



Datalogia

- 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

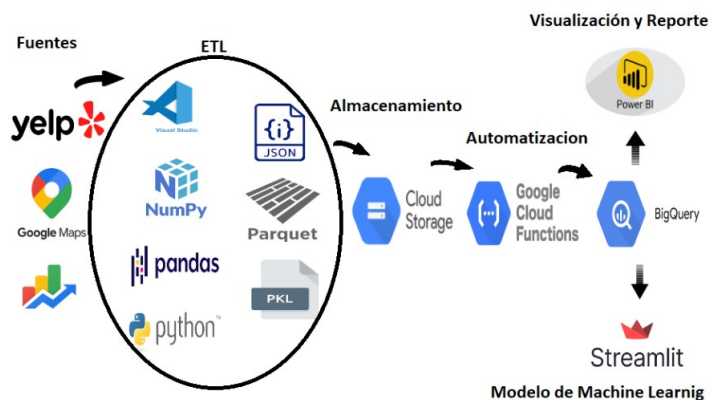
INFORMACION UTILIZADA:

Para el desarrollo del producto se empleó 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.



EXTRACCION Y TRANSFORMACION DE DATOS

Desde las fuentes los datos fueron tratados, mediante lenguajes y sistemas de procesamiento para poder usarse, cambiando su estructura, tipo y agrupamiento entendiendo que los datos son relevantes para su interpretación posterior en las siguientes etapas

ALMACENAMIENTO DE DATOS

Los datos resultantes fueron cargados en sistemas de almacenamiento de Google Cloud Storage, y para la creación de tablas Big query que permiten la toma de datos a intervalos programados entre otras tareas hacen posible sostener las fuentes procesadas necesarias

AUTOMATIZACION DE DATOS

Mediante Google Cloud Function fue posible automatizar la carga de datos externos, como también su actualización, permitiendo un continuo proceso mediante el cual los datos son actuales, por ejemplo entender el valor de la franquicia McDonald's en Google Finance interpretando su valor de mercado y relacionándolo a las opiniones de clientes.

MODELO DE MACHINE LEARNING

NLTK (Natural Language Toolkit) y Random Forest se utilizaron como modelos de machine learning usando los datos de almacenados en Google Cloud Storage.

DIAGRAMA ENTIDAD RELACION

