

PROYECTO DE AULA

FUNDAMENTOS DE CIENCIA DE DATOS

Avances del proyecto – repositorio GitHub

María Bernarda Salazar Sánchez, Ph.D.
Departamento de Ingeniería de Sistemas
Universidad de Antioquia, Colombia
bernarda.salazar@udea.edu.co

Objetivo:

Fomentar la aplicación práctica de los contenidos de las Unidades 2 y 3 del curso, consolidando el inicio del proyecto de aula mediante la creación y gestión de un repositorio en GitHub.

Orientaciones para el entregable I:

El resultado del Análisis exploratorio **deberá ser realizado en un notebook** y publicado en un **repositorio de GitHub** desarrollado para el proyecto. Recuerde que el repositorio debe incluir un archivo **README** que explique claramente autores, título del proyecto, objetivo y distribución del repositorio. Se debe incluir una carpeta con los datos originales y con los datos que ha seleccionado para este proyecto (mínimo 3 variables continuas y 3 discretas).

Para la creación del notebook del análisis exploratorio se pueden apoyar en los ejemplos sugeridos. El análisis requerido en este entregable deberá abordar técnicamente al menos:

1. Base de datos construida o seleccionada

- Mínimo 6 variables: 3 continuas y 3 discretas.
- Carpeta /data/ con datos originales y datos depurados para el proyecto.
- Tabla descriptiva de las variables (nombre, tipo, definición, fuente).

2. Notebook de análisis exploratorio (EDA)

- Estadísticos descriptivos de las variables seleccionadas.
- Visualizaciones univariadas (histogramas, boxplots, gráficas de barras).
- Visualizaciones bivariadas/multivariadas (dispersión, correlaciones, tablas cruzadas).
- Interpretaciones breves que expliquen hallazgos o patrones.

3. Detección y análisis de atípicos

- Aplicar al menos dos técnicas (ej. IQR, Z-score, Grubbs, DBSCAN).
- Explicar los resultados: qué valores fueron detectados, posibles causas y si deben eliminarse, transformarse o mantenerse.

Fecha de entrega:

La fecha máxima de entrega es **martes siete (07) de octubre de 2025**. La entrega se realiza por medio del repositorio del proyecto de aula que has creado para este curso.

Evaluación:

Esta actividad relacionada con el proyecto de aula **corresponde al 15%** de la nota final curso.

Rúbrica de evaluación:

CRITERIO	DESCRIPCIÓN	%
Organización del repositorio en GitHub	Claridad en la estructura del repositorio, orden de carpetas y calidad del README.	15
Base de datos y variables	Cumplimiento en la selección de al menos 6 variables (3 continuas y 3 discretas), organización de la base y documentación de variables.	20
Análisis exploratorio de datos (EDA)	Nivel de desarrollo de estadísticos descriptivos, visualizaciones y hallazgos iniciales.	25
Detección y análisis de atípicos	Aplicación de técnicas de detección, interpretación de resultados y decisiones frente a los datos atípicos.	20
Interpretación y comunicación de resultados	Claridad, profundidad y coherencia de las conclusiones presentadas en el notebook y el README.	20

Recomendaciones generales

- **Organicen la información de referencia:** utilicen tablas o esquemas comparativos para sintetizar la información de los artículos o fuentes que respalden el proyecto.

- **Documenten las variables:** presenten una tabla donde se especifique nombre, tipo (continua o discreta), definición y fuente de cada variable seleccionada para la base de datos.
- **Favorezcan la visualización:** usen diagramas o esquemas cuando sea pertinente, de manera que se facilite la comprensión de la estructura del proyecto y de las relaciones entre variables.

Penalizaciones

- **Entrega tardía:** cada hora de retraso posterior a la fecha y hora límite implicará una deducción de 0,5 unidades sobre la calificación del entregable.
- **README incompleto o ausente:** la falta de un archivo README, o su entrega sin cumplir las indicaciones dadas (autores, título, objetivo y estructura del repositorio), implicará una deducción de 0,5 unidades.

Ejemplos de notebooks con EDA

Para orientar el desarrollo del entregable, se sugieren los siguientes recursos prácticos que presentan ejemplos de análisis exploratorio de datos:

1. [Tutorial EDA para Data Science](#) – Introducción con ejemplos básicos de exploración y visualización.
2. [Intro to Exploratory Data Analysis in Python](#) – Explicación progresiva del proceso de EDA con código comentado.
3. [Step by Step Data Preprocessing & EDA](#) – Guía paso a paso que combina limpieza, preprocesamiento y análisis exploratorio.