Capstone Project - The Battle of Neighborhoods

Bushra

1. Introduction

The parent always considers choosing the appropriate school for their kids. Therefore, they tend to ask friends, family, and even neighbors, to help them make the right decision. However, this becomes a challenge, if the parent is moving to another city. Parents at that moment don't have enough information about the schools in that city, and maybe they don't know anyone there. When the parent is moved, the first step is to choose a neighborhood that they will live in. Their choice depends on several factors as the type of houses, the distance from the workplace, and the most important is the availability of schools in that neighborhood or a nearby neighborhood. Therefore, the parent has to explore the schools according to the neighborhood.

1.2 Business Problem

The main goal of this project is to help parents choose the right neighborhood when they are moved. This will be based on the available schools in that neighborhood. Therefore, in this project, I will explore the neighborhoods of Toronto city regarding the available schools in each neighborhood. Schools have different categories, which are: elementary school, high School, middle School and etc. The number of schools in each category will be presented as well.

2.The Methodology

The data used from Wikipedia which contains all neighborhood in Toronto city with their postal code.

From this link: https://en.wikipedia.org/wiki/List_of_postal_codes_of_Canada:_M

Then, I scraped the data using BeautifulSoup library, after that, I use another data set for getting the Latitude Longitude of each neighborhood.

From this link: http://cocl.us/Geospatial_data

However, the available schools in each neighborhood and their categorizes are extracted from foursquare.

2.1 Data Preprocessing

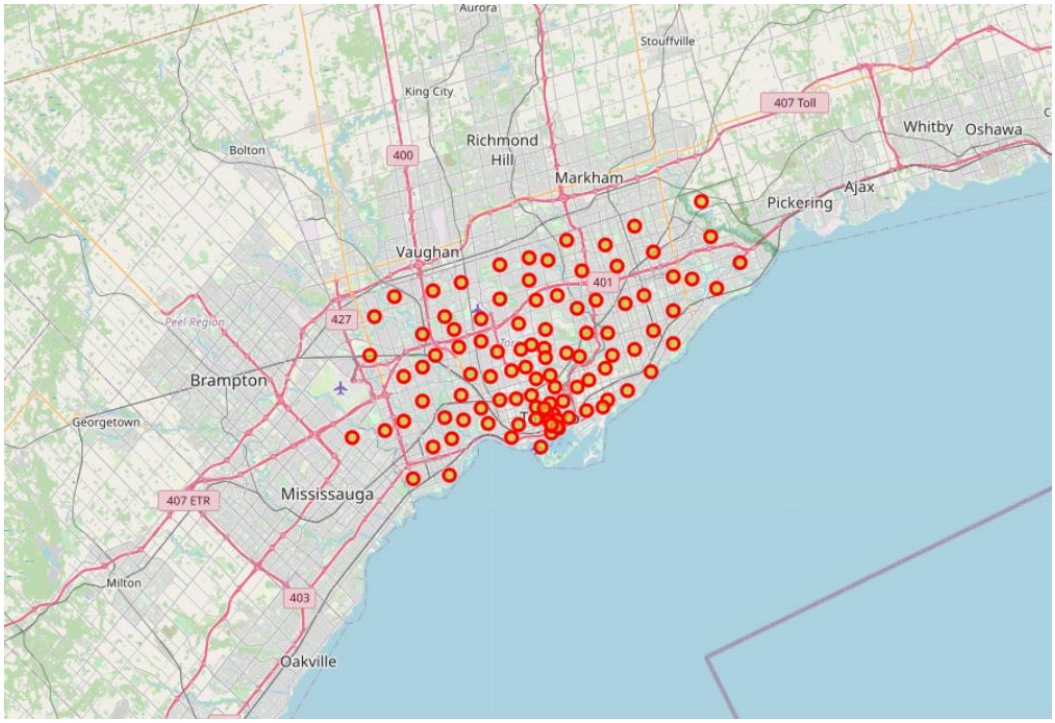
First, I Scrape The Data and save it into panda data frame, Saving the scarped data into panda frame, and handle missing values by delete all rows that have ' Not assigned ' value, Group neighborhood by postal code, and Merge the original data frame with the Geospatial data frame to get a complete dataset.

The result as below**:**

	Borough	Neighborhood	Postal Code	Latitude	Longitude
0	Scarborough	Malvern / Rouge	M1B	43.806686	-79.194353
1	Scarborough	Rouge Hill / Port Union / Highland Creek	M1C	43.784535	-79.160497
2	Scarborough	Guildwood / Morningside / West Hill	M1E	43.763573	-79.188711
3	Scarborough	Woburn	M1G	43.770992	-79.216917
4	Scarborough	Cedarbrae	M1H	43.773136	-79.239476

2.2 Explore the Neighborhoods

First by Get the coordinate information of Toronto then create map of Toronto using latitude and longitude values:



2.3 Explore Schools in the Neighborhoods:

- Step 1: Assign Foursquare access information
- Step 2: Get the information from Foursquare

  Category Id of schools is: 4bf58dd8d48988d13b941735
- Step 3: Add the extracted information to the data frame based on the postal code.
-

The result of this phase:

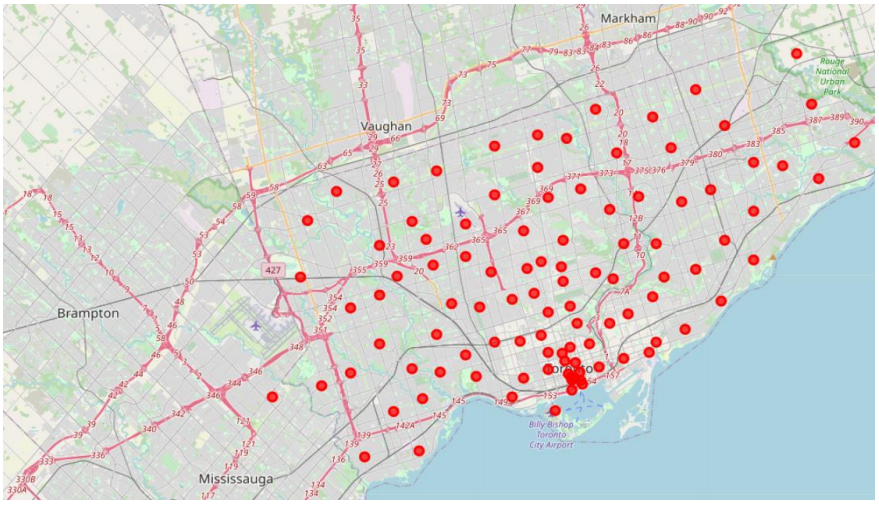
	Neighborhood	Neighborhood Latitude	Neighborhood Longitude	Venue	Venue Latitude	Venue Longitude	Venue Category
0	Kennedy Park / Ionview / East Birchmount Park	43.727929	-79.262029	St. Maria Goretti Catholic School	43.730201	-79.266135	School
1	Kennedy Park / Ionview / East Birchmount Park	43.727929	-79.262029	SCAS	43.728427	-79.256372	School
2	Clarks Corners / Tam O'Shanter / Sullivan	43.781638	-79.304302	Scarborough Pauline Johnson YMCA Before and Af...	43.785020	-79.303703	Daycare
3	Clarks Corners / Tam O'Shanter / Sullivan	43.781638	-79.304302	Stephen Leacock Collegiate Institute - Gym	43.785057	-79.300964	High School
4	Steeles West / L'Amoreaux West	43.799525	-79.318389	Scarborough Beverly Glen YMCA Before and After...	43.798805	-79.323136	Daycare

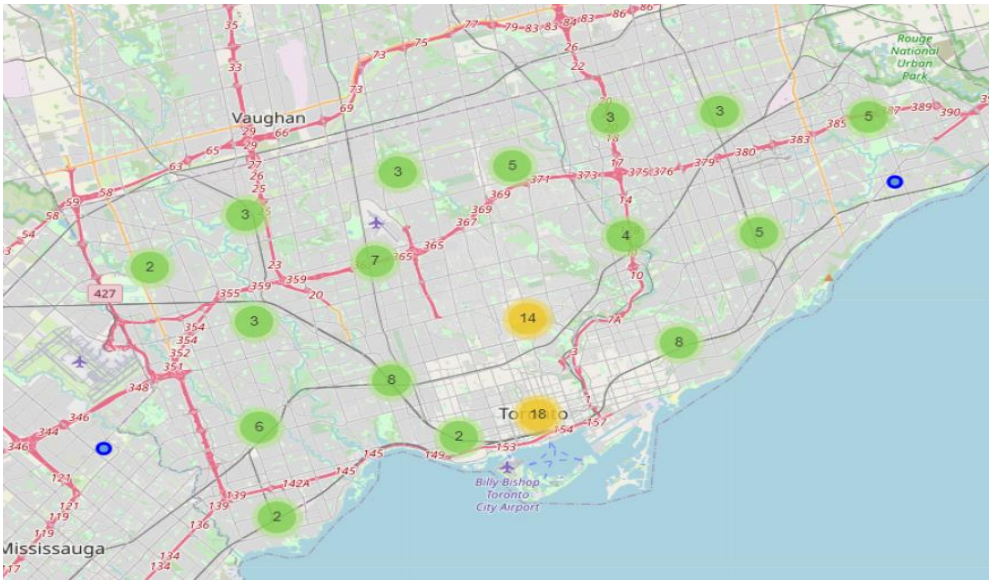
3 The Result

3.1 Cluster Neighbourhoods Based on Schools

- Step 1: Run k-means to cluster the neighbourhood into 4 clusters

- Step 2: Create a new data frame that includes the cluster as well as the top 10 venues for each neighbourhood
- Step 3: Visualize the resulting clusters on a map





The result of each phase of this project has been presented above. However, parents know can use this data to see the available of schools in Toronto neighborhood. They can see the type of schools as well, and the number of schools in each neighborhood.

The table below shows the top 5 Neighborhood that have the highest number of schools. Commerce Court / Victoria Hotel has the highest number of schools which

is 15 schools. Followed by First Canadian Place / Underground city, Toronto Dominion Centre / Design Exchange, St. James Town and University of Toronto / Harbord.

4. Discussion

The use of foursquare was useful for this type of projects, that depends mainly on locations. However, I noticed that they have 31 categories under school venue.

However, I recommend performing similar projects as the following:

1- A project that focus on one type of school and compare between them.

2- A project that compare between number of schools and the size of the neighbourhood. Is there any relationship between them?

5. Conclusion

To conclude, the choose of the right school is one of parent concerns. This will raise, if the parent does not know the area or moving to a new area, because their choose of a neighbourhood often depend on the chosen school. Therefore, this project is designed to help parents to explore the schools in Toronto regarding its neighbourhoods. The results show that Commerce Court / Victoria Hotel has the highest number of schools, that counted as 15 schools. As a recommendation, to do slimier project that compare between number of schools and the size of the neighbourhood.