



Zillow Housing Data Analysis for Horizon Real Estate

1 Business Understanding

2 Data Understanding

3 Modelling

4 Recommendations



Business Understanding

There are 3 main important things in real estate;

1 Location.

2 Condition.

3 Price.



Problem Statement

Real Estate can be a very risky business to new investors but lucrative when the risks are understood. The risks can be bad location, high prices and unpredictable nature of the Market.



Overview

Horizon Real Estate is looking to penetrate the market and build residential houses in the United States. There is need to identify the top five locations suitable for residential houses and have a high return on investment.



Business Objective

- Identifying Zip codes to invest in based on return on investment
- Identifying the volatility of home values in different areas
- Identifying the influence of different factors on home values
- Predict home values to decide on investment



Data Understanding

The Data was sourced from Zillow Housing Data.

The Data contains;

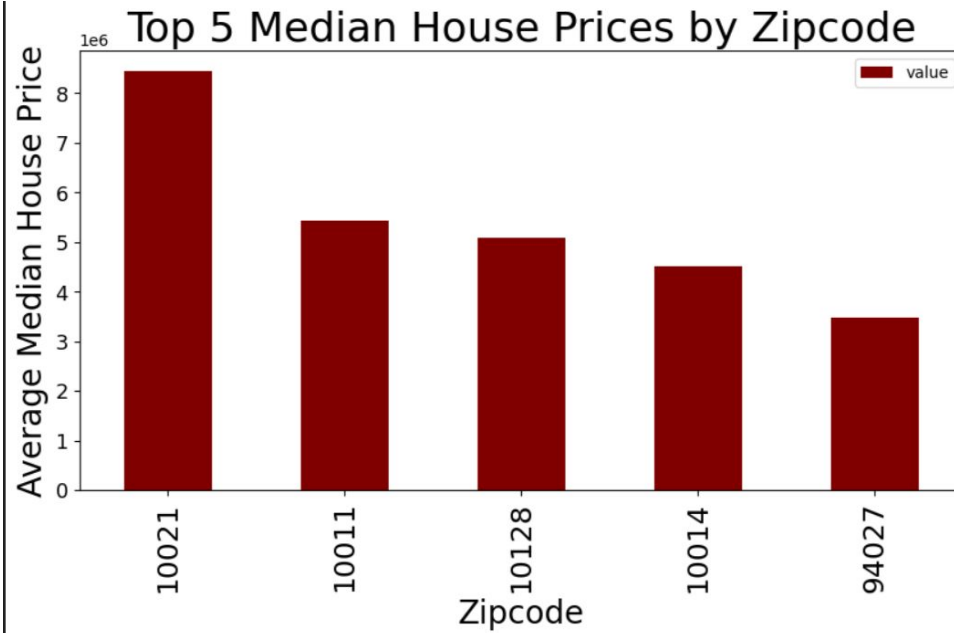
- Median house prices.
- Regions where the houses are located.
- City names and metropolitan areas where the houses are located.



Column	Description
RegionID	Unique region identifier
Region Name	Names of the region(Zip codes)
City	City names for the Region
State	Names of the states
Metro	Names of metropolitan areas
County Name	Names of counties
Size Rank	Rank of Zip codes by urbanization
Date Columns (265 Columns)	Median house prices across the years

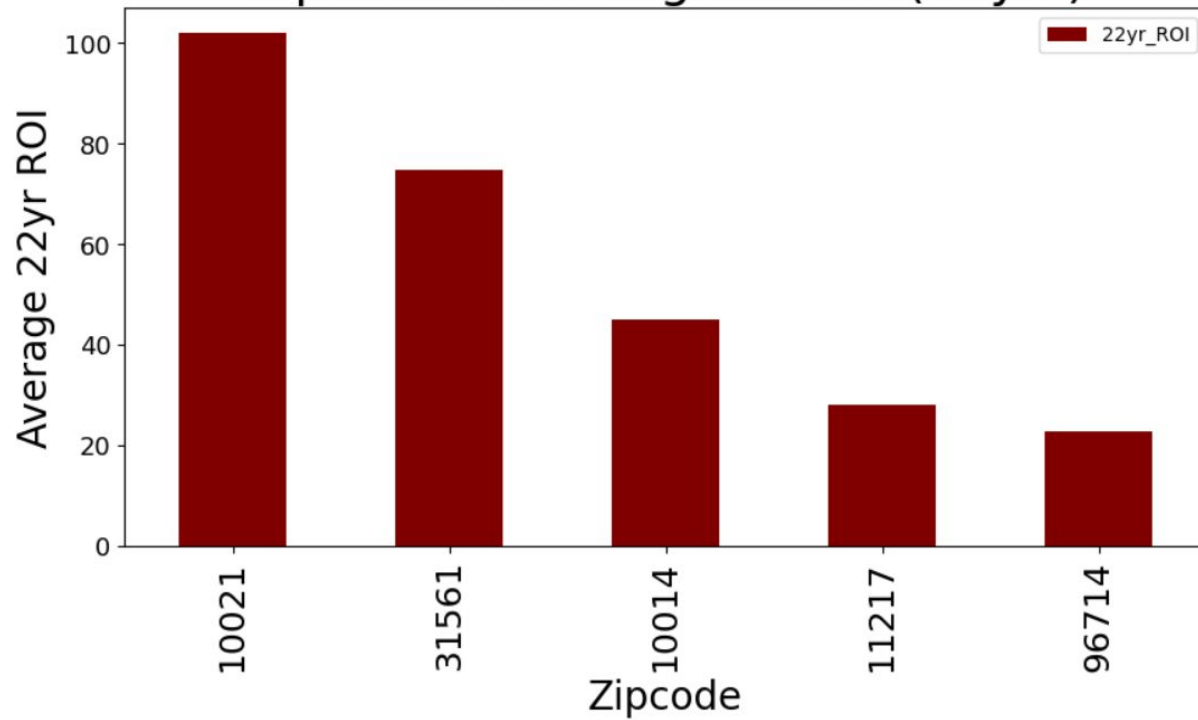


Data Analysis



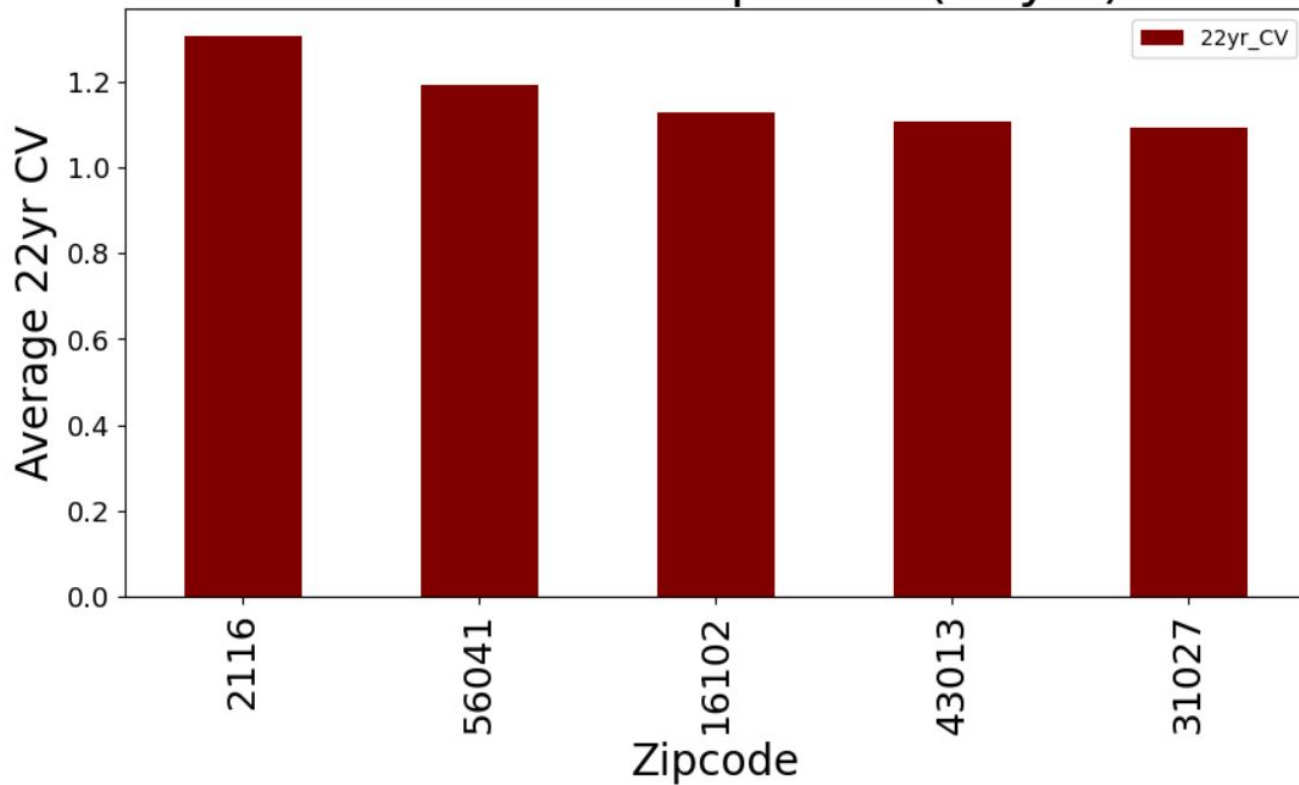
Atherton (California) has the highest house prices followed by Palm Beach (Florida) and Snowmass Village (Colorado)

Zipcodes with highest ROI(22yrs)



The zipcodes with the highest ROI in 22 years are: * Upper East Side- New York(10021)

Most Volatile Zipcodes(22yrs)



The zipcodes with the highest price volatility in 22 years is: * 56041 New UlmMinnesota

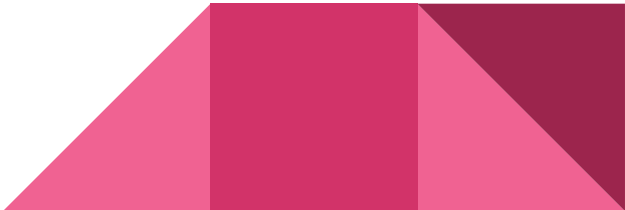
Modelling

Modelling Comparison


Columns	Akaike information criteria	Bayesian Information Criteria
Moving Average	5246.84	5257.29
Auto Regressive	4152.85	4163.31
ARMA	3941.84	3955.78
ARIMA	3578.14	3588.58
SARIMA	3539.1	3546.05
SARIMA (Tuned)	2838.21	2880.28

Model Evaluation

Forecast Technique	Mean Absolute Percentage Error
Dynamic Forecast	4.24
Non-Dynamic Forecast	0.02



Conclusion

- The prices of houses also differ with the city ranks. Homes in highly-ranked cities have high median prices.
 - The top three cities with the highest median house prices are Atherton (California), Palm Beach(Florida), and Snowmass Village (Colorado).
 - The cities with the highest ROI in 22 years are New York, Georgia state and Hawaii.
 - Using the model, it can also deduce that the growth of home value is bound to slow down considerably in the coming years
- 

Recommendations

- The company should focus on locations with great beaches and outdoor activities since houses in these locations fetch highest prices as this is seen from cities in California and Florida.
- Invest in the cities, New York and California, which has the highest Return On Investment.
- Sell at a discounted price to increase sales over the next few years due to the slowing down of real estate price in the foreseeable future.

