Analysis of the Classification of Indian Hotel under World-Class Segment

By- Jeeten Patel

1. Introduction

1.1. Background

India (Hindi: Bhārat), officially the Republic of India, is a country in South Asia. It is the seventh-largest country by area, the second-most populous country, and the most populous democracy in the world. Bounded by the Indian Ocean on the south, the Arabian Sea on the southwest, and the Bay of Bengal on the southeast.

Ministry of Tourism has a voluntary scheme for Classifying/ Re-classifying Hotels under various categories namely Heritage Grand, Heritage Classic, Heritage Basic, 5 Star Deluxe, 5 Star and 4 Star. The aims and objectives of this scheme are to provide world-class standard services to tourists.

1.2. Problem

This project elaborated on a Machine Learning concepts that attempt to analyze the segmentation of world-class standard services hotel & restaurants located across India and tries to understand what is popular around them and what they have to offer to tourist who is contemplating to choose stay. This analysis is not only limited to tourist but also investors seeking business opportunities can take advantage as it provides an analysis of the surrounding within a 1km radius of the Hotel.

1.3. Interest

The deciding factor would be on how lively, supportive, vibrant and unique each of the locations around the hotel would be when compared to each other and considering that Hotel service & features. The business problem in this study assumes that it will assist people who are interested in creating a projection of potential life and activities in these hotel neighborhoods or if anyone interested in a potential location for a business opportunity.

2. Data acquisition and cleaning

2.1. Data source

Data source- https://data.gov.in/resources/approved-hotelsrestaurantsair-catering-unitstime-share-resortsapartmentsconvention (CSV format). These Hotel types are specifically classified in categories: Heritage Grand, Heritage Classic, Heritage Basic, 5 Star Deluxe, 5 Star and 4 Star. The above data has been published by the Ministry of Tourism for Classifying/ Re-classifying Hotels under various categories at the project level.

2.2. Data cleaning

The irrelevant data columns: Phone, Fax, Email ID and Website have been removed. The Nominatim library from geocoders.geopy package has been used to find the Longitude and Latitude of Hotel Locations.

For avoiding getting results outside India and solving duplicate Hotel name problems. A new column Dummy-1 has been created combining Hotel Name & State columns, which ultimately taken as input for geocoders.geopy package for producing Latitude and Longitude column. The purpose of creating this Dummy-1 column is to encounter two potential problems with the Hotel Name which arises during analysis. First, there are three 'Country Inn & Suites By Carlson' located differently in India and secondly located abroad also few of them are Country Inn & Suites By Carlson, Prineville, USA and Country Inn & Suites by Radisson, Portland Delta Park, USA. Hence, we solved the problem using the Dummy-1 column. After fixing these problems, two outliers are found in the data which are fixed manually. These arises as in original dataset State name are missing.

	Hotel Name	Address	State	Type	Rooms	Latitude	Longitude	Dummy-1
0	The Oberoi Cecil	"Ambedkar ChowkChaura Maidan, Shimla - 171004,...	HIMACHAL PRADESH	Heritage Grand	75	31.103160	77.154990	The Oberoi Cecil- Heritage Grand
1	The Lalit Grand Palace	"Gupkar Road, Srinagar, JAMMU AND KASHMIR"	JAMMU AND KASHMIR	Heritage Grand	112	34.079870	74.862381	The Lalit Grand Palace- Heritage Grand
2	Hotel Fateh Prakash Palace	"The City Palace Complex, Udaipur - 313001, RA...	RAJASTHAN	Heritage Grand	30	24.575035	73.683426	Hotel Fateh Prakash Palace- Heritage Grand
3	Shiv Niwas Palace	"The City Palace ComplexUdaipur, Udaipur, RAJA...	RAJASTHAN	Heritage Grand	36	24.574100	73.683758	Shiv Niwas Palace- Heritage Grand
4	Savoy Hotel	"77, Sylks Road, Ootacamund, TAMIL NADU"	TAMIL NADU	Heritage Grand	40	11.414861	76.692884	Savoy Hotel - Heritage Grand

2.3. Feature selection

After data cleaning, there are 306 samples and 7 features in the given data frame along with one dummy column named Dummy-1 for analysis purposes.

The Foursquare API to find the top 20 venues in the neighborhoods of these Hotels. This will help understand the nature of life around these locations. For illustration purposes analyze the popular venues located within a 1km radius is considered.

```
hm.info()
```

```
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 306 entries, 0 to 305
Data columns (total 8 columns):
Hotel Name    306 non-null object
Address       306 non-null object
State         306 non-null object
Type          306 non-null object
Rooms         306 non-null int64
Latitude      306 non-null float64
Longitude     306 non-null float64
Dummy-1       306 non-null object
dtypes: float64(2), int64(1), object(5)
memory usage: 19.2+ KB
```

```
hm.shape
```

```
(306, 8)
```

For this project, analysis is carried using serval libraries:

```
import requests # library to handle requests
import pandas as pd # library for data analsysis
import numpy as np # library to handle data in a vectorized manner
import random # library for random number generation
import seaborn as sns # plotting library
import matplotlib.pyplot as plt # plotting library
import folium # plotting library
%matplotlib inline
```

3. Exploratory Data Analysis

3.1. Distribution of Hotels

3.1.1. Classification Ranking

Heritage: This category will cover hotel in Residences/Havelies/Hunting Lodges/Castles/Forts/ Palaces built prior to 1950. The hotel should have a minimum of 5 rooms (10 beds).

Heritage Classic: This category will cover hotels in Residences/Havelies/Hunting Lodges/Castles/Forts/ Palaces built Prior to 1935. The hotel should have a minimum of 15 rooms (30 beds).

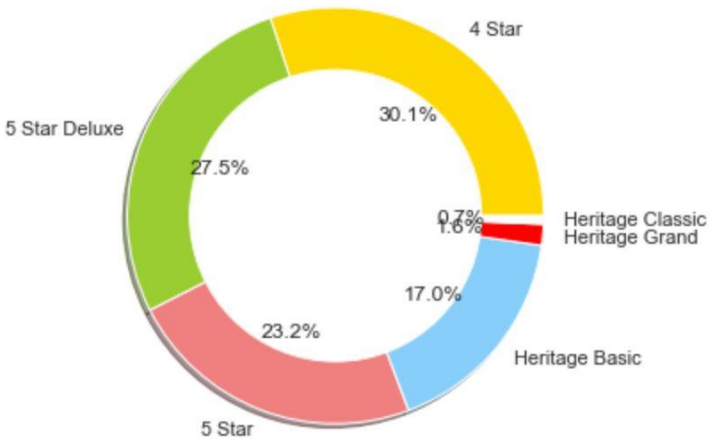
Heritage Grand: This category will cover hotels in Residence/Havelies/Hunting Lodges/Castles/Forts/ Palaces built prior to 1935. The hotel should have minimum of 15 rooms (30 beds).

5-Star Deluxe, 5-Star and 4-Star: These categories are properties that offer their guests the highest levels of luxury through personalized services, a vast range of amenities, and sophisticated accommodations. In general hotel prices and luxury increasing as we move from 1-Star to 5-Star Deluxe Hotels. Moreover, there is no universal standard for hotel ratings each country follows its Classification.

Few links below to read further:
http://tourism.gov.in/sites/default/files/051820120222241_0.pdf
http://tourism.gov.in/hotels-restaurants.com

3.1.2. Pie-Chart Distribution:

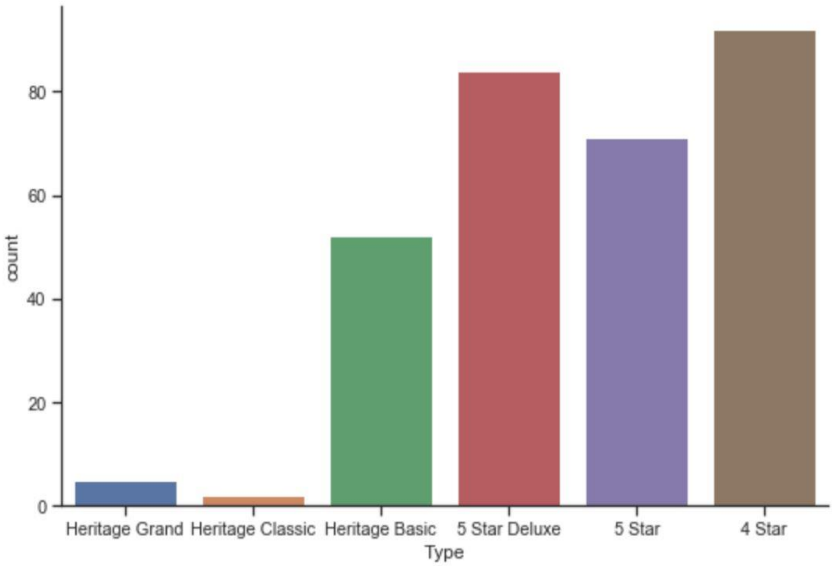
The number of 4-Star Hotels are highest in India out of these five segments. While 5-Star Deluxe are on 2nd highest. There are very low number of Heritage Classic and Heritage Grand 0.7% and 1.6% respectively while comparing to other segments.



3.1.3. **Bar-Chart Distribution:**

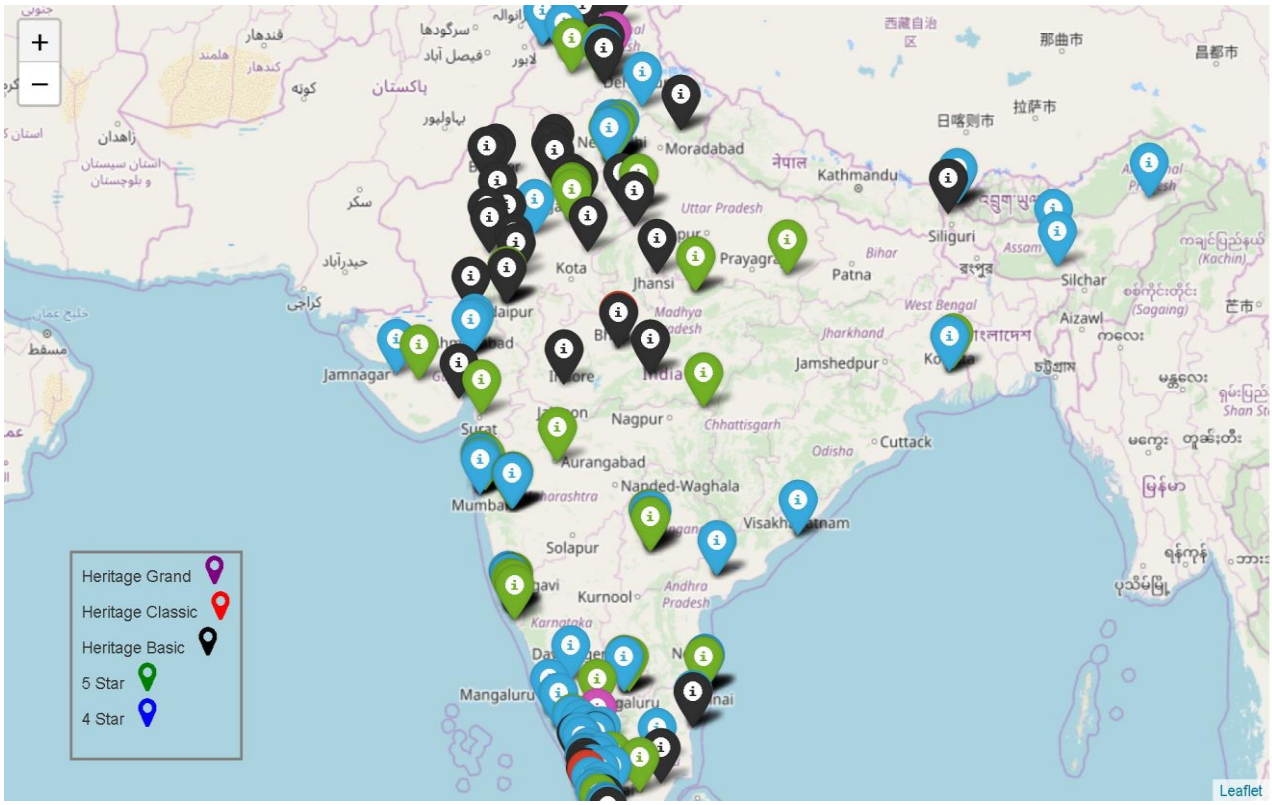
x-axis: Type of Hotel

y-axis: Number of Hotels



3.1.4. **<u>Folium Map Distribution:</u>**

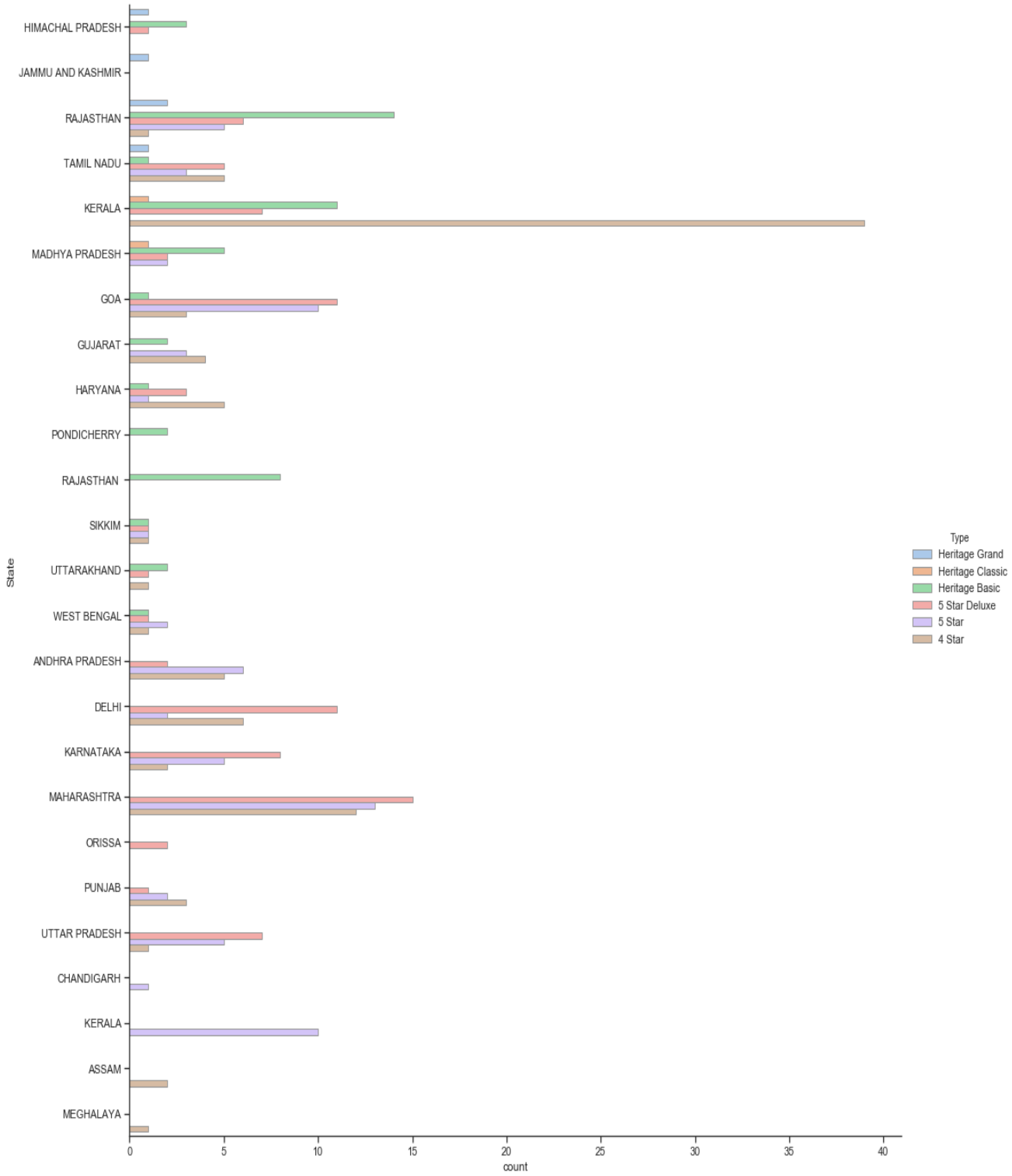
Python Folium library is used to visualize these 306 Hotels locations using Latitude and longitude values to get the visual as below:



We find that South India is heavily concentrated with premier Hotels. On the other hand, as we move from west to east India number significantly reduces due to several geographic and financial reasons which are not part of this analysis.

3.1.5. **State-wise Distribution:**

In Kerala, the number of 4 Star Hotels are far exceeded while comparing to other states. As per Maharashtra, the top three premier segments combined are greater than any other state, the majority of them are in Mumbai which is the financial, commercial and entertainment capital of India



3.2. **Relationship between Rooms and Hotel Type**

3.2.1. **Mean**

The figure below shows the means of Rooms along with Type of Hotels. We find 5-star Deluxe Mean to be highest. Followed by 5 Star as 2nd highest.

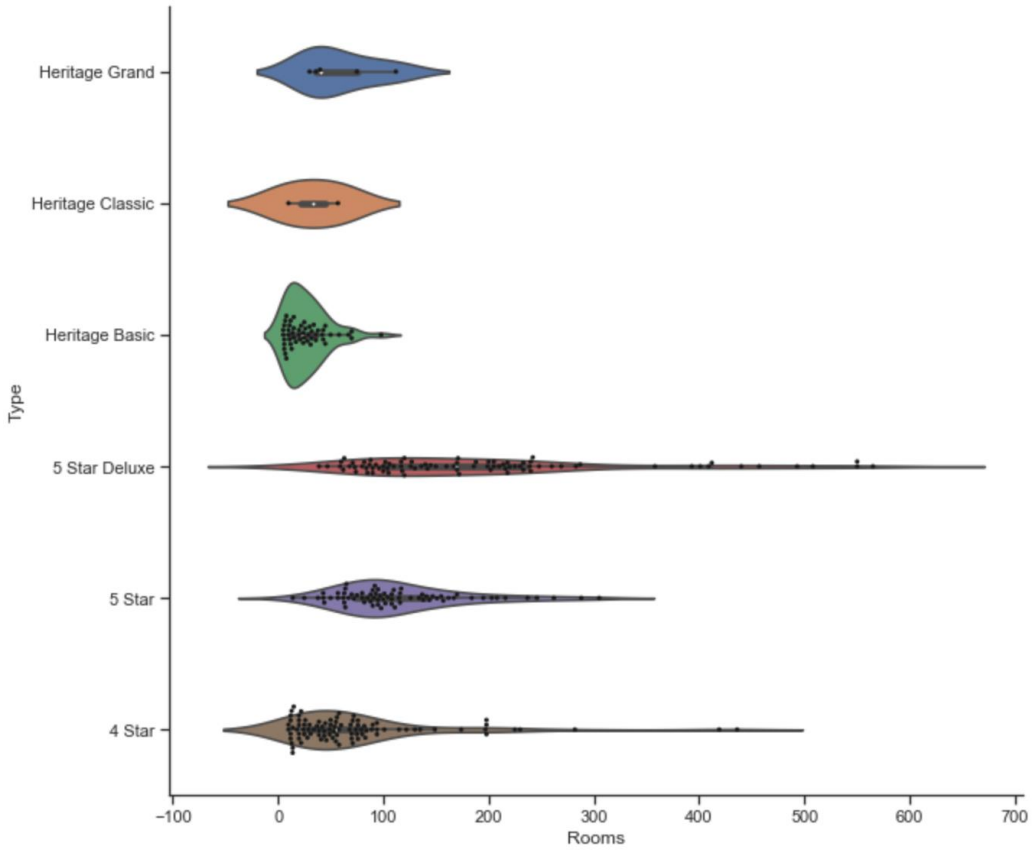
Mean

```
hm.groupby(['Type'])['Rooms'].mean()
```

```
1]: Type
    4 Star               74.347826
    5 Star              114.985915
    5 Star Deluxe       194.785714
    Heritage Basic       26.596154
    Heritage Classic     33.500000
    Heritage Grand       58.600000
    Name: Rooms, dtype: float64
```

3.2.2. **Violin Plot**

Violin Plot is plotted between Rooms and Type column to understand the distribution of Rooms across the dataset. Violin plot is a method to visualize the distribution of numerical data of different variables. It is similar Box Plot but with a rotated plot on each side, giving more information about the density estimate on the y-axis.



4. K-means Clustering: Algorithm

4.1. Analyse Each Neighbourhood of Hotel Location

The Foursquare API is used to find out the top 20 most common venues in the radius of 1000 meters for each Hotel location. In summary 272 unique categories were returned by Foursquare API. The table below shows the list of top 20 venue category for each hotel location:

Neighborhood	Type	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	...	11th Most Common Venue	12th Most Common Venue	13th Most Common Venue	14th Most Common Venue	Cc
ATS Residency	Hotel	Bakery	Indian Restaurant	Bus Station	Other Great Outdoors	Shopping Mall	Track	Field	Falafel Restaurant	...	Fast Food Restaurant	Women's Store	Event Space	M
Ahilya Fort	Indian Restaurant	Historic Site	River	Resort	Women's Store	Farmers Market	Electronics Store	Event Space	Falafel Restaurant	...	Duty-free Shop	Field	Fish & Chips Shop	M
Ajit Bhawan	Hotel	Indian Restaurant	Café	Pizza Place	American Restaurant	Bakery	Shopping Mall	Multiplex	Fish & Chips Shop	...	Fast Food Restaurant	Field	Women's Store	M
Aman, New Delhi	Indian Restaurant	Hotel	Café	Lounge	Tibetan Restaurant	Food & Drink Shop	Food	Fast Food Restaurant	Bar	...	Korean Restaurant	Snack Place	Indian Chinese Restaurant	C
Beach Hotel	Resort	Spa	Beach	Hotel	Food	Flower Shop	Flea Market	Fish Market	Fish & Chips Shop	...	Food & Drink Shop	Food Court	Fast Food Restaurant	Fa M

Above analysis based on frequency as shown below.

Let's print each neighborhood along with the top 20 most common venues

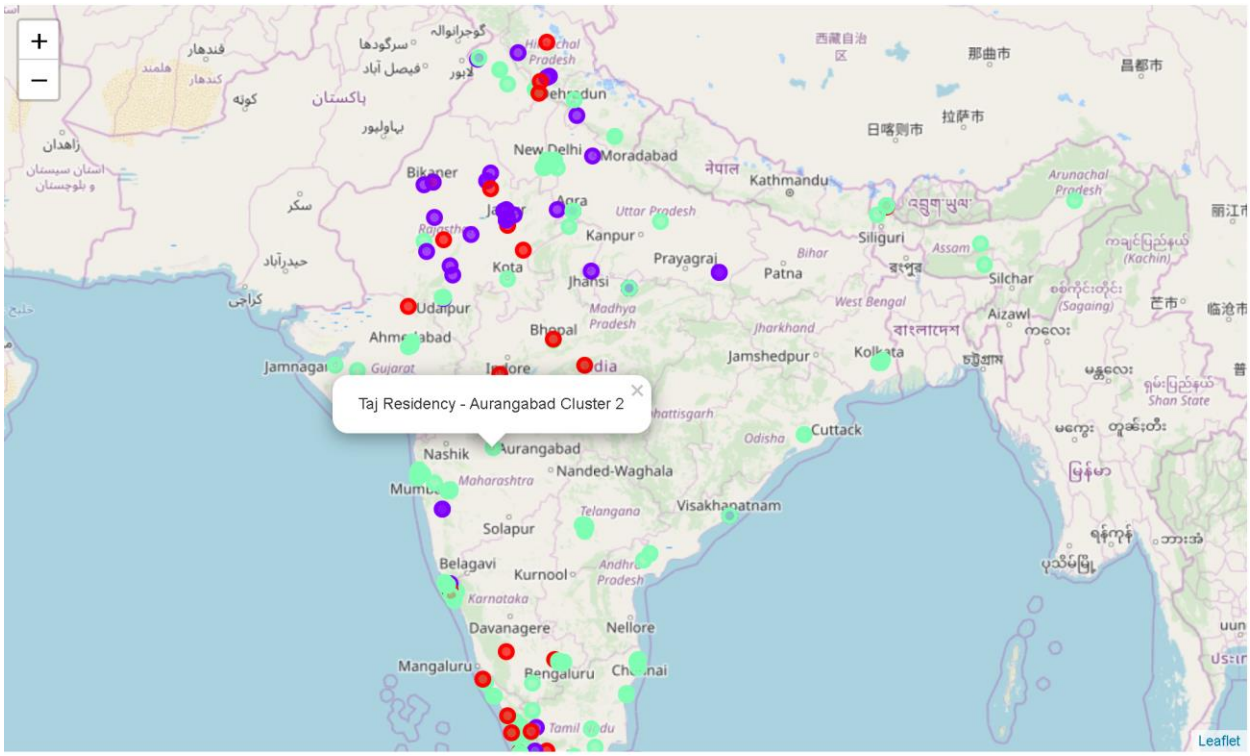
```
num_top_venues = 20

for hood in hotels_venues_grouped['Hotel Name']:
    print("----"+hood+"----")
    temp = hotels_venues_grouped[hotels_venues_grouped['Hotel Name'] == hood].T.reset_index()
    temp.columns = ['venue','freq']
    temp = temp.iloc[1:]
    temp['freq'] = temp['freq'].astype(float)
    temp = temp.round({'freq': 2})
    print(temp.sort_values('freq', ascending=False).reset_index(drop=True).head(num_top_venues))
    print('\n')
```

```
----ATS Residency----
                         venue  freq
0                        Hotel  0.25
1          Other Great Outdoors  0.12
2             Indian Restaurant  0.12
3                         Track  0.12
4                 Shopping Mall  0.12
5                   Bus Station  0.12
6                        Bakery  0.12
7                Nature Preserve  0.00
8                    Music Venue  0.00
9                   Neighborhood  0.00
10  Northeast Indian Restaurant  0.00
11                  Night Market  0.00
12                   Music Store  0.00
13                        Museum  0.00
14                    Nightclub  0.00
15                    Multiplex  0.00
16               Nightlife Spot  0.00
```

4.2. **Run k-means to cluster the neighbourhood into 3 clusters**

K-means clustering is one of the simplest and popular unsupervised machine learning algorithms. A cluster refers to a collection of data points aggregated together because of certain similarities. The ‘means’ in the K-means refers to averaging of the data; that is, finding the centroid.

For this analysis, K is set to be 3 clusters. K can be set higher based on the objective. Based on analysis these 3 clusters will provide similar supportive, the vibrant and unique ambiance around the hotel. If a tourist wants a similar experience they may choose from the same cluster and vice versa..



Cluster 1

```
_venues_merged['Cluster Labels'] == 0, hotels_venues_merged.columns[[1] + list(range(3, hotels_venues_merged.shape[1]))]].head()
```

8]:

	Hotel Name	1st Most Famous Venue	2nd Most Famous Venue	3rd Most Famous Venue	4th Most Famous Venue	5th Most Famous Venue	6th Most Famous Venue	7th Most Famous Venue	8th Most Famous Venue	9th Most Famous Venue	10th Most Famous Venue	11th Most Famous Venue	12th Most Famous Venue	1 M Fam Ve
1	Ahilya Fort	Indian Restaurant	Historic Site	River	Yoga Studio	Event Space	Falafel Restaurant	Farm	Farmers Market	Fast Food Restaurant	Field	Eastern European Restaurant	Fish & Chips Shop	Ma
4	Beach Hotel	Hotel	Spa	Beach	Farm	Farmers Market	Fast Food Restaurant	Field	Fish & Chips Shop	Yoga Studio	Event Space	Fish Market	Flea Market	Flc S
12	Chokhi Dhani	Hotel	Restaurant	Event Space	Falafel Restaurant	Farm	Farmers Market	Fast Food Restaurant	Field	Yoga Studio	Eastern European Restaurant	Fish & Chips Shop	Fish Market	Ma
14	Club Mahindra Lakeview	Tea Room	Waterfall	Yoga Studio	Fish & Chips Shop	Falafel Restaurant	Farm	Farmers Market	Fast Food Restaurant	Field	Fish Market	Electronics Store	Flea Market	Flc S
15	Cochin Seaport Hotel	Beach	Bed & Breakfast	Falafel Restaurant	Farmers Market	Fast Food Restaurant	Field	Fish & Chips Shop	Fish Market	Yoga Studio	Flea Market	Flower Shop	Food	Foc D S

Cluster 2

```
hotels_venues_merged.loc[hotels_venues_merged['Cluster Labels'] == 1, hotels_venues_merged.columns[[1] + list(range(3, hotels_v
```

	Hotel Name	1st Most Famous Venue	2nd Most Famous Venue	3rd Most Famous Venue	4th Most Famous Venue	5th Most Famous Venue	6th Most Famous Venue	7th Most Famous Venue	8th Most Famous Venue	9th Most Famous Venue	10th Most Famous Venue	11th Most Famous Venue	12th M Fam Ve
7	Bissau Palace	Coffee Shop	Indian Restaurant	Sculpture Garden	Food	Historic Site	Yoga Studio	Field	Falafel Restaurant	Farm	Farmers Market	Fast Food Restaurant	Fis Ch S
13	Chomu Palace	Hotel	Yoga Studio	Event Space	Farm	Farmers Market	Fast Food Restaurant	Field	Fish & Chips Shop	Fish Market	Flea Market	Flower Shop	F
17	Country In & Suites, Ajmer	Market	Monument / Landmark	Design Studio	Farm	Farmers Market	Fast Food Restaurant	Field	Fish & Chips Shop	Fish Market	Yoga Studio	Falafel Restaurant	Flo S
19	Country Inn & Suites by	Resort	Yoga	Fish & Chips	Falafel	Farm	Farmers	Fast Food	Field	Fish	Electronics	Flea	Flo

Cluster 3

```
hotels_venues_merged.loc[hotels_venues_merged['Cluster Labels'] == 2, hotels_venues_merged.columns[[1] + list(range(3, hotels_v
```

	Hotel Name	1st Most Famous Venue	2nd Most Famous Venue	3rd Most Famous Venue	4th Most Famous Venue	5th Most Famous Venue	6th Most Famous Venue	7th Most Famous Venue	8th Most Famous Venue	9th Most Famous Venue
0	ATS Residency	Track	Indian Restaurant	Other Great Outdoors	Bus Station	Shopping Mall	Bakery	Farm	Farmers Market	Fast Food Restaurant
2	Ajit Bhawan	Indian Restaurant	Café	Bakery	Pizza Place	American Restaurant	Multiplex	Shopping Mall	Fish & Chips Shop	Farm
3	Aman, New Delhi	Hotel	Café	Lounge	Tibetan Restaurant	Food & Drink Shop	Food	Fast Food Restaurant	Bar	Bakery
5	Bell Tower Resorts Pvt.Ltd.	Indian Restaurant	Pizza Place	Clothing Store	Yoga Studio	Fish Market	Farm	Farmers Market	Fast Food Restaurant	Field
6	Best Western Amrutha Castle	Indian Restaurant	Chinese Restaurant	Hotel	Ice Cream Shop	Coffee Shop	Restaurant	South Indian Restaurant	Lounge	Fast Food Restaurant

5. Conclusion

The relationship between Number of Hotels & Type of Hotel and Number of rooms & Type of Hotel as discussed in Exploratory Data Analysis are significant. Analysis it further can give meaning insights. As per someone with business seeking objective can provide competitive edge. Folium Map provides significant observations how number of premier Hotels diminish from West to East. K-means algorithm provides three clusters based on similarities of supportive, vibrant and unique ambience around the hotel.

6. Future directions

By setting K to be higher further analysis can be done as it depends on the objective of the user as shown below. This analysis was done with the top 20 most common venues this can be further increased. Instead of using most common venues one can use the highest rated venues or something else. One can also visualize State-wise or for the city using Folium Map. The same analysis can be done for 1-star, 2-star, 3-star hotels.

Cluster 4 (Set k=4). Remove # below.

```
#hotels_venues_merged.loc[hotels_venues_merged['Cluster Labels'] == 3, hotels_venues_merged.columns[[1] + list(range(5, hotels_
```

Cluster 5 (Set k=5). Remove # below.

```
#hotels_venues_merged.loc[hotels_venues_merged['Cluster Labels'] == 4, hotels_venues_merged.columns[[1] + list(range(3, hotels_
```