Coursera Capstone Project

Coursera IBM Data scientist Certification

ThangarajDeivasikamani

26-JAN-2019

Project Brief

In this project, we will study in details the area classification using Foursquare data and machine learning segmentation and clustering. The aim of this project is to segment areas of Bangalore and Chennai based on the most common places captured from Foursquare.

Using segmentation and clustering, we hope we can determine:

- The similarity or dissimilarirty of both cities
- classification of venues located inside wheher its Restaurant,shops,Hotel, Clothstore,etc..

Table of Contents

1.Introduction

Bangalore and Chennai are two major cities in India. Both cities are south part of Indina and become a more attention for residential, job employment, tourism, education, shopping and sports activity. Both cities are well known in India, and become the top choice for local and foreign communities.

Brief information about both cities:

Bangalore: Bangalore officially known as Bengaluru is the capital city of the South Indian state of Karnataka.It has a population of over ten million, making it a megacity and the third most populous city and fifth most populous urban agglomeration in India.Bengaluru the Silicon valley of India is one of the best city for IT professionals. It is home to many educational and research institutions in India.(source: https://en.wikipedia.org/wiki/Bangalore)

Chennai: Chennai also known by its former name Madras is the capital of the Indian state of Tamil Nadu. Located on the Coromandel Coast off the Bay of Bengal, it is the biggest cultural, economic and educational centre of south India.According to the 2011 Indian census, it is the sixth most populous city and fourth-most populous urban agglomeration in India.It is also known as the Detroit of India because of its automobile industry which produces more than 40% of the auto parts and motor vehicles in India. (source: https://en.wikipedia.org/wiki/Chennai)

2. Business Problem/Objective

In this project, we will study in details the area classification using Foursquare data and machine learning segmentation and clustering. The aim of this project is to segment areas of Bangalore and Chennai based on the most common places captured from Foursquare.

Using segmentation and clustering, we hope we can determine:

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

3. Data

The data (Postal Code) acquired from the following webpage pages:

india's open dataset - https://data.gov.in/catalog/locality-based-pincode can be used to fetch the areas of Bangalore and chennai

- About the dataset

you can use the temporary key to fetch the subset of information but if you want to have a full look at the full data you would need to login to the website a key is provided for each user which can be used to fetch the data from the API the data can be fetched in xml/json/csv format for our analysid we would be fetching the csv format data

- Alternate source of Location information(Only postal code)

1. https://www.mapsofindia.com/pincode/india/karnataka/bangalore/
2. https://www.mapsofindia.com/pincode/india/tamil-nadu/chennai/

The above data are restructure to csv file for easier manipulation and reading.

- latitude and longitude

Will be download using Google api.
Another aspect to consider for this project is the Foursquare data. I believe that the data as good as provided, meaning although we are using Foursquare data for segmentation and clustering, the amount and accuracy of data captured can't 100% determine correct classification in real world.

To start, let's get and look at the data.

Bangalore data

Out[5]:

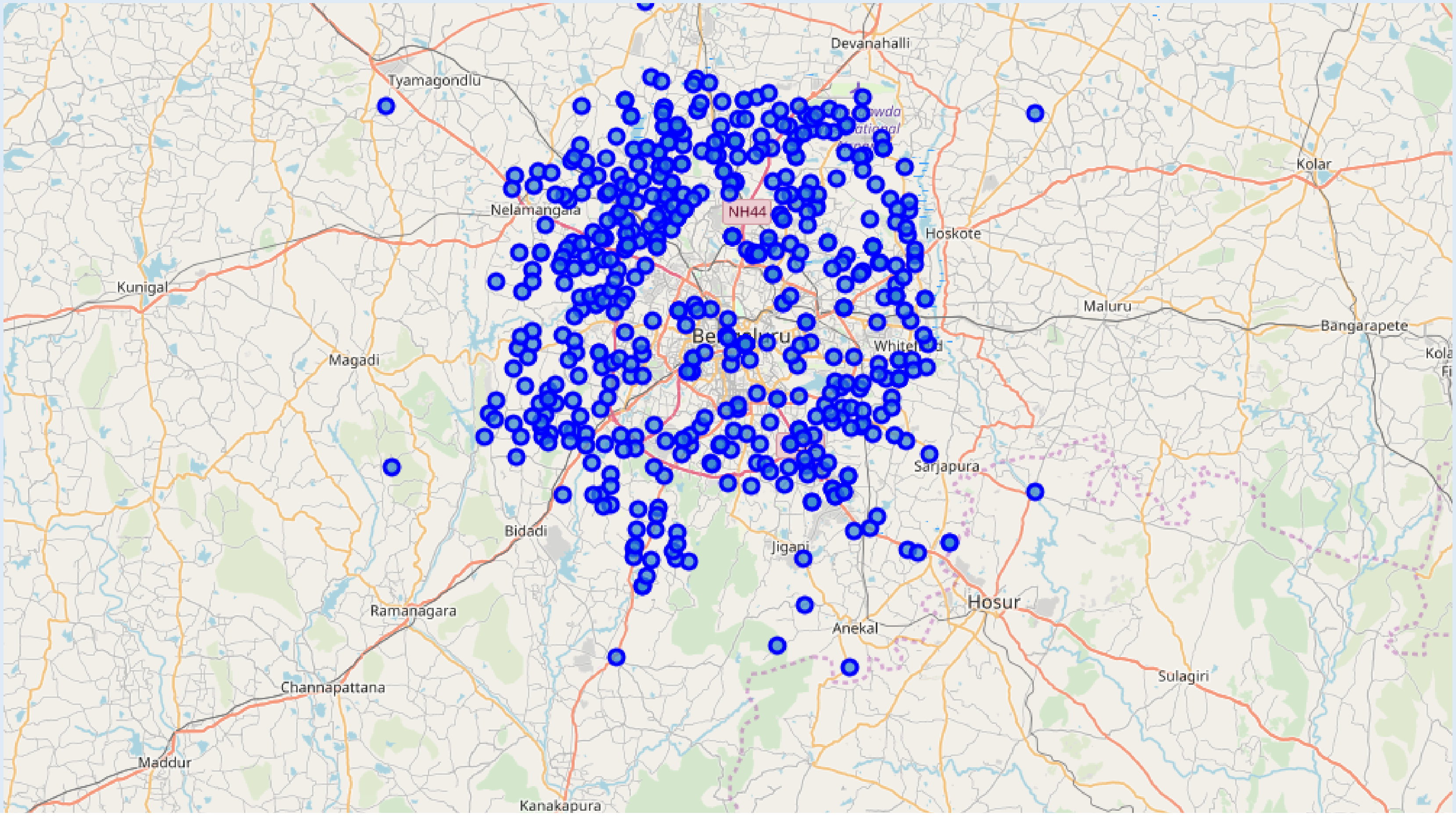
	Location	Officename	Pincode	Sub-distname	Districtname	StateName	Latitude	Longitude
0	Bangalore North	Vidhana Soudha S.O	560001	Bangalore North	BANGALORE	KARNATAKA	12.971599	77.594563
1	Bangalore North	HighCourt S.O	560001	Bangalore North	BANGALORE	KARNATAKA	12.971599	77.594563
2	Bangalore North	Dr. Ambedkar Veedhi S.O	560001	Bangalore North	BANGALORE	KARNATAKA	12.971599	77.594563
3	Bangalore North	Legislators Home S.O	560001	Bangalore North	BANGALORE	KARNATAKA	12.971599	77.594563
4	Bangalore North	Rajbhavan S.O (Bangalore)	560001	Bangalore North	BANGALORE	KARNATAKA	12.971599	77.594563
5	Bangalore North	Mahatma Gandhi Road S.O	560001	Bangalore North	BANGALORE	KARNATAKA	12.971599	77.594563
6	Bangalore North	Bangalore Bazaar S.O	560001	Bangalore North	BANGALORE	KARNATAKA	12.971599	77.594563
7	Bangalore North	Bangalore G.P.O.	560001	Bangalore North	BANGALORE	KARNATAKA	12.971599	77.594563
8	Bangalore North	Bangalore City S.O	560002	Bangalore North	BANGALORE	KARNATAKA	12.971599	77.594563
9	Bangalore South	Bangalore Corporation Building S.O	560002	Bangalore South	BANGALORE	KARNATAKA	12.971599	77.594563

Chennai data

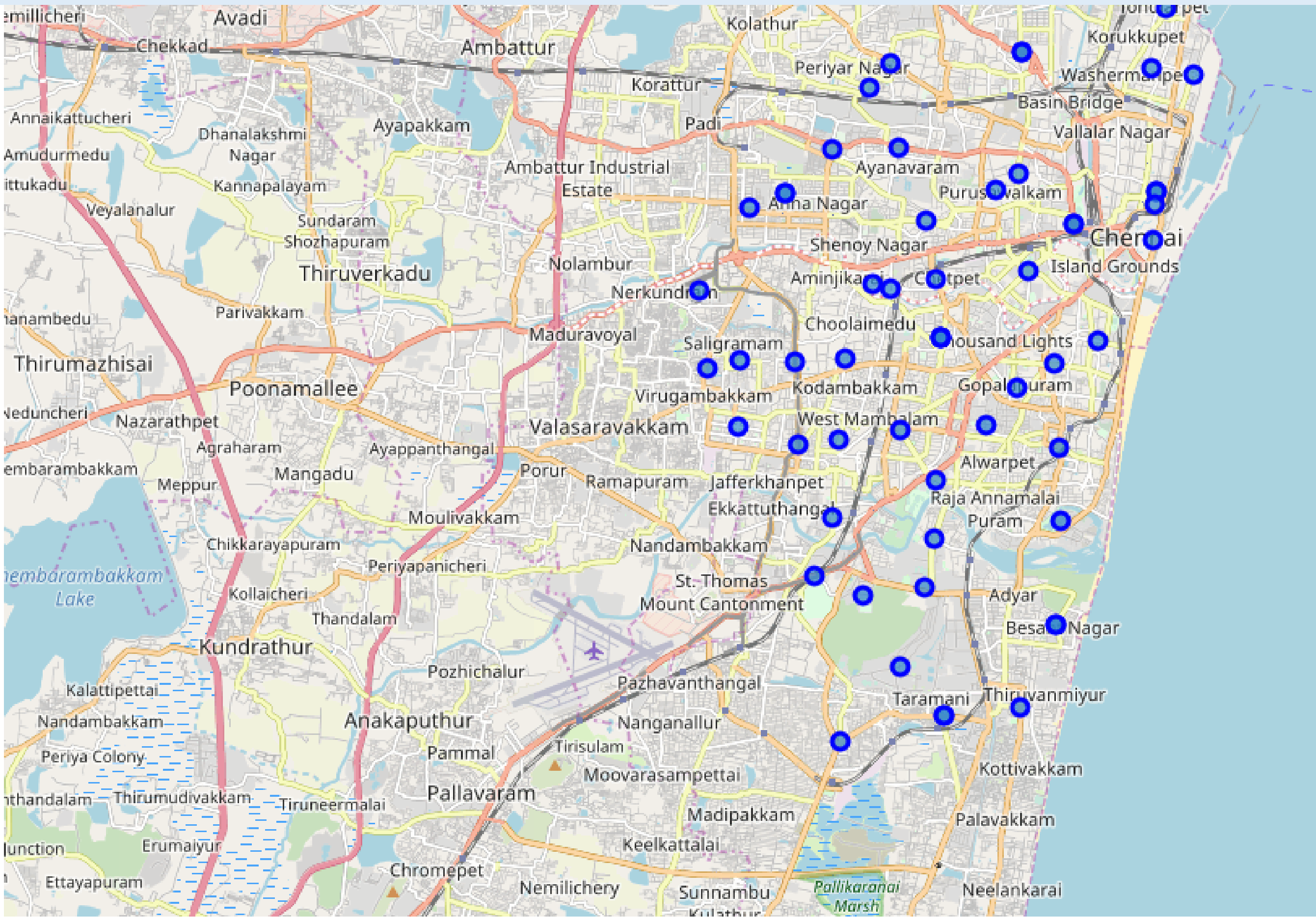
Out[6]:

	Location	Officename	Pincode	Sub-distname	Districtname	StateName	Latitude	Longitude
0	Parrys	Chennai G.P.O.	600001	Fort - Tondiarpet	CHENNAI	TAMIL NADU	13.089576	80.288228
1	Chennai	Anna Road H.O	600002	Egmore - Nungambakkam	CHENNAI	TAMIL NADU	13.082680	80.270718
2	Parrys	Park Town H.O	600003	Fort - Tondiarpet	CHENNAI	TAMIL NADU	13.089576	80.288228
3	Mylapore	Mylapore H.O	600004	Mylapore - Triplicane	CHENNAI	TAMIL NADU	13.036791	80.267630
4	Tiruvallikkeni	Tiruvallikkeni S.O	600005	Mylapore - Triplicane	CHENNAI	TAMIL NADU	13.058711	80.275706
5	NUNGAMBAKKAM	Greams Road S.O	600006	Egmore - Nungambakkam	CHENNAI	TAMIL NADU	13.059537	80.242479
6	Vyasarpadi	Vepery S.O	600007	Fort - Tondiarpet	CHENNAI	TAMIL NADU	13.118319	80.259439
7	Egmore	Egmore S.O	600008	Egmore - Nungambakkam	CHENNAI	TAMIL NADU	13.073226	80.260921
8	Fort St George	Fort St George S.O	600009	Fort - Tondiarpet	CHENNAI	TAMIL NADU	13.079644	80.287449
9	Kilpauk	Kilpauk S.O	600010	Perambur - Purasawakkam	CHENNAI	TAMIL NADU	13.083607	80.239206

Bangalore data map



Chennai data map



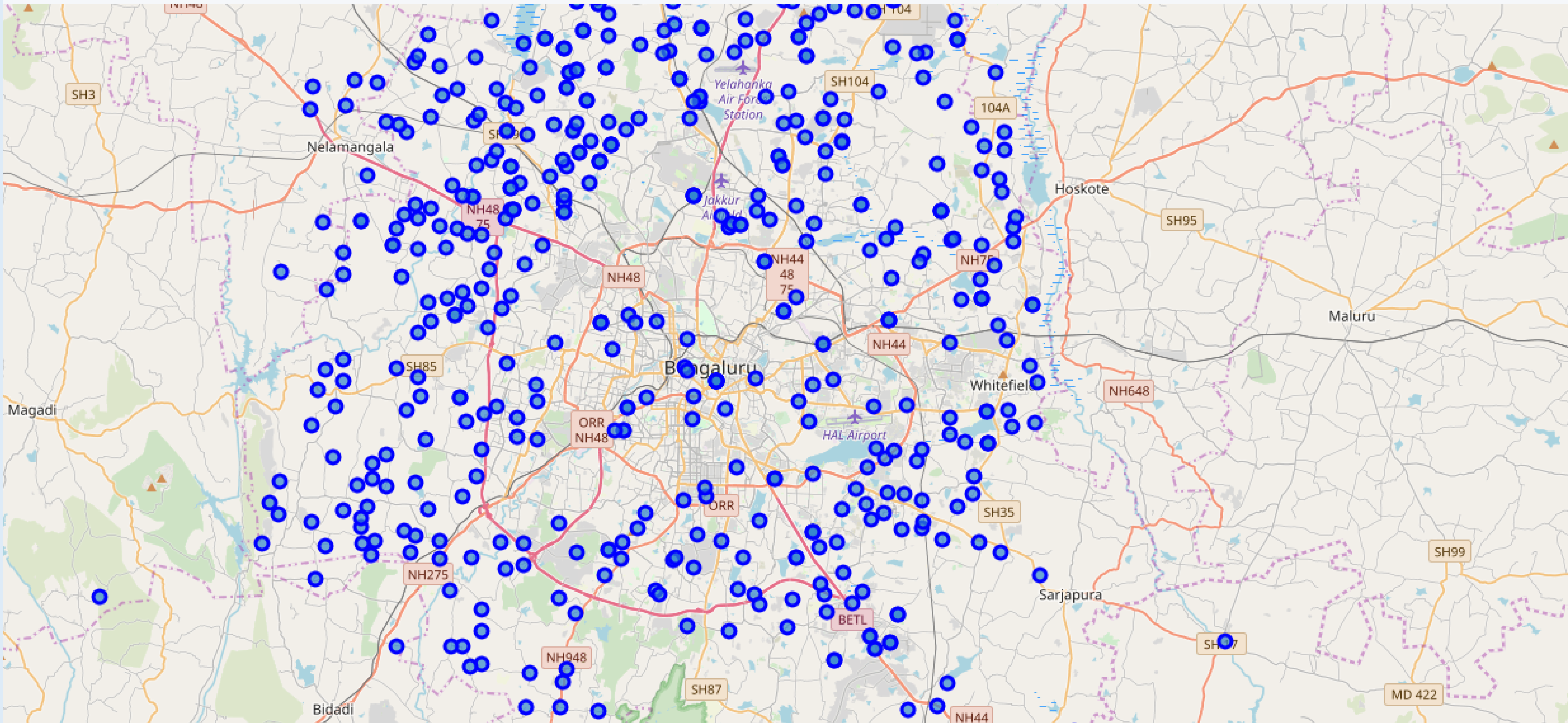
4. Methodology

In this project, I will use the K-Mean clusters.

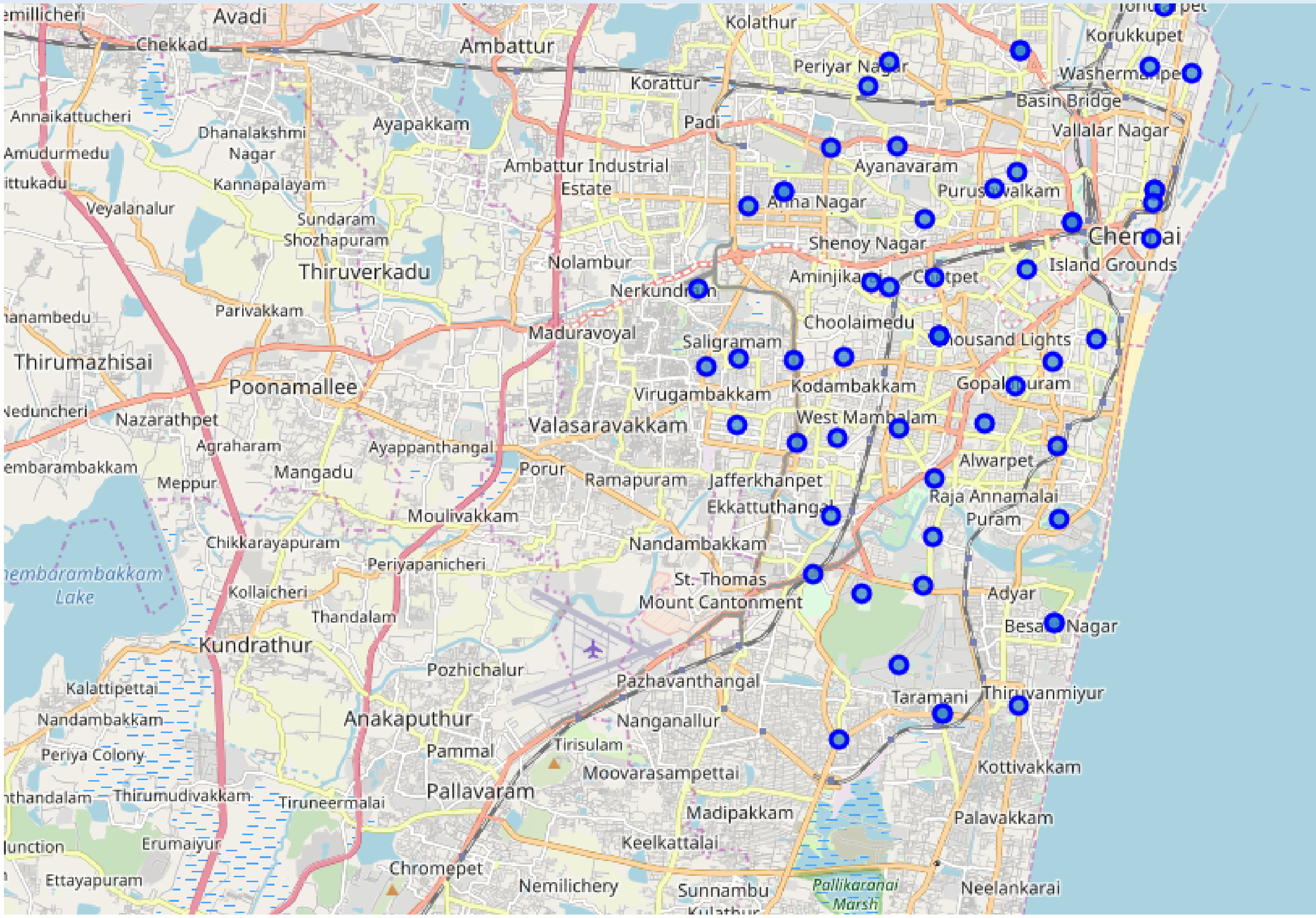
Above, we have done convert postal code and addresses into their equivalent latitude and longitude values.
Then we will use the Foursquare API to explore neighborhoods in both cities, Bangalore and Chennai
After that, explore function to get the most common venue categories in each neighborhood,
and 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 Bangalore and Chennai and their emerging clusters.

Based on dataframe analysis above, Bangalore south have the highest number of location within it Banaglore city.

Bangalore sub set data map



Chennai subset data map



5.Foursquare API Explore Function

Explore the Bangalore South location venues

In [25]:
```
import requests # library to handle requests
from pandas.io.json import json_normalize # tranform JSON file into a pandas dataframe

#explore the first neighborhood in our dataframe
#Get the neighborhood's latitude and longitude values.
neighborhood_latitude = bangalore_south_data.loc[0, 'Latitude'] # neighborhood latitude value
neighborhood_longitude = bangalore_south_data.loc[0, 'Longitude'] # neighborhood longitude value
neighborhood_name = bangalore_south_data.loc[0, 'Location'] # neighborhood name

#get the top 100 venues that are in bangalore_south within a radius of 1000 meters
LIMIT = 100 # limit of number of venues returned by Foursquare API
radius = 1000 # define radius
url = 'https://api.foursquare.com/v2/venues/explore?&client_id={}&client_secret={}&v={}&ll={},{}&radius={}&limit={}'.format(
    CLIENT_ID,
    CLIENT_SECRET,
    VERSION,
    neighborhood_latitude,
    neighborhood_longitude,
    radius,
    LIMIT)

#Send the GET request and examine the resutls
results = requests.get(url).json()
```

6.Explore a All venues and cat

In [27]:
```
#function to repeat the same process to all area
def getNearbyVenues(names, latitudes, longitudes, radius=500):

    venues_list=[]
    for name, lat, lng in zip(names, latitudes, longitudes):
        print(name)

        # create the API request URL
        url = 'https://api.foursquare.com/v2/venues/explore?&client_id={}&client_secret={}&v={}&ll={},{}&radius={}&limit={}'.format(
            CLIENT_ID,
            CLIENT_SECRET,
            VERSION,
            lat,
            lng,
            radius,
            LIMIT)

        # make the GET request
        results = requests.get(url).json()["response"]['groups'][0]['items']
        # return only relevant information for each nearby venue
        venues_list.append([(
            name,
            lat,
            lng,
            v['venue']['name'],
            v['venue']['location']['lat'],
            v['venue']['location']['lng'],
            v['venue']['categories'][0]['name']) for v in results])

    nearby_venues = pd.DataFrame([item for venue_list in venues_list for item in venue_list])
```

Bangalore Venue details

Out[27]:

	Area	Area Latitude	Area Longitude	Venue	Venue Latitude	Venue Longitude	Venue Category
0	560034	12.919954	77.625689	New Sagar Veg & Non-Veg Restaurant	12.919247	77.629621	Indian Restaurant
1	560034	12.919954	77.625689	Vishal Supermarket	12.915644	77.625360	Department Store
2	560034	12.919954	77.625689	New Sagar Veg & Non-Veg Restaurant	12.919247	77.629621	Indian Restaurant
3	560034	12.919954	77.625689	Vishal Supermarket	12.915644	77.625360	Department Store
4	560034	12.915011	77.670077	Kaikondanahalli lake	12.915391	77.673300	Lake

Chennai venue

Anna Nagar west	16	16	16	16	16	16
Ashok Nagar	11	11	11	11	11	11
Ayanavaram	5	5	5	5	5	5
Besant Nagar	58	58	58	58	58	58
Chennai	32	32	32	32	32	32
Chetpet	13	13	13	13	13	13
Egmore	8	8	8	8	8	8
Engineering College, Guindy	2	2	2	2	2	2
Fort St George	5	5	5	5	5	5
Gopalapuram	35	35	35	35	35	35
Guindy	6	6	6	6	6	6
High Court of Madras	9	9	9	9	9	9
Indian Institute of Technology Madras	4	4	4	4	4	4
Jawahar Nagar	5	5	5	5	5	5
Kalaignar Karunanidhi Nagar	6	6	6	6	6	6
Kilpauk	8	8	8	8	8	8
Kodambakkam	5	5	5	5	5	5
Kodungaiyur	1	1	1	1	1	1
Kotturpuram	7	7	7	7	7	7
Koyambedu	5	5	5	5	5	5
Mylapore	10	10	10	10	10	10

Bangalore unique cat details

In [29]:
```
#check how many categories were returned for Bangalore location
print('There are {} uniques categories in Bangalore.'.format(len(bangalore_south_venues['Venue Category'].unique())))
bangalore_south_venues.groupby('Venue').count()
```

```
There are 182 uniques categories in Bangalore.
```

Out[29]:

	Area	Area Latitude	Area Longitude	Venue Latitude	Venue Longitude	Venue Category
Venue						
19th main smokin adda	1	1	1	1	1	1
24/7 @ Lalit Ashok	1	1	1	1	1	1
24th Main	1	1	1	1	1	1
3654 Chinese Hut	2	2	2	2	2	2
4 Cue Snook	1	1	1	1	1	1
6 Ballygunge Place	1	1	1	1	1	1
6th AVENUE HOTEL	1	1	1	1	1	1
7-11 Kudsan Coffe	4	4	4	4	4	4
7th Cross DPS Bus Stand	1	1	1	1	1	1

Chennai unique cat details

In [30]:

```
#check how many categories were returned for Chennai location
print('There are {} uniques categories in Chennai.'.format(len(chennai_venues['Venue Category'].unique())))
chennai_venues.groupby('Area').count()
```

```
There are 110 uniques categories in Chennai.
```

Out[30]:

	Area Latitude	Area Longitude	Venue	Venue Latitude	Venue Longitude	Venue Category
Area						
Aminjikarai	5	5	5	5	5	5
Anna Nagar	34	34	34	34	34	34
Anna Nagar west	16	16	16	16	16	16
Ashok Nagar	11	11	11	11	11	11
Ayanavaram	5	5	5	5	5	5
Besant Nagar	58	58	58	58	58	58
Chennai	32	32	32	32	32	32
Chetpet	13	13	13	13	13	13
Egmore	8	8	8	8	8	8
Engineering College, Guindy	2	2	2	2	2	2
Fort St George	5	5	5	5	5	5

7. Analyze Each Neighborhood

Analyze Bangalore

[31]:

```
# one hot encoding
bangalore_onehot = pd.get_dummies(bangalore_south_venues[['Venue Category']], prefix="", prefix_sep="")

# add neighborhood column back to dataframe
bangalore_onehot['Venue'] = bangalore_south_venues['Venue']

# move neighborhood column to the first column
fixed_columns = [bangalore_onehot.columns[-1]] + list(bangalore_onehot.columns[:-1])
bangalore_onehot = bangalore_onehot[fixed_columns]
#examine the new dataframe size after one hot encoding
print('{} rows were returned after one hot encoding.'.format(bangalore_onehot.shape[0]))

#group rows by neighborhood and by taking the mean of the frequency of occurrence of each category
bangalore_grouped = bangalore_onehot.groupby('Venue').mean().reset_index()

#examine the new dataframe size after one hot encoding
print('{} rows were returned after grouping.'.format(bangalore_grouped.shape[0]))
bangalore_onehot.head()
```

```
1296 rows were returned after one hot encoding.
740 rows were returned after grouping.
```

Bangalore Cat details

Out[31]:

	Venue	ATM	Accessories Store	Airport	Airport Gate	Airport Lounge	Airport Service	Airport Terminal	American Restaurant	Andhra Restaurant	Arcade	Art Gallery	Art Studio	Arts & Crafts Store	Asian Restaurant	Athletics & Sports	Auto Garage	Auto Workshop	BBQ Joint	Badminton Court	Bakery	Bar	Bed & Breakfast
0	New Sagar Veg & Non-Veg Restaurant	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
1	Vishal Supermarket	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
2	New Sagar Veg & Non-Veg Restaurant	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
3	Vishal Supermarket	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
4	Kaikondanahalli lake	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0

Chennai cat details

(740, 9)

Out[43]:

	Venue	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	19th main smokin adda	Food & Drink Shop	Cluster Labels	Department Store	Electronics Store	Eastern European Restaurant	Duty-free Shop	Dumpling Restaurant	Dog Run
1	24/7 @ Lalit Ashok	Cluster Labels	Indian Restaurant	Deli / Bodega	Electronics Store	Eastern European Restaurant	Duty-free Shop	Dumpling Restaurant	Dog Run
2	24th Main	Cluster Labels	Indian Restaurant	Deli / Bodega	Electronics Store	Eastern European Restaurant	Duty-free Shop	Dumpling Restaurant	Dog Run
3	3654 Chinese Hut	Chinese Restaurant	Farm	Electronics Store	Eastern European Restaurant	Duty-free Shop	Dumpling Restaurant	Dog Run	Diner
4	4 Cue Snook	Pool Hall	Cluster Labels	Deli / Bodega	Electronics Store	Eastern European Restaurant	Duty-free Shop	Dumpling Restaurant	Dog Run

8. Cluster Neighborhoods

Cluster the Bangalore venues

In [46]:
```
# set number of clusters
kclusters = 5

bangalore_grouped_clustering = bangalore_grouped.drop('Venue', 1)

# run k-means clustering
kmeans = KMeans(n_clusters=kclusters, random_state=0).fit(bangalore_grouped_clustering)

# check cluster labels generated for each row in the dataframe
print(len(kmeans.labels_))
kmeans.labels_[0:8]
#create a new dataframe that includes the cluster as well as the top 10 venues for each neighborhood.
bangalore_merged = ban_neighborhoods_venues_sorted
#add clustering labels
bangalore_merged['Cluster Labels'] = kmeans.labels_

#merge Bangalore_grouped with Bangalore__data to add latitude/longitude for each neighborhood
bangalore_merged = bangalore_merged.join(bangalore_south_venues.set_index('Venue'), on='Venue')

bangalore_merged.head() # check the last columns!
```

```
740
```

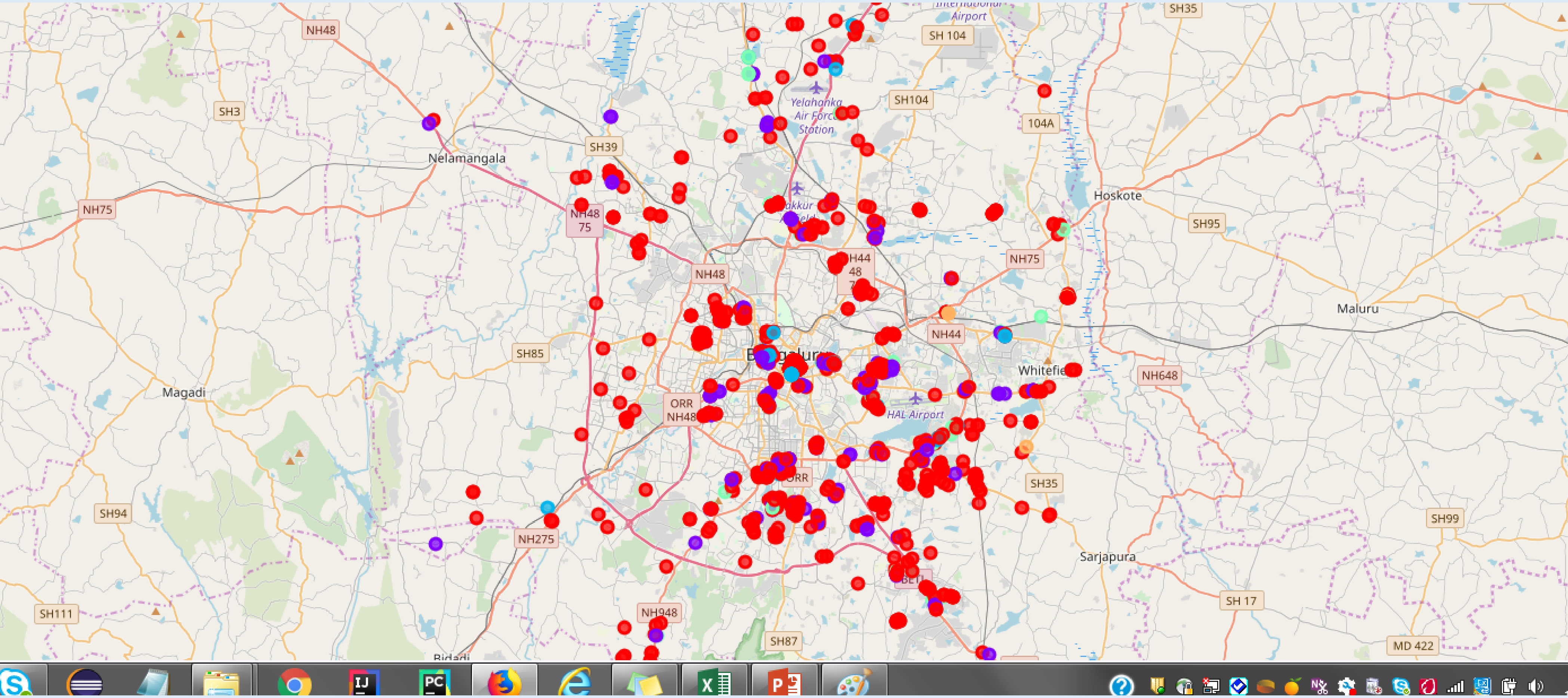
Bangalore cluster details

740

Out[46]:

	Venue	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	Cluster Labels	Area	Area Latitude	Area Longitude	Venue Latitude	Venue Longitude	Venue Category
0	19th main smokin adda	Food & Drink Shop	Cluster Labels	Department Store	Electronics Store	Eastern European Restaurant	Duty-free Shop	Dumpling Restaurant	Dog Run	0	562157	13.002233	77.550167	13.004082	77.549602	Food & Drink Shop
1	24/7 @ Lalit Ashok	Cluster Labels	Indian Restaurant	Deli / Bodega	Electronics Store	Eastern European Restaurant	Duty-free Shop	Dumpling Restaurant	Dog Run	1	560074	12.993533	77.578740	12.991838	77.581854	Indian Restaurant
2	24th Main	Cluster Labels	Indian Restaurant	Deli / Bodega	Electronics Store	Eastern European Restaurant	Duty-free Shop	Dumpling Restaurant	Dog Run	1	562149	12.911162	77.588582	12.908630	77.585748	Indian Restaurant
3	3654 Chinese Hut	Chinese Restaurant	Farm	Electronics Store	Eastern European Restaurant	Duty-free Shop	Dumpling Restaurant	Dog Run	Diner	0	562149	12.911162	77.588582	12.911861	77.586066	Chinese Restaurant
3	3654 Chinese Hut	Chinese Restaurant	Farm	Electronics Store	Eastern European Restaurant	Duty-free Shop	Dumpling Restaurant	Dog Run	Diner	0	562149	12.915347	77.588448	12.911861	77.586066	Chinese Restaurant

Bangalore cluster map details



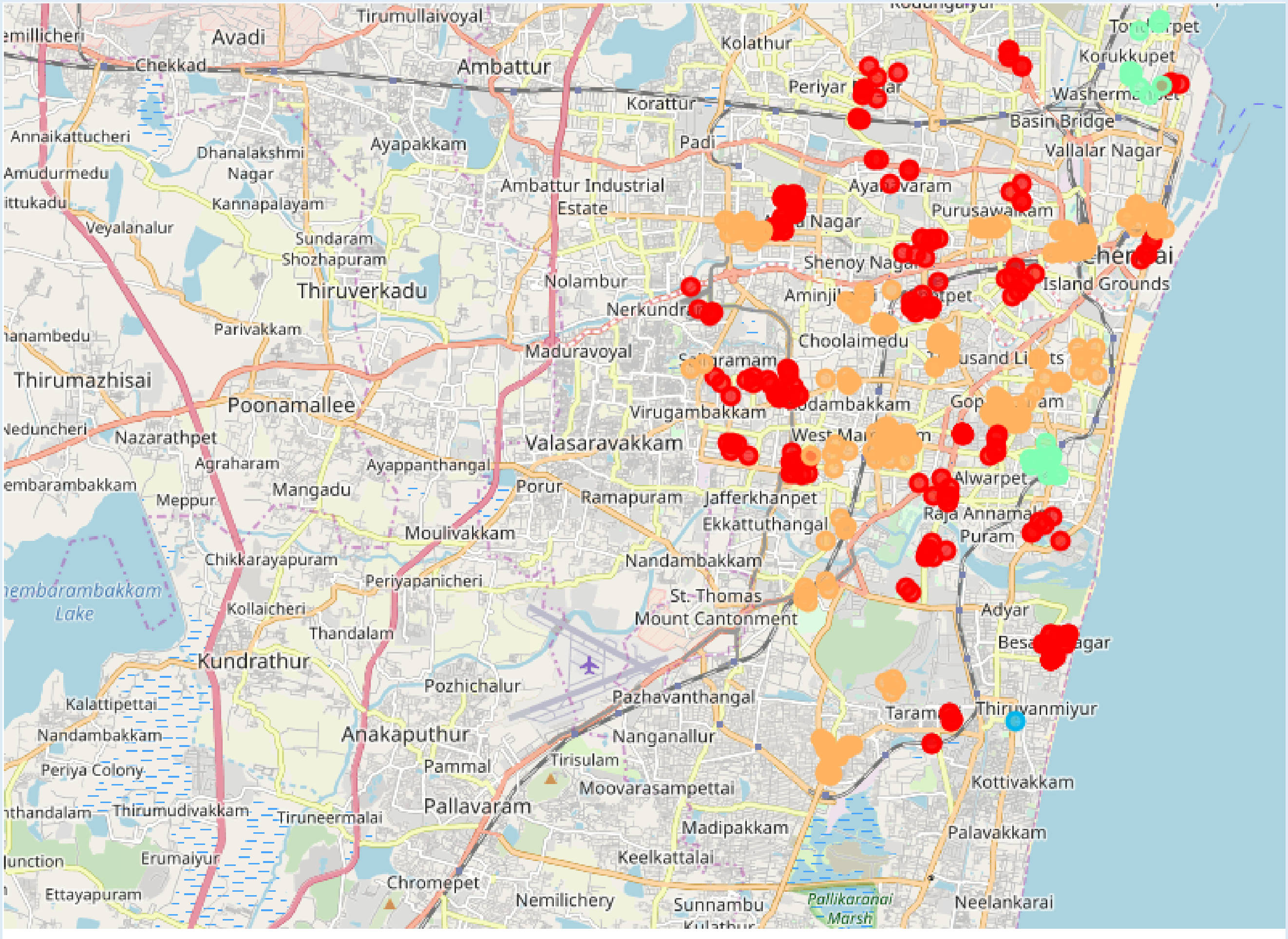
Chennai cluster details

47

Out[50]:

	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	Cluster Labels	Area Latitude	Area Longitude	Venue	Venue Latitude	Venue Longitude	Venue Category
0	Aminjikarai	Indian Restaurant	Bus Station	Restaurant	Fast Food Restaurant	Pizza Place	Health & Beauty Service	Harbor / Marina	Convenience Store	4	13.07056	80.227949	KFC	13.072318	80.227674	Fast Food Restaurant
0	Aminjikarai	Indian Restaurant	Bus Station	Restaurant	Fast Food Restaurant	Pizza Place	Health & Beauty Service	Harbor / Marina	Convenience Store	4	13.07056	80.227949	Domino's Pizza	13.070424	80.223845	Pizza Place
0	Aminjikarai	Indian Restaurant	Bus Station	Restaurant	Fast Food Restaurant	Pizza Place	Health & Beauty Service	Harbor / Marina	Convenience Store	4	13.07056	80.227949	Mehta Nagar Bus Stop	13.067977	80.226835	Bus Station
0	Aminjikarai	Indian Restaurant	Bus Station	Restaurant	Fast Food Restaurant	Pizza Place	Health & Beauty Service	Harbor / Marina	Convenience Store	4	13.07056	80.227949	Seamount Restaurant	13.067913	80.226601	Restaurant
0	Aminjikarai	Indian Restaurant	Bus Station	Restaurant	Fast Food Restaurant	Pizza Place	Health & Beauty Service	Harbor / Marina	Convenience Store	4	13.07056	80.227949	Apoorva Sangeetha	13.069872	80.224589	Indian Restaurant

Chennai Cluster map details



9. Result

Banaglore Result

Cluster 1

In [52]:
```
bangalore_merged.loc[bangalore_merged['Cluster Labels'] == 0, bangalore_merged.columns[[1] + list(range(5, bangalore_merged.shape[1]))]]
```

Out[52]:

	1st Most Common Venue	5th Most Common Venue	6th Most Common Venue	7th Most Common Venue	8th Most Common Venue	Cluster Labels	Area	Area Latitude	Area Longitude	Venue Latitude	Venue Longitude	Venue Category
0	Food & Drink Shop	Eastern European Restaurant	Duty-free Shop	Dumpling Restaurant	Dog Run	0	562157	13.002233	77.550167	13.004082	77.549602	Food & Drink Shop
3	Chinese Restaurant	Duty-free Shop	Dumpling Restaurant	Dog Run	Diner	0	562149	12.911162	77.588582	12.911861	77.586066	Chinese Restaurant
3	Chinese Restaurant	Duty-free Shop	Dumpling Restaurant	Dog Run	Diner	0	562149	12.915347	77.588448	12.911861	77.586066	Chinese Restaurant
4	Pool Hall	Eastern European Restaurant	Duty-free Shop	Dumpling Restaurant	Dog Run	0	560064	13.121561	77.621189	13.121841	77.622968	Pool Hall
6	Chinese Restaurant	Duty-free Shop	Dumpling Restaurant	Dog Run	Diner	0	562157	12.909117	77.576717	12.906354	77.578447	Chinese Restaurant
7	Coffee Shop	Eastern European Restaurant	Duty-free Shop	Dumpling Restaurant	Dog Run	0	560073	13.062109	77.483103	13.060387	77.484875	Coffee Shop
7	Coffee Shop	Eastern European Restaurant	Duty-free Shop	Dumpling Restaurant	Dog Run	0	562157	13.061867	77.484259	13.060387	77.484875	Coffee Shop
7	Coffee Shop	Eastern European	Duty-free Shop	Dumpling	Dog Run	0	562162	13.061867	77.484259	13.060387	77.484875	Coffee Shop

Chennai result details

Cluster 1

In [68]:

```
chennai_merged.loc[chennai_merged['Cluster Labels'] == 0, chennai_merged.columns[[1] + list(range(5, chennai_merged.shape[1]))]]
```

Out[68]:

	1st Most Common Venue	5th Most Common Venue	6th Most Common Venue	7th Most Common Venue	8th Most Common Venue	Cluster Labels	Area Latitude	Area Longitude	Venue	Venue Latitude	Venue Longitude	Venue Category
1	Department Store	Gym / Fitness Center	Malay Restaurant	Chinese Restaurant	Bakery	0	13.089136	80.209562	Shankar Chat Bhandar	13.086593	80.210253	Snack Place
1	Department Store	Gym / Fitness Center	Malay Restaurant	Chinese Restaurant	Bakery	0	13.089136	80.209562	Kumarakom	13.090402	80.212701	Indian Restaurant
1	Department Store	Gym / Fitness Center	Malay Restaurant	Chinese Restaurant	Bakery	0	13.089136	80.209562	Jack 'N' Jill	13.092006	80.210388	Fast Food Restaurant
1	Department Store	Gym / Fitness Center	Malay Restaurant	Chinese Restaurant	Bakery	0	13.089136	80.209562	amala mess	13.085722	80.208533	Fast Food Restaurant
1	Department Store	Gym / Fitness Center	Malay Restaurant	Chinese Restaurant	Bakery	0	13.089136	80.209562	Nuts 'n' Spices	13.091817	80.210298	Department Store
1	Department Store	Gym / Fitness Center	Malay Restaurant	Chinese Restaurant	Bakery	0	13.089136	80.209562	French Loaf	13.088752	80.212764	Bakery
1	Department Store	Gym / Fitness	Malay Restaurant	Chinese Restaurant	Bakery	0	13.089136	80.209562	Cafe Coffee Day	13.091986	80.210386	Coffee Shop

10.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. 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: Bangalore: Restaurant
- Cluster 2: Bangalore: Electronics Store
- Cluster 3: Bangalore: Hotel
- Cluster 4: Bangalore: Duty-free Shop
- Cluster 5: Bangalore: Clothing Store
- Cluster 1: Chennai: Department Store
- Cluster 2: Chennai: Pizza Place
- Cluster 3: Chennai : Middle Eastern Restaurant
- Cluster 4: Chennai: Vegetarian / Vegan Restaurant
- Cluster 5: Chennai : Indian Restaurant

What is lacking at this point is a systematic, quantitative way to identify and distinguish different venues 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 location.

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.

11.Conclusion

Using Foursquare API, we can capture the 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 venues located inside the city whether it is Restaurant,shops,Hotel, Clothstore or others In conclusion, both cities Bangalore and Chennai are the center of attraction among 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.

12. Acknowledgement

In this project,we have to acknowledge the data science course provided by IBM powered by Coursera

Thank you.
ThangarajDeivaikamani

Thank You