# Where Should Tom Open His New Coffee Shop

## 1. Introduction

Tom wants to start a coffee shop in Toronto. But he is a beginner in this area, and he does not want to open the coffee at a neighborhood where there are many coffee shops and very competitive. However, if he open the coffee shop at a neighbhood where there is no coffee shop, such neighborhood may be lack of residents and tourists which can be a problem for the revenue. Therefore, he wants to open his coffee shop at the neighborhood which has the number of coffee shop closest to the mean number of coffee shop in all neighborhood of Toronto. Moreover, Tom is not very confident with the taste of his coffee, therefore, he prefer the neighborhood with average coffee shop rating less than 7.5/10.

## 2. Data acquisition and cleansing

## 2.1 Data source

The information of each neighborhood in Toronto were acquired from Wikipedia (<https://en.wikipedia.org/wiki/List_of_postal_codes_of_Canada:_M>).Their longitudinal and latitudinal data were obtained from Foursquare. Locations and name of the coffee shop in Toronto were also acquired from Foursquare database.

## 2.2 Data cleansing

Data scrapped from Wikipedia was converted into pandas dataframe. The dataframe consists of 3 columns: Postal code, borough, and neighborhood. There are some missing data in the borough column. Such row with missing borough was dropped. However, when the data of neighborhood is missing but not borough; the neighborhood will be assigned a name same as its borough. As some neighborhoods share the same postal code, the neighborhoods with same postal code were combined together and separated by comma.

After the neighborhood data was cleaned, the corresponding locations of them can be acquired using Foursquare databased. Next the adjacent coffee shops were searched. Exotic neighborhoods with no coffee shop were neglected. For all the coffee shop in Toronto, some of them do not have a rating which is meaningless to Tom when he considers the rating problem. Therefore, the coffee shops without ratings were dismissed. However, they were still considered when counting the number of coffee shop in each neighborhood.

## 3. Exploratory data analysis

## 3.1 Information of the neighborhood

After the neighborhood information of Toronto were scrapped from Wikipedia, they were visualized in Figure 1. As we can see, neighborhoods are densely located near the harbor.

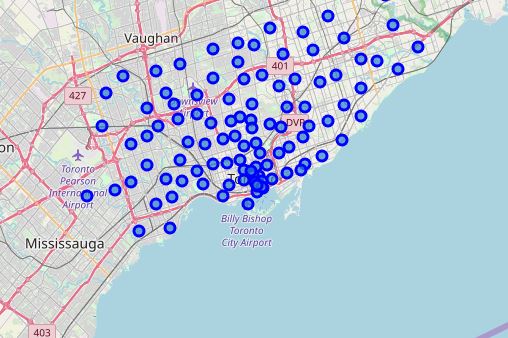


Figure Map of Toronto with neighborhood information

## 3.2 Information of the coffee shop

The information of the coffee shops in Toronto including the neighborhoods they are in and they exact location are shown in Table 1. They are visualized in the map of Toronto and shown in Figure 2 (blue dot). It can be noted that there is almost no coffee shop in the inland area.

Table Selected coffee shop information



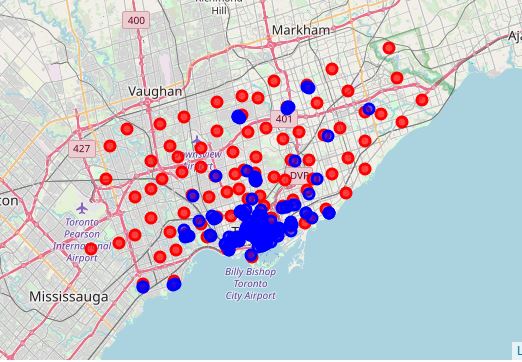


Figure Map of Toronto with information of coffee shop

## 3.3 Information of the coffee shop at each neighborhood

The number of the coffee shop was counted and ranked by descending order as shown in Figure 3. Tom’s coffee shop will be roughly located in the neighborhood with mean value of coffee shop in Toronto which is around 10.

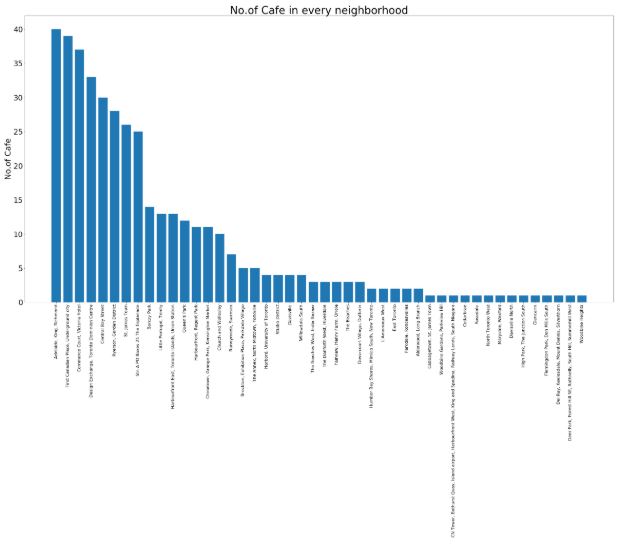
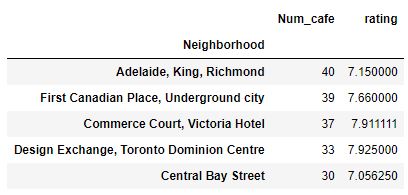


Figure Number of coffee shop in each neighborhood

## 3.4 Rating of the coffee shop

After we have the ID of every coffee shop, their rating in Foursquare can be obtained. Some ratings are missed and therefore they are meaningless to the final decision. Coffee shop without rating was dropped and the average ratings based on neighborhood were calculated and summarized in Table 2.

Table Summary of number and mean rating of coffee shop in each neighborhood



## 3.5 Location of the Tom’s coffee shop

After the number of coffee shop in each neighborhood were ranked by descending order, the average ratings were examined from middle of the list. The first neighborhood with average rating less than 7.5 are selected as the location of the Tom’s coffee shop. The neighborhood where Tom’s coffee shop is located is highlighted in black color and shown in Figure 4 Map of Toronto with location of Tom’s coffee shop

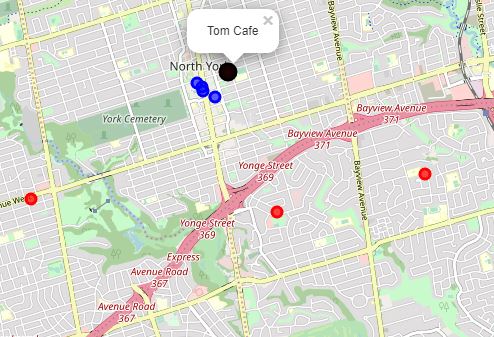


Figure Map of Toronto with location of Tom’s coffee shop

## 4. Conclusion

In this study, the location of Tom’s coffee shop was chosen based on the number of coffee shop in each neighborhood as well as their average ratings. After data acquisition and cleansing. So basic information were summarized and visualized. Finally, after having the information needed for the prediction, the desired location of the coffee shop was selected.