**Explore restaurants in Toronto**

Bharat Bhushan

Jan 19, 2021

**1. Introduction**

**1.1 Background**

Toronto is the capital city of the Canadian province of Ontario. it is the most populous city in Canada and the fourth most populous city in North America. Also, it is one of the main hub of business and tourism in Canada attracting lots of tourists from all around the world. As a part of Applied Capstone part in IBM Data Science course, I worked on this project where I explored different neighbourhoods in Toronto so as to get recommendations for a person planning to open a new restaurant in Toronto.

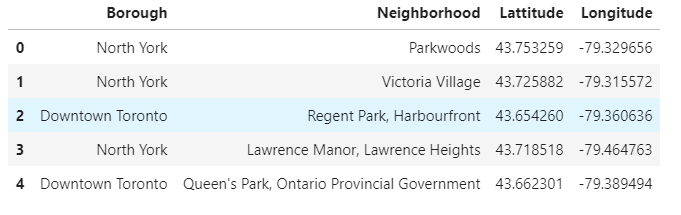
**1.2 Problem**

To explore different neighbourhoods in Toronto city using data fetched from Foursquare like list of restaurants in the neighbourhood, their categories and count in each neighbourhood and come up with useful insights to help a person who wants to open a new restaurant in Toronto in taking some of key decisions e.g. location, restaurant category.

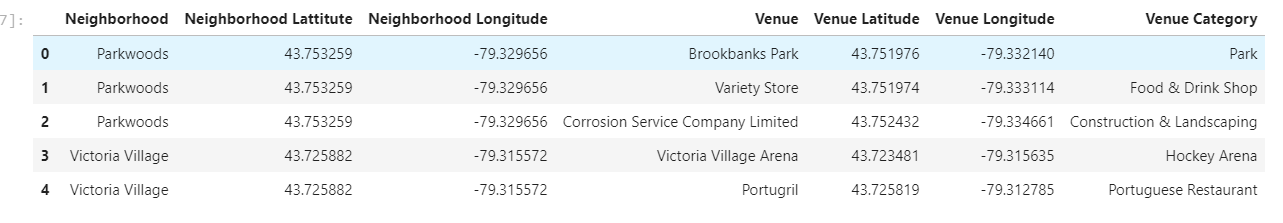
**2. Data collection and cleaning**

**2.1 Data sources**

I used a python web scraping code to collect information of different neighborhoods in Toronto from wiki page : [**Toronto Postal Codes**](https://en.wikipedia.org/wiki/List_of_postal_codes_of_Canada:_M). Then, I used excel containing co-ordinates of each neighbourhood in Toronto provided in IBM Data Science Capstone course on Coursera and combined this data to create a dataframe :



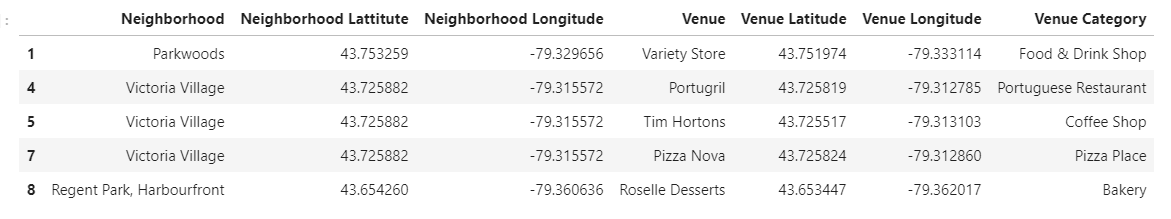
Using this data, I used Foursquare API to find venues of different categories in each neighbourhood in Toronto. It gave a dataset of 2100 venues :



**2.2 Data Cleaning**

As we are interested in restaurants business, next step is to segregate the data related to restaurants from all the venues.

Inspected the venue categories and found that most of the restaurants have a keyword from list of words e.g. ‘Diner’,’Coffee’,Pizza etc. Spent some time in identifying the keywords and then filtered data of all venues in Toronto to data related to restaurants. Final data has 1070 restaurants in Toronto city located in different neighbourhoods.



The data can be further analysed and transformed to derive the features needed to be fed to an unsupervised machine learning algorithm to generate groups of similar neighbourhoods.

**3. Methodology**

**3.1 Exploratory Data Analysis**

**Visualize all restaurants**

Used co-ordinates of different venues and created a map using Folium maps to visualize all restaurants using markers :

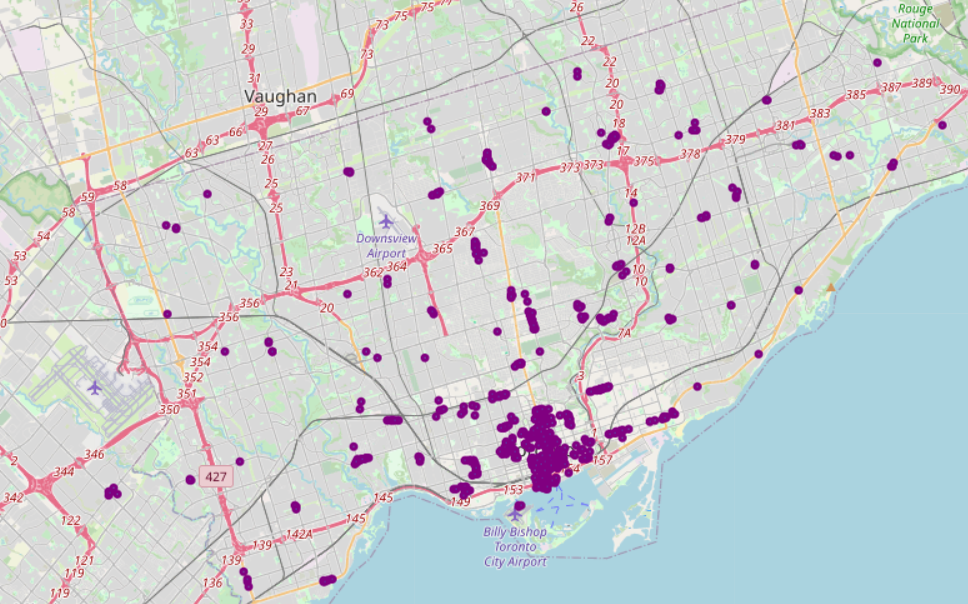


Fig 1 – Restaurants in Toronto

**Venue counts in neighbourhoods**

Bar plot in Fig2 below shows the counts of restaurants in top 20 neighbourhoods with highest count. Neighbourhoods in borough Downtown are the ones with highest number of restaurants as compared to other areas of the city.

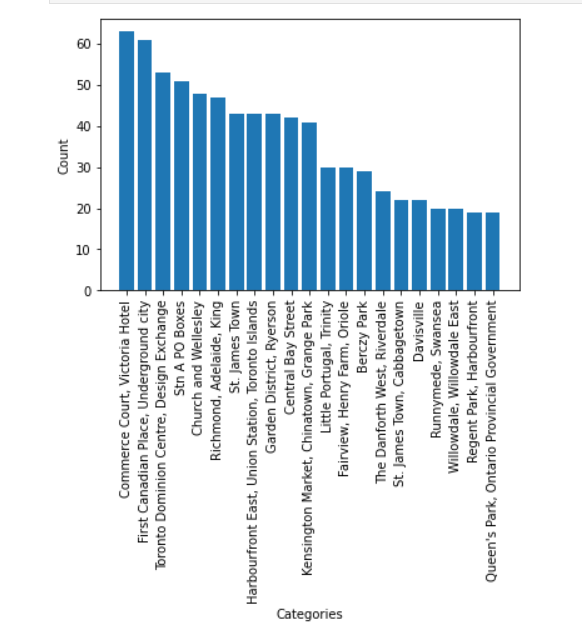


Fig2 – Counts of restaurants in Tornto neighbourhoods

**Categories of restaurants**

Bar plot in Fig3 below shows top 20 most common restaurant categories in city of Toronto. Coffee shops is the most common category. There is almost equal distribution of restaurants of different cuisines e.g. Japanese, Italian, Thai, Greek and American.

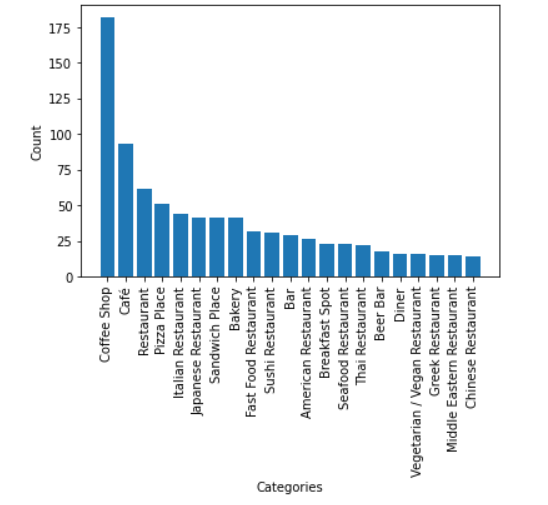


Fig3 – Counts of categories of restaurants

**3.2 Feature selection**

As restaurant category is categorical data, it can’t be used for building model. Converted this data to numeric using one hot encoding and then calculated mean value for each of the category for each neighbourhood creating 76 features.

Another feature is the number of venues in each of neighbourhood which is also calculated and added to the dataset.

This dataset can be given to a machine learning clustering algorithm to divide the data into different clusters based on these features for further study.

**3.3 Model generation**

Used k-means clustering to cluster the data into 2 clusters using above features. Data is divided into clusters of 68 and 13 neighborhoods. Fig4 shows the different groups of neighbourhoods in different colors on a Folium map.

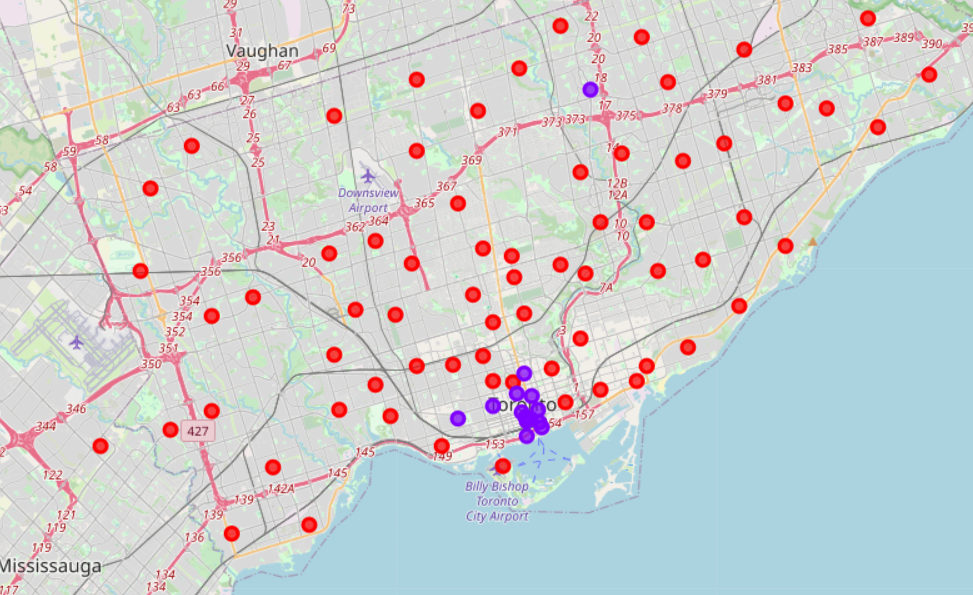


Fig4 – Clusters of neighbourhoods

**6. Results and Discussions**

Data is divided into 2 groups :

1. Most of the neighbourhoods [13 out of 82] in downtown seem to belong to this group. Number of restaurants in these neighbourhoods is relatively higher as compared to other neighbourhoods. Also, there are restaurants belonging to variety of categories/cuisines in most of the neighbourhoods. Coffee shop is the most common categories of restaurants. These neighbourhoods seem to attract diverse variety of customers. So, if you want to open a restaurant where number visiting customers is high but at the same time, the market competition is also high – then you can choose a neighbourhood from this category.
2. This is a bigger group and comprises of [68 out of 82] neighbourhoods. Most of these neighbourhoods have relatively very less number of restaurants. So, a neighbourhood with lesser number of customers and also less market competition can be chosen from this list.

Details related to neighbourhood in each group and most common categories in each neighbourhood are dumped in form of an excel file.

**6. Conclusion**

Analysed different neighbourhoods in city of Toronto based on restaurants in the city, categories of restaurant and count of restaurants in the neighbourhood. List of neighbourhoods is divided into two groups and results are visualized on a Folium map showing neighbourhoods of different groups in different colors. Pros and cons of choosing a neighbourhood from each group are also discussed.