# Capstone Project
# Report

## Introduction

When someone ponders the possibility of moving, especially to an entirely different country such as Canada, it would be helpful to have the ability to learn about different neighborhoods and their features prior to making the major decision and actual move. What distinguishing characteristics make Toronto, in particular, a possibly desirable place to live? Having access to such information would be especially valuable during the decision-making process for possible new immigrants to Canada. Data Science and its methodologies, such as developing a recommender system, can help facilitate the research process for individuals contemplating moving to Canada, and in particular, the Toronto region.

Upon closer research, it is possible to utilize Data Science techniques and tools to create a recommendation system that would help lay out geographical data for potential immigrants to Canada. Would moving to the Toronto region be cost-effective and appealing? This question can be addressed by analyzing data related to the various features that define different neighborhoods of Toronto, Canada.

## Source of Data

For this project, we needed access to the data set containing geographical information about the neighborhoods in Toronto, Canada. The neighborhoods within this region can be identified using postal codes and the geographical coordinates. This information is available using the Wikipedia website: 'https://en.wikipedia.org/wiki/List_of_postal_codes_of_Canada:_M'. We also needed information identifying unique characteristics about the neighborhoods in Toronto, Canada. The Foursquare application is useful for this aspect of data collection, and the

Four Square APIs were used to study different neighborhoods in Toronto. To make this project more manageable, the boroughs specifically focused upon were Downtown Toronto, East Toronto, West Toronto and Central Toronto.

**Methodology**

In order to determine which neighborhood in Toronto would be the most desirable to possibly move to, we needed to run exploratory analyses and cluster the neighborhoods in Toronto. As mentioned in the Data section, the boroughs focused upon were Downtown Toronto, East Toronto, West Toronto and Central Toronto.

First, the Wikipedia website 'https://en.wikipedia.org/wiki/List_of_postal_codes_of_Canada:_M' was scraped and the data was wrangled and cleaned. Then geographical locations for the various neighborhoods were obtained from the dataset located at: https://cocl.us/Geospatial_data.

Next, the Four Square APIs were used to acquire information about the various features of the neighborhoods in Toronto. This exploratory analysis provided high-level information about the venues available in the boroughs of Toronto. The venues we were most interested in related to health care, education, fine dining, entertainment and retail store options.

K-Means Clustering was utilized to segment the Toronto boroughs into five different clusters. We studied each cluster and determined which one had the highest value of offerings or features, and this would be identified as the neighborhood with the most significant appeal to a potential Canadian immigrant.


## Results

| Groups | Arts and Entertainment | Athletics and Sports | Educational | Food and Dine | Gym and Fitness Center | Health and Beauty Service | Hospitals | Others | Recreation | Stores and Utilities | Transportation | Facilities |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| C3 | 8.272727 | 0.727273 | 48.636364 | 68.090909 | 2.545455 | 0.909091 | 4.64E+00 | 3.727273 | 2.545455 | 12.272727 | 0.454545 | 152.818182 |
| C1 | 5.090909 | 1 | 21.545455 | 70.818182 | 3 | 1.181818 | 4.44E-16 | 2.454545 | 2.727273 | 13.272727 | 0.181818 | 121.272727 |
| C4 | 4 | 1.666667 | 12.5 | 53.333333 | 2.5 | 1.166667 | 3.33E-01 | 2.166667 | 7.666667 | 9.666667 | 1.333333 | 96.333333 |
| C2 | 1.5 | 1.833333 | 15.5 | 30.333333 | 2.333333 | 0.666667 | 1.67E-01 | 1.666667 | 6 | 6.666667 | 0 | 66.666667 |
| C5 | 0.25 | 0.75 | 9.75 | 7.5 | 0.5 | 0.5 | 5.00E-01 | 1.5 | 2.75 | 2.75 | 0.75 | 27.5 |

From the data analyses, you can see that Cluster 3 has the highest number of appealing venues and features in Toronto. Cluster 3 has the following neighborhoods, as provided in the table on the following page. Cluster 3 has neighborhoods specifically located in the Downtown Toronto region, and depending upon which features that the potential immigrant to Canada is interested in, there are several options for possible neighborhoods to move to. The Richmond neighborhood may be more appealing to someone who is looking for more options in arts and

entertainment, in addition to having plenty of fine dining options in the area (1 km to the center of the neighborhood).

| Neighborhood | Postcode | Borough | Latitude | Longitude | Arts and Entertainment | Athletics and Sports | Educational | Food and Dine | Gym and Fitness Center | Health and Beauty Service |
|---|---|---|---|---|---|---|---|---|---|---|
| Church and Wellesley | M4Y | Downtown Toronto | 43.66586 | -79.38316 | 9 | 0 | 50 | 70 | 2 | 2 |
| Ryerson,Garden District | M5B | Downtown Toronto | 43.657162 | -79.378937 | 6 | 0 | 50 | 65 | 2 | 1 |
| St. James Town | M5C | Downtown Toronto | 43.651494 | -79.375418 | 8 | 0 | 50 | 65 | 3 | 0 |
| Central Bay Street | M5G | Downtown Toronto | 43.657952 | -79.387383 | 9 | 0 | 51 | 63 | 2 | 2 |
| Adelaide,King,Richmond | M5H | Downtown Toronto | 43.650571 | -79.384568 | 13 | 0 | 51 | 60 | 4 | 2 |
| Design Exchange,Toronto Dominion Centre | M5K | Downtown Toronto | 43.647177 | -79.381576 | 8 | 2 | 48 | 71 | 3 | 1 |
| Commerce Court,Victoria Hotel | M5L | Downtown Toronto | 43.648198 | -79.379817 | 7 | 2 | 49 | 70 | 3 | 0 |
| Harbord,University of Toronto | M5S | Downtown Toronto | 43.662696 | -79.400049 | 8 | 0 | 49 | 71 | 0 | 0 |
| Chinatown,Grange Park,Kensington Market | M5T | Downtown Toronto | 43.653206 | -79.400049 | 8 | 0 | 44 | 74 | 3 | 0 |
| Stn A PO Boxes 25 The Esplanade | M5W | Downtown Toronto | 43.646435 | -79.374846 | 6 | 2 | 44 | 69 | 3 | 0 |
| First Canadian Place,Underground city | M5X | Downtown Toronto | 43.648429 | -79.38228 | 9 | 2 | 49 | 71 | 3 | 2 |

**Discussion**

As can be confirmed by the results above, Cluster 3 is the most desirable area in terms of the total venues and features available according to the Four Square database. The specific breakdown of venues and features can be viewed in terms of the categories per column and geographical location, and the potential immigrant to Canada can determine which categories are more critical to his or her needs in the neighborhood he or she would like to move to.

## Conclusion

Determining whether to make a major move to an entirely different country such as Canada can be overwhelming. This project took a preliminary step in making a tool that would help simplify the decision-making process for an individual who would consider moving to Toronto, Canada as an example. This methodology can be applied to other locations as well, and help create a more methodical approach in studying the geographical appeal to moving to a brand new location. What would be helpful in a future supplementary study is looking at the financial aspects of the five different clusters and seeing which would be the most cost-effective neighborhood. Affordability wasn't taken into consideration during this study, as we focused upon the venues offered per neighborhood across Toronto. Still, being able to apply Data Science techniques in a practical manner is extremely useful in creating a recommendation tool for potential immigrants and helping to alleviate concerns about which borough would be the best fit for someone moving to Toronto, Canada.