

Proyecto de Inversión Gastronómica y Afines en Mercado De Estados Unidos



Integrantes

Víctor Vargas
Guillermo del Rio
Michael Martinez Chinchilla
Julian Scarpeccio
Benjamin Zelaya

Índice

Índice	1
1 Entendimiento de la situación propuesta	2
2 Objetivos	2
2.1 objetivos específicos	2
2.2 Objetivos del grupo	2
3 Alcance y fuera de alcance	3
3.1 Alcance	3
3.2 Fuera de alcance	3
4 Análisis preliminar de la calidad del dato	3
5 KPI y Métricas Asociadas	3
6 Solución Propuesta	4
6.1 Medidas Estratégicas	5
6.2 Medidas Operativas	5
7 Metodología Propuesta	5
7.1 Cronograma General (GANTT)	6
8 Stack tecnológico	6
8.1 Data Pipeline	7
8.1.1 Diagrama flujo de dato	7
8.1.2 Data Ingest	7
8.1.3 Conexión con Databricks	7
8.1.4 ETL	7
8.1.5 Conexión con sql database	8
8.1.6 Conexión con Power BI	8
9 Roles Y Responsabilidades	8
10 Creación del repositorio en Github	8

1. Entendimiento de la situación propuesta

El cliente, que forma parte de un conglomerado de empresas en el sector de restaurantes y afines, busca realizar un análisis exhaustivo de las opiniones de los usuarios en las plataformas Yelp y Google Maps. Este análisis se centrará en hoteles, restaurantes y otros negocios relacionados con el turismo y el ocio. El cliente también está interesado en identificar ubicaciones estratégicas para abrir nuevos locales de restaurantes y negocios relacionados, así como desarrollar un sistema de recomendación de restaurantes que permita a los usuarios descubrir nuevas experiencias basadas en sus preferencias previas.

Para llevar a cabo este proyecto, el cliente ha contratado a Icon Data Science Consulting como consultores externos. La tarea principal es estudiar la relación entre las reseñas de diferentes tipos de negocios, que incluyen restaurantes, hoteles, y otros servicios. Además, se analizarán datos como la cantidad de locales de restaurantes, sus nombres y franquicias, así como el tipo de oferta gastronómica que ofrecen. El objetivo es evaluar la viabilidad de introducir una propuesta gastronómica completamente nueva o incorporar una franquicia existente en ubicaciones estratégicas.

Se realizará un análisis preliminar de la cantidad de ofertas gastronómicas por tipo en un conglomerado específico de estados, que fueron estratégicamente seleccionados para obtener un marco de referencia y tomar decisiones bien fundamentadas.

El proyecto implica recopilar, depurar y disponibilizar información relevante de diferentes fuentes para analizar la relación entre la cantidad de locales de restaurantes, nombres de cada uno de ellos, franquicias y tipo de oferta gastronómica. A través de reportes, dashboards y el entrenamiento de un modelo de machine learning de clasificación, se resolverá una demanda y/o oportunidad de inversión específica relacionada con los objetivos del proyecto.

2. Objetivos

2.1. objetivos específicos

- Recopilar y depurar datos de diferentes fuentes para crear una base de datos (DataWarehouse).
- Realizar un análisis exploratorio de los datos para encontrar relaciones.
- Crear un dashboard interactivo y visualmente atractivo que integre los resultados del análisis exploratorio de datos.
- Entrenar y poner en producción un modelo de machine learning para proponer una oportunidad de inversión en el sector.

2.2. Objetivos del grupo

- Investigar y analizar los conjuntos de datos de locales gastronómicos y sus franquicias en cada uno de los conglomerados para identificar tendencias.
- Entender la correlación entre los tipos de locales gastronómicos, franquicias y reviews con el fin de proporcionar una base sólida para las decisiones futuras relacionadas con la oportunidad de inversión.
- Proveer información relevante y confiable a la empresa para respaldar su toma de decisión sobre el emplazamiento de un nuevo local comercial

3. Alcance y fuera de alcance

3.1. Alcance

- Recopilación, limpieza y análisis de datos de restaurantes, su tip, franquicias, ubicaciones, población y reseñas estratificando en 3 grupos de estados su densidad poblacional
 - Alta
 - Media
 - Baja

3.2. Fuera de alcance

- La implementación real del costo del emplazamiento del restaurante y/o franquicia en cuanto a construcción, diseño, plazo de entrega de obra, contratación de personal. El proyecto se enfoca en proporcionar información y análisis para respaldar la decisión de invertir en una nueva oferta gastronómica y/o franquicia, pero la implementación concreta queda bajo responsabilidad de la empresa/cliente.
- Análisis exhaustivos de aspectos económicos, logísticos y regulatorios asociados con la habilitación, obligaciones impositivas locales y de ese índole.

4. Análisis preliminar de la calidad del dato

Se realizó un análisis preliminar y transformaciones de los conjuntos de datos dados por la empresa y obtenidos externamente para así observar la calidad de los datos, concatenar archivos, limpiar valores nulos, observar tipos de datos y realizar gráficos que nos ayuden a comprender la información.

Podemos observar los archivos dentro de la carpeta sprint_1 llamado ETL y EDA.

5. KPI y Métricas Asociadas

- **Tasa de Satisfacción del Cliente:**
 - **Definición:** Este KPI mide el porcentaje de clientes satisfechos en función de las reseñas y opiniones recopiladas en plataformas como Yelp y Google.
 - **Fórmula:** Tasa de Satisfacción del Cliente = (Número de Reseñas Positivas (valoración ≥ 4 estrellas) / Total de Reseñas) x 100
 - **Objetivo:** Evaluar la satisfacción general de los clientes con las franquicias de restaurantes en cada categoría de estados.
- **Cantidad de Sucursales por Conglomerado de Estados:**
 - **Definición:** Este KPI permite identificar cuántas sucursales tendrás por conglomerado de estados.

- **Objetivo:** Determinar la expansión y presencia de las franquicias en cada categoría de estados, lo que puede indicar oportunidades de crecimiento o saturación del mercado.
- **Porcentaje de Restaurantes con Alta Calificación:**
 - **Definición:** Este KPI muestra el porcentaje de restaurantes que tienen una calificación por encima de un umbral específico (por ejemplo, 4 estrellas).
 - **Fórmula:** $\text{Porcentaje de Restaurantes con Alta Calificación} = (\text{Número de Restaurantes con Rating} \geq \text{Umbral}) / \text{Total de Restaurantes} \times 100$
 - **Objetivo:** Evaluar la calidad general de las franquicias en cada categoría de estados y destacar las que tienen una alta calificación.
- **Top 5 Franquicias por Conglomerado de Estados:**
 - **Definición:** Este KPI identifica las mejores franquicias por conglomerado de estados.
 - **Objetivo:** Identificar las franquicias más exitosas en términos de presencia y calificación en cada categoría de estados, lo que puede servir como referencia para futuras inversiones.
- **Tasa de Penetración de Mercado por Estado:**
 - **Definición:** Este KPI mide la cuota de mercado que tienen las franquicias en cada uno de los tres tipos de estados: densamente poblados, de población intermedia y de menor población. Indica la proporción de restaurantes de franquicias con respecto al total de restaurantes en cada estado.
 - **Objetivo:** Identificar dónde las franquicias tienen una presencia más sólida y evaluar su participación en el mercado en función de la densidad de población.
 - **Fórmula:** $\text{Tasa de Penetración de Mercado por Estado} = (\text{Número de Restaurantes de Franquicias} / \text{Total de Restaurantes}) \times 100$.

6. Solución Propuesta

Teniendo en cuenta que la empresa es la encargada de tomar la última decisión en administrar sus recursos, se propone los siguientes pasos para que pueda invertir en el sector de restaurantes y afines.

- Creación de dashboards y reportes visuales para presentar la información de manera clara y concisa a quienes toman decisiones en la empresa.
- Desarrollar un modelo de machine learning para predecir la viabilidad de implementar nuevos locales comerciales en los diferentes estados.
- Generar un plan de acción sugerido de acuerdo a las interpretaciones del estudio, con medidas estratégicas y operativas si es necesario.

6.1. Medidas Estratégicas

- **Selección de Ubicaciones Estratégicas:** Basándose en los hallazgos del estudio y las recomendaciones del modelo de machine learning, la empresa debería definir las ubicaciones específicas donde planea abrir nuevos locales comerciales. Esto incluiría identificar ciudades, barrios o áreas precisas.
- **Estrategia de Marca y Concepto:** Definir la estrategia de marca y el concepto para los nuevos locales. Esto implica decidir si se introducirá una nueva oferta gastronómica o se incorporará una franquicia existente, así como definir la propuesta de valor y el enfoque de marketing.
- **Inversión en Marketing y Publicidad:** Planificar una estrategia de marketing sólida para promocionar los nuevos locales. Esto podría incluir campañas publicitarias, presencia en redes sociales, programas de fidelización de clientes y eventos promocionales.

6.2. Medidas Operativas

- **Adquisición de Bienes Raíces y Permisos:** Iniciar la búsqueda y adquisición de propiedades o espacios comerciales en las ubicaciones estratégicas identificadas. Asegurarse de obtener los permisos y licencias necesarios para operar.
- **Diseño y Construcción:** Si es necesario, llevar a cabo el diseño y la construcción de los locales comerciales. Esto incluiría la planificación de la distribución, la decoración y la infraestructura necesaria.
- **Selección de Proveedores y Abastecimiento:** Establecer acuerdos con proveedores de alimentos y bebidas, equipos de cocina y otros suministros necesarios. Asegurarse de mantener una cadena de suministro eficiente.

7. Metodología Propuesta

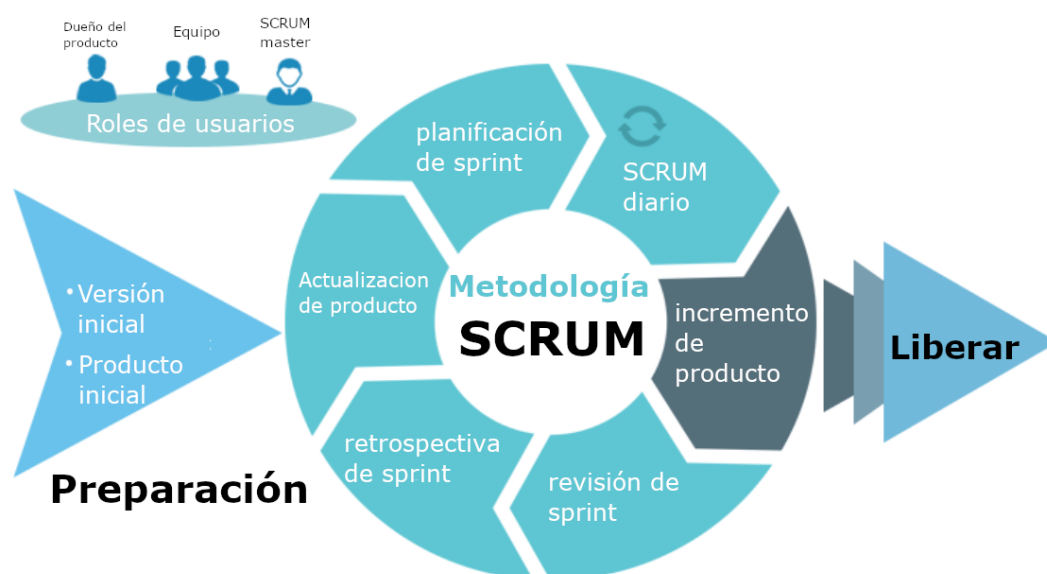


Figura 1: Metodología SCRUM

En este proyecto, estamos empleando la metodología ágil mediante el marco de trabajo Scrum utilizando el software de administración de proyectos trello. Esto conlleva la incorporación de los valores y conceptos ágiles en nuestro enfoque de desarrollo. Estamos aplicando la estructura definida por Scrum en términos de roles, eventos, artefactos y reglas para la organización y gestión efectiva del trabajo colaborativo.

7.1. Cronograma General (GANTT)



Figura 2: Tareas pendientes

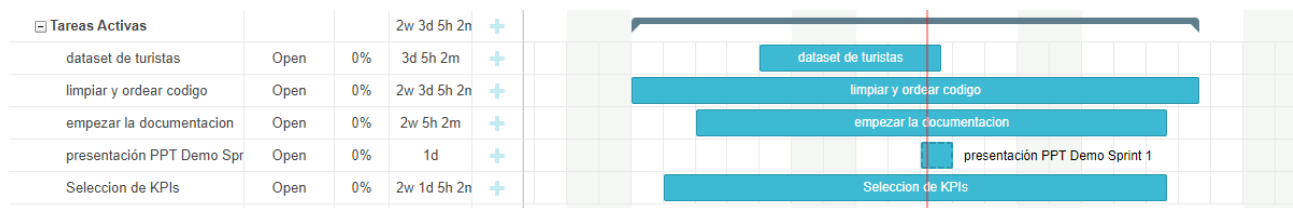


Figura 3: Tareas en proceso

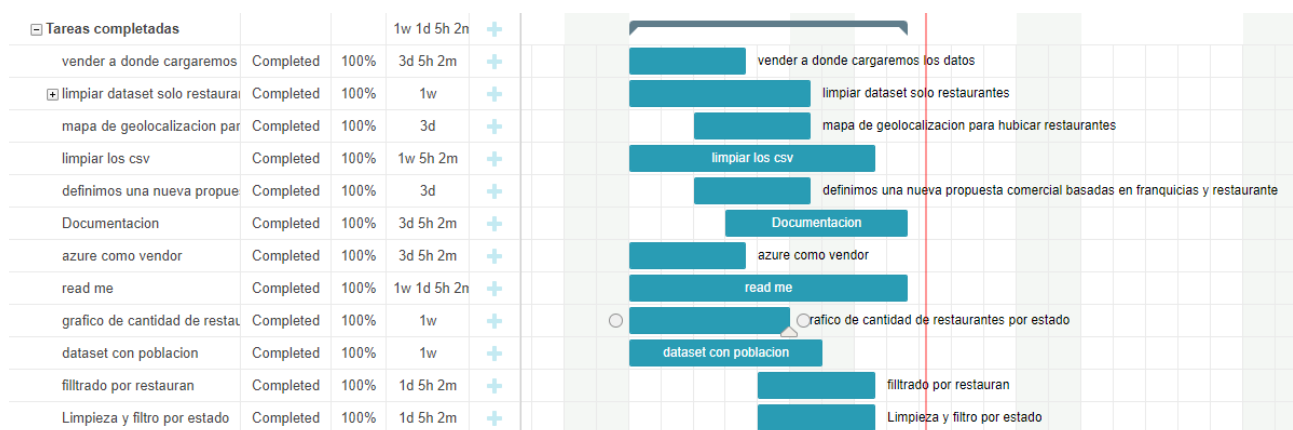


Figura 4: Tareas completadas

8. Stack tecnológico

Para llevar a cabo nuestro proyecto hemos seleccionado las siguientes tecnologías:

- **Trabajo diario:** python, google meet, github.
- **Ingeniería de datos:** Microsoft Azure, Python, sql.
- **Análisis y visualización de datos:** Power Bi, python.
- **Modelo de machine learning:** Python.
- **Gestión de proyectos:** Trello.

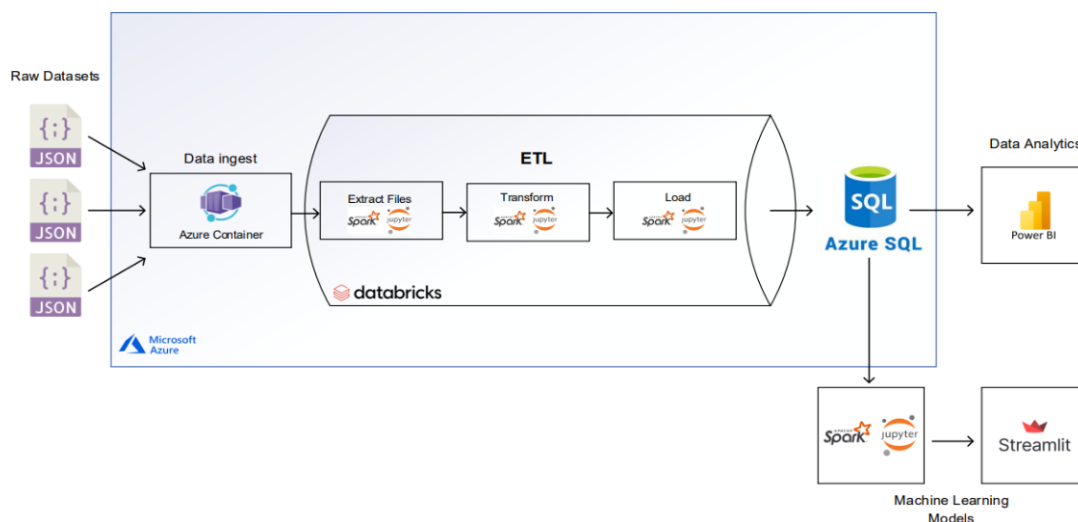


Figura 5: Diagrama de flujo

8.1. Data Pipeline

8.1.1. Diagrama flujo de dato

En esta sección se estructurará el flujo del dato desde la recepción hasta la salida del ETL.

8.1.2. Data Ingest

Los datos entregados por la empresa y extraídos por nuestro equipo mediante api y web scraping se descargaron y son almacenados de manera temporal en el localhost de nuestra máquina.

Dado que trabajaremos sobre el esquema de Microsoft Azure se creará un contenedor donde se almacenarán los datasets sin procesar en la nube. Para esto, fue necesario crear una cuenta de trabajo en el portal de Azure. En dicha cuenta se crea un grupo de recursos donde incluimos una cuenta de almacenamiento con un contenedor.

8.1.3. Conexión con Databricks

Una vez almacenados los datasets en el contenedor de Azure se procede a realizar la conexión con Databricks, nuestro lugar de trabajo principal.

En el grupo de recursos previamente creado se añade un workspace de Databricks. Ahí se creará un clúster que permite computar nuestros datos (Single Node 10.4 LTS Apache Spark 14 GB Memory, 4 Cores), el criterio de selección es en base al alcance de nuestros recursos.

Dentro de Databricks creamos un Notebook y lo conectamos con el clúster. En dicho Notebook establecemos las variables necesarias para la conexión con el contenedor.

8.1.4. ETL

Se realizará todo el proceso de extracción, transformación y carga de los datos hacia el data warehouse.

8.1.5. Conexión con sql database

Creada la SQL Database de Azure se realizará la conexión con Databricks por medio del protocolo jdbc.

8.1.6. Conexión con Power BI

La conexión se realiza mediante el conector de Azure SQL Database de PowerBI. Se ingresan las credenciales del servidor de base de datos y se cargan los datos ya sea por Direct Query o Import Data.

9. Roles Y Responsabilidades

Víctor Vargas → Data Engineer

Guillermo del Rio → Data Science

Michael Martinez → Data Engineer

Julian Scarpeccio → Data Science

Benjamin Zelaya → Data Analyst

10. Creación del repositorio en Github

Para poder visualizar los trabajos de forma remota y conciliar la evolución de versiones, se adoptó la tecnología git , por lo tanto se creó un repositorio colaborativo en la plataforma github, donde se podrá crear, modificar y visualizar los contenidos que se irán requiriendo a lo largo del proyecto.

[REVIEWS AND RECOMMENDATIONS](#)