

# EDA Bicicletas Públicas

## ANALISIS DE DATOS

### Introducción

El estudio tiene como objetivo el análisis de los datos registrados del Sistema de Transporte Público de Bicicletas de la Ciudad de Buenos Aires en el portal BA Data con respecto a los recorridos realizados por los usuarios del mismo durante el primer semestre del 2021 a fin de colaborar en la identificación de desafíos para mejorar el mencionado sistema.

### Contexto

#### *¿Qué es Ecobici?*

**Ecobici** es el Sistema de Transporte Público de Bicicletas de la Ciudad de Buenos Aires, que tiene estaciones automáticas para que los usuarios puedan moverse en bici las 24 horas, todos los días del año. El sistema tiene una forma sencilla de utilizar con registración previa a través de una app móvil o de forma presencial, se elige el pase que se prefiera y finalmente se busca una bici en la estación más cercana comenzando a disfrutar de un medio de transporte rápido, ecológico y saludable de transporte.

### Análisis de los datos

Para permitir que las estaciones de bicicletas con tecnología automática lleguen a más barrios y respondan mejor a la demanda de los usuarios mediante la instalación de nueva infraestructura, la incorporación de más bicicletas y la agilización del proceso de retiro y devolución se realizara un análisis exploratorio de los datos obtenidos en el primer semestre de 2021. del uso del sistema.

Para esta evaluación vamos a trabajar con un dataset del portal BA data.

<https://data.buenosaires.gob.ar/dataset/bicicletas-publicas>

La exploración de dichos datos nos guiara a través del análisis completo para la mejora integral del sistema respondiendo algunas de las preguntas como:

1. ¿Dónde viajan los usuarios de **Ecobicis**?
2. ¿Cuándo viajan?
3. ¿Qué tan lejos van?
4. ¿Qué estaciones son las más populares?
5. ¿En qué días de la semana se realizan la mayoría de los viajes?

*Cabe aclarar que tras un análisis de los datos históricos en el periodo 2015al 2020 que se hayan en el portal BA data con respecto al recorrido y a los usuarios del sistema de Ecobicis se encuentran inconsistencias en el registro de los datos como duplicación en el registro de id de usuario en diferentes años, nombres diferentes de campos en diferentes años que se refieren a la misma información, diferentes campos para el mismo dataset, y etc. A modo de sugerencia se aconsejaría estandarizar los datos históricos.*

Más allá de la nota al margen anterior, en virtud de establecer características actuales del uso del sistema se tomará en cuenta el ultimo dataset aportado por el portal de BA Data que es recorridos-realizados-2021.csv.

Este set de datos cuenta con las siguientes variables:

**ID:** Clave de identificación única del registro.

**Estado cerrado:** Estado de finalización del recorrido.

**Duración:** tiempo del recorrido en segundos. I

**Id de estación de inicio:** clave de identificación de la estación de inicio del recorrido.

**Fecha de inicio:** Fecha de inicio del recorrido.

**Nombre de estación de inicio:** Nombre de la estación de inicio del recorrido.

**Fecha de fin:** Fecha de inicio del recorrido.

**Id de estación de fin de viaje:** clave de identificación de la estación de fin del recorrido.

**Nombre de estación de fin de viaje** Nombre de la estación de inicio del recorrido.

**ID de ciclista:** Clave de identificación del usuario del sistema

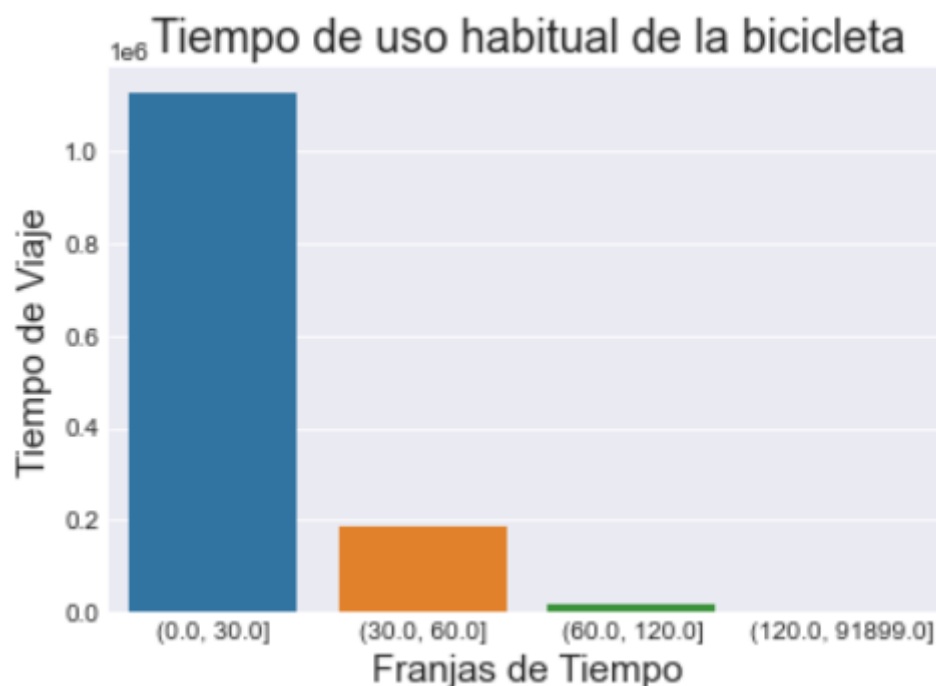
**Tipo de ciclista** Tipo de usuario

**Modelo de bicicleta** Modelo de bicicleta

**Origen de viaje:** origen del viaje

## 1-¿Cuántas horas se suele utilizar la bicicleta en los recorridos?

Como mencionamos anteriormente, contamos con una columna llamada 'Duración' que habla de la duración de cada viaje que cubre, que es en segundos. En primer lugar, lo convertiremos en minutos, luego crearemos buckets para agrupar los viajes en 0–30 minutos, 30–60 minutos, 60–120 minutos, 120 minutos y más. Luego, tracemos un gráfico para ver cuántas horas se recorren en bicicleta habitualmente.



Hay una gran cantidad de ciclistas que viajan menos de media hora por viaje y la mayoría menos de 1 hora.

## 2-Misma ubicación inicial y final versus diferente ubicación inicial y final

Observamos que en los datos que hay algunos viajes que comienzan y terminan en el mismo lugar. Analicemos cuántos para determinar su importancia dentro del marco general del análisis. Esto incide en el tiempo de uso de la bicicleta ya que no es la mismo realizar un recorrido de ida que uno de ida y vuelta.



## 3-Patrón de recorridos mensual.

En este apartado, básicamente, estamos trazando un gráfico de series de tiempo para ver la tendencia del número de viajes realizados por día y la tendencia del número total de tiempo que las bicicletas estuvieron en uso por día.

A continuación, se explica el proceso de confección del grafico paso a paso:

1. La fecha y hora de inicio del viaje la convertimos para poder manipular los componentes de la fecha como el mes, por ejemplo
2. Agrupar los datos por días del mes y contar el número de ocurrencias para trazar viajes por día.
3. Repita los pasos 2 y 3, pero la única diferencia esta vez es que sumamos los datos en lugar de contar para obtener la duración total de los viajes por día.
4. Se traza ambos datos en un solo gráfico utilizando el método de eje doble para su visualización y comparación.



Comparemos el grafico con el calendario donde constan los feriados del mes de abril y los días que corresponden a fines de semana.

ABRIL						
DOM	LUN	MAR	MIE	JUE	VIE	SAB
				1	2	3
4	5	6	7	8	9	10
11	12	13	14	15	16	17
18	19	20	21	22	23	24
25	26	27	28	29	30	

- FERIADOS DE ABRIL
- **Días 28 ,29 ,30 ,31 ,1 ,2 ,3 ,4 .** Pascuas Judías
- **Día 1.** Jueves Santo Festividad Cristiana
- **Día 2.** Día del Veterano en la Guerra de Malvinas
- **Día 2.** Viernes Santo Festividad Cristiana

El número de corredores disminuye considerablemente los feriados y los fines de semana. La cantidad de tiempo que los ciclistas andan en bicicleta disminuye acompaña esta disminución.

## 4- Total de viajes por día de semana.

Si ahora graficamos a nivel global la distribución de la cantidad de viajes de acuerdo al día de la semana,

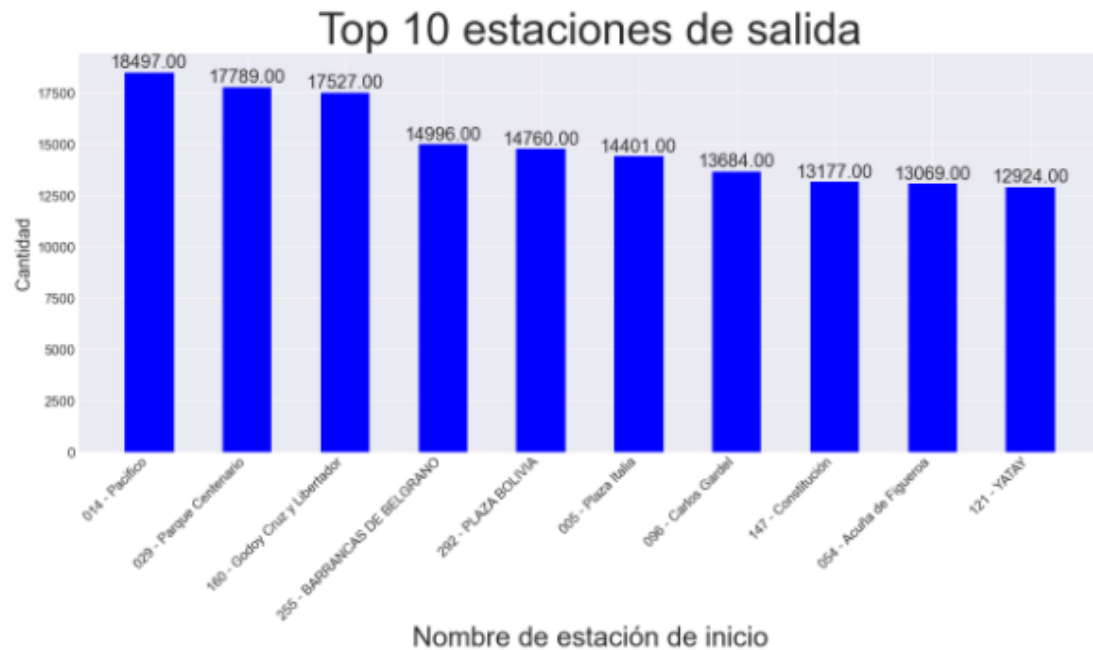


Se observa una clara disminución de la cantidad de recorridos los fines de semana a nivel global ya no mensual.

Se confirma la tendencia del grafico anterior del mes de abril a nivel global.

## 5a- Top 10 estaciones de salida

Esto es bastante sencillo, obtenemos las ocurrencias de cada estación de inicio y dividimos para obtener los primeros 10 valores y luego graficamos lo mismo.



Este grafico es útil para conocer las estaciones donde quizás sea necesario ampliar la infraestructura disponibilizando mas bicicletas.

## 5b- Top 10 estaciones finales.



## Conclusiones

- Hay una gran cantidad de ciclistas que viajan menos de media hora por viaje y la mayoría menos de 1 hora.
- El 93% de los viajes son tienen diferente estación inicial y final.
- El número de corredores disminuye considerablemente los feriados y los fines de semana. La cantidad de tiempo que los ciclistas andan en bicicleta disminuye acompaña esta disminución.
- Se confirma con otro grafico agrupando la cantidad de viajes por día de semana confirmando la disminución del uso de las bicis los fines de semana. Por ejemplo, este dato serviría para determinar periodos adecuados de realización de mantenimiento del sistema. También sugiere un uso del sistema más intenso relacionado a actividades laborales.
- Top 10 estaciones de salida:
  1. 014 - Pacifico
  2. 029 - Parque Centenario
  3. 160 - Godoy Cruz y Libertador
  4. 255 - BARRANCAS DE BELGRANO
  5. 292 - PLAZA BOLIVIA
  6. 005 - Plaza Italia
  7. 096 - Carlos Gardel
  8. 147 - Constitución
  9. 054 - Acuña de Figueroa
  10. 121 - YATAY
- Top 10 estaciones de llegada:
  11. 160 - Godoy Cruz y Libertador
  12. 014 - Pacifico
  13. 029 - Parque Centenario
  14. 255 - BARRANCAS DE BELGRANO
  15. 292 - PLAZA BOLIVIA
  16. 096 - Carlos Gardel
  17. 147 - Constitución
  18. 054 - Acuña de Figueroa
  19. 008 - Congreso
  20. 005 - Plaza Italia