

THE BATTLE OF THE NEIGHBORHOODS IN THE CITY OF TORONTO, ONTARIO, CANADA

INTRODUCTION

THE LOCATION

Ontario is a province located in east-central Canada, covering more than 1 million square kilometres, which borders USA and the Great Lakes. According to STATCAN [1], it had a population of 14.57 million people as of 2019. Canadas' appeal as an immigration nation has been increasing over the past two decades, with Ontario being the most popular destination between July 1, 2017 and June 30, 2018, as well as also being the most popular destination for the year 2019 (Erin Duffin, 2019) [2].

Some important facts about Ontario include:

- Their temperatures varying from 30 degrees Celsius in summer to -40 degrees Celsius in winter,
- Housing one fifth of the worlds fresh water
- Their industries vary from crop cultivation, to mineral mining, to automobile manufacturing as well as technological development.
- Has a range of cultures that are celebrated, with festivals like the Caribbean Carnival, Oktoberfest, Canadian Aboriginal and so on.

PURPOSE OF THE PROJECT

The aim of this project is to help immigrants in determining the neighbourhood that best suits them upon arrival to the province of Ontario, within Toronto city. Their decision on which neighbourhood to choose from, would be based on an analysis of venues in each neighbourhood. This project is for people looking for ease of access to Cafe, Schools, Supermarket, hospitals and so on.

PROBLEM TO BE SOLVED

Sorting list of venues for clusters of neighbourhoods in order to determine which neighbourhood has the most venues when it comes to parks, restaurants, malls and so on.

DATA DESCRIPTION

To consider this problem, the list of data considered were:

- The Ontario dataset scrapped from Wikipedia on week 3 of this course, was used. The dataset consists of the Postal code, Borough, and Neighbourhood.

Data Link: https://en.wikipedia.org/wiki/List_of_postal_codes_of_Canada:_M

The data was then cleaned, in order to remove any unnecessary data that was found i.e. incomplete data for any borough or neighbourhood.

Another dataset for latitude and longitude was gotten from the data link below: http://cocl.us/Geospatial_data

- The Foursquare API was used to gather data on venues, which include their names, location and many more features.

Foursquare is a location data provider with varying information about venues and events within any area of interest geographically. As such, the platform would be used to get data on venues, through an API call.

- The data retrieved from the Foursquare API call includes:

- a. The Neighbourhoods
- b. The Neighbourhoods' Latitude and Longitude
- c. Venues
- d. Name of the various venues
- e. Latitude and Longitude of each venue
- f. Venue Category

METHODOLOGY

We would be making use of the Foursquare API features to fetch out venues for each neighbourhood. As a result of the http request limitations, the **limit** for number of places per neighbourhood parameter, was set to 100 and the search **radius** was set to 500 for this project.

After setting the **limit** and **radius** for making the call, a class was made to get venues featuring their names, latitude, longitude, and radius as the objects of the class.

The next step, would be to explore the data to figure out how many venues are situated in each neighbourhood, then find the common venues, sort them in a descending order. This would be done to figure out which venues is the most common for each neighbourhood. Figuring this out would help us filter out the neighbourhoods with the most noise (restaurants, bars, clubs and malls).

Finally, we make a cluster of 5, in order to explore neighbourhoods and find the similarities within these neighbourhoods.

The analysis of the data begins by a simple explanatory data analysis. Using the class defined, I was able to get a count of all the venues situated in each neighbourhood.

The next step was to categorize each venue using the dummies method found from the Pandas Data Frame. Doing this enabled me to be able to find the top 5 most frequent venues situated in each neighbourhood.

Next, was grouping the top 10 common venues for each neighbourhood. This was done, so as to enable a proper analysis of the neighbourhoods when clustering them together.

RESULTS AND DISCUSSION

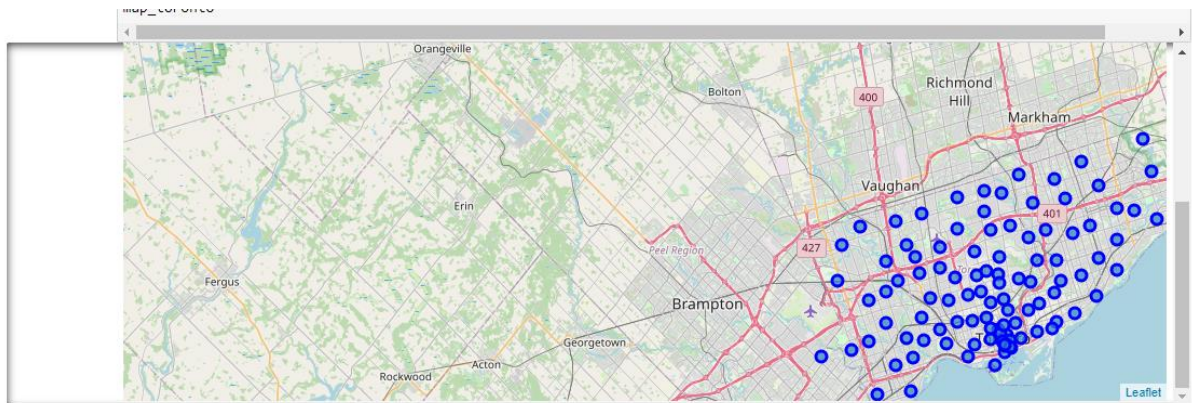


Fig 1. Map of Toronto.

Figure 1 shows the map of Toronto, with blue markers placed on the neighbourhoods situated in the city of Ontario.



Fig 2. Grouping of neighbourhoods.

Figure 2 above shows how to grouping of the neighbourhoods was done according to the amount of venues located in each neighbourhood.

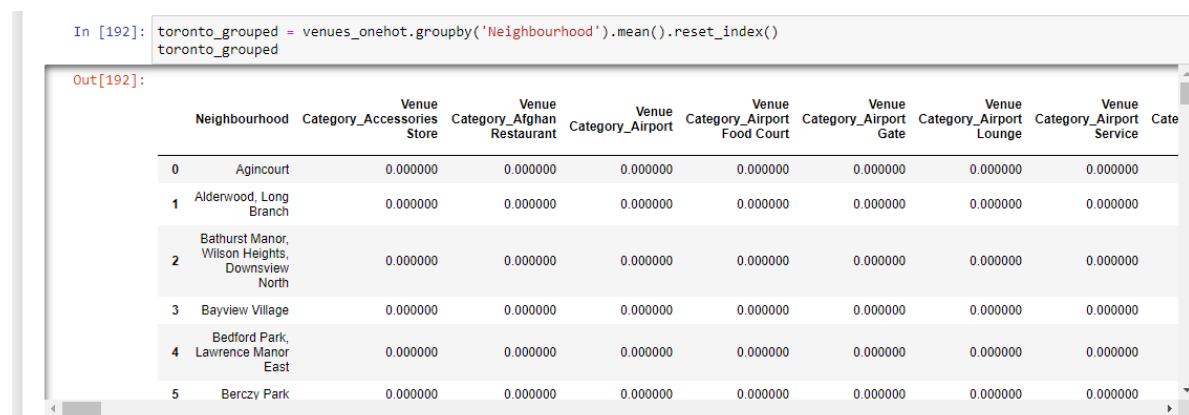


Fig 3. Grouping the neighbour “dummies” according to their mean

```

----Agincourt----
      venue  freq
0  Venue Category_Lounge  0.25
1  Venue Category_Breakfast Spot  0.25
2  Venue Category_Skating Rink  0.25
3  Venue Category_Latin American Restaurant  0.25
4  Venue Category_Accessories Store  0.00

----Alderwood, Long Branch----
      venue  freq
0  Venue Category_Pizza Place  0.29
1  Venue Category_Gym  0.14
2  Venue Category_Skating Rink  0.14
3  Venue Category_Sandwich Place  0.14
4  Venue Category_Coffee Shop  0.14

----Bathurst Manor, Wilson Heights, Downsview North----

```

Figures 3 and 4 shows how the top 5 frequent venues for each neighbourhood was gotten. This was accomplished by making use of the dummies method, grouping them according to their mean, then writing to function to fetch out the frequent venues (5).

```
In [201]: # add clustering labels
neighbourhoods_venues_sorted.insert(0, 'Cluster_Labels', kmeans.labels_)
toronto_merged = df_merge

# merge toronto_grouped with toronto_data to add latitude/longitude for each neighborhood
toronto_merged = toronto_merged.join(neighbourhoods_venues_sorted.set_index('Neighbourhood'), on='Neighbourhood')

toronto_merged.head() # check the last columns!
```

Figure 5 shows how I carried out the K-means clustering method, in order to determine the similar neighbourhoods, as well as their respective common venues between. This would help in making a decision for the immigrant on which neighbourhood he/she would like to reside in, and which neighbourhood he/she would never consider based on their preference as determined from the analysed data.

The aim of the project was to examine different neighbourhoods in the city of Toronto, Ontario for immigrants, so as to ensure they make the right choice for the neighbourhood they choose to live in. From the above project, it can be seen that there is a clear analysis of the

neighbourhoods, giving the immigrant a clear overview of how each neighbourhood is constructed. With this analysis, the immigrant would be able to make a confirmed choice of his/her preferred choice of neighbourhood.

REFERENCES

- [1] <https://www12.statcan.gc.ca/census-recensement/2016/dp-pd/hlt-fst/pd-pl/Comprehensive.cfm>
- [2] <https://www.statista.com/statistics/444906/number-of-immigrants-in-canada/#:~:text=Number%20of%20immigrants%20in%20Canada%2C%20by%20province%202019&text=Ontario%20was%20the%20province%20with,there%20in%20the%20same%20year>