

Report

Battle of Neighborhoods

Finding the best Neighborhoods to open a Gymnasium in Hyderabad

Date: 21-7-2020

Report by: Aryan Desai

1. Introduction:

1.A. Background:

We are in the year 2020, almost a new decade! And there is no denying that an increasing number of people are turning health conscious every year. Everyone dreams of having the perfect beach body- a ripped look or a muscular physique, and a lot of them are ready to put in work too! Exposure from social media and other digital platforms have opened new roads to information for people and a whole lot of them are sticking to healthy diets and striving to burn calories. Spending time at a gym and pumping iron has become something of a trend these days. So, everybody's looking for a gym! Healthcare clubs (or Gyms as people like to call it) have become a huge business. Look around, I bet you'll find a few in your own locality. Although this is a good thing for the public, **it is a business hitch for the gym owners.**

1.B Business Problem:

Now, we understand that competition is the wellspring of improvement but too much competition will only result in a loss for all parties. And the problem here is that too many Gyms in vicinity to each other causes loss in business and no scope for new gyms to open. So, this project focuses on **finding the best possible locality for one to open a Gymnasium**. This is done by comparing the neighborhoods and the number of gyms

available in them. I have chosen Hyderabad as my city of application because I'm personally familiar with it. But do note this project can be applied to any city. By the end of this project we will find the localities that are most favorable for running a Gymnasium in Hyderabad.

2. Data Acquisition:

For this project I have used the following data: **List of Suburbs of Hyderabad, Foursquare data** on the required locations. Acquiring this data was simple and was done through web scraping. The list of Suburbs of Hyderabad was obtained from the Wikipedia page. And Foursquare data from the Foursquare API. Link to the Wikipedia page:

https://commons.wikimedia.org/wiki/Category:Suburbs_of_Hyderabad,_India

The list obtained is:

Subcategories

This category has the following 54 subcategories, out of 54 total.

A

- Abids (1 C, 13 F)
- Alwal (1 C, 1 F)
- Ameerpet, Hyderabad (3 C, 21 F)

B

- Bandlaguda, Rangareddy (1 C, 2 F)
- Banjara Hills (3 C, 25 F)
- Basheerbagh (1 C, 7 F)
- Begumpet (5 C, 9 F)
- Boduppal (2 F)
- Bolarum (3 C, 1 F)

C

- Cavalry Barracks, Hyderabad (1 C)
- Chikkadpally (7 F)

D

- Dabirpura (1 C)
- Dilsukhnagar (1 C, 2 F)
- Domalguda (3 C)

E

- Erragadda (3 F)

G

- Gachibowli (4 C, 17 F)
- Gajularamaram (2 F)

- Ghatkesar (1 C, 2 F)
- Golconda (5 C, 4 F)

H

- Hayathnagar (1 C, 14 F)
- HITEC City (5 C, 29 F)
- Hyderguda (2 F)
- Hydershakote (14 F)

J

- Jubilee Hills (3 C, 8 F)

K

- Kachiguda (1 C, 4 F)
- Khairatabad (1 C, 2 F)
- Kotli, Hyderabad (3 C, 7 F)
- Kukatpally (16 F)

L

- L. B. Nagar (16 F)

M

- Madhapur (1 C, 19 F)
- Malakpet (3 C, 2 F)
- Malkajgiri (3 C, 6 F)
- Manikonda (8 F)
- Masab Tank (4 F)
- Mehdiplotnam (1 C)
- Miyapur (5 F)

- Moazzam Jahl Market (16 F)
- Moula-Ali (3 C, 5 F)

N

- Nacharam (1 C, 4 F)
- Nadergul (1 C)
- Nagole, Hyderabad (4 F)
- Nampally (2 C, 10 F)
- Narayanguda (1 C, 4 F)
- Nizampet (2 C, 32 F)

O

- Old City (Hyderabad, India) (8 C, 26 F)

P

- Pedda Amberpet (1 F)

S

- Sanathnagar (8 F)
- Sanjeeva Reddy Nagar (10 F)
- Shamirpet (3 C, 5 F)
- Shamshabad (1 C, 4 F)
- Sitaphalmandi (1 C, 1 F)
- Somajiguda (5 F)

T

- Tarnaka (1 C, 6 F)
- Trimulgherry (1 C, 3 F)

Foursquare API was used with the help of URL. The website of foursquare that we use here is:

<https://developer.foursquare.com/>

3. Methodology:

Libraries used:

As a database, GitHub repository is used in this project.

The master data which is a list of suburbs of the city of Hyderabad was acquired from the internet.

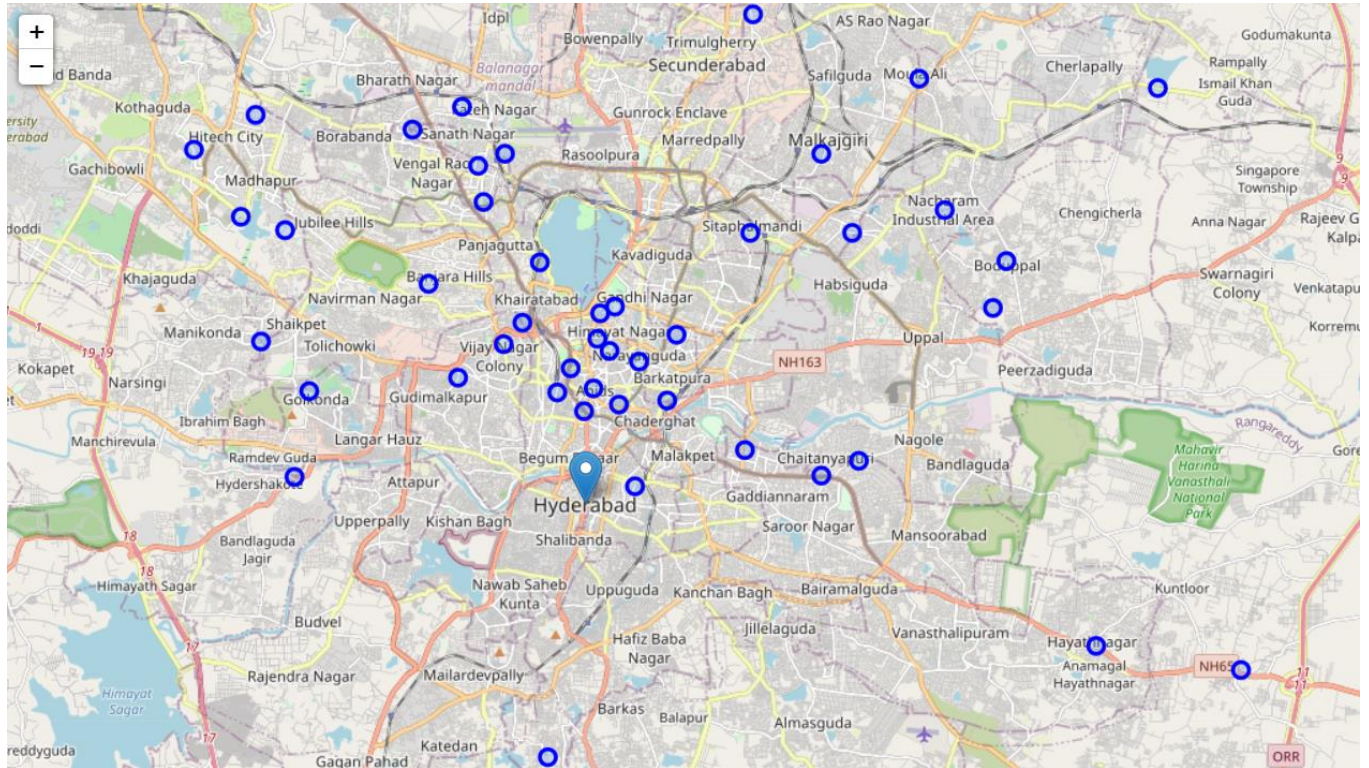
In this code I had to use multiple python libraries such as:

- **Folium:** To visualize geographic details of Hyderabad and its neighborhoods.
- **Geopy:** To locate the coordinates of the localities.
- **Geocoder:** To convert addresses to geographical information(lat/long)

For web scraping, **BeautifulSoup** library is used. This is the head of the list we scrape from the Wikipedia page.

Locality	
0	Abids
1	Alwal
2	Ameerpet, Hyderabad
3	Bandlaguda, Rangareddy
4	Banjara Hills

After loading the data about the neighborhoods into the Dataframe, we generate a map focusing on the city of Hyderabad and the localities to examine what kind of spread we are going to face.



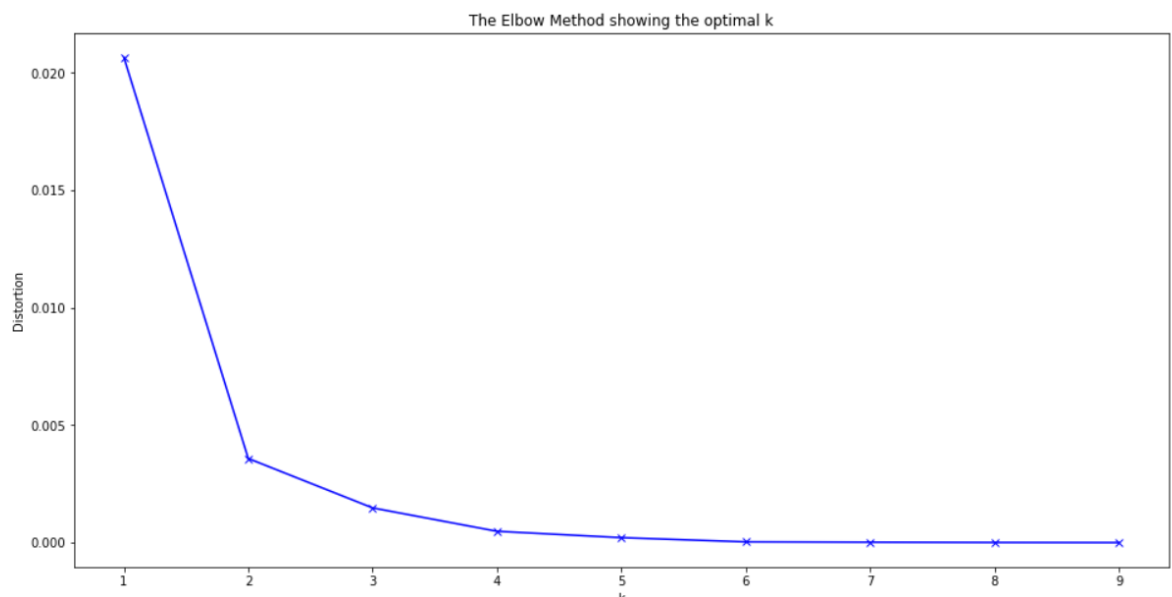
FourSquare API is used to get information on the popular destinations in those localities. We find the **top 100 most common venues** in each locality within a 2kms range. We find that there is a total of 151 unique categories of venues. So, we group them based on their localities to get the following dataframe.

	Locality	ATM	Accessories Store	Afghan Restaurant	Airport	Airport Food Court	Airport Lounge	Airport Service	American Restaurant	Arcade	...	Steakhouse	Supermarket	Taxi Stand	Tea Room	Temple	Th Restaurant
0	Abids	0.00	0.0	0.00	0.0	0.0	0.0	0.0	0.00	0.0	...	0.00	0.0	0.0	0.0	0.00	0.0
1	Alwal	0.25	0.0	0.00	0.0	0.0	0.0	0.0	0.00	0.0	...	0.00	0.0	0.0	0.0	0.00	0.0
2	Ameerpet, Hyderabad	0.00	0.0	0.00	0.0	0.0	0.0	0.0	0.01	0.0	...	0.00	0.0	0.0	0.0	0.00	0.0
3	Bandlaguda, Rangareddy	0.00	0.0	0.25	0.0	0.0	0.0	0.0	0.00	0.0	...	0.00	0.0	0.0	0.0	0.00	0.0
4	Banjara Hills	0.00	0.0	0.00	0.0	0.0	0.0	0.0	0.02	0.0	...	0.01	0.0	0.0	0.0	0.01	0.0

Our project only deals with the gyms in the localities so we find its occurrences in each one.

We also check if there's enough information on gyms in our dataset and go ahead.

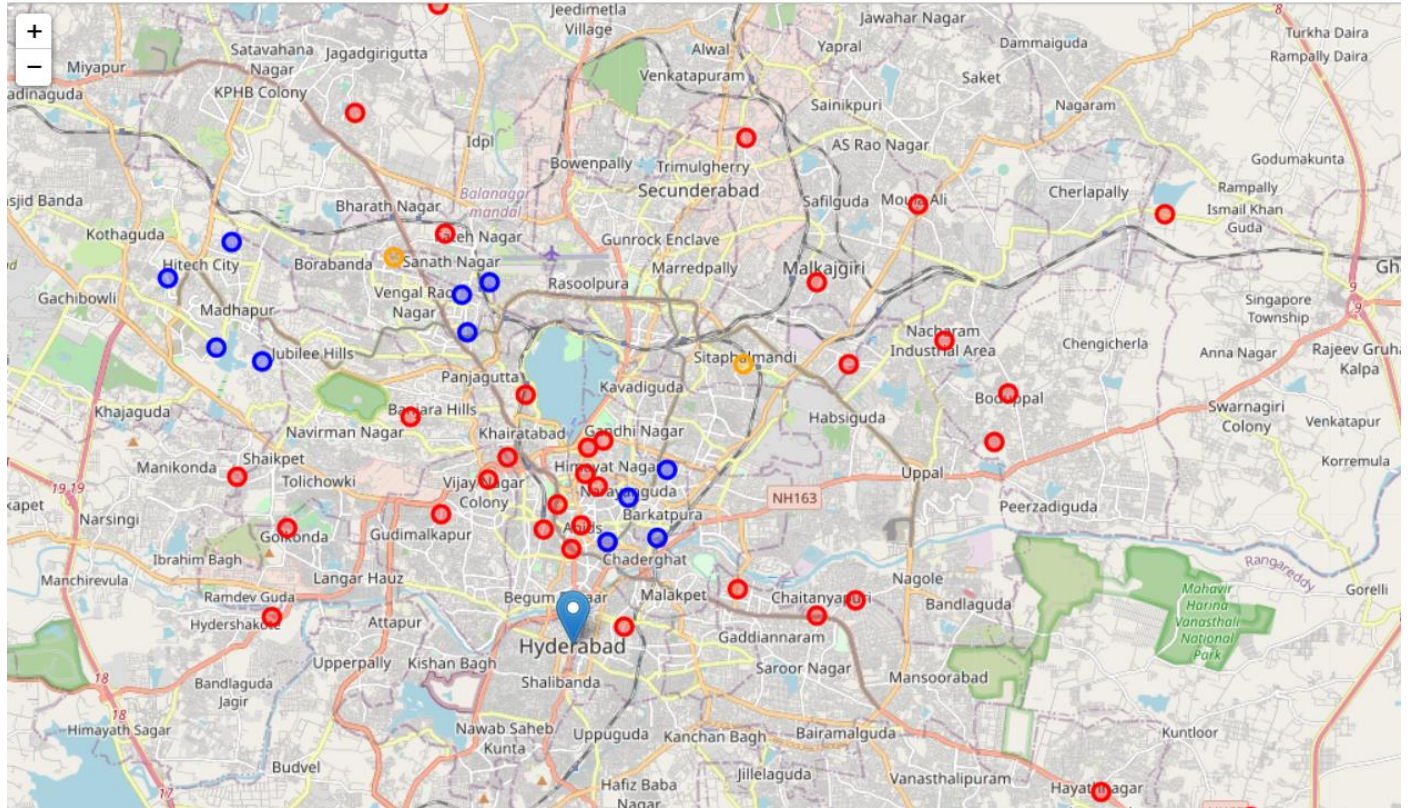
Now we move forward to clustering the localities. We initialize with 3 clusters (i.e., $k=3$) and segment them. But before commencing the mapping we use the elbow method to check if we used the right no of clusters (3).



We can observe that the “elbow” is the number 3 which is optimal for this case.

Now we can run a K-Means using k as the number 3

To further analyze our data, we have a look at the **map of localities after clustering** them. We shade the circlemarkers of the localities into different colors (Red, Blue, Orange) so as to distinguish them.



Cluster analysis:

As we can see in the map the clusters are widespread.

After conducting K-means Clustering we design 3 new dataframes, one for each cluster.

We find that there 39 localities in cluster 1:

	Locality	Gym	Cluster Label	Latitudes	Longitudes
29	Madhapur	0.010000	0	17.45694	78.39013
26	Koti, Hyderabad	0.014286	0	17.38594	78.48338
2	Ameerpet, Hyderabad	0.010000	0	17.43535	78.44861
24	Kachiguda	0.019608	0	17.38688	78.49553
23	Jubilee Hills	0.020000	0	17.42865	78.39762
46	Sanjeeva Reddy Nagar	0.014085	0	17.44438	78.44724
6	Begumpet	0.019608	0	17.44729	78.45396
19	HITEC City	0.010000	0	17.44823	78.37429
15	Gachibowli	0.010000	0	17.43181	78.38636
41	Narayanguda	0.011765	0	17.39642	78.48842
10	Chikkadpally	0.032787	0	17.40301	78.49792

2 localities in cluster 2:

	Locality	Gym	Cluster Label	Latitudes	Longitudes
14	Erragadda	0.052632	1	17.45333	78.43034
48	Sitaphalmandi	0.095238	1	17.42800	78.51700

And 11 localities in cluster 3:

	Locality	Gym	Cluster Label	Latitudes	Longitudes
43	Old City (Hyderabad, India)	0.0	2	17.408300	78.478500
32	Manikonda	0.0	2	17.401390	78.391630
49	Somajiguda	0.0	2	17.420720	78.463000
33	Masab Tank	0.0	2	17.400930	78.453620
47	Shamshabad	0.0	2	17.242020	78.425870
34	Mehdipatnam	0.0	2	17.392630	78.442190
35	Miyapur	0.0	2	17.421020	78.582440
36	Moazzam Jahi Market	0.0	2	17.384480	78.474420
45	Sanathnagar	0.0	2	17.458760	78.443100
37	Moula-Ali	0.0	2	17.465770	78.560180
38	Nacharam	0.0	2	17.433510	78.566730
39	Nagole, Hyderabad	0.0	2	17.372426	78.544543
44	Pedda Amberpet	0.0	2	17.321150	78.642370
31	Malkajgiri	0.0	2	17.447370	78.535200
40	Nampally	0.0	2	17.388970	78.467330
42	Nizampet	0.0	2	17.518330	78.381860
0	Abids	0.0	2	17.389800	78.476580

4. Results:

The output from the dataframe and the code suggest the following:

1. Cluster 1 has a mean occurrence of 0.016.
2. Cluster 2 has a mean occurrence of 0.074 which is the highest among all 3.
3. Cluster 3 has a mean occurrence of 0.00 which is the **lowest**.

The number of localities in cluster 1,2 ,3 are 11,2 and 39 respectively.

The clusters with the least no of gyms are favorable for opening one and in this case it is cluster number 3.

5. Discussions:

This Project gives us the location with the least number of gyms. But it doesn't take all factors into consideration, such as population of the locality, surrounding atmosphere, the kind of gym, etc. Therefore, making a decision solely based on these results is not optimal.

I recommend that the user take a good look at all the parameters affecting his business before starting one. Any location close to the center of the city would be suitable because of its proximity from other locations. The standard of living of that area is yet another factor one must absolutely consider.

Nevertheless, this project suffices for finding the neighborhood with the least competition. And this code can be applied to any city with a few minor changes.

6. Conclusion:

Thereby after analyzing all the results and modifying any required area, we can conclude that Cluster 3 has the localities most favorable to open a Gymnasium. Therefore, it is the **Neighborhoods in Cluster 3** that are recommended if one is willing to open a Gym.

Thank you