

Informe del Proyecto Final Henry

Introducción

El Proyecto Final de Henry consiste en un análisis exhaustivo de distintas fuentes de datos que busca proporcionar insights valiosos mediante procesos de ETL (Extracción, Transformación y Carga), el uso de técnicas avanzadas de análisis de datos con dashboards analíticos, y sistemas de recomendación y machine learning.

En nuestro proyecto en cuestión hemos simulado ser una consultora de gastronomía de lujo llamada “Helumaro” que asesora a un cliente para sugerirle el lugar y el tipo de restaurante con más posibilidades de tener éxito en el segmento de gastronomía de lujo. El proyecto se centra en varios aspectos del análisis de datos de mercado y utiliza diferentes herramientas y tecnologías para lograr sus objetivos.

El proyecto se divide en tres grandes partes o sprints: en primer lugar un análisis preliminar de los datos principales (bases de datos de Google y Yelp) para delimitar los lineamientos del proyecto.

En el segundo sprint nos enfocamos en la disponibilidad de los datos con dos objetivos principales: por un lado un proceso de ETL para la limpieza y transformación de los mismos en base a nuestros objetivos y por el otro la carga y automatización de los mismos en los servicios de la nube.

Objetivos del Proyecto

El objetivo principal del Proyecto Final Henry es analizar datos de diversas fuentes para obtener insights sobre el mercado, específicamente en el contexto de los restaurantes de lujo de los distintos estados de Estados Unidos.

Además, se implementan sistemas de recomendación y análisis de sentimientos para mejorar la experiencia del usuario y para monitorear tres KPIs que servirán como parámetros de éxito y rentabilidad del negocio.

A continuación se detallarán los objetivos de cada etapa.

ETL (Extracción, Transformación y Carga)

Los procesos ETL son esenciales para la limpieza y preparación de los datos antes de su análisis. El proyecto incluye varios notebooks dedicados a diferentes fuentes de datos. En este proyecto trabajamos con dos fuentes de datos principales: una base de datos de Google y una de Yelp que tienen una gran cantidad de información sobre los restaurantes en Estados Unidos que existen en sus plataformas y las reseñas que los usuarios han hecho sobre ellos.

La información estaba dividida en una gran cantidad de archivos, sin embargo, luego del proceso de ETL hemos reducido la misma a únicamente tres tablas que están

interconectadas entre sí y que pueden observarse adecuadamente en el modelo de entidad relación del repositorio.

Cada uno de estos procesos para cada uno de los archivos de la base de datos, está documentado en su respectivo archivo .ipynb donde se puede observar el detalle de lo realizado y la justificación de la eliminación o modificación de información. De esta manera hemos proporcionado una guía sobre la estructura del ETL y los pasos seguidos.

Análisis Exploratorio del Mercado

El análisis exploratorio del mercado se enfoca en determinar cómo se encuentra el mercado de la gastronomía de lujo en los distintos estados de Estados Unidos. El trabajo se realizó a través de varios notebooks y gráficos que proporcionan una visión detallada de los datos recopilados distribuyéndolos por estado y por tipo de cocina entre otras variables.

Algunos de los análisis realizados incluyen un web scrapping de distintas páginas con información de restaurantes de estrellas michelin para obtener una idea comparativa de la situación del mercado de gastronomía de lujo.

Análisis Massachusetts

En este segmento podemos encontrar información de por qué consideramos que Massachusetts es uno de los mejores estados para invertir en el segmento de gastronomía de alto lujo.

Massachusetts combina un gran poder adquisitivo, muy por encima de la media de los estados de Estados Unidos junto con una escasa competencia en restaurantes de lujo. Siendo un poco más precisos hemos observado que en dicho estado no existe ningún restaurante con estrella Michelin. Este dato, si bien no es determinante en absoluto, sí nos otorga un parámetro del estado de situación del mercado.

Dashboard

El dashboard es una herramienta interactiva que permite a los usuarios explorar los datos de manera más dinámica. En él se podrán observar, de manera detallada, toda la información al respecto sobre el análisis hecho. Los distintos gráficos y filtros permiten al usuario una exploración muy amplia de todas las variables. Asimismo podrá observar la evolución de los KPIs a lo largo del tiempo.

La documentación en el README de esta sección detalla cómo utilizar el dashboard y las funcionalidades disponibles.

Sistema de Recomendación y Análisis de Sentimientos

Esta sección del proyecto tiene dos grandes pilares: por un lado el análisis de sentimiento y por el otro el modelo de Machine Learning.

En el análisis de sentimiento la tarea que realizamos fue procesar el lenguaje natural de las reseñas para determinar el sentimiento de la misma. De esta manera normalizamos una

variable compleja como es una opinión escrita y la transformamos en un valor numérico que luego fue utilizada en el modelo de Machine Learning.

El modelo de machine learning se centra en determinar la probabilidad de éxito de un restaurante en base a la elección de las distintas variables disponibles. También analiza la incidencia de las reseñas y su relación con las características del comercio para delimitar la estrategia de negocio más eficiente y rentable.

Conclusiones

El Proyecto Final Henry demuestra cómo una combinación de análisis exploratorio, procesos ETL eficientes y técnicas avanzadas de machine learning pueden proporcionar insights valiosos y mejorar la experiencia de un grupo inversor en la industria de restaurantes de lujo. La documentación y los notebooks proporcionados permiten una comprensión detallada de los métodos y resultados obtenidos.

Hemos observado que la gastronomía de lujo en Estados Unidos ofrece grandes oportunidades de inversión en el estado de Massachusetts por los distintos factores que hemos comentado en el informe.

Este informe resume los aspectos más importantes del proyecto, pero cada sección contiene una documentación más detallada que puede ser consultada para una comprensión más profunda de cada análisis y proceso.