Business venture to maximize revenue

Table of Contents

Abstract

In an over-crowded city like Bengaluru in India, coming up with a food venture is like a double-edged sword. On one hand we see a high number of restaurants and eateries which would mean a high competition and on the other hand an increasing population in the IT capital of India would mean an opportunity to venture into a profitable business.

Introduction

As one seeks to venture into a profitable food business, the success and failure of the venture would depend on the present competition in the area, type of the venue, footfall in the neighborhood. Data Science with ML can help uncover and segment a city (like Bengaluru) to answer and identify areas which are similar in nature and the opportunity cost to venture into a specific type of food business is low and the returns can be high. **This analysis is primarily targeted towards individuals / groups who would like to venture into a food business** and would like to know details which give them an upper hand before start of any business.

Data Acquisition

The data was obtained by scraping through a publicly available Wikipedia page which lists all the regions in a city of Bengaluru and neighborhoods within each region. Using geopy , co-ordinates of the neighborhoods were identified. Neighborhoods whose co-ordinates are unidentified will be cleaned or manually updated. Cases where the co-ordinates differ from the actual co-ordinates were updated manually. This clean data includes details like **Neighborhood, Region, Latitude, Longitudes**.

For each region in the city of Bengaluru, the identified number of neighborhood counts are as follows.

Region	Neighbourhood
Central	11
Eastern	8
North-Eastern	6
Northern	8
South-Eastern	7
Southern	8
Southern suburbs	6
Western	9

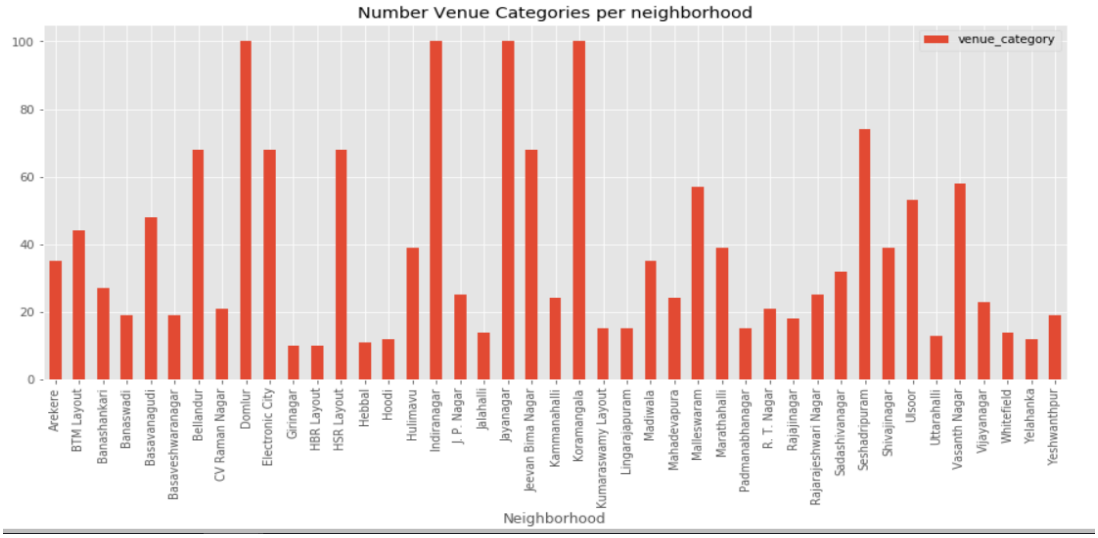
A brief view of a clean data is as follows

	Neighbourhood	Region	latitude	longitude
0	Cantonment area	Central	13.0196	77.5096
1	Domlur	Central	12.9625	77.6382
2	Indiranagar	Central	12.9733	77.6405
3	Malleswaram	Central	13.0163	77.5587
4	Rajajinagar	Central	12.9882	77.5549

The data was further processed with Fairsquare API to identify the nearby venues in each of the neighborhood. The dataframe now contains details like **Neighborhood, Region, Latitude, Longitudes, <Venue categories>.** Data was cleaned with neighborhood with lower than 10 venues to avoid outliers. This data was segmented / clustered, and the resulting data was examined to identify similar areas and number of venues in each of these areas in a cluster.

	Neighbourhood	Region	latitude	longitude	venue_name	venue_longitude	venue_latitude	venue_category
0	Cantonment area	Central	13.019567	77.509589	Maruthi Theatre	13.018795	77.505283	Indie Movie Theater
1	Cantonment area	Central	13.019567	77.509589	bhat canteen	13.016796	77.504154	Fast Food Restaurant
2	Domlur	Central	12.962467	77.638196	Lavonne	12.963909	77.638579	Café
3	Domlur	Central	12.962467	77.638196	Barbeque Nation	12.962684	77.641599	BBQ Joint
4	Domlur	Central	12.962467	77.638196	Smoke House Deli	12.965584	77.641498	Deli / Bodega

More than 180 unique venue categories were identified with the help of Fairquare API.

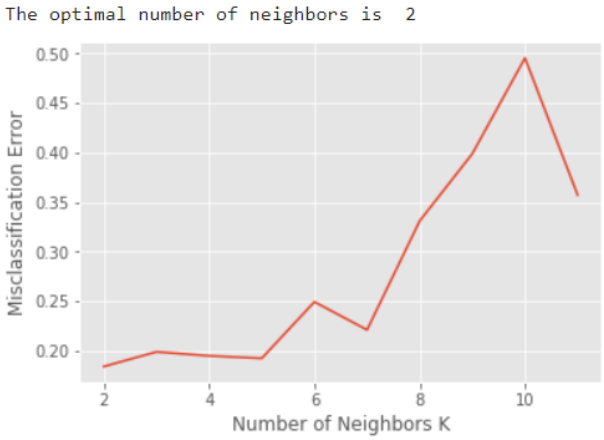


Methodology

K nearest neighbor (knn), a supervised learning algorithm, was used to cluster the neighborhoods into groups of similar areas. The independent variables in our case were the venue categories identified using the FourSquare API.

Statistical mean of the venue categories per neighborhood forms the inputs in identifying the cluster details. Choosing the right value of cluster required the dataset to run against different values ranging from 2 to 12 and calculating the optimal 'k' with the least error.

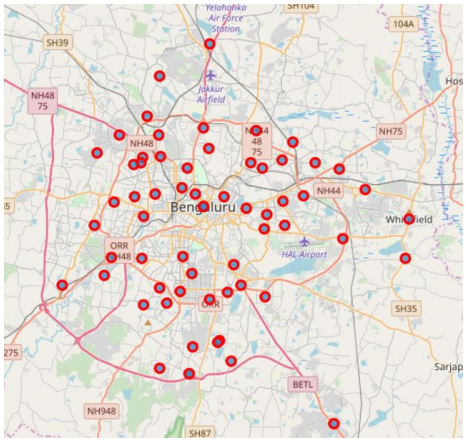
KCluster	MSE
2	0.184127
3	0.198889
4	0.194870
5	0.192381
6	0.249315
7	0.221093
8	0.330952
9	0.398730
10	0.495238
11	0.357143



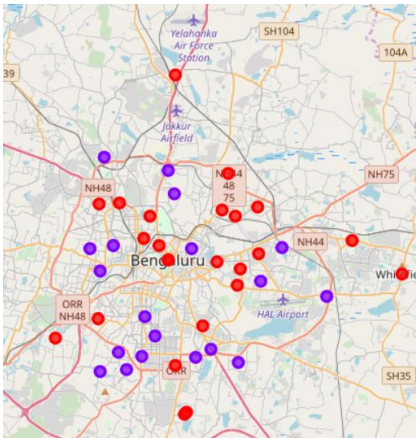
Optimal value for number of cluster was identified at 2 where the error is the lowest. Kmeans methodology was executed against the dataset with the optimal number of 'k' and the further analysis was carried out. The results of the same are discussed in the following sections of the Report.

Results

Based on kmeans methodology, with optimal value of k, the following 2 clusters were identified.

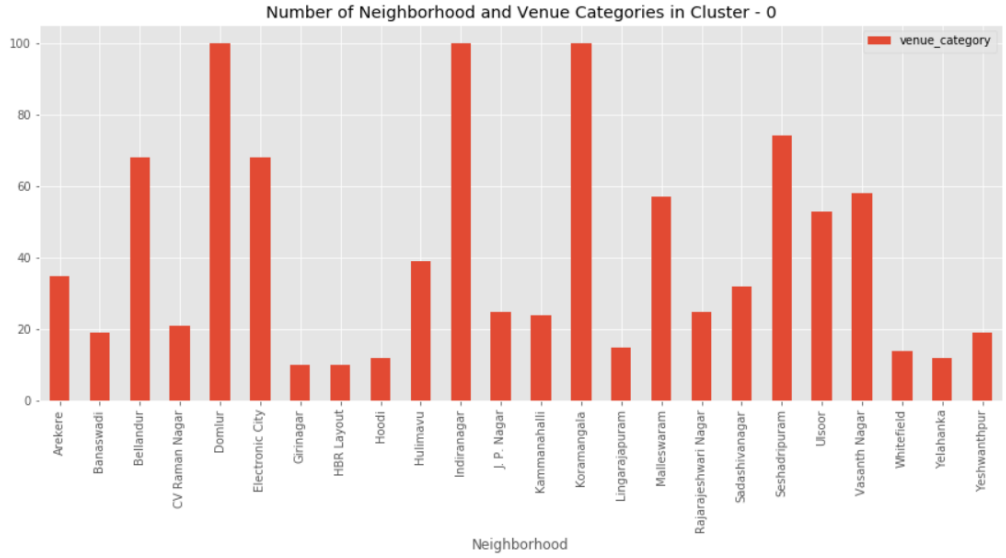


Neighborhoods in Bengaluru

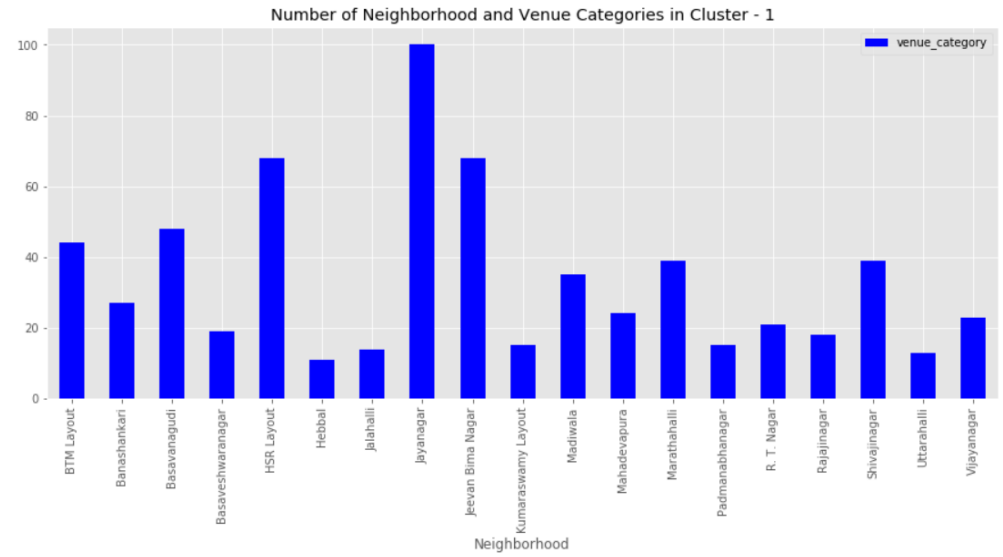


Clusters in the city of Bengaluru

Details of neighbourhoods in each cluster



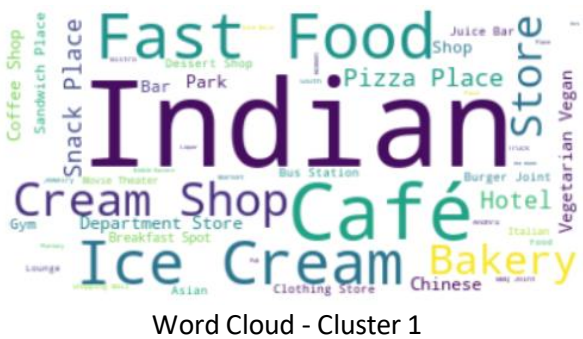
Cluster – 0
(Count – 24)



Cluster – 1
(Count – 19)

Given the venue_category that we receive using FourSquare API, word cloud was the optimum choice for visualizing different key words that is repeated in each of these clusters.

Word Cloud - Cluster 0



Word Cloud - Cluster 1

Discussion

As per the results, it is observed that neighborhoods in both the clusters have a high presence of **Indian Restaurants and Café.** Though restaurant catering to other cuisines are very less. Categories related to Ice-creams, Pizza are moderately available in each of the clusters.

Conclusion

As per the business problem, most common venue is an Indian restaurant in either of the cluster, hence it would be advisable to not venture into a similar category. Restaurant catering to a specific cuisine other than Indian is less in competition and would be advisable to venture into.