1. Introduction

This project is viewed from the perspective of an owner of an Italian restaurant. The said owner wants to open up another restaurant in Toronto where he/she currently has no outlets. The goal of this analysis is to find out which neighborhood(s) are ideal for opening an Italian restaurant.

The ideal audience for this would be Italian restaurant owners trying to set up shop in a new location and for investors to check the feasibility of the project.

2. Data

The data is gathered from Wikipedia and Foursquare.
Wikipedia provides the names and borough of the different neighborhoods in Toronto.
The data from Foursquare is used to gain information on the above mentioned neighborhoods, specifically the number of Italian restaurants in the neighborhoods.

By leveraging the data from the foursqaure API we get to know the places where there is an abundance of Italian restaurants and areas where there aren't many. This data can be leveraged to pick an ideal location to set up shop.
Example of the neighborhood data is as follows:

Postal Code ⇕	Borough ⇕	Neighborhood ⇕
M1A	Not assigned	
M2A	Not assigned	
M3A	North York	Parkwoods
M4A	North York	Victoria Village
M5A	Downtown Toronto	Regent Park, Harbourfront
M6A	North York	Lawrence Manor, Lawrence Heights
M7A	Downtown Toronto	Queen's Park, Ontario Provincial Government
M8A	Not assigned	
M9A	Etobicoke	Islington Avenue
M1B	Scarborough	Malvern, Rouge
M2B	Not assigned	
M3B	North York	Don Mills
M4B	East York	Parkview Hill, Woodbine Gardens
M5B	Downtown Toronto	Garden District, Ryerson
M6B	North York	Glencairn
M7B	Not assigned	
M8B	Not assigned	
M9B	Etobicoke	West Deane Park, Princess Gardens, Martin Grove, Islington, Cloverdale
M1C	Scarborough	Rouge Hill, Port Union, Highland Creek
M2C	Not assigned	
M3C	North York	Don Mills
M4C	East York	Woodbine Heights
M5C	Downtown Toronto	St. James Town
M6C	York	Humewood-Cedarvale

Example of the data the foursquare API will provide
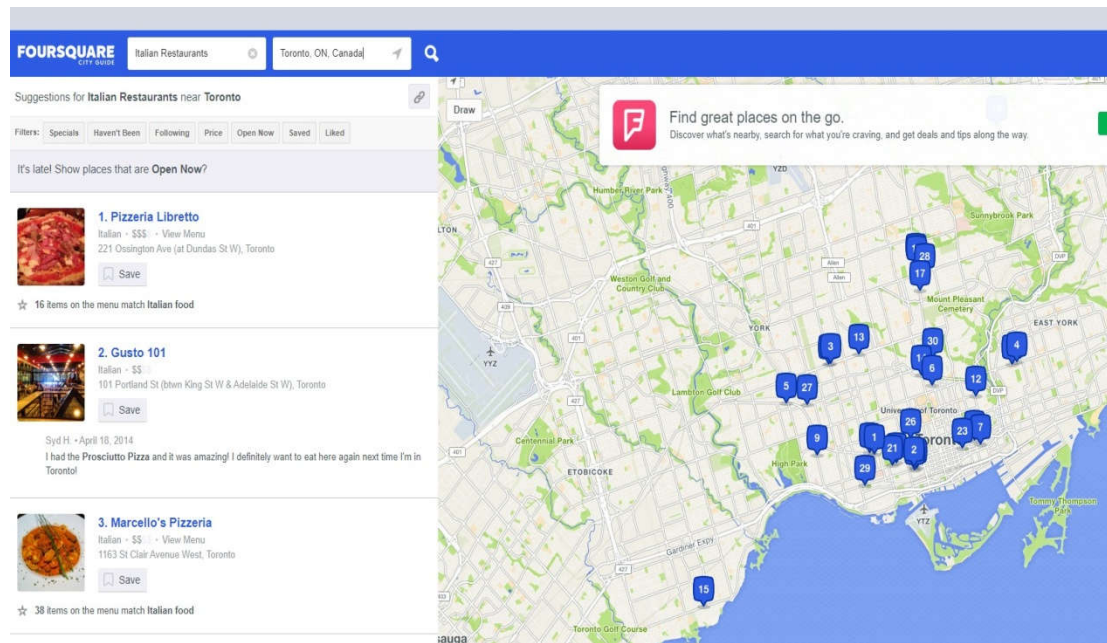
3. Methodology

3.1 Data Collection

There are two main data sources from which data must be collected, Wikipedia and Foursquare.

The data from Wikipedia gives us the names of the neighborhoods in Cannada and the data from the Foursquare API gives us information on each neighborhood

3.2 Transforming the data

After aquiring the data from passing each neighborhood name into the Foursquare API our data looks like this

	Neighborhood	Neighborhood Latitude	Neighborhood Longitude	Venue	Venue Latitude	Venue Longitude
0	Parkwoods	43.753259	-79.329656	VIA CIBO | italian streetfood	43.754067	-79.357951
1	Parkwoods	43.753259	-79.329656	Sorento Restaurant	43.726575	-79.341989
2	Parkwoods	43.753259	-79.329656	Paisano	43.754946	-79.349580
3	Parkwoods	43.753259	-79.329656	Remezzo Italian Bistro	43.778649	-79.308264
4	Parkwoods	43.753259	-79.329656	Beiruti Grand Cafe	43.773002	-79.332533

3.2.1 Visualising all the restaurants on the map
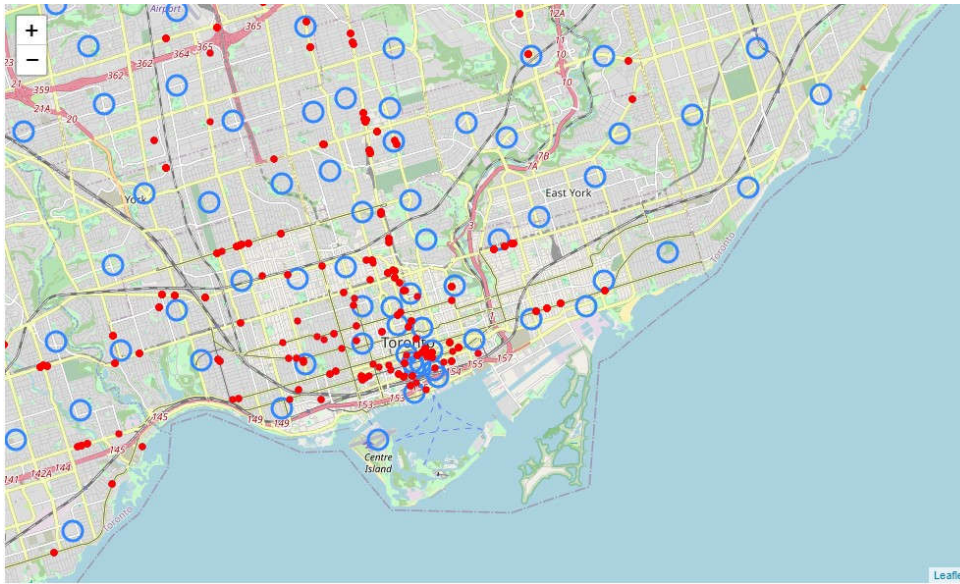
The above data shows us all the Italian Restaurants in Toronto, but does not hold much value for our analysis. Therefore, we group the data based on the neighborhood and aggregate it based on the count. The data looks something like this:

Neighborhood	Neighborhood Latitude	Neighborhood Longitude	Venue	Venue Latitude	Venue Longitude
Downsview	40	40	40	40	40
Don Mills	20	20	20	20	20
Willowdale	20	20	20	20	20
Parkwoods	10	10	10	10	10
Scarborough Village	10	10	10	10	10

3.3 Analyzing the Data

The goal of the analysis is the find the 5 most Italian Restaurant dense areas and the 5 areas with the least density of Italian Restaurants

Using the data we have, which is in sorted form we can pull out these 10 neighborhoods. The top 5 neighborhoods are:

```
top_5_neighborhoods = list(count_df_sorted.iloc[:5,0])
print("The top 5 neighborhoods are :")
print(top_5_neighborhoods)
```

```
The top 5 neighborhoods are :
['Downsview', 'Don Mills', 'Willowdale', 'Parkwoods', 'Scarborough Village']
```

The bottom 5 neighborhoods are:

```
bottom_5_neighborhoods = list(count_df_sorted.tail(5).iloc[:,0])
print("The bottom 5 nrighborhoods are: ")
bottom_5_neighborhoods.reverse()
print(bottom_5_neighborhoods)
```

```
The bottom 5 nrighborhoods are:
['Woburn', 'Cedarbrae', 'York Mills, Silver Hills', 'First Canadian Place, Underground ci
```

4.Results

By taking into consideration the assumptions stated above, the green and yellow areas of the map are ideal places to open up an italian restaurant in toronto with Green being the highly dense and yellow being less dense in terms of number of Italian Restaurants per neighborhood. The blue dots are all the neighborhoods in toronto.
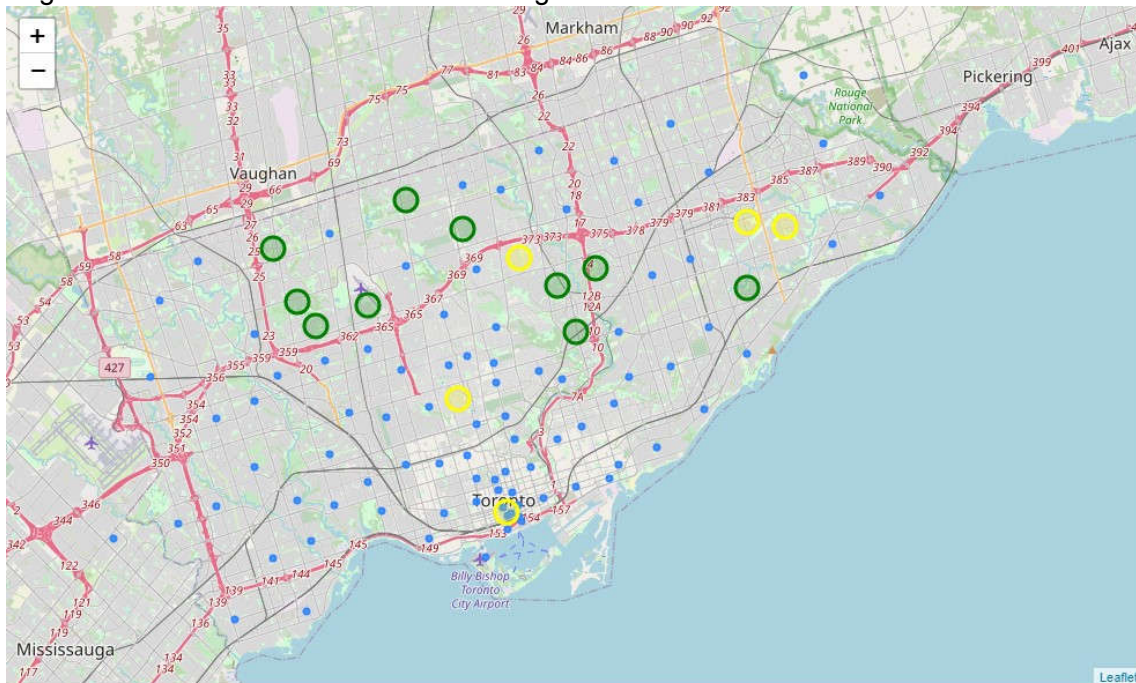
5.Discussions

This study can further be improved by acquiring information about the footfall of the italian restaurants which will show us the true density value of Italian Restaurant lovers in each neighborhood, after normalization.

Another type of data that could be useful is the reviews of such restaurants which can help us beat the competition by overcoming their flaws and coming out withour own unqiue solution to those flaws.

6.Conclusion

In conclusion, I would like to state that, based on the assumptions, the areas marked in green on the map are the most ideal areas to set up an Italian Restaurant and the ares in yellow are potential markets

I also would like to state that the ideal market strategy would be to set up shop in one of the 'green' neighborhoods and then slowly expand to the 'yellow' neighborhoods as the 'green' neighborhoods already have proof of concept of Italian Restaurants. But this conclusion would require data that I have mentioned in the Discussion section of the report.