

A thick dark grey vertical bar runs down the left side of the page. To its right, several thin, curved lines in black and grey sweep upwards and outwards from the bottom left corner.

26-4-2024

Yelp & Google maps- Reviews and Recommendations

Sprint 3

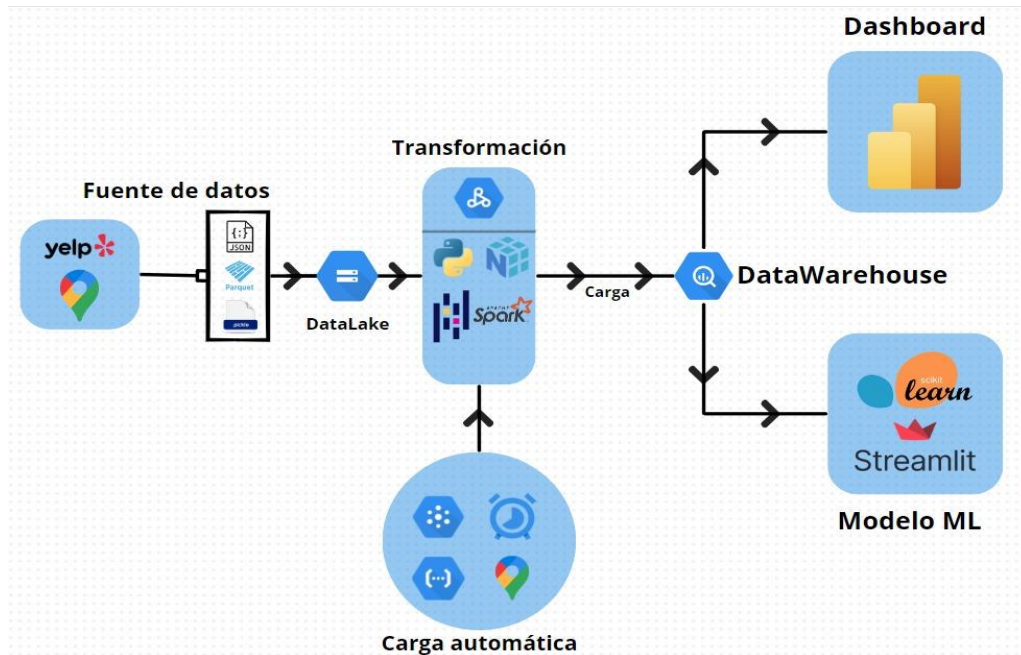
Kensit Cortes
Juan Carlos Sánchez
Facundo Blanco
José Efraín Pazos
Rafael Miranda

DATA FIVEBUSINESS CONSULTANS

Contenido

Pipeline del proyecto	2
Dashboard del proyecto.....	2
Modelo entidad relación.....	3
Modelo de machine learning	4

Pipeline del proyecto



El diagrama muestra el flujo final de trabajo que se obtuvo en Google Cloud, como herramienta para darle manejo a la data, empezando por la construcción del datalake en Google Cloud Storage, pasando a Dataproc y utilizando Google Cloud Functions para realizar el etl y el pipeline automatizado, posterior a esto se almacena la data limpia en Big query, que sería nuestro DataWarehouse y posterior a esto se hace la conexión automática con Power BI para la elaboración del tablero. De Big query, también obtenemos la información para el desarrollo de los modelos de Machine Learning, con la librería Scikit-Learn para realizar el código y con Streamlit para hacer el despliegue de los mismos

Dashboard del proyecto

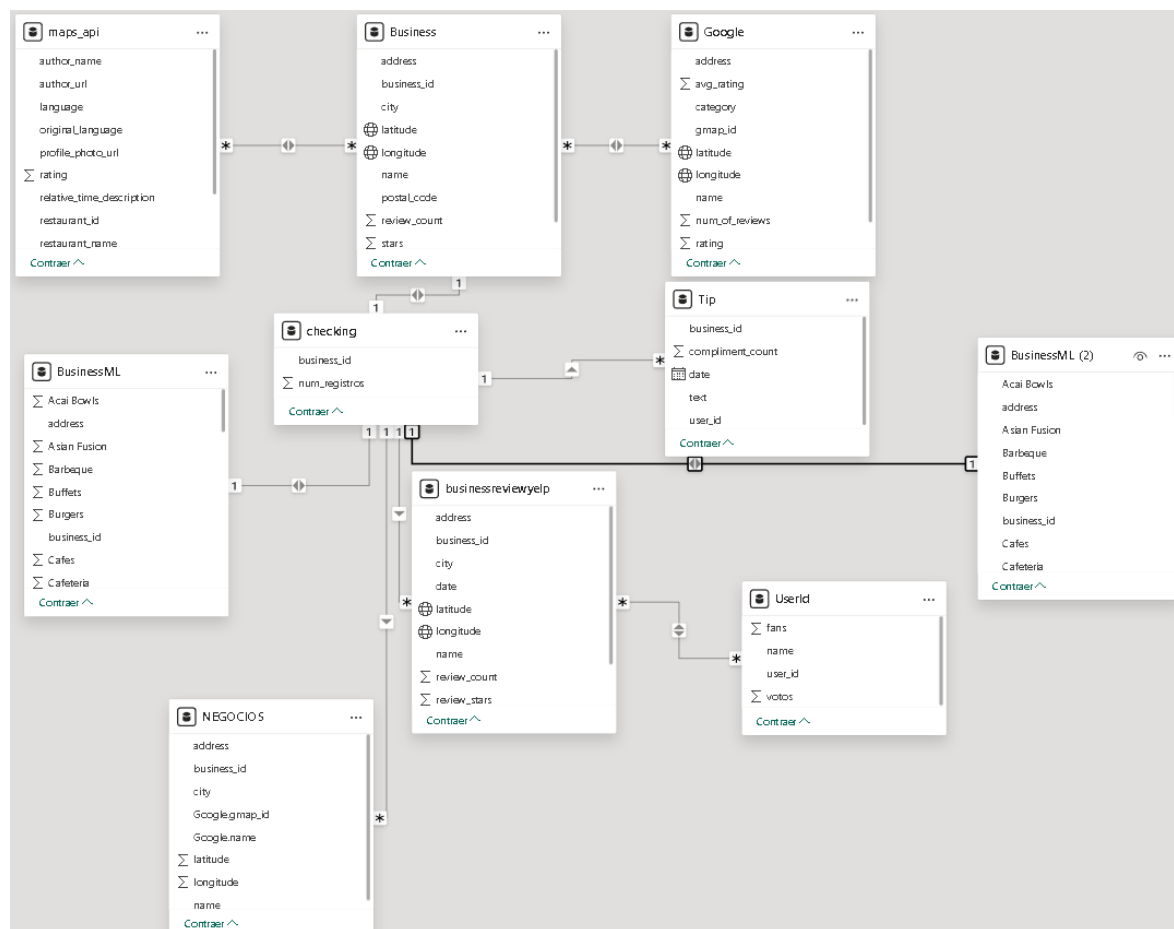
En el dashboard podemos encontrar el análisis de la página de Yelp, Google y los KPI's.

La primera página es el análisis de Yelp, donde podemos encontrar la calificación promedio de los negocios, sus visitas totales y el filtro por rango de calificaciones y por ciudades. De esta manera se puede hacer el análisis del promedio de las calificaciones por negocio y por ciudad. la cantidad total de sus calificaciones y el promedio de estas.

La segunda página es un análisis de la plataforma de Google, donde se puede hacer el análisis de los negocios a través de las calificaciones que han dado a través de Google Maps. Se puede observar el promedio de calificaciones, las reseñas totales y una relación entre la cantidad total de reseñas y el promedio de estas.

Por último, encontramos la visualización de los análisis de los KPI's escogidos

Modelo entidad relación



El modelo entidad-relación abarca una variedad de relaciones clave, que incluyen desde uno a muchos, muchos a muchos y uno a uno. Este enfoque proporciona una estructura sólida para representar las interconexiones entre los diferentes elementos del conjunto de datos.

En nuestro caso, hemos implementado un esquema de estrella que se centra en una tabla principal de hechos llamada "Business". Esta tabla es el núcleo de nuestro modelo y está conectada a casi todos los demás componentes a través de su clave primaria, **business_id**. Además, hemos establecido relaciones de muchos a muchos con las tablas de "Maps Api" y "Google", utilizando el campo "name" como punto de conexión.

Este diseño nos permite capturar de manera efectiva las relaciones complejas entre los diversos elementos de datos, lo que facilita el análisis y la comprensión de la información en su totalidad. Además, al implementar un esquema de estrella, estamos optimizando la estructura de nuestra base de datos para consultas eficientes y resultados precisos.

Modelo de machine learning

El primer modelo de machine learning, propone un modelo que puede predecir la factibilidad de colocar un nuevo negocio y saber si será aceptado por los clientes, teniendo en cuenta la ciudad y 3 tipos de comida que el restaurante tenga como optativa para vender y tenga buena aceptación por el cliente.

El modelo de machine learning se hizo bajo el modelo de árboles de decisión, que permite saber si la ciudad y las 3 categorías de restaurantes serán aceptados por los clientes, tomando la cantidad de estrellas, como indicador de una buena o mala decisión para invertir.

El segundo modelo es un sistema de recomendación de restaurantes para los usuarios, para darle a conocer la posibilidad de conocer nuevos sabores basados en experiencias previas, con la similitud de coseno para calcular que tan similares son los usuarios entre si en función de sus preferencias de restaurantes