

# Tarea 1

Juan Zamora O.

Septiembre 2023

## 1 Objetivo

El objetivo de esta actividad es aplicar de manera exitosa los contenidos revisados respecto del procesamiento, medidas resumen de datos y su interpretación.

## 2 El conjunto de datos Boston Housing.

El conjunto de datos Boston Housing es un conjunto de datos ampliamente utilizado en el campo del análisis de datos y la estadística para estudiar y analizar la relación entre distintas características de las viviendas en el área metropolitana de Boston y sus precios. Este conjunto de datos fue creado por el profesor Harrison y David L. Rubinfeld en 1978, y se ha convertido en un recurso fundamental para la enseñanza y la investigación en el ámbito de la regresión y el aprendizaje automático<sup>1</sup>.

El conjunto de datos contiene 506 registros, cada uno representando una vivienda, más 14 variables, incluyendo 13 características (predictores) y una variable objetivo (MEDV):

---

<sup>1</sup>Descargar desde archivo adjunto a este documento.

Item	Descripción
CRIM	Tasa de criminalidad per cápita por ciudad
ZN	Proporción de terreno residencial dividido en zonas para lotes de más de 25,000 pies cuadrados
INDUS	Proporción de acres comerciales no minoristas por ciudad
CHAS	Variable ficticia de Charles River (1 si el tramo limita con el río, 0 en caso contrario)
NOX	Concentración de óxidos nítricos (partes por 10 millones)
RM	Número promedio de habitaciones por vivienda
AGE	Proporción de unidades ocupadas por sus propietarios construidas antes de 1940
DIS	Distancias ponderadas a cinco centros de empleo de Boston
RAD	Índice de accesibilidad a carreteras radiales
TAX	Tasa de impuesto a la propiedad de valor total por \$10,000
PTRATIO	Proporción alumno-profesor por localidad
B	$1000(Bk - 0.63)^2$ , donde Bk es la proporción de personas de ascendencia afroamericana por ciudad
LSTAT	Porcentaje de población de menor status
MEDV	Valor medio de las viviendas ocupadas por sus propietarios en miles de dólares

### 3 Modalidad de trabajo y entregable

Las/os estudiantes deberán organizarse en equipos de máximo 3 integrantes. Solo uno de los integrantes deberá hacer entrega del documento final a través del Aula. Este documento deberá ser un archivo HTML generado a partir de un cuadernillo (notebook) Jupyter. Deberá indicar claramente los/as integrantes y tener enumeradas las respuestas. El código asociado a cada una deberá estar escrito en una y solo una celda de código y además mostrar la salida correspondiente. A su vez, junto a cada código y salida asociada, deberá agregar una celda de texto donde se entregue de manera clara y breve la interpretación de la salida observada. Ejemplos de este ejercicio de análisis han sido abordados en la ayudantía.

### 4 Actividades

1. Cargar el conjunto de datos 'Boston Housing' y mostrar las primeras 5 filas del conjunto de datos

2. Generar Estadísticas descriptivas básicas (una de localización y una dispersión)
3. Crear un gráfico de dispersión para todas las variables
4. Calcular la correlación de Pearson entre las variables
5. Crear un gráfico de correlación de Pearson para cada par de variables
6. Visualizar la matriz de correlación con un mapa de calor
7. Crear un pairplot para visualizar las relaciones entre las variables más correlacionadas
8. Crear un histograma para cada característica
9. Crear un diagrama de cajas para cada característica
10. Crear un gráfico de dispersión para ver cómo se relaciona el valor medio de la vivienda (MEDV) con las variables más correlacionadas
11. Calcular la asimetría y curtosis de las variables
12. Crear gráficos de densidad para cada característica
13. Separar los datos en dos grupos basados en la variable CHAS
14. Calcular la media y varianza para cada grupo
15. Calcular la media y varianza total
16. Calcular la varianza intra-grupo
17. Calcular la varianza inter-grupo
18. Interpretar qué indican los valores de ambas varianzas