Business Relocation Analysis and Assistance

1. Introduction

One of the most important decisions an entrepreneur needs to take is to choose or to move his business location. What motivates business relocation is always money. Entrepreneur is avoiding money loss or perceiving new opportunities of money gain [1]. When choosing a business location, many factors should be considered and well studied. In this project we will consider two of them. The first one is **foot traffic**. Depending on business nature, entrepreneur may need more or less pedestrian traffic. Besides, depending on business, nearby **competitors** may be beneficial or not for a business [2].

The goal of this project is to provide an analysis and visualizations to help entrepreneur who is willing to choose his business location to better understand foot traffic and competitiveness in every neighborhood in a city of his choice.

This analysis will be applied on the city of Toronto, CA and as an illustrative example we will suppose that business is about a restaurant. But the same analysis could be applied to any other city and with any other business.

2. Data

Two main sources of data will be used in this project. The first one is a Wikipedia web page [3]. We will scrape this page to get a complete list of Toronto neighborhood as well as corresponding postal codes. The second source of our data is Foursquare API [4]. The API will provide us information about venues in Toronto like Venue name, address, category and geospatial information. This data will be used to first calculate the total number of venues for every neighborhood. Then, we will use it to calculate for a given address the number of competitors based within a chosen radius. Besides, the category data about venues is useful to filter competitors of a given business and to make neighborhood clustering. Another csv file that contains correspondence between postal code, latitude and longitude will be mainly used to get geospatial data for every neighborhood.

3. Methodology

a. Web scraping and datafile creation

We begin with scraping wikipedia page [3] to extract all neighborhood, borough in Toronto city and their postal codes. Not assigned borough are discarded. A data frame is created.

b. CSV file preprocessing

The data frame created above is preprocessed and neighborhood with the same postal code are merged. Not assigned neighborhood are replaced by borough.

c. Add Longitude and latitude data

A csv file with latitude and longitude data is used to get geospatial data for neighborhood.

d. Explore neighborhood in Toronto

To explore Toronto city, we begin by getting its geospatial coordinates using geopy api [5]. Then, we draw a map with neighborhood markers. In the second part, we begin with getting venues data from Foursquare API. This data includes venues names, category, latitude, longitude and postal code. Since the API returns data based on geospatial data and not on neighborhood information, venue postal code is used to verify corresponding neighborhood.

e. Delete duplicate venues

Since we want to get the maximum venes, our result contains a considerable number of duplicate venues associated to different neighborhoods. To delete duplicate, we proceed as follows: For every venue, if its postal code is null, then we will associate it to the nearest neighborhood. Otherwise, we will extract the first part of its postal code. Postal codes are cleaned and normalized.

f. Analyze venues and competitors distribution in neighborhoods

In this section we analyze venues distribution in neighborhoods and provide functions to draw maps useful to localize best opportunities for a new business or a new franchise. To do this, we analyze neighborhoods crowdedness w.r.t total number of venues. We create a dataframe that contains for every neighborhood columns= postalcode and crowdedness. Since we can't get exact value of crowdedness, we will get estimated value using available data. We will calculate for every neighborhood the distance to the nearest neighborhood. Then, we

divide the number of venues in this neighborhood by this distance. Finally, we normalize data to get values between 0 and 1. Next, we draw a choropleth map to visualize neighborhoods with crowdedness index. the goal of the next section is to get for a given address: the number of venues as well as the number of competitors within a chosen radius. We get geospatial data from the address and calculate distance between two points using geopy api [5].

g. Cluster neighborhood w.r.t top-5 venues categories:

We first create one hot representation of venues categories for venues. Then, we get occurrences of every category. We then get top-5 categories for a venue to get an idea about neighborhood profile. Therefore, we get 10th more common categories for every neighborhood and use KMeans model to cluster neighborhood to five clusters. We then visualize the result in a map using different color marker for every cluster.

4. Results

In this section, we invoke the results of our work.

a. Toronto city neighborhood:

Figure 1. is a Toronto city map with neighborhood markers. In this map we see how some neighborhood are close and others are distant. This gives us the idea to consider area for every neighborhood to calculate crowdedness.

b. Crowdedness in Toronto city:

After calculating the crowdedness index for every neighborhood, choropleth map in Figure 2. illustrates our results. This helps to locate neighborhood with big foot traffic.

c. Neighborhood profile:

Figure 3. is an illustration of neighborhood profiling based on venues categories that are most present. This profile gives an idea about business types in neighborhood and helps to make a decision about business categories that best suit entrepreneur’s business.

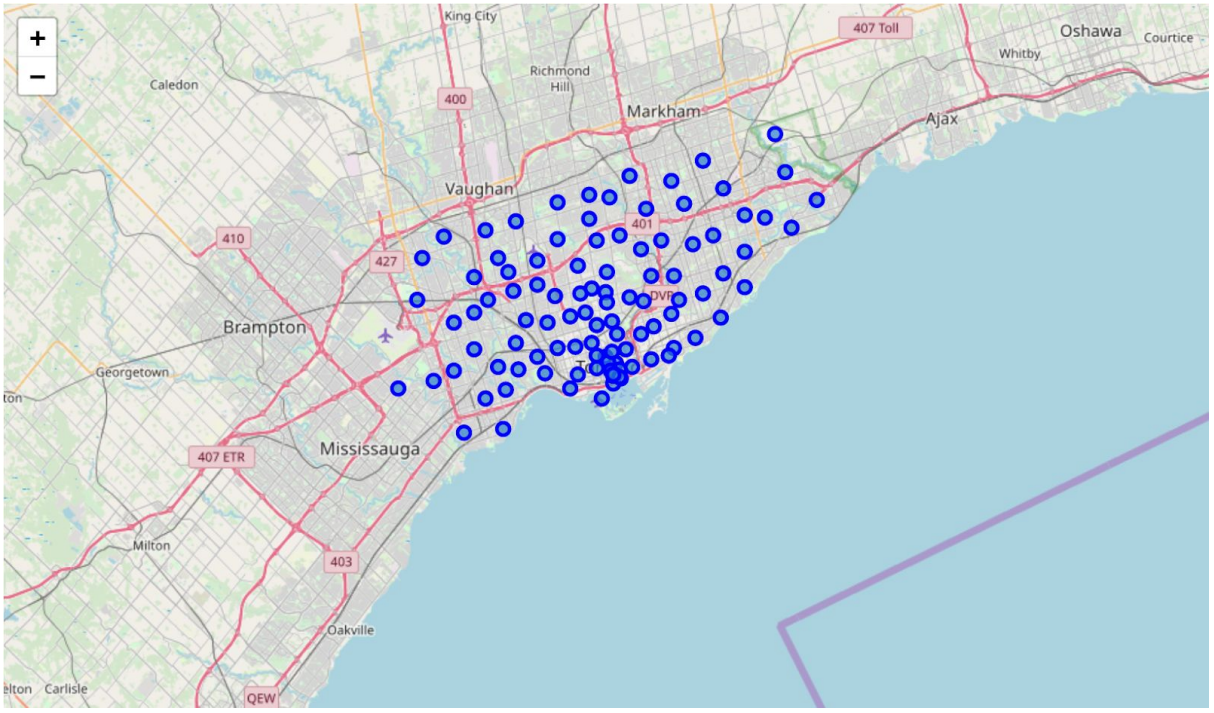

Figure1. Toronto City map with neighborhood markers

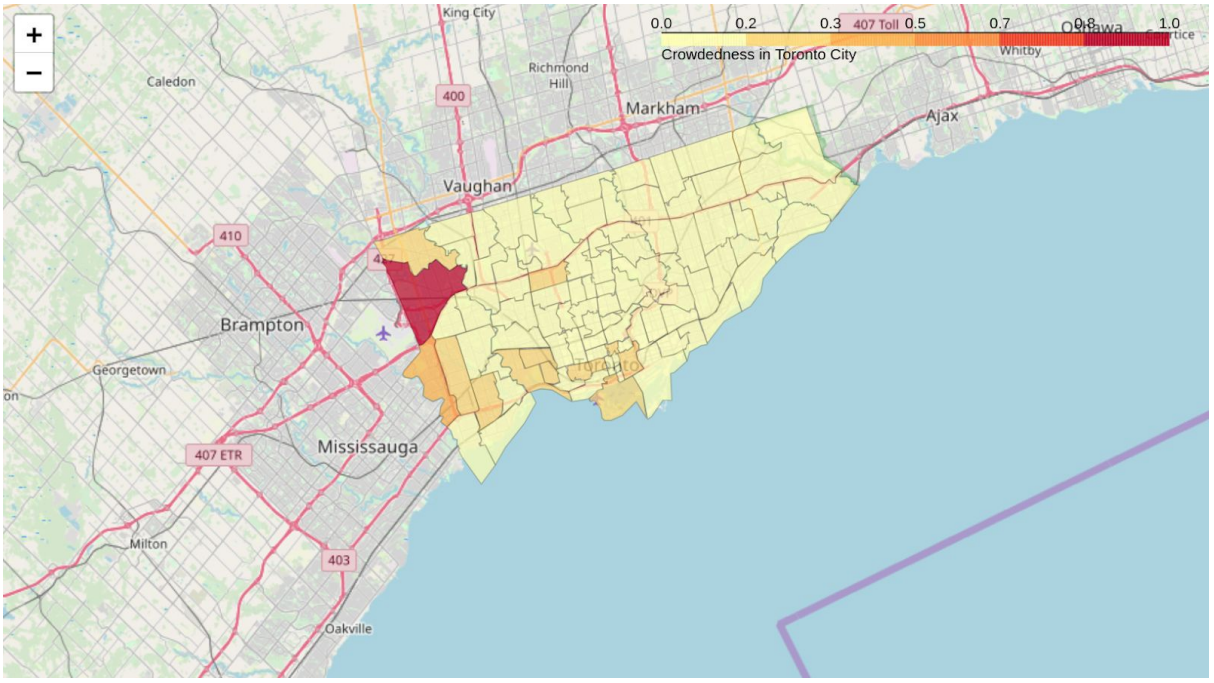

Figure 2. Crowdedness in Toronto city

```
----Rouge,Malvern----
          venue  freq
0           Zoo  0.12
1      Pharmacy  0.12
2  Dessert Shop  0.12
3    Campground  0.12
4           Spa  0.12


----Highland Creek,Rouge Hill,Port Union----
                venue  freq
0               Beach  0.14
1  Italian Restaurant  0.14
2           BBQ Joint  0.14
3     Cosmetics Shop  0.14
4  Athletics & Sports  0.14


----Guildwood,Morningside,West Hill----
                 venue  freq
0          Pizza Place  0.18
1         Burger Joint  0.09
2                 Park  0.09
3  Fried Chicken Joint  0.09
4       Breakfast Spot  0.09
```

Figure 3. Neighborhood profile

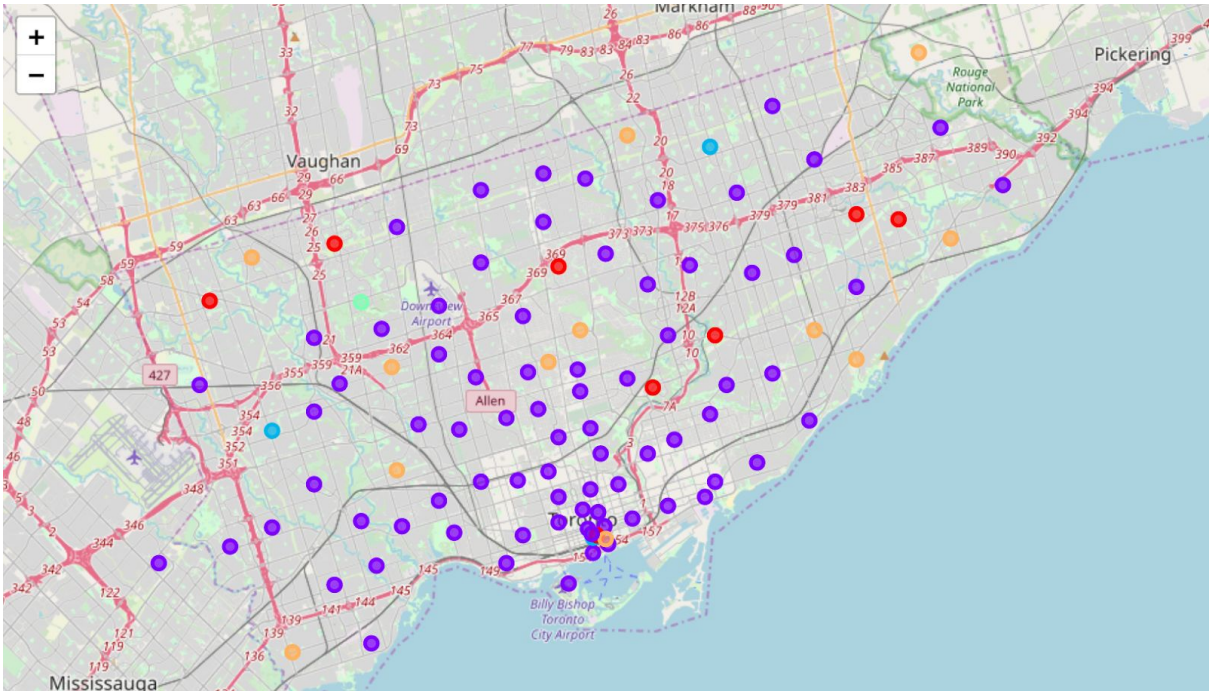

Figure 4. Neighborhood clustering

	PostalCode	Borough	Neighborhood	VenueName	VenueCategory	VenueLatitude	VenueLongitude
228	M5B	Downtown Toronto	Ryerson,Garden District	Elgin And Winter Garden Theatres	Theater	43.653394	-79.378507
351	M5C	Downtown Toronto	St. James Town	Fahrenheit Coffee	Coffee Shop	43.652384	-79.372719
352	M5C	Downtown Toronto	St. James Town	Pearl Diver	Gastropub	43.651481	-79.373600
353	M5C	Downtown Toronto	St. James Town	Triple A Bar (AAA)	BBQ Joint	43.651658	-79.372720
355	M5C	Downtown Toronto	St. James Town	Versus Coffee	Coffee Shop	43.651213	-79.375236
358	M5C	Downtown Toronto	St. James Town	Cambridge Suites Toronto	Hotel	43.651836	-79.378107
364	M5C	Downtown Toronto	St. James Town	Vintage Conservatory	Speakeasy	43.652229	-79.373487

Figure 5. Nearby venues list

	PostalCode	Borough	Neighborhood	VenueName	VenueCategory	VenueLatitude	VenueLongitude
347	M5C	Downtown Toronto	St. James Town	Gyu-Kaku Japanese BBQ	Japanese Restaurant	43.651422	-79.375047
349	M5C	Downtown Toronto	St. James Town	Terroni	Italian Restaurant	43.650927	-79.375602
350	M5C	Downtown Toronto	St. James Town	GEORGE Restaurant	Restaurant	43.653346	-79.374445
354	M5C	Downtown Toronto	St. James Town	Mystic Muffin	Middle Eastern Restaurant	43.652484	-79.372655
359	M5C	Downtown Toronto	St. James Town	NAMI	Japanese Restaurant	43.650853	-79.375887
361	M5C	Downtown Toronto	St. James Town	The Carbon Bar	Restaurant	43.653367	-79.374965
363	M5C	Downtown Toronto	St. James Town	La Bettola Di Terroni	Italian Restaurant	43.651993	-79.378056

Figure 6. Nearby competitors list

d. Neighborhood Clustering:

Neighborhood clustering is a way to get insight about similar neighborhood w.r.t venues categories (Figure 4.)

e. Nearby and competitors venues list:

The goal of this part is to explore a specific place. Given a radius, an address and competitors categories, two lists of nearby and competitors venues are extracted (Figure 5. and Figure 6.)

5. Discussion

In this part we discuss one limitation and one perspective in our work. The approach that we followed supposes that we can get all venues or at least most of them. But, this is not the case, which causes biases in our data and our analysis. As we have seen for example for crowdedness analysis, downtown regions are not crowded as expected.

We used a Foursquare free account. Other subscriptions are more flexible and can offer more possibilities. For example, information like popularity and rating give more information and enable deeper analysis of surrounding venues for a given address.

6. Conclusion

In this project, we offered an analysis of crowdedness and competitiveness in Toronto city. We made an analysis of venues. We clustered venues w.r.t their categories and implemented functions to get venues informations about surrounding venues of a given address. We also selected competitors within a given radius. This project can be extended to other places and other competitors categories.

7. References

[1] https://www.inc.com/encyclopedia/relocation.html

[2] https://www.entrepreneur.com/slideshow/299849

[3] https://en.wikipedia.org/wiki/List_of_postal_codes_of_Canada:_M

[4] https://developer.foursquare.com/

[5] https://geopy.readthedocs.io/en/stable/