

1 Project Overview & Business Case

2 Data Sources

3 Data Loading & Preparation

4 Data Cleaning, Renaming & Feature Engineering

5 Exploratory Data Analysis (EDA)

6 Interactive Dashboard in Tableau

7 Conclusion

Code ▾

# Urban Mobility in Buenos Aires: Analysis with R & Tableau (Ecobici 2024)

Emilio Nahuel Pattini

2026-01-16

## 1 Project Overview & Business Case

This project analyzes full-year 2024 trip data from Buenos Aires' public bike-sharing system, **Ecobici**, to uncover usage patterns, peak demand, and opportunities for enhancing sustainable urban mobility.

As cities globally adopt greener transport, understanding bike-sharing behaviour is key to optimizing station placement, fleet management, and reducing congestion and emissions.

Using the most recent complete dataset (~3.56 million trips), this analysis delivers actionable insights for smart city planning and sustainable mobility strategies worldwide.

### Key Objectives:

- Identify peak hours, days, and seasonal patterns in bike usage.
- Map the most popular origin and destination stations and neighborhoods.
- Explore trip durations and user demographics (where available).
- Provide data-driven recommendations for system improvements.

This reproducible report demonstrates skills in data cleaning, exploratory analysis, visualization, and business storytelling using R and Tableau.

## 2 Data Sources

This analysis uses official open data from the Government of Buenos Aires.

### 1. Ecobici Trips 2024

- File: `ecobici_recorridos_realizados_2024.csv`
- Source: <https://data.buenosaires.gob.ar/dataset/bicicletas-publicas> (<https://data.buenosaires.gob.ar/dataset/bicicletas-publicas>)
- Size: ~3.56 million rows (complete year 2024)
- Key columns (after renaming to English):
  - `trip_id`: Unique trip ID
  - `duration_seconds`: Trip duration (seconds)
  - `start_datetime` / `end_datetime`: Start and end times
  - `origin_station_id` / `origin_station_name`: Origin station
  - `origin_longitude` / `origin_latitude`: Origin coordinates
  - `destination_station_id` / `destination_station_name`: Destination station
  - `destination_longitude` / `destination_latitude`: Destination coordinates
  - `user_id`, `bike_model`, `gender`

### 2. Ecobici Stations (New System)

- File: `estaciones_ecobici_nuevo_sistema.csv`
- Source: <https://data.buenosaires.gob.ar/dataset/estaciones-bicicletas-publicas> (<https://data.buenosaires.gob.ar/dataset/estaciones-bicicletas-publicas>)

- Note: Downloaded but not used due to complete mismatch in station IDs (0% successful join). The trips dataset already includes comprehensive station names and coordinates for every trip, ensuring full coverage.

### Data Quality Note

An attempt to join neighborhood, commune, and placement information from the stations file failed (0% match). The analysis relies entirely on the built-in station names and coordinates from the trips file. Neighborhood-level insights can be added in future updates with a synced stations dataset.

All data is public and open. 2025 data is not yet fully available as of January 2026. Column names have been standardized to English for clarity and international readability.

## 3 Data Loading & Preparation

The raw Ecobici trips data (2024) is loaded quietly. Original column names are in Spanish, as provided by the open data portal.

## 4 Data Cleaning, Renaming & Feature Engineering

In this step I:

- Rename all columns to English for consistency and international readability.
- Clean and normalize station names.
- Create new derived features (duration in minutes, start hour, weekday, weekend flag, month).
- Filter out invalid or outlier trips (e.g., trips shorter than 1 minute, longer than 24 hours, or missing key values).

After cleaning, I display the structure and basic statistics of the final dataset.

Show

```
## Rows: 3,234,209
## Columns: 22
## $ trip_id                               <dbl> 20428222, 20431744, 20424802, 20427241, 20...
## $ duration_seconds                      <dbl> 568, 1355, 680, 466, 1176, 1906, 695, 492, ...
## $ start_datetime                         <dttm> 2024-01-23 18:36:00, 2024-01-23 22:41:20, ...
## $ origin_station_id                     <dbl> 513, 460, 137, 99, 68, 17, 284, 432, 26, 5...
## $ origin_station_name                   <chr> "San Martin II", "Beiro Y Segurola", "Azop...
## $ origin_station_address                <chr> "Av. San Martín 5129", "Segurola 3194", "A...
## $ origin_longitude                     <dbl> -58.49074, -58.51193, -58.36749, -58.43541...
## $ origin_latitude                      <dbl> -34.59713, -34.60750, -34.61560, -34.59610...
## $ end_datetime                          <dttm> 2024-01-23 18:45:28, 2024-01-23 23:03:55, ...
## $ destination_station_id               <dbl> 498, 382, 150, 206, 68, 186, 432, 284, 32, ...
## $ destination_station_name             <chr> "Habana", "Biarritz", "Rodrigo Bueno", "Fi...
## $ destination_station_address          <chr> "Gral. José Gervasio Artigas 4298 (y Haban...
## $ destination_longitude                <dbl> -58.49496, -58.47726, -58.35547, -58.43734...
## $ destination_latitude                 <dbl> -34.58660, -34.60543, -34.61875, -34.58495...
## $ user_id                             <dbl> 992557, 320782, 861425, 320714, 1041602, 9...
## $ bike_model                           <chr> "FIT", "FIT", "FIT", "FIT", "ICONIC", "FIT...
## $ gender                               <chr> "MALE", "FEMALE", "FEMALE", "OTHER", "MALE...
## $ duration_minutes                    <dbl> 9.466667, 22.583333, 11.333333, 7.766667, ...
## $ hour_start                           <int> 18, 22, 15, 17, 21, 22, 19, 21, 17, 12, 11...
## $ weekday                              <ord> Tue, Tue, Tue, Tue, Tue, Tue, Tue, Tu...
## $ day_type                            <chr> "Weekday", "Weekday", "Weekday", "Weekday"...
## $ month                                <ord> January, January, January, January, Januar...
```

Show

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

### Key Results After Cleaning

- The final dataset contains **3,234,209 valid trips** (after removing invalid records).
- Trip durations in minutes have a median of ~16.1 minutes and an average of ~21.7 minutes, which is typical for urban bike-sharing (short commutes with some longer rides).
- Most trips are between ~10 and 26 minutes (between 1st and 3rd quartile).

### Data Reduction Summary

Stage	Number of Trips	% of Original
Original (raw)	3,559,284	100%
After cleaning	3,234,209	~90.8%
Removed (invalid)	~325,075	~9.1%

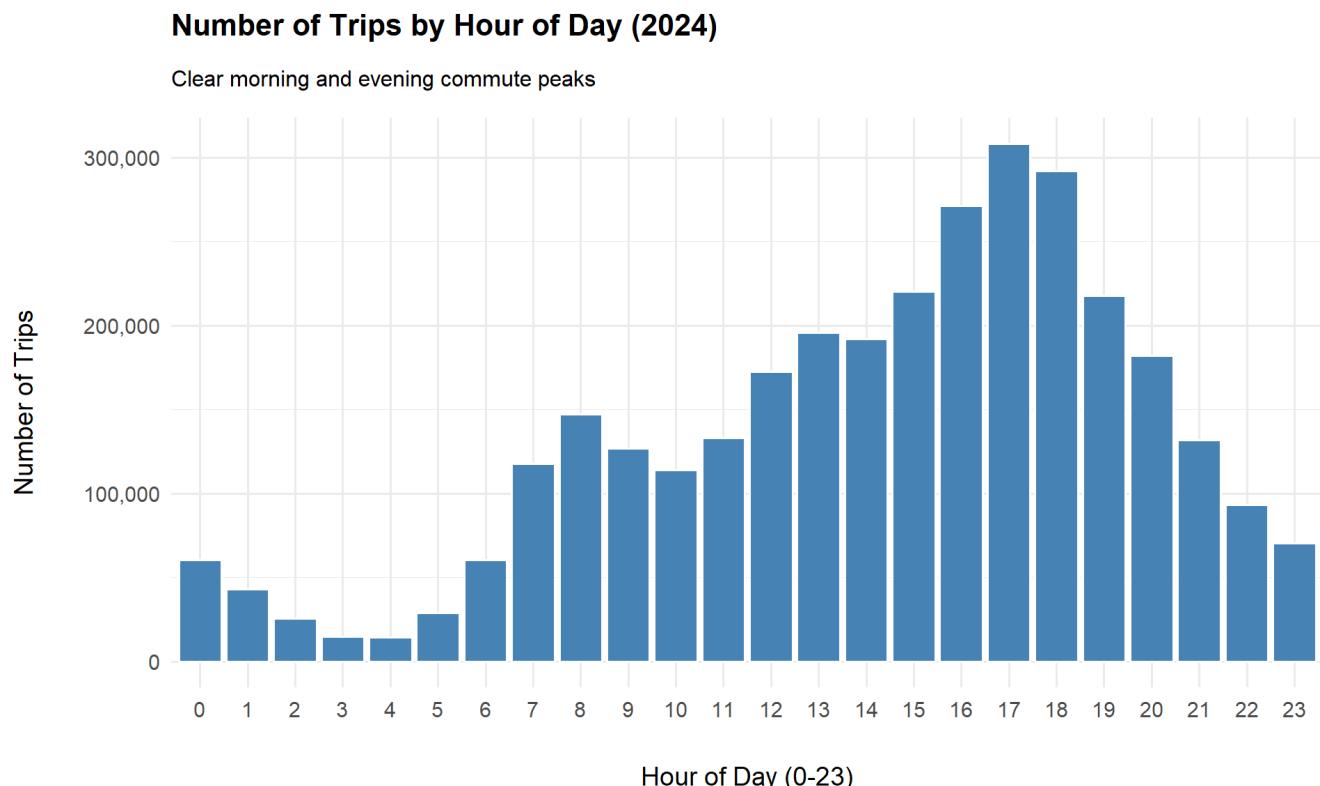
This cleaning process removed approximately 325,000 invalid or outlier records (~9% of the original data), which is normal and expected in large-scale public datasets like this one. The remaining data is clean, consistent, and ready for deeper analysis.

## 5 Exploratory Data Analysis (EDA)

### 5.1 Peak Hours of Usage

This bar chart shows trips started by hour of the day (0–23), highlighting daily usage patterns and rush hours.

Show



#### Key Insights

- Evening peak (16–18 hs) dominates with **271,500–308,297 trips per hour** — strong commuting home pattern.
- Secondary morning peak (7–9 AM): **117,583–147,119 trips** — entry commute.
- Midday (12–15 hs) steady at **172,710–220,351 trips** — mix of errands and leisure.

#### Potential Recommendations

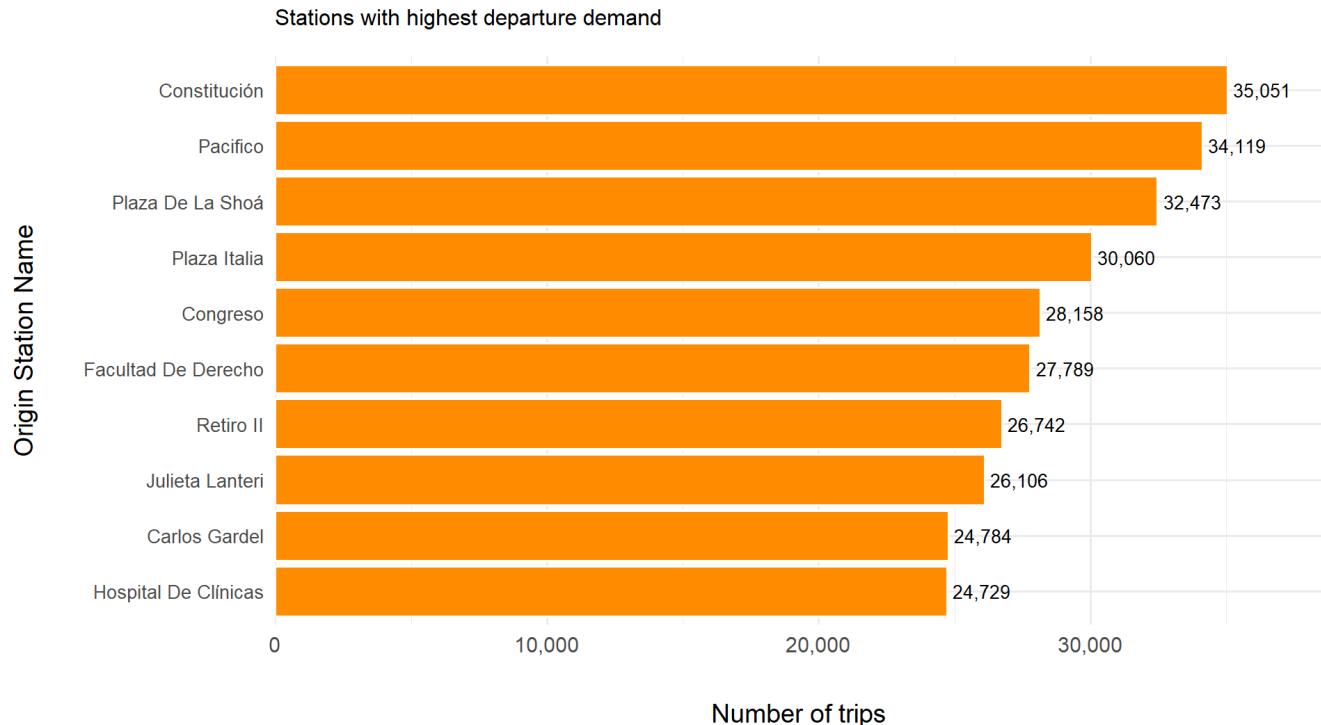
- Increase bike availability during evening peak (16–18 hs) to meet highest demand.
- Explore incentives for returning bikes in high-demand periods.

### 5.2 Top 10 Most Popular Origin Stations

This horizontal bar chart shows the 10 most used origin stations in 2024 (cleaned names), revealing high-demand departure points.

Show

## Top 10 Origin Stations by Number of Trips (2024)



### Key Insights

- Top station: **Constitución** (35,051 trips), followed by **Pacifico** (34,119) and **Plaza De La Shoá** (32,473).
- The top 10 stations account for approximately 8.9% of all trips — showing concentrated demand in a few locations.
- Many top stations align with transport hubs or central areas, supporting commuting focus.

### Potential Recommendations

- Prioritize bike supply and rebalancing near top origin stations (e.g., Constitución, Pacifico) during peak hours.
- Consider incentives for returning bikes to these busy departure points.

### Note on Analysis Focus

This analysis prioritizes **origin stations** as the key indicator of departure demand (where bikes are needed most). Destination stations reflect arrival patterns and can be explored in future analyses for full trip flows or net bike movement.

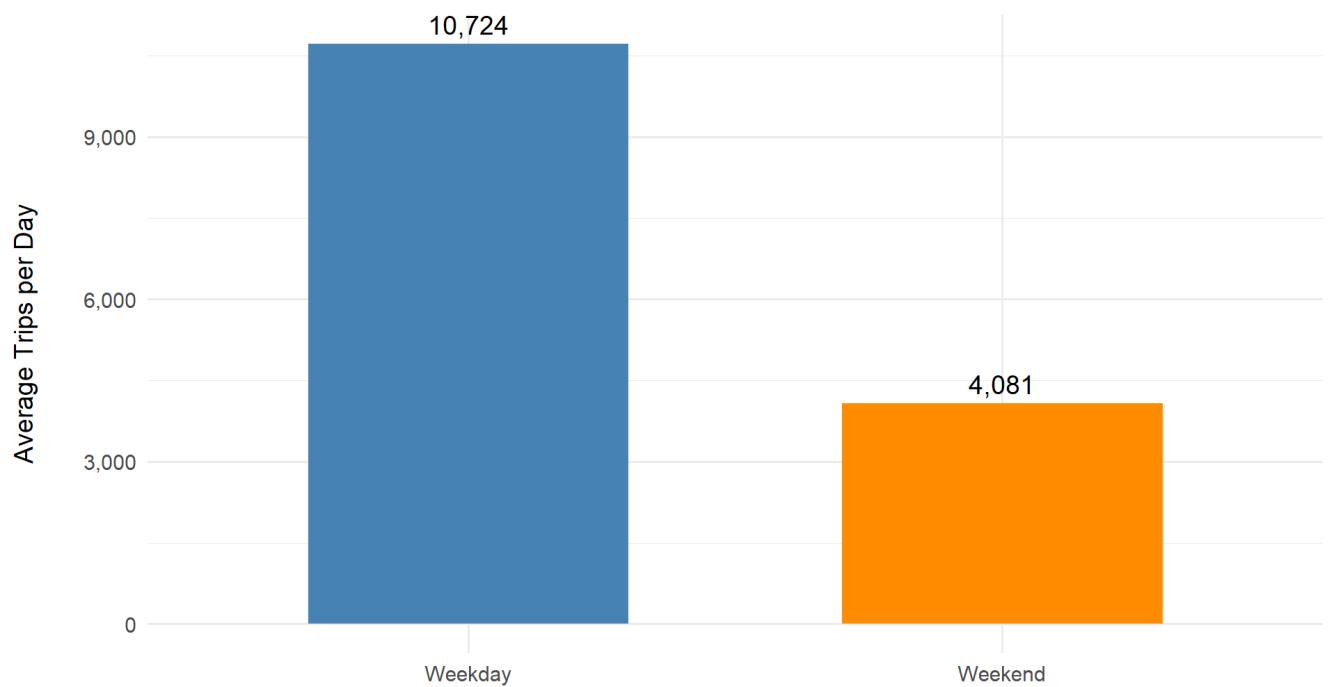
## 5.3 Average Daily Trips: Weekday vs Weekend (Normalized)

This bar chart shows the average number of trips per day on weekdays vs weekends (normalized by unique days to account for weekdays having ~2.5× more days).

Show

## Average trips per Day: Weekday vs Weekend (2024)

Normalized by number of unique days in each category



### Key Insights

- Weekday average is significantly higher than weekend, confirming **daily commuting** as the primary use case.
- Weekend usage is lower but still substantial, suggesting some **recreational or leisure activity** on Saturdays and Sundays.
- The weekday-to-weekend ratio (~2.6:1 per day) reinforces the commuting pattern seen in peak hours analysis.

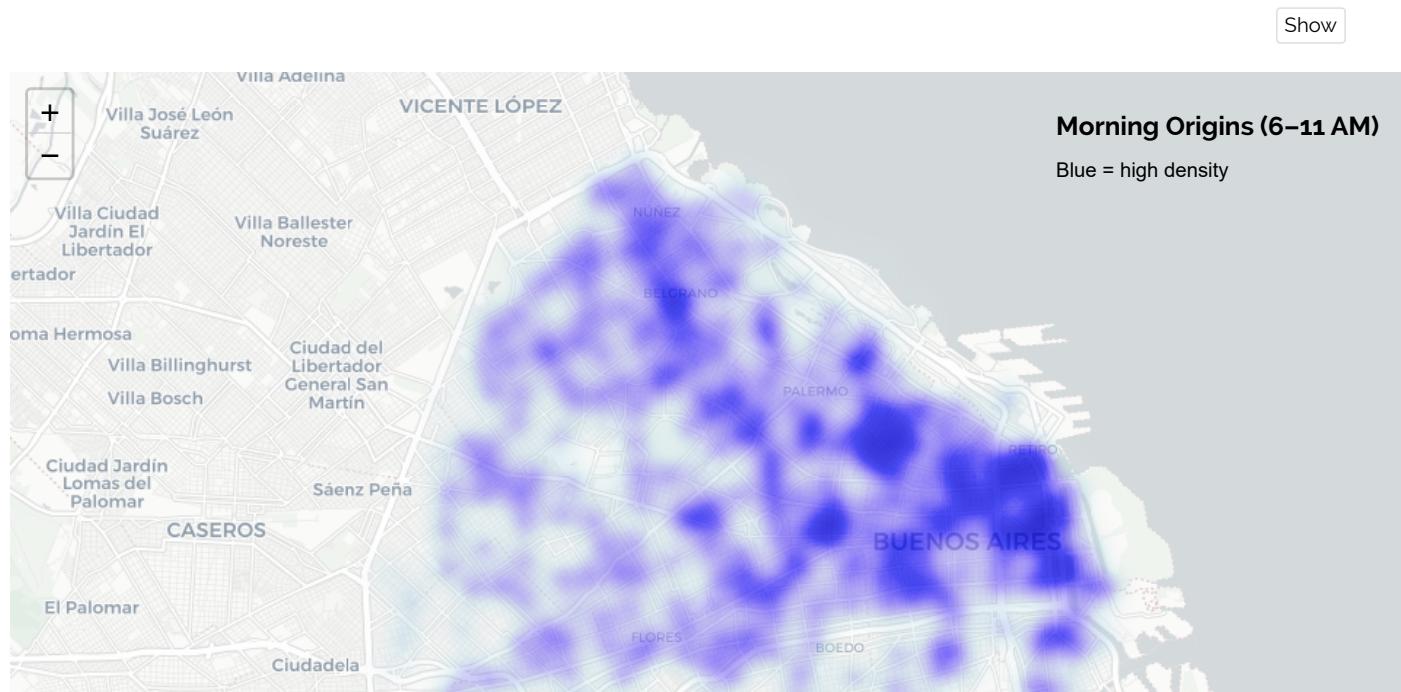
### Potential Recommendations

- Prioritize bike availability and rebalancing on weekdays during peak hours.
- Explore weekend promotions to boost recreational use.

## 5.4 Spatial Heatmaps of Trip Origins and Destinations

These interactive heatmaps visualize the geographic distribution of trip **origins** (blue) and **destinations** (red) in Buenos Aires (2024 sample).

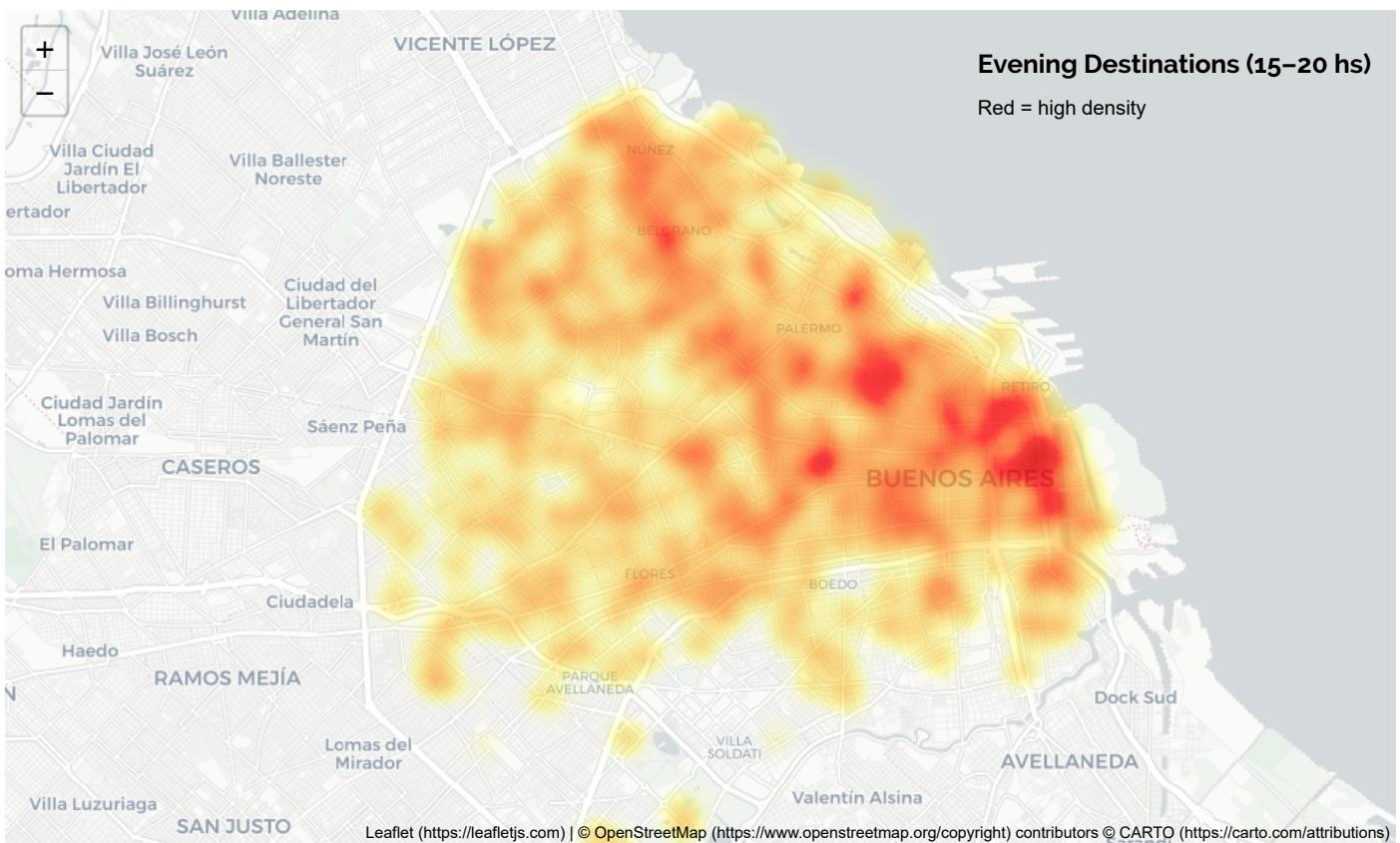
### 5.4.1 Morning Origins Heatmap (6–11 hs)





#### 5.4.2 Evening Destinations Heatmap (15–20 hs)

Show



#### Key Insights – Commuting Flows: Morning Origins vs Evening Destinations

- Morning origins (blue) show high concentration in central and northern areas (e.g., Microcentro, Palermo, Recoleta), while evening destinations (red) maintain high density in the center but with more dispersion across the city.
- This pattern indicates a **commuting flow** where people start trips in central/residential areas in the morning and return to more dispersed residential zones in the evening.
- The strong central saturation in both periods suggests that the core of the city acts as a major hub for both departures and arrivals.

#### Potential Recommendations

- Prioritize bike availability in **central high-density clusters** during morning departures and evening arrivals to meet persistent demand in the center.
- Use dynamic rebalancing to redistribute excess bikes from the center (where they accumulate during the day) to more dispersed residential areas in the evening, ensuring availability for return trips.

#### Note on sample

A random sample of 10,000 trips was used for interactive performance; analysis of the full dataset shows identical spatial patterns.

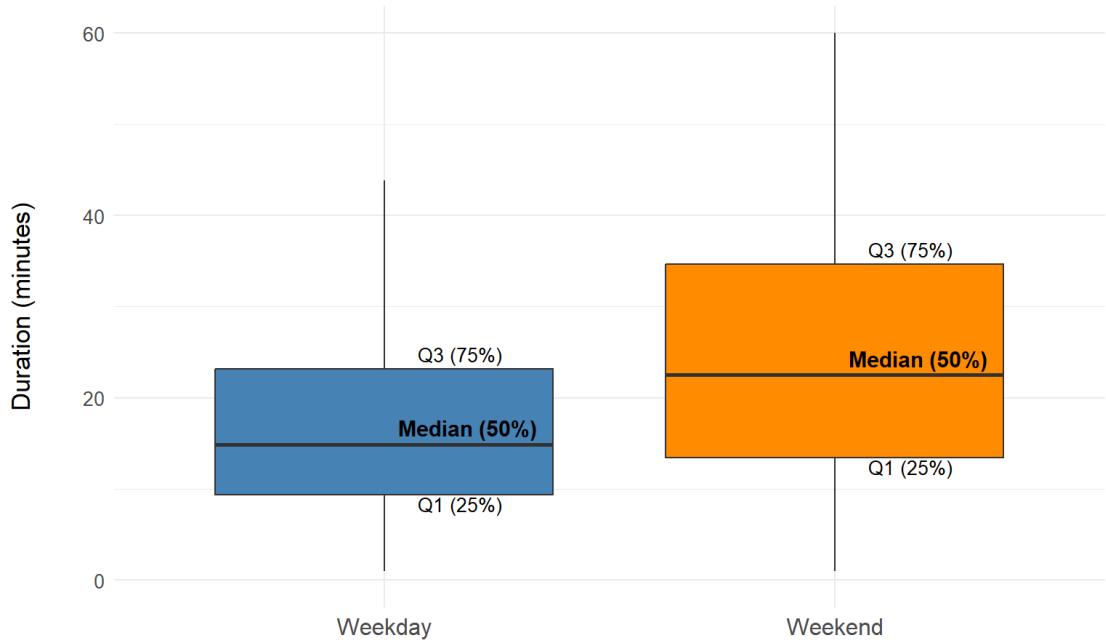
### 5.5 Trip Duration Analysis

This boxplot shows the distribution of trip durations in minutes by day type, revealing typical usage lengths.

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).

### Key Insights

- Trip durations are generally short, with a median of ~16.1 minutes and an average of ~21.7 minutes – typical for urban bike-sharing (quick commutes or errands).
- Weekday trips tend to be **slightly shorter** (more focused commuting), while weekend trips show **greater variation** (likely including leisure or longer recreational rides).
- Most trips (75% or more) are under ~26 minutes, indicating efficient, short-distance use.

### Potential Recommendations

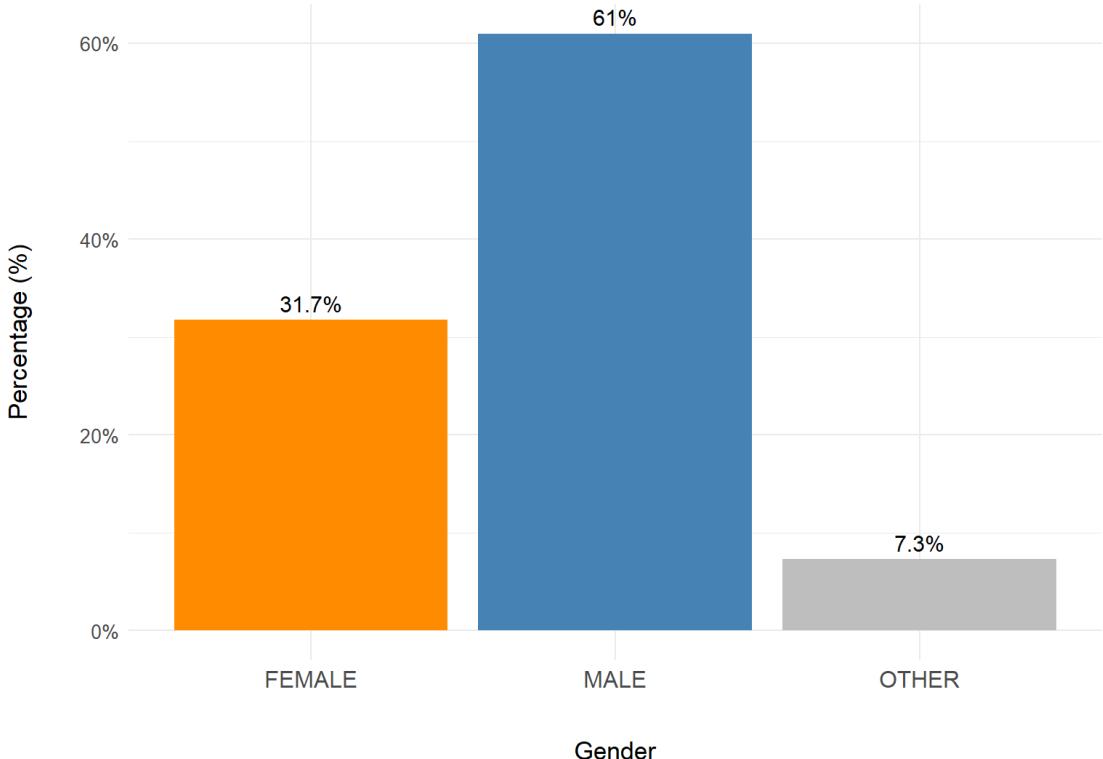
- Leverage the existing free 30-minute limit on weekdays by promoting short, efficient trips (e.g., campaigns encouraging commuting under 30 minutes to avoid charges).
- Explore extending the free time or number of trips on weekends to boost recreational usage, as weekend trips show more variation and potential for longer leisure rides.

## 5.6 Usage by Gender

This bar chart shows the percentage of trips by gender, providing demographic insights.

Show

## Percentage of Trips by Gender (2024)



### Key Insights – Usage by Gender

- Male users account for the majority of trips (~60.8%), while female users represent ~31.6%.
- A small percentage (~7.3%) are classified as “Other”, with negligible missing values (0.3%).
- The gender imbalance suggests opportunities to increase female participation through targeted campaigns, safety improvements, or inclusive marketing.

### Potential Recommendations

- Launch initiatives to boost female usage (e.g., women-only events, safety features, or promotions aimed at female commuters).
- Monitor gender trends over time to assess the impact of inclusivity efforts.

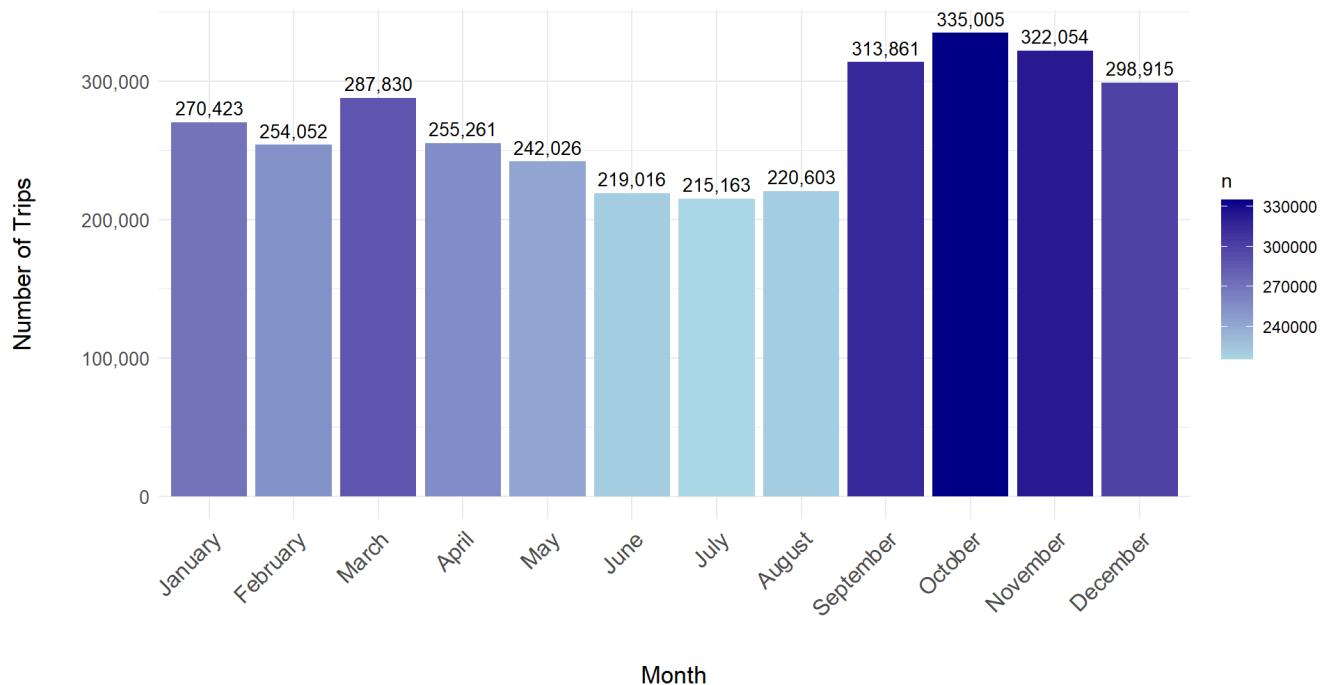
## 5.7 Trips by Month (Seasonality)

This bar chart shows the total number of trips per month in 2024, revealing seasonal patterns in usage.

Show

## Total Trips by Month (2024)

Seasonal usage patterns throughout the year



### Key Insights

- Usage peaks in the **late spring and early summer** (September–December), with the highest counts in **October** (335,005 trips) and **November** (322,054 trips).
- The lowest usage occurs in **winter** (June–August), with the minimum in **July** (215,163 trips), likely due to colder weather in Buenos Aires.
- Overall, there is a clear seasonal pattern: higher ridership during warmer months (late spring/early summer) and lower in winter, typical for outdoor urban bike-sharing in the Southern Hemisphere.

### Potential Recommendations

- Increase bike supply and marketing efforts during the **peak season** (September–December) to capitalize on higher demand.
- Consider reduced operations or targeted incentives in winter months (June–August) to maintain engagement during lower ridership periods.

## 5.8 Net flow by Station (Origins vs Destinations)

This horizontal bar chart shows the net flow of bikes per station (destinations minus origins, that is, arrivals minus departures) in 2024.

Positive values (blue) indicate bike excess (more arrivals, that is, destinations).

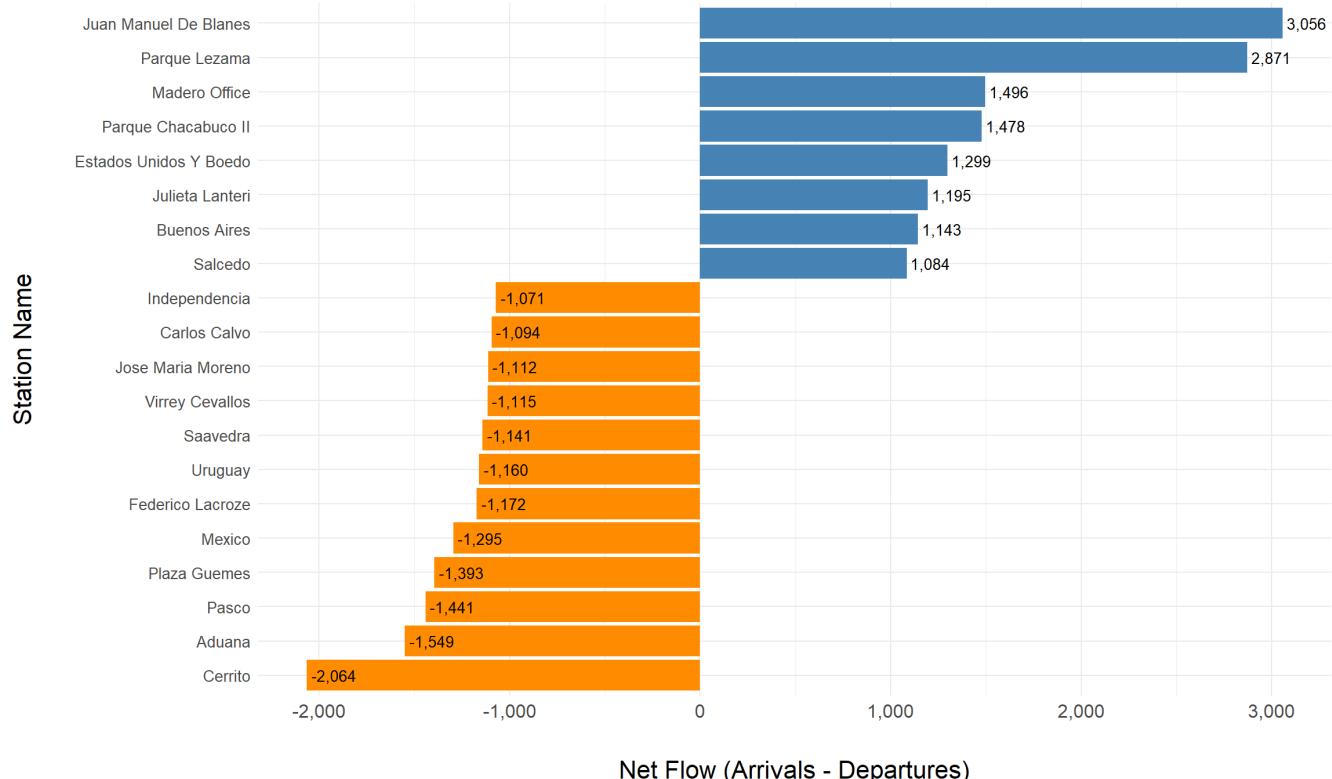
Negative values (orange) indicate bike shortage (more departures, that is, origins).

Stations are ordered from highest positive net flow (top) to lowest negative (bottom).

Show

## Net Bike Flow by Station (Top 20 Extremes)

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



### Key Insights

- Stations like **Juan Manuel De Blanes** (+3,056) and **Parque Lezama** (+2,871) have the highest positive net flow (excess arrivals, that is, destinations) — they accumulate bikes and need removal.
- Stations like **Cerrito** (-2,064) and **Aduana** (-1,549) have the most negative net flow (excess departures, that is, origins) — they lose bikes and need addition.
- The top extremes highlight stations that require frequent rebalancing to maintain availability.

### Potential Recommendations

- Prioritize rebalancing for stations with positive net flow (remove excess bikes).
- Prioritize rebalancing for stations with negative net flow (add bikes).
- Use predictive models to anticipate daily net flow based on time of day and historical patterns.

## 6 Interactive Dashboard in Tableau

To make the analysis more interactive and exploratory, I created a Tableau dashboard with the cleaned dataset.

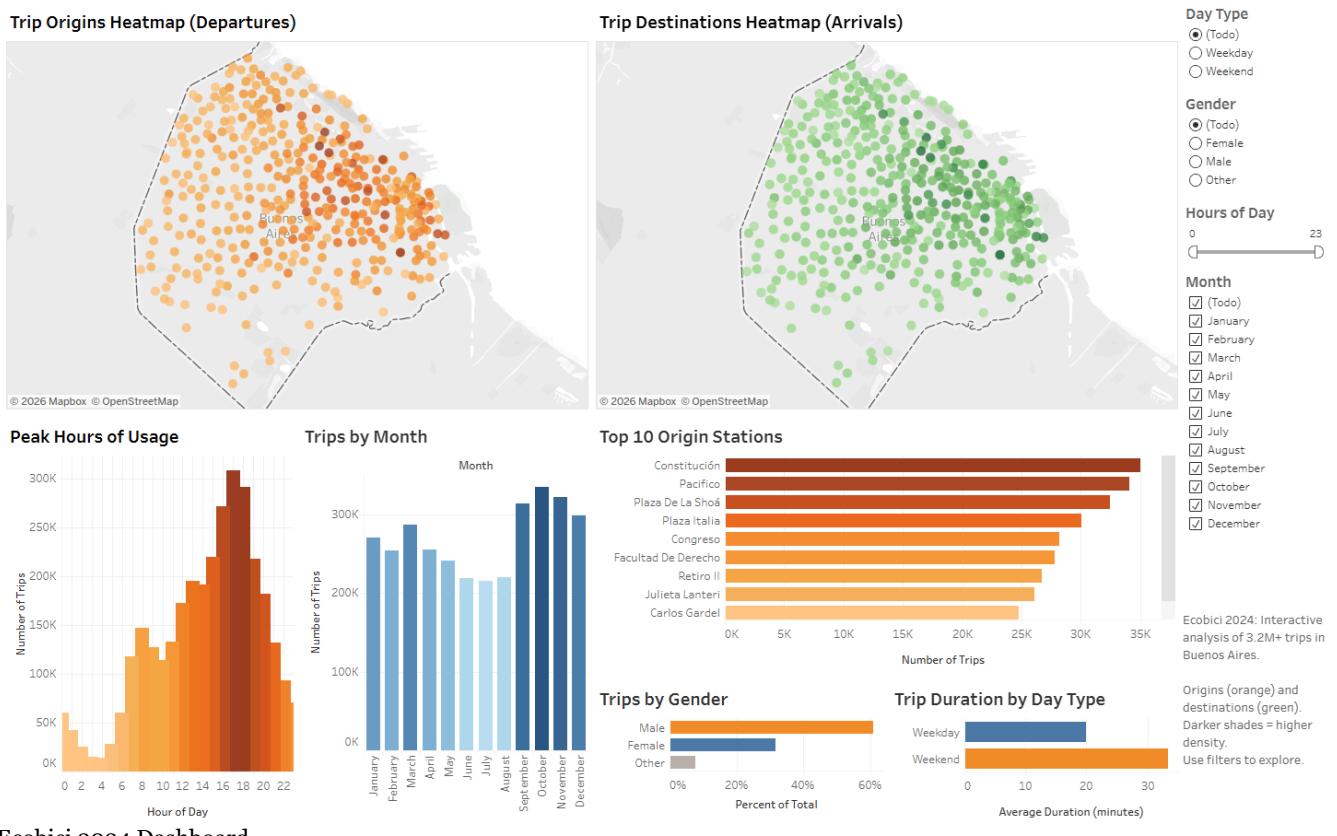
### Dashboard Highlights

- Origins and Destinations heatmaps (interactive with filters for Hour of Day, Day Type, Month, and Gender).
- Peak hours bar chart.
- Trips by month bar chart.
- Top 10 origin stations.
- Trip duration boxplot by day type.
- Usage by gender bar chart.

### Link to the full interactive dashboard

View the Ecobici 2024 Dashboard on 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](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))

### Screenshot



## Ecobici 2024 Dashboard

# 7 Conclusion

This project analyzed over 3.2 million Ecobici trips from 2024 to uncover key usage patterns in Buenos Aires' public bike-sharing system.

## Main Findings

- Strong commuting focus: Evening peak (16–18 hs, up to 308k trips/hour) and secondary morning peak (7–10 AM).
- Weekday average ~10,724 trips/day vs ~4,081 on weekends (ratio ~2.6:1).
- Top origin stations (Constitución 35k, Pacífico 34k) and spatial heatmaps show concentrated demand in central areas (Microcentro, Palermo, Recoleta).
- Morning origins cluster in residential/central zones, evening destinations disperse, confirming bidirectional commuting.
- Trip durations: median ~16.1 min, average ~21.7 min — most under 30 min.
- Gender distribution: Male ~60.8%, Female ~31.6%.
- Seasonal peaks in late spring/early summer (October 335k), lowest in winter (July 215k).
- Net flow extremes (e.g., Juan Manuel De Blanes +3,056, Cerrito -2,064) highlight rebalancing needs.

## Lessons Learned

- Open data sources can have inconsistencies (e.g., station ID mismatches) — relying on built-in coordinates ensured complete coverage without external dependencies.
- Normalization techniques (such as per-day averages) are critical to avoid misleading conclusions when comparing groups with unequal sizes.
- Sampling (10,000 trips) strikes a practical balance between interactive performance and analytical accuracy in large-scale visualizations.
- Consistent terminology (e.g., origins/destinations instead of mixing with departures/arrivals) and clear explanations significantly improve readability and stakeholder understanding.

## Next Steps

- Incorporate real-time data for dynamic rebalancing.
- Explore user demographics (age, gender) and predictive modeling for demand forecasting.

- Expand to 2025 data when available.

This reproducible report demonstrates skills in R (tidyverse, leaflet), data cleaning, EDA, visualization, interactive dashboards (Tableau), and business insights — ready for data analyst roles.

**View the full rendered report online** (with interactive heatmaps):

Open Ecobici 2024 Report on my website (<https://www.emilionahuelpattini.com/en/data/data-analysis/projects/ecobici-2024/urban-mobility-analysis.html>)

(Interactive version with zoom, pan, and tooltips — no download needed.)

## **Interactive Dashboard**

Explore the full dataset interactively in Tableau Public:

View the Ecobici 2024 Dashboard ([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](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))

## **Download PDF Report**

For a printable version:

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

(Full report with all visuals and insights — 12 pages)