Location Analysis for Shopping Malls in Hyderabad

By Siva Chaitanya M

Introduction

The city of Hyderabad is one of the most densely populated metropolitan cities in the Indian subcontinent. Leveraging its growth in population over the past 5 decades, businesses have seen a rising interest of modern choices among Hyderabadis. Shopping malls have sprouted all around the city, from the outskirts to its core, giving the people weekend timeouts, movie buffs a continuous stream of movies to watch in the mall's sophisticated layouts. Businesses gain from putting up a stall in a mall where the footfall is high due to the various choices a mall provides in the small area. This project helps in analyzing the density of various shopping malls around the city, in different neighborhoods.

Problem Statement:

What is my competition for opening up a new shopping mall in Hyderabad?

Problem Description:

Competition makes or breaks a business. Competition impacts a company's business strategy. Some may choose to go head-on, and some may choose to play it safe. It depends on the risk levels taken by the stakeholders and the kind of marketing they choose to do. As shopping malls contain huge variety of businesses within them, it is important to understand the level of uncertainty in the market they are faced with from their peers.

Target Audience:

For shopping mall owners and investors, it is very important to understand where the competition lies while considering to open up a new shopping mall.

Data Collection & Resources

<u>Description of the Data:</u>

The data that forms the basis of this analysis contains the location of various malls in the city along with the localities to pinpoint where the competition lies. The following data was collected for this analysis:

<u>Location</u>: Latitude and longitude with the help of one library named, Geocoder in order to obtain the values for city neighborhoods and another API named, Foursquare API for exploring the surrounding venues in a particular locality.

<u>Neighborhood</u>: Localities in the city, this data was scarped from a Wikipedia article (Link in reference section) which contained all the neighborhood information required to carry out further analysis.

Methodology

<u>Data procurement and cleaning:</u>

- Neighborhoods in Hyderabad data:

    As stated earlier, the data for the location of the localities in and around Hyderabad are procured from a Wikipedia article in the following link:

    [List of neighbourhoods in Hyderabad - Wikipedia](#)

    I have used the *Beautiful Soup* library in order to extract the required information about the various neighbourhoods in Hyderabad from the Wikipedia html page. Post that, I have converted the obtained list of neighbourhood names into a pandas dataframe using the python *pandas* library.

- Location data of Neighborhoods:

    The latitudes and longitudes for the above neighborhoods are obtained using a python library named, *Geocoder*. This library uses various APIs from different platforms like google, ArcGIS etc to obtain the required location information. The location data thus obtained after utilizing the above library, is then fed into the previous dataframe.

The pandas dataframe with neighborhoods and relevant location data looks like this:

```
In [211]:   hyd_data

executed in 33ms, finished 3 hours ago
```

	Areas	Lat	Long
0	Ameerpet	17.43535	78.44861
1	Begumpet	17.44729	78.45396
2	SR Nagar	17.49808	78.46963
3	Prakash Nagar	17.36069	78.41634
4	Punjagutta	17.42616	78.45210
...	...	...	...
239	Narsingi	17.39114	78.35790
240	Puppalguda	17.39861	78.38234
241	Shamshabad	17.27439	78.41095
242	Rajiv Gandhi International Airport	17.25466	78.42266
243	Umdanagar	17.29595	78.41551

244 rows × 3 columns

Using the *folium* library to visualise the neighborhoods on a map of Hyderabad city:
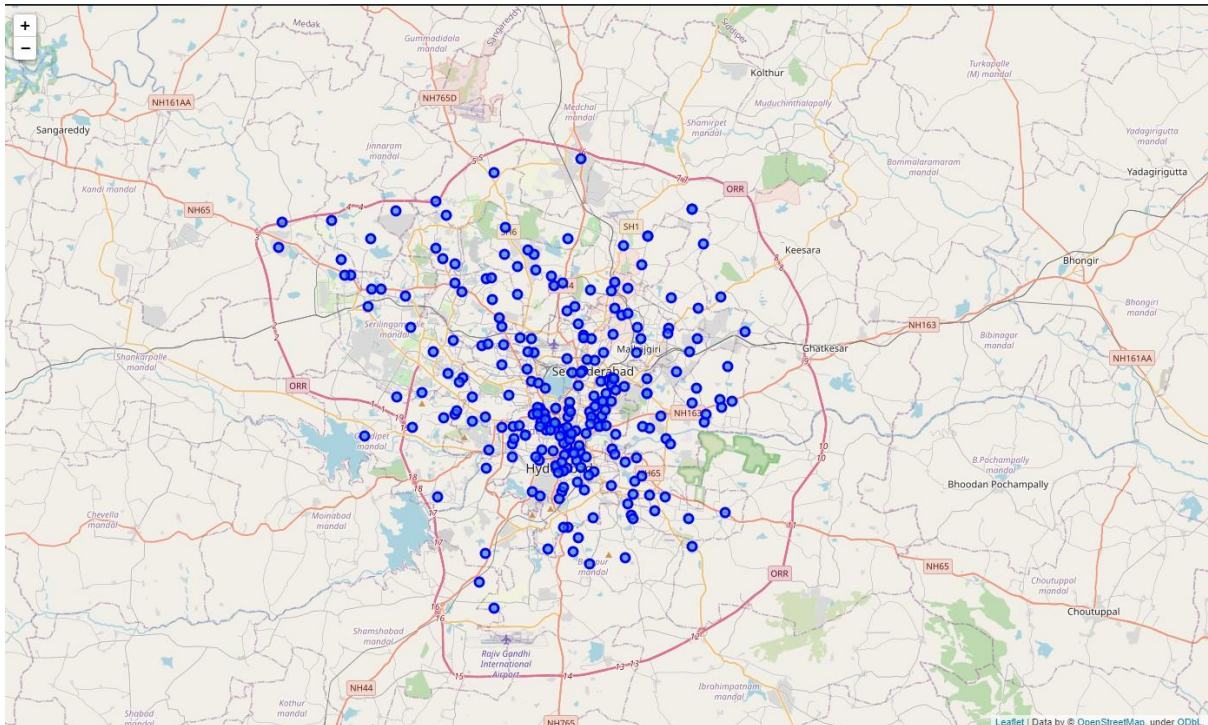
Exploring the Neighborhoods with Foursquare API:

- Foursquare API:

    Foursquare has one of the largest database of 120+ million places and is used by over 150,000 developers worldwide. With the help of Foursquare API, we will obtain many categories of the venue data such as names, locations, menus and even photos.

    After finding the list of neighborhoods and their locations, we then connect to the Foursquare API to gather information about venues inside each neighborhood and in a radius of 2000 meters. The information obtained per venue is as follows:

    - Name of the Neighborhood
    - Latitude and Longitudes of the Neighbor
    - Name of the Venue
    - Venue Latitude and Longitude
    - Category of the Venue

The resultant dataframe after merging the neighborhoods with their corresponding venues is:


```
In [297]: hyd_venues.head(25)

executed in 26ms, finished a few seconds ago
```

	Neighborhood	Neighborhood Latitude	Neighborhood Longitude	Venue	Venue Latitude	Venue Longitude	Venue Category
0	Ameerpet	17.43535	78.44861	Kakatiya Deluxe Mess	17.433435	78.447090	Diner
1	Ameerpet	17.43535	78.44861	Mekong	17.437151	78.454301	Chinese Restaurant
2	Ameerpet	17.43535	78.44861	10 Downing Street	17.435868	78.457443	Pub
3	Ameerpet	17.43535	78.44861	Kebabs & Kurries	17.432374	78.457585	Indian Restaurant
4	Ameerpet	17.43535	78.44861	Meridian Resturant	17.427355	78.452159	Fast Food Restaurant
5	Ameerpet	17.43535	78.44861	ITC Kakatiya	17.432514	78.457353	Hotel
6	Ameerpet	17.43535	78.44861	Minerva Coffee Shop	17.428839	78.455573	Indian Restaurant
7	Ameerpet	17.43535	78.44861	Chopsticks	17.429910	78.445674	Chinese Restaurant
8	Ameerpet	17.43535	78.44861	Blue Fox	17.437054	78.445912	Indian Restaurant
9	Ameerpet	17.43535	78.44861	Naturals Ice Cream	17.428536	78.439620	Ice Cream Shop
10	Ameerpet	17.43535	78.44861	Chutney's	17.426219	78.447569	Vegetarian / Vegan Restaurant
11	Ameerpet	17.43535	78.44861	Marigold	17.437116	78.454506	Hotel
12	Ameerpet	17.43535	78.44861	Southern Mirchi	17.426419	78.448560	South Indian Restaurant
13	Ameerpet	17.43535	78.44861	Lamakaan	17.421017	78.447474	Performing Arts Venue
14	Ameerpet	17.43535	78.44861	PVR Cinemas	17.426516	78.453261	Multiplex
15	Ameerpet	17.43535	78.44861	Vivanta by Taj	17.443498	78.460556	Hotel
16	Ameerpet	17.43535	78.44861	Himalayas	17.426359	78.451211	Bookstore
17	Ameerpet	17.43535	78.44861	Q-Mart	17.426681	78.438616	Department Store
18	Ameerpet	17.43535	78.44861	Lifestyle	17.436356	78.456529	Clothing Store
19	Ameerpet	17.43535	78.44861	Hotel Green Park	17.436276	78.453892	Hotel
20	Ameerpet	17.43535	78.44861	Big Byte	17.430177	78.438352	Bakery
21	Ameerpet	17.43535	78.44861	GVK One	17.419411	78.448069	Shopping Mall
22	Ameerpet	17.43535	78.44861	Chinese Pavilion	17.423456	78.448695	Asian Restaurant
23	Ameerpet	17.43535	78.44861	INOX Movies	17.419391	78.448119	Multiplex
24	Ameerpet	17.43535	78.44861	Aqua	17.423612	78.462541	Hotel Bar

- Analyzing the Neighborhoods:

  In order to figure out the distribution of the shopping malls around each neighborhood, we have to first encode the categorical values of the '*Venues*' dataset from the above mentioned dataframe using *get_dummies* method of the *pandas* library, which converts all categorical values into binary values of the numerical type. The resulting dataframe is as follows:

  Now we can go ahead and create a new data frame with only the neighborhood and their corresponding mean number of shopping malls under a new dataframe, which looks like this:


- ## Clustering using K-Means:

    Now we will perform clustering on the data by using k-means clustering. K-Means clustering is an unsupervised learning algorithm which identifies k number of centroids, and then allocates every data point to the nearest cluster, while keeping the centroids as small as possible. It is one of the simplest and popular machine learning algorithms and is particularly suited to solve the problem for this project. Before we proceed, we must find the optimum value for k, so that we achieve the most performant model for the given dataset. With the use of elbow method, we can see that the value of 4 for k is the best estimate to use in this situation.

    ### The Elbow Method using Distortion
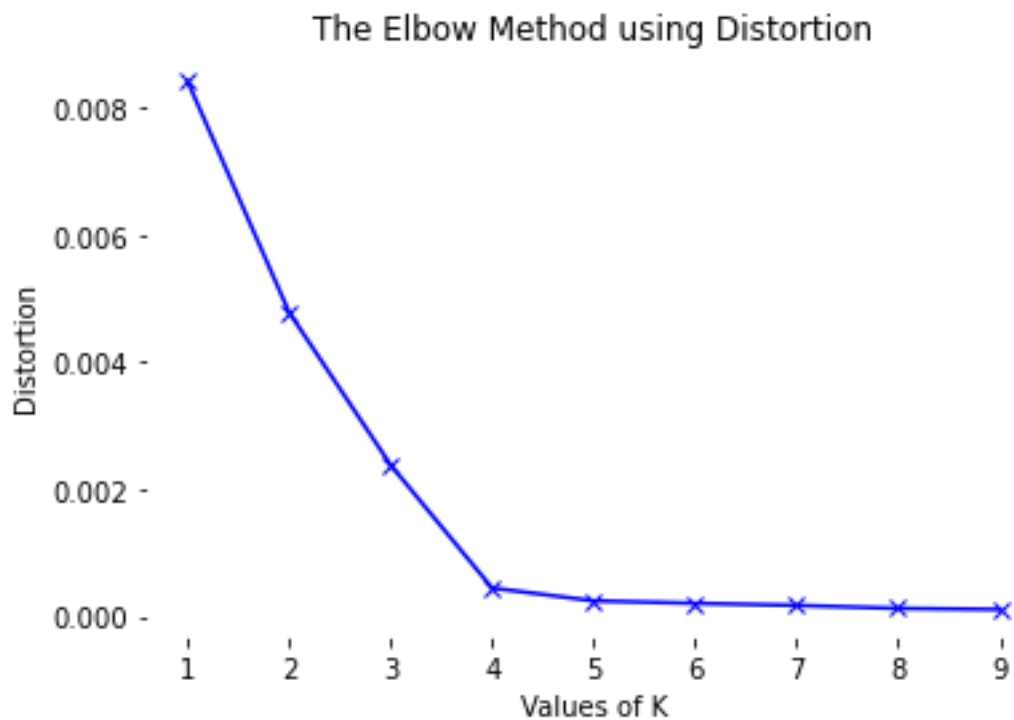
    We will cluster the neighborhoods into 4 clusters based on their frequency of occurrence for shopping malls. The results will allow us to identify the neighborhoods having fewer number of shopping malls, which will help us to answer the question as to which localities are most suitable to open new shopping malls.

Results

The results from the k-means clustering show that we can categorize the neighborhoods into 4 clusters based on the frequency of occurrence of number of shopping malls:

- Cluster 0: Neighborhoods with almost no shopping malls in 94 localities
- Cluster 1: Neighborhoods with moderate number shopping malls in 79 localities
- Cluster 2: Neighborhoods with high number shopping malls in 11 localities
- Cluster 3: Neighborhoods with low number of shopping malls in 60 localities

The results of the clustering are visualized in the map below with cluster 0 in red color, cluster 1 in purple color, cluster 2 in blue green color and cluster 3 in dull green color.
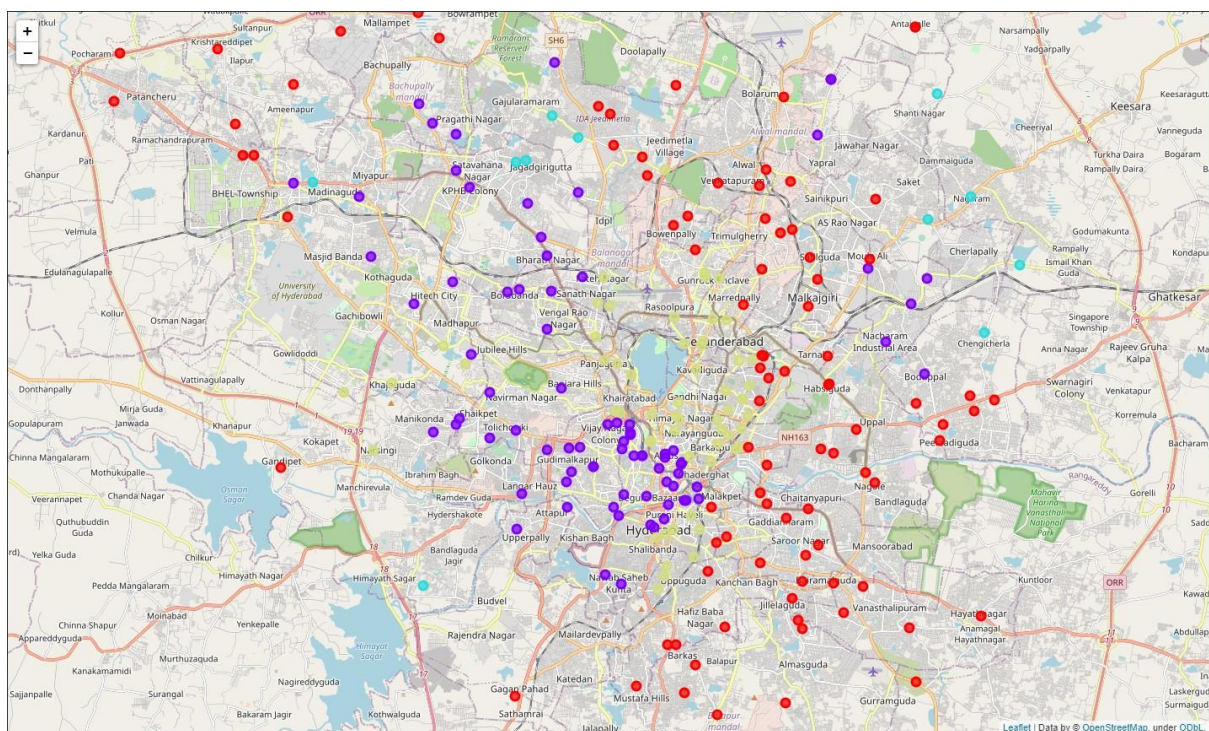
Discussions

As observations noted from the map in the Results section, a considerable amount of shopping malls is concentrated around the left part or western part of the city, with the highest number in cluster 1 and a few in cluster 2. On the other hand, cluster 0 and cluster 3 have almost none or very low number shopping malls in their neighborhoods which represents a great opportunity and high potential areas to open new shopping mall. And with the cluster 0 and cluster 3 being concentrated mostly over the eastern part of the city, having a good number of national highways going through them, they are a strong locality to consider. Therefore, this project recommends franchise/property owners to capitalize on these findings to open new shopping malls in the neighborhoods in cluster 0 with little to no competition and in cluster 3 with moderate competition which can be overcome by some

unique selling propositions to stand out in the competition. Lastly, it is advised to avoid neighborhood cluster 1 and 2 which already have a high concentration of shopping malls and suffering from intense competition.

Conclusion

In this project, we have gone through the process of identifying the business problem, specifying the data required, extracting and preparing the data, performing machine learning by clustering the data into 4 clusters based on their similarities, and lastly providing recommendations to the relevant stakeholders and investors regarding the best location to open a new shopping mall. To answer the business question that was raised in the introduction section, the answer proposed by this project is: The neighborhoods in cluster 0 and cluster 3 are the most preferred locations to open a new shopping mall. The findings of this project will help the relevant stakeholders to capitalize on the opportunities on high potential locations while avoiding overcrowded areas suffering from high competition.

References

Category: List of neighborhoods in Hyderabad. Retrieved from:

https://en.wikipedia.org/wiki/List_of_neighbourhoods_in_Hyderabad

Foursquare Developers Documentation. Retrieved from:

https://developer.foursquare.com/docs