Capstone Project Final Report

IBM Data Science Professional Certificate

Kaylene Robinson

December 2019

1. Introduction

1.1 Background

Toronto, the capital of Ontario Province, is the most populated city on Canada, with a span of 2,731,571 citizens. Toronto also serves as one of the worlds most diverse place to live and is a primary destinations for Canadian immigrants.

As a result, Toronto's geographical locations is split into different municiplaities. Each municipality has a an historical identity and are represented in their names. For example, East York, Etobicoke, Forest HIll, Mimico, and North York are few of hundred neighborhoods in Toronto.

1.2 Problem

In this project we examined Toronto neighbohoods to group them into similar clusters. Some factors we determined may influence these clusters were restaurants, events, parks and schools.

1.3 Interest

Those who would potentially intersested in the school may be people looking for new residency, emigrating from a different country, or tourists. This study may help those prospects make a decision on what attributes their new neighborhood have, and can help them make the best choice.

2. Data Cleaning & Selection

2.1 Data Sources

Geographical data for the Toronto neighborhoods was obtained through this Wikipedia link: https://en.wikipedia.org/wiki/List_of_postal_codes_of_Canada:_M (https://en.wikipedia.org/wiki/List_of_postal_codes_of_Canada:_M).

We also extracted geographical coordinates of post codes from the following CSV file: http://cocl.us/Geospatial_data (http://cocl.us/Geospatial_data).

These links contain data that details the locations used to receive information for venues in the area. Following this data, we also requested venue data for each neighborhood using Foursquare API to cluster the neighborhoods.

2.2 Data Cleaning

The data downloaded from the aforementioned sources were combined into one table. Using pandas, the data was coded into a data frame and all rows labeled "Not Assigned" were ignores. Neighborhoods with the dame postal code were merged.

2.3 Feature Selection

A table with the postal code, borough, neighborhood, latitude and longitude was created. The table shown below consisted of 103 rows and 5 columns.

Table 1

	Postcode	Borough	Neighbourhood	Latitude	Longitude
0	M4W	Downtown Toronto	Rosedale	43.6796	-79.3775
1	M4X	Downtown Toronto	Cabbagetown,St. James Town	43.668	-79.3677
2	M4Y	Downtown Toronto	Church and Wellesley	43.6659	-79.3832
3	M5A	Downtown Toronto	Harbourfront	43.6543	-79.3606
4	M5B	Downtown Toronto	Ryerson,Garden District	43.6572	-79.3789

3. Methodology

3.1 Exploratory Data Analysis

Using the exploratory data analysis approach, I was able to visualize the dataset and summarize their main characteristics. Once the data was cleaned, features from the data set were identified. To represent the data we used the BeautifulSoup python package to read the Wikipedia data and transform it into a data frame using pandas.

We then extracted the CSV file for the geographical coordinates, cleaned this data, and combined it the data frames together.

3.2 Data Exploration

Using the python folium library, we were able to generate a visual map of the Toronto neighborhoods.

Using Foursquare API we were able to explore the venues in each neighborhood. We found the top 50 venues of each neighborhood and created a new data frame to display these venues.

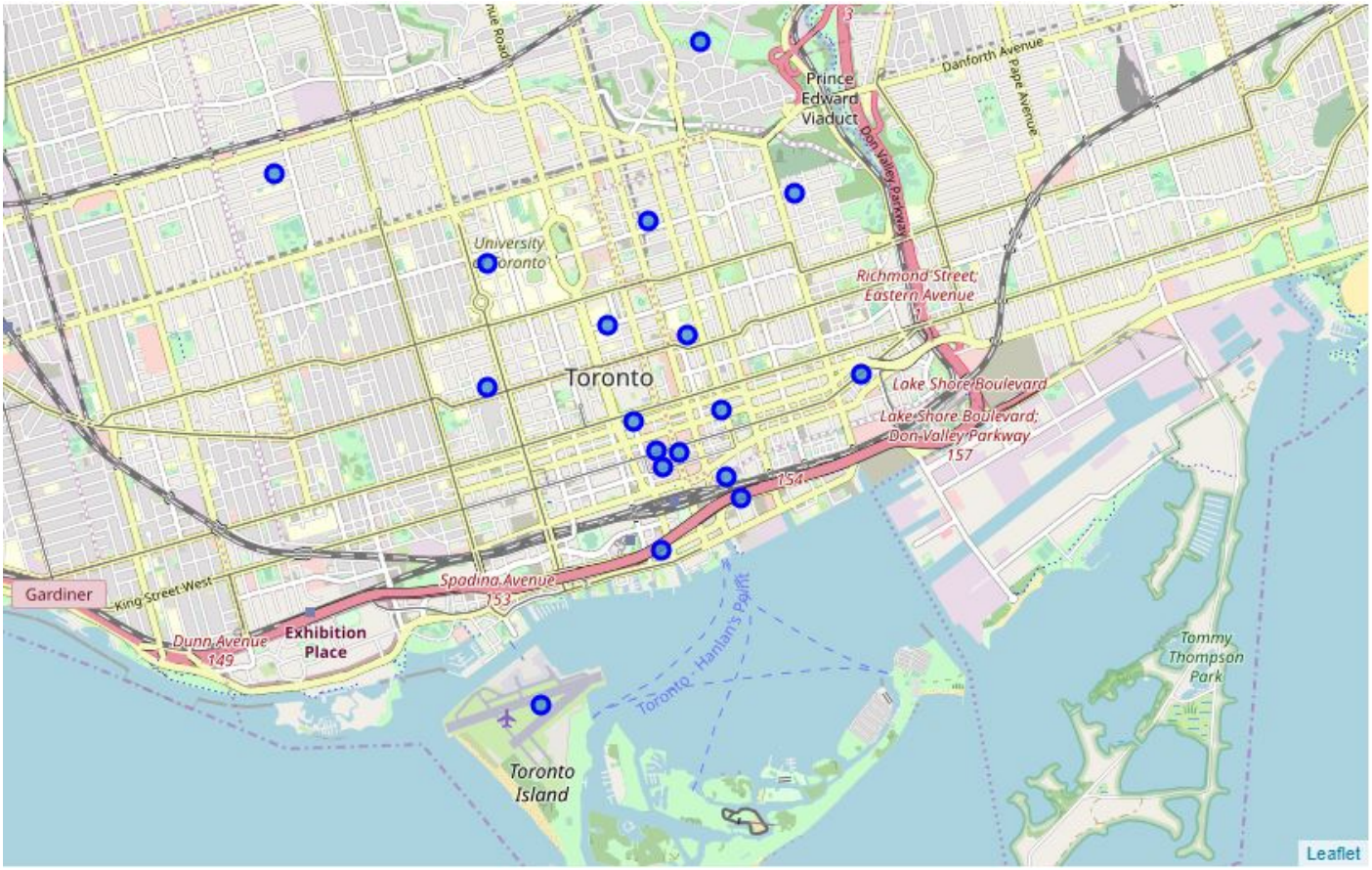
Tabel 2

	Neighborhood	1st Most Common Venue	2nd Most Common Venue	3rd Most Common Venue	4th Most Common Venue	5th Most Common Venue	6th Most Common Venue	7th Most Common Venue	8th Most Common Venue	9th Most Common Venue	10th Most Common Venue
0	Adelaide,King,Richmond	Coffee Shop	Café	Steakhouse	Thai Restaurant	Bar	Restaurant	Bakery	Burger Joint	Asian Restaurant	Sushi Restaurant
1	Berczy Park	Coffee Shop	Bakery	Steakhouse	Cheese Shop	Beer Bar	Seafood Restaurant	Cocktail Bar	Farmers Market	Café	Museum
2	CN Tower,Bathurst Quay,Island airport,Harbourf...	Airport Terminal	Airport Lounge	Airport Service	Harbor / Marina	Boat or Ferry	Boutique	Plane	Coffee Shop	Sculpture Garden	Airport Gate
3	Cabbagetown,St. James Town	Coffee Shop	Park	Pub	Italian Restaurant	Café	Pizza Place	Bakery	Restaurant	Gym / Fitness Center	Playground
4	Central Bay Street	Coffee Shop	Italian Restaurant	Ice Cream Shop	Sandwich Place	Burger Joint	Café	Dessert Shop	Bakery	Bar	Chinese Restaurant

4. Results

Below the clusters of each neighborhood are illustrated following with a map pinpointing each neighborhood.

Folium Map



5. Discussion & Observations

Based on the results of the most common venues in each neighborhood, we categorized similar neighborhoods into 6 clusters. Organizing the clusters can help those interested choose which area is best for them.

Table 3

Neighborhood	1st Most Common Venue	2nd Most Common Venue	3rd Most Common Venue	4th Most Common Venue	5th Most Common Venue	6th Most Common Venue	7th Most Common Venue	8th Most Common Venue	9th Most Common Venue	10th Most Common Venue
Adelaide,King,Richmond	Coffee Shop	Cafe	Steakhouse	Thai Restaurant	Bar	Restaurant	Bakery	Burger Joint	Asian Restaurant	Sushi Restaurant
Berczy Park	Coffee Shop	Bakery	Steakhouse	Cheese Shop	Beer Bar	Seafood Restaurant	Cocktail Bar	Farmers Market	Cafe	Museum
CN Tower,Bathurst Quay,Island airport,Harbourfront West,King and Spadina,Railway Lands,South Niagara	Airport Terminal	Airport Lounge	Airport Service	Harbor / Marina	Boat or Ferry	Boutique	Plane	Coffee Shop	Sculpture Garden	Airport Gate
Cabbagetown,St. James Town	Coffee Shop	Park	Pub	Italian Restaurant	Cafe	Pizza Place	Bakery	Restaurant	Gym / Fitness Center	Playground
Central Bay Street	Coffee Shop	Italian Restaurant	Ice Cream Shop	Sandwich Place	Burger Joint	Cafe	Dessert Shop	Bakery	Bar	Chinese Restaurant
Chinatown,Grange Park,Kensington Market	Cafe	Bar	Dumpling Restaurant	Vietnamese Restaurant	Chinese Restaurant	Coffee Shop	Mexican Restaurant	Bakery	Gaming Cafe	Burger Joint
Christie	Grocery Store	Cafe	Park	Athletics & Sports	Italian Restaurant	Diner	Nightclub	Convenience Store	Candy Store	Restaurant
Church and Wellesley	Coffee Shop	Japanese Restaurant	Sushi Restaurant	Gay Bar	Restaurant	Cafe	Pub	Gym	Hotel	Gastropub
Commerce Court,Victoria Hotel	Coffee Shop	Cafe	Hotel	Restaurant	American Restaurant	Seafood Restaurant	Bakery	Steakhouse	Gym	Deli / Bodega
Design Exchange,Toronto Dominion Centre	Coffee Shop	Cafe	Hotel	Bar	Restaurant	American Restaurant	Seafood Restaurant	Steakhouse	Italian Restaurant	Gastropub
First Canadian Place,Underground city	Coffee Shop	Cafe	Steakhouse	Restaurant	Hotel	Asian Restaurant	Burger Joint	Deli / Bodega	Japanese Restaurant	American Restaurant
Harbord,University of Toronto	Cafe	Restaurant	Sandwich Place	Bookstore	Japanese Restaurant	Italian Restaurant	Bar	Bakery	Sushi Restaurant	Nightclub
Harbourfront	Coffee Shop	Pub	Bakery	Park	Mexican Restaurant	Breakfast Spot	Cafe	Theater	Dessert Shop	Brewery
Harbourfront East,Toronto Islands,Union Station	Coffee Shop	Aquarium	Hotel	Cafe	Fried Chicken Joint	Restaurant	Brewery	Scenic Lookout	Italian Restaurant	Bakery
Rosedale	Park	Playground	Trail	Wings Joint	Department Store	Eastern European Restaurant	Dumpling Restaurant	Donut Shop	Diner Restaurant	Dog Run
Ryerson,Garden District	Coffee Shop	Clothing Store	Fast Food Restaurant	Middle Eastern Restaurant	Cafe	Cosmetics Shop	Bakery	Sporting Goods Shop	Japanese Restaurant	Italian Restaurant
St. James Town	Cafe	Coffee Shop	Restaurant	Clothing Store	Hotel	Italian Restaurant	Bakery	Beer Bar	Diner	Cosmetics Shop
Stn A PO Boxes 25 The Esplanade	Coffee Shop	Restaurant	Cafe	Hotel	Beer Bar	Seafood Restaurant	Japanese Restaurant	Italian Restaurant	Bakery	Cheese Shop

6. Conclusion

In this project, we analyzed and prepared data to explore neighborhoods in Toronto. Using K-mean clustering, a machine learning algorithm we were able to cluster each neighborhood. Also, we provided suggestions to those who would be interested in living or traveling in Toronto. Furthermore, one may be able to use other algorithms as a means to cluster neighborhood and may compare the different results of each algorithm.