Explore City – Hyderabad

Mahesh Sunke

13th June 2020

Capstone Project – The Battle of Neighborhoods – Final Assignment Presentation

PUBLIC

Introduction:

- 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 choosing good and comfortable neighborhood out of numbers of other neighborhoods in Hyderabad.
- Hyderabad is most happening city in India. So many 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 transportation availability, near by attractions, safety, house rent, cost of living, malls, restaurants, super markets and etc.
- This model will be a systematic solution which can help the people in exploring and choosing better neighborhood.

Data requirements :

To build the systematic model for the requirement: we require

- Neighborhood details in the city
- Population
- Attractions
- Transport hubs details
- Infrastructure
- Companies/Organizations
- Average house rent/properties
- Localities near by
- Facilities
- Political/Regional commanded neighborhoods if any. and etc.

Data collection :

We have various sources for collecting above data. I have used below. .

- Open source data websites such as:
  - Wikipedia, census report websites, etc.
  - India 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
- 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. .

Hyderabad Data Set:

	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

Foursquare API:

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. Through Foursquare API, I could able to create following data frame:

	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

Methodology:

Clustering Approach:

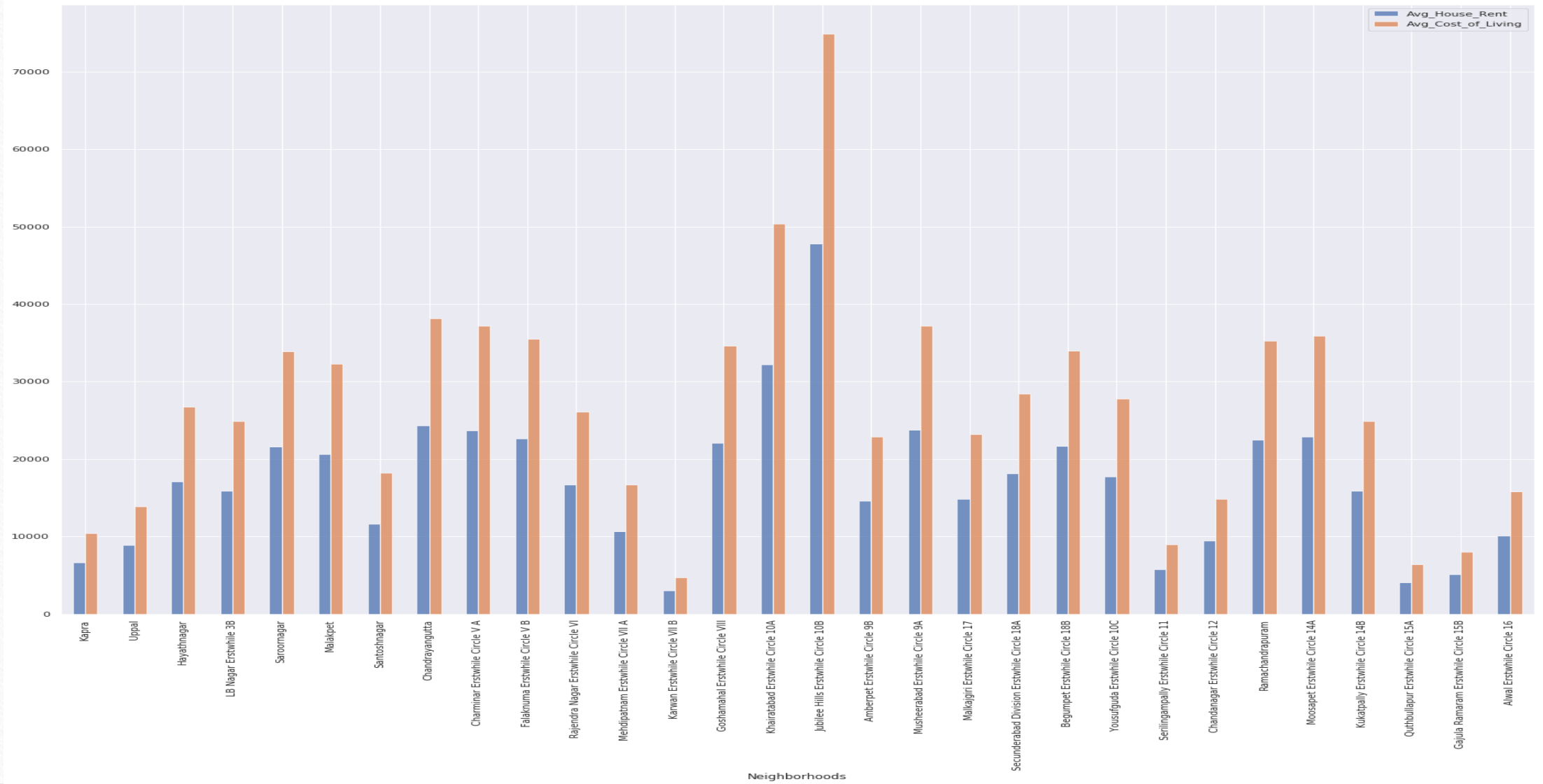
Scrapping the data from different sources and then combining it to form a single-ton dataset is a difficult task. To do so, we need to explore the current state of dataset and then list up all the features needed to be fetched. Once the data was ready, I have used K-means clustering algorithm to analyze 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

Results:

- I further deeper 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


Avg_House_Rent
Avg_Cost_of_Living
70000
60000
50000
40000
30000
20000
10000
0
Kapra
Uppal
Hayathnagar
LB Nagar Erstwhile 3B
Saroornagar
Malakpet
Santoshnagar
Chandrayangutta
Charminar Erstwhile Circle V A
Falaknuma Erstwhile Circle V B
Rajendra Nagar Erstwhile Circle VI
Mehdipatnam Erstwhile Circle VII A
Karwan Erstwhile Circle VII B
Goshamahal Erstwhile Circle VIII
Khairatabad Erstwhile Circle 10A
Jubilee Hills Erstwhile Circle 10B
Amberpet Erstwhile Circle 9B
Musheerabad Erstwhile Circle 9A
Malkajgiri Erstwhile Circle 17
Secunderabad Division Erstwhile Circle 18A
Begumpet Erstwhile Circle 18B
Yousufguda Erstwhile Circle 10C
Serilingampally Erstwhile Circle 11
Chandanagar Erstwhile Circle 12
Ramachandrapuram
Moosapet Erstwhile Circle 14A
Kukatpally Erstwhile Circle 14B
Quthbullapur Erstwhile Circle 15A
Gajula Ramaram Erstwhile Circle 15B
Alwal Erstwhile Circle 16
Neighborhoods

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 data frames.
- 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 beginning and there's more to explore. . Happy Learning.

Thank you !

Mahesh Sunke