**Introduction**

For my final capstone, I wanted to look at the main city where a lot of my family lives, Toronto, to see where the best location within the city would be to open a new restaurant. In my hypothetical situation, I am assisting a client who desires to open a new restaurant somewhere in Toronto but wants an ideal spot with limited competition from similar establishments.

By using Foursquare combined with data science techniques, I will recommend the best area to start a new restaurant with the fewest surrounding competitors to give my client’s restaurant the best opportunity to succeed. Foursquare has venue traits combined with location data that I will use to map out the city, and then generate a data set that can be used to show best locations for the new business.

**Data**

The data used was extracted from Foursquare through their developer API in combination with Python programming. The first step to the data gathering process was the retrieval of the geographical coordinates of the neighborhoods across the city limits. From the location data, the venues and their characteristics could be scraped and inserted into a data frame. This data frame will help separate the venues into the different neighborhoods. Later, the similar venues with similar traits such as restaurant type, are sorted and grouped. After some k-means clustering, the map could be further improved with color markers based on cluster label.

**Methodology**

The main data science technique used was k-means clustering that classified and grouped the venues by neighborhood and venue type. This made it possible to see what types of venues were where and it was also easy to see the rankings of most common and least common venue types in each cluster and neighborhood.

**Results**

The resulting data frame shows the 38 neighborhoods within Toronto each with rankings of the 10 most common venues within the vicinity of the neighborhoods. It also has cluster labels from the k means clustering. The resulting map contains location markers for all of the neighborhoods within the Toronto city limits that is colored based on the cluster labels. Now the client is able to see the locations of possible venue sites along with the breakdown of the most common venue types in each neighborhood.

**Discussion**

The resulting data set and map are useful tools that enable the client to visualize the best location for their new venue, but the data itself obtained from Foursquare does have some issues. Some of the venues are labeled purely as “restaurants”, while others have more specific labels that specify the type of food they serve. In some cases, these labels overlap and may characterize the same venue multiple times, which might skew the resulting table for some venues types to be more or less common. These overlapping labels occur several times in different instances. Ignoring the possible skew, in my hypothetical situation, my client decides that he/she wants to open a “bar” type venue, in which case the data display shows 2 ideal locations having “bar” as the 10th most common venue type. From there, they can see on the map which location might be better suited based on the specific coordinates in the city.

**Conclusion**

In conclusion, after analyzing venues based on location and type within Toronto neighborhoods, we can see that Runnymede, Swansea and Brockton, Parkdale Village, Exhibition Place are the 2 neighborhoods with the least amount of venues considered “bars” and from the map generated can see that they are both located in the western part of the city. In future analyses, we could use more specific statistical techniques to see the numerical differences between neighborhoods.