**Introduction**

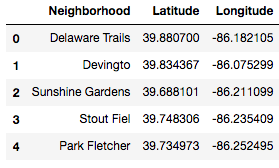
A family owned brewery has outgrown their small operation in Danville, Indiana – a small town about 20 miles West of Indianapolis. The family has decided to relocate the business to Indianapolis and is looking for which location will yield the most success. With the brewery being a football themed establishment, the family would like the new location to be in close proximity to Lucas Oil Stadium, which is the home stadium of the Indianapolis Colts. The exercise aims to identify the category of venues that are most popular near Lucas Oil Stadium along with whether or not there is any nearby competition of other breweries in the vicinity.

**Data**

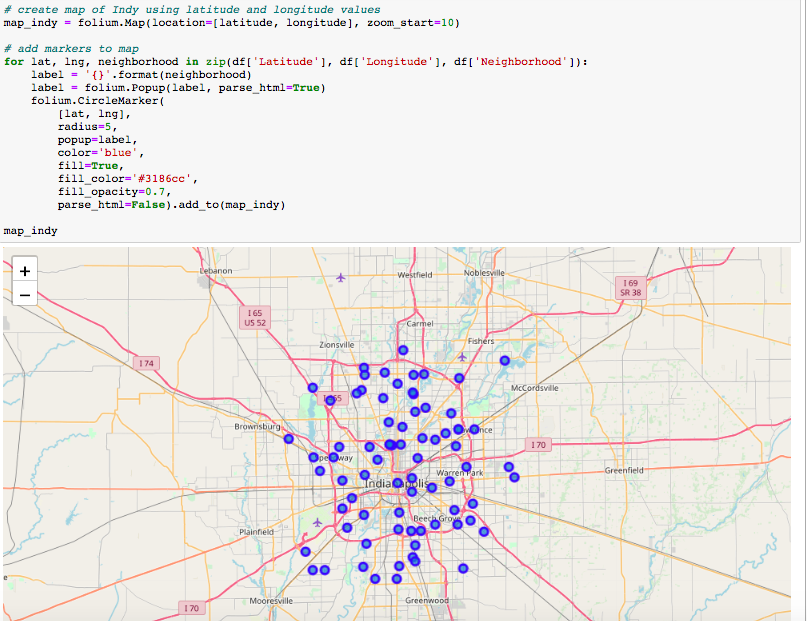
There are two sources of data that was used in this study. The first is a government dataset that provides Indianapolis neighborhoods along with the latitude and longitude. The second piece of information is extracted from Foursquare API, which returns all venue information within a radius of each neighborhood center.

**Methodology**

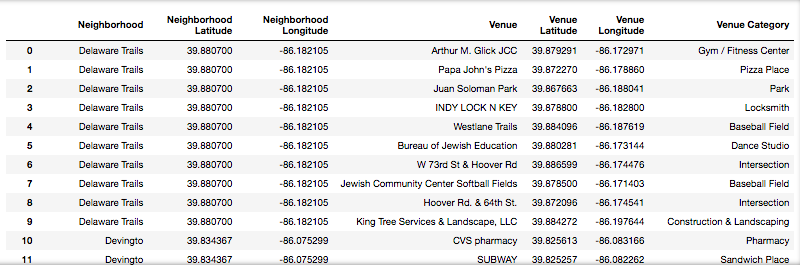
This project’s methodology consists of 3 main components: establish Indianapolis Neighborhood information, use Foursquare API to extra venues in each Neighborhood, and analyze neighborhood venues near Lucas Oil Stadium. First I was able to create a pandas datafame that contained 77 Indianapolis neighborhoods along with their latitude and longitude. Below are the first five neighborhoods in the dataset:



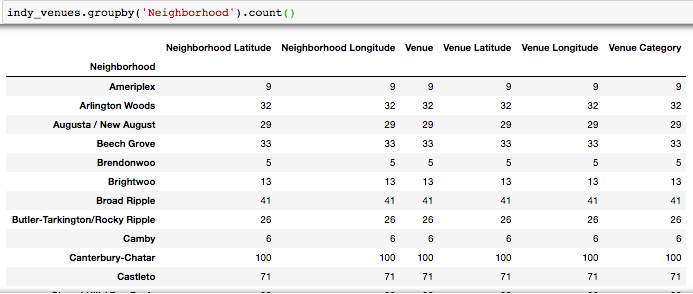
Mapping the neighborhood data using the Folium package:



The information pulled from Foursquare API contained all the venues, venue categories, and venue latitude/longitude within a 1 mile radius. I then added this information to the existing dataframe and computed the count of total venues for each neighborhood that was pulled from Foursquare API. Below is the result of adding Foursquare API data to the existing dataframe:



Here is a snapshot of the result of the total venues extracted from Foursquare API by grouping the dataframe above by Neighborhood:

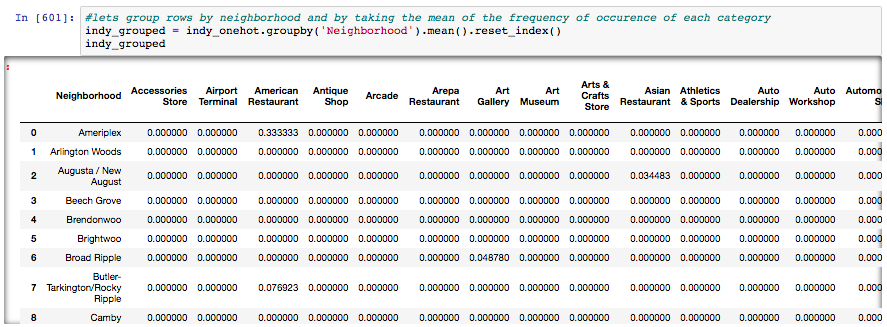
****

Next, I computed how many unique categories exist from the Foursquare API data. In other words, the results below represent how many different venue type/category exist in the Indianapolis neighborhoods:

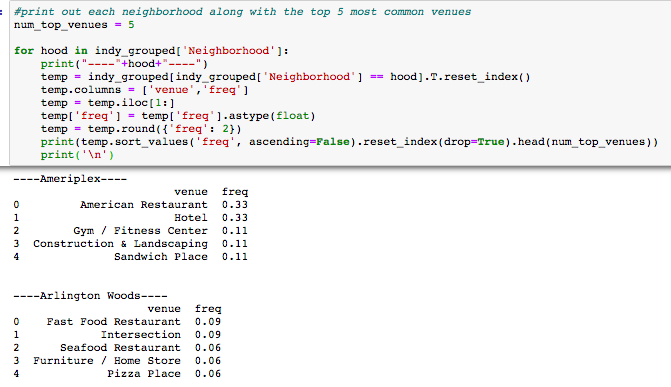
../Desktop/Screen%20Shot%202019-11-16%20at%209.49.15%20AM.png

The next process in this study was to analyze the neighborhood information individually. To do this, I had to convert the venue categories into dummy variables and group the data by neighborhood and generate a new dataframe. The results are below:

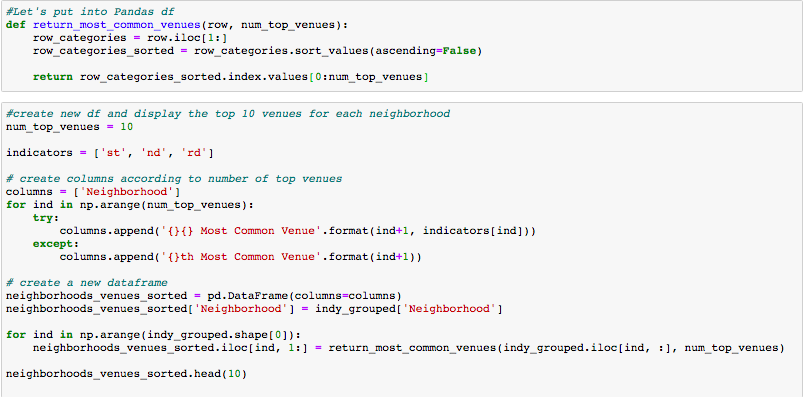




The code below computes the frequency of each venue category per neighborhood. For example, in the Ameriplex neighborhood, American Restaurants and Hotels are the two most popular types of venues. In the Arlington Woods neighborhood, the most popular type of venue is Fast Food.

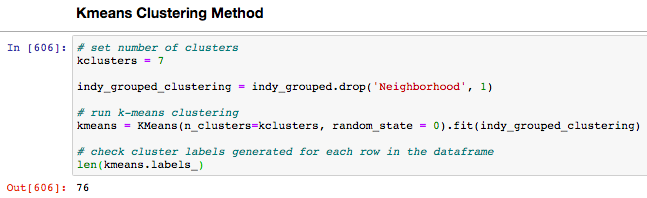


Next, I created a function that returns the most common venues in each neighborhood. This function will then be used in the code below that generates a new dataframe that represents the top 10 most common in each neighborhood.

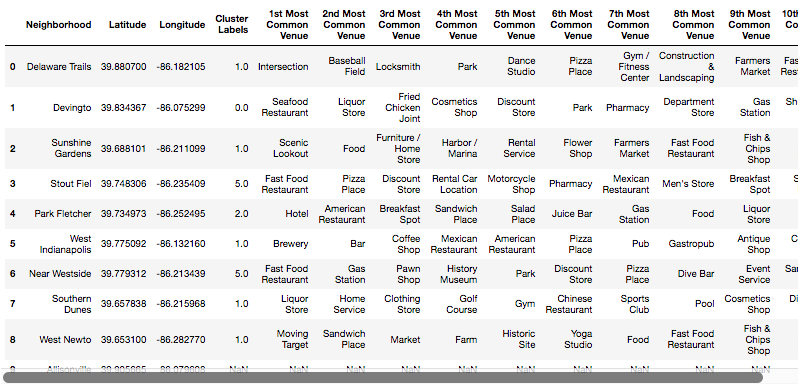




The next process in this study was to begin the Kmeans Clustering method to cluster the neighborhood data. Clustering is a classification method that aim to partition the input datasets into groups. The data in each group are very similar to each other, and very different than data in the other groups. Kmeans is one of the simplest and most common algorithms for clustering data. Below is a snapshot of the code used to generated cluster labels. The cluster labels represent which cluster a neighborhood falls in based on venue category similarities.

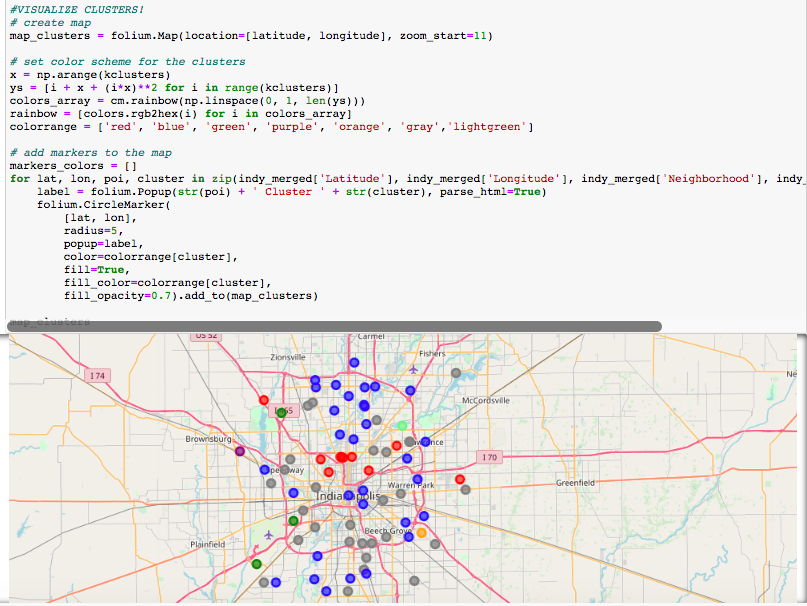


I chose to use 7 clusters because I thought it resulted in the most useful information. Entering the label information into the final dataframe:

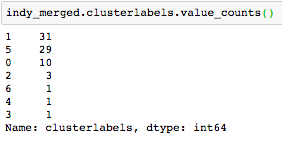


**Results**

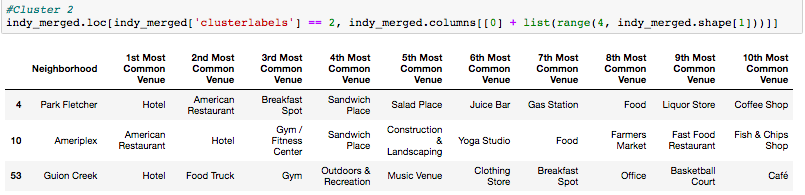
Finally, after running the k-means clustering and adding the results to the final dataframe, we can plot the clusters using Folium. Each color represents a different cluster.

****

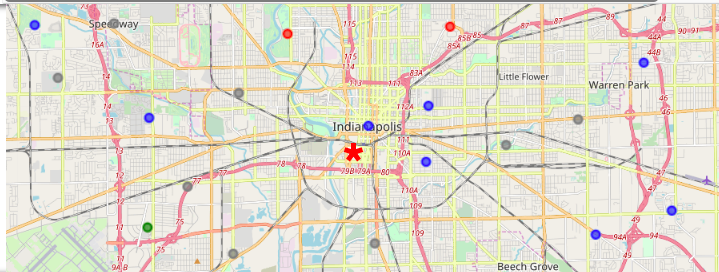
How many neighborhoods fall into each cluster?

****

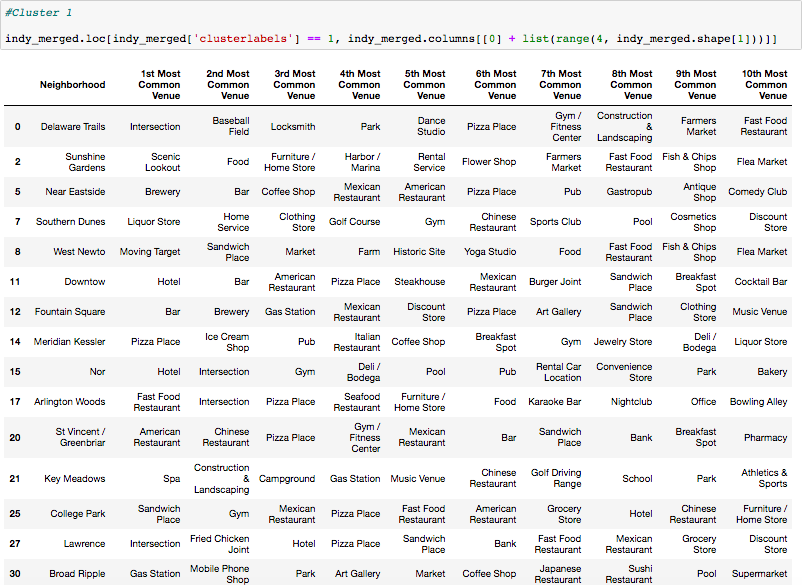
Next, I generated a dataframe for each cluster that shows which neighborhoods fall into that cluster, along with the top 10 most common venues in that neighborhood. For example, as seen below with Cluster 2 – there are three neighborhoods that fall into it: Park Fletcher, Ameriplex, and Guion Creek. When observing the most common venues in that cluster, you can gather that these neighborhoods were clustered together because they have similar top venues – namely Hotels and American Restaurants.

****

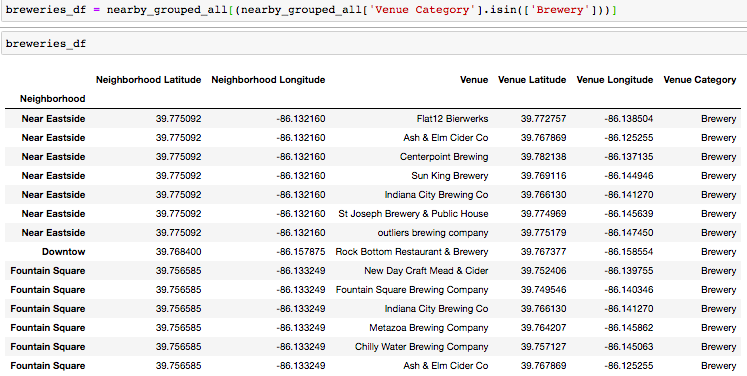
The ultimate goal of this exercise was to investigate the viability of moving and opening a family owned brewery close to Lucas Oil Stadium (LOS). There are many factors that go into determining that viability, and one of them is looking into the nearby competition. If we zoom into the map we can see that the approximate location of LOS (red star), we can see that it would likely fall into the blue cluster (Cluster 1).



The three nearby neighborhoods, which are in cluster 1, are Near Eastside, Downtown, and Fountain Square. Let’s look at the top 10 most common venues in Cluster 1 and analyze those 3 neighborhoods.



As we can see from the dataframe above, breweries are in the top 2 categories for Near Eastside and Fountain Square neighborhoods. Below is the result of all the breweries in the Near Eastside, Downtown, and Fountain Square Neighborhoods.

****

**Discussion**

Based on the results, if the family moved their brewery near Lucas Oil Stadium, the brewery location would most likely fall into Cluster 1. The three closest neighborhoods, which are Near Eastside, Downtown, and Fountain Square, turn out to have breweries as quite a common venue. There are at least 14 breweries in those neighborhoods combined. There might be more if I increased the radius of influence when pulling info from Foursquare API. Measuring the distance on google maps, this puts at least 14 breweries within a 3 mile radius of Lucas Oil Stadium.

While there are a lot of breweries existing near the stadium, there is at least 1 venue that falls in the brewery category in the Downtown neighborhood – which is Rock Bottom Restaurant and Brewery. Thus, if the family does decide to move the brewery near Lucas Oil Stadium, moving it to the Downtown neighborhood would be the best option give the locations of surrounding competition.

**Conclusion**

In conclusion, the family would face some stiff competition if they decide to move the brewery near the stadium as there are at least 14 other breweries within a 3 mile radius of the stadium. If they do decide to move near the stadium, the downtown neighborhood might be the best location as it has the least amount of breweries in that Cluster. However, giving the fact that the neighborhood is Downtown Indianapolis, the price or buying/leasing space is likely amongst the highest of all neighborhoods in Indy, which could explain why there is a lack of breweries in that neighborhood. Examining the price per sqft for buying/leasing space in Indianapolis neighborhoods would be another useful study for helping the family make a decision.