Sana Khan

IST 718

Lab 2

Objective

The objective of this Lab is to use time series data to forecast which zip code is the best investment opportunity for the Syracuse Real Estate Investment Trust (SREIT). The data provided is Zillow housing data from 1996 – 2023. The dataset has the housing value for each month of the time frame for zip codes in the United States.

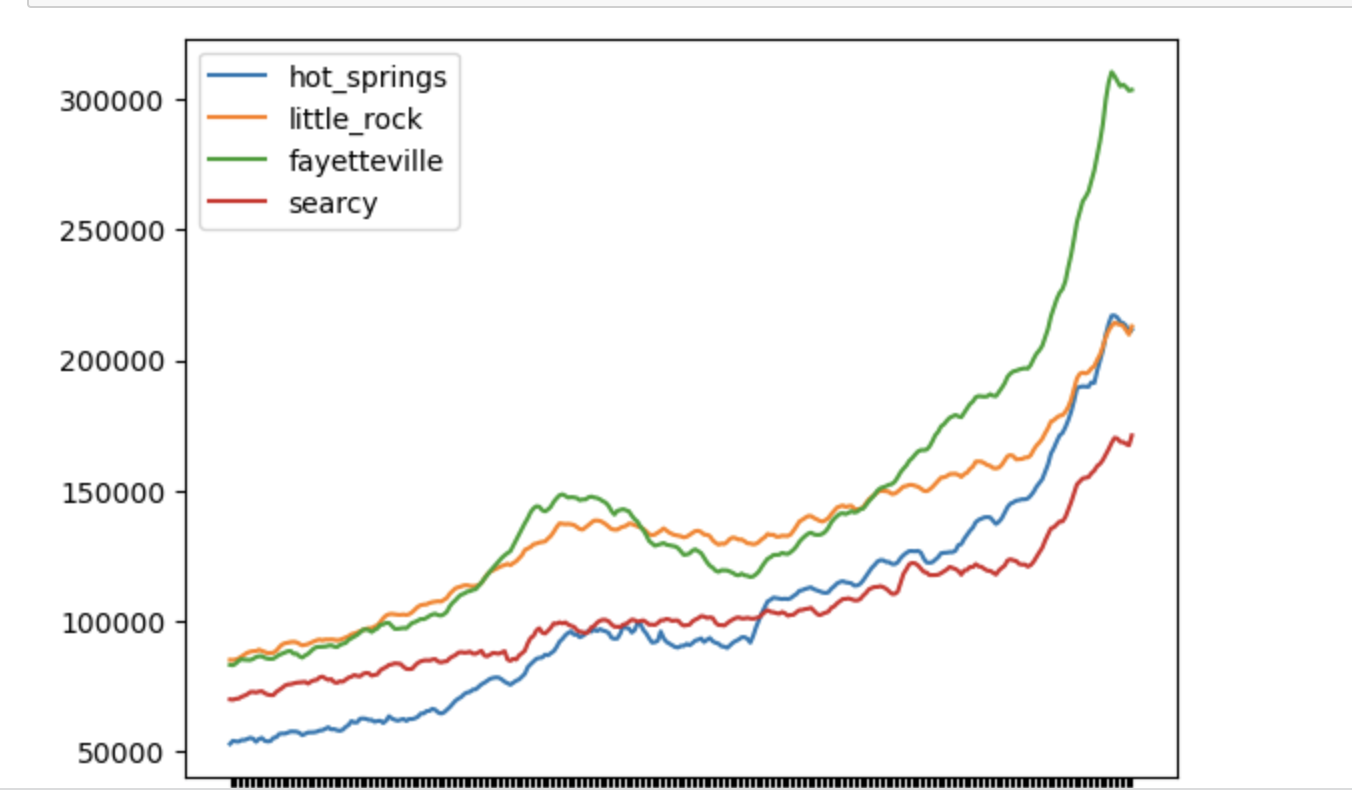
Data Cleansing

The initial data frame has 29,540 rows and 322 columns. The text columns are below:

|  |
| --- |
| * RegionID |
| * SizeRank |
| * RegionName |
| * RegionType |
| * StateName |
| * State |
| * City |
| * Metro |
| * CountyName |
|  |

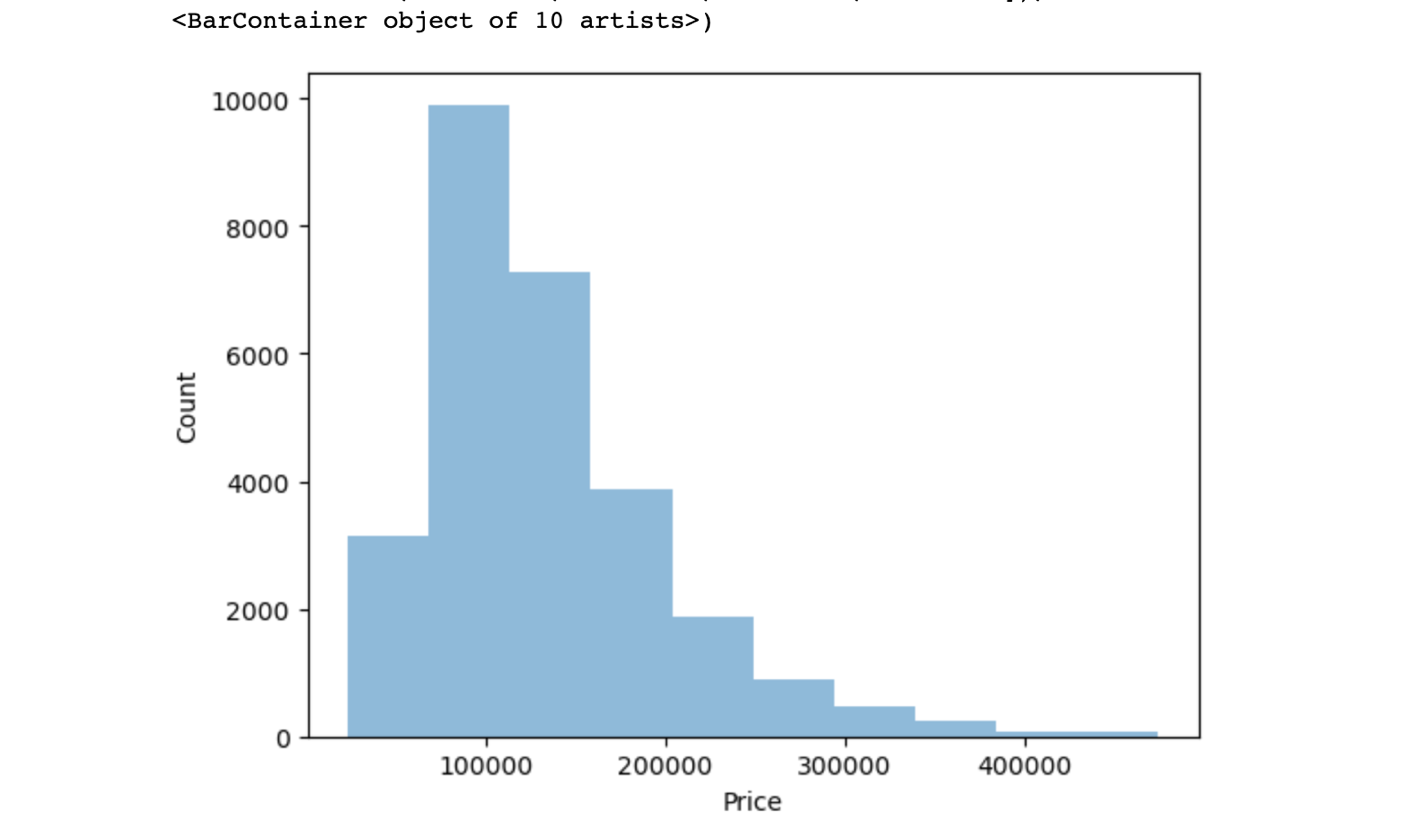
The remaining columns are the data columns, with the value of the home in the rows.

To prepare the data frame, any null values were dropped from the set. To plot the time series data for the Hot Springs, Little Rock, Fayetteville, Searcy metro areas, I subsetted the data based on the metro. I then took an average of the months and used that to create the following time series graph:

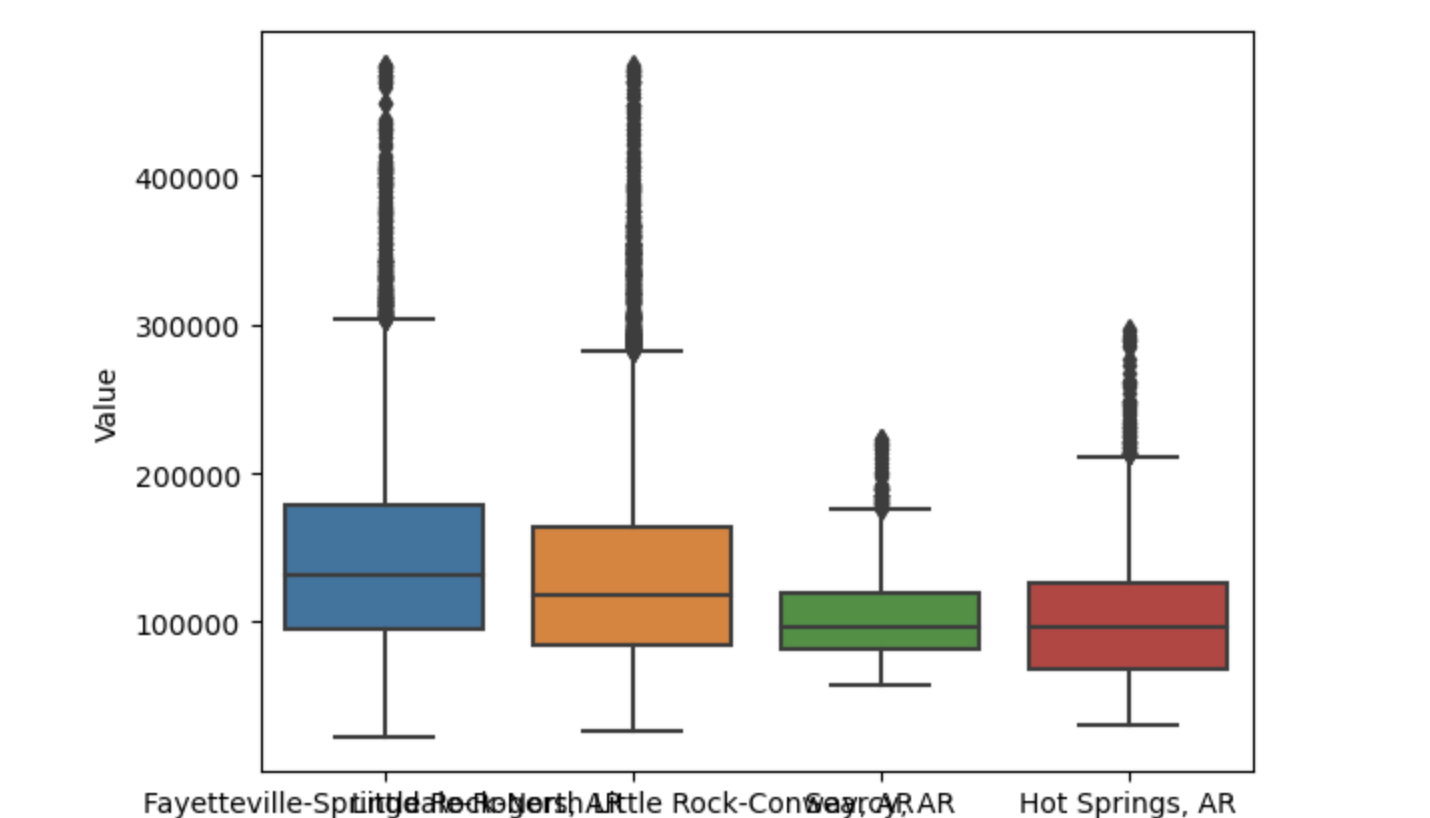


We can see that there is a rise in average home prices from 2004 – 2007 then a sharp decline which is due to the 2008 financial crisis, where home values were greatly impacted. We can also see that since the 2008 decline, home prices have been rising with the greatest increase in Fayetteville. There is also a huge jump in home prices after 2020, when demand for homes had skyrocketed, which meant the cost of homes had jumped as well. You can see the decline in 2023, which shows that home values have started to decrease after the pandemic boom.

I then transposed the dataframe and created a histogram to see the distribution of home prices.



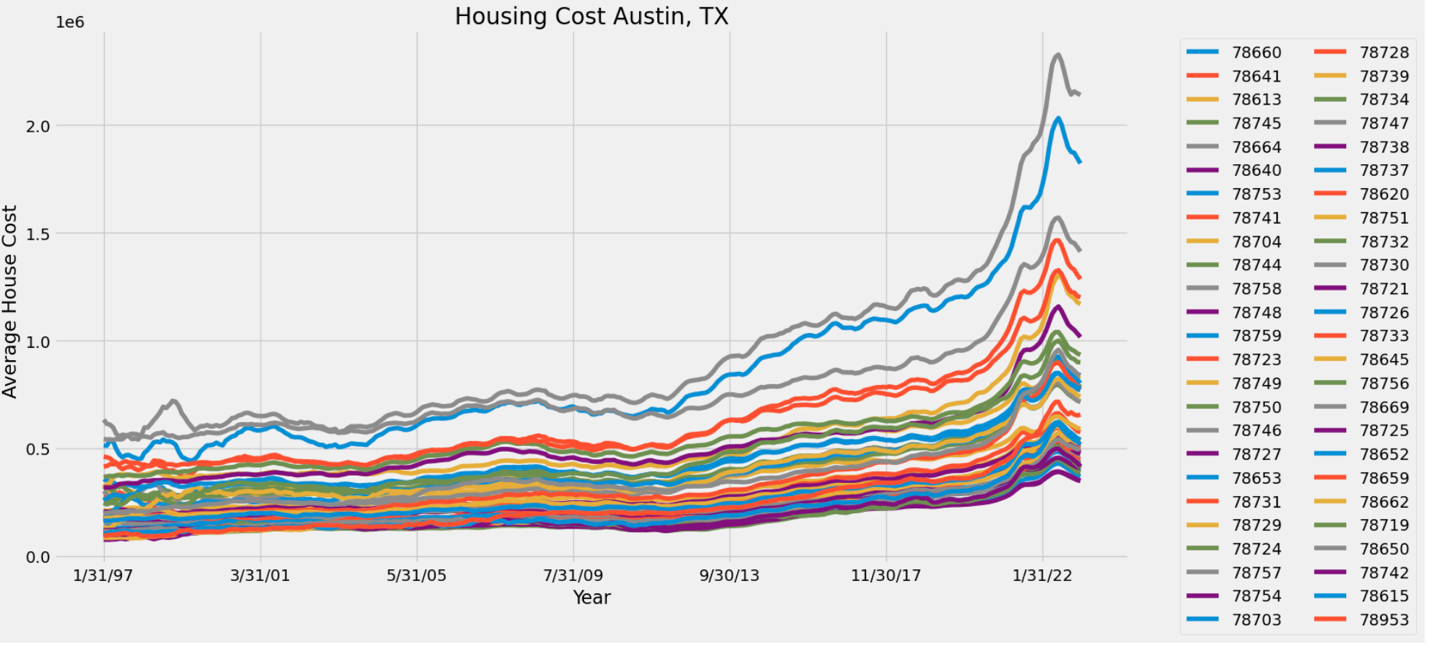
Based on the histogram, the median cost of a house is between 100,000 – 150,000. I then made a boxplot of the cost of homes based on the metro.



Based on the boxplot, Fayetteville and Little Rock had the widest variability in home prices, with the prices going up to half a million. Hot springs had less variability, with Searcy being the least. Searcy may be a good option to buy a home, but other factors such as crime, poverty and transportation should be looked at to determine if this is a good area to purchase a home.

To create the model I select Austin,TX as my city to potentially invest in. I decided to do this, because I live in Austin and am familiar with the market and have seen how much the cost of a home has increased here. I also choose to do this because when I tried to run the ARIMA/prophet model on the entire data set I ran out of RAM and my computer was unable to complete the action.

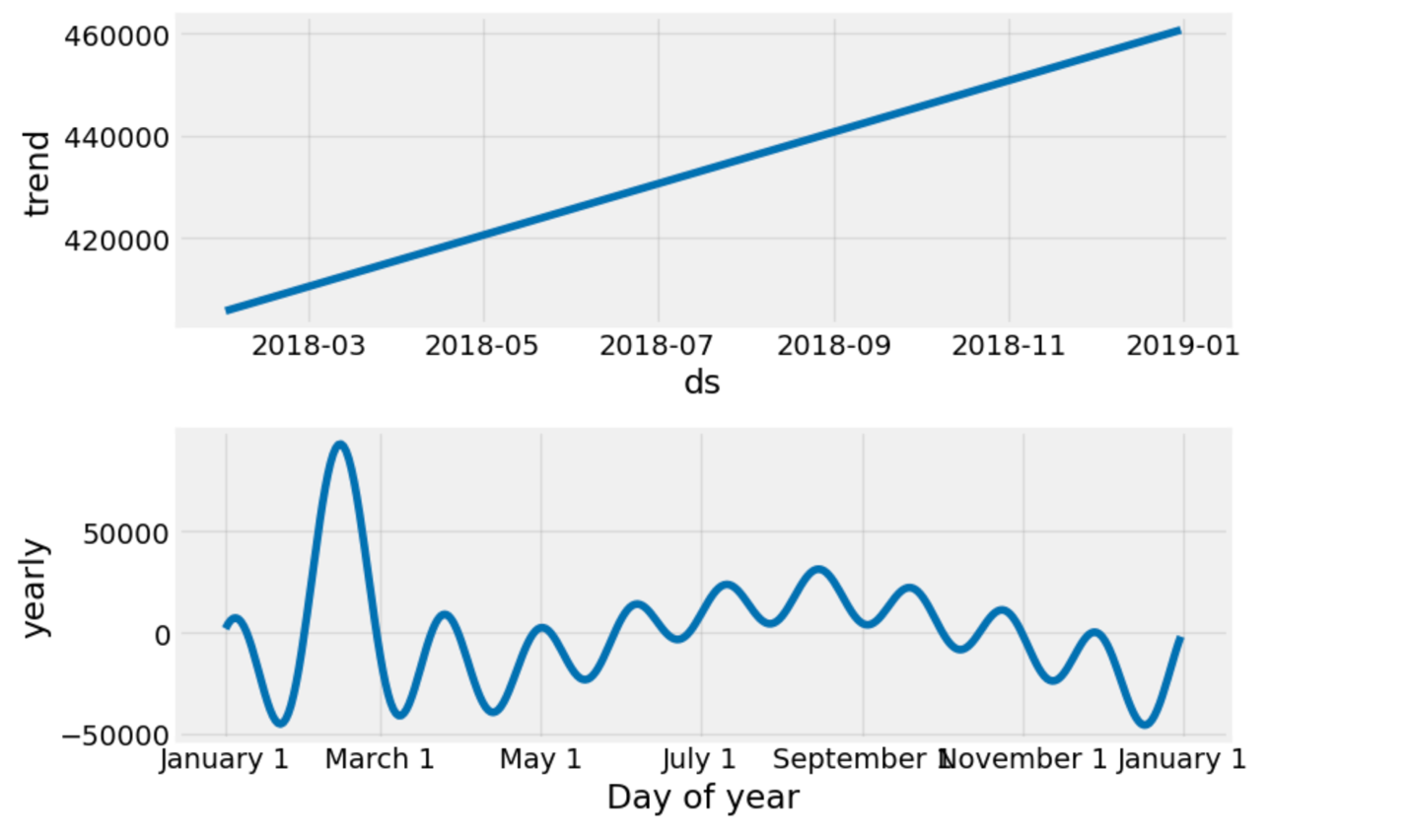
I initially plotted the zip codes in the Austin metro area, and I see there are similar trends across the zip codes.



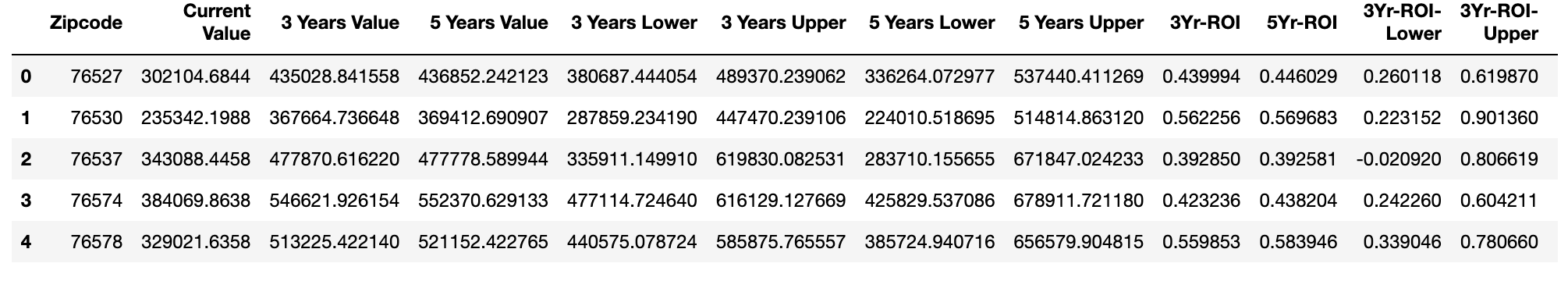
Austin data shows that the housing market was steady during the 2008 financial crisis but it shows a decline in prices in 2011. After 2011- 2012 we can see a steady increase in the price of homes with a sharp increase during the pandemic. This graph is a bit difficult to make sense of since the colors overlap.

I then split up the training data to be data from 1997 – 2017 and the testing data to be 2018.

I tried to use the Prophet Model to forecast home prices in 2018 however, this graph is also difficult to interpret. Prophet did show us that the overall trend is an increase in home prices for 2018, with a sharp increase in early spring. It looks like the best time to buy would be in the late fall or early winter.



I also used the ARIMA model to gather more insights. The model showed the following 5 zipcodes as the best to invest in:



Questions to answer:

* What technique/algorithm/decision process did you use to down sample?
  + I decided to use the ARIMA model because the prophet model was difficult for me to interpret and ARIMA gave the results of what would be the best investment opportunity. I did choose to focus on a specific metro which is Austin, TX since I am familiar with this market. I’ve lived here for over 25 years and have seen how the real estimate industry has boomed in this metro.
* What three zip codes provide the best investment opportunity for the SREIT?
  + The best zip codes provided by the model are actually outside of the Austin City limits but I would recommend 76530,76578,76257
* Why?
  + I would pick these because I have seen the suburbs of Austin grow at a large rate. There is more room for inventory at a lower cost in these areas.