

A photograph of a small, blue wooden boathouse or cabin situated on stilts in a large body of water. A wooden pier extends from the bottom center towards the boathouse. The water is a vibrant turquoise color. In the background, a distant shoreline with buildings is visible under a bright, slightly cloudy sky.

Real Estate Investment & California Price Prediction

THIP RATTANAVILAY

BELLEVUE UNIVERSITY / DSC-680

PROBLEM TO BE SOLVED

- HIGH MARKET PRICE
- MEDIAN HOUSEHOLD INCOME
- AFFORABLE HOUSING
- INVESTMENTS OPPORTUNITY



A photograph showing a person's hands interacting with a tablet. The tablet screen displays several blue bar charts. The person is wearing a dark blue suit jacket over a light-colored plaid shirt. In the background, there is a large green plant and a yellow cylindrical object on the left.

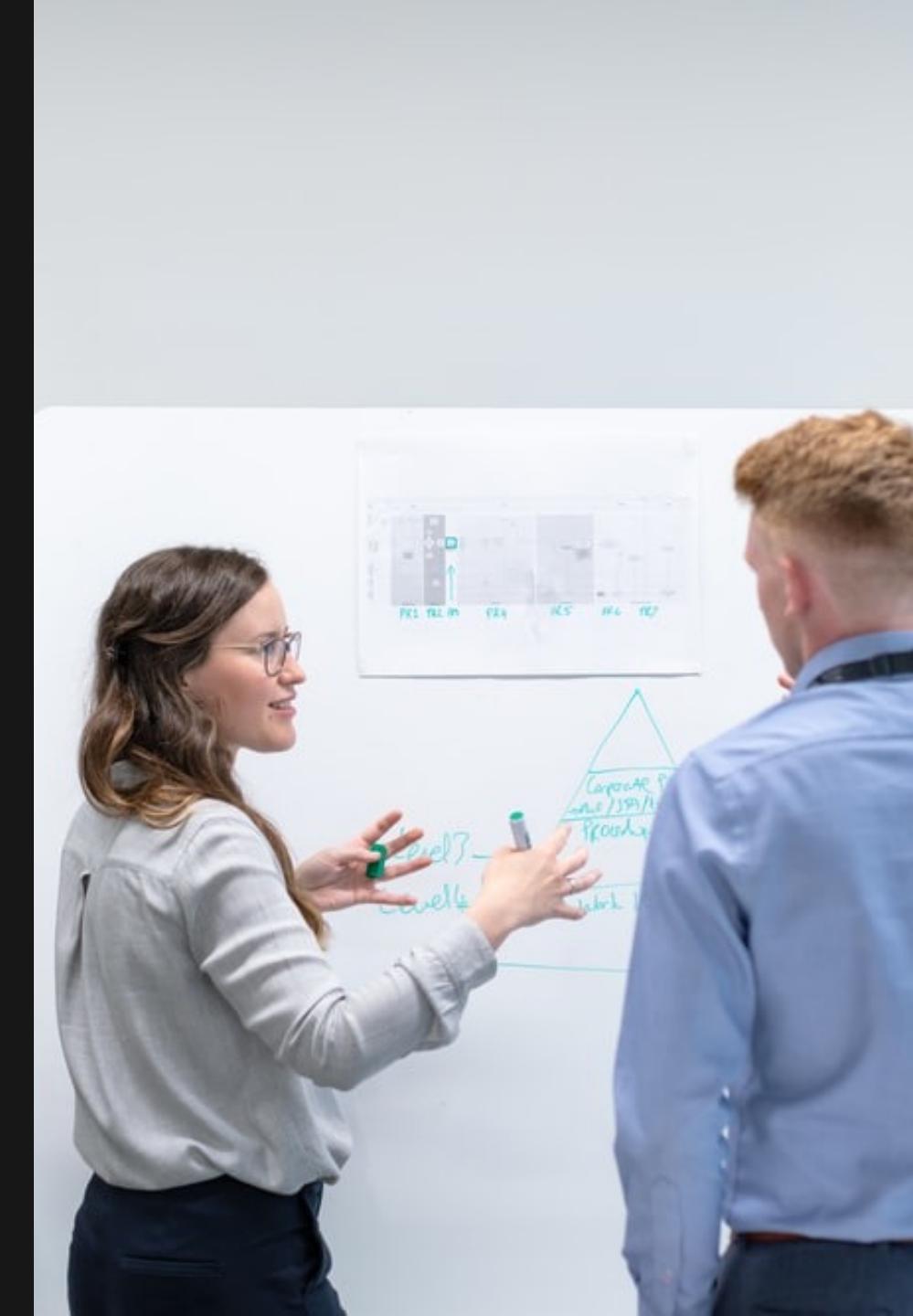
THE DATA

Data processing with two files (CA housing and Zillow)

Houses sold in 2016 (Zillow Transaction data) - 90k Properties. California averaged more than 200,000 new homes annually, with multifamily housing accounting for more of the housing production.

Properties data– 27 Million Properties (58 attributes)

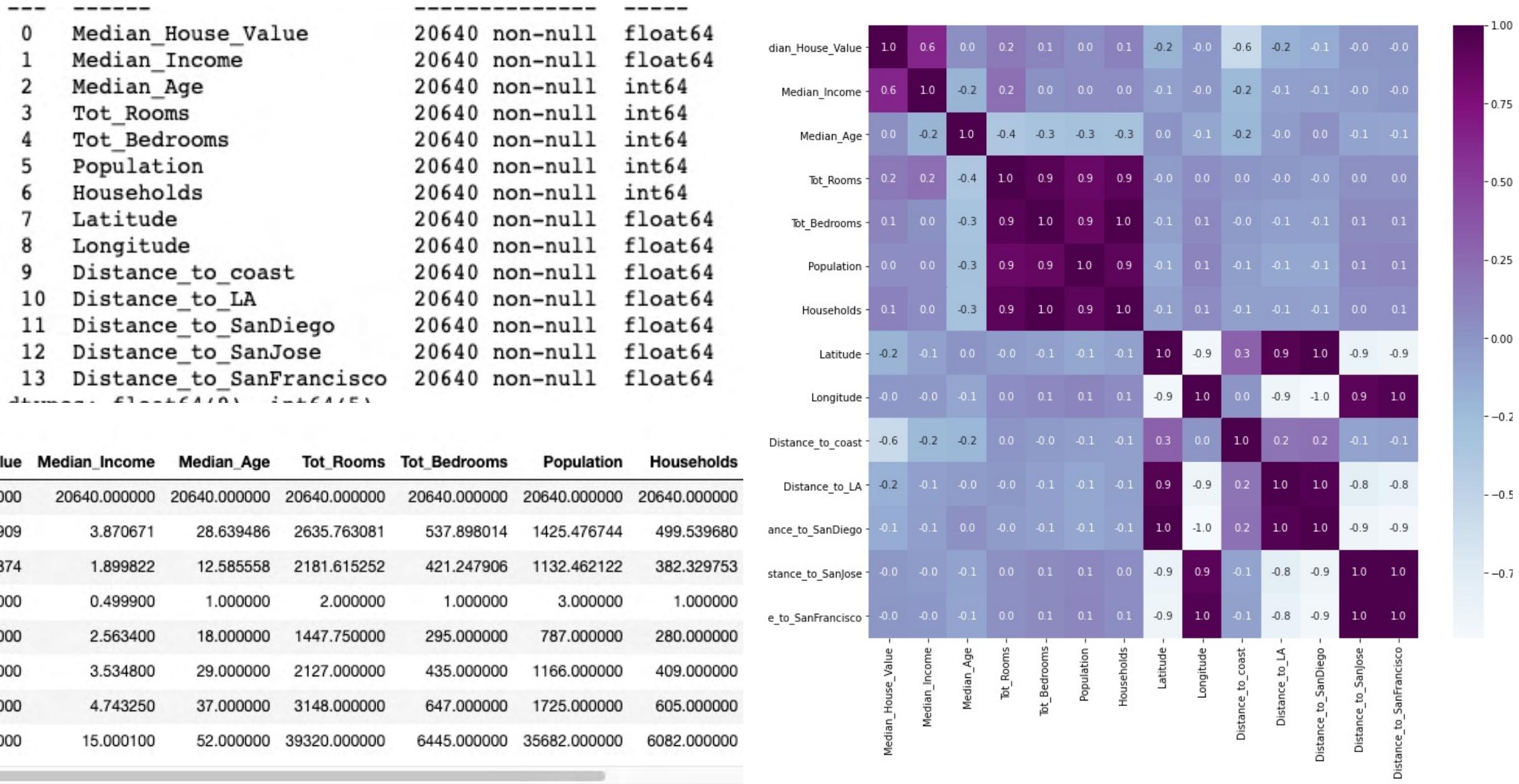
- Mortgage Records Attached
- Tax Assessments
- Physical features of the Property(Bed Count , Living Area, Age of the building)
- Location Details about the properties
- Amenities available in the property



METHOD & APPROACH



Exploratory Data Analysis





Missing Values & Duplications

	unique count
Median_House_Value	3842
Median_Income	12928
Median_Age	52
Tot_Rooms	5926
Tot_Bedrooms	1928
Population	3888
Households	1815
Latitude	862
Longitude	844
Distance_to_coast	12590
Distance_to_LA	12590
Distance_to_SanDiego	12590
Distance_to_SanJose	12590
Distance_to_SanFrancisco	12590

Let's dive into the data and look for unique values to ensure there are no duplicates or missing values.

Feature Engineering and Feature selection



The target variable is Median_House_Value

Dataset - California_Houses

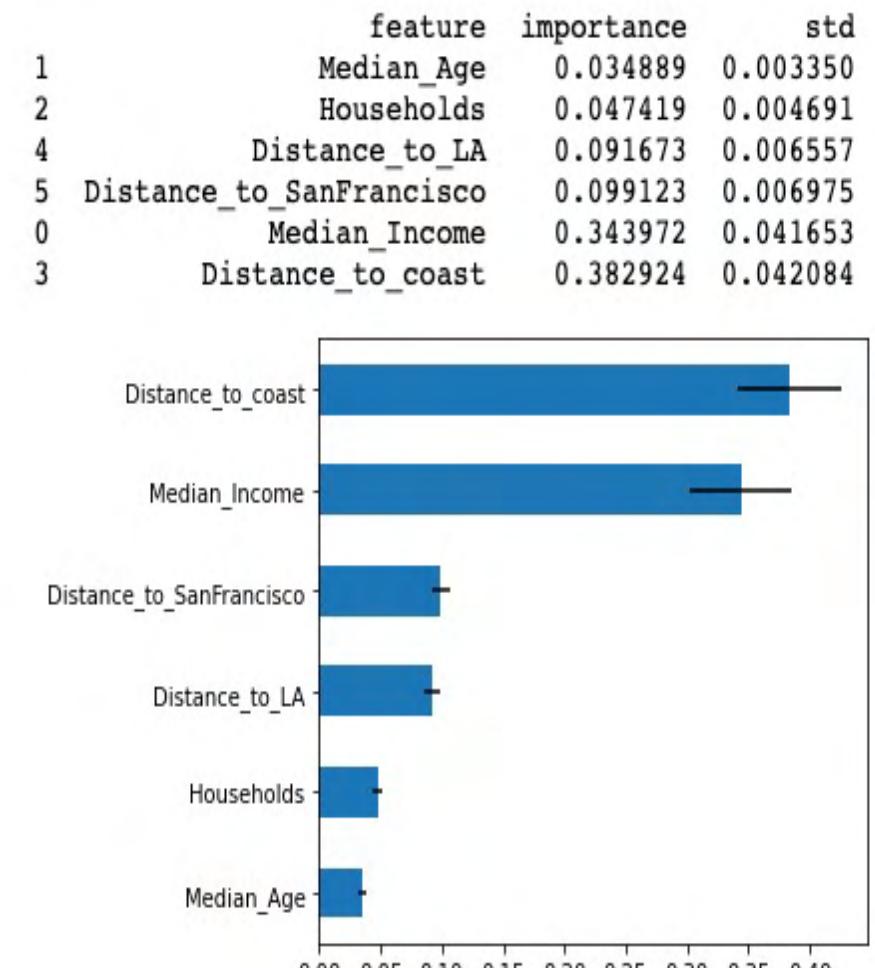
- Median_House_Value
- Median_Income
- Median_Age
- Tot_Rooms
- Tot_Bedrooms
- Population
- Households
- Latitude
- Longitude
- Distance_to_coast
- Distance_to_SanDiego
- Distance_to_SanJose
- Distance_to_SanFrancisco

Dataset – Zillow_California

- UnemploymentRate
- MedianMortageRate
- MedianMortageRate
- MedianSoldPrice_AllHomes.California

USED THE FOLLOWING TECHNIQUES

- Multiple Linear Regression
- Regression Tree Model
- Random Forest
- Lasso Regression
- Ridge Regression





Modeling



Multiple Linear Regression

MULTIPLE LINEAR REGRESSION:

Regression statistics

Mean Error (ME) : 0.0045

Root Mean Squared Error (RMSE) : 0.3183

Mean Absolute Error (MAE) : 0.2474

Mean Percentage Error (MPE) : -0.0365

Mean Absolute Percentage Error (MAPE) : 2.0652



Lasso Regression

LASSO:

Regression statistics

Mean Error (ME) : 0.0045

Root Mean Squared Error (RMSE) : 0.3183

Mean Absolute Error (MAE) : 0.2474

Mean Percentage Error (MPE) : -0.0365

Mean Absolute Percentage Error (MAPE) : 2.0652



Ridge Regression

Regression statistics

Mean Error (ME) : 0.0045

Root Mean Squared Error (RMSE) : 0.3183

Mean Absolute Error (MAE) : 0.2474

Mean Percentage Error (MPE) : -0.0365

Mean Absolute Percentage Error (MAPE) : 2.0652

RIDGE:



Regression Tree

REGRESSION TREE MODEL:

Regression statistics

Mean Error (ME) : -0.0049

Root Mean Squared Error (RMSE) : 0.4417

Mean Absolute Error (MAE) : 0.3343

Mean Percentage Error (MPE) : -0.1088

Mean Absolute Percentage Error (MAPE) : 2.7865



Random Forest

RANDOM FOREST:

Regression statistics

Mean Error (ME) : 0.0007

Root Mean Squared Error (RMSE) : 0.2300

Mean Absolute Error (MAE) : 0.1647

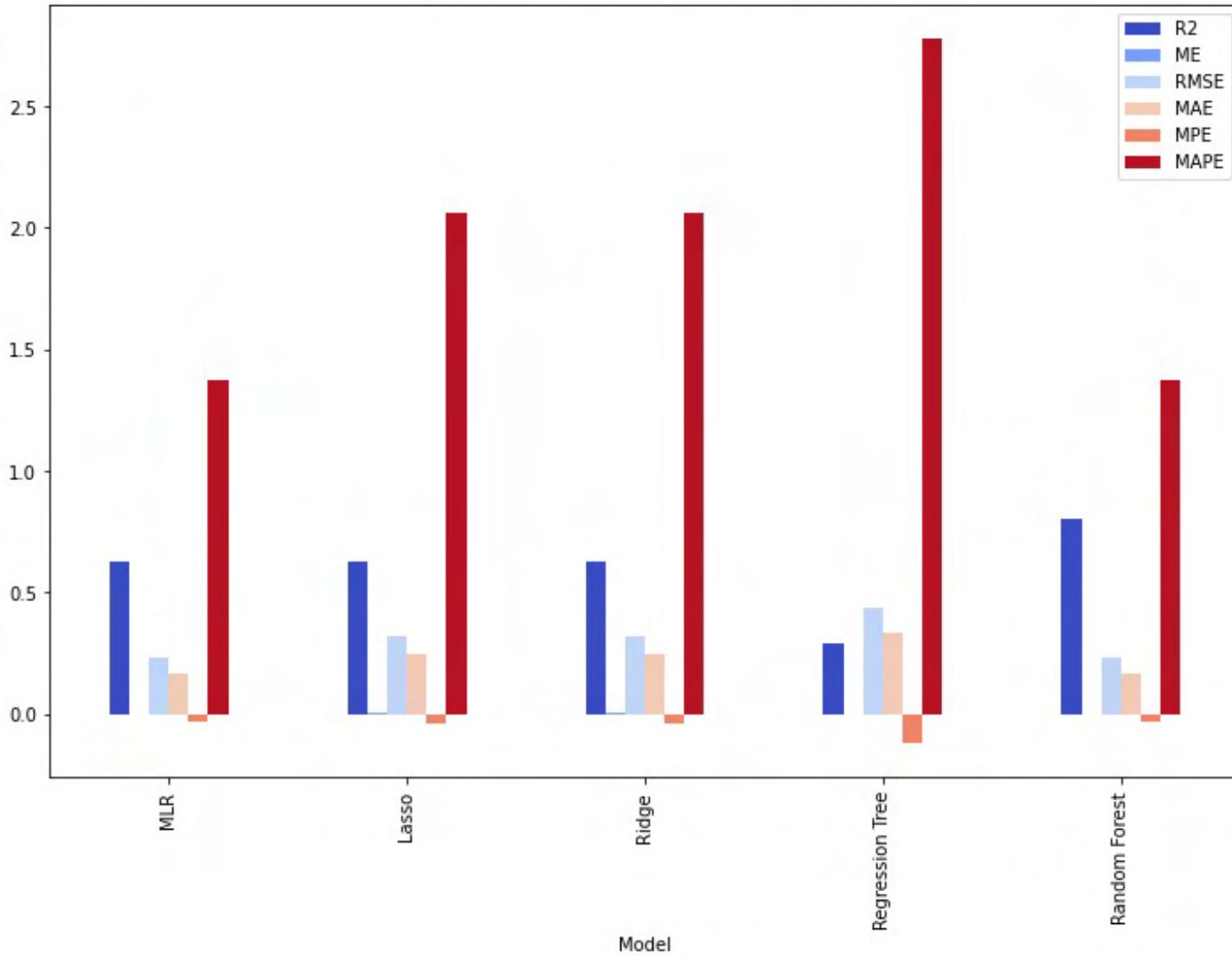
Mean Percentage Error (MPE) : -0.0341

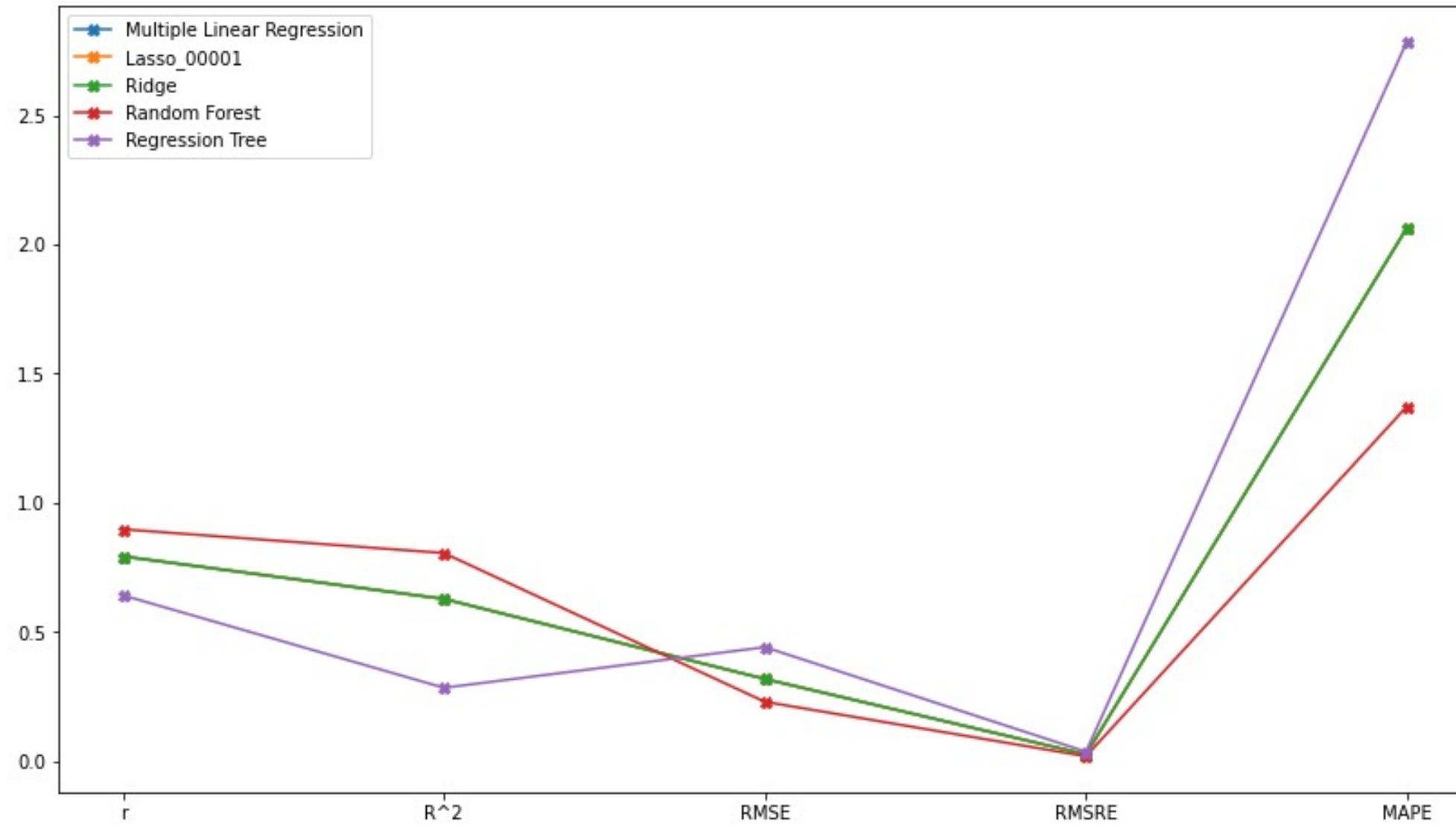
Mean Absolute Percentage Error (MAPE) : 1.3736

Summary Metrics Table of Each Model

	Model	R2	ME	RMSE	MAE	MPE	MAPE
0	MLR	0.6283	0.0007	0.2300	0.1647	-0.0341	1.3736
1	Lasso	0.6283	0.0045	0.3183	0.2474	-0.0365	2.0652
2	Ridge	0.6282	0.0045	0.3183	0.2474	-0.0369	2.0653
3	Regression Tree	0.2886	-0.0059	0.4403	0.3330	-0.1172	2.7772
4	Random Forest	0.8059	0.0007	0.2300	0.1647	-0.0341	1.3736

FIGURE 1: METRICS OF REGRESSION MODELS

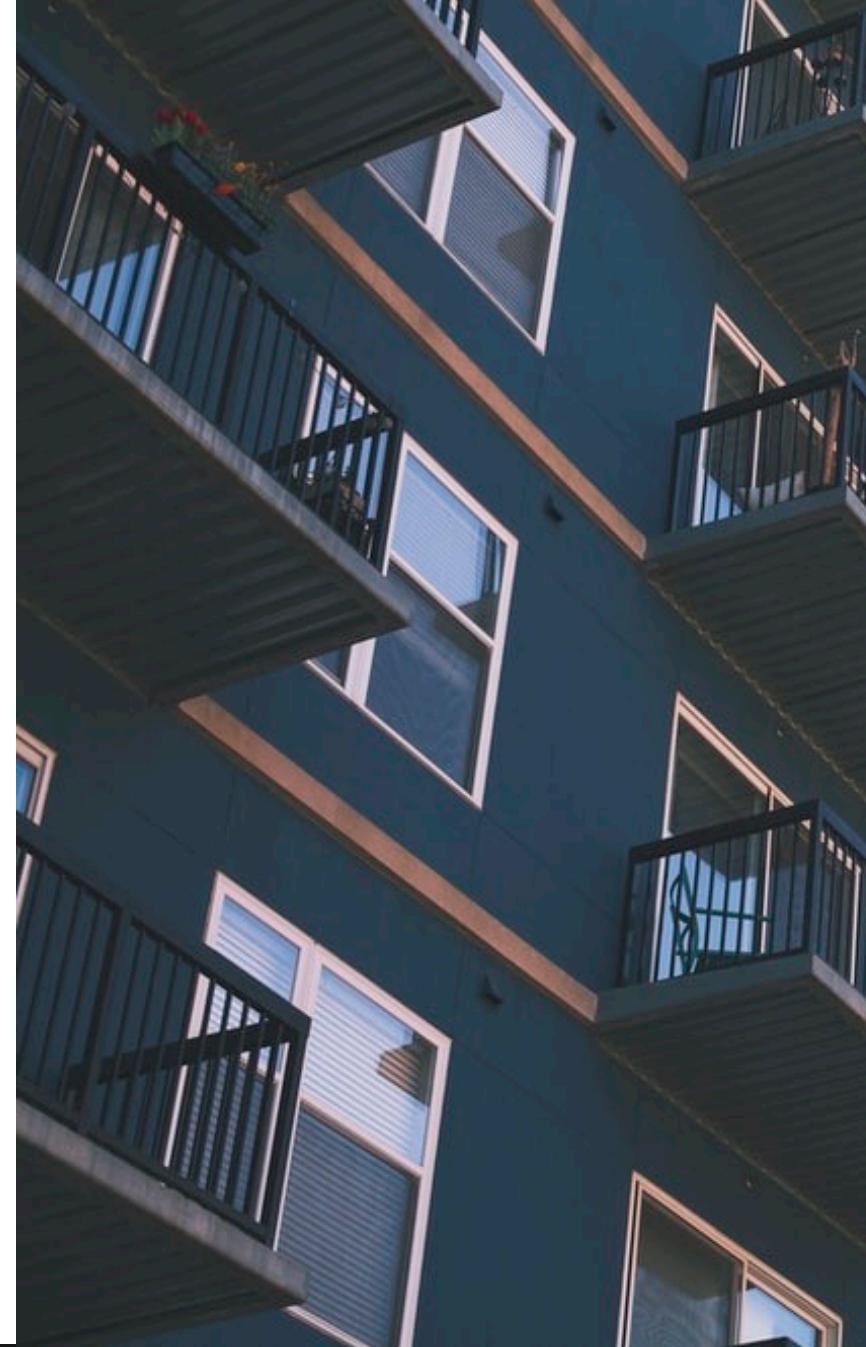




Results for equation MedianPriceSold

Results for equation MedianPriceSold

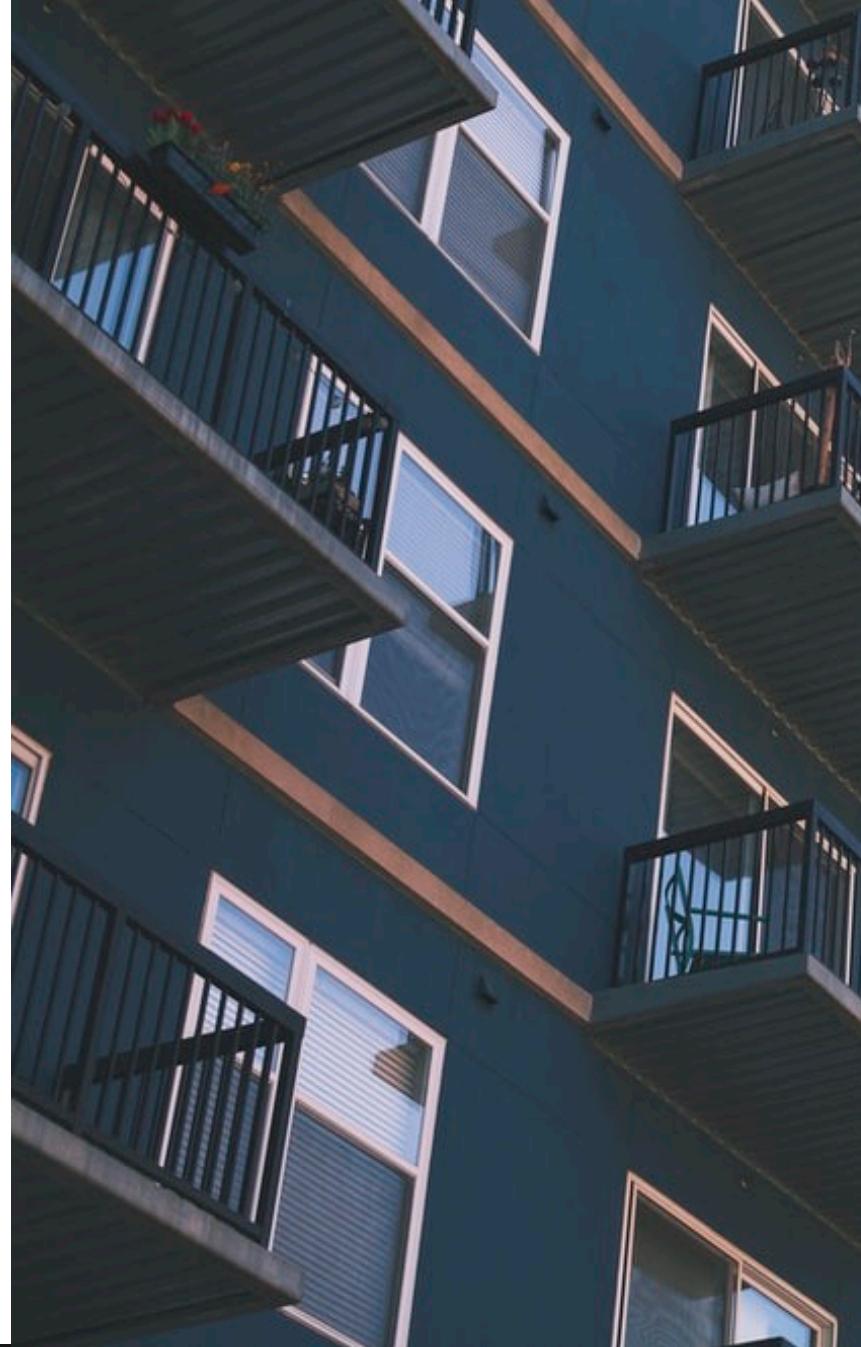
	coefficient	std. error	t-stat	prob
const	0.808931	0.266405	3.036	0.002
L1.MedianPriceSold	-0.807157	0.135795	-5.944	0.000
L1.MedianMortgageRate	0.038633	0.066047	0.585	0.559
L1.UnemploymentRate	0.131373	0.056569	2.322	0.020
L2.MedianPriceSold	-0.521436	0.151130	-3.450	0.001
L2.MedianMortgageRate	0.077491	0.078102	0.992	0.321
L2.UnemploymentRate	0.134453	0.076314	1.762	0.078
L3.MedianPriceSold	-0.214626	0.140026	-1.533	0.125
L3.MedianMortgageRate	0.132654	0.085020	1.560	0.119
L3.UnemploymentRate	0.403773	0.082506	4.894	0.000
L4.MedianPriceSold	-0.024349	0.120712	-0.202	0.840
L4.MedianMortgageRate	-0.062845	0.078180	-0.804	0.421
L4.UnemploymentRate	0.299898	0.097189	3.086	0.002
L5.MedianPriceSold	-0.070050	0.082012	-0.854	0.393
L5.MedianMortgageRate	0.063569	0.067940	0.936	0.349
L5.UnemploymentRate	-0.070137	0.079234	-0.885	0.376



Results for equation MedianMortgageRate

Results for equation MedianMortgageRate

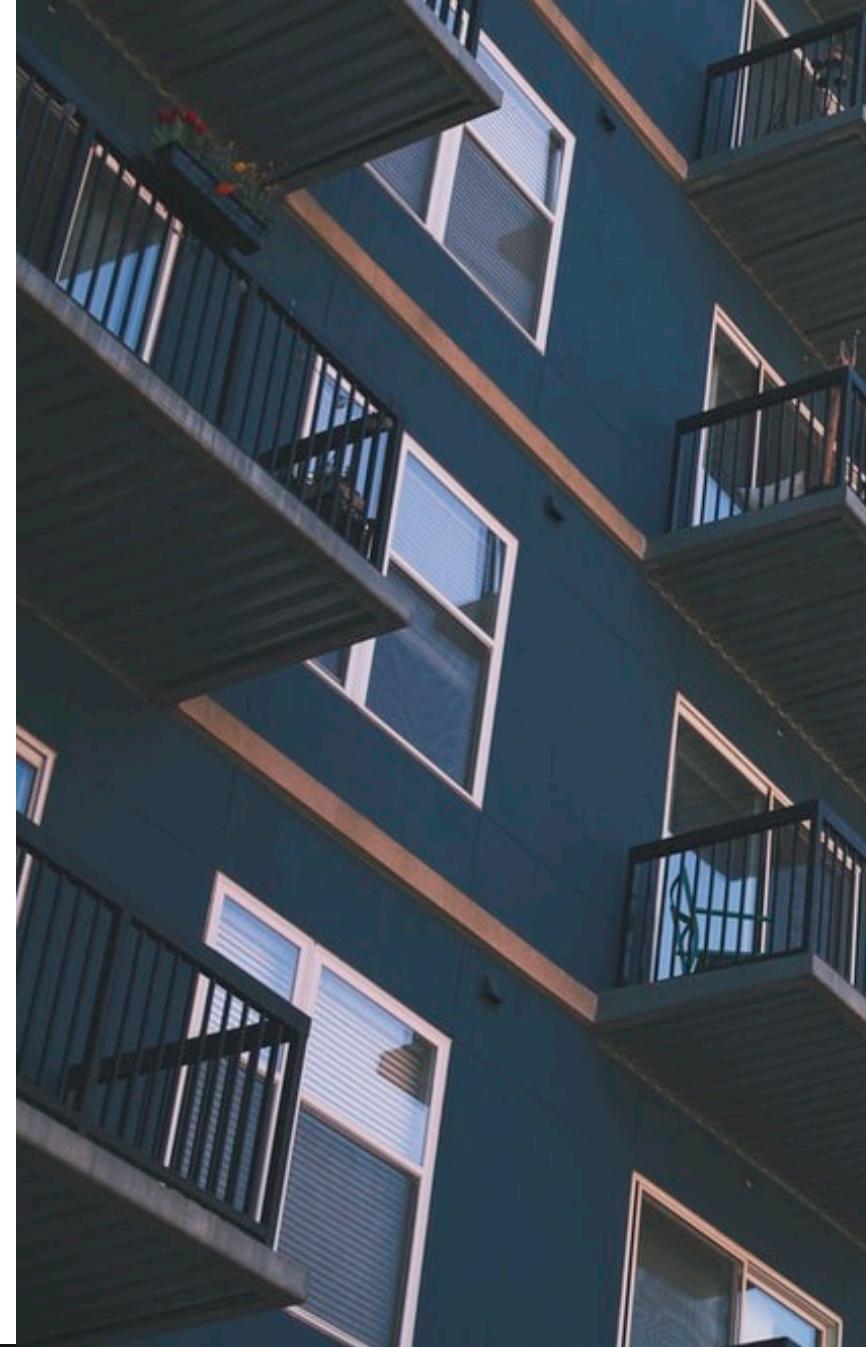
	coefficient	std. error	t-stat	prob
const	0.360321	0.494555	0.729	0.466
L1.MedianPriceSold	-0.163778	0.252091	-0.650	0.516
L1.MedianMortgageRate	-0.700044	0.122609	-5.710	0.000
L1.UnemploymentRate	0.081409	0.105014	0.775	0.438
L2.MedianPriceSold	0.204519	0.280558	0.729	0.466
L2.MedianMortgageRate	-0.636471	0.144989	-4.390	0.000
L2.UnemploymentRate	-0.002091	0.141670	-0.015	0.988
L3.MedianPriceSold	0.302594	0.259945	1.164	0.244
L3.MedianMortgageRate	-0.356814	0.157831	-2.261	0.024
L3.UnemploymentRate	0.140902	0.153165	0.920	0.358
L4.MedianPriceSold	0.507503	0.224089	2.265	0.024
L4.MedianMortgageRate	-0.047946	0.145134	-0.330	0.741
L4.UnemploymentRate	0.229897	0.180422	1.274	0.203
L5.MedianPriceSold	0.383198	0.152247	2.517	0.012
L5.MedianMortgageRate	-0.094017	0.126125	-0.745	0.456
L5.UnemploymentRate	0.189584	0.147091	1.289	0.197



Results for equation UnemploymentRate

Results for equation UnemploymentRate

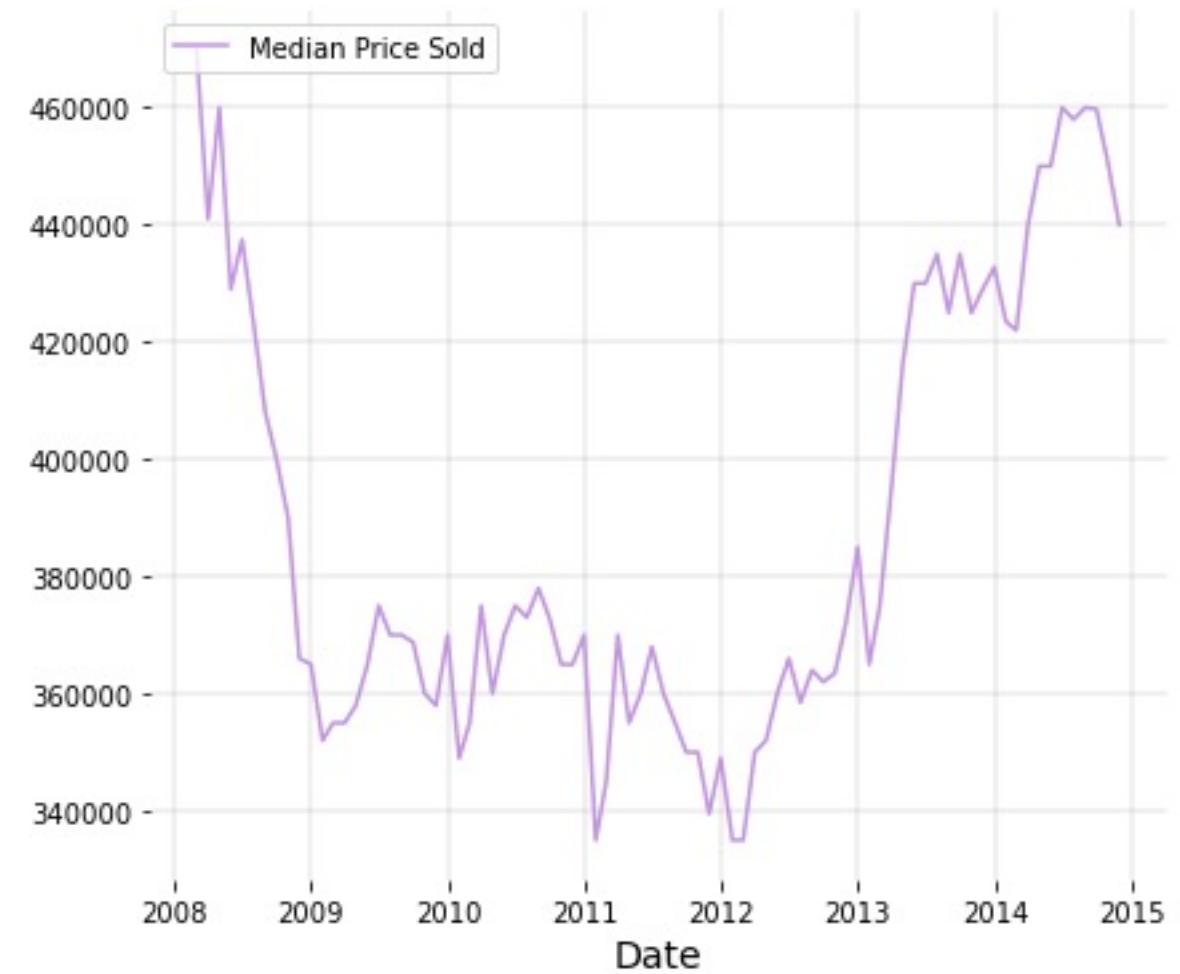
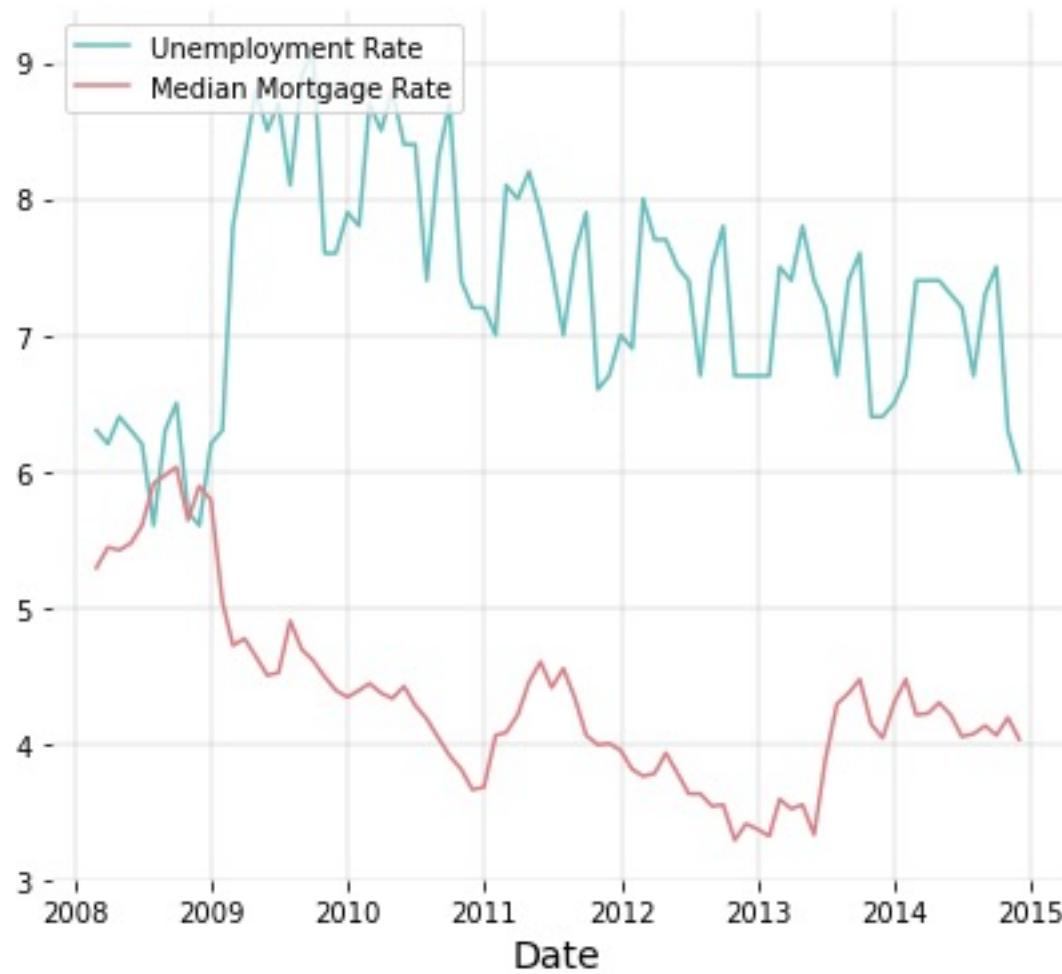
	coefficient	std. error	t-stat	prob
const	2.789087	0.542785	5.138	0.000
L1.MedianPriceSold	0.322277	0.276676	1.165	0.244
L1.MedianMortgageRate	0.077964	0.134567	0.579	0.562
L1.UnemploymentRate	-0.956789	0.115255	-8.301	0.000
L2.MedianPriceSold	0.306579	0.307919	0.996	0.319
L2.MedianMortgageRate	0.179395	0.159129	1.127	0.260
L2.UnemploymentRate	-0.974701	0.155486	-6.269	0.000
L3.MedianPriceSold	-0.244309	0.285296	-0.856	0.392
L3.MedianMortgageRate	-0.137769	0.173223	-0.795	0.426
L3.UnemploymentRate	-0.763859	0.168102	-4.544	0.000
L4.MedianPriceSold	-0.540047	0.245943	-2.196	0.028
L4.MedianMortgageRate	-0.219592	0.159288	-1.379	0.168
L4.UnemploymentRate	-0.800094	0.198017	-4.041	0.000
L5.MedianPriceSold	-0.509329	0.167095	-3.048	0.002
L5.MedianMortgageRate	-0.037279	0.138425	-0.269	0.788
L5.UnemploymentRate	-0.563706	0.161436	-3.492	0.000

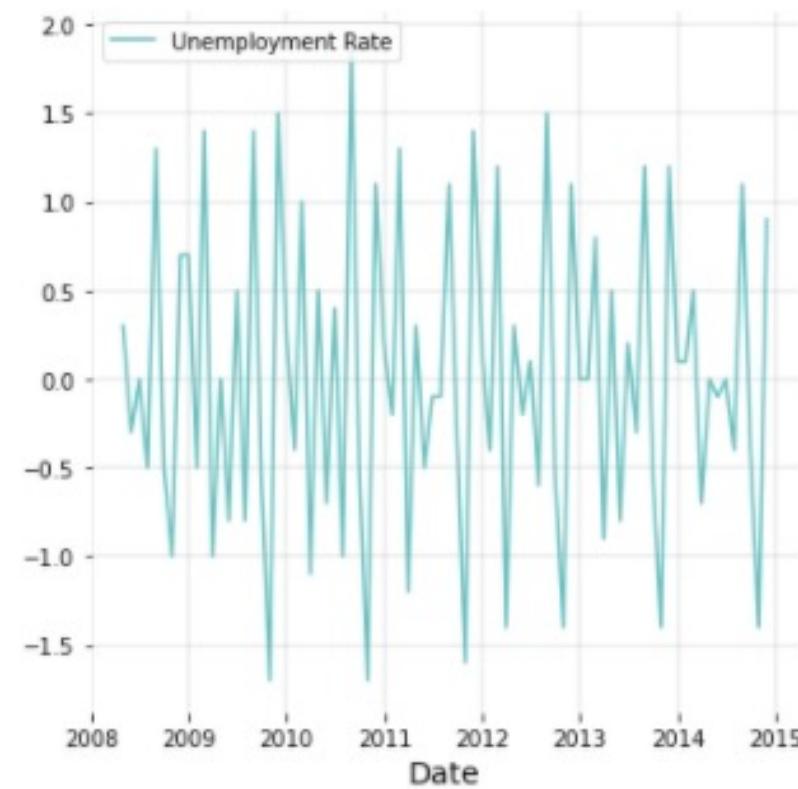


Correlation matrix of residuals

	MedianPriceSold	MedianMortgageRate	UnemploymentRate
MedianPriceSold	1.000000	- 0.072780	0.289204
MedianMortgageRate	-0.072780	1.000000	0.250494
UnemploymentRate	0.289204	0.250494	1.000000

Zillow Housing Training Data





Fine Tune

- Careful inspection of null values
- Ensemble learning outperforms a single model
- Minimizing the categorical attribute levels



RESULT



**RANDOM FOREST
PERFORMED THE BEST**

Random Forest scored the highest R2 value of 80.59% best fitting the data set. Thus, Random Forest deemed the best model for the prediction of California house prices.



Investment in California

A photograph of a row of colorful, multi-story Victorian-style houses in San Francisco. The houses are built close together, featuring intricate architectural details like decorative trim, multiple gables, and arched windows. The colors of the houses vary, including shades of blue, yellow, and red. In the foreground, there are some green trees and bushes. The sky is clear and blue.

California Housing Market Remains Solid

September 2021

438,190 | **Existing Home Sales**
-10.5% YTY **+16.8% YTD % change**



Price

\$808,890
+13.5% Y2Y



**Unsold
Inventory Index**

1.9 months
-5.0% Y2Y



**Median Days
on Market**

10 days
-9.1% Y2Y



September 2021 Southern California Housing

-8.3% YTY
+18.6% YTD

\$755,000
+15.0% YTY



**Unsold
Inventory Index**

1.9 months

Existing Home Sales
% change

Existing Home Price
% change



**Median Days
on Market**

10 days



**Sales Price to
List Price Ratio**

101.7%



September 2021 Central California Housing

-6.2% YTY

+7.5% YTD

Existing Home Sales

% change

\$450,000

+15.4% YTY

Existing Home Price

% change



**Unsold
Inventory Index**

1.9 months



**Median Days
on Market**

8 days



**Sales Price to
List Price Ratio**

101.7%



September 2021 Bay Area California Housing

-7.0% YTY
+26.5% YTD

\$1,290,000
+21.7% YTY



**Unsold
Inventory Index**

1.6 months

Existing Home Sales
% change

Existing Home Price
% change



**Median Days
on Market**

11 days



**Sales Price to
List Price Ratio**

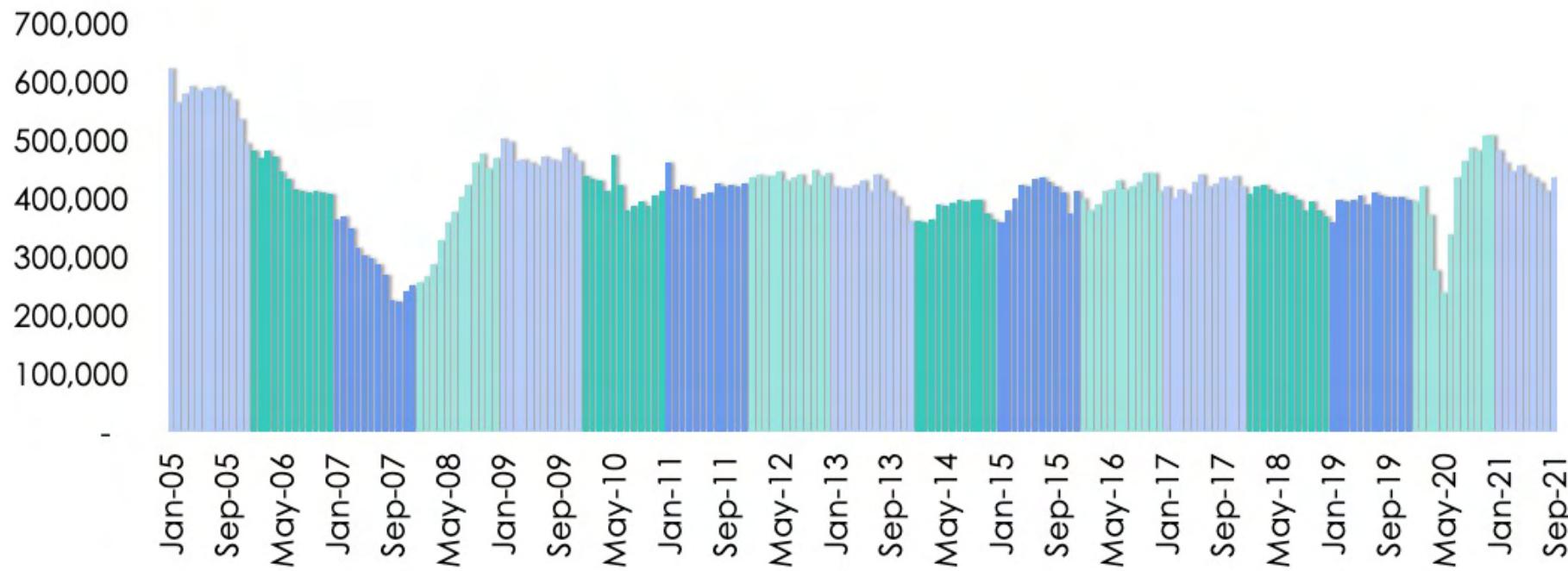
105.4%

A photograph of a row of colorful Victorian-style houses in San Francisco, likely Alamo Square. The houses are multi-story with intricate architectural details, gables, and decorative trim. They are painted in various colors, including blues, yellows, and reds. The scene is set against a clear blue sky.

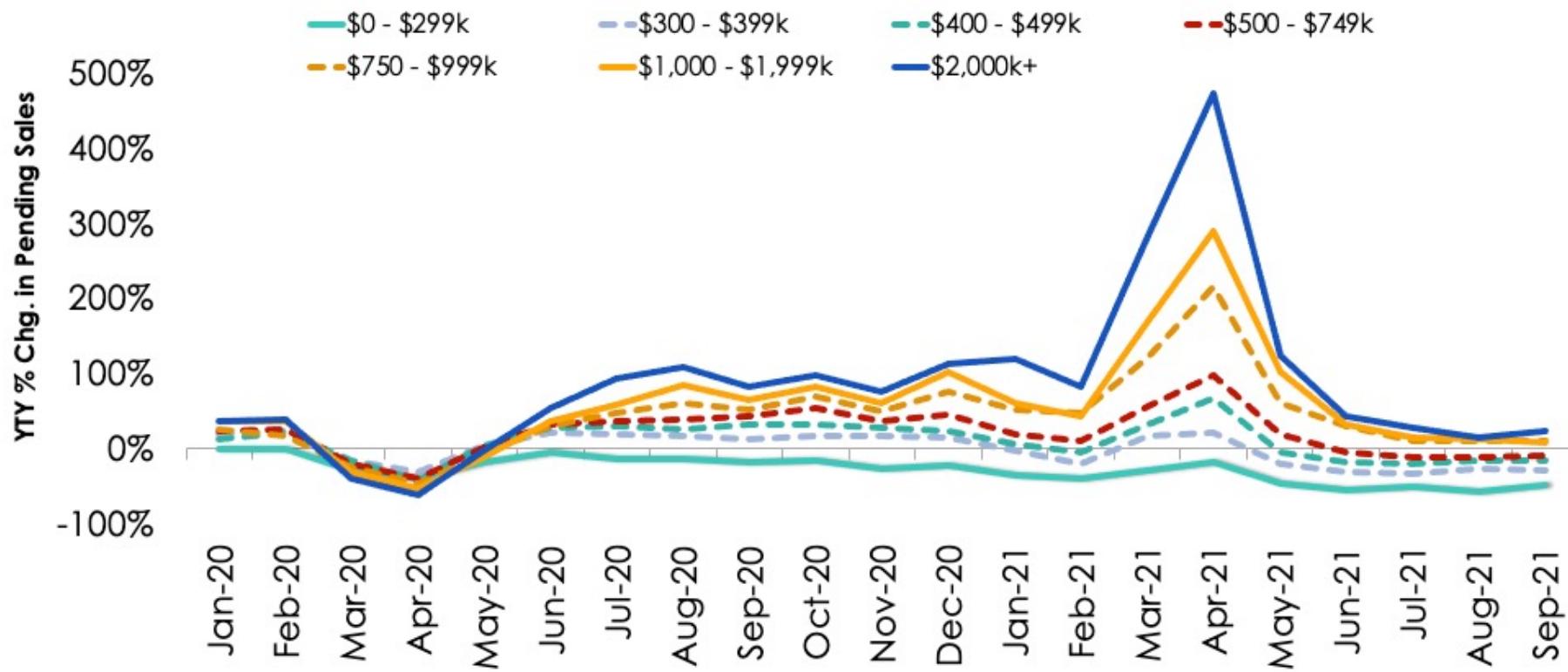
California Housing Market Sales

Sales bounced back after 4 months of decline

California, September 2021 Sales: 438,190 Units, +16.8% YTD, -10.5% YTY, +5.6% MTM



Pending sales at the top end grew modestly

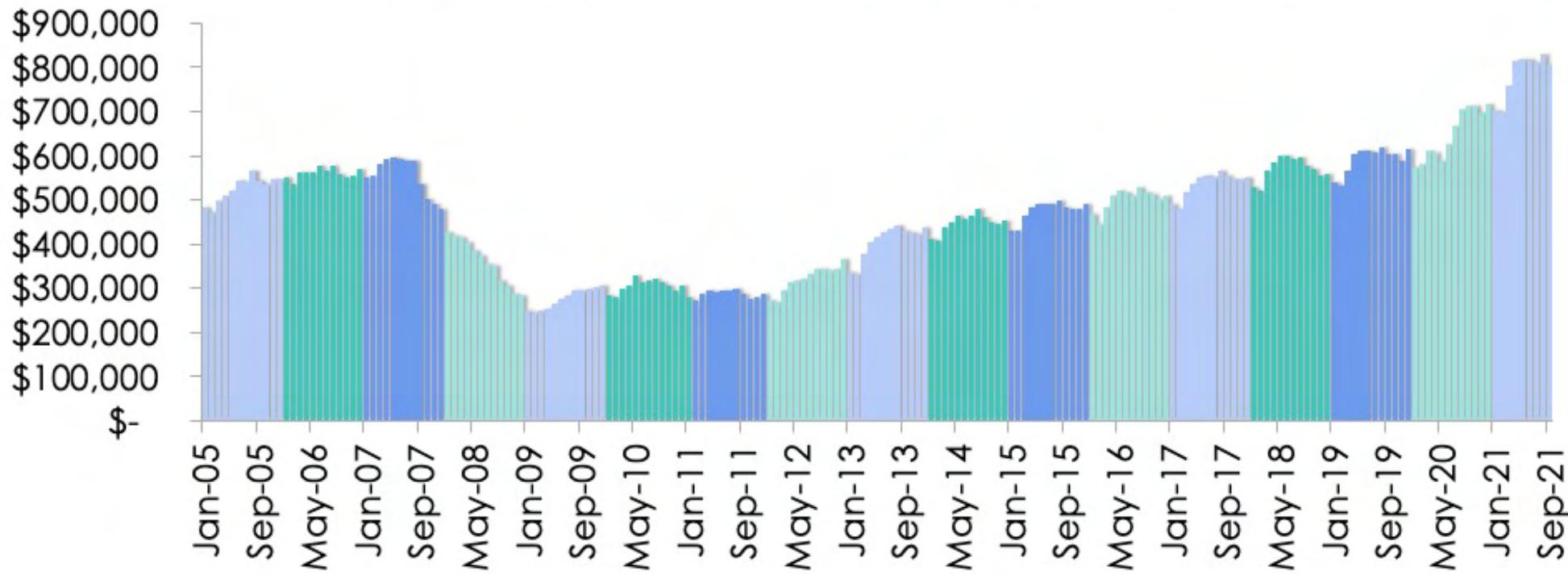


A photograph of a row of colorful Victorian-style houses in San Francisco, likely Alamo Square. The houses are multi-story with intricate architectural details, gables, and decorative trim. They are painted in various colors, including blues, yellows, and reds. The scene is set against a clear blue sky.

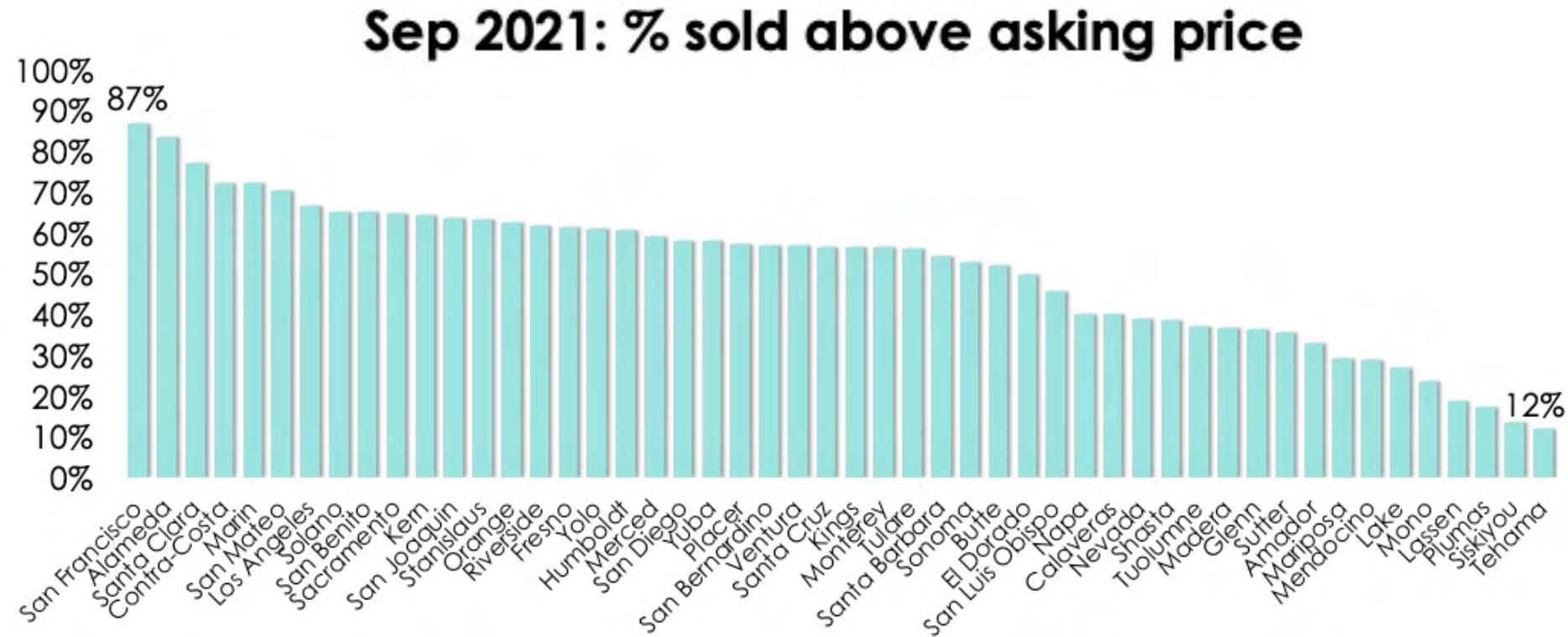
California Housing Market Price

California median home price retreated after hitting a new peak

California, August 2021: \$808,890, -2.3% MTM, +13.5% YTY



Homes sold above asking price - by county



A dark, semi-transparent overlay covers the background image of a row of colorful Victorian-style houses in San Francisco. The houses are built close together, featuring intricate architectural details like decorative trim, multiple gables, and varied rooflines. Some have porches and flower boxes. The colors range from light blues and yellows to darker tones like red and brown. In the foreground, there are some green trees and bushes.

Should Investors or Home buyer Invest in California Housing Market?



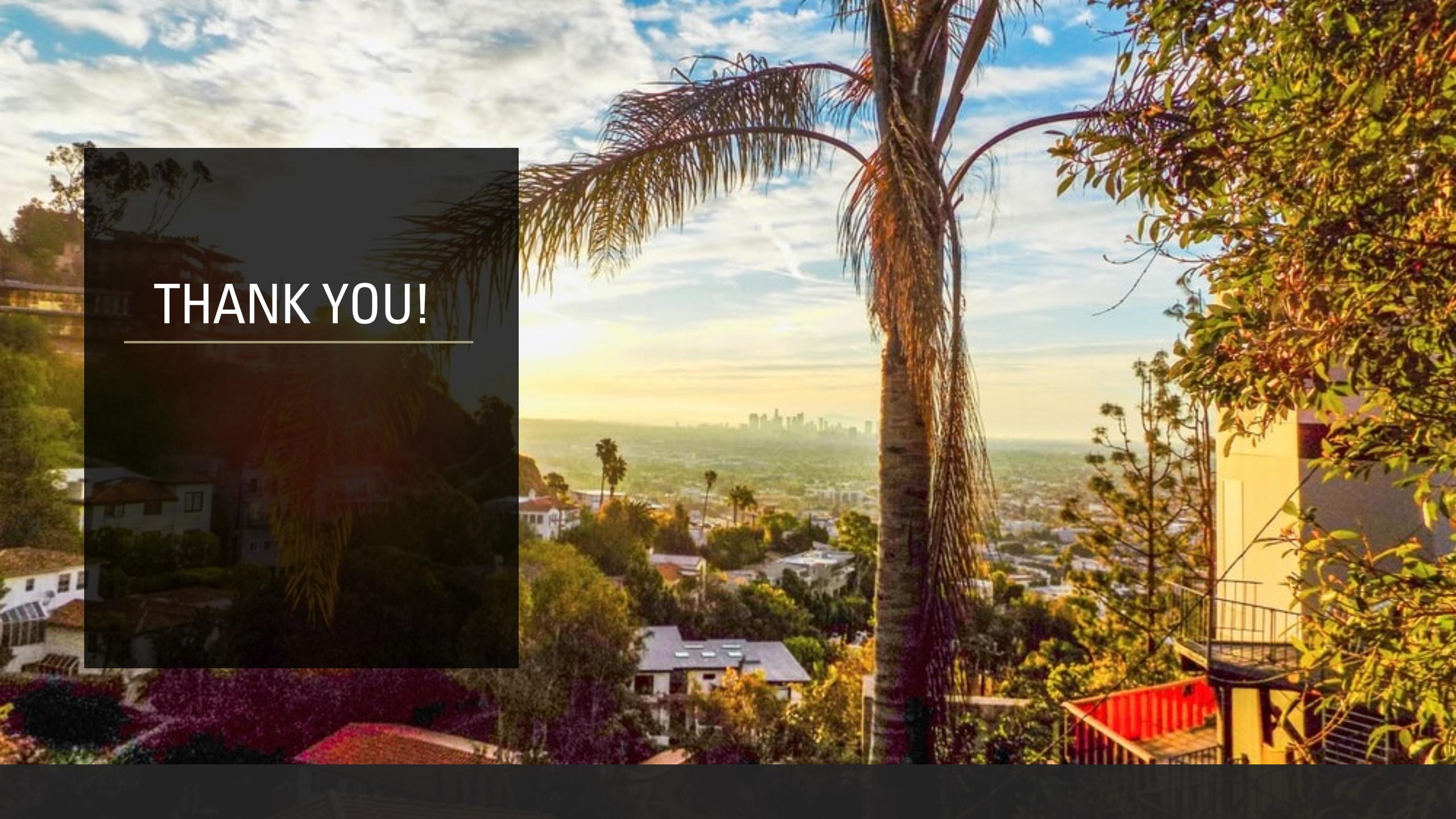
The Answer is ...





Save your money!

The data shows that the housing market in California is way too high and that your mortgage and interest rate will be too high to manage, and you'll end up filling for bankruptcy. For those seeking to invest stay far way from California as possible.



THANK YOU!



References
