

Data Science Capstone Project

...

Battle of Neighborhoods

By Smriti Chaudhury

Introduction : Business Problem

In this project, the chosen city is New Delhi. One of the country's largest urban agglomerations, Delhi sits alongside the *Yamuna River*, the tributary of *Ganges River*, about 100 miles south of the *Great Himalayas*. Delhi is the administrative, legislative and judicial capital of India and surrounds the metropolitan region as well as surrounding rural regions.

I have lived in Delhi since my birth and have travelled to different places in the city. It comprises people coming from various other states having different religions who live in harmony which is one of the great things about the city. It has many landmarks from historical monuments like *Lal Quila (Red Fort)*, *Humayun's Tomb* to famous temples like *Lotus Temple*, *Akshardham Temple* and various famous attractions like *Agrasen Ki Baoli*, *National Museum*, *National Zoological Park*, *National Rail Museum* etc.,

Objective

The objective of this project is to choose the best neighbourhood where a Restaurant can be set up. Delhi is a hotspot for Continental, Thai, Mexican and Chinese food as well. Keeping all this in mind, the focus of this capstone project will be to:

- Choose a neighbourhood among the busiest ones popularity wise and distance wise.
- The stakeholders who want to open a different cuisine restaurant such as Italian, Thai or Mexican around famous attractions of the city would be interested in the results of this project.

Data Acquisition

To compute the dataset, the following data will be needed :

1. List of neighbourhoods in Delhi.

Source : https://en.wikipedia.org/wiki/Neighbourhoods_of_Delhi

2. Coordinates of all neighbourhoods and venues

Source : GeoPy Nominatim Geocoding

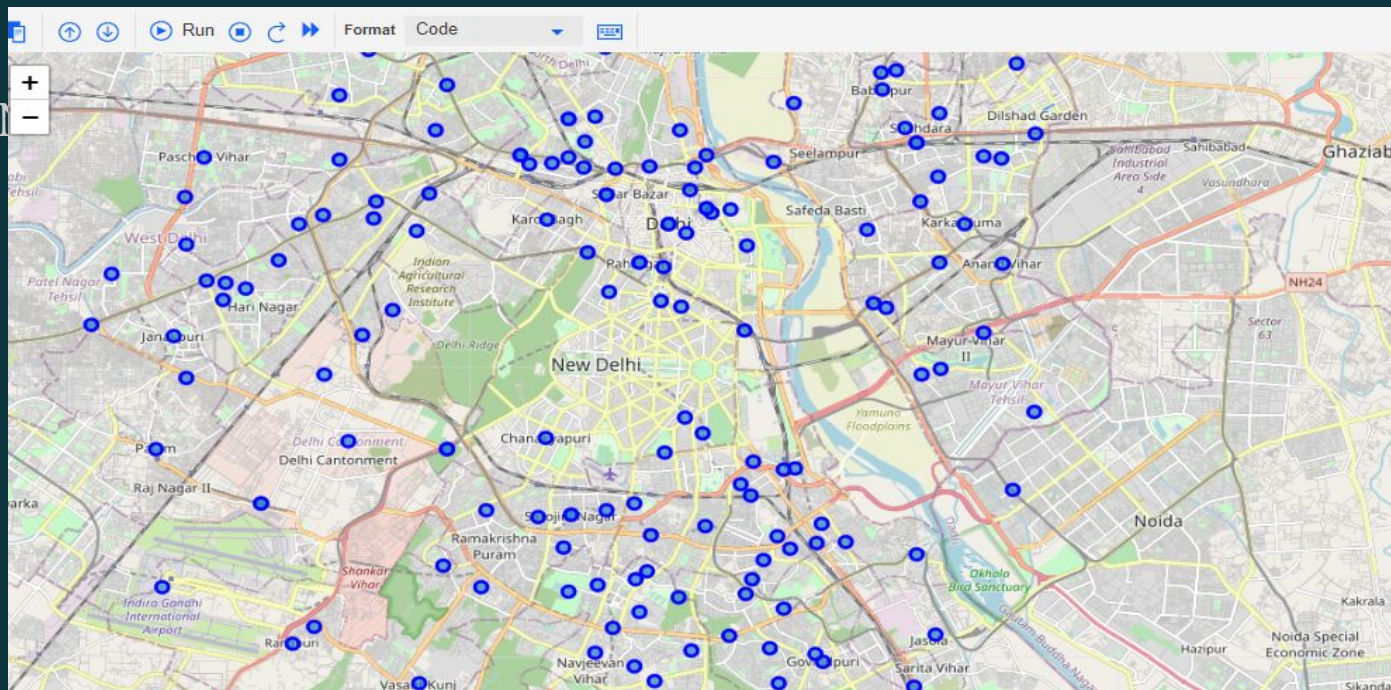
3. Number of restaurants and their type and location in every neighbourhood.

Source: Foursquare API <https://developer.foursquare.com/>

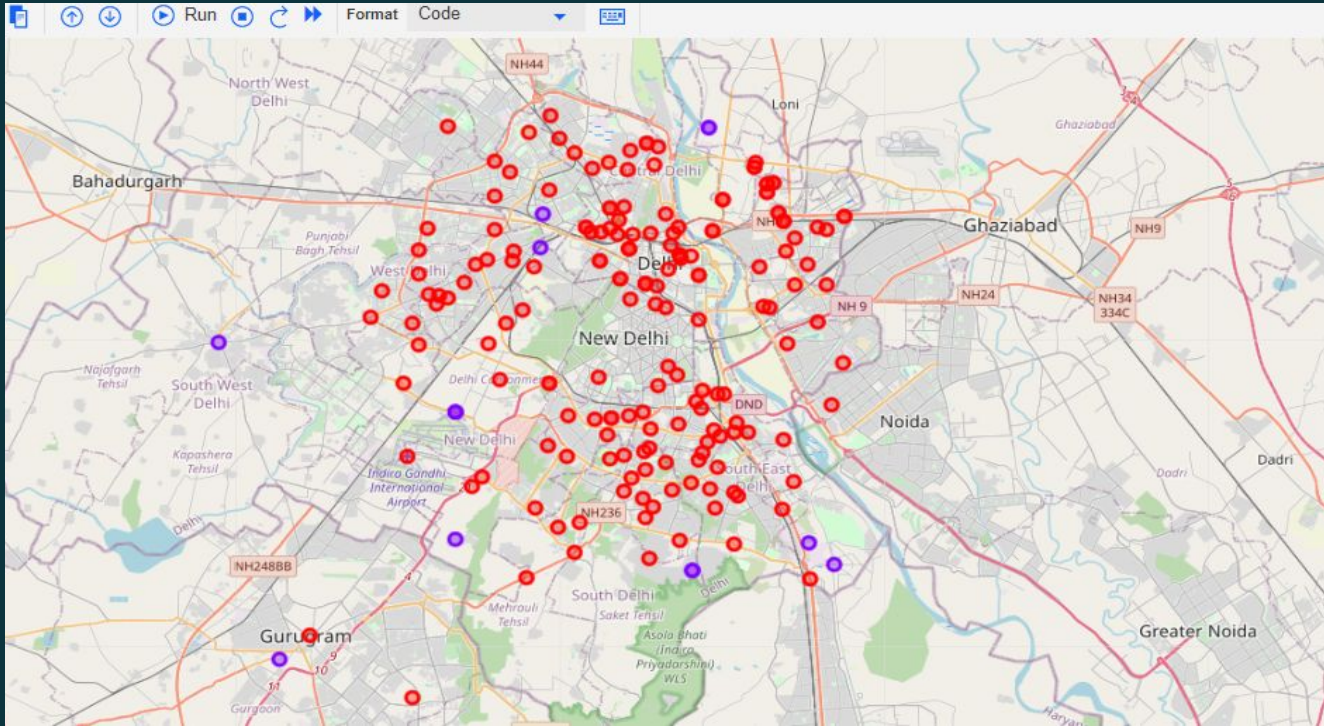
Methodology

1. Collect the data. Gathering data from webpages to build the dataset of neighborhoods in the city.
2. Using this data, a folium map of the Delhi neighbourhoods marked on it will be created.
3. Explore each of neighbourhoods and their venues using Foursquare location data.
4. Each cluster will be examined and a decision will be made regarding which cluster fits the shareholder's requirements.
5. Use a for loop to see what is the adequate number of clusters that should be formed by silhouette coefficient.
6. The ones that fit the requirements will be further explored and shortlisted based on how small their respective distances to the center or Delhi are.

Results

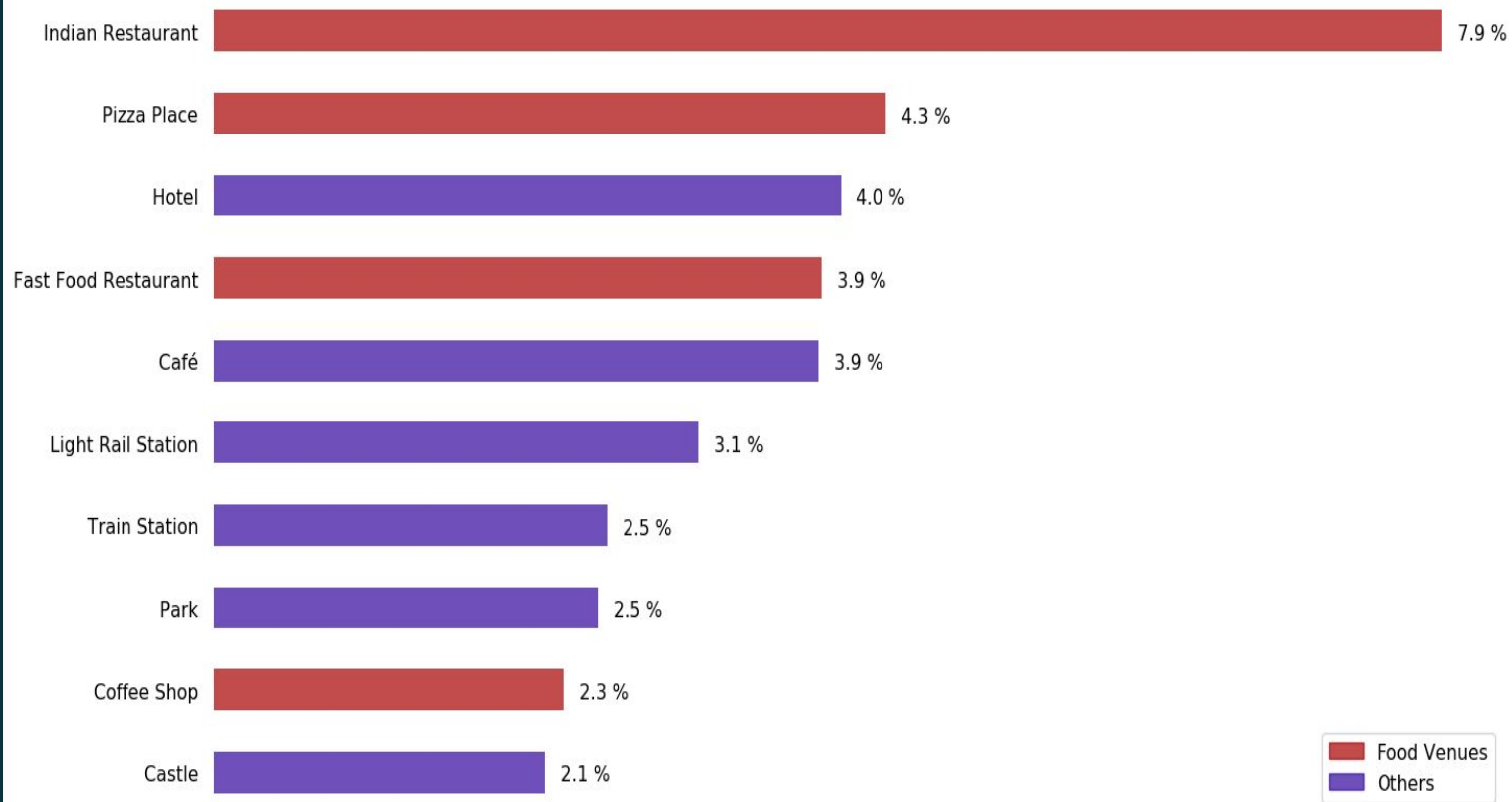


Map of Delhi with neighborhoods marked on it.



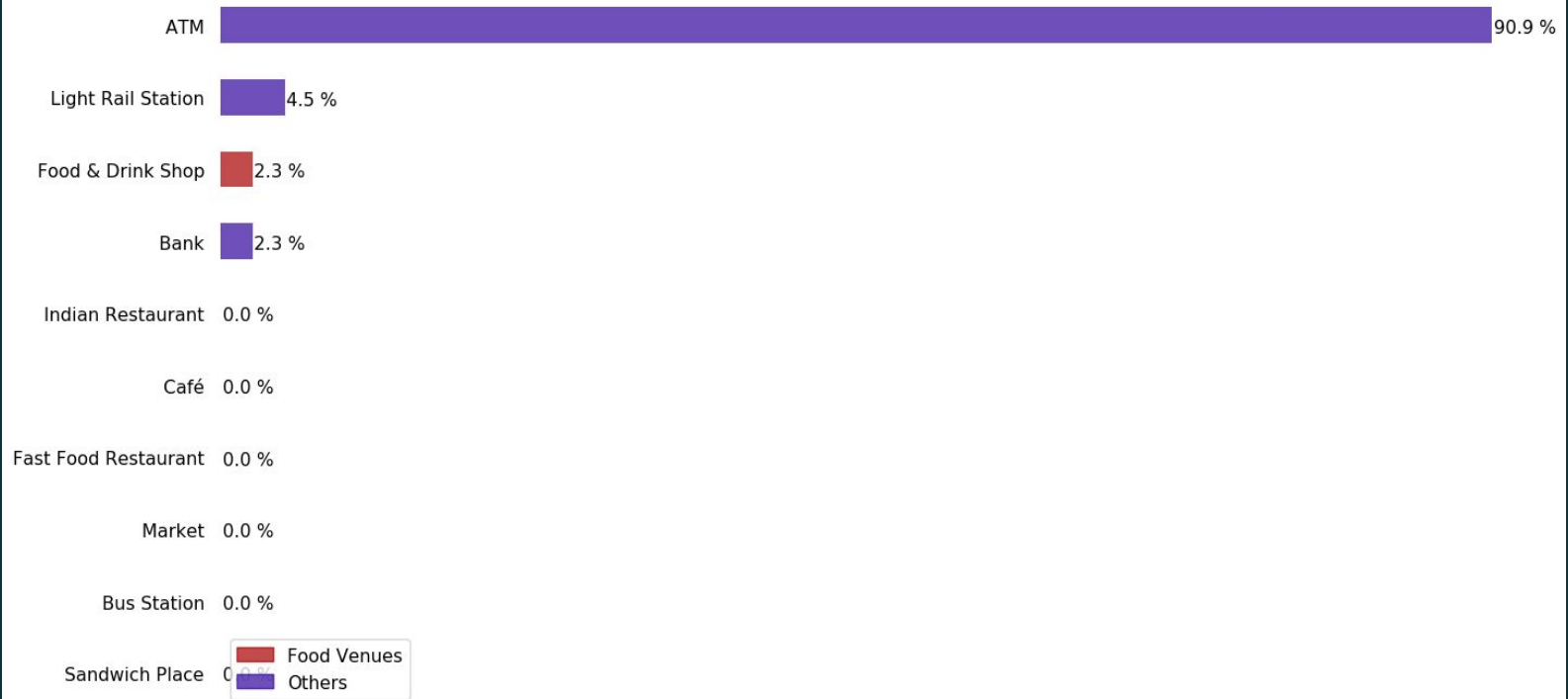
Map of Delhi with Clusters are color coded.

Ten Most Prevalent Venues of Cluster 0 (in % of all venues)



Ten Most Prevalent Venues of Cluster 1

(in % of all venues)



Discussion

Looking at the map of Delhi, Cluster 0 has a proximal number of neighborhoods and an efficient number of centres are there to attract many customers for setting up a multi cuisine restaurant. Even though Cluster 0 has restaurants and cafes but the main idea behind this is to set up a different cuisine restaurant like Italian, Mexican, Thai, or Mughlai. Top 5 venues are listed in Cluster 0 for each neighborhood and the neighborhoods which have many restaurants already are excluded.

The above figure shows all the neighborhoods in Cluster 0 and enables the stakeholders and businessmen to choose a location as per their convenience and requirements by looking at each neighborhood selected in Cluster 0 which have greater number of famous attractions, landmarks and most visited places to attend so the restaurant incomes can go up.

Conclusion

The objective of this project was to identify the best potential neighbourhoods in Delhi, India where a multi cuisine restaurant can be set up. All the required neighbourhood data was either scraped of the internet or obtained using a geolocator. After the neighbourhoods were visualized on a folium map, their venues were explored using Foursquare location data. Based on the frequency of occurrences of different venue types, the neighbourhoods were divided into two groups with the help of k-means clustering. The clusters were examined and the best one in which a restaurant could be set up was chosen. The neighbourhoods were filtered further based on proximity to existing restaurants and distance from the center of the city. As touched upon earlier, the results of the analysis highlight potential neighbourhoods where a multi cuisine restaurant may be opened solely based on geographical location and proximity to competitors.