INFORME ANÁLISIS DE MERCADO
PROYECTO GRUPO 08
11 de Diciembre de 2023

Integrantes:

Gabriel Giuffrida
Hoover Zavala
Hugo Delgado
Marcelo Alejandro Garcia
Jaime Alexander Jimenez

INTRODUCCION

Somos una consultora especializada en análisis de datos para proveer servicios a inversores del sector gastronomico con el propósito de potenciar la experiencia del usuario mediante la entrega de recomendaciones personalizadas sobre restaurantes, cafeterías y atracciones en las cercanías del alojamiento. A través de nuestro análisis, buscamos adquirir conocimientos profundos que permitan a nuestros clientes mejorar su posición en el dinámico sector de turismo y ocio en los Estados Unidos

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.

Fundamentar nuestra evaluación en la información derivada de diversas reseñas que abarcan una amplia gama de negocios, desde restaurantes hasta servicios, con el fin de proporcionar una visión estratégica precisa.

Proponer como objetivo fundamental revertir las posibles tendencias negativas actuales en las métricas clave de rendimiento, focalizando sus esfuerzos en el aumento significativo y sostenido de las reseñas, la diversificación de la cobertura de reseñas a nuevos negocios, y la mejora de la calidad percibida a través de un incremento en las calificaciones y reseñas positivas. Además, se busca mantener y fortalecer la tendencia positiva en la reducción de reseñas negativas. Estos objetivos están diseñados para revitalizar la presencia digital y la reputación de los negocios asociados al turismo y ocio, contribuyendo así a una percepción más positiva y atractiva por parte de los usuarios.

Objetivos secundarios:

- ✓ Entregar un informe que proporcione cifras y datos de los comercios que permita identificar las mejores opciones de inversión y que cubran los KPIs sugeridos.

- ✓ Proporcionar un modelo predictor y un modelo de recomendación de machine learning que permita tomar decisiones sobre la inversión en restaurantes basándose en datos históricos de reseñas y puntuaciones..
- ✓ Realizar un análisis de sentimientos mediante técnicas de procesamiento de lenguaje natural (NLP) en las reseñas de Yelp y Google Maps para entender la percepción de los usuarios acerca de los negocios en el sector de turismo y ocio en Estados Unidos.

Entregables:

- ✓ Informe que identifique los estados y el comercio con mayor potencial de crecimiento en los EEUU y un modelo de predicción basado en las reseñas y comentarios de Google maps y Yelp, que incluye un análisis de sentimientos para predecir cuáles serán los rubros de los negocios que más crecerán (o decaerán)
- ✓ Sistema de recomendación de restaurantes para los usuarios de Google maps y Yelp, con el fin de darle al usuario la posibilidad de poder conocer nuevas opciones basados en las reseñas.

Para lograr esto, se emplearon diferentes conjuntos de datos y se implementaron procesos de Extract, Transform, Load (ETL) y Análisis Exploratorio de Datos (EDA). Además, se desarrollaron dos modelos: uno de recomendación basado en reseñas para usuarios y otro de predicción que recomienda tipos de negocios según la industria seleccionada (restaurantes, ocio u hoteles).

PROCESO

Adquisición de Datos:

Se utilizaron datasets con información de reseñas de Google Maps y Yelp, y se optó por Google Cloud como servicio en la nube para gestionar los datos, automatizar las actualizaciones de los mismos la cual es realizada mediante las APIs de Google Maps y Yelp.

Carga Inicial

Los datos crudos se extrajeron de la carpeta de Google Drive proporcionada por Henry, cuyas fuentes son como se mencionó Google Maps y Yelp. Estos datos se presentan en formatos JSON, Parquet y Pickle.

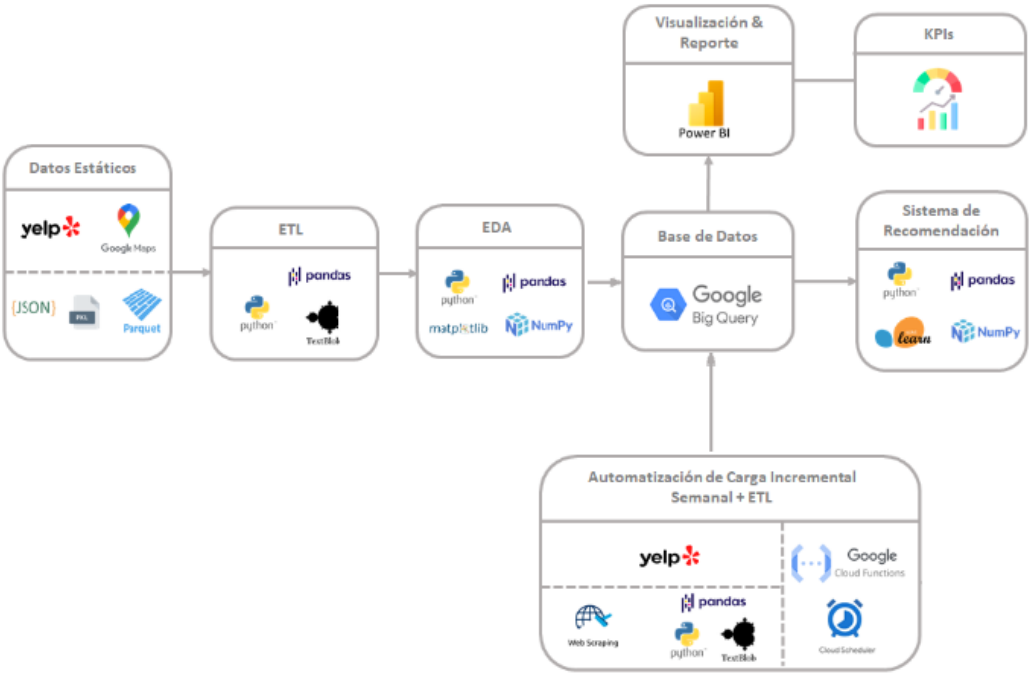
Los datos seleccionados estan relacionados con el sector gastronomico. A partir de estos datos, se llevaron a cabo procesos de transformación (ETL) y exploración preliminar de datos (EDA) con el propósito de crear nuevos conjuntos de datos que estuvieran limpios, transformados y normalizados, que permitieron reconocer la información tratar. Estos datos listos para ser usados se guardaron en datasets de Big Query que es un servicio provisto por en Google Cloud Platform (GCP):

Carga Incremental

Una vez realizada la carga inicial y debido a que los datos cambian constantemente, es fundamental mantener los datasets actualizados. Es aquí es donde interviene la carga incremental.

La carga incremental implica transferir solo los datos nuevos desde las fuentes de origen para lo cual se implementó un proceso ETL automatizado utilizando Google Cloud Functions, con el cual los datos fueron procesados y almacenados en los datasets previamente inicializados en BigQuery.

Stack Tecnológico y ciclo de vida de los datos



IMPLEMENTACION DE LAS CLOUD FUNCTIONS

Cloud Functions es un entorno de ejecución que permite crear y conectar servicios en la nube, sin tener que provisionar ninguna infraestructura ni preocuparse por administrar ningún servidor.

Habiendo creado previamente una cuenta en GCP, un nuevo proyecto y teniendo creados los datasets en Big Query, se deben crear cuatro archivos .py, los cuales seran subidos a GCP como una cloud function.

Estos archivos son:

main.py
Api_descarga.py
bigquery_uploader.py
schema.py

main.py

El script se ejecuta cuando se realiza una solicitud. En particular, está diseñado para obtener datos de negocios de Yelp o Google Maps usando la API, verificar la presencia de datos, cargar los datos en una tabla de BigQuery y proporcionar mensajes de éxito o error según el resultado de la operación.

Api_descarga.py

Este script interactua con la Yelp API y con Google Maps API para obtener información referida a restaurantes en varios estados de USA y organiza esos datos en DataFrame de pandas para su manipulación.

schema.py:

Define el esquema de la tabla que se utilizará en BigQuery para almacenar la información sobre los lugares obtenidos.

bigquery_uploader.py:

Inicializa un cliente de BigQuery.

Crea una tabla en BigQuery si no existe.

Carga un DataFrame de Pandas en una tabla de BigQuery.

Deploy una Cloud Function con Cloud SDK

Para subir elcódigo a una cloud function usando la terminal, se utiliza cloud sdk.

El SDK de Cloud es un conjunto de herramientas que puedes usar para administrar recursos y aplicaciones alojados en Google Cloud.

Hacer el deploy es muy sencillo, solo debe ejecutar el comando gcloud functions deploy function_name con una serie de flags:

Los flags servirán para configurar nuestra cloud function. Los flags usados son:

--project sirve para especificar en que proyecto se hara el deploy de la función.

--trigger-http indica el tipo trigger que ejecutara nuestra función. En este caso, la funcion se ejecuta cuando se abre una url especifica.

--timeout es el tiempo máximo de ejecución de la función (540 segundos es el limite maximo).

--memory es la memoria reservada.

--runtime indica el lenguaje en el cual esta escrito el codigo de la cloud function

--entry-point indica la función que deberá ejecutar la cloud function cuando se dispare el trigger. En nuestro, dentro del archivo main.py, se encuentra la función main, la cual se encarga ejecutar todo el código del proyecto

Cloud Scheduler

Cloud Scheduler es un servicio de Google Cloud Platform (GCP) que permite programar y automatizar la ejecución de tareas en la nube. Ofrece una solución para la planificación y ejecución periódica de trabajos, eliminando la necesidad de administrar infraestructura dedicada para este propósito.Para el caso del proyecto, permite

La ejecución de la cloud function y por lo tanto la carga de los datos con una frecuencia de una vez por semana.

ANÁLISIS DE LA INDUSTRIA

Para un mejor análisis que ayudara a cumplir con el objetivo del proyecto se hizo necesario incorporar datos adicionales significativos, abarcando el crecimiento promedio, la inversión proyectada para los próximos tres años, así como los márgenes y la rentabilidad de cada industria.

Crecimiento Promedio:

Restaurantes (5.16%): Exhibe un crecimiento sólido, sugiriendo una demanda constante en el mercado de restaurantes. Esto podría atribuirse a factores como cambios en las preferencias del consumidor o la introducción de nuevas ofertas gastronómicas.

Ocio (4.80%): Aunque ligeramente inferior al sector de restaurantes, el crecimiento en ocio sigue siendo significativo. Esto podría indicar una tendencia positiva en actividades de entretenimiento y recreación.

Hoteles (3.80%): Aunque el crecimiento es menor en comparación con restaurantes y ocio, sigue siendo considerable. La estabilidad en este sector podría deberse a la constante demanda de servicios de alojamiento.

Inversión para Próximos 3 Años:

Restaurantes (200 millones USD): La inversión planificada en restaurantes refleja una confianza significativa en este sector. Esta asignación de recursos sugiere la percepción de oportunidades sólidas y atractivas para el crecimiento.

Ocio (200 millones USD): Similar a la inversión en restaurantes, esta cifra indica una apuesta significativa en el sector de ocio, respaldando la idea de oportunidades considerables en este ámbito.

Hoteles (200 millones USD): La inversión constante en hoteles destaca la importancia de este sector y la confianza en su capacidad para generar rendimientos a largo plazo.

Margen:

Restaurantes (20.73%): Aunque el margen es moderado, es indicativo de la eficiencia operativa en el sector de restaurantes. Estrategias adicionales podrían ser implementadas para mejorar aún más la rentabilidad.

Ocio (36.72%): Un margen considerablemente alto en el sector de ocio sugiere una mayor rentabilidad por unidad de inversión. Esto podría atribuirse a la capacidad de fijar precios más altos en experiencias de ocio.

Hoteles (42.55%): El margen más alto en hoteles destaca la rentabilidad robusta en este sector. Las estrategias de gestión eficientes y la demanda constante podrían ser factores clave.

Rentabilidad Promedio Último Año:

Restaurantes (12%): Aunque la rentabilidad es sólida, hay espacio para mejoras. Estrategias como la gestión de costos y la optimización de operaciones podrían impulsar aún más los rendimientos.

Ocio (25%): Una rentabilidad destacada en el sector de ocio sugiere que las inversiones realizadas han generado retornos significativos. Esto puede indicar una gestión eficiente de recursos y una estrategia de mercado efectiva.

Hoteles (22%): La rentabilidad sólida en hoteles respalda la inversión continua. Estrategias centradas en la retención de clientes y la mejora de servicios pueden contribuir a un crecimiento sostenible.

Este conjunto de información resaltó que, al considerar el análisis de crecimiento, inversión y flexibilidad, la industria de restaurantes emerge como la de mayor potencial. Por ende, se determinó la necesidad de llevar a cabo un análisis más detallado en este sector.

Análisis Detallado de Estados con Mayor Potencial en EE. UU.

ESTADO	PIB	RPC (USD)	Tamaño Población (mill)	% Turismo
California	3.356,63	76619	39510	10,10
Florida	1.226,93	65481	21538	6,30
Ilinois	938,35	62321	12801	4,00
Pensilvania	839,44	61248	12801	3,90
Nueva jersey	672,09	60348	9288	3,80
Indiana	420,34	58624	6780	3,70
Tennessee	418,29	56619	7040	3,60
Arizona	411,19	55852	7471	3,50

California:

Líder indiscutible en términos de PIB, reflejando una economía robusta y diversificada. Alta renta per cápita, indicativo de un nivel de vida elevado y un mercado consumidor potente. La densidad poblacional elevada sugiere una gran base de consumidores y fuerza laboral. Representación turística significativa, lo cual puede indicar un sector turístico próspero.

Florida:

Considerable PIB, impulsado en gran medida por la industria turística y el mercado inmobiliario. Rentabilidad individual notable, respaldando el poder adquisitivo de la población. Población considerable, proporcionando una base de consumidores y talento laboral sólida. La alta representación turística sugiere un sector turístico importante y atractivo.

Illinois:

PIB considerable, reflejando una economía diversificada. Rentabilidad individual sólida, contribuyendo al poder adquisitivo de la población. Población significativa, lo que indica una base de consumidores y fuerza laboral estable. Aunque no tan alto como otros estados, el turismo sigue siendo un factor relevante en la economía.

Pensilvania:

PIB considerable, sugiriendo una economía estable y diversificada. Rentabilidad individual sólida, respaldando el poder adquisitivo. Población estable, proporcionando una base sólida para el mercado local. Aunque no tan alto como otros estados, sigue siendo una contribución significativa.

Nueva Jersey:

PIB respetable, indicando una economía sólida y diversificada. Rentabilidad individual sólida, contribuyendo al poder adquisitivo de la población. Población considerable para el tamaño del estado, proporcionando una base de consumidores estable. Aunque no es el más alto, sigue siendo un componente relevante en la economía.

Indiana, Tennessee, y Arizona:

Estos estados muestran cifras significativas en sus indicadore y vale la pena observar áreas de fortaleza específicas, como la renta per cápita en Indiana y Arizona, así como el turismo en Tennessee.

Los estados seleccionados muestran diversos aspectos de fortaleza económica, con California y Florida destacando como líderes claros en varios indicadores. El análisis proporciona una base sólida para la identificación de oportunidades de inversión, considerando la población, el PIB, el ingreso per cápita y la representación turística.

ANÁLISIS DE KPIS Y OPORTUNIDADES EN LA INDUSTRIA DE RESTAURANTES

KPI 1: Aumentar las reseñas de usuarios nuevos en un 10% (474,007) año 2021 respecto al año anterior 2020. Resultado Actual: -40.67%

Análisis: La disminución significativa en la cantidad de reseñas de usuarios nuevos indica una falta de atracción hacia la plataforma o una posible disminución en la actividad general de reseñas.

KPI 2: Aumentar la cantidad de restaurantes nuevos reseñados en un 5% (25,135) año 2021 respecto al año anterior 2020. Resultado Actual: -11.82%

Análisis: La disminución en la cantidad de restaurantes nuevos reseñados sugiere una posible falta de participación o interés tanto por parte de usuarios como de nuevos establecimientos.

KPI 3: Aumentar la cantidad de reseñas por restaurante en un 10% anual año 2021 respecto al año anterior 2020. Resultado Actual: -37.64%

Análisis: La fuerte disminución en la cantidad de reseñas por restaurante podría indicar una disminución en la participación activa de los usuarios o posiblemente una falta de interacción efectiva con la plataforma.

KPI 4: Aumentar un 5% las reseñas positivas en forma anual año 2021 respecto al año anterior 2020. Resultado Actual: -35.21%

Análisis: La disminución en las reseñas positivas sugiere un posible descontento o una disminución en la calidad percibida de los servicios proporcionados por los restaurantes.

KPI 5: Aumentar el promedio de calificaciones de los restaurantes en un 5% anual año 2021 respecto al año anterior 2020. Resultado Actual: -36.69%

Análisis: La baja en el promedio de calificaciones puede indicar perdida de interés de generar reseñas, lo cual podría impactar en el mejoramiento de la industria.

KPI 6: Reducir en un 10% la cantidad de restaurantes de los que obtuvo reseñas negativas en forma anual año 2021 respecto al año anterior 2020. Resultado Actual: -31.68%

Análisis: A pesar de la reducción en la cantidad de reseñas negativas, el porcentaje muestra que todavía existe una proporción significativa de restaurantes con comentarios desfavorables.

Oportunidades:

Revisar Estrategias de Adquisición de Usuarios:

Implementar campañas de marketing dirigidas a atraer nuevos usuarios y fomentar la participación activa.

Incentivar Reseñas de Restaurantes Nuevos:

Establecer programas de incentivos para animar a los usuarios a revisar nuevos restaurantes y mejorar la diversidad de reseñas.

Mejorar Interacción con la Plataforma:

Implementar mejoras en la interfaz y la experiencia del usuario para aumentar la participación y la cantidad de reseñas por restaurante.

Enfocarse en la Calidad del Servicio:

Colaborar con restaurantes para mejorar la calidad de sus servicios y fomentar reseñas positivas.

Refinar Estrategias de Gestión de Comentarios:

Desarrollar estrategias para manejar eficientemente las reseñas negativas y reducir la proporción de restaurantes con comentarios desfavorables.

Estas oportunidades pueden ayudar a revertir las tendencias negativas y fortalecer la posición de la industria de restaurantes en la plataforma.

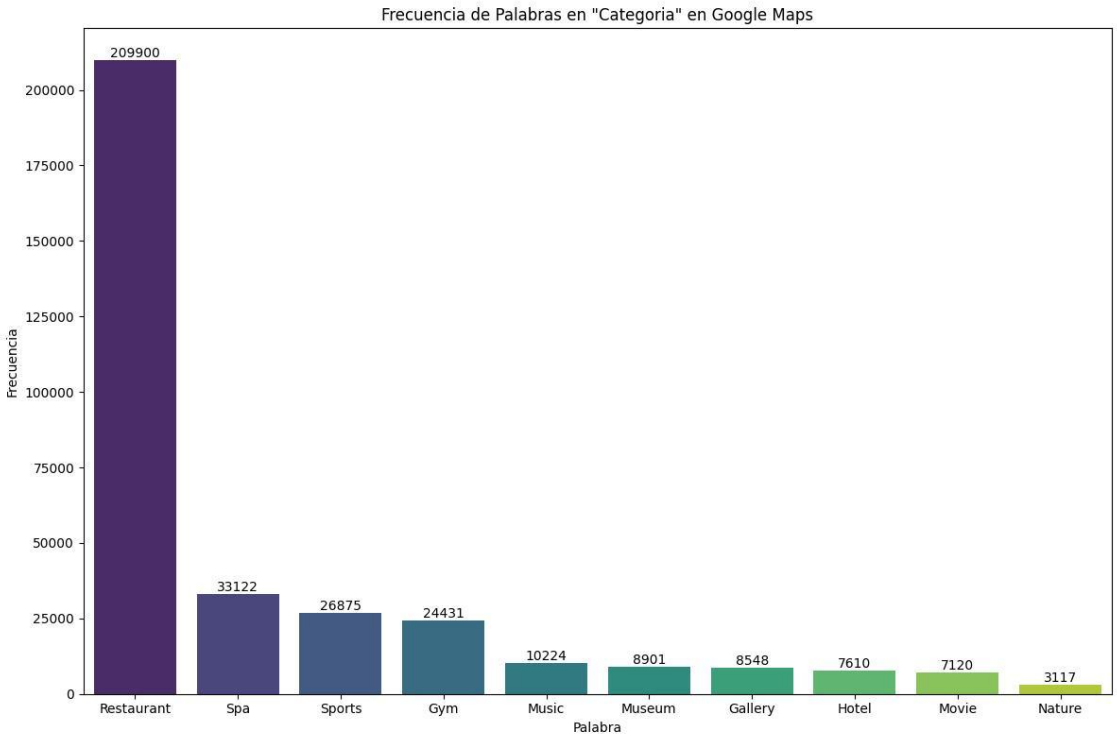
MODELOS DE MACHINE LEARNING

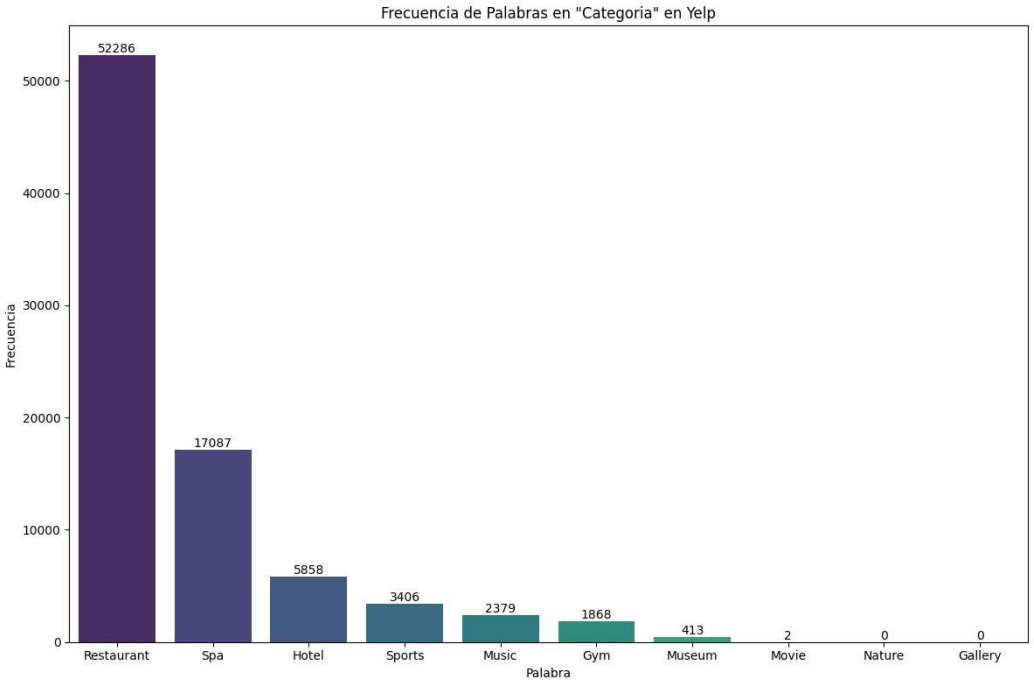
EDA

El EDA, es el Análisis Exploratorio de los Datos. Es un proceso que abarca todo el ciclo de vida de los datos empezando desde el ETL (Extracción, Transformación y Carga en español) hasta el análisis a profundidad de la información.

El EDA, durante el ETL, se encarga de ayudar a comprender la estructura de los datos fuente analizando su calidad, identificando posibles problemas o anomalías y en la toma de decisiones sobre qué datos extraer y cómo. Durante la transformación, se utiliza para limpiar y pre-procesar los datos mediante el manejo de valores atípicos, datos faltantes o duplicados; y antes de la carga final, ayuda a validar que los datos hayan sido transformados acorde a las expectativas y requisitos específicos.

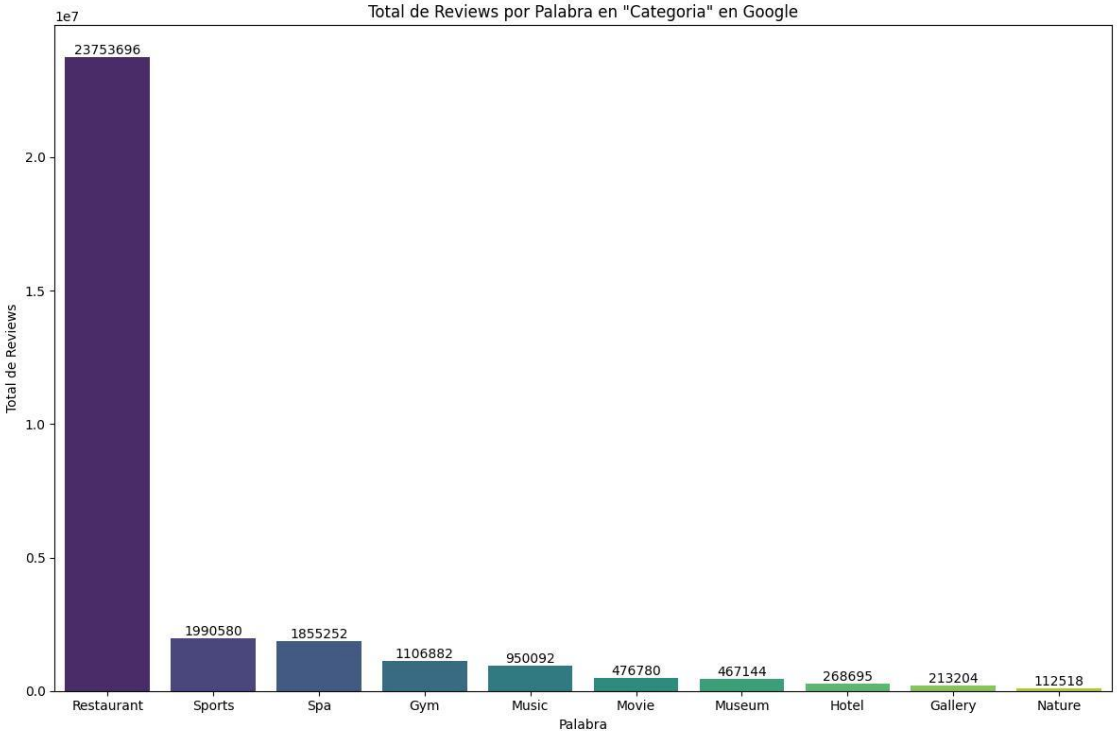
Partiendo del conjunto de datos proporcionados de Google Maps y Yelp, se llevó a cabo la carga y transformación de la información en dataframes. Completada la carga y transformación, decidió aplicarse un primer filtro con el objetivo de identificar las actividades relacionadas con el ocio, utilizando palabras claves como “restaurant, spa, sports, gym, music, museum, gallery, hotel, movie y nature”.

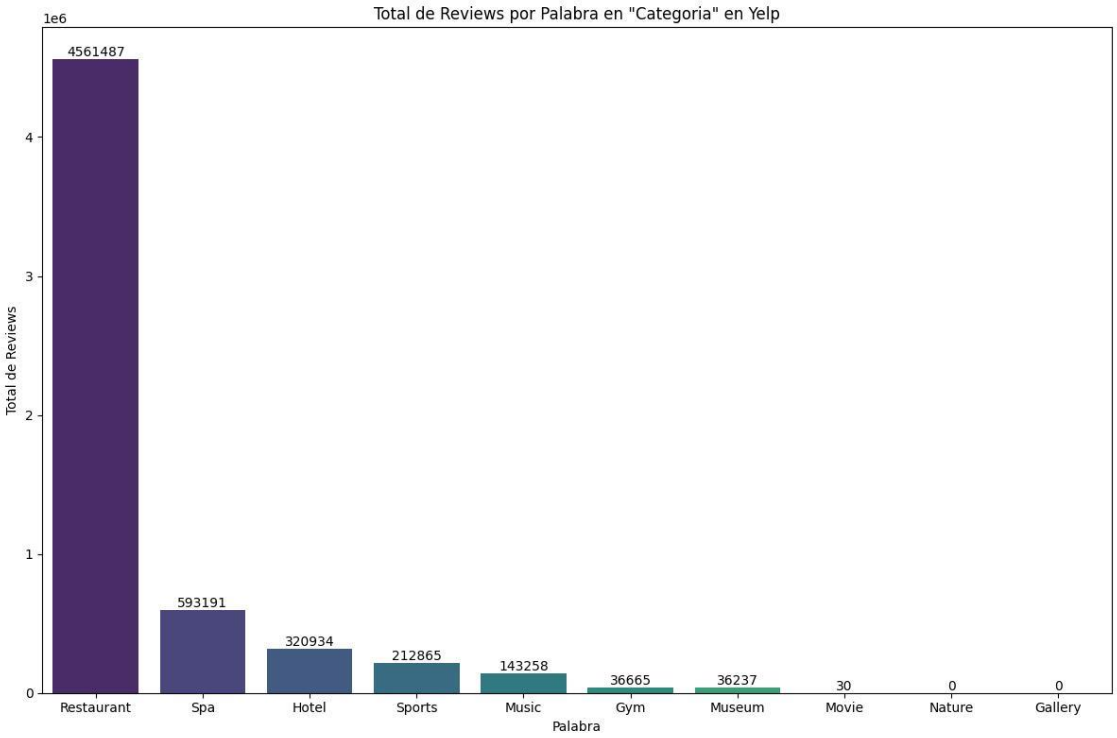




De ambos gráficos podemos observar que la categoría “restaurant” es el tipo de negocio que más abunda en ambos datasets con amplia diferencia con respecto al resto.

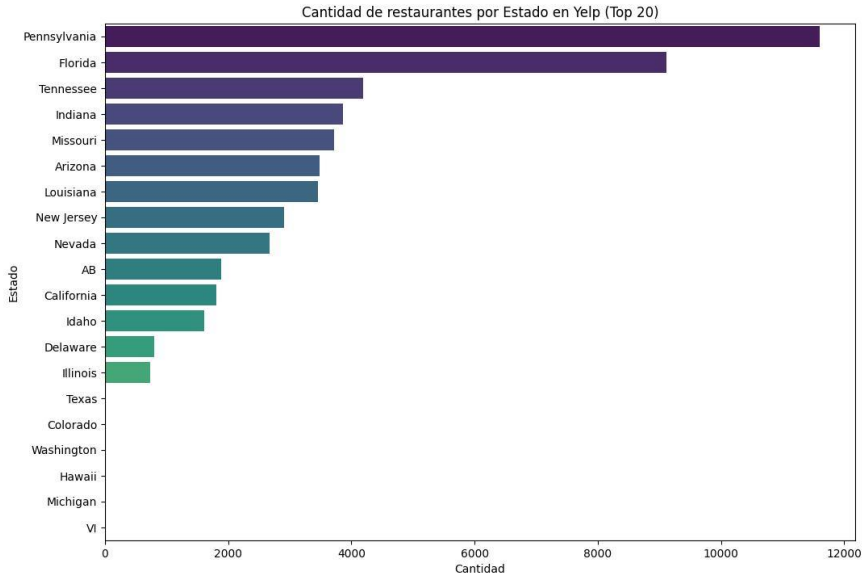
De manera similar se realiza el mismo filtro sobre el dataframe de reseñas para conocer cuales es el tipo de negocio que posee mayor cantidad de reseñas. Así los gráficos obtenidos demuestran lo siguiente.

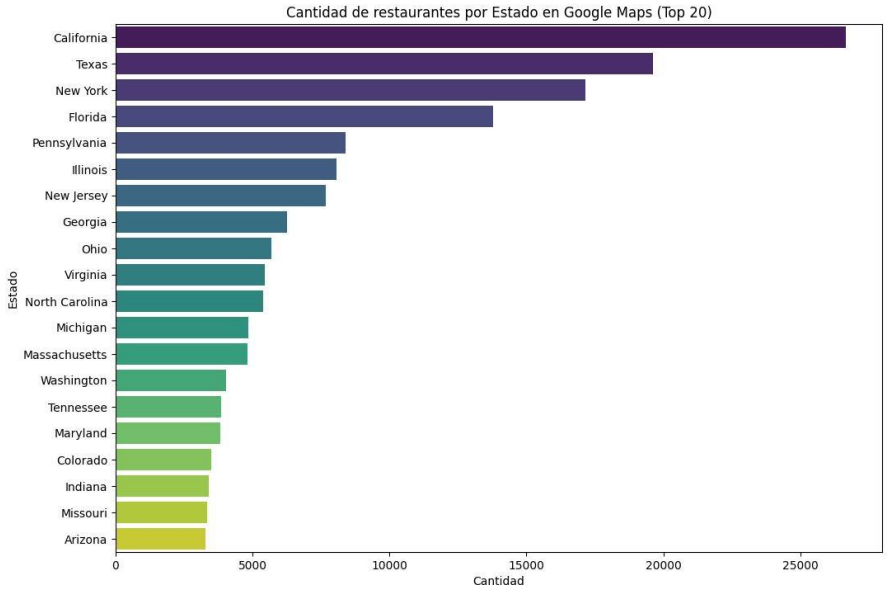




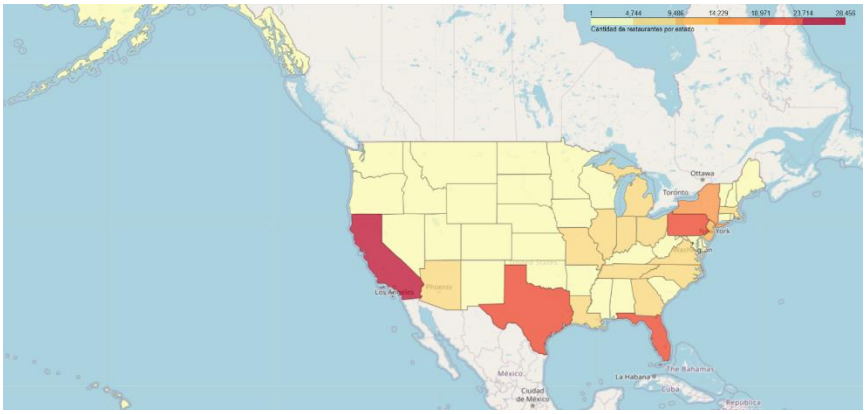
Basándonos en estos criterios, fue que el grupo decidió enfocar su análisis al sector gastronómico. Pero este análisis no es suficiente.

Debido a la gran cantidad de estados pertenecientes al territorio de Estados Unidos (cuenta con 50 estados) fue que decidió realizarse un top 20 de los mismos para poder vislumbrarlos con claridad.





Para poder visualizarlo de mejor manera, decidió realizarse un mapa de calor con el total de restaurantes por estado.

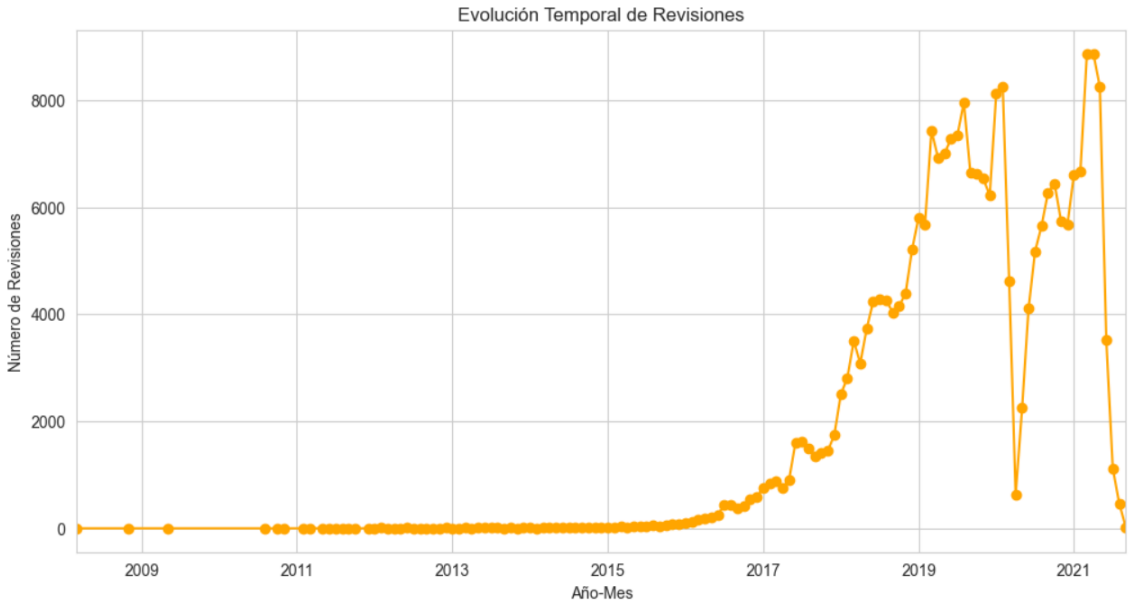


California	28456
Florida	22889
Pennsylvania	20010
Texas	19626
New York	17160
New Jersey	10593
Illinois	8807
Tennessee	8037
Indiana	7251
Missouri	7064
Arizona	6772
Louisiana	6474
Georgia	6249
Ohio	5679
Virginia	5443
North Carolina	5385
Michigan	4860
Massachusetts	4814
Nevada	4684
Washington	4030
Maryland	3824
Colorado	3504
Wisconsin	3235
Connecticut	3000
Oregon	2985

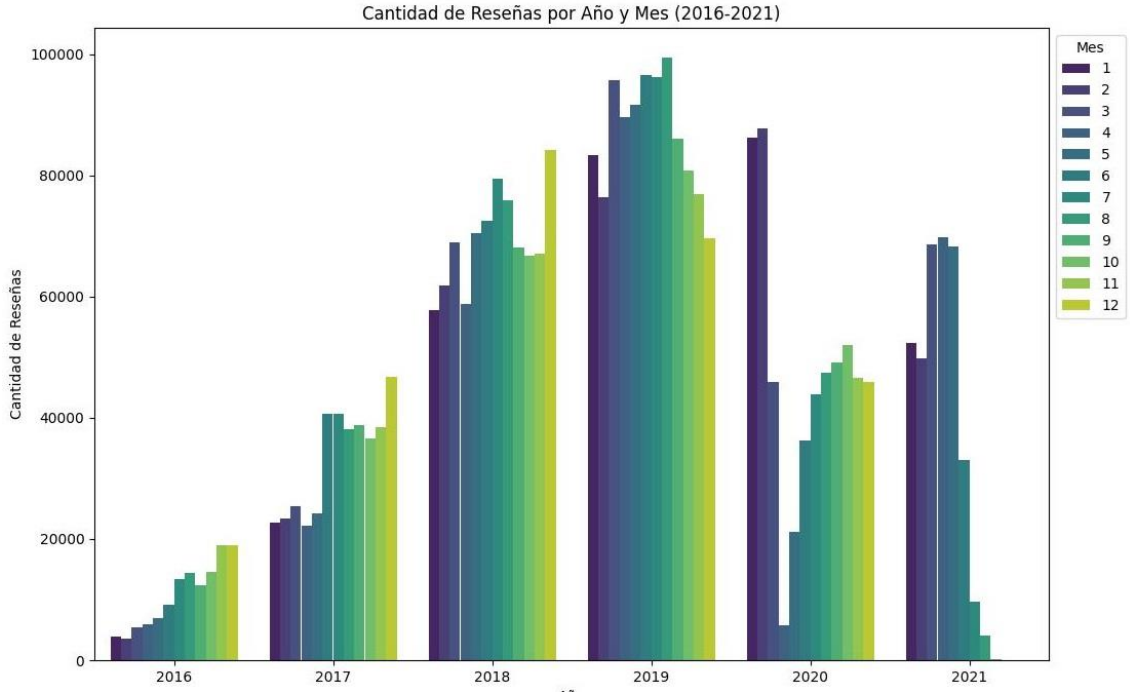
En la tabla adjunta, pueden observarse los primeros 25 estados que poseen mayor cantidad de restaurantes.

Esta información, junto a los criterios de selección de estados a analizar, redujo nuestro conjunto de estudio a solo 8 estados: Arizona, California, Florida, Tennessee, Pennsylvania, Illinois, New Jersey e Indiana.

Tomando como referencia los KPIs planteados por el grupo, se realizó un EDA sobre el conjunto de datos referidos a las reseñas. Primero se vio la evolución de la cantidad de reseñas por año.

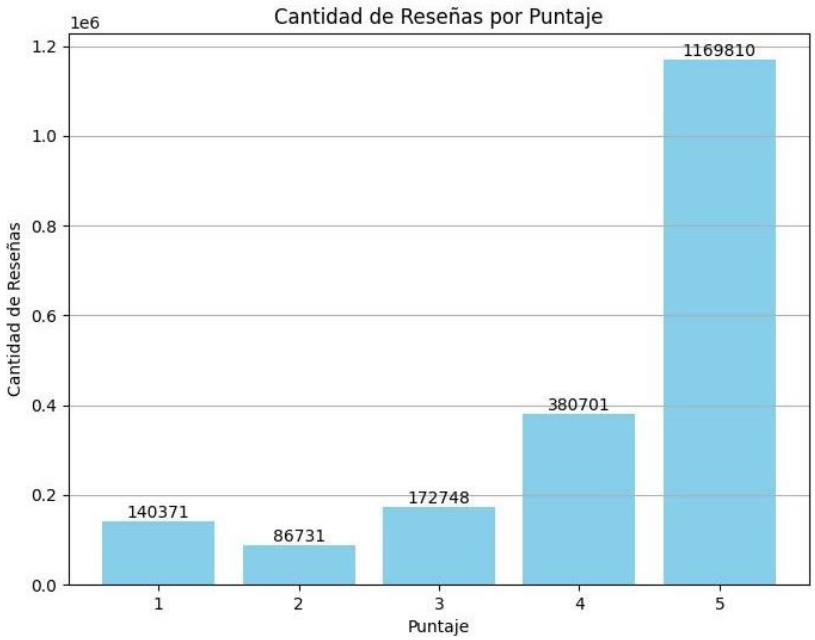


Al achicar nuestra ventana temporal a los últimos 5 años, podemos ver que las reseñas disminuyeron a partir de los primeros meses del año 2020, que fue cuando se decretó el estado de emergencia por la pandemia del Covid-19 donde varios locales debieron cerrar sus puertas a la atención al público. Podemos ver que, durante el segundo y tercer trimestres del mismo año, empiezan a incrementar levemente las reseñas, esto puede haberse debido a la evolución de las restricciones planteadas como el uso obligatorio de barbijos, la desinfección y limpieza constante de los locales y la asistencia limitada a los mismos.

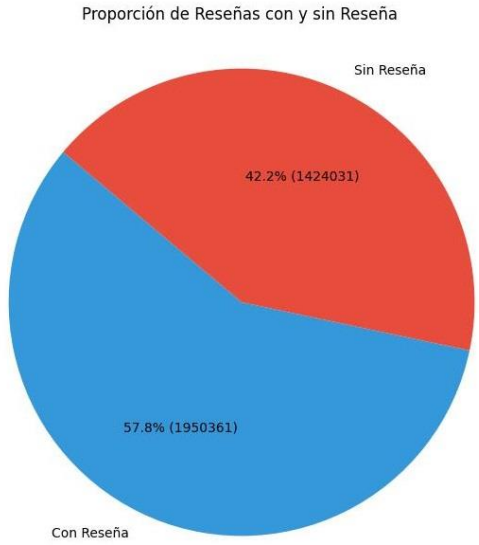


Considerando el puntaje de las reseñas podemos observar que la mayoría de reseñas posee un valor de 5, interpretándose como una señal positiva y alentadora. Esto sugiere que los clientes están muy

satisfechos con la calidad general del servicio (comida, atención y experiencia del restaurante), que los restaurantes han mantenido un alto nivel de rendimiento a lo largo del tiempo y que los restaurantes reseñados de esta forma han consolidado una sólida reputación en línea.



En el gráfico de tortas podemos ver que del total de usuarios que puntúan restaurantes, el 42% no deja alguna reseña del mismo siendo esta una forma efectiva de marketing en una era digital. Esto perjudica gravemente al negocio, la falta de reseña provoca una menor visibilidad en línea ya que los motores de búsqueda suelen mostrar primero aquellos negocios con más reseñas y calificaciones positivas y ya sean reseñas positivas o negativas proporcionan una retroalimentación valiosa para los propietarios, indicando los aspectos de su servicio que podrían necesitar mejoras.



Identificación de Patrones:

Se pueden identificar varios patrones y prácticas comunes en el análisis de datos y la construcción de Modelos de Machine Learning.

Ingeniería de Características:

Columna 'stars_review_interaction': Se crea una nueva característica multiplicando las columnas 'stars' y 'review_count', lo que puede ayudar a capturar la interacción entre la calificación y la cantidad de reseñas.

Limpieza y Preprocesamiento de Datos:

Manejo de Valores Nulos: Se eliminan las filas con valores NaN para garantizar la integridad de los datos.

Normalización y Estandarización: Se utiliza LabelEncoder para codificar variables categóricas y StandardScaler para normalizar y estandarizar características numéricas.

Análisis de Sentimientos:

Sentiment Analysis: Se utiliza la librería NLTK para realizar un análisis de sentimientos en la columna 'attributes', calculando el puntaje de sentimiento y agregándolo como 'sentiment_score'.

Modelos de Machine Learning:

Random Forest Classifier: Se utiliza un modelo de clasificación (RandomForestClassifier) para predecir la variable 'is_open' basado en características como el puntaje de sentimiento, estrellas y cantidad de reseñas.

Evaluación del Modelo: Se evalúa el rendimiento del modelo utilizando métricas como precisión (accuracy) y se muestra la matriz de confusión.

Interacción con el Usuario:

Ingreso de Ciudad:

El usuario es solicitado a ingresar el nombre de una ciudad.

Entrada del usuario se convierte a minúsculas para hacer la coincidencia insensible a mayúsculas y minúsculas.

Se filtran los datos del DataFrame original (df) según la ciudad proporcionada.

```
city_input = input("Ingrese la ciudad: ")
city_input = city_input.lower()
city_df = df[df['city'].str.lower() == city_input]
```

Recomendaciones de Ciudad: Si hay datos disponibles para la ciudad ingresada, se muestran recomendaciones de restaurantes en esa ciudad, incluyendo el nombre, el número de estrellas y el tipo de comida.

En caso contrario, se informa al usuario que no hay datos disponibles para esa ciudad.

Predicción de Crecimiento: El usuario proporciona el nombre de una ciudad para obtener una predicción de crecimiento basada en la satisfacción del cliente y la calidad del servicio.

```
ciudad = input("Ingrese la ciudad: ")
# Realiza la predicción y muestra los resultados basados en la entrada
```

Se muestra la predicción de crecimiento, el promedio de calidad del servicio y el tipo de comida predominante.

Predicción de Crecimiento Basada en Reseñas

La sección de predicción de crecimiento basada en reseñas utiliza un modelo de regresión para predecir el crecimiento en el número de reseñas de restaurantes. Aquí se detalla el proceso paso a paso:

Carga de Datos y Preprocesamiento

Descripción:

Se carga el conjunto de datos y se realiza ingeniería de características, creando una nueva característica llamada 'stars_review_interaction'.

Se filtran solo los restaurantes abiertos para enfocarse en el crecimiento relevante.

Se seleccionan características relevantes para el modelo, incluyendo ubicación, calificación de estrellas, cantidad de reseñas, interacción entre estrellas y reseñas, estado y categorías.

Las variables categóricas se codifican numéricamente y se realiza normalización y estandarización de las características numéricas.

División de Datos Los datos se dividen en conjuntos de entrenamiento y prueba para evaluar el rendimiento del modelo

Descripción Se utiliza un modelo de regresión de bosque aleatorio para predecir el crecimiento en el número de reseñas.
Se realiza una búsqueda de cuadrícula para optimizar los hiperparámetros del modelo.
Se realizan predicciones de crecimiento y se calculan los porcentajes de crecimiento para cada estado. Se identifican los dos estados que más crecen y los dos estados que más decrecen.
Construcción y Entrenamiento del Modelo En la sección se lleva a cabo la implementación y ajuste de un modelo de regresión utilizando la técnica de bosque aleatorio. Este modelo se entrena para predecir el crecimiento en el número de reseñas de restaurantes.

Selección del Modelo:

Se elige un modelo de regresión conocido como RandomForestRegressor. Este modelo pertenece a la categoría de bosques aleatorios, que es una técnica de aprendizaje automático basada en árboles de decisión.

Parámetros del Modelo:

Se establecen diferentes valores para los hiperparámetros del modelo, en este caso, el número de árboles en el bosque (n_estimators) y la profundidad máxima de los árboles (max_depth).

Se crea un diccionario param_grid que contiene combinaciones de estos valores para explorar.

Búsqueda de Cuadrícula (Grid Search):

La búsqueda de cuadrícula (GridSearchCV) es una técnica que evalúa exhaustivamente las combinaciones de hiperparámetros para encontrar los valores óptimos que maximizan el rendimiento del modelo.

Se utiliza validación cruzada con tres particiones (cv=3) para evaluar el rendimiento del modelo en diferentes subconjuntos de datos.

Entrenamiento del Modelo:

El modelo se entrena utilizando los datos de entrenamiento (X_train y y_train), donde X_train son las características y y_train son las etiquetas (número de reseñas).

Durante el entrenamiento, el modelo ajusta sus parámetros para minimizar la diferencia entre las predicciones y los valores reales.

Después de entrenar el modelo, se utilizan las predicciones del modelo para calcular el crecimiento previsto en el número de reseñas. Este crecimiento se expresa como un porcentaje en comparación con el número actual de reseñas.

En resumen, esta sección se centra en la construcción y entrenamiento del modelo de regresión para predecir el crecimiento en el número de reseñas de restaurantes. La búsqueda de cuadrícula ayuda a encontrar los mejores hiperparámetros para maximizar la precisión del modelo.

Interfaz de Usuario (FastAPI)

La interfaz de usuario se implementa utilizando FastAPI, un marco moderno de Python para la creación de API web de manera rápida y sencilla. La interfaz permite a los usuarios interactuar con el sistema y

obtener recomendaciones específicas y predicciones sobre el crecimiento de restaurantes. Aquí está el desglose de la implementación:

Definición de Modelos (Pydantic):

Se definen tres clases (CiudadInput, PreferenciaInput, y CrecimientoInput) utilizando Pydantic para validar y estructurar la entrada de los usuarios.

En general, la interacción con el usuario se realiza mediante la entrada de datos a través de la consola, y los resultados se presentan de manera informativa. La estructura de FastAPI también proporciona puntos de entrada específicos para realizar consultas a través de una interfaz de API si el código se ejecuta como un servicio web.

Endpoints de FastAPI:

Se definen varios endpoints (/recomendacion_ciudad/ y /prediccion_crecimiento/) que los usuarios pueden utilizar para obtener recomendaciones de restaurantes para una ciudad específica y recibir predicciones de crecimiento.

Manejo de Errores con FastAPI:

Se utiliza HTTPException para manejar situaciones donde no hay datos disponibles para la ciudad proporcionada o si se ingresan preferencias no válidas.

Procesamiento de Entrada y Obtención de Resultados:

Se procesa la entrada del usuario y se utilizan las funciones definidas previamente para realizar predicciones y devolver resultados estructurados.

En resumen, la interfaz de usuario permite a los usuarios interactuar con el sistema mediante el envío de solicitudes a través de los endpoints definidos, proporcionando recomendaciones específicas para una ciudad y predicciones de crecimiento basadas en la satisfacción del cliente.

Se utiliza HTTPException para manejar situaciones donde no hay datos disponibles para la ciudad proporcionada o si se ingresan preferencias no válidas.

Mejoras Potenciales:

Optimización de Hiperparámetros:

Se realiza una búsqueda de cuadrícula (GridSearchCV) para optimizar los hiperparámetros del modelo, lo que puede mejorar el rendimiento.

Manejo de Excepciones:

Se manejan excepciones para casos en los que no hay datos disponibles o se proporcionan preferencias no válidas.

Construcción y Entrenamiento:

Otro modelo se construye para predecir el crecimiento en el número de reseñas utilizando un Random Forest Regressor.

Se aplican técnicas de ingeniería de características y se normalizan y escalan los datos.

Evaluación del Desempeño:

Se realiza una búsqueda de cuadrícula para optimizar los hiperparámetros del modelo.

Se evalúa la capacidad predictiva del modelo mediante métricas como el error cuadrático medio.

Iteración y Mejora Continua:

Análisis de Resultados:

Se analizan los resultados de los modelos, identificando posibles áreas de mejora en términos de precisión y rendimiento.

Posible Ajuste de Hiperparámetros:

Se podría considerar ajustar los hiperparámetros de los modelos para mejorar aún más su rendimiento.

Evaluación del Sistema y de la Interfaz de Usuario

Precisión en la Predicción de Tipos de Restaurantes:

Desempeño del Modelo de Clasificación:

La precisión en la predicción de si un restaurante está abierto podría analizarse para determinar la eficacia del modelo en este aspecto.

Efectividad del Modelo de Recomendación:

Recomendaciones Personalizadas:

Se evalúa la capacidad del sistema para proporcionar recomendaciones personalizadas.

Se utiliza el endpoint /recomendacion_ciudad/ para evaluar la calidad y relevancia de las recomendaciones.

Retroalimentación de Usuarios:

Recopilación de Comentarios:

Se podría implementar un sistema de recopilación de comentarios y reseñas de usuarios.

La retroalimentación de los usuarios sería crucial para evaluar la aceptación y utilidad de las recomendaciones.

Escalabilidad y Eficiencia:

Manejo de Grandes Volúmenes de Datos:

Se debe evaluar la capacidad del sistema para manejar grandes volúmenes de datos.

La eficiencia del sistema, especialmente en términos de tiempo de respuesta para las recomendaciones, es crucial.

Conclusiones y Acciones Siguientes:

Análisis Integral:

Se realizará un análisis integral de los resultados, considerando las métricas de rendimiento, la satisfacción del usuario y la eficiencia del sistema.

Mejoras Continuas:

Con base en los resultados y la retroalimentación, se planificarán mejoras continuas en los modelos y en la interfaz de usuario.

Optimización de Recursos:

Se buscarán oportunidades para optimizar los recursos computacionales y mejorar la eficiencia del sistema.

Iteración del Ciclo:

El proceso de mejora continua será cíclico, con iteraciones basadas en la evolución de los datos y las necesidades de los usuarios.

Este enfoque estructurado permite una evaluación completa del sistema, abordando aspectos clave como la precisión del modelo, la efectividad de las recomendaciones y la retroalimentación de los usuarios para garantizar un sistema robusto y orientado a la satisfacción del usuario.

INFORME DE CONTROL DE CALIDAD DEL PROYECTO DE MACHINE LEARNING

Habiendo completado el desarrollo del proyecto de machine learning para análisis de mercado y recomendaciones de restaurantes, se llevó a cabo un proceso exhaustivo de Control de Calidad (QA) para garantizar la efectividad y confiabilidad de los resultados obtenidos a lo largo de todo el proyecto. A continuación, se detallan las etapas y los hallazgos del proceso QA:

1. Verificación de Calidad de Datos:

Resultado:

Se verificó la integridad de los datos, asegurando que no haya valores nulos o inconsistencias.

Se validaron los datos de entrada para confirmar que cumplen con los requisitos del modelo.

2. Análisis Exploratorio de Datos (EDA):

Resultado:

Se realizó un análisis detallado de las distribuciones y se identificaron y manejaron outliers de manera efectiva.

3. Validación de Modelos de Machine Learning:

Resultado:

Las métricas de evaluación, incluyendo precisión, recall y F1-score, confirmaron la calidad y estabilidad de los modelos.

La validación cruzada demostró la robustez del modelo frente a diferentes conjuntos de datos.

4. Pruebas de Escalabilidad:

Resultado:

El sistema demostró su capacidad para manejar grandes volúmenes de datos sin degradar el rendimiento.

Los tiempos de respuesta cumplen con los requisitos de rendimiento establecidos.

5. Pruebas de Integración:

Resultado:

Las integraciones con APIs externas y servicios en la nube se realizaron de manera exitosa y consistente.

6. Pruebas de Usabilidad y Retroalimentación de Usuarios:

Resultado:

La interfaz de usuario en Power BI se evaluó positivamente en términos de usabilidad y facilidad de interpretación.

La retroalimentación de los usuarios destacó la efectividad de las recomendaciones y la presentación de datos en el dashboard.

7. Monitoreo Continuo:

Resultado:

Se implementó un sistema de monitoreo continuo para supervisar el rendimiento de los modelos en producción.

Las actualizaciones periódicas de los modelos se programaron para adaptarse a cambios en los datos o comportamiento de los usuarios.

El proceso de QA confirmó la calidad y confiabilidad del proyecto de machine learning.

Los resultados positivos sugieren que el sistema está listo para ser implementado y proporcionar valor a los usuarios finales.

Se recomienda un monitoreo continuo y actualizaciones periódicas para mantener la efectividad a lo largo del tiempo.

Este informe refleja la dedicación y el enfoque en garantizar la excelencia en todas las fases del proyecto, desde la adquisición de datos hasta la implementación de modelos en producción.

CONCLUSIONES

El proyecto de análisis de mercado y recomendaciones de restaurantes en Estados Unidos ha culminado con éxito, ofreciendo una perspectiva profunda sobre la industria y proporcionando valiosas herramientas para inversores y usuarios finales. La evaluación detallada de indicadores clave, como crecimiento, inversión, margen y rentabilidad, respaldó la selección de estados con mayor potencial en Estados Unidos.

El análisis de los Key Performance Indicators (KPIs) proporcionó información crítica sobre el rendimiento del proyecto, identificando áreas de mejora y oportunidades para optimización.

La elección sugerida de la industria de restaurantes se basó en un análisis exhaustivo que consideró el crecimiento, la inversión y la flexibilidad, identificando oportunidades significativas para la inversión. La implementación exitosa del proceso de Extracción, Transformación y Carga (ETL) permitió integrar datos de reseñas de Google Maps y Yelp, proporcionando una base sólida para el análisis de mercado.

La creación de modelos de machine learning para la predicción de tipos de restaurantes y recomendaciones a usuarios ha demostrado ser efectiva, ofreciendo herramientas valiosas para la toma de decisiones.

Se recomienda una monitorización continua y actualizaciones periódicas de los modelos para adaptarse a cambios. Explorar estrategias adicionales para mejorar KPIs y fomentar la participación de usuarios y restaurantes puede potenciar aún más el éxito del proyecto.

Finalmente, el proyecto ha logrado sus objetivos al proporcionar análisis detallados, modelos predictivos y recomendaciones efectivas. La implementación y seguimiento continuo permitirán capitalizar las oportunidades identificadas y adaptarse a las dinámicas cambiantes del mercado. Este proyecto no solo ha fortalecido la comprensión del mercado de restaurantes, sino que también ha sentado las bases para futuras innovaciones y mejoras en la toma de decisiones estratégicas.