

HOME LLC

Objective

To build a data science model that explains how these factors impacted home prices over the last 20 years by using publicly available data for key supply-demand factors that influence US home prices nationally.

Data

Two datasets are collected for this assignment :-

1. Supply Data
2. Demand Data

These datasets contain quarterly data on key supply-demand factors that influence US home prices nationally in the last 20 years and are collected from Kaggle.

1. Supply Data

Supply File has 6 features :-

- **DATE:** The date of the observation. (2003 - 2023)
- **PERMIT:** New Privately-Owned Housing Units Authorized in Permit-Issuing Places: Total Units (Thousands of Units, Seasonally Adjusted Annual Rate). This variable represents the number of new housing units authorized for construction in permit-issuing places.
- **MSACSR:** Monthly Supply of New Houses in the United States (Seasonally Adjusted). It indicates the monthly supply of new houses available in the United States.
- **TLRESCONS:** Total Construction Spending: Residential in the United States (Millions of Dollars, Seasonally Adjusted Annual Rate). This variable represents the total construction spending on residential projects.
- **EVACANTUSQ176N:** Housing Inventory Estimate: Vacant Housing Units in the United States (Thousands of Units, Not Seasonally Adjusted). It provides an estimate of the number of vacant housing units in the United States.

- **CSUSHPISA:** S&P/Case-Shiller U.S. National Home Price Index (Index Jan 2000=100, Seasonally Adjusted). This variable serves as a proxy for home prices and represents the home price index for the United States.

2. Demand Data

Demand File also has 6 features :-

- **INTDSRUSM193N:** Interest Rates, Discount Rate for United States (Billions of Dollars, Seasonally Adjusted Annual Rate). This variable represents the interest rates or discount rates for the United States.
- **UMCSENT:** University of Michigan: Consumer Sentiment. It measures the consumer sentiment index based on surveys conducted by the University of Michigan.
- **GDP:** Gross Domestic Product (Billions of Dollars, Seasonally Adjusted Annual Rate).
- **MORTGAGE15US:** 30-Year Fixed Rate Mortgage Average in the United States (Percent, Not Seasonally Adjusted). It indicates the average interest rate for a 30-year fixed-rate mortgage.
- **MSPUS:** Median Sales Price of Houses Sold for the United States (Not Seasonally Adjusted)
- **CSUSHPISA:** S&P/Case-Shiller U.S. National Home Price Index (Index Jan 2000=100, Seasonally Adjusted). This variable serves as a proxy for home prices and represents the home price index for the United States.

The S&P Case-Shiller U.S. National Home Price Index (CSUSHPISA) is used as a stand-in for housing prices and serves as the outcome variable in this analysis. These datasets offer valuable insights into different supply and demand elements and their potential impact on housing prices in the United States. Conducting an analysis of this data and constructing a data science model can aid in discovering patterns and connections between these factors and fluctuations in home prices spanning the last twenty years.

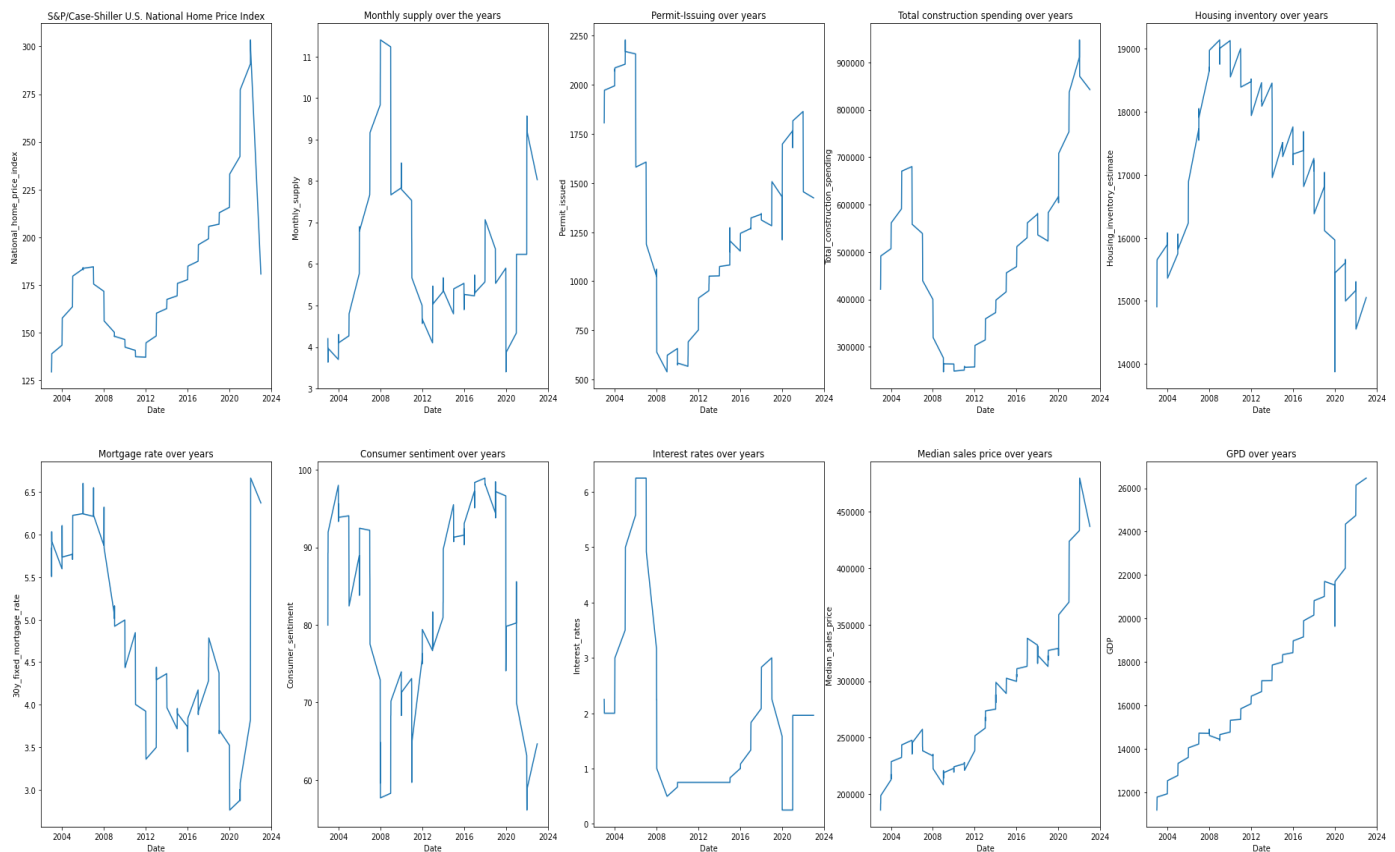
For better understanding, columns name have been changed :-

- CSUSHPISA: National_home_price_index
- MSACSR : Monthly_supply

- PERMIT: Permit_issued
- TLRESCONS : Total_construction_spending
- EVACANTUSQ176N : Housing_inventory_estimate
- MORTGAGE30US : 30y_fixed_mortgage_rate
- UMCSENT : Consumer_sentiment
- INTDSRUSM193N : Interest_rates
- MSPUS : Median_sales_price
- DATE : Date

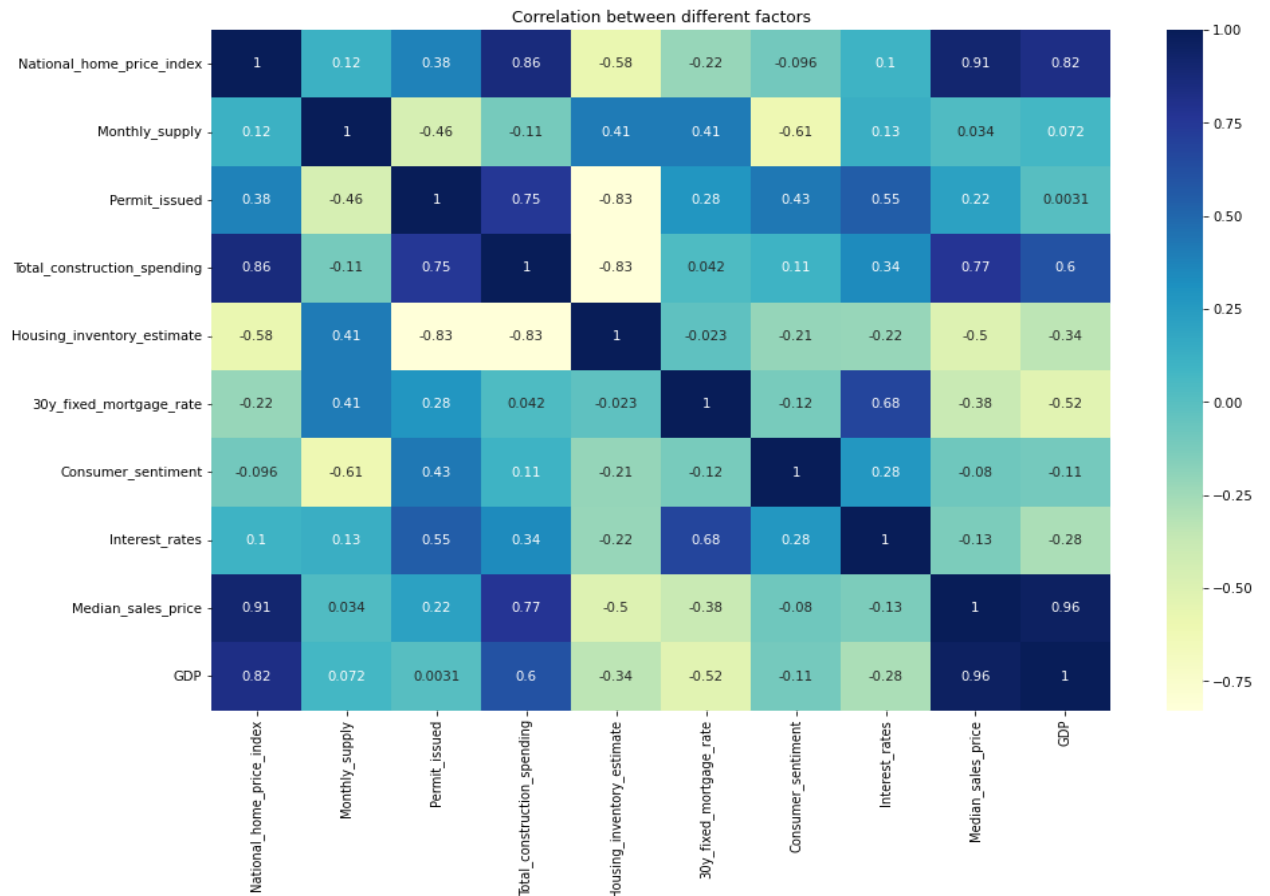
Exploratory Data Analysis

1. Understanding trends for each factors over the years



2. Correlation Heatmap

In order to understand which factor is influencing the CSUSHPISA(S&P/Case-Shiller U.S. National Home Price Index), a correlation analysis is conducted with the results shown in below figure.



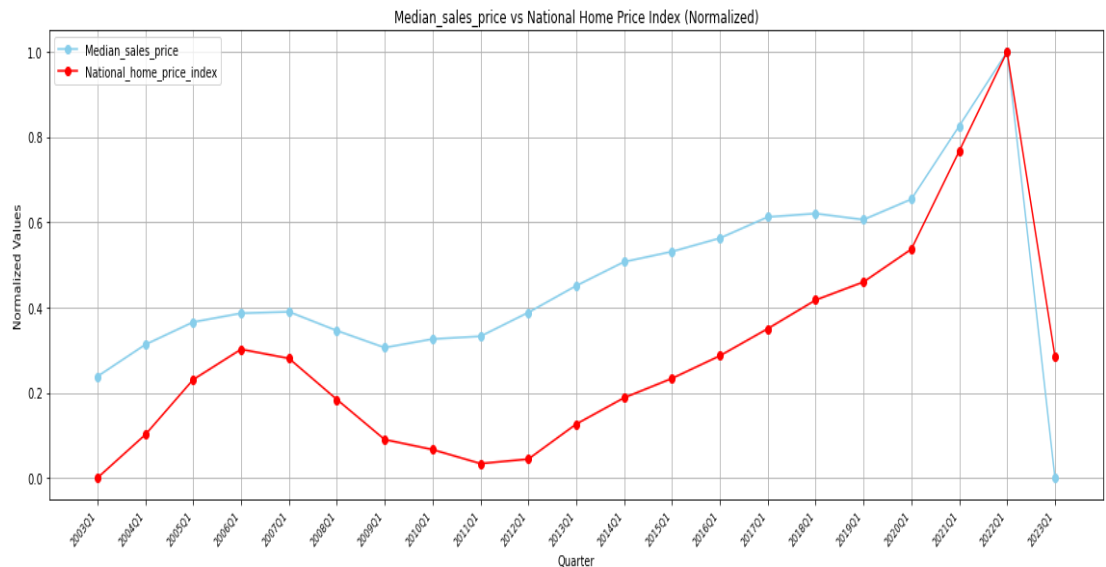
- **Monthly_supply**(Monthly Supply of New Houses): There exists a slight positive connection (0.12) between the monthly supply of new homes and the National Home Price Index (CSUSHPISA). This suggests that when the supply of new houses increases, it may have a minor beneficial impact on home prices.
- **Permit_issued**(New Privately-Owned Housing Units Authorized): A moderate positive relationship (0.38) is observed between the authorization of new housing units and the National Home Price Index (CSUSHPISA). This implies that a higher number of authorized housing units could positively affect home prices.

- **Total_construction_spending**(Total Construction Spending: Residential): A strong positive correlation (0.86) is evident between total spending on residential construction projects and the National Home Price Index (CSUSHPISA). This indicates a strong association between increased construction spending and higher home prices.
- **Housing_inventory_estimate**(Housing Inventory Estimate: There is a moderate negative correlation (-0.58) between the estimated count of vacant housing units and the National Home Price Index (CSUSHPISA). This suggests that a higher number of vacant housing units may exert a downward trend on home prices.
- **30y_fixed_mortgage_rate**(30-Year Fixed Rate Mortgage Average): A weak negative connection (-0.22) is identified between mortgage interest rates and the National Home Price Index (CSUSHPISA). This implies that higher mortgage rates are associated with slightly lower home prices.
- **Consumer_sentiment**(University of Michigan: Consumer Sentiment): There is a weak negative correlation (-0.096) between consumer sentiment and the National Home Price Index (CSUSHPISA). Lower consumer sentiment may have impacted slightly lower home prices, although the relationship is weak.
- **Interest_rates**(Interest Rates, Discount Rate): Positive correlation (0.1): There is a weak positive relationship between interest rates or discount rates and CSUSHPISA(National_home_price_index). Higher rates are associated with slightly higher home prices.
- **Median_sales_price**(Median Sales Price of Houses Sold): A strong positive correlation (0.91) is observed between the median sales prices of houses sold and the National Home Price Index (CSUSHPISA). Higher median sales prices are strongly associated with increased home prices.
- **GDP** (Gross Domestic Product): A strong positive correlation (0.82) is evident between GDP and the National Home Price Index (CSUSHPISA). Higher GDP is strongly tied to higher home prices.

3. Quarterly Trend Analysis on House Price Index

To understand how the House Price Index is changing in the last 20 years, we have plotted a quarterly line plot of highly correlated factors with the National House Price Index.

- **Median Sales Price vs National Home Price Index (correlation coefficient ~0.91)**

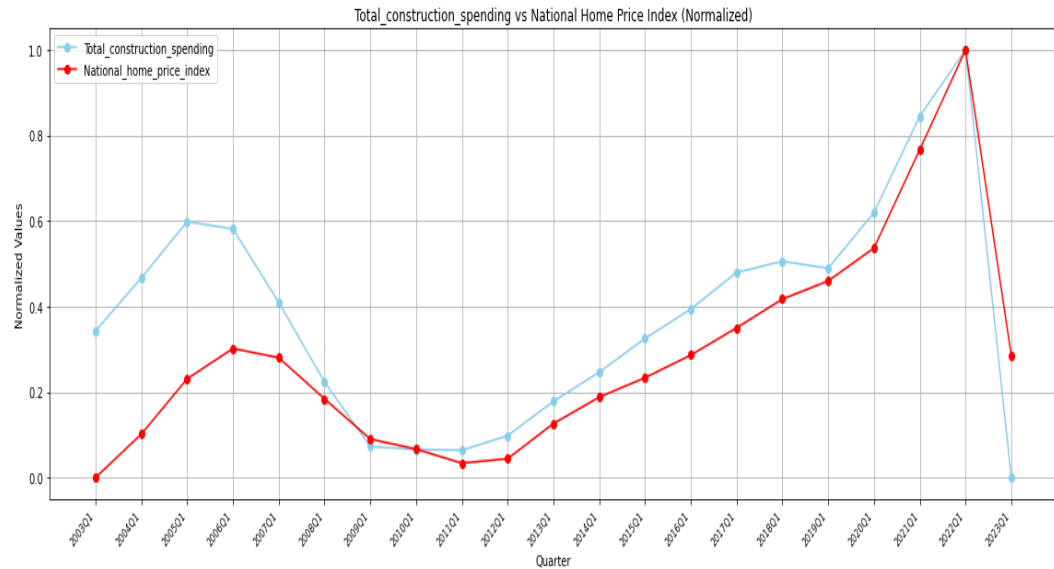


The median sale price, which represents the midpoint of house sale prices in the United States (with half selling above and half selling below), is the most important factor. Elevated median sale prices are closely linked to increased home prices as we can see the above figure.

- **Total Construction Spending vs National Home Price Index (correlation coefficient ~0.86)**

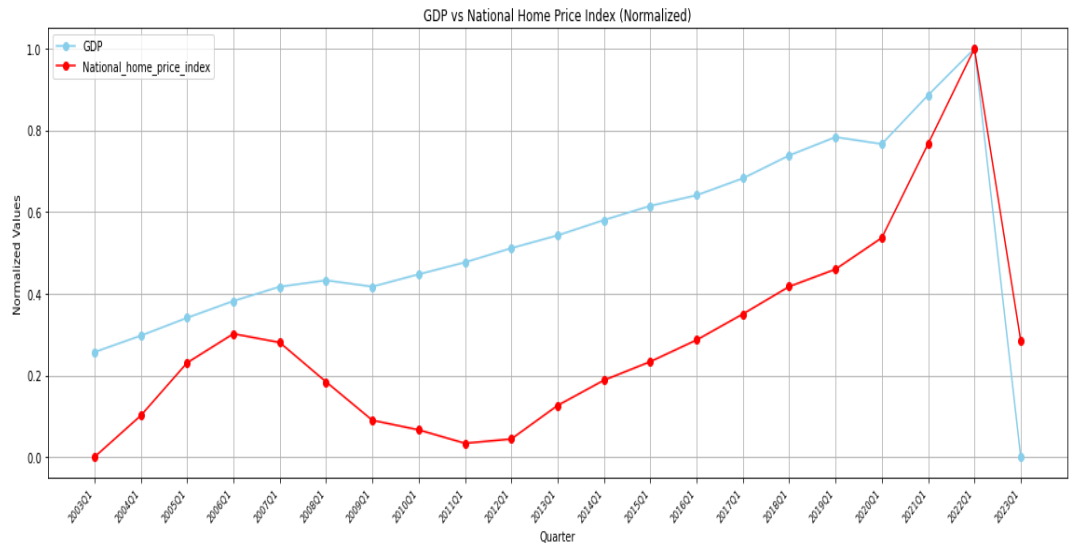
Total Construction Spending in the United States signifies the complete expenditure on residential construction projects. As per Investopedia, residential construction spending accounts for approximately half of the total construction spending in the U.S., and the health of the housing market can be gauged by monitoring new home construction, which typically increases when consumers have confidence in their employment and economic circumstances. The total

spending on residential construction exhibits a robust positive relationship with home prices. This implies that increased expenditure on residential construction tends to drive up home prices.



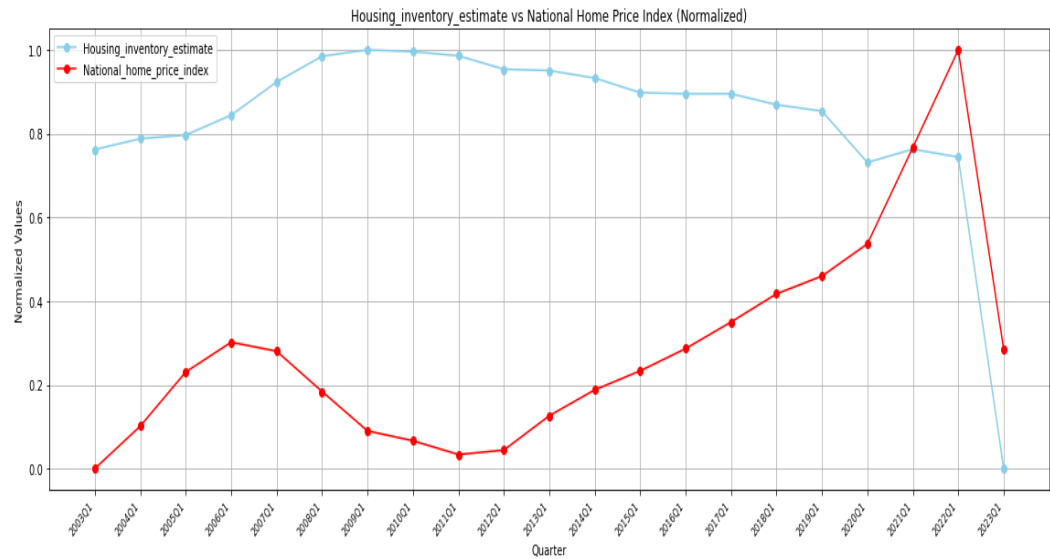
- **GDP vs National Home Price Index(correlation coefficient ~0.82)**

GDP typically rises when the value of goods and services sold by domestic producers to foreign countries exceeds the value of foreign goods and services purchased by domestic consumers, resulting in a trade surplus. The housing market and the broader economy are intricately connected. When real estate prices increase, homeowners often feel more confident in their investments and more willing to spend. This, in turn, leads to increased investment in constructing new homes, contributing to the overall growth of the gross domestic product; similar trends can be seen in the plotted graph below.



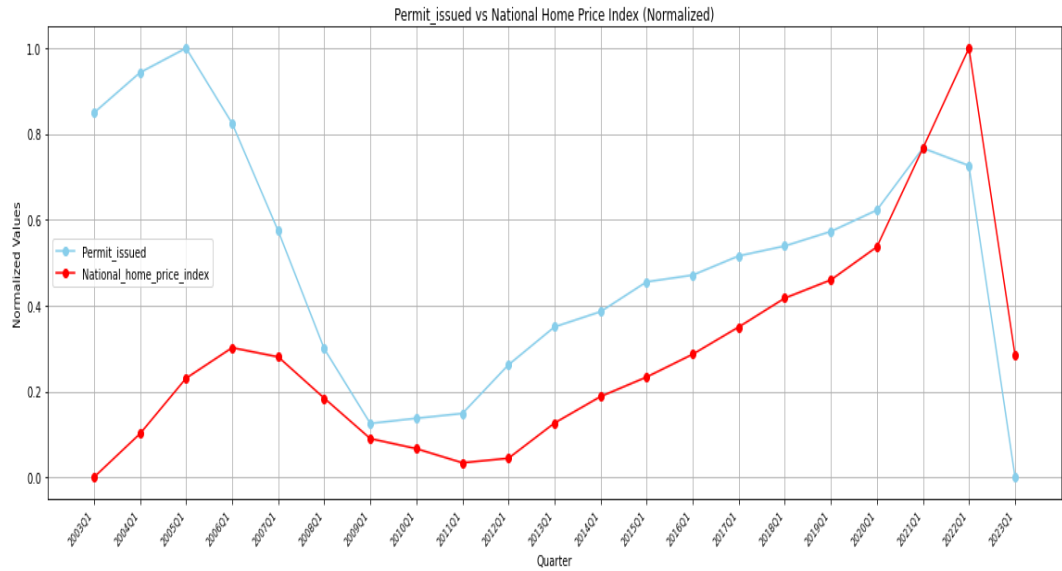
- **Housing Inventory Estimate vs National Home Price Index(correlation coefficient ~ -0.58)**

The housing inventory estimate of vacant housing units in the United States can impact the S&P/Case-Shiller U.S. National Home Price Index. A shortage of available homes in the market can lead to price increases, often resulting in competitive bidding scenarios. In such cases, a particular property may become highly sought-after, leading multiple potential buyers to engage in bidding wars, each attempting to outbid the others by offering higher purchase prices. In simpler terms, if the number of vacant housing units rises, it implies that property values may face downward pressure, potentially causing a decline in home prices which can be seen in quarterly trend plot.



- **Permit Issued vs National Home Price Index(correlation coefficient ~0.38)**

Permit Issued is an economic indicator that quantifies the quantity of new privately-owned housing units sanctioned through building permits in areas where permits are issued. It serves as a gauge for assessing the robustness of the housing market and the broader economy. The issuance of residential building permits can be indicative of consumer confidence and financial stability. There exists a moderate positive correlation between the number of new privately-owned housing units authorized and home prices. This implies that when construction permits for more housing units are granted, it tends to elevate home prices. This phenomenon occurs because a reduction in the supply of homes, as well as constraints on labor and building materials, results in price increases.



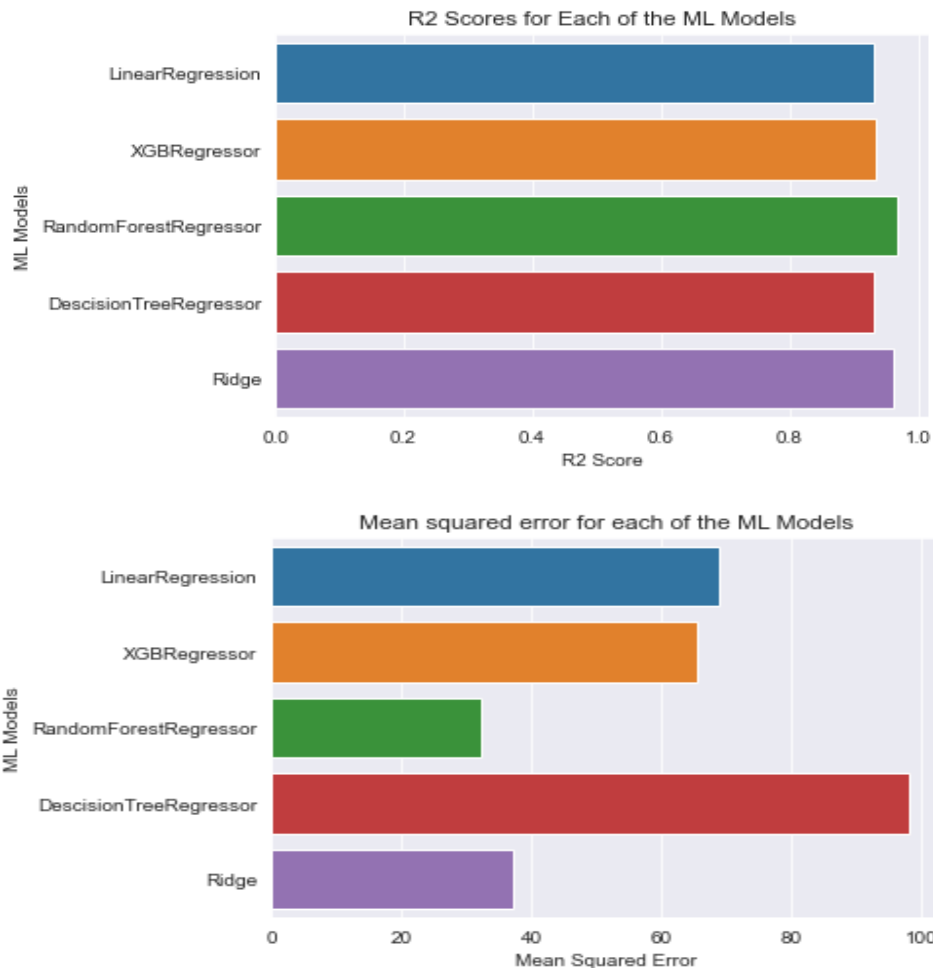
ML Model

In this analysis, we have used five models :-

1. Linear Regression
2. XGB Regressor
3. Random Forest Regressor
4. Decision Tree Regressor
5. Ridge

Out of all these Random Forest Regressor Models performed best with best R2 score of 0.9675 and least mean squared error value of 32.33.

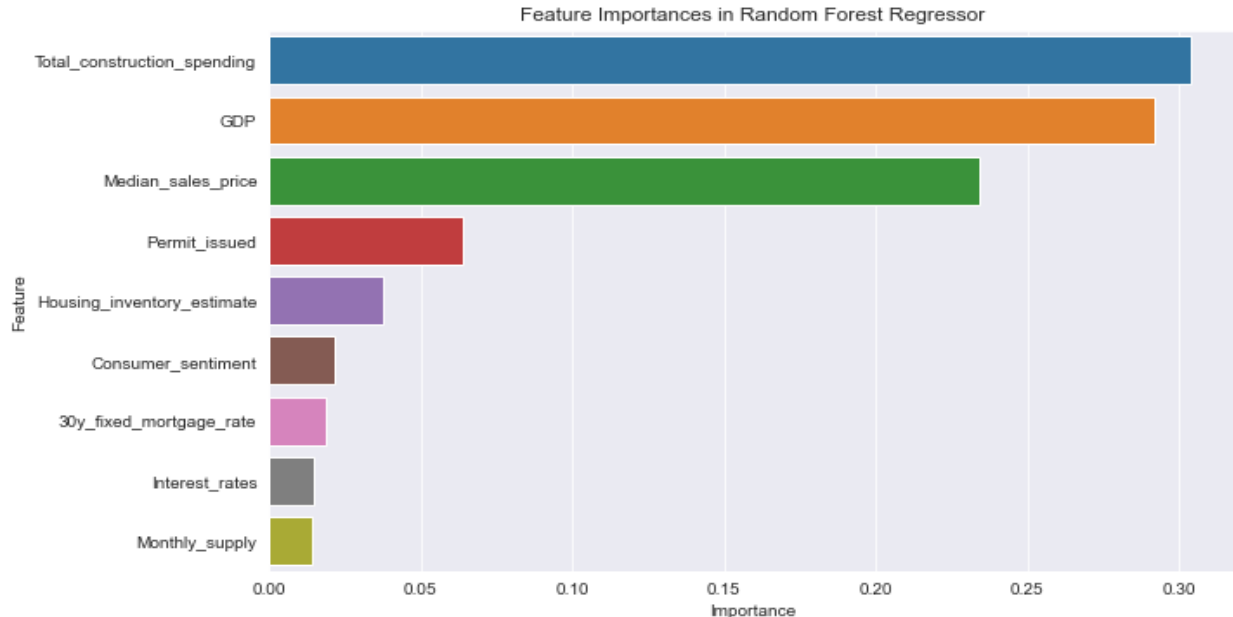
	Models	R2 Score		Models	MSE
0	LinearRegression	0.9307	0	LinearRegression	68.8879
1	XGBRegressor	0.9341	1	XGBRegressor	65.4954
2	RandomForestRegressor	0.9675	2	RandomForestRegressor	32.3335
3	DescisionTreeRegressor	0.9012	3	DescisionTreeRegressor	98.2115
4	Ridge	0.9626	4	Ridge	37.2031



Model Evaluation

After evaluating these all five models by using R2 score and mean squared error(MSE), We are able to find the model with best accuracy and least error. As we know A lower MSE indicates better performance, as it reflects smaller prediction errors and The R-squared score helps us understand how well the independent variables (features) explain the variation in the dependent variable (target). In this evaluation, the Random Forest Regressor model demonstrated strong performance. The Mean Squared Error (MSE) for the testing dataset was 32.32, indicating that prediction errors were relatively low. Additionally, the R-squared score stood at 0.9675, signifying that the model can account for approximately 96.75% of the variability observed in the target variable. After examining the coefficients of the Random Forest Regression model, we are able to understand the significance and influence of each feature on the predicted target variable. Here are some notable findings:-

Feature	Importance
Total_construction_spending	0.303657
GDP	0.292082
Median_sales_price	0.233893
Permit_issued	0.064084
Housing_inventory_estimate	0.037668
Consumer_sentiment	0.021352
30y_fixed_mortgage_rate	0.018615
Interest_rates	0.014440
Monthly_supply	0.014208



All the features are significantly impacting the predicted home price index but the features like Total construction spending, GDP, Median Sales Price and Permit issued which also had the highest correlation with national home price index have more impact than rest of the factors.

Insights

Supply Factors

- Total construction spending on residential projects (TLRESCONS) demonstrates a strong positive correlation with home prices. This indicates that increased construction

spending is closely linked to higher home prices. The reason behind this is straightforward: construction expenses encompass building materials, labor costs, and other expenditures, ultimately raising the overall cost of houses.

- The S&P/Case-Shiller U.S. National Home Price Index (CSUSHPISA) exhibits a very weak positive correlation with the monthly supply of new homes (MSACSR). Given the feeble correlation, it can be inferred that there is a negative association (due to increased supply). This implies that an upsurge in the supply of new homes may have a minor adverse impact on housing prices.
- The number of authorized housing units (PERMIT) is moderately positively connected to home values. This means that a higher count of authorized housing units could contribute to elevated property prices. This is because when more housing units receive authorization for construction, it affects the availability of homes, materials, and labor.
- The estimated number of vacant housing units (EVACANTUSQ176N) exhibits a moderate negative correlation with home prices. This implies that a higher quantity of vacant housing units may exert downward pressure on home prices. This occurs because an increase in housing supply tends to lead to a reduction in home prices.

Demand Factors

- Gross Domestic Product (GDP) shows a strong positive correlation with home prices. Higher GDP is closely connected to increased home prices.
- The median sales price of houses sold (MSPUS) exhibits a strong positive correlation with home prices. Higher median sales prices are strongly associated with elevated home prices.
- The average interest rate for a 30-year fixed-rate mortgage (MORTGAGE30US) demonstrates a weak negative correlation with home prices. This suggests that when mortgage rates are higher, it's associated with slightly lower home prices. An increase in the federal funds rate can lead to higher mortgage rates, and elevated mortgage rates can dampen demand for home purchases, leading to a decrease in home prices.
- Consumer sentiment (UMCSENT) exhibits a weak negative correlation with home prices. Lower consumer sentiment is linked to slightly lower home prices. When

consumers have confidence in the economy, they are more inclined to invest in significant expenses like homes, which can drive up home prices.

- Data indicates that interest rates or discount rates (INTDSRUSM193N) have a weak positive association with home prices, suggesting a scaling or trailing issue. In reality, interest rates are negatively correlated with house prices. Higher interest rates tend to result in lower home prices.

Conclusion

After analyzing the correlation, different ML Models and Random forest Regressor coefficients , we conclude that :-

- Supply-related factors, such as housing inventory and the authorization of housing units, positively impact home prices. Increased spending on residential construction projects also significantly contributes to higher home prices.
- On the other hand, demand-related factors like mortgage interest rates have a negative effect on home prices. Elevated mortgage rates and reduced consumer sentiment are associated with slightly lower home prices.
- Economic factors, including GDP, Total Construction Spending, interest rates, play a pivotal role in determining home prices. A robust economy characterized by higher GDP and somewhat lower interest rates tends to bolster higher home prices.
- The median sales price of houses sold strongly correlates with home prices, illustrating the significance of market dynamics and buyer behavior in shaping home price trends.
- These insights hold value for various stakeholders in the real estate sector, such as homebuyers, sellers, developers, and policymakers. Understanding the factors that influence home prices can facilitate well-informed decisions regarding investments, financing, and economic strategies.