

# **Factors Influence Hotel Classification in India**

**reducing class imbalance on heritage hotels**

*- Suhruth Yambakam*

# Chapter 1: Introduction

## 1.1 Overview

The dramatic rise of international tourism, from the pursuit of the privileged few only half a century ago to a truly worldwide phenomenon today, has seen the sector become a bedrock of our global economy, representing 9% of the world's GDP, 30% of service exports and one in every eleven jobs.

Although travelers become increasingly adventurous, seeking new and unexplored destinations, they still demand certain indicators of what they are about to experience. Official hotel classification systems have long provided such indications/information relating to accommodation.

There is no worldwide standard for official hotel classification systems, and there may well never be one due to the incredible diversity of the environmental, socio-cultural, economic and political contexts in which they are embedded. However, there are without doubt commonalities which unite accommodation of various standards across the globe. Identifying such commonalities, as well as differences, can help destinations in establishing or revising their classification systems in such a way that they are relevant and useful to consumers, hotels, intermediaries and destinations alike.

In this regard, this report takes an in-depth look at the criteria of hotel classification between Heritage and Star categories. Besides the comparison of criteria, the report provides a general overview of the factors influencing the classification task, their benefits and challenges, and offers general guidance on areas to consider when setting up an official classification system.

## 1.2. Purpose

Irrespective of the external criteria of a government while classifying a hotel there could be some simple internal factors that will affect the hotel classification. It could as simple as to be a *hotel name* and complex as a the hotel uses any *vip contact number*. This project takes a deeper look into various internal factors that affect hotel classification in India. It ponders on the class imbalance that occurs while building a classification engine and keys places to look into. The scope of the project will only limit it to the features of an individual hotel but there could be many external and other factors that influence the classification task.

There will be perspectives that we are going to solve this classification problem. Some of the goals or milestones that we will achieve would be:

- Analyzing bias made during classification of hotels by star category.
- Influence of using VIP contact numbers by a specific hotels.
- Effect of hotel chains in classifying other hotels with less influence.
- Study various geographical aspects that make an impact on hotel category.

# Chapter 2: Literature Survey

## 2.1. Existing Problem

Hotel classification systems originated with the purpose of informing travelers what level of service expected from different hotels. Such a system of classifying hotels is used in most of the hotels across the world. Hotel classification determine both the standard of the hotel offering and the ability to purchase hotel accommodation that meets that expectation on time. This allows the consumer to recognize hotel quality easily, differentiate all levels of facilities and services which offers great transparency - vital requirements for both business and leisure tourists. Amongst other things, it is beneficial to the tourists, the hotels and tourism industry.

In india , the government, through Hotel and Restaurant Accreditation Committee (HRACC), which is a special agency of the Ministry of Tourism, carries out hotel classification. The process can also be unofficial. Generally states do not intervene in the classification process in such process. In a country like India, where an official classification system exists, the government's main objective is to have a control over the hotels.

### ***History:***

Nine years after India became independent, the Government of India established the "Hotel standards and rate structure committee" in 1956 with the objective of promoting tourism and establishing hotel classification system considering the international standards (2006). The committee also suggested suitable rate structure for the hotel industry, and choose the use of stars as a symbol for point based hotel classification system. However no major changes were made in the system over decades and subsequently many users reported about the system because of its failure to guarantee customer satisfaction.

The Ministry of tourism revised the norms again in 2013 on the basis of suggestions provided by the Hotel Industry so as to make the Hotel Industry more customers focused with increased level of service. The new guidelines also aim at providing the guest all the information about the services offered by the hotel on their website. However, these guidelines were not well accepted by the Industry. (Sharma 2013) in a report published in Economic Times summarized views of the Hospitality Leaders who had the concern that the new guidelines are full of minute details like hooks for curtains, water sprays in bathrooms blackout curtains which are used in evaluating the Hotels and these details are of little value in evaluating hotels. It was further pointed out that the new guidelines of the Ministry of Tourism were too low what the hotel brand standards have to offer

The Ministry of tourism guidelines classifies hotels based on star categorization (one, two, three, four and five star) hotels and heritage category hotels (heritage grand, heritage classic and heritage basic) hotels. The inspection of the Hotels is carried out by Hotel & Restaurant Approval

&Classification Committee which inspects and assesses the hotel on the basis of the facilities and services offered. The committee checks the projects on the basis of the following parameters for allocating the star category to the hotel (MOT 2013). These parameters include, location details, status of ownership, site details, project details (size, area, facilities, energy conservation, etc), blue Prints of all areas, proposed capital structure, skills development initiatives (running training courses).

The part two of the norms deal with the level of services that the hotel should offer. These facilities are further classified as guest room facilities, bath room facilities, public area facilities, facilities for disabled, food and beverage facilities, facilities in kitchen, guest services, safety and security, communication facility, eco friendly practices etc.

### ***Problem Statement:***

This Provides a complete background of the hotel classification in India and the factors that influence the classification system. Despite of the general external factors there are many internal factors that influence the classification task. In this project we are going to figure out key areas where the hotel classification making a bias in other words creating a class imbalance.

## **2.2. Proposed Solution**

The objective of this study is to study the key areas a classification system should look into while classifying a hotel into its star category. Further, the study identifies the geographical influencers of the hotel class that is opted due to the classifier. The study finally concludes the major takeaway feature that may cause a class imbalance in the hotel classification task.

The dataset used for solving the problem is taken from the official government website [data.gov.in](https://data.gov.in) which consists features such as hotel name and corresponding geographical location, contact details, category it falls into and number of rooms it comprises of. The data includes hotels of the type 4 star, 5 star , 5 star delux which comes under star category and heritage basic , heritage classic and heritage grand which falls into heritage category.

As the hotels under categories 3 star, 2 star and 1 star were not listed in the dataset I constrained the solution to the categories heritage and star only. Even without these data points there is a huge change of class bias that could happen which will we see in detail.

Some of the sub-problems that we are going to solve are:

- Identifying the top influencers in specific factor that affect the classification with respect to the category.
- Determining key statistics on the influencer to conclude the amount of bias that occurred during the process.
- Making sure to reduce the class imbalance produced.

## Chapter 3: Theoretical Analysis

### 3.1. Zonal Divisions in India

The states of India have been grouped into six zones having an Advisory Council "to develop the habit of cooperative working" among these States. Zonal Councils were set up wide Part-III of the States Reorganization Act, 1956. The North Eastern States' special problems are addressed by another statutory body - The North Eastern Council, created by the North Eastern Council Act, 1971. The present composition of each of these Zonal Councils is as under:

- **Northern Zonal Council**, comprising Chandigarh, Delhi, Haryana, Himachal Pradesh, Jammu and Kashmir, Ladakh, Punjab, and Rajasthan;
- **North Eastern Council**, comprising Assam, Arunachal Pradesh, Manipur, Meghalaya, Mizoram, Nagaland and Tripura; The State of Sikkim has also been included in the North Eastern Council wide North Eastern Council (Amendment) Act, 2002 notified on 23 December 2002.
- **Central Zonal Council**, comprising the States of Chhattisgarh, Madhya Pradesh, Uttarakhand and Uttar Pradesh;
- **Eastern Zonal Council**, comprising Bihar, Jharkhand, Odisha, and West Bengal;
- **Western Zonal Council**, comprising Dadra and Nagar Haveli and Daman and Diu, Goa, Gujarat, and Maharashtra.
- **Southern Zonal Council**, comprising Andhra Pradesh, Karnataka, Kerala.

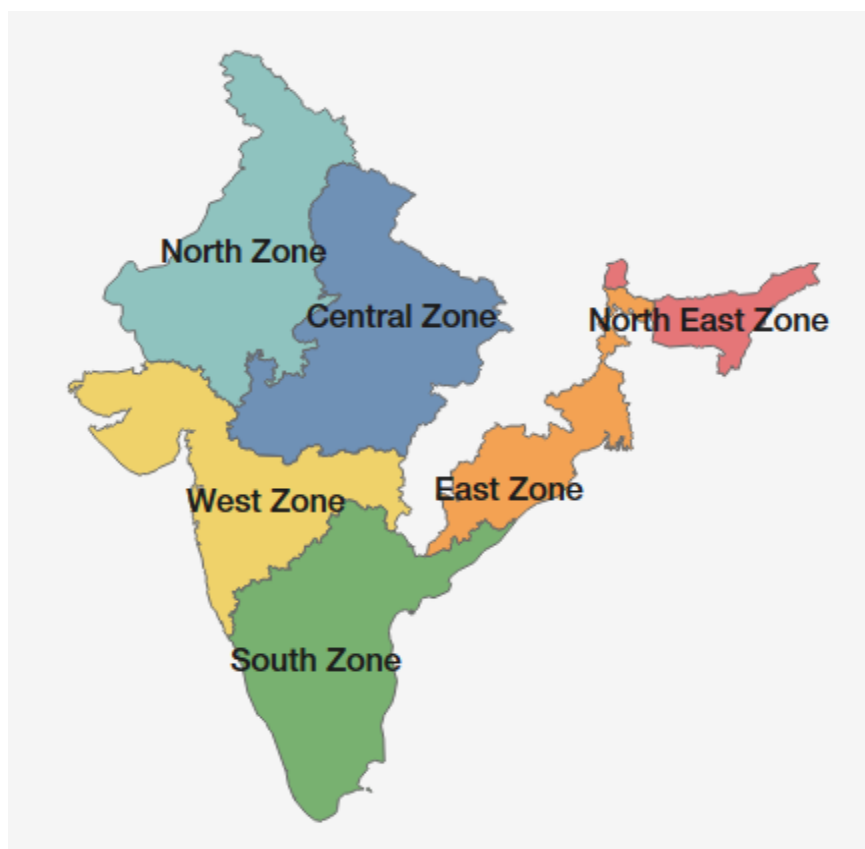


Figure1. Zonal Division in India

# Chapter 4: Experimental Investigations

## 4.1. Data Source Preparation

HOTEL NAME	ADDRESS	STATE	PHONE	FAX	EMAIL ID	WEBSITE	TYPE	Rooms
The Oberoi Cecil	"Ambedkar ChowkChaura Maidan, Shimla - 171004, HIMACHAL PRADESH"	HIMACHAL PRADESH	0177-2804848	0177-2811024	reservations@oberoigroup.com	www.oberoihotels.com	Heritage Grand	75
The Lalit Grand Palace	"Gupkar Road, Srinagar, JAMMU AND KASHMIR"	JAMMU AND KASHMIR	1.94E+09	1.94E+09	srinagar@thelalit.com	NA	Heritage Grand	112
Hotel Fateh Prakash Palace	"The City Palace Complex, Udaipur - 313001, RAJASTHAN"	RAJASTHAN	0294-2528016-9	0294-2528006	mgrfpp@udapur.hrhindia.com	www.hrhindia.com	Heritage Grand	30
Shiv Niwas Palace	"The City Palace ComplexUdaipur, Udaipur, RAJASTHAN"	RAJASTHAN	2.94E+09	2.94E+09	maltidutta50@yahoo.com	NA	Heritage Grand	36
Savoy Hotel	"77, Sylks Road, Ootacamund, TAMIL NADU"	TAMIL NADU	4.23E+08	4.23E+08	savoy.ooty@tajhotels.com	NA	Heritage Grand	40

Table 1. Head of the raw data

The above table provide with the first 5 rows of the data set that I used to solve the task. Each entry or row in the data tells about a specific hotel in India. It consists of the Hotel name, Address location, State, Phone and Fax numbers, Website and Email addresses, Type (category it belongs to) and Number of rooms it comprises of.

Perhaps you have noticed that there is only one numerical column provided in this data set. So I added a hotel id to uniquely identify the given hotel and create a secondary numerical attribute.

### ***Geo-coding address locations:***

As the scope of the project doesn't allow us to use any programming languages I manually geo coded the address locations using geo coders available in internet. one more challenge was to make the address accurately to the current address location which might have changes (since the dataset was older version).

Finally I mapped the corresponding latitude and longitude locations to a hotel name and saved in a excel sheet named *latlng mapper*.

### ***State aliases and capital mapper:***

I added a couple more sheets to map all of the corresponding capitals for a state and also state abbreviations to use in the visualizations.

Finally all of these experimental features are added to the dataset I started creating a join in tableau. The ones that are related to the state are mapped by the state column which itself is mapped using hotel name. I also created an empty column to map Telangana State into the map to get a perfect view for the data source version.

approved hotels is made of 4 tables. ⓘ

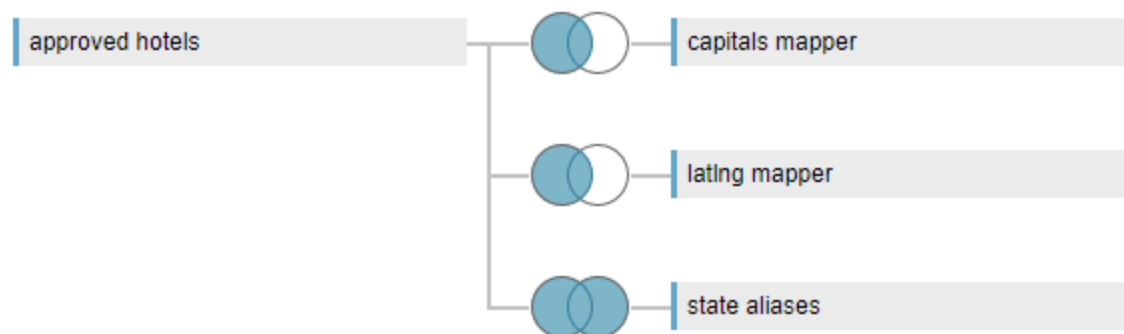


Figure 2. Data join structure

## Chapter 5: Solution Control Flow Chart

Every one has their own way of solving a data science problem. The following is the flow chart that I used in my problem solving process.

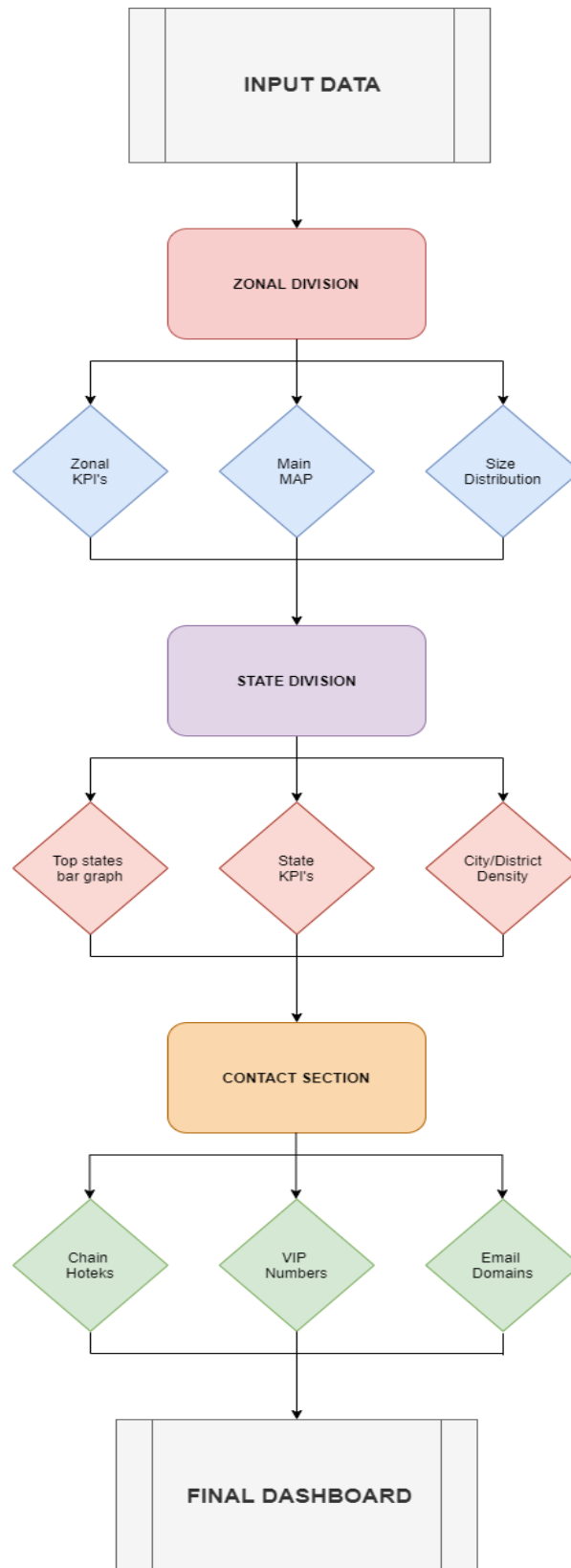


Figure 3: Flow chart



# Chapter 6: Results & Findings

## 6.1. Storytelling Dashboard

### ***Zonal Level Section :***

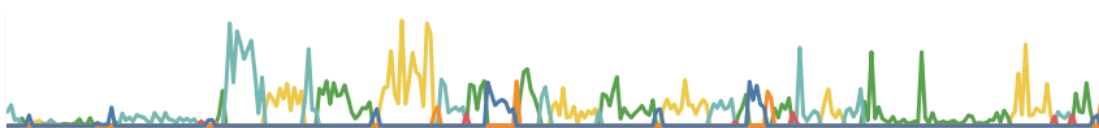
Firstly I made a group named *Zone by State* using the State column in the data set and then created a map using the state as shown in *figure1* above. Then I mapped each and every data point (i.e. hotel Id ) to a different map to communicate data about each individual hotel. By using tableau actions I created a zone filter action which will zoom into the zone and show the data points precisely. I also created a distribution of hotel size by using our two numerical columns namely hotel id and rooms which could be filtered by zone. Finally I included some relevant KPI's that would show the percentage of hotels by zone and type each.

### **Hotel Classification in India**

#### **Percentage of Rooms by Type**

4 Star	5 Star	5 Star Deluxe	Heritage Basic	Heritage Classic	Heritage Grand
21.49%	24.74%	48.60%	4.11%	0.20%	0.87%

#### **Hotel Size Distribution**



Select a zone



Figure 4: Zonal Section

Some of the key takeaways here are

- Hotels belonging to the star category seems to occupy more and hence there could be a bias occurred while classification.
- A few hotels are only large sized.
- Looking at the main map, we see that the hotel concentration is more on North and South. So a bias may also occur in hotels of this region

## State Level Section :

Following is the state level analysis of the category star and heritage separately. The reason why I splitted it into two different types is to understand if the bias itself separates. Here we can see top 10 state with respect to its type base on the hotel size. Also I created a density map based on the city or district field which I extracted from the address location and calculated their concentration. In the middle of the two I included a KPI that indian no of hotels that are near to the capital.

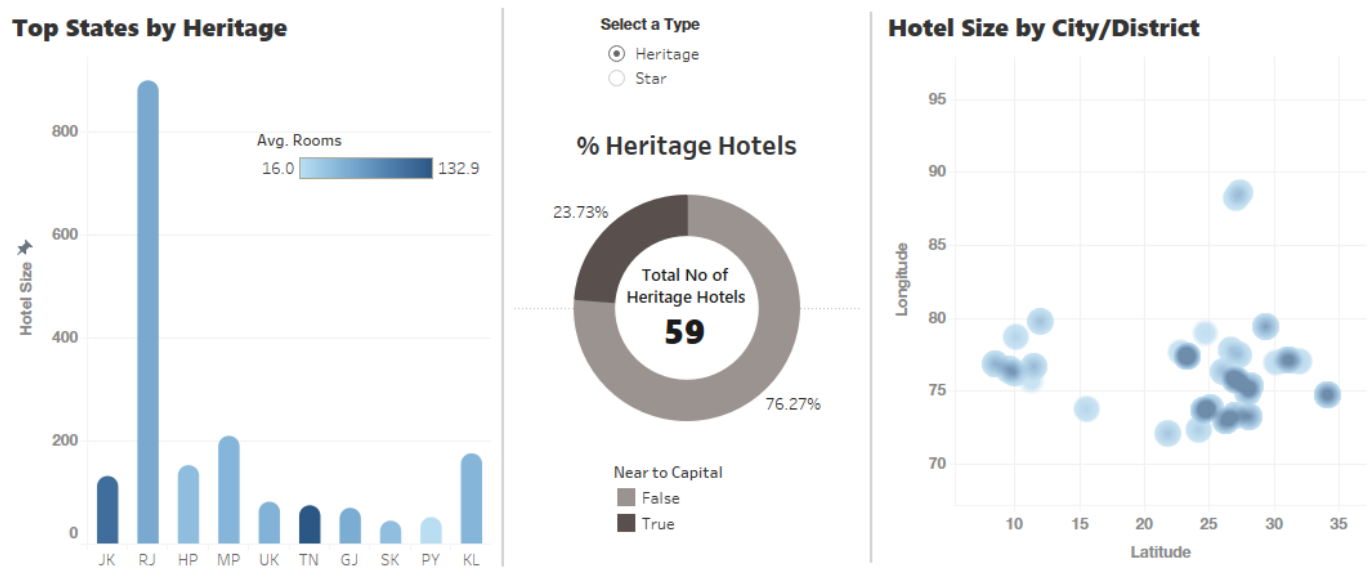


Figure 4: State Section

Some of the key p[oints here are:

- There doesn't seem a lot of bias with respect to the distance from capital of a specific state.
- Rajasthan comprised largest portion (51.95%) of all of the top hotels and 2nd when comes to average.
- Bias due to an Individual city or district increases when we work on specific state.

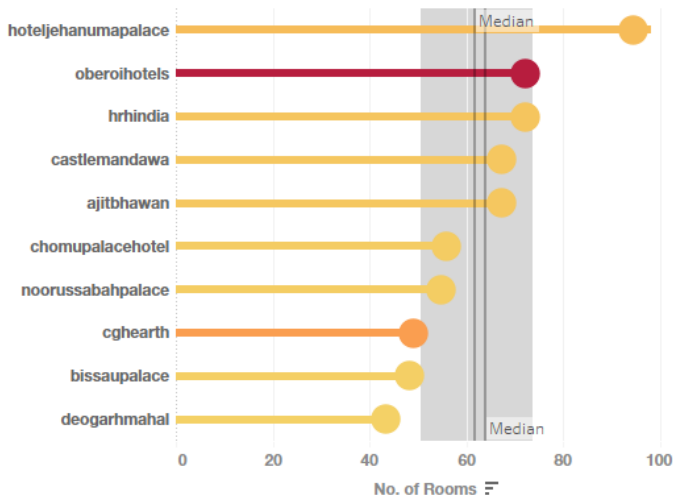
## Contact Section :

In this section I focused on the website and email domains. The hotels that are using the same website domain could be considered to belong to a same chain. So I listed out the top chain hotels by type. Also I calculated the top email domains by type. I included some non interactive KPI's about usage of VIP phone numbers and a distribution of fax number usage.

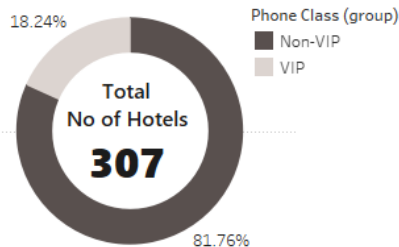
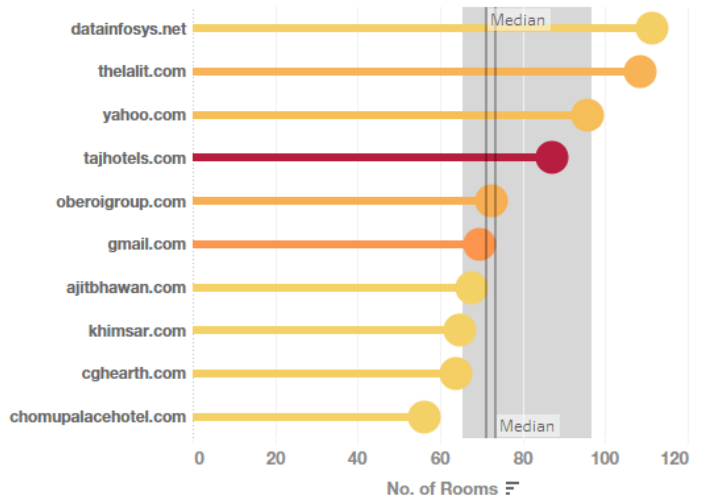
Some key points are:

- When it comes to star category *taj hotels* stands first with respect to both website and email domains.
- A very less people are using VIP numbers which brings up the context to study the Non VIP numbers in depth.

### Top Website Domains by Heritage



### Top Email Domains by Heritage



### Distribution of Fax Usage

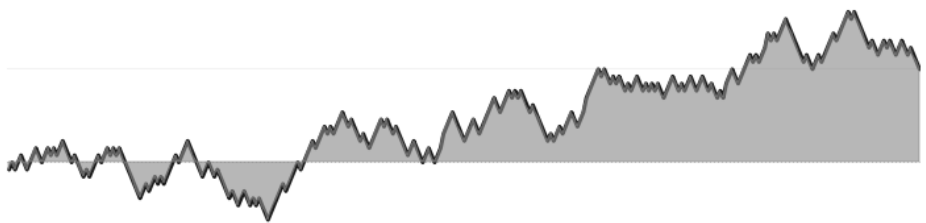


Figure 6: Contact Section

## Chapter 7 : Advantages and Disadvantages

### Advantages:

- This helps in reducing the bias that might be created while approaching the classification task.
- We could recommend model builders or data scientists has to work on certain areas that are likely to produce class imbalance.
- This dashboard also helps in to find out that instead of working on whole data we can work on single or group of attributes of precise classification.

### Disadvantages:

- Due to the more focus on the class imbalance ther might be a chance that ther could be undefitted classifier to be created
- One can also take another percpetion rather more focus on bias which leads to imbalance.

## Chapter 8: Applications

This project is only a visual analysis of the hotel classification, its influencers and bias produced. But one needs to do a comprehensive analysis using a programming language such as R or Python. Nevertheless this could be useful in the following ways:

- To identify the key areas while working in a classification task that uses similar or same dataset.
- To avoid categorical or class bias produced due to many factors or features present in the dataset.
- Analyzing most important attributes that contribute to each category through KPI's.

## Chapter 9 : Conclusions

Hotel classification is a vast domain of its own with which I explore a little bit of it. This is to summarise all of the points we have seen so far. Firstly we looked at the historical and legal scope of the project pretending that these hotel classifications are all official done by Government of India. We checked the previous methods that are taken and the approach that I am going to solve the problem. Then we take a look in to the geographical divisions of state across India. Thereafter we prepared, cleaned the dataset for the use case. With which we entered into the problem and created various stages namely zonal section, state section and contact section. We have looked into sub goals that we have solved with respect each section and their key take aways. Finally we looked into the advantages, disadvantages and applications of the project by an individual and a data driven person.

## Chapter 10: Future Scope

In this project I worked with a dataset that is a little bit outdated but still it gave us some valuable insights on the areas to work when doing the classification task. But I believe a much new dataset would serve better in this context. The factors we choose to study could not tell the bias in a greater amount of detail. Adding more features or imputing some others from the existing ones would be better to solve the problem. In the history section we have seen the major takeaways that government considers while hotel classification. In there it mentions about the services provided by the hotel and its quality. So this could be an extended area which can be further analyzed in terms of influencing the class imbalance.

Also you can take a deep look into the Address locations of each hotel to find some insights. Or you can take a look at the Non VIP phone numbers in a deeper sense. It all depends on you as an individual to solve this problem. We all are here to explore so I hope my short findings will lead you through a much better insights.

## References/ Bibilography:

- i. [Making Flow Happen: Dashboards that Persuade Inform and Engage by Jeff Pettiross, User experience designer, Tableau Software](#)
- ii. [Hotel Guidance Form 19-01-2018 by H & R Division, Ministry of Tourism, Government of India](#)
- iii. [Administrative Divisions of India by Wikipedia](#)
- iv. [The Antomy of URL by Blair Millen \(Also known as deopud\)](#)
- v. [Chapter 1. The Seven Stages of Visualizing Data by Ben Fry - Oreilly](#)
- vi. [Qualitative Case Study Methodology: Study Design and Implementation for Novice Researchers by Pamela Elizabeth Baxter and Susan M Jack, McMaster University](#)
- vii. [Morning News: Benefits of using VIP phone numbers by Saahil Khan, July 27, 2020 in Technology](#)

## **Youtube video :**

- i. [Tableau Zen Masters Tricks and Tips - Chats Edition by Andy Kriebel](#)
- ii. [Next level Viz in Tooltip by Tableau Software](#)

## **Websites for Geo-Coding :**

- i. [www.gps-coordinates.net/gps-coordinates-converter](http://www.gps-coordinates.net/gps-coordinates-converter)
- ii. [www.latlong.net/convert-address-to-lat-long.html](http://www.latlong.net/convert-address-to-lat-long.html)
- iii. <https://stevemorse.org/jcal/latlon.php>

## **Books:**

- i. Data Visualization: A Handbook for Data Driven Design by Andy Kirk
- ii. Undestanding Comics: The Invisible art by Scott Mc Cloud









