



Trabajo Integrador

Argentina Programa 4.0



Estudiante: Sergio Plaza

Comisión: Jueves 17 a 19 hs

Docentes: Melisa Frisoli

Fecha de entrega: 24/08/2023

Resumen

La empresa "BikeShare" tiene como objetivo mejorar el servicio y aumentar la satisfacción de los usuarios de su servicio. Para analizar la situación, se solicitó el análisis de la base de datos de uso de sus sistemas, la cual contiene información sobre los viajes realizados, como la duración, estaciones de inicio y fin, fechas y horas de lo mismos, entre otros datos.

De acuerdo a la solicitud de la gerencia, se identificaron y analizaron los siguientes aspectos del servicio, dentro del marco de tiempo que abarca la base de datos:

- a. **Cantidad de viajes de los usuarios de tipo Miembro y de tipo Casual que duran 5 minutos o más.** Se observó que la cantidad de viajes realizados por usuarios Miembro es mayor a la cantidad por usuarios Casuales, en una relación 3.5:1.
- b. **Bicicleta en la que se realizó el viaje más largo, y duración del mismo.** Se identificó el ID de la unidad en cuestión, notando en el registro que el lapso transcurrido desde su retiro hasta su devolución fue de aproximadamente 24 hs.
- c. **ID's de las 3 estaciones con la mayor duración de viajes, que comienzan y terminan ellas mismas.** Se presenta en esta sección un gráfico de barras con la información indicada.
- d. **ID's de las 5 estaciones de partida que poseen los viajes de mayor duración, de usuarios de tipo Miembro.** Se presenta en esta sección un gráfico de barras con la información indicada.
- e. **Nombre de la estación en donde comienzan y terminan la mayoría de los recorridos.** Se identificaron ambas estaciones de forma individual, observando que en ambos casos la estación es la misma (ID 31200), concluyendo así que es la más utilizada por todos los usuarios.
- f. **Geolocalización de las dos estaciones más populares** (las que tienen mayor número de viajes que parten de ellas). De la información obtenida en el mapa, se confirma que dichas estaciones se encuentran ubicadas en el centro neurálgico de la ciudad, debido al volumen de empresas y negocios que se encuentran en la zona.
- g. **Visualización de los primeros registros de la base de datos**, con todos los viajes disponibles de usuarios tipo "Miembro", y **representación gráfica comparativa de la cantidad de viajes efectuados, agrupados por día de la semana**, para ese tipo de usuarios. Se detectó que en fines de semana el uso de las unidades descende en promedio un 22% aproximadamente, con respecto a los días hábiles.

Adicionalmente, el equipo de análisis incluyó además los siguientes ítems, que podrían ser de utilidad:

- h. **ID's de las 3 estaciones con mayor duración de viajes que comienzan y terminan en ellas mismas, a nivel histórico** (información agregada). Complementando el ítem 3, se observó que el resultado en este caso es distinto al de aquel.
- i. **Representación gráfica comparativa de la cantidad de viajes agrupados por mes, para ambos tipos de usuario.** Se identificó que durante los meses de invierno el uso del servicio descende considerablemente, y a su vez asciende de la misma forma en primavera / verano. Esta información es útil para optimizar la gestión del mantenimiento de bicicletas y estaciones, de acuerdo a la intensidad de uso.

1. Objetivos y metodología de trabajo

El objetivo del análisis de la información mencionada en el apartado anterior es entender mejor el comportamiento de los usuarios y las tendencias en el uso de las bicicletas compartidas.

La gerencia de BikeShare utilizará estos resultados para mejorar el servicio y hacerlo más eficiente.

Para el estudio de la base de datos, se ha utilizado el lenguaje de programación Python + SQL, incluyendo las librerías necesarias para: Importar la información de acuerdo a filtros específicos; realizar cálculos entre distintos campos (columnas) para generar nueva información; generar gráficas y mapas; entre otras.

El **flujo de trabajo** aplicado responde al siguiente orden de actividades:

1. Importación y análisis exploratorio de datos
2. Limpieza de datos si es necesario
3. Procesamiento de la información de acuerdo a los requerimientos
4. Presentación de resultados y conclusiones

2. Análisis exploratorio de datos

La base de datos contiene 2 tablas, a saber:

trip_data: Contiene toda la información relativa al uso del servicio, incluyendo ID de bicicleta utilizada, fechas de inicio y fin de sesión de uso, tiempo transcurrido (duración) entre inicio y fin de cada sesión, estaciones de inicio y fin de cada sesión, y tipo de usuario (miembro o casual). Registros totales: 1.226.767

bikeshare_stations: Incluye información relativa a los IDs, nombres, coordenadas geográficas, y estado (abierta o cerrada) de todas las estaciones que posee la empresa. Registros totales: 144

Durante el procesamiento de la tabla *trip_data* se detectaron 3 registros en donde no estaba identificado si el usuario fue Member o Casual, indicando en su lugar el valor "Unknown". Dichos registros fueron excluidos del análisis acerca de la distribución de viajes por tipo de miembro. A su vez se detectó que al menos el ID de 1 bicicleta no fue registrado por el sistema correctamente; sin embargo, dicha anomalía no afectó a los análisis, pero se sugiere investigar más a fondo el hecho con el personal encargado de la programación y mantenimiento del sistema.

Se presentan a continuación los resultados a las consultas realizadas por la Gerencia:

a) Cantidad de viajes de los usuarios de tipo Miembro y de tipo Casual que duran 5 minutos o más, en porcentaje y valores específicos (Figura 1):

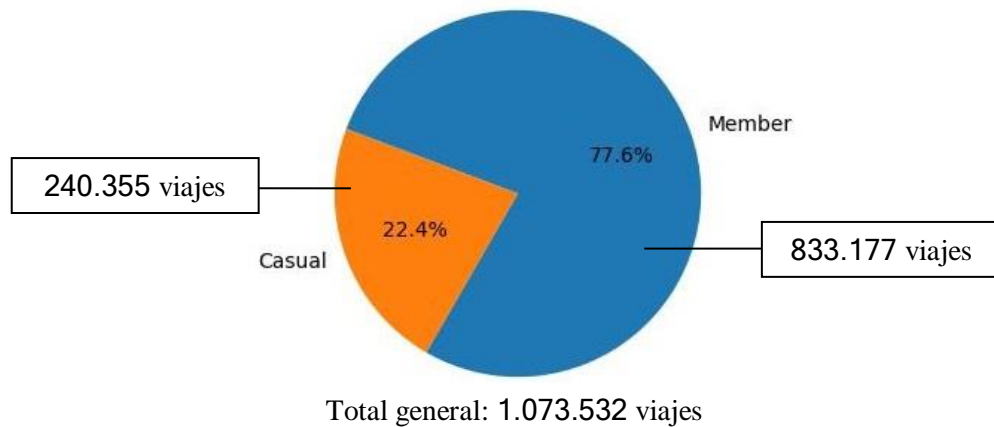


Figura 1: Cantidad de viajes superiores a 5 minutos, según tipo de usuario

b) Bicicleta en la que se realizó el viaje más largo, y duración del mismo:

Se detectó que la bicicleta **ID W00379** es la que registró el lapso de tiempo más extenso entre su retiro y su devolución, siendo el mismo de **1439 minutos** (23:59 hs).

c) ID's de las 3 estaciones con la mayor duración de viajes, que comienzan y terminan en ellas mismas:

Los ID de las 3 estaciones consultadas, de mayor a menor duración (en viajes individuales) son: 31221, 31705, 31617 (Figura 2).

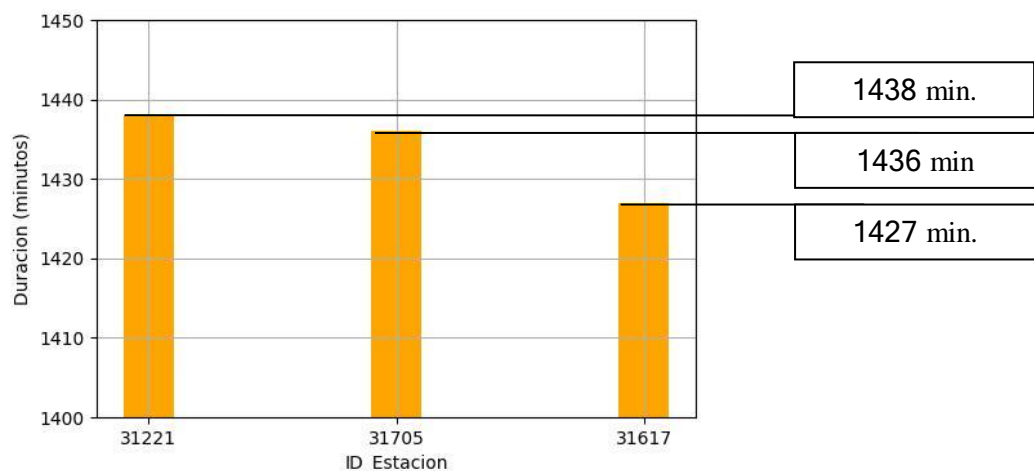


Figura 2: Estaciones con la mayor duración de viajes individuales, que comienzan y terminan en ellas mismas.

d) ID's de las 5 estaciones de partida que poseen los viajes de mayor duración , de usuarios de tipo Miembro:

Los ID en cuestión son los siguientes, de mayor a menor duración: 31232, 31400, 31104, 31200 y 31102 (Figura 3).

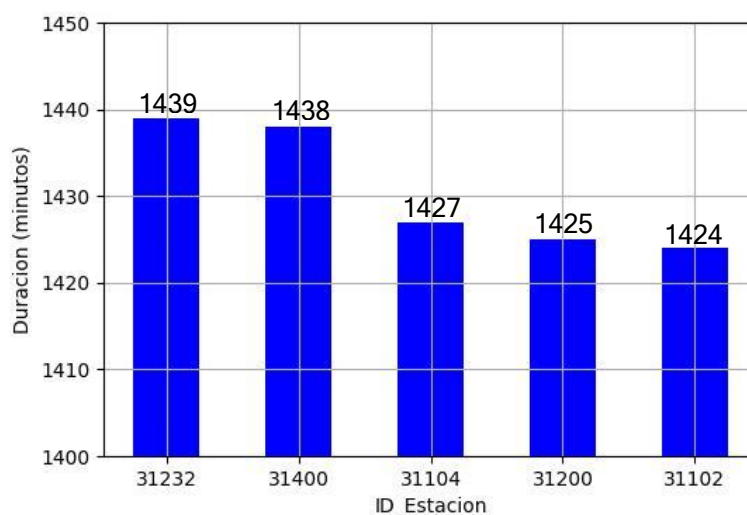


Figura 3: Estaciones de partida que poseen los viajes de mayor duración (usuarios tipo Miembro).

e) Nombre de la estación en donde comienzan y terminan la mayoría de los recorridos:

Se detectó que la estación **Massachusetts Ave & Dupont Circle NW** (ID 31200), es tanto la que posee la mayor cantidad de partidas (49093) como de llegadas (53910).

f) Geolocalización de las 2 estaciones con mayor cantidad de partidas (Figura 4):

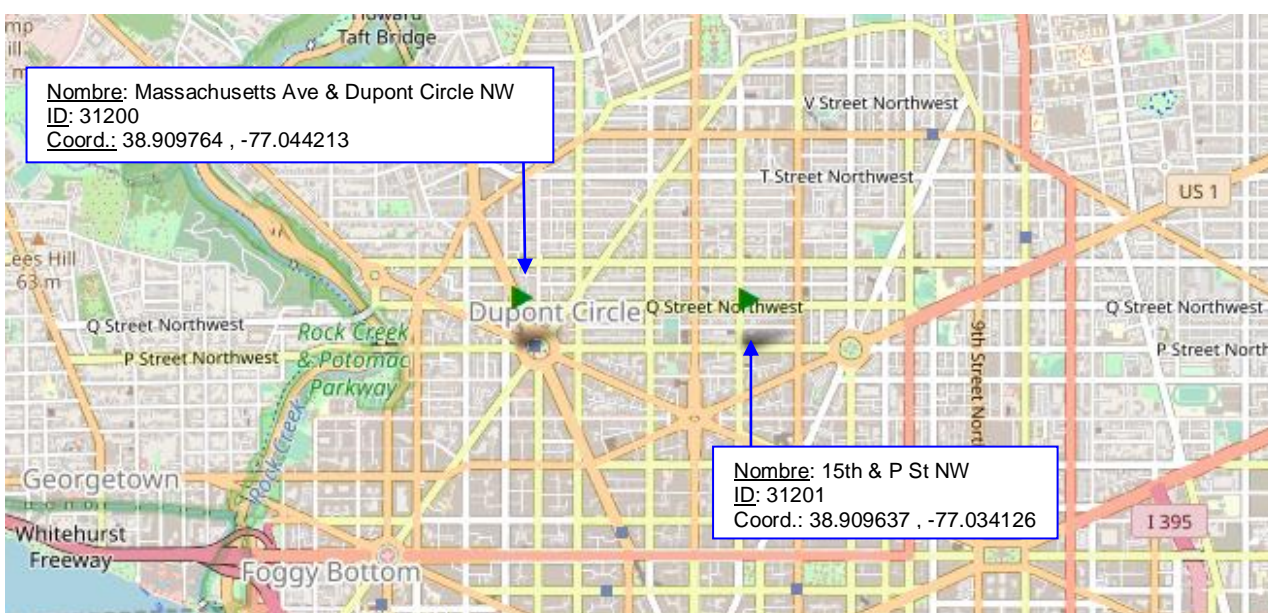


Figura 4

g) Visualización de los primeros registros de la base de datos de usuarios Miembro (Figura 5), y representación gráfica comparativa de la cantidad de viajes efectuados por los mismos, agrupados por día de la semana (Figura 6):

index	duration	start_date	end_date	start_station	end_station	bike_number	member_type
0	3548	2011-01-01 00:01:29	2011-01-01 01:00:37	31620	31620	W00247	Member
2	562	2011-01-01 00:06:13	2011-01-01 00:15:36	31400	31104	W00357	Member
3	434	2011-01-01 00:09:21	2011-01-01 00:16:36	31111	31503	W00970	Member
5	158	2011-01-01 00:32:33	2011-01-01 00:35:11	31605	31618	W01033	Member
6	560	2011-01-01 00:35:48	2011-01-01 00:45:09	31203	31201	W00766	Member

Figura 5

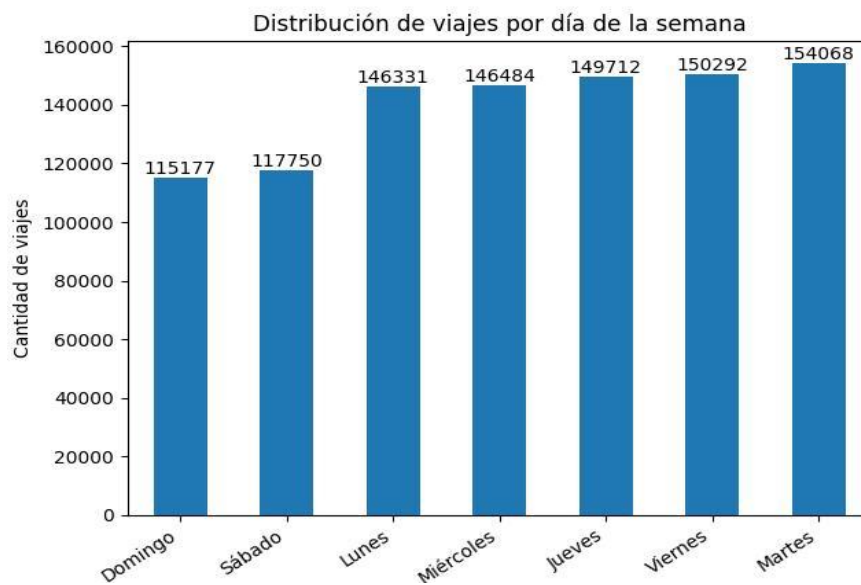


Figura 6

h) ID's de las 3 estaciones con mayor duración de viajes que comienzan y terminan en ellas mismas, a nivel histórico.

Los ID de las 3 estaciones en cuestión, de mayor a menor duración en horas (en viajes agrupados, históricamente) son: 31217, 31200, 31235 (Figura 7).

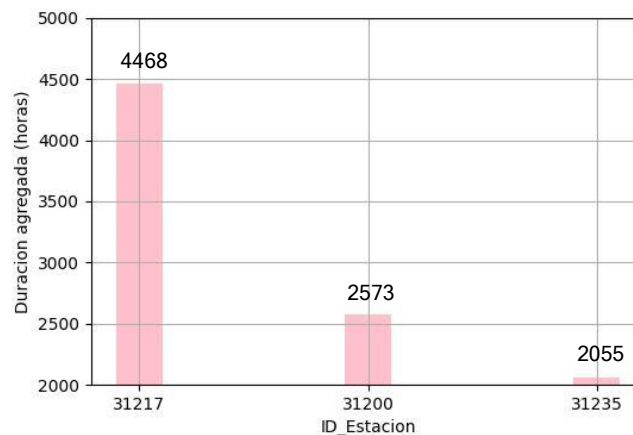


Figura 7

i) Representación gráfica comparativa de la cantidad de viajes agrupados por mes, para ambos tipos de usuario (Figura 8):

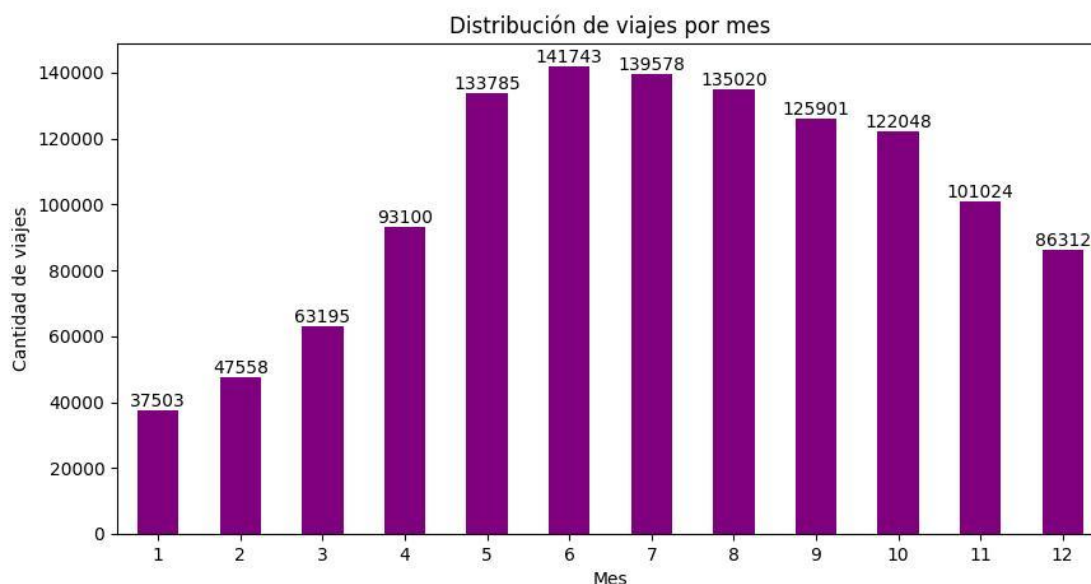


Figura 8

3. Conclusiones generales

En base al análisis de los resultados presentados en el apartado anterior, se han arribado a las siguientes conclusiones:

Usuarios Miembro vs Usuarios Casuales:

Es una buena señal que del total de viajes realizados de más de 5 minutos, casi un 80% fueron realizados por usuarios Miembro, lo cual significa que el público valora el servicio y los beneficios adicionales que reportan el estar registrado como Miembro. En este sentido, se deduce que la estrategia que aplicó la compañía fue satisfactoria.

Duración de viajes más largos:

Se detectaron varios casos en donde el tiempo transcurrido entre el retiro de una bicicleta hasta la devolución de la misma es prácticamente 24 horas. Dado que no es viable que una persona haga un viaje de esa duración, se deduce que dichas unidades permanecen en un domicilio hasta el día siguiente, cuando son retornadas a alguna estación. Se sugiere analizar más a fondo esta situación, por ejemplo mediante la aplicación de una encuesta de satisfacción del cliente en donde se incluyan preguntas relacionadas a esta cuestión, y así poder identificar oportunidades de mejora y/o necesidad de nuevas estaciones en puntos estratégicos. Los ID de las estaciones recomendadas para realizar las encuestas son los siguientes: 31232, 31400, 31104, 31200, 31102, 31221, 31705 y 31617.

Zonas más populares y estaciones más utilizadas:

Se determinó que las 2 estaciones más populares y de mayor demanda son las que se encuentran en el microcentro de la ciudad, cuyos ID son 31200 y 31201. En segundo lugar se ubican las estaciones 31235 y 31217, que reportan a nivel agregado la mayor cantidad de tiempo en viajes que comienzan y terminan en ellas mismas.

Habiendo identificado estas estaciones, sería una acción adecuada para mejorar el servicio ampliar dichas estaciones y/o ubicar nuevas en las cercanías, para evitar así posibles cuellos de botella y/o saturación de estaciones. A su vez es importante darle prioridad a estas estaciones en cuanto a la asignación de recursos para el mantenimiento de las mismas y de las bicicletas.

Frecuencia de uso según día de semana y mes del año:

Se detectó que si bien hay diferencia entre el volumen de viajes realizados en días hábiles y en fines de semana, la misma no es significativa. Se destaca que el volumen adicional en días hábiles puede corresponder a que los usuarios aprovechen el servicio para trasladarse desde y hacia sus respectivos trabajos.

En cuanto a la frecuencia de uso según los meses del año, lógicamente se detecta que en los meses fríos (enero, febrero y marzo) el uso de las estaciones desciende hasta menos de la mitad con respecto a los meses más cálidos (junio, julio, agosto, septiembre y octubre), y siendo meses de uso intermedio los meses de abril, noviembre y diciembre. Esta información es valiosa para gestionar de manera eficaz y eficiente el mantenimiento de las estaciones y bicicletas, incrementando el personal técnico que realiza estas labores para los meses cálidos, y disminuyéndolo para los meses fríos, realizando dichas transiciones de forma gradual durante los meses de uso intermedio. Esto beneficiará a la compañía, ya que no habrá personal ocioso durante los meses de bajo uso, y a su vez estará preparada para afrontar la suba de la demanda hacia mediados de año, evitando que el servicio se interrumpa por fallas en el sistema y/o falta de bicicletas en buenas condiciones, lo cual generaría demoras e inconvenientes para los usuarios frecuentes del servicio.

4. Recomendaciones

El equipo de análisis sugiere realizar un estudio enfocado en las estaciones menos demandadas (de menor cantidad de partidas y llegadas), así como de las estaciones cerradas, en donde se incluya información sobre fecha de cierre y motivo del mismo.

Este estudio podrá ayudar a analizar la viabilidad de continuar operando aquellas estaciones que no están siendo utilizadas suficientemente, y proveer fundamentos para decidir el cierre y/o reubicación de las mismas hacia zonas más demandadas y requeridas.

5. Bibliografía

- Documentación oficial de Python. *Python Software Foundation. Recuperado en 2023, de: <https://www.python.org/doc/>.*
- Think Python 2, Allen Downe.