 | Professional&OtherPlaces | Advertising Agency | FourCats Media | 262150489 | 26/03/17 07:41 | Sunday | EM |
| 40.4373283 | -3.6798999 | Professional&OtherPlaces | Advertising Agency | Forward Media | 50983030 | 20/02/17 09:01 | Monday | EM |

*Tabla 8:Muestra del conjunto de datos principal con las variables dia y timelot.*

En la ilustración 28 se representa la distribución de cada una de las 9 categorías de alto nivel, sobre cada una de las 6 franjas horarias. Se observa que en los intervalos horarios **EE** (18:00:00-20:59:59) y **E** (21:00:00-23:59:59) la cantidad de tweets generados es muy inferior al resto de categorías (**EE**:1469 tweets y **E**: 2160 tweets). De esta forma, de los 70.684 tweets que conforman el conjunto de datos, los tweets generados en intervalo **E** se corresponde con un 3,05% del total de tweets, mientras que los tweets generados dentro del intervalo **EE** suponen un 2,07% del total de tweets.

La franja horaria con el mayor número de tweets es la **M** (10:00:00-13:59:59) con un total de 25,87%, seguido por la franja horaria **N** (00:00:00-05:59:59) con un 25,27%. Las franjas horarias **EM** (06:00:00-09:59:59) y **A**(14:00:00-17:59:59) contiene respectivamente un 24,52% y 19,21 del total.

Al analizar cada uno de los intervalos temporales, se puede observar que la categoría **Outdoors&Recreation** es la que presenta mayor número de tweets para todos los intervalos temporales excepto para el intervalo **EE** (18:00:00-20:59:59) en donde la categoría **Food** es la dominante.

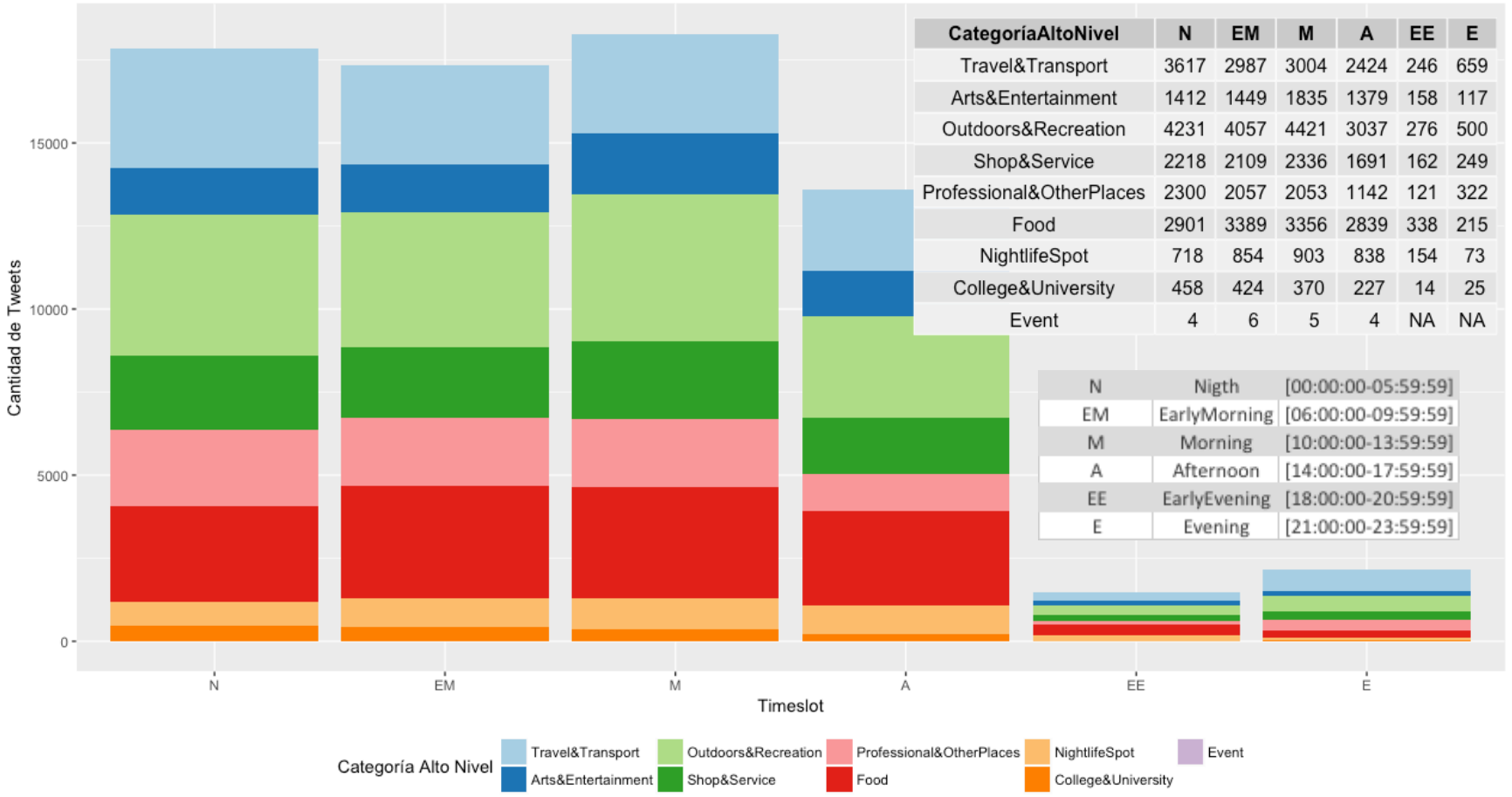
Evaluando individualmente cada una de las categorías se observa las categorías **Professional&OtherPlaces, College&University y Travel&Transport** alcanza su máximo en el intervalo **N** (00:00:00-05:59:59).

La categoría **Arts&Entertainment, Shop&Service, NightlifeSpot y Outdoors&Recreation** alcanzan el mayor número de tweets en el intervalo **M** (10:00:00-13:59:59).

La categoría **Food** y **Event** alcanza su máximo en el intervalo **EM** (06:00:00-09:59:59).

En la ilustración 29 se muestra el agrupamiento del total de tweets en función de las categoría de alto nivel y del día de la semana en el cual se generó dicho tweet.

Se aprecia como la categoría dominante para los viernes, en cualquier intervalo horario, se corresponde con **Outdoors&Recreation.** Los sábados se aprecia que la categoría dominante, para cada una de las franjas horarias, se corresponde con la categoría **Food.** El mismo comportamiento se observa los domingos, en donde la categoría dominante se corresponde con **Travel&Transport** seguida por **Food**.




| CategoríaAltoNivel | N | EM | M | A | EE | E |
|---|---|---|---|---|---|---|
| Travel&Transport | 3617 | 2987 | 3004 | 2424 | 246 | 659 |
| Arts&Entertainment | 1412 | 1449 | 1835 | 1379 | 158 | 117 |
| Outdoors&Recreation | 4231 | 4057 | 4421 | 3037 | 276 | 500 |
| Shop&Service | 2218 | 2109 | 2336 | 1691 | 162 | 249 |
| Professional&OtherPlaces | 2300 | 2057 | 2053 | 1142 | 121 | 322 |
| Food | 2901 | 3389 | 3356 | 2839 | 338 | 215 |
| NightlifeSpot | 718 | 854 | 903 | 838 | 154 | 73 |
| College&University | 458 | 424 | 370 | 227 | 14 | 25 |
| Event | 4 | 6 | 5 | 4 | NA | NA |

| | | |
|---|---|---|
| N | Nigth | [00:00:00-05:59:59] |
| EM | EarlyMorning | [06:00:00-09:59:59] |
| M | Morning | [10:00:00-13:59:59] |
| A | Afternoon | [14:00:00-17:59:59] |
| EE | EarlyEvening | [18:00:00-20:59:59] |
| E | Evening | [21:00:00-23:59:59] |

*Ilustración 28:Distribución número de tweets agrupados por categoría de alto nivel.*

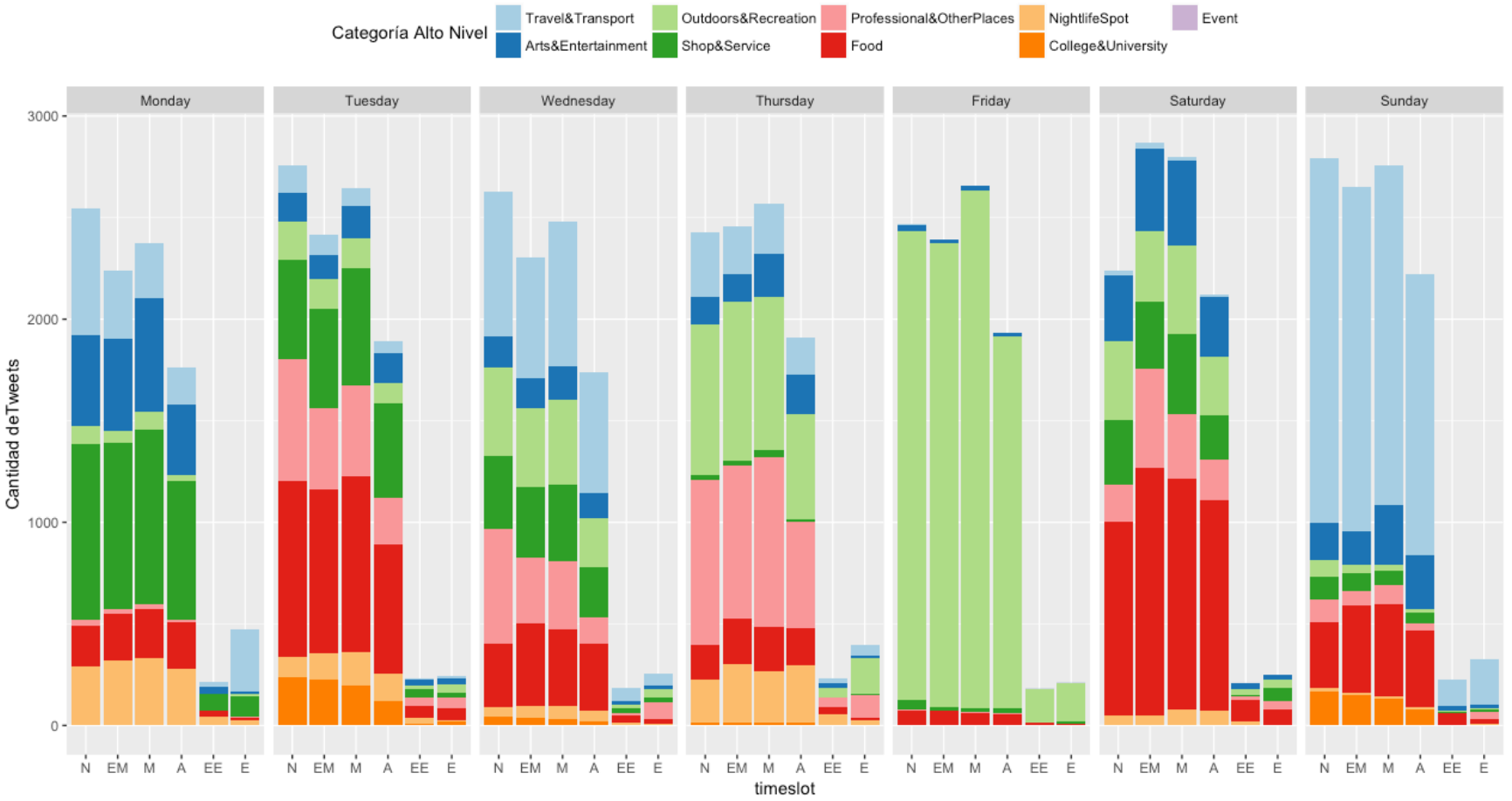



*Ilustración 29:Distribución de la cantidad de tweets agrupados por la categoría de alto nivel y el día de la semana.*

En este capítulo se busca extraer los patrones de trayectorias espacio-temporales para la categorías de bajo nivel asociados a los tweets que conforman el conjunto de datos.

Los patrones de trayectoria espacio-temporales pueden ser usados para obtener información de comportamientos frecuentes de los individuos en términos de espacio y tiempo. De esta forma, el problema del minado de trayectoria teniendo en cuenta las componentes de espacio y tiempo se puede modelar como una extensión de los algoritmos del minado de secuencias (Sequence Mining).

Dada una ruta de viaje generada por los individuos a lo largo del periodo de recolección de los datos, el algoritmo de minado de patrones secuenciales encontrará todos aquellos patrones cuya frecuencias no sea menor que el valor de **Support mínimo** definido en el parámetro de entrada del algoritmo tal y como se estableció en el algoritmo Apriori.

Se puede definir un patrón de trayectoria o ruta frecuente de viaje ($FTR$) como una secuencia de ubicaciones visitadas juntas con una frecuencia no menor al valor del **Support** y donde cada una de las rutas de viajes frecuentes, se presentan en una formato secuencial tal y como se presenta a continuación.

$$FTR_{u,} = c_{t_0} \rightarrow c_{t_1} \rightarrow \cdots \rightarrow c_{t_n}, donde\ t_n < t_{n+1}$$

En donde $c$ corresponde a la categoría de bajo nivel del tweet visitada por el individuo $u$ en el instante $t$.

Existen varios algoritmos que permiten el minado de patrones secuenciales, tales como cSPADE, PrefixSpan y adaptaciones del algoritmo Apriori. Para este estudio se usará el paquete de R TraMineR [11] el cual permite el minado de secuencias a través de la implementación de una adaptación del algoritmo prefix-tree-based search [7] [8].

A modo de ejemplo, se muestran las siguientes rutas de viajes. En cada ruta se muestran las distintas ubicaciones visitadas en forma secuencial por cada uno de los 8 individuos y la duración en minutos entre cada una de las ubicaciones.

$TR_1 =$ (Santa Iglesia Catedral de Santa María la Real de la Almudena) $-$ 632.6833
$-$ (Puerta de Alcalá) $-$ 1456.533
$-$ ( Estación de Madrid Puerta de Atocha) $-$ 1424.017
$-$ ( Templo de Debod)

$TR_2 =$ (Royal Thai) $-$ 24522.68 $-$ (Arepa Olé "Las Tablas") $-$ 32705.48
$-$ (Estación de Madrid Puerta de Atocha) $-$ 25899.37
$-$ (Estación de Madrid Puerta de Atocha)

[11] http://traminer.unige.ch/index.shtml

$TR_3 =$ (Polideportivo La Masó) − 29534.38 − (Cinesa Méndez Álvaro)

$TR_4 =$ (Starbucks) − 555.8333 − (Plaza de Manuel Becerra)

$TR_5 =$ (Palacio de Cristal del Retiro) − 3217.2 − (Primark)

$TR_6 =$ (Puerta del Sol) − 89.05 − (Puerta de Alcalá) − 14.2
− (Palacio Real del Pardo) − 984.0833 − (Casino de Madrid)

$TR_7 =$ (Puerta de Alcalá) − 1424.2 − (Estación de Madrid Puerta de Atocha)

$TR_8 =$ (Puerta del Sol) − 24539.05 − (Puerta de Alcalá) − 1424.2
− (Estación de Madrid Puerta de Atocha)

Dada estas 8 rutas, el algoritmo de minado determina las rutas de viajes frecuentes ($FTR$) a partir de un valor de mínimo de Support, que se pasa como parámetro dentro del algoritmo. Así, si el valor de frecuencia indicado es igual o mayor a 2, el algoritmo extrae los siguientes subconjuntos de rutas $FTR$.

$$FTR_1 = \text{(Puerta de Alcalá)} \longrightarrow \text{(Estación de Madrid Puerta de Atocha)}$$

La $FTR_1$ surge ya que este esté subconjunto cumple para el valor de frecuencia indicado (frecuencia =3) dado que aparece tanto en el $TR_1$, $TR_7$ y $TR_8$.

Otra $FTR$ que cumple con el parámetro de búsqueda es la siguiente:

$$FTR_2 = \text{(Puerta del Sol)} \longrightarrow \text{(Puerta de Alcalá)}$$

Este subconjunto aparece dos veces en el listado de las 8 rutas de viajes, tal como se puede observar a partir de la $TR_6$ y $TR_8$.

El valor de Support para una ruta de viaje frecuente esta dato por $S(FTR_n) = \frac{f(FTR_n)}{count(\sum_{i=1}^{n} TR_i)}$,

De esta forma para el $FTR_1$ y $FTR_2$ el valor del Support está dado por $S(FTR_1) = \frac{3}{8} = 0.375$ y $(FTR_2) = \frac{2}{8} = 0.25$.

En valor del **Confidence** de una regla $X \rightarrow Y$ indica la probabilidad de que un individuo que visite la ubicación $X$ visite seguidamente la ubicación $Y$.El valor de Confidence puede ser entendido como la probabilidad condicional $P(Y|X)$.

## 2. Minado de trayectorias secuenciales en función de días de la semana para categorizaciones de bajo nivel.

Para analizar esta categoría teniendo en cuenta la componente temporal, se presenta a continuación el gráfico de proporción del total de tweets categorizados a bajo nivel, en donde se muestran las 5 categorías más populares (de las 5.758 diferentes que conforman el conjunto de datos) agrupadas por día de la semana y los intervalos horarios definidos.

Se aprecia como la Estación de Madrid-Puerta de Atocha aparece dentro de las 5 categorías más populares para todos los intervalos temporales y días de la semana, mostrando así la alta frecuencia de visita durante cualquier día e intervalo horario.

Otras ubicaciones aparecen dentro de las 5 categorías de bajo nivel más importantes en días e intervalo horarios específicos. Por ejemplo, el Aeropuerto de Madrid, aparece siempre dentro de las 5 categorías más frecuentes durante todos los días día de las semana en los intervalos **N** (00:00:00-05:59:59). y **E** (21:00:00-23:59:59), observando una mayor proporción de tweets en este último intervalo horario. Ambos intervalos horarios pueden coincidir con los intervalo de mayor afluencia de viajeros en el aeropuerto.

Así mismo, la ubicación Palacio Real de Madrid aparece únicamente el domingo en el intervalo horario **EE** (18:00:00-20:59:59). El Estadio Vicente Calderón aparece en las 5 ubicaciones más frecuentadas el lunes en el intervalo **EM**. El estadio Santiago Bernabéu aparece de forma frecuente dentro de algunos intervalos horarios, destacando especialmente el domingo en el intervalo horario **A**(14:00:00-17:59:59) y jueves en el intervalo **EE** (18:00:00-20:59:59), coincidiendo tanto el día como los intervalos horarios con partidos en este estadio. Las otras apariciones de esta ubicación dentro de los intervalos temporales definidos pueden surgir a partir de visitas turísticas al estadio.

De igual forma, la ubicación Puerta del Sol, se encuentra presente dentro de las 5 ubicaciones más visitadas en la mayoría de los días de la semana para cada uno de los intervalos horarios, destacando una proporción de tweets muy similar en cada intervalo temporal – día.

Tal como se observó en la ilustración 10, la categoría Bankia está presente dentro de las 5 categorías más visitadas para todos los días de la semana e intervalo horario.

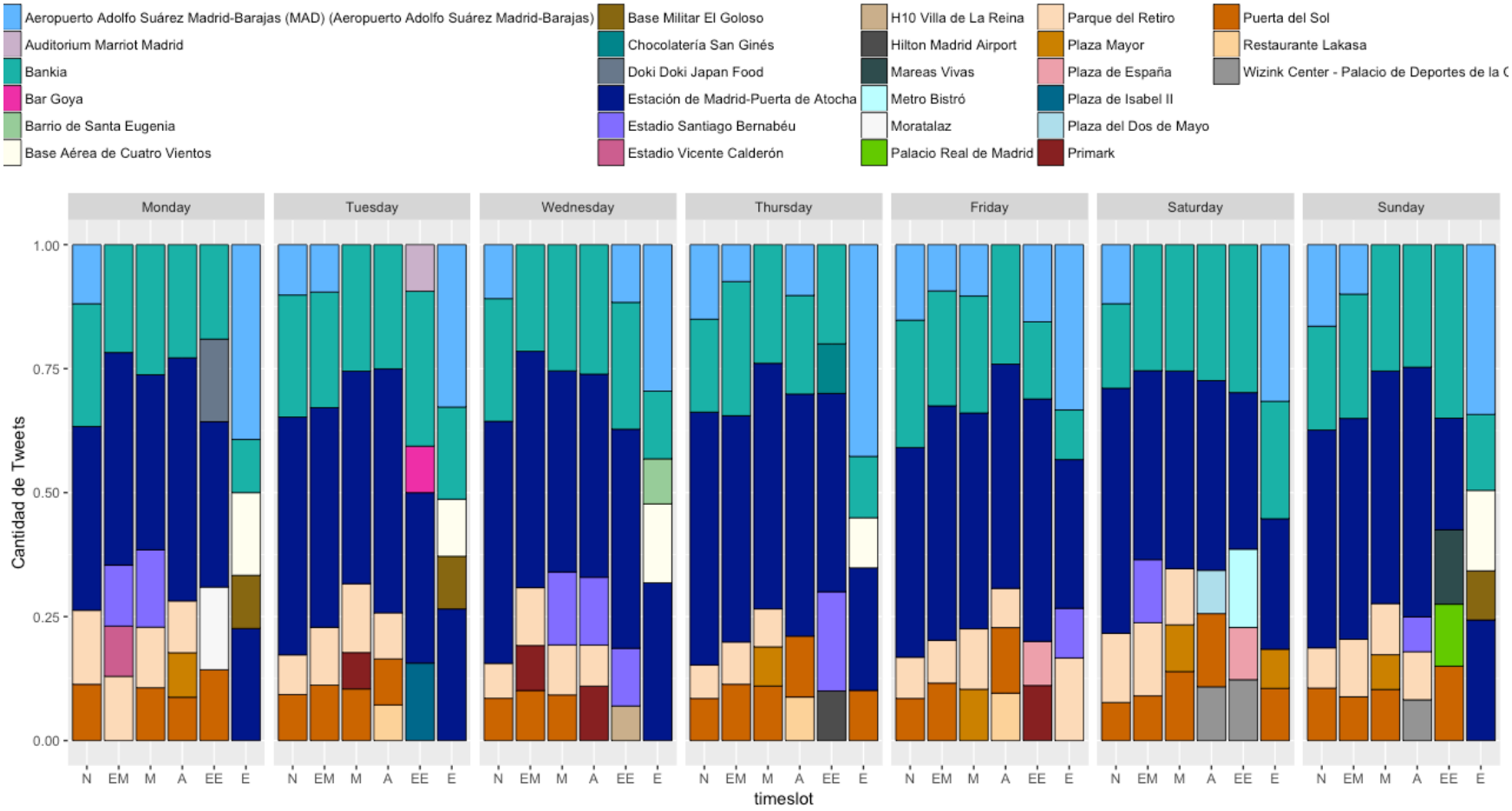



*Ilustración 30: Proporción de Tweets para las 5 categorías de bajo nivel más frecuentes agrupados por días y franja horaria.*

Los patrones secuenciales de trayectoria teniendo en cuenta la categorización de bajo nivel para de cada uno de los días de la semana se obtienen a través del código R que se puede consultar en el siguiente link.

Para obtener cada una de las trayectorias, se usó como parámetro de entrada del algoritmo para la obtención de patrones secuenciales, un valor de Support mínimo de **0.001**. En las reglas obtenida se han eliminado aquellos patrones de rutas que no muestran patrones claros de desplazamientos, es decir, la visita a la ubicación *X* implica visitar seguidamente la misma ubicación *X* ($X \rightarrow X$).

En las tablas de que se muestran a continuación, se listan trayectorias (Reglas) para las categorías de bajo nivel que cumplen con el valor del Support mínimo establecido para cada uno de los 7 días de la semana.

Para cada una de las rutas de viajes frecuentes, se indica la frecuencia en que dicha trayectoria aparece como una subtrayectoria en el listado de ubicaciones visitadas de forma secuencial, el valor de Support asociado a la misma y el valor de Confidence y Lift.

| Día | Reglas | Frecuencia | Confidence | Lift | Support |
|---|---|---|---|---|---|
| Monday | (Real Madrid Official Store) => (Estación de Madrid Puerta de Atocha) | 7 | 0,14 | 0,95420561 | 0,0013712 |
| **Monday** | **(Palacio Real de Madrid) => (Estación de Madrid Puerta de Atocha)** | **10** | **0,12345679** | **0,84145115** | **0,00195886** |
| Monday | (Plaza de España) => (Estación de Madrid Puerta de Atocha) | 7 | 0,109375 | 0,74547313 | 0,0013712 |
| **Monday** | **(Parque Juan Carlos I) => (Estación de Madrid Puerta de Atocha)** | **7** | **0,101449275** | **0,69145334** | **0,0013712** |
| Monday | (Palacio Real de Madrid) => (Puerta del Sol) | 7 | 0,086419753 | 2,77467195 | 0,0013712 |
| **Monday** | **(Plaza Mayor) => (Puerta del Sol)** | **9** | **0,081081081** | **2,60326364** | **0,00176298** |
| Monday | (Starbucks) => (Estación de Madrid Puerta de Atocha) | 6 | 0,078947368 | 0,53808587 | 0,00117532 |
| **Monday** | **(Primark) => (Estación de Madrid Puerta de Atocha)** | **7** | **0,075268817** | **0,51301377** | **0,0013712** |
| Monday | (Palacio Real de Madrid) => (Plaza Mayor) | 6 | 0,074074074 | 3,40674007 | 0,00117532 |
| **Monday** | **(Palacio Real de Madrid) => (Santa Iglesia Catedral de Santa María la Real de la Almudena)** | **6** | **0,074074074** | **18,0070547** | **0,00117532** |
| Monday | (Puerta del Sol) => (Estación de Madrid Puerta de Atocha) | 11 | 0,06918239 | 0,47153017 | 0,00215475 |
| **Monday** | **(Plaza Mayor) => (Estación de Madrid Puerta de Atocha)** | **7** | **0,063063063** | **0,42982235** | **0,0013712** |
| Monday | (Aeropuerto Adolfo Suárez Madrid Barajas (MAD) (Aeropuerto Adolfo Suárez Madrid Barajas)) => (Estación de Madrid Puerta de Atocha) | 8 | 0,046511628 | 0,31701183 | 0,00156709 |
| **Monday** | **(Puerta del Sol) => (Aeropuerto Adolfo Suárez Madrid Barajas (MAD) (Aeropuerto Adolfo Suárez Madrid Barajas))** | **7** | **0,044025157** | **1,30667691** | **0,0013712** |
| Monday | (Puerta del Sol) => (Bankia) | 7 | 0,044025157 | 0,55220744 | 0,0013712 |
| **Monday** | **(Puerta del Sol) => (Plaza Mayor)** | **7** | **0,044025157** | **2,02476061** | **0,0013712** |
| Monday | (Parque del Retiro) => (Palacio de Cristal del Retiro) | 6 | 0,035294118 | 2,72994652 | 0,00117532 |
| **Monday** | **(Estación de Madrid Puerta de Atocha) => (Parque del Retiro)** | **12** | **0,016021362** | **0,48111207** | **0,00235064** |
| Monday | (Estación de Madrid Puerta de Atocha) => (Plaza Mayor) | 12 | 0,016021362 | 0,73683831 | 0,00235064 |
| **Monday** | **(Bankia) => (Puerta del Sol)** | **6** | **0,014742015** | **0,47332066** | **0,00117532** |
| Monday | (Estación de Madrid Puerta de Atocha) => (Puerta del Sol) | 9 | 0,012016021 | 0,38579742 | 0,00176298 |
| **Monday** | **(Estación de Madrid Puerta de Atocha) => (Plaza del Dos de Mayo)** | **7** | **0,009345794** | **0,62776685** | **0,0013712** |
| Monday | (Estación de Madrid Puerta de Atocha) => (Real Madrid Official Store) | 7 | 0,009345794 | 0,95420561 | 0,0013712 |

*Tabla 9: Patrones secuenciales de bajo nivel para los lunes.*

| Día | Reglas | Frecuencia | Confidence | Lift | Support |
|---|---|---|---|---|---|
| Tuesday | (Plaza Mayor) => (Bankia) | 10 | 0,114942529 | 1,51170488 | 0,00199045 |
| **Tuesday** | **(Plaza de España) => (Bankia)** | **7** | **0,101449275** | **1,33424387** | **0,00139331** |
| Tuesday | (Templo de Debod) => (Estación de Madrid Puerta de Atocha) | 7 | 0,1 | 0,68353741 | 0,00139331 |
| **Tuesday** | **(Parque Juan Carlos I) => (Estación de Madrid Puerta de Atocha)** | **6** | **0,090909091** | **0,62139765** | **0,00119427** |
| Tuesday | (Puerta del Sol) => (Estación de Madrid Puerta de Atocha) | 14 | 0,083333333 | 0,56961451 | 0,00278662 |
| **Tuesday** | **(Primark) => (Bankia)** | **9** | **0,081818182** | **1,07605902** | **0,0017914** |
| Tuesday | (Parque del Retiro) => (Estación de Madrid Puerta de Atocha) | 14 | 0,075268817 | 0,51449053 | 0,00278662 |
| **Tuesday** | **(Puerta del Sol) => (Bankia)** | **12** | **0,071428571** | **0,9394166** | **0,00238854** |
| Tuesday | (Palacio Real de Madrid) => (Templo de Debod) | 6 | 0,069767442 | 5,00730897 | 0,00119427 |
| **Tuesday** | **(Primark) => (Estación de Madrid Puerta de Atocha)** | **7** | **0,063636364** | **0,43497835** | **0,00139331** |
| Tuesday | (Parque del Retiro) => (Bankia) | 7 | 0,037634409 | 0,49496144 | 0,00139331 |
| **Tuesday** | **(Bankia) => (Puerta del Sol)** | **9** | **0,023560209** | **0,70456245** | **0,0017914** |
| Tuesday | (Estación de Madrid Puerta de Atocha) => (Puerta del Sol) | 15 | 0,020408163 | 0,61030126 | 0,00298567 |
| **Tuesday** | **(Estación de Madrid Puerta de Atocha) => (Parque del Retiro)** | **12** | **0,016326531** | **0,44099188** | **0,00238854** |
| Tuesday | (Bankia) => (Plaza Mayor) | 6 | 0,015706806 | 0,90702293 | 0,00119427 |
| **Tuesday** | **(Bankia) => (Plaza de España)** | **6** | **0,015706806** | **1,14363761** | **0,00119427** |
| Tuesday | (Estación de Madrid Puerta de Atocha) => (Aeropuerto Adolfo Suárez Madrid Barajas (MAD) (Aeropuerto Adolfo Suárez Madrid Barajas)) | 8 | 0,010884354 | 0,40208083 | 0,00159236 |
| **Tuesday** | **(Estación de Madrid Puerta de Atocha) => (Estación de Madrid Puerta de Atocha) (Estación de Madrid Puerta de Atocha) (Estación de Madrid Puerta de Atocha)** | **7** | **0,00952381** | **2,17489177** | **0,00139331** |
| Tuesday | (Estación de Madrid Puerta de Atocha) => (Starbucks) | 7 | 0,00952381 | 0,67391013 | 0,00139331 |
| **Tuesday** | **(Estación de Madrid Puerta de Atocha) => (Parque Juan Carlos I)** | **6** | **0,008163265** | **0,62139765** | **0,00119427** |
| Tuesday | (Estación de Madrid Puerta de Atocha) => (Plaza Mayor) | 6 | 0,008163265 | 0,47140511 | 0,00119427 |
| **Tuesday** | **(Estación de Madrid Puerta de Atocha) => (Plaza de España)** | **6** | **0,008163265** | **0,59438036** | **0,00119427** |

*Tabla 10: Patrones secuenciales de bajo nivel para los martes.*

| Dia | Reglas | Frecuencia | Confidence | Lift | Support |
|---|---|---|---|---|---|
| Wednesday | (Estación de Madrid Puerta de Atocha)-(Real Madrid Official Store) => (Estación de Madrid Puerta de Atocha) | 5 | 0,5 | 3,51526163 | 0,0010337 |
| **Wednesday** | **(Círculo de Bellas Artes) => (Bankia)** | **7** | **0,162790698** | **2,08864351** | **0,00144718** |
| Wednesday | (Real Madrid Official Store) => (Estación de Madrid Puerta de Atocha) | 11 | 0,152777778 | 1,07410772 | 0,00227414 |
| **Wednesday** | **(Templo de Debod) => (Bankia)** | **5** | **0,089285714** | **1,14555703** | **0,0010337** |
| Wednesday | (Plaza Mayor) => (Bankia) | 7 | 0,079545455 | 1,02058717 | 0,00144718 |
| **Wednesday** | **(Palacio Real de Madrid) => (Estadio Santiago Bernabéu)** | **5** | **0,069444444** | **2,13950814** | **0,0010337** |
| Wednesday | (Aeropuerto Adolfo Suárez Madrid Barajas (MAD) (Aeropuerto Adolfo Suárez Madrid Barajas)) => (Estación de Madrid Puerta de Atocha) | 8 | 0,066115702 | 0,46482798 | 0,00165392 |
| **Wednesday** | **(Puerta del Sol) => (Estación de Madrid Puerta de Atocha)** | **9** | **0,063829787** | **0,4487568** | **0,00186066** |
| Wednesday | (Plaza Mayor) => (Estación de Madrid Puerta de Atocha) | 5 | 0,056818182 | 0,39946155 | 0,0010337 |
| **Wednesday** | **(Parque del Retiro) => (Bankia)** | **8** | **0,054054054** | **0,69352642** | **0,00165392** |
| Wednesday | (Primark) => (Bankia) | 7 | 0,048275862 | 0,61939084 | 0,00144718 |
| **Wednesday** | **(Parque del Retiro) => (Estación de Madrid Puerta de Atocha)** | **6** | **0,040540541** | **0,28502121** | **0,00124044** |
| Wednesday | (Estadio Santiago Bernabéu) => (Bankia) | 6 | 0,038216561 | 0,49032759 | 0,00124044 |
| **Wednesday** | **(Puerta del Sol) => (Plaza de Cibeles)** | **5** | **0,035460993** | **2,7665294** | **0,0010337** |
| Wednesday | (Primark) => (Estación de Madrid Puerta de Atocha) | 5 | 0,034482759 | 0,24243184 | 0,0010337 |
| **Wednesday** | **(Primark) => (Plaza Mayor)** | **5** | **0,034482759** | **1,89537618** | **0,0010337** |
| Wednesday | (Estadio Santiago Bernabéu) => (Estación de Madrid Puerta de Atocha) | 5 | 0,031847134 | 0,22390201 | 0,0010337 |
| **Wednesday** | **(Bankia) => (Primark)** | **11** | **0,029177719** | **0,97332846** | **0,00227414** |
| Wednesday | (Estación de Madrid Puerta de Atocha) => (Puerta del Sol) | 13 | 0,018895349 | 0,64820427 | 0,00268762 |
| **Wednesday** | **(Estación de Madrid Puerta de Atocha) => (Parque del Retiro)** | **11** | **0,015988372** | **0,52253889** | **0,00227414** |
| Wednesday | (Estación de Madrid Puerta de Atocha) => (Real Madrid Official Store) | 10 | 0,014534884 | 0,97646156 | 0,0020674 |
| **Wednesday** | **(Bankia) => (Aeropuerto Adolfo Suárez Madrid Barajas (MAD) (Aeropuerto Adolfo Suárez Madrid Barajas))** | **5** | **0,013262599** | **0,53017515** | **0,0010337** |
| Wednesday | (Estación de Madrid Puerta de Atocha) => (Estadio Santiago Bernabéu) | 8 | 0,011627907 | 0,35824322 | 0,00165392 |
| **Wednesday** | **(Estación de Madrid Puerta de Atocha) => (Primark)** | **8** | **0,011627907** | **0,38789094** | **0,00165392** |
| Wednesday | (Estación de Madrid Puerta de Atocha) => (Palacio Real de Madrid) | 7 | 0,010174419 | 0,68352309 | 0,00144718 |
| **Wednesday** | **(Estación de Madrid Puerta de Atocha) => (Aeropuerto Adolfo Suárez Madrid Barajas (MAD) (Aeropuerto Adolfo Suárez Madrid Barajas))** | **5** | **0,007267442** | **0,29051749** | **0,0010337** |
| Wednesday | (Estación de Madrid Puerta de Atocha) => (Puerta de Alcalá) | 5 | 0,007267442 | 1,00436047 | 0,0010337 |
| **Wednesday** | **(Estación de Madrid Puerta de Atocha) => (Real Madrid Official Store)-(Estación de Madrid Puerta de Atocha)** | **5** | **0,007267442** | **3,19569239** | **0,0010337** |

*Tabla 11: Patrones secuenciales de bajo nivel para los miércoles.*

| Dia | Reglas | Frecuencia | Confidence | Lift | Support |
|---|---|---|---|---|---|
| Thursday | (Hotel Praga) => (Estación de Madrid Puerta de Atocha) | 5 | 0,357142857 | 2,42759295 | 0,00100766 |
| **Thursday** | **(Museo Nacional Centro de Arte Reina Sofía (MNCARS)) => (Estación de Madrid Puerta de Atocha)** | **5** | **0,15625** | **1,06207192** | **0,00100766** |
| Thursday | (Círculo de Bellas Artes) => (Estación de Madrid Puerta de Atocha) | 5 | 0,108695652 | 0,73883264 | 0,00100766 |
| **Thursday** | **(Palacio Real de Madrid) => (Estación de Madrid Puerta de Atocha)** | **8** | **0,108108108** | **0,73483895** | **0,00161225** |
| Thursday | (Palacio Real de Madrid) => (Bankia) | 6 | 0,081081081 | 1,13972896 | 0,00120919 |
| **Thursday** | **(Templo de Debod) => (Bankia)** | **5** | **0,078125** | **1,09817635** | **0,00100766** |
| Thursday | (Plaza Mayor) => (Estación de Madrid Puerta de Atocha) | 7 | 0,075268817 | 0,51162174 | 0,00141072 |
| **Thursday** | **(Parque Juan Carlos I) => (Bankia)** | **5** | **0,073529412** | **1,03357774** | **0,00100766** |
| Thursday | (Parque Juan Carlos I) => (Estación de Madrid Puerta de Atocha) | 5 | 0,073529412 | 0,49979855 | 0,00100766 |
| **Thursday** | **(Puerta del Sol) => (Estación de Madrid Puerta de Atocha)** | **11** | **0,063218391** | **0,42971186** | **0,00221685** |
| Thursday | (Primark) => (Estación de Madrid Puerta de Atocha) | 5 | 0,060240964 | 0,40947351 | 0,00100766 |
| **Thursday** | **(Aeropuerto Adolfo Suárez Madrid Barajas (MAD) (Aeropuerto Adolfo Suárez Madrid Barajas)) => (Estación de Madrid Puerta de Atocha)** | **9** | **0,054878049** | **0,37302038** | **0,00181378** |
| Thursday | (Plaza Mayor) => (Palacio Real de Madrid) | 5 | 0,053763441 | 3,60505667 | 0,00100766 |
| **Thursday** | **(Parque del Retiro) => (Plaza Mayor)** | **6** | **0,052631579** | **2,80814941** | **0,00120919** |
| Thursday | (Parque del Retiro) => (Estación de Madrid Puerta de Atocha) | 5 | 0,043859649 | 0,29812545 | 0,00100766 |
| **Thursday** | **(Puerta del Sol) => (Bankia)** | **5** | **0,028735632** | **0,40392693** | **0,00100766** |
| Thursday | (Bankia) => (Puerta del Sol) | 8 | 0,02266289 | 0,64628309 | 0,00161225 |
| **Thursday** | **(Estación de Madrid Puerta de Atocha) => (Puerta del Sol)** | **13** | **0,017808219** | **0,50784128** | **0,00261991** |
| Thursday | (Bankia) => (Estación de Madrid Puerta de Atocha) | 6 | 0,016997167 | 0,11553417 | 0,00120919 |
| **Thursday** | **(Bankia) => (Puerta de Alcalá)** | **6** | **0,016997167** | **2,27945793** | **0,00120919** |
| Thursday | (Bankia) => (Plaza de España) | 5 | 0,014164306 | 1,11560772 | 0,00100766 |
| **Thursday** | **(Estación de Madrid Puerta de Atocha) => (Aeropuerto Adolfo Suárez Madrid Barajas (MAD) (Aeropuerto Adolfo Suárez Madrid Barajas))** | **8** | **0,010958904** | **0,33157367** | **0,00161225** |
| Thursday | (Estación de Madrid Puerta de Atocha) => (Palacio Real de Madrid) | 7 | 0,009589041 | 0,64298408 | 0,00141072 |
| **Thursday** | **(Estación de Madrid Puerta de Atocha) => (Plaza Mayor)** | **7** | **0,009589041** | **0,51162174** | **0,00141072** |
| Thursday | (Estación de Madrid Puerta de Atocha) => (Plaza de Castilla) | 5 | 0,006849315 | 0,97103718 | 0,00100766 |
| **Thursday** | **(Estación de Madrid Puerta de Atocha) => (Templo de Debod)** | **5** | **0,006849315** | **0,53103596** | **0,00100766** |

*Tabla 12:Patrones secuenciales de bajo nivel para los jueves.*

| Dia | Reglas | Frecuencia | Confidence | Lift | Support |
|---|---|---|---|---|---|
| Friday | (Estación de Madrid Puerta de Atocha)-(Plaza Mayor) => (Estación de Madrid Puerta de Atocha) | 5 | 0,83333333 | 6,33410973 | 0,00102145 |
| Friday | (Estación de Madrid Puerta de Atocha)-(Puerta del Sol) => (Estación de Madrid Puerta de Atocha) | 5 | 0,38461538 | 2,92343526 | 0,00102145 |
| Friday | (Santa Iglesia Catedral de Santa María la Real de la Almudena) => (Palacio Real de Madrid) | 5 | 0,26315789 | 13,4183114 | 0,00102145 |
| Friday | (Mercado de San Miguel) => (Estación de Madrid Puerta de Atocha) | 5 | 0,18518519 | 1,40757994 | 0,00102145 |
| Friday | (Mercado de San Miguel) => (Puerta del Sol) | 5 | 0,18518519 | 5,7737674 | 0,00102145 |
| Friday | (Plaza de Castilla) => (Estación de Madrid Puerta de Atocha) | 7 | 0,17948718 | 1,36426979 | 0,00143003 |
| Friday | (Real Madrid Official Store) => (Estación de Madrid Puerta de Atocha) | 7 | 0,17948718 | 1,36426979 | 0,00143003 |
| Friday | (Plaza de Cibeles) => (Estación de Madrid Puerta de Atocha) | 5 | 0,17857143 | 1,35730923 | 0,00102145 |
| Friday | (Puerta de Alcalá) => (Estación de Madrid Puerta de Atocha) | 5 | 0,12195122 | 0,92694289 | 0,00102145 |
| Friday | (Estadio Santiago Bernabéu) => (Estación de Madrid Puerta de Atocha) | 6 | 0,07894737 | 0,60007355 | 0,00122574 |
| Friday | (Primark) => (Estación de Madrid Puerta de Atocha) | 7 | 0,07777778 | 0,59118357 | 0,00143003 |
| Friday | (Puerta del Sol) => (Bankia) | 12 | 0,07643312 | 1,06897179 | 0,00245148 |
| Friday | (Puerta del Sol) => (Estación de Madrid Puerta de Atocha) | 12 | 0,07643312 | 0,58096293 | 0,00245148 |
| Friday | (Plaza Mayor) => (Puerta del Sol) | 7 | 0,07070707 | 2,20452937 | 0,00143003 |
| Friday | (Estadio Santiago Bernabéu) => (Bankia) | 5 | 0,06578947 | 0,92011278 | 0,00102145 |
| Friday | (Parque del Retiro) => (Estación de Madrid Puerta de Atocha) | 9 | 0,06382979 | 0,48516585 | 0,00183861 |
| Friday | (Palacio Real de Madrid) => (Bankia) | 6 | 0,0625 | 0,87410714 | 0,00122574 |
| Friday | (Plaza Mayor) => (Bankia) | 6 | 0,06060606 | 0,84761905 | 0,00122574 |
| Friday | (Plaza Mayor) => (Estación de Madrid Puerta de Atocha) | 6 | 0,06060606 | 0,46066253 | 0,00122574 |
| Friday | (Primark) => (Bankia) | 5 | 0,05555556 | 0,77698413 | 0,00102145 |
| Friday | (Palacio Real de Madrid) => (Plaza Mayor) | 5 | 0,05208333 | 2,57523148 | 0,00102145 |
| Friday | (Palacio Real de Madrid) => (Santa Iglesia Catedral de Santa María la Real de la Almudena) | 5 | 0,05208333 | 13,4183114 | 0,00102145 |
| Friday | (Aeropuerto Adolfo Suárez Madrid Barajas (MAD) (Aeropuerto Adolfo Suárez Madrid Barajas)) => (Estación de Madrid Puerta de Atocha) | 9 | 0,0505618 | 0,38431677 | 0,00183861 |
| Friday | (Plaza Mayor) => (Palacio Real de Madrid) | 5 | 0,05050505 | 2,57523148 | 0,00102145 |
| Friday | (Parque del Retiro) => (Palacio de Cristal del Retiro) | 7 | 0,04964539 | 4,26340674 | 0,00143003 |
| Friday | (Puerta del Sol) => (Primark) | 6 | 0,03821656 | 2,07855626 | 0,00122574 |
| Friday | (Parque del Retiro) => (Plaza Mayor) | 5 | 0,03546099 | 1,75334909 | 0,00102145 |
| Friday | (Puerta del Sol) => (Mercado de San Miguel) | 5 | 0,03184713 | 5,7737674 | 0,00102145 |
| Friday | (Bankia) => (Bankia)-(Bankia) | 10 | 0,02857143 | 3,17857143 | 0,0020429 |
| Friday | (Bankia) => (Plaza Mayor) | 9 | 0,02571429 | 1,27142857 | 0,00183861 |
| Friday | (Bankia) => (Puerta del Sol) | 9 | 0,02571429 | 0,80172884 | 0,00183861 |
| Friday | (Bankia) => (Primark) | 8 | 0,02285714 | 1,2431746 | 0,00163432 |
| Friday | (Estación de Madrid Puerta de Atocha) => (Puerta del Sol) | 13 | 0,02018634 | 0,62937651 | 0,00265577 |
| Friday | (Estación de Madrid Puerta de Atocha) => (Parque del Retiro) | 11 | 0,01708075 | 0,59298049 | 0,00224719 |
| Friday | (Bankia) => (Estadio Santiago Bernabéu) | 5 | 0,01428571 | 0,92011278 | 0,00102145 |
| Friday | (Bankia) => (Palacio Real de Madrid) | 5 | 0,01428571 | 0,72842262 | 0,00102145 |
| Friday | (Bankia) => (Parque del Retiro) | 5 | 0,01428571 | 0,49594732 | 0,00102145 |
| Friday | (Estación de Madrid Puerta de Atocha) => (Plaza de España) | 7 | 0,01086957 | 0,80615942 | 0,00143003 |
| Friday | (Estación de Madrid Puerta de Atocha) => (Aeropuerto Adolfo Suárez Madrid Barajas (MAD) (Aeropuerto Adolfo Suárez Madrid Barajas)) | 6 | 0,00931677 | 0,25621118 | 0,00122574 |
| Friday | (Estación de Madrid Puerta de Atocha) => (Plaza Mayor) | 6 | 0,00931677 | 0,46066253 | 0,00122574 |
| Friday | (Estación de Madrid Puerta de Atocha) => (Plaza del Dos de Mayo) | 6 | 0,00931677 | 0,61629176 | 0,00122574 |
| Friday | (Estación de Madrid Puerta de Atocha) => (Primark) | 6 | 0,00931677 | 0,50672878 | 0,00122574 |
| Friday | (Estación de Madrid Puerta de Atocha) => (Círculo de Bellas Artes) | 5 | 0,00776398 | 1,11778407 | 0,00102145 |
| Friday | (Estación de Madrid Puerta de Atocha) => (Jardines de Nuevos Ministerios) | 5 | 0,00776398 | 3,16705487 | 0,00102145 |

*Tabla 13: Patrones secuenciales de bajo nivel para los viernes.*

| Dia | Reglas | Frecuencia | Confidence | Lift | Support |
|---|---|---|---|---|---|
| Saturday | (Palacio de Cristal del Retiro) => (Parque del Retiro) | 13 | 0,160493827 | 4,29456671 | 0,00234699 |
| Saturday | (Primark) => (Estación de Madrid Puerta de Atocha) | 9 | 0,08490566 | 0,69569889 | 0,00162484 |
| Saturday | (Puerta del Sol) => (Bankia) | 13 | 0,068421053 | 0,98951491 | 0,00234699 |
| Saturday | (Plaza del Dos de Mayo) => (Estación de Madrid Puerta de Atocha) | 6 | 0,063829787 | 0,52300768 | 0,00108323 |
| Saturday | (Parque del Retiro) => (Estación de Madrid Puerta de Atocha) | 12 | 0,057971014 | 0,47500214 | 0,00216646 |
| Saturday | (Estadio Santiago Bernabéu) => (Estación de Madrid Puerta de Atocha) | 7 | 0,053435115 | 0,43783595 | 0,00126377 |
| Saturday | (Aeropuerto Adolfo Suárez Madrid Barajas (MAD) (Aeropuerto Adolfo Suárez Madrid Barajas)) => (Estación de Madrid Puerta de Atocha) | 7 | 0,050359712 | 0,41263675 | 0,00126377 |
| Saturday | (Parque del Retiro) => (Palacio de Cristal del Retiro) | 9 | 0,043478261 | 2,97316157 | 0,00162484 |
| Saturday | (Plaza Mayor) => (Bankia) | 6 | 0,042857143 | 0,61980604 | 0,00108323 |
| Saturday | (Plaza Mayor) => (Estación de Madrid Puerta de Atocha) | 6 | 0,042857143 | 0,3511623 | 0,00108323 |
| Saturday | (Plaza Mayor) => (Puerta del Sol) | 6 | 0,042857143 | 1,2493985 | 0,00108323 |
| Saturday | (Puerta del Sol) => (Estación de Madrid Puerta de Atocha) | 8 | 0,042105263 | 0,34500156 | 0,0014443 |
| Saturday | (Puerta del Sol) => (Plaza Mayor) | 7 | 0,036842105 | 1,45763158 | 0,00126377 |
| Saturday | (Parque del Retiro) => (Puerta del Sol) | 6 | 0,028985507 | 0,84500381 | 0,00108323 |
| Saturday | (Bankia) => (Parque del Retiro) | 11 | 0,028720627 | 0,76851957 | 0,00198592 |
| Saturday | (Bankia) => (Puerta del Sol) | 10 | 0,026109661 | 0,76116532 | 0,00180538 |
| Saturday | (Bankia) => (Palacio Real de Madrid) | 9 | 0,023498695 | 1,4624637 | 0,00162484 |
| Saturday | (Bankia) => (Plaza Mayor) | 9 | 0,023498695 | 0,92970906 | 0,00162484 |
| Saturday | (Bankia) => (Aeropuerto Adolfo Suárez Madrid Barajas (MAD) (Aeropuerto Adolfo Suárez Madrid Barajas)) | 8 | 0,020887728 | 0,83235344 | 0,0014443 |
| Saturday | (Estación de Madrid Puerta de Atocha) => (Parque del Retiro) | 11 | 0,016272189 | 0,43541863 | 0,00198592 |
| Saturday | (Bankia) => (Mercado de San Miguel) | 6 | 0,015665796 | 2,41035683 | 0,00108323 |
| Saturday | (Bankia) => (Plaza de España) | 6 | 0,015665796 | 1,42250567 | 0,00108323 |
| Saturday | (Estación de Madrid Puerta de Atocha) => (Puerta del Sol) | 9 | 0,013313609 | 0,38812675 | 0,00162484 |
| Saturday | (Estación de Madrid Puerta de Atocha) => (Puerta de Alcalá) | 7 | 0,01035503 | 1,72035125 | 0,00126377 |
| Saturday | (Estación de Madrid Puerta de Atocha) => (Plaza del Dos de Mayo) | 6 | 0,00887574 | 0,52300768 | 0,00108323 |
| Saturday | (Estación de Madrid Puerta de Atocha) => (Primark) | 6 | 0,00887574 | 0,46379926 | 0,00108323 |

*Tabla 14:Patrones secuenciales de bajo nivel para los sábados.*

| Dia | Reglas | Frecuencia | Confidence | Lift | Support |
|---|---|---|---|---|---|
| Sunday | (Estadio Vicente Calderón) => (Avenida De Los Poblados) | 15 | 0,133928571 | 10,7803571 | 0,00248468 |
| Sunday | {Museo Nacional del Prado} => {Parque del Retiro} | 7 | 0,122807018 | 3,01376409 | 0,00115952 |
| Sunday | (Avenida De Los Poblados) => (Estadio Vicente Calderón) | 9 | 0,12 | 6,46821429 | 0,00149081 |
| Sunday | {Puerta del Sol} => {Estación de Madrid Puerta de Atocha} | 18 | 0,090909091 | 0,71647282 | 0,00298161 |
| Sunday | (Palacio de Cristal del Retiro) => (Parque del Retiro) | 7 | 0,086419753 | 2,12079695 | 0,00115952 |
| Sunday | {Templo de Debod} => {Estación de Madrid Puerta de Atocha} | 7 | 0,081395349 | 0,64149311 | 0,00115952 |
| Sunday | (Palacio Real de Madrid) => (Estación de Madrid Puerta de Atocha) | 7 | 0,079545455 | 0,62691372 | 0,00115952 |
| Sunday | {Palacio Real de Madrid} => {Puerta del Sol} | 7 | 0,079545455 | 2,42533287 | 0,00115952 |
| Sunday | (Puerta del Sol) => (Bankia) | 15 | 0,075757576 | 1,00076255 | 0,00248468 |
| Sunday | {Plaza Mayor} => {Estación de Madrid Puerta de Atocha} | 9 | 0,055214724 | 0,43515834 | 0,00149081 |
| Sunday | (Parque del Retiro) => (Estación de Madrid Puerta de Atocha) | 11 | 0,044715447 | 0,35241143 | 0,0018221 |
| Sunday | {Plaza Mayor} => {Puerta del Sol} | 7 | 0,042944785 | 1,30938217 | 0,00115952 |
| Sunday | (Estadio Santiago Bernabéu) => (Estación de Madrid Puerta de Atocha) | 7 | 0,036649215 | 0,28883983 | 0,00115952 |
| Sunday | {Parque del Retiro} => {Palacio de Cristal del Retiro} | 8 | 0,032520325 | 2,42376794 | 0,00132516 |
| Sunday | (Bankia) => (Puerta del Sol) | 10 | 0,021881838 | 0,66717503 | 0,00165645 |
| Sunday | {Bankia} => {Aeropuerto Adolfo Suárez Madrid Barajas (MAD) (Aeropuerto Adolfo Suárez Madrid Barajas)} | 8 | 0,01750547 | 0,63281752 | 0,00132516 |
| Sunday | (Estación de Madrid Puerta de Atocha) => (Parque del Retiro) | 12 | 0,015665796 | 0,38444883 | 0,00198774 |
| Sunday | {Bankia} => {Palacio Real de Madrid} | 7 | 0,015317287 | 1,05080068 | 0,00115952 |
| Sunday | (Bankia) => (Plaza Mayor) | 7 | 0,015317287 | 0,56730343 | 0,00115952 |
| Sunday | {Estación de Madrid Puerta de Atocha} => {Puerta del Sol} | 11 | 0,014360313 | 0,4378445 | 0,0018221 |
| Sunday | (Estación de Madrid Puerta de Atocha) => (Estadio Santiago Bernabéu) | 8 | 0,010443864 | 0,33010266 | 0,00132516 |
| Sunday | {Estación de Madrid Puerta de Atocha} => {Estadio Vicente Calderón} | 7 | 0,009138381 | 0,49257507 | 0,00115952 |
| Sunday | (Estación de Madrid Puerta de Atocha) => (Museo Nacional del Prado) | 7 | 0,009138381 | 0,96786679 | 0,00115952 |
| Sunday | {Estación de Madrid Puerta de Atocha} => {Plaza Mayor} | 7 | 0,009138381 | 0,33845649 | 0,00115952 |

*Tabla 15:Patrones secuenciales de bajo nivel para los domingos.*

Se observa que la mayoría de los patrones de trayectorias obtenidos para los días de la semana se corresponden con desplazamiento típicos de recorridos de turistas, tal y como se observó en el análisis de asociación del capítulo anterior.

Alguna de las trayectorias obtenidas mediante este análisis que evidencian patrones de movilidad entre turistas se encuentran:(Palacio Real de Madrid)→ (Santa Iglesia Catedral de Santa María la Real de la Almudena); individuos que visitan el Palacio Real de Madrid visitan a continuación la Catedral de la Almudena. También, (Palacio Real de Madrid)→(Estadio Santiago Bernabéu), los individuos que visitan el Palacio Real de Madrid visitaran seguidamente el estadio Santiago Bernabéu.

Con una frecuencia de 18 repeticiones, la ruta más frecuente obtenida en este estudio ocurre el domingo y está dada por los individuos que visitan la Puerta del Sol lo cuales visitaran seguidamente la Estación de Madrid Puerta de Atocha: (Puerta del Sol) → (Estación de Madrid Puerta de Atocha). Esta ruta puede servir para sustentar una mejora de la frecuencia de servicios de transporte público que cubre esta ruta durante los domingos.

Otro patrón de interés observado entre las rutas obtenidas corresponde a individuos que los domingos visitan el Estadio Vicente Calderón y se desplazan a continuación a la Avenida De Los Poblados (Estadio Vicente Calderón) → (Avenida De Los Poblados). Esta trayectoria nos muestra patrones de desplazamiento de individuos que van a un partido de futbol el domingo. Otras trayectorias que muestran característica de desplazamiento con esquemas relacionados al futbol, también los domingos son: Estadio Santiago Bernabéu → Estación de Madrid Puerta de Atocha, Estación de Madrid Puerta de Atocha → Estadio Vicente Calderón, Estación de Madrid Puerta de Atocha → Estadio Santiago Bernabéu.

Es posible observar para otros días de la semana trayectorias con esquemas relacionados con el futbol, especialmente desde y hacia el Estadio Santiago Bernabéu. Dichas trayectorias, aunque pueden estar relacionadas a partidos de futbol, también podrían ser generadas por turistas que visitan el estadio.

A partir de las reglas obtenidas se generan grafos para poder observar de forma más clara los patrones de movimientos en función de los días de la semana. Cada vértice corresponde con una ubicación de bajo nivel. Las aristas muestran el desplazamiento y la frecuencia de dicho desplazamiento desde una ubicación hacia otra.

Se observa en todos los grafos, que las mayorías de las rutas convergen en la ubicación Estación de Madrid Puerta de Atocha y que provienen desde ubicaciones principalmente turísticas tales como Puerta del Sol, Plaza Mayor y Parque del Retiro entre otras.

A partir de los grafos construidos, se observa de forma más clara como el viernes es el día que muestra más aristas en comparación con todos los día de la semana, con un total de 41 trayectorias.

Es importante destacar la existencia de trayectorias entre ubicaciones que se dan de forma bidireccional tales como: Parque del Retiro↔ Palacio de Cristal, Estadio Vicente Calderón↔ Avenida De Los Poblados, Mercado de San Miguel↔ Puerta del Sol, Catedral de la Almudena↔ Palacio Real.

Es válido afirmar, que los desplazamientos entre las trayectorias que contienen ubicaciones relacionadas con medios de transporte como por ejemplo: Estación de Madrid Puerta de Atocha Aeropuerto↔ Adolfo Suarez Madrid Barajas, Estación de Madrid Puerta de Atocha↔ Real Madrid Oficial Store, se realicen por medio de dichos medios de transporte. En cambio, aquellas trayectorias que no contengan ubicaciones asociadas a medios de transporte, es válido inferir que muy probablemente dichos desplazamiento se han realizado a pie, especialmente en trayectos cortos tales como: Puerta del Sol↔ Plaza Mayor o Catedral de la Almudena ↔ Palacio Real.

Conociendo la secuencias de las visitas entre cada una de las ubicaciones, es posible indagar sobre ciertas características intrínsecas de estos desplazamiento a nivel espacial, de esta forma es posible evaluar la distancias que existe entre cada una de las rutas de viajes frecuentes obtenidas.

En la ilustración 34 se muestra la distribución de la distancia de las rutas de viajes frecuentes y la frecuencia de las mismas para cada uno de los días de la semana. El 51% del total de las rutas de viaje frecuentes están separadas por una distancia menor o igual a la mediana del total de las distancias entre rutas (2.412mts). Así mismo, el 81% de las rutas tiene una distancia menor o igual al valor de la media de las distancia entre cada una de las rutas (4.018mts).

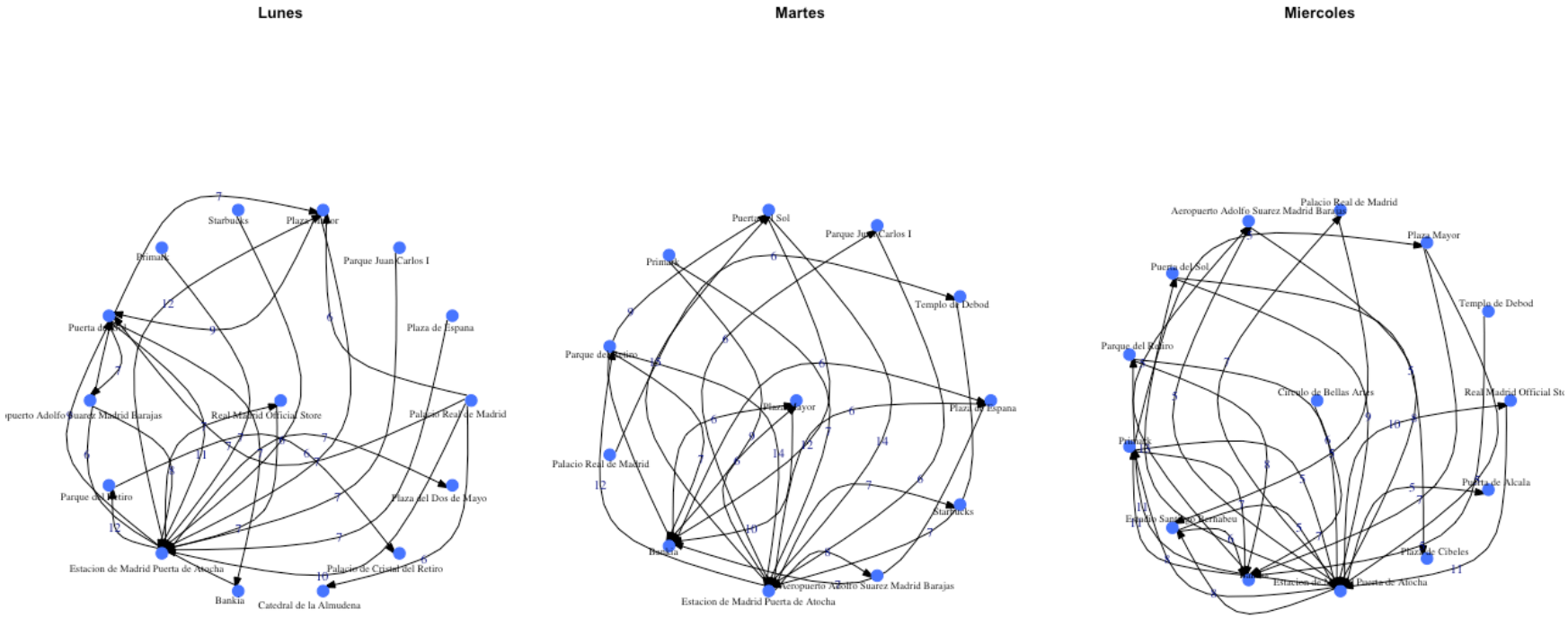



*Ilustración 31: Grafo de patrones secuenciales de bajo nivel para los Lunes, Martes y Miércoles.*

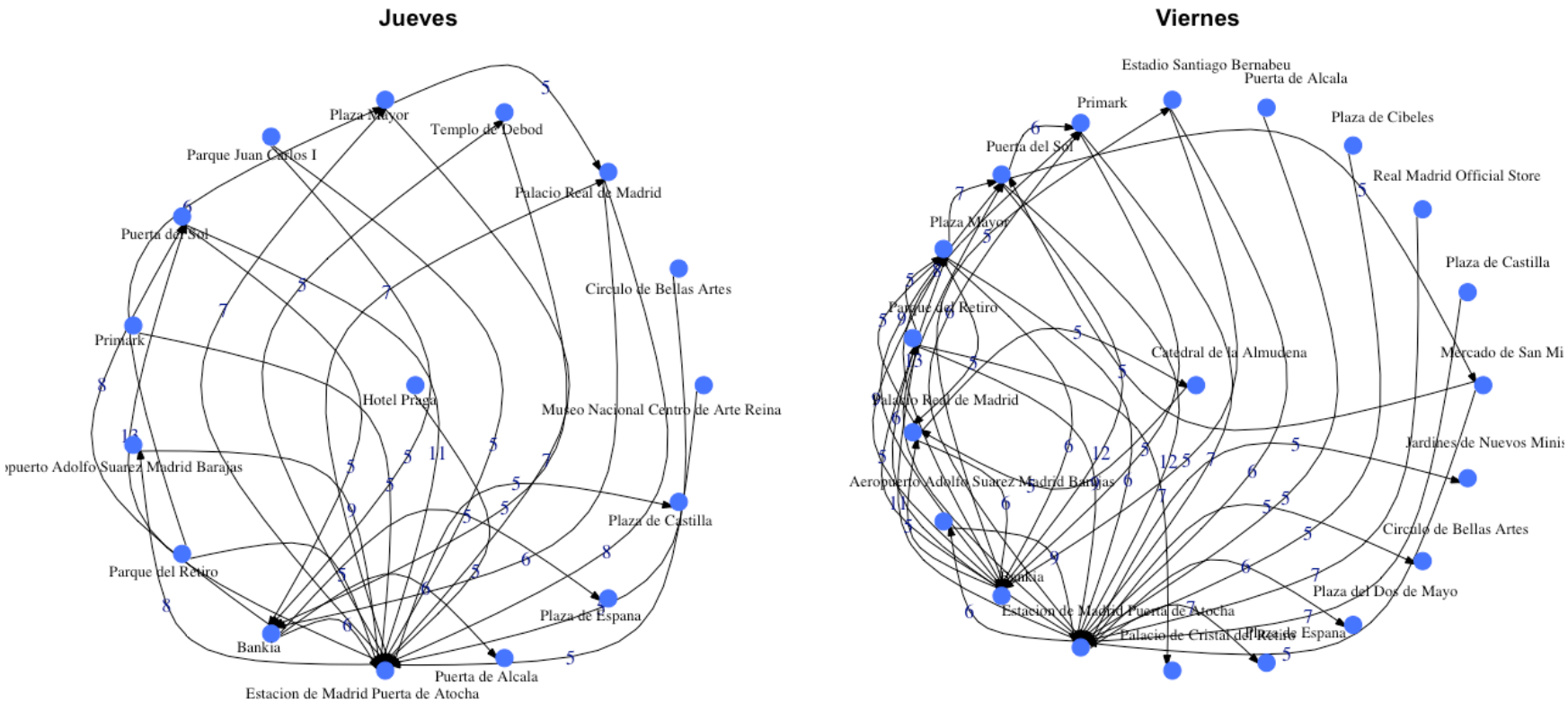



*Ilustración 32: Grafo de patrones secuenciales de bajo nivel para los Jueves y Viernes.*

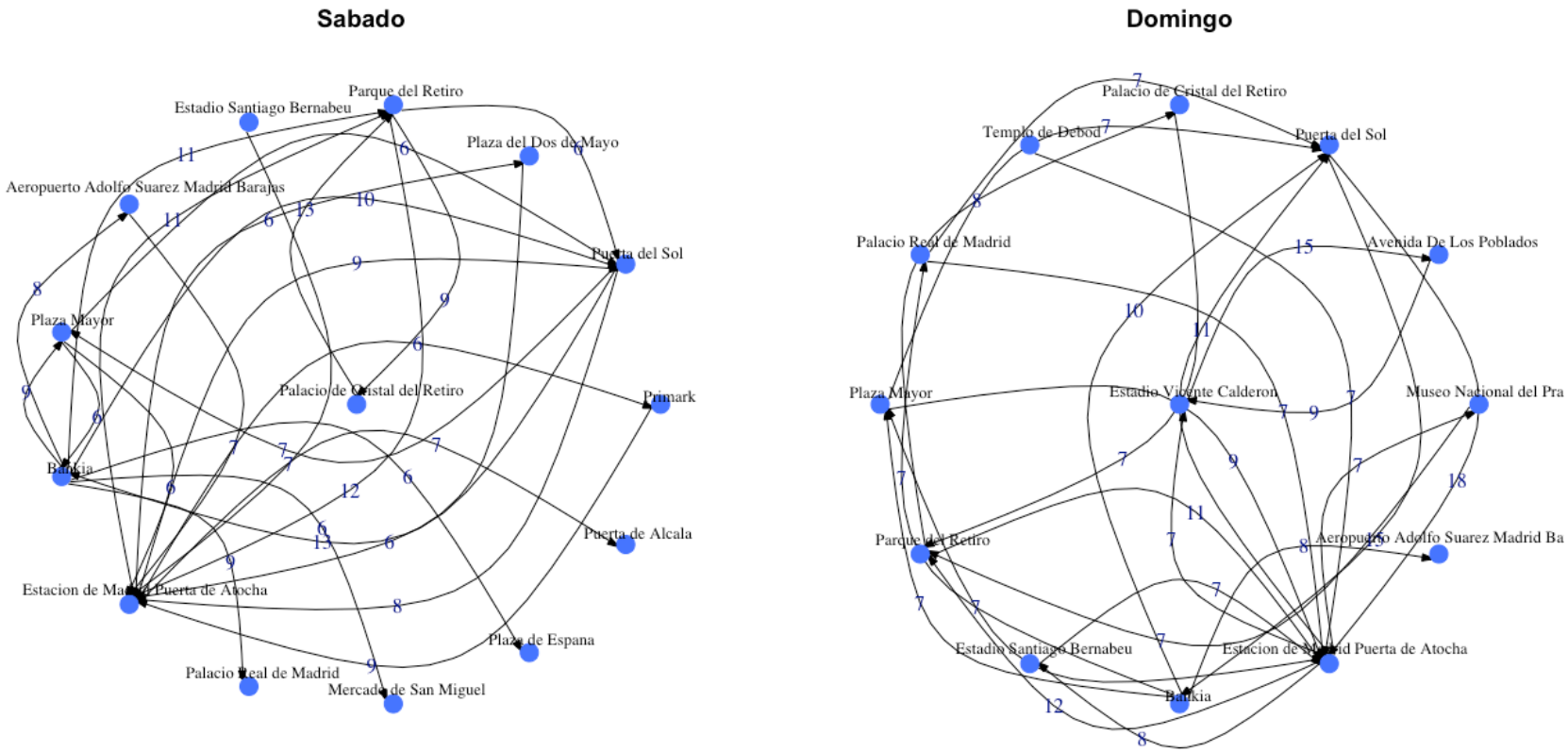



*Ilustración 33: Grafo de patrones secuenciales de bajo nivel para los sábados y domingos.*

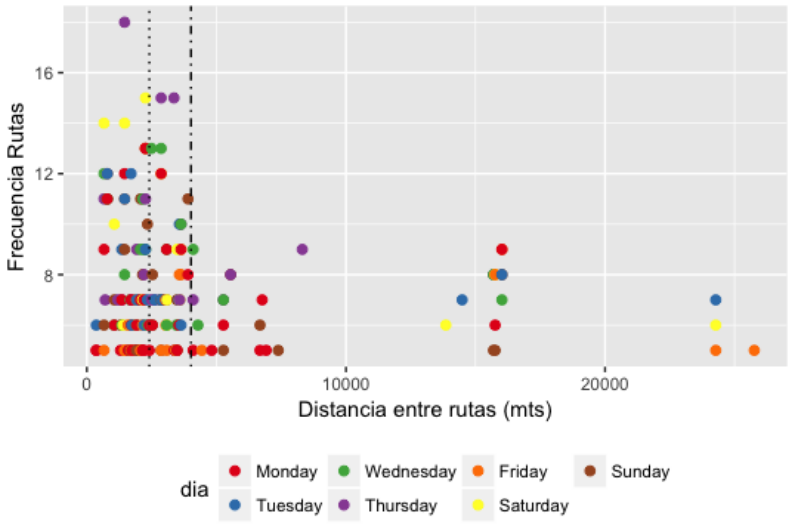



*Ilustración 34: Distribución de la distancia de las rutas en función de la frecuencia de las mismas y los días de la semana.*

Al evaluar la distancia de las rutas de viajes frecuentes, es posible ofrecer por ejemplo mejoras al servicio, asociado al minado de patrones de trayectorias. Por ejemplo, los viernes, una ruta de viaje frecuente es Puerta del Sol ↔ Mercado de San Miguel. Es posible, basado en este estudio, crear un motor de recomendaciones que permita, especialmente los viernes impactar con publicidad y ofertas asociadas a locales del Mercado de San Miguel y recomendaciones sobre la ruta más corta para ir al mercado de San Miguel y viceversa.

En general y a partir de los resultados de la caracterización de las trayectorias obtenidas, y de las ubicaciones que pueda visitar un individuo después de haber visitado anteriormente otras ubicaciones, es posible implementar un sistema de recomendaciones de viajes que:

1. Sugiera la ruta más corta entre dos ubicaciones: Esta información puede ser de ayuda para viajeros que quiera visitar de forma óptima la mayor cantidad de ubicaciones posibles empleado para ello el menor tiempo entre desplazamientos.
2. Sugerir la ruta más frecuente entre varias ubicaciones.
3. Predicción de la siguiente ubicación: Predecir un listado de ubicaciones en las cuales un individuo se pueda desplazar basado en sus recientes movimientos y modelos de patrones de trayectorias y de asociación. Esta información puede ser útil para anticipar posibles servicios en la futura ubicación que se va a visitar.
4. Gestión de tráfico: Predecir patrones de congestión de tráfico y mejorar los modelos de transporte dentro de la ciudad.

# CAPÍTULO VI – CONCLUSIONES

En este trabajo se ha presentado un metodología para la extracción y análisis de datos georreferenciados obtenidos a través de la red social Twitter. Esta metodología consiste en establecer una conexión por medio del API de Twitter y así poder obtener todos los tweets georreferenciados para su análisis.

Los datos georreferenciados por sí solo no presenta ningún significado en cuanto a la ubicación a la cual representa. Para ello, se ha empleado una metodología que permite la asignación de una semántica la cual permite caracterizar dichas ubicaciones en un contexto de bajo, medio y alto nivel; además de una semántica que caracteriza cada ubicación dentro del contexto geográfico de Madrid capital y que corresponde con cada uno de los barrios de la ciudad. La metodología ideada para la asignación de la categorización de alto, medio y bajo nivel, emplea la base de batos de ubicaciones de la red social Foursquare. De esta forma, cada una de las categorías correspondientes a cada uno de las ubicaciones se obtienen pasando por medio de la consulta API de dicha red social, las coordenadas geográficas de cada una de las ubicaciones.

Para llevar a cabo el estudio de detección de trayectorias y patrones de movilidad, se han empleado dos tipos de análisis; el análisis de asociación y el análisis secuencial. Como resultado de la aplicación del algoritmo Apriori en el análisis de asociación y el algoritmo prefix-tree-based search en al análisis secuencial, se han obtenido asociaciones y reglas de desplazamientos que evidencia la movilidad dentro de la ciudad de Madrid.

A partir del análisis de asociación por medio de las categorizaciones de bajo nivel se pueden destacar las siguientes 11 asociaciones más importantes en función de su valor de confidence.

| rules | support | confidence | lift |
|---|---|---|---|
| {Estación de Madrid-Puerta de Atocha,Plaza Mayor,Puerta de Alcalá} => {Puerta del Sol} | 0.00104991116136327 | 0.764705882352941 | 9.31947660954145 |
| {Puerta de Alcalá,Santa Iglesia Catedral de Santa María la Real de la Almudena} => {Palacio Real de Madrid} | 0.00104991116136327 | 0.777777777777778 | 17.0659457167091 |
| {Estación de Madrid-Puerta de Atocha,Palacio de Cristal del Retiro,Plaza Mayor} => {Parque del Retiro} | 0.00104991116136327 | 0.722222222222222 | 8.54928829403017 |
| {Aeropuerto Adolfo Suárez Madrid-Barajas (MAD) (Aeropuerto Adolfo Suárez Madrid-Barajas),Real Madrid Official Store} => {Estación de Madrid-Puerta de Atocha} | 0.00104991116136327 | 0.777777777777778 | 2.69277897668030384 |
| {Museo Nacional del Prado,Santa Iglesia Catedral de Santa María la Real de la Almudena} => {Estación de Madrid-Puerta de Atocha} | 0.00121143595541916 | 0.714285714285714 | 2.6631393298066 |
| {Santa Iglesia Catedral de Santa María la Real de la Almudena,Templo de Debod} => {Palacio Real de Madrid} | 0.00113067355839121 | 0.7 | 16.5408396946565 |
| {El Rincón Secreto} => {Estación de Madrid-Puerta de Atocha} | 0.00113067355839121 | 0.7 | 2.60987654320988 |
| {Aeropuerto Adolfo Suárez Madrid-Barajas (MAD) (Aeropuerto Adolfo Suárez Madrid-Barajas),Santa Iglesia Catedral de Santa María la Real de la Almudena} => {Palacio Real de Madrid} | 0.00121143595541916 | 0.681818181818182 | 16.1112074947953 |
| {Palacio de Cristal del Retiro,Santa Iglesia Catedral de Santa María la Real de la Almudena} => {Palacio Real de Madrid} | 0.00104991116136327 | 0.65 | 15.3593511450382 |
| {Palacio Real de Madrid,Parque del Retiro,Plaza Mayor} => {Puerta del Sol} | 0.00121143595541916 | 0.625 | 7.61687992125984 |
| {Mercado de San Miguel,Templo de Debod} => {Palacio Real de Madrid} | 0.00104991116136327 | 0.619047619047619 | 14.6279534714649 |

*Tabla 16: Mejores 11 reglas de asociación para categorización de bajo nivel en función del valor de confidence.*

Se observan que los resultados obtenidos muestran asociaciones entre ubicaciones frecuentadas por turistas. Es importante destacar que estas asociaciones obtenidas no representan la totalidad de los turistas, sino más bien a todos aquellos que hacen usos de las

redes sociales para documentar las visitas a diversas ubicaciones. Dentro de este grupo de turistas se pueden incluir: tecnológicos, jóvenes, etc.

De igual forma, se determinan las reglas de asociación para ubicaciones categorizadas a nivel medio. A continuación se muestran las 11 reglas de asociación obtenidas ordenadas según su valor de confidence.

| rules | support | confidence | lift |
|---|---|---|---|
| {Bar,Mediterranean Restaurant,Restaurant} => {Spanish Restaurant} | 0.0012921983524471 | 1 | 11.5181395348837 |
| {Bar,Clothing Store,Neighborhood,Plaza} => {Park} | 0.00104991116136327 | 1 | 5.1228796028134 |
| {Breakfast Spot,Park,Restaurant} => {Plaza} | 0.00137296074947504 | 1 | 3.30539241857982 |
| {Bus Station,Hotel,Spanish Restaurant} => {Plaza} | 0.00104991116136327 | 1 | 3.30539241857982 |
| {Cocktail Bar,Monument / Landmark,Park} => {Plaza} | 0.0012921983524471 | 1 | 3.30539241857982 |
| {Bar,Department Store,Neighborhood} => {Plaza} | 0.00104991116136327 | 1 | 3.30539241857982 |
| {Building,CafÈ,Train Station} => {Plaza} | 0.00104991116136327 | 1 | 3.30539241857982 |
| {Building,Clothing Store,Monument / Landmark} => {Plaza} | 0.00104991116136327 | 1 | 3.30539241857982 |
| {Building,Hotel,Monument / Landmark} => {Plaza} | 0.00113067355839121 | 1 | 3.30539241857982 |
| {CafÈ,Coffee Shop,Monument / Landmark} => {Plaza} | 0.00121143595541916 | 1 | 3.30539241857982 |
| {CafÈ,Clothing Store,Monument / Landmark} => {Plaza} | 0.00121143595541916 | 1 | 3.30539241857982 |

*Tabla 17: Mejores 11 reglas de asociación para categorización de ubicaciones a nivel medio con el mejor valor de confidence.*

Al igual que en las categorizaciones de bajo nivel, algunas de las reglas de asociaciones obtenidas se corresponden con desplazamientos de turistas. Además, se pueden observar varias reglas que reflejan movimientos cotidianos de la población, tales como entre restaurantes, estaciones de autobuses, bares, entre otros.

Seguidamente se ha estudiado la movilidad en función de las asociaciones ente los barrios de Madrid capital. Los resultados de este estudio muestran los movimientos entre los principales barrios y los cuales se encuentra influenciados por desplazamientos originado generalmente por estudiantes y turistas. Las 11 mejores reglas de asociaciones para la categorizaciones a nivel de barrio se muestra a continuación.

| rules | support | confidence | lift |
|---|---|---|---|
| {Casa de Campo,Cortes,Jerónimos,Palacio} => {Sol} | 0.00314973348408981 | 0.928571428571429 | 3.47884158201859 |
| {Atocha,Cortes,Embajadores,Jerónimos,Palacio} => {Sol} | 0.00323049588111775 | 0.869565217391304 | 3.25777807011774 |
| {Cortes,Embajadores,Jerónimos,Palacio} => {Sol} | 0.00428040704248102 | 0.854838709677419 | 3.20260602215607 |
| {Atocha,Cortes,Jerónimos,Recoletos,Sol} => {Palacio} | 0.00306897108706186 | 0.844444444444444 | 4.77003244120032 |
| {Atocha,Embajadores,Palacio,Recoletos} => {Sol} | 0.0033112582781457 | 0.836734693877551 | 3.13478032665411 |
| {Casa de Campo,Cortes,Jerónimos} => {Sol} | 0.00411888224842513 | 0.836065573770492 | 3.1322735051214 |
| {Atocha,Casa de Campo,Cortes,Palacio} => {Sol} | 0.00363430786625747 | 0.833333333333333 | 3.12203731719617 |
| {Atocha,Casa de Campo,Jerónimos,Palacio} => {Sol} | 0.00347278307220158 | 0.826923076923077 | 3.09802164552543 |
| {Cortes,Jerónimos,Justicia,Palacio} => {Sol} | 0.00363430786625747 | 0.818181818181818 | 3.06527300233806 |
| {Cortes,Jerónimos,Palacio,Universidad} => {Sol} | 0.00492650621870457 | 0.813333333333333 | 3.04710842158346 |
| {Casa de Campo,Cortes,Palacio} => {Sol} | 0.0051687934097884 | 0.810126582278481 | 3.03509450583121 |

*Tabla 18: Mejores 11 reglas de asociación para categorización de ubicaciones en función de los barrios de Madrid capital con el mejor valor de confidence.*

A continuación, mediante al análisis secuencial, se determinan la rutas de viajes más importantes para cada uno de los días de la semana. En la tabla que se muestra a

continuación, se indican las 20 rutas de viajes frecuentes más relevantes en función de la frecuencias que aparece como resultado de este conjunto de datos.

Se aprecia que la ruta de viaje más frecuente se da entre la Puerta del Sol hacia la estación Madrid Puerta de Atocha durante los domingos.

Es importante resaltar como la Estación de Madrid Puerta de Atocha actúa de punto de partida para la mayoría de rutas de viajes más relevantes.

| Día | Frecuencia | Confidence | Lift | Support | Ubicación1 | Ubicación2 |
|---|---|---|---|---|---|---|
| Sunday | 18 | 0,090909091 | 0,71647282 | 0,00298161 | Puerta del Sol | Estación de Madrid Puerta de Atocha |
| **Sunday** | **15** | **0,133928571** | **10,7803571** | **0,00248468** | **Estadio Vicente Calderón** | **Avenida De Los Poblados** |
| Sunday | 15 | 0,075757576 | 1,00076255 | 0,00248468 | Puerta del Sol | Bankia |
| **Tuesday** | **15** | **0,020408163** | **0,61030126** | **0,00298567** | **Estación de Madrid Puerta de Atocha** | **Puerta del Sol** |
| Tuesday | 14 | 0,083333333 | 0,56961451 | 0,00278662 | Puerta del Sol | Estación de Madrid Puerta de Atocha |
| **Tuesday** | **14** | **0,075268817** | **0,51449053** | **0,00278662** | **Parque del Retiro** | **Estación de Madrid Puerta de Atocha** |
| Saturday | 13 | 0,160493827 | 4,29456671 | 0,00234699 | Palacio de Cristal del Retiro | Parque del Retiro |
| **Saturday** | **13** | **0,068421053** | **0,98951491** | **0,00234699** | **Puerta del Sol** | **Bankia** |
| Friday | 13 | 0,020186335 | 0,62937651 | 0,00265577 | Estación de Madrid Puerta de Atocha | Puerta del Sol |
| **Wednesday** | **13** | **0,018895349** | **0,64820427** | **0,00268762** | **Estación de Madrid Puerta de Atocha** | **Puerta del Sol** |
| Thursday | 13 | 0,017808219 | 0,50784129 | 0,00261991 | Estación de Madrid Puerta de Atocha | Puerta del Sol |
| **Friday** | **12** | **0,076433121** | **1,06897179** | **0,00245148** | **Puerta del Sol** | **Bankia** |
| Friday | 12 | 0,076433121 | 0,58096293 | 0,00245148 | Puerta del Sol | Estación de Madrid Puerta de Atocha |
| **Tuesday** | **12** | **0,071428571** | **0,9394166** | **0,00238854** | **Puerta del Sol** | **Bankia** |
| Saturday | 12 | 0,057971014 | 0,47500214 | 0,00216646 | Parque del Retiro | Estación de Madrid Puerta de Atocha |
| **Tuesday** | **12** | **0,016326531** | **0,44099188** | **0,00238854** | **Estación de Madrid Puerta de Atocha** | **Parque del Retiro** |
| Monday | 12 | 0,016021362 | 0,48111207 | 0,00235064 | Estación de Madrid Puerta de Atocha | Parque del Retiro |
| **Monday** | **12** | **0,016021362** | **0,73683831** | **0,00235064** | **Estación de Madrid Puerta de Atocha** | **Plaza Mayor** |
| Sunday | 12 | 0,015665796 | 0,38444883 | 0,00198774 | Estación de Madrid Puerta de Atocha | Parque del Retiro |

*Tabla 19:Resumen de rutas frecuentes de viajes en función de la frecuencia.*

Una de las aplicaciones más relevantes de este estudio es en la aplicación de motores de recomendaciones a partir de los desplazamientos de los habitantes de una ciudad. Además, otras de las aplicaciones de este estudio pueden ser: gestión de tráfico, optimización de servicios de transporte público y acciones de marketing enfocadas a locales comerciales que se encuentran emplazados en las ubicaciones que un individuo pueda visitar.

# BIBLIOGRAFÍA


[1] D. Chiu, R for Data Science Cookbook, Birmingham: Packt Publishing, 2016.

[2] G. Lansley y P. A. Longley, «The geography of Twitter topics in London,» *Computers, Environment and Urban Systems,* vol. 58, pp. 85-96, July 2016.

[3] F. Luo, G. Cao, K. Mulligan y X. Li, «Explore spatiotemporal and demographic characteristics of human mobility via Twitter: A case study of Chicago,» *Applied Geography,* vol. 70, pp. 11-25, May 2016.

[4] E. Necula, «Analyzing Traffic Patterns on Street Segments Based on GPS Data Using R,» *Transportation Research Procedia,* vol. 10, pp. 276-285, July 2015.

[5] C. Comito, D. Falcone y DomenicoTalia, «Mining human mobility patterns from social geo-tagged data,» *Pervasive and Mobile Computing,* vol. 33, pp. 91-107, December 2016.

[6] J. Cranshaw, R. Schwartz, J. I. Hong y N. Sadeh, «The Livehoods Project: Utilizing Social Media to Understand the Dynamics of a City,» The 6th International AAAI Conference on Weblogs and Social Media, Dublin, 2012.

[7] J. Pei, J. Han, B. Mortazavi-Asl y H. Pinto, «PrefixSpan: Mining Sequential Patterns Efficiently by Prefix-Projected Pattern Growth,» ICDE, Burnaby, 2001.

[8] G. Ritschard, R. Bürgin y M. Studer, «Exploratory Mining of Life Event Histories.,» *Contemporary Issues in Exploratory Data Mining in Behavioral Sciences,* 2013.

[9] J. Heaton, «Comparing Dataset Characteristics that Favor the Apriori, Eclat or FP-Growth Frequent Itemset Mining Algorithms,» IEEE, Ft. Lauderdale, 2016.

[10] A. Gabadinho, G. Ritschard, M. Studer y N. S. Müller, «Mining sequence data in R with the TraMineR package: A user's guide,» Switzerland, 2011.

[11] D. Chiu, Machine Learning with R Cookbook, Birmingham: Packt Publishing, 2015, pp. 321-345.