



# **¡Felicitaciones!** **Has llegado al** **Proyecto Final**

**¡Dónde pones en práctica todo lo aprendido!**



# proyecto Final





# Presentación

Esta etapa consiste en la **simulación de un entorno laboral real**. Se busca potenciar habilidades tech y soft, fomentando el trabajo en equipo, de manera **independiente**. Así, el grupo deberá coordinarse, aprovechar las fortalezas individuales para lograr sinergia y avanzar en el desarrollo de un producto de data, respondiendo a los requerimientos planteados por un Product Owner, a partir de una propuesta de producto que debe nacer del grupo.

# Calendario



Semana	Lunes	Martes	Miércoles	Jueves	Viernes
1					<b>Kick OFF</b> PF 17 de nov
2	<b>FERIADO</b> 20 de nov	Job Prep Personal Brand 21 de nov	Daily 22 de nov	Job Prep Creación de CV 23 de nov	<b>DEMO</b> Planning & Retrospectiva 24 de nov
3	Dailys 27 de nov	Job Prep Personal Brand 28 de nov	Daily 29 de nov	Trabajo en equipo 30 de nov	<b>DEMO</b> Planning & Retrospectiva 1 de diciembre
4	Dailys 4 de diciembre	Job Prep Interviews 5 de diciembre	Daily 06 de diciembre		<b>FERIADO</b> 08 de diciembre
5	<b>DEMO FINAL</b> 11 de diciembre				

\* Presencial

\* Asíncrono

\* FERIADO



# Propuesta

Una empresa de **E-Commerce** de Argentina, está evaluando la posibilidad de expandirse al Brasil, y para ello necesita entender cómo es el mercado de E-Commerce de allá. Para ello, se consiguió data de 100k de órdenes desde el 2016 hasta el 2018, de distintos puntos de venta en Brasil. Esto se puede ver desde distintas dimensiones, **status de las ordenes, precios, pagos y performace de envíos hacia los usuarios, productos e incluso reviews de los mismos, escritos por diversos clientes**. Así mismo, se podrá encontrar un archivo de geolocalización con todos los códigos postales de Brasil, junto con tu latitud y longitud.



# Objetivos

- Realizar un proyecto de E-Commerce.
- Analizar la data de un la empresa Olist.
- Llevar a cabo una investigación del modelo de negocio.
- Crear una idea de producto para ofrecer soluciones de valor.



# ¿Qué esperamos?

## Algunas ideas....

Con base a los puntos de la **siguiente** slide, esperamos que puedan inspirarse para su producto final.

Es decir, tienen **la libertad de elección de los temas y el alcance que le quieran dar a su proyecto.**

Esto trae consigo un desafío más, ya que el proyecto no está acotado a pautas estrictas, sino que van a poder moldear el enfoque de acuerdo a sus preferencias.



# Algunas ideas de enfoque

**Ventas:** Con base en la data, podrías realizar una pequeña predicción de cómo serían las ventas a futuro.

**Delivery:** Si queremos entrar de lleno en este mercado, el tema de logística sería un punto importante a considerar y sería óptimo tener referencias sobre las principales métricas de logística.

**Productos:** Con base a las reviews, ¿podemos ver cuales son considerados los productos con mejor o peor calidad? Las 10 principales y las 10 últimas categorías de productos según los ingresos.

**Por ubicación geográfica:** Podrías agregar de ingresos por 'estado' Las 5 principales categorías de productos por 'estado' Número de clientes y vendedores por 'estado'

**Vendedores:** Analiza a los 10 mejores vendedores que tienen un gran desempeño dentro de las categorías.

**Pago:** Agregación del método de pago del cliente.

**La satisfacción del cliente:** Tendencia de compra.





# Ejemplos de posibles productos para tu proyecto E-Commerce

**Calculadora de costos de envío:** Determinar costo de envío de acuerdo a códigos postales y lejanía. Costos mínimos por producto/peso.

**Predicción de ventas:** Por estacionalidad, anualmente, mes a mes. A partir de estos datos, otros departamentos de la empresa pueden realizar acciones. Ej: estrategias de publicidad por parte de Marketing.

**Satisfacción del cliente:** Tendencias de compra y reviews.

# ¿En qué consiste cada Sprint?

Sprint #1	Sprint #2	Sprint #3
<b>Inicio del proyecto y empezando a trabajar con datos</b> <ul style="list-style-type: none"><li>• Kickoff del proyecto</li><li>• Entendimiento de las necesidades</li><li>• Documentar alcance, objetivo y entregables</li><li>• <u>Implementación</u> de stack propuesto</li><li>• Comienzo de EDA</li></ul>	<b>Data Engineering</b> <ul style="list-style-type: none"><li>• Arquitectura propuesta y diagrama</li><li>• Diccionario de datos</li><li>• Creación del DW</li><li>• Análisis de Datos</li><li>• MVP Dashboard</li><li>• Modelos y MVP producto ML</li></ul>	<b>Etapa de Analytics y Machine Learning</b> <ul style="list-style-type: none"><li>• Dashboard final</li><li>• Reportes</li><li>• Producto ML</li><li>• Ajustes necesarios al modelo</li><li>• Demo general del proyecto</li><li>• Entregable final</li><li>• Documentación</li></ul>

# Sprint 1





# Descripción del Sprint

En esta primera semana, se espera que realices una primera aproximación al proyecto y la temática que este plantea. Debes de hacer una robusta investigación y trabajo preliminar de los datos que disponen.

Por último, diseña un planeamiento de las siguientes etapas del proyecto.

## **Sprint #1**

### **Puesta en marcha el proyecto**

- Kick-off del proyecto
- Entendimiento de las necesidades
- Documentar alcance, objetivo y entregables
- Implementación del Stack Propuesto

# Hitos – Semana #1

**1. Entendimiento de la situación propuesta:** en equipo definen la propuesta y herramientas a utilizar.

**2. Objetivos:** específicos del trabajo y del grupo: establezcan los objetivos del proyecto y equipo.

**3. Alcance y fuera de alcance:** delimitan muy bien el alcance y las limitaciones de su solución.

**4. Crear repositorio en Github:** realicen una repositorio público en Github, usen la cápsula sobre Github que les compartimos.

**5. Solución propuesta:** en este paso definan lo siguiente:

- **Implementación Stack tecnológico** *¿Qué herramientas usarán?*
- **Metodología de trabajo** *¿Cómo será la manera de trabajar durante el todo el PF?*
- **Roles y responsabilidades** *¿Quién estará a cargo de qué?*

**6. Comienzo de EDA:** Deben de empezar con el análisis exploratorio de datos



# Github

Como verás, en este primer sprint tendrás que crear un repositorio en Github.

Como sabemos que este es un nuevo procedimiento, te invitamos a revisar el siguiente recurso haciendo click en el botón.

[Cápsula de Github](#)

# Sprint 2





# Descripción del Sprint

En esta segunda semana, se espera que los alumnos comiencen con el armado y construcción de toda la arquitectura del proyecto. Este es el sprint de Ingeniería de Datos. Es primordial que utilicen lo realizado en el sprint anterior, para poder trabajar de forma eficiente y acorde a los requerimientos de esta semana.

De este sprint depende que luego se puedan realizar tanto el análisis de datos como los modelos de Machine Learning.

## Sprint #2

### Trabajando los datos

- Creación del DW
- Análisis de datos de muestra
- MVP Dashboard
- MVP Producto ML





# Hitos - Semana #2

**1. Diseño adecuado del Modelo ER:** fase de modelado conceptual, donde se definen las entidades relevantes y sus relaciones antes de implementar físicamente la base de datos

**2. Documentación:** proporciona información detallada y estructurada sobre todos los aspectos del proyecto.

**3. Validación de datos:** proceso de asegurarse de que los datos de entrada sean correctos, relevantes y estén bien formados.

**4. Diccionario de Datos:** escriban una guía de referencia detallada que describe el conjunto de datos, sus características y cómo se deben utilizar y manejar los datos

**5. Arquitectura propuesta:** elaboren un plan o diseño detallado que describe cómo se organizan y conectarán los diferentes componentes del sistema.

**6. Creación del Data Warehouse:** deberán crear su Data Warehouse con todas las características.

**7. MVP Dashboard (Streamlit, Power Bi, Tableau, Looker):** presenten un borrador de su dashboard en la herramienta que seleccionaron.

# Sprint 3





# Descripción del Sprint

En esta última semana, se espera que desarrolles modelos de ML. Luego, con el/ los elegidos, realizar un producto orientado a datos. Además, tienes que realizar un reporte de **insights** y datos relevantes, recabados de la información con la que cuentan (puede ser incluida dentro del Streamlit).

Toda esta última presentación debe ir acompañada de un **storytelling**, el cual sea acorde a las necesidades de negocio, es decir, para tratar de resolver problemas o aportar soluciones a la empresa.

## Sprint #3

### **Etapas de Analytics y ML**

- Dashboard
- Insights y datos relevantes
- Producto ML



# Hitos - Semana #3

1. **Diseño de Reportes:** creación de documentos que presentan y explican los resultados de los análisis de datos de una manera clara, precisa y accesible.

2. **Dashboard versión final (e interactivo):** deberán presentar un dashbord interactivo que ayudará a la visualización de datos.

3. **Documentación:** presenten la documentación final.

4. **Funcionalidad/ interactividad :** recuerden que deberán proporcionar a los usuarios interactuar con los datos de forma significativa y productiva.

5. **Producto de ML en Streamlit:** aplicación web interactiva que permite a los usuarios interactuar con un modelo de ML.



# storytelling

En este sprint tendrán que hacer uso de habilidades blandas como es el *Storytelling*. Proceso de traducir análisis de datos en términos comprensibles con el fin de influir en una decisión o acción empresarial. El análisis de datos se centra en crear conocimientos valiosos a partir de los datos para proporcionar contexto y comprensión adicionales a una audiencia prevista.

⚠ Revisa la siguiente información si tienes dudas de cómo realizar este requerimiento



# Resumen de lo esperado (e ideas extra para implementar)

Los hitos son obligatorios. Las ideas no lo son, pero alentamos su implementación.

	sprint #1	sprint #2	sprint #3
Hitos	<b>Puesta en marcha del proyecto y definiciones iniciales:</b> <ul style="list-style-type: none"><li>• Tecnologías a usar</li><li>• Documento de alcance del proyecto</li><li>• EDA Preliminar, calidad del dato</li><li>• Repo Github</li><li>• Implementación Stack Propuesto</li></ul>	<b>Trabajando los datos</b> <ul style="list-style-type: none"><li>• Datawarehouse automatizado con carga inicial (La arquitectura NO puede ser solamente python + sql)</li><li>• Al menos 2 tablas de hechos y 5 dimensionales</li><li>• Diccionario de Datos</li><li>• MVP Dashboard</li><li>• Métricas</li></ul>	<b>Etapas de Analytics y ML</b> <ul style="list-style-type: none"><li>• Storytelling</li><li>• Dashboard (Streamlit o Power Bi, etc.)</li><li>• Insights encontrados sobre el negocio</li><li>• Recomendaciones de negocio</li><li>• Retomar hitos presentados en las anteriores demos</li><li>• Retoques finales en base a recomendaciones de HM y PO</li><li>• Producto ML en Streamlit</li></ul>
Ideas	<ul style="list-style-type: none"><li>• Seguimiento de tareas con un Diagrama Gantt</li></ul>	<ul style="list-style-type: none"><li>• Uso de herramientas Big Data como HDFS, Hive, Spark y/o servicios cloud</li><li>• Carga Incremental de Datos (puede ser en video)</li></ul>	



# Criterios DEMOS – DAILYS

Criterios para tener en cuenta en las demos y las dailys

- Siempre enciende la cámara.
- Entrar a la reunión unos minutos antes.
- Definir con anticipación el orden de presentación y los roles.
- Dejar listo todo lo que se va a compartir (y de forma pública).
- No leer textualmente las slides.
- En las dailys llevar dudas y consultas, pero sobre todo: posibles soluciones.
- Todos los integrantes del grupo participen tanto en demos como dailys.
- Repasar con el grupo la presentación antes de la demo.



# Es obligatorio revisar el siguiente material

## Workshops Asíncronos

- Workshop ANN
- Workshop Preprocesamiento de Datos
- Workshop Streamlit
- Workshop From Data to Viz

## Documentación

PF

- Rúbricas de evaluación de PO para Estudiantes
- Documentación General PF
- Consultor BI
- Disclaimer PF





Ten siempre a la mano el repositorio de  
este proyecto

# ¡Muchas Gracias!

**#soyhenry**

