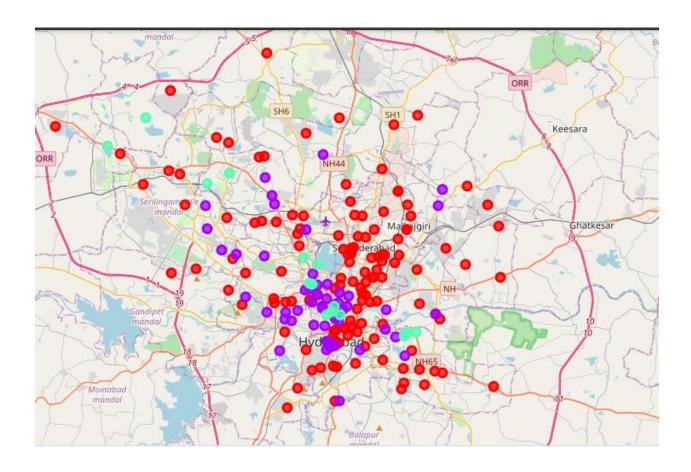
Coursera Capstone Project

The Battle of Neighborhoods – Hyderabad, India



Identifying lucrative neighborhoods in Hyderabad set up a restaurant

Introduction Section

Created By: Neha Madnekar

Introduction

There are lots of challenges to be faced while starting a restaurant business. Amongst all the decisions that are to be made location is a major discussion point and a prime concern. One needs to be very careful while choosing a location keeping an eye on the footfall available in that area and competition existing in the locality.

Hyderabad, popularly known as "City of Pearls", is a 400-year-old metropolitan. It is the sixth most populous urban agglomeration in India. Hyderabad is the most happening city in the south of India with a rich scope for food business. According to a survey, it is voted the second-best city in India for doing food business. It stands famously for food, shopping, technology, arts, and pearls.

Experiencing the social behavior revolution, City's cosmopolitan outlook and modern lifestyle has attracted people of diverse ethnicities and lifestyles. People here enjoy their leisure time with great zeal. Leisure time on weekends is given a serious thought. Hence all the cafes restaurants, theatres, parks, and amusement spots are assured, a house full crowd. Additionally, the growing social culture is boosting its growth at a faster pace. Hence, it is ideal for an individual/ group or a conglomerate restaurant chain because more people means more sales.

This project will help the audience address the following questions:

- 1. Which are some of the most lucrative localities in Hyderabad to set up a restaurant?
- 2. Which are some good areas in Hyderabad where a business can be set up with minimum competition?
- 3. How many clusters can the neighborhoods in the Hyderabad be divided into?

Business Problem: The objective of this project is to find all the major neighborhoods in Hyderabad and accordingly cluster them to identify lucrative localities to set up a restaurant. This project will be helpful to individual businessmen or restaurant-chain who want to identify optimal locations in Hyderabad to set up restaurants, hotels, food trucks etc.

Target Audience:

To summarize, this project is ideal for -

- 1. Individuals looking to start a restaurant
- 2. A restaurant-chain looking to grow its business
- 3. A conglomerate looking to diversify into restaurant business

Techniques Used:

- 1. Web Scraping: A list of all neighborhoods in Hyderabad is fetched from Wikipedia using several python requests along with their latitude and longitude details. Foursquare API is then used to access venues around these neighborhoods
- 2. Data Preprocessing and Cleansing: A data frame of all these neighborhoods mapped against their geo location details as well as venues is created indicating the presence/ absence of the venue category in that area
- Clustering: K-means clustering technique is then used to cluster all the neighborhoods based on the presence of all the venues except for the venue category – restaurants.
 Folium library is imported to help visualize the map of Hyderabad and superimpose the clustered data points onto it.

Data Requirements

This section has a total of four parts to represent all the data requirements that were fulfilled:

- 1. Neighborhood Data
- 2. Geolocation Data
- 3. Venue Data
- 4. Data Visualization

1. Neighborhood Data:

Hyderabad has many neighborhoods. The dataset with the list of all the localities in Hyderabad can be found on Wikipedia. There were total of 200 neighborhoods found in Hyderabad when the data from this Wikipedia dataset was fetched in the format shown below:

Link: https://en.wikipedia.org/wiki/Category:Neighbourhoods in Hyderabad, India

	Neighborhood
0	A. C. Guards
1	A. S. Rao Nagar
2	Abhyudaya Nagar
3	Abids
4	Adibatla
5	Adikmet
6	Afzal Gunj
7	Aghapura
8	Aliabad, Hyderabad
9	Alijah Kotla

2. Geolocation Data:

There was no latitude and longitude data found on the page. Therefore, the python package – Geocoder was used to fetch and map the neighborhoods to their respective latitudes and longitudes. These details were then transformed into a Data Frame with the format shown below:

	Neighborhood	Latitude	Longitude
0	A. C. Guards	17.395015	78.459812
1	A. S. Rao Nagar	17.411200	78.508240
2	Abhyudaya Nagar	17.337650	78.564140
3	Abids	17.389800	78.476580
4	Adibatla	17.235790	78.541300
195	Secunderabad	17.442000	78.501920
196	Serilingampally	17.482160	78.323000
197	Shah-Ali-Banda	17.357390	78.473200
198	Shahran Market	17.364890	78.476290
199	Shanker Mutt	17.399817	78.507919

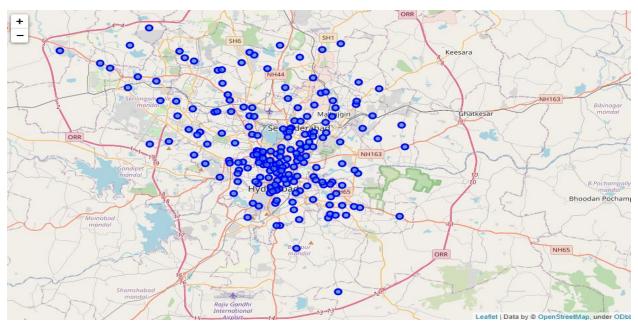
3. Venue Data

Next step was to fetch the details of the venues each of these neighborhoods using the Foursquare API. The contents of the dataset namely – Venue, Venue Latitude, Venue Longitude, Venue Category. A total of about 180 venues were extracted using the Foursquare API. A data frame was then created mapping all the venues against the neighborhoods in the format below.

	Neighborhood	Latitude	Longitude	VenueName	VenueLatitude	VenueLongitude	VenueCategory
0	A. C. Guards	17.395015	78.459812	Cafe Niloufer & Bakers	17.399715	78.462881	Café
1	A. C. Guards	17.395015	78.459812	Subhan Bakery	17.392412	78.464712	Bakery
2	A. C. Guards	17.395015	78.459812	Nizam club	17.403221	78.468729	Lounge
3	A. C. Guards	17.395015	78.459812	Prince Hotel	17.394736	78.442410	Indian Restaurant
4	A. C. Guards	17.395015	78.459812	Cream Stone Concepts	17.404284	78.481458	Ice Cream Shop

4. Data Visualization:

Python's package folium was then used to superimpose this data onto the map of Hyderabad



Methodology Used

The following is the step-by-step methodology that have been used to get the results:

1. <u>Web Scraping:</u> Perform this step using the information provided on the Wikipedia page and use several python requests and beautiful soup packages to retrieve this data in a Data Frame Format

	Neighborhood
0	A. C. Guards
1	A. S. Rao Nagar
2	Abhyudaya Nagar
3	Abids
4	Adibatla
5	Adikmet
6	Afzal Gunj
7	Aghapura
8	Aliabad, Hyderabad
9	Alijah Kotla

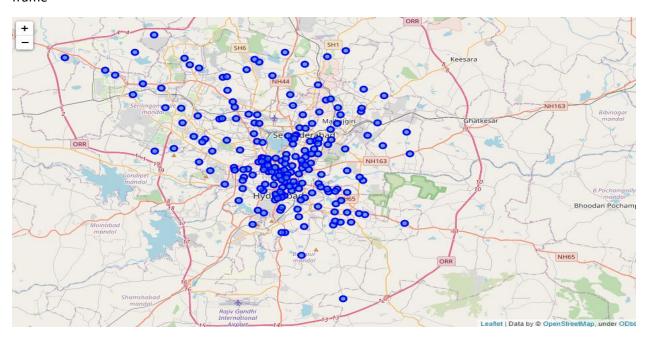
2. <u>Geolocations:</u> Use the Geocoder package to map the location details to latitude and longitude values to be able to fetch the venue details using Foursquare API

	Neighborhood	Latitude	Longitude
0	A. C. Guards	17.395015	78.459812
1	A. S. Rao Nagar	17.411200	78.508240
2	Abhyudaya Nagar	17.337650	78.564140
3	Abids	17.389800	78.476580
4	Adibatla	17.235790	78.541300
195	Secunderabad	17.442000	78.501920
196	Serilingampally	17.482160	78.323000
197	Shah-Ali-Banda	17.357390	78.473200
198	Shahran Market	17.364890	78.476290
199	Shanker Mutt	17.399817	78.507919

3. <u>Create Neighborhood Data Frame:</u> Populate all details (location, latitude and longitude) in a Pandas Data Frame

	Neighborhood	Latitude	Longitude	VenueName	VenueLatitude	VenueLongitude	VenueCategory
0	A. C. Guards	17.395015	78.459812	Cafe Niloufer & Bakers	17.399715	78.462881	Café
1	A. C. Guards	17.395015	78.459812	Subhan Bakery	17.392412	78.464712	Bakery
2	A. C. Guards	17.395015	78.459812	Nizam club	17.403221	78.468729	Lounge
3	A. C. Guards	17.395015	78.459812	Prince Hotel	17.394736	78.442410	Indian Restaurant
4	A. C. Guards	17.395015	78.459812	Cream Stone Concepts	17.404284	78.481458	Ice Cream Shop

4. <u>Visualize Maps:</u> Use the folium package to visualize the map of all the data sets in the data frame



- 5. <u>Venues:</u> Use the foursquare API to explore all the neighborhoods and get the venue details nearby the locations
- 6. <u>Create Venues Data Frame:</u> Merge the venues details against each of the neighborhoods create a larger D. In this DF, each venue is analyzed against the presence of venues and the most popular locations with more venues are chosen to be ideal

	Neighborhoods	ATM	Accessories Store	Afghan Restaurant	Airport	Airport Food Court	Airport Lounge	Airport Service	Airport Terminal	American Restaurant	 Sporting Goods Shop	Sports Bar	Stadium	Steakhouse
0	A. C. Guards	0	0	0	0	0	0	0	0	1	 0	0	0	1
1	A. S. Rao Nagar	0	0	0	0	0	0	0	0	1	 0	0	0	0
2	Abhyudaya Nagar	0	0	0	0	0	0	0	0	0	 1	0	0	0
3	Abids	0	0	0	0	0	0	0	0	1	 0	0	0	1
4	Adibatla	0	0	0	0	0	0	0	0	0	 0	0	0	0
195	Secunderabad	0	0	0	0	0	0	0	0	1	 0	1	0	0
196	Serilingampally	0	0	0	0	0	0	0	0	0	 0	0	1	0
197	Shah-Ali-Banda	0	0	0	0	0	0	0	0	1	 0	0	0	1
198	Shahran Market	0	0	0	0	0	0	0	0	1	 0	0	0	1
199	Shanker Mutt	0	0	0	0	0	0	0	0	1	 0	0	0	0
200 rows × 123 columns														

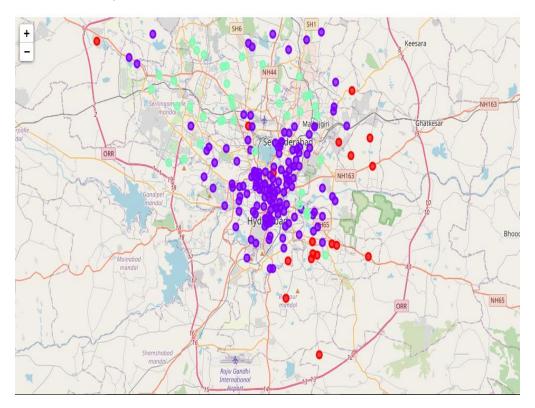
7. <u>Clustering:</u> Cluster the neighborhoods using K-means clustering technique and identify localities that already have higher number of restaurants to help answer the question as to which neighborhoods are most suitable to open new restaurants

Number of Clusters = 3

	Neighborhood	Restaurant	Cluster Labels	Latitude	Longitude
146	Nagaram, Medchal-Malkajgiri district	0	0	17.609930	78.491220
145	Nacharam	1	0	17.433510	78.566730
45	Chengicherla	0	0	17.437070	78.606840
56	Dundigal	0	0	17.593680	78.404020
137	Miyapur	0	0	17.421010	78.582460
116	Macha Bollaram	5	2	17.525910	78.376330
115	Lothkunta	4	2	17.494050	78.515140
173	Quthbullapur	5	2	17.505370	78.467490
175	Raidurg	5	2	17.424852	78.457255
96	Khajaguda	6	2	17.417620	78.365590

200 rows x 5 columns

8. <u>Visualize:</u> Use Python's Package folium to visualize the clusters by color coding 3 clusters created in the earlier step



Results

Attribute	Result
Number of neighborhood locations fetched	200
Total number of unique venue categories	122
Examples of venue categories	Café, Bakery, Lounge, Indian Restaurant, Ice Cream Shop, South Indian Restaurant, Hyderabadi Restaurant, Juice Bar, Scenic Lookout, Bistro etc.
Total number of localities with restaurants around:	187
Number of clusters:	2
Clustering based on	Existing presence of restaurants

Clustering Results:

Cluster	Restaurants	Number of Neighborhoods
0	Max 1	23
1	Max 3	139
2	Max 6	38

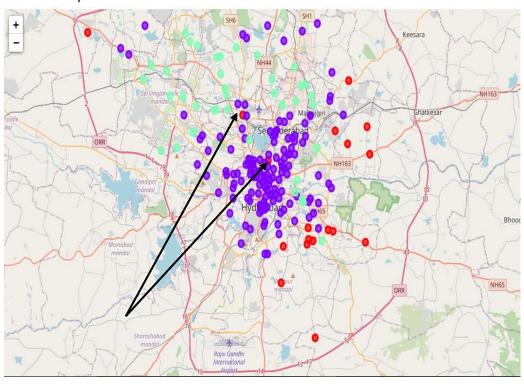
Discussion

As you may observe, all the data points belonging to the first cluster are on the borders of the city where they might not be any restaurants organically due to lower amount of population in that area.

Therefore, the first most lucrative areas to set up a restaurant would be the data points belonging to the cluster 0 that are well within the city limits.

In our results, we found two such data points:

- 1. Hayathnagar
- 2. Balapur

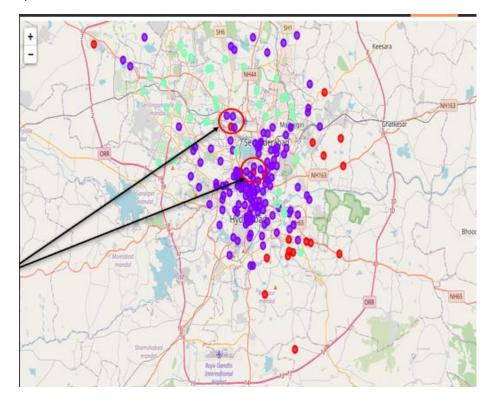


The second most lucrative list of possible areas for setting up a restaurant would be areas within the city limits nearby the data sets from the cluster 0 belonging to cluster 1. The reason behind this would be the population living in the neighborhoods of the first cluster might consider areas that might be far more popular near to them

Even though, the neighborhoods in cluster 2 have certain number of restaurants present, its always safer to set up shop in an area that already attracts a bit of crowd. Some of the areas in this bucket would be:

- 1. Domalguda
- 2. Hyderguda
- 3. Barkatpura

4. Ameerpet



The neighborhoods of cluster 2 should not be considered as there are already high number of restaurants present in the area.

Restaurants in these clusters are likely already suffering with immense amount of competition because of oversupply and high concentration of existing restaurants

Conclusion

The purpose of this project was to assess the neighborhoods in Hyderabad and create a Kmeans clustering model to suggest restauranters lucrative locations to set up a restaurant business. The neighborhoods data was extracted from multiple sources online including Wikipedia and Foursquare using its API. Through this project we were able to find that from the three clusters created, the cluster 2 had neighborhoods with the greatest number of restaurants in the area. Setting up a restaurant in an already saturated locality would create immense competition which the owner should definitely avoid. Therefore, this cluster was eliminated from the discussion.

We identified several localities in the cluster 0 and cluster 1 within the city limits which would be ideal for a restaurant setup. These areas had less competition and were well within the boundaries to ensure the presence of population density to ensure footfall.

Drawbacks

- 1. Major issue with this project is that Foursquare API returns very few number of venues
- 2. This project was a demonstration of the skills developed during the course of this certification. For setting up a restaurant there are several other factors that need to be considered like cost of rent, the surroundings around the shopping mall, the kind of people in the locality-if it's a luxurious area many people prefer going out, their lifestyle will be different from others and therefore spend a lot.
- 3. The demographic of the neighborhood should also be considered. That is, it the people living around the neighborhood prefer going out or ordering in etc.

Useful Links

Github Link

https://github.com/nehamadnekar/Capstone-Final-Project/blob/fd9426d795ce92ca024079361f32640bb5a1ca30/Capstone%20Final%20Project.ipy nb

NBViewer Link

https://nbviewer.jupyter.org/github/nehamadnekar/Capstone-Final-Project/blob/fd9426d795ce92ca024079361f32640bb5a1ca30/Capstone%20Final%20Project.ipy
nb

Jupyter Notebook Link

https://eu-gb.dataplatform.cloud.ibm.com/analytics/notebooks/v2/8fafab4c-1d6c-463b-8160-2441852973b2/view?access_token=36cb94890a6a6f1bd2f95c3fbadd7eb3c9b19d0f6488a6e5af 013eaa1c5d3316