IBM / Coursera Capstone Project – Week 5 – FINAL REPORT : The Battle of Neighborhoods

Student: Fodyé Cissé

I. Description of the problem: *Finding the best place to open a Pizza restaurant in the city of Toronto.*

Some stakeholders wanted to open a new Pizza chain Restaurants in the city of Toronto. They were looking for the best place to open their first new Pizza restaurant. They know that Toronto is a very dynamic city and that Pizza food is highly appreciated by the populations. They know that there are lots of Pizza restaurants in the city, but they would like to avoid unnecessary competition against existing Pizza places. So, they wanted to open their first Pizza restaurant in locations where there as less pizza restaurants.

As a data scientist, I have been asked by these stakeholders to resolve this business problem for them by finding the right places in the city of Toronto where they should start opening their Pizza restaurant.

II. Description of the Data

To resolve this problem, I will use the following data:

a. The Toronto city data from Wikipedia :
   https://en.wikipedia.org/wiki/List_of_postal_codes_of_Canada:_M
   From this Wikipedia page, I will extract all the Boroughs, Neighborhoods and Postal Codes of the Toronto city required to resolve this problem
b. The Toronto Geospace data: https://cocl.us/Geospatial_data. This is a csv file which contains all the geographical coordinates of all the Toronto Neighborhoods. That will be required to know all the latitudes and longitudes of all the Neighborhoods in Toronto.
c. The Foursquare API: http://developper.foursquare.com . I will use the Foursquare's explorer API tool to collect all the venues, etc nearby each location.

III. Methodology used to resolve the problem

a. First of all, like what we did in week 2 and week 3 labs, I will **collect** and **extract** all the Neighborhoods and boroughs data of Toronto from the Wikipedia page. Then, I will **clean** and **merge** these data with the Geospace data in one single Panda **DataFrame**.
b. Then, I will **explore** the data. I will first display all the Neighborhoods in a **Folium** map. I will use the **Foursquare API** tool to **extract** all the venues. My Foursquare account is a free account, so I will be limited to only 100 venues by the Foursquare API in a radius of 1000 meters. Once I get all the venues, I will **focus on the Pizza Places only**. Continuing the **exploration**, I will **merge** my first DataFrame with the venues returned by Foursquare. I will **analyze** these data focusing on the Pizza Places only. I will use the **matplotlib** library to plot the amount of Pizza Places we have per Borough.
c. And finally, I will use the **unsupervised machine learning algorithm, K-mean clustering**, to cluster the neighborhoods in different clusters of similarities. I will use first the **Elbow** method to determine the best K-mean value to use and will use the **Folium** package to get an overview of the neighborhoods by cluster on the map

The data collected from Wikipedia and Geospace are cleaned and merge into one DataFrame with all the Latitudes and Longitude for each Neighborhoods

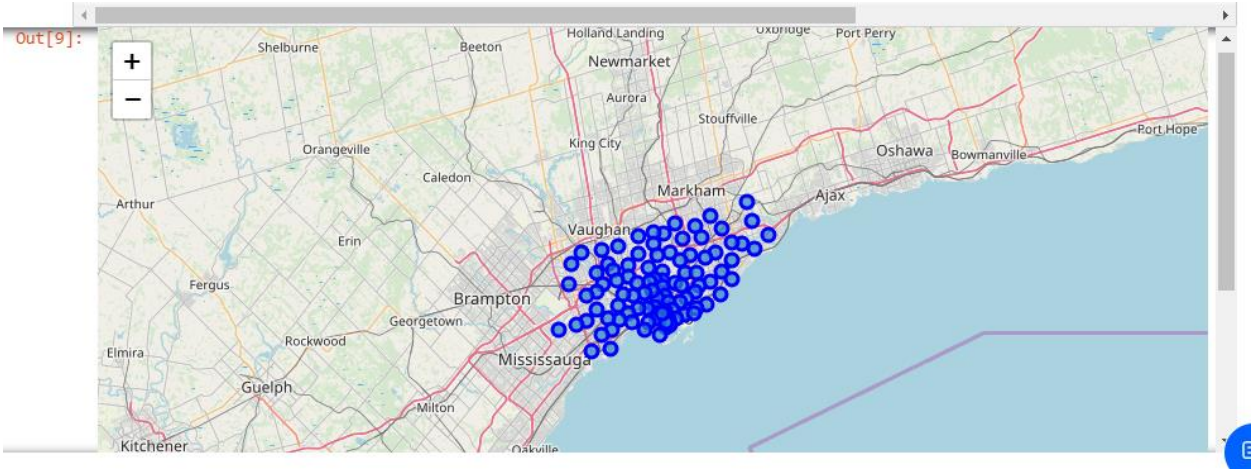
```
# The DataFrames are merged, but now, I need to move colums 'Latitude' and 'Longitude' to the end of the DF.
df_Toronto_Geodata= df_Toronto_Geodata.reindex(columns=['Postalcode','Borough','Neighborhood','Latitude','Longitude'])
df_Toronto_Geodata
```

]:

	Postalcode	Borough	Neighborhood	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

The Neighborhoods in a Folium map

Out[9]:



The venues extracted from the Foursquare API :

```
venues = results['response']['groups'][0]['items']

nearby_venues = json_normalize(venues) # flatten JSON

# filter columns
filtered_columns = ['venue.name', 'venue.categories', 'venue.location.lat', 'venue.location.lng']
nearby_venues =nearby_venues.loc[:, filtered_columns]

# filter the category for each row
nearby_venues['venue.categories'] = nearby_venues.apply(get_category_type, axis=1)

# clean columns
nearby_venues.columns = [col.split(".")[-1] for col in nearby_venues.columns]

nearby_venues.head()
```

	name	categories	lat	lng
0	Downtown Toronto	Neighborhood	43.653232	-79.385296
1	Japango	Sushi Restaurant	43.655268	-79.385165
2	Rolltation	Japanese Restaurant	43.654918	-79.387424
3	Sansotei Ramen 三草亭	Ramen Restaurant	43.655157	-79.386501
4	Nathan Phillips Square	Plaza	43.652270	-79.383516

Extracting the Pizza Place Data from the Data:

Extract Toronto Pizza Place Data

Entrée [27]:

```
toronto_PizzaPlace = toronto_grouped[["Neighborhood","Pizza Place"]]
toronto_PizzaPlace.head()
```

Out[27]:

	Neighborhood	Pizza Place
0	Adelaide, King, Richmond	0.040000
1	Agincourt	0.016129
2	Agincourt North, L'Amoreaux East, Milliken, St...	0.041667
3	Albion Gardens, Beaumond Heights, Humbergate, ...	0.088235
4	Alderwood, Long Branch	0.065217

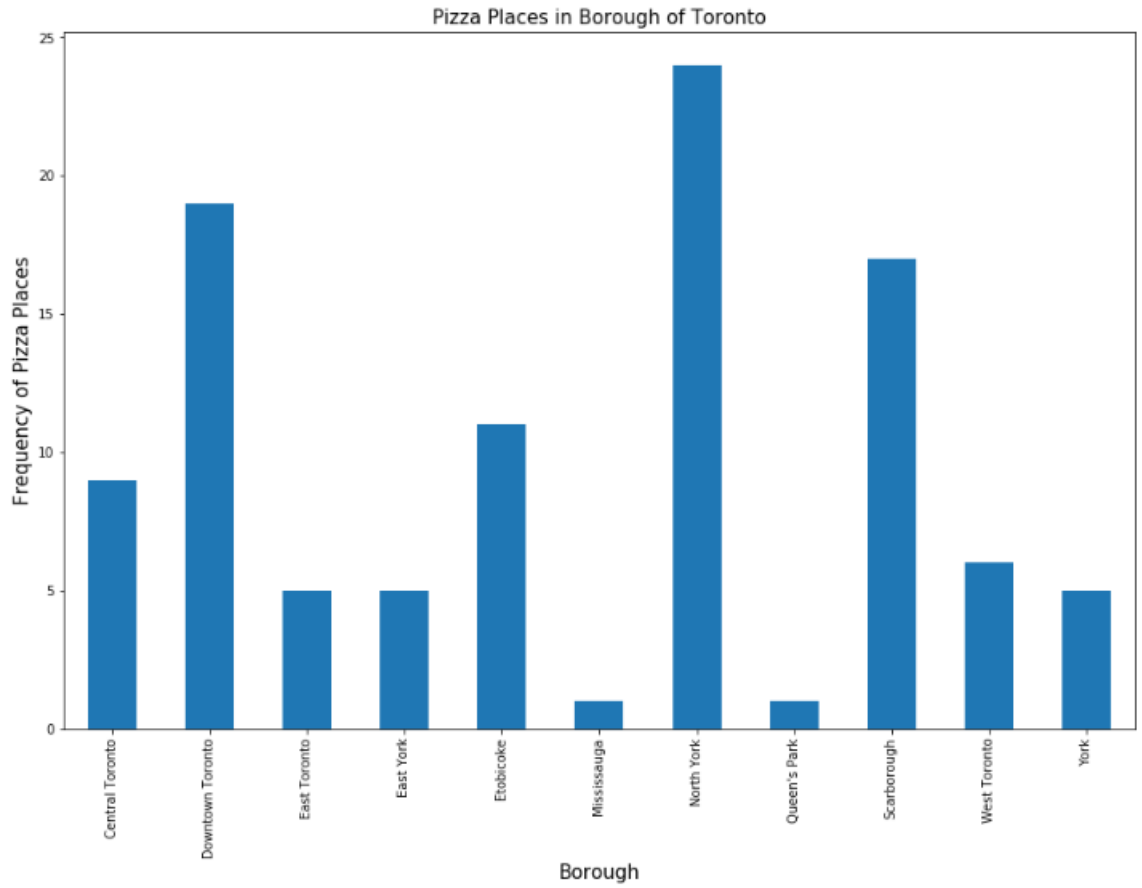
All the Pizza Place data merged with the Latitudes and Longitudes of all the Neighborhoods:

```
# Now let's merge the Toronto Geodat and the Toronto Pizza Places data per Neighborhood
toronto_PizzaPlace_merged = pd.merge(df_Toronto_Geodata, toronto_PizzaPlace, on='Neighborhood')

# Now, let's change the columns order
toronto_PizzaPlace_merged = toronto_PizzaPlace_merged.reindex(columns=['Postalcode','Borough','Neighborhood','Pizza Pl

toronto_PizzaPlace_merged.head()
```

:

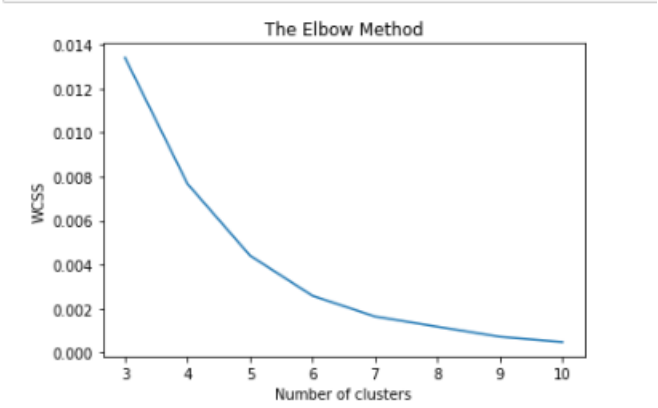
	Postalcode	Borough	Neighborhood	Pizza Place	Latitude	Longitude
0	M1B	Scarborough	Rouge, Malvern	0.058824	43.806686	-79.194353
1	M1C	Scarborough	Highland Creek, Rouge Hill, Port Union	0.000000	43.784535	-79.160497
2	M1E	Scarborough	Guildwood, Morningside, West Hill	0.121212	43.763573	-79.188711
3	M1G	Scarborough	Woburn	0.093750	43.770992	-79.216917
4	M1H	Scarborough	Cedarbrae	0.015152	43.773136	-79.239476

How many Pizza Places we have per Borough?



CLUSTERING:

Using the **Elbow** method to determine the best K value:



From this map, we can see that 6 is the best K value.

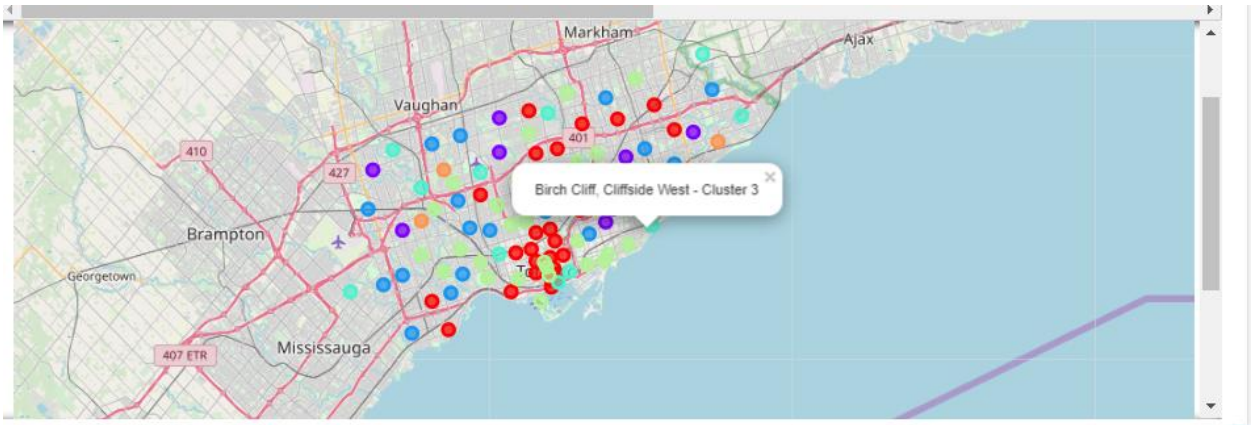
The Toronto Geodata, Pizza places data per Neighborhood merged with the Cluster labels

```
# Now let's merge the Toronto Geodata and the Toronto Pizza Places data per Neighborhood with Cluster labels
toronto_PizzaPlace_Clustering_merged2 = pd.merge(df_Toronto_Geodata, toronto_PizzaPlace_Clustering_merged[['Neighborhood','Pizza Place

toronto_PizzaPlace_Clustering_merged2.head()
```

:

	Postalcode	Latitude	Longitude	Borough	Neighborhood	Pizza Place	Cluster Labels
0	M1B	43.806686	-79.194353	Scarborough	Rouge, Malvern	0.058824	2
1	M1C	43.784535	-79.160497	Scarborough	Highland Creek, Rouge Hill, Port Union	0.000000	3
2	M1E	43.763573	-79.188711	Scarborough	Guildwood, Morningside, West Hill	0.121212	5
3	M1G	43.770992	-79.216917	Scarborough	Woburn	0.093750	1
4	M1H	43.773136	-79.239476	Scarborough	Cedarbrae	0.015152	0

Folium map with the Neighborhoods displayed by clusters :



Cluster 0:

```
# Cluster 0
toronto_PizzaPlace_Cluster0 = toronto_PizzaPlace_Clustering_merged2.loc[toronto_PizzaPlace_Clustering_merged2['Cluster Labels'] == 0]
print(toronto_PizzaPlace_Cluster0.shape)
toronto_PizzaPlace_Cluster0
```

```
(33, 7)
```

	Postalcode	Latitude	Longitude	Borough	Neighborhood	Pizza Place	Cluster Labels
51	M4X	43.667967	-79.367675	Downtown Toronto	Cabbagetown, St. James Town	0.010000	0
27	M3C	43.725900	-79.340923	North York	Flemingdon Park, Don Mills South	0.012346	0
71	M6A	43.718518	-79.464763	North York	Lawrence Heights, Lawrence Manor	0.010000	0
70	M5X	43.648429	-79.382280	Downtown Toronto	First Canadian Place, Underground city	0.010000	0
69	M5W	43.646435	-79.374846	Downtown Toronto	Stn A PO Boxes 25 The Esplanade	0.010000	0
67	M5T	43.653206	-79.400049	Downtown Toronto	Chinatown, Grange Park, Kensington Market	0.020000	0
66	M5S	43.662696	-79.400049	Downtown Toronto	Harbord, University of Toronto	0.010000	0
34	M4A	43.725882	-79.315572	North York	Victoria Village	0.017544	0
65	M5R	43.672710	-79.405678	Central Toronto	The Annex, North Midtown, Yorkville	0.020000	0
38	M4G	43.709060	-79.363452	East York	Leaside	0.013333	0
50	M4W	43.679563	-79.377529	Downtown Toronto	Rosedale	0.010000	0
39	M4H	43.705369	-79.349372	East York	Thorncliffe Park	0.010638	0
60	M5K	43.647177	-79.381576	Downtown Toronto	Design Exchange, Toronto Dominion Centre	0.010000	0
59	M5J	43.640816	-79.381752	Downtown Toronto	Harbourfront East, Toronto Islands, Union Station	0.010000	0

Cluster 1:

```
# Cluster 1
toronto_PizzaPlace_Cluster1 = toronto_PizzaPlace_Clustering_merged2.loc[toronto_PizzaPlace_Clustering_merged2['Cluster Labels'] == 1]
print(toronto_PizzaPlace_Cluster1.shape)
toronto_PizzaPlace_Cluster1
```

```
(11, 7)
```

	Postalcode	Latitude	Longitude	Borough	Neighborhood	Pizza Place	Cluster Labels
100	M9R	43.688905	-79.554724	Etobicoke	Kingsview Village, Martin Grove Gardens, Richv...	0.095238	1
3	M1G	43.770992	-79.216917	Scarborough	Woburn	0.093750	1
6	M1K	43.727929	-79.262029	Scarborough	East Birchmount Park, Ionview, Kennedy Park	0.076923	1
7	M1L	43.711112	-79.284577	Scarborough	Clairlea, Golden Mile, Oakridge	0.085714	1
8	M1M	43.716316	-79.239476	Scarborough	Cliffcrest, Cliffside, Scarborough Village West	0.073171	1
11	M1R	43.750072	-79.295849	Scarborough	Maryvale, Wexford	0.081633	1
24	M2R	43.782736	-79.442259	North York	Willowdale West	0.076923	1
35	M4B	43.706397	-79.309937	East York	Woodbine Gardens, Parkview Hill	0.076923	1
28	M3H	43.754328	-79.442259	North York	Bathurst Manor, Downsview North, Wilson Heights	0.075000	1
36	M4C	43.695344	-79.318389	East York	Woodbine Heights	0.085714	1
101	M9V	43.739416	-79.588437	Etobicoke	Albion Gardens, Beaumond Heights, Humbergate, ...	0.088235	1

Cluster 2:

```
# Cluster 2
toronto_PizzaPlace_Cluster2 = toronto_PizzaPlace_Clustering_merged2.loc[toronto_PizzaPlace_Clustering_merged2['Cluster Labels'] == 2]
print(toronto_PizzaPlace_Cluster2.shape)
toronto_PizzaPlace_Cluster2
```

```
(17, 7)
```

	Postalcode	Latitude	Longitude	Borough	Neighborhood	Pizza Place	Cluster Labels
92	M8Y	43.636258	-79.498509	Etobicoke	Humber Bay, King's Mill Park, Kingsway Park So...	0.060606	2
90	M8W	43.602414	-79.543484	Etobicoke	Alderwood, Long Branch	0.065217	2
84	M6S	43.651571	-79.484450	West Toronto	Runnymede, Swansea	0.055556	2
80	M6M	43.691116	-79.476013	York	Del Ray, Keelesdale, Mount Dennis, Silverthorn	0.055556	2
94	M9B	43.650943	-79.554724	Etobicoke	Cloverdale, Islington, Martin Grove, Princess ...	0.047619	2
79	M6L	43.713756	-79.490074	North York	Downsview, North Park, Upwood Park	0.060606	2
95	M9C	43.643515	-79.577201	Etobicoke	Bloordale Gardens, Eringate, Markland Wood, Ol...	0.054054	2
74	M6E	43.689026	-79.453512	York	Caledonia-Fairbanks	0.057143	2
0	M1B	43.806686	-79.194353	Scarborough	Rouge, Malvern	0.058824	2
102	M9W	43.706748	-79.594054	Etobicoke	Northwest	0.055556	2
33	M3N	43.761631	-79.520999	North York	Downsview Northwest	0.066667	2

Cluster 3:

```
# Cluster 3
toronto_PizzaPlace_Cluster3 = toronto_PizzaPlace_Clustering_merged2.loc[toronto_PizzaPlace_Clustering_merged2['Cluster Labels'] == 3]
print(toronto_PizzaPlace_Cluster3.shape)
toronto_PizzaPlace_Cluster3
```

```
(11, 7)
```

	Postalcode	Latitude	Longitude	Borough	Neighborhood	Pizza Place	Cluster Labels
96	M9L	43.756303	-79.565963	North York	Humber Summit	0.0	3
87	M7R	43.636966	-79.615819	Mississauga	Canada Post Gateway Processing Centre	0.0	3
16	M1X	43.836125	-79.205636	Scarborough	Upper Rouge	0.0	3
19	M2K	43.786947	-79.385975	North York	Bayview Village	0.0	3
97	M9M	43.724766	-79.532242	North York	Emery, Humberlea	0.0	3
1	M1C	43.784535	-79.160497	Scarborough	Highland Creek, Rouge Hill, Port Union	0.0	3
76	M6H	43.669005	-79.442259	West Toronto	Dovercourt Village, Dufferin	0.0	3
9	M1N	43.692657	-79.264848	Scarborough	Birch Cliff, Cliffside West	0.0	3
30	M3K	43.737473	-79.464763	North York	CFB Toronto, Downsview East	0.0	3
53	M5A	43.654260	-79.360636	Downtown Toronto	Harbourfront	0.0	3
56	M5E	43.644771	-79.373306	Downtown Toronto	Berczy Park	0.0	3

Cluster 4:

```
# Cluster 4
toronto_PizzaPlace_Cluster4 = toronto_PizzaPlace_Clustering_merged2.loc[toronto_PizzaPlace_Clustering_merged2['Cluster Labels'] == 4]
print(toronto_PizzaPlace_Cluster4.shape)
toronto_PizzaPlace_Cluster4
```

```
(28, 7)
```

	Postalcode	Latitude	Longitude	Borough	Neighborhood	Pizza Place	Cluster Labels
68	M5V	43.628947	-79.394420	Downtown Toronto	CN Tower, Bathurst Quay, Island airport, Harbo...	0.028571	4
45	M4P	43.712751	-79.390197	Central Toronto	Davisville North	0.030303	4
98	M9N	43.706876	-79.518188	York	Weston	0.041667	4
43	M4M	43.659526	-79.340923	East Toronto	Studio District	0.030000	4
57	M5G	43.657952	-79.387383	Downtown Toronto	Central Bay Street	0.030000	4
58	M5H	43.650571	-79.384568	Downtown Toronto	Adelaide, King, Richmond	0.040000	4
42	M4L	43.668999	-79.315572	East Toronto	The Beaches West, India Bazaar	0.030000	4
91	M8X	43.653654	-79.506944	Etobicoke	The Kingsway, Montgomery Road, Old Mill North	0.033333	4
14	M1V	43.815252	-79.284577	Scarborough	Agincourt North, L'Amoreaux East, Milliken, St...	0.041667	4
41	M4K	43.679557	-79.352188	East Toronto	The Danforth West, Riverdale	0.043956	4
88	M7Y	43.662744	-79.321558	East Toronto	Business Reply Mail Processing Centre 969 Eastern	0.040000	4
62	M5M	43.733283	-79.419750	North York	Bedford Park, Lawrence Manor East	0.039474	4
85	M7A	43.662301	-79.389494	Downtown Toronto	Queen's Park	0.026316	4
86	M9A	43.667856	-79.532242	Queen's Park	Queen's Park	0.026316	4
32	M3M	43.728496	-79.495697	North York	Downsview Central	0.028571	4

Cluster 5:

```
# Cluster 5
toronto_PizzaPlace_Cluster5 = toronto_PizzaPlace_Clustering_merged2.loc[toronto_PizzaPlace_Clustering_merged2['Cluster Labels'] == 5]
print(toronto_PizzaPlace_Cluster5.shape)
toronto_PizzaPlace_Cluster5
```

```
(3, 7)
```

	Postalcode	Latitude	Longitude	Borough	Neighborhood	Pizza Place	Cluster Labels
2	M1E	43.763573	-79.188711	Scarborough	Guildwood, Morningside, West Hill	0.121212	5
99	M9P	43.696319	-79.532242	Etobicoke	Westmount	0.125000	5
31	M3L	43.739015	-79.506944	North York	Downsview West	0.125000	5

IV. Conclusion

To recap:

The stakeholders wanted to avoid unnecessary competition against existing Pizza places. So, they wanted to open their first Pizza restaurant in locations where there as less pizza restaurants.

We can see from the K-Mean clustering that Cluster 3 is the cluster which has less Pizza Places.

So, the best places to open their first Pizza Restaurant are the Neighborhoods from **Cluster 3. In the Folium map, Cluster 3 Neighborhoods are neighborhoods displayed in the map in Cyan color.**

```
# Cluster 3
toronto_PizzaPlace_Cluster3 = toronto_PizzaPlace_Clustering_merged2.loc[toronto_PizzaPlace_Clustering_merged2['Cluster Labels'] == 3]
print(toronto_PizzaPlace_Cluster3.shape)
toronto_PizzaPlace_Cluster3
```

```
(11, 7)
```

:

	Postalcode	Latitude	Longitude	Borough	Neighborhood	Pizza Place	Cluster Labels
96	M9L	43.756303	-79.565963	North York	Humber Summit	0.0	3
87	M7R	43.636966	-79.615819	Mississauga	Canada Post Gateway Processing Centre	0.0	3
16	M1X	43.836125	-79.205636	Scarborough	Upper Rouge	0.0	3
19	M2K	43.786947	-79.385975	North York	Bayview Village	0.0	3
97	M9M	43.724766	-79.532242	North York	Emery, Humberlea	0.0	3
1	M1C	43.784535	-79.160497	Scarborough	Highland Creek, Rouge Hill, Port Union	0.0	3
76	M6H	43.669005	-79.442259	West Toronto	Dovercourt Village, Dufferin	0.0	3
9	M1N	43.692657	-79.264848	Scarborough	Birch Cliff, Cliffside West	0.0	3
30	M3K	43.737473	-79.464763	North York	CFB Toronto, Downsview East	0.0	3
53	M5A	43.654260	-79.360636	Downtown Toronto	Harbourfront	0.0	3
56	M5E	43.644771	-79.373306	Downtown Toronto	Berczy Park	0.0	3