Coursera Capstone Project Report

by Kahang Ngau

Part One: Introduction where you discuss the business problem and who would be interested in this project.

- The question that this project seeks to answer is to find which neighborhoods in Toronto could offer similar services if I had to move in the around cities, or even between cities in each county of Toronto.
- The target audience of this project would be someone who had to relocate themselves to different cities for some kinds of reason, like relocating jobs. They would like to see and compare the conditions of neighborhoods in around the city so that they can make the best choice.

Part Two: Data where you describe the data that will be used to solve the problem and the source of the data.

- This project will compile data frames for each financial center in the city of Toronto, Canada, by combining data scraped from several online sources. The data that I am going to use contains the neighborhoods, boroughs, and their respective longitude and latitude for each city.
- First, I will compile the data containing the top 5 most popular venues for each neighborhood which is used in a KMeans algorithm to cluster the neighborhoods together based on the similarity of the categories for their top 5 venues by using Foursquare location data to generate the map.
- This project only utilizes venues when grouping similarities. I plan to incorporate more data in the future such as proximity to water, population in each neighborhood, prices of houses, etc. to further create unique clusters to help identify similar neighborhoods

Part Three: Methodology section which represents the main component of the report where you discuss and describe any exploratory data analysis that you did, any inferential statistical testing that you performed, and what machine learnings were used and why.

- The first step will be creating a map and examining the different neighborhood of Toronto. As you can see below that a map is generated by applying the methods of plotly and mapbox to display the neighborhood of Toronto grouped by the postal code.

Neighborhoods of Toronto City
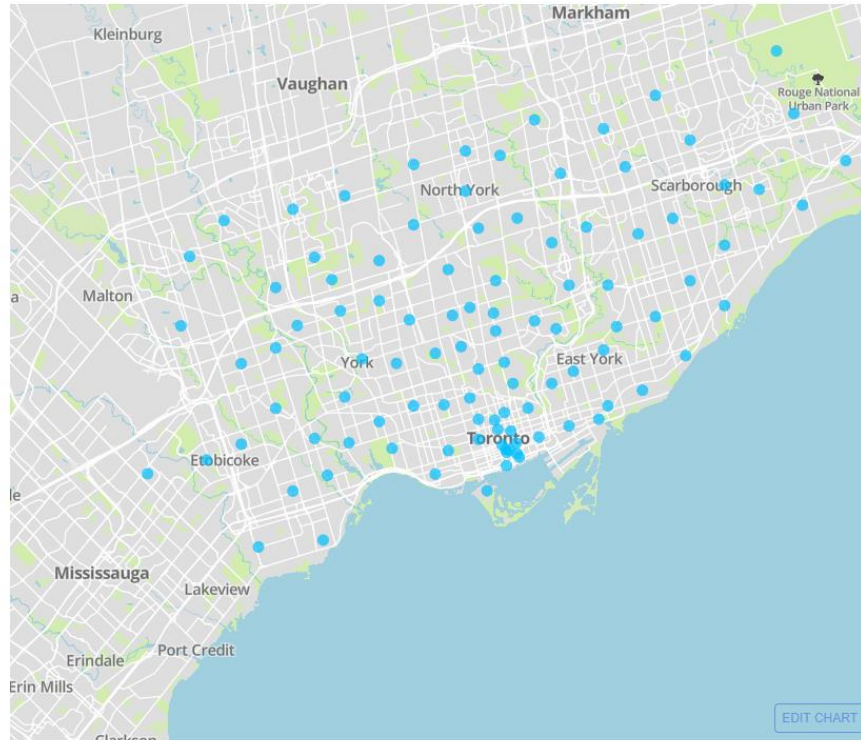
- Then another interactive chart is created for the neighborhoods and venues in and around Toronto, Canada. All the neighborhoods are highlighted in Blue while the venues are highlighted in Red.
- As expected, we can see that as you reach the inner parts of the city. The venues are found within tighter clusters around their respective neighborhoods. Hover in and around the map to see the different neighborhoods and their unique venues.
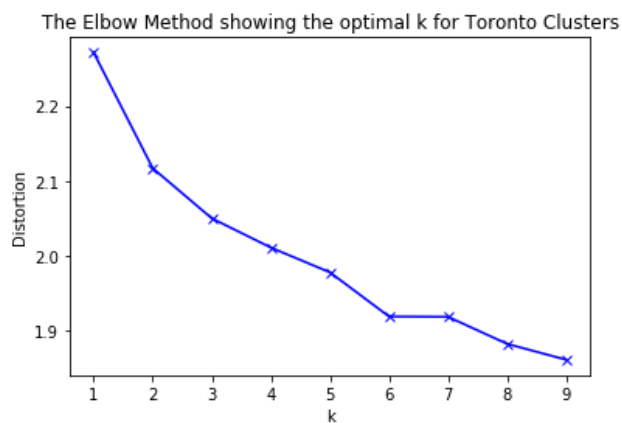
Neighborhoods and Venues of Toronto City
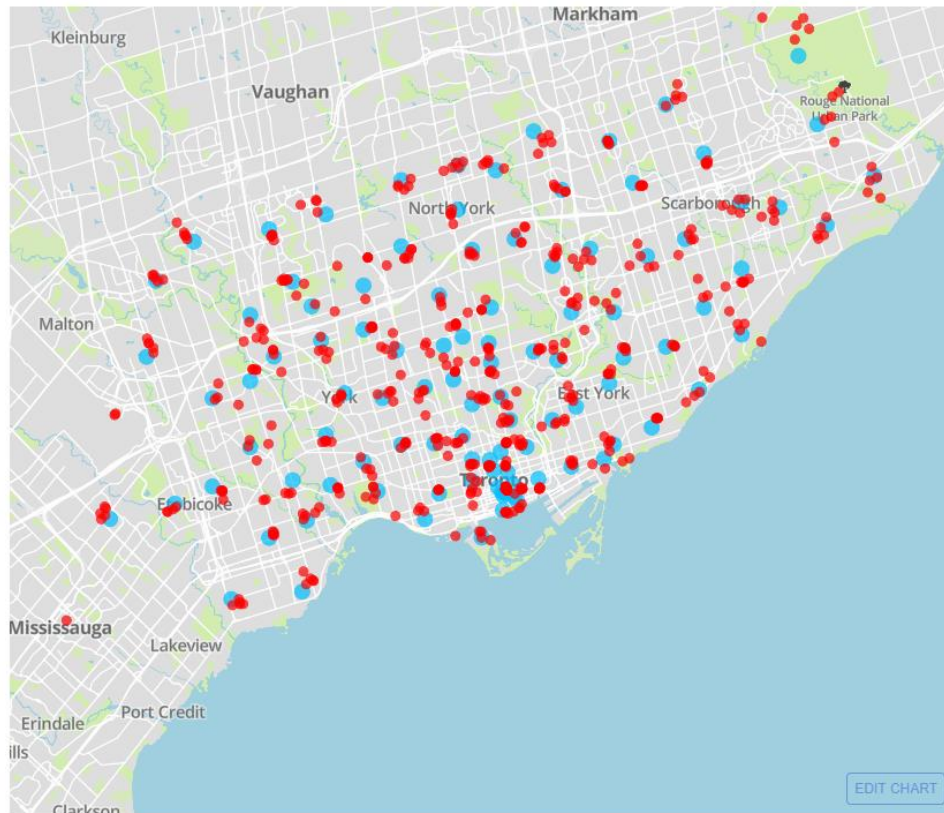
- Below is the map of the city Toronto which it displays the color code neighborhoods that are similar and dissimilar to one another based on the categories and conditions of their venues.
- Venues that are of the same color are similar to one another, on the other hand venues that are dissimilar are having different colors. I complied the data by using the cluster methods that categorize the venues and neighborhoods by their conditions.
- Then, I achieved this process by applying the foursquare API and past in the coordinates of each neighborhoods and to find venue information of the top 5 venues of their locations.
- I finally cluster the neighborhoods to see which are similar to each other based on the categories of the venues information.

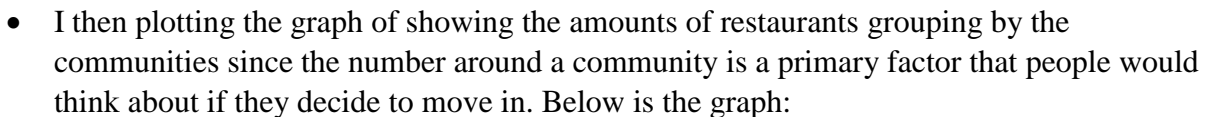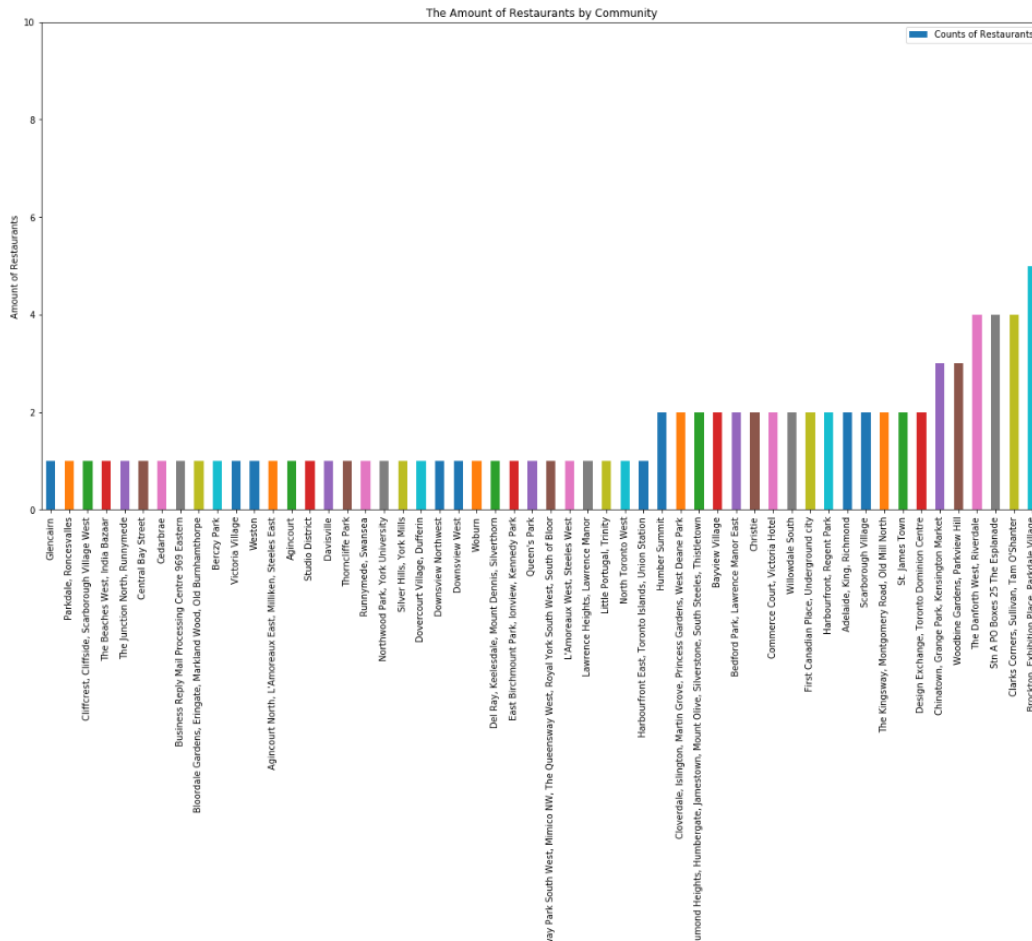Clusters of Neighborhoods by Venues in Toronto City
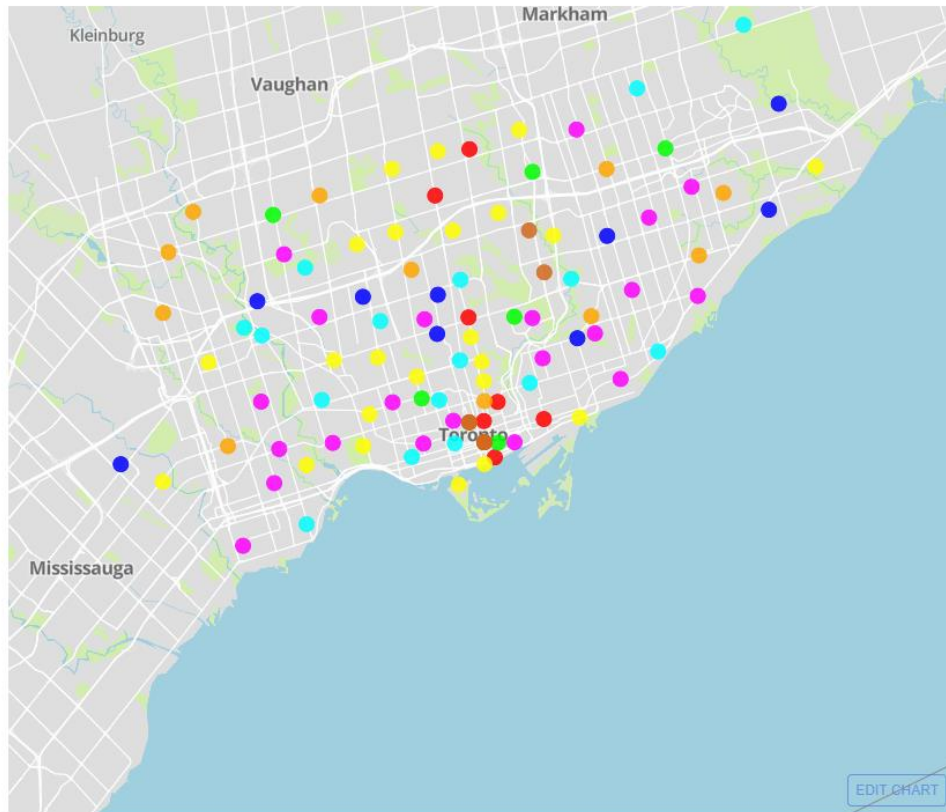
Part Four: Results section where you discuss the results.

- The result of my project finds the optimal K for the Toronto cluster. In order to do so, I applied the Elbow Method as shown below to find the best K which is 6.

The Elbow Method showing the optimal k for Toronto Clusters

Part Five: Discussion section where you discuss any observations you noted and any recommendations you can make based on the results.

- After we find out the best K to optimal the cluster in Toronto, I used plotly to plot the chart grouping by the venues of the communities. And we have the graph shown below:

- I then plotting the graph of showing the amounts of restaurants grouping by the communities since the number around a community is a primary factor that people would think about if they decide to move in. Below is the graph:

The Amount of Restaurants by Community

Part Five: Conclusion section where you conclude the report.

- We can see that Brockton Exhibition Place, Parkdale Village has the greatest amount of the restaurants in its community.
- If I were living on the outskirts of the city and working in the inner city and wanted to reduce my commute to work by moving to the inner city but wanted to move to a neighborhood that offers similar services, which neighborhood would I move to? So looking at the map, if I were living in the neighborhood whose point overlays Etobicoke label(Red data point), I would see the neighborhoods that were similar based off of their red color code. I would move to either of the two neighborhoods that are flanking the Toronto label which are colored red. Or, I would move to the neighborhood overlaying the East York label.

Clusters of Neighborhoods by Venues in Toronto City