



Hotel Apartment in Kochi, Kerala

– Analysis Report



MAY 28

Coursera Capstone Project
Authored by: Midhun Madhavan



Logo
Name

Introduction

BACKGROUND:

Kochi is a beautiful city situated on the Western Coast of India facing the Arabian Sea, and rightfully called the 'Queen of the Arabian Sea'. It has a rich network of backwaters and is the culture and heritage center of Kerala. Cochin known as the Queen of Arabian Sea, is the Commercial Capital of Kerala.

An enormous potential exists in Kerala, especially in Cochin for the hospitality industry. There are several IT parks, industries, many export companies, 100% literacy, highly developed social structure, and well laid-out communication facilities and transport infrastructure. These and a few other factors provide enormous scope for the growth of hospitality industry in Cochin.

BUSINESS PROBLEM:

My client wants to utilize the hospitality business potential of Kochi, and he is ready to open a hotel apartment in the region.

The objective of this project is to help the client in selecting a best location in Kochi to start his new hotel service apartment business. Choosing the right location is the key to the success of the business.

There are several factors needs to be considered for selecting the right location,

1. Accessibility, transportation facility etc.
2. Density of other hotel and hotel apartments
3. Industries and commercial establishments around the region

This project will attempt to explore patterns of suburbs of Kochi by using data science methodology like categorizing them into clusters in order to determine the best location to open a hotel apartment.

TARGET AUDIENCE:

This project is particularly useful to hotel businesspeople, property developers and investors looking to open or invest in new hotel apartments in Kochi.

Data

Based on the above factors we need following data to solve the problem.

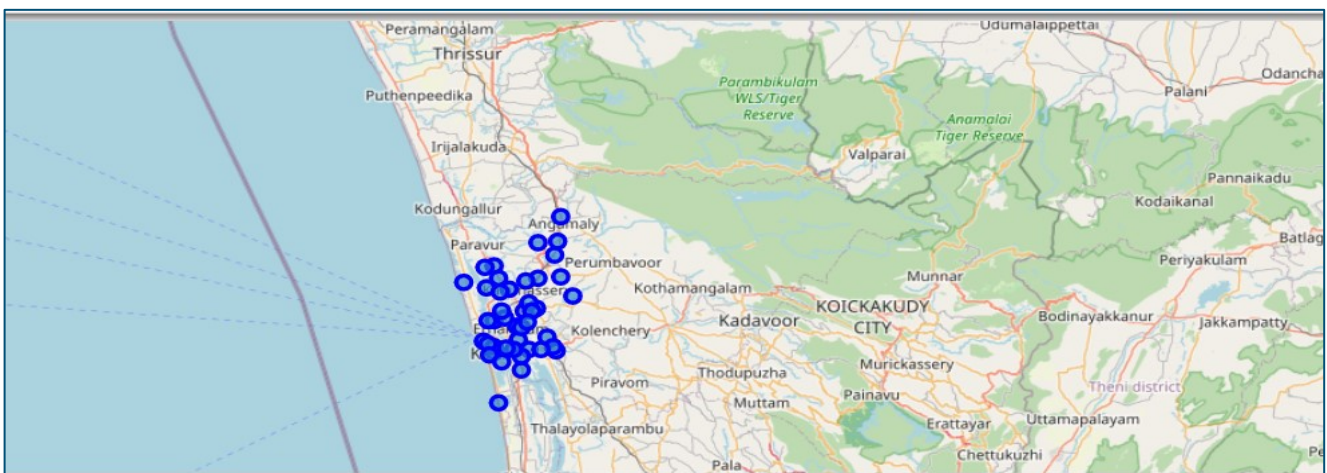
1. Suburbs of Kochi. Source of data is https://en.wikipedia.org/wiki/Category:Suburbs_of_Kochi
2. Location data, this data will be extracted from opensource.
3. Venue data –Venue queries will then be made by suburbs with Foursquare APIs. The resulting data regarding venue category will be used to observe commonality between subdistricts. We will use this data to perform clustering on the neighborhoods. K-means clustering algorithm will be used to find pattern between the suburbs.

Methodology

The following steps are performed in order to solve the business problem mentioned in above section.

DATA PREPARATION:

First step is to prepare and clean the data. The list of suburbs of Kochi is available in the Wikipedia page. The data is converted a data frame and then added locations data using Geocoder package. Then we visualized the data in a map using Folium package. This allowed us to perform a quick check to make sure that the geographical coordinates data returned by Geocoder are correctly plotted in the Kochi city.



EXPLORATORY DATA ANALYSIS:

Using the Foursquare API, we got the top 100 venues that are within a radius of 10 kilometers. Foursquare API provided us the venues with latitude and longitude.

Next step is to find out the categories and their numbers in each location which could attract the customers to select a hotel for short or long stay. We identified following venues categories which can be considered by a typical customer in selecting a hotel for stay,

- Transpiration facilities like bus stops, train stations etc.
- Shopping malls nearby
- Restaurants
- Cinema halls

These data are extracted and added to the data frame for each neighborhood.

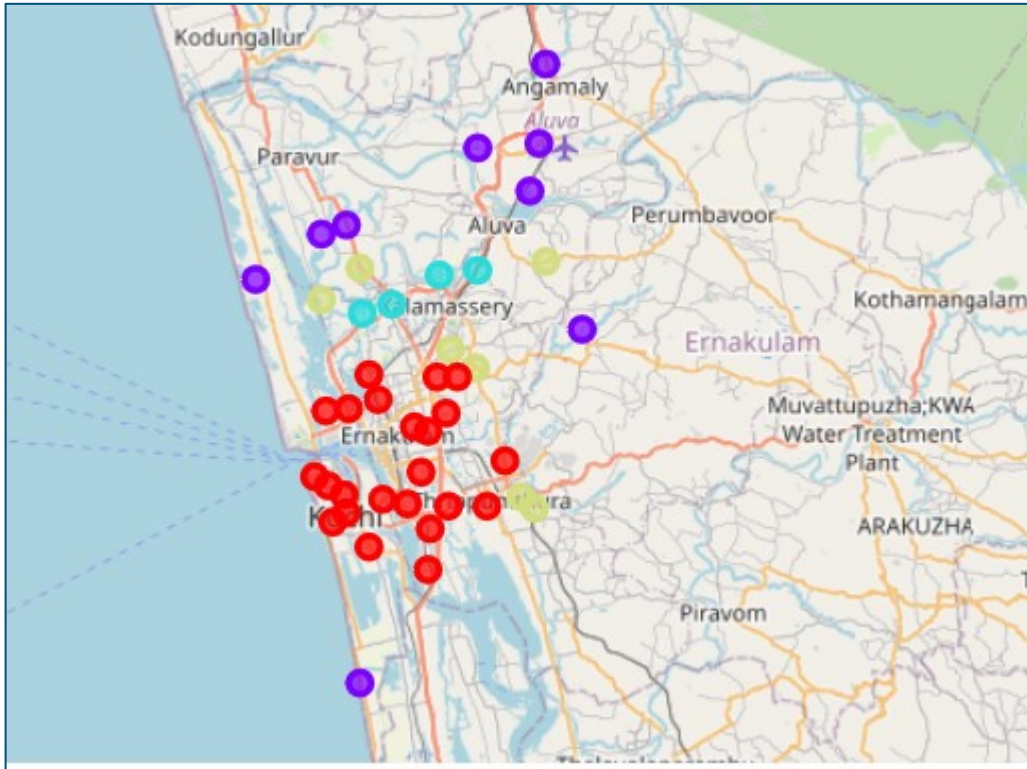
	Neighborhood	Latitude	Longitude	Number of Hotels	Number of Transporation points	Number of Restaurants	Number of Shopping Malls	Number of Cinemas
0	Alangad	10.847500	76.436090	0.0	1.0	1	0.0	0.0
1	Angamaly	10.203660	76.382680	4.0	1.0	8	0.0	2.0
2	Aroor	9.936010	76.261420	13.0	0.0	30	1.0	0.0
3	Chellanam	9.835260	76.270290	1.0	2.0	2	1.0	0.0
4	Chendamangalam	10.022316	76.339358	5.0	0.0	35	2.0	1.0

CLUSTER ANALYSIS:

To identify groups (clusters) with similar characteristics, the unsupervised learning method to our data, namely K-Means algorithm, was applied to our data. We clustered the neighborhoods into 4 clusters based on the factors mentioned above.

The results allowed us to identify which neighborhoods have higher number of hotels and what are the depending factors which decide the selection of a location.

Based on clustering results one map is created.



Results

Based on the analysis, we got the following results.

- Cluster 0: Neighborhoods has the high number of hotels.
- Cluster 1: Neighborhoods with very a smaller number of hotels,
- Cluster 2 and 3: Neighborhoods with smaller number of hotels but higher than Cluster 1

Discussions

We observed following points and dependencies with other features,

- Cluster 0 has the greatest number of hotels, and cluster 0 is closer to the city center. Cluster 0 also has the highest number of shopping malls, cinema and

Restaurants. This means location for hotel has a direct relation with facilities like restaurants, malls etc.

	Neighborhood	Latitude	Longitude	Number of Hotels	Number of Transportation points	Number of Restaurants	Number of Shopping Malls	Number of Cinemas	Cluster Labels
2	Aroor	9.936010	76.261420	13.0	0.0	30	1.0	0.0	0
7	Chilavannoor	9.961180	76.306590	10.0	0.0	32	2.0	1.0	0
12	Fort Kochi	9.957580	76.242390	13.0	0.0	31	1.0	0.0	0
13	Irumpanam	9.966870	76.357200	8.0	0.0	34	2.0	1.0	0
15	Karanakodam	9.988453	76.303426	10.0	0.0	28	2.0	1.0	0
16	Kochangadi	9.947600	76.260790	13.0	0.0	30	1.0	0.0	0
20	Kumbalam, Ernakulam	9.902200	76.310640	11.0	0.0	29	0.0	0.0	0
21	Maradu	9.940510	76.323950	10.0	0.0	32	2.0	1.0	0
22	Mattancherry	9.952060	76.250800	12.0	0.0	30	1.0	0.0	0
23	Mulavukad	9.999140	76.262410	10.0	0.0	31	2.0	1.0	0
24	Mundaveli	9.930700	76.253200	13.0	0.0	30	1.0	0.0	0

- Cluster 1 has very a smaller number of hotels, which means less competition but there are very fewer other facilities like restaurants, malls etc.

	Neighborhood	Latitude	Longitude	Number of Hotels	Number of Transportation points	Number of Restaurants	Number of Shopping Malls	Number of Cinemas	Cluster Labels
0	Alangad	10.84750	76.43609	0.0	1.0	1	0.0	0.0	1
1	Angamaly	10.20366	76.38268	4.0	1.0	8	0.0	2.0	1
3	Chellanam	9.83526	76.27029	1.0	2.0	2	1.0	0.0	1
5	Chengamanad, Ernakulam district	10.15354	76.34068	4.0	2.0	6	0.0	2.0	1
10	Chowwara	10.12815	76.37217	4.0	3.0	7	0.0	2.0	1
17	Koonammavu	10.10769	76.26171	1.0	0.0	4	0.0	0.0	1
19	Kottuvally	10.10315	76.24615	1.0	0.0	2	0.0	0.0	1
25	Nedumbassery	10.15669	76.37780	4.0	1.0	10	0.0	2.0	1
36	Twenty20 Kizhakkambalam	10.04626	76.40411	0.0	0.0	7	0.0	0.0	1
42	Vypin	10.07538	76.20702	1.0	0.0	1	0.0	0.0	1

- Cluster 2 has a smaller number of hotels, which means less competition. Also, the neighborhoods in this cluster has a considerable number of restaurants, malls.

	Neighborhood	Latitude	Longitude	Number of Hotels	Number of Transportation points	Number of Restaurants	Number of Shopping Malls	Number of Cinemas	Cluster Labels
6	Cheranallur	10.06085	76.28901	3.0	0.0	19	2.0	1.0	2
8	Choomikkara	10.08132	76.34155	4.0	0.0	17	2.0	1.0	2
18	Kothad	10.05561	76.27164	7.0	0.0	22	2.0	1.0	2
29	Pathalam	10.07817	76.31857	2.0	0.0	17	2.0	1.0	2

- Cluster 3 also has very a smaller number of hotels, which means less competition. Also, the neighborhoods in this cluster has a big number of restaurants, malls.

	Neighborhood	Latitude	Longitude	Number of Hotels	Number of Transportation points	Number of Restaurants	Number of Shopping Malls	Number of Cinemas	Cluster Labels
4	Chendamangalam	10.022316	76.339358	5.0	0.0	35	2.0	1.0	3
9	Chottanikkara	9.939097	76.375036	5.0	3.0	30	0.0	0.0	3
11	Edathala	10.086410	76.381810	6.0	0.0	25	0.0	0.0	3
14	Kadamakkudy	10.063520	76.246600	4.0	0.0	27	3.0	1.0	3
32	Thiruvankulam	9.946350	76.367460	6.0	3.0	27	1.0	0.0	3
34	Thrikkakkara South	10.033240	76.325190	5.0	0.0	35	2.0	1.0	3
39	Varappuzha	10.082610	76.270410	2.0	0.0	28	3.0	1.0	3

Conclusion

To conclude, the basic data analysis was performed to identify the most optimal locations for opening a hotel business in the city of Kochi. During the analysis, several important statistical features of the neighborhoods were explored and visualized.

Based on the analysis below mentioned clusters and neighborhoods are selected,

1. Cluster 3 is looking promising with not much competition and other customer attracting facilities like restaurant, malls and cinema. Consider locations like Kadamakkudy and Thiruvankulam.
2. If client is ready to accept the competition and provide a good quality service, then we can even consider the center of the city which is Cluster 0 where most of the hotels are located. Consider locations like Irumpanam and vaduthala.