**Pasos a Seguir**

* **Analizar el enfoque, que información de yelp y google maps vamos a enfocarnos.**
* **Armar un repositorio de github, para trabajar en grupo (Hacer pruebas antes de cómo funciona y probarlo todos).**
* **Empezar con ETL y después con el EDA.**
* **Hacer la división del trabajo (Ver en que cantidad de partes y ponerle prioridades)**
* **Definir qué podemos hacer hasta el 1 de enero.**
* **Ver en donde podemos alojar el proyecto (Google, AWS, Azure)**
* La base de datos depurada deberá contemplar por lo menos dos tipos diferentes de extracción de datos, ejemplo: datos estáticos, llamadas a una API, scrapping, entre otros.

**Pasos a seguir**

1. **Hacer ETL**
2. **EDA**
3. **ANALIZAR QUE PRECISAMOS DE LAS BASES REVIEWS Y METADATA SITIOS**
4. **2 API’S**
5. **MODELO MACHINE LEARNING -** Sistemas de recomendación
6. Mejoramiento de estrategias de marketing: campañas microsegmentadas
7. Datos adicionales a cruzar: Cotizaciones en bolsa, tendencias en redes sociales y medios de comunicación sobre comercios en expansíón.

**En que consiste**

# Sobre el proyecto final

Esta etapa consiste en la simulación de un entorno laboral real. Se va a buscar potenciar habilidades tech y soft, fomentando el trabajo en equipo, avanzar en el desarrollo de uno o más [productos de data](https://www.google.com/url?q=https://www.irion-edm.com/es/data-governance-insights-es/data-product-definicion-contextos-metadatos/&sa=D&source=editors&ust=1702657408834891&usg=AOvVaw2zNl6XVXvfwZtE0fWCnz_i), respondiendo a los requerimientos planteados por un [Product Owner](https://www.google.com/url?q=https://asana.com/es/resources/product-owner&sa=D&source=editors&ust=1702657408835194&usg=AOvVaw27cBKZQm1VniuaE1l_0L1I), a partir de una propuesta de producto que debe nacer del grupo.

La propuesta que hagan va a ser aceptada o modificada por el PO, que es encargado también de supervisar el avance, redireccionando esfuerzos (de ser necesario, semanalmente).

A su vez, a lo largo del desarrollo, los grupos contarán con el [seguimiento diario](https://www.google.com/url?q=https://www.atlassian.com/agile/scrum/standups&sa=D&source=editors&ust=1702657408835825&usg=AOvVaw0T8n4tuR-mp3AO228ywquB) de un HM, que podrá brindar apoyo sobre generalidades del proyecto y, eventualmente, consejos técnicos.

## ****Contexto****

"La opinión de los usuarios es un dato muy valioso, que crece día a día gracias a plataformas de reseñas. Su análisis puede ser determinante para la planificación de estratenias. Yelp es una plataforma de reseñas de todo tipo de negocios, restaurantes, hoteles, servicios entre otros. Los usuarios utilizan el servicio y luego suben su reseña según la experiencia que han recibido. Esta información es muy valiosa para las empresas, ya que les sirve para enterarse de la imagen que tienen los usuarios de los distintos locales de la empresa, siendo útil para medir el desempeño, utilidad del local, además de saber en qué aspectos hay que mejorar el servicio. Además, Google posee una plataforma de reseñas de todo tipo de negocios, restaurantes, hoteles, servicios, entre otros integrada en su servicio de localización y mapas, Google Maps. Los usuarios utilizan el servicio y luego suben su reseña según la experiencia vivida. Muchos usuarios leen las reseñas de los lugares a los que planean ir para tomar decisiones sobre dónde comprar, comer, dormir, reunirse, etc. Esta información es muy valiosa para las empresas, ya que les sirve para enterarse de la imagen que tienen los usuarios de los distintos locales de la empresa, siendo muy útil para medir el desempeño, utilidad del local, además de identificar los aspectos del servicio a mejorar.""

## ****Rol a desarrollar****

"Como parte de una consultora de data, nos han contratado para poder realizar un análisis del mercado estadounidense. Nuestro cliente es parte de un conglomerado de empresas de restaurantes y afines, y desean tener un análisis detallado de la opinión de los usuarios en Yelp y cruzarlos con los de Google Maps sobre hoteles, restaurantes y otros negocios afines al turismo y ocio, utilizando análisis de sentimientos, predecir cuáles serán los rubros de los negocios que más crecerán (o decaerán). Además, desean saber dónde es conveniente emplazar los nuevos locales de restaurantes y afines, y desean poder tener un sistema de recomendación de restaurantes para los usuarios de ambas plataformas para darle, al usuario por ejemplo la posibilidad de poder conocer nuevos sabores basados en sus experiencias previas. Pueden cambiar el tipo de comercio (no es necesario que sean restaurantes). "

## ****Propuesta de trabajo****

Recopilar, depurar y disponibilizar la información: Creación de una base de datos (DataWarehouse) de diferentes fuentes, tanto provistas por Henry como incorporadas por ustedes, corriendo en local o alojada en proveedores en la nube. La base de datos depurada deberá contemplar por lo menos dos tipos diferentes de extracción de datos, ejemplo: datos estáticos, llamadas a una API, scrapping, entre otros.

Reporte y análisis significativos de la(s) línea(s) de investigación escogidas: El análisis debe contemplar las relaciones entre variables y concluir, si es que existe, una relación entre estas, y los posibles factores que causan dicha relación en la realidad.

Entrenamiento y puesta en producción de un modelo de machine learning de clasificación no supervisado o supervisado: El modelo debe resolver un problema y conectar globalmente con los objetivos propuestos que se propongan como proyecto. .

## ****Ideas de análisis e implementación****

Mejoramiento de estrategias de marketing: campañas microsegmentadas

Sistemas de recomendación: Sobre algún lugar en particular, pueden ser comercios como restaurantes, hoteles, entre otros.

Datos adicionales a cruzar: Cotizaciones en bolsa, tendencias en redes sociales y medios de comunicación sobre comercios en expansíón.

## ****Datasets y fuentes complementarias****

"Los datos son extraídos de la plataforma de reseñas Yelp y de Google Maps, para Estados Unidos. Tiene información sobre la ubicación de los comercios, su categoría, puntajes promedios, si están abiertos o no, sobre los usuarios, las reseñas que hicieron, cuántas reseñas hicieron, cuántos votos han recibido esas reseñas entre otros. Se puede complementar el análisis con datasets adicionales que ofrezcan información que consideren pertinente al pedido (valores de acciones de las empresas, información geográfica adicional de los locales, etc). Recomendamos fuertemente el uso de NLP para procesar la información de las reseñas.

# Hitos y entregables

A lo largo de cada semana del desarrollo del proyecto, habrá un [conjunto de tareas a realizar](https://www.google.com/url?q=https://blog.ganttpro.com/es/que-es-un-hito-y-que-ejemplos-de-hitos-hay/&sa=D&source=editors&ust=1702657408836294&usg=AOvVaw3VDwQWQei9lb7fK5pdunzO), muchas de las cuales tendrán que definir ustedes mismos a partir del planteo de sus objetivos.

Sin embargo, alguna de esas tareas (incluyendo todas las subtareas que pudieran conformarlas) será crítica para pasar a la siguiente semana/ etapa del PF. Por ejemplo, sin realizar el ETL completo, no se puede dar inicio a la etapa de Analytics. Por este motivo, identificamos la compleción del ETL, que sabemos abarca muchas subtareas, como un hito de la primera semana.

Algunas tareas tienen como resultado productos, ya sean tangibles o intangibles, que dan una pauta de la concreción de esa tarea. Estos productos son los que identificamos como entregables.

### En resumen

Los hitos pueden definirse como las acciones que, una vez realizadas, dan una pauta del grado de avance en el cumplimiento de los objetivos. Mientras que los entregables son el producto de la realización de las tareas específicas a lo largo del proyecto, y pueden ser tangibles o intangibles.

# Importante

La documentación es un ítem fundamental en el desarrollo de cualquier proyecto de data y se espera que a lo largo de todo el PF, provean documentación para cada etapa. La misma deberá ser lo más detallada y clara posible. Cada semana contemplará elementos particulares a incluir en la documentación, pero en líneas generales, siempre deben documentar qué hicieron, con qué herramientas y fundamentar todas las decisiones que tomen con respecto al proyecto.

# Sprint #1: Puesta en marcha del proyecto y Trabajo con Datos

## Hitos

* 4 KPI’s
* Documentar alcance del proyecto
* EDA de los datos
* Repositorio en Github
* Implementación stack tecnológico
* Metodología de trabajo
* Diseño detallado
* Equipo de trabajo - Roles y responsabilidades
* Cronograma general - Gantt
* Análisis preliminar de calidad de datos

## Entregables

* Documentación
* Stack elegido y fundamentación
* Flujo de trabajo

# 👨‍💻 Sprint #2: Data Engineering

## Hitos

* ETL completo
* Estructura de datos implementada (DW, DL, etc). Pueden usar algún servicio
* Pipeline ETL automatizado
* Diseño del Modelo ER
* Pipelines para alimentar el DW
* Data Warehouse
* Automatización
* Validación de datos
* Documentación
* Diagrama ER detallado (tablas, PK, FK y tipo de dato)
* Diccionario de datos
* Workflow detallando tecnologías
* Análisis de datos de muestra
* MVP/ Proof of Concept de producto de ML ó MVP/ Proof of Concept de Dashboard

## Entregables

* Documentación y reporte realizado

# 📈 Sprint #3: Data Analytics + ML

* Diseño de Reportes/Dashboards
* KPIs
* Modelos de ML
* Modelo de ML en producción
* Documentación
* Selección del modelo, feature engineering
* Informe de análisis
* Video del proyecto realizado, para ser votado y, en caso de ganar, ser presentado en la graduación final.

## Hitos

* Dashboard final
* Producto/s de ML

## Entregables

* Dashboard
* Modelo de Machine Learning (grupos de 5 obligatorio)

# ✅ Última demo: consideraciones finales

\*La última demo es el viernes del tercer sprint

Para la última demo del Proyecto Final, se espera que los alumnos lleguen con los proyectos cerrados, implementando los últimos retoques/ características que los alumnos consideren, o que el PO les haya solicitado.

Es importante que se destine una parte de la última semana a trabajar en sus presentaciones, storytelling, verificar que todo funcione correctamente y completar documentación. Los repos deben quedar ordenados, y con un readme completo presentando el proyecto.

**Video del Proyecto Final**

La última tarea del pryecto final, es la de realizar un video explicativo del trabajo que realizaron en esta última etapa. La creación del video es obligatoria para todos los grupos (y si, van a poder subir esos videos a sus redes sociales y compartilos con quien lo deseen!), además, en una votación interna de la cohorte, los tres videos más votados, van a poder mostrarlos y presentar el día de la graduación.

**Requisitos del video:**

* Subir el video que realizaron para el PF --> extensión: 3 minutos de duración cómo máximo
* Enviar el enlace del mismo al formulario que se les va a facilitar.
* Deadline: jueves 10:00 GMT-3 de la semana siguiente a la demo final.
* Ese mismo jueves haremos una votación abierta a toda la cohorte (desde 10:30 hasta 12:00 GMT-3) Y les comunicaremos los equipos ganadores
* Aclaración: se considerarán solamente los grupos aprobados para exponer.