Opening a new restaurant in Delhi, India

Satish Agrahari

May 2020

1.Introduction

1.1. Business Problem

The objective of this project is to analyze the neighborhoods of Delhi to select the best location to open a restaurant.

1.2. Background

While Delhi has always been a food-loving with each region having its own special cuisine, they have never been very big on eating out in the past. But all that is changing now. The restaurant industry in Delhi has been growing at a rapid pace over the last decade or so and the growth story is set to continue for the next foreseeable future. Its increasing population, rapid urbanization, and growing awareness of western lifestyles and higher disposable income were some of the factors that contributed to the growth of the restaurant industry. Property developers, online food platform companies are taking advantage of this trend to build new restaurants to meet the demands. However not every restaurant marked success. There are few considerations particularly the location of the restaurant that determines whether the mall will be a success or failure

1.3. The target audience of this project

- Property Developers and investors who are looking to invest in Restaurants in Delhi.
- Those Online Food Platform Companies who are planning to invest in new restaurants in Delhi to improve their performance.
- Chefs or potential public who are aiming to open a new restaurant in Delhi.

1.4. *Using Data Science methodology techniques we will answer the following questions:*

- How is the frequency distribution of restaurants across Delhi?
- What is the effect of other venues on the frequency of restaurants in a locality?

2. Data acquisition and cleaning

2.1. Data Required

- List of neighborhoods in Delhi, the capital city of India.
- Latitude and longitude coordinates of each neighborhood. These are needed in order to plot the neighborhoods on the map and also to get the venues in each neighborhood.
- Venue Data, which will be used to perform clustering on neighborhoods.

2.2. Sources of Data and its extraction

- The neighborhood list data can be found in the following Wikipedia page link (https://en.wikipedia.org/wiki/Neighbourhoods_of_Delhi). By web scraping we can extract the data from the page, using python-requests and BeautifulSoup packages.
- The latitude and longitude (geographical coordinates) data of the neighborhoods can be extracted using the geocoder package of python.

However a very few neighborhoods geographical coordinates are not found using geocoder package which can be separately found

- Lastly we will use FourSquare API to get venue Data of neighborhoods. Foursquare has one of the largest databases of more than 105 million places worldwide.

2.3. Data cleaning

Data scraped from Wikipedia about neighborhoods of Delhi needed to be cleaned. There was a lot of unwanted data present. Using BeautifulSoup library I cleaned the data and created a dataframe containing boroughs and neighborhoods of Delhi.

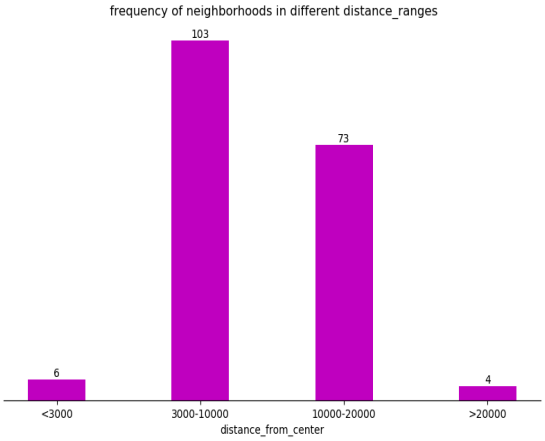
Geographical coordinates data obtained through the Geocoder library contained some outliers, i.e. coordinates that do not lie in Delhi. Using the maximum and minimum of Delhi's coordinates(lat, long) I found the wrong coordinates. However their quantity was low, hence I used Google search to find their real coordinates.

3. Methodology

Descriptive Data Analysis and modeling

3.1. Distribution of neighborhoods into different ranges

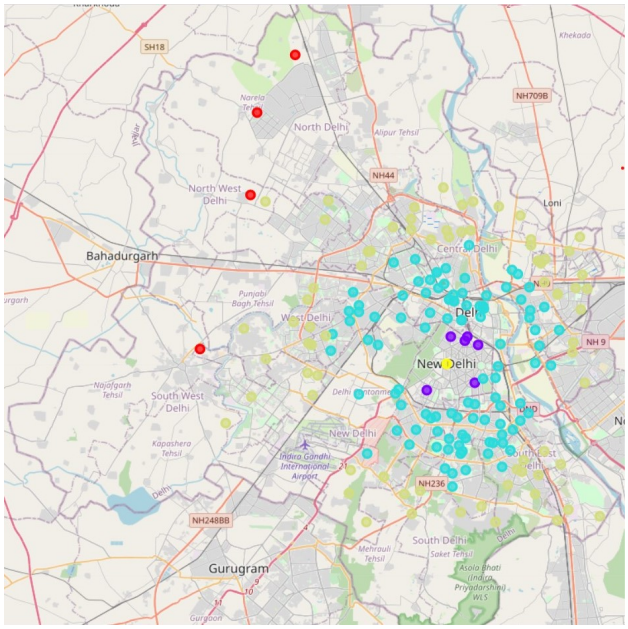
I distributed the neighborhoods into 4 groups ('<3000m', '3000-10000m', '10000-20000m', '>20000m'). Each group shows the number of neighborhoods in that range.



It was found that there are very few neighborhoods that are very close or very far from the city center. Most of them either lied in the range 3000-20000m.

We chose the neighborhoods lying in the range 0-10000m as these neighborhoods have easy access to the city(easy labor and

groceries), strong socio-economic dynamics, high population density.

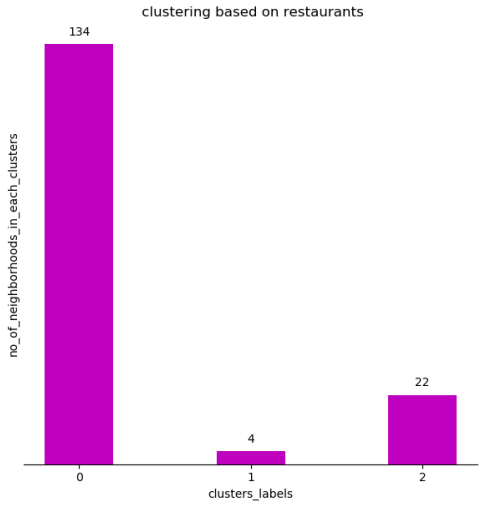


Here is the map showing the distribution of the neighborhoods into 4 clusters. The red ones are extremely far from the city center hence we excluded them from the test.

Also for final consideration, we took only groups 1 and 2, (i.e. 0-10000m range) for finding the best neighborhoods.

3.2. Distribution of neighborhoods into different clusters based on the number of restaurants

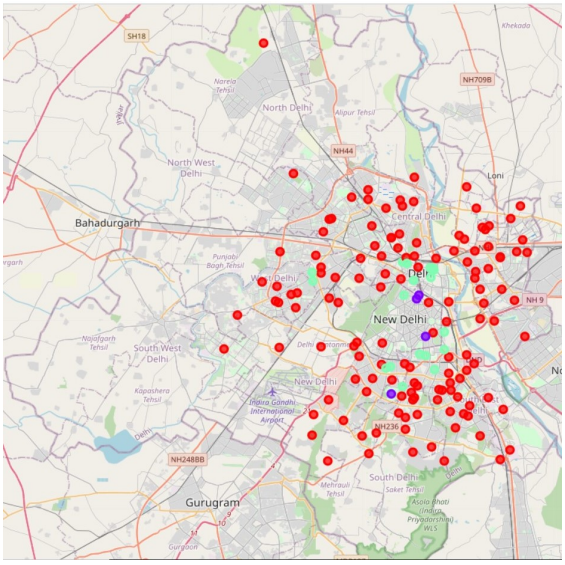
Using the KMeans clustering machine learning tool I distributed the neighborhoods into 3 clusters based on the frequency of restaurants in each neighborhood.



Here ‘cluster 0’ implies a low number of restaurants, ‘cluster 1’ implies a very large number of restaurants, and ‘cluster 2’ implies average number of restaurants.

It can be seen that most of the neighborhoods belong to ‘cluster 0’, which implies that there is a lot of scope for restaurants in Delhi.

.

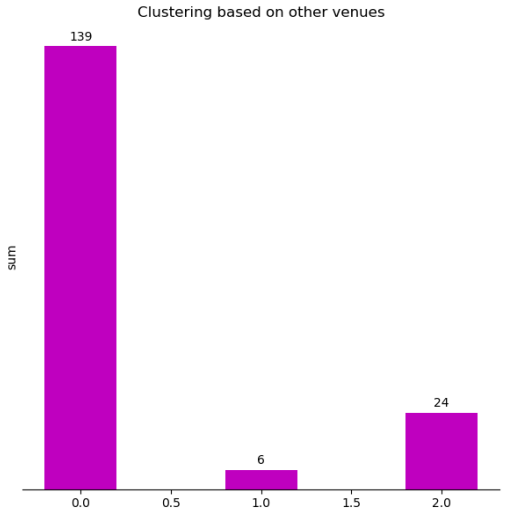


From the map, it can be seen that most of the high restaurant populated neighborhoods lie near the city center.

We selected the neighborhoods that belong to ‘cluster 0’ because we need little to no restaurants in the selected neighborhoods.

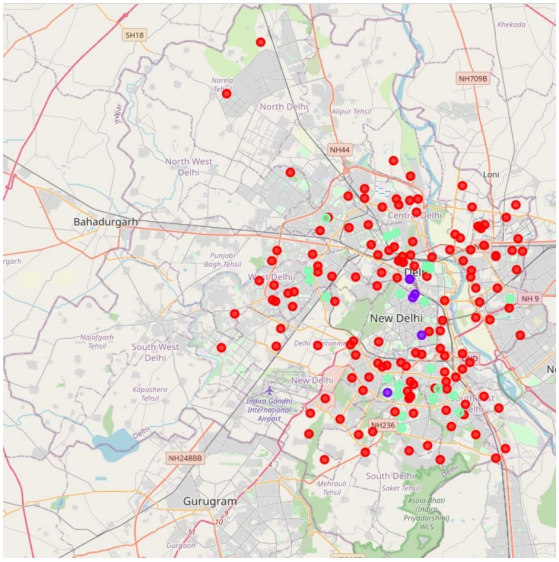
3.3. Distribution of neighborhoods into different clusters based on the number of other venues(other than restaurants)

Again using the KMeans clustering machine learning tool I distributed the neighborhoods into 3 clusters based on the frequency of venues in each neighborhood.



Here ‘cluster 0’ implies a low number of other venues, ‘cluster 1’ implies a very large number of other venues, and ‘cluster 2’ implies average number of other venues.

From the bar plot, we can see that most neighborhoods have a very low number of other venues, which is not so good because it shows low socio-economic dynamics.



Similar to restaurants clustering here also most of the densely populated areas lie close to the city center.

We selected the neighborhoods that belong to ‘cluster 1’ or ‘cluster 2’ because we need neighborhoods with high numbers of venues.

Collectively we selected the regions that followed all the criteria.

4. Result and Discussion

Our analysis shows that although there is a great number of restaurants in Delhi, there are potential areas that have low restaurant density. We found **11 neighborhoods** with a low number of restaurants and a high number of other venues within 10km range from Delhi center. Those are the potential neighborhoods based on the criteria we chose which offer a combination of **high population, closeness to the city center, strong socio-economic dynamics, *and* a number of pockets of low restaurant density**.

We found 11 neighborhoods namely **Naraina, Greater Kailash, Nehru Place, Hauz Khas, Malviya Nagar, Barakhamba Road, Safdarjung Enclave, Saket, Kamla Nagar, Shakti Nagar, Kailash Colony**, which fulfilled all the criteria.
This, of course, does not imply that those areas are actually optimal locations for a new restaurant! Purpose of this analysis was to only provide info on areas close to Delhi center but not crowded with existing restaurants - it is entirely possible that there is a very good reason for a small number of restaurants in any of those areas, reasons which would make them unsuitable for a new restaurant regardless of lack of competition in the area. Recommended zones should, therefore, be considered only as a starting point for more detailed analysis which could eventually result in a location that has not only no nearby competition but also other factors taken into account and all other relevant conditions met.

5. Conclusion

The purpose of this project was to identify Delhi neighborhood areas close to the center with a low number of restaurants in order to aid stakeholders in narrowing down the search for the optimal location for a new restaurant. By calculating restaurant density distribution from Foursquare data and distances from the city center using pyproj we created parameters to decide the optimality of a location. Clustering of those locations was then performed in order to find the best locations that can be used as starting points for final exploration by stakeholders.

The final decision on optimal restaurant location will be made by stakeholders based on specific characteristics of neighborhoods and locations in every recommended zone, taking into consideration additional factors like the attractiveness of each location (proximity to park or water), levels of noise/proximity to major roads, real estate availability, prices, social and economic dynamics of every neighborhood, etc.