Best Location for a new Barbershop in Toronto

Yunkun Yang

Aug 11st, 2020

1. Introduction
   1. Background

Social science research indicates that people’s appearance has a significant impact on their success. Halo effect, a psychological concept, says positive impression of a person can positively influence other’s opinion and feelings. More and more people realize the advantage one positive improvement on their appearance can bring. Among all the changes you can make, getting a haircut is easiest, especially before some special events, such as a friend’s wedding, or an interview. However, it is not always easy to find a barbershop in some neighborhood, and people have to drive very far before they can find any barbershop.

In this report, we will mainly discuss which neighborhood has the fewest barbershop and where is the most necessary place to open a new barbershop in Toronto. The requirement for this location is that we need the neighborhood with fewer than 4 barbershops and the population to be over 5000.

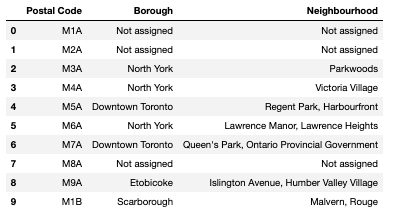
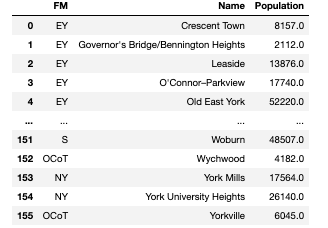
* 1. Interested Audience

The interested audience should be those investors who hope to open a new barbershop and those residents who lives in nearby neighborhoods.

1. Data Sources

All the neighborhoods data are scraped from Wikipedia pages. First page is the list of postal codes of Canada: M [[Link](https://en.wikipedia.org/wiki/List_of_postal_codes_of_Canada:_M)], from where we are able to load all postal codes of all neighborhoods in Toronto. Second page is the demographic of Toronto neighborhoods [[link](https://en.wikipedia.org/wiki/Demographics_of_Toronto_neighbourhoods)], from where we are able to obtain all neighborhood detailed data in GTA area such as population.

All scraped data will be like following:



Note:

By definition from Wikipedia, we know that the abbrev of each city is noted below.

- Old City of Toronto (OCoT)

- Scarborough (S)

- North York (NY)

- Etobicoke (E)

- York (Y)

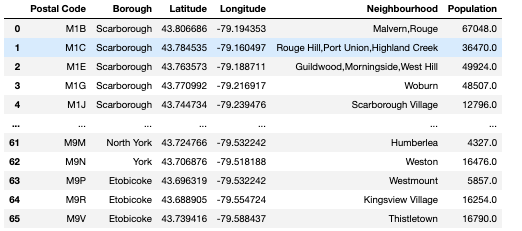
- East York (EY)

We are also able to obtain all spatial data with coordinates of neighborhoods in Toronto from the link [[here](http://cocl.us/Geospatial_data)]. With a map, we can easily get all detailed information about all neighborhoods in Toronto.

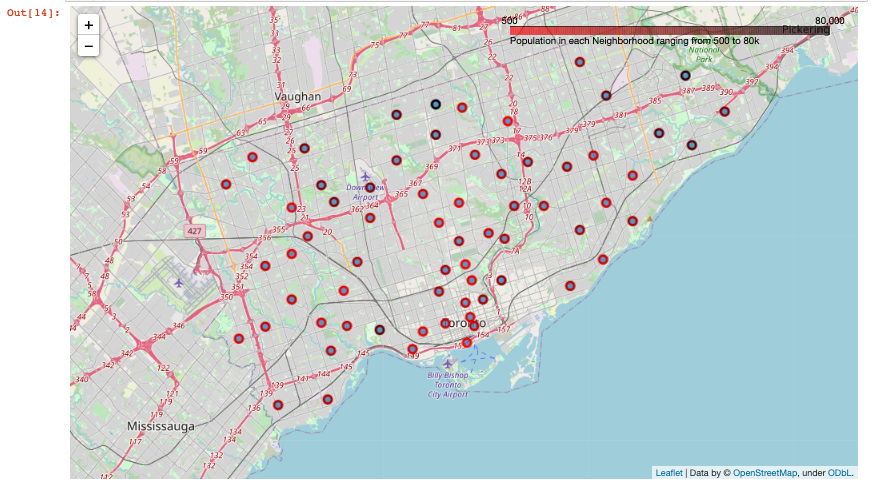
To get the venue data (all barbershop locations), we use Foursquare API to get those data. Foursquare API is a developer tool for us to scrape those real-world venues data. More details can be explored and found in the following link. [Foursquare](https://developer.foursquare.com/)

1. Exploratory Analysis

After we merge and cleaned the data, we are able to obtain a table with the locations and populations of each neighborhood.

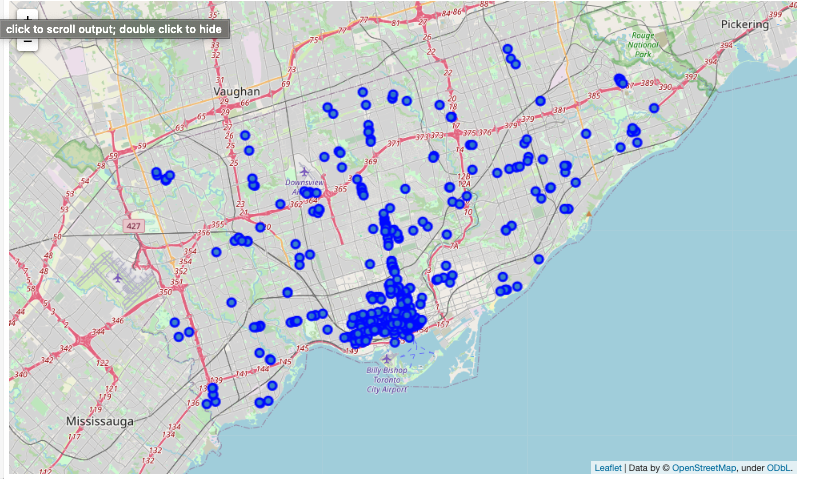


We are able to draw all the neighborhoods in the map to have a general idea about how this data appears.



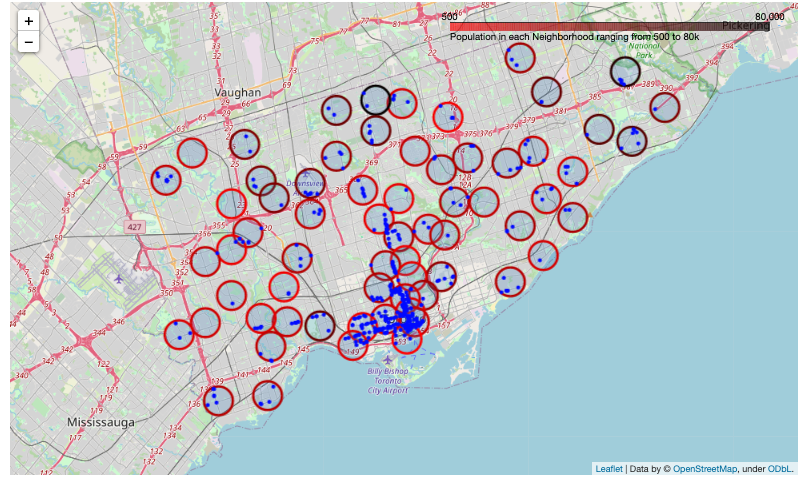
The color of dots indicates the number of population in each neighborhood. It can be seen that most of the population concentrate in Downtown, North York, Scarborough and Etobicoke. Population in Toronto Island, for example, is very small and appears to be a light dot in the map.

Secondly, after we get venue data using Foursquare API, we are also able to plot it in the map using blue dot.



We can see there are a lot of salons open in Downtown area and North York, but in other areas, salons still mainly concentrate in a few plazas.

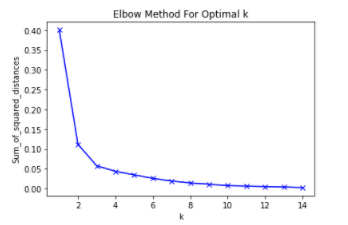
With the previous two maps overlaid together, we can somehow have a better sense on which neighborhood requires more new barbershop open.



Each red circle in the map is the circle with radius 1km around its center, and we can see there are overwhelming barbershops in Downtown area, greatly more than it should be while some neighborhoods in Etobicoke or Scarborough do not have enough barbershops around.

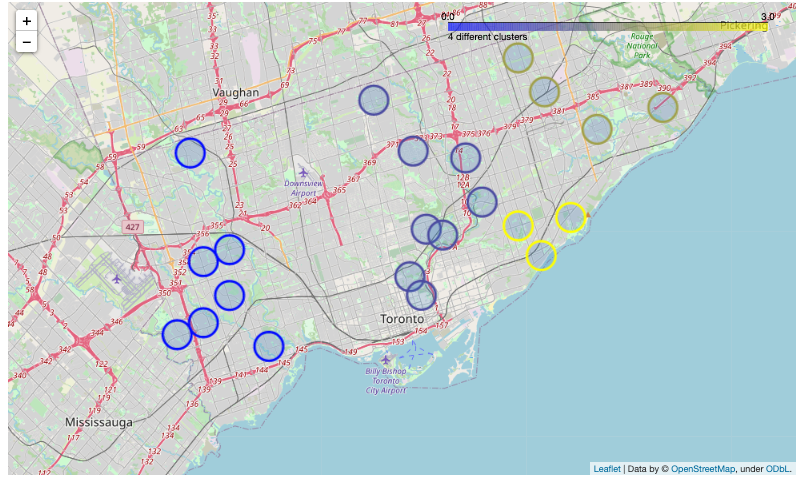
To tackle this problem, we will use k-means clustering techniques to cluster those neighborhoods with fewer barbershops and find the optimal location.

1. Modelling



After filtering those neighborhoods with more than 4 shops and fewer than 5000 people. We are able to run a k-means clustering to group those close neighborhoods together.

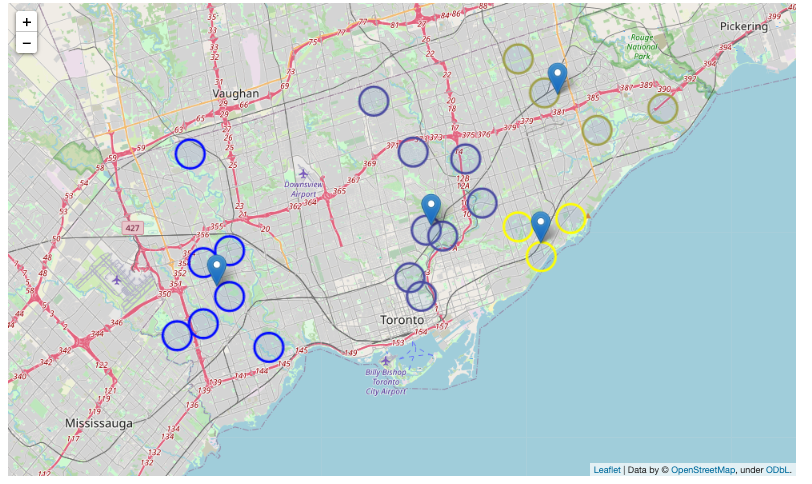
We first use elbow method to detect the best k for the dataset, as shown in the graph, I am able to find the k = 3.

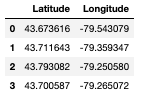
After modelling with k=3, we find that an extra cluster can be used so each different neighborhoods can be better separated. The final clusters can be shown in the map below.

To find the optimal location for each cluster, we decide to use the geometric median. Geometric median of a set of point is the median point which minimize the sum of the distance from the median point to each point.

1. Result and Discussion

We are able to find the optimal location has the latitude and longitude as stated below, and their following address are stated following



*5, Ravensbourne Crescent, Princess Anne Manor, Etobicoke, Etobicoke Centre, Etobicoke, Toronto, Golden Horseshoe, Ontario, M9A 2P7, Canada*

*90, Wicksteed Avenue, Leaside Business Park, Don Valley West, East York, Toronto, Golden Horseshoe, Ontario, M4G 4E8, Canada*

*Furniture Source, Shorting Road, Malvern West, Scarborough North, Scarborough, Toronto, Golden Horseshoe, Ontario, M1S 3V8, Canada*

*86, Aylesworth Avenue, Birchcliff, Scarborough Southwest, Scarborough, Toronto, Golden Horseshoe, Ontario, M1K 1A1, Canada*

The results are quite straightforward since they are just locations of where the barbershop should be opened. From the modelling perspective, since we remove those neighborhoods with more salons and fewer populations from our final modelling, adding those neighborhoods, may give different results.

What we are sure about this analysis is that those neighborhoods do require the new barbershops, but the actual location where those shops should be open can vary. More factors should be considered if an investor does hope to open the shop, such as the rent or the labor cost or even taxation in different cities. Different neighborhoods may contain different people who have very diverse preference, and that can also be a significant effect whether a new barbershop should be opened.

1. Conclusion

The purpose of the project is to detect which neighborhoods require new barbershop open and where is the optimal location. According to the Foursquare API, we are able to detect a few spots for those new shops, and they are already marked in the previous map. Without further information, investors can choose those spots or anywhere nearby to open a new shop.