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

1. Objective : To analyze 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. **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

2. **HNFSEPUSSA**: New One Family Homes for Sale in the United States
   Units: Thousands of Units,Seasonally Adjusted
   Frequency: Monthly, End of Month
   Source: https://fred.stlouisfed.org/series/HNFSEPUSSA

3. **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

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

5. **INTDSRUSM193N**: Interest Rates, Discount Rate for United States
   Units: Percent per Annum,Not Seasonally Adjusted
   Frequency: Monthly
   Source: https://fred.stlouisfed.org/series/INTDSRUSM193N

6. **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

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

8. **NA000334Q**: Gross Domestic Product
   Units: Millions of Dollars,Not Seasonally Adjusted
   Frequency: Quarterly
   Source: https://fred.stlouisfed.org/series/NA000334Q

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

10. **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

11. **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

12. **QUSR628BIS**: Real Residential Property Prices for United States

Units: Index 2010=100,Not Seasonally Adjusted

Frequency: Quarterly

Source: https://fred.stlouisfed.org/series/QUSR628BIS

13. **RSAHORUSQ156S**: Homeownership Rate in the United States

Units: Percent,Seasonally Adjusted

Frequency: Quarterly

Source: https://fred.stlouisfed.org/series/RSAHORUSQ156S

14. **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

15. **UNRATE**: Unemployment Rate

Units: Percent,Seasonally Adjusted

Frequency: Monthly

Source: https://fred.stlouisfed.org/series/UNRATE

16. **CPI (Adjusted Price):**

Source: https://www.fhfa.gov/DataTools/Downloads/Pages/House-Price-Index.aspx

17. **Median Home Prices (NSA):**

Units: Dollars

Source: https://www.fhfa.gov/DataTools/Downloads/Pages/House-Price-Index.aspx

3.2 Data Preparation:

- The NASDAQ Composite Index had a daily frequency, and it was converted to monthly data by averaging the daily values.
- Features such as Gross Domestic Product, Homeownership Rate in the United States, and Real Residential Property Prices for the United States had a quarterly frequency. To convert them to monthly data, the values for these features remained unchanged for the next two months.
- All the features were merged using the date as a common key.
- The assumption made for the S&P Case-Shiller data is that each data point is considered independently, without taking into account any temporal or sequential relationships between data points.

3.3 Exploratory Data Analysis:

An exploratory data analysis was conducted on the provided data to extract essential insights and identify significant features.

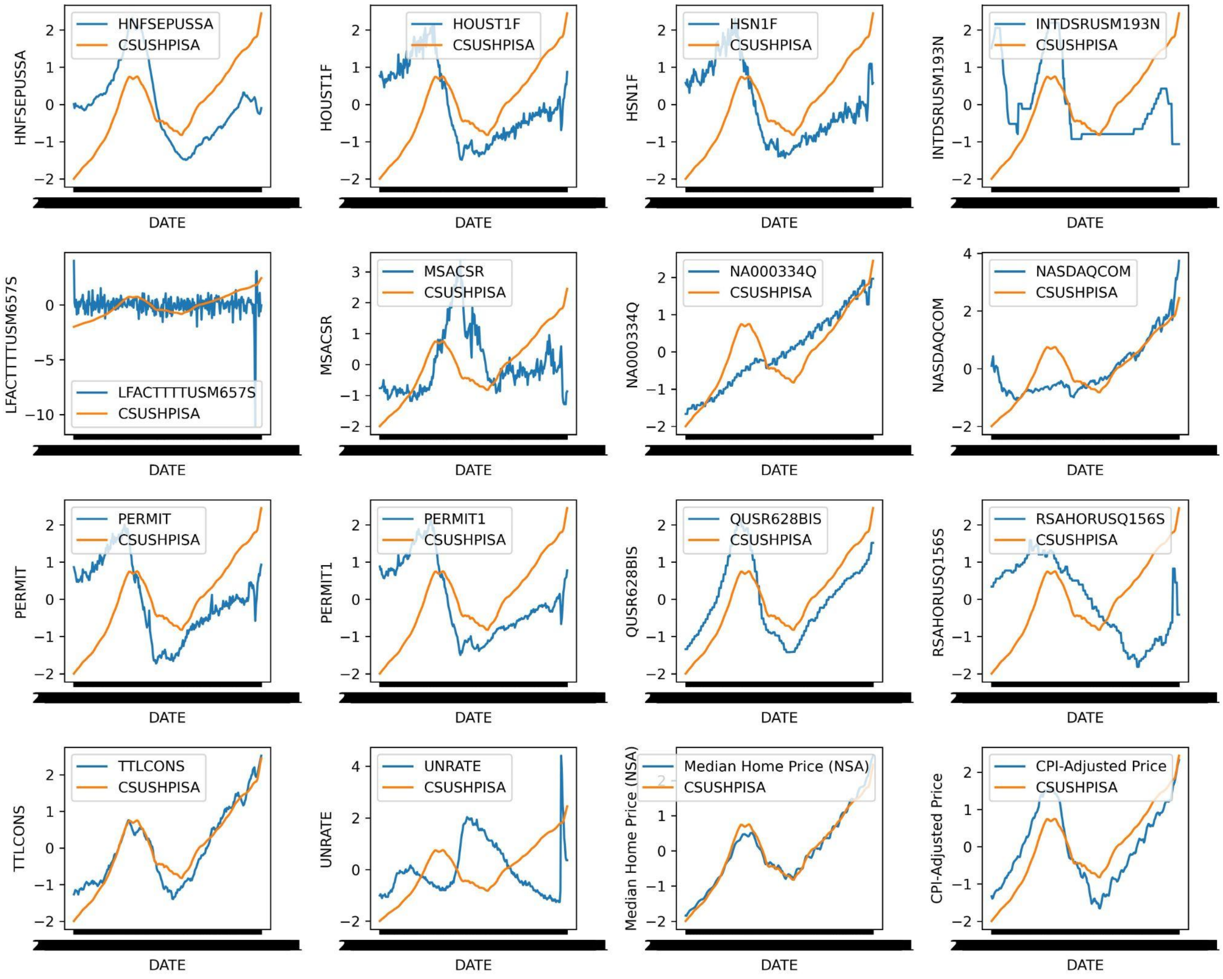
3.4 Model selection and Evaluation:

Lasso, Ridge, and Elastic Net were chosen for model development to assess the influence of the mentioned factors on the S&P/Case-Shiller Home Price Index. These models serve a dual purpose by aiding in feature selection and mitigating issues like multicollinearity and overfitting.

The model was assessed using R-squared (coefficient of determination), with an R-squared value approaching 1 indicating excellent model performance, while a value near 0 suggests poor model performance.

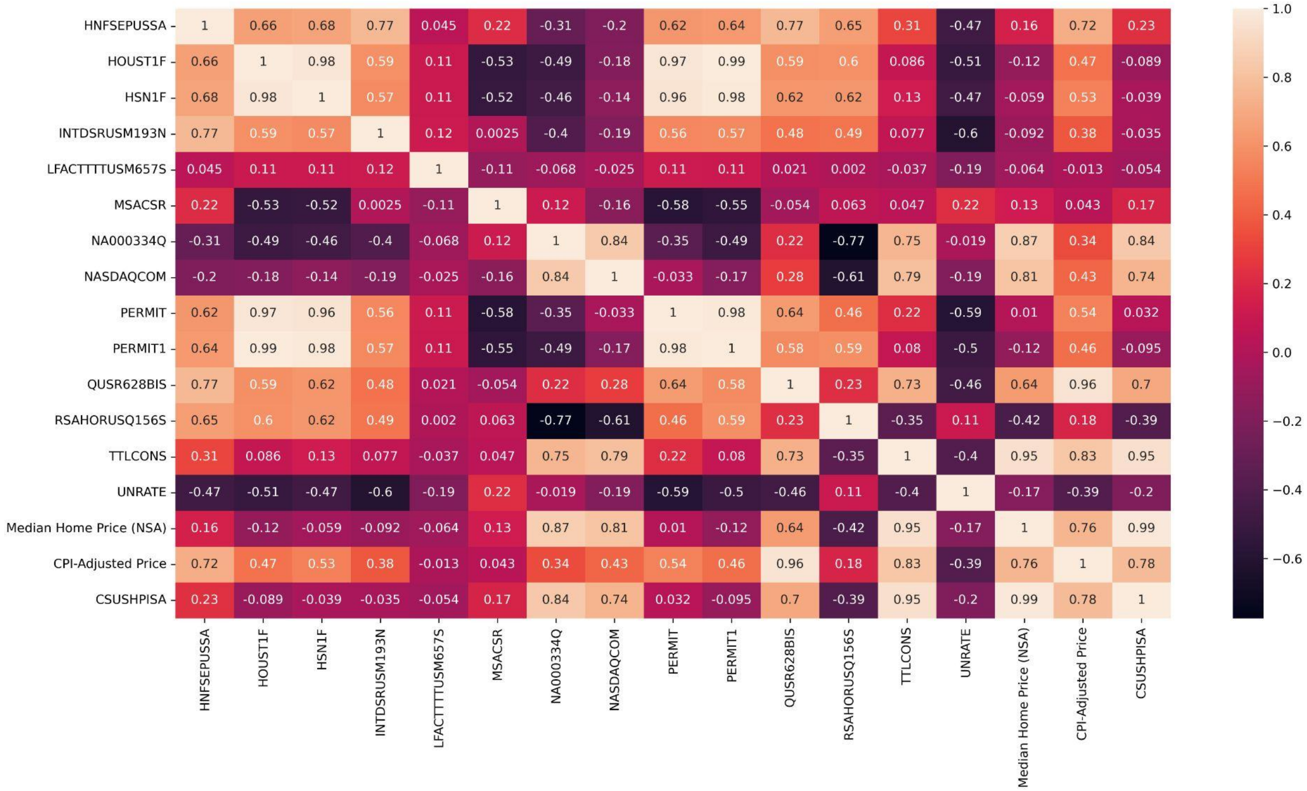
4. Result and Discussion

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 Median Home Price (NSA) 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 unemployment, meaning as one goes up, the other goes down.

4.2 Correlation Matrix:



	HNFSEPUSSA	HOUST1F	HSN1F	INTDSRUSM193N	LFACTTTTUSM657S	MSACSR	NA000334Q	NASDAQCOM	PERMIT	PERMIT1	QUSR628BIS	RSAHORUSQ156S	TTLCONS	UNRATE	Median Home Price (NSA)	CPI-Adjusted Price	CSUSHPISA
HNFSEPUSSA	1	0.66	0.68	0.77	0.045	0.22	-0.31	-0.2	0.62	0.64	0.77	0.65	0.31	-0.47	0.16	0.72	0.23
HOUST1F	0.66	1	0.98	0.59	0.11	-0.53	-0.49	-0.18	0.97	0.99	0.59	0.6	0.086	-0.51	-0.12	0.47	-0.089
HSN1F	0.68	0.98	1	0.57	0.11	-0.52	-0.46	-0.14	0.96	0.98	0.62	0.62	0.13	-0.47	-0.059	0.53	-0.039
INTDSRUSM193N	0.77	0.59	0.57	1	0.12	0.0025	-0.4	-0.19	0.56	0.57	0.48	0.49	0.077	-0.6	-0.092	0.38	-0.035
LFACTTTTUSM657S	0.045	0.11	0.11	0.12	1	-0.11	-0.068	-0.025	0.11	0.11	0.021	0.002	-0.037	-0.19	-0.064	-0.013	-0.054
MSACSR	0.22	-0.53	-0.52	0.0025	-0.11	1	0.12	-0.16	-0.58	-0.55	-0.054	0.063	0.047	0.22	0.13	0.043	0.17
NA000334Q	-0.31	-0.49	-0.46	-0.4	-0.068	0.12	1	0.84	-0.35	-0.49	0.22	-0.77	0.75	-0.019	0.87	0.34	0.84
NASDAQCOM	-0.2	-0.18	-0.14	-0.19	-0.025	-0.16	0.84	1	-0.033	-0.17	0.28	-0.61	0.79	-0.19	0.81	0.43	0.74
PERMIT	0.62	0.97	0.96	0.56	0.11	-0.58	-0.35	-0.033	1	0.98	0.64	0.46	0.22	-0.59	0.01	0.54	0.032
PERMIT1	0.64	0.99	0.98	0.57	0.11	-0.55	-0.49	-0.17	0.98	1	0.58	0.59	0.08	-0.5	-0.12	0.46	-0.095
QUSR628BIS	0.77	0.59	0.62	0.48	0.021	-0.054	0.22	0.28	0.64	0.58	1	0.23	0.73	-0.46	0.64	0.96	0.7
RSAHORUSQ156S	0.65	0.6	0.62	0.49	0.002	0.063	-0.77	-0.61	0.46	0.59	0.23	1	-0.35	0.11	-0.42	0.18	-0.39
TTLCONS	0.31	0.086	0.13	0.077	-0.037	0.047	0.75	0.79	0.22	0.08	0.73	-0.35	1	-0.4	0.95	0.83	0.95
UNRATE	-0.47	-0.51	-0.47	-0.6	-0.19	0.22	-0.019	-0.19	-0.59	-0.5	-0.46	0.11	-0.4	1	-0.17	-0.39	-0.2
Median Home Price (NSA)	0.16	-0.12	-0.059	-0.092	-0.064	0.13	0.87	0.81	0.01	-0.12	0.64	-0.42	0.95	-0.17	1	0.76	0.99
CPI-Adjusted Price	0.72	0.47	0.53	0.38	-0.013	0.043	0.34	0.43	0.54	0.46	0.96	0.18	0.83	-0.39	0.76	1	0.78
CSUSHPISA	0.23	-0.089	-0.039	-0.035	-0.054	0.17	0.84	0.74	0.032	-0.095	0.7	-0.39	0.95	-0.2	0.99	0.78	1

Considering the top 5 correlated factors with the S&P/Case-Shiller Home Price Index (CSUSHPISA):

- Median Home Price has a high correlation of **0.99** with CSUSHPISA.
- Total Construction Spending (TTLCONS) is closely correlated with a value of **0.95** with CSUSHPISA.
- Gross Domestic Product (NAA000334Q) shows a substantial correlation of **0.84** with CSUSHPISA.
- The Consumer Price Index (CPI) is notably correlated with a value of **0.78** with CSUSHPISA.
- The NASDAQ Composite Index (NASDAQCOM) exhibits a significant correlation of 0.**74** with CSUSHPISA.

4.3 Machine Learning Models:

The objective is not only to achieve a good R-squared but also to identify important parameters. Lasso regression, with an R-squared of **0.993**, tends to eliminate many features, which doesn't align with the objective. On the other hand, Ridge regression, 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, Elastic Net regression was chosen, combining aspects of both Lasso and Ridge. To find the right hyperparameters, manual tuning was conducted, resulting in an **alpha value** of **1** and an **l1_ratio** of **0.75**. This configuration achieved an R-squared of **0.970** and was successful in identifying important features, aligning with the overarching goal.

The highly important features selected by the Elastic Net model are:

- Median Home Price (NSA)
- NA000334Q
- TTLCONS
- QUSR628BIS
- CPI-Adjusted Price

- NASDAQCOM
- MSACSR

5. Conclusions:

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

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

- **Gross Domestic Product (NA000334Q):** A strong GDP indicates a robust economy, 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.

- **Real Residential Property Prices (QUSR628BIS):** 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.