Zillow Housing Data Analysis for Horizon Real Estate

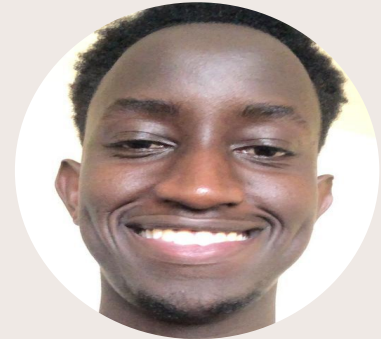By Databolical

Members

Bonface Mutua

Arnold Kalage

Calvince Ochieng

Getrude Obwoge

Susan Mungai

James Mponzi

01 Business Understanding

02 Data Understanding

03 Modeling

04 Recommendations

01

Business Understanding

"There are 3 main things important in real estate : location,location and location."

—Successful Real Estate Agent
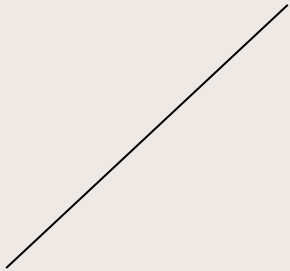
Problem Statement

Real Estate can be a very risky business to new investors but lucrative when the risks are understood. The risks can be bad locations, the unpredictable nature of the market, et cetera.

Overview

Horizon Real Estate, is looking to penetrate the market and build residential houses in the United States. We want to identify the top five locations suitable for residential houses and have a high return on the company's investment.

Business Objectives

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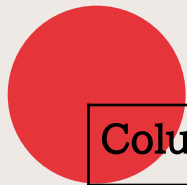- Predict home values to decide on investment

Data Understanding

02

Data Description

- The data was sourced from <u>Zillow Housing Data</u> and published on November 2019.
- The data contains:
- The median house prices through 22 years, i.e., 1996-2018.
- The regions where the houses are located.
- The city names and the metropolitan areas where these houses are located.

Column	Description
RegionID	Unique region identifier
RegionName	Names of the Regions (Zip codes)
City	City names for the regions
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

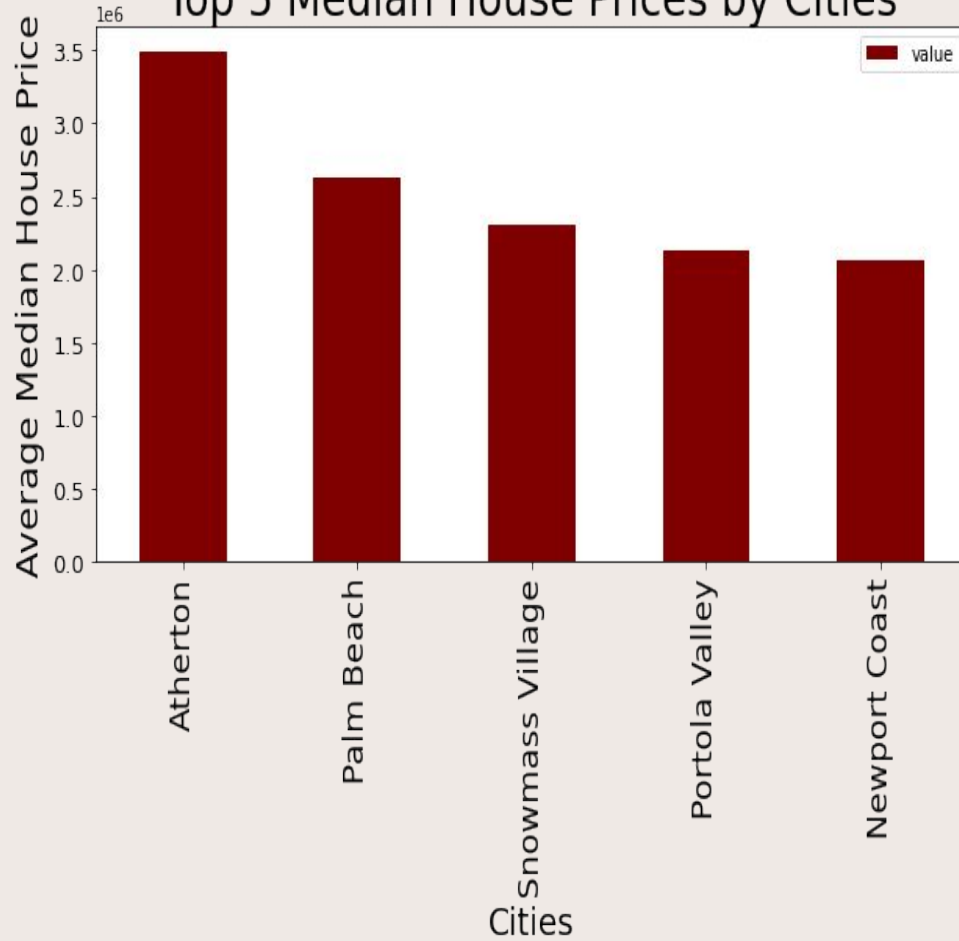# Trend of the Data

Home Value by Year
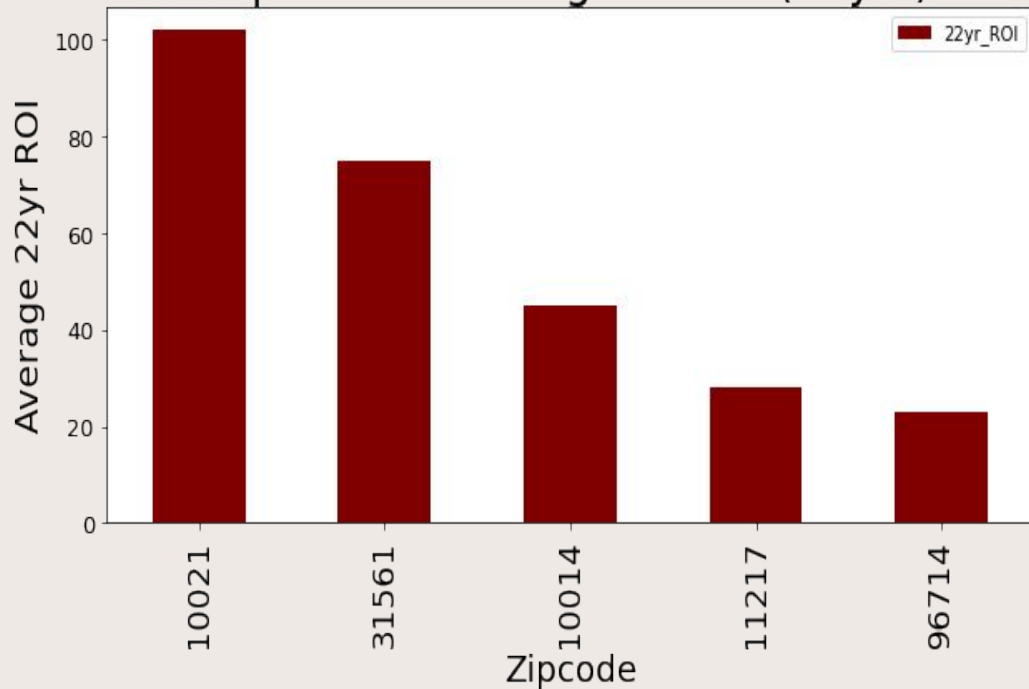
© 2022
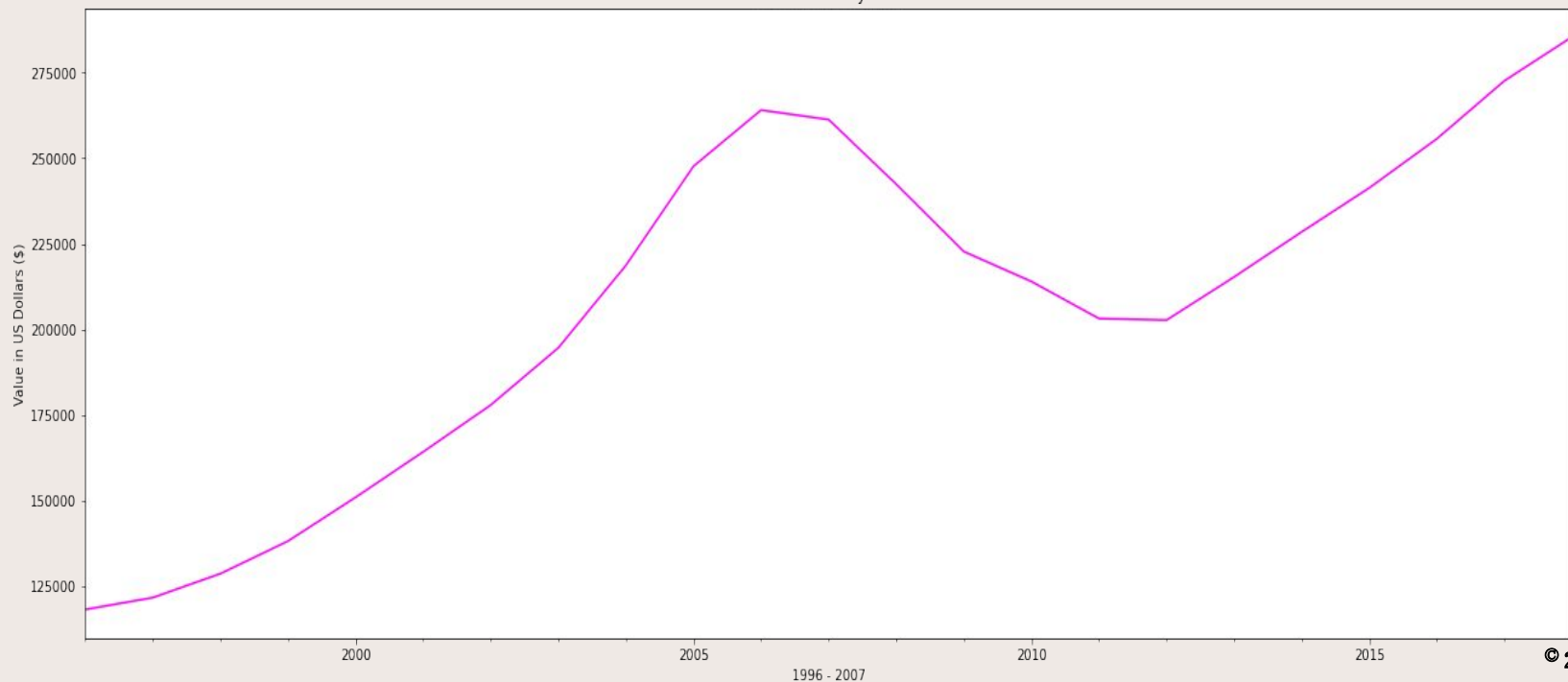
Data Analysis

03

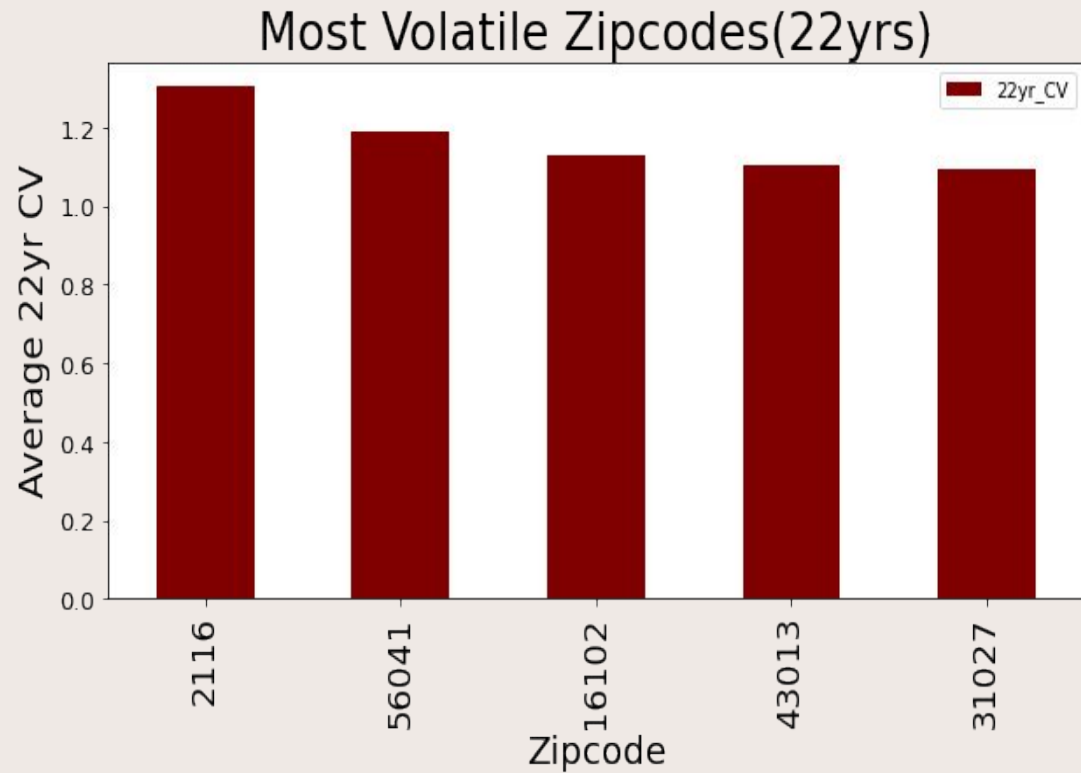Top 5 Median House Prices by Cities

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)

Average 22yr CV

22yr_CV

Zipcode: 2116, 56041, 16102, 43013, 31027

The zipcodes with the highest price volatility in 22 years is:

* 56041 New Ulm-Minnesota

Modelling

04

Model Comparison

Columns	AIC	BIC
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

- **AIC**- Akaike Information Criteria
- **BIC**-Bayesian Information Criteria

Model Evaluation

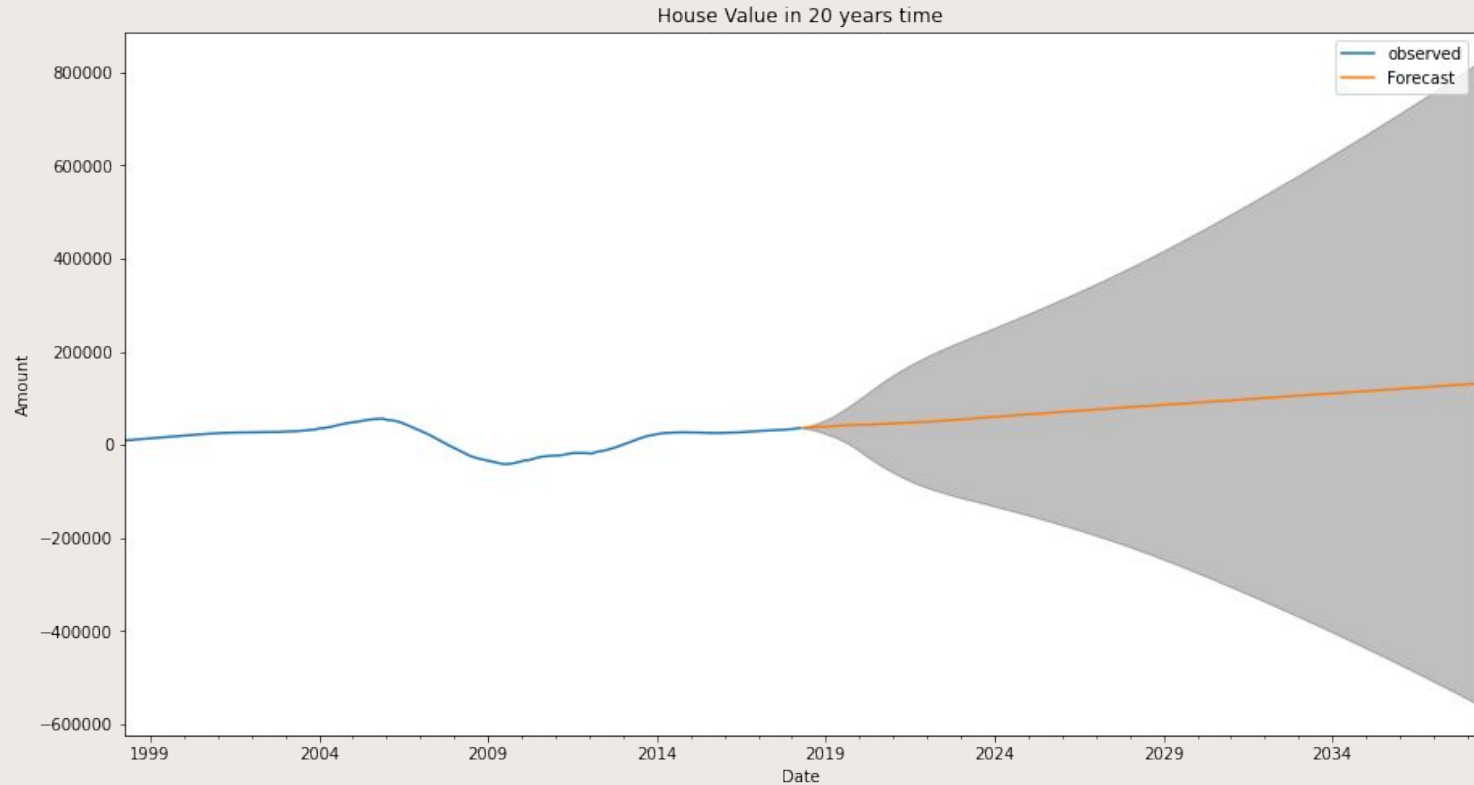| Forecast Technique	MAPE
Dynamic Forecast	4.24
Non-Dynamic Forecast	0.02

A MAPE less than 5% is considered as an indication that the forecast is acceptably accurate. The lower the MAPE, the better.

- **MAPE –** Mean Absolute Percentage Error

5 Year Forecast

House Value in 5 Years Time

20 Year Forecast

House Value in 20 years time
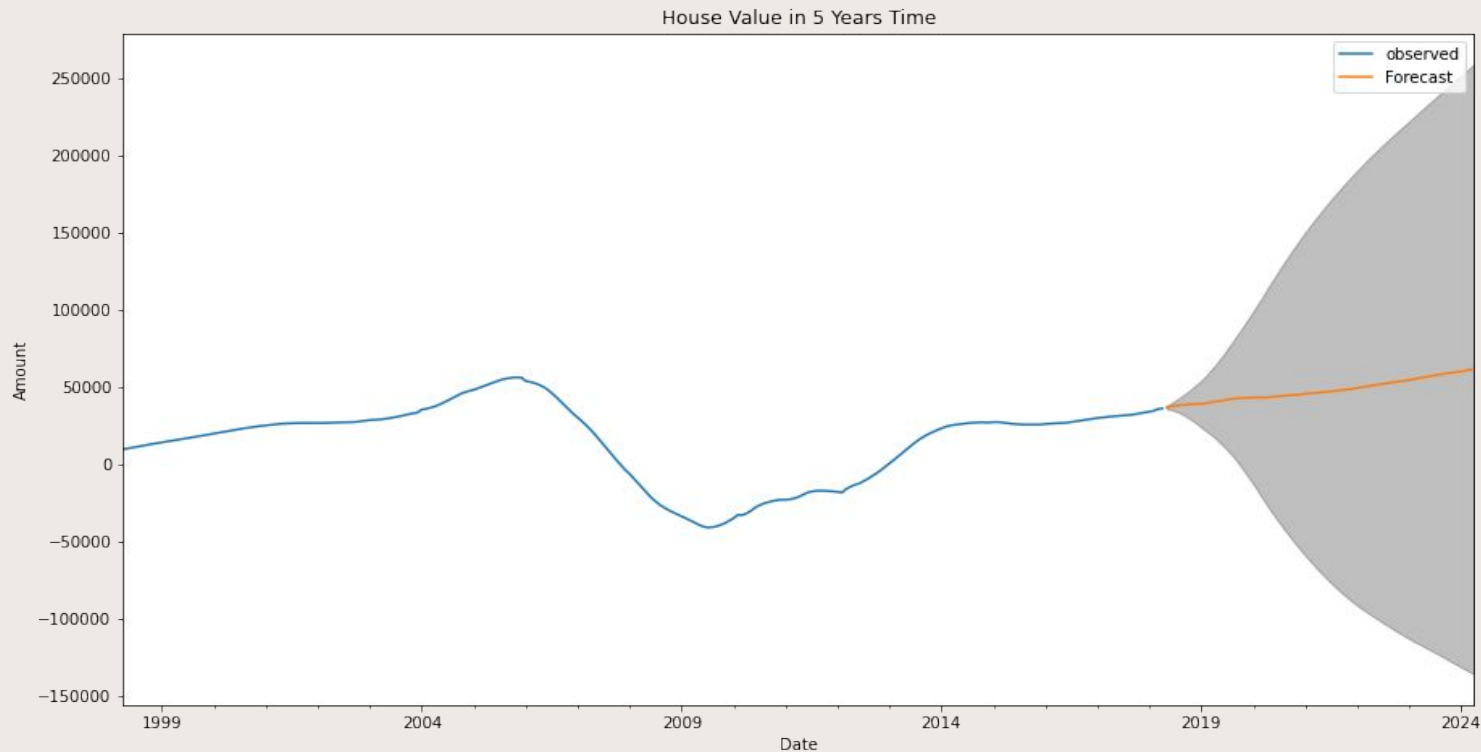
Conclusion & Recommendations

05

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 our model, we 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.

Thanks!

Do you have any questions?
Check our github repository