charles4data /
Forecasting-Real-Estate-Prices---Time-Series-Modeling

<> Code   ⊙ Issues   ⇅ Pull requests   ▷ Actions   ⊞ Projects   📖 Wiki   ⊙ Security   📈 Insights   ⚙

We are acting as consultants for a fictional real-estate investment firm. The firm has asked us what seems like a simple question: What are the top 5 best zip codes for us to invest in?

☆ **0** stars   ⑂ **0** forks   ⊙ **1** watching   ⌁ Activity

🌐 Public repository

⑂ **main** ▾   ···

⑂ Branches   ⬙ Tags

👤 **charles4data** update files   ···   1 minute ago   🕐 **17**

View code

☰ **README.md**   ✏

Time-Series-Analysis-For-Real-Estate-Investment

Business Understanding

Background

The U.S. real estate market has undergone significant tranasformation between 1996 to 2018.In recent years lower interest rates and government internation have restored stability in the market. Our project aims to analyze this period's real estate data to identify top-performing zip codes for short-term investments with high return on investment and low risk.

Business Problem

Waridi Investments a real estate investment seeks to identitfy the top 5 zip codes with the highest potential return on investment (ROI) within 5 years using the available data for maximum gain.

Objectives

1. To determine the top 5 zip codes that show highest potential return on investment(ROI) in 5 years.
2. To identify the top 5 states with the highest ROI.
3. To determine the best month to sell property to maximize ROI.

Data Understanding

We used the Zillow dataset for our analysis.The dataset consists of the following key features:

- **RegionID**: A unique index for each region.
- **RegionName**: The unique zip code of the region.
- **City**: The city in which the zip code is located.
- **State**: The state in which the zip code is located.
- **Metro**: The metropolitan area in which the zip code is located.
- **CountyName**: The county in which the zip code is located.
- **SizeRank**: A numerical rank of the size of the zip code, ranging from 1 to 14723.
- **1996-04 through 2018-04**: Monthly median housing sales values from April 1996 to April 2018, resulting in 265 data points for each zip code.

We will analyze this dataset to extract meaningful insights that will guide our forecasting process.

Data Overview

The dataset contained:

- Median monthly housing sales prices data for 265 zip codes in the USA, over the period of April 1996 through April 2018.
- 14,723 records and 272 variables.
- The features contain a mix of numerical and categorical variables.

Data Preprocessing

- Data cleaning including checking for duplicated rows, missing values, dropping unnecessary and renaming columns.
- Data analysis involved selecting top 10 states based on domain knowledge of the best states to invest in.
- Calculating Return on Investment (ROI) and Co-efficient of Variation(CV) to determine where to invest.
- Created visualizations of ROI and CV in relations to states and zip codes as well as time to average home sales price.

Exploratory Data Analysis

Selecting Top States for Investment

- Calculated ROI for different periods.
- Computed mean and standard deviation for ROI.
- Derived CV to assess investment risk.
- Identified top 10 states based on ROI analysis on graph shown below.
- State of Florida has the highest ROI for all specified periods.
- Even though New Jersey comes third in 22-year ROI, it performs poorly in the 5-year and 3-year ROI.

Comparative Analysis of States by ROI_all, ROI_5yr, and ROI_3yr
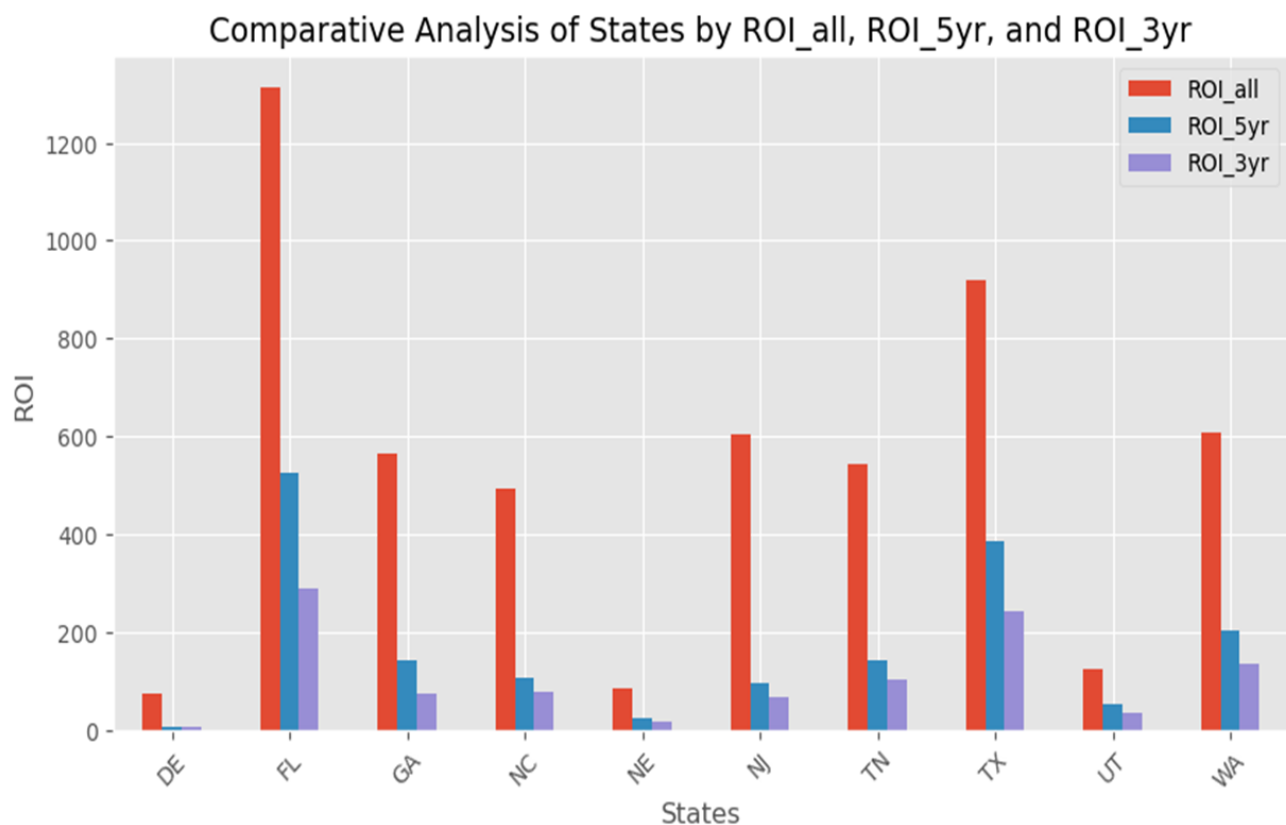
5-year ROI and CV

- We created bar charts showing comparison of differerent zip code between ROI and CV.

- 31527 ((Jekyll Island, Georgia) stands out again with a high 5-year ROI and CV.

- 7302 (Jersey City, New Jersey), 30317 (Atlanta, Georgia) and 78702 (Austin Texas) have high ROIs but higher than most CV values.
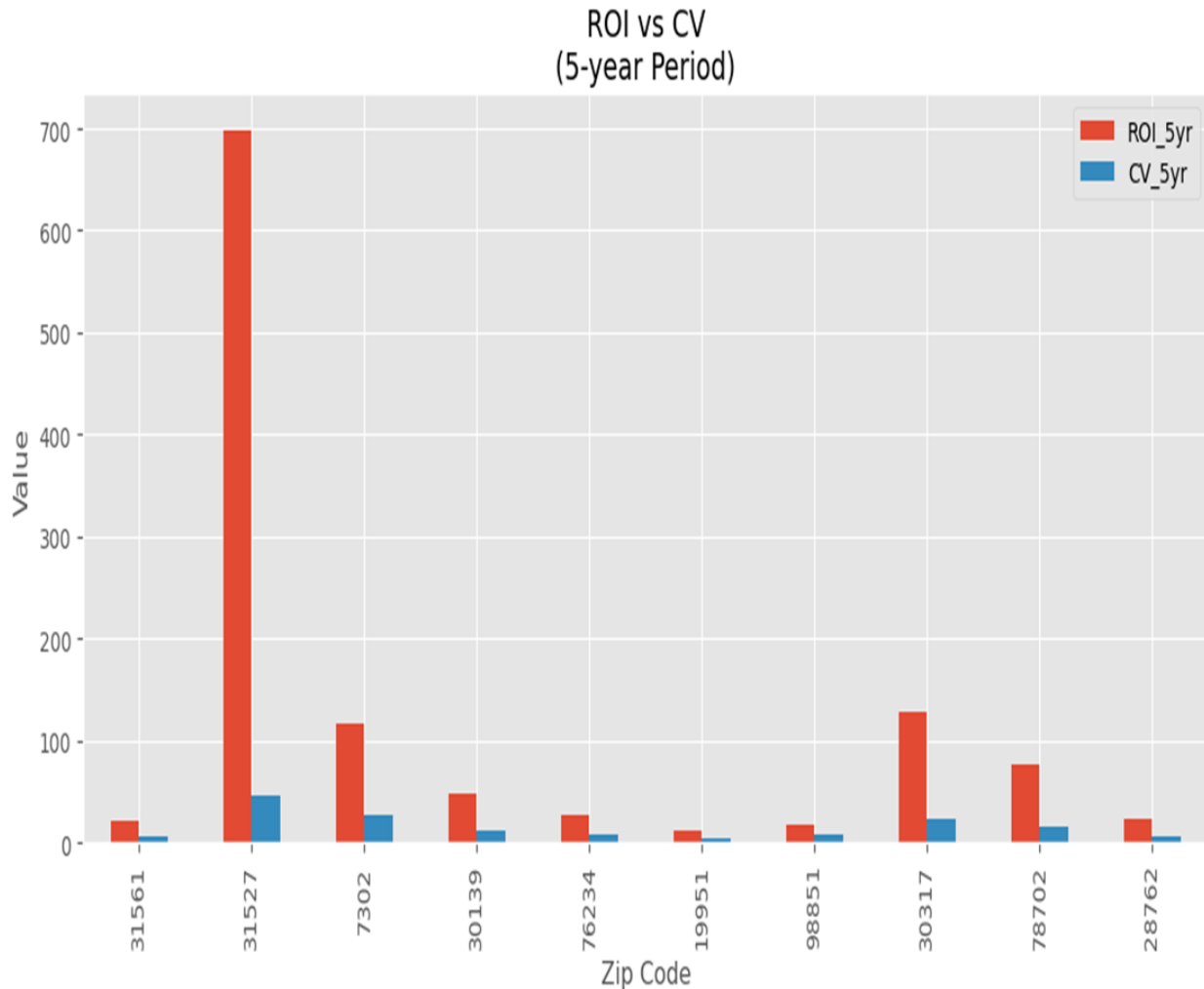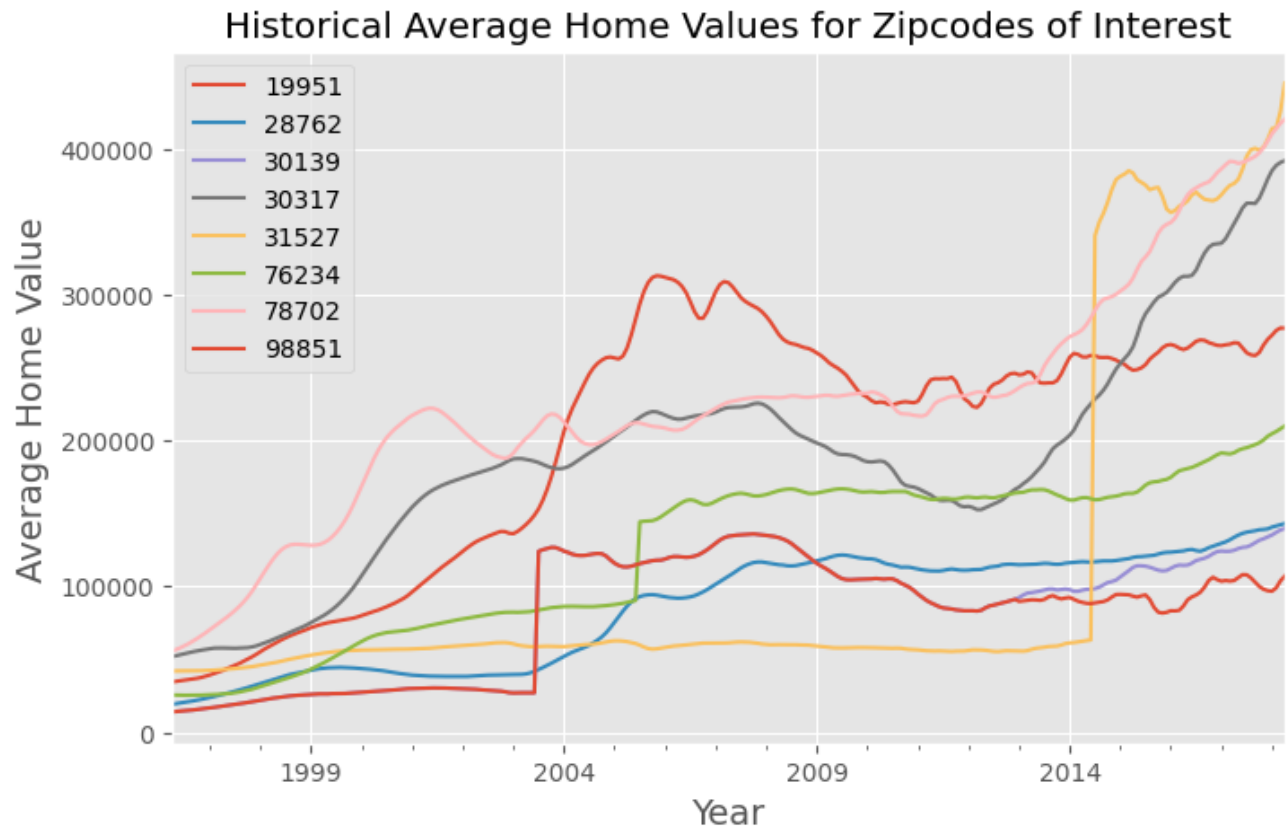

Data preparation for modeling:

The times series chart below shows the top 10 zipcode which met the criteria of our client from which modeling would be carried out to determine the top 5 zipcodes to invest in.

- Time series analysis shows an upward trend from 2012 onwards.

- There is a general decline in the average housing sales price from 2008 to 2012.

Time Series Modeling:

- Model used past home sale values to predict future sales.
- We sampled the top 10 zip codes based on ROI and CV.
- Revealed the top 5 zip codes based on RMSE and AIC metrics.
- The models used were SARIMAX and Facebook Prophet models.
- Facebook prophet models was able to produce better focus as by the graph below.
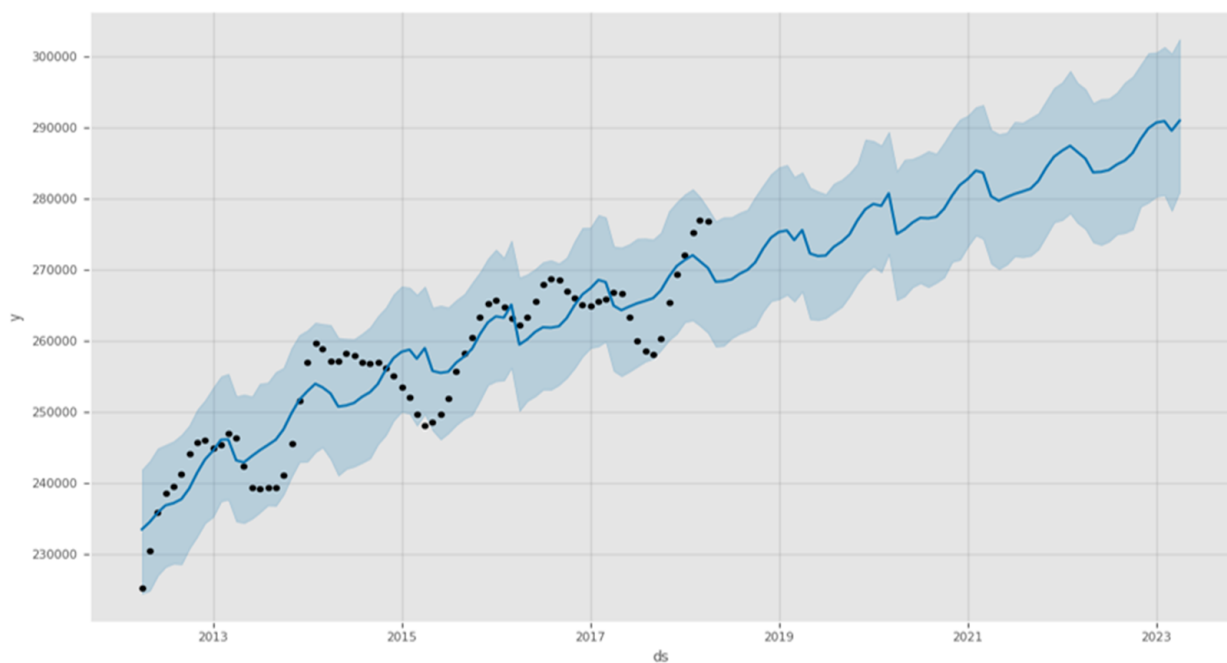
Conclusion

- **76234(Decatur, TX), 19951 (Harbeson, DE), 30317 (Atlanta, GA), 28762 (Old Fort, NC) and 98851 (Soap Lake, WA)** are top 5 zip codes to invest in because of high ROI and recourses should be put in them. The best months to sell to maximize profit are:
- December in Harbeson, DE.
- April in Old Fort, NC and Fairmount, GA.
- June in Decatur, TX, Soap Lake, WA.
- **Facebook prophet** is a better model in predicting future housing sales values. -The return on investment and the risk are not necessarily correlated. **Sometimes a high ROI could also mean a high risk and vice versa**

Recommendations

Based on our analysis of the real estate investment data, we have the following recommendations:

- Invest on the top 5 zip codes and allocate resources towards them.
- Obtain current data after 2018 for current predictions.
- There is need to consider outside factors (housing grade, property size, renovations) that may also affect Sales value.

Releases

No releases published
Create a new release

Packages

No packages published
Publish your first package

Languages

- **Jupyter Notebook** 100.0%