Boston Housing Price Analysis Report

---

1. Introduction

The Boston Housing dataset is a widely used dataset in regression analysis, particularly for predicting house prices based on various factors. This analysis aims to understand the key factors influencing housing prices and build a predictive model for median house value (MEDV) using statistical and machine learning techniques.

2. Dataset Overview

- **Number of Observations:** 506

- **Number of Features:** 14 (13 predictors + 1 target variable)

- **Target Variable:** MEDV (Median value of owner-occupied homes in $1000s)

Feature Description:

Feature	Description
CRIM	Per capita crime rate by town
ZN	Proportion of residential land zoned for large lots
INDUS	Proportion of non-retail business acres per town
CHAS	Charles River dummy variable (1 if tract bounds river; 0 otherwise)
NOX	Nitric oxide concentration (parts per 10 million)
RM	Average number of rooms per dwelling
AGE	Proportion of owner-occupied units built before 1940
DIS	Weighted distances to five Boston employment centers
RAD	Index of accessibility to radial highways
TAX	Full-value property tax rate per $10,000
PTRATIO	Pupil-teacher ratio by town
B	Proportion of African-Americans by town

Feature	Description
LSTAT	% lower status of the population
MEDV	Median value of owner-occupied homes in $1000s

3. Exploratory Data Analysis (EDA)

3.1 Distribution of Key Variables

- The target variable, **MEDV**, shows a slightly right-skewed distribution with a peak around $20,000.

- The **RM** feature (average number of rooms per dwelling) appears normally distributed, with most values ranging between 5 and 8.

- **LSTAT** (% lower status population) shows a left-skewed distribution, indicating that a majority of the areas have lower LSTAT values.

3.2 Correlation Analysis

A correlation heatmap was generated to identify relationships between variables. Key correlations include:

- **RM** has a strong positive correlation with **MEDV** (r = 0.7), indicating that larger homes tend to be more expensive.

- **LSTAT** has a strong negative correlation with **MEDV** (r = -0.74), suggesting that areas with a higher percentage of lower-status individuals tend to have lower house prices.

- **NOX** and **DIS** are moderately negatively correlated (r = -0.7), indicating that homes farther from employment centers have better air quality.

3.3 Boxplots and Histograms

- Boxplots for **MEDV** across different **CHAS** values showed that homes near the Charles River tend to have higher median values.

- Histograms for features like **AGE**, **TAX**, and **PTRATIO** provided insights into the distribution of housing characteristics across different towns.

4. Statistical Analysis

4.1 T-Test for CHAS

A two-sample t-test was conducted to determine if there is a significant difference in median house values for homes near the Charles River vs. those not near it.

- **Result:** The p-value was less than 0.05, indicating a statistically significant difference.

4.2 ANOVA for ZN Categories

An ANOVA test was performed to check if zoning (ZN) significantly affects house prices.

- **Result:** Zoning showed a significant effect on house prices ($p < 0.05$).

4.3 Multiple Linear Regression Model

A multiple linear regression model was built to predict **MEDV** using all available predictors.

- **Key Results:**
  - Adjusted R-squared: 0.73
  - Significant predictors: **RM**, **LSTAT**, **PTRATIO**, and **TAX**
  - Coefficients indicate that an increase in the average number of rooms leads to a significant increase in house prices, while higher LSTAT values decrease house prices.

5. Findings and Insights

- **RM** and **LSTAT** are the most significant factors affecting house prices in Boston.
- Homes near the Charles River tend to have higher median values.
- Areas with higher accessibility to highways (higher RAD values) tend to have slightly lower house prices.
- Increasing the number of rooms and improving socioeconomic conditions could significantly increase property values.

6. Recommendations

- Future development projects should focus on building larger homes with more rooms, as **RM** showed a strong positive correlation with house prices.
- Urban planners may consider improving socioeconomic conditions in low-status areas to boost property values.