


# Housing You Too Ltd Time Series

- GROUP MEMBERS;
- WILLIAM ITOTIA
- FRANK OYUGI
- ESTHER TERRY  
MUNENE
- FIRDOSA MOHAMMED




# Overview

The real estate market in the United States offers a wide range of housing options but there's little knowledge on how to value them to guide buyers and sellers.



It is therefore a huge advantage to know which properties will provide a substantial Return On Investment (ROI) and which properties will most likely bring losses before making a decision on the property to buy.



By using technical machine learning techniques, time series models can be built and analysed using historical information and thereafter, forecasting can be possible.

# Problem Statement

Housing You Too Ltd, a consultancy and real estate firm, has been approached by a client eager to invest in properties with the goal of achieving the highest possible Return on Investment (ROI). The client is looking for our expert guidance to identify the top 5 zip codes in California that promise the best ROI. To achieve this, we will utilize the extensive Zillow dataset, which contains historical real estate data.





# Objectives

1. Identify the 5 best zipcodes in California that will bring our clients the highest ROI.
2. Build a time series regression model to identify the 5 most profitable zipcodes.
3. Using the dataset and model created investigate the growth rate of the real estate industry in relation to sales value of properties.
4. Identify the 5 worst zipcodes in California to avoid in order to prevent incurring losses.



# Challenges

- 1.Data Ambiguity and Unpredictability: Since we are using real-world data, there is a possibility of ambiguity and unpredictability, which can lead to some inaccuracies in our analysis and predictions.
- 2.Changing Market Conditions: Real estate markets are influenced by various external factors such as economic shifts, policy changes, and unforeseen events like natural disasters or pandemics. These factors can alter market conditions rapidly, making historical data less reliable for future predictions.



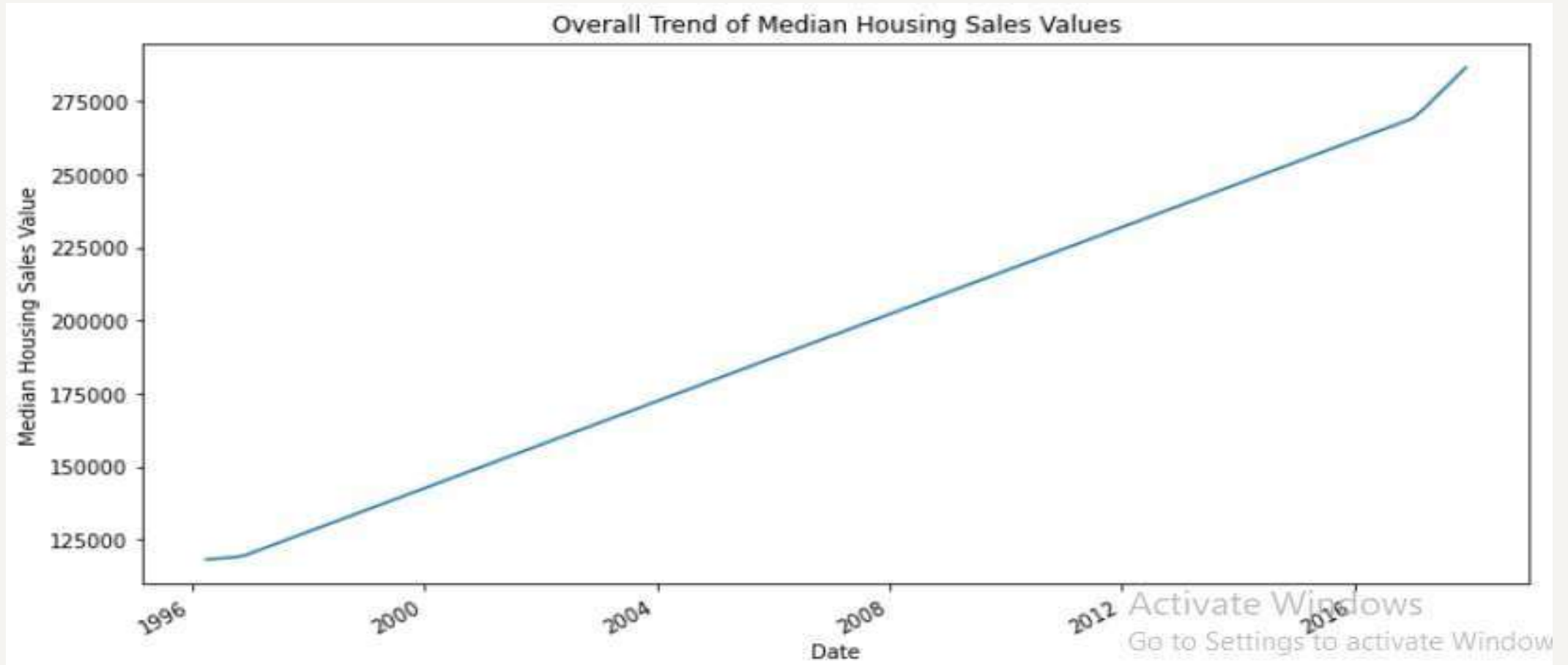


# Solutions

1. We will implement robust data cleaning and preprocessing techniques to handle missing values, outliers, and inconsistencies. We will also use advanced statistical methods and machine learning algorithms that can manage noisy data and enhance prediction accuracy.
2. Integrate real-time data sources and economic indicators into the analysis to account for current market conditions. Use adaptive models that can update predictions based on the latest data.



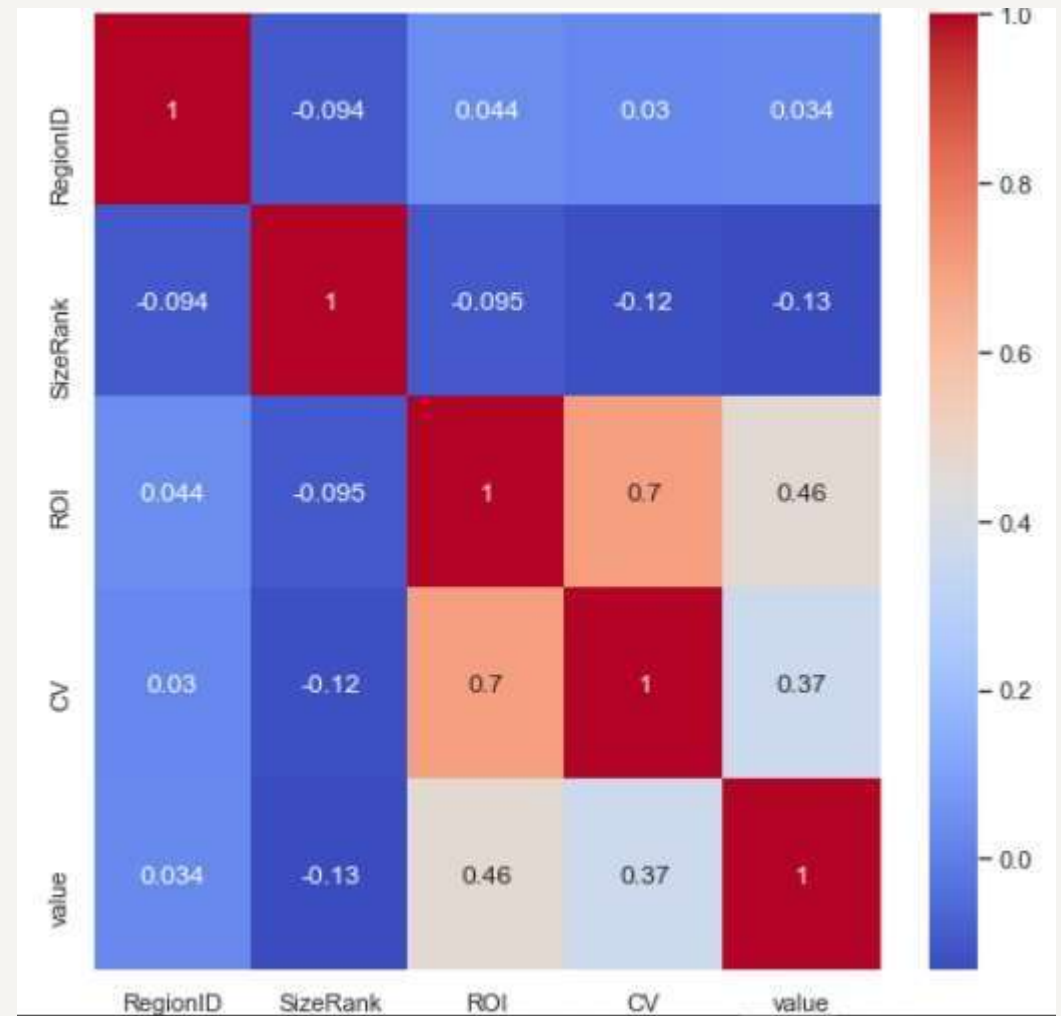
# Trend Of Housing Sale Values



As we can see the trend is an increasing linear trend, where we can see the median housing sales value increases every year.

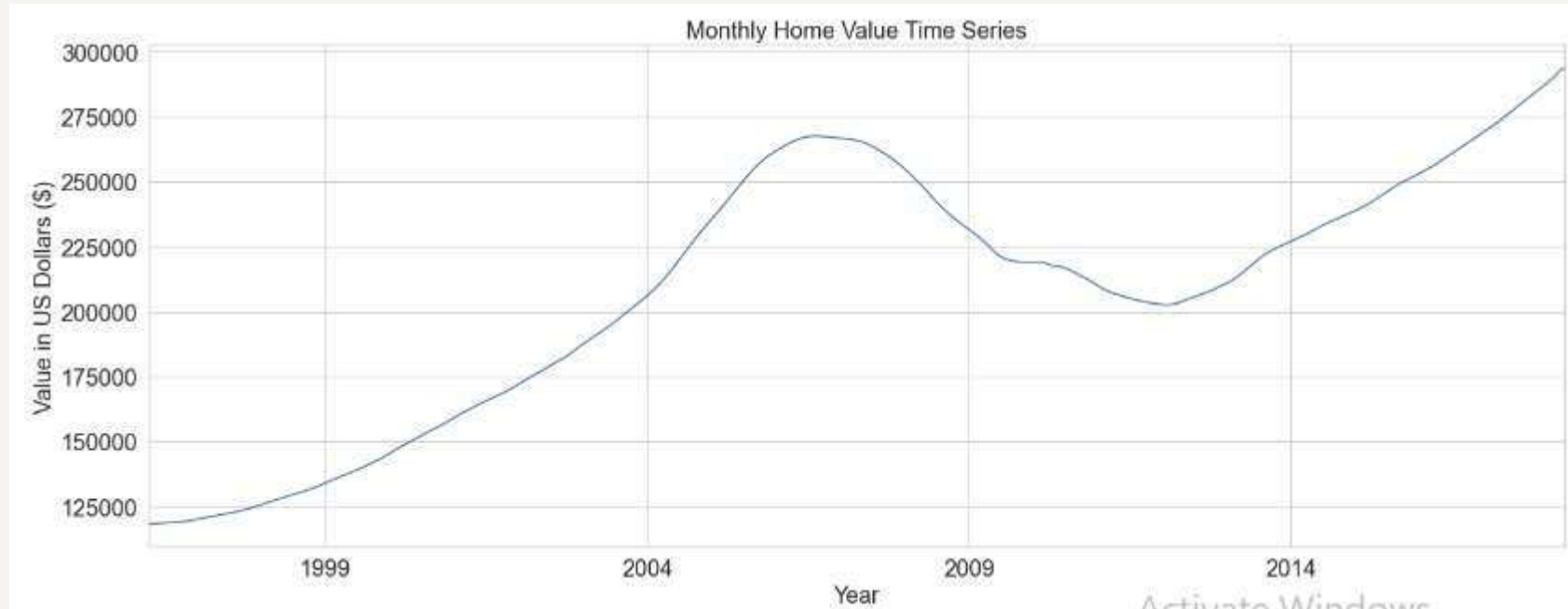
# Relationship Between Variables

This heatmap shows that most of the features exhibit weak relationships with each other, except for ROI and CV, which display a strong relationship.



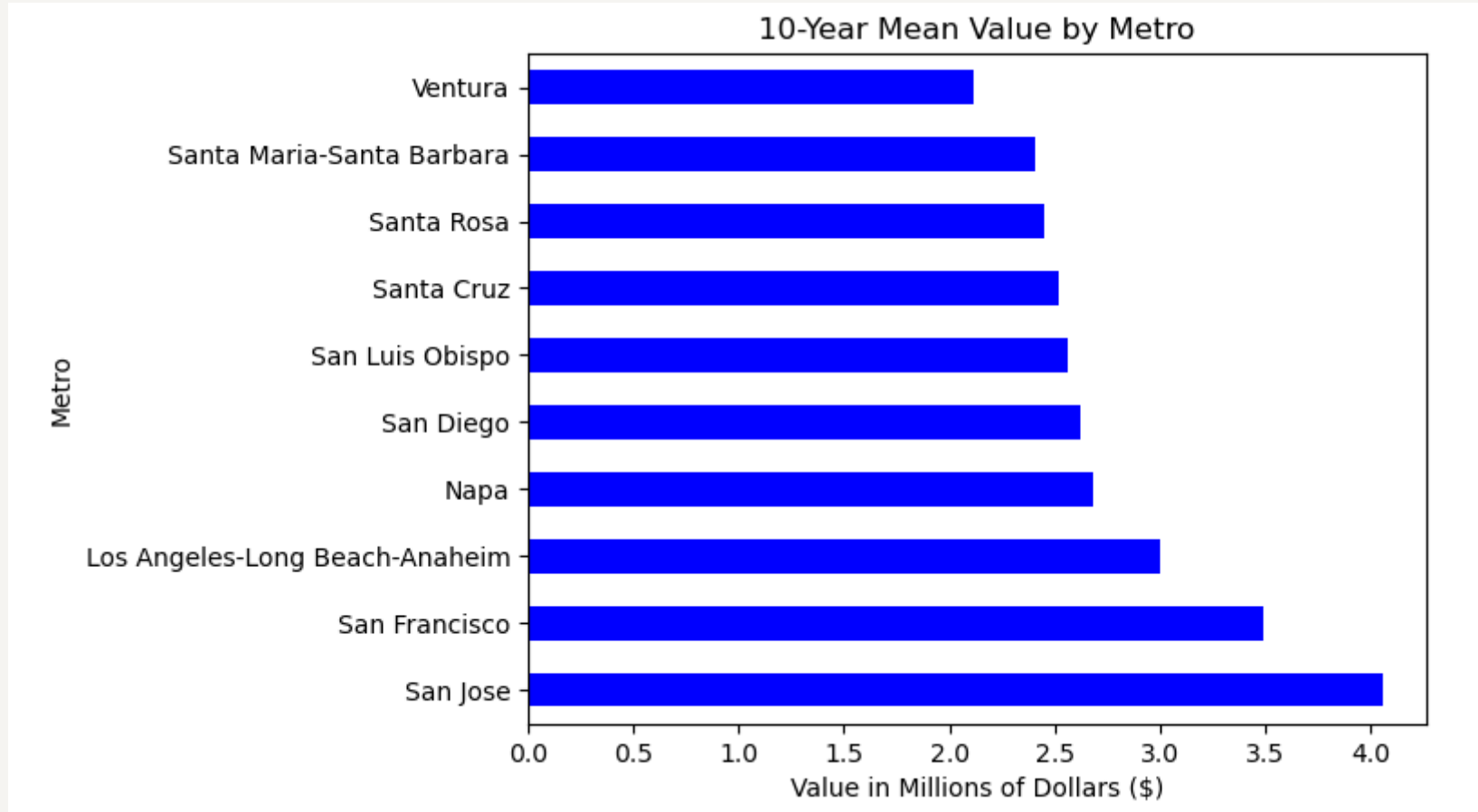


# Monthly Home Value Time Series



The plot of the housing prices indicates an overall upward trend from 1996 to around 2007, followed by a downward trend until approximately 2013, and then an upward trend again.

The top 10 California metros with highest average sales values.



# Conclusion

The top 5 zipcodes to invest in are:  
93241,93505,93588,95341 and  
95485.

The bottom 5 zipcodes to invest in  
are: :91008,95436,91916,91101  
and 91204.





# Recommendations

1. Invest in these 5 zipcodes: 93241, 93505, 93588, 95341 and 95485.
2. Do not invest in these 5 zipcodes: 91008, 95436, 91916, 91101 and 91204.
3. Increased use of digital marketing strategy to promote sales in future. The sales team should also consider discounts in future so as to entice customers thereby increasing sales.
4. Since there is probability of decline in sales in the future, the company should avoid overstocking.

