Housing Data : Linear Regression

BY

CHIRAG CHAUDHARI

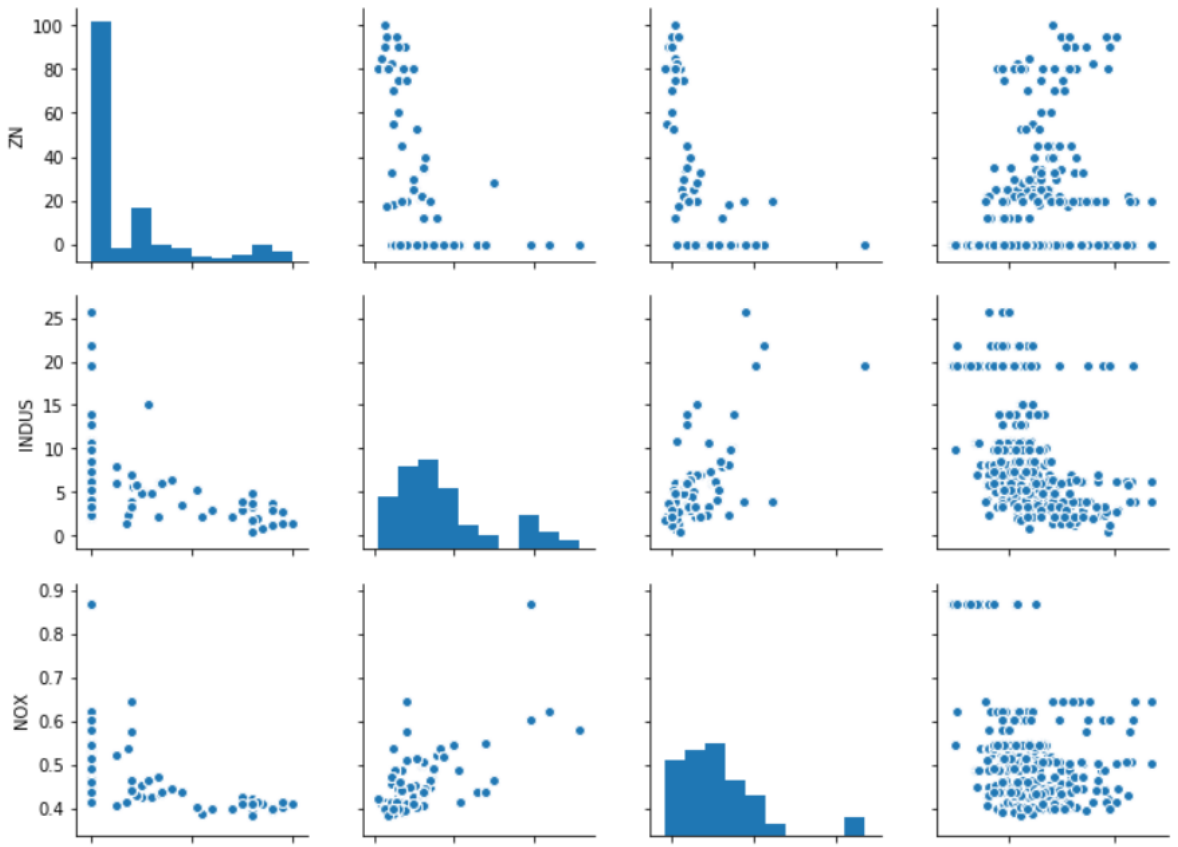
Objective:

The data set contains basic housing data such as proportion of residential land per zone, average number of rooms per dwelling etc. Linear regression is to be plotted to find out the relation between different variables considered in dataset.

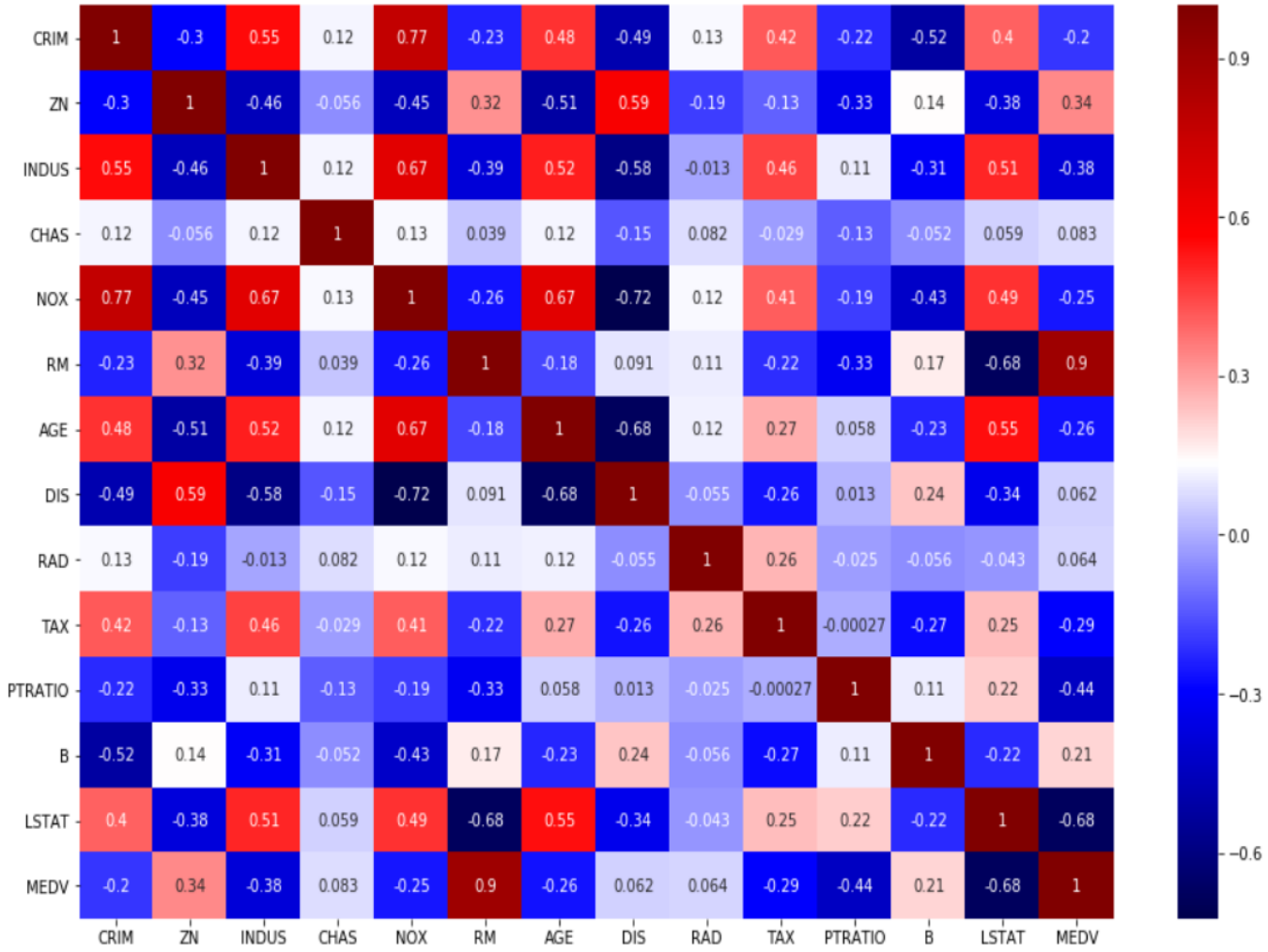
Data Understanding:

Variable Name/ Column Name	Description
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 pertown
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 prior to 1940
DIS	weighted distances to five Boston employment centres
RAD	index of accessibility to radial highways
TAX	full-value property-tax rate per 10,000
PTRATIO	pupil-teacher ratio by town
B	1000(Bk - 0.63)^2 where Bk is the proportion of blacks by town
LSTAT	Percent lower status of the population
MEDV	Median value of owner-occupied homes in1000's

Exploratory data analysis:



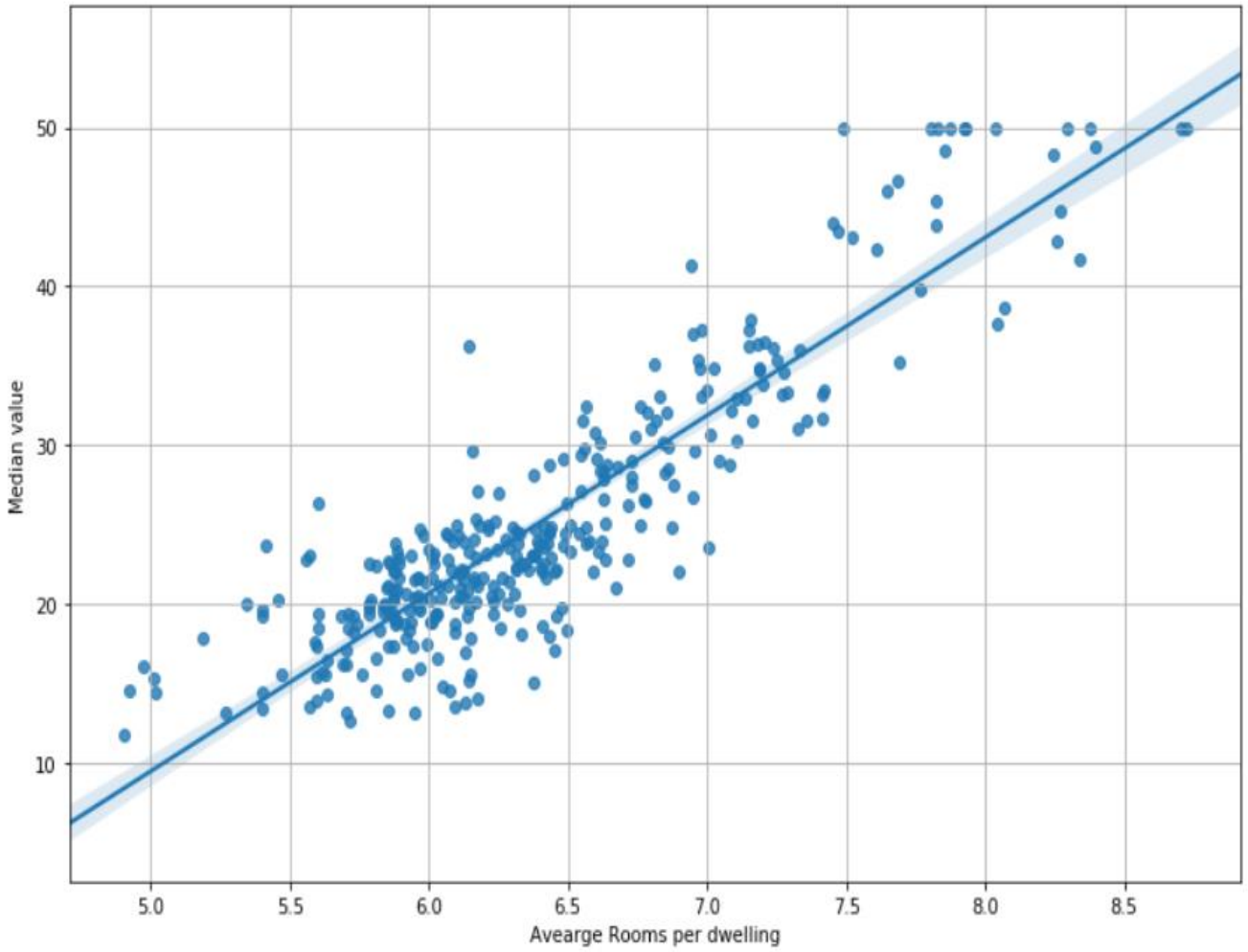
Scatter pair plots



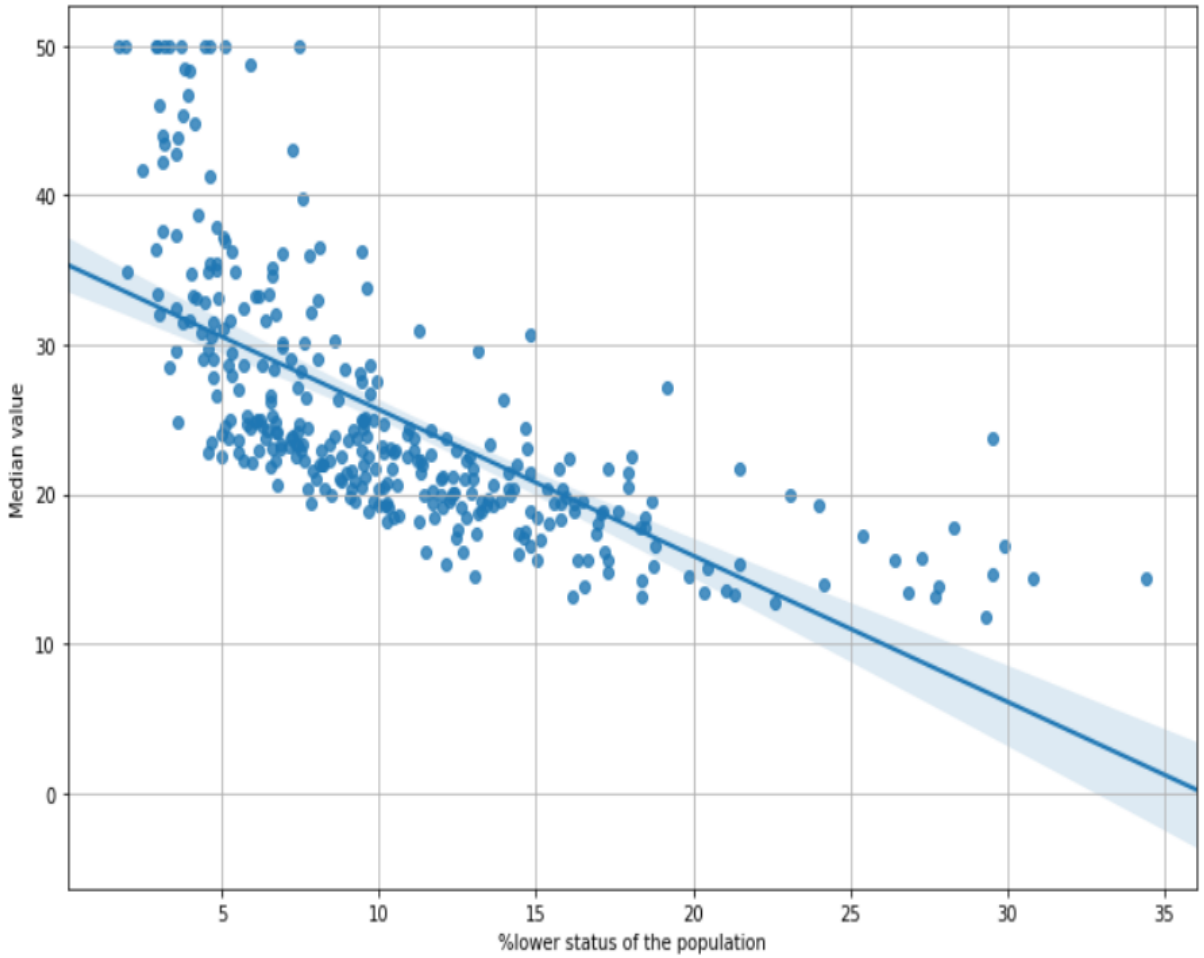
Correlation

Linear Regression:

Linear regression is useful for finding relationship between continuous variables. One is predictor or independent variable and other is response or dependent variable. Linear regression looks for statistical relationship but not deterministic relationship. Relationship between two variables is said to be deterministic if one variable can be accurately expressed by the other. For example, using temperature in degree Celsius it is possible to accurately predict Fahrenheit. Statistical relationship is not accurate in determining relationship between two variables. For example, relationship between height and weight.



Linear Regression plot between Average Rooms per Dwelling and Median Value



Linear Regression plot between %lower status of the population and Median Value

Thank You