

# Generalidades PF --> Students

🕒 8 min. read · 📖 [View original](#)

---

## Proyectos Finales

### 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, de manera independiente. Así, el grupo deberá coordinarse, aprovechar las fortalezas individuales para lograr sinergia y avanzar en el desarrollo de uno o más [productos de data](#), respondiendo a los requerimientos planteados por un [Product Owner](#), a partir de una propuesta de producto que debe nacer del grupo.

Entonces, como punto de partida, el grupo de alumnos deberá analizar la temática y los datos disponibles para formular una propuesta de proyecto a desarrollar.

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](#) de un HM, que podrá brindar apoyo sobre generalidades del proyecto y, eventualmente, consejos técnicos. No obstante, la coordinación,

decisiones y ejecución de tareas son enteramente responsabilidad del equipo en sí.

## Hitos y entregables

A lo largo de cada semana del desarrollo del proyecto, habrá un [conjunto de tareas a realizar](#), 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**

En esta semana deben realizar un análisis del proyecto seleccionado y los datos disponibles. En base al entendimiento que logren de la temática, deben proponer como encararla, brindando una solución o herramientas desarrolladas por ellos mismos para acercarse a dicha solución.

Esta propuesta deberá contemplar los siguientes ítems:

- Entendimiento de la situación actual

En la propuesta debe quedar manifiesto un adecuado manejo de la problemática, deben poder contextualizarla y expresar posibles análisis/ soluciones en torno a la misma.

- Objetivos

Los objetivos deben ser acciones concretas (verbos) que describan claramente lo que buscan lograr con el proyecto. Desarrollar, crear, hacer, etc.

- Alcance

Las temáticas suelen ser amplias y pueden admitir tratamientos mucho más abarcativos en extensión y magnitud de lo que puede realizarse durante el desarrollo del proyecto.

Es por esto que deberán delimitar su trabajo definiendo el alcance y las tareas/desarrollos que puedan considerar importantes para la integridad del proyecto pero que por complejidad o tiempo, estén fuera de alcance.

Esto último pueden plantearlo como posibilidades de continuidad del proyecto.

- Objetivos y KPIs asociados (planteo)

Del entendimiento de la problemática surgirán cuestiones que se buscarán resolver con el trabajo o las herramientas desarrolladas. Estas cuestiones, formuladas como objetivos, admitirán la creación de [KPIs](#) para evaluar su cumplimiento. Es una tarea muy abarcativa y a la vez muy específica en torno tanto a la problemática como al enfoque elegido.

Por ejemplo:

- Temática: generación de CO<sub>2</sub>.

Un enfoque puede ser medir el impacto de la agricultura y el consumo de carne animal en la generación de CO<sub>2</sub>. Un objetivo puede ser reducir la producción de carnes animales en un 8% para 2027 y ese mismo objetivo evaluarse mediante un KPI.

Entonces, la medición del mismo va a poder permitir tomar decisiones de negocio basadas en datos.

- Temática: aviones.

Comparar el retraso de todos los vuelos de la aerolínea de un semestre con respecto a otros. Además, compararlo con el promedio de la industria. Si se encuentra por encima, proponer una reducción de, por ejemplo, 2% para el semestre que viene, e ir aumentando ese porcentaje de reducción en un 1% por los próximos 5 años.

- Repositorio Github

Armar un repositorio de Github para trabajar colaborativamente con todo el grupo. Debe ser público para que lo pueda ver tanto el mentor como el Product Owner. Van a tener que llevar adelante diferentes branches y controles de versiones de su propio trabajo.

- Solución propuesta

Deben detallar qué tareas harán para cumplir los objetivos de trabajo propuestos previamente y cómo lo harán (metodologías de trabajo, forma de organización, distribución de tareas, roles de cada uno dentro del equipo, etc). También, deben detallar qué productos surgirán de su trabajo y en qué etapa los presentarán, teniendo en cuenta los requerimientos generales (entregables esperados) para cada etapa del proyecto.

A su vez, deben realizar una estimación de tiempo para cada tarea, contemplando los tiempos de ejecución globales y los hitos previstos para cada semana; y plasmar esa estimación en un diagrama de Gantt.

Una parte muy importante de la solución propuesta, es con qué herramientas (stack tecnológico) van a realizar la

arquitectura del proyecto. Para esto, lo que van a tener que hacer es seleccionar una pequeña porción de los datos que disponen y realizar un proceso de limpieza y transformación utilizando las herramientas que planean implementar. Esto les dará una idea de cómo funcionarán en el proyecto completo y les permitirá tener un mejor abordaje para futuras tareas. Hay que tener en cuenta que, como este ítem va a ser una presentación previa de lo que van a trabajar en el segundo sprint, el PO puede dar el OK o determinar cuál es el mejor camino para que tomen. Esto les va a permitir adelantar trabajo de la segunda semana, ya que no se va a tener que esperar hasta la segunda demo para verificar si la arquitectura cumple con los requisitos del PO.

Finalmente, como en Data es muy importante trabajar con datos de calidad, deberán incluir en su informe un análisis sobre los datos con los que van a trabajar (metadatos), detallandolos lo más posible: fuentes y confiabilidad de las mismas, qué representa cada columna de cada dataset, tipos de datos, método de adquisición, fecha de adquisición y última actualización, etc.

## 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

En la continuación de la primera semana, se espera que trabajen montando la infraestructura de su proyecto, con pipelines para realizar el proceso de ETL apuntando a estructuras de tipo Data Warehouse, Datalake o Datalakehouse, contemplando la carga incremental de datos.

Deben usar herramientas de big data y/o servicios cloud de su preferencia. En caso de que el grupo esté integrado por cinco personas, es obligatorio su utilización. De ser menos, se espera que también puedan usar estas herramientas. El PO, de ser necesario, les indicará si la arquitectura del proyecto es acorde a lo que se espera de ustedes. Por ejemplo, no se puede solo trabajar los datos con Python y almacenarlos localmente.

En el caso de usar modelos relacionales en sus estructuras de almacenamiento, deben entregar un diseño adecuado y detallado del modelo entidad relación, especificando las tablas, relaciones y tipos de datos adoptados.

En el caso de que vayan a utilizar modelos no relacionales, debería hacer una explicación de porque consideran su implementación por sobre los otros modelos, siempre respaldando las decisiones que tomen.

Siguiendo con la idea de adelantar trabajo, como en el sprint 1, en este segundo sprint, los alumnos van a tener que realizar un análisis exhaustivo en una muestra representativa de los datos. Esto puede incluir identificación de valores atípicos, distribuciones de variables y correlaciones preliminares. Así ya tendrán una idea de las características y peculiaridades de los datos completos y se les podrá dar feedback.

Diseño de Dashboard (Prueba de Concepto): Crear una versión simplificada del dashboard que se planea implementar, incluida la conexión con DW. Incorporar algunas de las visualizaciones preliminares y datos de muestra. Esto les permitirá explorar la interacción y la estructura del dashboard antes de trabajar con datos completos. Y, de nuevo, el PO podrá dar el visto bueno antes de la demo final (después de la cuál ya no se puede seguir modificando en base a recomendaciones).

La otra alternativa, serían el/ los Productos de ML: Antes de implementar los productos interactivos utilizando Streamlit (o la herramienta que sea), crear versiones de prueba con datos de ejemplo. Así permitirá identificar posibles problemas de usabilidad y funcionalidad antes de presentar la versión final.

## 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

En la última semana, se espera que armen un dashboard interactivo, junto con un análisis de los datos que hayan trabajado. Deben incluir los KPI 's que determinaron como importantes para el análisis realizado, y preparar un storytelling con el mismo.

Para los grupos de cinco personas, o en el caso de que el proyecto lo requiera, va a ser obligatorio tener por lo menos implementados modelos, y se esté comenzando a trabajar en un producto de ML (el cuál puede ser un MVP esta semana, y luego ser terminado para la última Demo). Ejemplos de productos: calculadora de costos de envíos según código postal, clasificación de magnitud de sismos, recomendación de días y horarios para pagar tarifas más bajas.

- 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.