Analysis of the segmentation of Indian World-class standard services Hotel & Restaurants

A. Data Analysis

By- Jeeten Patel


Introduction

Introduction

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.

Problem

- To Analyze the segmentation of world-class standard services hotel & restaurants located across India
- 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.

Interest

- How lively and vibrant each of the locations around the hotel with its features will be when compared to other
- Will assist tourist or if anyone interested in a potential location for a business opportunity.

Data acquisition and cleaning

Data acquisition and cleaning

- Data source- https://data.gov.in/resources/approved-hotelsrestaurantsair-catering-unitstime-share-resortsapartmentsconvention (CSV format)
- The above data has been published by the Ministry of Tourism for Classifying/ Re-classifying Hotels under various categories at the project
- The Hotel are specifically classified in five premier categories: Heritage Grand, Heritage Classic, Heritage Basic, 5 Star Deluxe, 5 Star and 4 Star
- The Nominatim library from geocoders.geopy package has been used to find the Longitude and Latitude of Hotel Locations.
- The irrelevant data columns: Phone, Fax, Email ID and Website have been removed.
- After data cleaning, there are 306 samples and 7 features in the given data frame along with one dummy column named as Dummy-1 for analysis purpose.

```
hm.info()
```

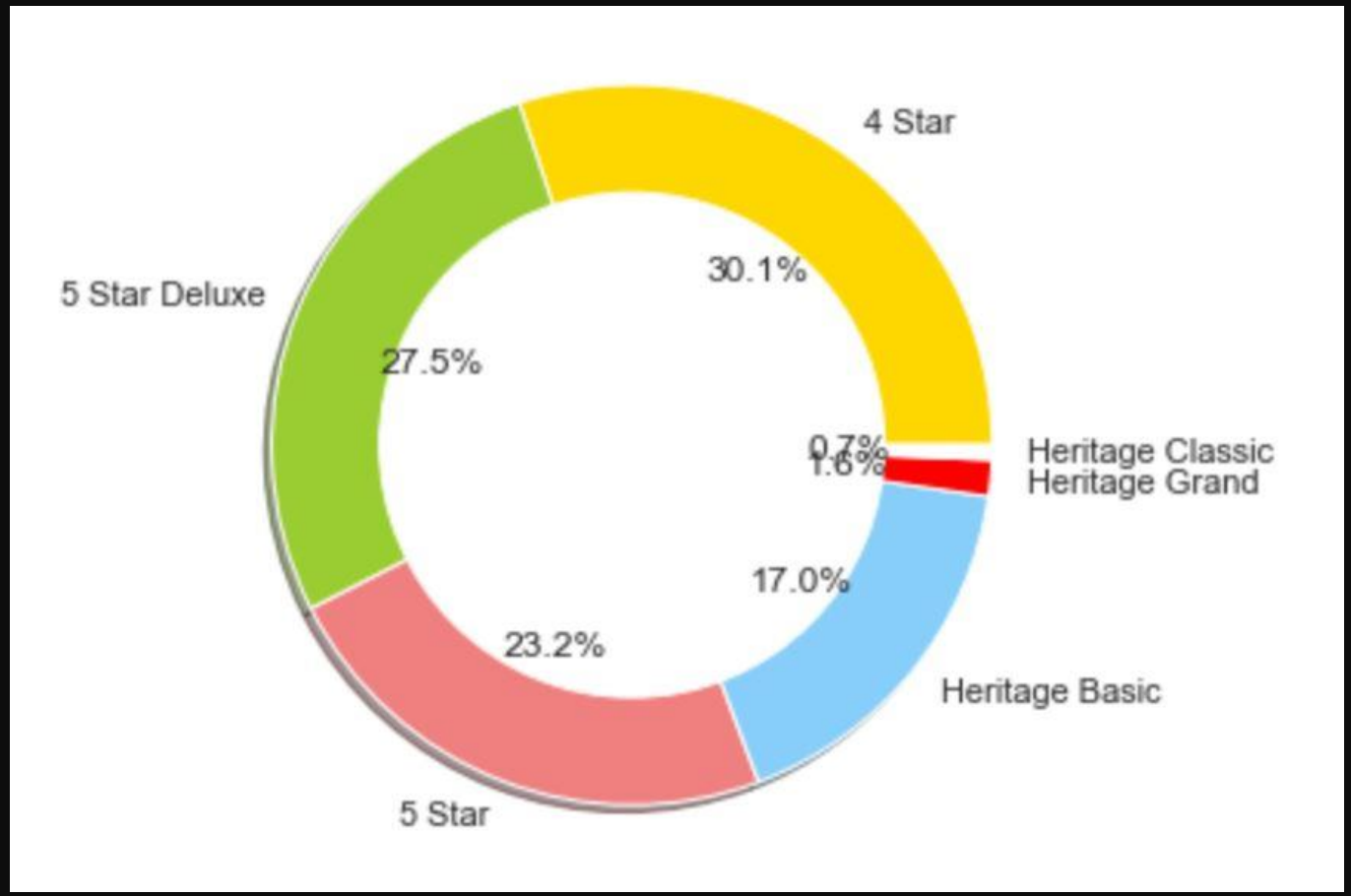
```
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 306 entries, 0 to 305
Data columns (total 8 columns):
Hotel Name    306 non-null object
Address       306 non-null object
State         306 non-null object
Type          306 non-null object
Rooms         306 non-null int64
Latitude      306 non-null 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)
```

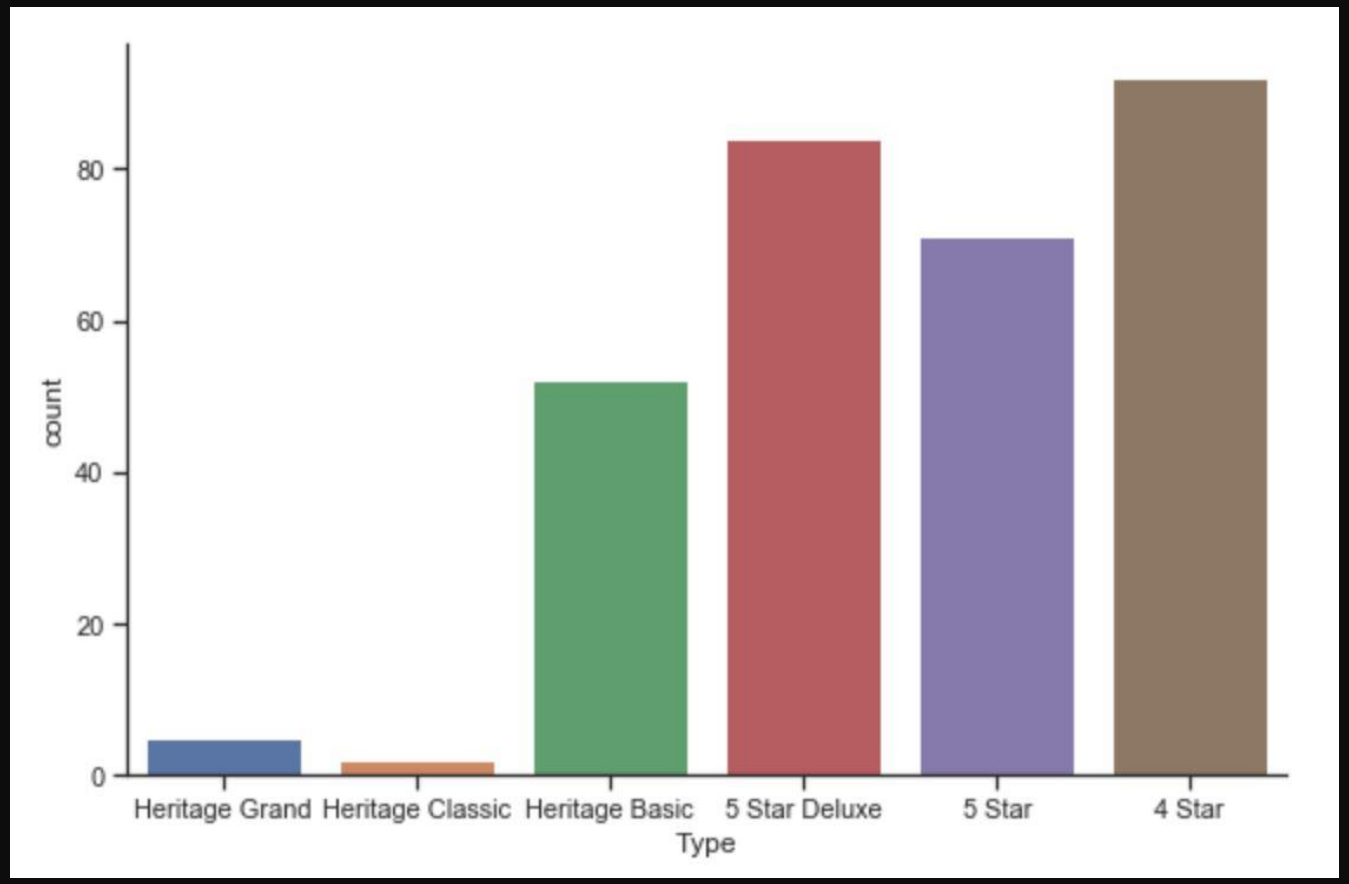

Exploratory Data Analysis

Distribution of Hotels: Pie-Chart Distribution

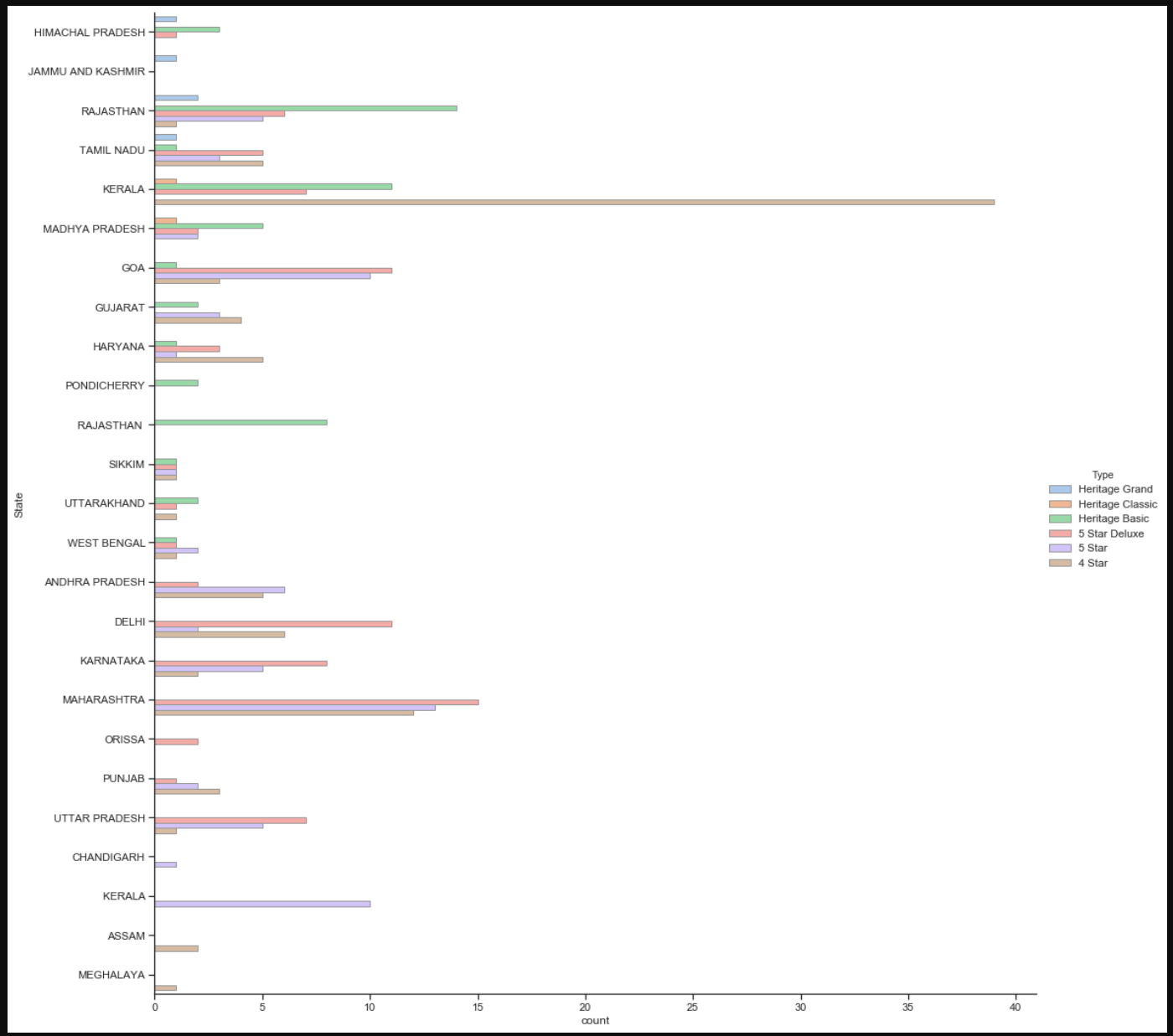


- The number of 4-Star Hotels are found to be highest in India out of these five segments. While 5-Star Deluxe are 2nd highest.
- There are very low number of top premier Hotel segment i.e. Heritage Classic and Heritage Grand 0.7% and 1.6% respectively while comparing to other segments.

Distribution of Hotels: Bar-Chart Distribution



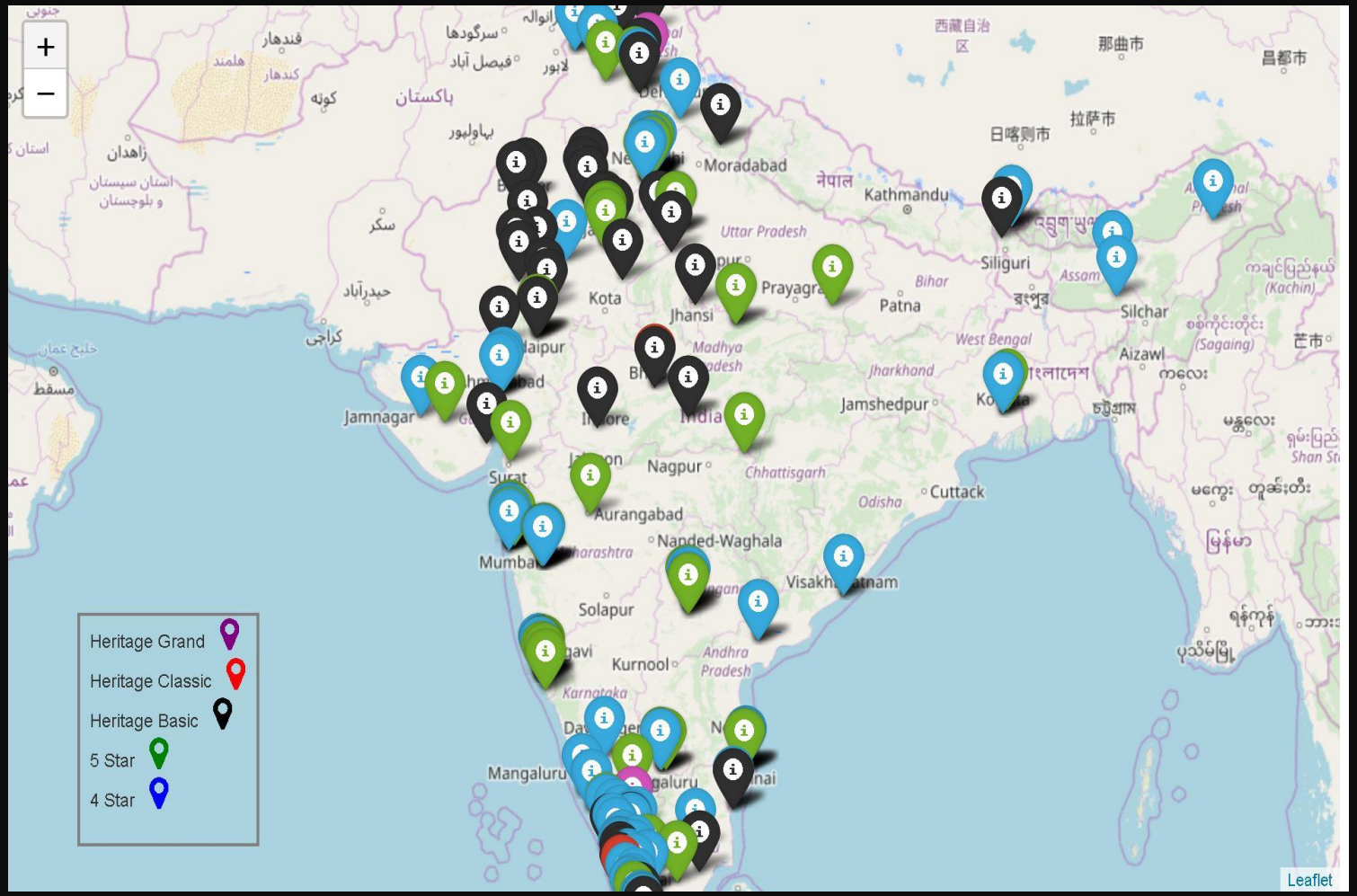
- x-axis: Hotel Type
- y-axis: Number of Hotels



Distribution of Hotels: State-wise Distribution

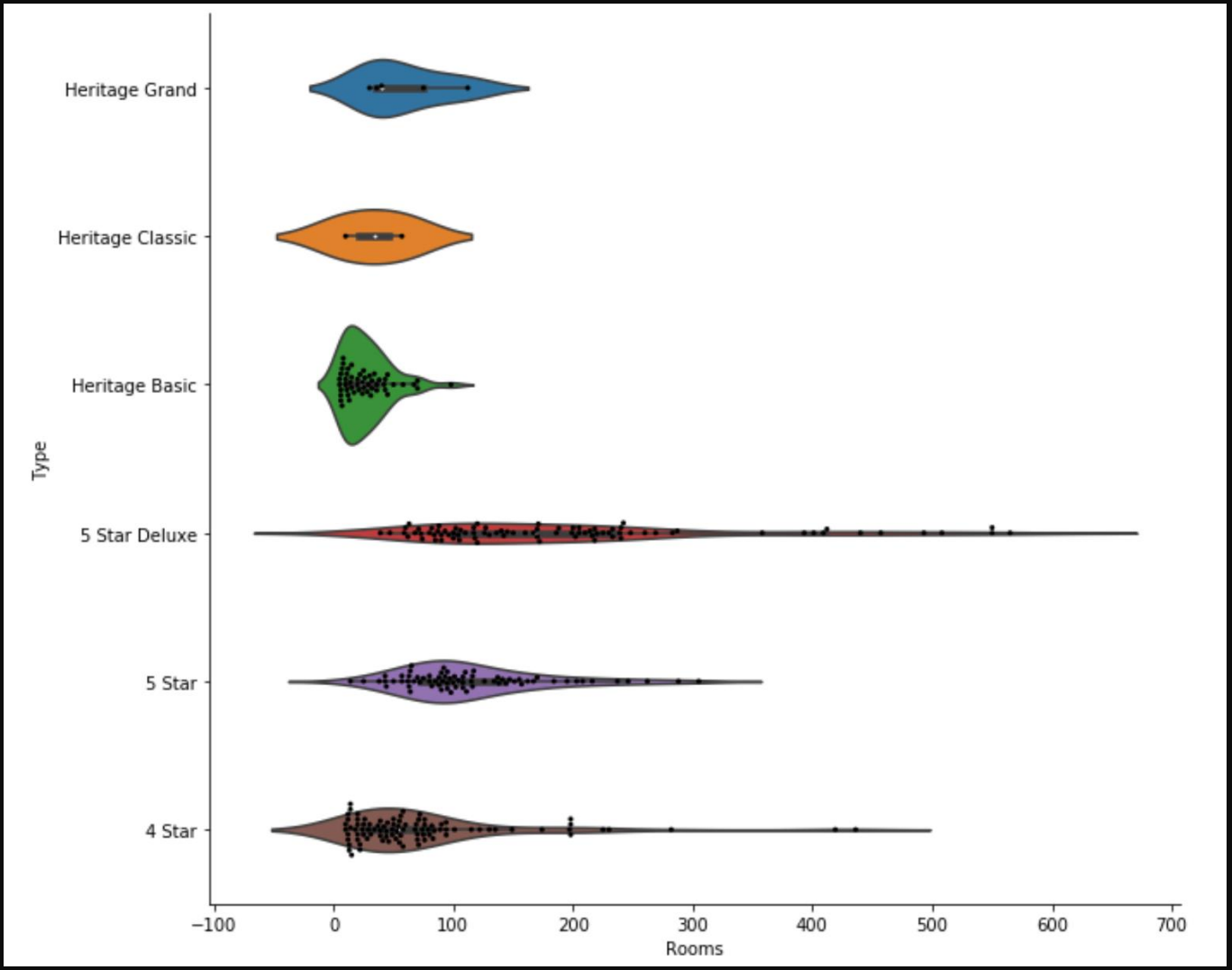
- x-axis: State
- y-axis: Number of Hotels

Distribution of Hotels: Folium Map Distribution



We find that South India is heavily concentrated with premier Hotels. On 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.

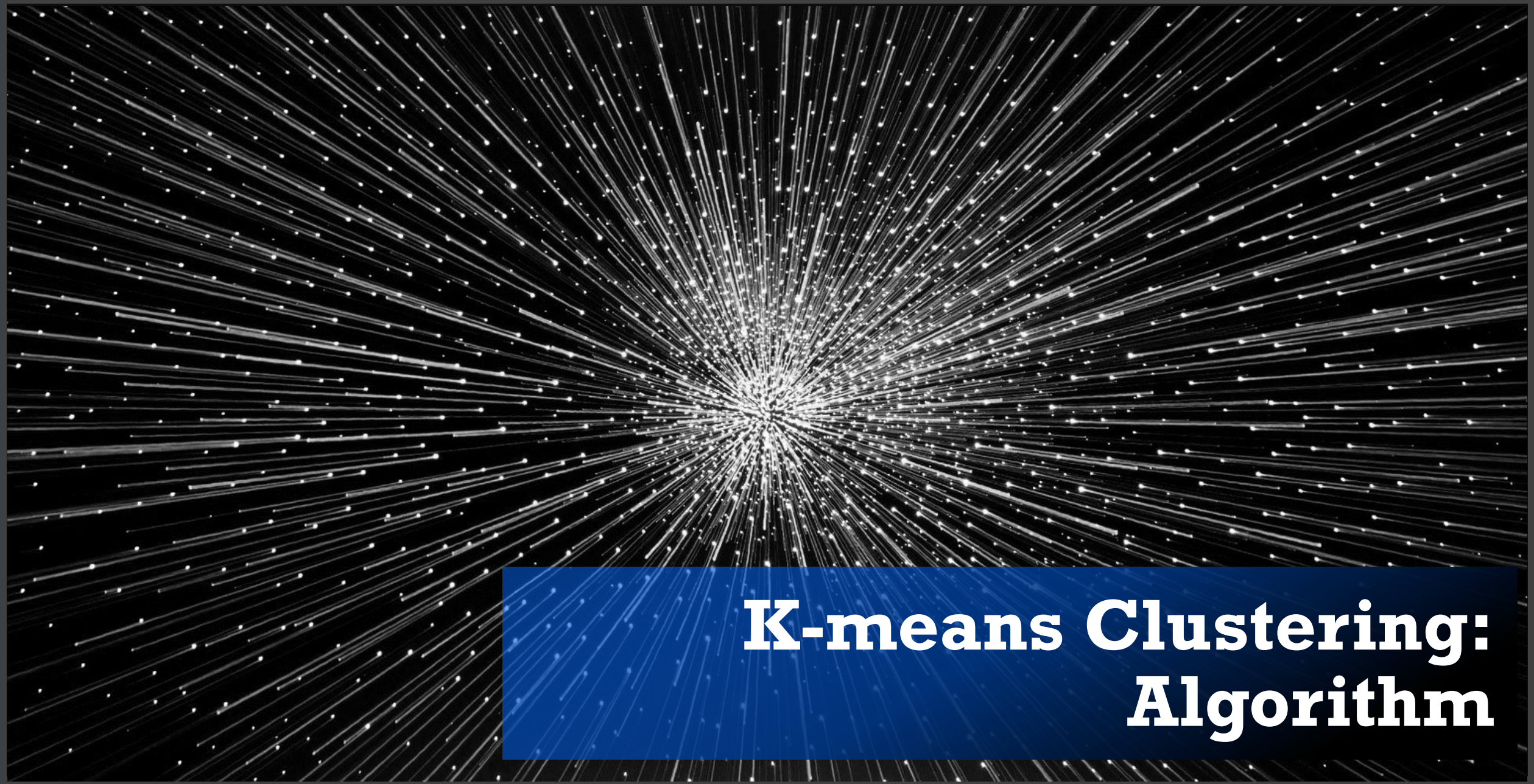
Distribution of Rooms: Box-Plot Distribution



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.

Mean:

4 Star	75
5 Star	115
5 Star Deluxe	195
Heritage Basic	27
Heritage Classic	33
Heritage Grand	59



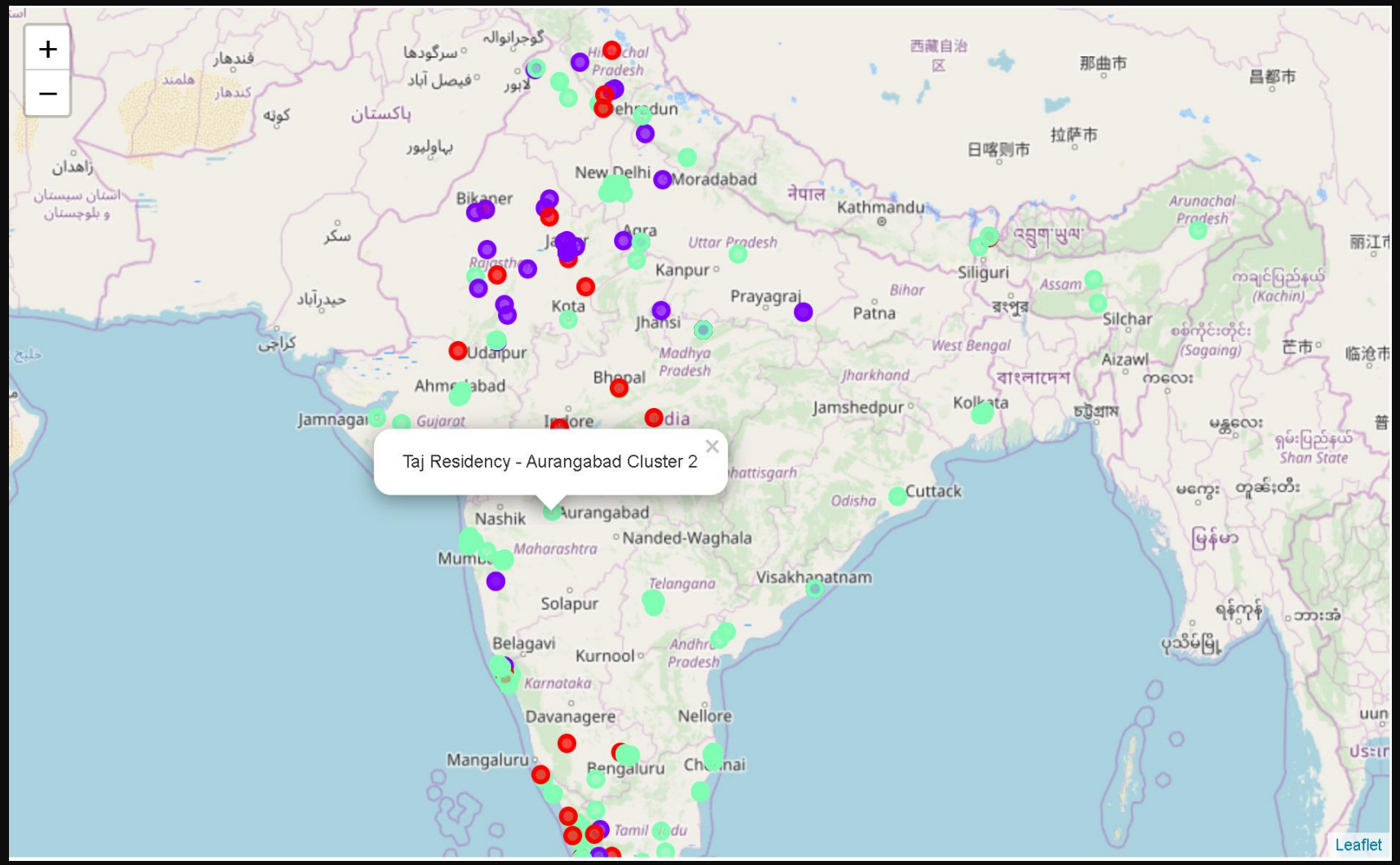
K-means Clustering: Algorithm

- Foursquare API Analysis

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

- ❑ Top 20 common venue in the radius 1000 meter for each Hotel location found using the Foursquare API
- ❑ In summary 272 unique categories were returned by Foursquare API.

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.

Run k-means to cluster the neighbourhood into 3 clusters

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

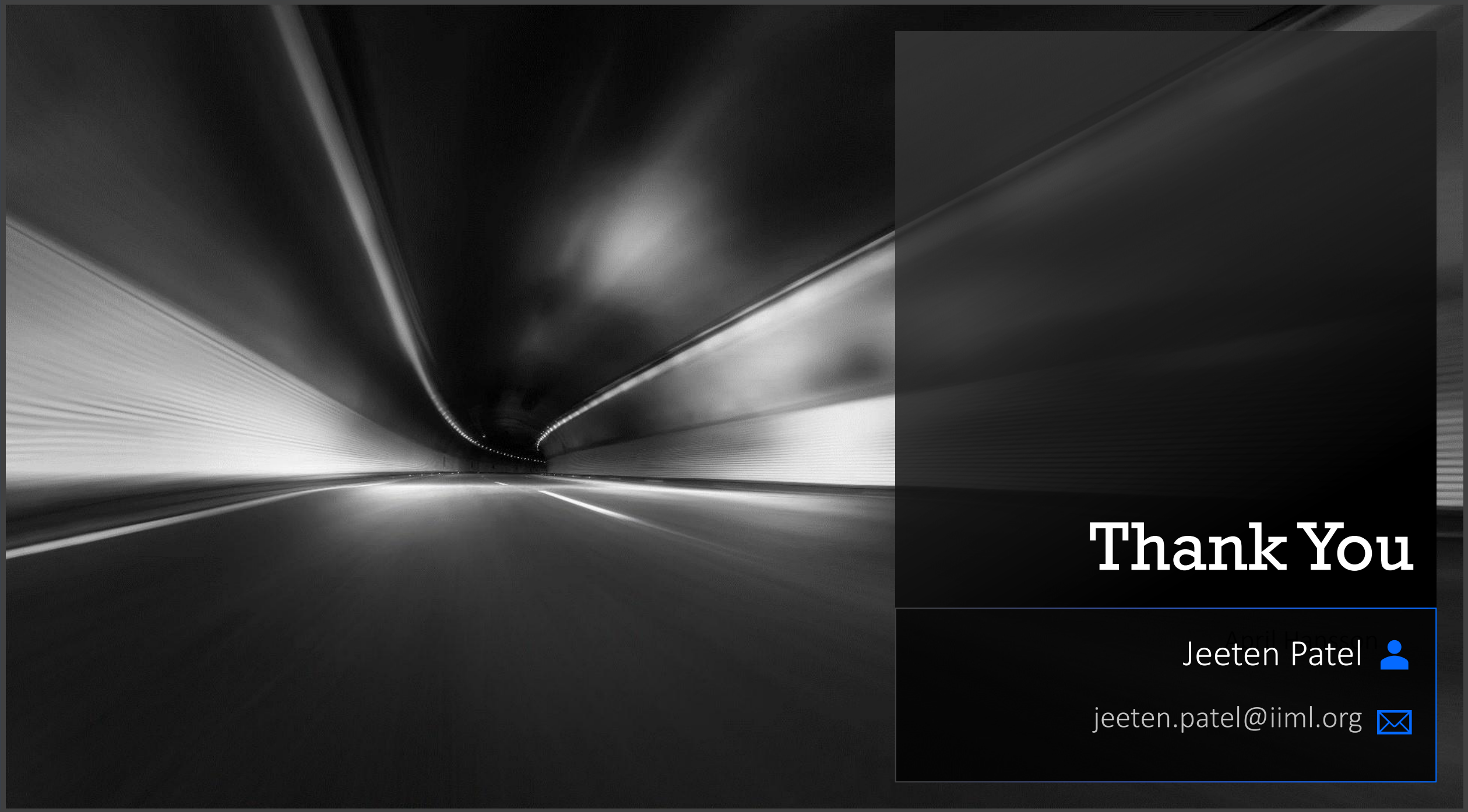
For this analysis K is set to be 3 clusters. K can be set higher based on objective. Based on analysis these 3 clusters will provide similar supportive, vibrant and unique ambience around the hotel. If a tourist want similar experience they may choose from same cluster and vice versa.

Conclusions

- 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.

Future directions

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



Thank You

Jeeten Patel

jeeten.patel@iiml.org