# IBM APPLIED DATA SCIENCE CAPSTONE

## *OPENING A CAFÉ IN DELHI, INDIA*

BY: HRITHIK SHUKLA

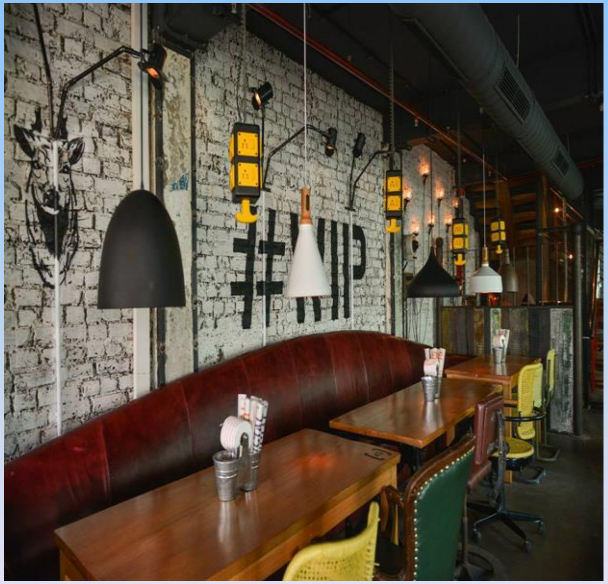
JUNE 2020

# INTRODUCTION:

Delhi is one of the major cities of India and it has one of the most vibrant varieties of café in the world. This city is famous for its food and foodies, you can say food is really important to the people of Delhi. This means the café business flourishes in this city, making it one of the most attractive destinations for restaurant franchises and entrepreneurs.

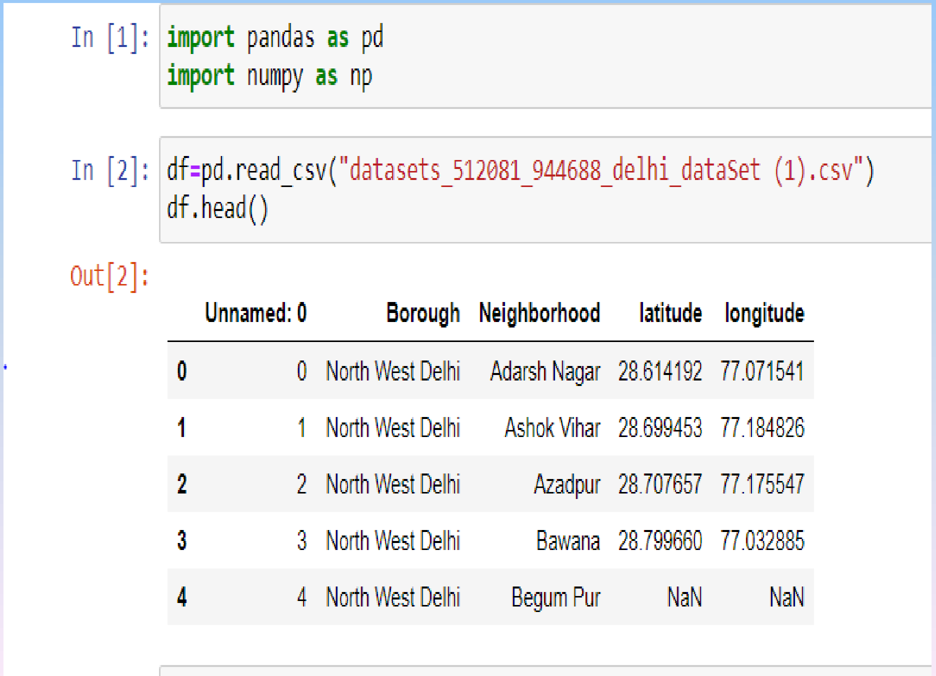
# BUSINESS PROBLEM:

- Delhi is a big city and has a ton of options when it comes to café places. For anyone new to the café business it will be difficult survive as there is a lot of competition in this market.
- If an entrepreneur wants to open a café in Delhi at what places should he/she open his/her business so that they face the least competition. There are lots of neighborhoods in Delhi we will have to conduct analysis to see which area can be considered for opening a café. We can leverage the foursquare data by calling data of all the café situated in the specific neighborhoods.



# METHODOLOGY:

- **Importing neighborhood data and performing data wrangling:**
- **Latitude and longitude of neighborhoods:**

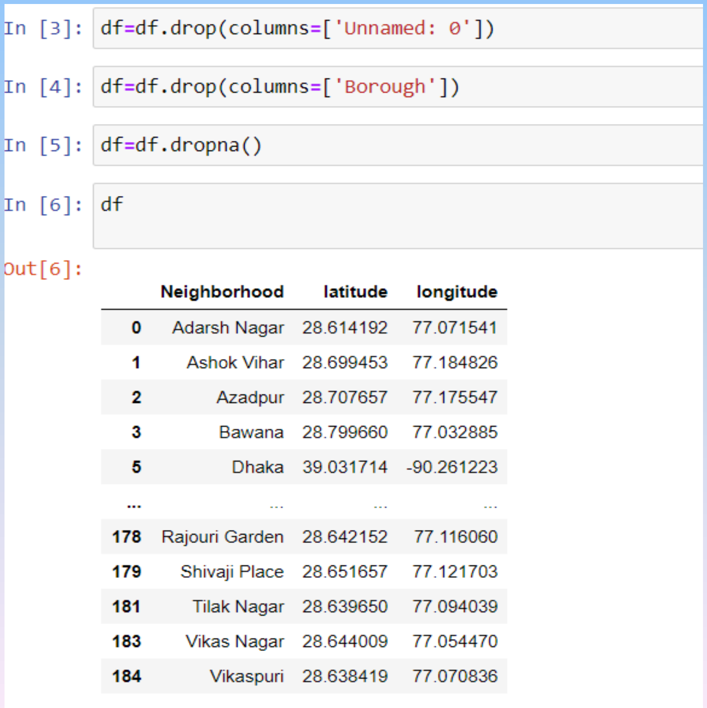


```
In [1]: import pandas as pd
        import numpy as np
```

```
In [2]: df=pd.read_csv("datasets_512081_944688_delhi_dataSet (1).csv")
        df.head()
```

Out[2]:

| | Unnamed: 0 | Borough | Neighborhood | latitude | longitude |
|---|---|---|---|---|---|
| 0 | 0 | North West Delhi | Adarsh Nagar | 28.614192 | 77.071541 |
| 1 | 1 | North West Delhi | Ashok Vihar | 28.699453 | 77.184826 |
| 2 | 2 | North West Delhi | Azadpur | 28.707657 | 77.175547 |
| 3 | 3 | North West Delhi | Bawana | 28.799660 | 77.032885 |
| 4 | 4 | North West Delhi | Begum Pur | NaN | NaN |

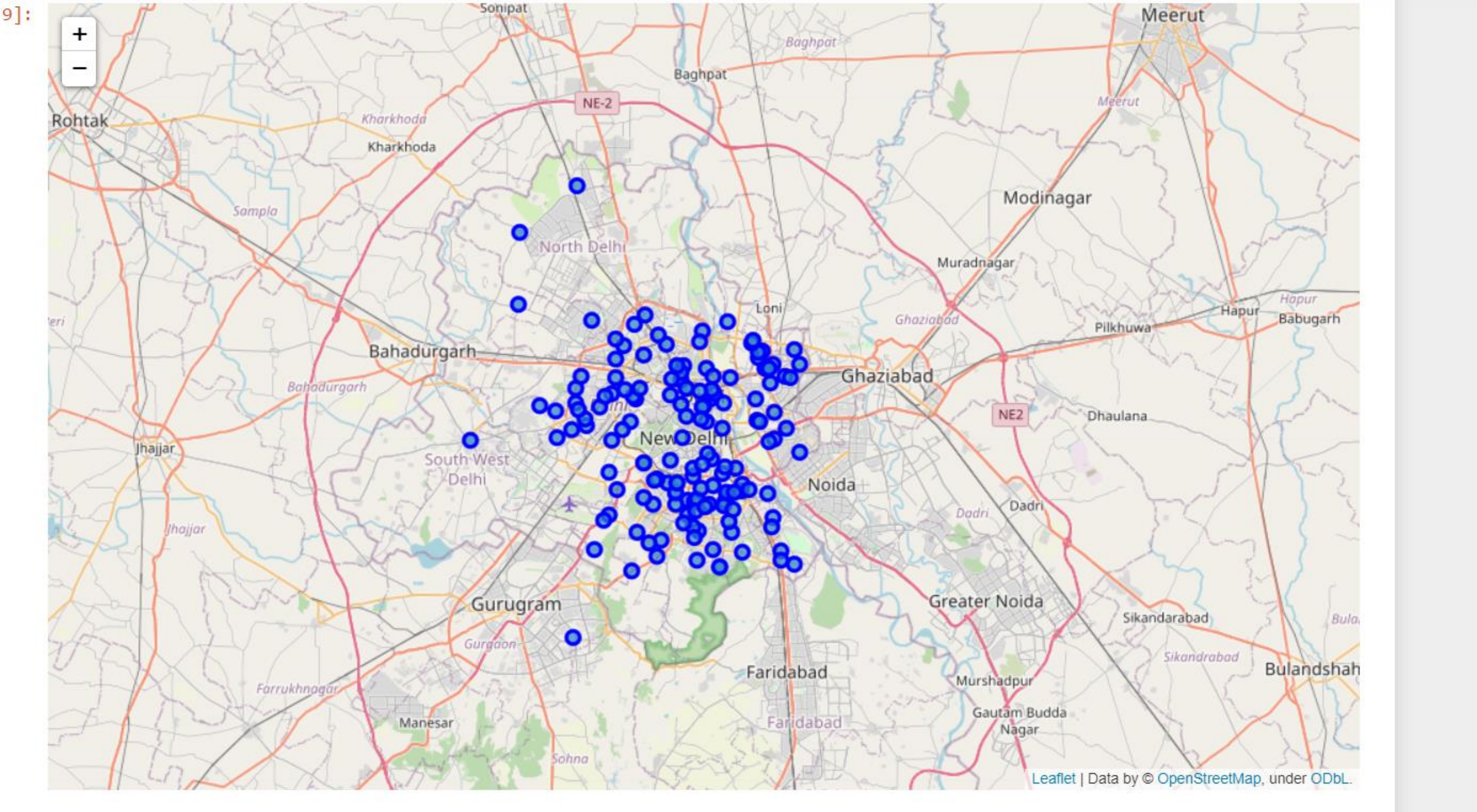


```
In [3]: df=df.drop(columns=['Unnamed: 0'])

In [4]: df=df.drop(columns=['Borough'])

In [5]: df=df.dropna()

In [6]: df
```

Out[6]:

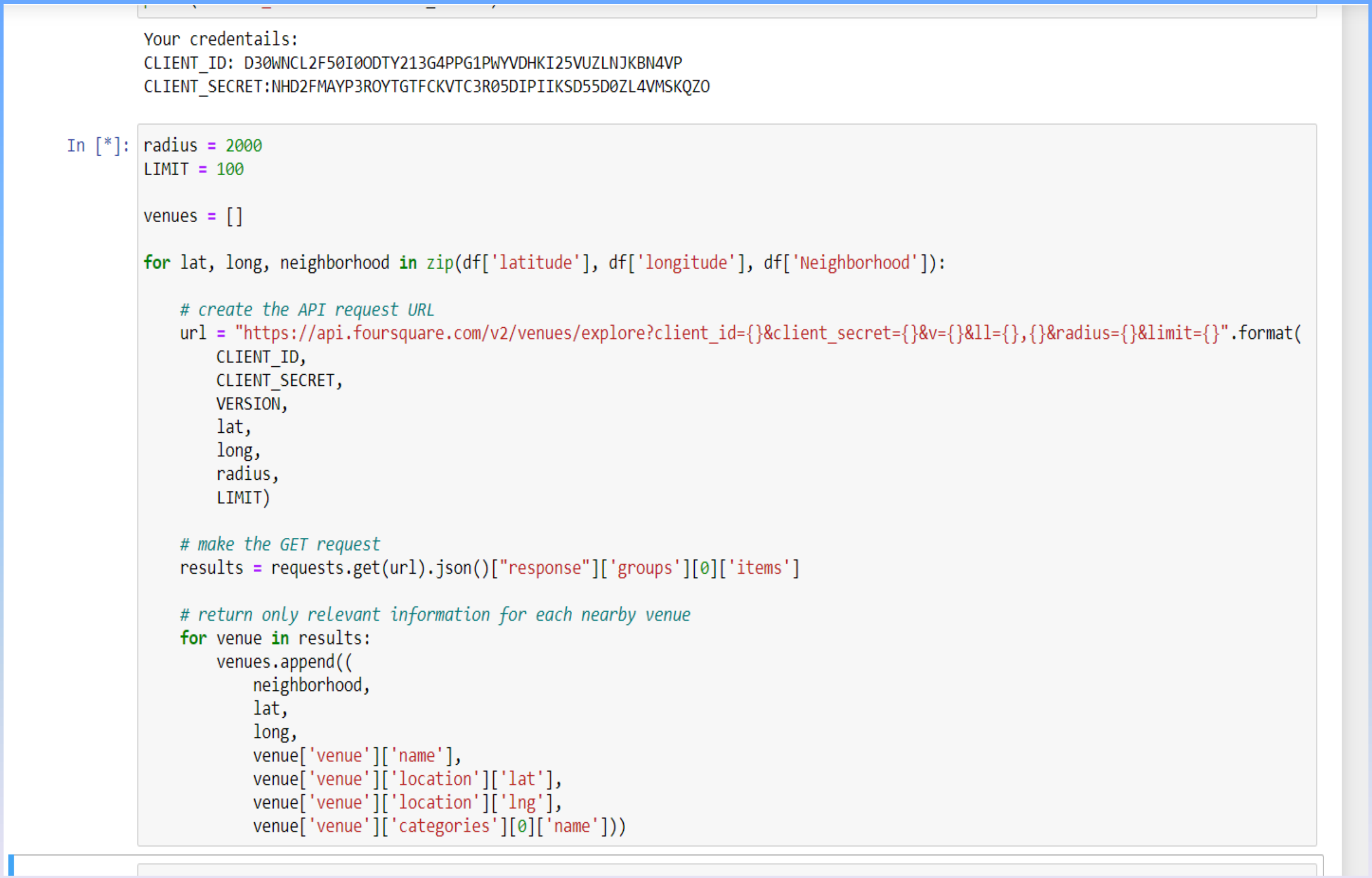
| | Neighborhood | latitude | longitude |
|---|---|---|---|
| 0 | Adarsh Nagar | 28.614192 | 77.071541 |
| 1 | Ashok Vihar | 28.699453 | 77.184826 |
| 2 | Azadpur | 28.707657 | 77.175547 |
| 3 | Bawana | 28.799660 | 77.032885 |
| 5 | Dhaka | 39.031714 | -90.261223 |
| ... | ... | ... | ... |
| 178 | Rajouri Garden | 28.642152 | 77.116060 |
| 179 | Shivaji Place | 28.651657 | 77.121703 |
| 181 | Tilak Nagar | 28.639650 | 77.094039 |
| 183 | Vikas Nagar | 28.644009 | 77.054470 |
| 184 | Vikaspuri | 28.638419 | 77.070836 |

# Visualizing the data:



Leaflet | Data by © OpenStreetMap, under ODbL.

# Calling foursquare API to get venue data:



```
Your credentails:
CLIENT_ID: D30WNCL2F50I0ODTY213G4PPG1PWYVDHKI25VUZLNJKBN4VP
CLIENT_SECRET:NHD2FMAYP3ROYTGTFCKVTC3R05DIPIIKSD55D0ZL4VMSKQZO
```

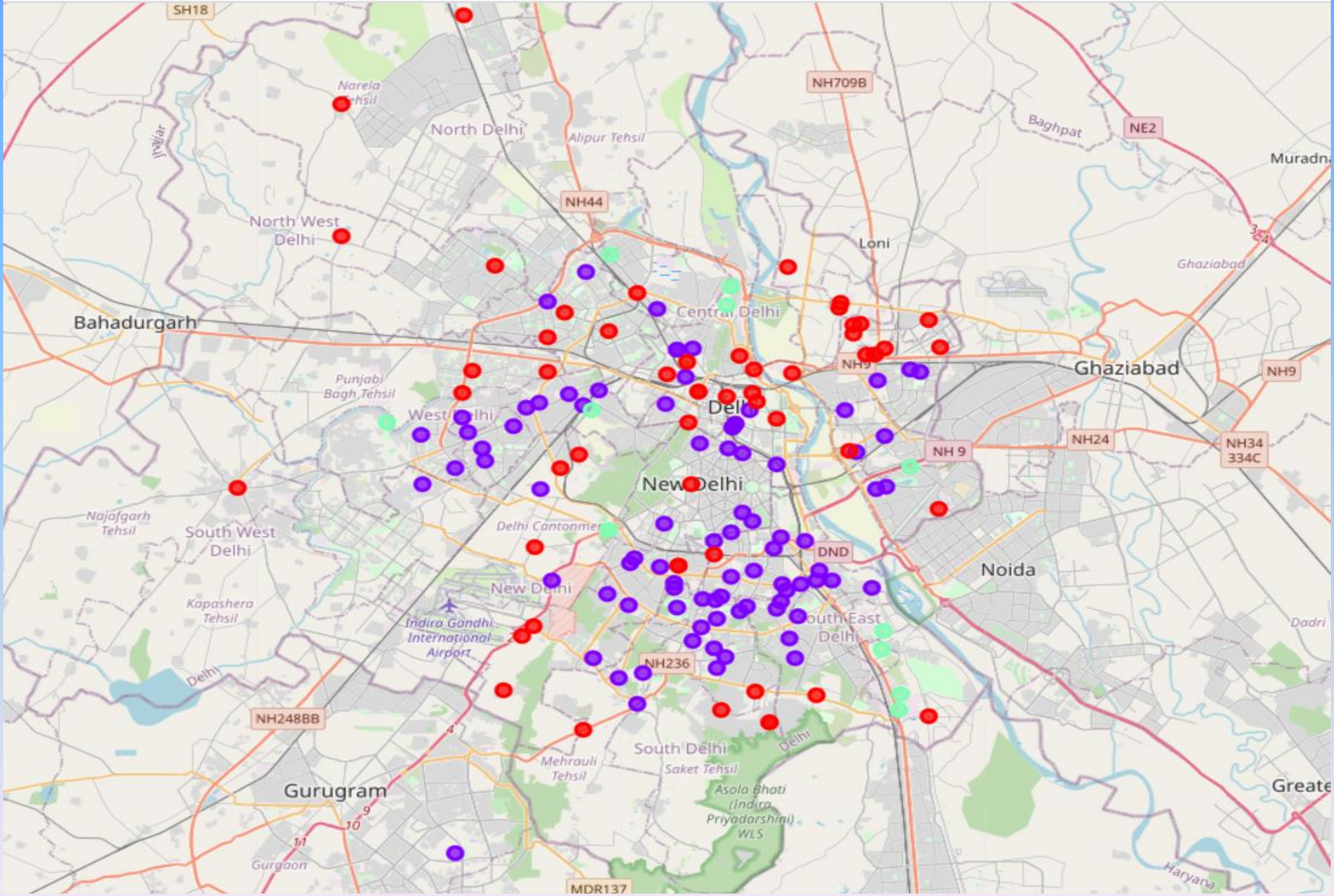
```
In [*]: radius = 2000
LIMIT = 100

venues = []

for lat, long, neighborhood in zip(df['latitude'], df['longitude'], df['Neighborhood']):

    # create the API request URL
    url = "https://api.foursquare.com/v2/venues/explore?client_id={}&client_secret={}&v={}&ll={},{}&radius={}&limit={}".format(
        CLIENT_ID,
        CLIENT_SECRET,
        VERSION,
        lat,
        long,
        radius,
        LIMIT)

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

    # return only relevant information for each nearby venue
    for venue in results:
        venues.append((
            neighborhood,
            lat,
            long,
            venue['venue']['name'],
            venue['venue']['location']['lat'],
            venue['venue']['location']['lng'],
            venue['venue']['categories'][0]['name']))
```

## Analyzing the venue data and preparing data for clustering:

- The foursquare API gives us various categories of venues pertaining to certain neighborhoods. We can count the number of venues in each category. We can check how many venues were returned for each neighborhood. Then we group the rows according to the neighborhoods and take mean of the frequency of occurrence of each venue category. We also perform one hot encoding of the data. Now as the scope of this project is limited to cafés so we filter the café data as venue category for neighborhoods. By doing all this we are actually preparing data for clustering.

# Performing clustering:

Below map shows the 3 clusters. Cluster 0 is purple, Cluster 1 is mint green and Cluster 2 is red.



# Result:

The neighborhoods were clustered using the K-Means clustering algorithm into 3 clusters according to the frequency of occurrence of “café” in the respective neighborhoods.

- After examining the clusters:
- **Cluster 0:** The number of cafés in these neighborhoods is moderate. It has a moderate level of competition.
- **Cluster 1:** It can be seen that this cluster of neighborhoods has the highest no. of cafés. Hence, the completion in these neighborhoods will be the highest amongst the 3 clusters.
- **Cluster 2:** It can be seen that in this cluster the number of cafés in these neighborhoods is low. Hence, It has low level of competition.

# CONCLUSION:

During this project, we have identified the business problem clearly, the data required was specified also extraction and cleaning of data was carried out, performing machine learning by clustering the data into 3 clusters, on the basis of their similar features. Then appropriate recommendations were provided to the target audience of the project, which were café franchises and entrepreneurs who want to open a café. After the analysis it was found that cluster 2 is the best suited cluster for opening new café, the neighborhoods in this cluster have a very low level of café concentration therefore providing low competition to a new café. These findings will help our target audience to make a knowledgeable decision on how they can avoid high competition and choose a place with low to moderate level of competition.

# References:

- https://www.kaggle.com/shaswatd673/delhi-neighborhood-data
- https://foursquare.com/developers/apps
- https://traveltriangle.com/blog/best-cafes-in-delhi/
- https://www.scoopwhoop.com/27-instagrammable-cafes-in-delhi/

# THANK YOU