

BATTLE OF NEIGHBOURHOODS

Ranjeet Kumar

27 March 2020

Topic: Identifying Neighbourhood in Central Delhi for Opening a Restaurant

1. Introduction

1.1 Background

New Delhi is India's capital and a hub of tourists attracting thousands of tourists every day. New Delhi serves as the seat of all three branches of the Government of India. Population density in New Delhi is about 19,300 Delhiites per square mile (7,400 per square kilometre). It is primarily divided into multiple districts Central Delhi, East Delhi, New Delhi, North Delhi, North East Delhi, North West Delhi, South Delhi, South West Delhi, West Delhi.

New Delhi presents itself as wonderful location for profitable businesses including restaurants.

1.2 Problem

A Food & Beverage entrepreneur is looking for a suitable location in Central Delhi to open a new restaurant to expand their business. Currently company has restaurants opened in other cities of India. Client is particularly interested to identify a suitable neighbourhood in Central Delhi for restaurant business.

I will focus this analysis on identifying the neighbourhoods where business will be good with less competition. Objective is to short list 4-5 such places so that client can decide the best once based on other factors.

1.3 Target Audience

Any entrepreneur interested in opening up a restaurant in Central Delhi

2. Data Acquisition and cleaning

2.1 Data Sources

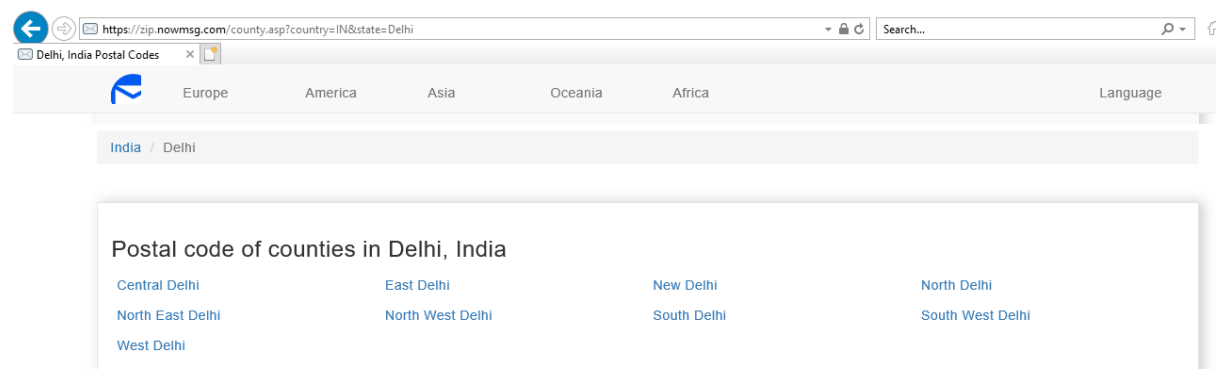
There were two main datasets that were used for this project.

First Dataset: List of all the neighbourhoods in Central Delhi

First Dataset: List of neighbourhoods in Central Delhi: The web page <http://zip.nowmsg.com> contains the postal code of each district and its corresponding location. I used BeautifulSoup to scrap the page and get a table with four columns: Ward, District, Latitude and Longitude. I will be using web scrapping tool BeautifulSoup for extracting the data in the form of a table from this page -

<https://zip.nowmsg.com/city.asp?country=IN&state=Delhi&county=Central%20Delhi>

Data available on this site contains lists of all places in Central Delhi and their Postal code, Place Name, Community, Country, State, Latitude, Longitude and a link to map.



Postal Codes Of Central Delhi, Delhi, India

A.G.C.R.	A.K.Market	Ajmeri Gate Extn.	Anand Parbat
Anand Parbat Indl. Area	Bank Street (Central Delhi)	Baroda House	Bengali Market
Bhagat Singh Market	Connaught Place	Constitution House	Dada Ghosh Bhawan
Darya Ganj	Delhi High Court	Delhi High Court Extension Counter	Desh Bandhu Gupta Road
Election Commission	Foreign Post Delhi IBC	Gandhi Smarak Nidhi	Guru Gobind Singh Marg
Hauz Qazi	I.A.R.I.	I.P.Estate	Inderpuri
Indraprastha H.O	Jama Masjid	Janpath	Karol Bagh
Krishi Bhawan	Lady Harding Medical College	Master Prithvi Nath Marg	Minto Road
Multani Dhanda	National Physical Laboratory	Nirman Bhawan	North Avenue
Pahar Ganj	Pandara Road	Parliament House	Patel Nagar (Central Delhi)
Patel Nagar East	Patel Nagar South	Patel Nagar West	Patiala House
Pragati Maidan	Pragati Maidan Camp	Rail Bhawan	Rajender Nagar
Rashtrapati Bhawan	Sansad Marg H.O	Sansadiya Soudh	Sat Nagar
Secretariat North	Shastri Bhawan	South Avenue	Supreme Court
Swami Ram Tirth Nagar	Udyog Bhawan	Union Public Service Commission	

[India](#) / [Delhi](#) / [Central Delhi](#) / [Darya Ganj](#)

Postal Codes Of Darya Ganj, Central Delhi, Delhi, India

Postal Codes	City / Place Name	Community	County	State	Latitude	Longitude	Map
110002	Darya Ganj	New Delhi	Central Delhi	Delhi	28.6453	77.2456	Map

Second Dataset: List of different venues in the neighbourhoods of Central Delhi:

This dataset will be formed using the Foursquare API. I will use the Foursquare location data to explore different venues in each neighbourhood of Central Delhi. Using the Foursquare location data, I can get information about these venues and analyse the neighbourhoods of Central Delhi easily based on this information.

I will be using these two datasets to solve the business problem of finding the best place to open a restaurant within Central Delhi, primarily by identifying the neighbourhoods in the Central Delhi, fetching the venues around them, using clustering to cluster these into multiple clusters and finally identify the cluster with less concentration of restaurants and further cleansing the identified locations to arrow down the target areas.

	Neighborhood	Neighborhood Latitude	Neighborhood Longitude	Venue	Venue Latitude	Venue Longitude	Venue Category
0	A.G.C.R.	28.6453	77.2456	Aap Ki Pasand	28.644786	77.240550	Food & Drink Shop
1	A.K.Market	28.6417	77.2132	The Drunkyard Cafe	28.641451	77.215506	Tibetan Restaurant
2	A.K.Market	28.6417	77.2132	The exotic roof top Restaurant	28.641039	77.213634	Indian Restaurant
3	A.K.Market	28.6417	77.2132	쉼터	28.641495	77.213152	Korean Restaurant
4	A.K.Market	28.6417	77.2132	Sita Ram Diwan Chand Chole Bhature	28.642324	77.210417	Food

2.2 Data Pre-processing

After obtaining the 2 datasets, pre-processing was needed so that it can be used in clustering algorithm. I pre-processed **Central_venues** data frame using **one-hot encoding** tool. The pre-processed data was stored in a data frame named **Central_onehot**.

It is ready to be used for clustering using K-Means technique. Next step was to further refine the data that belongs to Restaurant category and hence other venue related data was discarded from **Central_onehot** data frame. I filtered venues from **Central_onehot** data frame that are restaurants only and stored the output in a new data frame named **Central_restaurants**. This new data frame will now be used for clustering algorithm.

Also, a data frame named **venues_sorted** is created listing all the neighbourhoods of Central Delhi along with their respective 5 most common venues. This dataset would eventually help in visualising the solution. First 5 rows of this data frame are depicted in figure below:

	Neighborhood	1st Most Common Venue	2nd Most Common Venue	3rd Most Common Venue	4th Most Common Venue	5th Most Common Venue
0	A.G.C.R.	Food & Drink Shop	Tibetan Restaurant	Hotel	Food	Fast Food Restaurant
1	A.K.Market	Hotel	Indian Restaurant	Café	Hostel	Tibetan Restaurant
2	Ajmeri Gate Extn.	Food & Drink Shop	Tibetan Restaurant	Hotel	Food	Fast Food Restaurant
3	Anand Parbat	Hotel	Fast Food Restaurant	Tibetan Restaurant	Pizza Place	Food & Drink Shop
4	Anand Parbat Indl. Area	Ice Cream Shop	Chaat Place	Mobile Phone Shop	Pharmacy	Bar

3. Methodology

In **Central_restaurants** data frame, I also added a column containing total number of restaurants in that neighbourhood in order to help us in clustering using K-Means algorithm.

I used K-Means clustering algorithm to make 5 clusters of datasets using **Central_restaurants** data frame as input. After the clusters were made, I merged the first dataset and the **venues_sorted** data frame and inserted cluster labels also. The result data frame was named **Central_merged**.

	County	Place	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	Central Delhi	A.G.C.R.	28.6453	77.2456	1.0	Food & Drink Shop	Tibetan Restaurant	Hotel	Food	Fast Food Restaurant
1	Central Delhi	A.K.Market	28.6417	77.2132	3.0	Hotel	Indian Restaurant	Café	Hostel	Tibetan Restaurant
2	Central Delhi	Ajmeri Gate Extn.	28.6453	77.2456	1.0	Food & Drink Shop	Tibetan Restaurant	Hotel	Food	Fast Food Restaurant
3	Central Delhi	Anand Parbat	28.6431	77.2197	4.0	Hotel	Fast Food Restaurant	Tibetan Restaurant	Pizza Place	Food & Drink Shop
4	Central Delhi	Anand Parbat Indl. Area	28.6551	77.1833	1.0	Ice Cream Shop	Chaat Place	Mobile Phone Shop	Pharmacy	Bar

Next part was Analysis of each cluster to get the correct neighbourhoods. I calculated total number of neighbourhoods and total number of restaurants for each cluster. Then I calculated Restaurant/Neighbourhood ratio and found that this ratio was lowest for cluster with cluster label=1. Hence this cluster was chosen for further analysis.

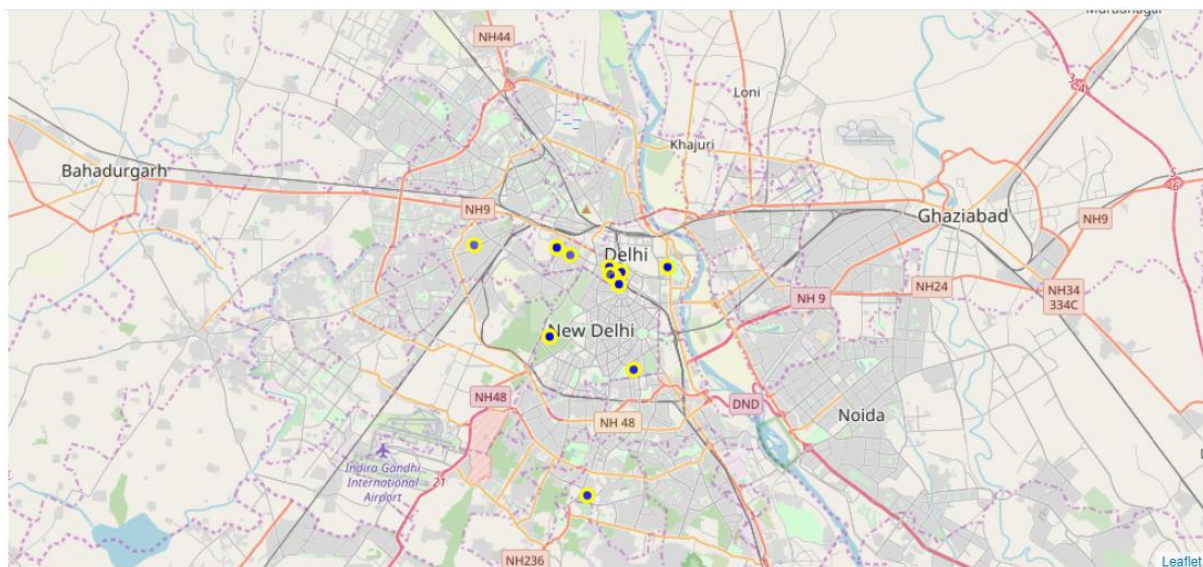
identified that cluster 1 has 16 neighbourhoods but just 1 restaurant. Further this list was drilled down based on areas where we have similar joints offering food like Hotels, Cafes etc. After dropping those venues, I narrowed down to 6 final venues suitable for restaurant opening with no competition.

The final dataset contains all the information about these remaining 6 neighbourhoods:

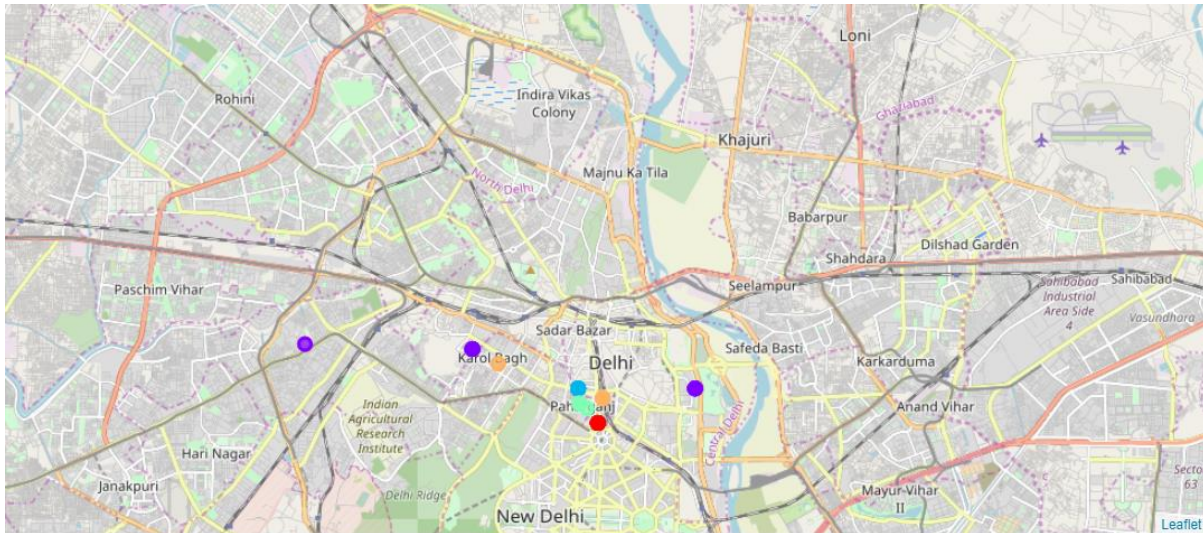
	County	Place	Latitude	Longitude	Cluster Labels	1st Most Common Venue	2nd Most Common Venue	3rd Most Common Venue	4th Most Common Venue	5th Most Common Venue
0	Central Delhi	Anand Parbat Indl. Area	28.6551	77.1833	1.0	Ice Cream Shop	Chaat Place	Mobile Phone Shop	Pharmacy	Bar
1	Central Delhi	Bank Street (Central Delhi)	28.6551	77.1833	1.0	Ice Cream Shop	Chaat Place	Mobile Phone Shop	Pharmacy	Bar
2	Central Delhi	Desh Bandhu Gupta Road	28.6551	77.1833	1.0	Ice Cream Shop	Chaat Place	Mobile Phone Shop	Pharmacy	Bar
3	Central Delhi	Guru Gobind Singh Marg	28.6551	77.1833	1.0	Ice Cream Shop	Chaat Place	Mobile Phone Shop	Pharmacy	Bar
4	Central Delhi	Master Prithvi Nath Marg	28.6551	77.1833	1.0	Ice Cream Shop	Chaat Place	Mobile Phone Shop	Pharmacy	Bar
5	Central Delhi	Sat Nagar	28.6551	77.1833	1.0	Ice Cream Shop	Chaat Place	Mobile Phone Shop	Pharmacy	Bar

4. Data Visualisation

A map of Central Delhi was generated using a visualisation library - **Folium**. All the neighbourhoods of Central Delhi have been marked on the map.



After using the clustering algorithm and creating 5 different clusters where each neighbourhood belong to one of these clusters, the new map of Central looked like this:



In the above map, 5 different colours, one for each cluster are used for representing each neighbourhood in Central Delhi.

5. Result

Foursquare data shows that there are many restaurants in Central Delhi to identify the areas with lower density, I used K-means clustering algorithm and segmented our neighbourhood dataset into 5 clusters. Based on the analysis, following outcomes were derived:

Total number of neighbourhoods in **cluster 0** is 18
 Total number of restaurants in this cluster is 216
 Ratio of Restaurant/Neighbourhood in this cluster is 12.0

Total number of neighbourhoods in **cluster 1** is 16
 Total number of restaurants in this cluster is 1
 Ratio of Restaurant/Neighbourhood in this cluster is 0.0625

Total number of neighbourhoods in **cluster 2** is 8
 Total number of restaurants in this cluster is 56
 Ratio of Restaurant/Neighbourhood in this cluster is 7.0

Total number of neighbourhoods in **cluster 3** is 5
 Total number of restaurants in this cluster is 46
 Ratio of Restaurant/Neighbourhood in this cluster is 9.2

Total number of neighbourhoods in **cluster 4** is 4
 Total number of restaurants in this cluster is 13
 Ratio of Restaurant/Neighbourhood in this cluster is 3.25

Based on Restaurant per neighbourhood ratio, cluster 1 seems to have lowest restaurant density. Cluster 1 has 16 neighbourhoods but just 1 restaurant. Hence cluster 1 is the perfect cluster to open a restaurant in any neighbourhood which I further narrowed down based on similar joints in the area.

6. Discussion

Further this list was drilled down based on areas where we have similar joints offering food like Hotels, Cafes etc. After dropping those venues, I narrowed down to 6 final venues suitable for restaurant opening with no competition.

	County	Place
0	Central Delhi	Anand Parbat Indl. Area
1	Central Delhi	Bank Street (Central Delhi)
2	Central Delhi	Desh Bandhu Gupta Road
3	Central Delhi	Guru Gobind Singh Marg
4	Central Delhi	Master Prithvi Nath Marg
5	Central Delhi	Sat Nagar

7. Conclusion

Objective of this project was to identify the neighbourhood in Central Delhi where restaurant competition is low making it an attractive location for new entrants. By deriving the restaurant density distribution based on Foursquare data and clustering technique, I have drilled down to the cluster with lowest density and have narrowed down to 6 neighbourhoods which are good for starting a new restaurant. Client can review the locations identified to further drill down based on factors like availability of places, rent etc. This concludes the project - The Battle of Neighbourhoods.