

## Hito 1 - Diseño

- Para realizar este proyecto debes haber cursado previamente todos los módulos de la carrera Data Science.
- Luego de resolver el hito, comprime la carpeta con lo solicitado y sube el `.zip` a la plataforma.
- El proyecto debe ser desarrollado en equipo, debe estar compuesto por un mínimo de 4 y un máximo de 6 integrantes. **No se podrá trabajar de forma individual**
- En cada hito, los y las estudiantes tendrán una nota individual y una grupal.
- Puntaje Total del Hito: 10 puntos.

### Objetivo

- Definir el equipo de trabajo, el tema de investigación, su motivación y plan de trabajo.

### Descripción

El primer hito del proyecto de la carrera Data Science tiene como objetivo sentar las bases de lo que será el proceso de investigación. Para esto, se evaluarán los siguientes aspectos:

#### Equipo de trabajo

Creación del equipo de trabajo, definición de roles y responsabilidades de cada integrante. Un integrante podría tener más de un rol.

Roles propuestos:

- Líder del equipo.
- Analista de datos.
- Control de calidad y validación de datos.
- Ingeniero/a de modelamiento.
- Ingeniero/a de Big Data.
- Visualización de datos.
- Documentador/a.

Se sugiere asignar funciones según fortalezas de cada integrante del equipo. La evaluación individual considerará el rol definido para cada integrante del equipo.

## Selección del tema y motivación

El éxito del proyecto dependerá del tema de investigación elegido. Es importante que discutan con el equipo docente la viabilidad del proyecto, verifiquen la calidad de los datos y el potencial de obtener resultados en el periodo de tiempo que se dispone.

Cada equipo deberá definir un solo tema a investigar y exponer la motivación en la elección.

El tema puede ser un proyecto propuesto o propio del equipo de trabajo, se incentiva trabajar con la segunda opción.

## Planificación de la investigación

El grupo debe tener previamente definido el o los requerimientos de su proyecto, así como los datos a utilizar. El equipo debe ser capaz de proponer una estrategia analítica con los siguientes puntos: Pregunta de investigación; hipótesis de trabajo; definición del vector objetivo; estrategias analíticas a nivel descriptivo, modelación y predicción de trabajo.

Se deben agregar todos los conceptos que sean relevantes para el desarrollo de la investigación y tener, en caso de ser necesario, las siguientes consideraciones:

- Encriptación de datos sensibles. ¿Qué datos serán transformados?
- Tratamiento de datos faltantes: ¿Qué técnicas de imputación se usarán o se eliminarán los registros?
- Muestreo: ¿Se trabajará con la población o una muestra? Explicar el motivo de la decisión.
- Análisis exploratorio: ¿Existen patrones importantes en los datos?, ¿Qué tipos de datos se tienen?, ¿Qué tipo de dato es el vector objetivo?, ¿Dispersión y distribución de los datos?, entre otros.
- Modelación adecuada para el vector objetivo: clasificación o regresión. Supervisado o no supervisado.

En esta etapa se proponen los primeros modelos candidatos y sus métricas de evaluación, los cuales deben estar justificados y con una agenda de entrenamiento preliminar.

## Entrega

Los puntos anteriormente descritos que componen el hito 1 serán evaluados con un máximo de 10 puntos divididos en un documento y una exposición.

### *Documentación*

Los grupos deben entregar un documento en donde se detalle los roles del equipo de trabajo, tema seleccionado, motivación y planificación de la investigación. **El documento debe tener máximo 5 páginas.**

El puntaje se dividirá de la siguiente manera:

- Creación del equipo y definición de roles y responsabilidades. **(1 Punto)**
- Descripción del tema y motivación. **(1 Punto)**
- Planificación de la investigación: Definir la pregunta de investigación; hipótesis de trabajo; definición del vector objetivo; estrategias analíticas a nivel descriptivo, modelación y predicción de trabajo. Se deben agregar todos los conceptos que sean relevantes para el desarrollo de la investigación. **(3 Puntos)**

### *Exposición*

Presentar la idea del proyecto con al menos la participación del 75% de los integrantes del grupo de trabajo presentes en la clase.

La presentación será evaluada según los siguiente criterios:

- **Evaluación grupal**
  - Planificación de la investigación, definición de la pregunta de investigación, hipótesis de trabajo, definición del vector objetivo, estrategias analíticas, procesamiento de datos y análisis descriptivo. **(2 Puntos)**
  - Calidad de la presentación e información expuesta. **(1 Punto)**
- **Evaluación Individual**
  - Dominio del tema y capacidad de exponer claramente los conceptos. **(2 Puntos)**

### **Consideraciones:**

- Si el o la estudiante no participa de la presentación su puntaje será de **0 puntos**.
- Se dispondrá de 15 minutos para exponer el estado de avance del proyecto



**Importante:** Cada estudiante deberá subir a la plataforma el documento que detalla la planificación del proyecto (documento) y la presentación en un archivo comprimido antes que comience la clase donde será presentado este hito.

## Aprobación del proyecto

Al finalizar la presentación el docente decidirá si el tema es viable para pasar al siguiente hito. En esta etapa podrá sugerir modificaciones o reemplazo total del tema escogido.