**A logo with blue and grey text

Description automatically generated**

**PROYECTO FINAL: GOOGLE - YELP**

**SPRINT I**

Puesta en marcha del proyecto y Trabajo con Dato

**Grupo 1**

* **CACERES, Juan Patricio**
* **GARCIA, Robertino**
* **LOPEZ, Stiven**
* **MARTÍNEZ, María Gabriela**
* **RENGIFO, Aylin Lizett**

**ENTENDIMIENTO DE LA SITUACIÓN ACTUAL**

En la actualidad digital, las revisiones han adquirido una relevancia crucial en las decisiones de los consumidores, ejerciendo una influencia omnipresente en la percepción de productos y servicios. La construcción de una reputación en línea se ha convertido en un campo de batalla donde las opiniones de los usuarios son armas poderosas y la visibilidad en línea se forja mediante reseñas recomendadas en la red social más importante del mercado.

En calidad de consultores de datos especializados, nos enfocamos en analizar la información recopilada de plataformas de reseñas como lo son Yelp y Google Maps. Nuestro objetivo primordial es descubrir patrones y tendencias en las opiniones de los usuarios, desentrañando así su experiencia, necesidades y expectativas respecto a diversos servicios. Este análisis de datos proporciona una visión detallada que, al aplicar técnicas avanzadas de análisis de sentimientos y aprendizaje automático nos permite ofrecer asesoría y recomendaciones a posibles inversores.

Este proyecto buscará implementar un modelo de recomendación fundamentado en análisis de sentimientos, así como construir un modelo predictivo que evalúe el potencial de crecimiento de negocios basado en patrones extraídos de las opiniones de los usuarios. Además, se pretende diseñar un dashboard interactivo que permita visualizar los resultados obtenidos, facilitando así la toma de decisiones informadas por parte del PO.

**OBJETIVO PRINCIPAL**

Desarrollar una infraestructura completa que centralice la recopilación, depuración y accesibilidad de datos detallados provenientes de las reseñas de los usuarios con el propósito de llevar a cabo un análisis exploratorio.

**Objetivos Específicos**

1. **Garantizar la calidad y coherencia de los datos** mediante la ejecución de un proceso integral que abarque la Extracción, transformación y Carga (ETL, así como el Análisis Exploratorio de Datos (EDA). Este enfoque asegurará la fiabilidad y consistencia de la información obtenida proporcionando una base sólida para el posterior análisis y aplicación en el desarrollo del proyecto.
2. **Desarrollar un modelo de Maching Learning,** modelo de recomendación donde se utilizarán métodos de análisis de sentimientos para categorizar las reseñas y prever las preferencias de los usuarios.
3. **Crear un dashboard interactivo** que facilite la visualización y exploración de los resultados del análisis realizado a las reseñas de los usuarios.

**ALCANCE**

El enfoque del proyecto consiste en suministrar a empresarios detalles sobre restaurantes de interés, priorizando los estados con mayor ***Producto Interno Bruto (PIB)*** en los ***Estados Unidos***

Estos estados son:

* Illinois
* California
* New York
* Texas
* Florida

Esta elección se fundamenta en un análisis exhaustivo llevado a cabo mediante la exploración de diversas páginas web y la aplicación de técnicas de web scraping.

**METODOLOGÍA DE TRABAJO**

El ***modelo Scrum***, seleccionado para nuestra metodología de trabajo, propone que cada iteración (Sprint) inicie con una breve sesión de planificación y concluya con una revisión del trabajo realizado. El objetivo es lograr un producto final que cumpla con los requisitos del cliente.

Este enfoque nos brinda la capacidad de abordar el proyecto en un entorno dinámico y adaptable, permitiendo ajustes según las necesidades emergentes.

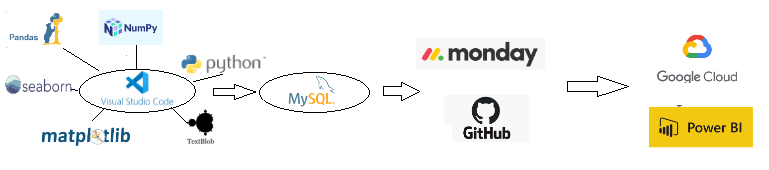
El proyecto se organiza en un total de 3 Sprints, programados semanalmente con las presentaciones al Product Owner y Scrum Master los días viernes. En estas sesiones, se proporciona un informe de avance junto con el producto entregable correspondiente. El Product Owner (cliente) revisa y realiza correcciones, las cuales se incorporan como requisitos para la siguiente etapa del proyecto.

**EQUIPO DE TRABAJO Y ROLES**

Debido a la implementación de una metodología de trabajo por fases, los roles experimentarán cambios a lo largo del proyecto, ajustándose a las necesidades específicas de cada SPRINT. No obstante, se detallan a continuación los roles asignados a cada integrante:

| Cáceres, Juan Patricio | Data Science |
| --- | --- |
| García, Robertino | Data Enginner |
| López, Stiven | Data Science |
| Martínez, María Gabriela | Data Enginner |
| Rengifo, Aylin Lizett | Data Analitycs |

**TECNOLOGÍAS A UTILIZAR:**

Se presentan las tecnologías a utilizar:

Es posible la incorporación de nuevas tecnologías a medida que se avance en el proyecto.

**ANÁLISIS EXPLORATORIO PRELIMINAR**

Llevamos a cabo una primera exploración analítica de los datos proporcionados por Yelp y Google con el propósito de comprender los tipos de datos con los que estamos trabajando, así como evaluar su calidad y posibles transformaciones a las que podrían ser sometidos. La intención no es realizar un análisis exhaustivo en esta etapa, sino obtener una visión inicial.

En cuanto a los archivos de Yelp, se dispone de información sobre usuarios (Data\_User, Data\_Reviews) y sobre los negocios (Business, Checkin, Tips). En los archivos de Google, contamos también con Reviews, archivos por State e información sobre los Business de cada región.

Se determina que estamos lidiando con una cantidad considerable de datos, por lo que será necesario gestionarlos como Big Data. Para abordar esto, se emplearán herramientas de Google Cloud, permitiendo trabajar en un entorno capaz de manejar eficazmente esta voluminosa cantidad de datos.

**KPI**

En base a la propuesta y el análisis preliminar realizado, se proponen los siguientes KPIs:

1. **Nombre:** Tasa de aprobación de clientes

**Descripción**: Mide la proporción de clientes que administraron una calificación positiva en las estrellas o rating (de 4 estrellas para adelante).

**Fórmula:**

**Periodicidad:** Mensual

**Objetivo:** Mantener una tasa de aprobación de clientes del 65% a lo largo de los siguientes 11 meses.

1. **Nombre:** Tasa de insatisfacción de clientes

**Descripción:**  Mide la proporción de clientes que administraron una calificación negativa en las estrellas o rating (por debajo de 2 estrellas).

**Fórmula:**

**Periodicidad:** Mensual

**Objetivo:** Mantener la tasa insatisfacción de los clientes en un 25% o menos a lo largo de los meses, asegurando una experiencia positiva y satisfactoria de manera consistente.

1. **Nombre:** Tasa de crecimiento de reseñas.

**Descripción:** Mide la variación porcentual del número total de reseñas entre un periodo específico de tiempo.

**Fórmula:**

**Periodicidad:** Anual

**Objetivo:** Incrementar la tasa de crecimiento de reseñas en un 4% anual

1. **Nombre:** Cantidad de restaurantes calificación positiva

**Descripción:** Mide la cantidad total de restaurantes que han recibido una calificación positiva, considerando las reseñas que cumplen con el criterio, es decir, que hayan sido calificados con 4 estrellas/rating o más.

**Fórmula:**

**Periodicidad:** Mensual

**Objetivo:** Incrementar la cantidad de restaurantes con calificación positiva en un 2% mensual

**PLANIFICACIÓN DE TAREAS (SPRINT I)**

La estructuración del plan de trabajo se llevará a cabo mediante la creación de un ***Diagrama de Gantt*** en ***Monday***. Al momento de elaborar este informe, nos hallamos en el día viernes de la primera semana, concluyendo la entrega del Informe de Avance I para dar inicio al Sprint II

