

# PROYECTO FINAL / 2023



## *Datalogia data consulting-Equipo*

Dante Alejandro Gomez – Data Analytics

Hector Fernando Calisaya- Data Engineer

Nicolas Fernando Tablon- Data Engineer

Daniela Florencia Gutierrez- Data Science

# CONTEXTO

Somos una consultora llamada Datalogia donde nuestro enfoque se centra en el análisis de datos recopilados de la plataforma Google Map y Yelp . El objetivo principal de este análisis será mostrar a nuestro cliente patrones y tendencias que necesita saber para tomar decisiones estratégicas a **la hora de cubrir las siguientes necesidades :**

**1. Evaluación de oportunidades:** Las reseñas de usuarios pueden proporcionar información valiosa sobre la popularidad y la calidad de los locales en consideración. Al analizar las opiniones y comentarios de los clientes, el cliente puede obtener una perspectiva más amplia sobre las oportunidades de inversión o expansión en diferentes ubicaciones o tipos de restaurantes.

**2. Identificación de tendencias:** Al revisar las reseñas de múltiples restaurantes, el cliente puede detectar tendencias emergentes en la industria de la comida, como tipos de cocina en crecimiento, platos populares o conceptos innovadores. Esta información puede ser útil para adaptar la estrategia de expansión a las preferencias cambiantes del mercado.

**3. Evaluación de la satisfacción del cliente:** Las reseñas de usuarios pueden brindar una idea clara sobre cómo se sienten los clientes respecto a la experiencia en un local específico. Esto ayuda al cliente a evaluar el nivel de satisfacción de los clientes y a comprender mejor los aspectos que son apreciados o que necesitan mejorar en los locales considerados para expandir.

**4. Mitigación de riesgos:** Al contar con reseñas de usuarios, el cliente puede reducir el riesgo de tomar decisiones basadas únicamente en datos financieros o estadísticas. Las experiencias reales compartidas por los clientes pueden proporcionar una visión más objetiva sobre los aspectos operativos y de servicio que no siempre se reflejan en los informes financieros.

**5. Respuesta a la demanda del mercado:** Al analizar las reseñas de los usuarios, el cliente puede identificar oportunidades para cubrir necesidades insatisfechas en el mercado. Esto podría implicar identificar áreas geográficas con una alta demanda de un determinado tipo de local(fast food, restaurants, take away,etc) o comprender las preferencias específicas de los clientes en una región determinada.

# Objetivos

## Objetivo general

Brindarle al cliente un servicio online donde pueda consultar con datos limpios y depurados que fueron extraídos de la plataforma de Google y Yelp con las reseñas de usuarios . Para dicho servicio se utilizará un sistema de recomendación basado en el análisis de sentimiento para evaluar cómo reseñas positivas o negativas. También para una mejor visualización y exploración se implementará un dashboard interactivo.

## Objetivo específicos

**Garantizar la disponibilidad de datos limpios :** al momento de extraer datos de la plataforma de google y yelp se encuentran en otros formatos como json, parquet o pickle (pkl) por lo cual se debe hacer una limpieza para que el resultado sean datos coherentes y de calidad.

**Realización de una análisis exploratorio de las reseñas :** este análisis ayudará a enfocarnos en patrones , comportamientos o tendencias que los usuarios de google y yelp tienen sobre los locales y los servicios que estos brindan.

**(Adicional) Modelo predictivo de machine learning :** realizar una técnica de análisis de sentimiento de los clientes con lo cual ,permitirá clasificar estas reseñas y predecir preferencias . Con esto se podrá brindar un servicio más óptimo y desarrollar estrategias personalizadas a la hora de expandirse.

**Desarrollar un dashboard interactivo:** para una mejor visualización de los datos recopilados se creará un dashboard( panel de gestión o panel de información) interactivo en el cual se podrá explorar el resultado del análisis proporcionando métricas medibles, gráficos y mapas que le serán útil para la toma de decisiones.

# Alcance

El proyecto se centrará en las franquicias de Mcdonald's ya instaladas, de los 51 estados Estados Unidos será de interés el estado de New York para aquella población que posee una cuenta de google maps y una yelp , en las que se registra sus reseñas y ubicaciones de estas franquicias.

# Formulación de KPI'S

| Nombre de KPI                             | Descripción  | Cálculo   | Periodicidad | Objetivo   |
|---|--|---|--------------|--|
| Tasa de reseñas positivas y negativas     | Porcentaje de reseñas con calificaciones altas o bajas | (total de reseñas positivas o negativas)/total de reseñas *100  | Mensual      | Encontrar negocios con reseñas positivas que indiquen un buen servicio o calidad |
| Distribución de los locales en Nueva York | Distancia entre los locales de la franquicia           | longitud1, latitud1 vs longitud1, latitud 2   | Mensual      | identificar los locales cercanos o lejanos al futuro nuevo local                 |
| Tasa de éxito según la categoría          | Porcentaje de categorías mas solicitadas               | (total de categoría en cuestión/total de categorías)*100  | Mensual      | Identificar aquellas categorías mas solicitadas                                  |
| Promedio de clientes                      | Porcentaje de clientes que dejan su reseña             | (cantidad de clientes periodo actual – cantidad de clientes periodo anterior / cantidad de clientes total del periodo anterior)*100 | Mensual      | Reconocer los locales que tienen un crecimiento positivo                         |

# Solución Propuesta

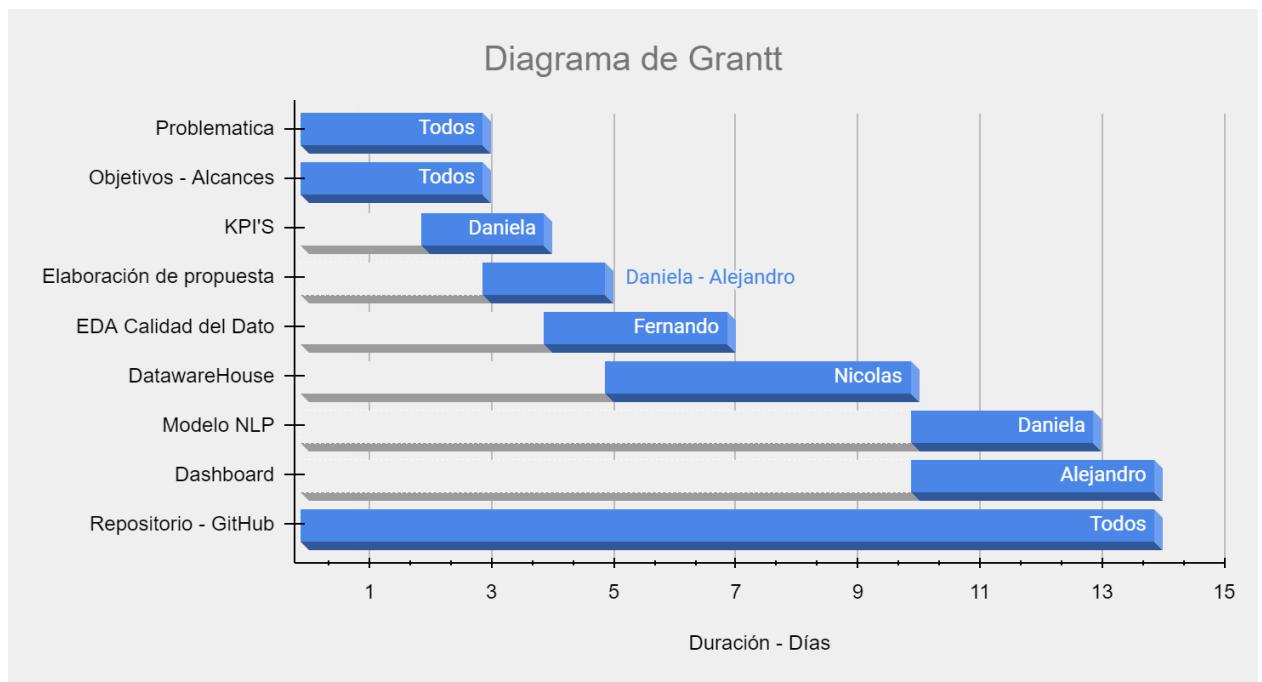
## Repositorio de GitHub



## Stack tecnológico

| Herramientas          | Descripción  |
|-----------------------|--|
| VSC                   | Entorno de desarrollo integrado para programación  |
| Python                | Lenguaje de programación que posee muchas librerías como pandas , numpy , seaborn para exploración de datos. |
| Google Cloud Platform | Plataforma de Google para almacenamiento(Cloud Storage) y procesamiento de datos (Big Query)                 |
| Power Bi              | Herramienta de visualización e interactiva de datos  |

# Planificación del proyecto



# Metodología de trabajo

Dia 1: kick off del proyecto , se reunió con Henry mentor y asignaciones de roles de cada integrante

se reviso los Sprint, la problemática y objetivos. Explicación de las reuniones diarias pactadas en cierto horario

Dia 2: Se puso en marcha la exploración a simple vista de los datos proporcionados con la herramienta VSC en Python

Dia 3: Formulacion de los entregables solicitados por parte del product owner

Dia 4: Preparación de la demo

Dia 5: Presentación de la demo ante Henry mentor

# Equipo de trabajo

| Nombre del integrante – Rol   | Responsabilidad   |
|---|---|
| Hector Fernando Calisaya – Data Engineer<br><br>Nicolas Fernando Tablon – Data Engineer | Realizar una selección del stack, con la herramienta adecuada para la implementación de medidas apropiadas para garantizar la disponibilidad de datos limpios, asegurando que sea la información necesarias y suficientes para la realización del análisis. |
| Daniela Florencia Gutierrez – Data Science  | Desarrollar un modelo de machine Learning preciso y eficaz que permita identificar aquellos patrones para predecir las preferencias individuales.   |
| Dante Alejandro Gomez – Data Analytics  | Desarrollar una interfaz interactiva la cual que brinde al cliente la posibilidad de explorar los resultados del análisis de las reseñas recolectadas. La finalidad es facilitar la visualización con gráficos y filtros para tomar decisiones acertadas.   |