**Problem Statement:**

Find publicly available data for key factors that influence US home prices nationally. Then, build a data science model that explains how these factors impacted home prices over the last 20 years.  
Use the S&P Case-Schiller Home Price Index as a proxy for home prices: [[fred.stlouisfed.org/series/CSUSHPISA](http://fred.stlouisfed.org/series/CSUSHPISA)]  
  
Note: Do not use pre-existing datasets for this assignment. Most factors that influence home prices at a national level can be individually sourced from public websites.

**Data Sources:**

These are the following sources that I used to collect data for each variable:

Home price index: <https://fred.stlouisfed.org/series/CSUSHPISA#0>

GDP data: <https://fred.stlouisfed.org/series/GDP>

Unemployment rate: [https://fred.stlouisfed.org/series/UNRATE#](https://fred.stlouisfed.org/series/UNRATE)

Consumer sentiment: <https://fred.stlouisfed.org/series/UMCSENT#0>

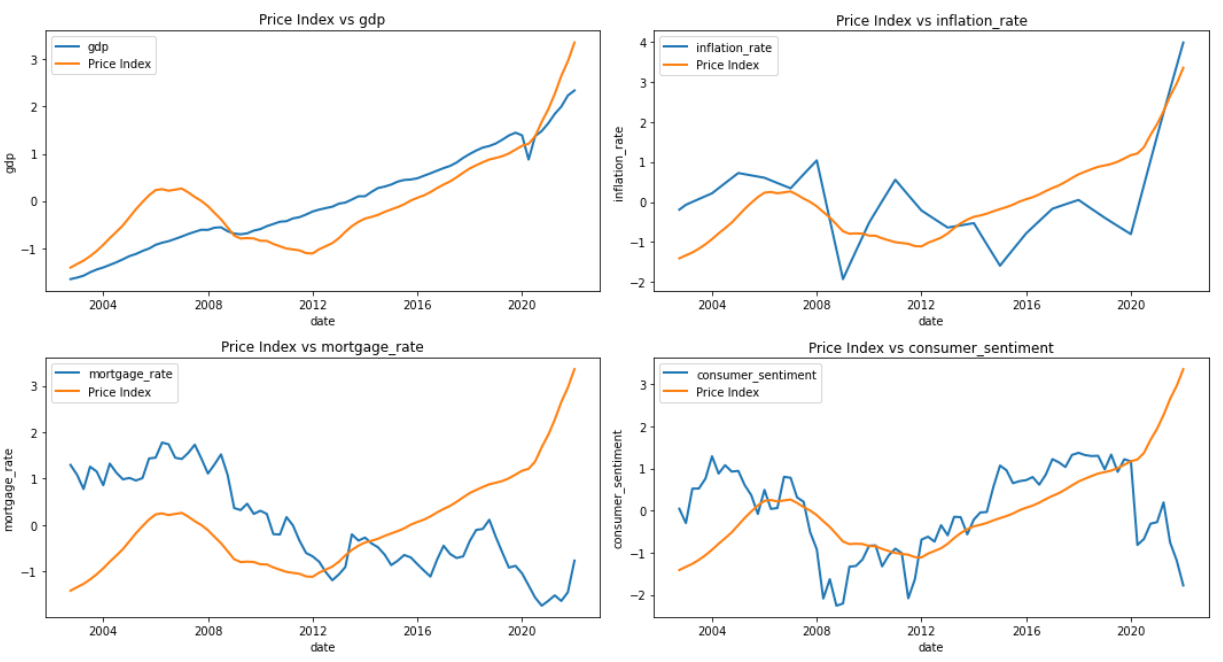
Mortgage rates: <https://fred.stlouisfed.org/series/MORTGAGE30US#0>

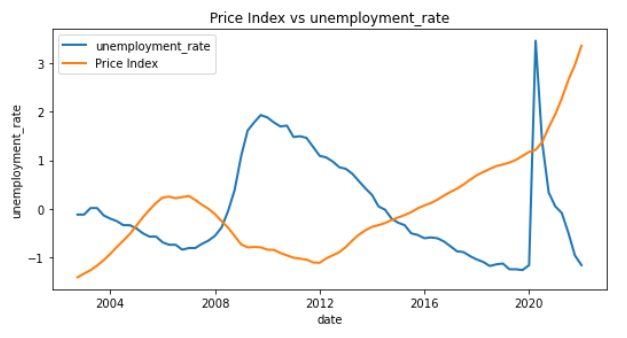
Inflation rates: <https://fred.stlouisfed.org/series/T30YIEM>

The collected data is trimmed between 2002-07-01 to 2022-01-01 for analysis.

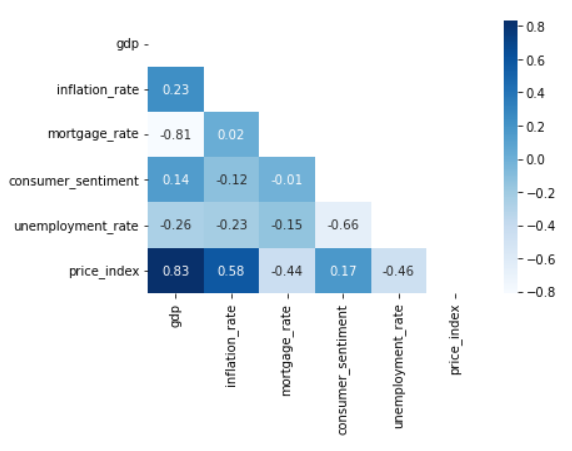
**Visualisation:**

Following are the line graphs showing the relationship between each variable and house price index:





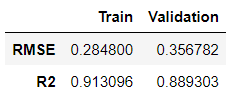
This is the correlation heat map of the variables:



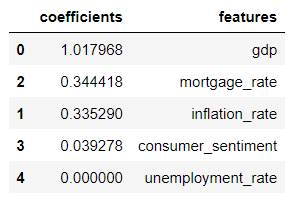
It seems from the graphs that the GDP has the highest impact on the house price index and Consumer Sentiment & Unemployment Rate has the lowest impact.

**Modelling:**

I used lasso model to predict the price index which achieved the following accuracy on testing and training data:

****

After getting the most influential coefficients from the best Lasso model, the following is the result achieved:

****