The Battle of Neighborhoods - Finding a better Neighborhood for constructing new venues in Hyderabad.

1. Introduction

This Project is about Exploring Venues in Neighbourhoods in Hyderabad. This can help people to explore best facilities in their Neighbourhood.

People can decide on their own that, which Neighbourhood is best for Restaurants, Motel, Bakery, Cafe... etc.

This Project also helps People who are migrating to new Places can decide which Neighbourhood is best to live and which Neighbourhood have best facilities for their use.

People can also check Schools, Hospitals... etc, have in their Neighbourhood. According to facilities in Neighbourhood they can know how much a house prices ranges. They can decide best Neighbourhood..

People who are trying set up any Restaurants, Motels, Bakery, Cafe.. etc. They can decide in which Neighbourhood is best for such venues and where Peoples goes more i.e, in which Neighbourhood.

2. Data Section

Used Datalink in Project

Foursquare API Data

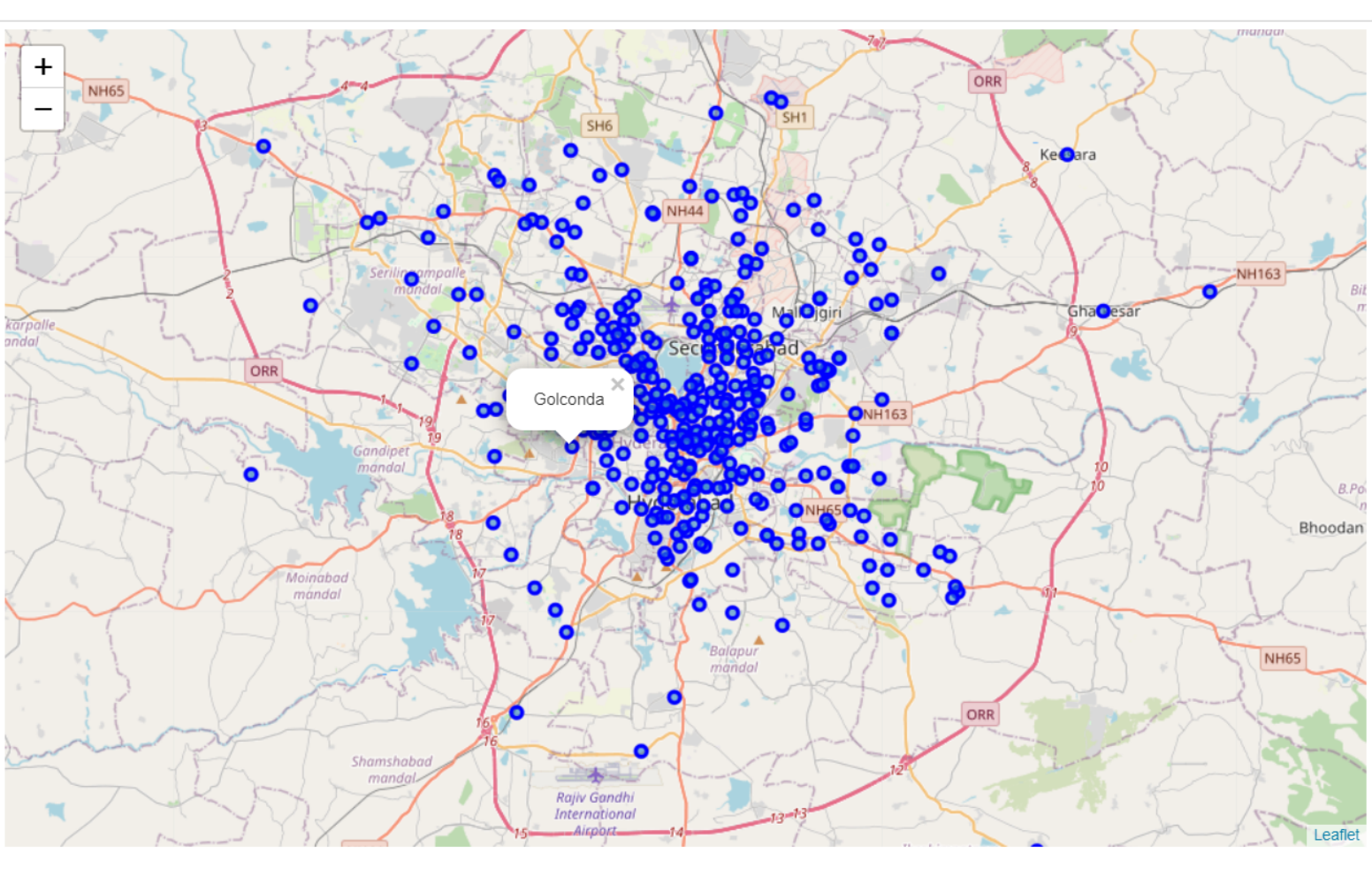
We will need data about different venues in different neighborhoods of that specific borough. In order to gain that information we will use "Foursquare" locational information. Foursquare is a location data provider with information about all manner of venues and events within an area of interest. Such information includes venue names, locations, menus and even photos. As such, the foursquare location platform will be used as the sole data source since all the stated required information can be obtained through the API.

After finding the list of neighborhoods, we then connect to the Foursquare API to gather information about venues inside each and every neighborhood. For each neighborhood, we have chosen the radius to be 500 meter.

The data retrieved from Foursquare contained information of venues within a specified distance of the longitude and latitude of the postcodes. 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

Map of Hyderabad



3. Methodology Section

Libraries used for this Project

- Pandas: For Data manipulation and Analysis. It is used to manipulat data in tables.
- Nominatim: To find locations by giving Neighbourhood names.
- Folium: To display map and cluster in Neigbourhood.
- KMeans: To cluster similar Venues Categories.
- Matplolib: To Visualize the plot.

Venues with location in Neighborhood of Hyderabad

```
Hyderabad_venues.head()
```

	Neighborhood	Neighborhood Latitude	Neighborhood Longitude	Venue	Venue Latitude	Venue Longitude	Venue Category
0	Gachibowli	17.443622	78.351964	Karachi Bakery	17.442930	78.355336	Bakery
1	Gachibowli	17.443622	78.351964	Gachibowli Stadium	17.445982	78.348245	Stadium
2	Gachibowli	17.443622	78.351964	creamstone	17.442998	78.355475	Ice Cream Shop
3	Gachibowli	17.443622	78.351964	Chettinaduvilas	17.442858	78.356053	Food Truck
4	Gachibowli	17.443622	78.351964	Mustang terrace	17.442840	78.355320	Italian Restaurant

Groupby Neighborhood and mean of Venues

```
Hyderabad_grouped = Hyderabad_onehot.groupby('Neighborhood').mean().reset_index()
Hyderabad_grouped
```

	Neighborhood	ATM	Accessories Store	Afghan Restaurant	Airport Terminal	American Restaurant	Andhra Restaurant	Antique Shop	Arcade	Arts & Crafts Store	...	Tea Room	Tennis Court	Tex-Mex Restaurant	Theater	Theme Par
0	A C Guards	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.000000	...	0.0	0.0	0.0	0.0	0.
1	AG Office	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.000000	...	0.0	0.0	0.0	0.0	0.
2	Abids	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.000000	...	0.0	0.0	0.0	0.0	0.
3	Abids Road	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.000000	...	0.0	0.0	0.0	0.0	0.
4	Adarsh Nagar	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.142857	...	0.0	0.0	0.0	0.0	0.
...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...
332	West Marredpally	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.000000	...	0.0	0.0	0.0	0.0	0.
333	Yapral	1.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.000000	...	0.0	0.0	0.0	0.0	0.
334	Yellareddyguda	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.000000	...	0.0	0.0	0.0	0.0	0.
335	Yousufguda	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.000000	...	0.0	0.0	0.0	0.0	0.
336	Zamistanpur	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.000000	...	0.0	0.0	0.0	0.0	0.

337 rows × 170 columns

Clustering

To compare the similarities of two cities, we decided to explore neighborhoods, segment them, and group them into clusters to find best neighborhoods in Hyderabad. To be able to do that, we need to cluster data which is a form of unsupervised machine learning: k-means clustering algorithm.

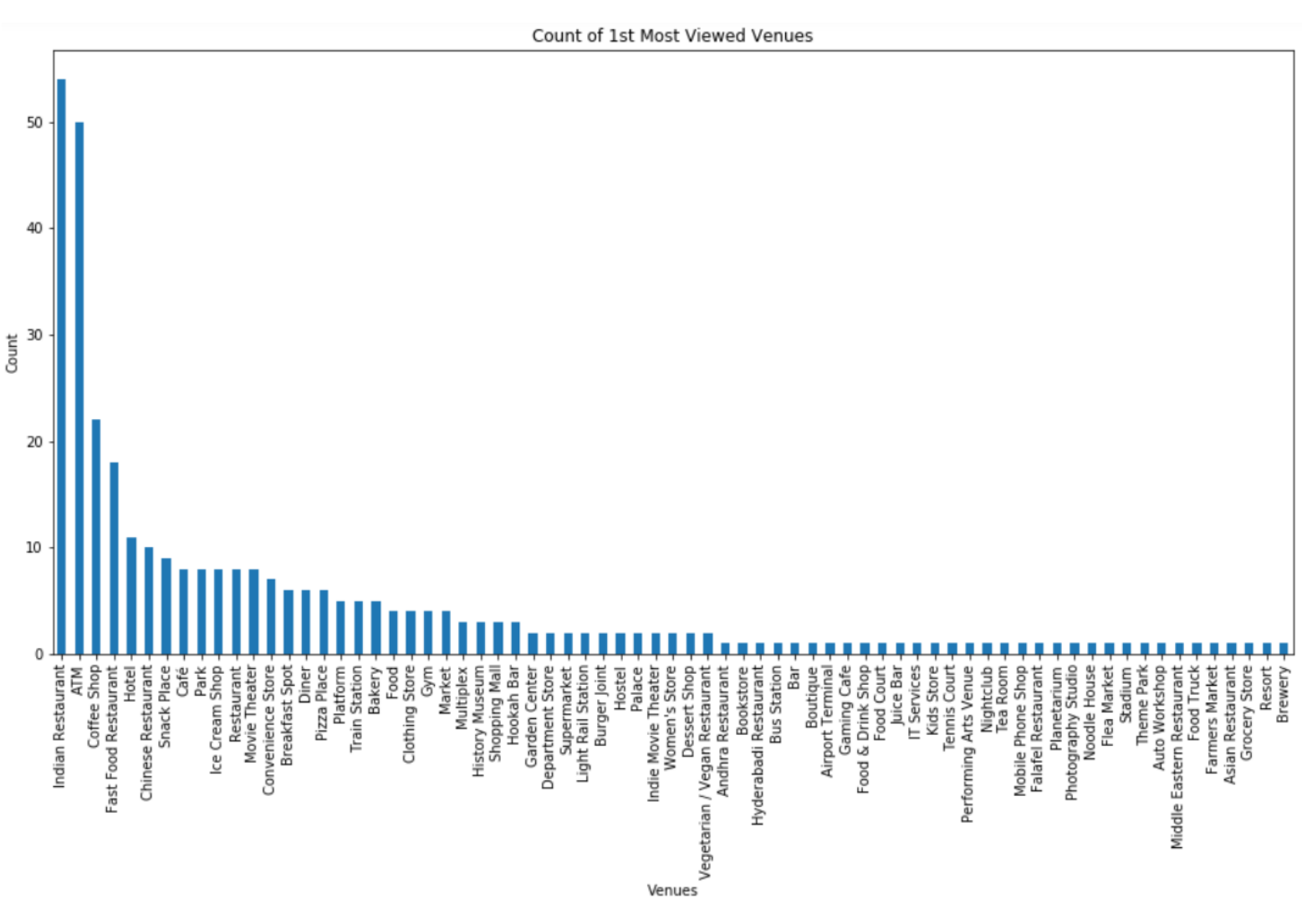
Using K-Means

	Areas	Latitude	Longitude	Cluster Labels	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	Gachibowli	17.4436	78.352	4	Café	Italian Restaurant	Department Store	Food Truck	Stadium	Indian Restaurant	Ice Cream Shop	Bakery	Hotel	Grocery Store
1	Shamshabad	17.2611	78.3932	4	Park	Café	Dhaba	Women's Store	Diner	Fast Food Restaurant	Farmers Market	Falafel Restaurant	Electronics Store	Dumpling Restaurant
2	Kukatpally	17.4931	78.4054	2	Indian Restaurant	Fast Food Restaurant	Pizza Place	Shop & Service	Metro Station	Middle Eastern Restaurant	Bakery	BBQ Joint	Ice Cream Shop	Food
3	Mallapur	17.4405	78.5789	4	Diner	Women's Store	Fishing Store	Fast Food Restaurant	Farmers Market	Falafel Restaurant	Electronics Store	Dumpling Restaurant	Donut Shop	Dog Run
4	Mallapur	17.4405	78.5789	4	Diner	Women's Store	Fishing Store	Fast Food Restaurant	Farmers Market	Falafel Restaurant	Electronics Store	Dumpling Restaurant	Donut Shop	Dog Run
5	Habsiguda	17.4155	78.5427	0	Indian Restaurant	Restaurant	Bakery	Sandwich Place	Park	Women's Store	Diner	Falafel Restaurant	Electronics Store	Dumpling Restaurant
6	Jubilee Hills	17.4308	78.4103	4	Indian Restaurant	Coffee Shop	Hookah Bar	Ice Cream Shop	Frozen Yogurt Shop	Bowling Alley	Bistro	Café	Juice Bar	Bar
7	Secunderabad	17.4691	78.5059	4	Hookah Bar	Sandwich Place	Ice Cream Shop	Café	Diner	Fast Food Restaurant	Farmers Market	Falafel Restaurant	Electronics Store	Dumpling Restaurant
8	Banjara Hills	17.4177	78.4399	4	Coffee Shop	Indian Restaurant	Café	Deli / Bodega	Sandwich Place	Bar	Bakery	Hookah Bar	Garden Center	Gym
9	Manikonda	17.4037	78.3766	0	Pizza Place	Restaurant	Women's Store	Dhaba	Farmers Market	Falafel Restaurant	Electronics Store	Dumpling Restaurant	Donut Shop	Dog Run

Using credentials of Foursquare API features of near-by places of the neighborhoods would be mined. Due to http request limitations the number of places per neighborhood parameter would reasonably be set to 100 and the radius parameter would be set to 500.

4. Results Section

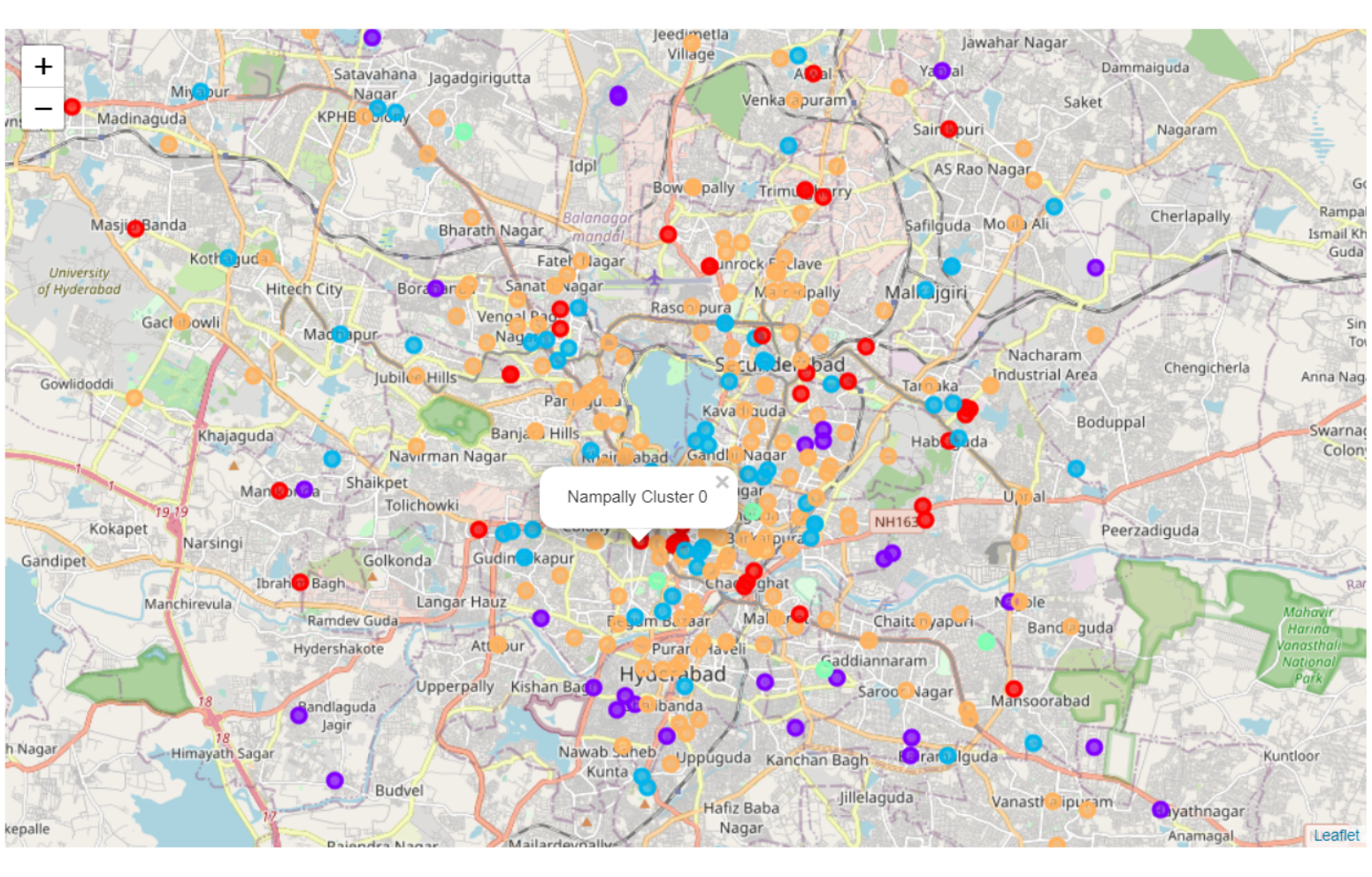
Most Common Venue Count in Neighborhood of Hyderabad



Foursquare API

The Foursquare Places API provides location based experiences with diverse information about venues, users, photos, and check-ins. The API supports real time access to places, Snap-to-Place that assigns users to specific locations, and Geo-tag.

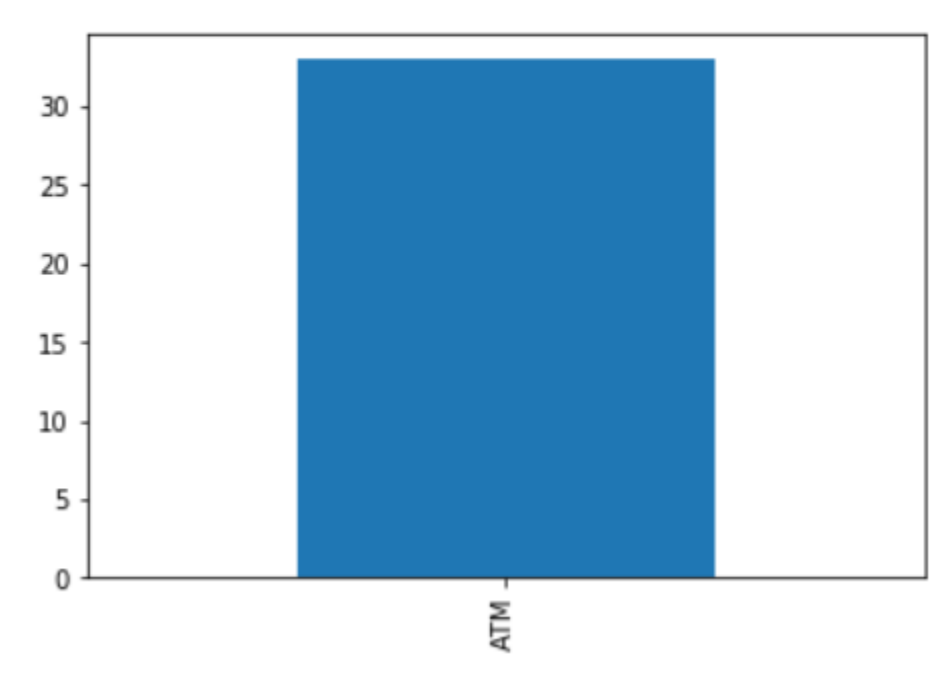
Cluster Map for Hyderabad



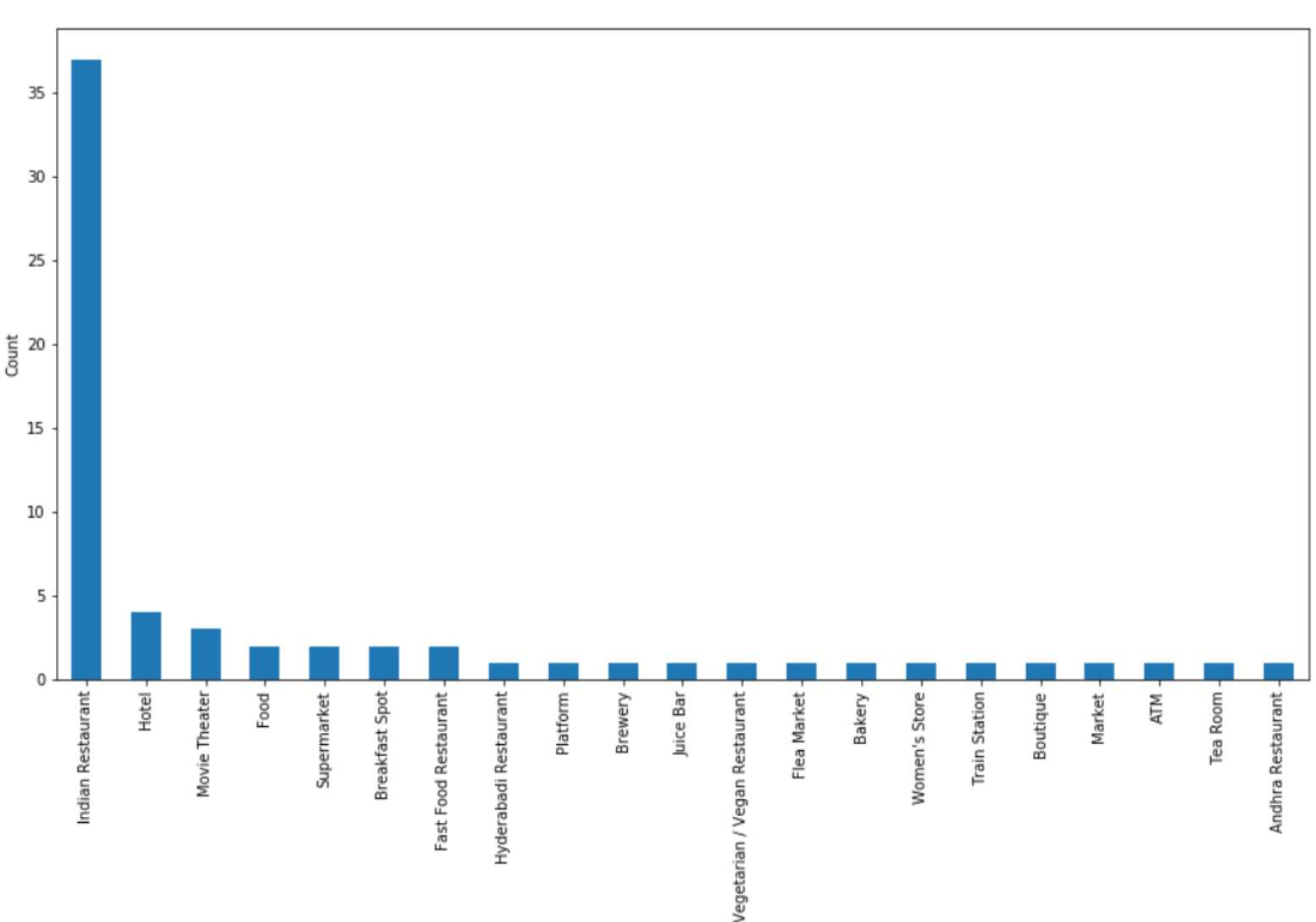
1st Most Common Venue for Cluster 1

```
Cluster_1['1st Most Common Venue'].value_counts().plot(kind = 'bar')
```

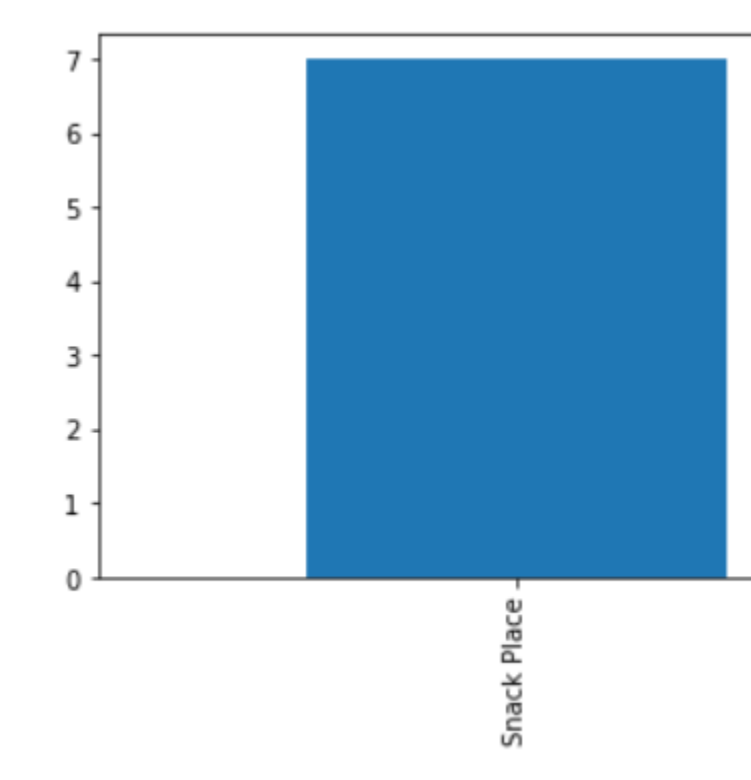
<matplotlib.axes._subplots.AxesSubplot at 0x17d62be36c8>



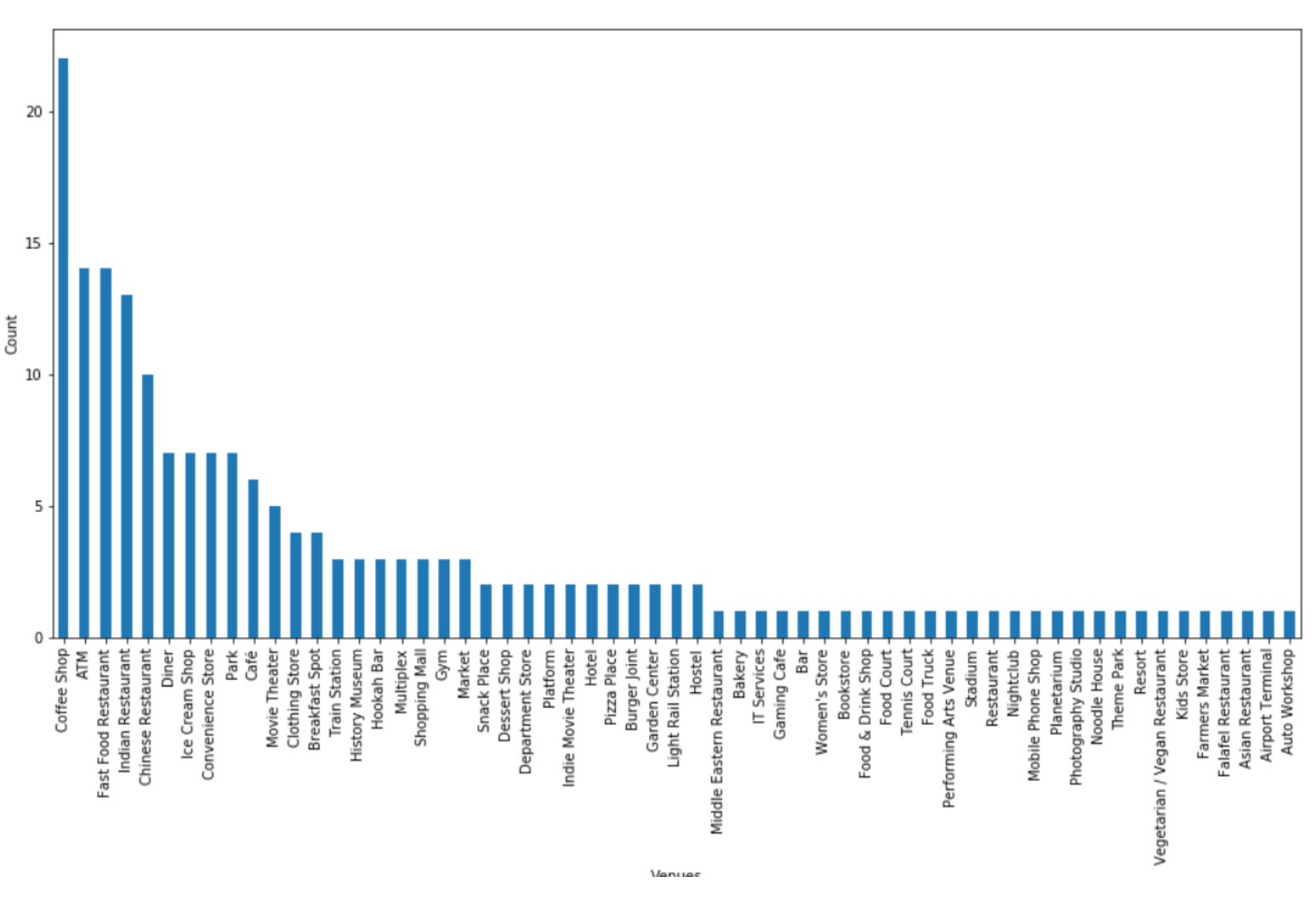
1st Most Common Venues for Cluster 2



1st Most Common Venue for Cluster 3



1st Most Common Venues for Cluster 4



5. Discussion Section

Problem trying to Solve

The main motive of the Project is to know which Neighbourhood is best to open Restaurants, Bakery... etc. People who what to open these can know which Neighbourhood is good. And Which facilities to recommend in that Neighbourhood.

- This Problem can be solved by taking data that which Neighbourhood is good according to previous data.
- By clustering same facilites in Neighbourhood we can decide the best venues to open.

6. Conclusion Section

In this project, using k-means cluster algorithm I separated the neighborhood into 5 different clusters and for 337 different lattitude and logitude from dataset, which have very-similar neighborhoods around them. The charts represent the Venues in different Clusters.

By Clusters we can say which Venue is most common in Neighborhood and can also say which neighborhood is famous for.

I feel rewarded with the efforts and believe this course with all the topics covered is well worthy of appreciation. This project has shown me a practical application to resolve a real situation that has impacting personal and financial impact using Data Science tools. The mapping with Folium is a very powerful technique to consolidate information and make the analysis and decision better with confidence.

In Future we can use this finding for implementing new Venues for Neighborhood.

Github link for this Project