# Recommendation of a location to open a new Restaurant in Kolkata

Keshab Das

12th April, 2020

## 1. Introduction

### 1.1. Background

Kolkata (formerly Calcutta) is the capital of India's state West Bengal and it is one of the oldest metro cities in India. Kolkata used to be India's capital under the British rule from 1773–1911. Being a heritage city, it holds rich value in literature and artistic thought and it is called the 'City of Joy'. There are huge number of schools, colleges, industries present in the city and people from around the world visit the place to see the beauty of the city. The native people known as Bengali are very fond of foods and sweets. As Food is one of the 3 daily essentials for human being, opening a restaurant in this city is a great choice of business. We can find restaurants of most of the popular cuisines in this city. So, for opening a restaurant in this city we should know the best location to open it keeping in mind that there should be very few numbers of restaurants nearby and the place is a popular one to attract customers.

### 1.2. Problem

The main objective of this project is to determine a location in city Kolkata suitable for opening a new restaurant. Here we will determine the location based on mainly two criteria: The location should be close to the centre of the City and there should be very few or no restaurants nearby.

### 1.3. Interest

Any individual who is interested to open a new restaurant or any food company who is interested to open a new branch will be the audience for this project. Choosing the best location to open a new restaurant will surely be able to give a boost to the business.

## 2. Data

### 2.1. Data Sources

Based on the definition of our business problem we will be requiring the data of different locations of city Kolkata. To find out the prime and popular locations of Kolkata based on their pin code we have used the below website resources:
https://www.mapsofindia.com/
http://pincode.india-server.com/

From the above resources we have created an CSV file of all popular locations of Kolkata along with Pin codes and stored the data in google drive that it can be used easily in our project. The link to the CSV file is:

https://drive.google.com/uc?export=download&id=1qCRjM5S6RzSNFT1MvUAszuBTxBlSZEvI

Below is a snapshot of the data:

| City | PostalCode | Location |
|---|---|---|
| Kolkata | 700020 | A.J.C.Bose Road |
| Kolkata | 700046 | Abinash Chaowdhury Lane |
| Kolkata | 700005 | Ahiritola |

Besides the above data, following data sources will be needed to extract/generate the required information:

- **Nomination Geocoder** to get the latitude and longitude information of all the locations including the centre of the City Kolkata.
- **Foursquare API** to get all the restaurants located around the different locations of City Kolkata.

## 2.2. Data Cleaning

We already have collected the data for different locations of city Kolkata based on different postal codes. So, now we will try to get the latitude & longitude details of the locations using Nomination. After populating the latitude & longitude details for all the locations we could find that for some of the locations/postal code the latitude/longitude data is not populated. So here we need to drop/eliminate the location records for which the latitude/longitude data is not populated as it will be an issue for our later analysis process.

As per our described problem we need to select the best location depending upon its distance from the centre of the city. So, after populating the latitude/longitude details we need to calculate the distance of each and every location from the centre of the city as it will be a necessary feature for our data analysis.

Now using Foursquare API, we will be exploring all the nearby venues for all of our locations to get the current existing restaurants data. Here we will try to explore every location with a radius of 1000m that no venue gets missed and we can cover up most of the area of the city. Here we will be working with only the restaurant's data so after exploring all the locations we need to eliminate the other venues except restaurants.

As we have taken the radius to explore the locations as 1000m, there may be duplicate restaurants present in our data. So, we will be removing the duplicate restaurant records from our data and will keep only those location records for which the restaurant is the nearest. No finally we will get our desired data frame with all the restaurant's data along with their nearest locations and the location's distance from the centre of the city Kolkata.

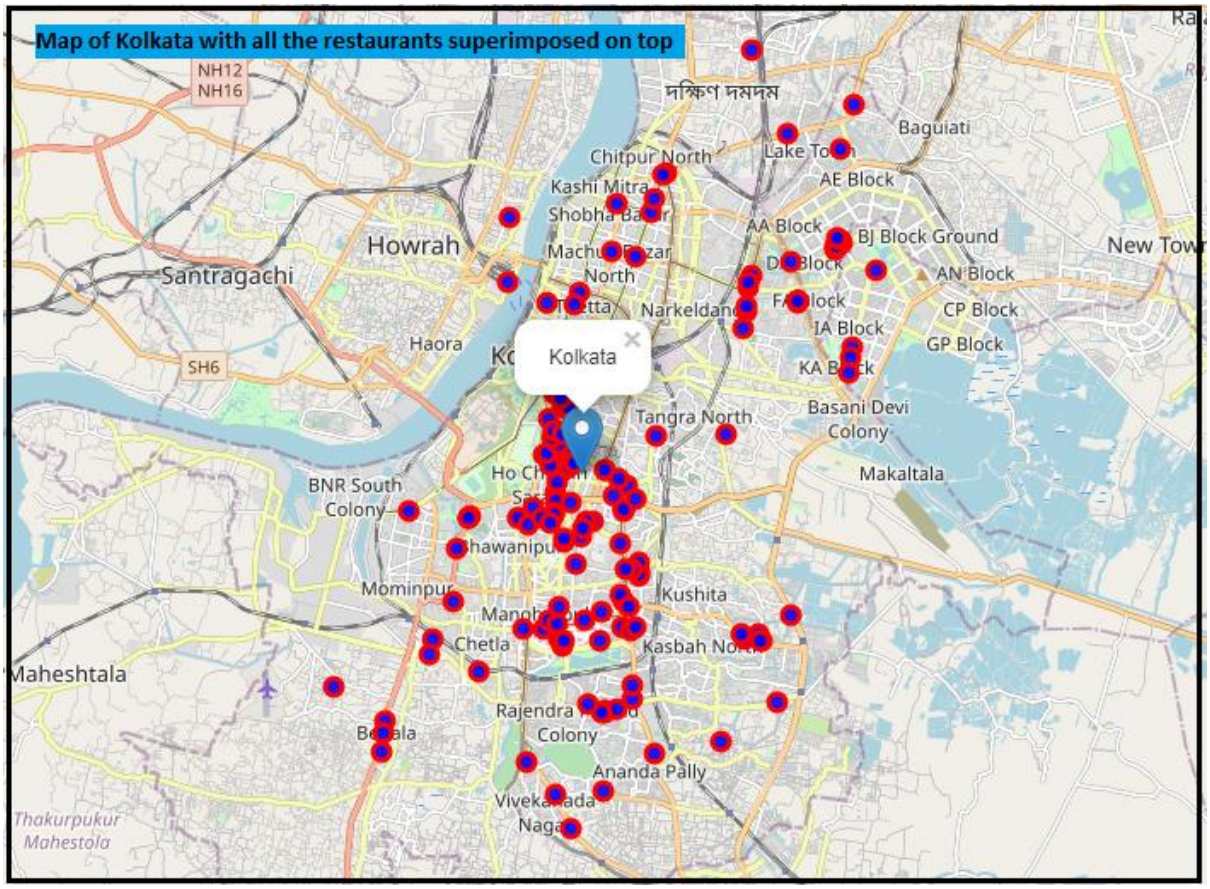
## 2.3. Feature Selection

After collecting all the required data and cleaning it up, we will have the all the required features present which we will be using for our further analysis. The main features that we will be using for our analysis and visualizations are: Location, Location latitude, Location Longitude, Venue, Location's Distance from centre.

# 3. Methodology

## 3.1. Exploratory data analysis

After preparing the data and extracting the required features from it we will get all the restaurants' data with their latitude and longitude details. So first we visualized the data by creating a map of city Kolkata with all the Restaurants superimposed on top as displayed below:




Now we are performing some basic explanatory data analysis and derive some additional info from our data. Let's count the number of restaurants in every location candidate and create a new data frame with the location and restaurant's count. After that we will be merging the restaurant count per location with our earlier data frame. So finally, we will have a data frame with all the locations name, latitude, longitude, distance from the city centre and total restaurants count. A snippet of the final data frame will be look like this:

| | Location | Latitude | Longitude | Distance from Centre | Restaurant Count |
|---|---|---|---|---|---|
| 0 | Baghajatin | 22.483994 | 88.375583 | 7.097440 | 0.0 |
| 1 | Bakery Road | 22.546598 | 88.327455 | 3.013952 | 0.0 |
| 2 | Ballygunge | 22.525881 | 88.366047 | 2.371353 | 12.0 |
| 3 | Ballygunge Sc College | 22.525881 | 88.366047 | 2.371353 | 0.0 |
| 4 | Barabazar | 22.590188 | 88.351242 | 5.011128 | 1.0 |

From the final data frame, we will be taking only those locations for modelling for which there are not more than 2 restaurants nearby and the location is not more than 5km away from the city centre.

## 3.2. Algorithm

In the final step we will focus on most promising locations and within those create clusters of locations that meet some basic requirements established in discussion with stakeholders: we will take into consideration locations with no more than two restaurants nearby and it should be within 5km of the centre of the city.

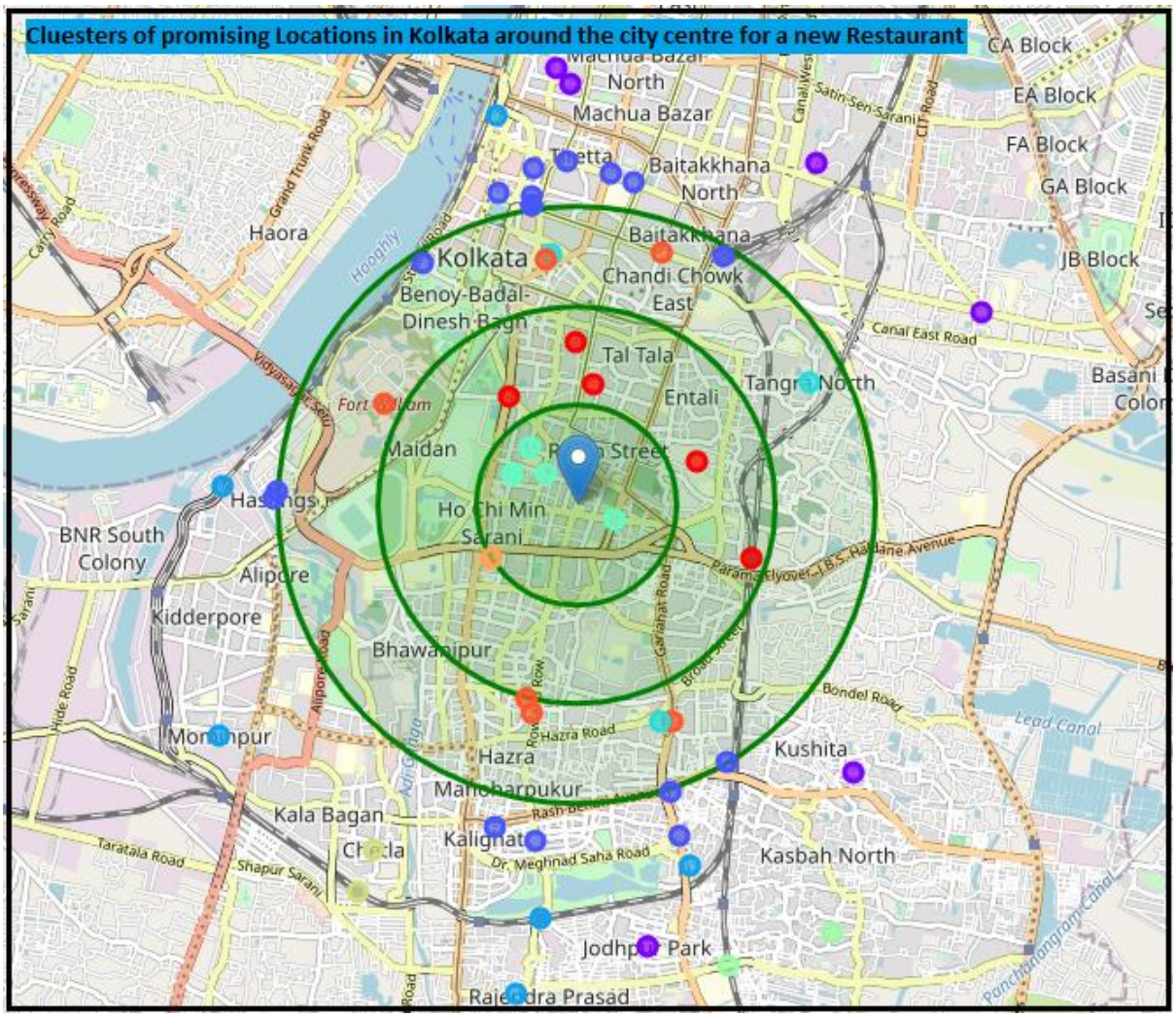
For clustering the considered locations here, we will be using **k-means clustering** and will create a total of 10 clusters.

We will be using k-means clustering here to segment the locations because here we need to group the locations based on some similar characteristics: Their distance from the city centre and Number of restaurants nearby.

After clustering we visualized the resulting clusters which is our final result:

## 4. Results

After clustering we visualized the resulting clusters which is our final result (3 circles indicating distance of 1km, 2km and 3km from the centre of the City):




From the above visualized data, we can see that there are 4 locations within 1km radius of the city centre which belong to a cluster for which there must be less than 2 restaurants nearby.

From the cluster labels we could find that they belong to cluster 5 and the address are as follows:

Shakespeare Sarani, Kolkata - 700017
Mall Road, Kolkata – 700080
Little Russel Street, Kolkata – 700071
Middleton Row, Kolkata – 700071

From our analysis we can see that there are almost ~200 restaurants present in Kolkata around 120 prime locations. So, the density of the restaurants present throughout the

city is light and there are almost 56 locations where there are not more than 2 restaurants nearby (Considering the location is not more than 5km away from the centre of the city). From our visualized data we have got one cluster (cluster 5) nearest to the centre of the city and there are total 5 locations for which no restaurants are present nearby. As per our business problem these 5 locations can be the most promising locations to open a new restaurant. As we already have the postal code details for all the locations so we were able to derive the exact address of the locations which can be present to Stakeholders for their interest.

## 5. Discussion

After considering our conditions to find out the best location we have created 10 clusters here for all the suitable locations. If we compare our raw data with the visualized data, we can see that the locations were clustered depending upon the distance from the city centre beside their nearby restaurant count. As for example if we look into the locations for cluster 2, we can see that all the locations are almost more than 3km away from the centre and they are spread out through the different areas of the city. If the neighbourhoods for these locations are popular and have popular venues around them, these can also be considered as promising locations for our stakeholders.

## 6. Conclusion

The goal of this project was to identify the best location in city Kolkata nearest to city centre to open a new restaurant depending upon the restaurant density in that location in order to help stakeholders to narrow down their search for the best location. We have collected postal code data for Kolkata to identify all the locations of the city and used Foursquare data to get all the restaurants details for those locations. We have analysed the data and extracted the locations distance from the centre of the city and also calculated the density of restaurants for each and every location. Finally, we have used clustering technique to cluster the filtered out top locations to achieve our goal. While using Foursquare API to get the nearby venues for locations we could notice that it could not able to return all of the venues accurately for some locations. If this could be improved our analysis could return better results.