

# **Capstone Project - Battle of the Neighborhoods in the city of Hyderabad, Telangana, India**

Sandeep Kumar Vengala

## **Table of contents:**

- Introduction: Business Problem
- Data
- Methodology
- Analysis
- Results and Discussion
- Conclusion

## **Introduction: Business Problem Statement**

In this project we will try to find an optimal location for opening a restaurant. Specifically, this report will be targeted to stakeholders interested in opening any food-joint/restaurant in Hyderabad, Telangana India.

Since there are lots of restaurants in Hyderabad, we will try to detect locations that are not already crowded with restaurants. We are also particularly interested in neighborhoods which are similar to the best neighborhoods in terms of amenities. We would also prefer locations as close to city centre as possible, assuming that first two conditions are met.

We will use our data science technical expertise to generate a few most promising neighborhoods based on these criteria. Advantages of each area will then be clearly expressed so that best possible final location can be chosen by stakeholders.

## **Data**

Based on definition of our problem, factors that will influence our decision are:

- number of existing restaurants in the neighborhood (any type of restaurant)
- variety of amenities in the neighborhood, if any
- distance of neighborhood from city centre

Neighborhoods have been defined based on names of post offices as given in the form of a table in the website: <http://pincode.india-server.com/cities/hyderabad/>

Following data sources will be needed to extract/generate the required information:

- names of all neighbourhoods (post office names) will be extracted from the above-mentioned website using Pandas package in python
- approximate addresses of centres of these neighbourhoods will be obtained using GeoPy Geocoder package in python
- number of restaurants and their type and location in every neighbourhood will be obtained using Foursquare API

## Data Cleaning:

The table present in the website was procured using the Pandas 'read\_html()' method:

List of all Hyderabad post offices with pincode	Unnamed: 1_level_0	Unnamed: 2_level_0	Unnamed: 3_level_0
S.No.	Post office	Office type	Pincode
0	1.0 A.Gs Office	S.O	500004
1	2.0 A.Gs. Staff Quarters	S.O	500045
2	3.0 Administrative Buildings	S.O	500007
3	4.0 Airforce Academy	S.O	500043
4	5.0 Aliabad	B.O	500101
5	6.0 Allampalli .	S.O	501102
6	7.0 Alwal	S.O	500010
7	8.0 Ambernagar	B.O	500044
8	9.0 Amberpet	S.O	500013
9	10.0 Anandbagh	S.O	500047
10	11.0 Anandnagar	S.O	500004
11	12.0 Anantaram	B.O	500101
12	13.0 Andhra Mahila Sabha	S.O	500044
13	14.0 Ankireddipalli	B.O	501301
14	15.0 Ankushapur	B.O	501301

**DataFrame gathered from website**

The DataFrame was then cleaned to only include names of the post offices which are the names of the neighborhoods:

Neighborhood_Names	
0	A.Gs Office
1	A.Gs. Staff Quarters
2	Administrative Buildings
3	Airforce Academy
4	Aliabad
5	Allampalli .
6	Alwal
7	Ambernagar
8	Amberpet
9	Anandbagh
10	Anandnagar
11	Anantaram
12	Andhra Mahila Sabha
13	Ankireddipalli
14	Ankushapur
15	Aoc Records

**DataFrame after dropping unwanted columns**

The names of neighbors were then extracted from the above DataFrame and a new dataframe was created consisting of the neighborhood names and their latitude and longitude.

	name	latitude	longitude
0	Aliabad	17.345630	78.472680
1	Alwal	17.502229	78.508858
2	Ankushapur	18.833897	79.668823
3	Ashoknagar	17.952459	79.949425
4	Badangpet	17.322375	78.528034
5	Bahadurpura	17.357067	78.454542
6	Balanagar Township	17.465711	78.449525
7	Banjara Hills	17.417746	78.439901
8	Barkatpura	17.392612	78.496968
9	Bazarghat	17.402787	78.460449
10	Begumpet	17.444020	78.462482
11	Bhaskar Nagar	17.501314	78.512777
12	BITS	17.547231	78.572562
13	Boduppall	17.420075	78.582784
14	Bogaram	17.304280	79.074132
15	Bolarum	17.533399	78.515577

**DataFrame after adding latitude and longitude**

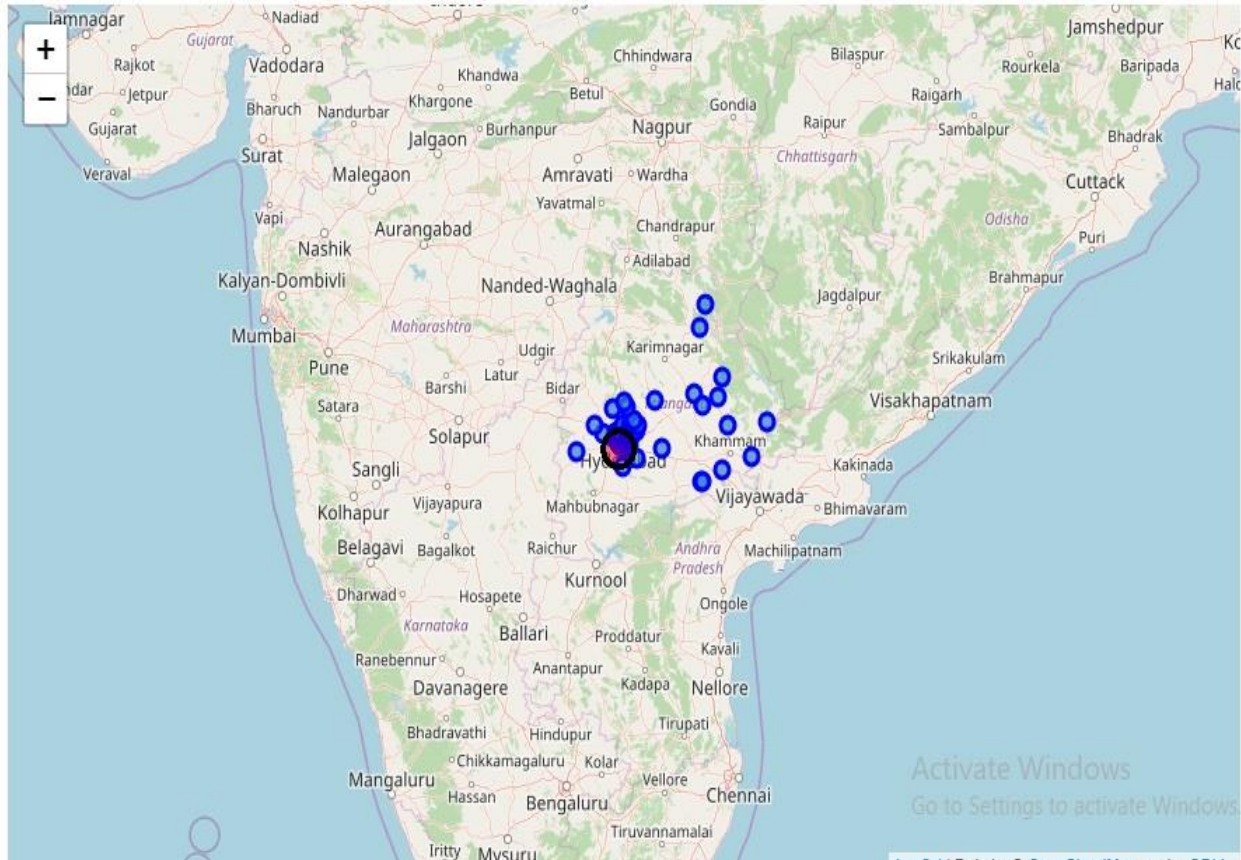
## Methodology

In the cells below the following format of the project will be followed in order to explore the neighborhoods in Hyderabad and try to analyze and suggest an optimal location for opening a food-joint/restaurant.

1. Get list of all neighborhoods from the website listing names of postal offices all over Hyderabad.
2. Obtain latitudes and longitudes for all these neighborhoods.
3. Check for any outliers, i.e. postal office names which aren't within 25Kms from the centre of Hyderabad and remove these data points.
4. Send GET requests to Foursquare API to get list of venues (maximum 100) within the vicinity of all neighbors (vicinity is defined within a radius of 1000m from the neighborhood).
5. Remove all neighborhoods which do not have more than 10 venues in their vicinities since there isn't enough data for accurate clustering using K-Means Clustering.
6. One-hot encode the data and feed the dataframe to K-Means clustering Algorithm to form 10 clusters from the list of available neighborhoods.
7. Create visualizations using the cluster number of all neighborhoods to better understand the data create valuable insights. Here, we try to find out the best neighborhoods amongst the available list of neighborhoods and try to identify its cluster.
8. Again, send GET requests to Foursquare API to get list of venues (maximum 100) of 'section = food' to obtain only venues related to food joints within the vicinity of all selected neighborhoods (vicinity is defined within a radius of 1000m from the neighborhood).
9. Create visualizations to find out which neighborhoods have maximum food joints and which ones have maximum types of food joints.
10. Suggest optimal locations for opening a restaurant based on all the statistical analysis done.

## Data Visualization

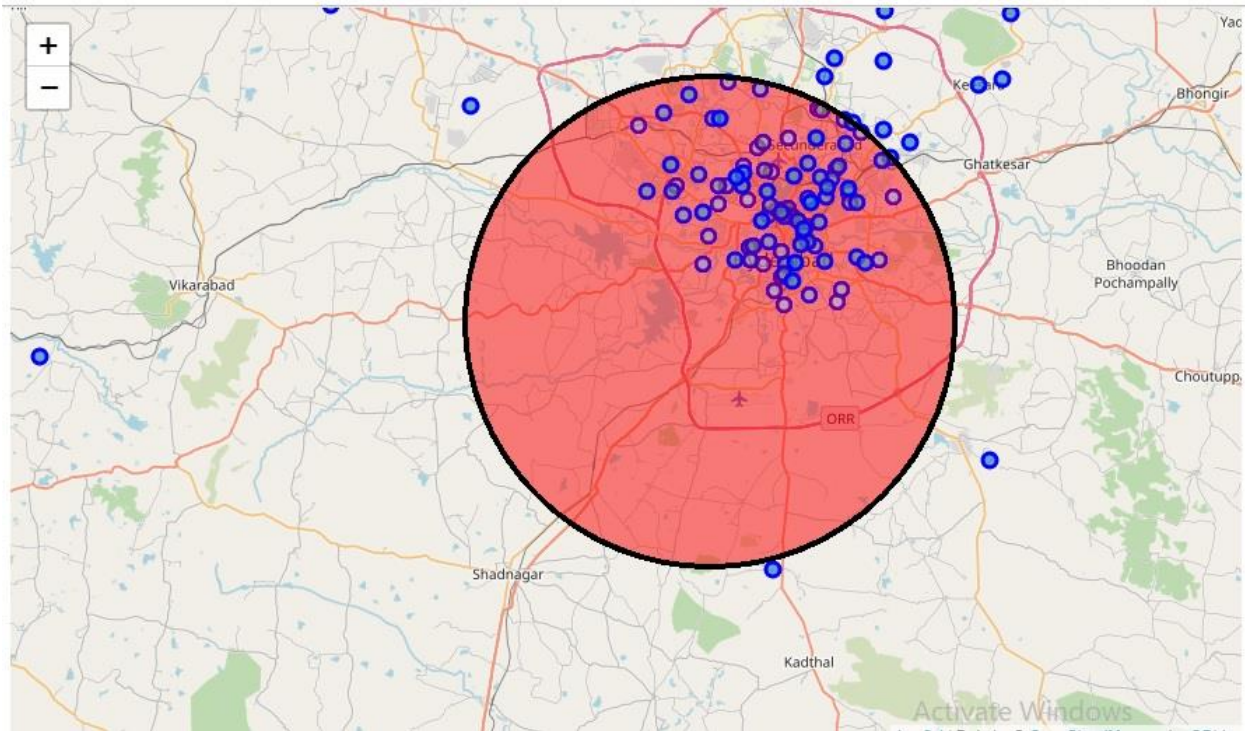
The obtained latitude and longitude values for all the neighborhoods are then used to plot all postal offices/neighborhoods in the map of Maharashtra. It is observed that a lot of these are far away from the centre of the city. The circle with a 25km radius from the centre of the city has been marked for reference.



**Map with all postal offices**



The previous map has been appropriately zoomed to highlight the relevant neighborhoods in Hyderabad.

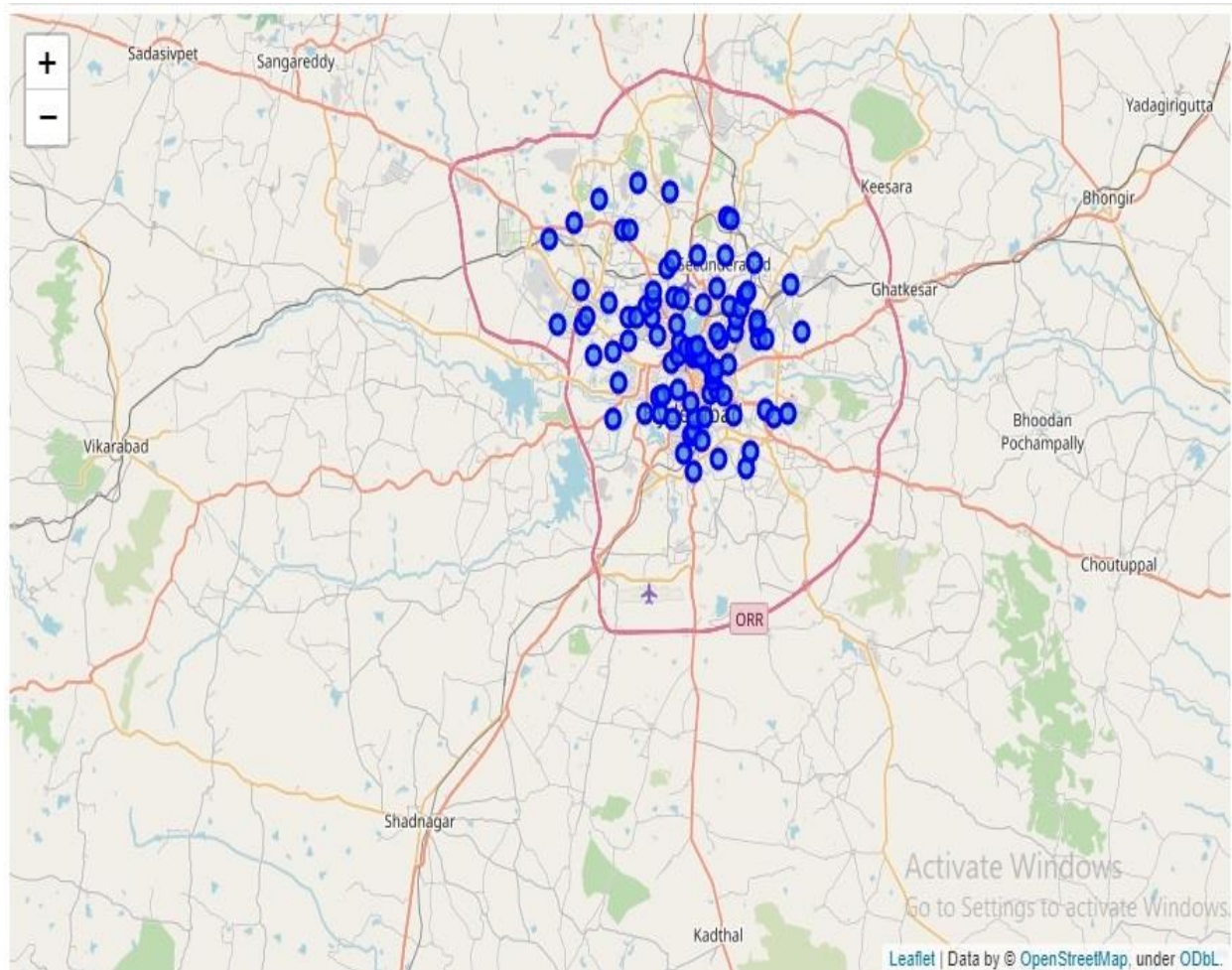


Thus, the dataframe is shortened to only those neighborhoods which are within 25Kms from the city centre and the final dataframe is as follows:

	name	latitude	longitude	dist_from_centre
0	Hydershahkote	17.357552	78.395056	6.053577
1	Rajendranagar (K.V.Rangareddy)	17.361597	78.427675	6.975046
2	Falaknuma	17.331841	78.465849	7.423260
3	Kishanbagh	17.360833	78.442748	7.659282
4	CRPF	17.318871	78.475286	7.912198
5	Bahadurpura	17.357067	78.454542	8.111570
6	Karwan Sahu	17.373489	78.441210	8.779740
7	Aliabad	17.345630	78.472680	8.794500
8	Golconda	17.383507	78.401141	8.873587
9	Kulsumpura	17.373727	78.445194	9.008066
10	Murad Nagar	17.348612	78.475765	9.248297
11	Uppuguda	17.340814	78.484183	9.609638
12	Moghalpura	17.357528	78.477084	9.929229
13	Mangalhat	17.378769	78.460440	10.365114
14	High Court	17.369165	78.472133	10.374933
15	Kanchanbagh	17.328056	78.500169	10.723105



The map of Hyderabad with relevant neighborhoods is as follows:



The dataframe created after getting the response from the Foursquare API as nearest venues within 1000m from a neighbourhood is as follows:

	Neighborhood	Neighborhood Latitude	Neighborhood Longitude	Venue	Venue Category	Venue Latitude	Venue Longitude
25	Aliabad	17.345630	78.472680	Hotel Shah Ghouse	Snack Place	17.347899	78.470986
26	Aliabad	17.345630	78.472680	Star Bawarche	Asian Restaurant	17.347700	78.470613
27	Aliabad	17.345630	78.472680	Scoops Ice Creams & Chat	Ice Cream Shop	17.343173	78.475036
28	Aliabad	17.345630	78.472680	Classic Gym	Gym	17.345489	78.476223
29	Aliabad	17.345630	78.472680	Anand Pan Shop	Smoke Shop	17.343041	78.468887
30	Aliabad	17.345630	78.472680	Kamdhenu Traders	Department Store	17.343593	78.468325
31	Aliabad	17.345630	78.472680	MedPlus	Pharmacy	17.347991	78.477954
32	Aliabad	17.345630	78.472680	Lee's Chinese Restaurant	Chinese Restaurant	17.347883	78.466605
33	Aliabad	17.345630	78.472680	cafe needs	Café	17.345678	78.479279
34	Aliabad	17.345630	78.472680	Volta Pan Shop	Food & Drink Shop	17.350301	78.478755
47	Moghalpura	17.357528	78.477084	Chowmahala Palace	History Museum	17.359300	78.471617
48	Moghalpura	17.357528	78.477084	Pista house	Indian Restaurant	17.356609	78.473637
49	Moghalpura	17.357528	78.477084	Charminar	Monument / Landmark	17.361555	78.474677
50	Moghalpura	17.357528	78.477084	Govind Ki Idly	South Indian Restaurant	17.364132	78.473771

The above table is grouped by neighborhood names and the count of every column is taken to find out the number of venues present in every neighborhood. Neighborhoods which have less than 10 venues are discarded since there isn't enough data to cluster these neighborhoods properly.

The resulting table further filters the possible neighborhoods and the final neighborhoods which shall be grouped into clusters.

The resulting table is one-hot encoded in order to create all unique 'Venue Category' values into columns with a value of 1 or 0.

One(1) meaning that it falls under that category.

The table created after one-hot encoding is as follows:

	Neighborhood_Name	Accessories Store	Afghan Restaurant	Airport	American Restaurant	Andhra Restaurant	Arcade	Arts & Crafts Store	Asian Restaurant	Athletics & Sports	...	Sporting Goods Shop	Sports Bar	Stad
0	Aliabad	0	0	0	0	0	0	0	0	0	...	0	0	
1	Aliabad	0	0	0	0	0	0	0	1	0	...	0	0	
2	Aliabad	0	0	0	0	0	0	0	0	0	...	0	0	
3	Aliabad	0	0	0	0	0	0	0	0	0	...	0	0	
4	Aliabad	0	0	0	0	0	0	0	0	0	...	0	0	

5 rows × 143 columns

This dataframe is then grouped by the neighborhood name again and the mean of every column is taken as the values for every row after grouping. The resulting table is as follows:

	Neighborhood_Name	Accessories Store	Afghan Restaurant	Airport	American Restaurant	Andhra Restaurant	Arcade	Arts & Crafts Store	Asian Restaurant	Athletics & Sports	...	Sporting Goods Shop	Sports Bar	Stad
0	Aliabad	0.000000	0.000000	0.00	0.000000	0.000000	0.000000	0.000000	0.100000	0.000000	...	0.000000	0.000000	
1	Banjara Hills	0.000000	0.000000	0.00	0.017241	0.000000	0.000000	0.000000	0.017241	0.000000	...	0.000000	0.000000	
2	Barkatpura	0.000000	0.000000	0.00	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	...	0.000000	0.000000	
3	Bazarghat	0.000000	0.000000	0.00	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	...	0.000000	0.000000	
4	Begumpet	0.000000	0.000000	0.05	0.000000	0.000000	0.050000	0.000000	0.050000	0.000000	...	0.000000	0.000000	
5	Central Secretariat	0.000000	0.000000	0.00	0.000000	0.000000	0.000000	0.025000	0.000000	0.000000	...	0.000000	0.000000	
6	Cyberabad	0.000000	0.000000	0.00	0.000000	0.050000	0.000000	0.000000	0.100000	0.000000	...	0.000000	0.000000	
7	Dr.B R Ambedkar O.U	0.000000	0.000000	0.00	0.000000	0.000000	0.000000	0.027778	0.000000	0.000000	...	0.027778	0.000000	
8	Film Nagar	0.000000	0.000000	0.00	0.000000	0.000000	0.000000	0.000000	0.043478	0.000000	...	0.000000	0.000000	
9	Gagan Mahal	0.000000	0.000000	0.00	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	...	0.031250	0.000000	
10	High Court	0.000000	0.000000	0.00	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	...	0.000000	0.000000	
11	Himayathnagar	0.000000	0.000000	0.00	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	...	0.025000	0.000000	
12	Hyderabad Jubilee	0.000000	0.000000	0.00	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	...	0.000000	0.000000	
13	Hyderabad Public School	0.000000	0.000000	0.00	0.000000	0.000000	0.025000	0.000000	0.000000	0.000000	...	0.025000	0.000000	
14	Hyderguda	0.000000	0.000000	0.00	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	...	0.020000	0.000000	
15	IICT	0.000000	0.000000	0.00	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	...	0.000000	0.000000	

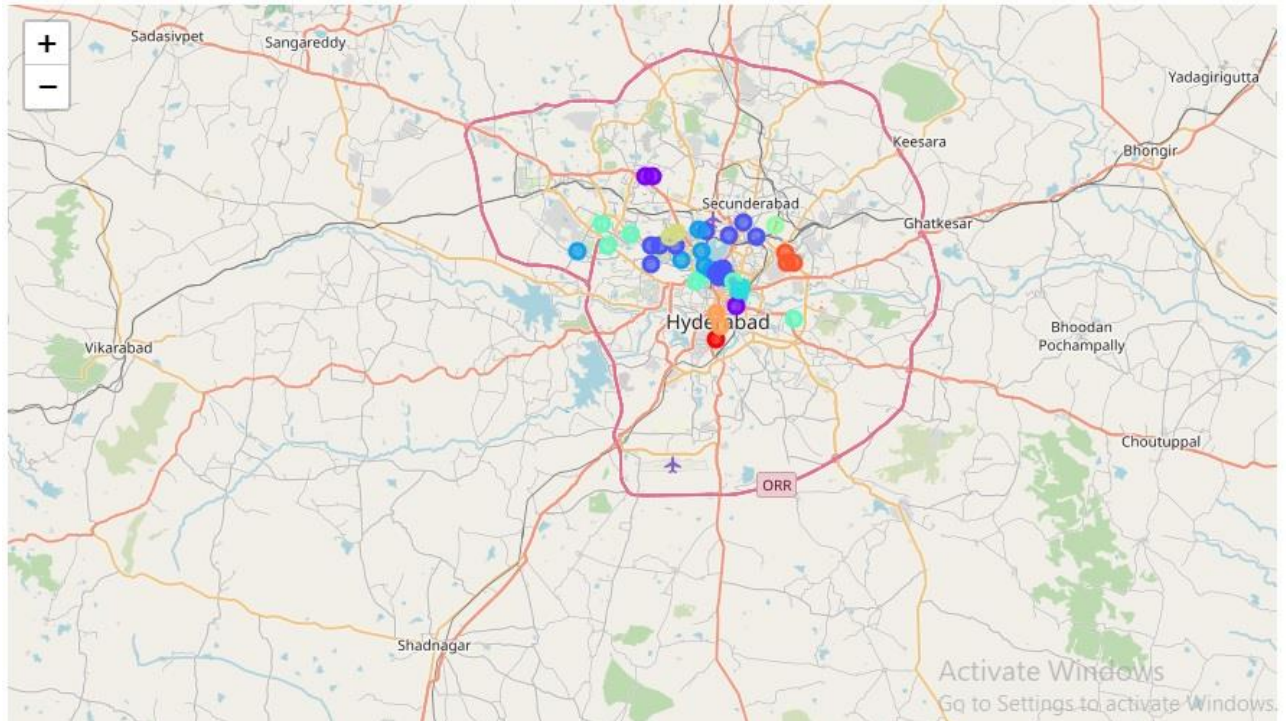
The above dataframe is fed to the K-Means clustering algorithm after dropping the 'Neighborhood\_Name' column to create clusters amongst these 41 neighborhoods.

A dataframe is created which consists of the names of the neighborhoods, their latitudes and longitudes as well as their assigned cluster number. The dataframe is as follows:

	name	latitude	longitude	dist_from_centre	Cluster_Number
0	Aliabad	17.345630	78.472680	8.794500	0
1	Banjara Hills	17.417746	78.439901	13.266589	3
2	Barkatpura	17.392612	78.496968	14.076201	4
3	Bazarghat	17.402787	78.460449	12.605049	5
4	Begumpet	17.444020	78.462482	16.818313	2
5	Central Secretariat	17.408266	78.468948	13.583746	3
6	Cyberabad	17.430929	78.369898	14.542945	5
7	Dr.B R Ambedkar O.U	17.408743	78.475459	13.999356	2
8	Film Nagar	17.414057	78.410503	12.281829	2
9	Gagan Mahal	17.409415	78.480091	14.338214	2
10	High Court	17.369165	78.472133	10.374933	8
11	Himayathnagar	17.402214	78.483624	13.922391	2
12	Hyderabad Jubilee	17.430519	78.419332	14.186551	2
13	Hyderabad Public School	17.444566	78.456737	16.653768	3
14	Hyderguda	17.401279	78.480489	13.636556	2
15	IICT	17.423913	78.539015	19.701343	9

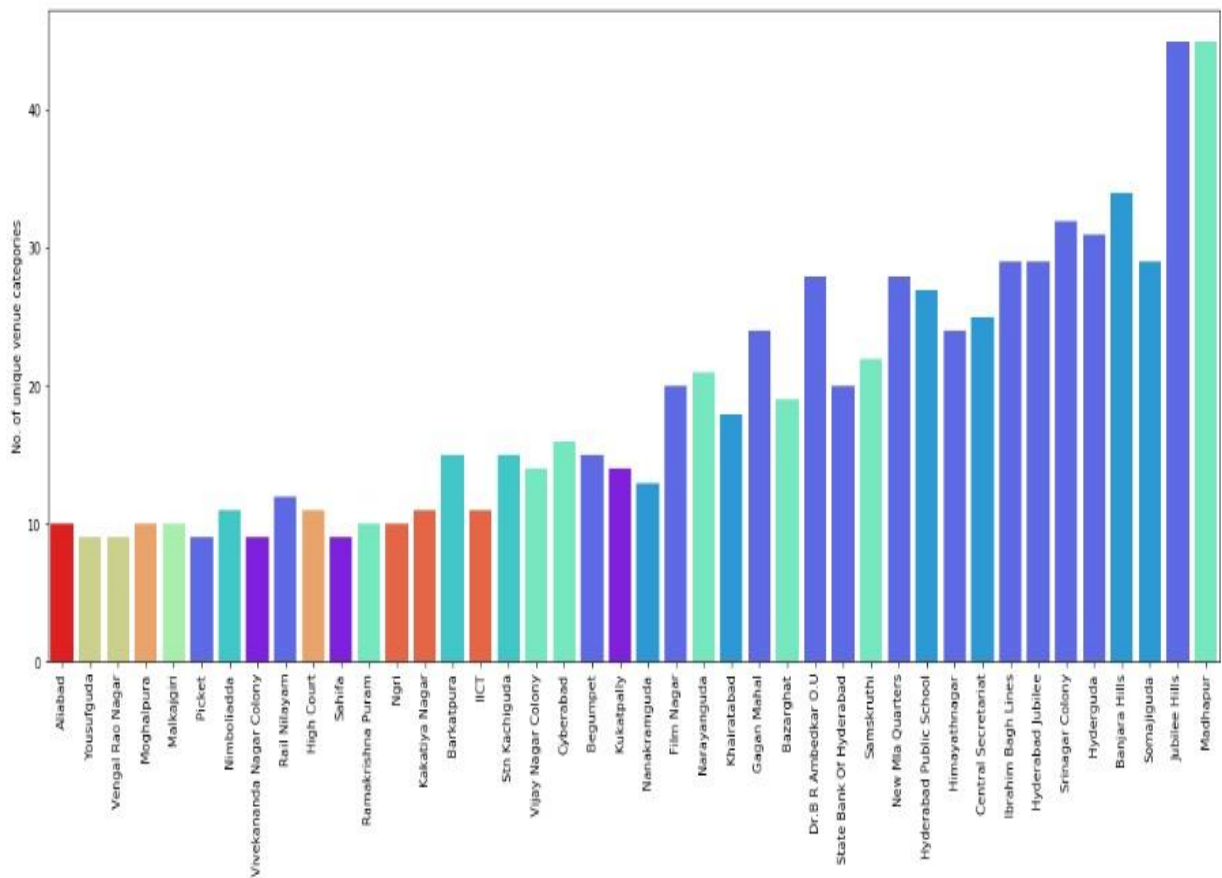


The map of Hyderabad is plotted with circle markers for every neighborhood which are color coded according to the cluster number of every neighborhood. The map is as follows:



## Analysis

A bar graph was plotted using the number of unique venue categories along the Y-axis with the corresponding neighborhood name on the X-axis. This is done to estimate the quality of a neighborhood based on the assumption that the best neighborhood will have the highest diversity of amenities which in this case is the number of unique venue categories. The bar plot has been color coded such that all bars of the same color correspond to neighborhoods in the same cluster. The bar graph is as follows:



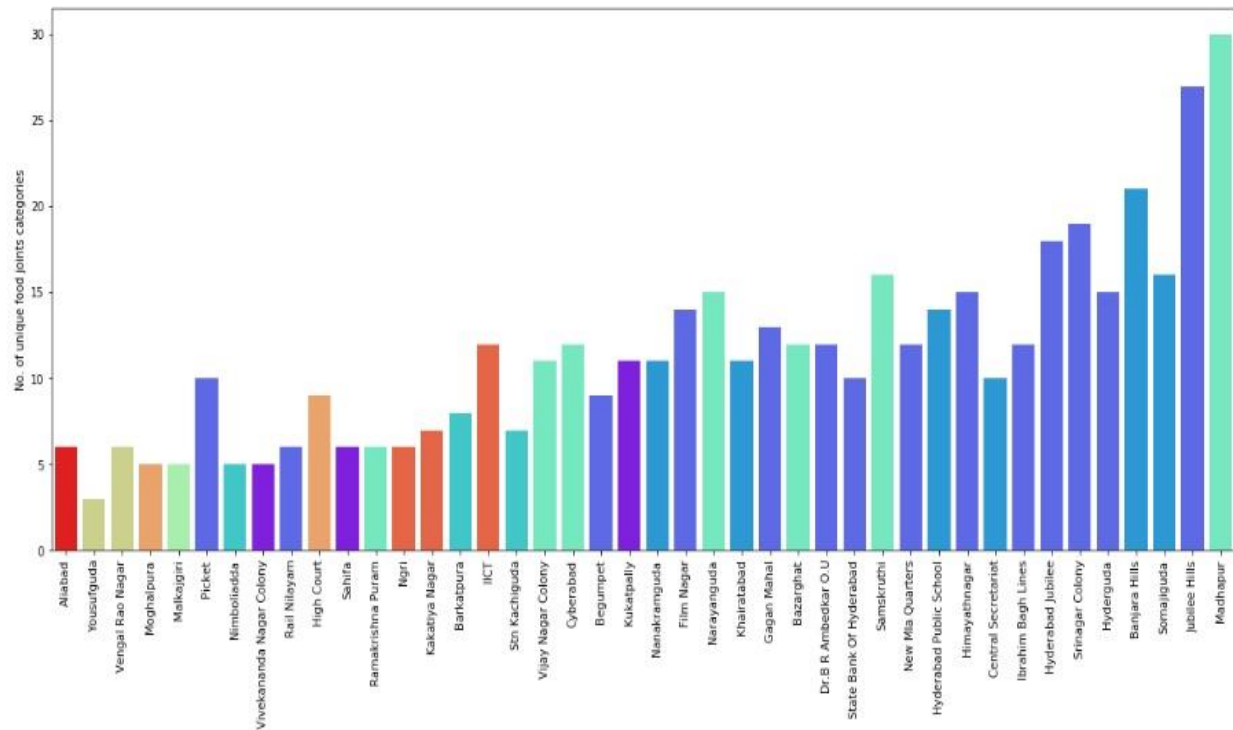
From the above plot, we can conclude that clusters pertaining to Jubilee Hills and Madhapur, i.e. cluster number 2 and 5 respectively can definitely be considered the best clusters since they consist of neighborhoods with the highest diversity of amenities.



Now, in order to judge neighborhoods on the basis of food-specific venues, the Foursquare API is again used to request for venues pertaining to foods, for all the final candidate neighborhoods and resulting dataframe is as follows:

	Neighborhood	Neighborhood Latitude	Neighborhood Longitude	Venue	Venue Latitude	Venue Longitude	Venue Category
0	Aliabad	17.345630	78.472680	Hotel Shah Ghouse	17.347899	78.470986	Snack Place
1	Aliabad	17.345630	78.472680	Star Bawarche	17.347700	78.470613	Asian Restaurant
2	Aliabad	17.345630	78.472680	Lee's	17.346575	78.476291	Chinese Restaurant
3	Aliabad	17.345630	78.472680	Pizza & Burgers	17.342968	78.468934	Sandwich Place
4	Aliabad	17.345630	78.472680	Balaji Sri Kishan Mithai Bhandar	17.343762	78.468352	Restaurant
5	Aliabad	17.345630	78.472680	Lee's Chinese Restaurant	17.347883	78.466605	Chinese Restaurant
6	Aliabad	17.345630	78.472680	cafe needs	17.345678	78.479279	Café
7	Banjara Hills	17.417746	78.439901	Guilt Trip	17.417401	78.443219	Deli / Bodega
8	Banjara Hills	17.417746	78.439901	Ofen	17.415806	78.443039	Bakery
9	Banjara Hills	17.417746	78.439901	Cafe Cubano	17.416990	78.442824	Café
10	Banjara Hills	17.417746	78.439901	Zafran Exotica	17.410325	78.437299	Indian Restaurant
11	Banjara Hills	17.417746	78.439901	Hard Rock Cafe Hyderabad	17.419400	78.448533	American Restaurant
12	Banjara Hills	17.417746	78.439901	Mamagoto	17.410408	78.437340	Asian Restaurant
13	Banjara Hills	17.417746	78.439901	WOFL	17.424944	78.438148	Breakfast Spot
14	Banjara Hills	17.417746	78.439901	The Lounge (CCD)	17.414837	78.448048	Café
15	Banjara Hills	17.417746	78.439901	Ziti	17.424055	78.438872	Restaurant

Now, the bar graph is plotted with the total number of food-joints/restaurants in a neighborhood on the Y-axis with the corresponding neighborhood name on the X-axis. This is done in order to estimate the saturation of a neighborhood with respect to the number of food-joints present in the neighborhood since we want to detect locations that are not already crowded with restaurants. The bars are color-coded according to the cluster number here as well. The bar graph is as follows:



## Results and Discussion

Based on the bar graph showing the number of unique venue categories for each neighborhood, we concluded that clusters pertaining to Jubilee Hills and Madhapur, i.e. cluster numbers 2 and 5 respectively can definitely be considered the best clusters since they consist of neighborhoods with the highest variety of amenities.

Thus, prospective stakeholders would like to open up their restaurant in similar neighborhoods since they have the largest customer base and also have a huge demand. However, it is possible that these neighborhoods might be saturated with venues. Therefore, we also look at the plot showing the number of unique venues for each and every neighborhood. Here we see that while the best neighborhoods in terms of diversity of amenities such as Jubilee Hills, Banjara Hills, Madhapur etc. are saturated with a large number of venues, there exist neighborhoods which are part of the same cluster yet have lesser number of venues. These neighborhoods might be optimal for opening up restaurants since they fall to the same cluster and are therefore similar to neighborhoods with highly diverse amenities yet have a smaller number of venues.

Instead of looking at the total number of venues of a certain neighborhood to determine the saturation, it would be better to simply look at the total number of food joints / restaurants in that neighborhood. Thus, we look at the bar graph showing the number of venue categories pertaining to food for each neighborhood. We can see that our initial assumption is still correct and neighborhoods in cluster 2 and 5 have the highest variety of food joints are thereby considered the best neighborhoods.

So, the optimal location to open up restaurants in Hyderabad would be neighborhoods in the **best clusters** i.e. clusters 2 and 5 with the **least number of existing food joints**. The ideal candidate neighborhoods can be seen from the plot of number of unique venues related to food for every neighborhood. Here we see that the neighborhoods with the least number of food joints belonging to cluster 2 and 5 are:

1. Rail Nilayam
2. RamaKrishna Puram
3. Picket
4. Begumpet
5. Cyberabad

## Conclusion

The purpose of this project was to identify areas/neighborhoods in Hyderabad close to the city-centre with low number of restaurants in order to aid stakeholders in narrowing down the search for optimal location for opening of a food-joint/restaurant. By calculating venue density distribution from Foursquare data we have first identified general clusters that justify further analysis and created groups of similar neighborhoods, and then generated extensive visualizations to justify how good a certain cluster is based on characteristics (diversity of amenities) of the neighborhoods in the clusters.

Then the Foursquare data was again used in order to find venues pertaining to food for all the neighborhoods to determine the saturation of a neighborhood with respect to number of food joints. **Optimal locations were identified as those neighborhoods which are part of the best clusters (having neighborhoods with most diverse amenities) yet have the least number of food joints.** The optimal locations obtained are:

1. Rail Nilayam
2. RamaKrishna Puram
3. Picket
4. Begumpet
5. Cyberabad

Final decision on optimal restaurant location will be made by stakeholders based on specific characteristics of neighborhoods and their locations, taking into consideration additional factors like attractiveness of each location (proximity to park or water), levels of noise / proximity to major roads, real estate availability, prices, social and economic dynamics of every neighborhood etc.