



# BEST PLACE FOR GROCERIES WAREHOUSE

- APPLIED DATA SCIENCE PROFESSIONAL CERTIFICATE
- VISHAL DHIMAN
- (DVISHAL231995@GMAIL.COM)

# 1. INTRODUCTION/ BUSINESS PROBLEM:

- There is a contractor who buys groceries from villagers and supplies it to the restaurants and other kind of places in Scarborough. Transporting groceries from long distance to different locations in Scarborough, costs more. So he thought he should build a warehouse in Scarborough near the restaurants and shops to whom he can deliver groceries so he can save his transportation money and time as well.

## - ADVANTAGES OF BUILDING WAREHOUSE NEAR CUSTOMERS:

- Building warehouse near his customers, can deliver them fresh items any time.
- Cooks in restaurants can get fresh groceries early morning and starts cooking without waiting so long for the groceries.
- There would be no delay in groceries delivery.
- After supplying fresh items without any delay build customer's trust.

## - TARGET AUDIENCE

- My analysis will educate someone who wants to build warehouse in the neighbourhood from where he can deliver the items to their customers quickly and saves lots of transportation money. It will give more business to the warehouse owner.
- This analysis is also helpful to those who want to open new business like restaurants, bars or shops in the neighbourhood. They can check competition level in that neighbourhood or what is more popular in that neighbourhood and then decide which neighbourhood have more chances of profit for the business.



## 2. DATA:

- As the contractor already made his mind to build warehouse in Scarborough, so for this we will need the data of location coordinates of the Scarborough and its neighbourhoods. Data like latitudes and longitudes of Scarborough and its neighbourhoods.
- Firstly, we will find the neighbourhoods of Scarborough using the list of Postal Codes.

## - SOURCE FOR LIST OF POSTAL CODES :

“[https://en.wikipedia.org/wiki/List\\_of\\_postal\\_codes\\_of\\_Canada:\\_M](https://en.wikipedia.org/wiki/List_of_postal_codes_of_Canada:_M)”

- This website consists of a table that contains all the postal code of Canada with Borough and Neighbourhoods.
- So by using python we will extract the data of that table and clean it to make it ready to use.

# - EXPORTING TABLE DATA FROM WEBSITE INTO DATAFRAME

Article Talk Read Edit View history Search Wikipedia

## List of postal codes of Canada: M

From Wikipedia, the free encyclopedia

This is a list of postal codes in Canada where the first letter is M. Postal codes beginning with M are located within the city of Toronto in the province of Ontario. Only the first three characters are listed, corresponding to the Forward Sortation Area.

Canada Post provides a free postal code look-up tool on its website,<sup>[1]</sup> via its applications for such smartphones as the iPhone and BlackBerry,<sup>[2]</sup> and sells hard-copy directories and CD-ROMs. Many vendors also sell validation tools, which allow customers to properly match addresses and postal codes. Hard-copy directories can also be consulted in all post offices, and some libraries.

### Toronto - FSAs [edit]

Note: There are no rural FSAs in Toronto, hence no postal codes start with M0.

Postcode	Borough	Neighbourhood
M1A	Not assigned	Not assigned
M2A	Not assigned	Not assigned
M3A	North York	Parkwoods
M4A	North York	Victoria Village
M5A	Downtown Toronto	Harbourfront
M5A	Downtown Toronto	Regent Park
M6A	North York	Lawrence Heights
M6A	North York	Lawrence Manor
M7A	Queen's Park	Not assigned
M8A	Not assigned	Not assigned
M9A	Etobicoke	Islington Avenue
M1B	Scarborough	Rouge
M1B	Scarborough	Malvern

-- Website Postal codes Table --

### Extracting data for Postal Codes Table from website data

```
# extracting header and cells data from the table
header = [head.findAll(text=True)[0].strip() for head in table.findAll('th')]
data = [[td.findAll(text=True)[0].strip() for td in tr.findAll('td')] for tr in table.findAll('tr')]
data = [row for row in data if len(row) == 3]

# Make DataFrame
raw_data = pd.DataFrame(data, columns=header)
raw_data.head()
```

	Postcode	Borough	Neighbourhood
0	M1A	Not assigned	Not assigned
1	M2A	Not assigned	Not assigned
2	M3A	North York	Parkwoods
3	M4A	North York	Victoria Village
4	M5A	Downtown Toronto	Harbourfront

-- Dataframe with website table data --

## - SOURCE FOR NEIGHBOURHOOD LOCATIONS DATA:

“[https://cocl.us/Geospatial\\_data](https://cocl.us/Geospatial_data)”

- Then we will get the geolocation data of all the Neighbourhoods in “toronto\_m.geospatial\_data.csv” file from “[https://cocl.us/Geospatial\\_data](https://cocl.us/Geospatial_data)”.
- This file contains Postal codes, latitudes and longitudes data.

### Download geolocation data set for the Postal Codes

```
url = 'https://cocl.us/Geospatial_data'
!wget -q -O 'toronto_m.geospatial_data.csv' url
locs = pd.read_csv(url)
locs.head()
```

'wget' is not recognized as an internal or external command, operable program or batch file.

	Postal Code	Latitude	Longitude
0	M1B	43.806686	-79.194353
1	M1C	43.784535	-79.160497
2	M1E	43.763573	-79.188711
3	M1G	43.770992	-79.216917
4	M1H	43.773136	-79.239476



## - VENUES DATA USING “FOURSQUARE” :

We also need the information of all the venues in each neighbourhood, and that information we can find using ‘Foursquare’. Foursquare will provide almost every information about the venues in the neighbourhood. And after cleaning the data we have data like this.

	Postal Code	Neighbourhood	Latitude	Longitude	Venue	Venue Category	Venue Summary	Distance
0	M1B	Rouge, Malvern	43.806686	-79.194353	Images Salon & Spa	Spa	This spot is popular	595
1	M1B	Rouge, Malvern	43.806686	-79.194353	Caribbean Wave	Caribbean Restaurant	This spot is popular	912
2	M1B	Rouge, Malvern	43.806686	-79.194353	Staples Morningside	Paper / Office Supplies Store	This spot is popular	735
3	M1B	Rouge, Malvern	43.806686	-79.194353	Wendy's	Fast Food Restaurant	This spot is popular	600
4	M1B	Rouge, Malvern	43.806686	-79.194353	Wendy's	Fast Food Restaurant	This spot is popular	387

### 3. METHODOLOGY:

In this section we will discuss about the data analysis that we did on the data extracted from the websites or internet and also which machine learning technique we used to get the best location for the groceries warehouse.

- Filtering and cleaning of data.
- Connecting to Foursquare to retrieve venues data for each neighbourhood.
- Processing and cleaning of retrieved data.
- Machine Learning technique used for finding best location for warehouse.

- Filtering and cleaning of Data:

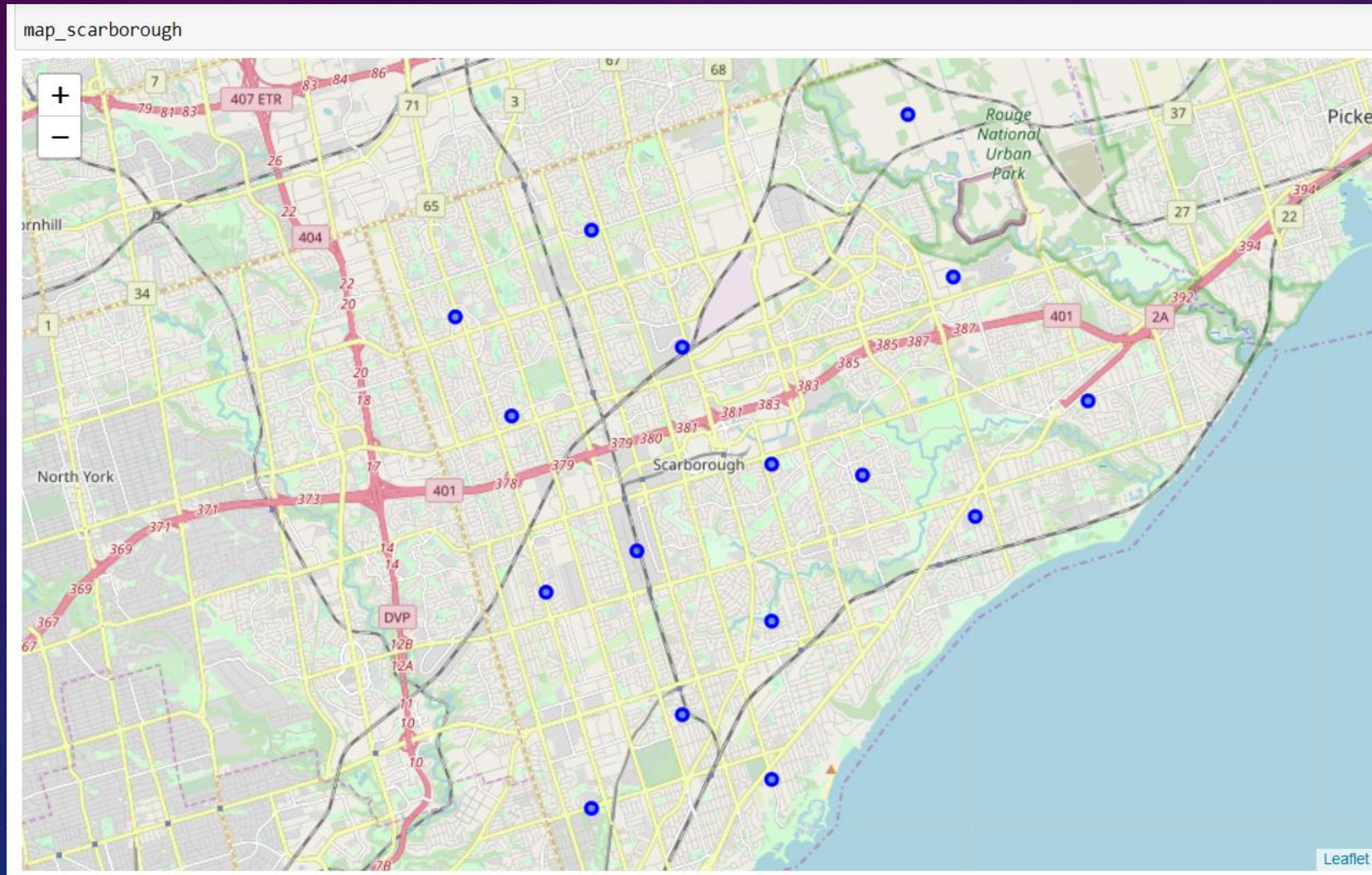
As the contractor already made his mind to build a warehouse in Scarborough, so we only selected the Scarborough data from the dataframe.

```
scarborough = toronto[toronto['Borough'] == 'Scarborough']  
scarborough
```

	Postal Code	Borough	Neighbourhood	Latitude	Longitude
0	M1B	Scarborough	Rouge, Malvern	43.806686	-79.194353
1	M1C	Scarborough	Highland Creek, Rouge Hill, Port Union	43.784535	-79.160497
2	M1E	Scarborough	Guildwood, Morningside, West Hill	43.763573	-79.188711
3	M1G	Scarborough	Woburn	43.770992	-79.216917
4	M1H	Scarborough	Cedarbrae	43.773136	-79.239476
5	M1J	Scarborough	Scarborough Village	43.744734	-79.239476
6	M1K	Scarborough	East Birchmount Park, Ionview, Kennedy Park	43.727929	-79.262029
7	M1L	Scarborough	Clairlea, Golden Mile, Oakridge	43.711112	-79.284577
8	M1M	Scarborough	Cliffcrest, Cliffside, Scarborough Village West	43.716316	-79.239476
9	M1N	Scarborough	Birch Cliff, Cliffside West	43.692657	-79.264848
10	M1P	Scarborough	Dorset Park, Scarborough Town Centre, Wexford ...	43.757410	-79.273304
11	M1R	Scarborough	Maryvale, Wexford	43.750072	-79.295849
12	M1S	Scarborough	Agincourt	43.794200	-79.262029



## - Scarborough Map with its Neighbourhoods





- Connecting to Foursquare to retrieve Venues Data:

Now we need the venues details for each neighbourhood.

After finding the list of neighborhoods, we then connect to the Foursquare API to gather information about venues inside each and every neighbourhood. For each neighbourhood, we have chosen the radius to be 1000 meter. It means that we have asked Foursquare to find venues that are at most 1000 meter far from the center of the neighbourhood. After cleaning, data would look like this.

	Postal Code	Neighbourhood	Latitude	Longitude	Venue	Venue Category	Venue Summary	Distance
0	M1B	Rouge, Malvern	43.806686	-79.194353	Images Salon & Spa	Spa	This spot is popular	595
1	M1B	Rouge, Malvern	43.806686	-79.194353	Caribbean Wave	Caribbean Restaurant	This spot is popular	912
2	M1B	Rouge, Malvern	43.806686	-79.194353	Staples Morningside	Paper / Office Supplies Store	This spot is popular	735
3	M1B	Rouge, Malvern	43.806686	-79.194353	Wendy's	Fast Food Restaurant	This spot is popular	600
4	M1B	Rouge, Malvern	43.806686	-79.194353	Wendy's	Fast Food Restaurant	This spot is popular	387

- One Hot Encoding Process:

Now we will make dataframe of selected venues to whom contractor supplies groceries and then do encoding on “Venue Category” column using OneHot encoding process. This process will make different feature columns of Venue Category. After encoding we will integrate all restaurant columns to “Total Restaurant” column and all joint columns to ‘Total Joint’ column. Now the dataset is ready for the Machine Learning Technique.

	Bakery	Breakfast Spot	Diner	Fish Market	Food & Drink Shop	Fruit & Vegetable Store	Grocery Store	Noodle House	Pizza Place	Sandwich Place	Total Restaurants	Total Joints
Neighbourhood												
Agincourt	2	1	0	0	0	0	0	1	1	1	17	0
Agincourt North, L'Amoreaux East, Milliken, Steeles East	2	0	0	0	0	0	0	2	2	0	10	0
Birch Cliff, Cliffside West	0	0	1	0	0	0	0	0	0	0	4	0
Cedarbrae	4	0	0	0	0	0	1	0	1	0	8	3
Clairlea, Golden Mile, Oakridge	2	0	1	0	0	0	1	0	1	1	3	0
Clarks Corners, Sullivan, Tam O'Shanter	1	0	0	0	0	0	1	1	2	2	12	1

- Machine Learning Technique (KMeans Clustering):

Now using KMeans Clustering, we will make 5 clusters. We selected only 5 clusters because we think 5 clusters are enough for our dataset. Then we will update our dataset with column names and index representing group or cluster. Also we will add new column “Total Sum” that stores the sum of all the venue category values of each group.

```
from sklearn.cluster import KMeans # import KMeans Clustering
```

```
kmeans = KMeans(n_clusters=5, random_state=0).fit(scar_onehot)
```

← # KMeans Clustering

	Bakery	Breakfast Spot	Diner	Fish Market	Food & Drink Shop	Fruit & Vegetable Store	Grocery Store	Noodle House	Pizza Place	Sandwich Place	Total Restaurants	Total Joints	Total Sum
G0	0.000000	0.333333	0.000000	0.000000	0.333333	0.000000	0.000000	0.000000	2.333333	0.333333	2.333333	1.333333	7.000000
G1	1.666667	0.333333	0.000000	0.333333	0.000000	0.000000	1.333333	0.000000	1.333333	0.333333	9.333333	3.000000	17.666667
G2	2.000000	1.000000	0.000000	0.000000	0.000000	0.000000	1.000000	1.000000	2.000000	2.000000	21.000000	0.000000	30.000000
G3	0.714286	0.142857	0.428571	0.000000	0.000000	0.142857	0.571429	0.142857	0.571429	0.714286	4.285714	0.142857	7.857143
G4	1.500000	0.000000	0.000000	0.000000	0.000000	0.000000	0.500000	1.500000	2.000000	1.000000	12.500000	0.500000	19.500000



- Sorting the DataFrame according to 'Total Sum' column.

After sorting the dataframe it shows that neighbourhoods falls under the 'G2' group are the best locations for building new Groceries Warehouse.

```
means = means.sort_values(by='Total Sum', axis=0, ascending=False)
means
```

	Bakery	Breakfast Spot	Diner	Fish Market	Food & Drink Shop	Fruit & Vegetable Store	Grocery Store	Noodle House	Pizza Place	Sandwich Place	Total Restaurants	Total Joints	Total Sum
G2	2.000000	1.000000	0.000000	0.000000	0.000000	0.000000	1.000000	1.000000	2.000000	2.000000	21.000000	0.000000	30.000000
G4	1.500000	0.000000	0.000000	0.000000	0.000000	0.000000	0.500000	1.500000	2.000000	1.000000	12.500000	0.500000	19.500000
G1	1.666667	0.333333	0.000000	0.333333	0.000000	0.000000	1.333333	0.000000	1.333333	0.333333	9.333333	3.000000	17.666667
G3	0.714286	0.142857	0.428571	0.000000	0.000000	0.142857	0.571429	0.142857	0.571429	0.714286	4.285714	0.142857	7.857143
G0	0.000000	0.333333	0.000000	0.000000	0.333333	0.000000	0.000000	0.000000	2.333333	0.333333	2.333333	1.333333	7.000000



- Neighbourhoods with there group number:

```
summary.columns=['Neighbourhood', 'Group']  
print(kmean.labels_)  
summary
```

```
[2 4 3 1 3 4 0 1 3 0 0 3 1 3 3 3]
```

	Neighbourhood	Group
0	Agincourt	2
1	Agincourt North, L'Amoreaux East, Milliken, St...	4
2	Birch Cliff, Cliffside West	3
3	Cedarbrae	1
4	Clairlea, Golden Mile, Oakridge	3
5	Clarks Corners, Sullivan, Tam O'Shanter	4
6	Cliffcrest, Cliffside, Scarborough Village West	0
7	Dorset Park, Scarborough Town Centre, Wexford ...	1
8	East Birchmount Park, Ionview, Kennedy Park	3
9	Guildwood, Morningside, West Hill	0
10	Highland Creek, Rouge Hill, Port Union	0
11	L'Amoreaux West	3
12	Maryvale, Wexford	1
13	Rouge, Malvern	3
14	Scarborough Village	3
15	Woburn	3

## 4. RESULTS:

- After Machine Learning Technique; KMeans Clustering, we got our result.

And we found that Group 'G2' got the highest 'Total sum' value that means it is the nearest neighbourhood to most of the venues and having the highest chances of profit.

Best Neighbourhood is : Agincourt

### The Best Result is : G2

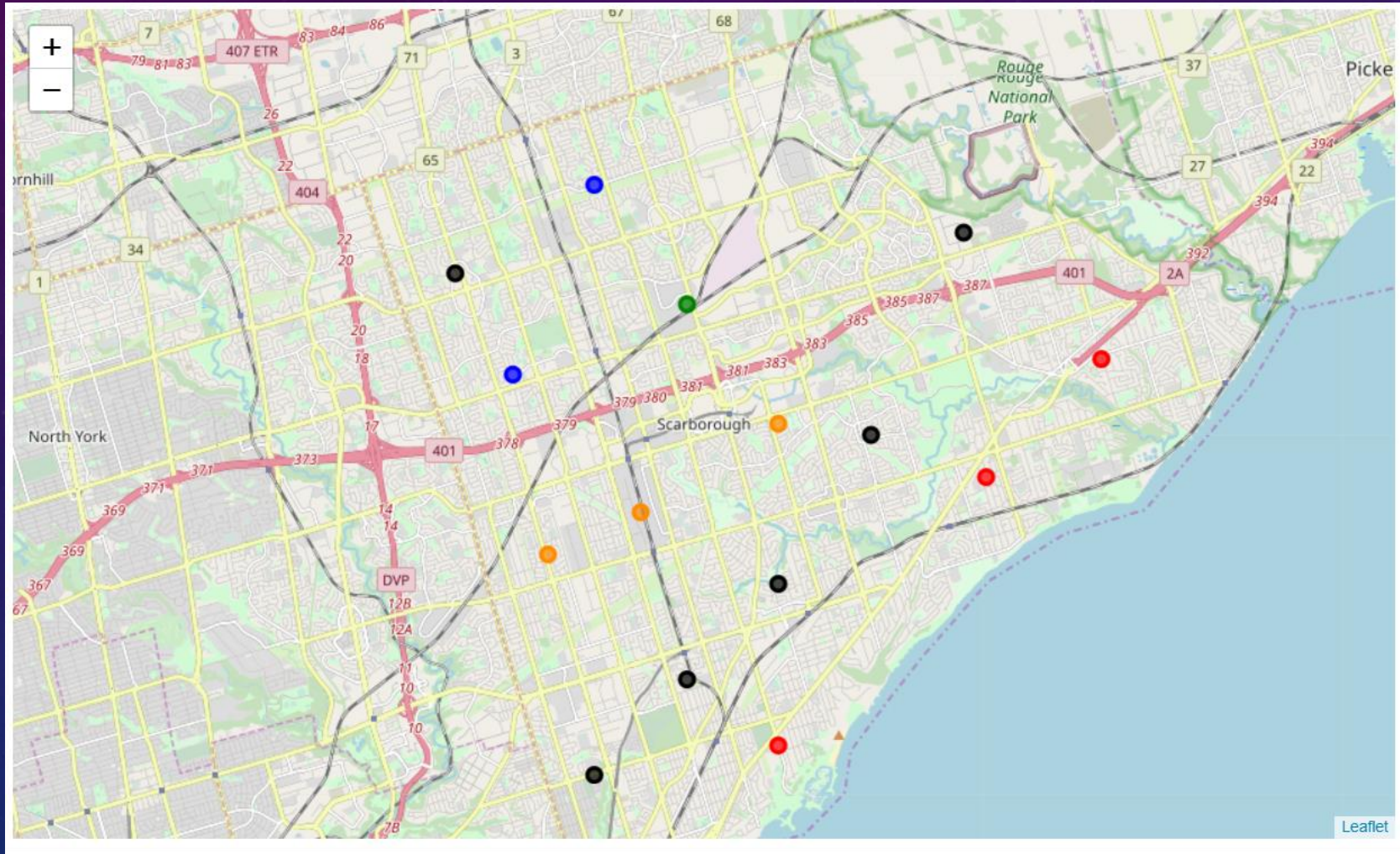
```
best = summary[summary['Group'] == 2]['Neighbourhood'][0]  
print('Best Neighbourhood for building Groceries Warehouse is : ',best)
```

Best Neighbourhood for building Groceries Warehouse is : Agincourt

- Map of Scarborough including its Neighbourhoods in different colors representing different groups.
- 1<sup>st</sup> Best Neighbourhood in Green color.
- 2<sup>nd</sup> Best Neighbourhood in Blue color.
- 3<sup>rd</sup> Best Neighbourhood in Orange color.
- 4<sup>th</sup> Best Neighbourhood in Black color.
- 5<sup>th</sup> Best Neighbourhood in Red color.



- Map of Scarborough with its Neighbourhoods in different colors.





## 5. CONCLUSION:

- **We recommend groceries contractor to build his Groceries Warehouse in “Agincourt” neighbourhood for best results and profit.**

**Best Neighbourhood details:**

**Name:** Agincourt

**Latitude:** 43.7942

**Longitude:** -79.262029



THANK YOU