

CENTRO DE INVESTIGACIÓN Y DOCENCIA ECONÓMICAS, A.C.



EXPLORANDO LOS FACTORES QUE INFLUYEN EN LA DEMANDA DIFERENCIADA
POR GÉNERO EN UN SISTEMA DE BICICLETAS COMPARTIDAS MEXICANO: UN
ANÁLISIS ESPACIO-TEMPORAL MEDIANTE RANDOM FOREST

TESINA

QUE PARA OBTENER EL GRADO DE
MAESTRA EN MÉTODOS PARA EL ANÁLISIS DE POLÍTICAS PÚBLICAS

PRESENTA

CRISTINA MARTIN DEL CAMPO NUÑEZ

DIRECTORA DE LA TESINA: DRA. DANIELA ALEJANDRA MOCTEZUMA OCHOA

AGUASCALIENTES, AGS.

2024

Agradecimientos

A mi madre, Beatriz, y a mi padre, Rogelio por haberme forjado como la persona que soy.

A mi hermana Julieta, por ser mi mayor cómplice y mi más grande ejemplo de dedicación.

A Gerardo, por todo el amor, apoyo y motivación en esta etapa.

A mis amigos, Jazmín, Erick, Richard, Korina, Raymundo e Itzel, por ser mi equipo en los estudios y en las aventuras.

A mi directora, la Dra. Daniela Moctezuma, por toda la paciencia, tiempo y conocimientos que compartió conmigo.

A mi lector, el Dr. Jaime Sainz, por su interés y confianza en mi tema desde el inicio.

Al CIDE y a todos sus docentes por la invaluable oportunidad de formarme en política pública y adentrarme en los métodos cuantitativos.

Al CONACYT por financiar mis estudios y mi estadía en Aguascalientes.

*Dedico este logro a todas las mujeres,
comprometiéndome a continuar la
búsqueda constante de producir
conocimiento con perspectiva de género.*

Resumen

La brecha de género en el uso de sistemas de bicicletas compartidas obstaculiza la realización plena de los beneficios asociados a su utilización como alternativa de transporte urbano. Aunque diversos factores, tanto individuales como sociales y ambientales, influyen en el uso de la bicicleta, existe escasa investigación en la región latinoamericana que analice de manera diferenciada los efectos de las características espaciales del entorno urbano en hombres y mujeres. El presente estudio tiene como objetivo identificar y validar los factores que afectan de manera diferenciada la demanda por género en el sistema MiBici de Guadalajara. Para ello, se propone una estructura metodológica que utiliza datos de viajes realizados en octubre de 2019 para entrenar un modelo con el algoritmo de *Random Forest*. Este modelo, utiliza como datos de entrada la información del entorno urbano de las estaciones de origen y destino, para generar como salida el género del viajero. A partir de estos datos, el modelo permite determinar la importancia de cada variable en el proceso de clasificación, es decir, identifica qué variables son cruciales para determinar si un viaje fue realizado por una mujer o un hombre. Aunado a lo anterior, se busca validar los resultados mediante los datos de la Encuesta de Percepción MiBici del mismo año, así como a través de una visualización del cruce de variables urbanas y la densidad de viajes por género en mapas generados con un sistema de información geográfica. Aunque los resultados de esta última validación cualitativa por medio de visualización de mapas no fueron concluyentes, los resultados sugieren que la infraestructura ciclista, como las vialidades de prioridad ciclista, las ciclo vías y los ciclopuertos, es de gran importancia para el modelo, al igual que la distancia al transporte público masivo, lo cual es consistente con la escasa literatura previa. Estos hallazgos tienen importantes implicaciones para el diseño de políticas de movilidad urbana inclusivas y la planificación de infraestructuras de transporte.

Índice

1. Introducción.....	1
2. Revisión de literatura.....	3
2.1 Sistemas de Bicicletas Compartidas y factores que afectan su demanda	3
2.2 Brecha ciclista y diferencias de uso por género en la movilidad.....	5
2.3 Aplicación de algoritmos de Machine Learning para el análisis de la demanda	6
3. Área de estudio y descripción de los datos.....	8
3.1 Información del SBC	9
3.2 Información del entorno.....	14
3.3 Encuesta de Percepción MiBici 2019	17
4. Metodología y desarrollo.....	18
4.1 Random Forest.....	18
4.2 Análisis de la encuesta	22
4.3 Análisis visual y cualitativo	22
4.4. Desarrollo del modelo.....	23
5. Resultados y discusión	26
5.1 Análisis del desempeño del modelo.....	26
5.2 Análisis de la importancia de los factores.....	28
5.3 Validación comparativa con Encuesta Mibici.....	38
5.4 Validación visual y cualitativa por medio de mapas.....	42
6. Conclusiones y recomendaciones.....	47
Bibliografía.....	49
Anexo 1	57

Índice de figuras

Figura 1. Distribución de las estaciones del sistema MiBici en 2023	8
Figura 2. Histórico de viajes realizados en el sistema de diciembre de 2014 a octubre de 2023	9
Figura 3. Diagrama de flujo del preprocesamiento de datos	10
Figura 4. Distribución de la cantidad de viajes por hora y por sexo para el mes de octubre 2019	11
Figura 5. Pirámide de edades de los usuarios anuales del sistema Mibici en octubre 2019	12
Figura 6. Proporción de viajes por sexo según su duración	13
Figura 7. Cantidad de viajes por cada estación y promedio según día de la semana y horario definido por género.....	14
Figura 8. Ejemplo de buffer de 400 metros de radio trazado a partir de cada estación del sistema	15
Figura 9. Matriz de correlaciones entre las variables del entorno	16
Figura 10. Diagrama de flujo de la estructura metodológica propuesta.....	18
Figura 11. Ejemplo de un árbol de decisión de 4 capas	19
Figura 12. Ejemplo del funcionamiento de un Random Forest.....	20
Figura 13. Proceso metodológico	23
Figura 14. Comparativo del desempeño de las métricas para el conjunto de prueba de los diferentes modelos propuestos con y sin variables temporales	28
Figura 15. Variación de la importancia de las variables para los cuatro ejercicios del modelo mensual que incluyen variables temporales	29
Figura 16. Importancia promedio de las variables del modelo mensual que incluye variables temporales.....	30
Figura 17. Importancia promedio de las variables del modelo mensual que excluye variables temporales.....	32
Figura 18. Variación de la importancia de las variables para los cuatro ejercicios de los modelos por horario que incluyen variables temporales	33
Figura 19. Importancia promedio de las variables de los modelos por horario que incluyen variables temporales	34

Figura 20. Pirámide de edades de las personas encuestadas en la encuesta de percepción MiBici 2019.....	38
Figura 21. Porcentaje de ocupación de las personas encuestadas según género	39
Figura 22. Modos de transporte con los que suelen completar el viaje de MiBici las personas encuestadas según género.....	40
Figura 23. Asociación entre la distribución de las calles de prioridad ciclista y la densidad de viajes originados según género.....	43
Figura 24. Asociación entre la concentración de las calles de prioridad ciclista por buffer y el porcentaje de viajes originados por género	44

Índice de tablas

Tabla 1. Factores que afectan la demanda de viajes en los SBC agrupados en categorías	4
Tabla 2. La demanda de movilidad por género.....	5
Tabla 3. Número de observaciones y su proporción según género de los distintos conjuntos de datos.....	24
Tabla 4. Métricas en el conjunto de entrenamiento	27
Tabla 5. Métricas en el conjunto de prueba	27
Tabla 6. Variables de mayor importancia en el origen y destino para los diferentes modelos.....	37
Tabla 7. Porcentaje de personas encuestadas por género según el propósito principal del viaje	41
Tabla 8. Resumen de los resultados	45

1. Introducción

En los últimos años, las bicicletas y los Sistemas de Bicicletas Compartidas (SBC), se han consolidado como un modo de transporte popular en las ciudades, con múltiples beneficios individuales y colectivos, como la promoción de la salud física y mental, la recuperación del espacio público, la mejora la calidad del aire y la reducción de la congestión (PAHO, 2020; Naciones Unidas, 2022).

A pesar de su creciente popularidad, sencillez y asequibilidad, no se han alcanzado los beneficios colectivos en su totalidad y los beneficios individuales no se han distribuido de manera equitativa entre hombres y mujeres. En Latinoamérica, por ejemplo, a pesar de que las mujeres constituyen aproximadamente la mitad de la población, representan menos del 30% de los viajes en bicicleta. Esta cifra es considerada baja en contraste con Países Bajos y Alemania, donde las mujeres representan la mitad o más de los desplazamientos en bicicleta (Díaz y Rojas, 2017; Pucher y Buehler, 2008).

La literatura destaca diversos factores que influyen en el uso de la bicicleta, desde aspectos individuales y sociales hasta características físicas y ambientales del entorno (Emond, Tang y Handy, 2009). Entre estos factores, el género ha sido identificado como uno de los determinantes más significativos en el uso de la bicicleta y los SBC. Además, se ha encontrado, al menos en contextos anglosajones, que el uso de los SBC está fuertemente correlacionado con las características espaciales del entorno urbano, como la diversidad de usos de suelo y actividades presentes, la densidad poblacional, la presencia de infraestructura y atractores o puntos de interés (El-Assi et al., 2017; Wang et al., 2015; Buck y Buehler, 2012; Faghih-Imani et al., 2014; Faghih-Imani et al., 2017).

Estas características del entorno urbano inmediato a las estaciones de bicicletas podrían tener una influencia significativa en la disparidad de uso de estas, un tema que ha sido escasamente estudiado en la literatura y aún más escasamente explorado en la región latinoamericana y en México. Con el fin de comprobar esta hipótesis, se utilizaron los datos abiertos oficiales del sistema MiBici, el SBC que desde 2014 opera en Guadalajara, la segunda ciudad más grande del país; además de información proporcionada por el Instituto de Planeación y Gestión del Desarrollo del Área Metropolitana de Guadalajara (IMEPLAN) y el Instituto Nacional de

Estadística y Geografía (INEGI). Estos datos se utilizaron para generar un modelo con *Random Forest* (RF), un algoritmo de *Machine Learning* (ML); y de esta forma evaluar la relevancia de cada variable en el funcionamiento del modelo. De esta manera, se buscó identificar los factores más influyentes en la clasificación por género de los viajes realizados en el SBC, para posteriormente, contrastarlos y validarlos, por un lado, con los resultados obtenidos por una encuesta de percepción realizada por WRI y por otro, de manera cualitativa con mapas del entorno urbano elaborados con un sistema de información geográfica.

Los resultados son relevantes para planificadores y responsables de la política de movilidad, proporcionando una mejor comprensión de los factores que afectan el uso diferenciado de la bicicleta por género. Estos resultados ofrecen insumos para proponer políticas que reduzcan la brecha de género y promuevan intervenciones eficaces y eficientes en infraestructura ciclista. Además, la información sobre los factores que afectan el uso de la bicicleta por género podría ser útil para la expansión de sistemas similares o la implementación en otras ciudades.

El presente documento se organiza de la siguiente forma, la sección 2 revisa la literatura relacionada con los factores que afectan el uso de los SBC y cuáles impactan específicamente en la movilidad femenina haciendo énfasis en las características del espacio construido y características temporales. En la sección 3 se presentan los antecedentes del sistema MiBici y la estadística descriptiva tanto de las diferencias de género en el sistema como de las características espaciales de su contexto. En la sección 4 se presenta la metodología empleada y se explica el desarrollo del modelo para la identificación de las variables con mayor influencia, posteriormente en la sección 5 se discuten los resultados de los modelos de RF, contrastándolos con los de la Encuesta de Percepción MiBici 2019 y con un análisis cualitativo del entorno urbano mediante mapas generados en un sistema de información geográfica. Finalmente, en la sección 6 se presentan las conclusiones y las recomendaciones para futuras investigaciones.

2. Revisión de literatura

En esta revisión de literatura, se explora la conexión entre los factores urbanos y la disparidad de género en el uso de los SBC. Se exploran estudios previos que abordan temas relacionados con la demanda de los SBC, el género y la movilidad, así como investigaciones que utilizan algoritmos de *Machine Learning* para aproximar la demanda y comprender los factores que influyen en ella.

2.1 Sistemas de Bicicletas Compartidas y factores que afectan su demanda

Los SBC han surgido como una respuesta a los desafíos de la movilidad urbana. Estos sistemas, implementados tanto por operadores privados como por autoridades locales, ofrecen al público un número determinado de bicicletas para su uso durante un periodo específico, ya sea con tarifas o de forma gratuita (Transport Canada, 2009; Castellanos et al., 2019 y PAHO, 2020). Su origen se remonta a 1967 con la implementación del primer sistema de este tipo en Ámsterdam. A lo largo de cuatro generaciones y medio siglo, la proliferación de estos sistemas ha sido notable, estimándose que para 2021 existían más de tres mil en todo el mundo (PBSC, 2021), con 92 registrados en la región latinoamericana en 2019, incluyendo 15 en México (LatinoSBP.org, 2019). A pesar de su crecimiento, su adopción no ha sido tan extendida pues se ve condicionada por diferentes factores.

En el estudio de Eren y Uz (2020), se llevó a cabo una revisión de la literatura enfocada en los factores que afectan la demanda de los SBC. Los autores categorizaron estos factores en seis grupos, detallados en la **Tabla I**. Dentro de su investigación, se resaltan especialmente los efectos positivos vinculados a la presencia de infraestructura ciclista, estaciones de transporte público y diversos puntos de interés como áreas verdes, instituciones educativas, museos, centros comerciales, restaurantes y hoteles.

Tabla 1

Factores que afectan la demanda de viajes en los SBC agrupados en categorías

Clima	Ambiente construido y uso de suelo	Transporte público
<ul style="list-style-type: none"> • Estaciones • Tipo de clima • Precipitación • Viento • Humedad • Temperatura 	<ul style="list-style-type: none"> • Infraestructura ciclista • Ambiente construido • Sistema de Bicicletas Públicas • Uso de suelo • Área Natural 	<ul style="list-style-type: none"> • Red de Transporte Público • Transporte Público • Viajes
Localización de las estaciones y buffers	Efectos sociodemográficos	Efectos temporales y de seguridad
<ul style="list-style-type: none"> • 200 m. • 250 m. • 300 m. • 400 m. • 500 m. • 800 m. • 1,000 m. • 3,000 a 4,000 m. 	<ul style="list-style-type: none"> • Genero • Edad • Educación • Ingresos • Tenencia 	<ul style="list-style-type: none"> • Día de la semana • Celebraciones • Horarios • Uso de casco

Fuente: Elaboración propia a partir de Eren y Uz (2020).

En el mismo contexto, es relevante destacar los criterios y la importancia asociada a la delimitación de *buffers* o áreas de influencia por estación. Estas áreas circulan con un radio definido en torno a cada estación del SBC, son fundamentales para analizar la mayoría de los factores previamente mencionados y comprender cómo la ubicación influye en la demanda de las estaciones. En ese sentido, para su delimitación se da, ya sea en función de la distancia promedio entre las estaciones o bien, dada en función de una distancia “caminable”.

La mayoría de los estudios sobre la demanda de los SBC se centran en análisis generales a nivel de estación. Considerando que varios sistemas ofrecen suscripciones diarias, semanales y anuales, han surgido análisis que consideran esta segmentación entre usuarios casuales y usuarios anuales, revelando disparidades en intensidad, patrones de uso y sensibilidad ante factores externos (Vogel et al., 2014; Faghieh-Imani et al., 2017; Sun et al. 2018; Jain et al, 2018). No obstante, existe un vacío significativo en la literatura, ya que son escasas o inexistentes las investigaciones que abordan de manera diferenciada los efectos de ciertos factores en hombres y mujeres.

2.2 Brecha ciclista y diferencias de uso por género en la movilidad

Se ha observado que en varios países la participación de mujeres en el ciclismo es notoriamente inferior a la de los hombres. (Garrard et al., 2012; Shaw et al., 2016). Pellicer-Chenoll et al. (2021) atribuyen estas disparidades a una “cultura ciclista”, explicando las marcadas diferencias entre lugares como Países Bajos o Alemania, donde el uso de la bicicleta es más equitativo, y regiones como la latinoamericana, donde prevalece una brecha muy pronunciada.

Sin embargo, la cultura no es el único factor determinante. Una revisión sistemática de Ravensbergen et al. (2018), especialmente basada en estudios estadounidenses, identifica dos explicaciones principales para esta brecha de género en el ciclismo. En primer lugar, las mujeres manifiestan una mayor preocupación por la seguridad y preferencia por infraestructura segregada. En segundo lugar, las responsabilidades duales de las mujeres, tanto reproductivas como productivas, pueden hacer que el uso de la bicicleta sea complicado, especialmente en viajes asociados al trabajo doméstico y de cuidado.

Este impacto del género en los patrones de movilidad se refleja en el análisis de Jaimurzina et al. (2017) (ver **Tabla 2**) donde evidencia que la movilidad de las mujeres es más compleja debido a los roles que desempeña, generando viajes más variados y dependientes de necesidades distintas, incluso de terceros.

Tabla 2

La demanda de movilidad por género

	Mujeres	Hombres
Motivos De viaje	Diversos, en función de sus tareas laborales, sociales y relacionadas con las tareas de hogar	Principalmente laborales
Distancia	Trayectos concentrados, de corta a mediana distancia, frecuentemente bastante cerca del hogar	Distancias medianas a largas
Paradas	Múltiples paradas durante el trayecto	Limitadas, relacionadas, generalmente, con las responsabilidades laborales
Origen-destino	Fragmentado, viajes en zigzag	Trayectos directos de ida y vuelta por razones laborales
Horarios	Variables	Durante las horas pico
Viajes acompañados	Frecuentemente acompañados de niños y personas mayores y de movilidad reducida.	Menos que las mujeres
Carga	Coches de bebé, bultos o compras	Sin carga particular

Fuente: Jaimurzina, Muñoz y Pérez, 2017, p.13.

Desafortunadamente, estas diferencias en la movilidad rara vez se consideran. Wang y Akar (2019) señalan la escasez de estudios que exploren cómo diversos factores, especialmente aquellos relacionados con el entorno urbano, influyen de manera diferenciada en los SBC según el género. En la región latinoamericana, Higuera-Mendieta et al. (2021) coinciden en señalar que la mayoría de las investigaciones carecen de una perspectiva de género, limitándose a incluir el género como una variable de control adicional.

En el análisis de la relación entre género, características urbanas y SBC, el estudio de Wang y Akar (2019) en el sistema *CitiBike* de New York utiliza modelos binomiales negativos, revelando resultados esperados. La presencia de ciclovías segregadas y de bicigestacionamientos se asocia positivamente con un mayor uso por parte de las mujeres. Además, la bicicleta es utilizada por mujeres para diversos fines más allá de los utilitarios, demostrando una asociación positiva entre la presencia de áreas verdes en un radio de 500 metros y una mayor cantidad de viajes femeninos. El número de intersecciones viales se correlaciona de manera diferente, de modo que una menor densidad vehicular implica velocidades más bajas que a su vez se asocia con una mayor atracción de viajes femeninos. Por el contrario, un mayor número de estaciones de camiones o entradas al metro resultaron ser predictores negativos de viajes realizados por mujeres (Wang y Akar, 2019).

2.3 Aplicación de algoritmos de Machine Learning para el análisis de la demanda

El *Machine Learning (ML)* o Aprendizaje Automático constituye una rama de la Inteligencia Artificial. Se caracteriza por la capacidad de un sistema para adaptarse a nuevas circunstancias mediante algoritmos que detectan y extrapolan patrones a partir de los datos proporcionados (Russel y Norvig, 2004). Este proceso no depende únicamente de programación explícita, sino que implica aprender a partir de los datos percibidos y sus propiedades, como lo son sus características y etiquetas, siendo las características las variables independientes y las etiquetas las dependientes. En ese sentido, el ML busca, a partir de características, aprender a predecir clases o clasificaciones de los datos, especialmente aquellas difíciles de medir o calcular directamente (Jung, 2022). Se distinguen tres tipos principales de tipos: aprendizaje supervisado, no supervisado y por refuerzo.

La implementación de ML en el análisis de los SBC ha experimentado un auge significativo gracias a los avances tecnológicos y la generación masiva de datos en tiempo real. En una revisión sistemática realizada por Albuquerque et al. (2021) sobre enfoques de aprendizaje automático aplicados en SBC, se identificaron dos tendencias principales. La primera implica el uso de algoritmos de aprendizaje no supervisado para agrupación o clasificación, mientras que la segunda se centra en el aprendizaje supervisado para predicciones. Los estudios sugieren que la utilidad de los algoritmos de ML en los SBC se centra en el análisis y pronóstico de la demanda, la predicción de flujos y en el rebalanceo de bicicletas entre las estaciones.

Respecto a los métodos y algoritmos empleados, la investigación de Albuquerque et al. (2021), revela que, para clasificación, se utilizan comúnmente algoritmos como *K-Means* y *clustering* jerárquico, en ocasiones acompañados de algoritmos de detección de comunidades. Otro segmento importante emplea máquinas de soporte vectorial (SVM), redes neuronales (NN), árboles de decisión (DT), y el ampliamente aplicado *Random Forest* (RF), especialmente en la predicción de la demanda y exploración del impacto de diversos factores en ella.

Estudios comparativos han favorecido a RF por su alta capacidad para predecir con gran exactitud la demanda de los SBC (Yang et al, 2016; Bacciu, 2017; Wang y Kim, 2018; Feng y Wang, 2017; Ashqar et al., 2020; Ruffieux et al., 2017). Asimismo, la combinación de RF con otros algoritmos como *k-means* (Xu et al, 2019) o *SVM* (Bacciu et al, 2017) también han demostrado buenos resultados en las predicciones. Las variables temporales y climatológicas son comúnmente empleadas para predecir la demanda, sin embargo, estudios más exhaustivos que, además de predecir la demanda, buscan identificar y clasificar los factores más relevantes que la afectan, suelen incluir además variables espaciales como el uso de suelo, puntos de interés y presencia de infraestructura, así como variables demográficas.

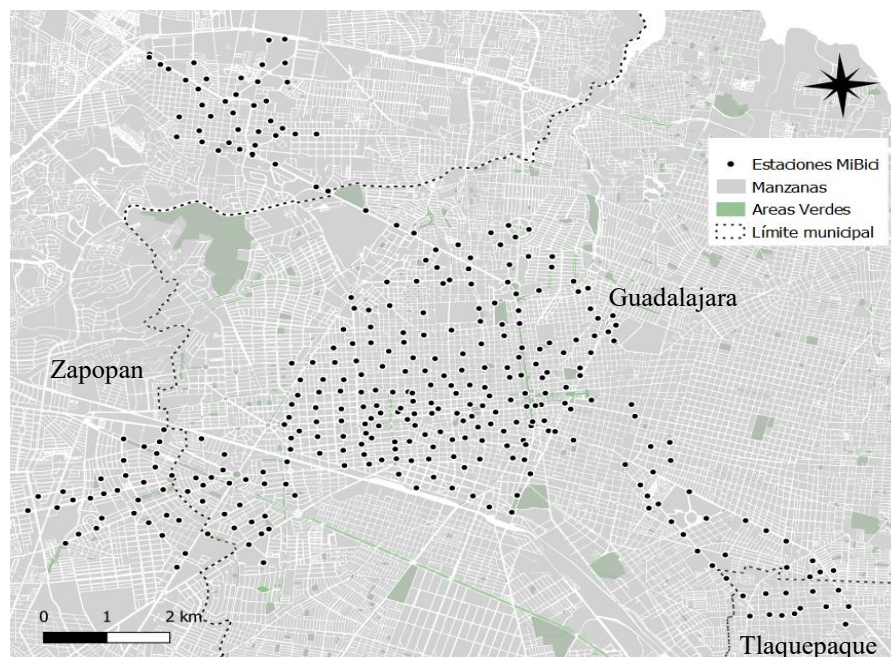
Ejemplos específicos, como el estudio de Du et al. (2019) en Shanghai, con RF, resaltan la influencia de factores como el área residencial, áreas verdes, y tamaño de la población en la frecuencia de uso de bicicletas sin anclaje en diferentes periodos del día. Del mismo modo, Guidon, et al (2020) en su estudio para expandir un SBC eléctrico identificaron, mediante RF, la importancia de variables como lugares de empleo, tamaño de la población, presencia de bares y restaurantes y distancia a la estación de tren principal para predecir la demanda en dos ciudades suizas.

3. Área de estudio y descripción de los datos

El análisis de esta investigación se fundamenta en datos del SBC “MiBici” del Área Metropolitana de Guadalajara (AMG), la tercera metrópoli más grande del país (SEDATU, CONAPO e INEGI, 2023). Este SBC, el segundo en el país, opera en algunos municipios de esta ciudad gracias a los esfuerzos tanto de colectivos ciclistas como de dependencias gubernamentales. Fue inaugurado en 2014 con menos de 90 estaciones (Instituto de Movilidad y Transporte del Estado de Jalisco, 2015) y nueve años después, el sistema cuenta con más de tres mil vehículos y más de 300 estaciones repartidas en los municipios de Guadalajara, Zapopan y Tlaquepaque (MiBici, s.f.-a), como se muestra en la **Figura 1**.

Figura 1

Distribución de las estaciones del sistema MiBici en 2023

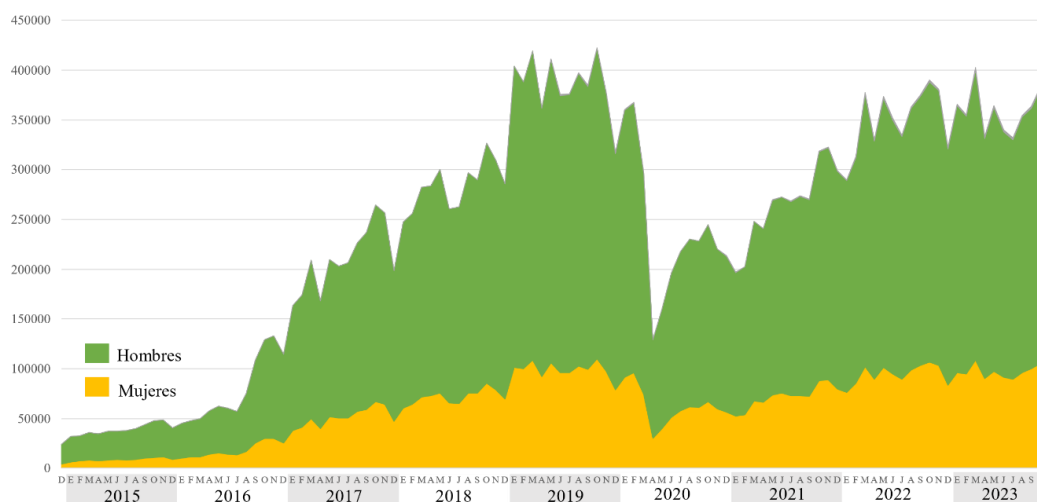


Fuente: Elaboración propia con información de MiBici (s.f.-b).

En este lapso, el sistema ha acumulado más de 25.8 millones de viajes (MiBici, s.f.-b). La **Figura 2** muestra la tendencia de aumento en el uso hasta principios de 2020, cuando la pandemia SARS-CoV-2 y las medidas de aislamiento y restricción derivadas de ella interrumpieron este crecimiento. Posterior a ello, aunque la cantidad de viajes va en ascenso, aún no se logran alcanzar los niveles de uso prepandemia.

Figura 2

Histórico de viajes realizados en el sistema de diciembre de 2014 a octubre de 2023



Fuente: Elaboración propia con datos de MiBici (s.f.-b).

La **Figura 2** también revela que, en los casi 10 años del sistema, este se ha usado predominantemente por hombres, pues las mujeres han hecho en promedio únicamente el 26% de los viajes. Este desequilibrio ha llevado a la implementación de esfuerzos para aumentar la participación femenina como el programa *Mi Pasaje Apoyo a Mujeres*, el cual beneficiaba a jefas de familia con una membresía gratuita (Jalisco, 2021).

Actualmente, el sistema ofrece dos tipos de membresías: anual por \$457, elegida por más 86 mil personas y temporal, con opciones de 1 día por \$100, 3 días por \$200 y 7 días por \$351. Ambos tipos permiten viajes ilimitados de 30 minutos en un horario de 5:00 a 00:59 (MiBici, s.f.-c).

3.1 Información del SBC

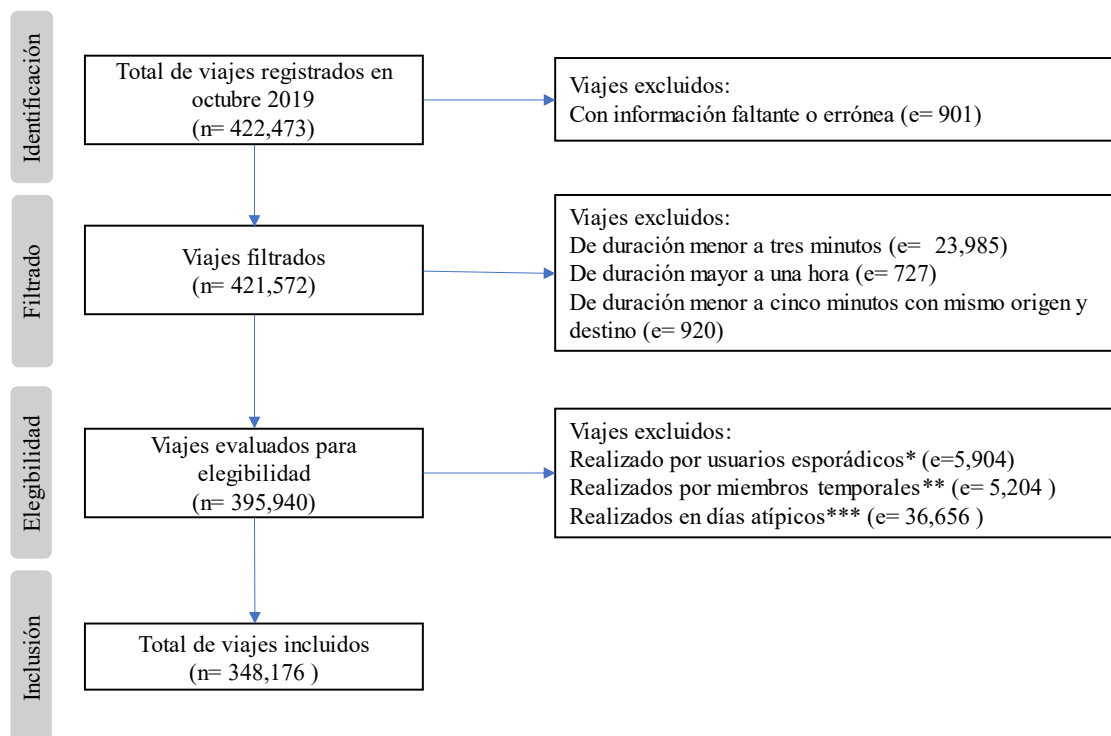
El sitio oficial de datos abiertos del sistema MiBici, ofrece información disponible en formato .csv agrupada por mes y por año. Las observaciones de cada conjunto de datos son los viajes realizados en el mes que incluyen la información de las estaciones de origen y destino, el tiempo de inicio y fin, el género y edad del usuario, así como claves únicas de identificación del usuario y del viaje.

Para investigar los factores que influyen en la demanda por género, se obtuvieron los datos de todos los viajes realizados en octubre del 2019 debido a que, por un lado, fue el mes que registró una mayor cantidad de viajes no solo en 2019 sino en toda la historia del sistema. Y, por otro lado, este mes es cercano a la fecha de levantamiento de la encuesta de percepción de WRI, lo cual es de utilidad con fines de comparación y/o validación.

En octubre de 2019, se registraron 422,473 viajes. Se realizó una limpieza de estos datos para eliminar registros erróneos, incompletos o que quedan fuera del ámbito de este estudio, este preprocesamiento de datos está ilustrado en la **Figura 3**. Para la parte de filtrado de datos, se consideraron criterios similares a los empleados por Nickkar et al (2019) para eliminar registros que puedan representar problemas de anclaje, devolución inmediata por problemas con el vehículo o incluso por alguna operación de balanceo.

Figura 3

Diagrama de flujo del preprocesamiento de datos

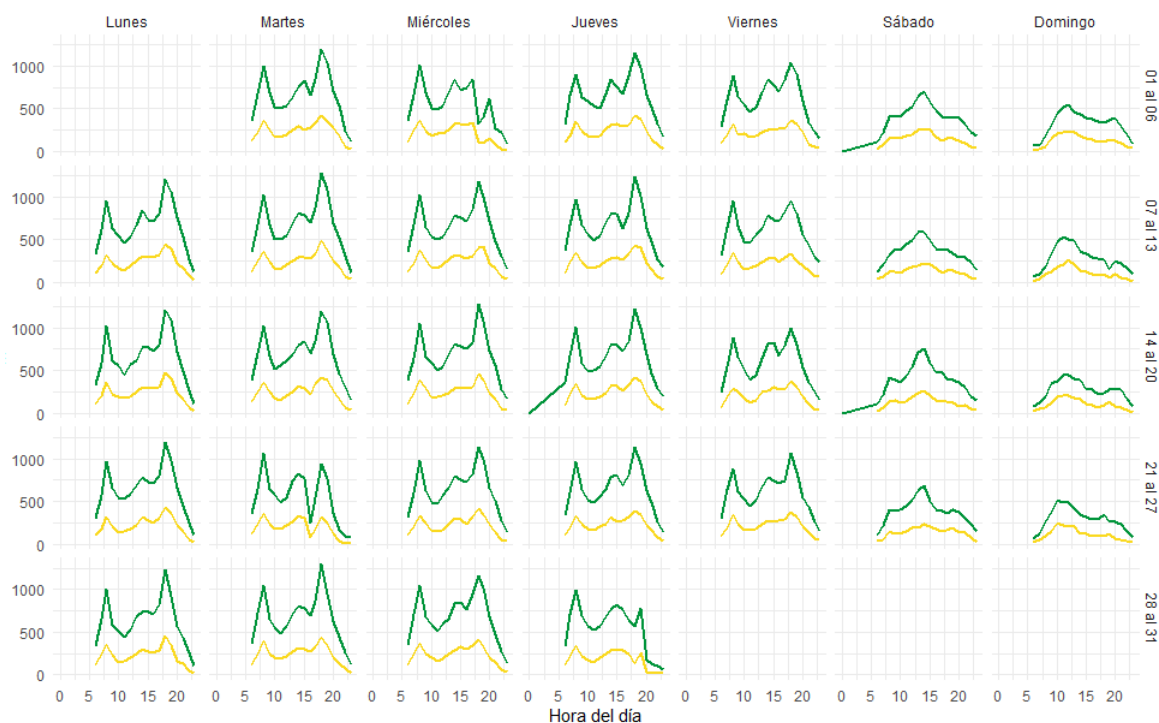


Notas: *Realizan menos de 3 viajes al mes, **realizan viajes únicamente en un lapso de hasta 7 días contiguos. ***miércoles 2, martes 22 y jueves 31. Fuente: Elaboración propia.

A continuación, para evaluar la elegibilidad, se ha demostrado un comportamiento diferenciado entre usuarios temporales y miembros anuales (Vogel et al., 2014; Faghih-Imani et al., 2017; Sun et al. 2018; Jain et al, 2018), por lo que se excluyen los viajes realizados por usuarios que todos sus viajes se hayan realizado únicamente en un lapso de 7 días contiguos. Así mismo, considerando la **Figura 4** que presenta la distribución de la cantidad total de viajes por día, por hora y por sexo, donde además de observarse las diferentes dinámicas entre días típicamente laborales y fines de semana, así como entre horas pico y horas valle, permite identificar días con un comportamiento atípico para excluir todos esos viajes.

Figura 4

Distribución de la cantidad de viajes por hora y por sexo para el mes de octubre 2019



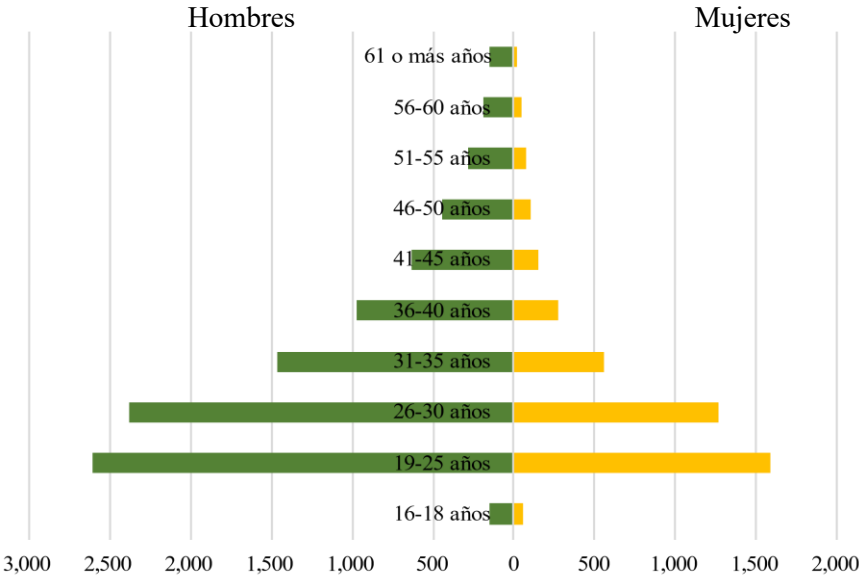
Fuente: Elaboración propia con datos de MiBici (s.f.-b).

El preprocesamiento de los datos tiene como objetivo la estandarización para optimizar el desempeño del algoritmo. En consecuencia, para la implementación del algoritmo se trabajó con 348,176 viajes realizados por 13,404 usuarios considerados miembros anuales y usuarios frecuentes del sistema. Del total, el 30.8% corresponde a mujeres y el 69.2% a hombres, resultando en un promedio diario de 3,231 viajes femeninos y 9,204 masculinos. La **Figura 5**

muestra la distribución por sexo y edad de los usuarios, para ambos sexos, el grupo de 16 a 18 años es reducido, posiblemente debido a la necesidad de tener tarjeta de crédito o débito para el pago del servicio. En cambio, el grupo de 19 a 30 concentra el grueso de la población para ambos sexos, posteriormente se observa un comportamiento inversamente proporcional, donde a mayor edad, menor cantidad de usuarios de ese grupo.

Figura 5

Pirámide de edades de los usuarios anuales del sistema Mibici en octubre 2019

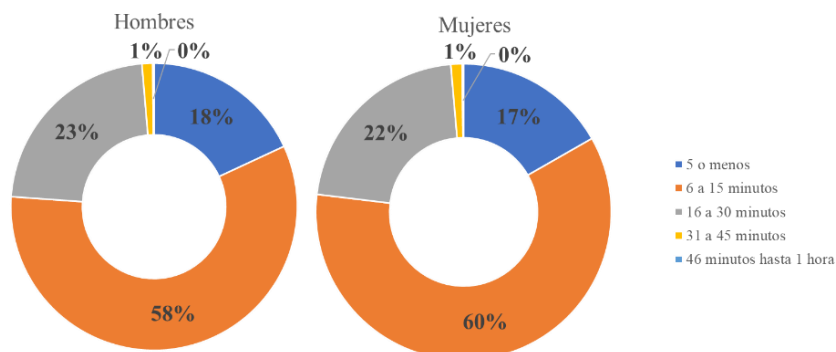


Fuente: Elaboración propia con datos de MiBici (s.f.-b).

Considerando lo expuesto en la revisión de literatura sobre las disparidades de género en la movilidad, se presenta la **Figura 6**, la cual muestra la duración de los viajes según el género. Se destaca que, tanto para hombres como mujeres, más de la mitad de los viajes tienen una duración de 6 a 15 minutos. En los intervalos de viajes con una duración inferior a 5 minutos y de 16 a 30 minutos, la proporción de hombres es apenas supera en un punto porcentual al de las mujeres. En otras palabras, la diferencia en cuestión de duración de viaje, y posiblemente en distancia, entre hombres y mujeres es casi nula.

Figura 6

Proporción de viajes por sexo según su duración



Fuente: Elaboración propia con datos de MiBici (s.f.-b).

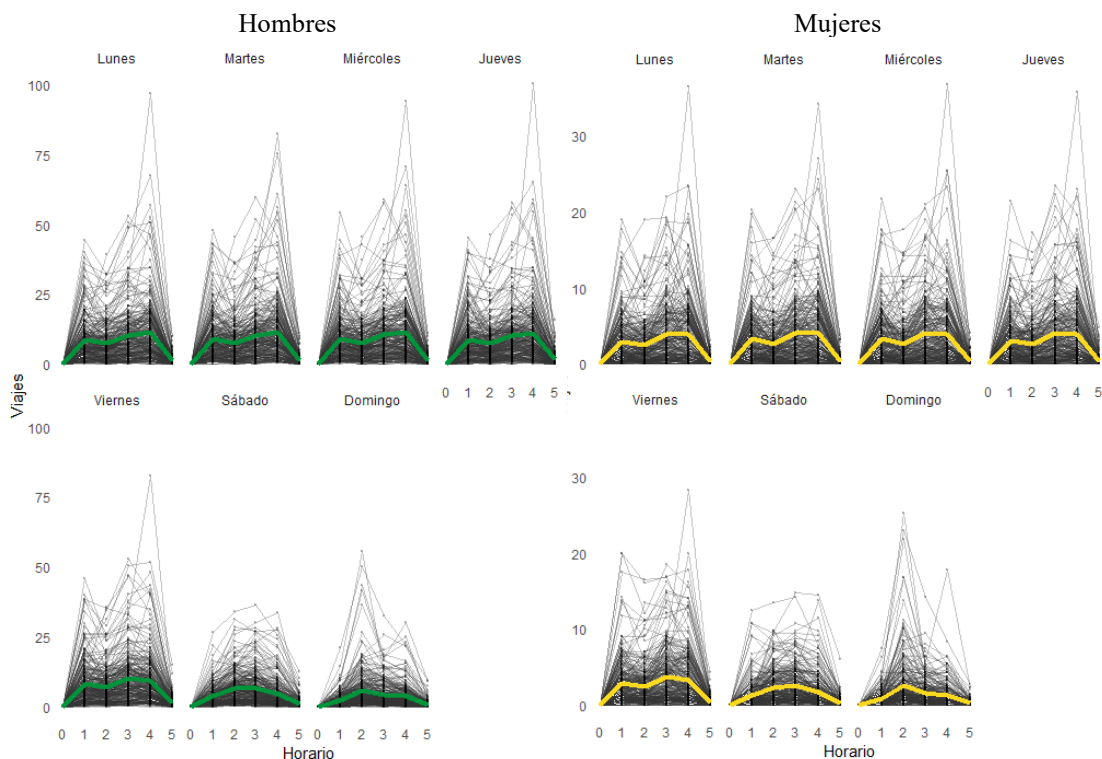
Retomando lo presentado en la **Figura 6**, se observa que, en los días típicamente laborales la actividad del sistema está estrechamente vinculada con los horarios de entrada (aproximadamente a las ocho) y salida (alrededor de las seis), mientras que los sábados y domingos, no solo disminuye drásticamente el uso del sistema, sino que los picos de actividad no son tan pronunciados. Además, se destaca que, en el caso de las mujeres, estos picos no son tan marcados, indicando un uso más estable a lo largo de las horas.

Considerando estos patrones de viaje a lo largo del día y para mitigar el riesgo de sobreajuste en la creación del modelo con RF, se llevó a cabo una categorización con base en trabajos similares (El-Assi et al, 2017; Du et al., 2019; Zhou et al., 2019; Guidon et al., 2020). Cada día se dividió en seis segmentos horarios: Madrugada (12:00 am – 5:59 am); Pico Matutino (6:00 am – 9:59 am); Mediodía (10:00 am – 1:59 pm); Tarde (2:00 pm – 5:59 pm); Pico Vespertino (6:00 pm – 9:59 pm) y Noche (10:00 pm – 11:59 pm).

La **Figura 7** presenta una visión detallada del comportamiento por día y horario de los viajes. Este enfoque tipo “grafico de espagueti” ofrece una representación visual de las trayectorias de las 274 estaciones y el promedio del sistema. Aunque la correspondencia exacta de cada trayectoria a una estación específica no es posible debido a la densidad de datos, se destaca que, tanto para hombres como para mujeres, algunas estaciones muestran un comportamiento significativamente diferente al promedio general.

Figura 7

Cantidad de viajes por cada estación y promedio según día de la semana y horario definido por género



Fuente: Elaboración propia con datos de MiBici (s.f.-b).

Este hallazgo motiva la investigación más profunda sobre factores subyacentes que generan estas variaciones. En consonancia con las observaciones de Du et al (2019) sobre la heterogeneidad de los patrones de uso de los SBP donde algunas áreas experimentan escaso uso mientras que otras registran una alta demanda, se plantea la hipótesis de que la frecuencia de uso guarda una fuerte correlación con factores geográficos del entorno.

3.2 Información del entorno

El portal oficial de datos abiertos de MiBici no solo ofrece información de los viajes, sino que también proporciona una base de datos que incluye detalles sobre la ubicación geográfica de cada estación. La información se cargó en el sistema de información geográfica QGIS,¹ donde

¹ <https://qgis.org/es/site/>

se ejecutó el proceso de geometría vectorial para crear un buffer de 400 metros alrededor de cada estación, definido a partir del concepto de “radio caminable” que ONU-Hábitat (2022) define como “un círculo con un radio de 400 m. que indica un umbral práctico y realista de la distancia máxima en que el espacio público debe encontrarse de las personas”. En la **Figura 8** se ilustra un ejemplo.

Figura 8

Ejemplo de buffer de 400 metros de radio trazado a partir de cada estación del sistema



Fuente: Elaboración propia con datos de MiBici (s.f.-b); INEGI (s.f. y 2020), Gobierno de Jalisco (s.f.) e IMEPLAN 2019).

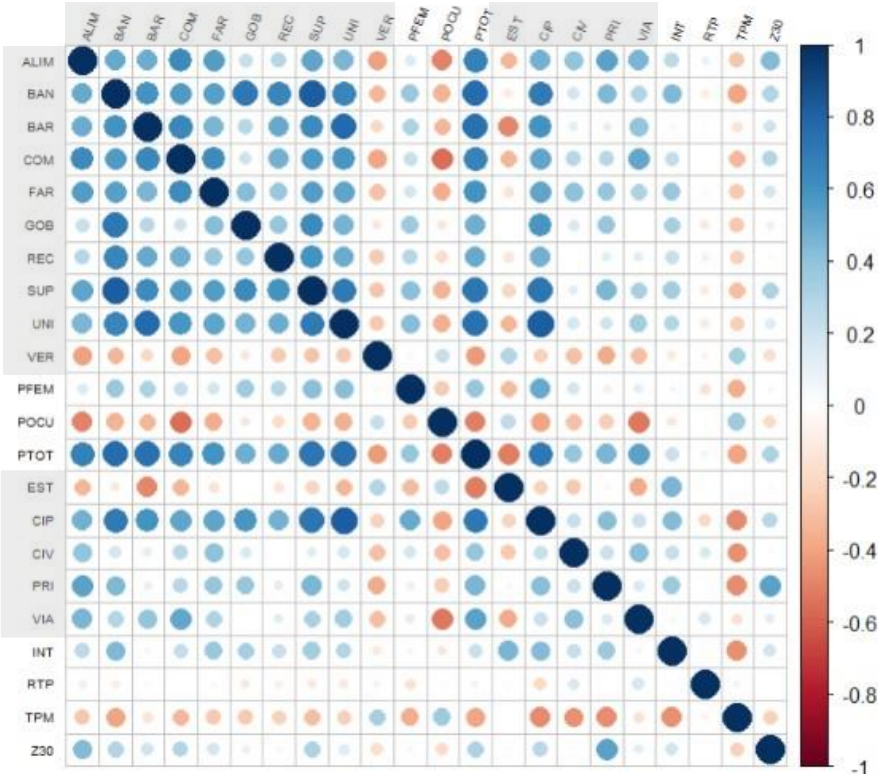
Asimismo, la **Figura 8** presenta la traza urbana, así como puntos de interés como bancos, escuelas, restaurantes y bares, áreas verdes y estaciones de transporte público, entre otros. Estos elementos son considerados posibles factores que afectan la demanda de los SBC según la literatura previa. Estos datos se recopilaron de información proporcionada tanto por el gobierno de Guadalajara e IMPEPLAN e INEGI, dando como resultado 23 variables.

Para determinar la presencia de algún tipo de equipamiento o unidad económica, se contabilizó el número total dentro del *buffer*. De manera similar, para variables como ciclovías o áreas verdes, se sumó la longitud o el área dentro del *buffer*. Respecto a cuestiones demográficas se contabilizó la información de las manzanas cuando el centroide de estas se encontraba dentro del *buffer*. Estas variables se agruparon en 4 categorías: (1) Actividades y Uso de Suelo; (2) Demográficos; (3) Infraestructura Ciclista; y (4) Transporte Público y Movilidad.

En la **Figura 9** se presentan las correlaciones entre las variables del entorno. Cabe destacar que ninguna supera una correlación de 0.85 y que las correlaciones más altas se encuentran entre variables de la misma categoría. También se observan correlaciones significativas entre la presencia de infraestructura ciclista y la cantidad de unidades económicas. En el **Anexo 1** se encuentra el detalle y estadística descriptiva de cada variable del entorno.

Figura 9

Matriz de correlaciones entre las variables del entorno



Fuente: Elaboración propia con datos de MiBici (s.f.-b); INEGI (s.f. y 2020), Gobierno de Jalisco (s.f.) e IMEPLAN 2019).

Finalmente, para construir la base de datos utilizada para el algoritmo de RF, cada viaje se considera como una observación, y se le añade toda la información de las 4 categorías mencionadas anteriormente tanto de la estación de origen como la de destino. Esto resulta en una base final de 348,176 viajes con las variables de género, día de la semana, horario y 23 variables para la estación origen del viaje y 23 para la estación destino. En otras palabras, se tiene una base de 348,176 observaciones con 49 variables en total.

3.3 Encuesta de Percepción MiBici 2019

Paralelamente a la construcción del conjunto de datos con información del sistema MiBici, se gestionó a través de la Plataforma Nacional de Transparencia² la base de datos de respuestas de la Encuesta de Percepción MiBici 2019. Esta encuesta forma parte del estudio realizado por WRI México, el Instituto Metropolitano de Planeación (IMPELAN), BKT Bici Pública SA. de C.V y el extinto Instituto de Movilidad y Transporte del Gobierno del Estado de Jalisco. La base de datos incluye las respuestas de 3,371 personas, con un 67% hombres y un 31% mujeres, recopiladas en línea durante el periodo del 12 al 23 de diciembre del 2023.

La encuesta recaba información demográfica (género, edad, grado de estudios, ocupación, residencia e ingresos), aspectos relacionadas con la movilidad de los encuestados (principal medio de transporte, gasto en transporte y acceso a los diferentes medios), detalles sobre los viajes realizados en el sistema (propósito, destinos, tiempo, frecuencia, alternativas), opiniones (ventajas, desventajas, satisfacción, áreas de mejora, obstáculos, percepción de seguridad, experiencias), entre otros temas.

Aunque no constituye una muestra representativa, resulta valiosos conocer los resultados, sirviendo, así como un marco de referencia para interpretar la selección de variables más significativas generadas por el algoritmo RF.

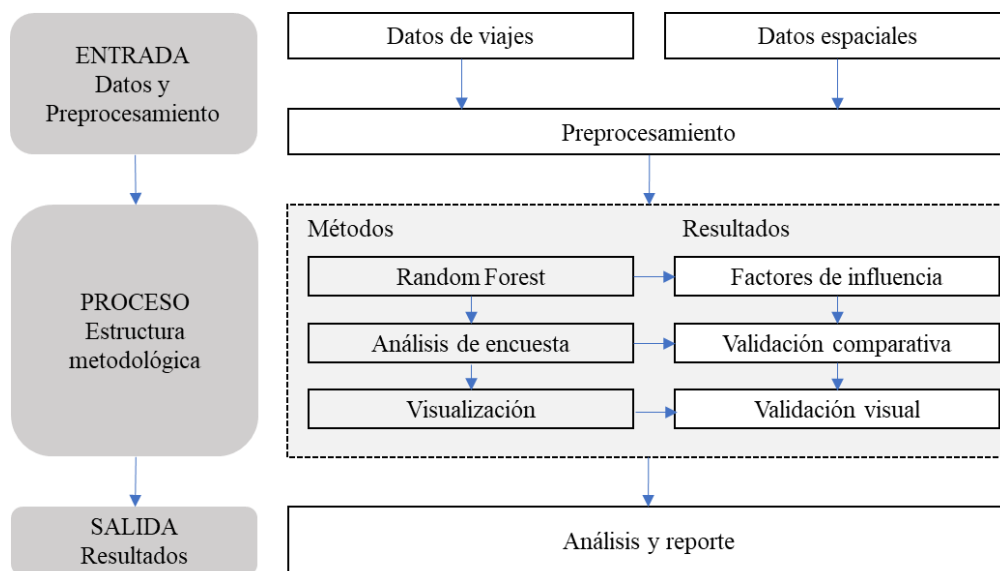
² Solicitud de información 142411722000383

4. Metodología y desarrollo

Para lograr la identificación, comprensión y validación de los factores que impactan de manera diferenciada la demanda del sistema MiBici, se propone una estructura metodológica de tres escalas. La estructura incorpora (i) la aplicación de algoritmos *Random Forest* para obtener la importancia de las variables, (ii) análisis y contraste con los resultados de la encuesta de percepción de WRI, y (iii) visualización de los resultados principales. En la **Figura 10** se muestra un diagrama esquemático del proceso.

Figura 10

Diagrama de flujo de la estructura metodológica propuesta



Fuente: Elaboración propia.

4.1 Random Forest

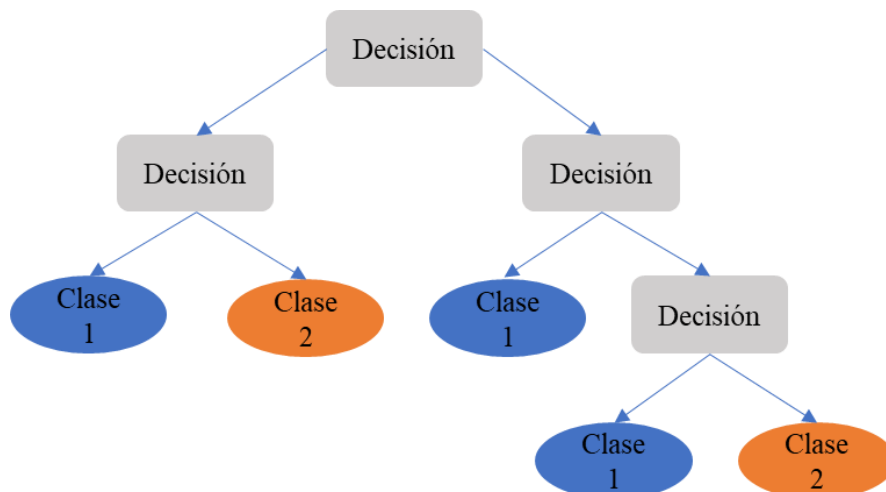
Random Forest (RF) o Bosques Aleatorios, es un algoritmo de aprendizaje supervisado no paramétrico para resolver problemas de clasificación y regresión (Breiman, 2001). Este algoritmo funciona eficazmente con grandes volúmenes de datos, además no se requiere una suposición sobre la distribución estadística de estos, haciéndolo adecuado en el caso de relaciones no lineales, interacciones categóricas y colinealidad entre variables. (Catani et al., 2013; Ashqar et al., 2020; Du et al., 2019; Strobl, et al., 2007).

RF se clasifica dentro de los métodos de ensamble (*ensemble methods*), que consisten en combinar las decisiones de varios modelos simples para mejorar el rendimiento final, ya sea por medio de *bagging* o de *boosting* (Genuer et al, 2010; Strobl et al, 2007; Zhou et al., 2019). Además, RF utiliza la técnica conocida como agregación de Bootstrap (*bagging*), donde se realiza un muestro del conjunto de datos de entrenamiento con reemplazo (Zhou et al., 2019).

En el contexto de RF, se combinan un conjunto de árboles de clasificación o regresión (CART, por sus siglas en inglés) (Hasan et al., 2016; Genuer et al., 2010). Estos árboles son estructuras jerárquicas que se construyen particionando recursivamente la muestra, desde el nodo raíz hasta el nodo terminal, generando grupos cada vez más homogéneos (Grömping, 2009). La **Figura 11** ilustra el proceso de los árboles, donde cada nodo representa una decisión basada en características elegidas al azar, buscando la mejor división según un criterio, como la ganancia de Gini, que maximiza la reducción de la impureza global de los nodos buscando hacer cada vez más homogéneos los grupos. La selección de la siguiente variable de división es condicional a las variables de predicción anteriores en la misma rama del árbol, pero independiente de otras. (Yang et al., 2016; Grömping, 2009; Strobl, et al., 2007).

Figura 11

Ejemplo de un árbol de decisión de 4 capas



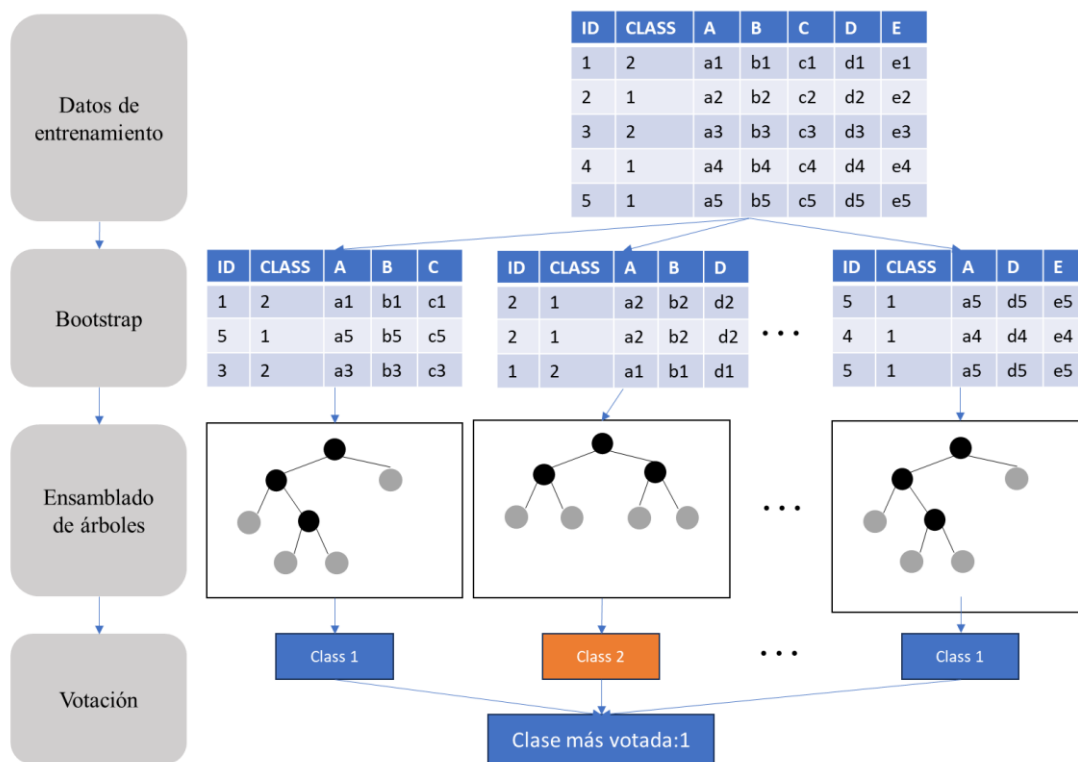
Fuente: Elaboración propia con información de Yang et al. (2016).

Ahora, considerando que RF combina varios de estos árboles, el proceso para su creación, resumido en los siguientes 5 pasos (Grekousis, et al., 2022), se presenta en la **Figura 12**:

1. De un conjunto de entrenamiento dado, se extraen n muestras aleatoriamente con reemplazo (muestra *bootstrap*).
2. De cada muestra con k predictores, se selecciona aleatoriamente un subconjunto ($m < k$) y se desarrolla un CART.
3. Cada árbol crece hasta su mayor extensión sin poda.
4. Se calcula el resultado de predicción clasificación de cada árbol.
5. Se utiliza la clase/voto más frecuente para crear el resultado final.

Figura 12

Ejemplo del funcionamiento de un Random Forest



Fuente: Elaboración propia con información de Humboldt-Universität zu Berlin (2023).

Un RF es aleatorio por dos razones, por un lado, cada árbol se basa en un subconjunto aleatorio de observaciones y por otro, cada división dentro de cada árbol se crea basándose en un subconjunto aleatorio de variables (Grömping, 2009; Sathishkumar y Cho, 2020; Strobl et al., 2007). A pesar de que cada árbol es inestable, en conjunto contribuyen a lograr un sesgo y varianza relativamente bajos (Ngo, 2022) así como a minimizar el riesgo de sobreajuste (Guidon et al. 2020).

Adicional a su capacidad para predecir y clasificar, RF lleva a cabo una selección directa de características durante el proceso de clasificación (Hasan, et al., 2016). Por consiguiente, se emplea extensamente no solo como una herramienta predictiva, sino también como una técnica eficaz para evaluar y clasificar la influencia de diversos parámetros del modelo (Yang, et al., 2016; Grömping, 2009; Grekousis, et al., 2022; Guidon et al., 2020; Ashqar et al., 2020; Du, et al., 2019; Sathishkumar y Cho, 2020; Genuer et al., 2010; Hasan, et al., 2016).

Existen varias métricas para medir la importancia, siendo dos de las más utilizadas el índice de importancia derivada de Gini y el índice de importancia por permutación (PIM). La primera de ellas se basa en la impureza del nodo de Gini que cuantifica la heterogeneidad del conjunto de datos y se calcula a través de la contribución de cada variable a la reducción de la impureza de Gini a lo largo de todos los nodos en los que se utiliza para dividir (Hasan et al., 2016), de modo que cuanto mayor sea la reducción de impureza que aporta una variable, más crucial es para la toma de decisiones del modelo en el proceso de clasificación.

En paralelo, el índice de importancia de permutación (PIM) se centra en evaluar cómo la alteración de los valores de una variable impacta en la predicción del modelo (Strobl, et al., 2007; Sathishkumar y Cho, 2020). Considerando que RF construye los árboles con muestras Bootstrap, resulta que deja un conjunto que no es usado para el crecimiento de ese árbol conocido como “fuera de la bolsa” (*out of the bag - OOB*) (Archer & Kimes, 2008). Para medir la importancia de una característica específica en el árbol, se barajan aleatoriamente los valores de esta característica en las muestras OOB y se compara la precisión de la clasificación entre las muestras OOB intactas y las muestras OOB con la característica particular (Hasan et al., 2016).

Strobl et al. (2007), encontraron que las medidas de importancia de las variables se ven afectadas por el número de categorías y la escala de medición de las variables predictoras. Cuando se trata de variables continuas o con un gran número de clases, las importancias basadas en impurezas pueden inducir el error (Scikit-learn Developers, s.f.). En este escenario, las variables que representan distancias o áreas podrían verse favorecidas en términos de importancia en comparación con la variable que indica el número de estaciones dentro de un radio. Esto se debe a que las primeras pueden presentar hasta 274 valores diferentes, mientras que la variable de

estación solo cuenta con 28 posibles valores. Sin embargo, para mayor experimentación, en este trabajo se aplican ambas medidas.

4.2 Análisis de la encuesta

En contraste con otros enfoques que se centran en comparar el rendimiento del algoritmo RF con otras técnicas o variaciones internas mediante medidas como R2, MSE, MAE, F1, *Accuracy*, entre otras (Yang, et al., 2016; Bacciu, 2017; Feng y Wang, 2017; Wang y Kim, 2018; Ashqar et al., 2020; Ruffieux et al., 2017), en este trabajo se busca analizar la relevancia de las variables descriptivas, más allá de construir un modelo con precisión alta.

Por lo tanto, la metodología propuesta se diferencia al incorporar un análisis comparativo con otra fuente de datos, en este caso, los resultados de una encuesta de percepción. Aunque la encuesta no es de una muestra representativa, su utilidad radica en proporcionar un conjunto externo de datos que puede servir para contrastar y, en cierta medida, validar los resultados obtenidos mediante el RF.

La capacidad de desglose de respuestas por género y otras variables demográficas proporcionada por la encuesta es esencial para contextualizar y validar los resultados. Aunque las fechas de recopilación son ligeramente diferentes, la temporalidad cercana entre la encuesta y los datos del RF sugiere una representatividad contextual, contribuyendo así a la solidez y relevancia de nuestras conclusiones en relación con los factores urbanos que afectan la demanda diferenciada de sistemas de bicicletas públicas entre hombres y mujeres. Así mismo la implementación de este análisis comparativo ofrece la oportunidad de generar nuevas hipótesis permitiendo una aproximación integral, más profunda y contextualizada.

4.3 Análisis visual y cualitativo

La visualización de la información es igualmente una herramienta muy poderosa. En este último método se busca una validación visual de los resultados obtenidos a través del RF. Esta validación se materializa mediante la creación de mapas que representan gráficamente la distribución y concentración de los viajes realizados tanto por hombres como mujeres, así como la presencia espacial de los factores identificados como relevantes por el RF. Para lograr esto, se aprovecharon los datos geospaciales con los que previamente se alimentó al modelo.

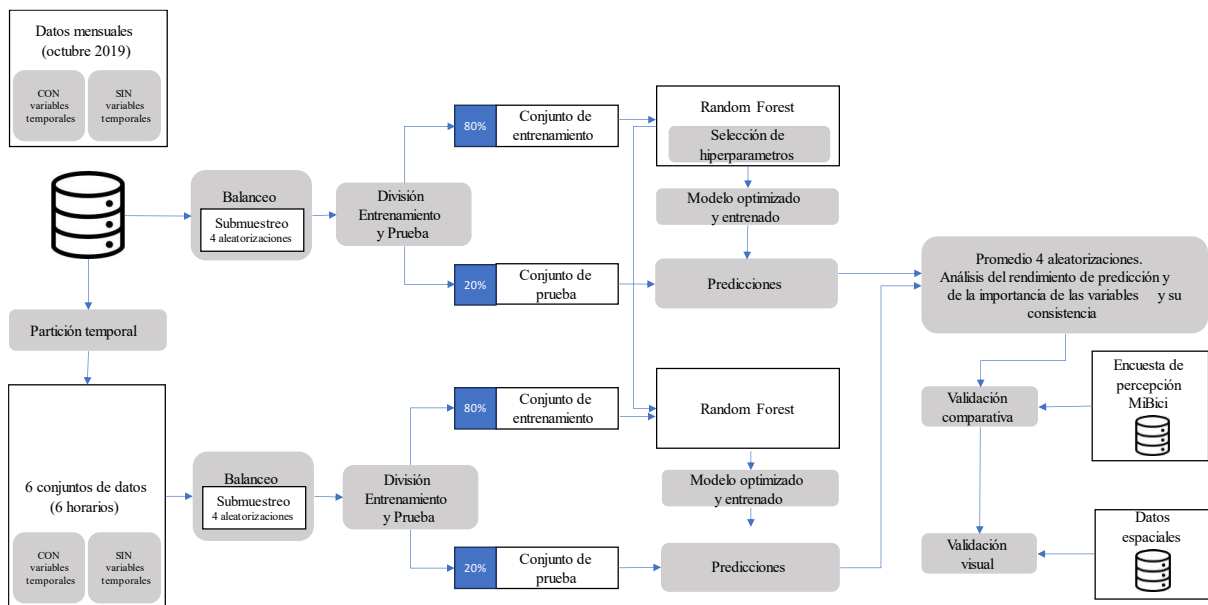
La visualización cartográfica no solo sirve como una herramienta de validación, sino que también facilita la identificación de patrones espaciales y revela posibles correlaciones visuales entre la presencia de ciertos factores urbanos y las pautas de movilidad diferenciadas por género. Este enfoque refuerza y complementa los resultados cuantitativos del RF, proporcionando una comprensión más holística y contextualizada de los factores urbanos que contribuyen a las disparidades en la demanda de sistemas de bicicletas públicas entre hombres y mujeres.

4.4. Desarrollo del modelo

La **Figura 13** presenta todo el procedimiento experimental de este estudio. La investigación utiliza dos tipos de datos: mensuales y por horario, en los que los datos mensuales se dividen en 6 en función horario. Todos los modelos de clasificación se desarrollan de forma independiente empleando el lenguaje de programación Python y diversas librerías según se requiera, principalmente *scikit-learn (sklearn)*, una herramienta esencial para el aprendizaje automático.

Figura 13

Proceso metodológico



Fuente: Elaboración propia.

El modelo generado con el algoritmo de RF, tiene como objetivo clasificar si un viaje es realizado por un hombre o una mujer, empleando las variables del entorno propuestas. Con los datos recopilados de octubre de 2019, se preparan dos modelos con datos mensuales: uno que

considera las variables de día y horario, y otro que las excluye. Para los datos segmentados por horario, se descarta el horario de madrugada debido a la falta de observaciones y se preparan algunos modelos con la variable del día de la semana y otros sin ella, lo que suma un total de 10 modelos.

Dada la naturaleza del problema abordado, es necesario equilibrar la base de datos para evitar sesgos que afecten negativamente el proceso de generalización y puedan perjudicar a la clase minoritaria, en este caso, las mujeres. Por lo tanto, se realiza un submuestreo aleatorio de los viajes realizados por hombres, repitiendo esta acción 4 veces para cada uno de los 12 modelos, lo que garantiza la consistencia y la robustez de los resultados. Una vez desarrollados los 12 modelos con los cuatro diferentes conjuntos balanceados, se procede a compararlos, promediarlos y analizar detalladamente la importancia de las variables en cada configuración.

Para llevar a cabo los diferentes ejercicios, el proceso de entrenamiento y prueba se lleva a cabo de la siguiente manera. Con la muestra balanceada, se codifican las variables y se normalizan todos los valores. Los conjuntos de entrenamiento y prueba se conforman por la división aleatoria de los datos en dos secciones: 20% para prueba y 80% para entrenamiento. La **Tabla 3** presenta las dimensiones de los diferentes conjuntos de datos.

Tabla 3

Número de observaciones y su proporción según género de los distintos conjuntos de datos

Modelo	Inicial (H/M)	Balanceado (H/M)	Train	Test
Mensual	348,176 (74/26)	180,916 (50/50)	144,732	348,176
Madrugada	2 (100/0)	-	-	-
Pico Matutino	73,695 (74/26)	37,940 (50/50)	30,352	7,588
Mediodía	73,814 (74/26)	38,616 (50/50)	30,892	7,724
Tarde	95,022 (73/27)	51,458 (50/50)	41,166	10,292
Pico Vespertino	93,548 (74/26)	48,236 (50/50)	38,588	9,648
Noche	12,095 (81/19)	4,666 (50/50)	3,732	934

Fuente: Elaboración propia.

Para obtener un mejor rendimiento del algoritmo RF, es el usuario quien define los parámetros de número de árboles de decisión y número de parámetros seleccionadas para cada árbol (Ashqar et al., 2020). En este caso, para la definición de los hiperparámetros óptimos se usó la función de GridSearchCV la cual emplea el algoritmo de optimización combinatoria *Grid Search*. Con

ello, los mejores parámetros se establecieron, para el número de estimadores o árboles fue de mil, para el máximo número de parámetros para tener en cuenta en las particiones resultó *'sqrt'*, es decir que se usarán el cuadrado de la cantidad de características para cada división, el criterio para las particiones debía ser "*Gini*" y la profundidad máxima de 12, esto para evitar el sobreajuste.

5. Resultados y discusión

5.1 Análisis del desempeño del modelo

En esta sección, se evalúa el desempeño promedio de los modelos utilizando cuatro métricas: *accuracy*, *precision*, *recall* y *F1*. Estas métricas se derivan de la matriz de confusión, la cual compara las etiquetas verdaderas con las predichas por el algoritmo de clasificación categorizándolas en verdaderos positivos (*VP*) y verdaderos negativos (*VN*) donde coinciden la etiqueta verdadera con la etiqueta de la clasificación y, falsos positivos (*FP*) y falsos negativos (*FN*) donde se encuentran errores de clasificación.

Zhou et al. (2019) explican que la métrica de *Accuracy* o precisión se refiere a la proporción de observaciones predichas correctamente respecto al total de observaciones ($\frac{VP+VN}{VP+FP+VN+FN}$). *Precision* o precisión, similar a la anterior, se refiere a la proporción de clasificaciones que coinciden con la clase positiva real todas las predicciones positivas producidas por el modelo ($\frac{VP}{VP+FP}$). *Recall* o sensibilidad es la proporción de observaciones positivas que fueron identificadas como tal ($\frac{VP}{VP+FN}$). *F1 score* es interpretada como el promedio ponderado de *precisión* y *recall* ($2 * \frac{Precision*Recall}{Precision+Recall}$). Dichas métricas son obtenidas tanto para los conjuntos de entrenamiento como para los de prueba para identificar sobreajuste o subajuste y para evaluar el rendimiento y la capacidad de generalización de los modelos.

Las Tablas 4 y 5 presentan las métricas resultantes tanto en el conjunto de entrenamiento como en el de prueba. Se observa que, en todos los casos, el rendimiento con el conjunto de prueba es inferior. En la etapa de entrenamiento el rendimiento de los modelos mensuales es el más bajo respecto a los otros horarios, sin embargo, en la etapa de prueba se mantiene casi en el mismo puntaje a diferencia del resto. El modelo nocturno muestra la mayor diferencia entre las métricas de ambas etapas dado que tiene un rendimiento casi perfecto durante el entrenamiento para luego experimentar una caída drástica en la fase de prueba, siendo esto una posible evidencia de sobreajuste derivado de la cantidad de observaciones de los diferentes conjuntos.

Tabla 4*Métricas en el conjunto de entrenamiento*

	Con Variables Temporales				Sin Variables Temporales			
	<i>Accuracy</i>	<i>Precision</i>	<i>Recall</i>	<i>F1</i>	<i>Accuracy</i>	<i>Precision</i>	<i>Recall</i>	<i>F1</i>
Mensual	68%	72%	60%	66%	67%	71%	58%	64%
Pico Matutino	83%	88%	77%	82%	82%	87%	76%	81%
Mediodía	77%	82%	74%	78%	77%	80%	73%	76%
Tarde	75%	79%	68%	73%	74%	77%	67%	72%
Pico Vespertino	76%	80%	68%	74%	74%	79%	67%	72%
Noche	92%	96%	87%	91%	90%	94%	86%	90%

Fuente: Elaboración propia.**Tabla 5***Métricas en el conjunto de prueba*

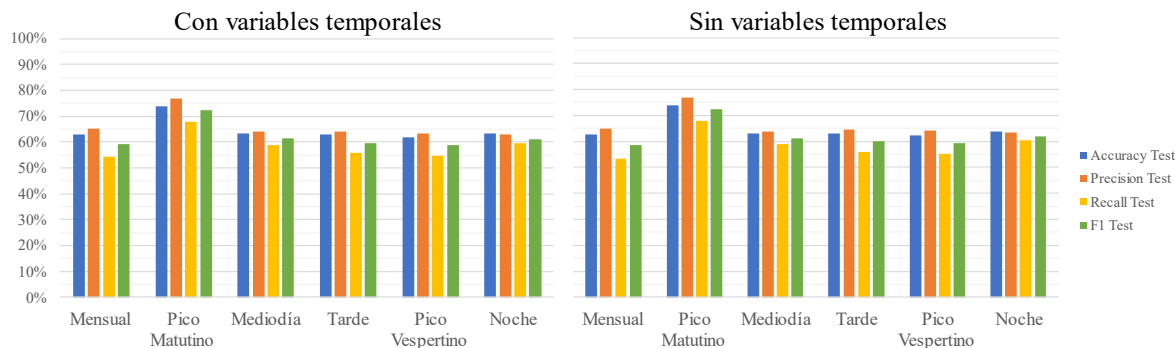
	Con Variables Temporales				Sin Variables Temporales			
	<i>Accuracy</i>	<i>Precision</i>	<i>Recall</i>	<i>F1</i>	<i>Accuracy</i>	<i>Precision</i>	<i>Recall</i>	<i>F1</i>
MENSUAL	63%	65%	54%	59%	63%	65%	53%	59%
Pico Matutino	74%	77%	68%	72%	74%	77%	68%	72%
Mediodía	63%	64%	59%	61%	63%	64%	59%	61%
Tarde	63%	64%	56%	60%	63%	64%	56%	60%
Pico Vespertino	62%	64%	55%	59%	62%	64%	55%	59%
Noche	63%	63%	60%	61%	64%	63%	60%	62%

Fuente: Elaboración propia.

Si bien en el conjunto de entrenamiento al incluir las variables temporales se obtiene un mejor puntaje, en el conjunto de prueba esto no sucede pues los puntajes son casi iguales entre los modelos con y sin variables temporales. La **Figura 14** facilita la comparación de las métricas promedio derivadas de los modelos propuestos en la prueba (*test*). Se observa una tendencia similar en los puntajes de los modelos con y sin variables temporales. Destaca especialmente que el modelo para el pico matutino se posiciona como el único que experimenta mejoría significativa en su desempeño respecto al modelo mensual.

Figura 14

Comparativo del desempeño de las métricas para el conjunto de prueba de los diferentes modelos propuestos con y sin variables temporales



Fuente: Elaboración propia.

En promedio entre los modelos para el conjunto de prueba, se tiene un 65% de *accuracy* y un 62% de *F1*. Esto se traduce en que, en promedio, el 65% de todas las predicciones son correctas, lo cual no es un rendimiento particularmente bueno ya que presenta dificultades para clasificar correctamente. Así mismo, si hacemos un análisis más detallado considerando cada una de las clases, resulta que en promedio el *recall* para las mujeres en los modelos con variables temporales es de 71% mientras que para los hombres es de 58. En el caso de los modelos sin variables temporales se reporta un *recall* de 70% y 59% respectivamente, esta brecha indica que el modelo es mejor para clasificar a las mujeres que a los hombres pues presenta una tasa de falsos negativos más alta en hombres.

5.2 Análisis de la importancia de los factores

La importancia de las características calculadas con Impureza de Gini, se normaliza para asegurar que la suma total de las importancias sea 1, proporcionando una perspectiva relativa de la contribución de cada característica al modelo. Un valor más alto asignado a una característica específica indica mayor relevancia en la capacidad del modelo de hacer predicciones. Por otro lado, en la importancia de permutación, una mayor magnitud del valor significa que la permutación de esa característica afecta negativamente a la precisión del modelo por lo que se trata de una característica importante. Existe la posibilidad de valores negativos que indican que el modelo no depende fuertemente de esta característica. Asimismo, cabe

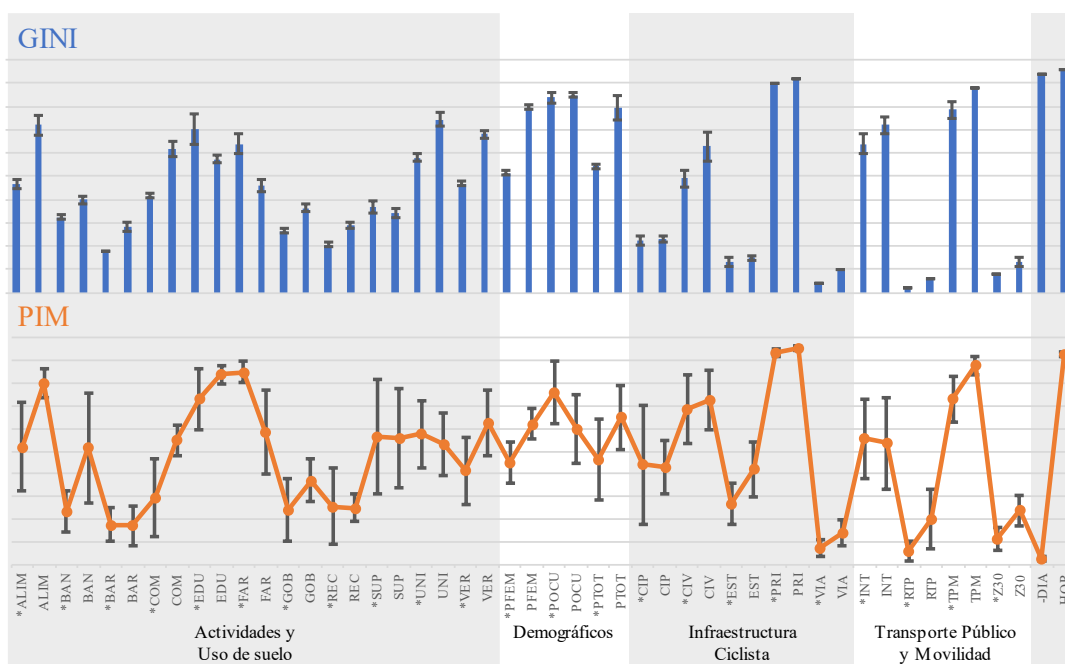
resaltar que PIM es más robusto en el sentido de que no es tan sensible ni a la distribución de los datos ni al ruido o correlaciones entre datos ni al sobreajuste como Gini.

Considerando que se tienen 4 ejercicios distintos por cada modelo, no se considera el puntaje específico de Gini o PIM, sino la posición de cada variable en relación con el total. Esto implica que la variable con el puntaje de importancia más bajo tendrá la posición 1, y a medida que la importancia aumenta, también lo hace su posición en el ranking.

La **Figura 15** permite observar el promedio y la variación de los puntajes de importancia calculados a partir de la impureza de Gini (arriba) y la permutación (abajo). Se puede observar consistencia en la métrica de GINI, mientras que con PIM se muestra una mayor variación entre los cuatro ejercicios, aun así, ambos índices coinciden en señalar como las variables de mayor importancia la presencia de calles de prioridad ciclista tanto en el origen como en el destino (*PRI y PRI) así como el horario (-HOR).

Figura 15

Variación de la importancia de las variables para los cuatro ejercicios del modelo mensual que incluyen variables temporales



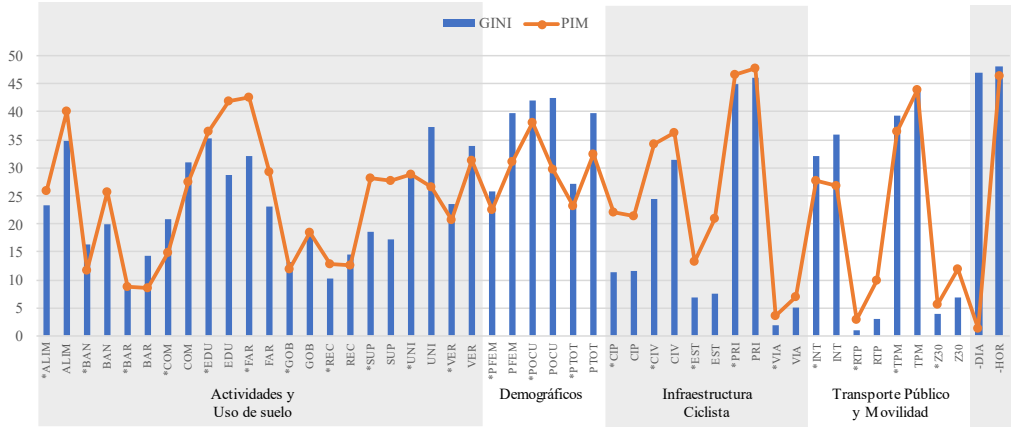
Nota: En el eje x las variables con * se refieren a variables en la estación de Origen. Fuente: Elaboración propia.

Las vialidades de prioridad ciclista son “aquellas donde las personas conductoras de vehículos no motorizados tienen prelación sobre vehículos automotores, ya sea por la segregación de un carril exclusivo o con señalamiento de uso compartido” (Reglamento de Movilidad, Transporte y Seguridad Vial de Guadalajara, 2023, Artículo 6). Considerando esta definición y la literatura existente, tiene sentido que estas vialidades tengan un impacto diferenciado, dado que las mujeres tienden a preferir circular en infraestructura ciclista más que sus contrapartes masculinas. Sin embargo, sería razonable esperar que la variable de ciclovías tuviera una importancia aún mayor debido a esta preferencia.

La **Figura 16** permite la comparación de los promedios de ambas métricas presentadas previamente. Ambos índices muestran un comportamiento similar, destacando dada su importancia, además de la variable previamente mencionada, la distancia mínima a una estación de transporte público masivo en ambos puntos (TPM y *TM), así como la población ocupada en el origen (*POCU). Estos resultados concuerdan con lo encontrado por Wang y Akar (2019), quienes hallaron una asociación negativa entre los viajes femeninos y la densidad de empleo, así como las estaciones de metro, ambas explicadas por el hecho de que los hombres, en mayor medida que las mujeres, realizan más desplazamientos para ir al trabajo y, por ende, tienen una mayor probabilidad de integrar la bicicleta con otros medios de transporte público. A su vez coinciden con los resultados de Mamdoohi y Amini (2021) que establecen que las paradas de transporte público afectan negativamente el uso de la bicicleta de las mujeres.

Figura 16

Importancia promedio de las variables del modelo mensual que incluye variables temporales



Nota: En el eje x las variables con * se refieren a variables en la estación de Origen. Fuente: Elaboración propia.

De manera similar, tanto en el origen como en el destino, las rutas de transporte público (*RTP y RTP) y la presencia de la Vía Recreativa (*VIA y VIA) no parecen ser determinantes en la clasificación del género del viajero. Aunque se esperaría que ambas variables de transporte público tuvieran un efecto negativo similar, es posible que la construcción de la variable con los metros de ruta de rutas troncales de autobuses y SITREN, no capture la información de manera tan efectiva como lo haría el número de paradas. Asimismo, la nula importancia de la existencia de la Vía Recreativa,³ podría explicarse por la baja proporción de viajes que genera en su corto horario de servicio (domingos de 8 de la mañana a 2 de la tarde).

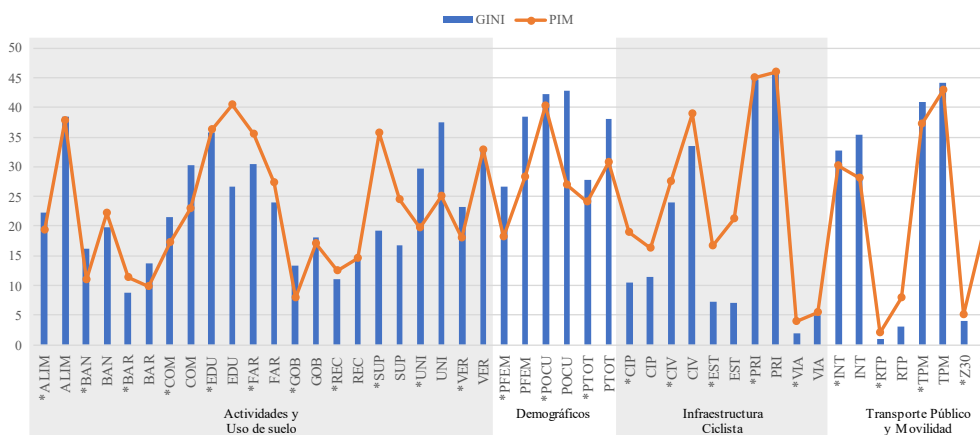
Por otro lado, la mayor discrepancia entre estas métricas se observa con la variable del día de la semana (-DIA), ya que, por impureza de Gini esta resulta ser la segunda variable con mayor importancia mientras que para el índice de permutación resulta ser la de menor relevancia para el modelo. Otras variables de importancia según el índice PIM son las pertenecientes a la categoría de Actividades y Uso de Suelo como las farmacias en el origen (*FAR) y los servicios educativos (EDU) y los restaurantes y cafeterías (ALIM), ambos en el destino.

Al repetir el ejercicio y excluir las variables temporales, resulta la **Figura 17** donde se observan resultados muy similares al ejercicio anterior, por lo que no se detallan en este documento. Además, dado que los modelos por horarios sin variables temporales mostraron un menor sobreajuste y, al igual que en el modelo mensual, la variable del día de la semana (-DIA) se posicionó entre la primera o tercera variable de menor importancia en la mitad de los ejercicios realizados donde se incluyó, se omite el análisis de los modelos por horarios que incluyen la variable temporal.

³ “Conjunto de ejes y vialidades del Municipio que, de forma temporal, en los días y horas así establecidos, se transforman en espacios para ser usados por la población para la práctica de actividades recreativas...” (Reglamento de la Vía RecreActiva del Municipio de Guadalajara, 13 de diciembre de 2005, Artículo 4).

Figura 17

Importancia promedio de las variables del modelo mensual que excluye variables temporales



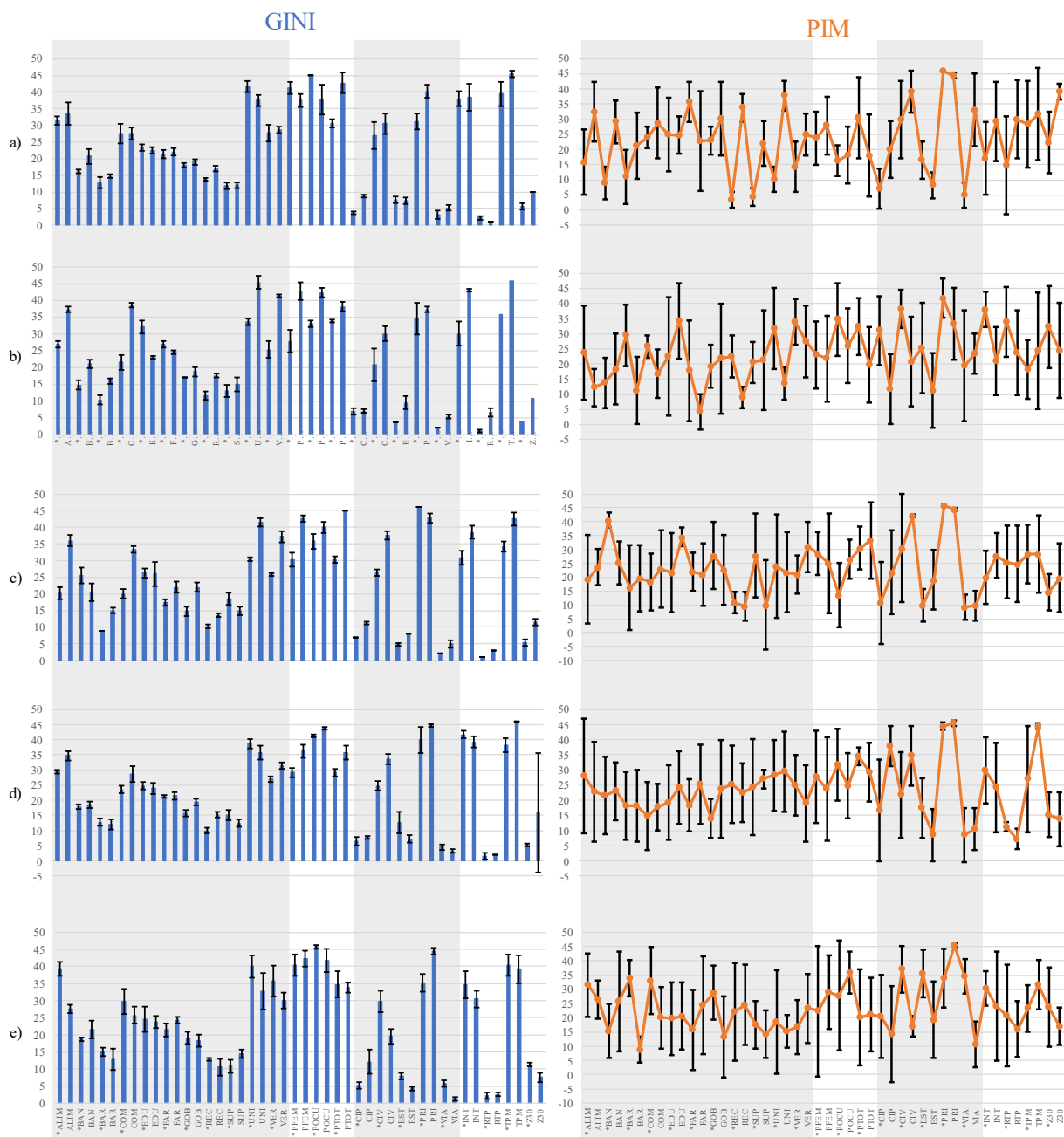
Nota: En el eje x las variables con * se refieren a variables en la estación de Origen. Fuente: Elaboración propia.

Por lo tanto, al centrarnos en los modelos por horarios que excluyen variables temporales, observamos la **Figura 18**. La primera impresión revela algunas variaciones entre los diferentes horarios y en comparación con el modelo mensual, así como una consistencia en las importancias de GINI y una marcada variación en PIM.

En todos los modelos, medidos por la métrica de impureza de Gini, la distancia mínima a una estación de transporte público masivo en el destino (TPM) se mantiene como una de las variables de mayor importancia. A su vez, todos los modelos concuerdan al identificar como variables de menor importancia las rutas de transporte público (*RTP) y la presencia de la Vía Recreativa (*VIA), especialmente en la estación de origen. En el contexto de la métrica por permutación, se observa una variación considerable entre los diferentes ejercicios, así como entre los diferentes horarios. Algunas variables, con gran importancia en un horario, pueden carecer de relevancia en otro, sin embargo, todos ellos coinciden nuevamente en presentar las calles de prioridad ciclista en el origen y destino (*PRI y PRI) como las dos variables de mayor importancia. La **Figura 19** permite comparar los promedios de ambas métricas para así identificar coincidencias entre estas.

Figura 18

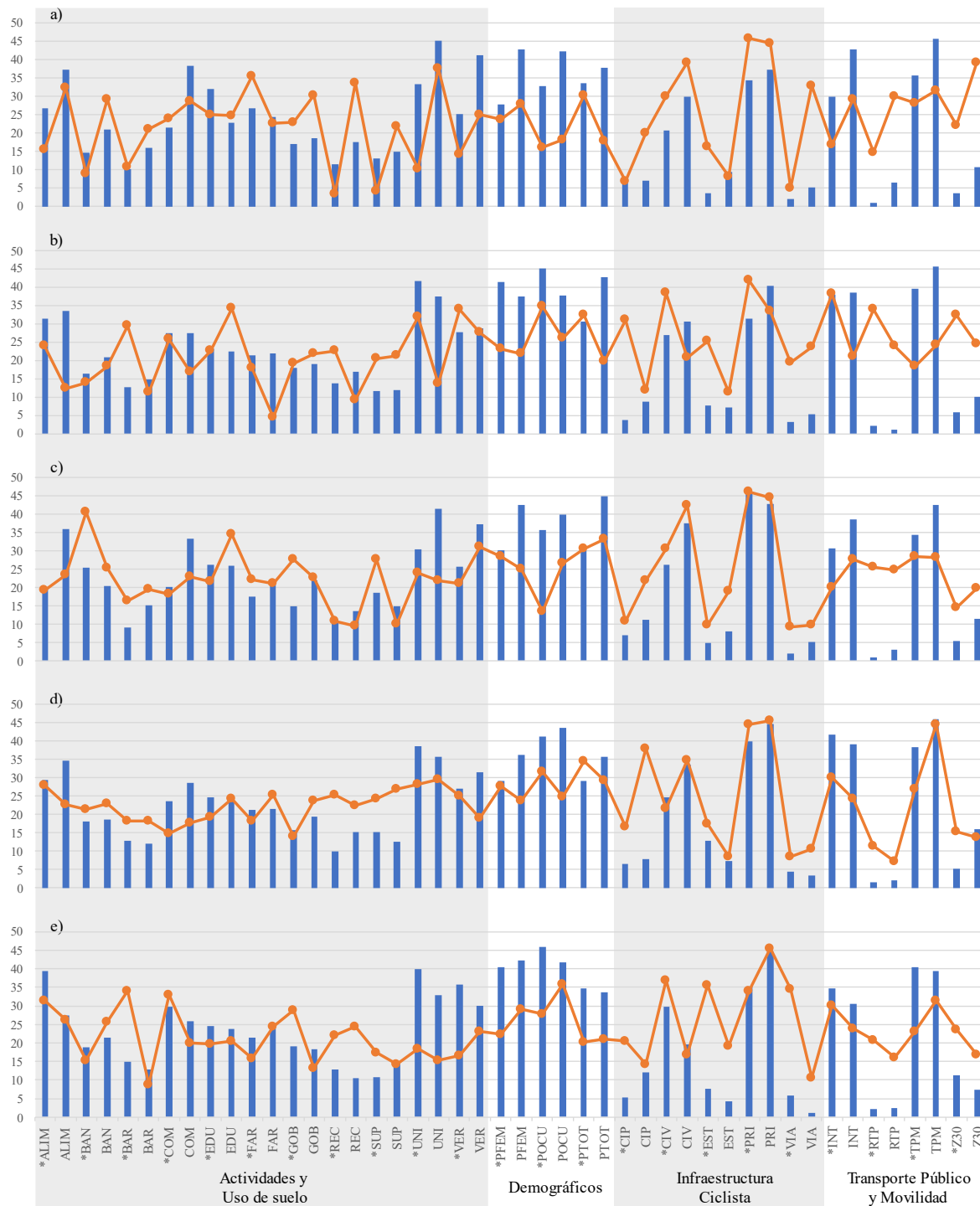
Variación de la importancia de las variables para los cuatro ejercicios de los modelos por horario que incluyen variables temporales



Notas: Los incisos se refieren a) Pico Matutino; b) Mediodía; c) Tarde; d) Pico Vespertino y e) Noche. En el eje x las variables con * se refieren a variables en la estación de Origen. Fuente: Elaboración propia.

Figura 19

Importancia promedio de las variables de los modelos por horario que incluyen variables temporales



Notas: Los incisos se refieren a) Pico Matutino; b) Mediodía; c) Tarde; d) Pico Vespertino y e) Noche. En el eje x las variables con * se refieren a variables en la estación de Origen. Fuente: Elaboración propia.

En el modelo del Pico Matutino, se destaca que las variables en el destino son más influyentes que en el origen. Por un lado, el índice que se deriva de la impureza de Gini resalta la distancia mínima a una estación de transporte público masivo como la más importante, tanto en el origen como en el destino (*TPM y TPM). Por otro lado, el índice PIM posiciona a las vialidades de prioridad ciclista como la variable principal en ambas ubicaciones (*PRI y PRI).

Otras variables relevantes según el índice GINI incluyen la distancia mínima a una universidad (UNI), el número de intersecciones viales (INT), la población femenina (PFEM) y la ocupada (POCUP). Un estudio previo (Wang y Akar, 2019) encontró una relación inesperada entre una corta distancia a un campus universitario y más viajes de mujeres. También destacó la importancia positiva del número de intersecciones en la cantidad de viajes femeninos debido a su relación con menos tránsito y velocidades más bajas.

En cambio, la importancia por permutación consideró de mayor importancia en el destino las ciclovías y la zona 30. Según Reglamento de Movilidad, Transporte y Seguridad Vial de Guadalajara (2023) una Zona 30 se refiere a “el área de accesibilidad determinada con señalamientos y dispositivos viales para reducir la velocidad a un máximo de treinta kilómetros por hora” (Art. 51). Por lo que, al igual que las ciclovías, podría estar relacionada con una mayor preferencia por parte de las mujeres debido a lo mostrado en otros estudios (Wang y Akar, 2019; Dill et al., 2014; Garrard et al., 2007)

En el modelo del Mediodía, se observa que el entorno del origen tiene una importancia similar al del destino según el índice Gini. Esta métrica destaca como variables de importancia en el origen a la población ocupada (*POCU), universidades (*UNI), población femenina (*PFEM) y transporte público masivo (*TPM). En el destino, se mantienen importantes el transporte público masivo (TPM), las intersecciones (INT) y la población ocupada (POCU), pero resaltan las calles de prioridad ciclista (PRI) y la población total (PTOT). Para la métrica de PIM, se observa que hay más variables relevantes en el origen que en el destino, como la infraestructura ciclista (*PRI y *CIV), las intersecciones (*INT), la población ocupada (*POCU) y las áreas verdes (*VER). La literatura ha sugerido que las áreas verdes influyen de manera positiva en el uso de los SBC (Su et al., 2014; Chen et al, 2017; Zhao y Li, 2017), sin embargo, Wang y Akar

(2019) sostienen que, dado que las mujeres tienden a usar la bicicleta para uso recreativo, los parques y destinos de recreación las benefician más.

Si bien esta premisa incluye a los sitios recreativos, en este caso, la variable *REC y REC que incluye servicios de esparcimiento culturales, deportivos y otros, desde compañías de espectáculos artísticos y culturales, museos, sitios históricos, zoológicos y hasta parques con juegos electrónicos, no resultaron relevantes para ningún modelo. En futuras investigaciones, tal y como lo sugiere Wang y Akar (2019) se podría explorar la influencia de cada sitio diferente.

En el caso de la tarde, aunque la métrica derivada de Gini señala a las vialidades de prioridad ciclista en el origen (*PRI) como la variable principal, las demás variables más relevantes se encuentran en el destino. Entre estas variables destacan algunas que ya habían sido importantes en horarios anteriores. La métrica PIM coincide en señalar a la misma variable como la más importante, reconociendo su relevancia tanto en la estación de origen como en la de destino. Otras variables de importancia son los bancos en el origen (*BAN) y, en tres de los cuatro ejercicios realizados, los servicios educativos (EDU), la población total (PTOT) y el área verde (VER) en el destino resultaron ser de gran importancia.

Al respecto, la existencia de bancos no ha sido mencionada en la literatura previa como una variable de impacto en el uso de los SBC, sin embargo, podría ser una variable implicada en viajes para mandados. La variable de servicios educativos incluye desde escuelas de educación básica y media, técnico superior, de oficios y otro tipo de servicios como escuelas de arte, de deporte y de idiomas, posiblemente habría que desarrollar un ejercicio futuro con esta variable más desagregada.

En el modelo de pico vespertino, tanto GINI como PIM identifican a las calles de prioridad ciclista tanto en el origen como en el destino (*PRI y PRI) como variables de alta importancia. Específicamente, en el puntaje de importancia por permutación, cuatro de las primeras cinco variables de mayor importancia corresponden a la infraestructura ciclista. Contrario a los resultados de Wang y Akar (2019) donde la existencia de sitios para aparcar bicicletas fue la variable que contribuyó más a minimizar la brecha de género debido a que las mujeres se sienten

más cómodas cerca de otros ciclistas, el modelo del pico vespertino es el único horario donde los ciclopuertos⁴ (CIP) cobran relevancia.

Finalmente, en el modelo nocturno resaltan, para la métrica derivada de Gini variables en su mayoría demográficas, mientras que para la métrica derivada de las permutaciones la mayoría de las variables fueron referentes a infraestructura ciclista, de modo que por primera vez el número de estaciones del sistema MiBici dentro del radio del origen (*EST) fue de importancia.

La **Tabla 6** resume las variables de mayor importancia en ambas métricas, tanto para el origen como para el destino, de los diferentes modelos.

Tabla 6

Variables de mayor importancia en el origen y destino para los diferentes modelos

Global				Pico Matutino 6:00 am a 9:59 am				Mediodía 10:00 am a 1:59 pm			
Origen		Destino		Origen		Destino		Origen		Destino	
GINI	PIM	GINI	PIM	GINI	PIM	GINI	PIM	GINI	PIM	GINI	PIM
*PRI	*PRI	PRI	PRI	*TPM	*PRI	TPM	PRI	*POCU	*PRI	TPM	EDU
*POCU	*FAR	TPM	TPM	*PRI	*FAR	UNI	CIV	*UNI	*CIV	PTOT	PRI
*TPM	*POCU	POCU	EDU	*PTOT	*PTOT	INT	Z30	*PFEM	*INT	PRI	VER
*EDU	*TPM	PFEM	ALIM	*UNI	*CIV	PFEM	UNI	*TPM	*POCU	INT	POCU
*FAR	*EDU	PTOT	CIV	*POCU	*TPM	POCU	REC	*INT	*VER	POCU	Z30

Tarde 2:00 pm a 5:59 pm				Pico Vespertino 6:00 pm a 9:59 pm				Noche 10:00 pm a 11:59 pm			
Origen		Destino		Origen		Destino		Origen		Destino	
GINI	PIM	GINI	PIM	GINI	PIM	GINI	PIM	GINI	PIM	GINI	PIM
*PRI	*PRI	PTOT	PRI	*INT	*PRI	TPM	PRI	*POCU	*CIV	PRI	PRI
*POCU	*BAN	PRI	CIV	*POCU	*PTOT	PRI	TPM	*TPM	*EST	PFEM	POCU
*TPM	*CIV	TPM	EDU	*PRI	*POCU	POCU	CIP	*PFEM	*VIA	POCU	TPM
*INT	*PTOT	PFEM	PTOT	*UNI	*INT	INT	CIV	*UNI	*PRI	TPM	PFEM
*UNI	*TPM	UNI	VER	*TPM	*UNI	PFEM	UNI	*ALIM	*BAR	PTOT	ALIM

Nota: La descripción de las variables se observa en el Anexo 1. Se muestran en negritas las variables que coinciden en importancia en ambos índices. *Fuente:* Elaboración propia.

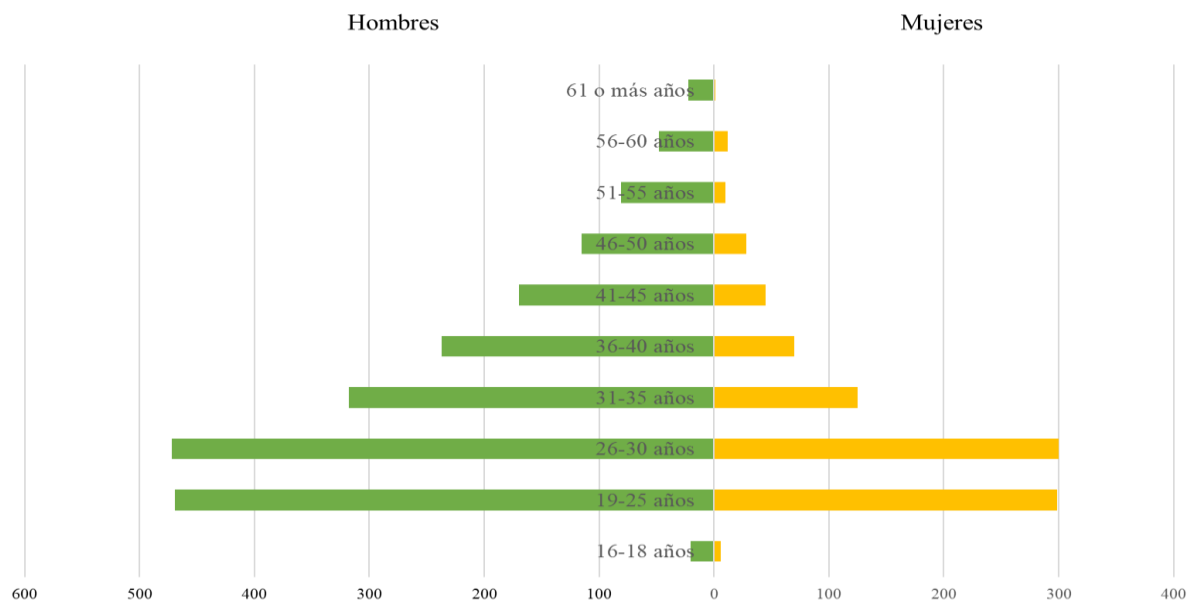
⁴ Espacio destinado al resguardo exclusivo de bicicletas. (Reglamento de Movilidad, Transporte y Seguridad Vial de Guadalajara, 2023, Artículo 6).

5.3 Validación comparativa con Encuesta Mibici

La encuesta de Percepción MiBici 2019 incluye respuestas de 3,371 personas. Sin embargo, debido a la delimitación que se usó en el modelo, en este caso únicamente se contemplan las respuestas de usuarios que indicaron usar el sistema al menos una vez cada 15 días, obteniendo así 2,911 personas, de los cuáles el 67% corresponden a hombres. La **Figura 20** presenta la distribución por género y edad, la cual se asemeja a la pirámide de edades de los usuarios que realizaron viajes en octubre, presentada en la sección dos, con la principal diferencia de que en el caso de la encuesta el grupo de 26 a 30 años es casi igual que el de 19 a 25 años.

Figura 20

Pirámide de edades de las personas encuestadas en la encuesta de percepción MiBici 2019

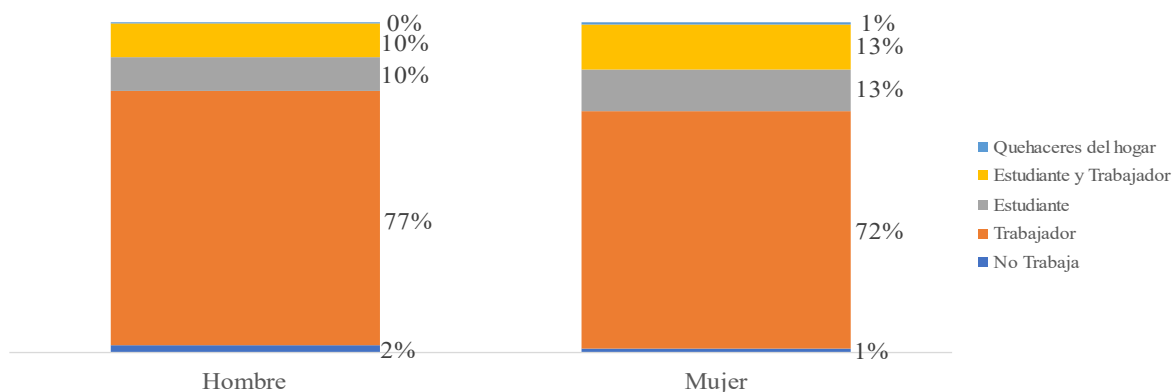


Fuente: Elaboración propia con datos de WRI (2019).

La encuesta incluye además una pregunta sobre su ocupación. La **Figura 21** expone los porcentajes según la ocupación y el género de los encuestados. Se observa que existe un poco más de diversidad en cuanto a la ocupación de las mujeres, destacando que uno por ciento de ellas se dedica a los quehaceres del hogar lo cual podría afectar sus necesidades de movilidad como se observó en la sección dos.

Figura 21

Porcentaje de ocupación de las personas encuestadas según género



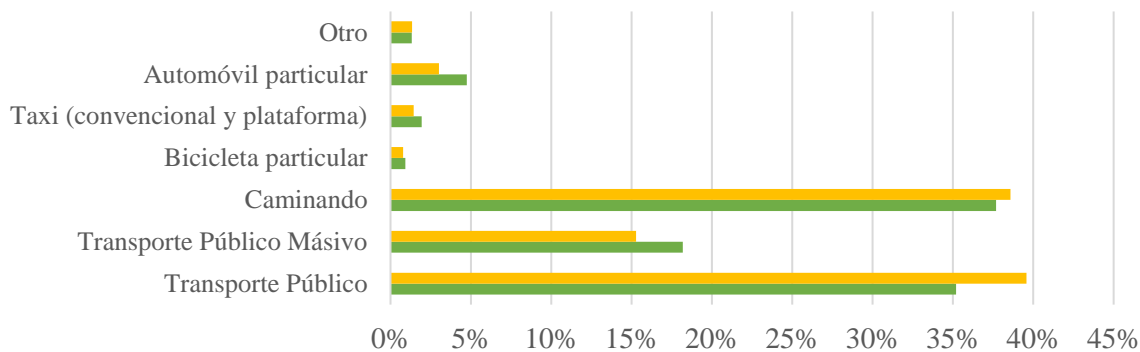
Nota: La categoría “No trabaja” agrupa a quienes contestaron Desocupado, Jubilado o en Blanco; “Trabaja” agrupa empleado, empleado de gobierno, negocio propio y otro. *Fuente:* Elaboración propia con datos de WRI (2019).

Además de estas características demográficas, se puede obtener información sobre el comportamiento de estos usuarios respecto al sistema. Por ejemplo, tanto hombres como mujeres tienen como principal destino el municipio de Guadalajara, seguido de Zapopan y finalmente Tlaquepaque. También resulta que más de la mitad de los encuestados caminan menos de 5 minutos, lo que se traduce en distancias cortas dentro de un radio caminable y se adapta a los 400 metros de radio definidos en esta investigación.

En relación con sus desplazamientos, la **Figura 22** revela que los usuarios de MiBici suelen finalizar sus viajes caminando o utilizando alguna forma de transporte público. Se destaca que las mujeres muestran un mayor índice de utilización del transporte público (autobuses, camiones y Pretren) para complementar sus trayectos. Sin embargo, sorprendentemente, la variable relacionada con las rutas de este transporte no demostró relevancia en ningún modelo de RF. Por el contrario, la distancia a una estación de transporte público masivo fue consistentemente una variable importante tanto en el origen como en el destino, lo cual podría atribuirse al hecho de que un 3% más de hombres que mujeres completan sus viajes utilizando esta modalidad. Además, el uso del automóvil como complemento del viaje podría estar relacionado con el hecho de que el 60% de los hombres encuestados afirmaron tener acceso a un automóvil en su hogar, en comparación con el 52% de las mujeres. Los resultados de un estudio previo en Terán (Mamdoohi y Amini, 2021) sugerían que el contar con al menos un vehículo personal en el hogar implicaba un menor uso de la bicicleta y que las mujeres eran más sensibles a esto.

Figura 22

Modos de transporte con los que suelen completar el viaje de MiBici las personas encuestadas según género



Nota: El color verde se refiere a hombres y el amarillo a mujeres. La categoría de Otros engloba motocicleta, transporte de personal y respuestas en blanco; Transporte Público incluye autobuses, camiones y PreTren; y Transporte Público Masivo incluye Macrobus, Tren Liero y Trolebus. Fuente: Elaboración propia con datos de WRI (2019).

Una diferencia adicional entre hombres y mujeres en cuanto al uso de MiBici se relaciona con su uso durante los fines de semana. Mientras que el 70% de las mujeres utilizan el sistema estos días, los hombres presentan una proporción menor, registrando un 67%, si bien en los modelos de RF la variable de día de la semana con la métrica GINI obtenía un puntaje alto, con la métrica de PIM carecía de total relevancia.

La **Tabla 7** revela los propósitos principales de viaje, donde se constata que, si bien para ambos sexos el principal motivo de viaje es ir al trabajo, el porcentaje de mujeres que lo hacen con este fin es ligeramente menor. Por otro lado, se observa que, en los propósitos de ir a la escuela, por entretenimiento o con fines sociales, las mujeres presentan un mayor porcentaje que los hombres. Esto podría relacionarse con el hecho de que las variables relacionadas con universidades y áreas verdes tengan importancia en varios modelos.

Además de estos propósitos principales, un 21% de los hombres y un 25% de las mujeres mencionaron que también utilizaban MiBici para realizar viajes sociales o de entretenimiento, así como un 15% tanto de hombres como de mujeres señalaron que lo usaban para la realización de trámites y mandados. Esta distribución de motivos de viaje podría estar relacionada con lo expuesto en la sección 2 y podría tener implicaciones en la elección de ciertas estaciones de origen y destino que satisfagan las necesidades de viaje específicas de cada grupo demográfico.

Tabla 7*Porcentaje de personas encuestadas por género según el propósito principal del viaje*

PROPÓSITO PRINCIPAL	HOMBRE	MUJER
Asistir a la vía recreativa	0.7%	0.9%
Hacer ejercicio con MiBici	1.6%	1.8%
Hacer trámites/ mandados	9.4%	8.2%
Ir a la casa	6.7%	7.8%
Ir a la escuela	7.9%	9.7%
Ir a reuniones laborales/ personales	5.2%	4.9%
Ir al trabajo	44.9%	42.3%
Ir de compras	0.6%	0.7%
Por diversión	1.2%	1.1%
Social/ entretenimiento	9.1%	10.1%
Vinculación con transporte público	9.7%	9.6%
Otro	2.7%	2.6%

Fuente: Elaboración propia con datos de WRI (2019).

En cuanto a la circulación, se observan diferencias entre hombres y mujeres en sus preferencias. Mientras que el 6% de los hombres suele circular por avenidas, solo el 3% de las mujeres lo hace. Del mismo modo, el 30% de los hombres opta por circular por calles locales, en comparación con el 24% de las mujeres. A pesar de estas diferencias, tanto hombres como mujeres muestran una clara preferencia por las ciclovías, aunque esta elección es más marcada entre las mujeres, representando un 62% en contraste con el 55% de los hombres. Este comportamiento podría estar relacionado con la aversión al riesgo observada en las mujeres, mencionada en la sección 2 y con las importancias registradas para vialidades de prioridad ciclista, ciclovías, zonas 30 y número de intersecciones.

Relacionado a ello, la infraestructura ciclista, que comprende las ciclovías y carriles de prioridad, fueron evaluados por los encuestados en una escala de 1 a 5, donde 5 representa una calificación excelente y 1 pésima. En materia de seguridad se obtuvo un promedio de 3.18, la cobertura de la red fue la peor calificada con un 2.71 en promedio, el entendimiento de los señalamientos con 3.19 y el estado físico de la infraestructura con 2.84. En general, los hombres

otorgaron calificaciones más altas a estos cuatro aspectos, lo cual podría influir en la utilización de dicha infraestructura.

Además, es importante destacar que la falta de infraestructura ciclista fue identificada como uno de los principales obstáculos en el sistema por el 21% de los encuestados, seguido por el estado de las vialidades, mencionado por el 19% de los participantes.

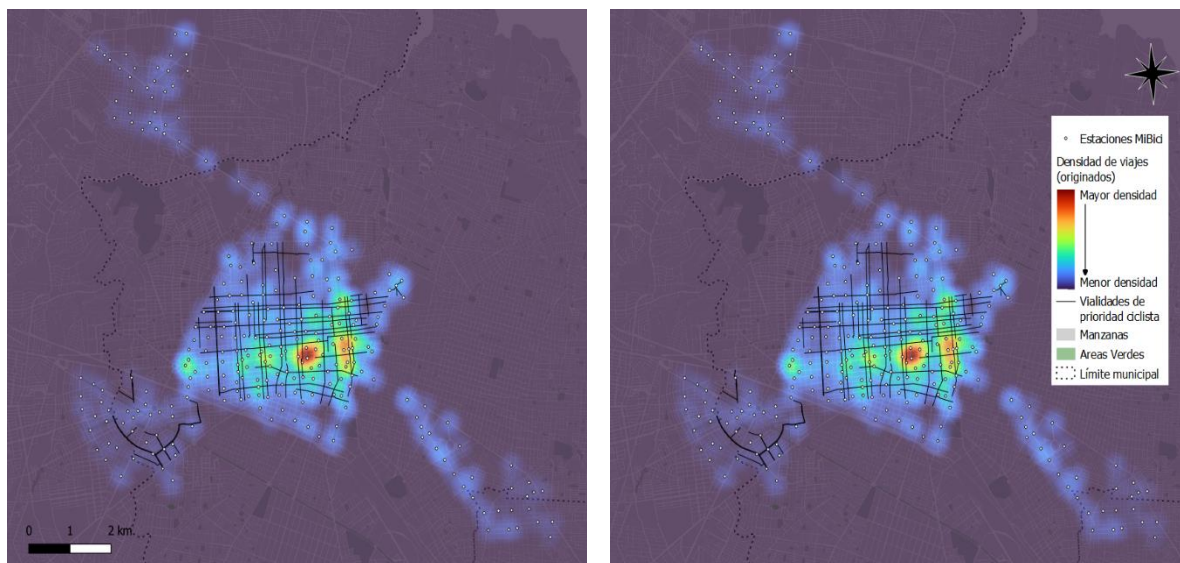
5.4 Validación visual y cualitativa por medio de mapas

Para clarificar la asociación entre las variables del entorno urbano y la cantidad de viajes por cada grupo, se realiza un cruce visual de estas variables y así poder comprender en qué sentido y magnitud impacta cada variable en el hecho de que un viaje sea realizado por un hombre o una mujer.

Las calles de prioridad ciclista emergieron como una variable importante tanto en el modelo mensual como en algunos modelos específicos. La **Figura 23** presenta, en el lado derecho, la ubicación de este tipo de infraestructura ciclista, así como la densidad de viajes originados por hombres. Por otro lado, en el lado izquierdo se muestra la densidad de viajes originados por mujeres. En esta representación, el color rojo indica una mayor densidad de viajes en esa área. Sin embargo, se observa una distribución muy similar tanto para hombres como para mujeres, lo que sugiere que no hay una relación clara entre la densidad de viajes y la existencia de esta infraestructura.

Figura 23

Asociación entre la distribución de las calles de prioridad ciclista y la densidad de viajes originados según género

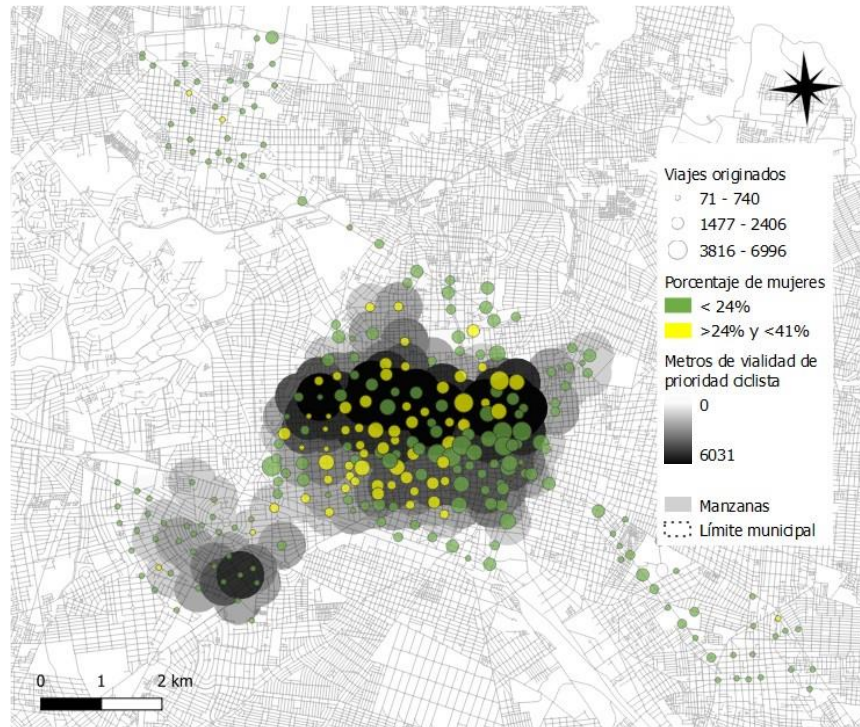


Fuente: Elaboración propia.

Dado que aparentemente tanto hombres como mujeres utilizan las mismas estaciones en casi la misma medida, la **Figura 24** muestra la misma información desde otra perspectiva. En esta figura, los buffers se categorizan en una escala de grises según la cantidad de metros de prioridad ciclista dentro de cada uno. Un color negro indica una mayor cantidad de metros de prioridad ciclista. Además, se muestran dos atributos en cada estación: el color de cada estación refleja el porcentaje de viajes por género, donde las estaciones con un porcentaje de viajes hecho por mujeres superior al 24% (el promedio) se muestran en color amarillo, mientras que las que están por debajo se muestran en verde. Por otro lado, el tamaño de las estaciones está determinado por la cantidad total de viajes originados en cada una. Esta combinación de atributos permite ponderar entre la cantidad de viajes y la predominancia de un género en ellos.

Figura 24

Asociación entre la concentración de las calles de prioridad ciclista por buffer y el porcentaje de viajes originados por género



Fuente: Elaboración propia.

Sin embargo, esta visualización tampoco aclara una relación entre la cantidad de viajes por género y la existencia de calles de prioridad ciclista. Esta situación persiste en los diferentes horarios y para diferentes variables, lo que dificulta profundizar en el impacto de las variables resultantes del modelo RF. Basado en la falta de claridad de las Figuras 22 y 23, se puede intuir que, posiblemente el proceso de balanceo de la muestra, donde se tomaron muchas menos observaciones de hombres de las que existían, pudo sesgar de alguna manera las importancias de las variables.

A modo de resumen de esta sección se presenta la **Tabla 8** que compara los resultados de los modelos RF, con los resultados de la encuesta y lo dicho en la literatura para las variables donde se presentaron más coincidencias.

Tabla 8

Resumen de los resultados

Categoría	VARIABLES	Literatura	GINI	PIM	Encuesta	Mapas
Temporales	Día (DIA)	Los hombres viajan durante las horas pico mientras que las mujeres tienen días y horarios más variables	Importante mensual y en horarios	Sin importancia mensual ni horaria	Un mayor porcentaje de mujeres utiliza el sistema en fin de semana. (H 67% - M 70%)	
	Horario (HOR)		Importante mensual y en horarios	Importante mensual y en horarios	Sin información	
Infraestructura ciclista	Ciclopuertos (CIP)	Las mujeres se sienten más cómodas cerca de otros ciclistas	Sin importancia mensual ni horaria	Sin importancia mensual, importante en pico vespertino	Sin información	
	Ciclo vías (CIV)	Las mujeres tienden a preferir circular en infraestructura ciclista más que sus contrapartes masculinas.	Medianamente importante en global y muy importante en los horarios	Medianamente importante en global y muy importante en los horarios	Las mujeres prefieren circular por ciclo vías (H 55% - M 62%)	No concluyente
Transporte y movilidad	Vialidades de prioridad ciclista (PRI)		Importante mensual y en horarios	Importante mensual y en horarios	Sin información	
	Número de intersecciones viales (INT)	Mayor número de intersecciones implica menor densidad vehicular y mayor atracción de viajes femeninos.	Medianamente importante en mensual e importante en horarios	Sin importancia mensual ni horaria	Más hombres que mujeres prefieren circular por avenidas (H 5% - M 3%) o calles locales (H3 0% - M 24%)	
	Rutas de transporte público (RTP)	Afectación negativa en el uso de las mujeres, pero positiva en los hombres por ser más propensos a realizar viajes de trabajo	Sin importancia mensual ni horaria	Sin importancia mensual ni horaria	Las mujeres muestran un mayor índice de utilización del transporte público para complementar sus trayectos. (H 35% - M 40%)	
	Distancia a estación de transporte público masivo (TPM)		Importante mensual y en horarios	Importante mensual y en horarios	Los hombres más que las mujeres, completan sus viajes utilizando transporte público masivo (H 18% - M 15%)	

Categoría	VARIABLES	Literatura	GINI	PIM	Encuesta	Mapas
Demográficas	Población ocupada (OCU)	Asociación negativa entre los viajes femeninos y la densidad de empleo	Importante mensual y en horarios	Medianamente importante mensual sin importancia horaria	Mayor porcentaje de hombres que su propósito principal es ir al trabajo (H 45% - M 42%)	
Actividades y uso de suelo	Distancia a una universidad (UNI)	A menor distancia a un campus universitario, mayor cantidad de viajes de mujeres	Medianamente importante en mensual e importancia horaria	Sin importancia mensual ni horaria	Mayor porcentaje de mujeres que su propósito principal es ir a la escuela (H 8% - M 10%)	No concluyente
	Metros cuadrados de área verde (VER)	Influyen de manera positiva en el uso de los SBC, pero aún más en el uso por parte de las mujeres debido al uso recreativo que le dan	Medianamente importante en mensual y horarios	Sin importancia mensual y medianamente importante al mediodía	Mayor porcentaje de mujeres que su propósito principal es ir sociales o de entretenimiento (H 9% - M 10%)	

Fuente: Elaboración propia.

6. Conclusiones y recomendaciones

Este estudio se centró en identificar y validar las variables del entorno urbano que impactan en el uso de bicicletas compartidas de manera diferenciada por género. Se propone una metodología innovadora que consiste en entrenar un algoritmo de *Random Forest* con la información de más de 348 mil viajes realizados en el sistema de bicicletas compartidas de la ciudad de Guadalajara, y 23 variables del entorno urbano.

Se evaluó el rendimiento de diferentes modelos y con base en las métricas de importancia derivada del índice de Gini y el índice de importancia de permutación se identificaron las variables que tuvieron una mayor importancia para clasificar el género de los usuarios que realizaron viajes. Estas métricas nos permiten conocer las variables que más influyen tanto en la separación de las clases como en el rendimiento del modelo, sin embargo, su interpretación requiere un análisis adicional para comprender completamente cómo las variables afectan la clasificación, si de manera negativa o positiva y en que magnitud.

Por ello, posteriormente se realizó una validación con los resultados de una encuesta de percepción realizada en el mismo año por WRI y finalmente se intentó una segunda validación visual por medio de un cruce espacial de variables. Esta última validación no tuvo los resultados esperados ya que no se logró ver una relación clara entre las variables, muy posiblemente por un sesgo introducido con la operación de submuestreo realizada para balancear la muestra que, por la naturaleza del problema, tenía una sobre representación de viajes masculinos.

Los resultados obtenidos revelaron que la infraestructura ciclista, incluyendo las vialidades de prioridad ciclista, las ciclovías y los ciclopuertos, juegan un papel crucial en la determinación del género del usuario que realiza un viaje. Asimismo, se encontró que la distancia al transporte público masivo también influye significativamente, así como el hecho que dependiendo el horario del día las variables de importancia se van modificando. Estos hallazgos son consistentes con la literatura previa, respaldando la robustez de nuestras conclusiones.

En resumen, este estudio proporciona una visión detallada y actualizada de los factores que influyen en la demanda diferenciada por género en un sistema de bicicletas compartidas en el contexto urbano mexicano. Estos hallazgos abonan a la literatura sobre la brecha de género en los sistemas de bicicleta pública y pueden ser utilizados por autoridades locales y planificadores

urbanos para mejorar la accesibilidad y la equidad en el transporte público, promoviendo así una movilidad más sostenible y equitativa para todos los ciudadanos.

En trabajos futuros se podrían utilizar datos GPS para conocer el entorno no solo de las estaciones de origen y destino sino de todo su trayecto. Asimismo, se podrían refinar las clasificaciones considerando el ciclo de vida y no solo el género, suponiendo que ello impacta directamente en la movilidad del usuario.

Bibliografía

- Albuquerque, V., Sales Dias, M., & Bacao, F. (2021). Machine learning approaches to bike-sharing systems: A systematic literature review. *ISPRS International Journal of Geo-Information*, 10(2), 62. <https://doi.org/10.3390/ijgi10020062>
- Archer, K. J., & Kimes, R. V. (2008). Empirical characterization of random forest variable importance measures. *Computational Statistics & Data Analysis*, 52(4), 2249–2260. <https://doi.org/10.1016/j.csda.2007.08.015>
- Ashqar, H. I., Elhenawy, M., & Rakha, H. A. (2020). Modeling bike counts in a bike-sharing system considering the effect of weather conditions. In *arXiv [cs.CY]*. <http://arxiv.org/abs/2006.07563>
- Bacciu, D., Carta, A., Gnesi, S., & Semini, L. (2017). An experience in using machine learning for short-term predictions in smart transportation systems. *Journal of Logical and Algebraic Methods in Programming*, 87, 52–66. <https://doi.org/10.1016/j.jlamp.2016.11.002>
- Breiman, L. (2001). Random Forests. *Machine Learning*, 45(1):5-32
- Buck, D., & Buehler, R. (2011). *Bike lanes and other determinants of capital bikeshare trips*. 91st Transportation Research Board Annual Meeting Nacto.org. <https://nacto.org/wp-content/uploads/2012/02/Bike-Lanes-and-Other-Determinants-of-Capital-Bikeshare-Trips-Buck-et-al-12-3539.pdf>
- Castellanos, S., De la Lanza, I., Bray Sharpin, A., Lleras, N., Lo Re, L., & Amezola R., D. C. (2019). *Guía para la estructuración de sistemas de bicicletas compartidas*. Inter-American Development Bank. <http://dx.doi.org/10.18235/0002391>
- Catani, F., Lagomarsino, D., Segoni, S., and Tofani, V. (2013). Landslide susceptibility estimation by random forests technique: sensitivity and scaling issues, *Nat. Hazards Earth Syst. Sci.*, 13, 2815–2831, <https://doi.org/10.5194/nhess-13-2815-2013>.
- Díaz, R., y Rojas, F. (2017). *Mujeres y Ciclismo Urbano: Promoviendo políticas Inclusivas de movilidad en América Latina*. Banco Interamericano de Desarrollo (BID).

<https://publications.iadb.org/es/mujeres-y-ciclismo-urbano-promoviendo-politicas-inclusivas-de-movilidad-en-america-latina>

- Dill, J., Goddard, T., Monsere, C., McNeil, N. (2014). Can Protected Bike Lanes Help Close the Gender Gap in Cycling? Lessons from Five Cities. *Urban Studies and Planning Faculty Publications and Presentations*. 123. https://pdxscholar.library.pdx.edu/usp_fac/123
- Du, Y., Deng, F., & Liao, F. (2019). A model framework for discovering the spatio-temporal usage patterns of public free-floating bike-sharing system. *Transportation Research. Part C, Emerging Technologies*, 103, 39–55. <https://doi.org/10.1016/j.trc.2019.04.006>
- El-Assi, W., Salah Mahmoud, M., & Nurul Habib, K. (2017). Effects of built environment and weather on bike sharing demand: a station level analysis of commercial bike sharing in Toronto. *Transportation*, 44(3), 589–613. <https://doi.org/10.1007/s11116-015-9669-z>
- Emond, CR., Tang, W. y Handy, SL. (2009). Explaining Gender Difference in Bicycling Behavior. *Transportation Research Record*, 2125(1), 16-25. <https://doi.org/10.3141/2125-03>
- Eren, E., & Uz, V. E. (2020). A review on bike-sharing: The factors affecting bike-sharing demand. *Sustainable Cities and Society*, 54(101882), 101882. <https://doi.org/10.1016/j.scs.2019.101882>
- Faghih-Imani, A., Eluru, N., El-Geneidy, A. M., Rabbat, M., & Haq, U. (2014). How land-use and urban form impact bicycle flows: evidence from the bicycle-sharing system (BIXI) in Montreal. *Journal of Transport Geography*, 41, 306–314. <https://doi.org/10.1016/j.jtrangeo.2014.01.013>
- Faghih-Imani, Ahmadreza; Eluru, Naveen; and Paleti, Rajesh. (2017). "How Bicycling Sharing System Usage is Affected by Land Use and Urban Form: Analysis from System and User Perspectives". *Civil & Environmental Engineering Faculty Publications*. 15. https://digitalcommons.odu.edu/cee_fac_pubs/15
- Feng, Y., & Wang, S. (2017). A forecast for bicycle rental demand based on random forests and multiple linear regression. 2017 IEEE/ACIS 16th International Conference on Computer and Information Science (ICIS), 101-105. DOI: 10.1109/ICIS.2017.7959977

- Garrard, J., Rose, G., & Lo, S. K. (2008). Promoting transportation cycling for women: The role of bicycle infrastructure. *Preventive Medicine*, 46(1), 55–59. <https://doi.org/10.1016/j.ypm.2007.07.010>
- Garrard, J., Rose, G., & Lo, S. K. (2008). Promoting transportation cycling for women: The role of bicycle infrastructure. *Preventive Medicine*, 46(1), 55–59. <https://doi.org/10.1016/j.ypm.2007.07.010>
- Genuer, R., Poggi, J.-M., & Tuleau-Malot, C. (2010). Variable selection using random forests. *Pattern Recognition Letters*, 31(14), 2225–2236. <https://doi.org/10.1016/j.patrec.2010.03.014>
- Gobierno de Guadalajara. (s.f.). Mapa Medio Ambiente. <https://mapa.guadalajara.gob.mx/medio-ambiente>
- Grekousis, G., Feng, Z., Marakakis, I., Lu, Y., & Wang, R. (2022). Ranking the importance of demographic, socioeconomic, and underlying health factors on US COVID-19 deaths: A geographical random forest approach. *Health & Place*, 74(102744), 102744. <https://doi.org/10.1016/j.healthplace.2022.102744>
- Grömping, U. (2009). Variable Importance Assessment in Regression: Linear Regression versus Random Forest. *The American Statistician*, 63(4), 308–319. <http://www.jstor.org/stable/25652309>
- Guidon, S., Reck, D. J., & Axhausen, K. (2020). Expanding a(n) (electric) bicycle-sharing system to a new city: Prediction of demand with spatial regression and random forests. *Journal of Transport Geography*, 84(102692), 102692. <https://doi.org/10.1016/j.jtrangeo.2020.102692>
- Hasan, M., Nasser, M., Ahmad, S. and Molla, K. (2016) Feature Selection for Intrusion Detection Using Random Forest. *Journal of Information Security*, 7, 129-140. doi: 10.4236/jis.2016.73009
- Higuera-Mendieta, D., Uriza, P. A., Cabrales, S. A., Medaglia, A. L., Guzman, L. A., & Sarmiento, O. L. (2021). Is the built-environment at origin, on route, and at destination

- associated with bicycle commuting? A gender-informed approach. *Journal of Transport Geography*, 94(103120), 103120. <https://doi.org/10.1016/j.jtrangeo.2021.103120>
- Humboldt-Universität zu Berlin. (2023). 8. Image classification - Random Forests. *Introduction to Remote Sensing*. Department of Geography. Humboldt-Universität zu Berlin. https://pages.cms.hu-berlin.de/EOL/geo_rs/index.html
- INEGI. (2020). *Censo de Población y Vivienda 2020*. <https://www.inegi.org.mx/programas/ccpv/2020/>
- INEGI. (s.f.). *Directorio Estadístico Nacional de Unidades Económicas 11/2019*. <https://www.inegi.org.mx/app/mapa/denue/>
- Instituto de Movilidad y Transporte del Estado de Jalisco. (2015). Plan Integral de Movilidad y Transporte Sustentable del Área Metropolitana de Guadalajara. https://plan.jalisco.gob.mx/wp-content/uploads/2022/12/Plan_Integral_de_Movilidad_Urbana_Sustentable_del_AMG_2015_2.pdf
- Instituto Metropolitano de Planeación. (2019). *SIG metro*. <https://sigmetro.imeplan.mx/>
- Jaimurzina, A., Muñoz, C. y Perez, G. (2017). *Género y transporte: experiencias y visiones de política pública en América Latina*. Comisión Económica para América Latina y el Caribe (CEPAL). https://repositorio.cepal.org/bitstream/handle/11362/43125/1/S1700969_es.pdf
- Jain, T., Wang, X., Rose, G., & Johnson, M. (2018). Does the role of a bicycle share system in a city change over time? A longitudinal analysis of casual users and long-term subscribers. *Journal of Transport Geography*, 71, 45–57. <https://doi.org/10.1016/j.jtrangeo.2018.06.023>
- Jalisco. (2021). *Convocatoria Mi Pasaje Apoyo a Mujeres*. Gobierno de Jalisco. <https://www.jalisco.gob.mx/es/gobierno/comunicados/convocatoria-mi-pasaje-apoyo-mujeres>
- Jung, A. (2022). *Machine Learning: Foundations, Methodologies, and Applications, The Basic*. Springer.

- LatinoSBP.org. (2019). Bike Sharing Systems in Latin America Annual Report. <https://cff-prod.s3.amazonaws.com/storage/files/uevrH77rhjqqsj4oFxYGswAfarXaVAL8rMMxJfc7.pdf>
- Mamdoohi, A. R., & Amini, V. (2021). Bicycle Demand: A Gender Analysis for Tehran, Iran. *International Journal of Transportation Engineering*, 8(3), 211-223. doi: 10.22119/ijte.2020.119098.1391
- MiBici. (s.f.-a). *Acerca de MiBici*. MiBici. <https://www.mibici.net/es/acerca-de-mibici/>
- MiBici. (s.f.-b). *Datos Abiertos*. MiBici. <https://www.mibici.net/es/datos-abiertos/>
- MiBici. (s.f.-c). *Preguntas Frecuentes*. MiBici. <https://www.mibici.net/es/preguntas-frecuentes/>
- Naciones Unidas. (2022). Día Mundial de la Bicicleta 3 de junio. <https://www.un.org/es/observances/bicycle-day>
- Ngo, T.-T. T., Pham, H. T., Acosta, J. G., & Derrible, S. (2022). Predicting bike-sharing demand using Random Forest. *Journal of Science and Transport Technology*, 13–21. <https://doi.org/10.58845/jstt.utt.2022.en65>
- Nickkar, A., Banerjee, S., Chavis, C., Bhuyan, I. A., & Barnes, P. (2019). A spatial-temporal gender and land use analysis of bikeshare ridership: The case study of Baltimore City. *City Culture and Society*, 18(100291), 100291. <https://doi.org/10.1016/j.ccs.2019.100291>
- ONU-Habitat. (2022, 12 de agosto). El Radio Caminable. ONU-Habitat. <https://onuhabitat.org.mx/index.php/el-radio-caminable>
- PAHO. (2020). Bicycle-sharing Systems across the United States of America. PanAmerican Health Organization (PAHO). <https://doi.org/10.37774/9789275122143>
- PBSC Urban Solutions. (2021). The Meddin Bike-Sharing World Map. PBSC Urban Solutions. <https://www.pbsc.com/blog/2021/10/the-meddin-bike-sharing-world-map>
- Pellicer-Chenoll, M., Pans, M., Seifert, R., López-Cañada, E., García-Massó, X., Devís-Devís, J., & González, L.-M. (2021). Gender differences in bicycle sharing system usage in the

- city of Valencia. *Sustainable Cities and Society*, 65(102556), 102556.
<https://doi.org/10.1016/j.scs.2020.102556>
- Pucher, J., & Buehler, R. (2008). Making cycling irresistible: Lessons from the Netherlands, Denmark and Germany. *Transport Reviews*, 28(4), 495–528.
<https://doi.org/10.1080/01441640701806612>
- Ravensbergen, L., Buliung, R., & Laliberté, N. (2019). Toward feminist geographies of cycling. *Geography Compass*, 13(7). <https://doi.org/10.1111/gec3.12461>
- Reglamento de la Vía RecreActiva del Municipio de Guadalajara. Art. 4. 2005.
<https://www.viarecreactiva.com/wp-content/uploads/2017/09/reg.viarecreactivaguadalajara.pdf>
- Reglamento de Movilidad, Transporte y Seguridad Vial de Guadalajara. Art. 6. 2023.
<https://transparencia.guadalajara.gob.mx/sites/default/files/reglamentos/RegMovilidadTransporteSeguridadVialGuadalajara.pdf>
- Russel, S. y Norvig, P. (2004). Inteligencia Artificial Un enfoque Moderno. 2ª Edición Pearson Prentice Hall.
- S. Ruffieux, N. Spycher, E. Mugellini and O. A. Khaled. (2017). "Real-time usage forecasting for bike-sharing systems: A study on random forest and convolutional neural network applicability," *Intelligent Systems Conference (IntelliSys)*, London, UK, pp. 622-631, doi: 10.1109/IntelliSys.2017.8324359
- Sathishkumar, V.E. & Cho, Y. (2020). Season wise bike sharing demand analysis using random forest algorithm. *Computational Intelligence*. <https://doi.org/10.1111/coin.12287>
- Scikit-learn Developers. (s.f.). *Feature importances with a forest of trees*. Scikit-learn Developers. https://scikit-learn.org/stable/auto_examples/ensemble/plot_forest_importances.html
- Secretaría de Desarrollo Agrario, Territorial y Urbano (SEDATU), Consejo Nacional de Población (CONAPO) e Instituto Nacional de Estadística y Geografía (INEGI). (2023). *Metrópolis de México 2020*.
https://www.gob.mx/cms/uploads/sedatu/MM2020_19102023.pdf

- Shaw, C., Russell, M., Keall, M., MacBride-Stewart, S., Wild, K., Reeves, D., Bentley, R., & Woodward, A. (2020). Beyond the bicycle: Seeing the context of the gender gap in cycling. *Journal of Transport & Health*, 18(100871), 100871. <https://doi.org/10.1016/j.jth.2020.100871>
- Strobl, C., Boulesteix, A.-L., Zeileis, A., & Hothorn, T. (2007). Bias in random forest variable importance measures: Illustrations, sources and a solution. *BMC Bioinformatics*, 8(1). <https://doi.org/10.1186/1471-2105-8-25>
- Sun, F., Chen, P., & Jiao, J. (2018). Promoting public bike-sharing: A lesson from the unsuccessful Pronto system. *Transportation Research. Part D, Transport and Environment*, 63, 533–547. <https://doi.org/10.1016/j.trd.2018.06.021>
- Transport Canada. (2009). Bike Sharing Guide. <http://mobility-workspace.eu/wp-content/uploads/bsg.pdf>
- Vogel, M., Hamon, R., Lozenguez, G., Merchez, L., Abry, P., Barnier, J., Borgnat, P., Flandrin, P., Mallon, I., and Robardet, C. (2014). From bicycle sharing system movements to users: a typology of v'elov cyclists in lyon based on large-scale behavioural dataset. *Journal of Transport Geography*, 41, 280–291. <https://doi.org/10.1016/j.jtrangeo.2014.07.005>
- Wang, B., & Kim, I. (2018). Short-term prediction for bike-sharing service using machine learning. *Transportation Research Procedia*, 34, 171–178. <https://doi.org/10.1016/j.trpro.2018.11.029>
- Wang, K., & Akar, G. (2019). Gender gap generators for bike share ridership: Evidence from Citi Bike system in New York City. *Journal of Transport Geography*, 76, 1–9. <https://doi.org/10.1016/j.jtrangeo.2019.02.003>
- Wang, X., Lindsey, G., Schoner, J. E., & Harrison, A. (2015). Modeling bike share station activity: Effects of nearby businesses and jobs on trips to and from stations. *Journal of Urban Planning and Development*. DOI:10.1061/(ASCE)UP.1943-5444.0000273
- Xu, H., Duan, F., & Pu, P. (2019). Dynamic bicycle scheduling problem based on short-term demand prediction. *Applied Intelligence*, 49(5), 1968–1981. <https://doi.org/10.1007/s10489-018-1360-6>

- Yang, Z., Hu, J., Shu, Y., Cheng, P., Chen, J., & Moscibroda, T. (2016). Mobility modeling and prediction in bike-sharing systems. *Proceedings of the 14th Annual International Conference on Mobile Systems, Applications, and Services*. pp. 165–178. <https://doi.org/10.1145/2906388.2906408>
- Zhou, X., Wang, M., & Li, D. (2019). Bike-sharing or taxi? Modeling the choices of travel mode in Chicago using machine learning. *Journal of Transport Geography*, 79(102479), 102479. <https://doi.org/10.1016/j.jtrangeo.2019.102479>

Anexo 1

Descripción y estadísticos de las variables del entorno empleadas

Categoría	Clave	Descripción	MIN	MAX	MEDIAN	MEAN	SD
	ALIM	núm. de UE de Servicios de preparación de alimentos y bebidas alcohólicas y no alcohólicas (7225) en el radio	2	305	54	65.42	50.74
	BAN	núm. de UE de Banca Múltiple (5221) en el radio	0	36	7	8.15	6.99
	BAR	núm. de UE de Centros nocturnos, bares, cantinas y similares (7224) en el radio	0	42	3	6.38	8.53
	COM	núm. de UE de Comercio al por menor de abarrotes, alimentos, bebidas, hielo y tabaco (461) y en minisúpers (462112) en el radio	0	406	31	47.34	56.41
	EDU	núm. de UE de Servicios educativos (menos nivel superior) (611 menos 6113) en el radio	0	38	10	11.92	7.31
Actividades y Uso de Suelo	FAR	núm. de UE de Comercio al por menor de productos farmacéuticos y naturistas (46411) es el radio	0	117	7	11.64	17.67
	GOB	núm. de UE de Actividades legislativas, gubernamentales, de impartición de justicia y de organismos internacionales y extraterritoriales (93) en el radio	0	58	4	7.54	9.99
	REC	núm. de UE de Servicios de esparcimiento culturales y deportivos, y otros servicios recreativos (71) en el radio	0	30	4	5.82	5.15
	SUP	núm. de UE de Comercio al por menor en supermercados (46211) y en tiendas departamentales (462210) en el radio	0	32	3	4.81	5.2
	UNI	distancia lineal mínima a una UE de Escuelas de educación superior (6113) en el radio	0	25	2	3.24	4.28
	VER	m2 de área verde en el radio	0	146,025	10,472	16,985	20,080.1

Categoría	Clave	Descripción	MIN	MAX	MEDIAN	MEAN	SD
	PFEM	% de población femenina en el radio	43%	56%	52%	52%	2%
Demográficos	POCU	población ocupada media (población flotante) en el radio	305	19,341	4,470	5,174	3,341.72
	PTOT	población total en el radio	378	7,745	2,368	3,804	1,624.59
	EST	núm. de estaciones del sistema dentro del radio	0	10	3	3.489	2.169
	CIP	núm. de ciclopuestos en el radio	0	37	0	3.52	7.35
Infraestructura Ciclista	CIV	m. de ciclovia en el radio	0	3,369.70	652.1	803	808.8
	PRI	m. de calles de prioridad ciclista en el radio	0	6,031	1,084	1,657	1,772.63
	VIA	m. de ViaReactiva en el radio	0	1,443.50	0	238.8	366.51
	INT	núm.. de intersecciones viales en el radio	100	1,412	586	612	198.76
Transporte Público y Movilidad	RTP	m. de corredores troncales en el radio	0	6,420	0	924.8	1,679.01
	TPM	distancia lineal mínima a una estación de transporte público masivo (MACROBUS y SITEUR	24.56	5,142.07	1,685.19	1,835.9	1,405.81
	Z30	m2. de Zonas 30 en el radio	0	493,608	0	57,853	102,996.26

Fuente: Elaboración propia.