York Region, Canada

Analysis for Opening a new Restaurant

1. Introduction

- **Background**

  Is a regional municipality in Southern Ontario, Canada, between Lake Simcoe and Toronto. It replaced the former York County in 1971, and is part of the Greater Toronto Area and the inner ring of the Golden Horseshoe. The regional government is headquartered in New market.

  The 2016 census population was 1,109,909, with a growth rate of 7.5% from 2011 to 2016. The Government of Ontario expects its population to surpass 1.5 million residents by 2031.

- **Problem**

  data might contribute to analyze and decide which Neighborhood is the best to open a restaurant in order to get a good profit , also the data would help to decide what kind of restaurant should be opened Asian , Italian etc…

- **Interest**

  An American Business Man is interested in opening a restaurant in York region

  He wants to know what neighborhood is better in order to have a successful restaurant and have a good profit ,also he would like to know what kind of food is preferred to serve there This also helps with the success of the Restaurant

2. Data acquisition and cleaning

- **Data sources**

  First we need all the neighborhoods with its postal code , we took them from Wikipedia https://en.wikipedia.org/wiki/List_of_postal_codes_of_Canada:_M

  I scraped the data and add it to a data frame

[277]:

	Postal Code	Borough	Neighborhood
0	M1A	Not assigned	NaN
1	M2A	Not assigned	NaN
2	M3A	North York	Parkwoods
3	M4A	North York	Victoria Village
4	M5A	Downtown Toronto	Regent Park, Harbourfront

A csv file for the latitude and longitude with the postal codes of Canada https://cocl.us/Geospatial_data also downloaded into a data frame

[280]:

	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

Also we need the statistics of the population of Canada and its neighborhoods with the

Average salary and second most common language which may help us to decide what kind of restaurant we need to open this also we found in Wikipedia

https://en.wikipedia.org/wiki/Demographics_of_Toronto_neighbourhoods

after scraping the data and downloaded into a data frame

[168]:

	Name	FM	Census Tracts	Population	Land area (km2)	Density (people/km2)	% Change in Population since 2001	Average Income	Transit Commuting %	% Renters	Second most common language (after English) by name	Second most common language (after English) by percentage	Map
5	Amesbury	NY	0280.00, 0281.01, 0281.02	17318	3.51	4934	1.1	27546	16.4	19.7	Spanish (6.1%)	06.1% Spanish	NaN
6	Armour Heights	NY	0298.00	4384	2.29	1914	2.0	116651	10.8	16.1	Russian (9.4%)	09.4% Russian	NaN
7	Banbury	NY	0267.00	6641	2.72	2442	5.0	92319	6.1	4.8	Unspecified Chinese (5.1%)	05.1% Unspecified Chinese	NaN
8	Bathurst Manor	NY	0297.01, 0310.01, 0310.02	14945	4.69	3187	12.3	34169	13.4	18.6	Russian (9.5%)	09.5% Russian	NaN
10	Bayview Village	NY	0305.01, 305.02	12280	4.14	2966	41.6	46752	14.4	15.6	Cantonese (8.4%)	08.4% Cantonese	NaN

- **Data Cleaning**

  Data downloaded from the csv and scraped from the Wikipedia source for the neighborhoods were combined into one table , first I removed the rows in the first data frame when the BOROUGH is not assigned as it as no value to me , then I merged any rows which has the same postal codes , after that I created my data frame after including only York neighborhoods and exclude all others

[282]:

	Postal Code	Borough	Neighborhood	Latitude	Longitude
17	M2H	North York	Hillcrest Village	43.803762	-79.363452
18	M2J	North York	Fairview, Henry Farm, Oriole	43.778517	-79.346556
19	M2K	North York	Bayview Village	43.786947	-79.385975
20	M2L	North York	York Mills, Silver Hills	43.757490	-79.374714
21	M2M	North York	Willowdale, Newtonbrook	43.789053	-79.408493

After the data frame is ready I used the foursquare API in order to get all the venues for York region using the previous data frame and printing the venues were returned by Foursquare then I added the categories of the venues

[273]:

	Neighborhood	Neighborhood Latitude	Neighborhood Longitude	Venue	Venue Latitude	Venue Longitude	Venue Category
0	Hillcrest Village	43.803762	-79.363452	Eagle's Nest Golf Club	43.805455	-79.364186	Golf Course
1	Hillcrest Village	43.803762	-79.363452	AY Jackson Pool	43.804515	-79.366138	Pool
3	Hillcrest Village	43.803762	-79.363452	Duncan Creek Park	43.805539	-79.360695	Dog Run
4	Fairview, Henry Farm, Oriole	43.778517	-79.346556	The LEGO Store	43.778207	-79.343483	Toy / Game Store
5	Fairview, Henry Farm, Oriole	43.778517	-79.346556	DAVIDsTEA	43.777593	-79.345089	Tea Room

After I got the data frame of York venues with its categories I filtered it out and created my data frame with restaurant categories in order to use it in clustering and statistics later , also another data frame created which has all venues except the restaurant categories

The Restaurant venue data frame :

1]:

	Neighborhood	Neighborhood Latitude	Neighborhood Longitude	Venue	Venue Latitude	Venue Longitude	Venue Category
2	Hillcrest Village	43.803762	-79.363452	Villa Madina	43.801685	-79.363938	Mediterranean Restaurant
11	Fairview, Henry Farm, Oriole	43.778517	-79.346556	New York Fries - Fairview Mall	43.778605	-79.343577	Restaurant
17	Fairview, Henry Farm, Oriole	43.778517	-79.346556	Moxie's Classic Grill	43.777779	-79.343185	American Restaurant
24	Fairview, Henry Farm, Oriole	43.778517	-79.346556	Thai Express	43.777990	-79.344091	Restaurant
30	Fairview, Henry Farm, Oriole	43.778517	-79.346556	Heart Sushi	43.777203	-79.343805	Japanese Restaurant

The all Other Venues data Frame except restaurants:

:

	Neighborhood	Neighborhood Longitude	Venue	Venue Latitude	Venue Longitude	Venue Category
0	Hillcrest Village	-79.363452	Eagle's Nest Golf Club	43.805455	-79.364186	Golf Course
1	Hillcrest Village	-79.363452	AY Jackson Pool	43.804515	-79.366138	Pool
3	Hillcrest Village	-79.363452	Duncan Creek Park	43.805539	-79.360695	Dog Run
4	Fairview, Henry Farm, Oriole	-79.346556	The LEGO Store	43.778207	-79.343483	Toy / Game Store
5	Fairview, Henry Farm, Oriole	-79.346556	DAVIDsTEA	43.777593	-79.345089	Tea Room

Also for the population data frame after scraping the data from Wikipedia , it must be cleaned and all the not necessary columns should be removed in my case **Census Tracts** , **Land area (km2)**, **% Change in Population since 2001**, **% Renters** and **Map** , the other cleaning process is the column FM sould be filtered in order to distinguish the regions I queried only the values FM=="NY" or FM=="Y" or FM=="EY for York Region and set the index as Name of Neighborhood

Now I have cleared data frame as below

3]:

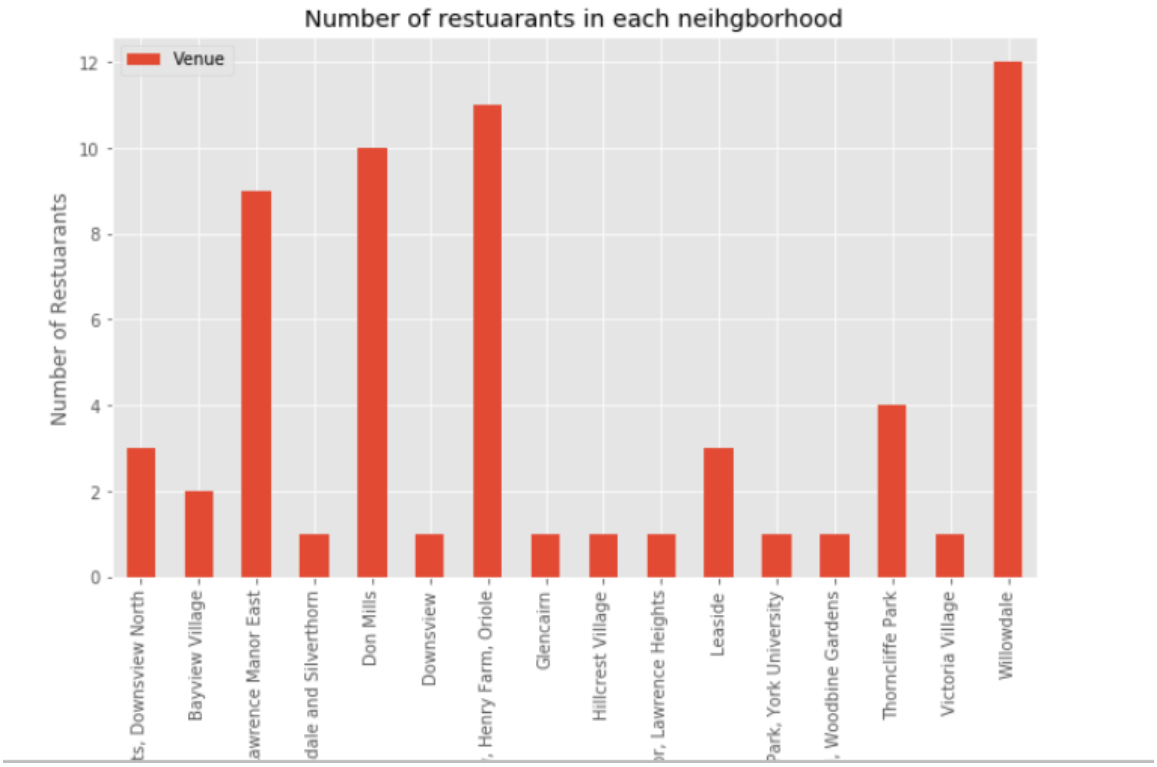
	Name	FM	Population	Density (people/km2)	Average Income	Transit Commuting %	Second most common language (after English) by name	Second most common language (after English) by percentage
8	Bathurst Manor	NY	14945	3187	34169	13.4	Russian (9.5%)	09.5% Russian
40	Don Mills	NY	21372	2377	47515	10.8	Unspecified Chinese (3.9%)	03.9% Unspecified Chinese
44	Downsview	NY	36613	2270	26751	14.4	Italian (11.7%)	11.7% Italian
67	Henry Farm	NY	2790	3066	56395	15.6	Mandarin (3.9%)	03.9% Mandarin
91	Leaside	EY	13876	4938	82670	9.7	Bulgarian (0.4%)	00.4% Bulgarian
168	Willowdale	NY	43144	5618	39895	15.6	Cantonese (7.9%)	07.9% Cantonese
169	Wilson Heights	NY	13732	3317	37978	15.9	Filipino (6.2%)	06.2% Filipino

3. Exploratory Data Analysis

- First in order to know how many Restaurant in the neighborhood I used my data frame for venue restaurants in order to count the number grouped by the neighborhood then get all the neighbors which has the smallest number of restaurants which are :

'Bathurst Manor, Wilson Heights, Downsview North, Bayview Village, Del Ray, Mount Dennis, Keelsdale and Silverthorn' , 'Downsview', 'Glencairn','Hillcrest Village','Lawrence Manor, Lawrence Heights','Leaside,'Northwood Park, York University' ,'Parkview Hill, Woodbine Gardens','Parkview Hill, Woodbine Gardens','Victoria Village

So many well let us see the bar plot and see the relationship between neihgborhoods and number of restaurants



Now for modeling I created a data frame for all the restaurant categories and their neighborhoods then hot decoded them as following

[299]:

	Neighborhood	American Restaurant	Asian Restaurant	Caribbean Restaurant	Chinese Restaurant	Comfort Food Restaurant	Dim Sum Restaurant	Fast Food Restaurant	Greek Restaurant	Indian Restaurant	Indonesian Restaurant	Italian Restaurant	Japanese Restaurant	Korean Restaurant
2	Hillcrest Village	0	0	0	0	0	0	0	0	0	0	0	0	0
11	Fairview, Henry Farm, Oriole	0	0	0	0	0	0	0	0	0	0	0	0	0
17	Fairview, Henry Farm, Oriole	1	0	0	0	0	0	0	0	0	0	0	0	0
24	Fairview, Henry Farm, Oriole	0	0	0	0	0	0	0	0	0	0	0	0	0
30	Fairview, Henry Farm, Oriole	0	0	0	0	0	0	0	0	0	0	0	1	0

Then calculated the mean value

[300]:

American estaurant	Asian Restaurant	Caribbean Restaurant	Chinese Restaurant	Comfort Food Restaurant	Dim Sum Restaurant	Fast Food Restaurant	Greek Restaurant	Indian Restaurant	Indonesian Restaurant	Italian Restaurant	Japanese Restaurant	Korean Restaurant	Mediterranean Restaurant	M Resta
0.000000	0.0	0.0	0.0	0.000000	0.0	0.0	0.000000	0.000000	0.0	0.000000	0.0	0.0	0.0	
0.000000	0.0	0.0	0.5	0.000000	0.0	0.0	0.000000	0.000000	0.0	0.000000	0.5	0.0	0.0	
0.111111	0.0	0.0	0.0	0.111111	0.0	0.0	0.111111	0.111111	0.0	0.222222	0.0	0.0	0.0	
0.000000	0.0	0.0	0.0	0.000000	0.0	0.0	0.000000	0.000000	0.0	0.000000	0.0	0.0	0.0	
0.000000	0.2	0.1	0.1	0.000000	0.1	0.0	0.000000	0.000000	0.0	0.100000	0.2	0.0	0.0	

Now also i calculated the top five restaurants for each neighborhood

Then I need to create a data frame for the top ten most common restaurants in each neighbor hood

	Neighborhood	1st Most Common Restuarant	2nd Most Common Restuarant	3rd Most Common Restuarant	4th Most Common Restuarant	5th Most Common Restuarant	6th Most Common Restuarant	7th Most Common Restuarant	8th Most Common Restuarant	9th Most Common Restuarant	10th Most Common Restuarant
0	Bathurst Manor, Wilson Heights, Downsview North	Sushi Restaurant	Restaurant	Middle Eastern Restaurant	Vietnamese Restaurant	Indian Restaurant	Asian Restaurant	Caribbean Restaurant	Chinese Restaurant	Comfort Food Restaurant	Dim Sum Restaurant
1	Bayview Village	Chinese Restaurant	Japanese Restaurant	Vietnamese Restaurant	Indonesian Restaurant	Asian Restaurant	Caribbean Restaurant	Comfort Food Restaurant	Dim Sum Restaurant	Fast Food Restaurant	Greek Restaurant
2	Bedford Park, Lawrence Manor East	Italian Restaurant	Thai Restaurant	Comfort Food Restaurant	Greek Restaurant	Indian Restaurant	American Restaurant	Restaurant	Sushi Restaurant	Middle Eastern Restaurant	Mexican Restaurant
3	Del Ray, Mount Dennis, Keelsdale and Silverthorn	Restaurant	Vietnamese Restaurant	Indonesian Restaurant	Asian Restaurant	Caribbean Restaurant	Chinese Restaurant	Comfort Food Restaurant	Dim Sum Restaurant	Fast Food Restaurant	Greek Restaurant
4	Don Mills	Asian Restaurant	Restaurant	Japanese Restaurant	Italian Restaurant	Caribbean Restaurant	Chinese Restaurant	Dim Sum Restaurant	Indonesian Restaurant	Comfort Food Restaurant	Fast Food Restaurant

Then I queried the neighborhoods which has the least number of restaurants and what are the most common trends

[70]:

	Neighborhood	1st Most Common Restuarant	2nd Most Common Restuarant	3rd Most Common Restuarant	4th Most Common Restuarant	5th Most Common Restuarant	6th Most Common Restuarant	7th Most Common Restuarant	8th Most Common Restuarant	9th Most Common Restuarant	10th Most Common Restuarant
0	Bathurst Manor, Wilson Heights, Downsview North	Sushi Restaurant	Restaurant	Middle Eastern Restaurant	Vietnamese Restaurant	Indian Restaurant	Asian Restaurant	Caribbean Restaurant	Chinese Restaurant	Comfort Food Restaurant	Dim Sum Restaurant
1	Bayview Village	Chinese Restaurant	Japanese Restaurant	Vietnamese Restaurant	Indonesian Restaurant	Asian Restaurant	Caribbean Restaurant	Comfort Food Restaurant	Dim Sum Restaurant	Fast Food Restaurant	Greek Restaurant
3	Del Ray, Mount Dennis, Keelsdale and Silverthorn	Restaurant	Vietnamese Restaurant	Indonesian Restaurant	Asian Restaurant	Caribbean Restaurant	Chinese Restaurant	Comfort Food Restaurant	Dim Sum Restaurant	Fast Food Restaurant	Greek Restaurant
5	Downsview	Korean Restaurant	Vietnamese Restaurant	Indonesian Restaurant	Asian Restaurant	Caribbean Restaurant	Chinese Restaurant	Comfort Food Restaurant	Dim Sum Restaurant	Fast Food Restaurant	Greek Restaurant
7	Glencairn	Japanese Restaurant	Vietnamese Restaurant	Indonesian Restaurant	Asian Restaurant	Caribbean Restaurant	Chinese Restaurant	Comfort Food Restaurant	Dim Sum Restaurant	Fast Food Restaurant	Greek Restaurant
8	Hillcrest Village	Mediterranean Restaurant	Vietnamese Restaurant	Indonesian Restaurant	Asian Restaurant	Caribbean Restaurant	Chinese Restaurant	Comfort Food Restaurant	Dim Sum Restaurant	Fast Food Restaurant	Greek Restaurant

4. Modeling Using K Mean Clustering

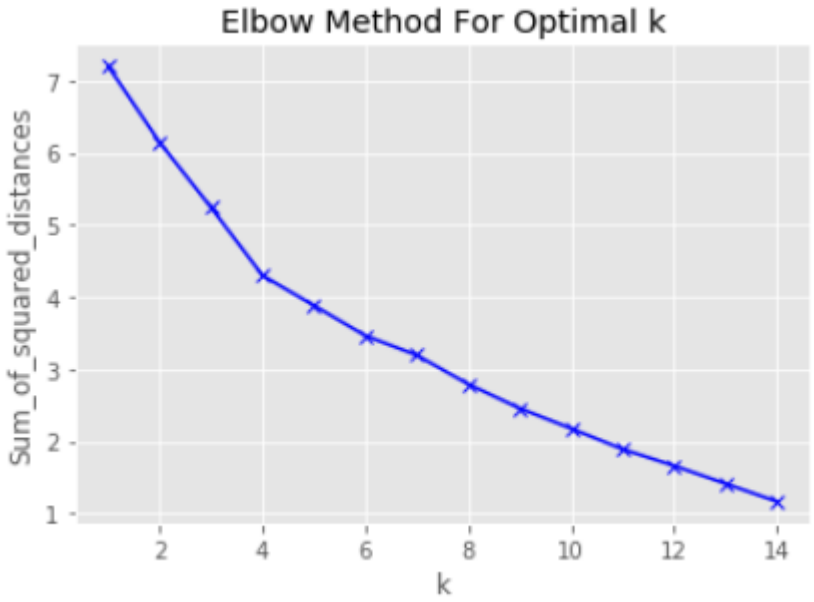
In order to cluster the restaurants in York region using the top ten common restaurants we used K mean Clustering which is very good method for the calculation and cluster the data with it is similar data besides it is not complicated and easy to use

- First we need to find the best K

Using the elbo method as a method o evaluation the model

```
K = range(1,15)
for k in K:
    km = KMeans(n_clusters=k)
    km = km.fit(York_grouped_clustering)
    Sum_of_squared_distances.append(km.inertia_)
```

317]:
```
plt.plot(K, Sum_of_squared_distances, 'bx-')
plt.xlabel('k')
plt.ylabel('Sum_of_squared_distances')
plt.title('Elbow Method For Optimal k')
plt.show()
```



And after trying K=6, k=8,k=9

It was found that the best k is 6 so we will need 6 clusters as our data is not that big

So I used the data frame after dropping the neighborhood column and fit the data into the Kcluster method and generated the cluster labels then inserted them into the data frame

[304]:

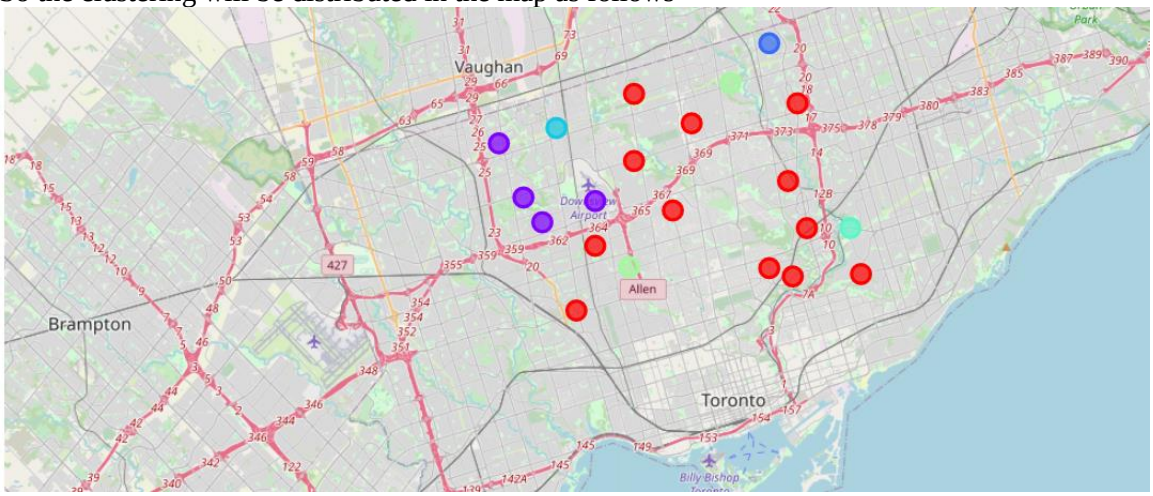
	Postal Code	Borough	Neighborhood	Latitude	Longitude	Cluster Labels	1st Most Common Restuarant	2nd Most Common Restuarant	3rd Most Common Restuarant	4th Most Common Restuarant	5th Most Common Restuarant
17	M2H	North York	Hillcrest Village	43.803762	-79.363452	2.0	Mediterranean Restaurant	Vietnamese Restaurant	Indonesian Restaurant	Asian Restaurant	Caribbean Restaurant
18	M2J	North York	Fairview, Henry Farm, Oriole	43.778517	-79.346556	0.0	Fast Food Restaurant	Japanese Restaurant	Restaurant	American Restaurant	Asian Restaurant
19	M2K	North York	Bayview Village	43.786947	-79.385975	5.0	Chinese Restaurant	Japanese Restaurant	Vietnamese Restaurant	Indonesian Restaurant	Asian Restaurant
22	M2N	North York	Willowdale	43.770120	-79.408493	0.0	Ramen Restaurant	Sushi Restaurant	Restaurant	Vietnamese Restaurant	Middle Eastern Restaurant
24	M2R	North York	Willowdale	43.782736	-79.442259	0.0	Ramen Restaurant	Sushi Restaurant	Restaurant	Vietnamese Restaurant	Middle Eastern Restaurant

After cleaning the data and removing NAN values and change cluster label type to integer the final data frame is as below

[306]:

	Postal Code	Borough	Neighborhood	Latitude	Longitude	Cluster Labels	1st Most Common Restuarant	2nd Most Common Restuarant	3rd Most Common Restuarant	4th Most Common Restuarant	5th Most Common Restuarant	6th Most Common Restuarant	7th Most Common Restuarant	8th Most Common Restuarant
17	M2H	North York	Hillcrest Village	43.803762	-79.363452	2	Mediterranean Restaurant	Vietnamese Restaurant	Indonesian Restaurant	Asian Restaurant	Caribbean Restaurant	Chinese Restaurant	Comfort Food Restaurant	Dim Sum Restaurant
18	M2J	North York	Fairview, Henry Farm, Oriole	43.778517	-79.346556	0	Fast Food Restaurant	Japanese Restaurant	Restaurant	American Restaurant	Asian Restaurant	Indonesian Restaurant	Caribbean Restaurant	Chinese Restaurant
19	M2K	North York	Bayview Village	43.786947	-79.385975	5	Chinese Restaurant	Japanese Restaurant	Vietnamese Restaurant	Indonesian Restaurant	Asian Restaurant	Caribbean Restaurant	Comfort Food Restaurant	Dim Sum Restaurant
22	M2N	North York	Willowdale	43.770120	-79.408493	0	Ramen Restaurant	Sushi Restaurant	Restaurant	Vietnamese Restaurant	Middle Eastern Restaurant	Japanese Restaurant	Indonesian Restaurant	Fast Food Restaurant
24	M2R	North York	Willowdale	43.782736	-79.442259	0	Ramen Restaurant	Sushi Restaurant	Restaurant	Vietnamese Restaurant	Middle Eastern Restaurant	Japanese Restaurant	Indonesian Restaurant	Fast Food Restaurant

So the clustering will be distributed in the map as follows



Examining the clusters

Cluster 1

[112]:

	Neighborhood	Longitude	Cluster Labels	1st Most Common Restuarant	2nd Most Common Restuarant	3rd Most Common Restuarant	4th Most Common Restuarant	5th Most Common Restuarant	6th Most Common Restuarant	7th Most Common Restuarant	8th Most Common Restuarant	9th Most Common Restuarant	10th Most Common Restuarant
18	Fairview, Henry Farm, Oriole	-79.346556	0	Fast Food Restaurant	Japanese Restaurant	Restaurant	American Restaurant	Asian Restaurant	Indonesian Restaurant	Caribbean Restaurant	Chinese Restaurant	Comfort Food Restaurant	Dim Sum Restaurant
22	Willowdale	-79.408493	0	Ramen Restaurant	Sushi Restaurant	Restaurant	Vietnamese Restaurant	Middle Eastern Restaurant	Japanese Restaurant	Indonesian Restaurant	Fast Food Restaurant	Greek Restaurant	Asian Restaurant
24	Willowdale	-79.442259	0	Ramen Restaurant	Sushi Restaurant	Restaurant	Vietnamese Restaurant	Middle Eastern Restaurant	Japanese Restaurant	Indonesian Restaurant	Fast Food Restaurant	Greek Restaurant	Asian Restaurant
26	Don Mills	-79.352188	0	Asian Restaurant	Restaurant	Japanese Restaurant	Italian Restaurant	Caribbean Restaurant	Chinese Restaurant	Dim Sum Restaurant	Indonesian Restaurant	Comfort Food Restaurant	Fast Food Restaurant
27	Don Mills	-79.340923	0	Asian Restaurant	Restaurant	Japanese Restaurant	Italian Restaurant	Caribbean Restaurant	Chinese Restaurant	Dim Sum Restaurant	Indonesian Restaurant	Comfort Food Restaurant	Fast Food Restaurant
28	Bathurst Manor, Wilson Heights, Downsview North	-79.442259	0	Sushi Restaurant	Restaurant	Middle Eastern Restaurant	Vietnamese Restaurant	Indian Restaurant	Asian Restaurant	Caribbean Restaurant	Chinese Restaurant	Comfort Food Restaurant	Dim Sum Restaurant
35	Parkview Hill, Woodbine Gardens	-79.309937	0	Fast Food Restaurant	Vietnamese Restaurant	Thai Restaurant	Asian Restaurant	Caribbean Restaurant	Chinese Restaurant	Comfort Food	Dim Sum Restaurant	Greek Restaurant	Indian Restaurant

It is noticed that the popular categories of restaurant in these neighborhood are Sushi , fast food , Chinese and Indian

Cluster 2

This neighborhood also prefer Asian food such as Chinese and Japanese and Greek

Cluster 3

this neighborhood prefers Vietnamese and Thai food

Cluster 4

These neighborhoods prefer Thai , sushi and Ramen with fast-food

Cluster 5

this Neighborhood prefers Portuguese , Vietnamese and Indonesian food

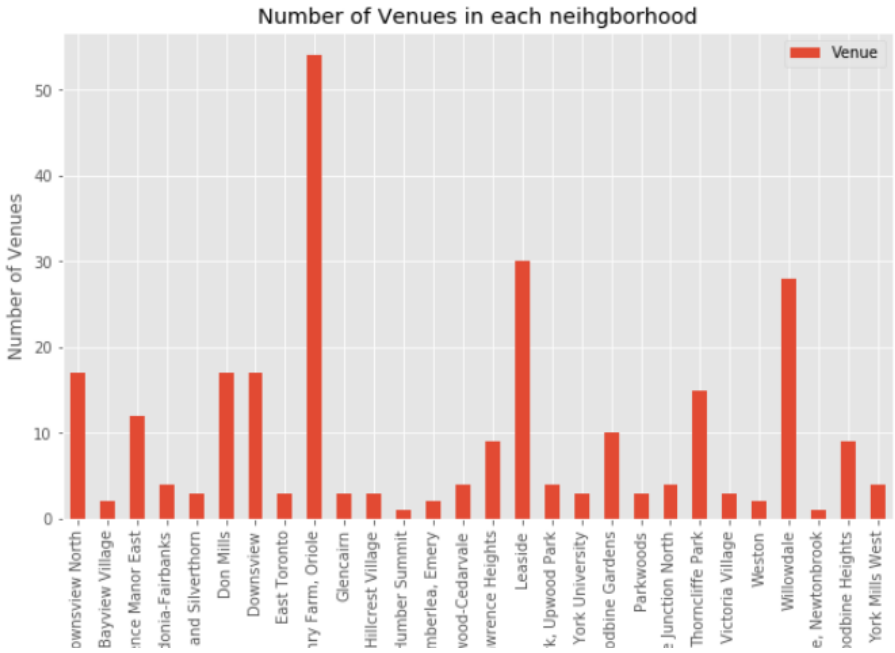
Cluster 6

These Two prefer Chinese , Japanese and Vietnamese

now we finished with similar venues as restaurant categories

we need to cluster all other venues except restaurants and see how many venues are there in each neighborhood

as we can see in this bar visualization diagram



the neighborhoods which have the most number of venues are (the more exist the better as the restaurant will have a lot of visitors which will be hungry after any activity)

Bathurst Manor, Wilson Heights, Downs view North

Don Mills

Downsview

airview, Henry Farm, Oriole

Willowdale

Leaside

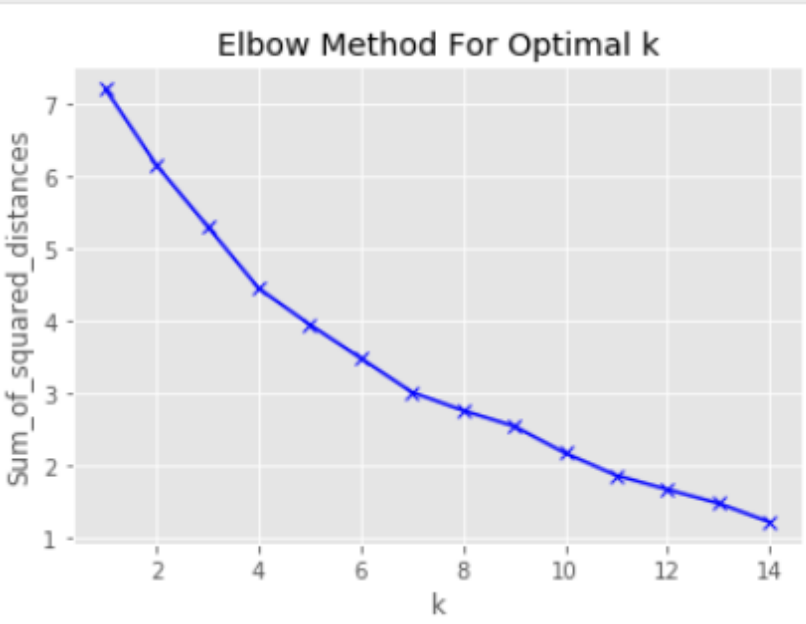
Then I repeated all the other steps onehot coding finding the mean values and group by the neighborhood then findout the top ten most common venues after clearing the data the final data frame is as below

[146]:

	Neighborhood	1st Most Common Venue	2nd Most Common Venue	3rd Most Common Venue	4th Most Common Venue	5th Most Common Venue	6th Most Common Venue	7th Most Common Venue	8th Most Common Venue	9th Most Common Venue	10th Most Common Venue
0	Bathurst Manor, Wilson Heights, Downsview North	Coffee Shop	Bank	Frozen Yogurt Shop	Pharmacy	Shopping Mall	Deli / Bodega	Ice Cream Shop	Diner	Pizza Place	Supermarket
5	Don Mills	Coffee Shop	Beer Store	Gym	Bubble Tea Shop	Sandwich Place	Discount Store	Concert Hall	Bike Shop	Sporting Goods Shop	Gym / Fitness Center
6	Downsview	Grocery Store	Park	Liquor Store	Baseball Field	Shopping Mall	Gym / Fitness Center	Snack Place	Bus Stop	Hotel	Home Service
24	Willowdale	Coffee Shop	Pizza Place	Café	Sandwich Place	Shopping Mall	Juice Bar	Butcher	Ice Cream Shop	Movie Theater	Hotel

Modeling with Kmean

As previous I used the elbo method and testing some values in order to find the best K for clustering the venues

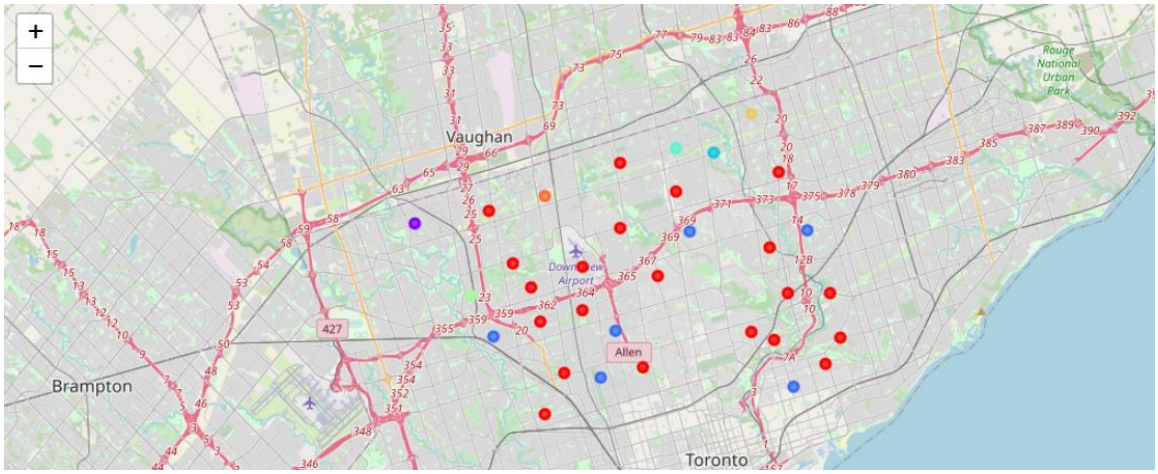


After testing the best value for K is 8 , using and executing the K mean model after fitting the data and insert cluster labels we will have the result below :

[154]:

	Postal Code	Borough	Neighborhood	Latitude	Longitude	Cluster Labels	1st Most Common Venue	2nd Most Common Venue	3rd Most Common Venue	4th Most Common Venue	5th Most Common Venue	6th Most Common Venue	7th Most Common Venue	8th Most Common Venue	9th Most Common Venue	C
17	M2H	North York	Hillcrest Village	43.803762	-79.363452	6	Pool	Golf Course	Dog Run	Yoga Studio	Dessert Shop	Coffee Shop	Concert Hall	Construction & Landscaping	Convenience Store	C
18	M2J	North York	Fairview, Henry Farm, Oriole	43.778517	-79.346556	0	Clothing Store	Coffee Shop	Bank	Women's Store	Bakery	Burrito Place	Movie Theater	Distribution Center	Metro Station	
19	M2K	North York	Bayview Village	43.786947	-79.385975	3	Café	Bank	Yoga Studio	Discount Store	Concert Hall	Construction & Landscaping	Convenience Store	Cosmetics Shop	Curling Ice	
21	M2M	North York	Willowdale, Newtonbrook	43.789053	-79.408493	4	Home Service	Yoga Studio	Diner	Coffee Shop	Concert Hall	Construction & Landscaping	Convenience Store	Cosmetics Shop	Curling Ice	
22	M2N	North York	Willowdale	43.770120	-79.408493	0	Coffee Shop	Pizza Place	Café	Sandwich Place	Shopping Mall	Juice Bar	Butcher	Ice Cream Shop	Movie Theater	

And the distribution of the clusters using folium map is



Explore the clusters

Cluster 1 the most trended venues beer shop pizza place, clothing and cafe shop

we can see that Fairview, Henry Farm, Oriole and Willowdale and Don Mills are in tis cluster so they have similar venues and as Willowdale has a big number of restaurants and they are all successful then in result all the neighborhoods in this cluster are in the save side

Cluster 2 The most common venues Pizza Place,Fried Chicken Joint and Clothing Store

Cluster 3 The most common venues Park , pizza and coffee shop

Cluster 4 The most common venues café ,bank and yoga

Cluster 5 The most common venues park, convenience store, yoga

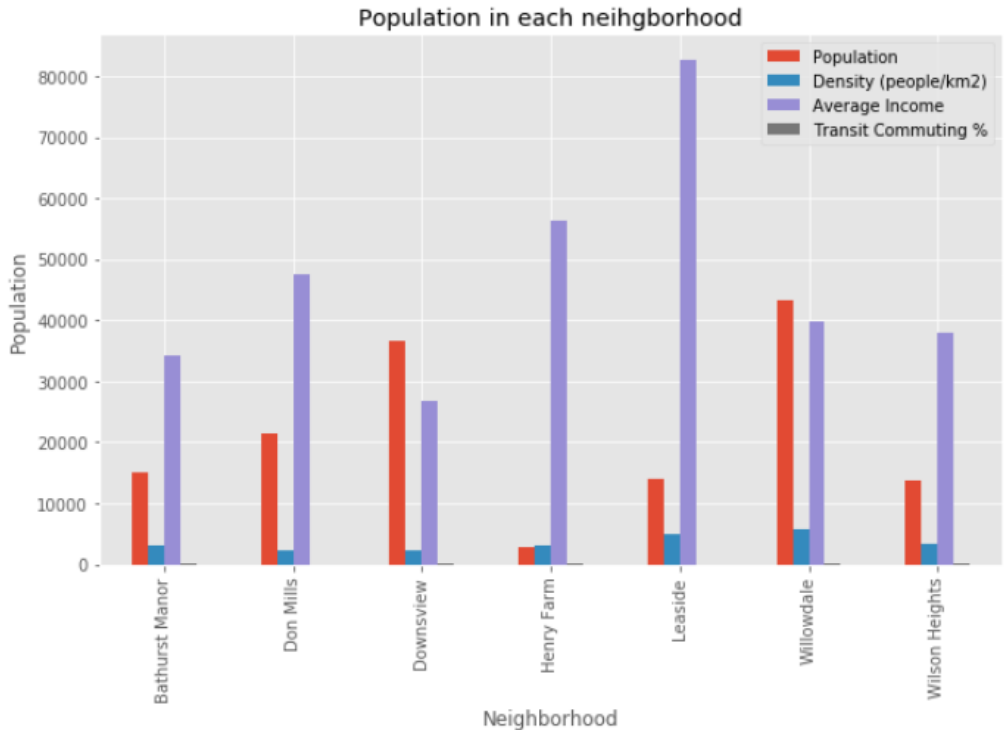
Cluster 6 The most common venues food service, baseball field and yoga

Cluster 7 The most common venues pool ,golg ,dog run

Cluster 8 The most common venues Coffee shop, message studio and bar

After all that I executed the final part which is using all the previous information such as clustering all the statistics after scraping the population information from the Wikipedia about the York region

Using the bar plot to compare the neighborhoods with their average salary population and some other attributes as follows



Looking at the plot we can see that Willowdale has the biggest population and a good average salary

also Downsview has a good number of population and average salary

Leaside is not convenient as it has small number of population and it is the most important attribute but the salary average is high it can be cuz there is no big number of people live there

5. comparing results

by comparing the most populated and best salary average neighbor hoods with the biggest number Venues in each neighborhood the result is as follows

[227]:

Neighborhood	Venue
Don Mills	17
Downsview	17
Leaside	30
Willowdale	28

These are the best neighborhoods using the population salary average statistics and all other venues except restaurants similar clusters

Then I calculated the number of restaurants for these neighborhoods

[228]:

Neighborhood	Venue
Don Mills	10
Downsview	1
Leaside	3
Willowdale	12

We can see after comparing the two tables the best neighborhoods are Downs view and leaside , don mils also good choice but it has too many restaurants so is willowdale

Then I calculated the type of restaurants of these two neighborhoods

[329]:

	Neighborhood	Venue	Venue Category
184	Downsview	Sean's Twisted	Korean Restaurant
217	Leaside	The Leaside Pub	Restaurant
224	Leaside	Kintako Japanese Restaurant	Sushi Restaurant
245	Leaside	Mucho Burrito Fresh Mexican Grill	Mexican Restaurant

After that comparing the clusters I can see that they all are in two clusters as

26	M3B	North York	Don Mills	43.745906	-79.352188	0	Asian Restaurant	Restaurant	Japanese Restaurant	Italian Restaurant	Caribbean Restaurant	Chinese Restaurant	Dim Sum Restaurant	Indonesian Restaurant	Comfort Food Restaurant
27	M3C	North York	Don Mills	43.725900	-79.340923	0	Asian Restaurant	Restaurant	Japanese Restaurant	Italian Restaurant	Caribbean Restaurant	Chinese Restaurant	Dim Sum Restaurant	Indonesian Restaurant	Comfort Food Restaurant
30	M3K	North York	Downsview	43.737473	-79.464763	1	Korean Restaurant	Vietnamese Restaurant	Indonesian Restaurant	Asian Restaurant	Caribbean Restaurant	Chinese Restaurant	Comfort Food Restaurant	Dim Sum Restaurant	Fast Food Restaurant
31	M3L	North York	Downsview	43.739015	-79.506944	1	Korean Restaurant	Vietnamese Restaurant	Indonesian Restaurant	Asian Restaurant	Caribbean Restaurant	Chinese Restaurant	Comfort Food Restaurant	Dim Sum Restaurant	Fast Food Restaurant
32	M3M	North York	Downsview	43.728496	-79.495697	1	Korean Restaurant	Vietnamese Restaurant	Indonesian Restaurant	Asian Restaurant	Caribbean Restaurant	Chinese Restaurant	Comfort Food Restaurant	Dim Sum Restaurant	Fast Food Restaurant
33	M3N	North York	Downsview	43.761631	-79.520999	1	Korean Restaurant	Vietnamese Restaurant	Indonesian Restaurant	Asian Restaurant	Caribbean Restaurant	Chinese Restaurant	Comfort Food Restaurant	Dim Sum Restaurant	Fast Food Restaurant
38	M4G	East York	Leaside	43.709060	-79.363452	0	Sushi Restaurant	Restaurant	Mexican Restaurant	Vietnamese Restaurant	Indian Restaurant	Asian Restaurant	Caribbean Restaurant	Chinese Restaurant	Comfort Food Restaurant

Leaside , willodale and don mils cluster 1

Downsview cluster 2 based on this information as for venues except restaurants all these neighborhoods are in the same cluster so they have similar venues so tey all have similar number of customers and frequency on the other hand Downsview belongs to another cluster not similar to Willowdale for example in case of most common restaurants so this will help us to exclude all the common similar restaurants in cluster 1 and look only in cluster 2

6. Conclusion

looking at it all we can recommend to open a restaurant in Donsview as it has only one Korean restaurant

1-is neighborhood has a good population and good average salary

2-it has 11 percent of Italian native language speakers

3- and based on clustering the most trended restaurant categories are Asian categories

4-finally it has a good number of other venues

And there are few choices based on this Information to choose what kind of food is served

It could be Chinese or Vietnamese based on clustering or Italian Restaurant as there are no Italian food in this region and there are 11 percent of people speaks Italian.