Capstone - The Battle of Neighborhoods Report

Contents

1 Introduction 2

1.1. Background 2

1.2. Problem 3

1.3. Interest 3

2 Data Acquisition & Cleaning 3

2.1 Data Acquisition 3

2.2 Data Cleaning 4

3 Methodology 5

3.1 Exploratory Data Analysis 5

3.1.1 Neighborhoods in Hyderabad 5

3.2 Modeling 6

4 Results 7

5 Discussion 10

6 Conclusion 10

1 Introduction

1.1. Background

I am an Indian and lot of my friends live in various parts of the world. I am taking this specific scenario based on my own experience because not just one or two, but my many friends asked me about places to live in Hyderabad since they wanted to move to the city. Of course, they are all in Information technology industry and Hyderabad is one of the biggest growing cities in India when it comes to IT advancement.

Since I have been living in Hyderabad for more than 10 years now, I have seen the increased surge in IT industries and number of small, medium and big IT companies like Amazon, Google, Microsoft, Deloitte, OnePlus to name a few expanding their footprint in the city of pearl and continuing in 2019 -2020.

Now, because of this many people have either moved or planning to moving in Hyderabad for work and people who don't know much about the city either rely on their friends, if any, to know about the city and places they could select to reside having amenities of their interest close by or look for specific information on web site or sometimes even come to the city first and then start searching for information or end up relying upon brokers.

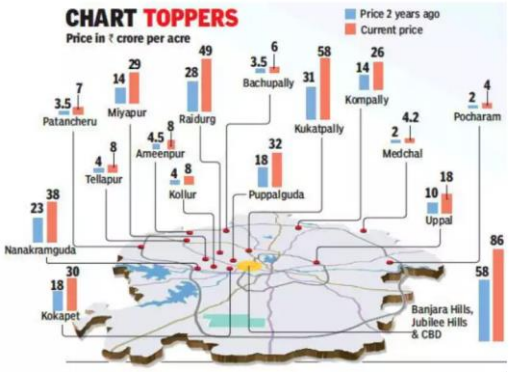
What if you are abroad and wanted to relocate, wouldn't it be nice to have some key information about the neighborhoods in Hyderabad before hand to help you taking the decision on moving to the neighborhood having things that you would consider to be close to your place?

Should you also invest in the city or may be open a business here….? Hyderabad has grown tremendously in last 10 years as evident from this chart that shows the 100% growth in land prices in last two years…

Price movement in key micro-markets

LOCALITY	AVERAGE 'ASK' RATE (PER SQ FT)	QOQ CHANGE	YOY CHANGE
Beeramguda	2,600-3,100	3%	15%
Manikonda	3,600-4,400	4%	14%
Kukatpally	4,100-5,400	4%	14%
Nizampet	2,800-3,500	2%	12%
Miyapur	3,700-4,600	4%	13%
Gachibowli	4,800-5,800	4%	10%
Kondapur	4,900-5,800	4%	15%
Chandanagar	3,600-4,800	4%	16%
Pragati Nagar	3,100-3,700	4%	7%
Nallagandla	4,900-5,700	5%	10

Source – www.99acres.com



Source – timesofindia.indiatimes.com

1.2. Problem

This project aims to find specific places of interest in and around various neighborhoods of Hyderabad to enable a family or individuals moving to Hyderabad either from within India or from any part of the world to view neighborhoods and close by venues of their interest and take a decision where they would like to either buy or rent a hose.

We are going to need neighborhoods listing of Hyderabad and their coordinates information in order to call foursquare API for listing specific areas of interest. I'll be considering specific categories places of which would be listed in an around neighborhoods based on the below table.

These are the specific categories I have selected to be necessary for anyone to know to decide on a living place.

Category	Category ID
School	4bf58dd8d48988d13b941735
Hospital	4bf58dd8d48988d196941735
Pharmacy	4bf58dd8d48988d10f951735
Stationery Store	52f2ab2ebcbc57f1066b8b21
Indian Restaurant	4bf58dd8d48988d10f941735
Convenience Store	4d954b0ea243a5684a65b473
Chinese Restaurant	4bf58dd8d48988d145941735

These categories ID are taken from foursquare API documentations from here : https://developer.foursquare.com/docs/resources/categories . More categories can be added if needed for analysis.

1.3. Interest

People who are considering relocating to Hyderabad will be interested to identify & explore its neighborhoods and specific venues around each neighborhood in order to decide on a location based.

2 Data Acquisition & Cleaning

2.1 Data Acquisition

The data required for this project is a list of Hyderabad neighborhoods with their latitude and longitude information.

The first source of this dataset is a scraped from Wikipedia page that contains the listing of neighborhoods in Hyderabad.

@link1 - https://en.wikipedia.org/wiki/Category:Neighbourhoods_in_Hyderabad,_India

Each neighborhood listed on this page points to neighborhood information page having more information about the neighborhood including its latitude and longitude information. For example, more information of "Abids" neighborhood is linked from the above to its own page –

@link2 - https://en.wikipedia.org/wiki/Abids.

Using these two web pages data will be scrapped and dataset with the following columns would be created for Hyderabad city.

- **Neighborhood** – Name of the neighborhood in Hyderabad
- **Latitude** – Latitude of the neighborhood
- **Longitude** – Longitude of the neighborhood

2.2 Data Cleaning

The data cleaning would be done in two steps using **Beautiful Soup** library in python.

The first step would be to scrap the data from @link1 and stored in tabular format. After the web scrapping, string manipulation would be performed to get the neighborhoods names and their corresponding links information.

The second step would involve going through each neighborhood one by one and web scrapping the linked page(@link<X>) to extract latitude and longitude information. For each neighborhood information page, we'll extract the page content and perform string manipulation to extract Latitude and Longitude Information for the final dataset. We could see the progress of individual neighborhoods being processed from fig 2.2.

```
Processing : A. S. Rao Nagar
Processing : A.C. Guards
Processing : Abhyudaya Nagar
Processing : Abids
Processing : Adikmet
Processing : Afzal Gunj
Processing : Aghapura
Processing : Aliabad, Hyderabad
Processing : Alijah Kotla
Processing : Allwyn Colony
Processing : Alwal
Processing : Amberpet
Processing : Ameerpet
Processing : Ashok Nagar, Hyderabad
Processing : Asif Nagar
Processing : Attapur
Processing : Azamabad, Hyderabad
Processing : Azampura
Processing : Badichowdi
```

Fig -2.2 Individual neighborhoods being processed

Using the final dataset containing the neighborhoods in Hyderabad along with latitude and longitude information, we can find specific venues within 3km meter radius of each neighborhood by connecting to the Foursquare API. This would return a json file containing venues of all categories taken up to be listed in each neighborhood and converted to a data frame.
The data frame will be used to create clusters using K-mean clustering algorithm to cluster similar neighborhoods together.

The final dataset of neighborhoods in Hyderabad looks like as below in figure 2.2

	Neighborhood	Latitude	Longitude
0	A. S. Rao Nagar	17.283333	78.316667
1	A.C. Guards	17.396944	78.456944
2	Abhyudaya Nagar	18.991378	72.844164
3	Abids	17.366111	78.476111
4	Adikmet	17.406944	78.512778

Fig 2.2 Subset of the final dataset of Hyderabad Neighborhoods

3 Methodology

3.1 Exploratory Data Analysis

3.1.1 Neighborhoods in Hyderabad

We have quite several neighborhoods identified in the city and that probably will help in making a close observation of the places in and around. There are 199 neighborhoods in the Hyderabad and they are visualized on a map using folium on python (see fig 3.1.1).

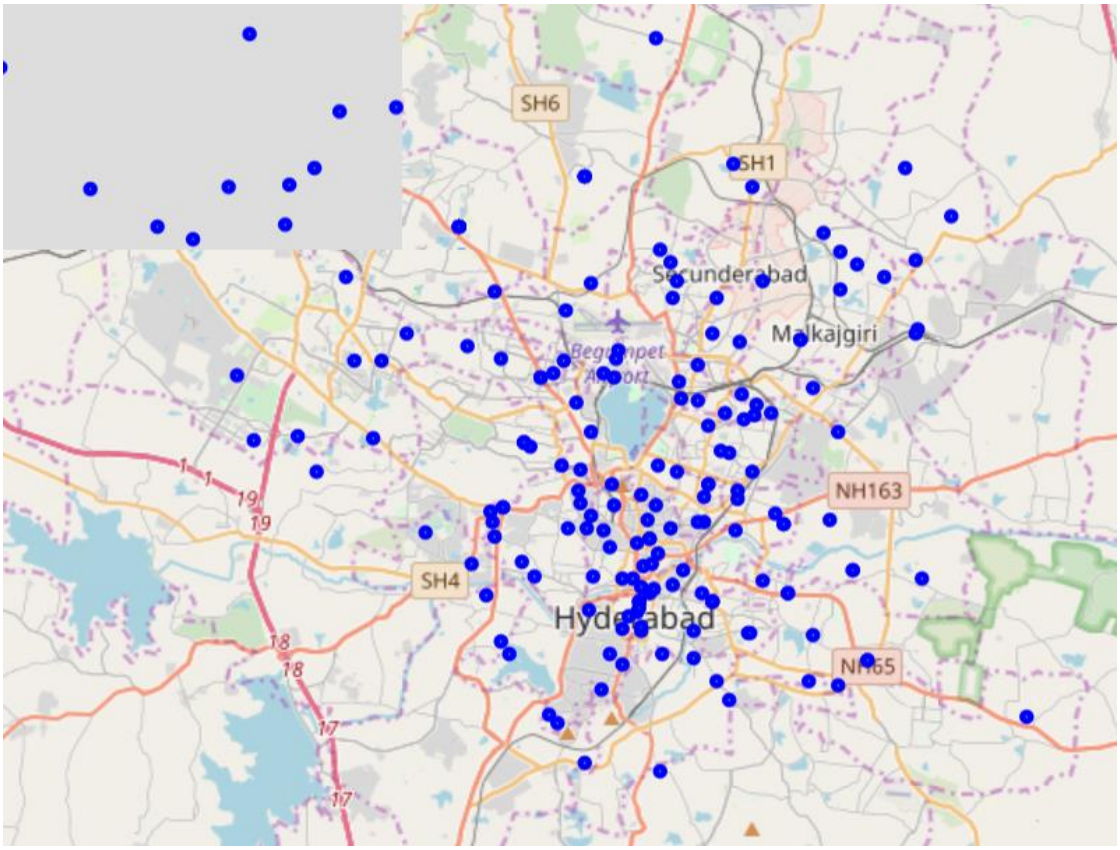


Fig 3.1 Neighborhoods in Hyderabad, India

If needed one can probably focus on top 10 or 20 neighborhoods in Hyderabad and use the API to get more details in and around those neighborhoods.

3.2 Modeling

Using the final dataset containing the neighborhoods in Hyderabad along with the latitude and longitude, we can find specific venues that we selected within a 2KM radius of each neighborhood by connecting to the Foursquare API. This returns a json file containing all the venues in each neighborhood which is converted to a panda's data frame. This data frame contains all the venues along with their coordinates and category (see fig 3.2.1).

	Neighborhood	Neighborhood_Latitude	Neighborhood_Longitude	Venue	Venue_Latitude	Venue_Longitude	Venue_Category
0	A.C. Guards	17.396944	78.456944	DineHill	17.405256	78.451674	Indian Restaurant
1	A.C. Guards	17.396944	78.456944	Paradise Restaurant	17.403602	78.452848	Indian Restaurant
2	A.C. Guards	17.396944	78.456944	Prince Hotel	17.394736	78.442410	Indian Restaurant
3	A.C. Guards	17.396944	78.456944	Sandarshini	17.405330	78.451656	Indian Restaurant
4	A.C. Guards	17.396944	78.456944	Al-humdulillah Hotel	17.390728	78.462992	Indian Restaurant
5	A.C. Guards	17.396944	78.456944	Sarvi	17.412698	78.449704	Indian Restaurant
6	A.C. Guards	17.396944	78.456944	Hyderabad House	17.402775	78.456992	Indian Restaurant
7	A.C. Guards	17.396944	78.456944	Kamat Hotel	17.404330	78.467602	Indian Restaurant
8	A.C. Guards	17.396944	78.456944	Paradise Food Court	17.412113	78.465443	Indian Restaurant

Fig 3.2.1 Venue details of each Neighborhood

One hot encoding is done on the venues data. (One hot encoding is a process by which categorical variables are converted into a form that could be provided to ML algorithms to do a better job in prediction). The Venues data is then grouped by the Neighborhood and the mean of the venues are calculated, finally the 10 common venues are calculated for each of the neighborhoods.

We see that we have quite a number of restaurants venue returned from the data and it might make sense to exclude restaurants from the data to focus on rest of the categories. There is abundance of restaurants available in Hyderabad and should not be of any issue finding one.

The resulting dataset looks much more meaningful after excluding restaurants category. Fig 3.2.2

	Neighborhood	Latitude	Longitude	Cluster Labels	1st_Most_Common_Venue	2nd_Most_Common_Venue	3rd_Most_Common_Venue	4th_Most_Common_Venue
0	A.C. Guards	17.396944	78.456944	0	Hospital	Convenience Store	Medical Center	Grocery Store
1	Abhyudaya Nagar	18.991378	72.844164	0	Hospital	School	Pharmacy	Medical School
2	Abids	17.366111	78.476111	2	Hospital	School	Trade School	College Administrative Building
3	Adikmet	17.406944	78.512778	1	Convenience Store	Hospital	Pharmacy	Trade School
4	Afzal Gunj	17.373333	78.470833	2	Hospital	Chaat Place	School	College Administrative Building
5	Aghapura	17.388333	78.464444	0	Hospital	Hotel	Chaat Place	School
6	Alijah Kotla	17.357222	78.476944	2	Hospital	School	Trade School	College Administrative Building
7	Allwyn Colony	17.483333	78.416667	4	Hospital	Trade School	College Administrative Building	College Classroom
8	Amberpet	17.390833	78.523611	1	Convenience Store	Hospital	School	Trade School
9	Ameerpet	17.437500	78.448200	0	Hospital	Pharmacy	Convenience Store	Trade School

Fig: 3.2.2

To help people find similar neighborhoods we will be clustering similar neighborhoods using K - means clustering which is a form of unsupervised machine learning algorithm that clusters data based on predefined cluster size. We will use a cluster size of 5 for this project that will cluster neighborhoods into 5 clusters. The reason to conduct a K- means clustering is to cluster neighborhoods with similar venues together so that people can shortlist the area of their interests based on the venues/amenities around each neighborhood.

4 Results

After running the K-means clustering we can access each cluster created to see which neighborhoods were assigned to each of the five clusters. See figure 4.0

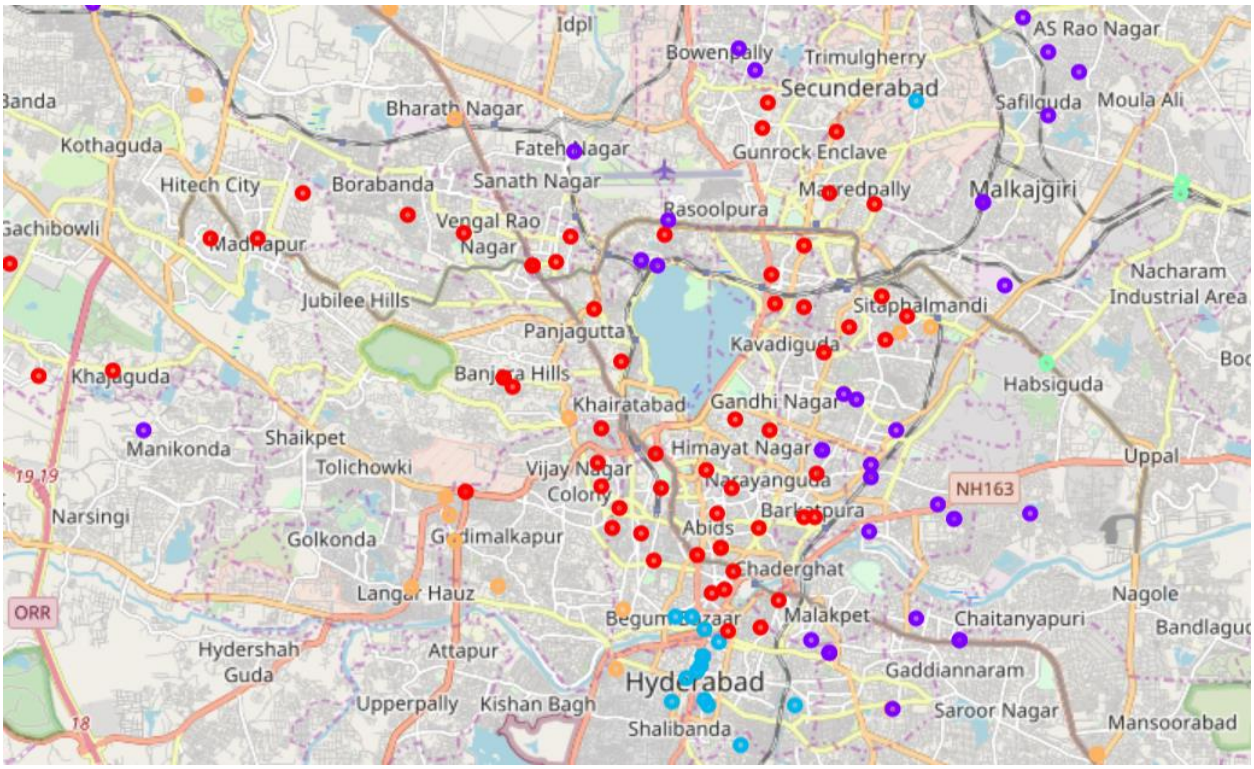


Fig – 4.0 – 5 clusters in Hyderabad

We see that cluster 0 shown in red is the biggest cluster also dense in the center and Top 1, 2 & 3 venues should have all the venues of interest covered by anyone moving to Hyderabad.

1st_Most_Common_Venue	1st_Most_Common_Venue
Hospital	63
Hotel	1
Preschool	1

2nd_Most_Common_Venue	2nd_Most_Common_Venue
Chaat Place	6
Convenience Store	23
Department Store	2
Food	1
High School	3
Hospital	1
Hotel	14
Irani Cafe	1
Medical Center	2
Pharmacy	8
Preschool	1
School	1
Student Center	1
Trade School	1

	3rd_Most_Common_Venue
3rd_Most_Common_Venue	
Chaat Place	6
College Administrative Building	1
Convenience Store	13
Department Store	1
Emergency Room	1
Food	2
Food Court	1
Grocery Store	3
High School	8
Hospital	1
Hotel	6
Irani Cafe	2
Medical Center	2
Pharmacy	5
School	4

5 Discussion

The aim of this project is to help people who want to relocate to the Hyderabad in India, they can choose the neighborhoods to which they want to relocate based on the common venues any family would want to have close by.

For a family I feel that the neighborhoods in Cluster 0 are more suitable due to the venues in that cluster, these neighborhoods have common venues as categorized which is ideal for a family. The choices of neighborhoods may vary from person to person.

6 Conclusion

This project helps a person get a better understanding of the neighborhoods with respect to the required venues in that neighborhood. It is always helpful to make use of technology to stay one step ahead i.e. finding out more about places before moving into a neighborhood. We have just taken family into consideration for people moving from outside Hyderabad to have a better sense of the neighborhood they would want to move considering venues of their interest proximity. This data set can also be looked from the perspective of neighborhoods crime statistics and Airbnb price predictor based on location….something to consider in future analysis.