**Análisis de Datos**

**Analista de datos:** Luis Fernando Martínez

**Fecha de creación:** 2024/12/08

**Fecha de actualización:** 2024/12/08

**Introducción**

En este documento se evidenciará el proceso de análisis de datos del proyecto. Se realizará un análisis descriptivo sobre cada archivo csv procesado correspondiente a los meses desde junio de 2023 a mayo de 2024, posteriormente se ejecutará un análisis descriptivo utilizando los datos consolidados en la tabla “consolidado\_junio2023\_mayo2024\_raw” del dataset “capstone\_project\_data” alojado en BigQuery, y para finalizar se llevará a cabo un análisis descriptivo utilizando R.

**Preguntas de negocio:**

* ¿Existe una relación entre el tipo de miembro y el tiempo de viaje?
* ¿Existe una relación entre el tipo de miembro y el día de la semana en que se utiliza el servicio?
* ¿Existe una relación entre el tipo de miembro y la cantidad de viajes?
* ¿Existe una relación entre el tipo de miembro y el tipo de bicicleta que se usa para el servicio?
* ¿Existe una relación entre el tipo de miembro y las distancias recorridas en los viajes?
* ¿Existe una relación entre el tipo de miembro y las estaciones utilizadas?
* ¿Existe alguna relación entre el tipo de miembro y las rutas utilizadas?

**Preguntas para el análisis descriptivo:**

* ¿Qué tipo de usuario utiliza con más frecuencia el servicio por mes?
* ¿Qué tipo de usuario utiliza con más frecuencia el servicio en todo el dataset?
* ¿Cuál es el tiempo de viaje promedio por mes de todos los usuarios?
* ¿Cuál es el tiempo de viaje promedio por mes de los usuarios miembros y casuales?
* ¿Cuál es el tiempo de viaje promedio en todo el dataset?
* ¿Cuál es el tiempo de viaje promedio en todo el dataset de los usuarios miembros y casuales?
* ¿Qué duración tuvo el viaje más extenso en tiempo de cada mes?
* ¿Qué duración tuvo el viaje más extenso en tiempo de cada mes de usuarios miembros y casuales?
* ¿Qué duración tuvo el viaje menos extenso en tiempo de cada mes?
* ¿Qué duración tuvo el viaje menos extenso en tiempo de cada mes de usuarios miembros y casuales?
* ¿Qué duración tuvo el viaje más extenso en tiempo de todo el dataset?
* ¿Qué duración tuvo el viaje más extenso en tiempo de todo el dataset de usuarios miembros y casuales?
* ¿Qué duración tuvo el viaje menos extenso en tiempo de todo el dataset?
* ¿Qué duración tuvo el viaje menos extenso en tiempo de todo el dataset de usuarios miembros y casuales?
* ¿Cuál es el día de la semana en que más se utiliza el servicio?
* ¿Cuál es el día de la semana en qué más utilizan el servicio los usuarios miembros y casuales?
* ¿Cuál es el día de la semana en que menos se utiliza el servicio?
* ¿Cuál es el día de la semana en que menos utilizan el servicio los usuarios miembros y casuales?
* ¿Cuál es el tiempo de viaje promedio por día?
* ¿Cuál es el tiempo de viaje promedio por día de usuarios miembros y casuales?
* ¿Cuántos viajes se realizan por día?
* ¿Cuántos viajes realizan por día los usuarios miembros y los casuales?
* ¿Qué tipo de bicicleta es la que más se usa?
* ¿Qué tipo de bicicleta es la que más usan los usuarios miembros y casuales?
* ¿Qué tipo de bicicleta es la que menos se usa?
* ¿Qué tipo de bicicleta es la que menos usan los usuarios miembros y casuales?
* ¿Cuál es la distancia más larga recorrida en un solo viaje en cada mes?
* ¿Cuál es la distancia más larga recorrida en un solo viaje en cada mes de los usuarios miembros y casuales?
* ¿Cuál es la distancia más corta recorrida en un solo viaje en cada mes?
* ¿Cuál es la distancia más corta recorrida en un solo viaje en cada mes de los usuarios miembros y casuales?
* ¿Cuál es la distancia más larga recorrida en un solo viaje en todo el dataset?
* ¿Cuál es la distancia más larga recorrida en un solo viaje en todo el dataset de los usuarios miembros y casuales?
* ¿Cuál es la distancia más corta recorrida en un solo viaje en todo el dataset?
* ¿Cuál es la distancia más corta recorrida en un solo viaje en todo el dataset de los usuarios miembros y casuales?
* ¿Cuál es la distancia promedio recorrida en cada mes?
* ¿Cuál es la distancia promedio recorrida en cada mes por los usuarios miembros y casuales?
* ¿Cuál es la distancia promedio recorrida en todo el dataset?
* ¿Cuál es la distancia promedio recorrida en todo el dataset por usuarios miembros y casuales?
* ¿Cuál es la estación más utilizada por mes?
* ¿Cuál es la estación más utilizada por mes por los usuarios miembros y casuales?
* ¿Cuál es la estación más utilizada en todo el dataset?
* ¿Cuál es la estación más utilizada en todo el dataset por usuarios miembros y casuales?
* ¿Cuál es la ruta más utilizada por mes?
* ¿Cuál es la ruta más utilizada por mes por los usuarios miembros y casuales?
* ¿Cuál es la ruta más utilizada en todo el dataset?
* ¿Cuál es la ruta más utilizada en todo el dataset por usuarios miembros y casuales?
* ¿Cuál es la ruta menos utilizada por mes?
* ¿Cuál es la ruta menos utilizada por mes por los usuarios miembros y casuales?
* ¿Cuál es la ruta menos utilizada en todo el dataset?
* ¿Cuál es la ruta menos utilizada en todo el dataset por usuarios miembros y casuales?

**Análisis descriptivo**

Para realizar el análisis descriptivo se empieza por analizar cada mes individuamente, es decir, aplicar fórmulas y tablas dinámicas a cada archivo según sea necesario.

Junio de 2023

Este caso es especial porque será el único archivo para el cual se aplicarán todas las fórmulas de Excel manualmente, esto con el objeto de practicarlas y enfrentar diferentes situaciones donde es necesario anidar fórmulas. Los demás archivos tendrán menos fórmulas aplicadas a mano y se optará por tablas dinámicas para acelerar el trabajo.