

1 Panoramica del Progetto e Caso di Business

2 Fonti dei Dati

3 Caricamento e Preparazione dei Dati

4 Pulizia dei Dati, Rinomina e Creazione di Variabili

5 Analisi Esplorativa dei Dati (EDA)

6 Dashboard Interattivo in Tableau

7 Conclusione

Code ▾

# Mobilità Urbana a Buenos Aires: Analisi con R e Tableau (Ecobici 2024)

Emilio Nahuel Pattini

2026-01-16

## 1 Panoramica del Progetto e Caso di Business

Questo progetto analizza i dati dei viaggi dell'intero anno 2024 del sistema pubblico di bike-sharing di Buenos Aires, **Ecobici**, per individuare modelli di utilizzo, picchi di domanda e opportunità per migliorare la mobilità urbana sostenibile.

Man mano che le città nel mondo adottano trasporti più ecologici, comprendere il comportamento nell'uso delle biciclette condivise è fondamentale per ottimizzare la collocazione delle stazioni, la gestione della flotta e ridurre congestione ed emissioni.

Utilizzando il dataset più recente e completo (~3,56 milioni di viaggi), questa analisi fornisce insight concreti per la pianificazione delle smart city e strategie di mobilità sostenibile a livello globale.

### Obiettivi Principali:

- Identificare le ore, i giorni e i modelli stagionali di maggiore utilizzo delle biciclette.
- Mappare le stazioni e i quartieri più popolari di origine e destinazione.
- Esplorare la durata dei viaggi e la demografia degli utenti (quando disponibile).
- Fornire raccomandazioni basate sui dati per migliorare il sistema.

Questo report riproducibile dimostra competenze in pulizia dei dati, analisi esplorativa, visualizzazione e storytelling aziendale utilizzando R e Tableau.

## 2 Fonti dei Dati

Questa analisi utilizza dati aperti ufficiali del Governo della Città di Buenos Aires.

### 1. Viaggi Ecobici 2024

- File: `ecobici_recorridos_realizados_2024.csv`
- Fonte: <https://data.buenosaires.gob.ar/dataset/bicicletas-publicas> (<https://data.buenosaires.gob.ar/dataset/bicicletas-publicas>)
- Dimensione: ~3,56 milioni di righe (anno completo 2024)
- Colonne chiave (dopo rinomina in inglese):
  - `trip_id` : ID univoco del viaggio
  - `duration_seconds` : Durata del viaggio (secondi)
  - `start_datetime` / `end_datetime` : Ora di inizio e fine
  - `origin_station_id` / `origin_station_name` : Stazione di origine
  - `origin_longitude` / `origin_latitude` : Coordinate di origine
  - `destination_station_id` / `destination_station_name` : Stazione di destinazione
  - `destination_longitude` / `destination_latitude` : Coordinate di destinazione
  - `user_id`, `bike_model`, `gender`

### 2. Stazioni Ecobici (Nuevo Sistema)

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

- Nota: Scaricato ma non utilizzato a causa di totale incompatibilità negli ID delle stazioni (0% join riuscito). Il dataset dei viaggi include già nomi e coordinate complete delle stazioni per ogni viaggio, garantendo copertura totale.

### Nota sulla Qualità dei Dati

Il tentativo di unire informazioni su quartiere, comune e collocazione dal file delle stazioni è fallito (0% corrispondenza). L'analisi si basa interamente sui nomi e coordinate integrati nel file dei viaggi. In futuri aggiornamenti sarà possibile aggiungere insight a livello di quartiere con un dataset delle stazioni sincronizzato.

Tutti i dati sono pubblici e aperti. I dati del 2025 non sono ancora completamente disponibili a gennaio 2026. I nomi delle colonne sono stati standardizzati in inglese per maggiore chiarezza e leggibilità internazionale.

## 3 Caricamento e Preparazione dei Dati

I dati grezzi dei viaggi Ecobici (2024) vengono caricati in modo silenzioso. I nomi originali delle colonne sono in spagnolo, come forniti dal portale di dati aperti.

## 4 Pulizia dei Dati, Rinomina e Creazione di Variabili

In questa fase:

- Rinomino tutte le colonne in inglese per garantire coerenza e leggibilità internazionale.
- Pulisco e normalizzo i nomi delle stazioni.
- Creo nuove variabili derivate (durata in minuti, ora di inizio, giorno della settimana, indicatore weekend, mese).
- Filtro i viaggi non validi o anomali (ad esempio, viaggi inferiori a 1 minuto, superiori a 24 ore o con valori chiave mancanti).

Dopo la pulizia, mostro la struttura e le statistiche di base del dataset finale.

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
```

### Risultati Chiave Dopo la Pulizia

- Il dataset finale contiene **3.234.209 viaggi validi** (dopo la rimozione dei record non validi).
- Le durate dei viaggi in minuti hanno una mediana di ~16,1 minuti e una media di ~21,7 minuti, tipica dei sistemi di bike-sharing urbani (spostamenti brevi con alcuni percorsi più lunghi).
- La maggior parte dei viaggi è compresa tra ~10 e 26 minuti (tra il primo e il terzo quartile).

## Sintesi della Riduzione dei Dati

Fase	Numero di Viaggi	% dell'Originale
Originale (grezzo)	3.559.284	100%
Dopo la pulizia	3.234.209	~90,8%
Rimossi (non validi)	~325.075	~9,1%

Questo processo di pulizia ha rimosso circa 325.000 record non validi o anomali (~9% dei dati originali), il che è normale e previsto nei dataset pubblici di larga scala come questo. I dati rimanenti sono puliti, coerenti e pronti per un'analisi più approfondita.

## 5 Analisi Esplorativa dei Dati (EDA)

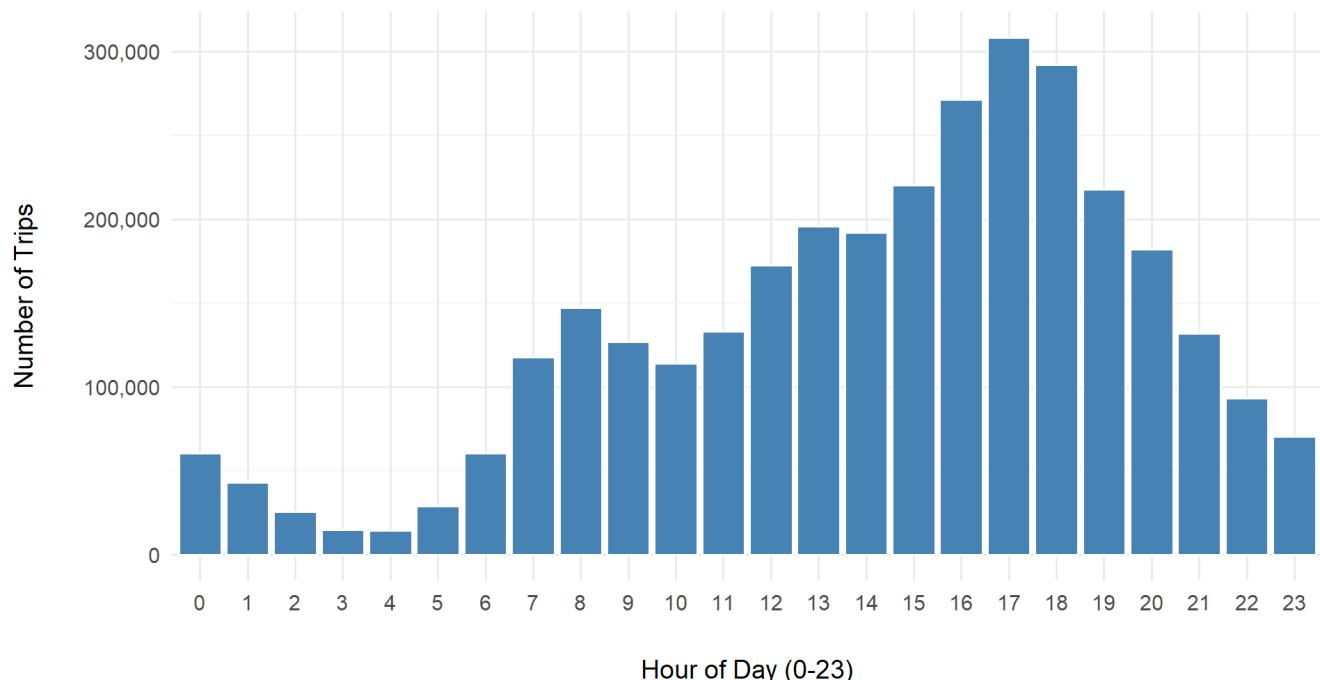
### 5.1 Ore di Punta di Utilizzo

Questo grafico a barre mostra i viaggi iniziati per ora del giorno (0–23), evidenziando i modelli di utilizzo quotidiano e le ore di punta.

Show

Number of Trips by Hour of Day (2024)

Clear morning and evening commute peaks



### Principali Risultati

- Il picco serale (16–18 h) domina con **271.500–308.297 viaggi all'ora** — forte modello di rientro a casa.
- Picco secondario al mattino (7–9 h): **117.583–147.119 viaggi** — spostamento di ingresso al lavoro/studio.
- Mezzogiorno (12–15 h) stabile a **172.710–220.351 viaggi** — mix di commissioni e tempo libero.

### Raccomandazioni Potenziali

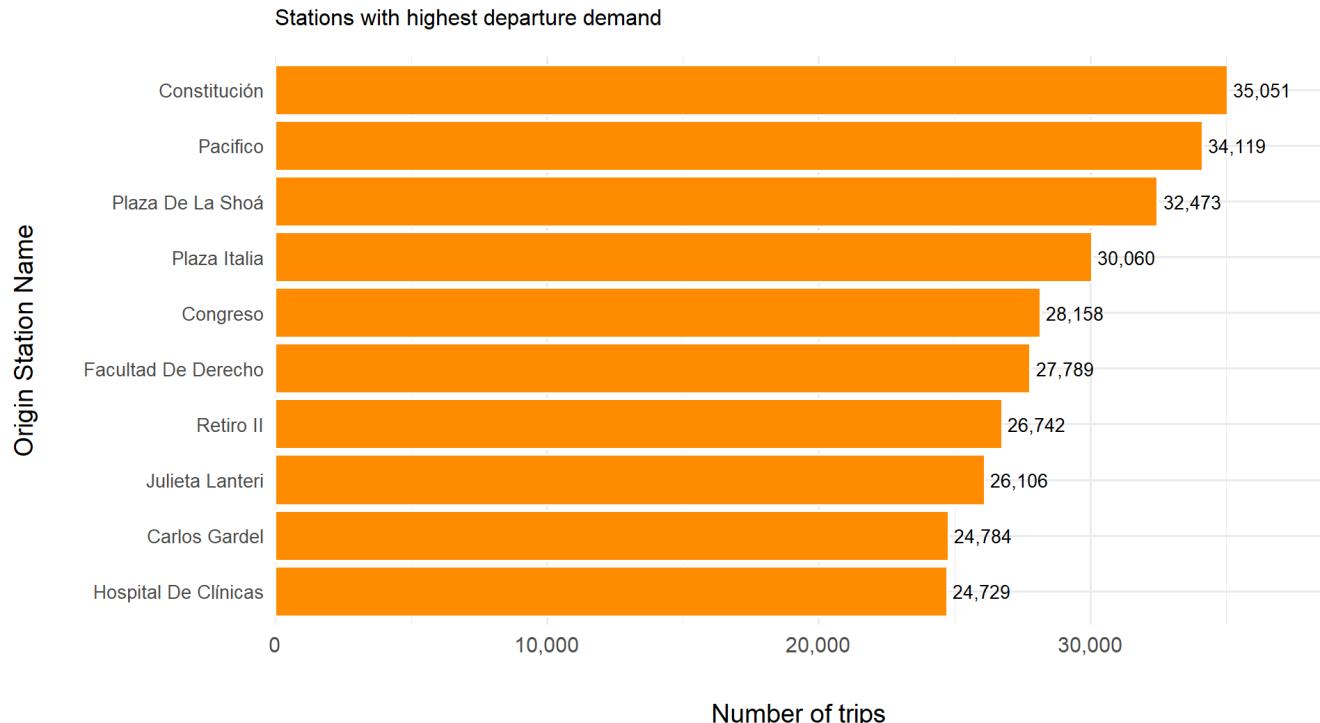
- Aumentare la disponibilità di biciclette durante il picco serale (16–18 h) per soddisfare la massima domanda.
- Esplorare incentivi per la restituzione delle biciclette nei periodi di alta domanda.

### 5.2 Top 10 Stazioni di Origine Più Popolari

Questo grafico a barre orizzontali mostra le 10 stazioni di origine più utilizzate nel 2024 (nomi puliti), rivelando i punti di partenza a maggiore domanda.

Show

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



### Risultati Chiave

- Stazione principale: **Constitución** (35.051 viaggi), seguita da **Pacífico** (34.119) e **Plaza De La Shoá** (32.473).
- Le 10 stazioni principali rappresentano circa l'8,9% di tutti i viaggi — mostrando una domanda concentrata in poche località.
- Molte delle stazioni più utilizzate coincidono con nodi di trasporto o aree centrali, a supporto degli spostamenti pendolari.

### Raccomandazioni Potenziali

- Dare priorità alla disponibilità di biciclette e al riequilibrio vicino alle principali stazioni di origine (es. Constitución, Pacífico) durante le ore di punta.
- Considerare incentivi per la restituzione delle biciclette in questi punti di partenza molto frequentati.

### Nota sul Focus dell'Analisi

Questa analisi dà priorità alle **stazioni di origine** come principale indicatore della domanda di partenza (dove le biciclette sono più necessarie).

Le stazioni di destinazione riflettono i modelli di arrivo e possono essere esplorate in analisi future per studiare i flussi completi dei viaggi o il movimento netto delle biciclette.

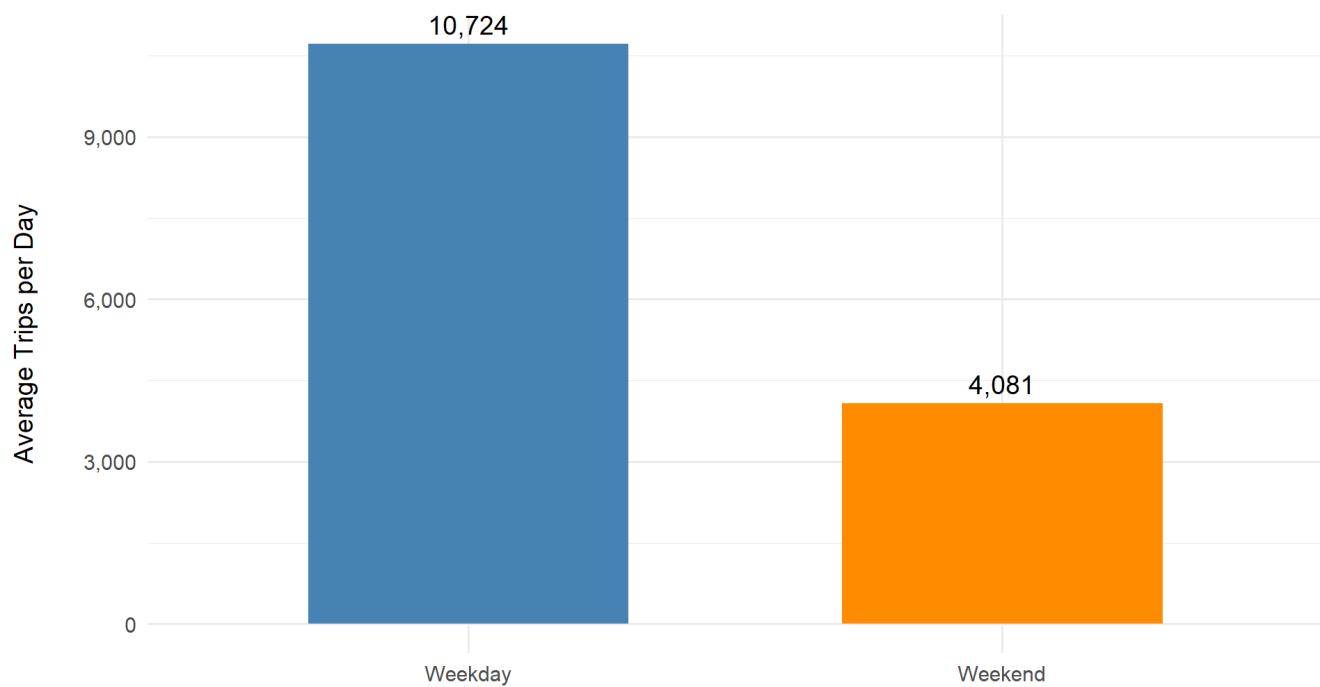
## 5.3 Viaggi Giornalieri Medi: Giorni Feriali vs Weekend (Normalizzato)

Questo grafico a barre mostra il numero medio di viaggi al giorno nei giorni feriali rispetto ai weekend (normalizzato per i giorni unici, considerando che i giorni feriali sono  $\sim 2,5 \times$  più numerosi).

Show

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

Normalized by number of unique days in each category



### Risultati Chiave

- La media dei giorni feriali è significativamente più alta rispetto al weekend, confermando il **pendolarismo quotidiano** come principale caso d'uso.
- L'utilizzo nel weekend è più basso ma comunque consistente, suggerendo una certa **attività ricreativa o di svago** il sabato e la domenica.
- Il rapporto tra giorni feriali e weekend (~2,6:1 al giorno) rafforza il modello di pendolarismo osservato nell'analisi delle ore di punta.

### Raccomandazioni Potenziali

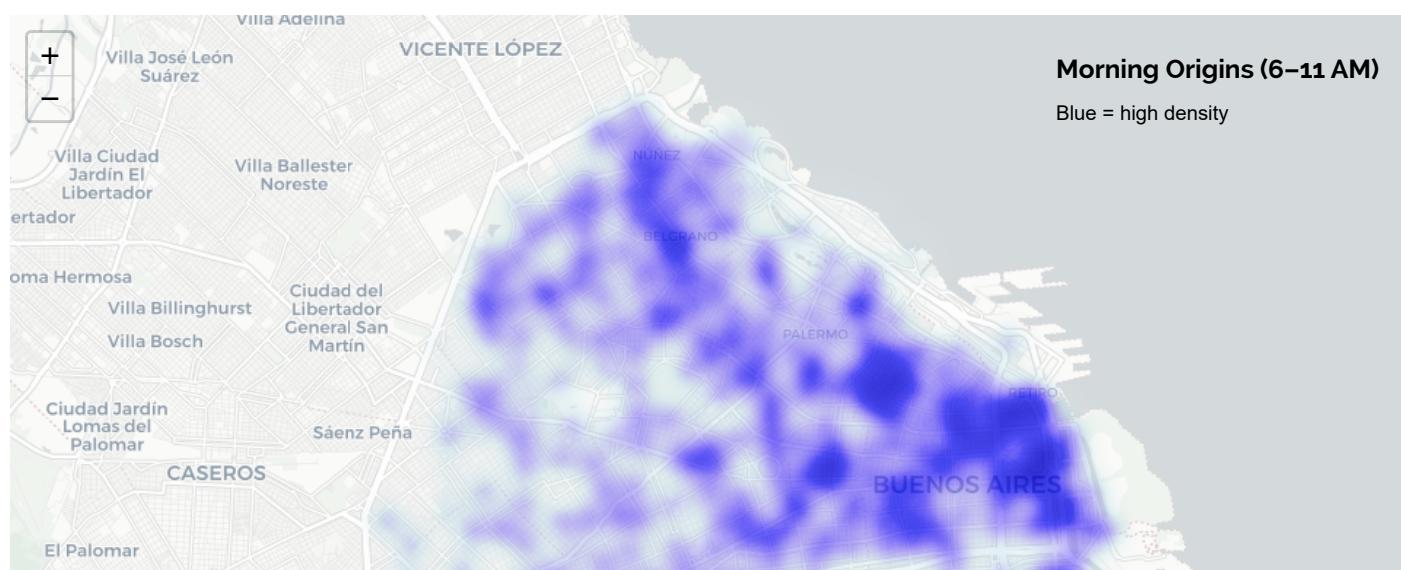
- Dare priorità alla disponibilità di biciclette e al riequilibrio nei giorni feriali durante le ore di punta.
- Esplorare promozioni nel weekend per incentivare l'uso ricreativo.

## 5.4 Heatmap Spaziali delle Origini e Destinazioni dei Viaggi

Queste heatmap interattive visualizzano la distribuzione geografica delle **origini** (blu) e delle **destinazioni** (rosso) dei viaggi a Buenos Aires (campione 2024).

### 5.4.1 Heatmap delle Origini Mattutine (6–11 h)

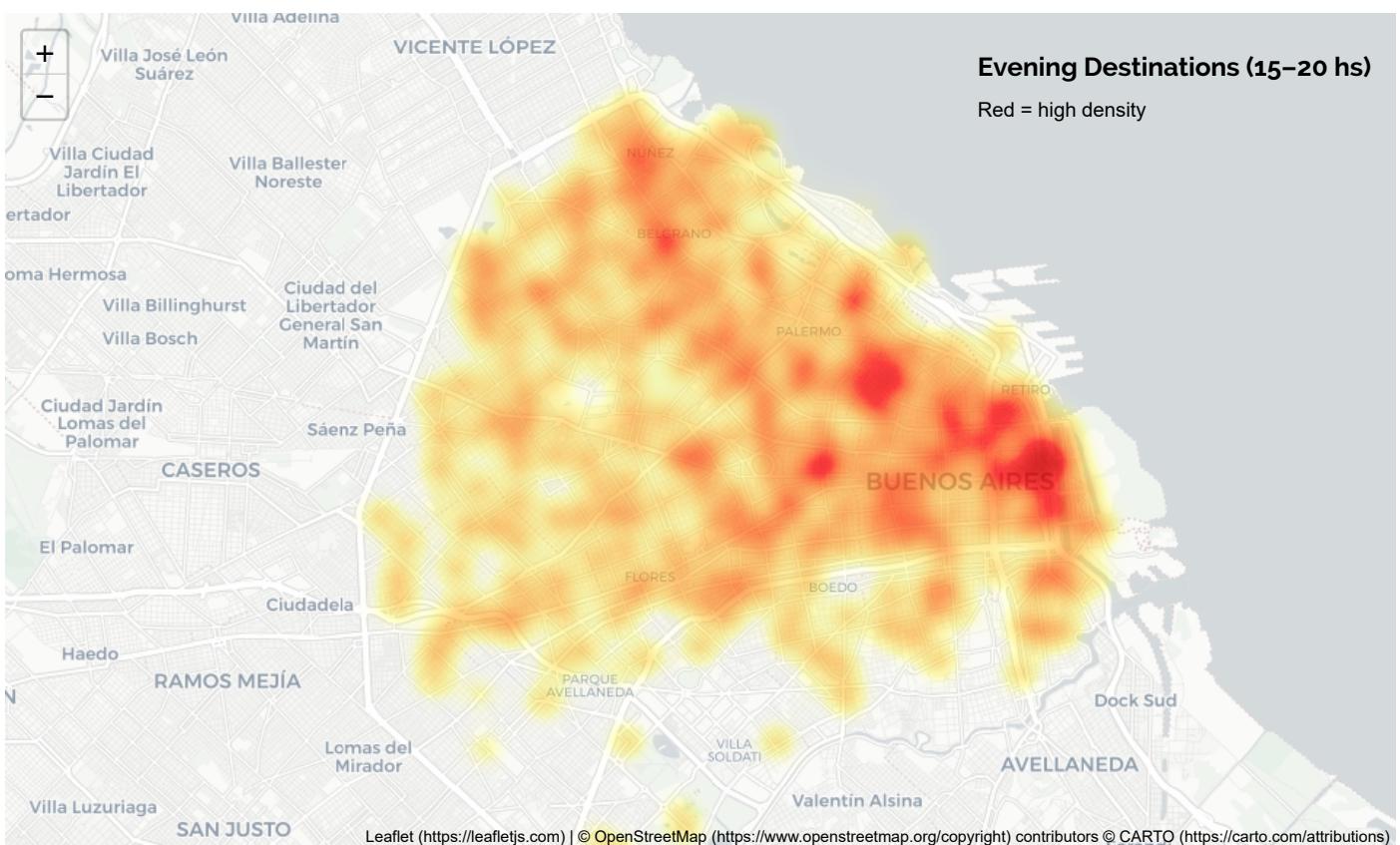
Show





#### 5.4.2 Heatmap delle Destinazioni Serali (15–20 h)

Show



#### Risultati Chiave – Flussi Pendolari: Origini Mattutine vs Destinazioni Serali

- Le origini mattutine (blu) mostrano un'alta concentrazione nelle aree centrali e settentrionali (es. Microcentro, Palermo, Recoleta), mentre le destinazioni serali (rosso) mantengono un'elevata densità nel centro ma con maggiore dispersione in tutta la città.
- Questo modello indica un **flusso pendolare** in cui le persone iniziano i viaggi nelle zone centrali/residenziali al mattino e rientrano in aree residenziali più disperse la sera.
- La forte saturazione centrale in entrambi i periodi suggerisce che il nucleo della città funge da grande hub sia per le partenze che per gli arrivi.

#### Raccomandazioni Potenziali

- Dare priorità alla disponibilità di biciclette nei **cluster centrali ad alta densità** durante le partenze mattutine e gli arrivi serali per soddisfare la domanda persistente nel centro.
- Utilizzare un riequilibrio dinamico per redistribuire l'eccesso di biciclette dal centro (dove si accumulano durante il giorno) verso aree residenziali più disperse la sera, garantendo disponibilità per i viaggi di ritorno.

#### Nota sul campione

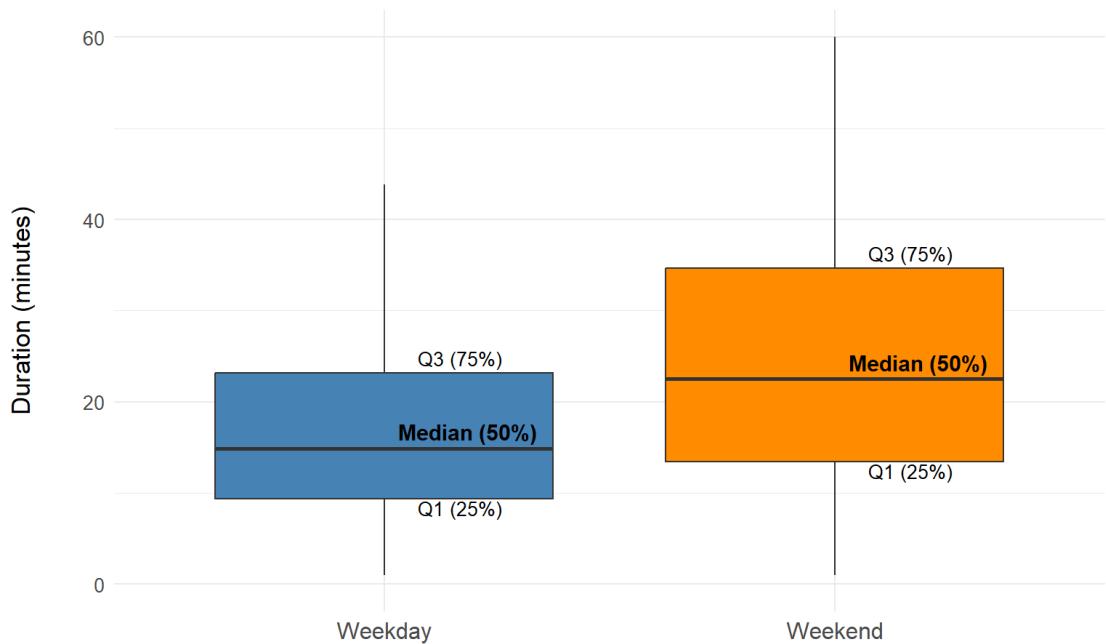
È stato utilizzato un campione casuale di 10.000 viaggi per migliorare le prestazioni interattive; l'analisi dell'intero dataset mostra modelli spaziali identici.

#### 5.5 Analisi della Durata dei Viaggi

Questo boxplot mostra la distribuzione delle durate dei viaggi in minuti per tipo di giorno, rivelando lunghezze di utilizzo tipiche.

## 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).

### Risultati Chiave

- Le durate dei viaggi sono generalmente brevi, con una mediana di ~16,1 minuti e una media di ~21,7 minuti – tipico dei sistemi di bike-sharing urbani (spostamenti rapidi o commissioni).
- I viaggi nei giorni feriali tendono ad essere **leggermente più brevi** (più focalizzati sul pendolarismo), mentre quelli del weekend mostrano **maggior variazione** (probabilmente includendo svago o percorsi ricreativi più lunghi).
- La maggior parte dei viaggi (75% o più) dura meno di ~26 minuti, indicando un utilizzo efficiente e di breve distanza.

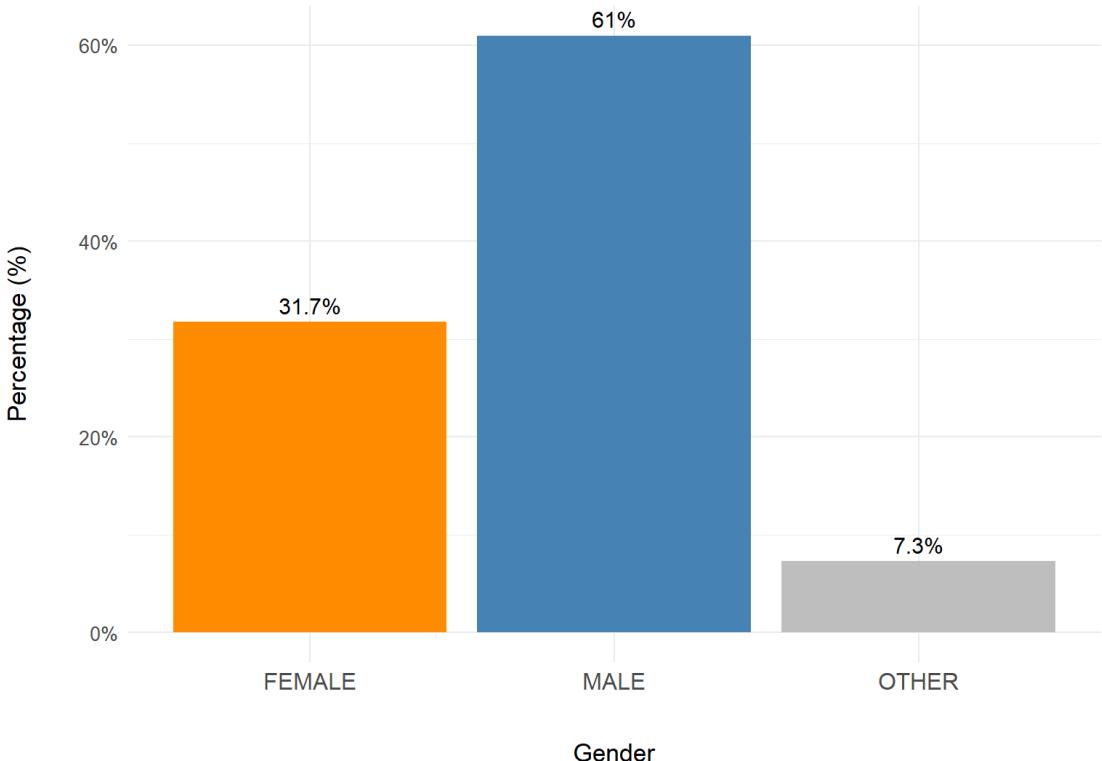
### Raccomandazioni Potenziali

- Sfruttare il limite gratuito esistente di 30 minuti nei giorni feriali promuovendo viaggi brevi ed efficienti (es. campagne che incoraggino spostamenti inferiori a 30 minuti per evitare costi).
- Esplorare l'estensione del tempo gratuito o del numero di viaggi nel weekend per incentivare l'uso ricreativo, dato che i viaggi del fine settimana mostrano maggiore variazione e potenziale per percorsi di svago più lunghi.

## 5.6 Utilizzo per Genere

Questo grafico a barre mostra la percentuale di viaggi per genere, fornendo informazioni demografiche.

## Percentage of Trips by Gender (2024)



### Risultati Chiave – Utilizzo per Genere

- Gli utenti maschi rappresentano la maggioranza dei viaggi (~60,8%), mentre le utenti femmine costituiscono ~31,6%.
- Una piccola percentuale (~7,3%) è classificata come “Altro”, con valori mancanti trascurabili (0,3%).
- Lo squilibrio di genere suggerisce opportunità per aumentare la partecipazione femminile attraverso campagne mirate, miglioramenti della sicurezza o marketing inclusivo.

### Raccomandazioni Potenziali

- Lanciare iniziative per incrementare l'utilizzo femminile (es. eventi riservati alle donne, misure di sicurezza o promozioni rivolte alle pendolari).
- Monitorare le tendenze di genere nel tempo per valutare l'impatto degli sforzi di inclusività.

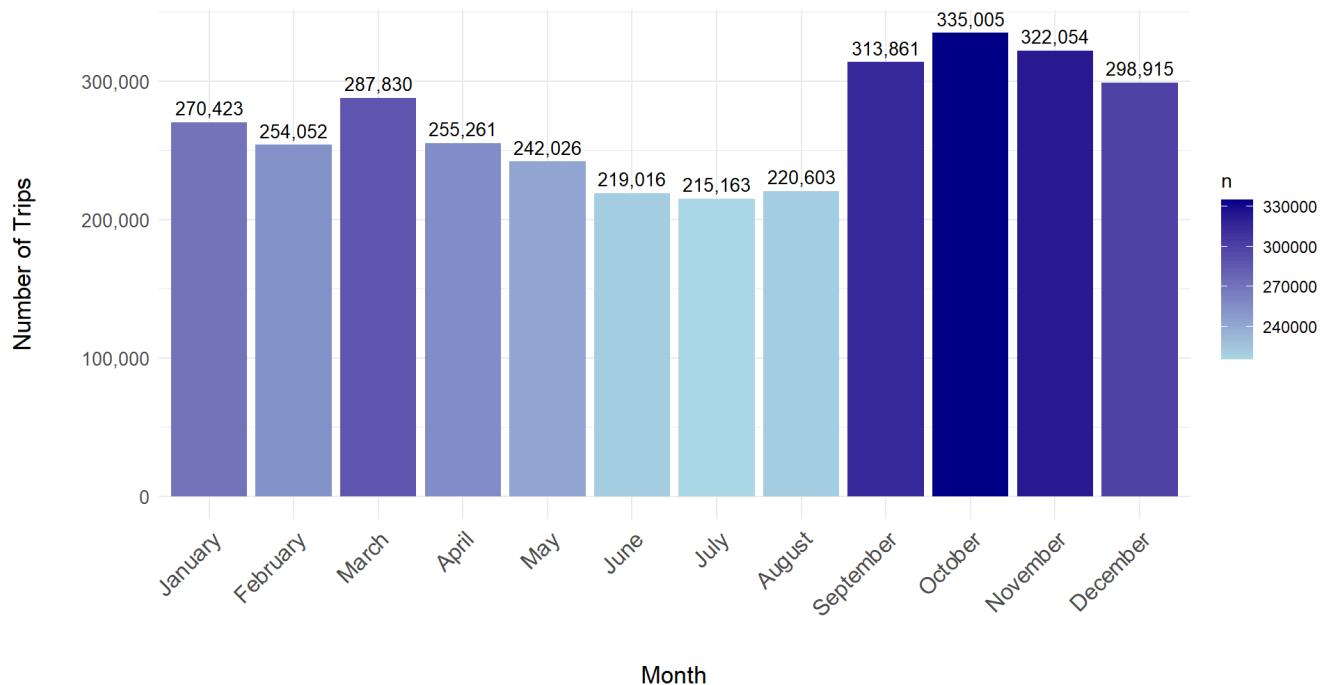
## 5.7 Viaggi per Mese (Stagionalità)

Questo grafico a barre mostra il numero totale di viaggi per mese nel 2024, rivelando modelli stagionali di utilizzo.

Show

## Total Trips by Month (2024)

Seasonal usage patterns throughout the year



### Risultati Chiave

- L'utilizzo raggiunge il picco in **tarda primavera e inizio estate** (settembre–dicembre), con i valori più alti in **ottobre** (335.005 viaggi) e **novembre** (322.054 viaggi).
- Il minor utilizzo si registra in **inverno** (giugno–agosto), con il minimo in **luglio** (215.163 viaggi), probabilmente a causa del clima più freddo a Buenos Aires.
- Nel complesso, emerge un chiaro modello stagionale: maggiore utilizzo nei mesi caldi (tarda primavera/inizio estate) e minore in inverno, tipico dei sistemi di bike-sharing urbani all'aperto nell'emisfero australe.

### Raccomandazioni Potenziali

- Aumentare la disponibilità di biciclette e gli sforzi di marketing durante la **stagione di punta** (settembre–dicembre) per sfruttare la maggiore domanda.
- Considerare operazioni ridotte o incentivi mirati nei mesi invernali (giugno–agosto) per mantenere il coinvolgimento durante i periodi di minore utilizzo.

## 5.8 Flusso Netto per Stazione (Origini vs Destinazioni)

Questo grafico a barre orizzontali mostra il flusso netto di biciclette per stazione (destinazioni meno origini, cioè arrivi meno partenze) nel 2024.

I valori positivi (blu) indicano un eccesso di biciclette (più arrivi, cioè destinazioni).

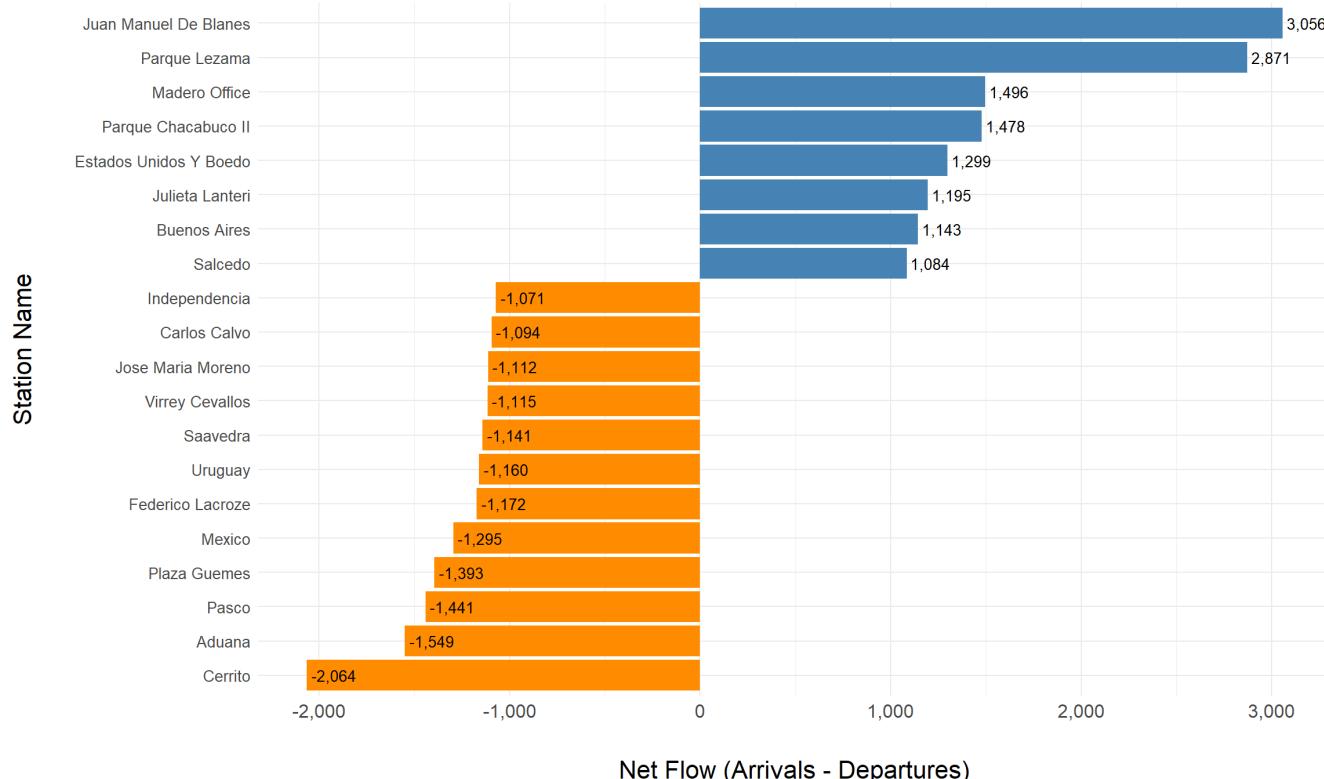
I valori negativi (arancione) indicano una carenza di biciclette (più partenze, cioè origini).

Le stazioni sono ordinate dal flusso netto positivo più alto (in alto) al più basso negativo (in basso).

Show

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

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



### Risultati Chiave

- Stazioni come **Juan Manuel De Blanes** (+3.056) e **Parque Lezama** (+2.871) hanno il flusso netto positivo più alto (eccesso di arrivi, cioè destinazioni) — accumulano biciclette e necessitano di rimozione.
- Stazioni come **Cerrito** (-2.064) e **Aduana** (-1.549) presentano il flusso netto negativo più alto (eccesso di partenze, cioè origini) — perdono biciclette e necessitano di aggiunta.
- Gli estremi principali evidenziano stazioni che richiedono frequente riequilibrio per mantenere la disponibilità.

### Raccomandazioni Potenziali

- Dare priorità al riequilibrio nelle stazioni con flusso netto positivo (rimuovere biciclette in eccesso).
- Dare priorità al riequilibrio nelle stazioni con flusso netto negativo (aggiungere biciclette).
- Utilizzare modelli predittivi per anticipare il flusso netto giornaliero in base all'ora del giorno e ai modelli storici.

## 6 Dashboard Interattivo in Tableau

Per rendere l'analisi più interattiva ed esplorabile, è stato creato un dashboard in Tableau con il dataset pulito.

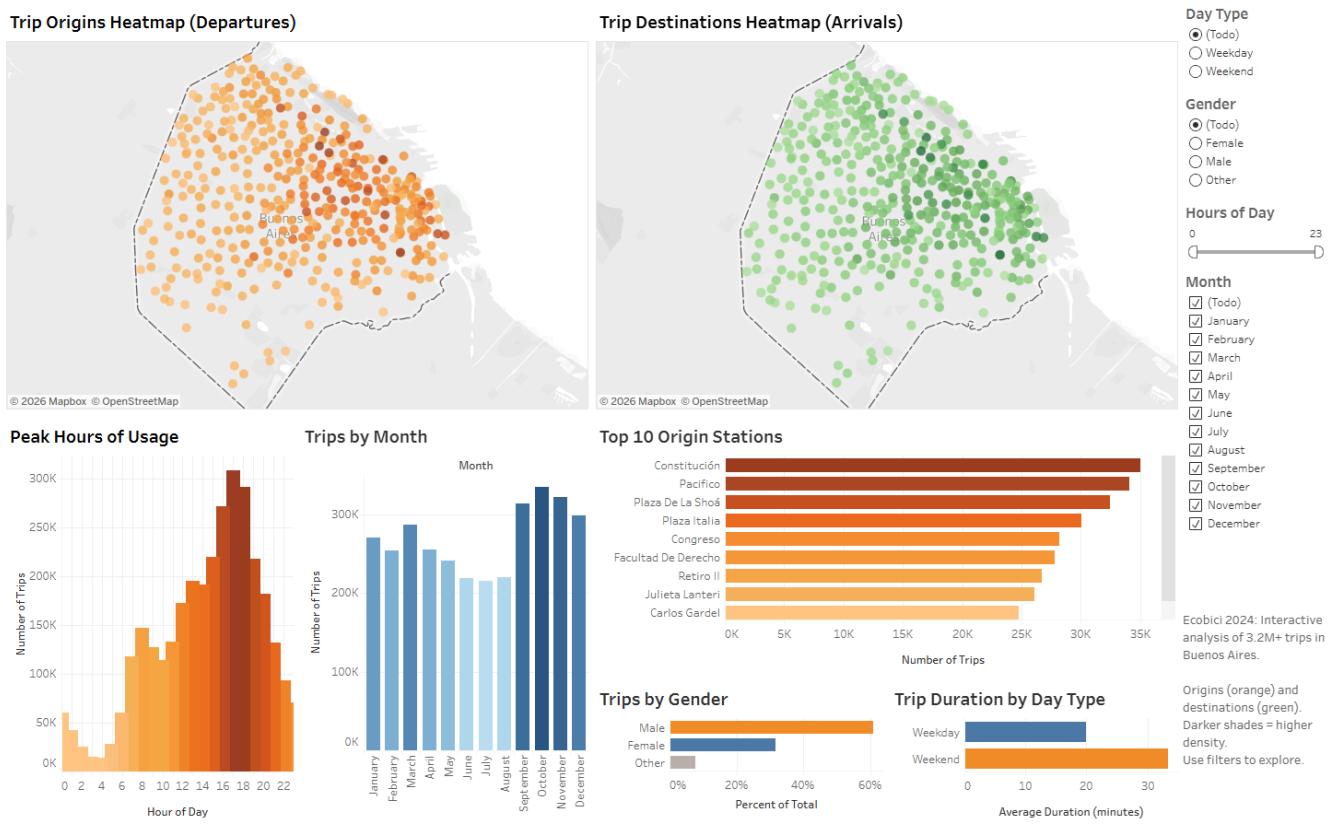
### Punti Salienti del Dashboard

- Heatmap di origini e destinazioni (interattive con filtri per ora del giorno, tipo di giorno, mese e genere).
- Grafico a barre delle ore di punta.
- Grafico a barre dei viaggi per mese.
- Top 10 stazioni di origine.
- Boxplot della durata dei viaggi per tipo di giorno.
- Grafico a barre dell'utilizzo per genere.

### Link al dashboard interattivo completo

Visualizza il Dashboard Ecobici 2024 su 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 Conclusioni

Questo progetto ha analizzato oltre 3,2 milioni di viaggi Ecobici del 2024 per individuare i principali modelli di utilizzo nel sistema pubblico di bike-sharing di Buenos Aires.

## Principali Risultati

- Forte focus sul pendolarismo: picco serale (16–18 h, fino a 308k viaggi/ora) e picco secondario mattutino (7–10 h).
- Media nei giorni feriali ~10.724 viaggi/giorno vs ~4.081 nei weekend (rapporto ~2,6:1).
- Principali stazioni di origine (Constitución 35k, Pacífico 34k) e heatmap spaziali mostrano domanda concentrata nelle aree centrali (Microcentro, Palermo, Recoleta).
- Le origini mattutine si concentrano in zone residenziali/centrali, le destinazioni serali si disperdono, confermando il flusso pendolare bidirezionale.
- Durata dei viaggi: mediana ~16,1 min, media ~21,7 min — la maggior parte sotto i 30 min.
- Distribuzione per genere: Maschi ~60,8%, Femmine ~31,6%.
- Picchi stagionali in tarda primavera/inizio estate (ottobre 335k), minimi in inverno (luglio 215k).
- Estremi di flusso netto (es. Juan Manuel De Blanes +3.056, Cerrito -2.064) evidenziano necessità di riequilibrio.

## Lezioni Apprese

- Le fonti di open data possono presentare inconsistenze (es. incompatibilità degli ID delle stazioni) — l'uso delle coordinate integrate ha garantito copertura completa senza dipendenze esterne.
- Le tecniche di normalizzazione (come le medie per giorno) sono fondamentali per evitare conclusioni fuorvianti quando si confrontano gruppi di dimensioni diverse.
- Il campionamento (10.000 viaggi) raggiunge un equilibrio pratico tra prestazioni interattive e accuratezza analitica nelle visualizzazioni su larga scala.
- Terminologia coerente (es. origini/destinazioni invece di partenze/arrivi) e spiegazioni chiare migliorano significativamente la leggibilità e la comprensione degli stakeholder.

## Prossimi Passi

- Incorporare dati in tempo reale per il riequilibrio dinamico.
- Esplorare la demografia degli utenti (età, genere) e modelli predittivi per la previsione della domanda.

- Espandere al dataset 2025 quando sarà disponibile.

Questo report riproducibile dimostra competenze in R (tidyverse, leaflet), pulizia dei dati, EDA, visualizzazione, dashboard interattivi (Tableau) e business insights — pronto per ruoli da data analyst.

#### **Visualizza il report completo online** (Con heatmap interattive):

Apri il Report Ecobici 2024 sul mio sito web (<https://www.emilionahuelpattini.com/it/data/data-analysis/projects/ecobici-2024/urban-mobility-analysis.html>)

(Versione interattiva con zoom, pan e tooltips — nessun download necessario.)

#### **Dashboard Interattivo**

Esplora l'intero dataset in modo interattivo su Tableau Public: Visualizza il 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](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))

#### **Scarica il Report in PDF**

Per una versione stampabile:

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

(Report completo con tutte le visualizzazioni e gli insight — 12 pagine)