1. Introduction

1.1 Background

This project aims to monitor and analyze data related to Toronto city's characteristics such as Postal codes, boroughs and neighborhoods combined with Foursquare location data. Data describing Toronto's wider area were extracted from Wikipedia and latitudes – longitudes from Foursquare through API calls.

1.2 Challenge

A major problem about establishing a new shop, restaurant or office, is the location. The location can affect dramatically the productivity and development of a business. That is the reason why a business analysis model should include location data analysis and manipulate those data in a way that provides valuable insights.

1.3 Interest

There are some factors such as population density, public transport, presence of other competitors, borough's or neighborhood's rent value, that highly affect the development and sustainability of a business. By monitoring all the related data, the decision about the location of a new business is definitely improved.

2. Data acquisition and cleaning

The data required for this analysis where extracted from the following link: https://en.wikipedia.org/wiki/List_of_postal_codes_of_Canada:_M. Beautiful soup library was also one of the tools that helped data manipulation. After the page was downloaded as an html file, the first table (containing the related data) was accessed (through an iteration) and converted to a pandas data frame. First five (5) rows displayed below.

[12]:

	PostalCode	Borough	Neighborhood
0	M1A\n	Not assigned\n	Not assigned\n
1	M2A\n	Not assigned\n	Not assigned\n
2	M3A\n	North York\n	Parkwoods
3	M4A\n	North York\n	Victoria Village
4	M5A\n	Downtown Toronto\n	Regent Park, Harbourfront

After having the data frame, a data cleaning process was necessary. The "\n" suffix was removed and any missing values were also excluded. The result was a (103,3) shaped data frame with 11 boroughs and 103 neighborhoods.
First five (5) rows of the data frame:

[14]:

	PostalCode	Borough	Neighborhood
2	M3A	North York	Parkwoods
3	M4A	North York	Victoria Village
4	M5A	Downtown Toronto	Regent Park, Harbourfront
5	M6A	North York	Lawrence Manor, Lawrence Heights
6	M7A	Downtown Toronto	Queen's Park, Ontario Provincial Government

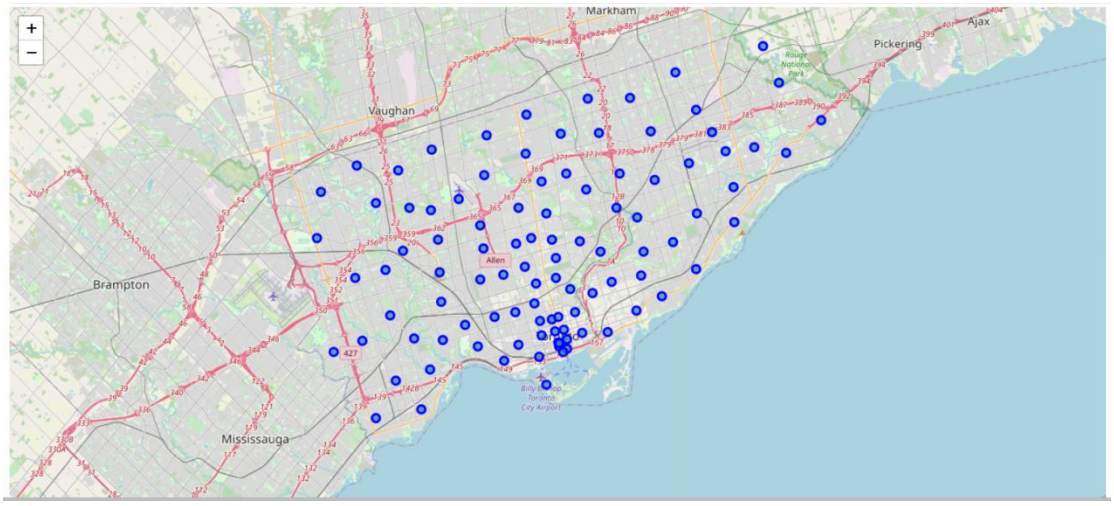
The third stage was the location data addition. This was achieved by using a geolocator fun ction and creating two (2) new columns in the data frame in which the latitudes and longit udes were stored. Only one (1) missing value found and was excluded. First five rows of the updated data frame:

	PostalCode	Borough	Neighbourhood	latitude	longitude
2	M3A	North York	Parkwoods	43.7545	-79.3300
3	M4A	North York	Victoria Village	43.7276	-79.3148
4	M5A	Downtown Toronto	Regent Park, Harbourfront	43.6555	-79.3626
5	M6A	North York	Lawrence Manor, Lawrence Heights	43.7223	-79.4504

3. Methodology and Results

In this section, I will analyze the steps taken for location data analysis and the results. After the main data manipulation was completed, as shown in the previous chapters, I had to use Foursquare API credentials in order to monitor the statistics and venues for my area of study.

A map of Toronto with each borough was made.



The data frame was processed again, in a way that reveals the venues in each neighborhood and specifically the most popular venue. The first five (5) neighborhood's and venue's coordinates are provided in the table below.

	Neighborhood	Neighborhood Latitude	Neighborhood Longitude	Venue	Venue Latitude	Venue Longitude	Venue Category
0	Parkwoods	43.7545	-79.3300	Brookbanks Park	43.751976	-79.332140	Park
1	Parkwoods	43.7545	-79.3300	Variety Store	43.751974	-79.333114	Food & Drink Shop
2	Victoria Village	43.7276	-79.3148	Victoria Village Arena	43.723481	-79.315635	Hockey Arena
3	Victoria Village	43.7276	-79.3148	Portugril	43.725819	-79.312785	Portuguese Restaurant
4	Victoria Village	43.7276	-79.3148	Tim Hortons	43.725517	-79.313103	Coffee Shop

My analysis and business proposition was based in the following examples of top trending venues in each neighborhood. The frequency of each venue is maybe one of the major factors which have to be related with the location choice. Also, the population is a crucial factor and the neighborhood's location in comparison with other neighborhoods. For example, if there is a neighborhood with no restaurants in its trending list, but with very low population and a large distance from other neighborhoods, it may not be an efficient location for a new restaurant.

On the other hand, a neighborhood that has many restaurants may also have the appropriate audience/target group which could support a diverse idea on the same field.

For this project, I will consider as a major factor the lack of restaurants in a neighborhoods trending list.

4. Results

In this chapter there are some results of the prior analysis. In the table below, the first five rows of the data frame which represent each neighborhood are displayed. Furthermore, the most common venues are shown (descending by columns).

	Neighborhood	1st Most Common Venue	2nd Most Common Venue	3rd Most Common Venue	4th Most Common Venue	5th Most Common Venue	6th Most Common Venue	7th Most Common Venue	8th Most Common Venue	9th Most Common Venue	10th Most Common Venue
0	Agincourt	Breakfast Spot	Badminton Court	Latin American Restaurant	Skating Rink	Farmers Market	Escape Room	Ethiopian Restaurant	Event Space	Falafel Restaurant	Fast Food Restaurant
1	Alderwood, Long Branch	Convenience Store	Dance Studio	Sandwich Place	Coffee Shop	Gym	Pub	Pizza Place	Deli / Bodega	Curling Ice	Fish Market
2	Bathurst Manor, Wilson Heights, Downsview North	Mediterranean Restaurant	Fried Chicken Joint	Deli / Bodega	Coffee Shop	Middle Eastern Restaurant	Pizza Place	Health Food Store	Donut Shop	Fish Market	Fish & Chips Shop
3	Bayview Village	Flower Shop	Dog Run	Gas Station	Trail	Park	Event Space	Electronics Store	Escape Room	Ethiopian Restaurant	Falafel Restaurant
4	Bedford Park, Lawrence Manor East	Pizza Place	Italian Restaurant	Sandwich Place	Coffee Shop	Butcher	Café	Liquor Store	Restaurant	Sushi Restaurant	Pub

I took five (5) neighborhoods which in my opinion represent a good establishing location for a new restaurant in the city of Toronto. Those neighborhoods are :

```
----Alderwood, Long Branch----
              venue  freq
0       Dance Studio  0.14
1        Coffee Shop  0.14
2        Pizza Place  0.14
3  Convenience Store  0.14
4               Pub  0.14
```

In those neighborhoods, the trending venues are closely related to entertainment. There is a decent frequency in restaurants (0.14 Pizza place). That fact may indicate that the specific area is a place where people visit for entertaining activities.

```
----Bayview Village----
         venue  freq
0  Flower Shop   0.2
1  Gas Station   0.2
2         Park   0.2
3        Trail   0.2
4      Dog Run   0.2
```

In this area there are no restaurants in trending venues but also the park and dog run indicates an area with enough free space for sports and hobbies. Also, the high frequency of gas stations may be an indicator about high traffic which means that a restaurant with take-away and drive-through services, could be a considerable option.

```
----Birch Cliff, Cliffside West----
                  venue  freq
0                  Café  0.25
1          Skating Rink  0.25
2  General Entertainment  0.25
3       College Stadium  0.25
4           Music Venue  0.00
```

This is a neighborhood similar to the first group. The frequencies indicate the presence of teenagers and young people. This could be an appropriate place for a fast-food restaurant.

```
----Cedarbrae----
          venue  freq
0        Lounge  0.33
1  Gaming Cafe   0.33
2         Trail  0.33
3  Yoga Studio   0.00
4  Music Venue   0.00
```

In this example, a restaurant could work as a supplementary service to gaming cafes. Many young people who spend hours in gaming cafes could be also potential customers in a low-budget restaurant.

```
----Central Bay Street----
                       venue  freq
0                Coffee Shop  0.21
1             Bubble Tea Shop  0.04
2              Sandwich Place  0.04
3          Italian Restaurant  0.04
4  Middle Eastern Restaurant  0.04
```

In this last example, we have a central street which obviously has many health regulated establishments. Although, the frequency in each one except coffee shops, is low. That means that the main activities in this area are related to entertainment and food & beverages. So, having a unique idea of a restaurant may fit in this area.

5. Conclusion

In conclusion, I want to mention the importance of data analysis in business planning. Having the appropriate data set could be a major advantage in setting a new business. I tried to combine the results from the exploratory analysis of the data set and correlate the trends with potential market demands.

The location of a business, especially in customer service sector, directly affects its productivity and development. Using all those methods and tools could be a game-changing factor for any business, company or entrepreneur.