The Battle of Neighborhood: Bangalore VS Hyderabad

Siddharth Datthathreya

Email: siddharthdathathreya@gmail.com

1. Introduction

1.1 Background

Hyderabad and Bangalore, they are the metropolitan cities of my country called India. They are known for their technological advances and development in the IT Sector. The people who wants to come to this place is mainly for the jobs and luxurious lives. For that we need to know what are the frequent venues and where it will be. This places are also known for the rich hotels and every corner is having restaurants/ street foods. As I have lived my life in both of these states , I would like to recommend this is the best places to live, here is why:

1. You'll have great society
2. Meet other like-minded travelers and make new friends around the world!
3. So many opportunities that you can grab and many restaurants and many, many more.

1.2 Business Understanding/Problem Description:

As the people doesn't know where to live or start to work in a country, I am representing in this report how to start and which colony is having all the resources such as utilities, groceries, gym, IT services etc. And one more thing

- Which is the good place to start?
- Where we can we get all the resources as I mentioned above?
- In the given neighborhood, is there any good rating services like online-ordering food is available?

As people who migrate from other states, other countries, these two are the most recommended places stated by the TOI in our country for the luxurious life and cleanliness. The Bangalore is also known as Silicon City of India. Hyderabad is getting developed in much more stages nowadays. I will try to answer the following questions

- Where is the good place to start in Hyderabad or Bangalore?
- Where are the *'amenities'* located?
- Recent activities and Most common venues
- Suggests the best place to start either in Hyderabad or in Bangalore.

1.3 Target Audience

This project will serve two groups of audience:

4. **Travelers:** Help them make an informed decision while choosing a place by providing an in-depth analysis of place and their neighborhood.
5. **Non Residents:** Provide useful information and models which can help them where to open their first/next place to live.

2. Data Requirements

I analyzed in this project: **Hyderabad and Bangalore**. Following are the datasets used in the project:

6. [Hyderabad Pincode](#)
7. [Bangalore Pincode](#)

2.1 Hyderabad Pin code:

Using the web scrape method using beautifulsoup I have extracted the following columns:

- Neighborhood
- Pin code
- District

Below is a snapshot of the Hyderabad dataset:

	Neighborhood	District	Postal
0	A.Gs Office S.O	Hyderabad	500004
1	A.Gs. Staff Quarters S.O	Hyderabad	500045
2	Administrative Buildings S.O	Hyderabad	500007
3	Aliabad B.O	Hyderabad	500078
4	Allembylines S.O	Hyderabad	500087
5	Alwal S.O	Hyderabad	500010
6	Ambernagar B.O	Hyderabad	500044
7	Amberpet S.O	Hyderabad	500013
8	Anandbagh S.O	Hyderabad	500047
9	Anandnagar S.O (Hyderabad)	Hyderabad	500004
10	Anantaram B.O	Hyderabad	500015

After getting the latitudes and longitudes the values will be like:

Out[3]:

	Neighborhood	District	Postal	Latitude	Longitude
0	A.Gs Office S.O	Hyderabad	500004	17.406847	78.462384
1	A.Gs. Staff Quarters S.O	Hyderabad	500045	17.412195	78.530606
2	Administrative Buildings S.O	Hyderabad	500007	17.385044	78.486671
3	Aliabad B.O	Hyderabad	500078	17.340111	78.472958
4	Allembylines S.O	Hyderabad	500087	17.385044	78.486671
5	Alwal S.O	Hyderabad	500010	17.496702	78.506623
6	Ambernagar B.O	Hyderabad	500044	17.412641	78.509998
7	Amberpet S.O	Hyderabad	500013	17.392018	78.516256
8	Anandbagh S.O	Hyderabad	500047	17.456768	78.542660
9	Anandnagar S.O (Hyderabad)	Hyderabad	500004	17.411341	78.456367
10	Anantaram B.O	Hyderabad	500015	17.385044	78.486671

2.2 Bangalore Pin code:

Using the web scrape method using beautifulsoup I have extracted the following columns:

- Neighborhood
- Pin code
- District

Below is a snapshot of the Hyderabad dataset:

Out[1]:

	Neighborhood	District	Postal
0	A F Station Yelahanka S.O	Bangalore	560063
1	Adugodi S.O	Bangalore	560030
2	Agara B.O	Bangalore	560034
3	Agram S.O	Bangalore	560007
4	Amruthahalli B.O	Bangalore	560092
5	Anandnagar S.O (Bangalore)	Bangalore	560024
6	Anekal S.O	Bangalore	562106
7	Anekalbazar B.O	Bangalore	562106
8	Arabic College S.O	Bangalore	560045
9	Ashoknagar S.O (Bangalore)	Bangalore	560050

After getting the latitudes and longitudes using Google maps API the values will be like:

Out[2]:

	Neighborhood	District	Postal	Latitude	Longitude
0	A F Station Yelahanka S.O	Bangalore	560063	13.137243	77.610284
1	Adugodi S.O	Bangalore	560030	12.942004	77.608304
2	Agara B.O	Bangalore	560034	12.920707	77.641442
3	Agram S.O	Bangalore	560007	12.957993	77.630838
4	Amruthahalli B.O	Bangalore	560092	13.065879	77.604206
5	Anandnagar S.O (Bangalore)	Bangalore	560024	13.031328	77.591313
6	Anekal S.O	Bangalore	562106	12.708922	77.697065
7	Anekalbazar B.O	Bangalore	562106	12.971599	77.594563
8	Arabic College S.O	Bangalore	560045	13.029023	77.620763
9	Ashoknagar S.O (Bangalore)	Bangalore	560050	12.936969	77.560867
10	Attibele S.O	Bangalore	562107	12.752169	77.751625

3. Methodology:

3.1 Data Collection:
- The Hyderabad and Bangalore pin code were scraped from the respective sites. (https://finkode.com/ap/hyderabad.html) and (https://finkode.com/ka/bangalore.html)
- We used Foursquare API to get the venues around the hostel.
- We acquired the latitudes and longitudes data by using Geocoding API from google cloud.

3.2 Analytic Approach:
I took two approaches in the project.

Firstly, I used exploratory data analysis(EDA) to find the given dataset is reliable or not by using the data preprocessing methods and normalize the given json. Directly, by using the BeautifulSoup package, we are able to retrieve the data from the given site as mentioned above. After retrieving, I used the Geocoding API from Google maps with the relevant credentials and found the latitudes and longitudes. Later I plotted those in the map using Folium package.

Secondly, I used prescriptive analytics to help finding a perfect place to live whether it might Hyderabad or Bangalore. I will use clustering(K-Means). I will find the most common venue of a particular place where we / the people can actually live in that place with luxurious life with all the amenities present in the neighborhood.

4. Analysis

4.1 Exploratory Data Analysis

First, we have to gather the places as the dataset regarding the Hyderabad and Bangalore is not easily found. So, I have acquired the data from the sites mentioned above and filled it with latitudes and longitudes using the Geocoding API from Google Maps.

The below dataset is about Hyderabad district and it is acquired from the site mentioned above:

Out[3]:

	Neighborhood	District	Postal	Latitude	Longitude
0	A.Gs Office S.O	Hyderabad	500004	17.406847	78.462384
1	A.Gs. Staff Quarters S.O	Hyderabad	500045	17.412195	78.530606
2	Administrative Buildings S.O	Hyderabad	500007	17.385044	78.486671
3	Aliabad B.O	Hyderabad	500078	17.340111	78.472958
4	Allembylines S.O	Hyderabad	500087	17.385044	78.486671
5	Alwal S.O	Hyderabad	500010	17.496702	78.506623
6	Ambernagar B.O	Hyderabad	500044	17.412641	78.509998
7	Amberpet S.O	Hyderabad	500013	17.392018	78.516256
8	Anandbagh S.O	Hyderabad	500047	17.456768	78.542660
9	Anandnagar S.O (Hyderabad)	Hyderabad	500004	17.411341	78.456367
10	Anantaram B.O	Hyderabad	500015	17.385044	78.486671

The below dataset is about Bangalore district and it is acquired from the site mentioned above:

Out[3]:

	Neighborhood	District	Postal	Latitude	Longitude
0	A.Gs Office S.O	Hyderabad	500004	17.406847	78.462384
1	A.Gs. Staff Quarters S.O	Hyderabad	500045	17.412195	78.530606
2	Administrative Buildings S.O	Hyderabad	500007	17.385044	78.486671
3	Aliabad B.O	Hyderabad	500078	17.340111	78.472958
4	Allembylines S.O	Hyderabad	500087	17.385044	78.486671
5	Alwal S.O	Hyderabad	500010	17.496702	78.506623
6	Ambernagar B.O	Hyderabad	500044	17.412641	78.509998
7	Amberpet S.O	Hyderabad	500013	17.392018	78.516256
8	Anandbagh S.O	Hyderabad	500047	17.456768	78.542660
9	Anandnagar S.O (Hyderabad)	Hyderabad	500004	17.411341	78.456367
10	Anantaram B.O	Hyderabad	500015	17.385044	78.486671

To acquire those datasets, first we need to scrape the data from the table using BeautifulSoup (bs4 package) and converting it into html parser. Later we list it in the form of the lists and store it in a variable. After acquiring the lists, convert those lists into data frame using the pandas' package. Remember we are using python language for the analysis of data. Later we import the json library and google maps package for accessing the latitudes and the longitudes of the given dataset as we have seen above. I used the credentials key, got from the google cloud console. It will take five minutes to gather all the latitudes and longitudes. When we are gathering the data, it will be stored in the form of JSON. We need to convert that into list using the json package. And store it in the data frame. The given data will be formed as we are seeing above.

After the analysis, we are finding the number of collisions in the given data set. A **collision** or **clash** is a situation that occurs when two distinct pieces of data have the same hash value, checksum, fingerprint, or cryptographic digest. For the Hyderabad data set we got 14 and for the Bangalore data set we got 9. The total number of rows of Hyderabad and Bangalore data set are 202 and 269 respectively. As I am using the pin code of the separate districts, I will be getting these values.
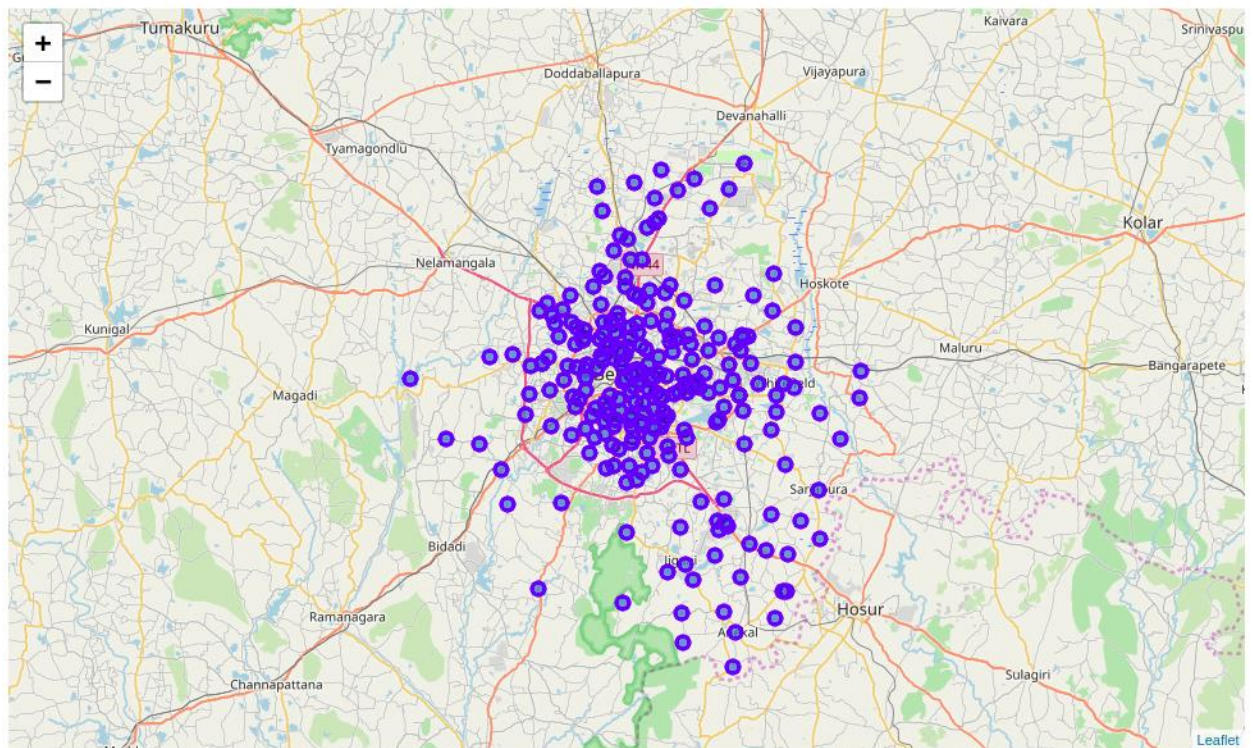
Later we will plot those formed data into the map using the Folium API.

The below plotted map is for Hyderabad district:
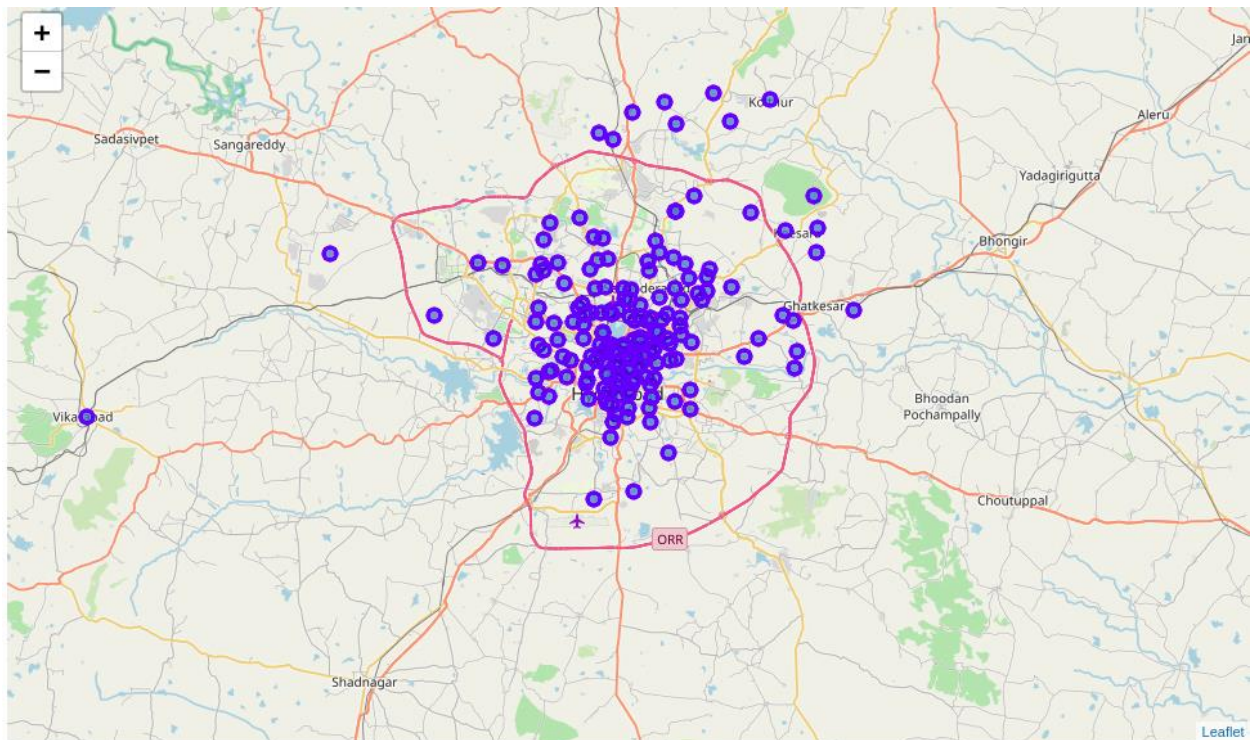
The below plotted map is for Bangalore district:

Collisions : 9

To access the venues from the exact pin code address we need to use the FourSquare API. FourSquare API allows us to access the venues and most common venues based on the people's ratings and contact numbers and a lot more. We can use this for acquiring the data regarding the neighborhood places for the amenities and such. To activate the FourSquare APi, we need to use the credential ID and credential key which will be provided from FourSquare developer's console. We have to login to access those details or sign up. We need the category table from the pin code so I have acquired the data set.

After acquiring the data set, proceed to store it in another data frame so that the original data set won't be lost. The explored dataset will be converted into one-hot encoding for the prediction analysis therefore we will convert it. The below are the datasets which are converted into one-hot encoding I.e. acquired from the explored data frame.

The below is for Bangalore and Hyderabad one-hot encoded data set:

0	A F Station Yelahanka S.O	0.0	0.0	0.0	0.0	0.333333	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0
1	Adugodi S.O	0.0	0.0	0.0	0.0	0.000000	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0
2	Agara B.O	0.0	0.0	0.0	0.0	0.000000	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0
3	Agram S.O	0.0	0.0	0.0	0.0	0.000000	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0
4	Amruthahalli B.O	0.0	0.0	0.0	0.0	0.000000	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0

0	A.Gs Office S.O	0.0	0.0	0.0	0.0	0.0	0.000	0.0	0.000	0.0	0.0	0.0	0.033333	0.033333	0.0
1	A.Gs. Staff Quarters S.O	0.0	0.0	0.0	0.0	0.0	0.000	0.0	0.000	0.0	0.0	0.0	0.000000	0.000000	0.0
2	AP Police Academy PO	0.0	0.0	0.0	0.0	0.0	0.125	0.0	0.000	0.0	0.0	0.0	0.000000	0.000000	0.0
3	Administrative Buildings S.O	0.0	0.0	0.0	0.0	0.0	0.000	0.0	0.000	0.0	0.0	0.0	0.066667	0.000000	0.0
4	Aliabad B.O	0.0	0.0	0.0	0.0	0.0	0.000	0.0	0.125	0.0	0.0	0.0	0.125000	0.000000	0.0

We have converted it into one-hot encoding because we have to find the prediction at which place the most common venue will be there for the clustering the labels. That is why, we have converted the categorical table into one-hot encoded form.

After that we will find the places which are having the most common venue by using the function created in the code and we will get the Bangalore and Hyderabad data set as seen below:

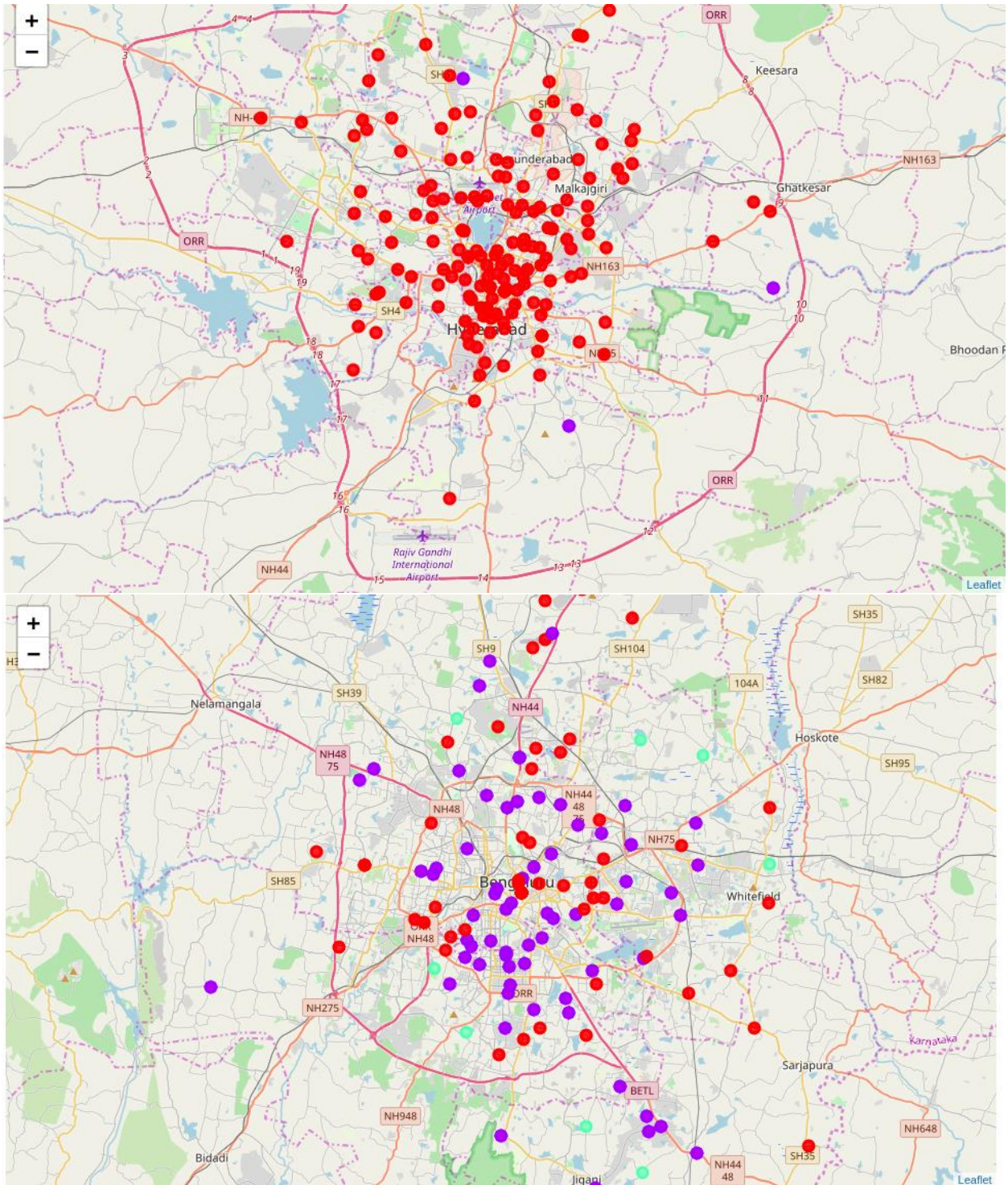|	Neighbourhood	1st Most Common Venue	2nd Most Common Venue	3rd Most Common Venue	4th Most Common Venue	5th Most Common Venue	6th Most Common Venue	7th Most Common Venue	8th Most Common Venue	9th Most Common Venue	10th Most Common Venue
0	A F Station Yelahanka S.O	Snack Place	Hot Spring	Airport Service	Dive Bar	Fast Food Restaurant	Farmers Market	Farm	Falafel Restaurant	Event Space	Electronics Store
1	Adugodi S.O	Indian Restaurant	Lounge	Café	Dessert Shop	Multiplex	Donut Shop	Coffee Shop	Clothing Store	Tea Room	Bookstore
2	Agara B.O	Indian Restaurant	Café	Italian Restaurant	Ice Cream Shop	Maharashtrian Restaurant	North Indian Restaurant	Snack Place	Bar	BBQ Joint	Pizza Place
3	Agram S.O	Indian Restaurant	Pizza Place	Business Service	Women's Store	Diner	Farm	Falafel Restaurant	Event Space	Electronics Store	Duty-free Shop
4	Amruthahalli B.O	Convenience Store	Shoe Store	Ice Cream Shop	Buffet	Café	Women's Store	Doner Restaurant	Farm	Falafel Restaurant	Event Space

	Neighbourhood	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	A.Gs Office S.O	Multiplex	Indian Restaurant	Coffee Shop	Hyderabadi Restaurant	Hotel	Hotel Bar	Ice Cream Shop	Performing Arts Venue	Science Museum	Café
1	A.Gs. Staff Quarters S.O	Indian Restaurant	Park	Dumpling Restaurant	Yoga Studio	Fast Food Restaurant	Farmers Market	Falafel Restaurant	Electronics Store	Donut Shop	Diner
2	AP Police Academy PO	Coffee Shop	River	Cricket Ground	Kebab Restaurant	Antique Shop	Hotel	IT Services	Café	Donut Shop	Fast Food Restaurant
3	Administrative Buildings S.O	Coffee Shop	Bus Station	Juice Bar	Indian Restaurant	Bakery	Chaat Place	Pizza Place	Indie Movie Theater	South Indian Restaurant	Multiplex
4	Aliabad B.O	Gym	Train Station	Smoke Shop	Asian Restaurant	Snack Place	Café	Bakery	Ice Cream Shop	Diner	Fast Food Restaurant

4.2 Clustering

After the above step, we have to find the K value for the the clustering analysis of the K-mean Clustering algorithm. Therefore, we will be using the silhouette method to find the optimized k-value. For the Bangalore data set we will get 3 as the optimized value, whereas for the Hyderabad data set we will get 2 as the optimized value for the K. We will find it by plotting it in the graph as shown below for the Bangalore and Hyderabad data set respectively.
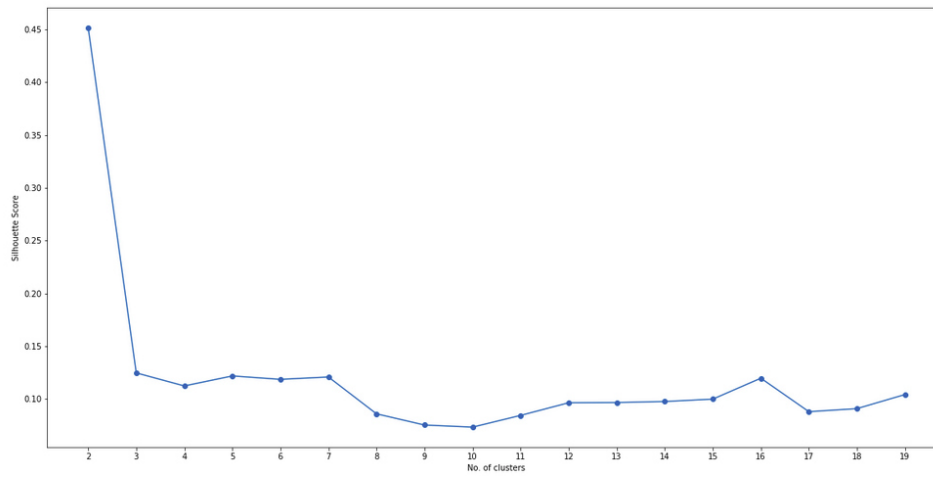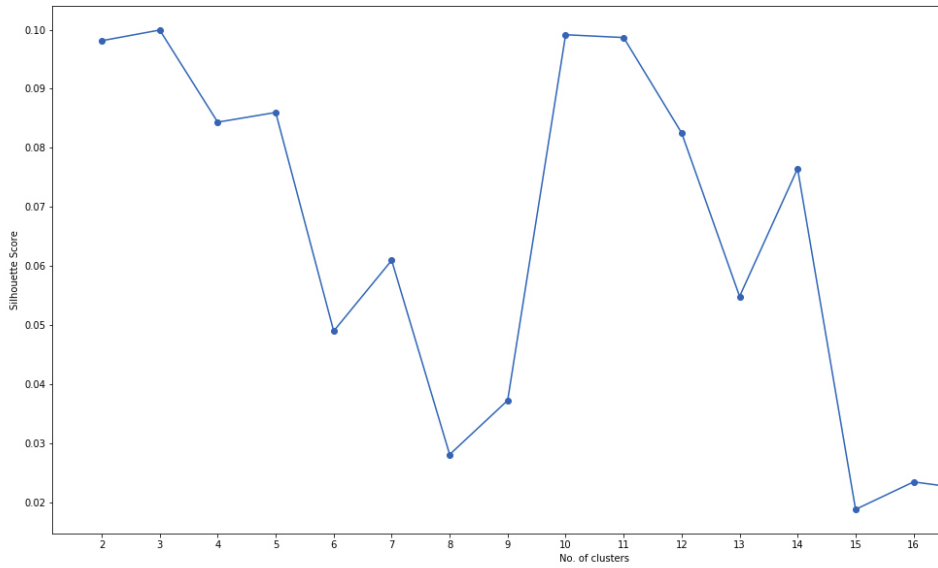
Now we will geo-visualize the data using the K-means Clustering algorithm for both the data sets Hyderabad and Bangalore respectively:

We will see that the cluster labels are maximum up to 2 for Bangalore whereas for Hyderabad it is just 1 because of the places are very highly near to each other and the recommendation venue is very high in those places.

Below is the final cluster data set for the Bangalore and Hyderabad:

	Neighborhood	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	10th Most Common Venue
38	Bidrahalli B.O	77.715877	2	ATM	Paper / Office Supplies Store	Halal Restaurant	Diner	Farm	Falafel Restaurant	Event Space	Electronics Store	Duty-free Shop	Dumpling Restaurant
54	Chikkabettahalli B.O	77.551398	2	ATM	Bus Station	Convenience Store	Café	Farmers Market	Farm	Falafel Restaurant	Event Space	Electronics Store	Duty-free Shop
69	Doddagubbi B.O	77.674249	2	Pizza Place	Pool	Women's Store	Diner	Farm	Falafel Restaurant	Event Space	Electronics Store	Duty-free Shop	Dumpling Restaurant
71	Doddakallasandra S.O	77.557580	2	ATM	Department Store	Juice Bar	Doner Restaurant	Fast Food Restaurant	Farmers Market	Farm	Falafel Restaurant	Event Space	Electronics Store
95	Hennagara B.O	77.675087	2	ATM	Field	Farmers Market	Farm	Falafel Restaurant	Event Space	Electronics Store	Duty-free Shop	Dumpling Restaurant	Dry Cleaner
102	Hosakerehalli B.O	77.536256	2	ATM	Pizza Place	Toll Booth	Department Store	Diner	Farm	Falafel Restaurant	Event Space	Electronics Store	Duty-free Shop

	Neighborhood	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	10th Most Common Venue
20	Badangpet B.O	78.529083	1	Pharmacy	Yoga Studio	Department Store	Fast Food Restaurant	Farmers Market	Falafel Restaurant	Electronics Store	Dumpling Restaurant	Donut Shop	Diner
81	Jeedimetla B.O	78.458642	1	Pharmacy	Pizza Place	Yoga Studio	Department Store	Farmers Market	Falafel Restaurant	Electronics Store	Dumpling Restaurant	Donut Shop	Diner
147	Pratapsingaram B.O	78.663985	1	Pharmacy	Yoga Studio	Department Store	Fast Food Restaurant	Farmers Market	Falafel Restaurant	Electronics Store	Dumpling Restaurant	Donut Shop	Diner

5. Result

We got a glimpse on where to start to live in these cities which I have taken as my project I.e Hyderabad and Bangalore. In Hyderabad, the best place to start is, having three places whereas, in Bangalore there are eight places to start with. All of the mentioned places are having all the amenities and IT sectors are booming in both of these districts. I conclude that no matter, where we start, start from a place where it is comfortable for you in terms of language, society etc. which is of course, out of presentation topic. The above-mentioned places are best places to start whether it is Bangalore or Hyderabad.

6. Discussion

According to the above analysis, the best place to start to live in Bangalore are Bidrahalli, Chikkabettahalli, Doddagubi, Doddakallasandra, Hennagara, Hosakerahalli, Hulimangala and Kadugodi. And the best place to start to live in Hyderabad are Badangpet, Jeedimetla and Pratapsingaram.In those above places all of the required essentials are available in the given neighborhood.