The Government of Massachusetts is planning a real estate development project and wants to understand the factors influencing house prices in Boston. They have historical data of various neighborhoods, including crime rate, average number of rooms, percentage of lower-status population, and accessibility to highways.

As a Data Scientist, you are assigned to build a prediction system that will help estimate the median value of homes (MEDV) based on these characteristics. The authorities want reliable predictions before launching their housing policy.

Dataset Link: https://www.kaggle.com/datasets/schirmerchad/bostonhoustingmlnd

1. Problem Implementation (5 Marks)

- Load and explore the Boston Housing dataset.
- Preprocess the data: handle missing values, separate features (X) and target (y), and scale features.
- Split the dataset into training and testing sets.
- Build a Random Forest Regressor to predict housing prices.
- Evaluate the model using $R^2$ score and error metrics (MSE, MAE).

2. Visualization and Inference (3 Marks)

Perform and interpret visualizations such as:

- Plot a histogram of the target variable (MEDV).
- Scatter plot of LSTAT vs MEDV.
- Heatmap of correlations between features and MEDV.
- Write two insights explaining which factors strongly influence housing prices.

3. Test Case Implementation (2 Marks)

Create at least two test cases using sample digit images (or rows from dataset):

- Case 1: A neighborhood with high rooms per dwelling, low crime rate, and high accessibility → predict house price.
- Case 2: A neighborhood with high poverty percentage (LSTAT), low average rooms, and high crime rate → predict house price.
- Show model predictions for both.