Introduction

Mumbai and Delhi are two major cities in India. Delhi is national capital of India while Mumbai is very well known as economic captial of India. Both cities are very popular in world for jobs, tourism, livelihood etc. Lots of people come around these cities to make career in these.

Problem Statement ¶

We will study area classification using Foursquare API data and ML segmentation and clustering. The aim is to segment areas of Delhi and Mumbai based on the most common places captured from Foursquare in India which can help someone to identify place according to their needs, interests for residence.

Using segmentation and clustering, we hope we can determine:

1) the similarity or dissimilarirty of both cities 2) Area classificaotion located inside the city whether it is residential, tourism places, or others

Data

We needed the data for each city. We acquired the areas and picodes from following two links.

-- For Mumbai (https://www.mapsofindia.com/pincode/india/maharashtra/mumbai/ (https://www.mapsofindia.com/pincode/india/maharashtra/mumbai/))

-- For Delhi (https://www.whatsuplife.in/delhi/blog/zip-pin-postal-code-pincodes-delhi/ (https://www.whatsuplife.in/delhi/blog/zip-pin-postal-code-pincodes-delhi/))

These HTML pages will be parsed to fetch the area pincodes and will be converted to csv.

Data is in form of Area along with their Pincodes for each city. We will further fetch the lattitude and longitude for each areas and store to a DataFrame for analysis and also to a separate CSV file to avoid scrapping again.

This data (Area, Pincode, City, Latitude, Longitude) will be help to identify common places using FS API.

Following is a sample of dataframe for Mumbai City with latitude and logntidue

```
df.head(10)
```

	Area	Pincode	District	Latitude	Longitude
2	Agripada	400011	Mumbai	18.975302	72.824898
3	Airport	400099	Mumbai	19.090201	72.863808
4	Ambewadi	400004	Mumbai	19.186776	72.859313
5	Andheri	400053	Mumbai	19.120371	72.848043
6	Andheri East	400069	Mumbai	19.115883	72.854202
7	Andheri Railway station	400058	Mumbai	19.120371	72.848043
8	Antop Hill	400037	Mumbai	19.020761	72.865256
9	Asvini	400005	Mumbai	18.900867	72.815941
10	Azad Nagar	400053	Mumbai	19.165798	72.955893
11	B P t colony	400003	Mumbai	19.101937	72.861599

Following is the sample of dataframe for Delhi City

```
dfd.head(10)
```

	Area	Pincode	District	Latitude	Longitude
5	Bank Street	110005	Central Delhi	28.651718	77.221939
6	Baroda House	110001	Central Delhi	28.615804	77.230020
7	Bengali Market	110001	Central Delhi	28.629465	77.232185
9	Connaught Place	110001	Central Delhi	28.631383	77.219792
13	Delhi High court	110003	Central Delhi	28.608764	77.236435
14	Desh Bandhu gupta road	110005	Central Delhi	28.645182	77.223261
15	Election Commission	110001	Central Delhi	28.448733	77.028976
17	Guru Gobind singh marg	110005	Central Delhi	28.533065	77.208283
20	I.A.r.i.	110012	Central Delhi	28.648948	77.183660
21	I.P.estate	110002	Central Delhi	28.614868	77.196937

Methodology

Above, we have done convert addresses into their equivalent latitude and longitude values. Then we will use the Foursquare API to explore neighborhoods in both cities, Mumbai and Delhi

After that, explore function to get the most common venue categories in each neighborhood, and then use this feature to group the neighborhoods into clusters

K-means clustering algorithm will be use to complete this task. And also, the Folium library to visualize the neighborhoods in Mumbai and Delhi and their emerging clusters.

Based on dataframe analysis above, we found out that 400001 area in Mumbai and 110001 area in Delhi are both have the highest number of area within cities.

Following is the list of different Venues and Categories found for 400001

```
Bazargate
M.P.t.
Mumbai.
Stock Exchange
Town Hall
(460, 7)
```

	Area	Area Latitude	Area Longitude	Venue	Venue Latitude	Venue Longitude	Venue Category
0	Bazargate	18.928665	72.832264	Trishna	18.928619	72.832356	Seafood Restaurant
1	Bazargate	18.928665	72.832264	Jehangir Art Gallery	18.927606	72.831464	Art Gallery
2	Bazargate	18.928665	72.832264	Chhatrapati Shivaji Maharaj Vastu Sangrahalaya...	18.926864	72.832559	History Museum
3	Bazargate	18.928665	72.832264	Sabyasachi	18.926990	72.833185	Boutique
4	Bazargate	18.928665	72.832264	Kala Ghoda Café	18.928515	72.832354	Café

Following is the list of different venues and cateogries found for 110001

```
Baroda House
Bengali Market
Connaught Place
Election Commission
Janpath
Krishi Bhawan
North Avenue
Parliament House
Patiala House
Pragati Maidan
Rail Bhawan
Sansad Marg
Secretariat North
Shastri Bhawan
Supreme Court
(417, 7)
```

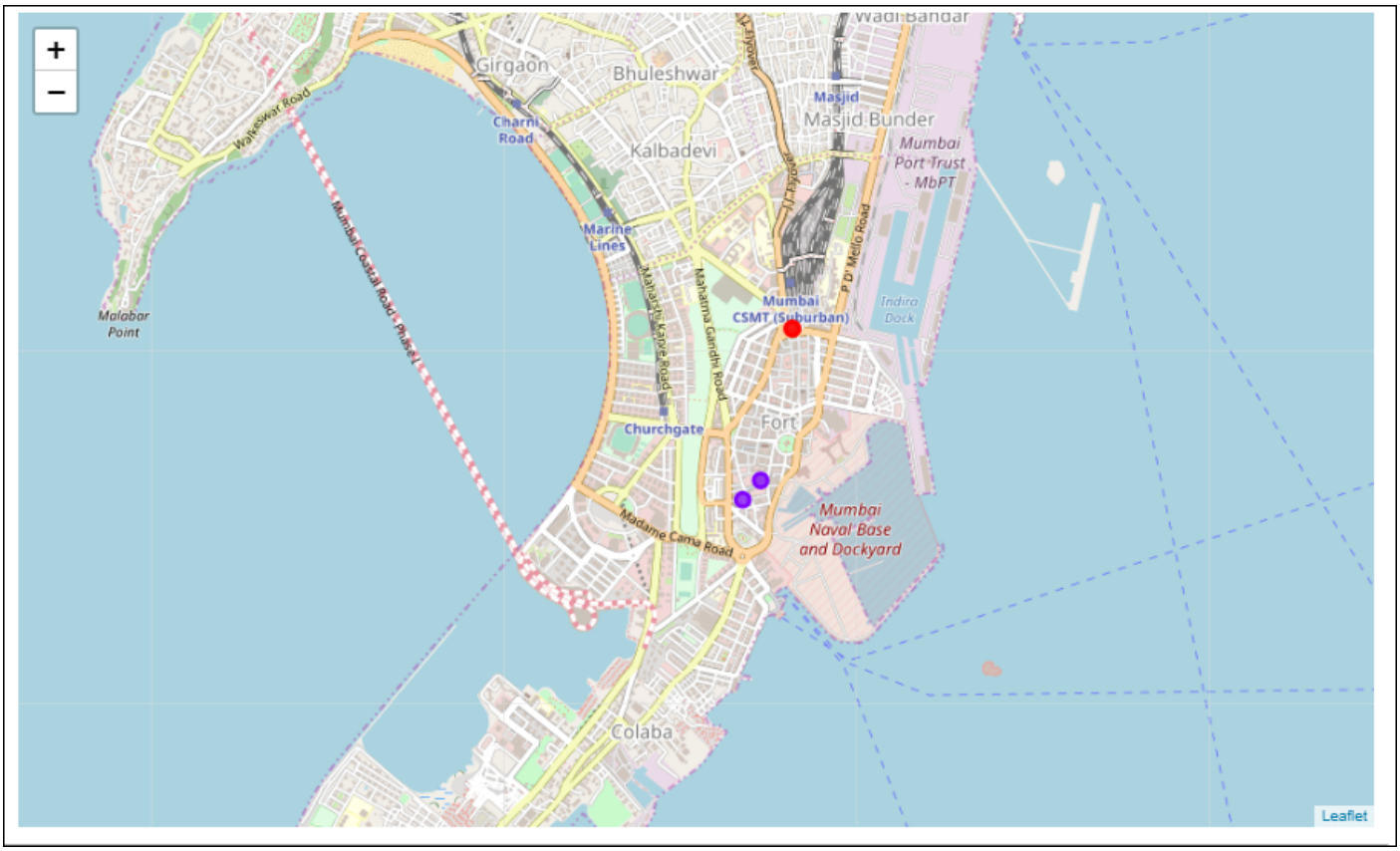
	Area	Area Latitude	Area Longitude	Venue	Venue Latitude	Venue Longitude	Venue Category
0	Baroda House	28.615804	77.23002	India Gate | इंडिया गेट (India Gate)	28.612796	77.229207	Monument / Landmark
1	Baroda House	28.615804	77.23002	Amar Jawan Jyoti | अमर जवान ज्योति (Amar Jawan...	28.612980	77.228247	Sculpture Garden
2	Baroda House	28.615804	77.23002	Andhra Bhavan Canteen	28.617095	77.225721	Indian Restaurant
3	Baroda House	28.615804	77.23002	Gulati Restaurant	28.608010	77.229989	Indian Restaurant
4	Baroda House	28.615804	77.23002	National Gallery Of Modern Art | राष्ट्रीय आधु...	28.609411	77.234585	Art Gallery

Analysing the Mumbai and Delhi area to get most common places by Foursquare API

Mumbai - 4000001

	Area	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
0	Bazargate	Indian Restaurant	Café	Chinese Restaurant	Bakery	Bar	Seafood Restaurant	Coffee Shop	Pizza Place
1	M.P.t.	Indian Restaurant	Restaurant	Bakery	Coffee Shop	Ice Cream Shop	Chinese Restaurant	Café	Bar
2	Mumbai.	Indian Restaurant	Café	Fast Food Restaurant	Train Station	Bakery	Coffee Shop	Seafood Restaurant	Multiplex
3	Stock Exchange	Indian Restaurant	Café	Chinese Restaurant	Seafood Restaurant	Bar	Coffee Shop	Fast Food Restaurant	Bakery
4	Town Hall	Indian Restaurant	Café	Fast Food Restaurant	Train Station	Bakery	Coffee Shop	Seafood Restaurant	Multiplex

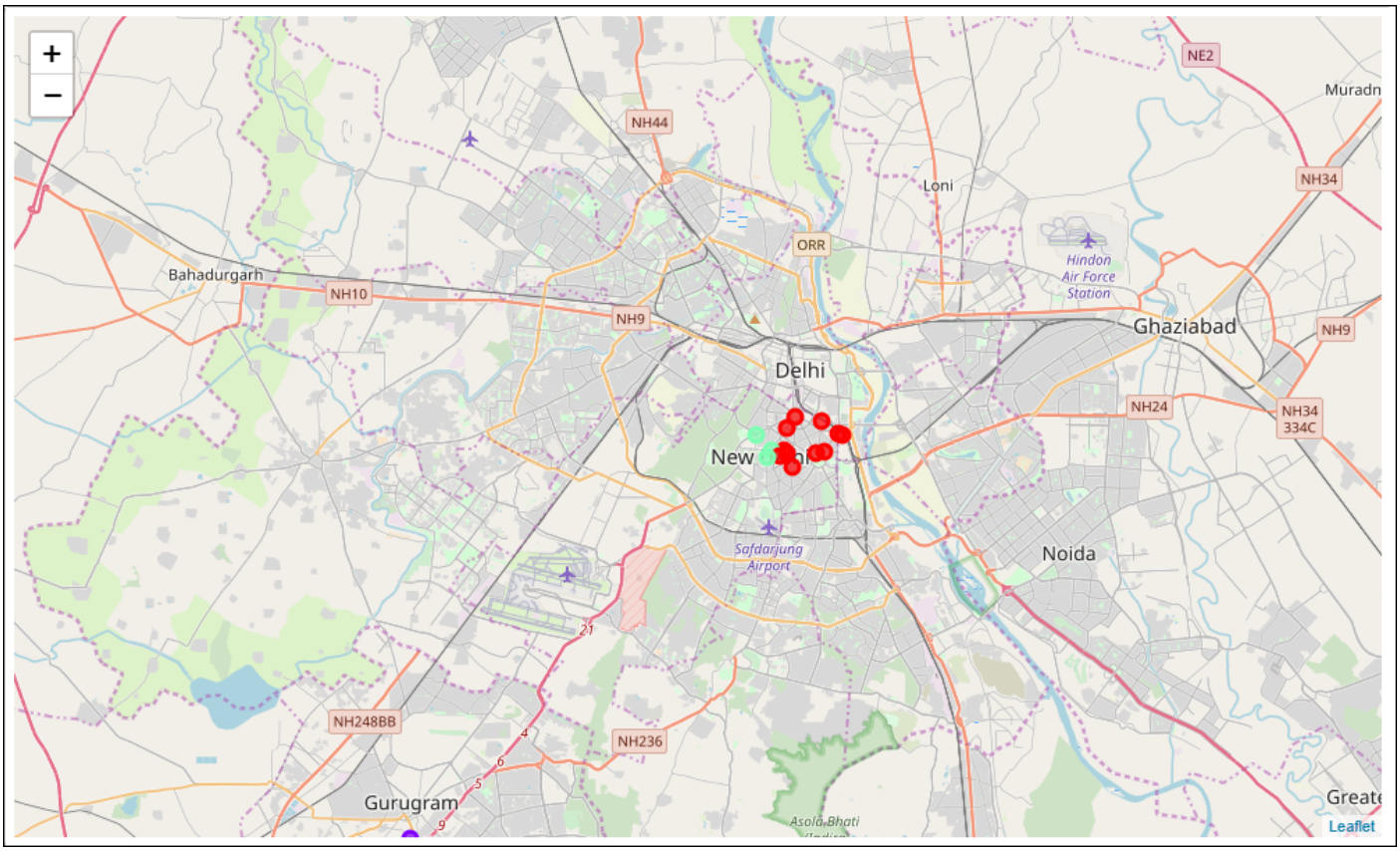
	Area	Pincode	District	Latitude	Longitude	Cluster Labels	1st Most Common Venue	2nd Most Common Venue	3rd Most Common Venue	4th Most Common Venue	5th Most Common Venue	6th Most Common Venue	7th Most Common Venue	8th Most Common Venue
0	Bazargate	400001	Mumbai	18.928665	72.832264	1	Indian Restaurant	Café	Chinese Restaurant	Bakery	Bar	Seafood Restaurant	Coffee Shop	Pizza Place
1	M.P.t.	400001	Mumbai	19.121958	72.909294	2	Indian Restaurant	Restaurant	Bakery	Coffee Shop	Ice Cream Shop	Chinese Restaurant	Café	Bar
2	Mumbai.	400001	Mumbai	18.938771	72.835335	0	Indian Restaurant	Café	Fast Food Restaurant	Train Station	Bakery	Coffee Shop	Seafood Restaurant	Multiplex
3	Stock Exchange	400001	Mumbai	18.929832	72.833361	1	Indian Restaurant	Café	Chinese Restaurant	Seafood Restaurant	Bar	Coffee Shop	Fast Food Restaurant	Bakery
4	Town Hall	400001	Mumbai	18.938771	72.835335	0	Indian Restaurant	Café	Fast Food Restaurant	Train Station	Bakery	Coffee Shop	Seafood Restaurant	Multiplex



Delhi - 110001

	Area	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
0	Baroda House	Indian Restaurant	Smoke Shop	Pool	Chinese Restaurant	Sculpture Garden	Food & Drink Shop	Concert Hall	Furniture / Home Store
1	Bengali Market	Theater	Café	Hotel	Indian Restaurant	Art Gallery	Bakery	Salon / Barbershop	Historic Site
2	Connaught Place	Indian Restaurant	Café	Hotel	Coffee Shop	Chinese Restaurant	Bar	Lounge	Fast Food Restaurant
3	Election Commission	Tourist Information Center	Hotel	Lawyer	Golf Course	Wine Bar	Government Building	Flea Market	Food & Drink Shop
4	Janpath	Indian Restaurant	Hotel	History Museum	Hotel Bar	Restaurant	Japanese Restaurant	Government Building	Jewelry Store

	Area	Pincode	District	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
0	Baroda House	110001	Central Delhi	28.615804	77.230020	0	Indian Restaurant	Smoke Shop	Pool	Chinese Restaurant	Sculpture Garden	Food & Drink Shop	Concert Hall	Furniture / Home Store
1	Bengali Market	110001	Central Delhi	28.629465	77.232185	0	Theater	Café	Hotel	Indian Restaurant	Art Gallery	Bakery	Salon / Barbershop	Historic Site
2	Connaught Place	110001	Central Delhi	28.631383	77.219792	0	Indian Restaurant	Café	Hotel	Coffee Shop	Chinese Restaurant	Bar	Lounge	Fast Food Restaurant
3	Election Commission	110001	Central Delhi	28.448733	77.028976	1	Tourist Information Center	Hotel	Lawyer	Golf Course	Wine Bar	Government Building	Flea Market	Food & Drink Shop
4	Janpath	110001	Central Delhi	28.610086	77.218247	0	Indian Restaurant	Hotel	History Museum	Hotel Bar	Restaurant	Japanese Restaurant	Government Building	Jewelry Store


NE2
Muradn
NH44
Loni
NH34
ORR
Hindon
Air Force
Station
Bahadurgarh
NH10
NH9
Ghaziabad
NH9
Delhi
NH24
NH34
334C
New
Safdarjung
Airport
Noida
NH248BB
NH236
Gurugram
Asola Bhati
Leaflet

Result

```
#Cluster 1 for Mumbai
m_merged.loc[m_merged['Cluster Labels'] == 0, m_merged.columns[[2] + list(range(5, m_merged.shape[1]))]]
```

	District	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
2	Mumbai	0	Indian Restaurant	Café	Fast Food Restaurant	Train Station	Bakery	Coffee Shop	Seafood Restaurant	Multiplex
4	Mumbai	0	Indian Restaurant	Café	Fast Food Restaurant	Train Station	Bakery	Coffee Shop	Seafood Restaurant	Multiplex

```
#Cluster 2 for Mumbai
m_merged.loc[m_merged['Cluster Labels'] == 1, m_merged.columns[[2] + list(range(5, m_merged.shape[1]))]]
```

	District	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
0	Mumbai	1	Indian Restaurant	Café	Chinese Restaurant	Bakery	Bar	Seafood Restaurant	Coffee Shop	Pizza Place
3	Mumbai	1	Indian Restaurant	Café	Chinese Restaurant	Seafood Restaurant	Bar	Coffee Shop	Fast Food Restaurant	Bakery

```
#Cluster 3 for Mumbai
m_merged.loc[m_merged['Cluster Labels'] == 2, m_merged.columns[[2] + list(range(5, m_merged.shape[1]))]]
```

	District	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
1	Mumbai	2	Indian Restaurant	Restaurant	Bakery	Coffee Shop	Ice Cream Shop	Chinese Restaurant	Café	Bar

```
#Cluster 1 for Delhi
d_merged.loc[d_merged['Cluster Labels'] == 0, d_merged.columns[[2] + list(range(5, d_merged.shape[1]))]]
```

	District	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
0	Central Delhi	0	Indian Restaurant	Smoke Shop	Pool	Chinese Restaurant	Sculpture Garden	Food & Drink Shop	Concert Hall	Furniture / Home Store
1	Central Delhi	0	Theater	Café	Hotel	Indian Restaurant	Art Gallery	Bakery	Salon / Barbershop	Historic Site
2	Central Delhi	0	Indian Restaurant	Café	Hotel	Coffee Shop	Chinese Restaurant	Bar	Lounge	Fast Food Restaurant
4	Central Delhi	0	Indian Restaurant	Hotel	History Museum	Hotel Bar	Restaurant	Japanese Restaurant	Government Building	Jewelry Store
5	Central Delhi	0	Indian Restaurant	Hotel	Restaurant	Spa	History Museum	Hotel Bar	Lounge	Hotel Pool
8	Central Delhi	0	Art Gallery	Indian Restaurant	Pool	Sculpture Garden	Snack Place	Park	Concert Hall	Furniture / Home Store
9	Central Delhi	0	Theater	Pool	Coffee Shop	Chinese Restaurant	Udupi Restaurant	Train Station	Art Gallery	Art Museum
10	Central Delhi	0	Hotel	Spa	Indian Restaurant	Chinese Restaurant	History Museum	Restaurant	Wine Bar	Music Venue
11	Central Delhi	0	Hotel	Indian Restaurant	Café	Chinese Restaurant	Bar	Coffee Shop	Lounge	Italian Restaurant
13	Central Delhi	0	Indian Restaurant	Hotel	Restaurant	History Museum	Spa	Government Building	Music Venue	Hotel Pool
14	Central Delhi	0	Art Gallery	Theater	Café	Arcade	Burger Joint	Furniture / Home Store	Concert Hall	Performing Arts Venue

```
#Cluster 2 for Delhi
d_merged.loc[d_merged['Cluster Labels'] == 1, d_merged.columns[[2] + list(range(5, d_merged.shape[1]))]]
```

	District	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
3	Central Delhi	1	Tourist Information Center	Hotel	Lawyer	Golf Course	Wine Bar	Government Building	Flea Market	Food & Drink Shop

```
#Cluster 3 for Delhi
d_merged.loc[d_merged['Cluster Labels'] == 2, d_merged.columns[[2] + list(range(5, d_merged.shape[1]))]]
```

	District	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
6	Central Delhi	2	Indian Restaurant	Asian Restaurant	Theater	Stadium	Spiritual Center	Garden	Smoke Shop	Historic Site
7	Central Delhi	2	Government Building	Garden	Hotel Bar	Food & Drink Shop	Hostel	Music Venue	History Museum	Spiritual Center
12	Central Delhi	2	Music Venue	Museum	Tea Room	Spiritual Center	Garden	Government Building	Wine Bar	Golf Course

Discussion

Based on cluster for each cities, we believe that classification for each cluster can be done better with calculation of venues categories (most common) in each cities. Refering to each clsuter, we can't deterimine clearly what represent in each cluster by using Foursquare - Most Common Venue data.

However, for the sae of this project we assumed each cluster as follow:

-- Cluster 1: Mumbai: Tourism -- Cluster 2: Mumbai: Residental -- Cluster 3: Mumbai: Mix -- ster 1: Delhi: Residental -- ster 2: Delhi: Tourism -- ster 3: Delhi: Sport

It is tough to identify 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.

We believe that the classification we propose is an encouraging step towards a quantitative and systematic comparison of the different cities. Further studies are indeed needed in order to relate the data acquired, then observe it to more meaningful and objective results.

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 dissimilarirty of both cities classification of area located inside the city whether it is residential, tourism places, or others In conclusion, both cities Mumbai and Delhi are the center of attraction in India. 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 classitification 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. We hope in the future, a method to determine it can be establish and explore for references.