Sammy Pardes
IST 718
Week Lab 6
8/20/2021

Introduction

The Syracuse Real Estate Investment Trust (SREIT), is looking for a new investment opportunity. Being based in Syracuse, they are looking to invest in the state of New York. Specifically, they are looking at real estate opportunities in New York City.

The SREIT would like to only consider zip codes where at least 94% of the population has Internet access. In the past year and a half, it has become overwhelmingly apparent how important Internet access is for the modern way of life. According to realtorparty.com, several studies have been conducted that indicate increased Internet access positively impacts housing prices and may increase the value of a home.

About the Data and Data Cleaning

The initial dataset pulled from Zillow contained 30,464 observations of 300 variables. The majority of the columns were date values between 1996 and 2020.

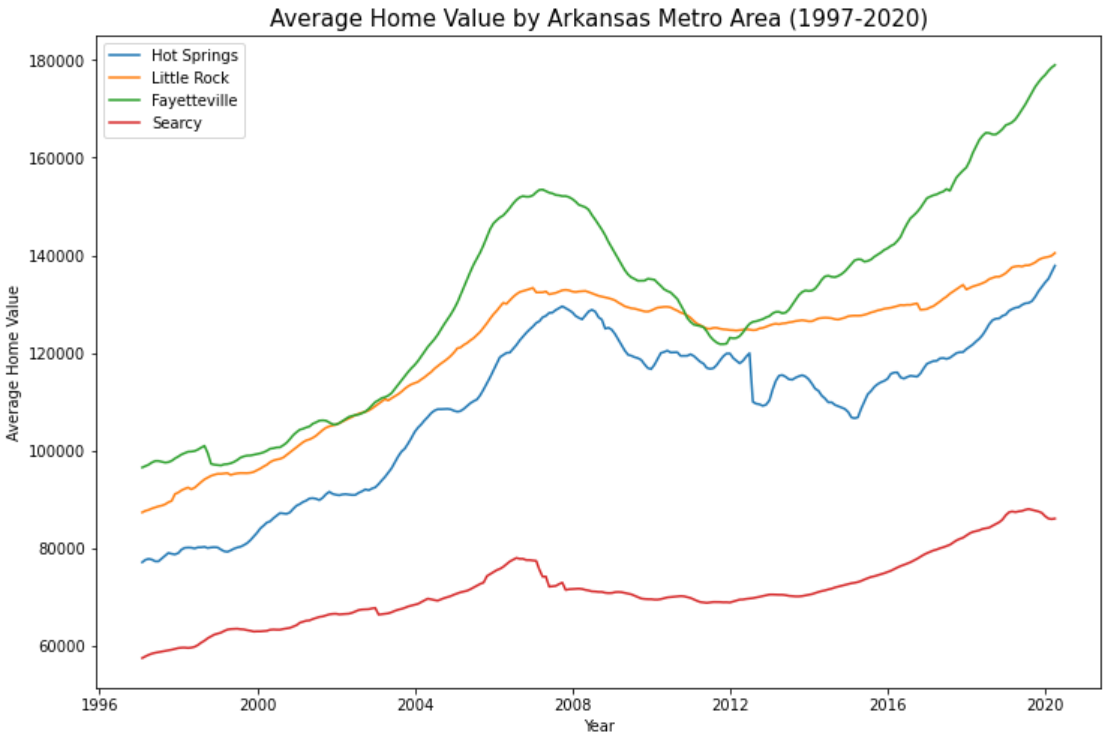
The melt() function was used to combine all of the dates into one column. The dates were also converted to datetime data types with the to_datetime() function. This left a data frame with 8,865,024 rows and 11 columns.

Part I: Arkansas

To begin understanding the Zillow data, an initial analysis was performed on a subset of housing prices in Arkansas from 1997 to present day. Four specific metro areas of Arkansas were extracted and put into their own data frames. These included Hot Springs, Little Rock, Fayetteville, and Searcy.

For each metro area, the data frames were grouped by the Date variable and the mean of the HomeValue was calculated for each date. The cleansed data frames for each metro area contained only two columns; date and respective average home value.

Finally, the seaborn package was used to create a time series plot of all four metro areas. The visualization below shows the change in average home value from 1997-2020 for Hot Springs, Little Rock, Fayetteville, and Searcy. While each area saw decreasing home values around the 2008 financial crisis, Searcy seemed to be least negatively impacted. Although, the average home value in Searcy is consistently much lower than the other three areas. Since about 2012, the average home value in Fayetteville has steadily increased and has surpassed Little Rock and Hot Springs by a significant margin.



Part II: SREIT Analysis

Housing data from 1997-2018 was extracted from the cleansed and melted data set described above. From this, rows where the State variable was NY were saved into a new data frame. Then, only those in the New York-Newark-Jersey City metro area were retained for analysis. There were 524 unique zip codes in the NYC metro area.

Since the SREIT determined that only areas with high Internet access were candidates for investing, broadband data from data.cityofnewyork.us was used to pull zip codes where the percentage of the population without Internet access was less than or equal to 6%.

The Date column was renamed to “ds” and HomeValue was updated to “y” to be compatible with Prophet. All other columns were dropped. The eligible zip codes were then matched with the NYC data. This left 12 zip codes for the SREIT to potentially invest in.

Data before 2018 was grouped by the zip code and put into a NDFrame for training. 2018 housing information was also grouped and reserved for testing. Each “group” contained 252 observations in the respective zip code, one for every month from January 1997 - December 2017.

Using a for loop, each grouped DF in the high Internet access data set was fitted with a Prophet model. For each zip code, a forecast was created that predicted 12 additional months (for 2018). A plot was created for each zip code in the training data showing historical and predicted

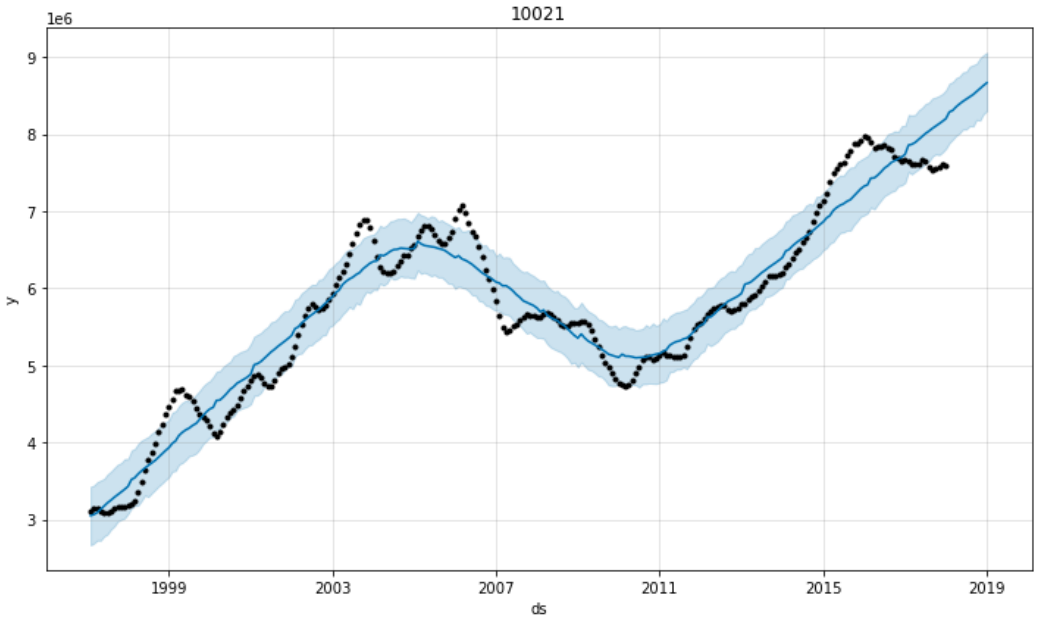
home values. A new data frame was created containing the y_hat value for all the zip codes for each month 1997-2018.

Results and Recommendations

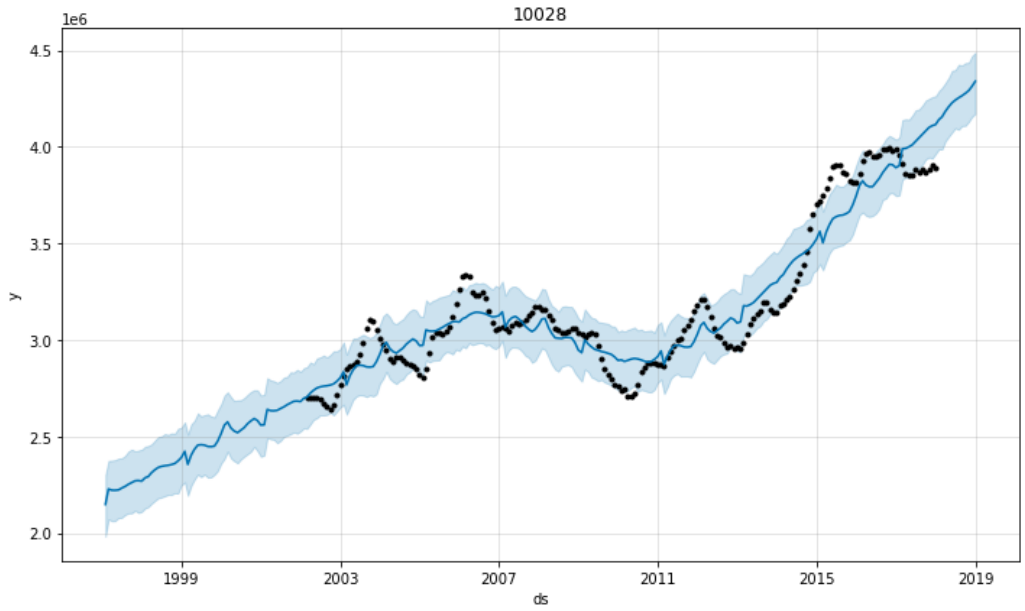
The SREIT decided to invest in the three NYC zip codes with high rates of Internet access that the Prophet model predicted would be the greatest at the end of 2018. These zip codes are 10011, 10021, and 10028. These zip codes encompass the Chelsea, Upper East Side, and Yorkville neighborhoods of New York City.

The SREIT also chose these three zips because they had sufficient historical data and the top three largest average differences in home value each month. This meant that, on average, these three zip codes increased each month by the most amount of money.

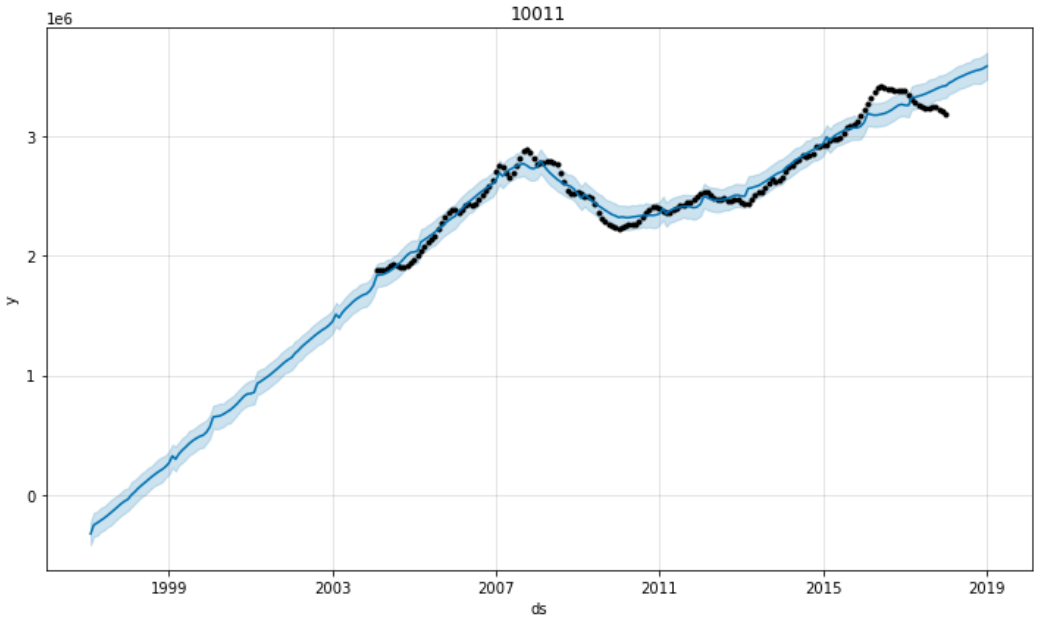
At the end of 2018, the Prophet model predicted that homes in the Upper East Side would be worth an average of $8,666,019.



In Yorkville the average home value was predicted to reach $4,339,353 by December 2018.



The third highest predicted average home value zip code was Chelsea. There, the Prophet estimated homes would reach an average of $3,581,183 nearing the end of 2018.



Conclusion

For all three zip codes, the home values were a bit overestimated. The RMSE in December 2018 for all 12 zip codes was $476,933.

Although the forecasts average out to predict the general trend of the home values, the Prophet models tended to smooth the data and missed capturing the peaks and valleys that occured over time.

Depending on the risk level SREIT is willing to take on, these zip codes may provide viable investment opportunities. I recommend that the SREIT use this analysis as a guideline for investing but they should ultimately consult a real estate expert to make their final decision.