

OPENING A NEW RESTAURANT IN DELHI, INDIA

PULKIT BHAMBRI

June 18, 2020

1. INTRODUCTION

1.1 Background

In India and especially Delhi, restaurants having Indian cuisine are the most popular. Delhi has people from different parts of India, thus the city has different types of food traditions; its cuisine is influenced by the various cultures. Punjabi cuisine is common, due to the dominance of Punjabi communities. Delhi cuisine is actually an amalgam of different Indian cuisines modified in unique ways. This is apparent in the different types of street food available. Indian restaurants in Delhi comprise of these uniquely modified Indian cuisines ranging from delicacies like Dal Makhani from Punjab to Biryani from Hyderabad. In the current scenario where people prefer to go out more opening a new Indian Restaurant in the neighbourhoods of Delhi can be very profitable.

1.2 Problem

This project aims to segment the neighbourhoods of Delhi on the basis of presence of Indian restaurants in them so that the perfect neighbourhoods for the opening of a new Indian restaurant can be determined where the restaurant won't get lost in a cluster of other Indian restaurants and also not face a lot of competition, hence being exclusive in the location and profitable for the owner.

1.3 Interest

Any business owner or an individual looking to open a new Indian restaurant in Delhi would be very interested in knowing the best location for his/her restaurant. Existing restaurant owners which aim to make more profits can use this model for relocation of their restaurants as well. This model will also show how Indian restaurants are the most common restaurants in Delhi.

2. DATA ACQUISITION AND CLEANING

2.1 Data sources

This project will require:

- Dataset containing the neighbourhoods of Delhi.
- Location data which can help us get venues nearby to each neighbourhood.

Dataset containing the neighborhoods of Delhi as well as their latitude and longitude coordinates can be found on Kaggle.com on the following link: https://www.kaggle.com/shaswatd673/delhi-neighborhood-data/data?select=delhi_dataset.csv

To get venues nearby to each neighbourhood we will leverage the Foursquare API. The Foursquare API will provide many categories of venue data, we are particularly interested in the Indian Restaurant category in order to help us to solve the business problem put forward.

2.2 Data cleaning

We find that the latitude and longitude information for 33 neighbourhoods is not available hence we will remove the rows associated with those neighbourhoods.

As the neighbourhood dataset already contains an index column we will have to remove that column named 'unnamed: 0' so that we have a single indexing for our dataset.

The Foursquare API call will return a cleaned dataset in a json file format which we will have to convert into a pandas data frame.

In the next sections we will use the cleaned datasets and process them further for modeling and applying visualization techniques to get clearer insights into the data we have.

3. METHODOLOGY

3.1 Exploratory Data Analysis

First of all we will need the dataset containing the neighbourhoods in Delhi. The mentioned dataset is available on Kaggle.com with the latitude and longitude coordinates of each neighbourhood as well. In the case where coordinate information is not available we can easily obtain it using the Open Code Geocode package. The data is in the form of a csv (comma separated value) file and in the next step is converted to a pandas data frame for further analysis. After populating the Pandas data frame we will plot these neighbourhoods on a map of Delhi using the Folium package. This helps us to check if the coordinate information in the dataset is correct or not. The plot is shown below:

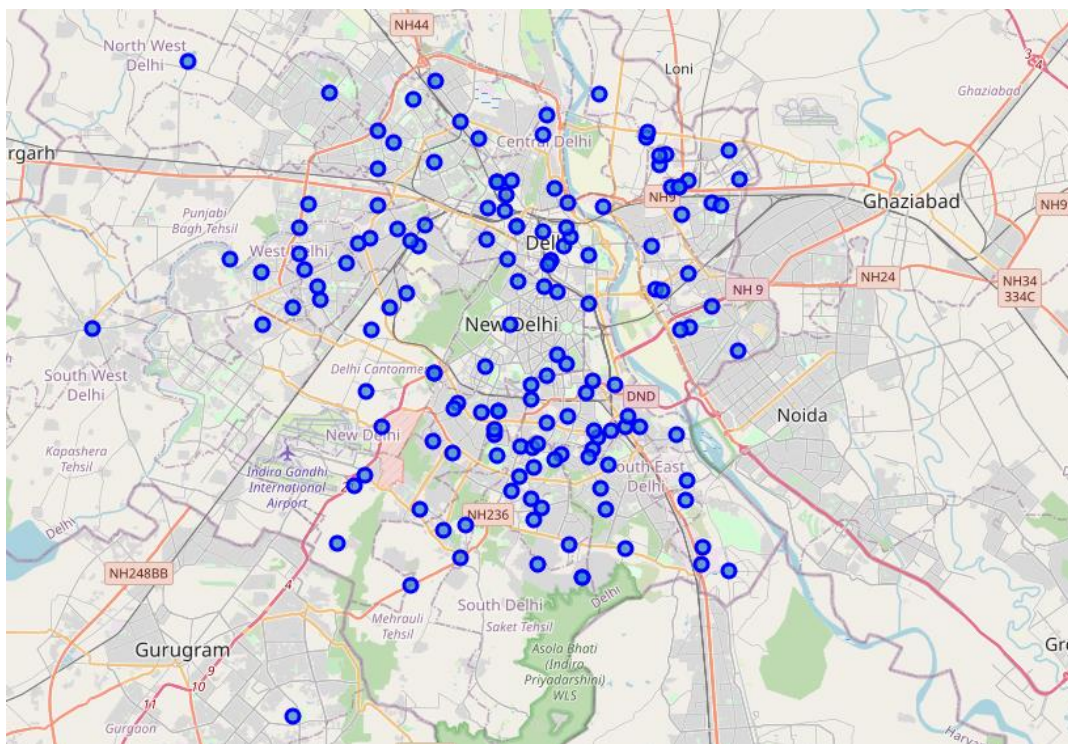


Fig.1 Plot of neighbourhoods of Delhi

Next we will leverage the Foursquare API to get the top 100 venues that are within a radius of 1000m from the neighbourhoods. We need to register a Foursquare Developer Account in order to obtain the Foursquare ID and Foursquare secret key. We then make API calls to Foursquare passing in the geographical coordinates of the neighbourhoods in a Python loop. Foursquare will return the venue data in JSON format and we will extract the venue name, venue category, venue latitude and longitude. With the data, we can check how many venues were returned for each neighbourhood and examine how many unique categories can be curated from all the returned venues. From this we can examine which venue is the most common in Delhi and it is Indian Restaurants with 316 instances. Then, we will analyse each neighbourhood by grouping the rows by neighbourhood and one hot encoding the venues by each neighbourhood and then taking the mean of the frequency of occurrence of each venue category. By doing so, we are also preparing the data for use in clustering. Since we are analysing the “Indian Restaurants” data, we will filter the “Indian Restaurants” as venue category for the neighbourhoods.

3.2 Machine Learning

The Machine Learning that is being applied is K-Means clustering which is a portioning algorithm. K-Means divides the data into K non overlapping clusters without any cluster internal structure. It tries to minimize the intra cluster distances and maximize the inter cluster distances. K-Means clustering is chosen here as it is one of the best clustering techniques and in this case we need to segment the neighbourhoods on the basis of occurrence of Indian Restaurants in them. We have the one hot encoded data of Indian Restaurants by neighbourhood. We will now apply the K-Means clustering algorithm to segment the data in 4 clusters (number of clusters can be determined by the elbow method which is a plot of within cluster sum of squares (WCSS) and number of clusters (K), the elbow point in the graph is selected as the correct number of clusters). Once we obtain the cluster labels we will add it to the data frame and visualize the map with clustered neighbourhoods.

RESULTS

The results from the k-means clustering show that we can categorize the neighbourhoods into 3 clusters based on the frequency of occurrence for “Indian Restaurants”:

- Cluster 0: Neighbourhoods with highest number of Indian Restaurants
- Cluster 1: Neighbourhoods with low number of Indian Restaurants
- Cluster 2: Neighbourhoods with moderate concentration of Indian Restaurants
- Cluster 3: Neighbourhoods with least number of Indian Restaurants

The results of the clustering are visualized in the map below with cluster 0 in red colour, cluster 1 in purple colour, cluster 2 in mint green colour and cluster 3 in yellow colour

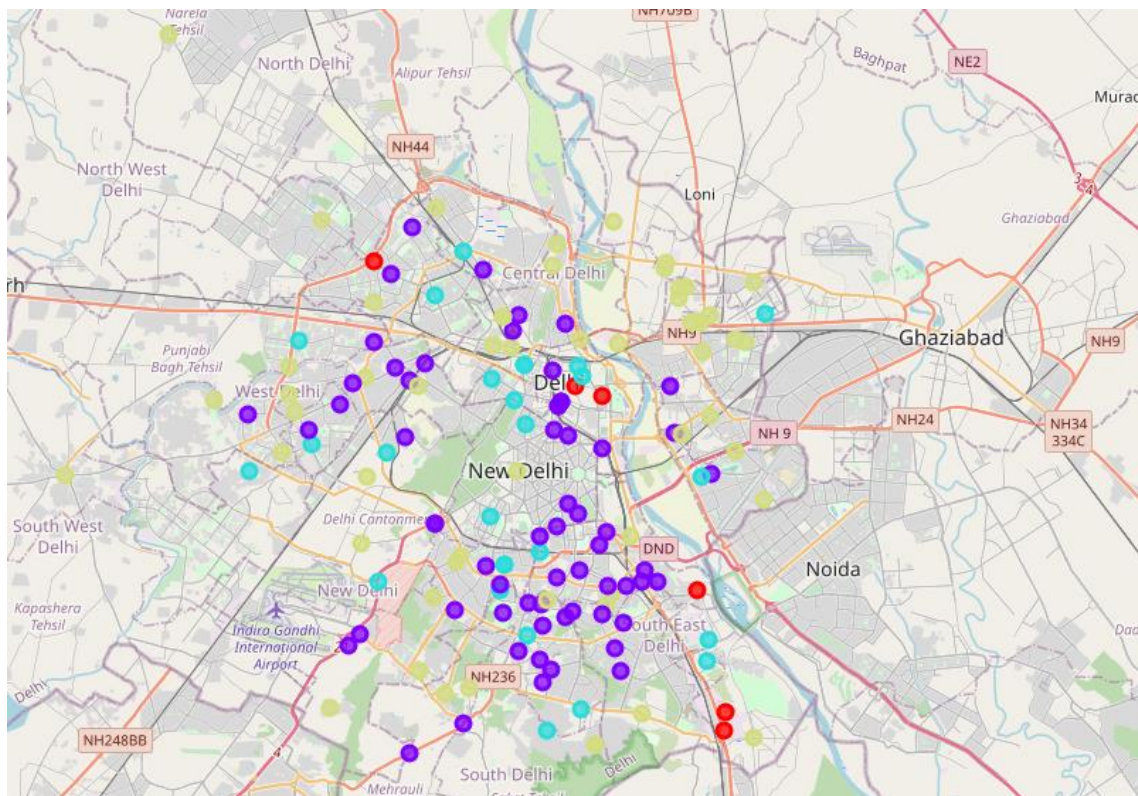


Fig.2 Plot of neighbourhoods clustered on the basis of concentration of Indian Restaurants in them.

Further we will examine each cluster and take the mean of “Indian Restaurants” column of each cluster to know what the clusters depict. The results are as follows:

- Cluster 0 : (Neighbourhoods with highest number of Indian Restaurants)

Mean = 0.376542

Colour code = Red

Neighbourhoods =

Badarpur
Chawri Bazaar
Daryaganj
Gautampuri
Okhla
Pitam Pura

- Cluster 1 : (Neighbourhoods with low number of Indian Restaurants)

Mean = 0.104003

Colour code = Purple

Neighbourhoods =

Alaknanda	Laxmi Nagar
Ashok Vihar	Lodi Colony
Barakhamba Road	Maharani Bagh
Chandni Chowk	Mahipalpur
Chandni Chowk	Malviya Nagar
Chhattarpur	Mayur Vihar
Chittaranjan Park	Moti Nagar
Civil Lines	Munirka
Connaught Place	Nehru Place
Defence Colony	Netaji Nagar
Dhaura Kuan	New Friends Colony
Dhaura Kuan	Nizamuddin West
Fateh Nagar	Paharganj
Friends Colony	Paharganj
Geeta Colony	Patel Nagar
Ghitorni	Pragati Maidan
Golf Links	Punjabi Bagh
Greater Kailash	Rajouri Garden
Green Park	Rangpuri
Gulmohar Park	Roshanara Bagh
Hauz Khas Village	Safdarjung Enclave
Inderpuri	Saket
Jangpura	Sarvodaya Enclave
Jor Bagh	Shahpur Jat
Kailash Colony	Shalimar Bagh
Kamla Nagar	Shivaji Place
Khan Market	Siri Fort
Khirki Village	South Extension
Kirti Nagar	Sriniwaspuri
Kohat Enclave	Vikaspuri
Lahori Gate	Wazirabad
Lajpat Nagar	

- Cluster 2 : (Neighbourhoods with moderate number of Indian Restaurants)

Mean = 0.202319

Colour code = Mint Green

Neighbourhoods =

Adarsh Nagar	
Anand Vihar	
Azadpur	
Bara Hindu Rao	
Bara Hindu Rao	Khanpur
Chanakyapuri	Kotwali
Dariba Kalan	Krishna Nagar
Dilshad Garden	Naraina
Gole Market	Palam
Hauz Khas	Paschim Vihar
INA Colony	Patparganj
Jasola	Sainik Farm
Jhandewalan	Sarita Vihar
Karol Bagh	Sarojini Nagar
Keshav Puram	Sarojini Nagar

- Cluster 3 : (Neighbourhoods with least number of Indian Restaurants)

Mean = 0.003222

Colour code = Yellow

Neighbourhoods =

Ashok Nagar	Mandoli	Sant Nagar
Babarpur	Maujpur	Sarai Kale Khan
Bali Nagar	Meera Bagh	Sarai Rohilla
Bawana	Mehrauli	Shahdara
Brij Puri	Model Town	Shahdara
Dayanand Colony	Moti Bagh	Shakarpur
Delhi Cantonment	Najafgarh	Shakti Nagar
Dhaka	Nand Nagri	Shakti Nagar
East Vinod Nagar	Narela	Shastri Nagar
East of Kailash	Naveen Shahdara	Shastri Nagar
Gandhi Nagar	Neeti Bagh	Shastri Park
Govindpuri	Nehru Nagar	Sonia Vihar
Jahangirpuri	Nehru Vihar	Tilak Nagar
Jaitpur	Pandav Nagar	Timarpur
Janakpuri	Pratap Nagar	Tughlaqabad
Jhilmil Colony	Preet Vihar	Vasant Kunj
Kabir Nagar	Pul Bangash	Vasant Vihar
Kalkaji	Rajokri	Vasundhara Enclave
Kashmiri Gate	Rama Krishna Puram	Vikas Nagar
Kingsway Camp	Rani Bagh	Vishwas Nagar
Kirby Place	Rithala	Vivek Vihar
Kishangarh Village	Sangam Vihar	Yamuna Vihar

5. DISCUSSION

Based on the results we can see that neighbourhoods in cluster 3 have the least amount of Indian restaurants so these neighbourhoods seem ideal for opening a new restaurant as it would not face competition from other restaurants of the same cuisine. Cluster 1 has a low number of Indian restaurants too and can be considered as well. Choice of clusters can depend on the availability of space as well as the type of population in those areas (financially and cuisine preference). Investors and business owners who are looking to open a new restaurant are advised to avoid the neighbourhoods of cluster 0 as they have the highest concentration of Indian restaurants.

The same analysis can be applied to any category of venues obtained from the Foursquare API be it a new Arcade or a Shopping Mall. We considered just one category here but the project can be used with multiple venue categories as well for instance we can use all the venue categories which belong to restaurants and analyse the concentration of restaurants in Delhi irrespective of their cuisine.

6. CONCLUSION

In this project, we have gone through the process of identifying the business problem, specifying the data required, extracting and preparing the data, performing machine learning by clustering the data into 4 clusters based on their similarities, and lastly providing recommendations to the relevant stakeholders i.e. business owners and investors regarding the best locations to open a new Indian restaurant.

To answer the business question that was raised in the introduction section, the answer proposed by this project is: The neighbourhoods in cluster 3 are the most preferred locations to open a new Indian restaurant.

The findings of this project will help the relevant stakeholders to capitalize on the opportunities on high potential locations while avoiding overcrowded areas in their decisions to open a new Indian restaurant.