**INSY 5378: Project 3**

**Prediction of House price Index using Crime Rate Data**

**Group Members:**

Sathya Narayanan Manivannan

Sandeep Ramesh

Vaithiraj Sokkalingam

**Introduction:**

The primary goal of the project is to predict the house price index using the crime rate data. Our aim was to assess the crime rate and relate to the rise and fall of house price index.

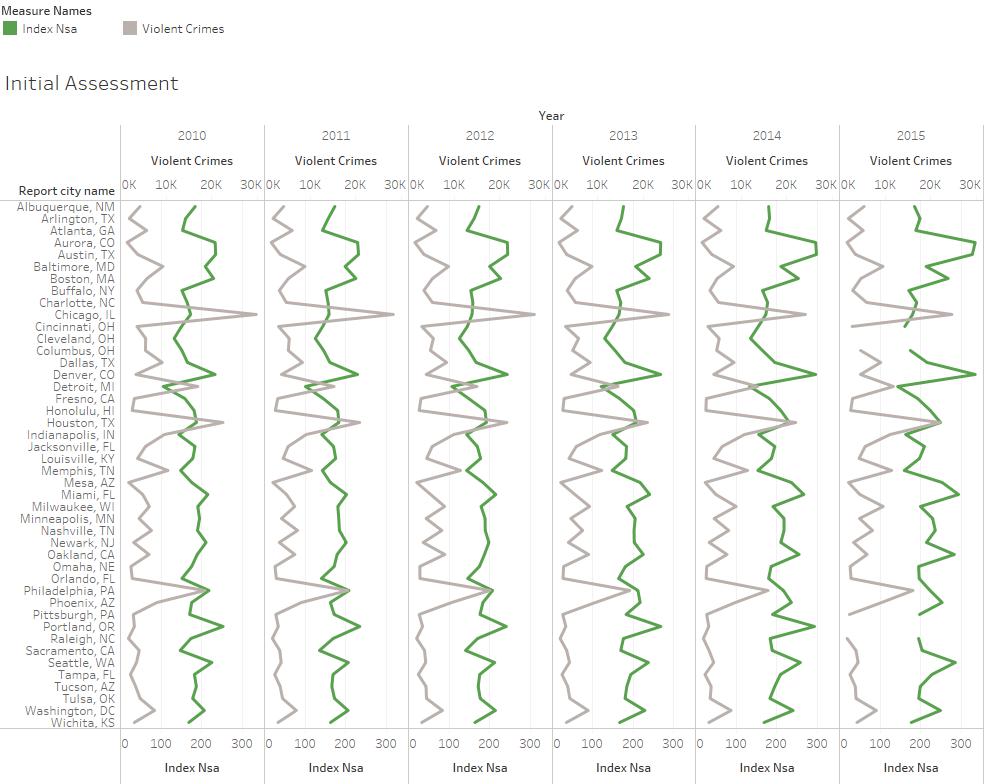
**Data Collection and Preparation:**

Data was collected from <https://www.kaggle.com/marshallproject/crime-rates> and <https://catalog.data.gov/dataset/fhfa-house-price-indexes-hpis> for the initial preparation.

Year and City names were used as a common column between the 2 data sets to merge. This was done in Python and the resulting set of data was converted into a Pandas data frame for manipulative operations. Missing values were dropped to maintain consistency.

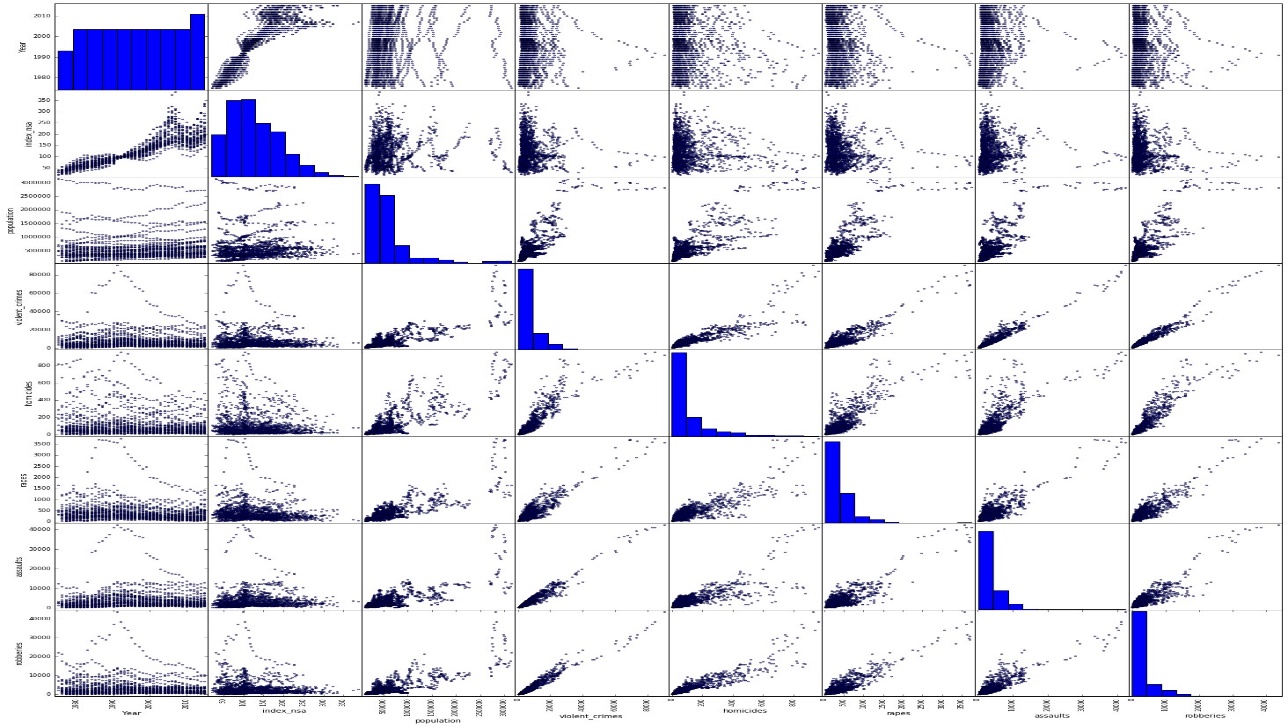
Now the data set is a collection of feature names like Year, City Name, Population, Rapes, Robberies, Assault, homicides.

**Initial Assessment:**

****

Through this initial assessment we can tell that decrease in violent crimes results in an increase in house price index and vice versa.

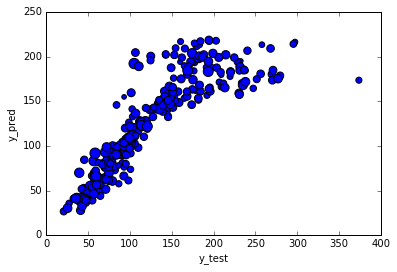
**Scatter Matrix:**



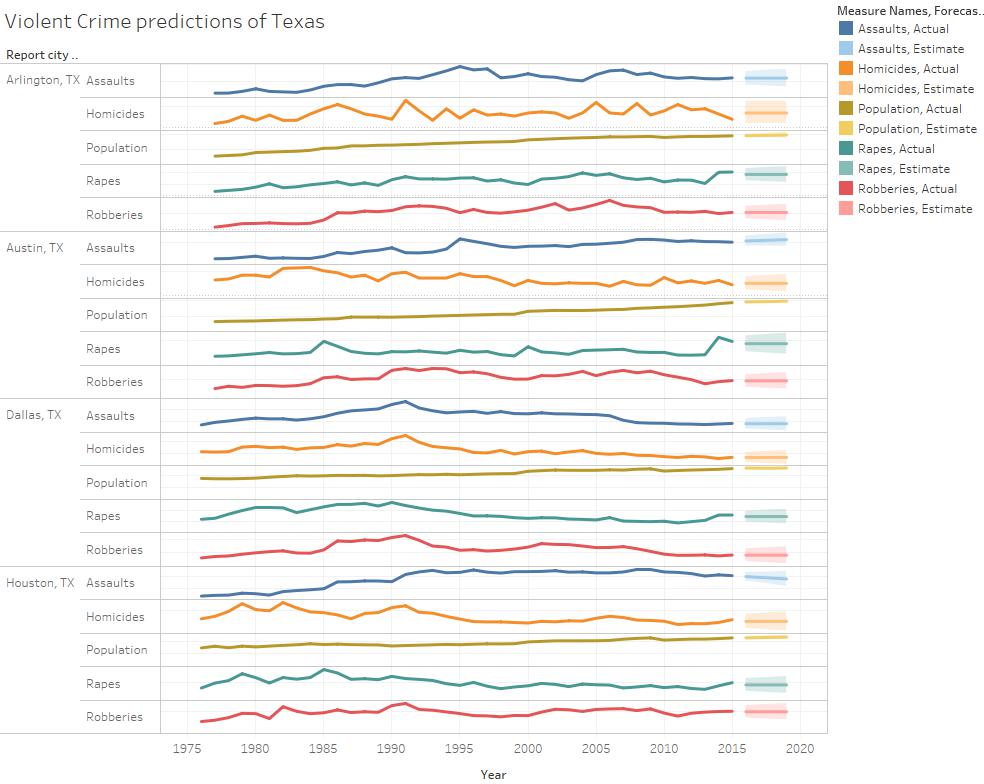
Through this scatter matrix we can infer that HPI does not have linear relationship with other features and hence will not give a high prediction error. Also Violent Crimes feature is a sum of rapes, robberies, assaults and homicides. Hence we eliminate violent crimes feature from our data set.

**Model Training and Testing**:

We have used OLS to train and test our model. We have also split our data to 80:20 test size to train and test our model. After fitting our model through Linear Regression , we have used 20% data to test our model prediction. R square score of 76% tells us that the accuracy is good enough.

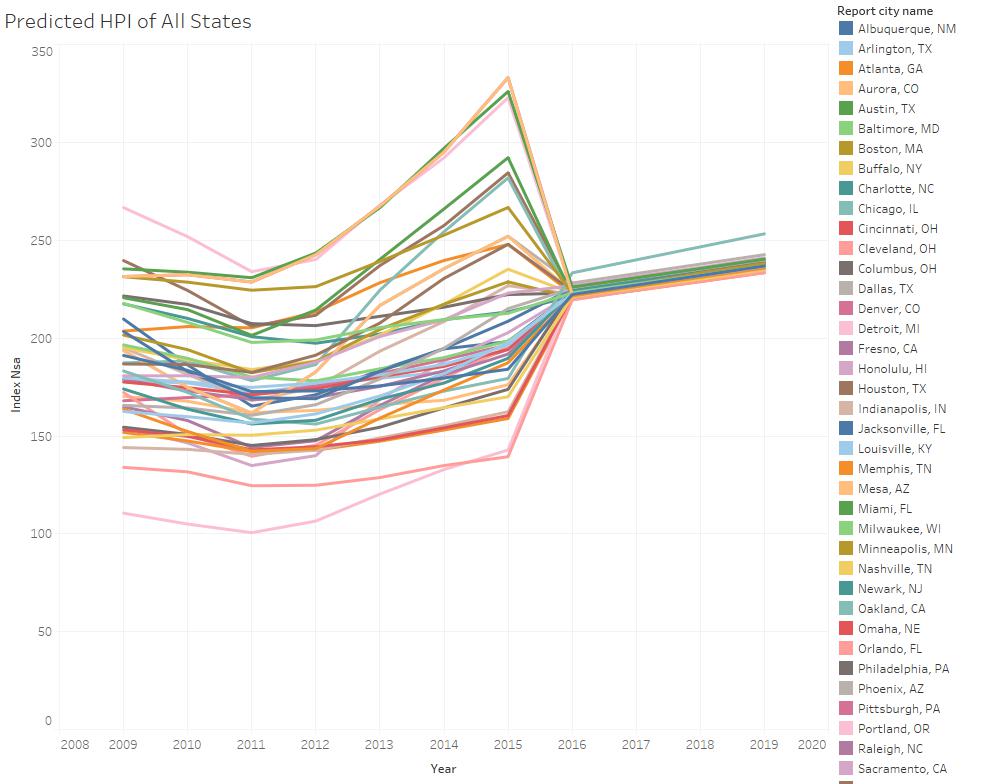


Above plot depicts the y\_test vs y\_pred values. As the values are in a linear progression , we can safely say most of the values have been predicted correctly.



**Prediction of data from 2016 to 2019:**

Tableau was used to predict our values of our independent features from 2016 to 2019. We have used this data again to predict the HPI for 4 step ahead forecast using the coefficients we obtained from training our model. This was done manually in MS Excel and the data was fed again into Tableau to create a better visualization.



Through these predicted values we can say that the HPI below the range of 250 is safe to be invested and bought.

**Inferences**:

* Through this experiment, we have been able to conclude that we can get an approximation of the HPI using crime rate data.
* Increase in HPI is directly proportional to increase in population and decrease in violent crimes.
* Alternatively, drop in HPI could be a possibility of surge in violent crimes.
* Predictions can be used by Investors or Home buyers for better business decisions.