1. Objective

To analyse publicly available data on economic, demographic, and real estate indicators to build a predictive model that explains the impact of these factors on the S&P/Case-Shiller Home Price Index, a key indicator of U.S. home prices, over the last two decades.

2. Introduction

The S&P CoreLogic Case-Shiller Home Price Indices play a crucial role in tracking the price levels of single-family homes in the United States. These indices offer valuable insights into the ever-changing housing market, enabling us to monitor and understand the dynamics of home prices. The S&P CoreLogic Case-Shiller U.S. National Home Price Index is a key component of this system, providing a comprehensive view of the overall value of single-family homes nationwide. It achieves this by aggregating data from nine different regions, and its data is updated on a monthly basis. By focusing on city indices, we can also delve into the average price changes in specific geographic markets, covering 20 major metropolitan areas. These areas are further grouped into two composites, one with 10 metro areas and another that encompasses all 20. One important aspect of these indices is their ability to measure percentage changes in housing market prices while maintaining a constant level of quality, ensuring that variations due to factors such as house types, sizes, or physical characteristics are excluded from the calculations.

3. Data and Methodology

3.1 Data Collection:

The features were identified by conducting a literature survey of The S&P CoreLogic Case-Shiller Home Price Indices. Most of the data for the corresponding features was collected from https://fred.stlouisfed.org/

1. UNRATE: Unemployment Rate Units: Percent, Seasonally Adjusted
   Frequency: Monthly Source: https://fred.stlouisfed.org/series/UNRATE

2. CSUSHPISA: S&P/Case-Shiller U.S. National Home Price Index Units: Index Jan 2000=100,Seasonally Adjusted
   Frequency: Monthly Source: https://fred.stlouisfed.org/series/CSUSHPISA

3. PERMIT: New Privately-Owned Housing Units Authorized in Permit-Issuing Places: Total Units Units : Thousands of Units, Seasonally Adjusted Annual Rate
   Frequency: Monthly Source: https://fred.stlouisfed.org/series/PERMIT

4. PERMIT1: New Privately-Owned Housing Units Authorized in Permit-Issuing Places: Single-Family Units Units : Thousands of Units, Seasonally Adjusted Annual Rate
   Frequency: Monthly Source: https://fred.stlouisfed.org/series/PERMIT1

5. MSACSR: Monthly Supply of New Houses in the United States Units: Months' Supply, Seasonally Adjusted
   Frequency: Monthly Source: https://fred.stlouisfed.org/series/MSACSR

6. TTLCONS: Total Construction Spending: Total Construction in the United States Units: Millions of Dollars, Seasonally Adjusted Annual Rate
   Frequency: Monthly Source: https://fred.stlouisfed.org/series/TTLCONS

7. NASDAQCOM: NASDAQ Composite Index Units: Index Feb 5, 1971=100,Not Seasonally Adjusted
   Frequency: Daily, Close Source: https://fred.stlouisfed.org/series/NASDAQCOM

8. LFACTTTTUSM657S: Active Population: Aged 15 and over: All Persons for United States Units: Growth rate previous period , Seasonally Adjusted
   Frequency: Monthly Source: https://fred.stlouisfed.org/series/LFACTTTTUSM657S

9. HSN1F: New One Family Houses Sold: United States Units: Thousands, Seasonally Adjusted Annual Rate
   Frequency: Monthly Source: https://fred.stlouisfed.org/series/HSN1F

10. HOUST1F: New Privately-Owned Housing Units Started: Single-Family Units Units : Thousands of Units ,Seasonally Adjusted Annual Rate
    Frequency: Monthly Source: https://fred.stlouisfed.org/series/HOUST1F

11. The Labor Force Participation Rate (LFPR) https://pib.gov.in/PressReleaseIframePage.aspx?PRID=1966154

12. Housing starts (New Housing Project) https://towardsdatascience.com/linear-regression-on-housing-csv-data-kaggle-10b0edc550ed

13. Industrial Production: Cement (The industrial production (IP) index measures the real output of all relevant establishments located in the United States) https://fred.stlouisfed.org/series/INDPRO/

14. Personal Income & Outlays csv https://alfred.stlouisfed.org/release?rid=54

15. New Privately-Owned Housing Units Completed: Total Units https://fred.stlouisfed.org/series/computsa

3.2 Data Preparation:

When combining datasets that contain information recorded monthly and quarterly, it's common to encounter missing or null values due to differences in the frequency of data collection. To handle these gaps in the merged dataset, the missing values can be replaced or filled in with the mode value, which refers to the most frequently occurring value in the respective dataset columns. This method helps to ensure completeness and consistency in the combined data.

3.3 Exploratory Data Analysis:

An exploratory data analysis was performed on the given dataset to extract crucial insights and recognize important features. Using the ECDF (Empirical Cumulative Distribution Function) specification can aid in visualizing histograms more effectively, providing a clearer understanding of the values. Additionally, employing regression for plotting can help in examining variations within the dataset

3.4 Model selection and Evaluation:

We employed linear regression, random forest, and XGBoost for evaluation purposes. We utilized various metrics including Mean Squared Error (MSE), Root Mean Squared Error (RMSE), R-squared (R2), and adjusted R-squared. Among these, the random forest model exhibited a significantly lower MSE of 8.33 and an exceptional adjusted R2 of 0.99. This performance was notably superior to XGBoost, despite XGBoost having higher MSE..

4. Result and Discussion

4.1 Exploratory Data Analysis

HNFSEPUSSA is following the same trend as of CSUSHPISA Even a small change in the NA000334Q can make a big difference in the S&P Home Price Index (HPI), and they tend to move in a positive direction together. The Labor Force Participation Rate (LFPR) is closely linked to the S&P Home Price Index (HPI), and this relationship is characterized by a positive correlation. TTLCONS is directly associated with the S&P Home Price Index (HPI). CPI Adjusted Price is following a similar trend to the S&P Home Price Index (HPI). UNRATE is inversely correlated with employment rate , meaning as one goes up, the other goes down.

4.2 Correlation Matrix:

New Privately-Owned Housing Units Started: Single-Family Units : has a high correlation of 0.94 with CSUSHPISA. Total Construction Spending (TTLCONS) is closely correlated with a value of 0.4 with CSUSHPISA. Shows a substantial correlation of 0.84 with CSUSHPISA. new one family house is notably correlated with a value of 0.55 with CSUSHPISA. The NASDAQ Composite Index (NASDAQCOM) exhibits a significant correlation of 0.26 with CSUSHPISA.

4.3 Machine Learning Models:

The objective is not only to achieve a good R-squared but also to identify important parameters. Linear regression, with an R-squared of 0.993, tends to eliminate many features, which doesn't align with the objective. On the other hand, xg bost, with an R-squared of 0.997, primarily focuses on maximizing R-squared and reducing overfitting and multicollinearity but falls short in feature identification. To strike a balance random forest was chosen,. Give less mse with with an R-squared of 0.997

5. Conclusions

- Upon analysing the data through three distinct processes, namely EDA (Exploratory Data Analysis), correlation matrix examination, and machine learning modelling, a set of common and highly important features emerge. These features are:

- Personal Income & Outlays: It directly reflects changes in home prices, which have a significant impact on the S&P Home Price Index (HPI).

- Employment rate, which often leads to increased demand for housing and higher home prices, influencing the S&P HPI.

- Total Construction Spending (TTLCONS): It signifies the level of construction activity, which affects housing supply and demand, consequently impacting the S&P HPI.

- New privately own housing Directly measures property prices, affecting the value of the S&P HPI.

- CPI-Adjusted Price: Reflects changes in housing costs, impacting home prices and the S&P HPI.

- NASDAQ Composite Index (NASDAQCOM): The performance of tech companies can influence economic growth, job creation, and housing demand, affecting the S&P HPI.

- Monthly Supply of New Houses (MSACSR): The supply of new houses relative to demand impacts home prices, which, in turn, influences the S&P HPI.

MECE FRAME WORK

Factors Influencing US House Prices

Economic Factor

- Growth in the Economy
- Unemployment
- Customer Trust Rates
- Offering
- GDP
- Home Sales Economy Mirror
- Supply and Demand
- Advance
- Compliance
- Competition
- Extinct
- Surplus Productivity

Location

- Neighborhoods
- Highways
- Attractions
- Schools
- Area Desirability
- Crime Rate

Government

- Government laws
- Property Taxes

Banks

- Mortgage Availability
- Interest Rates