

# Applied Data Science



A Project Report on

## **“The Battle of Neighbourhoods: Exploring similar neighbourhood between Bangalore and Hyderabad”**

**Submitted By: Sankalp Joshi**

**Project Report submitted in partial fulfilment of the requirement for the  
award of the IBM Data Science Professional Certificate**

On Coursera



**(<https://www.coursera.org/>)**

# **TABLE OF CONTENTS:**

- 1. Introduction**
- 2. Data**
- 3. Methodology**
- 4. Analysis**
- 5. Results and Discussion**
- 6. Conclusion**

# Introduction

## **Business problem:**

Think of a scenario where you were living in Bangalore, India. But due to increase in the demand of the Data Scientist job in Hyderabad, you planned to move to Hyderabad. But since you don't any idea about the different areas of Hyderabad and the fact that you actually like the current area of Bangalore where you are residing due to all the amenities that you receive in that area, you want to look for a similar place with almost same amenities in Hyderabad also.

In this project, we are trying to solve such a problem. We are going to explore the similarity between neighbourhoods in Bangalore and Hyderabad, and will try to come up with some insights as a reference to which neighbourhoods in Hyderabad should one choose if he/she is moving from a particular area in Bangalore to Hyderabad.

To complete this task, we are going to use Foursquare location data to get the most common venue categories in each neighbourhood of both the city, Bangalore as well as Hyderabad, and then based on this we will group the neighbourhoods in clusters with k-means clustering Machine Learning Algorithm, and present the clustering result on a map generated with Folium library.

## About the Dataset

For this project, since the data set was not readily available, I used the postal code data for both the cities and searched for their respective Geolocation for the Latitude and Longitude data of each neighborhoods in both cities using the Geocoder package of Python. Then various data cleaning methods were applied in order to generate the final data set of both cities.

Then to generate the data about the venue places of each area in both cities within 1km, along with their respective Geolocation (Latitude & Longitude), I've used **Foursquare**.

Sample of each dataset used is shown below:


1) The neighborhood data of Bangalore is extracted from <https://www.mapsofindia.com/pincode/india/karnataka/bangalore/> ', after data cleaning process the sample Bangalore neighborhood data will look like this:

A screenshot of a web browser displaying a table with five rows of neighborhood data for Bangalore. The table has five columns: an index, Area, Pincode, State, and District. The data is as follows:

|   | Area                  | Pincode | State     | District  |
|---|-----------------------|---------|-----------|-----------|
| 1 | A F station yelahanka | 560063  | Karnataka | Bangalore |
| 2 | Adugodi               | 560030  | Karnataka | Bangalore |
| 3 | Agara                 | 560034  | Karnataka | Bangalore |
| 4 | Agram                 | 560007  | Karnataka | Bangalore |
| 5 | Air Force hospital    | 560007  | Karnataka | Bangalore |

Fig: Neighborhood of Bangalore Data

2) The neighbourhood data of Hyderabad is extracted from <https://www.mapsofindia.com/pincode/india/tehangana/hyderabad/> ', after data cleaning process the sample Hyderabad neighborhood data will look like this:

A screenshot of a web browser displaying a table with five rows of neighborhood data for Hyderabad. The table has five columns: an index, Area, Pincode, State, and District. The data is as follows:

|   | Area                     | Pincode | State     | District  |
|---|--------------------------|---------|-----------|-----------|
| 1 | A.G.college              | 500030  | Telangana | Hyderabad |
| 2 | A.Gs office              | 500004  | Telangana | Hyderabad |
| 3 | A.Gs. staff quarters     | 500045  | Telangana | Hyderabad |
| 4 | Administrative Buildings | 500007  | Telangana | Hyderabad |
| 5 | Afzalgunj                | 500012  | Telangana | Hyderabad |

Fig: Neighbourhood of Hyderabad Data

3) The Latitude and Longitude data for both the places which is extracted using the geocoders library of Python. The after extracting the Geolocation of each area of both places the final dataset will look like:

|   | Area               | Postal Code | City      | Latitude | Longitude |
|---|--------------------|-------------|-----------|----------|-----------|
| 0 | Adugodi            | 560030      | Bangalore | 12.9428  | 77.6104   |
| 1 | Agara              | 560034      | Bangalore | 13.0387  | 77.6468   |
| 2 | Air Force hospital | 560007      | Bangalore | 12.964   | 77.6275   |
| 3 | Amruthahalli       | 560092      | Bangalore | 13.0665  | 77.5966   |
| 4 | Anandnagar         | 560024      | Bangalore | 13.0334  | 77.5895   |

Fig: Geolocation Data of each Neighborhood in both cities

4) After we get the working data Bangalore neighbourhoods and Hyderabad neighborhoods, the Foursquare API is used to get the most common venue categories nearby each neighborhood in Manhattan and Seattle. And then group the neighborhood in clusters based on the most common venue categories with k-mean clusters Machine Learning Algorithm. For generating the nearby venues, we will choose a radius of 1km and limit to the top 100 venues.

|   | City      | Neighborhood | Neighborhood Latitude | Neighborhood Longitude | Venue              | Venue Latitude | Venue Longitude | Venue Category |
|---|-----------|--------------|-----------------------|------------------------|--------------------|----------------|-----------------|----------------|
| 0 | Bangalore | Adugodi      | 12.942847             | 77.610416              | PVR IMAX           | 12.934595      | 77.611321       | Movie Theater  |
| 1 | Bangalore | Adugodi      | 12.942847             | 77.610416              | Lot Like Crêpes    | 12.936421      | 77.613284       | Creperie       |
| 2 | Bangalore | Adugodi      | 12.942847             | 77.610416              | Koramangala Social | 12.935518      | 77.614097       | Lounge         |
| 3 | Bangalore | Adugodi      | 12.942847             | 77.610416              | Tommy Hilfiger     | 12.934552      | 77.611347       | Clothing Store |
| 4 | Bangalore | Adugodi      | 12.942847             | 77.610416              | PVR Cinemas        | 12.934389      | 77.611184       | Multiplex      |

Fig: Foursquare Location Data

## **Methodology**

In this project, we are trying to explore the similarity between neighborhoods in Bangalore and Hyderabad. So, the stakeholders of this project would be talents living in Bangalore who are thinking to shift to Hyderabad in search of new jobs.

To complete this task, we will do the following steps in this project:

- Data Extraction: Including Bangalore neighborhood data, Hyderabad Neighborhood data, venues nearby each neighborhood which will be generated by Foursquare API.
- Data Cleaning Processes: Including removal of Unwanted rows, Changing Column names, and Resetting Index.
- Exploratory data analysis to better understand our working dataset.
- Creating Map of Bangalore and map of Hyderabad with Folium library.
- Clustering of the neighborhoods: We will use K means clustering machine learning algorithm since it is fast and especially when there are too many variables.
- Presenting the clustering result on the map.

## **Analysis**

### **Exploratory Data Analysis:**

- The venues dataset has venues for 285 unique neighbourhoods, out of which 171 are from Bangalore and rest 114 are from Hyderabad.
- Foursquare API returned 100 venues for very less no. of the neighborhoods, but returned less than 100 venues for most of the neighborhoods.
- There are 274 unique categories generated from the dataset. Bangalore has 246 unique venue categories and Hyderabad has 164 unique venue categories.
- Visualize the top 25 categories, most of the top 25 categories are related to Restaurant, Cafe and other food related shop.

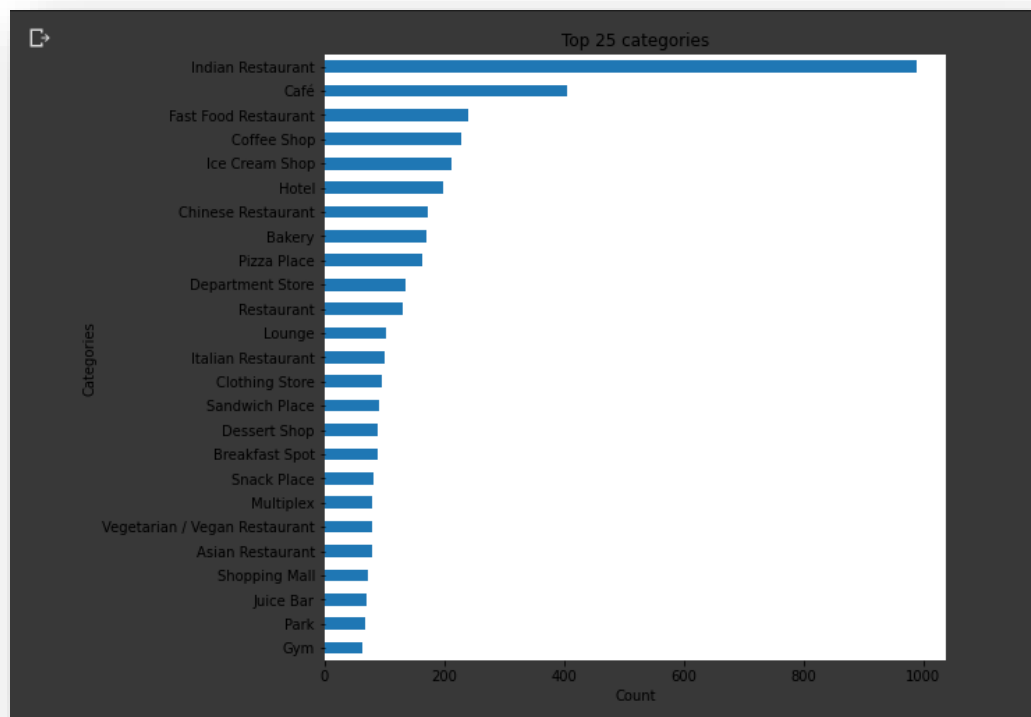


Fig: Top 25 Categories

- All those categories whose count is 1

```
array(['Toll Booth', 'University', 'Mountain', 'Basketball Court',
      'Other Nightlife', 'Museum', 'Noodle House', 'Design Studio',
      'Food Service', 'Social Club', 'Bike Rental / Bike Share',
      'Butcher', 'Currency Exchange', 'Pakistani Restaurant',
      'Irish Pub', 'Chaat Place', 'Art Museum', 'Outdoors & Recreation',
      'Bagel Shop', 'Bridal Shop', 'Wine Shop', 'Laser Tag',
      'Lighthouse', 'Arts & Entertainment', 'Pool Hall',
      'English Restaurant', 'Bike Shop', 'Speakeasy', 'Henan Restaurant',
      'Dive Bar', 'Airport Food Court', 'Maharashtrian Restaurant',
      'Construction & Landscaping', 'Bus Line', 'Memorial Site',
      'Airport Terminal', 'Harbor / Marina', 'Recreation Center',
      'Southern / Soul Food Restaurant', 'Animal Shelter', 'Castle',
      'Discount Store', 'Chettinad Restaurant', 'Duty-free Shop',
      'New American Restaurant', 'Coworking Space', 'Turkish Restaurant',
      'Nightlife Spot', 'Zoo', 'Event Service', 'Dumpling Restaurant',
      'Caribbean Restaurant', 'Parsi Restaurant',
      'Modern European Restaurant', 'Photography Studio',
      'Kebab Restaurant', 'Auto Workshop', 'Financial or Legal Service',
      'Farm', 'Frozen Yogurt Shop', 'Gift Shop', 'Theme Park'],
      dtype=object)
```

- Top 10 categories for each Neighborhood

In this section, we will first use the one hot encoding to each category into a separate column and calculate the frequency of occurrence of each category. Then we will slice the top 10 categories for clustering. Below are the top 10 categories of the first 5 neighborhoods.

| Neighborhood | 1st Most Common Venue | 2nd Most Common Venue | 3rd Most Common Venue | 4th Most Common Venue   | ...               | 6th Most Common Venue | 7th Most Common Venue | 8th Most Common Venue | 9th Most Common Venue | 10th Most Common Venue      |
|--------------|-----------------------|-----------------------|-----------------------|-------------------------|-------------------|-----------------------|-----------------------|-----------------------|-----------------------|-----------------------------|
| 0            | Adugodi               | Indian Restaurant     | Lounge                | Café                    | Dessert Shop      | ...                   | Multiplex             | Juice Bar             | Donut Shop            | Brewery                     |
| 1            | Afzalgunj             | History Museum        | Food Court            | Food                    | Food Truck        | ...                   | Diner                 | Women's Store         | Event Service         | English Restaurant          |
| 2            | Agara                 | Pharmacy              | Restaurant            | Bakery                  | Music Venue       | ...                   | Garden                | Coffee Shop           | Electronics Store     | Eastern European Restaurant |
| 3            | Air Force hospital    | Café                  | Juice Bar             | South Indian Restaurant | Pizza Place       | ...                   | Bar                   | Gift Shop             | Dry Cleaner           | Discount Store              |
| 4            | Alabad                | Chinese Restaurant    | Auto Garage           | Gym                     | Food & Drink Shop | ...                   | Snack Place           | Asian Restaurant      | Café                  | Smoke Shop                  |

5 rows x 11 columns

Fig: Top 10 categories for each Neighborhood

- Use K means to Cluster Neighborhoods into 4 clusters on map

In this section, we will first create maps for Bangalore and Hyderabad, do the k means clustering and then superimpose the clustering result on the maps of Bangalore and Hyderabad to for visualization.

Clustered Map of Bangalore:

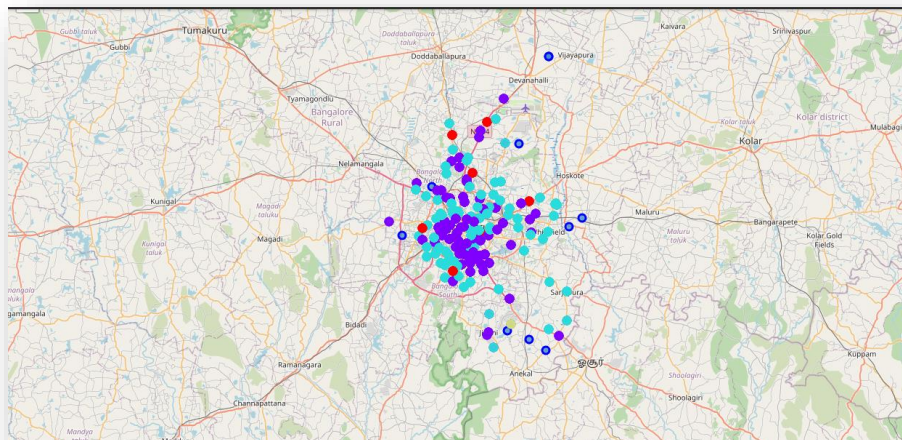


Fig: Clustered Map of Bangalore



### Clustered Map of Hyderabad:

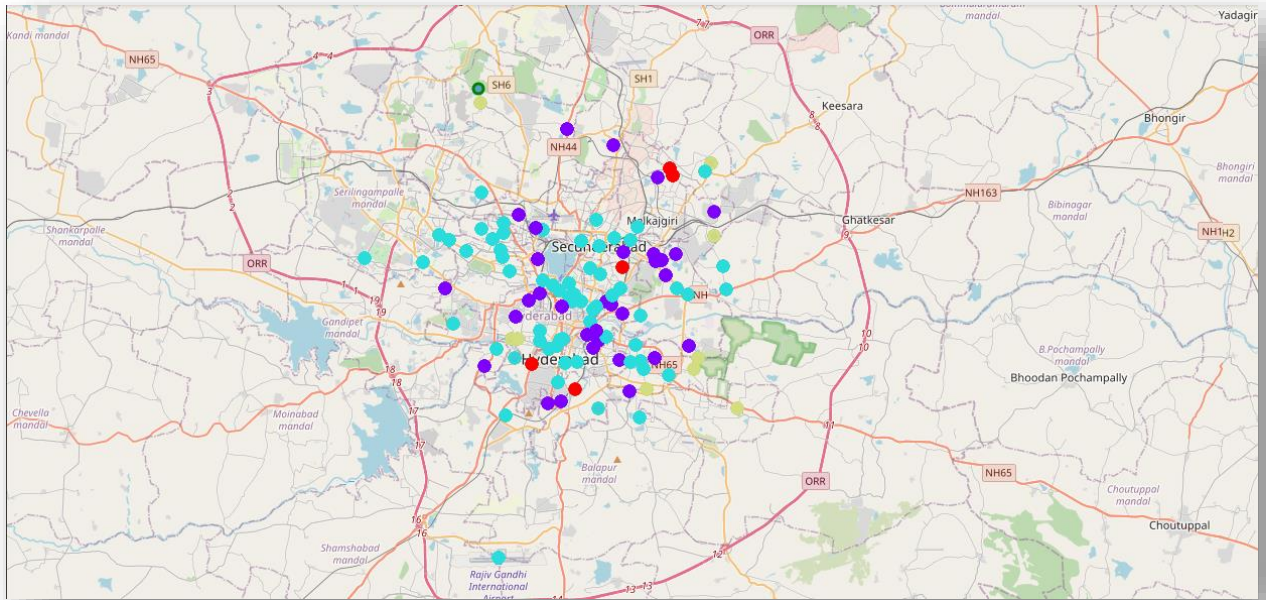


Fig: Clustered Map of Hyderabad

From the map above, we can see that the neighborhoods in Bangalore are mostly in two clusters, which is cluster 1 and cluster 2.

## **Results and Discussion**

As we can clearly see from the above clustered map that for both the cities the majority of neighbourhoods lies in either cluster 1 or cluster 2. So, for this project we will only going to conclude our results based on those 2 clusters.

After Googling about different posh areas of Bangalore it was found that Jayanagar, Indiranagar and M G road are among the top posh area in Bangalore according to various website. After checking out which clusters are the three neighborhoods belongs to it was found that

Jayanagar = Cluster 1

Indiranagar = Cluster 2

MG Road = Cluster 2

So, we further found out the neighbourhood in Hyderabad with same cluster number and the result were:

If someone is moving to Hyderabad from Jayanagar, the similar neighborhoods in Hyderabad that he/she could possibly choose are:

- |                             |                      |
|-----------------------------|----------------------|
| 1. Amberpet                 | 14. New Nallakunta   |
| 2. Bazarghat                | 15. Ngri             |
| 3. Chanchalguda             | 16. Old Malakpet     |
| 4. Falaknuma                | 17. Sahifa           |
| 5. Huda Residential complex | 18. Saidabad         |
| 6. I.E.nacharam             | 19. Sainagar         |
| 7. I.M.colony               | 20. Sakkubai Nagar   |
| 8. Jama I osmania           | 21. Sanathnagar I e  |
| 9. Jeedimetla               | 22. Saroornagar      |
| 10. Kakatiya Nagar          | 23. Sitaphalmandi    |
| 11. Murad Nagar             | 24. Snehapuri Colony |
| 12. Nallakunta              | 25. Somajiguda.      |
| 13. New Maruthi nagar       |                      |

And if someone is moving to Hyderabad from MG road or Indiranagar, the similar neighborhoods in Hyderabad that he/she could possibly choose are:

- |                         |                       |
|-------------------------|-----------------------|
| 1. Ashoknagar           | 18. Hyderabad Jubilee |
| 2. Afzalgunj            | 19. Jubilee Hills     |
| 3. Attapur              | 20. Kanchanbagh       |
| 4. Badangpet            | 21. Khairatabad       |
| 5. Balapur              | 22. L B nagar         |
| 6. Banjara Hills        | 23. Lallaguda         |
| 7. Barkatpura           | 24. Mg Road           |
| 8. Begumpet             | 25. Madhapur          |
| 9. Bholakpur            | 26. Malakpet Colony   |
| 10. Boduppal            | 27. Mangalhat         |
| 11. Dattatreya Colony   | 28. Moghalpura        |
| 12. Dhoolpet            | 29. Nanakramguda      |
| 13. Dilsukhnagar Colony | 30. Narayanguda       |
| 14. Erragadda           | 31. New MLA quarters  |
| 15. Gagan Mahal         | 32. P & t colony      |
| 16. Gandhinagar         | 33. Rahmath Nagar     |
| 17. Golconda            | 34. Rail Nilayam      |

35. Rajendranagar  
36. Saidabad Colony  
37. Secunderabad  
38. Shyam Nagar  
39. South Banjara hills  
40. Srinagar Colony

41. Srinivasapuram  
42. Uppal  
43. Yakutpura  
44. Yousufguda  
45. Zamistanpur

Since there are too many neighborhoods in Bangalore, we did not look deep into each one, instead we focused on the top three neighborhoods in Bangalore, which are:

- Jayanagar
- Indiranagar, and
- MG Road

And we did find the similar neighborhood in Hyderabad for each of those 3 neighborhoods in Bangalore.

## **Conclusion**

For the final conclusion, after Googling about different areas of both the cities and from the result obtained from above algorithms. We can suggest that:

1) If someone is moving from Jayanagar Area of Bangalore, Then the top 3 similar neighborhoods in Hyderabad that one could choose are:

- Amberpet
- Falaknuma, and
- Chanchalguda

2) If someone is moving from Indiranagar or M G Road area in Bangalore, Then the top 3 similar neighborhoods in Hyderabad that one could choose are:

- Banjara Hills
- Jubilee Hills, and
- Uppal

In the end, in this project, we are trying to find out the similar neighborhoods between Bangalore and Hyderabad, intend to provide an insight to those who would like to move from Bangalore to Hyderabad, to start a new career.

And luckily, there do exist some neighborhoods in Hyderabad similar to those in Bangalore as shown above.

But it should also be well acknowledged that the result of this analysis is basically according to the categories of venues nearby the centre of the neighborhood, there are some other factor you may also want to take into consideration before you make up your mind, things like weather, commuting, crime rate, and so on.

Hope you had a good time through this journey of exploring similar neighborhoods.