

Proyecto Final

Sprint 2

2023



8 JUNIO

Analytix - Data consulting

EQUIPO SCRUM

- Matías Cejas
- David Soler
- Juan Arieta
- Ana María García
- Luis Bouchez

Contenido

- Producto..... 3
 - Nombre del producto: ReviewTrends..... 3
 - Descripción:..... 3
 - Componentes del producto 3
 - 1- Informe de Análisis de satisfacción del cliente 3
 - 2- Modelo de predicción y recomendación 3
 - 3- Tablero de Métricas Clave..... 3
 - Bono Ofrecido (Asesoría estratégica) 4
 - Asesoría para inversionistas..... 4
 - Asesoría para empresarios..... 4
 - Soporte técnico prioritario..... 4
- Desarrollo del producto 4
 - 1- Información utilizada 4
 - 2- Modelo de data escogido : Data Warehouse..... 4
 - 3- Modelo Entidad-Relación..... 5
- PIPELINE DEL PROYECTO DE DATOS..... 1
 - 1. Extracción y Transformación de Datos 1
 - 2. Almacenamiento de Datos..... 2
 - 3. Automatización de tareas 2
 - 4. Análisis de Datos en la Nube..... 2
 - 5. Modelos de Machine Learning..... 2
 - 6. Visualización y Reportes..... 2

Producto

Nombre del producto: ReviewTrends

Descripción:

ReviewTrends es una solución integral que combina análisis avanzado, machine learning y estrategias de gestión de la reputación para garantizar la disponibilidad de datos limpios y estructurados, provenientes de plataformas de reseñas como Yelp y Google Maps. Nuestro objetivo es brindar información relevante y confiable para respaldar la toma de decisiones estratégicas tanto en la gestión de la reputación como en la inversión en negocios rentables.

Componentes del producto

1- Informe de Análisis de satisfacción del cliente

- Análisis de datos de satisfacción del cliente basado en las calificaciones promedio, el número de reseñas y el texto de las reseñas.
- Generación de un informe detallado que resume los hallazgos y proporciona recomendaciones específicas para mejorar la experiencia del cliente.

2- Modelo de predicción y recomendación

- Desarrollo de un modelo de recomendación en la plataforma para ofrecer recomendaciones personalizadas a los clientes y mejorar la satisfacción del cliente.
- Desarrollo de un modelo predictivo utilizando técnicas de machine learning para predecir el crecimiento de los negocios, que brinde una visión clara del éxito futuro de los negocios.

3- Tablero de Métricas Clave

- Desarrollo de un tablero interactivo que muestre las métricas clave de manera visualmente atractiva y fácil de entender, permitiendo a los usuarios monitorear y analizar de manera efectiva el desempeño de los negocio en función de datos relevantes.

Con este tablero, los usuarios pueden optimizar su estrategia, mejorar la experiencia del cliente y lograr un crecimiento sólido y sostenible para su negocio.

Bono Ofrecido (Asesoría estratégica)

- Asesoría para inversionistas

Para aquellos inversionistas que buscan identificar y seleccionar negocios rentables, el informe les brinda una visión detallada sobre la situación del negocio principal en comparación con sus competidores. Les ayuda a evaluar el potencial de crecimiento y rentabilidad de cada negocio, considerando métricas clave como calificaciones promedio, número de reseñas y categorías de negocio.

- Asesoría para empresarios

Para los empresarios que desean mejorar la experiencia de usuario en sus negocios, el informe de competencia les brinda una evaluación completa sobre cómo se posicionan en relación con sus competidores. Destaca las fortalezas y debilidades en términos de calificaciones, reseñas y categorías de negocio. A partir de estas recomendaciones estratégicas, los empresarios pueden identificar áreas de mejora y enfoque para elevar la experiencia de usuario en sus negocios.

- Soporte técnico prioritario

Acceso a un canal de soporte técnico exclusivo y priorizado por 6 meses.

Desarrollo del producto

1- Información utilizada

Para el desarrollo del proyecto se utilizó la siguiente información

- CARPETA metadata_sitios – GOOGLE MAPS
- CARPETA reviews_estados – GOOGLE MAPS
- ARCHIVO business.pkl – YELP
- ARCHIVO review.json – YELP
- ARCHIVO user.parquet – YELP
- ARCHIVO checkin.json – YELP
- ARCHIVO tip.json – YELP

En el diccionario adjunto a este informe puede revisar a detalle cada archivo.

2- Modelo de data escogido : Data Warehouse

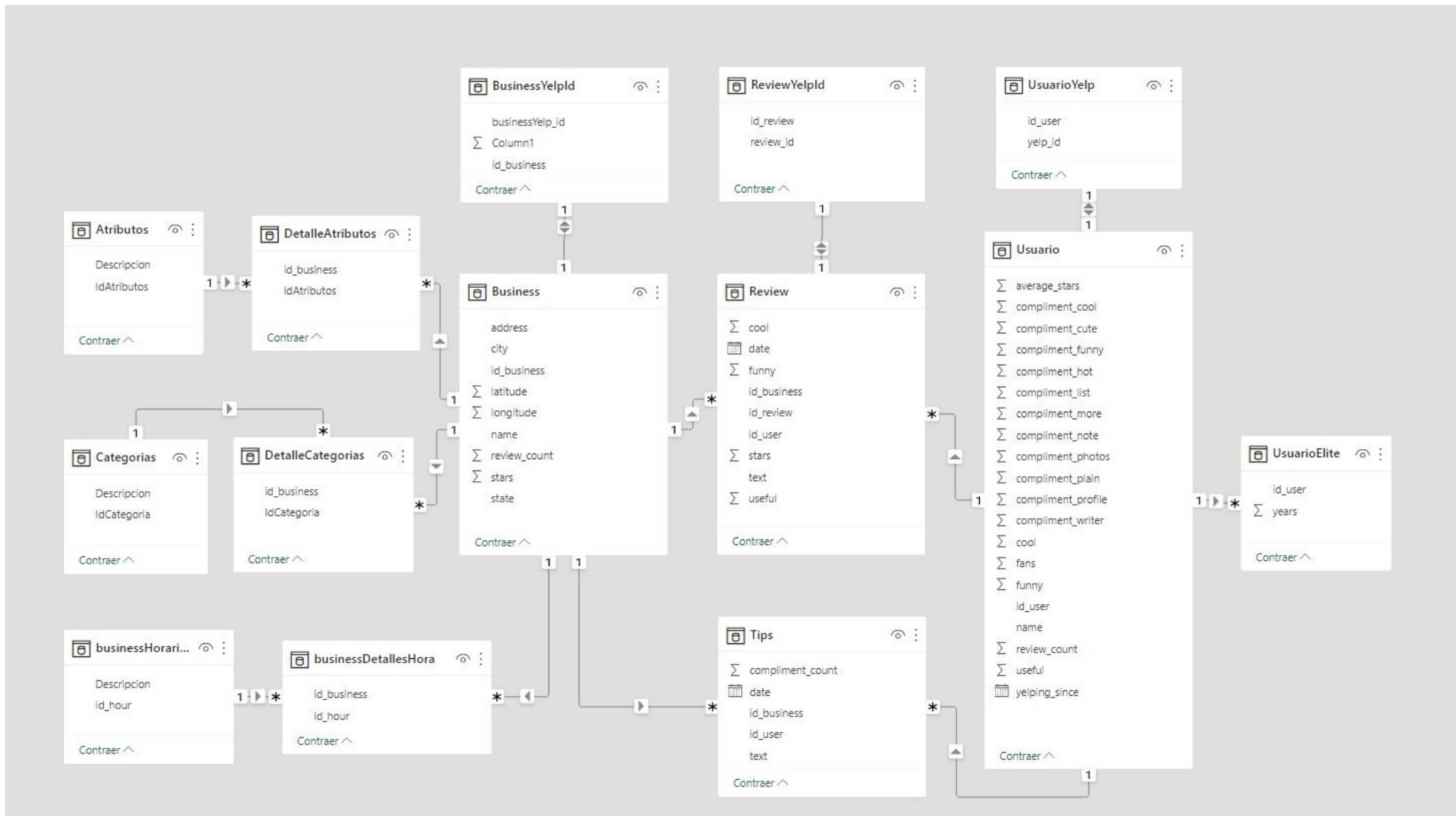
En este informe, exploraremos el uso de un modelo de data warehouse para gestionar los datos de Google Maps y Yelp. Analizaremos el proceso de extracción,

transformación y carga (ETL) de los datos, así como la integración en un esquema predefinido basado en tablas relacionales. Esto permitirá identificar relaciones y patrones entre los datos, brindando una visión más completa y valiosa. Al utilizar este modelo, se facilitará la toma de decisiones informadas y se obtendrá una ventaja competitiva. La data warehouse nos permite almacenar y analizar eficientemente grandes volúmenes de datos estructurados y semiestructurados. Con él, podemos garantizar la consistencia y calidad de los datos, así como la seguridad y control de acceso. Este enfoque brinda la flexibilidad de realizar análisis exploratorios y generar informes predefinidos. Además, podemos optimizar el rendimiento de las consultas y análisis complejos. En resumen, el uso de un modelo de data warehouse para los datos de Google Maps y Yelp mejora la capacidad de análisis, la toma de decisiones y la generación de informes en el entorno empresarial actual.

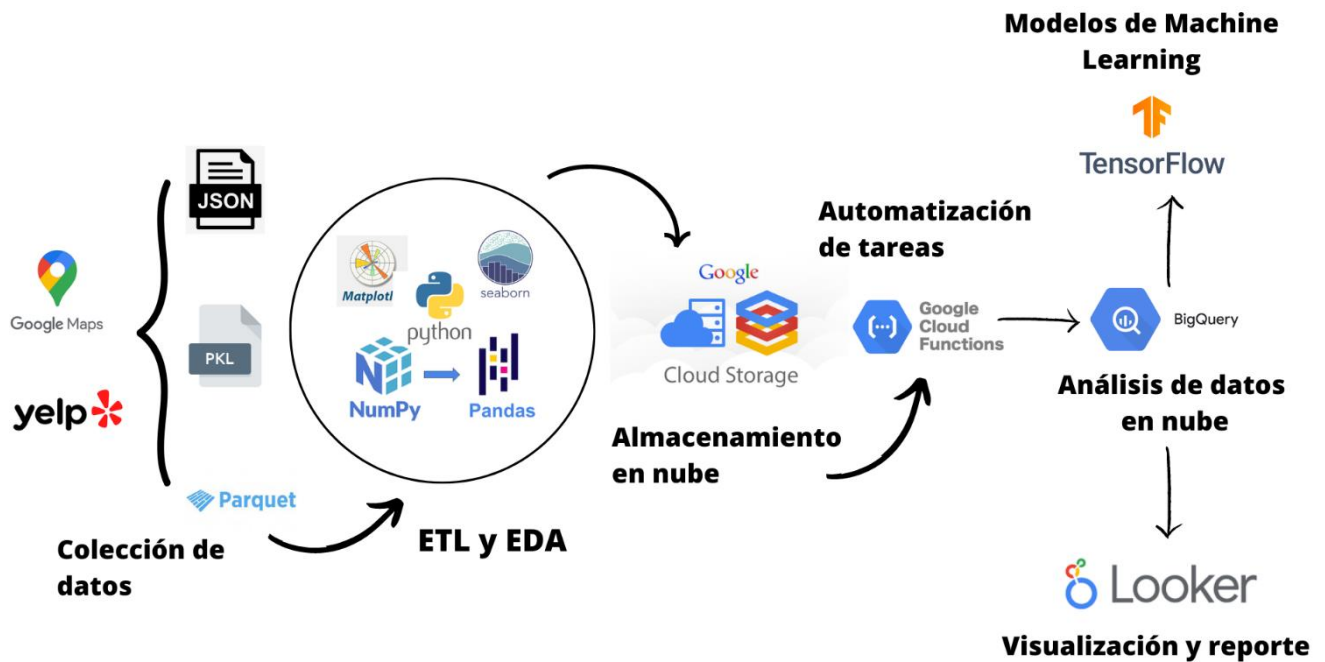
3- Modelo Entidad-Relación

El modelo ER representa la estructura y las interacciones de los datos de Yelp y Google Maps. Incluye entidades principales como "business", "User" y "Review" con sus atributos relevantes. Este modelo es esencial para la gestión de la reputación y la toma de decisiones estratégicas en la inversión en negocios rentables. Proporciona una visión clara de la información clave necesaria para respaldar estas decisiones. Al comprender la estructura de datos, se pueden tomar medidas efectivas para mejorar la reputación y maximizar la rentabilidad.

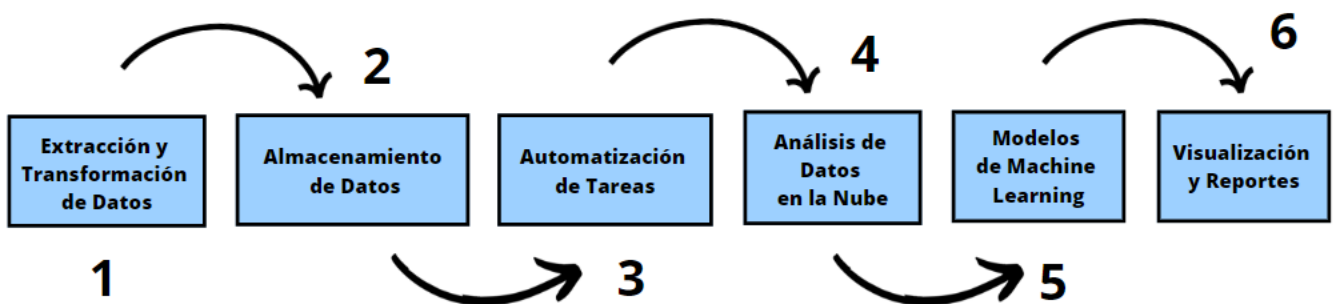
.



PIPELINE DEL PROYECTO DE DATOS



WORK FLOW



Basándonos en la arquitectura del proyecto y el esquema del work flow:

1. Extracción y Transformación de Datos

Se utilizó Python con las bibliotecas Pandas, NumPy, Matplotlib y Seaborn para realizar el proceso de ETL (Extract, Transform, Load) y realizar el análisis exploratorio de datos (EDA) de los datos de las plataformas de reseñas Yelp y Google Maps.

2. Almacenamiento de Datos

Se utilizará Google Cloud Storage como servicio de almacenamiento para almacenar los datos limpios y estructurados.

3. Automatización de tareas

Con Google Cloud Function se pretende automatizar tareas como la programación de extracción de datos, limpieza y transformación periódica de datos, y actualización de modelos de machine learning.

4. Análisis de Datos en la Nube

Con BigQuery como almacén de datos analíticos en la nube se ejecutarán consultas y se realizarán análisis avanzados en tiempo real de los datos almacenados en Cloud Storage, aprovechando la velocidad de procesamiento y la potencia de esta tecnología.

5. Modelos de Machine Learning

TensorFlow se usará como biblioteca principal para desarrollar y entrenar modelos de machine learning.

Usando los datos almacenados en BigQuery para alimentar y entrenar los modelos de TensorFlow.

6. Visualización y Reportes

Se usará Google Looker como herramienta de visualización y reporte para crear paneles interactivos y personalizados, conectándolo desde BigQuery.

Esta arquitectura propuesta combina las tecnologías mencionadas para permitir la extracción, transformación, análisis, modelado y visualización de datos provenientes de plataformas de reseñas, brindando una solución integral para el análisis y gestión de la reputación.