Home LLC report

07 September 2023 11:16 AM

Problem statement:
Find publicly available data for key *supply-demand* 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.

The following are the common demand and supply factors that influence any Nation's Home prices.

A. Demand Factors

i. Economic Conditions
ii. Interest Rates
iii. Demographics
iv. Consumer Confidence
v. Government Policies
vi. Market Speculations.

B. Supply Factors

i. Construction and Development
ii. Land Availability
iii. Labor And Material Costs
iv. Regulation and Permitting
v. Market Speculations
vi. Existing Home inventory
vii. Natural Disasters
viii. Interest Rates
ix. Home Builder Confidence

Data Collection:

The table below is about the factors and its details Available publicly and why it is chosen for this study. All data available in the **DataSet Folder.**

Abbreviation	Frequency	Column	Article title	CSV file name	Reference	Explanation
HPI	Monthly	HomePriceIndex	S&P/Case-Shiller U.S. National Home Price Index	CSUSHPISA	S&P Dow Jones Indices LLC, S&P/Case-Shiller U.S. National Home Price Index [CSUSHPISA], retrieved from FRED, Federal Reserve Bank of St. Louis; https://fred.stlouisfed.org/series/CSUSHPISA, August 31, 2023.	The S&P CoreLogic Case-Shiller Home Price Indices are the leading measures of U.S. residential real estate prices, tracking changes in the value of residential real estate nationally.
CCI	Monthly	ConsumerConfidenceIndex	Consumer Price Index for All Urban Consumers: Housing in U.S. City Average	CPIHOSSL	U.S. Bureau of Labor Statistics, Consumer Price Index for All Urban Consumers: Housing in U.S. City Average [CPIHOSSL], retrieved from FRED, Federal Reserve Bank of St. Louis; https://fred.stlouisfed.org/series/CPIHOSSL, September 1, 2023	The Consumer Price Index (CPI) is a measure of the average change over time in the prices paid by urban consumers for a market basket of consumer goods and services. Indexes are available for the U.S. and various geographic areas. Average price data for select utility, automotive fuel, and food items are also available.
CSENT	Monthly	ConsumerSentiment	Consumer Sentiment	UMCSENT	"Surveys of Consumers, University of Michigan, University of Michigan: Consumer Sentiment © [UMCSENT], retrieved from FRED, Federal Reserve Bank of St. Louis, (Accessed on date)"	Consumer sentiment is a statistical measurement of the overall health of the economy as determined by consumer opinion. It takes into account people's feelings toward their current financial health,

						the health of the economy in the short-term, and the prospects for longer-term economic growth, and is widely considered to be a useful economic indicator.
IRPI	Quarterly	InterestRatesAndPriceIndexes	Interest Rates and Price Indexes; Contract Rate on 30-Year, Fixed-Rate Conventional One-to-Four-Family Residential Mortgage Commitments	BOGZ1FL073165103Q	Interest Rates and Price Indexes; Contract Rate on 30-Year, Fixed-Rate Conventional One-to-Four-Family Residential Mortgage Commitments, Level (BOGZ1FL073165103Q) | FRED73165103Q	Interest Rates and Price Indexes; Contract Rate on 30-Year, Fixed-Rate Conventional One-to-Four-Family Residential Mortgage Commitments
CLI	Monthly	CompositeLeadingIndicator	Leading Indicators OECD: Leading indicators: CLI: Amplitude adjusted for the United States	USALOLITOAASTSAM	OECD descriptor ID: LOLITOAA OECD unit ID: STSA OECD country ID: USA All OECD data should be cited as follows: OECD,"Main Economic Indicators - complete database"Main Economic Indicators(database) http://dx.doi.org/10.1787/data-00052-en(Accessed on date)Copyright, 2016, OECD. Reprinted with permission.	The composite leading indicator (CLI) is designed to provide early signals of turning points in business cycles showing fluctuation of the economic activity around its long term potential level. CLIs show short-term economic movements in qualitative rather than quantitative terms.
LC_yearly	Annual	LaborCost_yearly	Unit Labor Costs for Retail Trade: Building Material and Supplies Dealers (NAICS 4441) in the United States	IPUHN4441U101000000	U.S. Bureau of Labor Statistics, Unit Labor Costs for Retail Trade: Building Material and Supplies Dealers (NAICS 4441) in the United States [IPUHN4441U101000000], retrieved from FRED, Federal Reserve Bank of St. Louis; https://fred.stlouisfed.org/series/IPUHN4441U101000000, September 1, 2023.	Unit Labor Costs for Retail Trade: Building Material and Supplies Dealers (NAICS 4441) in the United States
NUR	Monthly	NationalUnemploymentRate	us-national-unemployment-rate	us-national-unemployment-rate	https://www.macrotrends.net/1316/us-national-unemployment-rate	us-national-unemployment-rate
HS	Monthly	HousingStarts	housing-starts-historical-chart	housing-starts-historical-chart	https://www.macrotrends.net/1314/housing-starts-historical-chart	Housing starts measures the annualised change in the number of new residential buildings that began construction during the reported month.
DGDP	Monthly	DebtToGDPRatio	debt-to-gdp-ratio-historical-chart	debt-to-gdp-ratio-historical-chart	https://www.macrotrends.net/1381/debt-to-gdp-ratio-historical-chart	historical data comparing the level of gross domestic product (GDP) with Federal Debt.
GDPGR	Annual	GDPGrowthRate_yearly	GDPGrowthRate	GDPGrowthRate	https://stats.oecd.org/Index.aspx?QueryId=60703#	The annual average rate of change of the gross domestic product (GDP) at market prices based on constant local currency, for a given national economy, during a specified period of time. It expresses the difference between GDP values from one period to the next as a proportion of the GDP from the earlier period, usually multiplied by 100.

	Monthly	NAHB Builder Confidence	Data Scrapped using developer mode	consolidated_permit_confidence	https://www.mortgagenewsdaily.com/data/builder-confidence	builder perceptions of current single-family home sales and sales expectations for the next six months as “good,” “fair” or “poor.”
	Monthly	Building Permits	Data Scrapped using developer mode		https://www.mortgagenewsdaily.com/data/building-permits	Building Permits data provides an estimate on the number of homes planning on being built.

Data Cleaning/ Consolidation:

The codes are available in the following files.

buildingPe…

merging cs…

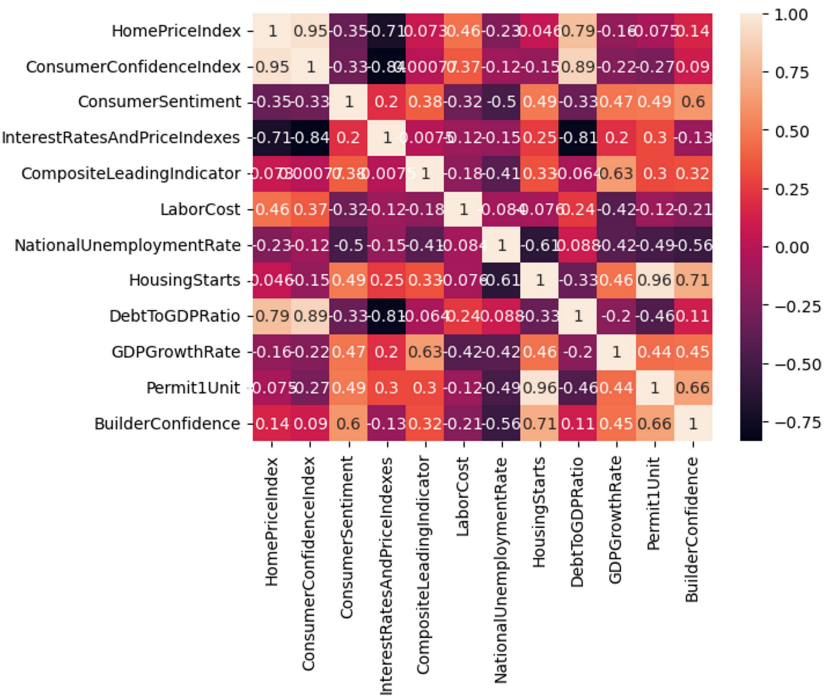
Exploratory Data Analysis and Identifying Feature Importance

EDA

The Correlation between The target(HomePriceIndex) and all the features can be viewed in the plot below.

Values between -0.2 to 0.2 indicate Zero correlation.

Inference: ConsumerConfidenceIndex shows a high correlation(0.95) with the target. Other likely-Higher correlated Features are DebtToGDPRatio(0.79),InterestRatesAndPriceIndexes(0.71), LaborCost(0.46), and ConsumerSentiment(0.35) .These features shows strong positive linear relationship with our target variable.



From <http://localhost:8888/lab/tree/JupyterLab%20folders/HomeLLCSep23/EDA.ipynb>

Since our problem is a supervised regression problem, Linear regression, XGB and SHAP method are used to find the feature importance.

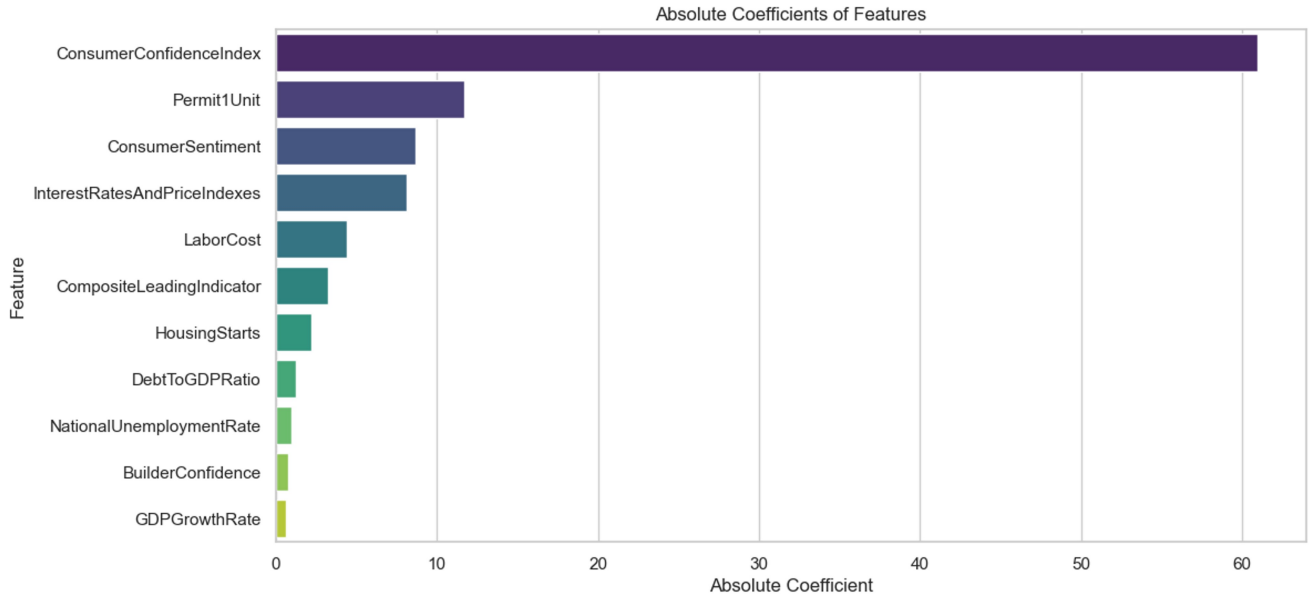
Model 1- Linear regression :

Linear regression is used for identifying feature importance in regression problems because it quantifies how each feature contributes to the target variable in a straightforward and interpretable way. The coefficients represent the feature's impact, aiding in the selection of the most influential features for prediction and model simplification.

Features	Coefficient
ConsumerConfidenceIndex	60.955484
ConsumerSentiment	-8.675500
InterestRatesAndPriceIndexes	8.106888
CompositeLeadingIndicator	3.221504
LaborCost	4.421297
NationalUnemploymentRate	-0.996055
HousingStarts	2.221966
DebtToGDPRatio	1.207567
GDPGrowthRate	0.604441
Permit1Unit	11.725068
BuilderConfidence	-0.737380

From <http://localhost:8888/lab/tree/JupyterLab%20folders/HomeLLCSep23/EDA.ipynb>

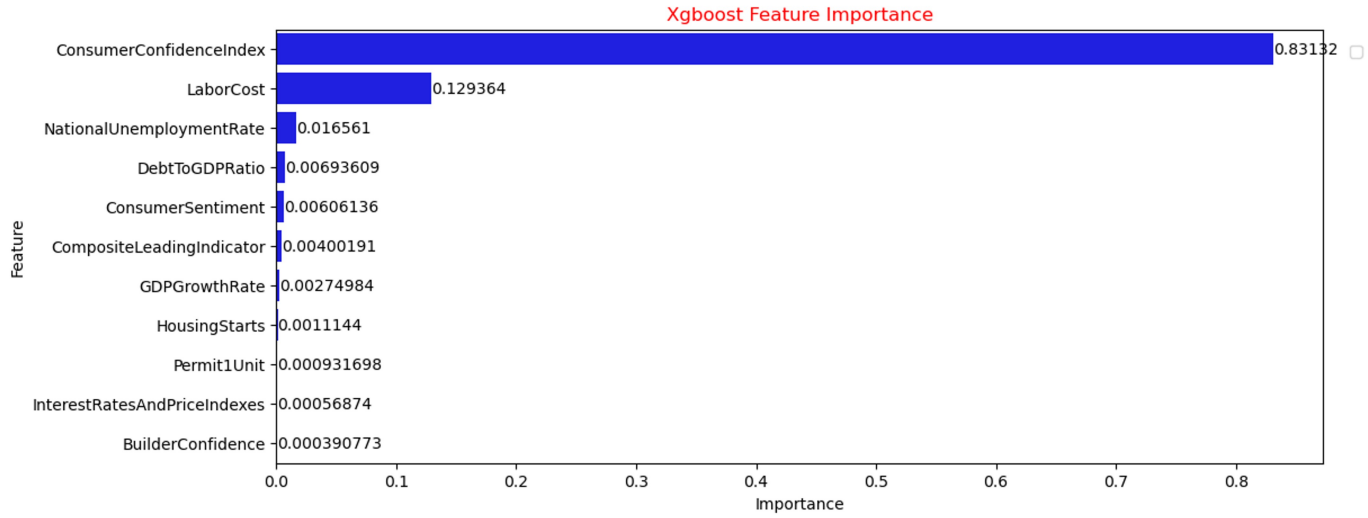
Absolute Coefficient values help in the determination of Feature importance.



From <http://localhost:8888/lab/tree/JupyterLab%20folders/HomeLLCSep23/EDA.ipynb>

Model 2- XGB:

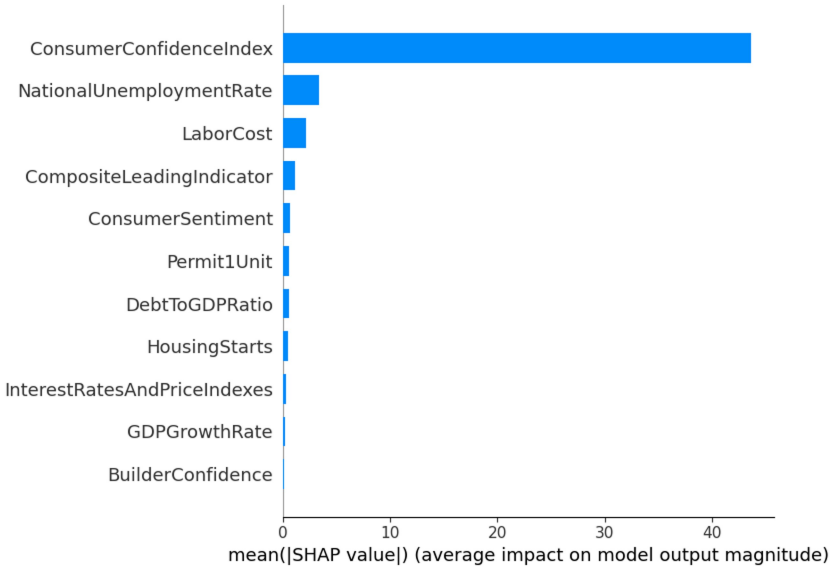
XGBoost is favored for feature importance in regression tasks because it excels at capturing complex interactions between features. Its ensemble of decision trees assigns importance scores, aiding in effective feature selection.



From <http://localhost:8888/lab/tree/JupyterLab%20folders/HomeLLCSep23/EDA.ipynb>

Method-SHAP:

SHAP (SHapley Additive exPlanations) is a popular method for identifying feature importance in regression problems. It is preferred because it offers comprehensive and interpretable insights into how each feature contributes to model predictions. SHAP values are based on cooperative game theory and provide a way to fairly distribute the contribution of each feature to the prediction outcome.



From <http://localhost:8888/lab/tree/JupyterLab%20folders/HomeLLCSep23/EDA.ipynb>

Feature Importance Comparison across models

Model/Method	**Linear Regression**	**XGB**	**SHAP Method**
Inference	1. ConsumerConfidenceIndex 2. Permit1Unit 3. ConsumerSentiment 4. InterestRatesAndPriceIndexes 5. LaborCost	1. ConsumerConfidenceIndex 2. LaborCost 3. NationalUnemploymentRate 4. DebtToGDPRatio 5. ConsumerSentiment	1. ConsumerConfidenceIndex 2. NationalUnemploymentRate 3. LaborCost 4. CompositeLeadingIndicator 5. ConsumerSentiment

Result:
The Analysis(human intervention-Part) of the above table shows that the following are the important features that influence US home prices *nationally*.

1. ConsumerConfidenceIndex
2. LaborCost
3. ConsumerSentiment

GITHUB REPO link:
https://github.com/janumirra/HOME_LLC_Sep2023/tree/main