Final Report | Capstone Project – Business Explorer

1. Introduction:

In any city, if someone is looking to open a restaurant then, where would you recommend that they open it? Similarly, if a contractor is trying to start their own business, where would you recommend that they set up their office?. The above questions define our business problem.
In other way find the best location or nearby area in neighborhoods in a large city of our choice.

1.1 Business Problem:

Nowadays. Find the best location or area to open a new business is a difficult task.
Location is the primary factor in starting a new business because it's affecting your business if you didn't choose the right location then your new business is not successful.

1.2 Target Audience of this project:

This project a particularly useful to those who want to open something new. Mean that open a new restaurant, shops, shopping malls, etc. Then a person is looking for the right area or location. But choosing the right place may be very difficult in a large city or city may be new for a person who wants to open a restaurant.
For example, the person knows the nearby frequency of restaurants, shops, or general stores. Then definitely, he will select the best location for new business

1.3 Solution:

This project aims to analyze the nearby areas in neighborhoods in the city of Delhi (we chose Delhi city for our problem). This means find popular venues category in Neighborhoods in Delhi (popular venues like a clothing store, gym, bus stations, other restaurants).
So, We have to segmenting or clustering neighborhoods in Delhi based on neighborhood information (number of different popular venues category).
And, then finally represent each cluster in graph form like bar charts using this graph. A person can easily choose a Cluster (The group of Neighborhoods areas) based on requirements.

The person can easily explore boroughs or neighborhood areas in the selected cluster.

2. Data Section:

Data link- https://en.wikipedia.org/wiki/List_of_neighbourhoods_of_Delhi
This is a Wikipedia page and we need to scrap this Wikipedia page using python packages for web scraping like BeautifulSoup, lxml.
Data link- https://www.kaggle.com/shaswatd673/delhi-neighborhood-data
But Delhi neighborhoods dataset already present on kaggle website. In this project, we download the Delhi Neighbourhoods dataset from above Kaggle website data link.

2.1 Data Description:

In data section part retrieve list of borough, neighborhoods in the city of Delhi.

1. Borough
2. Neighborhood
3. Neighborhood latitude (geographical coordinates for finding exact location)
4. Neighborhood longitude (geographical coordinates for finding exact location)

If we didn't find geographical coordinates then we can also use python packages (geopy) geolocator to convert address in string format into latitude or longitude and then based on these geographical coordinates fetch popular venue category(like a clothing store, Indian restaurant) in neighborhoods.

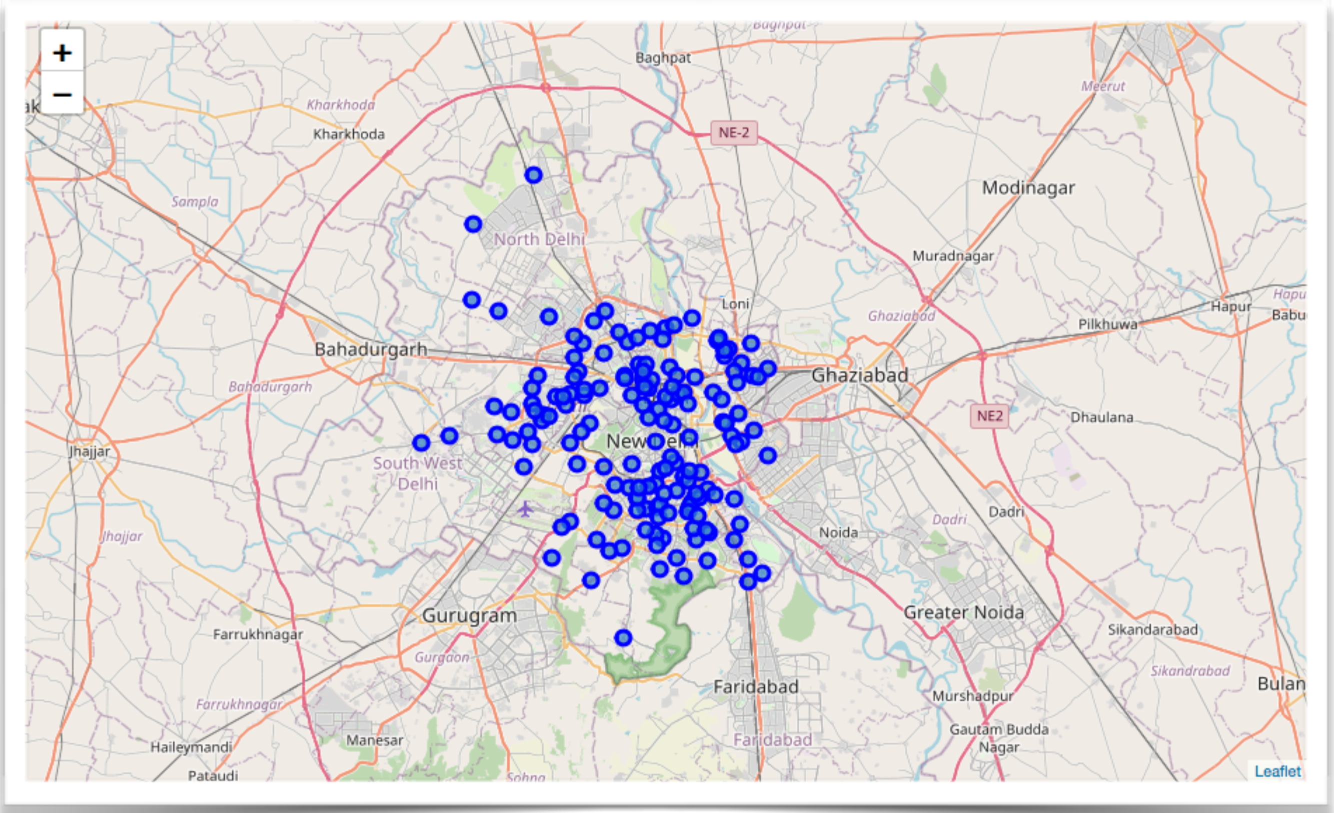
2.2 Foursquare API Data:

We will need data about different venues in different neighborhoods of that specific borough. 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 the neighborhood, we then connect to the Foursquare API to gather information about venues inside every
neighborhood.
For each neighborhood, we have chosen the radius to be 500 meters. The data retrieved from Foursquare contained information on venues within a specified distance of the longitude and latitude of the postcodes. The information obtained per venue as follows:

1. Borough

2. Neighbourhood
3. Neighbourhood Latitude
4. Neighbourhood Longitude
5. Venue
6. Name of the venue e.g. the name of a store or restaurant
7. Venue Latitude
8. Venue Longitude
9. Venue Category

Map Of Neighborhoods In Delhi



3. Methodology Section:

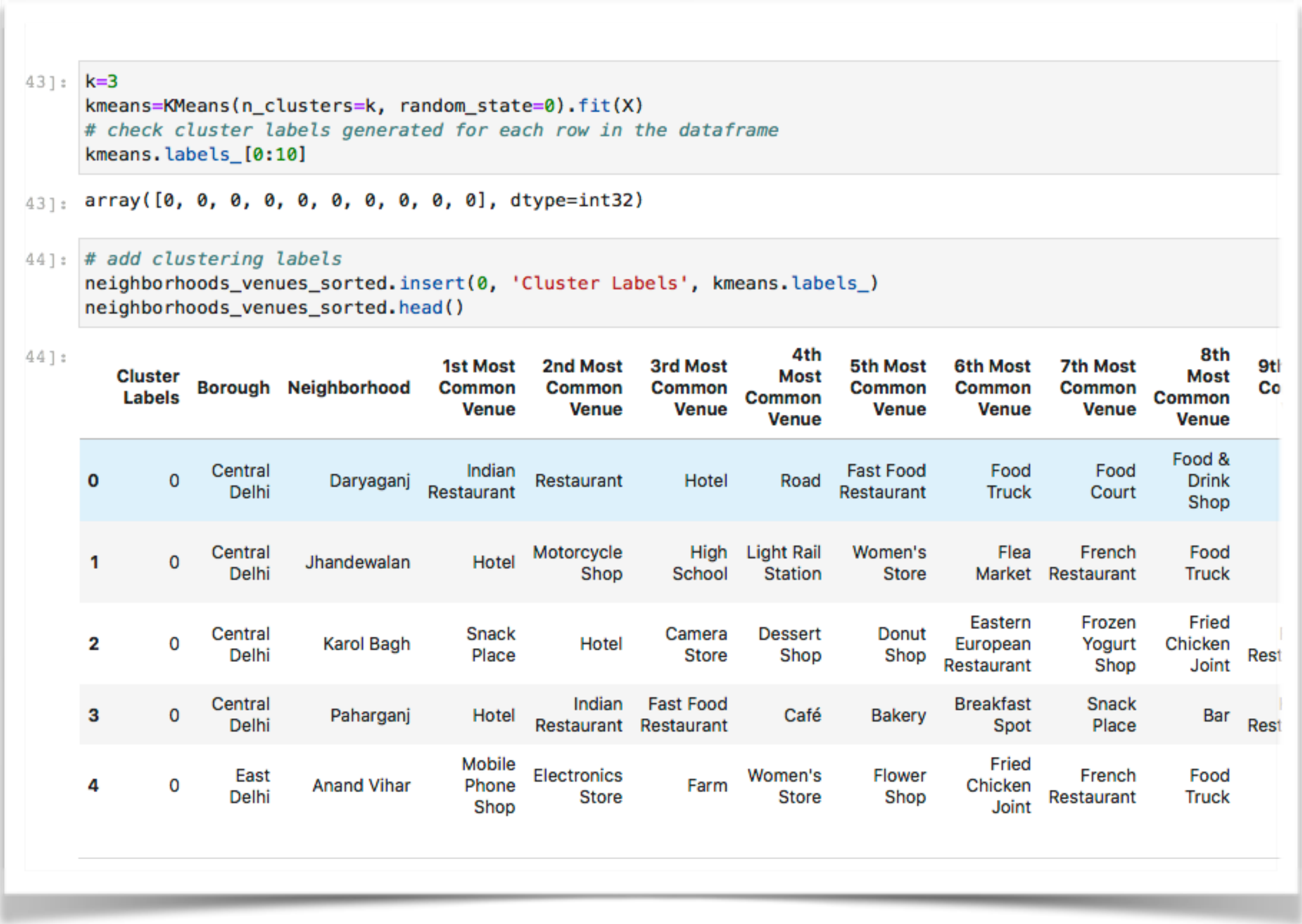
Clustering Approach:

First of all, we used the clustering approach to solve the above business problem. we decided to segment and cluster neighborhoods in Delhi based on neighborhood information.
To be able to do that, we need to cluster data which is a form of unsupervised machine learning: k-means clustering algorithm.
In k means clustering algorithm. We need to define a number of clusters to create and this is also a disadvantage of the k means clustering algorithm to predict the value of k (number of clusters).
how do we find the optimal value of k?

Silhouette method:

We will be using the silhouette method for finding optimal value of k.
For more information about the silhouette method then visit below link
Link-Silhouette method
For every k value, we will use k means clustering and then, find the average silhouette score. Silhouette score range from -1 to 1
the optimal k value is the one that has the highest average silhouette score.

Using K-Means Clustering Approach:



```
43]: k=3
     kmeans=KMeans(n_clusters=k, random_state=0).fit(X)
     # check cluster labels generated for each row in the dataframe
     kmeans.labels_[0:10]

43]: array([0, 0, 0, 0, 0, 0, 0, 0, 0, 0], dtype=int32)

44]: # add clustering labels
     neighborhoods_venues_sorted.insert(0, 'Cluster Labels', kmeans.labels_)
     neighborhoods_venues_sorted.head()

44]:
```

	Cluster Labels	Borough	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	9t Co
0	0	Central Delhi	Daryaganj	Indian Restaurant	Restaurant	Hotel	Road	Fast Food Restaurant	Food Truck	Food Court	Food & Drink Shop	
1	0	Central Delhi	Jhandewalan	Hotel	Motorcycle Shop	High School	Light Rail Station	Women's Store	Flea Market	French Restaurant	Food Truck	
2	0	Central Delhi	Karol Bagh	Snack Place	Hotel	Camera Store	Dessert Shop	Donut Shop	Eastern European Restaurant	Frozen Yogurt Shop	Fried Chicken Joint	Rest
3	0	Central Delhi	Paharganj	Hotel	Indian Restaurant	Fast Food Restaurant	Café	Bakery	Breakfast Spot	Snack Place	Bar	Rest
4	0	East Delhi	Anand Vihar	Mobile Phone Shop	Electronics Store	Farm	Women's Store	Flower Shop	Fried Chicken Joint	French Restaurant	Food Truck	

Find nearby popular venues of each neighborhood in Delhi using credentials of foursquare API. Due to HTTP request limitations, the number of popular venues per neighborhood parameter would reasonably be set to 100 and, the radius parameter would be set to 500 meters.

4. Results Section:

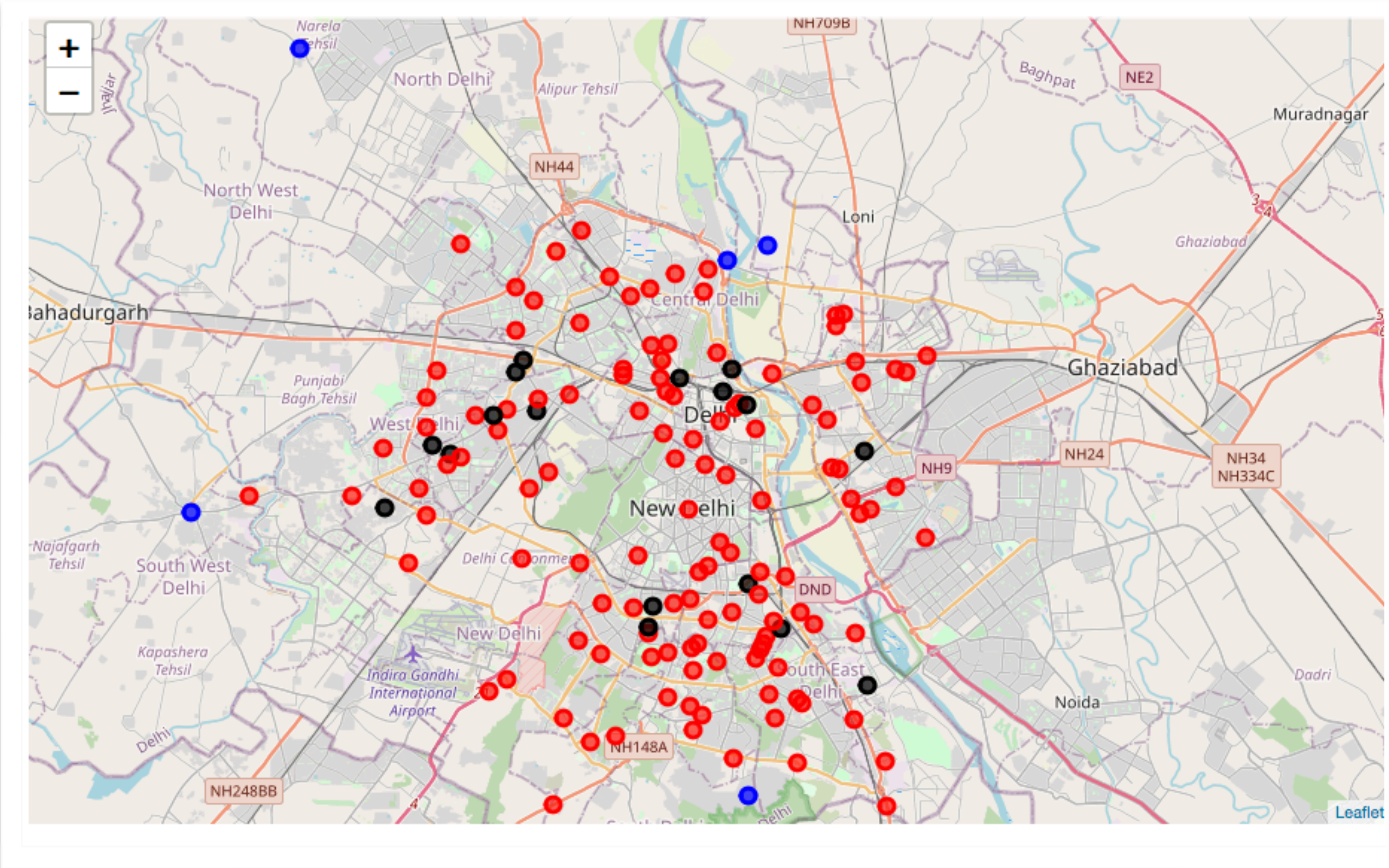
The result from k means clustering show that we categorizes the Neighborhooods in to 3 clusters based on frequency of different venue categories.

-> Cluster 0: Neighborhoods with high frequency of different venue categories

-> Cluster 1: Neighborhoods with low frequency of different venue categories

-> Cluster 2: Neighborhoods with moderate frequency of different venue categories

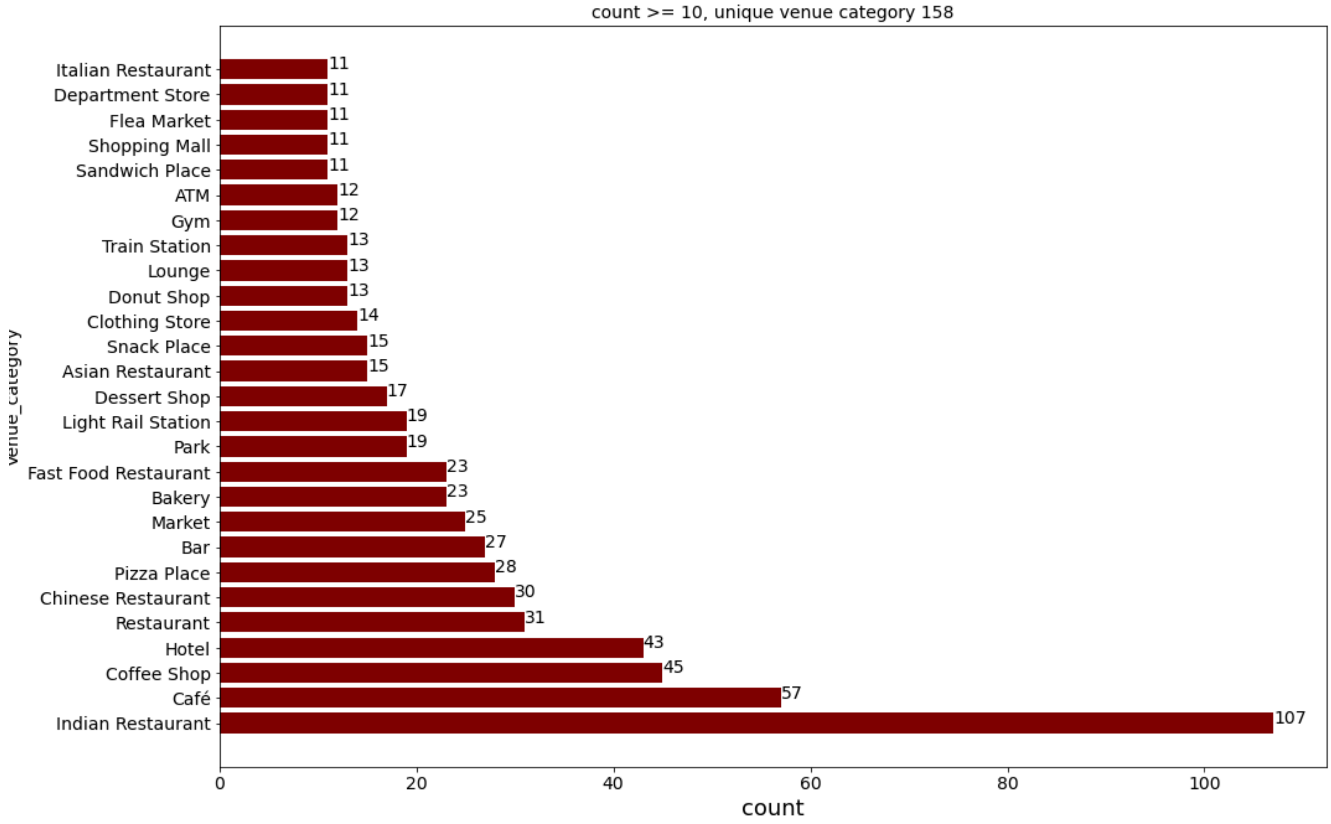
Map Of Clusters of Neighborhoods In Delhi



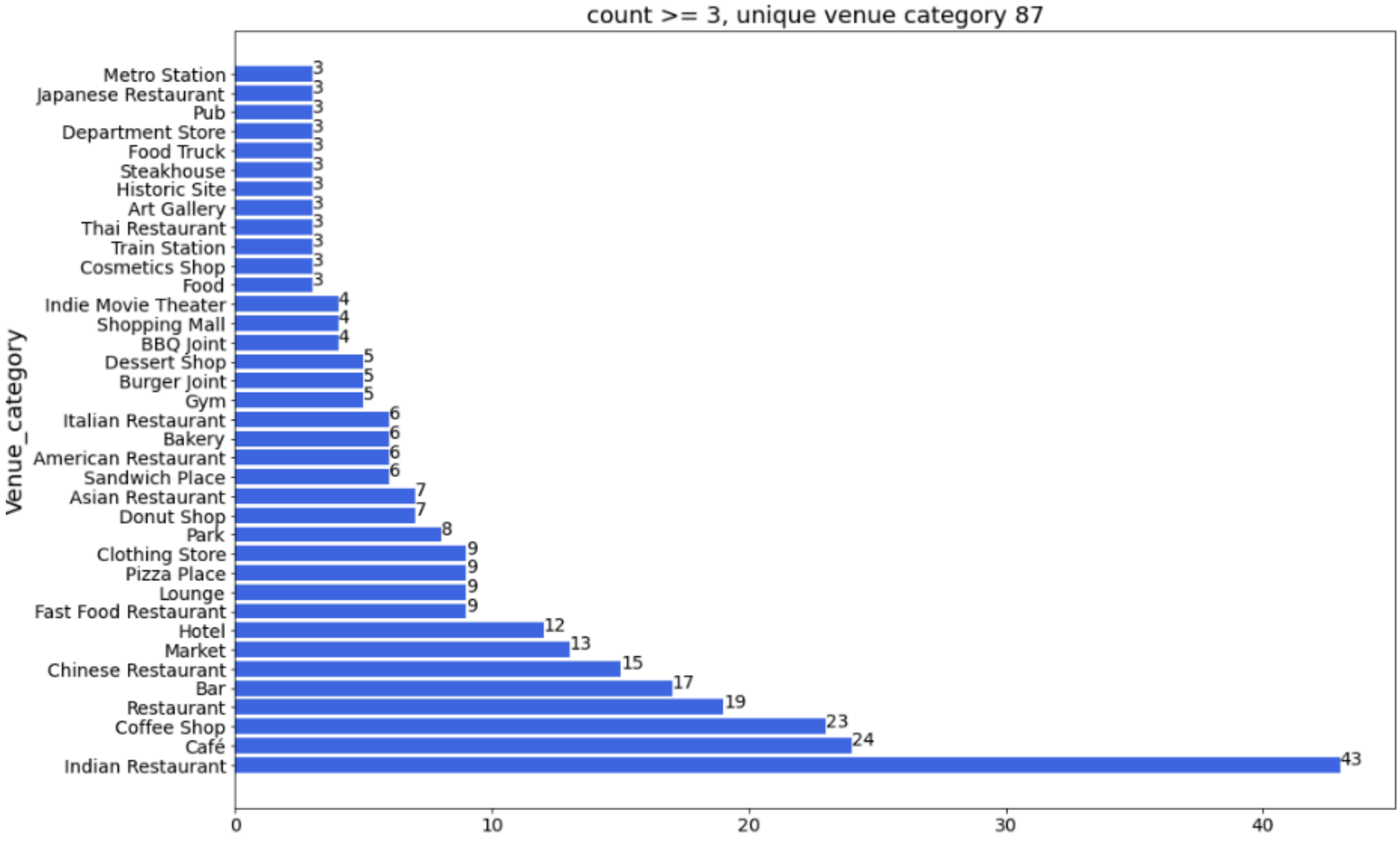
Each circular mark in above map represent neighborhood in Delhi and different colour show present in different cluster. Red colour for cluster 0, blue colour for cluster 1, black colour for cluster 2.

For each cluster represent different venue categories and also frequency of each venue category using bar graph.

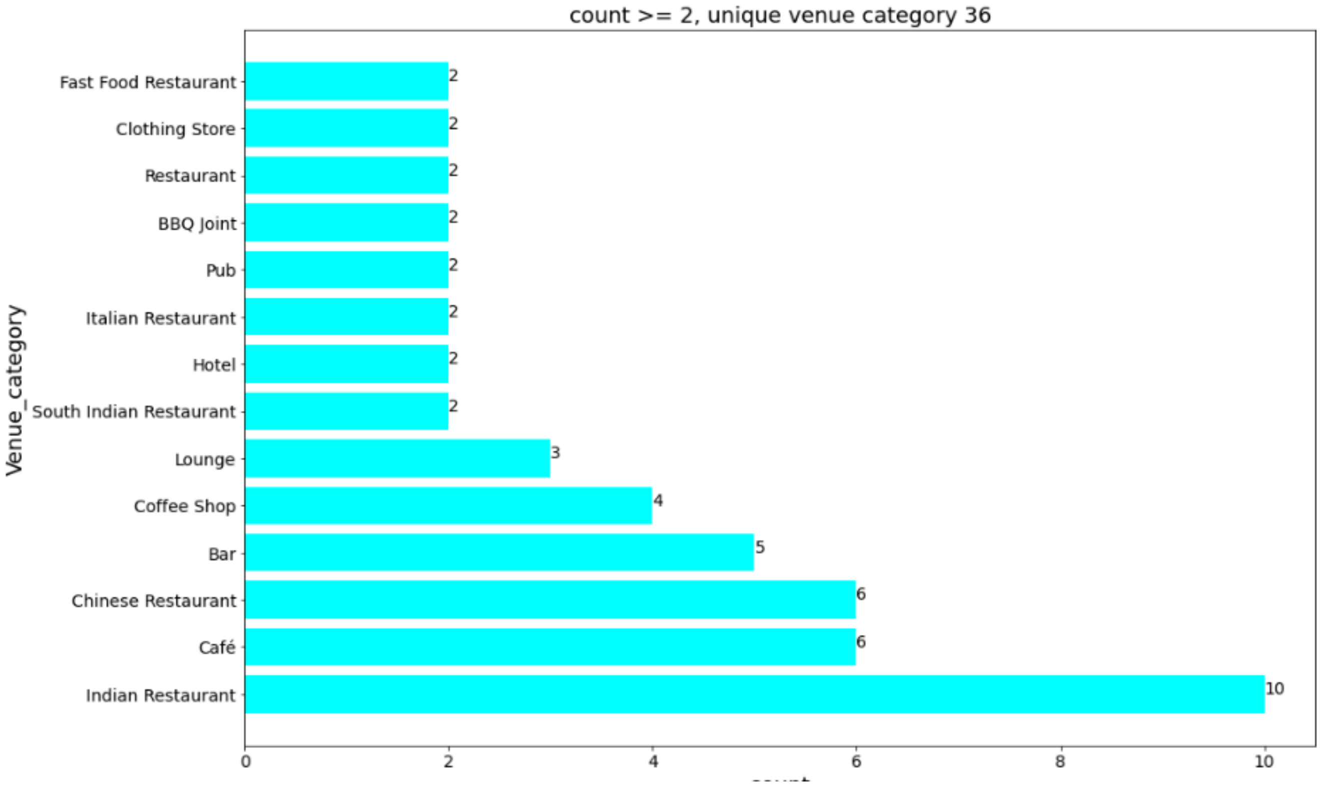
Explore Cluster Wise:



Explore Borough Wise:



Explore Neighbor Wise:



These above three graph shows either explore cluster wise and then explore borough or neighbor wise in selected cluster.

5. Discussion:

As observations noted from a map in the result section, most of the popular venues fall in neighborhoods in cluster 0. In cluster 1 have a low frequency of different Venues, cluster 2 with a moderate frequency of venues.
Suppose a person wants to open an Indian restaurant in Delhi.
Where would you recommend opening a restaurant?
After analyzing each cluster in the result section, as we know that cluster 0 and cluster 2 have a high frequency of Indian restaurants. In cluster 1 has a low number to no Indian restaurant.
Cluster 1 has a very high frequency of ATM (traffic area).
This is also a plus point for opening an Indian restaurant. If you want to know the exact location, Then you can also explore borough or neighbor in the selected cluster.
Here for understanding, we have taken only one example Indian restaurant. So we will assume any venue category with the replacement of Indian restaurants.

6. Limitations And Suggestions For Future Research:

In this project, we consider one factor i.e. frequency of different venue categories, there are other factors such as population and income of residents that could influence the location decision of a new venue.
However, To the best knowledge of this researcher, such data are not available to the neighborhood level required by this project.
Future research could devise a methodology to estimate such data to be used in clustering algorithm to determine preferred locations to open a new venue.

7. Conclusion:

In this project, We have gone through a process of identifying the business problem, specifying the data required, extracting and preparing data, performing machine learning by clustering the data into three clusters based on similarities.
Lastly, provide recommendations to the relevant stakeholders. And those who want to open fast-food restaurants, clothing stores, gyms, etc. To answer a business question that had raised in the introduction section, the answer proposed by this project.

Libraries Which are Used to Develop the Project:

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.

Geocoder: To retrieve Location Data.

Matplotlib: Python Plotting Module.

References:

Foursquare Developers Documentation.
https://developer.foursquare.com/docs

Geocoder to retrieve geographical coordinates from address.https://www.kite.com/python/docs/geopy.geocoders