**Stack elegido y fundamentación**

**Fase 1: Recolección de Datos**

**Fuente de Datos:** El uso de APIs como Yelp y Google Maps es adecuado para obtener datos geolocalizados y de reseñas. Estas fuentes son ricas en datos tanto estructurados como no estructurados, lo que permite hacer un análisis completo de las reseñas y las características de los lugares. **Almacenamiento Inicial**: Optar por almacenamiento en la nube como AWS S3 o Google Cloud Storage es apropiado, ya que ofrecen escalabilidad y flexibilidad para manejar grandes volúmenes de datos. Las bases de datos MySQL se utilizan para una estructura de datos más organizada y accesible para consultas.

***Fase 2: Procesamiento y Limpieza***

**Herramientas:** El uso de Pandas es esencial para la manipulación y limpieza de datos. La eliminación de duplicados y la transformación de texto son pasos cruciales para asegurar la calidad de los datos antes del análisis.

**ETL**: El proceso ETL es fundamental para garantizar que los datos sean extraídos, transformados y cargados de manera eficiente. Python es ideal para este tipo de tareas, y el uso de bibliotecas como requests, scrapy y API calls permite integrar diversas fuentes de datos de manera automatizada.

***Fase 3: Análisis y Modelado***

**Análisis Exploratorio de Datos (EDA):** La visualización mediante herramientas como Matplotlib, Seaborn y Power BI facilita la comprensión de los datos a través de gráficos interactivos y KPIs. Las nubes de palabras (word clouds) permiten explorar rápidamente los términos más comunes en las reseñas. Análisis de Sentimientos: El análisis de sentimientos es fundamental para obtener insights sobre la percepción de los usuarios. Herramientas como TextBlob, VADER y spaCy ayudan a detectar emociones y opiniones.

**Machine Learning**: Los modelos predictivos con Scikit-learn permitirán identificar tendencias y predecir puntuaciones futuras, lo cual es valioso para entender la evolución del negocio. Los sistemas de recomendación, tanto colaborativos como basados en contenido, aportan valor al proporcionar recomendaciones personalizadas a los usuarios.

***Fase 4: Almacenamiento y API***

**Bases de Datos**: El uso de MySQL para consultas rápidas es adecuado, **Desarrollo de API**: FastAPI es una excelente opción para exponer los modelos como endpoints REST debido a su velocidad y facilidad de integración. El despliegue en plataformas como Render o AWS Lambda asegura que los modelos estén accesibles de manera eficiente.

**Fase 5: Visualización y Producto Final**

**Dashboards y Reportes**: Power BI es una herramienta potente para crear dashboards interactivos que permiten a los usuarios finales explorar los datos de manera visual. Matplotlib y Streamlit son útiles para crear prototipos rápidos de visualizaciones en la web.

