

# **Monitoreo de las estaciones de bicicletas más concurridas de Buenos Aires en el período 2020**

**REALIZADO POR**

Ariel Hernandez  
Barbara Vilaro  
Camila Casanovas

**PEDIDO POR**

Gobierno de la Ciudad de Buenos Aires

# ÍNDICE

<b>Introducción</b>	<b>1</b>
Equipo de analistas	1
Objetivo del análisis	2
<b>Bases de datos</b>	<b>3</b>
Dataset	3
Entidades identificadas	4
Tablas realizadas	4
Diagrama Entidad Relación	5
<b>Dashboard</b>	<b>7</b>
Alcance	7
Usuarios	7
Transformaciones realizadas en Power BI	7
Modelo relacional en Power BI	10
Medidas calculadas	10
Segmentaciones elegidas	11
Descripción del tablero final	11
<b>Conclusión</b>	<b>12</b>
Recomendaciones finales	12

# INTRODUCCIÓN

El análisis realizado en este proyecto se centra en el servicio de bicicletas ofrecido por el Gobierno de la Ciudad de Buenos Aires. En el mismo se analiza el nivel de servicio de las estaciones que conforman la red de puntos de salida y entrada de cada bicicleta con el fin de conocer si la cantidad de módulos -puntos de recepción de bicicletas- es suficiente.

Esta inquietud surge del mismo ente gubernamental por la necesidad de conocer si existen aglomeraciones en las estaciones o si el usuario final puede, en la mayoría de los casos, dejar la bicicleta en la estación más cercana sin la necesidad de esperar a que se desocupe alguno o tenga que trasladarse a otra estación para devolverla. Sumado al creciente auge que tiene este medio de transporte en otras ciudades del mundo y que, se intenta promover para disminuir de algún modo el tránsito de vehículos en la Ciudad.

El análisis comienza primeramente con la limpieza y transformación de los datos crudos obtenidos y finaliza con un tablero de control dinámico que visualiza las principales conclusiones del estudio. A su vez, se anexan en este documento, algunas recomendaciones a realizar a partir de los resultados encontrados.

## ***Equipo de analistas***

El equipo de desarrollo del proyecto está integrado por diferentes perfiles con experiencia en el sector privado. Esto permite encarar el proyecto de una manera más objetiva y distante, basando su análisis y conclusiones en los datos encontrados.

Sin embargo, al tratarse de un ámbito estatal es posible que se deban tener en cuenta otras variables, de carácter social, que permitan ampliar la visión del proyecto en pos de mejorar la calidad de vida de los ciudadanos. Tal análisis cualitativo se deja a criterio de quien utilice el dashboard.

### ***Objetivo del análisis***

El motivo principal de este análisis es conocer cuáles son las estaciones de bicicletas que más alquileres de bicicletas tuvieron en 2020. De esta manera se podría aconsejar, al ente pertinente, si sería útil poner más módulos en tales estaciones o evaluar la posibilidad de dividir la misma, para equilibrar la atención de los clientes.

# BASE DE DATOS

## Dataset

El análisis se realizó con los datos recopilados por el mismo ente gubernamental y que ofrece para su descarga gratuita en su página web. Asimismo se toman en cuenta sólo los datos del año 2020 ya que es la información más reciente del que se tiene registro completo.

Esta base de datos consta de 1.048.576 filas y 15 columnas, mostrando información sobre todos los viajes realizados durante el año. En la tabla siguiente se muestra un esquema de cómo es la base de datos original tal como fue descargada y los tipos de datos que contiene:

Campo	Descripción del campo	Tipo de datos
periodo	Es el año de la base de datos.	Int
id_usuario	Es un número único que identifica a cada usuario.	Int
fecha_origen_recorrido	Es el registro del momento en que el usuario busca una bicicleta en un módulo.	Timestamp
id_estacion_origen	Es la identificación única de cada estación desde donde se busca una bicicleta.	Int
nombre_estacion_origen	Es el nombre con el que habitualmente se la conoce a cada estación, desde donde se busca una bicicleta.	Text
long_estacion_origen	Es una de las coordenadas necesarias para ubicar la posición geográfica exacta desde donde se busca la bicicleta, expresada en grados, minutos y segundos. Va de Este a Oeste.	Decimal
lat_estacion_origen	Es otra de las coordenadas necesarias para ubicar la posición geográfica exacta desde donde se busca la bicicleta, expresada en grados, minutos y segundos. Va de Norte a Sur.	Decimal
domicilio_estacion_origen	Es el lugar a donde va el usuario local para buscar su bicicleta	Text
duracion_recorrido	Es el tiempo que el usuario usó la bicicleta	Int
fecha_destino_recorrido	Es el registro del momento en que el usuario estaciona una bicicleta en un módulo.	Timestamp

id_estacion_destino	Es la identificación única de cada estación desde donde se recepciona una bicicleta.	Int
nombre_estacion_destino	Es el nombre con el que habitualmente se la conoce a cada estación, desde donde se recepciona una bicicleta.	Text
long_estacion_destino	Es una de las coordenadas necesarias para ubicar la posición geográfica exacta desde donde se recepciona la bicicleta, expresada en grados, minutos y segundos. Va de Este a Oeste.	Decimal
lat_estacion_destino	Es otra de las coordenadas necesarias para ubicar la posición geográfica exacta desde	Decimal

#### Referencias:

- Int: número entero
- Text: datos de texto
- Decimal: número con decimales
- Timestamp: Fecha y horario exacto en donde se realizó cada acción

#### **Entidades identificadas**

A partir del dataset anterior se identificaron las siguientes entidades:

1. Usuario: es la persona que utiliza el servicio de bicicletas
2. Estaciones: son los puntos de retiro o devolución de las bicicletas
3. Viajes: es la actividad registrada por cada usuario, cada vez que utiliza el servicio.

#### **Tablas realizadas**

Las entidades antes mencionadas permiten ordenar la información de una manera entendible y útil para el análisis, con lo que sirvieron de base para crear las siguientes tablas y así clasificar los datos:

##### Estación

Contiene toda la información acerca de cada una de las estaciones. Se crea esta tabla para centrar toda la información que en el dataset se repite en cada viaje. De esta manera se puede realizar el análisis utilizando el ID de estación y luego relacionarlo con esta tabla. La clave primaria elegida es id\_estacion ya que es un dato único para cada estación y ésta no se repite.

id_estacion (PK)	nombre_estacion	direccion_estacion	long_estacion	lat_estacion
Int	Text(50)	Text(100)	Decimal(15,2)	Decimal(15,2)

### Alquileres

Es el conjunto de los viajes realizados desde una determinada estación a otra, considerando la fecha y el horario en que se registró cada acción realizada por el usuario. En esta tabla fue necesario crear un registro único e irreplicable para que sirviera como referencia de clave primaria, con lo cual se crea una nueva columna con la correlación de números enteros que identifiquen cada viaje. Es por esto que id\_recorrido es la clave primaria y id\_usuario es la clave foránea, ya que se relaciona con la tabla anterior y es clave primaria de la misma.

duracion_recorrido	id_estacion_origen	fecha_origen_recorrido	fecha_destino_recorrido	id_estacion_destino	id_usuario (FK)	id_recorrido (PK)
Int	Int	Timestamp	Timestamp	Int	Int	Int

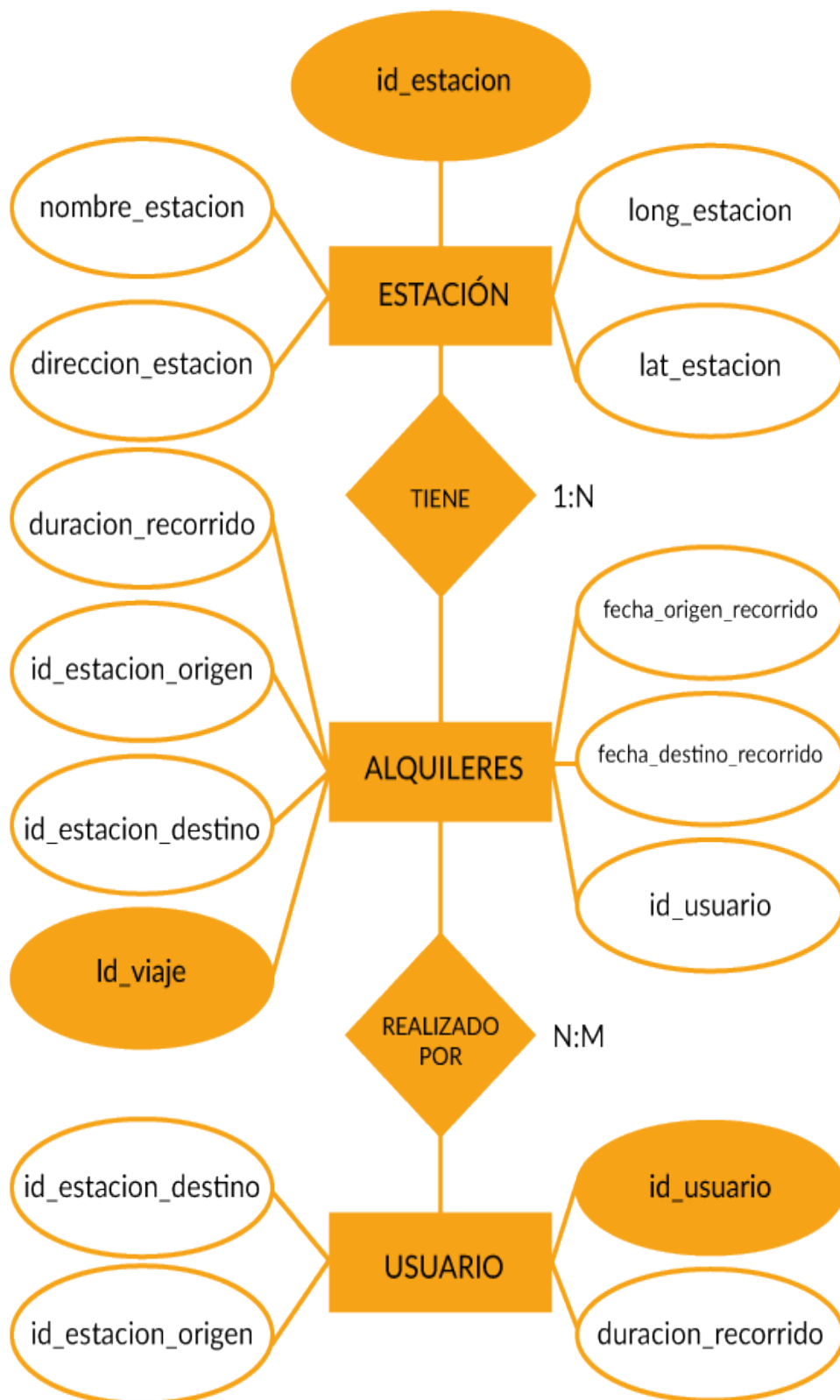
### Usuario

Comprenden todos los datos relacionados a las personas que utilizan el servicios de Bicicletas, en el mismo no se registran datos de carácter personal, pero si se tienen en cuenta los movimientos realizados por cada usuario, durante el 2020, en relación a este servicio. La clave primaria es id\_usuario.

id_usuario (PK)	id_estacion_origen	id_estacion_destino	duracion_recorrido
Int	Int	Int	Int

### Diagrama Entidad Relación

Este diagrama permite saber cómo se relacionan cada una de las entidades. Para nuestro proyecto es el siguiente:



Las relaciones entre entidades son las siguientes:

Estación - Alquileres: una a muchas

Alquileres- Usuarios: muchos a muchos



# DASHBOARD

## **Alcance**

El alcance de este análisis se extiende a todas las estaciones de bicicletas ubicadas en la Ciudad de Buenos Aires y a todos los usuarios que utilicen el servicio, para analizar si existen cuellos de botellas en la prestación.

## **Usuarios**

Este proyecto está dirigido a quienes tengan la potestad de hacer cambios en el servicio de préstamos de bicicletas en Buenos Aires, ya que el mismo brinda información acerca de las estaciones y el flujo de movimiento.

El dashboard presentado les permitirá visualizar oportunidades de mejoras en dónde podrían aplicarse reformas en el servicio, tales como sumar más módulos en ciertos días pico o grandes transformaciones cómo evaluar separar una estación en dos cercanas, para evitar aglomeraciones o cuellos de botellas.

## **Transformaciones realizadas en Power BI**

Las transformaciones realizadas fueron:

1. Se creó una nueva columna con la que se modificaron todos los datos de ID de estación de origen y llegada, ya que algunos datos estaban mal registrados. Para ello, se dividió la columna por el delimitador "-" y se renombraron las dos nuevas columnas como "id\_estacion\_origen" y "nombre\_estacion\_origen", lo mismo para "id\_estacion\_destino" y "nombre\_estacion\_destino" (Ver figura 1).
2. Se creó la tabla "coordenadas" (Ver figura 4) y se agregaron dos columnas, una que nos indica la longitud de id y otra indicándonos la latitud del id, el cual es utilizado para calcular de manera exacta las estaciones dentro de nuestro gráfico "Mapa de estaciones":

●  $\text{longitud\_id} = \text{left}(\text{Coordenadas}[\text{long\_estacion}], 8) * 10$   
(Ver figura 2)

● `latitud_id = left(Coordenadas[lat_estacion],8)*10`

(Ver figura 3)

Figura 1

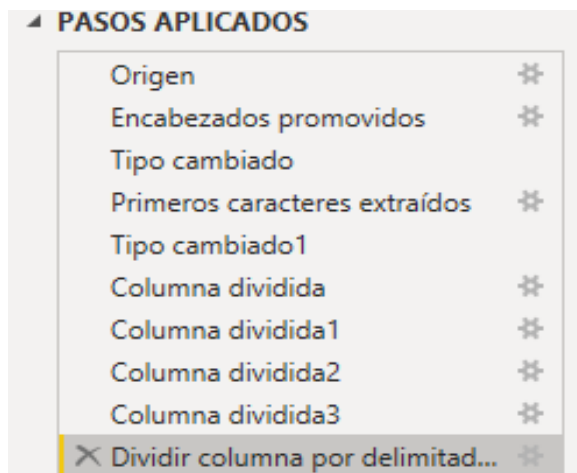


Figura 2

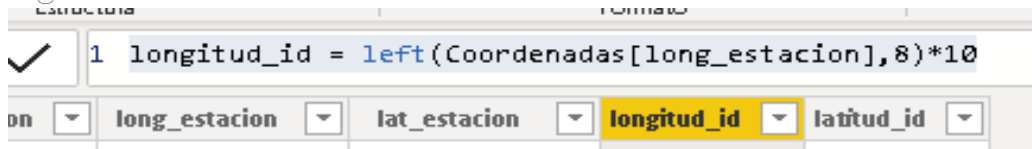


Figura 3

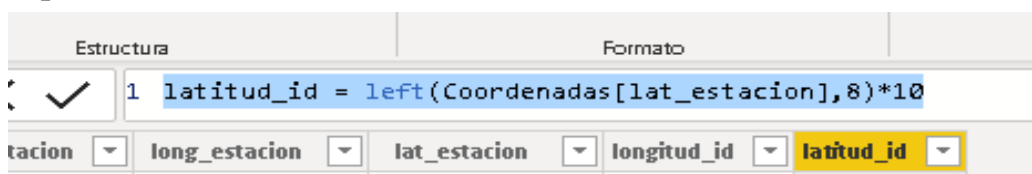
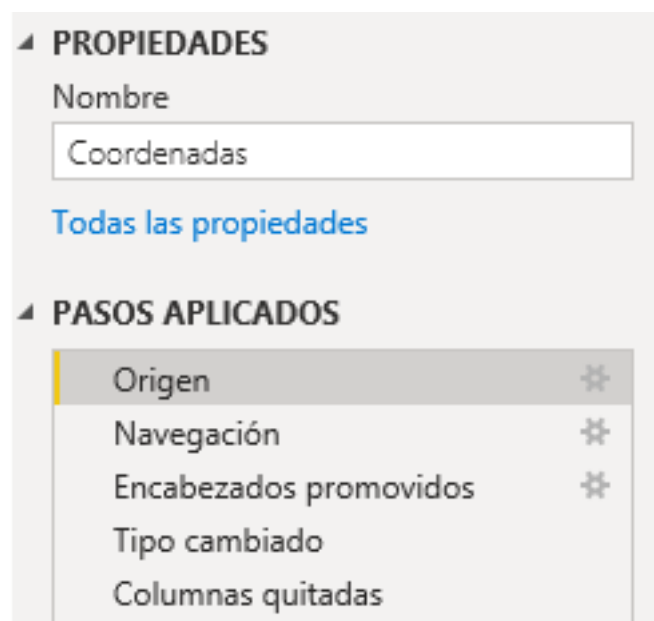


Figura 4



3. Creamos una nueva tabla llamada “Calendario”, la cual utilizamos para calcular y separar de manera numérica el “día”, “mes” y “trimestre” donde el usuario hizo uso de la bicicleta:

- Calendario = CALENDAR(MIN(Alquileres[fecha\_origen\_recorrido]),TODAY()) (Ver figura 5)
- Mes = MONTH(Calendario[Fecha origen]) (Ver figura 6)
- Dia = DAY(Calendario[Fecha origen]) (Ver figura 7)
- Trimestre = QUARTER(Calendario[Fecha origen]) (Ver figura 8)

4. En la tabla “Alquileres” agregamos las siguientes columnas, las cuales se utilizan para calcular las medidas y la realización de los graficos del dashboard final: (Ver figura 9)

- Hora Inicio = RIGHT(Alquileres[fecha\_origen\_recorrido],8)
- Hora Termino = RIGHT(Alquileres[fecha\_destino\_recorrido],8)
- Fecha inicio = left(Alquileres[fecha\_origen\_recorrido],10)
- duracion\_recorrido\_horas = Alquileres[duracion\_recorrido]/3600

Figura 5



Figura 6

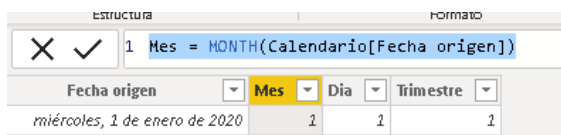


Figura 7

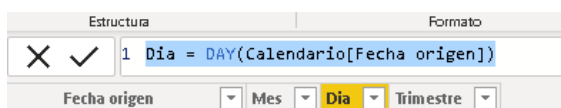


Figura 8

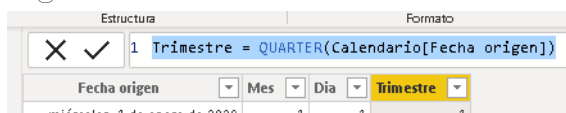
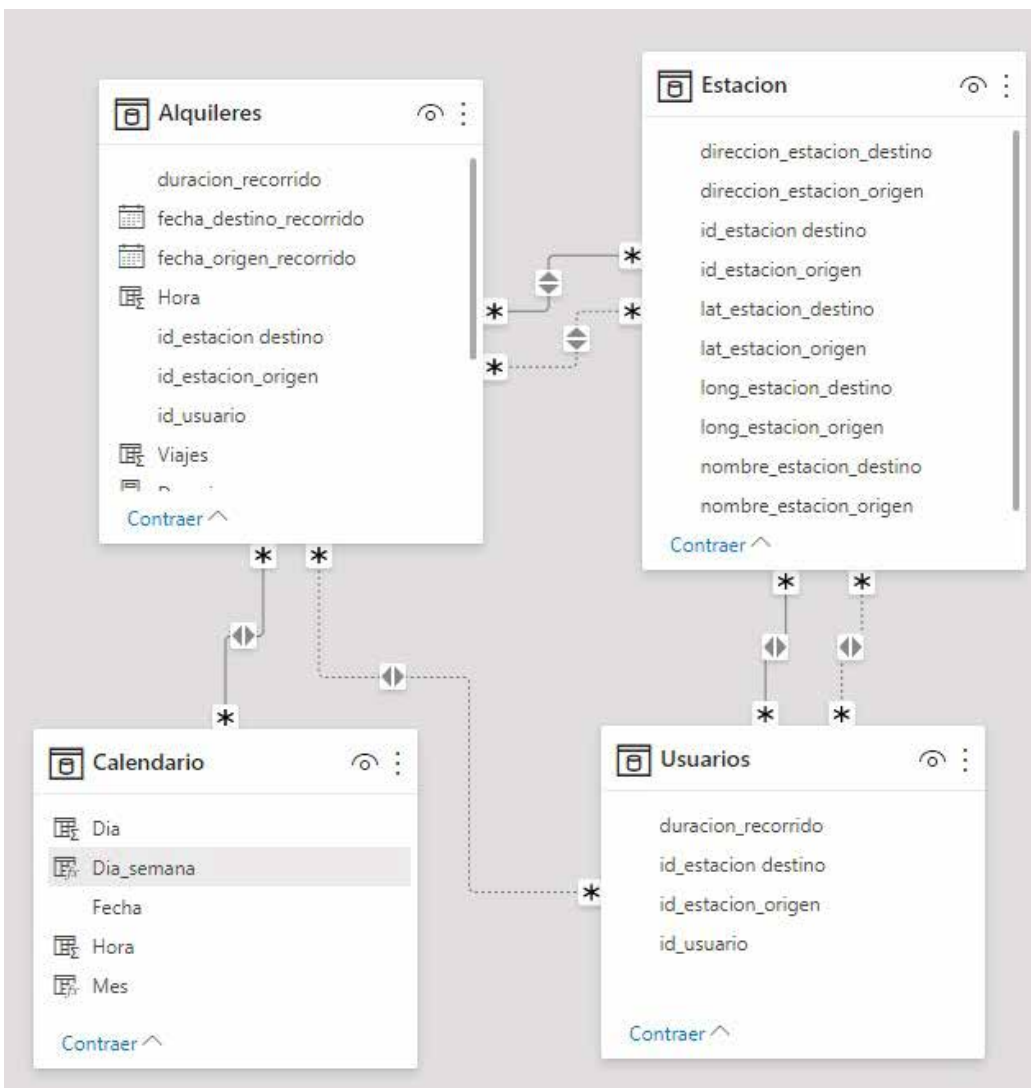


Figura 9

<b>Hora Inicio</b> ▼	<b>Hora Termina</b> ▼	<b>Fecha inicio</b> ▼	<b>duracion_recorrido_horas</b> ▼
----------------------	-----------------------	-----------------------	-----------------------------------

### Modelo relacional en Power BI

En la pestaña “Todas las tablas” se termina de definir el modelo de entidad relacional. La herramienta no consiguió reconocer las relaciones y por tal motivo se la relaciona de manera manual. El resultado fue el siguiente diagrama:



### Medidas calculadas

Una vez obtenido el diagrama y la interacción de los datos, se analizan las medidas adicionales que serán útiles para el análisis del dashboard, las cuales se detallan a continuación:

- Medida = GEOMEAN(Alquileres[duracion\_recorrido])

- Medida 3 = CALCULATE([Recuento de Viaje],PREVIOUS-MONTH(Alquileres[fecha\_destino\_recorrido].[Date]))
- MesPrevio = CALCULATE([Recuento de Viaje],PREVIOUS-MONTH(Alquileres[fecha\_destino\_recorrido].[Mes]))
- Recuento de id\_estacion\_origen para 2 =CALCULATE(COUNTA('Estacion'[id\_estacion\_origen]),'Estacion'[id\_estacion\_origen] IN { 2 })
- Recuento de Viaje =COUNTA('Alquileres'[Viajes])
- Total acumulado de Recuento de Viaje en Mes =CALCULATE(-COUNTA('Alquileres'[Viajes]), FILTER(ALLSELECTED('Calendario'[Mes]),ISONORAFTER('Calendario'[Mes], MAX('Calendario'[Mes]), DESC) ))

Todas estas medidas fueron utilizadas para representar en el dashboard final “análisis global” representado en los distintos gráficos de la solapa “gráficos 1”.

### ***Segmentaciones elegidas***

Las segmentaciones realizadas según nuestro enfoque de análisis están relacionadas al tiempo, para analizar los datos según los meses del año, y así establecer si la utilización del servicio depende del clima.

### ***Descripción del tablero final***

Realizamos la visualización de nuestro dashboard en dos solapas, las que nombramos como “Gráficos 1” y “Gráficos 2”

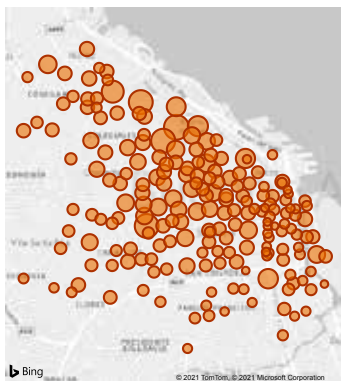
En la pestaña “Gráfico 1” se realiza un análisis de monitoreo global de las estaciones. Con tres kpi's en la parte superior: Viajes Totales, Viajes por Día y Duración promedio. Además, se visualiza la zona geográfica en donde se ubican, a través del visor de mapa, donde el tamaño del círculo indica la cantidad de viajes. También, aparecen el ranking de las principales estaciones de origen y destino. Finalmente la cantidad de viajes por mes.

## Monitoreo de Estaciones de Bicicletas - Buenos Aires

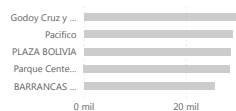


Estacion

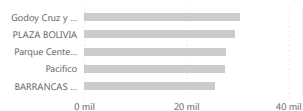
Todas



Top Estaciones de Origen



Top Estaciones de Destino



Viajes por Mes

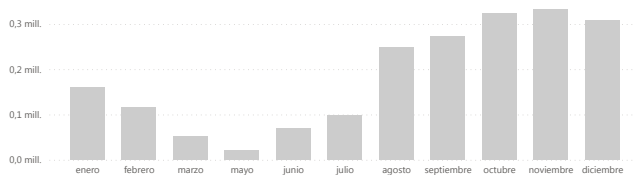
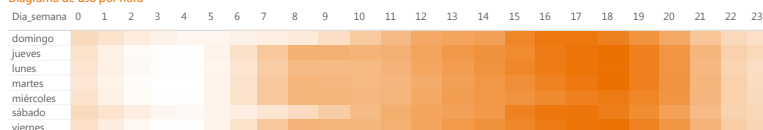
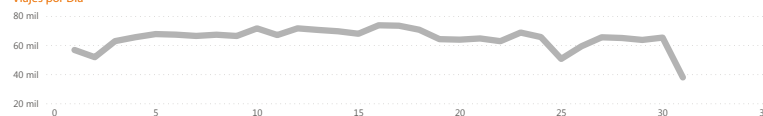


Diagrama de uso por hora



Viajes por Día



### Conclusión

Los principales resultados que arroja la visualización de los datos son:

1. Las estaciones Godoy Cruz y Libertador, Pacífico y Plaza Bolivia, tienen mayor movimiento que las demás, cercanas a ciertos lugares con gran circulación en la Ciudad. Sin embargo, en general todas tienen un movimiento más o menos similar. En verano se producen la menor cantidad de viajes por periodo vacacional.
2. En abril no se registraron viajes por la cuarentena estricta. Posteriormente, subieron proporcionalmente hasta alcanzar el pico máximo en noviembre.

Considerando las estaciones mayormente concurridas mencionadas anteriormente:

3. La mayor aglomeración de transacciones se produce por la tarde, coincidiendo con el horario de salida laboral entre 16:00 y 18:00 horas.
4. En general, el usuario devuelve la bicicleta a la misma estación donde la obtuvo. Cuando no se produce este patrón, las estaciones concurridas también están en zonas con alto movimiento en la ciudad.

### ***Recomendaciones finales***

Las sugerencias que, como equipo de análisis de datos, podemos dar son:

1. Aumentar la cantidad de módulos de las estaciones más cálida del año, con instalaciones nuevas o reutilizando los de estaciones menos concurridas.
2. Se deberían programar los mantenimientos necesarios durante el invierno, como puesta a punto para la temporada alta.
3. También se podría promover el uso de las bicicletas con algún tipo de beneficio, en estaciones con menos actividad.

