Best Areas to Visit for Busy Coffee-Loving Travelers

In Toronto, Canada

Introduction & Business Problems

- Dexter loves visiting non-chain cafes whenever he travels to a new city which he does frequently due to his consultant work. He also loves taking photos from his visits to share with his family.

- Dexter decides to start his analysis using the data for the city of Toronto since he wants to visit Toronto soon. But once his analysis program works, he believes he can easily use it for other cities around the world he will get to visit in near future.

- Dexter would like to be able to identify an area or two with most cafes and coffee shops that are not available outside of Toronto and that are close to each other that would make it easier for him to visit multiple venues with limited time.
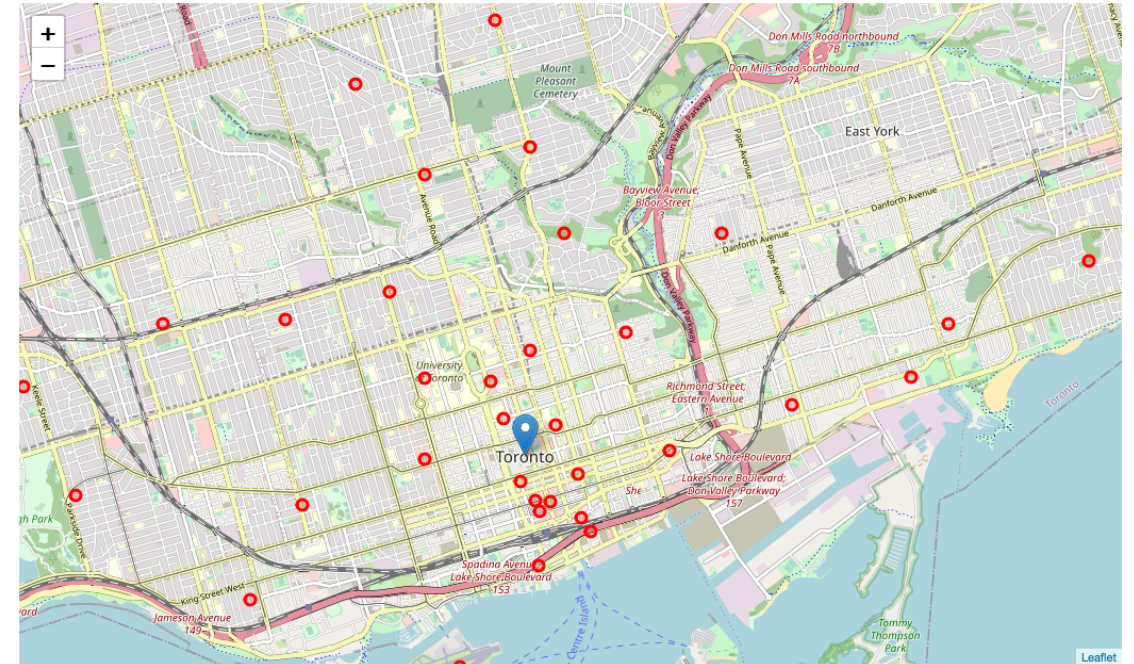
Data Description

- The postal codes for Toronto
  - https://en.wikipedia.org/wiki/List_of_postal_codes_of_Canada:_M
  - Lists postal codes and neighborhood names needed to search venues in Toronto

- Geopspacial data for Canada postal codes
  - http://cocl.us/Geospatial_data
  - CSV format converted to Pandas dataframe
  - Contains data on the postal codes, longitude and latitude information

- Foursquare API for venue information
  - Toronto venue data will be obtained from the Foursquare API
  - Requires latitudes and longitudes of locations of interests
  - Locations of cafes, public transportation stops, and parks and other venues in these neighborhoods to identify best neighborhoods for Dexter to visit

Methodology

- Obtain longitude and latitude data of Toronto neighborhoods to query Foursquare API for venues

- Identify the coffee shops and café venues

- Focus analysis on non-chain coffee venues
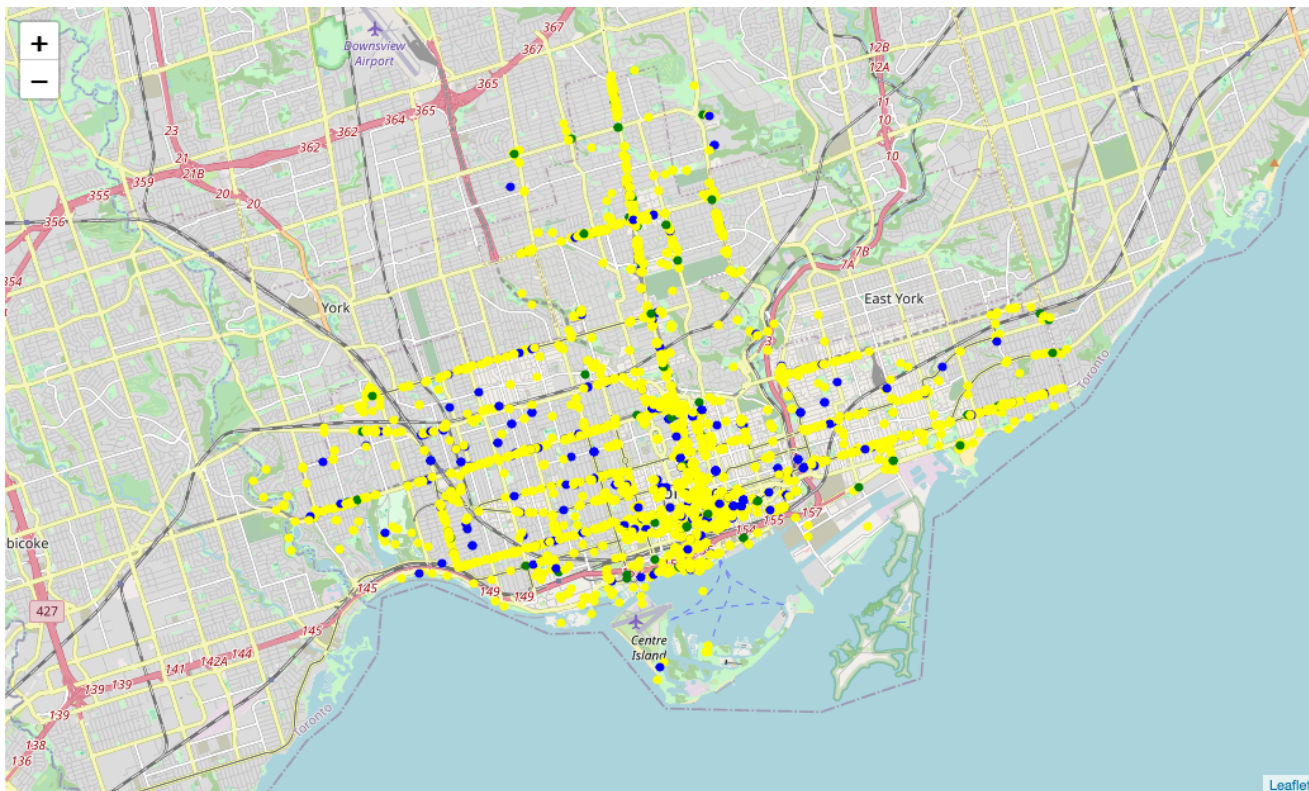
- Apply DBSCAN for density analysis

Methodology

- Merge the geospatial data for postal codes and Toronto neighborhood names to gain the coordinates of the postal codes for Toronto neighborhoods

- Put the locations of coffee venues on the Toronto map.
- Use the Folium library again.
- Visualize different colors for the coffee shops and cafés since these are our main interests at the moment.
  - Blue: non-(international)-chain coffee shops and café venues
  - Green: Chain venues such as 'Starbucks' and 'Tim Hortons'
  - Yellow: all other venues

Methodology

Visualization of Venues of Interest

Analysis

- Identify a couple of concentrations of coffee shops and café venues

```
[49]: toronto_venues.head()
```

```
[49]:
```

	Neighborhood	Neighborhood Latitude	Neighborhood Longitude	Venue	Venue Latitude	Venue Longitude	Venue Category
0	Regent Park, Harbourfront	43.65426	-79.360636	Roselle Desserts	43.653447	-79.362017	Bakery
1	Regent Park, Harbourfront	43.65426	-79.360636	Tandem Coffee	43.653559	-79.361809	Coffee Shop
2	Regent Park, Harbourfront	43.65426	-79.360636	Cooper Koo Family YMCA	43.653249	-79.358008	Distribution Center
3	Regent Park, Harbourfront	43.65426	-79.360636	The Distillery Historic District	43.650244	-79.359323	Historic Site
4	Regent Park, Harbourfront	43.65426	-79.360636	Impact Kitchen	43.656369	-79.356980	Restaurant

```
[50]: toronto_venues.shape
```

```
[50]: (3741, 7)
```

```
coffee_venues = toronto_venues[(toronto_venues['Venue Category'] == 'Coffee Shop') | (toronto_venues['Venue Category'] == 'Café')]
len(coffee_venues)
```

```
498
```

```
coffee_venues_cnt = pd.DataFrame(coffee_venues.groupby('Venue').size().sort_values(ascending=False))
coffee_venues_cnt.head()
```

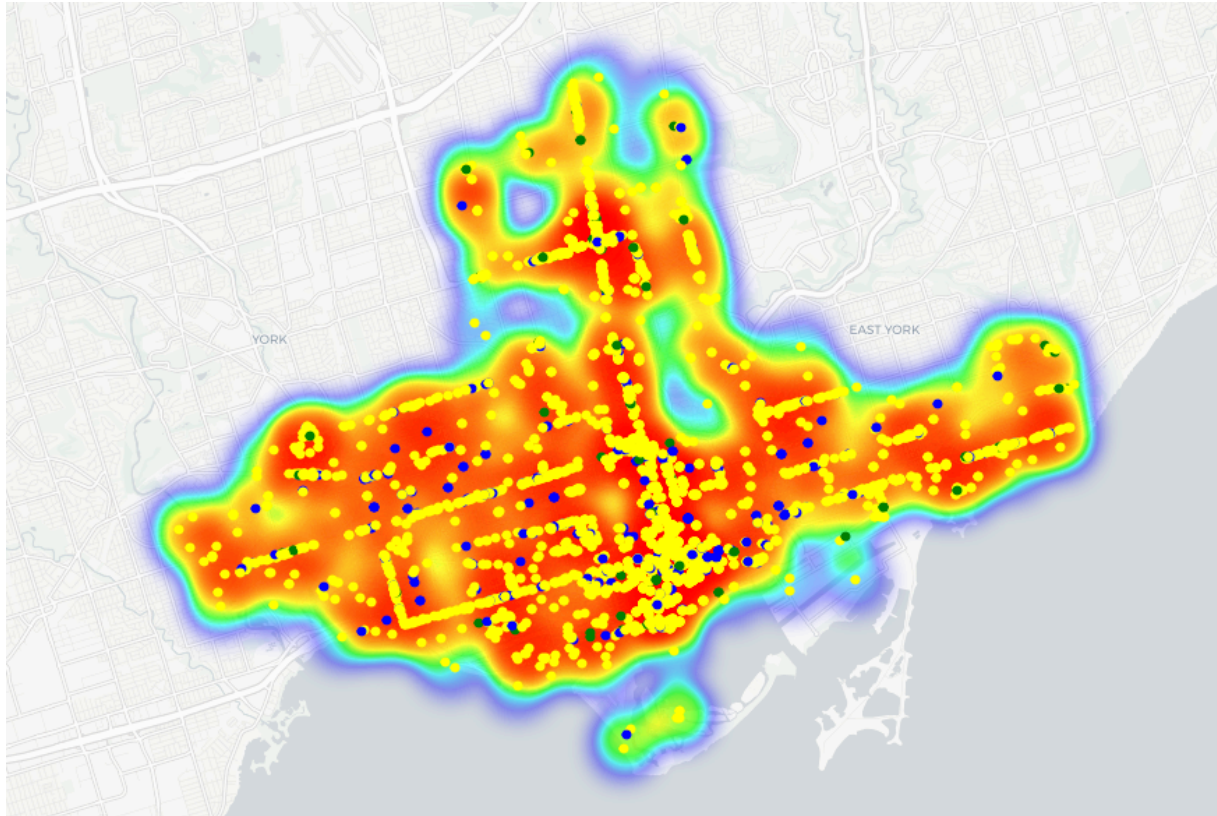|	0
Venue	
Starbucks	79
Tim Hortons	33
Aroma Espresso Bar	18
Jimmy's Coffee	12
Pilot Coffee Roasters	11

```
len(coffee_venues_cnt)
```

```
151
```

Analysis

- Number of venues in Toronto: 3,741
- Number of coffee venues: 498
- Number of unique coffee venues: 151
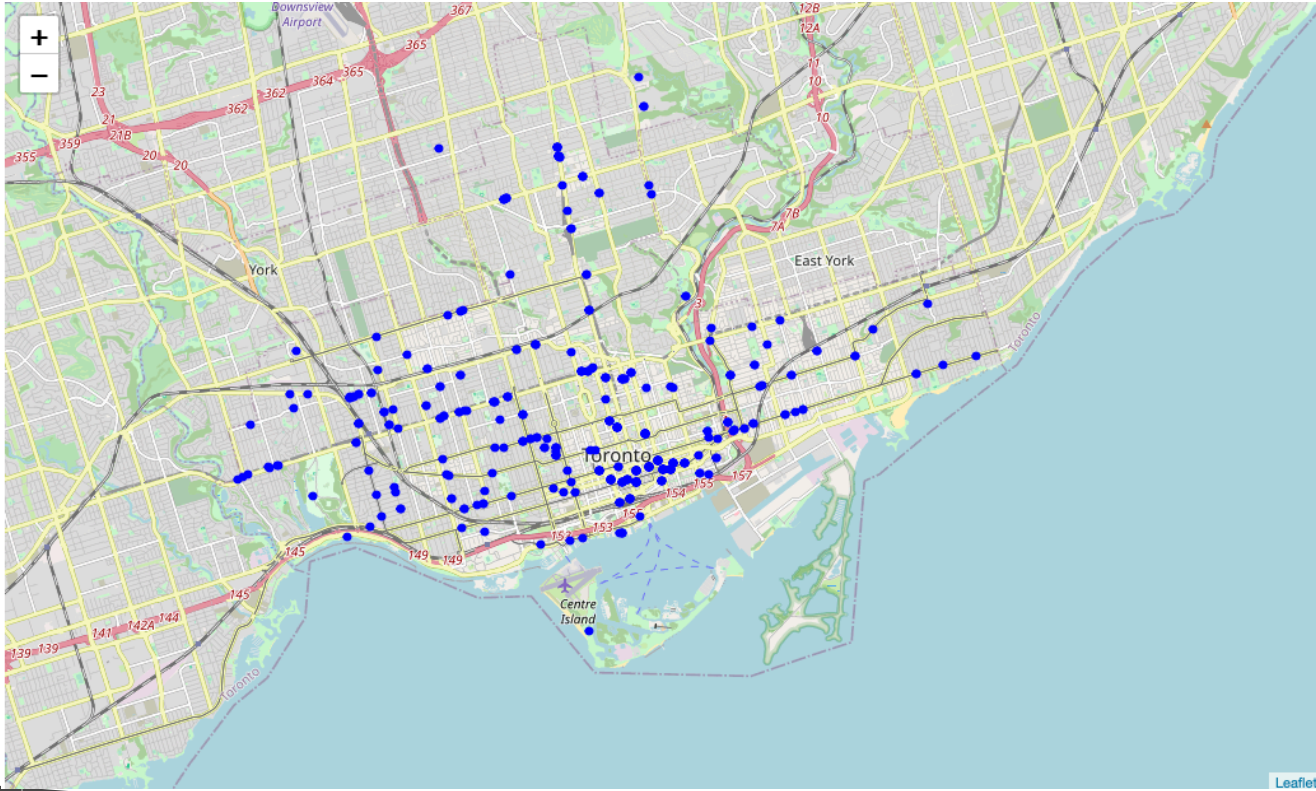
- Analysis about coffee venues out of all venues

- Number of venues in Toronto: 3,741

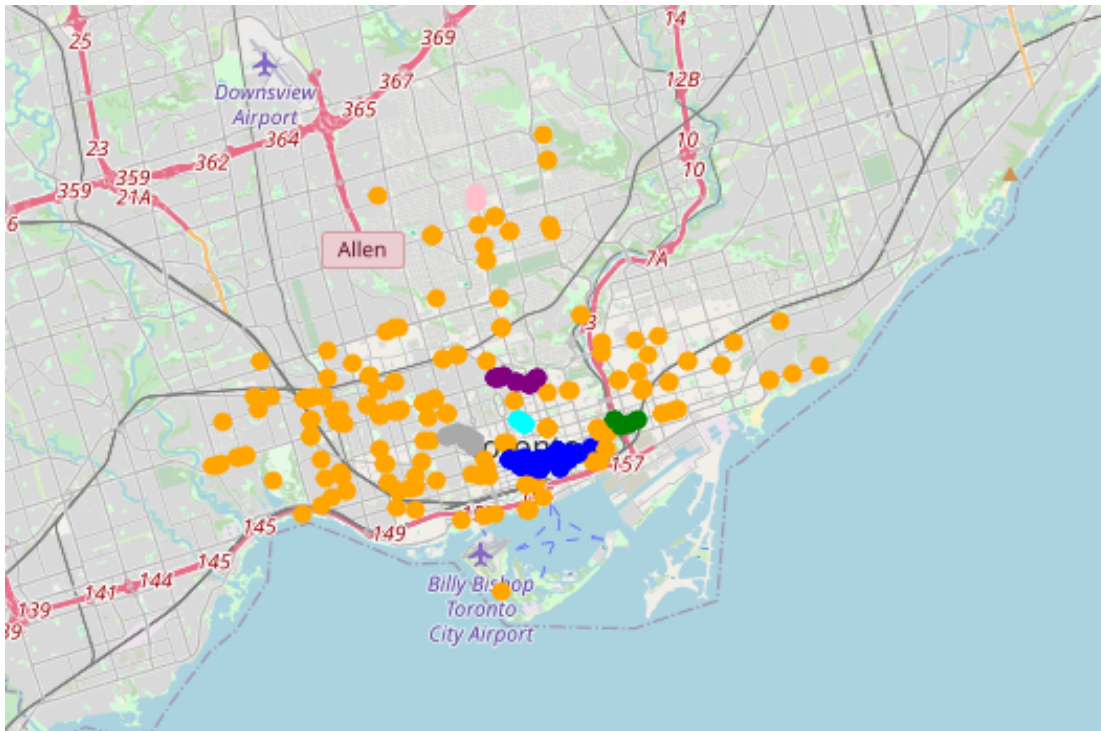- Number of coffee venues: 498

- Number of unique coffee venues: 151

Analysis

- Use of heatmap for distribution of venues

Analysis

- Narrow the analysis to just non-chain, local coffee venues

```
# Compute DBSCAN
db = DBSCAN(eps=0.15, min_samples=10).fit(Clus_dataSet)
core_samples_mask = np.zeros_like(db.labels_, dtype=bool)
core_samples_mask[db.core_sample_indices_] = True
labels = db.labels_
non_chain_coffee_data["Clus_Db"]=labels
```
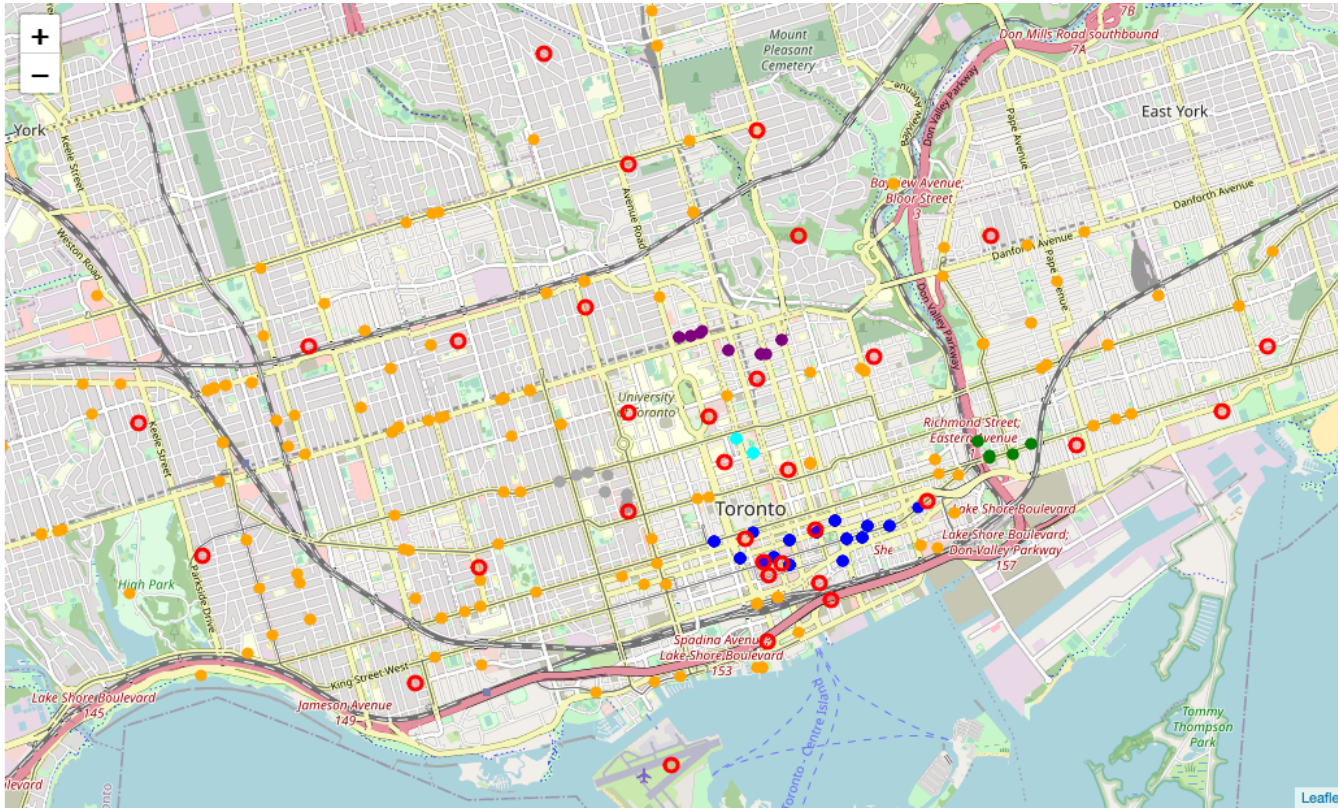
```
non_chain_coffee_data["Clus_Db"][0:20]
```

```
0     0
1    -1
2     0
3    -1
4    -1
5     1
6     0
7     0
8     1
9    -1
10    0
11   -1
12    0
13    1
14   -1
15    0
16    1
17    1
18    2
19    2
Name: Clus_Db, dtype: int64
```

```python
def cluster_color(cluster):
    return {
        -1: 'orange',
        0: 'blue',
        1: 'green',
        2: 'cyan',
        3: 'purple',
        4: 'darkgray',
        5: 'pink'
    }[cluster]
```

Analysis

- Perform DBSCAN clustering
- Visualize the result

- Count the venues in each cluster.

Cluster	Count of Non-chain Coffee Venues
-1	201
0	87
1	10
2	10
3	21
4	17
5	10

- Neighborhoods with coffee venues in Cluster 0

```
Neighborhood
St. James Town                                          13
Stn A PO Boxes                                          12
Berczy Park                                             11
Commerce Court, Victoria Hotel                          10
Toronto Dominion Centre, Design Exchange                 8
Regent Park, Harbourfront                                7
First Canadian Place, Underground city                   7
Richmond, Adelaide, King                                 6
Garden District, Ryerson                                 6
Harbourfront East, Union Station, Toronto Islands        4
Central Bay Street                                       3
```
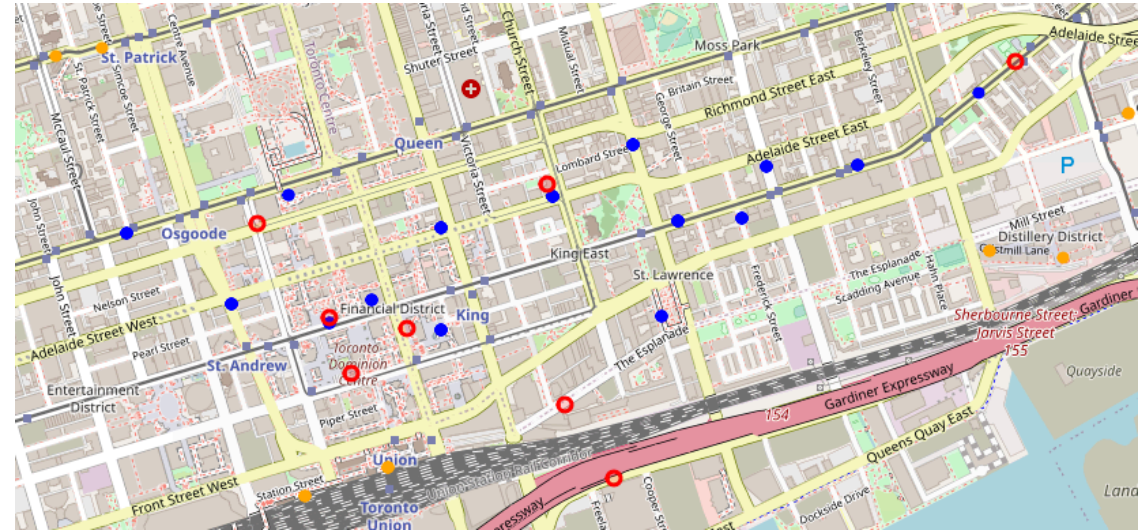
Analysis

- Visualize the clusters in relation to neighborhoods

- Count the venues in clusters

- Identify neighborhoods with coffee venues in cluster group 0

Results

- Top three areas
  - St. James Town
  - Stn A PO Boxes
  - Berczy Park
- Highest congregation of local coffee shops and cafes in proximity

Conclusion

- Our research show that the **St. James Town**, **Stn A PO Boxes** and **Berczy Park** neighborhoods, all in Downtown Toronto, are top three areas for coffee lovers such as Dexter to enjoy real Toronto java scene when only limited amount of time can be afforded to enjoy with high density of coffee venues that are not international chains.

- This method can easily be applied to other cities and for other types of venues as well.