



Bienvenidos a la documentación de nuestro proyecto Grupal!

Nº. Grupo: 11

Participantes:

- Nelson Joel Stanich
- Matias Bonfanti
- Leon Yohel Espitia
- Geronimo Maggi

Rubro: Gastronomía.

Territorio: Estados Unidos.

Propuesta del Proyecto

Qué es Yelp.

Yelp es una plataforma dirigida a consumidores de varias partes del mundo, nuestro enfoque se concentrará explícitamente en el territorio **estadounidense** dado su tamaño y relevancia.

La plataforma nos ofrece información, fotos y especialmente reseñas acerca de los negocios locales sin importar el tipo o el tamaño. Además, a través ella se puede contactar con el negocio para diferentes acciones, tales como: requerir una cotización, entrar en lista de espera, fijar una cita, hacer una reservación e incluso comprar un producto o servicio.

1. Problemáticas.

1.1 Reviews (reseña).

Dentro de la plataforma **Yelp** existe un campo llamado 'Review' (*a partir de ahora 'reseña'*), el cual da la posibilidad a los clientes de expresar sus opiniones, sobre el local basados en sus experiencias. Estos comentarios pueden ser de carácter **positivo o negativo**. Además, pueden abarcar diferentes aspectos tales como: Calidad, servicio, precio, tiempos, etc.

Estas reseñas resultan de **gran valor y utilidad** para aquellos usuarios con intención sobre los negocios, permitiéndoles generar un grado de previa confianza y/o expectativa ante el o los productos o servicios ofrecidos por el negocio.

De igual manera, las reseñas brindan a los administradores y directivos, la oportunidad de generar métricas y otros indicadores, que permiten **medir** el cumplimiento de sus objetivos, además de la percepción que tienen los clientes de sus negocios.

Como paso inicial de nuestro análisis, hemos identificado algunas situaciones que afectan la calidad del dato ('reseñas') entre ellas encontramos:

- Vacío: Reseñas sin diligenciar.
- Mal diligenciado: Problemas gramaticales.
- Falsos: Sesgos en la información.

Como manejo a esta problemática *encontramos la necesidad de analizar profundamente cada rubro y **proyectar** los tipos de negocios que pueden **crecer o decrecer**.*

1.2 Inserción de nuevos participantes en el negocio.

Ante un posible capital de inversión, se planea determinar las zonas con mayores posibilidades de éxito en la inserción de nuevas unidades de negocio.

Frente a esta problemática, nuestro equipo realizará un análisis exhaustivo de los datos, con el fin de fundamentar una decisión que resuelva la incertidumbre de nuestro cliente.

1.3 Sistema de recomendación.

Se identifica como una problemática, la falta de un sistema que le permita a los clientes encontrar un negocio o local adecuado a sus gustos y requerimientos.

Dada la amplia oferta de establecimientos en el rubro gastronómico, se plantea ofrecer a los usuarios interesados un **sistema de recomendación** que brinde información veraz y oportuna al momento de decidir una opción de compra o visita. Encontramos gran utilidad de este sistema, puesto que se podría mejorar la experiencia del usuario a través de este servicio.

2. Objetivos

A lo largo de todo este proyecto y de este documento, se seguirá un hilo conductor: **problemática-objetivo-solución**. Es decir, para cada problemática identificada, se establecerá un objetivo que le satisfaga, sumado a esto, se seleccionaran herramientas que permitan lograr los objetivos correspondientes a cada problemática (de estas hablaremos más adelante) y se ejecutará una solución.

Si seguimos el hilo, aquí podremos los siguientes objetivos:

2.1 Aplicar técnicas de análisis de datos que permitan categorizar las opiniones de los clientes las cuales están basadas en sus experiencias en el establecimiento.

Con este análisis pretendemos determinar si el negocio está **Creciendo o Decreciendo en términos de popularidad**.

2.2 Generar una proyección de la probabilidad de apertura de nuevos participantes de negocio basados en el análisis previo de los datos. Determinando así, las **zonas** dentro del territorio norteamericano donde tenga **potencial, la inserción de nuevos establecimientos**.

2.3 Crear un **sistema de recomendación** que facilite la búsqueda de un establecimiento basado en el análisis de los datos y las preferencias de los usuarios.

3.Alcance del Proyecto

El alcance de este proyecto está dado al cumplimiento de los objetivos planteados. Sin embargo, queremos destacar las tareas principales e irremplazables que hacen parte del proceso para el alcance de las metas establecidas. Para ello hemos definido el siguiente camino:

Fase 1:

El paso inicial el cual es definido de avanzar con cualquier otra fase del proyecto, es precisamente revisar los datos.

Esta fase se basa principalmente en un proceso **ETL** (Extract, Transform, Load). La cual será aplicada en el conjunto de datasets extraídos de la página oficial de Yelp y provista por el staff de Henry.

A estos conjuntos de datos, se les hará un análisis exploratorio, se sacarán conclusiones y se procederá con su debido tratamiento; enfocándonos en las tareas de limpieza, normalización, verificación de sus campos, detección de outliers, y demás pasos fundamentales del proceso.

Al finalizar esta fase, se desea generar una **Base de Datos** y crear **tablas relacionales** que permitan la continuidad del proyecto, y de esta manera, proponer un análisis minucioso que provea los resultados esperados y así lograr el cumplimiento de los objetivos planteados en el proyecto.

Fase 2:

Siguiendo el hilo de este documento, la segunda fase viene alineada con la resolución del primer problema que se detectó, el cual está relacionado al tema de reseñas. Para esto, aplicaremos los conceptos que conocemos en el campo de Machine Learning.

Aquí podemos ver el paso a paso propuesto:

- Gestión de features para aplicación del modelo ML (Categorización de variables alfanuméricas).
- Selección del modelo ML.
- Entrenamiento y Testeo del modelo de ML seleccionado.
- Obtener predicciones basadas en el modelo ML aplicado.
- Análisis final de la predicción.

Fase 3:

En esta fase nos centraremos especialmente en la parte **analítica** de la Base de Datos obtenida a través del cumplimiento de las fases anteriores.

Sabemos que las últimas dos problemáticas que se nombraron con anterioridad, deben ser abarcadas con el fin de mostrar diferentes visualizaciones, gráficos, dashboards para dar cumplimiento a los objetivos propuestos.

Para esto se realizará lo siguiente:

- Creación de métricas para la apropiada evaluación del proyecto.
- Análisis y evaluación de resultados - KPI'S.
- Armado de los dashboards interactivos.
- Creación de storytelling.

Fase 4:

Aquí ya habremos llegado al final de nuestro proyecto y lo que nos quedará por hacer será sacar diferentes conclusiones del proyecto.

4.KPI'S

La definición de KPIs debe ir de la mano de los objetivos propuestos anteriormente. Con estos, evaluaremos el éxito, el rendimiento de los objetivos, desempeño, entre otros. Presentamos aquí los KPIs y mencionaremos a que objetivo corresponde.

Desempeño del equipo de trabajo: Mediremos la productividad del equipo, contrastando las **tareas realizadas** por cada uno de los integrantes, contra las **tareas asignadas** en el diagrama de Gantt.

Con respecto al objetivo 1:

CSAT (Customer Satisfaction Score) mediremos la satisfacción de los clientes con el producto, servicio o empresa. Donde podemos evaluar si el cliente **está satisfecho con su experiencia**.

Con respecto al objetivo 2:

Predicción de Zonas Potenciales: Análisis del desarrollo comercial de las zonas en estudio. Evaluando la cantidad de aperturas vs. cierres de locales, teniendo como parámetro los últimos años, ampliando el rango de tiempo si es necesario para contemplar el comportamiento de los datos en pandemia. (rango aproximado 5 años)

Con respecto al objetivo 3:

Uso del sistema de recomendaciones:

Este KPI nos permitirá evaluar el sistema de recomendaciones, para ello estableceremos una **cantidad mínima de usuarios**, medibles en un **término a definir** a partir de la implementación del sistema.

Grado de eficacia del sistema de recomendación: Con este KPI evaluaremos el grado de precisión del sistema. Para cumplirlo establecemos un **indicador favorable** de la credibilidad del sistema **superior al 80%**; lo mediremos a través de una campaña de feedback (encuesta) que será aplicada a la mayoría de usuarios que utilicen el sistema.

5. Repositorio GitHub

Todo el proyecto estará en un repositorio de Github. Aquí dejamos el enlace:

[Enlace al Repositorio](#)

6. Soluciones

Recordamos el hilo que venimos siguiendo en este documento (y que seguiremos a lo largo de todo el proyecto). Problemática-Objetivo-Solucion. Ya identificamos cuales son los problemas principales, definimos los objetivos que se deben cumplir, vimos también el alcance de todo el proyecto, juntos con los KPI 's que se propusieron. Teniendo todo esto, quedan de documentar algunos aspectos importantes para que se lleve a cabo la resolución de nuestro proyecto. Estos son:

6.1 Stack Tecnológico

Dentro de la oferta tecnológica nuestro grupo ha escogido trabajar con las siguientes herramientas tecnológicas:

- Docker
- AirFlow (Flujo de trabajo en herramientas tecnológicas)
- Minlo (Data Lake - Almacenamiento de datos)
- Apache Spark (ETL)
- Postgres (Sistema relacional de base de datos)
- Power BI (Visualización)
- Python - Streamlit (App Web- Sistema de recomendaciones)

6.2 Roles

La elaboración del proyecto será realizada en conjunto con el equipo, distribuyendo las tareas y responsabilidades de forma equitativa y según los skills de cada uno de los integrantes.

La conformación de los roles de trabajo será la siguiente:

- Bonfanti Matias, (Data Engineer)
- Espitia Leon, (Data Analyst)
- Maggi Geronimo (Data Engineer)
- Stanich, Joel (Data Analyst)

Cabe destacar que el cumplimiento de los roles implica tareas diversas y que en ocasiones será necesario la asistencia por parte de otros participantes.

6.3 Metodología de trabajo

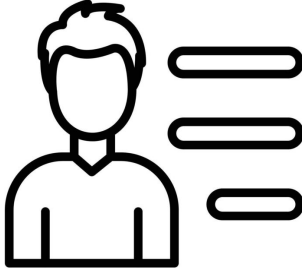
Como Metodología de Trabajo para el desarrollo del proyecto utilizaremos el modelo de SCRUM.

La **metodología Scrum** es un proceso para llevar a cabo un conjunto de tareas de forma regular con el objetivo principal de trabajar de manera colaborativa, es decir, para fomentar el trabajo en equipo. Con este método de trabajo lo que se pretende es alcanzar el mejor resultado de un proyecto determinado.

En Scrum se van realizando entregas regulares y parciales del trabajo final, de manera prioritaria y en función del beneficio que aportan dichas entregas a los receptores del proyecto. Por este motivo, es una metodología especialmente indicada para proyectos complejos, con requisitos cambiantes y en los que la innovación y la flexibilidad son protagonistas.

7. Análisis preliminar de la calidad de los datos

El conjunto de datos



6.9 millones de opiniones



150.346 empresas



27 Estados

- 908,915 revisiones de 1,987,897 usuarios
- Más de 1.2 millones de atributos comerciales como horas, estacionamiento, disponibilidad y ambiente.
- Check-ins agregados a lo largo del tiempo para cada una de las 131,930 empresas

Sólo están incluidas empresas que han tenido al menos 3 revisiones en menos de 14 días. Por lo tanto, el número de recuento de revisiones puede diferir del número de revisiones reales para cualquier usuario.

¿Qué revisiones están incluidas en los archivos de revisiones?

Solo las opiniones que Yelp recomendó en el momento de la recopilación de datos están incluidas.

8.Conclusión

Habiendo mencionado y explyado todos los aspectos necesarios para que se lleve a cabo, en nombre de todo el equipo agradecemos a todos los lectores de este documento.

Muchas gracias.