



NEXUSANALYTICS

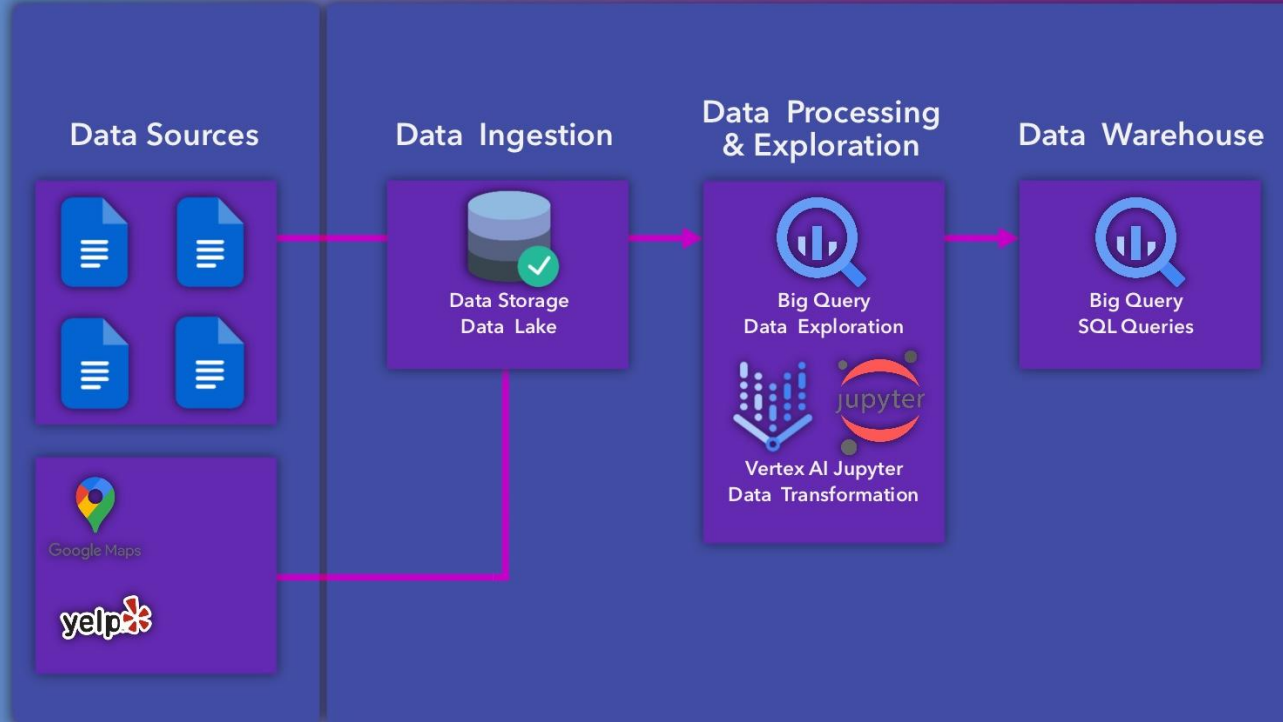
ETL

Yelp: business, reviews, check_in.

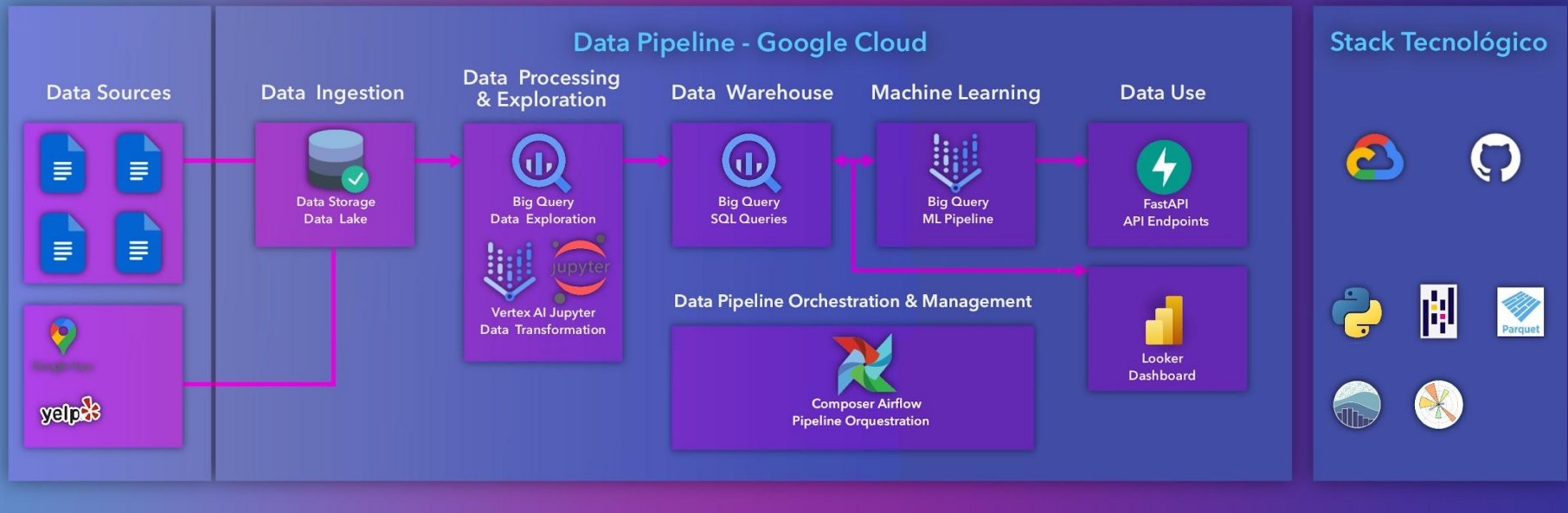
Google maps: metadata, reviews.

- Selección de columnas de interés.
- Filtrado de los datos para Restaurantes y Estaciones de Servicio.
- Filtrado de Restaurantes por comida étnica.
- Limpieza y transformación de datos.
- Aplicación de Análisis de Sentimientos en Reviews de usuarios.

Pipeline Processo ETL



Pipeline General



EDA

Para este Análisis exploratorio se tomaron como muestras los datos disponibles del año 2022.

- Tipos de variables.
- Comportamiento individual de variables.
- Relación entre variables.
- Registros por dataset.

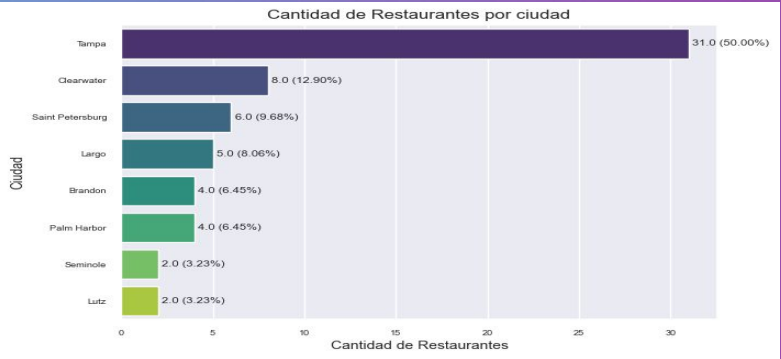
Yelp:

- **df_business**: 466 registros
- **df_review**: 23 registros
- **df_cheack_in**: 12 registros

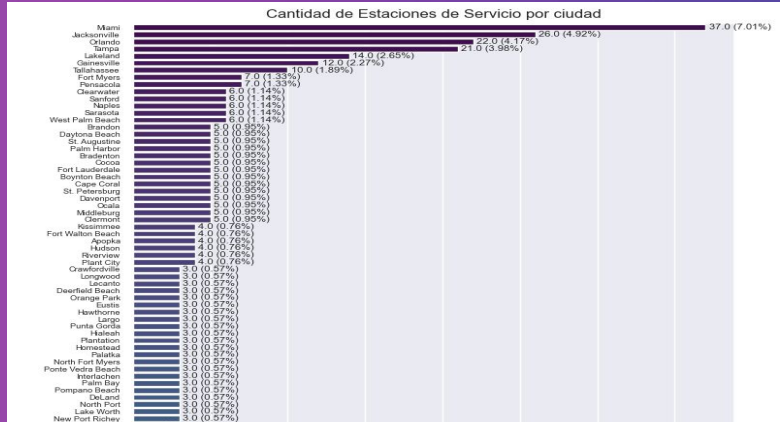
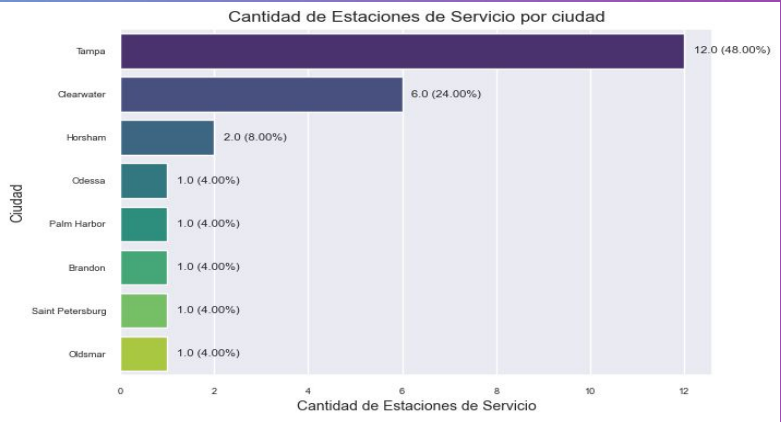
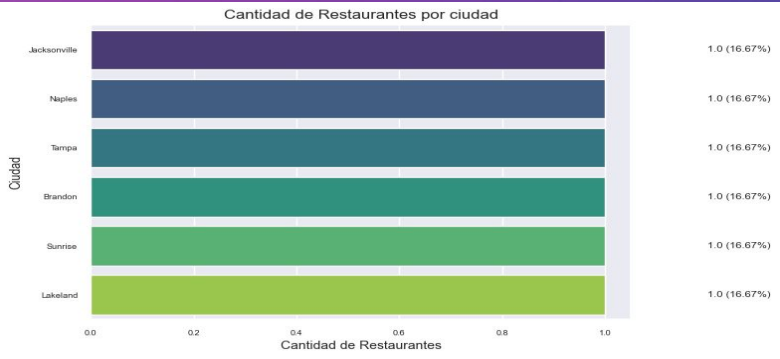
Google maps:

- **df_metadata**: 534 registros
- **df_review_states**: 2096 registros

Yelp:



Google maps:



Yelp:



Google maps:



Algunas Observaciones:

- Las categorías más populares son comida mexicana, italiana y china.
- La mayor cantidad de reviews se registran entre las 16 y 2 hs.
- Mayor cantidad de reviews tiende a mejorar la puntuación promedio.
- Tampa y Brandon aparecen entre las ciudades con mayor densidad de restaurantes en ambos dataset.
- Hay una gran diferencia en la cantidad de registros de Estaciones de Servicio entre ambos datasets.

Diagrama ER

Yelp:

review	
review_id	string
user_id	string
business_id	string
stars	int
date	string
hour	string
sentiment_analysis	int

checkin	
business_id	string
date	object
hour	object

business	
business_id	string
names	string
address	string
city	string
state	string
postal_code	string
latitude	float
longitude	float
stars	float
review_count	int
categories	object
hours	object

Google maps:

metadata	
name	string
address	string
city	string
state	string
postal_code	string
latitude	float
longitude	float
stars	float
review_count	float
categories	object
hours	object
gmap_id	string
url	string

review_states	
user_id	int
name	string
stars	int
text	string
gmap_id	string
date	string
hour	string
time	int
sentiment_analysis	int



Próximas acciones

- Extracción de Datos de Apis Yelp y Google.
- Automatización de procesos.
- Unificar información de las distintas fuentes.
- Desarrollo de sistema de recomendación con ML.
- Desarrollo de Api para consumir el modelo de ML.
- Coneccion de dashboard a Big Query.

Gracias!



NEXUSANALYTICS