Introduction/Business Problem

During a meeting with the home owners association in Huntsville, Alabama, there was a discussion about what to do with an undeveloped plot of land in the neighborhood. One member suggested the installation of a dog park. During the pitch, the member who suggested the dog park noted that while there may be a little bit of maintenance in running a dog park, it would raise property values. After some debate, the rest of the members of the HOA decided it would have to be decided in the next meeting and they requested more information to really identify if building a dogpark is worth it for property values in the neighborhood. The analysis will be done using the foursquare API alongside the API provided by the real estate marketplace, Zillow.

Data

To do this analysis, I will gather up data on multiple zip codes in the Huntsville area using Foursquare's API. I will use data clustering techniques to identify zip codes that are similar to one another in terms of nearby venues and I will then identify if any of those neighborhoods contain a dog park within 3 miles. The zipcodes with dog parks will be compared to the zip codes within their cluster. I will then utilize the historical data of average 3 bedroom home values by zip code provided by Zillow. Using foursquare, I will identify when the dog park was created by getting the details of the venue. I will then track the value of the homes in the different zip codes overtime and compare how the values changed after the installation of the dog park compared to how values changed for other neighborhoods.

After the data is collected, I will compare and plot data showing how home values compared after the installation of a dog park versus other zip codes.

Methodology/Results

To begin, I extracted the zipcodes data from the wikipedia page for Hutnsville, Alabama. I was able to use the wikipedia python package to extract zipcode information from the html. The data of different zip codes was used to find latitude and longitude information for each zip code using geopy. After this, several zip codes were thrown out as they were not currently in use, they were just assigned to the city of Huntsville so it can grow into them. Following this, similar to how we performed our labs, I used foursquare to find venues within the different areas. This data was then broken into a cluster shown in Figure 1. This ended up leaving me with one particularly large cluster so I took all of the zip codes from that cluster and clustered them again as shown in Figure 2. This left us with a smaller main cluster that includes many neighborhoods.

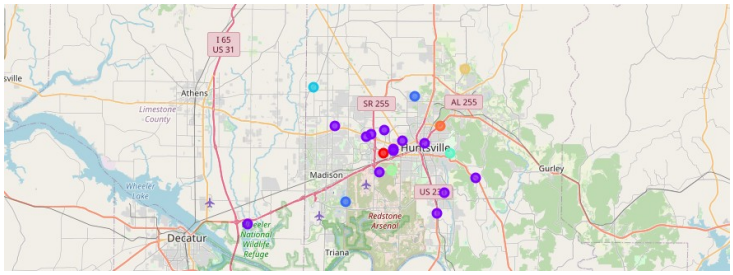

Figure 1: Map of Huntsville, AL with the initial clustering of zip codes base don similar venues.

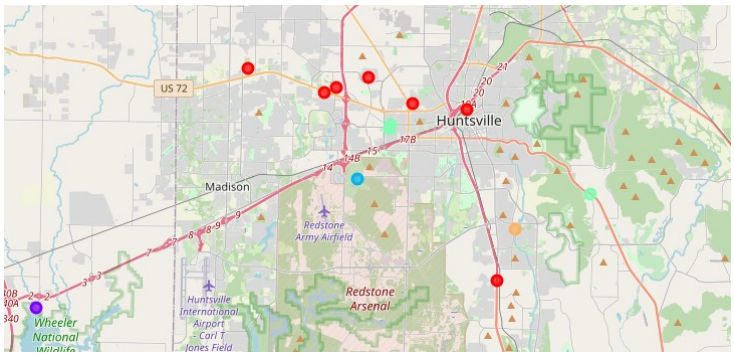

Figure 2: Map of Huntsville, split by zip code, following the second round of clustering.

The final cluster selected was the clusters shown in red in Figure 2. Data was then downloaded from Zillow's website. The data being looked at comes from the value of a 3 bedroom home within that zip code. The data provided occurs over time. We are most interested in seeing how that data changes over time after the installation of a dogpark. To this end, we extracted zip code housing data for the differente neighborhoods. We lost a few hits because Zillow didn't provide home values for those zip codes. In the end, we found 4 different zip codes that fell within our cluster and only one of them contained a dog park. The 35806 Zip Code contains a large dog park/walking trail. From what I was able to find, the dog park opened up in late 2014. Figure 3 shows a comparison of the increase in property values starting shortly before the dog park was built.

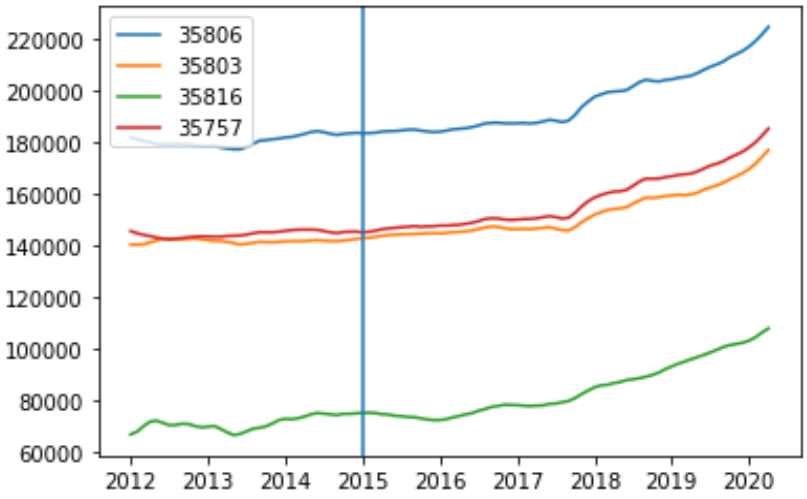

Figure 3: Comparing the home values over time based on Zillow's algorithm. The vertical line shows when the dog park opened.

Discussion

In the end, from the data we worked with, it's difficult to identify a noticeable correlation between a dog park installation and the home values in the neighborhood. That's not to say it wouldn't be an excellent idea as it would enable camaraderie and friendships among the neighbors as it provides a place for their dogs to play and expel energy. I would prefer to do an analysis with the specific neighborhoods but unfortunately Zillow doesn't always contain information on specific neighborhood data. In the end it seems, we don't have a choice but to petition the neighborhood council with a well reasoned argument. If I had more time to work on this project, I would like to redo a similar analysis

but perhaps on a larger city. The reason I think this migh tbe wise is that larger cities will tend to have more data on websites like Zillow as well as foursquare. I believe this would lead to more precise real estate values and we would also likely be able to find more data on specific neighborhoods.

Having said that, it is interesting to see how consistent Huntsville's real estate market has grown the past few years. It appears to me that while certain zip codes have lower property values on average, all zip codes in Huntsville, AL appear to be raising somewhat consistently relative to each other.

Conclusion

In the end, we will have to appeal to the neighborhood council using emotions. Our data unfortunately did not show any specific changes in the pricing of the surrounding neighborhood. Unfortunately some of our data sources came up short and weren't specific enough. But having said that, there's still plenty of reasons to find to build a dog park in the neighborhood.