Where should the new coffee shop open?

1. Introduction

    1.1 Background
    To identify are there any opportunities to open a new coffee shop in the neighborhood in Toronto City, setting the central point as "City Hall", the center of the city. Based on the JSON data file extraction, there are 50 coffee shops based on the API search. There are a total of 10 boroughs and 103 neighborhoods from the data. There are multiple neighborhoods under the same postal code and group into one postal code and treat as the same neighborhoods by analyze the selected neighborhoods situation.

    1.2 Problem
    How can people understand the existing situation of coffee shop location distribution and identify the potential neighborhoods with less competition, so that it's a good location to open the new coffee shop?

    1.3 Interest
    Using the Geocoder Python package to obtain the geodata for the postal code grouped neighborhoods. Then merging those data with postal code data get from Wiki. Then create the visualization maps to understand the location of the existing coffee shop. Then using clustering to identify the neighborhoods with less coffee then plotting those existing and potential locations of selected neighborhoods to better understand the surrounding of the suggested location or find any insight from the result.

2. Data acquisition and cleaning

    2.1 Data sources
    Mainly 2 sources:
    1) The Postal code data from Wikipedia by using BeautifulSoup
    2) The coffee shop data from Foursquare via API

    2.2 Data cleaning
    1) As there are 10 boroughs and 103 neighborhoods, however, some borough where are null, dropped those records with the borough. There are multiple neighborhoods are sharing the same postal code, grouping those neighborhoods into one combined neighborhood.
    2) The Foursquare searching geodata by the term 'coffee shop' data, then transform the coffee shop JSON data file into dataframe_filtered. Since the

Postal code form JSON file is 6-character, which Wiki data source with first 3 digits only. Need to update the dataframe_filtered postal code to the first 3 characters only.

2.3 Data selection
The key data required from the two data sources are the Postal code, grouped neighborhoods, latitude, longitude, Coffee Shop (name, neighborhoods, the latitude, and the longitude)

3. Exploratory Data Analysis
    3.1 As-is coffee shop location based on the dataframe_filtered based on the transformed coffee shop file. Red dot as the City Hall (as Central point for reference)

Out[28]:
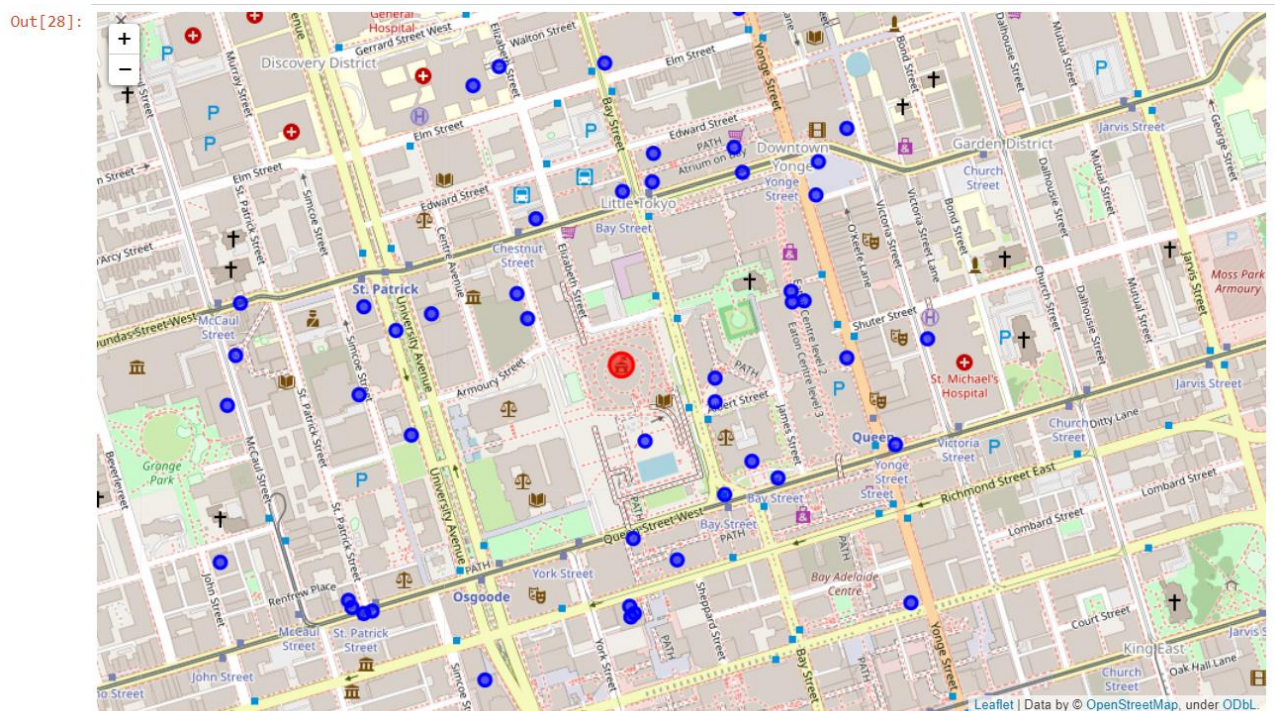
Also listed down the name of those coffee shops, seems Timothy's World Coffee brand is a chain store with 5 shops.

```
In [26]:  ### Check are there any chain coffee shop
          dataframe_filtered.name.value_counts()

Out[26]:  Timothy's World Coffee                 5
          Love Shop                              1
          Smokes'N Coffee                        1
          Coffee Court                           1
          HotBlack Coffee                        1
          Square Shop!                           1
          F5 Shop                                1
          Starbucks                              1
          Future Shop ASW                        1
          Gift Shop At St Michael's Hospital     1
          Seattle's Best Coffee                  1
          The Hippie Grow Shop                   1
          Urbana Coffee Co.                      1
          Coffee Antidote 9th Floor              1
          BRIKA Popup Shop at The Hudson Bay     1
          Coffee office                          1
          Onyx Barber Shop                       1
          Fresh Start Coffee Company             1
          Sushi Shop                             1
          Bulldog Coffee                         1
```

Using the Bar chart to visualize the total number of coffee shops by grouped neighborhoods. (Rotated as below)
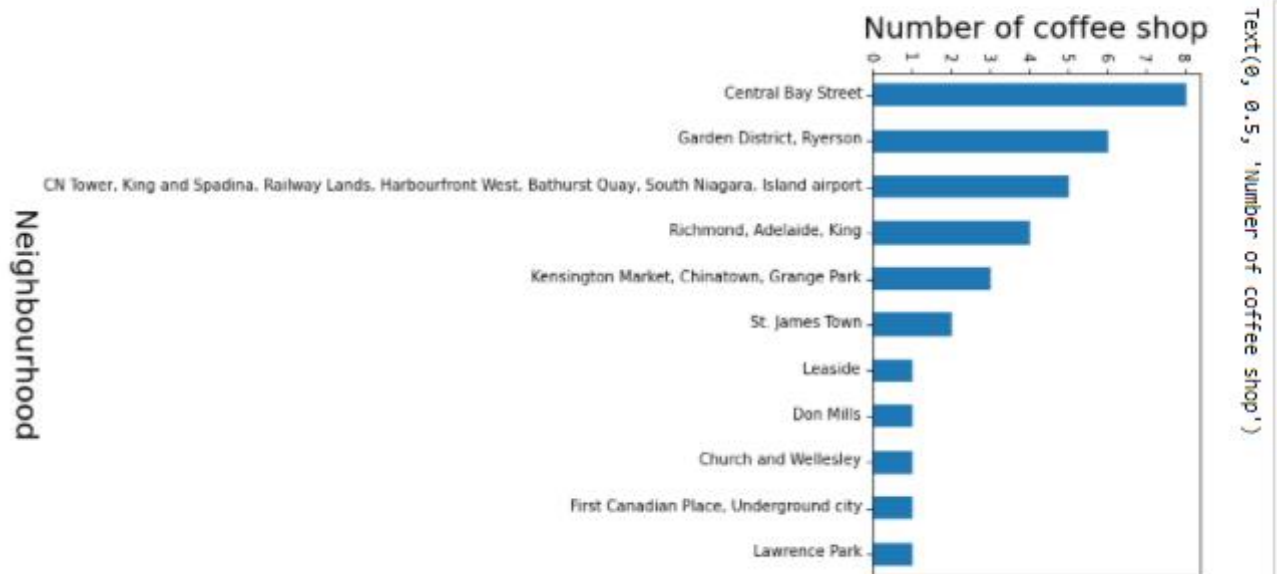
3.2 To-be target coffee shop possible location.

3.2.1 Clustering neighborhood to identify the potential neighborhood that still can accommodate a new coffee shop. Try to group neighborhoods into 6 clusters. Then showing the neighborhoods' clusters on the map and also examining different clusters. To looks for neighborhoods' clusters with the relatively fewer coffee shop, then deep diving the neighborhoods on the map.

The 6 neighborhoods' clusters as below:
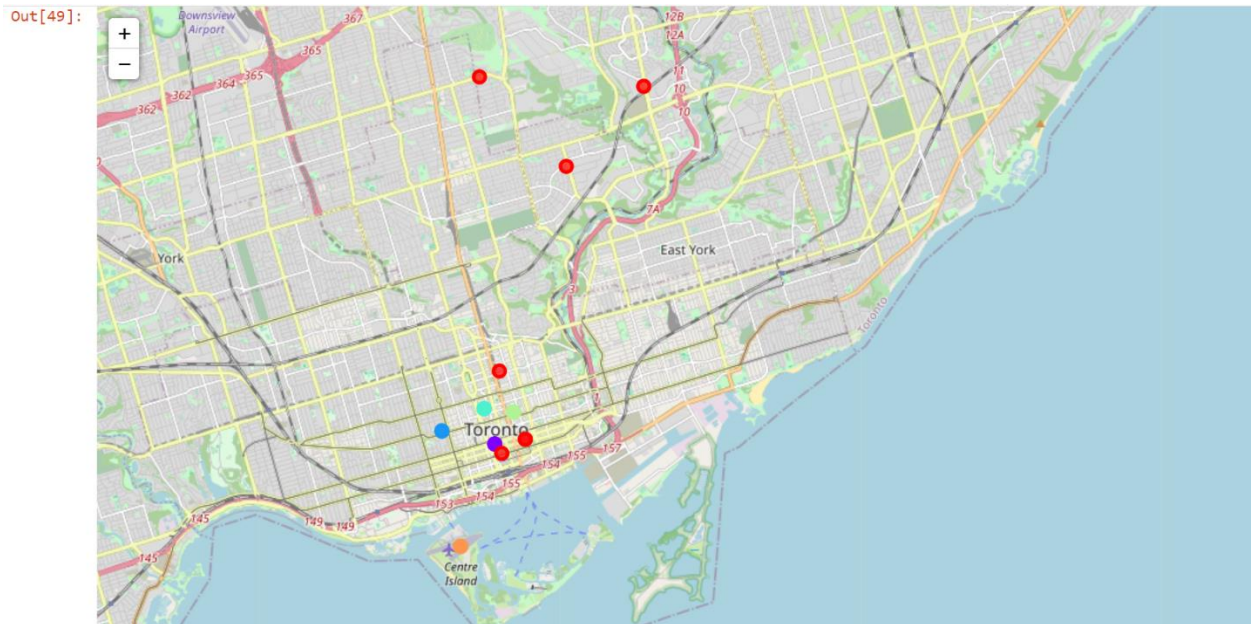
Examining the 6 neighborhoods' clusters:

Cluster 1 – with 8 coffee shop already

```
In [50]: ### Cluster 1
         neighbourhood_coffee_data.loc[neighbourhood_coffee_data['label'] == 0]
```

Out[50]:

	Postal Code	Borough	Neighbourhood	Latitude	Longitude	label	M3B	M4G	M4N	M4Y	M5B	M5C	M5G	M5H	M5T	M5V	M5X
0	M3B	North York	Don Mills	43.745906	-79.352188	0	1	0	0	0	0	0	0	0	0	0	0
1	M3C	North York	Don Mills	43.725900	-79.340923	0	1	0	0	0	0	0	0	0	0	0	0
2	M4G	East York	Leaside	43.709060	-79.363452	0	0	1	0	0	0	0	0	0	0	0	0
3	M4N	Central Toronto	Lawrence Park	43.728020	-79.388790	0	0	0	1	0	0	0	0	0	0	0	0
4	M4Y	Downtown Toronto	Church and Wellesley	43.665860	-79.383160	0	0	0	0	1	0	0	0	0	0	0	0
11	M5C	Downtown Toronto	St. James Town	43.651494	-79.375418	0	0	0	0	0	0	1	0	0	0	0	0
12	M5C	Downtown Toronto	St. James Town	43.651494	-79.375418	0	0	0	0	0	0	1	0	0	0	0	0
33	M5X	Downtown Toronto	First Canadian Place, Underground city	43.648429	-79.382280	0	0	0	0	0	0	0	0	0	0	0	1

Cluster 2 – With 4 coffee shop only

```
In [51]:   ### Cluster 2
           neighbourhood_coffee_data.loc[neighbourhood_coffee_data['label'] == 1]
```

Out[51]:

	Postal Code	Borough	Neighbourhood	Latitude	Longitude	label	M3B	M4G	M4N	M4Y	M5B	M5C	M5G	M5H	M5T	M5V	M5X
21	M5H	Downtown Toronto	Richmond, Adelaide, King	43.650571	-79.384568	1	0	0	0	0	0	0	0	1	0	0	0
22	M5H	Downtown Toronto	Richmond, Adelaide, King	43.650571	-79.384568	1	0	0	0	0	0	0	0	1	0	0	0
23	M5H	Downtown Toronto	Richmond, Adelaide, King	43.650571	-79.384568	1	0	0	0	0	0	0	0	1	0	0	0
24	M5H	Downtown Toronto	Richmond, Adelaide, King	43.650571	-79.384568	1	0	0	0	0	0	0	0	1	0	0	0

Cluster 3 – With 3 coffee shop only

```
In [52]:   ### Cluster 3
           neighbourhood_coffee_data.loc[neighbourhood_coffee_data['label'] == 2]
```

Out[52]:

	Postal Code	Borough	Neighbourhood	Latitude	Longitude	label	M3B	M4G	M4N	M4Y	M5B	M5C	M5G	M5H	M5T	M5V	M5X
25	M5T	Downtown Toronto	Kensington Market, Chinatown, Grange Park	43.653206	-79.400049	2	0	0	0	0	0	0	0	0	1	0	0
26	M5T	Downtown Toronto	Kensington Market, Chinatown, Grange Park	43.653206	-79.400049	2	0	0	0	0	0	0	0	0	1	0	0
27	M5T	Downtown Toronto	Kensington Market, Chinatown, Grange Park	43.653206	-79.400049	2	0	0	0	0	0	0	0	0	1	0	0

Cluster 4 – With 8 coffee shop already

```
In [53]:   ### Cluster 4
           neighbourhood_coffee_data.loc[neighbourhood_coffee_data['label'] == 3]
```

Out[53]:

	Postal Code	Borough	Neighbourhood	Latitude	Longitude	label	M3B	M4G	M4N	M4Y	M5B	M5C	M5G	M5H	M5T	M5V	M5X
13	M5G	Downtown Toronto	Central Bay Street	43.657952	-79.387383	3	0	0	0	0	0	0	1	0	0	0	0
14	M5G	Downtown Toronto	Central Bay Street	43.657952	-79.387383	3	0	0	0	0	0	0	1	0	0	0	0
15	M5G	Downtown Toronto	Central Bay Street	43.657952	-79.387383	3	0	0	0	0	0	0	1	0	0	0	0
16	M5G	Downtown Toronto	Central Bay Street	43.657952	-79.387383	3	0	0	0	0	0	0	1	0	0	0	0
17	M5G	Downtown Toronto	Central Bay Street	43.657952	-79.387383	3	0	0	0	0	0	0	1	0	0	0	0
18	M5G	Downtown Toronto	Central Bay Street	43.657952	-79.387383	3	0	0	0	0	0	0	1	0	0	0	0
19	M5G	Downtown Toronto	Central Bay Street	43.657952	-79.387383	3	0	0	0	0	0	0	1	0	0	0	0
20	M5G	Downtown Toronto	Central Bay Street	43.657952	-79.387383	3	0	0	0	0	0	0	1	0	0	0	0

Cluster 5 – With 6 coffee shop already

```
In [54]:  ### Cluster 5
          neighbourhood_coffee_data.loc[neighbourhood_coffee_data['label'] == 4]
```

Out[54]:

	Postal Code	Borough	Neighbourhood	Latitude	Longitude	label	M3B	M4G	M4N	M4Y	M5B	M5C	M5G	M5H	M5T	M5V	M5X
5	M5B	Downtown Toronto	Garden District, Ryerson	43.657162	-79.378937	4	0	0	0	0	1	0	0	0	0	0	0
6	M5B	Downtown Toronto	Garden District, Ryerson	43.657162	-79.378937	4	0	0	0	0	1	0	0	0	0	0	0
7	M5B	Downtown Toronto	Garden District, Ryerson	43.657162	-79.378937	4	0	0	0	0	1	0	0	0	0	0	0
8	M5B	Downtown Toronto	Garden District, Ryerson	43.657162	-79.378937	4	0	0	0	0	1	0	0	0	0	0	0
9	M5B	Downtown Toronto	Garden District, Ryerson	43.657162	-79.378937	4	0	0	0	0	1	0	0	0	0	0	0
10	M5B	Downtown Toronto	Garden District, Ryerson	43.657162	-79.378937	4	0	0	0	0	1	0	0	0	0	0	0

Cluster 6 – With 5 coffee shop already

```
In [55]:  neighbourhood_coffee_data.loc[neighbourhood_coffee_data['label'] == 5]
```

Out[55]:

	Postal Code	Borough	Neighbourhood	Latitude	Longitude	label	M3B	M4G	M4N	M4Y	M5B	M5C	M5G	M5H	M5T	M5V	M5X
28	M5V	Downtown Toronto	CN Tower, King and Spadina, Railway Lands, Har...	43.628947	-79.39442	5	0	0	0	0	0	0	0	0	0	1	0
29	M5V	Downtown Toronto	CN Tower, King and Spadina, Railway Lands, Har...	43.628947	-79.39442	5	0	0	0	0	0	0	0	0	0	1	0
30	M5V	Downtown Toronto	CN Tower, King and Spadina, Railway Lands, Har...	43.628947	-79.39442	5	0	0	0	0	0	0	0	0	0	1	0
31	M5V	Downtown Toronto	CN Tower, King and Spadina, Railway Lands, Har...	43.628947	-79.39442	5	0	0	0	0	0	0	0	0	0	1	0
32	M5V	Downtown Toronto	CN Tower, King and Spadina, Railway Lands, Har...	43.628947	-79.39442	5	0	0	0	0	0	0	0	0	0	1	0

4. Clustering the Neighborhoods

   4.1 Group neighborhoods into 6 clusters
      Refer to 3.2 for details.

   4.2 Examining Clusters results
      Refer to 3.2 for details.

   4.3 Identify the potential clusters.
      Based on the identified clusters 2, 3 (neighborhoods – 'Richmond, Adelaide, King', 'Kensington Market, Chinatown, Grange Park') create the new data frame new_shop capture the corresponding Latitude, Longitude. Then based on the cluster 2, 3 (neighborhoods – 'Richmond, Adelaide, King', 'Kensington Market, Chinatown, Grange Park') existing coffee shop name, corresponding Latitude, Longitude from the merged_coffee_data source from JSON data file.

5. Conclusion

   5.1 The target neighborhoods, 'Richmond, Adelaide, King' or 'Kensington Market, Chinatown, Grange Park' both only have 4 and 3 coffee shops.

   Red spots – City Hall, Green spots – existing coffee shops in the neighborhoods and Blue spots – potential location after neighborhood clustering.
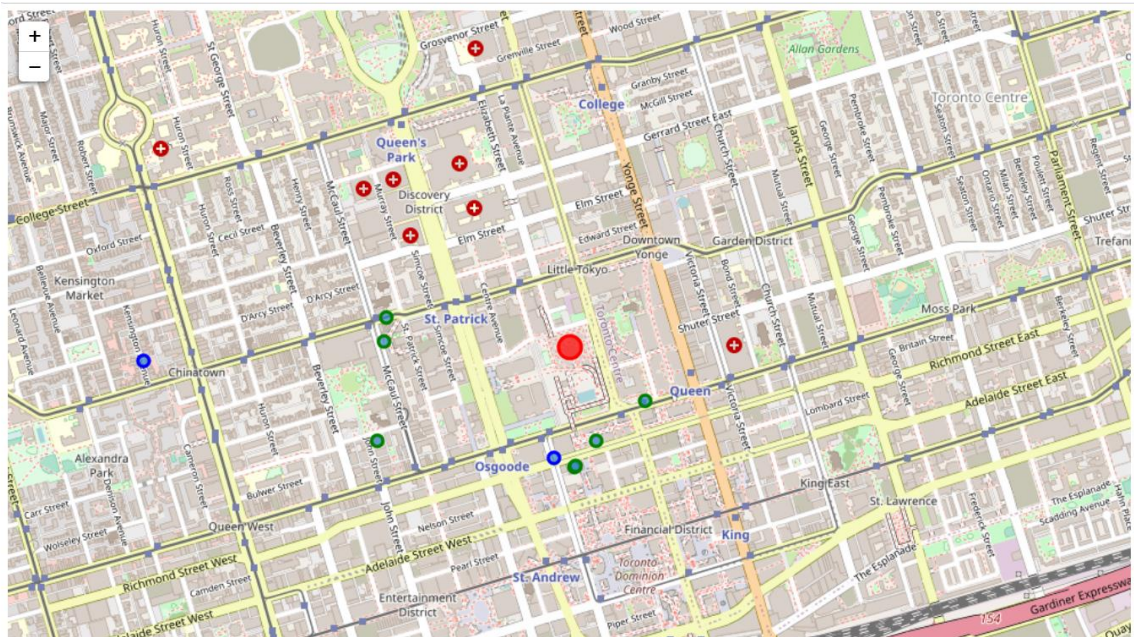

   5.2 Deep diving from the map in 2 potential locations

Option 1 - 'Richmond, Adelaide, King' most close to City Hall. Based on the observation below, there are 2 coffee shops very close, and the other already nearby. However, the potential location is near the Charge point and the Four Seasons Centre for the Performing Arts. There may have room for a new coffee shop after the performing or waiting for charging.

Out[67]:
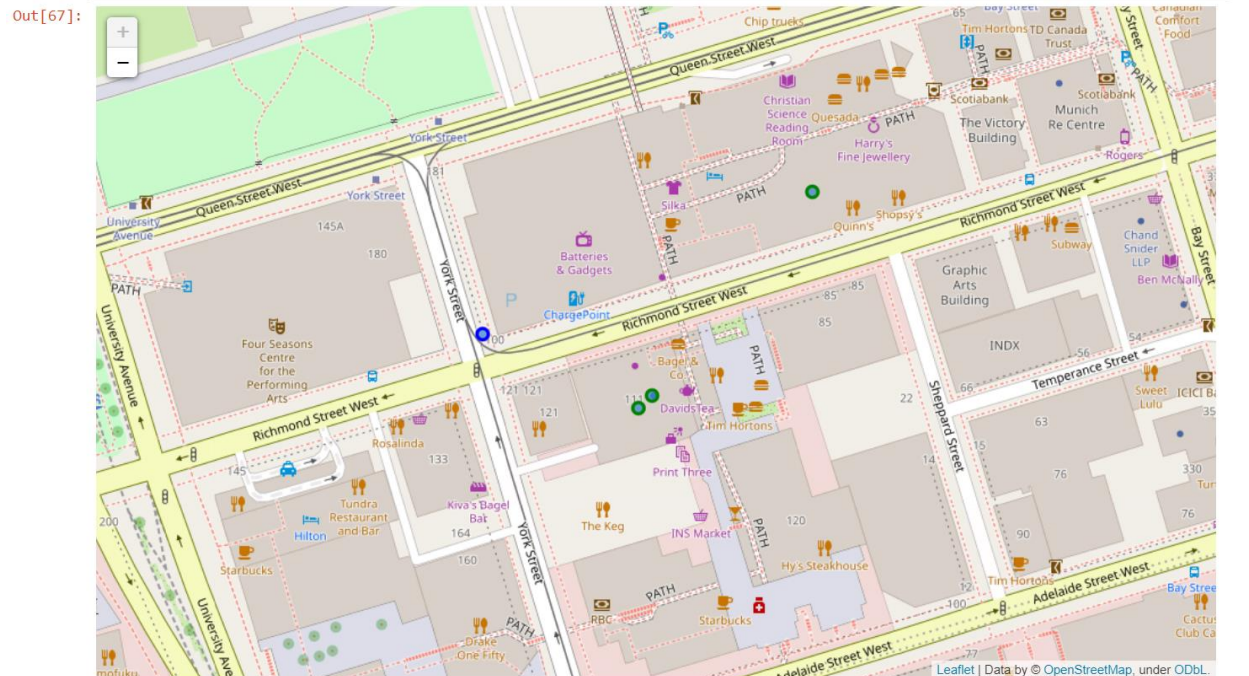
Option 2 - 'Kensington Market, Chinatown, Grange Park' which a bit far from City Hall. Based on the observation below, no coffee shop in the neighborhood, and there are few salons nearby. This may bring the business opportunity to open a new coffee shop in this neighborhood to fulfilling the needs (gaps).
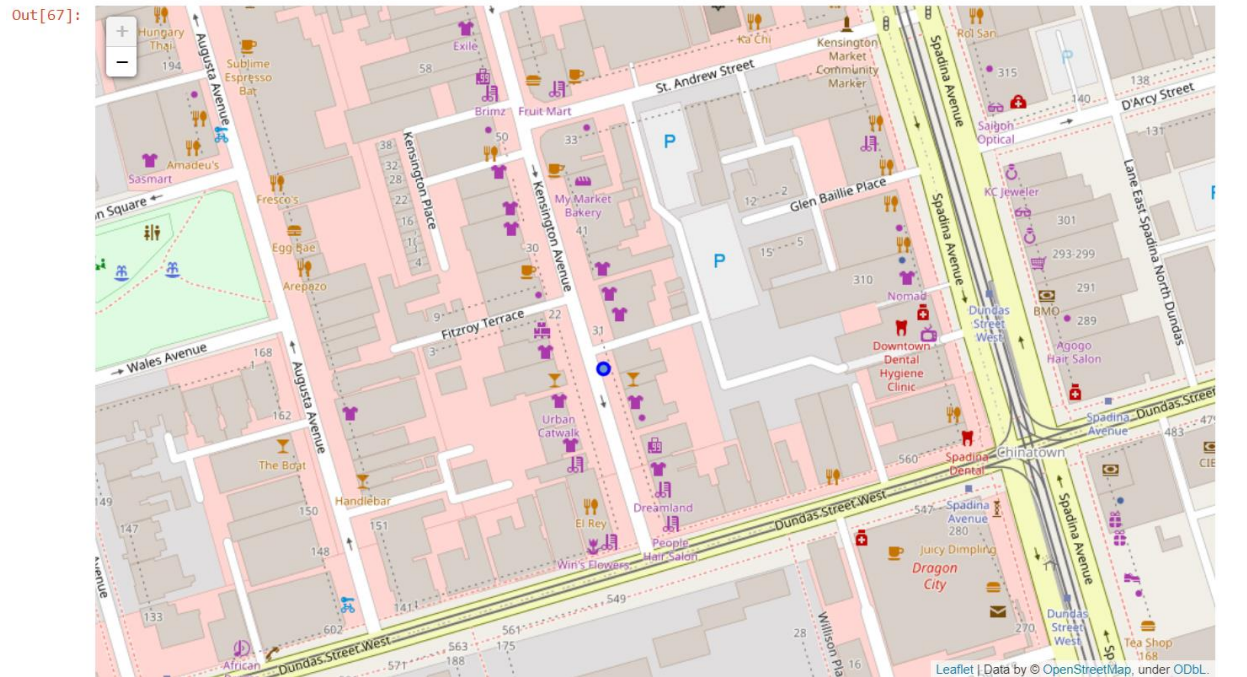
Still, many other criteria also need to take into consideration, this will further discuss and explain in the next sections. Like the consumer preference, as we know China Town may mainly visit with Canadian Chinese, will tea house even more wildly accept that a coffee shop. This needs to further discuss in the future.

6. Limitation

6.1 Data sources accuracy and up-today

Given the extracted list of coffee shop counts only has 1 Starbucks, first, it's not reasonable. Given the map of the close up of the 'Richmond, Adelaide, King', also showing there are Starbucks does not display with the green dot, that means not all the Starbuck shops include from the data source.

```
In [26]: ### Check are there any chain coffee shop
         dataframe_filtered.name.value_counts()

Out[26]: Timothy's World Coffee          5
         Love Shop                       1
         Smokes'N Coffee                 1
         Coffee Court                    1
         HotBlack Coffee                 1
         Square Shop!                    1
         F5 Shop                         1
         Starbucks                       1
```

'Richmond, Adelaide, King' Zoom-in

6.2 Will nearby restaurants also offer coffee, for example, Italian Restaurants in the neighborhood may also offer coffee. Need to think of further obtain the drinks menu in the neighborhood.

6.3 Other criteria also not taking into account yet, post-pandemic, the number of changes about the performing art center audience. The number of users visiting the charging point, the demographic of citizen/visitor in the area. Also, the rental in the area etc.

7. Future directions

7.1 Additional testing – grouping clusters lower than or higher than 6 to compare the locate neighborhood to verify if there are other potential neighborhood identify.

7.2 Include additional data mention in 6.2, 6.3 into account for further analysis before making decision.

7.3 Potential rival location (e.g., Timothy's World Coffee, Starbucks, etc.) should also specific into the map if the data available.