**Process and Results**

Process

The process of model building followed the following steps

* Importing Libraries and Dataset
* Pre-Processing the Dataset
* EDA
* Pipeline and Model Building

Importing Libraries and Dataset

Various libraries for the model building pre-processing, EDA and others were loaded at each and every step of the code the libraries used were pandas, Numpy, Sklearn etc.

Pre-Processing the Dataset

The nan values were dropped the description of each values were noted and various important components of the dataset were included. The Nasdaq value was converted to float as it was in String type.

EDA

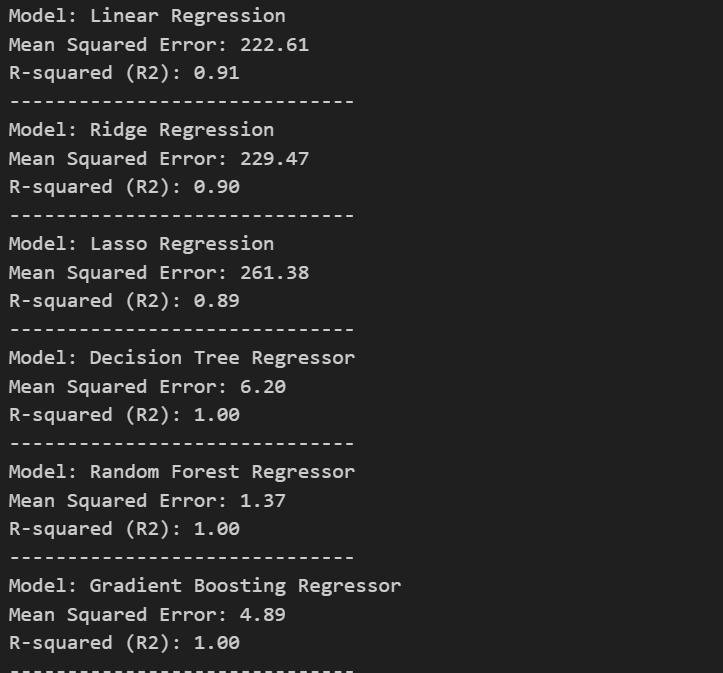
The trends of the all the columns were understood using the line plot. The heatmap was potted for understanding the correlation between the variables

Model Building

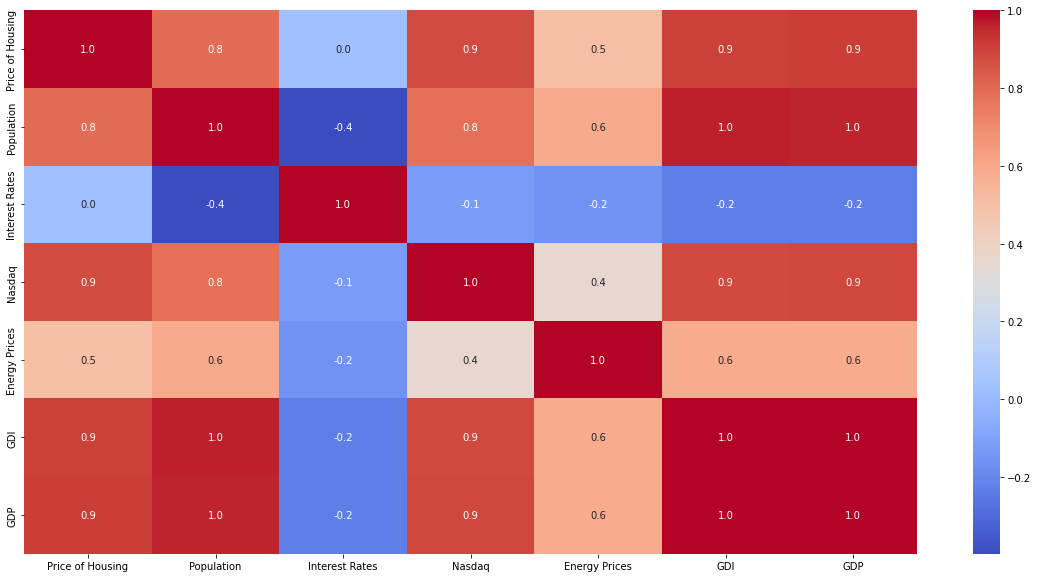
Before the model fitting the data was splitted into train and test. The pipeline was created which include first the scaling and then the model fitting was carried out. A loop was used to fit various models and to get the Mean Squared Error and R-squared value. The GDP values was removed from the dataset as it was tampering the score values. Thinks so due to curse of dimensionality. The various models scores are given in the Result section.

Results

Model scores



Correlation Map:



The correlation plot basically helps us to tell that there is a positive correlation between the various factors that are included in our dataset from the population, energy prices, GDP, GDI but there is not so correlation between interest rates and Housing Prices that may be due low availability of data or due to its small time effect on the housing prices. Other all factors have a positive correlation with the housing price and effect it high as most of the values are over 0.5 value. For model section the model scores are given and the Random forest regressor performs the best. The score of the model was tampered due to the GDP column at the first but later then was adjusted.

Conclusion

The various factors such as the GDP, GDI, Population, Interest Rates and Energy prices have positive correlation with the housing prices in America