# **Introduction**

Currently, you live in West Toronto and you really enjoy living there. Due to recent circumstances pertaining to the Coronavirus, you lost your previous job and have had to find another one. Fortunately, you received an offer this past week. However, you must accept or reject it by the end of this week. Unfortunately, you were offered a job in East Toronto. Due to distance, traffic, Subway System routes, and other variables impacting duration of travel to work, you cannot conveniently live in West Toronto while working in East Toronto.

Pre-accepting or rejecting the written job offer, you want to compare life in East Toronto – commerce, housing, and much more – to life in West Toronto. To compare the venues and commerce in both areas, I will construct a K-Means algorithm used to cluster similar neighborhoods throughout the city of Toronto. If East Toronto and West Toronto fall into the same clusters – with respect to the venues and commerce – then you will continue to look into other aforementioned variables (Example: Housing). Otherwise, you will decline the job offer.

# **Data**

## Data Sources

A few sources were utilized when developing this algorithm. The first source was the following link -- <https://en.wikipedia.org/wiki/List_of_postal_codes_of_Canada:_M>. A table with Toronto postal codes (starting with the letter “M”), as well as respective boroughs neighborhoods, was retrieved from this page. The second source was the following link – <https://cocl.us/Geospatial_data>. This link contains a data frame with coordinates (latitudes and longitudes) corresponding to each of the Toronto zip codes. The rest of the data was collected using the Foursquare API. Data collected using this API includes venue names, their respective coordinates, and their respective “categories” (Examples: Pizza Shop, Coffee Shop, Grocery Store).

## Data Cleansing

After data was collected from the first source (<https://en.wikipedia.org/wiki/List_of_postal_codes_of_Canada:_M>), all rows with “Not Assigned” for the Borough column were deleted from the table. Data was also cleaned when organizing data for the cluster visualization. Some of the cluster labels were “NaN” – these rows were deleted from the data frame used for the visualization. Additionally, the cluster labels were converted from float values to integer values. This conversion allowed the clusters to be viewed within the final map.

# **Methodology**

## Exploratory Data Analysis

In the “Exploratory Data Analysis” portion of my notebook, the following steps were completed:

* 1. I looked at different venues, their coordinates, and their respective “types.”
  2. I grouped the venues by neighborhood.
  3. I compared existing venue types across each of the neighborhoods.
  4. I calculated a “frequency” for each of the venues in each of the neighborhoods.
  5. I displayed the top 5 most frequent venues in each of the neighborhoods for visual comparison.

No major takeaways were extracted from this exploratory data analysis. My high-level takeaways were the following:

* Common venue types varied across each of the neighborhoods.
* Restaurants, Coffee Shops, and Grocery Stores were some of the most common venues across all neighborhoods.

## K-Means Algorithm

In this analysis, the K-Means Algorithm was utilized to group similar neighborhoods on the basis of commerce/venues. This algorithm was chosen because I wanted to “cluster” groups based on similarity/distance metrics for three reasons:

* As a developmental exercise.
* As a means of comparing East Toronto neighborhoods with West Toronto neighborhoods.
* K-Means Algorithm is a simple clustering algorithm to implement.

Other clustering algorithms considered include hierarchical clustering, agglomerative clustering, and DBSCAN. However, amongst these algorithms and the K-Means algorithm, the K-Means algorithm was the easiest. As a result, I utilized this algorithm.

# **Results**

In this study, it was found that East Toronto venues and West Toronto are similar venue-wise. All of the neighborhoods from each of the boroughs belonged to cluster 4. On the next page, you will see images supporting this result.

A screen shot of a computer

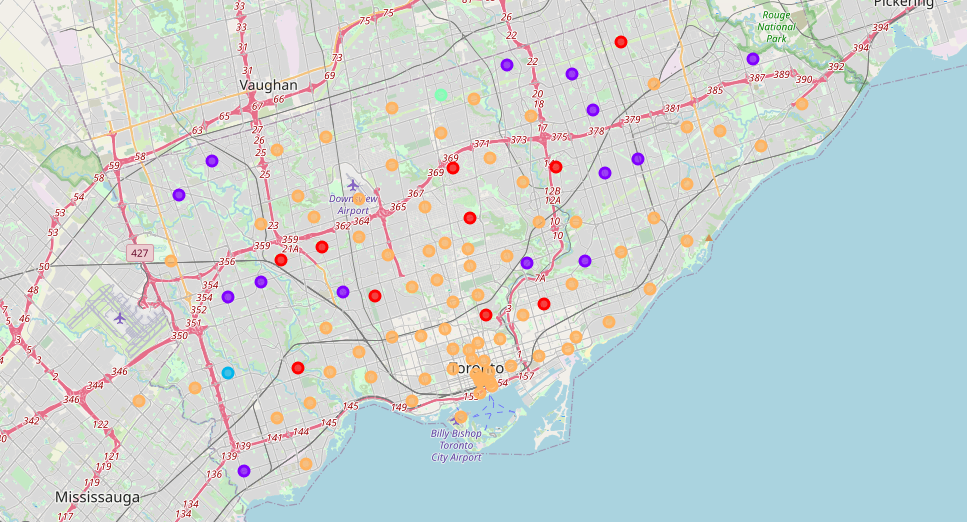
Description automatically generated

Table 1: Comparison of East Toronto and West Toronto Neighborhoods

Figure 1: Map with Neighborhoods Clustered by Color

# **Discussion**

Provided the results of the study, I would recommend further comparing the life of East Toronto and West Toronto, in the scenario provided in the introduction. In the future, when conducting a study like this, I would recommend normalizing metrics where applicable. In this study, no metrics were normalized. Additionally, I would label the map (Figure 1) with boroughs for clarity. While the map does support the results of the study, it is difficult to discern where “East Toronto” and “West Toronto” are located on the map.

# **Conclusion**

In this study, you had – in theory – previously lost your job and acquired a job offer in East Toronto, while living in West Toronto. Commuting to East Toronto is too inconvenient, so working in East Toronto would require living there. Additionally, you really like West Toronto, so you are not inclined to move. You will only move and accept the job offer if life in East Toronto is comparable to life in West Toronto. In this study, I compared the commerce of the two boroughs and found that the two areas are similar via K-Means Clustering. Provided these results, you are now compelled to compare other aspects of life – such as housing – to determine whether or not moving to East Toronto is warranted.