Capstone Project- Battle of Neighborhoods

Opening a Scandinavian Restaurant in Toronto, Canada

Introduction/Business Problem

The inspiration for this subject comes from personal experience, having explored Scandinavian countries over the past years. I love their culture and food, but there is little to no opportunity to experience any of that outside of Scandinavia. The few places that do have certain elements of this cuisine lack the hygge feeling that makes the picture whole.

This project aims to find the best location to open such a restaurant in Toronto. While there is little competition, I will be analyzing restaurants with similar cuisine, such as places specialized in serving fish or French restaurants, whoch also offer seafood and their dessert choices are very similar.

This project aims to find the perfect location to open such a restaurant in Toronto. It could also be used by anyone else thinking about opening an unusual restaurant with little to no direct competition in the city.

Data

For this project I need to use a combination of data that is available online and also the Foursquare API.

- List of postal codes and neighbourhoods in Canada (https://en.wikipedia.org/wiki/List_of_postal_codes_of_Canada:_M)
- Geospaial data in a csv format to get the longitude and latitude values used for plotting (https://cocl.us/Geospatial_data)
- Foursquare API data about venues, which will help me categorize and group the restaurants

The data section is inspired by the practice notebook from the course.

Methodology

Data transformations

I started by gathering the data, cleaning and combining the necessary data sources to get the data ready for analysis and clustering.

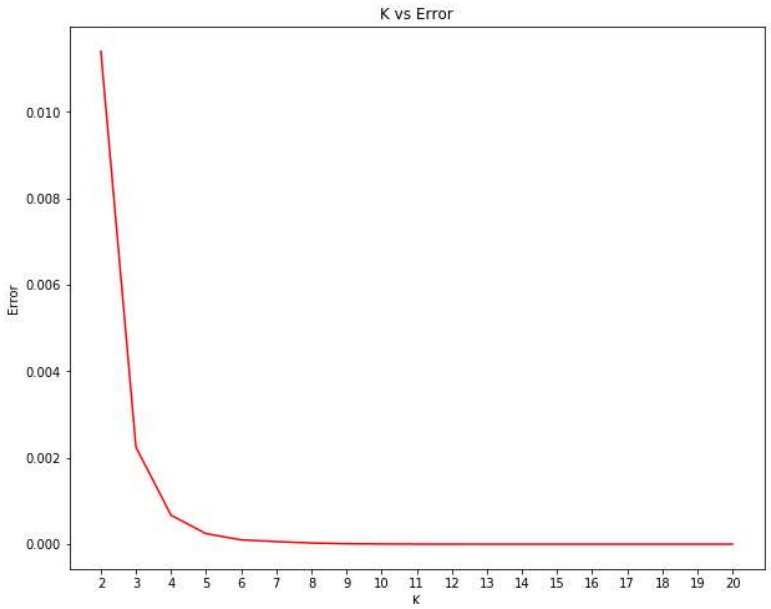
After getting the base data from the wikipedia page, I merged it with the Geocoder csv data to get the whole picture of Toronto geographically. As a next step I extracted the venue fata from Foursquare and merged it with the previous dataframe, thus making the picture whole and my data ready for analysis. I used an older Foursquare API version, since the corona outbreak changed the restaurant landscape and I an optimist, I would like to have a picture of the world before and after the pandemic.

As I previously mentioned, I found no direct competition, since there are no Scandinavian restaurants per se in the data, so I resorted to grouping venues which serve similar cousines. I chose venues serving fish and French restaurants and potential competitors, created a modified category in my data for them and then worked with this modified category when clustering.

For the machine learning part I first created dummy data with one hot encoding to make my data numerical and suitable for the algorithm. I then simplified this dataset to only show neighborhoods and the mean frequency of potential competitor restaurants.

K -Means Clustering

I decided to do the clustering based on the neighborhoods that had similar averages of potential competitors in the area. For deciding upon the optimum k value, I resorted to the Elbow Point technique, which suggested that a k of 4 would be ideal in my case.



I then created a map using the Folium package to visualize the clusters.

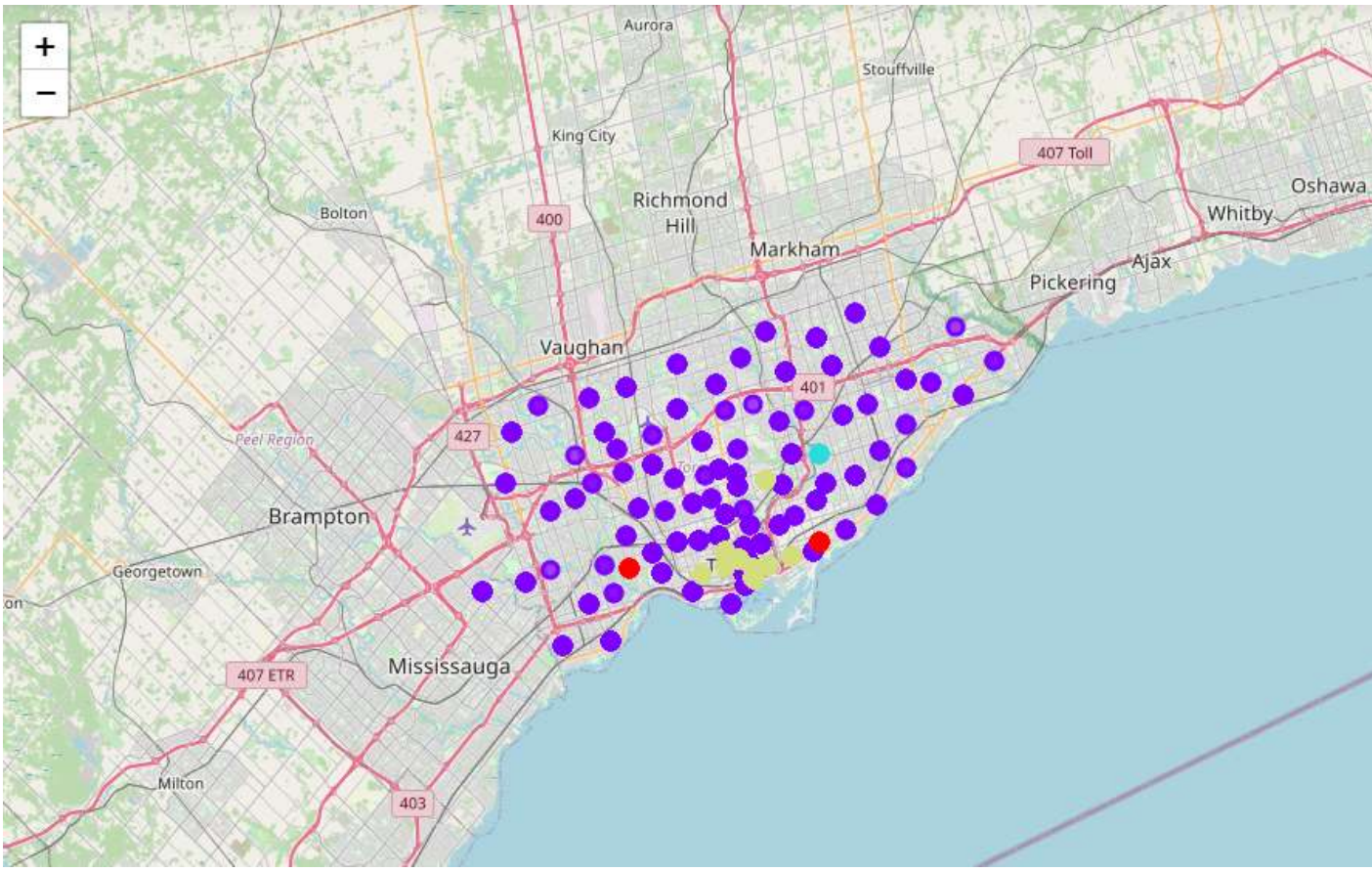
This is very important, since a location close to the sea would go perfectly with the vibe of the restaurant. When deicing upon the perfect location, I looked at the number of potential competitors per cluster and that, together with the location is very valuable when choosing where our Scandinavian restaurant should be.

Results

There are certain relevant findings to consider for someone who would like to open a Scandinavian restaurant in Toronto. First of all, there is no direct competition, meaning that when analyzing competition, one can only resort to choosing venues with a similar profile.

I hereby created a new label, called “Potential Competition” which is basically a group of venue that serve fish and French cuisine.

After assigning a competitor score to each neighborhood, I was able to create 4 neighborhood clusters based on the competition in the neighborhoods.



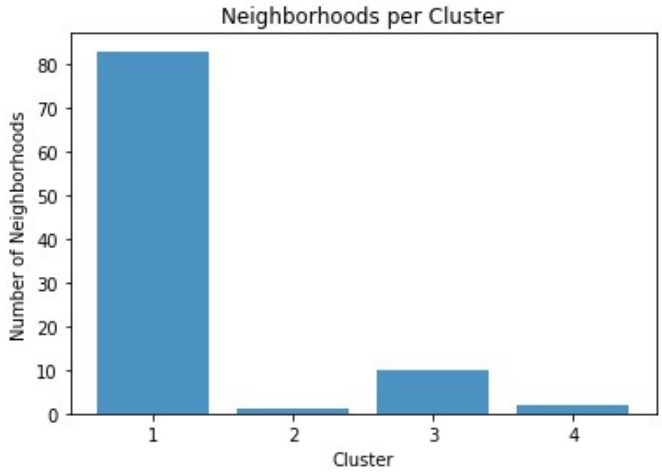
Cluster 1 -purple

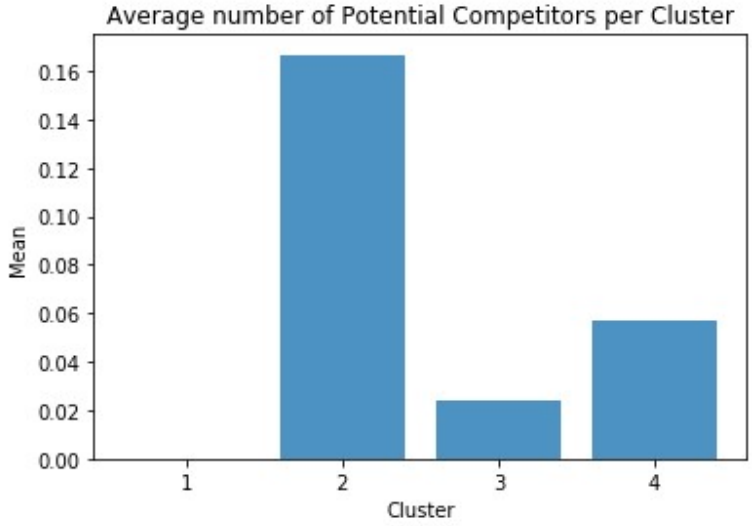
Cluster 2 – light blue

Cluster 3 – yellow

Cluster 4 – red

I then analyzed each cluster and compared them in terms of neighborhoods in each cluster and competitors in each cluster.





We have a very large area with virtually no competition, but also with no natural scenery, which is important for the hygge factors. So even though this blank area could be further explored, it would be difficult to find a beautiful place there.

Then we have cluster number 2, which contains only 1 neighborhood and the competition is high. There is definitely a preference for this cuisine in the area, however, if we look at the map, we can see that this area also lacks the natural scenery, so I'll stick to the other alternatives.

Clusters 3 and 4 are closer to the sea, with the competition being a bit higher for cluster 4, which is composed of only 2 neighborhoods. If we close in, we'll see that cluster 3 is next to a university and other tourist hotspots, which I'd like to avoid in order to have a cozy atmosphere.

In this case the map can help us decide, since if we zoom in, we'll that one of cluster 4's locations is next to a park and close to a beach. Even though the competition is higher, than for the other cluster, this seems to be the perfect setting and perhaps the demand is higher too.

Discussion

Based on the data above we have a very large area with no potential competitors at all, shown with purple on the map. This could be further explored by analyzing demographic data. However, the lack

of competition could also indicate that this area wouldn't work at all and it's not by the sea, which is an important aspect.

The most promising is cluster number 3, with little competition, but a very nice location. The cluster with the most competition should also be further analyzed, since it could indicate a strong preference and demand for such food by the ihabitants of the area.

Given the fact that there is no direct competition, surveys and other data could be used to make this project more accurate and detailed.