Location for new Toronto Coffee Shop

Ahmed Fahim

June 14, 2021

1. Introduction

1.1. Background

Coffee is the most popular adult beverage in Canada. Canadians love their coffee more than beer, wine, soda, and even tap water! According to a study, men consumed on average about 3 cups of coffee a day, while women consumed about 2.4 cups. Coffee is mostly consumed at home, and the most common time for a cup of coffee in Canada is around breakfast. When it comes to coffee taken outside of the home, the majority of Canadians prefer to do so at a restaurant or at work. Nevertheless, coffee shops are still very popular in Canada, with two of three Canadians visiting Coffee shops at least once a month.

Despite all this popularity, the location of a coffee shop still plays great impact on its success (or failure). An article by startmycoffeeshop, gives a great introduction to anyone choosing a spot for a new coffee shop. According to the article, there are a number of factors to watch out for when choosing location for setting up a new coffee shop. The number and type of businesses were deemed as particularly important. So for anyone looking to start a new coffee shop in the Toronto area, it would be a good idea if we could segment the place, and spot the desirable areas.

1.2. Problem

The analysis proposed only requires data on the businesses that operate in the Toronto area. The article mentions dozens of pointers in choosing location for coffee shop; however, the main ones are summarized below:

1. Most of the coffee shop business is taken by popular chain stores like Tim Horton's and Starbucks. So for new coffee shops, it should be as far as possible from these chain stores.
2. The main clientele for coffee shops consists of office workers picking up a coffee on their way to work. Hence, coffee shops should try to place themselves as close as possible to business centers and offices.

2. Data acquisition and Cleaning

2.1. Data source

First and foremost we need a map of the Toronto area. This was obtained from the Toronto Open Data Portal. After this we need to get the businesses and their locations using the Foursquare API. But before we can do that we need to cover the Toronto area with a grid to span the whole area without missing anything. This was done using the python h3 library using a cell area resolution of 8, which covered the whole Toronto area with hexagonal grids with a side length of a little less than 500 meters. The cell resolution was selected such that limit on the number of results returned by the API call was never reached, ensuring that our list is complete. And then we used the Foursquare API centered at the centroids of each of the hexagons to locate the businesses.

Finally we used geopy Nominatim to convert our desirable location coordinates into actual addresses.

2.2. Data Cleaning

The data obtained from the Toronto Open Data Portal had many columns but only area name, area code and geometry (i.e. the polygons, or shapes, that define the outline of the neighborhoods in Toronto) columns were chosen. The other columns were irrelevant for our analysis.

The data returned by the Foursquare API also had many irrelevant information that was also removed. The data that remained include the name of the business, latitude, longitude and the category of business. As we wanted our list to be complete, in many places there were overlaps in our search which resulted in duplicate entries which were later removed. Moreover, some of the names had characters like ‘ and ` which caused difficulty in printing and displaying the results. These characters were removed from the names.

3. Methodology

3.1. Exploratory Data analysis

Having obtained the required data our first step is to visualize the data to spot any obvious trends and relationships. As most of the data is geographical it makes sense to first of all visualize the geographical distribution of the data. We will be using the python library folium for this purpose.

First of all, we will visualize the distribution of coffee shops in the area. We used a choropleth map to show the number of coffee shops in each neighborhood, and on top of that we used marker clusters to show the individual coffee shops.

The figure below shows that most of the neighborhoods in Toronto have very few, if any, coffee shops. However, some on the other hand has over a hundred. The top two neighborhoods for coffee shops are Waterfront Communities – The island and Bay Street Corridor. Both of these neighborhoods are quite densely populated with lots of businesses.

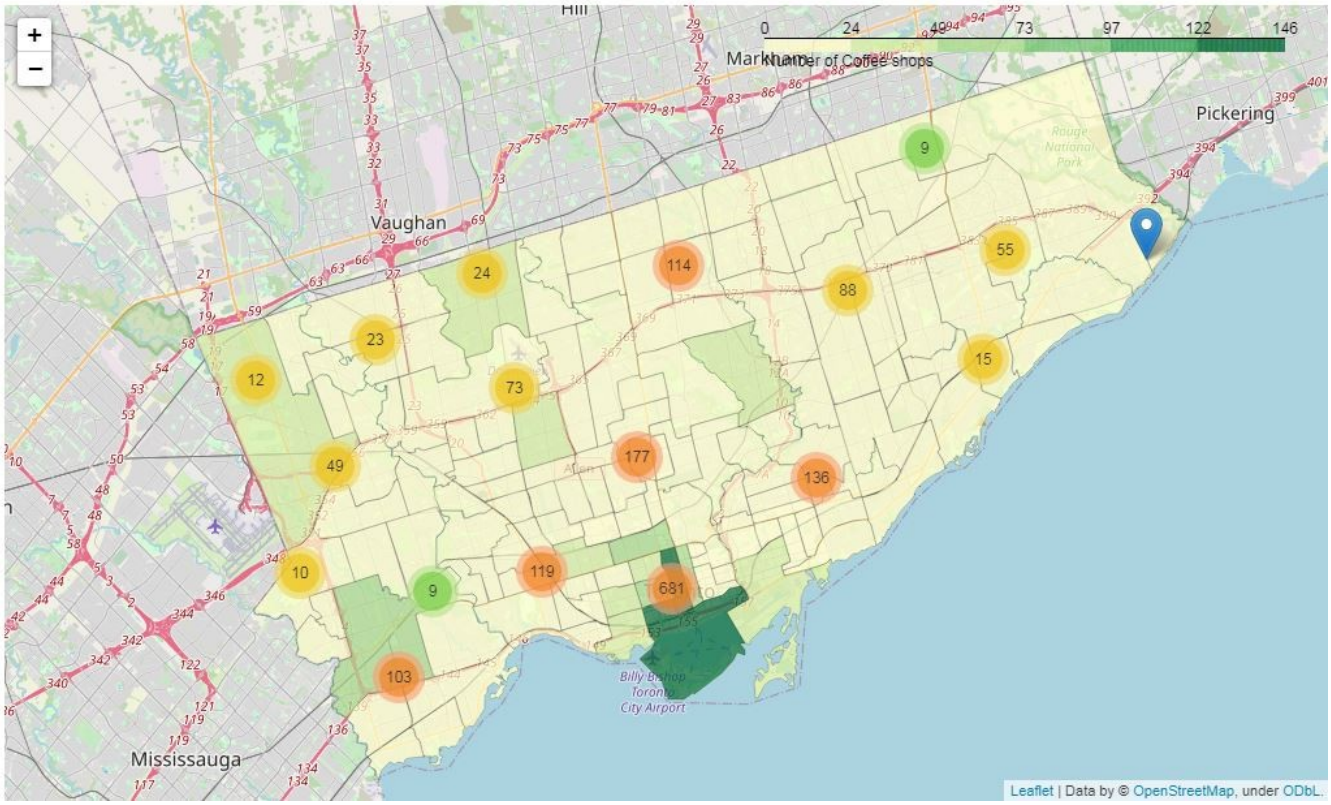

Figure 1: Distribution of Coffee shops in Toronto

We will use a similar visualization for offices, i.e. a choropleth map to show the neighborhood level distribution and with marker clusters on top to show individual offices. Once again, this shows that most neighborhood has very few offices, which would explain the lack of coffee shops for most of the Toronto area. We also notice that the top two neighborhoods for coffee shops are also the top two neighborhoods for offices.

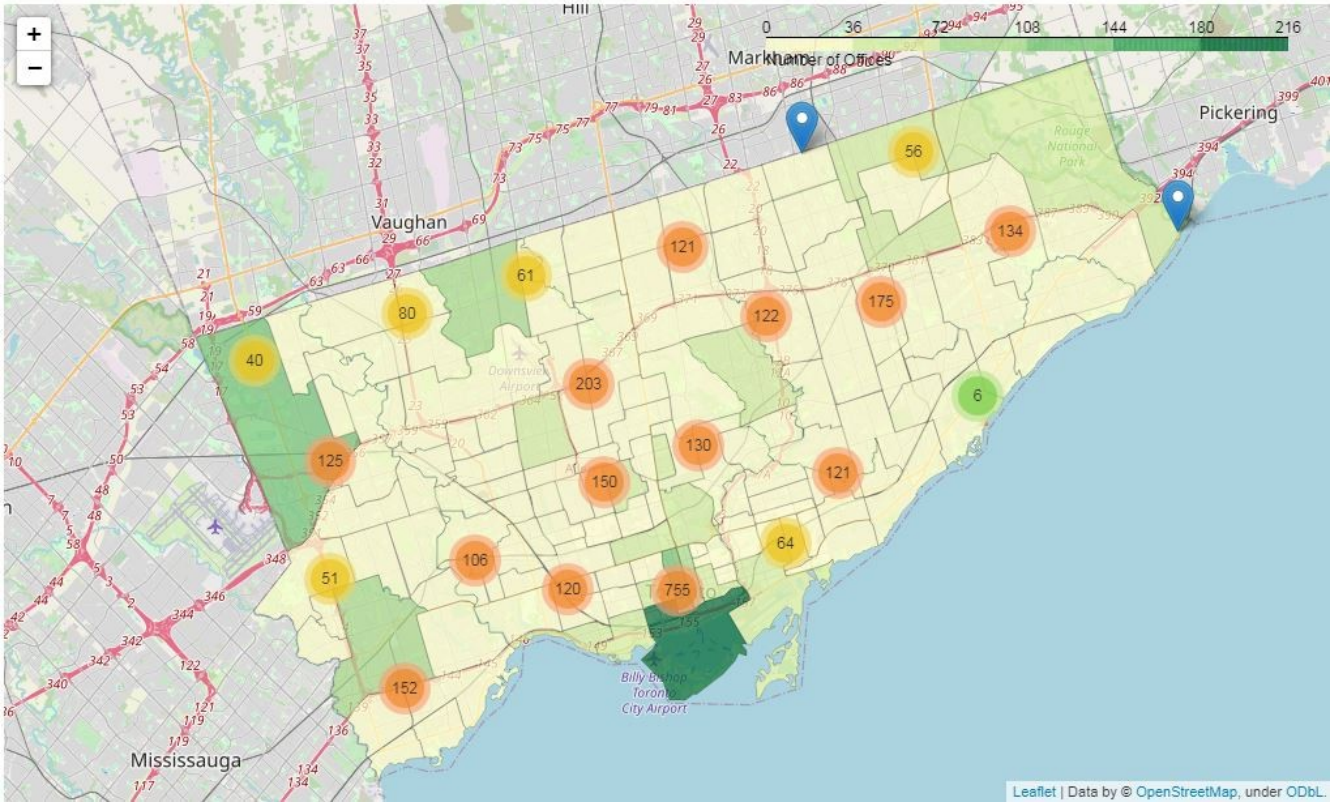

Figure 2: Distribution of offices in Toronto

Next we need to identify the chain coffee stores, because it is needed for the analysis proposed by the article. From the list and pie chart below we can see that a number of coffee shops make up a large portion of the market. Tim Horton's, which is the most popular coffee shop in Canada, make up over 20 % of coffee shops; Starbucks is second with just below 14% of the market.

We can see from the list that many coffee shops have 6 or fewer coffee shops in the area. Hence, this is taken as the cutoff point to determine chain stores. That is, coffee shops that have more than 6 stores in the Toronto area are taken as chain coffee stops that must be avoided. The chain coffee shops make up over 50 % of the market.

```
Total number of coffee shops in Toronto is 1698.

Tim Hortons has 344 coffee shops in Toronto.
Starbucks has 224 coffee shops in Toronto.
Second Cup has 80 coffee shops in Toronto.
Country Style has 41 coffee shops in Toronto.
Coffee Time has 38 coffee shops in Toronto.
Aroma Espresso Bar has 32 coffee shops in Toronto.
Timothys World Coffee has 27 coffee shops in Toronto.
La Prep has 15 coffee shops in Toronto.
Balzacs Coffee has 6 coffee shops in Toronto.
Tim Hortons / Esso has 6 coffee shops in Toronto.
Jimmys Coffee has 6 coffee shops in Toronto.
Pilot Coffee Roasters has 6 coffee shops in Toronto.
Dark Horse Espresso Bar has 6 coffee shops in Toronto.
Delimark Cafe has 5 coffee shops in Toronto.
Wallace Espresso has 4 coffee shops in Toronto.
Java Joes has 4 coffee shops in Toronto.
Coffee time has 4 coffee shops in Toronto.
DAVIDsTEA has 4 coffee shops in Toronto.
Sam James Coffee Bar has 4 coffee shops in Toronto.
```

Figure 3: List of chain coffee shops

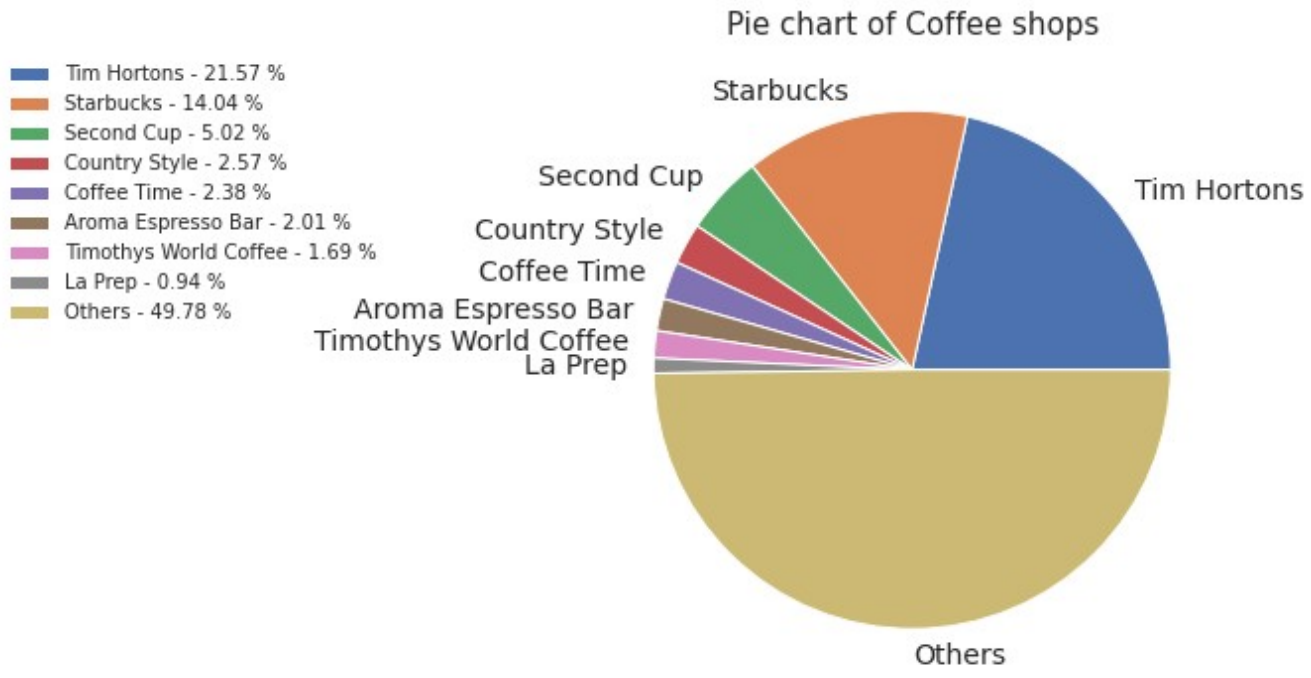

Figure 4: Pie chart of coffee shops

3.2. Scoring, Clustering, and Heat map

Now that we've done our preliminary analysis, its time to spot some favorable locations. To accomplish this we will cover the entire area with a grid and find locations on the grid that are desirable for setting up coffee shops. And in order to quantify the desirability of location we will assign it a score. The choice of which scoring function to use was somewhat arbitrary, but it should follow our intuitions and the guidelines given in the article.

The score is calculated using the distances to nearby businesses; only those that are within 500 meters from the center of our grid are considered 'nearby businesses'. We calculate the inverse of the distance to the nearby businesses and scale it by factor depending on the type of business. For example chain coffee shops are scaled by a factor of -1.9; regular coffee shops are scaled by a factor -0.6; and finally offices are scaled by a factor of +1.5. The sum of all these weightings make up the score for a particular location. The scores for all locations are then scaled so that it lies in the interval 0 to 1.

After all the grid points have been scored, we now cluster them according to the score. We used python yellowbrick library to find the best number of clusters to use in kmeans clustering. Next we chose the clusters that had the highest scores and plotted the points as a heat map with weightings onto the map of Toronto. The map is shown below, where we plotted the clusters labeled 1 and 3.

The map below shows that the best spot for a new coffee shop is located within the hub of offices and other coffee shops.

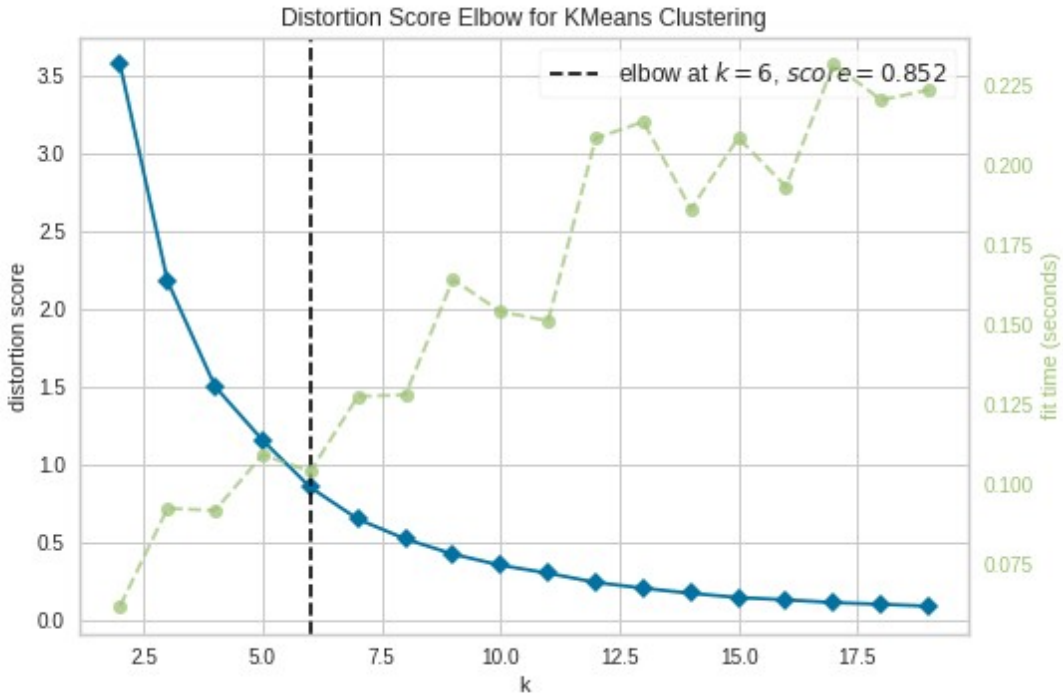

Figure 5: Finding best value for k using yellowbrick

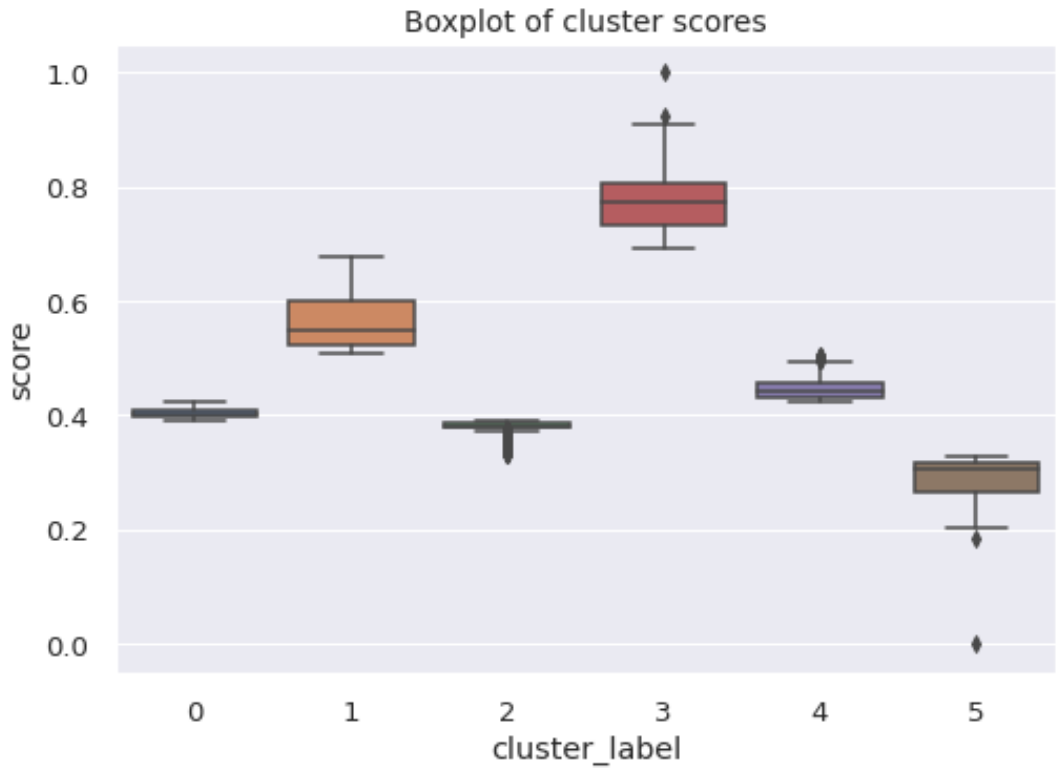

Figure 6: Boxplot of cluster scores

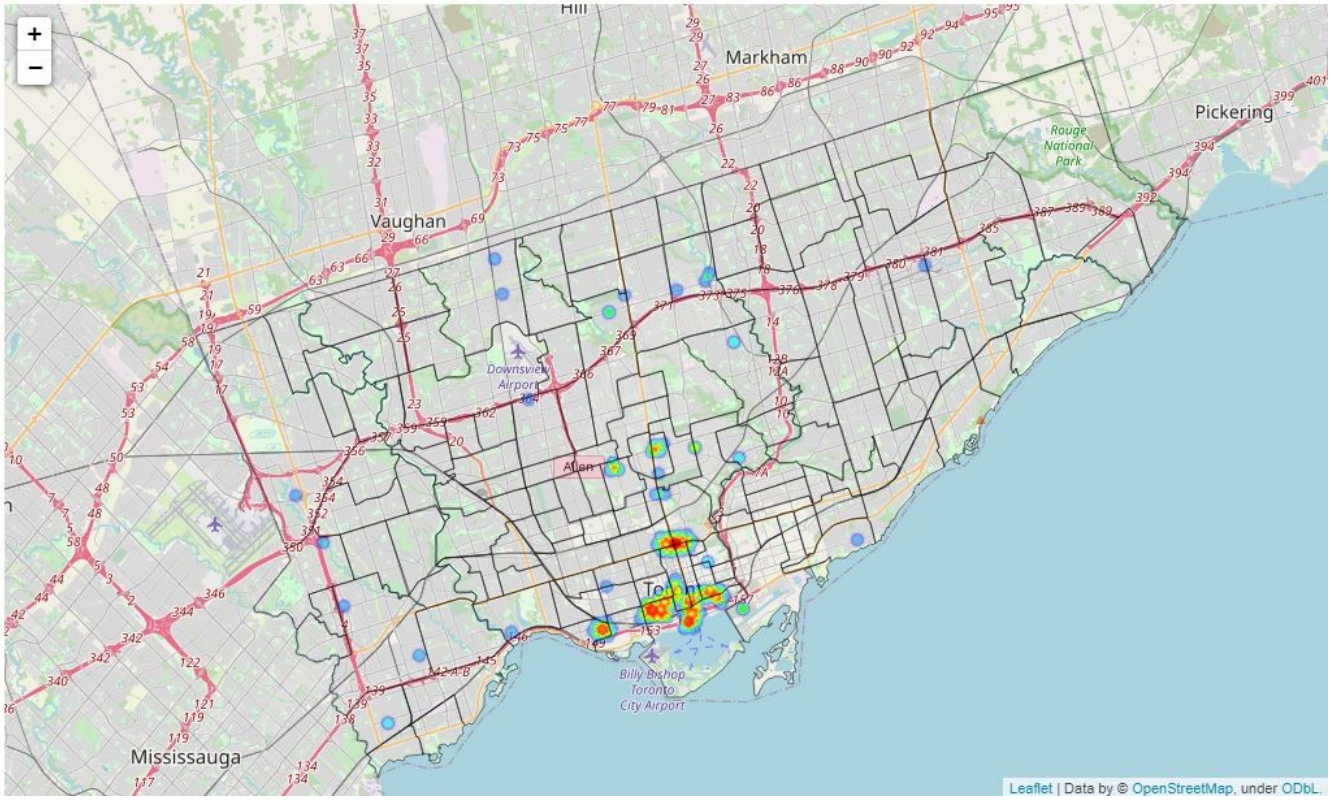

Figure 7: Heatmap of potential locations

4. Results

The analysis revealed locations that might be good for new coffee shops. The best locations are given below.

```
The no. 1 location for coffee shop with a score of 1.00 out of 1 is:
Hudson's Bay Centre, Park Road, Yorkville, University–Rosedale, Old Toronto, Toronto

The no. 2 location for coffee shop with a score of 0.93 out of 1 is:
RBC Waterpark Place III, 85, Harbour Street, Harbourfront, Spadina–Fort York, Old Toronto, Toronto

The no. 3 location for coffee shop with a score of 0.91 out of 1 is:
Jacobs & Co. Steakhouse, Brant Street, Fashion District, Spadina–Fort York, Old Toronto, Toronto

The no. 4 location for coffee shop with a score of 0.89 out of 1 is:
72, Fraser Avenue, Little Tibet, Spadina–Fort York, Old Toronto, Toronto

The no. 5 location for coffee shop with a score of 0.82 out of 1 is:
Joey & Toby Tanenbaum Opera Centre, Nicholson Lane, St. Lawrence, Toronto Centre, Old Toronto, Toronto
```

Figure 8: List of addresses for new coffee shop

5. Discussion

The analysis is contingent on a number of key assumptions which are given below:

- The guidelines given in the article by startmycoffeeshop are correct.
- The truncated version of guidelines used in this analysis was reasonably correct.
- The scoring function used to score the grid points was appropriate.

Of these assumptions, the first one is most likely to be met. The second one should also be reasonably met. The third point however, is open to discussion. Nevertheless, I tried a number of scoring function having various functional forms, but they all seemed to point towards roughly the same locations. Hence, I also believe that my analysis was correct.

Nevertheless, choosing an actual location will require considerable human judgment as well. For example the place might be too expensive one’s choice.

6. Conclusion

In this study I analyzed the Toronto area in search for suitable spots for a new coffee shop. Bulk of the data was obtained using the Foursquare API. And the method of finding the best location for coffee shop was taken from the web article by startmycoffeeshop. I then transformed the main points from the article into a set of scoring criteria, which I evaluated for each point in the hexagonal grid that spanned the entire area. This resulted in a grid with a score assigned to each point that denoted the desirability

of the location for a new coffee shop. Finally I clustered the points and plotted the high scoring grid points onto a heat map.