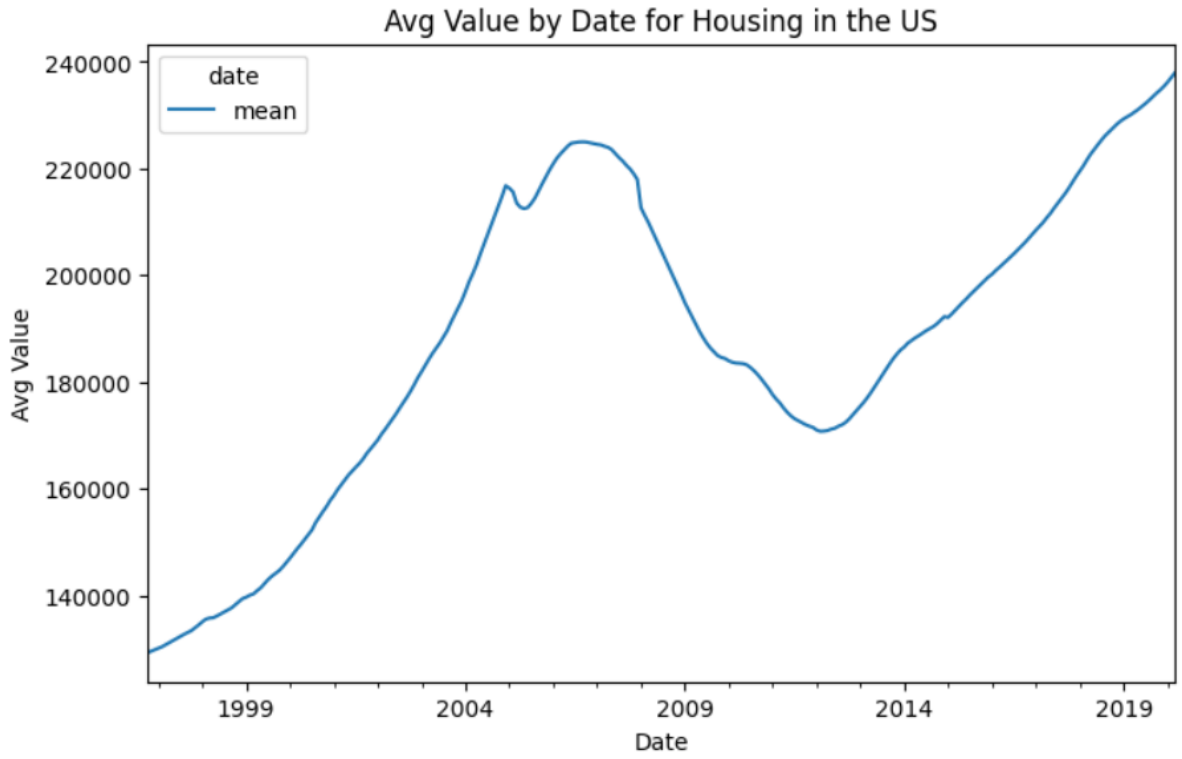
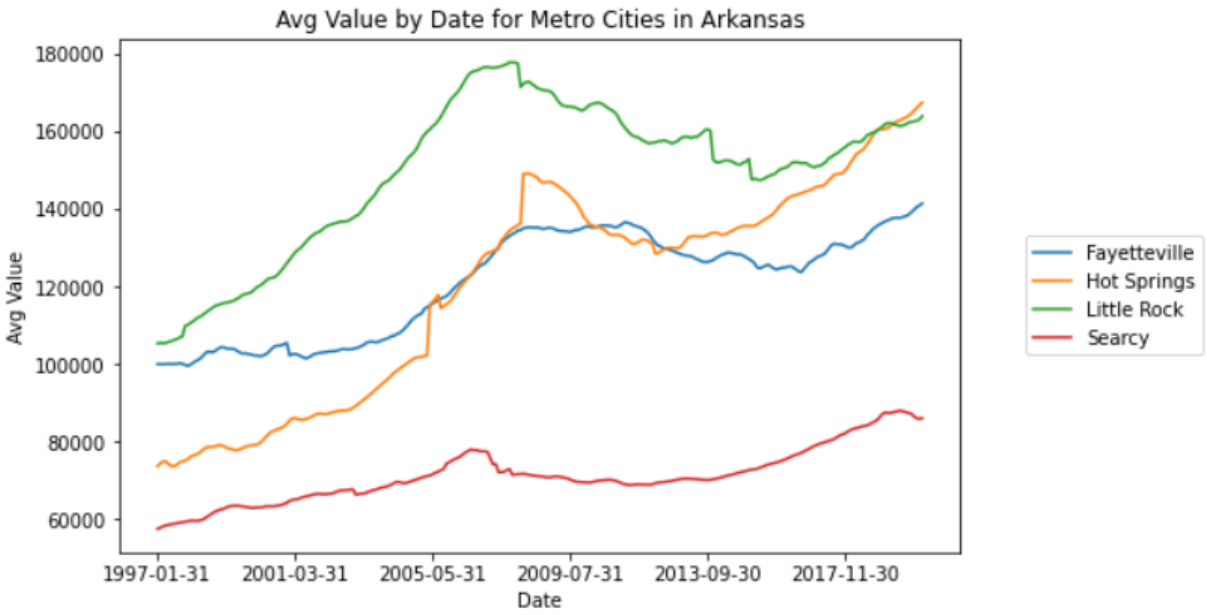
Rohini Shrivastava

Lab Week 6

We have a dataset taken from Zillow with 300 columns (291 of which are dates) and 30,464 rows. It covers 30,464 zip codes from 14,862 cities and 51 states (including Washington DC). We created a plot to see how the overall prices for houses has changed. Due to the 2008 stock market crash, many houses started to become foreclosed. It caused many distressed sales to occur (https://money.cnn.com/2012/03/27/real_estate/home-prices/index.htm).

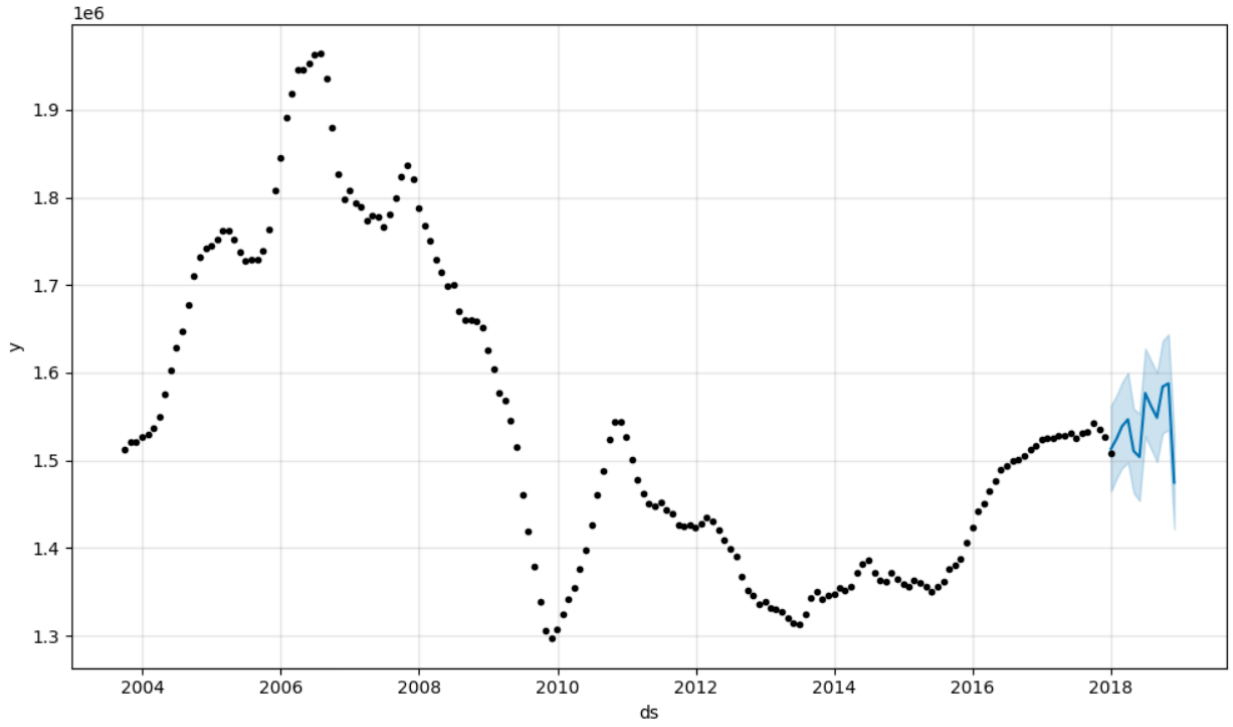


We calculated the average of the housing prices for all the zip codes located in Searcy, Hot Springs, Fayetteville and Little Rock. From there we created a time series plot showing how the prices for the housing in that area has changed since 2017.



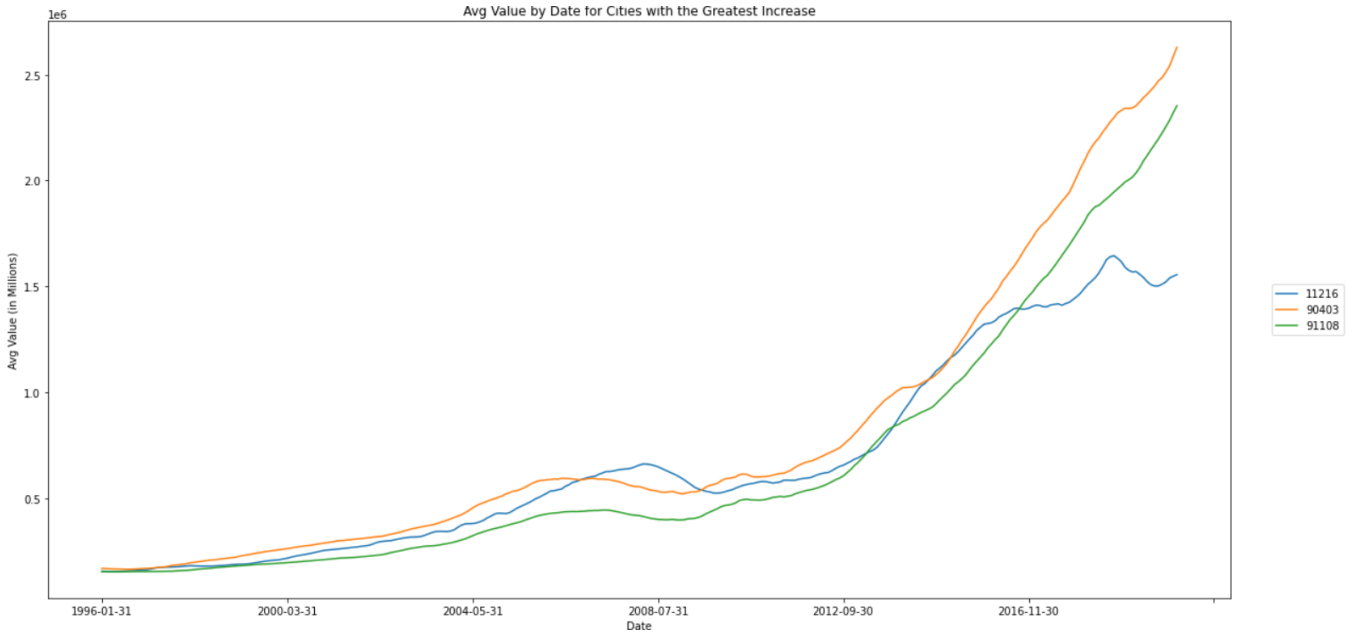
Unfortunately, the forecasting model I created did not work as expected when using Prophet when calculating it with the whole dataset, so we were only able to do it using the zip code 10025, which is for New York City, New York.

After running the model through prophet we graphed the data points as seen below. We then calculated the mean absolute error, root mean square error, mean square error, and R-Squared. Our MAE and RMSE was 72306.57 and 84,391.11 respectively. When comparing it to the average house price in New York City (which after calculating the 2018 prices seems to be 1,467,031.17), 72K-84K is not that high of a margin. However that is still too much for me so I don't believe this forecast is that great. Also our R-Squared is -0.20, showing that it's a really poor fit.



MAE	72306.566
MSE	7121589993.19
RMSE	84391.11
R-Squared	-20.58

We calculated the three zip codes with the greatest percent increase from March 2000 to March 2017. We calculated this by looking through the dataframe and calculating the difference between the price on 2017-03-31 and subtracting it from the price on 2000-03-31, and then dividing it by the price from 2000-03-31. We found 91108 had a 637% increase, 90403 had a 591% increase and 11216 had a 552% increase. 91108 is in Pasadena, CA, 90403 is in Santa Monica, CA, and 11216 is Brooklyn, NY.



For curiosities sake I also calculated the lowest price change using the same method. 36105 had a 75% decrease, 44049 had a 60.3% decrease, and 38112 with a 57.7% decrease. 36105 is in Montgomery, AL, 44049 is in Kipton, OH, and 38112 is in Memphis, TN.