**Boston Hosue Rent:**

**CSV file-**  'https://raw.githubusercontent.com/training-ml/Files/refs/heads/main/boston\_house\_rent.csv

**Features –**

| **CRIM** | **ZN** | **INDUS** | **CHAS** | **NOX** | **RM** | **AGE** | **DIS** | **RAD** | **TAX** | **PTRATIO** | **B** | **LSTAT** |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |

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 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 10.000usd

PTRATIO pupil-teacher ratio by town

B 1000(Bk-0.63)^2 where Bk is the proportion of blacks by town

LSTAT % lower status of the population

**Label –** Price

**Shape of the dataset -** 506 rows × 14 columns

As checked dataset rent is having continuous values, so we need to run the model on Regression model

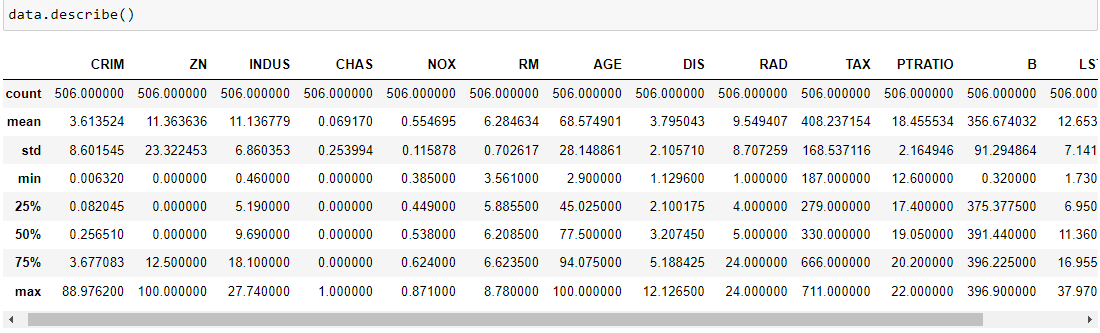
So for this model we are going to use Regressor model

* data.sample(n=8) checked and confirmed all the data having Continuous data

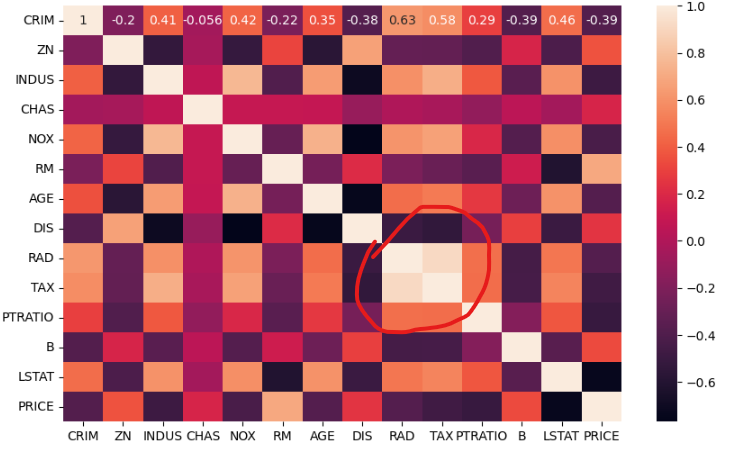
Checking the Nulls

* data don’t have any nulls value, data is clean.

Describe the data

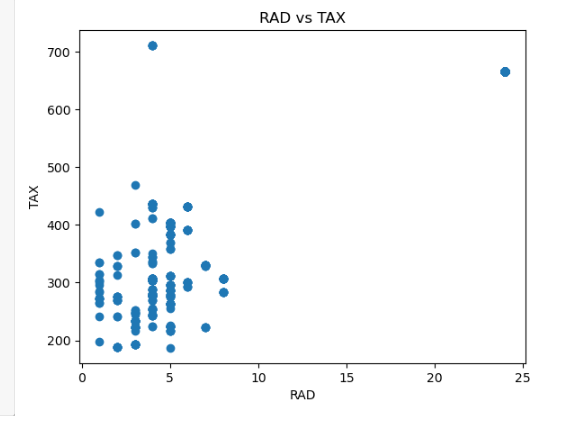


Heatmap:



* As checked Heatmap there is correlation with the label of RAD and TAX, so if it is really there we can delete one of them, for proofing we are plotting scatter plot.
* If scatter plot of x axis increase then y axis also be increase, that time only we can delete one of them, (lowest value will be delete)

Scatter Plot:



* As per scatter plot we didn’t see as much relationship in between RAD and TAX, so we are not deleting any column and not taking any risk.
* there is not a close relationship

Model building using AdaBoost:

* Training the module using and initializing Ada Boost
* Accuracy of the training data is - R2 score for train is 0.909059
* Model evaluation on test data - R2 score for test is 0.795472

As checked there is room for improvement in test data but in train data we have trying to improve that accuracy

#using Randomized searchCV

Giving parameters for Randomized SearchCV

* {'n\_estimators':[47,50,60,70], 'learning\_rate':[0.09,0.25,0.30]}

Predicting the y\_test data for R2 score

After using the Randomized SearchCV the testing score is increases as 3-4 %, now the R2score is

* Model evaluation on test data after tuning- R2 score for test is 0.8367