

Capstone Project - The Battle of Neighborhoods

By Prahlad

May 25, 2020

1. Introduction

1.1 Background

This project is sponsored by one of the big restaurant chains from Western India. They are interested in opening a multi-cuisine Indian restaurant in Delhi. Accordingly, this report has been prepared to evaluate the best neighborhood to open a new restaurant in one of the Delhi's localities. The restaurant should be in the neighborhood with good footfall and high business potential. This report will use the Foursquare API in order to analyse the data about the locations of existing restaurants in various neighborhoods and find out about trending venues and neighbourhoods that might suggest popular neighbourhood to open a new Indian restaurant.

Delhi is the capital city of India. The population of Delhi is around 20 million. The town is divided into nine districts (Boroughs). Delhi is a historical town. The area around Delhi was probably inhabited before the second millennium BCE and there is evidence of continuous inhabitation since at least the 6th century BCE. The city is believed to be the site of [Indraprastha](#), the legendary capital of the [Pandava's](#) in the Indian epic *Mahabharata*. The earliest architectural relics date back to the [Mauryan](#) period (c. 300 BCE). Delhi is a favourite destination for domestic (25 million tourists per year) and foreign tourists (5 million per year).

Delhi and New Delhi are used interchangeably to refer to the NCT- National Capital Territory of Delhi. The NCR- National Capital Region is a much larger entity comprising the entire NCT along with adjoining districts in neighbouring states of UP, Haryana and Rajasthan.

With Delhi's diverse culture, come the diverse food joints. Delhi is the birthplace of the hugely popular Mughlai cuisine. This is the style of cooking prevalent during the Mughal era and is defined by the use of whole and ground spices. Some of the signature Mughlai dishes include *kebabs*, *koftas*, *pilafs*, and *biryani*. Delhi is also famous for its street food which includes *parathas*, *chaats*, and *kulfi*.

While India has always been a food-loving country, with each region, having its own special cuisine, Indians have never been very big on eating out. But all that is changing now. The restaurant industry in India has been growing at a rapid pace over the last decade or so and the growth story is set to continue in the near foreseeable future. Delhi is the leader in this rapid growth with variety of choices for food lovers from across the globe. Apart from Indian

cuisines, international chains like McDonalds, Domino's Pizza Hut, KFC, *Burger King*, *Dunkin Donuts*, *Taco Bell*, *Krispy Kreme*, cafes like *Starbucks* and *Costa Coffee* have set up shop in Delhi.

1.2 Map of Districts of Delhi(Equivalent to Boroughs)



1.3 Problem Definition

One of our clients from Mumbai, who runs a big chain of Indian restaurants, is interested in opening a branch in Delhi. He has signed a contract with our firm to prepare a project report to suggest the best neighborhood to open the new Indian restaurant in Delhi.

For this project we will be using the descriptive and predictive capability of the much hyped data science tools and geospatial coordinate analysis with the help of Foursquare API.

The Project will focus on the following points to arrive at final solution:

- a) Analysis of trending venues across neighbourhoods of Delhi

- b) Neighborhoods already having large number of Indian restaurants
- c) Neighbourhoods deficient in supply of Indian restaurants - the demand supply gap
- d) Neighbourhoods overcrowded with multi-category restaurants including Indian restaurants - excess supply over demand.
- e) Finally Zero on to the best neighborhoods in Delhi for opening a new Indian Restaurant based on demand supply gap and closeness to popular venues.

1.4 Target Audience

One of the big food chains from Mumbai, which runs a chain of Indian Restaurants in Mumbai, and other towns of Western India, is the key stake holder sponsoring this project.

Other business people interested in setting up a new restaurant in Delhi may also be interested in the report.

2. Data Acquisition and Cleaning

2.1 Data Sources

Based on the preliminary investigations the raw data from the following sources will be collected for this project:

i) Delhi Wikipedia page to extract the districts (Boroughs) and Neighborhood data
https://en.wikipedia.org/wiki/List_of_districts_of_Delhi

ii) Data from Kaggle – Neighborhoods of Delhi
<https://www.kaggle.com/shaswatd673/delhi-neighborhood-data>

Restaurants of Delhi
<https://www.kaggle.com/shrutimehta/zomato-restaurants-data>

iii) Data from Foursquare : <https://developer.foursquare.com/>

Foursquare APIs will be used to get all the trending venues in each neighborhood.

- iv) Python Libraries and tools as per the requirement will be imported or installed

2.2 Data Cleaning

2.2.1 Web Scrapping to Find the Boroughs and Neighborhoods

The Boroughs and Neighborhood data was scraped from Delhi Wikipedia page using the BeautifulSoup and lxml. The parent children relationship in html was used to retrieve the data. This retrieved data was converted into a tabular format in panda's data frame.

There were 185 records in tabular format with 185x2 structure i.e. 185 rows and 2 columns.

The sample Table format is given below:

	Borough	Neighborhood
1.	North West Delhi	Ashok Vihar
2.	North West Delhi	Azadpur
3.	North West Delhi	Bawana
4.	North West Delhi	Begum Pur
5.	North West Delhi	Dhaka
6.	North West Delhi	Jahangirpuri
7.	North West Delhi	Karala

2.2.2 Finding the Latitude and Longitude for Each Neighborhood

By using Nominatim the latitudes and longitudes were extracted for each neighborhood.

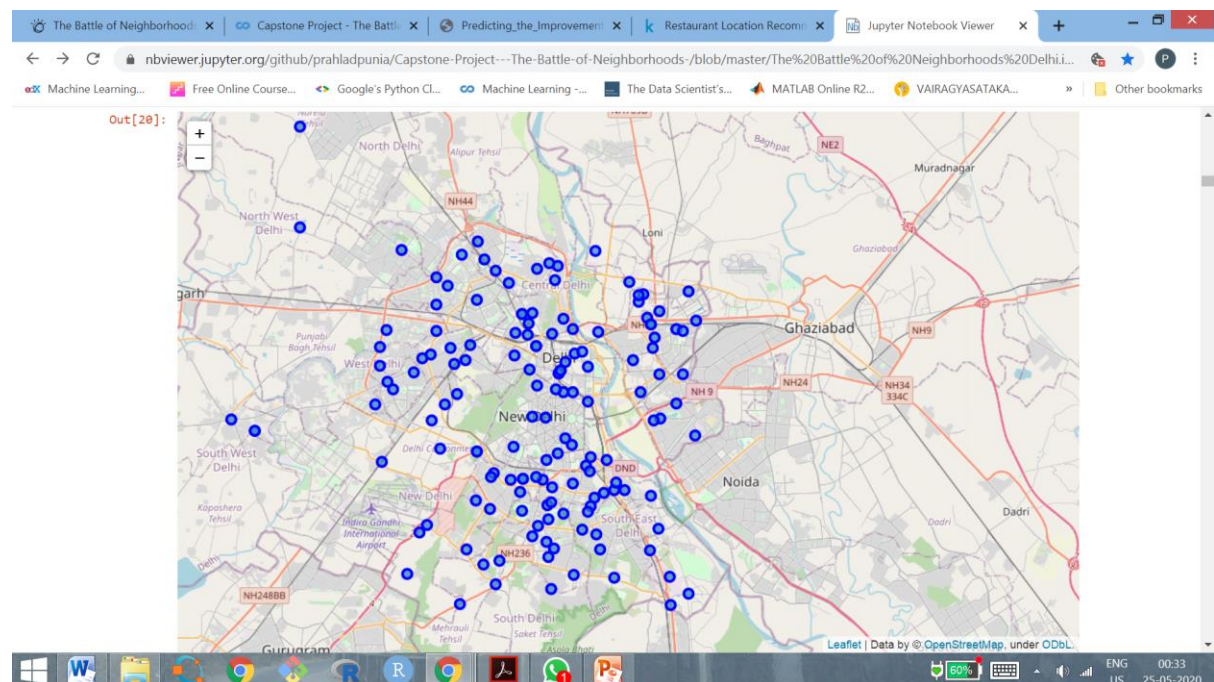
The two data sets that is Borough / Neighborhoods and the lat-long were joined to create a single pandas data frame.

Few records having NaN values were dropped.

This problem of NaN came because the lat –long coordinates could not be established correctly by the code due to lack of either spelling mismatch or similar names existing somewhere else. In few cases even Longitudes and latitudes were far off from the Delhi’s lat-long coordinates. A sample data frame is as given below:

	Borough	Neighborhood	latitude	longitude
0	North West Delhi	Adarsh Nagar	28.714401	77.167288
1	North West Delhi	Ashok Vihar	28.699453	77.184826
2	North West Delhi	Azadpur	28.707657	77.175547
3	North West Delhi	Bawana	28.799660	77.032885
4	North West Delhi	Dhaka	28.708698	77.205749

Finally the neighbourhoods were imposed on Delhi map to get the visual picture of the data



3. Exploring Data Analysis

3.1 Data Structure

The structure of Data was checked for further analysis.

i) Data Type

Data Types:

Borough **object**
Neighborhood **object**
latitude **float64**
longitude **float64**
dtype: **object**

ii) Data Validation

The data was manually checked to ascertain that all Latitude and Longitude values are in the vicinity of Delhi Latitude and Longitude and there are no out liars.

4. Foursquare API

4.1 Setting up Developer Account and Checking the Credentials

A developer account was setup with foursquare and credentials were checked.

```
CLIENT_ID = xxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxx
CLIENT_SECRET = xxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxx
VERSION = '20200524'
print ('Your credentials :')
print ('CLIENT_ID: ' + CLIENT_ID)
print ('CLIENT_SECRET:' + CLIENT_SECRET)
```

4.2 Fetching Data from Foursquare API

The foursquare account was used to fetch 100 venues within radius of 500 meters.

- a) The metadata was extracted using the get result and JASON.
- b) Categories of Venues were extracted by using the get_category_type
- c) JSON file was converted into pandas data frame by first filtering all the columns and rows and cleaning the columns. This will generate the complete database of nearby venues of Delhi.

d) The Data Frame is as follows:

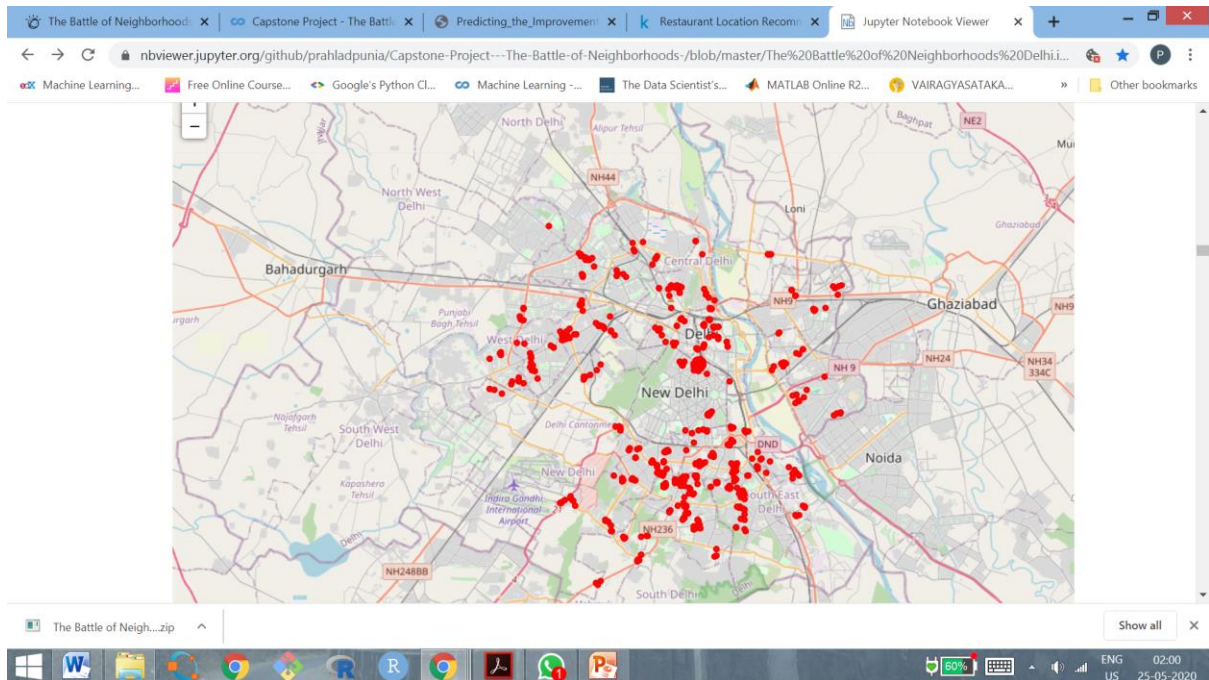
	name	categories	lat	long
0	Vishyavidyalaya Metro Station@Entry gate #1 n	Train Station	28.715596	77.170981
1	Adarsh Nagar Metro Station	Light Rail Station	28.716598	77.170436
2	Modern Age Computers	IT Services	28.716500	77.162900
3	My Idea Store	Mobile Phone Shop	28.717487	77.170922

4.3 Data frame for Delhi Venues with All Neighborhoods with Restaurants

Next a data Frame for Delhi Venues with all neighborhoods was created as shown below (849x8)

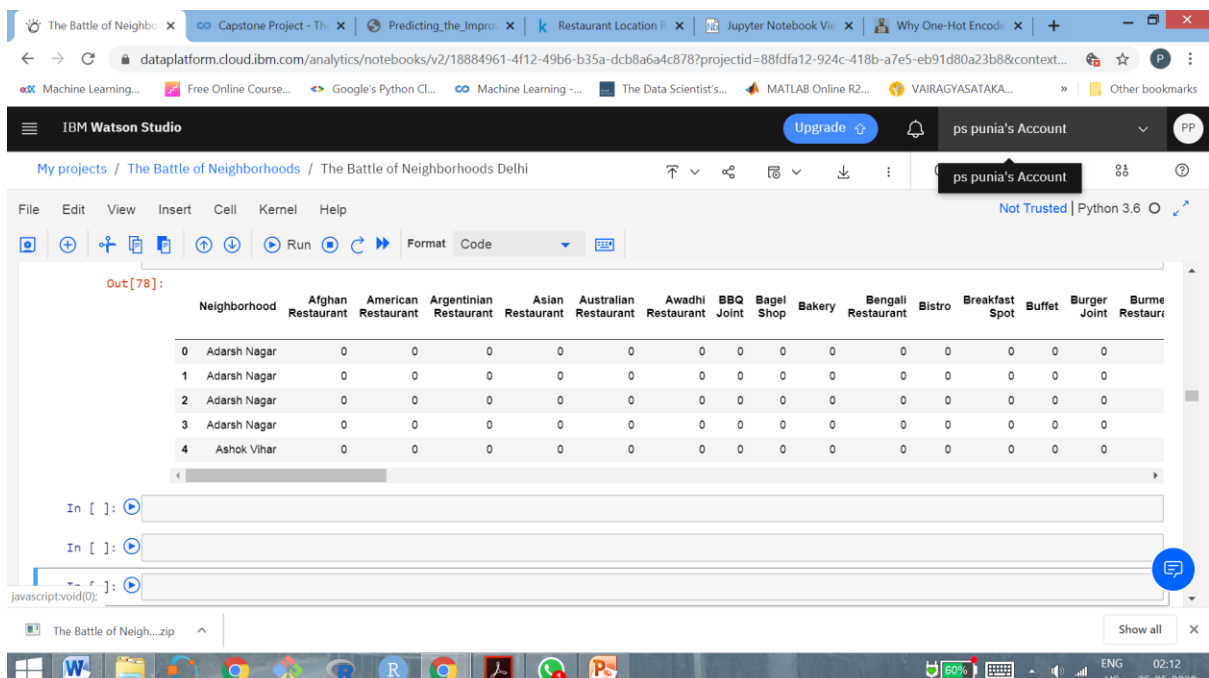
	Neighborhood	Neighborhood Latitude	Neighborhood Longitude	Venue	Venue Latitude	Venue Longitude	Venue Category
0	Adarsh Nagar	28.614193	77.071541	Eagle Boys Pizza	28.615595	77.070784	Pizza Place
1	Adarsh Nagar	28.614193	77.071541	Bikanerwala	28.613391	77.076084	Indian Restaurant
2	Adarsh Nagar	28.614193	77.071541	Bikano East Patel Nagar	28.616190	77.066978	Fast Food Restaurant
3	Adarsh Nagar	28.614193	77.071541	McDonald's	28.616330	77.067034	Fast Food Restaurant
4	Adarsh Nagar	28.614193	77.071541	Nat Khat Caterers	28.699630	77.187832	Indian Restaurant

4.4 Delhi Map with all Restaurants



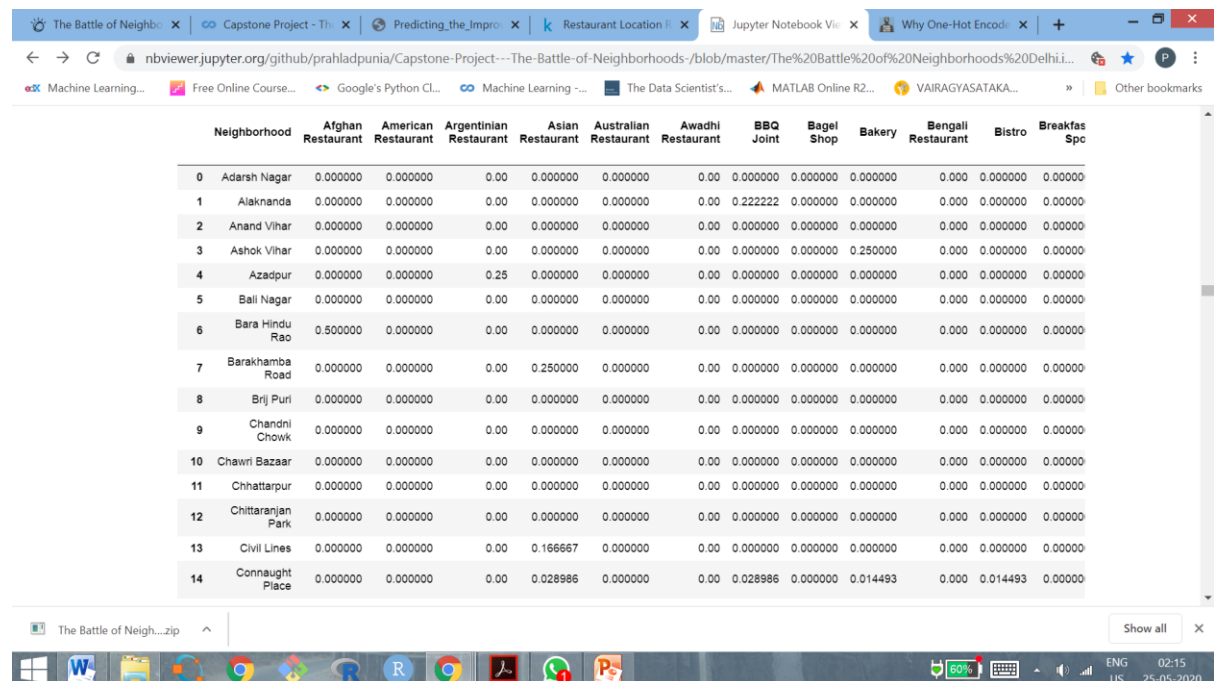
4.4 One Hot Encoding

This was done to make the categories of restaurants into categorical data for easier analysis.



4.5 Grouping the Data by Mean of the Frequency

This data was grouped by mean of the frequency of each category.



	Neighborhood	Afghan Restaurant	American Restaurant	Argentinian Restaurant	Asian Restaurant	Australian Restaurant	Awadhi Restaurant	BBQ Joint	Bagel Shop	Bakery	Bengali Restaurant	Bistro	Breakfast Spc
0	Adarsh Nagar	0.000000	0.000000	0.00	0.000000	0.000000	0.00	0.000000	0.000000	0.000000	0.000	0.000000	0.00000
1	Alaknanda	0.000000	0.000000	0.00	0.000000	0.000000	0.00	0.222222	0.000000	0.000000	0.000	0.000000	0.00000
2	Anand Vihar	0.000000	0.000000	0.00	0.000000	0.000000	0.00	0.000000	0.000000	0.000000	0.000	0.000000	0.00000
3	Ashok Vihar	0.000000	0.000000	0.00	0.000000	0.000000	0.00	0.000000	0.000000	0.250000	0.000	0.000000	0.00000
4	Azadpur	0.000000	0.000000	0.25	0.000000	0.000000	0.00	0.000000	0.000000	0.000000	0.000	0.000000	0.00000
5	Bali Nagar	0.000000	0.000000	0.00	0.000000	0.000000	0.00	0.000000	0.000000	0.000000	0.000	0.000000	0.00000
6	Bara Hindu Rao	0.500000	0.000000	0.00	0.000000	0.000000	0.00	0.000000	0.000000	0.000000	0.000	0.000000	0.00000
7	Barakhamba Road	0.000000	0.000000	0.00	0.250000	0.000000	0.00	0.000000	0.000000	0.000000	0.000	0.000000	0.00000
8	Brij Puri	0.000000	0.000000	0.00	0.000000	0.000000	0.00	0.000000	0.000000	0.000000	0.000	0.000000	0.00000
9	Chandni Chowk	0.000000	0.000000	0.00	0.000000	0.000000	0.00	0.000000	0.000000	0.000000	0.000	0.000000	0.00000
10	Chawri Bazaar	0.000000	0.000000	0.00	0.000000	0.000000	0.00	0.000000	0.000000	0.000000	0.000	0.000000	0.00000
11	Chhattarpur	0.000000	0.000000	0.00	0.000000	0.000000	0.00	0.000000	0.000000	0.000000	0.000	0.000000	0.00000
12	Chittaranjan Park	0.000000	0.000000	0.00	0.000000	0.000000	0.00	0.000000	0.000000	0.000000	0.000	0.000000	0.00000
13	Civil Lines	0.000000	0.000000	0.00	0.166667	0.000000	0.00	0.000000	0.000000	0.000000	0.000	0.000000	0.00000
14	Connaught Place	0.000000	0.000000	0.00	0.028986	0.000000	0.00	0.028986	0.000000	0.014493	0.000	0.014493	0.00000

5. K Means Clustering Algorithm

Clustering algorithm k-means was applied to the data. **K-means clustering** is one of the simplest and popular unsupervised machine learning **algorithms**. In other words, the **K-means algorithm** identifies **k** number of centroids, and then allocates every data point to the nearest cluster, while keeping the centroids as small as possible. K=3, 4, 5, 6 was tried. K=5 is the best. Results for 4, 5, and 6 here were the same. This may be due to small database.

The array generated by k-means is:

`array ([3, 4, 4, 0, 4, 4, 2, 4, 4, 3], dtype=int32) (for k=6)`

5.1 DataFrame - Clusters & top 10 Venues for each Neighborhood

IBM Watson Studio interface showing a Jupyter Notebook output. The output displays a DataFrame with columns: Borough, Neighborhood, latitude, longitude, Cluster_Labels, and 1st Most Common Venue through 9th Most Common Venue. The data is filtered for North West Delhi.

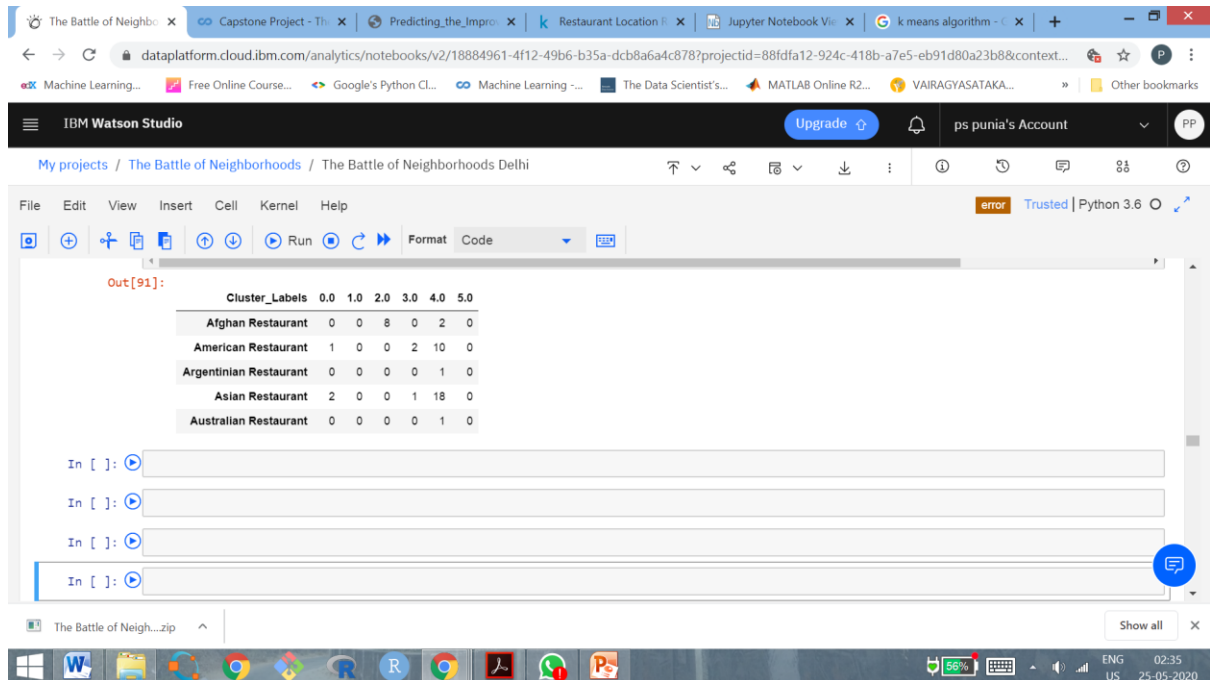
	Borough	Neighborhood	latitude	longitude	Cluster_Labels	1st Most Common Venue	2nd Most Common Venue	3rd Most Common Venue	4th Most Common Venue	5th Most Common Venue	6th Most Common Venue	7th Most Common Venue	8th Most Common Venue	9th Most Common Venue
0	North West Delhi	Adarsh Nagar	28.714401	77.167288	3.0	Fast Food Restaurant	Pizza Place	Indian Restaurant	Vegetarian / Vegan Restaurant	Dumpling Restaurant	Dhaba	Dim Sum Restaurant	Diner	Doner Restaurant
1	North West Delhi	Ashok Vihar	28.699453	77.184826	0.0	Indian Restaurant	Bakery	Diner	Falafel Restaurant	Dhaba	Dim Sum Restaurant	Doner Restaurant	Donut Shop	Dumpling Restaurant
2	North West Delhi	Azadpur	28.707657	77.175547	4.0	Cafe	Argentinian Restaurant	Indian Restaurant	Restaurant	Vegetarian / Vegan Restaurant	Eastern European Restaurant	Dim Sum Restaurant	Diner	Doner Restaurant
7	North West Delhi	Keshav Puram	28.688926	77.161683	4.0	Gastropub	Indian Restaurant	Cafe	Bakery	Food Truck	Food Stand	Food Court	Food	Fish & Chips Shop
9	North West Delhi	Kohat Enclave	28.698041	77.140539	0.0	Indian Restaurant	Bakery	Food Court	Food	Eastern European Restaurant	Dhaba	Dim Sum Restaurant	Diner	Doner Restaurant

5.2 Map of Clusters

IBM Watson Studio interface showing a Jupyter Notebook output. The output displays a map of Delhi with clusters of points colored by Cluster_Labels (0, 1, 2, 4, 7, 9). The map shows the distribution of these clusters across the city, with labels for various areas like Bahadurgarh, New Delhi, Noida, and Ghaziabad.

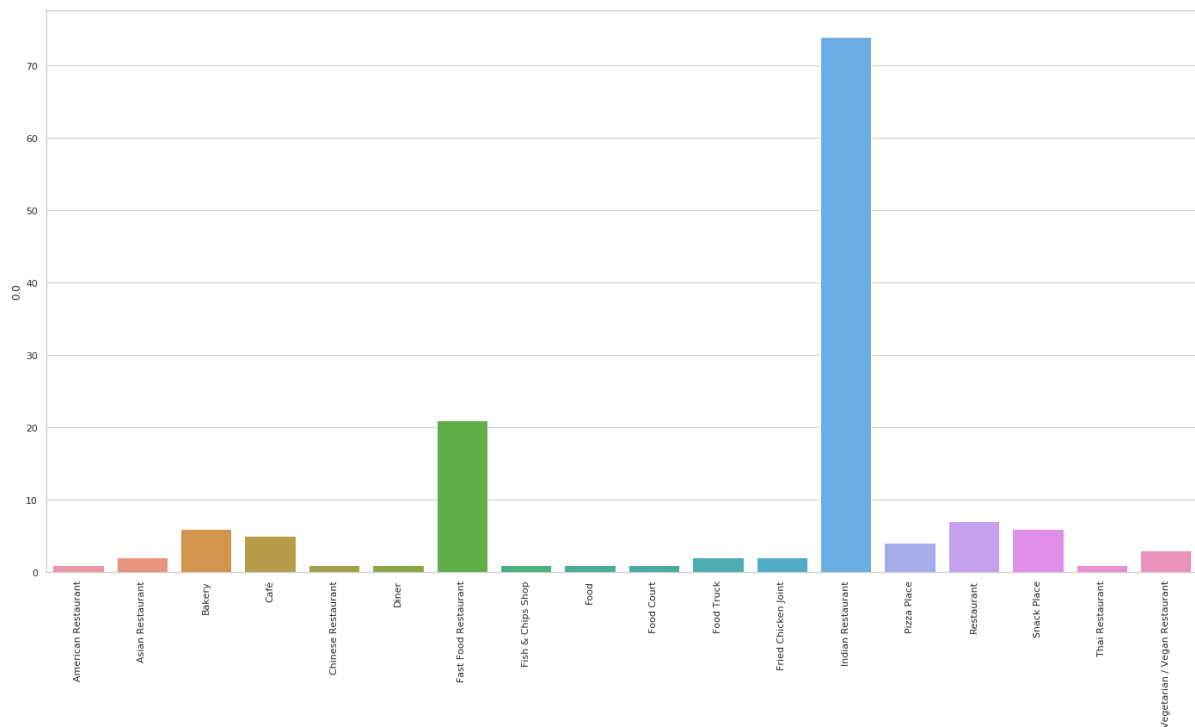
5.3 Analysing Each Cluster Data

The cluster data for each of the six clusters was analysed and the total number of each category were tabulated:

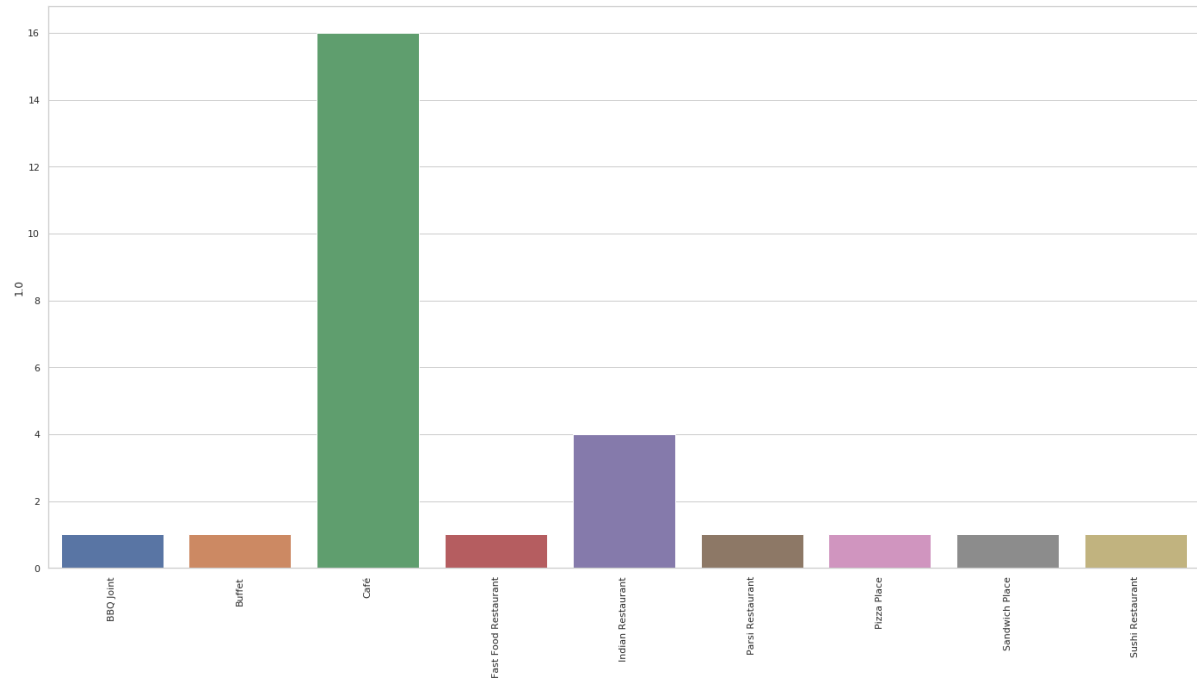


5.4 Plotting the Cluster Data on Bar Plot

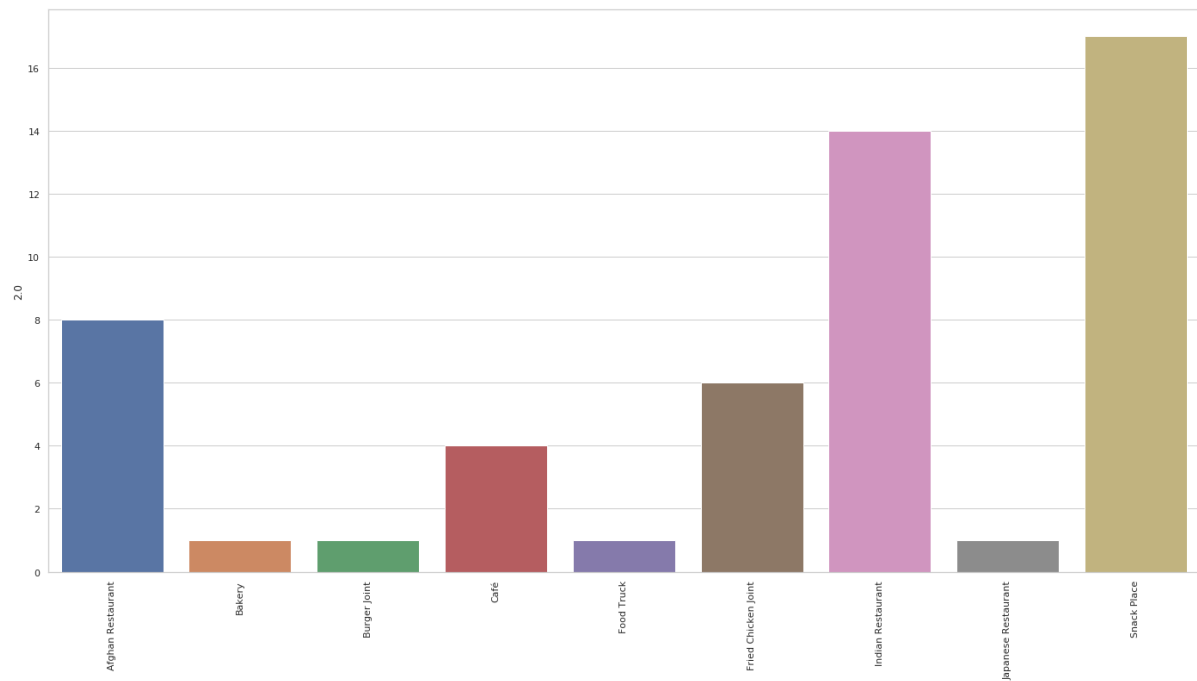
5.4.1. The cluster 0



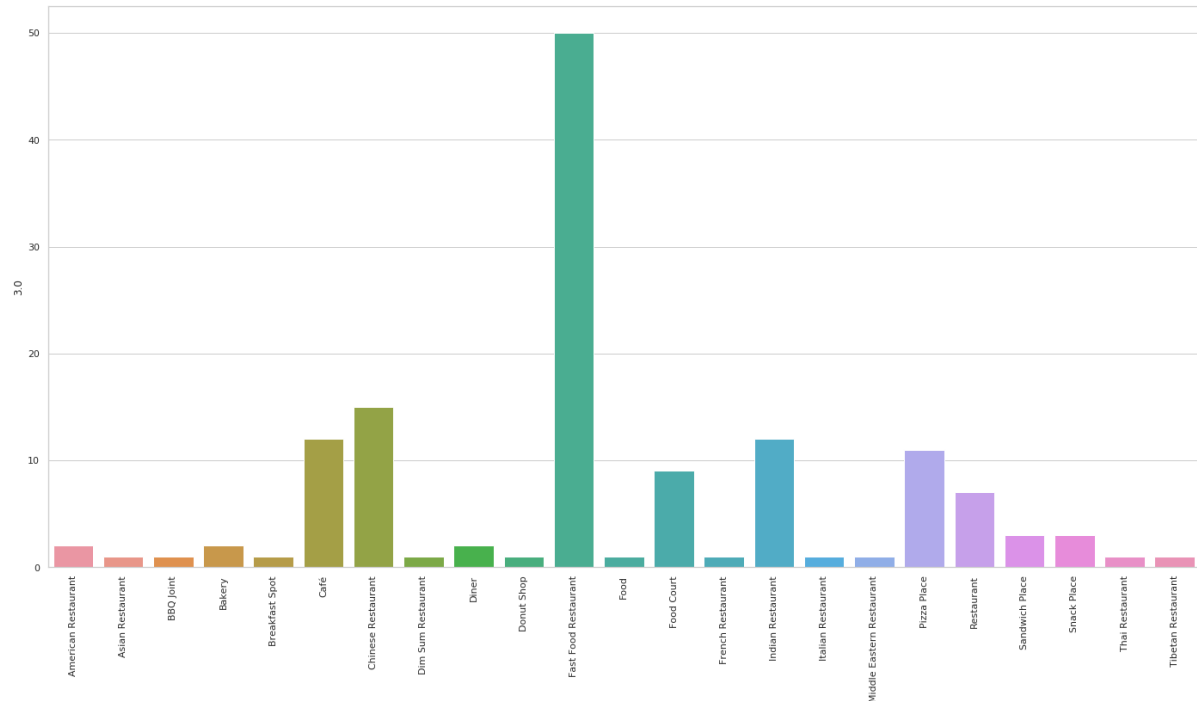
5.4.2 Cluster 1



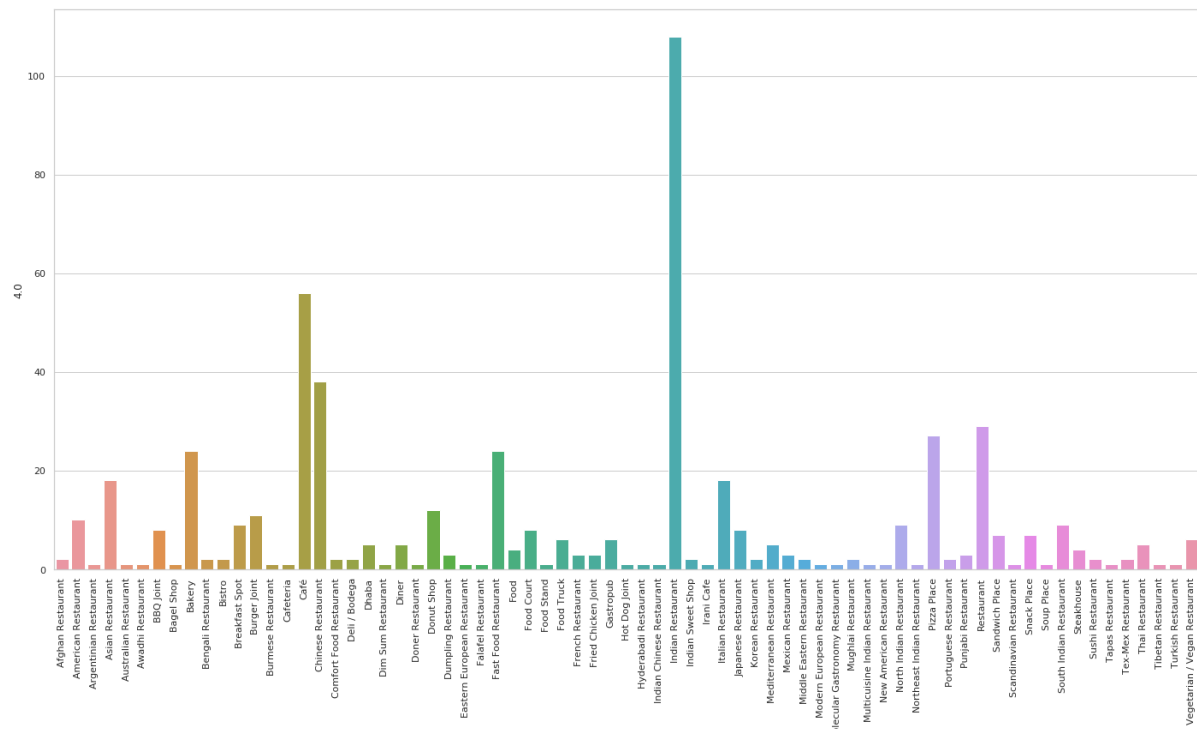
5.4.3 Cluster 2



5.4.4 Cluster 3



5.4.5 Cluster 4



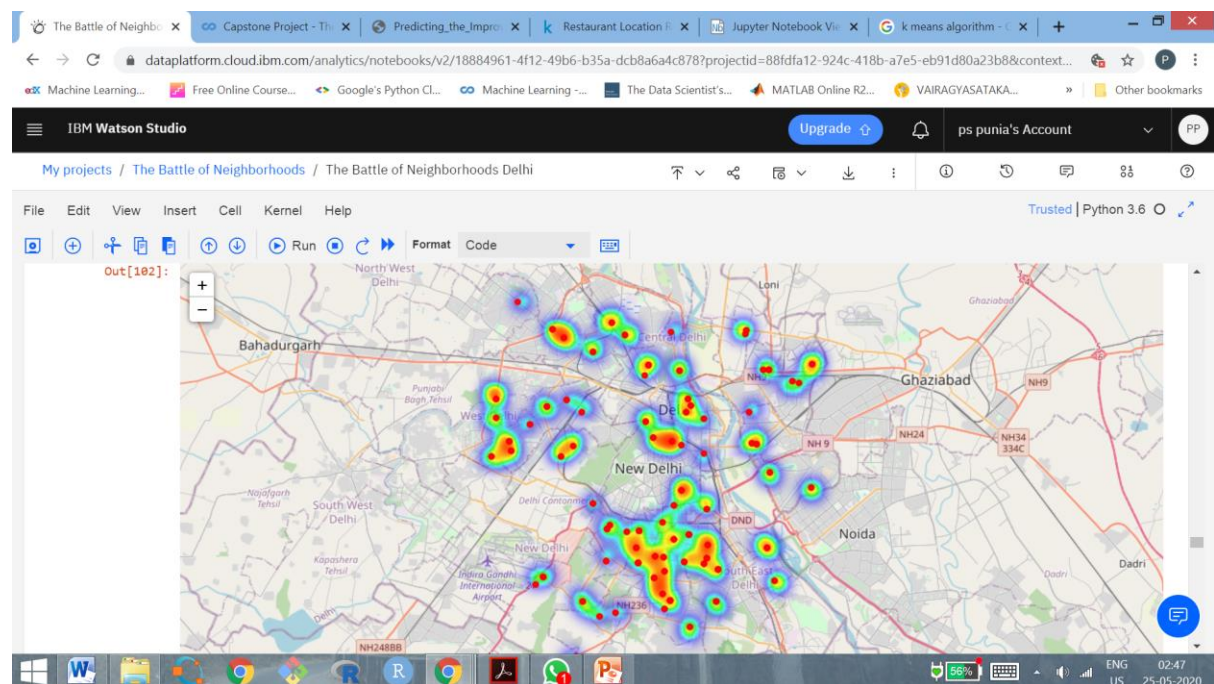
5.4.6 Analysing the Clusters

These 5 clusters were analysed by putting a constraint of Indian / other restaurants to arrive at the right location for Indian Multi Cuisine restaurants

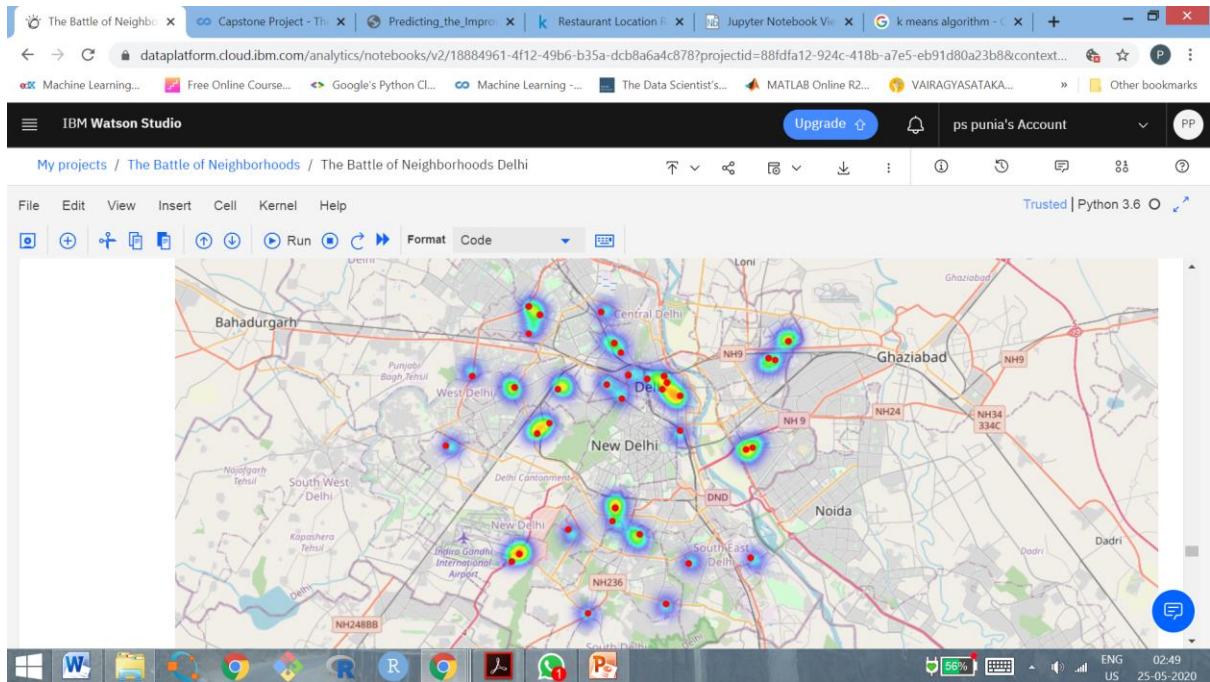
1. Frequency/ Density of the Indian Restaurants in each neighborhood. Avoid neighborhoods with >30% density of Indian Restaurants
2. Filter out the neighborhoods > 70 percentile of density for non-Indian Restaurants. Avoid neighborhoods having very high density of Restaurants.
3. Zero in on the most popular neighborhoods
4. Examine remaining neighborhoods which are close to popular neighborhoods and have less supply of Indian Restaurants
5. Finally zero in on the top neighborhoods for a new restaurant where demand supply gap exists and is closer to popular neighborhoods.

The results are visually shown in the next two maps.

5.4.7 Clusters on the Map of Delhi with All Type of Restaurants



5.4.7 Clusters with Only Indian restaurants



5.4.8. Most Popular Neighborhoods Delhi

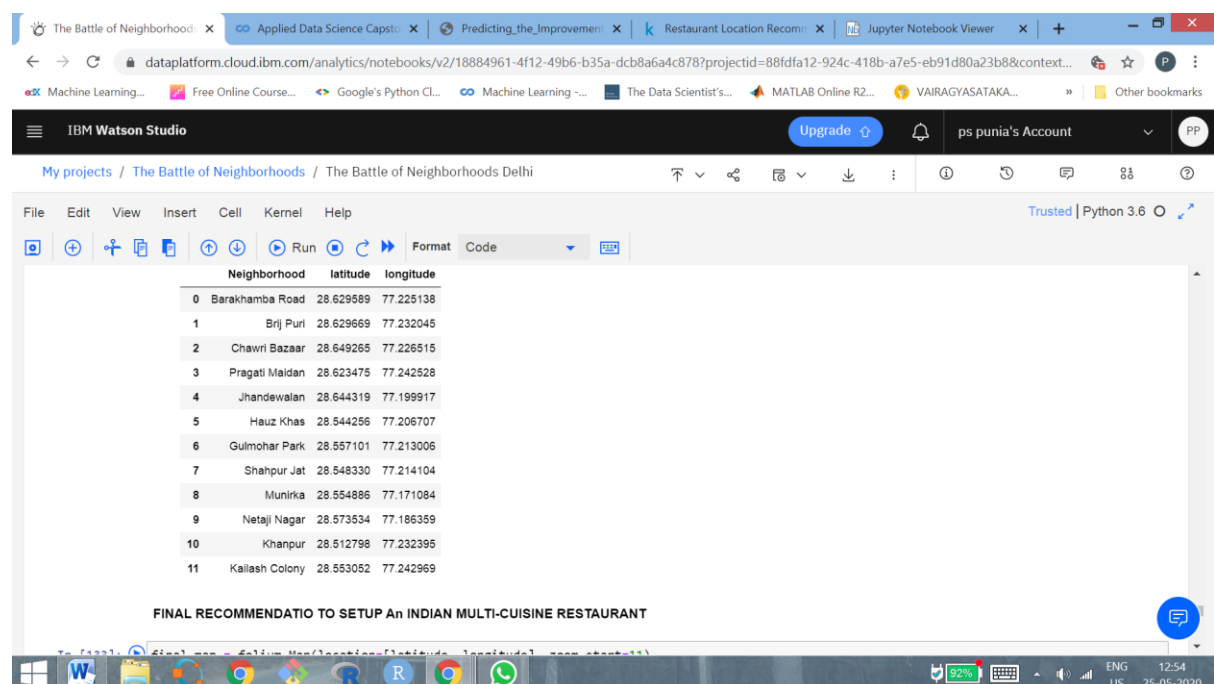
The following emerged as the most Popular Neighborhoods of Delhi:

```
array(['Connaught Place', 'Khirkhi Village', 'Hauz Khas Village'], dtype=object)
```

The neighbourhoods in these were further analysed for demand of Indian Restaurants and at the same time no oversupply of other Restaurants.

Five Top neighbourhoods from each were identified. These were combined from the three neighbourhoods to finally arrive at 12 neighborhoods where the new restaurant can be setup.

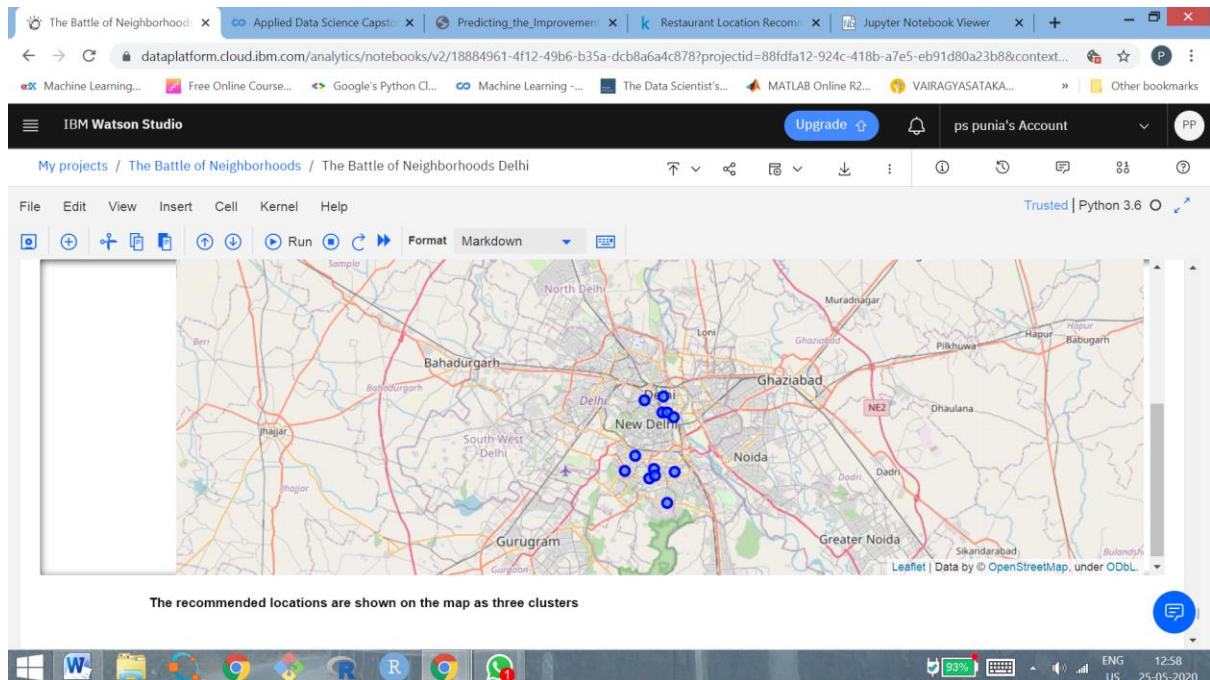
5.4.9 FINAL RECOMMENDATIO TO SETUP An INDIAN MULTI-CUISINE RESTAURANT



The screenshot displays the IBM Watson Studio interface. At the top, there's a browser window with multiple tabs, including 'The Battle of Neighborhoods', 'Applied Data Science Capstone', 'Predicting the Improvement', 'Restaurant Location Recommendation', and 'Jupyter Notebook Viewer'. The main interface shows a Jupyter Notebook with a table of 11 neighborhoods in Delhi. The table has columns for 'Neighborhood', 'latitude', and 'longitude'. Below the table, the text 'FINAL RECOMMENDATIO TO SETUP An INDIAN MULTI-CUISINE RESTAURANT' is visible. The bottom of the screen shows a Windows taskbar with various application icons and system tray information.

	Neighborhood	latitude	longitude
0	Barakhamba Road	28.629589	77.225138
1	Brij Puri	28.629669	77.232045
2	Chawri Bazaar	28.649265	77.226515
3	Pragati Maidan	28.623475	77.242528
4	Jhandewalan	28.644319	77.199917
5	Hauz Khas	28.544256	77.206707
6	Gulmohar Park	28.557101	77.213006
7	Shahpur Jat	28.548330	77.214104
8	Munirka	28.554886	77.171084
9	Netaji Nagar	28.573534	77.186359
10	Khanpur	28.512798	77.232395
11	Kailash Colony	28.553052	77.242969

5.4.9 Final Recommended Sites on Map of Delhi



6. Conclusion

1. Total Delhi Neighborhoods selected were 165 with 849 Restaurants within radius of 0.5 kms.
2. The neighbourhoods were further clustered or segmented on the number of restaurants and category of Restaurants i.e. French, Fast-food, Chinese, Italian, Indian, Thai and Japanese etc.; 116 neighborhood with 75 categories
3. Finally using K-Means algorithm 5 clusters were obtained.
4. Each of these clusters had a good variety of Restaurants e.g. Italian, Afghani, Australian etc.;
5. Out of the collection we further filtered down the Indian Restaurants.
6. The clusters had following distributions

Cluster	No of Indian Restaurants
0	5
1	10
2	75
3	125
4	1
5	0

7. The clusters 2 and 3 are selected for analysis as they have the highest density of Indian restaurants.
8. Next a Heat map to analyse the density of restaurants in other neighborhoods was generated - 56 neighborhoods with less number of Indian restaurants and at the same time over all low density of restaurants were selected.
9. Finally selected Top 3 neighborhoods i.e. Canaught Place (new name Rajiv Chowk), Hauz Khas Village and Khirki Village. As the potential venues.
10. Next Top 5 neighborhoods from these three were selected
11. The data in row 10 above was added together finally to arrive at 12 Neighborhoods that fulfil the business objective

Limitations of the results The results may not be fully accurate because of:

1. The dataset is very small. I tried with $k=3, 4, 5, 6$ and 7 but the results are mostly same with no change in final neighborhoods. To get better results we need a larger database of restaurants.
2. NCR the region in and around Delhi has 5000 restaurants listed on Zomato but we could find data only for 849.
3. Few neighborhoods (15 of them) Latitude and Longitude were beyond Delhi's lat long and had to be eliminated
4. Foursquare API does not have the most recent information on all restaurants in the city of Delhi
5. Only 100 Venues were selected initially which for a town like Delhi is very small.

References

1. Wikipedia page Neighborhoods of Delhi-
https://en.wikipedia.org/wiki/Neighbourhoods_of_Delhi
2. Coursera course Notes by Alex Aklson
3. <https://www.kaggle.com/shrutimehta/zomato-restaurants-data>
4. zomato restaurants dataset and project on neighborhoods in the city of New Delhi by Zayed Rais
5. Restaurant Location Recommender by Kumar Shashwat
6. The projects on Toronto Neighborhood Analysis submitted as part of this course
7. Finally the Discussion Forum was a great help. Thank you to each one.