

Recopilación y preparación de datos

Ciencia de datos aplicada: Segundo entregable

Objetivo:

A partir de la propuesta de proyecto presentada en el primer entregable, se espera que en este entregable se avance en la adquisición del conjunto de datos que será utilizado así como una primera etapa de análisis exploratorio de datos (EDA). Este entregable permitirá al equipo visualizar y entender la estructura y calidad de los datos, así como dejar preparados los datos para un análisis posterior.

Consigna:

Se propone que cada equipo desarrolle la etapa de adquisición y análisis exploratorio de datos, entregando una **notebook** como la que se presenta a continuación a modo de template: [\[Modelo de entrega con un dataset de referencia\]](#)

Esta entrega incluirá información importante vinculada a los datos que se utilizarán para el entrenamiento posterior de los modelos de *Machine Learning*. Algunos de los principales tópicos que podrá incluir se presentan a continuación:

- **Descripción del dataset:**
 - Origen y formato,
 - Variables incluidas (y significado),
 - Justificación de la elección del dataset.
- **Análisis exploratorio básico**
 - Tipos de variables,
 - Distribuciones, valores faltantes, outliers,
 - Estadísticas descriptivas,
 - Visualizaciones iniciales (gráficos básicos).
- **Diagnóstico y mejoras en la calidad de los datos:**
 - Identificación de problemas: datos faltantes, inconsistencias, duplicados, etc.
 - Decisiones tomadas para limpiar o transformar los datos.

- **Transformaciones realizadas**
 - Normalización o estandarización si corresponde,
 - Codificación de variables categóricas,
 - Generación de nuevas variables,
 - Filtrado o selección de variables si es necesario.
- **Reflexión final**
 - Qué decisiones se tomaron y la justificación,
 - Dificultades encontradas,
 - Sigüientes pasos proyectados.

Formato de entrega: Se propone la entrega de una Notebook (.ipynb o link a Google Colab)

Fecha límite de entrega: 06/10/2025

Evaluación: Se valorará la calidad del diagnóstico, la pertinencia de las transformaciones realizadas, la claridad del código y la documentación.