HOTEL BOOKING DEMAND

A Minor Project Report

in partial fulfillment of the degree

Bachelor of Technology
in
Computer Science & Artificial Intelligence

By

2003A52094	A.SAI CHARAN REDDY
2003A52095	G.MANIKANTA REDDY
2003A52073	K.SAIKUMAR

Under the Guidance of

Mrs.L.Mounika Asst.professor CSE Department

Submitted to

SCHOOL OF COMPUTER SCIENCE & ARTIFICIAL INTELLIGENCE
SR UNIVERSITY, ANANTHASAGAR, WARANGAL

April, 2023.

SCHOOL OF COMPUTER SCIENCE & ARTIFICIAL INTELLIGENCE

CERTIFICATE

This is to certify that this project entitled **“HOTEL BOOKING DEMAND**" is the bonafied work carried out by **A.Saicharan,G.Manikanta,K.Saikumar** as a Minor Project for the partial fulfillment to award the degree **BACHELOR OF TECHNOLOGY** in **COMPUTER SCIENCE & ARTIFICIAL INTELLIGENCE** during the academic year 2022-2023 under our guidance and Supervision.

Mrs.L.Mounika
Assisstant Professor,
SR University,
Ananthasagar, Warangal.

Dr. M.Sheshikala
Assoc. Prof. & HOD (CSE),
SRUniversity,
Ananthasagar, Warangal.

External Examiner

ACKNOWLEDGEMENT

We owe an enormous debt of gratitude to our project guide **Mrs.L.Mounika Assistant Professor** as well as Head of the CSE Department **Dr. M.Sheshikala, Associate Professor** for guiding us from the beginning through the end of the Minor Project with their intellectual advices and insightful suggestions. We truly value their consistent feedback on our progress, which was always constructive and encouraging and ultimately drove us to the right direction.

We express our thanks to project co-ordinators **Dr. P Praveen, Assoc. Prof., and Dr. Mohammed Ali Shaik, Asst. Prof.** for their encouragement and support.

We wish to take this opportunity to express our sincere gratitude and deep sense of respect to our beloved Dean, **Dr. C.V. Guru Rao,** for his continuous support and guidance to complete this project in the institute.

Finally, we express our thanks to all the teaching and non-teaching staff of the department for their suggestions and timely support.

A.Saicharan Reddy

G.Manikanta Reddy

K.Saikumar

ABSTRACT

In this Project 'Hotel Booking Demand', an accurate booking cancellation forecast by which user know the things related to hotel bookings very earlier. Booking cancellation has a significant effect on revenue which essentially affects request the board choices in the inn business. To reduce the cancellation effect, the hotel applies the cancellation model as the key to addressing this problem with the machine learning-based system developed. By combining data science tools and capabilities with human judgement and interpretation, this project aims to demonstrate how the predictive analysis of the model can contribute to synthesizing and predict about booking cancellation forecasting. Furthermore, this project aims, by detailing the full prediction & analysis, to give relaxation to user who want to apply in particular hotel. By Implement Various Algorithms like Random Forest, Decision Tree, etc. to classify the data and also use Confusion Matrix to separate categorical data in particular type, user can know the prediction up to the desired level. It prevents the hotel as well as Tourists to poor dealing of room. User/Customer have to enter certain field by which this model detects his prediction about the cancelation.

Table of Contents

LIST OF FIGURES

1.INTRODUCTION

1.1 OVERVIEW

Hotels plays an important role for any person or traveller who are travelling from one destination to another. Hotel play an important role for tourists whether the tourist is local or international. Hotel provides many best services to the customer such as parking area, food, room service and also it provides services that is offered by customer. By providing these services Hotels take the valuable feedback from the customers. By these feedbacks Hotels maintains their reputation in the city/area. If the services are poor, the bookings of that hotel are low and if the services are awesome then high bookings in that hotel takes place. In this model, the prediction possibility of a booking for a hotel based on different factors and also try to predict if they need special requests based on different features. In this project the dataset which we are using contains both Resort and City Hotel data. Here we use many popular Machine Learning Algorithms like Decision Tree, Random Forest. To predict the cancellation chances. This Model gives the prediction of Hotel Booking Cancellation up to the certain level ofaccuracy i.e., 76% (Approx.).

1.2 PROBLEM STATEMENT

Predict whether a booking will be canceled or not to allow the hotel's manager to make better decisions in order to improve overbooking strategies and cancellation policies.

With the increasing day to day trend of hotel cancellation at border time, it affects the local as well as international tourists more and more. Customer/Tourists don't have any idea about the pre-cancellation scenario of particular hotel at the time of booking of that hotel.

1.3 EXISTING SYSTEM

Reservation is done via phone calls or by visit in person to the hotel reservation office. The guest's personal details such as Name, gender, Age and Duration of visit or stay, are entered during booking made. Then the booking officer asks to prepare room for guest before his/ her check in date. The data and documents are transferred manually to the appropriate office for compilation of the guest's file. On the day of check in date the file is transferred to the reception. On the day of checking in the guest is given the key to his room, she/ he also specify if room service is needed.

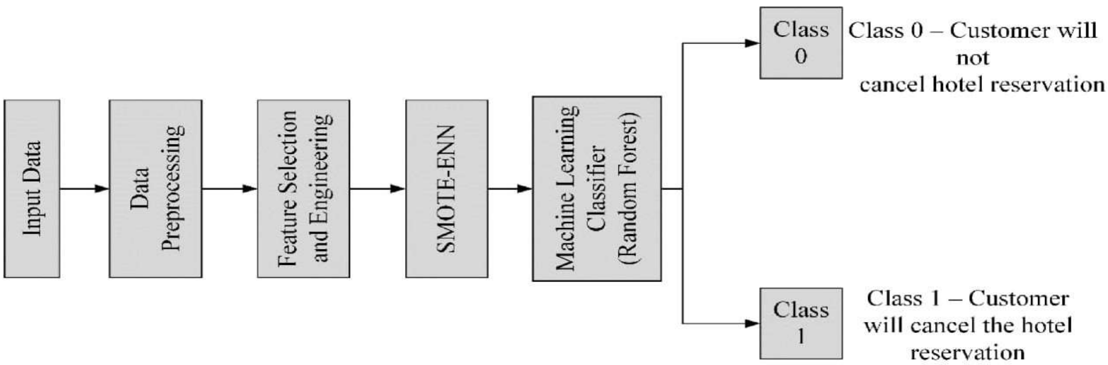
1.4 PROPOSED SYSTEM

Based on the above-mentioned reasons, we propose an system for a hotel booking system. The main idea of the proposed expert system is to interconnect the hotel with places, activities and interesting events around, and based on user preferences to suggest suitablehotel services and suitable activities for the hotel guest. The proposed expert system will be connected to a simple questionnaire to detect guest preferences . The system consists of a knowledge base which contains IF-THEN rules for determining which hotel services and activities around the hotel are more or less suitable for Hotel guest.

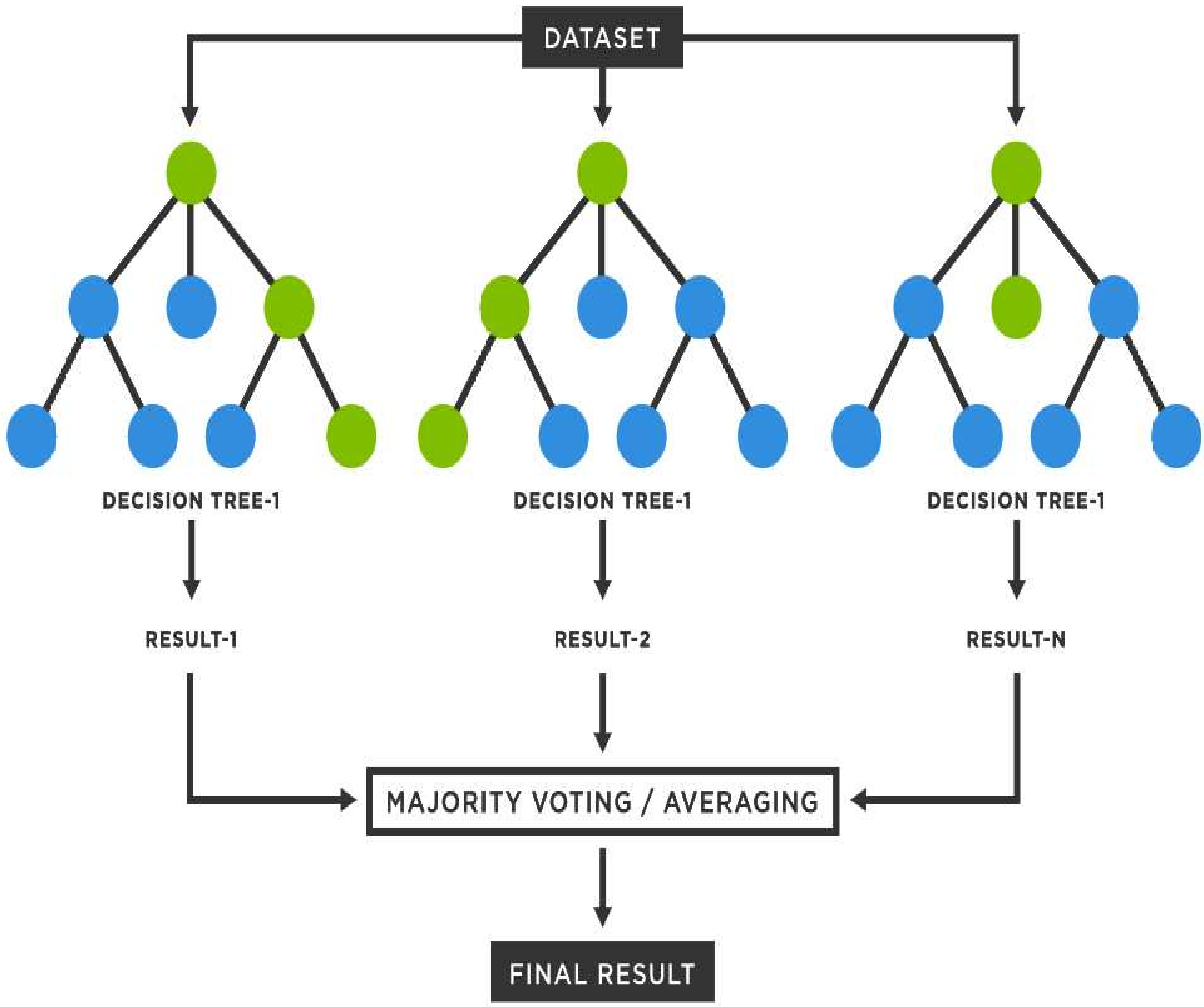
1.5 OBJECTIVES DESCRIPTION

- Provide people seeking hotel rooms with accurate information about available accommodations.
- Allow booking of rooms without errors and without creating conflicts.
- To maximize occupancy
- To provide up to date information about the status of reservations
- To allow staff to easily update information in the system, and have the system inform everyone who needs to know

1.6 OVERALL ARCHITECTURE



From fig 1.6.1 the architecture describes how the process of hotel booking demand goes from starting to end. The model stores the information whether the Rome should booked ornot.



From fig 1.6.2 It Describes about RandomForest and DecisionTree Architechture.

2.LITERATURE SURVEY

A. Predicting Hotel Bookings Cancellation with a Machine Learning

Classification Booking cancellation is a very common thing in today's world, which can cause severe losses to the business owners. This paper describes how AI is used to identify which booking can be cancelled and prevent some losses. The machine learning model should be evaluated in the real time environment for accuracy. Predictionmodel of hotel booking cancellation no doubt the issue that can be resolved in the contextof Design Science Research (DSR), as it need to develop an artifact, here in this particular case, a form of Revenue Management System (RMS), fulfilling the two requirements of DSR.

B. Rising rate of Cancellation in the Hotel Industry

The growing trend of Hotels Industry is beneficial for Hotels but there are some problems too such as Rising Rate of Cancellation. The user cancels the booking of the hotel after seeing the reviews given by the people who booked the hotel already. In Some case hotel owners treat the customer in bad way, this also affect the reputation as well as cancellations. The growing trend of Hotels Industry is beneficial for Hotels but there are some problems too such as Rising Rate of Cancellation. The user cancels the booking of the hotel after seeing the reviews given by the people who booked the hotel already. Now a days, people expecting a better accommodation at the Hotel site, if people found any lag in accommodation then they give poor rating of that Hotel. So, if we looking at percentage of cancellation then we found that the percentage of cancellation is increasingday by day. From a survey Cancellation rate rose from under 33% in 2014 to 40% in 2018. Also, during the COVID-19 pandemic this percentage gets increased because peoples book their Hotels in very earlier time and after changing situation day by day.

C. Aspect based Sentiment Oriented Summarization of Hotel Reviews

The reviews and the feedbacks of the customer play an important role in the image as well as the revenue system of the hotel. But most of the travelers don't read all reviews. The system analyzes the reviews and feedback by the customers. The feedbacks of the customer are gathered from the hotel's website and the stored as classes. As per the study, the model analyzes the overlooked information by the customers and takes some essential steps. Finally, after processing all the data collected an emotional analysis is done. The hotels thereby can take the required steps to improve their service.

D. Application Of machine Learning in Hotel Industry

A Critical Review The growth in IT industry also affects the Hotel Industry. However, this change is quite slow. Many researchers are focused on testing and applying new artificial intelligence technology and learning equipment in the hotel industry. The study offers a brief knowledge about the use of Machine Learning and its combined technology in the hotel and tourism industry. Machine learning is quite trending these days.

E. Social media on the Internet**:**

While there is a lack of a formal definition, "social media" can be generally understood as Internet-based applications that carry consumer-generated content which encompasses "media impressions created by consumers, typically informed by relevant experience and archived or shared online for easy access by other impressionable consumers (Blacks haw). Social media exists in a variety of forms, including Social Networking Sites, Video Sharing Sites, Photo Sharing, Collaborative Directories, Social type Sites, Content Voting Sites, Business Networking Sites and Social (Collaborative) Bookmarking Sites (Tang, 2011). Despite social media's significant

F. Communication rationale of user-generated content**:**

Because hospitality and tourism products are intangible with an experiential nature, e-word of mouth has become important in travel planning (Kim *et al.*, 2011). With development of Internet, the traditional word of mouth, such as verbal communication, mass media, advertisements cannot satisfy tourists requirements, but user-generated content supported through social media is “a mixture of fact and opinion, impression and sentiment, founded and unfounded tidbits, experiences and even rumor” (Blackshaw, 2006).Accordingly, many research studies also focus on Use-Generated Content (UGC) effect, in particular, online reviews and travel blogs (Kwok and Yu, 2013).

There is a specific standards to explain UGC: (1) On the premise of Internet publishing, (2) Creativity content and (3) Created by non-professionals (Zhao *et al.*, 2012). Accordingly speaking, online review is also a common form of UGC and applied to tourism marketing widely. Past research has explained the communication process by “Transmission theory of ternary”, that is say, both sides of communication and the content of comment is pivotal in purchase decision of potential customer (Hao, 2010). Generally, professional degree, opinion leader and relationship of sender and receiver are main indexes to measure the reliability of information sources (Gilly *et al.*, 1998); the quality and quantity of comments, tendentiousness and strength would change purchase intention of different level (Park *et al.*, 2007); Moreover, receiver’s professional and sensory ability and risk perception also take important positions in the communication process. In short, the major factors that influence the spread of UGC are comment volume, rating, valence, variance, dispersion and so on.

G. Online reviews and tourism reservation:

Many research studies focus on online reviews' effect, especially in destination and hospital. For example, O'Connor's analysis showed hotel location, room size, staff and cleanliness were important attributes for positive online review and few hotels responded to traveler's online reviews. Pantelidis used a content analysis approach to analyze about 2500 online reviews of 300 restaurants in London and summarized the key word from these sentences (Pantelidis, 2010). In addition, the relationship between the sales volume and reviews are proved to be significant, including positive attribute which can greatly stimulate reservation intention of potential customer (Ye *et al.*, 2010).

However, the report mapping the travel mind-the influence of social media joint published by Conrad advertising and You Gov concluded that social media's influence on the travel plan was not greater than our imagination (ChinaFace, 2011).

H .There are two key components about the research:

(1) The online traveler (hotel potential customers), who is driven by a number of personal and trip-related needs, (2) The online tourism domain, which is composed of informational entities provided by a number of "players", including individual consumers through means of social media, this tourism domain has a distinct semantic structure determined by the hyper textual nature of the internet and the tourism industry structure (Zheng and Bing, 2011). Thus,this study attempts to investigate the consistency between online information's demand and supply and find the key factors to influence hotel reservation

I. Through the analysis of several research papers:

It is possible to identify the key factors that lead users to provide a review on purchased tourism services. As main reasons in providing reviews by consumers, most of the studies, here analyzed, highlight aspects such as «quality of service», «customer satisfaction and dissatisfaction» and «social identity and sense of belonging to the community» (Crotts *et al.,* 2009; Swanson and Hsu, 2009; Kim *et al.,* 2009; Casaló *et al.,* 2010; Sun and Qu, 2011; Sánchez-García and Currás-Pérez, 2011; Nusair *et al.,* 2011; Bronner and Hoog, 2011; Casalo *et al.,* 2015; Gvili and Levy, 2016; Zervas *et al.,* 2017; Yen *et al.,* 2019; Yang, 2018b).

According to Bronner and Hoog (2011), the most frequent motivation to perform e-WOM is to provide useful information to allow others to make a satisfactory choice. Furthermore, some of the research findings reveal that negative reviews can be encountered more often than positive reviews. Yet, Swanson and Hsu (2009) argue that customers, who have experienced satisfying experiences, are not necessarily inclined to recommend the service or to persuade others to use the same services. In this line of research, Sánchez-García and Currás-Pérez (2011) adopt an unusual perspective in the literature by proposing that dissatisfaction, caused by perceivedservice failure, tends to trig for emotional processes that lead consumers to experience and manifest through their reviews specific emotions such as anger, disappointment which, in turn, may induce agents' behavioural response. Potential consumers are then inclined .

J. In terms of consumers' emotional involvement:

Hu and Kim (2018) examine the effects of e-WOM motivations on customers' e-WOM behaviour. The authors reveal important implications for the hotel sector and online marketing managers. First, self-enhancement and enjoyment is the most important drive for hotel guests to spread positive e-WOM. Moreover, according to the authors, positive online comments are related more to a pleasant stay rather than economic incentives. Hoteliers should ensure that eachguest has the opportunity

K. Several studies have also explored the impact:

That reviews, provided by consumers, have on the hotel sector. Specifically, authors tend to highlight how especially positive judgments can increase the number of bookings and, consequently, the productivity of the firm, thus providing hoteliers with important information for their marketing strategies. Qiang et al.(2009) empirically investigate the impact of online consumer-generated reviews on hotel room sales. Utilizing data collected from the largest travel website in China, the authors develop a fixed effect log-linear regression model to assess the influence of online reviews on the number of hotel room bookings. In details, the study shows that positive online reviews can significantly increase the number of bookings, while the variance or polarity of e-WOM (*i. e.* rather volatile judgments or their total absence) exerts a negative impact on the number of online sales. Qiang *et al.* (2011) conduct a further study to identify the impact of online user-generated reviews on business performance, using data extracted from a major online travel agency in China. The empirical findings show that traveller reviews have a significant impact on online sales. Notably, the variable that mostly affects the reservations is related to the average and variance of the judgments.

L.

In addition, Lin and Xu (2017) suggest that reviews not only have an effect on perceived reviewer trustworthiness, but also on brand attitude and purchase intention. A positive review enhances reviewer trustworthiness since it is viewed by potential consumers as being fair and believable, and it can predict a stronger purchase intention while vice versa a negative review. Mauri and Minazzi (2013) also assess that the prevalence of positive/negative comments will increase/decrease the hotel purchase intention and the level of potential consumers' expectations. Blal *et al.* (2014) analyse the tourist market in the city of London, through data on 319 hotels from *tripadvisor.com*. The results show a positive correlation between the ratings and volume of reviews, and hotels' revenue. Positive reviews generate an increase in sales, while negative reviews generate a decrease. Specifically, ratings have a larger effect on upper-tier hotels, while volume of reviews drives the lower-tier hotels' results.

M.

On the whole, several other studies have identified a link between the volume of reviews and hotels' revenue. Experts have advocated that increasing the volume of online reviews can help mitigate negative comments (Teixeira and Kornfeld, 2013), improve consumer perception (Viglia *et al.,* 2014), and eventually, improve operational performance (Kim *et al.,* 2015). Therefore, as stated by García *et al.* (2017), indicators such as rating and volume are able to influence consumers' willingness to pay.

A further line of research explores in what measure hotel guests reviews, posted on consumer-generated websites, have an influence on consumers' decision-making process and service expectations. Mauri and Minazzi (2013), through survey data gathered on students or young graduates in the main university cities in Northern Italy, reveal a positive correlation between hotel purchasing intention and customers' expectations with respect to the review rating. Moreover, hotel managers' responses to guests' reviews exert a negative impact on customers' purchasing intentions. Tsao *et al.* (2015) .

N.

Booking intentions amongst those individuals who are strongly inclined towards conformity. Conversely, a higher number of reviews prove to be more persuasive amongst those individuals who are characterized by a low degree of conformity. Another thread of research relates to the effects of social networks. In this respect, Ladhari and Michaud (2015) explore the influence of comments published on Facebook on friends' intentions to book a specific hotel, the trust and the attitude towards this hotel, and the perception about its website. A survey conducted on a sample of Canadian students, under the age of 35, has confirmed all the hypotheses on the remarkable influence of such comments that are able to drive users' decision-making process.

O.

There are several studies that show that guests' rating on hotels is a determinant variable that should be considered when implementing pricing policy. Qu (2014) find that the inclusion of additional reviews, obtained from other travel sites, can provide hotels more reliable information for assessing the effect of customers' opinions on business productivity. Torres *et al.* (2015) also explore the impact of a hotel rating and number of reviews on the value generated through online transactions. Through a sample of 178 hotels, representing various types of firms and brands within the United States, the authors assess that ratings in TripAdvisor, as well as the number of reviews, are positively correlated to the average size of each online booking transaction. Hence, regarding e-WOM volume, a large number of comments is desirable only for firms with positive ratings that meet clients' expectations. Indeed, for these firms, the positive effect of rating can beeven more strengthened by volume. Therefore, hospitality operators should make an effort to satisfy their clients as well as encourage them to publish online feedback. In this manner, firms' online reputation will increase allowing hotels to raise their prices and obtain a higher profitability for their business.

P.

On the demand side, several studies find that customers' rating boosts hotel performance and affects hotel room prices (Ogut and Onur Tas, 2012; Lu and Stepchenkova, 2012; Nieto *et al.,* 2014; Hernández-Maestro and Muñoz-Gallego, 2014; Viglia *et al.,* 2016b; Guizzardi *et al.,* 2017). Acar *et al.* (2012) remark that hotels need to follow effective and efficient promotional policies based upon effective dynamic pricing strategies. In this respect, Abrate and Viglia (2016) suggest that hospitality operators should adjust their prices in line with reviewers' evaluations about their accommodation. Hence, online reputation, by means of online customers'reviews, plays an increasing role in pricing making decisions.

In terms of dynamic pricing strategy, Viglia *et al.* (2016a) remark that an important indicator is the so-called «reference price». This price is a benchmark that consumers use to evaluate prices on the market and purchase a specific product. In particular, the authors show how pricing and discounting policies affect reference price formation.

Q.

A further drive of pricing policy relates to customers' characteristics and clustering. Abrate *et al.* (2012) analyze the dynamic price decisions within the hotel sector. The authors conclude that the inter-temporal pricing structure often depends on a price discrimination policy based on customers' clusters, stars rating, as an indicator of quality, and the number of services supplied. Empirical results show that when customers belong to the business cluster, the lowest prices seem to be set in the time span immediately preceding customer staying. On a weekend, when the leisure cluster is predominant, prices tend to increase when approaching to the check-in date. Wu *et al.* (2014) show that a randomized pricing strategy tends to generate higher profits than a flat pricing strategy. The authors suggest that the online retailer should adopt the discounted price for only one period and then return to the baseline price. When low- type consumers are more patient, the retailer should decrease the promotion frequency and, simultaneously, adopt a high price. Moreover, the online retailer should diminish the promotion frequency and the high price and, simultaneously, increase the low price to encourage high-type consumers, who value time more, to purchase at the high price.

R.

From a standard economics perspective, García *et al.* (2017) remark that hotels tend to implement their pricing strategies based upon forecasted levels of demand, demand price elasticity as well as competitors' prices. Yet, the higher volatility of the online market place makes more and more difficult to make predictions and forecasts demand pattern. The role of pricing policy in online transactions should able to maximize sellers' profits having in mind consumers' product evaluations (Kim *et al.,* 2009). As suggested by Guo *et al.* (2013a), an appropriate policy of market segmentation developed through an online reservation systemwould benefit both hotels and consumers. By reaching an optimal number of demand segments, firms would obtain higher profits, while consumers would gain considerable price discounts.

S. Supply and online pricing policy

As the standard economics theory suggests, the level of supply influences price dynamics. Prices tend to increase when there is a scarcity of hotels available to bookin in a certain area (Ibrahim and Atiya, 2016). Abrate and Viglia (2016) support that tactical price decisions tend to be influenced by the amount of online real-time competitors. On the same line of research, Balaguer and Pernìas (2013) conduct a study on the relationship between the numberof competitors and hotel prices in Spain. A higher density of competitors implies, on average, a lower price dispersion. The findings suggest that the entry of a new competitor in the neighbourhood will force a reduction in the optimal level of prices in the area. For midweekdays, the effect on price level will be higher if the new entrant offers the same quality of accommodation. The effect of a new hotel on the price level will be also lower at the weekend, where there is a higher proportion of potential consumers. Becerra *et al.* (2013) argue that the degree of local competition mitigates the effect of differentiation on pricing policy; but hotels characterized by higher quality services (expressed in terms of number of stars) can better withstand the entry of new competitors that would impose price cuts. Furthermore, Xie and Kwok (2017) provides insight on the impact of innovators as Airbnb on hotels, in a increasingly competitive hospitality market, driven by ever-changing technology and innovation. The study empirically considers in what measure Airbnb pricing policy affects the performance of nearby hotels through the lens of price difference and price dispersion.

T.

Demand and vary prices depending on the day of the week during the event. Pricing policy are also influenced by other factors such as the presence of large events and taxation, although the thread of this research is still scarce. As an example, Herrmann and Herrmann (2014) investigate hotel prices in Munich under the influence of the Oktoberfest. In particular,the author analyses how the event affects the daily online price level as well as hotels price differentials.

U.

Online Travel Agencies (OTA) also play an important role in tourism web marketing via online brokerage agencies (Internet Distribution Systems, IDS). These are portals where consumers can compare different offers, read the reviews and make a reservation. Nowadays, OTA have become a key distribution channel for many hospitality firms providing not only the possibility to hold reservations, but also a higher visibility to hotels. Several papers are devoted to explore the role of OTA in hospitality and tourism research (Guo *et al.,* 2013a; Blal and Sturman, 2014; Ling *et al.,* 2014; Mellinas *et al.,* 2015).

The use of OTA platforms is also a useful tool for monitoring and managing e-WOM (Yang, 2018a, b). Sorzabal *et al.* (2013) remark that IDS are an essential tool for the tourism sector since these provide a flexible way for changing prices at a real time. Raguseo *et al.*(2017) show that hotels listed on multiple OTA are able to boost sales revenue and operating profitability. The authors explore in what manner a higher visibility influences firms' business profitability. They find that those hotels that have chosen to advertise their rooms on a higher number of OTA are able to boost sales and business profitability and, thus, capture more economic value from their visibility on these distribution channels. On the opposite, the authors show that visibility on TripAdvisor in the form of either higher review ratings, larger review volumes, and higher variance in ratings and hoteliers' responses has no significant effect on revenue and business profit growth. As a further outcome, Xiang *et al.* (2017) indicate that, within the hotel sector.

3.DATA PRE-PROCESSING

3.1 Dataset Description:

This dataset is taken from UCI Machine Learning Repository. This dataset contains 32columns and 87230 rows. It contains 32 attributes, contains 492 null values

This data set contains booking information for a city hotel and a resort hotel, and includes information such as when the booking was made, length of stay, the number of adults, children, and/or babies, and the number of available parking spaces, among other things. All personally identifying information has been removed from the data.

```
[ ] df.describe()
```

	is_canceled	lead_time	arrival_date_year	arrival_date_week_number	arrival_date_day_of_month	stays_in_weekend_nights	stays_in_week_nights	adults	children	babies	is_repeated_guest	previous_cancella
count	119390.000000	119390.000000	119390.000000	119390.000000	119390.000000	119390.000000	119390.000000	119390.000000	119386.000000	119390.000000	119390.000000	119390.0
mean	0.370416	104.011416	2016.156554	27.165173	15.798241	0.927599	2.500302	1.856403	0.103890	0.007949	0.031912	0.0
std	0.482918	106.863097	0.707476	13.605138	8.780829	0.998613	1.908286	0.579261	0.398561	0.097436	0.175767	0.8
min	0.000000	0.000000	2015.000000	1.000000	1.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.0
25%	0.000000	18.000000	2016.000000	16.000000	8.000000	0.000000	1.000000	2.000000	0.000000	0.000000	0.000000	0.0
50%	0.000000	69.000000	2016.000000	28.000000	16.000000	1.000000	2.000000	2.000000	0.000000	0.000000	0.000000	0.0
75%	1.000000	160.000000	2017.000000	38.000000	23.000000	2.000000	3.000000	2.000000	0.000000	0.000000	0.000000	0.0
max	1.000000	737.000000	2017.000000	53.000000	31.000000	19.000000	50.000000	55.000000	10.000000	10.000000	1.000000	26.0

Fig 3.1.1

df.describe() tells about our dataset which contains rows and columns including 492 null values. The above diagram states that it contains 32 attributes and also contains null values ,this above diagram is shown before cleaing of the dataset.

3.2 DATA CLEANING

Data cleaning is the process of fixing or removing incorrect, corrupted, incorrectly formatted, duplicate, or incomplete data within a dataset. When combining multiple data sources, there are many opportunities for data to be duplicated or mislabeled.

```
import pandas as pd #data preprocessing
import numpy as np #linear algebra
import matplotlib.pyplot as plt #matplotlib work like MATLAB
import seaborn as sns#uses Matplotlib underneath to plot graphs
import warnings#to ignore warnings
warnings.filterwarnings('ignore')
```

```
## Importing Data
df = pd.read_csv("/content/drive/MyDrive/hotel_bookings.csv")
```

```
[ ] df.head()
```

	hotel	is_canceled	lead_time	arrival_date_year	arrival_date_month	arrival_date_week_number	arrival_date_day_of_month	stays_in_weekend_nights	stays_in_week_nights	adults	...	deposit_type	agent
0	Resort Hotel	0	342	2015	July	27	1	0	0	2	...	No Deposit	NaN
1	Resort Hotel	0	737	2015	July	27	1	0	0	2	...	No Deposit	NaN
2	Resort Hotel	0	7	2015	July	27	1	0	1	1	...	No Deposit	NaN
3	Resort Hotel	0	13	2015	July	27	1	0	1	1	...	No Deposit	304.0
4	Resort Hotel	0	14	2015	July	27	1	0	2	2	...	No Deposit	240.0

5 rows × 32 columns

Fig 3.2.1

The above figure 3.2.1 states that the data set is loaded into our collab .And we imported our our necessary libraries **.**

```
char.isnull().sum()
```

```
hotel                              0
arrival_date_month                 0
meal                               0
country                          488
market_segment                     0
distribution_channel               0
reserved_room_type                 0
assigned_room_type                 0
deposit_type                       0
customer_type                      0
reservation_status                 0
reservation_status_date            0
arrival_date_year                  0
stays_in_weekend_nights            0
adults                             0
children                           4
babies                             0
is_repeated_guest                  0
previous_cancellations             0
required_car_parking_spaces        0
total_of_special_requests          0
dtype: int64
```

fig 3.2.2

From the above figure we have 488 null values in country and 4 null values in children.

```
char_1.isnull().sum()
```

```
hotel                          0
arrival_date_month             0
meal                           0
country                        0
market_segment                 0
distribution_channel           0
reserved_room_type             0
assigned_room_type             0
deposit_type                   0
customer_type                  0
reservation_status             0
reservation_status_date        0
arrival_date_year              0
stays_in_weekend_nights        0
adults                         0
children                       0
babies                         0
is_repeated_guest              0
previous_cancellations         0
required_car_parking_spaces    0
total_of_special_requests      0
dtype: int64
```

Fig 3.2.3

Therefore we removed the null values.

3.3 DATA AUGMENTATION

Data augmentation is a set of techniques to artificially increase the amount of data by generating new data points from existing data. This includes making small changes to data or using deep lea rning models to generate new data points.

3.4 DATA VISUALIZATION

Data visualization is the graphical representation of information and data. By using visual elements like charts, graphs, and maps, data visualization tools provide an accessible way to see and understand trends, outliers, and patterns in data.

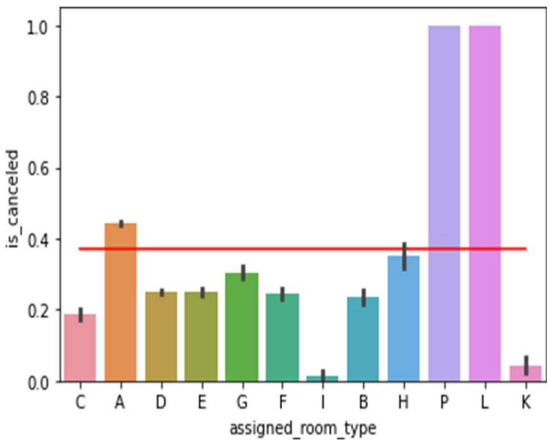


Fig.3.4.1

The above diagram about describes about types of rooms are avaliable in hotel,it visualizes about the people assigned rooms.In the hotel contains different types of rooms like C,A,D,E,G,F,I,B,etc

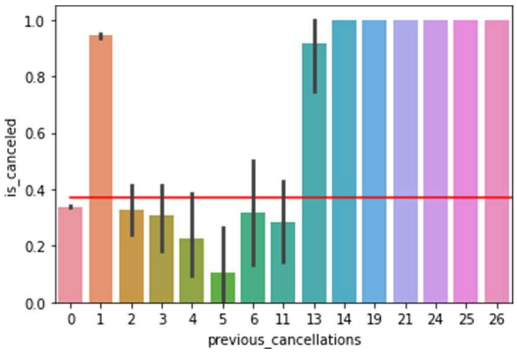


Fig 3.4.2

The above diagram visualizes about previous cancellations which key attribute of our project.our project aim is dectect previous cancellations.It shows the number of people cancelled.

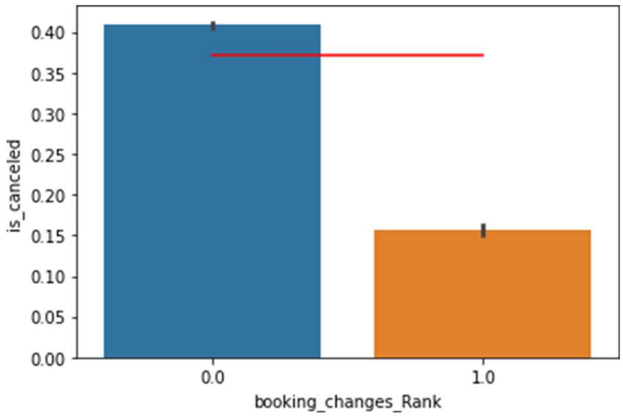


Fig 3.4.3

The above figure shows about the difference between cancelled and booking changes Number of changes/amendments made to the booking from the moment the booking was entered on the until the moment of check-in or cancellation

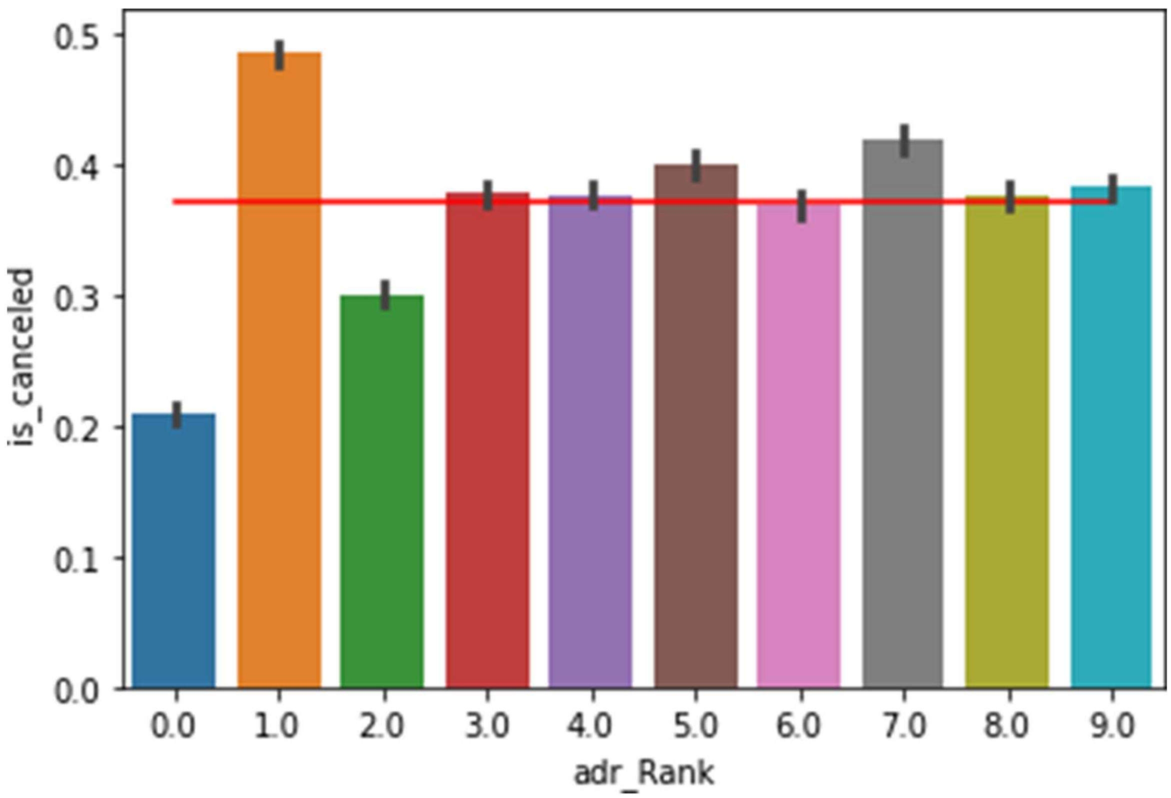


Fig 3.4.4

From above figure adr stands for average daily rate ,edr describes obout how many people using the hotel and which type of rooms they are using and the bar plot shows the difference between cancelled and adr rank

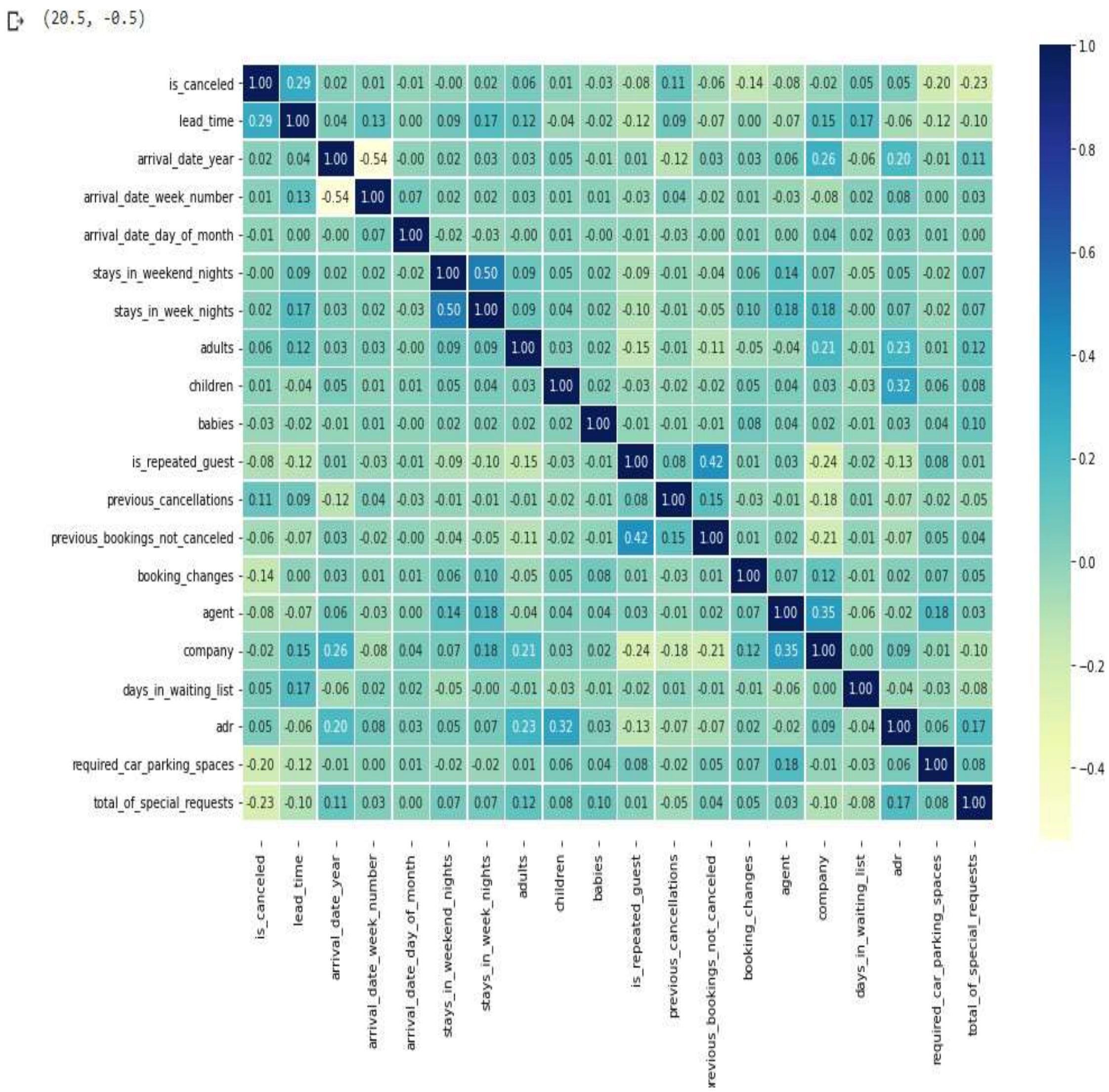


Fig 3.4.5

A correlation matrix is a table showing correlation coefficients between variables. Each cell in the table shows the correlation between two variables. A correlation matrix is used to summarizedata, as an input into a more advanced analysis, and as a diagnostic for advanced analyses.

From above figure we plotted the correlation matrix and we took the necessary measures to plot it.

We see that there are so many similar values and with this we can evaluate the model easily and visualize predominantly.

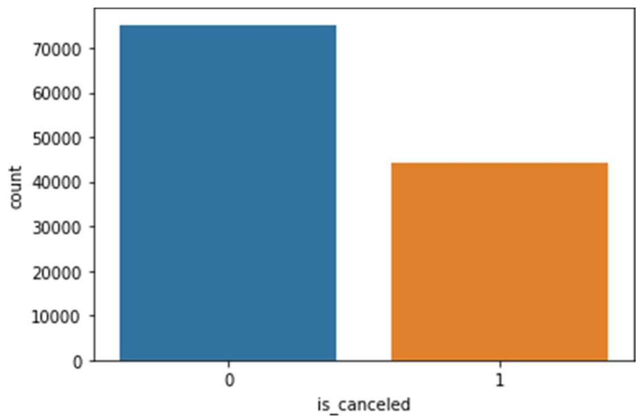


Fig 3.4.6

The above **count plot** it describes about whether the room is cancelled or not, it has two values 1 and 0. Here 1 states about cancelling of rooms and 0 states about not cancelling of rooms.

```
# Country univariant analysis only for top 10 countries
fig, axes = plt.subplots(1, 1, figsize=(16, 10))
_ = df['country'].value_counts(normalize=True).head(10).sort_values(ascending=True).plot(kind = 'barh', title='c', fontsize=14)
```

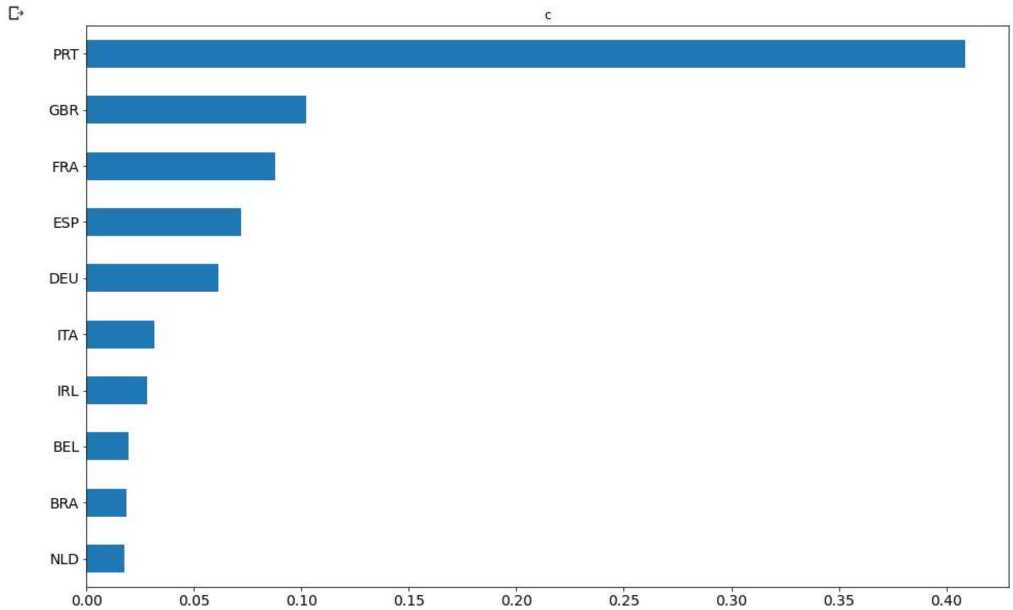


Fig 3.4.7

The above figure is country of origin it states that which country people are highest visitors of hotel and lowest visitors of hotel .

```
df.is_canceled.value_counts().plot.barh()
```

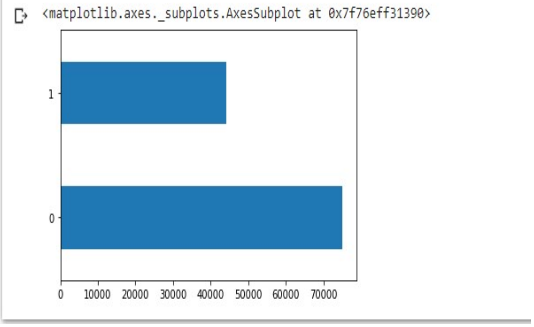


Fig 3.4.8

The above **bar plot** it describes about whether the room is cancelled or not, it has two values 1 and 0.

Here 1 states about cancelling of rooms and 0 states about not cancelling of rooms.

```
df.hist(figsize=(10,10),bins=50)
plt.show()
```

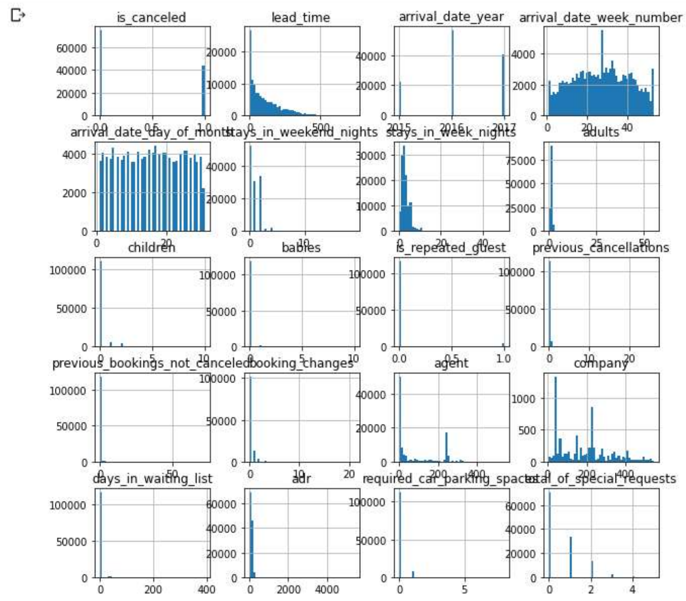


Fig.3.4.9

The above Histogram describes about variation of a bar chart in which data values are grouped together and put into different classes and there are 16 values and 16 figures showing different visualizations.

```
[102] sns.boxplot(x="agent",data=df)
```

```
<matplotlib.axes._subplots.AxesSubplot at 0x7f76db375390>
```

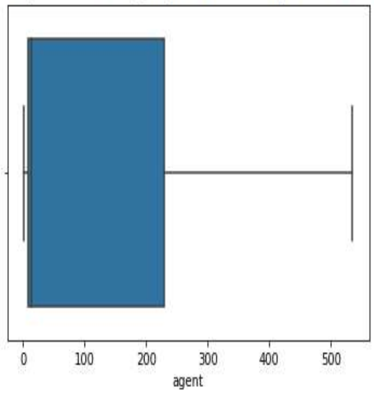


fig.3.4.10

The above box plot states that how many booking are happening through the agency and its all about 210 percent. The agency deals third party apps they are goibibo etc.

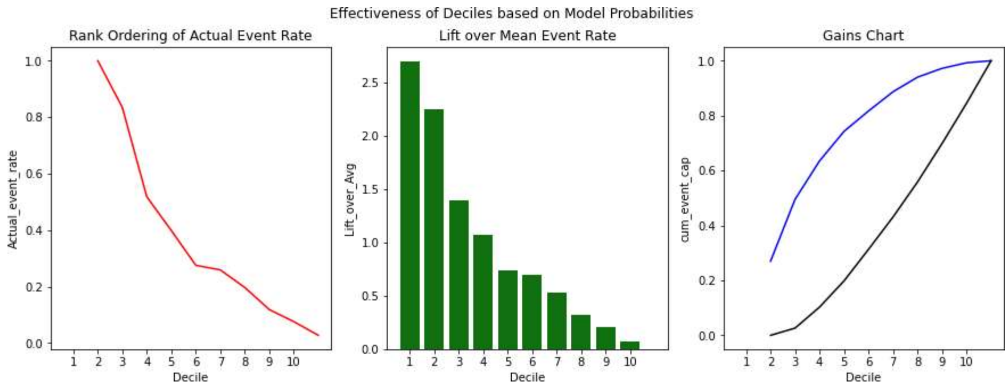


Fig.3.4.11

- Gain Chart Gain at a given decile level is the ratio of cumulative number of targets (events) up to that decile to the total number of targets (events) in the entire data set % oftargets (events) covered at a given decile level.
- Lift Chart It measures how much better one can expect to do with the predictive model comparing without a model. It is the ratio of gain % to the random expectation % at a given decile level. The random expectation at the xth decile is x%.

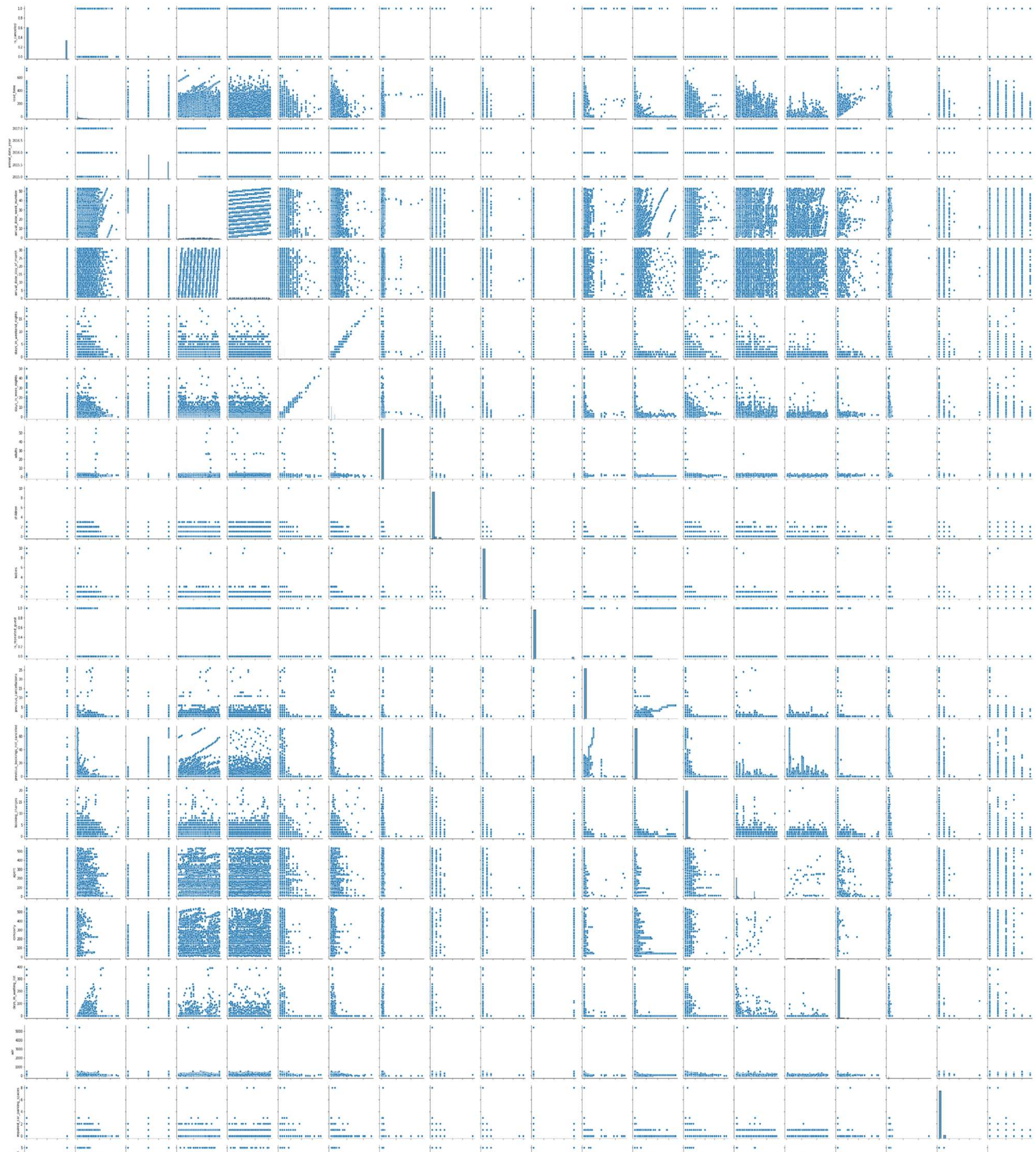

is_canceled
lead_time
arrival_date_year
arrival_date_week_number
arrival_date_day_of_month
stays_in_weekend_nights
stays_in_week_nights
adults
children
babies
is_repeated_guest
previous_cancellations
previous_bookings_not_canceled
booking_changes
agent
company
days_in_waiting_list
adr
required_car_parking_spaces
total_of_special_requests

Fig .3.4.12

Pairplot visualizes given data to find the relationship between them where the variables can be continuous or categorical. Plot pairwise relationships in a data-set. Pairplot is a module of seaborn library which provides a high-level interface for drawing attractive and informative statistical graphics.

By this pair plot hotel or resort management can visualize there issues and sort it out in a proper manner.

```
metrics.plot_confusion_matrix(dtree, X_test, y_test)
```

```
<sklearn.metrics._plot.confusion_matrix.ConfusionMatrixDisplay at 0x7f76de2b8fd0>
```

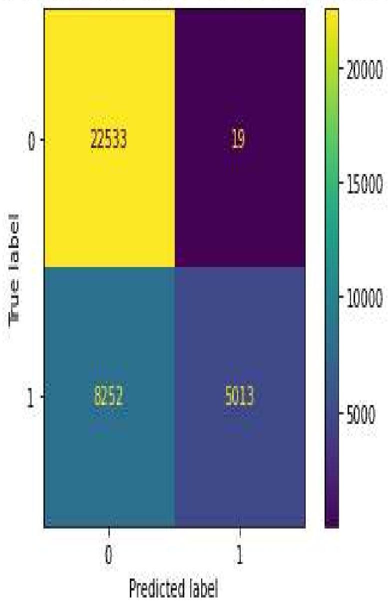


fig.3.4.13

It is a performance measurement for machine learning classification problemwhere output can be two or more classes.

The above confusion matrix using decision tree we gain 76% accuracy of the predictedmodel.

```
metrics.plot_confusion_matrix(rf, X_test,y_test)
```

```
<sklearn.metrics._plot.confusion_matrix.ConfusionMatrixDisplay at 0x7f76ddc13790>
```

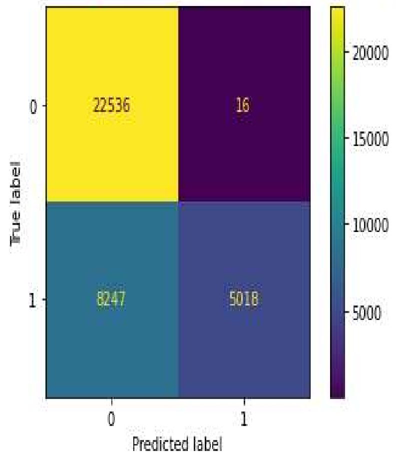


fig.3.4.14

It is a performance measurement for machine learning classification problemwhere output can be two or more classes.

The above confusion matrix using random forest we gain 76% accuracy of the predictedmodel.

4.Design

4.1 Requirement Specification (S/W &H/W):

4.1.1 Software requirements.

1. Programming language. Python.

2. Python libraries. Numpy, pandas.Scikit-learn,Matplotlib,Seaborn

3. Integrated development environment (IDE): Google colab

4. Data storage : CSV format.

4.1.2 Hardware requirements.

1. Processor: Intel i5 or higher.

2 Ram: 8GB are higher.

3 storage: 50GB or higher

4.2 UML Diagrams:

4.2.1 UML Class diagram:

```
+-------------------+          +------------------+          +------------------+
|    Reservation    |          |     Customer     |          |      Room        |
+-------------------+          +------------------+          +------------------+
| -status: String   |          | -type: String    |          | -type: String     |
| -statusDate: Date |          | -adr: double     |          | -price: double    |
| -waitListDays: int|          | -country: String |          | -isAvailable: bool|
| -reqParking: int  |          | -babies: int     |          | -reservedType: Str|
| -reqSpecial: int  |          | -agent: int      |          | -assignedType: Str|
| -isRepeatGuest:bool|         | -company: int    |          |                  |
| -prevCancel: int  |          |                  |          |                  |
| -prevBookNotCancel: int|     |                  |          |                  |
| -bookChanges: int |          |                  |          |                  |
| -depositType: String |       |                  |          |                  |
| -meal: String     |          |                  |          |                  |
| -marketSegment: String |     |                  |          |                  |
| -distChannel: String|        |                  |          |                  |
| -canceled: bool   |          |                  |          |                  |
| -leadTime: int    |          |                  |          |                  |
| -arrivalYear: int |          |                  |          |                  |
| -arrivalMonth: int|          |                  |          |                  |
| -arrivalWeek: int |          |                  |          |                  |
| -arrivalDay: int  |          |                  |          |                  |
| -weekendNights: int|         |                  |          |                  |
| -weekNights: int  |          |                  |          |                  |
+-------------------+          +------------------+          +------------------+
```

The above uml diagram states about "Booking," as all of the attributes provided pertain to a hotel booking and do not require any additional classes.

4.2.2 UML Activity Diagram:

```
+-------------------------+
|    Hotel Booking Demand  |
+-------------------------+
|                         ||
| + Country               |
| + Market Segment        |
| + Distribution Channel  |
| + Lead Time             |
| + Arrival Date          |
| + Stays in Weekends     |
| + Stays in Week Nights  |
| + Adults                |
| + Children              |
| + Babies                |
| + Meal                  |
| + Reserved Room Type    |
| + Assigned Room Type    |
| + Deposit Type          |
| + Agent                 |
| + Company               |
| + Required Parking      |
| + Special Requests      |
| + Previous Cancellations|
| + Previous Bookings     |
| + Days in Waiting List  |
| + Reservation Status    |
| + Reservation Status Date|
| + Is Repeated Guest     |
| + ADR                   |
| + Total of Guests       |
| + Is Canceled           |
|                         |
+-------------------------+
          |
          |
          v
+-------------------------+
|    Hotel Booking System |
+-------------------------+
|                         |
| -Hotel                   |
| -Booking                 |
| -Guest                   |
| -Payment                 |
|                         |
+-------------------------+
          |
          |
          v
+-------------------------+
|    Reservation Process  |
+-------------------------+
|                         |
| + Check Availability    |
| + Select Room           |
| + Confirm Booking       |
| + Take Payment          |
| + Issue Receipt         |
| + Update Reservation Status |
|                         |
+-------------------------+
```

The above diagram states about uml activity diagram

5.Implementation

5.1 Modules:

First we loaded the dataset into the google collab and upload to drive . uploaded the path of our csv file in collab. And we performed the basic methods and feature engineering and split into categorical into numerical ,found the outliers and completed the data pre processing and imputing the variables . we did feature selection then moved to bi variate analysis also called feature discretization and completed data visualization ploted different types of graphs and did the modelling of classification of algorithms

5.2 Overview Technology:

Python is a high-level, general-purpose programming language. Its design philosophyemphasizes code readability with the use of significant indentation. Python is dynamically-typed and garbage-collected. It supports multiple programming paradigms,including structured (particularly procedural), object-oriented and functional programming. It is often described as a "batteries included" language due to its comprehensive standard library

Colaboratory, or “Colab” for short, is a product from Google Research. Colab allows anybody to write and execute arbitrary python code through the browser, and is especially well suited to machine learning, data analysis and education. More technically, Colab is a hosted Jupyter notebook service that requires no setup to use, while providing access free of charge to computing resources including GPUs.

6.RESULTS

The goal of our hotel booking demand is to predict whether the booking will be canceledOr not.

By using classification random forest algorithm we got

Random Forest Metrics
Accuracy: 0.7692994946533769
Precision: 0.9968216130313866
Recall: 0.37828872973991706

f1_score: 0.5484452702333461

By using classification decision tree algorithm we got

Decision Tree Metrics
Accuracy: 0.769076137029902
Precision 0.9962241653418124
Recall 0.3779117979645684

f1_score 0.5479586817511067

7.Conclusion and Future Scope

Therefore we predicted our model good accuracy and we visualized with different graphes to understand. we cleaned and preprocessed the data and then we performed the exploratory data analysis to extract information from the data.

We want to build an application with better ui /ux design

8.Bibilography

https://github.com/SAI52094

9.References

[A]https://www.researchgate.net/publication/228254814_Hotel_Revenue_Management_-_A_Critical_Literature_Review

[B]https://scialert.net/fulltext/?doi=jas.2014.341.347

[c]https://www.scribd.com/document/521448466/A-LITERATURE-REVIEW-ABOUT-ONLINE-HOTEL-BOOKING-SYSTEM

[D]https://www.sciencedirect.com/science/article/abs/pii/S0969698918310282#

[E]https://www.redalyc.org/journal/289/28966251006/html/

[F]https://www.studymode.com/subjects/literature-review-for-a-hotel-management-system-page1.html

[G]https://www.academia.edu/37649550/HOTEL_MANAGEMENT_INFORMATION_SYSTEM_PROJECT_FINAL

[H] https://www.ipl.org/essay/Literature-Review-On-Hotel-Management-F3NU2X36C486

[I] https://www.studocu.com/en-gb/document/the-robert-gordon-university/research-methods/literature-review-on-hotel-industry-and-service-quality/1629975

[j]https://www.pdfprof.com/PDF_Image.php?idt=92787&t=39

[k]https://www.academia.edu/36435770/Chapter_2_REVIEW_OF_RELATED_LITERATURE_AND_STUDIES_This_chapter_presents_the_review_of_related_literature_and_studies_particularly_on_the_Proposed_Computerized_Hotel_Management_System_of