Report

Introduction

My client is planning to start a coffee shop business in Downtown Toronto. Prior to opening a coffee shop, choosing a location is one of the important tasks to success. A promising chosen location provides the owner with a stable traffic of customers and profits. Convenience and visibility matter a lot. Few people are willing to walk more than thirty minutes for a cup of coffee or some light refreshments. The location should be convenient and visible for target customers. It is good for people to have a breakfast or grab a cup of coffee on the way to work or to school. The surrounding land use is very important. Besides, train station and bus stop can boost passenger traffic in nearby streets. Apart from the movement of people, the number of competitors is also important. Restaurants, coffee shops, cafés and bakeries are little similar catering. They share the catering market. Some of them are the direct competitors to the new coffee shop.

My client wants to know where the right location is for their new coffee shop in Downtown Toronto.

Data

The postal codes of Canada are obtained from Wikipedia [1]. The geographical coordinates of each neighborhood from 'http://cocl.us'. Besides, the categories of venues are obtained from Foursquare [2].

Data cleaning:
Some borough information is missing. Any cells with a borough that is Not assigned is removed. Besides, those neighborhoods that exist in the same post code area are combined. Those Not assigned neighborhood columns are filled with its borough.

The distance between venues is calculated with their geographical coordinates. Venues will be segmented into several clusters with their geographical information. Venue categories such as breakfast, coffee shop, café and bakery are regarded as direct competitors. Restaurants and diners are regarded as restaurants. The remaining venues are regarded as other venues. As a result, there are three groups. They are restaurants, direct competitors and other venues.

Methodology

The total area of Downtown Toronto is about 17 km^2 [3]. The size of each neighborhood is very different, so the borough's area is segmented using a k-Means clustering algorithm. The algorithm minimizes the intra-cluster distances and maximizes the inter-cluster distances.

Venues in each cluster are grouped by three main categories such as restaurants, direct competitors and other venues. The first category named restaurants includes restaurant and diner. The second category includes breakfast, coffee shop, café and bakery. Since they are the direct competitors to the new coffee shop, the category is called direct competitors. The last category includes all the remaining venues, so the name is other venues.

In general, the more the number of venues, the higher the chance of people movement in the cluster will be. Thus, the number of venues except catering will be considered. Another factor will increase the movement of people in the nearby area is traffic interchange like train station. The number of train stations are concerned.

Any new coffee shop will face a big challenge of the catering market share. Hence, the number of the similar caterings and the direct competitors are also counted. The catering to other venue ratios in each cluster are compared.

Results

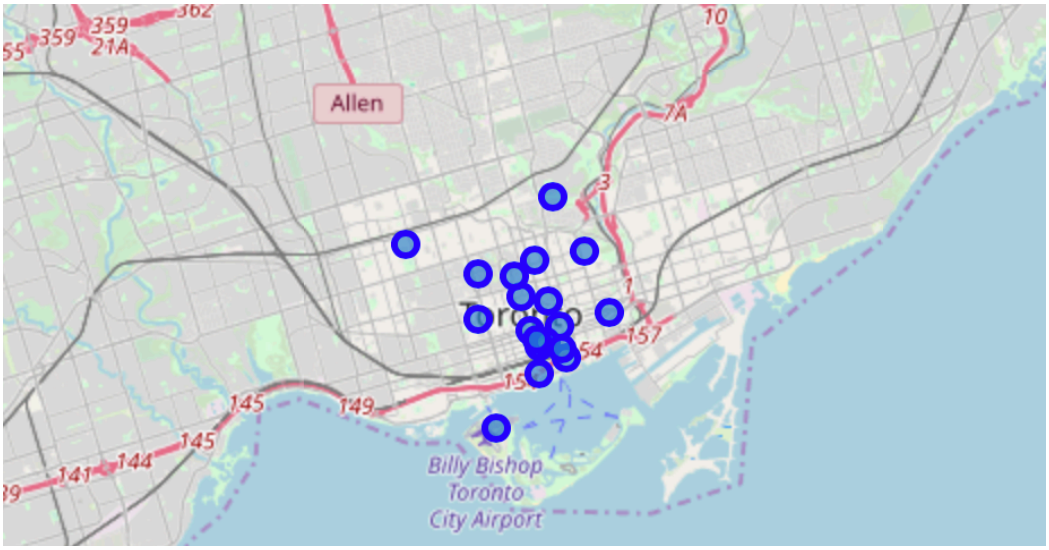

Figure 1 Neighbourhoods in Downtown Toronto

All neighborhoods are shown on the above map (Figure 1). There are 20 neighborhoods in Downtown Toronto.

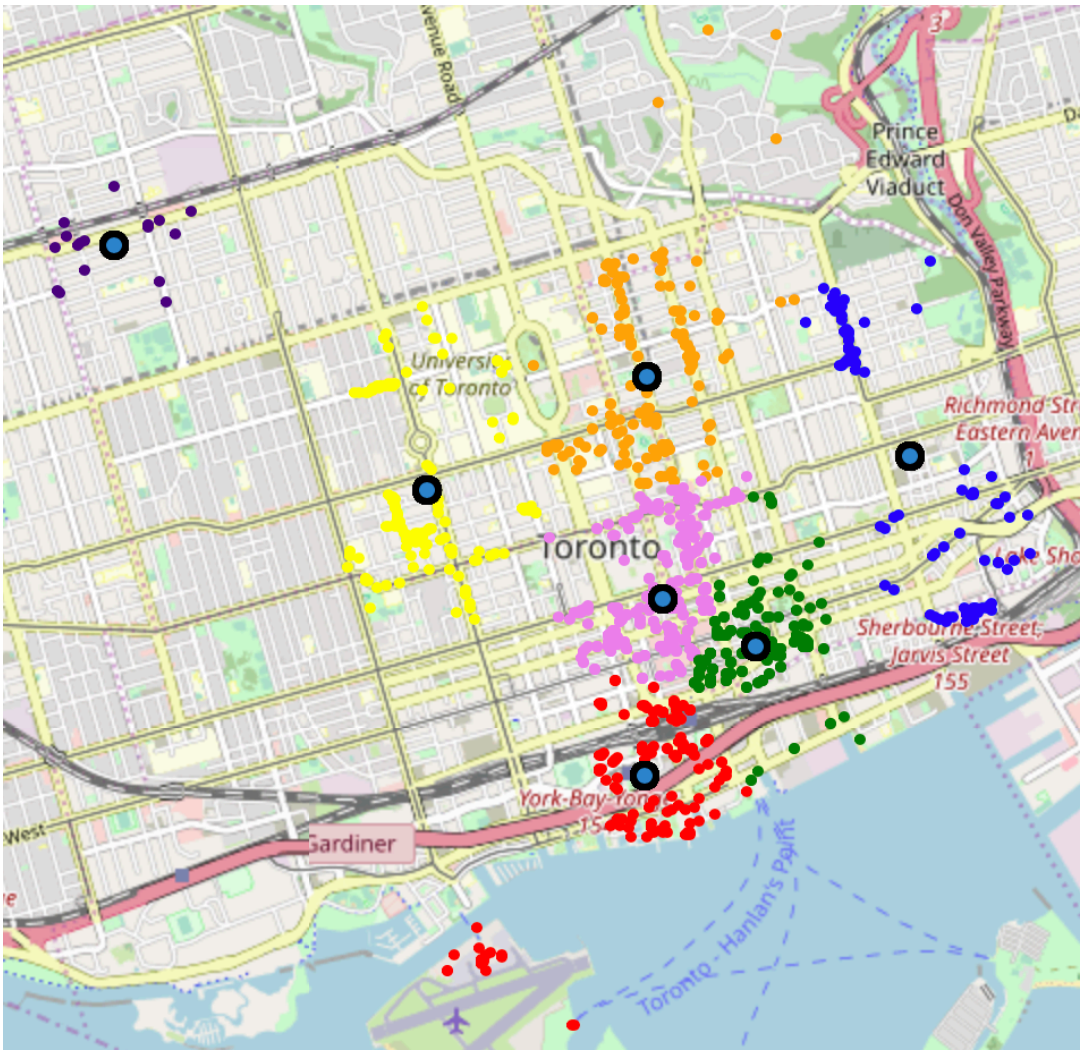

Figure 2 Clusters in Downtown Toronto

All venues in the borough segmented by k-Means clustering algorithm are shown in the figure 2 with respective color. Red, orange, yellow, green, blue, indigo and violet are for cluster 0, cluster 1, cluster 2, cluster 3, cluster 4, cluster 5 and cluster 6 respectively.

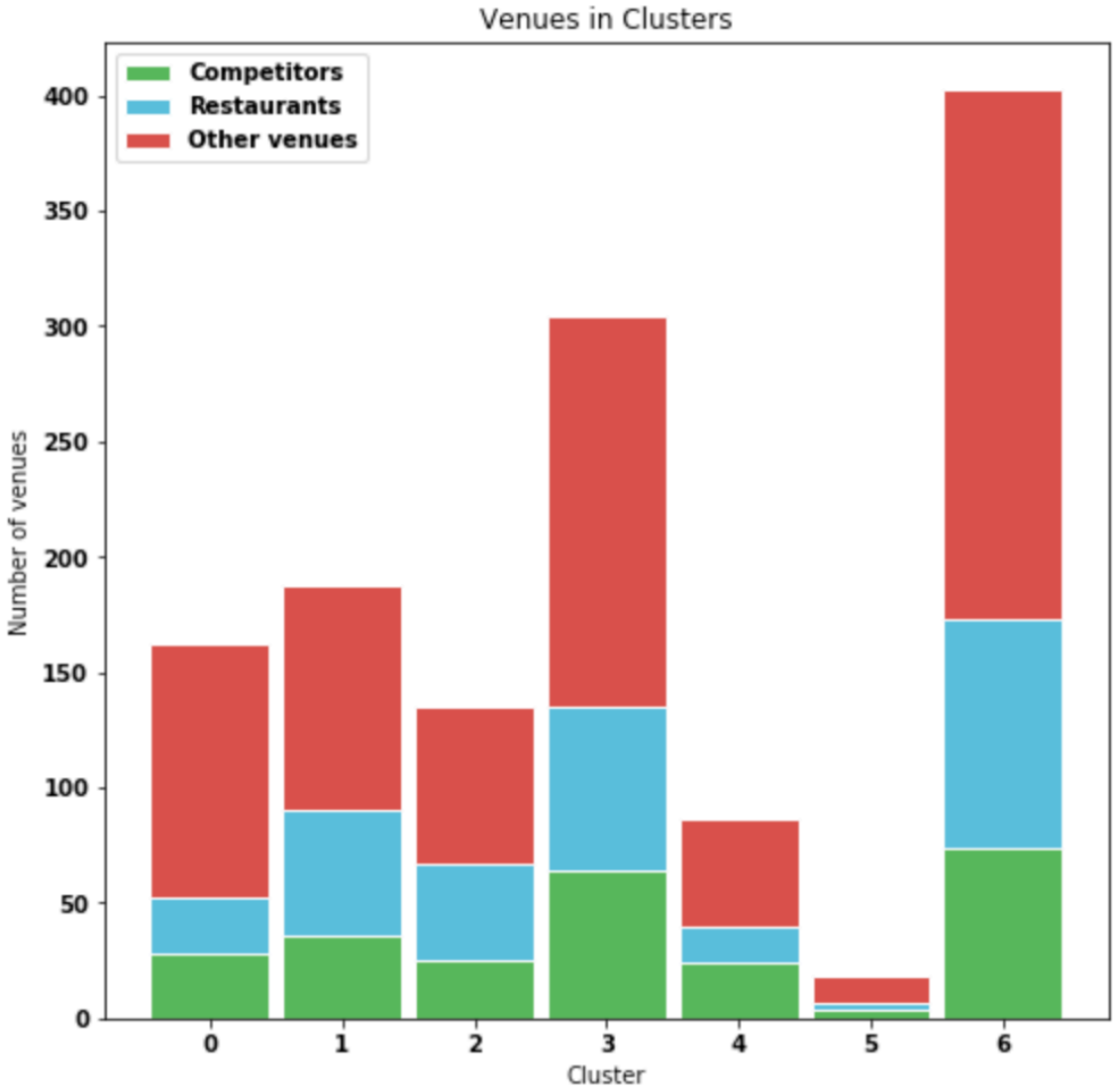

Figure 3 A bar chart of number of venues in clusters

In the figure 3, the Cluster 6 have the greatest number of venues. The top 3 of the greatest number of venues are in Cluster 6, Cluster 3 and Cluster 1.

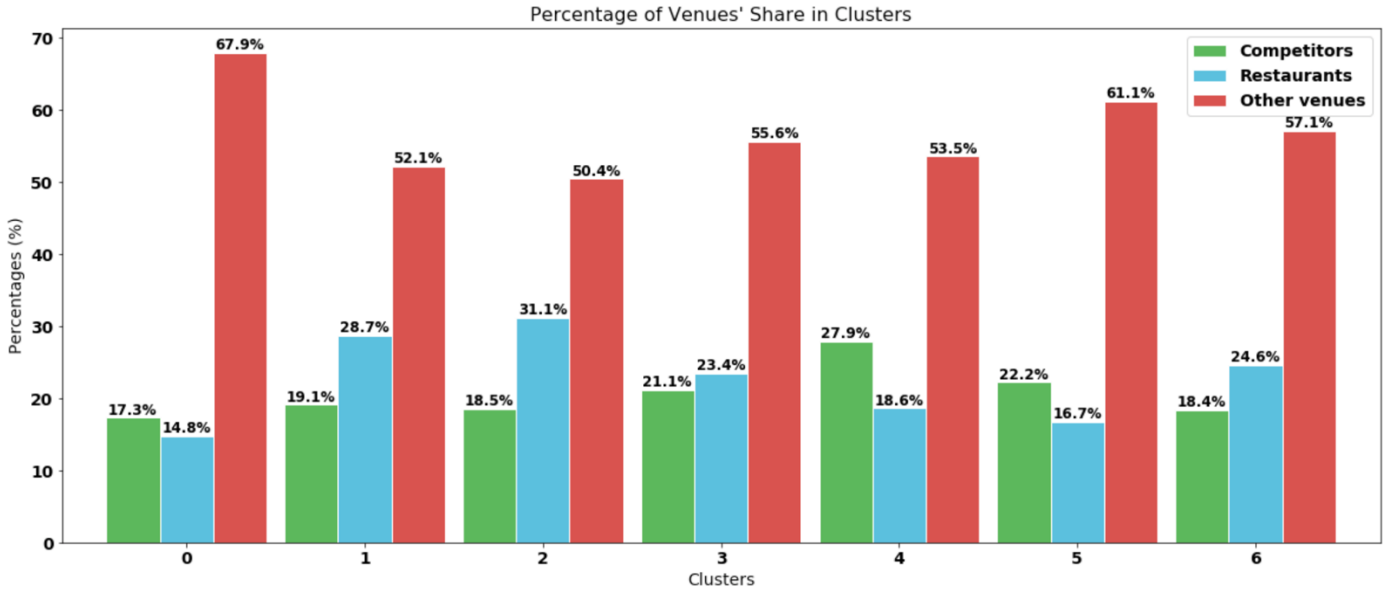

Figure 4 A bar chart of venues' share in clusters

In the figure 4, the lowest percentage of venues share in the category of competitors is Cluster 0. Besides, the lowest percentage of venues share in the category of restaurants is Cluster 0.

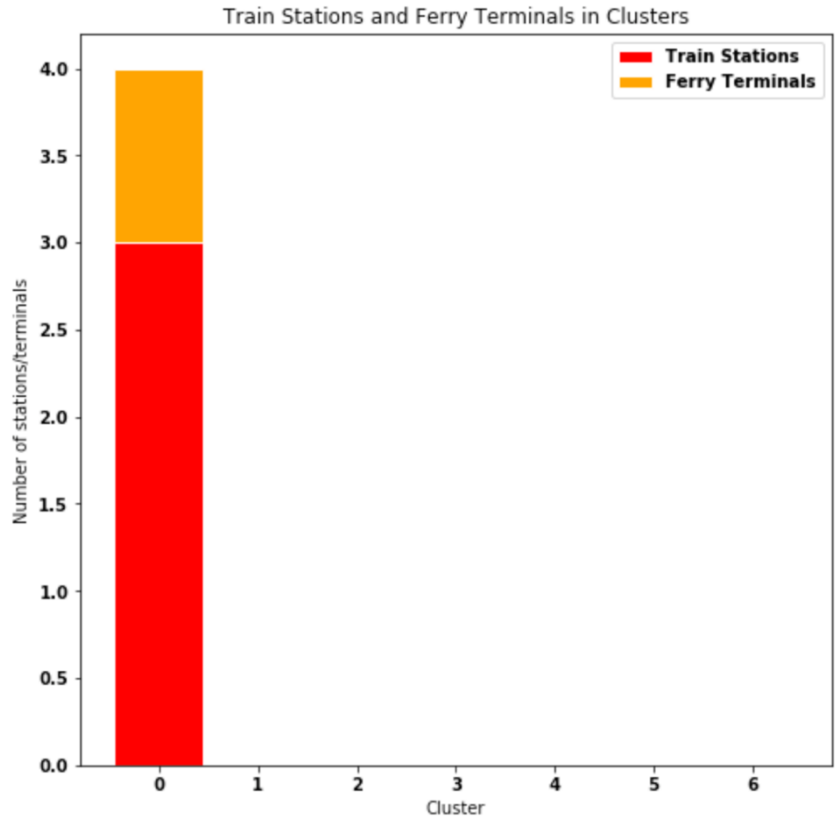

Figure 5 A bar chart of number of train stations and ferry terminals

The figure 5 shows that the Cluster 0 has 3 train stations and 2 ferry terminals while the other clusters does not have any train stations and ferry terminals.

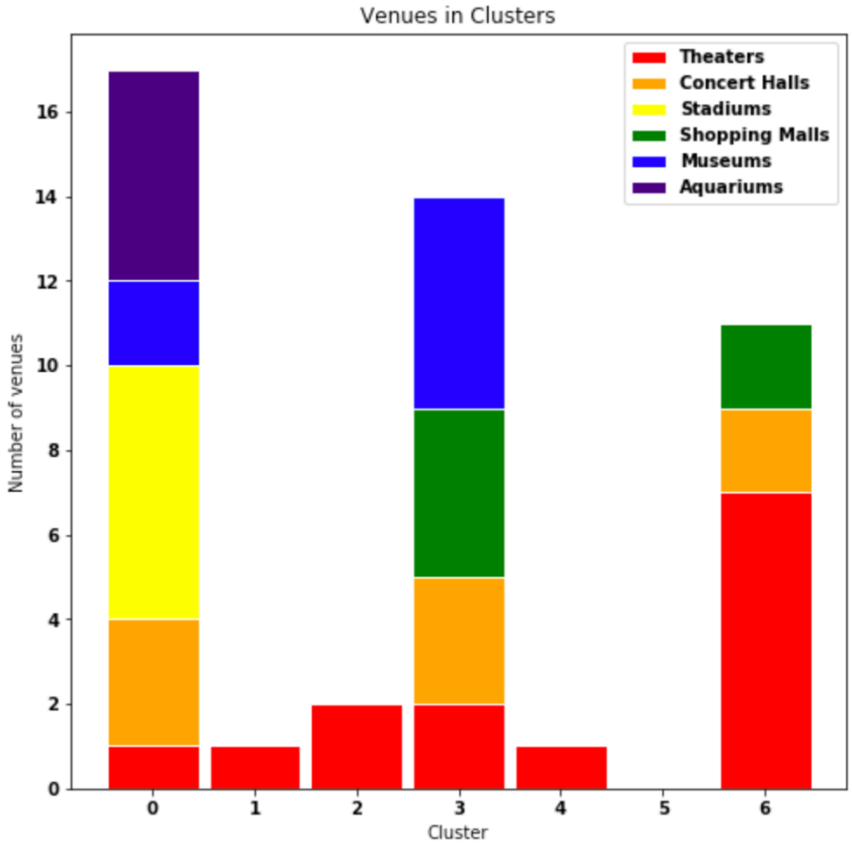

Figure 6 A bar chart of number of attractions

Some attractions such as theaters, concert halls, stadiums, shopping malls, museums and aquariums are counted. The number of them in each cluster are shown in the figure 6. The top 3 of the greatest number of attractions is Cluster 0, Cluster 3 and Cluster 6. The total number of attractions in Cluster 0, Cluster 3 and Cluster 6 are 17, 14 and 11 respectively.

Discussion

Slicing a big area into some smaller area is good for analyzing and comparing each area in order to find the right area to open a new coffee shop. In the figure 2, the area of each cluster is similar, but the density of venues in them are different. Some clusters have a larger number of venues, but some have smaller. It is important to survey how many venues nearby the new coffee shop. By observation in the figure 3, the Cluster 6 have the greatest number of venues. The top 3 of the greatest number of venues are in Cluster 6, Cluster 3 and Cluster 1. Apart from the different number of venues, there are different portions of the 3 main categories. In the figure 4, the lowest percentage of venues share in the category of competitors is Cluster 0. The percentage is 17.3%. Besides, the lowest percentage of venues share in the category of restaurants is Cluster 0. The percentage of the category of restaurants is 14.8%. Any new catering business will face a big challenge of market share. Those percentages in Cluster 0 and Cluster 5 have 32.1% and 38.9% respectively.

The second factor which will be advantageous to the movement of people in street is public transportation like train and ferry. The figure 5 shows that the Cluster 0 has 3 train stations and 2 ferry terminals while the other clusters does not have any train stations and ferry terminals.

Attractions are also important to increase the movement of potential customers in the nearby area. Theaters, concert halls, stadiums, shopping malls, museums and aquariums attract a lot of people to visit. The number of them in each cluster are shown in the figure 6. The top 3 of the greatest number of attractions is Cluster 0, Cluster 3 and Cluster 6.

Conclusion

By the analysis, the Cluster 0 has the lowest percentage of catering venues in the area. Besides, 3 train stations and a ferry terminal in the Cluster 0 is beneficial to raise people traffic. Finally, the Cluster 0 has the larger number of attractions. Thus, a new coffee shop will be suggested to open in the Cluster 0. The latitude and longitude coordinates of the cluster's center are 43.641395, -79.383690 respectively. A circle area with a radius of 700 meters about the cluster's center is shown in the figure 7 as below. A new coffee shop will be suggested to open inside the area.

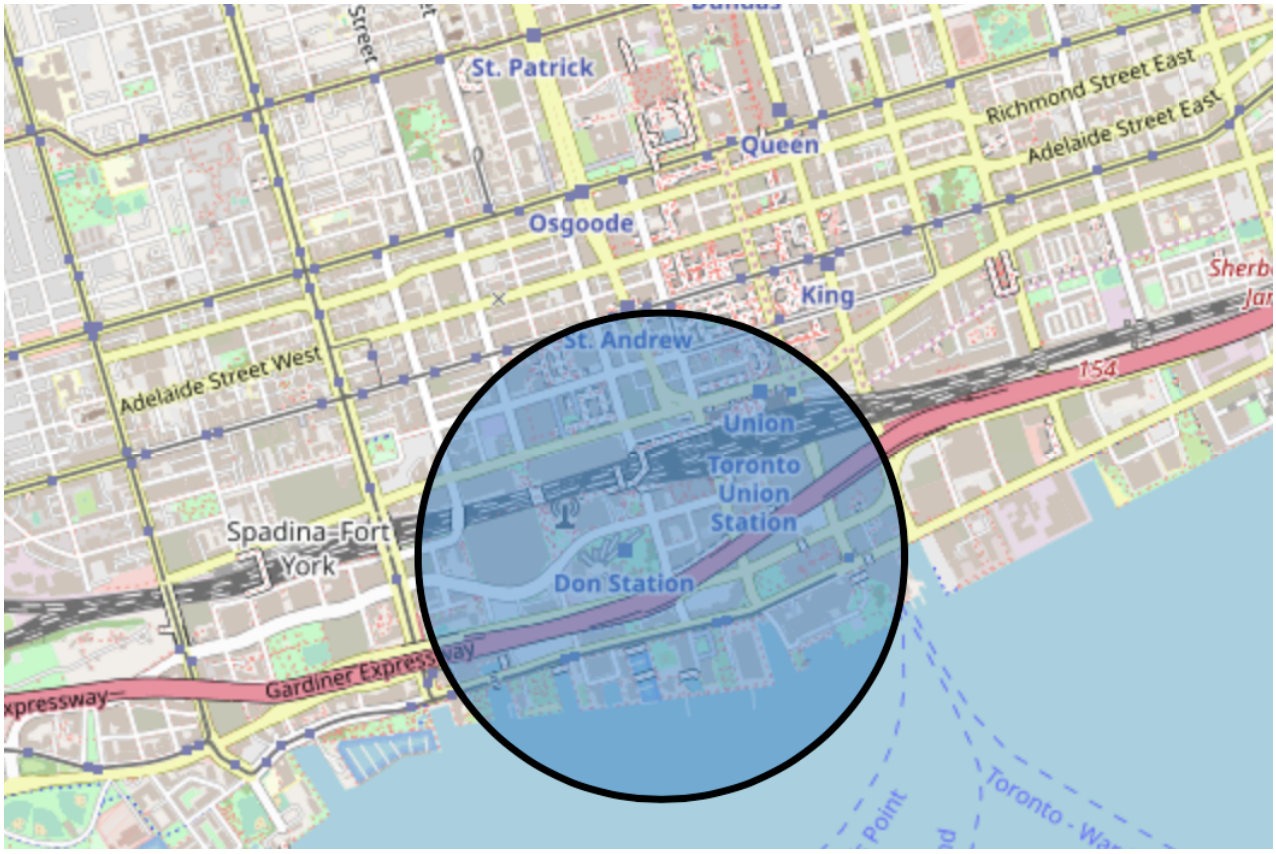

Figure 7 A suggested area for new coffee shop

References

1. https://en.wikipedia.org/wiki/List_of_postal_codes_of_Canada:_M
2. https://developer.foursquare.com
3. https://en.wikipedia.org/wiki/Downtown_Toronto