

## PROPUESTA DE PROYECTO

### RESTAURANTES EN EL ESTADO DE FLORIA, ESTADOS UNIDOS



Ornaldo Hernández Ramos  
Cristian Ignacio Papini  
Aaron Bernardo Martínez Vera  
Jhoeliel David Palma Salazar  
Janice Rocío Rico Sánchez

# ÍNDICE

Introducción .....	03
Entendimiento de la Situación Actual .....	04
Objetivos .....	05
Alcance .....	06
KPIs .....	07
Repositorio de GitHub .....	08
Solución Propuesta:	
Stack Tecnológico .....	09
Metodología de Trabajo .....	10
Cronograma General - Diagrama de Gantt .....	11
Equipo de Trabajo – Roles y Responsabilidades .....	12
Diseño Detallado – Entregables .....	13
Análisis Preliminar de Calidad de Datos .....	14

## INTRODUCCIÓN

Al estar en la era de la información y donde los datos más importantes y verídicos son los que provienen directamente de las personas, los reviews que se realizan a los restaurantes a lo largo y ancho de los Estados Unidos arrojan puntos importantes a ser tomados en cuenta para conocer las preferencias de los clientes potenciales, así como información valiosa a la hora de aperturar nuevos locales de comida.

Es por ello que se debe realizar un análisis de este mercado estadounidense, para apoyarse en la opinión de quienes tienen la experiencia de ir a estos lugares y degustar lo que ofrecen, sirviendo como ruta de conocimiento para quienes se disponen luego a hacer lo mismo.

El poder de las reseñas de los clientes deja un abanico de posibilidades a la hora de tomar decisiones no sólo para otros usuarios, sino también para los empresarios interesados en esta área comercial.

Google Maps ofrece gran cantidad de información acerca de restaurantes que detalla principalmente su localización, atributos y categorías. Asimismo se recogen las opiniones de los usuarios acerca de estos lugares, lo cual marca un hilo aprovechable para ser tomado en cuenta en posteriores soluciones.

## ENTENDIMIENTO DE LA SITUACIÓN ACTUAL

Cuando se habla de temas de turismo, esparcimiento y ocio, gran parte de las búsquedas que se realizan se concentran en el área de la comida! Además de buscar preferencias habituales, los comensales también quieren tener experiencias nuevas, no sólo cuando viajan sino también alrededor de la zona donde habitan, es por ello que contar con información completa y suficiente para tomar decisiones de dónde comer, se vuelve una necesidad.

Del mismo modo, los empresarios cada vez más están tomando en consideración un estudio de mercado para realizar sus inversiones, basado en información que proviene directamente de fuentes que se encuentran activas, haciendo uso de los servicios de restaurantes y que se mueven de un lugar a otro buscando conocer nuevos lugares.

Debido a lo planteado anteriormente es que decidimos armar herramientas que ayuden tanto a comensales como a empresarios a cumplir sus objetivos respecto al área de los restaurantes, específicamente en el Estado de Florida, Estados Unidos, colocando a disposición de los primeros una aplicación informativa e interactiva, de acuerdo a datos de partida introducidos por ellos mismos, para que obtengan recomendaciones puntuales, y para los segundos, análisis de mercado y aplicación de modelos de Machine Learning que les muestre dónde aperturar nuevos lugares de comida.

## OBJETIVOS

### General:

- Proveer información veraz y puntual de restaurantes en el Estado de Florida, Estados Unidos, a comensales y empresarios, para que tomen decisiones más acertadas, cada uno en su área.

### Específicos:

- Construir un Data Warehouse eficiente e integral, con data confiable.
- Analizar la opinión de los usuarios en Google Maps.
- Analizar las locaciones de los restaurantes con mejores ratings.
- Extraer KPIs para evaluar el desempeño de las herramientas generadoras de información.
- Dar recomendaciones a comensales y empresarios, tomando como base el resultado de los estudios previos.
- Presentar una herramienta interactiva para comensales que busquen recomendaciones de restaurantes.
- Conformar dashboards con análisis de datos relacionales para que los empresarios identifiquen oportunidades de negocio en el ámbito de los restaurantes.

## **ALCANCE**

A través de este Proyecto se aplicarán procesos de EDA y ETL a los datos suministrados de sitios de comida y reviews de usuarios de Google Maps, y se complementarán con otros obtenidos de diversas fuentes, logrando así una plataforma interactiva para que clientes puedan consultar opciones de restaurantes en el Estado de Florida, Estados Unidos, de acuerdo a datos clave introducidos.

De igual forma se generarán dashboards informativos para empresarios interesados en invertir en la categoría de los lugares de comida, con cruce de datos que darán como resultado cifras y gráficos entendibles y valiosos para la toma de decisiones.

## KPIs

### **Count Reviews (por categoría) / Date:**

En qué categoría voy a tener más información sobre la cual apoyar mi decisión? Más información = Más capacidad de decisión.

En qué categoría es que tengo más apoyo/seguimiento por parte de los clientes? Mayor contacto con el cliente = Mayor capacidad de establecer lo que al cliente le gusta o no.

### **Rating (por categoría) / Date:**

Qué categoría tiene mejor reputación?

Qué categoría es la más aclamada por la gente? No quisiera apostar por categorías poco reconocidas por el público.

### **Rating (por Estado) / Date:**

En qué Estado sería más conveniente abrir mi lugar, basado en la reputación de los ya existentes? No quisiera un público demasiado exigente con los mismos.

### **Count Reviews (por día) / Date:**

En qué días de la semana me conviene abrir el lugar?

Es conveniente tener abierto los 7 días de la semana? Más allá de que eso se define a través de un análisis económico-financiero, con la cantidad de reviews se puede tener una idea de cómo es el flujo a lo largo de la semana.

### **Rating (por precio) / Date:**

Qué tipo de mercado es el más aclamado por la gente: Económico, intermedio o de alto nivel?

Esto me sirve para, más allá de las ideas que pueda tener acerca del negocio, adaptar mi idea a la demanda actual del mercado.

## REPOSITORIO DE GITHUB

Para una mejor organización de los elementos utilizados y generados a lo largo de este proyecto, se disponibilizó un Repositorio en GitHub en la siguiente ubicación:

[https://github.com/janicerico/PG\\_Google\\_Maps.git](https://github.com/janicerico/PG_Google_Maps.git)



## **SOLUCIÓN PROPUESTA: STACK TECNOLÓGICO**

Actividades Diarias: WhatsApp Group, Google Meet, Google Docs, GitHub.

Ingeniería de Datos: Python (Pandas, Glob, Duckdb, Time, Polars).

Análisis y Visualización de Datos (comensales): Python (Streamlit).

Deploy (empresarios): AWS (SageMaker, Comprehend, Glue Databrew, S3, DynamoDB, RDS).

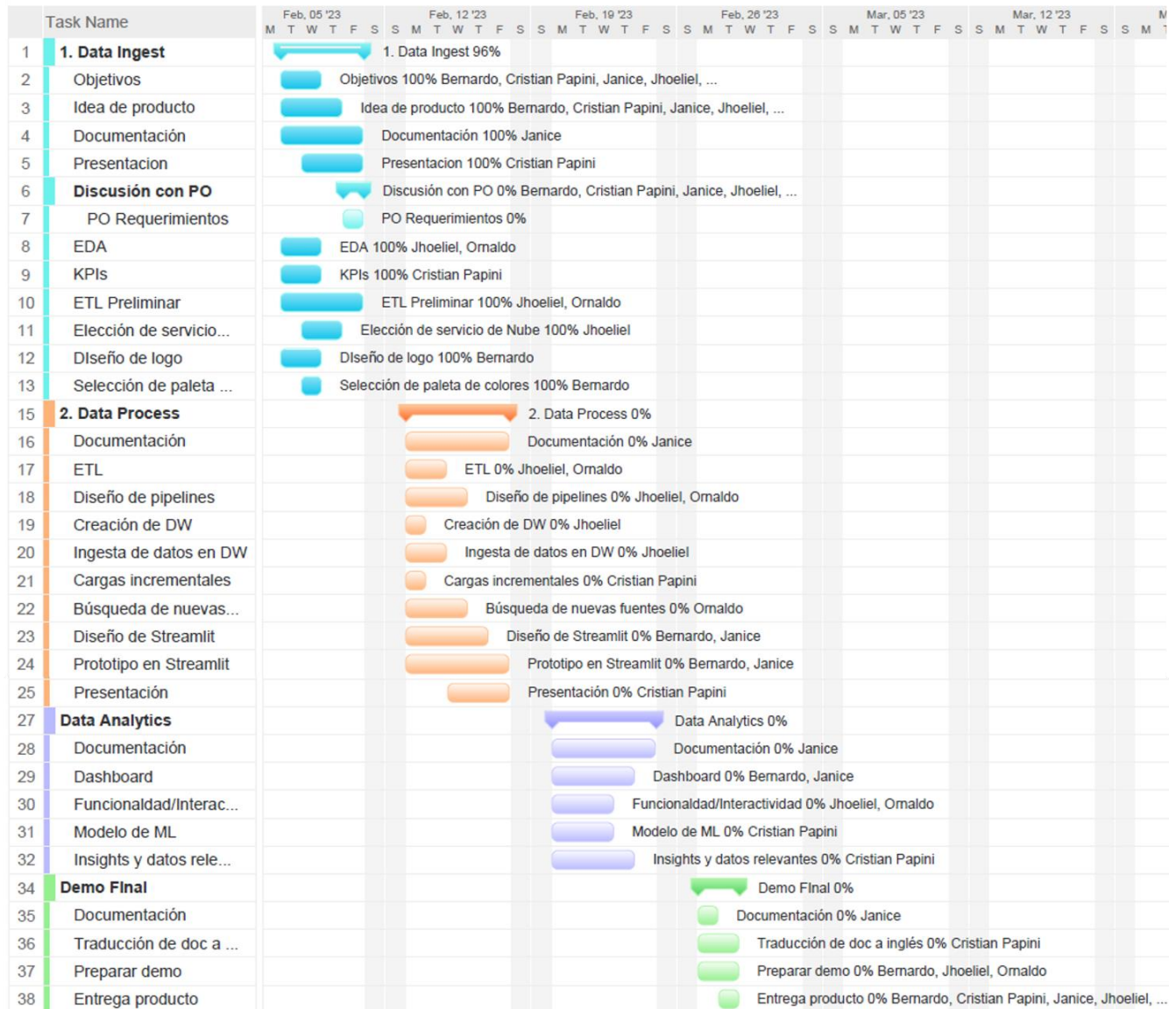
## **METODOLOGÍA DE TRABAJO**

Para la ejecución de este Proyecto, desde los primeros pasos de análisis, hasta el producto final, se estará llevando a cabo una Metodología Ágil de tipo Scrum, con dailies del equipo de trabajo donde se discuten los progresos del día anterior y las tareas para el actual.

Semanalmente se le reportarán al Product Owner los avances y metas logradas, así como recibir su retroalimentación.

Se establecieron metas para un mes (generales), por cada semana, y diarias. Para ello, se elaboró un Diagrama de Gantt, con el rol de cada miembro del equipo y los tiempos estimados para su ejecución.

# CRONOGRAMA GENERAL - DIAGRAMA DE GANTT



## **EQUIPO DE TRABAJO – ROLES Y RESPONSABILIDADES**

Los diferentes integrantes del equipos tendrán diferentes roles de acuerdo a sus habilidades y decisión personal de trabajar en cierta área, lo que aumentará la eficiencia de las tareas a realizar durante la ejecución del mismo. Es así como cada uno reportará de sus dudas y resultados a los demás miembros, de modo que se cumplan los objetivos planteados y los deadlines exigidos.

El trabajo quedará repartido de la siguiente manera:

Data Engineers: Jhoeliel Palma y Orinaldo Hernández.

Data Analytics: Aaron Martínez y Janice Rico.

Machine Learning: Cristian Papini.

Redacción de Informes: Janice Rico.

## **DISEÑO DETALLADO – ENTREGABLES**

Cada uno de los archivos de trabajo y visualizaciones serán almacenados en un Repositorio de GitHub, el cual contendrá:

- Jupyter Notebooks para los procesos de EDA, y ETL automatizado para la carga y tratamiento de los Datasets.
- Datasets finales a partir de los que se generarán las visualizaciones y análisis.
- Plataforma interactiva para realizar las consultas referentes a Restaurantes recomendados, por parte de los comensales.
- Dashboards con herramientas de AWS, incluyendo los KPIs generados, para los empresarios.

## ANÁLISIS PRELIMINAR DE CALIDAD DE DATOS

Los Datasets suministrados están divididos en varias carpetas con la siguiente distribución:

<b>Carpeta</b>	<b>Sub-Carpeta</b>	<b>Cantidad Archivos .json</b>	<b>Peso (Mb)</b>
reviews-estados	review-Alabama	12	467,90
	review-Alaska	4	142,30
	review-Arizona	14	636,00
	review-Arkansas	16	611,80
	review-California	18	762,90
	review-Colorado	16	751,50
	review-Connecticut	18	673,60
	review-Delaware	7	235,00
	review-District_of_Columbia	4	160,90
	review-Florida	19	886,30
	review-Georgia	13	575,50
	review-Hawaii	11	471,70
	review-Idaho	14	569,50
	review-Illinois	14	587,50
	review-Indiana	15	599,00
	review-Iowa	18	673,00
	review-Kansas	13	511,00
	review-Kentucky	11	429,50
	review-Louisiana	10	385,30
	review-Maine	8	290,00
	review-Maryland	16	669,30
	review-Massachusetts	16	640,60
	review-Michigan	15	624,70
	review-Minnesota	12	502,80
	review-Mississippi	14	486,70
	review-Missouri	11	455,90
	review-Montana	7	248,00
	review-Nebraska	13	468,30
	review-Nevada	12	520,00
	review-New_Hampshire	9	342,40
	review-New_Jersey	13	539,80
	review-New_Mexico	12	486,90
	review-New_York	18	759,90
	review-North_Carolina	15	656,20
	review-North_Dakota	4	144,90
	review-Ohio	13	538,50
	review-Oklahoma	11	445,80
	review-Oregon	15	644,90
	review-Pennsylvania	16	658,30
	review-Rhode_Island	6	227,20
	review-South_Carolina	14	575,60
	review-South_Dakota	5	172,60
	review-Tennessee	12	497,40
	review-Texas	16	713,50

	review-Utah	10	477,20
	review-Vermont	3	85,60
	review-Virginia	12	480,80
	review-Washington	13	569,80
	review-West_Virginia	8	264,80
	review-Wisconsin	12	454,10
	review-Wyoming	3	111,00
		<b>Total Peso (Mb)</b>	<b>24883,70</b>
metadata-sitios		11	<b>2832,30</b>

Respecto a la calidad de los datos, en el proceso de EDA, en este primer acercamiento de esta semana, se encontró lo siguiente:

En las tablas de *reviews*, columnas 'text', 'pics' y 'resp' se visualizaron gran cantidad de valores nulos, por lo que posteriormente, en el proceso de ETL se deben tomar decisiones para rellenar estos espacios con información extraída de otras fuentes, o eliminarlas.

En las tablas de *sitios*, la columna 'description' tiene gran cantidad de nulos, por lo que en la siguiente fase del proyecto se decidirá cómo rellenar esta información, ya que es necesaria al momento en que el comensal haga sus consultas. La columna 'price', se va a eliminar, ya que no contiene ningún dato válido (sólo caracteres). Más tarde habrá que adicionar otra columna que se refiera a esa información.

En general, y por la gran cantidad de Datasets, se les hará su respectivo ETL, uniendo las tablas, haciendo limpieza de ellas y terminando con una exportación a un tipo de dato .parquet para su mayor compresión y posterior uso en lo que sigue del proceso de Análisis y Machine Learning.