

## IBM - Applied Data Science Capstone Project Presentation

# The Battle of Neighborhoods: Exploring similar neighborhoods between Bangalore and Hyderabad

Submitted By: Sankalp Joshi

# 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 have 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.
- ❖ The neighborhood data of Bangalore is extracted from '<https://www.mapsofindia.com/pincode/india/karnataka/bangalore/>'
- ❖ The neighbourhood data of Hyderabad is extracted from '<https://www.mapsofindia.com/pincode/india/telangana/hyderabad/>'

- ❖ 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

# Methodology

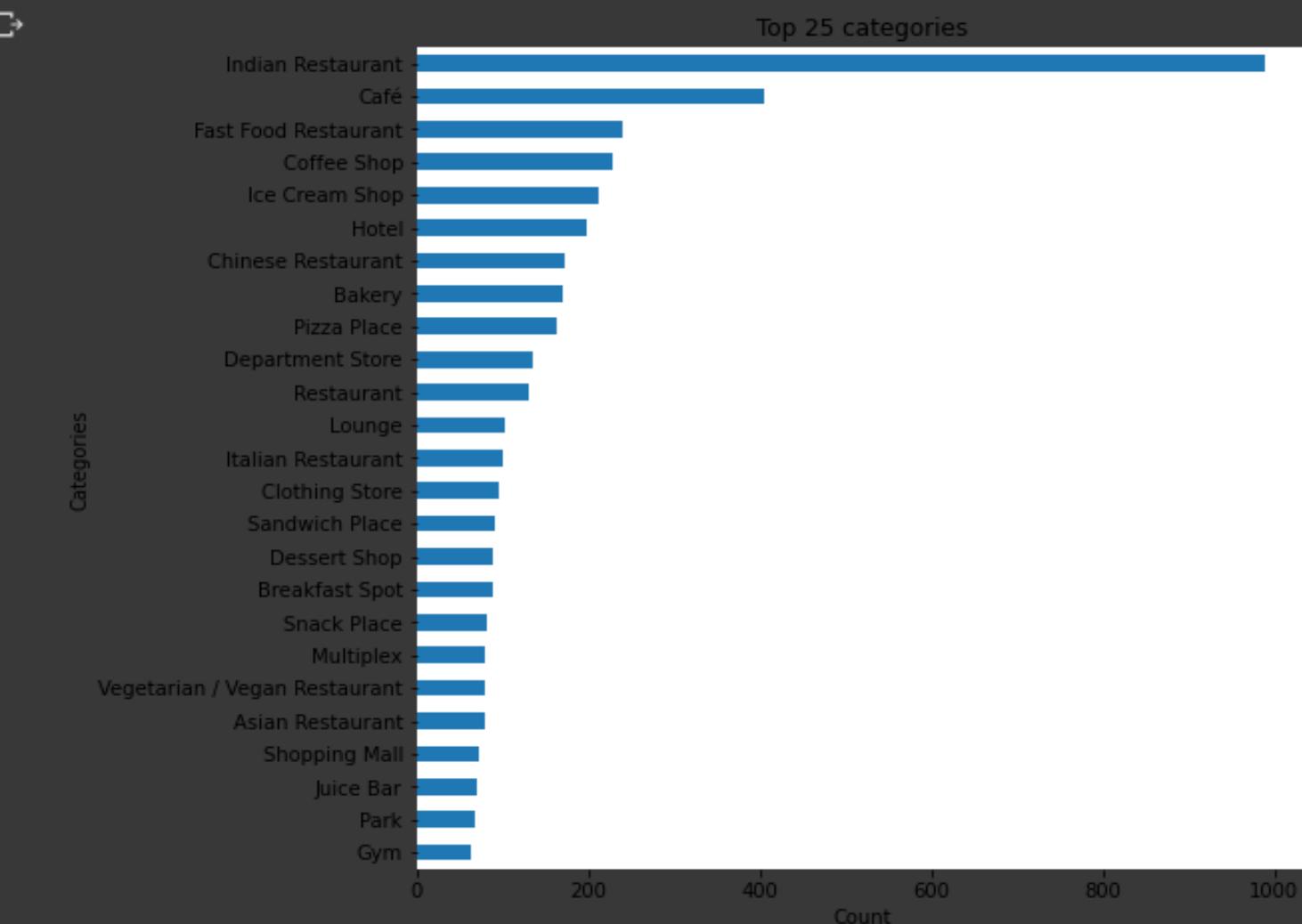
- ❖ Data Extraction: Including Bangalore neighborhood data, Hyderabad Neighborhood data, venues nearby each neighborhood which will be generated by Foursquare API.
- ❖ Data Cleaning Processes: Including removal of Unwanted rows, Changing Column names, and Resetting Index.
- ❖ Exploratory data analysis 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

## ❖ EDA

- ❖ 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.

## ❖ Top 25 Categories



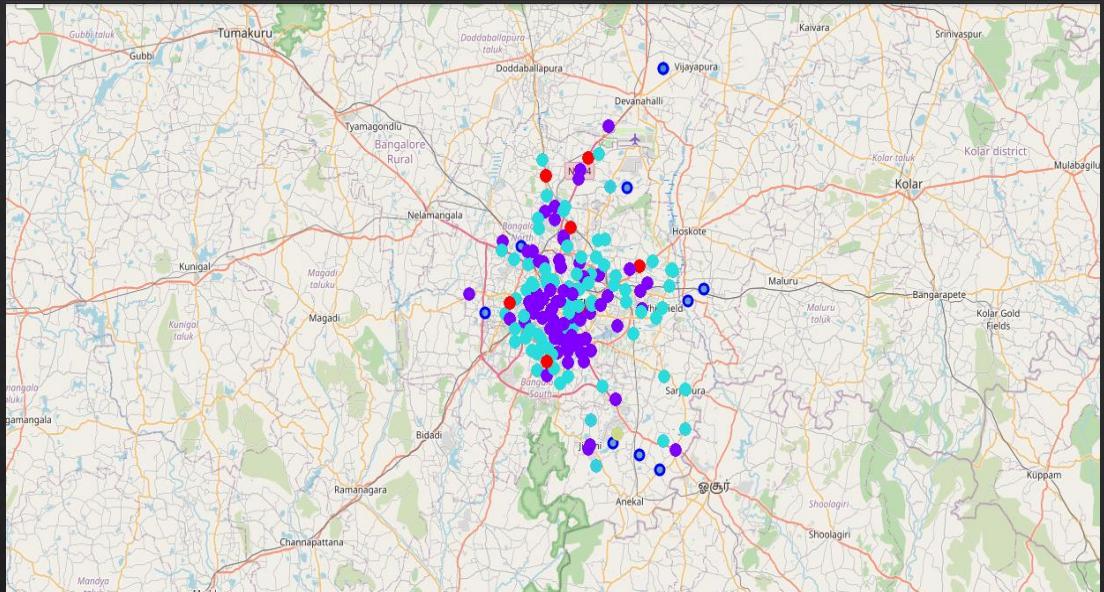
❖ Top 10 categories for each Neighborhood:

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.

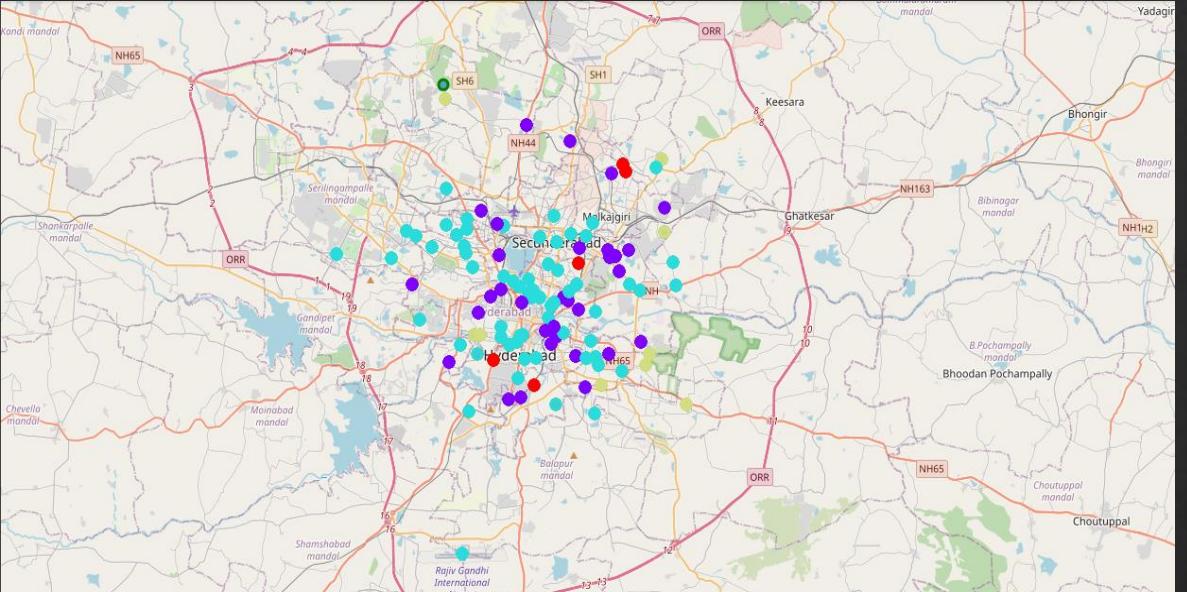
	Neighborhood	1st Most Common Venue	2nd Most Common Venue	3rd Most Common Venue	4th Most Common Venue	...	6th Most Common Venue	7th Most Common Venue	8th Most Common Venue	9th Most Common Venue	10th Most Common Venue
0	Adugodi	Indian Restaurant	Lounge	Café	Dessert Shop	...	Multiplex	Juice Bar	Donut Shop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

5 rows x 11 columns

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



Clustered Map of Bangalore



Clustered Map of Hyderabad

# 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

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

- ❖ 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
- ❖ 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

# Conclusion

- ❖ 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.

THANK YOU!