



DataMinds Solutions

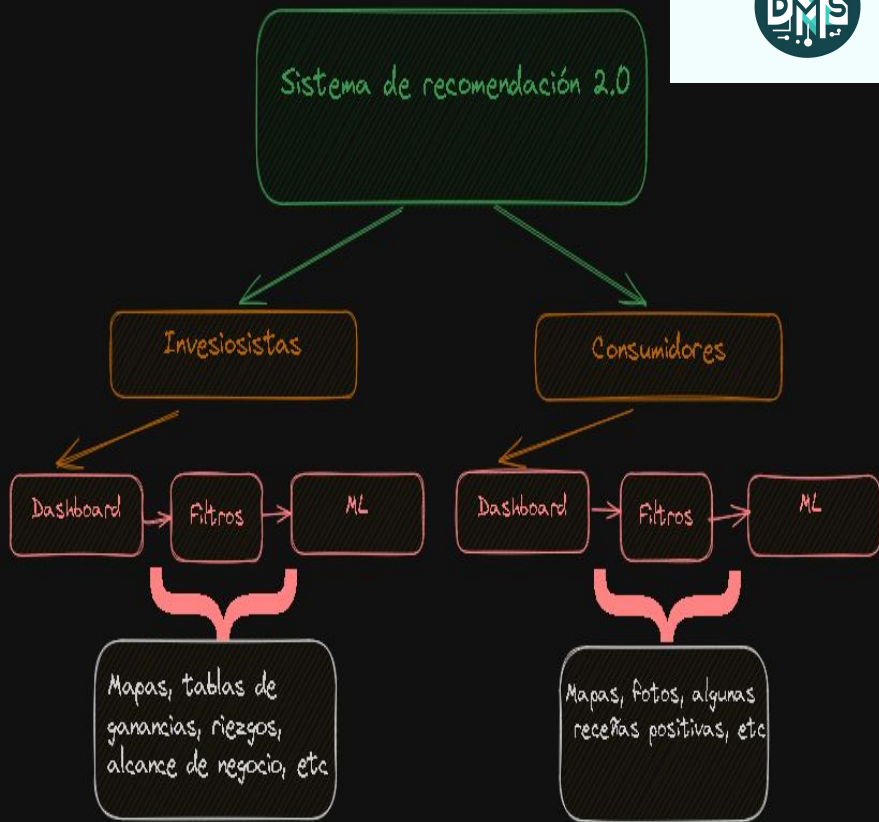
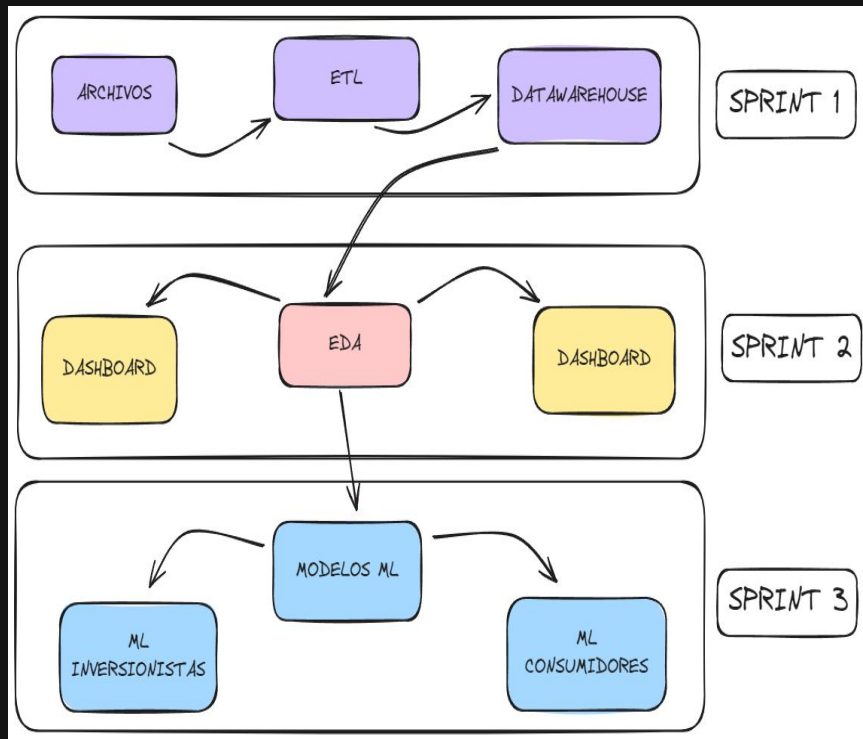
**Desbloqueando el Potencial Empresarial a través de Ingeniosas Ideas
e Innovación en Big Data**



Google Maps



Nuestra consultora analizará datos de Yelp y Google Maps para entender la satisfacción del cliente, prever tendencias de negocios y sugerir ubicaciones estratégicas para nuevos locales. Luego, desarrollaremos un sistema de recomendación personalizado basado en el historial del usuario. Los resultados se presentarán en un informe con recomendaciones clave.



Metodología Estándar

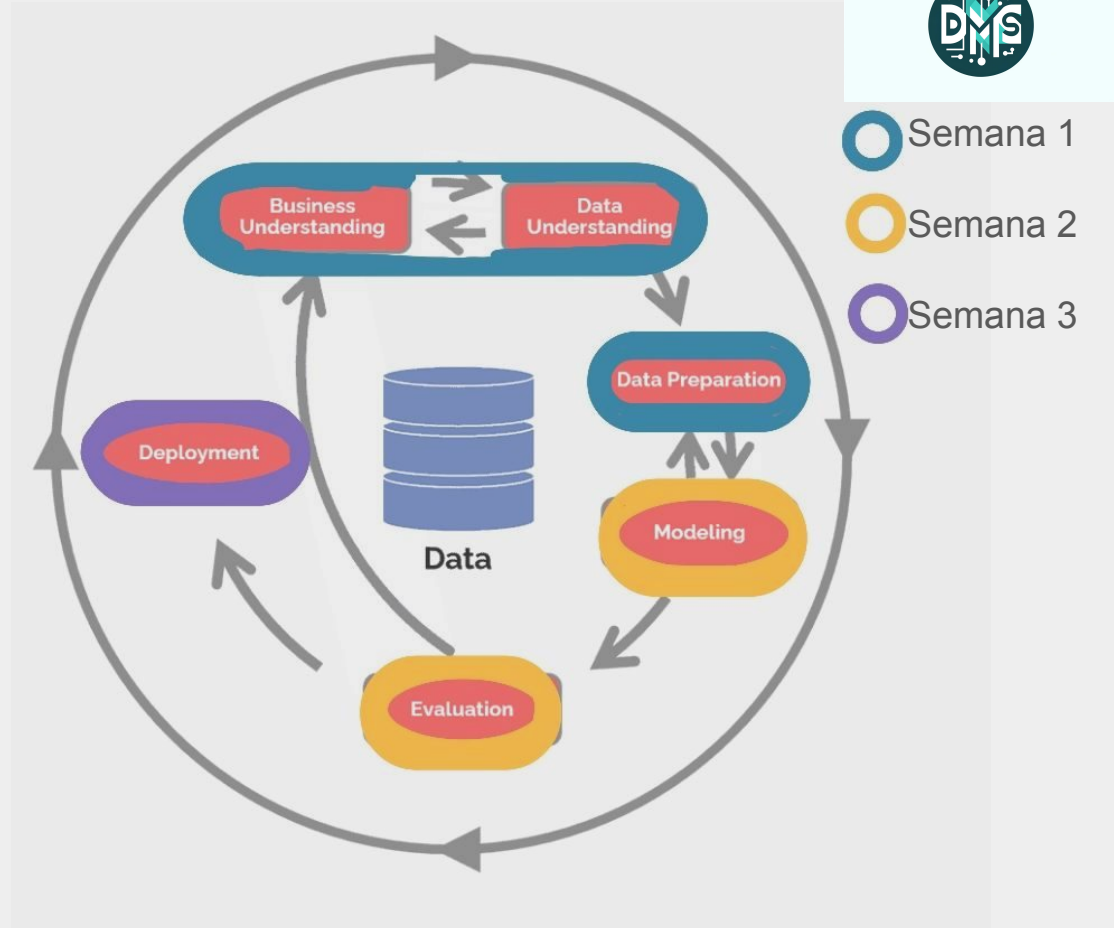


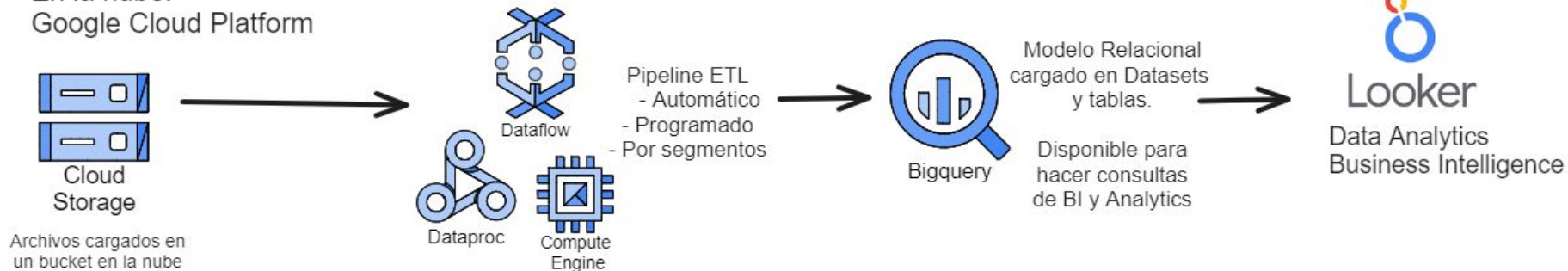


Diagrama Procedural

Local:



En la nube:
Google Cloud Platform





Data Lake

A la hora de cargar los datos en la nube optamos por la ingesta de los datos en google cloud storage, debido a su facilidad de uso

data_lake_pf_henry

Ubicación	Clase de almacenamiento	Acceso público	Protección
northamerica-northeast1 (Montreal)	Standard	No público	Ninguna

[OBJETOS](#)[CONFIGURACIÓN](#)[PERMISOS](#)[PROTECCIÓN](#)[CICLO DE VIDA](#)

Depósitos > data_lake_pf_henry

[SUBIR ARCHIVOS](#)[SUBIR CARPETA](#)[CREAR CARPETA](#)[TRANSFERIR LOS DATOS](#)

[EDITAR LA RETENCIÓN](#)[DESCARGAR](#)[BORRAR](#)

Filtrar solo por prefijo de nombre **Filtro** Filtrar objetos y carpetas

<input type="checkbox"/>	Nombre	Tamaño	Tipo	Fecha de creación
<input type="checkbox"/>	Google Maps/	—	Carpeta	—
<input type="checkbox"/>	Yelp/	—	Carpeta	—
<input type="checkbox"/>	google-cloud-dataproc-metainfo/	—	Carpeta	—



Procesamiento de Datos



**Cloud
Dataproc**

Decidimos usar Google Cloud Dataproc y Spark para el procesamiento de datos, ya que son herramientas que funcionan muy bien cuando se trata de gran cantidad de información





Data Warehouse




BigQuery


Utilizaremos BigQuery ya que proporciona una solución de almacenamiento de datos sin servidor y altamente escalable. Además nos va a permitir analizar grandes conjuntos de datos utilizando SQL con facilidad y velocidad









Repositorio de Github

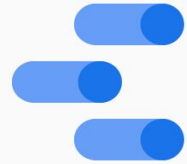
 **PF_Grupo06_Henry** Public Edit Pins Watch 0

main 2 Branches 0 Tags Add file Code

 **nicoMusante**

 Merge pull request [#14](#) from Zzepu/main ff1a44a · 2 hours ago 48 Commits

 diagramas	carpetas de ETL y EDA	yesterday
 documentación	ordenamos todo en carpetas	3 days ago
 notebooks	EDA google en profundidad finalizado	3 hours ago
 .gitignore	modificacion rutas, creacion archivo consumo final	yesterday
 README.md	Fixed structure	4 hours ago
 requirements.txt	Requirements file	4 days ago



Data Studio

Data Studio sera la herramienta a utilizar para crear dashboards interactivos.

KPI's



KPI's a utilizar

-Índice de Sentimiento Promedio por Año (ISP):

Definición: Este KPI mide el tono general de las reseñas recopiladas, proporcionando una indicación de si la percepción de los usuarios es positiva negativa o neutral durante el periodo de un año calendario.

Calculo de ISP: $(\text{Número total de reseñas positivas} - \text{Número total de reseñas negativas}) / \text{Número total de reseñas}.$

Formula : $(\text{ISP del año actual} - \text{ISP del año anterior}) / \text{ISP del año anterior}$

Objetivo: Mantener un índice de sentimiento positivo para asegurar una imagen favorable de los locales de la empresa.

- Tasa de Crecimiento de Reseñas Mensuales:

Definición: Este KPI evalúa la velocidad a la que las reseñas se están acumulando en Yelp y Google Maps en un período de tiempo específico.

Fórmula: $(\text{Número total de reseñas en el mes actual} - \text{Número total de reseñas en el mes anterior}) / \text{Número total de reseñas en el mes anterior}.$

Objetivo: Asegurarse de que la empresa está generando interés y participación continua en las plataformas, lo que podría indicar la relevancia y satisfacción de los clientes.

-Índice de Influencia del Usuario (IIU):

Definición: Este KPI mide la influencia de ciertos usuarios en la comunidad, considerando factores como la cantidad de seguidores, la frecuencia de reseñas, y la aceptación de sus recomendaciones por otros usuarios.

Fórmula: $(\text{Número de seguidores del usuario} * \text{Puntuación promedio de sus reseñas}) / \text{Número total de seguidores en la plataforma}.$

Objetivo: Identificar usuarios influyentes cuyas opiniones puedan tener un impacto significativo en la reputación de los locales y colaborar estratégicamente con ellos.

ANALISIS EDA



Estadística descriptiva

```
df_google.drop(columns=['user_id', 'latitude', 'longitude']).describe()
```

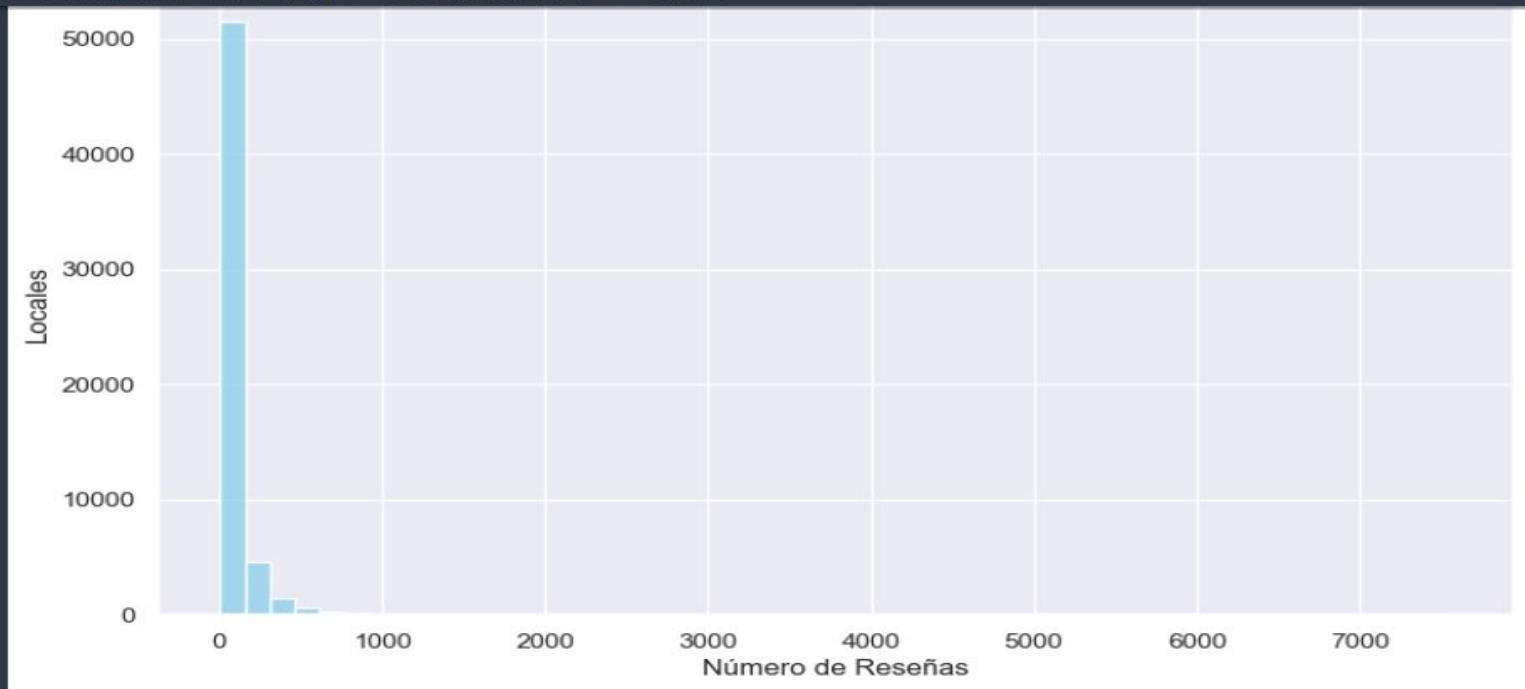
Python

	rating	avg_rating	num_of_reviews
count	82270	82270	82270
mean	4	4	142
std	1	0	341
min	1	1	5
25%	4	4	34
50%	5	4	57
75%	5	5	104
max	5	5	9998

- Viendo esta información vemos que contamos con 82.270 registros válidos en el Dataframe.
- En cuanto a los Ratings tenemos los siguientes datos :
 - La calificación mínima es 1, la calificación máxima es 5.
 - La calificación promedio es de 4, con una mediana de 5.
 - El 25% de las calificaciones son 4 o menos, el 50% son 5 y el 75% son 5. Esto sugiere que la mayoría de las calificaciones están en el rango superior.
- En cuanto al número de reviews tenemos los siguientes datos:
 - El promedio del número de reseñas es de 4, con una desviación estándar de 0.
 - La cantidad mínima de reseñas es 5, mientras que la máxima es 9998.
 - El 25% de los productos tienen al menos 34 reseñas, el 50% tienen al menos 57 reseñas y el 75% tienen al menos 104 reseñas. Esto indica una distribución sesgada hacia productos con más reseñas.



EDA profundo



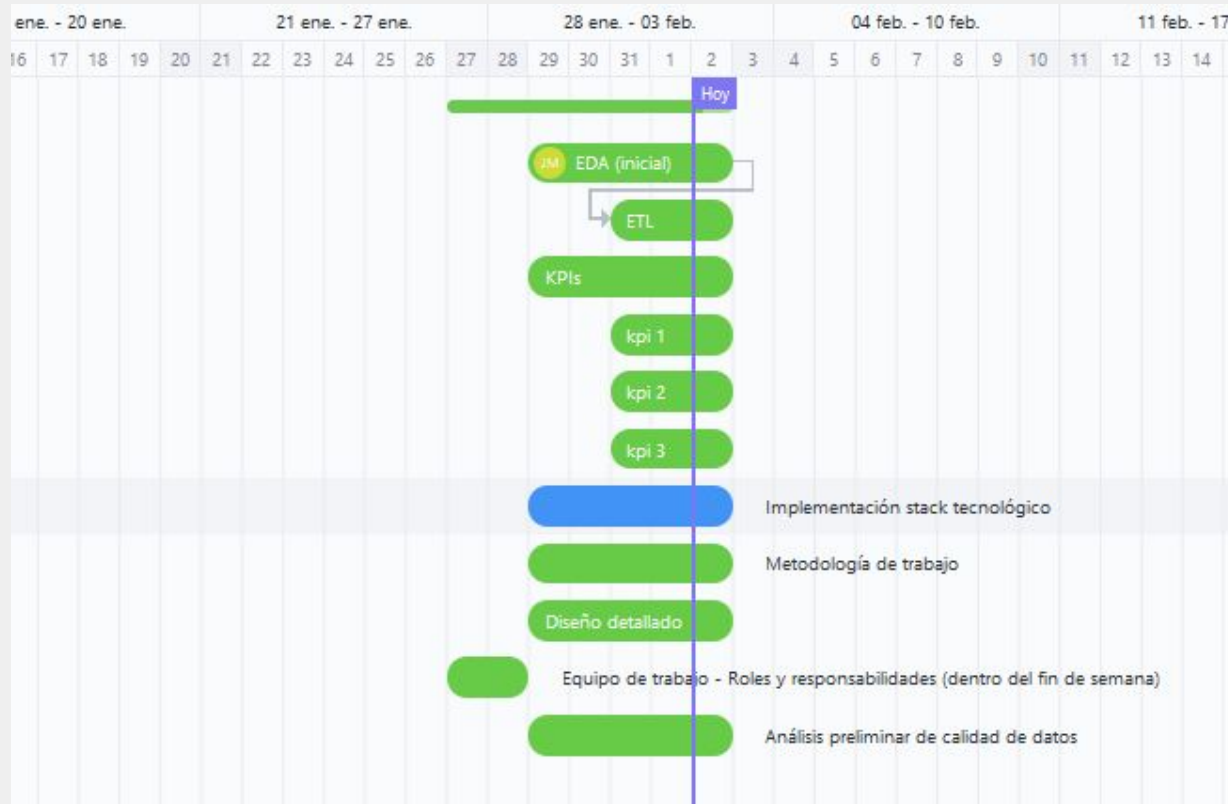
Aca podemos ver que se debe a que la inmensa mayoria de los locales tienen menos de 500 reviews, mientras que algunos superan los miles



Vertex.ai

Vertex AI será la herramienta a utilizar para entrenar los modelos de ML debido a la gran versatilidad e integración que ofrece

Diagrama de Gantt



ClickUp

DataMind Solutions STAFF



- 1..Project Manager...
- 2..Data Analyst...ETL..EDA..
- 3..Data Analyst...Dashboard
- 4..Data Architect..Deployment..Github, Google GCP, y todo lo relevante a buscar información sobre plataformas...escribir reportes e informes
- 5..Data Engineer...Machine Learning..

- 1__Chief Executive Officer__...**FER Abraham**
- 2__Data Engineer Department__...**THIAGO Sepulveda, Nico, Fer**
- 3__Business Intelligent Department__...**IGNACIO Waukuluk, Fer, Thiago**
- 4__Enterprise Architecture Department...**NICO Mussante, Juan**
- 5__Reserch and Developer Department__...**JUAN Mendoza, Ignacio**

Análisis Dashboard Usuarios Yelp!



Looker Studio

Analisis Dashboard KPI's



Looker Studio