



UNCUYO
UNIVERSIDAD
NACIONAL DE CUYO



FACULTAD DE
INGENIERÍA

TESINA FINAL DE GRADO

Licenciatura en Ciencias de la Computación

Análisis y visualización de datos de transporte para el mejoramiento de la movilidad urbana de Mendoza

Autor: | Gastón Ezequiel Cavallo
Tutor: | Dra. Ana Carolina Olivera

DICIEMBRE 2025

Facultad de Ingeniería
UNIVERSIDAD NACIONAL DE CUYO

Agradecimientos

Deseo expresar mi más sincero agradecimiento a todas aquellas personas e instituciones que, de una u otra forma, hicieron posible la culminación de esta tesina final de grado.

En primer lugar, quiero agradecer profundamente a mi Directora de Tesina: Dra. Ana Carolina Olivera, por su invaluable guía, paciencia, dedicación y los conocimientos compartidos a lo largo de todo este proceso. Su apoyo fue fundamental para el desarrollo de este trabajo.

Asimismo, extendo mi especial agradecimiento al Dr. Pablo Javier Vidal, por sus valiosos aportes y orientación en el área de programación y soporte realizados en esta tesina.

Quisiera destacar y agradecer especialmente a Juan José Palavecino, Director General de Gestión Documental, Transparencia y Acceso a la Información en el Ministerio de Economía de la Nación, por su generosidad al facilitarnos los datos del Sistema Único de Boleto Electrónico (SUBE), información crucial para el desarrollo de esta investigación.

Mi gratitud también se dirige a Javier Passera de la Sociedad de Transporte de Mendoza, por su disposición y colaboración al proporcionarnos los datos del sistema de bicicletas públicas biciTRAN. Su aporte fue de gran valor para el análisis realizado.

De igual manera, agradezco profundamente al Esp. Ing. Gustavo Luis Pastor, Ingeniero Senior de la Secretaría de Transporte de la Provincia de Mendoza y Docente de la Facultad de Ingeniería en la Cátedra Tránsito y Vialidad Urbana, por su tiempo, sus conocimientos y la información brindada, que enriquecieron significativamente este trabajo final de carrera.

No puedo dejar de mencionar el apoyo incondicional de mi familia y amigos, quienes con su aliento y comprensión hicieron más llevadero este camino, principalmente destacando a mi pareja, cuyo apoyo continuo me motivó a seguir adelante a lo largo de las dificultades encontradas en mis años de carrera. Su presencia constante fue una fuente de motivación invaluable.

Finalmente, agradezco a todas aquellas personas que, aunque no sean mencionadas directamente, contribuyeron de alguna manera a la realización de este trabajo.

Gastón Ezequiel Cavallo
Mendoza, diciembre de 2025

Resumen

La eficiencia de la movilidad urbana juega un papel fundamental en el desarrollo sostenible de las ciudades modernas. En particular, los sistemas de transporte público, tanto mono como multiusuarios, han tomado relevancia como catalizadores del mejoramiento del tráfico vehicular, tratando de disminuir la utilización del vehículo privado. En Mendoza, el continuo crecimiento de la población, la actividad económica y la distribución de los puntos de interés a lo largo de su geografía han planteado retos significativos en la gestión del sistema de transporte público urbano. Un aspecto importante es la falta de datos de cómo se trasladan las personas, en particular, matrices origen-destino actualizadas como consecuencia de factores externos, como el cambio de los sistemas de base de datos o la imposibilidad de acceder a datos sensibles para estudios de movilidad.

Esta tesina final de grado se centra en la construcción de matrices origen-destino a partir de datos reales de los distintos sistemas de transporte público y su utilización en la construcción de visualizaciones asociadas a la movilidad de las personas. Las matrices y sus visualizaciones constituyen una fuente valiosa de información para los tomadores de decisiones, permitiéndoles observar características espaciales y temporales que de otra forma no podrían capturarse a simple vista.

Por último y como parte de este trabajo final de carrera, se implementa un prototipo de página web interactiva que permite observar las visualizaciones obtenidas a partir de las matrices origen-destino construidas. Los resultados muestran que la integración de técnicas avanzadas de análisis de datos y visualización puede conducir a mejoras significativas en el análisis de la información de la movilidad.

Abstract

The efficiency of urban mobility plays a fundamental role in the sustainable development of modern cities. In particular, public transportation systems, whether single-user or multi-user, have gained relevance as catalysts for improving traffic flow by seeking to reduce private vehicle use. In Mendoza, the continuous growth of the population, economic activity, and the distribution of points of interest throughout its territory have posed significant challenges for the management of the urban public transportation system. One important aspect is the lack of updated data on how people move around the city, especially updated *origin-destination matrices*, due to external factors such as changes in database systems or the inaccessibility of sensitive data for mobility studies.

This final degree project focuses on the construction of *origin-destination matrices* using real data from different public transportation systems and their use in building visualizations related to people's mobility. These matrices and their associated visualizations represent a valuable source of information for decision-makers, allowing them to observe spatial and temporal characteristics that would otherwise be difficult to capture at a glance.

Finally, as part of this final degree project, a prototype of an interactive web page was developed to display the visualizations generated from the constructed *origin-destination matrices*. The results demonstrate that integrating advanced data analysis and visualization techniques can lead to significant improvements in the analysis of mobility information.

TABLA DE CONTENIDOS

	Página
Índice de Tablas	1
Índice de Figuras	1
1 Introducción	2
1.1. Motivación	2
1.2. Objetivos	3
1.3. Estructura general del documento	4
2 Marco Teórico y Antecedentes	5
2.1. Movilidad	5
2.1.1. Sistema de transporte público de pasajeros urbano	6
2.2. Demanda de transporte público	8
2.2.1. Sistema automatizado de cobro de Tarifas - AFC	9
2.2.2. Localización de vehículos automatizada - AVL	9
2.2.3. Matrices origen destino	10
2.3. Comparativa de métodos de inferencia OD y fundamentación del método Trip Chaining	12
2.4. Visualización	16
2.4.1. Matrices de correlación	16
2.4.2. Gráficos de barras o histogramas	17
2.4.3. Gráficos de torta	18
2.4.4. Mapas de calor	19
2.4.5. Diagramas de Sankey	20
3 Diseño y desarrollo del trabajo	22
3.1. Metodología general de trabajo	22
3.2. Fuentes de datos	25
3.3. Construcción de las Matrices Origen Destino	30

TABLA DE CONTENIDOS

3.3.1.	Análisis Exploratorio de Datos	30
3.3.2.	Elección y aplicación del modelo algorítmico MOD	31
3.4.	Visualización	38
3.4.1.	Visualizaciones gráficas y tablas	38
3.4.2.	Visualizaciones en mapas interactivos	40
3.5.	Tecnologías y recursos utilizados	41
4	Resultados y Discusiones	45
4.1.	Construcción de la matriz OD	45
4.2.	Prototipo de sistema web	48
4.2.1.	Visualizaciones gráficas y tablas	50
4.2.2.	Visualizaciones en mapas interactivos	64
4.3.	Resumen del capítulo	75
5	Conclusiones, desafíos y perspectivas futuras	78
5.1.	Conclusiones	78
5.2.	Dificultades encontradas	80
5.3.	Trabajos futuros	80
A	Anexos	82
A.1.	Código fuente de funciones destacadas	82
	Bibliografía	89
	Lista de Abreviaturas y Siglas	95

ÍNDICE DE TABLAS

TABLA	Página
2.1. Distancias máximas de transferencia utilizadas en la literatura especializada.	15
2.2. Tiempos máximos de transferencia utilizados en la literatura especializada.	16
3.1. Código de grupos, cantidad de líneas y número de paradas totales por grupo del sistema mendoTRAN.	26
3.2. Descripción de los campos y valores de los registros del conjunto de datos SUBE.	27
3.3. Descripción de los campos de los registros de las transacciones de biciTRAN.	28
4.1. Resultados por configuración del algoritmo de <i>trip chaining</i> . Se reportan viajes únicos, viajes con transbordo y transbordos inferidos para combinaciones de distancia máxima de transferencia (600-1400 m) y ventana temporal (1-3 h). Los conteos corresponden a la ejecución después del refinamiento descrito en la Sección 3.3.2.	48

ÍNDICE DE FIGURAS

FIGURA	Página
2.1. Foto ilustrativa de bus característico del Área Metropolitana de Mendoza	7
2.2. Foto ilustrativa de metrotranvía de Gran Mendoza	7
2.3. Foto ilustrativa de bicicletas compartidas de Gran Mendoza	8
2.4. Ejemplo de matriz origen-destino	11

2.5.	Diagrama de funcionamiento del algoritmo de Trip Chaining. B denota Estación de inicio (<i>Boarding</i>), A denota estación de destino (<i>Alighting</i>), MTD denota máxima distancia de transferencia (<i>Maximum Transfer Distance</i>)	14
2.6.	Matriz de correlación de viajes entre departamentos de Mendoza.	17
2.7.	Histograma de viajes a cada hora diferenciando por género.	18
2.8.	Gráfico de torta que muestra las proporciones de las categorías de descuento.	19
2.9.	Mapa de calor de viajes sobre un mapa cartográfico del área metropolitana de Mendoza.	20
2.10.	Diagrama de Sankey, muestra la proporción de viajes entrantes y salientes de cada departamento analizado en la provincia de Mendoza (mendoTRAN).	21
3.1.	Metodología de trabajo propuesta	24
3.2.	Publicidad de SUBE en Mendoza. Fuente Facebook mendoTRAN.	26
3.3.	Captura del conjunto de datos de mendoTRAN.	27
3.4.	Captura del conjunto de datos de biciTRAN.	28
3.5.	Ubicación de las estaciones de biciTRAN en la ciudad de Mendoza.	29
3.6.	Comparativa entre los valores de la columna NRO_TARJETA en mendoTRAN, en azul están agrupados los valores que no aportan información (nulos o de valor genérico) en naranja se encuentran los valores que son útiles para el algoritmo de <i>trip chaining</i> (TC).	31
3.7.	En rojo aparecen las paradas del conjunto de datos y su distribución casi lineal y cuadrículada debido al truncamiento en el tercer decimal de las coordenadas, mientras que en verde se aprecian las paradas en el punto exacto donde deben ir, las flechas indican la parada que fue inferida. . . .	32
3.8.	Distribución de viajes en la madrugada (00–05 h). Se observa la finalización del servicio a las 2 a.m. y su reanudación a las 4 a.m.	33
3.9.	Rutas posibles captadas por el algoritmo de selección	37
3.10.	Captura de las transacciones de la tarjeta que más viajes tenía del conjunto de datos, se observan dos rectángulos rojos que marcan los dos viajes que tuvo la tarjeta en el mismo día.	38
4.1.	Vista de los valores más comunes para el campo “NROTARJETA”.	46
4.2.	Vista previa de filas con valores nulos en distintos campos.	47
4.3.	Vista principal de la aplicación.	49
4.4.	Página de visualizaciones gráficas y tablas.	50
4.5.	Opciones del ítem Matrices Origen Destino en visualizaciones gráficas y tablas (mendoTRAN).	50

4.6. Matriz Origen Destino entre departamentos (mendoTRAN).	51
4.7. Opciones del filtro de la tabla matriz origen-destino entre departamentos (mendoTRAN).	51
4.8. Origen y destino entre departamentos Godoy Cruz, Maipú, Capital, Guaymallén (mendoTRAN).	52
4.9. Origen y destino entre departamentos unificando Godoy Cruz, Maipú, Capital, Guaymallén (GC M C G) (mendoTRAN).	53
4.10. Origen y destino entre departamentos - por género (mendoTRAN).	53
4.11. Opciones del filtro de la tabla matriz origen-destino entre departamentos - por género (mendoTRAN).	54
4.12. Opciones del ítem Histogramas en visualizaciones gráficas y tablas.	54
4.13. Comparativa de Viaje Único vs. Transbordo (mendoTRAN).	55
4.14. Distribución de viajes por hora y género (mendoTRAN).	55
4.15. Distribución de viajes por hora y género filtrando con un clic (mendoTRAN).	56
4.16. Distribución de viajes por hora y género filtrando con doble clic (mendoTRAN).	56
4.17. Distribución de viajes por hora y tipo de descuento (mendoTRAN).	57
4.18. Distribución de viajes con y sin descuento por hora (mendoTRAN).	58
4.19. Distribución de viajes por hora y tipo de usuario (biciTRAN).	59
4.20. Distribución de viajes por hora y género de usuario (biciTRAN).	59
4.21. Opciones del ítem Gráficos de torta en visualizaciones gráficas y tablas (mendoTRAN).	60
4.22. Opciones del ítem Gráficos de torta en visualizaciones gráficas y tablas (biciTRAN).	60
4.23. Viajes agrupados por género (mendoTRAN).	61
4.24. Viajes agrupados por género (biciTRAN).	61
4.25. Viajes agrupados por descuento (mendoTRAN).	62
4.26. Viajes agrupados por tipo de usuario y temporada (biciTRAN).	63
4.27. Diagrama de Sankey de viajes del conjunto de datos mendoTRAN.	64
4.28. Diagrama de Sankey de viajes del conjunto de datos biciTRAN.	64
4.29. Página de visualizaciones en mapas interactivos.	65
4.30. Visualización de paradas de llegada para una línea seleccionada y contador de descensos (mendoTRAN).	66
4.31. Visualización de <i>pop-up</i> en el mapa (mendoTRAN).	66
4.32. Visualización de la cantidad de salidas por hora (mendoTRAN).	67
4.33. Visualización de descensos por parada (mendoTRAN).	68
4.34. Pares de paradas con más viajes (mendoTRAN).	69

4.35. Visualización de destinos por departamento (mendoTRAN).	70
4.36. Visualización de viajes por hora con ascensos y descensos (mendoTRAN).	71
4.37. Visualización de mapa de calor de viajes para salidas (mendoTRAN).	72
4.38. Visualización de mapa de calor de viajes para llegadas (mendoTRAN).	72
4.39. Pares de estaciones con más duración de viaje (biciTRAN).	73
4.40. Visualización de mapa de calor de viajes para salidas filtrado por mes (biciTRAN).	74
4.41. Capacidad por estación (biciTRAN).	75

INTRODUCCIÓN

1.1. Motivación

La movilidad urbana sostenible se define como el conjunto de políticas, estrategias y soluciones orientadas a satisfacer las necesidades de desplazamiento de la población actual sin comprometer las posibilidades de las generaciones futuras. Este enfoque prioriza la eficiencia energética, la equidad social y la reducción de impactos ambientales, promoviendo modos de transporte alternativos como el transporte público eficiente, la movilidad activa (caminar y andar en bicicleta) y los sistemas de micromovilidad compartida.

Para avanzar hacia este modelo, resulta esencial optimizar la planificación y gestión de la red de transporte mediante el análisis de datos confiables y herramientas de visualización. En el caso de Mendoza, una ciudad que ha experimentado un crecimiento sostenido en población y actividad económica¹, el desafío de gestionar eficazmente el sistema de transporte se ha vuelto cada vez más complejo. La congestión vehicular, las emisiones contaminantes y la necesidad de mejorar la calidad de vida exigen soluciones basadas en evidencia.

Este proyecto de tesina se centra en el análisis y visualización de datos de transporte urbano con el objetivo de mejorar la toma de decisiones en políticas de movilidad. Para ello, se utilizan datos provenientes del sistema de pago electrónico del transporte público de Mendoza, que permiten construir matrices origen-destino (matrices OD, *origin-destination matrices*) y detectar patrones de viaje, horarios de alta demanda y

¹Fuente: Mendoza Gobierno

modos de transporte utilizados.

Las matrices origen-destino constituyen una herramienta clave para comprender cómo se desplazan los ciudadanos dentro de la ciudad, mientras que las visualizaciones gráficas y cartográficas facilitan la identificación de rutas, comportamientos y oportunidades de mejora. Además, se incorporan datos del sistema de bicicletas compartidas, ampliando el alcance del análisis hacia una visión más integrada de la movilidad urbana.

Como resultado, se desarrolla una aplicación interactiva de visualización de datos que permite a los tomadores de decisiones explorar la información de manera intuitiva y dinámica. Esta herramienta busca facilitar la identificación de problemas y oportunidades, contribuyendo a una planificación más eficiente y sostenible.

En síntesis, el propósito de esta tesina es doble: mejorar la experiencia de los usuarios del transporte público y optimizar la gestión del sistema de movilidad urbana en Mendoza. Al integrar análisis de datos avanzados y visualizaciones accesibles, se busca aportar soluciones concretas para una ciudad más conectada, equitativa y resiliente.

1.2. Objetivos

El objetivo principal de esta tesina final de grado consiste en evaluar algoritmos para la construcción de matrices origen-destino producto del análisis de las transacciones de las tarjetas de pago (*Smart Cards*).

Con base en lo anterior, como objetivos específicos se pueden mencionar:

- **Objetivo 1:** En lo que respecta a las Ciencias de la Computación, se busca entender, aplicar y evaluar modelos algorítmicos existentes sobre la construcción de matrices origen destino para ser utilizados como insumo en la predicción de la demanda futura del sistema de transporte público de pasajeros urbano (STPPU). La obtención de matrices origen destino para el STPPU permite observar la demanda del servicio y planear mejoras.
- **Objetivo 2:** En cuanto a la aplicación de los algoritmos, se espera implementar un sistema que permita visualizar la información de los datos de transporte público de pasajeros, posibilitando analizar el impacto de la movilidad urbana en el Gran Mendoza que sirva de soporte para los tomadores de decisiones.

1.3. Estructura general del documento

El presente documento se encuentra organizado de la siguiente manera: En este capítulo se presentan las motivaciones del trabajo, los objetivos generales y específicos.

- **Capítulo 2:** Brinda un marco teórico sobre conceptos asociados a movilidad urbana, en particular aquellos relacionados con el transporte público urbano de pasajeros y las tarjetas de pago utilizadas para abonar el servicio, la descripción de los algoritmos a utilizar en esta tesina final de grado y por último, conceptos de tipos de visualizaciones comúnmente utilizados en movilidad urbana. A su vez, este capítulo describe los trabajos relacionados en estas áreas.
- **Capítulo 3:** Detalla el diseño y desarrollo del trabajo, se explica el algoritmo canónico utilizado para la construcción de las matrices origen destino y sus variantes. Se incluye, también, la puesta en producción del sitio web asociado a esta tesina final de carrera.
- **Capítulo 4:** Expone los principales resultados obtenidos a partir de los diferentes experimentos realizados y el diseño e implementación de la herramienta de visualización. Además, se presenta un análisis de los resultados obtenidos.
- **Capítulo 5:** Presenta las conclusiones del trabajo, dificultades encontradas durante el transcurso del mismo y trabajos futuros.

MARCO TEÓRICO Y ANTECEDENTES

En este capítulo se introducen los conceptos asociados al transporte de pasajeros público urbano, los sistemas de tarjetas de pago más utilizados, el algoritmo seleccionado para la inferencia de destinos, la creación de matrices origen-destino y los tipos de visualización de datos utilizados.

2.1. Movilidad

La movilidad urbana se refiere a la capacidad de los habitantes de una ciudad para desplazarse de un lugar a otro de manera eficiente, segura y sostenible. Este concepto abarca todos los modos de transporte disponibles en una ciudad, incluidos vehículos privados, transporte público, bicicletas y desplazamientos a pie. La movilidad urbana es un componente clave del desarrollo urbano, ya que impacta directamente en la calidad de vida de los ciudadanos, la economía local y el medio ambiente (Mohammed and Oke, 2023).

La movilidad se proporciona a través de múltiples modalidades, incluyendo muchos viajes que requieren el uso de más de un tipo de transporte. Por ejemplo, ir en automóvil al trabajo o al colegio suele incluir un trayecto a pie. Muchos viajes incluyen no solo desplazamientos a pie, sino también transbordos a otros modos de transporte. Así pues, la movilidad forma parte inherente de una medida multimodal del rendimiento del sistema (Meyer and Institute of Transportation Engineers, 2016).

2.1.1. Sistema de transporte público de pasajeros urbano

Un sistema de transporte público de pasajeros urbano (STPPU) es una red organizada de servicios de transporte destinados a trasladar a las personas dentro de una ciudad o área metropolitana. Este sistema está diseñado para ser accesible, eficiente y asequible, y juega un papel crucial en la movilidad urbana y la sostenibilidad de las ciudades modernas (Meyer and Institute of Transportation Engineers, 2016).

Dentro de los componentes del STPPU se pueden destacar:

- **Vehículos de transporte:** Incluyen autobuses, tranvías, metros, trenes ligeros, bicicletas, monopatines eléctricos y en algunas ciudades, ferris. Estos vehículos están diseñados para transportar personas de manera eficiente.
- **Infraestructura:** Comprende las estaciones, paradas, vías y terminales necesarias para el funcionamiento del sistema. También incluye carriles exclusivos para autobuses y tranvías, así como instalaciones de mantenimiento y almacenamiento y otras como estaciones de bicicletas compartidas.
- **Redes y rutas:** El sistema suele estar organizado en una red de rutas predefinidas que conectan diferentes partes de la ciudad. Estas rutas están diseñadas para maximizar la cobertura y la eficiencia del transporte, permitiendo a los usuarios llegar a sus destinos con facilidad.
- **Sistema de Cobro de Tarifas:** Involucra los métodos y tecnologías utilizados para el pago de los viajes, como las tarjetas inteligentes, aplicaciones móviles y sistemas automatizados de cobro de tarifas.
- **Gestión y operación:** Incluye la planificación, regulación y supervisión del sistema de transporte. Esto abarca desde la programación de horarios hasta la gestión del tráfico y la respuesta a emergencias.
- **Tecnología de la Información y Comunicación (TIC):** Utiliza sistemas avanzados de información para proporcionar a los usuarios datos en tiempo real sobre horarios, rutas y posibles retrasos. Las TIC también facilitan la optimización de las operaciones y el mantenimiento del sistema.

En el Área Metropolitana de Mendoza o Gran Mendoza, el STPPU principal es conocido como mendoTRAN¹, un sistema de transporte público que unifica buses (Figura 2.1) y metrotranvías (Figura 2.2) y convive con un sistema de bicicletas compartidas

¹mendoTRAN

conocido como biciTRAN² (Figura 2.3). El sistema mendoTRAN tiene una distribución troncal, es decir, conecta diferentes puntos de la ciudad sin necesidad de pasar por el centro de la misma (distribución radial). Mientras que biciTRAN utiliza estaciones para el alquiler de las bicicletas y su posterior devolución.



Figura 2.1: Foto ilustrativa de bus característico del Área Metropolitana de Mendoza³.



Figura 2.2: Foto ilustrativa de metrotranvía de Gran Mendoza⁴.

²biciTRAN

³Fuente: Sitio oficial del Gobierno de Mendoza

⁴Fuente: Sitio oficial del Gobierno de Mendoza



Figura 2.3: Foto ilustrativa de bicicletas compartidas de Gran Mendoza⁵.

2.2. Demanda de transporte público

La demanda en un sistema de transporte público urbano abarca el número de usuarios del transporte, así como sus patrones de movimiento dentro de la ciudad. Este término engloba el total de pasajeros, las franjas horarias durante las cuales realizan sus viajes y sus elecciones de rutas.

Se puede definir un viaje en transporte público como el desplazamiento desde un punto de origen hasta un destino final. Por ejemplo, alguien podría salir de su casa (origen), tomar un autobús y llegar a su trabajo (destino). Estos viajes pueden incluir varias etapas, como transbordos entre diferentes medios de transporte, y son esenciales para trazar los patrones de movilidad urbana.

Entender la demanda es crucial porque ayuda a planificar y operar el sistema de transporte de manera eficiente. Permite diseñar rutas y horarios que satisfagan mejor las necesidades de los usuarios, identificar dónde se necesita invertir en infraestructura y mejorar la calidad del servicio. Todo esto contribuye a reducir la congestión vehicular y las emisiones de gases contaminantes, promoviendo un desarrollo urbano más sostenible (Wu et al., 2025).

Para medir la demanda, se pueden utilizar varias técnicas, como encuestas a los usuarios (Blum et al., 2010), análisis de datos de tarjetas de transporte (Barry et al., 2002), modelos estadísticos para inferir la cantidad de ascensos y descensos por estación (Hazelton, 2010), aplicaciones para teléfonos móviles (Barabino et al., 2024), y el uso

⁵Fuente: Sitio oficial del Gobierno de Mendoza

de sensores para monitorear en tiempo real el número de pasajeros (Wang et al., 2011). Esta tesina final de grado se enfoca principalmente en los datos de las tarjetas de pago y el uso de aplicaciones de pago para construir información sobre la demanda. A continuación, se introducen conceptos relacionados con la demanda.

2.2.1. Sistema automatizado de cobro de Tarifas - AFC

En todo el mundo, los sistemas de transporte han empezado a depender cada vez más del sistema automatizado de cobro de tarifas (AFC, *Automated Fare Collection*). Estos sistemas AFC varían en cuanto a sus aplicaciones, desde tarjetas inteligentes hasta sistemas basados en teléfonos inteligentes. Muchas ciudades utilizan una variedad de sistemas de tarjetas inteligentes, como la *Oyster card* de Londres, la *Smart Link* de Nueva York, la *Charlie card* de Boston, la *Yikatong* de Beijing y la *Octopus card* de Hong Kong (Zannat and Choudhury, 2019; Mohammed and Oke, 2023). En estos sistemas, los pasajeros registran su actividad al abordar o descender mediante la validación de la tarjeta inteligente o la activación del teléfono inteligente en un dispositivo dedicado, entre otros métodos, según el tipo de red y las políticas de las agencias de transporte. La información individual de los pasajeros es recopilada por el sistema AFC y almacenada por el operador de transporte (Nassir et al., 2011) y queda a resguardo del ente regulador. La información proporcionada constituye un recurso ininterrumpido, detallado y completo para el análisis de la demanda. Gracias a su precisión en cada operación individual, es posible emplearla con el fin de obtener una comprensión más profunda y clara sobre los patrones de movilidad dentro de nuestras urbes.

Los sistemas automatizados de cobro de tarifas pueden ser categorizados como:

- Abiertos: Solo se graban los ingresos de pasajeros, comunes en áreas donde las tarifas son fijas.
- Cerrados: Tanto las entradas como las salidas son registradas, típicos en lugares donde las tarifas dependen de la distancia recorrida.

En Mendoza, el sistema AFC es de tipo abierto y se conoce como Sistema Único de Boleto Electrónico (SUBE)⁶.

2.2.2. Localización de vehículos automatizada - AVL

El conocimiento exacto de la ubicación de los vehículos dentro de una red es crucial para una estimación precisa de los puntos de partida y llegada. En muchas

⁶Sistema de pagos para el transporte público de pasajeros.

infraestructuras, los sistemas AFC no registran información geoespacial asociada a las transacciones de pago. Sin embargo, este desafío se aborda a través de los sistemas de localización de vehículos automatizados (AVL, *Automated Vehicle Location*), que integran datos de GPS con los del sistema AFC para proporcionar una solución efectiva. Los dispositivos GPS facilitan la recolección de lugares de cobro en adición al tiempo guardado por el AFC. En algunos casos, pueden surgir problemas cuando se combinan las dos fuentes de datos. Primordialmente, las imprecisiones de los GPS pueden llevar a una errónea inferencia de las estaciones de salida o arribo. Si los datos de AVL no están disponibles, o son sumamente imprecisos, la localización de los vehículos puede ser inferida alternando fuentes, tales como horarios del transporte.

2.2.3. Matrices origen destino

Una matriz origen-destino es una herramienta utilizada en el análisis y la planificación del transporte para representar los patrones de desplazamiento entre diferentes ubicaciones dentro de un área geográfica. Cada celda de la matriz indica el número de viajes que se realizan desde una ubicación de origen específica a una ubicación de destino específica en un período de tiempo determinado. Esta matriz es fundamental para entender los flujos de tráfico y para la planificación de infraestructuras de transporte eficientes. Sin embargo, debido al tipo de estructura del sistema de transporte público, esto puede ser un desafío en la práctica.

Existen numerosos tipos de matrices origen-destino; cada tipo representa la forma de agrupar esos orígenes y destinos, obteniendo así como resultado la posibilidad de analizar las matrices bajo determinadas características como agrupamiento por hora, por día, por semana laboral, etc. Mohammed and Oke (2023) utiliza dos tipos de matrices origen-destino, a nivel de ruta, las cuales contienen el flujo de pasajeros desde las estaciones de origen hasta las estaciones de llegada para la misma ruta; y a nivel de red, las cuales contienen el flujo de pasajeros desde las estaciones de origen a las estaciones de destino, incluyendo las estaciones de transferencias a través de múltiples rutas. Diferentes tipos de datos han sido utilizados para estimar las matrices origen-destino, redes de telefonía celular (Wang et al., 2013), sistemas de peajes (Wang and Zhang, 2011), información de Sistema de Posicionamiento Global (GPS, Global Positioning System) y encuestas de viajes (Ge and Fukuda, 2016).

En el ámbito del transporte público, la implementación cada vez más común de sistemas AFC y AVL ha resultado en una mayor accesibilidad a los datos generados por las tarjetas inteligentes. Esta situación ha propiciado el desarrollo de innovadoras aplicaciones que aprovechan esta información, tanto histórica como en tiempo real

(Nassir et al. (2011); Alsger et al. (2015)), (Zúñiga et al. (2020); Ait-Ali and Eliasson (2022)).

Un sistema de matriz origen-destino, ya sea a nivel de toda la red o de una ruta individual, no suele estar disponible en la mayoría de las agencias de transporte público. La mayor ventaja de emplear datos de sistemas automáticos de recolección radica en su capacidad para establecer estas matrices origen-destino de manera eficiente para la agencia. Comparado con el método manual, estas ventajas no son únicamente cuantitativas, sino también cualitativas (Cui, 2006):

- El costo de obtención de la matrices origen-destino se reduce de forma notable.
- La matriz resultante se basa en una muestra de tamaño mucho mayor.
- El proceso es más adecuado para la automatización, por lo que puede ejecutarse con mayor rapidez y actualizarse con más frecuencia.
- Este procedimiento puede complementarse con encuestas dirigidas, de modo que se obtenga una visión más completa y rentable del comportamiento de los viajeros.

Estimaciones más precisas de las matrices origen-destino se pueden emplear eficazmente en la optimización de múltiples aspectos del transporte público, por ejemplo, infiriendo el propósito de los viajes (Alsger et al., 2018), estimación de precios y capacidad del transporte (Ait Ali et al., 2020), o estimando los costos de hacinamiento en los vehículos públicos (Hörcher et al. (2017); Yap et al. (2020)). La Figura 2.4 muestra un ejemplo de una matriz origen-destino.

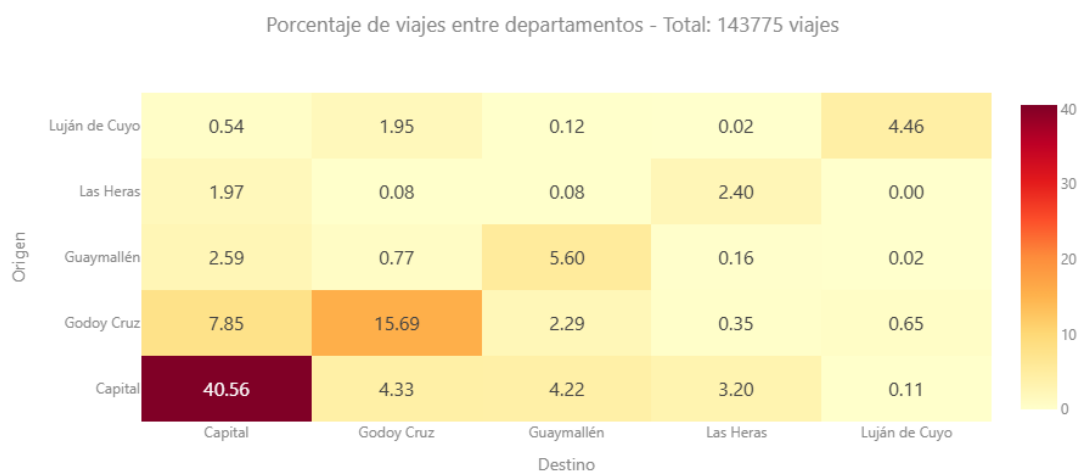


Figura 2.4: Ejemplo de matriz origen-destino

2.3. Comparativa de métodos de inferencia OD y fundamentación del método Trip Chaining

Existen múltiples modelos o algoritmos para la generación de matrices origen-destino y la inferencia de estaciones de destino o salida en los sistemas AFC. En primer lugar, los métodos basados en encuestas son costosos en términos de tiempo y dinero y pueden resultar muy inexactos tanto en escalas de espacio como de tiempo. Los métodos basados en aplicaciones de teléfonos móviles requieren de sistemas más sofisticados en las unidades de transporte público, como *wifi* o cruzar datos con operadores de telefonía móvil. En cambio, el uso de sensores de monitoreo para la estimación de las matrices origen-destino comprende un alto costo en infraestructura, *hardware* y mantenimiento para sostener el sistema. En cuanto a los modelos predictivos, si bien son variados y tienen muy buenos resultados, requieren de grandes cantidades de datos para obtener resultados confiables. Los métodos basados en tarjetas inteligentes ofrecen resultados más exactos y permiten una recolección de datos más eficiente a un costo mínimo; el sistema que provee los datos para esta tesina es de este tipo. Es por estas razones que se eligen estos métodos frente a los anteriores mencionados.

En esta sección se presenta la propuesta algorítmica utilizada para el desarrollo de este trabajo. En particular, se muestra el modelo general del algoritmo *trip chaining* (TC) que se utiliza como base para el modelo propuesto en esta tesina final de grado. Los sistemas AFC proporcionan gran cantidad de datos, como las coordenadas del inicio del viaje, la estación de destino, datos categóricos sobre el pasajero, como sexo, tipo de tarifa utilizada, etc. Estos datos recolectados pueden ser explotados mediante las matrices origen-destino.

El método de *trip chaining* fue introducido por James Barry en 2002 (Barry et al., 2002) para estimar la matriz origen-destino (matriz OD, *origin-destination matrix*) de los viajes del transporte público en la ciudad de *New York*, utilizando los datos de solo entrada del sistema AFC. Este método tiene una gran aceptación por su simplicidad y eficiencia en los sistemas de tarjetas inteligentes, tanto es así que ha sido utilizado en numerosos trabajos en el área (Mohammed and Oke, 2023).

El método tiene como objetivo inferir la localización de la parada de destino del viaje de cada pasajero. Esto es necesario cuando no es posible la obtención de la información de las paradas de cada pasajero a partir de los datos del sistema AFC, como suele ser el caso de la mayoría de los sistemas de transporte en la actualidad. Por lo tanto, es un paso esencial para generar una estimación inicial de los viajes entre paradas dentro de cada ruta. Asimismo, el método de TC puede ser utilizado para

deducir los transbordos, lo que ayuda a asignar tramos de viaje para completar toda la secuencia de viajes de un pasajero y, por lo tanto, forma parte integral de la creación de un modelo de origen-destino de red completa (Mohammed and Oke, 2023).

Barry et al. (2002) introdujo dos supuestos basados en el comportamiento de los viajes en transporte público que luego fueron probados utilizando los datos de una encuesta recolectados por el consejo de transporte de la Ciudad de Nueva York. Estos supuestos son:

- Un alto porcentaje de los viajeros regresan a la estación de destino de su viaje previo al iniciar su siguiente viaje.
- Un alto porcentaje de los viajeros finaliza su último viaje del día en la estación donde comenzaron el primero (conocido como el supuesto del viaje simétrico del día).

La idea detrás del segundo supuesto es que los usuarios comienzan su primer viaje en algún lugar cercano a su hogar, y eventualmente su último viaje tiene como objetivo su regreso a casa. La Figura 2.5 representa estos conceptos de manera simple; se muestran en verde el inicio de cada viaje y en rojo la estación de llegada o arribo. El círculo verde indica la zona de incidencia, es decir, la distancia máxima que se considera que el pasajero podría recorrer para ir a la siguiente estación, y en donde hay dos círculos superpuestos se podría decir que una misma estación de llegada es considerada una de salida para el siguiente viaje.

Este algoritmo es comúnmente implementado como una heurística a base de reglas. *Trip chaining* toma las estaciones de salida obtenidas de los AFC y AVL y crea zonas de incidencias alrededor de cada estación de salida para inferir la estación de arribo que precede a la transacción. De esta forma, si una estación de la línea de transporte en la cual viaja el pasajero está lo suficientemente cerca de la estación de salida de la siguiente transacción, se puede asumir que dicha estación es su estación de destino.

Zhao (2004) mejoró los supuestos propuestos por Barry et al. (2002) sugiriendo las siguientes tres hipótesis:

- h1** Los viajeros no utilizan otros tipos de transporte privado como autos, bicicletas o motocicletas entre dos viajes de transporte público sucesivos.
- h2** Hay un límite en la distancia que un viajero camina en un transbordo (En el manuscrito de Zhao la distancia es de 400 metros).
- h3** El último destino del día suele ser la misma estación que el primer viaje del día.

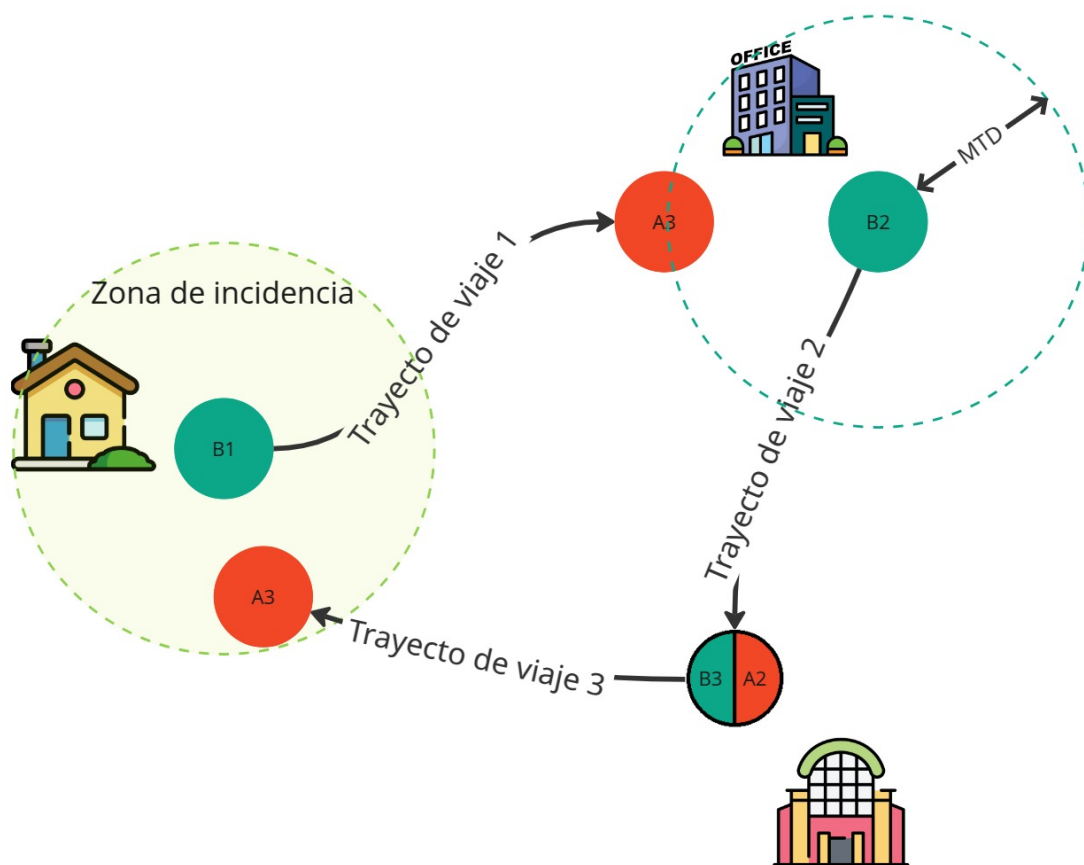


Figura 2.5: Diagrama de funcionamiento del algoritmo de Trip Chaining. B denota Estación de inicio (*Boarding*), A denota estación de destino (*Alighting*), MTD denota máxima distancia de transferencia (*Maximum Transfer Distance*)

Estas últimas dos suposiciones (h2-h3) son usadas ampliamente en las investigaciones de matrices origen-destino (la Tabla 2.1 agrupa algunas de ellas). La segunda conjetura (h2) de Zhao (2004) requiere dos parámetros: distancia y tiempo a pie. La distancia a pie es definida como la distancia máxima que los pasajeros pueden considerar caminar al siguiente transbordo. En la Tabla 2.1 se observan las distancias máximas utilizadas en diferentes investigaciones con sus respectivas ciudades, mientras que en la Tabla 2.2 se pueden observar los tiempos máximos de transferencias que fueron utilizados.

Las hipótesis propuestas por Zhao antes mencionadas son utilizadas en este trabajo como punto de partida para el desarrollo del algoritmo de TC. A su vez, las comparativas de distancias y tiempos (tablas 2.1 y 2.2) fueron utilizadas para la búsqueda de los mejores valores para los parámetros del algoritmo a aplicar en esta tesina final de grado.

Puesto que el algoritmo de TC utiliza el comienzo de un viaje para estimar el destino del viaje previo, se presenta una limitación cuando un usuario ha realizado

2.3. COMPARATIVA DE MÉTODOS DE INFERENCIA OD Y FUNDAMENTACIÓN DEL MÉTODO TRIP CHAINING

Distancia (metros)	Estudio	Ciudad
402	Zhao (2004)	Chicago, US
530	Alsger et al. (2016)	SEQ, Australia
640	Nunes et al. (2016)	Porto, Portugal
750	Gordon et al. (2013)	London, UK
800	Alsger et al. (2015), Nassir et al. (2011)	SEQ, Australia London, UK
900	Yan et al. (2019)	London, UK
1000	Munizaga and Palma (2012), Munizaga et al. (2014)	Santiago, Chile
1100	Cui (2006)	Chicago, US
2000	Trépanier et al. (2007)	

Tabla 2.1: Distancias máximas de transferencia utilizadas en la literatura especializada.

únicamente un viaje en el día (problema de la transacción única). Además del problema de la transacción única, otros inconvenientes encontrados son:

- Pasajeros caminan más que la distancia máxima.
- Errores en los datos como valores faltantes o transacciones duplicadas. La solución aceptada para estos errores es eliminar la transacción.
- Pasajeros cuyo destino del último tramo de viaje no es el mismo que el punto inicial de la jornada.
- El uso de modos de transporte privados entre dos transacciones de *smart card*.
- Cuando la distancia entre sucesivas paradas de destino y de salida es más larga de lo asumido.

Para aplicar el supuesto de simetría del viaje y eliminar las transacciones únicas en un día, es necesario definir el inicio y fin del día. Generalmente, la medianoche no es el momento en que el transporte público tiene su menor demanda. De acuerdo a las matrices origen-destino de Santiago, Chile, descritas en los estudios de Munizaga et al. (2014), luego de las 0:00 a.m. hay actividades que pertenecen a fines de las cadenas del día previo. Por esta razón, la hora de menor demanda, que corresponde a las 4:00 a.m. para Santiago, fue detectada y utilizada como la medianoche virtual. Este cambio reduce el error y la cantidad de viajes ignorados. Para mendoTRAN, no fue necesario este cambio dado que los autobuses operan en un rango horario distinto a 24 horas;

Tiempo (minutos)	Estudio
18	Barry et al. (2009), Ali et al. (2015), Bagchi and White (2005)
30	Liu and Zhou (2019), Munizaga et al. (2014), Munizaga and Palma (2012)
60	Alsger et al. (2016), Alsger et al. (2015), MOSALLANEJAD et al. (2019), N He et al. (2015), Yan et al. (2019)
90	Hofmann and O'Mahony (2005), Kumar et al. (2018), Nassir et al. (2011), Chu and Chapleau (2008), Gordon et al. (2013), Seaborn et al. (2009)
Variable	Yap et al. (2017)
<frecuencia de tránsito	Huang et al. (2020)

Tabla 2.2: Tiempos máximos de transferencia utilizados en la literatura especializada.

debido a esto, se tomó el inicio del día cuando comienzan las frecuencias de transporte, 4:00 a.m., y el fin del mismo al finalizar la última frecuencia, 3:00 a.m.

2.4. Visualización

La representación gráfica de datos proporciona una manera de ilustrar información abstracta. Permite detectar patrones, tendencias y correlaciones que de otra forma podrían pasar desapercibidos. Debido a la forma en que el cerebro humano procesa la información, utilizar gráficos para visualizar grandes cantidades de datos complejos es más fácil que en los informes, bases de datos o resultados de análisis tradicionales. La visualización de datos ayuda a representar datos que no son útiles en forma cruda, lo que es de particular ayuda cuando se trabaja con grandes cantidades de datos (*big data*).

A continuación se presentan las visualizaciones comúnmente utilizadas en el ámbito del reporte de información asociada a movilidad urbana, las cuales se han escogido para ser implementadas en este trabajo final de carrera.

2.4.1. Matrices de correlación

Las matrices de correlación a menudo son utilizadas para representar matrices origen-destino, ya que permiten detectar rápidamente patrones de flujo entre cada par de origen-destino. Estas son útiles para resumir grandes volúmenes de datos de viajes, destacando zonas de alta demanda y patrones de movilidad de manera intuitiva.

Por ejemplo, Geng and Zhang (2022) analizan en Pittsburgh y Nueva York la relación entre la centralidad de la red vial y la intensidad de movilidad humana mediante *tweets* geolocalizados. En su estudio, calculan diversas métricas y muestran un mapa de calor de la correlación por pares entre ellas.

La Figura 2.6 muestra un ejemplo de una matriz de correlación o matriz origen-destino, en donde se puede visualizar la proporción de viajes entre departamentos de Mendoza.

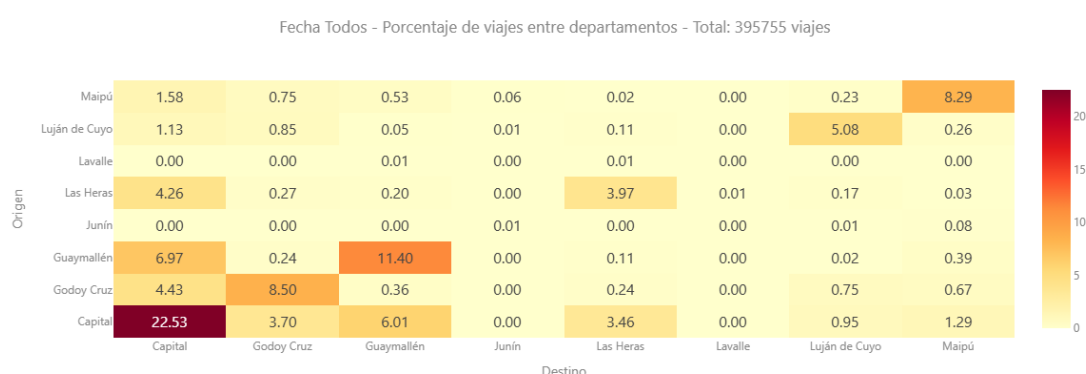


Figura 2.6: Matriz de correlación de viajes entre departamentos de Mendoza.

2.4.2. Gráficos de barras o histogramas

Los gráficos de barras consisten en barras o contenedores encerrados en un sistema de coordenadas cartesianas con ejes x e y . Estos emparejan grupos nominales usando barras separadas en un eje con valores escalados ordinalmente, indicados por la altura de las barras en el otro eje. Debido a que los gráficos de barras se especializan en entidades individuales, separando un grupo del otro, estos son ideales para comparaciones discretas entre grupos (Zhao and Gaschler, 2022). Los gráficos de barras e histogramas son frecuentes para comparar categorías o distribuciones de datos cuantitativos. Un ejemplo reciente es Molina-Campoverde et al. (2024), quienes estudian trayectorias GPS de usuarios para clasificar modos de transporte. En su análisis exploratorio presentan histogramas de velocidad para cada medio de transporte considerado.

La Figura 2.7 muestra un histograma construido en esta tesina; en él se pueden visualizar los viajes realizados a cada hora separando por el género de los usuarios; a su vez, cada barra contiene tres colores para distinguir entre las diferentes categorías (Femenino, Masculino y No especificado).

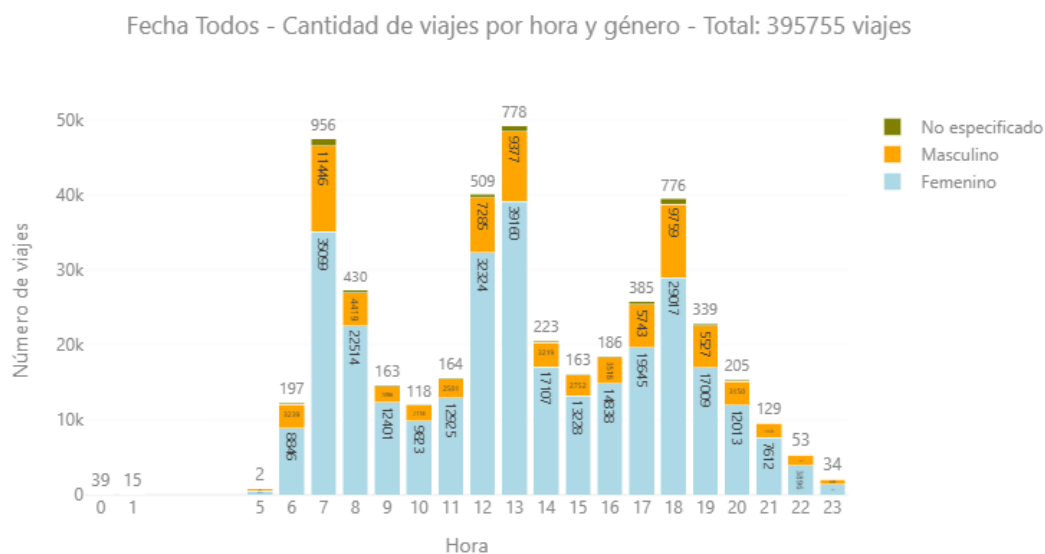


Figura 2.7: Histograma de viajes a cada hora diferenciando por género.

2.4.3. Gráficos de torta

Los gráficos de torta (*pie charts*) consisten en un círculo dividido por múltiples líneas radiales que parten desde un punto central en un sistema de coordenadas polares. Suelen usar ángulos, porciones o arcos (áreas radiales) y oscuridad para representar valores de grupos. Estos tipos de gráficos se utilizan frecuentemente para mostrar información de manera proporcional. Los gráficos circulares son superiores a las barras para hacer comparaciones cuando hay que combinar elementos, por ejemplo, al contrastar A+B con C+D (Zhao and Gaschler, 2022). En Silva et al. (2022) (herramienta de visualización ANTENNA sobre datos de telefonía móvil) cada nodo geográfico (por ejemplo, celdas hexagonales) se representa con un pequeño gráfico de torta. Cada sector del círculo codifica el número de viajes de un tipo específico: por ejemplo, viajeros que llegan (llegadas), salen (partidas) o pasan (itinerantes) por esa ubicación. El ángulo de cada porción es proporcional al conteo de ese tipo de actividad. De este modo, la torta resume en un vistazo la composición del flujo en cada área. El radio total de la torta refleja el volumen total de viajes. Esta visualización permite identificar rápidamente zonas que funcionan principalmente como origen (mayor sector de partidas), destino (sector de llegadas) o como puntos de paso.

La Figura 2.8 es un ejemplo de un gráfico de torta en el cual se muestran las proporciones de cada categoría de descuentos del conjunto de datos de mendoTRAN.

Fecha Todos - Proporción de viajes por tipo de descuento - Total: 144517 viajes

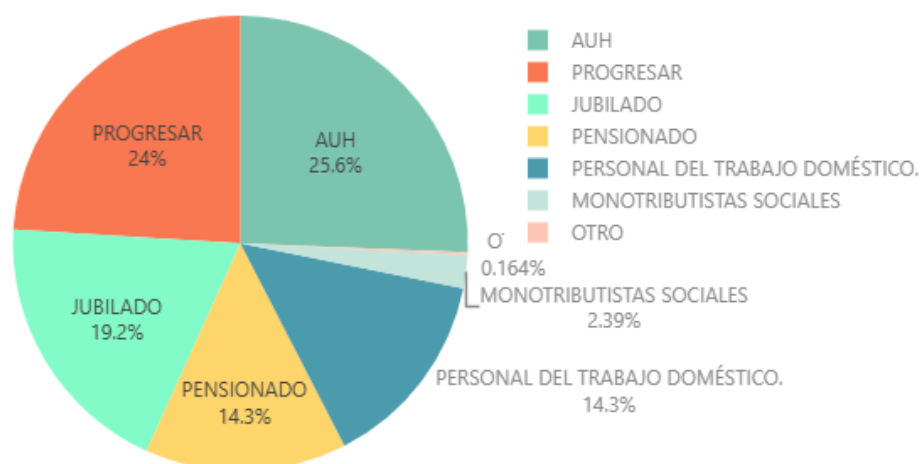


Figura 2.8: Gráfico de torta que muestra las proporciones de las categorías de descuento.

2.4.4. Mapas de calor

Desde el punto de vista de la clasificación de los métodos de presentación de datos realizada por MacEachren y DiBiase (MacEachren and DiBiase, 1991), los mapas de calor o *heat maps* pueden clasificarse como mapas continuos y suaves. La creciente popularidad de los mapas de calor proviene de su atractivo y de la facilidad de creación utilizando diversas bibliotecas de mapeo (Netek et al., 2019). A pesar de su aplicación común, no se ha evaluado si son una solución efectiva como mapas para una lectura rápida en un entorno web. De manera similar, no se ha verificado hasta qué punto su nivel de detalle (generalización) es un tema clave.

Los mapas de calor fueron importados a la cartografía desde las técnicas de visualización de datos, al igual que otros tipos de mapas ya bien establecidos en la cartografía, como los diagramas, gráficos, puntos o coropletas (Bertin, 1983). Los mapas de calor son visualizaciones para la representación gráfica de la densidad de fenómenos espaciales, generalmente medidos en puntos. DeBoer (2015) destaca el hecho de que el término en sí mismo no es unívoco; puede denotar tanto el mapa de densidad (independientemente del método utilizado) como el proceso de estimación de datos de punto a superficie (estimación de densidad de puntos)(Słomska-Przech et al., 2021).

Hu et al. (2022) analiza datos de bicicletas compartidas en Beijing y genera mapas de calor basados en análisis de densidad de *kernel* para visualizar la distribución espacial de viajes en bicicleta durante horas pico. El *heat map* muestra geográficamente cuántos viajes de bicicleta caen en cada ubicación, y ayuda a evidenciar patrones espaciales (por ejemplo: áreas céntricas vs. periféricas) y diferencias temporales en la movilidad urbana ciclista.

La Figura 2.9 muestra un mapa de calor de viajes sobre un mapa cartográfico del área metropolitana de Mendoza, en el cual se visualiza la densidad de salidas que ocurrieron a determinada hora en el sistema mendoTRAN. Las zonas pintadas en rojo indican una mayor cantidad de salidas de las paradas, mientras que las zonas en azul indican menor cantidad. La transparencia total del color indica que los alrededores no muestran salidas en ese momento.

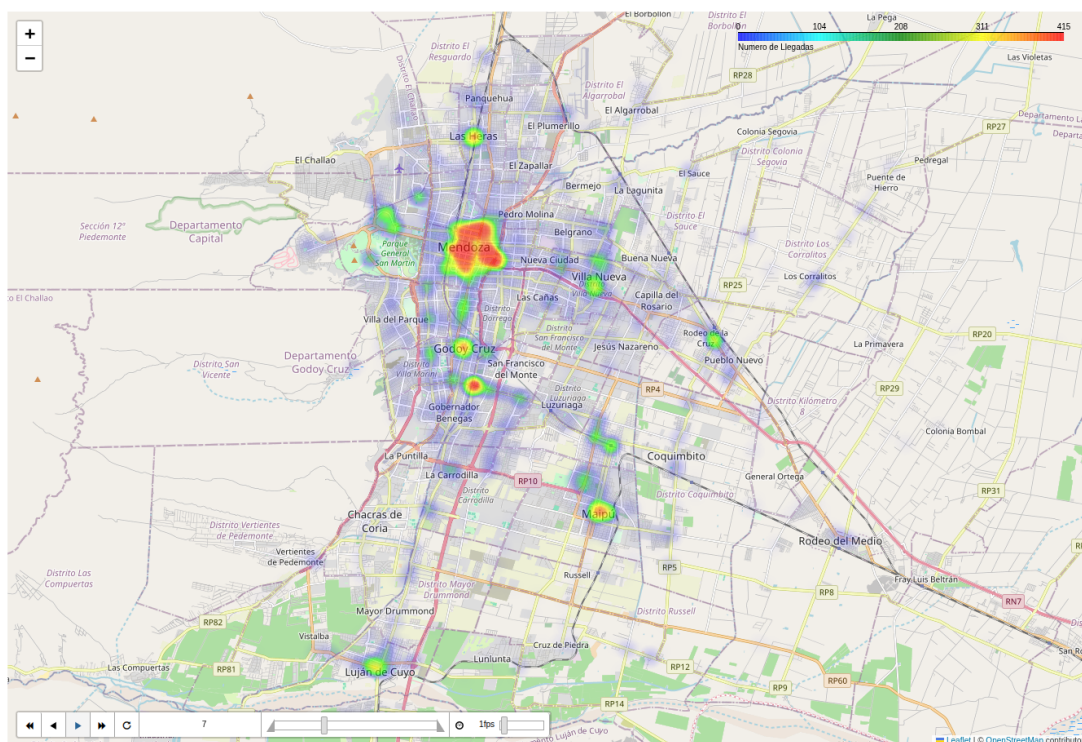


Figura 2.9: Mapa de calor de viajes sobre un mapa cartográfico del área metropolitana de Mendoza.

2.4.5. Diagramas de Sankey

Los diagramas de Sankey son utilizados para representar flujos de movilidad. Keller et al. (2022) utilizan los diagramas para analizar la eficiencia del transporte público, donde los nodos del diagrama representan el número de veces que un origen (nodo izquierdo) o un destino (nodo derecho) es utilizado en un viaje. El tamaño de un nodo

representa la suma de las conexiones entrantes o salientes. Las líneas conectan los orígenes y destinos entre sí y representan el número de conexiones de cada relación con su tamaño. Estos diagramas ayudan a analizar conexiones frecuentemente solicitadas y brindan un análisis cercano de la eficiencia de la red de transporte público.

La Figura 2.10 ilustra la relación de los viajes entrantes y salientes para cada departamento evaluado en la provincia de Mendoza en el sistema mendoTRAN; como se mencionaba previamente, un nodo o enlace de mayor tamaño indica una mayor afluencia de pasajeros.

Diagrama de Sankey para viajes entre departamentos

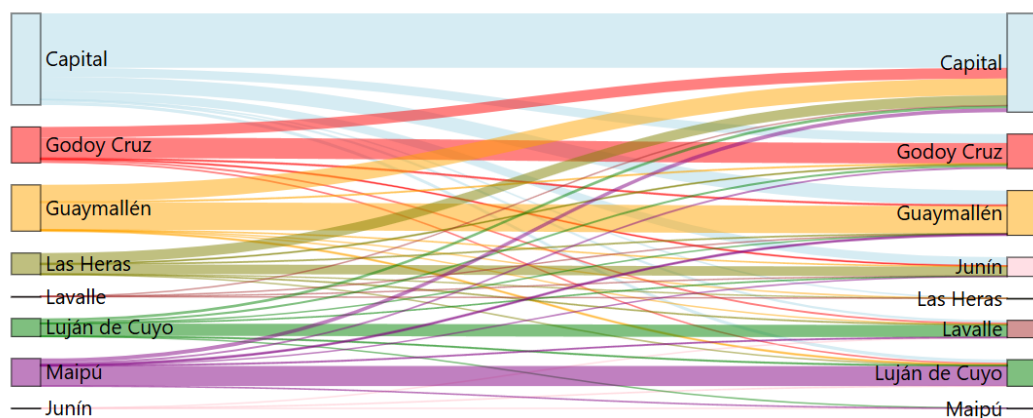


Figura 2.10: Diagrama de Sankey, muestra la proporción de viajes entrantes y salientes de cada departamento analizado en la provincia de Mendoza (mendoTRAN).

La investigación realizada sobre inferencia de destinos, generación de matrices origen-destino y las visualizaciones anteriormente mencionadas que son utilizadas dentro del prototipo de sistema para la ayuda a la toma de decisiones mostró una base sólida para la realización de los objetivos propuestos en esta tesina final de grado. En el siguiente capítulo se explica en más detalle qué datos se utilizaron, cómo se ejecutaron los algoritmos desarrollados, cómo se crearon las visualizaciones utilizadas y también la implementación del sistema de visualización objetivo.

DISEÑO Y DESARROLLO DEL TRABAJO

Para alcanzar los objetivos establecidos en esta tesina, es fundamental establecer una metodología sólida que posibilite la organización y administración sistemática de las diversas tareas y recursos involucrados.

Este capítulo detalla el diseño de la metodología de trabajo utilizada, junto a los métodos a través de los cuales se pretende llegar a obtener resultados (Sección 3.1), los datos utilizados (Sección 3.2), los pasos para la construcción de las matrices origen-destino (Sección 3.3), la generación de gráficos de visualización (Sección 3.4) y las tecnologías de hardware y software utilizadas en la implementación (Sección 3.5).

3.1. Metodología general de trabajo

En esta sección se describen todos los pasos establecidos como parte del proceso llevado a cabo para construir las matrices origen-destino y el visualizador de la información obtenida.

La secuencia de pasos definida en esta tesina final de grado se puede observar en la Figura 3.1. En primer lugar, a partir de lo analizado en el capítulo anterior, se planifican los pasos a seguir en el proyecto, tomando como punto de partida una recopilación de trabajos realizados por distintos investigadores. Se dedica la primera etapa al análisis de los trabajos previos y la selección de los métodos a implementar y desarrollar. Posteriormente, se solicita a la entidad encargada de la administración de transporte, el Ministerio de Transporte de la Nación ¹, los datos correspondientes a las

¹Ministerio de Transporte Argentina

transacciones de las tarjetas de pago del Sistema Único de Boleto Electrónico (SUBE) para la aplicación de los algoritmos y a la entidad encargada de la administración de los datos de transporte provincial, la Sociedad de Transporte de Mendoza (STM)², los datos correspondientes a las transacciones del sistema de bicicletas compartidas biciTRAN.

Después de recopilar los datos requeridos, se utilizan métodos de Análisis de Datos Exploratorio (EDA, *Exploratory Data Analysis*) (Mukhiya and Ahmed, 2020) en los conjuntos de datos, además de identificar las columnas y distinguir entre las que eran numéricas y las que contenían registros categóricos, enumerando los valores posibles de estas últimas. Luego, se eliminan las filas que tengan celdas nulas o vacías y aquellas con las longitudes o latitudes en 0, así como también aquellas que no presenten identificador de usuario o el mismo fuera un valor por defecto. Además, se eliminan columnas que contengan un solo valor a lo largo del conjunto, es decir, que no aportan datos relevantes. Después de realizar la limpieza del conjunto de datos, se selecciona el conjunto de datos de mendoTRAN y se comienza a calcular los orígenes y destinos mediante el algoritmo de TC. Para esto, es necesario un paso previo que consiste en determinar cuál transacción precede a otra. Debido a que las transacciones no cuentan con minutos y segundos, existen casos en los que hay más de una transacción para una misma hora. Este preprocesamiento se realiza construyendo y analizando listas ordenadas de posibles rutas.

Como se muestra en la Figura 3.1, finalmente se definen los métodos de visualización que se presentan, seleccionando aquellos que mejor se ajustan al objetivo del trabajo, es decir, que ofrecen datos enriquecidos para la toma de decisiones (visualizaciones estáticas) y también aquellos tipos de visualización que pueden ser insertados en una interfaz interactiva (visualizaciones dinámicas).

En las próximas secciones, se proporciona una descripción minuciosa y completa de cada fase que compone la metodología completa aplicada en este trabajo final de carrera, así como de las tecnologías y recursos empleados en su desarrollo.

²Sociedad de Transporte de Mendoza

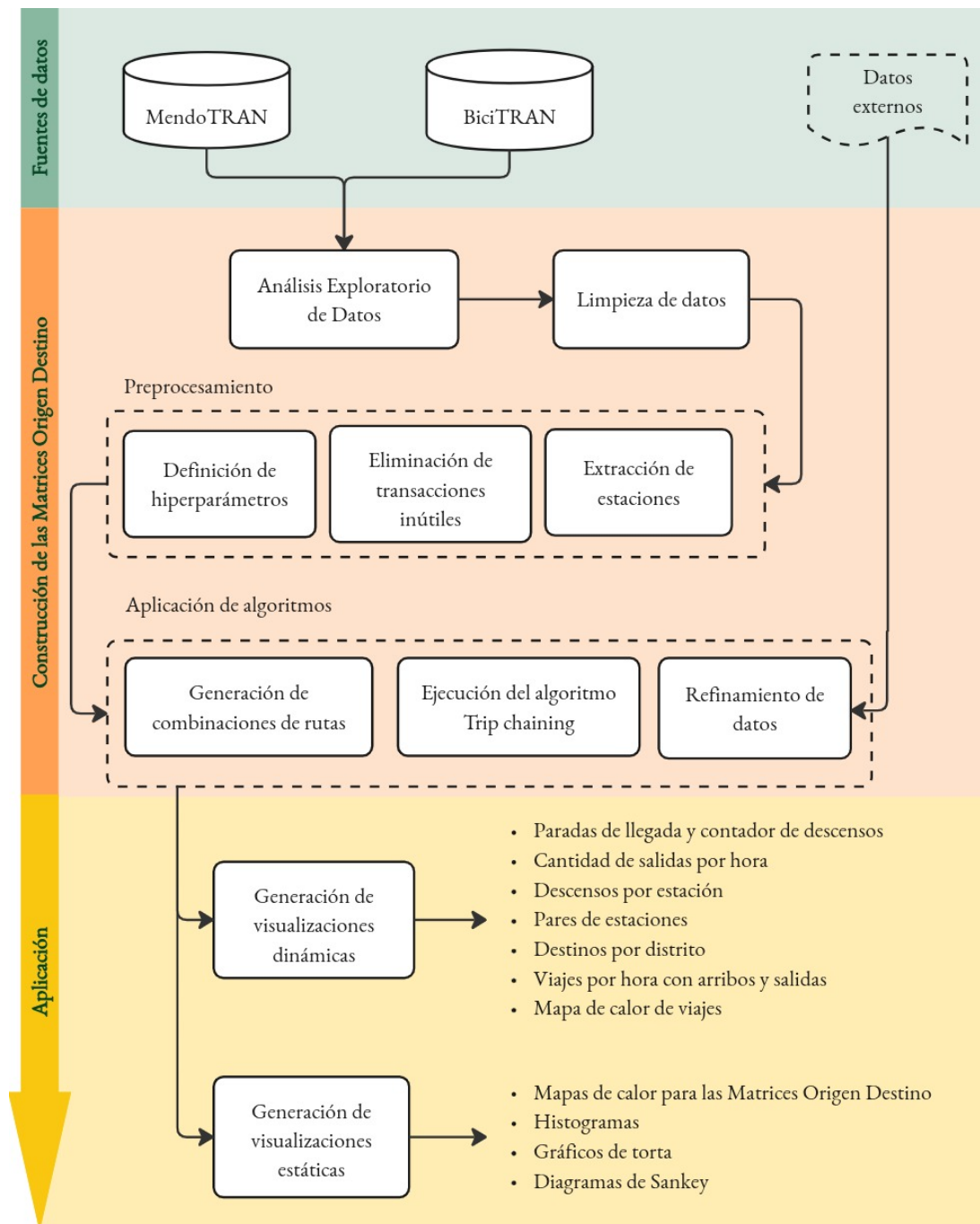


Figura 3.1: Metodología de trabajo propuesta

3.2. Fuentes de datos

El SUBE es un AFC utilizado en algunas provincias de Argentina, unificando así el sistema de cobro de transporte público. Al momento de adquirir la tarjeta, el usuario puede optar por registrar sus datos personales y acceder a tarifas diferenciales (estudiantes, jubilados y usuarios con tarifa social). Las transacciones son realizadas en el momento en que el usuario aproxima la tarjeta a un validador, dispositivo físico presente en las estaciones o dentro de los modos de transporte.

El sistema almacena los datos básicos de cada transacción (identificador único, fecha, horario y coordenadas donde se origina la transacción), sobre la tarjeta (usuario, número consecutivo de viaje de la tarjeta), sobre la tarifa aplicada (tarifa vigente, tipo de tarifa pagada, subsidio) y sobre el servicio (parada, modo, línea, ramal, sentido, vehículo, etc.). A su vez, estas transacciones pueden ser asociadas a los datos de GPS que suelen tener estas máquinas validadoras, lo que permite identificar para cada transacción la latitud y longitud donde se produjo. Entre las ventajas de estos sistemas se pueden encontrar que facilitan el proceso de movilidad en sí mismo al acelerar los tiempos de pago, permiten aplicar tarifas diferenciales según zonas y relevan al conductor de realizar dicha tarea (Pelletier et al., 2011). Por otro lado, su implementación permite la recopilación de una masiva cantidad de datos administrativos de dichas transacciones cuya potencialidad puede ser explotada a partir de un proceso de inteligencia de datos para analizar la movilidad a partir de un mejor entendimiento de patrones de viajes y operaciones sobre el sistema de transporte.

mendoTRAN es la red de transporte público del Área Metropolitana de Mendoza (AMM), utiliza como medio de pago el sistema SUBE (ver Figura 3.2). Este sistema entró en vigencia en Mendoza el 2 de enero de 2019³, depende directamente del Ente de Movilidad Provincial (EMOP), engloba nueve grupos de buses y uno de metro-tranvía, sumando así 382 líneas de transporte distribuidas como se pueden observar en la Tabla 3.1. La franja horaria de operación de la red de transporte va desde las 4 a.m. hasta las 3 a.m., por lo que, al no estar operando, no se generan datos en esas horas.

³Nota periodística: Más de tres años viajando con el mendoTRAN (Unidiversidad)

Código de Grupo	Cantidad de líneas por grupo	Número de Paradas por grupo
Grupo 100	8	199
Grupo 200	56	1614
Grupo 300	31	1030
Grupo 400	27	1119
Grupo 500	42	1107
Grupo 600	32	1231
Grupo 700	87	2143
Grupo 800	51	1872
Grupo 900	46	1628
Metrotranvía	2	50

Tabla 3.1: Código de grupos, cantidad de líneas y número de paradas totales por grupo del sistema mendoTRAN.



Figura 3.2: Publicidad de SUBE en Mendoza. Fuente Facebook mendoTRAN.

El conjunto de datos de las transacciones provistas por mendoTRAN consiste en 5 días de transacciones entre el 17 y el 21 de abril de 2023, contiene 3.842.663 registros y 16 columnas, cuyas características se detallan en la Tabla 3.2 y se visualizan en la Figura 3.3 de ejemplo, como se puede observar en esta última, cada registro contiene información que permite identificar inequívocamente cada transacción. Por cuestiones de salvaguarda de identidad del usuario de la tarjeta, solo se muestra un pequeño subconjunto de transacciones y sus identificaciones ofuscadas.

Campo	Valor
NROTARJETA	Contiene el identificador de la tarjeta que realizó la transacción, en este ítem, el valor por defecto o nulo es definido como “ACOM” .
JURISDICCION	Valor categórico que indica si la empresa que prestó el servicio pertenece al gobierno de la provincia o no. El único valor en el conjunto de datos encontrado fue “PROVINCIAL” .
ENTIDAD	Nombre de la empresa que prestó el servicio.
INTERNO	Identificador de la transacción.
CODIGOCONTRATO	Identificador del vehículo.
STATION_ID	Identificador único de la parada.
RAMAL	Número de línea del bus.
LINEA	Número y definición de la línea.
TIPO_TARIFA	Dato categórico para especificar si el usuario cuenta con algún tipo de descuento. Este dato contiene tres valores posibles: “Tarifa_desc_ATSF” y “Tarifa_desc_local” (para transacciones con descuento), “Tarifa_normal” (para el resto de transacciones).
MOTIVO_ATS	Valor categórico que especifica que tipo de descuento se aplicó. Aquí hay 7 valores diferentes “PROGRESAR” , “JUBILADO” , “AUH” , “PENSIONADO” , “OTRO” , “MONOTRIBUTISTAS SOCIALES” , “PERSONAL DEL TRABAJO DOMÉSTICO” y un valor nulo para las transacciones sin descuento.
SEXO	Género del usuario. Dato categórico con 3 valores posibles “Masculino” , “Femenino” o nulo en caso de no ser especificado.
LONGITUD	Latitud en que fue realizada la transacción.
LATITUD	Longitud en que fue realizada la transacción.
FECHA_HORA	Fecha y hora de la transacción en formato “DD/MM/AAAA HH” .

Tabla 3.2: Descripción de los campos y valores de los registros del conjunto de datos SUBE.

NROTARJETA	MODO	JURISDICCION	ENTIDAD	LINEA	RAMAL	INTERNO	TIPO_TARIFA	CODIGOCONTRATO	MOTIVO_ATS	SEXO	FECHA_HORA	ETAPA_RED_SUBE	LONGITUD	LATITUD	STATION_ID
DIFF3CEB082BBB6E	COL	PROVINCIAL	AUTOTR TRAPICHE SRL Y TRANSP GRAL B.MITRE SRL LITE	LINEA 700	700	18962CACF5D438FE952E	Tarifa_desc_local	568	NaN	F	20230418 14	NaN	-68.868	-33.010	M5381
9394A3ABD720B4E3	COL	PROVINCIAL	AUTOTR TRAPICHE SRL Y TRANSP GRAL B.MITRE SRL LITE	LINEA 700	700	18962CACF5D438FE952E	Tarifa_desc_ATSF	621	PERSONAL DEL TRABAJO DOMÉSTICO.	F	20230418 14	NaN	-68.868	-33.010	M5381
NaN	COL	PROVINCIAL	AUTOTR TRAPICHE SRL Y TRANSP GRAL B.MITRE SRL LITE	LINEA 700	700	18962CACF5D438FE952E	Tarifa_desc_local	583	NaN	M	20230418 14	NaN	-68.868	-33.010	M5381
NaN	COL	PROVINCIAL	AUTOTR TRAPICHE SRL Y TRANSP GRAL B.MITRE SRL LITE	LINEA 700	700	18962CACF5D438FE952E	Tarifa_desc_local	581	NaN	M	20230418 14	NaN	-68.868	-33.010	M5381
ACOM	COL	PROVINCIAL	AUTOTR TRAPICHE SRL Y TRANSP GRAL B.MITRE SRL LITE	LINEA 700	700	18962CACF5D438FE952E	Tarifa_normal	602	NaN	NaN	20230418 14	NaN	-68.868	-33.010	M5381

Figura 3.3: Captura del conjunto de datos de mendoTRAN.

BiciTRAN es un sistema automatizado de alquiler de bicicletas compartidas de la provincia de Mendoza, depende directamente del Ente de Movilidad Provincial (EMOP), cuenta con 89 estaciones distribuidas a lo largo de 5 departamentos (Figura 3.5). Este sistema tiene la característica de que funciona las 24 horas, aunque la reposición de las bicicletas solo funciona de 6 a.m. a 10 p.m. El conjunto de datos de las transacciones del sistema biciTRAN representa las transacciones del sistema desde marzo de 2024 a marzo de 2025, contiene 172.266 registros y 12 columnas, las cuales se detallan en la Tabla 3.3 y se visualizan en la Figura 3.4 de ejemplo. Se ha ofuscado el dato sensible *idUser* considerando el artículo 19 de protección de datos de la “Reglamentación del Sistema de Bicicletas Compartidas como servicio Anexo al servicio Público de Movilidad para el Área Metropolitana del Gran Mendoza” Res. 087 ^{4 5}.

Campo	Descripción
ID	Identificador de la transacción.
ID_USER	Contiene el identificador del usuario que realizó la transacción.
GENDER	Género del usuario. Dato categórico con 5 valores posibles “ Masculino ”, “ Femenino ”, “ No binario ”, “ No especificado ” u “ Otro ”.
USER_TYPE	Este campo indica si el usuario tiene domicilio en Mendoza (“ Local ”) o no (“ Turista ”).
TEST_USER	Columna que denota si el usuario es un usuario de prueba.
PLATE	Matrícula del vehículo utilizado en la transacción.
ORIGIN	Estación de origen de la transacción.
FROM_TIME	Momento en el que se originó la transacción. Tiene un formato “ AAAA-MM-DD hh:mm:ss ”.
DESTINATION	Estación de destino de la transacción.
TO_TIME	Momento en el que finalizó la transacción. Tiene un formato “ AAAA-MM-DD hh:mm:ss ”.
DURATION	Duración de la transacción.
STATUS	Estado de la transacción. Esta puede ser “ Finalizado ”, “ Finalizado automáticamente ”, “ Finalizado manualmente ” o “ Cancelado ”.

Tabla 3.3: Descripción de los campos de los registros de las transacciones de biciTRAN.

	id	idUser	gender	userType	testUser	plate	fromTime	toTime	duration	status	origin	destination
0	247729	117	F	Turista	False	BT388	2025-01-01 2:06:46	2025-01-01 2:34:16	00:27:29	Finalizado	72	8
1	247730	1	M	Turista	False	BT655	2025-01-01 2:07:05	2025-01-01 2:35:56	00:28:50	Finalizado	72	8
2	247754	8	NE	Turista	False	BT268	2025-01-01 3:34:46	2025-01-01 3:47:53	00:13:07	Finalizado	85	52

Figura 3.4: Captura del conjunto de datos de biciTRAN.

⁴Reglamentación del Sistema de Bicicletas Compartidas

⁵Reglamentación del Sistema de Bicicletas Compartidas (Anexo)

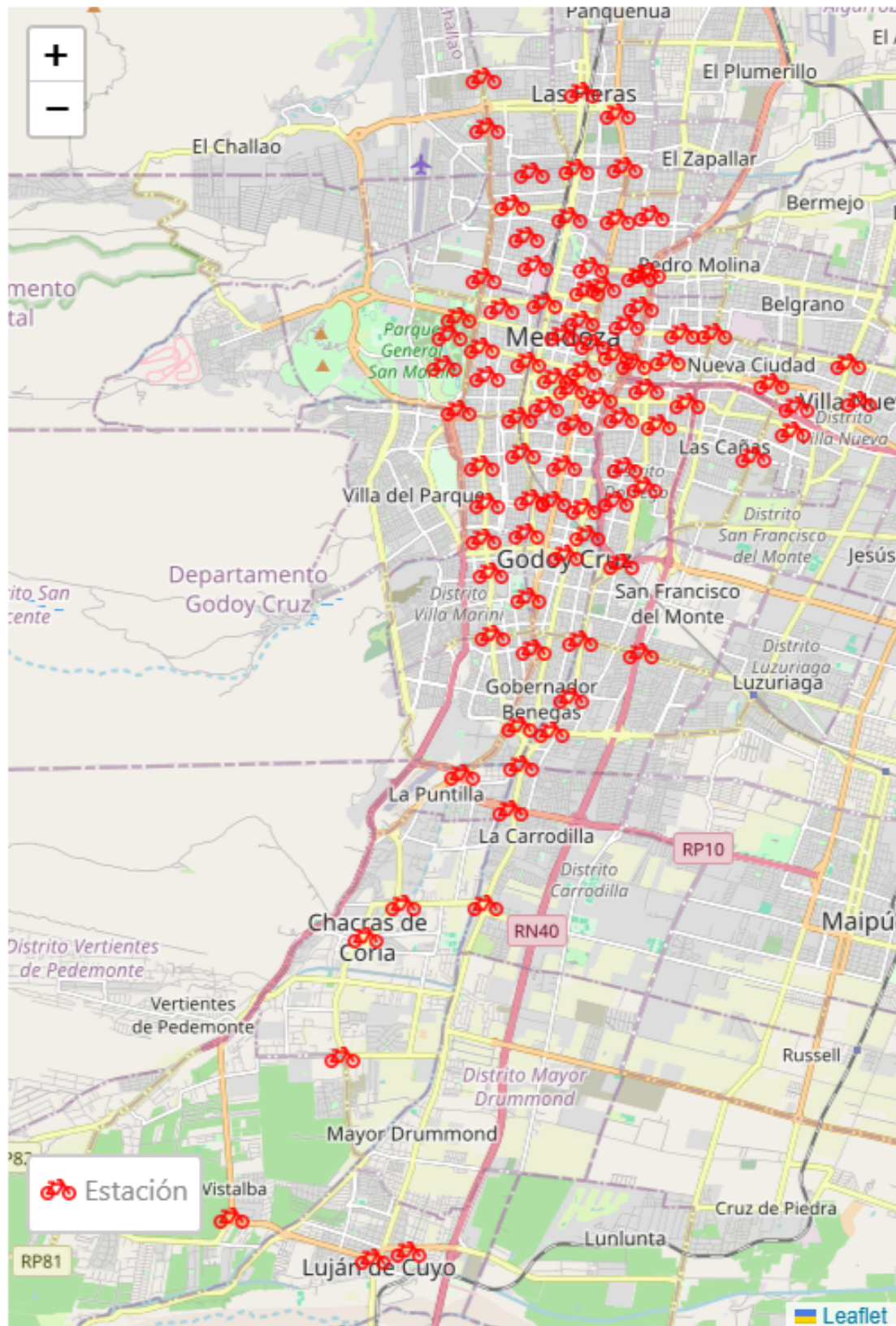


Figura 3.5: Ubicación de las estaciones de biciTRAN en la ciudad de Mendoza.

3.3. Construcción de las Matrices Origen Destino

A continuación se describe cómo se realizó la etapa de construcción de las matrices origen-destino para mendoTRAN. Iniciando por el análisis exploratorio de datos para crear una muestra uniforme adecuada para la implementación del algoritmo. Se describirá detalladamente el procedimiento empleado para descartar valores superfluos y eliminar columnas que no contribuyen con información significativa para las fases posteriores. A continuación, se explica la selección del algoritmo de *trip chaining* y su implementación en esta investigación. Este proceso se detalla a través de las fases de preprocesamiento de datos, la ejecución del método y el perfeccionamiento de los resultados, generando al final un conjunto de datos apto para ser utilizado en las fases de visualización.

3.3.1. Análisis Exploratorio de Datos

En esta sección se comentará el proceso de limpieza y exploración de datos realizado a las transacciones de SUBE previo a la etapa de ejecución del algoritmo de *trip chaining*. El objetivo de esta etapa fue generar un conjunto de datos libre de imperfecciones y adecuado para la ejecución del método. Como se dijo en el capítulo anterior, el *trip chaining* requiere que cada tarjeta tenga al menos dos viajes en un mismo día, no debe haber paradas con identificador nulo, no deben existir paradas con latitud o longitud nula y todas las transacciones deben tener fecha y hora válidas.

En primer lugar, se analizaron los valores de las columnas para detectar inconsistencias en los datos, por ejemplo, si las columnas eran de tipo categórico, los valores de las columnas no deberían diferir por mayúsculas, minúsculas, acentos o caracteres especiales. Así, se establece la convención para nombrar la categoría y se unifican valores similares. Luego, fue necesario eliminar valores nulos; si la fila tenía un valor nulo en algunas de las columnas clave como NRO_TARJETA, LATITUD, LONGITUD, HORA, FECHA, LINEA, esta se eliminaba. Uno de los inconvenientes de este conjunto de datos era que en la columna NRO_TARJETA, había valores genéricos (por ejemplo ACOM) que podían considerarse nulos; estos también eran eliminados. En la Figura 3.6, se presenta un análisis comparativo llevado a cabo en esta etapa, donde se observa la relación entre la proporción de valores que pueden ser anulados y los valores que podrían ser obtenidos. Una vez realizada la limpieza, se puede continuar con los siguientes pasos.

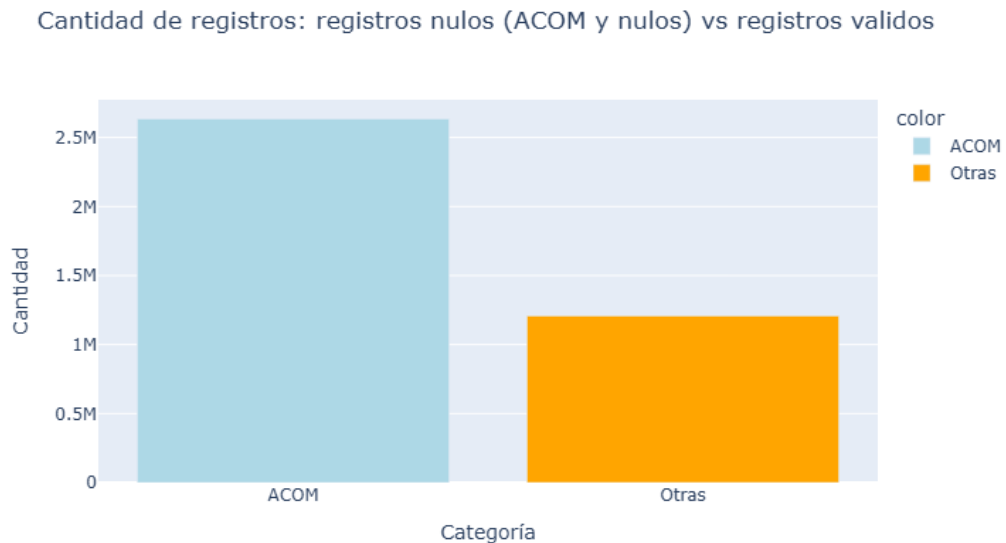


Figura 3.6: Comparativa entre los valores de la columna `NRO_TARJETA` en `mendoTRAN`, en azul están agrupados los valores que no aportan información (nulos o de valor genérico) en naranja se encuentran los valores que son útiles para el algoritmo de TC.

3.3.2. Elección y aplicación del modelo algorítmico MOD

Para la confección de las matrices origen-destino se utiliza el algoritmo de *trip chaining*. Este método fue seleccionado debido a su simplicidad en la práctica y su utilización en la bibliografía específica del tema (Sección 2.3). Para su aplicación, el algoritmo se dividió en tres etapas claramente definidas:

- Preprocesamiento de los datos de las transacciones de SUBE.
- Aplicación del algoritmo de *trip chaining* específicamente adaptado para los datos de SUBE.
- Análisis y refinamiento del algoritmo considerando como caso particular la ciudad de Mendoza.

La etapa de preprocesamiento consiste en obtener y preparar la información necesaria para la aplicación del algoritmo de *trip chaining*. En primer lugar, se extraen todas las coordenadas iniciales de cada transacción (latitud y longitud), utilizándolas para generar una lista preliminar de paradas. Es importante destacar que este proceso puede introducir errores, ya que las coordenadas provienen del momento exacto en que se registra la tarjeta y podrían variar ligeramente debido al movimiento del vehículo al partir de la parada.

Debido a las inexactitudes en las coordenadas dentro del conjunto de datos, es fundamental emplear fuentes externas para alinear las transacciones con las paradas correctas. Se recurre a la web de mendoTRAN y la interfaz de programación de aplicaciones (*API, application programming interface*) pública para acceder a todas las paradas oficiales. Dado que la amplia lista de paradas produce muchos datos irrelevantes, es decir, paradas sin transacciones registradas en el conjunto de datos, se busca una opción más eficaz. Se opta por identificar la parada correcta usando tanto las coordenadas como la línea de transporte de la transacción. Este método es sencillo pero preciso. Se agrupan todas las combinaciones de coordenadas y líneas desde el conjunto de datos y, por cada combinación, se consulta la *API* de mendoTRAN para identificar paradas cercanas. Estas paradas se clasifican según su proximidad, y se elige la más cercana que también pertenezca a la línea de transporte en cuestión.

La Figura 3.7 presenta la corrección final de coordenadas utilizando datos externos precisos. En rojo aparecen las paradas del conjunto de datos y su distribución casi lineal y cuadriculada debido al truncamiento en el tercer decimal de las coordenadas, mientras que en verde se aprecian las paradas en el punto exacto donde deben ir, las flechas indican la parada que fue inferida.

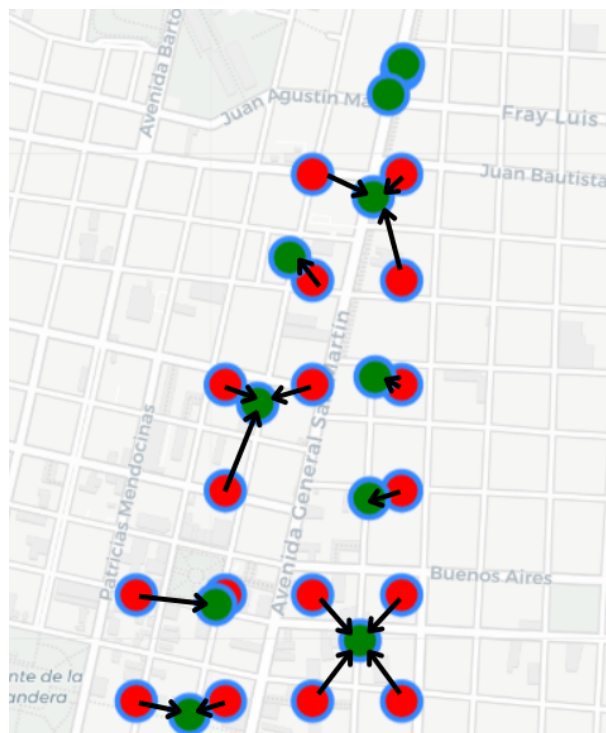


Figura 3.7: En rojo aparecen las paradas del conjunto de datos y su distribución casi lineal y cuadriculada debido al truncamiento en el tercer decimal de las coordenadas, mientras que en verde se aprecian las paradas en el punto exacto donde deben ir, las flechas indican la parada que fue inferida.

Considerando los horarios específicos del transporte público en la Área Metropolitana de Mendoza, se delimitó claramente el día operativo desde las 4:00 a.m. hasta las 3:00 a.m. del día siguiente, esto es comprobable mediante el análisis de las horas registradas en cada transacción. Para acotar el análisis, solo se verifica el número de transacciones entre las 00 y las 5 horas (Figura 3.8). Por último, se identificaron los días disponibles y se filtraron aquellas tarjetas que presentaban una sola transacción en todo el conjunto de datos, así como aquellas con múltiples transacciones totales, pero solo una por día, para manejarlas posteriormente.

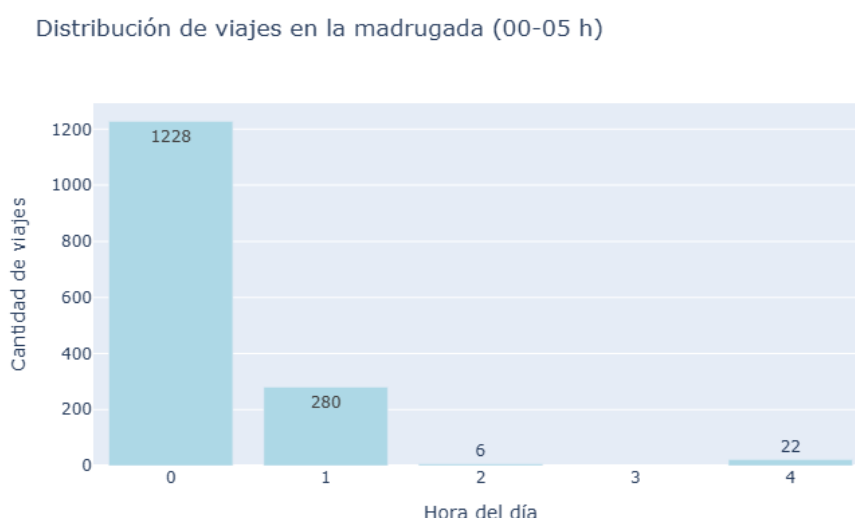


Figura 3.8: Distribución de viajes en la madrugada (00–05 h). Se observa la finalización del servicio a las 2 a.m. y su reanudación a las 4 a.m.

Dentro de la etapa de implementación del algoritmo de TC, se establecieron normas precisas basadas en los datos extraídos del sistema mendoTRAN. La verificación de su eficacia se llevó a cabo comparando distintas configuraciones alternativas, seleccionando finalmente aquellas que optimizaban al máximo el número total de transacciones incluidas en el análisis:

- *Distancia máxima a pie entre paradas: 1000 metros.*
- *Tiempo máximo entre transacciones consecutivas: 2 horas, considerando que, en el Área Metropolitana de Mendoza, el viaje más largo puede durar hasta hora y media y que el tiempo promedio entre frecuencias podría alcanzar media hora.*

Puesto que las transacciones proporcionadas por el Ministerio de Transporte estaban únicamente en formato de horas, sin la inclusión de minutos ni segundos, surgió

la necesidad de establecer un método preciso para organizar aquellas transacciones que compartieran la misma tarjeta, así como el mismo día y hora de registro. Con este fin, en esta tesina final de grado, se desarrollaron algoritmos específicos adaptados a los datos de SUBE en Mendoza, diseñados para posibilitar el ordenamiento de las transacciones. Este ordenamiento, a su vez, facilita tanto el proceso de asignación de destinos como el encadenamiento de los viajes.

El Algoritmo 3.1 genera múltiples listas clasificadas de transacciones, donde cada lista simboliza un itinerario potencial para la ruta en el día bajo análisis. Aunque el algoritmo no puede establecer el itinerario exacto, se puede afirmar que presenta el itinerario más óptimo y probable.

Algoritmo 3.1 Generar combinaciones de rutas

Entrada: df ▷ DataFrame de entrada, contiene las transacciones de una tarjeta

- 1: df_agrupado_por_dia $\leftarrow$ agruparPor(df, "fecha_ajustada")
- 2: rutas $\leftarrow$ []
- 3: **Para todo** grupo_dia $\in$ df_agrupado_por_dia **hacer**
- 4: ruta_del_dia $\leftarrow$ procesar_dia(grupo_dia)
- 5: agregarA(rutas, ruta_del_dia)
- 6: **Fin Para**
- 7: **return** rutas

Inicialmente, las transacciones se organizan por día virtual utilizando el campo fecha_ajustada, que se establece previamente de acuerdo con la ventana operativa especificada. Después, cada conjunto de transacciones agrupadas por día se procesa de forma individual.

Para procesar cada día, se utiliza el Algoritmo 3.2, en el cual las transacciones se agrupan nuevamente por hora y se procede acorde al tamaño de estos grupos:

- En el caso de que un grupo disponga de únicamente una transacción, esta se integra directamente en las rutas ya existentes o da lugar a la creación de una ruta nueva si no hay rutas previas disponibles.
- En un grupo de horario con múltiples transacciones, se elaboran todas las permutaciones posibles de dichas transacciones, las cuales se incorporan al final de las rutas actuales. Posteriormente, se evalúa cada secuencia generada al calcular su peso mediante la distancia total recorrida, implementando el Algoritmo 3.3. Después de esta evaluación, las rutas son ordenadas basándose en el peso asignado, y se seleccionan solo las tres mejores para evitar un incremento excesivo en la cantidad de casos a considerar.

Algoritmo 3.2 Algoritmo para procesar transacciones diarias

Entrada: grupo_dia ▷ Transacciones del dia agrupadas

- 1: df_agrupado_por_hora $\leftarrow$ agruparPor(grupo_dia, "hora")
- 2: rutas_por_dia $\leftarrow$ []
- 3: **Para todo** grupo $\in$ df_agrupado_por_hora **hacer**
- 4: **Si** tamaño(grupo) < 2 **Entonces**
- 5: **Si** tamaño(rutas_por_dia) = 0 **Entonces**
- 6: rutas_por_dia $\leftarrow$ [primerElemento(grupo)]
- 7: **Si no**
- 8: **Para todo** elemento $\in$ rutas_por_dia **hacer**
- 9: agregarA(elemento, primerElemento(grupo))
- 10: **Fin Para**
- 11: **Fin Si**
- 12: **Si no**
- 13: rutas_temporales $\leftarrow$ []
- 14: indices $\leftarrow$ [0, 1, ..., tamaño(grupo) - 1]
- 15: permutaciones $\leftarrow$ generarPermutaciones(indices)
- 16: **Para todo** permutacion $\in$ permutaciones **hacer**
- 17: ruta_temporal $\leftarrow$ []
- 18: **Para todo** valor $\in$ permutacion **hacer**
- 19: agregarA(ruta_temporal, grupo[valor])
- 20: **Fin Para**
- 21: agregarA(rutas_temporales, ruta_temporal)
- 22: **Fin Para**
- 23: nuevas_rutas $\leftarrow$ []
- 24: **Si** tamaño(rutas_por_dia) = 0 **Entonces**
- 25: rutas_por_dia $\leftarrow$ copiaDe(rutas_temporales)
- 26: **Si no**
- 27: **Para todo** elemento $\in$ rutas_por_dia **hacer**
- 28: **Para todo** ruta $\in$ rutas_temporales **hacer**
- 29: rutas_auxiliares $\leftarrow$ copiaDe(elemento)
- 30: concatenar(rutas_auxiliares, ruta)
- 31: agregarA(nuevas_rutas, rutas_auxiliares)
- 32: **Fin Para**
- 33: **Fin Para**
- 34: rutas_por_dia $\leftarrow$ copiaDe(nuevas_rutas)
- 35: **Fin Si**
- 36: rutas_por_dia $\leftarrow$ ordenar(rutas_por_dia, clave= calculaPesoRuta(x))
- 37: **Si** tamaño(rutas_por_dia) > 3 **Entonces**
- 38: rutas_por_dia $\leftarrow$ cortaLosPrimerosNElementos(rutas_por_dia, 3)
- 39: **Fin Si**
- 40: **Fin Si**
- 41: **Fin Para**
- 42: **return** rutas_por_dia

Para determinar el peso correspondiente a cada ruta producida, el Algoritmo 3.3 evalúa la suma de las distancias que el usuario cubre entre transacciones consecutivas. Durante este proceso, se identifica cuál parada en la ruta está más cercana a la parada ligada a la siguiente transacción. La distancia directa que conecta estas dos paradas se utiliza para establecer el peso de la ruta.

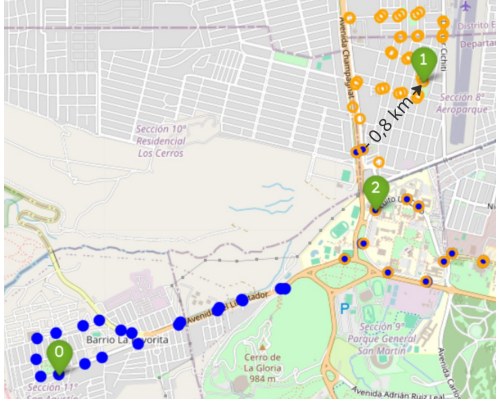
Algoritmo 3.3 Calcular peso de rutas

Entrada: df ▷ DataFrame de rutas, contiene las transacciones de una ruta

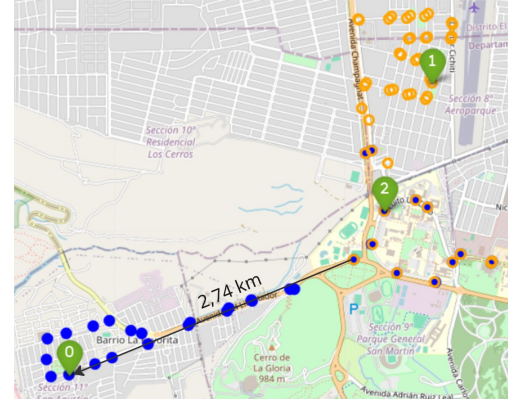
Entrada: viaje_circular ▷ Valor booleano que indica si la ultima transaccion del DataFrame corresponde a la ultima del dia

```
1: ruta ← copiaDe(df)
2: peso ← 0
3: primera_estacion ← ruta[0]
4: estacion_previa ← primera_estacion
5: Para todo elemento ∈ ruta[1,n] hacer
6:     peso ← peso + estacionMasCercanaAUnViaje(elemento.ESTACION_ID, estacion_previa.LINEA)[0]['distancia']
7:     estacion_previa ← element
8: Fin Para
9: Si viaje_circular Entonces
10:     peso ← peso + estacionMasCercanaAUnViaje( primera_estacion.ESTACION_ID, estacion_previa.LINEA)[0]['distancia']
11: Fin Si
12: return rutas
```

Las figuras 3.9a y 3.9b ilustran visualmente cómo el algoritmo escoge entre rutas alternativas, optando siempre por aquella con la menor distancia total caminada. En las figuras, se muestran tres paradas y los posibles viajes entre ellas, sabiendo que la parada final es la número 2, se debe determinar cuál es la parada que comenzó el viaje. Analizando el funcionamiento del algoritmo, se puede ver que en el primer caso, partiendo desde la parada 0, tenemos un peso o distancia recorrida total de 800 metros, mientras que en el segundo caso, partiendo desde la parada 1, el usuario debería haber recorrido 2,74 kilómetros. Debido a esto, el algoritmo toma el caso 1 como la ruta correcta.



(a) Caso 1. Recorrido de parada 0 a parada 1 y de parada 1 a parada 2.



(b) Caso 2. Recorrido de parada 1 a parada 0 y de parada 0 a parada 2.

Figura 3.9: Rutas posibles captadas por el algoritmo de selección

En un siguiente paso, se implementa el algoritmo completo de *trip chaining* (Algoritmo 3.4) que clasifica los viajes como de inicio, intermedio o final. Finalmente, se eliminan aquellas transacciones que no tienen un destino claramente definido. En el Apéndice A.1 se puede consultar el código en Python.

Algoritmo 3.4 Algoritmo de *trip chaining*

Entrada: path ▷ Grafo de entrada, contiene las transacciones ordenadas de una tarjeta

- 1: tiempo_previo ← Nulo
- 2: tme ← 2 ▷ Tiempo maximo de espera
- 3: md ← 0.8 ▷ Maxima distancia
- 4: **Para todo** fila ∈ ruta **hacer**
- 5: **Si** tiempo_previo ≠ Nulo & fila[“hora”] - tiempo_previo ≤ tme & fila[“distancia”] ≤ md **Entonces**
- 6: fila[“cadena”] ← 1 ▷ El valor 1 representa un tramo del viaje
- 7: **Si no**
- 8: fila[“cadena”] ← 0 ▷ El valor 0 representa el inicio del viaje
- 9: **Fin Si**
- 10: tiempo_previo ← fila[“hora”]
- 11: **Fin Para**

La Figura 3.10 muestra un ejemplo de cómo se ve el conjunto de datos posterior a la aplicación del algoritmo de *trip chaining*, se crea una columna “chain” que contiene valores booleanos, 0 para determinar un inicio de viaje y 1 para un viaje intermedio o transbordo. Por términos de simplicidad y un mejor análisis posterior, el fin de viaje está implícitamente definido por un valor 1 cuya siguiente transacción es 0 o nula (último viaje del grupo).

NROTARJETA	RAMAL	ORIGEN	HORA	FECHA_AJUSTADA	DESTINO	chain
IF6962C7E181B83AE023	201	M9399	9	2023-04-19	M15728	0
IF6962C7E181B83AE023	201	M15728	10	2023-04-19	M4217	1
IF6962C7E181B83AE023	201	M4217	10	2023-04-19	M15728	1
IF6962C7E181B83AE023	234	M2181	11	2023-04-19	M15728	1
IF6962C7E181B83AE023	201	M15728	11	2023-04-19	M4217	1
IF6962C7E181B83AE023	200	M4217	11	2023-04-19	M2212	1
IF6962C7E181B83AE023	204	M9996	11	2023-04-19	M2181	1
IF6962C7E181B83AE023	201	M15728	12	2023-04-19	M1018	1
IF6962C7E181B83AE023	201	M1018	15	2023-04-19	M2212	0
IF6962C7E181B83AE023	600	M9994	15	2023-04-19	M1035	1
IF6962C7E181B83AE023	201	M1028	17	2023-04-19	M9399	1

Figura 3.10: Captura de las transacciones de la tarjeta que más viajes tenía del conjunto de datos, se observan dos rectángulos rojos que marcan los dos viajes que tuvo la tarjeta en el mismo día.

3.4. Visualización

La etapa de visualización se puede dividir en dos subsecciones, la de visualizaciones gráficas y tablas estáticas, y la subsección de mapas dinámicos e interactivos.

3.4.1. Visualizaciones gráficas y tablas

Para las visualizaciones de esta sección se han utilizado los datos procesados de la Sección 3.3.1 y la librería *plotly*⁶.

Las visualizaciones que se utilizaron para datos de movilidad se listan a continuación.

- **Mapas de calor para las matrices origen-destino:** Utilizados para mostrar de manera gráfica y tabular las transacciones de ambos conjuntos de datos, en este trabajo se encuentran:
 - Orígenes y destinos entre departamentos.
 - Origen y destino entre departamentos Godoy Cruz, Maipú, Capital, Guaymallén.
 - Origen y destino entre departamentos unificando Godoy Cruz, Maipú, Capital, Guaymallén.
 - Origen y destino entre departamentos Gran Mendoza - por género. Este ítem tiene la posibilidad de filtrar por género.

⁶Página oficial de *Plotly*

- **Histogramas:** Empleados para representar las distintas categorías y visualizar qué cantidad de transacciones se les atribuyen a estas. Para el conjunto de datos de mendoTRAN las categorías utilizadas son “género” y “tipo de descuento”, mientras que para el conjunto de datos de biciTRAN las categorías son “género” y “tipo de usuario”:
 - Histograma: Viaje único y transbordo: Esta comparativa señala cuantas transacciones se consideran viajes de un tramo y cuántas han sido transbordos, es decir, le suceden a un viaje dentro de las horas y distancias máximas establecidas. Solo utilizado para mendoTRAN.
 - Histograma: Distribución de viajes por hora y género: Este histograma tiene a las horas como eje de abscisas y la cantidad de transacciones en el eje de ordenadas, a su vez, cada barra del histograma se divide por las diferentes categorías de géneros. Empleado en ambos conjuntos de datos.
 - Histograma: Distribución de viajes por hora y descuento: Este histograma tiene a las horas como eje de abscisas y la cantidad de transacciones en el eje de ordenadas, a su vez, cada barra del histograma se divide por las diferentes categorías de descuentos, no se incluyen las transacciones sin descuentos. Solo utilizado para mendoTRAN.
 - Histograma: Distribución de viajes con y sin descuento por hora: Esta visualización es similar a la anterior, solo que muestra el porcentaje de transacciones con descuento, englobando todas las categorías, y sin descuento. Solo utilizado para mendoTRAN.
 - Histograma: Distribución de viajes por hora y tipo de usuario: Esta visualización se desarrolló siguiendo los mismos patrones que las anteriores, donde las horas se encuentran en el eje de abscisas y la cantidad de viajes en el eje de ordenadas, dividiendo así, cuantos correspondían a cada tipo de usuario. Solo utilizado para biciTRAN.
- **Gráficos de torta:** Para visualizar la proporcionalidad de las categorías en las columnas de los conjuntos de datos, se han formulado los siguientes gráficos:
 - Gráfico de torta: Distribución de viajes por género. Empleado en ambos conjuntos de datos.
 - Gráfico de torta: Distribución de viajes por descuento. Solo utilizado para mendoTRAN.

- **Gráfico de torta:** Distribución de viajes por tipo de usuario. Solo utilizado para biciTRAN.
- **Diagramas de Sankey:** Para visualizar los patrones de movilidad entre departamentos en los conjuntos de datos, se utilizan diagramas de Sankey, aquí solo una visualización es elegible. Esta muestra nodos representando a los departamentos y enlaces entre ellos; para ambos elementos, el grosor o tamaño indica la magnitud de las incidencias.

Todas las visualizaciones del conjunto de datos de mendoTRAN pueden ser desglosadas o filtradas por día o mostrar todos los días. Mientras que las visualizaciones del conjunto de datos de biciTRAN pueden ser filtradas por mes.

3.4.2. Visualizaciones en mapas interactivos

Los mapas interactivos son muy frecuentes en los sistemas de visualización de datos para la toma de decisiones. A su vez, son muy empleados para la representación de patrones de movilidad. En este trabajo se han confeccionado los siguientes mapas para el conjunto de datos de mendoTRAN:

- **paradas de llegada y contador de descensos:** Dada una línea, imprime todas las paradas que esta recorre y al hacer clic en alguna de estas, el mapa muestra la cantidad de descensos que ocurrieron en la parada seleccionada.
- **Cantidad de salidas por hora:** Al elegir una parada, el mapa genera un histograma que ilustra la cantidad de salidas desde esa parada para cada hora.
- **Descensos por parada:** Dada una parada elegida como origen, muestra la cantidad de viajes que salieron de dicha parada y las paradas de destino y cantidad de descensos que tuvo en cada una de ellas.
- **Pares de paradas:** Muestra los pares de paradas de origen y destino con mayor cantidad de viajes.
- **Destinos por departamento:** Al seleccionar un departamento, muestra un mapa coroplético con todos los departamentos que tuvieron como destino los viajes iniciados en el departamento seleccionado.
- **Viajes por hora con ascensos y descensos:** Muestra las salidas, llegadas y total de transacciones (sumatoria de salidas y llegadas) realizados por hora de la parada seleccionada.

- **Mapa de calor de viajes:** En este mapa se tiene la opción de elegir entre visualizar las salidas o los orígenes. Dependiendo de la elección realizada, se presentará un mapa de calor interactivo que desplegará progresivamente la cantidad de transacciones realizadas cada hora.

Todas las visualizaciones de esta sección se construyeron utilizando el total de las transacciones, eliminando la posibilidad de desglosar por día.

En cuanto al conjunto de datos de biciTRAN, se han desarrollado los mapas detallados a continuación:

- **Pares de estaciones con más duración de viaje:** Dado un mes, muestra los pares de estaciones de origen y destino de los viajes con mayor duración.
- **Capacidad por estación:** Cada estación se representa en el mapa. Al elegir una estación, se revela la capacidad máxima de bicicletas que esta puede albergar.
- **Mapa de calor de viajes:** En este mapa interactivo, es necesario elegir entre las opciones de mostrar puntos de salida o de origen, así como el mes deseado para la visualización. Basado en la selección realizada, se presentará un mapa de calor que despliega de manera continua la cantidad de transacciones que tienen lugar cada hora.

3.5. Tecnologías y recursos utilizados

En esta sección se describen las tecnologías utilizadas para la implementación de los algoritmos y visualizaciones de esta tesina final de grado.

Como lenguaje de programación predefinido se eligió Python⁷, por su facilidad de manejo, rapidez y versatilidad para la manipulación de grandes volúmenes de datos.

Para este proyecto se han utilizado diversas bibliotecas que han ayudado en el desarrollo, entre ellas se pueden destacar:

- **Jupyter Notebook**⁸: los cuadernos de trabajo *Jupyter Notebook* permiten ejecutar código y visualizar la salida en un mismo entorno. Estos fueron utilizados para la exploración de datos y el desarrollo de la mayoría de las funciones en este trabajo.

⁷Página oficial de Python

⁸Página oficial de Jupyter Notebook

- **Pandas**⁹: es una herramienta de análisis y manipulación de datos de código abierto rápida, potente, flexible y fácil de usar, construida sobre el lenguaje de programación Python. Fue utilizada para la manipulación de los conjuntos de datos.
- **Numpy**¹⁰: lleva la potencia computacional de lenguajes como C y Fortran a Python, un lenguaje mucho más fácil de aprender y usar. Esta potencia conlleva simplicidad: una solución en NumPy suele ser clara y elegante. Esta librería fue utilizada mayormente como dependencia de otras librerías y además como base de algunas estructuras de datos empleadas en el proyecto, ya que permite manejar los datos numéricos con mayor eficiencia.
- **Folium**¹¹: facilita la visualización de datos manipulados en Python en un mapa interactivo. Permite tanto la vinculación de datos a un mapa para visualizaciones coropletas como el paso de visualizaciones vectoriales/raster/HTML enriquecidas como marcadores en el mapa. La biblioteca incorpora una serie de conjuntos de mosaicos de OpenStreetMap, Mapbox, etc., y admite conjuntos de mosaicos personalizados. Folium admite superposiciones de imagen, vídeo, GeoJSON y TopoJSON y tiene varias capas vectoriales integradas.
- **Plotly**¹²: La biblioteca de gráficos Python de Plotly crea gráficos interactivos con calidad de publicación. Se utilizó esta librería para la generación de la mayoría de las visualizaciones (todas aquellas que no tenían mapas incluidos).
- **Dash**: Esta librería pertenece a Plotly y es específica para el diseño y programación *frontend* usando lenguaje de Python. Ofrece las mismas visualizaciones de Plotly más un agregado de otras librerías como **dash-leaflet** que permite la integración de mapas de folium en la web.

También se utilizaron sitios externos para el desarrollo de algunas partes de este trabajo, como:

- **Render**¹³: Una plataforma para el despliegue de aplicaciones de manera simple y que incluye integración continua, es decir, permite actualizar la web automáticamente cuando se actualiza el repositorio.

⁹Página oficial de Pandas

¹⁰Página oficial de Numpy

¹¹Página oficial de Folium

¹²Página oficial de Plotly

¹³Render.com

- **AWS**¹⁴: Los servicios web de Amazon (AWS, *Amazon Web Services*) fueron utilizados por el mismo propósito que *Render*, para el despliegue de la aplicación, ya que consta de servidores de mayores prestaciones y permite el pago de los mismos bajo demanda.
- **DuckDNS**¹⁵: Es un servidor de nombres de dominios (*Domain Name System* (DNS)), fue útil para este trabajo debido a que las instancias de AWS no poseían IP estáticas y para accederlas se debía cambiar la IP cada vez que se reiniciaba la instancia.
- **Repo-View**¹⁶: Repo-View es una página web estática que utiliza *API* de GitHub¹⁷ para compartir visualizaciones de sitios.

El servidor web fue implementado utilizando la librería **Dash** de Plotly. Esta permite programar y servir fácilmente una página web utilizando como base un servidor Flask¹⁸. Si bien su programación fue puramente en Python, este *framework* permite la inyección de código JavaScript mediante la importación de una extensión.

En cuanto a los recursos utilizados, todo el procesamiento se ejecutó en una laptop Alienware X15-R2 con 32 GB de memoria, 3 TB de disco duro, procesador 12th Gen Intel® Core™ i7-12700H x 20 núcleos y tarjeta gráfica Nvidia RTX 3060. Para el alojamiento de la aplicación en un principio se optó por un servidor gratuito obtenido en render. El mismo cuenta con 512 MB de memoria RAM y 10% de una CPU compartida, de la cual se desconocen sus características. Luego, por cuestiones de rendimiento, se movió el servidor a una instancia *t2 large EC2* (*Elastic Compute Cloud*, por sus siglas en inglés) de Amazon con 8 GB de RAM y 2 vCPU.

Para evaluar el funcionamiento de los algoritmos seleccionados, es necesario contar con diferentes conjuntos de datos:

- a) **Sistema SUBE, Mendoza**: fue utilizado para el algoritmo TC y la generación de los gráficos de visualización de movilidad.
- b) **Sistema biciTRAN, Mendoza**: fue utilizado para la generación de los gráficos de visualización de movilidad.

¹⁴Página oficial de AWS

¹⁵Página oficial de DuckDNS

¹⁶Página del repositorio de Repo-View

¹⁷Web oficial de GitHub

¹⁸Página oficial de Flask

El código fuente de la aplicación y los algoritmos desarrollados para esta tesina se pueden encontrar en un repositorio privado de GitHub, el cual es accesible mediante el link de repoview: <https://ibrahemesam.github.io/repo-view>.

RESULTADOS Y DISCUSIONES

En este capítulo se presentan y analizan los principales resultados obtenidos a partir de la aplicación del algoritmo de inferencia de destinos (*trip chaining*), la construcción de matrices origen-destino y el desarrollo de herramientas de visualización aplicadas a los datos del sistema de transporte público de pasajeros urbano de Mendoza y del sistema de bicicletas compartidas biciTRAN.

El objetivo de este apartado es exponer de forma clara y organizada cómo se plasmaron en resultados los procedimientos metodológicos explicados en el capítulo anterior. Para ello, se muestran las distintas salidas intermedias y finales del procesamiento de datos, se describen las visualizaciones gráficas y geoespaciales implementadas, y se discuten los patrones de movilidad identificados, resaltando su relevancia para la planificación y gestión del sistema de transporte.

Además, se incluyen interpretaciones de los hallazgos que permiten poner en perspectiva la utilidad práctica de la información generada, evidenciando de qué forma estos insumos pueden servir como apoyo para la toma de decisiones y la implementación de políticas de movilidad urbana más eficientes y sustentables.

4.1. Construcción de la matriz OD

En esta sección se muestran los resultados de la ejecución integral de los algoritmos asociados a la construcción de las matrices origen-destino desglosados por las etapas recorridas:

- **Preprocesamiento:** El proceso de limpieza de datos resultó en una drástica

reducción de registros, pasando de un total inicial de 3.842.663 a 395.755 entradas, lo cual representa una disminución del 90 %. Los elementos identificados para su depuración incluyeron 2.420.079 identificadores de tarjeta que resultaron ser nulos, 214.939 registros que contenían el identificador “ACOM” (véase la Figura 4.1), y 443.264 filas en las que los campos de latitud y longitud contenían valores nulos (referirse a la Figura 4.2), además, 27 registros estaban duplicados. Tras la eliminación de estos datos incorrectos, se procedió a filtrar la información no útil para el análisis. En este paso, se identificaron 406.930 transacciones cuyo número de tarjeta fue empleado solo una vez, además, se detectaron 553 transacciones que constituían las únicas operadas en un día concreto una vez que se ajustaron las fechas mediante la aplicación de una ventana de tiempo virtual. Finalmente, el conjunto de datos quedó con 402.667 transacciones válidas; no obstante, cabe mencionar que en 8.063 de estas no fue posible determinar un destino, dado que no cumplían con las premisas necesarias para calificarlas como viajes. La etapa de inferencia de paradas requirió de múltiples consultas a la *API* de mendoTRAN, estas tomaron alrededor de 18 horas debido a la cantidad de paradas a consultar.

- **Construcción de la matriz:** El proceso de ordenar las transacciones por orden cronológico y de crear listas enlazadas para rutas internas (Algoritmo 3.1) llevó un total de 3 horas. En contraste, el método que determina los destinos utilizando las normas del algoritmo (Algoritmo 3.4) demoró, en promedio, 40 segundos por configuración. Este último procedimiento fue replicado para cada posible configuración. Los viajes que fueron inferidos para cada configuración se presentan en la Tabla 4.1.

NROTARJETA	count
ACOM	214938
962C7E181B83AE023	12
C8FF682ACCA60BF19	11
51880CE3BC8B1654F	11
DAFEB4217DBD76B74F	11
...	...
E4291D1C6CBEED8	1
EEEECE31081B4CD4	1
923CD26D66946B5	1
446B23074380F8B	1
A417FBD2191B859	1

Name: count, Length: 788267, dtype: int64

Figura 4.1: Vista de los valores más comunes para el campo “NROTARJETA”.

La Tabla 4.1 sintetiza 15 configuraciones del algoritmo de TC que combinan distancias máximas de transferencia entre 600 y 1400 m y ventanas temporales entre 1 y 3 horas. Para cada configuración se contabilizan: viajes únicos, viajes con transbordo y

NRO.TARJETA	MODO	JURISDICCION	LINEA	RAMAL	INTERNO	TIPO.TARIFA	CODIGO.CONTRATO	MOTIVO.ATS	SEXO	FECHA.HORA	ETAPA.RED.SUBE	LONGITUD	LATITUD
NaN	COL	PROVINCIAL	LINEA 700	700	5E329E038A4B2EB74CB9	Tarifa_desc_local	659	NaN	M	20230421 21	NaN	0.0	0.0
NaN	COL	PROVINCIAL	LINEA 700	700	5E329E038A4B2EB74CB9	Tarifa_desc_local	582	NaN	M	20230421 21	1.0	0.0	0.0

Figura 4.2: Vista previa de filas con valores nulos en distintos campos.

transbordos inferidos. El objetivo es identificar un punto de operación que maximice la detección plausible de transbordos sin sobredimensionar el emparejamiento de transacciones.

■ **Evaluación del efecto de la ventana temporal (tiempo de transferencia).**

A distancia fija, el paso de 1 a 2 horas concentra la mayor ganancia y de 2 a 3 horas las mejoras se atenúan. Por ejemplo, a 1000 m (TC3→TC8→TC13):

- los viajes únicos bajan **9,8%** de 309.885 a 279.523 y luego **6,9%** a 260.241;
- los viajes con transbordo suben **31,6%** de 37.142 a 48.863 y luego **13,8%** a 55.613;
- los transbordos inferidos suben **40,2%** de 47.655 a 66.831 y luego **17,8%** a 78.748.

El patrón se reproduce para 600, 800, 1200 y 1400 m con magnitudes similares: la ventana de **2 h** captura la mayor parte de los transbordos adicionales, mientras que **3 h** aporta incrementos menores acompañados por una caída adicional de viajes únicos.

■ **Evaluación del efecto de la distancia máxima (espacio de transferencia).**

A tiempo fijo de 2 horas (TC6-TC10) y al ampliar la distancia en pasos de 200 m se observan ganancias decrecientes:

- **600→800 m:** +4,5% en viajes con transbordo y +6,2% en transbordos inferidos; viajes únicos -1,9%.
- **800→1000 m:** +6,8% y +9,3%; -2,9%.
- **1000→1200 m:** variación casi nula en viajes con transbordo (-0,4%) y +3,7% en inferidos; -1,5% en únicos.
- **1200→1400 m:** +2,6% en viajes con transbordo y -0,7% en inferidos; variación mínima en únicos (+0,2%).

En síntesis, subir de **800 a 1000 m** todavía aporta señal (sobre todo en transbordos inferidos) con una merma acotada de viajes únicos. Más allá de **1000-1200 m** las mejoras son marginales y pueden introducir emparejamientos espurios si existen imprecisiones de georreferenciación.

Considerando los recorridos y espaciamientos de líneas publicados por mendoTRAN¹, el equilibrio entre detección de transbordos y cautela ante errores de posicionamiento, y el patrón de rendimientos decrecientes observado, se adopta la configuración **TC8 (1000 m, 2 h)** como punto de operación para el resto del trabajo. Esta combinación captura la mayor parte de los transbordos plausibles con una penalización moderada sobre viajes únicos y evita el sobreemparejamiento que pueden introducir distancias o ventanas más laxas. Empíricamente, el salto **1→2 h** añade aproximadamente **+32%** en viajes con transbordo y **+40%** en transbordos inferidos, mientras que **2→3 h** agrega sólo **+14-18%** adicionales con una nueva caída de viajes únicos, por lo que **2 h** es el mejor compromiso.

ID	Distancia (m)	Tiempo(h)	Cantidad de viajes únicos	Cantidad de viajes con trans- bordo	Transbordos inferidos
TC1	600	1	318321	33844	42437
TC2	800		313728	35626	45248
TC3	1000		309805	37142	47655
TC4	1200		306537	38324	49741
TC5	1400		303239	39525	51838
TC6	600	2	293299	43752	57551
TC7	800		287849	45741	61012
TC8	1000		283249	47411	63942
TC9	1200		279523	48683	66396
TC10	1400		275847	49914	68841
TC11	600	3	271330	51892	71380
TC12	800		265273	53941	75388
TC13	1000		260241	55613	78748
TC14	1200		256147	56897	81558
TC15	1400		252200	58091	84311

Tabla 4.1: Resultados por configuración del algoritmo de *trip chaining*. Se reportan viajes únicos, viajes con transbordo y transbordos inferidos para combinaciones de distancia máxima de transferencia (600-1400 m) y ventana temporal (1-3 h). Los conteos corresponden a la ejecución después del refinamiento descrito en la Sección 3.3.2.

4.2. Prototipo de sistema web

En esta sección se presenta el prototipo de la aplicación web que se ha desarrollado para esta tesina empleando las herramientas de software detalladas en la Sección 3.5.

¹recorridos de MendoTRAN

A continuación, se ofrece una descripción extensa de la navegación del sistema web y un análisis detallado de las diferentes vistas que pueden encontrarse en él.

Al acceder al prototipo web, el usuario se encuentra con la pantalla principal, ilustrada en la Figura 4.3. Esta vista proporciona una breve reseña sobre la aplicación desarrollada específicamente para este proyecto final de grado, además de presentar a los miembros del equipo.



Figura 4.3: Vista principal de la aplicación.

En esta ventana se puede seleccionar el conjunto de datos a utilizar y alternar los colores de la página entre modo oscuro o claro; estos selectores permanecen estáticos en la cabecera, de modo que pueden ser seleccionados y accionados en cualquier vista. Desde este punto hay dos posibles caminos seleccionando alguna de las pestañas disponibles:

- Visualizaciones gráficas y tablas.
- Visualizaciones en mapas interactivos.

4.2.1. Visualizaciones gráficas y tablas

La pestaña de visualizaciones gráficas y tablas redirige a una nueva sección compuesta de 4 pestañas internas, mostrando los diferentes tipos de visualizaciones estáticas (Figura 4.4).

- Matrices Origen Destino.
- Histogramas.
- Gráficos de torta.
- Diagramas de Sankey.

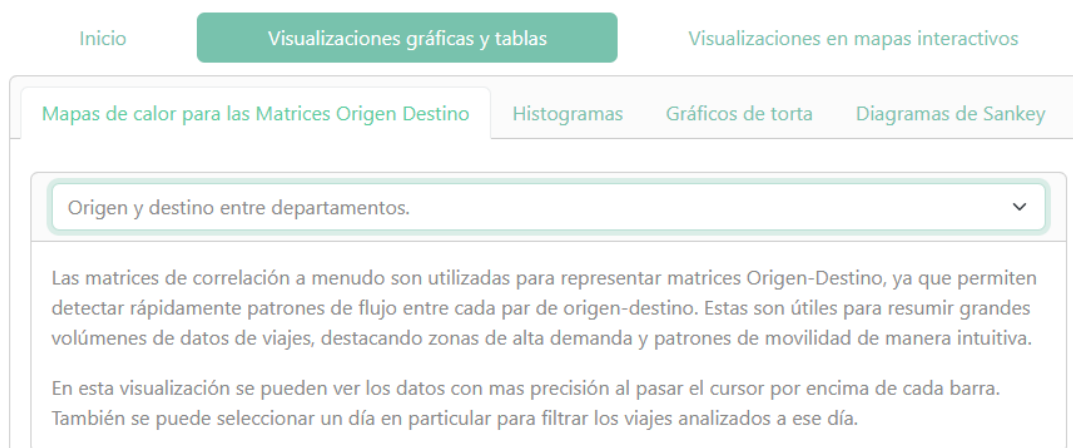


Figura 4.4: Página de visualizaciones gráficas y tablas.

4.2.1.1. Matrices Origen Destino

La pestaña inicial que aparece en la sección de visualizaciones gráficas y tablas está dedicada a los mapas de calor de las matrices origen-destino. Al acceder a ella, una lista de opciones se puede desplegar, permitiendo abrir distintos tipos de visualizaciones en función de la selección realizada (Figura 4.5).



Figura 4.5: Opciones del ítem Matrices Origen Destino en visualizaciones gráficas y tablas (mendoTRAN).

Al seleccionar el primer ítem llamado **Origen y destino entre departamentos** se puede ver una matriz de incidencia que tiene tanto en sus filas como en sus columnas los distintos departamentos analizados, tomando como origen los departamentos ubicados en las filas y como destino aquellos ubicados en las columnas (Figura 4.6). Las celdas presentan porcentajes que indican el volumen de desplazamientos entre distintos pares origen-destino. Estos porcentajes se visualizan mediante el color de las celdas: un amarillo suave señala una baja movilidad entre los departamentos, mientras que tonalidades más intensas de rojo reflejan un mayor flujo de viajes entre ellos, con valores que aumentan a medida que el color se intensifica. A su vez, los valores utilizados para calcular dicha matriz pueden ser filtrados utilizando el menú desplegable ubicado arriba de la tabla (Figura 4.7), permitiendo elegir un día en particular u observar todos los viajes juntos.

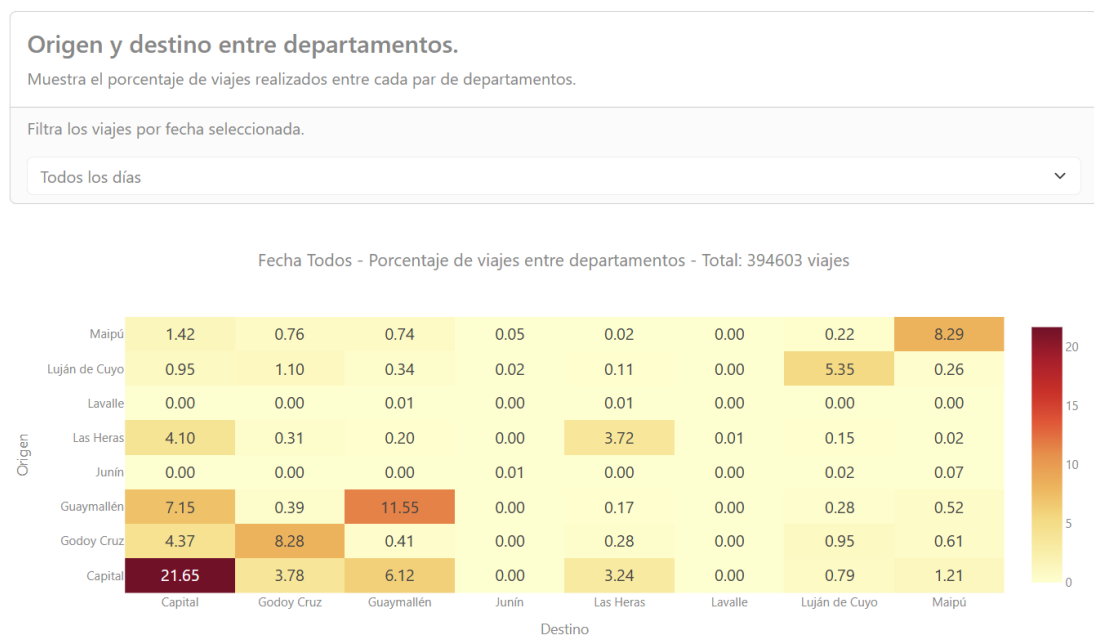


Figura 4.6: Matriz Origen Destino entre departamentos (mendoTRAN).

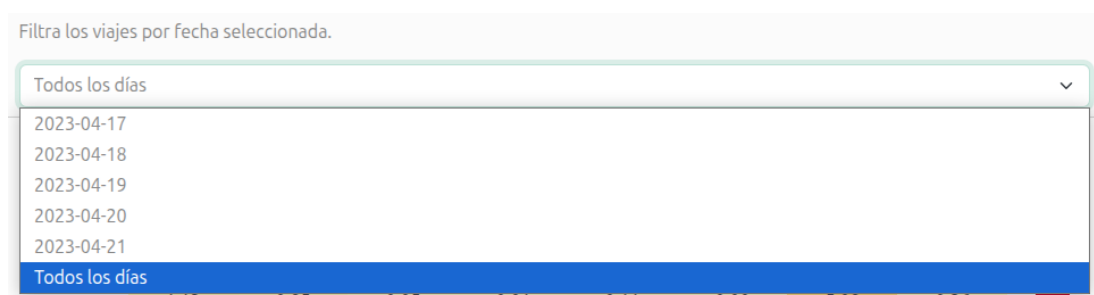


Figura 4.7: Opciones del filtro de la tabla matriz origen-destino entre departamentos (mendoTRAN).

El segundo ítem del menú desplegable corresponde a **Origen y destino entre departamentos Godoy Cruz, Maipú, Capital, Guaymallén**, es una representación similar a la anterior, pero con la salvedad de que el análisis realizado solo incluye los departamentos mencionados, desestimando visualizar el resto (Figura 4.8). Estos departamentos fueron escogidos porque incluyen la mayoría de los viajes, lo que facilita concentrarse en el análisis de los desplazamientos que ocurren entre ellos. Las funcionalidades y representaciones de este ítem son idénticas al ítem anterior.

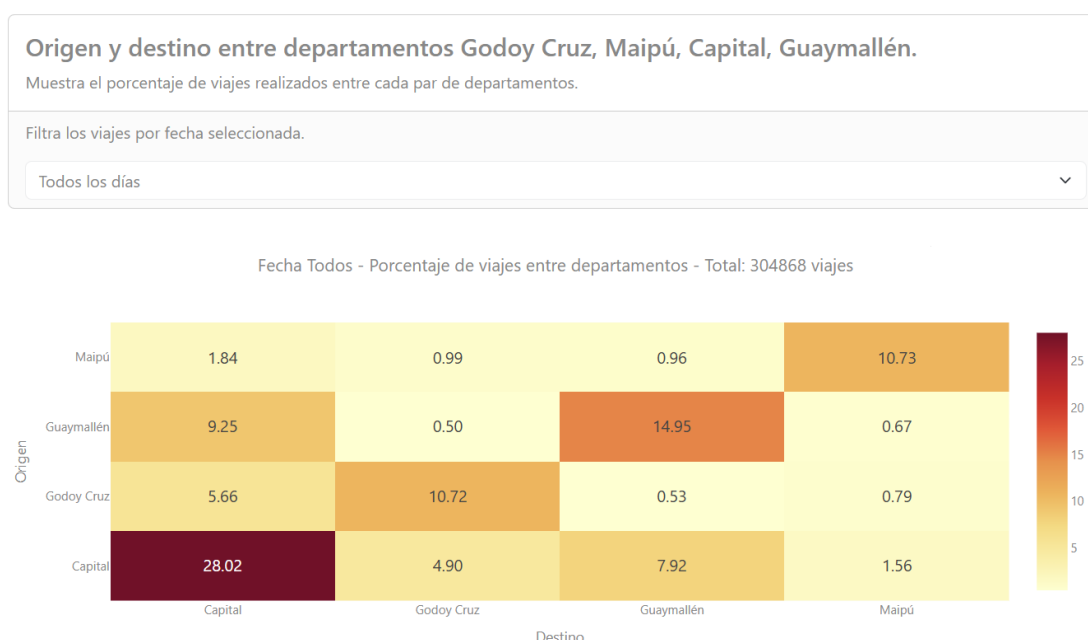


Figura 4.8: Origen y destino entre departamentos Godoy Cruz, Maipú, Capital, Guaymallén (mendoTRAN).

El tercer valor en el menú desplegable es **Origen y destino entre departamentos unificando Godoy Cruz, Maipú, Capital, Guaymallén (GC M C G)**, este es similar al primero, solo que unifica los departamentos del Gran Mendoza y los representa con las siglas (GC M C G) (Figura 4.9). De esta forma, permite visualizar cómo interactúan los departamentos más alejados con los más cercanos a la ciudad de Mendoza.

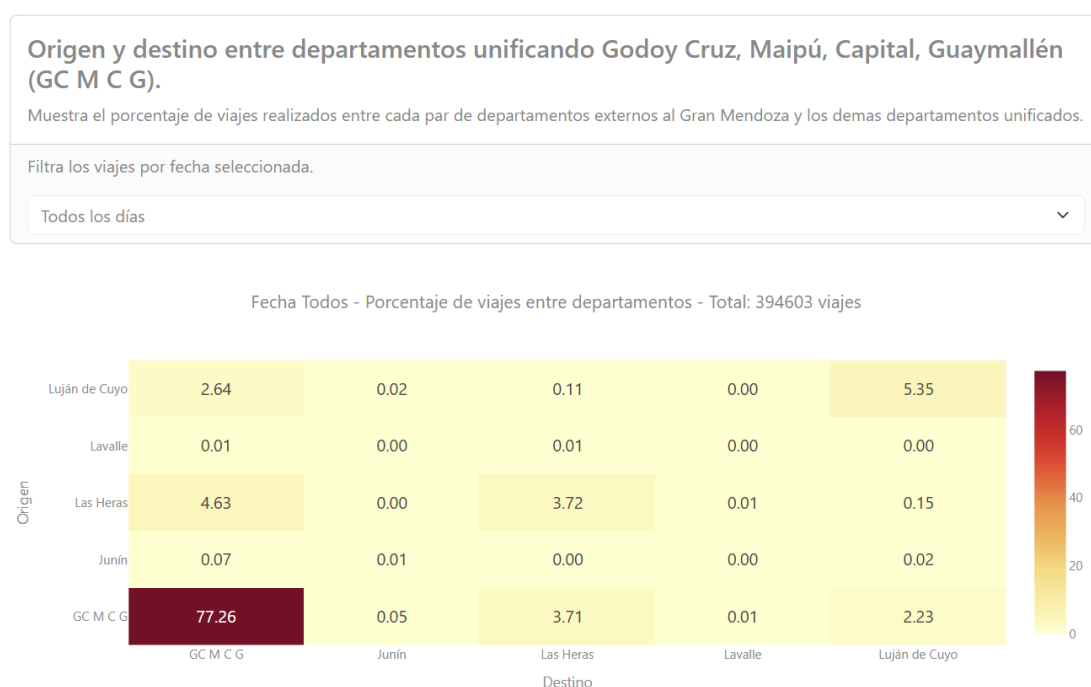


Figura 4.9: Origen y destino entre departamentos unificando Godoy Cruz, Maipú, Capital, Guaymallén (GC M C G) (mendoTRAN).

El último elemento del menú es **Origen y destino entre departamentos - por género** y representa las interacciones entre departamentos, permitiendo filtrar por género, además del filtrado por fecha. La Figura 4.11 ilustra la posibilidad de seleccionar fechas y géneros específicos, además de proporcionar la opción de elegir *Todos* en cualquiera de las categorías disponibles. Por otro lado, la Figura 4.10 presenta el gráfico que acompaña a esta visualización. Esta visualización resulta útil cuando se busca analizar patrones de movilidad diferenciados por género.

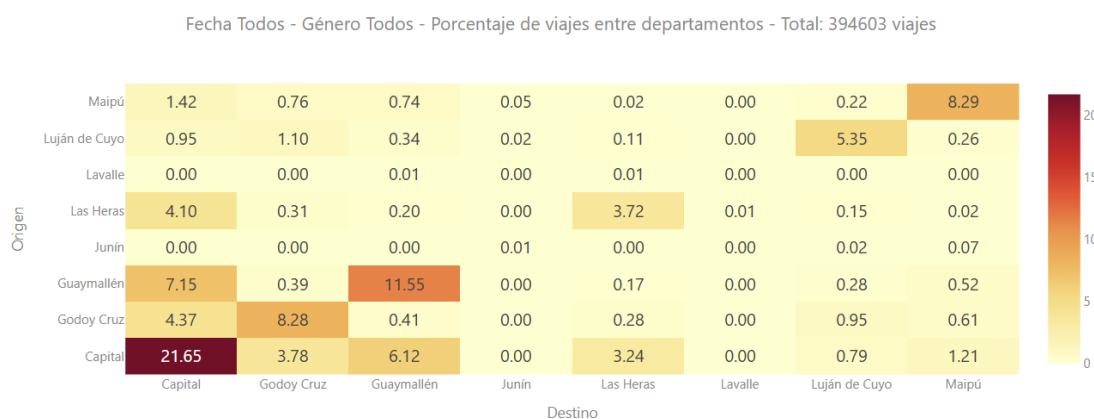


Figura 4.10: Origen y destino entre departamentos - por género (mendoTRAN).

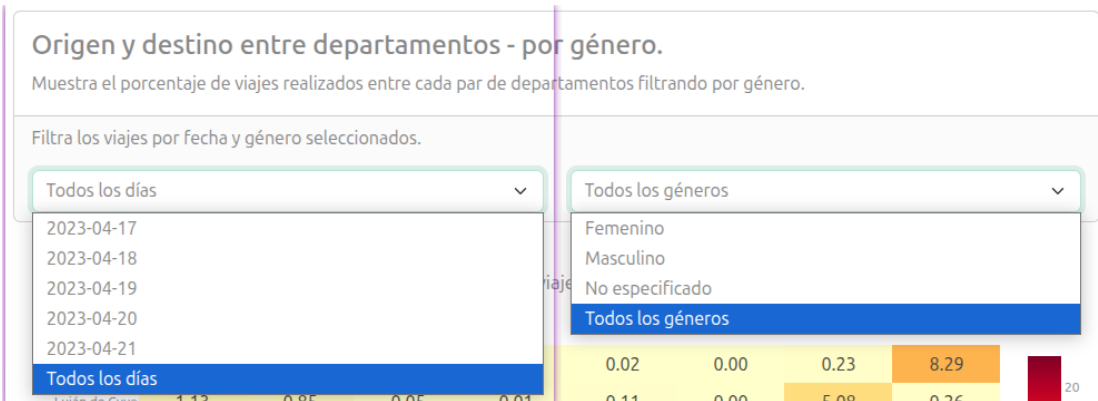


Figura 4.11: Opciones del filtro de la tabla matriz origen-destino entre departamentos - por género (mendoTRAN).

El conjunto de datos de biciTRAN cuenta con las mismas visualizaciones en esta pestaña, por lo que se omite su descripción en esta sección.

4.2.1.2. Histogramas

En la solapa Histogramas del prototipo web, se presentan cuatro alternativas visualizables (ver Figura 4.12). Estos histogramas ofrecen una representación gráfica de las distribuciones estudiadas en este trabajo de fin de carrera; es importante destacar que cualquier visualización en esta parte del informe proporciona la opción de filtrado por día específico de la semana o la selección de *Todos* para observar la distribución completa.

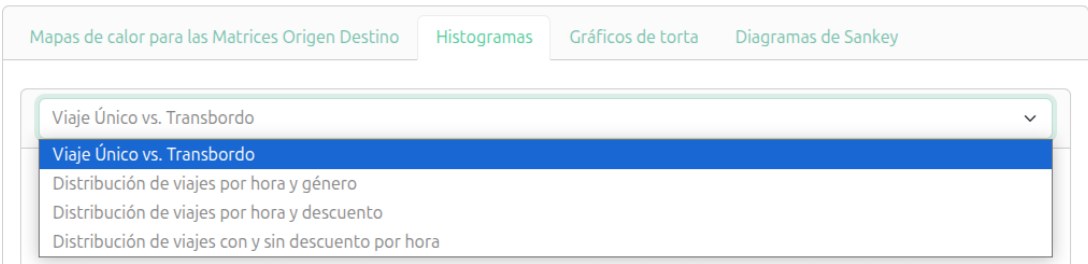


Figura 4.12: Opciones del ítem Histogramas en visualizaciones gráficas y tablas.

En esta pestaña, la primera visualización disponible es **Viaje Único vs. Viajes con transbordo** (Figura 4.13). Esta representación gráfica ilustra la comparación entre viajes categorizados como viajes únicos (compuestos por una sola transacción) y aquellos que contienen múltiples segmentos de viaje, dentro de los límites temporales y de distancia definidos por el algoritmo de *trip chaining*. El análisis proporcionado por esta visualización ayuda a los responsables de tomar decisiones a diseñar estrategias que

promuevan los transbordos y a evaluar su impacto. Cabe señalar que esta visualización se aplica exclusivamente al conjunto de datos mendoTRAN.

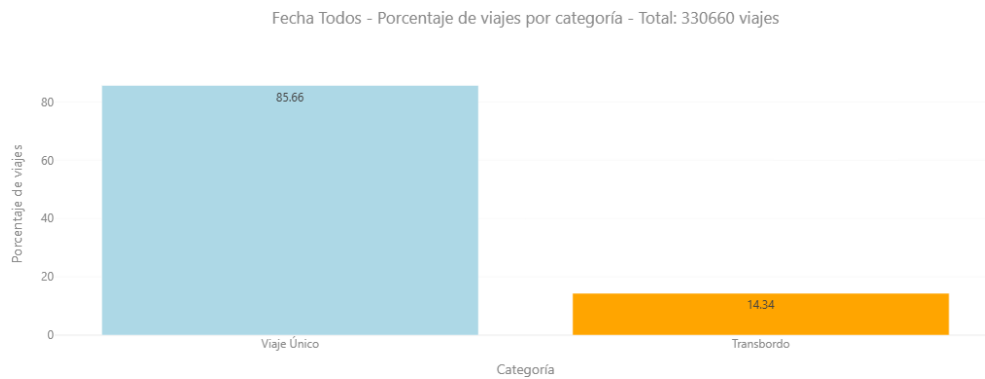


Figura 4.13: Comparativa de Viaje Único vs. Transbordo (mendoTRAN).

Del total de 331.515 transacciones analizadas, el 85,64% se clasificó como viaje único y el 14,36% como viajes con transbordo. Esto indica que cerca de uno de cada siete viajes requiere transferencia bajo los umbrales establecidos, información clave para focalizar mejoras en nodos de intercambio y frecuencias.

La segunda opción bajo la pestaña *Histogramas* es **Distribución de viajes por hora y género** (Figura 4.14). Este apartado ilustra cómo se distribuyen los viajes a lo largo de las horas de operación y divide cada barra para exhibir el conteo de categorías específicas. El análisis de esta representación gráfica revela quiénes son los usuarios del servicio en distintos momentos del día, lo cual resulta útil al considerar políticas más inclusivas y al determinar cuáles son las franjas horarias que podrían beneficiarse de mejoras específicas.

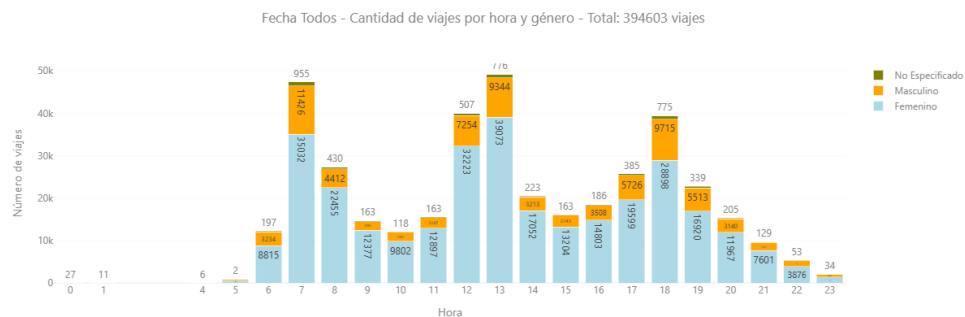


Figura 4.14: Distribución de viajes por hora y género (mendoTRAN).

Desde el punto de vista de los tomadores de decisiones, es posible concluir que el STPPU de Mendoza es utilizado mayoritariamente por mujeres. La distribución

horaria exhibe picos pronunciados a las 7-8 h, 12-13 h (el mayor) y 18 h las cuales coinciden con el ingreso y egreso de alumnos a los establecimientos educativos y el comienzo y fin de la jornada laboral, con valles en la madrugada y después de las 21 h. En todas las franjas predomina el segmento Femenino ($\approx 70 - 80\%$), mientras que Masculino se mantiene en el $20 - 25\%$ y No especificado resulta marginal. Este patrón respalda la priorización de oferta en 06:30-08:30, 12:00-14:00 y 17:30-19:00, así como medidas de gestión específicas en nodos de transferencia durante esas ventanas.

Esta representación gráfica, junto con todas las posteriores que presentarán diversas categorías apiladas en las barras, ofrece dos funciones principales: primero, se puede ocultar una categoría específica de la vista haciendo clic sobre ella (Figura 4.15); en segundo lugar, al realizar un doble clic, es posible enfocarse exclusivamente en la categoría deseada, mostrando solo esta en el gráfico (Figura 4.16).

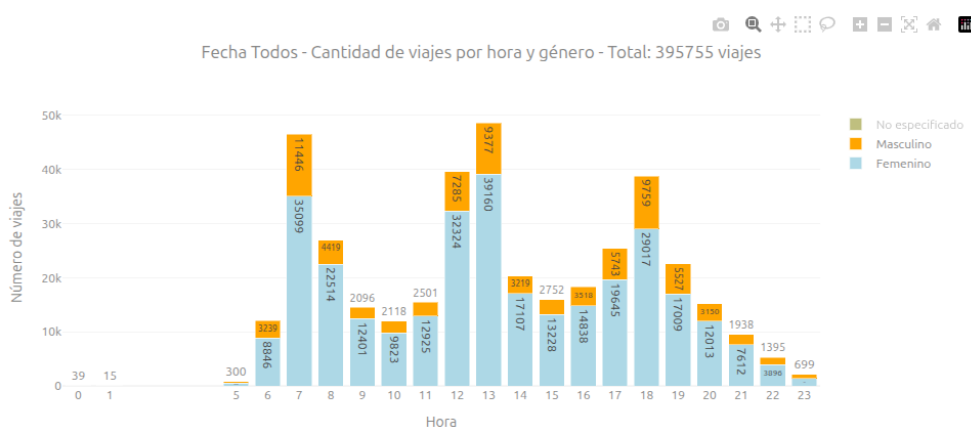


Figura 4.15: Distribución de viajes por hora y género filtrando con un clic (mendo-TRAN).

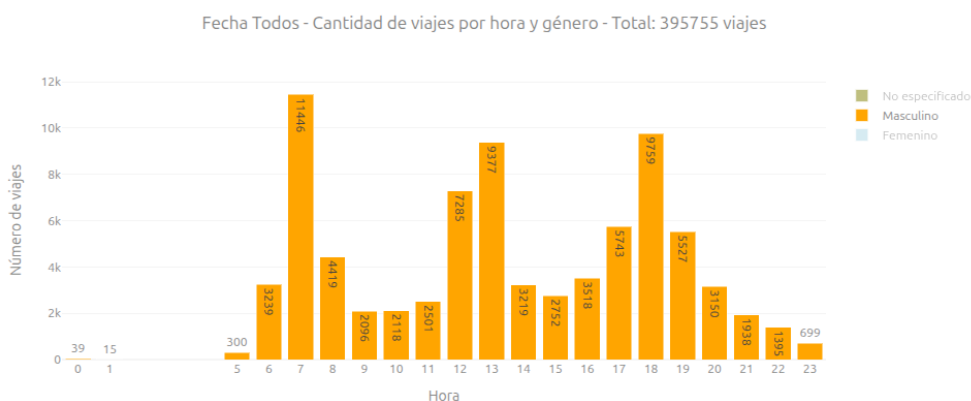


Figura 4.16: Distribución de viajes por hora y género filtrando con doble clic (mendo-TRAN).

La siguiente opción disponible en este segmento es la de **Distribución de viajes por hora y descuento** (Figura 4.17), este histograma permite visualizar la distribución de los viajes a cada hora, agrupando por tipo de descuento. Estos datos no contienen las transacciones sin descuento. Esta visualización solo aplica para el conjunto de datos de mendoTRAN. Una gráfica de este tipo permite analizar los patrones de demanda por tipo de usuario, de manera que se puedan analizar más eficientemente las políticas tarifarias y optimizar la planificación de frecuencias y recursos.

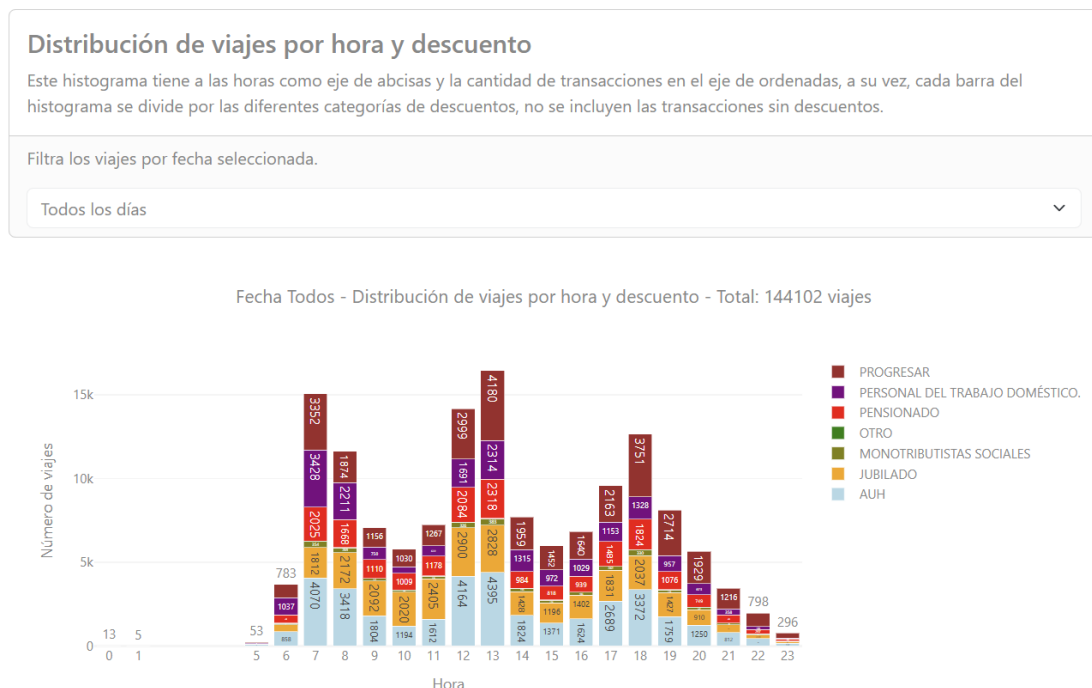


Figura 4.17: Distribución de viajes por hora y tipo de descuento (mendoTRAN).

En la Figura 4.17 se puede observar que las personas que se trasladan entre las 7 y las 8 de la mañana poseen mayoritariamente cuatro tipos de subsidios en el pasaje: Progresar, Personal de Trabajo Doméstico, Asignación Universal por Hijo (AUH) y jubilados. Los tomadores de decisiones pueden entonces dirigir su atención a estos segmentos a fin de mejorar la calidad de servicio para ellos.

Por último, la Figura 4.18 muestra la **Distribución de viajes con y sin descuento por hora**, en la que se puede ver la comparación entre la cantidad de viajes con descuento y la cantidad de viajes sin descuento a cada hora. Esta visualización solo aplica para el conjunto de datos de mendoTRAN. Este tipo de visualización, junto con la anterior mencionada, aporta perspectiva social y económica al análisis, permitiendo ver si se pueden aplicar tarifas diferenciadas por horario como estrategia para el fomento de la demanda.

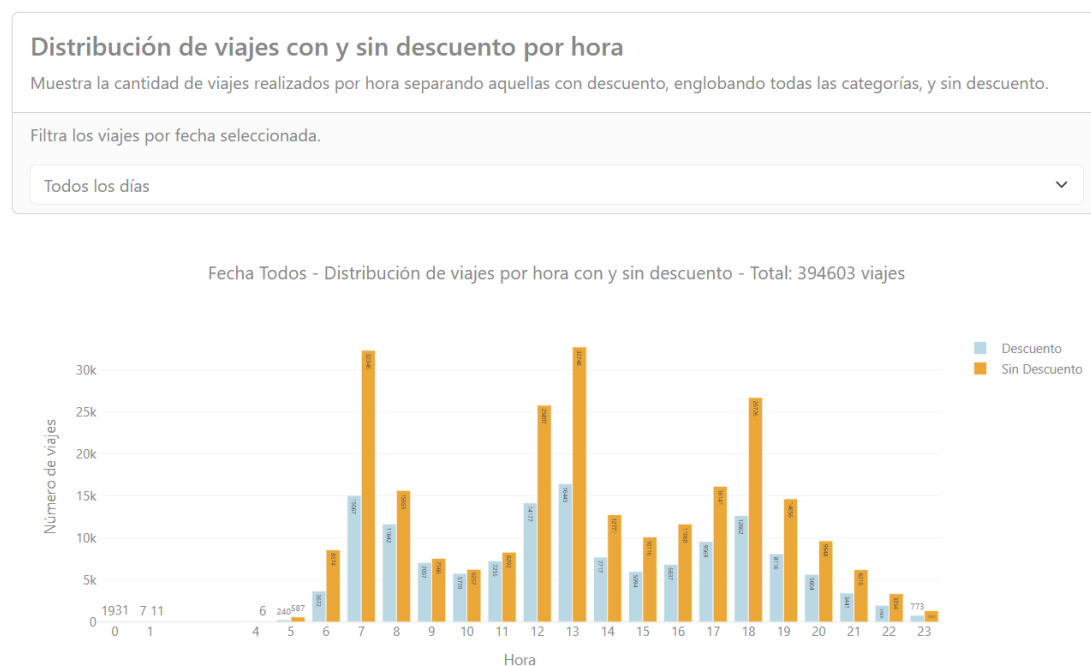


Figura 4.18: Distribución de viajes con y sin descuento por hora (mendoTRAN).

Considerando la Figura 4.18 la totalidad de viajes sin descuento supera ampliamente a los viajes con descuento en horas pico o punta; lo que podría reflejar que el STPPU es atractivo para aquellos sectores de la sociedad que no requieren un descuento en su tarifa de pasaje. Los viajes con descuento replican la estacionalidad, con mayor presencia relativa en mañana y mediodía, consistente con usos educativos y laborales. Si se tuviera información de otras épocas del año también se podría analizar la estacionalidad de la utilización del STPPU a fin de observar la incidencia turística del mismo.

Si se selecciona el conjunto de datos de biciTRAN, se pueden observar las visualizaciones de **Distribución de viajes por hora y género** (antes mencionada) y **Distribución de viajes por hora y tipo de usuario**, la cual consiste en un histograma que muestra la distribución de los tipos de usuarios “Local” o “Turista” que viajaron utilizando el sistema. La Figura 4.19 muestra esta última visualización. Esta gráfica permite visualizar la demanda turística del sistema, diferenciar patrones de uso, ayudar a diseñar tarifas diferenciadas y su impacto, y resulta útil para municipios y entes de turismo para entender cómo se mueven los visitantes dentro de la ciudad.

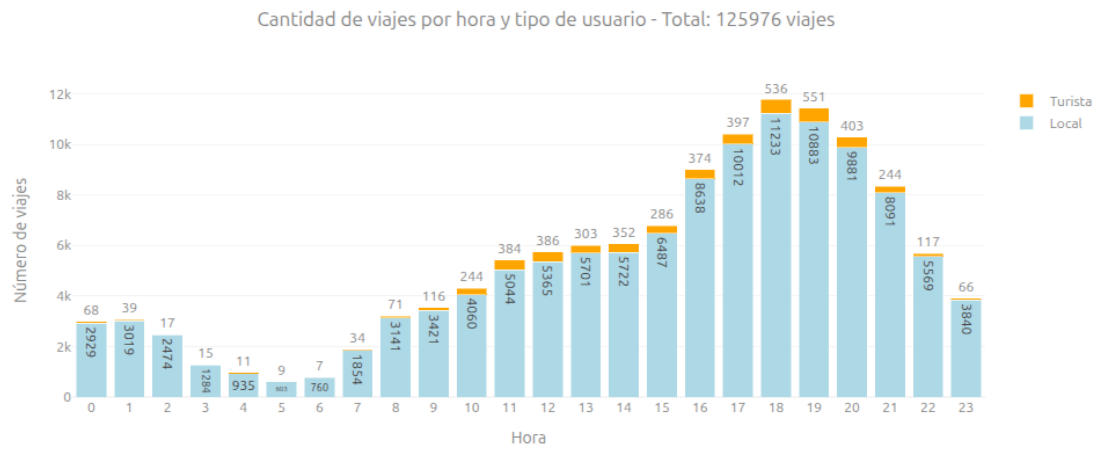


Figura 4.19: Distribución de viajes por hora y tipo de usuario (biciTRAN).

En la Figura 4.19 se puede ver que los usuarios locales concentran la demanda en todas las horas ($\approx 94 - 97\%$), con picos totales en 18-19 h y un hombro 12-14 h. La participación de turistas es baja, pero se incrementa al mediodía ($\approx 6\%$), mientras que en la madrugada y la noche es marginal. Este patrón sugiere priorizar redistribución para locales en 17:00-20:00 y asegurar disponibilidad en zonas turísticas 12:00-14:00.

La Figura 4.20 muestra como en el sistema de bicicletas compartidas biciTRAN la demanda se concentra en 17-19 h, con un pico en 18 h, mientras que la madrugada y la noche tardía registran volúmenes bajos. En todas las franjas predomina Masculino ($\approx 53 - 55\%$), seguido por Femenino ($\approx 44 - 46\%$) y una proporción marginal No especificado. Este patrón sugiere priorizar redistribución y disponibilidad de bicicletas entre 16:30-19:30 y medidas de seguridad/infraestructura que favorezcan el uso en horarios de descenso de luz.

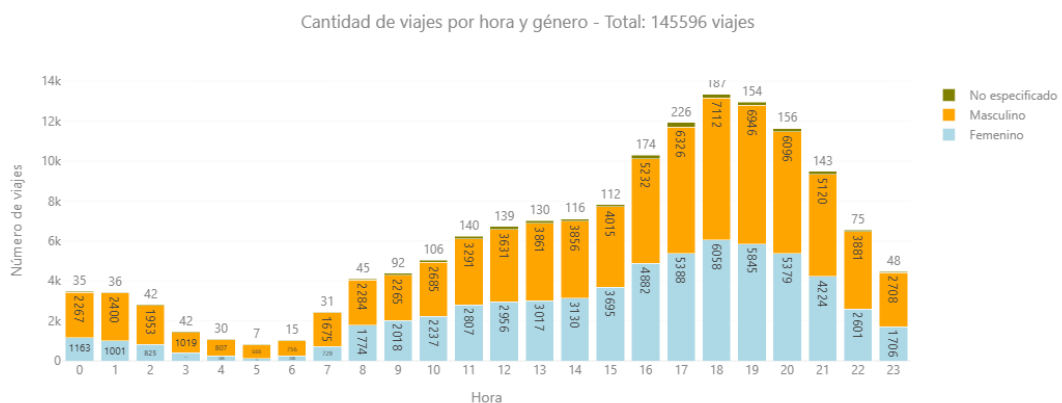


Figura 4.20: Distribución de viajes por hora y género de usuario (biciTRAN).

4.2.1.3. Gráficos de torta

La siguiente pestaña de las visualizaciones gráficas y tablas es la de *Gráficos de torta*, esta consta de dos visualizaciones útiles para analizar la proporcionalidad de las subcategorías dentro de dos posibles agrupaciones (Figura 4.21). Ambas gráficas pueden ser utilizadas para conocer más a los usuarios del sistema y diseñar mejores políticas de movilidad.

El conjunto de datos de biciTRAN incluye dos visualizaciones que, aunque diferentes de las mostradas para mendoTRAN (véase Figura 4.22), serán explicadas con más detalle posteriormente en esta sección.



Figura 4.21: Opciones del ítem Gráficos de torta en visualizaciones gráficas y tablas (mendoTRAN).

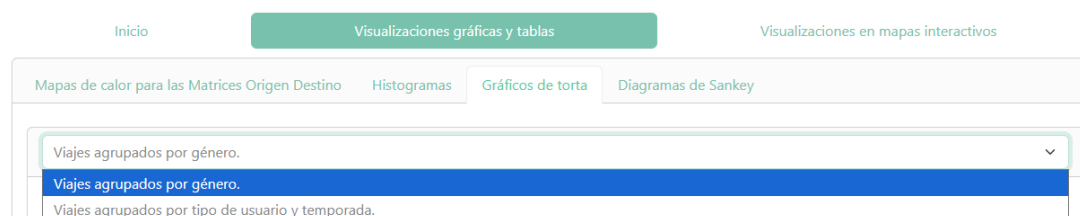


Figura 4.22: Opciones del ítem Gráficos de torta en visualizaciones gráficas y tablas (biciTRAN).

La primera, **Viajes agrupados por género**, la cual muestra la proporción de viajes realizados en la categoría género. En el sistema mendoTRAN (Figura 4.23), el 78,2 % de las validaciones corresponde a Femenino, el 20,3 % a Masculino y el 1,47 % a No especificado, evidenciando una marcada feminización del uso en la muestra analizada. En el sistema de bicicletas compartidas (Figura 4.24), el 55,5 % de los viajes corresponde a usuarios masculinos y el 42,9 % a femeninos (No especificado: 1,6 %), lo que evidencia una brecha moderada a favor de masculinos y sugiere enfocar intervenciones de seguridad e infraestructura para favorecer la participación femenina. Esta composición contrasta con el sistema mendoTRAN (predominio femenino), lo que sugiere patrones de uso diferenciados por modo, pudiendo ser causados por múltiples motivos y dejando una oportunidad de investigación y mejoras para ambos sistemas.

Viajes agrupados por género.

Muestra la proporción de viajes realizados en la categoría género.

Filtra los viajes por fecha seleccionada.

Todos los días

Fecha Todos - Proporción de viajes por género - Total: 394603 viajes

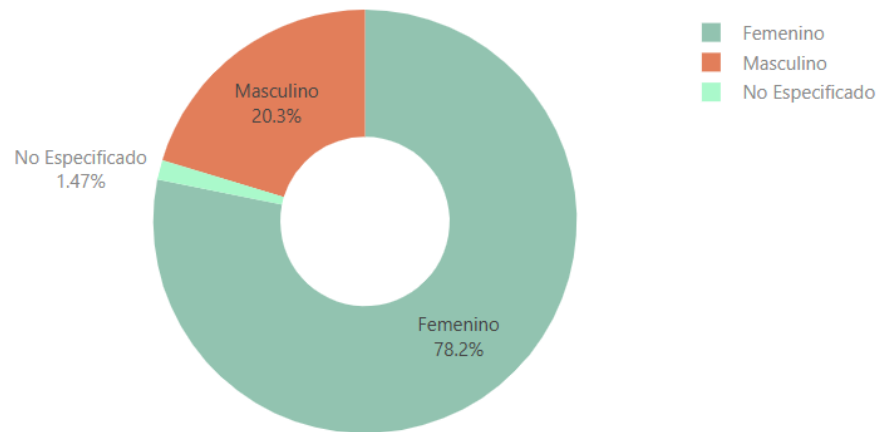


Figura 4.23: Viajes agrupados por género (mendoTRAN).

Proporción de viajes por género - Total: 145596 viajes

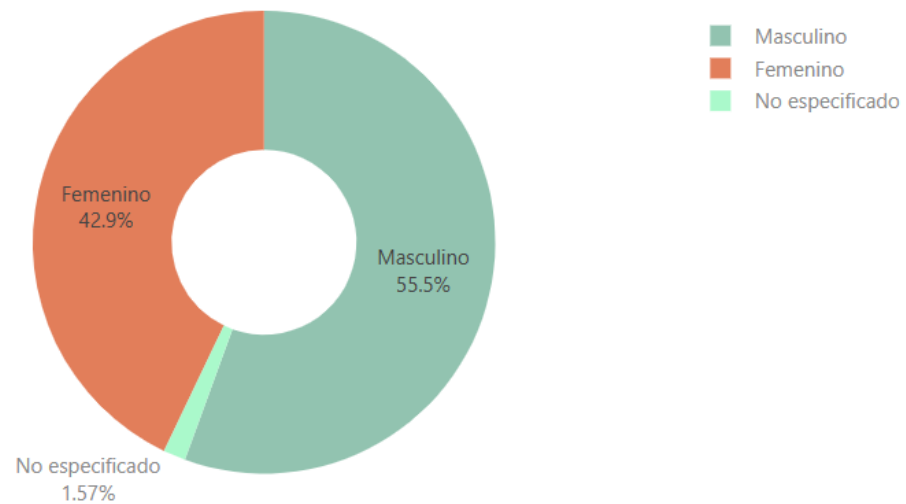


Figura 4.24: Viajes agrupados por género (biciTRAN).

La segunda opción, **Viajes agrupados por descuento**, la cual muestra la proporción de viajes realizados analizando solo aquellos que tienen algún tipo de descuento (Figura 4.25). Esta última visualización solo aplica para el conjunto de datos de mendoTRAN. Dentro de los viajes con descuento predominan AUH (25,6%) y PROGRESAR (24,0%), seguidos por Jubilado (19,2%) y Pensionado (14,3%); Personal del trabajo doméstico representa 14,3%. En conjunto, estos grupos justifican refuerzos operativos en 7-8 h, 12-14 h y franjas de media mañana con foco en accesibilidad y regularidad.



Figura 4.25: Viajes agrupados por descuento (mendoTRAN).

Para el conjunto de datos de biciTRAN, se pueden seleccionar dos visualizaciones posibles, la antes mencionada **Viajes agrupados por género**, y otra llamada **Viajes agrupados por tipo de usuario y temporada**, la cual, eligiendo entre las opciones “Todos los días”, “Anual” o “Por día de semana” permite agrupar el conjunto de datos y mostrar un gráfico de la distribución de tipo de usuario para cada grupo (Figura 4.26). La proporción de viajes turistas en biciTRAN presenta una estacionalidad marcada: alcanza 11,3% en enero y 13,8% en febrero, con un repunte en diciembre (3,6%), mientras que el resto del año se mantiene entre 0,6-2,3%. En consecuencia, el uso es mayoritariamente local (> 97 % en 10 de 12 meses), lo que sugiere priorizar la operación para desplazarse durante el año y reforzar capacidad, redistribución y señalización orientada a turistas en la temporada estival.



Figura 4.26: Viajes agrupados por tipo de usuario y temporada (biciTRAN).

4.2.1.4. Diagramas de Sankey

Para este último tipo de visualización en el conjunto de datos de mendoTRAN se puede filtrar por día de semana (Figura 4.27), mientras que en el conjunto de datos de biciTRAN se puede filtrar por tipo de usuario (Figura 4.28). Estas visualizaciones permiten ver rápidamente qué pares de departamentos concentran la mayor movilidad, y ayudan a priorizar mejoras de infraestructura, frecuencias y planificación de rutas troncales.

El diagrama de Sankey de la Figura 4.27 evidencia un patrón fuertemente centralizado: el mayor flujo corresponde a viajes internos en Capital ($\approx 89,2k$) y los corredores radiales con mayor volumen conectan Capital con Godoy Cruz, Guaymallén, Las Heras, Maipú y Luján de Cuyo. Existen además flujos tangenciales relevantes (p. ej., Guaymallén-Godoy Cruz, Maipú-Luján) que justifican mejorar la conectividad directa entre periferias para descongestionar el centro.

En cuanto al diagrama de Sankey de biciTRAN (Figura 4.28) evidencia un uso marcadamente local: el flujo dominante es Capital a Capital y, en menor medida, Godoy Cruz presenta altos viajes internos. Los interdepartamentales se concentran en los ejes Capital-Godoy Cruz y Capital-Guaymallén, mientras que las conexiones tangenciales son secundarias. Esto sugiere priorizar densificación de estaciones y continuidad de infraestructura en esos corredores, además de fortalecer enlaces laterales para evitar desvíos por el centro.

Diagrama de Sankey de viajes.

Muestra el flujo de viajes entre diferentes departamentos y categorías.

Filtra los viajes por fecha seleccionada.

Todos los días

Diagrama de Sankey para viajes entre departamentos

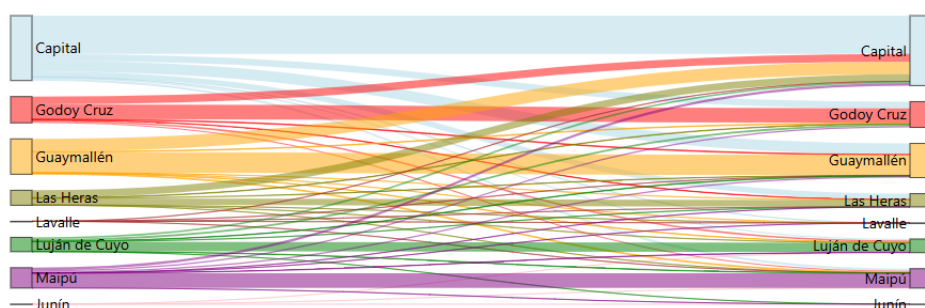


Figura 4.27: Diagrama de Sankey de viajes del conjunto de datos mendoTRAN.

Diagrama de Sankey para viajes entre departamentos

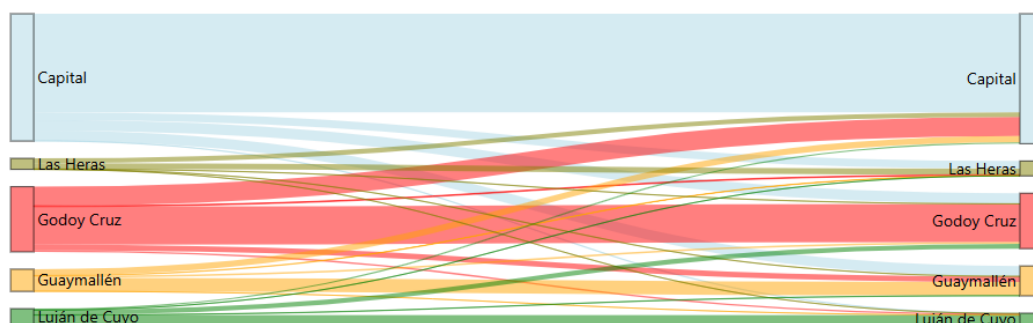


Figura 4.28: Diagrama de Sankey de viajes del conjunto de datos biciTRAN.

4.2.2. Visualizaciones en mapas interactivos

La pestaña de visualizaciones en mapas interactivos incluye un menú desplegable que ofrece una selección de las visualizaciones accesibles (Figura 4.29). Algunas de las opciones disponibles son las siguientes:

- Paradas de llegada para una línea seleccionada y contador de descensos (mendoTRAN).
- Cantidad de salidas por hora (mendoTRAN).
- Descensos por parada (mendoTRAN).

- Pares de paradas (mendoTRAN).
- Destinos por departamento (mendoTRAN).
- Viajes por hora con ascensos y descensos (mendoTRAN).
- Mapa de calor de viajes (mendoTRAN).
- Pares de estaciones con más duración de viaje (biciTRAN).
- Mapa de calor de viajes (biciTRAN).
- Capacidad por estación (biciTRAN).

Visualización de Información para el Transporte Público de Pasajeros Urbano MendoTran y BiciTran (Mendoza)

Trabajo Final de Carrera - Licenciatura en Ciencias de la Computación - Facultad de Ingeniería - Universidad Nacional de Cuyo

Alumno: Gastón Ezequiel Cavallo (cavallogaston@gmail.com)

Tutora: Dra. Ana Carolina Olivera (acolivera@conicet.gov.ar)

Se utilizan los datos de: MendoTran ▾

Inicio Visualizaciones gráficas y tablas Visualizaciones en mapas interactivos

Selector de visualizaciones en mapas interactivos

Elija una visualización dinámica para ver en el mapa

▾

Glosario de términos

Este es el glosario de términos utilizados en la visualización dinámica

Ascensos: Son los viajes que salen de una estación, sinónimos utilizados: salidas.

Descensos: Son los viajes que llegan a una estación, sinónimos utilizados: llegadas, arribos.

Figura 4.29: Página de visualizaciones en mapas interactivos.

4.2.2.1. Visualizaciones para mendoTRAN

La primera opción del menú desplegable es **Paradas de llegada para una línea seleccionada y contador de descensos**. Esta opción presenta un mapa que muestra la ubicación de todas las paradas a lo largo de la línea previamente escogida (Figura

4.30). Al hacer clic en una parada específica, aparece un *pop-up* que revela el número de personas que emplearon la línea seleccionada y descendieron en dicha parada (Figura 4.31), lo cual facilita comprender la demanda real dentro de la red. Esta visualización ayuda a detectar paradas de autobuses con alta demanda de bajadas, lo que puede indicar:

- Puntos de atracción importantes (escuelas, centros de trabajo, hospitales).
- Necesidad de infraestructura (refugios, paradas seguras, accesibilidad).

Estaciones de llegada y contador de descensos

Mapa de las estaciones de llegada y contador de descensos

Dada una línea, imprime todas las estaciones que esta recorre y al hacer click en alguna de estas, el mapa muestra la cantidad de descensos que ocurrieron en la estación seleccionada. Esta visualización utiliza el dataset completo de viajes.

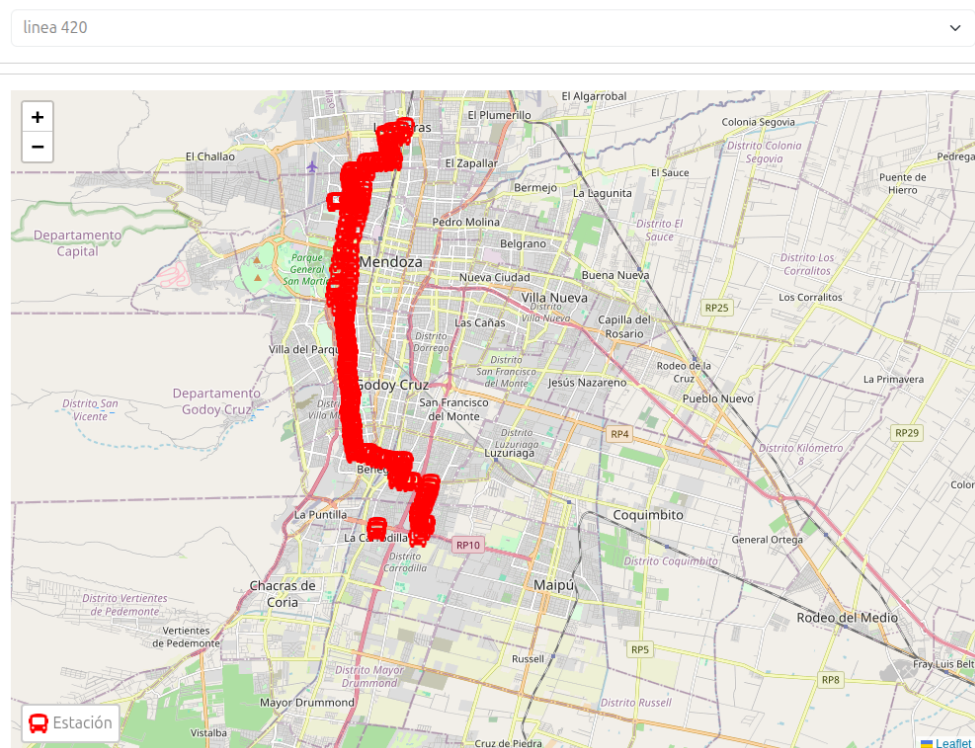


Figura 4.30: Visualización de paradas de llegada para una línea seleccionada y contador de descensos (mendoTRAN).



Figura 4.31: Visualización de *pop-up* en el mapa (mendoTRAN).

En la pestaña dedicada a las **visualizaciones en mapas interactivos**, se presenta la visualización de la **Cantidad de salidas por hora**. Este recurso incluye un mapa que detalla todas las paradas de autobús. Al seleccionar una parada específica, el mapa se actualiza mostrando un histograma que revela la cantidad de salidas desde esa parada a lo largo de las distintas horas del día (Figura 4.32). Además, el mapa agrupa las paradas dependiendo de su proximidad. Cuando un usuario se acerca a una región o selecciona un grupo de paradas, el mapa se ajusta ampliando la vista y demostrando un desglose pormenorizado de las paradas incluidas en ese grupo. Esta representación temporal de la demanda de ascensos para cada parada es valiosa para identificar discrepancias entre la oferta y la demanda de transporte.

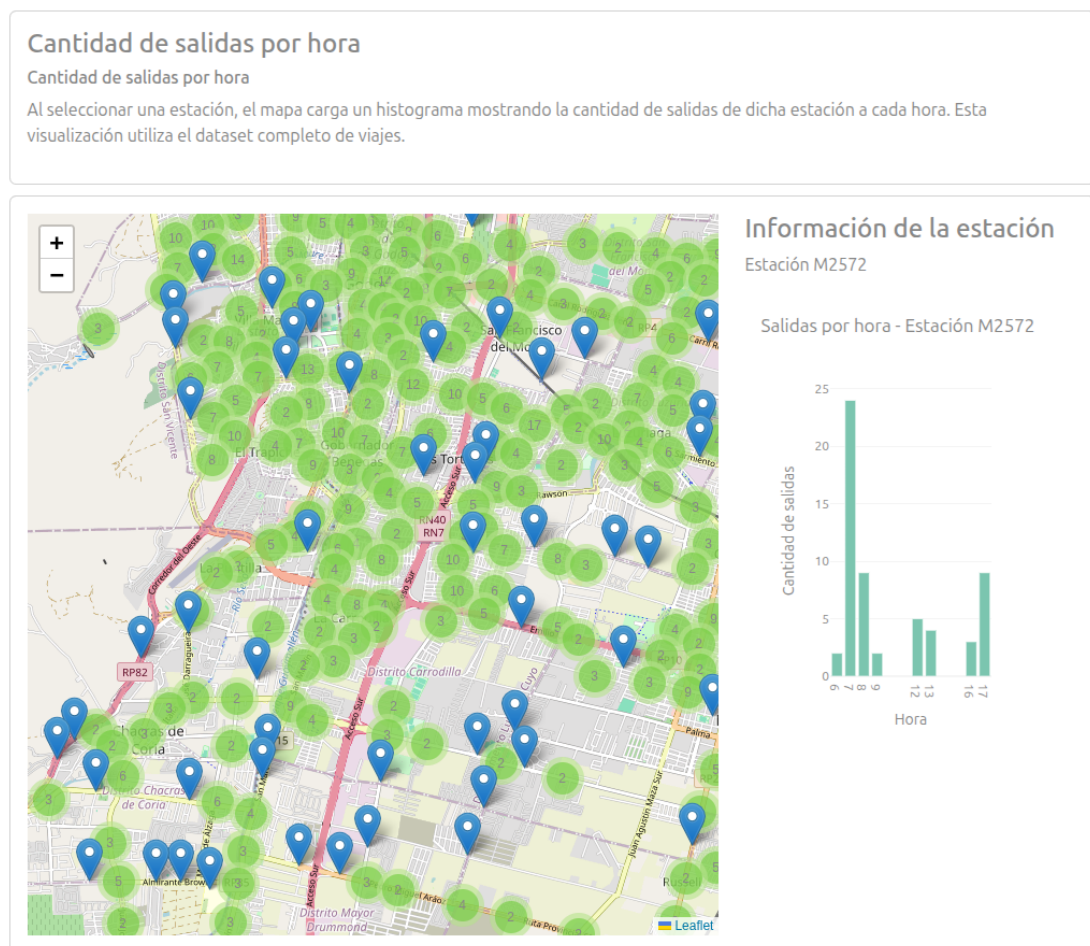


Figura 4.32: Visualización de la cantidad de salidas por hora (mendoTRAN).

En la visualización interactiva titulada **Descensos por parada**, se presenta un mapa donde una parada previamente seleccionada se resalta en azul, mientras que todas las paradas que recibieron viajes provenientes de dicha parada se indican en rojo (Figura 4.33). Al elegir una parada, se despliega el número de viajes asociados,

mostrando tanto el número de salidas desde la parada de origen como el número de llegadas a las paradas de destino. Esta herramienta visual resulta especialmente valiosa para examinar las conexiones origen-destino entre paradas, permitiendo destacar las rutas directas entre ellas. Además, facilita la comprensión de los principales flujos de tráfico a partir de una parada estratégica, la identificación de posibles rutas con alta demanda y la evaluación de la eficiencia de la red de transporte.

Descensos por estación

Dada una estación de origen, muestra los viajes originados en esa estación, las estaciones de destino que tuvieron dichos viajes y la cantidad de descensos que tuvieron en cada una de ellas

Dada una estación elegida como origen, el mapa muestra la estación seleccionada en celeste y las cantidad de viajes que salieron de dicha estación. Además, muestra las estaciones de destino y cantidad de descensos que tuvo en cada una de ellas. Esta visualización utiliza el dataset completo de viajes.

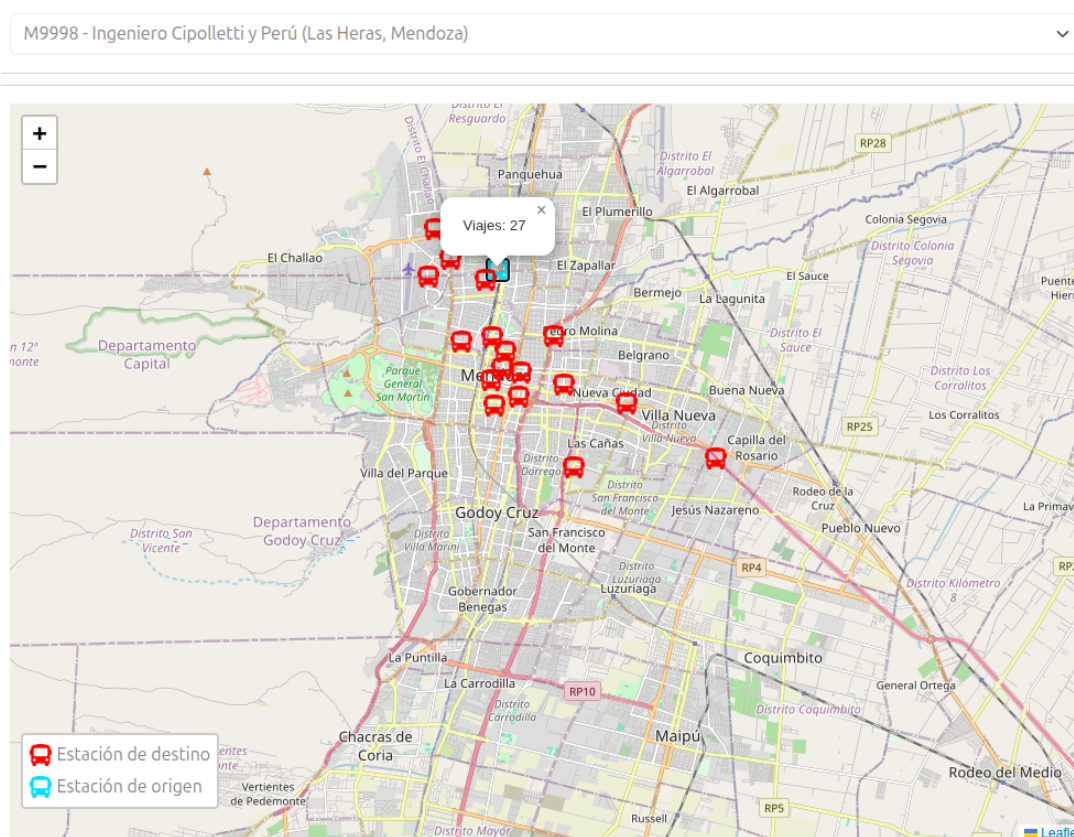


Figura 4.33: Visualización de descensos por parada (mendoTRAN).

Pares de paradas incluye un mapa donde se representan los 10 pares de paradas con el mayor número de desplazamientos mutuos, considerando los días seleccionados a través del filtro aplicado (Figura 4.34). Esta representación gráfica señala las rutas con mayor tráfico de pasajeros entre dos ubicaciones específicas. Identificar estas rutas de alta demanda es fundamental para orientar decisiones sobre inversiones prioritarias, tales como mejorar la frecuencia de los servicios, incrementar la cantidad de vehículos

disponibles, actualizar infraestructura importante, diseñar rutas directas más eficientes y evaluar la dependencia de ciertas conexiones estratégicas.

Pares de estaciones

Pares de estaciones de origen y destino con mayor cantidad de viajes

Muestra los pares de estaciones de origen y destino con mayor cantidad de viajes. En verde, la estación de origen y en azul, la estación de destino. Además, una línea que une ambas estaciones y el número de viajes que se realizaron entre ambas estaciones se muestra al pasar el cursor por encima de la línea. Esta visualización permite el filtrado por día.

Filtro de días, seleccione el día que desea ver

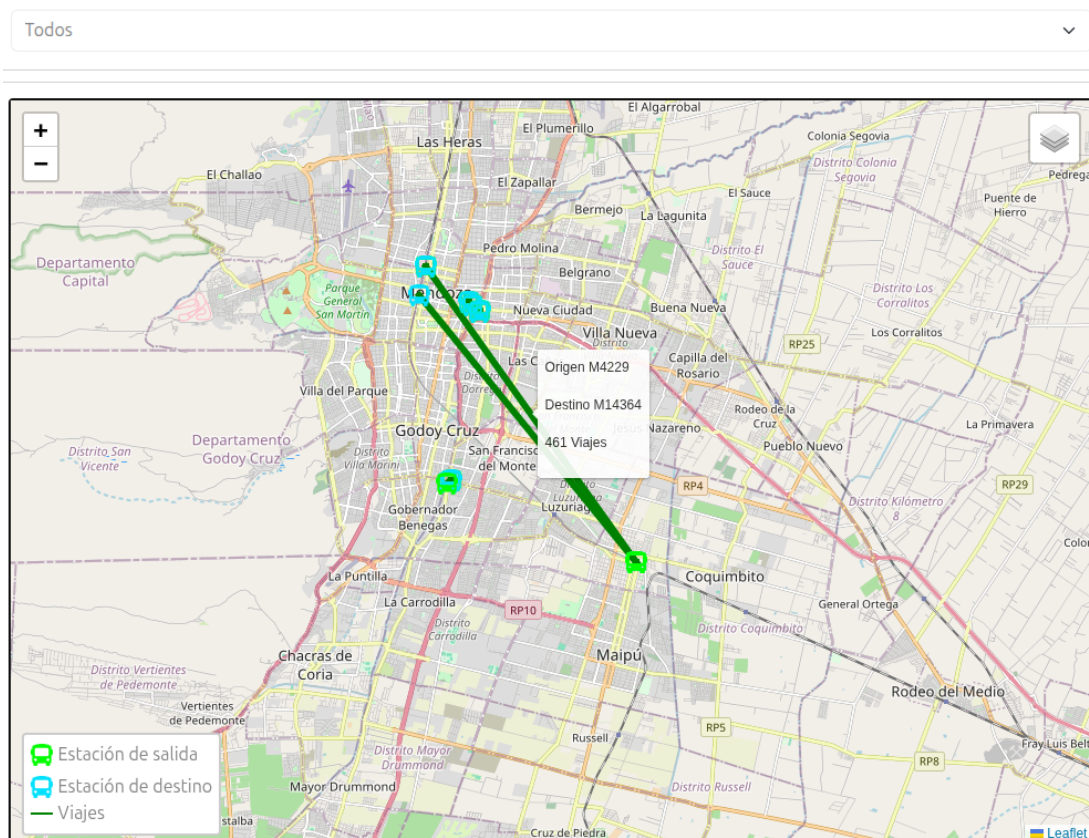


Figura 4.34: Pares de paradas con más viajes (mendoTRAN).

Por su parte, **Destinos por departamento** se trata de un mapa coroplético superpuesto sobre un mapa de la ciudad de Mendoza (Figura 4.35). Al elegir un departamento en la lista, el mapa se actualiza para mostrar un mapa coroplético que incluye todos los departamentos a los que se dirigieron los viajes iniciados en el departamento seleccionado. Esta herramienta es útil para examinar la dependencia territorial en mendoTRAN, es decir, qué áreas generan o atraen un número mayor de viajes, así como para identificar disparidades entre diferentes zonas. Se puede realizar zoom para su mejor visualización. El formato coroplético mejora la comprensión de la concentración espacial.

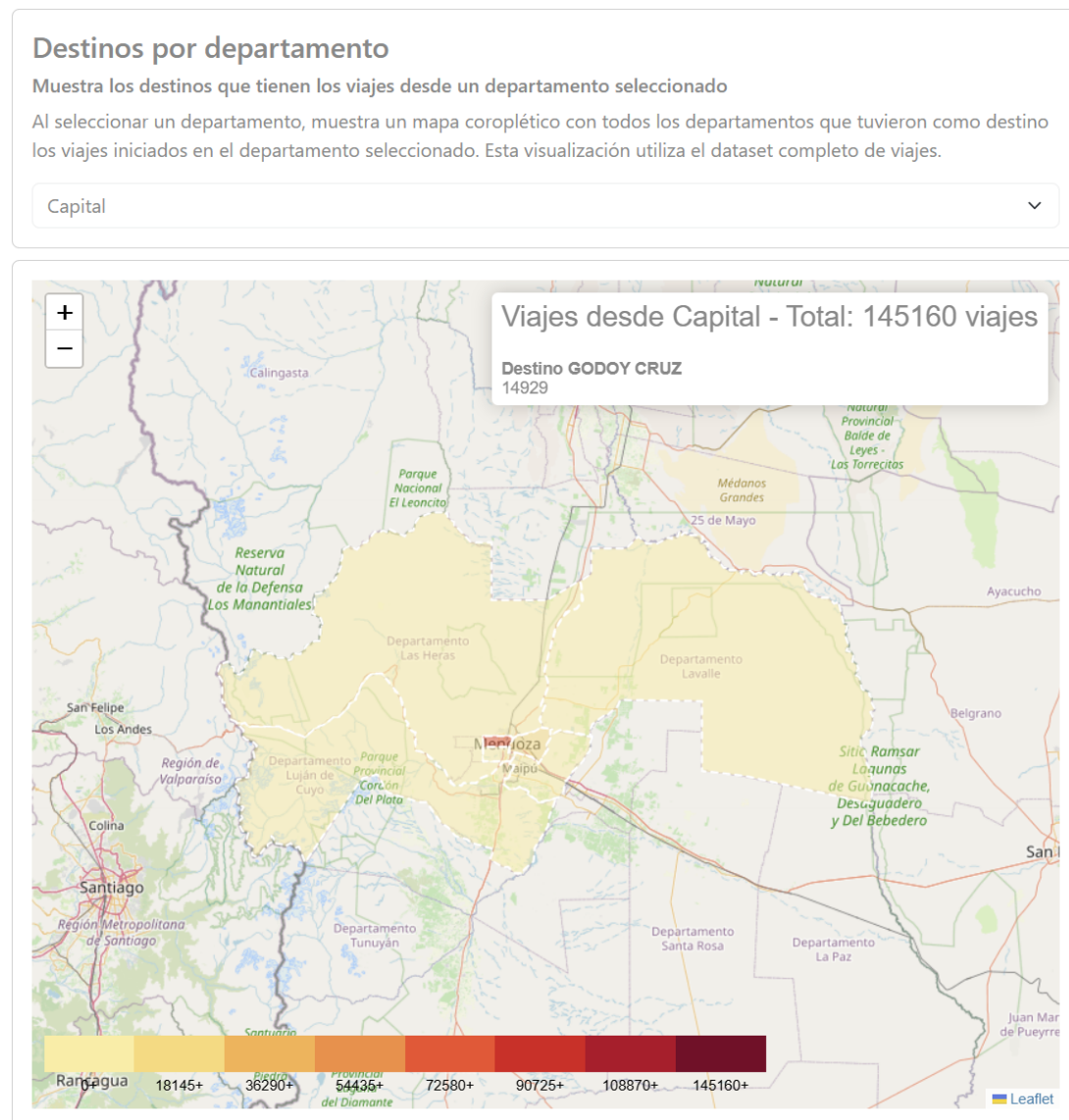


Figura 4.35: Visualización de destinos por departamento (mendoTRAN).

La representación **Viajes por hora con ascensos y descensos** forma parte del mismo conjunto de visualizaciones y se presenta mediante un mapa similar al de **Cantidad de salidas por hora**, pero su histograma asociado contrasta las salidas, llegadas y movimientos totales de la parada seleccionada (como se observa en la Figura 4.36). Destinada principalmente a analizar los movimientos en las paradas, la barra que representa los viajes totales permite distinguir visualmente los períodos operativos con mayor facilidad. Faculta la identificación de paradas con mayor circulación de pasajeros, posibilita la comparación detallada de los flujos de ascenso y descenso por cada hora, y proporciona una representación clara de la actividad de una parada a lo largo de varias horas del día.



Figura 4.36: Visualización de viajes por hora con ascensos y descensos (mendoTRAN).

Finalmente, se llega a la interfaz interactiva denominada **Mapa de calor de viajes**, que ofrece una representación dinámica de los desplazamientos en las paradas, mostrando un mapa de calor que se actualiza con datos cada hora. Esta herramienta permite elegir entre dos opciones: el mapa de calor de salidas (Figura 4.37) y el mapa de calor de llegadas (Figura 4.38). Esta visualización permite identificar de manera espacial y dinámica la concentración de movimientos en toda la red, facilitando la identificación intuitiva de los puntos calientes (*hot spots*) de actividad y su variación a lo largo de las horas.

CAPÍTULO 4. RESULTADOS Y DISCUSIONES

Mapa de calor de viajes

Mapa de calor de viajes por estación y hora de salida o llegada

En este mapa se debe seleccionar el tipo de viaje a desplegar (ascenso o descenso). Una vez realizada la selección se muestra el mapa de calor interactivo. El mapa interactivo presenta la cantidad de transacciones ocurridas a cada hora. Esta visualización dinámica utiliza como recurso todo el conjunto de datos disponible.

Filtros para el mapa de calor de viajes, seleccione salidas o llegadas

- ☒ Salidas de cada estación
- ☐ Llegadas de cada estación

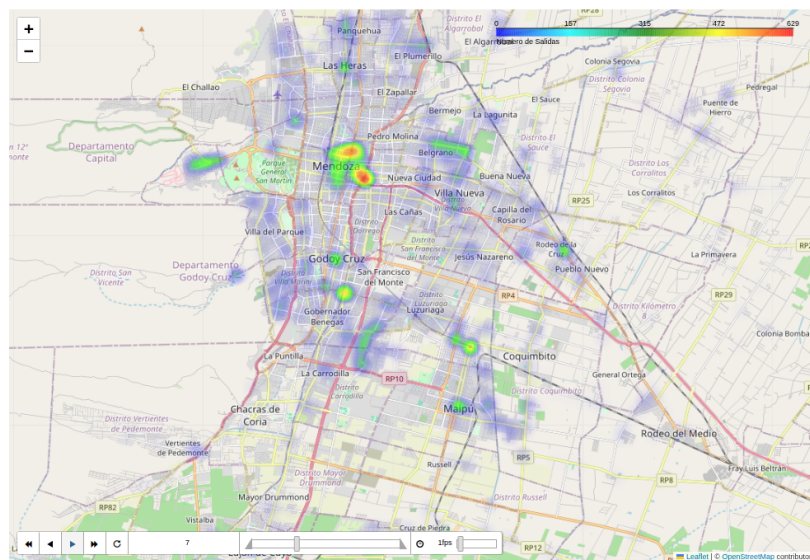


Figura 4.37: Visualización de mapa de calor de viajes para salidas (mendoTRAN).

Mapa de calor de viajes

Mapa de calor de viajes por estación y hora de salida o llegada

En este mapa se debe seleccionar el tipo de viaje a desplegar (ascenso o descenso). Una vez realizada la selección se muestra el mapa de calor interactivo. El mapa interactivo presenta la cantidad de transacciones ocurridas a cada hora. Esta visualización dinámica utiliza como recurso todo el conjunto de datos disponible.

Filtros para el mapa de calor de viajes, seleccione salidas o llegadas

- ☐ Salidas de cada estación
- ☒ Llegadas de cada estación

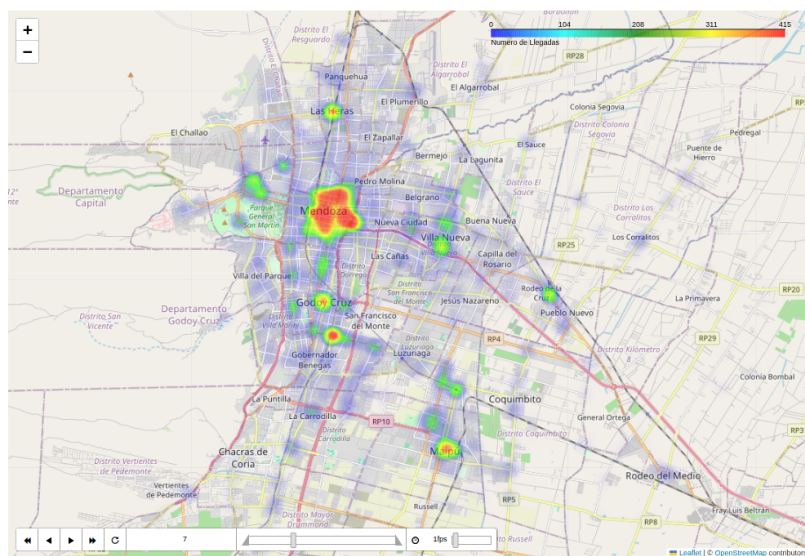


Figura 4.38: Visualización de mapa de calor de viajes para llegadas (mendoTRAN).

4.2.2.2. Visualizaciones para biciTRAN

Entre las visualizaciones del conjunto de datos de biciTRAN, se encuentra **Pares de estaciones con más duración de viaje**. Esta visual muestra las estaciones de bicicletas compartidas de origen y destino en viajes de mayor duración, conectadas por una línea que las une. La estación de origen está en verde, mientras que la de destino se presenta en azul. Al pasar el cursor sobre la línea, se revela la duración del viaje entre esas estaciones (véase la Figura 4.39). Esta visualización permite filtrar los datos seleccionando el mes de interés para el análisis. Además, facilita la identificación de rutas con uso prolongado, lo que puede indicar trayectos extensos, posibles problemas de redistribución de bicicletas o rutas preferidas por turistas o para recreación. Esto también permite analizar si hay un desequilibrio en estaciones con muchas salidas y pocas llegadas. La inclusión de los puntos turísticos de interés en el análisis de las rutas encontradas en este apartado podría mejorar en gran medida la explotación del sistema en la zona.

Pares de estaciones con más duración de viaje

Pares de estaciones de origen y destino con mayor duración de viajes

Muestra los pares de estaciones de origen y destino de los viajes con mayor duración. En verde, la estación de origen y en azul, la estación de destino. Además, una línea que une ambas estaciones y la duración del viaje que se realizó entre ambas estaciones se muestra al pasar el cursor por encima de la línea.

Filtro de meses, seleccione el mes que desea ver

Enero

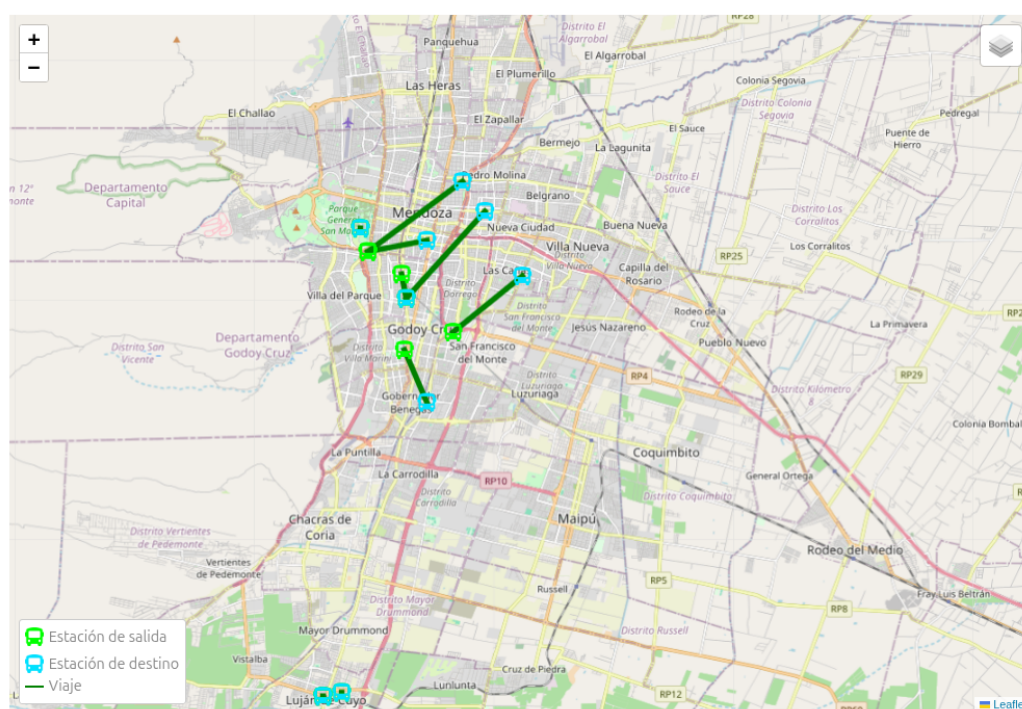


Figura 4.39: Pares de estaciones con más duración de viaje (biciTRAN).

La visualización **Mapa de calor de viajes** permite filtrar por mes (Figura 4.40). Puede ser utilizada para detectar la estacionalidad del uso del sistema de bicicletas y para evaluar si se necesita reforzar infraestructura ciclista o aumentar la capacidad de ciertas estaciones en temporada alta; también puede ser utilizada para capturar posible demanda insatisfecha.

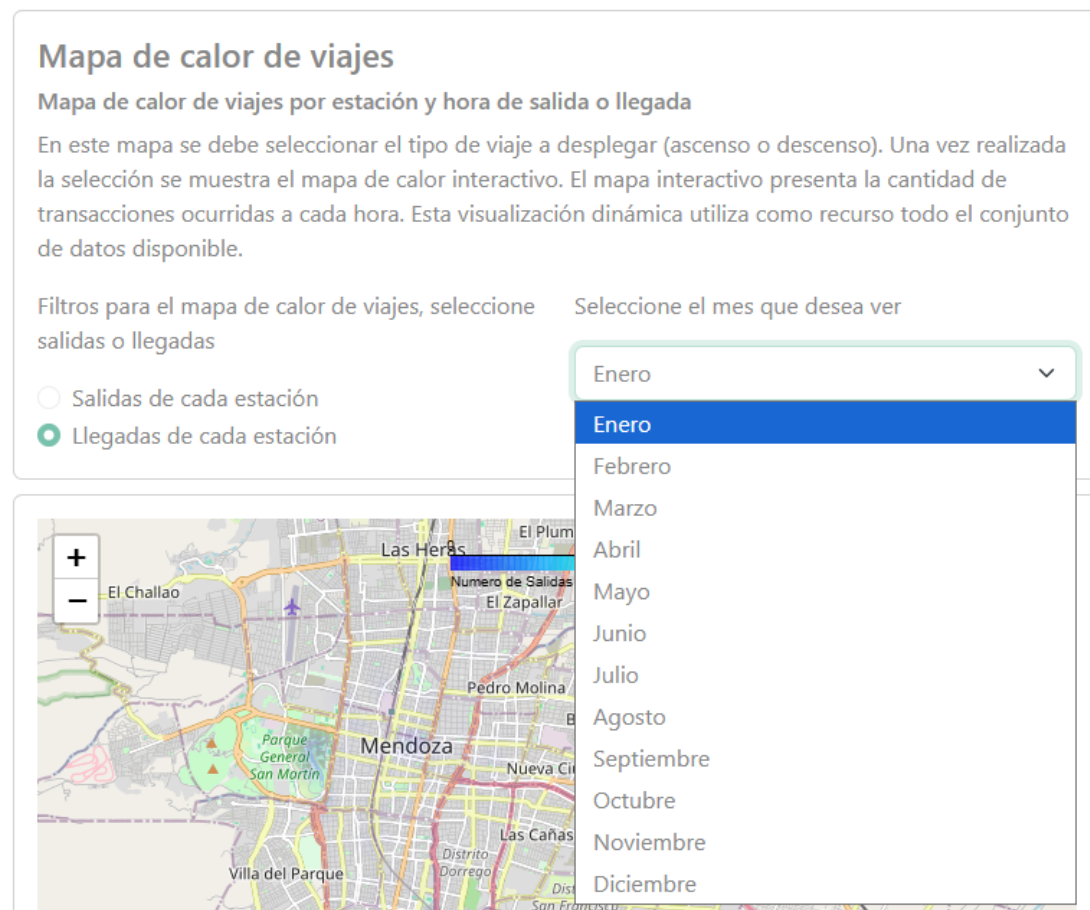


Figura 4.40: Visualización de mapa de calor de viajes para salidas filtrado por mes (bicíTRAN).

La última visualización es la de **Capacidad por estación**, que consiste en un mapa con las estaciones; al hacer clic en alguna de ellas se muestra la cantidad de bicicletas que la estación puede albergar (Figura 4.41). Este mapa permite comparar la capacidad instalada con la demanda real y ayuda a:

- Detectar cuellos de botella: estaciones que constantemente quedan vacías o saturadas.
- Planificar ampliaciones o redistribución de anclajes.
- Tomar decisiones sobre mantenimiento e inversiones.

Capacidad por estación

Muestra la capacidad de las estaciones de Bicitran

Muestra un mapa con la capacidad de las estaciones de Bicitran. La capacidad se refiere a la cantidad de bicicletas que cada estación puede albergar.

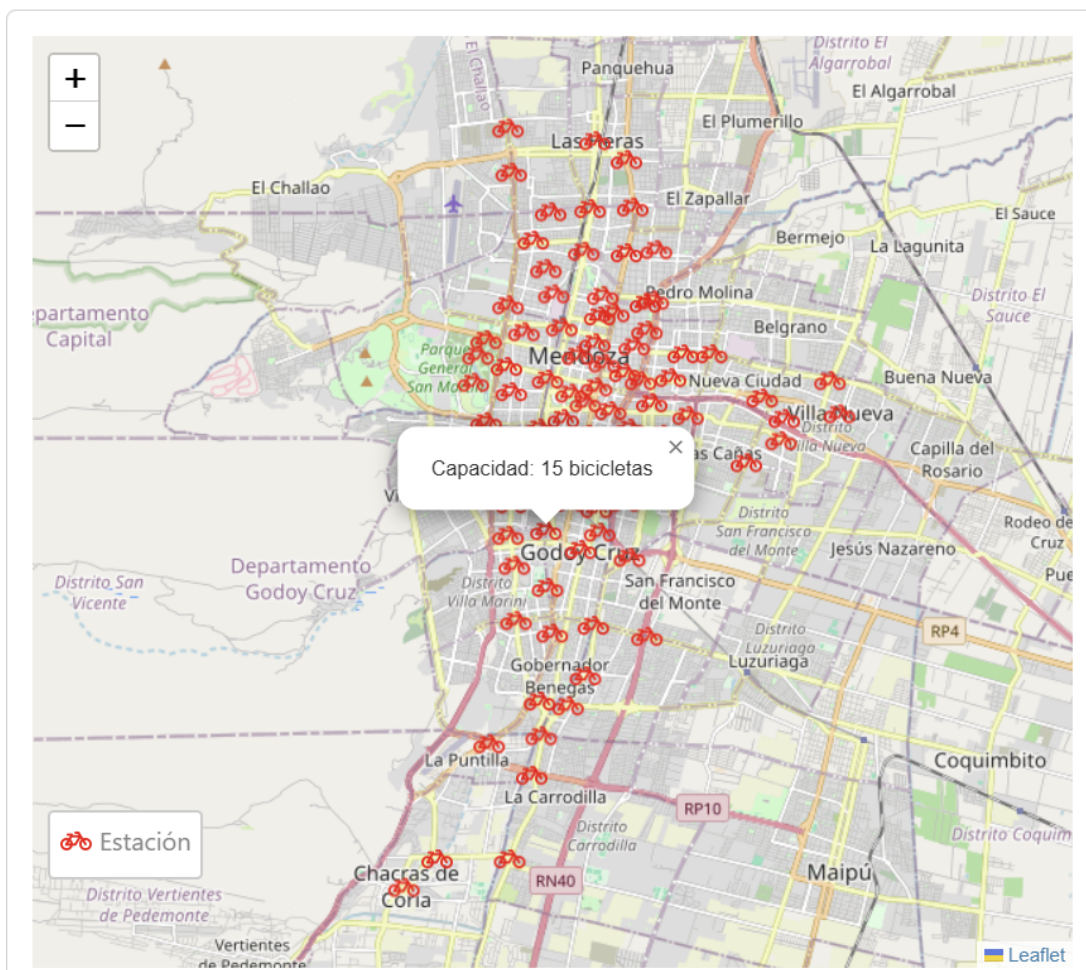


Figura 4.41: Capacidad por estación (biciTRAN).

4.3. Resumen del capítulo

El análisis de los resultados obtenidos a partir de las matrices origen-destino y las visualizaciones generadas permite identificar patrones significativos en el comportamiento de los usuarios del transporte público del AMM. En primer lugar, se observa una fuerte concentración de viajes en horarios pico para el sistema mendoTRAN, tanto por la mañana como por la tarde, lo cual coincide con los horarios laborales y escolares. Mientras que para el sistema biciTRAN se puede observar alta demanda en horarios

vespertinos, más específicamente entre las 17 y las 20 hs. Esta regularidad horaria se ve reflejada incluso en los histogramas de distribución por hora y género para ambos sistemas.

Del mismo modo, las disparidades de género destacan una participación considerablemente mayor del género femenino para el sistema mendoTRAN a lo largo de casi todo el horario de servicio. Esto podría estar vinculado a la distribución de trabajos o responsabilidades familiares, aunque sería necesario realizar un análisis demográfico más detallado para explorar estas observaciones en profundidad. Por otro lado, en cuanto al sistema biciTRAN, aunque la diferencia de género no es tan sobresaliente, se observa una predominancia leve de usuarios masculinos. Esta tendencia podría señalar posibles deficiencias en medidas de seguridad o infraestructura que deberían atenderse, especialmente para mejorar las condiciones de uso en horarios de escasa iluminación.

En el caso de mendoTRAN, se observa que una cantidad significativa de usuarios carecen de descuentos, indicando que personas que podrían preferir el uso de transporte privado deciden emplear el STPPU. Esto refleja la eficiencia del sistema, manteniendo una superior relación calidad/precio en comparación con los vehículos particulares. Esta inclinación sugiere una posible disminución en el uso de automóviles privados, contribuyendo así a la reducción de la contaminación causada por vehículos y promoviendo la sostenibilidad urbana.

Con la configuración seleccionada **TC8 (1000 m, 2 h)**, el *trip chaining* en mendoTRAN clasificó el 85,64% de las transacciones como viajes únicos y el 14,36% como transbordos, valor consistente con el análisis de sensibilidad que mostró rendimientos decrecientes al ampliar umbrales. La demanda presenta picos en 7-8 h, 12-13 h y 18 h, con predominio femenino en el total de validaciones y con los descuentos AUH y PROGRESAR explicando cerca de la mitad del universo bonificado, seguidos por jubilados y pensionados. Los flujos espaciales se organizan principalmente en desplazamientos intradepartamentales y en corredores radiales hacia Capital, mientras que las conexiones tangenciales son secundarias. En conjunto, los resultados sugieren que la red ofrece trayectos directos competitivos para el usuario promedio y que los transbordos se concentran en pocos corredores y nodos de intercambio.

En cuanto a la segmentación geográfica, las matrices OD por departamento revelan una fuerte dependencia entre los departamentos del Gran Mendoza para ambos sistemas. Esta concentración territorial reafirma la necesidad de un sistema eficiente de transporte metropolitano. Las visualizaciones en forma de mapas de calor y coropléticos facilitan la identificación de estos patrones y su comunicación a decisores.

Por su parte, las visualizaciones interactivas en mendoTRAN, como el registro

de pasajeros que abordan o dejan el transporte en cada parada, son esenciales para analizar la demanda y carga en lugares específicos, lo cual es crucial para mejorar las frecuencias, infraestructura y áreas de transbordo. Identificar los pares de paradas con mayor tráfico proporciona una valiosa herramienta para enfocar inversiones o aplicar estrategias de gestión de la demanda. En biciTRAN, el mapa interactivo que muestra las estaciones con mayor duración de viaje y el que señala la capacidad de dichas estaciones podrían señalar futuras estaciones a añadir que optimizarían tanto la eficiencia como la demanda del sistema.

En conjunto, los resultados no solo validan la utilidad de las técnicas implementadas, sino que también evidencian el potencial de los datos y herramientas de análisis para informar políticas públicas de movilidad basadas en evidencia.

CONCLUSIONES, DESAFÍOS Y PERSPECTIVAS FUTURAS

Este capítulo contiene las conclusiones principales de la presente tesina final de grado. Se resumen los objetivos cumplidos (Sección 5.1), los problemas y desafíos encontrados (Sección 5.2) y los posibles trabajos futuros (Sección 5.3).

5.1. Conclusiones

A lo largo del presente trabajo se desarrolló una solución integral para el análisis y visualización de datos del sistema de transporte público de Mendoza, con el objetivo de contribuir a la mejora de la movilidad urbana. Para ello, se diseñó e implementó una estructura modular compuesta por diversos algoritmos de extracción, transformación y carga de datos (*Extract, Transform and Load* (ETL)), junto con algoritmos de análisis orientados a la generación de matrices origen-destino y un visor web interactivo para la exploración de los resultados.

Uno de los principales logros fue la automatización del proceso de generación de matrices OD utilizando datos reales provistos por el sistema SUBE. Esta automatización permitió obtener representaciones más precisas y actualizadas de los flujos de pasajeros en la red de transporte, superando limitaciones de métodos tradicionales como encuestas manuales o simulaciones teóricas. Asimismo, la integración de visualizaciones geoespaciales interactivas en el visor desarrollado facilita la interpretación de los datos por parte de tomadores de decisiones, urbanistas y público en general, lo cual es fundamental para la planificación y evaluación de políticas de transporte.

El desarrollo de la herramienta se basó en tecnologías de software libre, lo cual no

solo garantiza su accesibilidad y posibilidad de réplica, sino que también promueve su evolución colaborativa. Además, se priorizó la eficiencia en el procesamiento de grandes volúmenes de datos, asegurando que el sistema pueda escalar en el futuro para incorporar información más detallada o en tiempo real.

Entre los principales aportes del trabajo se destacan:

- La generación automatizada de matrices OD a partir de datos de validaciones SUBE, empleando criterios bien fundamentados para la estimación de destinos.
- La creación de un pipeline ETL robusto que permite la limpieza, normalización y almacenamiento eficiente de datos de transporte.
- El desarrollo de un visor web accesible que facilita la exploración espacial y temporal de la demanda del sistema de transporte.

Este trabajo demuestra cómo el uso inteligente de datos puede ser una herramienta poderosa para abordar problemas complejos de movilidad urbana.

A su vez, con base en los objetivos mencionados en el capítulo 1.2, se han cumplido las metas propuestas en este trabajo:

- **Evaluar modelos algorítmicos existentes para la construcción de matrices OD aplicadas al sistema de transporte urbano.**

Este objetivo fue alcanzado mediante la implementación y adaptación del algoritmo de *trip chaining*, el cual permitió inferir destinos en viajes registrados únicamente con datos de ingreso (sistema SUBE). La aplicación de esta metodología permitió obtener matrices OD coherentes con patrones de movilidad urbana conocidos, y representa un avance frente a métodos manuales o encuestas esporádicas.

- **Implementar un sistema de visualización interactivo para analizar la información de movilidad urbana.**

Se logró desarrollar una aplicación web que integra diversas visualizaciones (tablas, gráficos y mapas interactivos) a partir de las matrices generadas. Esta herramienta ofrece una interfaz amigable y funcional para facilitar el análisis espacial y temporal de la movilidad, permitiendo a los tomadores de decisiones detectar patrones, zonas críticas y posibles oportunidades de mejora en la red de transporte.

5.2. Dificultades encontradas

Durante el desarrollo del trabajo, se presentaron diversos desafíos tanto técnicos como logísticos. A continuación, se listan las principales dificultades encontradas:

- Acceso a los datos: La obtención de los datos de SUBE fue uno de los limitantes principales de este trabajo, ya que no solo se consiguió una muestra poco representativa, sino que también era insuficiente para desarrollar sistemas más avanzados o abordar diferentes algoritmos relacionados en el campo de la inteligencia artificial. Esto conllevó posibles sesgos sociodemográficos y cobertura geográfica incompleta limitando el análisis realizado y la posibilidad de maximizar los resultados del sistema.
- Calidad e integridad de los datos: La cantidad y calidad pobre de los datos contribuyó a reducir en gran medida el conjunto de datos.
- La incapacidad de alojar el sistema en servidores gratuitos y la indisponibilidad de servidores universitarios resultó en costos adicionales para el estudiante, así como la necesidad de encontrar un patrocinador para sostener el servicio durante la evaluación de esta tesina final de grado.

5.3. Trabajos futuros

A partir del trabajo realizado, surgen múltiples líneas de acción que podrían ampliarse o profundizarse en futuras investigaciones:

- Mejorar la precisión del algoritmo de *trip chaining*: Utilizando datos complementarios como sensores AVL, información en tiempo real o datos de geolocalización de dispositivos móviles.
- Incluir predicción de demanda: Aplicando modelos de aprendizaje supervisado o series temporales que, a partir de las matrices OD históricas, puedan anticipar flujos futuros.
- Integrar otros sistemas de movilidad: Como monopatines eléctricos o transporte interurbano, para ofrecer una visión más holística del ecosistema de movilidad.
- Desarrollar una *API* pública: Que permita a desarrolladores, urbanistas o investigadores acceder a las matrices y visualizaciones para nuevas aplicaciones.

- Modificar los algoritmos para que sean aplicables a diferentes ciudades que emplean el sistema de pago SUBE

Este trabajo final de carrera complementa y refuerza aún más el aprendizaje obtenido a lo largo de la carrera. Las etapas de investigación y desarrollo contribuyeron a una formación más práctica en torno a los temas tratados en este trabajo y ofrecieron una muestra de la realidad que enfrenta un graduado de la Licenciatura en Ciencias de la Computación cuando se dedica al trabajo de investigación de un tema en específico.

Esta experiencia permitió poner en práctica herramientas y conceptos adquiridos durante la formación académica, integrando conocimientos de análisis de datos, programación, visualización de información y manejo de grandes volúmenes de datos reales. Además, trabajar con datos de movilidad urbana planteó desafíos concretos, como la validación de información, la interpretación de patrones y la necesidad de comunicar resultados de forma clara y accesible para diferentes tipos de usuarios.

A nivel personal, el desarrollo de esta tesina final de grado fortaleció competencias clave como la autonomía en la investigación, la gestión de tiempos, la resolución de problemas complejos y la capacidad de adaptación a tecnologías y marcos teóricos nuevos. En conjunto, este proceso consolidó habilidades técnicas y blandas que serán valiosas tanto para la continuación de mis estudios como para el desempeño profesional en proyectos interdisciplinarios que requieran soluciones innovadoras y basadas en evidencia.

Al mismo tiempo, este trabajo final de carrera no solo aporta una solución técnica funcional y replicable, sino que también sienta las bases para el desarrollo continuo de sistemas inteligentes de apoyo a la toma de decisiones en el ámbito del transporte público, con un claro enfoque en la mejora de la calidad de vida en Mendoza.



A.1. Código fuente de funciones destacadas

A continuación se muestra el código en Python correspondiente a las funciones mencionadas anteriormente, desarrolladas específicamente para el dataset de SUBE de mendoTRAN.

```

1 import pandas as pd
2 from tqdm import tqdm
3
4 def nearestStationOfATrip(startStation, line, stations,
5     busStops):
6     """
7     Encuentra la estacion mas cercana a una estacion de
8     inicio dentro de una linea de transporte.
9
10    Parametros:
11    - startStation: Identificador de la estacion de inicio
12      (ID de la estacion).
13    - line: Linea de transporte (por ejemplo, linea de
14      autobus o tren).
15    - stations: DataFrame que contiene informacion de las
16      paradas, incluyendo ID, latitud y longitud.
17    - busStops: DataFrame que indica las paradas de autobus
18      asociadas a cada linea.

```

```

13
14     Retorna:
15     - DataFrame con las paradas cercanas ordenadas por
16       distancia desde la estacion de inicio.
17     """
18     stops_match = busStops.loc[line] == 1
19     stops_list = stops_match[stops_match].index.tolist()
20     stationsWithLatLon = stations[stations['STATION_ID'].isin
21       (stops_list)]
22     startStationLatLon = stations[stations['STATION_ID'] ==
23       startStation].values[0]
24     stationsWithLatLon.insert(loc=len(stationsWithLatLon.
25       columns), column='dist', value=stationsWithLatLon.
26       apply(lambda x: distance.distance((x.LATITUD,x.
27         LONGITUD), (startStationLatLon[0],startStationLatLon
28         [1])).km, axis=1))
29     stationsWithLatLon = stationsWithLatLon.sort_values(by='
30       dist', ascending=True)
31     return stationsWithLatLon
32
33 def calculate_path_weight(df_path, round_trip=False):
34     """
35     Calcula el peso total de un camino basado en las
36     distancias entre paradas.
37
38     Parametros:
39     - df_path: Lista de objetos que representan un camino,
40       donde cada objeto contiene informacion
41         sobre la estacion y el ramal.
42     - round_trip: Booleano que indica si el camino es de
43       ida y vuelta. Si es True, se incluye
44         la distancia de regreso a la estacion
45         inicial.
46
47     Retorna:
48     - weight: Peso total del camino, calculado como la suma
49       de las distancias entre paradas.
50     """

```

```
39 path = df_path.copy()
40 weight = 0
41 first_station = path[0]
42 previous_station = first_station
43 for element in path[1:]:
44     weight += nearestStationOfATrip(element.STATION_ID,
45                                     previous_station.RAMAL, stations, busStops).iloc[0][
46                                         'dist']
47     if(element.STATION_ID == previous_station.STATION_ID):
48         weight += 100
49     previous_station = element
50 if round_trip:
51     weight += nearestStationOfATrip(first_station.
52                                     STATION_ID, element.RAMAL, stations, busStops).iloc
53                                     [0]['dist']
54 return weight
55
56 def add_destination_to_path(path):
57     """
58     Agrega informacion de destino y distancia a cada
59     elemento de un camino.
60
61     Parametros:
62     - path: Lista de objetos que representan un camino,
63         donde cada objeto contiene informacion
64         sobre la estacion y el ramal.
65
66     Retorna:
67     - path: Lista actualizada con informacion adicional de
68         destino ('destination') y distancia ('dist')
69         para cada elemento del camino.
70     """
71     previous_station = path[-1]
72     for i in range(len(path)):
73         element = path[i]
74         nearestStation = nearestStationOfATrip(element.
75                                                 STATION_ID, previous_station.RAMAL, stations,
76                                                 busStops).iloc[0]
77         previous_station['destination'] = nearestStation['
```

```

        STATION_ID']
69     previous_station['dist'] = nearestStation['dist']
70     previous_station = element
71     return path
72
73 def add_destination_to_trips(df, path, max_distance=0.8,
74                             max_time=2):
75     """
76     Agrega informacion de destino, distancia y cadena a un
77     DataFrame de viajes basado en un camino.
78
79     Parametros:
80     - df: DataFrame que contiene los datos de los viajes.
81     - path: DataFrame o lista que representa un camino, con
82       informacion de paradas, distancias y destinos.
83     - max_distance: Distancia maxima (en kilometros) para
84       considerar que dos viajes estan conectados.
85     - max_time: Tiempo maximo (en horas) entre viajes para
86       considerarlos parte de la misma cadena.
87
88     Retorna:
89     - df: DataFrame actualizado con las columnas 'chain', '
90       destination' y 'dist'.
91     - trips_amount: Numero total de viajes procesados.
92     """
93     chain = 0
94     previous_time = None
95     trips_amount = 0
96     for idx,row in path.iterrows():
97         if previous_time is not None and row.HORA -
98             previous_time <= max_time and row.dist <=
99             max_distance :
100             chain = 1
101         else:
102             chain = 0
103             trips_amount += 1
104         previous_time = row.HORA
105         df.at[row.name, 'chain'] = chain
106         df.at[row.name, 'destination'] = row.destination

```

```
99     df.at[row.name, 'dist'] = row.dist
100     return df, trips_amount
101
102 def generate_path_combinations(df):
103     """
104     Genera combinaciones de rutas basadas en los datos de
105     un DataFrame agrupados por día y hora.
106
107     Parametros:
108     - df: DataFrame que contiene los datos de los viajes.
109       Se espera que incluya las columnas:
110       - 'FECHA': Fecha para agrupar los viajes por día.
111       - 'HORA': Hora asociada a cada viaje.
112
113     Retorna:
114     - Una lista de rutas optimizadas por día, donde cada
115       ruta es una lista de combinaciones de viajes.
116     """
117     df['FECHA_AJUSTADA'] = df.apply(lambda x: x['FECHA'] - pd
118                                     .to_timedelta((1 if x['HORA'].dt.hour < 5 else 0),
119                                     unit='D'))
120     df_grouped_by_day = df.groupby(df['FECHA_AJUSTADA'].dt.
121                                   date)
122     paths = []
123     for name, day_group in df_grouped_by_day:
124         if len(day_group) < 2:
125             continue
126         df_grouped = day_group.groupby(['HORA'])
127         paths_per_day = []
128         i = -1
129         for name, group in df_grouped:
130             i += 1
131             if len(group) < 2:
132                 if len(paths_per_day) == 0:
133                     paths_per_day.append([group.iloc[0]])
134                 else:
135                     for element in paths_per_day:
136                         element.append(group.iloc[0])
137             else:
```

```

132     temp_paths = []
133     ids = [i for i in range(len(group))]
134     permutations = list(itertools.permutations(ids))
135     for permutation in permutations:
136         temp_path = []
137         for idx in permutation:
138             temp_path.append(group.iloc[idx])
139         temp_paths.append(temp_path)
140     new_paths = []
141     if len(paths_per_day) == 0:
142         paths_per_day = temp_paths.copy()
143     else:
144         for element in paths_per_day:
145             for path in temp_paths:
146                 paths_aux = element.copy()
147                 paths_aux.extend(path)
148                 new_paths.append(paths_aux)
149         paths_per_day = new_paths.copy()
150     paths_per_day = sorted(paths_per_day, key=lambda x:
151                             calculate_path_weight(x, round_trip=len(
152                                 df_grouped) == i))
153     if len(paths_per_day) > 3:
154         paths_per_day = paths_per_day[0:3]
155     paths.append(paths_per_day[0])
156     return paths
157
158 def trip_chain(df, dist, time):
159     """
160     Procesa un DataFrame de viajes para generar cadenas de
161     viajes basadas en distancias y tiempos maximos.
162
163     Parametros:
164     - df: DataFrame que contiene los datos de los viajes,
165         incluyendo informacion de paradas y tarjetas.
166     - dist: Distancia maxima (en kilometros) para
167         considerar que dos viajes estan conectados.
168     - time: Tiempo maximo (en horas) entre viajes para
169         considerarlos parte de la misma cadena.

```

```

165     Retorna:
166     - No retorna un valor explicito, pero genera un archivo
167       CSV con los datos procesados, incluyendo
168       las columnas de origen, destino y distritos asociados
169       .
170     """
171     df_aux = df.copy()
172     trips_amount = 0
173     print('distance: ', dist, 'time: ', time)
174     for _, trip in tqdm(trips_per_card_df.iterrows(), total=
175         len(trips_per_card_df)):
176         shortest_path_of_trips = generate_path_combinations(
177             df_aux[df_aux['NROTARJETA'] == trip.NROTARJETA])
178         for shortest_path_of_trip in shortest_path_of_trips:
179             parsed_path = add_destination_to_path(
180                 shortest_path_of_trip)
181             parsed_path = pd.DataFrame(parsed_path)
182             df_aux, trips_amount_aux = add_destination_to_trips(
183                 df_aux, parsed_path, dist, time)
184             trips_amount += trips_amount_aux
185     df_aux.dropna(subset=['destination'], inplace=True)
186     df_aux.rename(columns={'STATION_ID': 'STATION_ORIGIN', '
187         destination': 'STATION_DESTINATION'}, inplace=True)
188     df_aux['DISTRICT_ORIGIN'] = df_aux['STATION_ORIGIN'].map(
189         lambda x: full_stations[full_stations['STATION_ID'] ==
190             x]['STATION_DISTRICT'].values[0])
191     df_aux['DISTRICT_DESTINATION'] = df_aux['
192         STATION_DESTINATION'].map(lambda x: full_stations[
193             full_stations['STATION_ID'] == x]['STATION_DISTRICT'].
194         values[0])
195     df_aux.drop(columns=['dist'], inplace=True)
196     df_aux.to_csv('./datasets/
197         dataset_with_destinations_graph_'+str(dist)+'_'+str(
198             time)+'.csv', index=False)
199     print('trips amount: ', trips_amount)

```

BIBLIOGRAFÍA

- Ait-Ali, A. and Eliasson, J. (2022). The value of additional data for public transport origin–destination matrix estimation. *Public Transport*, 14:419–439.
- Ait Ali, A., Warg, J., and Eliasson, J. (2020). Pricing commercial train path requests based on societal costs. *Transportation Research Part A: Policy and Practice*, 132:452–464.
- Ali, A., Kim, J., and Lee, S. (2015). Travel behavior analysis using smart card data. *KSCE Journal of Civil Engineering*, 20(4):1532–1539.
- Alsger, A., Assemi, B., Mesbah, M., and Ferreira, L. (2016). Validating and improving public transport origin–destination estimation algorithm using smart card fare data. *Transportation Research Part C: Emerging Technologies*, 68:490–506.
- Alsger, A., Tavassoli, A., Mesbah, M., Ferreira, L., and Hickman, M. (2018). Public transport trip purpose inference using smart card fare data. *Transportation Research Part C: Emerging Technologies*, 87:123–137.
- Alsger, A. A., Mesbah, M., Ferreira, L., and Safi, H. (2015). Use of smart card fare data to estimate public transport origin–destination matrix. *Transportation Research Record: Journal of the Transportation Research Board*, 2535(1):88–96.
- Bagchi, M. and White, P. (2005). The potential of public transport smart card data. *Transport Policy*, 12(5):464–474. Road User Charging: Theory and Practices.
- Barabino, B., Coni, M., Di Francesco, M., Obino, A., and Ventura, R. (2024). Origin–destination matrices from smartphone apps for bus networks. *Public Transport*, 16(2):505–549.
- Barry, J. J., Freimer, R., and Slavin, H. (2009). Use of entry-only automatic fare collection data to estimate linked transit trips in new york city. *Transportation Research Record: Journal of the Transportation Research Board*, 2112(1):53–61.

- Barry, J. J., Newhouser, R., Rahbee, A., and Sayeda, S. (2002). Origin and destination estimation in new york city with automated fare system data. transportation research record. *Transportation Research Record*, 1817:183–187.
- Bertin, J. (1983). *Semiology of graphics*. University of Wisconsin press.
- Blum, J. J., Sridhar, A., and Mathew, T. V. (2010). Origin–destination matrix generation from boarding–alighting and household survey data. *Transportation Research Record: Journal of the Transportation Research Board*, 2183(1):1–8.
- Chu, K. K. A. and Chapleau, R. (2008). Enriching archived smart card transaction data for transit demand modeling. *Transportation Research Record: Journal of the Transportation Research Board*, 2063(1):63–72.
- Cui, A. (2006). Bus passenger origin–destination matrix estimation using automated data collection systems. Master’s thesis, Massachusetts Institute of Technology.
- DeBoer, M. (2015). Understanding the heat map. *Cartographic perspectives*, (80):39–43.
- Ge, Q. and Fukuda, D. (2016). Updating origin–destination matrices with aggregated data of gps traces. *Transportation Research Part C: Emerging Technologies*, 69:291–312.
- Geng, L. and Zhang, K. (2022). Correlation of road network structure and urban mobility intensity: An exploratory study using geo-tagged tweets. *ISPRS International Journal of Geo-Information*, 12(1):7.
- Gordon, J. B., Koutsopoulos, H. N., Wilson, N. H. M., and Attanucci, J. P. (2013). Automated inference of linked transit journeys in london using fare–transaction and vehicle location data. *Transportation Research Record: Journal of the Transportation Research Board*, 2343(1):17–24.
- Hazelton, M. L. (2010). Statistical inference for transit system origin–destination matrices. *Technometrics*, 52(2):221–230.
- He, L., Nassir, N., Trépanier, M., and Hickman, M. (2015). *Validating and calibrating a destination estimation algorithm for public transport smart card fare collection systems*, volume 52. CIRRELT.
- Hofmann, M. and O’Mahony, M. (2005). Transfer journey identification and analyses from electronic fare collection data. In *Proceedings. 2005 IEEE Intelligent Transportation Systems, 2005.*, pages 34–39.

- Hörcher, D., Graham, D. J., and Anderson, R. J. (2017). Crowding cost estimation with large scale smart card and vehicle location data. *Transportation Research Part B: Methodological*, 95:105–125.
- Hu, L., Wen, Z., Wang, J., and Hu, J. (2022). Spatial interaction analysis of shared bicycles mobility regularity and determinants: A case study of six main districts, beijing. *ISPRS International Journal of Geo-Information*, 11(9):477.
- Huang, D., Yu, J., Shen, S., Li, Z., Zhao, L., and Gong, C. (2020). A method for bus od matrix estimation using multisource data. *Journal of Advanced Transportation*, 2020:1–13.
- Keller, C., Glück, F., Gerlach, C. F., and Schlegel, T. (2022). Investigating the potential of data science methods for sustainable public transport. *Sustainability*, 14(7):4211.
- Kumar, P., Khani, A., and He, Q. (2018). A robust method for estimating transit passenger trajectories using automated data. *Transportation Research Part C: Emerging Technologies*, 95:731–747.
- Liu, J. and Zhou, X. (2019). Observability quantification of public transportation systems with heterogeneous data sources: An information-space projection approach based on discretized space-time network flow models. *Transportation Research Part B: Methodological*, 128:302–323.
- MacEachren, A. M. and DiBiase, D. (1991). Animated maps of aggregate data: Conceptual and practical problems. *Cartography and Geographic Information Systems*, 18(4):221–229.
- Meyer, M. D. and Institute of Transportation Engineers (2016). *Transportation planning handbook*.
- Mohammed, M. and Oke, J. (2023). Origin-destination inference in public transportation systems: A comprehensive review. *International Journal of Transportation Science and Technology*, 12:315–328.
- Molina-Campoverde, J. J., Rivera-Campoverde, N., Molina Campoverde, P. A., and Bermeo Naula, A. K. (2024). Urban mobility pattern detection: Development of a classification algorithm based on machine learning and gps. *Sensors*, 24(12):3884.
- MOSALLANEJAD, M., SOMENAHALLI, S., VIJ, A., and MILLS, D. (2019). An approach to distinguish destination from the alighting stop based on fare data. *Journal of the Eastern Asia Society for Transportation Studies*, 13:1348–1360.

- Mukhiya, S. K. and Ahmed, U. (2020). *Hands-On Exploratory Data Analysis with Python*.
- Munizaga, M., Devillaine, F., Navarrete, C., and Silva, D. (2014). Validating travel behavior estimated from smartcard data. *Transportation Research Part C: Emerging Technologies*, 44:70–79.
- Munizaga, M. A. and Palma, C. (2012). Estimation of a disaggregate multimodal public transport origin–destination matrix from passive smartcard data from santiago, chile. *Transportation Research Part C: Emerging Technologies*, 24:9–18.
- Nassir, N., Khani, A., Lee, S. G., Noh, H., and Hickman, M. (2011). Transit stop-level origin–destination estimation through use of transit schedule and automated data collection system. *Transportation Research Record: Journal of the Transportation Research Board*, 2263(1):140–150.
- Netek, R., Brus, J., and Tomecka, O. (2019). Performance testing on marker clustering and heatmap visualization techniques: a comparative study on javascript mapping libraries. *ISPRS international journal of geo-information*, 8(8):348.
- Nunes, A. A., Galvao Dias, T., and Falcao e Cunha, J. (2016). Passenger journey destination estimation from automated fare collection system data using spatial validation. *IEEE Transactions on Intelligent Transportation Systems*, 17(1):133–142.
- Pelletier, M.-P., Trépanier, M., and Morency, C. (2011). Smart card data use in public transit: A literature review. *Transportation Research Part C: Emerging Technologies*, 19(4):557–568.
- Seaborn, C., Attanucci, J., and Wilson, N. H. M. (2009). Analyzing multimodal public transport journeys in london with smart card fare payment data. *Transportation Research Record: Journal of the Transportation Research Board*, 2121(1):55–62.
- Silva, P., Maças, C., Correia, J., Machado, P., and Polisciuc, E. (2022). Antenna: A tool for visual analysis of urban mobility based on cell phone data. In *VISIGRAPP (3: IVAPP)*, pages 88–100.
- Słomska-Przech, K., Panecki, T., and Pokojski, W. (2021). Heat maps: Perfect maps for quick reading? comparing usability of heat maps with different levels of generalization. *ISPRS International Journal of Geo-Information*, 10(8).
- Trépanier, M., Tranchant, N., and Chapleau, R. (2007). Individual trip destination estimation in a transit smart card automated fare collection system. *Journal of Intelligent Transportation Systems*, 11(1):1–14.

- Wang, H. and Zhang, X. N. (2011). Estimation of origin-destination matrix with tolling data. *Applied Mechanics and Materials*, 50:239–244.
- Wang, M.-H., Schrock, S. D., Vander Broek, N., and Mulinazzi, T. (2013). Estimating dynamic origin-destination data and travel demand using cell phone network data. *International Journal of Intelligent Transportation Systems Research*, 11:76–86.
- Wang, W., Attanucci, J., and Wilson, N. (2011). Bus passenger origin-destination estimation and related analyses using automated data collection systems. *Journal of Public Transportation*, 14(4):131–150.
- Wu, Y., Shen, T., Li, Y., Kita, M., Yoshida, Y., Sho, K., Yuan, L., Yu, L., Jiang, R., and Long, Y. (2025). Exploring urban railway station-based attractiveness considering demographic-specific demands: Case study of odakyu line, japan. *Cities*, 158:105620.
- Yan, F., Yang, C., and Ukkusuri, S. V. (2019). Alighting stop determination using two-step algorithms in bus transit systems. *Transportmetrica A: Transport Science*, 15(2):1522–1542.
- Yap, M., Cats, O., and Van Arem, B. (2020). Crowding valuation in urban tram and bus transportation based on smart card data. *Transportmetrica A: Transport Science*, 16(1):23–42.
- Yap, M., Cats, O., van Oort, N., and Hoogendoorn, S. (2017). A robust transfer inference algorithm for public transport journeys during disruptions. *Transportation Research Procedia*, 27:1042–1049.
- Zannat, K. E. and Choudhury, C. F. (2019). Emerging big data sources for public transport planning: A systematic review on current state of art and future research directions. *Journal of the Indian Institute of Science*, 99(4):601–619.
- Zhao, F. and Gaschler, R. (2022). Graph schema and best graph type to compare discrete groups: Bar, line, and pie. *Frontiers in Psychology*, 13.
- Zhao, J. (2004). The planning and analysis implications of automated data collection systems : rail transit od matrix inference and path choice modeling examples. Master’s thesis, Massachusetts Institute of Technology. Department of Civil and Environmental Engineering.
- Zúñiga, F., Muñoz, J. C., and Giesen, R. (2020). Estimation and prediction of dynamic matrix travel on a public transport corridor using historical data and real-time information. *Public Transport*, 13(1):59–80.

