

- 1 Descripción del Proyecto y Caso de Negocio
- 2 Fuentes de Datos
- 3 Carga y Preparación de Datos
- 4 Limpieza de Datos, Renombrado y Creación de Variables
- 5 Análisis Exploratorio de Datos (EDA)
- 6 Panel Interactivo en Tableau
- 7 Conclusión

Movilidad Urbana en Buenos Aires: Análisis con R y Tableau (Ecobici 2024)

Code ▾

Emilio Nahuel Pattini

2026-01-16

1 Descripción del Proyecto y Caso de Negocio

Este proyecto analiza los datos de viajes de todo el año 2024 del sistema público de bicicletas compartidas de Buenos Aires, **Ecobici**, para descubrir patrones de uso, picos de demanda y oportunidades para mejorar la movilidad urbana sostenible.

A medida que las ciudades en todo el mundo adoptan transportes más ecológicos, comprender el comportamiento en el uso de bicicletas compartidas es clave para optimizar la ubicación de estaciones, la gestión de la flota y la reducción de la congestión y las emisiones.

Utilizando el conjunto de datos más reciente y completo (~3.56 millones de viajes), este análisis ofrece información accionable para la planificación de ciudades inteligentes y estrategias de movilidad sostenible a nivel global.

Objetivos Clave:

- Identificar las horas, días y patrones estacionales de mayor uso de bicicletas.
- Mapear las estaciones y barrios más populares de origen y destino.
- Explorar la duración de los viajes y la demografía de los usuarios (cuando esté disponible).
- Proporcionar recomendaciones basadas en datos para mejorar el sistema.

Este informe reproducible demuestra habilidades en limpieza de datos, análisis exploratorio, visualización y narración de negocio utilizando R y Tableau.

2 Fuentes de Datos

Este análisis utiliza datos abiertos oficiales del Gobierno de la Ciudad de Buenos Aires.

1. Viajes Ecobici 2024

- Archivo: `ecobici_recorridos_realizados_2024.csv`
- Fuente: <https://data.buenosaires.gob.ar/dataset/bicicletas-publicas> (<https://data.buenosaires.gob.ar/dataset/bicicletas-publicas>)
- Tamaño: ~3.56 millones de filas (año completo 2024)
- Columnas clave (tras renombrar al inglés):
 - `trip_id` : ID único del viaje
 - `duration_seconds` : Duración del viaje (segundos)
 - `start_datetime` / `end_datetime` : Hora de inicio y fin
 - `origin_station_id` / `origin_station_name` : Estación de origen
 - `origin_longitude` / `origin_latitude` : Coordenadas de origen
 - `destination_station_id` / `destination_station_name` : Estación de destino
 - `destination_longitude` / `destination_latitude` : Coordenadas de destino
 - `user_id` , `bike_model` , `gender`

2. Estaciones Ecobici (Nuevo Sistema)

- Archivo: `estaciones_ecobici_nuevo_sistema.csv`

- Fuente: <https://data.buenosaires.gob.ar/dataset/estaciones-bicicletas-publicas> (<https://data.buenosaires.gob.ar/dataset/estaciones-bicicletas-publicas>)
- Nota: Descargado pero no utilizado debido a incompatibilidad total en los IDs de estaciones (0% unión exitosa). El dataset de viajes ya incluye nombres y coordenadas completas de estaciones para cada viaje, asegurando cobertura total.

Nota de Calidad de Datos

Se intentó unir información de barrio, comuna y ubicación desde el archivo de estaciones, pero falló (0% coincidencia). El análisis se basa completamente en los nombres y coordenadas integrados en el archivo de viajes. En futuras actualizaciones se podrán añadir insights a nivel de barrio con un dataset de estaciones sincronizado.

Todos los datos son públicos y abiertos. Los datos de 2025 aún no están completamente disponibles a enero de 2026. Los nombres de columnas se han estandarizado al inglés para mayor claridad y legibilidad internacional.

3 Carga y Preparación de Datos

Los datos brutos de viajes Ecobici (2024) se cargan silenciosamente. Los nombres originales de las columnas están en español, tal como fueron provistos por el portal de datos abiertos.

4 Limpieza de Datos, Renombrado y Creación de Variables

En esta etapa:

- Renombro todas las columnas al inglés para mantener consistencia y legibilidad internacional.
- Limpio y normalizo los nombres de las estaciones.
- Creo nuevas variables derivadas (duración en minutos, hora de inicio, día de la semana, indicador de fin de semana, mes).
- Filtro viajes inválidos o atípicos (por ejemplo, viajes menores a 1 minuto, mayores a 24 horas o con valores clave faltantes).

Después de la limpieza, muestro la estructura y las estadísticas básicas del dataset final.

Show

```
## Rows: 3,234,209
## Columns: 22
## $ trip_id
## $ duration_seconds
## $ start_datetime
## $ origin_station_id
## $ origin_station_name
## $ origin_station_address
## $ origin_longitude
## $ origin_latitude
## $ end_datetime
## $ destination_station_id
## $ destination_station_name
## $ destination_station_address
## $ destination_longitude
## $ destination_latitude
## $ user_id
## $ bike_model
## $ gender
## $ duration_minutes
## $ hour_start
## $ weekday
## $ day_type
## $ month
```

Show

```
##      Min.   1st Qu.   Median     Mean   3rd Qu.   Max.
## 1.000    9.867   16.100   21.748   25.850 1439.133
```

Resultados Clave Después de la Limpieza

- El dataset final contiene **3.234.209 viajes válidos** (tras eliminar registros inválidos).
- Las duraciones de los viajes en minutos tienen una mediana de ~16,1 minutos y un promedio de ~21,7 minutos, lo cual es típico en sistemas de bicicletas compartidas urbanas (trayectos cortos con algunos recorridos más largos).

- La mayoría de los viajes están entre ~10 y 26 minutos (entre el primer y tercer cuartil).

Resumen de Reducción de Datos

Etapa	Número de Viajes	% del Original
Original (crudo)	3.559.284	100%
Después de limpieza	3.234.209	~90,8%
Eliminados (inválidos)	~325.075	~9,1%

Este proceso de limpieza eliminó aproximadamente 325.000 registros inválidos o atípicos (~9% de los datos originales), lo cual es normal y esperado en datasets públicos a gran escala como este. Los datos restantes están limpios, consistentes y listos para un análisis más profundo.

5 Análisis Exploratorio de Datos (EDA)

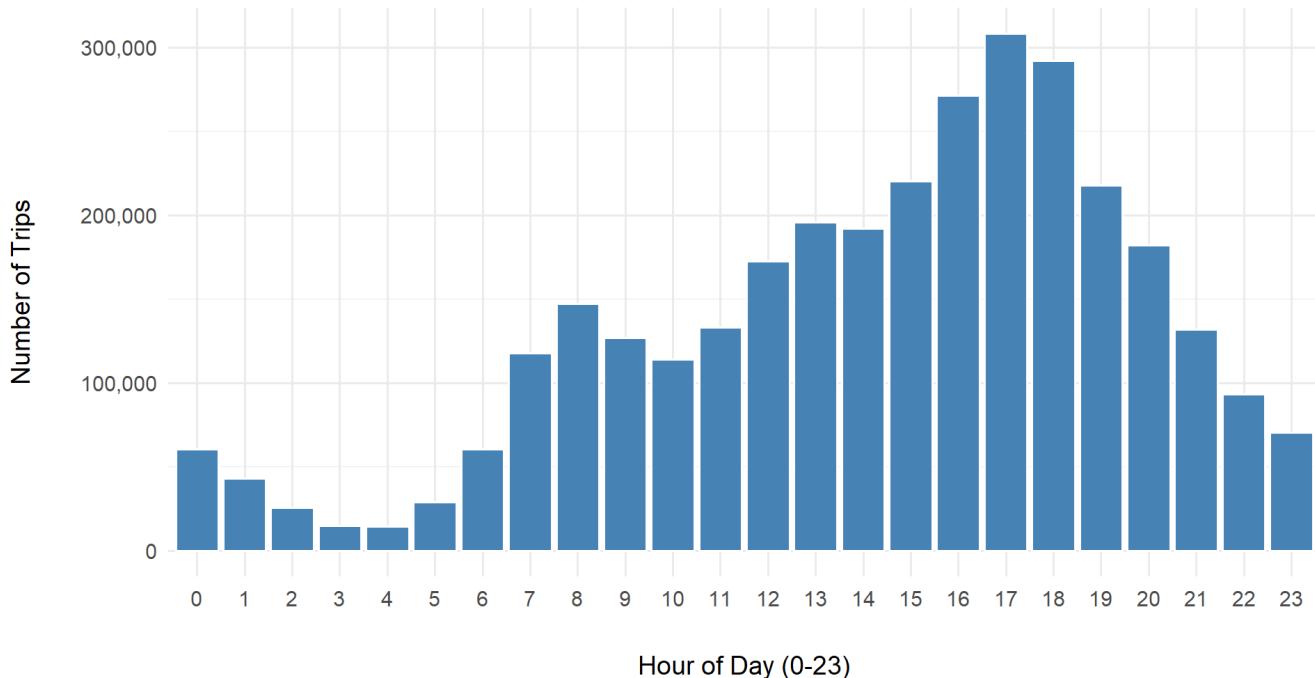
5.1 Horas Pico de Uso

Este gráfico de barras muestra los viajes iniciados por hora del día (0–23), destacando los patrones diarios de uso y las horas de mayor demanda.

Show

Number of Trips by Hour of Day (2024)

Clear morning and evening commute peaks



Principales Conclusiones

- El pico vespertino (16–18 hs) domina con **271.500–308.297 viajes por hora** — fuerte patrón de regreso a casa.
- Pico secundario en la mañana (7–9 hs): **117.583–147.119 viajes** — desplazamiento de entrada al trabajo/estudio.
- Mediodía (12–15 hs) estable en **172.710–220.351 viajes** — mezcla de trámites y ocio.

Recomendaciones Potenciales

- Aumentar la disponibilidad de bicicletas durante el pico vespertino (16–18 hs) para satisfacer la mayor demanda.
- Explorar incentivos para la devolución de bicicletas en períodos de alta demanda.

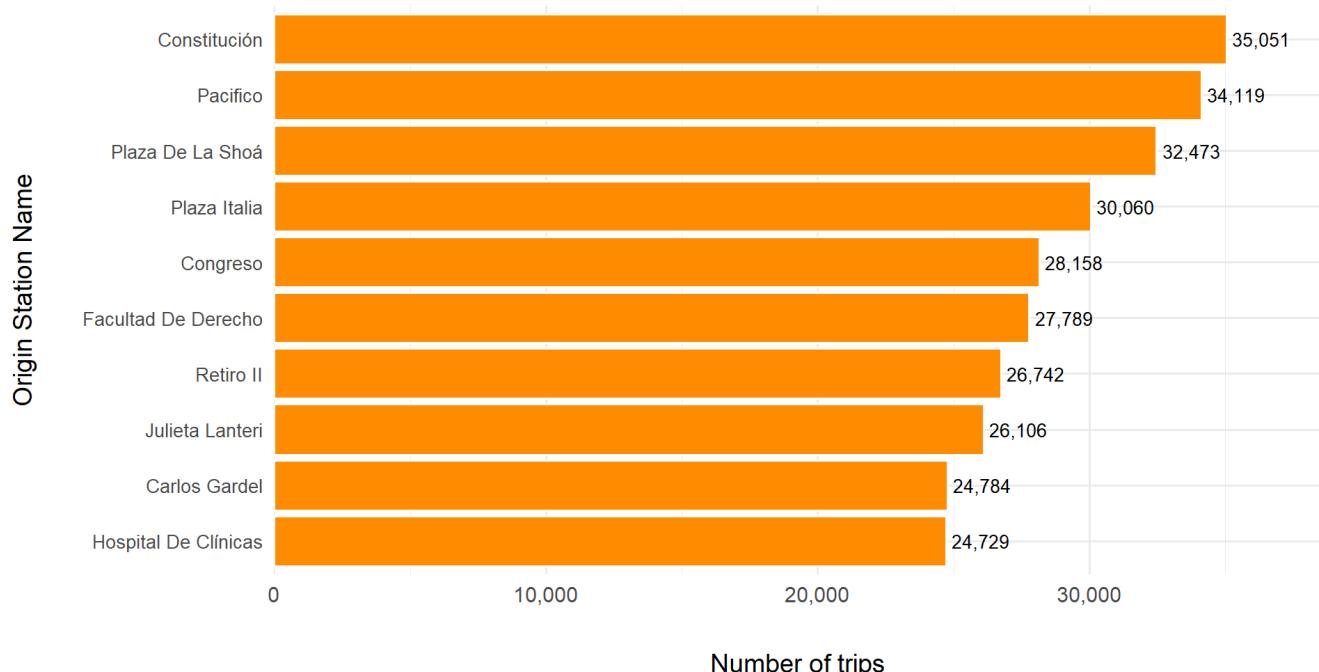
5.2 Top 10 Estaciones de Origen Más Populares

Este gráfico de barras horizontal muestra las 10 estaciones de origen más utilizadas en 2024 (nombres normalizados), revelando los puntos de partida de mayor demanda.

Show

Top 10 Origin Stations by Number of Trips (2024)

Stations with highest departure demand



Conclusiones Clave

- Estación principal: **Constitución** (35.051 viajes), seguida por **Pacífico** (34.119) y **Plaza De La Shoá** (32.473).
- Las 10 estaciones principales representan aproximadamente el 8,9% de todos los viajes — mostrando una demanda concentrada en pocos puntos.
- Muchas de las estaciones más utilizadas se alinean con nodos de transporte o áreas centrales, reforzando el enfoque en los desplazamientos diarios.

Recomendaciones Potenciales

- Priorizar el suministro de bicicletas y el reequilibrio cerca de las principales estaciones de origen (ej. Constitución, Pacífico) durante las horas pico.
- Considerar incentivos para la devolución de bicicletas en estos puntos de partida concurridos.

Nota sobre el Enfoque del Análisis

Este análisis prioriza las **estaciones de origen** como el principal indicador de la demanda de salida (donde más se necesitan bicicletas).

Las estaciones de destino reflejan patrones de llegada y pueden explorarse en futuros análisis para estudiar flujos completos de viajes o el movimiento neto de bicicletas.

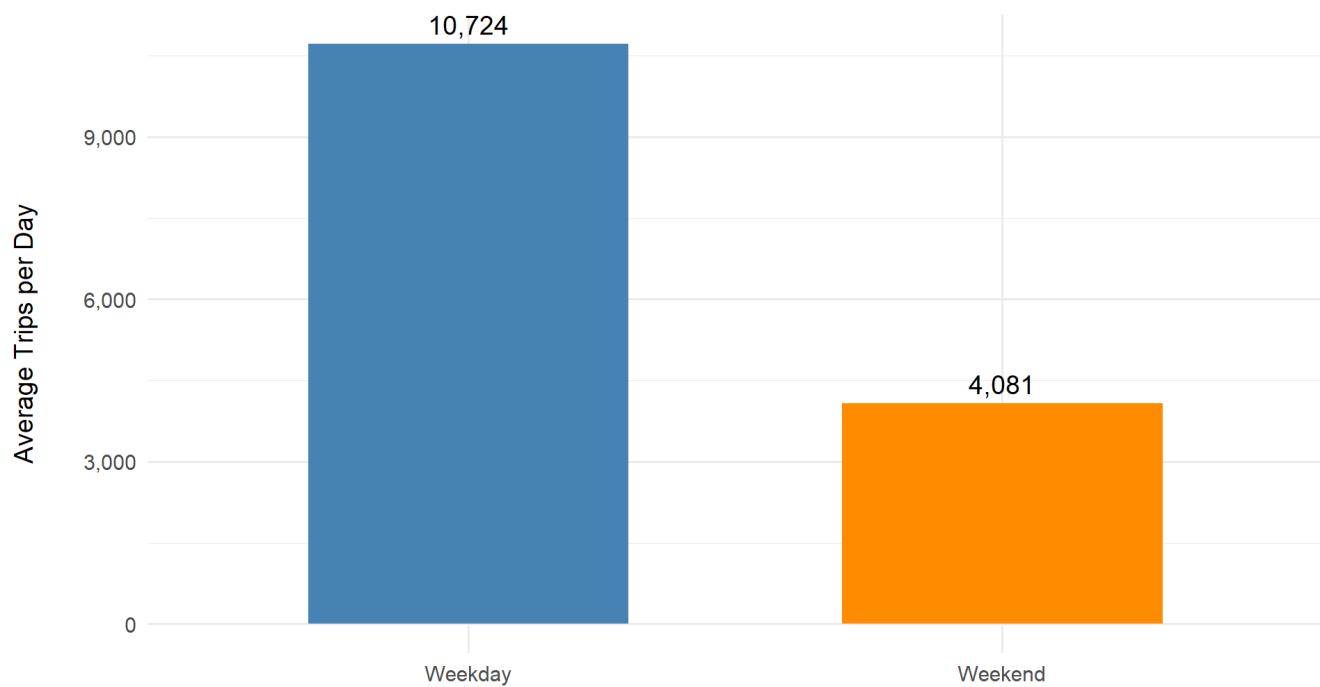
5.3 Promedio de Viajes Diarios: Días Laborales vs Fin de Semana (Normalizado)

Este gráfico de barras muestra el número promedio de viajes por día en días laborables frente a fines de semana (normalizado por días únicos para tener en cuenta que los días laborables son ~2,5× más numerosos).

 Show

Average trips per Day: Weekday vs Weekend (2024)

Normalized by number of unique days in each category



Conclusiones Clave

- El promedio de días laborables es significativamente mayor que el de fines de semana, confirmando el **desplazamiento diario** como principal caso de uso.
- El uso en fines de semana es menor pero igualmente considerable, lo que sugiere cierta **actividad recreativa o de ocio** los sábados y domingos.
- La relación entre días laborables y fines de semana (~2,6:1 por día) refuerza el patrón de desplazamiento observado en el análisis de horas pico.

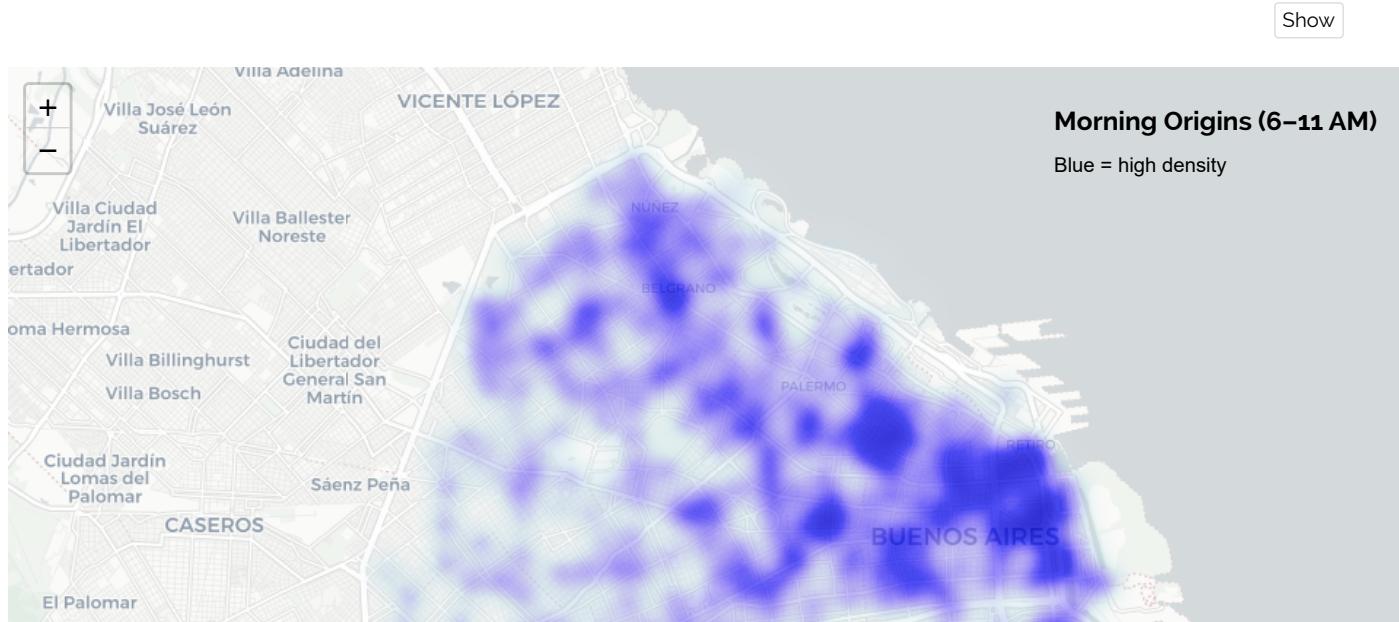
Recomendaciones Potenciales

- Priorizar la disponibilidad de bicicletas y el reequilibrio en días laborables durante las horas pico.
- Explorar promociones de fin de semana para impulsar el uso recreativo.

5.4 Mapas de Calor Espaciales de Orígenes y Destinos de Viajes

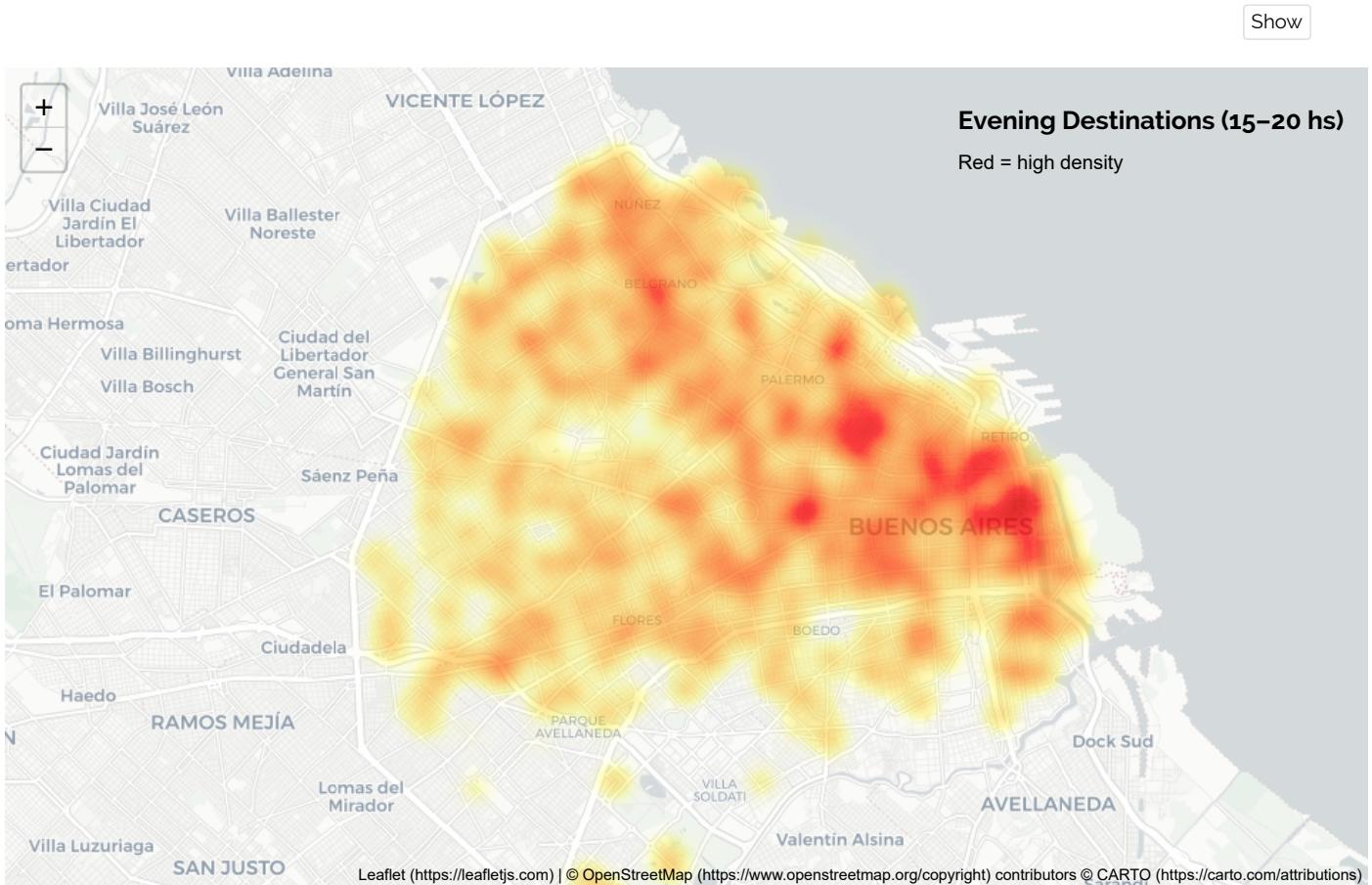
Estos mapas de calor interactivos visualizan la distribución geográfica de los **orígenes** (azul) y **destinos** (rojo) de los viajes en Buenos Aires (muestra 2024).

5.4.1 Mapa de Calor de Orígenes Matutinos (6–11 hs)





5.4.2 Mapa de Calor de Destinos Vespertinos (15-20 hs)



Conclusiones Clave – Flujos de Desplazamiento: Orígenes Matutinos vs Destinos Vespertinos

- Los orígenes matutinos (azul) muestran una alta concentración en áreas centrales y del norte (ej. Microcentro, Palermo, Recoleta), mientras que los destinos vespertinos (rojo) mantienen alta densidad en el centro pero con mayor dispersión por la ciudad.
- Este patrón indica un **flujo de desplazamiento** en el que las personas parten de zonas centrales/residenciales por la mañana y regresan a zonas residenciales más dispersas por la tarde.
- La fuerte saturación central en ambos períodos sugiere que el núcleo de la ciudad actúa como un gran hub tanto de partidas como de llegadas.

Recomendaciones Potenciales

- Priorizar la disponibilidad de bicicletas en **clusters centrales de alta densidad** durante las salidas matutinas y las llegadas vespertinas para satisfacer la demanda persistente en el centro.
- Usar un reequilibrio dinámico para redistribuir el exceso de bicicletas desde el centro (donde se acumulan durante el día) hacia áreas residenciales más dispersas en la tarde, asegurando disponibilidad para los viajes de regreso.

Nota sobre la muestra

Se utilizó una muestra aleatoria de 10.000 viajes para mejorar el rendimiento interactivo; el análisis del dataset completo muestra patrones espaciales idénticos.

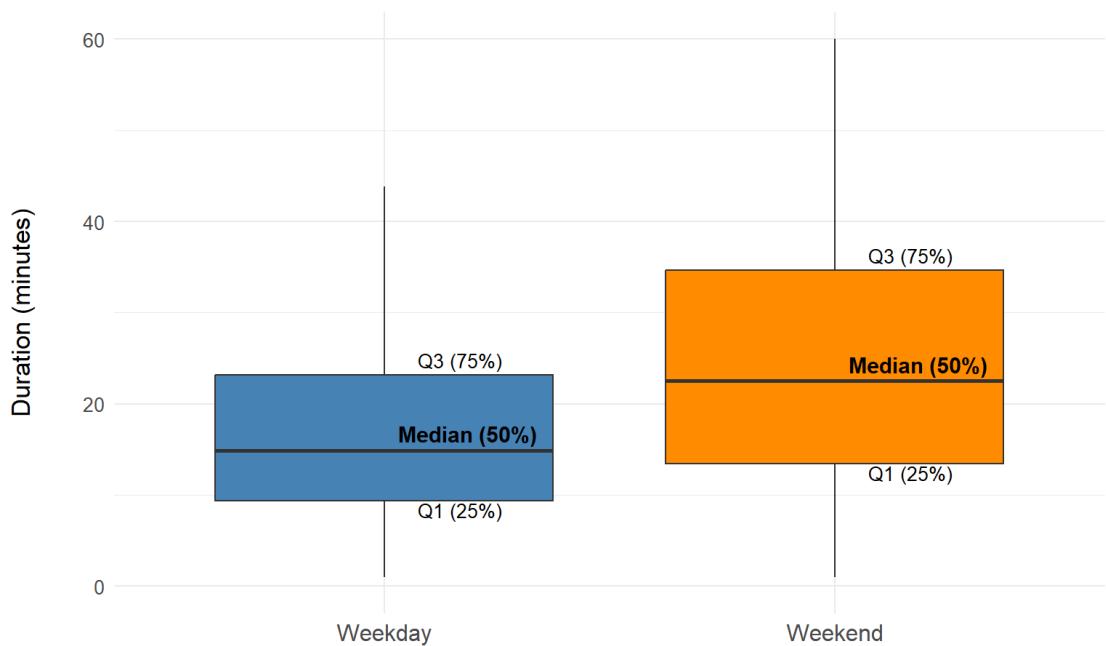
5.5 Análisis de Duración de Viajes

Este diagrama de caja muestra la distribución de las duraciones de los viajes en minutos según el tipo de día, revelando longitudes de uso típicas.

Show

Trip Duration Distribution by Day Type

Boxplot of duration in minutes (outliers removed for clarity)



Boxplot explanation: The box encloses the middle 50% of trips (from 25% to 75%).
The thick line inside is the median (50%). Whiskers show the typical range (excluding extreme outliers).

Conclusiones Clave

- Las duraciones de los viajes son generalmente cortas, con una mediana de ~16,1 minutos y un promedio de ~21,7 minutos – típico en sistemas de bicicletas compartidas urbanas (traslados rápidos o trámites).
- Los viajes en días laborables tienden a ser **ligevemente más cortos** (más enfocados en el desplazamiento), mientras que los viajes de fin de semana muestran **mayor variación** (probablemente incluyendo ocio o recorridos recreativos más largos).
- La mayoría de los viajes (75% o más) duran menos de ~26 minutos, lo que indica un uso eficiente y de corta distancia.

Recomendaciones Potenciales

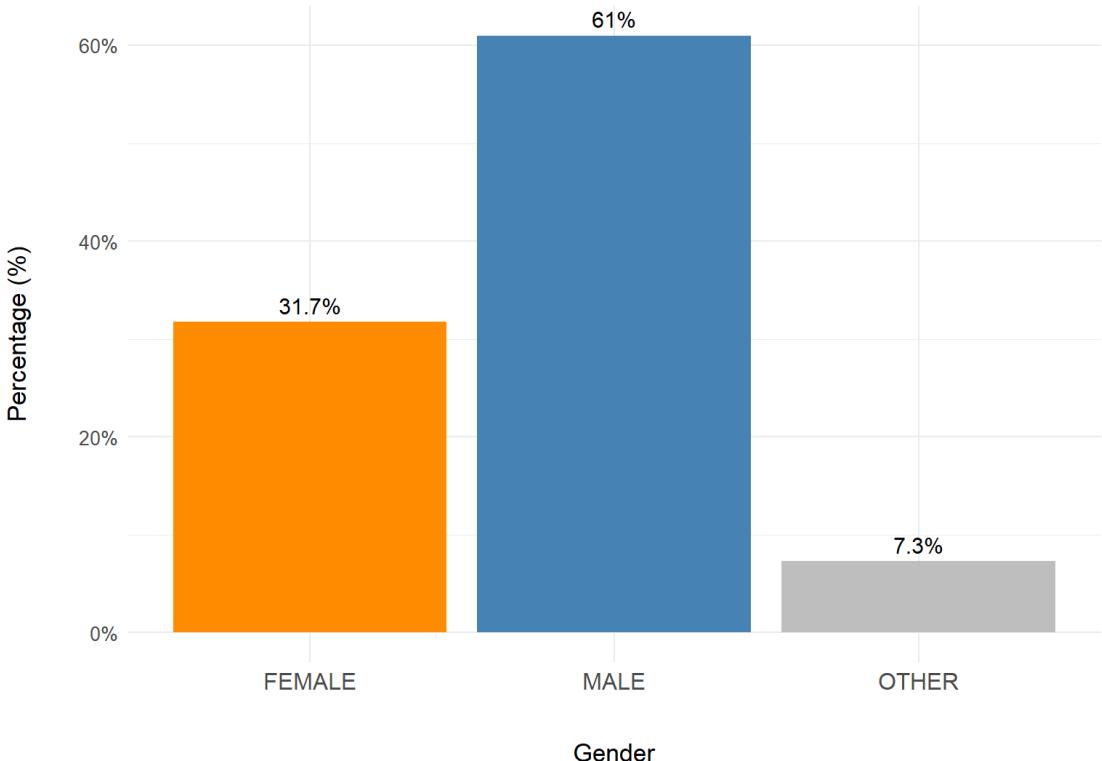
- Aprovechar el límite gratuito existente de 30 minutos en días laborables promoviendo viajes cortos y eficientes (ej. campañas que incentiven desplazamientos menores a 30 minutos para evitar cargos).
- Explorar la extensión del tiempo gratuito o del número de viajes en fines de semana para impulsar el uso recreativo, ya que los viajes de fin de semana muestran más variación y potencial para recorridos de ocio más largos.

5.6 Uso por Género

Este gráfico de barras muestra el porcentaje de viajes por género, proporcionando información demográfica.

Show

Percentage of Trips by Gender (2024)



Conclusiones Clave – Uso por Género

- Los usuarios masculinos representan la mayoría de los viajes (~60,8%), mientras que las usuarias femeninas constituyen ~31,6%.
- Un pequeño porcentaje (~7,3%) se clasifica como “Otro”, con valores faltantes insignificantes (0,3%).
- El desequilibrio de género sugiere oportunidades para aumentar la participación femenina mediante campañas específicas, mejoras de seguridad o marketing inclusivo.

Recomendaciones Potenciales

- Lanzar iniciativas para impulsar el uso femenino (ej. eventos exclusivos para mujeres, características de seguridad o promociones dirigidas a trabajadoras/estudiantes).
- Monitorear las tendencias de género a lo largo del tiempo para evaluar el impacto de los esfuerzos de inclusión.

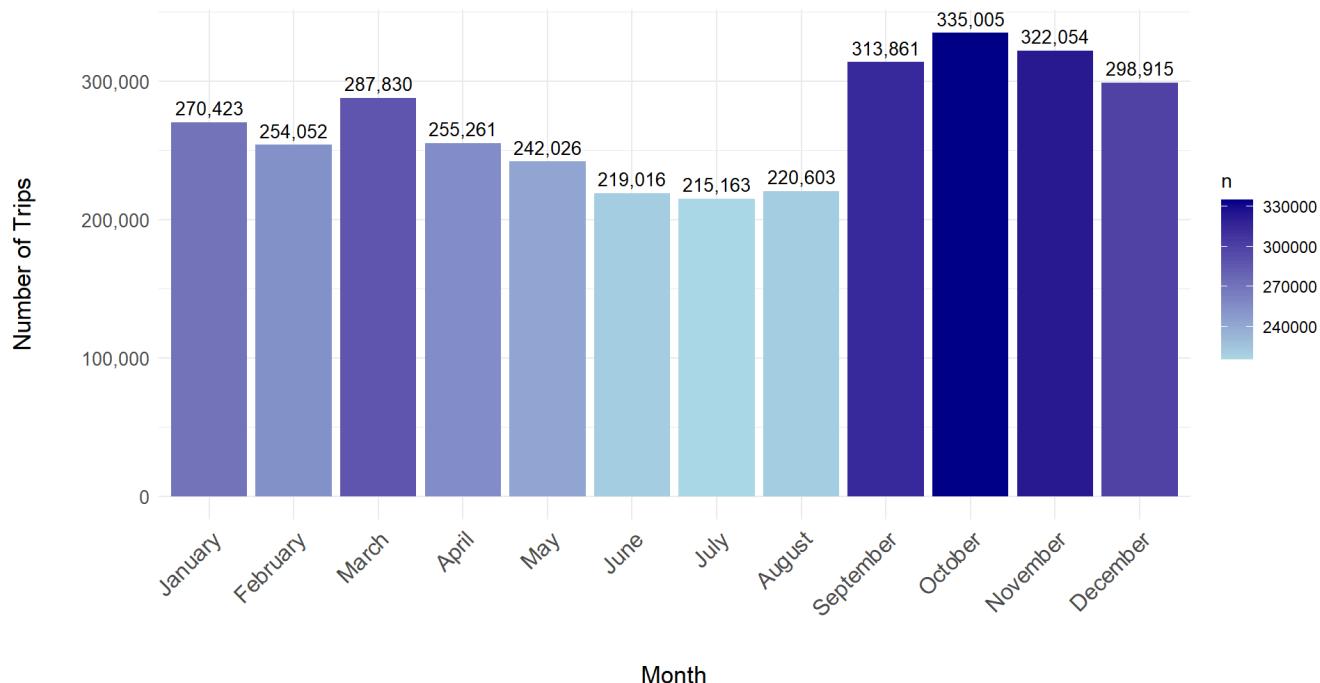
5.7 Viajes por Mes (Estacionalidad)

Este gráfico de barras muestra el número total de viajes por mes en 2024, revelando patrones estacionales de uso.

Show

Total Trips by Month (2024)

Seasonal usage patterns throughout the year



Conclusiones Clave

- El uso alcanza su pico en la **primavera tardía y el inicio del verano** (septiembre–diciembre), con los valores más altos en **octubre** (335.005 viajes) y **noviembre** (322.054 viajes).
- El menor uso ocurre en **invierno** (junio–agosto), con el mínimo en **julio** (215.163 viajes), probablemente debido al clima más frío en Buenos Aires.
- En general, existe un claro patrón estacional: mayor cantidad de viajes durante los meses cálidos (primavera tardía/verano temprano) y menor en invierno, típico de los sistemas de bicicletas compartidas al aire libre en el hemisferio sur.

Recomendaciones Potenciales

- Aumentar la disponibilidad de bicicletas y los esfuerzos de marketing durante la **temporada alta** (septiembre–diciembre) para aprovechar la mayor demanda.
- Considerar operaciones reducidas o incentivos específicos en los meses de invierno (junio–agosto) para mantener el compromiso durante los períodos de menor uso.

5.8 Flujo Neto por Estación (Orígenes vs Destinos)

Este gráfico de barras horizontal muestra el flujo neto de bicicletas por estación (destinos menos orígenes, es decir, llegadas menos partidas) en 2024.

Los valores positivos (azul) indican exceso de bicicletas (más llegadas, es decir, destinos).

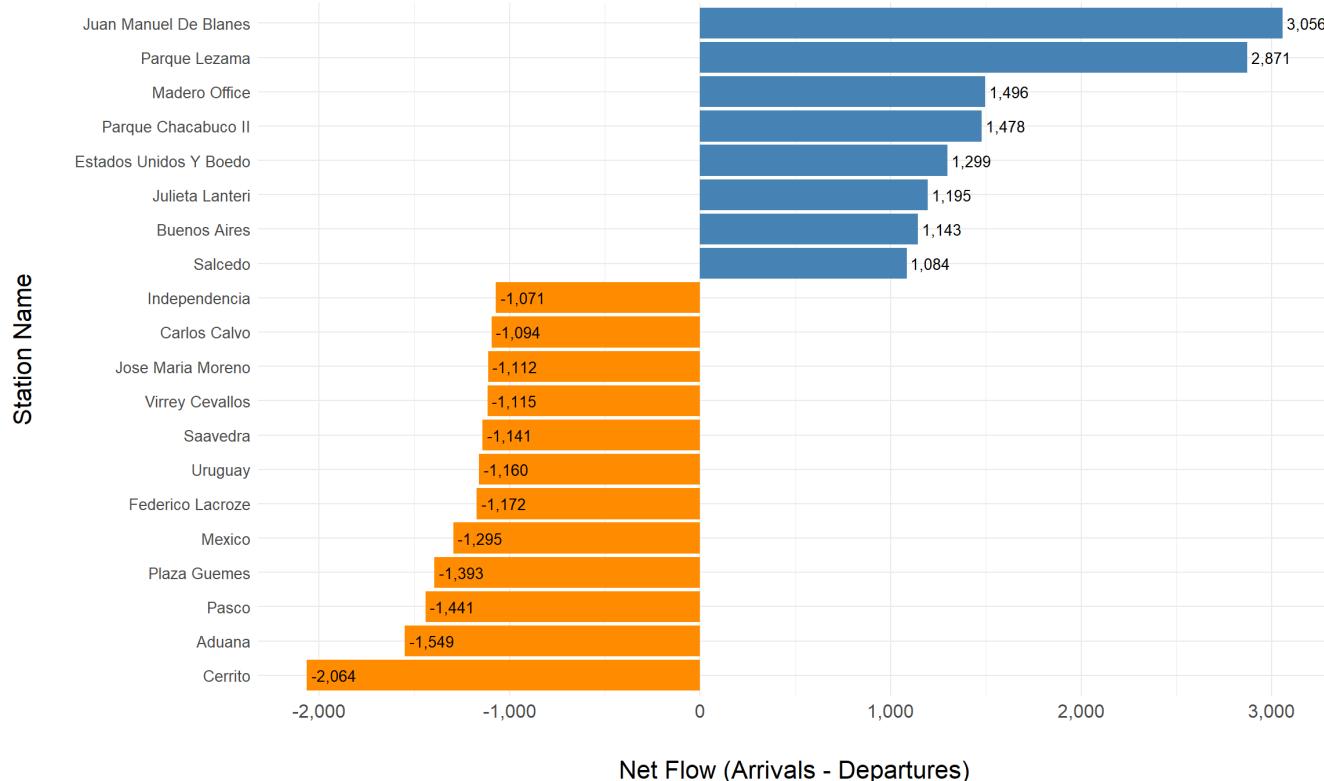
Los valores negativos (naranja) indican escasez de bicicletas (más partidas, es decir, orígenes).

Las estaciones están ordenadas desde el mayor flujo neto positivo (arriba) hasta el menor negativo (abajo).

Show

Net Bike Flow by Station (Top 20 Extremes)

Negative = excess departures (need addition), Positive = excess arrivals (need removal)



Conclusiones Clave

- Estaciones como **Juan Manuel De Blanes** (+3.056) y **Parque Lezama** (+2.871) tienen el mayor flujo neto positivo (exceso de llegadas, es decir, destinos) — acumulan bicicletas y requieren retiro.
- Estaciones como **Cerrito** (-2.064) y **Aduana** (-1.549) presentan el mayor flujo neto negativo (exceso de partidas, es decir, orígenes) — pierden bicicletas y necesitan reposición.
- Los extremos principales destacan estaciones que requieren reequilibrio frecuente para mantener la disponibilidad.

Recomendaciones Potenciales

- Priorizar el reequilibrio en estaciones con flujo neto positivo (retirar bicicletas en exceso).
- Priorizar el reequilibrio en estaciones con flujo neto negativo (agregar bicicletas).
- Usar modelos predictivos para anticipar el flujo neto diario según la hora del día y patrones históricos.

6 Panel Interactivo en Tableau

Para hacer el análisis más interactivo y exploratorio, se creó un panel en Tableau con el dataset limpio.

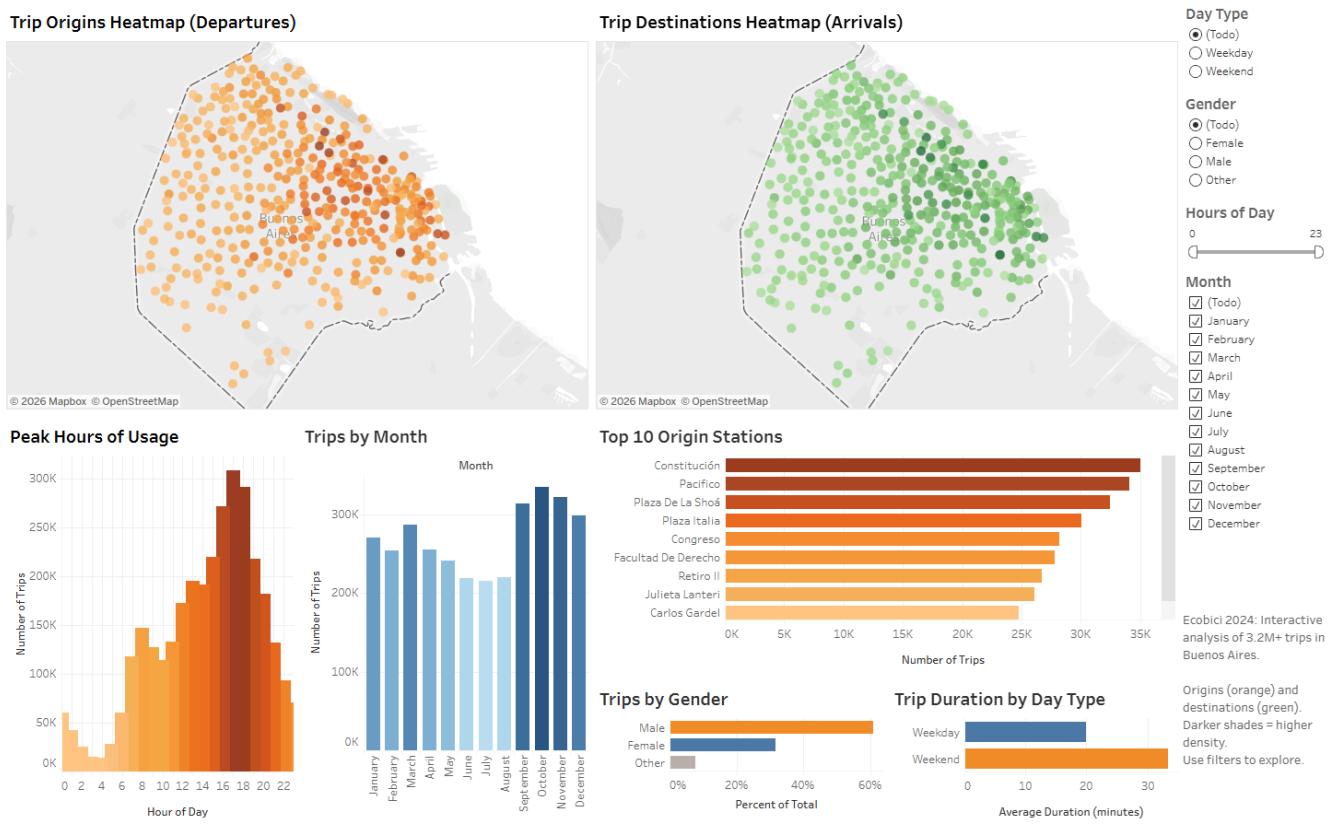
Aspectos Destacados del Panel

- Mapas de calor de orígenes y destinos (interactivos con filtros por hora del día, tipo de día, mes y género).
- Gráfico de barras de horas pico.
- Gráfico de barras de viajes por mes.
- Top 10 estaciones de origen.
- Diagrama de caja de duración de viajes por tipo de día.
- Gráfico de barras de uso por género.

Enlace al panel interactivo completo

Ver el Dashboard Ecobici 2024 en Tableau Public (https://public.tableau.com/views/Ecobici2024-InteractiveUrbanBike-SharingInsights/Ecobici2024-UrbanBike-SharingInsights?:language=es-ES&publish=yes&:sid=&:redirect=auth&:display_count=n&:origin=viz_share_link)

Captura de Pantalla



Ecobici 2024 Dashboard

7 Conclusión

Este proyecto analizó más de 3,2 millones de viajes de Ecobici en 2024 para descubrir patrones clave de uso en el sistema público de bicicletas compartidas de Buenos Aires.

Principales Hallazgos

- Fuerte enfoque en desplazamientos: pico vespertino (16–18 hs, hasta 308k viajes/hora) y pico secundario matutino (7–10 hs).
- Promedio en días laborables ~10.724 viajes/día vs ~4.081 en fines de semana (relación ~2,6:1).
- Principales estaciones de origen (Constitución 35k, Pacífico 34k) y mapas de calor espaciales muestran demanda concentrada en áreas centrales (Microcentro, Palermo, Recoleta).
- Los orígenes matutinos se agrupan en zonas residenciales/centrales, los destinos vespertinos se dispersan, confirmando el flujo bidireccional de desplazamientos.
- Duración de viajes: mediana ~16,1 min, promedio ~21,7 min — la mayoría menores a 30 min.
- Distribución por género: Hombres ~60,8%, Mujeres ~31,6%.
- Picos estacionales en primavera tardía/verano temprano (octubre 335k), mínimos en invierno (julio 215k).
- Extremos de flujo neto (ej. Juan Manuel De Blanes +3.056, Cerrito -2.064) destacan necesidades de reequilibrio.

Lecciones Aprendidas

- Las fuentes de datos abiertos pueden tener inconsistencias (ej. incompatibilidad de IDs de estaciones) — confiar en coordenadas integradas aseguró cobertura completa sin dependencias externas.
- Las técnicas de normalización (como promedios por día) son críticas para evitar conclusiones engañosas al comparar grupos de tamaños desiguales.
- El muestreo (10.000 viajes) logra un equilibrio práctico entre rendimiento interactivo y precisión analítica en visualizaciones a gran escala.
- La terminología consistente (ej. orígenes/destinos en lugar de mezclar con partidas/llegadas) y las explicaciones claras mejoran significativamente la legibilidad y la comprensión de los interesados.

Próximos Pasos

- Incorporar datos en tiempo real para el reequilibrio dinámico.

- Explorar demografía de usuarios (edad, género) y modelos predictivos para pronóstico de demanda.
- Expandir al dataset de 2025 cuando esté disponible.

Este informe reproducible demuestra habilidades en R (tidyverse, leaflet), limpieza de datos, EDA, visualización, dashboards interactivos (Tableau) e insights de negocio — listo para roles de analista de datos.

Ver el informe completo en línea (con mapas de calor interactivos):

Abrir el Informe Ecobici 2024 en mi sitio web (<https://www.emilionahuelpattini.com/es/data/data-analysis/projects/ecobici-2024/urban-mobility-analysis.html>)

(Versión interactiva con zoom, desplazamiento y tooltips — no requiere descarga.)

Panel Interactivo

Explora el dataset completo de manera interactiva en Tableau Public:

Ver el Dashboard Ecobici 2024 (https://public.tableau.com/views/Ecobici2024-InteractiveUrbanBike-SharingInsights/Ecobici2024-UrbanBike-SharingInsights?:language=es-ES&publish=yes&:sid=&:redirect=auth&:display_count=n&:origin=viz_share_link)

Descargar Informe en PDF

Para una versión imprimible:

Descargar PDF (<https://www.emilionahuelpattini.com/es/data/data-analysis/projects/ecobici-2024/urban-mobility-analysis.pdf>)

(Informe completo con todas las visualizaciones e insights — 12 páginas)