

## ACTIVIDAD SESIÓN FUNDAMENTOS DE APRENDIZAJE DE MÁQUINA

La empresa **XYJ Analytics**, dedicada al análisis de datos inmobiliarios, te ha contactado para desarrollar un modelo de *Machine Learning* que permita **predecir el precio de viviendas** a partir de un conjunto de datos recopilado en distintas ciudades.

El equipo de XYJ ha recopilado información como:

- Superficie en m<sup>2</sup>
- Número de habitaciones
- Antigüedad de la vivienda
- Barrio
- Precio de venta

Antes de crear modelos avanzados, quieren que realices una **exploración y análisis inicial del conjunto de datos**, prepares los datos y construyas un **modelo básico de regresión** para demostrar su aplicabilidad.

### INSTRUCCIONES:

#### 1. Carga y Exploración Inicial del Dataset (1 punto)

- Carga el archivo viviendas.csv.
- Muestra las primeras 5 filas del DataFrame.
- Describe la estructura del dataset (.info(), .describe()).

#### 2. Limpieza de Datos (1 punto)

- Detecta y elimina o imputa los valores faltantes.
- Asegúrate de que los tipos de datos estén correctamente definidos.

#### 3. Análisis Exploratorio de Datos () (2 puntos)

- Realiza gráficos para explorar las relaciones entre:
  - Superficie y precio

- Número de habitaciones y precio
- Detecta posibles outliers visualmente.

#### 4. Codificación de variables categóricas (1 punto)

- Convierte la variable barrio a variables numéricas utilizando One-Hot Encoding.

#### 5. División del Dataset (1 punto)

- Separa el dataset en un conjunto de entrenamiento (80%) y uno de prueba (20%).

#### 6. Entrenamiento de un Modelo de Regresión Lineal (2 puntos)

- Entrena un modelo de regresión lineal usando scikit-learn.
- Utiliza como variables predictoras: superficie, habitaciones, antigüedad y los barrios codificados.
- Muestra los coeficientes del modelo.

#### 7. Evaluación del Modelo (2 puntos)

- Calcula el **Mean Squared Error (MSE)** y el **R<sup>2</sup> (coeficiente de determinación)** sobre el conjunto de prueba.
- Comenta brevemente si el modelo tiene un buen desempeño.

#### INSTRUCCIONES ADICIONALES:

- Puntos totales = 10 puntos.
- Comprimir el archivo en formato .zip o .rar.
- Subir el archivo a la plataforma.
- Descarga el material complementario viviendas.csv