



# **Proyecto de Inversión Gastronómica y Afines en Mercado De Estados Unidos**

## **INTEGRANTES:**

**Víctor Vargas**

**Guillermo del Rio**

**Michael Martinez Chinchilla**

**Julian Scarpeccio**

**Benjamin Zelaya**

## **CONTENIDO DEL PROYECTO**

- 1. Entendimiento de la situación propuesta**
- 2. Objetivos del trabajo y del grupo**
- 3. Alcance y fuera de alcance**
- 4. Análisis preliminar de Calidad del dato**
- 5. Kpi y métricas asociadas**
- 6. Solución Propuesta**
- 7. Metodología Propuesta**
- 8. Definición del stack tecnológico**
- 9. Equipo de trabajo (Roles y responsabilidades)**
- 10. Creación del repositorio en Github**



## **Proyecto de Inversión Gastronómica y Afines en Mercado De Estados Unidos**

### **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**

### **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.

### **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**

### **Alcance:**

- La recopilación, limpieza y análisis de datos de restaurantes, tipos de restaurantes, franquicias, Estados, población, reviews para extraer información relevante.
- El análisis exploratorio de datos para identificar patrones preliminares y tendencias en los datos recopilados.
- La correlación entre los datos de tipos de locales gastronómicos, franquicias y reviews.
- Implementación de métricas e indicadores para evaluar diferentes aspectos de las ofertas gastronómicas y sus franquicias existentes, y su relación con los clientes, reviews, cantidad de restaurantes por estados.
- Se requerirá la incorporación de datasets adicionales que contengan información relevante y confiable.

### **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:**

Este KPI mide el porcentaje de clientes satisfechos en función de las reseñas y opiniones recopiladas en plataformas como Yelp y Google Maps.

Tasa de Satisfacción del Cliente = (Número de Reseñas Positivas / Total de Reseñas) x 100

- **Rentabilidad por Ubicación:**

Permite identificar cuáles son las ubicaciones más rentables y cuáles pueden requerir mejoras.

$$\text{Rentabilidad por Ubicación} = (\text{Ingresos} - \text{Costos}) / \text{Ingresos} \times 100$$

- **Porcentaje de Restaurantes con Alta Calificación:**

Este KPI muestra el porcentaje de restaurantes que tienen una calificación por encima de un umbral específico (por ejemplo, 4 estrellas). La fórmula es:

$$\text{Porcentaje de Restaurantes con Alta Calificación} = (\text{Número de Restaurantes con Rating} \geq \text{Umbral}) / \text{Total de Restaurantes} \times 100$$

- **Diferencia Promedio de Rating por Estado:**

identificar si la calidad de los restaurantes varía significativamente de un estado a otro.

$$\text{Diferencia Promedio de Rating por Estado} = \text{Promedio de "Rating" agrupado por "Estado"}$$

## **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.

## 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.

## 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



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.

## Cronograma General (GANTT)

## 8. Stack tecnológico

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

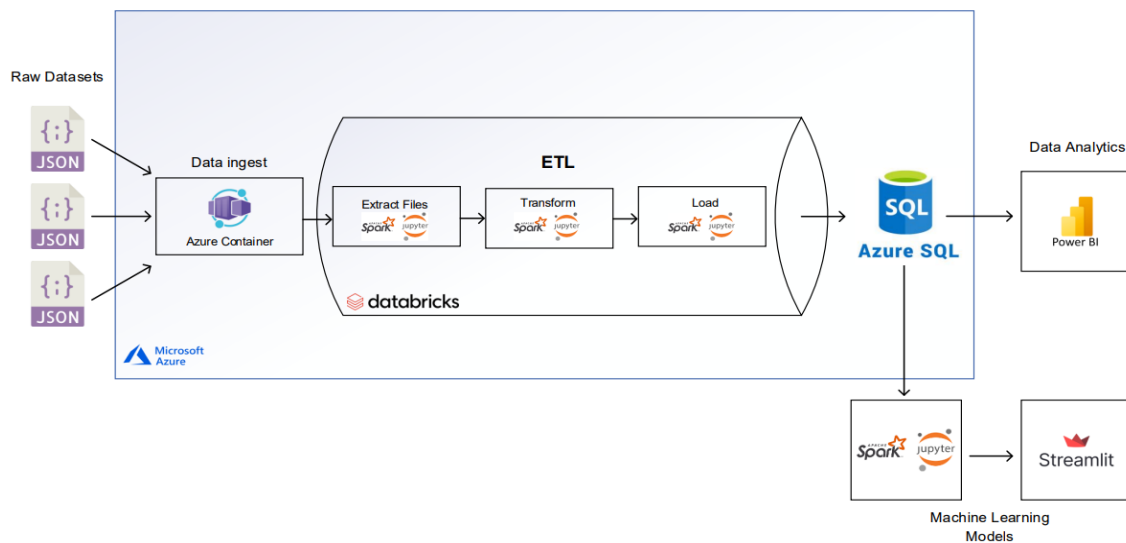
- **Trabajo diario:** python, google meet, github.
- **Ingeniería de datos:** Microsoft Azure, Python, sql.



- **Análisis y visualización de datos:** Power Bi, python.
- **Modelo de machine learning:** Python.
- **Gestión de proyectos:** Trello.

## Data Pipeline:

### Diagrama flujo de dato



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

### 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.

## **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.

## **ETL**

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

## **Conexión con sql database**

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

## **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.

**link:**

[https://github.com/gdelrio0410/PF\\_DS\\_REVIEWS\\_AND\\_RECOMMENDATIONS](https://github.com/gdelrio0410/PF_DS_REVIEWS_AND_RECOMMENDATIONS)