Capstone Project - Battle of Neighbourhoods

Applied Data Science Capstone

Table of contents

Introduction: Business Problem

In this project we will try to find an optimal location for a restaurant. Specifically, this report will be targeted to stakeholders interested in opening an **restaurant** in **Kolkata, West Bengal, India**.

Since there are lots of restaurants in Kolkata we will try to detect **locations that are not already crowded with restaurants**. We are also particularly interested in **areas with no restaurants in vicinity**. We would also prefer locations **as close to city center as possible**, assuming that first two conditions are met.

We will use our data science powers to generate a few most promising neighborhoods based on this criteria. Advantages of each area will then be clearly expressed so that best possible final location can be chosen by stakeholders.

Data

Based on definition of our problem, factors that will influence our decission are:

- number of existing restaurants in the neighborhood (any type of restaurant)
- distance to restaurants in the neighborhood, if any
- distance of neighborhood from city center

We decided to use regularly spaced grid of locations, centered around city center, to define our neighborhoods.

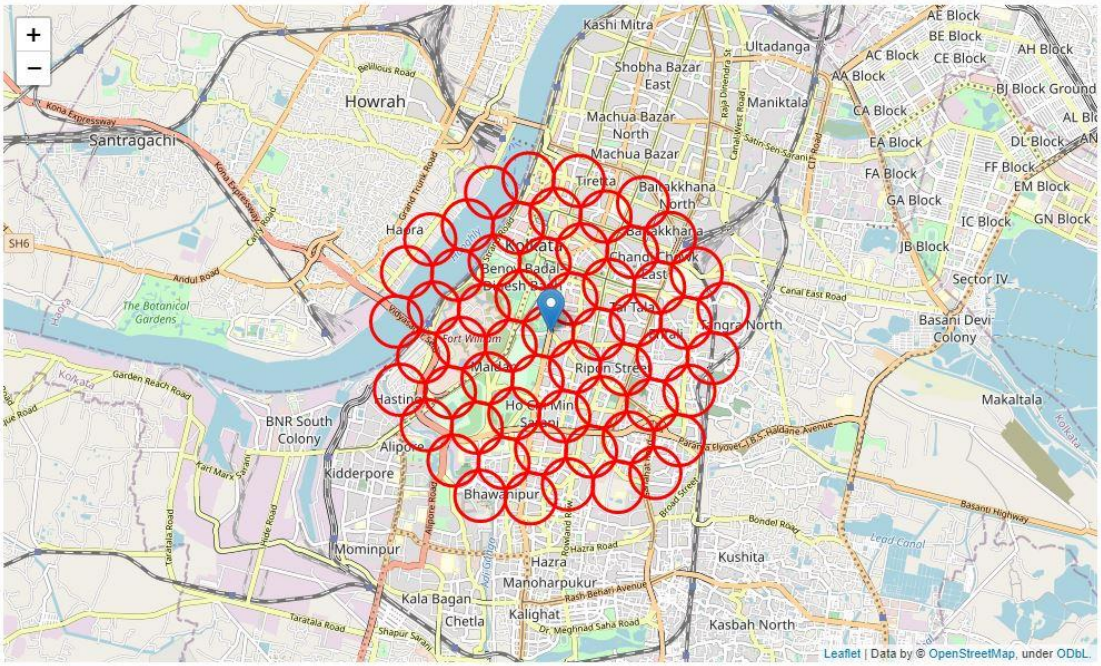
Following data sources will be needed to extract/generate the required information:

- centers of candidate areas will be generated algorithmically and approximate addresses of centers of those areas will be obtained using **geopy reverse geocoding**
- number of restaurants and their type and location in every neighborhood will be obtained using **Foursquare API**
- coordinate of Kolkata center will be obtained using **geopy geocoding** of well known Kolkata location (Rabindra Sadan/ Park Street/ Esplanade)

Methodology

1. 58 neighbourhood areas are randomly generated.

Out[270]:

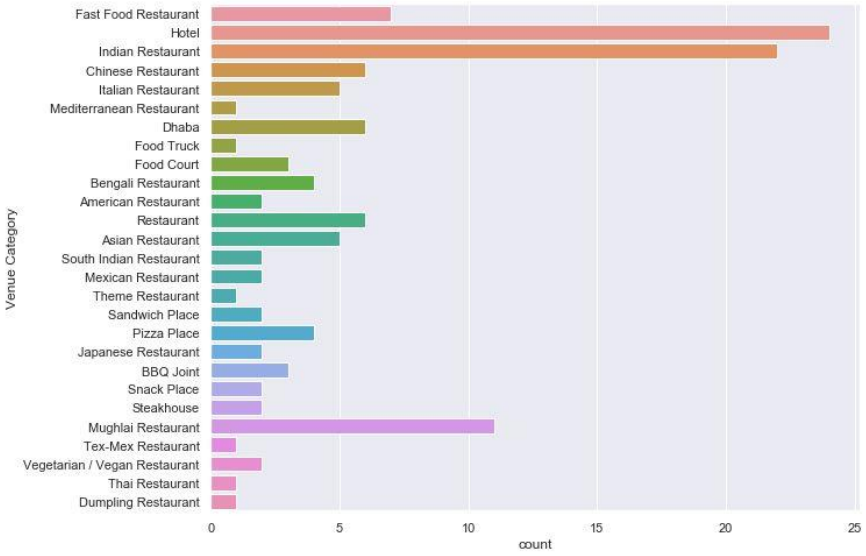


2. Address for each randomly generated neighbourhood centers are fetched using reverse geocoding.

Out[275]:

	Address	Latitude	Longitude
0	Hastings Post Office (Secondary), Bakery Road,...	22.547202	88.326939
1	Strand Road, Hastings, Kolkata, West Bengal, 7...	22.542173	88.331082
2	Taj Bengal, Belvedere Road, Alipore, Kolkata, ...	22.537145	88.335224
3	Sashi Sehkar Bose Row, Bhawanipur, Kolkata, We...	22.532118	88.339364
4	Vidyasagar Setu, Hastings, Kolkata, West Benga...	22.556906	88.326381

3. Total number of restaurants within 8km radius of Park Street, Kolkata



In this project we will direct our efforts on detecting areas of Berlin that have low restaurant density, particularly those with low number of restaurants. We will limit our analysis to area ~8 km around city center.

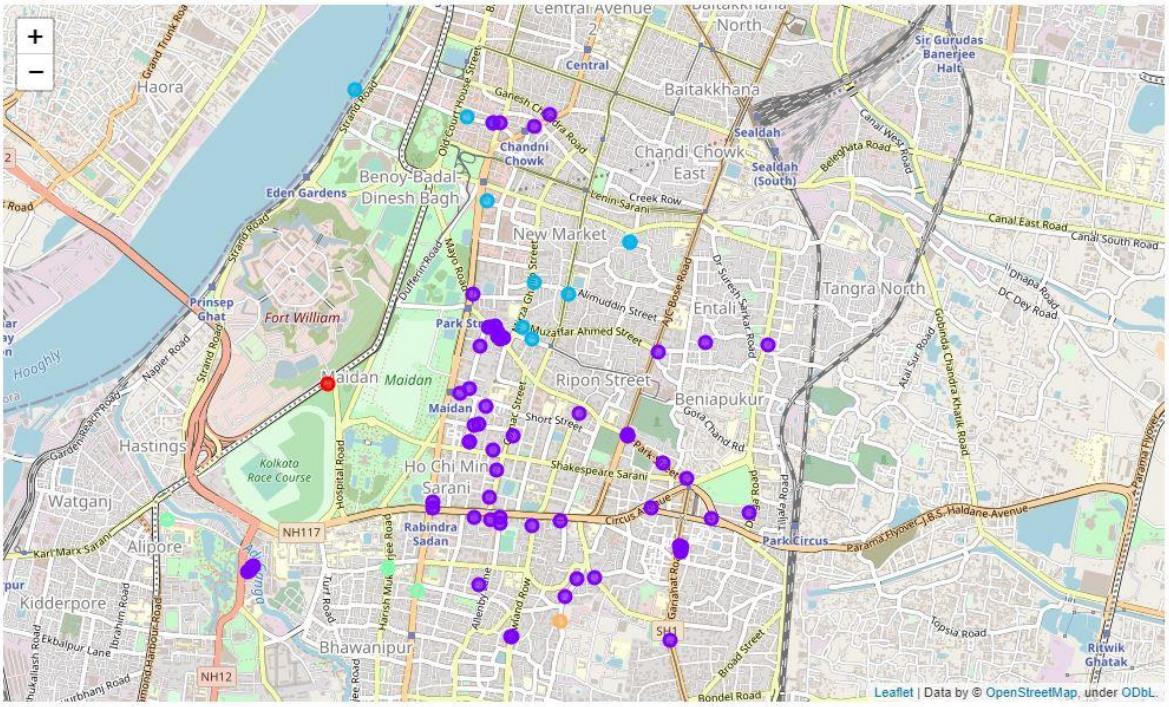
In first step we have collected the required **data: location and type (category) of every restaurant within 8km from Kolkata city center** (Park Street). We have also **identified restaurants** around each neighborhood.

In second and final step we will focus on most promising areas 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 in radius of 250 meters**.

We will present map of all such locations but also create clusters (using **k-means clustering**) of those locations to identify general zones / neighborhoods / addresses which would help the stakeholders in further decision making.

4. Visualizing the clusters



Results and Discussion

Therefore, from whatever limited data available, we could successfully identify 26 zones where the number of restaurants were less than 10 in their vicinity. Most of them had restaurant count even less than 5 in 400 m surroundings. What we observe is that Kolkata is comparatively less restaurant density. More data could have helped us to understand our regions in a much better manner.

Those location candidates were then clustered to create zones of interest which contain greatest number of location candidates. It was found that cluster 2 has 80% of the venues grouped together. Further observations reveal that 'Italian Restaurant', 'Bengali Restaurant', 'Asian Restaurant', 'Indian Restaurant', 'Chinese Restaurant', 'Mughlai Restaurant', 'South Indian Restaurant' are the most common types of restaurants. Hence the stakeholders can try setting up restaurants of different cultures other than the already present common ones.

Cluster 2

```
In [367]: kolkata_merged.loc[kolkata_merged['Cluster Labels'] == 1, kolkata_merged.columns[[1] + list(range(5, kolkata_merged.shape[1]))]]
```

Out[367]:

	Neighborhood Latitude	Venue Longitude	Venue Category	Distance Neighborhood	Cluster Labels	1st Most Common Venue	2nd Most Common Venue	3rd Most Common Venue
1	22.537145	88.334066	Hotel	129.320472	1	Italian Restaurant	Chinese Restaurant	Hotel
2	22.537145	88.334210	Indian Restaurant	113.193833	1	Italian Restaurant	Chinese Restaurant	Hotel
3	22.537145	88.334327	Chinese Restaurant	100.175214	1	Italian Restaurant	Chinese Restaurant	Hotel
4	22.537145	88.334498	Italian Restaurant	81.104265	1	Italian Restaurant	Chinese Restaurant	Hotel
5	22.537145	88.334241	Mediterranean Restaurant	109.758777	1	Italian Restaurant	Chinese Restaurant	Hotel
...	...	...	...	...	...	...	...	...
76	22.552085	88.363881	Indian Restaurant	403.597962	1	Mughlai Restaurant	Indian Restaurant	Vegetarian / Vegan Restaurant
77	22.576889	88.355538	Vegetarian / Vegan Restaurant	110.992094	1	Indian Restaurant	Vegetarian / Vegan Restaurant	Theme Restaurant
78	22.576889	88.356653	Indian Restaurant	235.668748	1	Indian Restaurant	Vegetarian / Vegan Restaurant	Theme Restaurant
79	22.576889	88.357136	Indian Restaurant	289.504131	1	Indian Restaurant	Vegetarian / Vegan Restaurant	Theme Restaurant
80	22.551720	88.371856	Chinese Restaurant	375.417768	1	Chinese Restaurant	Vegetarian / Vegan Restaurant	Theme Restaurant

65 rows × 8 columns

Result of all this is 26 zones containing largest number of potential new restaurant locations based on number of and distance to existing venues - both restaurants in general. This, of course, does not imply that those zones are actually optimal locations for a new restaurant! Purpose of this analysis was to only provide info on areas close to Kolkata center but not crowded with existing restaurants (particularly Italian) - it is entirely possible that there is a very good reason for 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 location which has not only no nearby competition but also other factors taken into account and all other relevant conditions met.

Conclusion

Purpose of this project was to identify Kolkata areas close to center with low number of restaurants in order to aid stakeholders in narrowing down the search for optimal location for a new restaurant. By calculating restaurant density distribution from Foursquare data, we have first identified general addresses that justify further analysis and then generated extensive collection of locations which satisfy some basic requirements regarding existing nearby restaurants. Clustering of those locations was then performed in order to create major zones of interest (containing greatest number of potential locations) and addresses of those zone centers were created to be used as starting points for final exploration by stakeholders.

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 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.