Finding the optimal spot for our authentic Italian Restaurant

# The Battle of Neighbourhoods

Sebastiaan Vrij

## Table of Contents

# Part 1: Introduction

In this project, I am creating a hypothetical scenario for a concept that there may not be enough Italian Restaurants in Toronto Area. With the purpose in mind, finding the location to open such a restaurant is one of the most important decisions for this entrepreneur and I am designing this project to help him find the most suitable location.

## Business Problem

In this project we will try to find the best locations to open this Italian restaurant. We will use our data science powers to find a few most promising neighbourhoods where there are not many Italian Restaurants yet.

## Target Audience

Our target stakeholders are businesspeople and investors that want to open an Italian restaurant in Toronto Canada.

# Part 2 Data Selection

Following data sources will be used to get the required information:

1. Wikipedia will be used scrap Toronto neighbourhoods;
2. Geospatial_Coordinates.csv will be used to get Latitude and Longitude information;
3. Foursquare API will be used to get restaurants data related to these 2 cities.

Above data sources will be used to get venues and Italian restaurants information in order to identify in which area has the most Italian restaurants and, this way, select the area with the least amount of restaurants.

## Data flow

1. First, it is used data from get city open data to get city information as well as latitude and longitude coordinates
2. Then, we created a data frame with borough and neighbourhood information. For Toronto, it is used Wikipedia to get the list of Postal Code of all Neighbourhoods in Toronto
3. And. for the neighbourhood in this data frame, it will be gathered the list of restaurants from Foursquare. With this information it is possible to come up with a total as well as draw the maps with Italian restaurants locations.

## Part 3 Methodology

The goal of this project is to come up with a simple study to identify area's in the city of Toronto, where Italian Restaurants are located. So we can define areas of opportunities to invest / start an new Italian Restaurant.

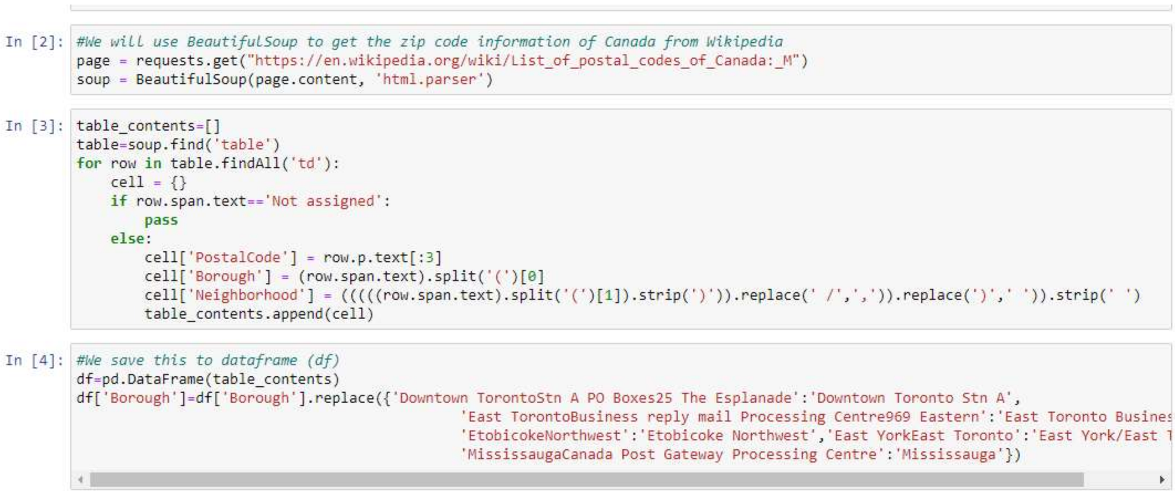
After that, it will be presented some number to justify the decision about which area has the most restaurant which helps us determine other area's where we could start our restaurant.

And finally, in the last part of this study, it is showed a map showing the spots where these Italian restaurants are located, and helps us to visualize the areas of opportunity for our restaurant.

## Part 4 - Analysis

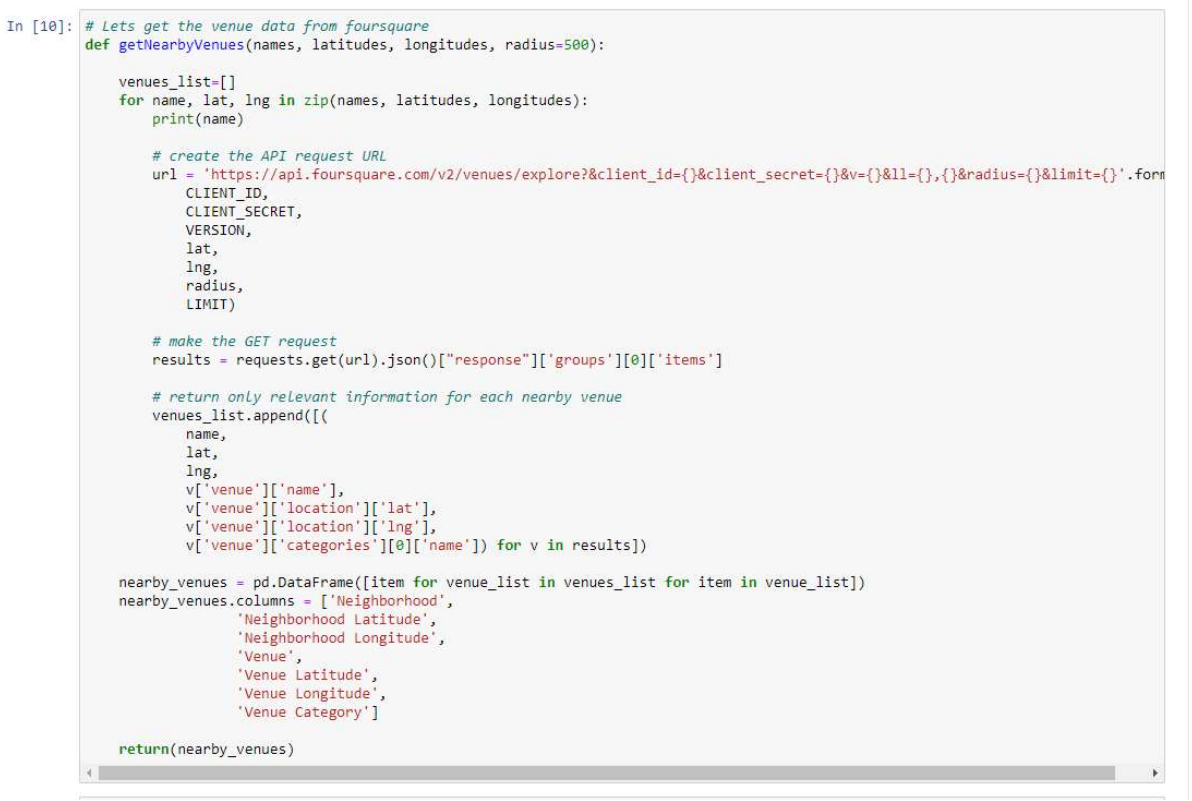
First, it is identified some basic geographic info from Toronto. Based on the zip code of the different area's.

Get geo information from Toronto from Wikipedia



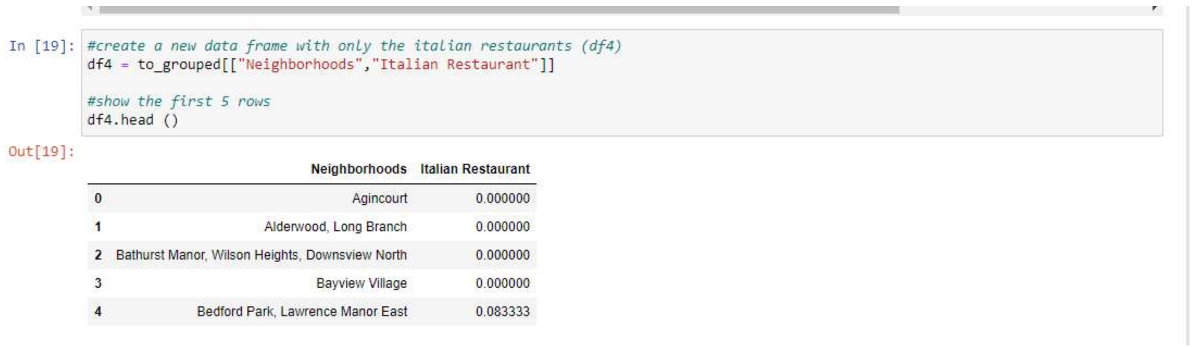
```
In [2]: #We will use BeautifulSoup to get the zip code information of Canada from Wikipedia
        page = requests.get("https://en.wikipedia.org/wiki/List_of_postal_codes_of_Canada:_M")
        soup = BeautifulSoup(page.content, 'html.parser')

In [3]: table_contents=[]
        table=soup.find('table')
        for row in table.findAll('td'):
            cell = {}
            if row.span.text=='Not assigned':
                pass
            else:
                cell['PostalCode'] = row.p.text[:3]
                cell['Borough'] = (row.span.text).split('(')[0]
                cell['Neighborhood'] = (((((row.span.text).split('(')[1]).strip(')')).replace(' /',',')).replace(')',' ')).strip(' ')
                table_contents.append(cell)

In [4]: #We save this to dataframe (df)
        df=pd.DataFrame(table_contents)
        df['Borough']=df['Borough'].replace({'Downtown TorontoStn A PO Boxes25 The Esplanade':'Downtown Toronto Stn A',
                                              'East TorontoBusiness reply mail Processing Centre969 Eastern':'East Toronto Busines
                                              'EtobicokeNorthwest':'Etobicoke Northwest','East YorkEast Toronto':'East York/East T
                                              'MississaugaCanada Post Gateway Processing Centre':'Mississauga'})
```

We extracted some venue data from foursquare



```
In [10]: # Lets get the venue data from foursquare
         def getNearbyVenues(names, latitudes, longitudes, radius=500):

             venues_list=[]
             for name, lat, lng in zip(names, latitudes, longitudes):
                 print(name)

                 # create the API request URL
                 url = 'https://api.foursquare.com/v2/venues/explore?&client_id={}&client_secret={}&v={}&ll={},{}&radius={}&limit={}'.form
                     CLIENT_ID,
                     CLIENT_SECRET,
                     VERSION,
                     lat,
                     lng,
                     radius,
                     LIMIT)

                 # make the GET request
                 results = requests.get(url).json()["response"]['groups'][0]['items']

                 # return only relevant information for each nearby venue
                 venues_list.append([(
                     name,
                     lat,
                     lng,
                     v['venue']['name'],
                     v['venue']['location']['lat'],
                     v['venue']['location']['lng'],
                     v['venue']['categories'][0]['name']) for v in results])

             nearby_venues = pd.DataFrame([item for venue_list in venues_list for item in venue_list])
             nearby_venues.columns = ['Neighborhood',
                           'Neighborhood Latitude',
                           'Neighborhood Longitude',
                           'Venue',
                           'Venue Latitude',
                           'Venue Longitude',
                           'Venue Category']

             return(nearby_venues)
```

We filtered it to only Italian restaurants.



```
In [19]: #create a new data frame with only the italian restaurants (df4)
         df4 = to_grouped[["Neighborhoods","Italian Restaurant"]]

         #show the first 5 rows
         df4.head ()
```

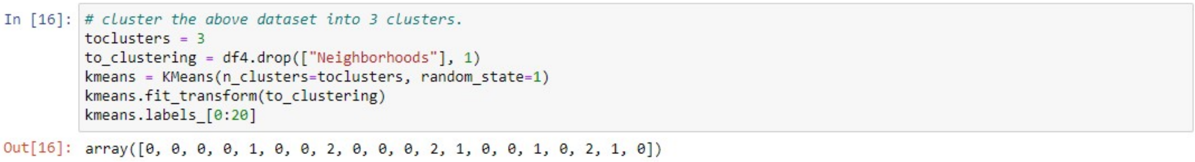
Out[19]:

| | Neighborhoods | Italian Restaurant |
|---|---|---|
| 0 | Agincourt | 0.000000 |
| 1 | Alderwood, Long Branch | 0.000000 |
| 2 | Bathurst Manor, Wilson Heights, Downsview North | 0.000000 |
| 3 | Bayview Village | 0.000000 |
| 4 | Bedford Park, Lawrence Manor East | 0.083333 |

Now in this new dataset we want to determine clusters to see if we can find areas where there are not many restaurants yet. We will do this with a type of analysis called K-Means.

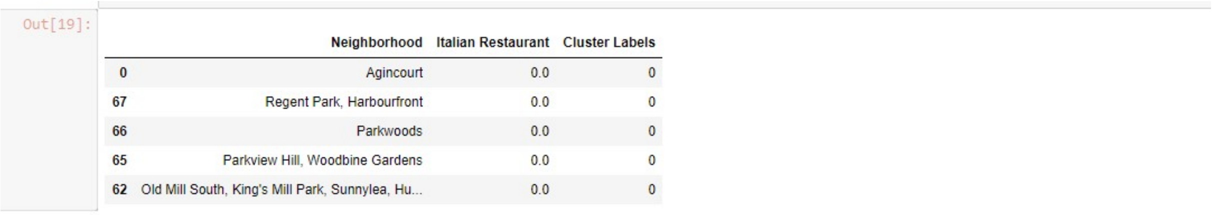
## K-Means

K-means clustering is a type of unsupervised learning, which is used when you have unlabeled data (i.e., data without defined categories or groups). The goal of this algorithm is to find groups in the data, with the number of groups represented by the variable K. The algorithm works iteratively to assign each data point to one of K groups based on the features that are provided. Data points are clustered based on feature similarity.



```
In [16]: # cluster the above dataset into 3 clusters.
         toclusters = 3
         to_clustering = df4.drop(["Neighborhoods"], 1)
         kmeans = KMeans(n_clusters=toclusters, random_state=1)
         kmeans.fit_transform(to_clustering)
         kmeans.labels_[0:20]

Out[16]: array([0, 0, 0, 0, 1, 0, 0, 2, 0, 0, 0, 2, 1, 0, 0, 1, 0, 2, 1, 0])
```

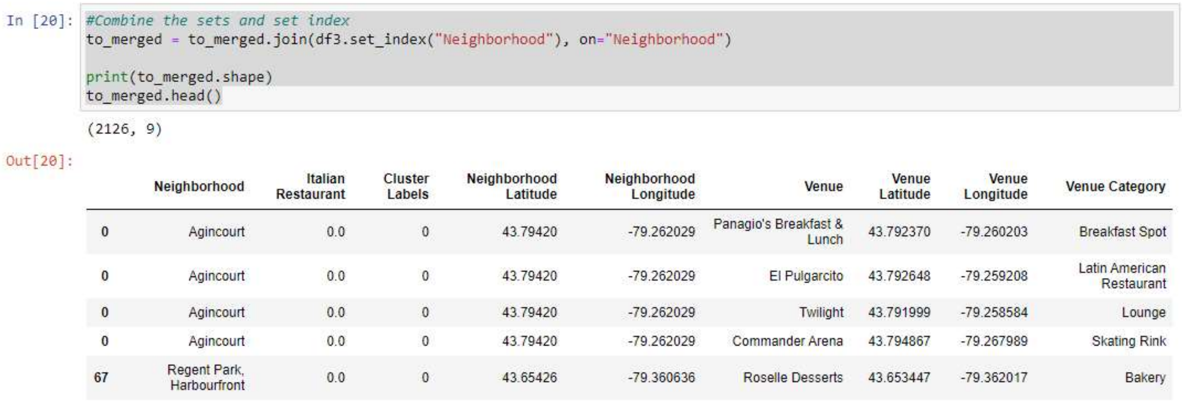
We can see that clusters 0,1 and 2 are being created.



Out[19]:

| | Neighborhood | Italian Restaurant | Cluster Labels |
|---|---|---|---|
| **0** | Agincourt | 0.0 | 0 |
| **67** | Regent Park, Harbourfront | 0.0 | 0 |
| **66** | Parkwoods | 0.0 | 0 |
| **65** | Parkview Hill, Woodbine Gardens | 0.0 | 0 |
| **62** | Old Mill South, King's Mill Park, Sunnylea, Hu... | 0.0 | 0 |

We combined this with the previous dataset to get one total data set.



```
In [20]: #Combine the sets and set index
         to_merged = to_merged.join(df3.set_index("Neighborhood"), on="Neighborhood")

         print(to_merged.shape)
         to_merged.head()

         (2126, 9)
```

Out[20]:

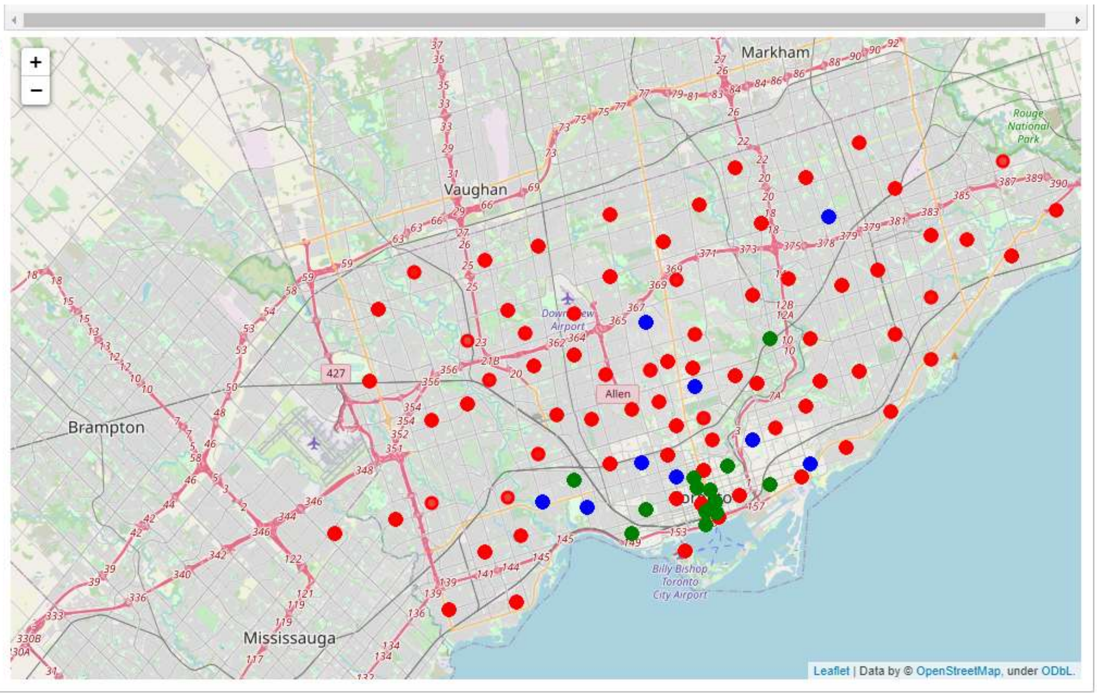
| | Neighborhood | Italian Restaurant | Cluster Labels | Neighborhood Latitude | Neighborhood Longitude | Venue | Venue Latitude | Venue Longitude | Venue Category |
|---|---|---|---|---|---|---|---|---|---|
| **0** | Agincourt | 0.0 | 0 | 43.79420 | -79.262029 | Panagio's Breakfast & Lunch | 43.792370 | -79.260203 | Breakfast Spot |
| **0** | Agincourt | 0.0 | 0 | 43.79420 | -79.262029 | El Pulgarcito | 43.792648 | -79.259208 | Latin American Restaurant |
| **0** | Agincourt | 0.0 | 0 | 43.79420 | -79.262029 | Twilight | 43.791999 | -79.258584 | Lounge |
| **0** | Agincourt | 0.0 | 0 | 43.79420 | -79.262029 | Commander Arena | 43.794867 | -79.267989 | Skating Rink |
| **67** | Regent Park, Harbourfront | 0.0 | 0 | 43.65426 | -79.360636 | Roselle Desserts | 43.653447 | -79.362017 | Bakery |

## Part 5 – Results

Now that we have create the clusters with K-means we want first find out in which cluster are the least amount of Italian restaurants. So we know where to invest.

Cluster 0 = Red
Cluster 1 = Blue
Cluster 2 = Green

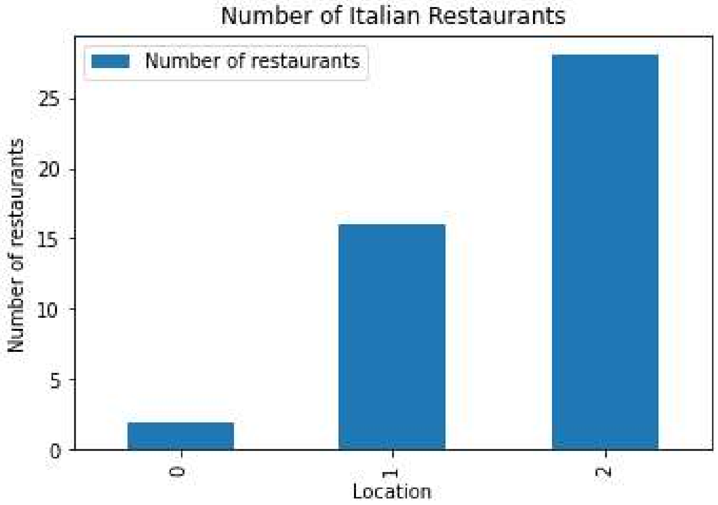
```
Out[23]:
```


+
−
Markham
Vaughan
Brampton
Mississauga
Rouge National Park
Downsview Airport
Billy Bishop Toronto City Airport
Leaflet | Data by © OpenStreetMap, under ODbL.

We can see tha

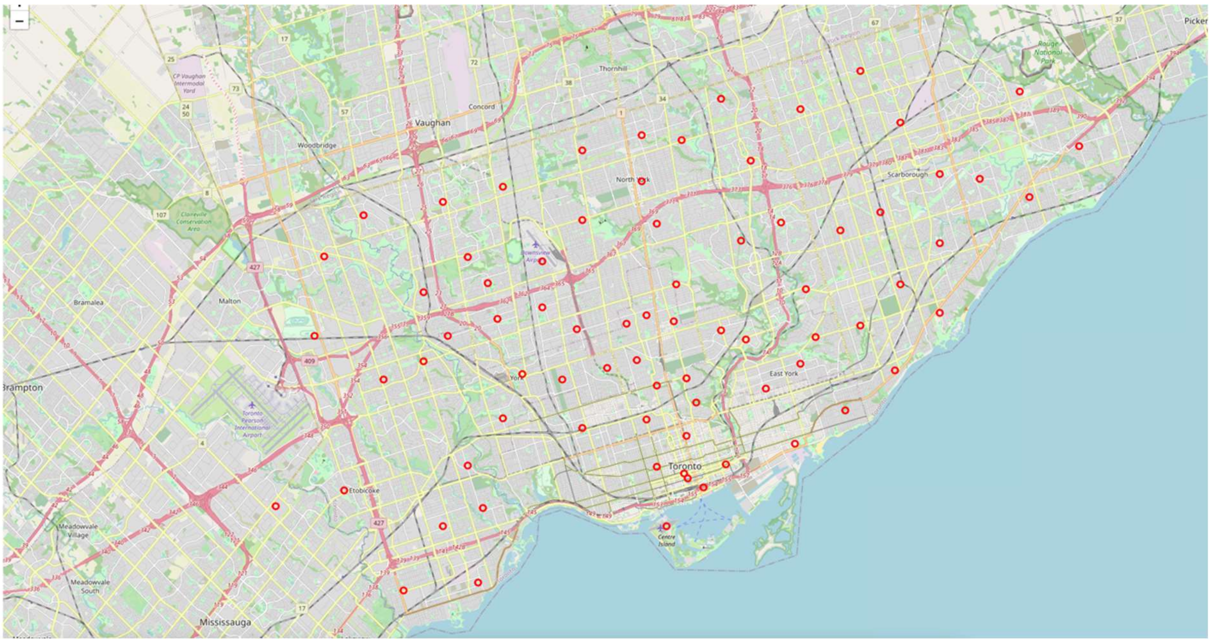
| Segment | Number of Italian Restaurants |
|---|---|
| 0 | 2 |
| 1 | 16 |
| 2 | 28 |

```
Out[37]: <function matplotlib.pyplot.show(close=None, block=None)>
```



We can see that segment 0 has the least number of restaurants. Therefore, we want to be in this area.

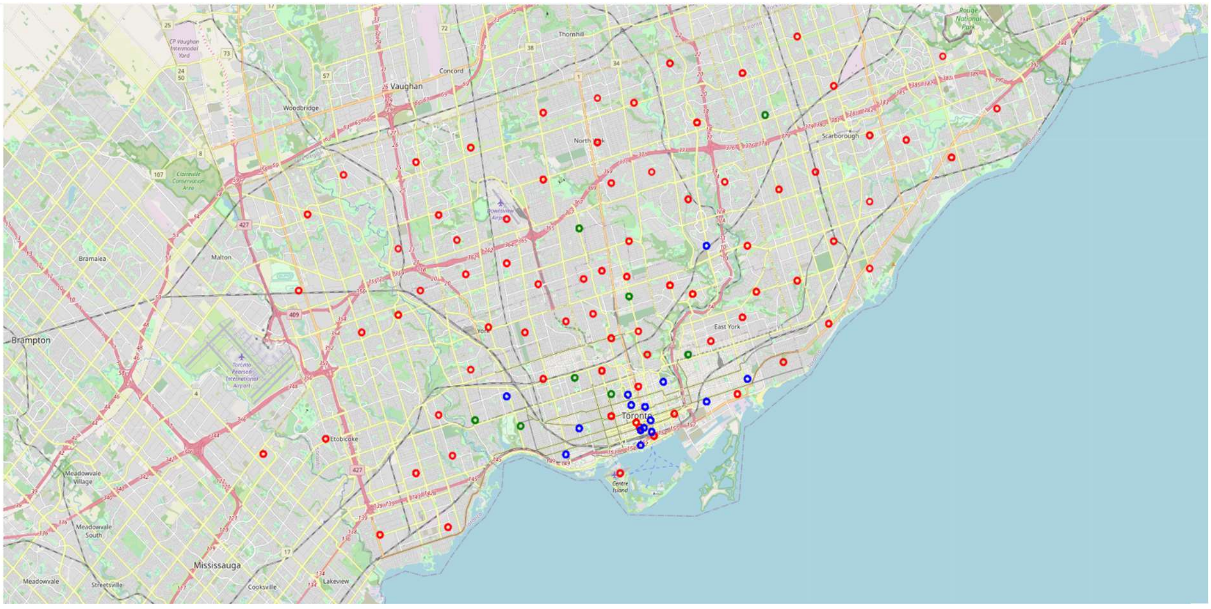
Below you can find how area 0 looks like



## Recommendations

Most of the Italian restaurants are in cluster 1 lowest in Cluster 0.
Looking at nearby venues it seems cluster 1 might be a good location as there are not a lot of Italian restaurants in these areas. We therefore recommend the entrepreneur to open an authentic Italian restaurant in these locations. If we look to the total map of all the area's

Area 0 = Red
Area 1 = Blue
Area 2 = Green



We might want to explorer the areas close to the blue and green areas first because there are likely to be more downtown.

## Appendix 1 List overview Area’s

Agincourt

Berczy Park
CN Tower, King and Spadina, Railway Lands, Harbourfront West, Bathurst Quay, South Niagara, Island airport
Birch Cliff, Cliffside West

Bathurst Manor, Wilson Heights, Downsview North

Alderwood, Long Branch

Downsview Central

Downsview East

Davisville North

First Canadian Place, Underground city

Bayview Village

Cliffside, Cliffcrest, Scarborough Village West

Clairville, Humberwood, Woodbine Downs, West Humber, Kipling Heights, Rexdale, El ms, Tandridge, Old Rexdale

Church and Wellesley

Humber Summit

Del Ray, Mount Dennis, Keelsdale and Silverthorn

Runnymede, The Junction North

The Kingsway, Montgomery Road, Old Mill North

Humberlea, Emery

Kingsview Village, St. Phillips, Martin Grove Gardens, Richview Gardens

Hillcrest Village

Dorset Park, Wexford Heights, Scarborough Town Centre

Don Mills North

Cedarbrae

Caledonia-Fairbanks

Fairview, Henry Farm, Oriole

Eringate, Bloordale Gardens, Old Burnhamthorpe, Markland Wood

Enclave of M4L

Downsview West

Dufferin, Dovercourt Village

York Mills West

Enclave of L4W

Guildwood, Morningside, West Hill

Golden Mile, Clairlea, Oakridge

Glencairn

Forest Hill North & West

Downsview Northwest

Scarborough Village

Richmond, Adelaide, King

Leaside

Lawrence Park

Lawrence Manor, Lawrence Heights

Moore Park, Summerhill East

Milliken, Agincourt North, Steeles East, L'Amoreaux East

Malvern, Rouge

Regent Park, Harbourfront

Parkwoods

Parkview Hill, Woodbine Gardens

Old Mill South, King's Mill Park, Sunnylea, Humber Bay, Mimico NE, The Queensway East, Royal York South East, Kingsway Park South East

Northwood Park, York University

New Toronto, Mimico South, Humber Bay Shores

North Park, Maple Leaf Park, Upwood Park

North Toronto West

Woodbine Heights

Roselawn

Thorncliffe Park

The Danforth  East

Victoria Village

Willowdale South

Wexford, Maryvale

Weston

Westmount

The Beaches

Summerhill West, Rathnelly, South Hill, Forest Hill SE, Deer Park

Steeles West, L'Amoreaux West

South Steeles, Silverstone, Humbergate, Jamestown, Mount Olive, Beaumond Heights, Thistletown, Albion Gardens

The Annex, North Midtown, Yorkville

Kensington Market, Chinatown, Grange Park

Kennedy Park, Ionview, East Birchmount Park

Woburn

Willowdale, Newtonbrook

Willowdale West

Rouge Hill, Port Union, Highland Creek

Mimico NW, The Queensway West, South of Bloor, Kingsway Park South West, Royal Y ork South West

Rosedale

Humewood-Cedarvale