



ENTENDIMIENTO DE LA SITUACIÓN ACTUAL

El conglomerado contratante del servicio requiere un estudio de mercados que le permita evaluar la viabilidad de inversión en hoteles, restaurantes y otros negocios afines al turismo y ocio. Para el conglomerado es importante utilizar la información de reseñas de todo tipo de negocios, restaurantes, hoteles, servicios, entre otros ya que éstos muestran la realidad del mercado, desempeño de la industria y servicio prestado.

Para determinar su participación se requiere el analizar datos de reseñas de todo tipo de negocios, restaurantes, hoteles y servicios e identificar las preferencias de los clientes sobre dónde comprar, comer, dormir, reunirse, etc.

OBJETIVOS

Objetivo del proyecto: Realizar un análisis de mercados que le permita evaluar la viabilidad de inversión para el año 2024 en hoteles, restaurantes y otros negocios afines al turismo y ocio al conglomerado contratante del servicio a partir de la información de reseñas de todo tipo de negocios, restaurantes, hoteles, servicios, entre otros.

Objetivo del equipo: Generar un informe y un modelo de recomendación que le permita al conglomerado identificar oportunidades de negocio a través del estudio de tendencias, patrones e información relevante de la industria para finales de la primera semana de diciembre del año 2023.

KPIs (PROPUESTA INICIAL SEGÚN OBJETIVO PLANTEADO POR EL PROYECTO)

1. Crecimiento anual de la cantidad de puntos nuevos creados por estado
2. Porcentaje de crecimiento anual de la cantidad de puntos nuevos creados por tipo de comercio
3. Top 5 o 10 (por definir) de los comercios de mayor representatividad y mejor valorados por estado
4. Top 10 de los aspectos a mejorar incluidos en las reseñas más importantes por comercio y clasificados estado
5. Aumentar la cantidad de reseñas por restaurantes, bares, etc un 10% en forma anual con respecto al año anterior
6. Aumentar un 5 % las reseñas positivas en forma anual con respecto al año anterior
7. Aumentar el promedio de calificaciones de los restaurantes, bares, etc., en un 5% en forma anual con respecto al año anterior
8. Reducir el tiempo promedio de respuesta a las reseñas de 24 horas a 12 horas con respecto al año anterior



ALCANCE

El proyecto se ejecutará teniendo en cuenta la información de las reseñas incluidas en las aplicaciones YELP y GOOGLE MAPS en el mercado de EEUU de hoteles, restaurantes y otros negocios afines al turismo u ocio, además de información adicional sobre cotizaciones en bolsa o tendencias en redes sociales o de medios de comunicación sobre comercios en expansión. El proyecto no incluye la información y/o análisis de tendencias o reseñas fuera del mercado de EEUU y de otras aplicaciones de reseñas diferentes a las de YELP y GOOGLE MAPS.

METODOLOGIA DE TRABAJO

El proyecto se trabajará bajo la metodología Scrum, desarrollando las siguientes actividades:

Planificación del Sprint: El equipo se reúne al inicio de cada Sprint (iteración) para planificar el trabajo que se realizará durante una (1) semana. El equipo estima el esfuerzo necesario para completar cada historia de usuario, utilizando puntos de historia o cualquier otra unidad de medida acordada.

Sprint Backlog: El equipo selecciona las historias de usuario que se compromete a completar durante el Sprint y las agrega al Sprint Backlog. Establece un objetivo claro para el Sprint, que ayuda a mantener el enfoque del equipo durante el período.

Reuniones Diarias (Daily Standup): El equipo se reúne diariamente durante 15 minutos para discutir el progreso, los obstáculos y los próximos pasos. Cada miembro del equipo responde a tres preguntas clave: ¿Qué hice ayer?, ¿Qué voy a hacer hoy? y ¿Hay algún impedimento en mi camino?

Desarrollo y Colaboración: El equipo trabaja en las historias de usuario seleccionadas durante el Sprint. Existe una colaboración cercana entre los miembros del equipo y comunicación constante para abordar posibles problemas y ajustar el enfoque según sea necesario.

Revisión del Sprint: Al final del Sprint, el equipo realiza una revisión para demostrar el trabajo completado al Product Owner y otros stakeholders. Se recopilan comentarios y se pueden realizar ajustes en el Product Backlog en función de las lecciones aprendidas durante el Sprint.

Retrospectiva del Sprint: El equipo reflexiona sobre el Sprint, discute lo que funcionó bien y lo que podría mejorarse. Se identifican acciones específicas para mejorar el proceso y la productividad en futuros Sprints.



DISEÑO DE LA SOLUCION

Arquitectura de datos: Se define cómo se estructurarán, procesarán y almacenarán los datos. Esto puede abarcar desde la selección de las bases de datos a utilizar hasta la planificación de la forma en que los datos fluyen a través del sistema.

Proceso de ETL: Este proceso incluyen la Extracción, Transformación y Carga (ETL). Esto incluye cómo se extraerán los datos de las fuentes, cómo se limpiarán y transformarán, y cómo se cargarán en el repositorio de datos.

Modelado de datos: Se prevé la construcción de modelos y análisis complejos, que incluye la estructura o el diseño inicial de estos modelos. Esto abarca las descripciones detalladas de los algoritmos a utilizar.

Diseño de la interfaz o herramientas de visualización: El proyecto incluye la presentación de los datos de manera visual y la construcción de interfaces.

Plan de pruebas y validación: Se establece cómo se probará la solución propuesta para garantizar su precisión, confiabilidad y validez.

DEFINICIÓN STACK TECNOLÓGICO

Herramientas de extracción de datos: Se utilizara Python para la recolección de datos desde diversas fuentes como bases de datos, APIs, archivos CSV, entre otros.

Plataformas de almacenamiento: Se utilizarán bases de datos o sistemas de almacenamiento de datos como MySQL, PostgreSQL, MongoDB, entre otros, donde se guardarán y gestionarán los datos recolectados.

Herramientas de procesamiento y limpieza: Se utilizarán frameworks o librerías para la transformación y limpieza de datos, como Pandas en Python o herramientas específicas de ETL (Extract, Transform, Load).

Herramientas de visualización y análisis: Plataformas o librerías para visualizar y analizar los datos, como Tableau, Power BI, Matplotlib, Seaborn, etc.

Herramientas de machine learning: Librerías o frameworks de machine learning como TensorFlow, Scikit-learn, PyTorch, entre otros, para implementar modelos predictivos o de análisis de datos.



EQUIPO DE TRABAJO - ROLES Y RESPONSABILIDADES

GRUPO	NOMBRE	ROL PRINCIPAL	ROL SECUNDARIO
1	Joaquin	Ingeniero datos	Analista datos
	Marcelo Alejandro Garcia	Ingeniero datos	Analista datos
2	Hugo Delgado	Machine learning	Ingeniero datos
	Jaime Alexander Jimenez	Machine learning	Analista datos
3	Gabriel Giuffrida	Analista datos	Ingeniero datos
	Hoover Zavala	Analista datos	Ingeniero datos

Analista de Datos:

Recolección y limpieza de datos: Obtener datos de diversas fuentes y asegurar su calidad y coherencia.

Análisis exploratorio de datos (EDA): Explorar los datos para identificar patrones, tendencias y relaciones significativas.

Generación de informes y visualizaciones: Crear informes visuales y narrativos para presentar hallazgos y tendencias.

Modelado predictivo básico: Desarrollar modelos predictivos simples para apoyar la toma de decisiones.

Colaboración interdisciplinaria: Trabajar con otros equipos (como ingenieros y científicos de datos) para comprender los requerimientos del negocio y apoyar la toma de decisiones basada en datos.

Ingeniero de Datos:

Diseño y mantenimiento de infraestructura: Crear y mantener la arquitectura y los sistemas para la gestión y procesamiento de datos.

ETL (Extract, Transform, Load): Extracción, transformación y carga de datos desde diferentes fuentes hacia los sistemas de almacenamiento.

Garantía de la calidad de datos: Asegurar la integridad, calidad y seguridad de los datos a lo largo del ciclo de vida.

Optimización de rendimiento: Mejorar la eficiencia y el rendimiento de los sistemas de datos.

Integración de sistemas: Asegurar la interoperabilidad entre diferentes sistemas y fuentes de datos.

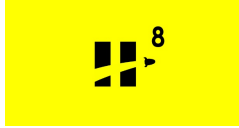
Ingeniero de Machine Learning:

Desarrollo y entrenamiento de modelos: Diseñar, implementar y entrenar algoritmos de machine learning para resolver problemas específicos.

Optimización de modelos: Ajustar y mejorar continuamente los modelos existentes para mejorar la precisión y la eficacia.

Despliegue de modelos: Integrar modelos en sistemas y aplicaciones en producción.

Investigación y desarrollo: Mantenerse al día con los avances en el campo y explorar nuevas técnicas o algoritmos.



GRUPO 08 PROYECTO FINAL
HENRY

Colaboración interdisciplinaria: Trabajar con equipos de negocio y de datos para comprender los objetivos del proyecto y traducirlos en soluciones de machine learning.