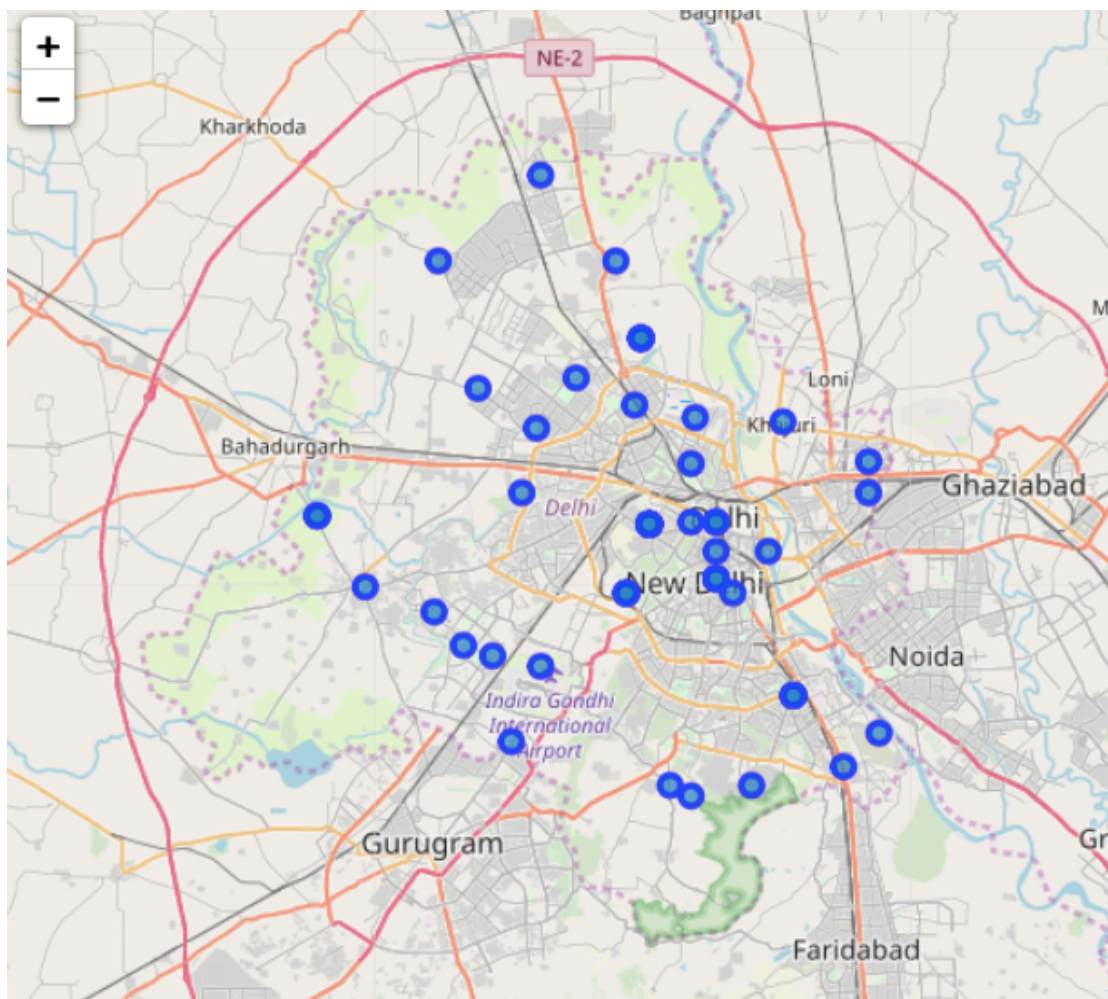


THE MOST APPROPRIATE LOCATION TO OPEN A RESTAURANT IN NEW DELHI (INDIA)

PREPARED FOR
IBM DATA SCIENCE PROFESSIONAL CERTIFICATE
(COURSERA)



PREPARED BY
MEHUL KUMAR

JULY 28TH, 2019

TABLE OF CONTENTS

| | <u>Page</u> |
|-----------------------------------|-------------|
| 1. Abstract..... | 1 |
| 2. Introduction | |
| 2.1 Background..... | 2 |
| 2.2 Problem..... | 2 |
| 2.3 Interest..... | 2 |
| 3. Data | |
| 3.1 Sources..... | 3 |
| 3.2 Cleaning..... | 3 |
| 3.3 Feature Selection..... | 4 |
| 4. Methodology | |
| 4.1 Explore Datasets..... | 5 |
| 4.2 Foursquare Location Data..... | 6 |
| 4.3 Pre-processing..... | 7 |
| 4.4 Clustering..... | 8 |
| 5. Results | |
| 5.1 Visualising Clusters..... | 9 |
| 5.2 Printing Clusters..... | 10 |
| 5.3 Answer to the problem..... | 12 |
| 6. Discussion..... | 13 |
| 7. Conclusion..... | 14 |
| 8. References..... | 15 |

1. Abstract

New Delhi is the capital city of India, a rapidly growing and developing nation with abundant of business opportunities. Opening a restaurant here is an attractive idea for any one who wants to extend their business to India.

If someone wants to open a restaurant in New Delhi to start/expand his/her business the first question that would come into their mind looking at the large population and choice of flavour of the "Delhites" is "Where is the most appropriate location to open a new restaurant in New Delhi".

When we consider this problem, we can create a map where the postal code is placed on New Delhi and each neighbourhood is clustered according to the venue density.

I collected the districts and neighbourhoods data from Open Government Data Platform India which provides the list of All India Pin-code directory with contact details along with Latitude and longitude, and found the coordinates for every neighbourhood. I did this manually and create a file.

Then I collected data of recommended venues inside 1000 meters radius of every neighbourhood, calculate the top10 most common venues by its category as features.

After reviewing our clusters we can choose the most appropriate location to open a restaurant in New Delhi and that mainly depend on the choice of restaurant type.

If someone is about to open a Chinese restaurant he/she may opt for Cluster 1 location that is mostly in South West Delhi.

2. Introduction

2.1 Background

This is the capstone project of IBM Data Science Professional Certificate. In this project, I will use the Foursquare location data to explore neighbourhoods in New Delhi, and to come up with a problem that I can use the Foursquare location data to solve.

2.2 Problem

I will explore the neighbourhoods in New Delhi and answer the question: “Where is the most appropriate location to open a new restaurant in New Delhi”.

2.3 Interest

New Delhi is the capital city of India, a rapidly growing and developing nation with abundant of business opportunities. Opening a restaurant here is an attractive idea for any one who want to extends its business to India. They would be very interested in this project.

3. Data

3.1 Data Sources

My main two data sources are:

- Open Government Data Platform India provides the list of All India Pincode directory with contact details along with Latitude and longitude.
- Venues data from Foursquare

Although the coordinates data was not available for New Delhi. So, I created a dataset for the Longitude and Latitude for all the postal codes available in New Delhi.

3.2 Data Cleaning

New Delhi is an urban district of Delhi which serves as the capital of India. First of all, I need to collect the districts and neighbourhoods data from Open Government Data Platform India provides the list of All India Pincode directory with contact details along with Latitude and longitude. and find the coordinates for every neighbourhood. I did this manually and create a file named DELHI_GEO.csv

The formatted data looks like this:

| | PostalCode | Region | Localities | Neighborhood | Latitude | Longitude |
|---|------------|------------------|---|-------------------------|-----------|-----------|
| 0 | 110001 | Central Delhi | Baroda House S.O,Bengali Market S.O,Bhagat Sin... | Connaught Place | 28.633300 | 77.216700 |
| 1 | 110002 | Central Delhi | A.G.C.R. S.O,Ajmeri Gate Extn. S.O,Darya Ganj ... | Darya Ganj | 28.633300 | 77.250000 |
| 2 | 110003 | Central Delhi | Delhi High Court Extension Counter S.O,Delhi H... | Aliganj | 28.650000 | 77.216700 |
| 3 | 110004 | Central Delhi | Rashtrapati Bhawan S.O | Rashtrapati Bhawan | 28.650000 | 77.216700 |
| 4 | 110005 | Central Delhi | Anand Parbat Indl. Area S.O,Anand Parbat S.O,B... | Lower Camp Anand Parbat | 28.650000 | 77.200000 |
| 5 | 110006 | Central Delhi | Hauz Qazi S.O,Jama Masjid S.O | Bara Tooti | 28.650000 | 77.216700 |
| 6 | 110007 | North Delhi | C.C.I. S.O,Delhi University S.O,Gulabi Bagh S.... | Birla Lines | 28.683300 | 77.200000 |
| 7 | 110008 | Central Delhi | Dada Ghosh Bhawan S.O,Patel Nagar East S.O,Pat... | Patel Nagar | 28.650000 | 77.216700 |
| 8 | 110009 | North West Delhi | Dr.Mukerjee Nagar S.O,G.T.B.Nagar S.O,Gujanwa... | G.T.B Nagar | 28.709895 | 77.202561 |
| 9 | 110010 | South West Delhi | 505 A B Workshop S.O,A F Palam S.O,Aps Colony ... | Delhi Cantt | 28.550000 | 77.266700 |

3.3 Feature Selection

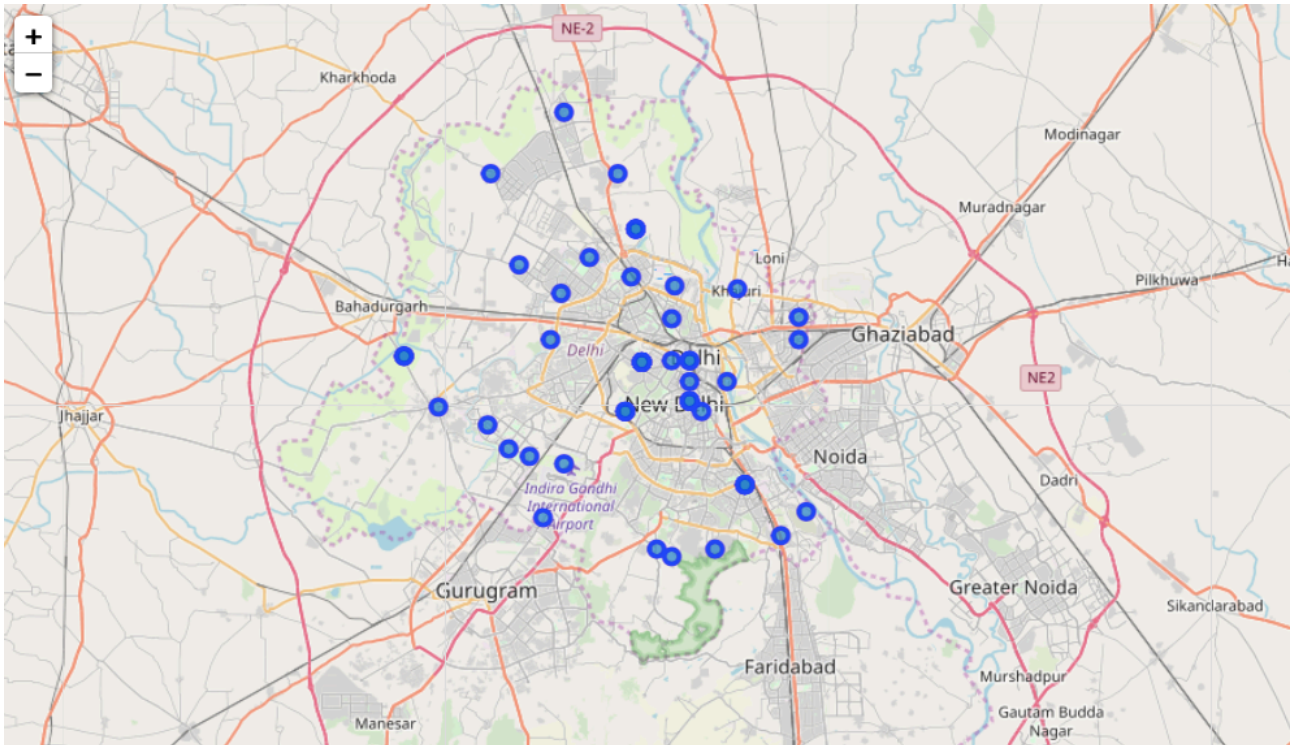
In this project, I will get data of recommended venues inside 1000 meters radius of every neighbourhood, calculate the top10 most common venues by its category as features.

| Feature Label | Type & Description |
|------------------------|---------------------------------|
| 1st Most Common Venue | str. Name of the vanue category |
| 2nd Most Common Venue | str |
| 3rd Most Common Venue | str |
| 4th Most Common Venue | str |
| 5th Most Common Venue | str |
| 6th Most Common Venue | str |
| 7th Most Common Venue | str |
| 8th Most Common Venue | str |
| 9th Most Common Venue | str |
| 10th Most Common Venue | str |

4. Methodology

4.1 Explore Dataset

Using folium library to create map with markers to establish a global view of the dataset, and examine the accuracy of coordinates.



It looks like every marker is in the right place. Markers cover almost whole New Delhi. We can identify some famous place Connaught Place, IGI Airport.

4.2 Examine Foursquare Location Data

Using Foursquare's `/venues/explore` API to get recommended venues for a specific location. To simplify the results set the `limit` property is 100 and `radius` is 1000.

This request should return a JSON data includes up to 100 venues for a coordinate.

```
results = requests.get(url).json()
results
```

```
{'meta': {'code': 200, 'requestId': '5d3f11e6f8953d00254febcb'},
 'response': {'suggestedFilters': {'header': 'Tap to show:',
  'filters': [{'name': 'Open now', 'key': 'openNow'}]},
  'headerLocation': 'N.D. Charge 1',
  'headerFullLocation': 'N.D. Charge 1, Delhi',
  'headerLocationGranularity': 'neighborhood',
```

4.3 Pre-processing Data

Firstly, use the venues data from Foursquare to create a DataFrame:

| | Neighborhood | Neighborhood Latitude | Neighborhood Longitude | Venue | Venue Latitude | Venue Longitude | Venue Category |
|---|-----------------|-----------------------|------------------------|--|----------------|-----------------|-------------------------|
| 0 | Connaught Place | 28.6333 | 77.2167 | Wenger's | 28.633412 | 77.218292 | Bakery |
| 1 | Connaught Place | 28.6333 | 77.2167 | HOTEL SARAVANA BHAVAN | 28.632319 | 77.216445 | South Indian Restaurant |
| 2 | Connaught Place | 28.6333 | 77.2167 | Starbucks Coffee | 28.632011 | 77.217731 | Coffee Shop |
| 3 | Connaught Place | 28.6333 | 77.2167 | Fabindia | 28.632012 | 77.217729 | Clothing Store |
| 4 | Connaught Place | 28.6333 | 77.2167 | Connaught Place कर्नाट प्लेस (Connaught Place) | 28.632731 | 77.220018 | Plaza |

Then create one-hot encoding to the venues categories, get mean of each category for every neighbourhood:

| | Neighborhood | ATM | Accessories Store | Afghan Restaurant | Arcade | Art Gallery | Asian Restaurant | Athletics & Sports | BBQ Joint | Bakery | ... | South Indian Restaurant | Spa | Sporting Goods Shop |
|---|-----------------|-----|-------------------|-------------------|--------|-------------|------------------|--------------------|-----------|--------|-----|-------------------------|-----|---------------------|
| 0 | Connaught Place | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 1 | ... | 0 | 0 | 0 |
| 1 | Connaught Place | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | ... | 1 | 0 | 0 |
| 2 | Connaught Place | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | ... | 0 | 0 | 0 |
| 3 | Connaught Place | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | ... | 0 | 0 | 0 |
| 4 | Connaught Place | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | ... | 0 | 0 | 0 |

Finally calculate the frequency of categories for each neighbourhood, we get the top10 most common venues for every neighbourhood:

| | PostalCode | Region | Localities | 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 |
|---|------------|---------------|---|-------------------------|----------|-----------|----------------|-----------------------|-----------------------|-----------------------|-----------------------|-----------------------|
| 0 | 110001 | Central Delhi | Baroda House S.O,Bengali Market S.O,Bhagat Sin... | Connaught Place | 28.6333 | 77.2167 | 1.0 | Café | Indian Restaurant | Chinese Restaurant | Bar | Coffee Shop |
| 1 | 110002 | Central Delhi | A.G.C.R. S.O,Ajmeri Gate Extn. S.O,Darya Ganj ... | Darya Ganj | 28.6333 | 77.2500 | 4.0 | Stadium | Women's Store | Fast Food Restaurant | Deli / Bodega | Department Store |
| 2 | 110003 | Central Delhi | Delhi High Court Extension Counter S.O,Delhi H... | Aliganj | 28.6500 | 77.2167 | 1.0 | Hotel | Restaurant | Hostel | Indian Restaurant | Motel |
| 3 | 110004 | Central Delhi | Rashtrapati Bhawan S.O | Rashtrapati Bhawan | 28.6500 | 77.2167 | 1.0 | Hotel | Restaurant | Hostel | Indian Restaurant | Motel |
| 4 | 110005 | Central Delhi | Anand Parbat Indl. Area S.O,Anand Parbat S.O,B... | Lower Camp Anand Parbat | 28.6500 | 77.2000 | 1.0 | Playground | Hotel | Jewelry Store | Fast Food Restaurant | Deli / Bodega |

4.4 Clustering

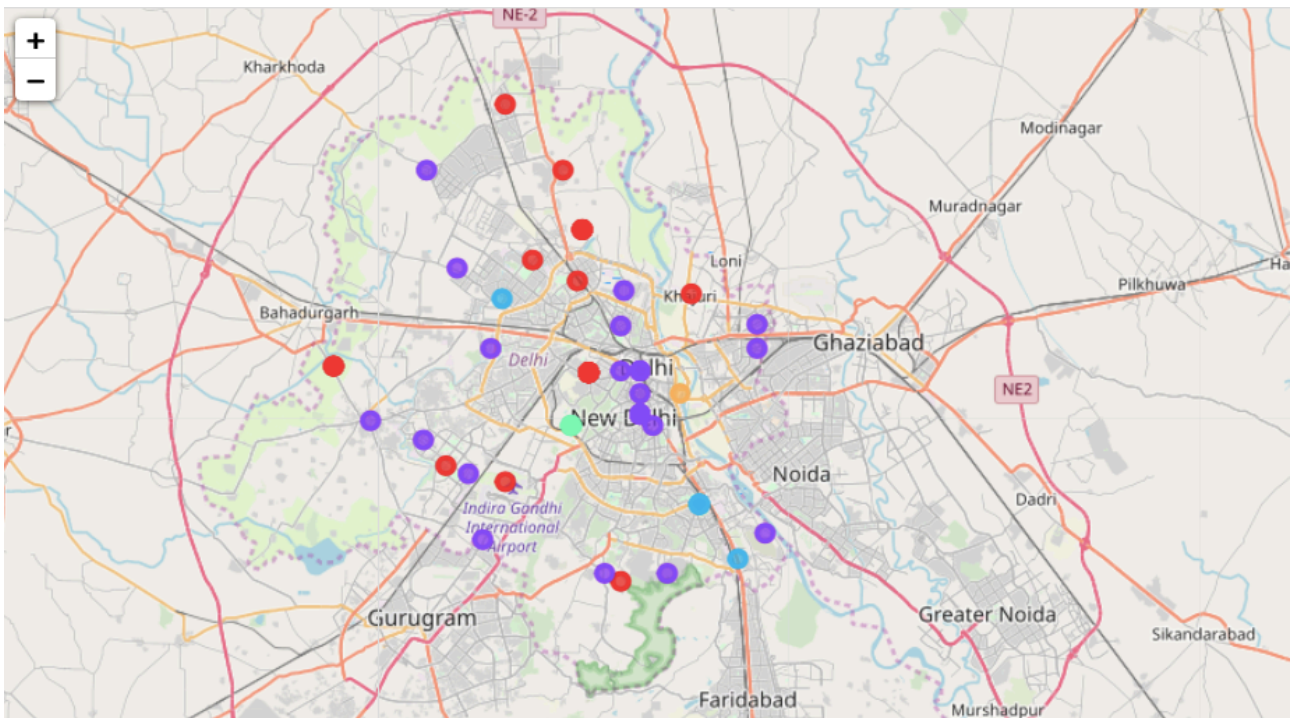
For this project, k-means is an appropriate clustering algorithm. Because we have a unlabelled dataset, so this is an unsupervised learning project. K-means clustering aims to partition n observations into k clusters in which each observation belongs to the cluster with the nearest mean. By clustering the neighbourhoods, we can find out the pattern in them, identify the identical neighbourhoods and see which is our target.

One difficulty of k-means is to determine the hyper-parameter k . In this project I will choose $k=5$, means clustering the neighbourhoods into 5 clusters. $k=5$ is an experience hyper-parameter.

5. Results

5.1 Visualize Clusters

Again, I use folium library to create a map with clustered markers:



There are two main clusters.

5.2 Printing Clusters

Cluster 1

| | Region | 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 |
|----|------------------|-----------|----------------|-----------------------|-----------------------|-----------------------|-----------------------|-----------------------|-----------------------|-----------------------|-----------------------|-----------------------|------------------------|
| 19 | New Delhi | 77.172600 | 0.0 | Chinese Restaurant | Coffee Shop | Indian Restaurant | Arcade | Donut Shop | Café | Fast Food Restaurant | Dessert Shop | Discount Store | Electronics Store |
| 20 | South West Delhi | 77.172600 | 0.0 | Chinese Restaurant | Coffee Shop | Indian Restaurant | Arcade | Donut Shop | Café | Fast Food Restaurant | Dessert Shop | Discount Store | Electronics Store |
| 21 | South West Delhi | 77.172600 | 0.0 | Chinese Restaurant | Coffee Shop | Indian Restaurant | Arcade | Donut Shop | Café | Fast Food Restaurant | Dessert Shop | Discount Store | Electronics Store |
| 22 | South West Delhi | 77.172600 | 0.0 | Chinese Restaurant | Coffee Shop | Indian Restaurant | Arcade | Donut Shop | Café | Fast Food Restaurant | Dessert Shop | Discount Store | Electronics Store |
| 23 | South Delhi | 77.172600 | 0.0 | Chinese Restaurant | Coffee Shop | Indian Restaurant | Arcade | Donut Shop | Café | Fast Food Restaurant | Dessert Shop | Discount Store | Electronics Store |
| 24 | South Delhi | 77.172600 | 0.0 | Chinese Restaurant | Coffee Shop | Indian Restaurant | Arcade | Donut Shop | Café | Fast Food Restaurant | Dessert Shop | Discount Store | Electronics Store |
| 25 | West Delhi | 77.172600 | 0.0 | Chinese Restaurant | Coffee Shop | Indian Restaurant | Arcade | Donut Shop | Café | Fast Food Restaurant | Dessert Shop | Discount Store | Electronics Store |
| 26 | West Delhi | 77.172600 | 0.0 | Chinese Restaurant | Coffee Shop | Indian Restaurant | Arcade | Donut Shop | Café | Fast Food Restaurant | Dessert Shop | Discount Store | Electronics Store |
| 27 | South West Delhi | 77.172600 | 0.0 | Chinese Restaurant | Coffee Shop | Indian Restaurant | Arcade | Donut Shop | Café | Fast Food Restaurant | Dessert Shop | Discount Store | Electronics Store |
| 28 | New Delhi | 77.172600 | 0.0 | Chinese Restaurant | Coffee Shop | Indian Restaurant | Arcade | Donut Shop | Café | Fast Food Restaurant | Dessert Shop | Discount Store | Electronics Store |

Cluster 2

| | Region | 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 |
|----|------------------|-----------|----------------|-----------------------|-----------------------|-----------------------|-----------------------|-----------------------|-----------------------|-----------------------|-----------------------|
| 0 | Central Delhi | 77.216700 | 1.0 | Café | Indian Restaurant | Chinese Restaurant | Bar | Coffee Shop | Clothing Store | Fast Food Restaurant | Pub |
| 2 | Central Delhi | 77.216700 | 1.0 | Hotel | Restaurant | Hostel | Indian Restaurant | Motel | Coffee Shop | Electronics Store | Department Store |
| 3 | Central Delhi | 77.216700 | 1.0 | Hotel | Restaurant | Hostel | Indian Restaurant | Motel | Coffee Shop | Electronics Store | Department Store |
| 4 | Central Delhi | 77.200000 | 1.0 | Playground | Hotel | Jewelry Store | Fast Food Restaurant | Deli / Bodega | Department Store | Dessert Shop | Discount Store |
| 5 | Central Delhi | 77.216700 | 1.0 | Hotel | Restaurant | Hostel | Indian Restaurant | Motel | Coffee Shop | Electronics Store | Department Store |
| 6 | North Delhi | 77.200000 | 1.0 | Breakfast Spot | Pizza Place | Playground | Indian Restaurant | Dessert Shop | Miscellaneous Shop | Donut Shop | Coffee Shop |
| 7 | Central Delhi | 77.216700 | 1.0 | Hotel | Restaurant | Hostel | Indian Restaurant | Motel | Coffee Shop | Electronics Store | Department Store |
| 8 | North West Delhi | 77.202561 | 1.0 | Pizza Place | Playground | Hotel | Hostel | Coffee Shop | Deli / Bodega | Department Store | Dessert Shop |
| 31 | East Delhi | 77.316700 | 1.0 | Chinese Restaurant | Sporting Goods Shop | Department Store | Pool | Food & Drink Shop | Athletics & Sports | Food | Gift Shop |
| 38 | North West Delhi | 77.033300 | 1.0 | Playground | Fast Food Restaurant | Deli / Bodega | Department Store | Dessert Shop | Discount Store | Donut Shop | Electronics Store |

Cluster 3

| | Region | 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 |
|----|------------------|-----------|----------------|-----------------------|-----------------------|-----------------------|-----------------------|-----------------------|-----------------------|-----------------------|-----------------------|-----------------------|
| 9 | South West Delhi | 77.266700 | 2.0 | Fast Food Restaurant | Indian Restaurant | Historic Site | Electronics Store | Women's Store | Department Store | Dessert Shop | Discount Store | Donut Shop |
| 10 | Central Delhi | 77.266700 | 2.0 | Fast Food Restaurant | Indian Restaurant | Historic Site | Electronics Store | Women's Store | Department Store | Dessert Shop | Discount Store | Donut Shop |
| 11 | Central Delhi | 77.266700 | 2.0 | Fast Food Restaurant | Indian Restaurant | Historic Site | Electronics Store | Women's Store | Department Store | Dessert Shop | Discount Store | Donut Shop |
| 12 | South Delhi | 77.266700 | 2.0 | Fast Food Restaurant | Indian Restaurant | Historic Site | Electronics Store | Women's Store | Department Store | Dessert Shop | Discount Store | Donut Shop |
| 13 | South Delhi | 77.266700 | 2.0 | Fast Food Restaurant | Indian Restaurant | Historic Site | Electronics Store | Women's Store | Department Store | Dessert Shop | Discount Store | Donut Shop |
| 14 | West Delhi | 77.266700 | 2.0 | Fast Food Restaurant | Indian Restaurant | Historic Site | Electronics Store | Women's Store | Department Store | Dessert Shop | Discount Store | Donut Shop |
| 15 | South West Delhi | 77.266700 | 2.0 | Fast Food Restaurant | Indian Restaurant | Historic Site | Electronics Store | Women's Store | Department Store | Dessert Shop | Discount Store | Donut Shop |
| 16 | South Delhi | 77.266700 | 2.0 | Fast Food Restaurant | Indian Restaurant | Historic Site | Electronics Store | Women's Store | Department Store | Dessert Shop | Discount Store | Donut Shop |
| 17 | West Delhi | 77.266700 | 2.0 | Fast Food Restaurant | Indian Restaurant | Historic Site | Electronics Store | Women's Store | Department Store | Dessert Shop | Discount Store | Donut Shop |
| 18 | South Delhi | 77.266700 | 2.0 | Fast Food Restaurant | Indian Restaurant | Historic Site | Electronics Store | Women's Store | Department Store | Dessert Shop | Discount Store | Donut Shop |
| 43 | South Delhi | 77.300000 | 2.0 | Indian Restaurant | Train Station | Women's Store | Food | Department Store | Dessert Shop | Discount Store | Donut Shop | Electronics Store |

Cluster 4

| | Region | 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 |
|----|------------------|-----------|----------------|-----------------------|-----------------------|-----------------------|-----------------------|-----------------------|-----------------------|-----------------------|-----------------------|-----------------------|------------------------|
| 40 | West Delhi | 77.1569 | 3.0 | Art Gallery | Market | Women's Store | Fast Food Restaurant | Department Store | Dessert Shop | Discount Store | Donut Shop | Electronics Store | Falafel Restaurant |
| 41 | North West Delhi | 77.1569 | 3.0 | Art Gallery | Market | Women's Store | Fast Food Restaurant | Department Store | Dessert Shop | Discount Store | Donut Shop | Electronics Store | Falafel Restaurant |
| 45 | South West Delhi | 77.1569 | 3.0 | Art Gallery | Market | Women's Store | Fast Food Restaurant | Department Store | Dessert Shop | Discount Store | Donut Shop | Electronics Store | Falafel Restaurant |
| 46 | South West Delhi | 77.1569 | 3.0 | Art Gallery | Market | Women's Store | Fast Food Restaurant | Department Store | Dessert Shop | Discount Store | Donut Shop | Electronics Store | Falafel Restaurant |
| 47 | South Delhi | 77.1569 | 3.0 | Art Gallery | Market | Women's Store | Fast Food Restaurant | Department Store | Dessert Shop | Discount Store | Donut Shop | Electronics Store | Falafel Restaurant |
| 48 | South Delhi | 77.1569 | 3.0 | Art Gallery | Market | Women's Store | Fast Food Restaurant | Department Store | Dessert Shop | Discount Store | Donut Shop | Electronics Store | Falafel Restaurant |

Cluster 5

| | Region | 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 |
|---|---------------|-----------|----------------|-----------------------|-----------------------|-----------------------|-----------------------|-----------------------|-----------------------|-----------------------|-----------------------|-----------------------|------------------------|
| 1 | Central Delhi | 77.25 | 4.0 | Stadium | Women's Store | Fast Food Restaurant | Deli / Bodega | Department Store | Dessert Shop | Discount Store | Donut Shop | Electronics Store | Falafel Restaurant |

5.3 Answer to the Question

The most common venue is Chinese Restaurant followed by Fast Food Restaurant and Indian Restaurant.

After reviewing our clusters we can choose the most appropriate location to open a restaurant in New Delhi and that mainly depend on the choice of restaurant type.

If someone is about to open a Chinese restaurant he/she may opt for Cluster 1 location that is mostly in South West and West Delhi.

So, I'm about to give a recommendation in most appropriate place to open a restaurant taken 3 options as restaurant type.

| Most Appropriate Location to Open a Restaurant in New Delhi(INDIA) | | | | |
|--|-----------|-------------|---------------|---------------|
| | Location: | Location 1 | Location 2 | Location 3 |
| Restraunt Type: | | | | |
| Chinese | | SW Delhi | NW Delhi | East Delhi |
| Fast Food | | South Delhi | SW Delhi | Central Delhi |
| Indian | | SW Delhi | Central Delhi | South Delhi |
| Others | | East Delhi | North Delhi | North Delhi |

6. Discussion

For this project, k-means works properly. We use the most common venues in neighbourhood, cluster the neighbourhoods into 5 clusters. The result is good for our problem.

Although the result is accurate, we can also notice that the clusters is not precise enough. Adding more features like population and average income would be helpful.

7. Conclusion

In this project, we need to use the location data from Foursquare to solve the problem “Where is the appropriate place to open a new restaurant in New Delhi”.

I collect the neighbourhoods data from Open Government Data Platform India which provides the list of All India Pin-code directory with contact details along with Latitude and longitude, and format it manually. Get venues data using Foursquare’s API. One-hot encode the venues’ categories and calculate the frequencies, then get TOP10 common venues for each neighbourhood.

I use k-means algorithm to cluster the dataset into 5 clusters. I use folium library to visualise the cluster results. By examining the results, I could answer the question of this project.

In the future, more features could help to improve the model to get a better result.

8. References

- [Open Government Data Platform India](#)
- [Foursquare API](#)
- [Python Notebook of all the analysis](#)

