Analysing Key Factors Influencing US Home Prices Over the Last 20 Years

Problem Statement:

Houses are one of the necessary needs of each and every person around the globe and therefore housing and real estate market is one of the markets which is one of the major contributors in the world's economy.

I have to find out the key factors that influence US home prices over the last 20 years.

So, I did some research and I found that unemployment rate, interest rate, active population, houses under construction, houses with completed construction and total house construction can influence the US home price.

Data Collection:

The most of data is publicly available. I have used the the S&P Case-Schiller Home Price Index.

And most of the data for corresponding features was collected from https://fred.stlouisfed.org/

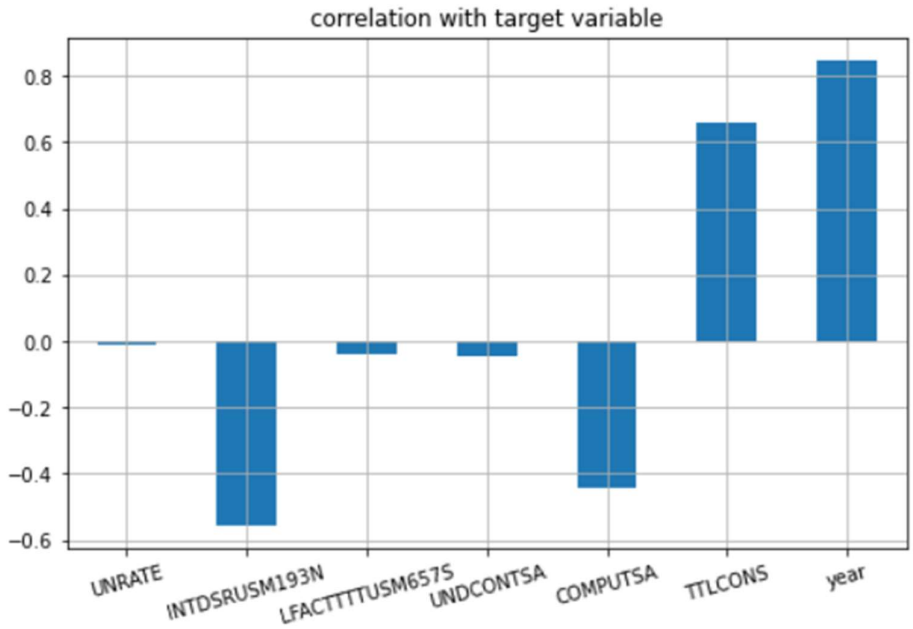
1. CSUSHPISA: US home prices
   Source: https://fred.stlouisfed.org/searchresults?st=CSUSHPISA+
2. UNRATE: Unemployment Rate
   Source: https://fred.stlouisfed.org/searchresults?st=UNRATE
3. INTDSRUSM193N: Interest Rate
   Source: https://fred.stlouisfed.org/searchresults?st=+INTDSRUSM193N+
4. LFACTTTTUSM657S: Active Population
   Source: https://fred.stlouisfed.org/searchresults?st=+++LFACTTTTUSM657S
5. UNDCONTSA: Under Construction (Privately owned unit under construction)
   Source: https://fred.stlouisfed.org/searchresults?st=UNDCONTSA++
6. COMPUTSA: Complete Construction (Privately owned housing unit complete construction)
   Source: https://fred.stlouisfed.org/searchresults?st=COMPUTSA+
7. TTLCONS: Total Construction
   Source: https://fred.stlouisfed.org/searchresults?st=TTLCONS+

Data Preparation:

I have collected data from the respective source and merge it in one table. Where the target variable is US home prices and others are the key features which may or may not influence the home prices.

Exploratory Data Analysis:

An exploratory data analysis was conducted on the data to get information about which features have significant impact on target (US home prices).



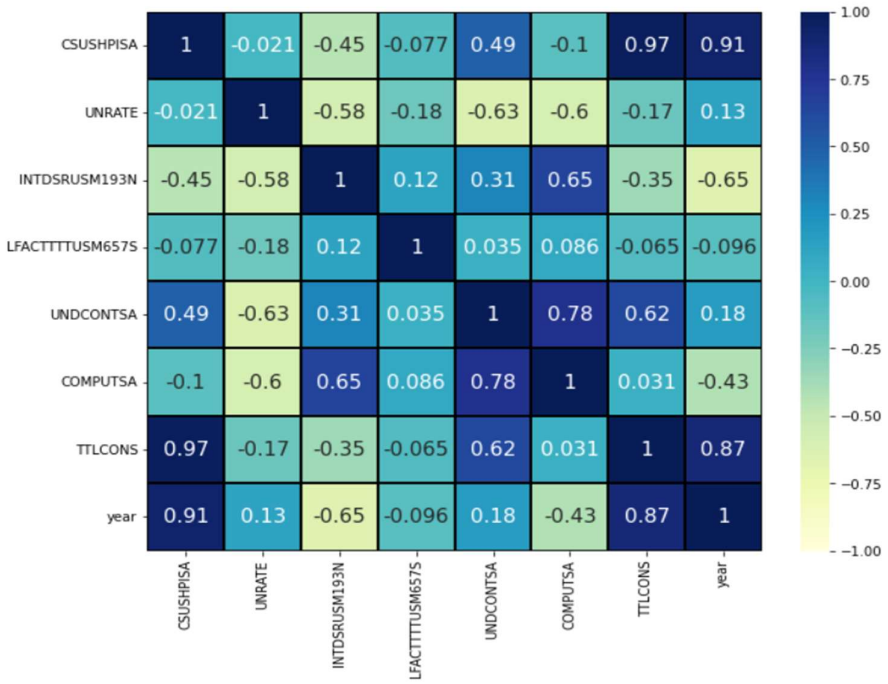
From the above graph we can say that, UNRATE: Unemployment Rate, INTDSRUSM193N: Interest Rate, LFACTTTTUSM657S: Active Population, UNDCONTSA: Under Construction (Privately owned unit under construction), COMPUTSA: Complete Construction (Privately owned housing unit complete construction) shows negative impact on the prices.

TTLCONS: Total Construction and years show positive impact on the US home price.

Model construction:

Linear Regression, Decision Tree Regressor, Random Forest Regressor models have been used and have checked the MSE, MAE, R^2, and RMSE for each model and on the basis of the values of R^2 and RMSE I have chosen the best fit model.

Result and Conclusion:

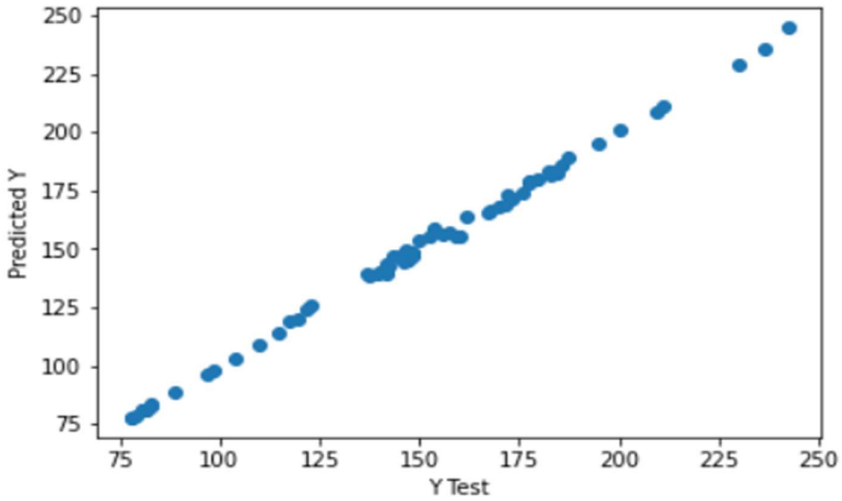


- As we can see that CSUSHPISA and TTLCONS are highly correlated with the Year. So, from that we can say that, US house price is highly influence with the change in years. As year changing the house prices are also increasing.

- TTLCONS also have high impact in US house price as we can see the correlation is **0.87.**

Machine Learning Models:

The **R^2 _score** for **Random Forest Regressor** was **0.998** and **RMSE** was **1.76** which is less that Linear Regressor and Decision Tree Regressor model. And which makes it the best fit model.



So, from this we can conclude that, as the years changes the price of houses also changes and it is showing the increase in the prices so we can predict that in coming years the price of US houses will also increase.