The Battle of Neighbourhoods

-: Final Report :-

Introduction/Business Problem:

Toronto, Canada and Bangalore, India are two major cities in the world. Both cities become a centre of attention for residential, job employment, tourism, education, shopping and sports activity.

In this project, we will study in details the area classification using Foursquare data and ML segmentation and clustering (K-Mean Algorithm). The aim of this project is to segment areas of Toronto and Bangalore based on the most common places captured from Foursquare.

Using segmentation and clustering, we hope we can determine:

- The similarity or dissimilarity of both cities
- Classification of area located inside the city whether it is residential, tourism places, or others

Data:

Toronto city data is already available and we have used that data to cluster the places in the week 3 assignment.

Bangalore city data is available in Wikipedia:
https://en.wikipedia.org/wiki/List_of_neighbourhoods_in_Bangalore

I have filtered the data and saved it the GitHub repository:
https://github.com/pritamroy11/Coursera_IBM-Watson/blob/master/week4-5/Bangalore_neighbourhoods.csv

Now, we have the neighbourhood details of the Bangalore city, I have used geocoder to get the latitude, longitude and FourSquare API for visualizing the data after segmentation and clustering using K-Means clustering algorithm

data of Toronto city:

```
In [33]: neighborhoods.head()
```

Out[33]:

	Postcode	Borough	Neighbourhood	Latitude	Longitude
37	M4E	East Toronto	The Beaches	43.676357	-79.293031
41	M4K	East Toronto	The Danforth West, Riverdale	43.679557	-79.352188
42	M4L	East Toronto	The Beaches West, India Bazaar	43.668999	-79.315572
43	M4M	East Toronto	Studio District	43.659526	-79.340923
44	M4N	Central Toronto	Lawrence Park	43.728020	-79.388790

data of Bangalore City:

```
In [8]: df.head()
```

Out[8]:

	Area	Neighborhoods	Latitude	Longitude	City
0	Central	Cantonment area	12.972442	77.580643	Bangalore
1	Central	Domlur	12.960992	77.638726	Bangalore
2	Central	Indiranagar	12.971891	77.641151	Bangalore
3	Central	Jeevanbheemanagar	12.962900	77.659500	Bangalore
4	Central	Malleswaram	13.003100	77.564300	Bangalore

Methodology:

1. I have converted the neighbourhood addresses into their equivalent latitude and longitude values.

2. Then I have used the Foursquare API to explore neighbourhoods in both the cities, using explore function to get the most common venue categories in each neighbourhood.

   Toronto:

```
In [39]: print(toronto_venues.shape)
         toronto_venues.head()
```

```
(1709, 7)
```

Out[39]:

	Neighborhood	Neighborhood Latitude	Neighborhood Longitude	Venue	Venue Latitude	Venue Longitude	Venue Category
0	The Beaches	43.676357	-79.293031	Glen Manor Ravine	43.676821	-79.293942	Trail
1	The Beaches	43.676357	-79.293031	The Big Carrot Natural Food Market	43.678879	-79.297734	Health Food Store
2	The Beaches	43.676357	-79.293031	Grover Pub and Grub	43.679181	-79.297215	Pub
3	The Beaches	43.676357	-79.293031	Upper Beaches	43.680563	-79.292869	Neighborhood
4	The Danforth West, Riverdale	43.679557	-79.352188	Pantheon	43.677621	-79.351434	Greek Restaurant

Bangalore:

```
In [37]: print(blore_venues.shape)
         blore_venues.head()
```

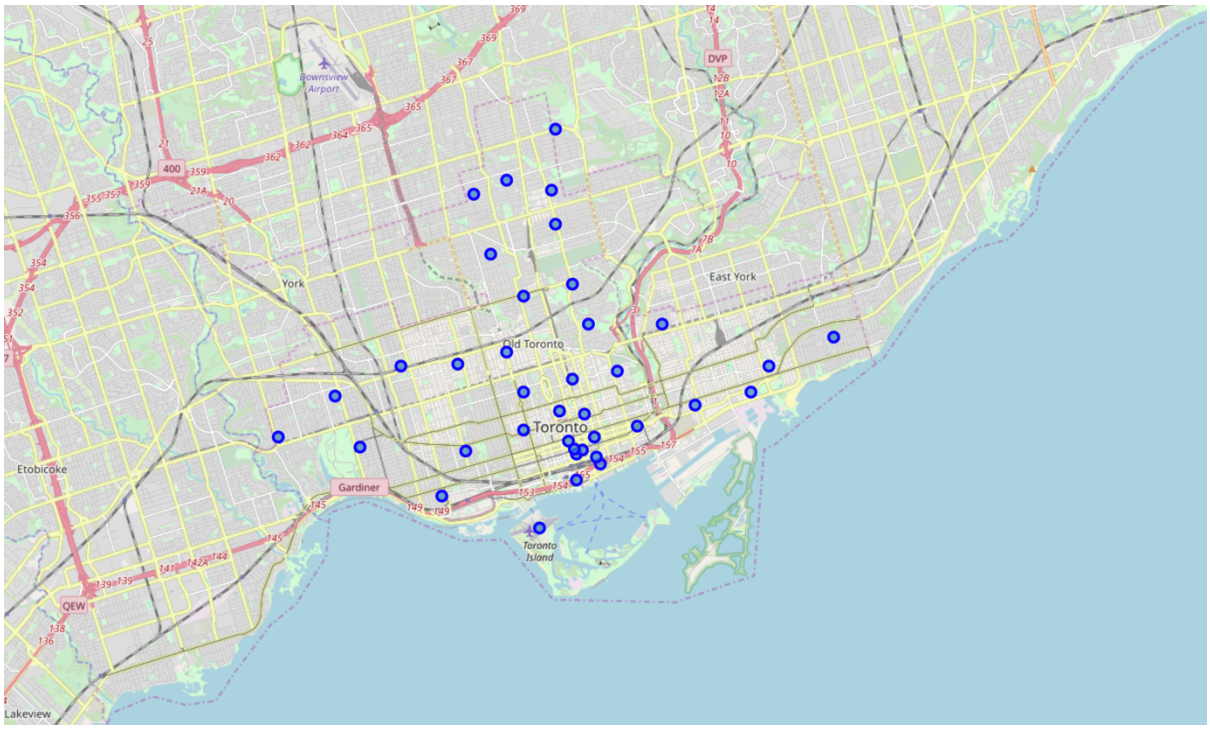
```
(607, 7)
```

Out[37]:

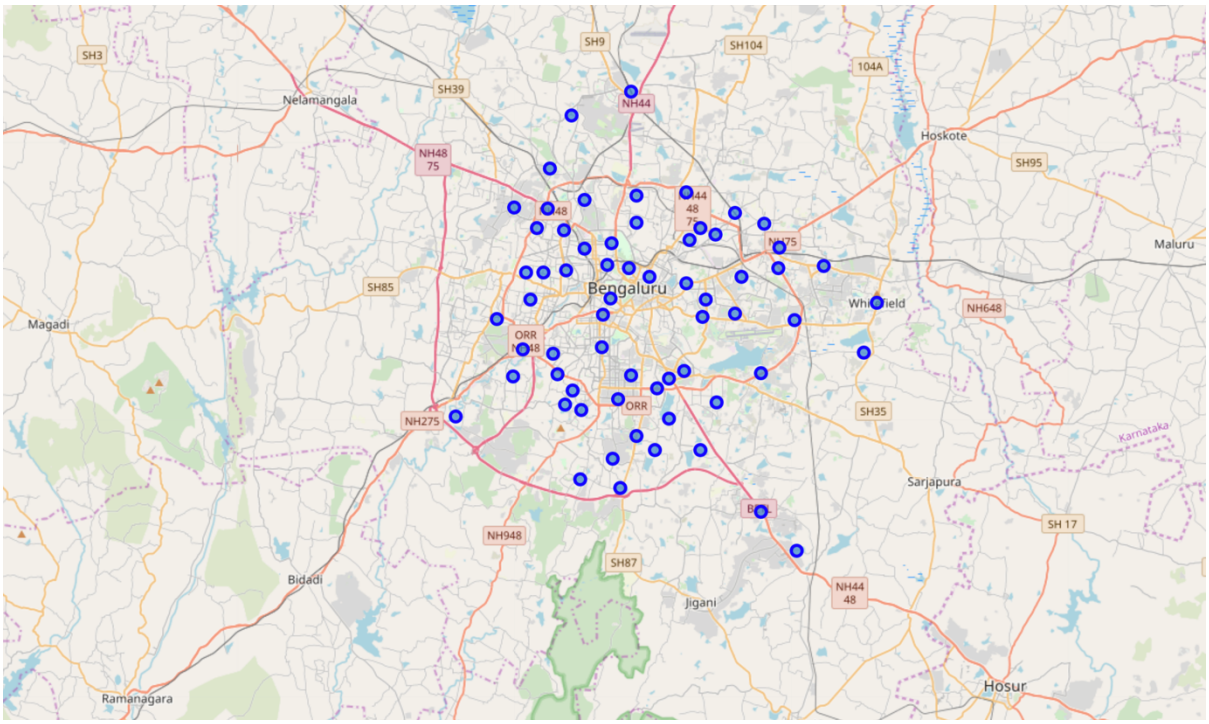
	Neighborhood	Neighborhood Latitude	Neighborhood Longitude	Venue	Venue Latitude	Venue Longitude	Venue Category
0	Cantonment area	12.972442	77.580643	Hotel Fishland	12.975569	77.578592	Seafood Restaurant
1	Cantonment area	12.972442	77.580643	Vasudev Adigas	12.973707	77.579257	Indian Restaurant
2	Cantonment area	12.972442	77.580643	Sapna Book House	12.976355	77.578461	Bookstore
3	Cantonment area	12.972442	77.580643	Adigas Hotel	12.973554	77.579161	Restaurant
4	Cantonment area	12.972442	77.580643	Kamat Yatrinivas	12.975985	77.578125	Indian Restaurant

3. The Folium library is used to visualize the neighbourhoods in Toronto and Bangalore with their respective clusters.

Toronto:



Bangalore:

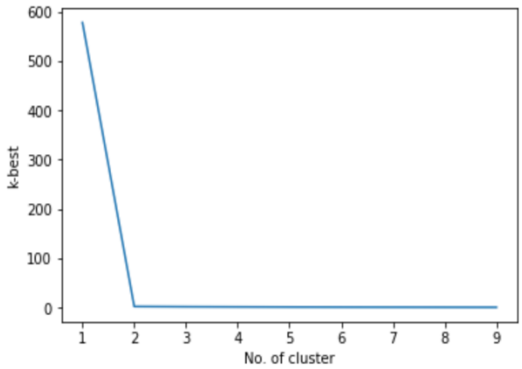


4. I have used the venue data after grouping the neighbourhoods for segmenting and clustering using K-Means Clustering Algorithm to find the clusters.

    a. Finding optimum K (cluster number) for the data using elbow graph

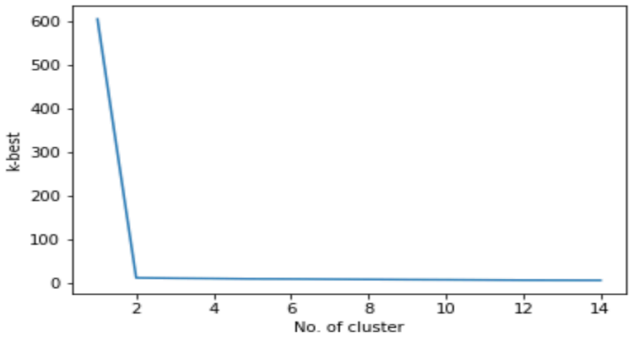
    Toronto:

```
In [49]: import matplotlib.pyplot as plt
k_best = {}
for k in range(1,10):
    km = KMeans(n_clusters=k,random_state=0)
    km.fit(toronto_grouped.drop('Neighborhood',axis=1))
    toronto_grouped['Cluster'] = km.labels_
    k_best[k] = km.inertia_

plt.figure()
plt.plot(list(k_best.keys()), list(k_best.values()))
plt.xlabel("No. of cluster")
plt.ylabel("k-best")
plt.show()
```



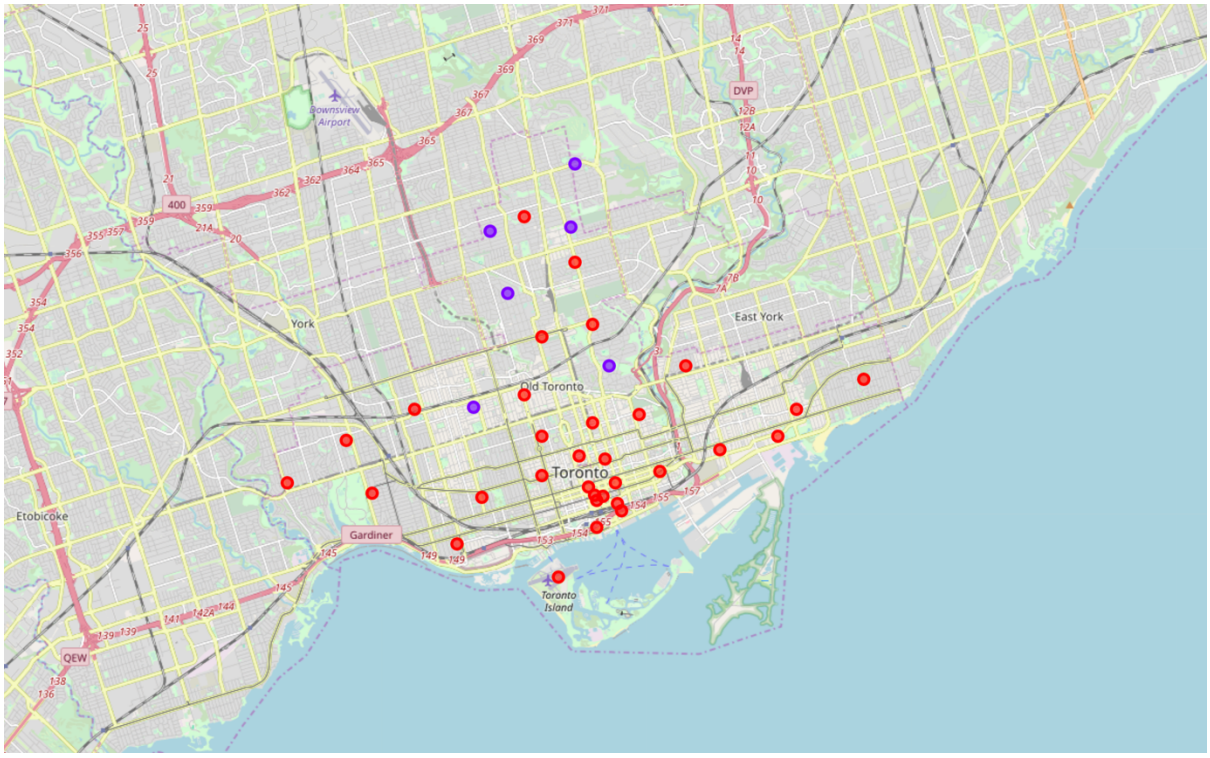
Bangalore:

In [48]:
```
import matplotlib.pyplot as plt
k_best = {}
for k in range(1,15):
    km = KMeans(n_clusters=k,random_state=0)
    km.fit(blore_grouped.drop('Neighborhood',axis=1))
    blore_grouped['Cluster'] = km.labels_
    k_best[k] = km.inertia_

plt.figure()
plt.plot(list(k_best.keys()), list(k_best.values()))
plt.xlabel("No. of cluster")
plt.ylabel("k-best")
plt.show()
print()
```

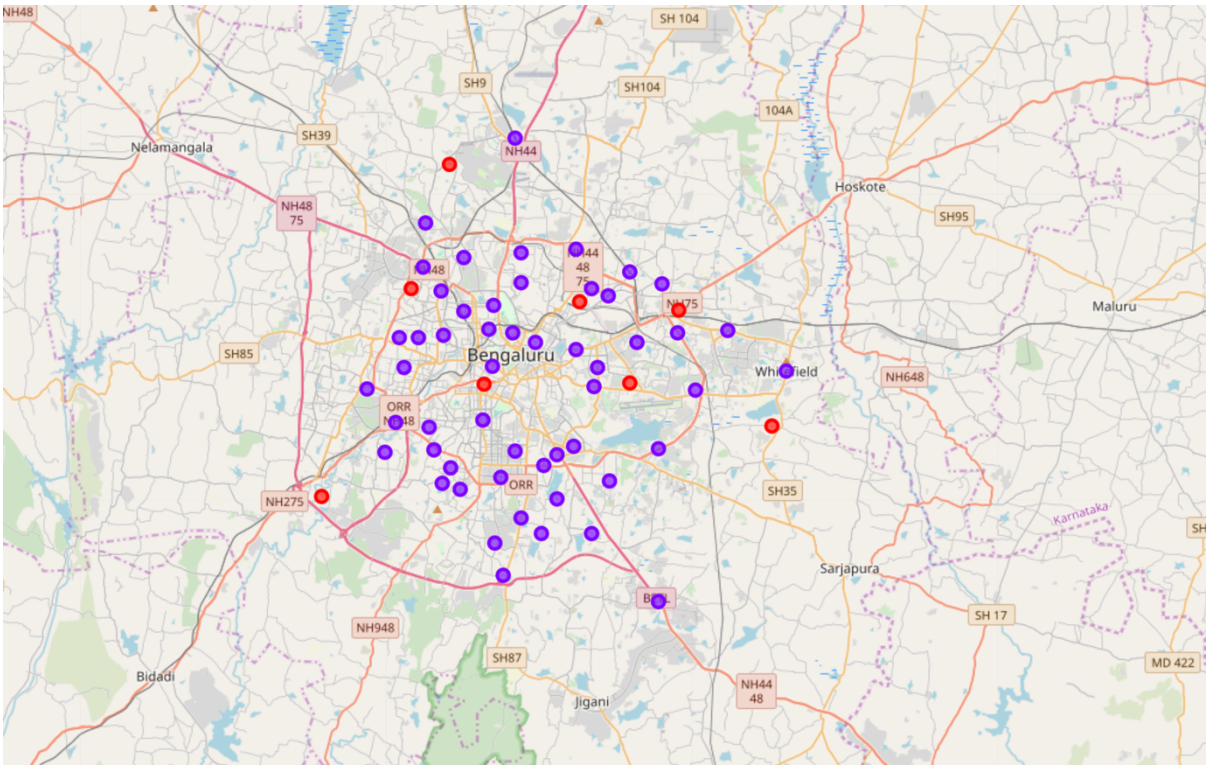


For both the cities, optimum cluster size is 2

Toronto Clusters:



Bangalore Clusters:



Results:

Cluster 1 for Toronto:

In [55]:
```
# Cluster 1 for Toronto
toronto_merged.loc[toronto_merged['Cluster Labels'] == 0, toronto_merged.columns[[2] + list(range(5, toronto_merged.shape[1]))]]
```

Out[55]:

	Neighbourhood	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	10t Co
37	The Beaches	0	Health Food Store	Trail	Pub	Dessert Shop	Falafel Restaurant	Event Space	Ethiopian Restaurant	Electronics Store	Eastern European Restaurant	Dumpli Restau
41	The Danforth West, Riverdale	0	Greek Restaurant	Coffee Shop	Italian Restaurant	Ice Cream Shop	Furniture / Home Store	Bookstore	Pub	Indian Restaurant	Sports Bar	Spa
42	The Beaches West, India Bazaar	0	Sushi Restaurant	Ice Cream Shop	Sandwich Place	Board Shop	Burger Joint	Burrito Place	Fast Food Restaurant	Fish & Chips Shop	Italian Restaurant	Steakh
43	Studio District	0	Café	Coffee Shop	Bakery	Italian Restaurant	American Restaurant	Bar	Fish Market	Coworking Space	Latin American Restaurant	Seafoo Restau
46	North Toronto West	0	Coffee Shop	Sporting Goods Shop	Clothing Store	Health & Beauty Service	Diner	Mexican Restaurant	Dessert Shop	Park	Gym / Fitness Center	Chines Restau
47	Davisville	0	Dessert Shop	Sandwich Place	Pizza Place	Coffee Shop	Gym	Sushi Restaurant	Café	Italian Restaurant	Restaurant	Pharma
48	Moore Park, Summerhill East	0	Playground	Trail	Wings Joint	Dessert Shop	Falafel Restaurant	Event Space	Ethiopian Restaurant	Electronics Store	Eastern European Restaurant	Dumpli Restau
49	Deer Park, Forest Hill SE, Rathnelly, South Hi...	0	Coffee Shop	Pub	Bagel Shop	Sports Bar	Vietnamese Restaurant	Liquor Store	Supermarket	Sushi Restaurant	Pizza Place	Light R Station
51	Cabbagetown, St. James Town	0	Coffee Shop	Café	Italian Restaurant	Restaurant	Pub	Bakery	Pizza Place	Market	Park	Japane Restau
52	Church and Wellesley	0	Coffee Shop	Gay Bar	Japanese Restaurant	Sushi Restaurant	Restaurant	Gastropub	Café	Hotel	Fast Food Restaurant	Mediter Restau
53	Harbourfront, Regent Park	0	Coffee Shop	Bakery	Park	Café	Mexican Restaurant	Restaurant	Pub	Gym / Fitness Center	Breakfast Spot	Theater

Cluster 2 for Toronto:

In [56]:
```
# Cluster 2 for Toronto
toronto_merged.loc[toronto_merged['Cluster Labels'] == 1, toronto_merged.columns[[2] + list(range(5, toronto_merged.sh
ape[1]))]]
```

Out[56]:

	Neighbourhood	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	10th Most Common Venue
44	Lawrence Park	1	Park	Swim School	Bus Line	Wings Joint	Diner	Falafel Restaurant	Event Space	Ethiopian Restaurant	Electronics Store	Eastern European Restaurant
45	Davisville North	1	Gym	Park	Convenience Store	Sandwich Place	Breakfast Spot	Clothing Store	Food & Drink Shop	Hotel	Wings Joint	Dumpling Restaurant
50	Rosedale	1	Park	Playground	Trail	Building	Wings Joint	Dumpling Restaurant	Discount Store	Dog Run	Doner Restaurant	Donut Shop
63	Roselawn	1	Garden	Home Service	Dim Sum Restaurant	Farmers Market	Falafel Restaurant	Event Space	Ethiopian Restaurant	Electronics Store	Eastern European Restaurant	Dumpling Restaurant
64	Forest Hill North, Forest Hill West	1	Jewelry Store	Sushi Restaurant	Park	Trail	Wings Joint	Discount Store	Dog Run	Doner Restaurant	Donut Shop	Eastern European Restaurant
75	Christie	1	Grocery Store	Café	Park	Restaurant	Nightclub	Italian Restaurant	Diner	Baby Store	Coffee Shop	Convenience Store

Cluster 1 for Bangalore:

In [79]:
```
#Cluster 1 for Blore
blore_merged.loc[blore_merged['Cluster Labels'] == 0, blore_merged.columns[[2] + list(range(5, blore_merged.shape
[1]))]]
```

Out[79]:

	Longitude	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
3	77.6595	Gym	Pizza Place	Donut Shop	Flea Market	Fish & Chips Shop	Falafel Restaurant	Electronics Store	Eastern European Restaurant	Women's Store
5	77.5758	Historic Site	Women's Store	Department Store	Dessert Shop	Dim Sum Restaurant	Diner	Donut Shop	Eastern European Restaurant	Electronics Store
15	77.6878	Women's Store	Cupcake Shop	Department Store	Dessert Shop	Dim Sum Restaurant	Diner	Donut Shop	Eastern European Restaurant	Electronics Store
18	77.7412	Women's Store	Electronics Store	Flower Shop	Flea Market	Fish & Chips Shop	Fast Food Restaurant	Falafel Restaurant	Eastern European Restaurant	Food Court
24	77.6306	Restaurant	Women's Store	Eastern European Restaurant	Flea Market	Fish & Chips Shop	Fast Food Restaurant	Falafel Restaurant	Electronics Store	Donut Shop
31	77.5561	Bus Station	Food Court	Department Store	Dessert Shop	Dim Sum Restaurant	Diner	Donut Shop	Eastern European Restaurant	Electronics Store
57	77.4827	Optical Shop	Indian Restaurant	Italian Restaurant	Electronics Store	Flower Shop	Flea Market	Fish & Chips Shop	Fast Food Restaurant	Falafel Restaurant
60	77.5338	ATM	Women's Store	Electronics Store	Flower Shop	Flea Market	Fish & Chips Shop	Fast Food Restaurant	Falafel Restaurant	Donut Shop

Cluster 2 for Bangalore:

```
In [80]: #Cluster 2 for Blore
blore_merged.loc[blore_merged['Cluster Labels'] == 1, blore_merged.columns[[2] + list(range(5, blore_merged.shape
[1]))]]
```

Out[80]:

	Longitude	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	77.580643	Bookstore	Seafood Restaurant	Restaurant	Women's Store	Falafel Restaurant	Flea Market	Fish & Chips Shop	Fast Food Restaurant	Electronics Store
1	77.638726	Café	Juice Bar	Sandwich Place	Pizza Place	BBQ Joint	Rajasthani Restaurant	Chinese Restaurant	Women's Store	Fast Food Restaurant
2	77.641151	Pub	Lounge	Ice Cream Shop	Indian Restaurant	Cocktail Bar	Restaurant	Clothing Store	Bakery	Italian Restaurant
4	77.564300	Breakfast Spot	Bakery	Indian Restaurant	Food & Drink Shop	Flower Shop	Flea Market	Fish & Chips Shop	Fast Food Restaurant	Falafel Restaurant
6	77.552500	Vegetarian / Vegan Restaurant	Snack Place	Chaat Place	Park	Indian Restaurant	Café	Women's Store	Fish & Chips Shop	Fast Food Restaurant
7	77.581300	Coffee Shop	Indian Restaurant	Ice Cream Shop	Café	Women's Store	Cafeteria	Plaza	Chinese Restaurant	Spa
8	77.578700	Asian Restaurant	Chinese Restaurant	Chaat Place	Food & Drink Shop	Lounge	Indian Restaurant	Health & Beauty Service	Flea Market	Dessert Shop
9	77.605700	Clothing Store	Tea Room	Donut Shop	South Indian Restaurant	Fast Food Restaurant	Market	Diner	Dim Sum Restaurant	Flower Shop
10	77.628600	Burger Joint	Bakery	Women's Store	Flower Shop	Flea Market	Fish & Chips Shop	Fast Food Restaurant	Falafel Restaurant	Food Court
11	77.592000	Coffee Shop	Chinese Restaurant	Hotel	Italian Restaurant	Steakhouse	Pizza Place	Pub	Fast Food Restaurant	Sandwich Place
12	77.676200	Café	Fast Food Restaurant	Kerala Restaurant	Breakfast Spot	Juice Bar	Multiplex	Coffee Shop	Pizza Place	Chinese Restaurant
13	77.663900	Indian Restaurant	Park	Shop & Service	Department Store	Eastern European Restaurant	Fish & Chips Shop	Fast Food Restaurant	Falafel Restaurant	Electronics Store

Discussion:

Based on cluster for each cities above, we believe that classification for each cluster can be done better with calculation of venues categories (most common) in each cities. Referring to each cluster, we can't determine clearly what represent in each cluster by using Foursquare.

However, we assumed each cluster as follow:

Cluster 1: ***Toronto:*** *Restaurants and Cafe*

Cluster 2: ***Toronto:*** *Tourist spots for roaming*

Cluster 1: ***Bangalore:*** *Tourist spots or places*

Cluster 2: ***Bangalore:*** *Mostly Restaurants and other shops*

What is lacking at this point is a systematic, quantitative way to identify and distinguish different district and to describe the correlation most common venues as recorded in Foursquare. The reality is however more complex: similar cities might have or might not have similar common venues. A further step in this classification would be to find a method to extract these common venues and integrate the spatial correlations between different of areas or district.

Conclusion:

Using Foursquare API, we can captured data of common places all around the world. Using it, we refer back to our main objectives, which is to determine the similarity or dissimilarity of both cities classification of area located inside the city whether it is residential, tourism places, or others.

In conclusion, both cities Toronto and Bangalore are the centre of attraction among peoples. However, to declare both cities are similar or dissimilar base on common venues visited is quite difficult. Both cities is similar in some venues also dissimilar in certain venues. And for classification based on common venues, again we must have more systematic or quantitative way to identify and declare this. Comparison can be made, but no such method or quantitative data to determine this.