CAPSTONE PROJECT

THE BATTLE OF NEIGHBOURHOOD

PRESENTATION

BHUPENDRA KUMAR YADAV

INTRODUCTION

BACKGROUND

Toronto demographics show that it is a large and ethnically diverse metropolis. It is the largest city in the Canada with a long history of international immigration. Toronto was home to nearly 2.93 million people in 2018. Over the last decade, the city has been growing faster than the region. The Toronto region continues to be by far the leading metropolitan gateway for legal immigrants admitted into the Canada.

This final project explores the best locations for restaurants throughout the city of Toronto. Potentially the owner of the new restaurant can have great success and consistent profit. However, as with any business, opening a new restaurant requires serious considerations and is more complicated than it seems from the first glance. In particular, the location of the restaurant is one of the most important factors that will affect whether it will have success or a failure. So our project will attempt to answer the questions “Where should the investor open a Restaurant?”, “Where should I go If I want good restaurant?” and “Which type of restaurant is most famous?”

BUSINESS PROBLEM

The objective of this Capstone project is to analyse and select the best locations in the city of Toronto to open a new restaurant. Using Data Science methodology and instruments such as Data Analysis and Visualization, this project aims to provide solutions to answer the business question: Where in the city of Toronto, where should the investor open a new Restaurant?

INTEREST

Expats who are considering to open Restaurant in Toronto will be interested to identify the place in Toronto and explore its neighbourhoods and common venues, types of restaurant in Toronto.

DATA ACQUISITION AND CLEANING

Data Acquisition

The data acquired for this project is a combination of data from two sources.

The first source of data is scraped from a Wikipedia page that contains the list of Toronto boroughs. This page contains additional information about the boroughs, the following are the columns:

- Postal Code: Postal code of Neighbourhoods
- Borough: Name of Borough
- Neighbourhood: Name of Neighbourhoods
- Link: https://en.wikipedia.org/wiki/List_of_postal_codes_of_Canada:_M

The second data source of the project uses a Toronto latitude and longitude according to the Postal Code. The dataset contains the following columns:

- Postal Code: Postal code of Neighbourhoods
- Latitude: Latitude of Neighbourhoods
- Longitude: Longitude of Neighbourhood
- Link: http://cocl.us/Geospatial_data

Data Cleaning

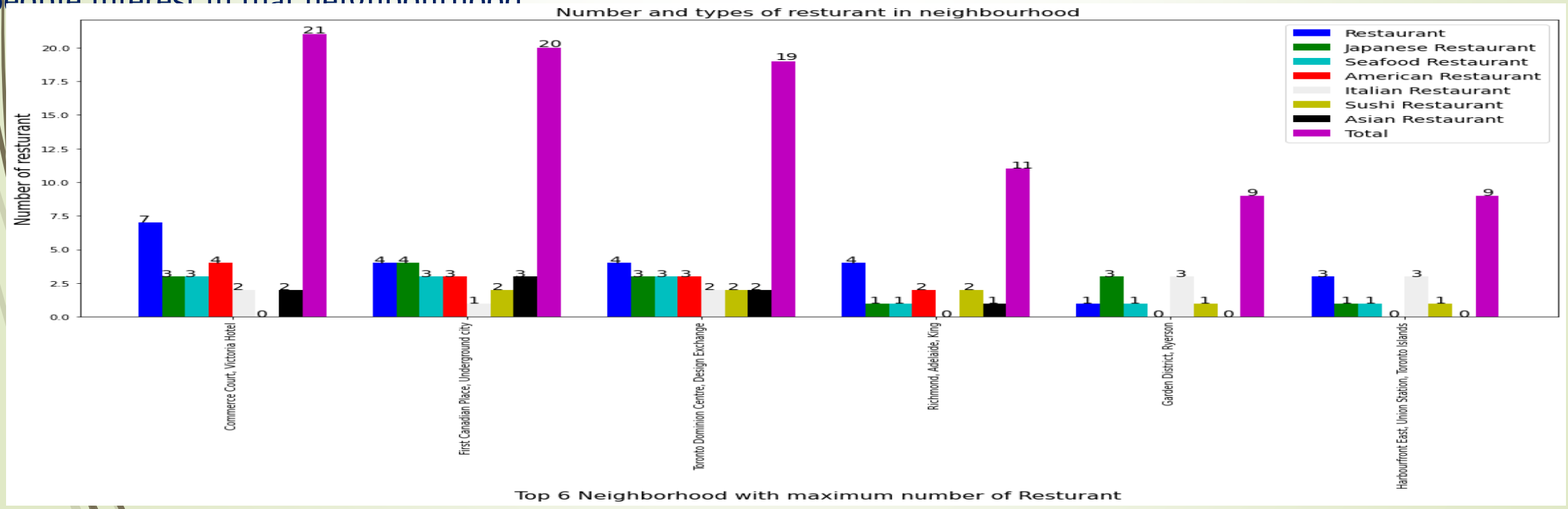
- The data preparation for each of the two sources of data is done separately. Neighbourhoods are merged according to their Postal Code.
- The second data is scraped from a Wikipedia page using the Beautiful Soup library in python. Using this library we can extract the data in the tabular format as shown in the website. After the web scraping, string manipulation is required to get the names of the boroughs in the correct form. This is important because we will be merging the two datasets together using Postal Code.
- The two datasets are merged on the Postal Code to form a new dataset that combines the necessary information in one dataset. The purpose of this dataset is to select the neighbourhoods only in Toronto.
- As dataframe is sorted only for Toronto, now we can explore Toronto. Folium is used visualise neighbourhood of Toronto on map.
- Foursquare API is used get venues and venue category of all neighbourhoods of Toronto. Two new columns with venue and its category is added.
- Using previous dataframe, venues category is sorted which are present in high number. It is done to obtain the trend of venue category in these neighbourhood.
- Now new dataset is formed with only restaurant and neighbourhood. It is sorted according to neighbourhood with maximum number of restaurant. Along with these type of restaurant which are in high number in each neighbourhood is obtained.

METHODOLOGY

EXPLORATORY DATA ANALYSIS

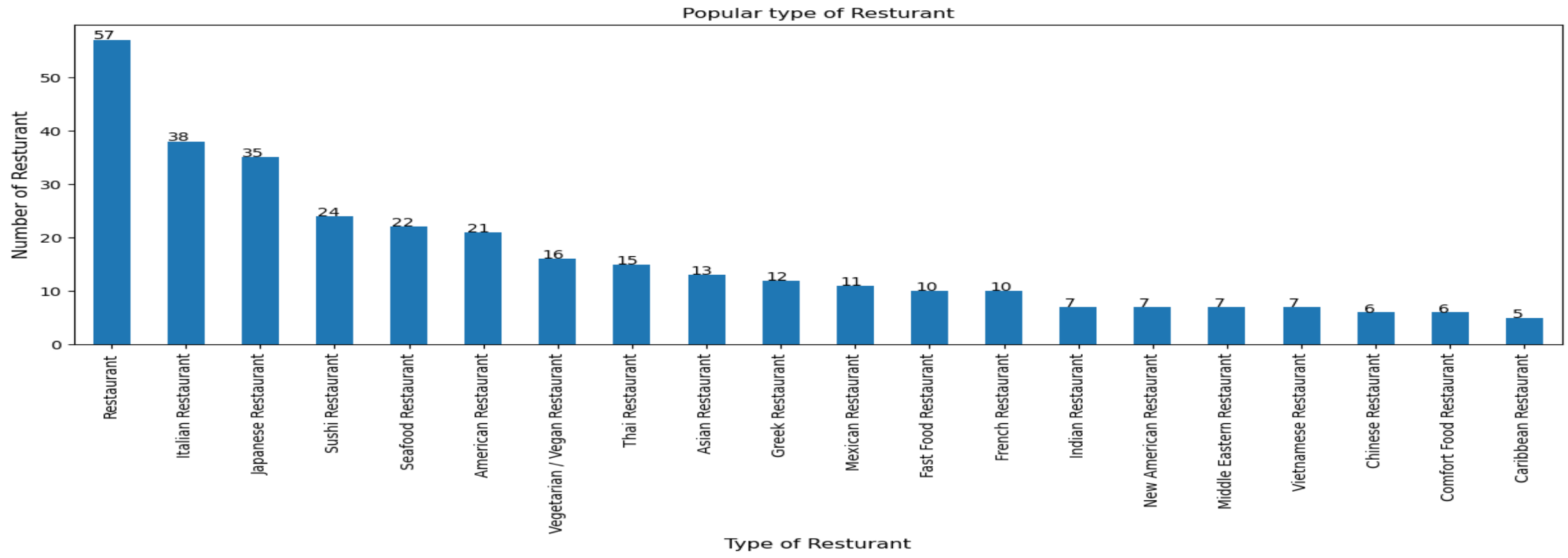
Graphical analysis of Top 6 Neighbourhoods

Matplotlib function is used to plot top 6 neighbourhood which have maximum number of restaurant. Type of restaurant along with it number in each neighbourhood is also plotted in bar graph. It will give clear idea about people interest in that neighbourhood



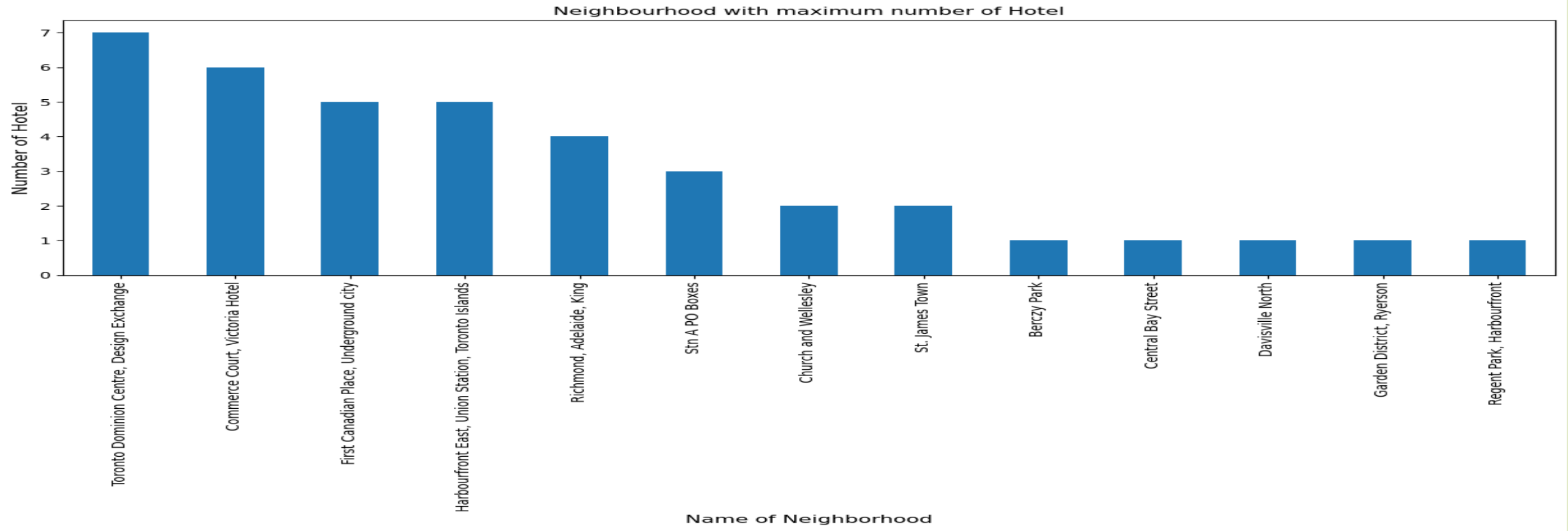
Popular type of restaurant

By plotting the dataframe we can visualize which type of restaurant has the highest number. It is clearly visible that Italian and Japanese restaurant are more famous in Toronto followed by Seafood and American.



Neighbourhoods with maximum number of hotel

Location of restaurant is affected by many factor. If it located near to the hotel then probability of visitor is very high. Toronto Dominion Centre, Commerce court and First Canadian place have most number of hotel.



As dataframe is sorted only for Toronto, now we can explore Toronto. Folium is used visualise neighbourhood of Toronto on map.

Foursquare API is used get venues and venue category of all neighbourhoods of Toronto. Two new columns with venue and its category is added.

MODELLING

Using the final dataset containing the neighbourhoods in Toronto along with the latitude and longitude, we can find all the venues within a 500 meter radius of each neighbourhoods by connecting to the Foursquare API. This returns a json file containing all the venues in each neighbourhoods which is converted to a pandas dataframe. This data frame contains all the venues along with their coordinates and category

Out[159]:

	Neighborhood	Neighborhood Latitude	Neighborhood Longitude	Venue	Venue Latitude	Venue Longitude	Venue Category
0	Regent Park, Harbourfront	43.65426	-79.360636	Roselle Desserts	43.653447	-79.362017	Bakery
1	Regent Park, Harbourfront	43.65426	-79.360636	Tandem Coffee	43.653559	-79.361809	Coffee Shop
2	Regent Park, Harbourfront	43.65426	-79.360636	Cooper Koo Family YMCA	43.653249	-79.358008	Distribution Center
3	Regent Park, Harbourfront	43.65426	-79.360636	Body Blitz Spa East	43.654735	-79.359874	Spa
4	Regent Park, Harbourfront	43.65426	-79.360636	Impact Kitchen	43.656369	-79.356980	Restaurant
5	Regent Park, Harbourfront	43.65426	-79.360636	Corktown Common	43.655618	-79.356211	Park
6	Regent Park, Harbourfront	43.65426	-79.360636	The Distillery Historic District	43.650244	-79.359323	Historic Site
7	Regent Park, Harbourfront	43.65426	-79.360636	Morning Glory Cafe	43.653947	-79.361149	Breakfast Spot
8	Regent Park, Harbourfront	43.65426	-79.360636	The Extension Room	43.653313	-79.359725	Gym / Fitness Center
9	Regent Park, Harbourfront	43.65426	-79.360636	Dominion Pub and Kitchen	43.656919	-79.358967	Pub

One hot encoding is done on the venues data. (One hot encoding is a process by which categorical variables are converted into a form that could be provided to ML algorithms to do a better job in prediction). The Venues data is then grouped by the Neighbourhoods and the mean of the venues are calculated, finally the 10 common venues are calculated for each of the neighbourhoods.

Out[151]:

	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	Toronto Dominion Centre, Design Exchange	Hotel	Restaurant	American Restaurant	Seafood Restaurant	Japanese Restaurant	Concert Hall	Sushi Restaurant	Italian Restaurant	Office	Thai Restaurant
1	Commerce Court, Victoria Hotel	Restaurant	Hotel	American Restaurant	Seafood Restaurant	Japanese Restaurant	Italian Restaurant	Concert Hall	Office	Thai Restaurant	Sushi Restaurant
2	Forest Hill North & West, Forest Hill Road Park	Sushi Restaurant	Concert Hall	Office	Hotel	Thai Restaurant	American Restaurant	Seafood Restaurant	Japanese Restaurant	Italian Restaurant	Restaurant
3	Moore Park, Summerhill East	Restaurant	Concert Hall	Office	Hotel	Thai Restaurant	American Restaurant	Seafood Restaurant	Sushi Restaurant	Japanese Restaurant	Italian Restaurant
4	First Canadian Place, Underground city	Hotel	Japanese Restaurant	Restaurant	American Restaurant	Seafood Restaurant	Concert Hall	Sushi Restaurant	Italian Restaurant	Office	Thai Restaurant

RESULTS

After running the K-means clustering we can access each cluster created to see which neighbourhoods were assigned to each of the five clusters. Looking into the neighbourhoods in the first cluster

The cluster one is the biggest cluster with 9 of the 15 neighbourhoods in the Toronto. Upon closely examining these neighbourhoods we can see that the most common venues in these neighbourhoods are Restaurants, Seafood Restaurants, Hotels and Italian Restaurants. Looking into the neighbourhoods in the second, third and fifth clusters, we can see these clusters have only one neighbourhoods in each. This is because of the unique venues in each of the neighbourhoods, hence they couldn't be clustered into similar neighbourhoods

Out[66]:

	Borough	Neighborhood	Cluster Labels	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	Downtown Toronto	Regent Park, Harbourfront	0	Restaurant	Concert Hall	Office	Hotel	Thai Restaurant	American Restaurant	Seafood Restaurant	Sushi Restaurant	Japanese Restaurant	Italian Restaurant
1	Downtown Toronto	Queen's Park, Ontario Provincial Government	0	Sushi Restaurant	Japanese Restaurant	Italian Restaurant	Concert Hall	Office	Hotel	Thai Restaurant	American Restaurant	Seafood Restaurant	Restaurant
2	Downtown Toronto	Garden District, Ryerson	0	Japanese Restaurant	Italian Restaurant	Office	Hotel	Thai Restaurant	Seafood Restaurant	Sushi Restaurant	Restaurant	Concert Hall	American Restaurant
4	East Toronto	The Beaches	0	Concert Hall	Office	Hotel	Thai Restaurant	American Restaurant	Seafood Restaurant	Sushi Restaurant	Japanese Restaurant	Italian Restaurant	Restaurant
6	Downtown Toronto	Central Bay Street	0	Italian Restaurant	Japanese Restaurant	Office	Hotel	Thai Restaurant	Sushi Restaurant	Concert Hall	American Restaurant	Seafood Restaurant	Restaurant
9	West Toronto	Dufferin, Dovercourt Village	0	Concert Hall	Office	Hotel	Thai Restaurant	American Restaurant	Seafood Restaurant	Sushi Restaurant	Japanese Restaurant	Italian Restaurant	Restaurant
17	East Toronto	Studio District	0	American Restaurant	Thai Restaurant	Seafood Restaurant	Italian Restaurant	Concert Hall	Office	Hotel	Sushi Restaurant	Japanese Restaurant	Restaurant
18	Central Toronto	Lawrence Park	0	Concert Hall	Office	Hotel	Thai Restaurant	American Restaurant	Seafood Restaurant	Sushi Restaurant	Japanese Restaurant	Italian Restaurant	Restaurant

The third cluster has one neighbourhoods which consists of Venues such as Restaurant, Concert Hall and Office.

Out[62]:

	Borough	Neighborhood	Cluster Labels	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
21	Central Toronto	Forest Hill North & West, Forest Hill Road Park	2	Sushi Restaurant	Concert Hall	Office	Hotel	Thai Restaurant	American Restaurant	Seafood Restaurant	Japanese Restaurant	Italian Restaurant	Restaurant

The fourth cluster has only one neighbourhoods in it, these neighbourhoods have common venues such as Restaurant, Concert Hall, Office, Hotel etc. Visualising the clustered neighbourhoods on a map using the folium library

Out[63]:

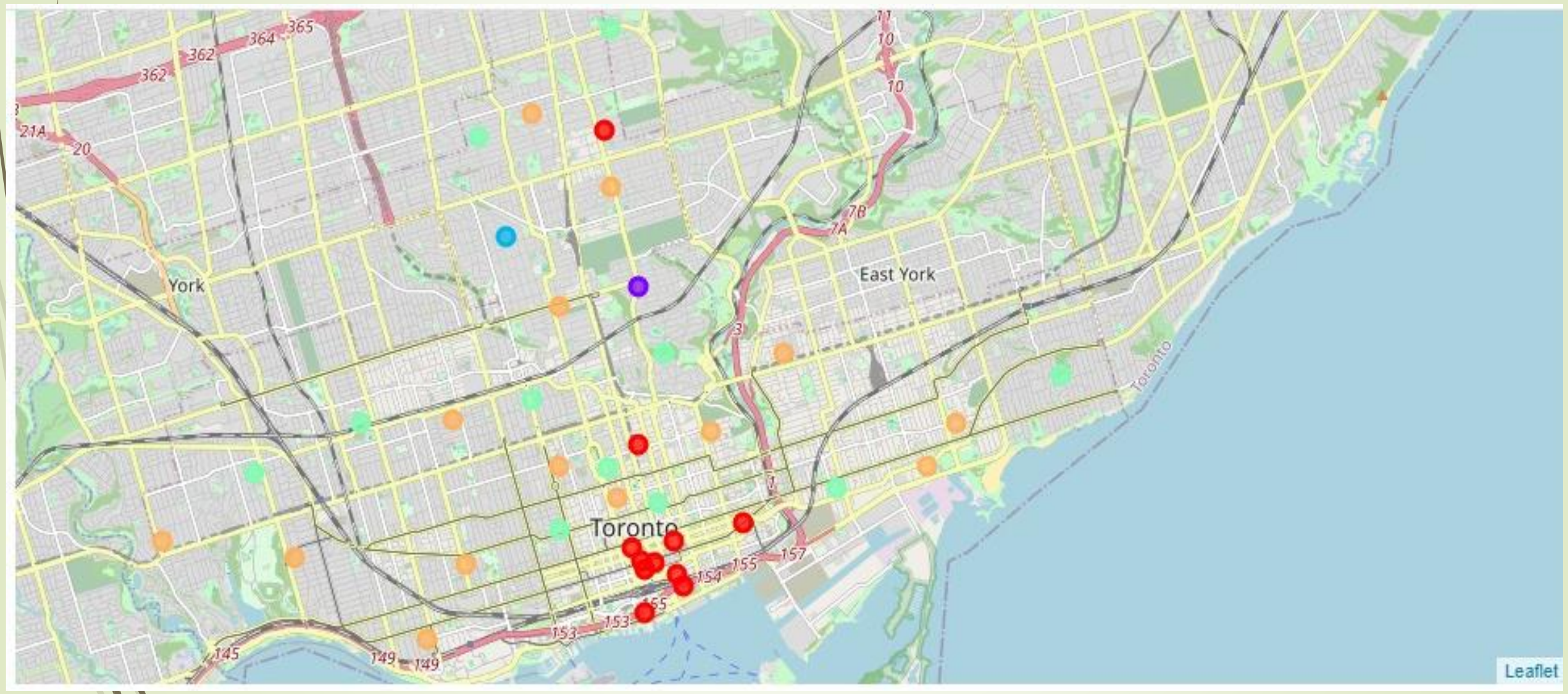
	Borough	Neighborhood	Cluster Labels	1st Most Common Venue	2nd Most Common Venue	3rd Most Common Venue	4th Most Common Venue	5th Most Common Venue	6th Most Common Venue	7th Most Common Venue	8th Most Common Venue	9th Most Common Venue	10th Most Common Venue
29	Central Toronto	Moore Park, Summerhill East	3	Restaurant	Concert Hall	Office	Hotel	Thai Restaurant	American Restaurant	Seafood Restaurant	Sushi Restaurant	Japanese Restaurant	Italian Restaurant

The fifth cluster has one neighbourhoods which consists of Venues such as Italian Restaurant, Sushi Restaurant

Out[68]:

	Borough	Neighborhood	Cluster Labels	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
7	Downtown Toronto	Christie	4	Italian Restaurant	Restaurant	Concert Hall	Office	Hotel	Thai Restaurant	American Restaurant	Seafood Restaurant	Sushi Restaurant	Japanese Restaurant
11	West Toronto	Little Portugal, Trinity	4	Restaurant	Japanese Restaurant	Italian Restaurant	Concert Hall	Office	Hotel	Thai Restaurant	American Restaurant	Seafood Restaurant	Sushi Restaurant
12	East Toronto	The Danforth West, Riverdale	4	Italian Restaurant	Restaurant	American Restaurant	Concert Hall	Office	Hotel	Thai Restaurant	Seafood Restaurant	Sushi Restaurant	Japanese Restaurant
14	West Toronto	Brockton, Parkdale Village, Exhibition Place	4	Italian Restaurant	Restaurant	Concert Hall	Office	Hotel	Thai Restaurant	American Restaurant	Seafood Restaurant	Sushi Restaurant	Japanese Restaurant
15	East Toronto	India Bazaar, The Beaches West	4	Sushi Restaurant	Italian Restaurant	Restaurant	Concert Hall	Office	Hotel	Thai Restaurant	American Restaurant	Seafood Restaurant	Japanese Restaurant
25	West Toronto	Parkdale, Roncesvalles	4	Italian Restaurant	Restaurant	Concert Hall	Office	Hotel	Thai Restaurant	American Restaurant	Seafood Restaurant	Sushi Restaurant	Japanese Restaurant
26	Central Toronto	Davisville	4	Sushi Restaurant	Italian Restaurant	Thai Restaurant	Seafood Restaurant	Restaurant	Concert Hall	Office	Hotel	American Restaurant	Japanese Restaurant
27	Downtown Toronto	University of Toronto, Harbord	4	Japanese Restaurant	Restaurant	Sushi Restaurant	Italian Restaurant	Concert Hall	Office	Hotel	Thai Restaurant	American Restaurant	Seafood Restaurant

Clustered neighbourhoods in City of Toronto each cluster is colour coded for the ease of presentation, we can see that majority of the neighbourhood's falls in the red cluster which is the first cluster. Three neighbourhoods have their own cluster (Light Green, Purple and Orange), these are clusters two three and five. The green cluster consists of only one neighbourhoods which is the 4th cluster.



DISCUSSION

The aim of this project is to help individual or businessmen to find a location to build a restaurant in a city of Toronto. Preference of choice may vary person to person. If a person want to open Italian Restaurant, American Restaurant or Japanese Restaurant then Downtown Toronto is best place to open it. Also we find that Italian and Japanese restaurant are more famous in Toronto followed by Seafood and American. Toronto Dominion Centre, Commerce court and First Canadian place these are the places with most number of Hotel, Offices and Concert Hall along with the variety of restaurant. These places have more number of people than other area so Restaurant is more likely to succeed here. We can take many other factor in account for more accurate prediction.

CONCLUSION

This project helps a person get a better understanding of the neighbourhoods with respect to the most common venues in that neighbourhood. It is always helpful to make use of technology to stay one step ahead i.e. finding out more about places to establish there restaurant business. We have just taken venue in our project for finding the best location for restaurant. In future more factor like parking, population, traffic and many other factor can be taken into account.