Where should Customer open a Restaurant?

Business requirement

A client would like to open a restaurant in a metropolitan city Kolkata in India and has asked where they should locate that store?

To do this analysis, we will use the following information:

- The foursquare API to retrieve the current venue in Kolkata.
- Zomato API from where we extract Restaurant name, rating, price etc.

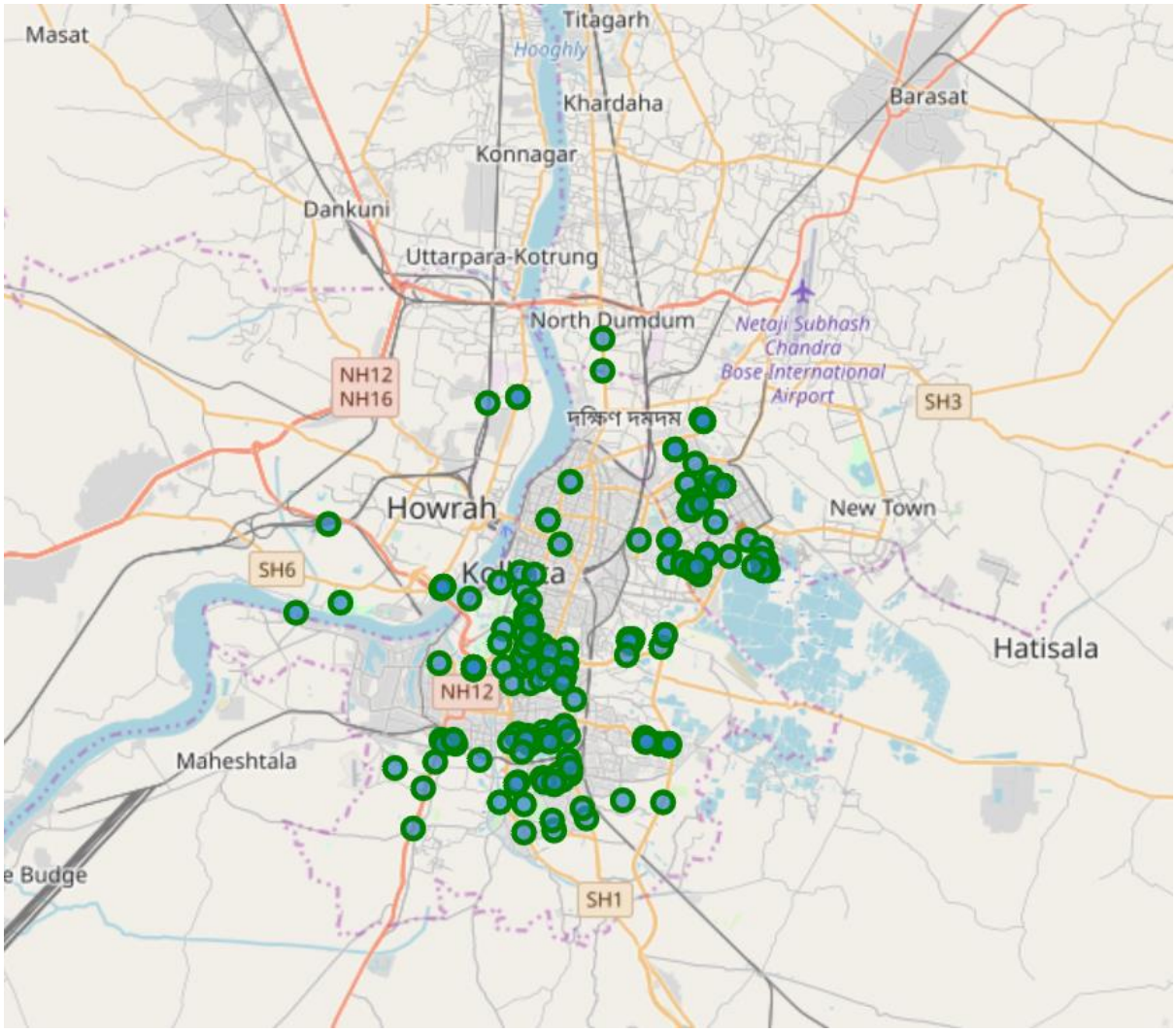
Using the above information, we will

- Visualize restaurants information with different dimention.
- Clustering based on user rating and average price.
- Cluster with low rating and high price will be the best area for new opening.

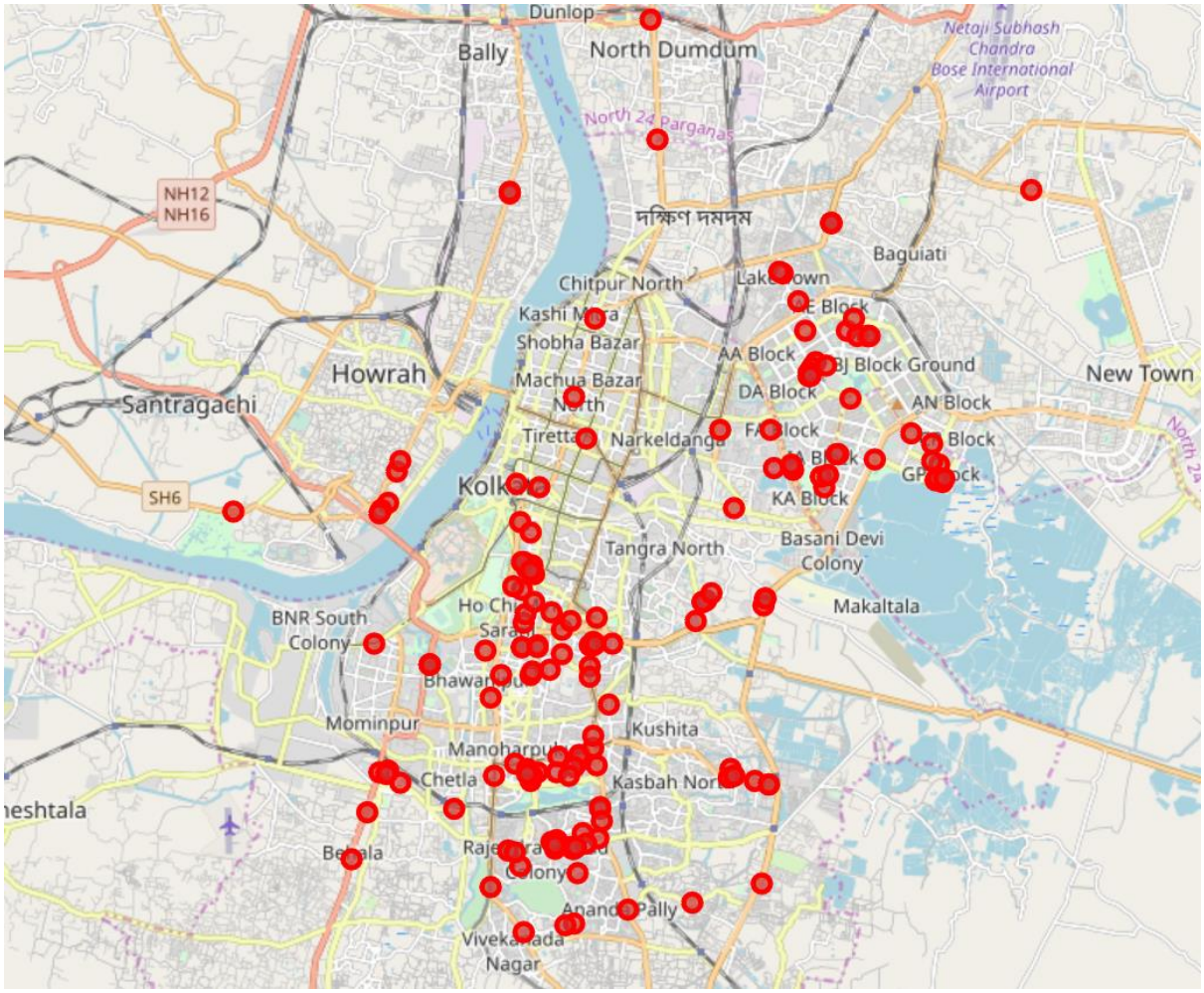
Data used in this project

To begin with, we will take a look at **Kolkata on the Map** using the `folium` library.
We will also fetch the data from **two different APIs**.

- **Foursquare API:** We will use the Foursquare API to fetch venues in Kolkata starting from the middle upto 10 KM in each direction.



- **Zomato API:** The Zomato API provides information about various venues including the complete address, user ratings, price for two people, price range and a lot more.



1. Data cleaning & Preparation

As a first step, we retrieved the data from two APIs (Foursquare and Zomato). We extract venue information from the centre of Kolkata, upto a distance of 4 Km. The latitude and longitude values are then used to fetch venue rating and price from Zomato.

Foursquare API requires client_id, and client_secret to function which can be accessed after creating a developer account. We will set the radius as 4 Kilometers. The version is a required parameter which defines the date on which we are browsing so that it retrieves the latest data.

The Zomato API allows using its search API to search for any given venue based on certain search filters such as query, latitude, longitude and more. Zomato also requires a Zomato user key which can be accessed with a developer account.

We'll use the name, lat, and lng values of various venues fetched from Foursquare API to use the search API and get more information regarding each venue.

The query will be the name of the venue. The start defines from what offset we want to start, so we'll keep it at 0. The count defines the number of restaurants we want to fetch. As we have the exact location coordinates, we'll fetch only one. We will supply the latitude and

longitude values. We will set the sorting criteria as real_distance so each time we get the venue we're searching based on location coordinates.

The below steps I follow for data cleaning and processing.

1. Checking latitude and longitude data types.
2. Convert its data type to numeric for graph.
3. Next combine Foursqaur and Zomato dataframe.To combine the two datasets, I'll have to check that the latitude and longitude values of each corresponding venue match. Thus, I'll round both the latitude and longitude values upto 4 decimal places. Then, I'll calculate the difference between the corresponding latitude and longitude values and see if the difference is less than 0.0004 which should ideally mean that the two locations are same.

	name	categories	lat	lng	venue	latitude	longitude	price_for_two	price_range	rating	address	lat_diff	lng_diff
0	Lalit Great Eastern Hotel	Hotel	22.5680	88.3500	The Tea Lounge - The Lalit Great Eastern	22.5679	88.3502	1500.0	3.0	3.7	The Lalit Great Eastern, 1 - 3, Old Court Hous...	-0.0001	0.0002
1	Blue & Beyond	Pub	22.5591	88.3533	Blue And Beyond	22.5591	88.3532	1600.0	3.0	3.9	The Lindsay, 8A & 8B, Lindsay Street, New Mark...	0.0000	-0.0001
2	Peter Cat	Indian Restaurant	22.5524	88.3525	Peter Cat	22.5524	88.3525	1000.0	3.0	4.3	18A, Park Street, Park Street Area, Kolkata	0.0000	0.0000
3	The Blue Poppy	Asian Restaurant	22.5485	88.3514	Blue Poppy-Thakali	22.5486	88.3512	500.0	2.0	4.3	Sikkim Commercial House, 1st Floor, 4/1 Middle...	0.0001	-0.0002

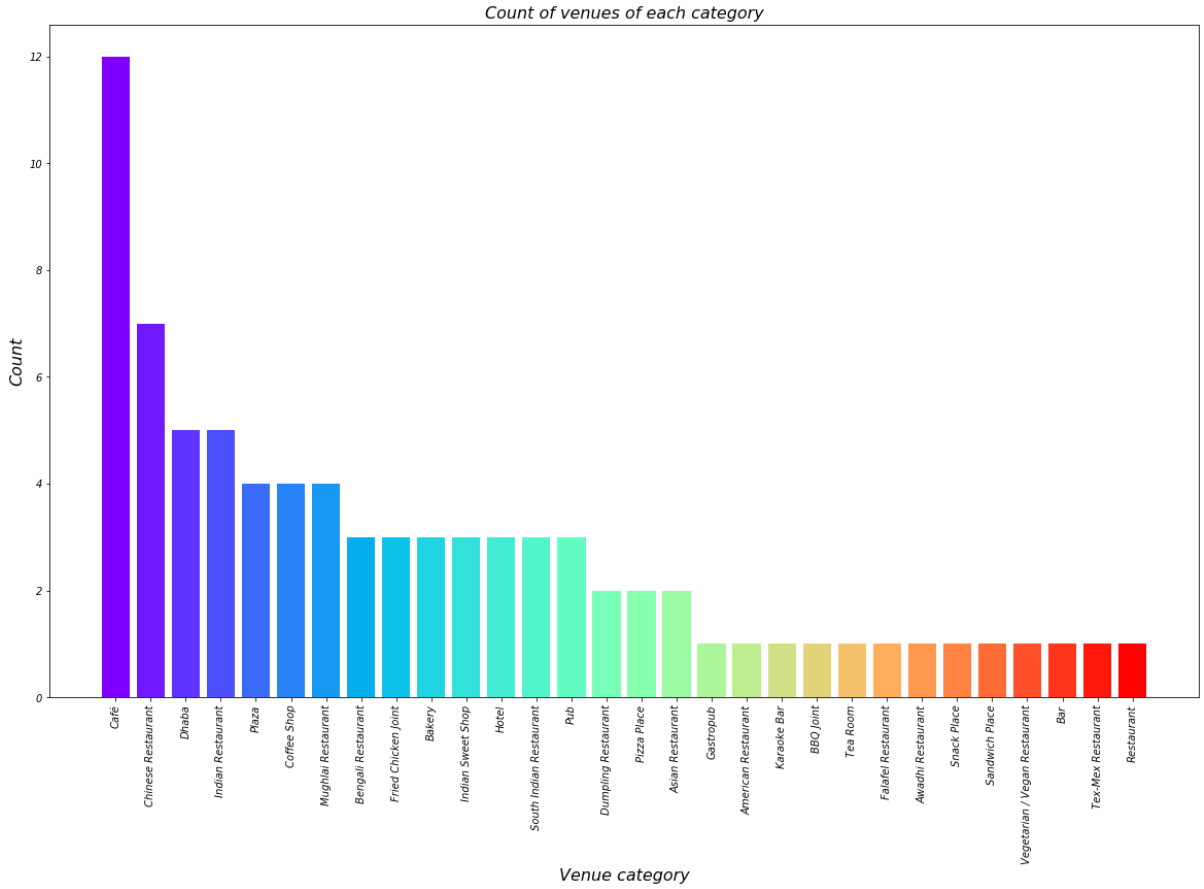
4. Checking Hotel name for duplicate. If found remove those.
5. Check for null value of the data. If found remove those.
6. We have price_for_two column in our data. I convert this column to avg_price column and added to data.

	name	categories	lat	lng	venue	latitude	longitude	price_for_two	price_range	rating	address	lat_diff	lng_diff	average_price
0	Lalit Great Eastern Hotel	Hotel	22.5680	88.3500	The Tea Lounge - The Lalit Great Eastern	22.5679	88.3502	1500.0	3.0	3.7	The Lalit Great Eastern, 1 - 3, Old Court Hous...	-0.0001	0.0002	750.0
1	Blue & Beyond	Pub	22.5591	88.3533	Blue And Beyond	22.5591	88.3532	1600.0	3.0	3.9	The Lindsay, 8A & 8B, Lindsay Street, New Mark...	0.0000	-0.0001	800.0
2	Peter Cat	Indian Restaurant	22.5524	88.3525	Peter Cat	22.5524	88.3525	1000.0	3.0	4.3	18A, Park Street, Park Street Area, Kolkata	0.0000	0.0000	500.0
3	The Blue Poppy	Asian Restaurant	22.5485	88.3514	Blue Poppy-Thakali	22.5486	88.3512	500.0	2.0	4.3	Sikkim Commercial House, 1st Floor, 4/1 Middle...	0.0001	-0.0002	250.0
4	Flurys	Bakery	22.5528	88.3526	Flurys	22.5528	88.3525	400.0	1.0	4.1	18, Park Street Area, Kolkata	0.0000	-0.0001	200.0

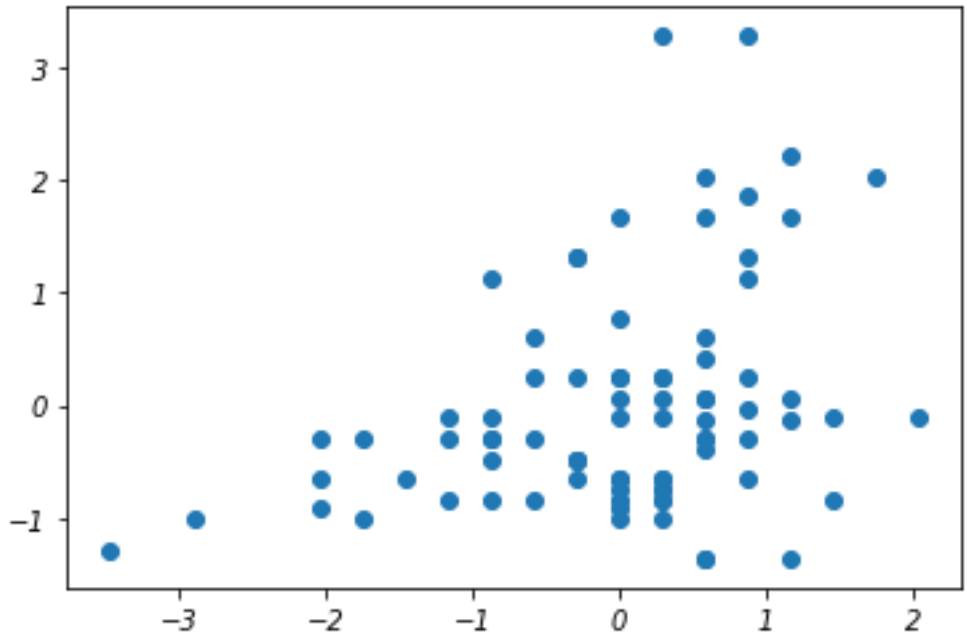
7. Next I remove all the columns which is not related to food as requirement is limited to resturannt.
8. As per preparation for clustering drop cloumns which is not required.
9. Drop the venues which have 0.0 rating as it means it's not been rated yet.

Analysis

We have various types of restaurants in the final dataset. Will look at the venues and check which are the majority venue categories in the list.

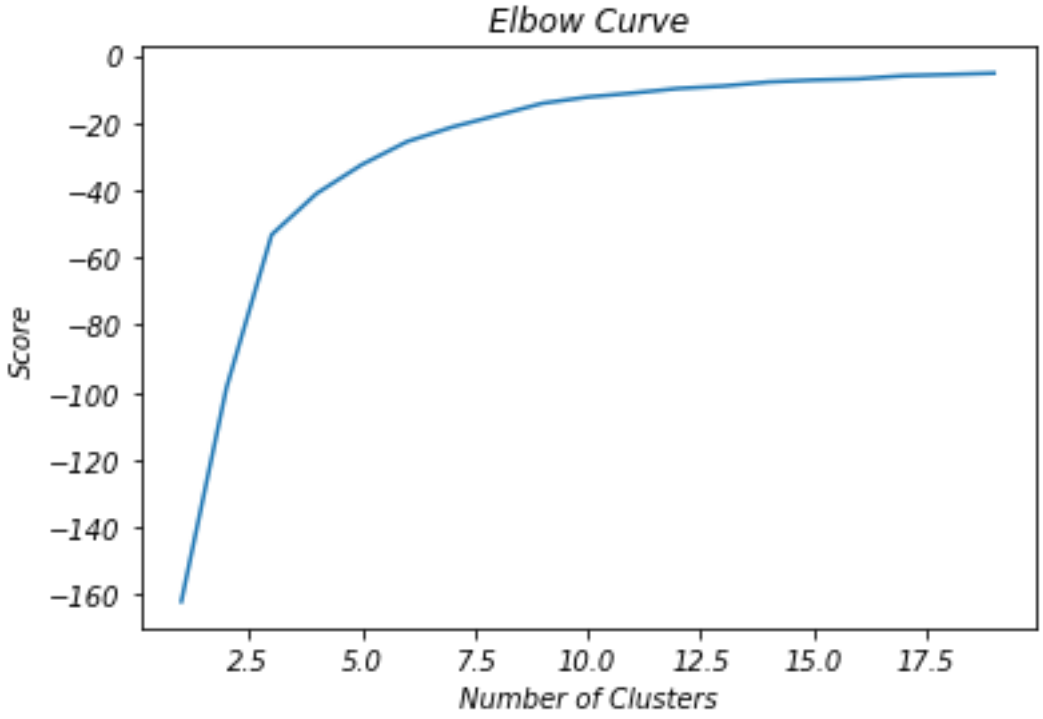


1. Since the rating and price are on different scales, they should be scaled before doing any analysis.will use StandardScaler from SciKitlearn Get a subset of our features and store in X_train. Let's try to cluster based on: rating and price.
2. VIsualize these relationships before plotting.



Clustering

First step is to determine how many clusters we should divide the neighbourhoods into. Let's use the visual Elbow method.



1. Looks like the elbow is at 6 clusters, so let's use that and re-run KMeans to get our Clusters.

	categories	venue	latitude	longitude	price_range	rating	address	average_price	Cluster Labels
0	Hotel	The Tea Lounge - The Lalit Great Eastern	22.5679	88.3502	3.0	3.7	The Lalit Great Eastern, 1 - 3, Old Court Hous...	750.0	4
1	Pub	Blue And Beyond	22.5591	88.3532	3.0	3.9	The Lindsay, 8A & 8B, Lindsay Street, New Mark...	800.0	4
2	Indian Restaurant	Peter Cat	22.5524	88.3525	3.0	4.3	18A, Park Street, Park Street Area, Kolkata	500.0	1
3	Asian Restaurant	Blue Poppy-Thakali	22.5486	88.3512	2.0	4.3	Sikkim Commercial House, 1st Floor, 4/1 Middle...	250.0	1
4	Bakery	Flurys	22.5528	88.3525	1.0	4.1	18, Park Street Area, Kolkata	200.0	3

2. Plot data with cluster number.
3. Received result with below format:

These venues for cluster 0 have mean price range of 214.29 and rating spread around 3.21

Result and Conclusion

A company who wants to open a new store can consider various cluster where average price is high and user rating is low.

Those area are potentially good for new opportunities.