

# This project was made for IBM course on data science.

## Were you will be, my coffee shop?

Here we will find a suitable place for our business coffee shop.

## 1 – Introduction

Toronto is a large city and therefore it has business activity. Accordingly, it would be good to give people the opportunity to discuss business over a cup of coffee at our coffee shop.

## 2 – Data

We will use the data from Wikipedia about Toronto's districts, find the coordinates to them, and use the Foursquare requests to get data about these places. And then we'll work with them with clustering, and choose the best place!

## 3 – Methodology

I will use the K-means clustering method, and I will configure it using the Elbow method to select the best number of clusters. Then I'll look at what features the clusters have and compare them.

---

I take the mailing code data from this [link](#).

I do it with the help of pandas, as a result the list of dataframes will be returned to us with all tables from the specified page, and having chosen the necessary one we will get this:

	Postal Code	Borough		Neighborhood
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		Regent Park, Harbourfront
...	...	...		...
175	M5Z	Not assigned		Not assigned
176	M6Z	Not assigned		Not assigned
177	M7Z	Not assigned		Not assigned
178	M8Z	Etobicoke	Mimico NW, The Queensway West, South of Bloor,...	
179	M9Z	Not assigned		Not assigned

Then filter by 'Not assigned' items.

	Postal Code	Borough	Neighborhood
2	M3A	North York	Parkwoods
3	M4A	North York	Victoria Village
4	M5A	Downtown Toronto	Regent Park, Harbourfront
5	M6A	North York	Lawrence Manor, Lawrence Heights
6	M7A	Downtown Toronto	Queen's Park, Ontario Provincial Government
...	...	...	...
160	M8X	Etobicoke	The Kingsway, Montgomery Road, Old Mill North
165	M4Y	Downtown Toronto	Church and Wellesley
168	M7Y	East Toronto	Business reply mail Processing Centre, South C...
169	M8Y	Etobicoke	Old Mill South, King's Mill Park, Sunnylea, Hu...
178	M8Z	Etobicoke	Mimico NW, The Queensway West, South of Bloor,...

I also have Geospatial\_Coordinates.csv

	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
...	...	...	...
98	M9N	43.706876	-79.518188
99	M9P	43.696319	-79.532242
100	M9R	43.688905	-79.554724
101	M9V	43.739416	-79.588437
102	M9W	43.706748	-79.594054

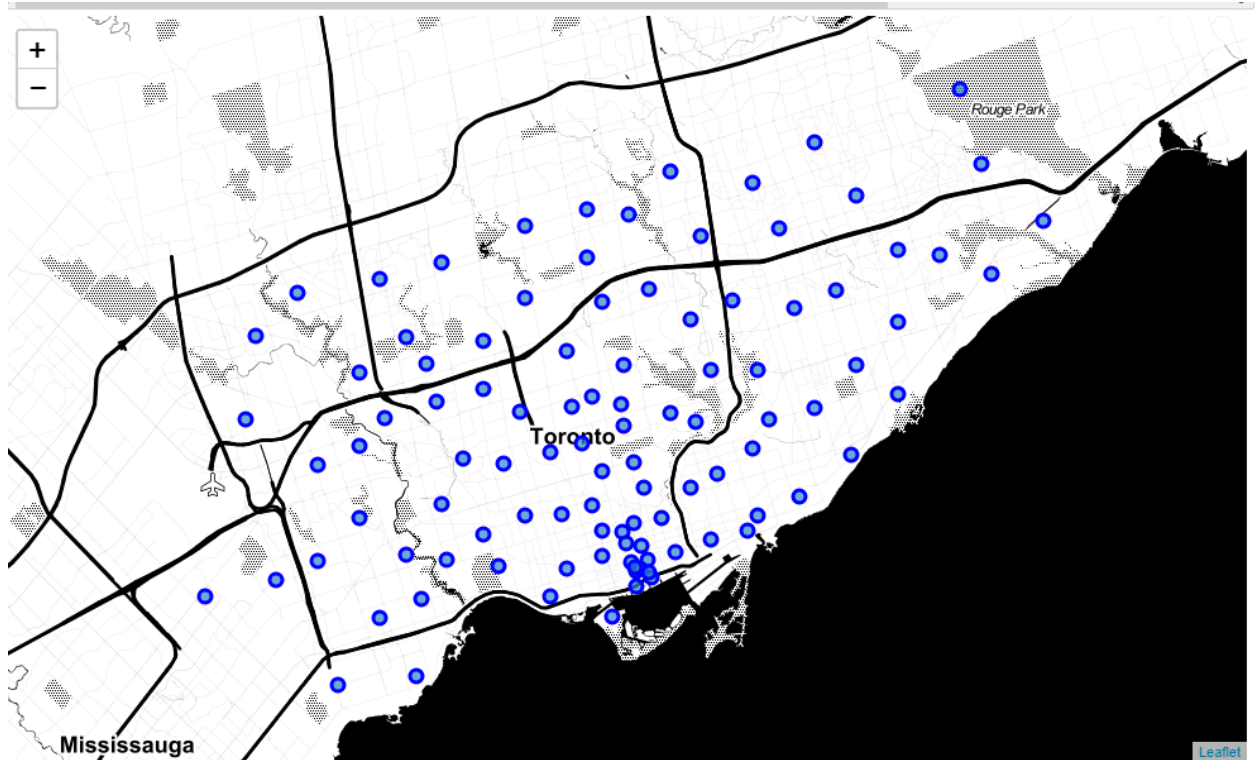
I'll take the coordinates from it and create a common dataframe.

	Postal Code	Borough	Neighborhood	Latitude	Longitude
0	M3A	North York	Parkwoods	43.753259	-79.329656
1	M4A	North York	Victoria Village	43.725882	-79.315572
2	M5A	Downtown Toronto	Regent Park, Harbourfront	43.654260	-79.360636
3	M6A	North York	Lawrence Manor, Lawrence Heights	43.718518	-79.464763
4	M7A	Downtown Toronto	Queen's Park, Ontario Provincial Government	43.662301	-79.389494
...	...	...	...	...	...
98	M8X	Etobicoke	The Kingsway, Montgomery Road, Old Mill North	43.653654	-79.506944
99	M4Y	Downtown Toronto	Church and Wellesley	43.665860	-79.383160
100	M7Y	East Toronto	Business reply mail Processing Centre, South C...	43.662744	-79.321558
101	M8Y	Etobicoke	Old Mill South, King's Mill Park, Sunnylea, Hu...	43.636258	-79.498509
102	M8Z	Etobicoke	Mimico NW, The Queensway West, South of Bloor,...	43.628841	-79.520999

I set the Nominatim geocoder and get the coordinates of the Toronto.

The geographical coordinate of Toronto are 43.6534817, -79.3839347.

Then I map Toronto and add district marker to it.



Then I get the name of the first neighborhood, its coordinates, and on them make a query on Foursquare to find out venues within a radius of 500 meters. The answer comes in the form of json.

```
{
  'meta': {
    'code': 200,
    'requestId': '5eea4215c6fd9436f69acb2a',
    'response': {
      'warning': {
        'text': "There aren't a lot of results near you. Try something more general, reset your filters, or expand the search area."
      }
    }
  },
  'headerLocation': 'Parkwoods - Donalda',
  'headerFullLocation': 'Parkwoods - Donalda, Toronto',
  'headerLocationGranularity': 'neighborhood',
  'totalResults': 2,
  'suggestedBounds': {
    'ne': {
      'lat': 43.757758604500005,
      'lng': -79.32343823984928
    },
    'sw': {
      'lat': 43.7487585955,
      'lng': -79.33587476015072
    }
  },
  'groups': [
    {
      'type': 'Recommended Places',
      'name': 'recommended',
      'items': [
        {
          'reasons': {
            'count': 0,
            'items': [
              {
                'summary': 'This spot is popular',
                'type': 'general',
                'reasonName': 'globalInteractionReason'
              }
            ]
          },
          'venue': {
            'id': '4e8d9dcd5fbbb6b3003c7b',
            'name': 'Brookbanks Park',
            'location': {
              'address': 'Toronto',
              'lat': 43.751976046055574,
              'lng': -79.33214044722958,
              'labeledLatLngs': [
                {
                  'label': 'display',
                  'lat': 43.751976046055574,
                  'lng': -79.33214044722958
                }
              ],
              'distance': 245,
              'cc': 'CA',
              'city': 'Toronto',
              'state': 'ON',
              'country': 'Canada',
              'formattedAddress': [
                'Toronto', 'Toronto ON', 'Canada'
              ]
            },
            'categories': [
              {
                'id': '4bf58dd8d4898d163941735',
                'name': 'Park',
                'pluralName': 'Parks',
                'shortName': 'Park',
                'icon': {
                  'prefix': 'https://ss3.4sqi.net/img/categories_v2/parks_outdoors/park_',
                  'suffix': '.png',
                  'primary': true
                },
                'photos': {
                  'count': 0,
                  'groups': []
                },
                'referralId': 'e-0-4e8d9dcd5fbbb6b3003c7b-0'
              }
            ],
            'reasons': {
              'count': 0,
              'items': [
                {
                  'summary': 'This spot is popular',

```

Then I'm processing it to get the pandas dataframe.

	name	categories	lat	lng
0	Brookbanks Park	Park	43.751976	-79.332140
1	Variety Store	Food & Drink Shop	43.751974	-79.333114

Then I do the same for every neighborhood and I group them.

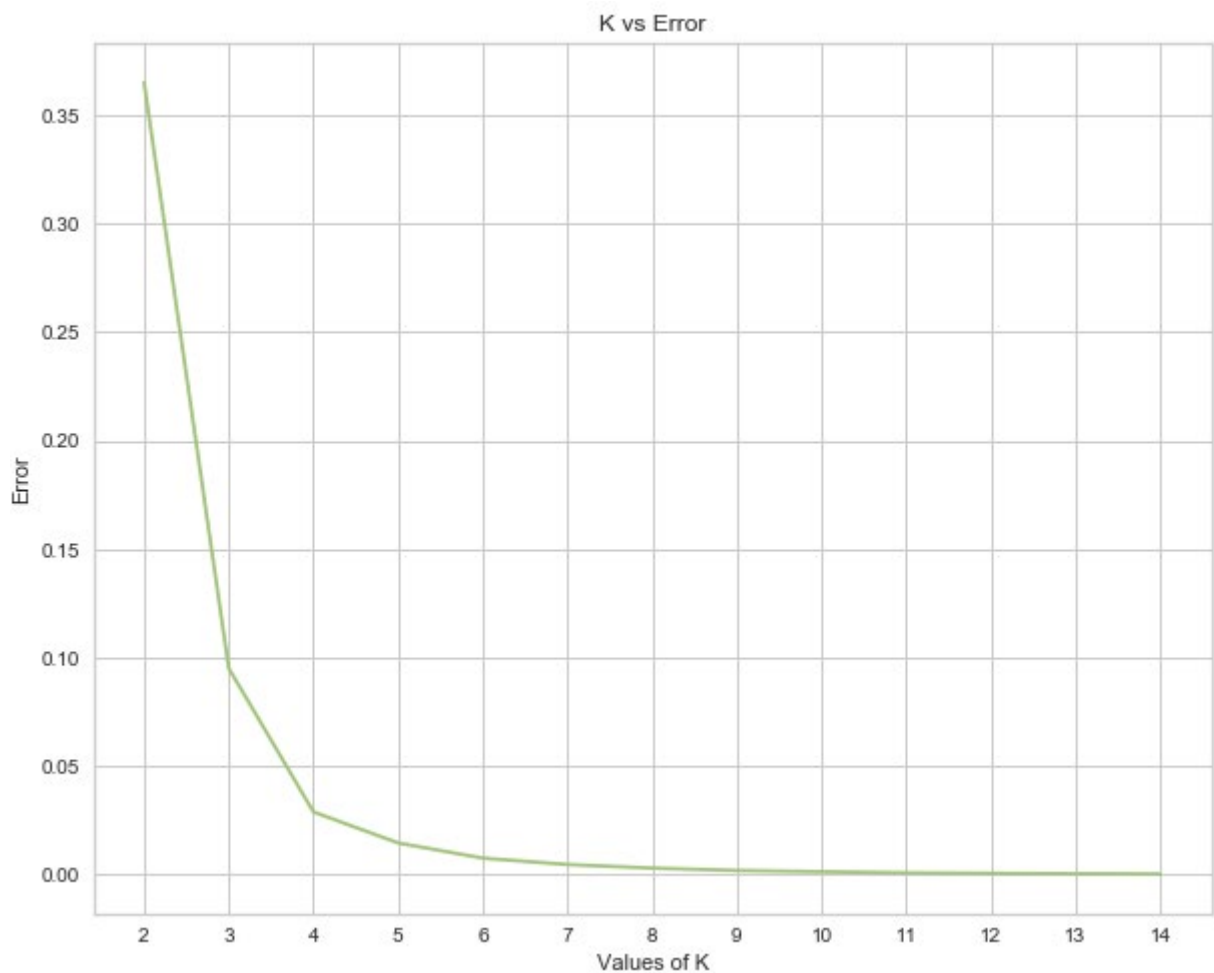
	Neighborhood	Neighborhood Latitude	Neighborhood Longitude	Venue	Venue Latitude	Venue Longitude	Venue Category
0	Parkwoods	43.753259	-79.329656	Brookbanks Park	43.751976	-79.332140	Park
1	Parkwoods	43.753259	-79.329656	Variety Store	43.751974	-79.333114	Food & Drink Shop
2	Victoria Village	43.725882	-79.315572	Victoria Village Arena	43.723481	-79.315635	Hockey Arena
3	Victoria Village	43.725882	-79.315572	Portugril	43.725819	-79.312785	Portuguese Restaurant
4	Victoria Village	43.725882	-79.315572	Tim Hortons	43.725517	-79.313103	Coffee Shop

	Neighborhood	Neighborhood Latitude	Neighborhood Longitude	Venue	Venue Latitude	Venue Longitude	Venue Category
	Agincourt	5	5	5	5	5	5
	Alderwood, Long Branch	8	8	8	8	8	8
	Bathurst Manor, Wilson Heights, Downsview North	20	20	20	20	20	20
	Bayview Village	4	4	4	4	4	4
	Bedford Park, Lawrence Manor East	23	23	23	23	23	23
	...	...	...	...	...	...	...
	Willowdale, Willowdale West	7	7	7	7	7	7
	Woburn	3	3	3	3	3	3
	Woodbine Heights	9	9	9	9	9	9
	York Mills West	4	4	4	4	4	4
	York Mills, Silver Hills	2	2	2	2	2	2

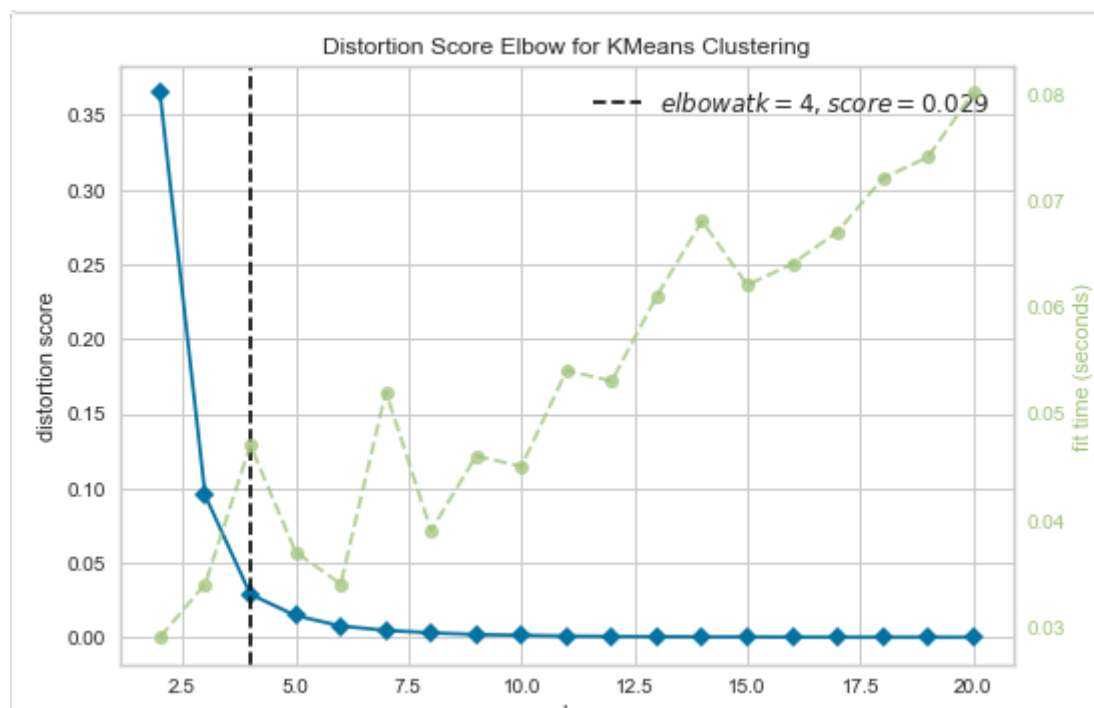
Then I look for unique categories, find among them coffee shops and group them.

	Neighborhoods	Coffee Shop
0	Agincourt	0.000000
1	Alderwood, Long Branch	0.125000
2	Bathurst Manor, Wilson Heights, Downsview North	0.100000
3	Bayview Village	0.000000
4	Bedford Park, Lawrence Manor East	0.086957

For better count of clusters we will use Elbow method.



While the number of clusters may seem obvious, it is better to check this with the KElbowVisualizer.

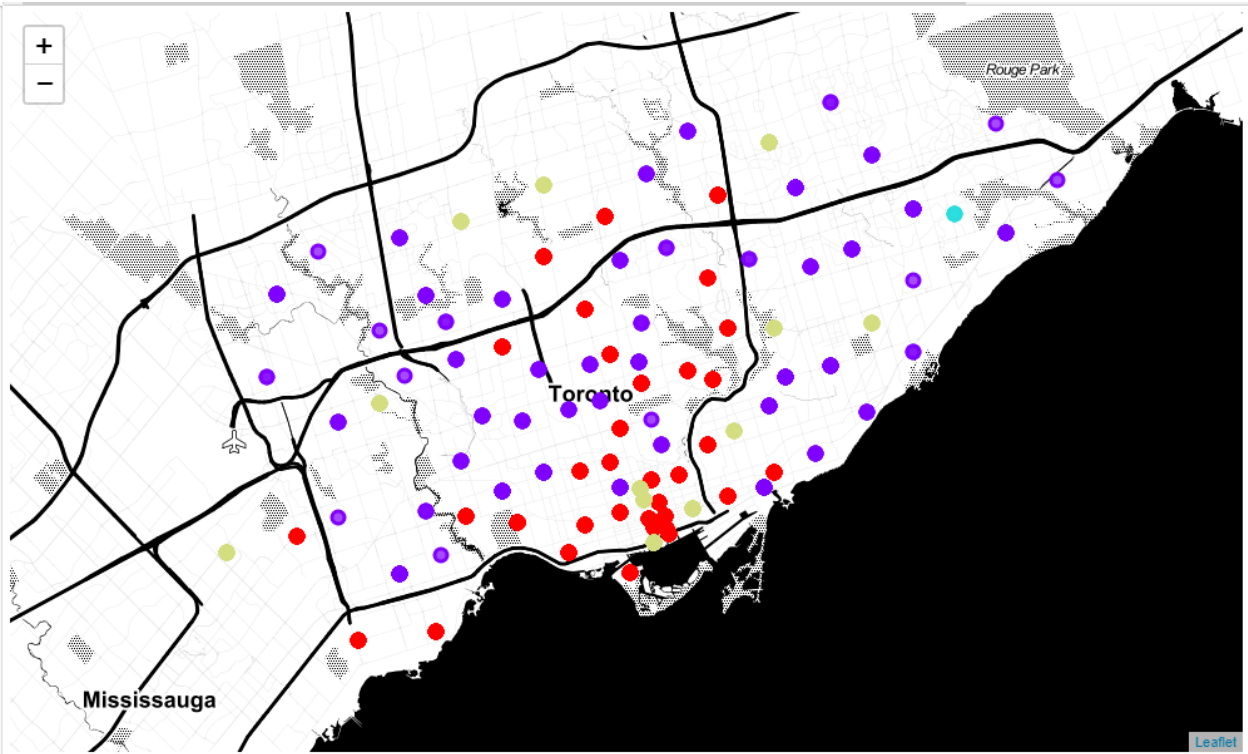


That's right, the best number of clusters for us is 4.

Then I run K-means, collect a dataframe with the necessary information and a cluster mark.

	Neighborhoods	Coffee Shop	Cluster Labels	Neighborhood Latitude	Neighborhood Longitude	Venue	Venue Latitude	Venue Longitude	Venue Category
31	Garden District, Ryerson	0.080000	0	43.657162	-79.378937	Salad King	43.657601	-79.381620	Thai Restaurant
75	Stn A PO Boxes	0.103093	0	43.646435	-79.374846	NAMI	43.650853	-79.375887	Japanese Restaurant
75	Stn A PO Boxes	0.103093	0	43.646435	-79.374846	GoodLife Fitness Toronto Street Women's Only	43.650400	-79.376700	Gym
75	Stn A PO Boxes	0.103093	0	43.646435	-79.374846	St. Lawrence Antique Market	43.649615	-79.371747	Antique Shop
75	Stn A PO Boxes	0.103093	0	43.646435	-79.374846	Indigospirit	43.648350	-79.380347	Bookstore
...	...	...	...	...	...	...	...	...	...
35	Harbourfront East, Union Station, Toronto Islands	0.130000	3	43.640816	-79.381752	Dollarama	43.641765	-79.379589	Discount Store
35	Harbourfront East, Union Station, Toronto Islands	0.130000	3	43.640816	-79.381752	Starbucks	43.643090	-79.383071	Coffee Shop
35	Harbourfront East, Union Station, Toronto Islands	0.130000	3	43.640816	-79.381752	Redpath Stage	43.638764	-79.383317	Music Venue
35	Harbourfront East, Union Station, Toronto Islands	0.130000	3	43.640816	-79.381752	Rainbow Reef	43.642260	-79.385994	Aquarium
42	Kennedy Park, Ionview, East Birchmount Park	0.166667	3	43.727929	-79.262029	Kennedy GO Station	43.732275	-79.262418	Train Station

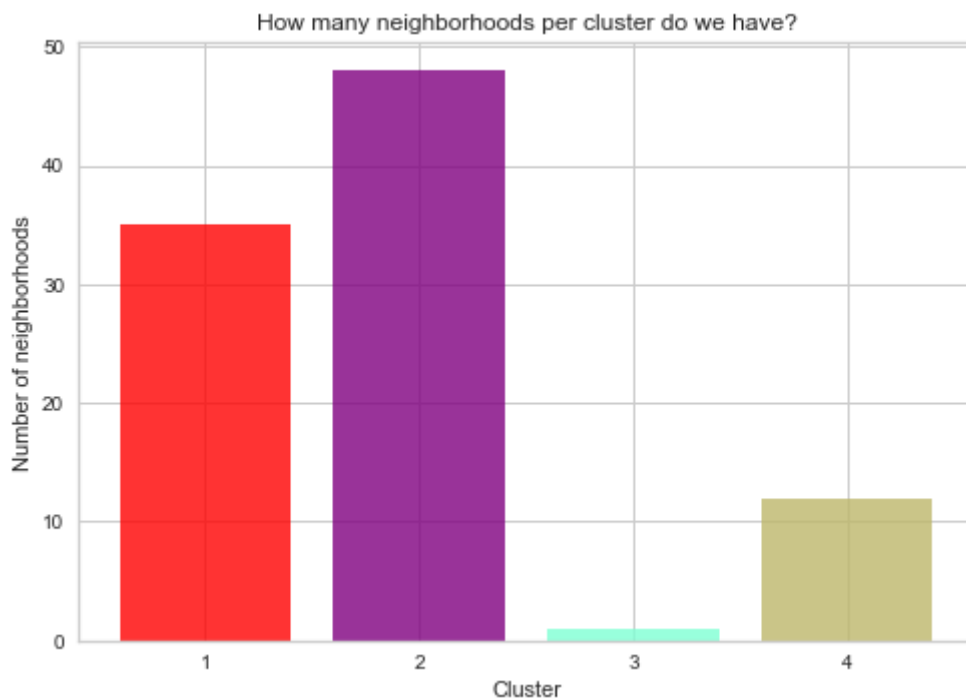
I'm putting the data on the map.



Looking at how many coffee shops each cluster has on average.

	Neighborhoods	Coffee Shop	Cluster Labels
0	Agincourt	0.000000	1
1	Alderwood, Long Branch	0.125000	0
2	Bathurst Manor, Wilson Heights, Downsview North	0.100000	0
3	Bayview Village	0.000000	1
4	Bedford Park, Lawrence Manor East	0.086957	0

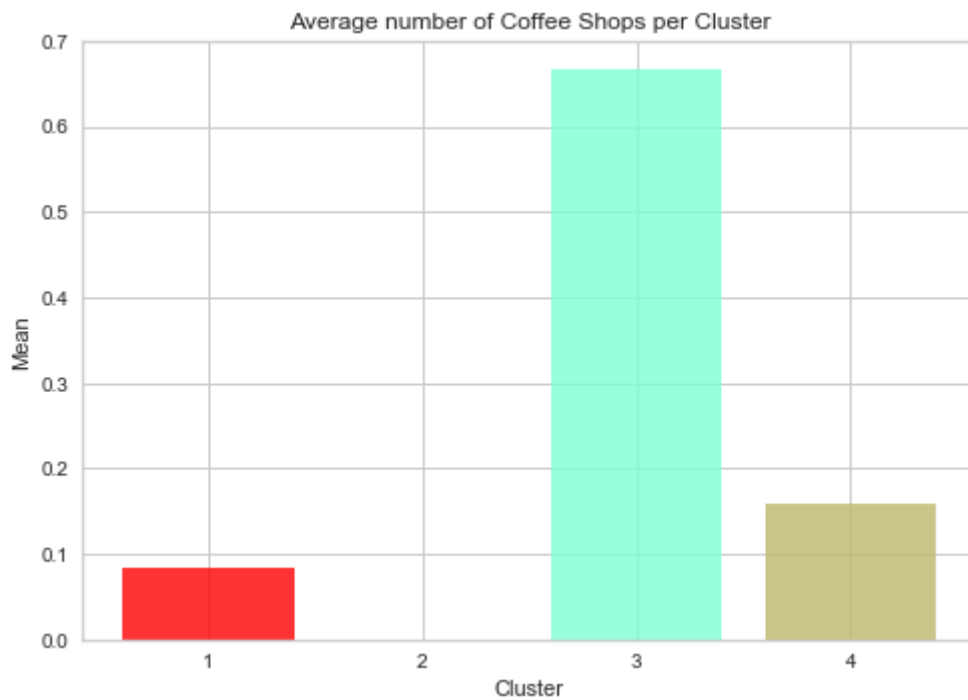
I look at how many neighborhoods each cluster has on average.



Looking at each cluster.

	Borough	Neighborhood	Coffee Shop	Cluster Labels	Neighborhood Latitude	Neighborhood Longitude	Venue	Venue Latitude	Venue Longitude	Venue Category
0	North York	Lawrence Manor, Lawrence Heights	0.083333	0	43.718518	-79.464763	Barrymore	43.715890	-79.468188	Furniture / Home Store
1	North York	Lawrence Manor, Lawrence Heights	0.083333	0	43.718518	-79.464763	Universal Lighting	43.721685	-79.466310	Furniture / Home Store
2	North York	Lawrence Manor, Lawrence Heights	0.083333	0	43.718518	-79.464763	Party City	43.719418	-79.459317	Miscellaneous Shop
3	North York	Lawrence Manor, Lawrence Heights	0.083333	0	43.718518	-79.464763	Rinx Real Entertainment Centre	43.717842	-79.462231	Event Space
4	North York	Lawrence Manor, Lawrence Heights	0.083333	0	43.718518	-79.464763	Bluenotes	43.718846	-79.465906	Clothing Store
...	...	...	...	...	...	...	...	...	...	...
1542	Downtown Toronto	Church and Wellesley	0.075000	0	43.665860	-79.383160	Town Inn Suites	43.669056	-79.382573	Hotel
1543	Downtown Toronto	Church and Wellesley	0.075000	0	43.665860	-79.383160	Joe Fresh	43.661956	-79.380160	Clothing Store
1544	Downtown Toronto	Church and Wellesley	0.075000	0	43.665860	-79.383160	Kokoni Izakaya	43.664181	-79.380258	Japanese Restaurant
1545	Downtown Toronto	Church and Wellesley	0.075000	0	43.665860	-79.383160	Starbucks	43.668071	-79.388367	Coffee Shop
1546	Downtown Toronto	Church and Wellesley	0.075000	0	43.665860	-79.383160	Flash	43.664319	-79.380190	Strip Club

Showing on a plot graph how many coffee shops each cluster has.



## 4 – Results

We can see that the first two clusters have the largest number of neighbourhoods and the smallest number of competitors simultaneously.

## 5 – Discussion

According to the results of the study, I would recommend opening a coffee shop first in the second cluster, as there are no competitors at all, and if the business will develop then move to the first cluster, which has competitors, but it is quite promising.

## 6 – Conclusion

Of course, it is not that simple, a business should be successful, with a good strategy and attractiveness for customers, but the very first step we have already done, found a place to start.