**IST718 – Big Data Analytics**

**Lab 2**

**David Doman**

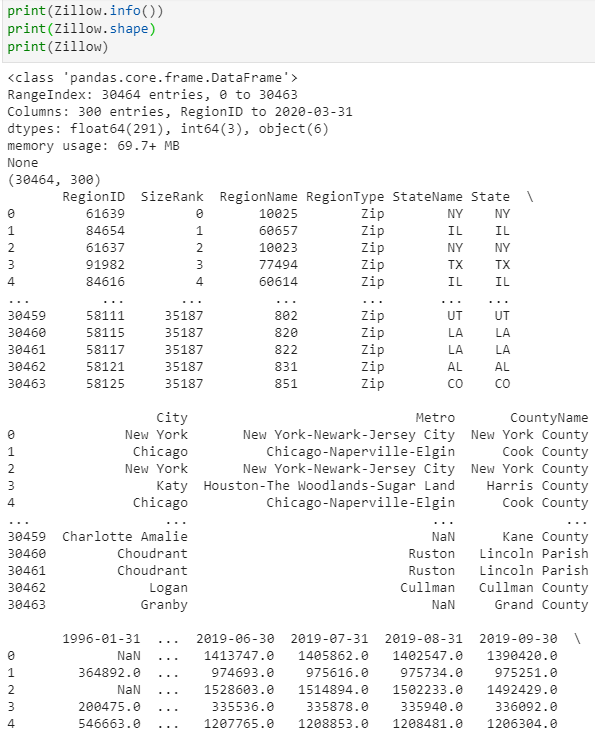
**Introduction**

When it comes to real estate, investors are always trying to find the next big opportunity, especially with respect to real estate. It is an area where if you are not looking forward to potential opportunities you can miss out greatly on very profitable opportunities. In this study, we will be looking at potential areas within the Arkansas area, particularly focusing on Hot Springs, Little Rock, Fayetteville, and Searcy to build models and project possible future investments.

**Dataset**

The dataset for this study comes from a popular source that we all know of and have probably used to look for places to live, Zillow. The dataset contains single family median home values.

Below is an example of the dataset:

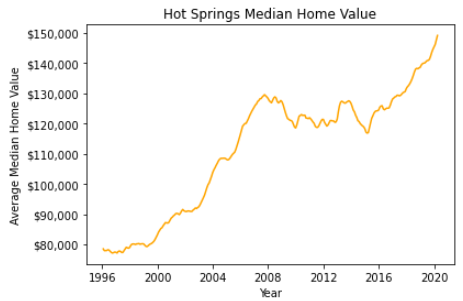


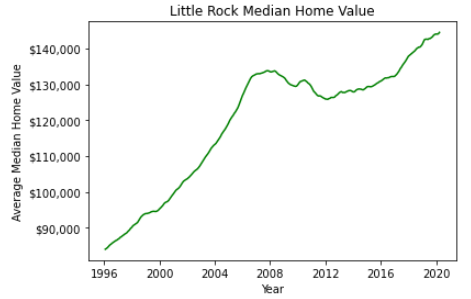
As shown above, the dataset contains categorical and numerical information. RegionName represents the zip codes, along with other location-based information and contained in the dataset is monthly median home value information. The data represented is from the year 1996 to the first few months of 2020. Not much data cleaning was needed. Removing N/A’s was the main task that needed to be performed.

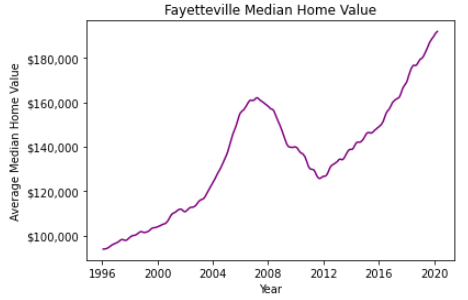
**Initial Analysis**

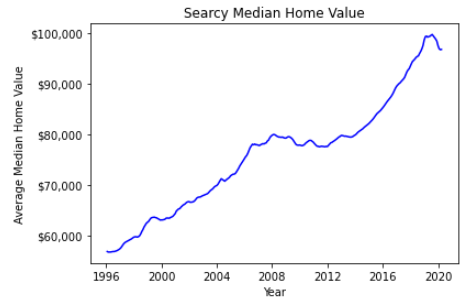
As stated above, 4 metro areas in Arkansas will be used to build the models for this analysis. These areas are Hot Springs, Little Rock, Fayetteville, and Searcy. The median home housing prices will be used to examine the potentially profitable areas to invest in future years.

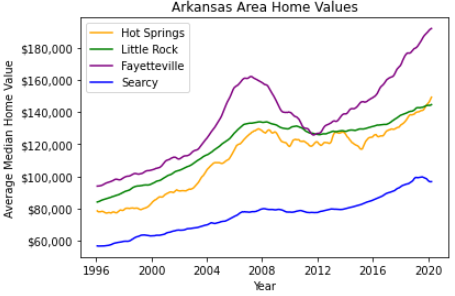
Shown below is the time series plot for each of these 4 areas:







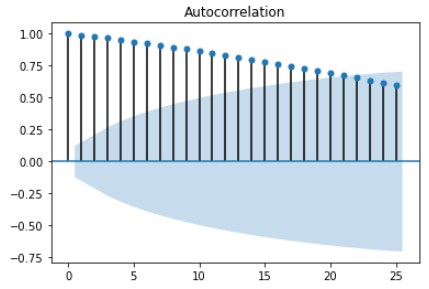


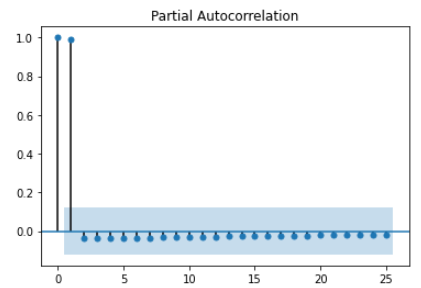


As shown from the graphs above, all 4 metro areas have seen an increase in the median home values over the past 24 or so years. Searcy is in a much lower home value area, seeing a steady increase with no substantial fluctuation throughout the years. Although there has been growth in Searcy it has not been much and is almost flat with a slight decrease within the past year or so. The other 3 areas have seen similar types of trends with spikes around the 2004-2008 time and then drop after the financial crisis of 2008. These 3 metro areas then again regained momentum and have been on the increase ever since 2012 after some recovery from the financial crisis. Hot Springs and Little Rock have had the most similar trends but in the very recent years Hot Springs has increased substantially, while Little Rock seems to have more of a flat trend. Hot Springs, for the first time in 24 years, has surpassed Little Rock in median home values and looks to be at a similar increase of the likes of Fayetteville. Fayetteville has had the highest median home values the past 24 years and has never been lower than any of the other 3 metro areas. Since 2012, Fayetteville’s increase in home value has been on an extreme increase.

The next step was to run the ACF and PACF plots. These models helped to provide insight into the final ARIMA model.

Below these plots are shown:

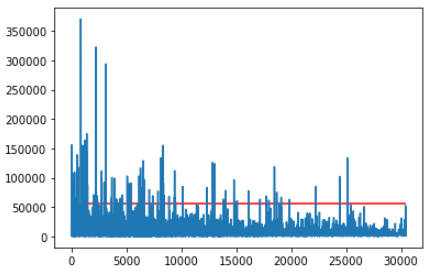




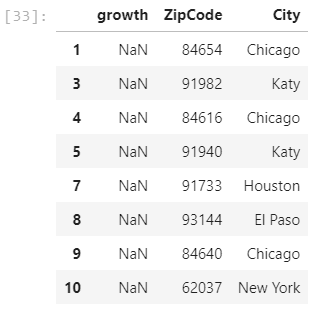
Below is the ARIMA model produced from the research and findings found from what was provided above.



The Mean Squared Error Model is shown below. There is not much that can be interpreted from this model, but it was necessary to run.



The last step was to look at potential growth in certain areas. The results show that Chicago zip codes seem popular so this should be an area of interest for investors.



**Conclusions**

As investors in Arkansas look for potential areas to invest, they could use the information provided above but I believe much more analysis would need to be done to make such substantial decisions. While results have been provided, no decisions should be made based off this analysis but only to use this as source.