**Battle of Neighborhoods – Report**

**1.Introduction**

**1.1 Background**

**Toronto**, city, capital of the province of [Ontario](https://www.britannica.com/place/Ontario-province), southeastern [Canada](https://www.britannica.com/place/Canada). It is the most populous city in Canada, a multicultural city, and the country’s financial and commercial center. The city is positioned on the edge of some of the best farmland in Canada, with a climate favorable to growing a wide range of crops, thereby making Toronto a [transportation](https://www.britannica.com/technology/transportation-technology), distribution, and manufacturing center. Most importantly, its central location, along with a host of political policies favoring [international trade](https://www.britannica.com/topic/international-trade), places this city with the greatest economic ties to, and influence from, the United States.

**1.2 Problem**

A user lives in one part of Toronto, he is quite comfortable there as he is aware of the nearby places and loves the amenities and venues that exist in the neighborhood. However the user has a received a job offer from a great company. However, given the far distance from your current place you unfortunately must move if you decide to accept the offer. So the user expects us to help him determine neighbourhoods on the other side of the that are exactly the same as his current neighbourhood, and if not perhaps similar neighbourhoods that are at least closer to his new job.

**2.Data**

**2.1 Data Sources**

Since there was no ready made data available for the Neighbourhoods we scrape the data from the Wikipedia page given below.

<https://en.wikipedia.org/wiki/List_of_postal_codes_of_Canada:_M,>

The data scraped from this link had Postal Code, Borough and Neighbourhood.

For investigating the neighbourhoods of Toronto we required latitudes and longitudes of the city, which we obtained from the geocoder package.

<https://geocoder.readthedocs.io/index.html>.

As for the information of the places we used the Foursquare API.

**2.2 Data Cleaning**

Firstly the data was not readily available to us, we had to scrape it. The data scraping was the first step to be done. The web scraping was done using the BeautifulSoup library of bs4 package.

Secondly ,the data we received had empty records, duplicate values for Boroughs and Neighbourhoods , unwanted symbols like ‘\n’ appended to it. All these were removed from the data for a more presentable and clean data.

Third, the latitudes and longitudes were extracted from the Geocoder package, All this data was in two different data frame, we had to combine the data into a single data frame and sorted.

**3.Methodology**

**3.1 Business Understanding**

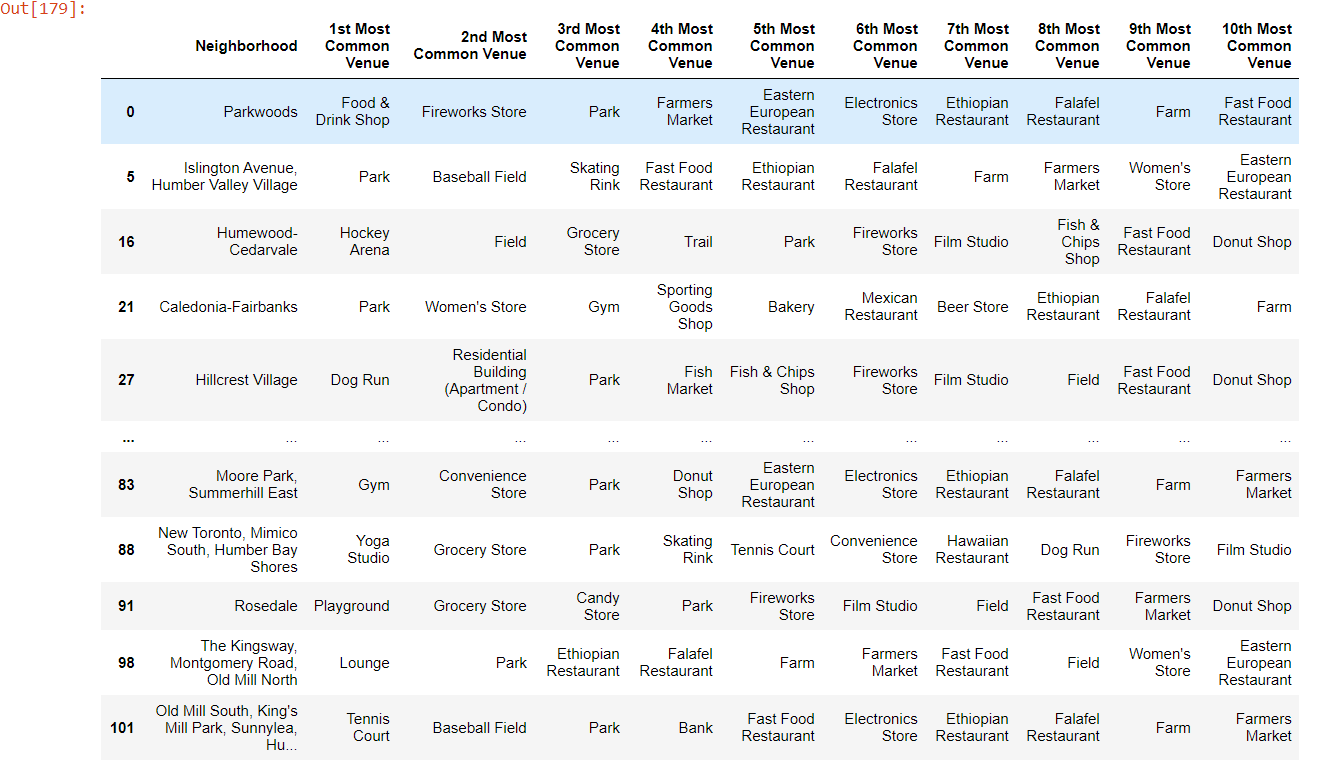
Our main aim is to find the neighbourhoods of different parts of Toronto, so that it becomes easy for the user in taking decisions.

**3.2 Analytic Approach**

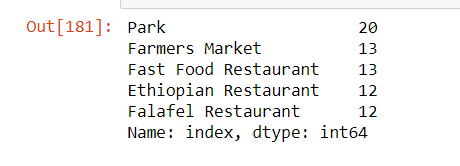
## Toronto has 4 Boroughs named East Toronto, West Toronto, Downtown Toronto and Central Toronto. First we explore the Queen's Park and Ontario Provincial Government neighbourhood of the Toronto data set to set up the functions and parameters. **We get the Parkwoods latitude and longitude values and then we find top 100 venues that are in Parkwoods within a radius of 500 meters.**

**3.3 Exploratory Data Analysis**

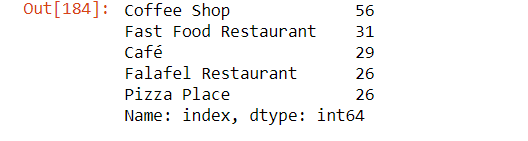
Here we use k means clustering model for training. We divide the city Toronto into two clusters. Cluster 1 has 20 neighbourhoods in it which have 20 parks, 13 Farmers Market, 13 Fast Food Restaurant, 12 Ethiopian Restaurants and 12 Falafel Restaurants.

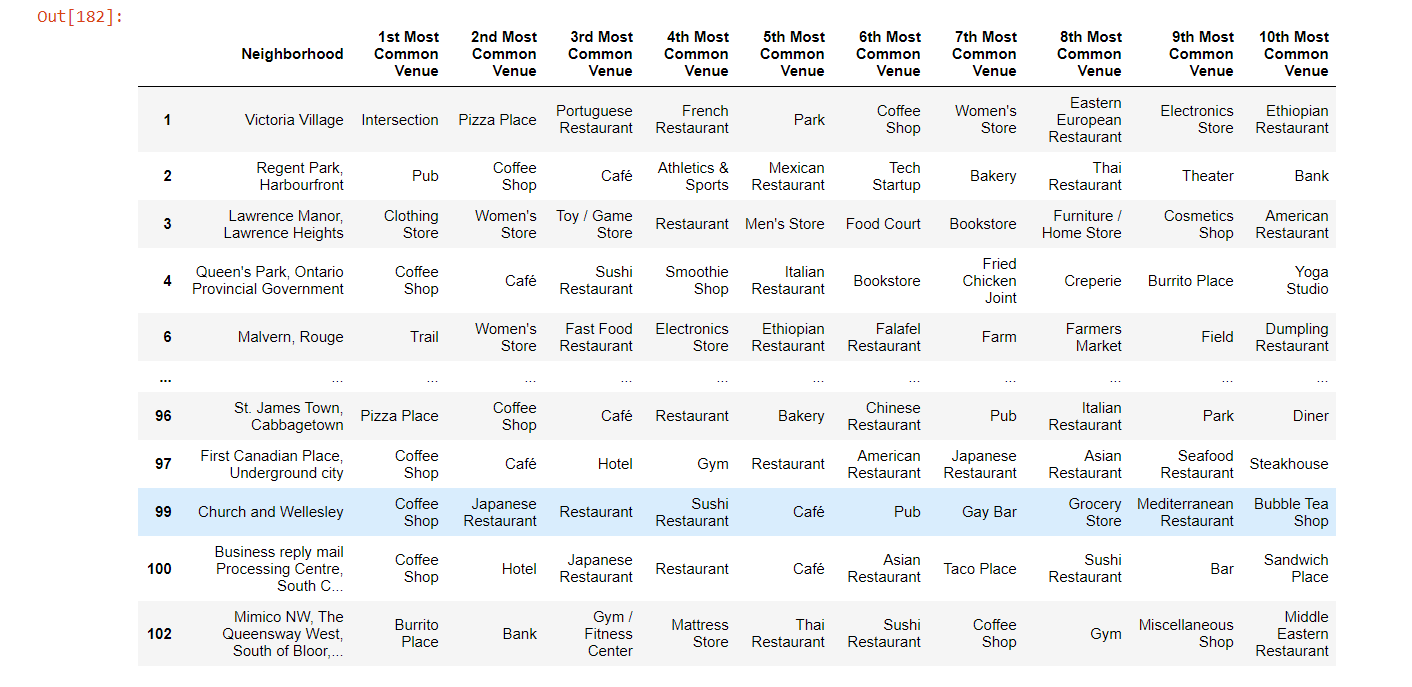
****

Neighbourhoods in Cluster 1



Cluster 2 has 82 neighbourhoods, which have 56 Coffee shops, 31 Fast Food Restaurant, 29 Cafes, 26 Falafel Restaurants , 26 Pizza Places.





Neighbourhoods of Cluster 2

**4.Results**

**The results show that cluster 1 has many parks, farmers markets and Ethiopian restaurants while cluster 2 has many coffee shops/café and Pizza place. Both places have many Fast Food and Falafel restaurants.**

**5.Discussions**

**Both the clusters have mostly the same venues in its neighbourhoods, cluster 1 has more entertainment and sport facilities while cluster 2 has more hangout places and both the places have equal amount of restaurants.**

**6.Conclusion**

**This analysis is performed on a limited data. This may be right or wrong. But if more amount of data is available there is scope of better results.**