RECOMMENDATION TO START NEW RESTAURANT BUSINESS IN GOA

Table of Contents

I. INTRODUCTION

A. DESCRIPTION & DISUCSSION OF THE BACKGROUND

Goa is a state in western India with coastlines stretching along the Arabian Sea. Its long history as a Portuguese colony prior to 1961 is evident in its preserved 17th-century churches and the area's tropical spice plantations.

The state of Goa, in India, is famous for its beaches and places of worship, and tourism is its primary industry. Tourism is generally focused on the coastal areas of Goa, with decreased tourist activity inland. Foreign tourists, mostly from Europe, arrive in Goa in winter, whilst the summer and monsoon seasons see many Indian tourists.

In recent years the restaurant industry in the Goa has undergone a period of growth. Considering Goa's diversity and ethnicity it is evident that starting a restaurant business is a good alternative. This article can serve as one of the guides to start a restaurant business based on a particular cuisine like Indian, Italian, Chinese, etc., by providing a specific location. The number of restaurants in a specific location categorized based on cuisine and population distribution based on ethnicity and culture are some of the features considered for analysis.

B. PROBLEM DESCRIPTION

A restaurant is a business which prepares and serves food and drink to customers in return for money, either paid before the meal, after the meal or with an open account. Goa is famous for its excellent cuisine. Its food culture includes an array of international cuisines influenced by the city's tourism history.

So, it is evident that to survive in such competitive market it is very important to strategically plan. Various factors need to be studied in order to decide on location such as;

1. Goa Population and demographics
2. Who are the competitors in that location?
3. Cuisine served / Menu of the competitors
4. Are there any venues like Tourist attractions, Entertainment zones, Parks etc., nearby where floating population is high?
5. Segmentation of the Talukas

And the list goes on...

Even though well-funded restaurant Company needs to choose the correct location to start its first venture. If this is successful, they can replicate the same in other locations. First move is very important, thereby choice of location is very important.

OBJECTIVE

The objective is to locate and recommend to the management which neighborhood of Goa will be best choice to start a restaurant based on cuisine. The management also expects to understand the rationale of the recommendations made.

SUCCESS CRITERIA

The success criteria of the project will be a good recommendation of neighborhood choice to Company based on lack of such restaurants considering cuisine as a factor in that location.

II. DATA

To build a recommendation model, following datasets and information are considered for analysis;

1. Scrapped Wikipedia using BeautifulSoup, to extract information about [12 Talukas](#) for detailed analysis.
2. I used Foursquare API to get information about available restaurants for a given city and Taluka in Goa. The API also provided information about restaurant styles based on cuisine.
3. Employed data provided by the Government of Goa from www.goa.gov.in to get more insights about Talukas of Goa. The data provided knowledge about the population density, employment and many more.

III. METHODOLOGY

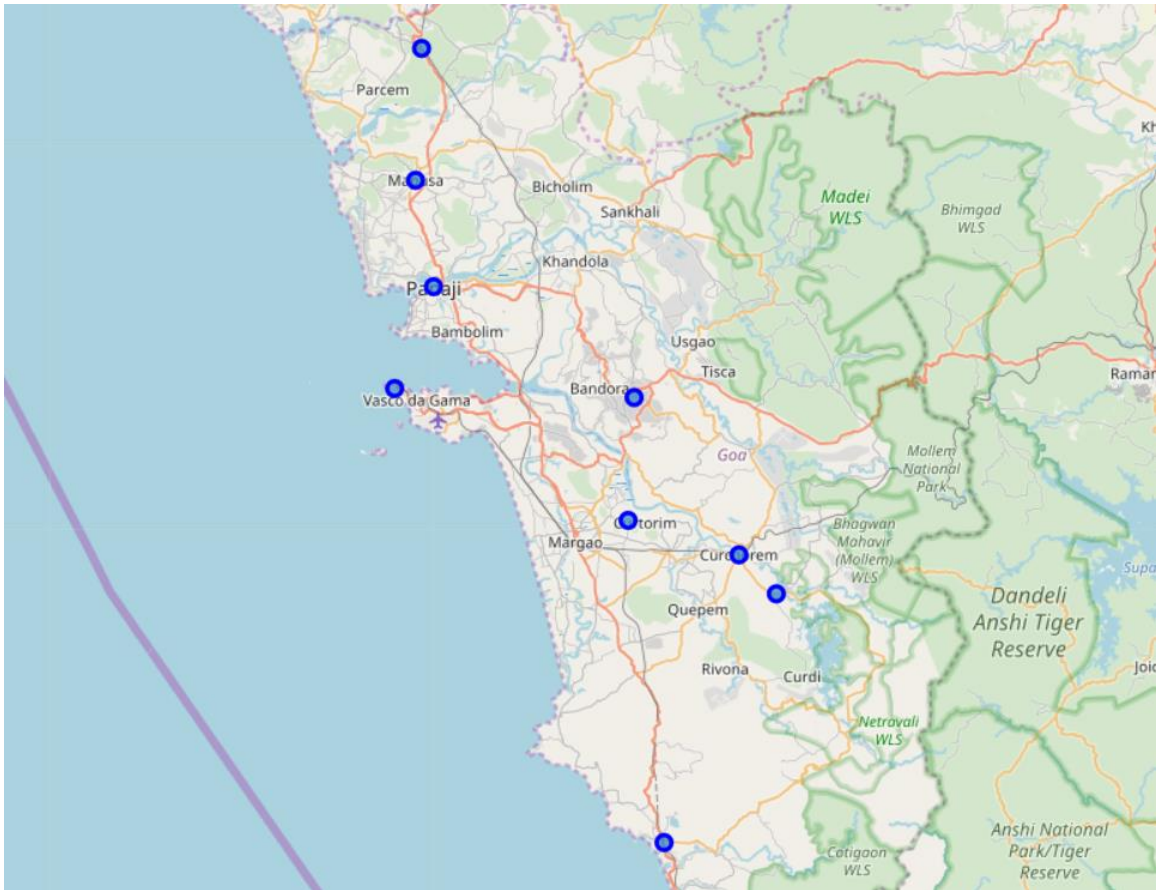BUSINESS UNDERSTANDING:

Our main goal is to get optimum location for new restaurant business in Goa for the company based on cuisine.

A. EXPLORATORY DATA ANALYSIS

Goa is mainly divided into 12 Talukas. We select major cities in each talukas. Geopy and folium libraries have been used to create a map to visualize neighborhoods of Goa.

All the blue markers on the map above are for cities in talukas.

B. PROBLEM APPROACH AND USING K-MEANS CLUSTERING

The process starts with scrapping the Wikipedia web page containing information about Goa's Taluka and its cities using library BeautifulSoup. During the process Pandas data frame is constructed having taluka and corresponding cities information.

	Taluka	Population	District	City
0	Bicholim	16986, 13651	North Goa, North Goa	Bicholim, Sanquelim
1	Canacona	12434	South Goa	Canacona
2	Salcete	16623, 87650	South Goa, South Goa	Cuncolim, Margao
3	Quepem	22730, 14795	South Goa, South Goa	Curchorem, Quepem
4	Bardez	40487	North Goa	Mapusa
5	Mormugao	94393	South Goa	Mormugao
6	Tiswadi	40017	North Goa	Panaji
7	Pernem	5289	North Goa	Pernem
8	Ponda	22664	South Goa	Ponda
9	Sanguem	6444	South Goa	Sanguem

Below is the data with Latitude and Longitude information for each city.

	Taluka	Population	District	City	Latitude	Longitude
1	Canacona	12434	South Goa	Canacona	15.0068	74.0395
2	Salcete	16623, 87650	South Goa, South Goa	Cuncolim, Margao	15.2923	74.0069
3	Quepem	22730, 14795	South Goa, South Goa	Curchorem, Quepem	15.2617	74.1083
4	Bardez	40487	North Goa	Mapusa	15.5927	73.8118
5	Mormugao	94393	South Goa	Mormugao	15.4082	73.793
6	Tiswadi	40017	North Goa	Panaji	15.499	73.8282
7	Pernem	5289	North Goa	Pernem	15.709	73.8168
8	Ponda	22664	South Goa	Ponda	15.4004	74.0116
9	Sanguem	6444	South Goa	Sanguem	15.2277	74.142

The neighborhoods in the resulting data frame has to be plotted on a map using folium library. In order to do so, latitude and longitude value for each neighborhood is determined using geopy library. After getting latitude and longitude values, it is merged with the original data frame containing neighborhood and taluka information appropriately. Finally, each neighborhood is plotted on folium map. Neighborhood with undetermined latitude and longitude values by geopy are excluded from evaluation.
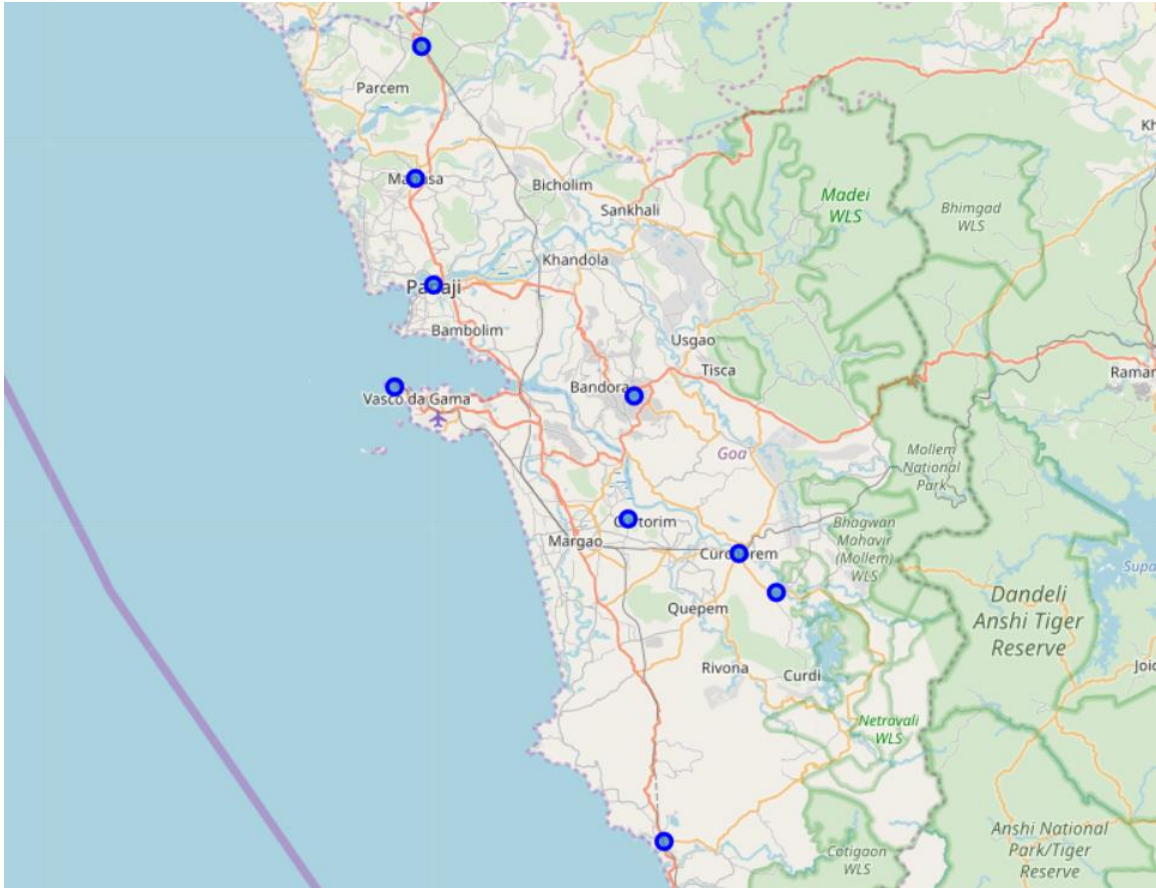
Now Foursquare API is used to explore the neighborhoods and segment them. First, data frame with all the venues information provided by foursquare for the given latitude and longitude values. Foursquare API returns 84 results.
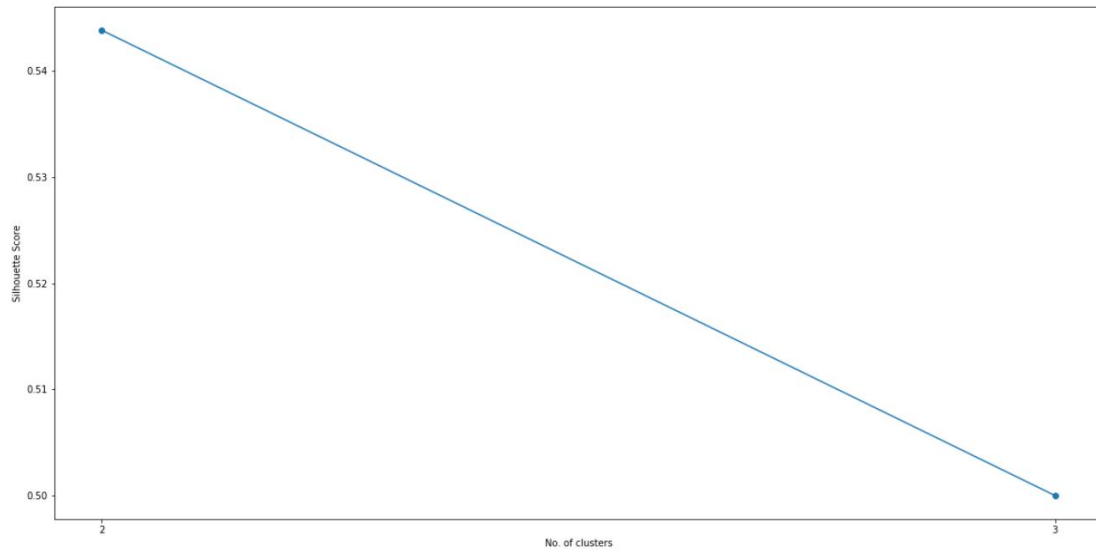
	Neighborhood	Neighborhood Latitude	Neighborhood Longitude	Venue	Venue Latitude	Venue Longitude	Venue Category
0	Canacona	15.006842	74.039489	Canacona Railway Station	15.006842	74.039379	Train Station
1	Canacona	15.006842	74.039489	Jasmine's Shack	15.005048	74.039935	Breakfast Spot
2	Canacona	15.006842	74.039489	Pai Chemist	15.008872	74.040630	Pharmacy
3	Canacona	15.006842	74.039489	Kadamba Bus Stand	15.008476	74.041817	Bus Station
4	Cuncolim, Margao	15.292336	74.006904	HDFC Bank	15.290946	74.007448	ATM
...	...	...	...	...	...	...	...
79	Pernem	15.708996	73.816760	Mahalaxmi Bhojanalay (NH17)	15.709986	73.813960	Seafood Restaurant
80	Ponda	15.400397	74.011610	St. Ana Bakery	15.401057	74.013129	Bakery
81	Ponda	15.400397	74.011610	Perfect Taste Fast Food	15.401699	74.011513	Fast Food Restaurant
82	Ponda	15.400397	74.011610	Ponda Fish Market	15.401696	74.008670	Fish Market
83	Ponda	15.400397	74.011610	Ponda Market	15.402069	74.008354	Market

Now, only restaurants are extracted from venue category list. The resulting data frame has 11 unique categories or cuisines available in Goa. Next, one hot encoding is performed on the resulting data frame for each neighborhood. The results have 19 unique restaurants in Goa with 11 different style of cuisines.

Rows are grouped by neighborhood to determine the frequency of occurrence of each restaurant. A new data frame is created with each row assigned for neighborhood and its corresponding top ten common restaurants based on cuisine.

	Neighborhood	American Restaurant	Asian Restaurant	Fast Food Restaurant	Indian Restaurant	Mughlai Restaurant	Portuguese Restaurant	Punjabi Restaurant	Restaurant	Seafood Restaurant	Southern / Soul Food Restaurant	Vegetarian / Vegan Restaurant
0	Mapusa	0.0000	0.0000	0.0	0.000	0.0000	0.0000	0.0000	0.000	1.0000	0.0000	0.0000
1	Panaji	0.0625	0.0625	0.0	0.375	0.0625	0.0625	0.0625	0.125	0.0625	0.0625	0.0625
2	Pernem	0.0000	0.0000	0.0	0.000	0.0000	0.0000	0.0000	0.000	1.0000	0.0000	0.0000
3	Ponda	0.0000	0.0000	1.0	0.000	0.0000	0.0000	0.0000	0.000	0.0000	0.0000	0.0000

Finally, k-means clustering is performed on the data frame to check the pattern for each neighborhood and get the information about the top ten common restaurants for each neighborhood. Before fitting the data frame best value of k for k-means clustering is determined by based on silhouette_score from sklearn.metrics. It is observed from the graph below that, k = 2 would yield more better results for the computation using k-means clustering.

Finally, k-means clustering is performed on the following data frame with k = 2, to determine the pattern of top 10 restaurants based on cuisine in every neighborhood.

	Neighborhood	1st Most Common Venue	2nd Most Common Venue	3rd Most Common Venue	4th Most Common Venue	5th Most Common Venue	6th Most Common Venue	7th Most Common Venue	8th Most Common Venue	9th Most Common Venue	10th Most Common Venue
0	Mapusa	Seafood Restaurant	Vegetarian / Vegan Restaurant	Southern / Soul Food Restaurant	Restaurant	Punjabi Restaurant	Portuguese Restaurant	Mughlai Restaurant	Indian Restaurant	Fast Food Restaurant	Asian Restaurant
1	Panaji	Indian Restaurant	Restaurant	Vegetarian / Vegan Restaurant	Southern / Soul Food Restaurant	Seafood Restaurant	Punjabi Restaurant	Portuguese Restaurant	Mughlai Restaurant	Asian Restaurant	American Restaurant
2	Pernem	Seafood Restaurant	Vegetarian / Vegan Restaurant	Southern / Soul Food Restaurant	Restaurant	Punjabi Restaurant	Portuguese Restaurant	Mughlai Restaurant	Indian Restaurant	Fast Food Restaurant	Asian Restaurant
3	Ponda	Fast Food Restaurant	Vegetarian / Vegan Restaurant	Southern / Soul Food Restaurant	Seafood Restaurant	Restaurant	Punjabi Restaurant	Portuguese Restaurant	Mughlai Restaurant	Indian Restaurant	Asian Restaurant

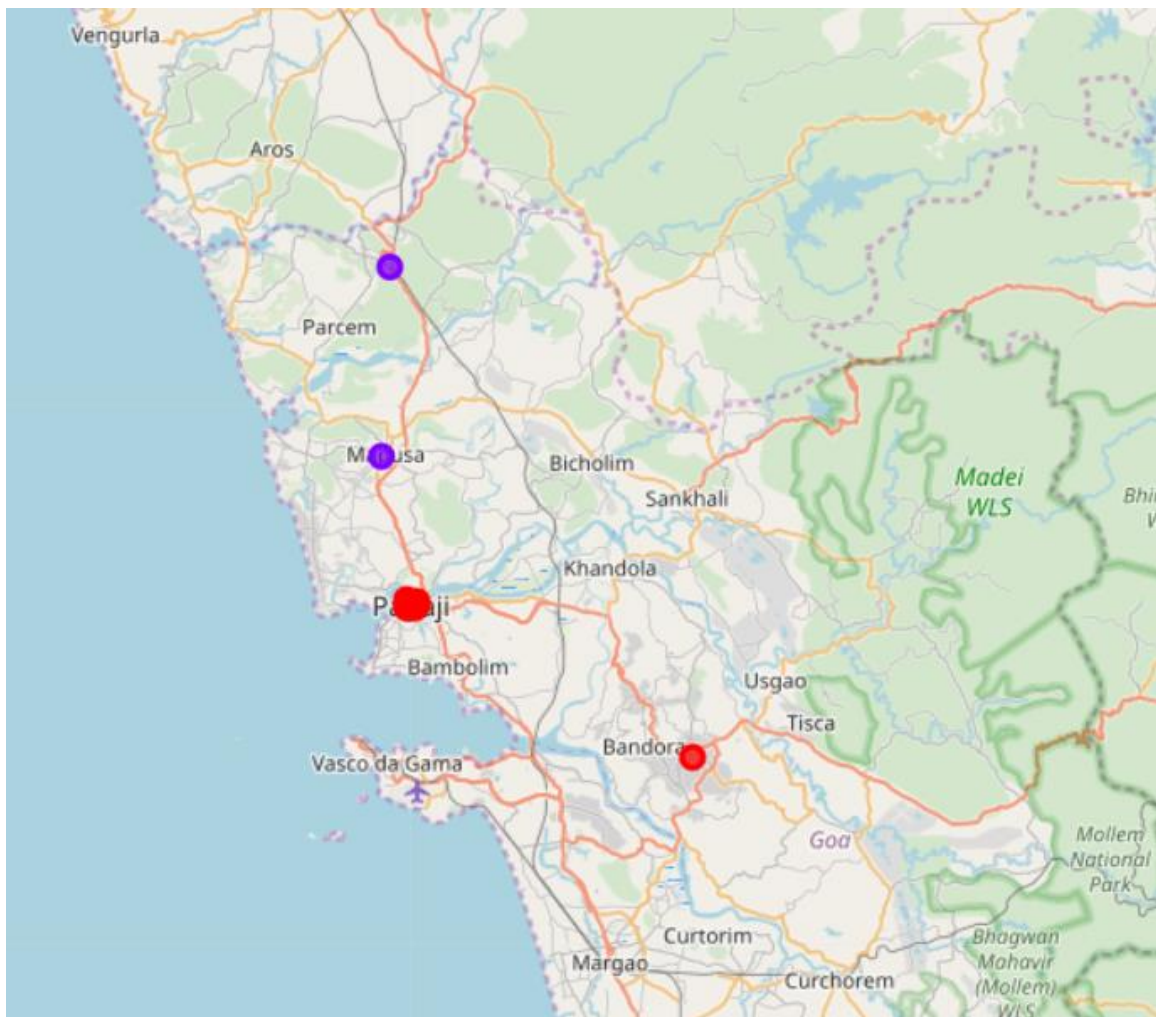|	Neighborhood	Neighborhood Latitude	Neighborhood Longitude	Venue	Venue Latitude	Venue Longitude	Venue Category	Cluster Labels	1st Most Common Venue	2nd Most Common Venue	3rd Most Common Venue	4th Most Common Venue	5th Most Common Venue	6th Most Common Venue
20	Mapusa	15.592674	73.811762	Kamlabai Restaurant	15.591089	73.808934	Seafood Restaurant	1	Seafood Restaurant	Vegetarian / Vegan Restaurant	Southern / Soul Food Restaurant	Restaurant	Punjabi Restaurant	Portuguese Restaurant
29	Panaji	15.498995	73.828214	Ritz Classic Family Restaurant	15.498470	73.826742	Indian Restaurant	0	Indian Restaurant	Restaurant	Vegetarian / Vegan Restaurant	Southern / Soul Food Restaurant	Seafood Restaurant	Punjabi Restaurant
30	Panaji	15.498995	73.828214	Delhi Darbar	15.500507	73.827109	Mughlai Restaurant	0	Indian Restaurant	Restaurant	Vegetarian / Vegan Restaurant	Southern / Soul Food Restaurant	Seafood Restaurant	Punjabi Restaurant
32	Panaji	15.498995	73.828214	George Bar & Restaurant	15.498843	73.828720	Indian Restaurant	0	Indian Restaurant	Restaurant	Vegetarian / Vegan Restaurant	Southern / Soul Food Restaurant	Seafood Restaurant	Punjabi Restaurant
38	Panaji	15.498995	73.828214	Goenchin	15.496678	73.827239	Asian Restaurant	0	Indian Restaurant	Restaurant	Vegetarian / Vegan Restaurant	Southern / Soul Food Restaurant	Seafood Restaurant	Punjabi Restaurant

III. RESULTS

NEIGHBORHOOD K-MEANS CLUSTERING ON MEAN OCCURRENCE OF VENUE CATEGORY:

All clusters follow unique pattern for top ten common restaurants for a particular neighborhood. The detail shows the number of neighborhoods assigned to each cluster. Cluster 1 has the highest neighborhoods 17 and cluster 2 has the least with 2.
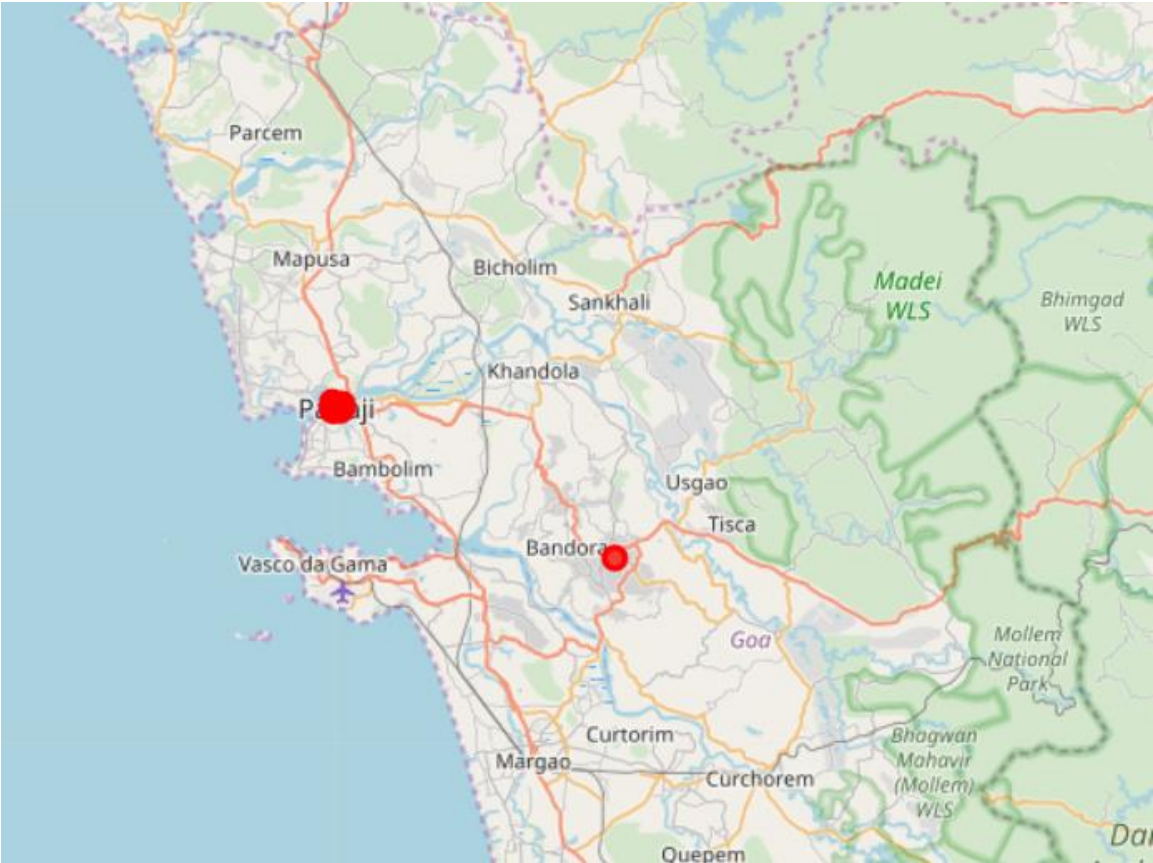
```
0    17
1     2
Name: Cluster Labels, dtype: int64
```

The clustered map is shown below;

The resulting data frames for each cluster are shown below.

- Cluster 1 (Most Common: Indian and Fast Food Restaurant)



| | Neighborhood | 1st Most Common Venue | 2nd Most Common Venue | 3rd Most Common Venue | 4th Most Common Venue | 5th Most Common Venue | 6th Most Common Venue | 7th Most Common Venue | 8th Most Common Venue | 9th Most Common Venue | 10th Most Common Venue |
|---|---|---|---|---|---|---|---|---|---|---|---|
| 29 | Panaji | Indian Restaurant | Restaurant | Vegetarian / Vegan Restaurant | Southern / Soul Food Restaurant | Seafood Restaurant | Punjabi Restaurant | Portuguese Restaurant | Mughlai Restaurant | Asian Restaurant | American Restaurant |
| 81 | Ponda | Fast Food Restaurant | Vegetarian / Vegan Restaurant | Southern / Soul Food Restaurant | Seafood Restaurant | Restaurant | Punjabi Restaurant | Portuguese Restaurant | Mughlai Restaurant | Indian Restaurant | Asian Restaurant |

- Cluster 2 (Most Common: Seafood Restaurant)
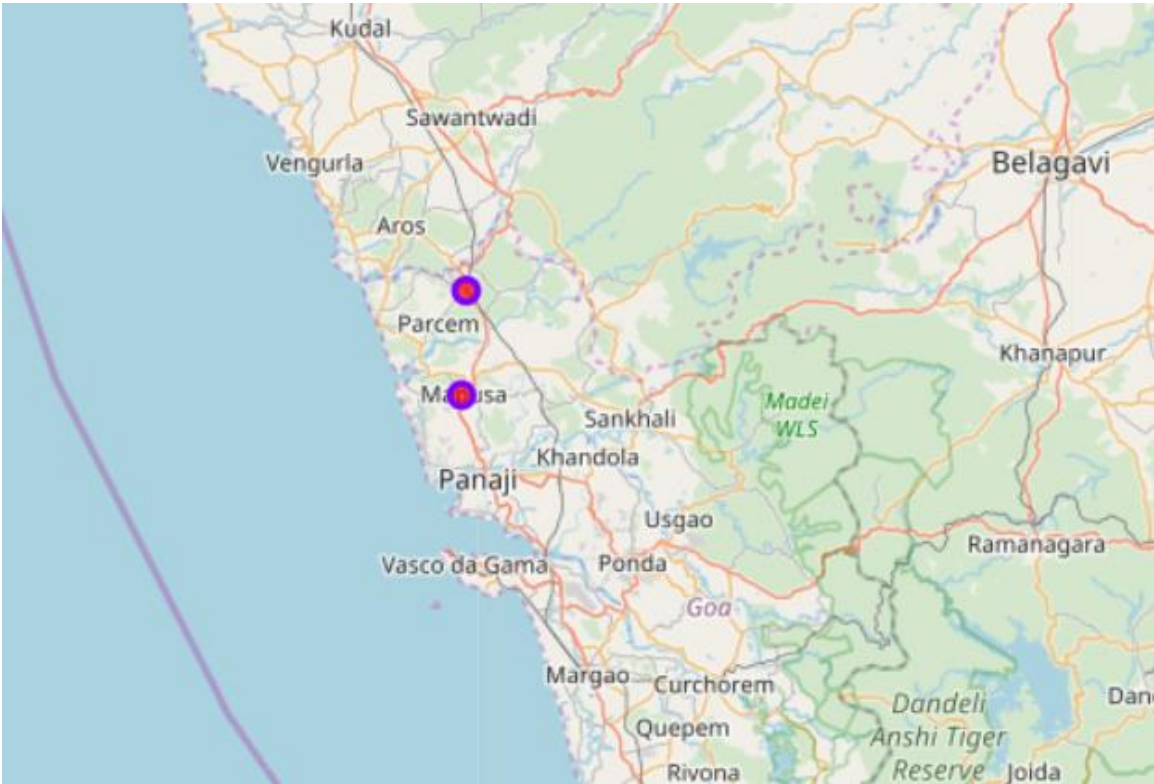


| | Neighborhood | 1st Most Common Venue | 2nd Most Common Venue | 3rd Most Common Venue | 4th Most Common Venue | 5th Most Common Venue | 6th Most Common Venue | 7th Most Common Venue | 8th Most Common Venue | 9th Most Common Venue | 10th Most Common Venue |
|---|---|---|---|---|---|---|---|---|---|---|---|
| 20 | Mapusa | Seafood Restaurant | Vegetarian / Vegan Restaurant | Southern / Soul Food Restaurant | Restaurant | Punjabi Restaurant | Portuguese Restaurant | Mughlai Restaurant | Indian Restaurant | Fast Food Restaurant | Asian Restaurant |
| 79 | Pernem | Seafood Restaurant | Vegetarian / Vegan Restaurant | Southern / Soul Food Restaurant | Restaurant | Punjabi Restaurant | Portuguese Restaurant | Mughlai Restaurant | Indian Restaurant | Fast Food Restaurant | Asian Restaurant |

## IV. DISCUSSION

The results can be approached in two ways;

1. *If company want to open a restaurant in preferred location and irrespective of cuisine, refer to that neighborhood in specific cluster and chose cuisine with the least common restaurant for better profits*
2. *If company wants to open a restaurant with a preferred cuisine and irrespective of location, refer to the cluster with the least number of restaurants with that specific cuisine and select one among the neighborhoods based on company's preference.*

## V. CONCLUSION

This analysis is performed on limited data. This may be right or may be wrong. But if good amount of data is available there is scope to come up with better results. If there are lot of restaurants probably there is lot of demand. Goa has so many restaurants, yet certain neighborhood or taluka doesn't have a specific cuisine restaurant available.