bnittalee / Time-Series-Model Public

Project 4

0 stars 0 forks

Star Watch

Code Issues Pull requests Actions Projects Security Insights Se

main

bnittalee Update Notebook.pdf ... now 18

View code

README.md

New Mexico, Albuquerque: Top 5 zip codes to invest in

By Brittney Nitta-Lee

Business and Data Understanding

My clients are real estate investors with a focus on properties in King County. Seeking to escape the cold weather, they are interested in exploring investment opportunities in Albuquerque, New Mexico. I will identify five potential zip codes for investment in the area and provide a list of recommendations along with suggested next steps.

Overview

The goal of this project is to analyze the historical real estate data for Albuquerque, New Mexico, and identify the top 5 zip codes for investment based on forecasted median home prices. A time series modeling approach is employed to predict future values, helping investors make informed decisions.

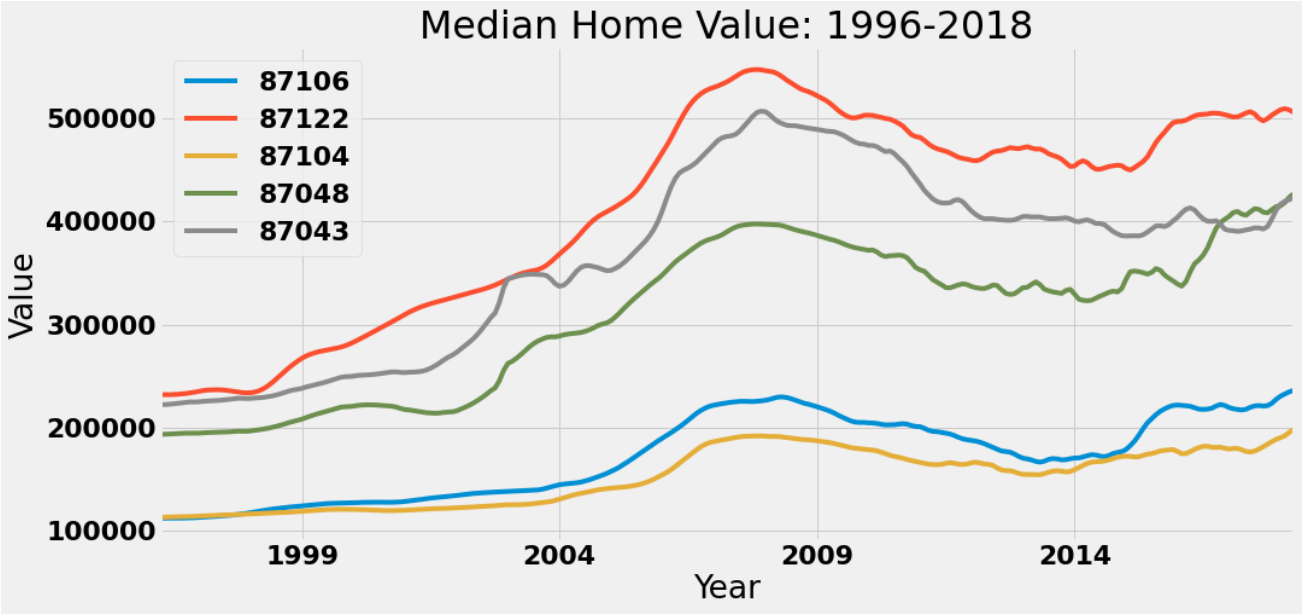
Data Understanding

In this time series modeling project, I used data from Zillow. The dataset includes 14,723 rows, each representing a zip code, and 272 columns. The data provides median sales for every zip code from April 1996 to April 2018.

Zip Code Selection

Firstly, I filtered the zip codes in the dataset by selecting Albuquerque in the Metro column and then sorted them based on their cumulative Return on Investment (ROI).
For assessing the accuracy of my model, I utilized the root mean squared error (RMSE) to compare it against my baseline naive model. The reason for using RMSE is that it quantifies the differences between predicted and actual values in the same units as the original data, facilitating a straightforward interpretation and comparison of various models.
Lastly, I transformed the data into a monthly time series format to create individual data points for forecasting future home values over the next four years.

Exploratory Data Analysis



Once I selected my final 5 zip codees, I created a visualization to see any trends or patterns. To make it easier to analyze and visualize the data, I resampled the time series to a monthly frequency with the start of each month as the observation point. This aggregated the data for each month and created a new dataframe with the data for the first day of each month. This was particularly useful for forecasting and filling in any gaps in the data.

As you can see, the visualization shows the median home price from April 1996 to Apirl 2018. The trend line has an upward pattern up until 2008 where you can see a downward trend. This is due to the Great Recession. The "Subprime Mortgage Crisis" was a period of time (2007 to 2010) when there was an increase in the number of high-risk mortgages that went into default and caused a ripple effect on the housing market and broader economy. This is important to highlight as I did not inlcuding this data into my modeling.

SARIMAX Model

FIn this project, I opted to implement a SARIMAX model as it is well-suited for capturing the seasonality in real estate prices, which tend to exhibit seasonal patterns. This model comprises autoregressive and moving average components and enables us to specify distinct orders for both. Additionally, we can introduce seasonal orders that permit further customization of the model to better align with the data for each zip code. One of the benefits of the SARIMAX model is that it provides interpretable results. To forecast future home values for the next four years, I utilized the resampled data as a data point.

Results

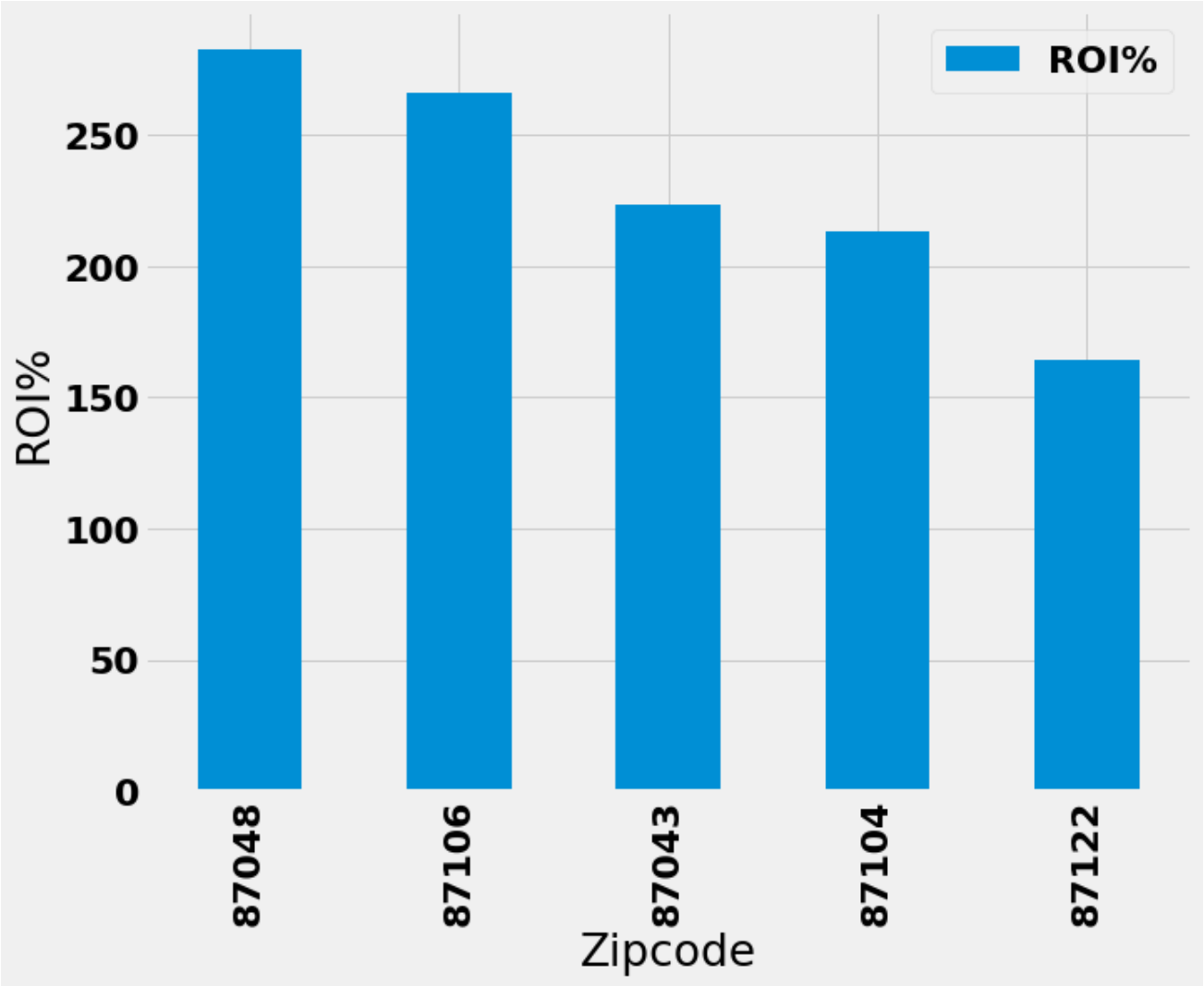
	Zipcode	City	2018 median value	rmse	low_conf	high_conf	forecast range	low_end	high_end	ROI%
0	87106	Albuquerque, NM	230500	28566.613682	236629.0	293381.0	56752.0	467129.0	523881.0	127.28
1	87122	Albuquerque, NM	508200	4276.872528	465108.0	543586.0	78478.0	973308.0	1051786.0	106.96
2	87104	Albuquerque, NM	190300	6118.499377	202629.0	239505.0	36876.0	392929.0	429805.0	125.86
3	87048	Corrales, NM	415000	16113.137999	447569.0	581744.0	134175.0	862569.0	996744.0	140.18
4	87043	Santa Ana Pueblo, NM	416700	16992.337405	375097.0	441879.0	66782.0	791797.0	858579.0	106.04

Each zipcode has the low and high confidence intervals, forecast range, the low end and high end of the forecast range and the ROI.

The low confidence interval and high confidence interval represent the lower and upper bounds of the range within which the true value of the predicted median home values is expected to fall. Basically, there's a 95% chance that the true median home value for a given zip code will fall within the range defined by the low and high confidence intervals.

The low end and high end of the forecast represents the lower and upper bounds, of the predicted range of median home values for a given zip code. These values are calculated using the SARIMAX model and historical data. They represent the range within which the true median home value is expected to fall based on the model's predictions.

The forecast range is the difference between the high end and low end values of the predicted range.



Recommendations

For those who are intersted investing in properties in New Mexico, there are the following zipcodes that has a high return on investment.

1. Corrales, NM (87048)
2. Albuquerque, NM (87106)
3. Santa Ana Pueblo, NM (87043) - Santa Ana Pubelo's median value in 2018 was 375,097 and $441,8790.
4. Albuquerque, NM (87104)
5. Albuquerque, NM (87122)

The model's inability to generate realistic forecasts resulted in identical forecast ranges for zipcodes 87048, 87106, 87104, and 87122. However, analyzing the ROI paints a different picture. Investors seeking property in New Mexico would be better off considering Corrales or Santa Ana Pueblo. A 2023 Zillow search revealed that Santa Ana Pueblo's home prices range from $1,795,950, while Corrales' prices range from $3,800,000. These wide ranges suggest that further investigation of these zipcodes could reveal intriguing insights.

Next Steps

1. Further investigation into Santa Ana Pueblo and Corrales.
2. Include external factors that may influence real estate prices, such as population growth or unemployment rates.
3. Investigate rapidly growing neighborhoods in New Mexico.

See the full analysis in the Jupyter Notebook

For additional info, contact Brittney Nitta-Lee at bnittalee@gmail.com

Repository Structure

```
├── .ipynb_checkpoints/
├── Data
├── Images
├── PDFS
├── .DS_Store
├── .gitattributes
├── Final-Notebook.ipynb
└── README.md
└── Zillow_data.csv
```

Releases

No releases published
Create a new release

Packages

No packages published
Publish your first package

Languages

- **Jupyter Notebook** 100.0%