|  |  |  |
| --- | --- | --- |
|  | | |
| HOUSE PRICE PREDICTION | | |
|  |  |  |

PRECITION OF HOUSE PRICE USING PYTHON

# IMPORTING LIBRARIES

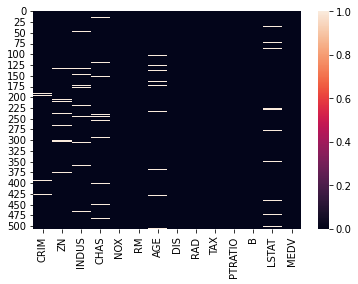
## Initially we have to import libraries ruquired for the project avaluation like numpy ,pandas ,visualization tools like seaborn,matplotlib,and required libraries imported according to the steps required.

## Data exploration

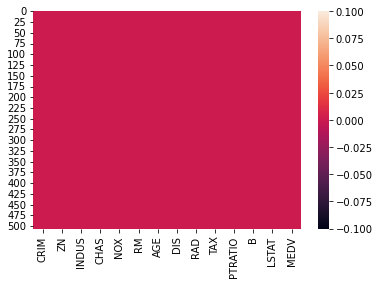
Now I have done explonatory data analysis,using pandas library and imported dataset which is in csv format later described by pandas library, and found the datatypes of the rows in the dataset.

DATA PREPROCESSING

Here I handled data cleansing using pandas library and simple imputer , here we will face little challenges which is good to know . For first four rows I used simple stats model to replace null values and replaced with mean of about its own row and the values converted into object type and remaining rows I tried with stats model but kernel was died while doing that so I used simple imputer technique it performed well. Here first row was changed into intiger type using simple imputer technique but remaining rows was not changed so I dropped. Here I willshow before and after null values.

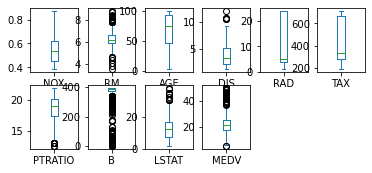


After removing null values:



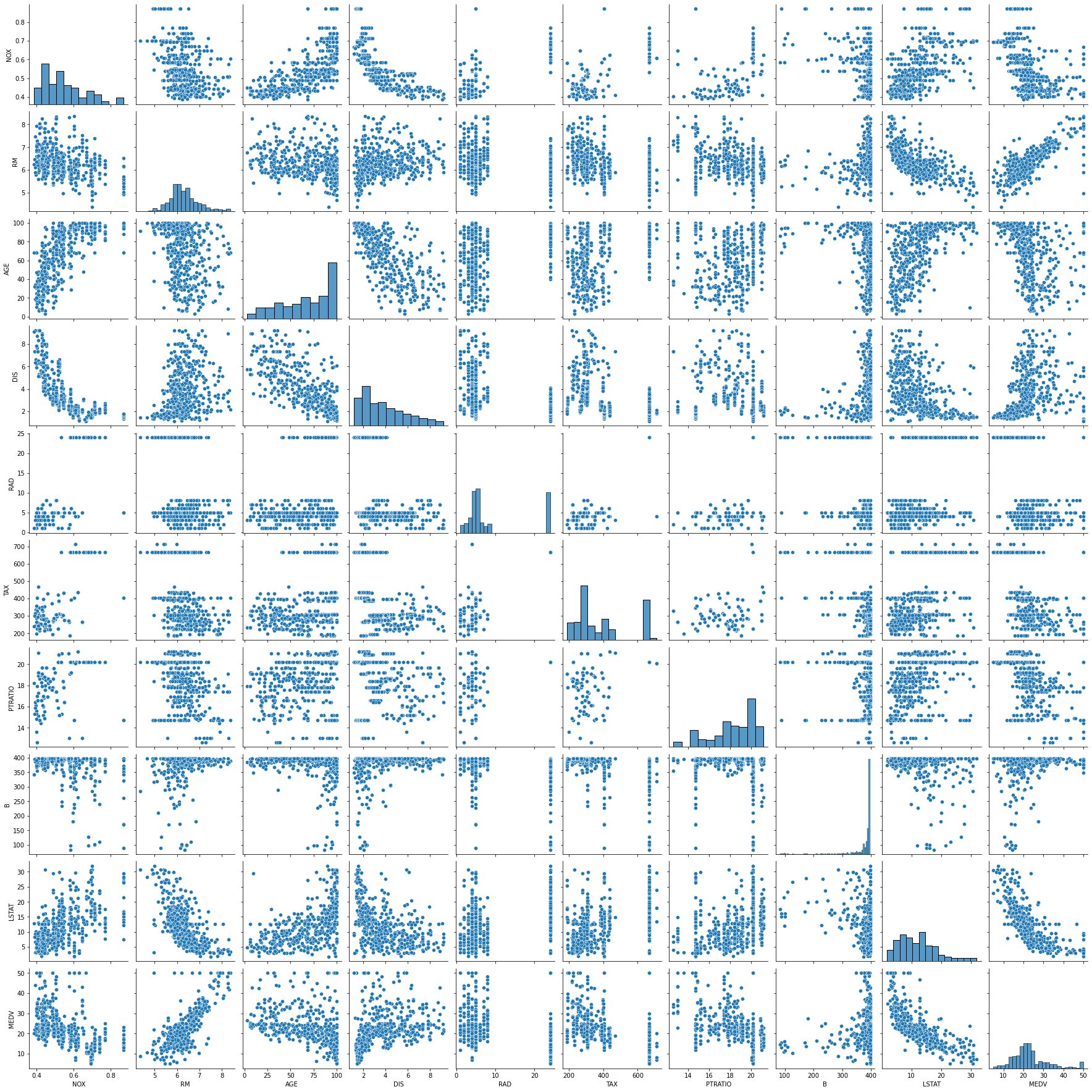
Removing outliers:

While removing outliers I used zscore technique to perform model more better here I changed few rows into integer using label encoder technique and assigned threshold value =3 and values above 3 has removed and below 3 was taken into account. Now I will show the values before outlier and after outliers.



After cleansing we can clearly observe that rows reduced from 506 to 465

Now im importing all the pairplots



Here now the splitted into training data and testing data using model selection

We can see that in the dataset

Model builing

Here I have imported required library for linear regression later I performed techniques and found the score about 77

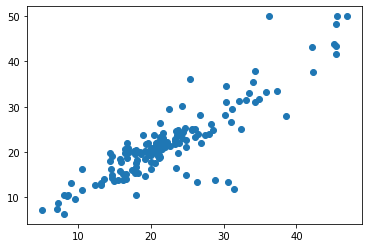
Next I moved to lasso and ridge regression techniques , both will get near to values of 77 percentage later moved to hyperparameter tuning models for lasso and ridge models both was reduced then I decided to not got for this models next I came to xgboost method with tuning method I got best score about 82 so I tried with individual values and got .99 for nearly n\_estimators of all values.

Now I came to xgboost method with 300 and calculated the scores and moved to decision tree regressor and support vecotor regressor and got best predictions for dcr of 100 and calculated remaing values

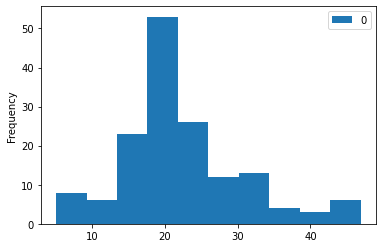
error:  
mean\_abosolute\_error 3.5454545454545454  
mean\_squared\_error 41.80961038961039  
root\_mean\_squared\_error 6.466035136744185

And r2\_score= 0.3976125595268577

Here I will show the scatter plot for actual value and predicted value



Histogram plot for the predicted value



Here I created dataframe for both actual and predicted values here I used crossvalscore,gridsearchcv with xgboost method and confusion for gridsearch cv with xgboost when trying to n\_estimators individually getting more predictions than tuning parameter same repeated to decision tree regressor. Here I recommond decision tree regressor with criterion mse and friedman\_man got great prediction about 100 percent.