



PROYECTO FINAL 2023

GOOGLE *YELP*

Datalogia

Data consulting-Equipo

- Dante Alejandro Gomez – Data Analytics
- Hector Fernando Calisaya- Data Engineer
- Nicolas Fernando Tablon- Data Engineer
- Daniela Florencia Gutierrez- Data Science

Producto en desarrollo

INFORMACION UTILIZADA:

Para el desarrollo del producto se empleo las siguientes carpetas y archivos:

- metadata_sitios
- review-estados /review-New_York
- business.pkl-Yelp
- review.json -Yelp
- tip.json -Yelp

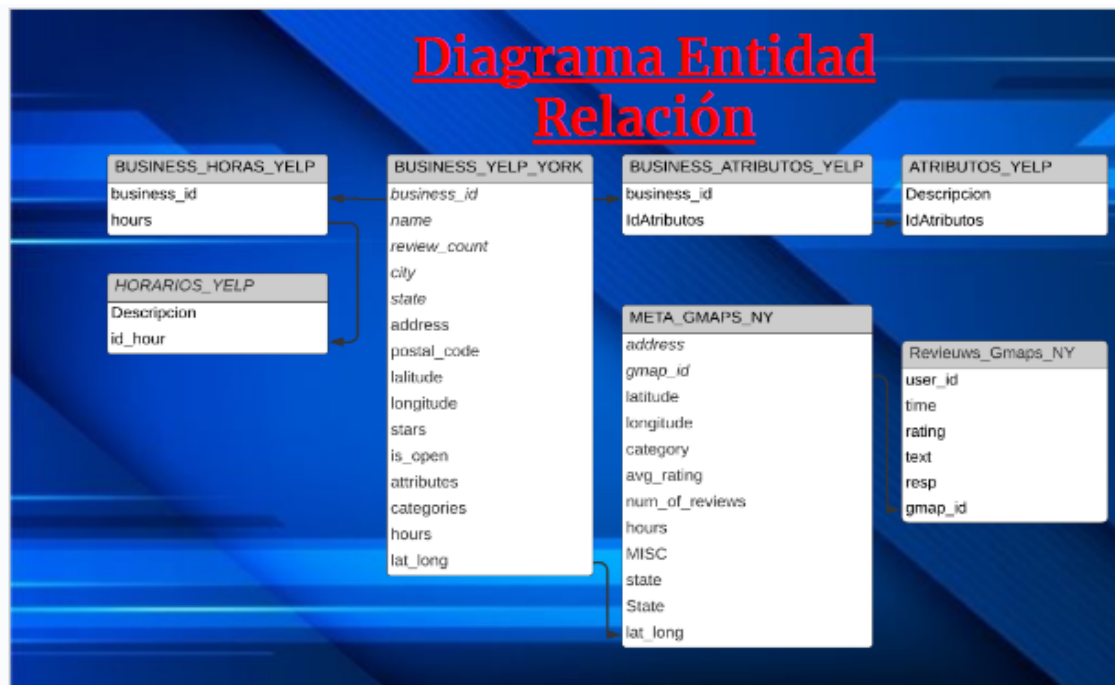
MODELO DE DATA ESCOGIDO : DATA WAREHOUSE

Es una solución centralizada para almacenar, gestionar y analizar grandes volúmenes de datos. En este caso ,se utilizo Google Cloud Storage que se suele usar para almacenar datos no estructurados: objetos o archivos de cualquier tipo(parquet, json o archivos csv) y de grandes volumenes en la nube de Google ya que podemos garantizar la calidad y seguridad de los datos junto al control de acceso.

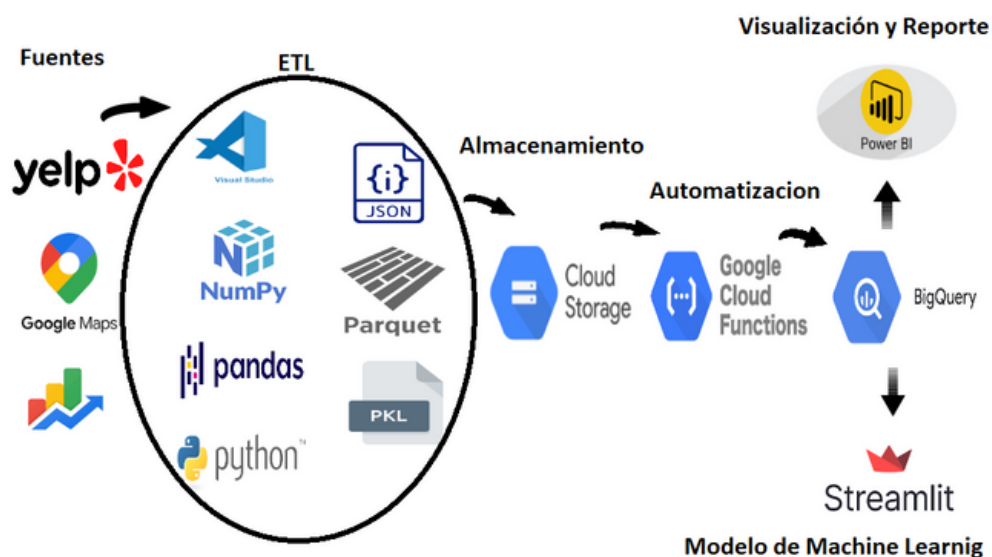
DER (DIAGRAMA ENTIDAD - RELACION)

El modelo DER respresenta la estructura de los datos de Yelp y Google Maps .Proporciona una vision de la información que seran de gran utilidad para la toma de decisiones

Diagrama Entidad Relación



PIPELINE DEL PROYECTO



EXTRACCION Y TRANSFORMACION DE DATOS(ETL)

Desde la fuente los datos fueron tratados, con las librerías de Numpy y Panda que ofrece el lenguaje de Python lo cual permitio hacer las transformaciones necesarias para disponibilizar los datos a los siguientes procesos.

ALMACENAMIENTO DE DATOS

GCP(Google Cloud Platform). Se creo la base de datos y se guardo la informacion en cloud storage y Big Query

AUTOMATIZACION DE DATOS

Se creo una función en automatizada la cual ingesta información de una API de finanzas y se guarda tanto en Cloud Storage como en Big Query. Esta función es llamada automáticamente cada 2 semanas.

MODELO DE MACHINE LEARNING(ADICIONAL)

NLTK (Natural Language Toolkit)y Random Forest se utilizaran como modelos de machine learning usando los datos de almacenados en Big Query.

Finalmente se utilizara Streamlit para representar el modelo en una API

VISUALIZACION Y REPORTE

Se usara Power BI como herramienta de visualizacion y reporte para crear paneles , filtros y graficos intectivos el cual estara disponible en una pagina web.

DIAGRAMA DE GANTT:

https://docs.google.com/spreadsheets/d/14UPUmRgiQqYQg5buaPHvRONM_tKcteJyPIlN3g6RlSg/edit#gid=0