Selection of best location to open a new restaurant in India

Minhaj Ahmed Ansari

Business Problem

Assuming that a Multi-national fast food chain wants to open its new chain of restaurant in India,

- Which location should they choose?
- Which type of cuisine should they serve?
- And most of all how to ensure people are easily able to reach the location.

Business Problem

I restricted my analysis to 5 major cities of India, which in a way are Economic, Technological and demographic centres of India. They are,

- Delhi
- Mumbai
- Chennai
- Kolkata
- Bengaluru

Key Criteria for selecting final location

- Optimal distance between city centre and the proposed location.
- Selection of cuisine.
- Density of local businesses at the proposed location.
- Traffic Conditions (also instrumental in case of online deliveries)

This analysis would be really helpful for the small businesses and start-ups who are looking to locate their stores/ restaurants at optimal location to get a jump start.

Data Sources

- An easy way to get location data is through Foursquare Places API's **explore** endpoint. The documentation on how to create foursquare developer account, rate limits and calling methods is available on the link here.

- Another data source that we will be using in this project will be **Traffic API** provided by Bing Maps. This API can be used to get traffic information on a particular set of location. A complete detailed documentation on this REST API is available here.

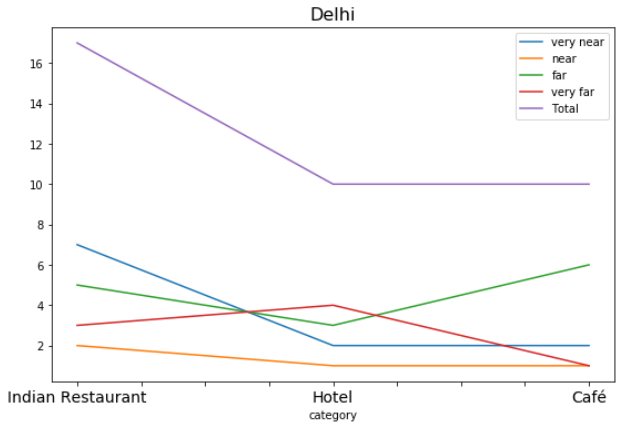
Cleaning Data

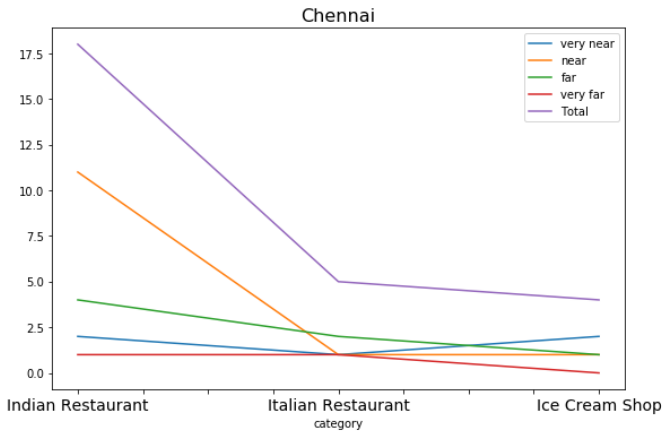
- Removing unnecessary columns
- Extracting Category of each venue
- Binning distance data
  - Very Near = 0 to 2 KM
  - Near = 2 to 4 KM
  - Far = 4 to 6 KM
  - Very Far = 6 to 8 KM
  - Outlier = 8 to 10 KM
- One hot encoding for
  - Categories
  - Binned distance data

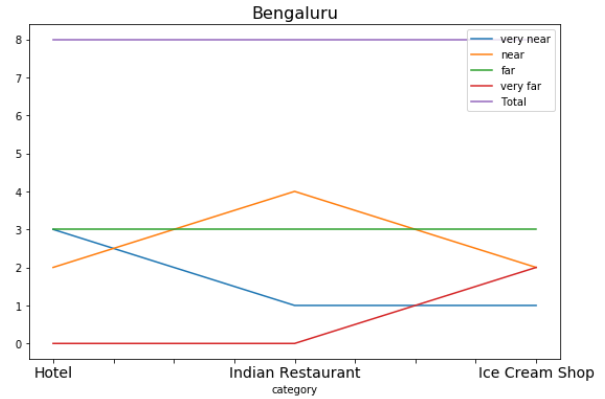
Finding Top 3 Categories and there spatial distribution

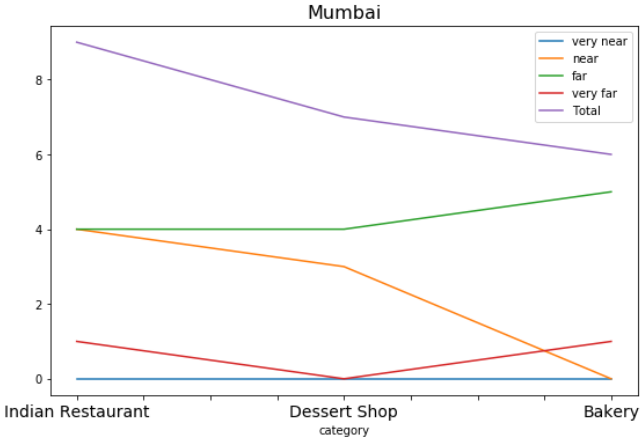
Exploratory Data Science

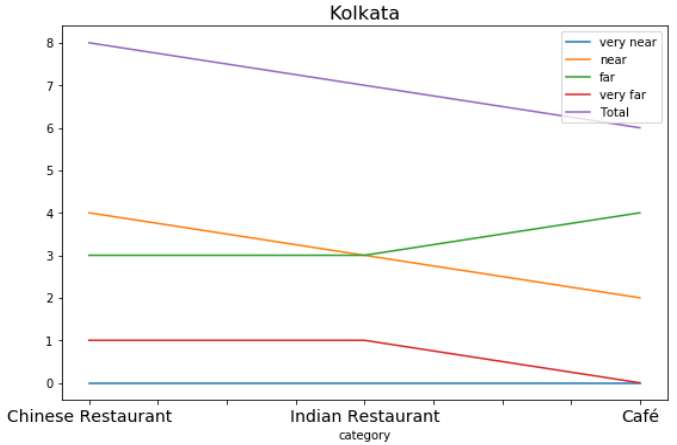
We can conclude from the graphs that, Indian Cuisine is the most recommended and favoured category among all at all the cities.











Exploratory Data Science

I identified that in almost all cities most of the venues were located either in 2 to 4 KM radius (near) or 4 to 6KM radius (far) region.

Distribution of venues

City	Very Near	Near	Far	Very Far	Outlier
Delhi	0.29	0.21	0.32	0.17	0.01
Mumbai	0.05	0.24	0.55	0.16	0.00
Chennai	0.12	0.56	0.21	0.08	0.03
Kolkata	0.05	0.31	0.39	0.22	0.03
Bengaluru	0.14	0.33	0.38	0.15	0.00

Note: The Foursquare API returns the great circle distance or the orthodromic distance. This distance is the shortest distance between two points on a sphere.

Having known about above detail we cannot use this distance as an indication of traffic density in any area or on between two locations. Here, the Traffic API which we discussed in Data section will come handy.

KMeans Clustering & Evaluation

I used elbow method to first determine the best k for each city and tried validating those with Silhouette Coefficient for each k.

CITY	ELBOW METHOD K	SILHOUETTE COEFFICIENT K
DELHI	4	14
MUMBAI	5	9
CHENNAI	4	8
KOLKATA	4	19
BENGALURU	5	8

The difference shows that; the global optimum widely varies from local optimum. Thus it would be prudent to proceed with global optimums.

Recommender System for Selecting best cluster

Criteria for Selection

1. **The cluster must not have any Indian Restaurant**
2. **The cluster should be 'nearest possible' to the city centre**
3. **The cluster should have 'highest possible density'**

Our recommender system should choose the optimal distanced cluster based on nearest possible distance, highest possible density and it must not have any Indian Restaurant.

Modelling and Final Recommendation

Recommender System for Selecting best cluster

The cluster 2 of Bengaluru city was suggested as best cluster to place our business based on our modelling and recommendation.

I inversed my direction data and multiplied it with cluster data. Providing me with a score which would be highest if the both distance and cluster density (count of venues) are optimised for their minimum and maximum value respectively. Mathematically we can represent it as,

$$\textit{Recommendation Score} = f(\textit{distance}, \textit{cluster density}) = \textit{distance}^{-1} * \textit{cluster venue counts}$$

City	Cluster label	Lat	Lng	Distance	Count	Score
Delhi	1	28.645444	77.186265	3008.75	8	0.002659
Mumbai	7	19.067727	72.86588	1639.6	5	0.00305
Chennai	5	13.043017	80.201891	4742.2	5	0.001054
Kolkata	3	22.537695	88.350328	4166.555556	9	0.00216
Bengaluru	2	12.972181	77.601005	2251.384615	26	0.011548

Predicting Traffic for our selected Cluster

Calculating Isochrone

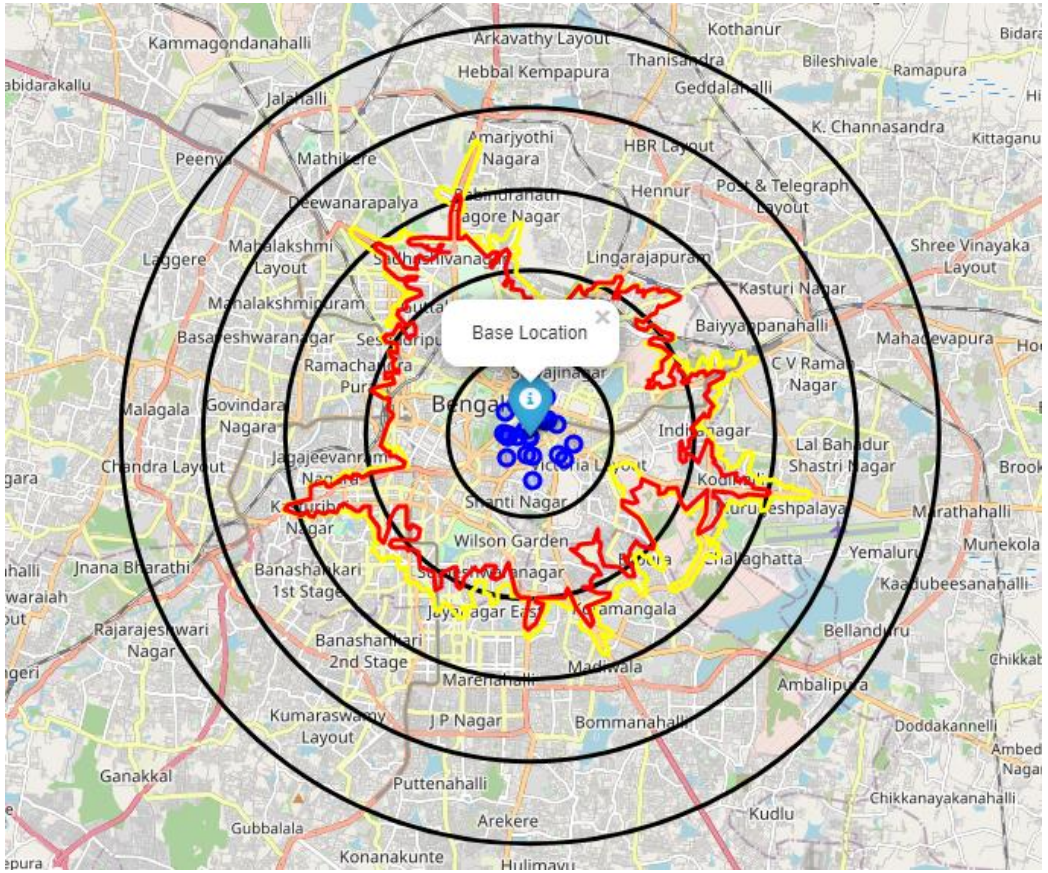
The Bing Maps Isochrone API provides time-specific, isoline polygons for the distance that is reachable from a given location and supports multiple modes of transportation (i.e., driving, walking, and public transit).

More details about this API is available here.

I chose the mean of latitude and longitude of best cluster as base location for calling this API. In real life scenario, businesses can choose the potential store locations to get results.

Predicting Traffic for our selected Cluster

Plotting Isochrone



- I have plotted **folium.Circle** for binned distance now choosing mean lat and long as base location.
- folim.Marker to denote mean lat and long location
- folium.CircleMarker to denote all cluster venues.
- And folium.PolyLine for plotting the isochrone diagram.

Conclusion & Future Directions

Some Key Findings

- Indian cuisine is still the most popular among people.
- Proximity to the city centre or square doesn’t really translates into the best location for business.
- The cluster 2 of Bengaluru city was suggested as best cluster to place our business.

Future scope

- The actual location of the restaurant may vary based on the availability of the space, rent rates, proximity to roads or highways etc.
- Another interesting study could be analysis of the crime data in the locality and use it as one of the modelling parameters.

THANKYOU

Minhaj Ahmed Ansari