

PROYECTO FINAL

Demo 3

Leonardo Cortés Marcelo Atencio Federico López Andrés Ruiz

Febrero 2024

Resumen

El siguiente documento presenta el informe de la tercera y última etapa del proyecto final a presentar en el Boot Camp de Data Science de ‘Soy Henry’. Esta etapa abarca la escogencia y desarrollo del modelo Machine Learning, el desarrollo final de un análisis gráfico en un dashboard, así como de la implementación de las KPIs planteadas y finalmente, la implementación tanto del modelo y del dashboard en un solo producto entregable al cliente final.

Índice

1. Ingenieria de características	1
1.1. Sentiment Analysis	2
1.2. Distancia Mínima	2
2. Modelo de Recomendación (ML)	2
2.1. Función <code>find_businesses</code>	2
2.2. Función <code>get_recommendations</code>	3
2.3. Función <code>get_business_info</code>	3
3. Modelo Entidad Relación	4
4. Dashboard Y KPIs	6
4.1. Análisis Global	6
4.2. Indicadores Clave de Desempeño (KPIs)	6
5. Producto Final	8
6. Conclusión	9

1. Ingenieria de características

En esta instancia procedemos a desarrollar un modelo de recomendación de negocios que, al ingresar una categoría y/o estado de negocio, podemos sugerir al inversionista el

negocio mejor calificado con las palabras más relevantes entre los usuarios, sumado a otras 5 recomendaciones de negocios similares.

1.1. Sentiment Analysis

Para llevar a cabo el análisis de sentimiento, utilizamos la función “`sentiment_analysis`” ubicada en nuestros recursos. Esta función evalúa cada uno de los comentarios de los usuarios y otorga una calificación de la siguiente manera:

- Valor 0 para comentarios negativos.
- Valor 1 para comentarios neutros.
- Valor 2 para comentarios positivos.

Es de nuestro especial interés, y especialmente para el inversionista, reducir la información a aquellos negocios cuya puntuación sea mayor que 4 y con comentarios netamente positivos (2). Esta información acotada se procesa en la siguiente etapa y se almacena en un dataframe denominado “`df.filtrado`”.

1.2. Distancia Mínima

A la información reducida, procedemos a dividirla por estados y almacenarla en dataframes individuales. Luego, ejecutamos la función “`distances`”, que calcula la distancia para cada negocio no turístico respecto al negocio turístico más cercano. La información resultante se almacena en un nuevo dataframe. Finalmente, unificamos toda la información procesada en una sola tabla.

Como la cantidad de registros es excesiva para realizar el modelo de Machine Learning, reducimos los datos a una muestra que permita tener una cantidad mínima de negocios por cada estado, categoría, distancia mínima no nula y cantidad de negocios cercanos menor a 3. Esta muestra se guarda como “`model.parquet`”.

2. Modelo de Recomendación (ML)

En esta etapa, desarrollaremos un modelo de recomendación de negocios. Al ingresar una categoría y/o estado de negocio, sugeriremos al inversionista el negocio mejor calificado junto con las palabras más relevantes entre los usuarios. Además, proporcionaremos otras 5 recomendaciones de negocios similares. Para realizar esta recomendación hemos escogido la similitud del coseno como modelo comparativo entre las características de los negocios.

Definiremos 3 funciones:

2.1. Función `find_businesses`

Esta función toma como argumento un estado y opcional una categoría de negocio y devuelve los tres negocios mejor rankeados y con mejor ubicación que estén dentro de las características tomadas como argumento.

Por ejemplo, si ingresamos:

- Estado: California
- Categoría: Pub

Obtendremos la siguiente información:

Business Name	State	Category
Black Diamond Tavern	California	Pub
Woodhouse Blending & Brewing	California	Pub
MCG Service LLC	California	Pub

si ingresamos:

- Estado: New York
- Categoría: None

Obtendremos la siguiente información:

Business Name	State	Category
Blue Mountain Reservation Trail Lodge	New York	Hostel
Rosie Dunn's Victorian Pub	New York	Pub
Sushi Nonaka	New York	Restaurant

2.2. Función `get_recommendations`

Esta función utiliza la similitud de coseno y toma como argumento el nombre de un negocio (generado desde la función `get_business_info`) y devuelve 5 negocios similares . Considera las variables de estado, categoría y la categoría de negocios no turísticos más cercanos

2.3. Función `get_business_info`

La función toma como argumento el nombre de un negocio y genera un wordcloud basado en los comentarios de los usuarios y la información básica del negocio (Nombre, Dirección, Ciudad, Estado, categoría, Lugar turístico más cercano y la distancia al mismo).

Por ejemplo, si ingresamos:

- Black Diamond Tavern”

Obtendremos la siguiente información:

- Nombre del negocio: *Black Diamond Tavern*
- Dirección: *Black Diamond Tavern, 42172 Moonridge Way, Big Bear Lake, CA 92315*
- Ciudad: *Big Bear Lake*
- Distancia: *0.839082*
- Negocio Turístico más Cercano: *Trails End*

- Categoría Turística: *Parque*
- Ranking: *4.100000*

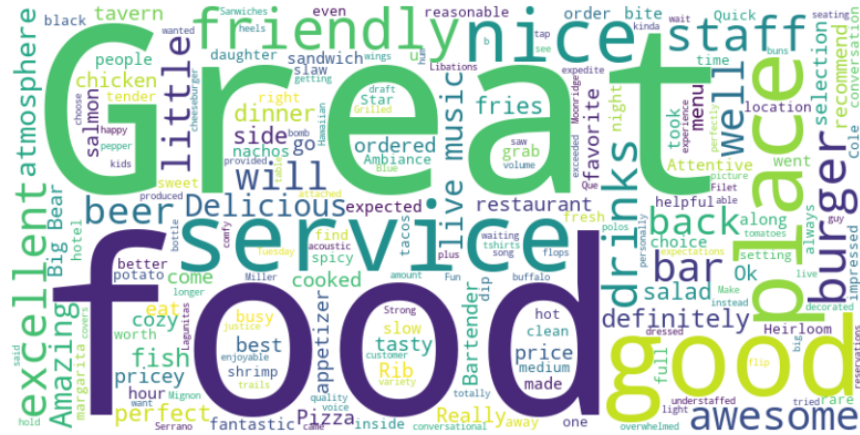


Figura 1: Wordcloud.

Adicional, entrega las recomendaciones dadas por la función `get_recommendations`

Business Name	State	Category
Tiburon Tavern	California	Restaurant
Huntsman Tavern	Nevada	Pub
HOB Tavern	New Jersey	Pub

3. Modelo Entidad Relación

Ya definidas las tablas para el modelo relacional, procedemos a visualizar tanto el diagrama final Entidad - Relación como las Primary Keys (PK) y Foreign Keys (FK) de vinculación.

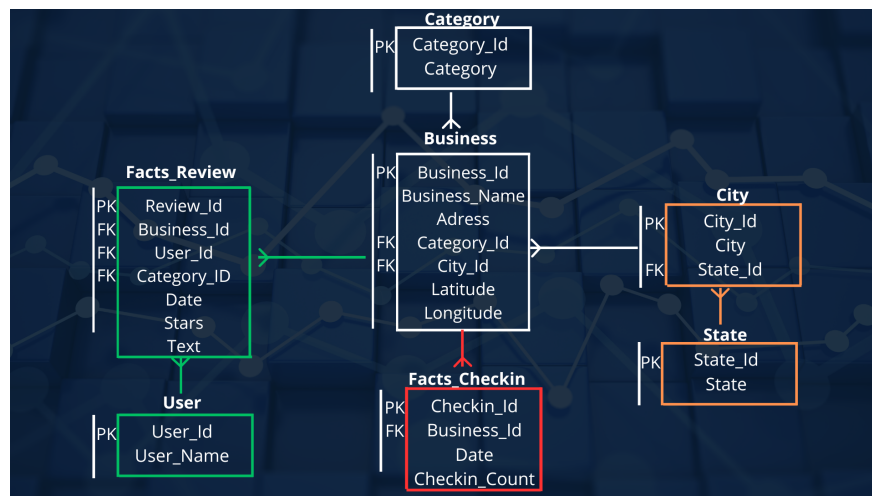


Figura 2: DER.

Donde podemos ver lo detallado:

■ **Tabla Category**

- Category_Id: Entero (Clave Primaria)
- Category: Cadena de caracteres

■ **Tabla City**

- City_Id: Entero (Clave Primaria)
- City: Cadena de caracteres
- State_Id: Entero (Clave Externa)

■ **Tabla State**

- State_Id: Entero (Clave Primaria)
- State: Cadena de caracteres

■ **Tabla User**

- User_Id: Objeto (Clave Primaria)
- User_Name: Objeto

■ **Tabla Business**

- Business_Id: Entero (Clave Primaria)
- Business_Name: Objeto
- Address: Objeto
- Category_Id: Entero (Clave Externa)
- City_Id: Entero (Clave Externa)
- Latitude: Entero
- Longitude: Entero

■ **Tabla Review**

- Review_Id: Entero (Clave Primaria)
- Business_Id: Entero (Clave Externa)
- Date: Fecha
- User_Id: Objeto (Clave Externa)
- Ranking: Entero
- Stars: Entero
- Text: Objeto
- Sentiment_Analysis: Entero

■ Tabla Checkin

- Checkin_Id: Entero (Clave Primaria)
- Date: Fecha
- Business_Id: Entero (Clave Externa)
- Checkin_Count: Entero

4. Dashboard Y KPIs

4.1. Análisis Global

Para la visualización de los datos, solamente dispondremos un tablero en el cual se puede ver una análisis global de la información: localización de los negocios en el mapa, cantidad de negocios por categoría, TOP 10 de ciudades con mayor cantidad de negocios y estados con mayor cantidad de negocios.

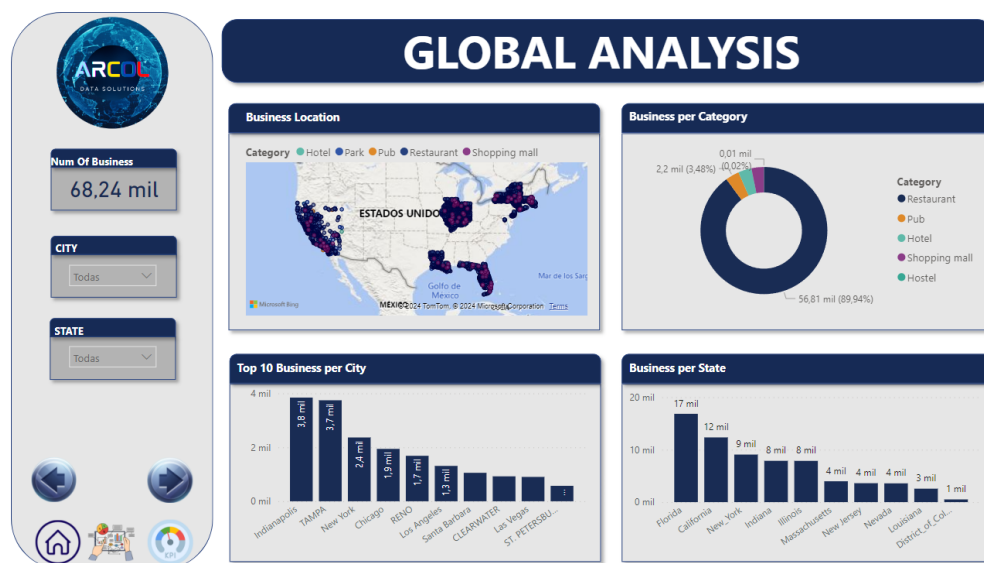


Figura 3: Global Analysis.

4.2. Indicadores Clave de Desempeño (KPIs)

En relación a los KPIs, hemos determinado 4 indicadores representativos para cada negocio. Estos indicadores se calculan en base a un período específico y se establecen objetivos en comparación con el período anterior.

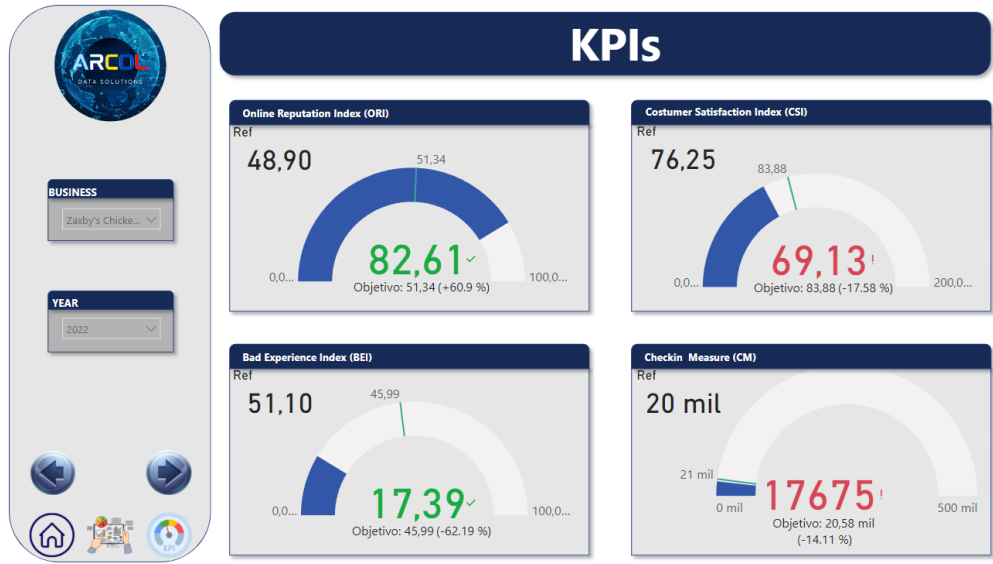


Figura 4: KPIs.

Índice de Reputación Online (ORI)

$$ORI = \left(\frac{\text{Cantidad de reseñas positivas (SA = 2)}}{\text{Total de reseñas}} \right) \times 100$$

Objetivo: Calculado anualmente, incrementar un 5 % la reputación online respecto al período anterior.

Índice de Satisfacción del Cliente (CSI)

$$CSI = \left(\frac{\text{Promedio de calificación de los usuarios}}{\text{Calificación máxima posible}} \right) \times 100$$

Objetivo: Calculado anualmente, incrementar un 10 % la satisfacción de los clientes respecto al período anterior.

Índice de Mala Experiencia por los Usuarios (BEI)

$$BEI = \left(\frac{\text{Cantidad de reseñas neutras y negativas (SA ≤ 1)}}{\text{Total de reseñas}} \right) \times 100$$

Objetivo: Calculado anualmente, disminuir un 10 % la mala experiencia de los usuarios respecto al período anterior.

Registros de Check-In (CM)

CM = Cantidad de Check-In registrados

Objetivo: Calculado anualmente, incrementar un 5 % la cantidad de Check-In registrados en el negocio en cuestión respecto al período anterior.

El inversor puede seleccionar un negocio específico y visualizar los indicadores para los años registrados, lo que le permitirá evaluar el desempeño de dicho negocio a lo largo del tiempo.

5. Producto Final

Finalmente, y luego de hacer distintas pruebas con el modelo de Machine Learning y los datos con los que alimenta el dashboard, se procede a realizar un documento “main.py” que contiene las funciones necesarias para ejecutar tanto el modelo ML realizado como la visualización del dashboard con los respectivos análisis y KPIs, en un solo producto final entregable, realizando el deploy del resultado final de nuestro proyecto a través de una “web app” codificada con la herramienta “streamlit”. Adicional, se crea una página de presentación dentro de la misma app en la que el usuario puede ver información del equipo de ARCOL Data Solutions.



Figura 5: ML Model.



Figura 6: Dashboard.

6. Conclusión

En conclusión, ARCOL Data Solutions presenta un enfoque integral y estratégico para el desarrollo de un modelo de recomendación de inversionistas para el sector turístico en Estados Unidos. Al aprovechar el poder del análisis de datos y las técnicas de aprendizaje automático, nuestro objetivo es brindar a nuestros clientes información valiosa y recomendaciones prácticas para tomar decisiones de inversión informadas. Nuestro equipo de profesionales experimentados está comprometido a brindar resultados de alta calidad y superar las expectativas del cliente. Esperamos tener la oportunidad de colaborar con usted y contribuir a su éxito en el panorama del mercado dinámico y competitivo.