

# **Data Science Capstone Project**

Analysing Business prospect for tourist venues in Kerala

**Author** : Sayani Roy

## Introduction

India is called a sub continent where 'Unity in diversity' is prominently visible. Several states of India represent several cultures. Tourism is one of most important business in India. Kerala is located in the southwestern most part of the country. All of 14 districts are as diverse as prosperous for tourism and new business.

In this analysis, we will analysis below details for state Kerala

1. Tourism details comparison among all districts
2. Business prospects for different districts.

## Description of the Data:

The data used are –

- Here, we are scraping data from the below page and converting to pandas dataframe

[https://en.wikipedia.org/wiki/List\\_of\\_districts\\_in\\_India](https://en.wikipedia.org/wiki/List_of_districts_in_India)

- Four Square API is being used to get venues details for neighborhoods in Kerala.

## Sample Data:

	State	Code	District	Headquarters	Population	Area in km2	Density per km2	Latitude	Longitude
0	Andhra Pradesh	AN	Anantapur	Anantapur	4083315	19130	213	14.685564	77.595406
1	Andhra Pradesh	CH	Chittoor	Chittoor	4170468	15152	275	13.217096	79.100677
2	Andhra Pradesh	CU	Kadapa	Kadapa	2884524	15359	188	14.477234	78.804932
3	Andhra Pradesh	EG	East Godavari	Kakinada	5151549	10807	477	17.321250	82.040714
4	Andhra Pradesh	GU	Guntur	Guntur	4889230	11391	429	16.314209	80.435028

	State	Code	District	Headquarters	Population	Area in km2	Density per km2	Latitude	Longitude
0	Kerala	AL	Alappuzha	Alappuzha	2121943	1415	1501	9.498067	76.338844
1	Kerala	ER	Ernakulam	Kakkanad	3279860	3063	1069	9.931233	76.267303
2	Kerala	ID	Idukki	Painavu	1107453	4356	254	9.813808	76.929735
3	Kerala	KN	Kannur	Kannur	2525637	2961	852	11.874477	75.370369
4	Kerala	KS	Kasaragod	Kasaragod	1302600	1989	654	12.507643	74.988213
5	Kerala	KL	Kollam	Kollam	2629703	2483	1056	8.893212	76.614143
6	Kerala	KT	Kottayam	Kottayam	1979384	2206	896	9.591441	76.522171
7	Kerala	KZ	Kozhikode	Kozhikode	3089543	2345	1318	11.258753	75.780411
8	Kerala	MA	Malappuram	Malappuram	4110956	3554	1058	11.072035	76.074005
9	Kerala	PL	Palakkad	Palakkad	2810892	4482	627	10.784703	76.653145
10	Kerala	PT	Pathanamthitta	Pathanamthitta	1195537	2652	453	9.264758	76.787041
11	Kerala	TS	Thrissur	Thrissur	3110327	3027	1026	10.517096	76.212578
12	Kerala	TV	Thiruvananthapuram	Thiruvananthapuram	3307284	2189	1509	8.524139	76.936638
13	Kerala	WA	Wayanad	Kalpetta	816558	2130	383	11.703206	76.083400

### Methodology:

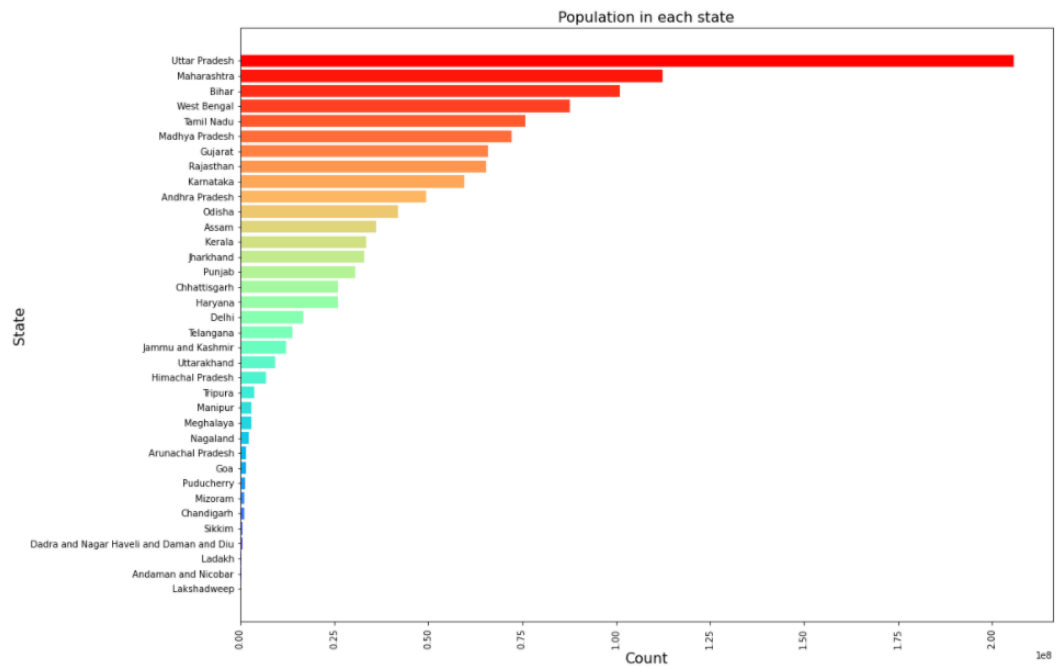
In this project India data is taken as a base which contains info of all states and the corresponding districts. Some exploratory data analysis applied –

- Bar graph to visualize all state population for India
- From the dataframe we will subset the dataset to get only the records for Kerala,
- Data cleaning with removal of any blank fields.
- Visualize all districts of Kerala using FOLIUM library.
- Bar graph to visualize number of tourist venues in 14 districts of Kerala
- Fetching Latitude and Longitude of venues using Geopy Library
- Normalizing the data fetched from Foursquare API.

Machine learning technique –

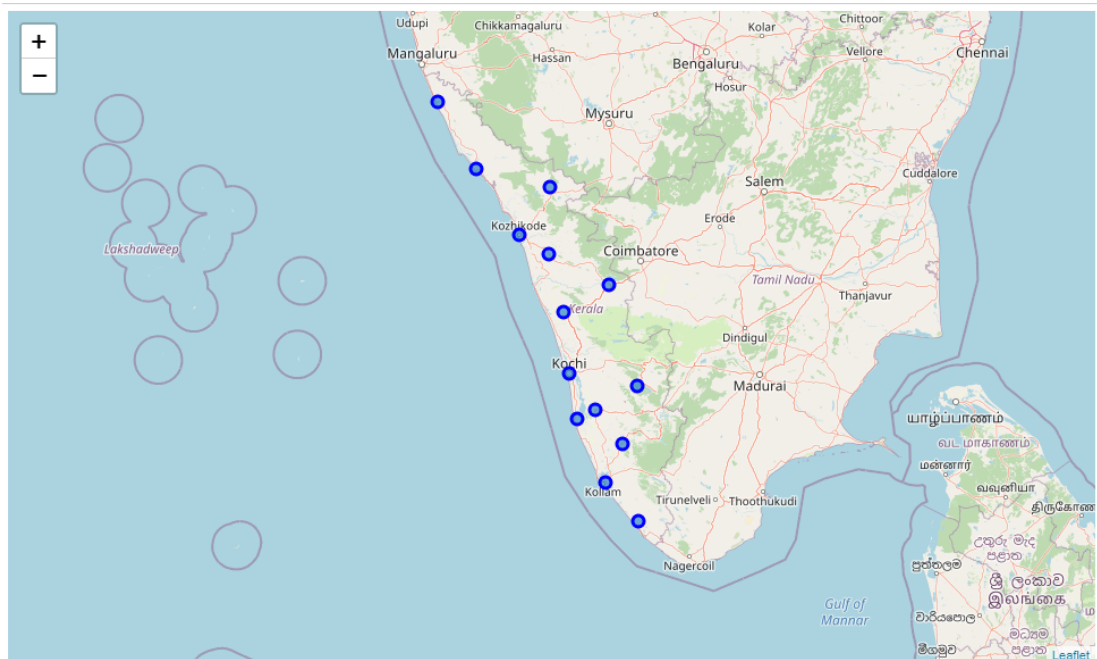
We will use Clustering technique to segment the venues and cluster them using K-MEANS clustering to get the clusters with similarities in their venues.

## Bar plot to visualize all state population for India

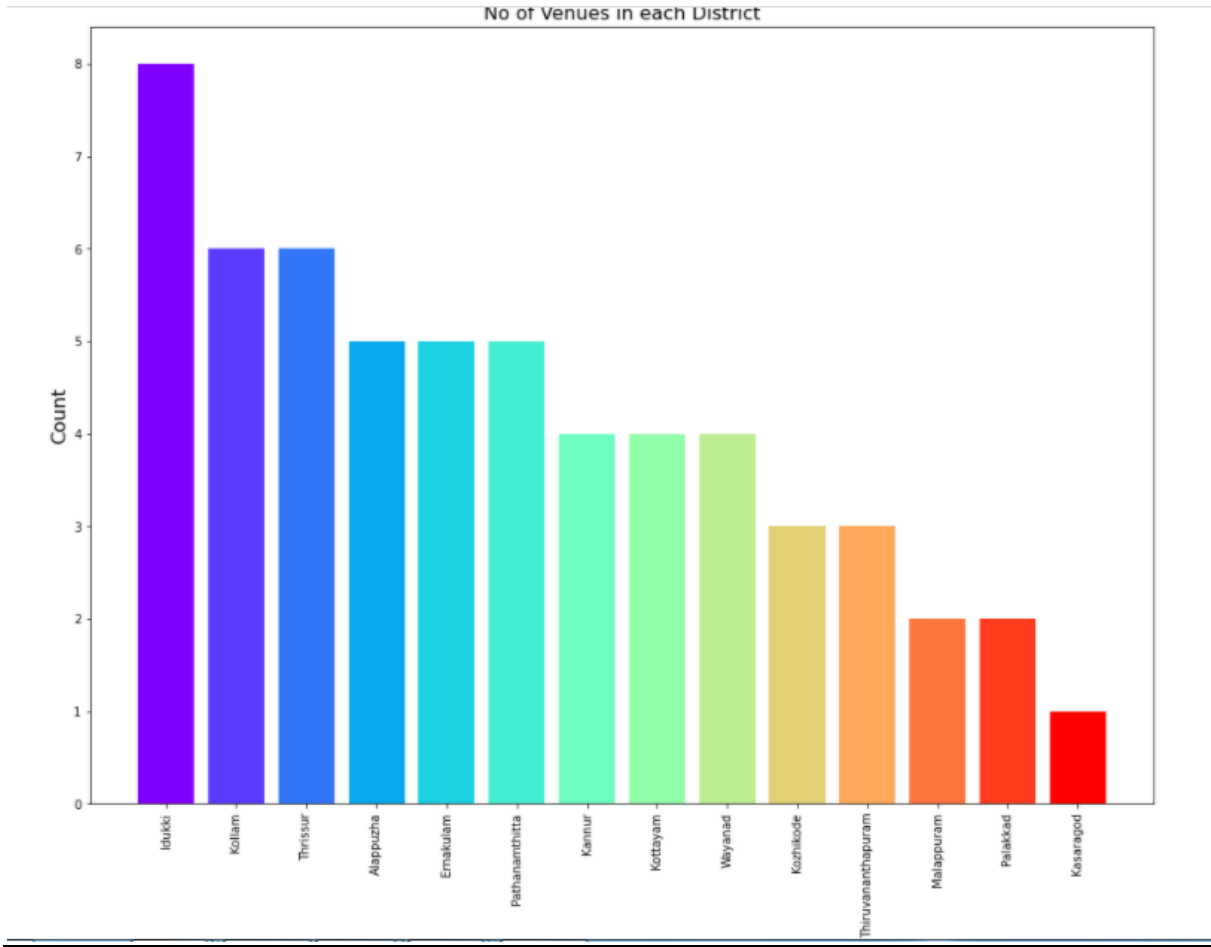


## Visualize all Districts of Kerala using FOLIUM library

Out[55]:



**Bar graph to visualize number of tourist venues in 14 districts of Kerala**



**The most common venue category in a particular district**

	District	1st Most Common Venue Category	2nd Most Common Venue Category	3rd Most Common Venue Category	4th Most Common Venue Category	5th Most Common Venue Category	6th Most Common Venue Category	7th Most Common Venue Category	8th Most Common Venue Category
0	Alappuzha	Beach	Lighthouse	Scenic Lookout	Lake	Zoo	Campground	Field	Garden
1	Ernakulam	Art Gallery	Stadium	Performing Arts Venue	Lake	Beach	Campground	Field	Garden
2	Idukki	Scenic Lookout	Performing Arts Venue	Roof Deck	Field	Garden	Zoo	Lake	Beach
3	Kannur	Beach	Historic Site	Zoo	Lighthouse	Campground	Field	Garden	Lake
4	Kasaragod	Historic Site	Zoo	Lighthouse	Beach	Campground	Field	Garden	Lake

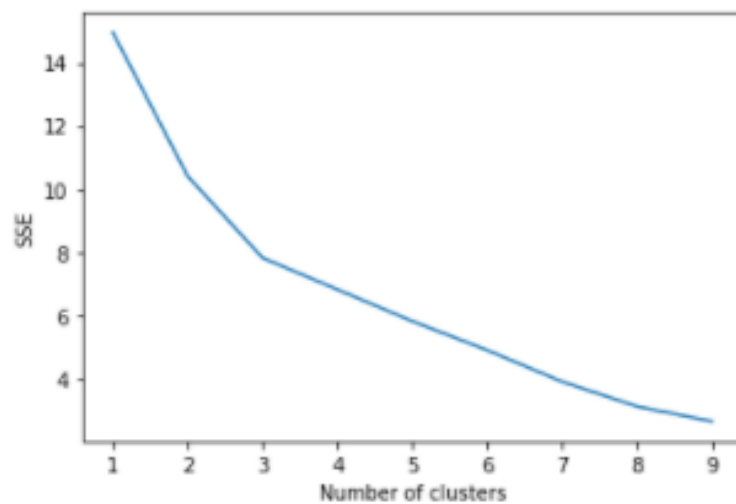
## The most common Business category in a particular district

32]:

	Venue	1st Most Common Business	2nd Most Common Business	3rd Most Common Business	4th Most Common Business	5th Most Common Business	6th Most Common Business	7th Most Common Business	8th Most Common Business	9th Most Common Business	10th Most Common Business
0	Alappuzha	Indian Restaurant	Restaurant	Vegetarian / Vegan Restaurant	Clothing Store	Eastern European Restaurant	Diner	Dessert Shop	Department Store	Convenience Store	Coffee Shop
1	Alappuzha Beach	Indian Restaurant	Vegetarian / Vegan Restaurant	Fried Chicken Joint	Eastern European Restaurant	Diner	Dessert Shop	Department Store	Convenience Store	Coffee Shop	Clothing Store
2	Alappuzha Light House	Indian Restaurant	Seafood Restaurant	Café	Vegetarian / Vegan Restaurant	Coffee Shop	Eastern European Restaurant	Diner	Dessert Shop	Department Store	Convenience Store
3	Baby Beach	Indian Restaurant	Vegetarian / Vegan Restaurant	Fried Chicken Joint	Eastern European Restaurant	Diner	Dessert Shop	Department Store	Convenience Store	Coffee Shop	Clothing Store
4	Bekal Fort	Hotel	Vegetarian / Vegan Restaurant	Fried Chicken Joint	Eastern European Restaurant	Diner	Dessert Shop	Department Store	Convenience Store	Coffee Shop	Clothing Store

## K-means clustering algorithm

Here elbow method is used to determine the optimum value of k to perform K-means clustering. The graph obtained is below.



## Result

### Each venue along with the top 5 most common business

#### ----Alappuzha----

	venue	freq
0	Indian Restaurant	0.75
1	Restaurant	0.25
2	Asian Restaurant	0.00
3	Mexican Restaurant	0.00
4	Italian Restaurant	0.00

#### ----Alappuzha Beach----

	venue	freq
0	Indian Restaurant	1.0
1	Asian Restaurant	0.0
2	Bakery	0.0
3	Italian Restaurant	0.0
4	Jewelry Store	0.0

#### ----Alappuzha Light House----

	venue	freq
0	Indian Restaurant	0.6
1	Seafood Restaurant	0.2
2	Café	0.2
3	Music Store	0.0
4	Italian Restaurant	0.0

#### ----Baby Beach----

	venue	freq
0	Indian Restaurant	1.0
1	Asian Restaurant	0.0
2	Bakery	0.0
3	Italian Restaurant	0.0
4	Jewelry Store	0.0

#### ----Bekal Fort----

	venue	freq
0	Hotel	1.0
1	Bakery	0.0
2	Indian Restaurant	0.0
3	Italian Restaurant	0.0
4	Jewelry Store	0.0

#### ----Black Beach----

	venue	freq
0	Indian Restaurant	0.50
1	Café	0.33
2	Restaurant	0.17
3	Music Store	0.00
4	Italian Restaurant	0.00

#### ----Calicut Beach----

	venue	freq
0	Restaurant	0.3
1	Vegetarian / Vegan Restaurant	0.2
2	Burger Joint	0.2
3	Indian Restaurant	0.2
4	Snack Place	0.1

```

----Calicut Beach----
      venue  freq
0      Restaurant  0.3
1  Vegetarian / Vegan Restaurant  0.2
2      Burger Joint  0.2
3      Indian Restaurant  0.2
4      Snack Place  0.1

```

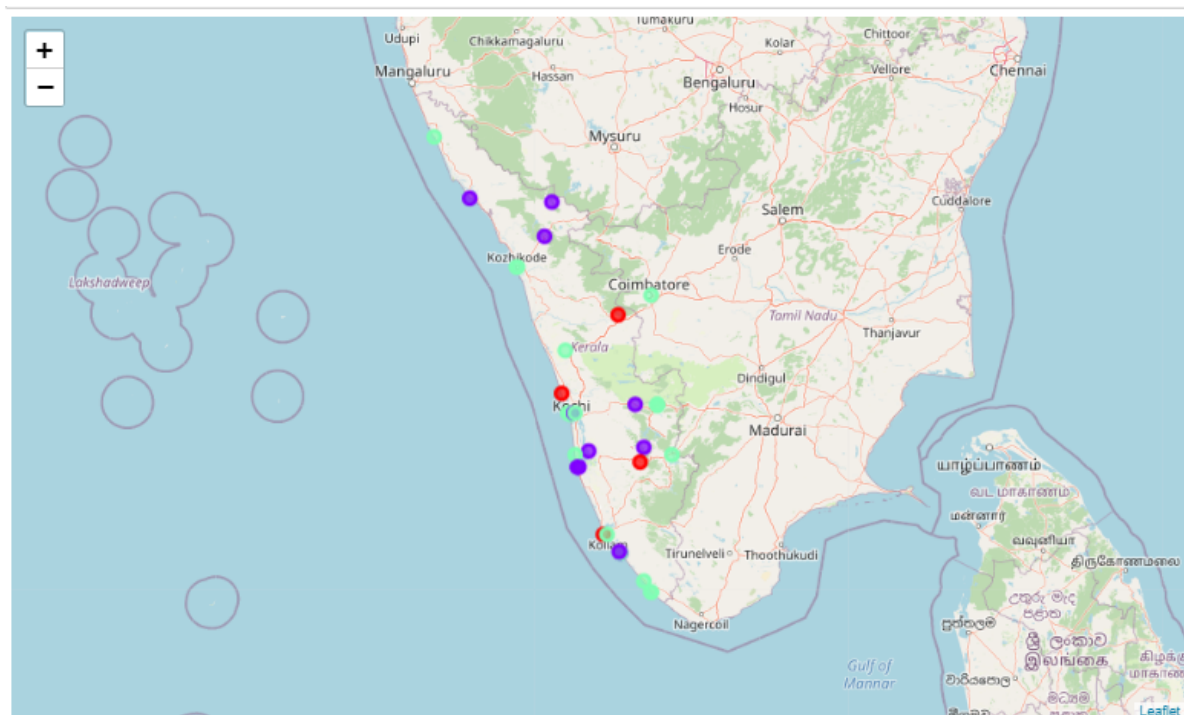
```

----Cherai Beach----
      venue  freq
0 Eastern European Restaurant  1.0
1      Asian Restaurant  0.0
2      Hotel  0.0
3      Italian Restaurant  0.0
4      Jewelry Store  0.0

```

### Representation of clusters in Map with different colours

The colors purple, green, and red represents three clusters respectively





## Discussion

The results show that the most common business in cluster one at the respective venues are Indian Restaurants. So Indian Restaurants are popular in these tourist venues and opening up a similar one can attract many tourists.

### Cluster 1

```
: nearby_business_merged.loc[nearby_business_merged['Cluster Labels'] == 0, nearby_business_merged.columns[[0] + list(range(4,
```

	Venue	1st Most Common Business	2nd Most Common Business	3rd Most Common Business	4th Most Common Business	5th Most Common Business	6th Most Common Business	7th Most Common Business	8th Most Common Business	9th Most Common Business	10th Most Common Business
1	Alappuzha	Indian Restaurant	Restaurant	Vegetarian / Vegan Restaurant	Clothing Store	Eastern European Restaurant	Diner	Dessert Shop	Department Store	Convenience Store	Coffee Shop
7	Alappuzha Light House	Indian Restaurant	Seafood Restaurant	Café	Vegetarian / Vegan Restaurant	Coffee Shop	Eastern European Restaurant	Diner	Dessert Shop	Department Store	Convenience Store
15	Kumarakom Backwaters	Boat or Ferry	Vegetarian / Vegan Restaurant	Coffee Shop	Fast Food Restaurant	Eastern European Restaurant	Diner	Dessert Shop	Department Store	Convenience Store	Clothing Store
24	Alappuzha Beach	Indian Restaurant	Vegetarian / Vegan Restaurant	Fried Chicken Joint	Eastern European Restaurant	Diner	Dessert Shop	Department Store	Convenience Store	Coffee Shop	Clothing Store
66	Durbar Hall Art Gallery	Indian Restaurant	Hotel	Gastropub	Bakery	Fried Chicken Joint	Seafood Restaurant	Restaurant	Breakfast Spot	Café	Coffee Shop
104	Vagamon Meadows	Indian Restaurant	Vegetarian / Vegan Restaurant	Fried Chicken Joint	Eastern European Restaurant	Diner	Dessert Shop	Department Store	Convenience Store	Coffee Shop	Clothing Store
131	Cheyaparra Falls, Munnar	Indian Restaurant	Café	Vegetarian / Vegan Restaurant	Coffee Shop	Eastern European Restaurant	Diner	Dessert Shop	Department Store	Convenience Store	Clothing Store
153	Baby Beach	Indian Restaurant	Vegetarian / Vegan Restaurant	Fried Chicken Joint	Eastern European Restaurant	Diner	Dessert Shop	Department Store	Convenience Store	Coffee Shop	Clothing Store
169	Varkala Beach	Indian Restaurant	Hotel	Seafood Restaurant	Café	Juice Bar	Vegetarian / Vegan Restaurant	Coffee Shop	Diner	Dessert Shop	Department Store
204	Black Beach	Indian Restaurant	Café	Restaurant	Vegetarian / Vegan Restaurant	Coffee Shop	Eastern European Restaurant	Diner	Dessert Shop	Department Store	Convenience Store
394	Lakkidi View Point	Indian Restaurant	Café	Vegetarian / Vegan Restaurant	Coffee Shop	Eastern European Restaurant	Diner	Dessert Shop	Department Store	Convenience Store	Clothing Store

The results show that the most common business in cluster two :

#### Cluster 2

```
: nearby_business_merged.loc[nearby_business_merged['Cluster Labels'] == 1, nearby_business_merged.columns[[0] + list(range(4, 11))]]
```

	Venue	1st Most Common Business	2nd Most Common Business	3rd Most Common Business	4th Most Common Business	5th Most Common Business	6th Most Common Business	7th Most Common Business	8th Most Common Business	9th Most Common Business	10th Most Common Business
20	Marari Beach	Hotel	Vegetarian / Vegan Restaurant	Fried Chicken Joint	Eastern European Restaurant	Diner	Dessert Shop	Department Store	Convenience Store	Coffee Shop	Clothing Store
28	Springr	Snack Place	Indian Restaurant	Market	Vegetarian / Vegan Restaurant	Clothing Store	Diner	Dessert Shop	Department Store	Convenience Store	Coffee Shop
34	David Hall	Hotel	Asian Restaurant	Café	Seafood Restaurant	Coffee Shop	Indian Restaurant	Italian Restaurant	Cajun / Creole Restaurant	Mediterranean Restaurant	Restaurant
59	Pepper House	Hotel	Seafood Restaurant	Vegetarian / Vegan Restaurant	Clothing Store	Eastern European Restaurant	Diner	Dessert Shop	Department Store	Convenience Store	Coffee Shop
92	Rajiv Gandhi indoor stadium	Indian Restaurant	Asian Restaurant	Hotel	Sandwich Place	Burger Joint	Fast Food Restaurant	Coffee Shop	Diner	Dessert Shop	Department Store
107	The Panoramic Getaway Hotel	Hotel	Vegetarian / Vegan Restaurant	Fried Chicken Joint	Eastern European Restaurant	Diner	Dessert Shop	Department Store	Convenience Store	Coffee Shop	Clothing Store
110	Kadathanadan Kaları Centre, Thekkady	Hotel	Indian Restaurant	Market	Vegetarian / Vegan Restaurant	Coffee Shop	Diner	Dessert Shop	Department Store	Convenience Store	Clothing Store
129	Tea Gardens, Munnar	Hotel	Vegetarian / Vegan Restaurant	Fried Chicken Joint	Eastern European Restaurant	Diner	Dessert Shop	Department Store	Convenience Store	Coffee Shop	Clothing Store
137	Punarjani Traditional Village	Hotel	Vegetarian / Vegan Restaurant	Fried Chicken Joint	Eastern European Restaurant	Diner	Dessert Shop	Department Store	Convenience Store	Coffee Shop	Clothing Store
158	Bekal Fort	Hotel	Vegetarian / Vegan Restaurant	Fried Chicken Joint	Eastern European Restaurant	Diner	Dessert Shop	Department Store	Convenience Store	Coffee Shop	Clothing Store
164	Kollam-Asramam ground	Asian Restaurant	Bakery	Juice Bar	Coffee Shop	Eastern European Restaurant	Diner	Dessert Shop	Department Store	Convenience Store	Clothing Store
189	South Cliff Beach	Restaurant	Hotel	Seafood Restaurant	Convenience Store	Pizza Place	Music Store	Juice Bar	Vegetarian / Vegan Restaurant	Cajun / Creole Restaurant	Dessert Shop

The results show that the most common business in cluster three :

#### Cluster 3

```
: nearby_business_merged.loc[nearby_business_merged['Cluster Labels'] == 2, nearby_business_merged.columns[[0] + list(range(4, 11))]]
```

	Venue	1st Most Common Business	2nd Most Common Business	3rd Most Common Business	4th Most Common Business	5th Most Common Business	6th Most Common Business	7th Most Common Business	8th Most Common Business	9th Most Common Business	10th Most Common Business
141	Mundakkayam	Asian Restaurant	Gift Shop	Bakery	Bistro	Boat or Ferry	Breakfast Spot	Burger Joint	Café	Cajun / Creole Restaurant	Fried Chicken Joint
160	Thirumullavaram Beach	Seafood Restaurant	Vegetarian / Vegan Restaurant	Clothing Store	Eastern European Restaurant	Diner	Dessert Shop	Department Store	Convenience Store	Coffee Shop	Cajun / Creole Restaurant
291	Malampuzha Dam and Garden	Tourist Information Center	Vegetarian / Vegan Restaurant	Fast Food Restaurant	Eastern European Restaurant	Diner	Dessert Shop	Department Store	Convenience Store	Coffee Shop	Clothing Store
343	Cheral Beach	Eastern European Restaurant	Vegetarian / Vegan Restaurant	Tourist Information Center	Fast Food Restaurant	Diner	Dessert Shop	Department Store	Convenience Store	Coffee Shop	Clothing Store

## **Conclusion**

In this project, we have analyzed the tourist venues for each Kerala District using of the Foursquare API. By K-means clustering algorithm, we have clustered the tourist spots based on exploring the frequency of the businesses that are present which could help us indicate a business opportunity that could be established in the locality so that the business could attract as many tourists as possible.