

March 2021

Deepali Sharma



# Capstone Project - The Battle of Neighbourhoods

## Introduction

New Delhi is the national capital of India. It has a population of 257,803 and is the largest commercial city in Northern India. New Delhi is a cosmopolitan city due to the multi-ethnic and multi-cultural presence of the vast Indian bureaucracy and political system. The city's capital status has amplified the importance of national events and holidays. The city is already among the world's fastest changing cities, with an ability to embrace technological change, and deal with rapid population growth while strengthening global connectivity. Traditionally speaking, a mixture of ancient Indian and Mughal styled cuisine is now considered as authentic Delhi cuisine. Street foods make a major part of the cuisine of the land. Top street food of Delhi is lassi, kebab, chole bhature, jalebis, falooda, samosa, butter chicken, Nihari, kulfi, and others.

## Data

For this project, we need the following data :

- List of New Delhi restaurants

- Locality
- Longitude
- Ratings

Description : The dataset contains the required information and we will use this dataset to explore various restaurants located in various parts of New Delhi.

## Approach

Collect the Zomato Kaggle dataset for India and extracting data for New Delhi. Using FourSquare API, we will find all venues/restaurants for each locality. Filter out the restaurants that are nearby locality. Use Aggregate Rating for restaurant to find the best and worst place to eat. Visualize the Ranking of localities using folium library in Python.

*Import libraries*

```
In [ ]: from geopy.geocoders import Nominatim
import numpy as np # library to handle data in a vectorized manner

import pandas as pd # library for data analysis
pd.set_option('display.max_columns', None)
pd.set_option('display.max_rows', None)

import json # library to handle JSON files

#!conda install -c conda-forge geopy --yes # uncomment this line if you hav
from geopy.geocoders import Nominatim # convert an address into latitude an

import requests # library to handle requests
from pandas.io.json import json_normalize # tranform JSON file into a panda

# Matplotlib and associated plotting modules
import matplotlib.cm as cm
import matplotlib.colors as colors

# import k-means from clustering stage
from sklearn.cluster import KMeans

#!conda install -c conda-forge folium=0.5.0 --yes # uncomment this line if
import folium # map rendering library

from bs4 import BeautifulSoup
import requests, os, string

print('Libraries imported.')
```

Libraries imported.

**The Dataset contains list of restaurants for different states of India. We will be working on the restaurants located in New Delhi.Using Locality we use can find out the restaurants.**

*Tranform the data into a pandas dataframe*

```
In [ ]: df = pd.read_csv('https://raw.githubusercontent.com/haanjiankur/Capstone-Pr
df.head()
```

Out[3]:

	Restaurant ID	Restaurant Name	Country Code	City	Address	Locality	Locality Verbose	Longitude
0	6317637	Le Petit Souffle	162	Makati City	Third Floor, Century City Mall, Kalayaan Avenu...	Century City Mall, Poblacion, Makati City	Century City Mall, Poblacion, Makati City, Mak...	121.027535
1	6304287	Izakaya Kikufuji	162	Makati City	Little Tokyo, 2277 Chino Roces Avenue, Legaspi...	Little Tokyo, Legaspi Village, Makati City	Little Tokyo, Legaspi Village, Makati City, Ma...	121.014101
2	6300002	Heat - Edsa Shangri-La	162	Mandaluyong City	Edsa Shangri-La, 1 Garden Way, Ortigas, Mandal...	Edsa Shangri-La, Ortigas, Mandaluyong City	Edsa Shangri-La, Ortigas, Mandaluyong City, Ma...	121.056831
3	6318506	Ooma	162	Mandaluyong City	Third Floor, Mega Fashion Hall, SM Megamall, O...	SM Megamall, Ortigas, Mandaluyong City	SM Megamall, Ortigas, Mandaluyong City, Mandal...	121.056475
4	6314302	Sambo Kojin	162	Mandaluyong City	Third Floor, Mega Atrium, SM Megamall, Ortigas...	SM Megamall, Ortigas, Mandaluyong City	SM Megamall, Ortigas, Mandaluyong City, Mandal...	121.057508

```
In [ ]: df_india = df[df['Country Code'] == 1]
df_NDLS = df_india[df_india['City'] == 'New Delhi']
df_NDLS.reset_index(drop=True, inplace=True)
df_NDLS.head()
```

Out[5]:

	Restaurant ID	Restaurant Name	Country Code	City	Address	Locality	Locality Verbose	Longitude	Latitude	C
0	18287358	Food Cloud	1	New Delhi	Aaya Nagar, New Delhi	Aaya Nagar	Aaya Nagar, New Delhi	0.000000	0.000000	
1	18216944	Burger.in	1	New Delhi	84, Near Honda Showroom, Adchini, New Delhi	Adchini	Adchini, New Delhi	77.196923	28.535382	Fa
2	313333	Days of the Raj	1	New Delhi	81/3, 1st Floor, Qutub Residency, Adchini, New...	Adchini	Adchini, New Delhi	77.197475	28.535493	S Cor
3	18384127	Dilli Ka Dhaba	1	New Delhi	66 A, Ground Floor, Sri Aurobindo Marg, Adchin...	Adchini	Adchini, New Delhi	77.198033	28.537547	
4	582	Govardhan	1	New Delhi	84, Adjacent Hero Motor Bike Showroom, Main Me...	Adchini	Adchini, New Delhi	77.196924	28.535523	

## Data Cleaning

```
In [ ]: #remove the unwanted columns and rows from dataset
df_Res= df_NDLS[df_NDLS.Longitude !=0.000000][['Restaurant Name','Locality']
```

```
In [ ]: df_Res = df_Res[df_Res['Aggregate rating'] !=0.0]
df_Res.head()
```

Out[7]:

	Restaurant Name	Locality	Longitude	Latitude	Cuisines	Aggregate rating	Rating text	Votes
1	Burger.in	Adchini	77.196923	28.535382	Fast Food	3.2	Average	46
2	Days of the Raj	Adchini	77.197475	28.535493	North Indian, Seafood, Continental	3.4	Average	45
3	Dilli Ka Dhaba	Adchini	77.198033	28.537547	South Indian, North Indian	2.6	Average	11
4	Govardhan	Adchini	77.196924	28.535523	South Indian, North Indian, Chinese	3.4	Average	238
5	Mezbaan Grills	Adchini	77.198122	28.538134	Mughlai	3.1	Average	8

```
In [ ]: address = 'New Delhi,India'

geolocator = Nominatim(user_agent="mum_explorer")
location = geolocator.geocode(address)
latitude = location.latitude
longitude = location.longitude
print('The geograpical coordinate of Mumbai City are {}, {}'.format(latitude, longitude))
```

The geograpical coordinate of Mumbai City are 28.6138954, 77.2090057.

Creating a map to show cluster of Restaurants

```

In [ ]: Delhi_Rest = folium.Map(location=[latitude, longitude], zoom_start=12)

X = df_Res['Latitude']
Y = df_Res['Longitude']
Z = np.stack((X, Y), axis=1)

kmeans = KMeans(n_clusters=5, random_state=0).fit(Z)

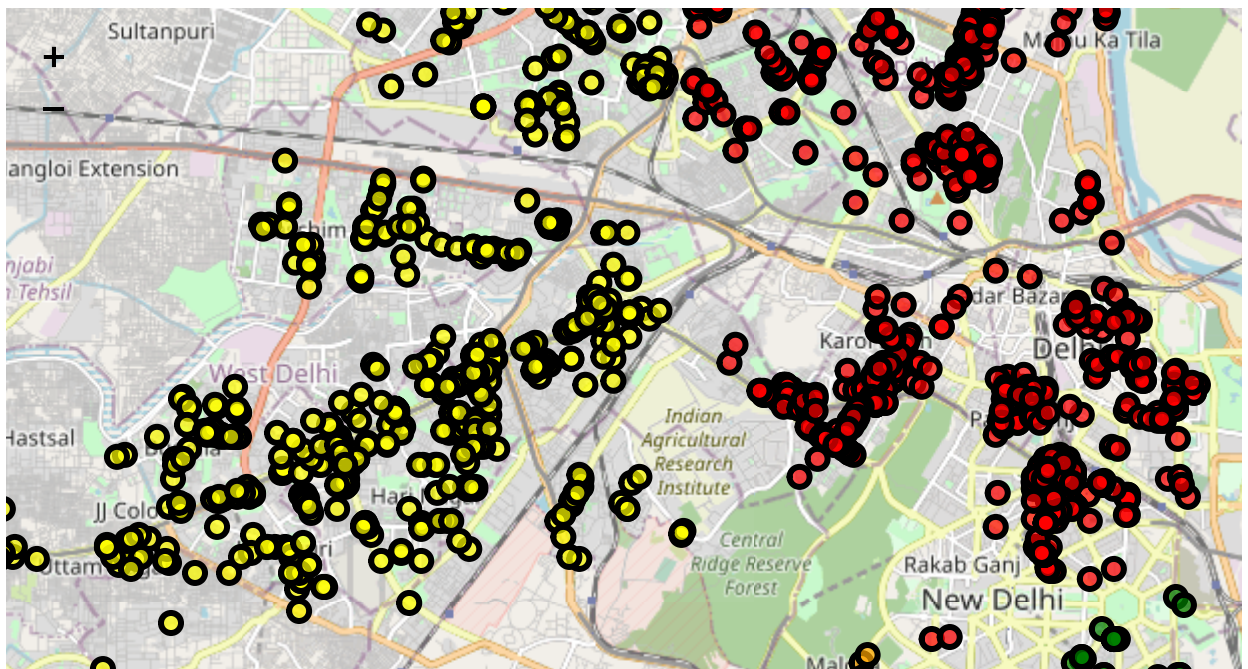
clusters = kmeans.labels_
colors = ['red', 'green', 'blue', 'yellow', 'orange']
df_Res['Cluster'] = clusters

for latitude, longitude, Locality, cluster in zip(df_Res['Latitude'], df_Res['Longitude'], df_Res['Locality'], df_Res['Cluster']):
    label = folium.Popup(Locality, parse_html=True)
    folium.CircleMarker(
        [latitude, longitude],
        radius=5,
        popup=label,
        color='black',
        fill=True,
        fill_color=colors[cluster],
        fill_opacity=0.7).add_to(Delhi_Rest)

Delhi_Rest

```

Out[9]:



```
In [ ]: df_Res.head()
```

Out[39]:

	Restaurant Name	Locality	Longitude	Latitude	Cuisines	Aggregate rating	Rating text	Votes	Cluster
0	Joey's Pizza	Azad Nagar	72.829976	19.126630	Pizza	4.0	Very Good	5145	2
1	Cafe Hydro	Borivali East	72.862381	19.221315	Asian	4.0	Very Good	156	1
2	The American Joint	Borivali West	72.841347	19.223840	Healthy Food, American, Burger, Salad	3.4	Average	170	1
3	The Fusion Kitchen	Borivali West	72.848923	19.254567	North Indian, Italian, Chinese, Mexican	4.7	Excellent	2083	1
4	145 Kala Ghoda	Fort	72.832585	18.927584	Fast Food, Beverages, Desserts	4.2	Very Good	1606	0

```
Plotting highest rated restaurant in top 10 locality of Delhi
```



```

In [ ]: import matplotlib.pyplot as plt
plt.figure(figsize=(9,5), dpi = 100)

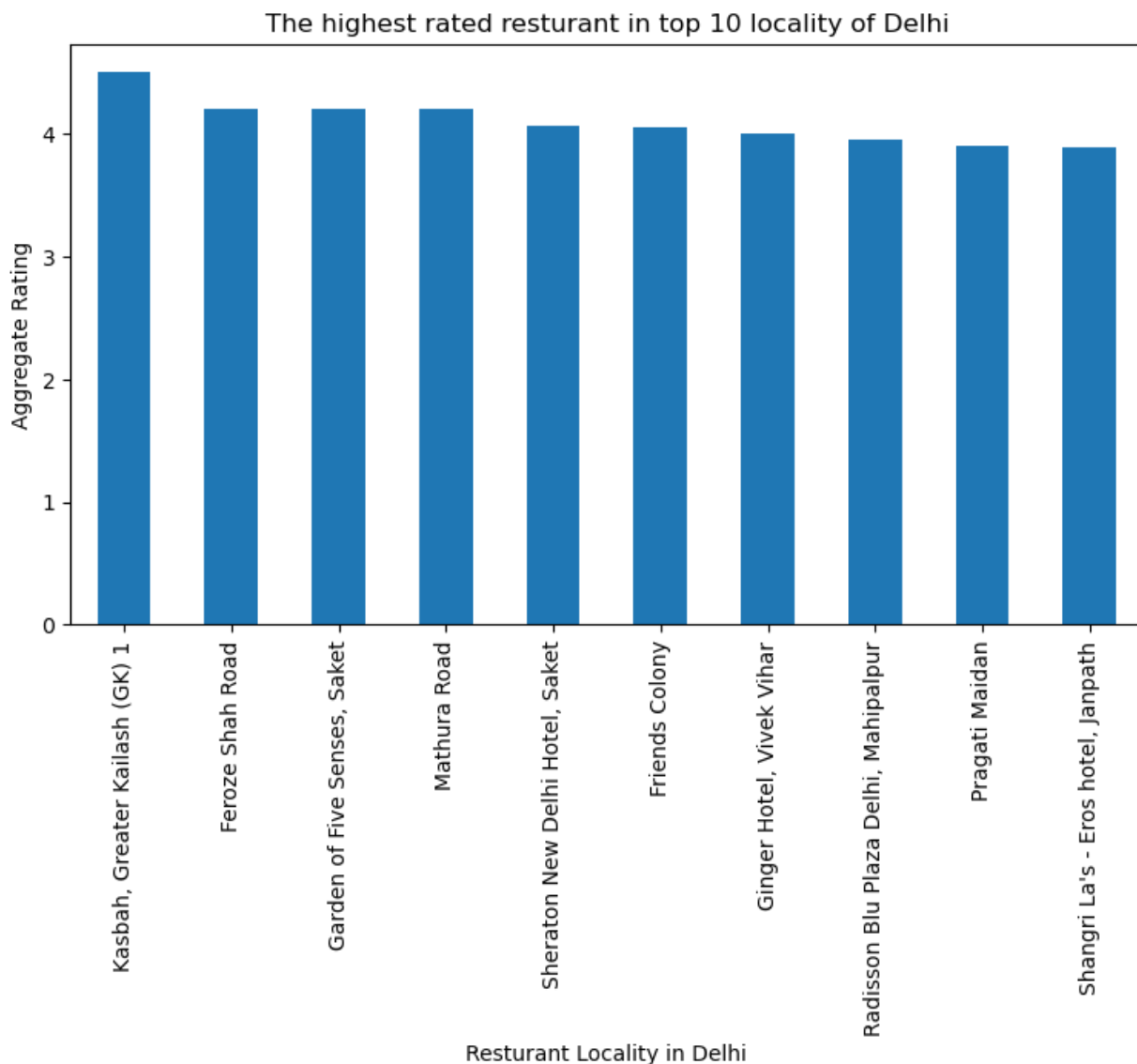
x=['Locality']
y=['Aggregate rating']
# title
plt.title('The highest rated resturant in top 10 locality of Delhi')
#On x-axis

#giving a bar plot
df_Res.groupby('Locality')['Aggregate rating'].mean().nlargest(10).plot(kind='bar')

plt.xlabel('Resturant Locality in Delhi')
#On y-axis
plt.ylabel('Aggregate Rating')

#displays the plot
plt.show()

```



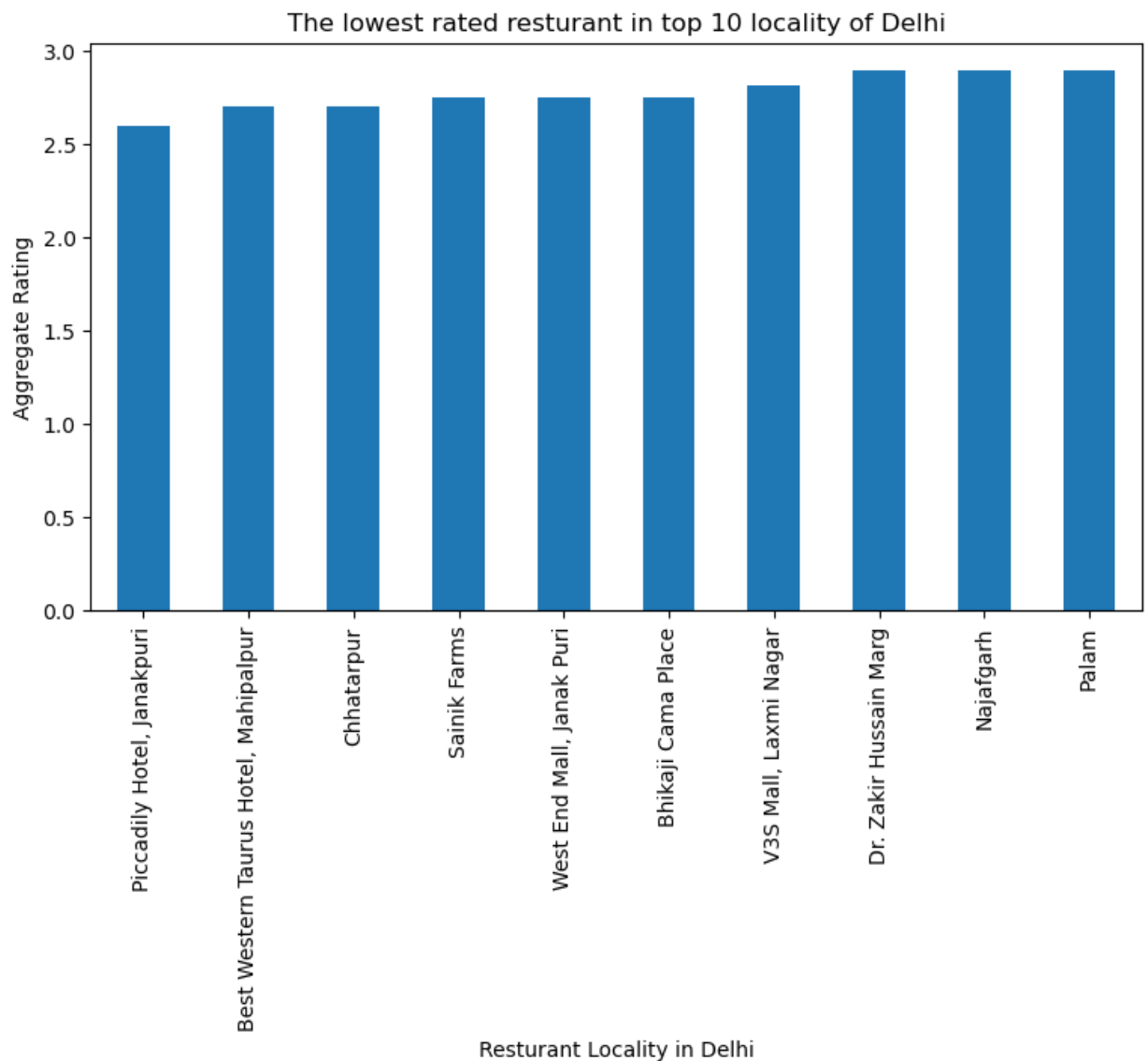


Plotting highest rated restaurant in top 10 locality of Delhi

```
In [ ]: import matplotlib.pyplot as plt
plt.figure(figsize=(9,5), dpi = 100)
# title
plt.title('The lowest rated resturant in top 10 locality of Delhi')
#On x-axis

#giving a bar plot
df_Res.groupby('Locality')['Aggregate rating'].mean().nsmallest(10).plot(kind='bar')

plt.xlabel('Resturant Locality in Delhi')
#On y-axis
plt.ylabel('Aggregate Rating')
#displays the plot
plt.show()
```



Plotting the highest number of Restaurant available in Locality of Delhi

In [ ]:

```

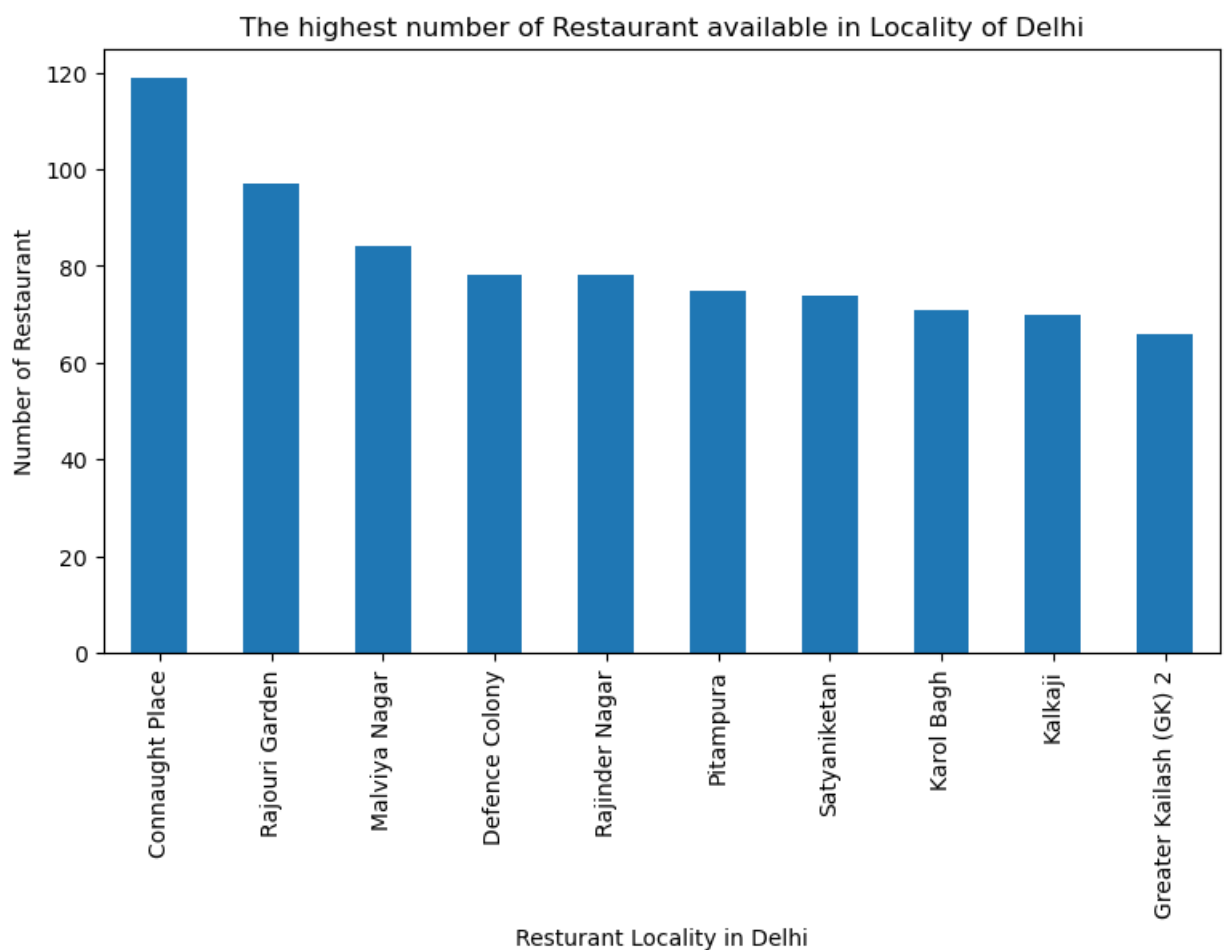
import matplotlib.pyplot as plt
plt.figure(figsize=(9,5), dpi = 100)
# title
plt.title('The highest number of Restaurant available in Locality of Delhi'
#On x-axis

#giving a bar plot
df_Res.groupby('Locality')['Restaurant Name'].count().nlargest(10).plot(kind='bar')

plt.xlabel('Resturant Locality in Delhi')
#On y-axis
plt.ylabel('Number of Restaurant')

#displays the plot
plt.show()

```



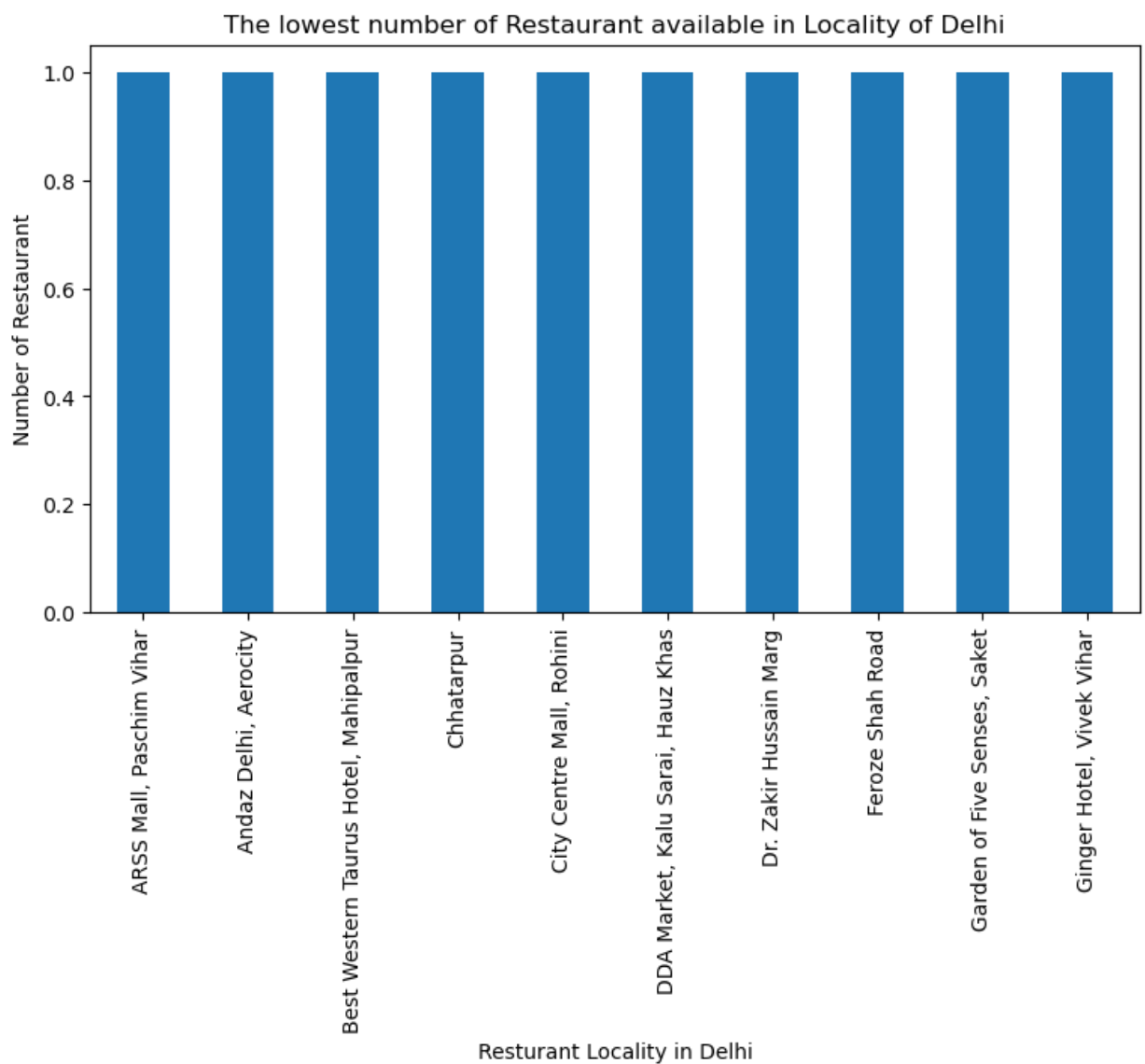
***Plotting the lowest number of Restaurant available in Locality of Delhi***

```
In [ ]:
import matplotlib.pyplot as plt
plt.figure(figsize=(9,5), dpi = 100)
# title
plt.title('The lowest number of Restaurant available in Locality of Delhi')
#On x-axis

#giving a bar plot
df_Res.groupby('Locality')['Restaurant Name'].count().nsmallest(10).plot(ki

plt.xlabel('Resturant Locality in Delhi')
#On y-axis
plt.ylabel('Number of Restaurant')

#displays the plot
plt.show()
```



## Data Transformation

```
In [ ]: df_Res_Loc = df_Res.groupby('Locality').count()['Restaurant Name'].to_frame()
df_Res_rating = df_Res.groupby('Locality')['Aggregate rating'].mean().to_frame()
df_Cuisines = df_Res.groupby(['Locality'])['Cuisines'].agg(', '.join).reset_index()
df_R = df_Res.groupby(['Locality'])['Rating text'].unique().agg(', '.join).reset_index()
df_V = df_Res.groupby(['Locality'])['Votes'].sum().to_frame()
df_Lat = df_Res.groupby('Locality').mean()['Latitude'].to_frame()
df_Lng = df_Res.groupby('Locality').mean()['Longitude'].to_frame()
df_final = pd.merge(df_Lat, df_Lng, on='Locality').merge(df_Res_Loc, on='Locality').merge(df_Res_rating, on='Locality').merge(df_Cuisines, on='Locality').merge(df_R, on='Locality').merge(df_V, on='Locality')

df_final = df_final[df_final['Aggregate rating'] != 0.000000]
df_final.columns = ['Locality', 'Lat', 'Lng', 'No_of_Restaurant', 'Cusines', 'Agg_Rating', 'Comments', 'No_of_Votes']
df_final.head()
```

Out[61]:

	Locality	Lat	Lng	No_of_Restaurant	Cusines	Agg_Rating	Comments	No_of_Votes
0	ARSS Mall, Paschim Vihar	28.668945	77.101544	1	North Indian, South Indian, Chinese, Mithai, F...	3.100000	Average	
1	Adchini	28.537063	77.197808	13	Fast Food, North Indian, Seafood, Continental,...	3.292308	Average, Good, Poor, Very Good	
2	Aditya Mega Mall, Karkardooma	28.656131	77.301266	4	Finger Food, North Indian, Mughlai, Pizza, Fas...	3.275000	Average, Good	
3	Aerocity	28.553077	77.104270	2	Fast Food, Italian, Pizza, North Indian, Conti...	3.200000	Average	
4	Aggarwal City Mall, Pitampura	28.690020	77.134650	3	North Indian, Chinese, Street Food, Mithai, No...	3.033333	Average	

```
In [ ]: df_final.shape
```

Out[62]: (240, 8)

Define Foursquare credentials and version

```
In [ ]: CLIENT_ID = 'BPCNO0YRKLDHVQH1OWE2A2FYGZH4OPSMRIL4X5YL5RRCBGHJ' # your Fours
CLIENT_SECRET = 'EMIQF1UMPHY1D5O4XD1UGJSWPQCBTGD3SXJAWS1S43QWV3TO' # your F
VERSION = '20180605' # Foursquare API version
LIMIT = 100 # A default Foursquare API limit value

print('Your credentails:')
print('CLIENT_ID: ' + CLIENT_ID)
print('CLIENT_SECRET: ' + CLIENT_SECRET)
```

Your credentails:

CLIENT\_ID: BPCNO0YRKLDHVQH1OWE2A2FYGZH4OPSMRIL4X5YL5RRCBGHJ

CLIENT\_SECRET:EMIQF1UMPHY1D5O4XD1UGJSWPQCBTGD3SXJAWS1S43QWV3TO

```
In [ ]: def getNearbyVenues(names, latitudes, longitudes):
    radius=500
    LIMIT=100
    venues_list=[]
    for name, lat, lng in zip(names, latitudes, longitudes):
        print(name)

        # create the API request URL
        url = 'https://api.foursquare.com/v2/venues/explore?&client_id={}&c
            CLIENT_ID,
            CLIENT_SECRET,
            VERSION,
            lat,
            lng,
            radius,
            LIMIT)

        # make the GET request
        results = requests.get(url).json()["response"]["groups"][0]["items"]

        # return only relevant information for each nearby venue
        venues_list.append([
            name,
            lat,
            lng,
            v['venue']['name'],
            v['venue']['location']['lat'],
            v['venue']['location']['lng'],
            v['venue']['categories'][0]['name'] for v in results])

    nearby_venues = pd.DataFrame([item for venue_list in venues_list for it
    nearby_venues.columns = ['Locality',
                            'Locality Latitude',
                            'Locality Longitude',
                            'Venue',
                            'Venue Latitude',
                            'Venue Longitude',
                            'Venue Category']

    return(nearby_venues)
```



## 2. Explore Neighborhoods in Delhi

```
In [ ]: # find the venues in all New Delhi Locality
new_Delhi_venues = getNearbyVenues(names=df_final['Locality'],
                                   latitudes=df_final['Lat'],
                                   longitudes=df_final['Lng']
                                   )
```

ARSS Mall, Paschim Vihar  
 Adchini  
 Aditya Mega Mall, Karkardooma  
 Aerocity  
 Aggarwal City Mall, Pitampura  
 Aggarwal City Plaza, Rohini  
 Alaknanda  
 Ambience Mall, Vasant Kunj  
 Anand Lok  
 Anand Vihar  
 Andaz Delhi, Aerocity  
 Ansal Plaza Mall, Khel Gaon Marg  
 Asaf Ali Road  
 Ashok Vihar Phase 1  
 Ashok Vihar Phase 2  
 Ashok Vihar Phase 3  
 Barakhamba Road  
 Basant Lok Market, Vasant Vihar  
 Bellagio, Ashok Vihar Phase 2  
 ...

```
In [ ]: new_Delhi_venues.head()
```

Out[70]:

	Locality	Locality Latitude	Locality Longitude	Venue	Venue Latitude	Venue Longitude	Venue Category
0	ARSS Mall, Paschim Vihar	28.668945	77.101544	Subway	28.669999	77.102546	Sandwich Place
1	ARSS Mall, Paschim Vihar	28.668945	77.101544	Pizza Hut	28.670321	77.103853	Pizza Place
2	ARSS Mall, Paschim Vihar	28.668945	77.101544	Café Coffee Day	28.670009	77.102480	Coffee Shop
3	ARSS Mall, Paschim Vihar	28.668945	77.101544	Baljeet's Amritsari Koolcha	28.665768	77.100481	Indian Restaurant
4	ARSS Mall, Paschim Vihar	28.668945	77.101544	Little Chef	28.670000	77.101459	Chinese Restaurant

```
In [ ]: new_Delhi_venues.groupby('Locality').count()
```

Out[71]:

	Locality Latitude	Locality Longitude	Venue	Venue Latitude	Venue Longitude	Venue Category
Locality						
ARSS Mall, Paschim Vihar	9	9	9	9	9	9
Adchini	6	6	6	6	6	6
Aditya Mega Mall, Karkardooma	10	10	10	10	10	10
Aerocity	7	7	7	7	7	7
Aggarwal City Mall, Pitampura	3	3	3	3	3	3
Aggarwal City Plaza, Rohini	7	7	7	7	7	7
Alaknanda	4	4	4	4	4	4
Ambience Mall, Vasant Kunj	46	46	46	46	46	46
Anand Lok	11	11	11	11	11	11

```
In [ ]: print('There are {} uniques categories.'.format(len(new_Delhi_venues['Venue
```

There are 207 uniques categories.

### 3. Analyze Each Neighborhood

```
In [ ]: # one hot encoding
new_Delhi_onehot = pd.get_dummies(new_Delhi_venues[['Venue Category']], pre

# add neighborhood column back to dataframe
new_Delhi_onehot['Locality'] = new_Delhi_venues['Locality']

# move neighborhood column to the first column
fixed_columns = [new_Delhi_onehot.columns[-1]] + list(new_Delhi_onehot.colu
new_Delhi_onehot = new_Delhi_onehot[fixed_columns]

new_Delhi_onehot.head()
```

Out[75]:

	Locality	ATM	Accessories Store	Afghan Restaurant	African Restaurant	Airport Lounge	Airport Service	Airport Terminal	American Restaurant	Arc
0	ARSS Mall, Paschim Vihar	0	0	0	0	0	0	0	0	
1	ARSS Mall, Paschim Vihar	0	0	0	0	0	0	0	0	
2	ARSS Mall, Paschim Vihar	0	0	0	0	0	0	0	0	
3	ARSS Mall, Paschim Vihar	0	0	0	0	0	0	0	0	
4	ARSS Mall, Paschim Vihar	0	0	0	0	0	0	0	0	

```
In [ ]: new_Delhi_onehot.shape
```

Out[76]: (3192, 208)

**Next, we will group rows by neighborhood and by taking the mean of the frequency of occurrence of each category**

```
In [ ]: new_Delhi_grouped = new_Delhi_onehot.groupby('Locality').mean().reset_index
new_Delhi_grouped
```

Out[77]:

	Locality	ATM	Accessories Store	Afghan Restaurant	African Restaurant	Airport Lounge	Airport Service	Airport Terminal	A
0	ARSS Mall, Paschim Vihar	0.111111	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	(
1	Adchini	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	(
2	Aditya Mega Mall, Karkardooma	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	(
3	Aerocity	0.000000	0.000000	0.000000	0.000000	0.142857	0.000000	0.142857	(
4	Aggarwal City Mall, Pitampura	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	(
5	Aggarwal City Plaza Rohini	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	(

**new size**

```
In [ ]: new_Delhi_grouped.shape
```

Out[81]: (239, 208)

```
In [ ]: ## print each Locality along with the top 5 most common venues

num_top_venues = 5

for hood in new_Delhi_grouped['Locality']:
    print("----"+hood+"----")
    temp = new_Delhi_grouped[new_Delhi_grouped['Locality'] == hood].T.reset
    temp.columns = ['venue', 'freq']
    temp = temp.iloc[1:]
    temp['freq'] = temp['freq'].astype(float)
    temp = temp.round({'freq': 2})
    print(temp.sort_values('freq', ascending=False).reset_index(drop=True).
    print('\n')
```

```
----ARSS Mall, Paschim Vihar----
              venue  freq
0      Indian Restaurant  0.22
1                      ATM  0.11
2      Coffee Shop      0.11
3  Multicuisine Indian Restaurant  0.11
4      Sandwich Place    0.11
```

```
----Adchini----
              venue  freq
0  Indian Restaurant  0.33
1          Café      0.33
2          Pub       0.17
3  Parsi Restaurant  0.17
4          ATM       0.00
```

```
----Aditya Mega Mall, Karkardooma----
```

### Function to sort venues in descending order

```
In [ ]: def return_most_common_venues(row, num_top_venues):
    row_categories = row.iloc[1:]
    row_categories_sorted = row_categories.sort_values(ascending=False)

    return row_categories_sorted.index.values[0:num_top_venues]
```

### Creating data frame to display top 10 venues for each neighborhood

```

In [ ]: num_top_venues = 10

indicators = ['st', 'nd', 'rd']

# create columns according to number of top venues
columns = ['Locality']
for ind in np.arange(num_top_venues):
    try:
        columns.append('{}{} Most Common Venue'.format(ind+1, indicators[ind]))
    except:
        columns.append('{}th Most Common Venue'.format(ind+1))

# create a new dataframe
neighborhoods_venues_sorted = pd.DataFrame(columns=columns)
neighborhoods_venues_sorted['Locality'] = new_Delhi_grouped['Locality']

for ind in np.arange(new_Delhi_grouped.shape[0]):
    neighborhoods_venues_sorted.iloc[ind, 1:] = return_most_common_venues(n

neighborhoods_venues_sorted.head()

```

Out[94]:

	Locality	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
0	ARSS Mall, Paschim Vihar	Indian Restaurant	ATM	Market	Pizza Place	Multicuisine Indian Restaurant	Coffee Shop	Sandwich Place	Refrigerator
1	Adchini	Indian Restaurant	Café	Parsi Restaurant	Pub	Farmers Market	Fried Chicken Joint	French Restaurant	
2	Aditya Mega Mall, Karkardooma	Indian Restaurant	Shopping Mall	Pizza Place	Multiplex	Hotel	Café	Fast Food Restaurant	Refrigerator
3	Aerocity	Hotel	Fast Food Restaurant	Airport Lounge	Airport Terminal	Coffee Shop	Rental Car Location	Women's Store	
4	Aggarwal City Mall, Pitampura	Department Store	Pizza Place	Chinese Restaurant	Fast Food Restaurant	Furniture / Home Store	Frozen Yogurt Shop	Fried Chicken Joint	Refrigerator

## 4. Cluster Neighborhoods

```
In [ ]: ## Cluster Locality  
## Run k-means to cluster the Locality into 5 clusters.  
  
# set number of clusters  
kclusters = 5  
  
new_Delhi_clustering = new_Delhi_grouped.drop('Locality', 1)  
  
# run k-means clustering  
kmeans = KMeans(n_clusters=kclusters, random_state=0).fit(new_Delhi_cluster  
  
# check cluster labels generated for each row in the dataframe  
kmeans.labels_[0:10]  
kmeans.labels_.shape
```

```
Out[95]: (239,)
```



```
In [ ]: # add clustering labels
New_Delhi_merged = df_final.head(239)
New_Delhi_merged['Cluster Labels'] = kmeans.labels_

# merge New_Delhi_grouped with df_Chinese to add latitude/longitude for each
New_Delhi_merged = New_Delhi_merged.join(neighborhoods_venues_sorted.set_index('id', inplace=True))

New_Delhi_merged.head()
```

<ipython-input-96-f2cba4c297e6>:3: SettingWithCopyWarning:  
A value is trying to be set on a copy of a slice from a DataFrame.  
Try using .loc[row\_indexer,col\_indexer] = value instead

See the caveats in the documentation: [https://pandas.pydata.org/pandas-docs/stable/user\\_guide/indexing.html#returning-a-view-versus-a-copy](https://pandas.pydata.org/pandas-docs/stable/user_guide/indexing.html#returning-a-view-versus-a-copy) ([https://pandas.pydata.org/pandas-docs/stable/user\\_guide/indexing.html#returning-a-view-versus-a-copy](https://pandas.pydata.org/pandas-docs/stable/user_guide/indexing.html#returning-a-view-versus-a-copy))

```
New_Delhi_merged['Cluster Labels'] = kmeans.labels_
```

Out[96]:

	Locality	Lat	Lng	No_of_Restaurant	Cusines	Agg_Rating	Comments	No.
0	ARSS Mall, Paschim Vihar	28.668945	77.101544	1	North Indian, South Indian, Chinese, Mithai, F...	3.100000	Average	
1	Adchini	28.537063	77.197808	13	Fast Food, North Indian, Seafood, Continental,...	3.292308	Average, Good, Poor, Very Good	
2	Aditya Mega Mall, Karkardooma	28.656131	77.301266	4	Finger Food, North Indian, Mughlai, Pizza, Fas...	3.275000	Average, Good	
3	Aerocity	28.553077	77.104270	2	Fast Food, Italian, Pizza, North Indian, Conti...	3.200000	Average	
4	Aggarwal City Mall, Pitampura	28.690020	77.134650	3	North Indian, Chinese, Street Food, Mithai, No...	3.033333	Average	

```

In [ ]: create map
clusters = folium.Map(location=[latitude, longitude], zoom_start=11)

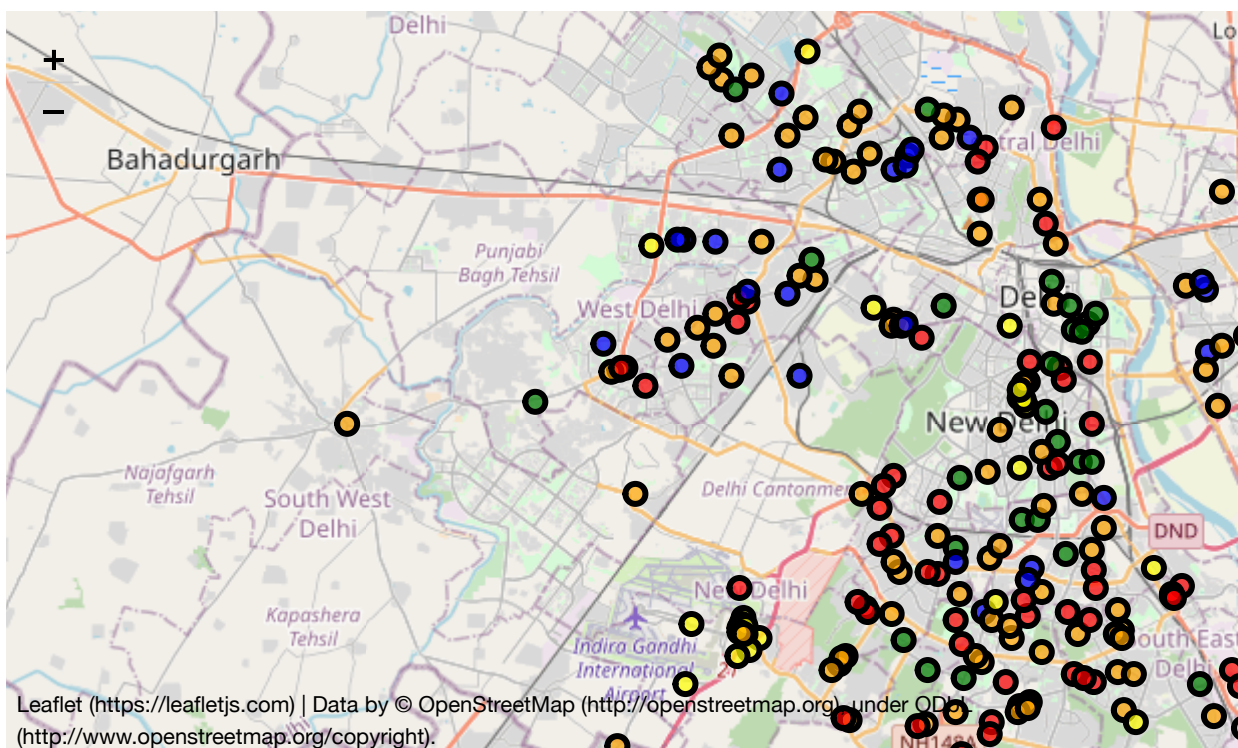
# create color scheme for the clusters
np.arange(kclusters)
[i + x + (i*x)**2 for i in range(kclusters)]
colors_array = cm.rainbow(np.linspace(0, 1, len(ys)))
rainbow = [colors.rgb2hex(i) for i in colors_array]
colors = ['red', 'green', 'blue', 'yellow', 'orange']

# add markers to the map
markers_colors = []
for lat, lon, poi, cluster in zip(New_Delhi_merged['Lat'], New_Delhi_merged['Lng'],
label = folium.Popup(str(poi) + ' Cluster ' + str(cluster), parse_html=True)
folium.CircleMarker(
    [lat, lon],
    radius=5,
    popup=label,
    color='black',
    fill=True,
    fill_color=colors[cluster],
    fill_opacity=0.7).add_to(map_clusters)

map_clusters

```

Out[101]:



## 5. Examine Clusters

### Cluster 1

```
In [ ]: New_Delhi_merged.loc[New_Delhi_merged['Cluster Labels'] == 0, New_Delhi_merged
```

```
Out[102]:
```

	Lat	Agg_Rating	Comments	No_of_Votes	Cluster Labels	1st Most Common Venue	2nd Most Common Venue	3rd Most Common Venue
<b>8</b>	28.555599	3.800000	Average, Very Good	2411	0	Café	Stadium	Dessert Shop
<b>9</b>	28.659870	3.281481	Average, Good	1309	0	Café	ATM	Convenience Store
<b>16</b>	28.630440	3.208333	Average, Good, Poor	1779	0	Bakery	Café	Indian Restaurant
<b>17</b>	28.557846	3.220000	Average, Good, Poor	1875	0	Café	Fast Food Restaurant	Pizza Place
<b>20</b>	28.568193	2.755556	Average, Poor	339	0	Lounge	Asian Restaurant	Café
<b>21</b>	28.568707	2.685714	Average,	1017	0	Garden	Bar	African Restaurant

## Cluster 2

```
In [ ]: New_Delhi_merged.loc[New_Delhi_merged['Cluster Labels'] == 1, New_Delhi_merged
```

```
Out[103]:
```

	Lat	Agg_Rating	Comments	No_of_Votes	Cluster Labels	1st Most Common Venue	2nd Most Common Venue	3rd Most Common Venue
1	28.537063	3.292308	Average, Good, Poor, Very Good	1560	1	Indian Restaurant	Café	Parsi Restaurant
12	28.641698	3.050000	Average	17	1	Hostel	Indian Restaurant	Movie Theater
23	28.656414	3.358182	Average, Good, Very Good	6668	1	Indian Restaurant	Snack Place	Dessert Shop
42	28.576161	2.983333	Average, Good, Poor	542	1	Indian Restaurant	Fast Food Restaurant	Dessert Shop
43	28.643320	3.287500	Average, Good, Poor, Very Good	2570	1	Indian Restaurant	Restaurant	Hotel
44	28.574288	3.496154	Average, Good, Poor, Very Good	15579	1	Indian Restaurant	Italian Restaurant	Coffee Shop
47	28.681582	3.046875	Average, Good, Poor	1224	1	Indian Restaurant	Afghan Restaurant	Gym
49	28.602169	2.900000	Average	7	1	Indian Restaurant	Golf Course	Italian Restaurant
55	28.617102	4.200000	Very Good	3010	1	Indian Restaurant	Women's Store	Fast Food Restaurant
71	28.641003	3.650000	Average, Good	367	1	Indian Restaurant	Hostel	Hotel
82	28.548371	3.137500	Average, Good, Poor	609	1	Café	Indian Restaurant	Shopping Mall
85	28.648885	3.561111	Average, Good, Very Good	7411	1	Indian Restaurant	Flea Market	Mosque
92	28.584294	3.100000	Average	35	1	Indian Restaurant	Bakery	Metro Station
97	28.662972	3.533333	Average, Very Good	293	1	Indian Restaurant	Fried Chicken Joint	Moving Target
99	28.648795	3.238028	Average, Good, Poor, Very Good	7282	1	Indian Restaurant	Fast Food Restaurant	Coffee Shop

	Lat	Agg_Rating	Comments	No_of_Votes	Cluster Labels	1st Most Common Venue	2nd Most Common Venue	3rd Most Common Venue
<b>113</b>	28.541242	2.984615	Average	817	1	Indian Restaurant	Bakery	Pizza Place
<b>114</b>	28.584615	3.303571	Average, Good, Poor, Very Good	6480	1	Indian Restaurant	Asian Restaurant	Deli / Bodega
<b>116</b>	28.486576	2.950000	Average	8	1	Indian Restaurant	Furniture / Home Store	Art Museum
<b>118</b>	28.646549	3.200000	Average	20	1	Indian Restaurant	Restaurant	Road
<b>133</b>	28.708185	3.344444	Average, Good, Very Good	568	1	Indian Restaurant	Park	Chinese Restaurant
<b>146</b>	28.535040	3.175000	Average, Good	60	1	Indian Restaurant	Bakery	Women's Store
<b>153</b>	28.608055	3.860000	Average, Good, Very Good	7978	1	Indian Restaurant	Ice Cream Shop	Smoke Shop
<b>165</b>	28.539464	2.900000	Average	14	1	Indian Restaurant	Restaurant	Hotel
<b>173</b>	28.714176	3.352083	Average, Good, Poor, Very Good	4514	1	Department Store	Fast Food Restaurant	Indian Restaurant
<b>199</b>	28.601779	3.700000	Average, Good, Very Good	964	1	Indian Restaurant	Hotel	Italian Restaurant
<b>204</b>	28.597038	3.533333	Average, Good, Very Good	902	1	Nightclub	Indian Restaurant	Shopping Mall
<b>209</b>	28.631362	3.537500	Average, Good	1666	1	Indian Restaurant	Bakery	Hotel Bar
<b>223</b>	28.619954	2.905714	Average, Poor	242	1	ATM	Indian Restaurant	Mobile Phone Shop

### Cluster 3

```
In [ ]: New_Delhi_merged.loc[New_Delhi_merged['Cluster Labels'] == 2, New_Delhi_merged
```

```
Out[104]:
```

	Lat	Agg_Rating	Comments	No_of_Votes	Cluster Labels	1st Most Common Venue	2nd Most Common Venue	3rd Most Common Venue
0	28.668945	3.100000	Average	117	2	Indian Restaurant	ATM	Market
2	28.656131	3.275000	Average, Good	434	2	Indian Restaurant	Shopping Mall	Pizza Place
4	28.690020	3.033333	Average	126	2	Department Store	Pizza Place	Chinese Restaurant
13	28.690182	3.090000	Average, Good, Poor	971	2	Dessert Shop	Smoke Shop	BBQ Joint
14	28.693969	3.378571	Average, Good	2343	2	Indian Restaurant	South Indian Restaurant	Pizza Place
15	28.691136	3.390909	Average, Good, Very Good	485	2	Indian Restaurant	Pizza Place	South Indian Restaurant
18	28.696302	3.300000	Average, Good	415	2	Indian Restaurant	South Indian Restaurant	Pizza Place
22	28.654105	3.085714	Average	80	2	Pizza Place	Gym	Gaming Cafe
33	28.657388	3.320000	Average, Good	636	2	Pizza Place	Indian Restaurant	Hotel
57	28.699837	3.376471	Average, Good, Very Good	2460	2	Pizza Place	Food Truck	Department Store
61	28.668181	3.400000	Average	33	2	Indian Restaurant	Pizza Place	Ice Cream Shop
65	28.556980	3.272414	Average, Good, Poor	5243	2	Indian Restaurant	Coffee Shop	Donut Shop
77	28.631728	3.060714	Average, Good, Poor, Very Good	1650	2	Indian Restaurant	Dessert Shop	Café
84	28.630942	3.122000	Average, Good, Poor, Very Good	1738	2	Indian Restaurant	Fried Chicken Joint	Pizza Place
98	28.651527	3.490909	Average, Good, Very Good	3222	2	Indian Restaurant	Pizza Place	Plaza

	Lat	Agg_Rating	Comments	No_of_Votes	Cluster Labels	1st Most Common Venue	2nd Most Common Venue	3rd Most Common Venue
103	28.652501	3.080000	Average, Good, Poor	2196	2	Indian Restaurant	Pizza Place	Convenience Store
104	28.656263	3.097619	Average, Good	965	2	Pizza Place	Accessories Store	Park
110	28.635391	3.040385	Average, Good, Poor, Very Good	2442	2	Indian Restaurant	Pizza Place	Convenience Store
126	28.602860	2.973214	Average, Good, Poor	1831	2	Bistro	Indian Restaurant	Pizza Place
140	28.627959	3.004167	Average	326	2	Indian Restaurant	Café	Snack Place
144	28.591092	3.200000	Average, Good, Poor	993	2	Indian Restaurant	Garden	Monument / Landmark
155	28.669160	3.148214	Average, Good, Poor	2735	2	Market	Indian Restaurant	Pizza Place
156	28.630734	3.185714	Average, Good	244	2	Indian Restaurant	Market	Pizza Place
160	28.712975	3.066667	Average	121	2	Indian Restaurant	Pizza Place	Multiplex
170	28.643757	3.575000	Average, Good	2879	2	Pizza Place	Fast Food Restaurant	Indian Restaurant
181	28.571849	3.316667	Average, Good	472	2	Fast Food Restaurant	Indian Restaurant	Snack Place
193	28.569983	3.133333	Average, Good, Poor	988	2	Coffee Shop	Indian Restaurant	Pizza Place
194	28.566717	3.303125	Average, Good, Very Good	2411	2	Coffee Shop	Market	Indian Restaurant
210	28.661133	3.775000	Good, Very Good	460	2	Indian Restaurant	Café	Pizza Place
231	28.637492	3.062264	Average, Good, Poor	1253	2	Indian Restaurant	Pizza Place	Beer Garden
235	28.652978	3.500000	Good	178	2	Fast Food Restaurant	Café	Pizza Place

#### Cluster 4



```
In [ ]: New_Delhi_merged.loc[New_Delhi_merged['Cluster Labels'] == 3, New_Delhi_merged
```

```
Out[105]:
```

	Lat	Agg_Rating	Comments	No_of_Votes	Cluster Labels	1st Most Common Venue	2nd Most Common Venue	3rd Most Common Venue
3	28.553077	3.200000	Average	59	3	Hotel	Fast Food Restaurant	Airport Lounge
10	28.554807	3.000000	Average	4	3	Hotel	Coffee Shop	Restaurant
19	28.548827	2.700000	Average	9	3	Hotel	Indian Restaurant	Accessories Store
35	28.590297	3.350000	Average, Good	127	3	Hotel	Indian Restaurant	Liquor Store
56	28.570246	4.050000	Average, Excellent	1949	3	Hotel	Sculpture Garden	Gym
60	28.665829	4.000000	Very Good	756	3	Hotel	Indian Restaurant	Cafe
69	28.550327	3.300000	Average	30	3	Hotel	Bed & Breakfast	Hotel Bar
70	28.590654	3.550000	Average, Good	302	3	Hotel	Restaurant	Light Rail Station
74	28.621267	3.416667	Average, Good	475	3	Hotel	Indian Restaurant	Italian Restaurant
83	28.552895	3.300000	Average	15	3	Hotel	Restaurant	Coffee Shop
111	28.618772	3.666667	Average, Good, Very Good	725	3	Hotel	Indian Restaurant	Restaurant
112	28.552036	3.500000	Good	26	3	Hotel	Restaurant	Coffee Shop
119	28.545033	2.960000	Average, Good	259	3	Hotel	Spa	Accessories Store
128	28.608252	2.945000	Average, Good	406	3	Fried Chicken Joint	Hotel	Women's Store
149	28.642910	2.996552	Average, Good, Poor, Very Good	2850	3	Hotel	Indian Restaurant	Cafe
162	28.725814	3.050000	Average	50	3	Hotel	Café	Women's Store
163	28.552711	3.100000	Average	7	3	Hotel	Hotel Bar	Restaurant

	Lat	Agg_Rating	Comments	No_of_Votes	Cluster Labels	1st Most Common Venue	2nd Most Common Venue	3rd Most Common Venue
<b>168</b>	28.543814	3.950000	Average, Excellent, Good, Very Good	1225	3	Hotel	Indian Restaurant	Hotel Ba
<b>169</b>	28.667261	3.200000	Average	76	3	Hotel	Indian Restaurant	Food Trucl
<b>174</b>	28.550802	3.250000	Average, Good	34	3	Hotel	Restaurant	Coffe Shop
<b>189</b>	28.620715	3.880000	Average, Good, Very Good	626	3	Hotel	Indian Restaurant	Italian Restaurant
<b>197</b>	28.592724	2.985714	Average	560	3	Hotel	Restaurant	Liquor Store
<b>205</b>	28.600171	3.714286	Average, Good, Very Good	1733	3	Hotel	Mediterranean Restaurant	Breakfast Spot
<b>213</b>	28.623619	3.466667	Average, Good	1141	3	Hotel	Italian Restaurant	Coffe Shop
<b>217</b>	28.535183	3.066667	Average	37	3	Hotel	Indian Restaurant	Women's Store
<b>220</b>	28.523519	3.700000	Good	281	3	Hotel	Steakhouse	Restaurant
<b>236</b>	28.648177	3.675000	Average, Good, Very Good	876	3	Indian Restaurant	Pizza Place	Parl
<b>238</b>	28.559928	3.075000	Average, Good	923	3	Indian Restaurant	Coffee Shop	Ba

## Cluster 5

```
In [ ]: New_Delhi_merged.loc[New_Delhi_merged['Cluster Labels'] == 4, New_Delhi_mer
Out[106]:
```

	Lat	Agg_Rating	Comments	No_of_Votes	Cluster Labels	1st Most Common Venue	2nd Most Common Venue	3rd Most Common Venue
5	28.700516	3.040000	Average, Good, Poor	594	4	Gym / Fitness Center	Shopping Mall	Pizza Place
6	28.527088	3.117391	Average, Good, Poor	1012	4	Coffee Shop	Steakhouse	Food & Drink
7	28.541298	3.425000	Average, Good, Very Good	2460	4	Coffee Shop	Fast Food Restaurant	Shopping Mall
11	28.562580	3.750000	Average, Good, Very Good	875	4	Fast Food Restaurant	Shopping Mall	Japanese Restaurant
24	28.649658	3.800000	Average, Excellent, Good, Very	1410	4	Indian Restaurant	Light Rail Station	Fast Food Restaurant