Coffee Shop Expansion in Downtown Toronto

Luying Pei

May 22, 2019

1. Introduction

1.1 Background

Downtown Toronto located in the city of Toronto, Ontario has always been a central location that is very popular among visitors and tourists from all over the world so there are a lot of business opportunities for small businesses such as shops, restaurants and bars. A small coffee shop located in Downtown Toronto has been doing very well in the past year and the owner of the coffee shop is looking to expand their business in other areas of the city. They currently have one location open on Central Bay Street, Downtown Toronto that is very popular and they are looking to open 1-2 new location(s) in other neighbourhoods of Downtown Toronto.

1.2 Problem

Due to the complex and varying characteristics of different neighbourhoods in Downtown Toronto, the owner of the coffee shop plans to hire a group of data scientists to determine the best locations for expanding his business with a list of criteria that may play key roles in terms of increasing the success of this business expansion. This list includes: 1. selecting a popular neighbourhood with a high number of venues (i.e. leading to a higher rate of customers visiting), 2. selecting a neighbourhood with lower potential competition (i.e. lower number of existing coffee shops), 3. considering the success of the current location, the owner would like to find a similar neighbourhood to increase the chance of success.

1.3 Interest

The owner of the coffee shop who hired the data scientists would be interested to see the results on the best location(s) for expanding his business and thus would be the main

audience of this project. Other interests may include other business owners or partners who wish to use this model to future business expansions in their own cities.

2. Data:

2.1 Data sources

The trusted data sources that will be used to conduct data exploration and other forms of data analysis as part of the methodology for the project include: Toronto Neighborhood Data from Wikipedia (https://en.wikipedia.org/wiki/List_of_postal_codes_of_Canada:_M); the data table containing the latitude and longitude coordinates of postal codes in Toronto compiled in the csv file that was provided in Week 3 assignment; the Foursquare Location Data (Venues) that will be used to cluster, segment, target, and position to craft recommendations for determining locations for the business expansion.

2.2 Data cleaning and selection

In order to approach this problem, as data scientists, we will first gather the data from reliable sources then clean up and sort the data to contain only relevant information we need to solve the problem. Then, we analyze the data to generate organized maps, tables and charts for us to visualize and draw conclusions based on our results. Finally, we will present our findings and our recommendations on the solution of the problem to our audience - the business owner. The details are summarized in a list below.

2.2.1 Start a Python notebook. Build the code to scrape the Wikipedia page of postal codes of Toronto (https://en.wikipedia.org/wiki/List_of_postal_codes_of_Canada:_M) and transform the data into a pandas dataframe containing all the postal codes matching the list of boroughs and neighbourhoods in Toronto. Clean up the dataframe by removing redundancies and irrelevant cells and reorganize data as needed.

2.2.2 Retrieve the geographical coordinates of postal codes in Toronto using the csv file (provided in Week 3 assignment) into a new dataframe that contains the latitude and longitude coordinates of postal codes and merge this dataframe with the first dataframe containing postal codes matching the boroughs and neighbourhoods in

Toronto to obtain a combined dataframe with the latitude and longitude coordinates of all the boroughs and neighbourhoods in Toronto.

2.2.3 Select Downtown Toronto with its neighbourhood names and latitude and longitude coordinates. Create map of Downtown Toronto to visualize the neighbourhoods using folium.Map

2.2.4 Use Foursquare API to retrieve lists of venues in each neighbourhood of Downtown Toronto. Count the number of venues in each neighbourhoods and select the neighbourhoods with an equal or higher number of venues than Central Bay Street – the current location of the business.

2.2.5 From the list of neighbourhoods selected with the high venue counts, use one-hot-encoding to generate a chart including the counts of each different venue type in each neighbourhood and summarize the finding in a frequency table. Exam the frequency of coffee shops in each neighbourhood and select the neighbourhoods with lower frequencies of coffee shops.

2.2.6 Explore and exam the selected neighbourhoods in Downtown Toronto and use Machine Learning and Statistical Analysis methods to evaluate the similarity between Central Bay Street (current business location) vs. the selected neighbourhoods as potential locations to open up more coffee shops. Use maps or graphs to help with visualization.

2.2.7 Based on results of data analysis and data visualization, draw conclusions and present findings and recommendations to the business owner.

3. Methodology:

3.1 Exploring all neighbourhoods in Downtown Toronto

After retrieving, cleaning up and merging data from the trusted data sources, we have compiled a list of 17 neighbourhoods located in Downtown Toronto sorted by postal codes, which is shown in the table below (Table 1). In addition, Figure 1 below shows a map of Downtown Toronto with its neighbourhoods marked. Some of the neighbourhoods are located closer together while others appear to be more spread out in distance.

Table 1. List of all neighbourhoods (17) in Downtown Toronto

	Postal Code	Borough	Neighbourhood	Latitude	Longitude
0	M4W	Downtown Toronto	Rosedale	43.679563	-79.377529
1	M4X	Downtown Toronto	Cabbagetown,St. James Town	43.667967	-79.367675
2	M4Y	Downtown Toronto	Church and Wellesley	43.665860	-79.383160
3	M5A	Downtown Toronto	Harbourfront,Regent Park	43.654260	-79.360636
4	M5B	Downtown Toronto	Ryerson,Garden District	43.657162	-79.378937
5	M5C	Downtown Toronto	St. James Town	43.651494	-79.375418
6	M5E	Downtown Toronto	Berczy Park	43.644771	-79.373306
7	M5G	Downtown Toronto	Central Bay Street	43.657952	-79.387383
8	M5H	Downtown Toronto	Adelaide,King,Richmond	43.650571	-79.384568
9	M5J	Downtown Toronto	Harbourfront East,Toronto Islands,Union Station	43.640816	-79.381752
10	M5K	Downtown Toronto	Design Exchange,Toronto Dominion Centre	43.647177	-79.381576
11	M5L	Downtown Toronto	Commerce Court,Victoria Hotel	43.648198	-79.379817
12	M5S	Downtown Toronto	Harbord,University of Toronto	43.662696	-79.400049
13	M5T	Downtown Toronto	Chinatown,Grange Park,Kensington Market	43.653206	-79.400049
14	M5V	Downtown Toronto	CN Tower,Bathurst Quay,Island airport,Harbourf...	43.628947	-79.394420
15	M5W	Downtown Toronto	Stn A PO Boxes 25 The Esplanade	43.646435	-79.374846
16	M5X	Downtown Toronto	First Canadian Place,Underground city	43.648429	-79.382280
17	M6G	Downtown Toronto	Christie	43.669542	-79.422564

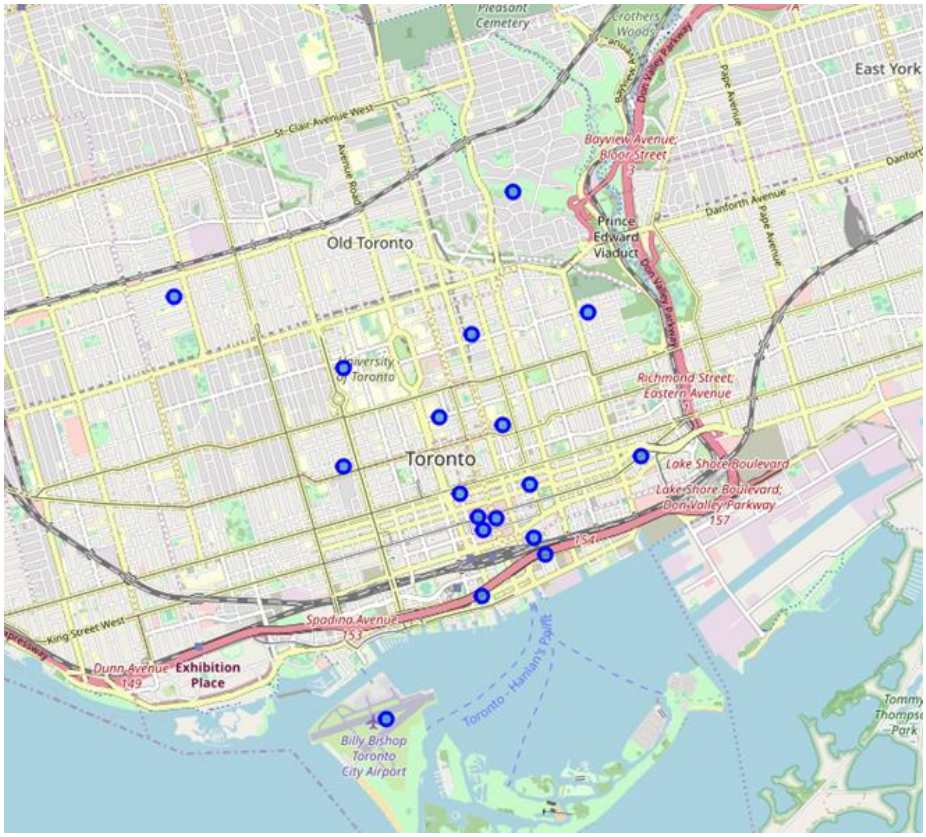

Figure 1. Map of all 17 neighbourhoods in Downtown Toronto

3.2 Number of venues occurring in each neighbourhood

We counted the number of venues occurring in each neighbourhood of Downtown Toronto and compiled the results in descending order in the table below (Table 2). From this table, we can see that there are 11 out of 17 neighbourhoods in Downtown Toronto that has an equal or higher number of venues than Central Bay Street where the original coffee shop is located, which has a venue count of 87. Next, we select these 11 neighbourhoods including Central Bay Street for further analysis.

	Neighbourhood	Venue
0	Adelaide,King,Richmond	100
8	Commerce Court,Victoria Hotel	100
16	St. James Town	100
15	Ryerson,Garden District	100
12	Harbourfront East,Toronto Islands,Union Station	100
10	First Canadian Place,Underground city	100
9	Design Exchange,Toronto Dominion Centre	100
5	Chinatown,Grange Park,Kensington Market	100
17	Stn A PO Boxes 25 The Esplanade	94
7	Church and Wellesley	88
4	Central Bay Street	87
1	Berczy Park	55
13	Harbourfront,Regent Park	48
3	Cabbagetown,St. James Town	44
11	Harbord,University of Toronto	34
6	Christie	16
2	CN Tower,Bathurst Quay,Island airport,Harbourf...	14
14	Rosedale	4

Table 2. Venue count in each neighbourhood of Downtown Toronto

3.3 Frequency of coffee shops in each neighbourhood

We used the ‘one hot encoding’ method to retrieve the different types of venues occurring in the selected neighbourhoods and then we calculated the frequency of each type of venue in each neighbourhood. Since we are interested mainly in coffee shops, we selected the frequency data specifically for coffee shops and compiled the values into Table 3 in

descending order. For better visualization, we also created a bar graph (Figure 2) showing the frequency of coffee shops in each of the selected neighbourhoods in Downtown Toronto.

Table 3. Frequency of coffee shops in each neighbourhood

	Neighbourhood	Coffee Shop
0	Chinatown,Grange Park,Kensington Market	0.040000
1	Adelaide,King,Richmond	0.060000
2	Church and Wellesley	0.068182
3	St. James Town	0.070000
4	First Canadian Place,Underground city	0.090000
5	Ryerson,Garden District	0.100000
6	Stn A PO Boxes 25 The Esplanade	0.106383
7	Commerce Court,Victoria Hotel	0.110000
8	Design Exchange,Toronto Dominion Centre	0.120000
9	Harbourfront East,Toronto Islands,Union Station	0.120000
10	Central Bay Street	0.160920

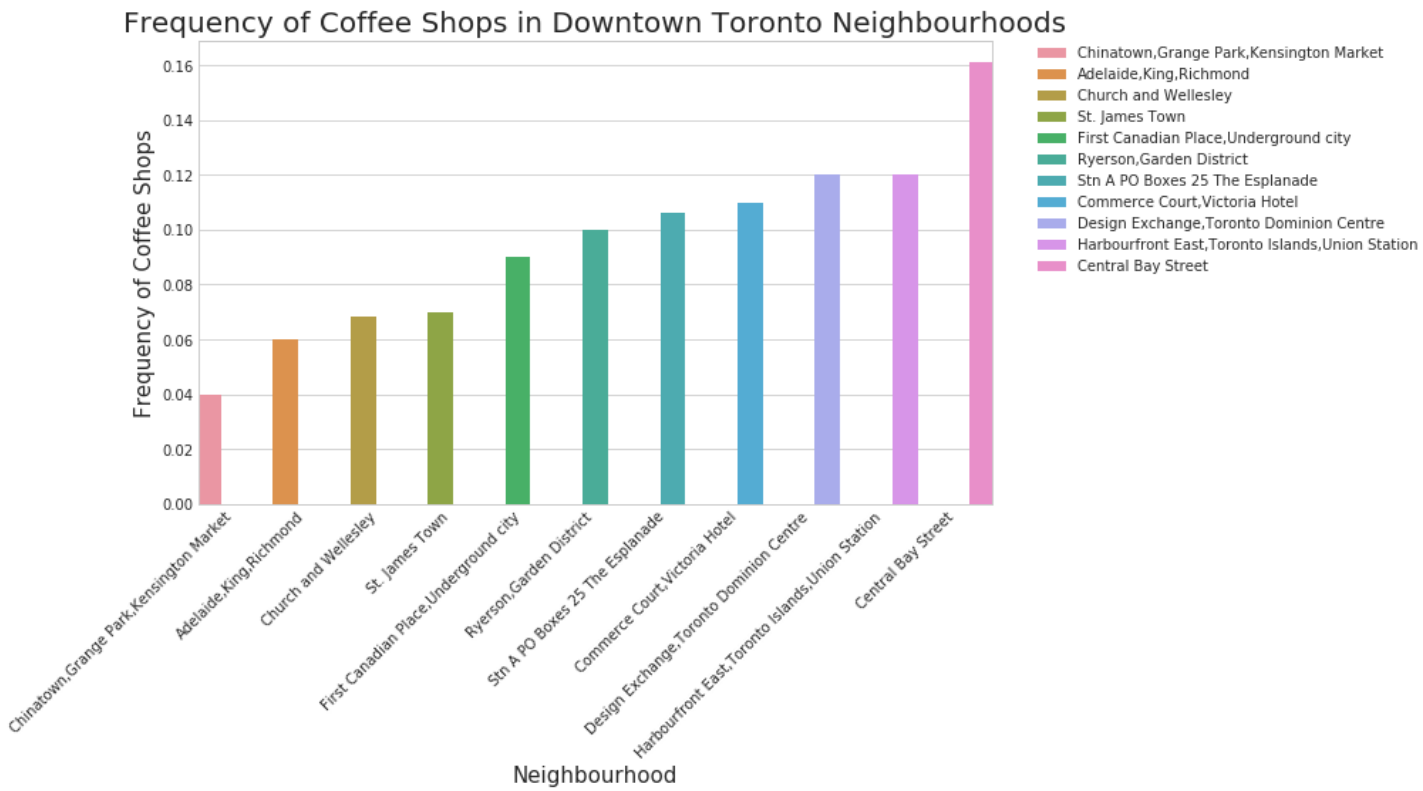

Figure 2. Frequency of coffee shops in each neighbourhood

3.4 Machine learning for similarities between neighbourhoods by k-means clustering

K-means clustering is a machine learning method vastly used for clustering in many data science applications and is especially useful for a fast discovery of insights from unlabeled data. In the real world, we can use this method for many applications such as customer segmentation and pattern recognition. Thus, in this project, we have decided to use this k-means clustering method to cluster the selected neighbourhoods in Downtown Toronto based on common venues in order to discover how similar or different they are from each other. For each neighbourhood, we ranked the venues based on frequency and compiled the ten most common venues into a table (Table 4). Then we performed the k-means clustering with the cluster number set to 2 based on the fact we only have 11 neighbourhoods selected. After merging the cluster labels with geographical coordinates of each neighbhourhood, we created a map (Figure 3) which shows all the neighbourhoods as labeled clusters with different colours. In Table 5, we are taking a closer look at the cluster that contains Central Bay Street to observe the neighbourhoods belong to the same cluster and share similarities with Central Bay Street where the original coffee shop is located.

Table 4. Rank of most common venues occurring in each neighbourhood

	Neighbourhood	1st Most Common Venue	2nd Most Common Venue	3rd Most Common Venue	4th Most Common Venue	5th Most Common Venue	6th Most Common Venue	7th Most Common Venue	8th Most Common Venue	9th Most Common Venue	10th Most Common Venue
0	Adelaide,King,Richmond	Coffee Shop	Café	Thai Restaurant	Bar	Steakhouse	American Restaurant	Burger Joint	Cosmetics Shop	Bakery	Hotel
1	Central Bay Street	Coffee Shop	Café	Italian Restaurant	Burger Joint	Middle Eastern Restaurant	Bar	Sushi Restaurant	Bakery	Chinese Restaurant	Japanese Restaurant
2	Chinatown,Grange Park,Kensington Market	Café	Vegetarian / Vegan Restaurant	Coffee Shop	Bakery	Mexican Restaurant	Bar	Dumpling Restaurant	Vietnamese Restaurant	Chinese Restaurant	Farmers Market
3	Church and Wellesley	Japanese Restaurant	Coffee Shop	Sushi Restaurant	Restaurant	Gay Bar	Gastropub	Men's Store	Mediterranean Restaurant	Hotel	Gym
4	Commerce Court,Victoria Hotel	Coffee Shop	Café	Hotel	Restaurant	American Restaurant	Bakery	Italian Restaurant	Deli / Bodega	Seafood Restaurant	Steakhouse
5	Design Exchange,Toronto Dominion Centre	Coffee Shop	Hotel	Café	Restaurant	Deli / Bodega	Steakhouse	Bakery	Seafood Restaurant	Gastropub	Italian Restaurant
6	First Canadian Place,Underground city	Coffee Shop	Café	Hotel	Bakery	Deli / Bodega	Burger Joint	Restaurant	Seafood Restaurant	Bar	Steakhouse
7	Harbourfront East,Toronto Islands,Union Station	Coffee Shop	Hotel	Aquarium	Italian Restaurant	Café	Fried Chicken Joint	Scenic Lookout	Pizza Place	Bakery	Restaurant
8	Ryerson,Garden District	Coffee Shop	Clothing Store	Cosmetics Shop	Café	Middle Eastern Restaurant	Ramen Restaurant	Sporting Goods Shop	Bubble Tea Shop	Pizza Place	Italian Restaurant
9	St. James Town	Coffee Shop	Café	Hotel	Restaurant	Cosmetics Shop	Bakery	Breakfast Spot	Gastropub	Clothing Store	Cocktail Bar
10	Stn A PO Boxes 25 The Esplanade	Coffee Shop	Restaurant	Café	Italian Restaurant	Cocktail Bar	Hotel	Seafood Restaurant	Beer Bar	Fast Food Restaurant	Bakery

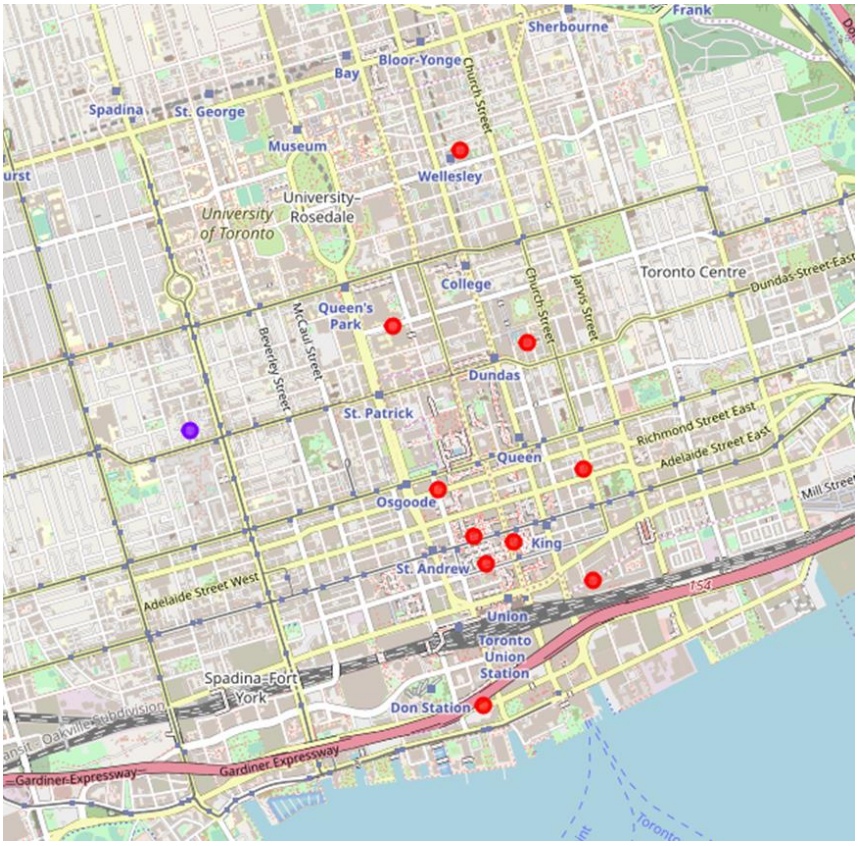

Figure 3. Neighbourhood clusters in Downtown Toronto

Table 5. Exploring cluster containing Central Bay Street

	Neighbourhood	Cluster Labels	1st Most Common Venue	2nd Most Common Venue	3rd Most Common Venue	4th Most Common Venue	5th Most Common Venue	6th Most Common Venue	7th Most Common Venue	8th Most Common Venue	9th Most Common Venue	10th Most Common Venue
0	Church and Wellesley	0	Japanese Restaurant	Coffee Shop	Sushi Restaurant	Restaurant	Gay Bar	Gastropub	Men's Store	Mediterranean Restaurant	Hotel	Gym
1	Ryerson,Garden District	0	Coffee Shop	Clothing Store	Cosmetics Shop	Café	Middle Eastern Restaurant	Ramen Restaurant	Sporting Goods Shop	Bubble Tea Shop	Pizza Place	Italian Restaurant
2	St. James Town	0	Coffee Shop	Café	Hotel	Restaurant	Cosmetics Shop	Bakery	Breakfast Spot	Gastropub	Clothing Store	Cocktail Bar
3	Central Bay Street	0	Coffee Shop	Café	Italian Restaurant	Burger Joint	Middle Eastern Restaurant	Bar	Sushi Restaurant	Bakery	Chinese Restaurant	Japanese Restaurant
4	Adelaide,King,Richmond	0	Coffee Shop	Café	Thai Restaurant	Bar	Steakhouse	American Restaurant	Burger Joint	Cosmetics Shop	Bakery	Hotel
5	Harbourfront East,Toronto Islands,Union Station	0	Coffee Shop	Hotel	Aquarium	Italian Restaurant	Café	Fried Chicken Joint	Scenic Lookout	Pizza Place	Bakery	Restaurant
6	Design Exchange,Toronto Dominion Centre	0	Coffee Shop	Hotel	Café	Restaurant	Deli / Bodega	Steakhouse	Bakery	Seafood Restaurant	Gastropub	Italian Restaurant
7	Commerce Court,Victoria Hotel	0	Coffee Shop	Café	Hotel	Restaurant	American Restaurant	Bakery	Italian Restaurant	Deli / Bodega	Seafood Restaurant	Steakhouse
9	Stn A PO Boxes 25 The Esplanade	0	Coffee Shop	Restaurant	Café	Italian Restaurant	Cocktail Bar	Hotel	Seafood Restaurant	Beer Bar	Fast Food Restaurant	Bakery
10	First Canadian Place,Underground city	0	Coffee Shop	Café	Hotel	Bakery	Deli / Bodega	Burger Joint	Restaurant	Seafood Restaurant	Bar	Steakhouse

3.5 Statistical analysis of neighbourhoods by Pearson correlation coefficient

Pearson correlation coefficient is a measure of the linear correlation between two variables in statistics, so we have decided to use this as one of the methods to evaluate the similarities between Central Bay Street and each of the selected neighbourhoods in Downtown Toronto based on the frequencies of different types of venues occurring in each neighbourhood. We have calculated Pearson correlation coefficients and p-values for each neighbourhood compared to Central Bay Street and listed them in Table 6. Figure 4 shows a bar graph representing the similarity between Central Bay Street and each of the other selected neighbourhoods in Downtown Toronto based on the calculated Pearson correlation coefficients.

Table 6. Pearson correlation coefficients and p-values for each neighbourhood vs. Central Bay Street.

	Neighbourhood	Coffee Shop	Pearson Correlation Coefficient	P-value
0	Chinatown,Grange Park,Kensington Market	0.040000	0.414713	7.124575e-09
1	Adelaide,King,Richmond	0.060000	0.571578	5.304815e-17
2	Church and Wellesley	0.068182	0.512197	1.990909e-13
3	St. James Town	0.070000	0.549027	1.464666e-15
4	First Canadian Place,Underground city	0.090000	0.682063	5.596237e-26
5	Ryerson,Garden District	0.100000	0.726425	8.293996e-31
6	Stn A PO Boxes 25 The Esplanade	0.106383	0.649456	6.105092e-23
7	Commerce Court,Victoria Hotel	0.110000	0.715710	1.477158e-29
8	Design Exchange,Toronto Dominion Centre	0.120000	0.714348	2.110117e-29
9	Harbourfront East,Toronto Islands,Union Station	0.120000	0.696131	2.050364e-27
10	Central Bay Street	0.160920	1.000000	0.000000e+00

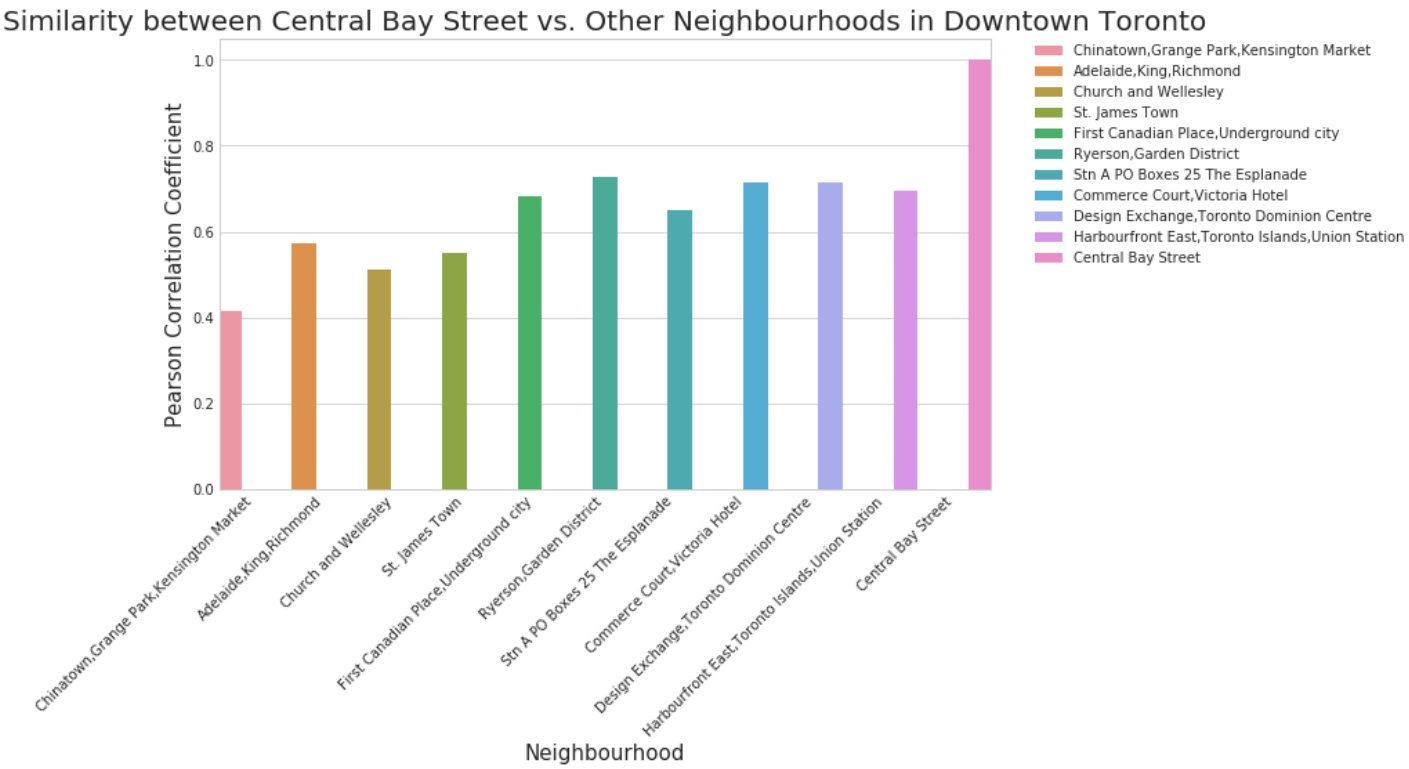

Figure 4. Comparison of Central Bay Street vs. other neighbourhoods based on Pearson correlation coefficient

4. Results and Discussion:

4.1 List of neighbourhoods in Downtown Toronto

After retrieving and cleaning up data from the reliable sources, we found that there are 17 neighbourhoods based on postal codes in the borough of Downtown Toronto, with varying distances in between each other as shown in Table 1 and Figure 1.

4.2 Selected neighbourhoods with top venue counts

Using Foursquare API, we retrieved a list of venues in each of the 17 neighbourhoods in Downtown Toronto. Knowing that the owner of the coffee shop wants to select locations that are popular with a high number of venues comparable to the current location on Central Bay Street that has a venue count of 87, we have selected 11 neighbourhoods from the list of 17 neighbourhoods in Downtown Toronto where the number of venues is equal or greater than 87 (including Central Bay Street, as shown in Table 2).

4.3 Ranking of neighbourhoods based on selected venue frequency

After retrieving the detailed information on venues using one-hot-encoding, we calculated the frequency of venues in each neighbourhood of Downtown Toronto. Then we took a closer look at the frequency of coffee shops and ranked the neighbourhoods in a descending order from the lowest to highest frequency of coffee shops (Table 3, Figure 2). This is very useful to us because we can then select the neighbourhoods that have lower frequencies of existing coffee shops: Chinatown/Grange Park/Kensington Market (0.04), Adelaide/King/Richmond (0.06), Church and Wellesley (0.07), St. James Town (0.07), and First Canadian Place/Underground City (0.09). By selecting these neighbourhoods with less existing coffee shops, we are hoping to expand the business in a space with less competition.

4.4 Neighbourhood clusters by k-means method

Using k-means clustering method, we have categorized the 11 selected neighbourhoods of Downtown Toronto into 2 different clusters, based on the type of venues and how common they are in each neighbourhood. We found 10 out of the 11 selected neighbourhoods fell into the same cluster (including Central Bay Street) and the left-over 1 neighbourhood was identified as its own separate cluster (Figure 3). This is an interesting and helpful observation for us because we now know that we can now filter out the single neighbourhood in the other cluster and focus on the rest of the neighbourhoods that are in the same cluster as Central Bay Street because they all share similar characteristics. All of the 5 neighbouhoods with lower frequencies of coffee shop that have been selected in the previous section (4.3) are included in the same cluster as Central Bay Street.

4.5 Comparison between Central Bay Street and other neighbourhoods using Pearson correlation coefficient

We performed a statistical analysis to evaluate the similarity between Central Bay Street and each of the other selected neighbourhoods by calculating the Pearson correlation coefficient and p-values (Table 6). As shown in Figure 4, from the 5 neighbourhoods with lower frequencies of coffee shops, we can select the 2 neighbourhoods with the highest values of Pearson correlation coefficients: Adelaide/King/Richmond (0.57) and First Canadian Place/Underground City (0.68) because this means that they are more similar to Central Bay Street based on this statistical test.

5. Conclusion:

Based on business problem, our data science team has conducted a detailed data analysis of the neighbourhoods in Downtown Toronto. From our results, we conclude that the best locations based on the coffee shop owner's criteria for expanding his business would be Adelaide/King/Richmond and First Canadian Place/Underground City. Therefore, we recommend the business owner to open new locations in these two neighbourhoods of Downtown Toronto. However, other locations can also be considered if the business owner decides to increase the scale of expansion or modify the criteria for expansion.