Applying Regression Model to Predict HPI using Mortgage, Inflation and Unemployment Rates

Masroor

Abstract

While there are many factors contributing to HPI - Housing Price Index (Measure of House Price changes), this project studies the relationship between Mortgage Rate, Inflation and Unemployment rates on HPI and uses these metrics to predict HPI using Regression model. This project used Historical data of '30 year fixed rate' Interest rate from freddiemac.com, Historical HPI data from data.gov and Inflation rate & Unemployment rate data from kaggle.com.

You may be thinking: If mortgage rate rise, HPI must fall, however as per the analysis using the data mentioned above, there is good but not strong correlation between mortgage and HPI. Correlation of unemployment rate on HPI is found to be weak.

Combining all three metrics (Mortgage Rate, Inflation Rate and Unemployment Rate), we then build a Regression model and examine the accuracy of Linear and Decision Tree Regression models using RMSE (Root Mean Squared Error) method.

Motivation

What comes to your mind when you think of housing price and factors affecting the price? I am sure you would be thinking of terms like Economy, GDP, Mortgage Rate, Job Market, Demand & Supply, Location.. Well, you are right !!! and it is one of the most complicated and dynamic domain of research. Housing and it's role in economy became even more significant after the recent economic recession in 2008, as many economists credited housing bubble as primary factor that drove America into recession.

As a student of Data Science, I wondered how factors like Mortgage Rate, Inflation Rate and Unemployment Rate are related to housing price and could these be used to build a machine learning model and predict HPI.

Dataset(s)

- **U.S Mortgage Rate Data:** Conventional, Conforming 30-Year Fixed Rate Mortgage Series Since 1971 - From Freddie Mac

  http://www.freddiemac.com/pmms/pmms_archives.html

- **FHFA House Price Indexes (HPIs):** This data is obtained by reviewing repeat mortgage transactions on single-family properties whose mortgages have been purchased or securitized by Fannie Mae or Freddie Mac since January 1975.

  https://catalog.data.gov/dataset/fhfa-house-price-indexes-hpis

- **Inflation and Unemployment Rates:** This dataset includes data on the economic conditions in the United States on a monthly basis since 1954. The unemployment rate represents the number of unemployed as a seasonally adjusted percentage of the labor force. The inflation rate reflects the monthly change in the Consumer Price Index of products excluding food and energy.

  https://www.kaggle.com/federalreserve/interest-rates

- **GEO JSON and KML data for the United States:** JSON file with boundary file for US States.

  http://eric.clst.org/tech/usgeojson/

Data Preparation and Cleaning

- **Converting Weekly and Monthly data into Quarterly Average:** Since the HPI data for U.S and 'USA or Census Division' are quarterly based - Mortgage, Inflation and Unemployment rate data was required to be converted to quarterly based average.

- **Data Filtering:** Since HPI data is available from 1975 to 2015, all the other metrics were required to be filtered accordingly.

- **Merging the data:** Mortgage Rate, Inflation Rate, Unemployment Rate and HPI were extracted from different data sources and hence it was required to merged. Data was indexed based on date to help in merge operation

- **Remove unwanted columns**: - After Data exploration and understanding the relationship, only few columns like 'place_name', 'U.S Interest Rate','Unemployment Rate', 'Inflation Rate', 'index_nsa' were needed for applying Regression Model

Research Question(s)

The goal of this project is to analyze HPI, Mortgage Rate, Inflation Rate and Unemployment Rate to determine the following:

- How is Housing Price Index dependent on Mortgage Rate?

- How does Housing Price Index vary based on locations?

- How is Housing Price Index dependent on Inflation and Unemployment Rates?

- Combining Mortgage Rate, Inflation Rate and Unemployment Rates to predict HPI

- Whether Linear Model or Decision Tree Model provide more accuracy to predict HPI using Mortgage Rate, Inflation Rate and Unemployment Rates?

Methods

One major advantage with given data set is availability over a long period. Chosen dataset has HPI data available from 1975 to 2015 (other data set have even longer history). This helped in analysis and building a model using large data set.

Method used in analyzing HPI dependency location wise: Choropleth map from Folium

Method used to study how Mortgage, Inflation and Unemployment rates changed over time: Line plot from Matplotlib

Method used in analyzing relationships between various metrics: Scatter Plot and Generating Correlation Coefficient using numpy.

Method used to train and predict HPI:

 a) Linear Regression Model

 b) Decision Tree Regression Model

Findings

U.S Mortgage Rate

Over the period of analysis 1975-2015, interest rates on the 30-year fixed-rate mortgage have ranged from as high as 17.73% in 1981 to as low as 3.35% in 2012.

We used interest rate data from Freddie Mac's Primary Mortgage Market Survey (PMMS) to examine historical mortgage rates.

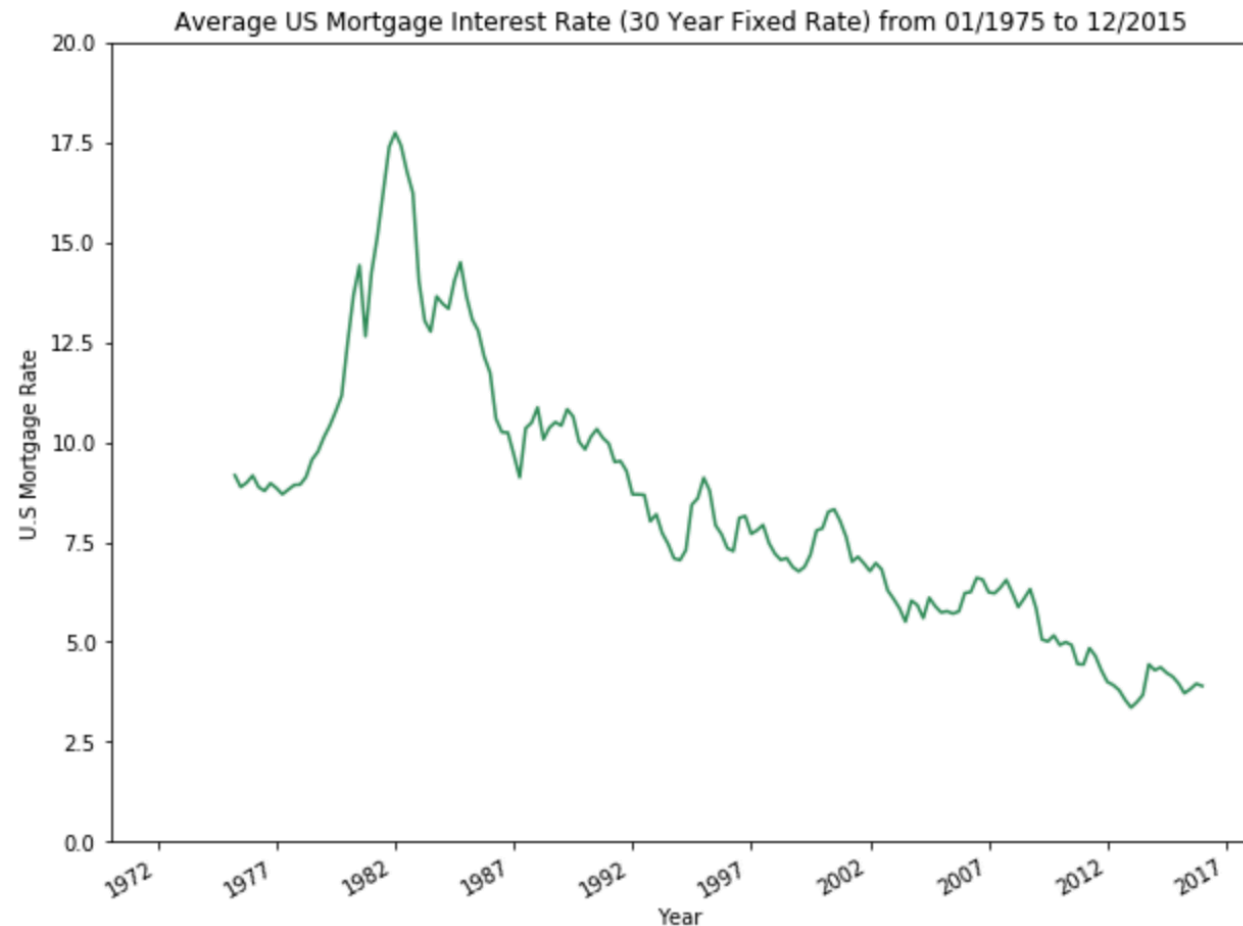U.S Mortgage rate fell tremendously after peaking in 1981.The long time average rate is 8.35%.

Fig 1: Average US Mortgage Interest Rate from 01/1975 to 12/2015

Findings

U.S Inflation and Unemployment Rates

The graph on the right shows U.S Inflation Rate in Red, it peaked at 13.3% during 1980 recession and lowest of 0.7% in 2010 as we were recovering from Financial Crisis of 2008.

The unemployment Rate is shown In Blue and as you would guess it has peaks soon after recessions of 1980 and 2008 started.
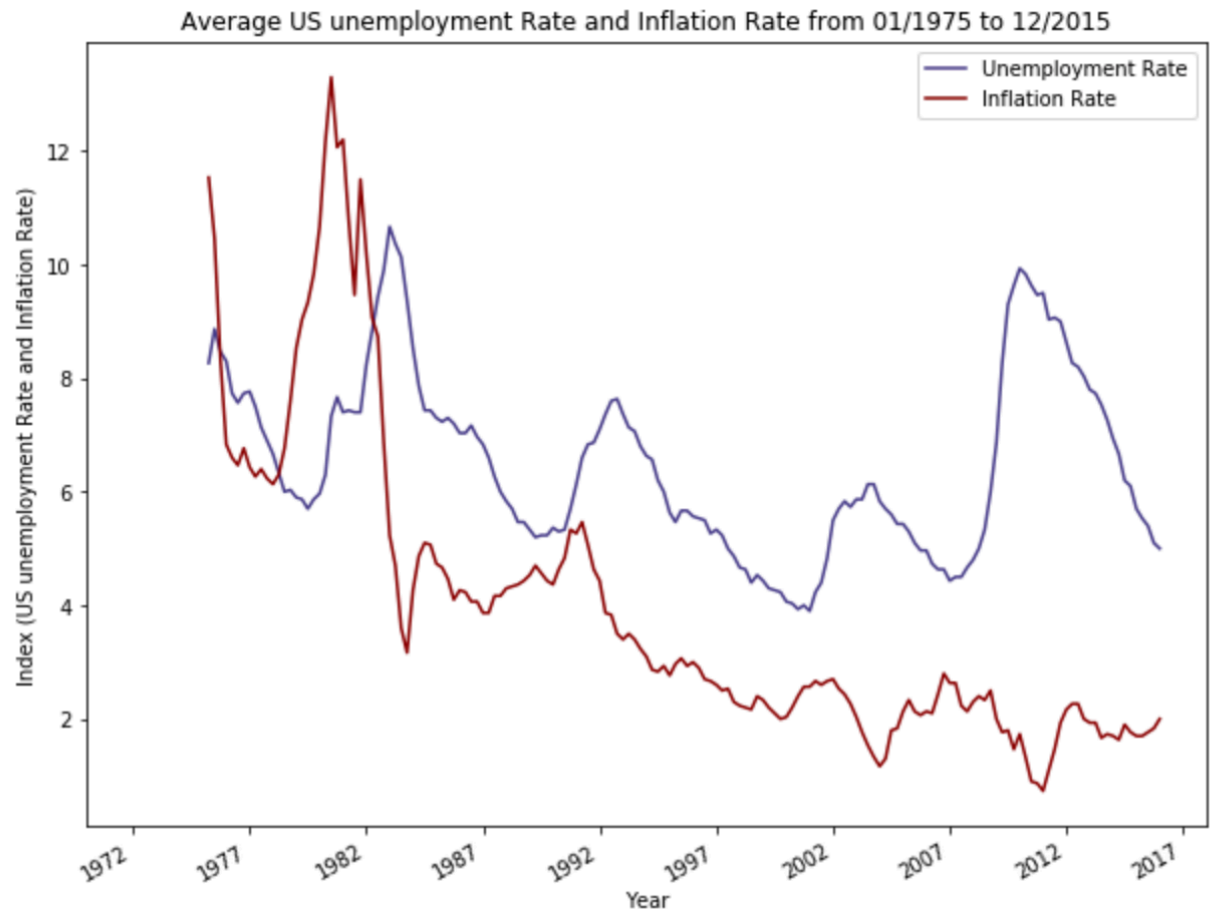
Fig 2: Average US Unemployment Rate and Inflation Rate from 01/1975 to 12/2015

Findings

How does Housing Price Index vary based on locations?

The graph on the right shows HPI from 1975 - 2015 for various divisions.

From the graph, it's seen HPI was very close for all Divisions around 1975 but varied drastically over time, peaking for Pacific Division in 2006. HPI is seen to be reaching max for most of the US divisions just before the 2008 Financial Crisis.
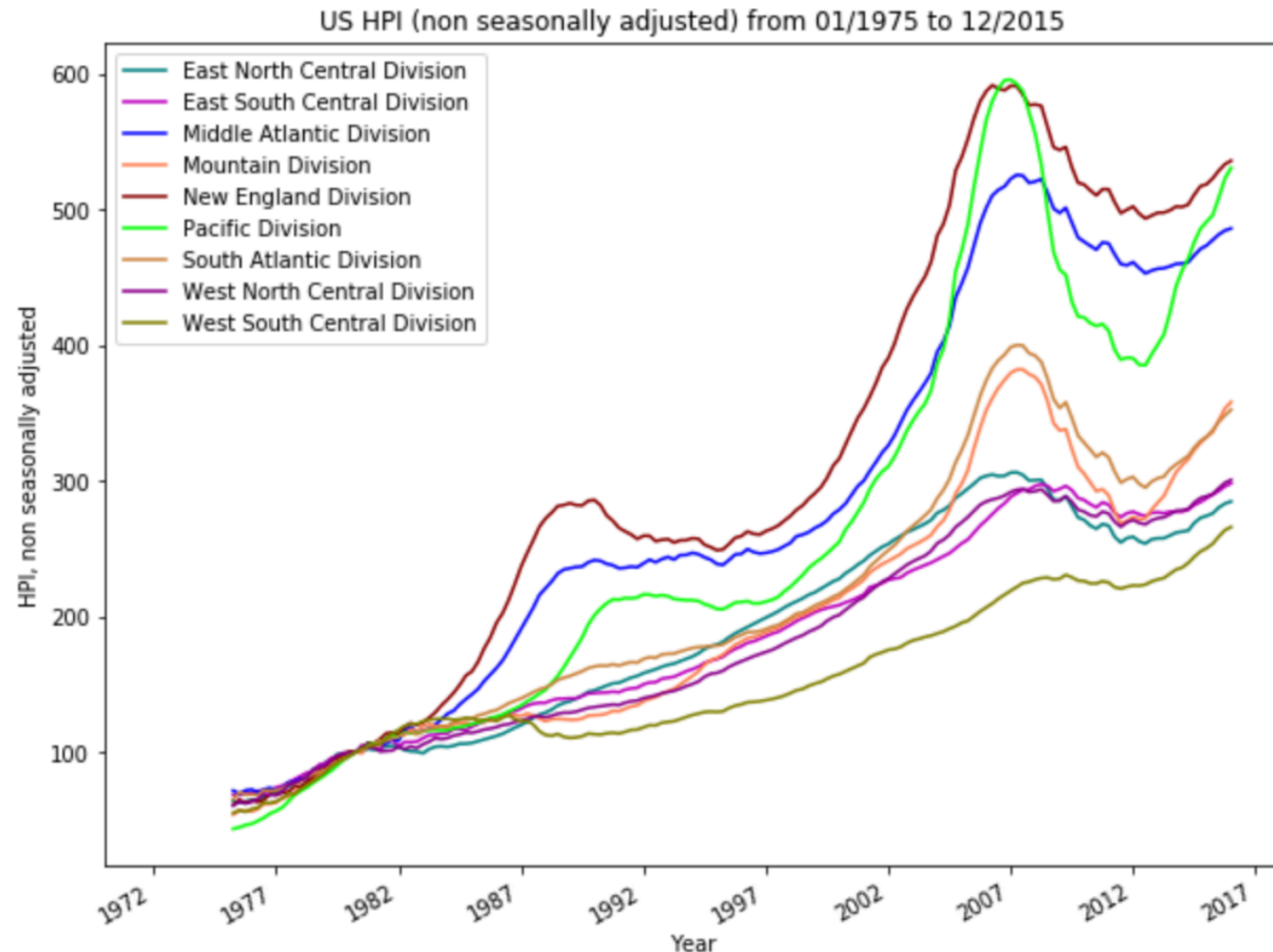

Fig 3: Line Plots for 9 U.S Divisions showing HPI from 01/1975 to 12/2015

Findings

HPI (nsa) for 2015 across United States of America. Top 10 States and their HPI are given below

place_name	index_nsa
District of Columbia	736.7975
Massachusetts	671.5575
New York	592.1950
Hawaii	538.1575
California	525.7025
New Jersey	477.9625
Maine	471.3725
Rhode Island	470.1650
Washington	463.0125
Vermont	447.5250

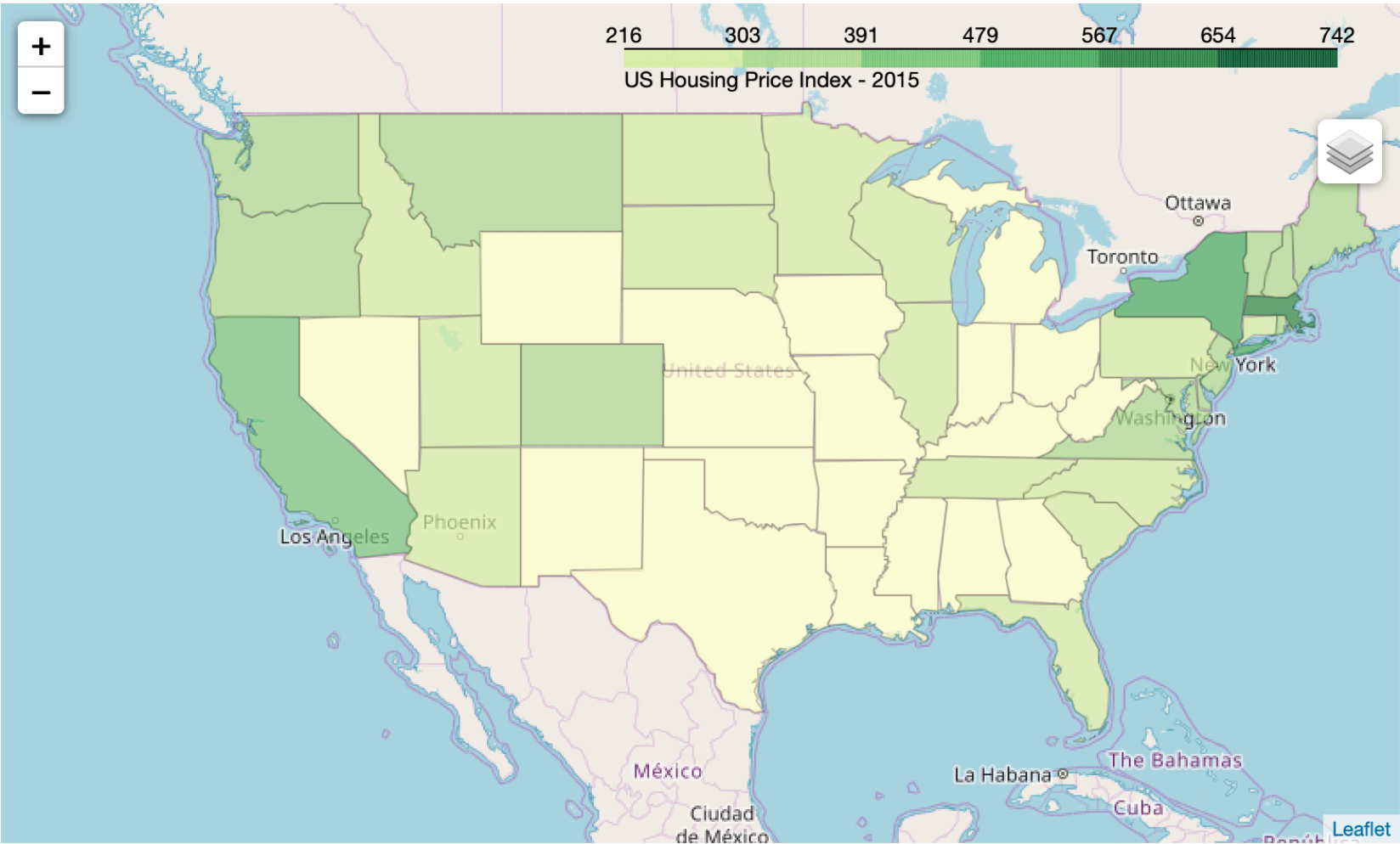Fig 4: HPI for 2015 across various States in US

Findings

How is Housing Price Index dependent on Mortgage Rate?

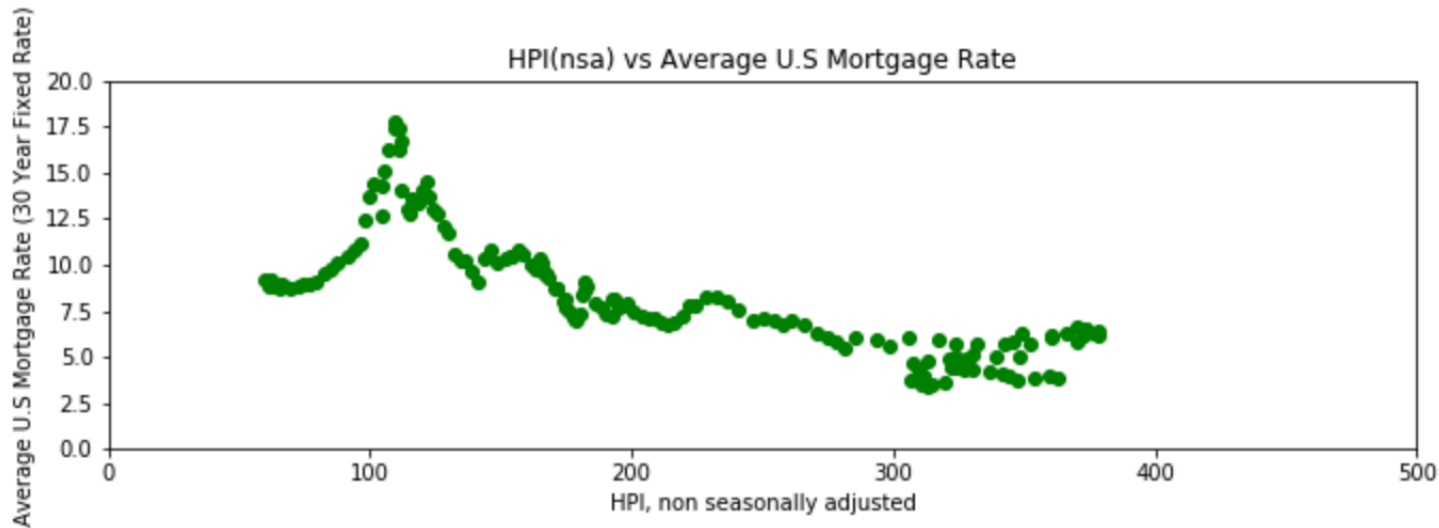HPI(nsa) vs Average U.S Mortgage Rate

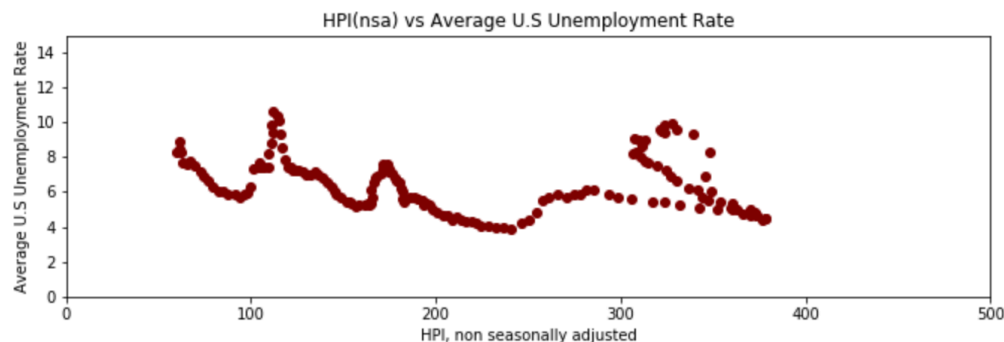Fig 5: Scatter Plot of HPI Vs U.S Mortgage Rate from 1975 to 2015

Corrcoef of HPI and Mortgage Rate
[[1.   -0.78606736]
 [-0.78606736 1.   ]]

Fig 6: Correlation coefficient of HPI and U.S Mortgage Rate

Analyzing the scatter plot and correlation factor shows negative relationship between Mortgage Rate and HPI, which means increase in Mortgage Rate may cause downfall in HPI, but a correlation of -0.78 suggests it's not a strong correlation.

Findings

How is Housing Price Index dependent on Inflation and Unemployment Rates?

HPI(nsa) vs Average U.S Unemployment Rate

```
Corrcoef of HPI and Unemployment Rate
[[ 1.          -0.24154172]
 [-0.24154172  1.        ]]
```


HPI(nsa) vs Average U.S Inflation Rate
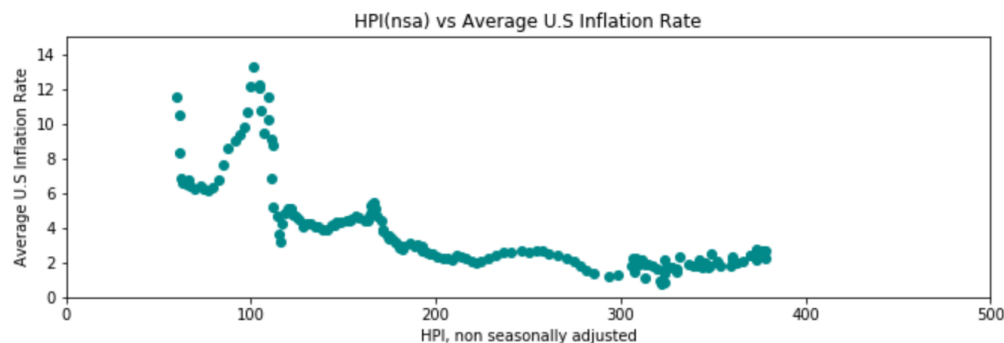
```
Corrcoef of HPI and Inflation Rate
[[ 1.          -0.75331819]
 [-0.75331819  1.        ]]
```

Fig 7: Scatter Plot of HPI Vs U.S Mortgage Rate from 1975 to 2015   Fig 8: Scatter Plot of HPI Vs U.S Mortgage Rate from 1975 to 2015

Scatter plot of Unemployment Rate and HPI does have negative correlation, but -0.24 suggests a weak correlation. Inflation does seem to have good correlation but like Mortgage Rate it doesn't have strong correlation.

Findings

Combining Mortgage Rate, Inflation Rate and Unemployment Rates to predict HPI - Linear Regression Model
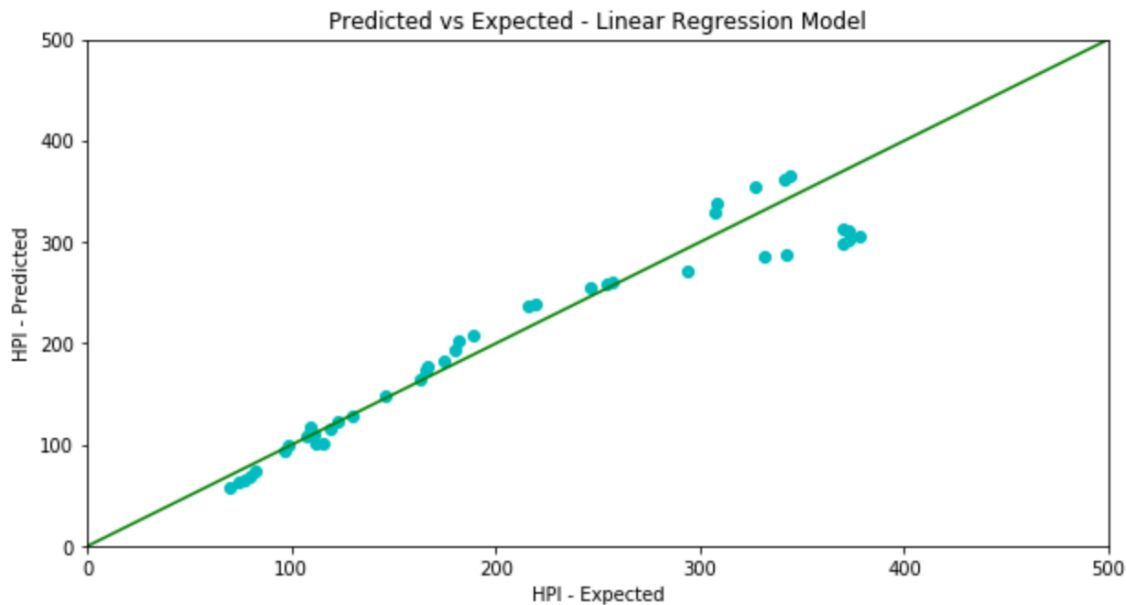

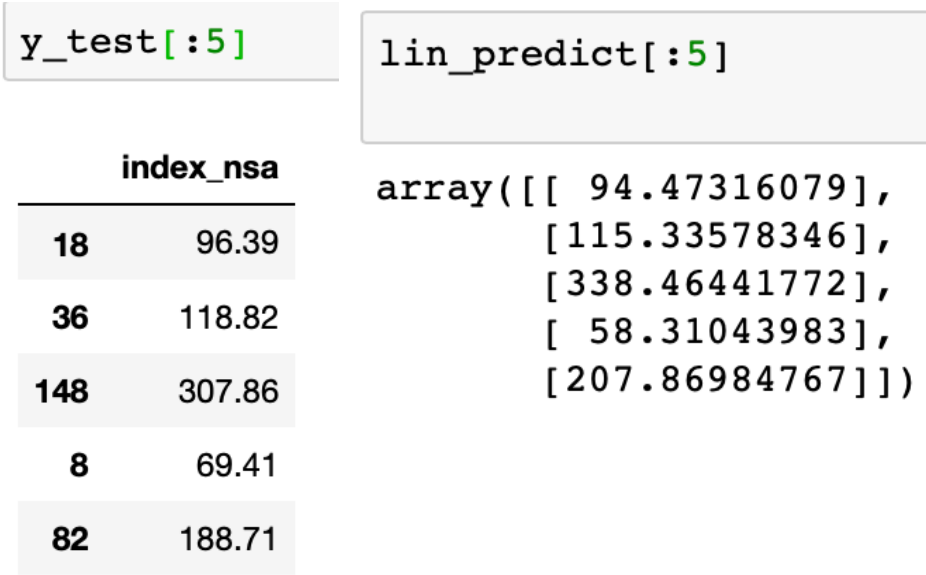Fig 9: Scatter Plot of Predicted vs Actual HPI using Linear Regression Model

Fig 10: Screenshot from notebook with first 5 values of Test data and Predicted Data

Linear Regression Model was applied to predict HPI using vast data on Mortgage Rate, Inflation Rate and Unemployment Rate. Fig 10 shows prediction is pretty close to the actual data. Fig 9 shows how predicted vs expected value fit against a linear line.

Findings

Combining Mortgage Rate, Inflation Rate and Unemployment Rates to predict HPI - Decision Tree Regression Model
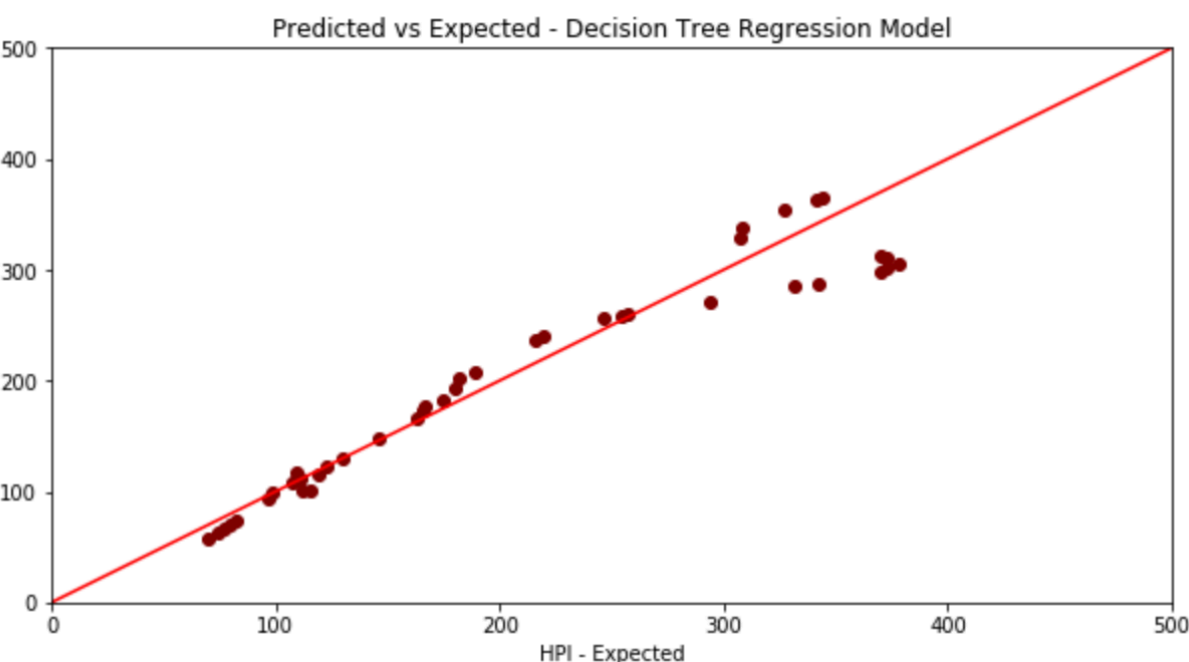

Fig 11: Scatter Plot of Predicted vs Actual HPI using Decision Tree Regression Model

Fig 12: Screenshot from notebook with first 5 values of Test data and Predicted Data

To compare the accuracy, Decision Tree Regression Model was also applied to predict HPI on Mortgage Rate, Inflation Rate and Unemployment Rate. Fig 12 shows prediction is more accurate than the predicted values in Fig 10 of slide 14. Comparing Fig 9 and Fig 11 clearly shows the dots are closer to the line in Decision Tree when compared to Linear Model.

Limitations

- 30 Year Fixed Rate Mortgage Rate is considered for analysis. It is required to check how Adjustable Rate varies with HPI and use it in machine learning if required.

- HPI is obtained by reviewing mortgage transactions on single-family properties only.

- Non Seasonally Adjusted HPI is used in the analysis, results may vary for Seasonally Adjusted HPI

Conclusions

Following is the conclusion based on the analysis:

- **How is Housing Price Index dependent on Mortgage Rate?**

  Mortgage Rate and HPI have negative correlation and correlation factor of -0.78 suggests good correlation but it doesn't seems strong (Assuming value greater than +/- 0.9 shows strong correlation). As per the analysis, mortgage rate is one of the metrics influencing HPI but may not have significant dependency.

- **How does Housing Price Index vary based on locations?**

  Housing Price is influenced by Location and it is clearly evident based on fig 3 and fig 4. DC, MA and NY have higher HPI compared to West South Central region, Off-course there are several factors that causes the dependency on location and it is out of scope for this project.

- **How is Housing Price Index dependent on Inflation and Unemployment Rates?**

  Like Mortgage Rate, Inflation have negative correlation of 0.75 and is one of the factor influencing HPI, but does not show a strong dependency. As per the data analysis, Unemployment Rate have weak negative correlation on HPI.

- **Combining Mortgage Rate, Inflation Rate and Unemployment Rates to predict HPI**

  Slide 14 and 15 shows result of Linear and Decision Tree Regression Models built using Mortgage Rate, Unemployment Rate and Inflation Rate.

- **Whether Linear Model or Decision Tree Model provide more accuracy to predict HPI using Mortgage Rate, Inflation Rate and Unemployment Rates?**

  As per the analysis, Linear Model has RMSE of 28.89 and Decision Tree Model has RMSE of 5.4. Based on RMSE values, Decision Tree Model performs better in predicting HPI using the chosen data.

Acknowledgements

- Collected the data from sites mentioned in slide 4

- No, I did not use informal analysis. This is purely based on my understanding and curiosity on the subject and going through the course materials.

- No, I did not receive feedbacks from Colleagues or Friends.

References

Data:

http://www.freddiemac.com/pmms/pmms_archives.html
https://catalog.data.gov/dataset/fhfa-house-price-indexes-hpis
https://www.kaggle.com/federalreserve/interest-rates

Domain Info:

https://www.fhfa.gov/Media/PublicAffairs/Pages/House-Price-Index-Frequently-Asked-Questions.aspx

http://www.freddiemac.com

https://www.valuepenguin.com/mortgages/historical-mortgage-rates

https://www.amerisbank.com/about-us/newsroom/why-is-the-housing-market-so-important/

Technical:

Course materials and Exercise notebook