

Desafío MELI

Autora: Dailys Maite Aliaga Reyes

Fecha: 02/06/2022

Desafío 1. BI

Análisis de los datos públicos del sistema ECOBICI publicados en el portal de Datos Abiertos del Gobierno de Buenos Aires.

Objetivo: Realizar un estudio sobre la evolución de los viajes en bicicleta de la Ciudad de Buenos Aires y ciclovías que actualmente hay disponibles, en busca de oportunidades de expansión.

Comprende el período 2013 a 2021.

Las características propias de los datos publicados en el portal, los cuales no están todos completos en la misma dimensión, hacen que los análisis de este informe se realicen en diferentes niveles.

Fuentes de datos:

Se obtuvieron los siguientes *datasets* del repositorio de datos de Argentina <https://data.buenosaires.gob.ar/dataset/>

Otras fuentes:

https://www.buenosaires.gob.ar/sites/gcaba/files/informe_movilidad_ciclista_2010-2020

- ✓ **Conteo ciclistas - GCBA:** Información estimada sobre la cantidad de viajes diarios y anuales en bicicleta, a partir de los operativos realizados por el Gobierno.

<https://data.buenosaires.gob.ar/dataset/conteo-de-viajes-en-bicicleta-por-la-ciudad>

En esta tabla solo tenemos los datos hasta 2019, investigando otras fuentes encontramos los valores de 2020 que fueron añadidos a la tabla.

- ✓ **Recorridos realizados:** Fueron obtenidas 8 tablas referentes a los años 2014 a 2021.

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

- ✓ **Ciclovías:** Referencia geográfica de las ciclovías de la Ciudad.

<https://data.buenosaires.gob.ar/dataset/ciclovias>

- ✓ **Estaciones:** Listado de las estaciones de bicicletas públicas

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

Skills Técnicos: Se utilizó *Microsoft Excel* y *RStudio* para el pre-procesamiento de los datos usando lenguaje *R*. Se utilizaron las librerías (dplyr); (tidyr) (lubridate) y (stringr). Para la construcción del *dashboard* se utilizó la herramienta gratuita de *Tableau Public*.

Pre-procesamiento:

Tablas: recorridos-realizados.csv y trips.csv

Origen, destino, horario, género y edad de los viajes realizados en el sistema de bicicletas públicas de la Ciudad.

- Tratamiento de tablas:
 - Fue criada una tabla única resultado de la UNIÓN de las tablas mencionadas recorridos-realizados y trips.
 - Fue calculado el tiempo de duración del viaje y adicionado a la tabla como una nueva columna.
 - Fue modificado el campo id_origen_estacion para que coincidiera con la tabla estaciones.
 - Fue creado Join entre las tablas (estaciones y recorridos_realizados) con relación muchos a muchos (n:n).

Tabla: Ciclovías.csv (3081 registros, 28 variables).

Referencia geográfica de las ciclovías.

- Análisis de NAs:

WKT	id	codigo	nomoficial	alt_izqini	alt_izqfin	alt_derini
0	0	0	0	0	0	0
alt_derfin	nomanter	nom_mapa	tipo_c	long	sentido	cod_sent
0	1718	2	0	0	16	0
observa	bicisenda	lado_ciclo	recorrid_x	ciclo_obse	tooltip_bi	red_jerarq
2450	0	2131	2903	529	0	0
red_tp	COMUNA	COM_PAR	COM_IMPAR	BARRIO	BARRIO_PAR	BARRIO_IMP
2937	0	0	0	236	19	19

- Fueron eliminadas las columnas con muchos NAs y que no eran relevantes para el estudio (nomanter: nombre anterior de la calle, observa: observaciones particulares de cada calle, ciclo_obse: observaciones)
- Tratamiento de coordenadas:
 - La columna WKT fue dividida en 2 y los datos fueron convertidos de string a double.
 - Fueron criadas dos columnas calculadas (Ponto 1 y Ponto 2) usando las latitudes y longitudes para crear las ciclovías en el mapa.

- La columna COMUNA fue transformada en String y los datos fueron remplazados por una concatenación de string: "Comuna" + Número.

Tabla: nuevas-estaciones-bicicletas-publicas.csv (229 registros, 8 variables).

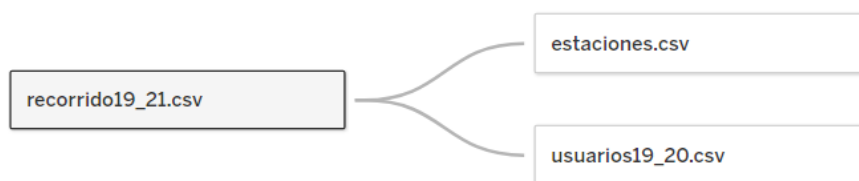
Listado de las estaciones de bicicletas públicas, el cual incluye nombre, ID, ubicación geográfica, dirección, latitud y longitud.

- Tratamiento de coordenadas:
 - La columna WKT fue dividida en 2 y los datos fueron convertidos de string a double.
 - Fue criada la columna calculada (Ponto_estaciones) usando las latitudes y longitudes para ubicar las estaciones en el mapa.

Tabla: usuarios_ecobici.csv (6 tablas con 5 variables cada).

Listado de usuarios del servicio, el cual incluye ID, edad, sexo del usuario, fecha y hora de registro en el sistema.

- Tratamiento:
 - Fue criada una tabla única resultado de la UNIÓN de las tablas usuarios_ecobici de los diferentes años.
 - Fue creado Join entre las tablas (usuarios y recorridos_realizados) con relación muchos a muchos (n:n).
 - Fue modificado el campo id_usuario de la tabla recorridos_realizados para que coincidiera con la tabla usuarios.
 - Fueron normalizados los valores de la columna género.



Dashboard: VisualBike

https://public.tableau.com/views/VisualBike_BA/Viajes?:language=es-ES&:display_count=n&:origin=viz_share_link

EcoBici es un sistema de bicicletas compartidas que funciona en la Ciudad de Buenos Aires. Actualmente cuenta con 229 estaciones.

Panel: General de viajes

En el dashboard, se eligió utilizar una única página que pueda mostrar la evolución de los viajes en bicicleta.



➤ Insights:

Durante el período analizado (2013 a 2020) existe un **aumento de 135% de los viajes anuales**, los años **2018, 2019 y 2020** comprenden la mayor cantidad de viajes, pues es donde parece que **el sistema alcanza una madurez en cuanto a uso**.

En general, hasta la segunda mitad de 2017, los viajes diarios se mantienen por debajo de 200 mil, para luego crecer exponencialmente y lograr un salto de más de **300 mil viajes diarios a mediados de 2019**.

La **duración media** de los viajes mensuales entre 2015 y 2018 es de unos **30 minutos** y entre 2019 y 2021 es de **20 minutos**.

2015 y 2016 son récord de uso respecto a la duración para luego caer en años siguientes. Es destacado que tanto en 2018 y 2019, que como se observó en el gráfico anterior, son años con gran cantidad de viajes realizados, los mismos tengan menor duración promedio. **Más viajes, pero más breves**.

Entre la segunda mitad del 2019 y principios del 2020, hay un marcado aumento de la duración de los viajes. Los usuarios reemplazaron en parte el transporte público por las bicicletas. Pero esto posiblemente sea resultado de los cambios que trajo la pandemia y no significa que sea un patrón.

Hasta 2018 inclusive el uso del sistema durante el día muestra una distribución pareja entre la mañana y la noche, con un pico bien significativo entre las 5 y 6 de la tarde.

A partir de 2019, se **vuelven a presentar cambio de hábitos**, disminuyendo la proporción de viajes durante la mañana, y acrecentándolos durante la tarde y noche, principalmente entre las 12 del mediodía y las 11 de la noche. Resalta el crecimiento sostenido durante la noche (a partir de las 20 horas).

Panel: Factores que influncian

En el dashboard, se eligió utilizar una única página que pueda mostrar la evolución de las ciclovías y estaciones que componen actualmente el sistema ECOBICI y se analizan los viajes realizados durante los años 2019 - 2020

- Mapa de ciclovías y estaciones



➤ Insights:

- A pesar de la extensión actual de las estaciones
- Se puede observar **escases de ciclovías y estaciones en barrios del oeste y del sur** de la Ciudad
- Además hay ciclovías que no tienen estaciones ECOBICI instaladas actualmente como por ejemplo: Mataderos, Liniers, Villa Luro y Villa Lugano. También existen estaciones aisladas con pocas ciclovías cercas.
- Si analizamos la cantidad de estaciones que tienen instaladas en su perímetro los barrios, vemos cierta **relación** entre la **disponibilidad de las estaciones** con la **cantidad de viajes** realizados.

- Palermo (29), San Nicolás (19), Recoleta (17) y Balvanera (13) son los barrios que tienen una mayor cantidad estaciones y la densidad de viajes está levemente concentrada.
- Geografía de los viajes:
 - **Insights:**
 - **Facultad de derecho** es la estación con **mayor cantidad de viajes iniciados y Constitución de viajes finalizados**.
 - California, Parque Las Heras, Plaza Italia, Pacífico y Parque Centenario mantienen un volumen destacado, aunque más lejano que el que administra los líderes antes mencionados.
 - Es posible ver que **las estaciones** donde se inician la mayoría de los viajes están ubicadas en la **zona Norte** de la ciudad.
- Densidad de viajes:
 - **Insights:**
 - Si analizamos geográficamente los viajes realizados podemos notar que en los barrios de **la zona Norte y Este de la Ciudad, hay mayor concentración** de éstos. Todo lo contrario en las zonas Oeste y Sur.
 - Habiendo revisado previamente la ubicación geográfica de las estaciones y de las ciclovías, podemos establecer una **relación clara entre el desarrollo del sistema en cuanto a su ubicación y la cantidad de viajes realizados**.
 - Un ejemplo de esto es el barrio de Palermo. Si bien es el de mayor superficie de la Ciudad, también lo es en cuanto a estaciones ECOBICI. Esto explica claramente el motivo por el cual tiene una alta densidad de viajes durante los años 2019-2020.
- Características de los usuarios:
 - **Insights:**
 - Es **mayor** el uso del sistema por parte de los **hombres**. Es de destacar que la población de la Ciudad de Buenos Aires está compuesta en proporciones inversas (menos hombres y más mujeres el último Censo Nacional).
 - La mayor cantidad de usuarios tienen entre 20 y 40 años de edad que suman el 68% del total del sistema.
 - Pasados los 40 años de edad, **la cantidad de usuarios decae a medida que crece la edad**.
 - No hay diferencias significativas respecto del género.
 - Independiente mente del genero 2019 fue un año de crecimiento exponencial. Ya 2020 nos muestra que la media de mujeres que hacían uso del sistema fue mayor que la de hombres.

- Los hombres también tienden a hacer recorridos más largos que las mujeres.

Observaciones sobre el Dashboard VisualBike:

Un adicional muy interesante sería usar encuestas realizadas a ciclistas para estimar los motivos de viajes. Con estos datos conseguiremos expandir el sistema ecobici y podrías relacionar el aumento de viajes con puntos estratégicos de la ciudad, como Universidades, empresas y más.

Relacionar también el uso de bicicletas con los accidentes de ciclistas usando este tipo de bicicletas nos ayudaría a tomar decisiones sobre la expansión del sistema y la depreciación de los activos.

Observaciones sobre presentación del Dashboard VisualBike:

En la presentación del Dashboard voy a construir una historia para contar sobre las conclusiones.

Conclusiones:

El sistema tiene un **crecimiento notorio entre 2019 y 2020**, el cual podría haber sido mayor de no ser la cuarentena y las medidas de salud.

Este crecimiento se observa tanto en la cantidad de viajes realizados, como en los usuarios activos del sistema.

1. La duración promedio de los viajes **cae** a partir de 2017 pero comienza a aumentar en 2019.
2. Hay un crecimiento en la proporción de los **viajes realizados después de las 20 horas**.

El **usuario** característico del sistema **es joven**, de 20 a 40 años y en menor medida, femenino.

Finalmente, si bien la red y el sistema se amplió en los últimos años, todavía **deja fuera a barrios del sur y del oeste de la Ciudad**.

Se nota que las zonas Norte, Centro y Este de la Ciudad hacen un uso sostenido del sistema.

Teniendo en cuenta que en las zonas mencionadas el transporte público es algo más abundante, sería de **importancia estratégica desarrollar** el sistema en las **zonas Sur y Oeste**. Allí el transporte es más escaso, y los viajes en bicicleta podrían ayudar a la **integración y comunicación de esos barrios**.