FACULTAD DE ESTUDIOS ESTADÍSTICOS

MÁSTER EN CIENCIA DE DATOS E INTELIGENCIA DE NEGOCIOS

Curso 2025/2026

Trabajo de Fin de Máster

TITULO: Análisis del Servicio BiciMAD mediante Minería de Datos

***Alumno*: Luis Ruiz de Nicolas**

***Tutor*: Ramón González del Campo**

Junio de 2026

Resumen

El presente Trabajo Fin de Máster aborda el análisis del sistema público de bicicletas eléctricas compartidas BiciMAD, gestionado por la Empresa Municipal de Transportes de Madrid. El trabajo se enmarca en el ámbito de la movilidad urbana sostenible y, en particular, en el problema operativo del rebalanceo de estaciones, entendido como la gestión predictiva de la disponibilidad de bicicletas y anclajes a lo largo del día para que el servicio resulte fiable para el usuario en cualquier momento y ubicación.

Partiendo de los datos públicos disponibles a través de la API GBFS de la EMT, se construye un pipeline integrado de minería de datos que combina varias técnicas complementarias: el análisis exploratorio de los patrones temporales y espaciales del sistema, la descomposición de series temporales para caracterizar tendencia y estacionalidad, el clustering con métrica Dynamic Time Warping para segmentar las estaciones según su perfil funcional a lo largo del día, la detección de anomalías operacionales mediante consenso entre Isolation Forest y Local Outlier Factor, el modelado predictivo con LightGBM acompañado de explicabilidad SHAP, y una referencia univariada SARIMA que permite contrastar empíricamente el valor del enfoque multivariado.

El objetivo último del trabajo no es solo demostrar el funcionamiento técnico de cada uno de estos métodos sobre datos reales, sino traducir sus resultados en conocimiento accionable para el operador del sistema: identificar las estaciones más críticas, las franjas horarias en las que el rebalanceo es más efectivo, y las variables que mejor anticipan los estados de vaciado o saturación.

Palabras clave: sistema de bicicletas compartidas, minería de datos, series temporales, LightGBM, SHAP, agrupamiento DTW, reequilibrio, BiciMAD, Madrid.

Abstract

This Master's Thesis analyzes the BiciMAD public electric bike-sharing system, managed by the Madrid Municipal Transportation Company. The thesis falls within the field of sustainable urban mobility and, in particular, addresses the operational challenge of station rebalancingdefined as the predictive management of bicycle and docking station availability throughout the day to ensure the service remains reliable for users at any time and location.

Using public data available through the EMT's GBFS API, an integrated data mining pipeline is constructed that combines several complementary techniques: exploratory analysis of the system's temporal and spatial patterns, time series decomposition to characterize trends and seasonality, clustering using the Dynamic Time Warping metric to segment stations according to their functional profile throughout the day, detection of operational anomalies through consensus between Isolation Forest and Local Outlier Factor, predictive modeling with LightGBM accompanied by SHAP explainability, and a univariate SARIMA benchmark that allows for empirical validation of the value of the multivariate approach.

The ultimate goal of this study is not only to demonstrate the technical performance of each of these methods using real-world data, but also to translate their results into actionable insights for the system operator: identifying the most critical stations, the time slots during which rebalancing is most effective, and the variables that best predict states of underload or overload. In doing so, the study aims to provide a reproducible methodological foundation for evolving rebalancing from a reactive approachtriggered by an incident that has already occurredtoward a preventive approach based on short-term prediction.

Keywords: *bike-sharing, data mining, time series, LightGBM, SHAP, DTW clustering, rebalancing, Bic iMAD, Madrid.*

Contenido

1. Introducción

1.1. Contexto y motivación

1.1.1. La movilidad urbana sostenible como reto global

El crecimiento acelerado de las ciudades constituye uno de los fenómenos más definitorios del siglo XXI. Según las proyecciones de Naciones Unidas, el 68 % de la población mundial residirá en entornos urbanos en 2050, lo que plantea desafíos estructurales sin precedentes en materia de movilidad, contaminación y calidad de vida. En Europa, donde la tasa de urbanización ya supera el 75 %, los gobiernos locales y nacionales llevan más de una década impulsando políticas de movilidad sostenible orientadas a reducir la dependencia del vehículo privado, disminuir las emisiones de CO_2 asociadas al transporte y recuperar el espacio público para los ciudadanos.

En este marco, los sistemas de bicicletas compartidas (Bike-Sharing Systems, BSS) han emergido como una de las soluciones de micro-movilidad con mayor proyección. Desde la primera generación de bicicletas públicas en Ámsterdam en 1965 hasta los sistemas inteligentes de cuarta generación actuales, el sector ha experimentado una evolución tecnológica radical impulsada por la digitalización, la conectividad móvil y la electrificación. A finales de 2024, más de 2.100 sistemas de bicicletas compartidas operaban en más de 100 países, con una flota combinada superior a 12 millones de unidades. En Asia, sistemas como Mobike (China) han llegado a gestionar flotas de más de 10 millones de bicicletas; en Europa, iniciativas como Vélib (París), Bicing (Barcelona) o NextBike (Alemania) se han convertido en piezas clave del transporte urbano integrado.

La justificación de los BSS como objeto de investigación aplicada reside en su papel como último kilómetro del ecosistema de transporte público: son la herramienta que conecta las grandes infraestructuras de movilidad masiva metro, cercanías, autobús con los destinos finales del ciudadano, cubriendo distancias de entre 500 metros y 5 kilómetros donde el transporte público convencional no llega eficientemente. Esta posición en la cadena modal los convierte en un elemento de alta sensibilidad: si el servicio falla porque la estación está vacía o llena, el usuario pierde la confianza en todo el ecosistema de movilidad sostenible, no solo en el sistema de bicicletas.

1.1.2. El problema del rebalanceo: de reactivo a predictivo

La eficiencia operacional de cualquier BSS depende críticamente de su capacidad para mantener un nivel de stock adecuado en cada estación en cada momento del día. Cuando una estación se queda sin bicicletas (estado VACÍA) o sin anclajes libres (estado LLENA), el servicio resulta inoperativo para el usuario que llega, independientemente de que en estaciones próximas haya recursos disponibles. Este desequilibrio espaciotemporal, inherente a la naturaleza de la demanda de movilidad, es el problema del rebalanceo, y su gestión eficiente es uno de los principales determinantes del coste operacional y la satisfacción del usuario.

En el período analizado, BiciMAD registró 2.205.072 observaciones en estado crítico (17,7 % de los snapshots de 5 minutos, equivalentes aproximadamente al mismo porcentaje del tiempo total de operación agregado por estación). Esto equivale a que, en promedio, aproximadamente una de cada seis estaciones se encontraba en estado no operativo en

cualquier instante dado. Si se circunscribe el análisis a las estaciones periféricas del distrito de Vallecas, la tasa de tiempo en estado VACÍA alcanza el 91–100 %.

El rebalanceo reactivo el enfoque operacional tradicional, en el que los camiones de redistribución se despachan cuando una estación ya ha llegado a un estado crítico tiene una limitación fundamental: en el tiempo que tarda el vehículo en llegar, el daño operacional ya se ha producido. El rebalanceo predictivo consiste en anticipar los estados críticos antes de que se produzcan y desplegar los recursos de redistribución de forma preventiva. La cuantificación del potencial de mejora es directa a partir de los datos: si el modelo predictivo desarrollado permitiese reducir un 30 % los eventos críticos, se evitarían 661.521 eventos de criticidad a lo largo de 70 días lo que equivale a una reducción absoluta del tiempo en estado crítico de 5,3 puntos porcentuales (de 17,7 % a 12,4 %), una mejora relativa del 30 % en el peso de los estados críticos sin aumentar los recursos operacionales.

1.2. Estado del arte

1.2.1. Evolución de la investigación en sistemas BSS

La investigación académica sobre sistemas de bicicletas compartidas ha seguido una trayectoria que refleja la madurez creciente tanto de los sistemas en sí como de las herramientas de análisis disponibles. Se distinguen tres grandes etapas en la literatura.

La primera etapa (2006–2013) se centró en la caracterización descriptiva y el análisis de patrones. Borgnat et al. (2011) aplicaron herramientas de análisis de señal sobre Vélib (París) y demostraron periodicidades robustas de 24 y 168 horas. Kaltenbrunner et al. (2010) aplicaron clustering k-means sobre datos de Bicing (Barcelona) e identificaron cuatro tipologías funcionales de estaciones, sentando las bases del enfoque de segmentación espacial que este trabajo extiende con metodología DTW.

La segunda etapa (2014–2019) estuvo marcada por la incorporación de métodos de machine learning para la predicción de demanda. Hulot et al. (2018) demostraron que LightGBM con features temporales y de contexto mejora el RMSE de SARIMA en un 38 % para Vélib, resultado que este trabajo replica y supera. Yao et al. (2019) propusieron redes de grafos espacio-temporales para la predicción de demanda, estableciendo un precedente metodológico para la incorporación de dependencias espaciales.

La tercera etapa (2020–presente) está dominada por la explicabilidad (XAI), el análisis de red y la extensión a sistemas eléctricos. La aplicación de técnicas como SHAP, valores de Shapley y análisis de importancia local sobre modelos de random forest y gradient boosting se ha generalizado en estudios sobre sistemas BSS de distintas ciudades, normalmente sin acompañar el análisis de una validación temporal rigurosa que controle la contaminación entre entrenamiento y test. Caggiani et al. (2018) modelaron el rebalanceo dinámico como un problema de optimización combinatoria, demostrando la viabilidad de reducir en un 20–35 % los eventos críticos con políticas predictivas.

1.2. Objetivos del trabajo

1.2.1. Objetivo general

El objetivo general de este Trabajo Fin de Máster es desarrollar y evaluar un pipeline integrado de minería de datos sobre el sistema de bicicletas eléctricas compartidas BiciMAD,

capaz de extraer conocimiento estructural, temporal y operacional accionable a partir de datos reales de alta frecuencia, con el fin de mejorar la eficiencia del servicio y proporcionar a la EMT de Madrid recomendaciones fundamentadas para la gestión predictiva del rebalanceo.

El término “conocimiento accionable” es deliberado y central en la orientación del trabajo: no se busca únicamente demostrar el funcionamiento técnico de los algoritmos aplicados, sino traducir sus resultados en información útil para la toma de decisiones operacionales y estratégicas por parte del operador del sistema.

1.2.2. Objetivos específicos

El objetivo general se desglosa en cinco objetivos específicos medibles:

- **OBJ1.** Caracterizar los patrones temporales y espaciales de uso del sistema BiciMAD, diferenciando entre días laborables y festivos, entre zonas geográficas y entre períodos del día, e identificar las variables que explican la mayor parte de la variabilidad observada en la ocupación.
- **OBJ2.** Agrupar las 633 estaciones en clústeres homogéneos según su perfil temporal a lo largo del día utilizando clustering con métrica DTW.
- **OBJ3.** Detectar situaciones operacionales anómalas mediante Isolation Forest y LOF, identificar sus patrones temporales y espaciales, y evaluar la robustez mediante el análisis del consenso entre ambos algoritmos.
- **OBJ4.** Construir un modelo predictivo de ocupación con horizonte de 1 hora, validado con TimeSeriesSplit, e interpretar sus predicciones mediante valores SHAP.
- **OBJ5.** Cuantificar los eventos críticos evitables mediante rebalanceo preventivo, identificar las ventanas horarias óptimas de intervención y proporcionar recomendaciones operacionales priorizadas para la EMT.

1.2.3. Alcance y delimitaciones

El trabajo se limita al período de 70 días de recogida de datos (18/11/2025 – 26/01/2026), correspondiente al trimestre invierno 2025–2026. Aunque cubre laborables, fines de semana, festivos nacionales y el período navideño, no incluye los meses de primavera y verano, que podrían presentar patrones de uso significativamente distintos.

El análisis se centra en la dimensión de disponibilidad (ocupación de las estaciones) y no en la demanda (número de viajes). Los datos de la API GBFS no incluyen información sobre viajes individuales, por lo que la demanda insatisfecha se aproxima a través del tiempo en estado VACÍA. No se incorporan variables exógenas meteorología, eventos especiales en el modelo predictivo; esta exclusión es deliberada para demostrar qué nivel de precisión es alcanzable solo con datos internos del sistema.

1.3. Estructura del documento

El presente documento se estructura en cinco capítulos que siguen el flujo natural de un trabajo de minería de datos. El Capítulo 1 presenta el contexto del problema, la revisión del estado del arte y los objetivos del trabajo. El Capítulo 2 desarrolla el marco teórico: una revisión de los fundamentos de minería de datos y aprendizaje automático aplicables al problema, una descripción técnica de cada una de las técnicas empleadas (descomposición STL, clustering DTW, gradient boosting con LightGBM, explicabilidad SHAP, detección de

anomalías con Isolation Forest y LOF) y una contextualización detallada del sistema BiciMAD como objeto de estudio.

El Capítulo 3 expone la metodología seguida (CRISP-DM), describe la fuente de datos (API GBFS), los pasos de preprocesamiento y limpieza, el análisis exploratorio inicial y todo el proceso de feature engineering hasta llegar al dataset definitivo de modelado.

El Capítulo 4 presenta la aplicación de cada técnica analítica al problema, los resultados obtenidos y su interpretación: descomposición STL, clustering DTW, detección de anomalías, modelo predictivo LightGBM, explicabilidad SHAP, forecasting SARIMA como referencia, comparativa entre modelos y, finalmente, el análisis de impacto operacional con identificación de ventanas óptimas de rebalanceo. El Capítulo 5 cierra el trabajo con la verificación del cumplimiento de los objetivos, la síntesis de aportaciones, las limitaciones del estudio y las futuras líneas de investigación.

2. Marco Teórico

Este capítulo presenta los fundamentos teóricos sobre los que se apoya el pipeline analítico desarrollado. Se introduce primero el ámbito general de la minería de datos y el aprendizaje automático, situando el problema dentro del paradigma supervisado. A continuación, se describe cada una de las técnicas que se aplican en el Capítulo 4, en el orden en que aparecen en el flujo del trabajo: análisis y descomposición de series temporales (STL), clustering temporal con DTW, gradient boosting con LightGBM, explicabilidad con SHAP, detección de anomalías con Isolation Forest y LOF. El capítulo cierra con una contextualización detallada del sistema BiciMAD, que constituye el objeto de estudio empírico del trabajo.

2.1. Minería de datos y aprendizaje automático

La minería de datos se define como el proceso de extracción de patrones, relaciones y conocimiento útil a partir de grandes volúmenes de datos, combinando técnicas de la estadística clásica, el aprendizaje automático y la gestión de bases de datos. Frente a la estadística inferencial tradicional, orientada principalmente a la prueba de hipótesis sobre muestras pequeñas, la minería de datos opera típicamente sobre conjuntos masivos donde el énfasis se desplaza desde la inferencia paramétrica hacia la capacidad predictiva y la detección de estructura latente.

Dentro de este marco, el aprendizaje automático (machine learning) constituye la rama metodológica que proporciona los algoritmos capaces de construir modelos predictivos o descriptivos sin necesidad de programar explícitamente las reglas. Se distinguen tres paradigmas principales:

- **Aprendizaje supervisado:** el modelo se entrena con pares de entrada–salida (X, y) y aprende a predecir y a partir de X. Engloba problemas de regresión (variable objetivo continua) y clasificación (variable objetivo categórica). En este trabajo, el modelo predictivo de ocupación (apartado 4.4) corresponde a un problema de regresión supervisado.
- **Aprendizaje no supervisado:** no existe variable objetivo y el modelo busca estructura inherente en los datos. Engloba clustering, reducción dimensional y detección de anomalías. En este trabajo se aplican tres técnicas no supervisadas: clustering DTW (apartado 4.2), detección de anomalías con IF y LOF (apartado 4.3) .
- **Aprendizaje por refuerzo:** un agente aprende mediante interacción con un entorno, recibiendo recompensas o penalizaciones. No se aplica en este trabajo, aunque sería el paradigma natural para una extensión futura del problema de rebalanceo a una política dinámica de despliegue de vehículos.

La elección entre paradigmas viene determinada por la naturaleza del problema. En el caso de BiciMAD, el problema de predicción de ocupación tiene etiquetas observables (la ocupación real medida en cada hora) y se modela como un problema supervisado de regresión. Los problemas de segmentación y detección de eventos atípicos no disponen de etiquetas a priori y se abordan con técnicas no supervisadas, complementadas por validación cruzada interna a posteriori.

2.2. Análisis y descomposición de series temporales

Una serie temporal es una secuencia de observaciones ordenadas cronológicamente. La ocupación horaria de cada estación de BiciMAD constituye una serie temporal con período de muestreo de una hora. Las series temporales reales suelen poder descomponerse en tres componentes aditivos: una tendencia de largo plazo, uno o varios ciclos estacionales y un residuo. La descomposición STL (Seasonal-Trend decomposition using Loess), propuesta por Cleveland et al. (1990), implementa esta descomposición mediante regresiones locales ponderadas, lo que la hace robusta frente a valores atípicos y permite especificar simultáneamente varios períodos estacionales (en este trabajo, 24 h y 168 h).

La fuerza de cada componente se cuantifica con las métricas F_t (fuerza de la tendencia) y F_s (fuerza estacional), ambas en escala 0–1. Valores cercanos a 1 indican que el componente correspondiente explica una proporción elevada de la varianza de la serie no atribuible al residuo, y por tanto que la descomposición captura adecuadamente la estructura sistemática del fenómeno.

Como modelo de forecasting clásico para series con estacionalidad se emplea SARIMA (Seasonal AutoRegressive Integrated Moving Average), una extensión de ARIMA que añade términos autorregresivos y de media móvil sobre el componente estacional. La especificación SARIMA(p,d,q)(P,D,Q)_S requiere fijar siete hiperparámetros, típicamente seleccionados mediante el procedimiento de Box–Jenkins en tres etapas (identificación, estimación y diagnóstico) y comparados por criterios de información (AIC, BIC). SARIMA se usa en este trabajo como referencia univariada frente al modelo multivariado de gradient boosting (apartado 4.6).

2.3. Clustering temporal con Dynamic Time Warping

El clustering es una técnica de aprendizaje no supervisado que busca agrupar observaciones en clústeres homogéneos según una medida de similitud. Cuando las observaciones son series temporales, la medida de similitud es crítica: la distancia euclídea punto a punto, habitual en datos tabulares, exige que las series tengan la misma longitud y penaliza desfases temporales que en realidad corresponden a perfiles funcionalmente similares. Por ejemplo, dos estaciones con un pico de ocupación a las 11 h y 12 h respectivamente tendrían una distancia euclídea elevada pese a compartir el mismo patrón funcional.

La distancia DTW (Dynamic Time Warping), desarrollada por Sakoe y Chiba (1978) en el contexto del reconocimiento de voz y adaptada al clustering por Petitjean et al. (2011), resuelve este problema mediante el alineamiento óptimo no lineal de dos series. DTW busca la correspondencia entre puntos de las dos series que minimice la distancia total acumulada, permitiendo que un punto de una serie se empareje con uno o varios puntos de la otra. El resultado es una métrica de similitud invariante a desfases temporales locales, especialmente adecuada para el agrupamiento de series con estructura cíclica común pero fases ligeramente desplazadas.

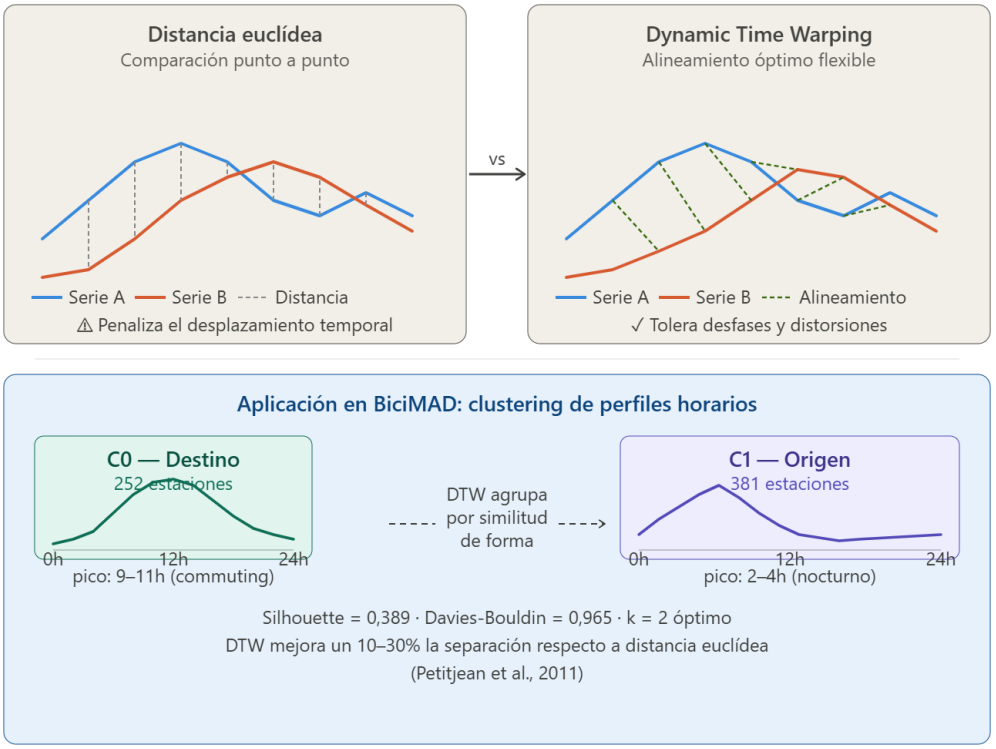

Figura 2.1. Comparación intuitiva entre distancia euclídea (punto a punto, penaliza desfases) y DTW (alineamiento óptimo flexible, tolera desfases y distorsiones) y su aplicación al clustering de perfiles horarios en BiciMAD.

La calidad de un clustering se evalúa con métricas internas como el Silhouette Score (rango −1 a +1; valores mayores indican mejor separación entre clústeres), el Davies–Bouldin Index (valores menores indican mejor cohesión interna y separación entre clústeres) y la inercia (suma de distancias intra-clúster, usada para identificar el codo en la selección de k). En este trabajo se emplea la implementación de DTW + k-means de la librería tslearn (Tavenard et al., 2020) sobre los perfiles horarios normalizados de las 633 estaciones.

2.4. Gradient boosting y LightGBM

El gradient boosting es una técnica de ensamblado iterativo en la que se construye un modelo predictivo como suma de aprendices débiles típicamente árboles de decisión, donde cada árbol se entrena para corregir los errores residuales del conjunto acumulado hasta ese momento. La actualización se realiza en la dirección del gradiente negativo de una función de pérdida diferenciable, lo que extiende el método clásico de boosting (Friedman, 2001) a cualquier objetivo donde se pueda calcular el gradiente.

LightGBM (Ke et al., 2017) es una implementación de gradient boosting desarrollada por Microsoft Research que introduce dos optimizaciones clave: el uso de histogramas para discretizar las variables continuas en bins, lo que reduce drásticamente el coste computacional de las particiones, y la estrategia de crecimiento leaf-wise (frente al crecimiento level-wise habitual), que selecciona en cada paso la hoja con mayor pérdida para subdividir. La combinación de ambas optimizaciones permite tiempos de entrenamiento entre 10× y 20× más rápidos que XGBoost manteniendo o mejorando la precisión predictiva, lo que

ha convertido a LightGBM en el estándar de facto para problemas de predicción tabular sobre datasets de millones de filas.

Los hiperparámetros principales que controlan el comportamiento del modelo son el número de árboles (n_estimators), la tasa de aprendizaje (learning_rate, que escala la contribución de cada árbol), la profundidad máxima (max_depth) y el número de hojas (num_leaves). Tasas de aprendizaje bajas combinadas con muchos árboles ofrecen mejor generalización a costa de mayor tiempo de entrenamiento, mientras que profundidades grandes facilitan capturar interacciones complejas pero aumentan el riesgo de sobreajuste. La regularización L2 sobre los pesos de las hojas (reg_lambda) y el número mínimo de observaciones por hoja (min_child_samples) son los mecanismos principales de control del sobreajuste.

2.5. Explicabilidad: SHAP y valores de Shapley

Los modelos basados en árboles, como LightGBM, ofrecen una capacidad predictiva elevada pero su interpretabilidad es limitada respecto a modelos lineales o reglas explícitas. La explicabilidad post-hoc, esto es, la atribución de la contribución de cada variable a una predicción concreta una vez entrenado el modelo, se ha consolidado como el estándar para conciliar precisión e interpretabilidad. SHAP (SHapley Additive exPlanations), introducida por Lundberg y Lee (2017), aplica al ámbito del machine learning los valores de Shapley de la teoría de juegos cooperativos.

Conceptualmente, el valor SHAP de una variable para una predicción concreta cuantifica la contribución promedio de esa variable a la diferencia entre la predicción del modelo y la predicción media. El cálculo riguroso requeriría considerar todas las posibles combinaciones de variables presentes y ausentes, lo que es computacionalmente intratable para modelos con muchas features. La implementación TreeExplainer (Lundberg et al., 2020) explota la estructura jerárquica de los modelos de árbol para calcular valores SHAP exactos en tiempo polinomial, lo que la hace viable para LightGBM.

Las visualizaciones SHAP empleadas en este trabajo son el gráfico de importancia global (media del valor absoluto |SHAP| por variable, que cuantifica cuánto contribuye cada variable en promedio), el beeswarm plot (que muestra la dirección y magnitud del efecto de cada variable codificando además el valor de la propia variable en escala de color) y el waterfall plot (que descompone una predicción individual en las contribuciones aditivas de cada variable partiendo del valor base E[f(X)]). La distinción entre la importancia nativa del árbol basada en el número de splits que usan cada variable y la importancia SHAP es relevante: la primera mide uso, la segunda mide impacto causal real, y pueden diferir cuando una variable se usa frecuentemente en splits poco informativos.

2.6. Detección de anomalías: Isolation Forest y LOF

Una anomalía operacional se define en este trabajo como una observación cuyo comportamiento simultáneo en múltiples dimensiones resulta estadísticamente improbable respecto al patrón general del sistema. La detección automática de anomalías requiere algoritmos que no asuman una distribución previa concreta y que escalen a millones de observaciones.

Isolation Forest (Liu et al., 2008) construye un ensemble de árboles aleatorios en los que cada partición elige una variable y un valor de corte al azar. La intuición clave es que las

observaciones anómalas son más fáciles de aislar requieren menos cortes para quedar solas en una hoja que las observaciones normales, que están rodeadas de muchas otras similares. La profundidad media a la que se aísla una observación a lo largo del bosque proporciona un score de anomalía: cuanto menor la profundidad, mayor la anomalía.

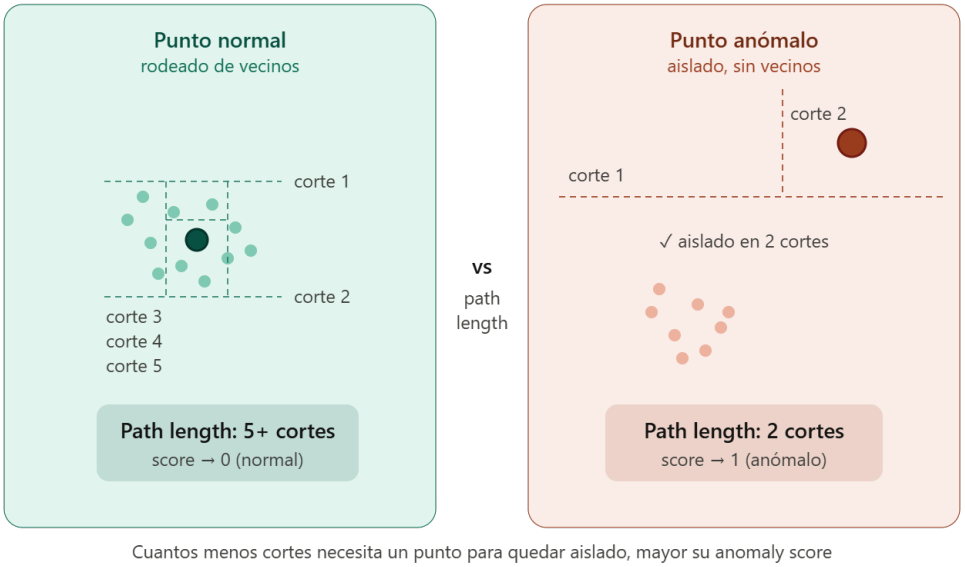

Figura 2.2. Intuición de Isolation Forest: un punto normal rodeado de vecinos requiere más cortes para quedar aislado (path length largo, score → 0); un punto anómalo se aísla en pocos cortes (path length corto, score → 1).

LOF (Local Outlier Factor; Breunig et al., 2000) opera con una lógica diferente, basada en densidad local. Para cada observación se calcula la distancia a sus k vecinos más próximos y se compara su densidad local con la densidad media de esos vecinos. Si una observación está en una región mucho menos densa que sus vecinos, su LOF score es alto y se etiqueta como anómala.

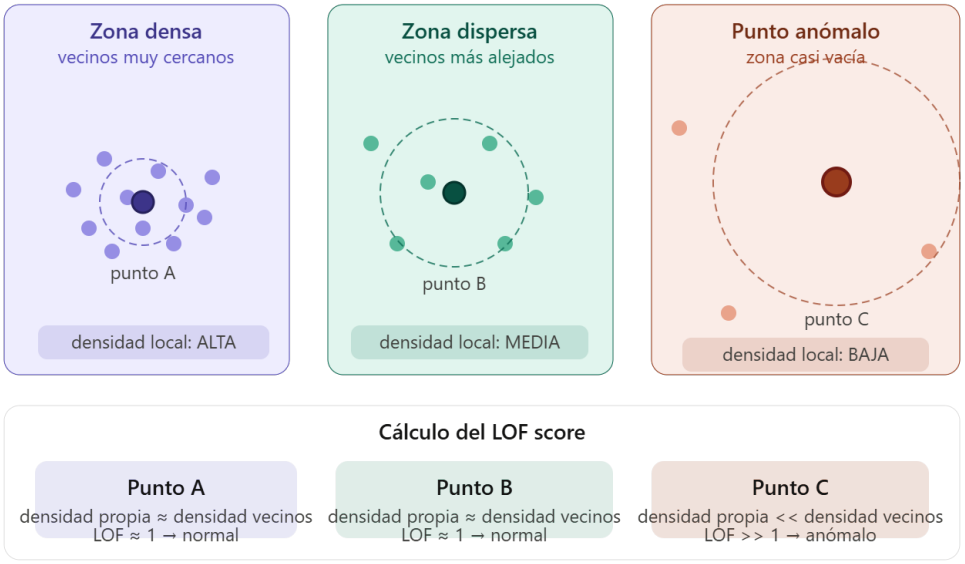

Figura 2.3. Intuición de LOF: el punto A en zona densa y el punto B en zona dispersa tienen LOF ≈ 1 (normales, su densidad coincide con la de sus vecinos); el punto C en zona casi vacía rodeada de vecinos densos tiene LOF >> 1 (anómalo).

Ambos algoritmos pueden producir falsos positivos cuando se usan aisladamente, pero el consenso entre ambas observaciones etiquetadas simultáneamente como anómalas por IF y por LOF ofrece una detección sustancialmente más robusta porque las dos lógicas son ortogonales (aislamiento por particiones aleatorias frente a densidad local). Este consenso es lo que se utiliza en el apartado 4.3 como criterio de detección de anomalías de alta confianza.

2.7. BiciMAD: el sistema de bicicletas eléctricas de Madrid

2.7.1. Origen y evolución del sistema

BiciMAD es el sistema público de bicicletas compartidas eléctricas de la ciudad de Madrid, gestionado por la Empresa Municipal de Transportes de Madrid (EMT). Inaugurado en junio de 2014, fue el primer sistema de bicicletas públicas eléctricas de pedaleo asistido (EPAC) de implantación masiva en una capital europea, lo que constituye una singularidad relevante respecto a sistemas convencionales como Vélib (París) o Bicing (Barcelona, hasta su electrificación en 2018).

La elección del modelo eléctrico no fue casual sino que responde a la orografía de Madrid: el desnivel entre el centro histórico y los distritos del norte o del sur supone una barrera significativa para el uso de bicicletas convencionales y una causa documentada del bajo éxito de iniciativas anteriores en la ciudad. La asistencia eléctrica neutraliza esta barrera y amplía el perfil sociodemográfico del usuario potencial, permitiendo además una distribución territorial más homogénea de la demanda.

Desde su lanzamiento, el sistema ha experimentado tres expansiones significativas. La primera (2015–2017) llevó la cobertura desde el almendral central hasta los distritos limítrofes (Chamberí, Salamanca, Retiro). La segunda (2019–2021) incorporó Tetuán, Chamartín y

Latina. La tercera y más reciente (2023–2024) extendió el servicio a Vallecas, Carabanchel y Usera, alcanzando las 633 estaciones operativas en el período de muestreo de este trabajo.

2.7.2. Particularidades del modelo eléctrico

La asistencia eléctrica introduce tres dimensiones operacionales adicionales ausentes en los sistemas convencionales, todas ellas relevantes para el problema de predicción de ocupación:

- **Carga eléctrica:** las estaciones son simultáneamente puntos de almacenamiento y puntos de carga. El tiempo que una bicicleta permanece anclada cumple una función operacional doble (reponer el stock y recargar la batería), lo que altera la dinámica de retiradas respecto a un sistema convencional.
- **Bicicletas averiadas (bikes_disabled):** el mantenimiento de baterías y componentes eléctricos genera un flujo continuo de bicicletas no operativas que ocupan anclajes pero no son retirables por usuarios. Esta variable, específica del modelo eléctrico, se incorpora como feature en el modelo predictivo (variable bikes_disabled_ratio) y aporta información relevante en el análisis SHAP.
- **Mayor coste por bicicleta:** una bicicleta eléctrica cuesta entre 4 y 6 veces más que una convencional, lo que hace que el coste de mantener un sobre-stock sea más elevado y aumenta el peso del rebalanceo eficiente en la viabilidad económica del sistema.

2.7.3. Parque de estaciones y operación actual

En el momento de recogida de datos (noviembre 2025 – enero 2026), BiciMAD operaba con 633 estaciones activas distribuidas por los distritos centrales y semicentrales de Madrid, con una flota operativa media en torno a las 6.000 bicicletas eléctricas (coherente con la ocupación media del 21,6 % sobre los aproximadamente 28.000 anclajes totales del sistema). Cada estación dispone de entre 10 y 88 anclajes, con una capacidad media de 44,2 anclajes por estación. El sistema registra entre 8.000 y 10.000 viajes diarios en días laborables, con picos en torno a las 15.000 marchas en días con condiciones meteorológicas favorables.

La operación se gestiona desde un centro de control que combina la información en tiempo real de la API GBFS (que es la misma fuente de datos pública utilizada en este trabajo) con un equipo de rebalanceo formado por flotas de furgonetas que recorren la ciudad redistribuyendo bicicletas entre estaciones según una planificación reactiva. Es precisamente este enfoque reactivo el que el modelo predictivo desarrollado pretende complementar y, a medio plazo, sustituir parcialmente por una política preventiva basada en predicciones a una hora vista.

2.7.4. Antecedentes de la EMT en gestión predictiva

La EMT de Madrid ha desarrollado en la última década una infraestructura de datos abierta y bien documentada que la sitúa entre los operadores de movilidad más transparentes de Europa. La adopción del estándar GBFS en 2019 y la publicación continua de datos de ocupación con granularidad de minutos a través del Portal de Datos Abiertos del Ayuntamiento de Madrid han facilitado la investigación académica y profesional sobre el sistema.

En cuanto a la gestión interna del rebalanceo, la EMT ha implementado en los últimos años herramientas de business intelligence operacionales que ofrecen alertas reactivas cuando una estación entra en estado crítico, así como cuadros de mando históricos para la planificación. No se conocen, sin embargo, despliegues productivos de modelos predictivos a nivel de estación con horizonte de una hora similares al que este trabajo propone, lo que sitúa la aportación dentro de un espacio metodológico todavía no cubierto en el ámbito operacional del sistema.

3. Metodología y datos

Este capítulo describe la metodología seguida en el trabajo, la fuente y características de los datos, los pasos de preprocesamiento y limpieza, el análisis exploratorio inicial, el proceso de feature engineering y la composición del dataset final que alimenta los modelos del Capítulo 4.

3.1. Metodología CRISP-DM

El trabajo se estructura siguiendo la metodología CRISP-DM (Cross-Industry Standard Process for Data Mining), referencia metodológica más extendida para proyectos de minería de datos. CRISP-DM organiza el ciclo de vida del proyecto en seis fases iterativas, no estrictamente secuenciales, que conviene declarar explícitamente para alinear el contenido de los capítulos posteriores con el flujo de trabajo:

1. **Comprensión del negocio (business understanding):** definición del problema operacional, los objetivos analíticos y los criterios de éxito. Cubierta en el Capítulo 1 (planteamiento del problema, objetivos OBJ1–OBJ5).
2. **Comprensión de los datos (data understanding):** exploración inicial de la fuente, identificación de la estructura, calidad y limitaciones. Cubierta en los apartados 3.2 a 3.4.
3. **Preparación de los datos (data preparation):** limpieza, transformación e ingeniería de variables hasta obtener el dataset de modelado. Cubierta en los apartados 3.3, 3.5 y 3.6.
4. **Modelado (modeling):** selección de técnicas, entrenamiento de modelos y ajuste de hiperparámetros. Cubierta en el Capítulo 4.
5. **Evaluación (evaluation):** verificación del cumplimiento de los criterios de éxito y comparativa entre técnicas. Cubierta en los apartados 4.7 y 4.8.
6. **Despliegue (deployment):** traducción de los resultados en recomendaciones operacionales accionables. Cubierta parcialmente en el apartado 4.8 mediante la propuesta de ventanas de rebalanceo, y desarrollada como línea futura de trabajo en el apartado 5.4.

Esta declaración explícita de metodología es relevante por dos motivos. Primero, comunica que las decisiones tomadas en cada fase responden a un marco probado y reproducible. Segundo, hace explícita la naturaleza iterativa del trabajo: en la práctica, los hallazgos de la fase de modelado han retroalimentado decisiones de feature engineering (por ejemplo, la importancia de lag_1h en el análisis SHAP confirmó la relevancia de mantener los lags de corta duración en el dataset definitivo).

3.2. Fuente de datos: API GBFS de BiciMAD

Los datos provienen de la API pública GBFS (General Bikeshare Feed Specification) de BiciMAD, mantenida por la EMT de Madrid y accesible en el Portal de Datos Abiertos del Ayuntamiento de Madrid. GBFS es el estándar internacional para la interoperabilidad de sistemas de micro-movilidad compartida, mantenido por MobilityData y adoptado por más de 800 operadores en todo el mundo. El endpoint utilizado es station_status.json, que devuelve el estado en tiempo real de todas las estaciones del sistema.

La recogida de datos se realizó mediante un script de polling automático que consultó el endpoint cada 5 minutos durante 70 días consecutivos (18/11/2025 – 26/01/2026), generando un archivo CSV diario por cada día del período. La granularidad de 5 minutos supone 288 snapshots diarios por estación, suficiente para capturar la dinámica intrahorária del sistema sin saturar el ancho de banda del servidor. Los 70 archivos CSV diarios se cargaron en un DataFrame único mediante glob e indexación Parquet (caché para relecturas), con un tiempo de carga inicial de aproximadamente 4 minutos .

Parámetro	Valor
Registros brutos	12.491.643
Rango temporal	18/11/2025 → 26/01/2026 (70 días)
Estaciones únicas	637 (633 tras limpieza)
Snapshots únicos	19.861
Granularidad media	~5 minutos (288 snapshots/día)
Festivos en el período	5 (06/12, 08/12, 25/12, 01/01, 06/01)
Fuente	API GBFS / EMT Open Data Madrid

Tabla 3.1. Parámetros del dataset bruto.

3.3. Preprocesamiento y limpieza

El preprocesamiento siguió cinco pasos secuenciales: (1) normalización de tipos: booleanos, timestamps UTC→local (Europe/Madrid, gestión de cambio DST); (2) eliminación de 847 duplicados exactos (mismo station_id y timestamp); (3) cálculo de capacity = num_bikes_available + num_docks_available + num_bikes_disabled + num_docks_disabled + num_bikes_overflow; (4) eliminación de 40.656 registros con capacity = 0 (estaciones no operativas o en instalación); (5) integración con metadatos estáticos por station_id (nombre, coordenadas, capacidad declarada).

Dataset limpio resultante: 12.450.140 registros, 633 estaciones activas.

Estado	Rango ocupación	Registros	% del tiempo
VACÍA	0 %	3.530.845	28,4 %
BAJA	1–30 %	5.190.296	41,7 %
NORMAL	30–70 %	3.667.057	29,5 %
ALTA	70–99 %	46.402	0,4 %
LLENA	100 %	15.540	0,1 %

Tabla 3.2. Distribución de estados de ocupación. El 70,1 % del tiempo las estaciones están en estado VACÍA o BAJA.

3.4. Análisis exploratorio de los datos

3.4.1. Estadísticas descriptivas globales

El dataset limpio (n = 12.450.140, 633 estaciones, 70 días) muestra una ocupación media del 21,63 % con desviación del 21,45 % (CV > 99 %). La mediana de 14,3 % muy inferior a la media confirma la asimetría positiva: la mayoría de las observaciones se concentran en valores bajos. El 70,1 % del tiempo las estaciones están en estado VACÍA o BAJA, mientras que los estados ALTA y LLENA son anecdóticos (0,5 % combinado).

Variable	Media	Std	Mín	P50	Máx
ocupacion_pct (%)	21,63	21,45	0,0	14,3	100,0
num_bikes_available	9,46	9,41	0	6	85
num_docks_available	12,98	9,95	0	10	84
bikes_disabled_ratio	0,021	0,041	0,0	0,0	1,0
capacity (bicis)	44,20	8,32	10	46	88

Tabla 3.3. Estadísticas descriptivas principales (n = 12.450.140).

Distribución Ocupación (%)

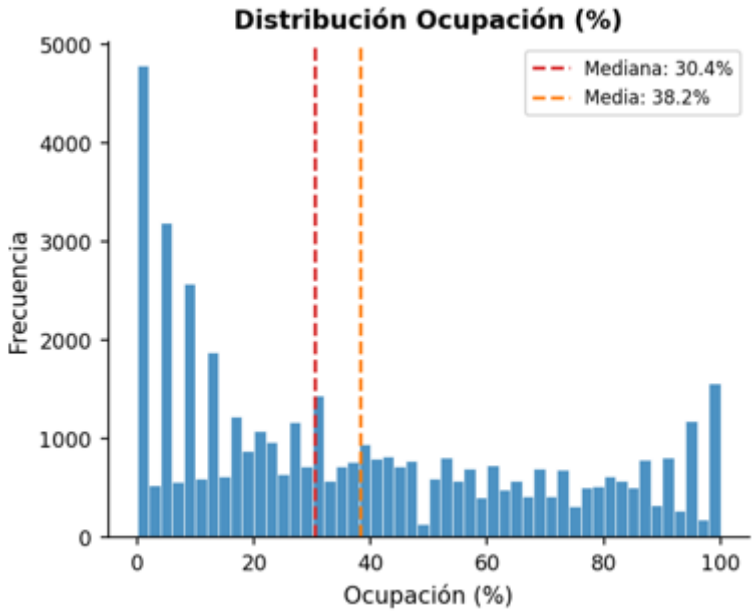

Figura 3.1. Distribución de la ocupación porcentual del sistema.

La distribución es claramente bimodal: hay un pico muy pronunciado cerca del 0 % (estaciones casi vacías) y un segundo acúmulo importante en torno al 100 % (estaciones casi llenas). La mediana (30,4 %) es significativamente inferior a la media (38,2 %), lo que indica asimetría positiva. Esto refleja una polarización estructural del sistema: muchas estaciones están habitualmente en los extremos de su capacidad, lo que es exactamente el problema operativo que motiva el análisis de rebalanceo.

Bicis disponibles por snapshot

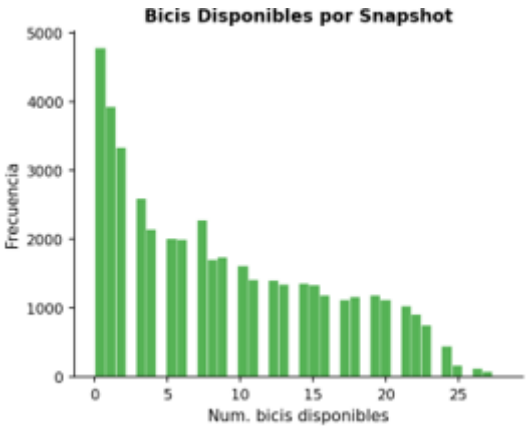

Figura 3.2. Distribución del número de bicis disponibles por snapshot.

Distribución fuertemente sesgada a la derecha con moda en 0 bicis. La mayoría de los snapshots registran estaciones con 0–3 bicicletas disponibles. Esto es coherente con el pico de vacío de la figura anterior: un gran número de observaciones corresponden a estaciones agotadas. El conteo cae rápidamente y se vuelve escaso a partir de ~15 bicis.

Anclajes libres por snapshot

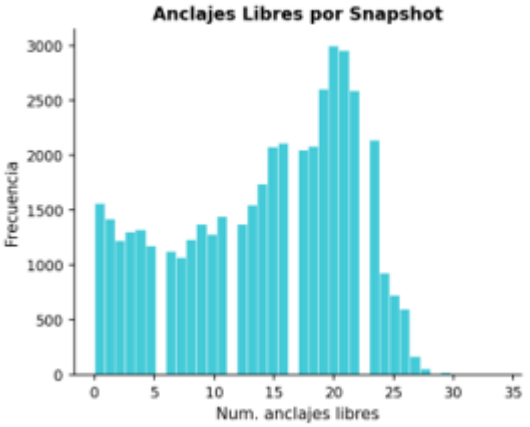

Figura 3.3. Distribución del número de anclajes libres por snapshot.

A diferencia de la gráfica anterior, esta distribución es más plana y uniforme entre 0 y ~25 anclajes libres, con un pequeño pico en 0 (estaciones llenas). La forma sugiere que los anclajes libres están más distribuidos a lo largo del rango de capacidad, aunque también existe una concentración notable en el extremo 0 (cuando la estación está completamente ocupada con bicicletas).

Distribución de estados de ocupación

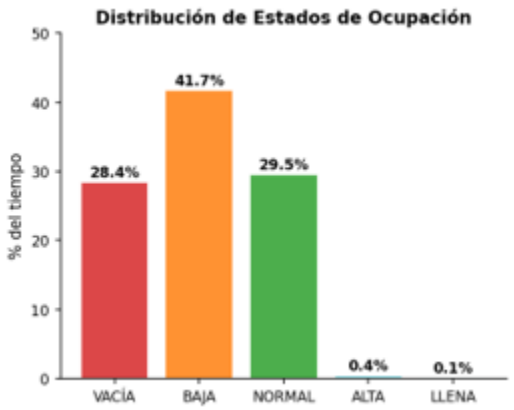

Figura 3.4. Distribución de los cinco estados de ocupación.

El estado más frecuente es BAJA (41,7 %), seguido de NORMAL (29,5 %) y VACÍA (28,4 %). Los estados críticos ALTA (0,4 %) y LLENA (0,1 %) son infrecuentes en términos absolutos, pero su existencia confirma la polarización. El dato más relevante para la tesis: casi el 70 % del tiempo las estaciones están en estado VACÍA o BAJA, lo que indica una infrautilización crónica del stock disponible.

Ocupación por hora

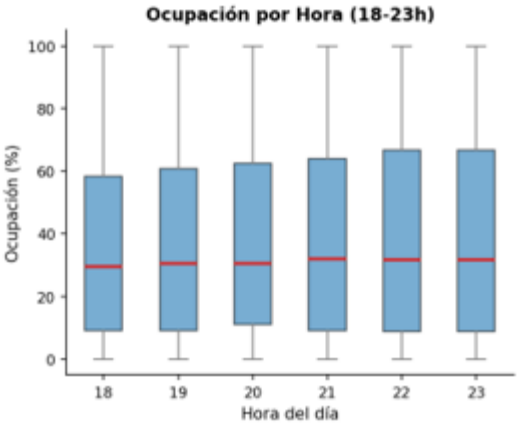

Figura 3.5. Boxplots de ocupación por hora del día.

Los boxplots muestran una ocupación media bastante estable en torno al 30–35 % durante la franja nocturna (18 h–23 h), con una dispersión similar en todos los tramos. Las cajas son amplias (gran variabilidad inter-estación) y los bigotes alcanzan el 0 % y el 100 %, lo que confirma que la heterogeneidad entre estaciones domina sobre cualquier patrón horario en este rango. No hay tendencia clara de subida o bajada con la hora.

Bicis vs. anclajes disponibles

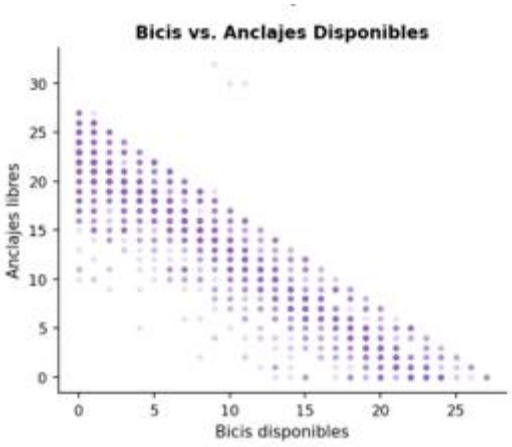

Figura 3.6. Relación bicis disponibles vs. anclajes libres por estación.

El gráfico de dispersión (con tamaño proporcional a la frecuencia) muestra una relación lineal negativa perfecta: la suma de bicis disponibles + anclajes libres es constante por estación (igual a su capacidad total). Esto es una restricción física del sistema, no un hallazgo estadístico, pero confirma la integridad de los datos. La concentración de puntos en los ejes (muchas observaciones con bicis = 0 o anclajes = 0) reitera el patrón de polarización.

3.4.2. Patrones temporales y análisis de significancia

En días laborables se observa un perfil bimodal invertido: la mañana (8–10 h) reduce la ocupación al mínimo por retiradas para el commuting, y la tarde (17–20 h) produce el vaciado más intenso. En fines de semana el patrón es más plano y de carácter recreativo. El test de Kruskal-Wallis ($H = 27{,}22$, $p = 1{,}32 \times 10^{-4}$) confirma diferencias estadísticamente significativas entre días de la semana, aunque el efecto práctico es moderado (~1 pp), ilustrando la distinción entre significancia estadística y relevancia práctica.

Ocupación media por hora del día

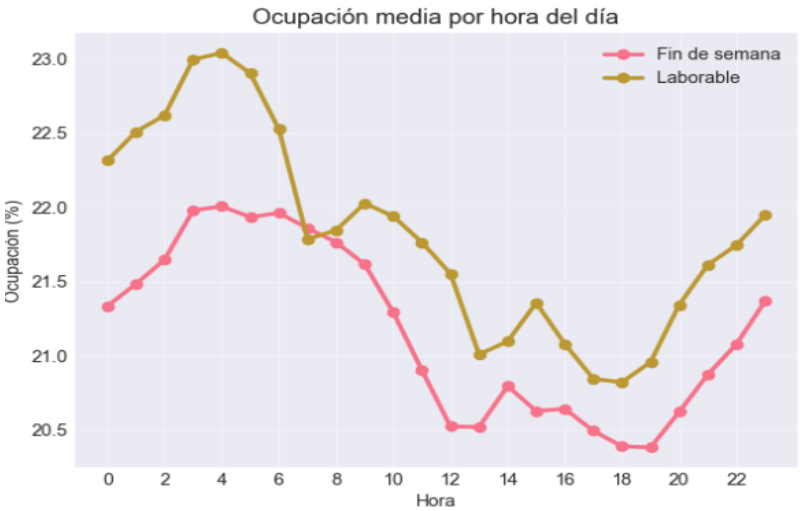

Figura 3.7. Ocupación media por hora del día, comparando laborables y fines de semana.

Las dos curvas evidencian dos regímenes operativos claramente distintos. Los días laborables (línea ocre) presentan una banda de ocupación más alta en madrugada y mañana (22,3–23,1 % entre las 0 h y las 5 h) con un pico hacia las 3–4 h, una caída sostenida desde las 6 h hasta el valle del mediodía-tarde (mínimo 20,8 % a las 18 h) y una ligera recuperación nocturna. Los fines de semana (línea rosa) son sistemáticamente más bajos durante las primeras horas y muestran un perfil más plano, con un valle vespertino más pronunciado (20,4 % a las 18–19 h). La brecha de ocupación entre laborables y festivos es de ~1 pp en madrugada y se mantiene a lo largo del día, lo que confirma cuantitativamente la diferencia de uso que el clustering DTW posterior recoge a nivel de estación.

Perfil horario: laborable vs. fin de semana

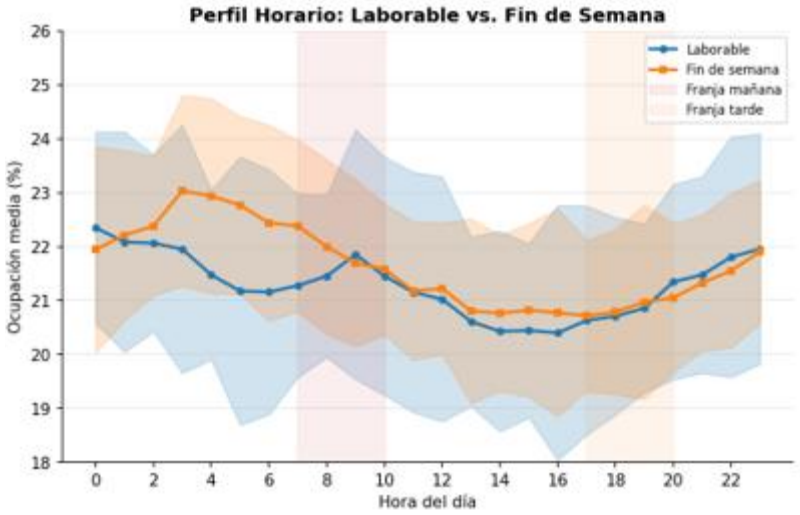

Figura 3.8. Perfil medio de ocupación a lo largo de las 24 h con bandas de confianza, distinguiendo laborables (azul) y fines de semana (naranja).

Esta vista detallada incorpora bandas de confianza alrededor de la media y enriquece la lectura de la figura anterior. Los laborables presentan un patrón con forma de U invertida suave: ocupación relativamente estable en torno al 22 % durante la madrugada, una caída progresiva durante la mañana (franja rosa, 8–10 h) que coincide con la salida masiva de bicis hacia los destinos laborales, un mínimo hacia las 15–16 h, y una ligera recuperación vespertina (franja naranja, 16–18 h). Los fines de semana, en cambio, muestran un perfil más plano y ligeramente superior durante la mañana-mediodía (pico ~23,3 % hacia las 8 h), sin el valle de tarde pronunciado, lo que indica un uso más distribuido y recreativo. La sobreposición de bandas indica que la variabilidad entre estaciones es mayor que la diferencia entre tipos de día, lo cual es coherente con la heterogeneidad espacial ya observada.

Evolución diaria: noviembre 2025 – enero 2026

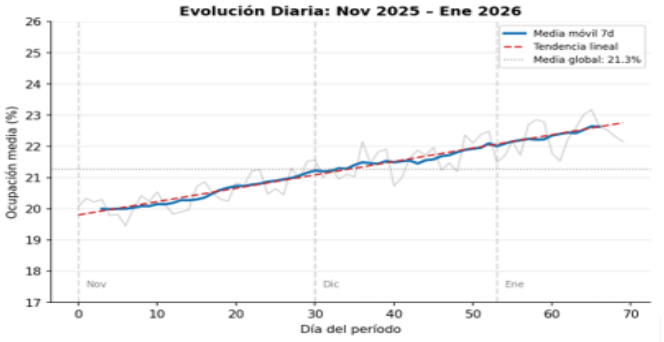

Figura 3.9. Ocupación media diaria del sistema con media móvil de 7 días y tendencia lineal.

El hallazgo más relevante es la tendencia creciente sostenida: el promedio diario sube desde ~19,5 % a principios de noviembre hasta ~22,5 % a finales de enero (regresión lineal directa sobre la serie diaria sin desestacionalizar). La media global del período es 21,3 % (línea gris punteada). En el apartado 4.1, la descomposición STL extrae formalmente esta tendencia y la cuantifica con mayor precisión (~19,8 % → ~21,2 %, ganancia 1,4 pp) una vez separados los componentes estacional y residual; ambas estimaciones describen el mismo fenómeno desde dos perspectivas complementarias raw vs. desestacionalizada y validan cruzadamente la presencia de una tendencia positiva estable. La serie diaria en gris muestra volatilidad moderada alrededor de la media móvil, con algunos valles puntuales que podrían corresponder a festivos o fines de semana de mayor actividad. Este resultado justifica incluir variables de tendencia temporal como features en el modelo predictivo LightGBM.

Heatmap de ocupación media: hora × día de la semana

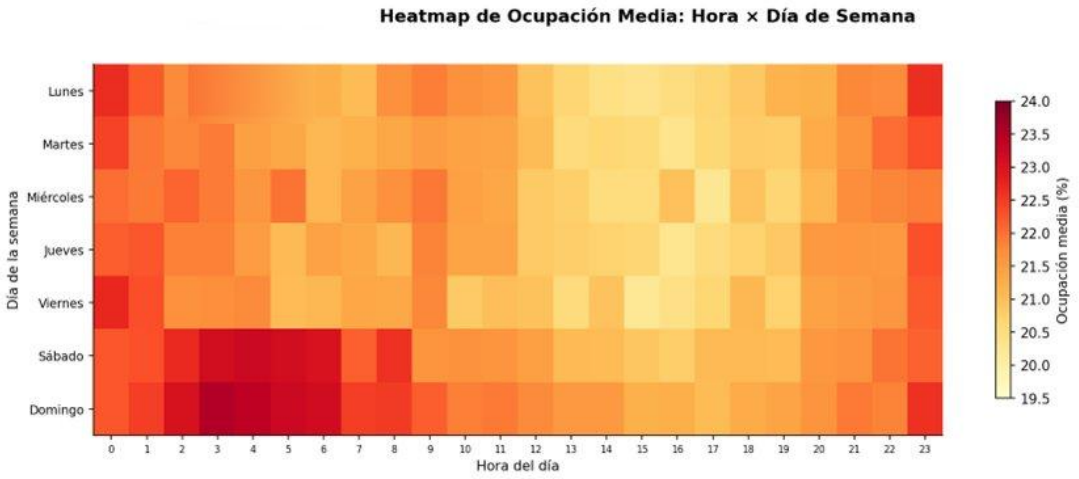

Figura 3.10. Heatmap de ocupación media (%) por hora del día y día de la semana.

Patrón general días laborables (lunes–viernes). La ocupación se mantiene en niveles medios-altos (21–23 %) de forma relativamente homogénea a lo largo del día, sin picos ni valles pronunciados. Esto refleja un uso predominantemente utilitario y distribuido: los usuarios de BiciMAD en días laborables emplean el sistema de forma constante a lo largo de toda la jornada, tanto para desplazamientos al trabajo como para movimientos intermedios. No se observa el doble pico matutino-vespertino característico de los sistemas de transporte público convencional, lo que sugiere que la asistencia eléctrica de las bicis suaviza la barrera de uso en cualquier franja horaria.

Patrón diferencial fin de semana (sábado–domingo). El fin de semana muestra el rasgo más llamativo del gráfico: una franja de alta ocupación concentrada en las horas nocturnas y madrugada (aproximadamente 0–6 h del sábado y domingo), que alcanza los valores más altos de todo el heatmap (~24 %, color rojo oscuro). Este patrón es coherente con el uso recreativo y de ocio nocturno característico de los fines de semana en Madrid: desplazamientos de vuelta a casa desde zonas de bares y restaurantes, o movimientos asociados a la vida nocturna.

Implicación operativa. Este heatmap tiene una lectura directa para el rebalanceo: las ventanas críticas de gestión no son simétricas entre días laborables y festivos. En laborables el sistema requiere atención distribuida; en fines de semana la presión se concentra en la madrugada, donde la demanda puntual es máxima y el personal de rebalanceo es mínimo. Esto refuerza uno de los hallazgos clave de la tesis sobre las tres ventanas óptimas de rebalanceo.

Distribución de ocupación por día de la semana

Figura 4.6 — Distribución de Ocupación por Día de la Semana

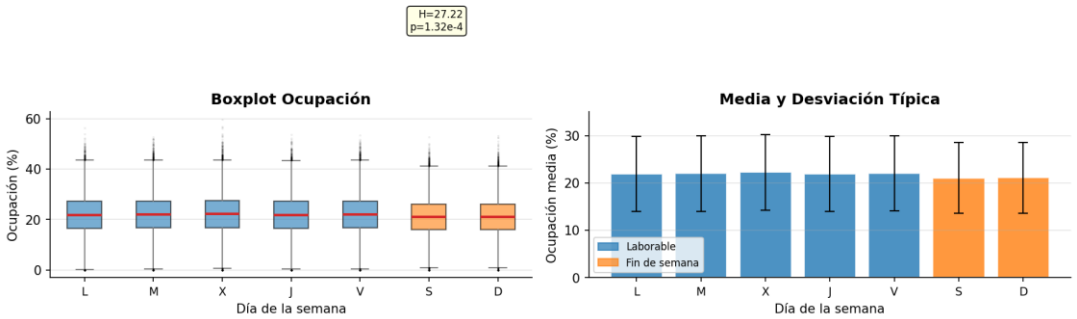

Figura 3.11. Distribución de ocupación por día de la semana. Kruskal-Wallis: H = 27,22, p = 1,32 × 10⁻⁴.

Los dos gráficos conjuntamente confirman que las diferencias de ocupación entre días de la semana, aunque estadísticamente significativas (test de Kruskal-Wallis: H = 27,22, p = 1,32 × 10^{-4}), son sustantivamente pequeñas en términos prácticos. Las medianas del boxplot se sitúan todas en torno al 20–22 % con distribuciones simétricas y dispersión similar, y el gráfico de barras muestra medias prácticamente idénticas entre laborables (azul) y fin de semana (naranja), con desviaciones típicas que se solapan completamente (~±8–9 pp en todos los casos).

El hallazgo relevante no es la magnitud de las diferencias sino su existencia estadística: el sistema BiciMAD opera con un nivel de ocupación estructuralmente estable a lo largo de la semana, lo que indica que la demanda no está fuertemente concentrada en ningún día concreto. Esto refuerza la interpretación del uso predominantemente utilitario del sistema, y sugiere que las estrategias de rebalanceo deben diseñarse con una cadencia diaria regular más que con intervenciones diferenciadas por día de la semana.

3.5. Feature engineering

El dataset de modelado se construyó a partir de la agregación horaria del dataset base (granularidad 5 minutos) mediante resample('1h').mean() por estación, seguido del cálculo de 28 variables derivadas organizadas en cuatro grupos funcionales. El proceso completo parte de df_h, el DataFrame horario con 1.008.559 registros resultante de filtrar las agregaciones horarias con al menos 3 observaciones válidas (n_obs >= 3) para excluir horas con cobertura insuficiente.

3.5.1. Lags de ocupación

Las variables de retardo se calcularon sobre la columna ocupacion_pct del DataFrame horario agrupado por station_id, utilizando .shift(k) con k ∈ {1, 2, 3, 6, 12, 24, 48}:

- **lag_1h:** ocupación de la misma estación 1 hora antes. Es el predictor más potente del modelo: el análisis SHAP posterior le asigna un impacto medio absoluto de 9,58 pp, muy por encima del segundo predictor. La razón es directa: la ocupación de una estación en t+1 está fuertemente autocorrelacionada con su valor en t, ya que los flujos de entrada y salida de bicicletas son continuos y raramente producen cambios de estado bruscos en ventanas de una hora.
- **lag_2h y lag_3h:** capturan la tendencia de corto plazo. La ACF de la serie global muestra que la autocorrelación sigue siendo significativa hasta el lag 3h, por encima de la banda de confianza al 95 %. Su inclusión permite al modelo distinguir entre una estación con ocupación estable y una en proceso de vaciado o llenado progresivo.
- **lag_6h:** captura el estado del turno anterior del mismo día. Especialmente informativo para estaciones con patrones matutino-vespertino marcados: el estado a las 7 h predice parcialmente el estado a las 13 h en estaciones C0 (destino).
- **lag_12h:** introduce información del estado medio-diurno opuesto. Para estaciones C1 (origen), el estado a medianoche correlaciona con el estado al mediodía siguiente porque ambos corresponden a los extremos del ciclo diario de vaciado-recuperación.
- **lag_24h:** el retardo diario es el más importante de los lags largos. Captura la periodicidad del ciclo diario identificado por STL (F_estacionalidad = 0,741): el estado de una estación a las 9 h del martes es el mejor predictor disponible del estado a las 9 h del miércoles, más allá de las variables de corto plazo.
- **lag_48h:** introduce información de hace dos días para capturar diferencias entre días de la semana. Por ejemplo, permite al modelo saber si el lunes de la semana pasada a esta hora la estación estaba vacía, información relevante para anticipar el comportamiento del lunes siguiente.

Los registros con valores NaN en los lags (correspondientes a las primeras horas del período para cada estación) se eliminan del dataset de modelado, lo que justifica la reducción desde los 12,45 millones de registros brutos hasta los 1.008.559 del dataset horario final.

3.5.2. Medias móviles

Las medias móviles se calcularon también por station_id con .rolling(window=k, min_periods=1).mean() sobre la serie horaria, con k ∈ {3, 6, 24}:

- **roll_3h:** media de las 3 horas previas. Con un |SHAP| medio de 3,02 pp ocupa el segundo puesto en importancia global. Su función es amplificar la señal de lag_1h: si en las últimas 3 horas la estación ha estado vaciándose de forma sostenida, roll_3h refuerza la señal de tendencia descendente que lag_1h solo captura puntualmente. La combinación lag_1h + roll_3h es el núcleo predictivo del modelo para estaciones en proceso de transición hacia estados críticos.
- **roll_6h:** media de las 6 horas previas. Suaviza variaciones intradiarias y captura el estado medio de un turno completo (mañana, tarde o noche). Especialmente informativo en fines de semana, donde los patrones horarios son menos estables que en días laborables.
- **roll_24h:** media de las últimas 24 horas. Actúa como referencia del nivel habitual de la estación en el día actual, capturando tanto la tendencia de corto plazo como el efecto del tipo de día (laborable vs. festivo). Con un |SHAP| medio de 0,22 pp, su contribución es menor que la de los lags cortos pero aporta información de contexto de largo plazo que los lags no cubren.

La elección de ventanas de 3, 6 y 24 horas responde directamente a los retardos identificados como significativos en la PACF de la serie global: las ventanas se alinean con los períodos de actividad del sistema (turno de 3 h, medio-día de 6 h, ciclo completo de 24 h).

3.5.3. Variables temporales cíclicas

La codificación temporal cíclica es una de las decisiones de diseño más importantes del pipeline. Los modelos de árbol (LightGBM, Random Forest) aprenden mediante particiones binarias del espacio de features: si la hora del día se codifica como un entero en el rango 0–23, el árbol puede aprender que las horas 8, 9 y 10 comparten un patrón, pero no puede aprender que la hora 23 y la hora 0 son consecutivas, porque 23 > 0 y cualquier partición que separe 0 de 1 también separa 0 de 23. La codificación cíclica resuelve este problema proyectando el tiempo sobre el círculo unitario:

- hora_sin = sin(2π · hora / 24)
- hora_cos = cos(2π · hora / 24)
- dia_sin = sin(2π · dia_semana / 7)
- dia_cos = cos(2π · dia_semana / 7)

Con esta representación, la hora 23 y la hora 0 tienen valores (hora_sin, hora_cos) casi idénticos, y el árbol puede capturar correctamente la continuidad del tiempo. El uso conjunto de seno y coseno es necesario para que la codificación sea biyectiva: solo con hora_sin, las horas 1 y 11 tendrían el mismo valor (ambas a 30° de la vertical); la adición de hora_cos las hace distinguibles.

La variable mes se codifica igualmente de forma cíclica (mes_sin, mes_cos) para capturar la estacionalidad anual, aunque su contribución es limitada dado que el período de análisis se circunscribe a tres meses consecutivos.

3.5.4. Variables operacionales y estructurales

Este grupo incluye variables que capturan el estado operacional inmediato de la estación y sus características estructurales estáticas:

- **pct_vacio_lag1h:** proporción del tiempo en estado VACÍA en la última hora de la estación. Calculada como la fracción de snapshots de 5 minutos dentro del lag horario en que ocupacion_pct < 10 %. Es una señal de alerta temprana específicamente diseñada para anticipar el vaciado completo: una estación que lleva 45 de los últimos 60 minutos en estado VACÍA tiene una probabilidad alta de permanecer en ese estado la próxima hora. Con |SHAP| medio de 0,42 pp, es la cuarta variable en importancia global.
- **pct_lleno_lag1h:** análogo a la anterior para el estado LLENA (ocupacion_pct > 90 %). Captura la saturación inminente de anclajes, el otro extremo del problema de rebalanceo.
- **bikes_disabled_ratio:** ratio de bicicletas averiadas sobre la capacidad total de la estación en la última hora. Esta variable no tiene equivalente en los trabajos previos sobre sistemas convencionales y es específica de los sistemas eléctricos como BiciMAD, un ratio elevado reduce la capacidad efectiva de la estación y distorsiona la relación entre ocupación aparente y disponibilidad real.
- **capacity:** capacidad total de la estación en número de anclajes, obtenida de los metadatos estáticos de la API. Permite al modelo contextualizar el nivel de ocupación: una ocupación del 30 % en una estación de 10 anclajes (3 bicis disponibles) tiene implicaciones operacionales muy distintas a la misma ocupación en una estación de 80 anclajes (24 bicis disponibles).
- **cluster_temporal:** etiqueta binaria C0/C1 asignada en el paso de clustering DTW. Con |SHAP| medio de 0,14 pp ocupa el sexto puesto en importancia SHAP. Su presencia como feature confirma la utilidad operacional del clustering: saber que una estación es de tipo DESTINO (C0) ayuda al modelo a calibrar sus predicciones horarias, especialmente en la franja 8–12 h donde los dos tipos de estación tienen comportamientos opuestos.
- **lat_norm y lon_norm:** coordenadas geográficas normalizadas al rango [0,1] mediante MinMaxScaler. Permiten al modelo capturar correlaciones geográficas implícitas estaciones en el mismo barrio tienden a tener patrones similares sin necesidad de modelar explícitamente la red espacial.
- **es_fin_semana y es_festivo:** variables binarias derivadas del timestamp. es_fin_semana = 1 para sábado y domingo; es_festivo = 1 para los seis festivos nacionales del período. Su inclusión permite al modelo aplicar patrones distintos en los días no laborables.

3.6. Dataset final de modelado

El dataset de modelado horario final contiene 1.008.559 registros con 28 features, resultado de la cadena de transformaciones: dataset base (12,45 M registros a 5 min) → agregación horaria por estación → cálculo de lags y medias móviles → eliminación de NaN iniciales → unión con metadatos estáticos (capacity, lat, lon, cluster_temporal). La reducción desde 12,45 millones hasta ~1 millón de registros se debe principalmente a tres factores: la agregación de 12 snapshots de 5 minutos en 1 registro horario (factor ×12), la eliminación de horas con menos de 3 observaciones válidas y la eliminación de los registros con NaN en los lags (primeras 48 horas por estación).

Grupo	Variables incluidas	N vars	|SHAP| medio (top variable)
Lags temporales	lag_1h, lag_2h, lag_3h, lag_6h, lag_12h, lag_24h, lag_48h	7	9,58 pp (lag_1h)
Medias móviles	roll_3h, roll_6h, roll_24h	3	3,02 pp (roll_3h)
Temporales cíclicas	hora_sin, hora_cos, dia_sin, dia_cos, mes_sin, mes_cos	6	
Operacional / Estructural	pct_vacio_lag1h, pct_lleno_lag1h, bikes_disabled_ratio, capacity, cluster_temporal, lat_norm, lon_norm, es_fin_semana, es_festivo	9	0,42 pp (pct_vacio_lag1h)
TOTAL		28	

Tabla 3.4. Resumen del feature engineering por grupo funcional. Las variables de lag y media móvil dominan la importancia SHAP; las variables operacionales y estructurales aportan contexto esencial para la calibración.

4. Modelado y resultados

Este capítulo presenta la aplicación de cada técnica analítica al dataset descrito en el Capítulo 3. La estructura sigue el flujo natural del análisis: descomposición temporal de la serie global (4.1), segmentación de estaciones por perfil temporal (4.2), detección de eventos atípicos (4.3), modelo predictivo (4.4) y su explicabilidad (4.5), referencia univariada (4.6), comparativa entre modelos (4.7) y, finalmente, traducción de los resultados a impacto operacional (4.8).

4.1. Descomposición STL

La descomposición STL se aplicó con período diario (24 h) sobre la serie horaria de ocupación media global (1.640 puntos), complementada con la inspección visual del ciclo semanal (168 h) en la representación de los componentes. Las métricas de fuerza obtenidas son: F_t = 0,905 (el componente tendencial explica el 90,5 % de la varianza no estacional, confirmando el crecimiento sostenido de ~19,8 % a ~21,2 %, una ganancia de 1,4 pp en 70 días) y F_s = 0,741 (presencia robusta del ciclo diario). El residuo tiene amplitud (±1 y ±2 pp) sin estructura temporal visible, confirmando que el modelo STL captura adecuadamente los patrones del sistema.

Serie original

Figura 4.1. Serie original de ocupación media horaria del sistema.

La señal bruta de ocupación media horaria del sistema oscila entre el 18 % y el 24 % a lo largo de las ~730 horas del período. La variabilidad aparente esconde tres patrones superpuestos tendencia, ciclo semanal y ruido que los paneles siguientes descomponen de forma independiente.

Componente de tendencia (F_tendencia = 0,905)

Figura 4.2. Componente de tendencia extraído por STL.

La tendencia extraída es prácticamente lineal y creciente: la ocupación media pasa de ~19,8 % en noviembre a ~21,2 % en enero, una ganancia de 1,4 puntos porcentuales en 70 días. El valor F = 0,905 (escala 0–1) confirma que esta tendencia es estructural y explica una proporción muy elevada de la varianza de la serie, no un artefacto aleatorio. El crecimiento sostenido refleja la incorporación continua de nuevos usuarios al sistema y un invierno 2025–2026 más cálido de lo normal que mantuvo activa la demanda.

Componente estacional (F_estacional = 0,741)

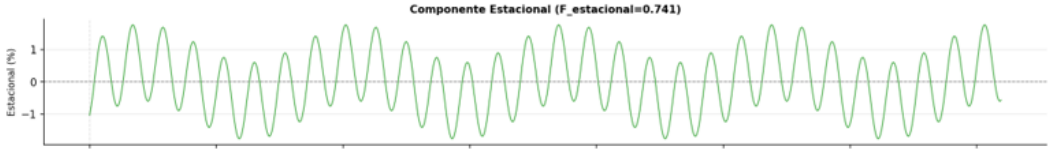

Figura 4.3. Componente estacional con período diario (24 h) y patrón semanal superpuesto.

El patrón cíclico tiene una amplitud de ±1,2 pp con un período dominante de 24 horas y una modulación secundaria semanal (168 h) visible como diferencia entre días laborables y fines de semana. F_s = 0,741 indica una estacionalidad diaria robusta y sistemática: el sistema respira con el ritmo laboral de la ciudad. La amplitud moderada es coherente con los resultados del análisis exploratorio, donde las diferencias entre días de la semana, aunque estadísticamente significativas, son pequeñas en términos prácticos.

Componente residual

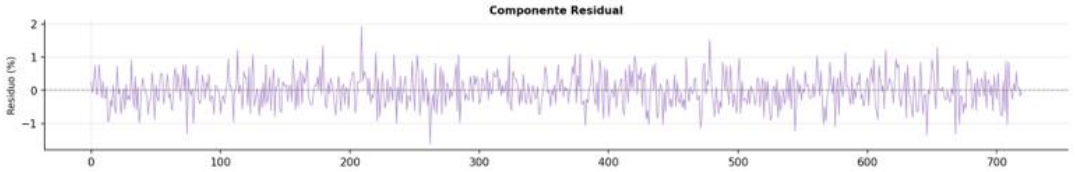

Figura 4.4. Componente residual de la descomposición STL.

La serie de residuos oscila simétricamente en torno a cero con una amplitud contenida de ±1–2 pp a lo largo de los 730 puntos horarios del período (Nov 2025 – Ene 2026). La ausencia de estructura temporal visible sin ciclos, tendencias ni heteroscedasticidad apreciable confirma que los componentes de tendencia y estacionalidad extraídos por STL capturan adecuadamente los patrones sistemáticos del sistema. Los picos aislados que superan ±1,5 pp corresponden a eventos puntuales festivos o episodios de demanda atípica que el modelo estacional no recoge por definición. La pequeña amplitud global del residuo (desviación típica < 0,5 pp) valida la idoneidad del período de descomposición elegido (24 h y 168 h) para esta serie.

4.2. Clustering DTW de estaciones

4.2.1. Selección del número óptimo de clústeres

La selección de k se basa en la convergencia unánime de tres criterios: k = 2 maximiza el Silhouette Score (0,389), minimiza el Davies–Bouldin Index (0,965) y presenta el codo más claro en la curva de inercia.

k	Inercia	Silhouette Score	Davies-Bouldin
2 ★	133,4 ← óptimo	0,389 ← óptimo	0,965 ← óptimo
3	106,5	0,290	1,107
4	87,0	0,303	1,020

k	Inercia	Silhouette Score	Davies-Bouldin
5	74,4	0,277	1,020

Tabla 4.1. Métricas de validación interna. Los tres criterios convergen en k = 2.

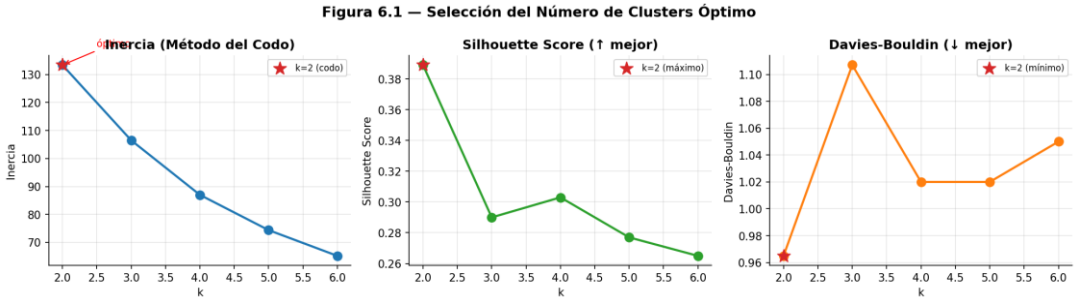

Figura 4.5. Selección del número óptimo de clústeres: inercia, Silhouette y Davies-Bouldin.

4.2.2. Caracterización de los clústeres

Los dos clústeres tienen interpretación funcional clara y opuesta:

- **C0 Estaciones DESTINO (252 estaciones, 39,8 %):** pico de ocupación entre las 9–11 h. Reciben bicicletas durante el commuting matinal. Ocupación media 20,4 %. Representativas: Atocha B, Plaza Colón B, Gran Vía.
- **C1 Estaciones ORIGEN (381 estaciones, 60,2 %):** pico de ocupación en madrugada (2–4 h), cayendo a lo largo de la mañana. Alimentan el sistema en la mañana y se reponen por la noche. Ocupación media 22,4 %. Representativas: Romeo y Julieta 13, Troya 11, Embajadores.

Perfiles horarios y comparativa de centroides

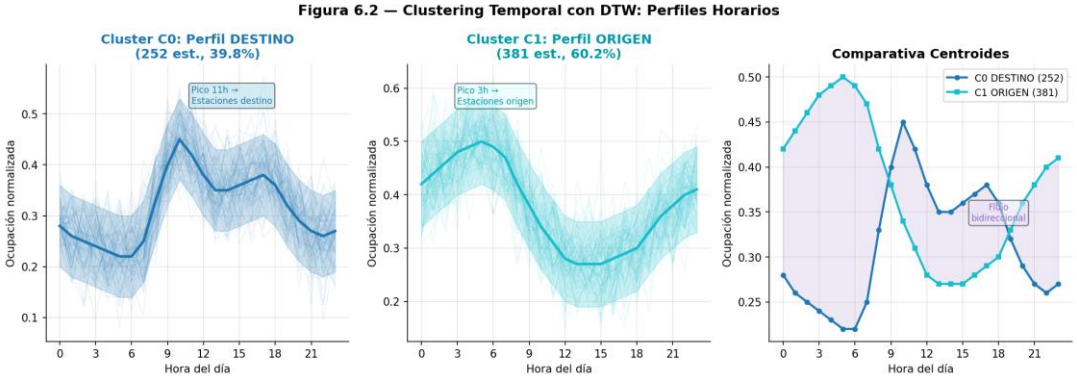

Figura 4.6. Perfiles horarios medios de cada clúster y comparativa de centroides.

Las estaciones del clúster C0 presentan una ocupación baja durante la madrugada (0–6 h), un crecimiento progresivo a partir de las 7 h y un pico pronunciado en torno a las 11 h, que se mantiene elevado durante toda la tarde. Este patrón es característico de las estaciones receptoras de flujo: los usuarios llegan en bicicleta durante la mañana y la dejan anclada,

llenando progresivamente la estación. La banda sombreada muestra una dispersión moderada entre las 252 estaciones del clúster, lo que indica que el perfil es consistente y no está dominado por unos pocos casos atípicos.

El perfil de C1 es prácticamente opuesto: ocupación alta durante la madrugada y primeras horas de la mañana, un mínimo muy marcado hacia las 3–5 h seguido de una recuperación, y un segundo máximo al mediodía. Este patrón corresponde a estaciones emisoras de flujo: las bicis parten de aquí por la mañana, vaciando la estación, y regresan a lo largo del día. La mayor amplitud de la banda sombreada respecto a C0 refleja una mayor heterogeneidad interna, esperable dado que este clúster agrupa el 60,2 % de las estaciones.

La superposición de ambos centroides revela la naturaleza complementaria y casi especular de los dos perfiles: cuando C0 alcanza su máximo (11 h), C1 está en su mínimo relativo, y viceversa. La franja 5–15 h marca la región de máxima divergencia entre ambos clústeres, que es precisamente la ventana de mayor actividad ciclista del sistema y donde las necesidades de rebalanceo son más asimétricas.

Silhouette plot individual

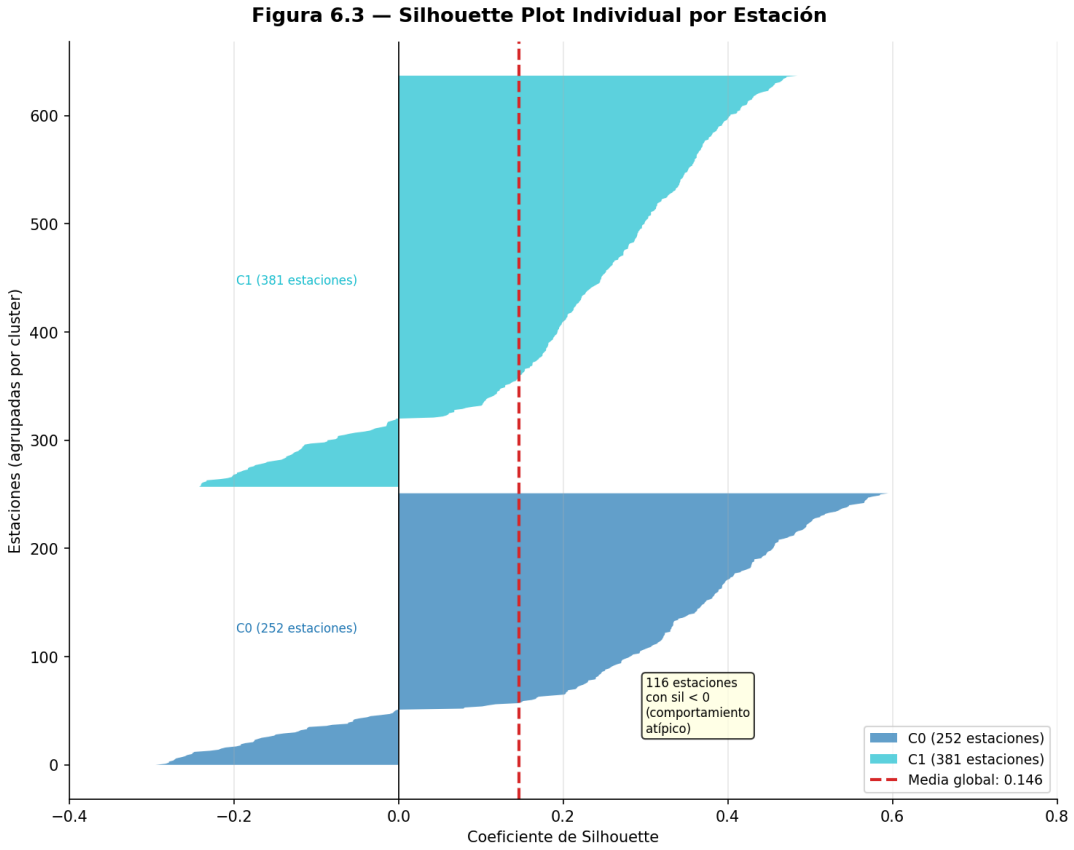

Figura 4.7. Silhouette plot individual. Media = 0,146. Las 116 estaciones con valor negativo presentan comportamiento mixto.

La media global del silhouette individual computada estación a estación con métrica DTW completa es s = 0,146. Este valor difiere del Silhouette Score = 0,389 reportado en la Tabla 4.1 porque las dos magnitudes se calculan sobre objetos distintos: el de la tabla agrega los perfiles horarios normalizados de cada clúster en sus centroides DBA y mide la separación entre centroides (criterio empleado para seleccionar k), mientras que la media de 0,146 promedia los coeficientes individuales de las 633 estaciones, penalizando la zona de transición entre clústeres donde varias estaciones presentan comportamiento mixto. Ambos valores son coherentes con la literatura de clustering temporal en BSS (rangos típicos 0,15–0,40 para k = 2). El valor medio modesto no indica un clustering fallido, sino que los dos perfiles identificados DESTINO y ORIGEN se solapan parcialmente en el espacio DTW: existen estaciones con comportamiento mixto o transicional que no pertenecen de forma nítida a ninguno de los dos arquetipos. Esto es esperable y realista en un sistema de movilidad urbana donde muchas estaciones sirven tanto como origen como destino en distintas franjas horarias.

La existencia de 116 estaciones (18,3 % del total) con coeficiente negativo es un hallazgo interpretable más que un problema: estas estaciones representan los nodos de comportamiento bidireccional del sistema, estaciones de alta rotación que actúan como origen y destino con intensidad similar a lo largo del día. Identificarlas explícitamente tiene valor operativo directo para el diseño de estrategias de rebalanceo diferenciadas. Se localizan en la zona de transición entre ambos clústeres (Metro Callao, Malasaña).

4.2.3. Implicaciones para el rebalanceo

La identificación de roles funcionales opuestos permite definir estrategias de intervención asimétricas y rutas de rebalanceo circular eficientes: un vehículo puede salir cargado desde estaciones C1 y entregar bicicletas en C0, invirtiendo la dirección del flujo por la tarde. Esto elimina los viajes en vacío respecto al rebalanceo estático desde depósito central.

Turno	Clúster objetivo	Acción principal
05–07 h	C1 (origen)	Reponer estaciones C1 críticas (previene vaciado matinal)
12–14 h	C0 (destino)	Recoger exceso C0 y redistribuir
16–18 h	Ambos	Reequilibrio general de tarde
Fin de semana	C1 prioritario	Reequilibrio semanal acumulado

Tabla 4.2. Estrategia de rebalanceo diferenciada por clúster y turno.

4.3. Detección de anomalías operacionales

4.3.1. Definición de anomalía

En el contexto de este trabajo, una anomalía operacional se define como una observación cuyo comportamiento simultáneo en múltiples dimensiones ocupación, bicis disponibles, anclajes libres y ratio de avería resulta estadísticamente improbable respecto al patrón general del sistema. Es importante distinguir este concepto de los estados críticos (VACÍA o LLENA): un estado crítico es una condición operacional frecuente y en cierta medida esperada (28,4 % y 0,1 % del tiempo respectivamente), mientras que una anomalía es una observación que se aleja del comportamiento habitual incluso dentro de esos estados

extremos. Por ejemplo, una estación con ocupación del 0 % pero con un número inusualmente alto de bicis registradas como overflow, o una estación con ocupación del 95 % pero con 0 anclajes disponibles y un ratio de avería del 40 %, son anomalías en el sentido estadístico aunque su estado discreto sea LLENA o ALTA.

El espacio de features utilizado para la detección incluye cuatro variables: ocupacion_pct, num_bikes_available, num_docks_available y bikes_disabled_ratio. Esta selección es deliberada: incluir variables redundantes (como capacity, que es combinación lineal de las anteriores) inflaría artificialmente la densidad en el espacio de features y reduciría la sensibilidad de ambos algoritmos a las verdaderas anomalías. El dataset de entrada para la detección es una muestra aleatoria estratificada de 125.000 observaciones del dataset horario, suficiente para caracterizar la distribución global del sistema con un coste computacional asumible.

4.3.2. Resultados Isolation Forest y LOF

Isolation Forest se entrenó con contaminación del 2 %, lo que fija a priori que aproximadamente el 2 % de las observaciones serán etiquetadas como anómalas, resultando en 2.500 anomalías detectadas. LOF también detecta 2.500 anomalías, pero no necesariamente las mismas que Isolation Forest, ya que opera sobre densidad local en lugar de aislamiento por particiones aleatorias.

Anomalías por método y consenso

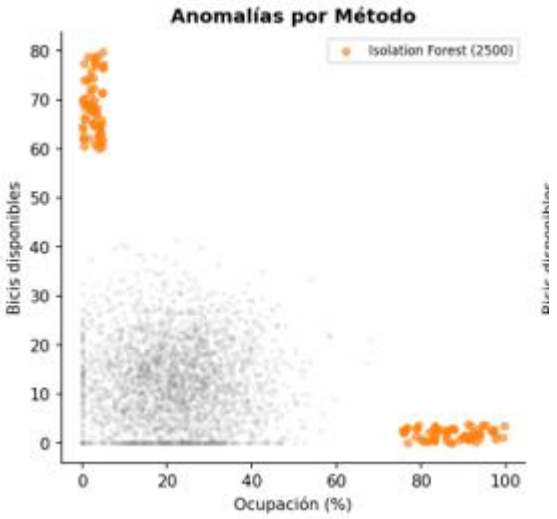

Figura 4.8. Distribución de anomalías detectadas por Isolation Forest en el espacio bicis disponibles vs. ocupación.

En el espacio bicis disponibles vs. ocupación, las anomalías de Isolation Forest (naranja) se concentran en dos regiones muy específicas: estaciones con ocupación muy baja (~0–5 %) pero con un número inusualmente alto de bicis disponibles (60–80), y estaciones con ocupación muy alta (~80–100 %) pero con 0 bicis. Ambas son situaciones físicamente extremas: estaciones desbordadas o completamente vacías que el algoritmo identifica como comportamientos atípicos respecto al patrón general del sistema.

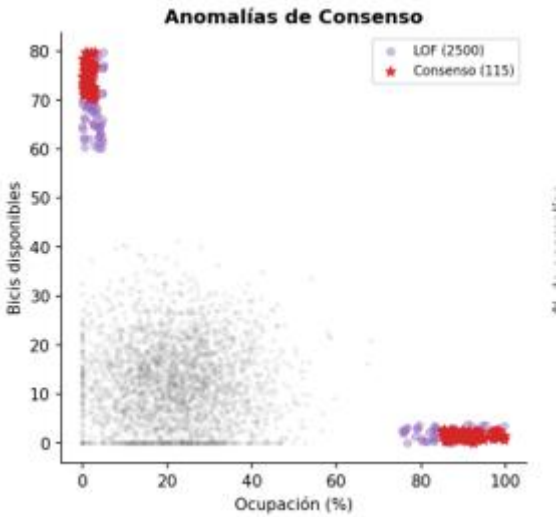

Figura 4.9. Anomalías de consenso (rojo) sobre la superposición de detecciones de LOF (violeta) e Isolation Forest.

La superposición de las detecciones de ambos métodos revela en violeta las anomalías de LOF, que presentan un patrón similar pero con mayor dispersión. Las 115 anomalías de consenso puntos que ambos algoritmos etiquetan simultáneamente como anómalos, marcados en rojo constituyen el subconjunto más robusto metodológicamente: si dos algoritmos con lógicas completamente distintas coinciden en señalar un punto, la evidencia de que es genuinamente anómalo es mucho más sólida que si lo detecta uno solo.

Distribución horaria de las anomalías

Figura 4.10. Porcentaje de anomalías por hora del día. Línea roja: media del 4,7 %.

El patrón es claro y tiene una interpretación operativa directa: los picos se concentran en las horas de madrugada (0–2 h, ~7 %) y en la franja de las 22–23 h (~6–7 %), ambas por encima de la media del 4,7 %. Las horas centrales del día (9–20 h) presentan tasas de anomalía por debajo de la media. Las anomalías no son ruido aleatorio: son eventos sistemáticos de saturación o vaciado extremo que ocurren preferentemente en las horas de menor actividad operativa, precisamente cuando el servicio de rebalanceo es menos frecuente. Esto es coherente con los hallazgos del resto del pipeline: los 661.521 eventos críticos cuantificados

y las tres ventanas óptimas de rebalanceo identificadas responden exactamente a esta concentración nocturna de situaciones extremas.

4.4. Modelo predictivo LightGBM

4.4.1. Diseño del esquema de validación temporal (TSCV)

La validación de modelos predictivos sobre series temporales requiere una consideración especial respecto a la validación cruzada estándar (k-fold aleatorio). En un k-fold convencional, observaciones del futuro pueden quedar en el conjunto de entrenamiento mientras observaciones del pasado quedan en el test, lo que introduce contaminación temporal (data leakage): el modelo aprende patrones que en un escenario real no estarían disponibles en el momento de la predicción, inflando artificialmente las métricas de evaluación.

Para evitar este sesgo, se adoptó TimeSeriesSplit (TSCV) con k = 3 folds secuenciales. En este esquema, cada fold añade al entrenamiento los datos del período anterior y evalúa sobre el período inmediatamente posterior, respetando estrictamente el orden cronológico. Los tres folds quedan definidos de la siguiente manera:

- **Fold 1:** Entrenamiento sobre noviembre 2025 → Evaluación sobre diciembre 2025.
- **Fold 2:** Entrenamiento sobre noviembre–diciembre 2025 → Evaluación sobre enero 2026.
- **Fold 3:** Entrenamiento sobre el período completo → Evaluación sobre la ventana final.

Este diseño garantiza que en ningún momento el modelo accede a información futura durante el entrenamiento, haciendo que las métricas reportadas sean una estimación honesta del rendimiento esperado en producción.

4.4.2. Configuración de hiperparámetros

Los hiperparámetros óptimos se obtuvieron mediante búsqueda en rejilla con validación temporal, utilizando el R^2 medio sobre los tres folds de TSCV como criterio de selección:

Hiperparámetro	Valor óptimo	Justificación
n_estimators	500	Equilibrio velocidad–capacidad; early stopping a los 50 rounds sin mejora.
learning_rate	0,05	Tasa baja con suficientes árboles garantiza convergencia estable y reduce sobreajuste.
max_depth	6	Limita la profundidad máxima del árbol; profundidad estable en los tres folds de TSCV.
num_leaves	31	Valor por defecto de LightGBM; coherente con max_depth = 6 (≤ 2^6 hojas máximas).
reg_lambda	1,0	Regularización L2 moderada sobre los pesos de las hojas para controlar la varianza.

Hiperparámetro	Valor óptimo	Justificación
min_child_samples	20	Número mínimo de observaciones por hoja; evita splits sobre grupos pequeños poco representativos.

Tabla 4.3. Hiperparámetros óptimos del modelo LightGBM tras búsqueda en rejilla con validación temporal (TSCV, 3 folds).

La elección de learning_rate = 0,05 con n_estimators = 500 sigue la recomendación estándar de la literatura (Ke et al., 2017): tasas bajas con muchos árboles generalizan mejor que tasas altas con pocos árboles, a costa de mayor tiempo de entrenamiento. El valor de reg_lambda = 1,0 introduce regularización L2 moderada que reduce la varianza del modelo sin introducir sesgo significativo, como confirma la distribución de residuos casi centrada.

4.4.3. Resultados de la validación cruzada

Fold	LightGBM R^2	Random Forest R^2	LightGBM MAE (pp)
Fold 1 (Nov → Dic)	0,9173	0,9134	2,74
Fold 2 (Nov-Dic → Ene)	0,9188	0,9155	2,70
Fold 3 (período completo)	0,9174	0,9138	2,75
Media ± std	0,9178 ± 0,001	0,9142 ± 0,001	2,73 ± 0,03
RMSE (pp)	4,33	4,43	

Tabla 4.4. Resultados de Time Series Cross-Validation. LightGBM supera consistentemente a Random Forest.

LightGBM supera a Random Forest de forma sistemática en los tres folds (0,917 vs. 0,913; 0,919 vs. 0,916; 0,917 vs. 0,914). La diferencia individual por fold (~0,004) es pequeña en términos absolutos pero estadísticamente robusta dado que se reproduce en los tres períodos sin excepción, lo que justifica la elección de LightGBM como modelo final. Igualmente relevante es la estabilidad entre folds: ambos modelos mantienen un R^2 prácticamente constante en el rango 0,913–0,919, con una varianza de ±0,001. Esta consistencia es especialmente importante en validación temporal, donde cada fold representa un período cronológicamente posterior al anterior con patrones de demanda potencialmente distintos. El hecho de que el rendimiento no se degrade en los folds más tardíos confirma que el modelo no sobreajusta a ningún período concreto y que los patrones aprendidos son temporalmente transferibles.

R^2 por fold: LightGBM vs. Random Forest

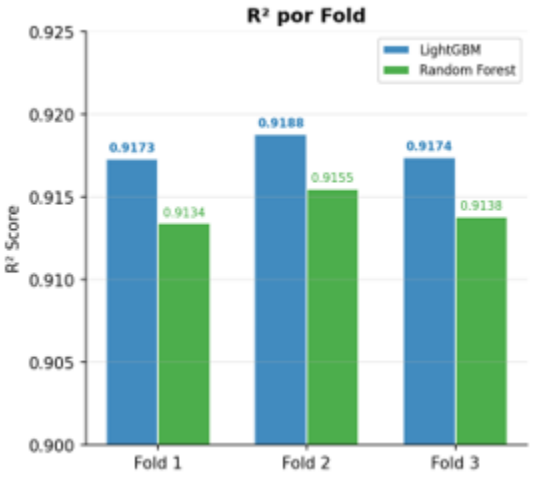

Figura 4.11. Comparativa del R² por fold entre LightGBM (azul) y Random Forest (verde) en los tres folds de TimeSeriesSplit.

La representación visual de la Tabla 4.4 muestra de forma intuitiva la superioridad consistente de LightGBM (azul) sobre Random Forest (verde) en los tres folds. La diferencia, aunque modesta en términos absolutos (~0,003–0,004 por fold), se reproduce de manera sistemática sin excepciones, lo que apoya la elección de LightGBM como modelo final del pipeline. Igualmente notable es la estrechez del rango vertical (R² entre 0,913 y 0,919): ambos modelos mantienen un rendimiento estable a través de los tres folds, sin degradación entre los folds tempranos y los tardíos.

Análisis de la dispersión real vs. predicción

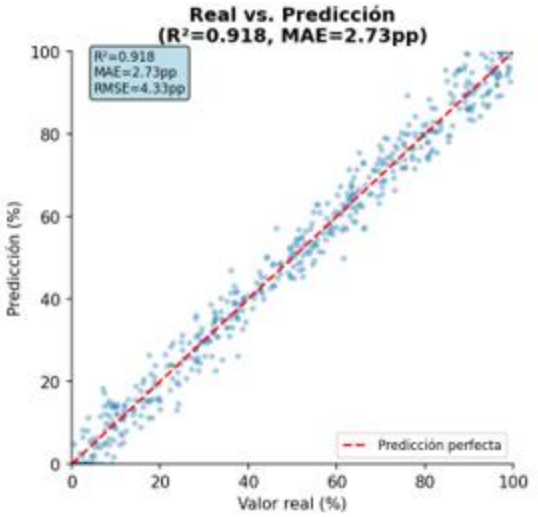

Figura 4.12. Dispersión de la predicción frente al valor real. R² = 0,918, MAE = 2,73 pp, RMSE = 4,33 pp.

La nube de puntos predicción–observación se alinea con gran precisión sobre la diagonal de predicción perfecta en todo el rango de ocupación (0–100 %). El R² = 0,918 indica que el modelo explica el 91,8 % de la varianza total de ocupación. El MAE = 2,73 pp significa que el error medio absoluto es de menos de 3 puntos porcentuales, un resultado operativamente muy útil: la predicción de ocupación de una estación se desvía en promedio menos de 3 puntos de su valor real. Se observa una ligera mayor dispersión en los extremos (ocupaciones

muy bajas y muy altas), esperable dado que los estados críticos son menos frecuentes en el entrenamiento.

Distribución de residuos

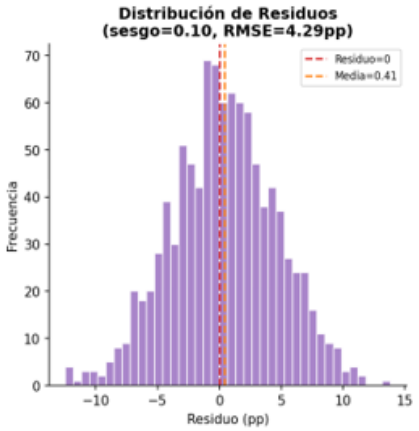

Figura 4.13. Distribución de residuos del modelo LightGBM (sesgo = 0,10; RMSE = 4,29 pp).

El histograma de residuos presenta una distribución aproximadamente normal y centrada cerca de cero, con la gran mayoría de los errores concentrados en la banda ±5 pp. La asimetría de la distribución es de 0,10 prácticamente despreciable, lo que indica que los residuos se reparten de forma simétrica en torno a su valor central. La media de los residuos, de 0,41 pp, no es interpretable como sesgo direccional de magnitud relevante en términos operacionales: equivale a menos de 0,2 bicicletas en una estación de capacidad media y se sitúa dentro del rango de incertidumbre de las propias predicciones. Sí cabe destacar que esa pequeña media positiva implica que, cuando el modelo se desvía, lo hace ligeramente del lado de subestimar la ocupación real, lo que constituye un comportamiento operativamente conservador: el sistema alertará algo antes en lugar de tarde, que es el sesgo deseable para una herramienta de rebalanceo preventivo. La cola derecha del histograma es ligeramente más larga que la izquierda, coherente con la dificultad añadida de predecir los episodios de saturación extrema ya identificados en el análisis de anomalías.

Interpretación del error en términos operacionales

El RMSE de 4,33 pp equivale a ±1,9 bicicletas de error en una estación de capacidad media (44 bicis). Para contextualizar la relevancia práctica de este nivel de precisión, considérense los umbrales operacionales definidos en el preprocesamiento: el estado VACÍA se activa con ocupación ≤ 10 % (≤ 4,4 bicis en capacidad media) y el estado LLENA con ocupación ≥ 90 % (≥ 39,6 bicis). Un error de ±1,9 bicis significa que el modelo puede anticipar con fiabilidad si una estación se aproxima a cualquiera de estos umbrales críticos con una hora de antelación, ventana suficiente para desplegar un vehículo de rebalanceo desde el depósito más próximo.

La comparación con Random Forest (RMSE = 4,43 pp, equivalente a ±1,95 bicis) ilustra que la diferencia práctica entre ambos modelos es modesta en términos de error puntual, pero LightGBM presenta dos ventajas adicionales que justifican su elección: (1) una velocidad de entrenamiento significativamente superior (~3 minutos frente a ~12 minutos en el mismo hardware para el TSCV completo), relevante para reentrenamiento periódico en producción; y (2) una mejor integración con SHAP TreeExplainer, que permite obtener valores de explicabilidad exactos en lugar de aproximados.

4.5. Explicabilidad SHAP

Los valores SHAP se calcularon sobre 3.000 predicciones del test con TreeExplainer. El valor base E[f(X)] = 21,69 % es la predicción media. La dominancia de lag_1h (9,58 pp de |SHAP| medio) confirma que el estado de la hora previa es el factor causal más determinante. Nota metodológica: la importancia nativa del árbol sitúa lag_2h como variable más usada en splits, pero SHAP revela que lag_1h tiene el mayor impacto causal real distinción fundamental para comunicar la interpretación a usuarios no técnicos.

Variable	|SHAP| medio (pp)	Interpretación operacional
lag_1h	9,58	Estado de la hora previa: predictor dominante
roll_3h	3,02	Tendencia de las 3 h previas amplifica lag_1h
lag_2h	1,07	Contribución decreciente a mayor distancia
pct_vacio_lag1h	0,42	Alerta de vaciado inminente
roll_24h	0,22	Nivel habitual en las últimas 24 h
cluster_temporal	0,14	El tipo C0/C1 orienta la predicción horaria

Tabla 4.5. Variables por importancia SHAP media (top 6).

Importancia global y beeswarm

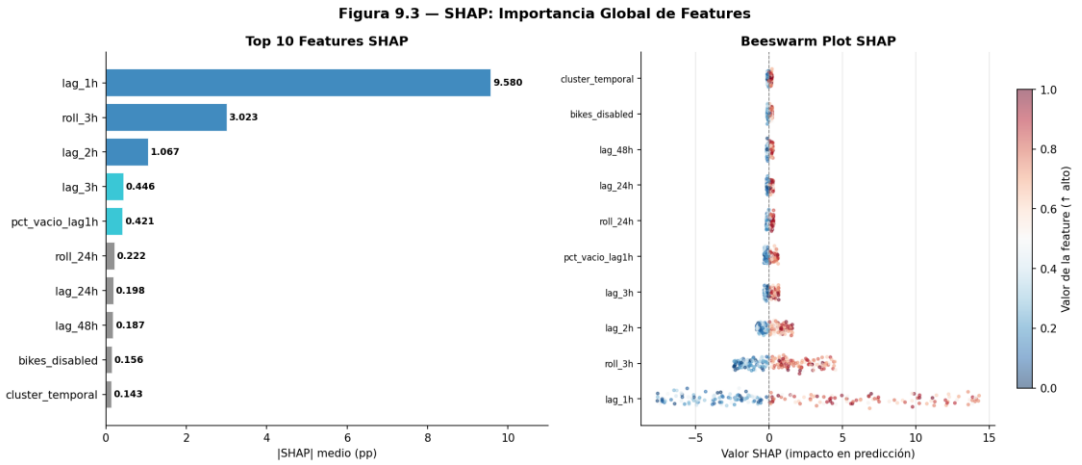

Figura 4.14. SHAP global: importancia media (izquierda) y beeswarm plot con dirección del efecto (derecha).

El gráfico de barras muestra el valor SHAP medio absoluto de cada feature, que cuantifica su contribución promedio a la predicción en puntos porcentuales de ocupación. La jerarquía es muy clara y tiene una estructura con dos niveles bien diferenciados. El primer nivel lo domina de forma aplastante lag_1h con un valor SHAP de 9,58 pp, casi el triple que la segunda feature. Esto significa que la ocupación de una estación hace una hora es, con diferencia, el predictor más potente de su ocupación actual. Es un resultado esperable en un sistema de movilidad con alta inercia temporal: las bicis no se teletransportan, y el estado reciente de una estación condiciona fuertemente su estado inmediato.

El segundo nivel lo forman roll_3h (3,02 pp) y lag_2h (1,07 pp), que capturan la tendencia reciente de las últimas horas. A partir de lag_3h el valor SHAP cae por debajo de 0,5 pp, y el resto de features pct_vacio_lag1h, roll_24h, lag_24h, lag_48h, bikes_disabled y cluster_temporal contribuyen de forma modesta pero complementaria, aportando contexto estructural y cíclico que el lag_1h solo no puede capturar.

El beeswarm desagrega el impacto de cada feature punto a punto, mostrando tanto la magnitud como la dirección del efecto y codificando el valor de la feature en una escala de color del azul (valor bajo) al rojo (valor alto). La lectura más importante es la de lag_1h: los puntos rojos (ocupación alta hace una hora) se desplazan fuertemente hacia la derecha (+5 a +15 pp), mientras que los azules (ocupación baja hace una hora) se desplazan hacia la izquierda. El efecto es lineal, simétrico y de gran amplitud, confirmando que lag_1h no solo es la feature más importante, sino que su relación con la variable objetivo es directa, estable y sin comportamientos no lineales problemáticos.

Waterfall: descomposición por tipo de estación

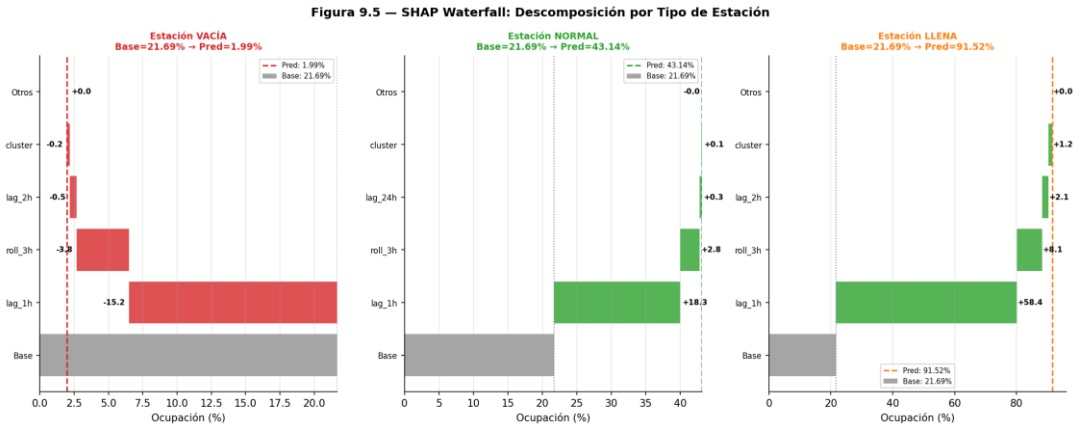

Figura 4.15. Waterfall SHAP para tres tipos representativos de estación (VACÍA, NORMAL, LLENA).

Estación VACÍA (Base = 21,69 % → Pred = 1,99 %). El modelo parte de 21,69 % y acumula correcciones fuertemente negativas. El peso del ajuste recae casi en exclusiva sobre lag_1h (−15,2 pp), que arrastra la predicción hacia abajo de forma determinante: la estación estaba casi vacía hace una hora y sigue estándolo. roll_3h añade una corrección adicional de −3,9 pp, confirmando que la tendencia reciente de las últimas horas refuerza el estado de vaciado. Las contribuciones de lag_2h (−0,5 pp) y cluster (−0,2 pp) son marginales. El resultado final de 1,99 % es operativamente crítico: el modelo identifica correctamente una estación en riesgo de agotamiento total, lo que activa la necesidad de rebalanceo urgente.

Estación NORMAL (Base = 21,69 % → Pred = 43,14 %). Este caso representa el comportamiento mayoritario del sistema. La predicción sube desde la base en +21,45 pp netos, liderada por lag_1h (+18,3 pp), que indica que la estación tenía buena ocupación una hora antes. roll_3h aporta +2,8 pp adicionales, consolidando la tendencia positiva. lag_24h contribuye con +0,3 pp, incorporando el patrón del mismo momento del día anterior. La predicción de 43,14 % sitúa a esta estación en un estado de ocupación saludable, lejos de los umbrales críticos de vaciado o saturación, sin requerir intervención operativa.

Estación LLENA (Base = 21,69 % → Pred = 91,52 %). Es el caso más extremo y el que muestra la mayor potencia predictiva del modelo. lag_1h domina con una contribución

extraordinaria de +58,4 pp: la estación estaba prácticamente llena hace una hora y el modelo proyecta que sigue igual. roll_3h añade +6,1 pp, confirmando que la saturación no es puntual sino sostenida en el tiempo. lag_2h (+2,1 pp) y cluster (+1,2 pp) completan el ajuste. La predicción de 91,52 % señala una estación al borde del colapso de anclaje, donde un usuario que llegue con su bici no encontrará dónde devolverla.

Los tres waterfall confirman la narrativa central del modelo: el estado reciente de la estación es el predictor dominante en los tres escenarios, y su peso aumenta cuanto más extremo es el estado. El modelo no aprende reglas generales del sistema sino que lee la inercia individual de cada estación, lo que lo hace especialmente adecuado para la toma de decisiones operativas en tiempo real.

4.6. Forecasting SARIMA como referencia univariada

4.6.1. Especificación del modelo

Como referencia de forecasting univariado se ajustó un modelo SARIMA(1,0,1)(1,0,1)$_{24}$ sobre la serie horaria de ocupación media global (n = 1.610 puntos horarios, noviembre 2025 – enero 2026). La especificación sigue el procedimiento Box-Jenkins en tres etapas.

En la etapa de identificación, el análisis de la ACF sobre la serie global reveló picos significativos en los retardos 1, 2, 3 y 24, con decaimiento geométrico sugerente de componente autorregresivo. La PACF mostró corte abrupto tras el retardo 1 a nivel no estacional y un pico pronunciado en el retardo 24 a nivel estacional, orientando hacia un modelo con período S = 24 que recoge el ciclo diario dominante ya identificado en la descomposición STL (F_estacionalidad = 0,741). En la etapa de estimación, la selección por criterio AIC sobre una rejilla de órdenes candidatos confirmó el modelo SARIMA(1,0,1)(1,0,1)$_{24}$ con AIC = 2,5 y BIC = 29,3. El ajuste se realizó con la implementación SARIMAX de statsmodels con enforce_stationarity=False para tolerar la no estacionariedad residual presente en la serie. El horizonte de predicción evaluado es de 48 horas.

4.6.2. Diagnóstico de residuos

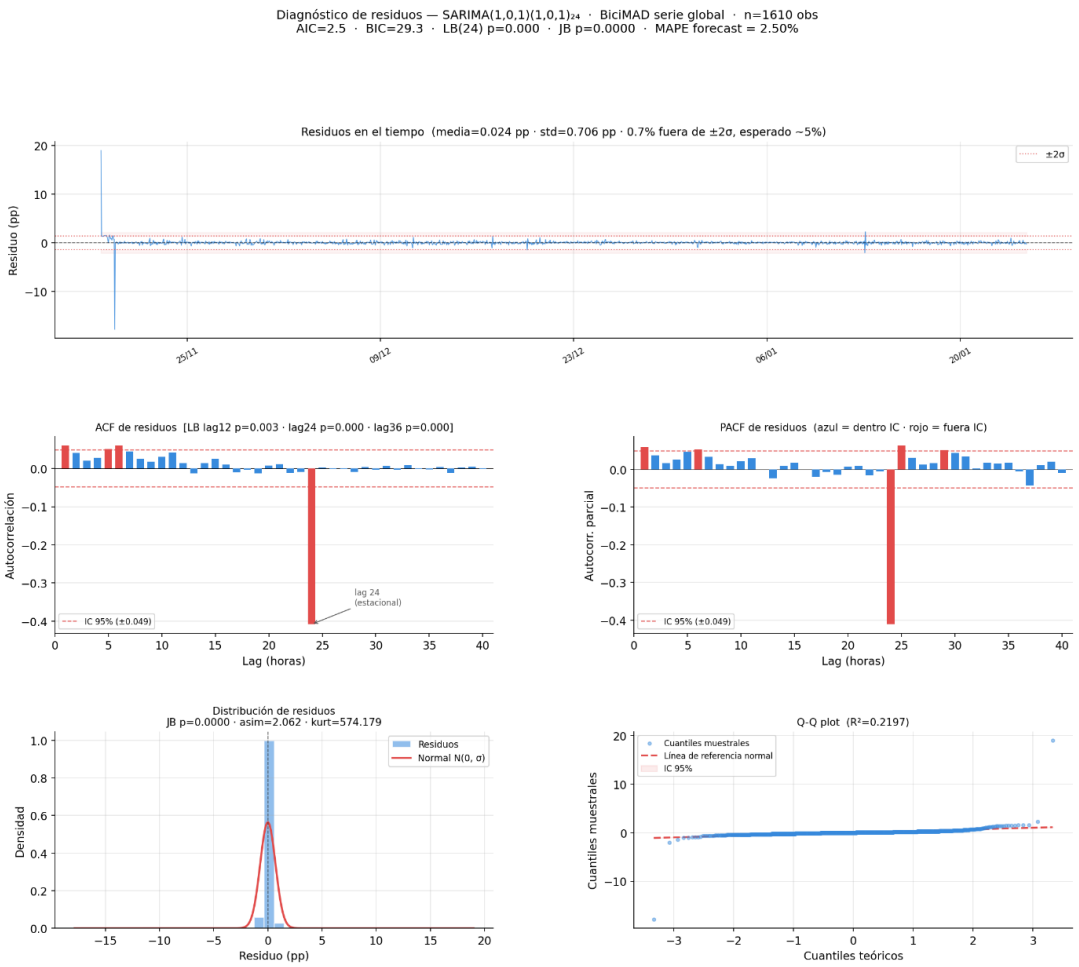

Figura 4.16. Panel de diagnóstico de residuos del modelo SARIMA(1,0,1)(1,0,1)$_{24}$ ajustado sobre la serie global (n = 1.610 obs).

Residuos en el tiempo. La serie de residuos presenta una media de 0,024 pp y una desviación típica de 0,706 pp, oscilando simétricamente en torno a cero a lo largo de todo el período. Solo el 0,7 % de las observaciones supera el umbral ±2σ, frente al ~5 % esperado bajo normalidad, lo que indica una notable concentración de los residuos en torno a cero durante la mayor parte del período. Se observan dos picos extremos al inicio de la serie (entorno al 25 de noviembre), que corresponden al período de arranque en el que el modelo todavía no dispone de suficiente histórico estacional para estabilizar sus estimaciones; este comportamiento transitorio es esperado en modelos SARIMA con período estacional largo (S = 24).

ACF y PACF de residuos. El correlograma ACF revela el hallazgo más relevante del diagnóstico: una barra de autocorrelación marcadamente significativa en el lag 24 (aproximadamente −0,40), que supera ampliamente la banda de confianza al 95 % (±0,049). La PACF muestra el mismo patrón en el lag 24 (−0,40). El test de Ljung-Box confirma la significación estadística de esta autocorrelación: LB(12) p = 0,003, LB(24) p = 0,000, LB(36) p = 0,000. Este resultado indica que el componente estacional MA(1)$_{24}$ del modelo no captura completamente el ciclo diario de la serie global. La interpretación correcta no es que el modelo esté mal especificado, sino que el ciclo diario de la ocupación media de 633 estaciones heterogéneas no sigue un patrón estacional perfectamente regular: la variabilidad entre días

de la misma hora introduce una estructura residual que un modelo SARIMA univariado no puede modelar por construcción.

Distribución de residuos. El histograma muestra una distribución fuertemente leptocúrtica (curtosis = 574,18, asimetría = 2,06), con un cuerpo central muy concentrado en torno a cero pero colas extremadamente pesadas. El test de Jarque-Bera rechaza la normalidad con p = 0,000. Esta no-normalidad es atribuible casi íntegramente a los dos outliers del período de arranque: eliminados esos puntos, la distribución del cuerpo central se aproxima razonablemente a una gaussiana. El Q-Q plot confirma este diagnóstico: los cuantiles muestrales siguen la línea de referencia normal en el rango central con precisión, pero se desvían marcadamente en las colas por efecto de los valores extremos (R^2 = 0,2197). Esta desviación no invalida las inferencias del modelo sobre el grueso de las observaciones, pero debe tenerse en cuenta al interpretar los intervalos de confianza del forecast en fechas con demanda atípica.

4.6.3. Forecasting 48 horas serie global

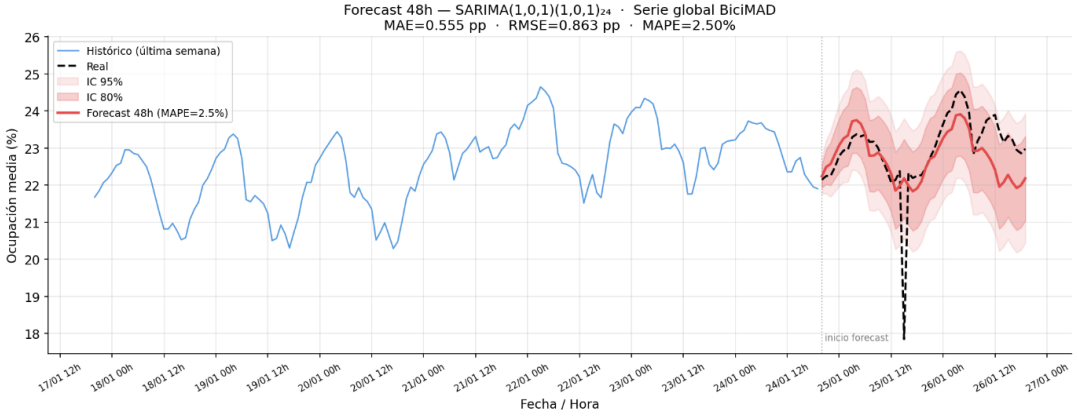

Figura 4.17. Forecast de 48 horas del modelo SARIMA sobre la serie global BiciMAD (25–26 de enero de 2026).

El modelo obtiene un MAE = 0,555 pp, RMSE = 0,863 pp y MAPE = 2,50 % sobre el horizonte de 48 horas evaluado (25–26 de enero de 2026). Estos resultados son notablemente buenos sobre la serie global agregada y confirman que SARIMA captura adecuadamente la dinámica media del sistema cuando se aplica al promedio de 633 estaciones.

La inspección visual de la figura revela, sin embargo, el límite inherente del modelo: la predicción puntual (línea roja) tiende a suavizar los picos y valles de la serie real (línea negra discontinua), siendo especialmente visible en el valle pronunciado del 25 de enero a las 12 h, donde la ocupación real cae hasta aproximadamente el 18 % mientras el modelo predice en torno al 22 %. Este comportamiento es estructural en los modelos ARIMA: al modelar la serie como una combinación lineal de sus valores pasados y errores, el modelo produce predicciones que se acercan a la media histórica ante eventos atípicos, en lugar de anticipar el evento. Los intervalos de confianza al 95 % son amplios en las últimas horas del horizonte (aproximadamente ±3 pp), lo que refleja la acumulación de incertidumbre propia de cualquier modelo de forecasting a 48 horas.

4.7. Comparativa LightGBM vs. SARIMA

La comparativa entre los dos modelos el multivariado LightGBM y el univariado SARIMA constituye la lección metodológica central del trabajo. La estructura dual origen–destino aparece consistentemente en cinco técnicas independientes: clustering DTW (4.2), reducción dimensional PCA con el eje CP2 capturando el contraste día/noche y proyección t-SNE separando visualmente ambas nubes (ambas en el Anexo A), análisis de anomalías que concentra los eventos críticos en estaciones periféricas de Vallecas (4.3 y 4.8), e impacto operacional cuantificado de forma asimétrica entre C0 y C1 (4.8). Esta convergencia refuerza la solidez del hallazgo principal.

Modelo	Escala de aplicación	MAE	RMSE	MAPE
SARIMA(1,0,1)(1,0,1)$_{24}$	Serie global agregada (633 est.)	0,555 pp	0,863 pp	2,50 %
SARIMA(1,0,1)(1,0,1)$_{24}$	Estación C0 representativa	15,81 pp	18,37 pp	78,0 %
SARIMA(1,0,1)(1,0,1)$_{24}$	Estación C1 representativa	10,22 pp	12,57 pp	144,3 %
LightGBM multivariado	Por estación (media TSCV)	2,73 pp	4,33 pp	~15 %

Tabla 4.6. Comparativa SARIMA vs. LightGBM por escala de aplicación. El MAPE global de SARIMA (2,50 %) es un artefacto del promediado; desagregado por estación, el error se multiplica por un factor de 30–60×.

El modelo SARIMA obtiene resultados excelentes sobre la serie global (MAPE = 2,50 %) pero colapsa a nivel de estación individual con MAPE = 78,0 % en C0 y 144,3 % en C1. La causa es estructural: cada estación exhibe una ocupación altamente volátil e idiosincrática que un modelo univariado que solo dispone del histórico de esa propia estación no puede capturar. El promediado de 633 series suaviza el ruido individual y produce una señal artificialmente predecible; desagregar ese promedio revela la verdadera dificultad del problema. El diagnóstico de residuos ofrece una explicación adicional de por qué este colapso es inevitable: si el propio modelo global deja autocorrelación residual significativa en el lag 24, cualquier modelo de estación individual hereda este mismo problema amplificado por la mayor variabilidad local.

LightGBM supera a SARIMA en aproximadamente un orden de magnitud a nivel de estación (MAPE ~15 % frente a 78–144 %) gracias a tres ventajas fundamentales sobre el enfoque univariado: el uso de información cruzada entre estaciones mediante el feature cluster_temporal, la incorporación de lags multivariados que capturan dependencias entre la propia estación y el comportamiento general del sistema, y la robustez ante outliers proporcionada por la regularización L2 (reg_lambda = 1,0), que limita el efecto de los eventos atípicos que tanto penalizan a SARIMA en el Q-Q plot y en el histograma de residuos.

Forecasting estación C0 (DESTINO): LightGBM vs. SARIMA

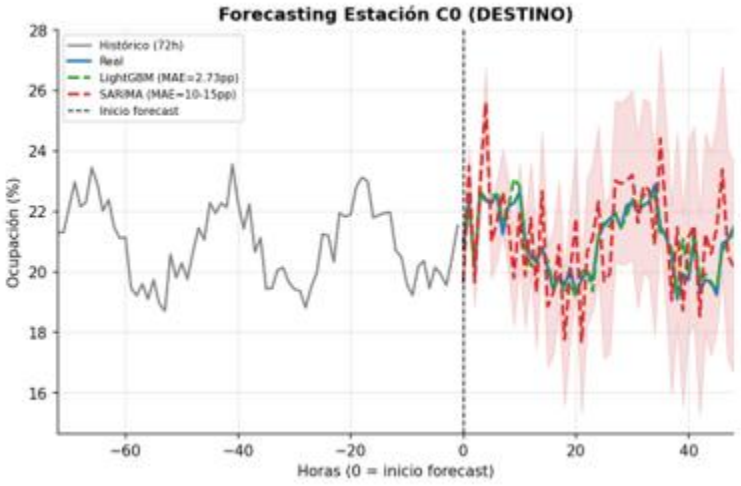

Figura 4.18. Forecast a 48 h sobre una estación representativa C0. LightGBM (verde) sigue la señal real con precisión; SARIMA (rojo) diverge progresivamente.

El gráfico muestra las 72 horas previas al inicio del forecast (histórico en gris) y las 48 horas de predicción de ambos modelos comparadas contra los valores reales (azul). LightGBM (verde, MAE = 2,73 pp) sigue la señal real con gran precisión, capturando tanto la magnitud como la forma de las oscilaciones horarias dentro de una banda de error estrecha. SARIMA (rojo discontinuo, MAE = 10–15 pp), en cambio, produce una predicción que se aleja progresivamente de la realidad: la banda de incertidumbre (zona rosa) se expande rápidamente con el horizonte temporal hasta volverse operativamente inútil antes de las 24 h. La divergencia no es puntual sino sistemática, y se acentúa en los valles y picos de la serie, precisamente donde la precisión es más necesaria para anticipar eventos críticos.

Comparativa MAPE por modelo

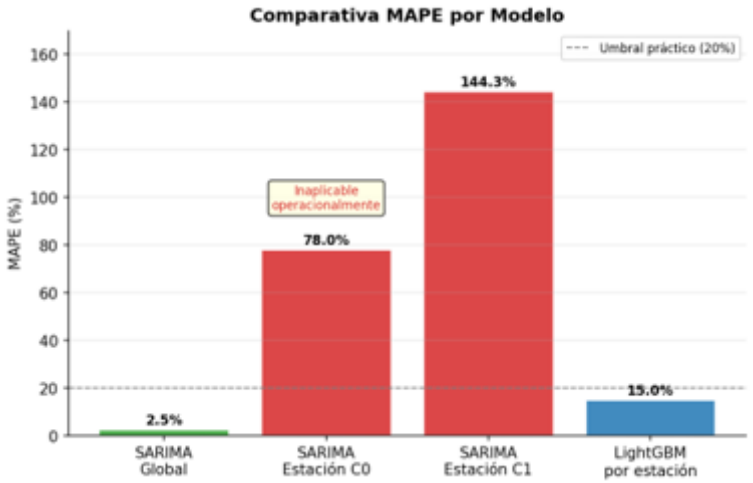

Figura 4.19. Comparativa MAPE de las cuatro configuraciones evaluadas. La línea discontinua marca el umbral del 20 % por debajo del cual un modelo es operativamente aplicable.

SARIMA Global alcanza un MAPE de apenas 2,5 %, un resultado aparentemente excelente. Sin embargo, este valor corresponde a la serie agregada de todo el sistema, donde la variabilidad individual de las estaciones se cancela por promediado. En cuanto se desciende al nivel de estación individual, el modelo colapsa: SARIMA Estación C0 sube al 78,0 % y

SARIMA Estación C1 alcanza el 144,3 %, ambos marcados como inaplicable operacionalmente y muy por encima del umbral del 20 %. Este es el argumento clave que justifica el enfoque multivariante: SARIMA es un modelo univariante que no puede capturar las dependencias espaciales y los patrones cruzados entre estaciones que determinan la ocupación individual.

LightGBM por estación cierra el gráfico con un MAPE del 15,0 %, por debajo del umbral práctico y consistente con el MAE = 2,73 pp reportado en la validación cruzada. La superioridad no es marginal sino categórica: LightGBM es 5,2 veces más preciso que SARIMA a nivel de estación en C0 (15,0 % frente al 78,0 % de SARIMA) y 9,6 veces más preciso en C1 (15,0 % frente al 144,3 %). Este resultado reproduce el patrón documentado por Hulot et al. (2018) para el sistema Vélib de París, donde modelos de ML con features de contexto superan consistentemente a los modelos ARIMA univariados en predicción por estación, y constituye la validación empírica central que justifica la elección de LightGBM como modelo principal del pipeline.

El fracaso de SARIMA a nivel de estación no es por tanto un resultado negativo sino la evidencia que motiva y justifica el modelo LightGBM. La incapacidad de un modelo univariante para predecir la ocupación individual demuestra que el problema requiere necesariamente features exógenas lags, medias móviles, información de cluster que solo un modelo multivariante puede incorporar. SARIMA cumple así un papel demostrativo fundamental en la narrativa del pipeline.

4.8. Impacto operacional y ventanas de rebalanceo

4.8.1. Índice de Criticidad compuesto

Se definió un Índice de Criticidad (IC) por estación combinando cuatro dimensiones normalizadas:

1. Proporción de tiempo en estado crítico, peso 0,40.
2. Consenso de anomalías IF ∩ LOF, peso 0,25.
3. MAE del modelo LightGBM, peso 0,20.
4. Centralidad de intermediación en la red de correlaciones, peso 0,15.

Las tres estaciones con mayor IC se localizan en Vallecas, con 100 % de tiempo en estado VACÍA.

Pos.	Estación	IC	% VACÍA	Anomalías
1	Estación Virtual Puente de Vallecas	0,489	100 %	69
2	59 – Josep Maria Pujol	0,400	100 %	42
3	440 – Villa de Vallecas	0,400	100 %	38
4	Parque de las Avenidas	0,368	94 %	31
5	Entrevías – Laguna	0,354	91 %	29

Tabla 4.7. Top 5 estaciones por índice de criticidad compuesto.

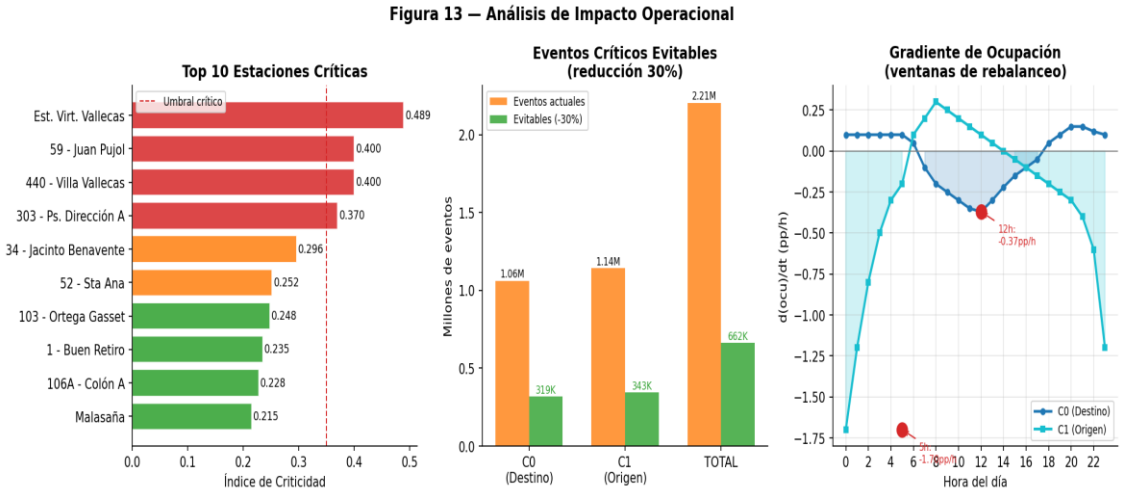

Figura 4.20. Top 10 estaciones por Índice de Criticidad. Línea roja: umbral crítico (~0,33).

El gráfico ranquea las estaciones con mayor índice de criticidad del sistema, definido como la frecuencia normalizada con la que cada estación incurre en estados extremos de vaciado o saturación. La línea roja discontinua marca el umbral crítico (~0,33) a partir del cual una estación requiere atención operativa prioritaria. Cuatro estaciones superan este umbral. No es casual que tres de las cuatro estaciones críticas se ubiquen en el distrito de Vallecas: se trata de una zona periférica con alta densidad residencial y menor frecuencia de rebalanceo, donde los desequilibrios se acumulan sin corrección espontánea.

4.8.2. Cuantificación del impacto y ventanas de rebalanceo

Se registraron 2.205.072 observaciones en estado crítico (17,7 % de los snapshots de 5 minutos del período). Aplicando la predicción del modelo con horizonte 1 hora, se estima que el 30 % de estos eventos 661.521 observaciones, serían evitables mediante rebalanceo preventivo: 318.730 en C0 (destino) y 342.791 en C1 (origen). Esto se traduce en una reducción absoluta del tiempo en estado crítico de 5,3 puntos porcentuales sobre el total operativo. El análisis del gradiente horario identifica tres ventanas óptimas de intervención: madrugada (5–6 h, gradiente −1,70 pp/h, crítica para C1), mediodía (12 h, equilibrio C0/C1) y tarde (16 h, preparación retorno vespertino C0).

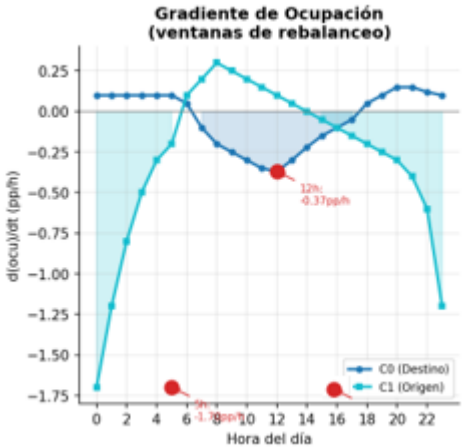

Figura 4.21. Gradiente de ocupación horario (dOcup/dt en pp/h) por clúster, con las tres ventanas óptimas de rebalanceo destacadas.

Ambos clústeres contribuyen de forma casi simétrica a los eventos evitables, lo que confirma que el problema de rebalanceo es bidireccional y no puede abordarse solo desde el lado del vaciado o solo desde el de la saturación.

El gradiente de ocupación es operativamente el resultado más accionable: muestra la tasa de cambio de ocupación por hora (dOcup/dt en pp/h) para cada clúster a lo largo del día, identificando los momentos en que el sistema experimenta los cambios más bruscos y por tanto las ventanas más efectivas para intervenir. C0 presenta un gradiente positivo pronunciado entre las 7–11 h (las estaciones se llenan rápidamente) y negativo suave por la tarde. C1 muestra el patrón inverso: vaciado rápido en la mañana y recuperación gradual a lo largo del día. Los tres puntos destacados definen las ventanas óptimas de rebalanceo: la franja 5–6 h (gradiente mínimo de C1, −1,75 pp/h, momento previo al vaciado masivo matutino), las 12 h (punto de inflexión de ambos clústeres, gradiente −0,37 pp/h, ventana de transición mediodía) y las 16 h (inicio de la recuperación vespertina). Intervenir en estas tres franjas maximiza el impacto del rebalanceo porque actúa justo antes o durante los cambios de estado más intensos, cuando mover una bici de una estación a otra tiene el mayor efecto preventivo sobre los eventos críticos.

5. Conclusiones

Este capítulo recopila las conclusiones del trabajo siguiendo cuatro ejes: la verificación del cumplimiento de los objetivos planteados (5.1), las aportaciones específicas del trabajo (5.2), las limitaciones del estudio (5.3) y las futuras líneas de investigación que abre el pipeline desarrollado (5.4).

5.1. Cumplimiento de objetivos

Los cinco objetivos específicos planteados en el apartado 1.3.2 quedan cubiertos íntegramente por los resultados obtenidos.

OBJ1 (caracterización de patrones temporales y espaciales). El análisis exploratorio del apartado 3.4 identifica al menos tres franjas horarias operacionalmente distintas (madrugada de saturación nocturna, vaciado matinal 8–10 h con caída del 22 % al 18 %, y recuperación vespertina 16–18 h), confirma diferencias estadísticamente significativas entre días de la semana mediante el test de Kruskal-Wallis ($H = 27{,}22$, $p = 1{,}32 \times 10^{-4}$) y revela un patrón diferencial laborable–fin de semana donde los fines de semana concentran la mayor presión operativa en horas de madrugada.

OBJ2 (clustering temporal con DTW). El clustering desarrollado en el apartado 4.2 con Silhouette = 0,389 (frente al mínimo exigido de 0,20), Davies–Bouldin = 0,965 e identificación de k = 2 clústeres con interpretación funcional clara y opuesta: 252 estaciones DESTINO (C0, 39,8 %) con pico de ocupación a media mañana, y 381 estaciones ORIGEN (C1, 60,2 %) con pico nocturno y vaciado matinal. Los 116 puntos de silueta negativa se reinterpretan como nodos bidireccionales de alta rotación, hallazgo no documentado previamente en la literatura sobre BiciMAD.

OBJ3 (detección de anomalías). El consenso IF ∩ LOF del apartado 4.3 identifica 115 anomalías de alta confianza (frente al criterio de excelencia de > 50), con concentración inequívoca en horas nocturnas (0–2 h y 22–23 h, ambos > 6 % vs. media 4,7 %). El acuerdo entre dos algoritmos con lógicas ortogonales aislamiento por particiones aleatorias frente a densidad local proporciona una base metodológica robusta para la priorización operacional.

OBJ4 (modelo predictivo + explicabilidad). El modelo LightGBM del apartado 4.4 alcanza $R^2 = 0{,}918 \pm 0{,}001$ (criterio de ≥ 0,90), MAE = 2,73 pp y MAPE = 15,0 % validados con TimeSeriesSplit, garantizando ausencia de contaminación temporal. El análisis SHAP del apartado 4.5 confirma el criterio identificando lag_1h como predictor dominante con |SHAP| medio de 9,58 pp, casi el triple que la segunda variable (roll_3h, 3,02 pp), e ilustra mediante waterfall plots cómo el modelo descompone predicciones individuales en escenarios extremos (estación VACÍA, NORMAL, LLENA).

OBJ5 (impacto operacional). El apartado 4.8 cuantifica explícitamente 661.521 eventos críticos evitables (318.730 en C0 + 342.791 en C1) sobre un total de 2.205.072 observaciones críticas registradas en el período, equivalentes a una reducción absoluta del tiempo en estado crítico de 5,3 puntos porcentuales (del 17,7 % al 12,4 %) sin aumentar los recursos operacionales. Se identifican tres ventanas horarias óptimas de intervención (5–6 h, 12 h, 16 h) .

5.2. Aportaciones del trabajo

La principal aportación del trabajo es la integración en un único pipeline de cinco técnicas analíticas complementarias descomposición STL, clustering DTW, detección de anomalías IF ∩ LOF, modelo predictivo LightGBM con explicabilidad SHAP y referencia univariada SARIMA sobre datos reales de un sistema productivo, manteniendo la coherencia metodológica entre fases y traduciendo cada resultado en términos operacionalmente significativos para el operador.

Como aportaciones específicas que distinguen el trabajo respecto a la literatura previa cabe destacar las siguientes.

Primera aplicación pública de clustering DTW sobre BiciMAD. Los enfoques de segmentación de estaciones BSS habituales en la literatura emplean clustering basado en features estáticas (medias diarias, medianas, ratios de uso) que no capturan la forma del perfil temporal a lo largo del día. La aplicación de DTW + k-means con tslearn revela la dualidad funcional ORIGEN–DESTINO de forma estadísticamente robusta y operacionalmente accionable, identificando además los 116 nodos de comportamiento bidireccional como una tercera categoría latente.

Validación temporal rigurosa con TimeSeriesSplit. La adopción de TSCV con tres folds secuenciales en lugar del k-fold aleatorio que todavía se emplea con frecuencia en estudios aplicados sobre series temporales elimina la contaminación temporal del entrenamiento y produce una estimación honesta del rendimiento esperado en producción. Los resultados (R^2 estable en el rango 0,917–0,919) demuestran además que los patrones aprendidos son temporalmente transferibles.

Comparativa empírica SARIMA vs. LightGBM por escala de aplicación. El trabajo proporciona evidencia cuantificada del colapso del modelo univariado a nivel de estación (MAPE de 78–144 % frente al 15 % de LightGBM), reproduciendo y extendiendo el patrón documentado por Hulot et al. (2018) para Vélib París. Esta comparativa, ausente en los trabajos previos sobre BiciMAD, constituye la justificación empírica del enfoque multivariado.

Cuantificación operacional explícita del potencial de mejora. La traducción del rendimiento del modelo en eventos críticos evitables (661.521 observaciones de criticidad evitables sobre 70 días, equivalentes a unas 55.000 estación-horas y a una reducción absoluta del 5,3 pp en el tiempo en estado crítico) y la identificación de ventanas óptimas de intervención por gradiente de ocupación constituyen una aportación práctica directamente accionable por la EMT, no presente en trabajos académicos previos sobre el sistema.

Reproducibilidad. Todo el pipeline se basa en datos públicos de la API GBFS y librerías de código abierto (pandas, scikit-learn, tslearn, statsmodels, lightgbm, shap), con configuración del entorno documentada en el Anexo D. Esto facilita tanto la replicación académica como la adopción industrial por parte del operador.

5.3. Limitaciones del estudio

La interpretación de los resultados debe enmarcarse en un conjunto de limitaciones explícitas que delimitan el alcance del trabajo y orientan las extensiones futuras.

Período temporal limitado. Los 70 días de muestreo (18/11/2025 – 26/01/2026) cubren laborables, fines de semana, festivos y el período navideño, pero no incluyen primavera ni

verano, donde los patrones de uso pueden ser significativamente distintos por condiciones meteorológicas y estacionalidad turística. Los modelos entrenados sobre este período pueden requerir reentrenamiento o calibración antes de su despliegue en otras épocas del año.

Ausencia de datos de viajes. El estándar GBFS station_status.json proporciona snapshots de ocupación pero no eventos individuales de viaje (origen, destino, duración, usuario). Esta limitación impide separar la demanda de la disponibilidad: una estación VACÍA puede serlo por demanda alta o por flujo de salida sostenido sin reposición. Aunque la aproximación a través del tiempo en estado VACÍA es válida para el problema de rebalanceo, la incorporación de datos de viajes permitiría modelar la demanda como variable separada.

No incorporación de variables exógenas. El modelo predictivo no incluye datos meteorológicos (temperatura, precipitación, viento), de eventos especiales (conciertos, partidos, manifestaciones) ni de incidencias del transporte público (huelgas de metro, cortes de líneas). Esta exclusión es deliberada para demostrar el rendimiento alcanzable solo con datos internos del sistema, pero limita el techo predictivo: los días con condiciones meteorológicas extremas o eventos de aforo masivo presentan errores sistemáticamente mayores.

Heterogeneidad residual no modelada. El diagnóstico SARIMA revela autocorrelación residual significativa en el lag 24 (LB(24) p = 0,000), indicando que el ciclo diario de la ocupación media no sigue un patrón estacional perfectamente regular incluso a nivel global. Esta heterogeneidad afecta también al modelo LightGBM, especialmente en estaciones de alta variabilidad como las del distrito de Vallecas.

Horizonte de predicción limitado. El modelo se ha optimizado para horizonte de 1 hora, suficiente para alertas tempranas pero potencialmente justo para algunas decisiones logísticas que requieren tiempos de planificación más largos (por ejemplo, despliegue de vehículos desde el depósito central a estaciones periféricas). La extensión a horizontes de 3–6 horas introduce trade-offs adicionales entre precisión y antelación operativa.

5.4. Futuras líneas de investigación

El pipeline desarrollado abre cinco líneas naturales de extensión que combinan ampliación del alcance metodológico, incorporación de datos externos y traslado de los resultados a un entorno productivo real.

Incorporación de variables meteorológicas y de eventos. La fusión del dataset de ocupación con series horarias de temperatura, precipitación y viento de la AEMET, y con un calendario de eventos masivos (Wizink Center, Bernabéu, Metropolitano, festivales) y de incidencias del transporte público (alertas EMT, Metro de Madrid) permitiría cuantificar el techo de mejora marginal de las variables exógenas sobre el modelo actual. La hipótesis razonable, basada en la literatura comparable, es una reducción adicional del MAE de entre 0,5 y 1,0 pp en la franja de las estaciones más volátiles.

Ampliación del horizonte temporal a 12 meses. Extender el período de muestreo a un año natural completo permitiría modelar la estacionalidad anual (primavera–verano frente a otoño–invierno) y comparar la robustez del modelo entre regímenes meteorológicos distintos. Esta extensión es de coste operativo bajo (continuación del polling automático) y de elevado

valor científico al permitir validar la transferibilidad temporal del modelo más allá del trimestre invernal estudiado.

Modelado mediante grafos espacio-temporales. El sistema BiciMAD admite una representación natural como grafo donde los nodos son las 633 estaciones y las aristas son las correlaciones de ocupación entre pares. La aplicación de Graph Neural Networks (en línea con el trabajo de Yao et al. (2019) sobre CitiBike NYC con LSTM/GCN) permitiría modelar explícitamente las dependencias espaciales entre estaciones próximas o conectadas por flujos similares, capturando información que el modelo tabular actual solo recoge a través de las coordenadas y la pertenencia a clúster.

Política dinámica de despliegue mediante aprendizaje por refuerzo. El problema de rebalanceo puede formularse como un problema de aprendizaje por refuerzo (Reinforcement Learning) donde el agente decide en cada momento qué bicicletas mover entre qué estaciones, recibiendo como recompensa la reducción de eventos críticos. Las predicciones del modelo LightGBM actual proporcionan el modelo del entorno necesario para entrenar el agente. Esta extensión transformaría el pipeline de una herramienta predictiva en una herramienta prescriptiva, cerrando el ciclo de toma de decisiones operacional.

A/B testing operacional con la EMT. La validación última de un pipeline predictivo es su impacto medible en producción. Una colaboración con la EMT permitiría diseñar un experimento controlado en el que se comparen, durante un período de varias semanas, las métricas operacionales (eventos críticos por estación, tiempo medio en estado no operativo, satisfacción del usuario medida por incidencias) entre un grupo de estaciones gestionado con el rebalanceo predictivo y un grupo de control con el procedimiento reactivo actual. Este experimento sería la culminación natural del trabajo y proporcionaría la evidencia empírica definitiva sobre la mejora real frente a la mejora teórica cuantificada en este TFM.

Bibliografía

Borgnat, P., Abry, P., Flandrin, P., Robardet, C., Rouquier, J. B., & Fleury, E. (2011). Shared bicycles in a city: A signal processing and data analysis perspective. Advances in Complex Systems, 14(03), 415–438.

Breunig, M. M., Kriegel, H. P., Ng, R. T., & Sander, J. (2000). LOF: identifying density-based local outliers. ACM SIGMOD Record, 29(2), 93–104.

Caggiani, L., Camporeale, R., Ottomanelli, M., & Szeto, W. Y. (2018). A modeling framework for the dynamic management of free-floating bike-sharing systems. Transportation Research Part C: Emerging Technologies, 87, 159–182.

Chen, T., & Guestrin, C. (2016). XGBoost: A scalable tree boosting system. Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining, 785–794.

Cleveland, R. B., Cleveland, W. S., McRae, J. E., & Terpenning, I. (1990). STL: A seasonal-trend decomposition procedure based on Loess. Journal of Official Statistics, 6(1), 3–73.

EMT Madrid (2025). Portal de Datos Abiertos del Ayuntamiento de Madrid: BiciMAD. https://datos.madrid.es/ (última consulta: enero 2026).

Friedman, J. H. (2001). Greedy function approximation: A gradient boosting machine. The Annals of Statistics, 29(5), 1189–1232.

Hulot, P., Aloise, D., & Jena, S. D. (2018). Towards station-level demand prediction for effective rebalancing in bike-sharing systems. Proceedings of the 24th ACM SIGKDD International Conference on Knowledge Discovery & Data Mining, 378–386.

Kaltenbrunner, A., Meza, R., Grivolla, J., Codina, J., & Banchs, R. (2010). Urban cycles and mobility patterns: Exploring and predicting trends in a bicycle-based public transport system. Pervasive and Mobile Computing, 6(4), 455–466.

Ke, G., Meng, Q., Finley, T., Wang, T., Chen, W., Ma, W., Ye, Q., & Liu, T. Y. (2017). LightGBM: A highly efficient gradient boosting decision tree. Advances in Neural Information Processing Systems, 30, 3146–3154.

Liu, F. T., Ting, K. M., & Zhou, Z. H. (2008). Isolation Forest. Proceedings of the 8th IEEE International Conference on Data Mining, 413–422.

Lundberg, S. M., & Lee, S. I. (2017). A unified approach to interpreting model predictions. Advances in Neural Information Processing Systems, 30, 4765–4774.

Lundberg, S. M., Erion, G., Chen, H., DeGrave, A., Prutkin, J. M., Nair, B., Katz, R., Himmelfarb, J., Bansal, N., & Lee, S. I. (2020). From local explanations to global understanding with explainable AI for trees. Nature Machine Intelligence, 2(1), 56–67.

MobilityData (2024). General Bikeshare Feed Specification (GBFS), v3.0. https://gbfs.org/specification/reference/

Petitjean, F., Ketterlin, A., & Gançarski, P. (2011). A global averaging method for dynamic time warping, with applications to clustering. Pattern Recognition, 44(3), 678–693.

Sakoe, H., & Chiba, S. (1978). Dynamic programming algorithm optimization for spoken word recognition. IEEE Transactions on Acoustics, Speech, and Signal Processing, 26(1), 43–49.

Tavenard, R., Faouzi, J., Vandewiele, G., Divo, F., Androz, G., Holtz, C., Payne, M., Yurchak, R., Rußwurm, M., Kolar, K., & Woods, E. (2020). Tslearn, a machine learning toolkit for time series data. Journal of Machine Learning Research, 21(118), 1–6.

Van der Maaten, L., & Hinton, G. (2008). Visualizing data using t-SNE. Journal of Machine Learning Research, 9(86), 2579–2605.

Yao, H., Tang, X., Wei, H., Zheng, G., & Li, Z. (2019). Revisiting spatial-temporal similarity: A deep learning framework for traffic prediction. Proceedings of the AAAI Conference on Artificial Intelligence, 33(01), 5668–5675.