

The Battle of Neighborhoods

Opening a Shopping Mall in Hyderabad , Telangana, INDIA

Capstone Project of IBM Data Science

-by Pranay Thatikonda, July 2021



1. Introduction

1.1 Background

A shopping center, is a group of shops built together, sometimes under one roof(wikipedia). Most of the shopping is done at the shopping malls by us because it is a place where you can buy what you want, have some snack or food in between and if you are tired you can watch a movie or have fun playing games or any activities, all at a single place. So, most of us prefer shopping at shopping malls instead of shopping any other store or outlet. It is a place where any startup or someone can showcase their product and let the market know about it. Now a days, flash mobs are being conducted in Malls, social media festivals are also celebrated by conducting flash mobs and shows. Real estate business, Restaurants, Groceries, Jeweleries, without leaving a corner each and every business is brought into these malls.

As a result this is a really huge benefit for the Businesses as well as owners/real-estate, Businesses will get into huge sales and owners of the outlets get the income as rent, as well as real estate developers get the profit as well as fame by constructing large and beautiful shopping malls.

However there are plenty number of malls which are located in the city of Hyderabad. So, selecting a suitable location for business as well as public is difficult. The location of the mall ultimately decides whether it will get into profits or losses.

1.2 Business Problem

The objective of the project " The Battle of Neighborhoods - Opening a Shopping Mall in Hyderabad , Telangana, India" is to explore and analyze the neighborhoods and the shopping malls in Hyderabad , Telangana and select the best locations in the city to open a new Shopping Mall. Using geocoder library to get the latitude and longitudes for neighborhoods. Foursquare api to explore neighborhoods, Data Science methodology and K means clustering this project will be a solution to the business problem : __In the city of Hyderabad , Telangana state of India, if someone is looking to open a Shopping Mall, where would you recommend that they open it?__

1.3 Target audience of this Project

Indian mall developers are looking to add over 65million square feet of new mall supply by 2022 end. Hyderabad one among the top cities comprise 11% share out of 7 cities. Hyderabad malls have average vacancy of 15%, lease rates of 100/- to 160/- rupees per square feet, At present, Hyderabad has over four million sq ft of the area, says an article by Times of India. Hyderabad expected to get 6 million sq ft of shopping mall space in 3 years, says Economic Times report. With the increasing supply of shopping malls selecting a location which will meet the customer demand is difficult. So, this project helps in finding a good location to open a new shopping mall to Business stakeholders, property developers, investors as well as customers.

2. Data

2.1 Below is the required data:

- List of neighborhoods in the city of Hyderabad. This will give us the scope of the areas to open shopping malls near to the residential areas and meet the demand.
- Latitude and Longitudes of the extracted neighborhood data. This will act as the input parameters to the foursquare api to explore a neighborhood.
- Shopping malls data, this will be the output from the foursquare api and help us to create clusters on the neighborhoods depending on the frequency of malls in a given radius of a particular neighborhood.

2.2 Sources and methods to get the data:

- The list of neighborhoods in Hyderabad is taken from a wikipedia page (https://en.wikipedia.org/wiki/List_of_neighbourhoods_in_Hyderabad), contains 31 neighborhoods in total, combining all the zones. By using web scraping technique we will get the neighborhood data from the wikipedia page.
- We will then get the Latitude and Longitude of each neighborhood using geocoder library and attach these coordinates to our neighborhood data
- Then with the help of Foursquare api calls we will send the coordinates of each neighborhoods and get the venues details of shopping malls and create the shopping malls data of each neighborhood.

This project will be completed with the help of data science skills, data cleaning, exploring, analyzing, visualizing using folium maps in particular. Also takes help from foursquare api to get the shopping mall details which are near to each neighborhood. And finally using a machine learning technique 'K-means clustering' to cluster malls into different categories.

2.3 Snapshot of datasets

Below is the screenshot of the head of the dataset that consists of list of neighborhoods in the column 1 and its coordinates fetched with the help of geocoder as Latitude and Longitude in columns 2 and 3.

This data will be used to send to foursquare api to fetch the venue categories that includes shopping mall categories as well if found any in particular neighborhood

	Neighborhood	Latitude	Longitude
0	Sanathnagar	17.456965	78.443478
1	Ghatkesar	NaN	NaN
2	Mehdipatnam	17.394263	78.434251
3	Balanagar	17.476746	78.422108
4	Malkajgiri	17.448344	78.528973

However we have dropped the rows that contains NaN's as those are anyways outskirts locations and had not found coordinates using geocoder library.

With the help of Foursquare api we are able to fetch the nearby venues and filtered shopping mall categories out of those venues and the frequency/mean of those malls present in the entire neighborhoods is shown as below.

	Neighborhoods	Shopping Mall
0	Alwal	0.000000
1	Amberpet	0.047619
2	Ameerpet	0.010000
3	Balanagar	0.142857
4	Dilsukhnagar	0.076923

3. Methodology

3.1 Data collection:

First, we need to fetch the list of neighborhoods of the city Hyderabad, Telangana, India from wikipedia page (https://en.wikipedia.org/wiki/List_of_neighbourhoods_in_Hyderabad), with the help of requests, BeautifulSoup libraries I was able to scrape the list of neighborhoods of unordered list and clean them using for loops and regular expressions and get the unique list and store in a dataframe.

Later we used geocoder library and fetched the Latitude and Longitudes of each neighborhoods, some of them which are located in the out skirts are removed from our dataframe as their coordinates are not having found using our library. Then we have merged Neighborhoods and coordinates into one single dataframe.

After getting the data required we have visualized the neighborhoods using folium maps, this can help us to locate the neighborhoods and verify that they are present inside the city of Hyderabad.

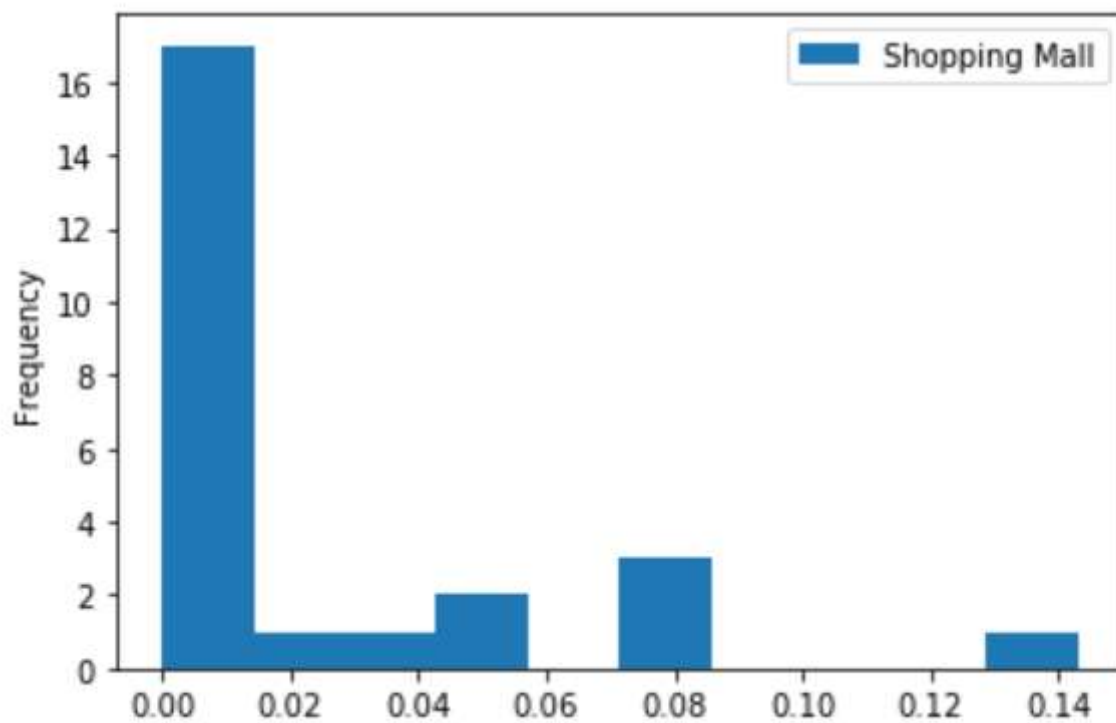
3.2 Analytic approach:

Here we have used FourSquare API to get the top 100 venues that are within the radius of 2 Kilometers by making API calls and passing the coordinates that we have fetched in the above section, using foursquare api credentials of developer account. Foursquare will return the venue data in JSON format and we extract useful information from it such as venue name, venue category, venue coordinates.

With this data we will create a dataframe and find unique types of venue categories and check number of venues each neighborhood has. After this we have analyzed each neighborhood by grouping them by neighborhood column and taking the mean of frequency of occurrence of each venue category.

Finally we will filter the 'Shopping Mall' venue category from the above created dataframe so that we can continue our methodology on our project to analyze shopping mall data.

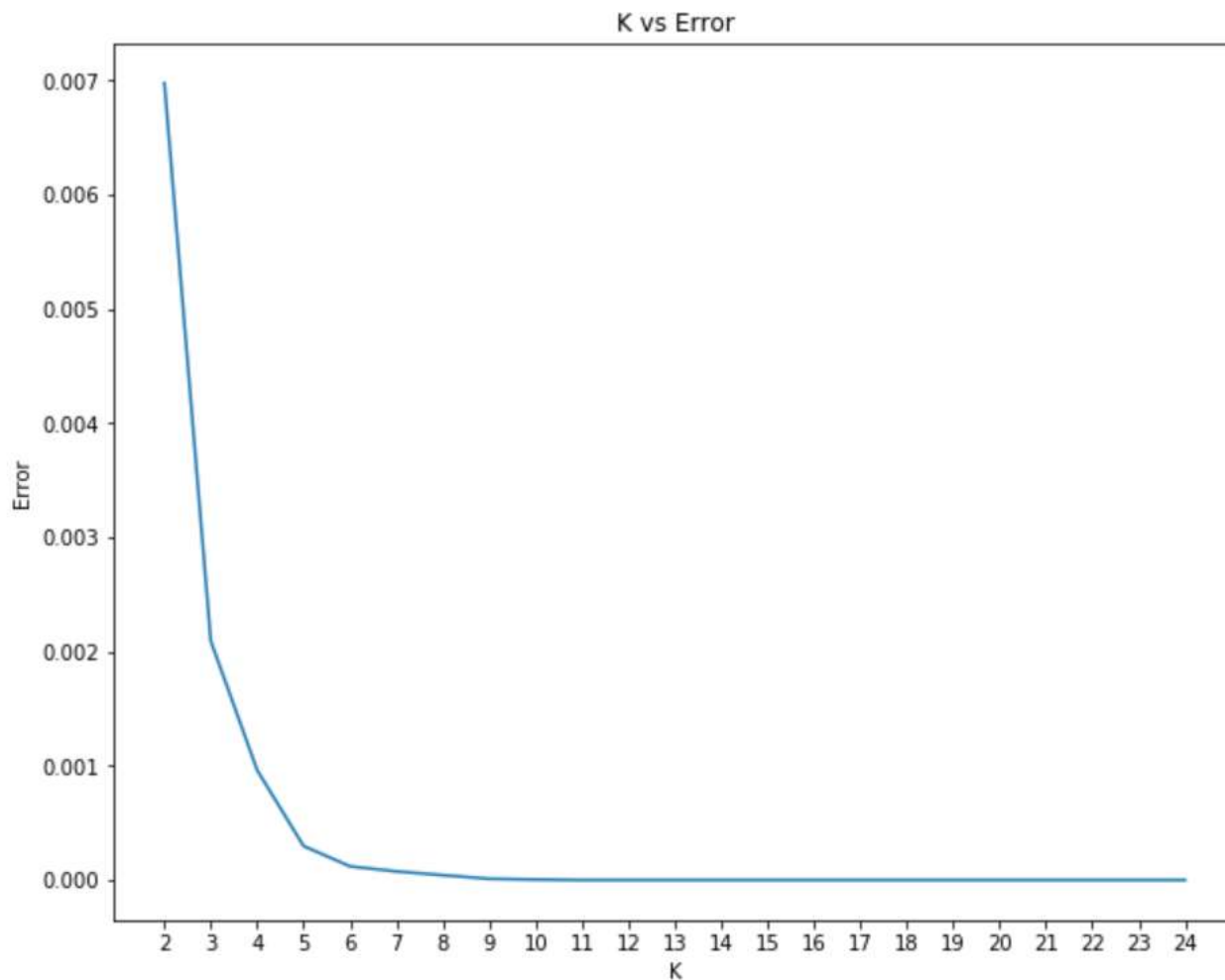
3.3 Exploratory data analysis:



First we plotted a histogram to check the presence of shopping malls in the city, X-axis represents 'mean of frequency of occurrence of Shopping Malls' and Y-axis represents 'Number of Neighborhoods'

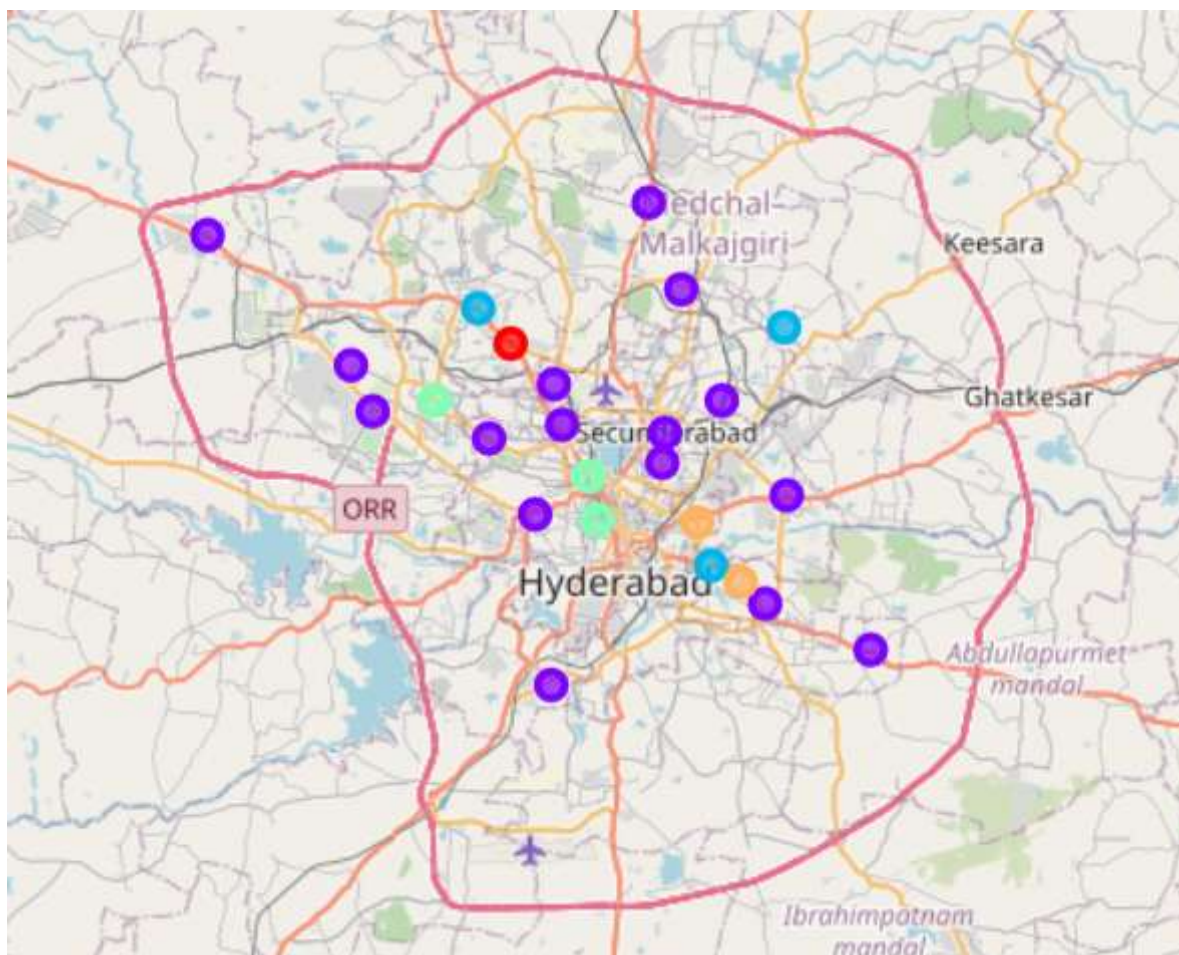
3.4 K-means Clustering:

It's the time to make clusters of Neighborhoods on basis of Shopping malls presence in the neighborhoods. We have used K-means clustering algorithm, K represents the number of clusters. We have found the K value using 'get_inertia' method by passing 2 to 25 as K value and finding the error for each model. We have drawn a line plot using k and error values as below one and the elbow point is considered as the optimal K value for K-means clustering, for us from below plot we can use K=5 as an optimal value



We have clustered the neighborhoods into 5 clusters based on the above curve. Then add the cluster labels to a column of our dataframe and create a map using folium and color the neighborhoods based on the cluster numbers. Below is the map obtained.

Based on the presence of number of shopping malls in each neighborhoods, this will help business to answer the question where to open a new shopping mall in the city of Hyderabad.



4. Results

The results obtained using K-means clustering based on 5 clusters of frequency of occurrence of "Shopping Mall" category in each neighborhoods are:

- Cluster 0 (Red color) : With only 1 Neighborhood with mean value more than all the clusters as it is an industrial area with small shopping malls around it
- Cluster 1 (Purple color) : With almost no shopping malls present in these neighborhoods, some of the areas might be having one or two shopping malls but our model did not detect it because of the radius(2000m) that we have chosen.
- Cluster 2,3,4 (Blue, Green, Orange) : These clusters 2,3 and 4 has enough number of shopping malls in these neighborhoods with different frequencies of mean. Neighborhoods in clusters 2 and 3 comes under main city or popular areas of the city Hyderabad.

5. Discussion

From the observations noted from the above map, most of the malls are located only at one third of the city neighborhoods of Hyderabad, with moderate to highest in number are from cluster 2,3 and 4. Apart from this in cluster - 0, it has small shopping malls with large in number when compared to all the neighborhoods. This represents a great opportunity and high potential to open a new shopping mall in this area, as the population in this area is also high as there is very little to no competition from the existing small malls.

Meanwhile shopping malls in cluster 1 are very very low and even does not exist in some neighborhoods, they are suffering from undersupply of shopping malls, we have to select a location such that it should be the centre of multiple neighborhoods to get more convenient to multiple neighborhoods.

We have also noted that there is no over supply of shopping malls in the city, that means more shopping malls are required for the city people. Finally, property developers, business investors are advised to avoid neighborhoods in cluster 2,3 and 4 as they are already having minimum number of shopping malls when compared to other cluster neighborhoods.

Also, it would have been an additional benefit if we had increased the radius to get the venues from foursquare api, as some of the neighborhoods are literally far from each other and we might have missed some of the malls situated in those areas.

6. Conclusion

In the above study, we have explored and analyzed various neighborhood venues, particularly shopping malls in Hyderabad city of Telangana, India using Data Science.

We used a dataset, neighborhoods fetched from wikipedia page, combining location data fetched using geocoder library, collected venue data from Foursquare API. We performed Exploratory Data Analysis and clustering on the datasets shown above in our pursuit of solution to our problem. We were able to find satisfactory answers to the questions we posed before the study.

The study is based on limited data, but it is nevertheless a significant step in looking at shopping malls data in the city of Hyderabad. This study can be reproduced easily for any other cities across the world.