Finding an optimal location for an Italian Restaurant

Prepared by: Mohamed Adel Saber

Introduction:

Business Problem: In this project we will try to find an optimal location for a restaurant. Specifically, this report will be targeted to stakeholders interested in opening an **Italian restaurant** in **Toronto**, Canada.

Since there are lots of restaurants in Toronto we will try to detect**:**

1- Locations that are not already crowded with restaurants.
2- We are also particularly interested in areas with minimum or no Italian restaurants in vicinity.
3- We would also prefer locations as close to Down Town as possible, assuming that first two conditions are met.

We will use our data science powers to generate a few most promising neighborhoods based on this criteria. Advantages of each area will then be clearly expressed so that best possible final location can be chosen by stakeholders.

Data:

Based on definition of our problem, factors that will influence our decision are:

- number of existing restaurants in the neighborhood (any type of restaurant)
- number of and distance to Italian restaurants in the neighborhood, if any
- distance of neighborhood from Down Town

We decided to use regularly spaced grid of locations, centered around Down Town, to define our neighborhoods.

Following data sources will be needed to extract/generate the required information:

- Every Neighborhood name of Toronto, Canada will be obtained using this data set 'https://www.toronto.ca/ext/open_data/catalog/data_set_files/2016_neighbourhood_profiles.csv'.

```
[2]: csv_path='https://www.toronto.ca/ext/open_data/catalog/data_set_files/2016_neighbourhood_profiles.csv'
df = pd.read_csv(csv_path,encoding='latin1')
df.head()
```

[2]:

	Category	Topic	Data Source	Characteristic	City of Toronto	Agincourt North	Agincourt South-Malvern West	Alderwood	Annex	Banbury-Don Mills	...	Willowdale West	Willowridge Martingrove Richvie
0	Neighbourhood Information	Neighbourhood Information	City of Toronto	Neighbourhood Number	NaN	129	128	20	95	42	...	37	
1	Neighbourhood Information	Neighbourhood Information	City of Toronto	TSNS2020 Designation	NaN	No Designation	No Designation	No Designation	No Designation	No Designation	...	No Designation	N Designatic
2	Population	Population and dwellings	Census Profile 98-316-X2016001	Population, 2016	2,731,571	29,113	23,757	12,054	30,526	27,695	...	16,936	22,15
3	Population	Population and dwellings	Census Profile 98-316-X2016001	Population, 2011	2,615,060	30,279	21,988	11,904	29,177	26,918	...	15,004	21,34
4	Population	Population and dwellings	Census Profile 98-316-X2016001	Population Change 2011-2016	4.50%	-3.90%	8.00%	1.30%	4.60%	2.90%	...	12.90%	3.80

5 rows × 145 columns

- Every Neighborhood location(latitude, longitude) will be obtained using (**geopy)**python library.

[3]:

	Neighbourhood	Latitude	Longitude
0	Agincourt North	43.808038	-79.266439
1	Alderwood	43.601717	-79.545232
2	Annex	43.670338	-79.407117
3	Banbury-Don Mills	43.751672	-79.370169
4	Bathurst Manor	43.665519	-79.411937

- To accurately calculate distances we need to create our grid of locations in Cartesian 2D coordinate system which allows us to calculate distances in meters (not in latitude/longitude degrees). Then we'll project those coordinates back to latitude/longitude degrees to be shown on Folium map. So I created functions to convert between WGS84 spherical coordinate system (latitude/longitude degrees) and UTM Cartesian coordinate system (X/Y coordinates in meters).

```
import shapely.geometry
import pyproj
import math

def lonlat_to_xy(lon, lat):
    proj_latlon = pyproj.Proj(proj='latlong',datum='WGS84')
    proj_xy = pyproj.Proj(proj="utm", zone=33, datum='WGS84')
    xy = pyproj.transform(proj_latlon, proj_xy, lon, lat)
    return xy[0], xy[1]

def xy_to_lonlat(x, y):
    proj_latlon = pyproj.Proj(proj='latlong',datum='WGS84')
    proj_xy = pyproj.Proj(proj="utm", zone=33, datum='WGS84')
    lonlat = pyproj.transform(proj_xy, proj_latlon, x, y)
    return lonlat[0], lonlat[1]

def calc_xy_distance(x1, y1, x2, y2):
    dx = x2 - x1
    dy = y2 - y1
    return math.sqrt(dx*dx + dy*dy)
```

- Then I calculated the distance from Downtown to each neighborhood.

[7]:

	Neighbourhood	Latitude	Longitude	X	Y	Distance From Downtown
2	Annex	43.670338	-79.407117	-5.307549e+06	1.050992e+07	3781.433005
4	Bathurst Manor	43.665519	-79.411937	-5.308253e+06	1.051056e+07	3893.566103
5	Bay Street Corridor	43.668865	-79.389126	-5.308014e+06	1.050787e+07	2222.603250
10	Broadview North	43.683924	-79.356964	-5.306028e+06	1.050389e+07	5220.545451
11	Casa Loma	43.678101	-79.409416	-5.306285e+06	1.051005e+07	4807.361366

- I utilized the Foursquare API to explore the boroughs and segment them. I designed the limit as **100 venues** and the radius **500 meter** for each Neighborhood from their given latitude and longitude information. Here is a head of the list Venues name, category, latitude and longitude information from **Foursquare API**.

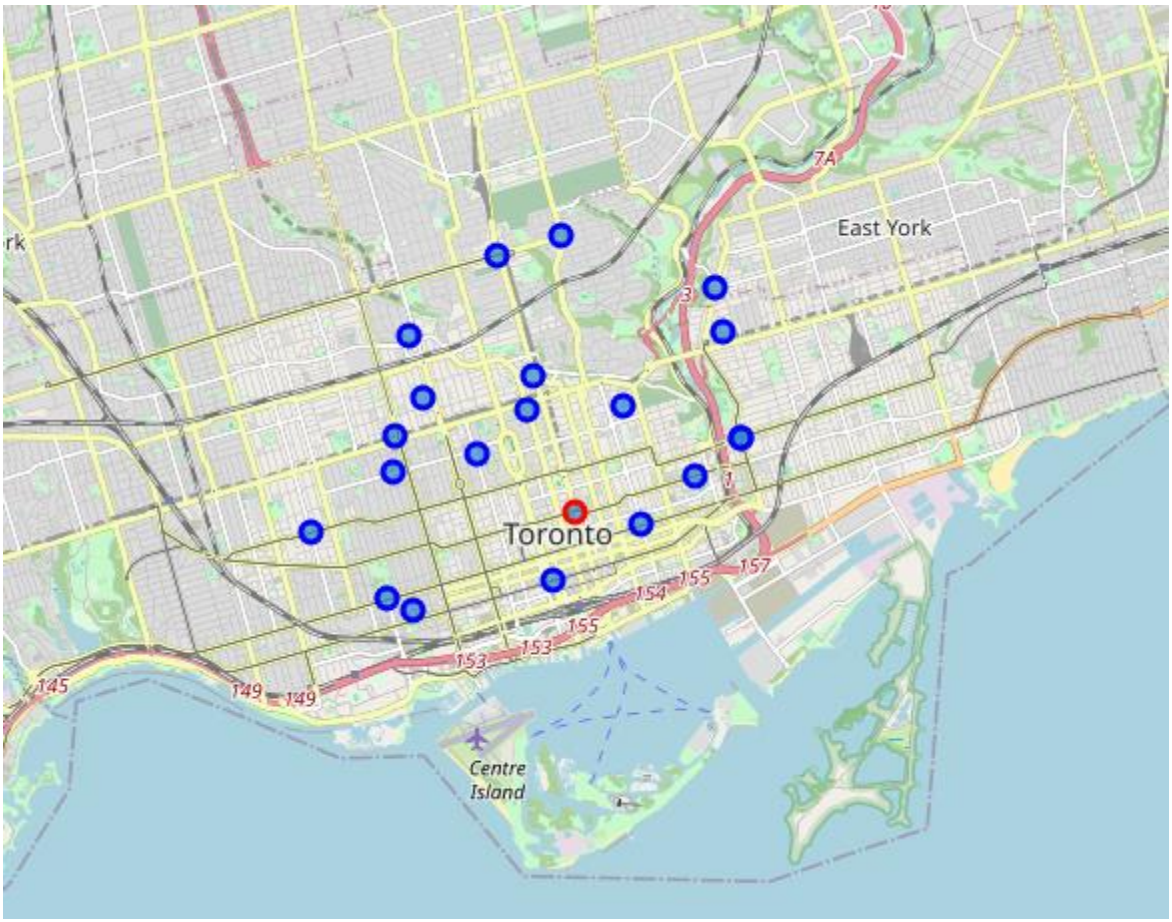
[13]:

	Neighborhood	Neighborhood Latitude	Neighborhood Longitude	Neighborhood X	Neighborhood Y	Neighborhood Distance From Downtown	Venue	Venue Latitude	Venue Longitude	Venue Category
0	Annex	43.670338	-79.407117	-5.307549e+06	1.050992e+07	3781.433005	Jean Sibelius Square	43.671426	-79.408831	Park
1	Annex	43.670338	-79.407117	-5.307549e+06	1.050992e+07	3781.433005	Fresh on Bloor	43.666755	-79.403491	Vegetarian / Vegan Restaurant
2	Annex	43.670338	-79.407117	-5.307549e+06	1.050992e+07	3781.433005	Roti Cuisine of India	43.674618	-79.408249	Indian Restaurant
3	Annex	43.670338	-79.407117	-5.307549e+06	1.050992e+07	3781.433005	Fuwa Fuwa Japanese Pancakes	43.665880	-79.407840	Pastry Shop
4	Annex	43.670338	-79.407117	-5.307549e+06	1.050992e+07	3781.433005	The Original Gyro Grill	43.666621	-79.405544	Greek Restaurant

Methodology:

In this project we will direct our efforts on detecting areas of Toronto that have low restaurant density, particularly those with low number of Italian restaurants. We will limit our analysis to area ~6km around Downtown.

I used python **folium** library to visualize geographic details of Toronto and its Neighborhoods and I created a map of Toronto with Neighborhoods superimposed on top(Downtown with the **red** color and the other neighborhood with **blue** color). I used latitude and longitude values to get the visual as below:



In first step we have collected the required **data for every restaurant within 6km from Toronto Downtown**

[19]:
```
restaurants_df.head()
```

[19]:

	Neighborhood	Neighborhood Latitude	Neighborhood Longitude	Neighborhood X	Neighborhood Y	Neighborhood Distance From Downtown	Venue	Venue Latitude	Venue Longitude	Venue Category
0	Annex	43.670338	-79.407117	-5.307549e+06	1.050992e+07	3781.433005	Fresh on Bloor	43.666755	-79.403491	Vegetarian / Vegan Restaurant
1	Annex	43.670338	-79.407117	-5.307549e+06	1.050992e+07	3781.433005	Roti Cuisine of India	43.674618	-79.408249	Indian Restaurant
2	Annex	43.670338	-79.407117	-5.307549e+06	1.050992e+07	3781.433005	The Original Gyro Grill	43.666621	-79.405544	Greek Restaurant
3	Annex	43.670338	-79.407117	-5.307549e+06	1.050992e+07	3781.433005	Real Thailand Restaurant	43.666557	-79.404657	Thai Restaurant
4	Annex	43.670338	-79.407117	-5.307549e+06	1.050992e+07	3781.433005	Famoso Neapolitan Pizzeria	43.666196	-79.406564	Italian Restaurant

We have also **identified Italian restaurants** (according to Foursquare categorization).

[25]:
```
italian_restaurants_df = toronto_venues[toronto_venues['Venue Category'].str.contains("Italian Restaurant")]
italian_restaurants_df.reset_index(drop=True, inplace=True)
```

[26]:
```
italian_restaurants_df.shape
```

[26]:
```
(33, 10)
```

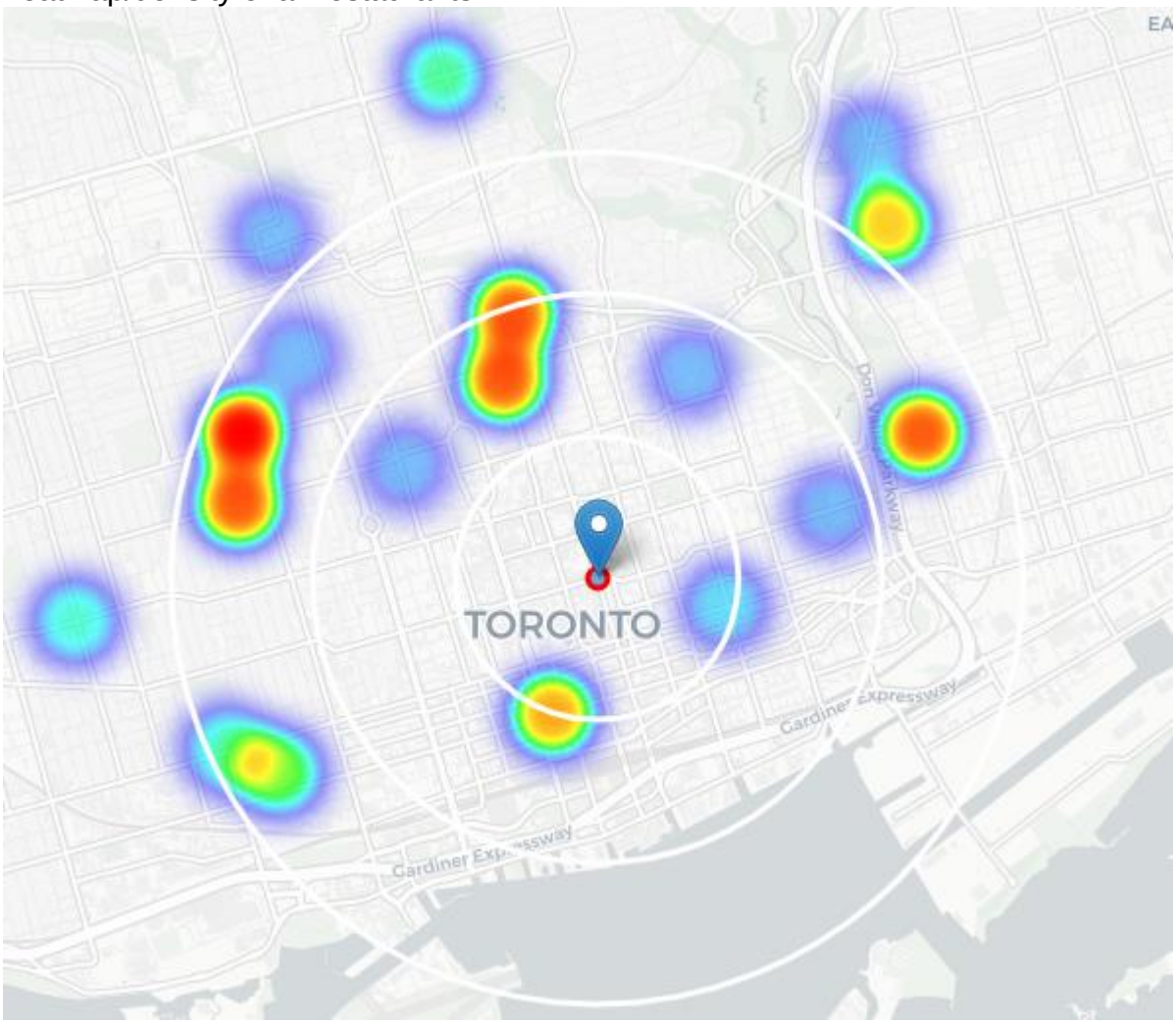
[27]:
```
italian_restaurants_df.head()
```

[27]:

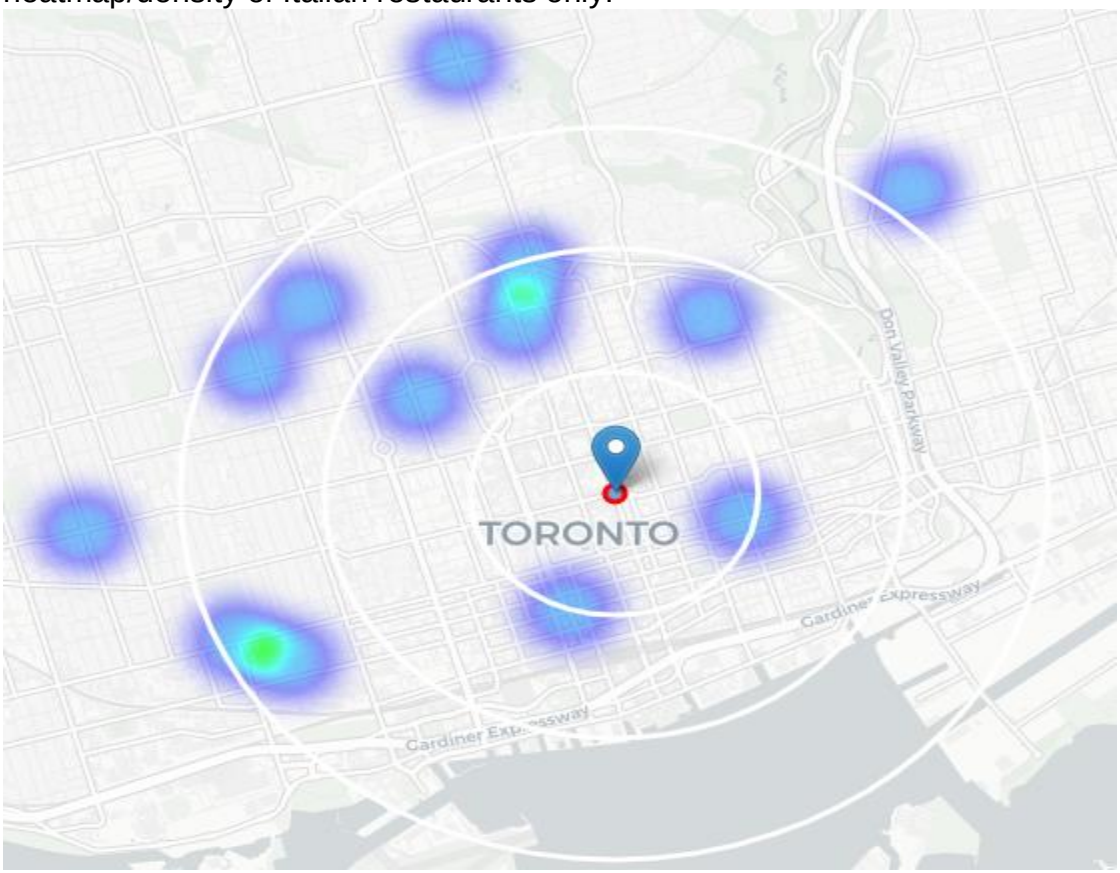
	Neighborhood	Neighborhood Latitude	Neighborhood Longitude	Neighborhood X	Neighborhood Y	Neighborhood Distance From Downtown	Venue	Venue Latitude	Venue Longitude	Venue Category
0	Annex	43.670338	-79.407117	-5.307549e+06	1.050992e+07	3781.433005	Famoso Neapolitan Pizzeria	43.666196	-79.406564	Italian Restaurant
1	Bathurst Manor	43.665519	-79.411937	-5.308253e+06	1.051056e+07	3893.566103	Famoso Neapolitan Pizzeria	43.666196	-79.406564	Italian Restaurant
2	Bay Street Corridor	43.668865	-79.389126	-5.308014e+06	1.050787e+07	2222.603250	Trattoria Nervosa	43.671019	-79.391081	Italian Restaurant
3	Bay Street Corridor	43.668865	-79.389126	-5.308014e+06	1.050787e+07	2222.603250	Blu Ristorante and Lounge	43.671685	-79.388614	Italian Restaurant
4	Bay Street Corridor	43.668865	-79.389126	-5.308014e+06	1.050787e+07	2222.603250	Buca	43.671917	-79.389236	Italian Restaurant

Second step in our analysis will be calculation and exploration of '**restaurant density**' across different areas of Toronto - we will use **heatmaps** to identify a few promising areas close to Downtown with low number of restaurants in general (*and* no Italian restaurants in vicinity) and focus our attention on those areas.

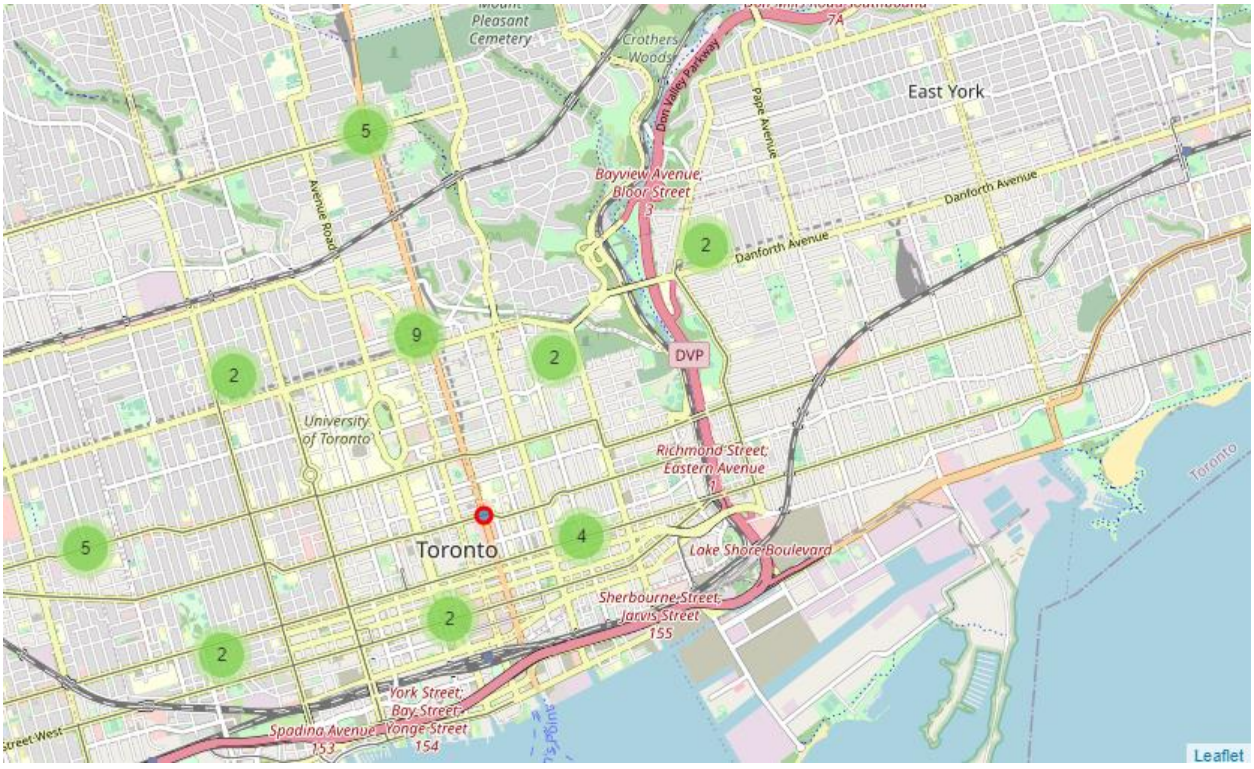
a- heatmap/density of all restaurants:



b- heatmap/density of Italian restaurants only:



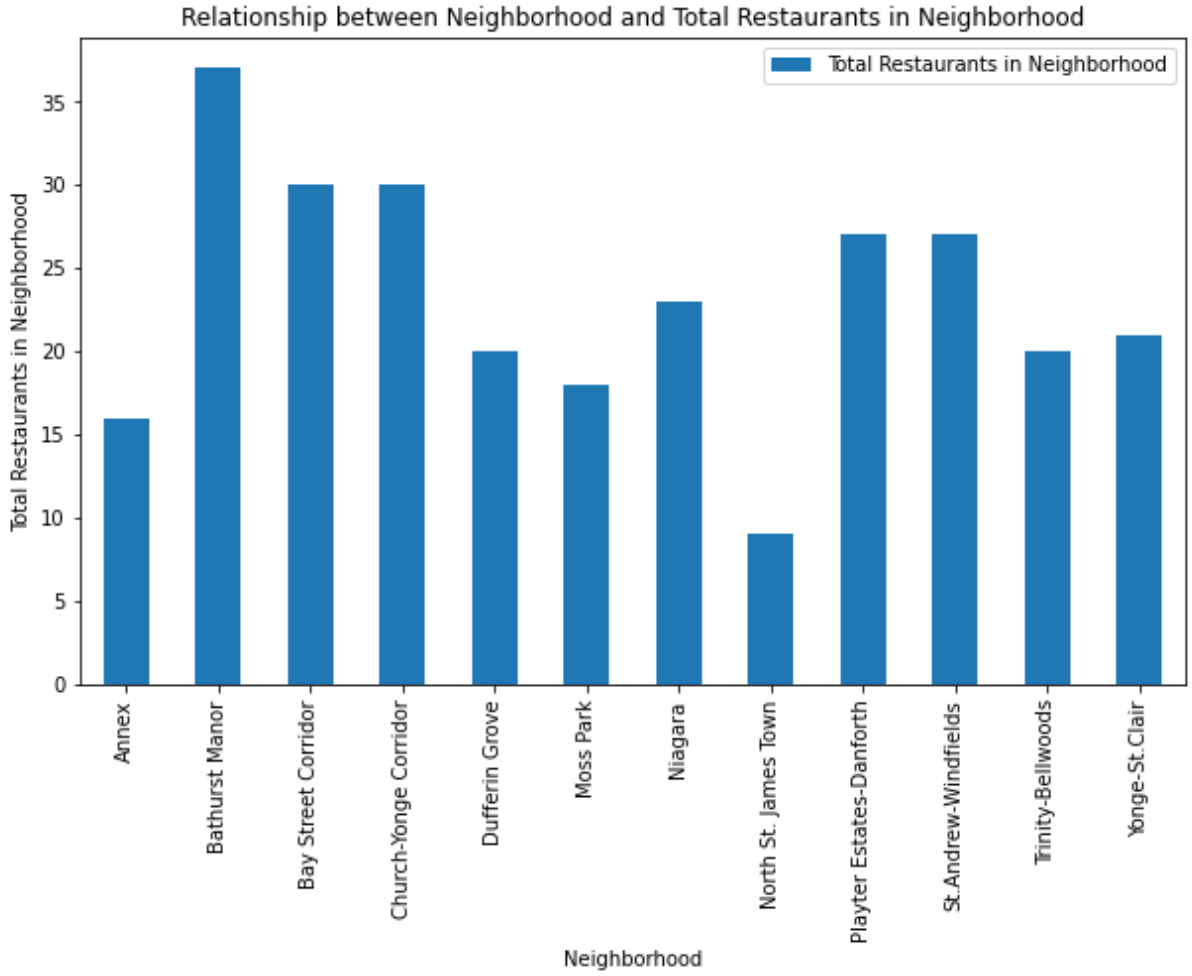
c- Distribution of the Italian restaurants:



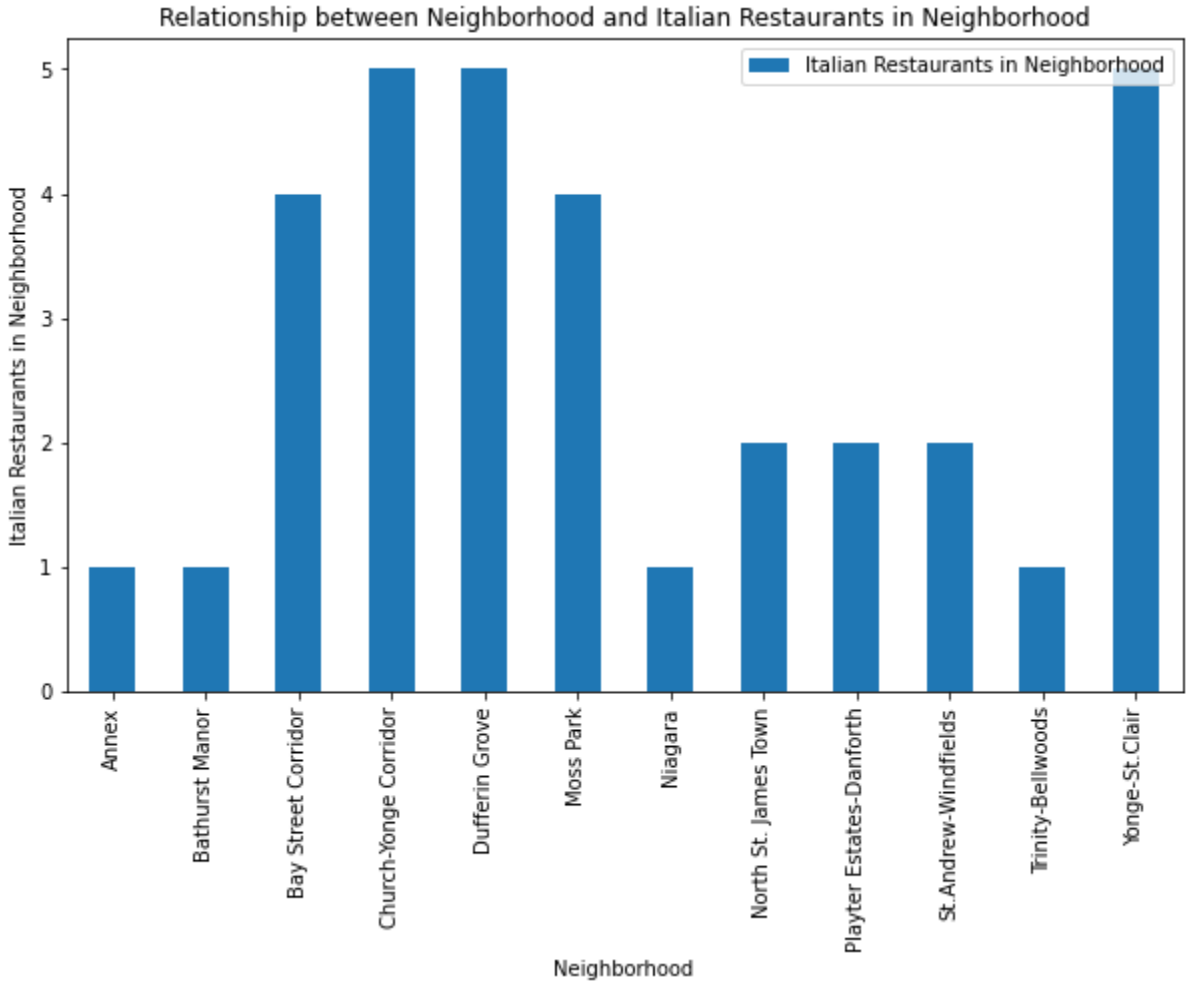
Then I created the table that contains all neighborhood useful information:

	Neighborhood	Neighborhood Distance from Downtown	Neighborhood Average Distance to Italian Restaurant	Italian Restaurants in Neighborhood	Total Restaurants in Neighborhood
0	Annex	3781.433005	665.860717	1	16
1	Bathurst Manor	3893.566103	633.570247	1	37
2	Bay Street Corridor	2222.603250	452.927237	4	30
3	Church-Yonge Corridor	2815.917939	420.395632	5	30
4	Dufferin Grove	5308.174412	439.791570	5	20
5	Moss Park	1327.800979	485.132152	4	18
6	Niagara	3777.940174	383.166767	1	23
7	North St. James Town	2300.852910	623.328355	2	9
8	Playter Estates-Danforth	4636.183478	480.191933	2	27
9	St.Andrew-Windfields	1440.871436	500.231430	2	27
10	Trinity-Bellwoods	4139.567767	227.509532	1	20
11	Yonge-St.Clair	5317.430127	310.412955	5	21

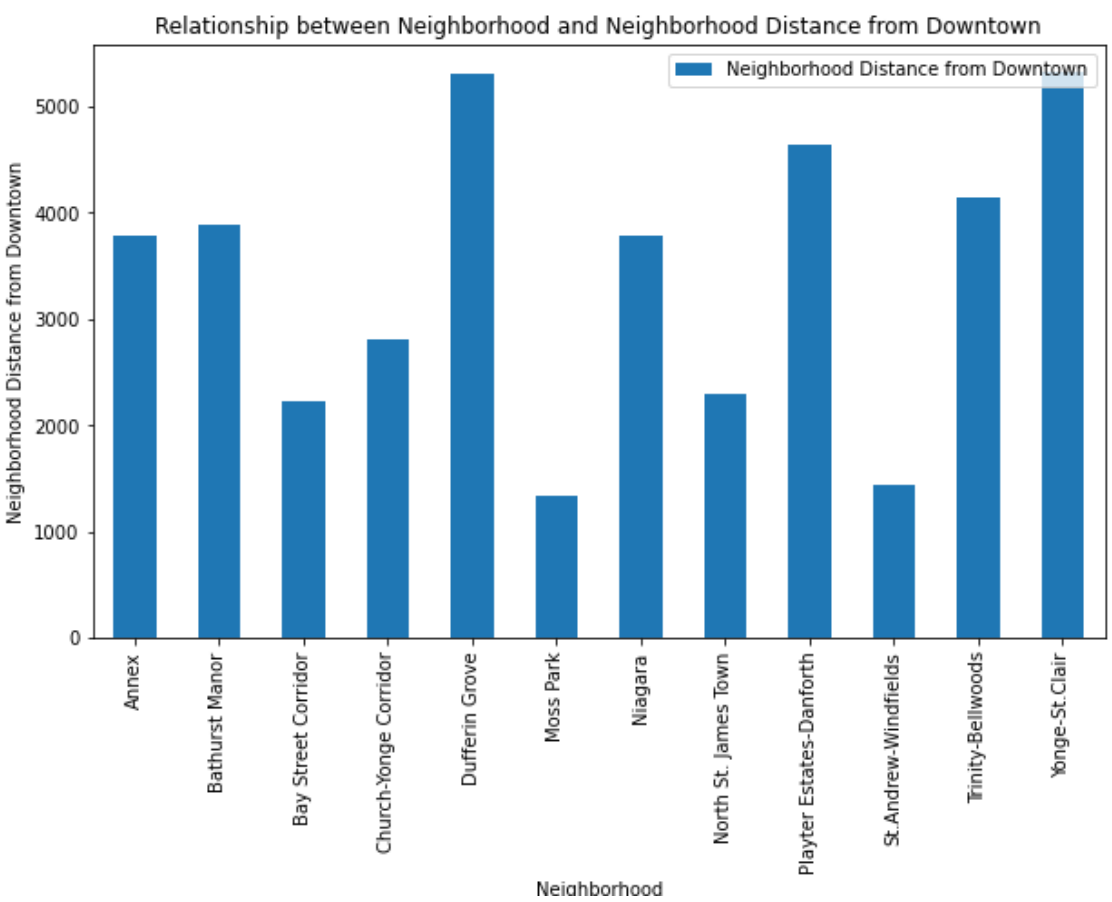
1- Relationship between Neighborhood and Total Restaurants in Neighborhood:



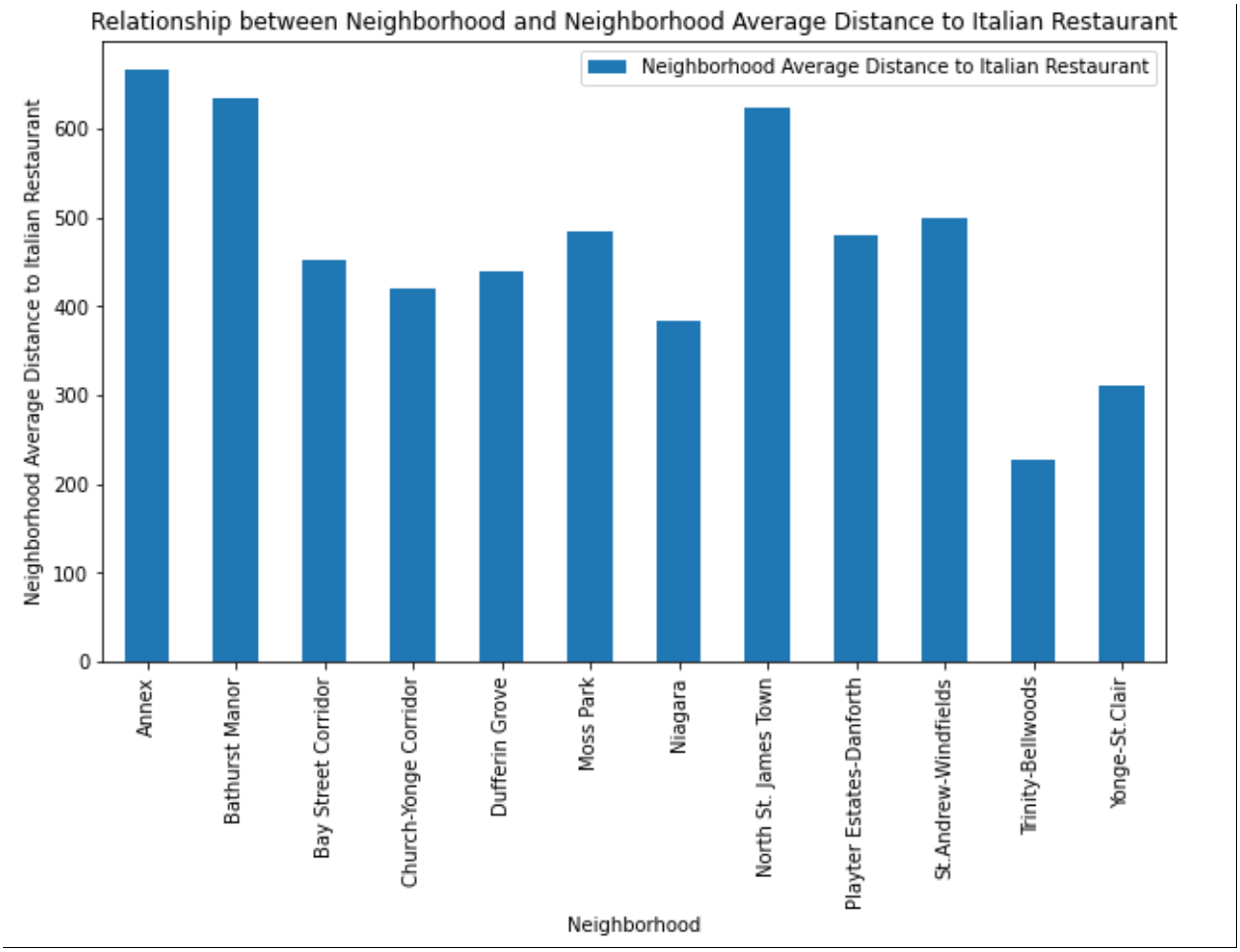
2- Relationship between Neighborhood and Italian Restaurants in Neighborhood:



3- Relationship between Neighborhood and Neighborhood Distance from Downtown:

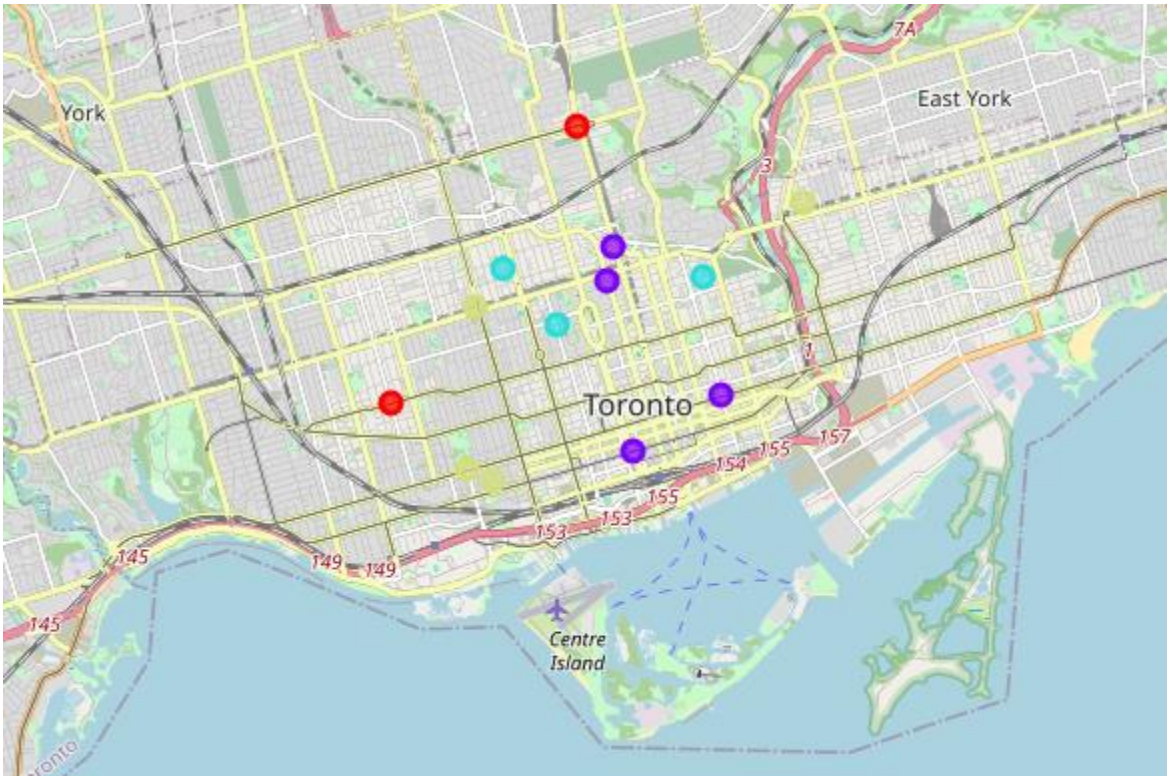


4- Relationship between Neighborhood and Neighborhood Average Distance to Italian Restaurant:



In third and final step we will focus on most promising areas and within those create **clusters of locations that meet some basic requirements** established in discussion with stakeholders: we will take into consideration locations with **smaller** distances from Downtown, **Larger** average distances to Italian restaurant, **smaller** numbers of Italian restaurants and **smaller** total numbers of restaurants.

We will present map of all such locations but also create clusters (using **k-means clustering**) of those locations



Results and Discussion:

Our analysis shows that although there is a great number of restaurants in Toronto (~13 00 Within 6km area from Downtown), there are pockets of low restaurant density fairly cl ose to Downtown. Highest concentration of restaurants was detected north and west fro m Downtown, so we focused our attention to areas south, south-east and east.
We considered only the neighborhoods within the range of 2.5km from Downtown, then we considered only the neighborhoods within **two or less** Italian restaurants, then we selected only neighborhoods with the minimum total numbers of restaurants.

we will filter our results to contains only neighborhoods with two or less italian restaurants and within the range of 2.5km

```
[45]: df_good_locations = df_Neiborhood[(df_Neiborhood['Italian Restaurants in Neighborhood'] <= 2) & (df_Neiborhood['Neighborhood D
df_good_locations
```

[45]:

	Neighborhood	Neighborhood Distance from Downtown	Neighborhood Average Distance to Italian Restaurant	Italian Restaurants in Neighborhood	Total Restaurants in Neighborhood
7	North St. James Town	2300.852910	623.328355	2	9
9	St.Andrew-Windfields	1440.871436	500.231430	2	27

we find that **North St. James Town** neighborhood is a good location as it has **smaller** total number of Restaurants in Neighborhood and **larger** average distance to the Italian Restaurants within a neighborhood

This led us to focus only on certain neighborhoods which will be a suitable places to open an Italian restaurant in (North St. James Town Neighborhood).

Conclusion:

Purpose of this project was to identify Toronto areas close to Downtown with low number of restaurants (particularly Italian restaurants) in order to aid stakeholders in narrowing down the search for optimal location for a new Italian restaurant. By calculating restaurant density distribution from Foursquare. Clustering of those locations was then performed in order to create major zones of interest for final exploration by stakeholders.

Final decision on optimal restaurant location will be made by stakeholders based on specific characteristics of neighborhoods and locations in every recommended zone, taking into consideration additional factors like attractiveness of each location (proximity to park or water), levels of noise / proximity to major roads, real estate availability, prices, social and economic dynamics of every neighborhood etc.