

Maximizar el impacto y la relevancia del turismo gastronómico, en el sector de la industria turística.

Análisis de mercado en Estados Unidos.



SOBRE NOSOTROS

DATUM TECH

NUESTRA VISIÓN: Ser conocidos como líderes en el campo de la consultoría especializada en análisis de datos, brindando soluciones innovadoras y de vanguardia que impulsen la toma de decisiones fundamentadas en empresas de diversas industrias. Nuestro objetivo es ser el socio preferido de nuestros clientes en su viaje hacia la transformación digital, generando un impacto positivo en la optimización de sus servicios y productos.

NUESTRA MISIÓN: Nos dedicamos a ofrecer servicios de consultoría especializada en análisis de datos y desarrollo de soluciones para empresas que buscan tomar decisiones informadas. Nuestra misión es ayudar a nuestros clientes a desbloquear el poder de sus datos, brindándoles conocimientos valiosos y soluciones prácticas que les permitan optimizar sus procesos, mejorar la eficiencia operativa y alcanzar resultados exitosos. Nos esforzamos por ser un socio estratégico, confiable y comprometido con el éxito de nuestros clientes, aportando experiencia técnica y conocimientos de negocio para impulsar su conocimiento y competitividad en un entorno digital en constante evolución.

NUESTRO EQUIPO DE TRABAJO

- Carlos Rivera: Data Analytics/Data Scientist
- Darwin Moreno: Data Engineer/Data Scientist
- Soledad Rozzatti: Data Analytics
- Stefany Rojas: Data Analytics
- Wilson Waigant: Data Engineer/Data Scientist



PRESENTACIÓN DEL PROYECTO

NUESTRO CLIENTE

TURFOOD es una empresa filial del conglomerado TURISK LTD. La cual busca, ofrecer experiencias gastronómicas únicas y memorables en el sector del turismo gastronómico. Por ello, TURFOOD ha decidido establecer una alianza estratégica con nuestra consultora, empleando nuestros servicios para llevar a cabo un análisis exhaustivo de datos provenientes de las plataformas Google Maps y Yelp. En busca de soluciones que impacten de manera positiva su portafolio en el sector del turismo gastronómico.

Durante el desarrollo de este proyecto proporcionaremos un análisis y comprensión de los datos extraídos de ambas plataformas, para transformarlos en señales que identifiquen patrones, tendencias y oportunidades de mejora.


Proveyendo de esta forma información valiosa para la creación de estrategias que mejoren la experiencia de los usuarios, optimicen los procesos internos y maximicen la rentabilidad de TURFOOD.

ANTECEDENTES

Según los informes publicados por la Organización Mundial del Turismo (www.unwto.org), el turismo en los Estados Unidos (EE.UU.) experimentó un registro de 45 millones de turistas en el año 2020, posicionándose en el cuarto lugar a nivel mundial en términos absolutos.

Más allá de ser evidente que los países de mayor tamaño suelen tener un mejor desempeño en cuanto al número absoluto de visitantes, al considerar el número de turistas en relación con la población de EE.UU., se obtiene una imagen más comparativa. Con 0.14 turistas por residente, EE.UU., se sitúa en el puesto 106 a nivel mundial y en el tercer lugar en la región de Norteamérica.





En el sector turístico, EE.UU. generó alrededor de 84,21 mil millones de dólares estadounidenses solo en el año 2020. Esta cifra representa aproximadamente el 0,36 por ciento de su producto interno bruto y constituye cerca del 100 por ciento de todos los ingresos del turismo internacional en Norteamérica.

Dentro del sector turístico, se identifica una rama específica conocida como turismo gastronómico.

El turismo gastronómico en EE.UU. es una tendencia en auge que puede potenciar y enriquecer la experiencia turística en el país. Esta forma de turismo se centra en la exploración de la diversidad culinaria y ofrece a los viajeros la oportunidad de descubrir y disfrutar de la rica oferta gastronómica que EE.UU. tiene para ofrecer.

Una de las características más destacadas de este tipo de turismo es la amplia variedad de platos típicos y especialidades regionales. Desde los sabores auténticos del sur hasta la fusión de influencias internacionales en las principales ciudades, el país ofrece una gama diversa de opciones culinarias para satisfacer todos los gustos. Los viajeros tienen la oportunidad de probar nuevos sabores, experimentar la cultura local a través de la comida y deleitarse con experiencias gastronómicas únicas.

Además, el turismo gastronómico no se limita solo a la comida, sino que también se trata de las experiencias y actividades que la rodean. Los turistas pueden participar en recorridos gastronómicos para explorar barrios históricos, probar platos tradicionales y aprender sobre la historia y la cultura local. También, pueden disfrutar de clases de cocina interactivas, visitas a viñedos y cervecerías artesanales, y participar en festivales y eventos culinarios que celebran la diversidad gastronómica del país. Por ello, el turismo gastronómico es un recurso fundamental para promover la industria turística en general, ya que se beneficia con el crecimiento de la demanda de hoteles, restaurantes y servicios relacionados.

En conclusión, el turismo gastronómico en EE.UU. tiene el potencial de potenciar y enriquecer la experiencia turística al ofrecer una variedad de sabores, experiencias culinarias únicas y la posibilidad de sumergirse en la cultura local a través de la comida.



OBJETIVOS

Objetivo general:

Implementar una herramienta de análisis de datos escalable para comprender mejor a los consumidores de determinados productos o servicios en el ámbito de la hotelería y gastronomía, y en base a ello tomar las mejores decisiones de negocio.

Objetivo específicos:

- Disponibilizar información relevante y de calidad, extraída desde fuentes externas, necesaria para realizar el análisis de mercado solicitado.
- Evaluar las emociones, actitudes y opiniones de los usuarios, para obtener información que permita comprender la imagen que los mismos tienen respecto de un determinado producto o servicio ofrecido por nuestro cliente.
- Predecir qué rubros de negocios tendrán mayor y menor crecimiento asociado.
- Determinar potenciales oportunidades para futuros nuevos locales.
- Crear un sistema de recomendación de locales pertenecientes al sector estudiado..

ALCANCE

El análisis se enfocará en los datos obtenidos desde dos plataformas externas virtuales (Google Maps y Yelp). El mismo tendrá en cuenta el análisis cuantitativo del nivel de satisfacción de los usuarios, que permitirá un mejor conocimiento de los consumidores. De esta forma se logrará una mejor toma de decisiones y estrategias de mercado, basadas en información precisa y de calidad.

Adicionalmente, el proyecto se sustenta en la implementación de un sistema de recomendación de restaurantes y afines basado en el conocimiento del interés de los usuarios.



METODOLOGÍA DE TRABAJO SCRUM

La metodología elegida para el desarrollo de este proyecto fue la “ Metodología Scrum”. Scrum es un proceso en el cual se aplican de manera regular un conjunto de buenas prácticas para trabajar colaborativamente, en equipo, y obtener el mejor resultado posible de un proyecto. Estas prácticas se apoyan unas a otras y su selección tiene origen en un estudio de la manera de trabajar de equipos altamente productivos.

Consiste en realizar entregas parciales y regulares del producto final, priorizadas por el beneficio que aportan al receptor del proyecto. Donde los requisitos pueden ser cambiantes, donde la innovación, la competitividad, la flexibilidad y la productividad son fundamentales.

Planilla de metodología SCRUM:

Reunión Diaria de Scrum: lunes a viernes 11 hs (GMT-3)

Reunión de Planificación del Sprint: lunes a viernes 17hs (GMT-3)

Sprint 1: Semana 1 Fecha de Inicio: 29 de mayo

Fecha de Fin: 2 de junio - Revisión y Retrospectiva

Sprint 2: Semana 2 Fecha de Inicio: 5 de junio

Fecha de Fin: 9 de junio - Revisión y Retrospectiva

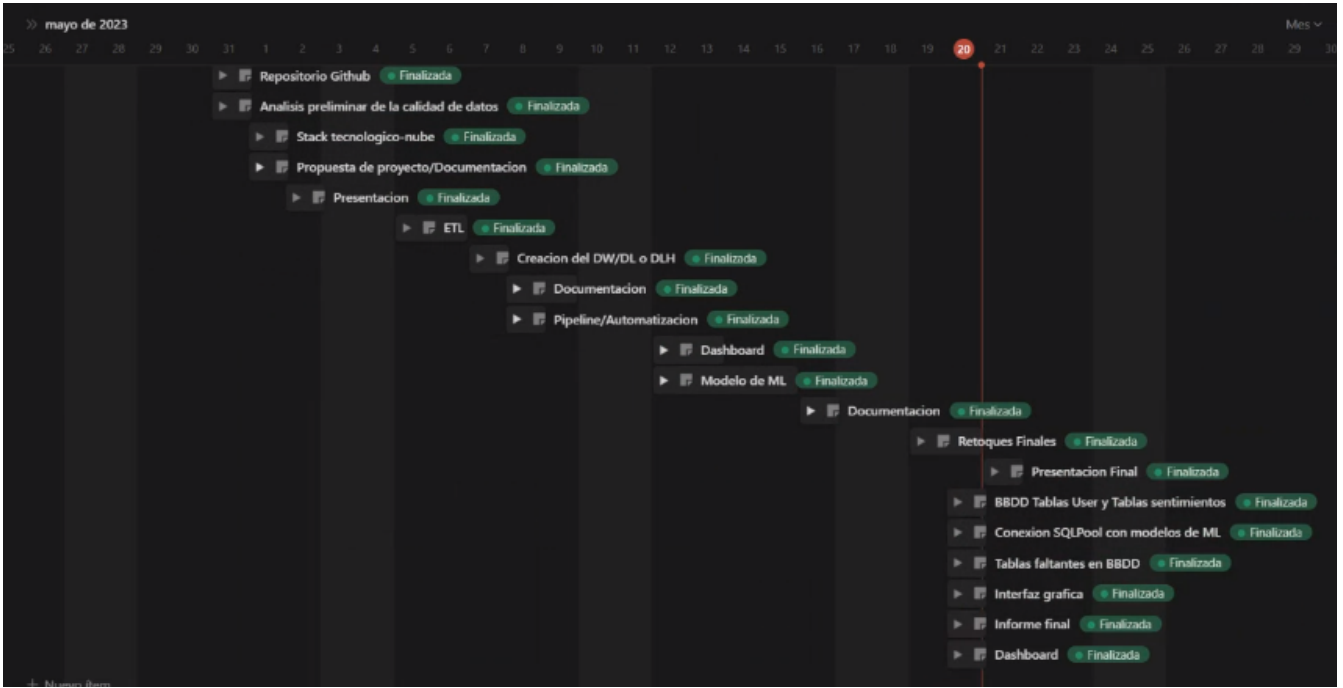
Sprint 3: Semana 3 Fecha de Inicio: 12 de junio

Fecha de Fin: 16 de junio - Revisión y Retrospectiva

Sprint 4: Semana 4 Fecha de Presentación Oficial del Proyecto: 21 de junio



Diagrama de Gantt:



Se puede acceder al diagrama, a los sprint y tareas definidas desde este [link](#).



INDICADORES CLAVE DE RENDIMIENTO PLANTEADOS

KPI - Key performance indicator

Los KPIs son herramientas esenciales para la gestión y el control efectivos, ya que brindan una visión clara y medible del desempeño y rendimiento de una organización.

Sus características principales son: específico, medible, alcanzable, relevante, tiempo determinado.

| KPI | Característica | Fórmula | Objetivo | Rango |
|------------------------------------|---|---|---|---------|
| Tasa de Crecimiento de Reseñas | Permite evaluar el impacto de las experiencias de los clientes en la participación activa de los mismos y la reputación y el éxito del negocio. | $\text{Tasa de Crecimiento de Reseñas} = \frac{(\text{Cantidad Reseñas Totales} - \text{Cantidad Reseñas Totales del mes anterior})}{\text{Cantidad Reseñas Totales del mes anterior}}$ | Aumentar la Tasa de Crecimiento de Reseñas en un 15%, mensualmente. | Mensual |
| Índice de Satisfacción del Cliente | Permite medir si un servicio cumplió con las expectativas del cliente | $\text{Índice de Satisfacción del Cliente} = \frac{(\text{Cantidad de Reseñas con rating } \geq 4)}{(\text{Cantidad de Reseñas Totales})}$ | Crear un índice de satisfacción 5% mensualmente. | Mensual |
| Rating Promedio | Permite evaluar el nivel de satisfacción | Rating Promedio = | Mantener un Rating Promedio igual o | Mensual |



| | | | | |
|--|--|--|--|---------|
| | del cliente respecto de los servicios ofrecidos. | (Suma total de Rating)/ (Cantidad Total de reseñas) | mayor a 4 para la categoría hotel, mantener un rating promedio igual o mayor a 4.5 para la categoría bakery, ambas mensualmente. | |
| Tasa de Reseñas Positivas | Proporciona una medida cuantitativa de la satisfacción del cliente, teniendo en cuenta su opinión. | Tasa de Reseñas Positivas = (Cantidad de Reseñas con Sentimiento Positivo Asociado) /(Cantidad Total de Reseñas) | Aumentar la Tasa de Reseñas Positivas en un 5%, mensualmente. | Mensual |
| Tasa de Crecimiento de Nuevos Clientes | Proporciona una visión general y una evaluación de la atracción de nuevos clientes. | Tasa de Crecimiento de Nuevos Clientes = (Cantidad Clientes Totales – Cantidad Clientes Totales del mes anterior) / (Cantidad Clientes Totales del mes anterior) | Aumentar la Tasa de Crecimiento de Nuevos Clientes en un 10% mensualmente. | Mensual |



STACK TECNOLÓGICO Y FLUJO DE TRABAJO

Nuestro enfoque se ha centrado en la adopción de Python como lenguaje de programación primario, respaldado por Azure como nuestro proveedor principal de servicios en la nube.

Para hacer un uso estratégico de los diversos servicios ofrecidos por Azure, garantizando una implementación ágil y robusta en el entorno de la nube, decidimos optar por **Azure Synapse Analytics**, la cual es una plataforma de análisis unificada y altamente escalable que nos ofrece numerosas ventajas para la integración de todos los recursos necesarios durante el desarrollo del proyecto.

1. Almacenamiento de datos:

Como sistema de almacenamiento de datos, utilizamos **Azure Data Lake Storage Gen2 (ADLS)**, un recurso altamente escalable y compatible con diversos formatos de datos, como Parquet y JSON, que son comúnmente utilizados para la recepción de archivos. Aprovechamos la capacidad de integración que Synapse Analytics nos ofrece con ADLS, permitiéndonos una gestión eficiente y unificada de los datos, optimizando nuestros procesos al conseguir un flujo de datos fluido y una infraestructura de trabajo robusta.

2. Procesamiento y Transformación de datos:

Al optar por Synapse Analytics, contamos con la posibilidad de disponer de un motor Apache Spark dentro de la misma plataforma. Permittiéndonos realizar actividades de procesamiento de datos de manera eficiente, sin necesidad de recurrir a plataformas externas. Y gozando de múltiples capacidades:

- **Procesamiento distribuido:** Apache Spark permite el procesamiento distribuido de datos, logrando trabajar con grandes volúmenes de datos de manera eficiente y paralela.



Consiguiendo así, la aceleración significativa del tiempo de procesamiento y mejora del rendimiento general de las operaciones.

- **Lenguajes de programación flexibles:** posibilidad de utilizar varios lenguajes de programación populares, como Scala y Python, para nuestras tareas de procesamiento y consultas. Brindando flexibilidad a la hora de la elección del lenguaje que se adapte mejor a nuestras necesidades y conocimientos.
- **Bibliotecas y funcionalidades avanzadas:** Apache Spark nos ofrece una amplia gama de bibliotecas y módulos que facilitan el análisis de datos, como Spark SQL para consultas SQL, Spark Streaming para procesamiento de datos en tiempo real, Spark MLlib para aprendizaje automático, y Spark GraphX para análisis de gráficos.
- **Escalabilidad y eficiencia:** El motor Apache Spark brinda la posibilidad de un auto escalado horizontal muy eficiente, a la hora de administrar los recursos de cómputo para manejar grandes volúmenes de datos y aumentar la capacidad de procesamiento según nuestra necesidad.
- **Integración con otros servicios de Azure:** Apache Spark se integra perfectamente con otros servicios de Synapse Analytics, permitiendo aprovechar la funcionalidad adicional y crear soluciones completas de extremo a extremo. Combinando Apache Spark con ADLS, Azure SQL Database y otros servicios para obtener una visión integral de nuestros datos.

De esta forma, logramos realizar una serie de transformaciones en los datos para asegurarnos de que estuvieran limpios y listos para su posterior análisis. Implementando diferentes técnicas, como la selección de columnas relevantes, la aplicación de filtros para eliminar datos no deseados y la creación de nuevas columnas basadas en cálculos o manipulaciones de los datos existentes. Estas transformaciones permitieron optimizar la calidad y estructura de los datos, preparándose adecuadamente para las etapas posteriores del análisis.



3. Carga de datos - Carga Incremental:

Una vez que los datos fueron transformados, nos enfocamos en cargarlos en una base de datos para su posterior análisis, **SQL Pool dedicado**, totalmente integrada dentro de la plataforma Synapse Analytics, que además nos permite optimizar recursos ya que sólo la mantenemos activa durante su consumo .

La integración de Synapse con SQL Pool dedicado y el motor de Apache Spark permite transferir de forma eficiente grandes volúmenes de datos entre el entorno de ejecución de Spark y el SQL Pool dedicado. El conector de SQL Pool dedicado para Apache Spark se implementa usando Scala y admite Scala y Python ofreciendo las siguientes características:

- Leer desde SQL Pool dedicado de Synapse: leer grandes conjuntos de datos desde tablas (internas y externas) y vistas de SQL Pool dedicado de Synapse.
- Soporte para la propagación de consultas.
- Escribir en SQL Pool dedicado de Synapse: ingerir grandes volúmenes de datos en tipos de tabla internos y externos.

Durante esta etapa, nos aseguramos de que los datos fueran almacenados de manera eficiente y confiable en la base de datos seleccionada. Organizamos los datos en tablas según su relevancia y utilizamos las capacidades de almacenamiento de la base de datos para realizar una carga eficiente. Este paso fue crucial para garantizar que los datos estuvieran disponibles y accesibles para su posterior análisis, creando una base sólida sobre la cual podemos obtener insights valiosos y tomar decisiones fundamentadas.

Para la carga incremental, se realiza una comparación de los nuevos datos con los datos existentes en el Data Lake mediante la utilización de columnas claves.

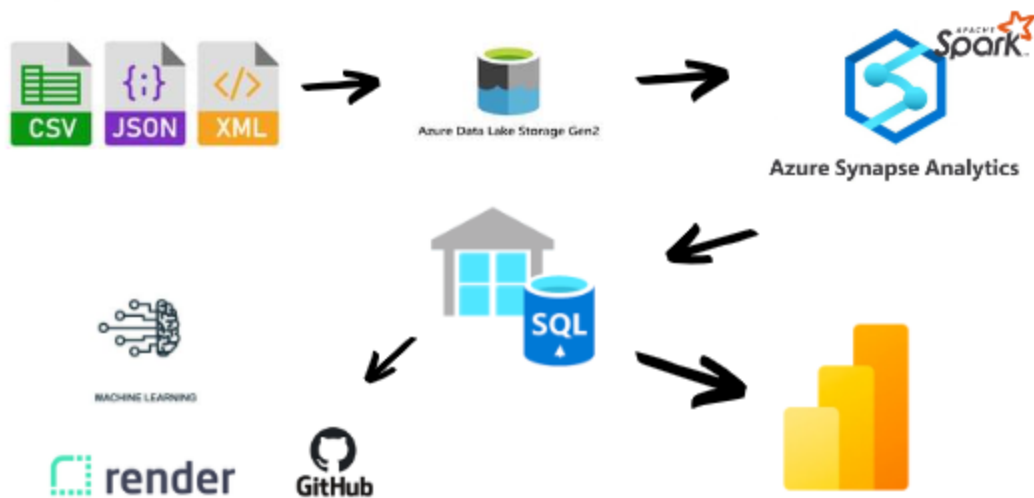
4. Disponibilización y Visualización de datos:

En cuanto a la visualización de los datos, Synapse Analytics ofrece una integración perfecta con Power BI, lo que facilita la visualización y el consumo de los datos analizados. Brindando la posibilidad de crear informes y paneles interactivos para obtener información valiosa de nuestros datos.



5. Github + Render

GitHub es una plataforma de desarrollo colaborativo que permite alojar y gestionar proyectos de código abierto. Render es un servicio de computación en la nube que ofrece una forma fácil de desplegar aplicaciones web, funciones sin servidor, bases de datos y otros recursos. Ambos servicios se integran entre sí y permiten crear flujos de trabajo automatizados para el desarrollo y la implementación de modelos. Con este método, se tiene un modelo de recomendación funcional y en línea, sin necesidad de emplear servidores ni infraestructura compleja. Además, se puede aprovechar la potencia de GitHub y Render para actualizar el modelo, escalar y monitorizar de forma sencilla.



MODELO DE RECOMENDACIONES

Un modelo de recomendaciones es una herramienta que ayuda a los usuarios a encontrar los lugares que más les gusten y se acomoden a sus necesidades, esto según sus preferencias o características. Los modelos de recomendaciones son muy útiles e importantes para ofrecer un servicio personalizado y mejorar la experiencia del usuario, así como para aumentar las ventas o la fidelización de los clientes. Con lo anterior en mente, Datum Tech desarrolló un modelo de recomendaciones amigable para el usuario y potente en sus resultados. A partir de ello se crearon dos endpoints con modelos diferentes.

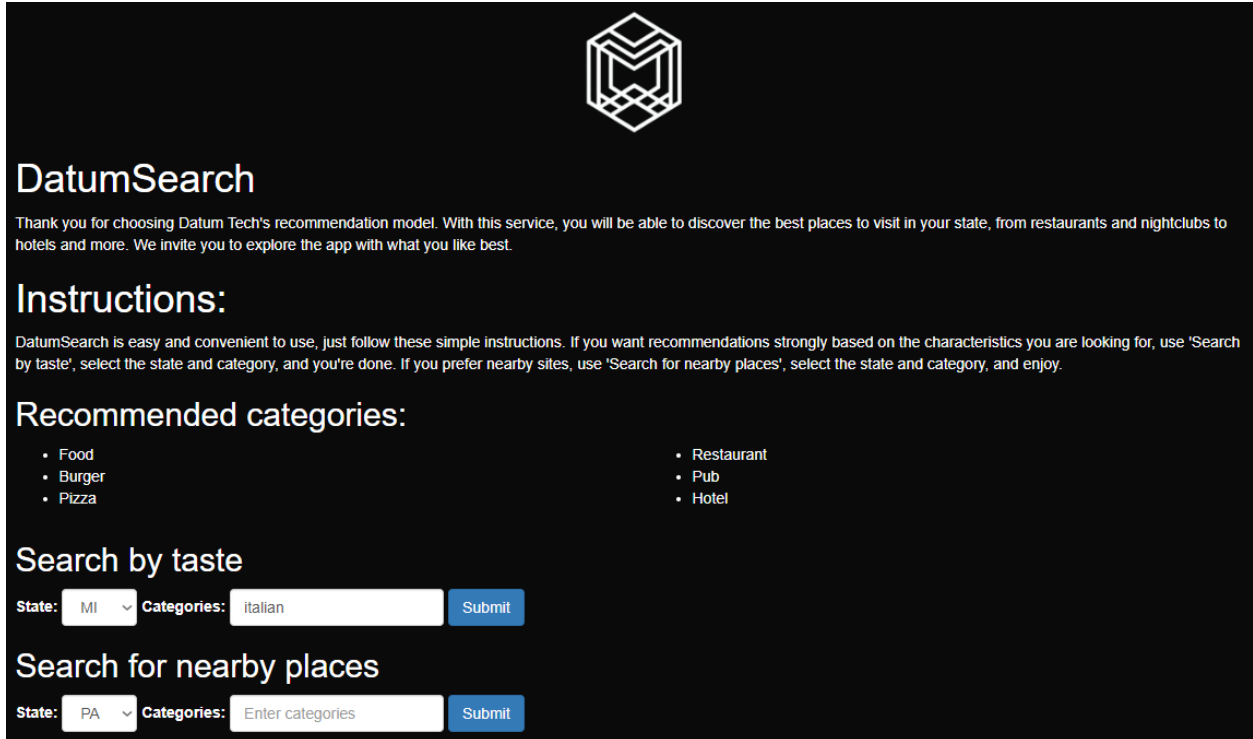
DatumSearch


DatumSearch es una interfaz de consultas desarrollada por Datum Tech que permite a los usuarios descubrir los mejores lugares para visitar en su estado. Con esta herramienta, el cliente podrá explorar una amplia variedad de opciones. La utilidad de DatumSearch radica en su capacidad para proporcionar recomendaciones personalizadas y actualizadas sobre los mejores lugares para visitar. Con esta herramienta, el cliente puede explorar una amplia variedad de opciones, desde restaurantes y clubes nocturnos hasta hoteles y más, todo basado en sus preferencias personales.

DatumSearch dispone de dos modelos de que el cliente puede realizar, esto incluye recomendaciones basadas en sus gustos, o en los lugares más cercanos en una ruta ubicados en una ruta, según las características deseadas. Lo anterior, permite al cliente descubrir nuevos lugares interesantes y emocionantes para visitar, sin tener que pasar horas buscando en línea o preguntando a amigos y familiares.



A continuación, una imagen de la interfaz:





DatumSearch

Thank you for choosing Datum Tech's recommendation model. With this service, you will be able to discover the best places to visit in your state, from restaurants and nightclubs to hotels and more. We invite you to explore the app with what you like best.

Instructions:

DatumSearch is easy and convenient to use, just follow these simple instructions. If you want recommendations strongly based on the characteristics you are looking for, use 'Search by taste', select the state and category, and you're done. If you prefer nearby sites, use 'Search for nearby places', select the state and category, and enjoy.

Recommended categories:

- Food
- Burger
- Pizza
- Restaurant
- Pub
- Hotel

Search by taste

State: Categories:

Search for nearby places

State: Categories:

Los modelos empleados son los siguientes:

Similitud del coseno

La similitud del coseno es una medida de la similitud entre dos vectores, funciona calculando el coseno del ángulo entre los dos vectores, que es igual al producto escalar de los vectores dividido por el producto de sus normas. El resultado es un valor entre -1 y 1, donde -1 indica una similitud nula, 0 indica una similitud media y 1 indica una similitud máxima. La similitud del coseno es de utilidad e importante en el modelo de recomendación porque permite calcular la afinidad entre los perfiles de los usuarios y las características que se quieren recomendar.

El modelo sigue los siguientes pasos para encontrar los mejores sitios adaptados al gusto del cliente:



-
-
- Filtra los datos según el estado y la categoría introducidos por el usuario.
 - Selecciona la información relevante para mostrar las recomendaciones (nombre, dirección, rating, categoría y atributos del sitio).
 - Calcula la matriz de similitud del coseno para encontrar el sitio que mejor se adapte a los parámetros introducidos del usuario.
 - Ordena los índices de los sitios según su similitud con la información que se toma de referencia, de mayor a menor.
 - Devuelve los primeros 5 lugares con mayor similitud a las características deseadas.

De esta manera, se genera una lista de 5 sitios similares, según su ubicación geográfica, su rating y su número de reseñas. Estos sitios pueden ser considerados como recomendaciones para el usuario que busca lugares para visitar en su estado y categoría de interés.

K-vecinos cercanos

El modelo de k-vecinos cercanos es un método de aprendizaje automático supervisado que se basa en la idea de que los objetos similares tienden a estar cerca unos de otros en el espacio de características. Este método consiste en asignar a un nuevo objeto el valor más frecuente entre sus k-vecinos más cercanos, donde k es un número entero positivo que se elige previamente. Para medir la distancia entre los objetos, se puede utilizar cualquier métrica adecuada, como la distancia euclidiana o la distancia de Manhattan. El modelo de k-vecinos cercanos es simple de implementar y puede adaptarse a diferentes tipos de problemas de clasificación o regresión.

El modelo sigue los siguientes pasos para encontrar los sitios más cercanos según el gusto del cliente:

- Filtra los datos según el estado y la categoría introducidos por el usuario.
- Selecciona la información relevante para mostrar las recomendaciones (nombre, dirección, rating, categoría y atributos del sitio).
- Emplea el modelo de k-vecinos para seleccionar los sitios más cercanos según las preferencias de los usuarios.



- Ordena los índices de los sitios según su similitud con la información que se toma de referencia, de mayor a menor.
- Devuelve los primeros 5 lugares más cercanos, según las características deseadas.

De esta manera, se genera una lista de 5 sitios similares, que, según su ubicación geográfica, se encuentran cerca los unos de otros.

Modo de uso

- Búsqueda por gusto: Si se desea recibir recomendaciones basadas en preferencias personales, utiliza la opción 'Búsqueda por gusto'. Para hacerlo, selecciona el estado en el que te encuentras e ingresa la categoría deseada (por ejemplo, comida, hamburguesas, pizza). Una vez que hayas seleccionado, DatumSearch generará una lista de recomendaciones basadas en tus gustos.

A continuación, un ejemplo en el estado de Michigan con preferencias de comida italiana:

Results

Tiramisu Italian Burgers

Address: Tiramisu Italian Burgers, 7599 Clyde Park Ave SW B, Byron Center, MI 49315

Average Rating: 4.8

Categories: Italian restaurant, American restaurant

Attributes: <https://www.google.com/maps/place/data=!4m2!3m1!1s0x8819b5ea8c173e39:0x3e49a36e39936c0c?authuser=-1&hl=en&gl=us>

Mangiamo Italian Grill

Address: Mangiamo Italian Grill, 107 W Michigan Ave, Saline, MI 48176

Average Rating: 4.2

Categories: Italian restaurant

Attributes: <https://www.google.com/maps/place/data=!4m2!3m1!1s0x883cbb40792158a7:0x7cc753b9a595a62?authuser=-1&hl=en&gl=us>

Boomers Pizza

Address: Boomers Pizza, 56300 Grand River Ave, New Hudson, MI 48165

Average Rating: 4.6

Categories: Pizza restaurant, Italian restaurant

Attributes: <https://www.google.com/maps/place/data=!4m2!3m1!1s0x882357ea7c37f555:0x821611c744735535?authuser=-1&hl=en&gl=us>



- Búsqueda por lugares cercanos: Si se prefiere recibir recomendaciones de lugares cercanos, se emplea la opción 'Búsqueda por lugares cercanos'. Para hacerlo, selecciona el estado en el que te encuentras y la categoría que deseas explorar. Una vez que hayas hecho tu selección, DatumSearch generará una lista de recomendaciones de lugares cercanos.

A continuación, un ejemplo en el estado de Florida con preferencias de un local de comida:

Results

Madeleina Cottage Restaurant

Address: 357 Corey Ave

Average Rating: 4.0

Categories: Burgers, Restaurants, Seafood, American (New), American (Traditional)

Attributes: BikeParking, BusinessAcceptsCreditCards, GoodForKids, HappyHour, HasTV, OutdoorSeating, RestaurantsGoodForGroups, RestaurantsReservations, RestaurantsTakeOut

Key West Shrimp Company

Address: 7217 Gulf Blvd, Ste 13

Average Rating: 4.0

Categories: Grocery, Seafood, Do-It-Yourself Food, Restaurants, Food, Specialty Food, Seafood Markets

Attributes: BikeParking, BusinessAcceptsCreditCards, Caters, DogsAllowed, RestaurantsDelivery, RestaurantsTakeOut, WheelchairAccessible

Dockside Dave's Grill

Address: 7141 Gulf Blvd

Average Rating: 4.0

Categories: Restaurants, Seafood

Attributes: BikeParking, BusinessAcceptsCreditCards, GoodForKids, HasTV, OutdoorSeating, RestaurantsGoodForGroups, RestaurantsTakeOut

Categorías recomendadas:

Se recomiendan una lista de categorías enfocadas en la gastronomía para ser utilizadas. DatumSearch se desarrolló con el objetivo de proporcionar información sobre los mejores lugares gastronómicos, los cuales, son esperados a ser descubiertos por intrépidos aventureros gastronómicos. Por lo anterior, DatumSearch así ha sido optimizado para el uso de búsquedas gastronómicas, si bien esta es su característica principal, no está limitado solo a ella, también genera resultados satisfactorios sobre locales como Pubs, Hoteles, entre otros.



En conjunto, DatumSearch es una herramienta poderosa y conveniente que permite al cliente descubrir los mejores lugares para visitar en su estado. Con sus dos modelos de búsqueda, 'Búsqueda por gusto' y 'Búsqueda por lugares cercanos', el cliente puede recibir recomendaciones personalizadas basadas en sus preferencias personales o en lugares más cercanos. Además, con sus instrucciones claras y fáciles de seguir, usar DatumSearch es rápido y sencillo. El cliente puede explorar una amplia variedad de categorías recomendadas, desde comida y bebida hasta alojamiento, para encontrar los lugares más interesantes y emocionantes para visitar. Recomendamos encarecidamente el uso de DatumSearch para descubrir los mejores lugares. Con su amplia variedad de opciones y recomendaciones personalizadas, estamos seguros de que el cliente encontrará muchos lugares nuevos e interesantes para explorar.

Análisis de sentimientos

El análisis de sentimientos es una tarea específica del procesamiento del lenguaje natural que consiste en identificar y extraer las opiniones, emociones y actitudes expresadas en un texto. Esta tarea tiene múltiples aplicaciones prácticas, como el monitoreo de la reputación de una marca, la detección de la satisfacción del cliente o la clasificación de reseñas de productos. Sin embargo, el análisis de sentimientos también plantea varios desafíos, como la ambigüedad, la ironía o el sarcasmo, que pueden dificultar la interpretación correcta de los textos. Para lograr un idóneo análisis sobre los sentimientos de las reseñas se emplearon las siguientes tecnologías:

1. NLTK (Natural Language Toolkit)

La librería NLTK de Python es una herramienta versátil y poderosa para el análisis de sentimientos. NLTK identifica y extrae la opinión o 'sentimiento' de un texto respecto a un tema o entidad. El análisis de sentimientos proporcionado por NLTK puede aplicarse a diversos campos, como los negocios, la política, la educación, la salud, entre otros. NLTK es una de las principales librerías de Python para el procesamiento del lenguaje natural (NLP por sus siglas en inglés) dada su eficacia y facilidad de uso, para tareas como, la tokenización, la etiquetación gramatical, el análisis sintáctico o la generación de texto. Entre sus funcionalidades, se encuentra el módulo `nltk.sentiment`, que contiene clases y métodos para realizar el análisis de sentimientos con diferentes enfoques y técnicas.



Uno de los componentes más destacados de este módulo es el analizador VADER (Valence Aware Dictionary and sEntiment Reasoner), el cual detecta el sentimiento de un texto con base en aspectos como la intensidad, la negación, el énfasis o el contexto. VADER utiliza un diccionario léxico que asigna valores numéricos a las palabras según su polaridad y su intensidad (positiva(1), negativa(-1) o neutra(0)). Además, incorpora reglas gramaticales y sintácticas para ajustar estos valores en función de la estructura del texto.

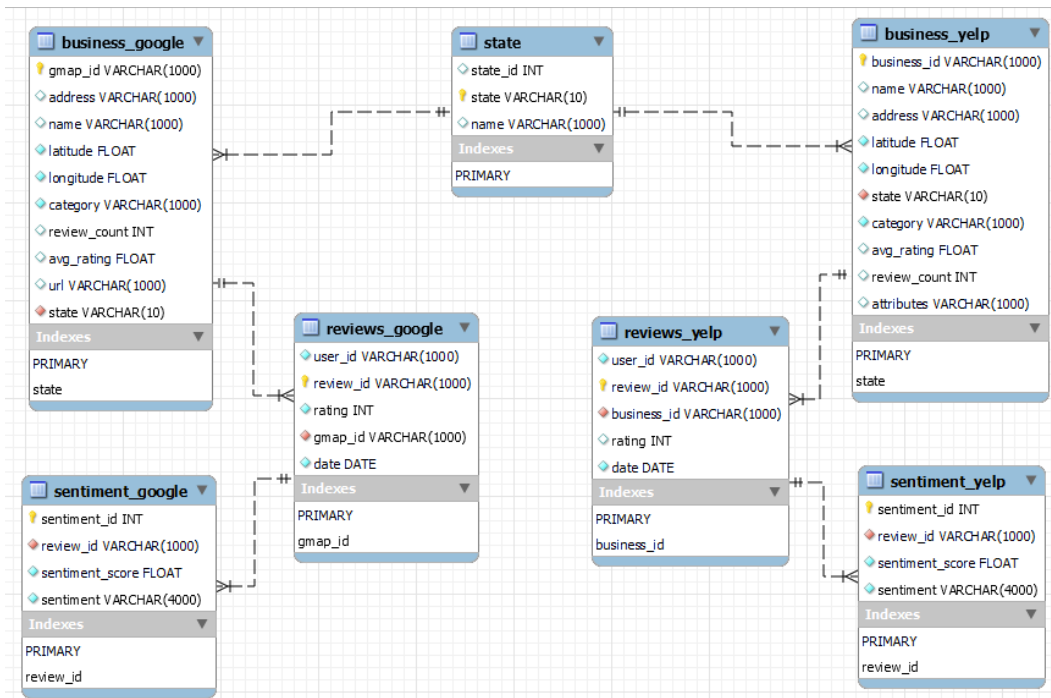
A continuación se muestra un ejemplo del resultado el análisis de sentimientos con NLTK usando Python:

| text | Puntaje de sentimiento | Sentimiento |
|---|------------------------|-------------|
| mixed feelings place. pros- great location. really heart everything. close attractions french quarter. history. many famous people stayed here. price great! sheets soft. saltwater pool amazing spoiled normal pools. staff friendly. cons- wifi bathroom light worked portion time. everything older outdated (carpet, tile, balcony, bedspread, elevator). ice box lower level. coffee makers refrigerators rooms. | 0.9523 | positivo |
| tied best indian philly. delivery mess (containers broken open, food cold) may restaurant's fault. tough say. solid 3-star now. | -0.4588 | negativo |
| manager friendly. staff seems run well there. food fresh, fries, potato cakes hot. food takes minutes asks us seat someone brings food table. never issue here. tend hit grocery shopping buy lot junk store. | 0.765 | positivo |
| must try!! order take rate servers always positive experience. place usually busy go food always hot pick up, someone always greets soon walk place clean welcoming. personally love drunken noodle dish get shrimp beef good portion size make 2 meals plus! husband gets chicken satay coconut rice loves it. definitely recommend place!! | 0.9798 | positivo |
| first glance restaurant seems italian restaurant. careful. menu lot items relationship italian cuisine. white table cloths also lend certain ethnic identity. restaurant food menu ole ole. quick note continue, tommy's warehouse district anyone else. setting gorgeous maintains characteristics buildings area. design decoration help relaxing step above. food truly star. never bad dish specific parts menu truly unique. always looked louisiana restaurant first. ordered turtle soup beet salad along linguini. food service fantastic. third time eaten perfect time. | 0.9809 | positivo |



MODELO DE ENTIDAD RELACIÓN

El modelo entidad relación es una herramienta fundamental para el diseño de bases de datos relacionales, ya que permiten almacenar, organizar y gestionar la información de la organización. Para una empresa dedicada a la gastronomía turística, como lo es Turfood, el modelo entidad relación permite definir las entidades que intervienen en su negocio como, los productos y servicios, las reservas, las facturas, etc. El modelo entidad relación también facilita el identificar los atributos de cada entidad y relación, como por ejemplo, el nombre y dirección del comercio, la valoración y la reseña del servicio, etc. Estos atributos son los que se convertirán en los campos de las tablas de la base de datos y en la base con la que se desarrollarán los kpi, el sistema de reseña y los modelos de machine learning. Además, el modelo entidad relación le permite a la empresa tener una visión clara y estructurada de su información, lo que le ayuda a optimizar sus procesos, mejorar el servicio al cliente, aumentar la rentabilidad y tomar mejores decisiones.



DICCIONARIO DE DATOS

Tabla 1: state

| Nombre | Descripción |
|----------|--|
| state_id | Atributo de identificación única. |
| state | Atributo de la abreviación del nombre del estado o región en donde se encuentra el comercio o establecimiento. |
| name | Atributo del nombre del estado o región en donde se encuentra el comercio o establecimiento. |

Fuente: Elaboración propia

La tabla 1 es una tabla de dimensión que tiene una importancia fundamental en el modelo de entidad relación, ya que permite conectar los dataset de google maps y yelp de manera óptima.

Tabla 2: business_google

| Nombre | Descripción |
|--------------|---|
| gmap_id | Atributo que indica la identificación única que Google Maps le asigna a la dirección física del comercio o establecimiento. |
| address | Atributo que indica la dirección física del comercio o establecimiento. |
| name | Atributo que indica el nombre del comercio o establecimiento que se quiere representar como una entidad en el modelo. |
| latitude | Coordenada de Latitud |
| longitude | Coordenada de Longitud |
| category | Atributo que indica la categoría del comercio o establecimiento. |
| review_count | Número de reseñas que aparecen sobre el comercio o establecimiento, que indican las opiniones o comentarios de los clientes o usuarios que han interactuado con ella. |



| | |
|------------|--|
| avg_rating | Valoración promedio. |
| url | URL de Google Maps que direcciona a la zona geográfica donde se localiza dicho comercio o establecimiento. |
| state | Atributo del nombre del estado o región en donde se encuentra el comercio o establecimiento |

Fuente: Elaboración propia

La tabla 2 en el modelo entidad-relación es de gran importancia, ya que permite organizar y gestionar de manera estructurada la información de los comercios o establecimientos. A través de sus atributos, como el nombre, la dirección, la categoría, las valoraciones y las reseñas, se facilita la búsqueda, el filtrado y la representación geográfica de los comercios. Además, proporciona datos clave para la toma de decisiones, la evaluación de la calidad y la integración con otras funcionalidades del sistema, mejorando la experiencia global del usuario.

Tabla 3: business_yelp

| Nombre | Descripción |
|-------------|---|
| business_id | Atributo identificador único del comercio o establecimiento, que asigna la plataforma Yelp. |
| name | Atributo que indica el nombre del comercio o establecimiento que se quiere representar como una entidad en el modelo. |
| address | Atributo que indica la dirección física del comercio o establecimiento. |
| latitude | Coordenada de Latitud |
| longitude | Coordenada de Longitud |
| state | Atributo del nombre del estado o región en donde se encuentra el comercio o establecimiento |
| category | Atributo que indica la categoría del comercio o establecimiento. |
| avg_rating | Valoración promedio. |



| | |
|--------------|---|
| review_count | Número de reseñas que aparecen sobre el comercio o establecimiento, que indican las opiniones o comentarios de los clientes o usuarios que han interactuado con ella. |
| attributes | Atributos adicionales asociados al negocio. |

Fuente: Elaboración propia

La tabla 3 permite identificar y relacionar de manera precisa y eficiente los diferentes comercios o establecimientos con sus respectivos atributos y características. Además, al incluir la valoración de los usuarios y el número de reseñas, esta tabla proporciona información relevante para la toma de decisiones de los usuarios, así como para el análisis y la gestión de los negocios

Tabla 4: reviews_google

| Nombre | Descripción |
|-----------|---|
| user_id | Atributo de identificación única del usuario en la plataforma. |
| review_id | Atributo de identificación única de identificación de reseña. |
| rating | Valoración del usuario sobre el comercio o establecimiento. |
| gmap_id | Atributo que indica la identificación única que Google Maps le asigna a la dirección física del comercio o establecimiento. |
| date | Atributo que indica la fecha en la que el comentario o reseña fue realizada por el usuario. |

Fuente: Elaboración propia

La tabla 4 es importante porque representa la información clave relacionada con las interacciones de los usuarios con los comercios o establecimientos en la plataforma Google Maps. En conjunto, esta tabla proporciona una visión completa de las interacciones de los



usuarios y sus opiniones sobre los comercios, lo cual es esencial para la toma de decisiones y la mejora de la plataforma.

Tabla 5: reviews_yelp

| Nombre | Descripción |
|-------------|---|
| user_id | Atributo de identificación única del usuario en la plataforma. |
| review_id | Atributo de identificación única de identificación de reseña. |
| business_id | Atributo identificador único del comercio o establecimiento, que asigna la plataforma Yelp. |
| rating | Valoración del usuario sobre el comercio o establecimiento. |
| date | Atributo que indica la fecha en la que el comentario o reseña fue realizada por el usuario. |

Fuente: Elaboración propia

La tabla 5 es de importancia ya que presenta información clave relacionada con las interacciones de los usuarios con los comercios o establecimientos en la plataforma Yelp. En conjunto, esta tabla proporciona una visión completa de las interacciones de los usuarios y sus opiniones sobre los comercios, lo cual es esencial para la toma de decisiones y la mejora de la plataforma.

Tabla 6: sentiment_google

| Nombre | Descripción |
|-----------------|---|
| sentiment_id | Atributo de identificación única del sentimiento en la plataforma. |
| review_id | Atributo de identificación única de identificación de reseña de la plataforma Google Maps.. |
| sentiment_score | Valoración otorgada por el modelo de sentimientos a la reseña. |
| sentiment | Atributo de clasificación del sentimiento. |

Fuente: Elaboración propia



La tabla 6 es de importancia, ya que presenta información clave relacionada con las opiniones de los usuarios respecto de los comercios o establecimientos en la plataforma Google Maps. Proporcionando los diferentes tipos de sentimientos asociados a las diferentes reseñas .

Tabla 7: sentiment_yelp

| Nombre | Descripción |
|-----------------|---|
| sentiment_id | Atributo de identificación única del sentimiento en la plataforma. |
| review_id | Atributo de identificación única de identificación de reseña de la plataforma Yelp. |
| sentiment_score | Valoración otorgada por el modelo de sentimientos a la reseña. |
| sentiment | Atributo de clasificación del sentimiento. |

Fuente: Elaboración propia

La tabla 7 es de importancia porque presenta información clave relacionada con las opiniones de los usuarios respecto de los comercios o establecimientos en la plataforma Yelp. Proporcionando los diferentes tipos de sentimientos asociados a las diferentes reseñas.

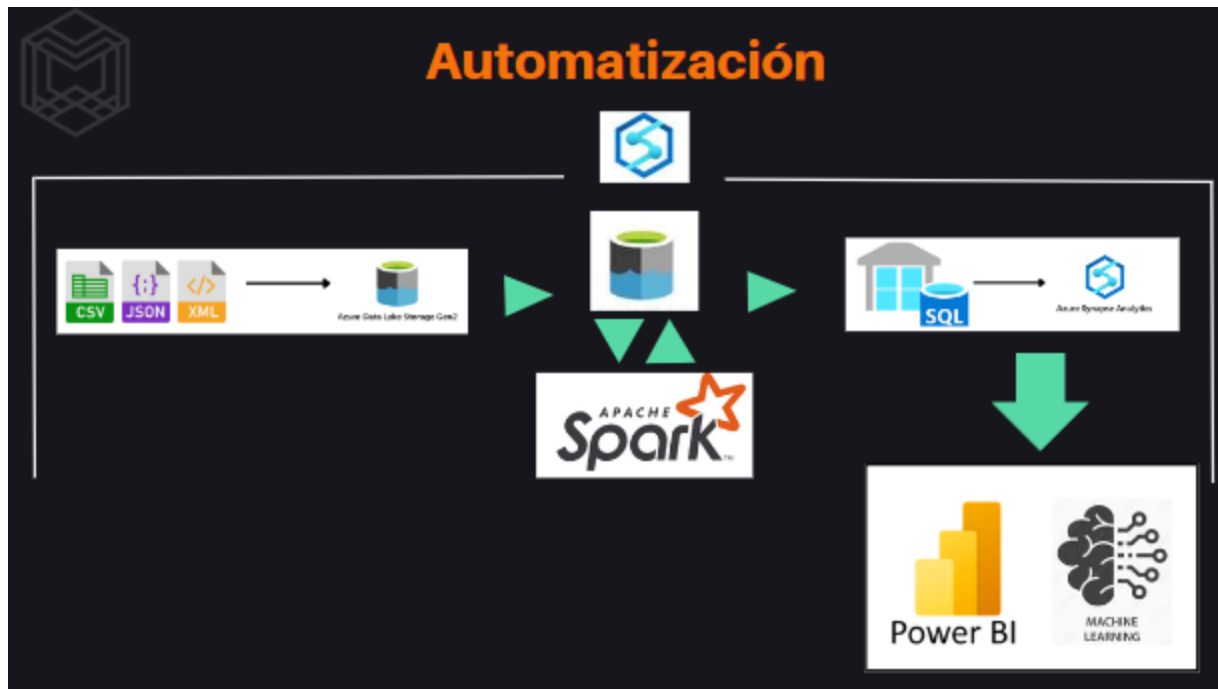
AUTOMATIZACIÓN

Al implementar Azure Synapse Analytics, contamos con la posibilidad de lograr la integración de todos los recursos utilizados a lo largo del desarrollo del proyecto.

A través de las conexiones integradas que puede establecer dicha plataforma con los recursos utilizados en el stack tecnológico, podemos gestionar actividades que involucran tanto al almacenamiento de datos (Azure Data Lake Storage Gen2), procesamiento, transformación y carga de datos (grupo Apache Spark situado en el espacio de trabajo de Synapse Analytics) y a la base de datos (SQL Pool dedicado en el espacio de trabajo de Synapse Analytics). Logrando así integrar desde la ingesta de datos en Azure Data Lake, la activación de la ejecución del proceso de transformación de los datos crudos, la carga incremental de nuevos datos, hasta la disponibilización de los datos ya listos para el consumo de los modelos de machine learning y las correspondientes consultas para generar las visualizaciones en Power BI.



De modo que la capacidad de automatización que proporciona Azure Synapse Analytics como plataforma unificada para la ingesta, preparación, administración y entrega de datos se observa muy beneficiosa.



ESTRATEGIAS PROPUESTAS

Tras un exhaustivo análisis de los datos recopilados, hemos identificado una serie de estrategias que se presentan como posibles soluciones para mejorar el desempeño de los negocios de nuestro valioso cliente, TURFOOD, en el sector de la industria turística. En particular, nos enfocamos en maximizar el impacto y la relevancia del turismo gastronómico en el mercado estadounidense.

Planteo de estrategias con base en los KPIs planteados.

Tasa de Crecimiento de Reseñas:

Al observar que la Tasa de Crecimiento de Reseñas tiene un comportamiento fluctuante en relación con las estacionalidades a lo largo del año calendario, como los períodos vacacionales, o en aquellos meses donde existe una mayor actividad social, como lo es el mes de diciembre, se sugiere:

- El lanzamiento de promociones con el objetivo de incentivar el consumo en dichos períodos. Estas promociones serían compensadas por el aumento del consumo.
- Además, se sugiere la implementación de programas de incentivos para alentar a los clientes a dejar reseñas, como descuentos en futuras visitas, regalos o puntos de fidelidad.
- También se sugiere lanzar promociones y concursos en las redes sociales para animar a los clientes a compartir sus experiencias y dejar reseñas.
- Otra estrategia a tener presente sería la colaboración con bloggers e influencers gastronómicos para aumentar la visibilidad y la cobertura en línea.

Al analizar la cantidad total de reseñas por estado, podemos observar que existe una mayor participación y/o predisposición de los clientes a dejar sus reseñas en determinados estados, como “Florida, Nueva York y California”, porque se sugiere:



-
-
- Tenerlos en cuenta a la hora de recomendar ciertas rutas gastronómicas que los incluyan.
 - También podrían ser tenidos en cuenta a la hora de analizar posibles localizaciones para la apertura de nuevos locales.
 - A su vez, en aquellos estados en donde no se observa una gran participación de los clientes a la hora de dejar sus reseñas, se sugiere reforzar estrategias que impulsen a los clientes a dejar su reseña, con el objetivo de sugerir mejoras en los servicios prestados.

Índice de Satisfacción del Cliente:

Al analizar el Índice de Satisfacción del Cliente, a lo largo de los meses, podemos observar que el mismo se mantiene por encima del 80% en términos generales. Por lo cual se sugiere fuertemente enfatizar en disminuir el Índice de Insatisfacción del Cliente, principalmente si nos enfocamos en la categoría “Hotel”.

Para ello se sugiere:

- Realizar encuestas de satisfacción después de cada experiencia gastronómica para identificar las principales áreas de insatisfacción y trabajar en ellas.
- Establecer programas de capacitación continua para el personal, enfocados en la atención al cliente, la resolución de problemas y la mejora de la experiencia general.
- Ofrecer programas de fidelización para incentivar a los clientes a regresar y mejorar su experiencia.
- Mejorar la calidad de las instalaciones y servicios ofrecidos, como la limpieza, la seguridad, la comodidad y la variedad de opciones.
- Establecer un sistema de gestión de quejas y sugerencias para atender rápidamente cualquier problema o comentario de los clientes.



Rating promedio por categoría:

Al analizar el Rating Promedio por categoría a lo largo de los meses, podemos observar que existen dos tendencias sobresalientes si ponemos atención principalmente en las categorías “Bakery y Hotel”. Si bien, en el caso de la categoría “Hotel” también se observa el menor rating promedio (coincidiendo con el mayor índice de insatisfacción anteriormente descrito), podemos apreciar que a lo largo de los meses, el rating promedio ha ido en aumento. Reforzando aún más las estrategias sugeridas anteriormente. En cuanto a la Categoría “Bakery”, observamos lo contrario, a lo largo de los meses el rating promedio ha ido en disminución. A pesar de haber comenzado con un buen rating promedio. Por lo que algunas sugerencias propuestas son:

- Implementar programas de mejora específicos para la categoría bakery, como capacitación adicional para el personal y mejoras en la calidad de los ingredientes o en la presentación de los productos.
- Adicionar líneas de productos gluten free y veganos, ya que las tendencias en el modo de alimentación impactan directamente sobre este sector gastronómico.
- Solicitar retroalimentación directa a los clientes en relación con los productos ofrecidos para tomar acciones correctivas.
- Destacar y promover la categoría “coffee”, a través de estrategias de marketing y publicidad.

Tasa de Reseñas Positivas:

Al observar la Tasa de Reseñas Positivas, podemos observar que, en términos generales, ha ido disminuyendo, principalmente en el período desde el mes de mayo hasta el mes de septiembre (período vacacional de verano). Por lo que se sugiere:

- Destacar las reseñas positivas existentes, mostrando las reseñas positivas que ya cuenta el local en su sitio web, redes sociales u otros canales de marketing. Resaltar testimonios y comentarios positivos de clientes satisfechos. Ya que de esta forma se podría generar confianza en los clientes potenciales y aumentar la probabilidad de que también dejen una reseña positiva.



-
-
- Implementar programas de respuesta lo más rápido posible a las reseñas que han dejado los clientes.
 - Lanzar preguntas abiertas en los espacios de interacción de las distintas redes sociales.
 - Agradecer siempre todo tipo de reseñas.

Tasa de Crecimiento de Nuevos Clientes:

Al observar la Tasa de Crecimiento de Nuevos Clientes, también podemos observar un comportamiento fluctuante. Que, como sucede con la Tasa de Crecimiento de Reseñas, estaría en directa relación con los períodos vacacionales.

Los estados en los que se observa mayor cantidad de clientes, coincide con aquellos estados donde se percibe una mayor participación activa de los clientes en cuanto al número de reseñas. Y en cuanto a la categoría con mayor cantidad de clientes asociados, se observa la categoría “Restaurant”. Probablemente, podría asociarse a que generalmente una comida en un restaurante nuevo se asocia a una nueva experiencia y el disfrute de un momento de relax y comunión. Por el contrario, la categoría en la cual se observa el menor número de clientes es “coffee”, porque podría enfatizarse en reforzar las estrategias sobre este sector.

Algunas sugerencias son:

- Participación en eventos y ferias comerciales para promocionar el producto o servicio y establecer relaciones con nuevos clientes potenciales.
- Establecer alianzas estratégicas con otras empresas o marcas complementarias para llegar a nuevos públicos y aumentar la visibilidad de la marca.
- Ofrecer promociones y descuentos para atraer a nuevos clientes y motivarlos a probar el producto o servicio.
- Ampliación de la carta ofrecida, con períodos de renovación de los productos.
- Presencia activa en redes sociales y colaboraciones con influencers y bloggers gastronómicos.
- Ofrecer promociones y descuentos exclusivos para nuevos clientes para incentivar su primera visita.

