2020

Finding best place to open Mughlai Restaurant in Kolkata

USING K-MEANS CLUSTERING AND FOUR-SQUARE API

SUDIPTA PAL

*****This document has been created in order to fulfil the requirement of the Capstone Project of Coursera.

Contents

Introduction

Background

Kolkata, formerly known as Calcutta, is the educational, commercial and cultural centre of the Eastern part of India, and is the third most populous metropolitan city of India. Kolkata is a pioneer in the field of drama, arts, theatre and literature with several Nobel laureates contributing to the Kolkata culture. But one thing people seems to forget about is the food of Kolkata. Somewhere between Rossogollas and lip-smacking fish preparations, many rich Bengali delicacies go unnoticed. And one such delicacy is the Kolkata Biryani. One can also blame the greater popularity of its southern counterpart - Hyderabadi Biryani, which might have for long prevented the Kolkata Biryani to flourish in all its glory. But Kolkata Biryani is slowly working its way out of the canals of the City of Joy and gaining the due recognition it has long deserved.

Business Problem

The aim of this project is to find out the best place in Kolkata to open a Mughlai Restaurant. In this project, leveraging venue data from Foursquare's 'Places AP' and 'k-means clustering' unsupervised machine learning algorithm, we will try to answer the question if someone want to open a Mughlai Restaurant in Kolkata which is the best for it.

Data

For this project, we need to have the below data:

- The List of the Neighbourhoods in Kolkata, India. This will help us to narrow down the place to a specific location to open the new restaurant.

- The Longitude and Latitude coordinates of the Kolkata Neighbourhoods. This will help us to plot and visualize each location on the map of Kolkata.

The data about the venues in these neighbourhoods, precisely related Mughlai Restaurant which will help us in clustering of neighbourhoods.

Data Sources

We have collected the neighbourhood’s data of Kolkata from Wikipedia using web scrapping. Then we have collected the latitude and longitude coordinates using Python Geocoder Package.

For the venue data we have used Foursquare API to make RESTful API calls to retrieve data about venues in different neighbourhoods.

Methodology

In this section we will download the required data using web scrapping and Foursquare API. We will analyse the data and perform K-means clustering find the best place to open a Mughlai Restaurant in Kolkata.

To kick off this project, we require the actual geographical co-ordinates of the city, Kolkata. We can look this up in the web but I have used geocoder to extract the actual geographical co-ordinates of Kolkata.

```
The polar coordinates of Kolkata are 22.5414185, 88.35769124388872.
```

Fetching neighbourhood's details of Kolkata from Wikipedia using Web Scrapping:

Using BeautifulSoup and request we have extracted the neighbourhood's details of Kolkata city.

	Neighborhood
1	Abhirampur
2	Agarpara
3	Ajoy Nagar
4	Alipore
5	Amodghata
6	Amtala
7	Anandapur, Kolkata
8	Andul
9	Ankurhati
10	Argari

Fig 1.1: List of Kolkata Neighbourhoods

This is the extracted data from Wikipedia about neighbourhoods of Kolkata. From Wikipedia we only able to get the neighbourhoods names but that's not sufficient. We need the geographical co-ordinates of each places in order to use this data with Foursquare API.

	Neighborhood	Latitude	Longitude
1	Abhirampur	22.684050	88.391650
2	Agarpara	22.489660	88.396400
3	Ajoy Nagar	22.526600	88.335100
4	Alipore	22.988010	88.388380
5	Amodghata	22.505220	88.399030
6	Amtala	22.514410	88.410320
7	Anandapur, Kolkata	22.570530	88.371240
8	Andul	22.610380	88.240010
9	Ankurhati	22.570530	88.371240
10	Argari	22.666470	88.366150
11	Ariadaha	22.472170	88.255460
12	Asuti	22.567630	88.344530
13	B. B. D. Bagh	22.567290	88.341060
14	Babughat	22.555080	88.246843
15	Badartala	22.604020	88.366370

Fig 1.2: List of Kolkata Neighbourhoods with geographical co-ordinates

Using geocoder package, we are able to extract the details of geographical co-ordinates of each location and created a dataframe.

Visualizing Neighbourhoods of Kolkata:

Using folium, we have visualized the map of Kolkata city and its neighbourhoods.

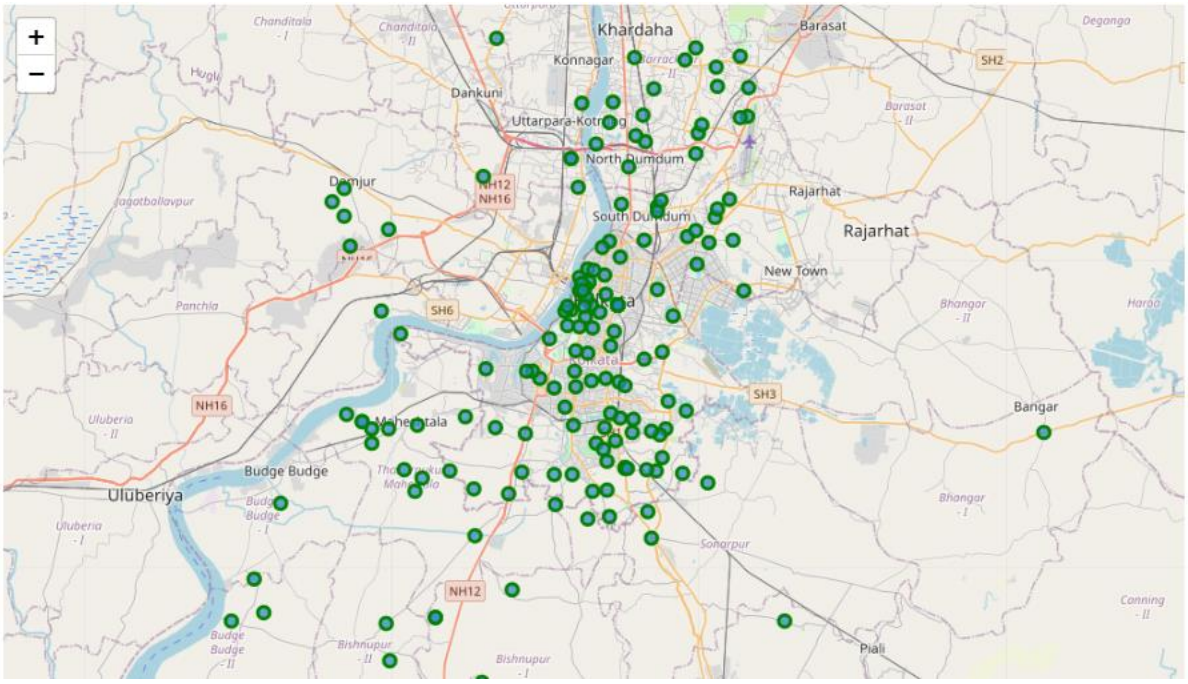

Fig 1.3 Map of Kolkata showing the neighbourhoods in green

Exploring the Neighbourhoods of Kolkata using Foursquare API

We have taken the radius as 2 KM and limit as 100 to extract the venues and their details in each neighbourhoods of Kolkata using Foursquare API. After extracting the venue details, we have created the venues dataframe which has 4285 observations and 7 features. We are able to extract 147 unique venue categories like Bakery, Pharmacy and as well as Mughlai Restaurant.

	Neighborhood	Latitude	Longitude	VenueName	VenueLatitude	VenueLongitude	VenueCategory
0	Abhirampur	22.68405	88.39165	Agarpara Railway Station	22.682886	88.385364	Train Station
1	Abhirampur	22.68405	88.39165	MedPlus	22.694171	88.402894	Pharmacy
2	Abhirampur	22.68405	88.39165	Shamvu da's tea shop	22.694959	88.379661	Bakery
3	Abhirampur	22.68405	88.39165	Events Bengal	22.694172	88.404686	Event Service
4	Abhirampur	22.68405	88.39165	MedPlus	22.700223	88.385307	Pharmacy

Analysing Neighbourhoods

After applying one-hot encoding we have managed to get 4285 observation and 148 features. By taking the sum of occurrence of each category we have grouped the rows of the neighbourhoods and We have found there are 67 Mughlai restaurants in Kolkata which quite high for a Mughlai restaurant number in a city. Now we have to find a place where the no of Mughlai restaurant low or moderate so that we can set up the new restaurant there and can operate without much competition with other Mughlai Restaurants.

	Neighborhoods	ATM	Airport	Airport Lounge	Airport Service	Airport Terminal	American Restaurant	Art Gallery	Art Museum	Arts & Crafts Store	Asian Restaurant	Athletics & Sports	Awadhi Restaurant	BBQ Joint	Ba
0	Abhirampur	0	0	0	0	0	0	0	0	0	0	0	0	0	1
1	Agarpara	0	0	0	0	0	0	0	0	0	0	0	0	0	0
2	Ajoy Nagar	0	0	0	0	0	1	0	0	0	1	0	2	0	0
3	Alipore	2	0	0	0	0	0	0	0	0	0	0	0	0	0
4	Amodghata	0	0	0	0	0	0	0	0	0	0	0	1	0	1
5	Amtala	0	0	0	0	0	0	0	0	0	0	0	0	0	0
6	Anandapur, Kolkata	0	0	0	0	0	0	0	1	0	0	0	0	0	1
7	Andul	0	0	0	0	0	0	0	0	0	0	0	0	0	0
8	Ankurhati	0	0	0	0	0	0	0	1	0	0	0	0	0	1
9	Argari	0	0	0	0	0	0	0	0	0	0	0	0	0	1
10	Asuti	0	0	0	0	0	0	0	0	0	1	0	0	1	2

Fig 1.4: Grouped by Category Dataset

Clustering Neighbourhoods

Firstly, we have created a new dataframe just only selecting neighbourhoods and Mughlai Restaurant Feature as these two only needed for further analysis.

Now we have to cluster these neighbourhoods to find the number of Mughlai restaurant in each cluster. From there we can identify the clusters having lower number of Mughlai Restaurant, so that we can set up a restaurant there. We have taken the number of clusters as 5 and applied clustering to segregate the neighbourhoods and form different clusters.

	Neighborhood	Mughlai Restaurant	Cluster Labels
0	Abhirampur	0	0
1	Agarpara	0	0
2	Ajoy Nagar	1	2
3	Alipore	0	0
4	Amodghata	1	2

Fig 1.5: The number of Mughlai Restaurant in a particular neighbourhoods and cluster labels.

We added the geographical data of each neighbourhoods and sorted the dataframe using the cluster levels. Using this we can visualize the cluster in the map of Kolkata.

	Neighborhood	Mughlai Restaurant	Cluster Labels	Latitude	Longitude
105	Durganagar, Kolkata	0	0	22.510840	88.372580
106	Duttapukur	0	0	22.569580	88.342570
107	East Kolkata	0	0	22.531860	88.326760
112	Fort William, India	0	0	22.619930	88.394180
113	Ganye Gangadharpur	0	0	22.609470	88.416060
114	Garden Reach	0	0	22.569870	88.351710
104	Dunlop, Kolkata	0	0	22.934720	88.371430
115	Garfa	0	0	22.650410	88.415660

Fig 1.6: Neighbourhood-Cluster data with geographical co-ordinates

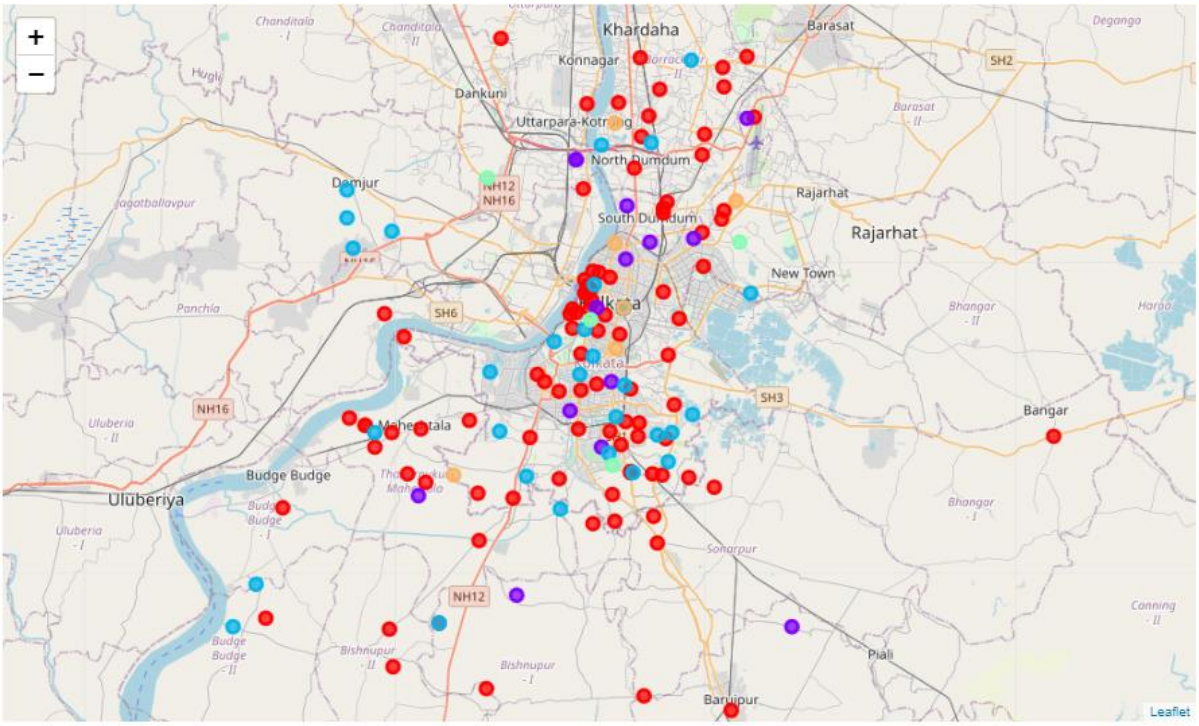

Fig 1.7: Showing different clusters

We have visualized the clusters on the map of Kolkata and we can say that cluster zero has more occurrence than any other clusters.

Ultimately, we have collected the results of our clustering.

- The neighbourhoods in cluster 0: 115
- The neighbourhoods in cluster 1: 15
- The neighbourhoods in cluster 2: 40
- The neighbourhoods in cluster 3: 4
- The neighbourhoods in cluster 4: 8

Results

From the result of the K-Means clustering we can cluster the neighbourhoods into 5 clusters:

- Cluster 0: It categorizes the neighbourhoods which are having zero/ very low number of Mughlai Restaurant.
- Cluster 1: It categorizes the neighbourhoods which are having low number of Mughlai Restaurant.
- Cluster 2: It categorizes the neighbourhoods which are having moderate number of Mughlai Restaurant.
- Cluster 3: It categorizes the neighbourhoods which are having high number of Mughlai Restaurant.
- Cluster 4: It categorizes the neighbourhoods which are having very high number of Mughlai Restaurant.

We have extracted the length of each cluster and cluster 0 has the largest number of neighbourhoods where the number of Mughlai Restaurant is very low to none.

Discussion

Cluster 1 and cluster 2 are also having neighbourhoods with low to moderate number of Mughlai Restaurant but compare to Cluster 0 the number of neighbourhoods are low. But we cannot discard those clusters as having lower number of Mughlai Restaurant the people of those neighbourhoods have a taste of Mughlai. If we want, we can turn the favour towards us and make a profit by opening a Mughlai Restaurant there also.

Conclusion

Even though overall Kolkata has large number of Mughlai Restaurant, cluster 0 having lower number of Mughlai Restaurant shows a great opportunity to build a successful business and make a profit out of it. As the trend is going, the hunger for Mughlai cuisine is keep increasing among the people of the city Kolkata. So, if anyone wants taking the advantage of lower number of Mughlai Restaurant in cluster 0 can set up a Mughlai Restaurant. Just keep in mind that we have used the number or the occurrence of Mughlai restaurant and analysed and categorized the data to come up with the results. There are other factors like cost of the project, population of the area, the average income of the population, type of shop like take away or in house sitting, can be added to more precise prediction.

_____