Capstone Project Report

Table of Contents

1. Introduction – Idea

Toronto is considered as financial hub of Canada. To set up a new office space in Toronto, the idea is to explore various options available in the neighbourhoods of Toronto.

Toronto is considered to be a safe city in comparison to other metropolises in the world. In an article in the Economist (2015), Toronto was ranked as the safest major city in North America and the eighth safest major city in the world. Not surprisingly, however, despite being ranked as a relatively safe city, Toronto has its fair share of crime. As is the case with any big city, some neighborhoods are considered to be less safe than others. Several reasons are attributed to higher crime such as lower income, higher unemployment, lower literacy and access to education, among other reasons.

An analysis of crime and neighbourhood data within Toronto will provide good understanding of the areas to be shortlisted for setting up a new office space.

For this project, the focus will be -

1. To explore the neighbourhoods of Toronto in terms of number of head offices in the areas, shopping facilities, Entertainment options.
2. To find number of crimes in the neighbourhoods of Toronto in terms of the total number of all types of crimes.

Stakeholders

The stakeholders for this idea will be any company or business house which wants to setup a new office space in Toronto.

2. Data

The following datasets will be used to meet the idea in section 1.

- Neighbourhood data for Toronto.

  Toronto neighbours data has to be derived from Wiki for Toronto.

- Foursquare location data to explore the venues around the cities and segment them.
- Latitude and Longitude details for Toronto postcodes.
- Toronto crime data for the year 2017.

From the above datasets, explore the financial institutions within a specified limit. Also, explore other common venues and cluster them using k-means algorithm. Also, explore the crime details within the neighbourhoods of Toronto.

3. Methodology

The data sets as explained in section 2 were used as the source for data exploration. 2 data frames were created – one for Toronto Neighbourhood details and the other for Crime data.

The Toronto Neighbourhood data was further augmented using Foursquare data. Foursquare data was used to identify and explore the 100 venues within 1000 metres range of Toronto. The query for exploring the neighbourhood data was set to 'Head Office' to identify the count of companies which have their head offices in Toronto neighbourhood.

A simple count of head offices of companies would imply the number of jobs being created through the financial hub city.

Data cleansing activity was performed on both the data frames. The categories of the neighbours were also requested using Foursquare.

The goal was to identify and cluster the office categories within the neighbourhoods.

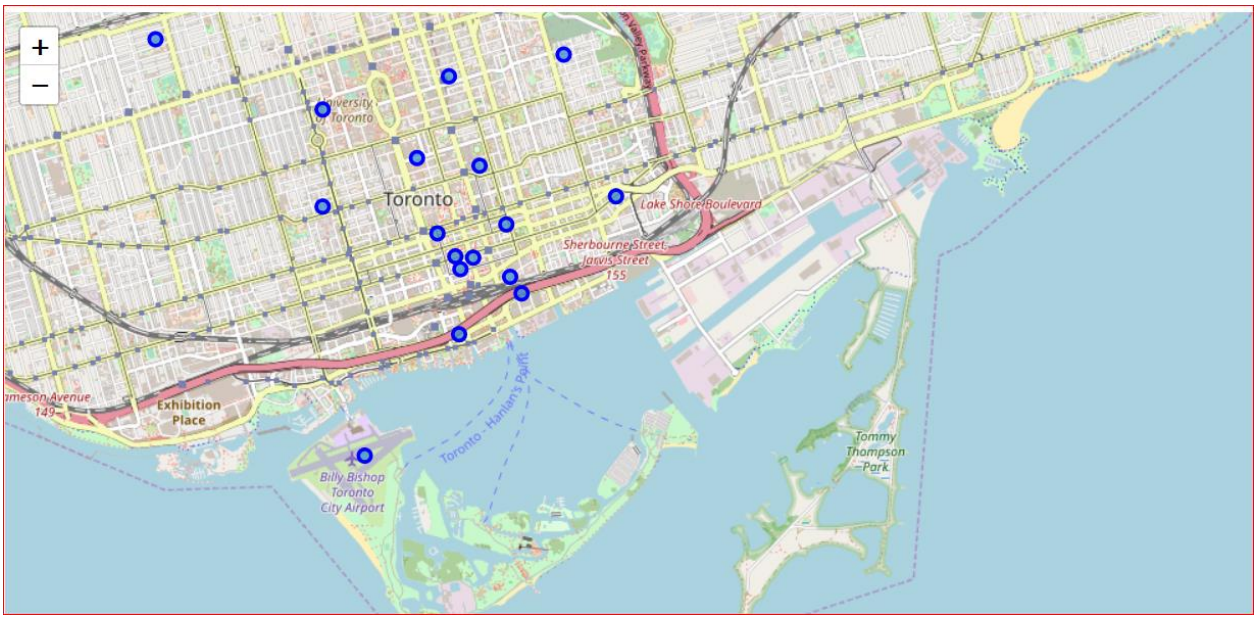
The Toronto crime dataset was used to visualize the areas which has significant crime activities during 2017. For visualization, python Folium library was used.

4. Results

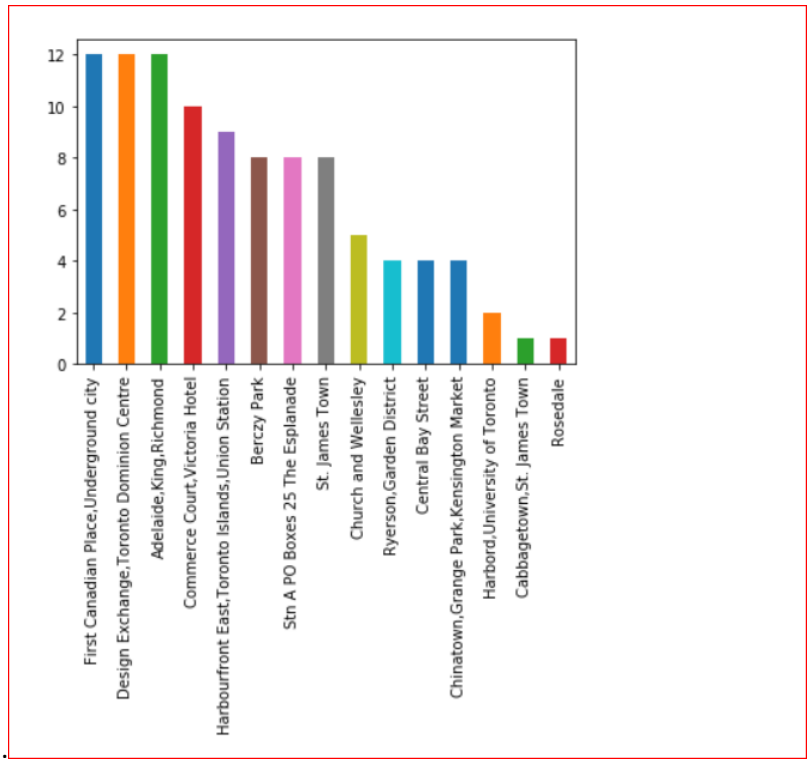
The visualization of data is divided into 2 parts as per the intention in section 1.

1. Toronto Neighbourhoods offices data exploration:

The following map depicts the neighbourhoods of Toronto.

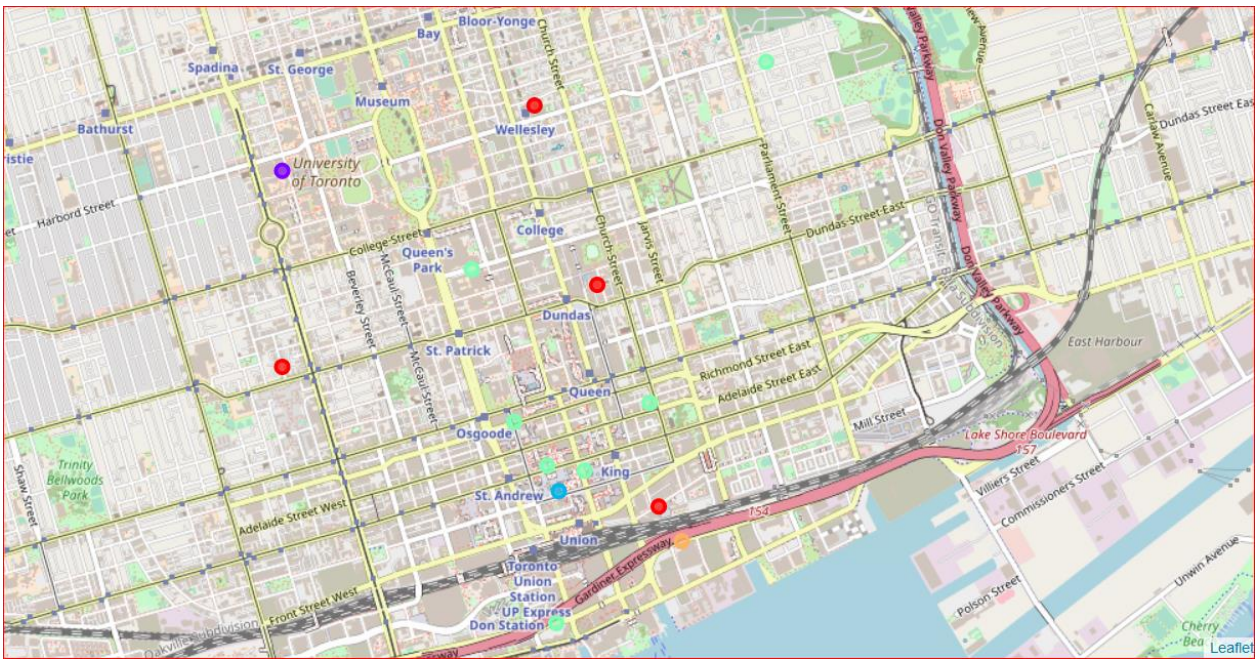


There is a significant count of company head offices near Toronto as shown in the below table. This suggests that opening a new office space in the neighbour of Toronto is a good option.



.

The Head Office categories were clustered to find the areas which has a good density of companies head offices. The nearby venues categories of the neighbours were also considered during k-means clustering. The below map shows the clustered data.

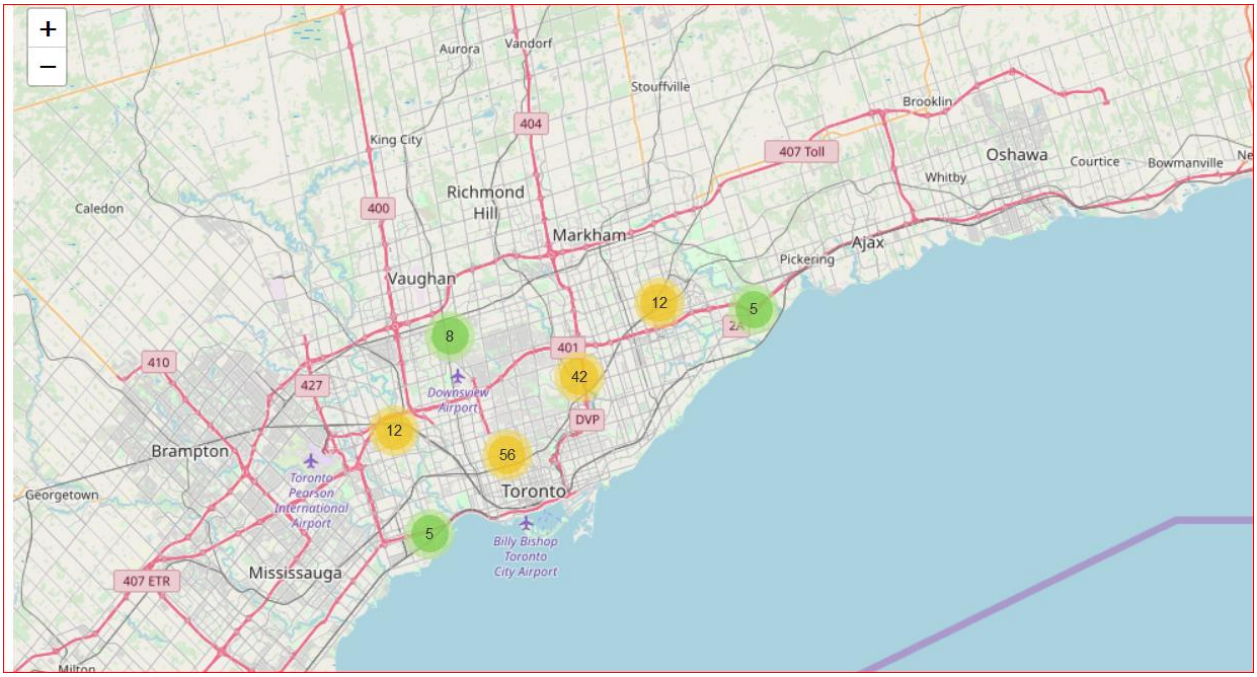


A sample of top 3 venues of nearby areas of cluster 3 (marked in GREEN dots in the map) is as below:

	Postcode	Latitude	Longitude	1st Most Common Venue	2nd Most Common Venue	3rd Most Common Venue	Cluster Labels
0	M4W	43.679563	-79.377529	Building	Pub	Office	3
1	M4X	43.667967	-79.367675	Office	Pub	Lounge	3
4	M5C	43.651494	-79.375418	Office	Pub	Lounge	3
6	M5G	43.657952	-79.387383	Office	Pub	Lounge	3
7	M5H	43.650571	-79.384568	Office	Lounge	Building	3
8	M5J	43.640816	-79.381752	Office	Lounge	Building	3
10	M5L	43.648198	-79.379817	Office	Building	Pub	3
14	M5X	43.648429	-79.382280	Office	Lounge	Building	3

2. Toronto Neighbourhoods crime data exploration:

Looking, at the crime data for the neighbourhoods of Toronto as below, it shows more crime numbers in the areas closer to station.



The data below shows the top crime counts per Neighbourhoods.

Neighbourhood	count	X	Y
Church-Yonge Corridor (75)	4699	-79.372192	43.670162
Waterfront Communities-The Island (77)	4362	-79.354797	43.653740
West Humber-Clairville (1)	3282	-79.554672	43.757866
Bay Street Corridor (76)	2744	-79.378822	43.668640
Moss Park (73)	2653	-79.355804	43.662113
York University Heights (27)	2538	-79.463386	43.787285

5. Discussion

It has been observed that, Toronto has a very high number of head offices of various companies in its neighbourhoods. The data shows that the crime rate count is also very high in the neighbourhood areas. A further exploration of data is required using other data sets like demographics, census, literacy, household income data etc.as the crime numbers will have to considered as per the population in that area.

6. Conclusion

It can be inferred that Toronto is a very good place to start office space. It has got supporting venues like pubs, Restaurants etc. in its nearby Neighbourhoods. From the above data and analysis, the area nearby University of Toronto is a good option to explore.