Parin Patel

Predicting Best Real Estate Investment Location – Based on Zillow data.

Introduction:

The last few years has seen a housing market driven partly by narrowing of home available on the market and low mortgage rates. According to CFRA Research's Global Director of Industry and Equity Research, Ken Leon, strong growth in the housing market was driven by higher levels of consumer confidence, higher household income, and employment growth. Since home growth is expected to level-off soon, Syracuse Real Estate Investment Trust (SREIT), may be wondering where their best bet chance investing in the United States is.

The goal of this study is to predict which zip codes provide the best investment for the Syracuse Real Estate Investment Trust (SREIT). In addition, we wanted to narrow down our results for SREIT to three final zip codes that would be the best investment opportunity.

Analysis:

Data Pre-Processing: Importing Datasets and Cleaning

For our analysis we imported the housing data described below from the following link:
http://files.zillowstatic.com/research/public/Zip/Zip_Zhvi_SingleFamilyResidence.csv

- House dataframe– a .csv that had data about ZHVI (Zillow Home Value Index). Interestingly, because of the encoding of this excel file, we had to specify when importing the file that it was a “"ISO-8859-1” encoding.

To learn more about the dataset, we used the information provided on the Zillow data's site.
https://www.zillow.com/research/data/

The data is about ZHVI (Zillow Home Value Index). According to the website, the ZHVI is a smoothed, seasonally adjusted measure of the typical home value and the market change across a given region and housing type.

The data has 30,415 rows (see Figure 1) and 293 columns. Each individual row contains a region within a county in the United States. The columns "RegionName" is denoted by that regions zipcode. The first 6 columns are related to the region/county/state/city/metron name information. In general, the raw data is sorted by in ascending order by the SizeRank column. The Zillow website does not specifically state what the parameters or definitions of this column are. It is likely that this number either is related to the average ZHVI per square foot for that region across all the years, the regions proximity to urban centers. Therefore, the highest ranked cities based on SizeRank have the largest available ZHVI per sq foot for potential urbanization. These smaller regions are typically close to city centers but have higher home value indices.

For the data types of the columns: there are 260 columns that are float64, 29 that are int64, and 4 columns that are object.

Review Data:

```
housedf.info()
```

```
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 30416 entries, 0 to 30415
Columns: 293 entries, RegionID to 2020-01
dtypes: float64(260), int64(29), object(4)
memory usage: 68.0+ MB
```

Figure 1

The four object columns are city/state/metro/CountyName.

Cleanup for this file was quite standard, we started by dropping the NA's from the dataset. This resulted in 12,306 rows remaining. We then checked for null values (NAN), and found none. Finally, we counted the number of unique zip codes and determined that there are 12,306 unique zipcodes in our dataset (Figure 2).

```
print(f'Unique zip codes: {housedf.RegionName.nunique()}')
```

```
Unique zip codes: 12306
```

Figure 2

Before we conduct a preliminary, exploratory analysis to narrow down our zip codes, we will explore some of the housing prices across the cities in the US.

Exploratory Analysis:

Exploring the Top 10 Cities: Mean Housing Prices Per Cities In US

We wanted to start by just trying to understand what the housing market price index looks like by cities. We did this by first focusing our data to the years from 1997 to 2017. And then finding the mean price of that city by averaging out price across the year range. We then narrowed it down to see what the 10 most expensive cities, according to housing indices, are in the US. You can see from figure 3 below that Aspen, Colorado is the most expensive housing market, with average prices around $3,062,168. This was followed by Palm Beach, CA where the average housing cost is $2,647,894

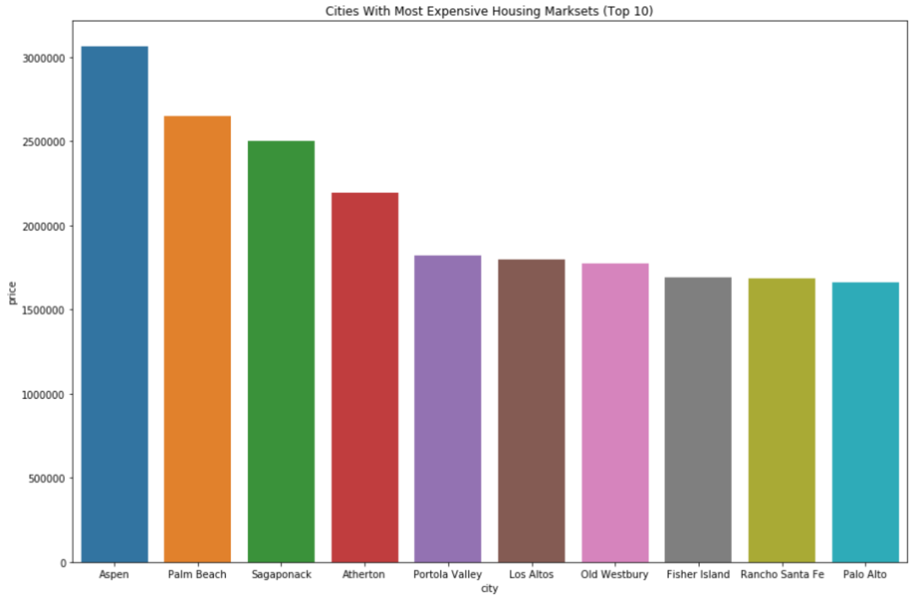


Figure 3

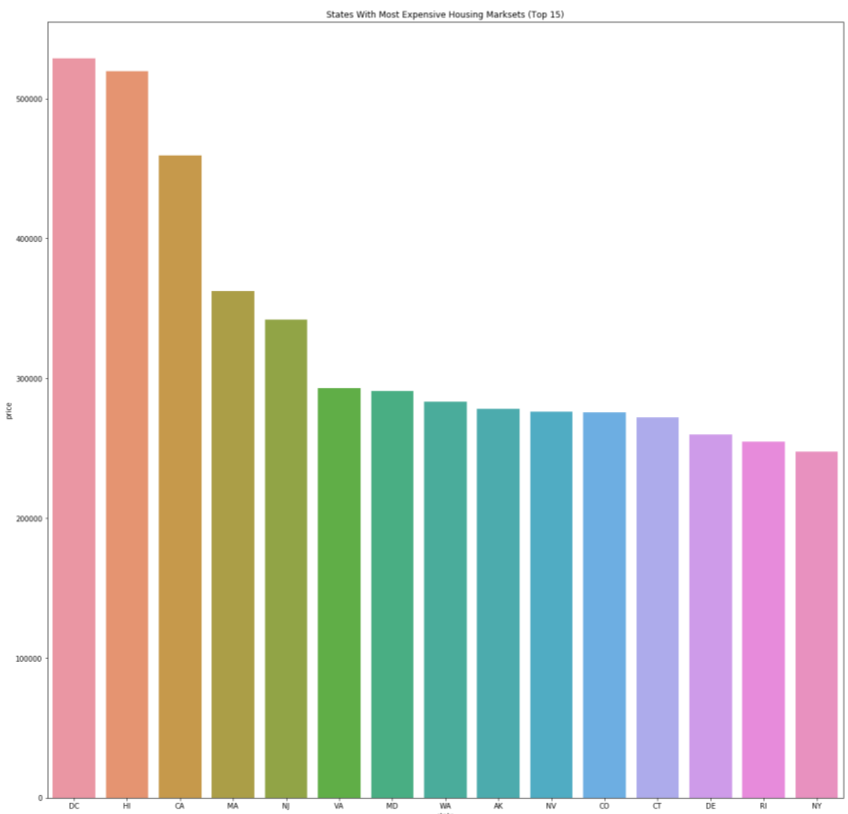


Figure 4

From Figure 4, we can see housing market prices by state/territory. Interesting, the Washington DC has the most expensive median home value, at $528,721.22. I was confused by this at first, so went to the Zillow website to investigate. I found the following link (https://www.zillow.com/washington-dc/home-values/) . Figure 5 below shows a screenshot of the website. After reviewing the site and reflecting upon my current housing situation in Washington DC, I believe DC has the highest median home value in US because their may not be as many entries for this region, as compared to other states. While DC definitely has many high-home values, it is likely that a few very high value homes are skewing the median home value. In the future, it might be a good idea to weigh the mean values to avoid this issue.

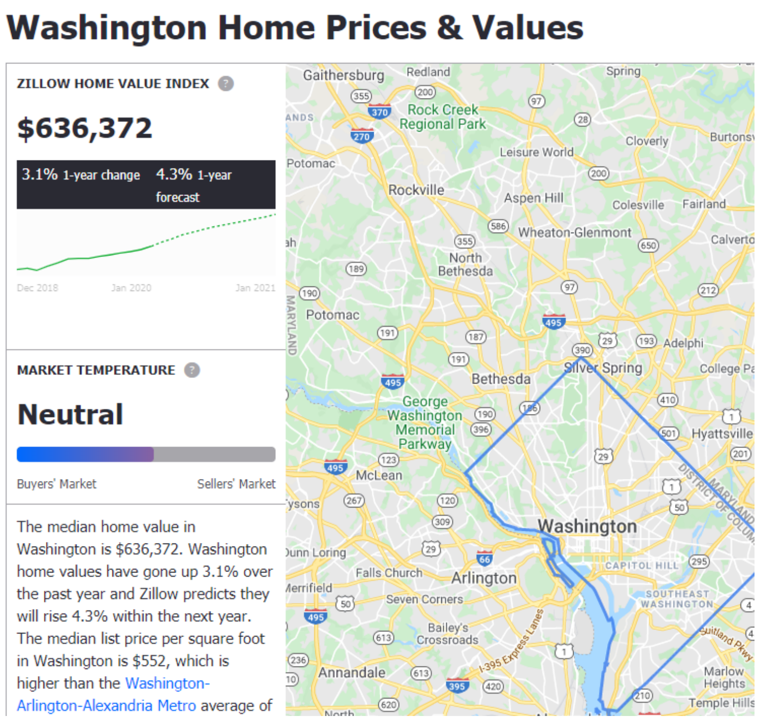


Figure 5

Determining Top Zip Codes with Highest Return on Investment.

For our preliminary analysis we wanted to figure how we should narrow down our focus by zip code. Since we can expect that 12,306 zip codes would be a lot for any prediction model to work with, we wanted to determine parameters for what characteristics could be used from the dataset. We determined that, after reviewing the Zillow website that we want to select the zip codes based on areas with the largest SizeRank. These areas would be close to urban centers but still have competitive home values. According to the UN, as more and more people moving to population centers, like cities, the growth of neighborhoods surrounding these areas will increase. This will lead to a growth in local economies aided by the growth of the nearby city.

To figure this out, we will first determine how many zip codes we could get if we capped the SizeRank at 25%. This cutoff value is about 3,921.5 with their being about 3,077 zip codes. By narrowing down the number of zip codes by focusing on the top 25%, we were able to reduce the number of zip codes to model from 12,306 to 3,077. You can see this in Figure 4.

```
#Calculate the 25% cutoff value.
sr_25 = housedf.SizeRank.quantile(q=0.25)
print(f'Size Rank 25% cutoff value: {sr_25}')

#Get data frame with top 25 zipcodes.
house_top25= housedf[housedf['SizeRank']<sr_25].drop(['RegionID','City',
                                                       'State','Metro','CountyName',
                                                       'SizeRank'],axis=1)
print(f'Amount of zipcodes: {len(house_top25)}')
```

```
Size Rank 25% cutoff value: 3921.5
Amount of zipcodes: 3077
```

Figure 6

Our next step is to narrow down the number of zip codes figure by trying to find zip codes around the average values around the median housing value. We chose to set the average cut off at 15% above and below the zip codes 1-year median house value. We chose this method based on the Investopedia article that addresses the purchasing power and associated risks of the target market. This is why, by selecting zip codes with a coefficient variation below the 65 percentiles, we are able to mitigate the number of risky zip codes in our final prediction. By setting the average cut off values at the 65th and 35th percentile, we determine the values to be below $ 378, 370.12 and above $236137.98 (see Figure 4).

```
house_top25['yr_avg']=house_top25.iloc[:,-12:].mean(skipna=True, axis=1)
#Get zipcodes with an average value 15% above the median and 15% below .
print(house_top25['yr_avg'].describe(),'\n')
#Calculate the 65% cutoff value (15% above).
q_65 = house_top25['yr_avg'].quantile(q=0.65)
print(f'Average Value 65% cutoff value: {round(q_65,2)}')
#Calculate the 35% cutoff value (15% below).
q_35 = house_top25['yr_avg'].quantile(q=0.35)
print(f'Average Value 35% cutoff value: {round(q_35,2)}')
```

```
count    3.077000e+03
mean     4.026637e+05
std      3.514643e+05
min      2.325358e+04
25%      2.005301e+05
50%      2.932240e+05
75%      4.710101e+05
max      3.597905e+06
Name: yr_avg, dtype: float64

Average Value 65% cutoff value: 378370.12
Average Value 35% cutoff value: 236137.98
```

Figure 7

Based on these narrowed parameters, we are able to narrow down the number of zip codes to 923. We will now move forward with this dataset to determine the relative risk of the zip codes.

Our next steps involve calculating our return on investment (ROI), Coefficient of variance (CV), standard deviation (std) of all the monthly values and calculate historical mean by region. We started by determining our ROI by narrowing the data between 1997 and 2017. We then determined the standard deviation of monthly housing values by zip code. Finally, we calculated the mean value of the zip code and the coefficient of variance. Our new values (STD, mean, ROI, and CV) were placed into a data table. Our next part requires reviewing the coefficients of variance to determine the zip codes with highest calculated ROI. Even through we were asked to determine the three best zip codes for investment, we will pick the top 5 zip codes right now. We will narrow our suggestion to three after we run our time series models. Figure 5 shows the five zip codes with the highest calculated ROI.

```
Zipcode : 55406
Location: Minneapolis, MN

Zipcode : 55418
Location: Minneapolis, MN

Zipcode : 29412
Location: Charleston, SC

Zipcode : 55104
Location: Saint Paul, MN

Zipcode : 63108
Location: Saint Louis, MO
```

Figure 8

In summary, we were able to ID the top 5 zip codes by choosing zip codes were the average home value was around the calculated median. We utilized the SizeRank column to help us narrow down the zip codes further - we assumed that SREIT would want to invest in areas that were larger/had a larger population density. This would, in the long term, mean more likelihood of growth, job opportunities, and urbanization. We finally calculated our ROI for each zip code and calculated the historical mean and calculated the measure of variability for each zip code.

Predictive Modeling: Time Series

Time Series Plots: Four Cities in Arkansas

Our initial data analysis started with creating individual time series plot for each of the four cities in Arkansas from 1997 to 2020. We then combined all four individual plots into one combined one.

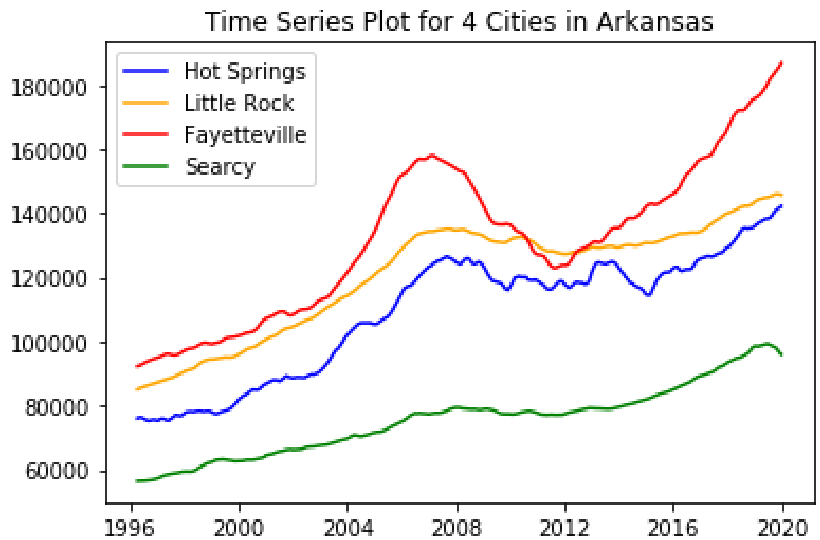


Figure 9

For each line you see above, we plotted the mean housing market value for each of the four Arkansas cities for each respective year from 1997 to 2020. We finally plotted these values across the time series plot. The results can be seen in Figure 9 above. Based on the plot, we can clearly see that the Fayetteville's housing market shot up after 2012 and is likely a hot market to enter in 2020. A good second and third option would be Little Rock and Hot Springs. Both of these markets, while not seeing as high of growth as Fayetteville, still seem to have a strong growing housing economy. More importantly, I would advise SREIT to steer clear of Searcy's housing market, since it has had the lowest growth over the years between

Time Series Plots: Narrowing Down Which Is the Best Zip Code to Invest In Using Forecasting

The final part of our project is to determine the three best zip codes for SREIT to invest in. We will want to use a timeseries analysis to rank the five zip codes we determined earlier to have the highest ROI , based on forecasted 10-year returns. From the beginning, we noticed that the wide-dimension of the zip code datatable was resulting in some warning and problems with packages. Therefore, we decided to resolve this first by making a function that would transform the data from a long-form time series data table into a separate list for each of the five zip codes. We will use this to then make five different data frames for the time series of each zip code with the frequency listed by year-month. This was done by looking at the descriptive statistics of the values for each of the zipcodes and then plotted against the time series. The below time series plot (figure 10) shows the median home value for the five zip codes: 55406, 55418, 29412, 55104, and 63108. We can see from the graph that the 63108, St. Louis, MO, has the highest median home value. This is then followed by 29412, Charleston, SC. The two Minneapolis zip codes (55406 and 55418) have similar patterns and prices over the years. This is expected since they are both regions around the same metropolitan area.

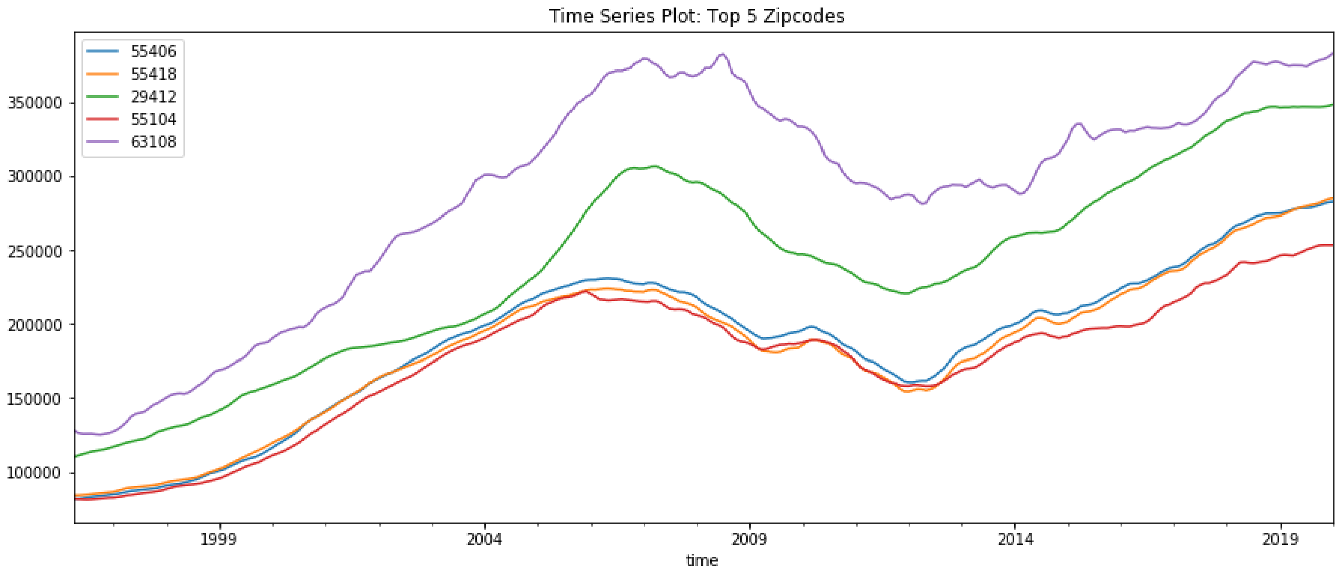


Figure 10

Our next section looks to find what we can expect as returns per month on our investment. If we can determine our expected average return value for each zip code, then we can determine which of the zip codes is the best value for SREIT. The following plots (Figure 11-15) are the percentage of returns for each of the individual zip codes.

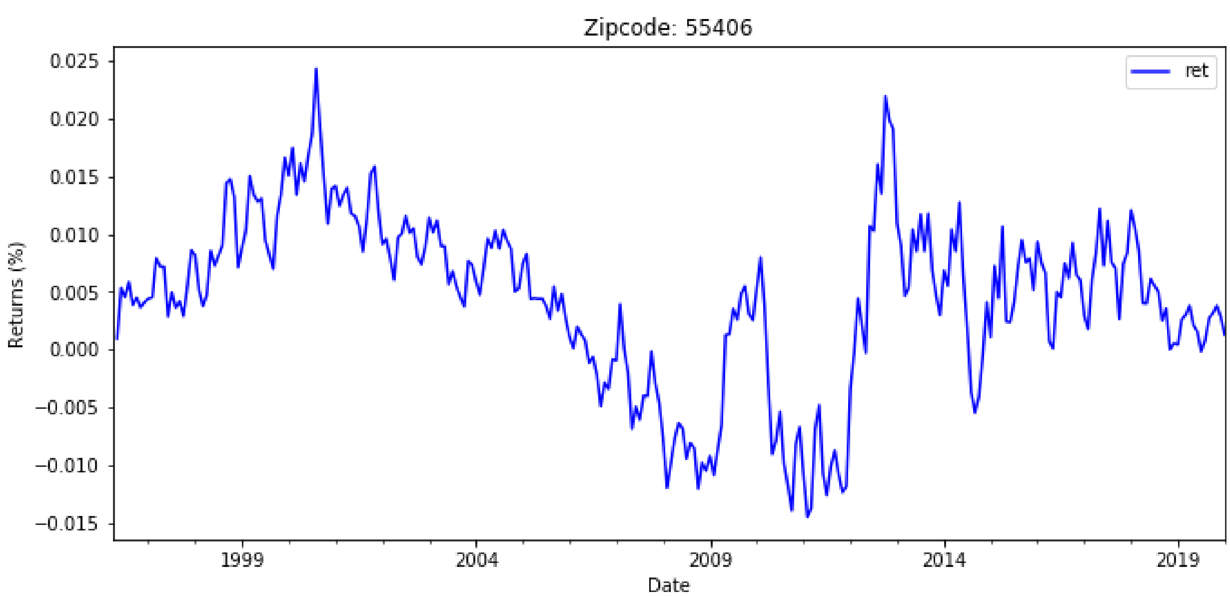


Figure 11

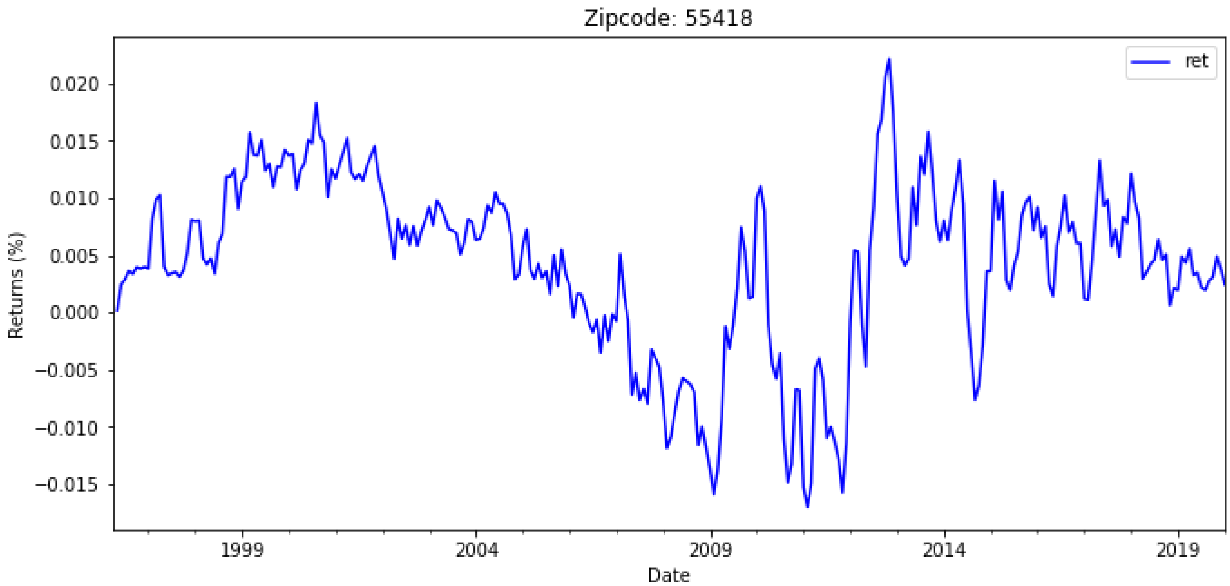


Figure 12

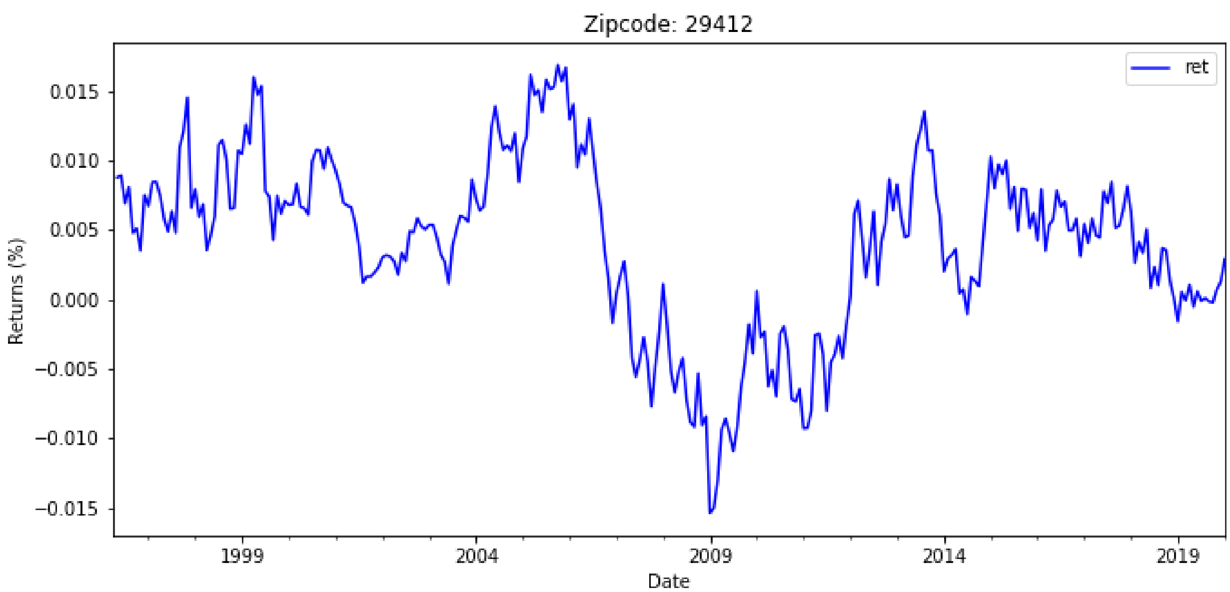


Figure 13

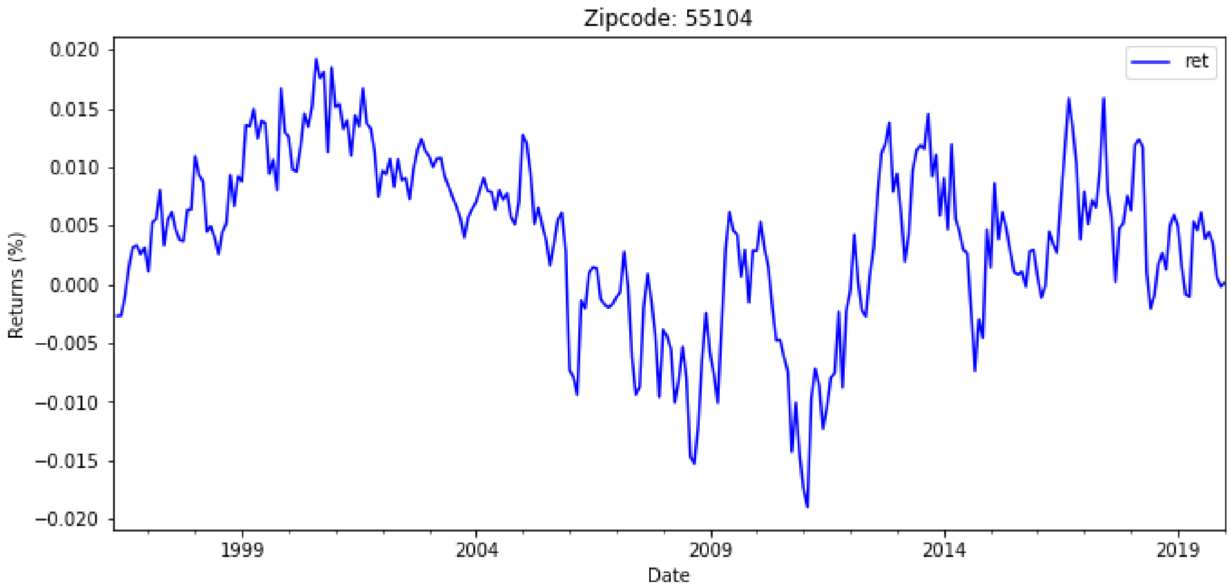


Figure 14

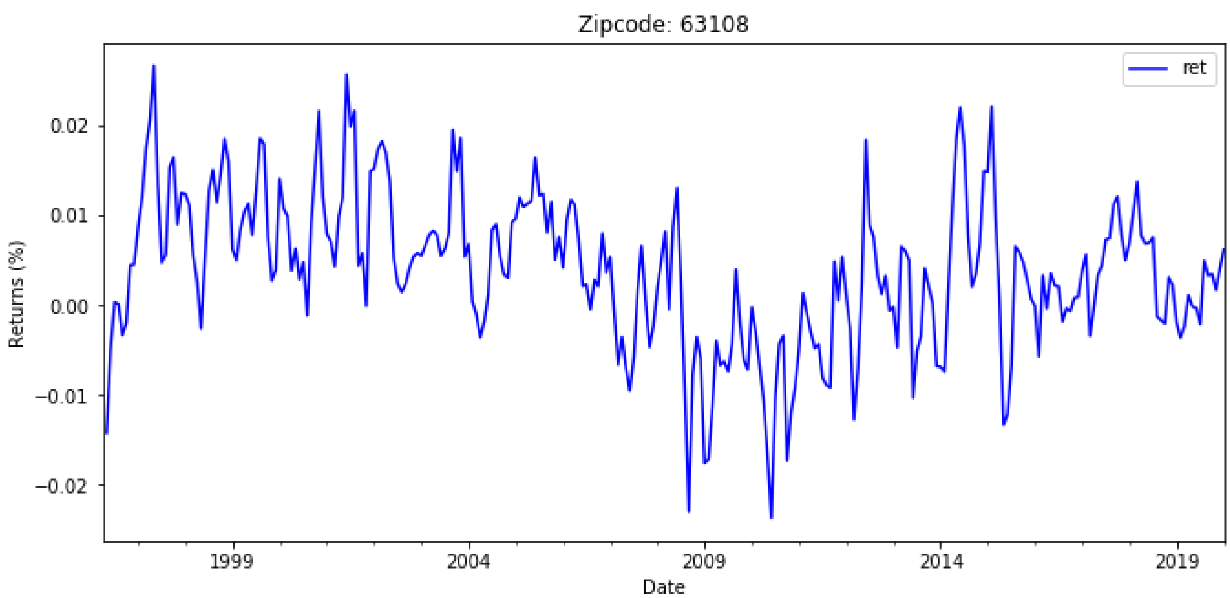


Figure 15

These individual ROI's show a potential for a wide variety of return on investments each month. It seems like Minneapolis and St. Paul (both in MN) have the lowest variability in returns. Taking this information, we not go into our ARIMA Model. Specifically, we will use ACF and PACF for each of the five zip codes to identify lag values that are in the ARMA mode. ACF describes how well the present value of the time series is related with its past values. This is different than PACF, which describes the correlation of the residuals with the next lag value. PACF helps us to understand if there is any hidden information in the residual which can be modeled in the next lag.

Zip Code 1 : 55406
Location: Minneapolis, MN

For the first analysis we reviewed the 55406-zip code for Minneapolis, MN. Figure 16 shows the ACR and PACF graphs for this zip. The results of our graph shows that we can limit the number of lags on the x-axis to 50 to make the plot easier to read. We can see that the ACF plot is tailing off towards the end. And that the PACF shows a single major spike at around 28 lag. Since the ACF plot is tailing off , there is a chance it will end up with AR and MA parameters in the final mode.

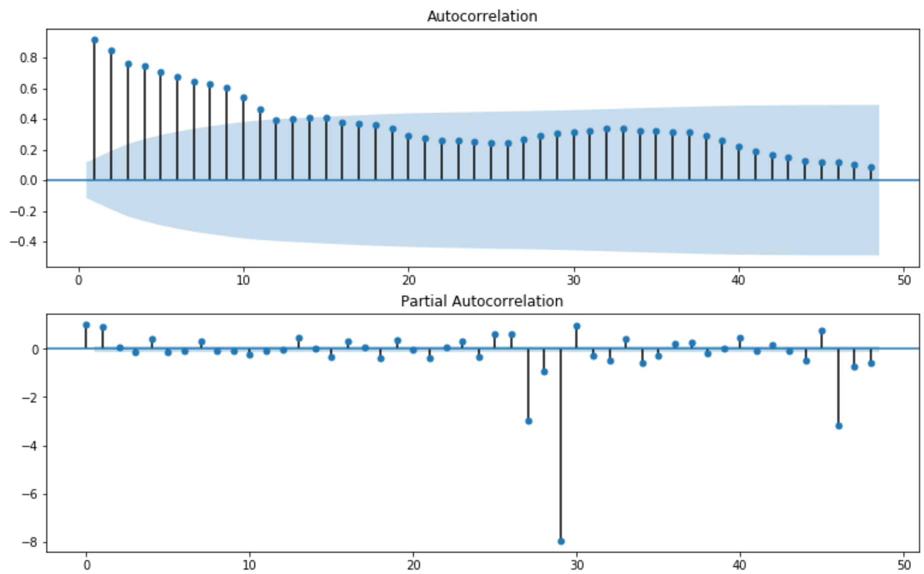


Figure 16 -- Zipcode 55406

The plot below shows the seasonal plot for the 55406-zip code. This plot seeks to find the seasonality of the ACF plot by differencing the rolling mean. When looking at the ACF plot we can see at statistically significant peak in the plot around lag 12. This is not seen in the other lags of the ACF plot, so it is likely that there is no seasonality to this. However, we can use the SARIMA model to get a better understanding.

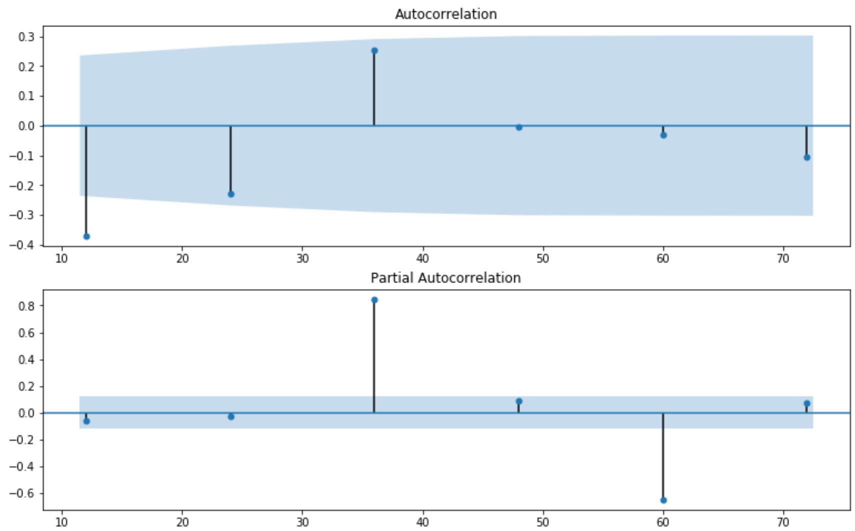


Figure 17- 55406 Seasonal Plot

Based on the results seen in below, Figure 18, the fitted, seasonal ARIMA model seems to capture most of the signal from the monthly returns. Our results show that first, the residuals are normally distributed (can see this in the Q-Q plot and the histogram). There is also a lot of background noise in the residual plot. We can see that the Ljung-Box Test has a p-value of 0. Sine this is less than 5%, we can reject the null hypothesis in favor of the alternative that the residuals are autocorrelated.

```
                                     SARIMAX Results
==========================================================================================
Dep. Variable:                              ret   No. Observations:                    249
Model:             SARIMAX(2, 0, 2)x(0, 0, 2, 12)   Log Likelihood                 1202.657
Date:                          Sun, 23 Feb 2020   AIC                           -2391.314
Time:                                  18:50:54   BIC                           -2366.691
Sample:                              05-01-1996   HQIC                          -2381.403
                                   - 01-01-2017
Covariance Type:                            opg
==============================================================================
                 coef    std err          z      P>|z|      [0.025      0.975]
------------------------------------------------------------------------------
ar.L1          0.1620      0.057      2.819      0.005       0.049       0.275
ar.L2          0.8110      0.056     14.499      0.000       0.701       0.921
ma.L1          0.9502      0.055     17.406      0.000       0.843       1.057
ma.L2          0.6863      0.047     14.459      0.000       0.593       0.779
ma.S.L12      -1.2066      0.070    -17.249      0.000      -1.344      -1.069
ma.S.L24       0.4030      0.061      6.580      0.000       0.283       0.523
sigma2      3.409e-06   3.15e-07     10.810      0.000    2.79e-06    4.03e-06
===================================================================================
Ljung-Box (Q):                      119.36   Jarque-Bera (JB):                11.47
Prob(Q):                              0.00   Prob(JB):                         0.00
Heteroskedasticity (H):               1.68   Skew:                            -0.31
Prob(H) (two-sided):                  0.02   Kurtosis:                         3.84
===================================================================================

Warnings:
[1] Covariance matrix calculated using the outer product of gradients (complex-step).
```

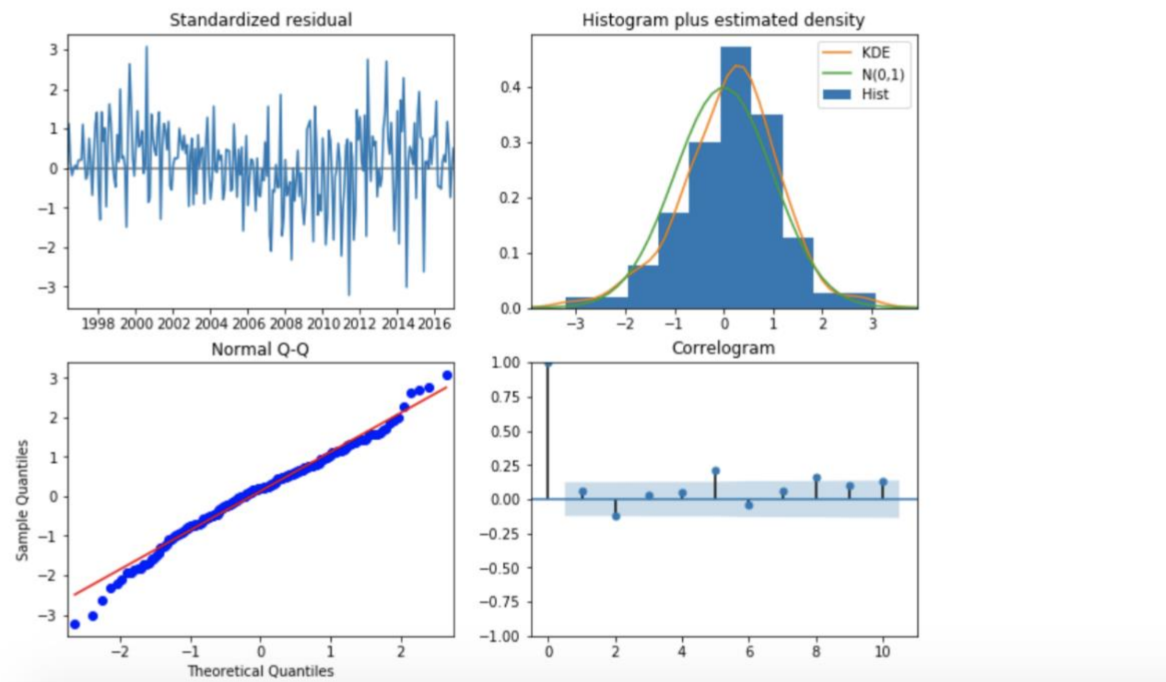


Figure 18

Our next steps involve trying to understand the amount of noise in our model. We calculate the RMSE of the training model, which is a metric which measures how of the data is noise and how much a signal worth considering. Specifically, for our case, we will perform a goodness of fit. Test by calculating the RMSE on the train data and the test data.

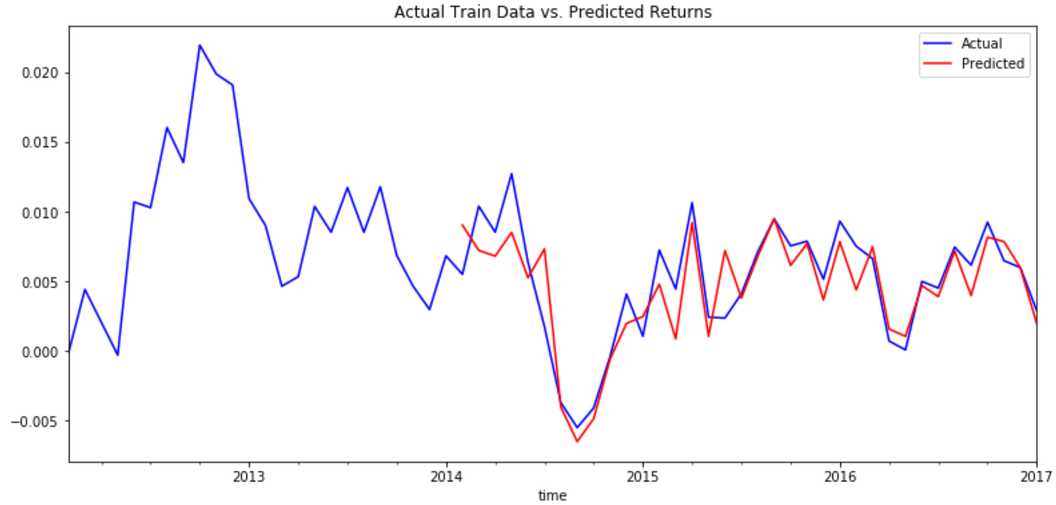


Figure 19 Zip code 55406

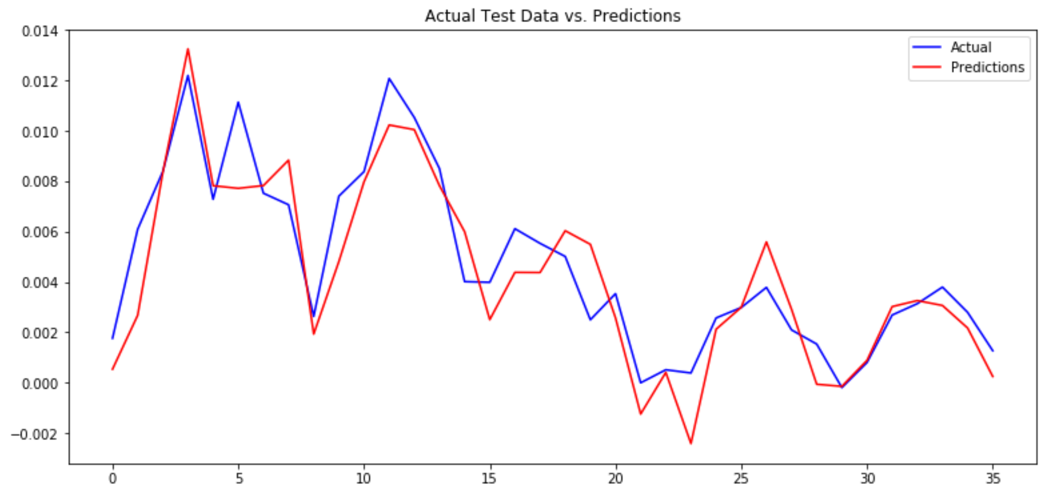


Figure 20 Zip Code 55406

Finally, we have created a final forecasting model that predicts the monthly returns for the 55406 zip code for the next 1,3,5, and 10 years.

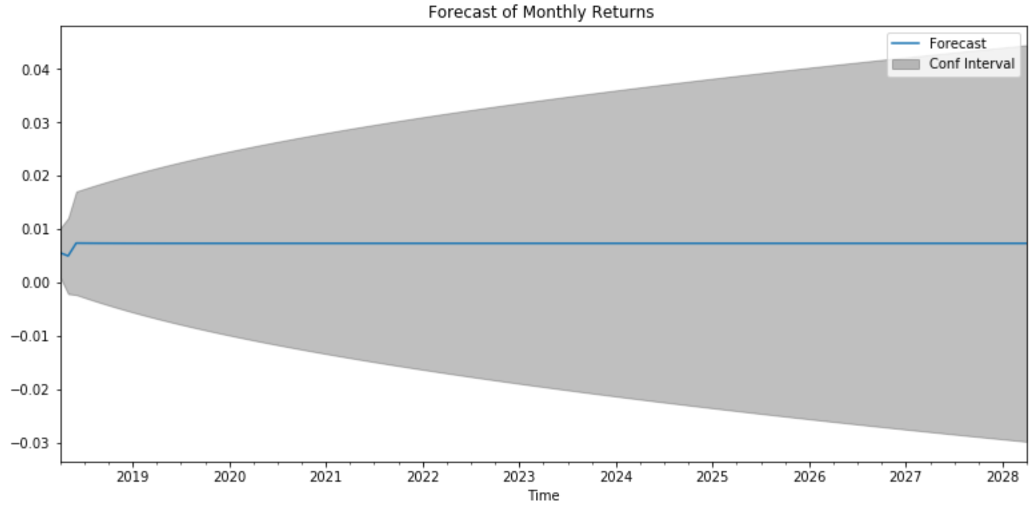


Figure 21 - Forecasting for 55406

```
Total expected return in 1 year: 8.68%
Total expected return in 3 years: 29.38%
Total expected return in 5 year: 54.01%
Total expected return in 10 years: 139.85%
```

Based on the results of this zip code for Minneapolis, MN we can expect a total long-term return of 139.85%. Overall, this is a good investment, but lets see how the others fare.

Zip Code : 55418
Location: Minneapolis, MN

For the first analysis we reviewed the 55418 -zip code for Minneapolis, MN. Figure 22 shows the ACR and PACF graphs for this zip. Based on the results of the ACF and PACF plots we can see that that the ACF is tailing off while the PACF spiked at about 43, and then started to tail off again. It is likely that we will have the AR and MA parameters in the final model.

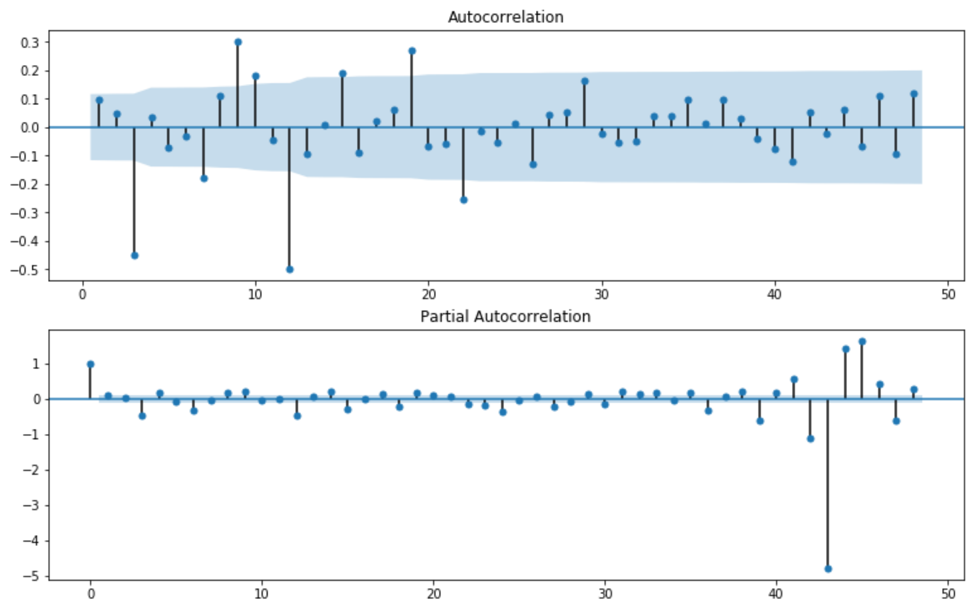


Figure 22 Zip Code 55418

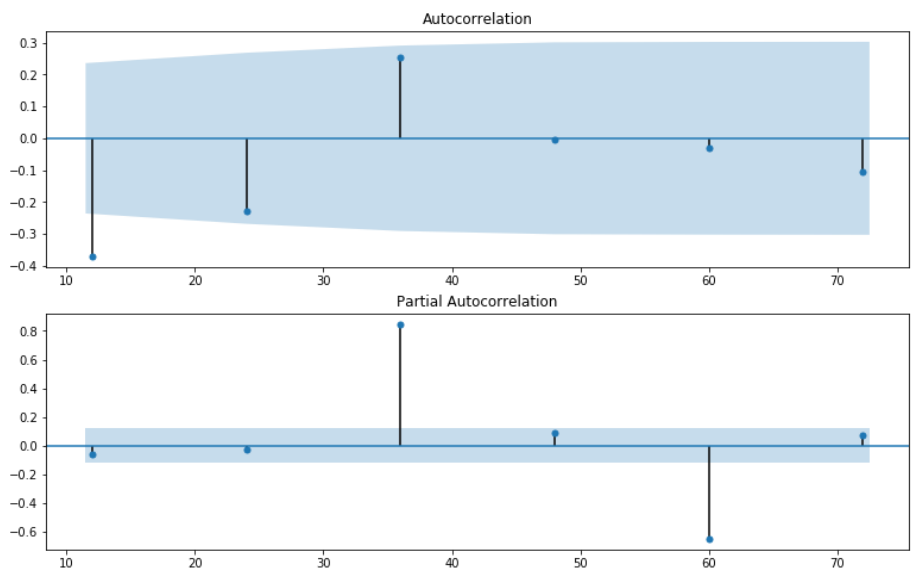


Figure 23 - Zip Code Seasonal Plot for 55418

The results of the seasonal plot for the 55418- zip code above (figure 23) show that the ACF and PACF graphs have a statistically significant autocorrelation on lag 12 in the ACF plot. Since this is again not repeated afterwards, there may not be seasonality in the data. We will then use the ACF to find the best non-seasonal and seasonal parameters for this zip code in our fitted SARIMA model.

```
                               SARIMAX Results
==============================================================================
Dep. Variable:                    ret   No. Observations:                  249
Model:               SARIMAX(1, 1, 4)   Log Likelihood                1114.766
Date:                Sun, 23 Feb 2020   AIC                          -2217.531
Time:                        18:58:29   BIC                          -2196.451
Sample:                    05-01-1996   HQIC                         -2209.045
                         - 01-01-2017
Covariance Type:                  opg
==============================================================================
                 coef    std err          z      P>|z|      [0.025      0.975]
------------------------------------------------------------------------------
ar.L1          0.5764      0.452      1.276      0.202      -0.309       1.462
ma.L1         -0.4354      0.441     -0.987      0.324      -1.300       0.429
ma.L2          0.0543      0.103      0.527      0.598      -0.148       0.257
ma.L3         -0.5468      0.080     -6.816      0.000      -0.704      -0.390
ma.L4          0.3391      0.180      1.883      0.060      -0.014       0.692
sigma2      7.291e-06   5.23e-07     13.953      0.000    6.27e-06    8.32e-06
===================================================================================
Ljung-Box (Q):                      163.48   Jarque-Bera (JB):                41.10
Prob(Q):                              0.00   Prob(JB):                         0.00
Heteroskedasticity (H):               4.38   Skew:                             0.49
Prob(H) (two-sided):                  0.00   Kurtosis:                         4.74
===================================================================================

Warnings:
[1] Covariance matrix calculated using the outer product of gradients (complex-step).
```

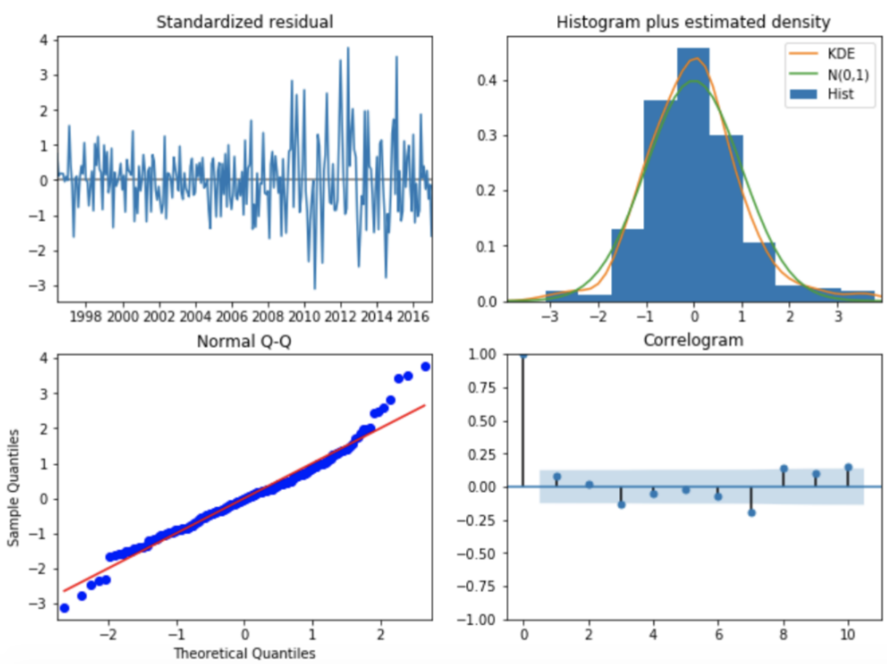


Figure 24 -- Zip code 55418

We can see that after fitting this model, from our above results ,this is a pretty good fitted model. We expect to have good predicted returns later. The residuals are definitely not distributed normally; however, they are autocorrelated. We can see this by the very low p-values of the Ljung-Box and Jarque-Bera tests. Both tests resulted in a probability of 0.00.

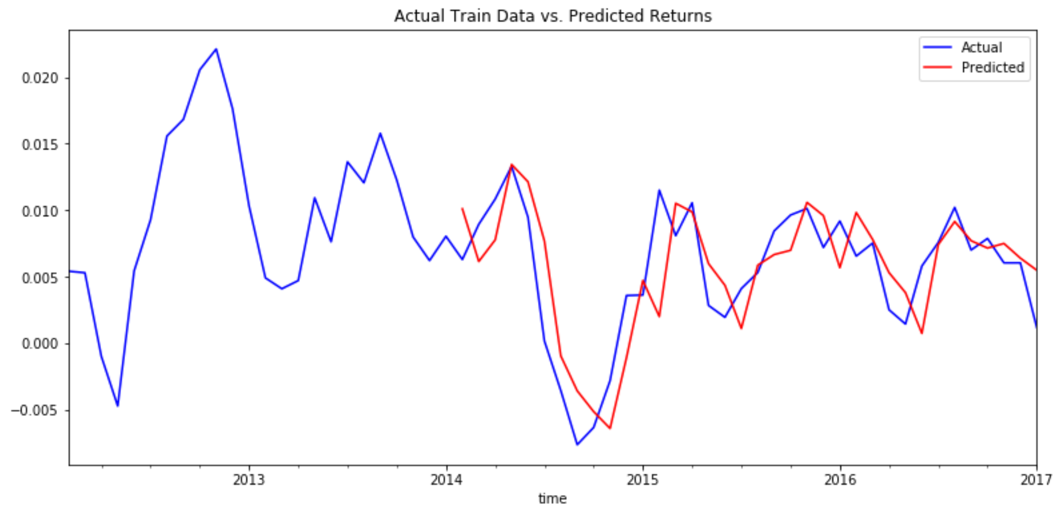


Figure 25- 55418

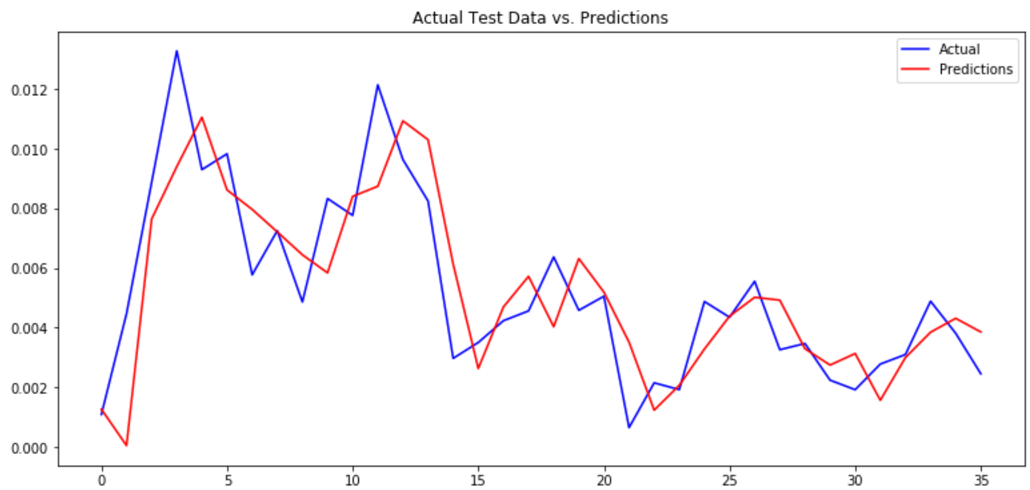


Figure 26- 55418

We can see a very low RMSE value on the test data than the training data. From the plot we see that the predictors are definitely not identical to the actual test data, however it follows the trend - which is good news!

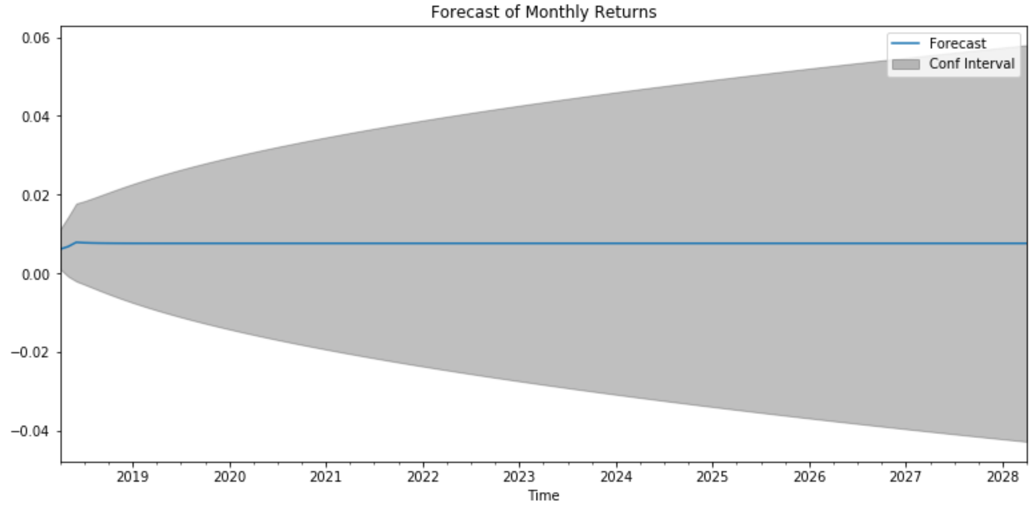


Figure 27- 55418 Zip Code Forcasting

```
Total expected return in 1 year: 9.33%
Total expected return in 3 years: 31.13%
Total expected return in 5 year: 57.27%
Total expected return in 10 years: 149.65%
```

Based on the results of this zip code for Minneapolis, MN we can expect a total long-term return of 149.65%. Overall, this is a good investment, and better than the first zip code.

Zip Code : 29412

Location: Charleston, SC

For the first analysis we reviewed the 29412-zip code for Charleston, SC. Figure 22 shows the ACR and PACF graphs for this zip. Based on the results of the ACF and PACF plots we can see that that the ACF is tailing off while the PACF spiked at about 43, and then started to tail off again. It is likely that we will have the AR and MA parameters in the final model.

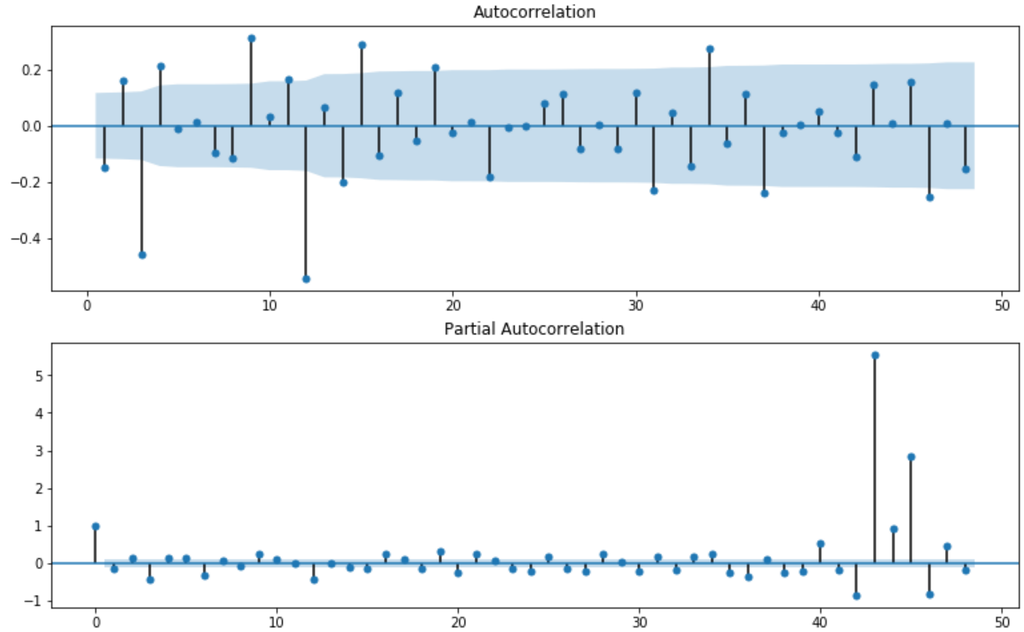


Figure 28 - Zip Code 29412

We can see from Figure 28, above, that the ACF plot tails off, and therefore, can likely end up with AR and MA parameters in the final model. However, the PACF cuts off after second lag.

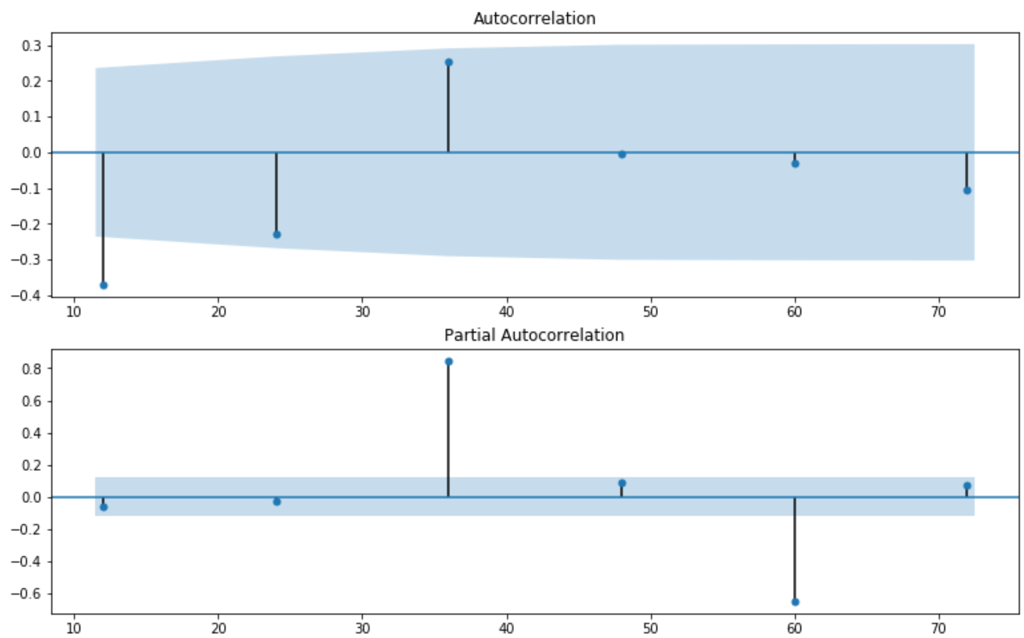


Figure 29 - Zip 29412 Seasonal Plots

From the seasoned plot of the ACF, we can see that there is a statistically significant autocorrelation on lag 12. Since there is nothing repeated afterwards, there might not be seasonality n the data.

```
                               SARIMAX Results
==============================================================================
Dep. Variable:                    ret   No. Observations:                  249
Model:               SARIMAX(0, 1, 4)   Log Likelihood                1193.822
Date:                Sun, 23 Feb 2020   AIC                          -2377.643
Time:                        18:34:50   BIC                          -2360.076
Sample:                    05-01-1996   HQIC                         -2370.571
                         - 01-01-2017
Covariance Type:                  opg
==============================================================================
                 coef    std err          z      P>|z|      [0.025      0.975]
------------------------------------------------------------------------------
ma.L1          0.0598      0.070      0.849      0.396      -0.078       0.198
ma.L2          0.1440      0.063      2.281      0.023       0.020       0.268
ma.L3         -0.4815      0.062     -7.723      0.000      -0.604      -0.359
ma.L4          0.1638      0.066      2.484      0.013       0.035       0.293
sigma2       3.82e-06   3.17e-07     12.045      0.000      3.2e-06    4.44e-06
===================================================================================
Ljung-Box (Q):                      177.41   Jarque-Bera (JB):                 2.05
Prob(Q):                              0.00   Prob(JB):                         0.36
Heteroskedasticity (H):               1.58   Skew:                             0.01
Prob(H) (two-sided):                  0.04   Kurtosis:                         3.44
===================================================================================

Warnings:
[1] Covariance matrix calculated using the outer product of gradients (complex-step).
```

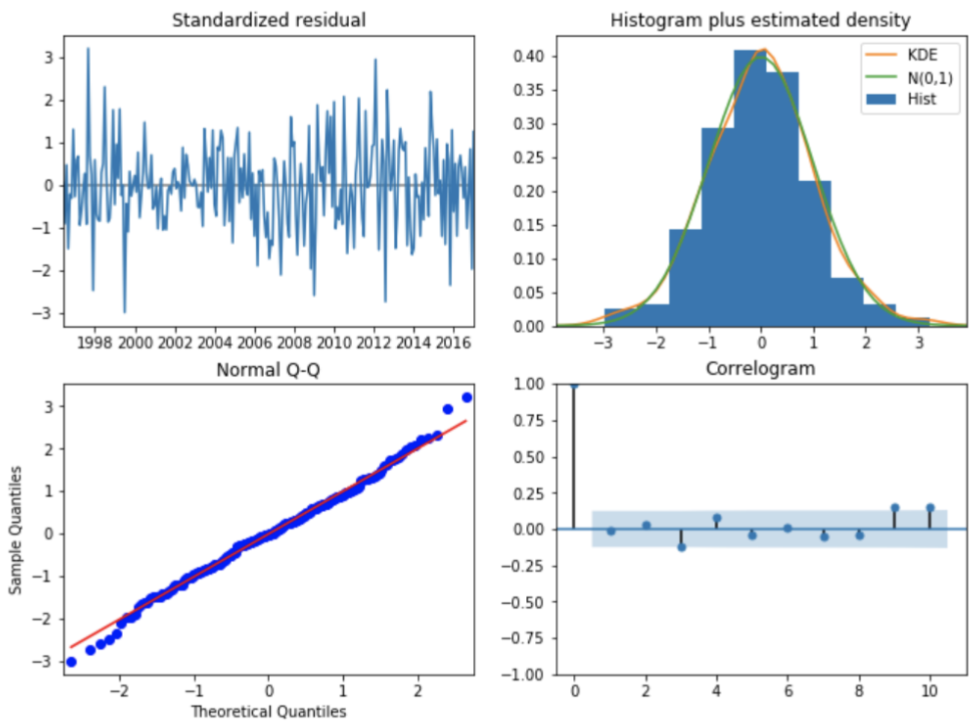


Figure 30 - Fitted SARIMA 39412

We can see based on the results of the SARIMA model above that this is the worst fitting model we have tested so far. The residuals are not normally distributed and are slightly autocorrelated, as you can see by the low p-values of the Ljung-Box tests. However, the p-value of the JB test is not low and the model has heteroskedasticity. Therefore, our model does not capture all of the signals properly and still needs to be improved further.

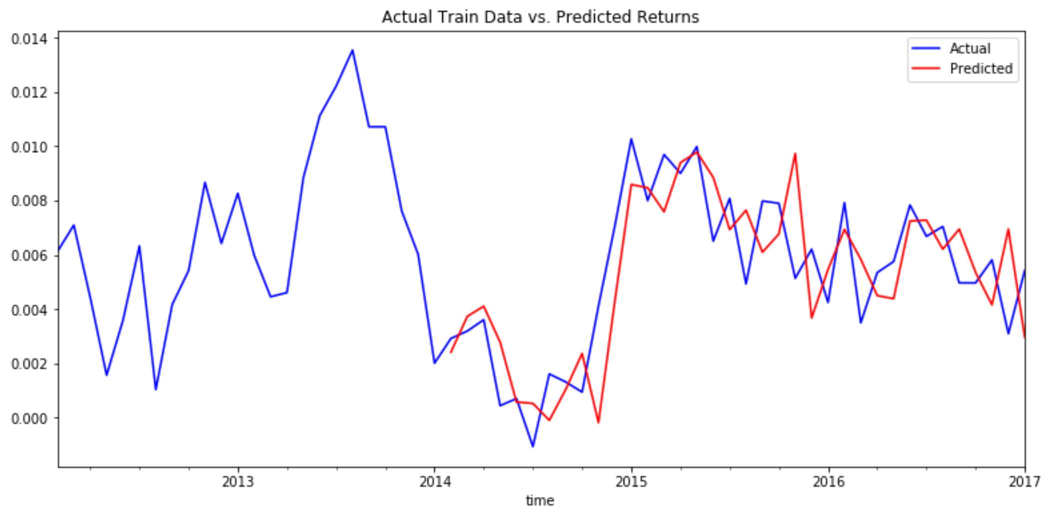


Figure 31- Train vs Predicted

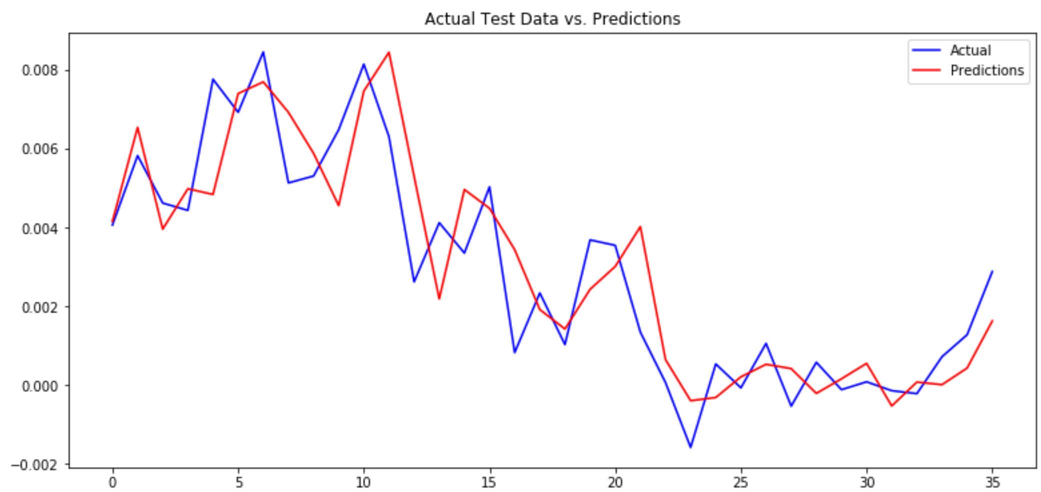


Figure 32 Test vs Predicted

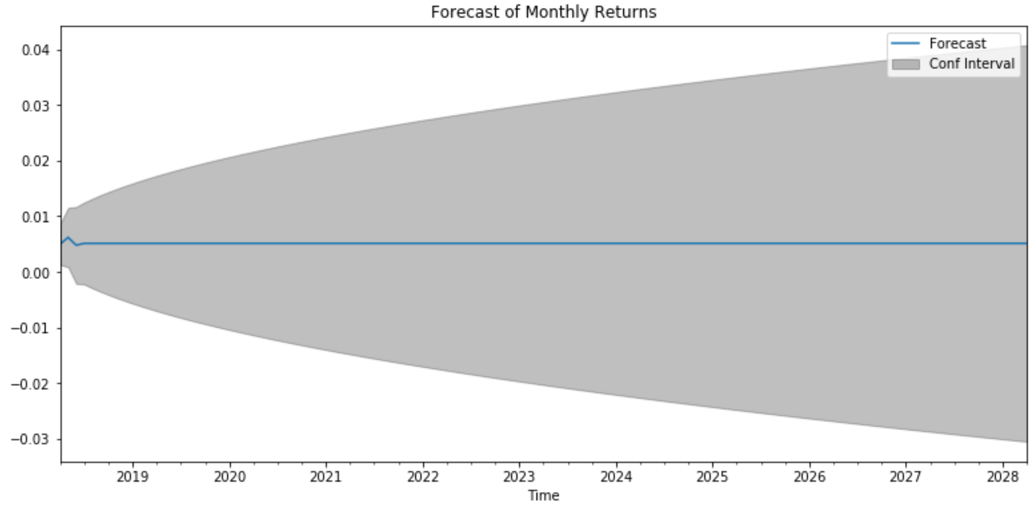


Figure 33- Forcasting 29412 Zip

```
Total expected return in 1 year: 6.37%
Total expected return in 3 years: 20.21%
Total expected return in 5 year: 35.85%
Total expected return in 10 years: 85.38%
```

Based on the results of this zip code for Charleston, SC we can expect a total long-term return of 85.38%. Overall, this is a good investment, and better than the first zip code.

Zip Code : 55104
Location: Saint Paul, MN

For the first analysis we reviewed the 55104-zip code for Saint Paul, MN. Figure 34 shows the ACR and PACF graphs for this zip. Based on the results of the ACF and PACF we can see that the ACF is tailing off - and therefore, we are quite confident that the MA(q) parameter is 0. We can also expect there to be a statistically significant values in the PACF after lag 43 which might entail a full ARMA model with p and q parameters.

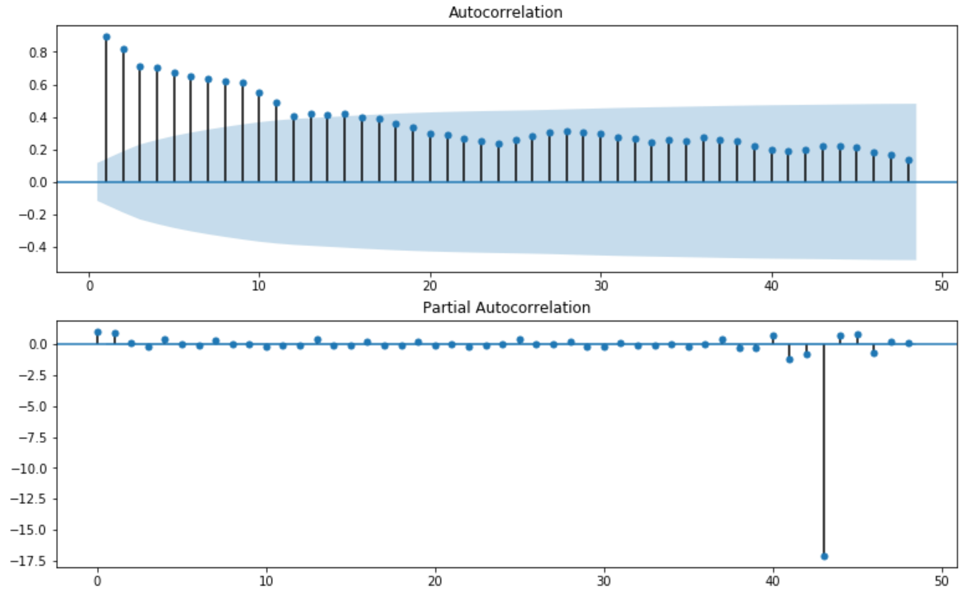


Figure 34 - 55104 Zip Code ACF

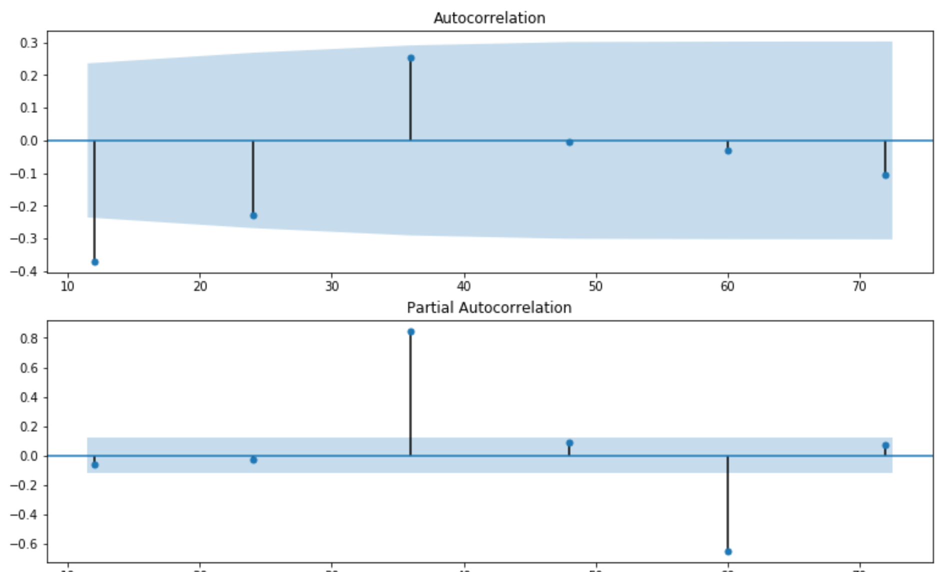


Figure 35 - 55104 Zip code

Initially, we would expect a 12-month seasonality since we can see statistically significant peaks in the ACF plots at around lag 12. But since there is no repetition in other lags, there is likely no seasonality in the results.

```
                               SARIMAX Results
==============================================================================
Dep. Variable:                    ret   No. Observations:                  249
Model:               SARIMAX(1, 1, 4)   Log Likelihood                1107.420
Date:                Sun, 23 Feb 2020   AIC                          -2202.840
Time:                        18:40:13   BIC                          -2181.760
Sample:                    05-01-1996   HQIC                         -2194.354
                         - 01-01-2017
Covariance Type:                  opg
==============================================================================
                 coef    std err          z      P>|z|      [0.025      0.975]
------------------------------------------------------------------------------
ar.L1         -0.0100      0.494     -0.020      0.984      -0.978       0.958
ma.L1         -0.0499      0.509     -0.098      0.922      -1.048       0.948
ma.L2          0.0064      0.055      0.115      0.908      -0.102       0.114
ma.L3         -0.5552      0.056     -9.835      0.000      -0.666      -0.445
ma.L4          0.1305      0.276      0.473      0.636      -0.410       0.671
sigma2      7.693e-06   6.65e-07     11.569      0.000    6.39e-06       9e-06
===================================================================================
Ljung-Box (Q):                       78.32   Jarque-Bera (JB):                 1.71
Prob(Q):                              0.00   Prob(JB):                         0.43
Heteroskedasticity (H):               2.40   Skew:                            -0.10
Prob(H) (two-sided):                  0.00   Kurtosis:                         3.36
===================================================================================

Warnings:
[1] Covariance matrix calculated using the outer product of gradients (complex-step).
```

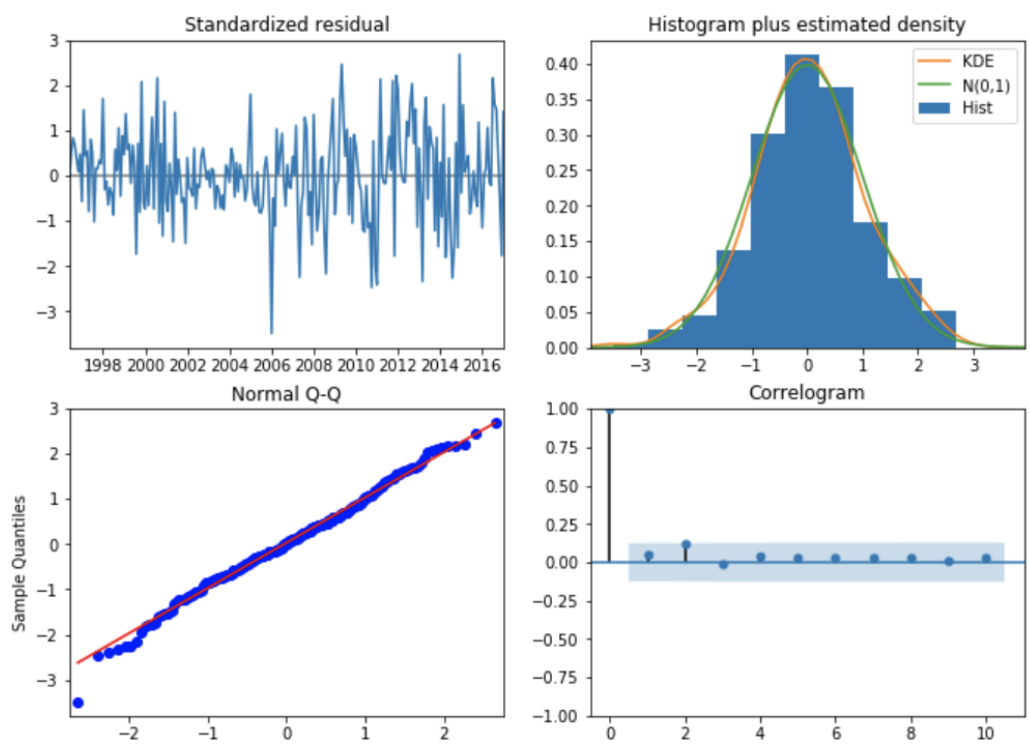


Figure 36 SARIMA Results – 55104

We can see that after fitting this model, from our above results ,this is a very good fitted model. We expect to have good predicted returns later. We can see from the histogram and Q-Q plot graphs above that the residuals are normally distributed. Based on the results of the Ljung-Box test, we can reject the null hypothesis in favor of the alternate hypothesis which shows that the data is autocorrelated at a 95% confidence level.

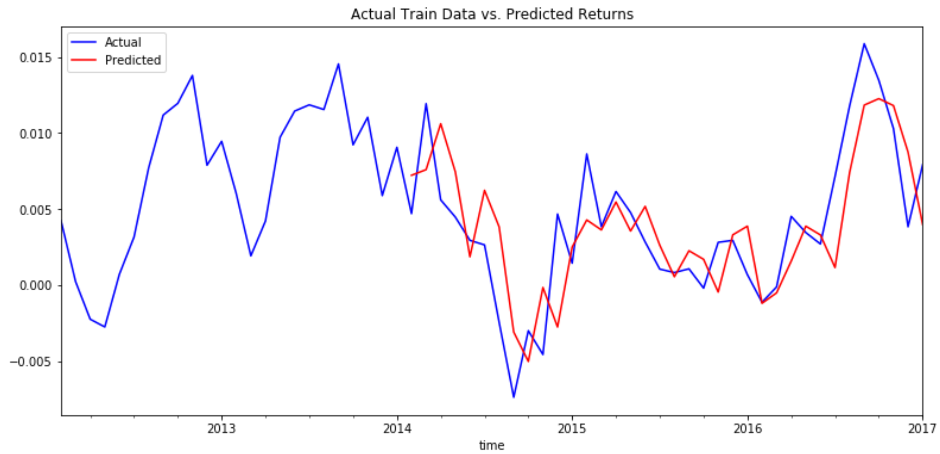


Figure 37

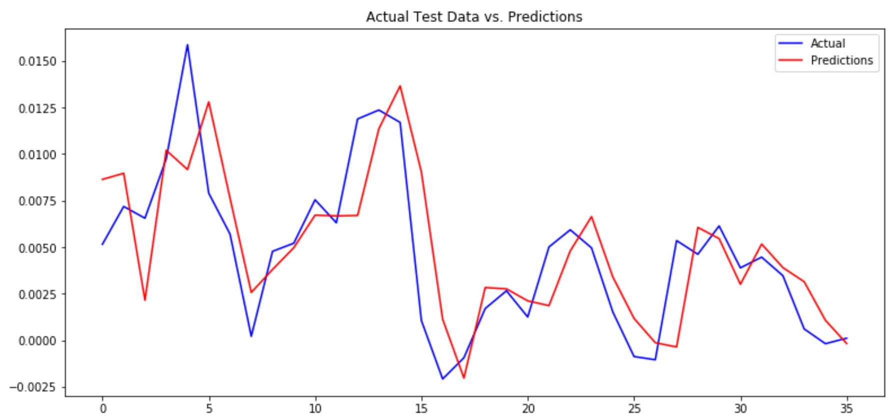


Figure 38

It is interesting the RMSE test data has a barely smaller value than the train data. However, we can see from the plots that the prediction pattern closely follows that hof the actual data - which is good news for our return forecast

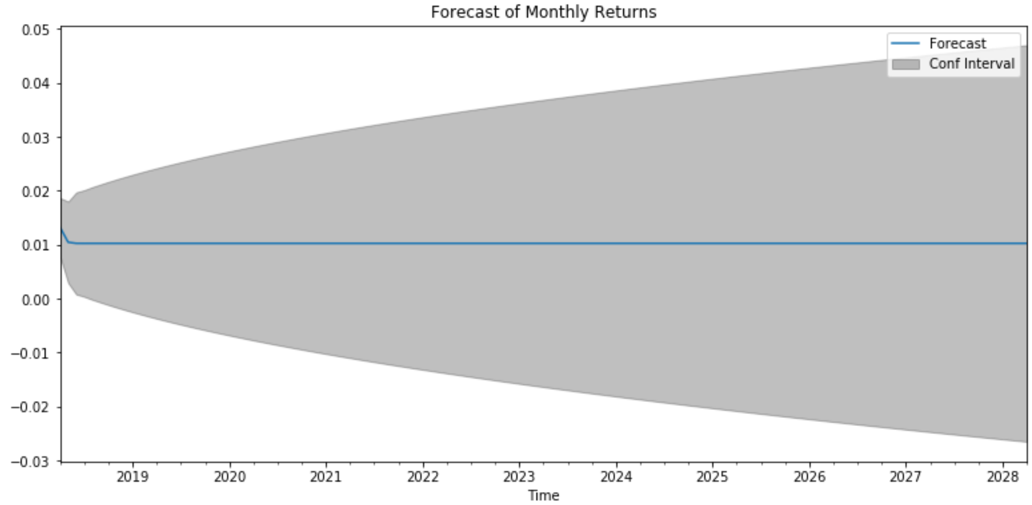


Figure 39 Forcasting 55104 Zip

```
Total expected return in 1 year: 13.36%
Total expected return in 3 years: 44.77%
Total expected return in 5 year: 84.87%
Total expected return in 10 years: 244.19%
```

Based on the results of this zip code for Saint Paul, MN we can expect a total long-term return of 244.19%. Overall, this is the best investment yet.

Zip Code : 63108

Location: Saint Louis, MN

For the first analysis we reviewed the 63108-zip code for Saint Louis, MN. Figure 34 shows the ACR and PACF graphs for this zip. Based on the results of the ACF and PACF we can see that the ACF is tailing off - and therefore, we are quite confident that the MA(q) parameter is 0. We can also expect there to be a statistically significant values in the PACF after lag 43 which might entail a full ARMA model with p and q parameters.

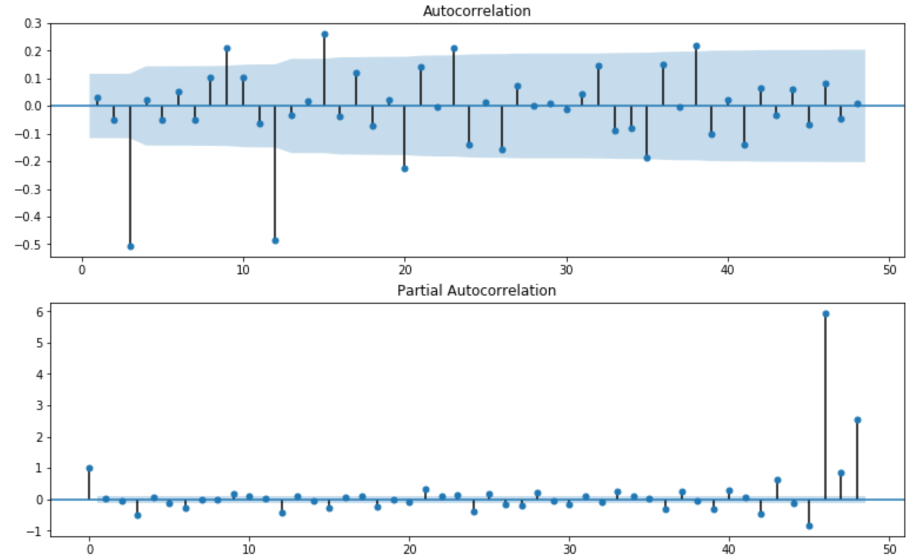


Figure 40 - 63108 Zip Code

There may be a 4-5-month seasonality given that there is a statistically significant peak in the ACF plot around lag 4-5. Because we can see another lag around 11, we can expect that there is seasonality in the results.

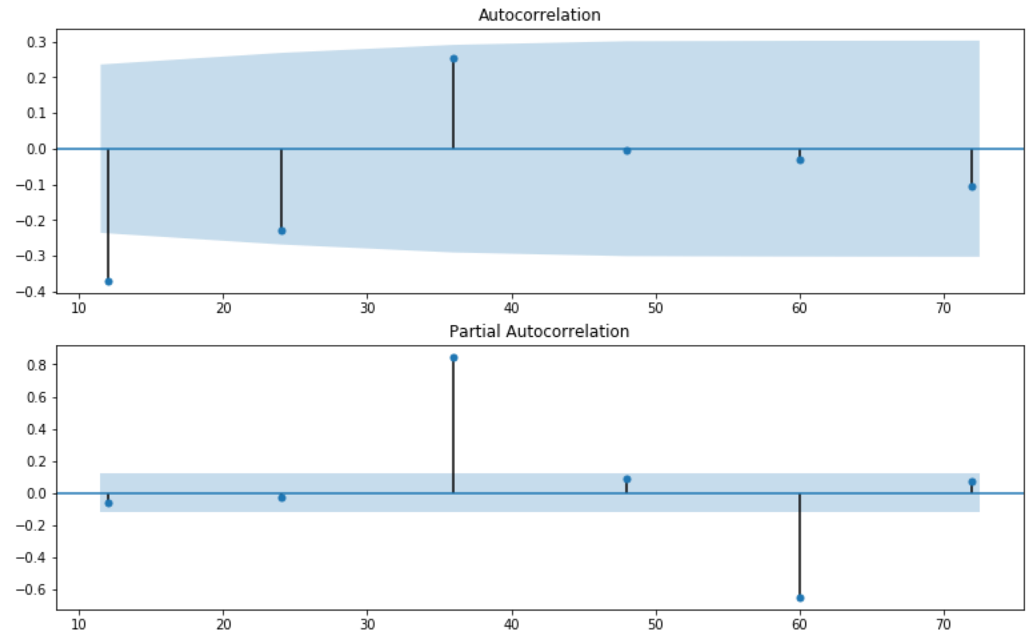


Figure 41 - ZipCode 62108

```
                               SARIMAX Results
==============================================================================
Dep. Variable:                    ret   No. Observations:                  249
Model:               SARIMAX(3, 1, 0)   Log Likelihood                 961.900
Date:                Sun, 23 Feb 2020   AIC                          -1915.800
Time:                        18:45:17   BIC                          -1901.747
Sample:                    05-01-1996   HQIC                         -1910.143
                         - 01-01-2017
Covariance Type:                  opg
==============================================================================
                 coef    std err          z      P>|z|      [0.025      0.975]
------------------------------------------------------------------------------
ar.L1          0.0136      0.051      0.267      0.790      -0.086       0.113
ar.L2         -0.0197      0.054     -0.362      0.717      -0.126       0.087
ar.L3         -0.5161      0.047    -10.895      0.000      -0.609      -0.423
sigma2      2.491e-05   2.41e-06     10.339      0.000    2.02e-05    2.96e-05
===================================================================================
Ljung-Box (Q):                      142.76   Jarque-Bera (JB):                 0.75
Prob(Q):                              0.00   Prob(JB):                         0.69
Heteroskedasticity (H):               1.19   Skew:                             0.02
Prob(H) (two-sided):                  0.42   Kurtosis:                         2.73
===================================================================================

Warnings:
[1] Covariance matrix calculated using the outer product of gradients (complex-step).
```

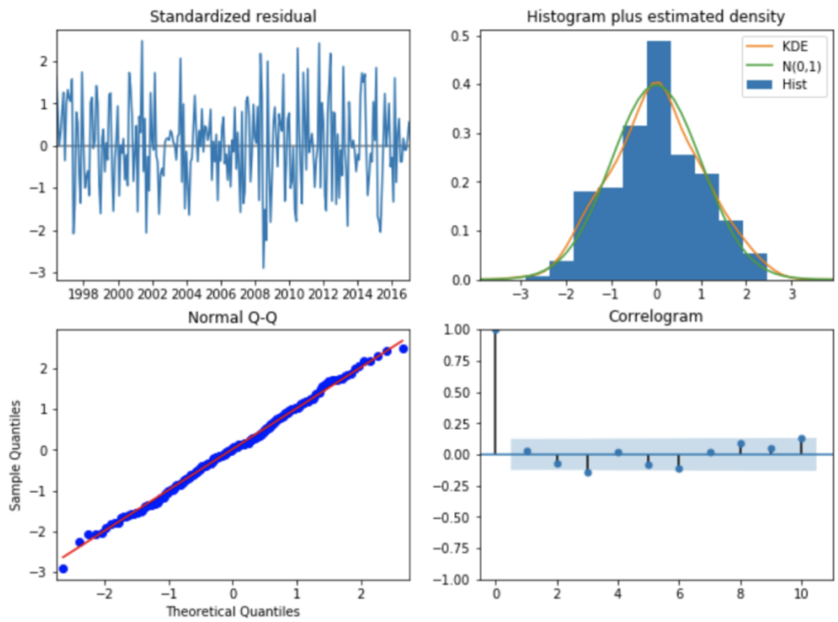


Figure 42- SARIMA Model - 63108

The SARIMA model is able to capture some of the signals in the data. The residuals are somwhat normally distributed and therefore, they are not autocorrelated. The p-value for Jarque-Bera (JB) test is high, therefore the residuals are very likely just noise.

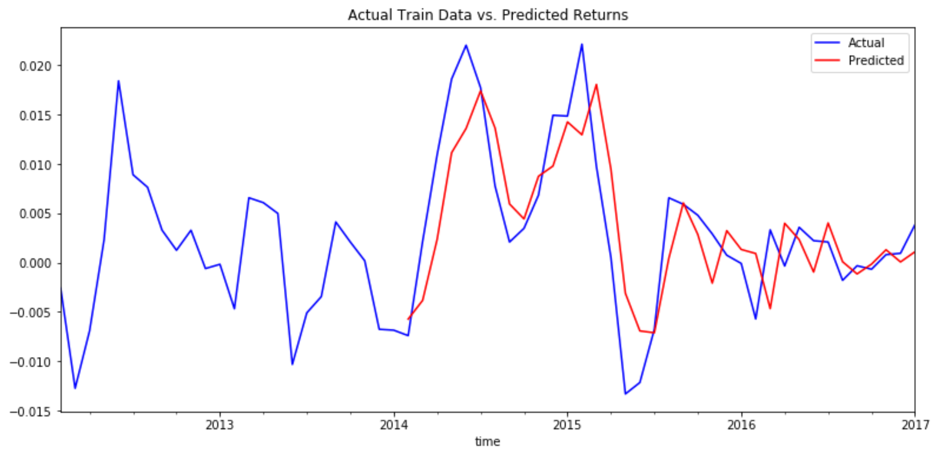


Figure 43

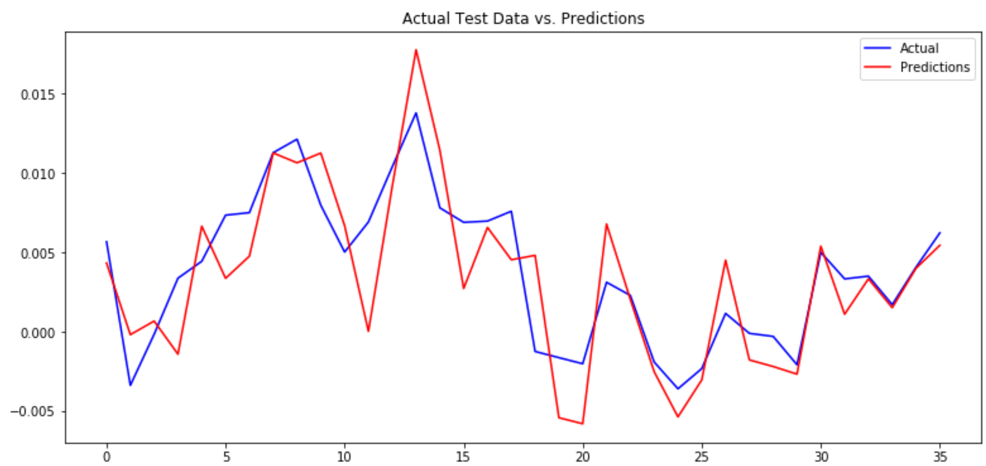
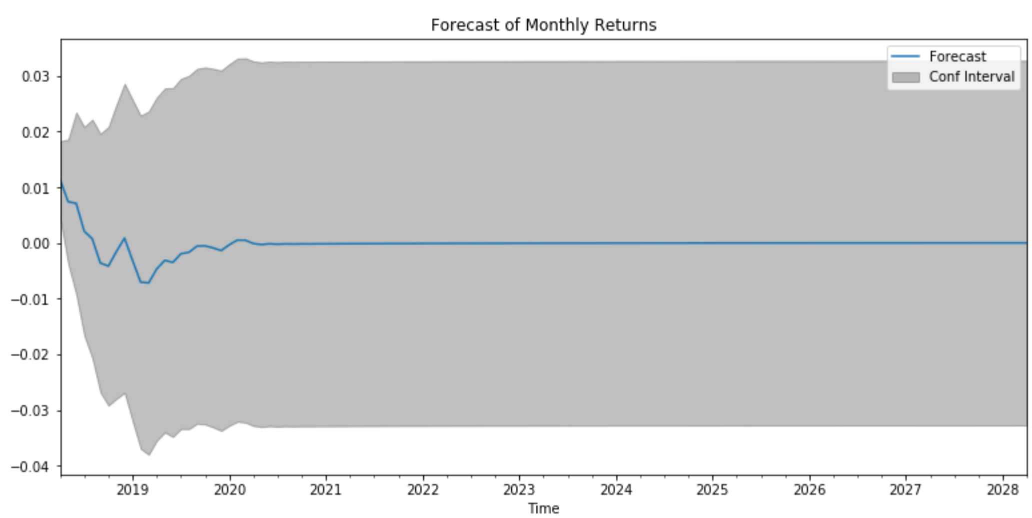


Figure 44




```
Total expected return in 1 year: 0.26%
Total expected return in 3 years: -1.73%
Total expected return in 5 year: -1.96%
Total expected return in 10 years: -2.1%
```

Based on the results of this zip code for Saint Louis, MN we can expect a total long-term return of -2.1%, which is terrible.

Results and Conclusion:

After performing time series analysis on the 5 zip codes and forecasting total returns for up to ten years, I recommend for the company to invest in the following 3 zip codes:

- 55104: Saint Paul, MN (244.19% 10-year total return)
- 55418: Minneapolis, MN (149.65% 10-year total return)
- 29412 : Charleston, SC (85.38% 10-year total return)

This is because, through these three locations, the SREIT will have the highest long term and short-term return.

References:

UN Report on Population movements to cities:
https://www.un.org/development/desa/en/news/population/2018-world-urbanization-prospects.html

Understanding ACF and PACF Tails:
https://stats.stackexchange.com/questions/241914/terms-cut-off-and-tail-off-about-acf-pacf-functions

Investing In Real Estate:
https://www.investopedia.com/ask/answers/031715/what-can-coefficient-variation-cov-tell-investors-about-investments-volatility.asp