

IT BROTHERS – Consultora especialista en ciencia de datos

INTEGRANTES DEL STAFF:

JIMÉNEZ, JOSE
ALONSO, LUCILA
ARGUMEDO, HÉCTOR
TALAVERA, RICARDO
UMBERT, NORBERTO

ÍNDICE

Introducción.....	3
Objetivo.....	3
Tareas realizadas.....	3
KPIs utilizados	4
Workflow.....	5
Cronograma.....	6

INTRODUCCIÓN

A través de este archivo, explicaremos lo desarrollado durante de esta semana, donde empezamos a trabajar desde los buckets creados en la nube para después trabajar con una herramienta de visualización. Desarrollamos también modelos de Machine Learning y reportes finales de información para la empresa Olist.

OBJETIVO

El objetivo de esta semana fue usar los datos, después del proceso de carga a la nube, en una herramienta de visualización para poder exponer mediante métricas y kpis las variables que nos permitan presentar a la empresa un reporte sobre cómo es la situación actual de la empresa según los datos dados.

TAREAS REALIZADAS

- Creación de API

Creamos una API de consulta, implementando FastAPI para hacer consultas a los datos almacenados en el Data Lake. Esta API esta dockerizada y desplegada con Google Cloud Run, de esta manera tenemos acceso web para poder ofrecer la posibilidad de consultar los datos y ,de ser necesario, descargar la información en formato JSON.

- BACK-UP

Como último proceso de ingeniería de datos, se realizó un bucket dentro de la nube, la cual la usamos para tener un backup de los datos obtenido a partir del Data Lake y los datos de Stage. Este backup, se programó con la herramienta Google Cloud Data Transfer, para poder ser actualizarlo en un día y hora ya preestablecido.

- MACHINE LEARNING

Se decidió hacer un análisis acerca del posicionamiento de sentimiento hacia la marca "Olist". Para lo cual se observó, que los datos proporcionados de la empresa acerca de los comentarios de clientes no presentaban una escala considerable para un modelamiento.

De acuerdo a lo anterior se obtuvieron datos adicionales provenientes de la red social “Twitter” y a través de “Google”.

Como primera etapa, se calcula el sentimiento hacia la marca, el cual nos da luces de la polaridad de los comentarios y una consolidación del posicionamiento de sentimiento hacia “Olist”.

Luego se procede a efectuar con algoritmos de clasificación un modelo ejecutable que se despliega en streamlit para su consumo.

- **VISUALIZACIÓN**

A partir de disponer de los datos ya normalizados, nos pusimos a trabajar con la parte de visualización con la herramienta POWER BI. Aquí importamos los datos desde BigQuery, en GCP , y siguiendo nuestro propósito del trabajo, nos propusimos generar visualizaciones. Siguiendo métricas y KPI's preestablecidos, nuestros dashboard pretenden dar una mejor comprensión de los datos y obtener información relevante. De esta manera, toda futura decisión puede ser respaldada por datos e información.

KPIS UTILIZADOS

- **KPIs - Logística**

El trabajo de las variables en este sector está enfocado a observar el desarrollo del plazo de entrega de los productos al cliente a lo largo del tiempo, analizando las cantidades de envíos realizados con éxito.

- **KPIs - Ventas**

El trabajo de las variables en este sector está enfocado en los ingresos obtenidos a partir de las ventas en la plataforma, ubicándolas a estas por distintas variables por ejemplo: Región, año y trimestre.

- **KPIs – Medios de pago**

El trabajo de las variables en este sector está enfocado en poder discriminar las formas de pagos de los clientes, mostrando así que medio de pago es la más utilizada y su variación de uso a lo largo del tiempo.

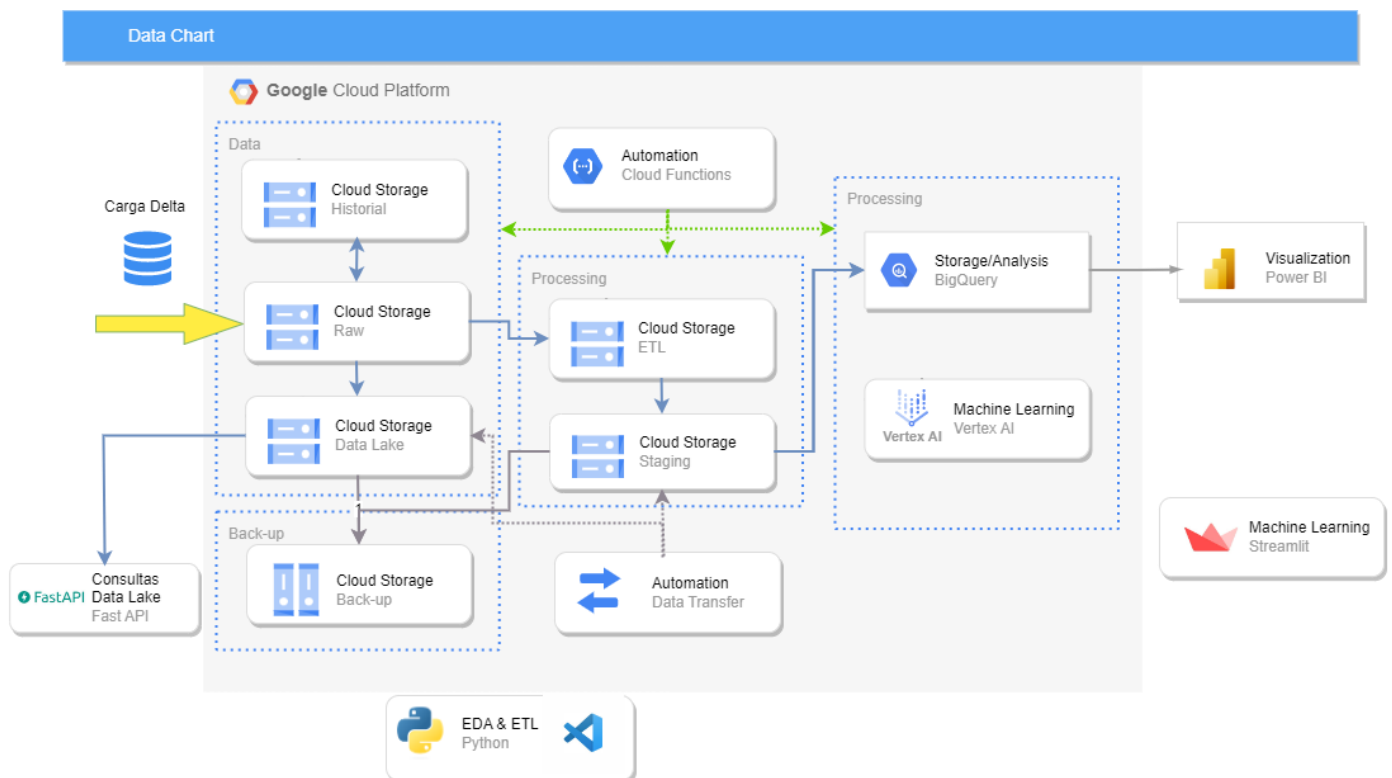
- **KPIs – Comentarios**

El trabajo de las variables en este sector está enfocado en los comentarios realizados por los usuarios de la plataforma, mismos que permiten tener una idea más cercana al nivel de satisfacción del cliente.

- **KPIs – Ordenes y Productos**

Permiten tener una evaluación anual y trimestral de cada región que realizó un pedido, ayudando a dilucidar la distribución y la evolución de los mismos.

WORKFLOW



CRONOGRAMA DE LA SEMANA

Tercer Semana Data Analytics + ML	Lunes 23 Enero	Martes 24 Enero	Miercoles 25 Enero	Jueves 26 Enero	Viernes 27 Enero	Horas trabajadas	Progreso
Diseño de Reporte/Dashboards	Lucila Héctor Ricardo	Lucila Héctor Ricardo				6	100%
Dashboard (conexión a BigQuery)		José Norberto				2	100%
KPI's		Ricardo Lucila Héctor	José Norberto			6	100%
Modelo de ML			Ricardo Héctor			8	100%
Modelo de ML en producción			Ricardo Héctor	Ricardo Héctor		8	100%
Documentación: Selección del modelo, feature engineering			Ricardo Héctor	Lucila Norberto José		6	100%
Informe de Análisis					Lucila Norberto José	2	100%