

IBM APPLIED DATA SCIENCE CAPSTONE

OPENING A CAFÉ IN DELHI, INDIA

BY: HRITHIK SHUKLA

JUNE 2020



INTRODUCTION:

Delhi is one of the major cities of India and it has one of the most vibrant varieties of café in the world. This city is famous for its food and foodies, you can say food is really important to the people of Delhi. This means the café business flourishes in this city, making it one of the most attractive destinations for restaurant franchises and entrepreneurs.

BUSINESS PROBLEM:

Delhi is a big city and has a ton of options when it comes to café places. For anyone new to the café business it will be difficult survive as there is a lot of competition in this market.

If an entrepreneur wants to open a café in Delhi at what places should he/she open his/her business so that they face the least competition. There are lots of neighborhoods in Delhi we will have to conduct analysis to see which area can be considered for opening a café. We can leverage the foursquare data by calling data of all the café situated in the specific neighborhoods.

TARGET AUDIENCE:

This will help anyone who wants to open a café in Delhi or wants to expand their franchise in Delhi. This will be beneficial to any entrepreneurs who are looking to invest in the restaurant business in Delhi. It will help them identify the best to open their business with less competition in their neighborhood.

DATA:

For solving this problem, we need the following data:

- List of the entire neighborhood in Delhi area. As the scope of this project is confined only to the Delhi region, it is the capital of India.
- We are going to need the latitude and longitude coordinates of all the neighborhoods in Delhi for plotting the map of neighborhoods.
- The list of all the cafés pertaining to the specific neighborhoods will be required to analyze the competition in the neighborhood.

Sources of Data:

The list of the neighborhood in Delhi is been obtained from the website 'www.kaggle.com', **Kaggle** is the world's largest data science community with powerful tools and resources to help you achieve your data science goals. On Kaggle people can find a lot of useful data sets. I even found the latitude and longitude data of the neighborhoods on Delhi on Kaggle itself.

Now coming to the venues pertaining to neighborhoods, I used the Foursquare API to call venues in the neighborhoods. **Foursquare City Guide**, commonly known as **Foursquare**, is a local search-and-discovery mobile app developed by Foursquare Labs Inc. the app provides personalized recommendations of places to go near a user's current location based on users' previous browsing history and check-in history.

Foursquare has one of the largest databases of 105+ million places and is used by over 125,000 developers.

METHODOLOGY:

- **Importing neighborhood data and performing data wrangling:**

The neighborhood data of Delhi was taken from **Kaggle**. The data was already in “.csv” format so we just had to import using panda. Then the data was put into a panda DataFrame. The data had a lot of null values so it had to be cleaned and prepared for the analysis.

- **Latitude and longitude of neighborhoods:**

The data already had the latitude and longitude of the neighborhoods so we did not need to fetch it from elsewhere. Coordinates were needed to call venue data of different neighborhoods from the foursquare API.

- **Visualizing the data:**

The folium library was used to visualize the various neighborhoods of Delhi on the map of Delhi. The coordinates were superimposed on the map of Delhi using the Folium library. This allowed us to check that all the coordinates of the data are correct according to the neighborhood.

- **Calling foursquare API to get venue data:**

So now we use foursquare API to call venues of the different neighborhoods within 2000 meter (2km) radius. A foursquare developer account is needed to call the venue data from foursquare API. We need to create an app in the account so that we can call data from it. Now we get client ID and client secret from the app created in the developer account. Then we make calls to the Foursquare API in a loop to get all the venues pertaining to the certain neighborhoods in Delhi.

- **Analyzing the venue data and preparing data for clustering:**

The foursquare API gives us various categories of venues pertaining to certain neighborhoods. We can count the number of venues in each category. We can check how many venues were returned for each neighborhood. Then we group the rows according to the neighborhoods and take mean of the frequency of occurrence of each venue category. We also perform one hot encoding of the data. Now as the scope of this project is limited to cafés so we filter the café data as venue category for neighborhoods. By doing all this we are actually preparing data for clustering.

- **Performing clustering:**

Now we perform clustering using the k-means algorithm. K-means clustering is a method of vector quantization, originally from signal processing that aims to partition n observations into k clusters in which each observation belongs to the cluster with the nearest mean serving as a prototype of the cluster. The neighborhoods are clustered into 3 clusters based on frequency of occurrence of “café” in the neighborhoods. The result will help us to analyze which area has higher concentration of “café” and which area has lower concentration of café. This will help us solve our business problem; it will tell us which area is best suited to open a new café.

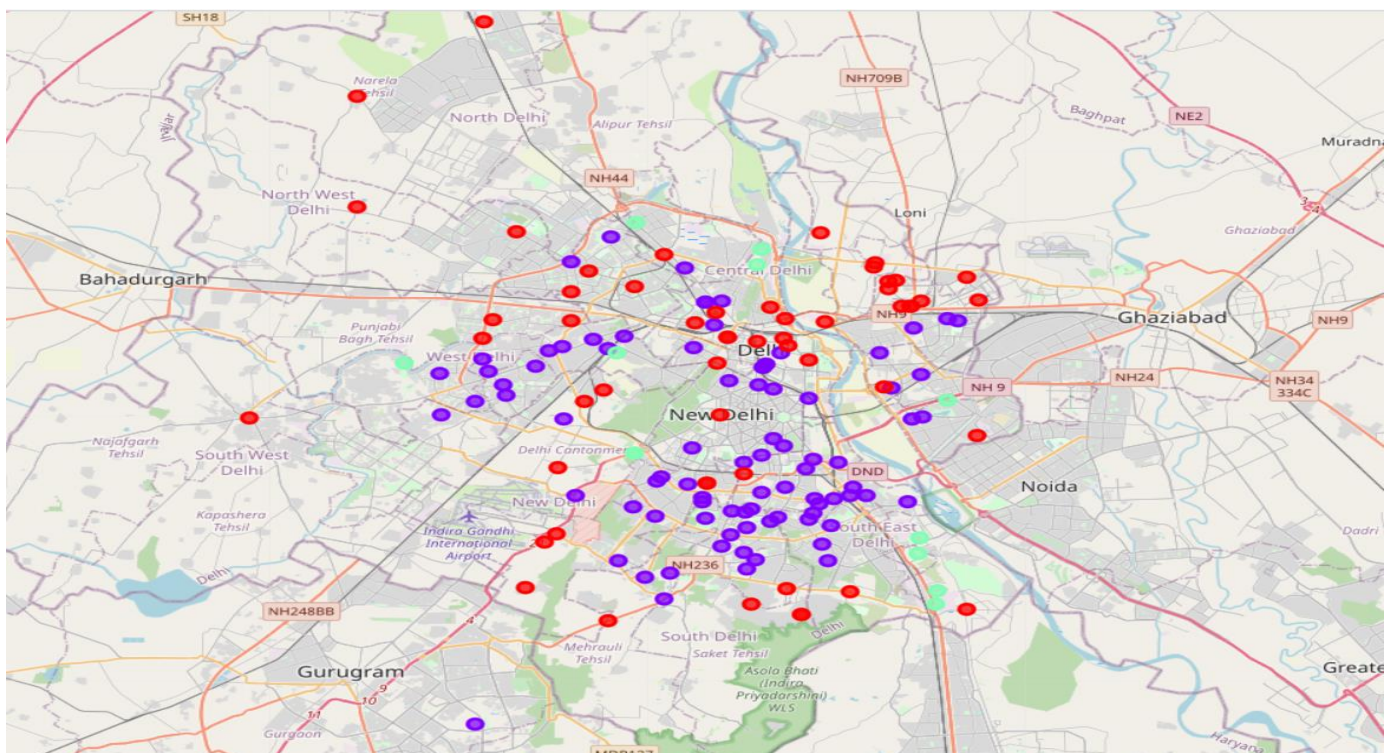
Result:

The neighborhoods were clustered using the K-Means clustering algorithm into 3 clusters according to the frequency of occurrence of “café” in the respective neighborhoods.

After examining the clusters:

1. **Cluster 0:** The number of cafés in these neighborhoods is moderate. It has a moderate level of competition.
2. **Cluster 1:** It can be seen that this cluster of neighborhoods has the highest no. of cafés. Hence, the completion in these neighborhoods will be the highest amongst the 3 clusters.
3. **Cluster 2:** It can be seen that in this cluster the number of cafés in these neighborhoods is low. Hence, It has low level of competition.

Below map shows the 3 clusters. Cluster 0 is purple, Cluster 1 is mint green and Cluster 2 is red.



Discussion:

We observe the map from the result section, most concentration of cafés in Delhi is situated in cluster 1, and cluster 0 has moderate number of cafés. On observing cluster 2 it has the lowest number of cafés in the neighborhoods. We can see that cluster 2 will be the best neighborhoods to open cafés as it has the least concentration of cafés in its neighborhoods. While in cluster 1 there will be a lot of competition as there is a lot of concentration of cafés in these neighborhoods. Therefore this project recommends entrepreneurs and café franchises to make use of these findings and open cafés in the neighborhoods of cluster 2, which has the least number of cafés in its neighborhoods. Entrepreneurs who can handle a little competition can open café in the cluster 0 which has a moderate level of competition. It is advised to avoid cluster 1 if someone wants to open a café as its neighborhoods have a lot of competition due to saturation of market.

CONCLUSION:

During this project, we have identified the business problem clearly, the data required was specified also extraction and cleaning of data was carried out, performing machine learning by clustering the data into 3 clusters, on the basis of their similar features. Then appropriate recommendations were provided to the target audience of the project, which were café franchises and entrepreneurs who want to open a café. After the analysis it was found that cluster 2 is the best suited cluster for opening new café, the neighborhoods in this cluster have a very low level of café concentration therefore providing low competition to a new café. These findings will help our target audience to make a knowledgeable decision on how they can avoid high competition and choose a place with low to moderate level of competition.

References:

1. <https://www.kaggle.com/shaswatd673/delhi-neighborhood-data>
2. <https://foursquare.com/developers/apps>
3. <https://traveltriangle.com/blog/best-cafes-in-delhi/>
4. <https://www.scoopwhoop.com/27-instagrammable-cafes-in-delhi/>