



UNIVERSITAT^{DE}
BARCELONA



Institut de
Formació Contínua-IL3
UNIVERSITAT DE BARCELONA

Máster en Big Data & Data Science

Trabajo Final de Máster (TFM)

Grupo 10

Análisis del sistema de bicicletas públicas MiBici mediante clusterización geoespacial

Autores:

Junjie Zhou

Sebastián Pazos

Nieves Quispe

Leonardo Aguilar

Tutor:

Ferran Arroyo Vendrell

Barcelona, 2024

© Todos los derechos reservados.

Índice

| | |
|---|-----------|
| Índice | 2 |
| 1. Resumen ejecutivo | 3 |
| 2. Objetivos | 4 |
| 2.1 Objetivo principal | 4 |
| 2.2 Antecedentes | 4 |
| 2.3 Punto de partida: contexto actual | 5 |
| 2.4 Stakeholders | 5 |
| 2.5 DAFO de la situación | 6 |
| 3. Desarrollo | 8 |
| 3.1 Gestión del equipo de trabajo..... | 8 |
| 3.2 Fuentes de información y benchmark | 8 |
| 3.3 Timing del proyecto | 9 |
| 3.3.1 Fases de desarrollo | 9 |
| 3.3.2 Análisis económico y payback | 10 |
| 3.3.3 Cronograma de actividades a realizar | 10 |
| 3.4 Hipótesis y planteamientos realizados | 11 |
| 3.5 Desarrollo y programación usada para realizar el objetivo marcado | 12 |
| 3.5.1 Pre-procesado | 12 |
| 3.5.2 División de estaciones por clústers | 12 |
| 3.6 Comparativa entre métodos de clustering tradicionales y modelos no supervisados de machine learning. | 15 |
| 4. Conclusiones | 16 |
| 4.1 Outputs..... | 16 |
| 4.1.1 Análisis de Movilidad Intercluster | 16 |
| 4.1.1.1 Flujo de viajes entre clusters | 16 |
| 4.1.1.2 Densidad de viajes por cluster por hora..... | 18 |
| 4.1.1.3 Densidad de viajes divididos por mañana y noche..... | 21 |
| 4.1.2 Análisis de Movilidad Intracluster: Caso Centro Urbano | 23 |
| 4.1.2.1 División de clusters en el Centro Urbano..... | 23 |
| 4.1.2.2 Distribución de estaciones por cluster | 23 |
| 4.1.2.3 Flujo de viajes entre clusters internos..... | 24 |
| 4.1.2.4 Densidad de viajes por cluster por hora..... | 25 |
| 4.1.2.5 Densidad de viajes divididos por mañana y noche..... | 28 |
| 4.2 Aplicación real..... | 30 |
| 4.3 Visualización de resumen..... | 31 |
| 4.4. Reflexión final | 31 |
| 5. Bibliografía y glosario de términos utilizados | 33 |
| 6. Anexos | 35 |
| 6.1 Diagrama de Sankey | 35 |
| 6.2 Métodos de clusterización | 36 |
| 6.3 Análisis descriptivo de viajes..... | 38 |

1. Resumen ejecutivo

MiBici es un sistema público de bicicletas compartidas que opera en los municipios de Guadalajara, Zapopan y Tlaquepaque. Su objetivo principal es ofrecer una alternativa de transporte sostenible, económica y eficiente, enfocada en reducir la congestión vehicular, mejorar la calidad del aire, y fomentar un estilo de vida saludable entre los ciudadanos.

El análisis geoespacial del sistema de bicicletas públicas MiBici en Guadalajara ha permitido identificar patrones clave de uso que revelan oportunidades para optimizar el servicio y aumentar su eficiencia operativa. Utilizando técnicas avanzadas de clusterización, se segmentaron las estaciones de bicicleta en diferentes zonas de la ciudad, destacando el *Centro Urbano* como el principal núcleo de movilidad. Más del 80% de los viajes se originan o terminan en este cluster, lo que evidencia la necesidad de priorizar la redistribución de bicicletas en esta área durante las horas pico. Los clusters periféricos, como *Tlaquepaque* y *Zapopan*, muestran una menor demanda, lo que sugiere la implementación de tarifas diferenciadas y promociones específicas para incentivar su uso en estos sectores.

Metodológicamente, el trabajo empleó técnicas de *clusterización jerárquica*, luego de probar sin éxito con otros métodos como *K-means* y *DBSCAN*. Se eligió la clusterización jerárquica por su capacidad de capturar mejor las relaciones espaciales entre las estaciones en áreas densamente interconectadas, como el *Centro Urbano*. A partir de los datos de viajes y utilizando la métrica de *haversine* para calcular las distancias geográficas, se identificaron grupos que permitieron analizar patrones de movilidad intracluster e intercluster. Estos clusters fueron luego utilizados para analizar el flujo de viajes y proponer mejoras específicas en la redistribución de bicicletas y la planificación operativa.

Además, se recomienda una redistribución dinámica de bicicletas que se ajuste a la demanda en tiempo real, complementada con mejoras en la infraestructura de estaciones en zonas de menor uso, como ciclovías y mejor iluminación. Esto, junto con campañas de marketing dirigidas a zonas y horarios de menor tráfico, podría equilibrar mejor la utilización del sistema. Paralelamente, las alianzas con empresas y centros educativos en clusters de alta actividad, promoviendo suscripciones grupales, garantizarían un flujo constante de usuarios durante las horas laborales.

En resumen, los ajustes en la redistribución de bicicletas, junto con mejoras en infraestructura y estrategias de promoción, permitirán a MiBici maximizar su eficiencia y sostenibilidad, consolidando su papel como un componente clave en la movilidad urbana de Guadalajara. Estos cambios no solo mejorarán la satisfacción de los usuarios actuales, sino que también incentivarán el uso en áreas y horarios de menor demanda.

2. Objetivos

2.1 Objetivo principal

Analizar el comportamiento de los usuarios del sistema de bicicletas públicas MiBici en Guadalajara, México, mediante la clusterización geoespacial, con el fin de identificar patrones de uso diferenciados entre días hábiles y no hábiles, así como entre distintas zonas de la ciudad y perfiles de usuarios (trabajo, turismo, compras), para optimizar la distribución de bicicletas y adaptar el servicio a las necesidades específicas de cada tipo de usuario.

2.2 Antecedentes

MiBici cuenta con estaciones distribuidas estratégicamente por la ciudad. Los usuarios pueden alquilar bicicletas a través de dos modalidades: suscripción anual o alquiler temporal. Las bicicletas deben devolverse a una estación después de 30 minutos de uso para evitar cargos adicionales. El sistema está disponible 24/7 para el anclaje de bicicletas, y los usuarios pueden liberarlas entre las 5:00 a.m. y las 00:59 a.m.

La creación de MiBici responde a la necesidad de ofrecer una opción de transporte urbano más verde, reducir la dependencia del automóvil, y crear conciencia sobre los beneficios del uso de la bicicleta para la salud y el medio ambiente. También se busca reducir el tráfico en una ciudad en rápido crecimiento y mejorar la movilidad en zonas de alta congestión.

El sistema comenzó en 2014 y ha experimentado una expansión considerable desde entonces. Actualmente, MiBici cuenta con aproximadamente 300 estaciones y 3,200 bicicletas en los tres municipios metropolitanos mencionados.

Crecimiento de los sistemas de bicicletas compartidas a nivel global

En las últimas dos décadas, los sistemas de bicicletas compartidas han proliferado en muchas ciudades del mundo como una respuesta a la necesidad de movilidad sostenible. Grandes urbes como París, Nueva York y Londres han adoptado este modelo como una forma de reducir la congestión vehicular, disminuir la contaminación y promover un estilo de vida saludable.

Guadalajara, México, se unió a esta tendencia con la creación de MiBici, un sistema que ha contribuido a la mejora de la infraestructura ciclista y la movilidad urbana en la ciudad.

Uso de técnicas de análisis de datos en la optimización de sistemas de bicicletas

Los avances en ciencia de datos y machine learning han permitido un análisis más detallado del comportamiento de los usuarios de bicicletas compartidas. A manera de ejemplo, un estudio reciente aplicó técnicas de clusterización al sistema City Bike en New York City para analizar el comportamiento de los usuarios y optimizar la operación del sistema. Este trabajo identificó patrones clave de uso relacionados con la ubicación geográfica de las estaciones y el tipo de día, proporcionando insights valiosos sobre cómo se utilizan las bicicletas en distintas zonas de la ciudad.

2.3 Punto de partida: contexto actual

MiBici opera actualmente con una extensa red de estaciones distribuidas estratégicamente por toda el área metropolitana de Guadalajara, ofreciendo servicios desde 2014 y facilitando millones de viajes cada año. Los datos públicos abarcan desde diciembre de 2014 hasta marzo de 2024, con registros detallados de cada viaje, incluyendo identificadores de estación de origen y destino, fechas y horarios de uso, y perfiles demográficos básicos de los usuarios. Para efectos de esta investigación, se utilizaron los datos del 2022 (cuando las restricciones por el COVID bajaron drásticamente) y 2023.

2.4 Stakeholders

A. Administradores del sistema MiBici

Aunque no están directamente involucrados, los administradores de MiBici podrían ser importantes stakeholders, ya que los hallazgos del proyecto podrían aportarles insights útiles sobre cómo mejorar la distribución de bicicletas, la planificación de nuevas estaciones, o incluso ajustar las tarifas y promociones basadas en el comportamiento del usuario.

B. Gobierno local (Ayuntamiento de Guadalajara)

El gobierno municipal puede estar interesado en los resultados, especialmente en lo que respecta a la mejora de la movilidad urbana, la planificación de infraestructura ciclista y las políticas de transporte sostenible. Los insights derivados del análisis podrían influir en decisiones futuras sobre inversión en ciclovías y estaciones de MiBici.

C. Usuarios del sistema MiBici

Aunque no están directamente involucrados en el análisis, los usuarios actuales y potenciales del sistema son un grupo de interés clave, ya que los resultados pueden conducir a mejoras en el servicio, como una mejor disponibilidad de bicicletas y estaciones, optimización del tiempo de espera y mejor experiencia de usuario.

D. Comercios locales

Negocios ubicados en zonas con alta afluencia de usuarios de bicicletas pueden beneficiarse si la planificación del sistema MiBici optimiza el tráfico de ciclistas en sus áreas. Pueden estar interesados en conocer patrones de uso para colaborar con promociones o servicios especiales para los usuarios de bicicletas.

2.5 DAFO de la situación

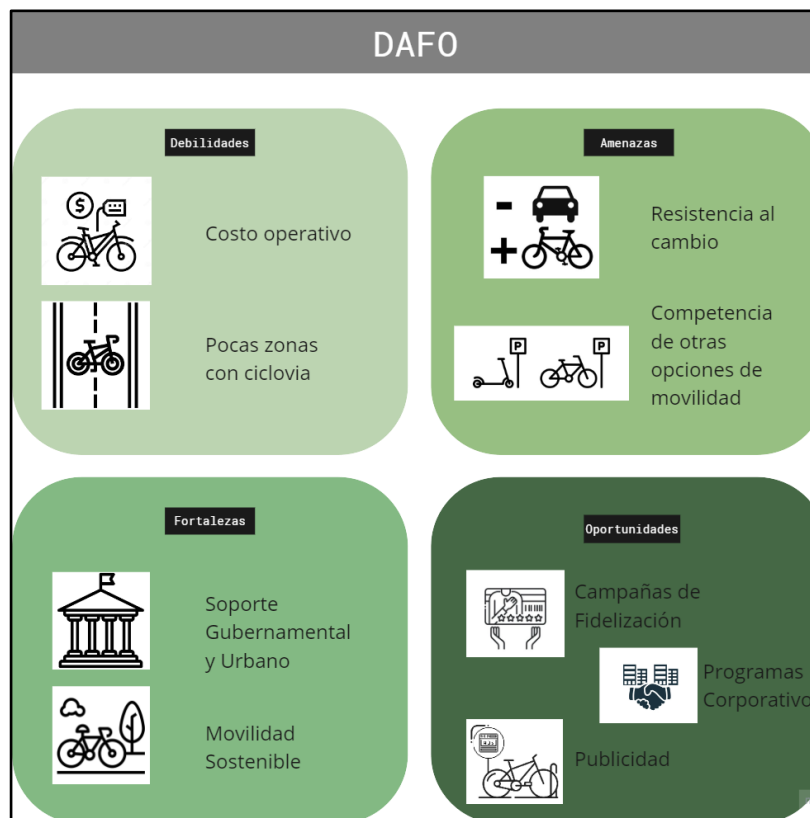


Figura 1: DAFO

Debilidades

Infraestructura limitada en algunas áreas: No todas las zonas de Guadalajara cuentan con ciclovías o infraestructura segura para los ciclistas, lo que limita el atractivo para nuevos usuarios en esas áreas.

Costo operativo elevado: El mantenimiento de las bicicletas y estaciones, así como la reubicación continua de bicicletas, puede generar altos costos operativos.

Amenazas

Resistencia cultural al cambio: Algunas personas pueden estar menos dispuestas a cambiar su medio de transporte tradicional por bicicletas compartidas, especialmente en una cultura orientada al automóvil.

Competencia de otras opciones de movilidad compartida: Servicios de movilidad como scooters y patinetes eléctricos pueden desviar a usuarios que buscan comodidad o velocidad.

Fortalezas

Soporte Gubernamental y Urbano: MiBici ya opera bajo el respaldo del gobierno local, lo que reduce riesgos regulatorios iniciales y genera confianza en su estabilidad.

Modelo de Movilidad Sostenible: MiBici ofrece una alternativa ecológica y saludable, con potencial de crecimiento en ciudades que buscan reducir la congestión y las emisiones.

Oportunidades

Campañas de Fidelización: Utilizar los datos de usuarios actuales para personalizar campañas y generar programas de fidelización (descuentos por uso frecuente, recompensas por referir a otros usuarios).

Programas Corporativos: Asociarse con empresas locales para ofrecer programas de membresía de bicicletas compartidas como un beneficio para empleados.

Publicidad: Ofrecer espacios publicitarios en las estaciones y bicicletas.

3. Desarrollo

3.1 Gestión del equipo de trabajo

| Integrante | Funciones |
|------------------|--|
| Junjie Zhou | <ol style="list-style-type: none">1. Enriquecimiento del dataset2. Trabajo técnico en la clusterización geoespacial.<ol style="list-style-type: none">a. Pruebas con distintos algoritmosb. Desarrollo de código de acuerdo con los objetivos propuestos |
| Sebastián Pazos | <ol style="list-style-type: none">1. Limpieza de datos2. Trabajo técnico en la clusterización geoespacial3. Aplicación del proyecto técnico en un ambiente de negocio |
| Nieves Quispe | <ol style="list-style-type: none">1. Redacción de la documentación, resumen ejecutivo y conclusiones2. Apoyo en la gestión del proyecto, incluyendo la realización del análisis DAFO |
| Leonardo Aguilar | <ol style="list-style-type: none">1. Propuesta del proyecto y selección del dataset2. Redacción de la documentación y conclusiones3. Aplicación del proyecto técnico en un ambiente de negocio |

Tabla 1: Gestión del equipo de trabajo

3.2 Fuentes de información y benchmark

Fuentes de información

El dataset utilizado en el análisis fue proporcionado por MiBici a través de su plataforma de datos abiertos [1]. Decidimos trabajar con los datos de 2022 y 2023, ya que en 2022 las restricciones por la pandemia de COVID-19 disminuyeron significativamente, lo que permitió un análisis más representativo del uso regular del sistema. Además, se utilizó Google Places API [2] para obtener información sobre lugares cercase a las estaciones de bicicletas, lo que se ayudó a identificar patrones de uso relacionados con sitios turísticos, comerciales y laborales.

Benchmark

Como parte del análisis comparativo, se revisaron estudios relevantes sobre sistemas de bicicletas públicas en otras ciudades:

1. City Bike en New York: En el estudio de Lee y Leung [3], se realizó una clusterización espacio-temporal del sistema City Bike utilizando el método Dynamic Time Warping

(DTW). Esta investigación inspiró la segmentación del uso de las bicicletas en Guadalajara según las variaciones geoespaciales y temporales.

2. BikeShare en Toronto: Un proyecto de GitHub [4] aplicó el algoritmo K-Means al sistema de bicicletas en Toronto. Aunque no se utilizó K-Means, este análisis brindó una referencia útil para estructurar el proceso de clusterización en diferentes zonas de la ciudad.

3.3 Timing del proyecto

3.3.1 Fases de desarrollo

A. Definición del proyecto

- **Responsables:** Todo el equipo
- **Descripción:** En esta fase se plantea el objetivo del proyecto, se define el problema a abordar y se selecciona el dataset adecuado para el análisis, asegurando que cuente con la información necesaria.

B. Preparación y limpieza de datos

- **Responsables:** Sebastián Pazos
- **Descripción:** Se lleva a cabo la limpieza del dataset, eliminando inconsistencias, valores nulos o erróneos, y preparando los datos para su posterior análisis. Aquí se asegura que los datos estén listos para el proceso de clusterización.

C. Enriquecimiento de datos

- **Responsables:** Junjie Zhou
- **Descripción:** En esta fase, se agregan variables adicionales que puedan enriquecer el análisis, proporcionando más detalles relevantes para la clusterización.

D. Clusterización geoespacial y análisis técnico

- **Responsables:** Junjie Zhou, Sebastián Pazos
- **Descripción:** Utilizando el algoritmo jerárquico, se realiza la clusterización geoespacial para identificar patrones de uso de las bicicletas en función de las zonas de la ciudad y los tipos de días. El análisis se centra en extraer insights clave sobre los movimientos y comportamiento de los usuarios.

E. Documentación y conclusiones

- **Responsables:** Nieves Quispe, Leonardo Aguilar
- **Descripción:** En esta fase, se redacta la documentación y las conclusiones del análisis. Además, se enfoca en cómo los resultados pueden aplicarse al contexto de negocio.

F. Creación del video de presentación y defensa del trabajo

- **Responsables:** Todo el equipo

- **Descripción:** Se elabora un video en el que se presenta el trabajo, argumentando y defendiendo las ideas más destacadas del proyecto. Esta fase incluye la síntesis de los hallazgos y la exposición visual y oral del análisis.

3.3.2 Análisis económico y payback

Desde una perspectiva económica, la tarifa de suscripción anual de MiBici en Guadalajara, fijada en 478 pesos mexicanos para 2024, representa una opción asequible en comparación con otros servicios de movilidad compartida y medios de transporte tradicionales. El sistema, que cuenta con más de 360 estaciones y casi 4,000 bicicletas, está orientado a fomentar la movilidad sostenible en la ciudad. Este modelo permite a los usuarios realizar viajes ilimitados de menos de 30 minutos, lo que convierte a MiBici en una alternativa económica frente al transporte público o privado. No obstante, las quejas relacionadas con la falta de bicicletas en horas pico y el estado de las unidades subrayan la importancia de mejorar la infraestructura y el servicio para maximizar el impacto positivo en la demanda.

En cuanto al análisis de payback, además de la economía directa en transporte, la implementación de sistemas como MiBici tiene un impacto positivo en la salud pública. El uso regular de bicicletas contribuye a la reducción de enfermedades cardiovasculares y obesidad, lo que a largo plazo genera un ahorro significativo en costos de atención médica. Investigaciones sugieren que el uso de la bicicleta está asociado con una disminución en la incidencia de enfermedades crónicas, lo que puede traducirse en ahorros en los gastos del sistema de salud de hasta 3,8 euros por cada kilómetro recorrido en bicicleta [7], debido a la menor necesidad de tratamiento de enfermedades relacionadas con el sedentarismo y contaminación. Así, la inversión inicial en la suscripción de MiBici puede recuperarse no sólo por la reducción en costos de transporte, sino también por los beneficios derivados de una mejor salud.

3.3.3 Cronograma de actividades a realizar

| Fase | Actividad | Responsable | Fecha de Inicio | Fecha de finalización |
|------------------------------------|--|-----------------|-----------------|-----------------------|
| 1. Definición del Proyecto | Definir el objetivo, el problema a abordar y seleccionar el dataset adecuado | Todo el equipo | 1-abr | 11-may |
| 2. Preparación y Limpieza de Datos | Limpieza de datos: eliminación de valores nulos, inconsistencias y preparación | Sebastián Pazos | 13-may | 31-may |

| | | | | |
|---|--|---------------------------------|--------|--------|
| 3. Enriquecimiento de Datos | Agregar variables adicionales relevantes para el análisis | Junjie Zhou | 1-jun | 14-jun |
| 4. Clusterización Geoespacial y Análisis Técnico | Aplicar algoritmo de clusterización jerárquico para analizar patrones de movilidad | Junjie Zhou, Sebastián Pazos | 15-jun | 17-jul |
| 5. Documentación y Conclusiones | Redactar la documentación y generar conclusiones basadas en el análisis realizado | Nieves Quispe, Leonardo Aguilar | 26-ago | 12-sep |
| 6. Creación del Video de Presentación y Defensa | Crear un video de síntesis y exposición del proyecto | Todo el equipo | 13-sep | 2-oct |

Tabla 2: Cronograma de actividades

3.4 Hipótesis y planteamientos realizados

Hipótesis

El uso del sistema de bicicletas públicas MiBici en Guadalajara presenta patrones de demanda diferentes entre días hábiles y no hábiles, así como entre distintas zonas geográficas de la ciudad, lo que sugiere la necesidad de ajustar la distribución de bicicletas y mejorar los servicios en áreas con mayor flujo de usuarios laborales y recreativos.

Planteamientos realizados:

1. Selección del mejor algoritmo de clusterización

Se realizó un análisis comparativo de diferentes algoritmos de clusterización, evaluando su capacidad para agrupar de manera efectiva las estaciones de bicicletas según el uso geoespacial. Finalmente, se seleccionó el algoritmo jerárquico debido a su capacidad para reflejar con precisión las estructuras subyacentes en los datos y facilitar la interpretación de los clusters.

2. Enriquecimiento del dataset con lugares de interés

Se integraron datos adicionales al dataset original, incluyendo centros educativos, lugares de interés turístico y comerciales cercanos a las estaciones de MiBici. Esto permitió describir de manera más detallada el perfil de cada cluster.

3. División en días hábiles y no hábiles:

Los datos fueron segmentados en días hábiles y no hábiles (fin de semana y días festivos) para identificar patrones diferenciados de uso. Esto facilita el análisis de cómo varía la demanda del sistema según el tipo de día.

4. Análisis en horas pico de la mañana y noche:

Se realizaron análisis específicos en las horas pico de la mañana y la noche para entender el comportamiento de los usuarios que utilizan el sistema en horarios laborales.

3.5 Desarrollo y programación usada para realizar el objetivo marcado

3.5.1 Pre-procesado

En el análisis descriptivo de los viajes en MiBici, se observan patrones claros de demanda según el tipo de día y la hora. Durante los días hábiles, los viajes son significativamente mayores en comparación con los fines de semana o días festivos, lo que sugiere un uso predominantemente laboral o educativo. Los días no hábiles y festivos presentan una demanda similar, razón por la cual se agrupan en una categoría común denominada "No Hábil". En cuanto a la duración de los viajes, en los días no hábiles se registran viajes ligeramente más largos, lo que podría reflejar un uso recreativo o turístico. Sin embargo, tanto en días hábiles como no hábiles, la mayoría de los viajes tienen una duración entre 5 y 10 minutos.

El análisis intradía revela variaciones importantes en la demanda a lo largo del día, especialmente en función del tipo de jornada. Durante los días hábiles, la franja de mayor uso es la tarde, con una demanda cercana a los 2 millones de viajes, seguida por la mañana pre-trabajo. Las franjas intermedias, como la mañana post trabajo y el almuerzo, presentan una demanda menor, lo que podría clasificarse como horarios valle. En los días no hábiles, la demanda se distribuye de manera más uniforme, alcanzando un pico similar en las franjas de almuerzo, tarde y noche, lo que indica un uso más relajado y no tan ligado a rutinas laborales. Para más detalle revisar anexos en sección 6.3.

3.5.2 División de estaciones por clústers

Uno de los objetivos clave es identificar grupos geográficos mediante la agrupación de estaciones de MiBici según su ubicación. Esto permite detectar zonas con alta densidad de estaciones o patrones geográficos que podrían influir en decisiones estratégicas, tanto operativas como de marketing, optimizando así la entrega del servicio.

Otro aspecto importante es la optimización de los recursos disponibles. Al analizar la demanda y la proximidad geográfica, es posible definir cómo distribuir de manera más eficiente los recursos, como bicicletas y estaciones de mantenimiento, asegurando que estén alineados con las necesidades del sistema.

Mejorar la accesibilidad también es una prioridad. Al comprender la cercanía de las estaciones a puntos de interés, es factible aumentar la conveniencia para los usuarios, lo que a su vez puede impulsar el uso del servicio y mejorar la experiencia del cliente.

Finalmente, es fundamental integrar el análisis del estado operativo de las estaciones. Evaluar su estado ayuda a entender cómo afecta la accesibilidad y el uso general de la red de bicicletas, permitiendo hacer ajustes que optimicen el rendimiento del sistema.

Para lograr los objetivos se evaluó clusterizar mediante tres métodos:

- K-means
- DBSCAN
- Jerárquico

En el análisis de K-Means, el método del codo indicó que **k=4** o **k=5** era el número óptimo de clústeres, ya que después de ese punto, la inercia disminuía marginalmente. Esto mostró que 4 o 5 clústeres eran suficientes para segmentar las estaciones en función de su proximidad geográfica, reflejando un agrupamiento lógico de las estaciones según su ubicación en la ciudad.

El análisis de silueta confirmó que **k=4** proporcionaba el mejor agrupamiento, con el coeficiente de silueta promedio más alto en ese punto. Esto indicó que las estaciones estaban bien agrupadas desde una perspectiva geográfica, sin tener en cuenta otros factores.

En el proceso de clusterización con **DBSCAN**, se experimentó con diferentes valores de los parámetros críticos **eps** (radio de vecindad) y **min_samples** (puntos mínimos en un clúster). Al ajustar el parámetro **eps** a valores menores, como 0.0005 y luego 0.35, se observó que los puntos considerados en la misma vecindad se reducían, lo que dio como resultado un mayor número de clústeres de lo esperado. El ajuste final a un valor de **eps=0.4** equilibró este comportamiento, logrando una segmentación más coherente con la estructura de los datos. Por otro lado, la reducción del parámetro **min_samples** a 3 permitió la formación de clústeres más pequeños, lo que ayudó a identificar agrupaciones más localizadas en los datos. El uso de la métrica **haversine**, al ser adecuada para datos de latitud y longitud, fue clave para medir correctamente las distancias geográficas en la superficie terrestre.

Sin embargo, a pesar de estos ajustes, el uso de DBSCAN mostró limitaciones en la identificación precisa de clústeres debido a la proximidad de las estaciones. Al tener todas las estaciones relativamente cercanas entre sí, el uso de un **eps** pequeño no aportó mejoras significativas en los resultados, sugiriendo que el algoritmo puede no ser el más adecuado para estos datos. Por lo tanto, se plantea que el **clustering jerárquico** sería más apropiado en este caso, ya que ofrece una visualización clara de los niveles de agregación y permite elegir de manera más flexible el número de clústeres tras la ejecución del algoritmo. Este enfoque podría ser más efectivo para capturar la estructura geográfica subyacente en los

datos de las estaciones, especialmente cuando las agrupaciones son densas o están muy próximas.

Caracterización

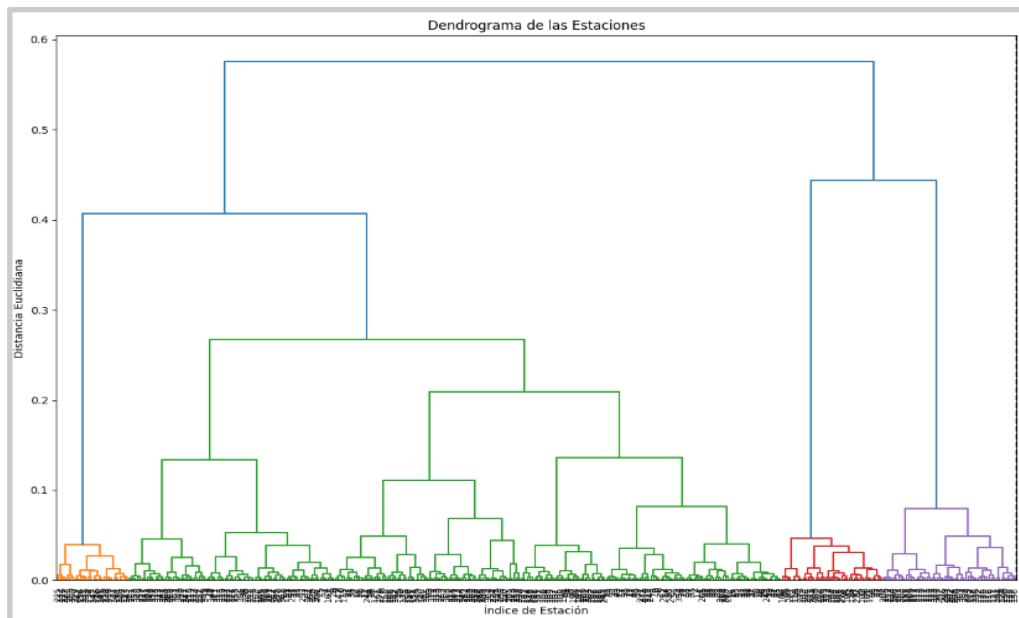


Figura 2: diagrama de dendograma

El dendrograma está compuesto por 4 clusters los cuales según la ubicación están nombrados de la siguiente manera:

Centro Urbano: Verde

Chapalita: Púrpura

Zapopan: Rojo

Tlaquepaque: Naranja

Según los clusters identificados **Centro urbano** es el centro de la ciudad, **Tlaquepaque** es un municipio al sureste de Guadalajara, **Zapopan** es otro municipio al noroeste de Guadalajara y **Chapalita** es una colonia (barrio) ubicado entre el municipio de Guadalajara y Zapopan.



Figura 3: Mapa de estaciones por cluster

3.6 Comparativa entre métodos de clustering tradicionales y modelos no supervisados de machine learning.

En la evaluación de métodos de clustering para la segmentación de estaciones de MiBici, enfrentamos desafíos específicos con las técnicas de **K-means** y **DBSCAN**, debido a la proximidad de las estaciones y la distribución geográfica de las mismas. K-means no logró una separación clara, especialmente en el área central de la ciudad, mientras que DBSCAN demostró ser demasiado sensible a las distancias cortas entre estaciones, generando una segmentación no deseada debido a la cercanía de aproximadamente 400 metros entre ellas.

Finalmente, **se optó por un enfoque de clustering jerárquico (clustering tradicional)**, que proporcionó una visualización más clara y flexible de las agrupaciones y se adaptó mejor a la densidad y proximidad de las estaciones. Para el análisis de clustering jerárquico, **inicialmente utilizó la distancia euclidiana**. Esta métrica calcula la distancia lineal más corta entre dos puntos en un plano, basada en el teorema de Pitágoras. Es adecuada para espacios donde las dimensiones tienen el mismo tipo de unidad y donde las diferencias en cada dimensión son comparables directamente, como es el caso en ciertos análisis de datos multidimensionales en espacios pequeños o de media escala.

Sin embargo, dado que las estaciones están distribuidas a lo largo de una región geográfica y que la precisión en la medición de distancias es crucial, se consideró esta aproximación. Para distancias geográficas, que implican puntos distribuidos sobre la superficie curva de la Tierra, **la métrica haversine resultó ser la más adecuada**. Esta métrica considera la curvatura de la Tierra, utilizando el radio terrestre para calcular el arco mínimo entre dos puntos dados sus coordenadas de latitud y longitud. Así, se garantiza una mayor precisión en el cálculo de distancias, especialmente útil para la planificación de rutas y la optimización de recursos en un sistema de transporte extendido como el de las estaciones de bicicletas compartidas.

4. Conclusiones

4.1 Outputs

4.1.1 Análisis de Movilidad Intercluster

4.1.1.1 Flujo de viajes entre clusters

A continuación se presenta la distribución de todos los viajes realizados entre los clusters:

Porcentaje de viajes

| nom_cluster_hc_destino | CENTRO URBANO | CHAPALITA | TLAQUEPAQUE | ZAPOPAN |
|------------------------|---------------|-----------|-------------|---------|
| nom_cluster_hc_origen | | | | |
| CENTRO URBANO | 84.55% | 1.23% | 1.03% | 0.45% |
| CHAPALITA | 2.47% | 2.68% | 0.00% | 0.00% |
| TLAQUEPAQUE | 1.29% | 0.00% | 1.75% | 0.00% |
| ZAPOPAN | 0.71% | 0.00% | 0.00% | 3.85% |

Figura 4: matriz de porcentaje de viajes

Porcentaje de viajes desde cluster origen

Cada fila suma 100%.

| nom_cluster_hc_destino | CENTRO URBANO | CHAPALITA | TLAQUEPAQUE | ZAPOPAN |
|------------------------|---------------|-----------|-------------|---------|
| nom_cluster_hc_origen | | | | |
| CENTRO URBANO | 96.90% | 1.40% | 1.18% | 0.52% |
| CHAPALITA | 47.93% | 52.04% | 0.01% | 0.01% |
| TLAQUEPAQUE | 42.40% | 0.01% | 57.59% | 0.00% |
| ZAPOPAN | 15.63% | 0.02% | 0.00% | 84.35% |

Figura 5: matriz de porcentaje de viajes desde cluster origen

De estos gráficos se puede desprender lo siguiente:

1. **El Centro Urbano de Guadalajara es el principal hub de movilidad:** La mayoría de los flujos de bicicletas, tanto de entrada como de salida, están concentrados en este cluster; 6 446 270 viajes inician ahí y 6 760 391 terminan ahí. Esta centralidad indica una alta dependencia de este cluster para la movilidad general en la ciudad, lo cual es consistente con la función de los centros urbanos como núcleos de actividad económica y social.
2. **Desigual distribución de flujos entre clusters:** Se observa que los flujos entre los clusters no son homogéneos. Mientras que hay conexiones significativas entre el Centro Urbano y los otros clusters (Chapalita, Zapopan, y Tlaquepaque), las conexiones directas entre estos clusters periféricos son menos pronunciadas. Esto sugiere que la mayoría de los viajes se concentran en movimientos entre el Centro

Urbano y los clusters periféricos, con una menor interconexión directa entre estos últimos. Este patrón es indicativo de un modelo de movilidad centrado en el núcleo urbano, donde los usuarios prefieren viajar hacia y desde el centro en lugar de entre los clusters periféricos.

3. **Patrones de movilidad en clusters periféricos:** Los clusters de Chapalita, Zapopan, y Tlaquepaque muestran una menor densidad de flujos en comparación con el Centro Urbano. Además, los flujos que parten de estos clusters parecen ser más unidireccionales hacia el Centro Urbano, con menos viajes de retorno. Esto podría implicar que estos clusters funcionan más como zonas residenciales o secundarias, con los usuarios utilizando las bicicletas principalmente para desplazarse hacia el centro por razones laborales o recreativas, y menos para desplazamientos dentro de estos clusters periféricos.
4. **Poca interacción directa entre los clusters periféricos (Chapalita, Tlaquepaque, y Zapopan).** La gran mayoría de los viajes entre estos clusters son insignificantes en términos absolutos y relativos. Por ejemplo, solo 4 viajes fueron registrados desde Tlaquepaque hacia Zapopan, y 3 desde Zapopan a Tlaquepaque, siendo la relación entre estos dos clusters la de menor flujo. Esto indica que los usuarios no suelen desplazarse directamente entre los clusters periféricos, lo cual sugiere que la infraestructura o la demanda no promueve la conectividad entre estos puntos.
5. **La mayoría de los viajes tienden a ser intracluster, es decir, tienen su origen y destino en el mismo cluster.** Llama la atención el caso de Zapopan, donde el 84.35% de viajes tiene esta característica, solo el 15.63% de los viajes se dirige hasta el Centro Urbano. Algunas justificaciones para este comportamiento son:
 - a. Zapopan es hogar de algunos de los centros comerciales más grandes y populares de la región, como Plaza Andares y Plaza Galerías, lo que lo convierte en un importante centro comercial y de entretenimiento dentro de la Zona Metropolitana de Guadalajara [5].
 - b. También alberga parques tecnológicos y universidades, como el Tec de Monterrey, Campus Guadalajara, y centros de investigación, lo que contribuye a un perfil urbano y de negocios destacado.
 - c. El transporte público, incluidas rutas de autobús y la Línea 3 del Tren Ligero, conecta Zapopan con otras áreas de Guadalajara, facilitando el tránsito y el flujo económico en la región [5].
 - d. Zapopan ha visto un rápido crecimiento residencial, especialmente en áreas como Valle Real y Santa Margarita, donde el desarrollo de fraccionamientos y edificios de departamentos ha sido notable. Este crecimiento responde tanto a la demanda de vivienda como a la expansión de la clase media y alta en la ZMG [6].

En el siguiente gráfico se destaca como las cinco rutas intercluster con mayor número de viajes involucran el Centro Urbano.



Figura 6: Mapa con rutas intercluster

4.1.1.2 Densidad de viajes por cluster por hora

Viajes iniciados por hora

Se graficó el flujo por hora para cada clúster, diferenciando la densidad de viajes de los días hábiles y no hábiles. Los días hábiles se refieren a los días de lunes a viernes que no sean feriados. Los no hábiles incluyen sábados y domingos, así como los feriados (tanto nacionales como los locales de Guadalajara).

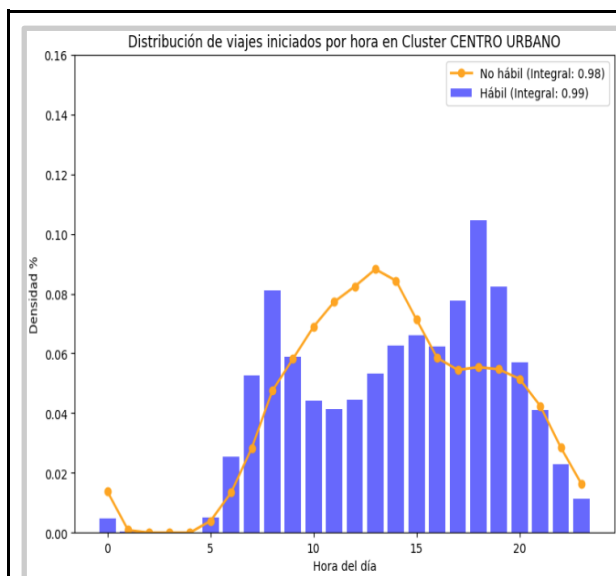


Figura 7: Viajes iniciados por hora en el Centro Urbano

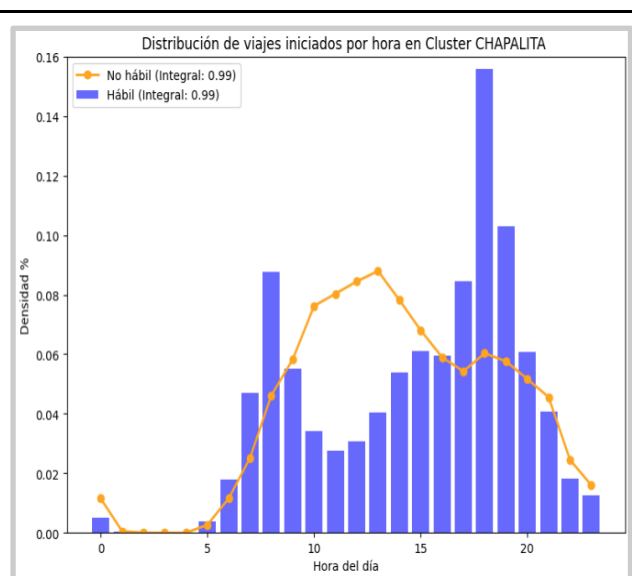


Figura 8: Viajes iniciados por hora en Chapalita

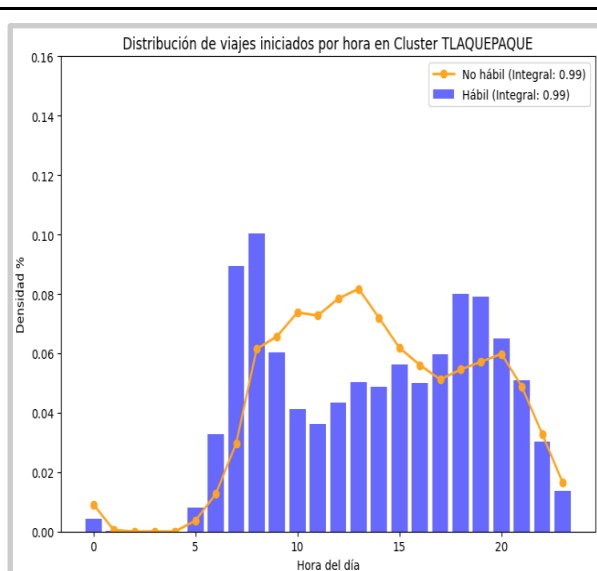


Figura 9: Viajes iniciados por hora en Tlaquepaque

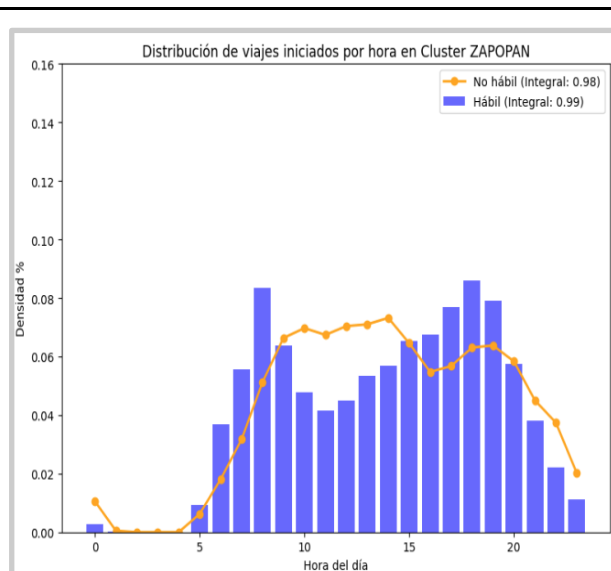


Figura 10: Viajes iniciados por hora en Zapopan

Los gráficos muestran algunas generalidades:

- Hay un pico notable en la mañana, alrededor de las 8-9 am, y otro pico por la tarde, alrededor de las 6-7 pm. Esto sugiere que el uso de bicicletas en días hábiles está fuertemente relacionado con los horarios de trabajo, con personas usando las bicicletas para desplazarse al trabajo y regresar a casa.
- En días no hábiles, la distribución es más uniforme a lo largo del día, con un aumento gradual que alcanza su máximo alrededor de las 12-1 pm y un descenso hacia la noche. Esto podría indicar un uso más recreativo o relacionado con actividades de ocio, en lugar de desplazamientos al trabajo.

También, se presentan algunos comportamientos específicos relacionados a los días hábiles.

- En los clusters **Centro Urbano** y **Chapalita** se utilizan más bicis para la salida de los trabajos en comparación con la entrada, lo que sugiere que los trabajadores prefieren llegar en medios alternativos ya que posiblemente no quieran llegar sudados a sus labores. Además, es entendible que opten por usar la bicicleta al salir cuando tienen más tiempo y la jornada laboral ha terminado.
- En **Tlaquepaque** hay un mayor pico en las horas de la mañana, a la entrada de los trabajos, que en la tarde y noche. Tlaquepaque es en gran parte una zona residencial. Es probable que muchas personas vivan en Tlaquepaque pero trabajan en otras áreas de la ciudad, como el Centro Urbano. Esto explicaría un uso intensivo de bicicletas por la mañana, cuando las personas se desplazan desde sus hogares hacia sus lugares de trabajo. Para la tarde, parece que la preferencia es optar por otros medios de transporte. Adicionalmente, Tlaquepaque, al ser más residencial, podría tener menos actividad comercial o de ocio en la tarde y noche en comparación con los otros cluster, lo que reduciría la necesidad de usar bicicletas en esos horarios.

- En **Zapopan**, el comportamiento es similar a lo que ocurre con el Centro Urbano y Chapalita. Sin embargo, los picos son más nivelados. El flujo a la hora de entrada a los trabajos es comparable a las horas de salida.

Viajes finalizados por hora

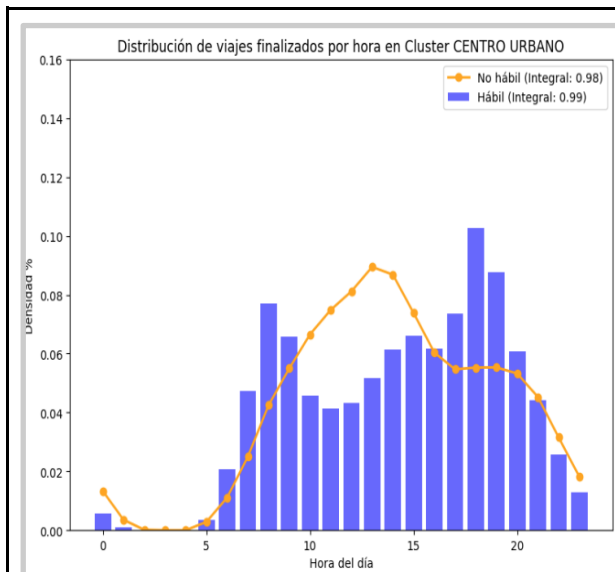


Figura 11: viajes finalizados por hora en Centro Urbano

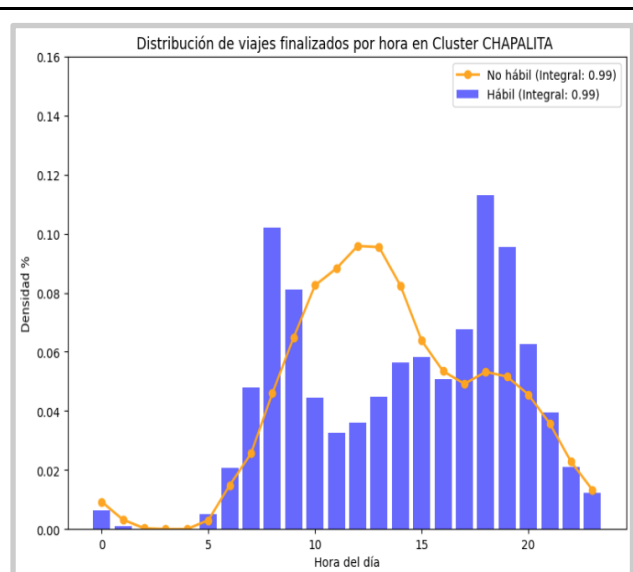


Figura 12: viajes finalizados por hora en Chapalita

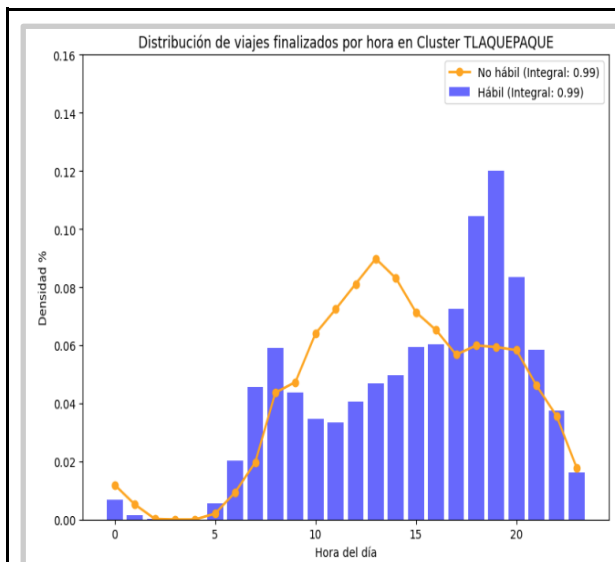


Figura 13: viajes finalizados por hora en Tlaquepaque

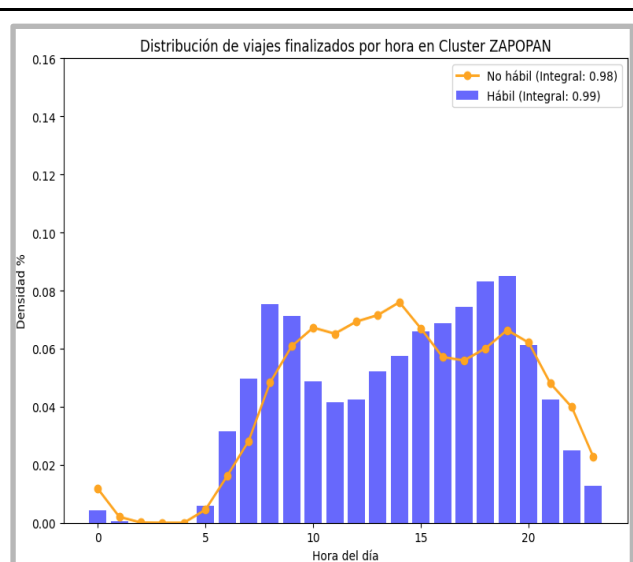


Figura 14: viajes finalizados por hora en Zapopan

Con base en los gráficos anteriores, se pueden extraer conclusiones:

- En el caso de **Centro Urbano** y **Chapalita**, hay dos picos claros en días hábiles: uno en la mañana (alrededor de las 8-9 am) y otro en la tarde (alrededor de las 6-7 pm).

Esto indica que las bicicletas se utilizan principalmente para desplazamientos relacionados con el trabajo. Hay un mayor uso de bicicletas en la tarde en comparación con la mañana, lo que sugiere que los trabajadores prefieren otros medios de transporte para ir al trabajo, pero optan por la bicicleta al finalizar la jornada.

- En el caso de **Tlaquepaque**, hay un mayor pico en las horas de la tarde, a la salida de los trabajos. Tlaquepaque es en gran parte una zona residencial; es probable que muchas personas vivan ahí pero trabajen en otras áreas de la ciudad, como el Centro Urbano. Parece que las personas se desplazan desde sus hogares hacia sus lugares de trabajo por otros medios pero por la tarde, sin la presión de una hora específica ni de no llegar sudado, parece que la preferencia por el uso de la bici incrementa.
- En **Zapopan** los picos de la mañana y la tarde están más nivelados, lo que indica que las personas en este cluster usan las bicicletas de manera más equitativa durante el día.

4.1.1.3 Densidad de viajes divididos por mañana y noche

Mañana

Porcentaje de viajes

| nom_cluster_hc_destino | CENTRO URBANO | CHAPALITA | TLAQUEPAQUE | ZAPOPAN |
|------------------------|---------------|-----------|-------------|---------|
| nom_cluster_hc_origen | | | | |
| CENTRO URBANO | 96.93% | 1.97% | 0.64% | 0.46% |
| CHAPALITA | 40.78% | 59.20% | 0.01% | 0.00% |
| TLAQUEPAQUE | 56.77% | 0.01% | 43.21% | 0.00% |
| ZAPOPAN | 15.94% | 0.03% | 0.00% | 84.03% |

Figura 15: matriz de porcentaje de viajes

Porcentaje de viajes desde cluster origen

| nom_cluster_hc_destino | CENTRO URBANO | CHAPALITA | TLAQUEPAQUE | ZAPOPAN |
|------------------------|---------------|-----------|-------------|---------|
| nom_cluster_hc_origen | | | | |
| CENTRO URBANO | 83.65% | 1.70% | 0.56% | 0.40% |
| CHAPALITA | 1.98% | 2.87% | 0.00% | 0.00% |
| TLAQUEPAQUE | 2.15% | 0.00% | 1.64% | 0.00% |
| ZAPOPAN | 0.81% | 0.00% | 0.00% | 4.24% |

Figura 16: matriz de viajes desde cluster origen

Centro Urbano es el principal con 1,302,940 viajes entrantes y salientes, lo que lo posiciona como el epicentro de la movilidad en la mañana. Representa el 83.65% de los viajes totales, una concentración alta de la movilidad en esta región.

Chapalita como destino es un área con un flujo significativo de usuarios que llegan en la mañana esto puede deberse a que Chapalita es una zona residencial con áreas de ocio como restaurantes y cafés.

Tlaquepaque y Zapopan registran una cantidad baja de viajes, tanto en origen como en destino en la mañana.

Noche

Porcentaje de viajes

| nom_cluster_hc_destino | CENTRO URBANO | CHAPALITA | TLAQUEPAQUE | ZAPOPAN |
|------------------------|---------------|-----------|-------------|---------|
| nom_cluster_hc_origen | | | | |
| CENTRO URBANO | 96.39% | 1.32% | 1.64% | 0.65% |
| CHAPALITA | 55.53% | 44.46% | 0.01% | 0.01% |
| TLAQUEPAQUE | 34.11% | 0.00% | 65.89% | 0.00% |
| ZAPOPAN | 17.81% | 0.01% | 0.00% | 82.18% |

Figura 17: matriz de porcentaje de viajes

Porcentaje de viajes desde cluster origen

| nom_cluster_hc_destino | CENTRO URBANO | CHAPALITA | TLAQUEPAQUE | ZAPOPAN |
|------------------------|---------------|-----------|-------------|---------|
| nom_cluster_hc_origen | | | | |
| CENTRO URBANO | 83.56% | 1.15% | 1.42% | 0.57% |
| CHAPALITA | 3.58% | 2.86% | 0.00% | 0.00% |
| TLAQUEPAQUE | 0.87% | 0.00% | 1.68% | 0.00% |
| ZAPOPAN | 0.77% | 0.00% | 0.00% | 3.55% |

Figura 18: matriz de viajes desde cluster origen

Centro Urbano se incrementa la cantidad de viajes de 1,302,940 a 1,553,188 viajes en la noche. En total concentra 83.56% de viajes por la noche, casi similar a la mañana 83.65 %.

Tlaquepaque se nota un aumento considerable en el número de viajes de origen y destino dentro de su zona, incrementando de 25,531 viajes en la mañana a 31,241 viajes en la noche. Este aumento probablemente sea debido a sus ofertas culturales, artesanales y gastronómicas durante la noche.

Entre **Zapopan** y **Tlaquepaque** tienen muy poca cantidad de viajes, esto probablemente se deba a la distancia que se tiene entre ellos.

Chapalita incrementa de 1,98% en la mañana a 3.58% viajes en la noche con destino al centro urbano probablemente se deba a la cantidad de comercios que se encuentran en el centro urbano.

4.1.2 Análisis de Movilidad Intracluster: Caso Centro Urbano

4.1.2.1 División de clusters en el Centro Urbano

El centro urbano cuenta con el 84% de los viajes realizados por los usuarios, es un porcentaje alto es por ese motivo que vamos a ampliar el análisis hacia el centro urbano y descubrir cómo están divididos los viajes respecto a las estaciones, cantidad de viajes por hora y por turno mañana y noche.

El análisis se realizará bajo la misma lógica de clusterización de la sección anterior mediante un cluster jerárquico definido como “Centro Urbano”.

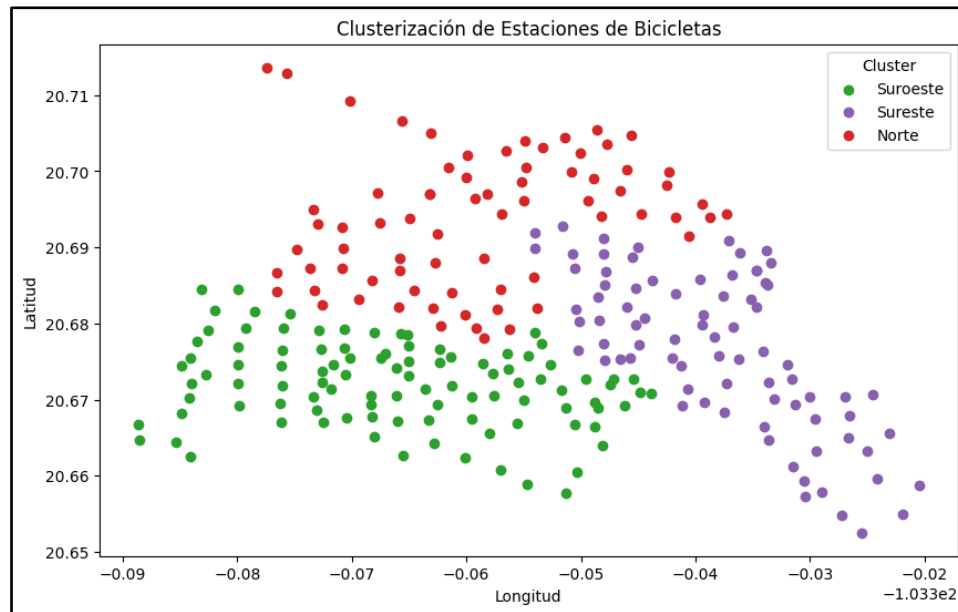


Figura 19: Clusters centro urbano

4.1.2.2 Distribución de estaciones por cluster

| Cluster | En servicio | Fuera de servicio |
|----------|-------------|-------------------|
| Sureste | 78 | 2 |
| Suroeste | 100 | 1 |
| Norte | 70 | 2 |

Tabla 3: Tabla distribución de estaciones por cluster

El centro urbano cuenta con 253 estaciones en total el cual está dividido de la siguiente forma: 101 estaciones pertenecientes al Cluster Suroeste, Cluster Sureste con 80 estaciones y Cluster Norte con 72 estaciones.

Actualmente se encuentran 5 estaciones fuera de servicio, representa el 2% del total de estaciones del centro urbano.

4.1.2.3 Flujo de viajes entre clusters internos

Porcentaje de viajes

| nom_cluster_hc_destino | Norte | Sureste | Suroeste |
|------------------------|--------|---------|----------|
| nom_cluster_hc_origen | | | |
| Norte | 49.76% | 20.41% | 29.83% |
| Sureste | 21.22% | 46.85% | 31.94% |
| Suroeste | 15.92% | 18.28% | 65.80% |

Figura 20: matriz de porcentaje de viajes en el Centro Urbano

Porcentaje de viajes desde cluster origen

Cada fila suma 100%.

| nom_cluster_hc_destino | Norte | Sureste | Suroeste |
|------------------------|--------|---------|----------|
| nom_cluster_hc_origen | | | |
| Norte | 12.69% | 5.20% | 7.60% |
| Sureste | 4.80% | 10.59% | 7.22% |
| Suroeste | 8.26% | 9.49% | 34.16% |

Figura 21: matriz de porcentaje de viajes desde cluster origen en el Centro Urbano

Cómo se compartió previamente, se utilizó el API de Google para ubicar centros de estudios, atracciones turísticas y centros comerciales que, posteriormente, se clasificaron dentro de estos clusters internos. Se buscó ubicar los centros de trabajo, pero no fue posible. Con los datos disponibles se desarrollaron los siguientes gráficos para entender la distribución de estos lugares de interés.

| Cluster | Centros educativos | Atracciones turísticas | Centros comerciales |
|----------|--------------------|------------------------|---------------------|
| Sureste | 70.4% | 14.7% | 14.9% |
| Suroeste | 52.2% | 21.5% | 26.3% |
| Norte | 69.7% | 20.1% | 10.2% |

Tabla 4: tabla de porcentajes de categorías de lugares de interés por cluster

Esta información apunta a las siguientes conclusiones:

- **Suroeste como clave de generación de movilidad.** Más de la mitad de los viajes en el Centro Urbano se originan en el Suroeste, específicamente el 51.91%, y es también el cluster donde más viajes se finalizan en el Centro Urbano. Además, el Suroeste presenta el mayor número de viajes internos con 2 193 081 viajes, lo que representa el 65.8% de los viajes realizados dentro de este clúster. Hay una alta concentración de actividad interna, lo que sugiere que los centros de trabajo y estudio mantienen a los usuarios dentro de la misma zona para la mayoría de sus desplazamientos. El

hecho de que los viajes inicien ahí en su mayoría también es indicador de una zona residencial considerable. Llama la atención que, en proporción, hay muchas más escuelas que centros turísticos y de comercio.

- **Sureste como un centro de turismo y de comercio.** El Sureste muestra una distribución de viajes más equilibrada, con un 46.85% de viajes internos (el menor de los tres), seguido de un 31.94% de viajes hacia el Suroeste y un 21.22% hacia el Norte. Esto sugiere que el Sureste actúa como una zona de tránsito donde personas de distintos lugares llegan a conocer o a comprar y luego se desplazan de vuelta a sus zonas de residencia. Algunos lugares que se encuentran aquí son:
 - Plaza de Armas
 - Palacio de Gobierno
 - Museo del Ejército y Fuerza Aérea Mexicana
 - Museo Regional de Guadalajara
 - Parque Morelos

En la figura # en el apartado 4.4.1 se puede observar como en el cluster Sureste, el mayor pico de viajes iniciados en este cluster se da en días no hábiles, que son aquellos que se relacionan en menor medida con el desplazamientos a centros de trabajo y estudio, y más a actividades de ocio y comerciales.

- **El Norte como destino de trabajo y estudio secundario.** Podría compararse al Suroeste en el sentido de que la proporción de lugares de turismo y centros comerciales es poca en comparación a las escuelas. La menor cantidad de viajes originados en el Suroeste se dirigen al Norte, esto podría tener que ver con el hecho de que se desarrollan actividades similares en ambos cluster.

4.1.2.4 Densidad de viajes por cluster por hora

Se graficó el flujo por hora para cada clúster interno del Centro Urbano, diferenciando la densidad de viajes de los días hábiles y no hábiles. Los días hábiles se refieren a los días de lunes a viernes que no sean feriados. Los no hábiles incluyen sábados y domingos, así como los feriados (tanto nacionales como los locales de Guadalajara).

Viajes iniciados por hora

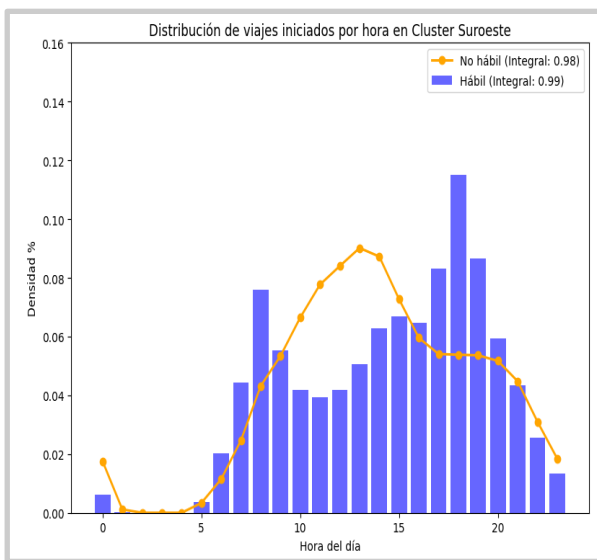


Figura 22: viajes iniciados por hora en el cluster Suroeste

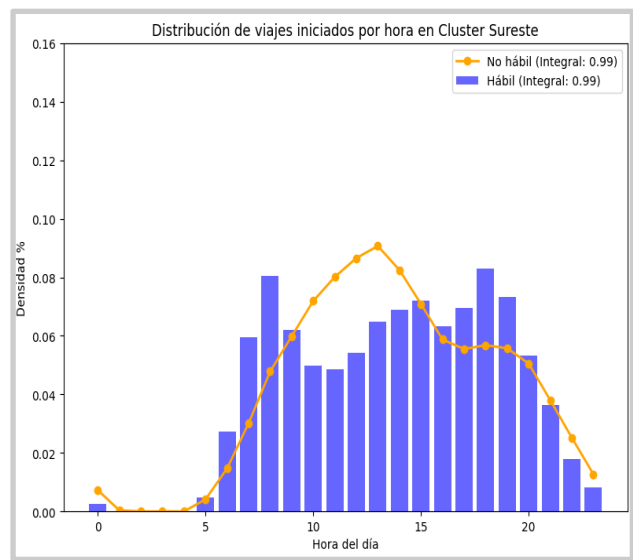


Figura 23: viajes iniciados por hora en el cluster Sureste

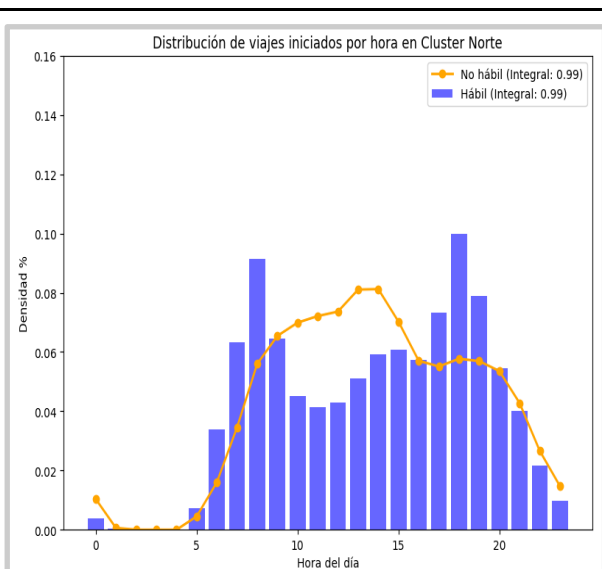


Figura 24: viajes iniciados por hora en el cluster Norte

Se pueden extraer algunas conclusiones para cada cluster:

1. **Cluster Suroeste y Norte:** estos clústeres parecen estar asociados a zonas con una mayor concentración de centros de trabajo y de estudio. El pico en la mañana coincide con los horarios de entrada a los trabajos y las escuelas (entre las 8 y 10 AM), y el pico en la tarde (entre 5 y 7 PM) coincide con la salida de estas actividades. Esto sugiere un patrón típico de día laboral, lo que indica que los usuarios en estos clústeres utilizan las bicicletas principalmente para desplazarse al trabajo o estudio.

2. **Cluster Sureste:** La distribución más dispersa y el pico más destacado en la tarde sugieren que este clúster podría estar relacionado con zonas de mayor actividad turística y comercial. Los turistas y compradores probablemente no sigan un horario estricto de 9 a 5, lo que explica los picos más suaves a lo largo del día.

Viajes finalizados por hora

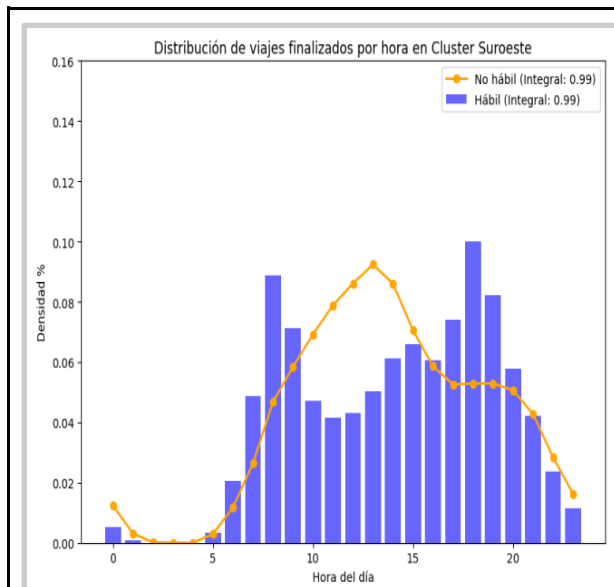


Figura 25: viajes finalizados por hora en el cluster Suroeste

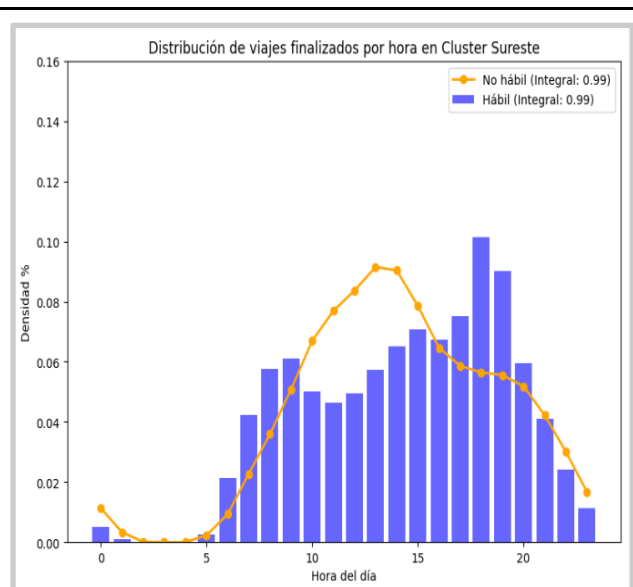


Figura 26: viajes finalizados por hora en el cluster Sureste

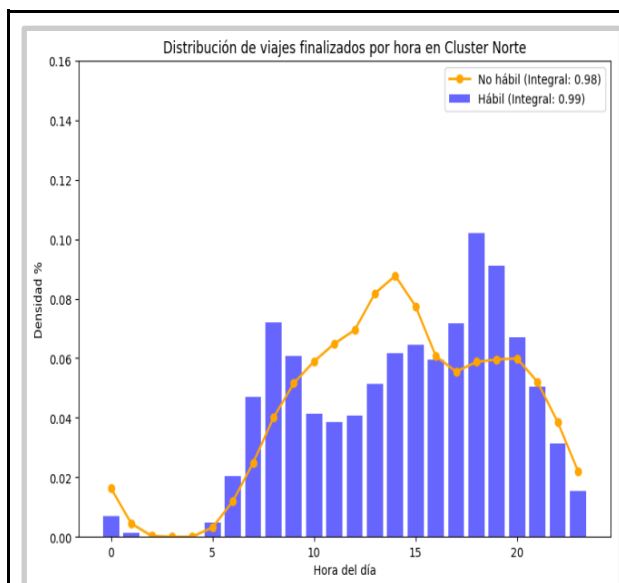


Figura 27: viajes finalizados por hora en el cluster Norte

La información apunta a algunas conclusiones para cada cluster:

1. **Cluster Suroeste y Norte:** El patrón de viajes finalizados por hora es similar al de los viajes iniciados. Esto indica que este clúster tiene un flujo significativo de personas que inician y terminan sus viajes dentro de la misma área. Los picos de finalización ocurren durante las horas de salida de trabajo (entre las 5 y 7 PM), lo que refuerza la idea de que este clúster está relacionado con lugares de trabajo y estudio. En el caso de
2. **Cluster Sureste:** A diferencia del Suroeste, en el Sureste los viajes finalizados muestran una distribución más dispersa, con un pico en la tarde (alrededor de las 4-6 PM), pero con mayor cantidad de viajes finalizados **durante la tarde y hasta la noche**. El patrón indica que los usuarios probablemente terminan sus viajes en este clúster después de actividades turísticas o comerciales. La distribución más uniforme de viajes finalizados sugiere que hay una menor dependencia de un horario laboral estricto, lo que concuerda con la idea de que **Centro 2** está relacionado con áreas turísticas o de ocio.

4.1.2.5 Densidad de viajes divididos por mañana y noche

Viajes iniciados en la mañana

Porcentaje de viajes

| nom_cluster_hc_destino | Norte | Sureste | Suroeste |
|------------------------|--------|---------|----------|
| nom_cluster_hc_origen | | | |
| Norte | 44.67% | 19.49% | 35.85% |
| Sureste | 21.92% | 39.03% | 39.05% |
| Suroeste | 13.08% | 15.43% | 71.48% |

Figura 28: matriz de viajes origen-destino con porcentaje de viajes en el Centro Urbano

Porcentaje de viajes desde cluster origen

| nom_cluster_hc_destino | Norte | Sureste | Suroeste |
|------------------------|--------|---------|----------|
| nom_cluster_hc_origen | | | |
| Norte | 13.26% | 5.78% | 10.64% |
| Sureste | 5.19% | 9.25% | 9.25% |
| Suroeste | 6.10% | 7.20% | 33.33% |

Figura 29: matriz de viajes origen-destino con porcentaje total en el Centro Urbano

Suroeste es el principal con 434 258 viajes entrantes y salientes, lo que lo posiciona como el cluster de centro urbano con más movilidad en la mañana. Representa el 33.33% de los viajes totales, una concentración alta de la movilidad en esta región.

Sureste como cantidad de viajes destino es el más bajo de los 3 cluster que llegan en la mañana esto puede deberse a que Chapalita es una zona residencial con áreas de ocio como restaurantes y cafés.

Norte según el porcentaje total de viajes dentro del norte se registra un 44.67% dentro del mismo cluster, el cluster como destino favorito para el norte es el suroeste, concentrando el 35.85% de viajes hacia el suroeste y dejando solo el 19.49% como destino destino en la mañana.

Viajes iniciados en la noche

Porcentaje de viajes

| nom_cluster_hc_destino | Norte | Sureste | Suroeste |
|------------------------|--------|---------|----------|
| nom_cluster_hc_origen | | | |
| Norte | 49.59% | 22.07% | 28.34% |
| Sureste | 23.14% | 47.49% | 29.36% |
| Suroeste | 17.37% | 19.76% | 62.88% |

Figura 30: matriz de viajes origen-destino con porcentaje de viajes en el Centro Urbano

Porcentaje de viajes desde cluster origen

| nom_cluster_hc_destino | Norte | Sureste | Suroeste |
|------------------------|--------|---------|----------|
| nom_cluster_hc_origen | | | |
| Norte | 12.27% | 5.46% | 7.01% |
| Sureste | 4.61% | 9.46% | 5.85% |
| Suroeste | 9.61% | 10.93% | 34.79% |

Figura 31: matriz de viajes origen-destino con porcentaje total en el Centro Urbano

Suroeste se incrementa la cantidad de viajes de 434 258 a 540 344 viajes en la noche. En total concentra 34.79% de viajes por la noche, casi similar a la mañana 33.33 %.

Norte muestra un aumento considerable en el número de viajes de origen y destino dentro de su zona, incrementando de 172 772 viajes en la mañana a 190 620 viajes en la noche. Este aumento probablemente sea debido a que el 69.7% de lugares identificados del cluster norte son Institutos, colegios y todo tipo de centro estudiantil, esto da a entender que prefieren usar la bicicleta más por la noche que por la mañana para asistir a sus centros educativos.

El **Sureste** es el cluster con menos cantidad de viajes de la noche, considerando que en la mañana era el cluster Norte .Esto puede deberse a que es el cluster con menor cantidad de destino de tipo escuela con tan solo 933 estaciones 933 representa el 24.08% de 3,875 estaciones de tipo escuela en el centro urbano.

4.2 Aplicación real

El análisis de MiBici destaca importantes oportunidades para mejorar la rentabilidad y la eficiencia operativa del sistema. La segmentación de usuarios, basada en la frecuencia de uso, permite diseñar ofertas personalizadas. Los usuarios más activos podrían beneficiarse de programas de lealtad que incrementen la frecuencia de uso, mientras que los usuarios menos activos o inactivos podrían ser incentivados con descuentos para reincorporarse al servicio. Esta diferenciación, alineada con los patrones de uso observados, asegura un flujo de ingresos más constante y optimiza la experiencia del usuario.

Además, el análisis espacial de las estaciones revela una marcada disparidad en la demanda entre zonas céntricas, como el *Centro Urbano*, y áreas periféricas con menor uso. La redistribución de bicicletas basada en la demanda es esencial para mejorar la eficiencia. Implementar tarifas diferenciadas por ubicación o franja horaria, como ofrecer descuentos en las estaciones menos activas durante horarios valle, podría equilibrar mejor el uso del servicio, reduciendo la congestión en zonas clave y maximizando la disponibilidad en áreas de menor actividad.

En cuanto a la demanda diaria, los días hábiles muestran un uso intensivo, especialmente en las horas de entrada y salida del trabajo. Esto presenta una oportunidad para establecer colaboraciones con empresas y centros educativos, ofreciendo suscripciones grupales que garanticen un flujo constante de usuarios. Durante los fines de semana, se pueden promover rutas recreativas o turísticas en las zonas periféricas, lo que aumentaría la utilización en estos periodos de menor demanda.

El uso de técnicas de clusterización geoespacial refuerza la necesidad de ajustar la disponibilidad de bicicletas según la demanda observada. Las estaciones con alta actividad deben ser priorizadas para asegurar la disponibilidad de bicicletas, mientras que aquellas con menor uso pueden beneficiarse de mejoras en la infraestructura, como ciclovías o iluminación, o de promociones que faciliten su acceso y aumenten su atractivo.

Como profundización de este trabajo se destaca que, contando con más información relacionada con el stock de bicicletas al momento de iniciar un viaje en una estación, se podría desarrollar un modelo predictivo del quiebre de stock de bicicletas en estaciones, considerando ubicación, tipo de día y hora, entregando así insight valioso en la gestión logística del sistema. También, mediante experimentos, se podría obtener una estimación de demanda en los sectores de mayor movimiento para mejorar el servicio.

En resumen, la combinación de una estrategia dinámica de redistribución de bicicletas, tarifas diferenciadas, y promociones dirigidas a usuarios específicos permite a MiBici maximizar tanto la eficiencia operativa como la satisfacción del cliente, al mismo tiempo que se asegura la sostenibilidad a largo plazo del sistema como un pilar clave en la movilidad urbana de Guadalajara.

4.3 Visualización de resumen

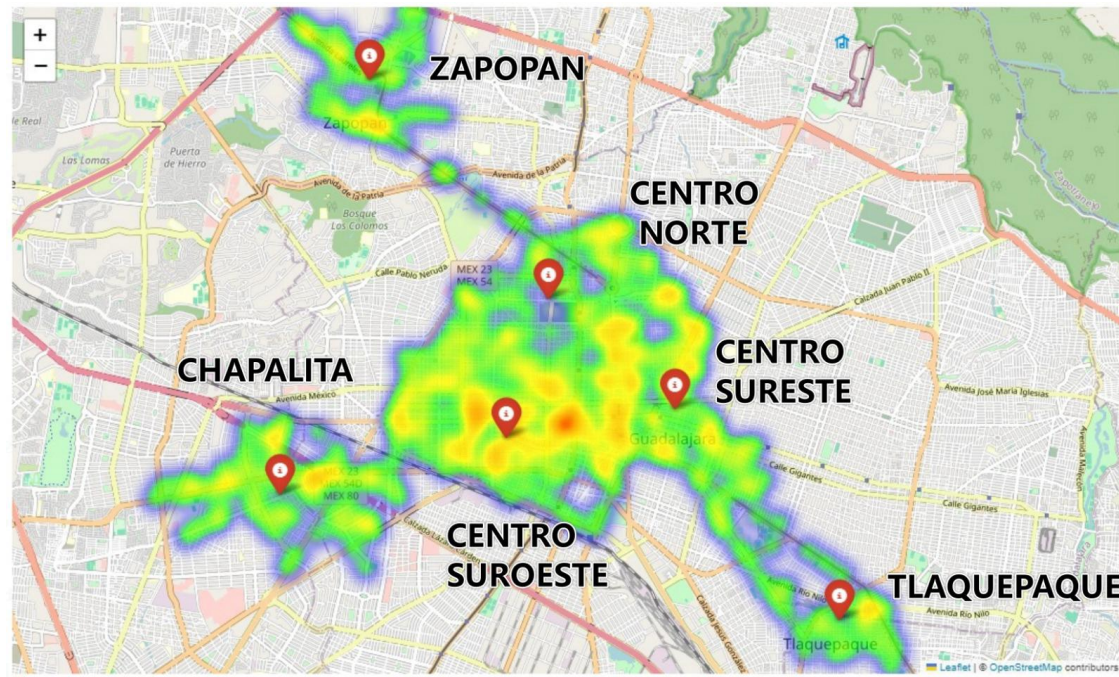


Figura 32: Mapa de calor

El mapa de calor generado presenta de manera clara las zonas con mayor y menor actividad dentro del sistema MiBici, utilizando los datos de los viajes realizados desde cada estación. Este gráfico resalta cómo el Centro Urbano, y en particular el *Suroeste*, concentra la mayoría de los viajes, lo que justifica la necesidad de priorizar la redistribución de bicicletas en estas zonas de alta demanda. Asimismo, permite identificar las estaciones con menor actividad, principalmente en los clusters periféricos, donde podrían implementarse mejoras en infraestructura o promociones para incentivar su uso.

Esta visualización respalda la propuesta de optimizar la distribución de bicicletas en función de la demanda y aplicar estrategias de marketing dirigidas a zonas con menos tráfico. Además, ofrece una base clara para decisiones estratégicas de expansión, mejoras operativas y alianzas comerciales, asegurando que el servicio sea eficiente y capaz de responder a las necesidades de los usuarios en Guadalajara.

4.4. Reflexión final

Inicialmente, se planteó la idea de analizar los viajes en función de las estaciones del año para identificar patrones de aumento o disminución en la actividad. Sin embargo, esta propuesta fue descartada debido a que Guadalajara no presenta estaciones climáticas bien definidas como en otras regiones. El clima en la ciudad es relativamente estable a lo largo del año, con variaciones mínimas entre las temporadas. Por lo tanto, incluir las estaciones del año como variable en el análisis no aportaría información relevante ni permitiría detectar cambios significativos en los patrones de uso del sistema de bicicletas compartidas.

En su lugar, se optó por considerar otras variables más representativas, como el análisis de viajes por hora y por turno mañana y noche.

Se consideró incluir topografía en nuestro análisis. Sin embargo, al analizar el área de estudio, se observó que se trata de una región mayormente plana, sin variaciones significativas en la altitud. Debido a esta falta de cambios topográficos relevantes, se optó por no incluir esta variable en el análisis final, ya que no aportaría valor significativo a los resultados.

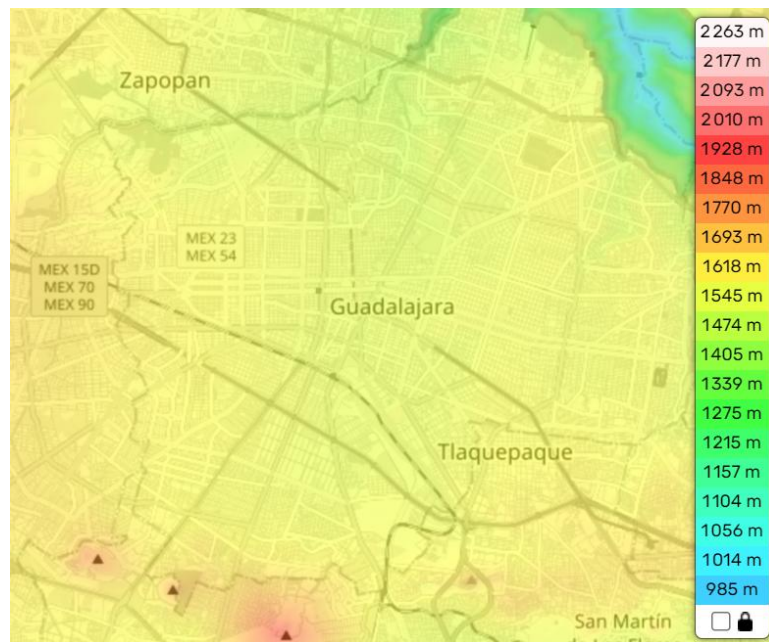


Figura 33: elevación del área investigada, tomada de topographic-map.com.

Por otra parte, a lo largo del desarrollo del TFM, uno de los principales desafíos fue definir un objetivo claro y viable que cumpliera con los estándares académicos y, al mismo tiempo, pudiera generar valor tangible para la compañía con la que inicialmente se había establecido contacto. Sin embargo, la falta de respuesta por parte de la empresa complicó esta interacción, lo que generó incertidumbre sobre la dirección que debía tomar el proyecto. Para superar esta dificultad, fue esencial el trabajo colaborativo del equipo, mediante reuniones periódicas para debatir diferentes enfoques, así como momentos de reflexión individual. Este proceso permitió a cada integrante aportar ideas que ayudaron a definir un objetivo sólido para el proyecto.

Un segundo obstáculo surgió tras recibir una baja calificación en la segunda entrega, lo que generó la necesidad de revisar profundamente las áreas de mejora. Esto motivó al equipo a aumentar la frecuencia de las reuniones y a dedicar más tiempo y esfuerzo en la corrección de errores y en la planificación de la tercera entrega. La coordinación interna y el compromiso individual de cada miembro del equipo permitieron enmendar el rumbo del trabajo, mejorando las debilidades identificadas y creando las sinergias necesarias para llevar el proyecto hacia un final exitoso. Este proceso de ajuste no solo fortaleció el resultado final, sino que también mejoró la dinámica de trabajo en equipo, asegurando que cada integrante contribuyera activamente al éxito del TFM.

5. Bibliografía y glosario de términos utilizados

Bibliografía

1. MiBici. Datos Abiertos [Internet]. 2023 [citado el 1 abr 2024]. Disponible en: <https://www.mibici.net/es/datos-abiertos/>
2. Google. Places API [Internet]. 2023 [citado 2024 jul 27]. Disponible en: <https://developers.google.com/maps/documentation/places/web-service>
3. Lee CKH, Leung EKH. Spatiotemporal analysis of bike-share demand using DTW-based clustering and predictive analytics [Internet]. Transp Res Part E Logist Transp Rev. 2023 Nov 20;172:102049. [citado 2024 jul 25]. Disponible en: <https://www.sciencedirect.com/science/article/pii/S1366554523003496>
4. Noernimat. Modeling Toronto Bikeshare Dataset Using K-Means Clustering [Internet]. GitHub. 2023 [citado 2024 jul 25]. Disponible en: https://github.com/noernimat/k-means_clustering_model
5. Gobierno de Zapopan. Guía Turística de Zapopan [Internet]. Zapopan: Gobierno de Zapopan; 2022 [citado el 17 ago 2024]. Disponible en: <https://www.zapopan.gob.mx/>
6. Instituto Nacional de Estadística y Geografía (INEGI). Censo de Población y Vivienda 2020 [Internet]. México: INEGI; 2020 [citado el 20 mar 2024]. Disponible en: <https://www.inegi.org.mx/>
7. Rabl A, de Nazelle A. Benefits of shift from car to active transport. *Transport Policy*. 2012;19(1):121–31.

Glosario

Clusterización Geoespacial: Se refiere a la agrupación de puntos geográficos basados en su proximidad espacial. Utilizando técnicas de data mining y algoritmos como K-Means o DBSCAN, se identifican patrones o concentraciones de datos en áreas geográficas específicas. Esta técnica se usa para analizar la distribución geográfica de datos y es útil en aplicaciones como estudios de mercado, planificación urbana, o el análisis de movilidad urbana.

Stakeholders: Son las personas, grupos o entidades que tienen interés o están afectadas por un proyecto, negocio o iniciativa. Los stakeholders pueden incluir clientes, empleados, inversores, proveedores, comunidades locales y reguladores. En el contexto de MiBici, por ejemplo, los stakeholders incluyen tanto a usuarios como a las autoridades municipales que apoyan el sistema.

Insights: Son descubrimientos clave o interpretaciones obtenidas a partir del análisis de datos. Un insight es un hallazgo relevante que proporciona una comprensión profunda sobre el comportamiento o tendencias, y que puede guiar decisiones estratégicas. Por ejemplo, en MiBici, un insight podría ser el hallazgo de que los usuarios prefieren ciertos horarios o estaciones específicas.

Scooters: Son vehículos de transporte personal ligeros y eléctricos, usados principalmente para recorridos cortos en áreas urbanas. Los scooters eléctricos han ganado popularidad como una solución de movilidad compartida en las ciudades. Representan una alternativa competitiva frente a las bicicletas compartidas, como en el caso de MiBici.

Soporte Gubernamental: Es el respaldo, financiero o normativo, que los gobiernos otorgan a ciertos proyectos o iniciativas. En MiBici, el soporte gubernamental se refiere a subsidios, facilidades para la infraestructura de ciclismo, o regulaciones que promuevan el uso de bicicletas como transporte sostenible.

Haversine: Es una fórmula matemática utilizada para calcular la distancia más corta entre dos puntos en la superficie de una esfera, dadas sus latitudes y longitudes. En geografía y sistemas de información geográfica (SIG), se usa para calcular distancias entre dos ubicaciones geoespaciales sobre la Tierra.

eps (epsilon): En el contexto de análisis de datos y clusterización, **eps** es un parámetro clave en el algoritmo DBSCAN (Density-Based Spatial Clustering of Applications with Noise). Representa la distancia máxima que se puede considerar para que dos puntos sean vecinos en un clúster. Es crucial para definir cuántos puntos son necesarios para formar un clúster basado en la densidad.

6. Anexos

6.1 Diagrama de Sankey

Los gráficos de Sankey tienen la ventaja de representar visualmente la proporcionalidad de los flujos. El grosor de las líneas es directamente proporcional a la cantidad de flujo que representan y el color de éstas representa el cluster de origen. En el siguiente diagrama se pueden observar los clusters identificados por los colores:

- a. Verde: Centro Urbano
- b. Morado: Chapalita
- c. Rojo: Zapopan
- d. Anaranjado: Tlaquepaque

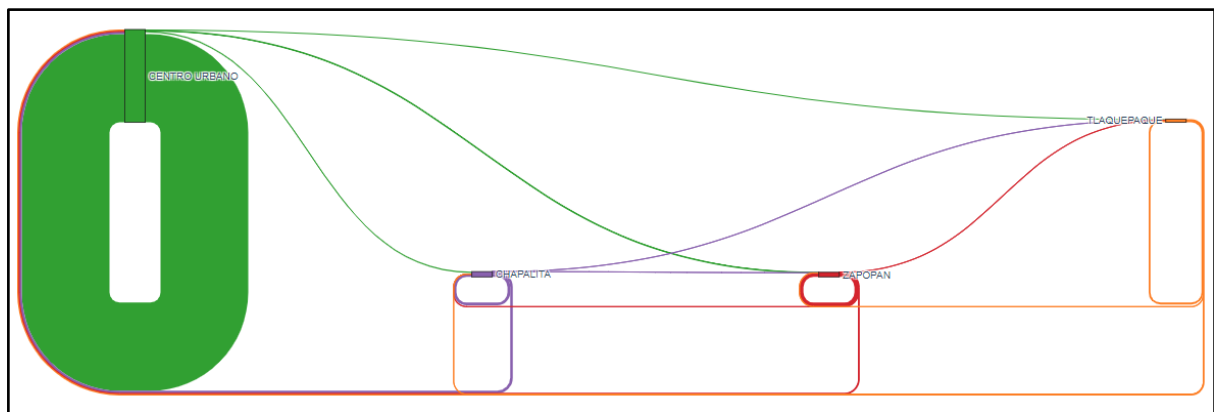


Figura 34: diagrama de Sankey para flujo de viajes intercluster

Asimismo, en el siguiente diagrama se muestra el flujo de viajes dentro del Centro Urbano

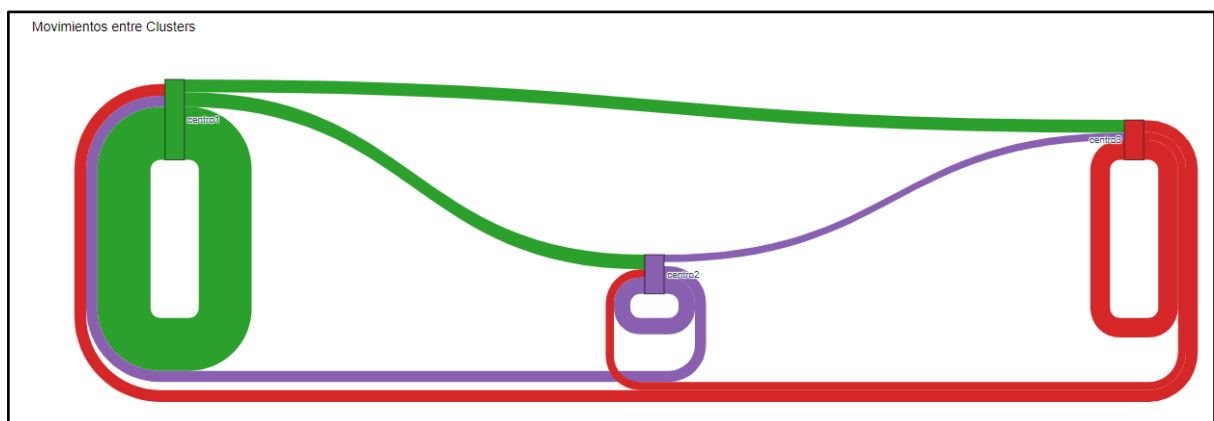


Figura 35: diagrama de Sankey para flujo de viajes en el Centro Urbano

6.2 Métodos de clusterización

K-means

Para clusterizar las estaciones de MiBici en Guadalajara primero se probó con K-means, luego de realizadas las pruebas de codo y silueta, obteniendo lo siguiente:

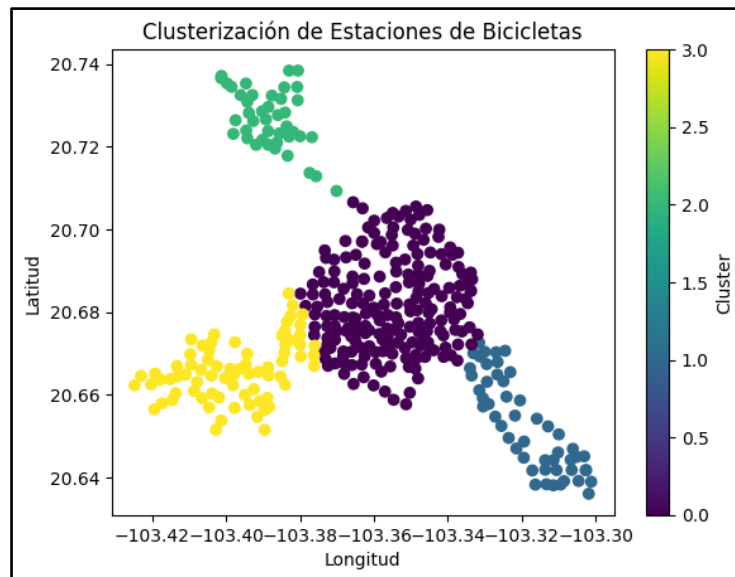


Figura 36: Clusterización k-means

Este resultado entrega 4 clusters separando las 4 grandes zonas de la ciudad donde opera la compañía. Este resultado no se eligió, pues las estaciones de lo que después será llamado como Centro Urbano no quedaban todas en el cluster central.

Una vez elegido el método de clustering jerárquico se probó con k-means para el centro urbano, obteniendo 2 clusters mediante el test del codo y el de silueta, al probar con estos números se obtiene lo siguiente:

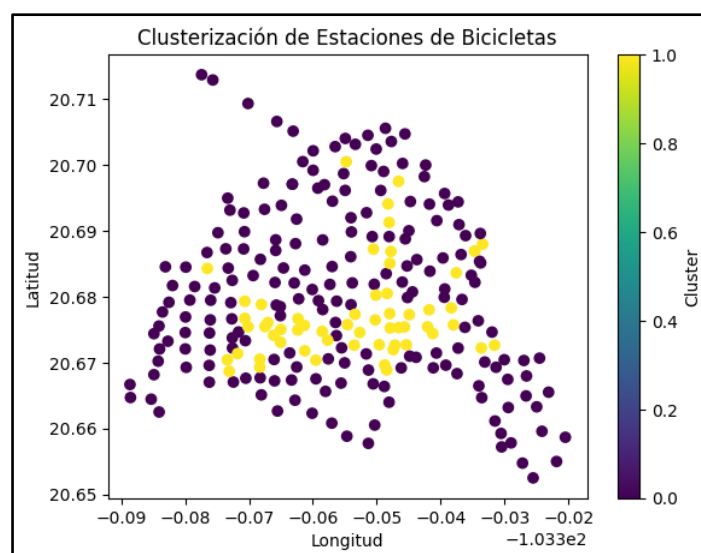


Figura 37: Clusterización k-means

K-means muestra 2 clusters pero ninguno está bien definido, por tanto este método de clusterización no es el elegido.

DBSCAN

Otro método que se probó fue DBSCAN. Después de calibrar el algoritmo con la curva del codo se obtuvo lo siguiente:

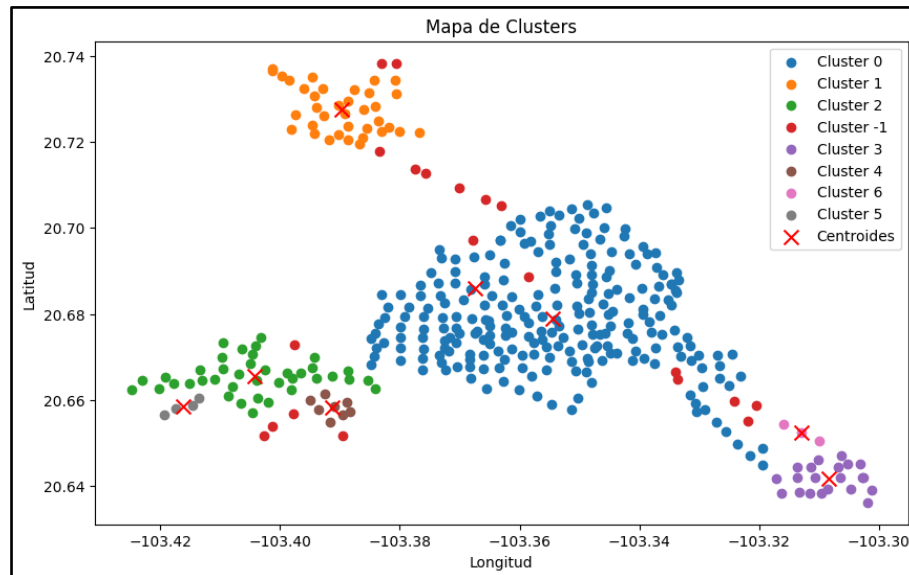


Figura 38: Clusterización Dbscan

El cluster -1 son considerados ruidos o outliers. Dado esto el clustering por dbscan fue descartado, pues se necesita tener un método que clasifique correctamente las estaciones y permita distinguir zonas.

Una vez definido el centro con el clustering jerárquico se procedió a clusterizar con este método, nuevamente obteniendo clusters que no agregaban valor al análisis que se buscaba realizar, como lo muestra la siguiente figura:

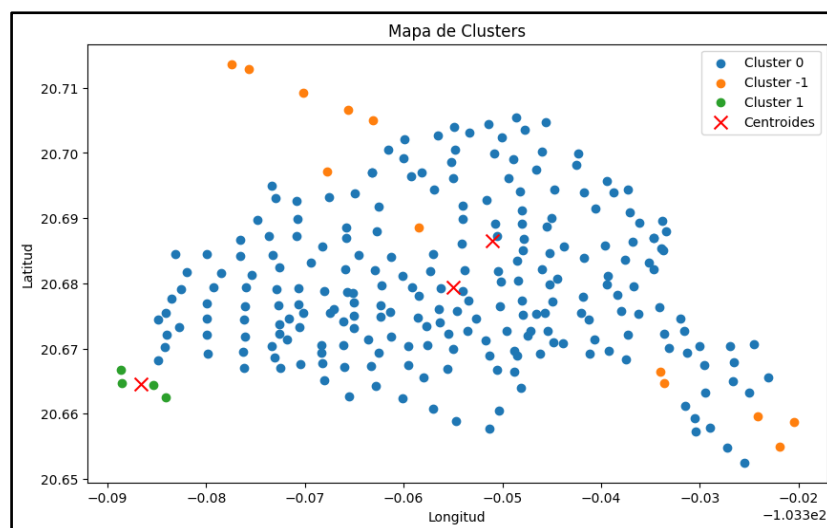


Figura 39: Clusterización Dbscan

DBSCAN logró identificar dos clusters. El tercer cluster, etiquetado como "Cluster -1", no está bien definido, ya que no cuenta con suficientes puntos vecinos dentro del radio establecido para formar un clúster.

Por este motivo, DBSCAN no se seleccionó como el mejor método de clusterización.

Jerárquico

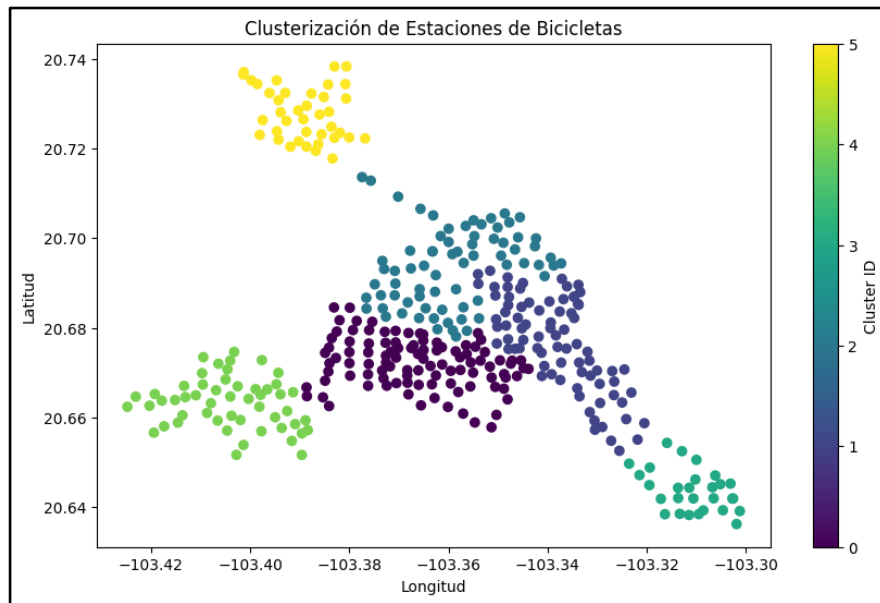


Figura 40: Clusterización Jerárquica

La clusterización jerárquica permitió tener 4 clusters claros que diferencian correctamente las distintas zonas. Este método también se ocupó para el cluster más grande, obteniendo dentro de éste 3 clusters. Finalmente, después de la aplicación de cluster jerárquico muestra claramente seis clusters bien definidos; por este motivo, se ha elegido este método para desarrollar el análisis de MiBici.

Para facilitar la interpretación, los seis clusters se dividieron en dos grupos: "Intercluster", que incluye los tres clusters en los extremos, e "Intracluster", que contiene los tres clusters centrales.

6.3 Análisis descriptivo de viajes

En el dataset de viajes el foco fue comprender cómo se distribuyen en el calendario los viajes y a lo largo del día. Para una vista respecto de la distribución de la demanda por día de la semana está el siguiente gráfico.

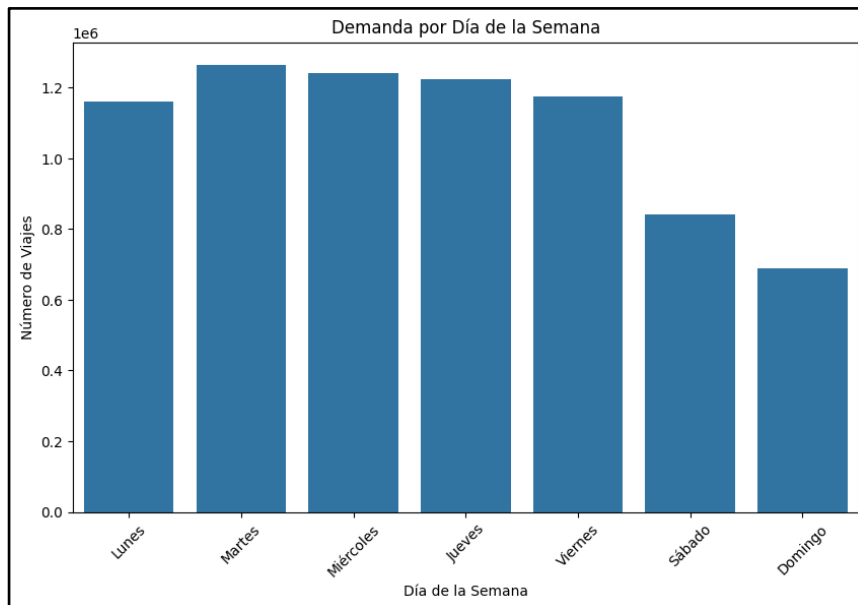


Figura 41: Demanda por día de la semana

Se observa una tendencia notable donde los días laborables (de lunes a viernes) presentan una mayor cantidad de viajes comparado con el fin de semana, lo que indica un uso intensivo del sistema probablemente para desplazamientos relacionados con el trabajo o la educación.

Como fuente de información externa se tomaron los feriados mexicanos y Guadalajara en un dataframe llamado:

mx_holidays: Variable importada sobre días festivos en México incluido festivos locales de Guadalajara.

Utilizando este dataframe, se tomaron los viajes por cada tipo de día (hábil, fin de semana y festivo) teniendo lo mostrado en el gráfico a continuación.

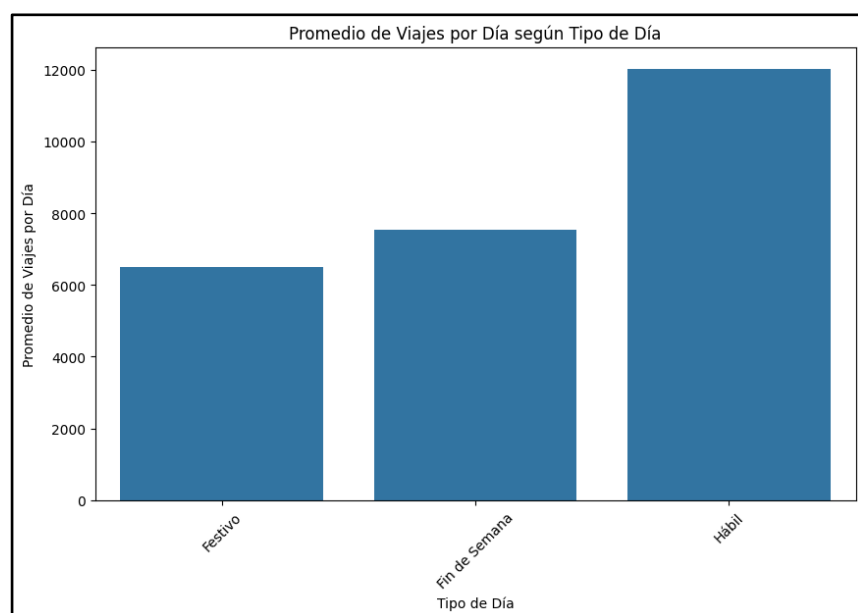


Figura 42: Promedio de viajes por día según tipo de día

Se aprecia que el mayor uso de MiBici es durante los días hábiles, luego los fines de semana y cerca están los días festivos. Dado su comportamiento similar es que Fin de semana y Festivo se agruparán en una categoría No Hábil, para facilitar el análisis.

En términos de duración de los viajes, clasificando por tipo de día se tiene que los días no hábiles tienen una duración levemente mayor que los viajes en días hábiles.

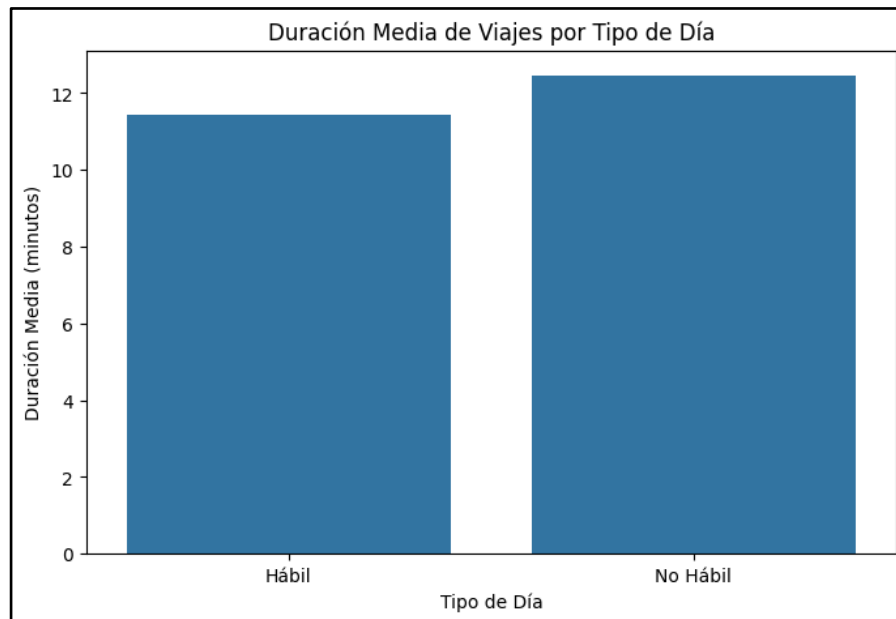


Figura 43: Duración media de viajes por tipo de día

Al hacer un doble click en la duración se aprecia que, aparte de la media, comparten los días hábiles y no hábiles una distribución similar, como lo muestran los gráficos.

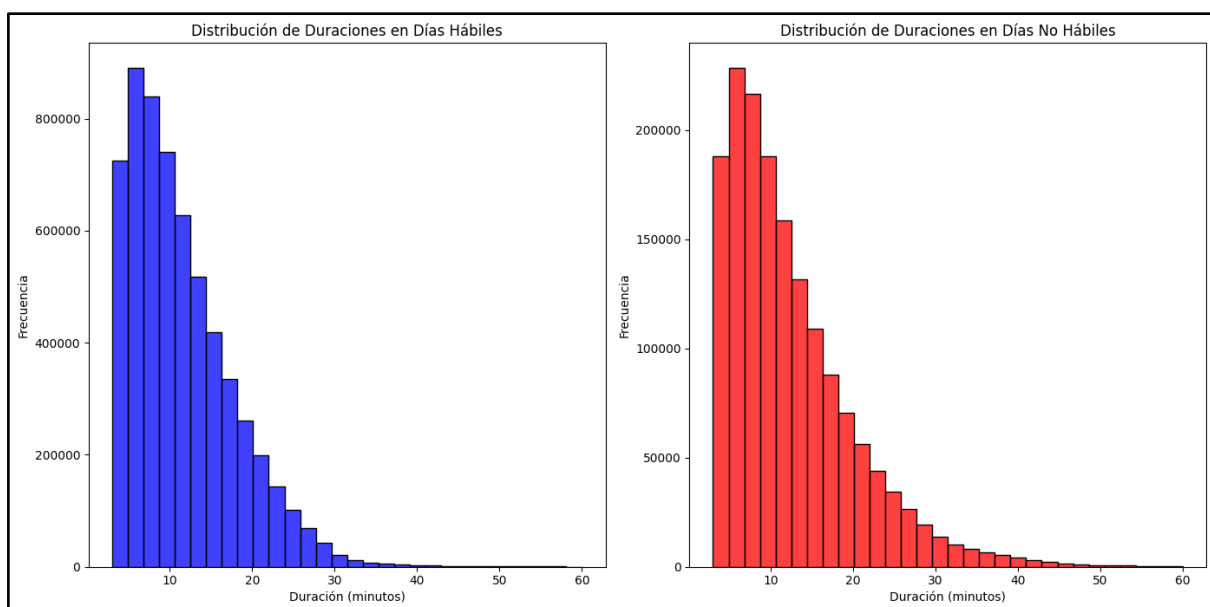


Figura 44: Distribución de duraciones en días hábiles y no hábiles

El gráfico de la izquierda muestra que en días hábiles la mayoría de los viajes son más cortos, con una moda clara alrededor de los 5 a 10 minutos. Esto sugiere que en días hábiles, los usuarios del sistema de bicicletas tienden a realizar viajes rápidos, posiblemente para desplazamientos de rutina como ir al trabajo o a la escuela.

El gráfico de la derecha muestra que aunque la moda también se encuentra en un rango de duración similar (alrededor de 5 a 10 minutos), la distribución es ligeramente más plana en comparación con los días hábiles, extendiéndose más hacia duraciones de 15 a 20 minutos. Esto podría indicar que en días no hábiles, los usuarios están más dispuestos a realizar viajes ligeramente más largos, quizás para actividades recreativas o turísticas.

Para el caso del análisis intradía se definieron 5 franjas horarias:

- **Mañana pre trabajo (Morning):** De 6:00 a.m. a 10:00 a.m. Cubre el periodo de desplazamientos matutinos y el inicio de actividades diurnas.
- **Mañana post trabajo (late morning):** de 10:00 a.m. a 1:00 p.m., cubre el periodo valle entre la llegada al trabajo y el almuerzo
- **Almuerzo (Midday):** De 1:00 p.m. a 4:00 p.m. Captura el periodo posiblemente más relajado del día, incluyendo la hora del almuerzo y actividades post-almuerzo.
- **Tarde (Afternoon):** De 4:00 p.m. a 8:00 p.m. Incluye el periodo de regreso a casa para muchas personas y el inicio de actividades de ocio después del trabajo.
- **Noche (Evening/Night):** De 8:00 p.m. a 12:00 a.m. Tiempo donde disminuye la actividad general, pero puede incluir salidas nocturnas o regresos tardíos a casa.

El uso de MiBici difiere bastante a lo largo del día, dependiendo de si el día es hábil o inhábil.

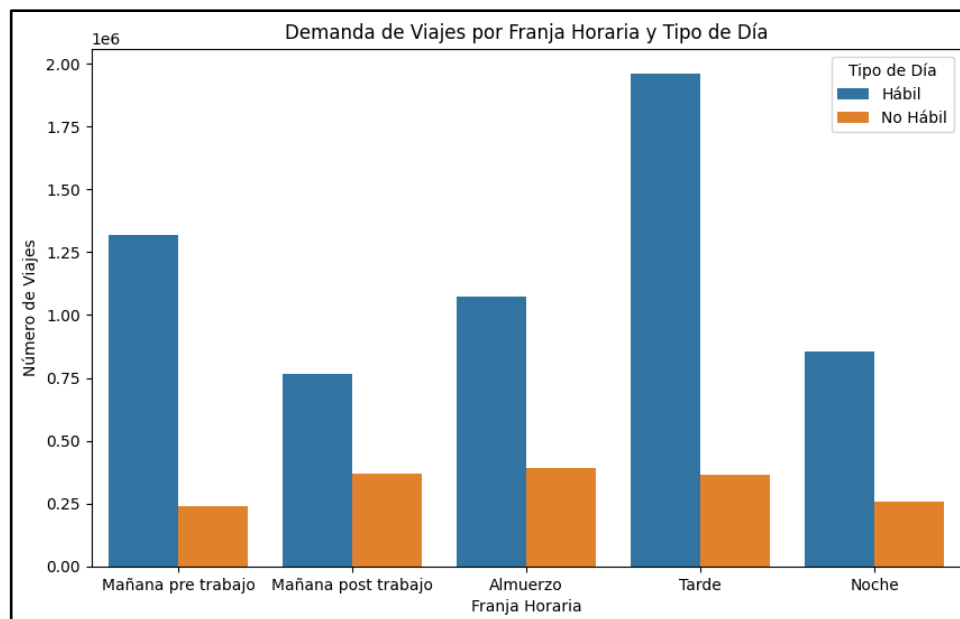


Figura 45: Demanda de viajes por franja horaria y tipo de día

Este gráfico ilustra la demanda de viajes en diferentes franjas horarias, diferenciando entre días hábiles y no hábiles. La demanda se mide en términos del número total de viajes realizados en cada franja horaria, mostrando claramente cómo varía el uso del servicio de bicicletas según el tipo de día y la hora del día.

Se puede observar que durante los días hábiles, la franja horaria con mayor demanda es la tarde, con un número de viajes que supera ampliamente al de otras franjas horarias, alcanzando cerca de 2 millones de viajes. Esto sugiere que muchas personas utilizan el servicio de bicicletas para regresar a casa o desplazarse después de sus actividades laborales o escolares. La segunda franja horaria con mayor demanda en días hábiles es la mañana pre-trabajo, con alrededor de 1.25 millones de viajes, lo cual indica que el servicio también es popular para el desplazamiento hacia el trabajo o estudios en las mañanas.

Las franjas horarias de mañana post trabajo y almuerzo también muestran una demanda considerable en días hábiles, aunque significativamente menor que la de la mañana pre trabajo y la tarde. La franja de la noche presenta una demanda similar a estas, lo que permitiría llamar horarios valle a mañana post trabajo, almuerzo y noche, mientras que se podría llamar horario peak a los viajes de tarde o mañana pre-trabajo.

En días no hábiles se observa una distribución más uniforme de la demanda a lo largo del día. Las franjas horarias de almuerzo, tarde y noche muestran una demanda relativamente similar, con cada una alcanzando cerca de 300 mil viajes. Esto podría indicar que en días no hábiles, las personas utilizan el servicio de bicicletas de manera más distribuida a lo largo del día, posiblemente para actividades recreativas o desplazamientos no relacionados con el trabajo.

df_places: Variable obtenida de Google Places API que contiene información de nombre, latitud, longitud y tipo de lugares en Guadalajara en tres categorías (escuelas, atracciones turísticas, centros comerciales).