Explore City – Hyderabad

Capstone Project – The Battle of Neighborhoods – Final Assignment Report

Mahesh sunke
13th June 2020

INTERNAL

1. INTRODUCTION:

Background:
The purpose of this Project is to help the people in exploring Hyderabad and its neighborhoods and It also helps the people to make smart and efficient decisions on chossing good and comfortable neighborhood out of numbers of other neighborhoods in Hyderabad.

Hyderabad City has a population of about 6.9 million, with about 9.7 million in Hyderabad Metropolitan Region, making it the fourth-most populous city and sixth-most populous urban agglomeration in India. With an output of US$74 billion, Hyderabad is the fifth-largest contributor to India's overall gross domestic product. The city has emerged as an Indian hub of pharmaceuticals and biotechnology. The formation of special economic zones and HITEC City dedicated to information technology has encouraged leading multinationals to set up operations in Hyderabad.

Problem description:
Hyderabad is most happening city in India. Somany people from all over the country and world comes to Hyderabad to explore opportunities. The people looks for better neighborhood in the city as per the facilites, localities, comforts and etc.

The people might look for:

- Better neighborhood
- Transportation availability(bus stops, railway stations, airport and etc.)
- Near by attractions
- Safety
- House Rent
- Cost of Living
- Malls, Super markets, Multiplexes and etc.

To address above, there should be a systematic solution which can help the people in exploring and choosing better neighberhood.

Target audience :

- Travellers/Visitors
- Employees
- Students
- The target for this project is basically for everyone who would like to explore Hyderabad.

2. DATA:

Data requirements :
To find a solution to the questions and build a suitable model, we need good amount of data. Data can answer so many things.

To build the systematic model for the requirement: we may required,

- Neighborhood details in the city
- Population
- Attractions
- Transport hubs details
- Infrastructure

- Companies/Organizations
- Average house rent/properties
- Localities near by
- Facilities
- Political/Regional commanded neighberhoods if any.
- Others.

Data collection :

We have various sources for collecting above data. Offcouse, some are open source and some may not. I have used:

- Open source data websites such as wikipedia, india gov website, census report websites, etc.
- Govt websites
- City Property websites to understand localities, facilities, house rent and etc.
- Social websites to understand any political/regional commanded neighborhoods
- Transportation department sites
- Others.

I couldn't able to find readily available data. So I have accessed above mentioned sources and gathered the data manually. Based on the collected information, prepared below table by using Pandas data frames. .

	Borough	Neighborhoods	Latitude	Longitude	Population	City	Avg_Income	Avg_House_Rent	Avg_Cost_of_Living
0	East Zone	Kapra	17.490340	78.572880	875944	Hyderabad	20366.98	6673.24	10457.42
1	East Zone	Uppal	17.393480	78.559647	752754	Hyderabad	27026.85	8855.34	13876.93
2	East Zone	Hayathnagar	17.329500	78.597640	483858	Hyderabad	52107.86	17073.14	26754.78
3	East Zone	LB Nagar Erstwhile 3B	17.443779	78.444643	537441	Hyderabad	48517.59	15896.79	24911.35
4	East Zone	Saroornagar	17.355581	78.535735	903197	Hyderabad	66025.22	21633.16	33900.65
5	South Zone	Malakpet	17.379778	78.499838	740568	Hyderabad	62869.10	20599.06	32280.14
6	South Zone	Santoshnagar	17.342310	78.504180	990929	Hyderabad	35523.90	11639.40	18239.74
7	South Zone	Chandrayangutta	17.325941	78.474493	672192	Hyderabad	74260.29	24331.38	38128.94
8	South Zone	Charminar Erstwhile Circle V A	17.361601	78.474675	406429	Hyderabad	72364.31	23710.16	37155.45
9	South Zone	Falaknuma Erstwhile Circle V B	17.333566	78.471895	87403	Hyderabad	69210.05	22676.67	35535.90
10	South Zone	Rajendra Nagar Erstwhile Circle VI	17.319120	78.391970	666293	Hyderabad	50892.87	16675.05	26130.94

Technology support:

- **FourSquare API:** I have used Four-square API as a prime data gathering source as it has a database of millions of places, especially their places API which provides the ability to perform location search, location sharing and details about a business and etc.
- After finding the list of neighborhoods, we then connected to the Foursquare API to gather information about venues inside each and every neighborhood.
- The data retrieved from Foursquare contained information of venues within a specified distance of the longitude and latitude of the neighborhoods.
- The information obtained per venue as follows:
    - Neighborhood
    - Neighborhood Latitude
    - Neighborhood Longitude
    - Venue
    - Name of the venue e.g. the name of a store or restaurant

- Venue Latitude
- Venue Longitude
- Venue Category

- Use of foursquare is focused to fetch nearest venue locations so that we can use them to form a cluster. Foursquare API leverages the power of finding nearest venues in a radius (in my case: 1500mts) and also corresponding coordinates, venue location and names. After calling, the following data frame is created:

	Borough	Neighborhoods	Neighborhood Latitude	Neighborhood Longitude	Venue	Venue Latitude	Venue Longitude	Venue Category
0	East Zone	Kapra	17.49034	78.572880	Apollo Pharmacy	17.490776	78.573792	Pharmacy
1	East Zone	Kapra	17.49034	78.572880	Big Bull Bakery	17.492283	78.565074	Bakery
2	East Zone	Kapra	17.49034	78.572880	Blue parks bar	17.483349	78.562201	Bar
3	East Zone	Kapra	17.49034	78.572880	Vijaya Plywood	17.477331	78.574005	Furniture / Home Store
4	East Zone	Kapra	17.49034	78.572880	Thalluri Theatres	17.477694	78.572620	Multiplex
5	East Zone	Uppal	17.39348	78.559647	Metro	17.391966	78.559706	Convenience Store
6	East Zone	Uppal	17.39348	78.559647	Nagole HMR Station	17.391336	78.558846	Light Rail Station
7	East Zone	Uppal	17.39348	78.559647	HMR Uppal Depot	17.390554	78.555524	Train Station
8	East Zone	Uppal	17.39348	78.559647	Uppal Bus Stand	17.401500	78.568899	Bus Station
9	East Zone	Uppal	17.39348	78.559647	Vegetable Market	17.401691	78.569595	Flea Market

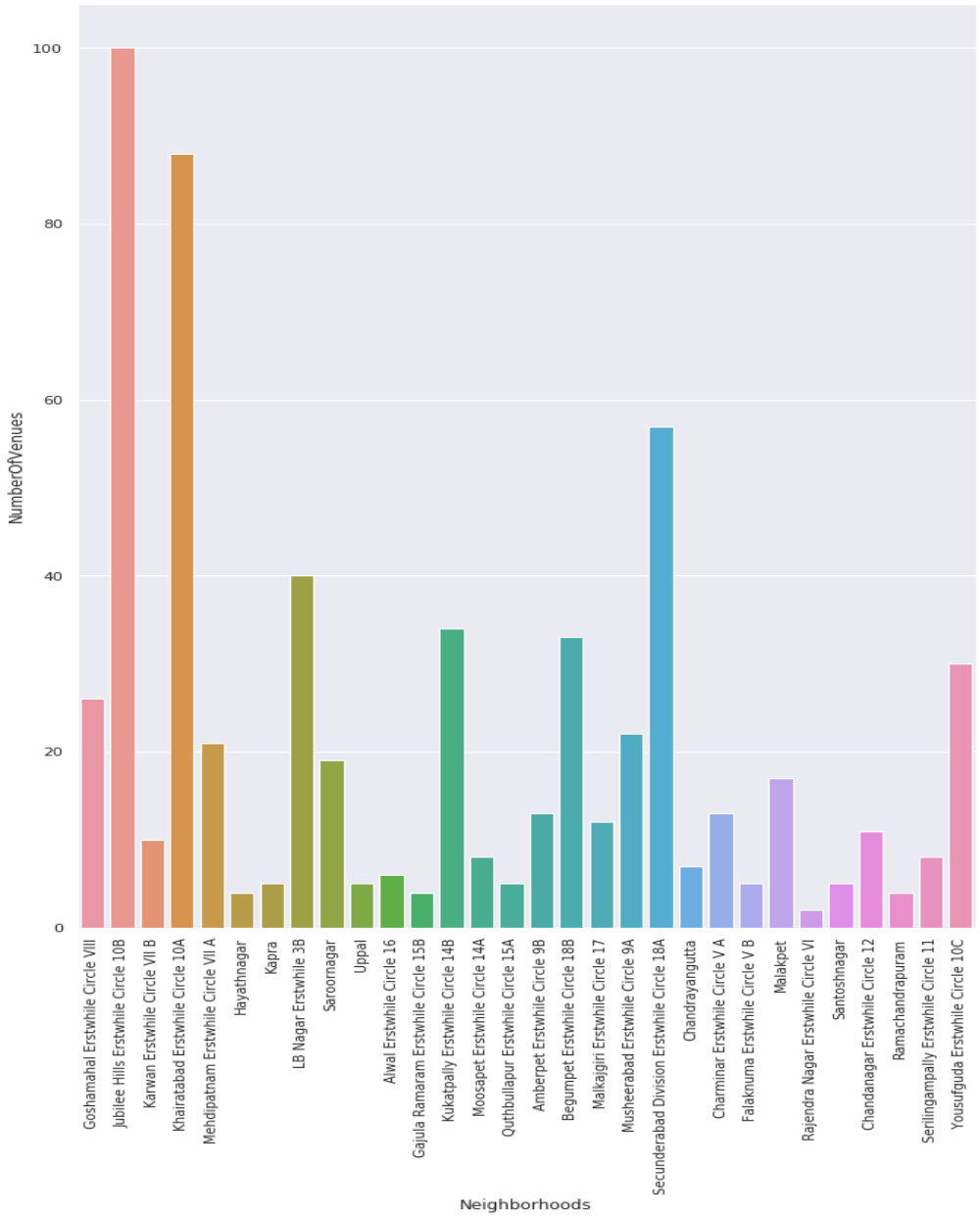
3. METHODOLOGY

Clustering Approach:

Scrapping the data from different sources and then combining it to form a single-ton dataset is a difficult task. To do so, we need to explore the current state of dataset and then list up all the features needed to be fetched. Once the data was ready, I have used K-means clustering algorithm to analyse clusters. Please find the most common venues that are near by Neighborhoods below

	Neighborhoods	1st Most Common Venue	2nd Most Common Venue	3rd Most Common Venue	4th Most Common Venue	5th Most Common Venue	6th Most Common Venue	7th Most Common Venue	8th Most Common Venue	9th Most Common Venue	10th Most Common Venue
0	Alwal Erstwhile Circle 16	Department Store	Indian Restaurant	Breakfast Spot	Bus Station	Grocery Store	Gym	Gastropub	Food & Drink Shop	Diner	Discount Store
1	Amberpet Erstwhile Circle 9B	Indian Restaurant	Food & Drink Shop	Café	Flea Market	Movie Theater	Food Court	Pizza Place	Sandwich Place	Convenience Store	Train Station
2	Begumpet Erstwhile Circle 18B	Indian Restaurant	Hotel	Clothing Store	Vegetarian / Vegan Restaurant	Light Rail Station	Cocktail Bar	Chinese Restaurant	Pizza Place	Pub	Café
3	Chandanagar Erstwhile Circle 12	Convenience Store	Train Station	Accessories Store	Pizza Place	Stadium	Multiplex	Diner	South Indian Restaurant	Bakery	Department Store
4	Chandrayangutta	Middle Eastern Restaurant	Italian Restaurant	Train Station	Indian Restaurant	Breakfast Spot	Resort	Food & Drink Shop	Diner	Discount Store	Donut Shop
5	Charminar Erstwhile Circle V A	History Museum	Bus Station	Indian Restaurant	Monument / Landmark	Coffee Shop	Food Court	Café	Juice Bar	Dessert Shop	Diner
6	Falaknuma Erstwhile Circle V B	Train Station	Resort	Italian Restaurant	Breakfast Spot	Indian Restaurant	Women's Store	Farmers Market	Flea Market	Fast Food Restaurant	Falafel Restaurant
7	Gajula Ramaram Erstwhile Circle 15B	ATM	Pizza Place	Café	Food & Drink Shop	Dessert Shop	Diner	Discount Store	Donut Shop	Electronics Store	Falafel Restaurant
8	Goshamahal Erstwhile Circle VIII	Indian Restaurant	Bakery	Hotel	Bus Station	Shoe Store	Department Store	Neighborhood	Farmers Market	Flea Market	Food
9	Hayathnagar	ATM	Pharmacy	Mattress Store	Department Store	Dessert Shop	Diner	Discount Store	Donut Shop	Electronics Store	Falafel Restaurant
10	Jubilee Hills Erstwhile Circle 10B	Ice Cream Shop	Indian Restaurant	Café	Coffee Shop	Brewery	Hookah Bar	Lounge	Dessert Shop	Bakery	Cocktail Bar

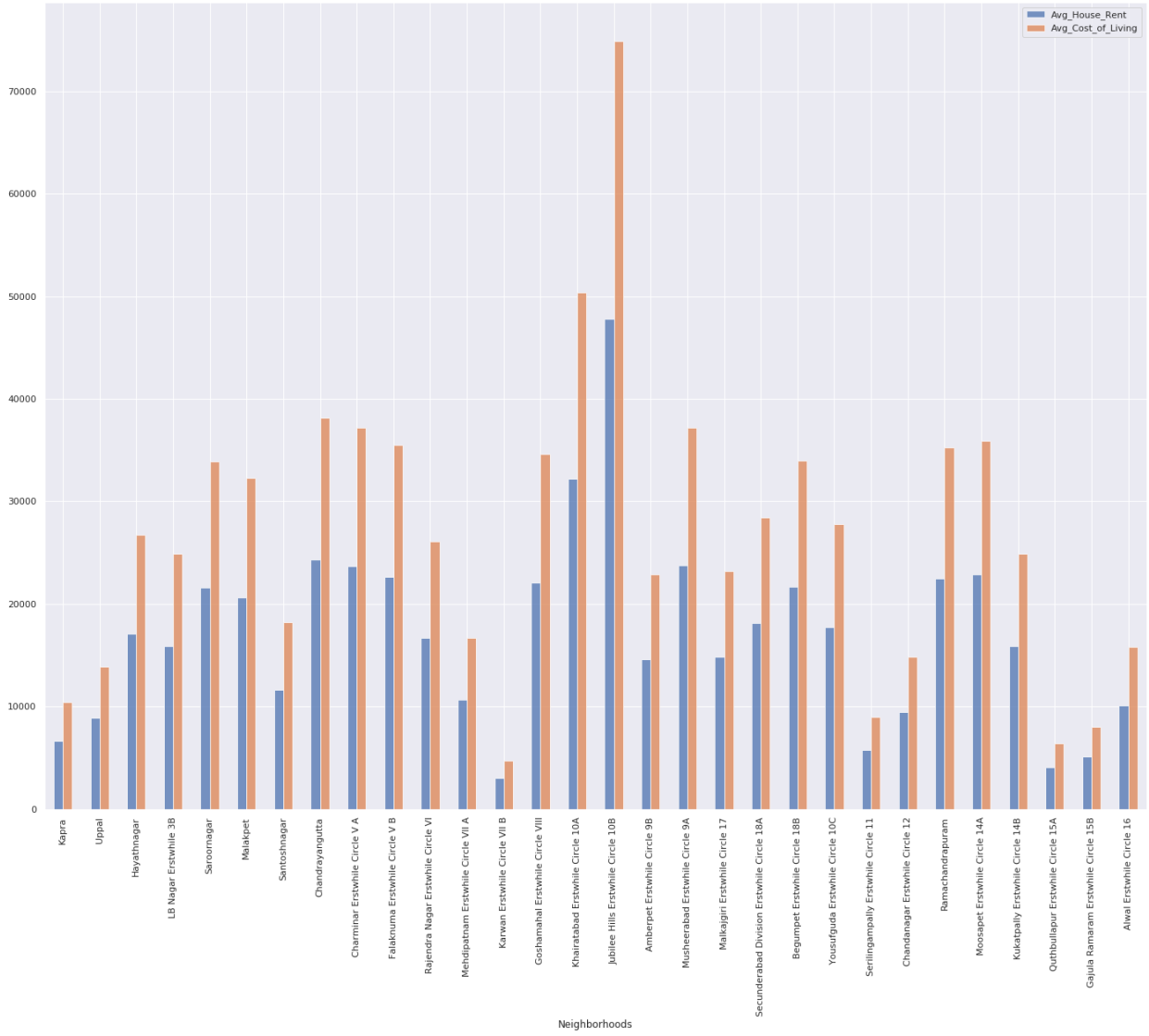
Exploring the dataset is important because it gives you initial insights and may help you to get partial idea of the answers that you are looking to find out from the data.



4. RESULTS SECTION:

- I further deeped into analysis to understand the Average House Rent and Cost of Living of the Neighborhoods.
- While exploring the dataset, Its observed that **Jubilee Hills** most number of venues and **Rajendra Nagar** has less.
- Also, Jubilee Hills has more house rents and cost of living.
- At the same time it has more venues.
- Average Hour Rent and Cost of Living of Neighborhoods can be found below.

Neighborhoods	Borough	Avg_House_Rent	Avg_Cost_of_Living
Kapra	East Zone	6673.24	10457.42
Uppal	East Zone	8855.34	13876.93
Hayathnagar	East Zone	17073.14	26754.78
LB Nagar Erstwhile 3B	East Zone	15896.79	24911.35
Saroornagar	East Zone	21633.16	33900.65
Malakpet	South Zone	20599.06	32280.14
Santoshnagar	South Zone	11639.40	18239.74
Chandrayangutta	South Zone	24331.38	38128.94
Charminar Erstwhile Circle V A	South Zone	23710.16	37155.45
Falaknuma Erstwhile Circle V B	South Zone	22676.67	35535.90
Rajendra Nagar Erstwhile Circle VI	South Zone	16675.05	26130.94
Mehdipatnam Erstwhile Circle VII A	Central Zone	10675.43	16729.14
Karwan Erstwhile Circle VII B	Central Zone	3017.32	4728.35
Goshamahal Erstwhile Circle VIII	Central Zone	22063.88	34575.62
Khairatabad Erstwhile Circle 10A	Central Zone	32164.52	50404.01
Jubilee Hills Erstwhile Circle 10B	Central Zone	47793.46	74895.63



5. DISCUSSION SECTION:

- Here, we are creating systematic model to explore city to understand which neighborhood has more number of venues, facilities, comforts and etc. The information has been explored and added above.
- We have also explored Average House Rents and Cost of Living of neighborhoods.
- These information may help people to take smart decisions on choosing comfort and convienient neighborhoods.

6. CONCLUSION:

- We explored Near by venues, Top venues and Common venues for Hyderabad Neighborhoods:
    - it gives good information about the nearest shops, restaurants, gyms, Malls, Multiplex theaters, Schools and etc.
- Also executed Avg house rent and Avg cost of living for the Neighborhoods.
- Based on all above, it gives good information to the people who would like to stay and explore Hyderabad.
- If we have good data of any other cities, then the model can be used to explore other cities also.

References:

Libraries & Packages which are used to explore above.

- Pandas: For creating and manipulating dataframes.
- Folium: Python visualization library would be used to visualize the neighborhoods cluster distribution of using interactive leaflet map.
- Scikit Learn: For importing k-means clustering.
- JSON: Library to handle JSON files.
- XML: To separate data from presentation and XML stores data in plain text format.
- Geocoder: To retrieve Location Data.
- Beautiful Soup and Requests: To scrap and library to handle http requests.
- Matplotlib: Python Plotting Module.

It is just begining and there's more to explore. . Happy Learning.

Thank you !