The Battle of Neighborhoods

(Predicting the best residential area in Hyderabad)

Sai Naga Poojitha Katta

May 8,2020

1. Introduction

Background

Hyderabad, besides being called the Pearl city of India and also one of the biggest metropolis. Hyderabad is among the global centres of information technology for which it is known as Cyberabad (Cyber City).

The Hyderabad Information Technology and Engineering Consultancy City, abbreviated as HITEC City, is an Indian information technology, engineering, health informatics, and bioinformatics, financial business district located in Hyderabad, Telangana state.

HITEC City is spread across 81 hectares (200 acres) of land under suburbs of Madhapur, Gachibowli, Kondapur, Manikonda and Nanakramguda, all the combined technology townships is also known as Cyberabad with a radius of 56.48 km(35.09 mi) surrounding approximate area of 15000 acres. HITEC City is within 2 kilometres (1.2 mi) of the residential and commercial suburb of Jubilee Hills.

Problem

An Employee from Delhi got placed in Google.Inc which is located in Kondapur Village(part if HITEC city) in Hyderabad. Now he needs to find a perfect neighbourhood in Hyderabad which fulfills all his requirement.

The best residential area according to the Employee has meet the below criteria:

- The Area should be near to his office, so that his travel time can be reduced.
- The House price should be feasible with his budget.
- The venues near to his residential area should have all the facilities like Gym, Deparmental Stores etc.

The Project aims to suggest the best residential area to the Employee, which meets all

his requirements.

Interest

This wil be useful for individual seeking for a best residential area which fits all his requirements, while moving to a new city and also for real estate agents to suggest the best place for their customers.

2. Data Sources

Data downloaded or scraped from multiple sources were combined into one table, which consists of 89 Neighbourhoods which were suitable for residential purpose

- I got the list of Neighbourhoods from the indiaserver.com which has list of all the Neighbourhoods of Hyderabad.
- I scraped the House pricing details from various real estate site like 99acres.com and maakan.com
- I used Forsquare API to get the most common venues of given Neighbourhood In Hyderabad.

3. Exploratory Data Analysis.

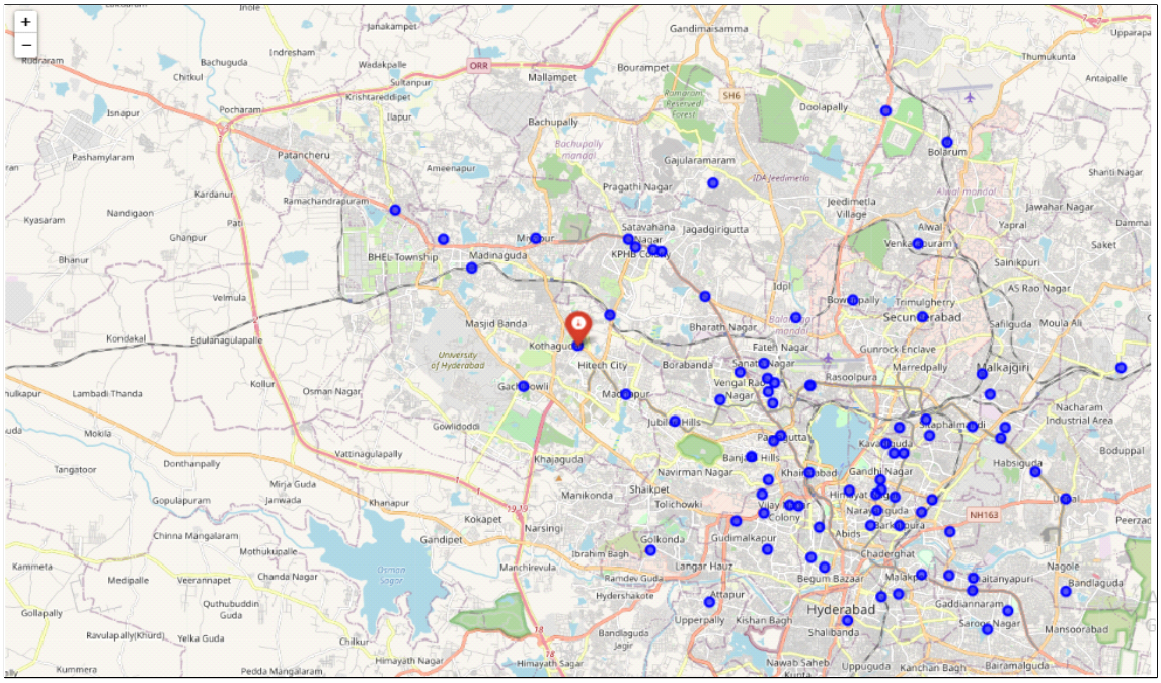
Data was downloaded and scraped from multiple resources and made into a form a .csv file which is available in my Github repository. The .csv file contains the columns *Neighbourhood nam*e and *Avg Price per sqft*.

	Neighbourhood	Avg Price per sqft
0	AP SE Board Hyderabad	4445.0
1	Ac Guards Hyderabad	6492.0
2	Ahmed Nagar Hyderabad	2343.0
3	Amberpet Hyderabad	2222.0
4	Ameerpet Hyderabad	9628.0

We get the latitude and longitude of each neighbourhod and the distance of each neighbourhood from the desired location using python *geopy* library.

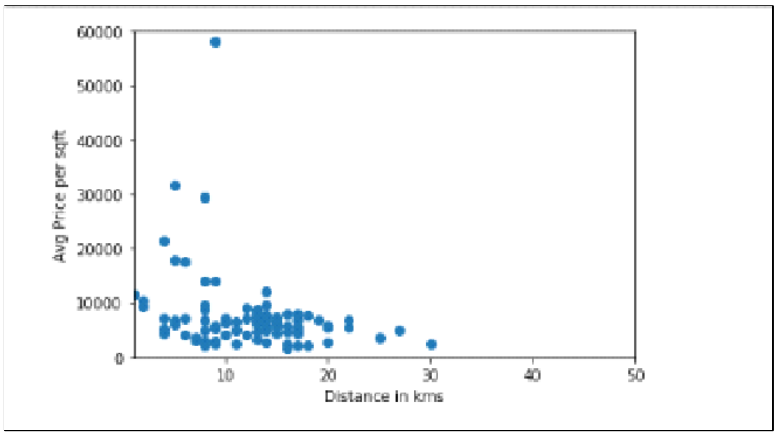
	Neighborhood	Avg Price per sqft	latitude	longitude	Distance in kms
0	AP SE Board Hyderabad	4445.0	17.493930	78.402257	4
1	Ac Guards Hyderabad	6492.0	17.399877	78.458051	11
2	Ahmed Nagar Hyderabad	2343.0	17.404066	78.444229	9
3	Amberpet Hyderabad	2222.0	17.390263	78.516481	17
4	Ameerpet Hyderabad	9628.0	17.437501	78.448251	8

I used python *folium* library to visualize geographic details of Google.Inc office in Hyderabad and all selected Neighbourhoods. I created a Map of Hyderabad in office is visualised with a red pin and all the blue colored circle markers represents the neighbourhoods.



Relation between price and distance

Plot a graph between distance and Avg price per sqft, check if there is any correlation between them.



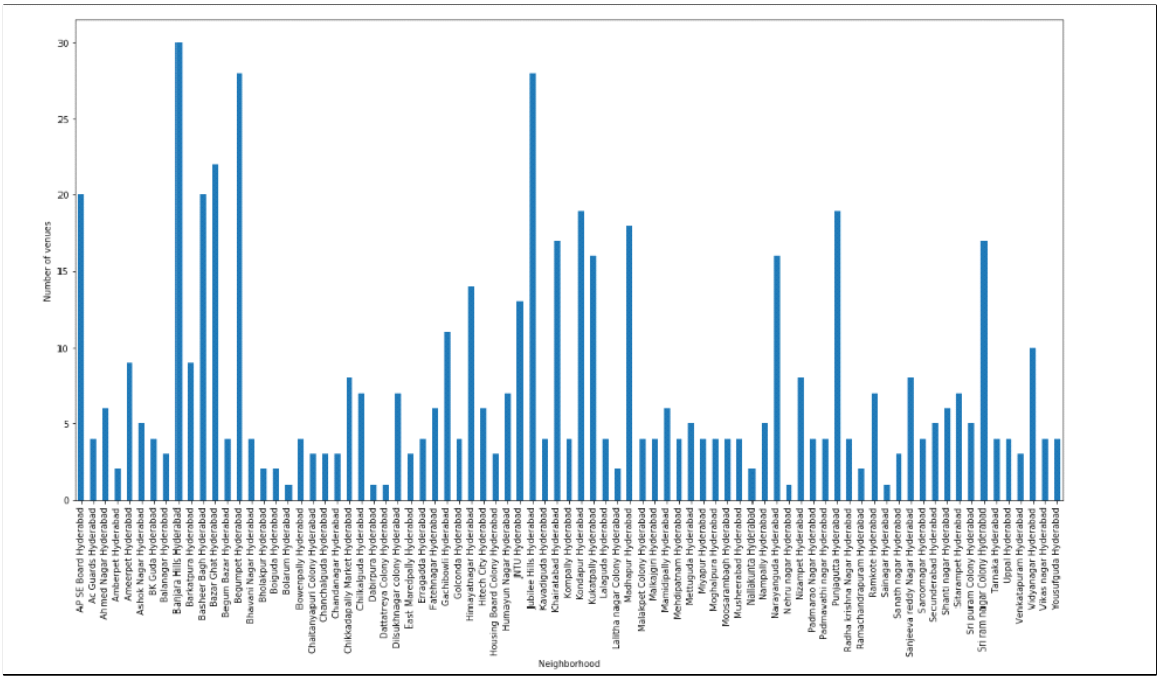
From the above figure, I can clearly see that there is no correlation between the distance and price.

Now,I used the Foursquare API to explore the venues of each Neighbourhood and segment them. I designed the limit as 100 venues and the radius 500 meter for each Neighbourhood from their given latitude and longitude informations. Here is table of the

list Venues name, category, latitude and longitude corresponding to their Neighbourhood.

	Neighborhood	Neighborhood Latitude	Neighborhood Longitude	Venue	Venue Latitude	Venue Longitude	Venue Category
0	AP SE Board Hyderabad	17.46383	73.42237	Paradise Restaurant	17.42884	78.484155	Indian Restaurant
1	AP SE Board Hyderabad	17.46383	73.42237	KIS Bakery	17.434710	78.426117	Bakery
2	AP SE Board Hyderabad	17.46383	73.42237	Chalanya Fine Dining	17.432000	78.482282	Food
3	AP SE Board Hyderabad	17.46383	73.42237	Cream Stone Corners	17.432500	78.484327	Ice Cream Shop
4	AP SE Board Hyderabad	17.46383	73.42237	Ohm Palace Hotel	17.435000	78.488167	Hotel Bar

Total, 591 venues were returned by Foursqaure API for all Neighbourhoods. We can see number of venues returned by Foursquare API for each neighbourhood from the below graph.



We see that Banjara hills, Jubilee hills and Begumpet are having the highest number of venues and Bolarum,Nizampet and Sainagar are having least number of venues. This doesn't mean that the neighbourhood isnt good enough, we can get more detailed information of each Neighbourhood by checking if the retrieved venues are the desired venues by the Employee.

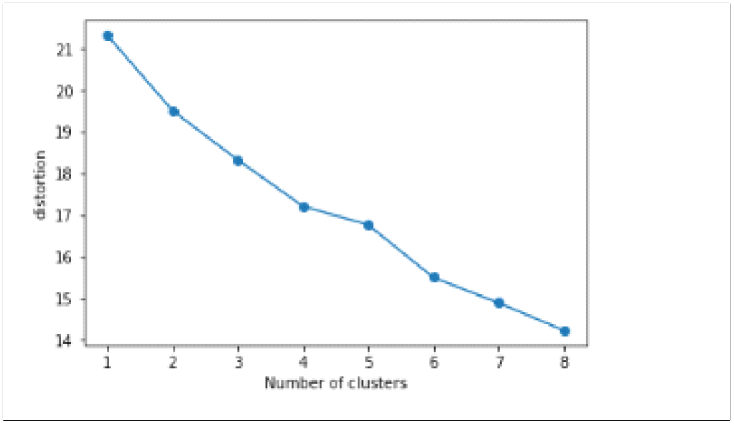
In summary of this graph 591 unique categories were returned by Foursquare, then I created a table which shows list of top 10 venue category for each borough in below table.

	Neighborhood	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	AP SE Board Hyderabad	Indian Restaurant	Fast Food Restaurant	Bakery	Clothing Store	Ice Cream Shop	Middle Eastern Restaurant	Metro Station	Pizza Place	Sandwich Place	Breakfast Spot
1	Ac Guards Hyderabad	Hyderabadi Restaurant	Grocery Store	Hotel	Bakery	Fish Market	Cosmetics Shop	Cupcake Shop	Deli / Bodega	Department Store	Dessert Shop
2	Ahmed Nagar Hyderabad	Café	Grocery Store	Vegetarian / Vegan Restaurant	Snack Place	Fish & Chips Shop	Convenience Store	Cosmetics Shop	Cupcake Shop	Deli / Bodega	Department Store
3	Amberpet Hyderabad	Movie Theater	Indian Restaurant	Women's Store	Fish & Chips Shop	Convenience Store	Cosmetics Shop	Cupcake Shop	Deli / Bodega	Department Store	Dessert Shop
4	Ameerpet Hyderabad	Indian Restaurant	Mobile Phone Shop	Electronics Store	Coffee Shop	Sandwich Place	Diner	Department Store	Fast Food Restaurant	Candy Store	Vegetarian / Vegan Restaurant
5	Ashok Nagar Hyderabad	Park	Dessert Shop	Restaurant	Brewery	Women's Store	Coffee Shop	Convenience Store	Cosmetics Shop	Cupcake Shop	Deli / Bodega
6	BK Guda Hyderabad	Indian Restaurant	Café	Bakery	Women's Store	Fish Market	Cosmetics Shop	Cupcake Shop	Deli / Bodega	Department Store	Dessert Shop
7	Balanagar Hyderabad	Fish Market	Market	Fish & Chips Shop	Convenience Store	Cosmetics Shop	Cupcake Shop	Deli / Bodega	Department Store	Dessert Shop	Diner
8	Banjara Hills Hyderabad	Coffee Shop	Indian Restaurant	Café	Furniture / Home Store	Deli / Bodega	Sandwich Place	Bakery	Hookah Bar	Ice Cream Shop	Farmers Market
9	Barkatpura Hyderabad	Coffee Shop	Gym	Bank	Motel	Café	Platform	Women's Store	Train Station	Frozen Yogurt Shop	Dumpling Restaurant

4. Modeling

We have some common venue categories in Neighbourhoods. So, I used unsupervised learning K-means algorithm to cluster the Neighbourhoods. K-Means algorithm is one of the most common cluster method of unsupervised learning, therefore I preferred to use this.

First, I will run K-Means to cluster the boroughs into 4 clusters because when I analyze the K-Means with elbow method it ensured me the 4 degree for optimum k of the K-Means.

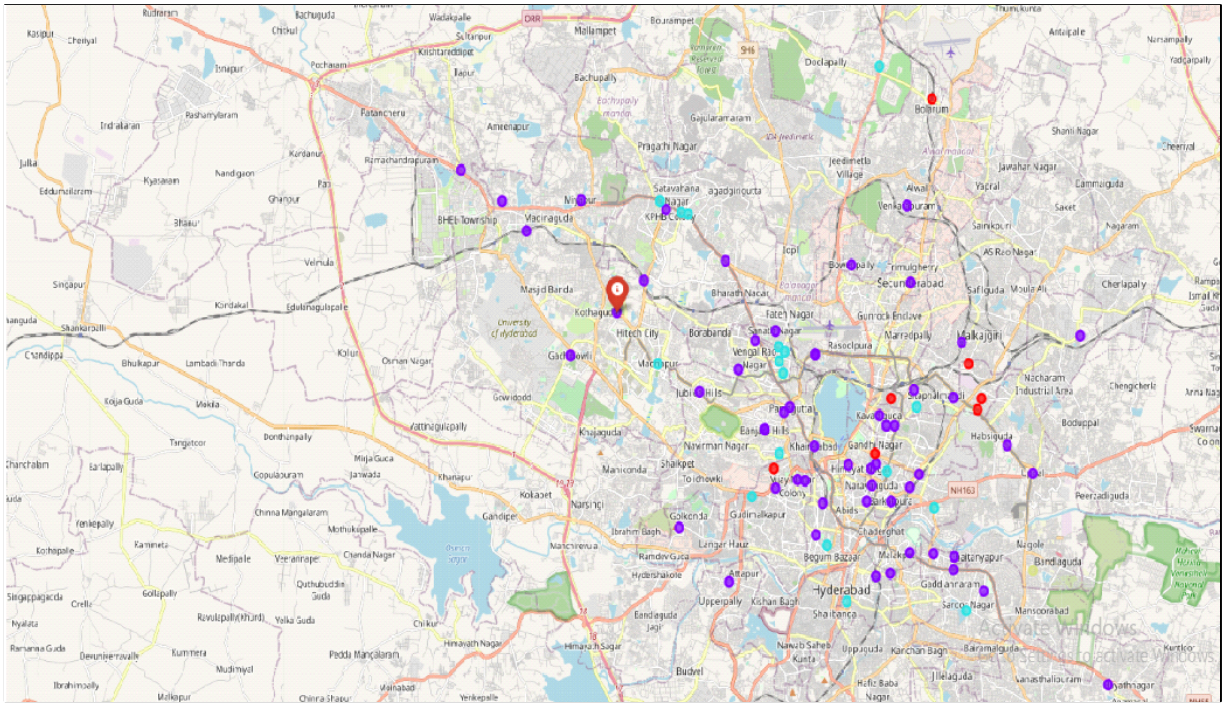


Although, the graph show almost linear line, I took K=4 as optimal value, since from K=4, I see that the value is getting normalised.

Then, I added the Cluster Labels to the table and merged the table with distance and Avg price per sqft details corresponding to each Neighbourhood, so that it would be easy to analyse. The below figure shows the merged table.

Neighborhood	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	Cluster Labels	Avg Price per sqft	Distance in kms
AB St Hasan Hyderabad	Indian Restaurant	Fast Food Restaurant	Bakery	Ice Cream Shop	Snack Place	Breakfast Spot	Pizza Place	Middle Eastern Restaurant	Sandwich Place	Shopping Mall	2	4445.0	4
Ab Gaurav Hyderabad	Hyderabad Restaurant	Hotel	Bakery	Vegetarian / Vegan Restaurant	Farmers Market	Comedy Club	Concert Hall	Convenience Store	Cupcake Shop	Deli / Bodega	1	6492.0	11
Ahmed Nagar Hyderabad	Snack Place	Café	Vegetarian / Vegan Restaurant	Grocery Store	Furniture / Home Store	Dumpling Restaurant	Coffee Shop	Garden	Comedy Club	Concert Hall	0	2543.0	9
Amberpet Hyderabad	Movie Theater	Indian Restaurant	Vegetarian / Vegan Restaurant	Electronics Store	Coffee Shop	Comedy Club	Concert Hall	Convenience Store	Cupcake Shop	Deli / Bodega	2	2222.0	17
Ameerpet Hyderabad	Indian Restaurant	Vegetarian / Vegan Restaurant	Supermarket	Diner	Electronics Store	Hotel	Sandwich Place	Food Truck	Food Court	Coffee Shop	2	9620.0	0
Ashok Nagar Hyderabad	Restaurant	Concert Hall	Brewery	Dessert Shop	Fast Food Restaurant	Vegetarian / Vegan Restaurant	Coffee Shop	Comedy Club	Convenience Store	Cupcake Shop	1	6106.0	13
BK Guda Hyderabad	Indian Restaurant	Café	Bakery	Vegetarian / Vegan Restaurant	Farmers Market	Comedy Club	Concert Hall	Convenience Store	Cupcake Shop	Deli / Bodega	2	3034.0	7

Again, I used Python *folium* library, to visualise each cluster in Map.



5. Discussion and Conclusion:

Now, I shall examine each cluster to find which cluster best fits Employee requirements.

As we know that Employee requirement, we need to check which cluster has gym and Departmental stores nearby.

I see that Cluster 2 has Gym in the nearest venues, so I search for Neighbourhoods in Cluster 2.

We see that Venkatapuram has Gym as most common venue and also has Park,Departmental Store,Farmers market, but the distance fro the location to office is 14km and has Avg price per sqft as 9628 Indian rupee. Where as, Kondapur in the same cluster has Gym, Departmental store and is within the Neighbourhood where the office is

located and has Avg price per sqft as 9342 Indian rupee, which is less than Venkatapuram.

So, we can conclude that Kondapur would be best Neighbourhood for the Employee to reside.

Again! It completely depends on the employee choice, if he finds the price is high and he can opt for a farther place of less hosue price, like Barkatpura which has Gym, bank and railway station nearby with distance from office is 15km and Avg price per sqft as 6406 Indian rupee.