Nicholas Lichtsinn

IST 718

Lab 6

**Introduction**

Choosing where to invest money is crucial for long term success and financial stability, most trust funds and financial planners have different algorithms and ways to predict return on investment and where clients should invest money.

This project will use time series forecasting models created with Python to determine the three zip codes that are the best investment opportunity for the Syracuse Real Estate Investment Trust.

**Analysis**

A picture containing chart

Description automatically generatedFirst I read in the dataframe with zillow data on average home value by zipcode over the last 20 years.

Chart, histogram

Description automatically generatedI removed n’s from the data as there were some missing values for some zip codes. I then created a tim series line graph for each of AR largest metro areas, Hot Springs, Little Rock, Fayetteville and Searcy. I aggregated the graphs onto one visualization shown below.

I then dropped the columns from the dataframe that would not be included in the ARIMA model. RegionID, City, State, CountyName, RegionType and StateName were removed.

Text

Description automatically generated with medium confidenceI created a 5 year and 3 year ROI for each zipcode.

Chart, bar chart

Description automatically generatedChart, line chart

Description automatically generatedI then changed the data from wide format to long using melt and changed sipcode to a string and the date column to datetime.

I plotted the average home value by month for all zipcodes and the top ten mean value zipcodes.

To limit the runtime for the processing of the model I took only the zip codes in the top 1% of ROI to use in the model and dropped the metro and sizerank columns so that I could feed the zip codes and values into a list to loop into a dictionary to use as a dataframe with each zipcode as a column and dates as rows.

I chose one zip code to use as a model to find the acf and pacf values.

Chart, line chart

Description automatically generatedA train dataset was created using 80% of the dataset and the remaining 20% was used to create a test dataset to test the accuracy of the models. Chart

Description automatically generatedI split the data into a train and test set with 80% of the instances used in the training set. And ran the arima model with a p, q and d of 1 to test the efficacy.

Chart, line chart

Description automatically generated

As you can see the model did not accurately predict the test set so I used the auto arima package to find the optimal order for the model. The resulting model was SARIMAX(0,2,0) this was then used in a loop for all zip codes to predict home values through 2024 with an upper and lower confidence interval.

A picture containing graphical user interface

Description automatically generatedBelow are the three most valuable zipcodes with the highest ROI based on the model.

A picture containing graphical user interface

Description automatically generated

A picture containing graphical user interface

Description automatically generated

**Conclusion**

To down sample I took the zip codes with the highest calculated 5 year ROI. After using the auto arima to calculate the optimal order for the model the zip codes that would provide the best investment were 90020, 90212, 90211. These all provided a predicted ROI of over 1.7 and based on the confidence intervals have very low risk involved for the investor. There were other zip codes that were cheaper to invest in if the trust is low on funds but given the exponential increase in housing prices in California in recent years this investment strategy makes sense if the trust has the necessary capital to invest.