Categakii / Zillow_Housing_Forecasting

<> Code   Issues   Pull requests   Actions   Projects   Wiki   Security   Insights   Settings

Zillow_Housing_Forecasting   Public

main

Branches   Tags

	Categakii Add files via upload ...	3 minutes ago	1
	House_price_prediction_fin...	Add files via upload	3 minutes ago
	House_price_prediction_fin...	Add files via upload	3 minutes ago
	README.md	Add files via upload	3 minutes ago
	ZILLOW REAL ESTATE INVE...	Add files via upload	3 minutes ago
	housing_image.jpg	Add files via upload	3 minutes ago

Go to file   Add file   Code

About

No description, website, or topics provided.

Readme

Activity

0 stars

1 watching

0 forks

Releases

No releases published
Create a new release

Packages

No packages published
Publish your first package

README.md

Zillow Real Estate Forecasting by Boma Yangu Inv

Overview

Boma Yangu real estate investment firm is seeking to make strategic real estate investments in specific geographic areas in the US. Their aim is to maximize returns on their investments by identifying regions with high growth potential, favorable market conditions, and strong demand for real estate properties. They have contracted our services to analyze relevant data and provide a recommendation on the top 5 best zip codes to invest in.

Objectives

- To provide a recommendation for the top 5 best zip codes for a real estate investment firm to invest in
- To develop a model that can accurately forecast expected returns from the property market over time

Data Understanding

This dataset was obtained from Zillow Research. It contains real estate property values from April 1996 to April 2018 for various Zipcodes. The columns include:

- RegionID: An identifier for the region.
- RegionName: The zip code.
- City: The city where the zip code is located.
- State: The state where the zip code is located.
- Metro: The metropolitan area related to the zip code.
- CountyName: The county where the zip code is located.
- SizeRank: A ranking based on the size or importance of the zip code.
- Monthly Data (e.g., 1996-04, 1996-05, ...): These columns represent the median real estate prices for a given zip code for each month from April 1996 to April 2018.

Metrics of Success

What defines our top best-performing zipcodes? Return On Investment: We will use the Return On Investment ratio to measure 'best'. The metrics we will be optimizing against/minimizing are the AIC and RMSE. The lower the AIC and RMSE, the better our model will be at forecasting.

Data Description

- The dataset contains 14722 rows and 272 columns
- Some columns in the dataset contain missing values
- Data has both continuous and categorical features comprising the following data types; objects, integers, floats
- The dataset is in wide format with the time periods appearing as columns. We converted it into a long format

Data Preparation

We dropped the unnecessary columns, 'Metro ' and 'CountyName' from our dataset to make it easier to work with. We also renamed the column, 'RegionName' to 'Zipcode'

Data Preprocessing

To make informed investment decisions, we enriched our dataset by calculating a few key metrics that provided insights into the historical performance and variability of housing prices for each zip code. Here is the breakdown of the metrics we used:

Historical Return on Investment (ROI):

This metric provides a measure of how much the value of a property in a given zip code has appreciated (or depreciated) over the entire span of our dataset. A higher ROI indicates that properties in this zip code have historically appreciated in value at a faster rate.

Standard Deviation of Monthly Values (std):

This metric measures the volatility or variability in monthly housing prices for each zip code over the time span of the dataset. A higher standard deviation indicates greater volatility and potential risk, but also potential reward.

Historical Mean Value (mean):

This metric provides the average monthly housing price for each zip code over the duration of our dataset. It gives us a general sense of the typical housing price in a given zip code.

Coefficient of Variance (CV):

The CV is a standardized measure of dispersion of a probability or frequency distribution. For our dataset, it provides a relative measure of the volatility in housing prices for each zip code. A higher CV indicates more volatility when compared to the mean.

Identifying Top 10 Investment Opportunities

To identify the Top investment opportunities, we examined the CV for each zip code to understand the volatility of housing prices. We also set an upper limit for CV using the 60th percentile. This filters out zip codes with higher-than-acceptable risk and we identified zip codes that offered the best historical ROI and also fit within the defined risk profile.

- The average CV across all zip codes was approximately 0.220
- Based on our risk tolerance, the upper CV limit was set at 0.240

- The top 10 zip codes showcased high ROIs while also aligning with the firm's risk profile.
- They were as follows:

Zipcode	Location
49309	Bitely, MI
40107	Boston, KY
48822	Eagle, MI
49265	Onsted, MI
49425	Holton, MI
29645	Gray Court, SC
66206	Leawood, KS
48835	Fowler, MI
48894	Westphalia, MI
15486	Franklin, PA

Exploratory Data Analysis

Distribution of Property Values for Top 10 Zipcodes

Distribution of Property Values
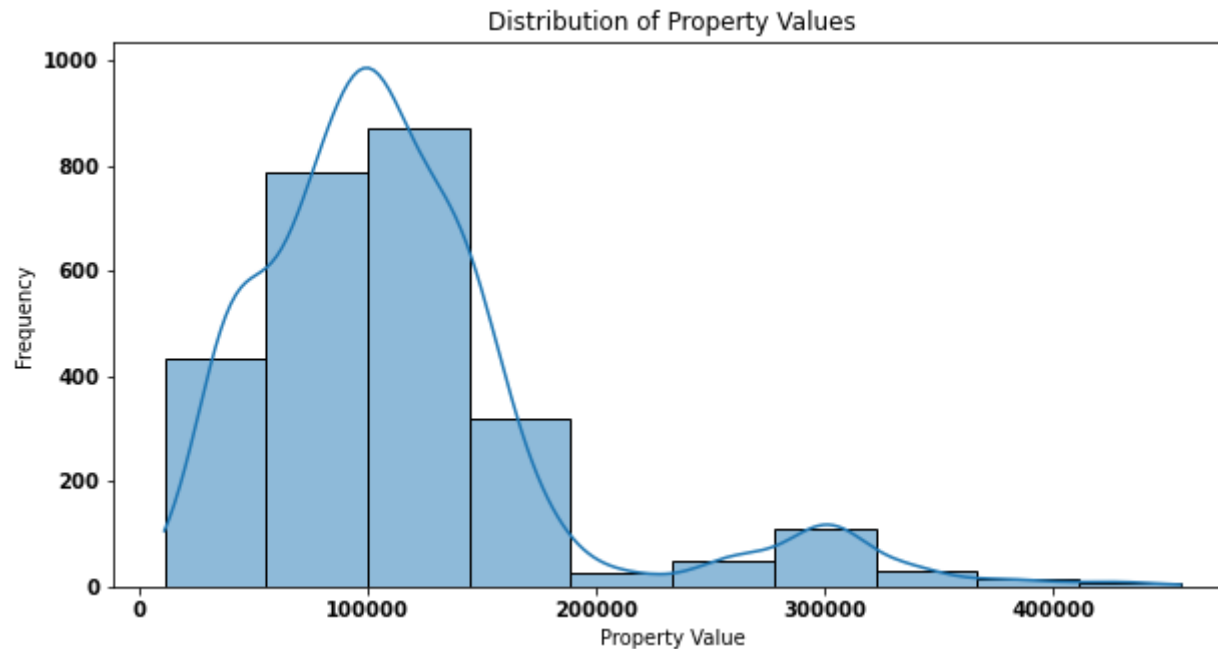
- From the plot, we can infer the common price ranges within these top-performing zip codes and identify any patterns or anomalies. This knowledge aids in understanding the typical property values in these areas and can inform investment decisions.

- The distribution of real estate values is right-skewed, with a long tail on the right side, indicating that the majority of the property values are concentrated in the lower price range.

- The shape of the distribution suggests that the real estate market consists of a mix of moderately priced properties and a smaller number of high-priced properties which form the tail of the distribution.

Time Series Plot for the Median Property Values of the Top 10 Zipcodes
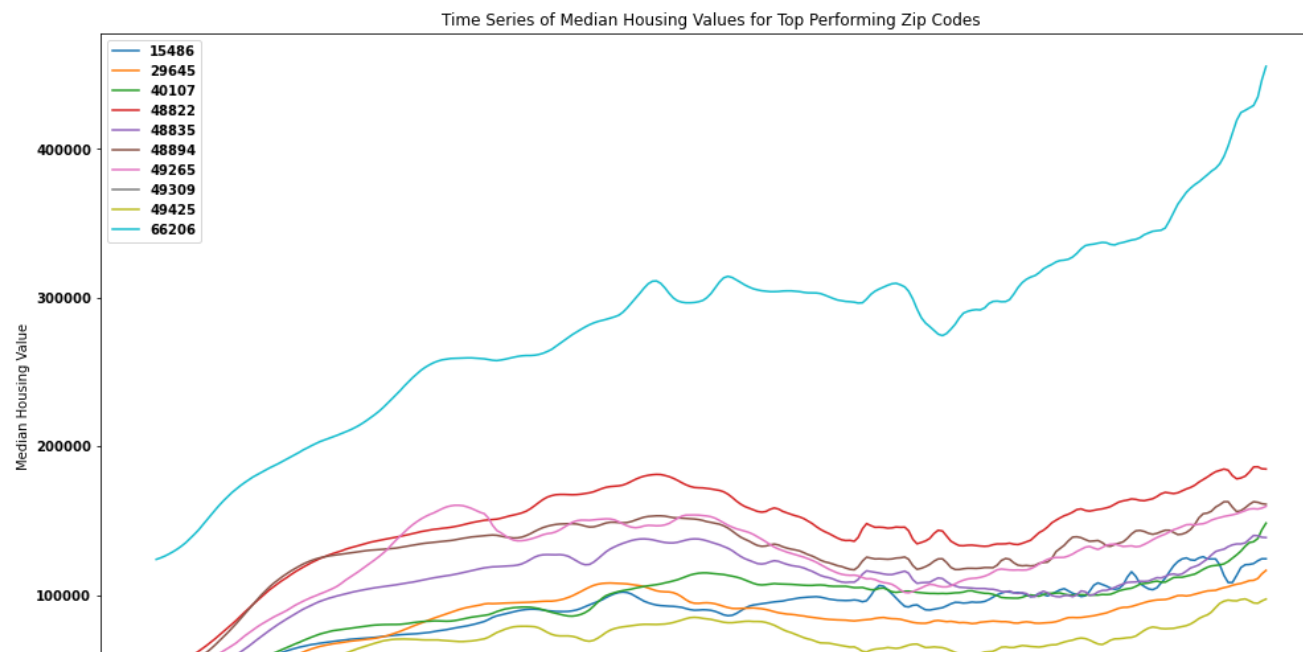
- From this plot, we observe the historical trends in property values for each zip code. Some zip codes show steady appreciation over time, while others might have more volatility or distinct growth phases. There were noticeable dips around 2008-2010, which corresponded with the global financial crisis. After this period, there's a clear recovery and steady growth for most Zipcodes.

Distribution of Property Values for Top 10 Zipcodes

Distribution of Property Values for top 10 Zip Codes

- From the plot, you can infer the median, variability, and potential outliers in property values for each zip code

Return on Investment for Top 10 Zipcodes

- From this visualization, we can infer that Zipcode 49309 has yielded the highest ROI over the years, followed by 40107

ROI Vs CV for Top 10 Zipcodes

- The scatter plot provides insights into the trade-off between risk (volatility in property values) and return (historical appreciation). Zip codes closer to the top-left corner offer higher ROI with lower risk, making them more desirable for investment.

Methods

We used the ARIMA time series modelling technique as our baseline model and improved it through SARIMA modelling. SARIMA extends the capabilities of ARIMA by incorporating seasonal components of the time series data into the model. We then used the best parameters to select the top 5 best zipcodes based on the predicted sales and return on investment uisng the following criteria:

- Treating the last data point of the time series as our Current Value.

- Projected Value taken as the summation of the current value and the forecasted value.

- Using the Current Value and the Projected Value to calculate the ROI for each zipcode.

- Sorting the zipcodes by ROIs to finally determine the top 5 best zipcodes.

Conclusion/ Interpretation

From our analysis, the best Zipcodes to invest in based on a combination of predicted sale price and positive predicted ROI are:

Zip Code 48822:

- Investment Value: $184,700

- Predicted Sale Price after 2 years: $197,993.26

- Predicted ROI: 7.20%

This Zipcode stands out with a promising ROI of 7.20%, indicating a potential return on investment. With an initial investment of $184,700, the projected sale price after 2 years suggests favorable growth potential. This could be attributed to positive market conditions and demand factors.

Zip Code 49309:

- Investment Value: $49,200

- Predicted Sale Price after 2 years: $51,777.49

- Predicted ROI: 5.24%

Zip code 49309 offers a solid investment option with a predicted ROI of 5.24%. Despite its lower initial investment value, the projected sale price after 2 years indicates notable growth potential. This suggests a favorable investment opportunity considering the moderate risk

Zip Code 29645:

- Investment Value: $116,500

- Predicted Sale Price after 2 years: $118,177.38

- Predicted ROI: 1.44%

Zip code 29645 presents a stable investment prospect with a predicted ROI of 1.44%. While the ROI is relatively modest, the expected increase in sale price after 2 years could provide steady returns, suitable for conservative investors seeking stable growth.

Zip Code 48835:

- Investment Value: $138,600

- Predicted Sale Price after 2 years: $138,618.56

- Predicted ROI: 0.01%

Zip code 48835 offers a nearly negligible ROI of 0.01%. While the projected growth is minimal, the initial investment value of $138,600 remains relatively stable. This might attract risk-averse investors who prioritize stability over high returns.

Zip Code 49265:

- Investment Value: $159,600

- Predicted Sale Price after 2 years: $158,830.09

- Predicted ROI: -0.48%

In contrast to the top 4 zipcodes, zip code 49265 displays a negative ROI of -0.48%, indicating a potential loss based on the projected sale price after 2 years. Investors might want to exercise caution with this zip code, as it suggests a declining property value trend.

Recommendations

- Zipcodes 48822, 49309, 29645, and 48835 stand out as the top investment choices based on their projected ROIs and anticipated property value growth over the next 2 years. Considering these insights, we strongly advise Boma Yangu to consider these options for their investment prospects. These selections offer a compelling blend of attractive sale prices and favorable returns on investment, all while accounting for Boma Yangu's risk threshold.

- For zip code 49265, we strongly urge Boma Yangu to approach it with caution, as it signals the possibility of incurring a loss.

Next Steps

- Collect more data with added exogenous variables like economic indicators, interest rates and demographic data that could potentially influence the housing price values, then model them in a SARIMAX model. This could lead to better predictions and better models overall.

- Build more complex and advanced models such as RNNs, e.g. LSTMs which make use of neural networks. These algorithms are more powerful than traditional time series. They are able to capture any complex patterns in our time series, hence leading to better forecasts/predictions.