

Ficha Técnica: Proyecto de Análisis de Datos

Título del Proyecto: City Bikes

Objetivos:

1. Calcular Métricas de Uso Diario:

- Determinar el número promedio de viajes realizados en un día típico.
- Calcular las estadísticas de duración de los viajes, incluyendo los valores máximos, mínimos, promedio y la desviación estándar.

2. Evaluar Métricas Históricas:

- Calcular el total de viajes realizados en el programa de bicicletas compartidas.
- Analizar el crecimiento del número de viajes diarios a lo largo del tiempo.
- Desglosar el total de viajes según el género/edad o tipo de suscripción de los usuarios.

3. Derivar Conclusiones y Recomendaciones:

- Identificar patrones y tendencias a partir de los datos analizados.
- Ofrecer recomendaciones estratégicas basadas en los hallazgos para la nueva CEO.

Equipo:

Individual

Herramientas y Tecnologías:

- Google BigQuery
- Google Slides
- Google Looker Studio

Lenguajes:

- SQL en BigQuery

Insumos:

citi_bike_trips.csv

Diccionario de datos:

tripduration: Duración del viaje en segundos desde el inicio hasta el final.

stoptime: Fecha y hora en que el viaje terminó.

start_station_id: Identificador único de la estación donde comenzó el viaje.

start_station_name: Nombre de la estación donde comenzó el viaje.

start_station_latitude: Latitud geográfica de la estación donde comenzó el viaje.

start_station_longitude: Longitud geográfica de la estación donde comenzó el viaje.

end_station_id: Identificador único de la estación donde terminó el viaje.

end_station_name: Nombre de la estación donde terminó el viaje.

end_station_latitude: Latitud geográfica de la estación donde terminó el viaje.

end_station_longitude: Longitud geográfica de la estación donde terminó el viaje.

bikeid: Identificador único de la bicicleta utilizada en el viaje.

usertype: Tipo de usuario: "Customer" (cliente) o "Subscriber" (suscriptor).

birth_year: Año de nacimiento del usuario.

gender: Género del usuario: female, male, unknown(desconocido/no especificado).

customer_plan: Plan de suscripción o tipo de membresía del cliente.

Procesamiento y análisis

Conectar/importar datos a herramientas

- Se creó el City-Bikes y el conjunto de datos Dataset en BigQuery.
- Tablas importadas: city_bike_trips.

Identificar y manejar valores nulos

- Se identifican valores nulos a través del comando SQL COUNTIF.
- Se eliminaron los valores nulos para la edad .
- **birth_year:** **4796 nulos** que representan el **9.59%** del total de los registros, quedando **45204** registros en total.

Identificar y manejar valores duplicados

- No se encontraron registros duplicados de viajes.

Identificar y manejar datos fuera del alcance del análisis

- **customer_plan:** se excluyó la columna ya que no tenía información en ninguno de los registros, solo 50000 valores nulos.
- Se calculó la edad de los usuarios con `EXTRACT(YEAR FROM CURRENT_DATE()) - birth_year` y se observó que habían **157** usuarios > 80 años , por la edad es posible que las fechas de nacimiento ingresadas no sean las correctas por lo cual se dejaron estos registros fuera del análisis.

Crear nuevas variables

- **trip_duration_minutos:** Se creó la variable duración del viaje en minutos.
- **stop_date:** se cambió el tipo de datos de la variable **stoptime** a **date** para tener solo la fecha del viaje.

Unir tablas

- Se creo una vista **view_city_bike_trips** , con las nuevas variables y excluyendo la que no era relevante para nuestro análisis.

Análisis Exploratorio

Agrupar datos

- Se importan los datos de BigQuery a Looker Studio.
- Se crearon score cards:
Total de viajes : **45.204**
Total de bicicletas: **14599**
Mínimo de viajes: **1 minuto.**
Máximo de viajes: **54.566 minutos.**
Promedio de viajes: **14,75 minutos.**
Desviación Estándar: **264,86 minutos.**

Visualizar las variables

- Se crearon gráficas de barras , circular y mapas para la visualización de variables en Looker Studio.

Conclusiones y Recomendaciones:

Patrones Identificados:

Crecimiento Rápido de Usuarios:

De 2015 a 2016, el número de usuarios creció exponencialmente, aumentando en más del 200%. Este crecimiento continuó en 2017, con un aumento del 147% en comparación con el año anterior. Sin embargo, en 2018, el número de usuarios disminuyó en un 43% en comparación con 2017.

Incremento en el Número de Bicicletas:

El número de bicicletas se mantuvo constante entre 2015 y 2016, a pesar del significativo aumento en usuarios. En 2017, la empresa aumentó drásticamente la cantidad de bicicletas en respuesta al crecimiento de usuarios, pero en 2018, aunque el número de bicicletas disminuyó, seguía siendo más alto que en 2015 y 2016.

Desajuste entre la Disponibilidad de Bicicletas y el Crecimiento de Usuarios:

En 2016, aunque el número de usuarios se triplicó, la cantidad de bicicletas no aumentó, lo que podría haber afectado la disponibilidad del servicio y la satisfacción del cliente. En 2017, la empresa parece haber corregido este desajuste aumentando significativamente la flota, pero en 2018, a pesar de la

disminución en usuarios, el número de bicicletas solo se redujo ligeramente, lo que podría haber resultado en un exceso de oferta en relación con la demanda.

Incremento en la Duración de los Viajes:

El aumento en la duración promedio de los viajes en 2018 podría indicar un cambio en la demanda o en el comportamiento de los usuarios. Es posible que los usuarios estén usando las bicicletas para viajes más largos.

Dominio de los Abonados:

La preferencia por el modelo de suscripción sugiere que el servicio de bicicletas compartidas es más popular entre los usuarios frecuentes. Esto puede ser un indicativo de una lealtad alta entre los abonados.

Uso Similar por Género:

El análisis revela una participación relativamente equilibrada entre hombres y mujeres en el uso de bicicletas compartidas. Los hombres representan el 37.1% de los usuarios, mientras que las mujeres constituyen el 34.5%. Un 28.4% de los usuarios no especificaron su género.

Recomendaciones Estratégicas:

Ofrece sugerencias al nuevo CEO basadas en los hallazgos.

- Se recomienda realizar un análisis detallado para ajustar la cantidad de bicicletas en función del crecimiento proyectado de usuarios. Si se espera que la cantidad de usuarios siga fluctuando, la empresa debería considerar implementar un sistema flexible para aumentar o reducir la flota según sea necesario.
- La disminución en el número de usuarios en 2018 sugiere que podría haber problemas en la retención de clientes. Se recomienda investigar las razones detrás de esta disminución, como la posible saturación del mercado, problemas de servicio, o cambios en la competencia, y tomar medidas para mejorar la retención y la experiencia del usuario.
- Aprovechar los datos históricos para segmentar a los usuarios y personalizar las ofertas, servicios y campañas de marketing según el comportamiento y las necesidades específicas de diferentes grupos de usuarios. Esto puede ayudar a mejorar la satisfacción y la lealtad, y

atraer a nuevos usuarios.

- Implementar un sistema de monitoreo continuo para mantener un equilibrio adecuado entre la oferta (número de bicicletas) y la demanda (número de usuarios). Esto permitirá reaccionar rápidamente a cambios en el mercado y evitar problemas de sobreoferta o escasez de bicicletas.
- Investigar las razones detrás del aumento en la duración de los viajes en 2018. Esto podría implicar una reevaluación de las rutas, la demanda o las necesidades de los usuarios. Se podrían realizar encuestas para entender mejor el comportamiento de los usuarios.
- Con un crecimiento en el número de usuarios y bicicletas, es crucial asegurar que la distribución de las bicicletas sea óptima. Implementar un sistema de gestión de bicicletas que asegure que las estaciones con alta demanda siempre tengan suficiente disponibilidad.
- Para capitalizar el balance entre los géneros y fomentar aún más la igualdad en el uso del servicio, se recomienda desarrollar estrategias que apelen tanto a hombres como a mujeres. Esto podría incluir la promoción de características del servicio que sean valoradas por ambos géneros. Además, es crucial reducir el porcentaje de usuarios con género no especificado, lo cual se puede lograr mediante mejoras en el proceso de registro, asegurando así una mayor precisión en la recopilación de datos demográficos. Esto permitirá personalizar mejor las campañas y mejorar la experiencia del usuario.
- Evaluar la posibilidad de ofrecer diferentes niveles de suscripción o promociones para atraer a más clientes y mantener el crecimiento en el número de abonados.
- Continuar monitoreando el crecimiento del número de viajes diarios y ajustar las estrategias de negocio y operativas en consecuencia. Utilizar análisis de tendencias para prever y prepararse para futuras demandas.
- Inversión en Infraestructura: Considerar inversiones en infraestructura para mejorar la experiencia del usuario y la eficiencia operativa, como estaciones más accesibles y una mejor distribución de bicicletas.