





DESCRIPCION



Nuestro proyecto consiste en recopilar, depurar y analizar datos de reseñas de Yelp y Google Maps, utilizando técnicas de **análisis de sentimientos** y **machine learning** para predecir los rubros de los negocios que experimentarán un crecimiento o declive.

Además, se busca determinar las ubicaciones más adecuadas para establecer nuevos locales comerciales y desarrollar un sistema de recomendación de negocios basado en las preferencias de los usuarios en ambas plataformas. Aunque nos enfocaremos principalmente en restaurantes, la metodología puede aplicarse a otros tipos de comercios.



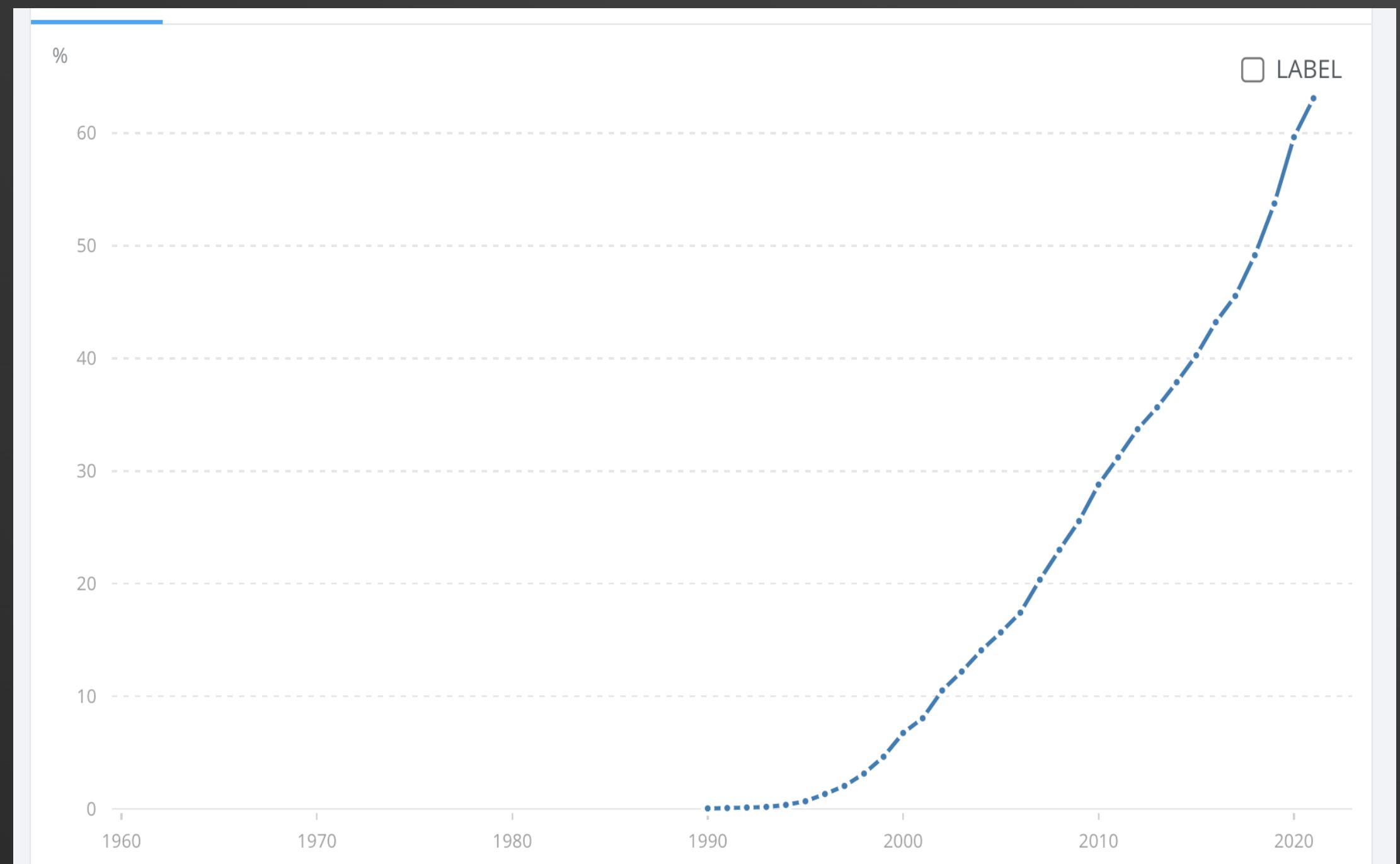


CONTEXTO

OPINIÓN USUARIOS
→ **GRAN CANTIDAD DE INFO**
 ↳ sobre la percepcion hacia
 diversos negocios

Permite evaluar: **DESEMPEÑO**
VISIBILIDAD
INTERACCION
DATOS DEMOGRAFICOS
Y GEOGRAFICOS

% individuos usando internet - THE WORLD BANK



OBJETIVO

Brindar un **análisis exhaustivo de la opinión de los usuarios** en Yelp y Google Maps.

Esto permitirá **identificar tendencias, predecir el crecimiento o decaimiento** de rubros comerciales y tomar decisiones estratégicas informadas para **mejorar decisiones de gestión e inversión de negocios.**

Además se buscará tener un **sistema de recomendación** de restaurantes para los usuarios de ambas plataformas para **darle al usuario la posibilidad de poder conocer nuevos negocios** basado en sus gustos y experiencias previas.





ALCANCE

RECOPILACIÓN DE DATOS

(extracción de datos)

DEPURACIÓN Y ALMACENAMIENTO DE DATOS

(Data Warehouse)

ANÁLISIS DE SENTIMIENTOS

(lenguaje natural NLP)

PREDICCIÓN DE TENDENCIAS

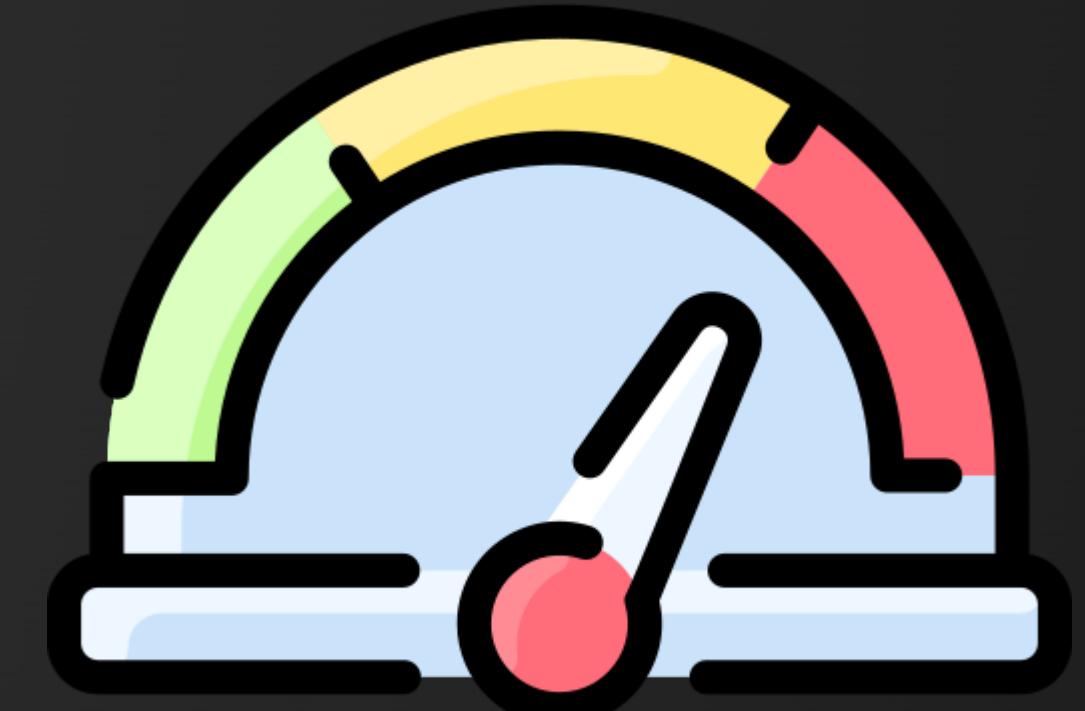
(modelos de machine learning)



*análisis adicional: cruzamiento con otros datasets

KPI's

- Determinar la satisfacción del cliente al utilizar el servicio de Yelp o Google.
- Determinar la retención y abandono de los clientes.
- Determinar los rubros de negocios que más crecieron o decayeron.
- Predecir la localización más conveniente de nuevos locales de estos negocios evaluados.
- Implementar un sistema de recomendaciones para usuarios.

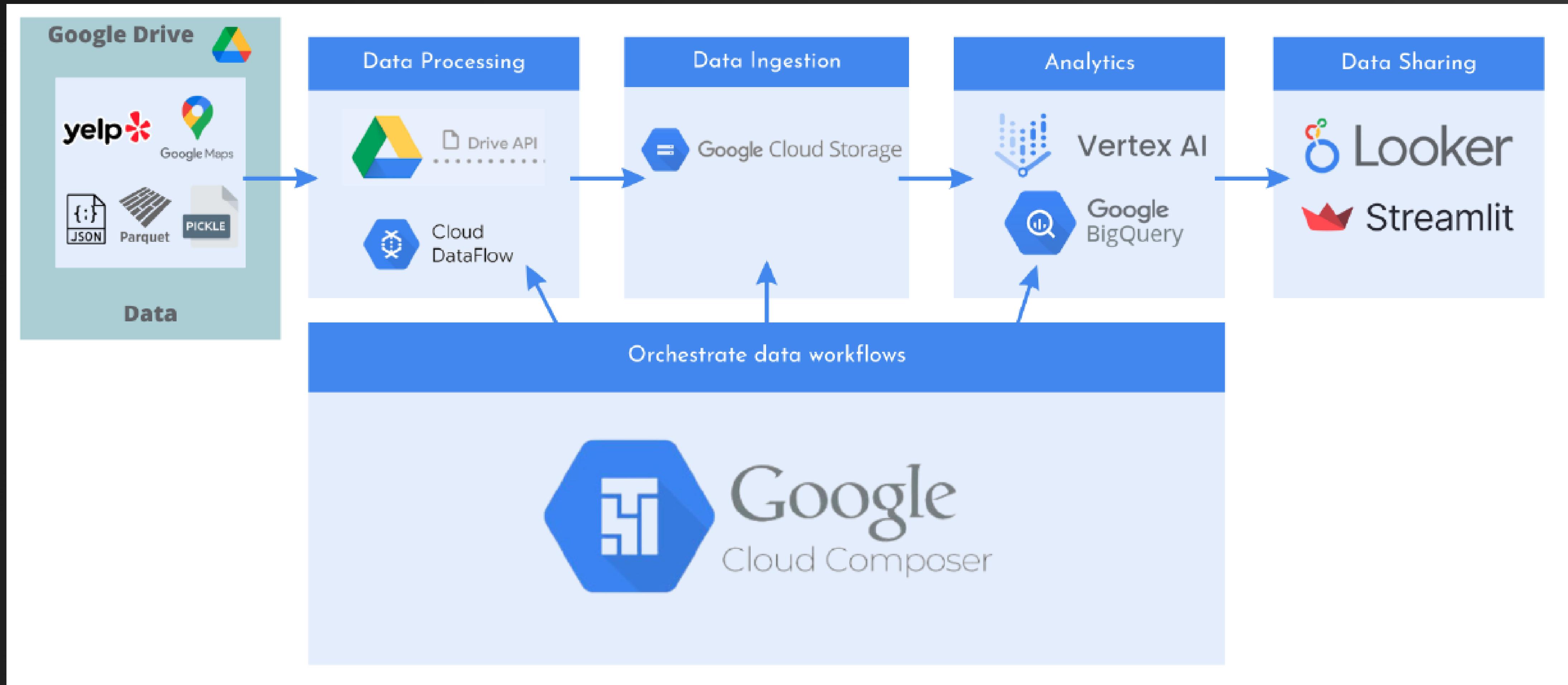


KPI	DESCRIPCION	FORMULA	FRECUENCIA
Promedio de calificación	Tasa variación porcentual del promedio de calificación de estrellas del 3%	$PCC = (\text{promedio estrellas año anterior}) - (\text{promedio estrellas año actual}) / (\text{promedio estrellas año actual}) * 100$	Anual
Cierre de sucursales	Porcentaje de sucursales cerradas por rubro	$CSC = (\text{cantidad de sucursales cerradas}) / (\text{cantidad de sucursales totales por rubro}) * 100$	Mensual
Satisfacción del cliente	Satisfacción del cliente por reseña	(calificación de estrellas)	Semanal
Abandono del cliente al negocio	Reducción del 5% la tasa de abandono global	$TAG: (\text{clientes perdidos}) / (\text{total de clientes al inicio del año}) * 100$	Anual

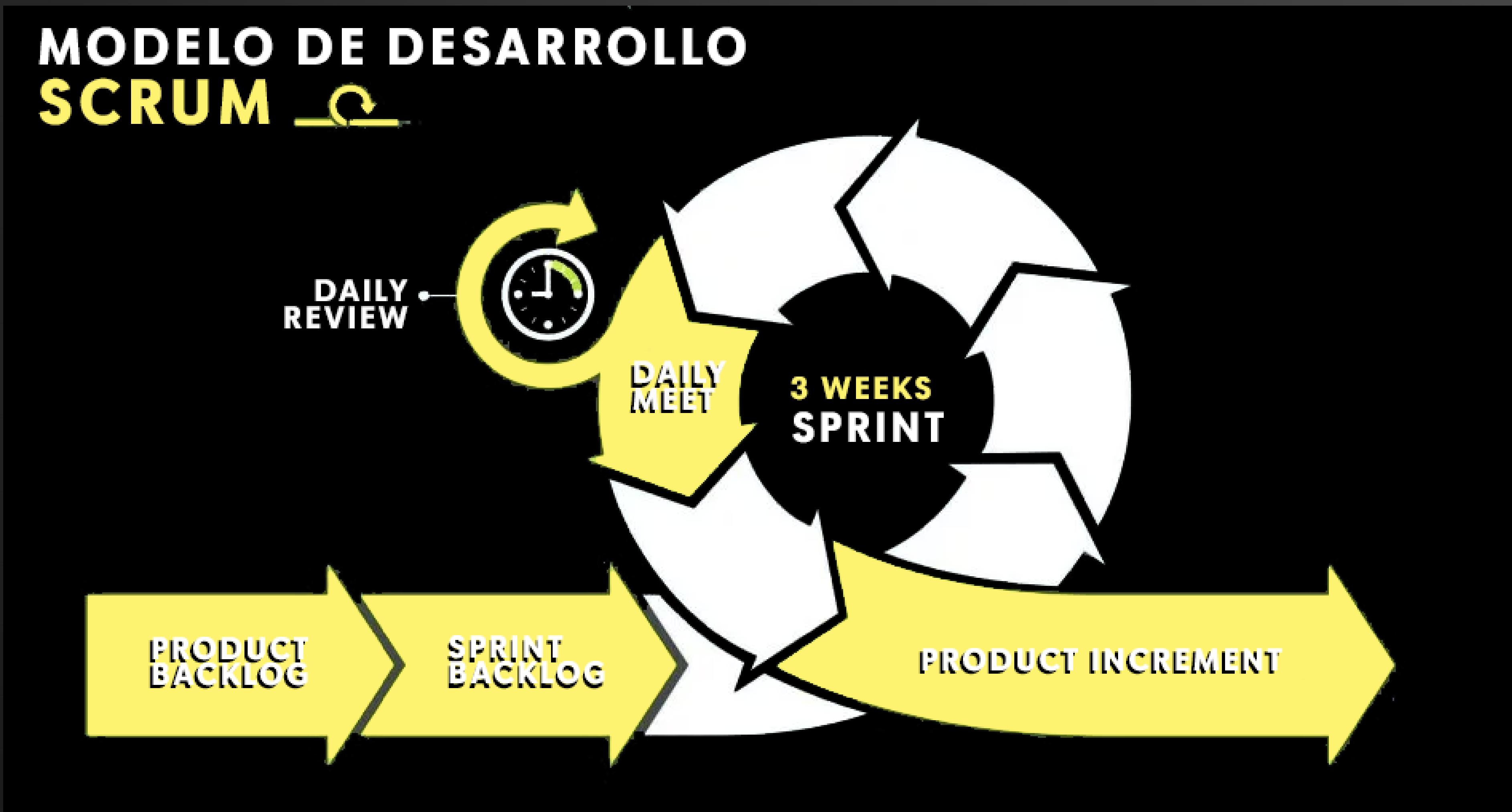
KPI	DESCRIPCION	FORMULA	FRECUENCIA
Capacidad de retención del cliente	Índice de tasa de retención de clientes	$IRTC = ((CE-CN) / CS) * 100$ CE = El número total de clientes al final del periodo CN = El número total de nuevos clientes que ha adquirido durante el periodo CS = El número total de clientes al inicio de un periodo	Mensual
Abandono de cliente mensual	Tasa de abandono de cliente	$TAC = (\text{total de clientes al principio del mes} - \text{total de clientes al final del mes}) / (\text{total de clientes al principio del mes}) * 100$	Mensual
Reducción de pérdida de cliente	Reducir los clientes perdidos en un 5%	$CP = (\text{total de clientes al inicio del mes(año)} - \text{total de clientes al final mes(año)})$	Mensual



↗ PROUESTA STACK TECNOLOGICO



↗ PROPUESTA METODOLOGIA



✓ PROPUUESTA EQUIPO DE TRABAJO

SCRUM ROLES:



PRODUCT OWNER:
CAROLINA VILLAGRA



SCRUM MASTER:
JULIAN MEDIAVILLA



TEAM MEMBERS:

CLARITZO PEREZ /data_analyst

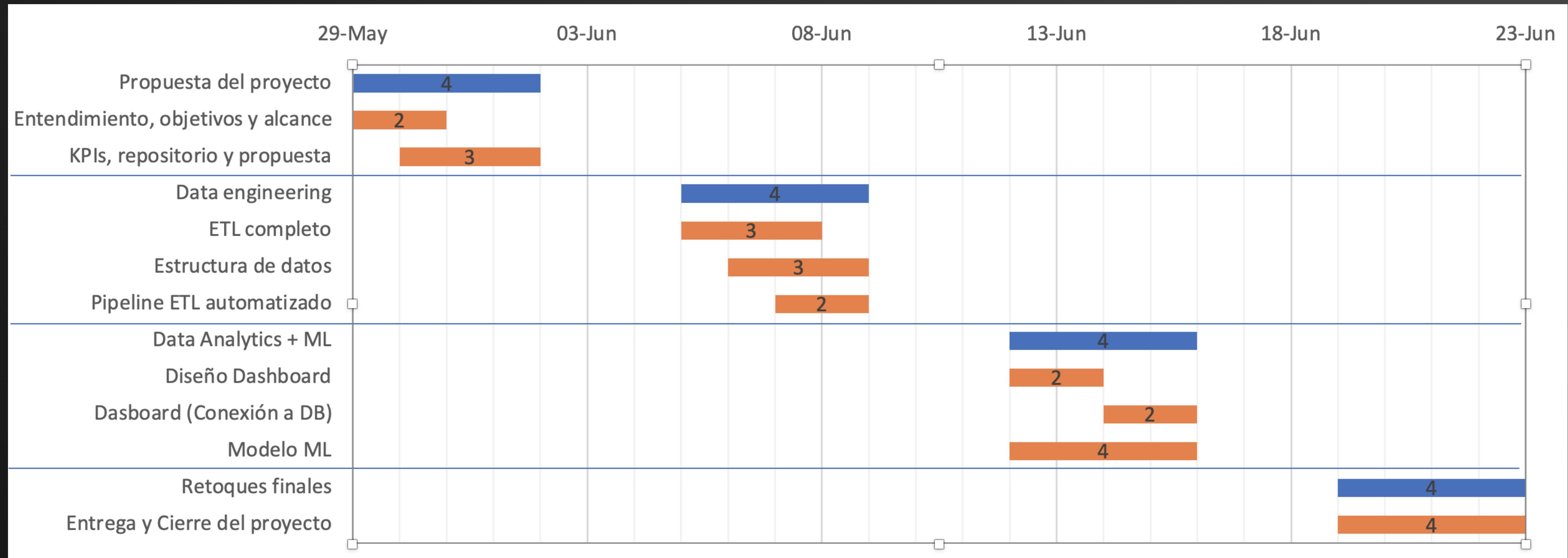
PAULA PALLARES /functional_analysis

BENJAMIN ZAMBELLI /data_engineer

BEDER RIVERA /data_engineer

GONZALO SCHWERDT /ml_engineer

↗ PROUESTA CRONOGRAMA GANTT





ANALISIS PRELIMINAR DE DATOS



Google Maps

⚡ Google Maps dataset metadata-sitios

Observaciones:

Hay 3.025.011 filas y 15 columnas.

Número de valores nulos:

"MISC": 690.834

"dirección": 80.511

"categoría": 17.419

"descripción": 2.770.722

"horario": 787.405

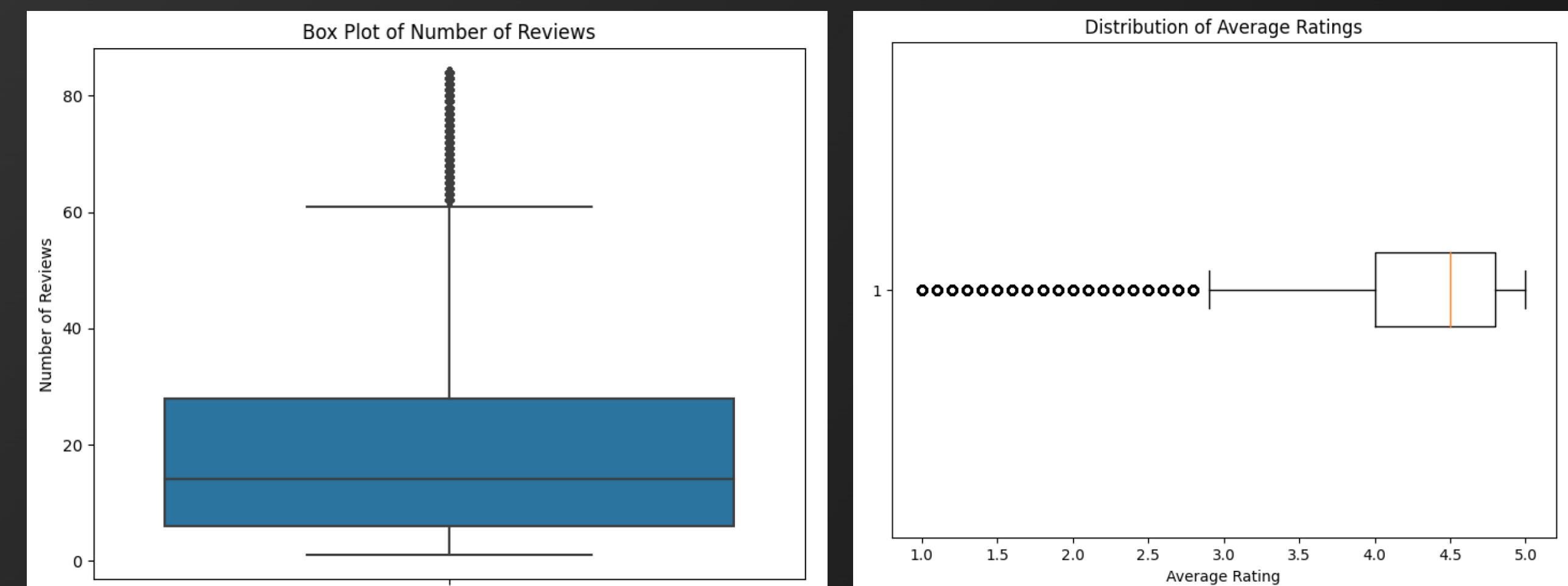
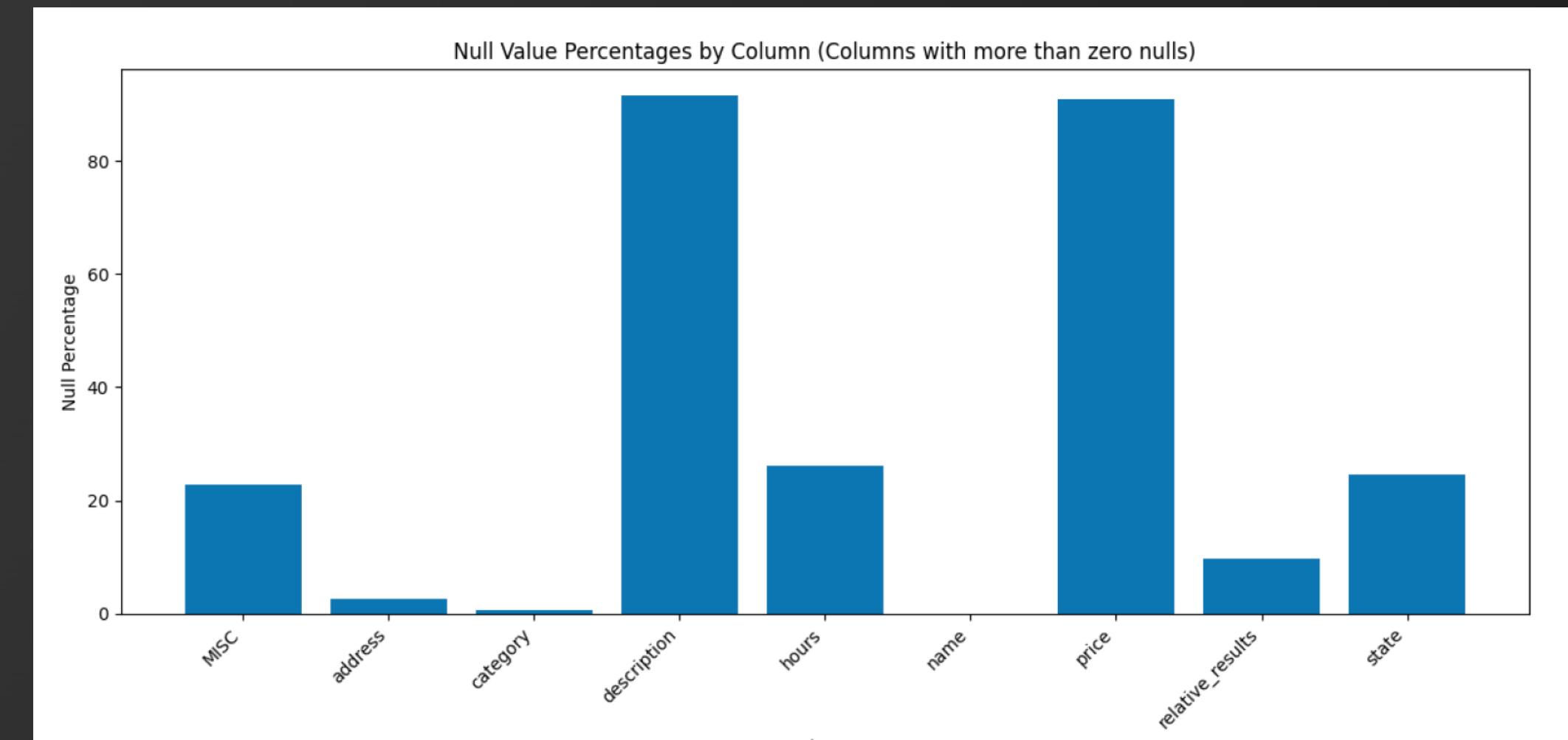
"nombre": 37

"precio": 2.749.808

"resultados_relativos": 295.058

"estado": 746.455

Parece haber 26.573 duplicados





ANALISIS PRELIMINAR DE DATOS



Google Maps dataset review-estados

Observaciones:

Hay 89.946.359 filas y 9 columnas.

Hay valores nulos en las columnas:

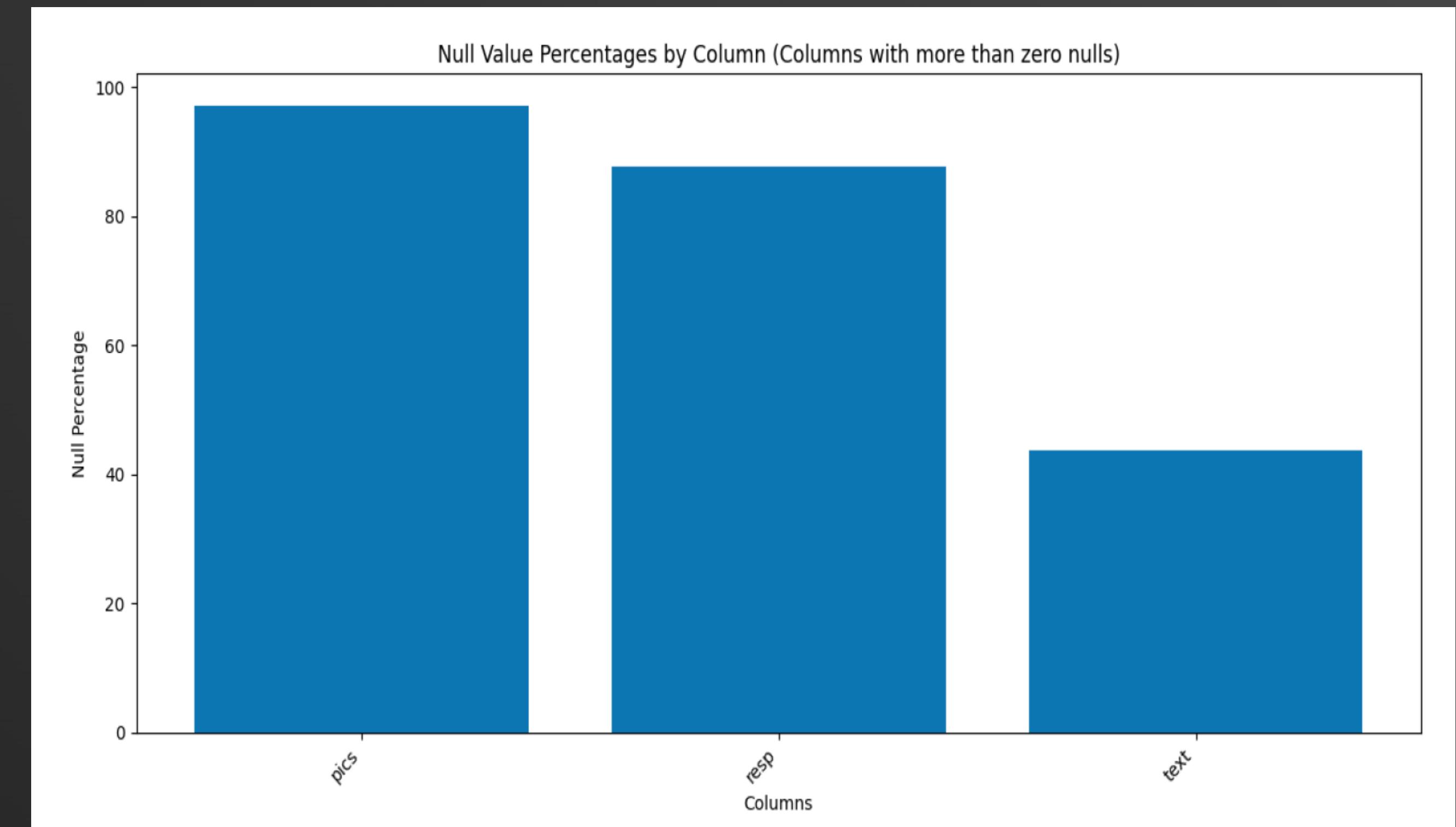
"imágenes" (87.450.680)

"respuesta" (78.917.249)

"texto" (39.307.744)

Parece haber 1.386.843 duplicados

La columna "tiempo" está en formato de sello de tiempo Unix, por lo que debemos convertirla a tipo de fecha para el análisis.



↗ ANALISIS PRELIMINAR DE DATOS



↗ Yelp! dataset **business.pkl**

Observaciones:

Hay 150.346 filas y 28 (14) columnas.

¡Las columnas se repiten dos veces!

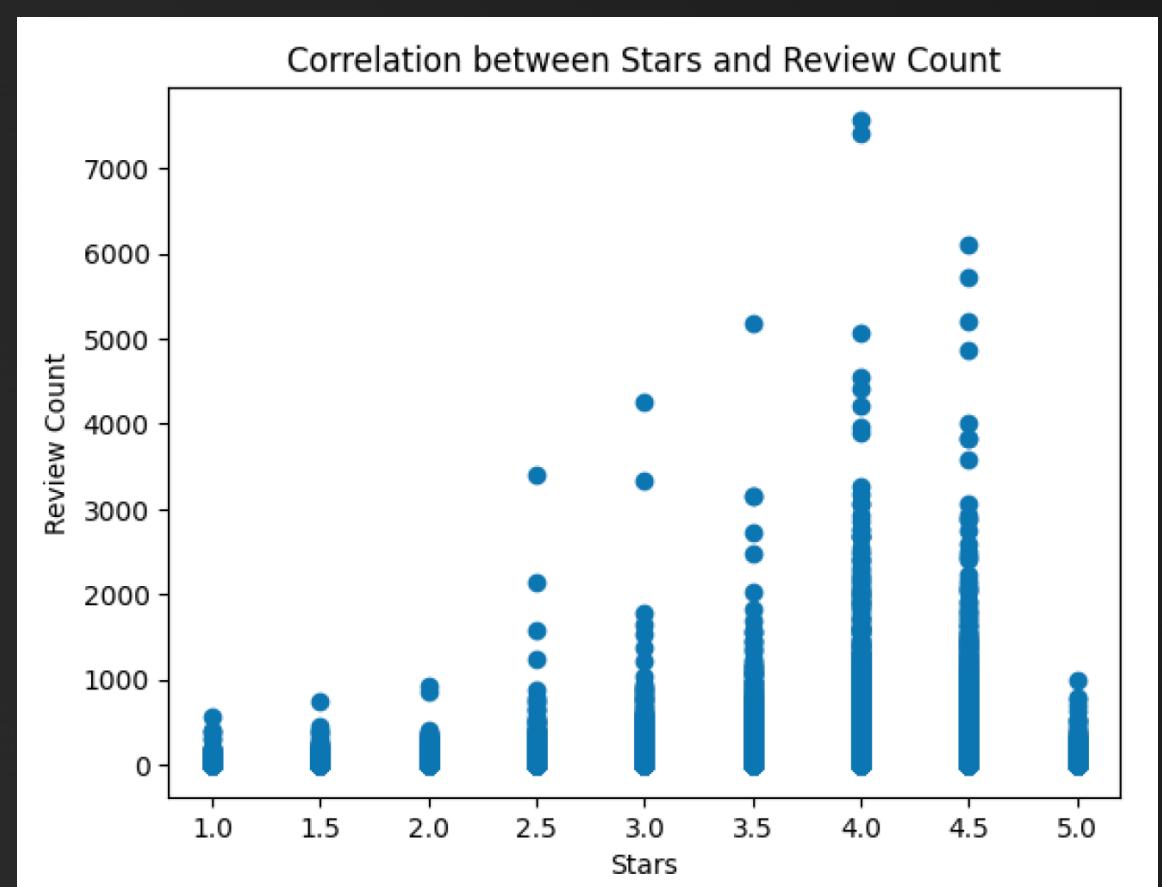
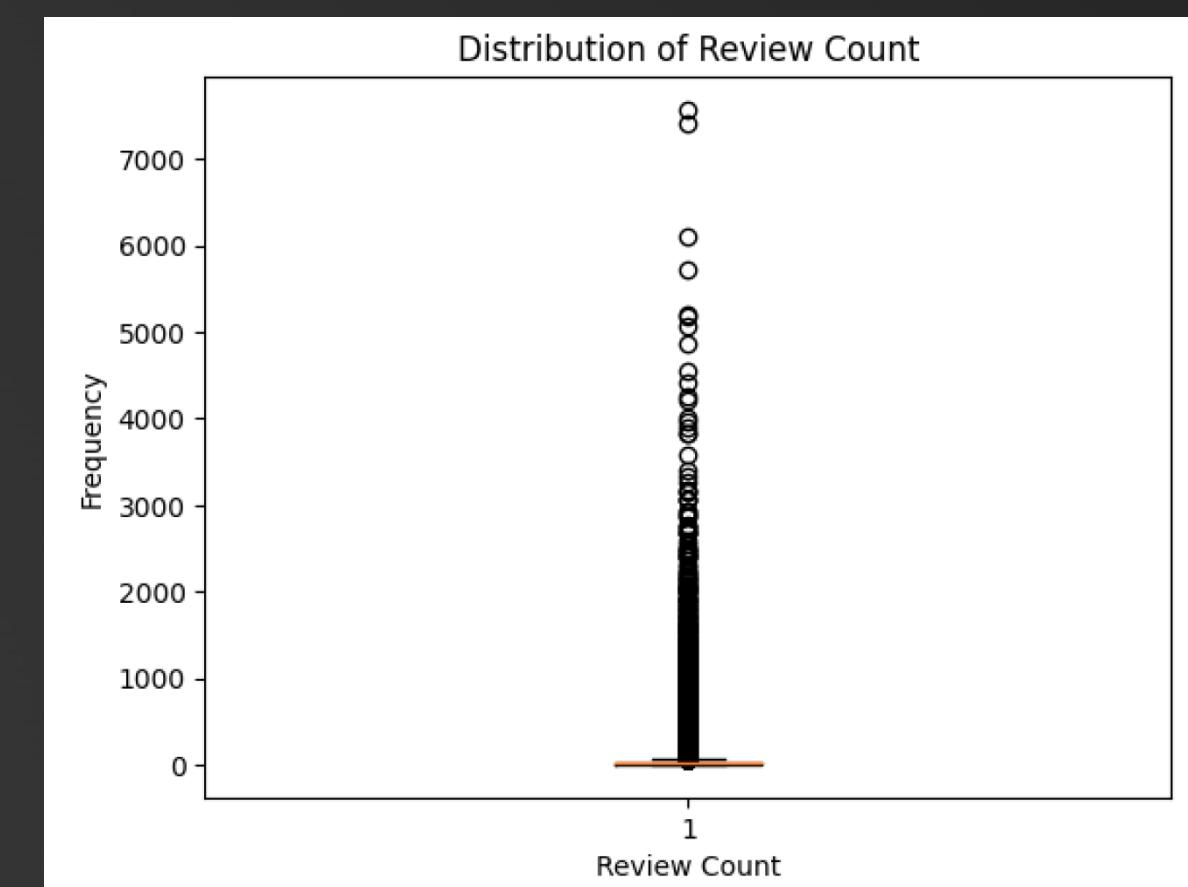
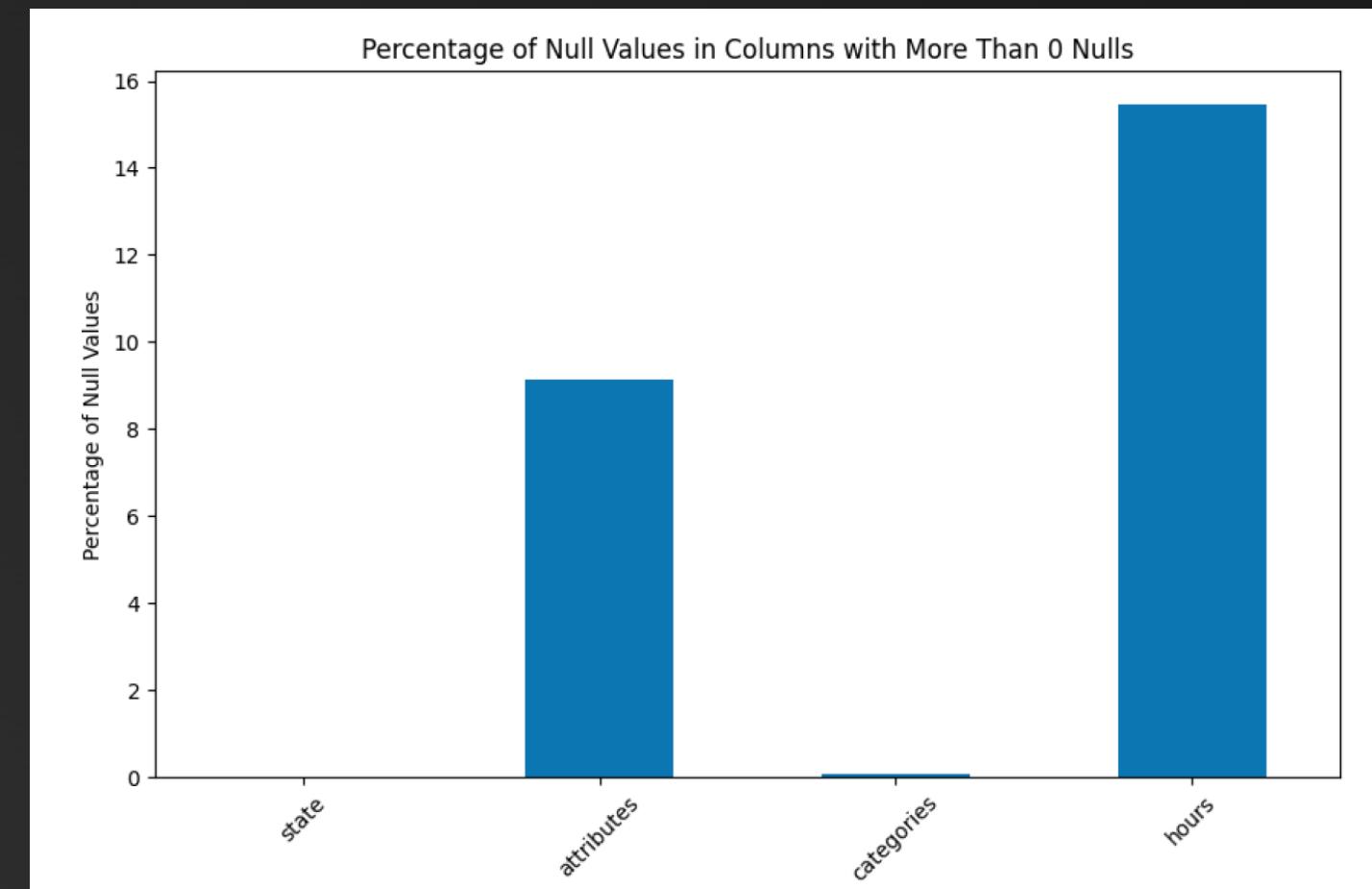
Aparentemente no hay duplicados

Todo es de tipo objeto. Las columnas 'estrellas', 'latitud' y 'longitud' deberían ser de tipo flotante, las columnas 'cantidad_de_reseñas' y 'está_abierto' deberían ser de tipo entero

Hay valores nulos en las columnas: estado (3), atributos (13744), categorías (103), horarios (23223)

Hay algunos lugares con muchas más reseñas (posibles valores atípicos).

Parece haber muchas reseñas entre 2.5 y 4.5 estrellas y menos entre 1 y 2, y específicamente 5.



↗ ANALISIS PRELIMINAR DE DATOS



↗ Yelp! dataset **review.json**

Observaciones:

Hay 6.990.280 filas y 9 columnas.
No parece haber valores nulos o duplicados.
La columna de fecha es de tipo cadena de texto.

↗ Yelp! dataset **user.parquet**

Observaciones:

Hay 2.105.597 filas y 22 columnas.
No parece haber valores nulos.
Parece haber 117.700 valores duplicados.



 CON ESTE TRABAJO APUNTAMOS A
COMPRENDER MEJOR LA IMAGEN QUE
TIENEN LOS USUARIOS DE SUS NEGOCIOS,
IDENTIFICAR ÁREAS DE MEJORA,
TOMAR DECISIONES MÁS INFORMADAS
Y ESTRATÉGICAS, Y **MEJORAR LA CALIDAD**
DE SUS SERVICIOS PARA SATISFACER LAS
NECESIDADES Y EXPECTATIVAS DE LOS
CLIENTES DE MANERA MÁS EFECTIVA.



GRACIAS 