



Proyectos Finales



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 semana 2.

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.



Semana 1: Propuesta de proyecto

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, por lo 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 CO₂.

Un enfoque puede ser medir el impacto de la agricultura y el consumo de carne animal en la generación de CO₂, un objetivo puede ser reducir la producción de carnes animales en un 50% para 2030 y ese mismo objetivo evaluarse mediante un KPI.

- 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. Debería 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, con qué herramientas (**stack tecnológico**) 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.

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).

- **Incluir stack tecnológico**
- **Metodología de trabajo**
- **Diseño detallado - Entregables**
- **Equipo de trabajo - Roles y responsabilidades**
- **Cronograma general - Gantt**
- **Análisis preliminar de calidad de datos**

Entregables

- Propuesta de proyecto



Semana 2: Data Engineering

En esta 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. Pueden usar herramientas de big data y/o servicios cloud de su preferencia y será considerado un plus.

En el caso de usar modelos relacionales en sus estructuras finales 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.

- **Diseño adecuado 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**

Hitos

- ETL completo
- Estructura de datos implementada (DW, DL, etc). Pueden usar algún servicio
- Pipeline ETL automatizado

Entregables

- Documentación

- Stack elegido y fundamentación
 - Flujo de trabajo
-



Semana 3: Data Analytics + ML (para grupos obligatorio)

Esta es la última semana completa para trabajar por lo que deberían tener un grado de concreción de objetivos medio/avanzado.

Se espera que armen un buen dashboard con un análisis de los datos que hayan trabajado. Deben incluir los KPI 's que hayan determinado como importantes para el análisis realizado y preparar un buen 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, aunque no el que vayan a recomendar finalmente.

- **Diseño de Reportes/Dashboards**
- **KPIs**
- **Modelo ML**
- **Modelo de ML en producción**
- **Documentación**
 - Selección del modelo, feature engineering
 - Informe de análisis

Hitos

- Dashboard (diseño)
- Dashboard (conexión a DB)
- ML en producción

Entregables

- Dashboard

- Modelo de Machine Learning (grupos de 5 obligatorio)

✓ Semana 4: Retoques finales

Las demos comienzan a partir del martes por lo que deberían llegar a esta etapa con los proyectos cerrados y usar este tiempo para trabajar en sus presentaciones, storytelling, verificar que todo funcione correctamente y completar documentación.

Sus repos deben quedar ordenados, y con un readme completo presentando el proyecto.

