Finding the best Neighborhoods for Foreign Students in Toronto

Capstone Project

Applied Data Science (IBM Certification)

MUHAMMAD ABAIDULLAH ANWAR

FEBRUARY 2021

Presentation Rundown

- Introduction
- Problem Description
- Data Description
- Methodology
- Results and Discussion
- Conclusion

Introduction

This Capstone Project is developed for IBM Certification Course on Applied Data Science offered by Coursera.

Toronto being the capital city of the Canadian province of Ontario has been recorded as one of the most populous city.

Population of 6,254,571 as of 2021 in Canada and the fourth most populous city in North America.

The city is surrounding the western end of Lake Ontario is an international center of business, finance, arts, and culture, education, and is recognized as one of the most multicultural and cosmopolitan cities in the world.

Problem Description

The immigrants during their initial few years and the foreign students always need to find a convenient locality to live which is close to the educational institutions, restaurants of their choices, and shopping facilities.

Objective: Identify the neighbourhoods of Toronto for the foreign students from India and Pakistan, particularly, to find a reasonable cost effective residence which is near to their universities, Indian or Pakistani dining facilities, and the shopping malls.

There are many cuisines which are difficult to distinguish as Indian or Pakistani and therefore it will mainly serve for both types of foreign students.

Data Description

City: Toronto (Canada)

Neighborhoods and Postal Code Data URL:
https://en.wikipedia.org/wiki/List_of_postal_codes_of_Canada:_M

Coordinate Data URL:

http://cocl.us/Geospatial_data

Category ID Data URL:

https://en.wikipedia.org/wiki/List_of_neighbourhoods_in_Toronto

Methodology

The Python is used to explore and analyze the data for multiple purpose. The necessary and required packages and libraries being used in the Capstone Project will be installed and imported.

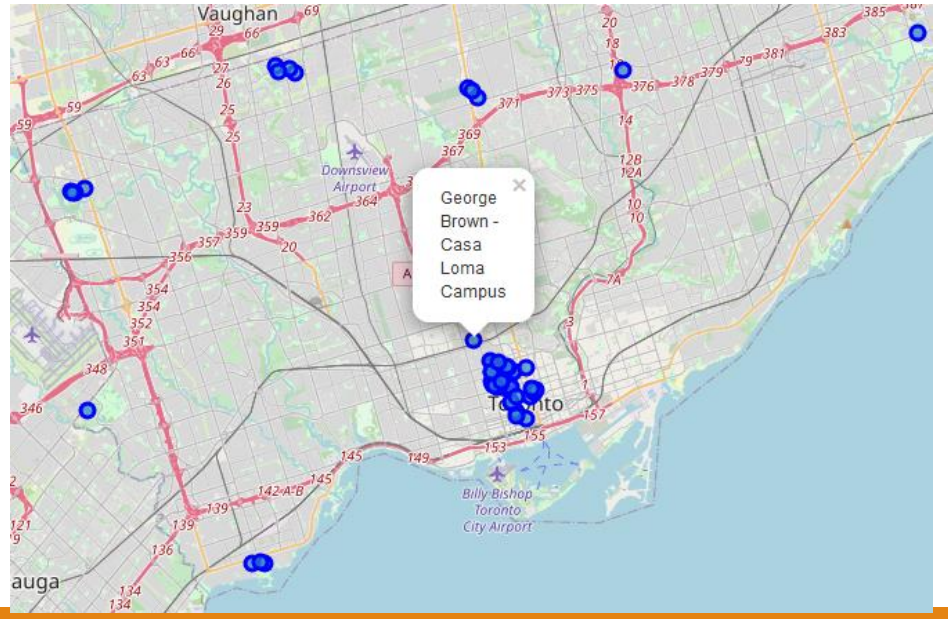
This data website provides a list of postal codes in Canada where the first letter is M. Postal codes beginning with M are located within the city of Toronto in the province of Ontario.

We used and processed the cells that have an assigned borough and ignore cells with a borough that is Not assigned. Also if a cell had a borough but a "Not assigned" neighborhood, then the neighborhood was assigned the same as the borough.

Methodology (Cont'd)

Since we needed the only the borough, neighbourhood, postal codes and geolocations (latitude and longitude) therefore end up selecting the columns that we needed.

We used the Colleges and University Category ID to show the map of Toronto city with 56 higher educational institutions.



Methodology (Cont'd)

Similarly, we wrote the codes for finding and displaying the Indian restaurants (143), Pakistani restaurants (13), and Halal restaurants (58) in the Toronto city.

We also explored the region of interest of Toronto (DENC) as follows:

- **D - Downtown**
- **E - East**
- **N - North**
- **C – Central**

Methodology (Cont'd)

One Hot Encoding to work with our categorical datatype of the venue categories. This was required to convert the categorical data into numeric data.

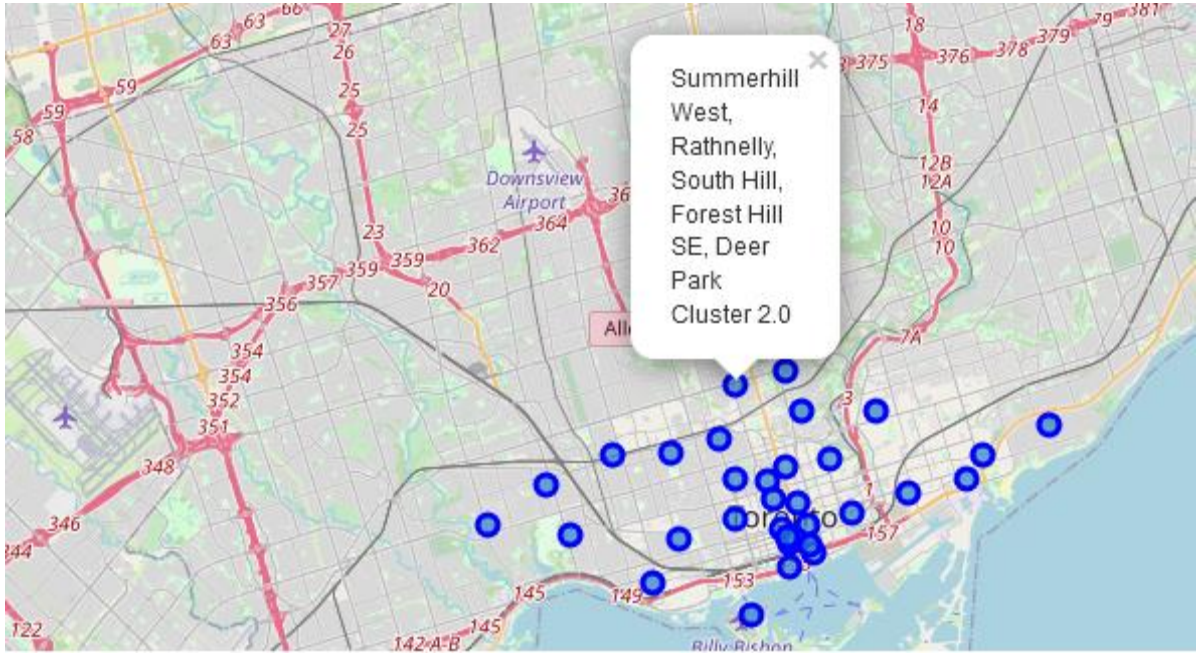
We identified, ranked, and label the top ten (10) venues categories in our neighborhood. A snapshot of the output is shown here.

:

	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	Berczy Park	College Classroom	Indian Restaurant	Student Center	North Indian Restaurant	Community College	General College & University	University	College Math Building	College Library	College Lab
1	Brockton, Parkdale Village, Exhibition Place	Trade School	College Theater	College Lab	College Residence Hall	College Rec Center	College Quad	College Math Building	College Library	College Gym	College Football Field
2	Central Bay Street	College Academic Building	Indian Restaurant	University	Student Center	College Science Building	College Administrative Building	College Lab	Medical School	Government Building	College & University
3	Christie	College Classroom	University	College Rec Center	College Quad	College Math Building	College Library	College Lab	College Gym	College Football Field	College Engineering Building
4	Church and Wellesley	University	Indian Restaurant	General College & University	College Communications Building	Halal Restaurant	High School	College Residence Hall	Trade School	Performing Arts Venue	College Track

Clustering using KMeans and Visualization

Machine Learning algorithm KMeans to build five (5) clusters of neighbourhoods. The clusters are shown in here.



Results and Discussion

We generated five (05) clusters and the largest cluster is fourth cluster that includes thirteen (13) neighborhoods and the second largest cluster is the third cluster which includes seven (07) clusters.

The clusters include the information of higher educational institutions and Indian, Pakistani, and Halal restaurants in the Toronto City.

Conclusion

We presented the analysis of Colleges, Universities, and different restaurants that has always been useful information for foreign students when come to Toronto for their studies.

This information is also useful for the expatriate families living in the city.

We explored both the cities based on their postal codes and then extrapolated the common venues present in each of the neighborhoods finally concluding with clustering similar neighborhoods together.

We could see that each of the neighborhoods in both the cities have a wide variety of experiences to offer which is unique in its own way.