

RECOMMENDATION SYSTEM: FAST FOOD FRANCHISES



MINDFUL DATA

Indice

Indice.....	2
Introducción.....	3
Necesidad del cliente.....	3
Objetivos.....	4
Alcance.....	5
Roles.....	5
Metodología de trabajo.....	6
Organización de las tareas.....	7
Diagrama de Gantt.....	8
Tecnologías a utilizar.....	9
Workflow.....	10
EDA preliminar.....	11
KPIs.....	16

Introducción

En la era digital, la opinión de los usuarios se ha convertido en un recurso invaluable para las empresas de todo tipo. El constante flujo de reseñas y comentarios que fluyen a través de plataformas de reseñas como Yelp y Google Maps proporciona una ventana única hacia la percepción que los clientes tienen de los negocios, restaurantes, hoteles y servicios. Estas opiniones, cargadas de experiencias personales, se han convertido en un factor determinante en la planificación de estrategias empresariales.

Yelp, conocida como una de las plataformas líderes en el ámbito de las reseñas, ofrece a los usuarios la oportunidad de compartir sus vivencias en una amplia gama de establecimientos. Desde restaurantes exquisitos hasta servicios esenciales, los clientes comparten sus pensamientos y calificaciones, generando una fuente de información esencial para las empresas. Estas opiniones no solo ayudan a las empresas a comprender cómo son percibidas por su clientela, sino que también sirven como un medio para evaluar su desempeño y determinar áreas de mejora.

Por otro lado, Google Maps, una herramienta omnipresente para la localización y navegación, ha incorporado una función de reseñas que permite a los usuarios dejar sus impresiones sobre negocios y lugares de interés. Esta característica se ha convertido en un recurso crucial para aquellos que buscan información sobre dónde comprar, comer, dormir o reunirse. Las opiniones en Google Maps no solo brindan información sobre la calidad de los establecimientos, sino que también influyen en las decisiones de los usuarios a la hora de elegir entre las diversas opciones disponibles.

En este contexto, exploraremos cómo las opiniones de los usuarios en Yelp y Google Maps se han convertido en un valioso activo permitiendo que las empresas puedan medir su desempeño, comprender la utilidad de sus locales y, sobre todo, identificar áreas de mejora en sus servicios. A lo largo de esta exploración, descubriremos cómo estas plataformas han transformado la forma en que los consumidores toman decisiones y cómo las empresas utilizan esta información para impulsar sus estrategias comerciales.

Necesidad del cliente

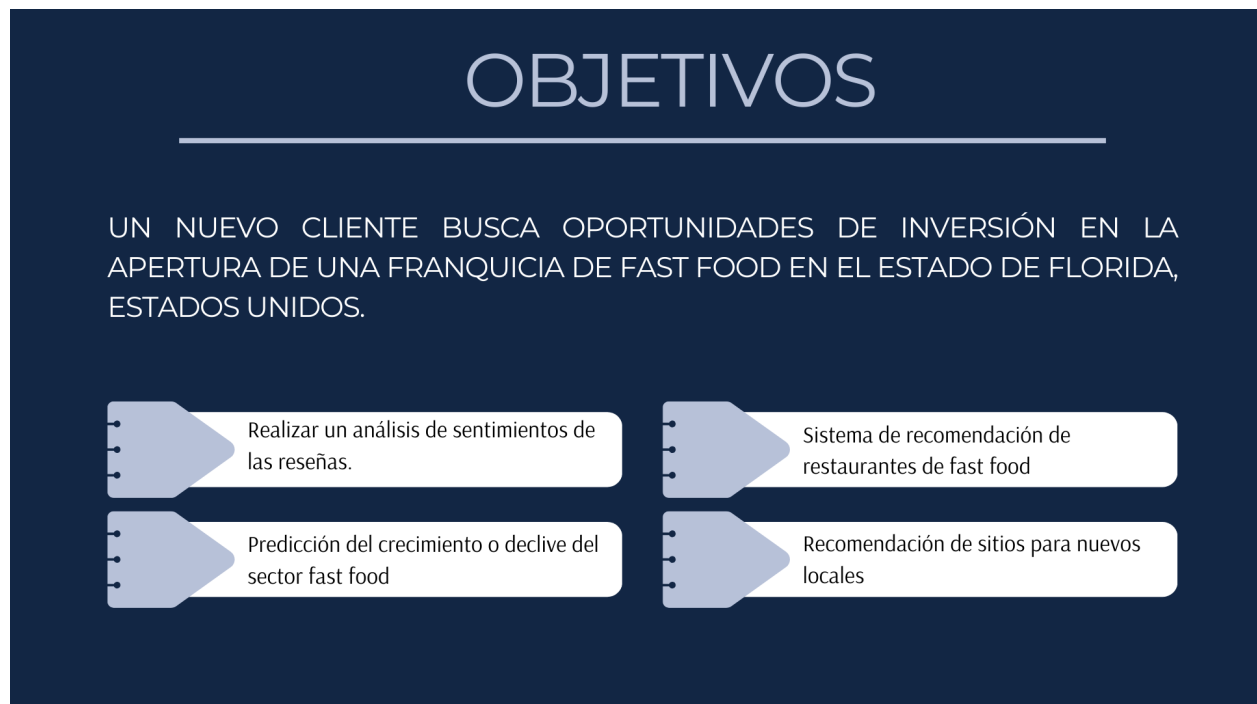
Somos la consultora Mindful Data y ha llegado a nuestra consultora un nuevo cliente, un inversor gastronómico sin experiencia previa en la industria de la comida rápida, que busca oportunidades de inversión en la apertura de una franquicia de fast food en el estado de Florida, Estados Unidos. Este informe tiene como objetivo analizar las necesidades específicas del cliente y proporcionar recomendaciones para satisfacer sus requerimientos.

Objetivos

El objetivo de este proyecto es realizar un análisis del mercado estadounidense específicamente del estado de Florida en el sector de restaurantes de comida rápida. El cliente que nos ha contratado es un inversor que planea instalar una franquicia de comida rápida en Florida y quiere que nuestra consultora de datos (Mindful Data) realice un análisis de dos fuentes principales de información: las opiniones de los usuarios en Yelp y Google Maps para recomendarle qué franquicia de fast food le conviene adquirir.

Para cumplir con este objetivo, nuestro equipo se ha propuesto las siguientes metas:

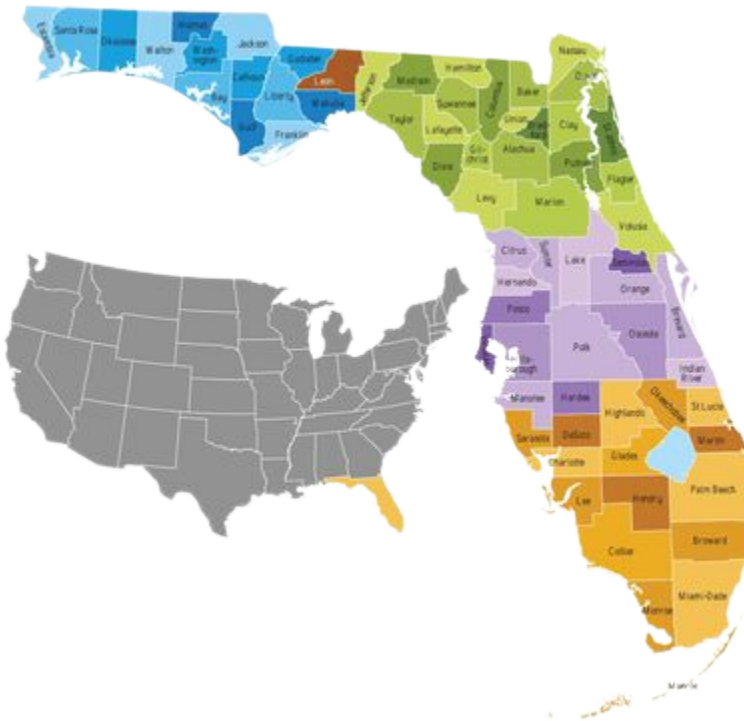
- ❖ Realizar un análisis de sentimientos y opiniones de las reseñas de Yelp y Google Maps para comprender la percepción de los usuarios sobre el sector fast food en Florida, Estados Unidos.
- ❖ Predicción del crecimiento o declive del sector fast food basándonos en datos históricos. Para ello, utilizaremos diferentes técnicas de machine learning.
- ❖ Sistema de recomendación de restaurantes de fast food.
- ❖ Recomendación de sitios para nuevos locales mediante un modelo de machine learning, se identifican las ciudades dentro del estado de Florida para emplazar nuevos locales del sector en función de reseñas positivas.



Alcance

El cliente, un inversor gastronómico sin experiencia previa en ningún rubro, ha manifestado su interés en establecer una franquicia de fast food en el estado de Florida, Estados Unidos. Como parte de Mindful Data Solutions, se ha definido el alcance del proyecto para abordar sus necesidades y garantizar una planificación y ejecución efectiva.

Por lo tanto, este proyecto se centrará en el análisis de las reseñas de los últimos 5 años de Yelp y Google Maps acerca de las cadenas de fast food más importantes de Estados Unidos con el objetivo de identificar cuál es la mejor opción a la hora de adquirir una franquicia y determinar en qué ciudad del estado de Florida conviene colocarla.



Roles

Dentro de nuestra Mindful Data Solutions hemos definido el grupo de trabajo para darle a curso a este proyecto, y los roles a cumplir por cada uno de los integrantes del mismo.

Cabe mencionar, que como la metodología de trabajo es por etapas, los roles pueden verse sujetos a modificación en función de las necesidades del proyecto.

- ❖ Data/Cloud Engineer: Sebastián Di Nesta, Joaquín Laurencio
- ❖ Data Analytics: Williams Amaro, Betiana Lopez Andueza
- ❖ Data Scientist: Nathalie Saravia



Metodología de trabajo

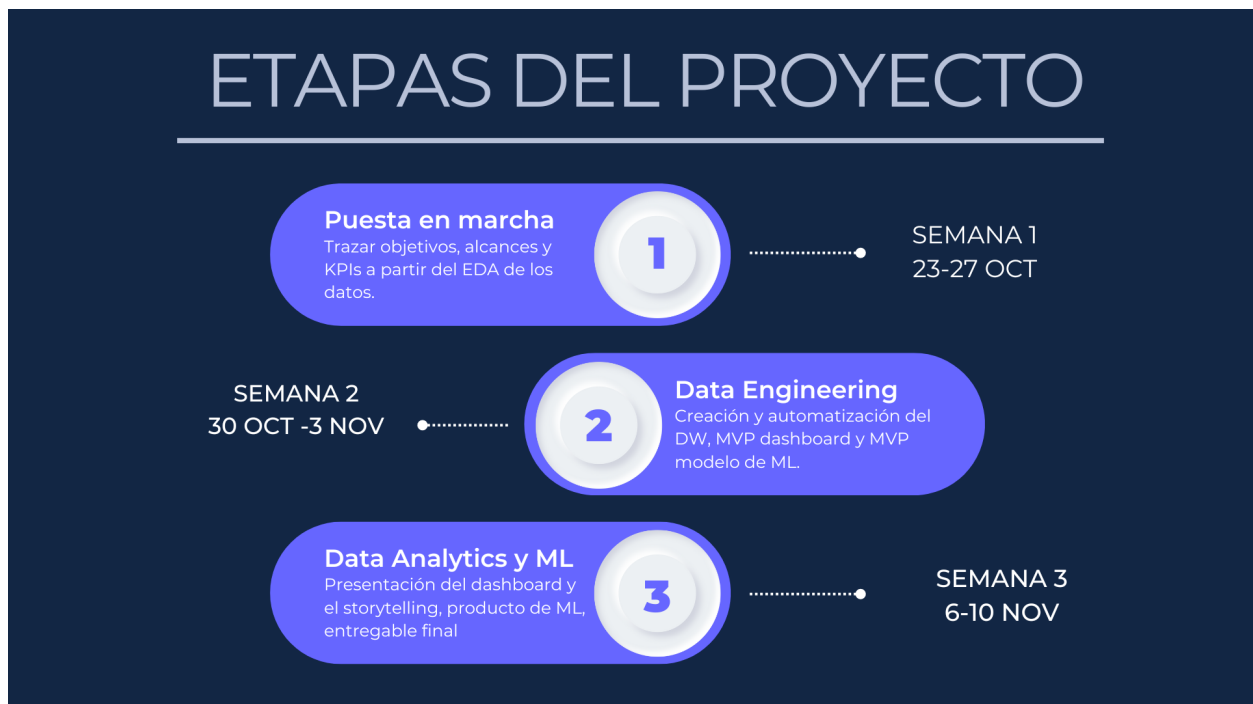
Se adopta la metodología de trabajo Scrum con entregas parciales sobre los avances del proyecto.

Se realizarán reuniones de seguimiento (dailies) en conjunto con el Scrum Master, en donde se revisarán las tareas realizadas hasta el momento, se establecerán cuáles se realizarán a continuación y se darán a conocer los problemas o dificultades que se están teniendo para poder encontrar una solución y avanzar en el proceso hacia el producto final.

El proyecto se ha dividido en 3 etapas o sprints, cada uno con una duración de una semana. En cada una de ellas se presenta un informe de avance con su correspondiente producto (entregable). Al final de cada sprint se revisan los errores o cambios que el cliente (Product Owner) solicite para incorporar estos requerimientos en la siguiente instancia.

A lo largo de las semanas se presentarán:

- Semana 1: Esta etapa constituye la puesta en marcha del proyecto y el tratamiento de los datos. Se trazan los objetivos, los alcances del proyecto así como los KPIs a evaluar. Además, se realiza el EDA preliminar de los datos.
- Semana 2: Etapa de Data Engineering. Se crea, se implementa y se automatiza el datawarehouse. Además, se plantean los MVPs del dashboard y de los modelos de Machine Learning preliminares.
- Semana 3: Etapa final. Se completa principalmente el dashboard y se presenta el producto de Machine Learning. Además se realiza la entrega de la documentación y del repositorio completo.



Organización de las tareas

Se implementó la utilización de Trello como herramienta para la gestión del proyecto y las tareas. A través de un sistema de tableros y tarjetas fuimos asignando responsabilidades, estableciendo plazos y nos permitió realizar un seguimiento del progreso de nuestro proyecto.



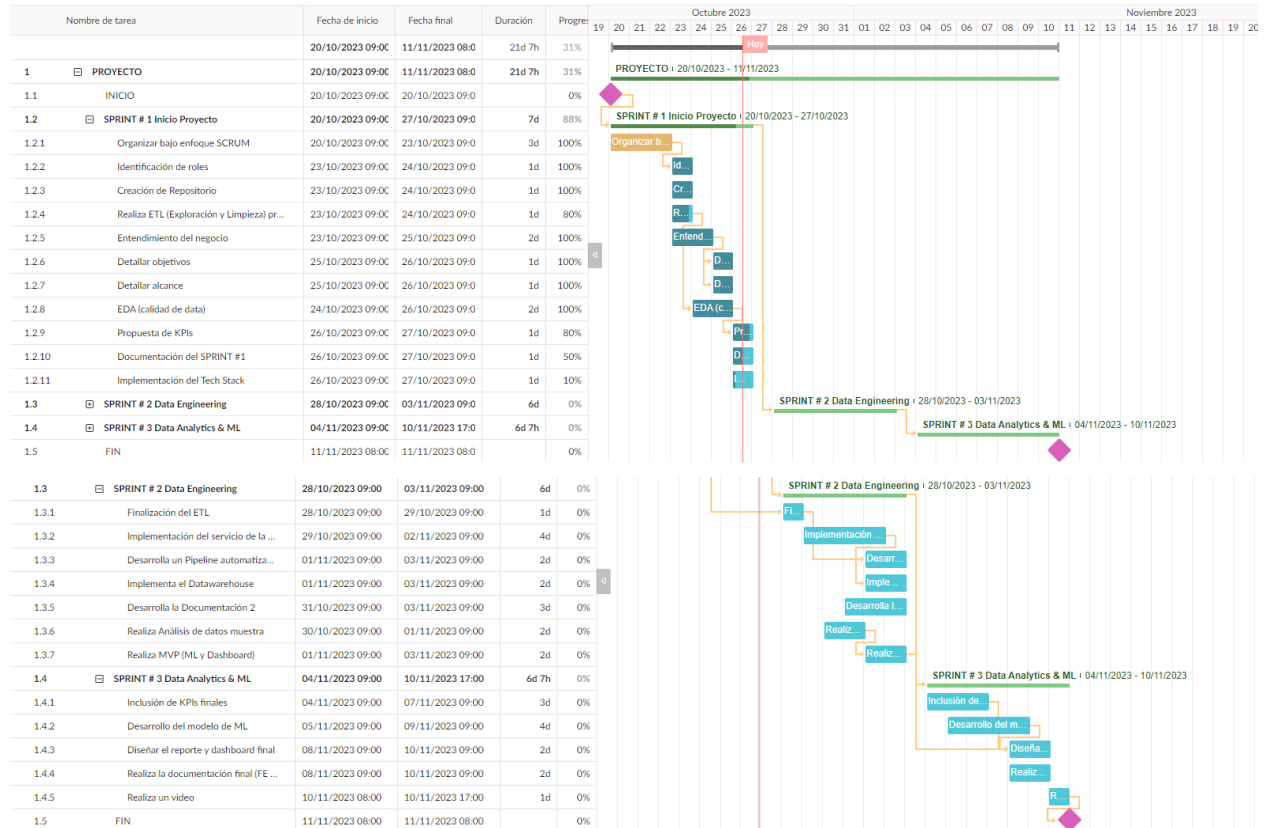
Diagrama de Gantt

A continuación se presenta el diagrama de Gantt correspondiente a la distribución de las actividades a lo largo del proyecto.

Conforme a la entrega del primer sprint (27/10/2023) se designan las siguientes actividades.

- Organización bajo la metodología SCRUM
- Identificación de roles
- Creación de Repositorio
- ETL (Exploración y Limpieza)
- Entendimiento del negocio
- Detallar objetivos
- Detallar alcances
- EDA (calidad de data)
- KPIs propuestos
- Documentación
- Stack tecnológico

INFORME DE AVANCE: PROYECTO FINAL - SEMANA 1



Para acceder a la visualización del diagrama de Gantt en la herramienta Gantt Pro haz click en el siguiente enlace: [GanttPro](#)

Tecnologías a utilizar

Para la realización del proyecto se seleccionaron diversas tecnologías, cada una con un propósito específico..

- Trello: como plataforma para la gestión de las tareas
- Visual Studio Code: como medio para realizar el EDA preliminar en local
- Kaggle: como medio para realizar el EDA preliminar.
- Google Colab: como medio para realizar el EDA preliminar.
- GitHub: como repositorio del proyecto

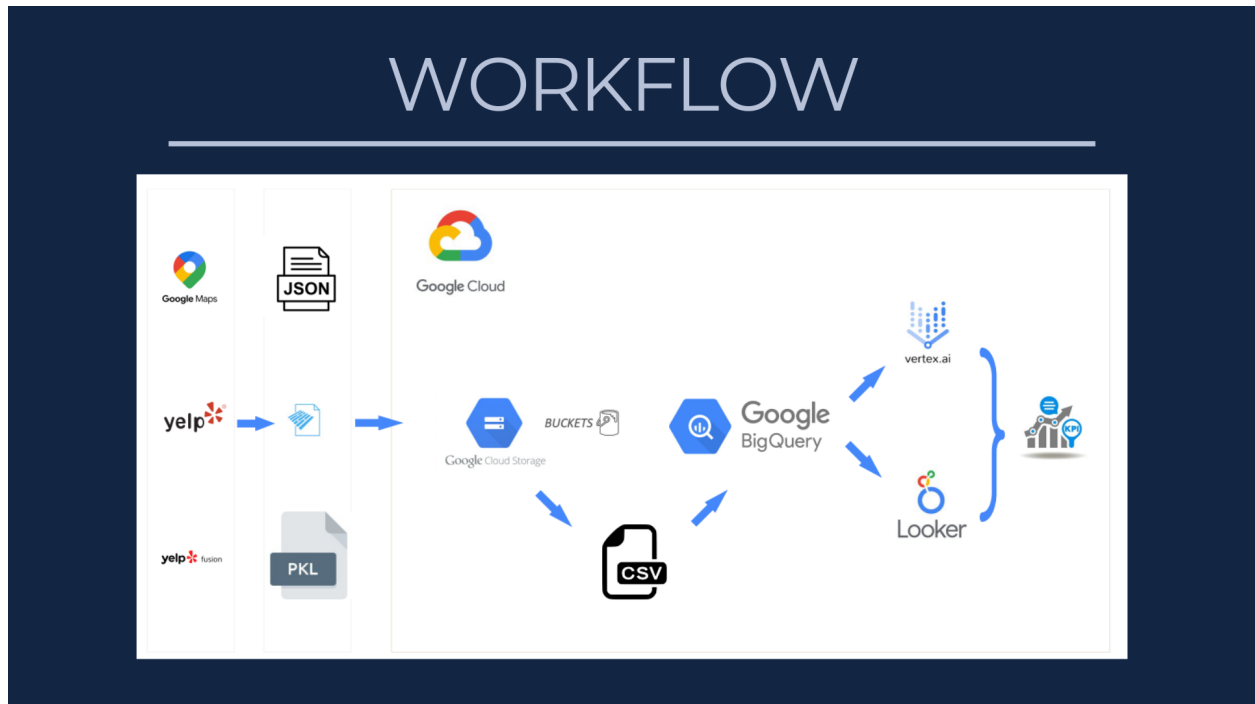


Workflow

Optamos por Google Cloud como servicio en la nube para gestionar los datos, automatizar procesos y para el procesamiento de aprendizaje automático :

- Cloud Storage Buckets como datalake
- Vertex Workbench para realizar el ETL de los datos
- BigQuery, como datawarehouse para almacenar y procesar los datos.
- Vertex AI para implementar los modelos de aprendizaje.
- Looker Studio como herramienta de visualización de datos y KPIs.

Cabe destacar que es posible que se modifiquen y/o incorporen nuevas tecnologías a medida que se avance en el desarrollo del proyecto.



EDA preliminar

En esta sección se realiza un primer análisis de los datos provistos por YELP y Google Maps. El objetivo de esta instancia es identificar los tipos de datos, conocer su calidad y los potenciales procesos de transformación a los que deberá ser sometida la data. Cabe mencionar, que en esta instancia no se intenta realizar un análisis exploratorio en profundidad, sino más bien tener un primer panorama de los dataset con su descripción general.

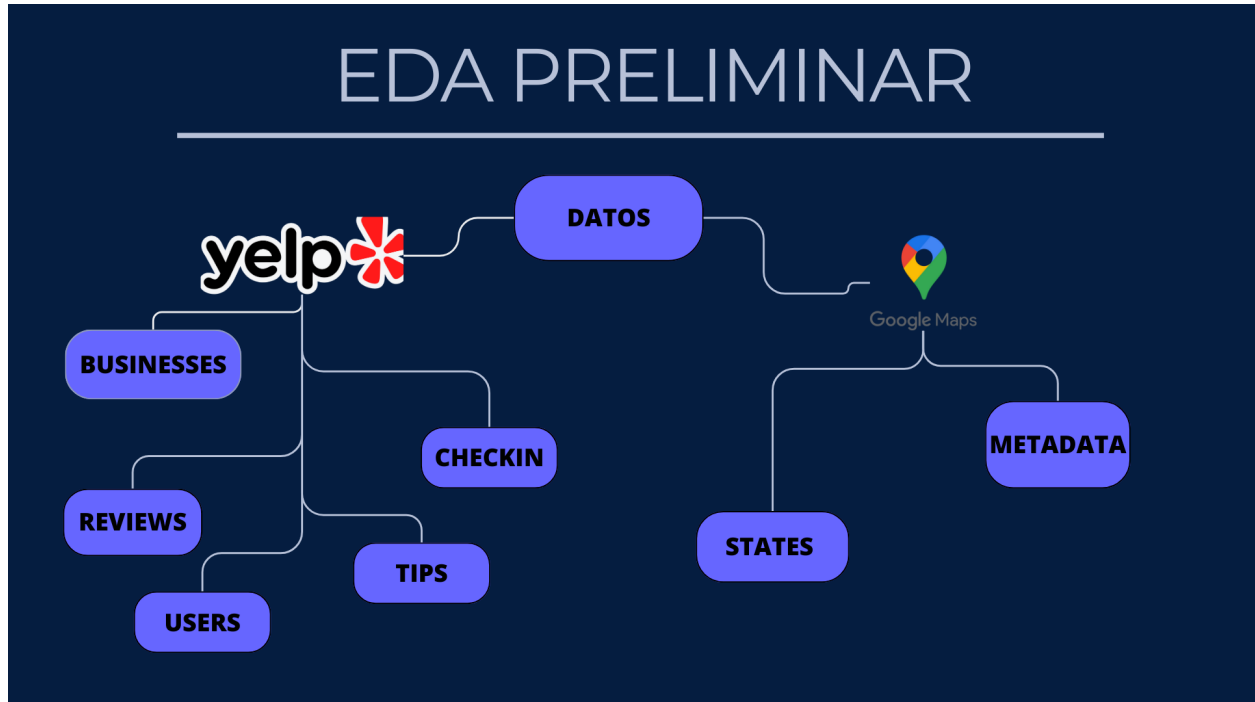
Contamos con la información de dos grandes fuentes de datos proveniente de las plataformas Yelp y Google Maps.

En el caso de Yelp la información provista se encuentra dividida en 5 datasets:

- Business.pkl: con información sobre los negocios de Yelp
- Checkin.json: con registros en el negocio
- Review.json: reseñas completas con el usuario y el negocio al que corresponden
- Tip.json: consejos escritos por el usuario
- user.parquet : con información sobre los usuarios.

En el caso de Google Maps la información provista se encuentra dividida en varios datasets:

- Review: son las reviews, se encuentran separados por estados, en el caso de Florida son 19 archivos .json.
- Metadata: información sobre los negocios, se encuentran en 11 archivos json.



En los archivos `yelp.ipynb` y `google_maps.ipynb` se hace un tratamiento de los datos para realizar el EDA de una manera más simplificada.

Para el archivo `yelp.ipynb`:

Businesses:

- Se eliminan columnas que se encuentran duplicadas y vacías
- Se debe filtrar la información para el estado de Florida. Evidenciamos ciertos problemas en la clasificación, al filtrar por el estado de Florida (columna `state`) observamos ciudades que no pertenecen al estado, por lo que se decide clasificar con código postal identificando los códigos postales correspondientes al estado de Florida y corroborando con un mapa de calor a través de la longitud que los datos corresponden efectivamente al estado de interés. Se eliminan en este paso algunos datos que donde la latitud y la longitud eran outliers y no se encontraban dentro del estado.
- Se eliminan columnas que en primera instancia no utilizaremos .
- Se revisan las categorías en primera instancia de restaurantes para filtrar.
- Se exporta el csv para el EDA preliminar.

Reviews:

- Hacemos un merge entre reviews y businesses haciendo uso del `business_id` para centrarme en los negocios que ya tengo filtrados para Florida con la categoría de restaurantes.
- Eliminamos las columnas que no utilizaré en esta instancia.
- Filtramos las reviews más actuales dentro de los últimos 5 años haciendo uso de la columna `date`.
- Exportamos a csv para el EDA preliminar.

Users:

- Seleccionamos las columnas de interés.
- Filtramos los users_id que se encuentran en los datasets filtrados para el estado de Florida y la categoría de restaurantes.
- Exportamos a csv para el EDA preliminar.

Para el archivo google_maps.ipynb:

Metadata:

- Trabajamos primero un archivo de metadata, luego a través de una función automatizamos el proceso.
- Se busca que se lean todos los json de la carpeta metadata
- Seleccionamos las columnas de interés
- Eliminamos vacíos de columna category y address que no me permitirán filtrar por el estado de Florida y la categoría de restaurantes.
- Se buscan los registros de address donde se menciona la palabra Florida o el estado es FL.
- Se eliminan duplicados
- Se concatenan los 11 json de la carpeta metadata.
- Exportamos a csv para el EDA preliminar.

Reviews:

- Se leen y concatenan los 19 json que se presentan para el estado de Florida.
- Se seleccionan las columnas de interés
- Se hace un merge con metadata para obtener los negocios y asociarlos a las reseñas.
- Se filtran las reseñas en función de las reseñas más actualizadas (de los últimos 5 años)
- Exportamos a csv para el EDA preliminar.

Para el EDA vamos a trabajar la información de Yelp y Google Maps por separado, al menos para esta instancia preliminar.

Identificamos en ambos casos:

- Restaurantes con una media de calificación de 5.0
- Restaurantes con mayor número de restaurantes en el estado
- Restaurantes por categoría
- Restaurantes por ciudad

Identificamos la categoría de fast food restaurante como la segunda categoría más popular y ahora centraremos nuestro análisis específicamente filtrando por restaurantes de fast food. Para cada plataforma presentamos:

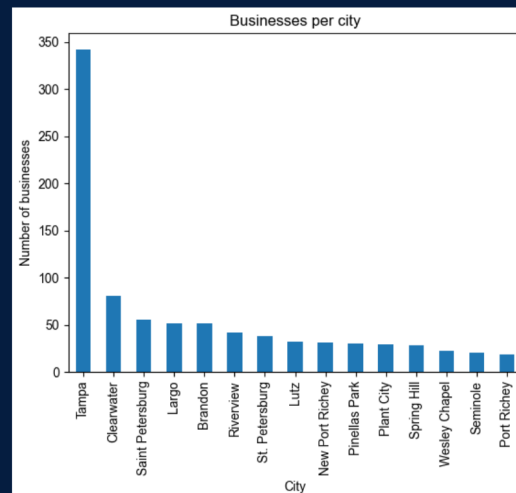
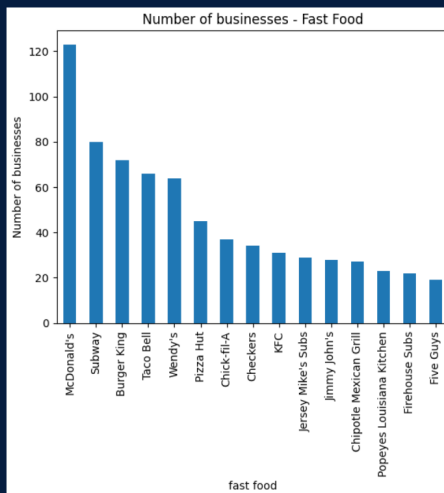
- Restaurantes de fast food con mayor número de restaurantes en el estado

INFORME DE AVANCE: PROYECTO FINAL - SEMANA 1

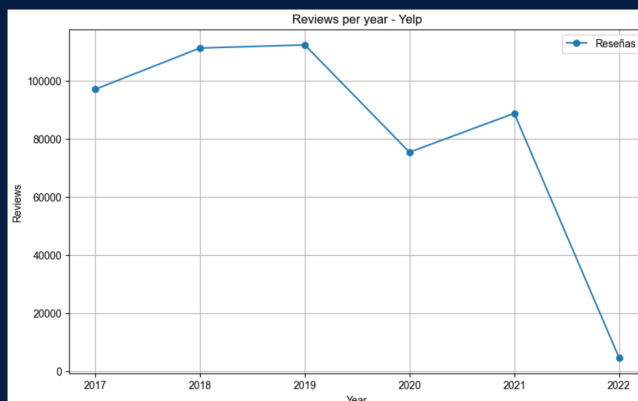
- Restaurantes de fast food por ciudad dentro del estado
- Reviews por año
- Palabras más populares por reseña

Se presentan a continuación, los resultados para Yelp:

EDA - YELP

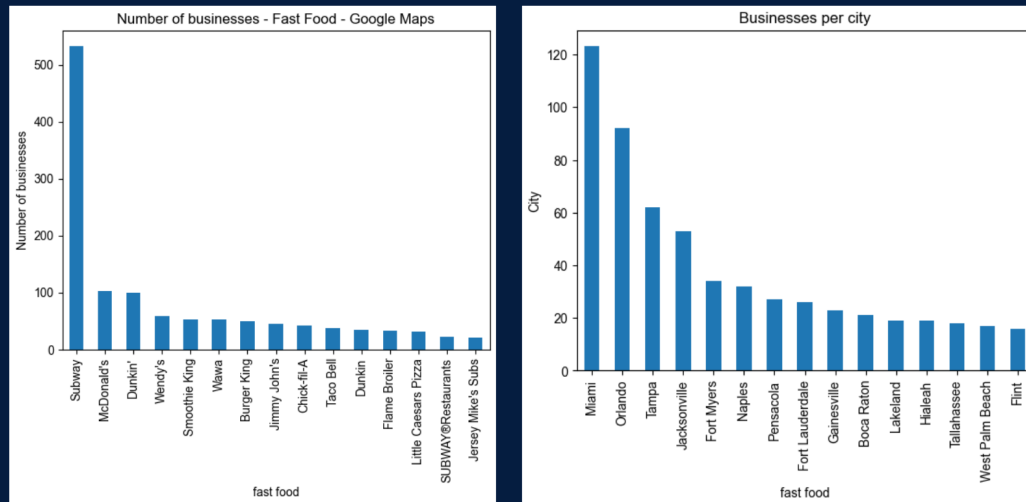


EDA - YELP

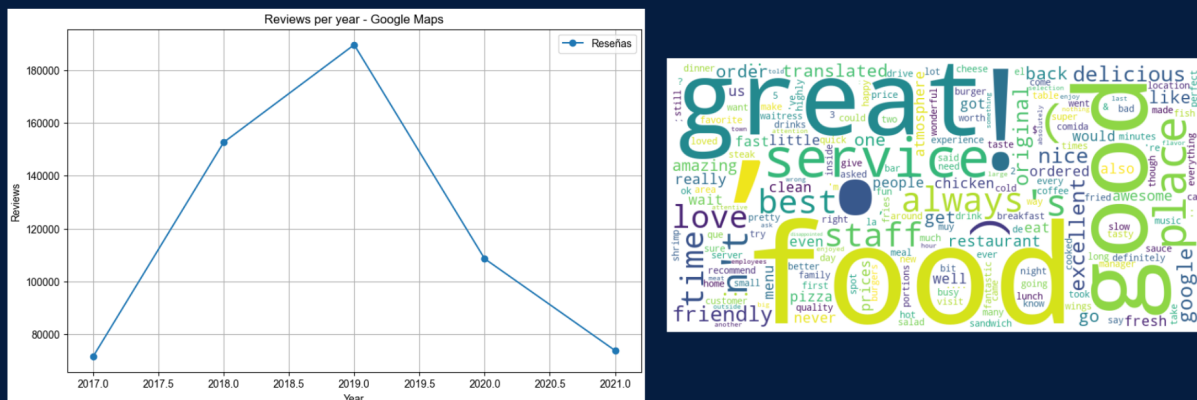


Se presentan a continuación, los resultados para Google Maps:

EDA - GOOGLE MAPS



EDA - GOOGLE MAPS



Cabe mencionar que para los avances de este proyecto, se adicionará información proveniente de la API de Yelp y de la API de Google Maps. En esta primera semana, se consiguieron las keys para ambas APIs y se comenzó el análisis preliminar del formato de la información para incorporarla la próxima semana.

KPIs

- ❖ Crecimiento de Reseñas Positivas: Este KPI se centra en el aumento porcentual en el número de reseñas positivas en comparación con el año anterior
Objetivo: Aumento del 5% en el número de reseñas positivas de Yelp y Google para los negocios de Fast food en comparación al año anterior.

Fórmula: $(\text{Cantidad de reseñas año actual} - \text{Cantidad de reseñas año anterior}) / (\text{reseñas anterior})$

- ❖ Disminución de Reseñas Negativas: Este KPI se centra en la disminución porcentual en el número de reseñas positivas en comparación con el año anterior.
Objetivo: Disminución del número de reseñas negativas del 5% de Yelp y Google para los negocios de Fast food es menor en comparación al año anterior .

Fórmula: $(\text{Cantidad de reseñas año anterior} - \text{Cantidad de reseñas año actual}) / (\text{reseñas anterior})$

- ❖ Índice de satisfacción de usuarios de los principales negocios de fast food (según Yelp y Google) para recomendar una cadena de fast food: Mide la satisfacción global de los usuarios con un negocio, combinando las calificaciones y ponderaciones de las reseñas de los usuarios en las plataformas Yelp y Google.
Objetivo: Aumentar el Índice de Satisfacción del Usuario en un 10% durante el próximo semestre.

Fórmula:

$$\text{Índice de Satisfacción del Usuario} = (\text{Ponderación_Yelp} * \text{Rating_Yelp} + \text{Ponderación_Google} * \text{Rating_Google}) / (\text{Ponderación_Yelp} + \text{Ponderación_Google})$$

- Ponderación_Yelp: Peso en proporción al total (1- total google)
- Rating_Yelp: Puntuación promedio de Yelp para un restaurante o categoría de restaurantes.
- Ponderación_Google: Peso en proporción del total (1- total yelp)
- Rating_Google: Puntuación promedio de Google para un restaurante o categoría de restaurantes.

INFORME DE AVANCE: PROYECTO FINAL - SEMANA 1

- ❖ Índice de Franquicia Estratégica en Florida: Identificación de áreas estratégicas dentro del estado de Florida para la adquisición de una franquicia del rubro fast food.
Objetivo: Identificar 3 áreas estratégicas para franquicias de comida rápida en Florida en intervalos semestrales.
Forma: Localización geográfica de los restaurantes de fast food con stars mayor a 4 para identificar posibles lugares de inversión.
Fórmula: (Número de restaurantes de comida rápida con rating > 4 en ubicaciones estratégicas) / (Número total de restaurantes de comida rápida en Florida).
- ❖ Índice de distribución de visitas por hora: Mide el comportamiento de los visitantes a lo largo del día y puede ayudar a la empresa a tomar decisiones estratégicas.
Objetivo: Lograr una distribución más uniforme de visitas a lo largo del día durante el próximo trimestre.
Fórmula: (Visitas en una hora específica / Total de visitas en el período) x 100

KPI	Descripción	Fórmula	Objetivo
Crecimiento de reseñas positivas	Este KPI se centra en el aumento porcentual en el número de reseñas positivas en comparación con el año anterior	$\left[\frac{\text{Cantidad de Reseñas Positivas en el Año Actual} - \text{Cantidad de Reseñas Positivas en el Año Anterior}}{\text{Cantidad de Reseñas Positivas en el Año Anterior}} \right] \times 100$	Aumento del 5% en el número de reseñas positivas para los negocios de Fast food en comparación al año anterior.
Disminución de reseñas negativas	Este KPI se centra en la disminución porcentual en el número de reseñas positivas en comparación con el año anterior	$\left[\frac{\text{Cantidad de Reseñas negativas en el Año Anterior} - \text{Cantidad de Reseñas negativas en el Año Actual}}{\text{Cantidad de Reseñas Negativas en el Año Anterior}} \right] \times 100$	Disminución del 5% en el número de reseñas negativas de los negocios de Fast food es menor en comparación al año anterior .

Satisfacción de Clientes en Fast Food	Mide la satisfacción global de los usuarios con un negocio, combinando las calificaciones y ponderaciones de las reseñas de los usuarios en las plataformas Yelp y Google.	$\frac{\text{Índice de Satisfacción del Usuario} = (\text{Ponderación_Yelp} * \text{Rating_Yelp} + \text{Ponderación_Google} * \text{Rating_Google})}{(\text{Ponderación_Yelp} + \text{Ponderación_Google})}$	Aumentar el Índice de Satisfacción del Usuario en un 10% durante el próximo semestre
Índice de Franquicia Estratégica en Florida	Mide la proporción de restaurantes de comida rápida con una calificación de estrellas mayor a 4 en ubicaciones estratégicas dentro del estado de Florida.	$\frac{(\text{Número de restaurantes de comida rápida con rating} > 4 \text{ en ubicaciones estratégicas})}{(\text{Número total de restaurantes de comida rápida en Florida})}$	Identifica 3 áreas estratégicas para franquicias de comida rápida en Florida en intervalos semestrales.
Índice de distribución de visitas por hora	Mide el comportamiento de los visitantes a lo largo del día y puede ayudar a la empresa a tomar decisiones estratégicas	$\text{Índice de Distribución de Visitas por Hora} = \frac{(\text{Visitas en una hora específica})}{(\text{Total de visitas en el período})} \times 100$	Lograr una distribución más uniforme de visitas a lo largo del día durante el próximo trimestre