

## Context

The problem at hand is to predict the housing prices of a town or a suburb based on the features of the locality provided to us. In the process, we need to identify the most important features in the dataset. We need to employ techniques of data preprocessing and build a linear regression model that predicts the prices for us.

## Data Description

Each record in the database describes a Boston suburb or town. The data were drawn from the Boston Standard Metropolitan Statistical Area (SMSA) in 1970. Detailed attribute information can be found below-

Attribute Information (in order):

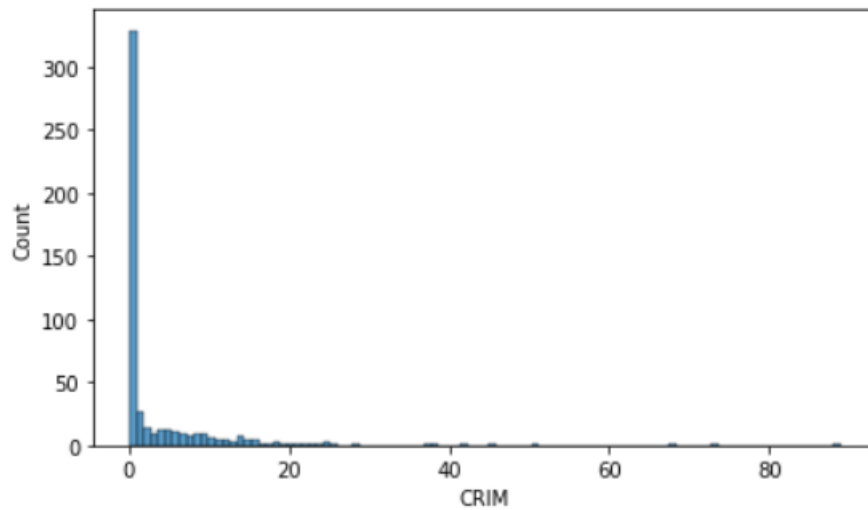
- CRIM: per capita crime rate by town
- ZN: proportion of residential land zoned for lots over 25,000 sq. ft.
- INDUS: proportion of non-retail business acres per town
- CHAS: Charles River dummy variable (= 1 if tract bounds river; 0 otherwise)
- NOX: nitric oxides concentration (parts per 10 million)
- 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 dollars
- PTRATIO: pupil-teacher ratio by town
- LSTAT: %lower status of the population
- MEDV: Median value of owner-occupied homes in 1000 dollars.

1. **Load the dataset**
2. **Check the shape of the dataset**
3. **Get the info regarding column datatypes**
4. **Get summary statistics for the numerical columns**
5. **Exploratory Data Analysis**
6. 

---
7. **Plot the distribution plots for all the numerical features and list your observations.**

**Hint : Use Seaborn**

**One of the plot that you will get would be as shown below :**

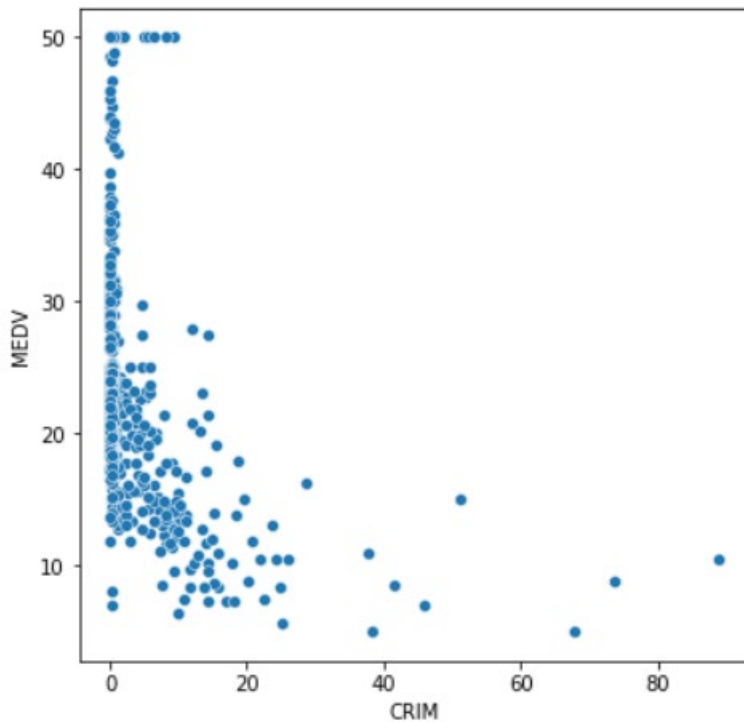


You will get similar plots for all the columns.

**Comment on the observations**

8. Plot the scatterplots for features and the target variable MEDV and list your observations.

One of the plot after running the cell would look like below :



You will get similar plots for all the columns for target variable.

**9. Plot the correlation heatmap and list your observations.**

**10. Split the dataset**

**11. Model Building (sklearn)**

**12. Get the scores on training and Testing data set and comment on it.**

**13. Get the coefficient and comment on it.**