**Analyzing Toronto Borough to Determining Business Venue Locations**

1. **Introduction**
   1. **Background**

Toronto is a capital city in the province of Ontario Canada, being one of the fastest-growing cities in the northern part of America. The growth of the metro area population as of 2021 is estimated to be 6,255,000 [1]. The increase in immigrants entering the country has a great effect on the country's economy. Toronto also is among the leading financial industry center in Northern America, which has helped the economic growth of the city, with a desirable quality of life among its residents, and has provided a highly viable business environment [2]. Toronto has six neighborhood districts also known as Boroughs, from each borough contains 140 official neighborhoods that are recognized by the city [3]. The neighborhoods are assigned with different alphanumeric postal codes (zip code), while some share the same postal code. The postal codes help to determine which borough each neighborhood belongs to, and this helps to distinguish between neighborhoods and also the common activities that are peculiar to them. Therefore, different categories and ranges of businesses can be examined based on the neighborhood to establish the best locality to start-up business investment.

* 1. **Problem**

Extracting existing data that contains information about the neighborhood according to their postal codes and the existing business to determine the best venue for small-medium enterprise (SME) investors.

* 1. **Interest**

Considering problems in determining the best location to invest, investors are more interested in targeting the best venue to maximize their customer base with fewer competitors. This helps in analyzing which borough has the most densely populated neighborhoods and exploring the nearby venues businesses to determine which is the best neighborhood to invest in.

1. **Data Acquisition and Cleaning**

* 1. **Data Source**

The data source used for this project is web source data, where the postal codes for boroughs in the city of Toronto and their respective neighborhoods in the province of Ontario Canada, were extracted using the python built-in library BeautifulSoup. These data were found on [Wikipedia](https://en.wikipedia.org/wiki/List_of_postal_codes_of_Canada:_M), with the use of the python library BeautifulSoup, which enables the data to be downloaded into a python data frame that was used for further analysis.

* 1. **Data Wrangling and Cleaning**

Data downloaded using the BeautifulSoup library from the web have lots of special characters and unassigned values. The BeautifulSoup enables the downloaded data to be grouped in columns. The columns were split into postal codes and Boroughs, with the boroughs and neighborhoods on the same column. The first approach was to write a python script to separate the boroughs from the neighborhoods such that they are different columns. Postal that are not assigned to the borough and neighborhood are dropped from the data frame.

Further data cleaning was carried out because it was observed from the data that there are boroughs with multiple neighborhoods, which were separated by special characters. A comma character was used to separate multiple neighborhoods while other characters were dropped.

* 1. **Data Description**

Before cleaning the data there were 180 indexes and 2 columns, after cleaning there are 103 indexes and 3 columns. For the sake of analysis, the location of each neighborhood will be needed to provide further information about venues that are close to each neighborhood. A python library named GEOCODER was used to download the location latitude and longitudinal distances for each neighborhood, this will be discussed in the methodology.

1. **Methodology** 
   1. **Exploratory Analysis**

The data extracted from the web and saved in a python data frame contains the postal code, borough, neighborhood, and the latitude and longitude distance.

**GEOCODER**

The geocoder is a python library that estimates the geographical coordinate of each neighborhood on the map. Considering the large data of neighborhoods for Toronto, the first 12 geographical coordinates values of the neighborhoods (Table 1) were printed, the latitudinal values are seen to be positive values while the longitudinal values are in negative values depending on their location on the map. For this project, we decided to consider the borough that contains a high number of neighborhoods to be used for further analysis (Figure 1). There are three boroughs with high neighborhoods which are; North York, Downtown Toronto, and Scarborough. These three boroughs will be compared to their nearby venues using Foursquare API that will be discussed later, to further assist in choosing the best location to start a business for SMEs.



Table 1: Geographical coordinates (Latitutde and Longitude) for the first 12 neighborhoods using Geocoder

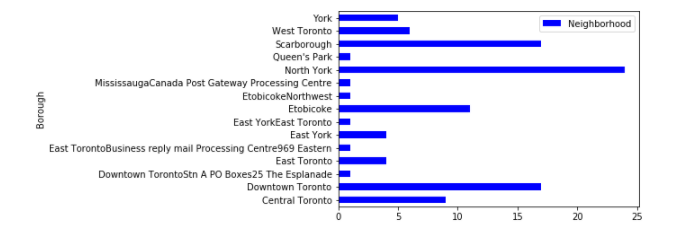


Figure 1: Neighborhood for each Borough

**FOLIUM**

Folium is a python library used to visualize the geographical details for each borough in Toronto city. The details display the borough and neighborhood with each point containing various locations on the map (figure 2).

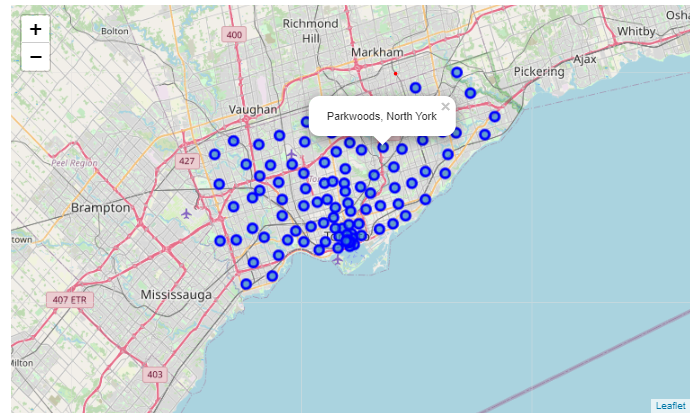


Figure 2: Location of boroughs and their Neighborhoods in Toronto city

The boroughs were extracted individually from the whole Toronto borough to visualize their locations on the map using folium. Figures 3, 4, and 5 show the Neighborhood map of Downtown Toronto, North York, and Scarborough respectively.

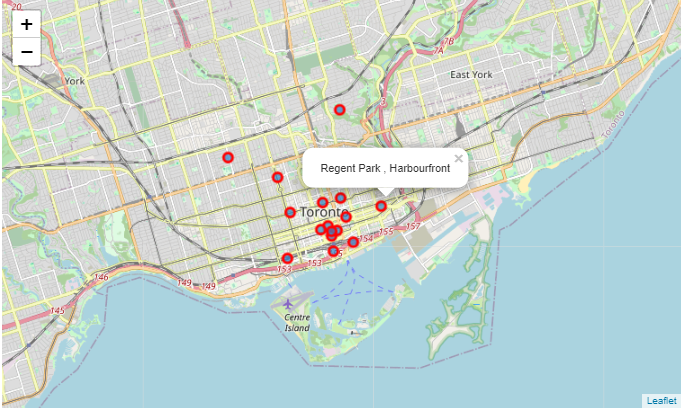


Figure 3: location of Downtown Toronto Neighborhoods

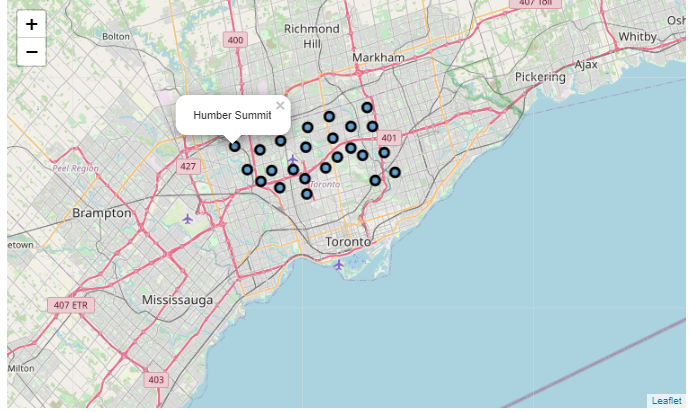


Figure 4: Location of North York Neighbrhoods

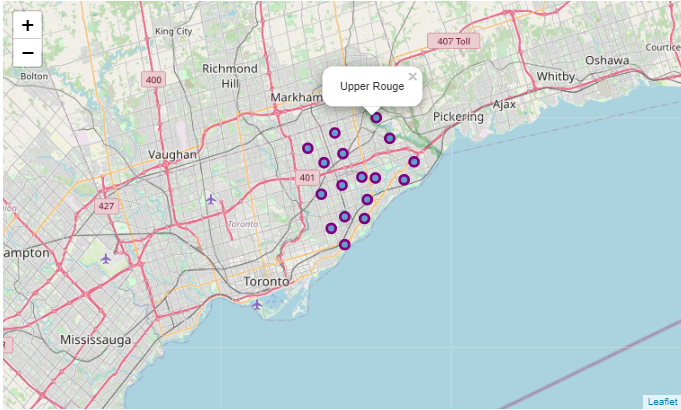


Figure 5: Location of Scarborough Neigborhoods

After visualizing the three different boroughs, I further explore the data by getting the first 100 nearby businesses venues that are within 700 meters radius. To achieve this task I used a foursquare web application programmable interface (API).

* 1. **Foursquare API**

To use the foursquare API, an account was created which makes it possible to generate a client ID and SECRET. These credentials were used to generate an access code that enables the connection of the geographical coordinates of neighborhoods to generate their nearby business venues. The results of the venues are in JSON format, to print the result into a data frame a function was written that shows the company's name, categories, and coordinates (Table 2,3, &4).

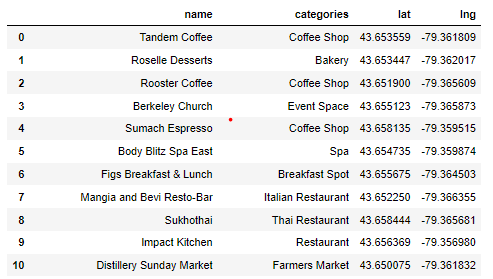


Table 2: First 10 result for Downtown Toronto nearby venues with total 88 venues

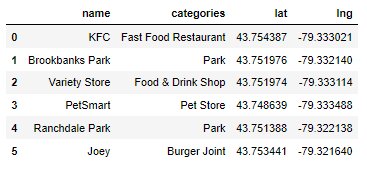


Table 3: North York Nearby venues with total 6 venues

Table 4: Nearby Venues for Scarborough with total of 6 venues



The nearby venues returned from each borough show that Downtown Toronto has more venues compared to North York and Scarborough. Further analysis was carried out on Downtown Toronto using foursquare to see the most common venue and the clusters using the unsupervised K-Means algorithm.

1. **Results**

The nearby venues for downtown Toronto were analyzed in K-means clusters. The clusters were divided into 5 different unsupervised learning algorithms. The purpose of the clustering is to show the different venues categories according to the most common venue, which will help to determine which is the most common business in the borough and what neighborhood.

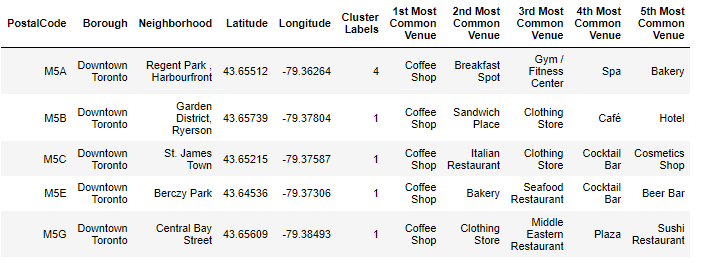


Table 5: Five most common venue in the Downtown Toronto borough

From the above table, it can be seen that the coffee shop is the first most common venue, also the restaurants are common businesses. The aim was to know which neighborhoods in the borough are more concentrated with business venues. The clusters group is shown in fig 6., it can be observed that most of the venues are in the first clusters which is the purple on the map. Among the list of the neighborhoods are the Graden District Harbourfront, St James Town, Berczy Park, and Central Bay Street.

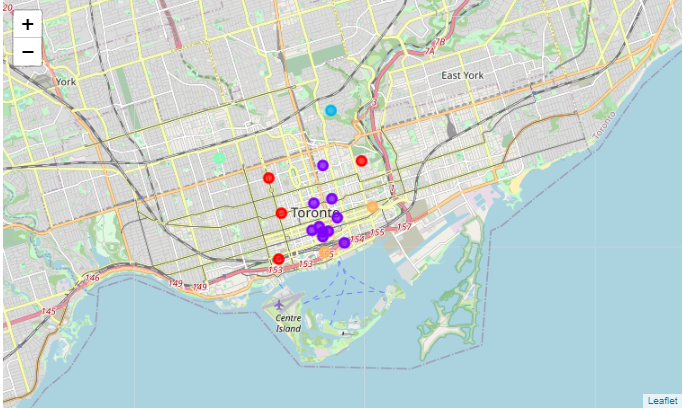


Figure 6. Clusters of venues in Downtown Borough

1. **Discussion**

The main focus of this project is to help explore the business opportunity venue for investors in the city of Toronto Canada. As there are too many boroughs growing faster in the city, to analyze the venues I downloaded the venues using the foursquare API, the highest venue was Downtown Toronto, which I then analyzed further. Downtown Toronto returned 88 venues as this may vary based on the foursquare web update. I used the Kmeans algorithm to segment the downtown Toronto venues into 5 different clusters, the clusters were used to identify the categories of businesses around the neighborhood. From the result, the most common business is the coffee shop followed by the restaurant, I will think this is because of the corporate business in downtown Toronto, as it is common for people to get coffee before the start of work or during lunch break.

1. **Conclusion**

The city of Toronto's fast growth can make it highly competitive for businesses and investors, with the analysis can give the business investor on what type of business and location where businesses are less competitive and fast-growing. Although this paper did not look into the revenue of business and cost expense according to their location, this can be recommended for future research as the aim of this project is to show the neighborhood with the most business opportunities.

**Reference**

[1] PR & Content Manager | RE/MAX Canada Lydia McNutt is an award-winning editor. (2020, December 7). *Toronto Population is Fastest Growing in North America: RE/MAX Canada News*. RE/MAX Canada. https://blog.remax.ca/toronto-population-is-fastest-growing-in-north-america/.

[2] City of Toronto. “Strong Economy.” *City of Toronto*, 17 June 2021, [www.toronto.ca/business-economy/invest-in-toronto/strong-economy/](http://www.toronto.ca/business-economy/invest-in-toronto/strong-economy/).

[3] Toronto - Wikipedia