PROFESSIONAL TRAINING REPORT

entitled

EDA ANALYSIS ON HOTEL BOOKING DATASET

Submitted in partial fulfillment of the requirements for the award of Bachelor of Engineering degree in Computer Science and Engineering

by

POLISETTY NARENDRA KUMAR

41731091

DEPARTMENT OF COMPUTER SCIENCE AND ENGINEERING
SCHOOL OF COMPUTING

SATHYABAMA

INSTITUTE OF SCIENCE AND TECHNOLOGY
(DEEMED TO BE UNIVERSITY)
Accredited with Grade "A++" by NAAC
JEPPIAAR NAGAR, RAJIV GANDHISALAI,
CHENNAI – 600119

OCTOBER 2023

SATHYABAMA

INSTITUTE OF SCIENCE AND TECHNOLOGY

(DEEMED TO BE UNIVERSITY)

Accredited with A++ Grade by NAAC

Jeppiaar Nagar, Rajiv Gandhi Salai,

Chennai – 600 119

www.sathyabama.ac.in

DEPARTMENT OF COMPUTER SCIENCE AND ENGINEERING

BONAFIDE CERTIFICATE

This is to certify that this Professional Training is the bonafide work of polisetty Narendra kumar who carried out the project entitled “EDA analysis on hotel booking dataset” under my supervision from June 2023 to October 2023.

Internal Guide
Mrs.YUGHA R

Head of the Department
Dr.S.VIGNESHWARI, M.E., Ph.D .,

Submitted for Viva voce Examination held on ____________________

Internal Examiner **External Examiner**

DECLARATION

I polisetty Narendra kumar(41731091), hereby declare that the Professional Training Report-I entitled "EDA on analysis on hotel booking dataset" done by me under the guidance of **Mrs.YUGHA R** submitted in partial fulfilment of the requirements for the award of Bachelor of Engineering degree in Computer Science and Engineering.

DATE:

PLACE: **SIGNATURE OF THE CANDIDATE**

ACKNOWLEDGEMENT

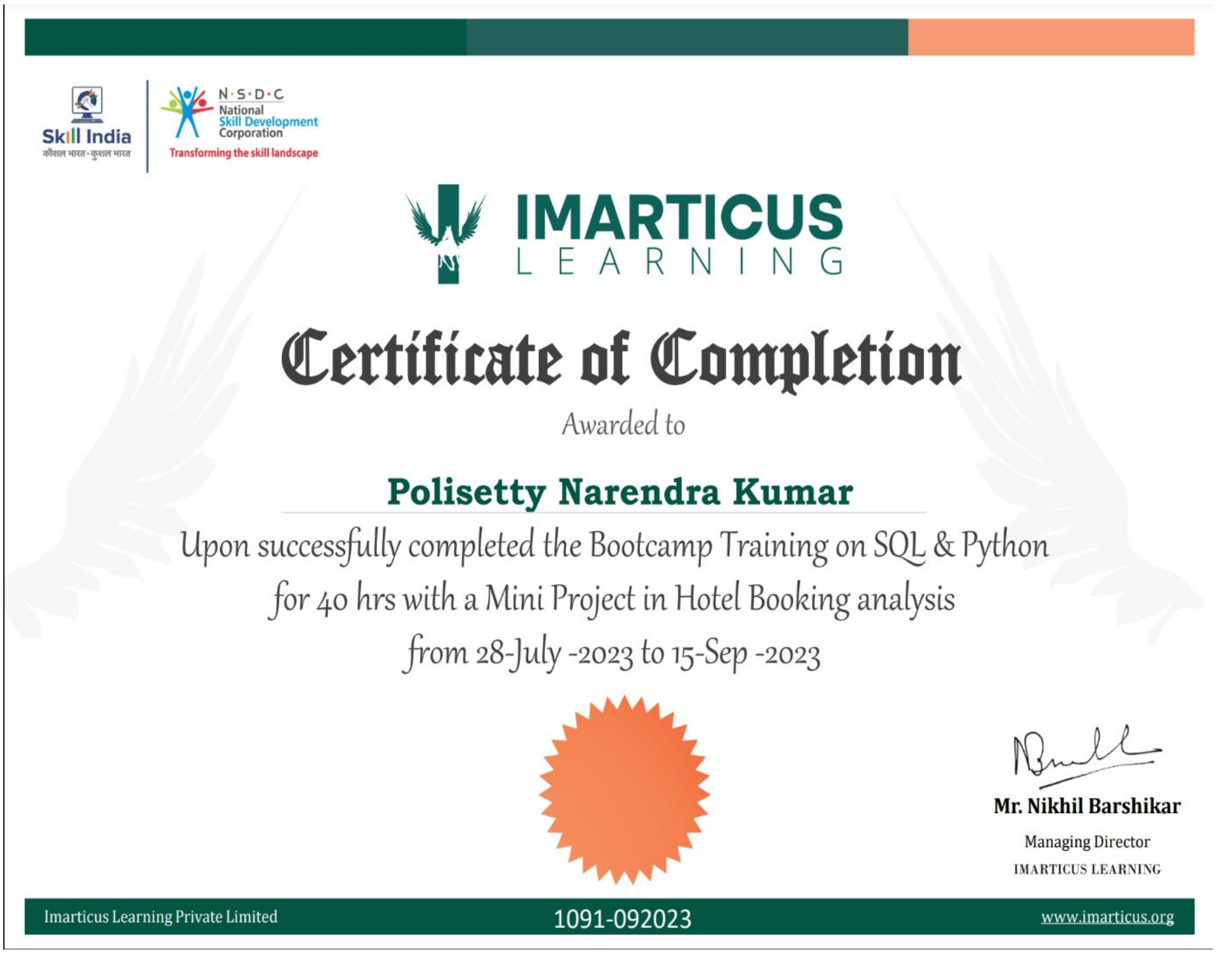
I am pleased to acknowledge my sincere thanks to **Board of Management** of **SATHYABAMA** for their kind encouragement in doing this project and for completing it successfully. I am grateful to them.

I convey my thanks to **Dr. T.Sasikala M.E., Ph.D.**, **Dean**, School of Computing, **Dr.S.Vigneshwari M.E., Ph.D., Head of the Department of Computer Science and Engineering** for providing me necessary support and details at the right time during the progressive reviews.

I would like to express my sincere and deep sense of gratitude to my Internal Guide **Mrs.YUGHA R** for her valuable guidance, suggestions and constant encouragement which paved way for the successful completion of my phase-1 professional Training.

I wish to express my thanks to all Teaching and Non-teaching staff members of the **Department of Computer Science and Engineering** who were helpful in many ways for the completion of the project.

SAMPLE COURSE CERTIFICATE



N·S·D·C National Skill Development Corporation

Transforming the skill landscape

IMARTICUS LEARNING

Certificate of Completion

Awarded to

Polisetty Narendra Kumar

Upon successfully completed the Bootcamp Training on SQL & Python for 40 hrs with a Mini Project in Hotel Booking analysis from 28-July -2023 to 15-Sep -2023

Mr. Nikhil Barshikar

Managing Director

IMARTICUS LEARNING

Imarticus Learning Private Limited

1091-092023

www.imarticus.org

ABSTRACT

Hotel booking analysis involves the systematic examination of various facets related to hotel reservations and occupancy. It encompasses the study of booking patterns, channels, lead times, cancellation rates, and revenue management. Understanding these aspects enables hotels to make data-driven decisions, optimize pricing strategies, and improve customer experiences. One crucial feature of hotel booking analysis is customer segmentation, which allows hotels to personalize their services and marketing efforts. By categorizing guests based on demographics, preferences, and booking behaviors, hotels can tailor their offerings, thereby enhancing guest satisfaction and loyalty. The digital age has ushered in an era of online reviews and ratings. Hotel booking analysis extends to monitoring and analyzing guest feedback on platforms such as TripAdvisor and Yelp. This feedback loop empowers hotels to identify areas for improvement, respond promptly to guest concerns, and maintain a positive online reputation. Competitor analysis is another vital component, enabling hotels to benchmark their performance against rivals. By analyzing competitor pricing, occupancy rates, and customer reviews, hotels can uncover opportunities for competitive advantage. Predictive analytics plays a significant role in hotel booking analysis. It helps forecast future booking trends, enabling hotels to proactively adjust staffing, marketing strategies, and inventory management to meet anticipated demand. Data visualization tools are essential in presenting complex booking data in a comprehensible manner. Visualizations like charts, graphs, and dashboards provide clear insights into booking patterns, helping hotels make informed decisions. he hospitality industry is constantly evolving, driven by changing consumer preferences, market dynamics, and technological advancements. In this context, hotel booking analysis emerges as a vital practice for hotels and accommodation providers worldwide. This abstract provides a concise overview of the significance, key features, and outcomes of hotel booking analysis.

ABSTRACT

TABLE OF CONTENTS

LIST OF FIGURES

CHAPTER 1

INTRODUCTION

1.1 OVERVIEW

When running a successful and demanding hospitality business, most hotel owners like a hotel that is running at full capacity and bringing in sizeable revenue. Most of the time hotel booking cancellations can be hurtful to business owners; although sometimes there are genuine reasons for guests to do so. These last-minute cancellations can result in lost revenue unless some measures are undertaken to mitigate the loss. The purpose of this project is to analyze Hotel Bookings data, investigate cancellations, and their underlying patterns; and suggest measures that can be implemented to reduce cancellations and secure revenue.

As per an written on Booking.com, the first thing that hotel owners can do is to take a closer look at their property's specific cancellation patterns and understand guest behavior patterns. Backed by this research Benjamin Verot suggests in his article some steps that owners can execute while setting up a robust cancellation policy.

- Requiring credit/debit card deposits
- Using length of stay restrictions
- Offering low rates/discounts for direct bookings
- Adopting a cautious overbooking strategy

CHAPTER 2

LITERATURE REVIEW

2.1 SURVEY

Hotel pricing strategies have been a common focus of research. This includes dynamic pricing, where hotels adjust their room rates based on factors like demand, time to arrival, and competitor pricing.

Studies have examined how dynamic pricing impacts consumer behavior and hotel revenue management.

Loyalty programs offered by hotels have been examined for their impact on customer retention and repeat bookings. Studies often assess the effectiveness of loyalty rewards in encouraging brand loyalty.

Research may also explore the influence of loyalty program features, such as points, discounts, and exclusive benefits.

The COVID-19 pandemic has significantly disrupted the hotel industry. Studies conducted during the pandemic assess how traveler behavior, booking patterns, and preferences have evolved in response to the crisis.

Emerging technologies such as artificial intelligence (AI) and chatbots are being explored for enhancing the hotel booking experience. AI-driven personalization and chatbot-assisted bookings are areas of interest.

Remember that the literature on hotel booking analysis is continuously evolving, and new research findings may have emerged since my last knowledge update in September 2021. Researchers and practitioners in this field should stay current with the latest developments and trends to make informed decisions and improve hotel booking systems and strategies.

CHAPTER 3

REQUIREMENTS ANALYSIS

3.1 OBJECTIVE OF THE PROJECT

Predict future booking demand by analyzing historical booking data, seasonal trends, and external factors (e.g., holidays, events) to optimize room availability and pricing strategies. Segment customers based on booking behavior, demographics, and preferences to tailor marketing efforts and services for different customer groups effectively. Analyze the booking process, user behavior on the hotel's website or booking platform, and customer feedback to identify and address pain points, ultimately increasing the conversion rate. Analyze guest reviews, ratings, and feedback to identify areas for improvement in guest satisfaction and service quality. Optimize room allocation across various booking channels, ensuring efficient utilization of available inventory while avoiding overbooking and underbooking. Assess the impact of loyalty programs on customer retention, repeat bookings, and overall revenue. Identify strategies to enhance program effectiveness. Analyze the influence of sustainability practices and eco-friendly certifications on booking decisions. Identify opportunities to attract environmentally conscious travelers. Monitor booking cancellation trends and implement strategies to mitigate revenue loss due to cancellations and no-shows. Ensure compliance with relevant legal and regulatory requirements in the hotel booking process, particularly data protection and privacy regulations. Assess the profitability of different booking channels, distribution partners, and marketing initiatives to optimize the allocation of resources and marketing budgets. Develop strategies for long-term growth and sustainability by analyzing market trends and customer preferences. Identify strategies to retain existing customers and encourage repeat bookings, such as personalized offers and loyalty programs. Evaluate the efficiency of internal processes related to hotel bookings, including reservations, check-in, and customer support.

3.2 REQUIREMENTS

- Data Source
- Data Cleaning Tools
- Data visualization tools
- Statistical Software
- Documentation
- Data Privacy and Security Measures
- Domain knowledge

3.2.1 *HARDWARE REQUIREMENTS*

- Operating System : Windows 8,Windows 10, Mac OS
- Memory : 8gb ram
- Processor : Intel Core i5
- Hard-Disk : Minimum 30gb

3.2.2 *SOFTWARE REQUIREMENTS*

- Software : 3.11.4
- IDE : Jupyter

CHAPTER 4

DESIGN DESCRIPTION OF PROPOSED PROJECT

4.1.1 PROPOSED METHODOLOGY

Designing a system for hotel booking analysis involves defining the architecture, components, and processes needed to gather, store, analyze, and visualize data effectively. Here's a proposed system for hotel booking analysis

Data Sources: Gather data from various sources, including your hotel's booking database, online travel agencies (OTAs), customer reviews, and external datasets (e.g., weather data, local events).

Data Cleaning: Implement automated processes to handle missing values, outliers, and data quality issues.

Feature Engineering: Create new features or transform existing ones to enhance the dataset's quality and usefulness for analysis.

Data Visualization: Utilize data visualization tools and libraries (e.g., Matplotlib, Seaborn, Tableau) to perform EDA and gain insights into the data.

Descriptive Statistics: Calculate summary statistics to understand data distributions and characteristics.

Model Development: If predictive modeling is part of your analysis, train machine learning models (e.g., regression, classification) to address specific research questions (e.g., demand forecasting, price optimization).

Model Evaluation: Assess model performance using appropriate metrics and cross-validation techniques.

4.1.2 *Various Stages*

Stage 1 : Data collection

Gather data from various sources, such as websites, mobile apps, reservation systems, and customer feedback.

Collect information on booking dates, room types, prices, guest demographics, and more.

Stage 2 : Data Cleaning and Preprocessing

Clean and prepare the collected data by removing duplicates, handling missing values, and ensuring data consistency.

Transform data into a suitable format for analysis, including data normalization and feature engineering.

Stage 3 : Exploratory Data Analysis (EDA)

Conduct exploratory analysis to gain insights into the data.

Visualize key metrics, trends, and patterns in booking data using graphs, charts, and statistical techniques.

Identify outliers and anomalies that may require further investigation.

Stage 4 : Demand Forecasting

Use historical booking data to forecast future demand for rooms.

Utilize time series analysis, regression models, or machine learning algorithms to make accurate predictions.

Stage 5 : Pricing Optimization

Analyze pricing strategies and revenue management techniques to maximize profitability.

Adjust room rates dynamically based on demand, seasonality, and competitive pricing.

Stage 6 : Customer Sentiment Analysis

Analyze customer reviews and feedback to understand their satisfaction levels.

Use sentiment analysis and natural language processing (NLP) to extract insights from textual data.

Stage 7 : Booking Funnel Analysis

Analyze the various stages of the booking process, from search to confirmation.

Identify drop-off points and bottlenecks in the booking funnel to optimize the user experience.

Stage 8 : Reporting and Visualization

Create reports and dashboards to communicate findings and insights to stakeholders.

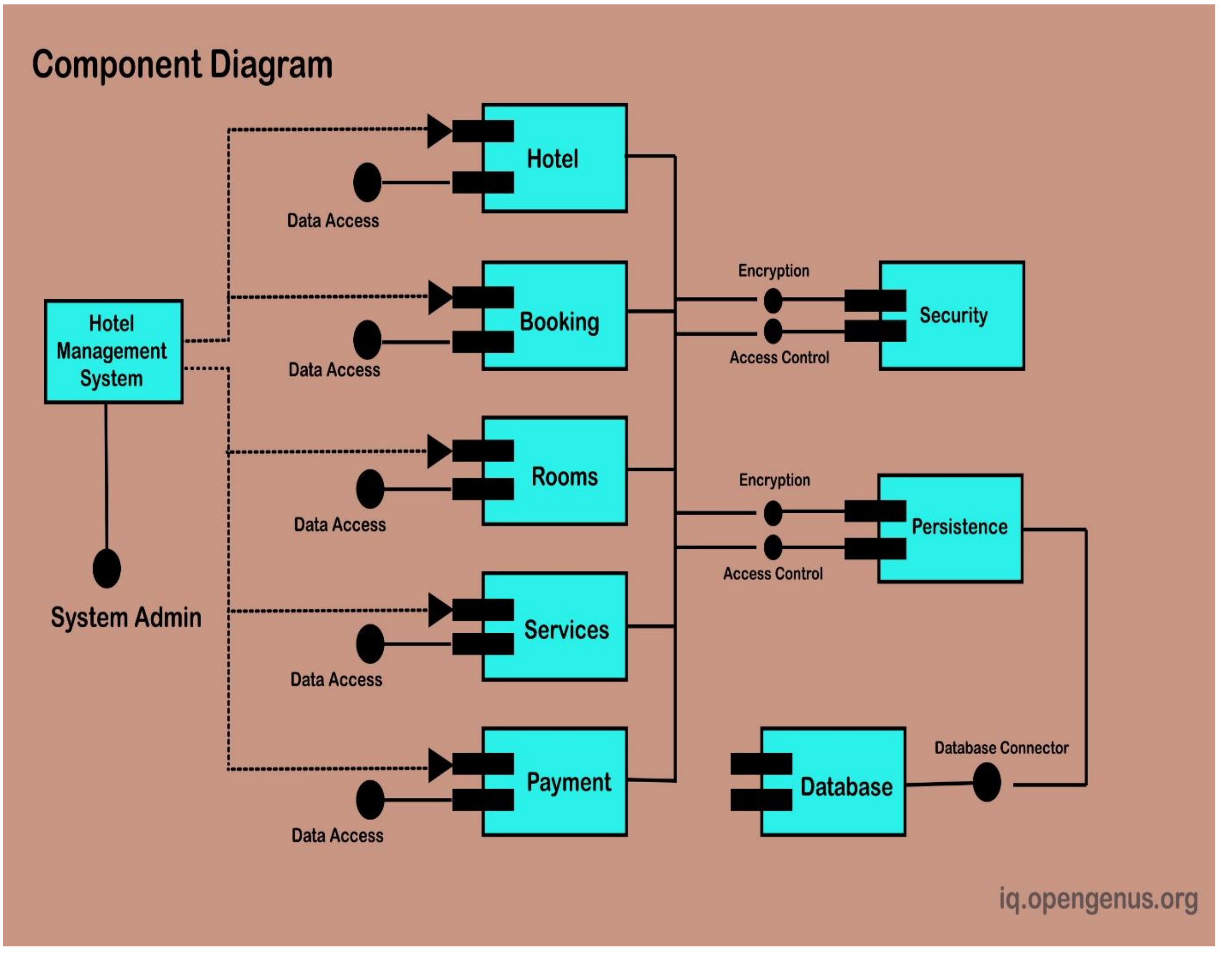
Use data visualization tools to present information in a clear and actionable format.

Stage 9 : Compliance and Data Security

Ensure that all data handling and analysis comply with data privacy regulations (e.g., Protect sensitive customer information.

4.1.3 *Internal or Component design structure*

Figure 4.1.3 :



This diagram depicts the components, available and required interfaces, ports, and linkages between Services, Booking, Rooms, Hotel, and Customers in a Hotel ManagementSystem.This sort of graphic is used to represent systems using Service Oriented Architecture (SOA) in Component-Based Development (CBD). The UML component diagram for a Hotel Management System depicts the organization and wiring of physical components in a system.

4.1.4 Working principles :

Figure 4.1.4.1: Importing libraries

```
jupyter hotel booking analysis Last Checkpoint: 15/09/2023 (autosaved)
File Edit View Insert Cell Kernel Widgets Help      Not Trusted | Python 3 (ipykernel)

In [2]: #Importing Libraries
        import pandas as pd
        import matplotlib.pyplot as plt
        import seaborn as sns
        import warnings
        import seaborn as sns
        warnings.filterwarnings('ignore')
```

Here we import the libraries like pandas, matplotlib, seaborn

Figure 4.1.4.2: Loading the data set

```
In [4]: #Loading the Dataset
        df = pd.read_csv(r"C:\Users\karan\Downloads\hotel_bookings.csv")
```

here we load the hotel booking dataset

Figure 4.1.4.3: Dealing with the missing values

```
In [13]: df.isnull().sum()

Out[13]: hotel                                0
         is_canceled                          0
         lead_time                            0
         arrival_date_year                    0
         arrival_date_month                   0
         arrival_date_week_number             0
         arrival_date_day_of_month            0
         stays_in_weekend_nights              0
         stays_in_week_nights                 0
         adults                               0
         children                             4
         babies                               0
         meal                                 0
         country                            488
         market_segment                       0
         distribution_channel                 0
         is_repeated_guest                    0
         previous_cancellations               0
         previous_bookings_not_canceled       0
         reserved_room_type                   0
         assigned_room_type                   0
         booking_changes                      0
         deposit_type                         0
         agent                            16340
         company                         112593
         days_in_waiting_list                 0
         customer_type                        0
         adr                                  0
         required_car_parking_spaces          0
         total_of_special_requests            0
         reservation_status                   0
         reservation_status_date              0
         dtype: int64
```

We have 3 features with missing values. so for all the missing values, we will just replace it with 0.

Figure 4.1.4.4: Non-null count and data types

```
In [14]: df.info()

<class 'pandas.core.frame.DataFrame'>
RangeIndex: 119390 entries, 0 to 119389
Data columns (total 32 columns):
 #   Column                          Non-Null Count   Dtype
---  ------                          --------------   -----
 0   hotel                           119390 non-null  object
 1   is_canceled                     119390 non-null  int64
 2   lead_time                       119390 non-null  int64
 3   arrival_date_year               119390 non-null  int64
 4   arrival_date_month              119390 non-null  object
 5   arrival_date_week_number        119390 non-null  int64
 6   arrival_date_day_of_month       119390 non-null  int64
 7   stays_in_weekend_nights         119390 non-null  int64
 8   stays_in_week_nights            119390 non-null  int64
 9   adults                          119390 non-null  int64
 10  children                        119386 non-null  float64
 11  babies                          119390 non-null  int64
 12  meal                            119390 non-null  object
 13  country                         118902 non-null  object
 14  market_segment                  119390 non-null  object
 15  distribution_channel            119390 non-null  object
 16  is_repeated_guest               119390 non-null  int64
 17  previous_cancellations          119390 non-null  int64
 18  previous_bookings_not_canceled  119390 non-null  int64
 19  reserved_room_type              119390 non-null  object
 20  assigned_room_type              119390 non-null  object
 21  booking_changes                 119390 non-null  int64
 22  deposit_type                    119390 non-null  object
 23  agent                           103050 non-null  float64
 24  company                         6797 non-null    float64
 25  days_in_waiting_list            119390 non-null  int64
 26  customer_type                   119390 non-null  object
 27  adr                             119390 non-null  float64
 28  required_car_parking_spaces     119390 non-null  int64
 29  total_of_special_requests       119390 non-null  int64
 30  reservation_status              119390 non-null  object
 31  reservation_status_date         119390 non-null  datetime64[ns]
dtypes: datetime64[ns](1), float64(4), int64(16), object(11)
memory usage: 29.1+ MB
```

We can see different data types for different columns.

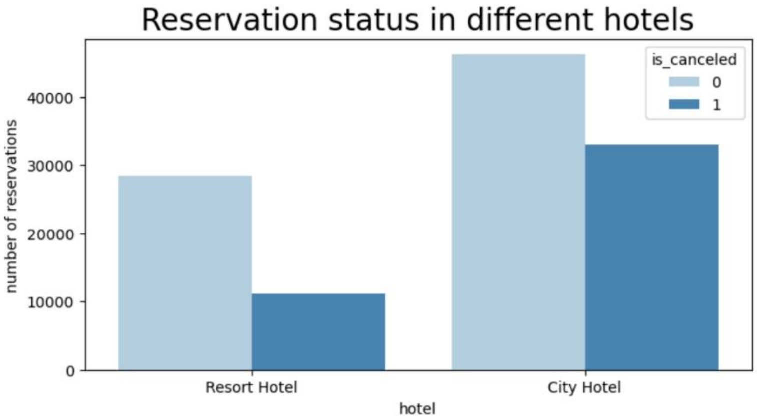
Figure 4.1.4.5: Cancelled and Non-Cancelled

```
In [16]: cancelled_perc = df['is_canceled'].value_counts(normalize = True)
         print(cancelled_perc)

         plt.figure(figsize = (5,4))
         plt.title('Reservation status count')
         plt.bar(['Not canceled','canceled'],df['is_canceled'].value_counts(), edgecolor = 'k',width = 0.7)
         plt.show()

         0    0.628648
         1    0.371352
         Name: is_canceled, dtype: float64
```

Booking got canceled 37% of the time. While booking guests did checked_in(did not canceled booking) about 63% of the time.

Figure 4.1.4.6: Booking ratio of resort hotel vs city hotel

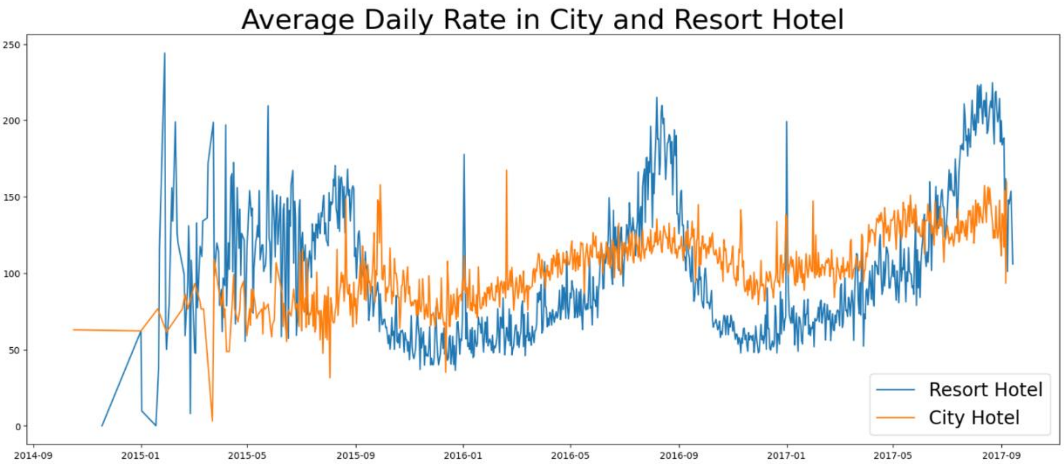
```
In [17]: plt.figure(figsize = (8,4))
         ax1= sns.countplot(x = 'hotel', hue = 'is_canceled',data = df, palette = 'Blues')
         legend_labels,_ = ax1. get_legend_handles_labels()
         plt.title('Reservation status in different hotels',size = 20)
         plt.xlabel('hotel')
         plt.ylabel('number of reservations')

Out[17]: Text(0, 0.5, 'number of reservations')
```



More than 60% of the people booked the city hotel than compare to the resort hotel

Figure 4.1.4.7: Average daily rate in city hotel and resort hotel

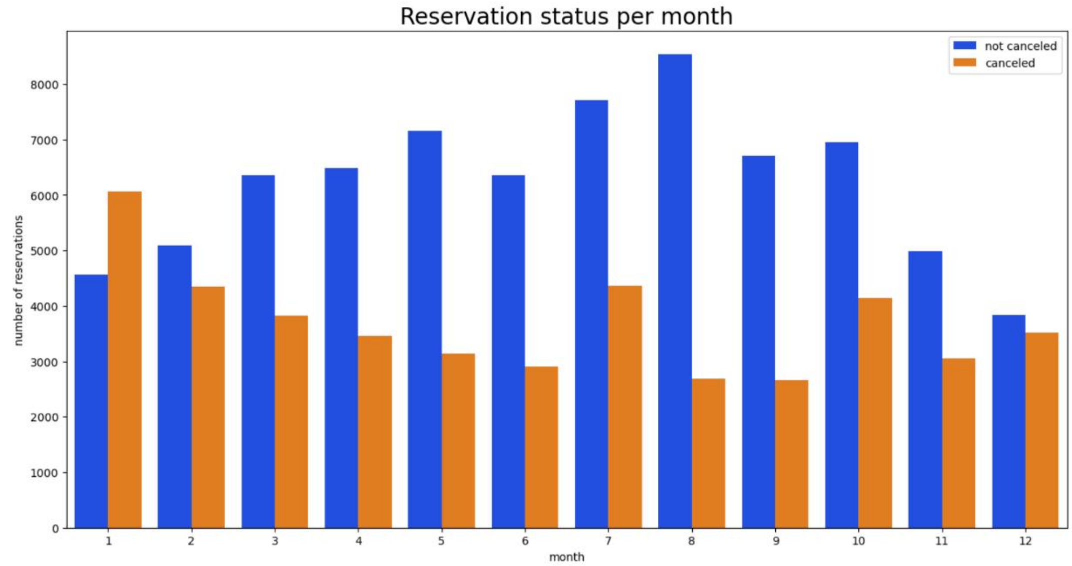
```
In [21]: plt.figure(figsize = (20,8))
         plt.title('Average Daily Rate in City and Resort Hotel', fontsize = 30)
         plt.plot(resort_hotel.index,resort_hotel['adr'], label = 'Resort Hotel')
         plt.plot(city_hotel.index,city_hotel['adr'], label = 'City Hotel')
         plt.legend(fontsize = 20)
         plt.show()
```



In the figure 4.1.4.7 we show the average daily rate in city and resort hotel from the year 2014 -17

Figure 4.1.4.8: Reservation status per month

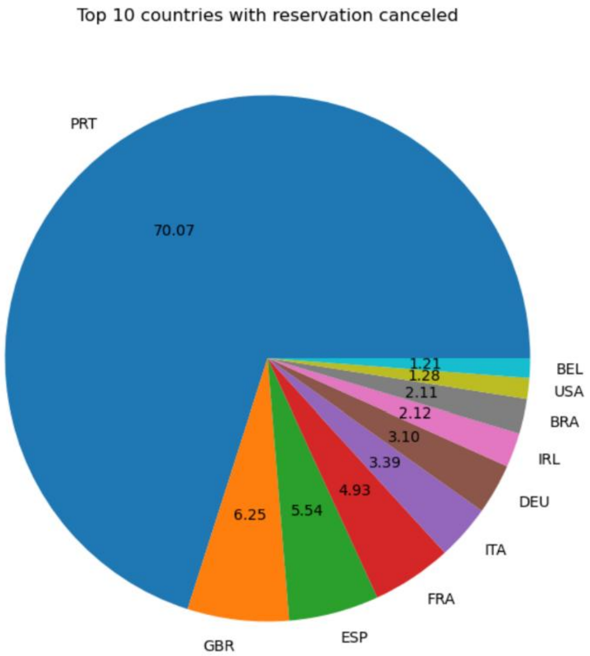
```
In [22]: df['month'] =df['reservation_status_date'].dt.month
         plt.figure(figsize = (16,8))
         ax1 = sns.countplot(x = 'month',hue = 'is_canceled',data = df,palette = 'bright')
         legend_labels,_ = ax1. get_legend_handles_labels()
         plt.title('Reservation status per month',size = 20)
         plt.xlabel('month')
         plt.ylabel('number of reservations')
         plt.legend(['not canceled','canceled'])
         plt.show()
```



The above figure shows the reservation status according to the month. Here blue color represents the not cancelled and orange color represents the cancelled.

Figure 4.1.4.9: Top 10 countries with reservation cancelled

```
In [24]: cancelled_data = df[df['is_canceled'] == 1]
         top_10_country = cancelled_data['country'].value_counts()[:10]
         plt.figure(figsize = (8,8))
         plt.title('Top 10 countries with reservation canceled')
         plt.pie(top_10_country,autopct = '%.2f',labels = top_10_country.index)
         plt.show()
```



The above figure shows the top 10 countries with reservation cancelled. The highest reservation canceled status is portugul and the lowest is Belgium.

Figure 4.1.4.10: Market segments

```
In [25]: df['market_segment'].value_counts()

Out[25]: Online TA        56402
         Offline TA/TO    24160
         Groups           19806
         Direct           12448
         Corporate         5111
         Complementary      734
         Aviation           237
         Name: market_segment, dtype: int64
```

The above figure shows about the booking platforms like online TA, offline TA, groups,direct,corporate and aviation

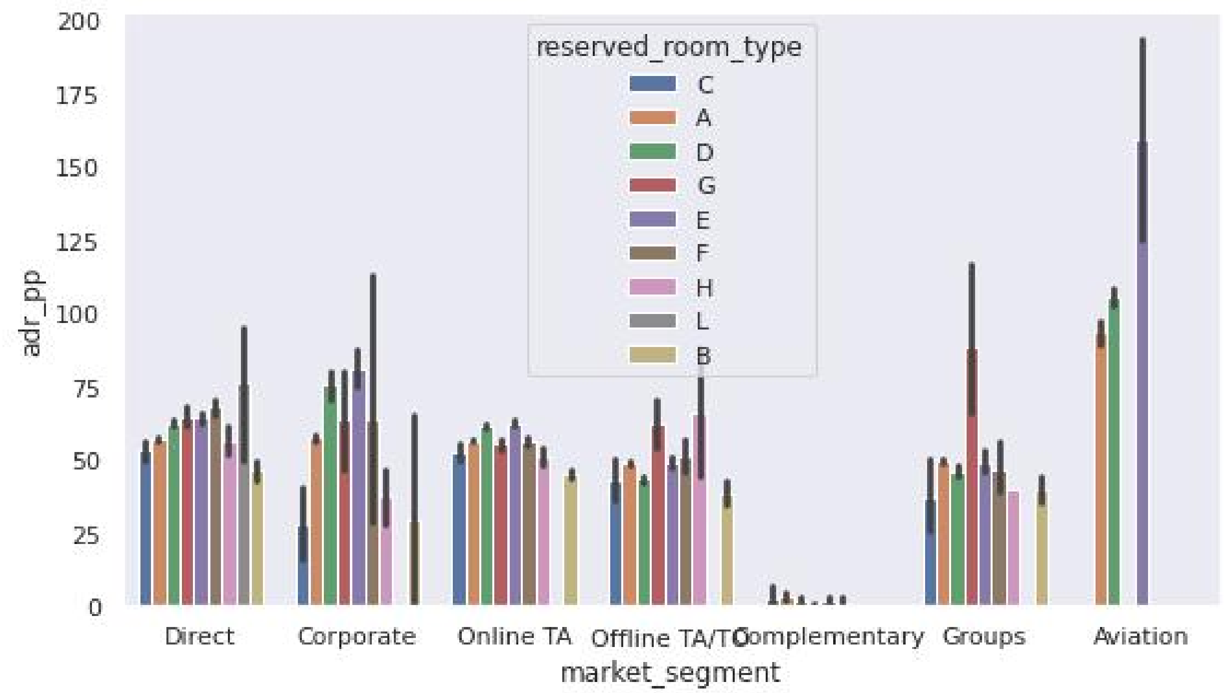
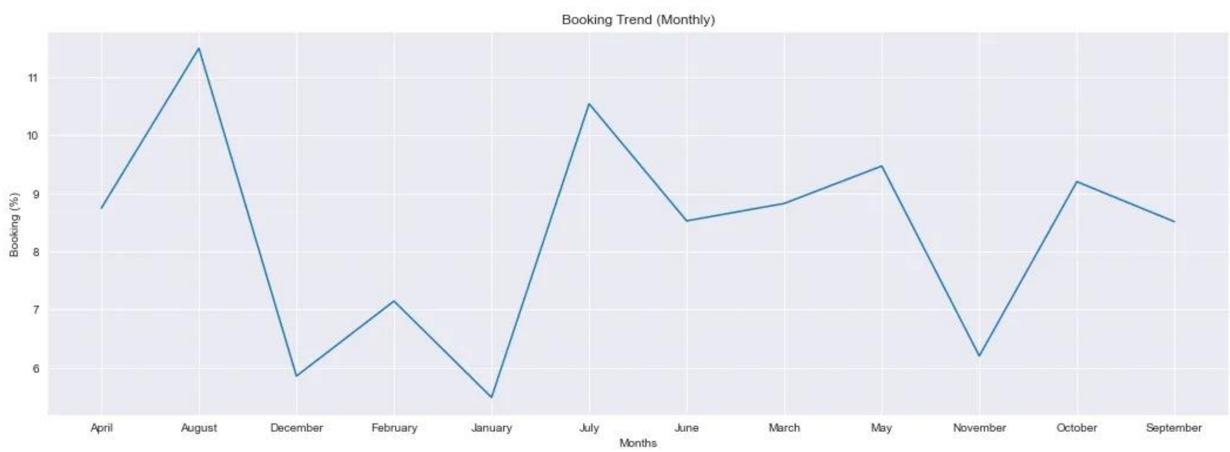


Figure 4.1.4.11: Busiest month for hotels

```
In [27]: new_order = ['January', 'February', 'March', 'April', 'May', 'June', 'July', 'August', 'September',
                      'October', 'November', 'December']
         sorted_months = df_not_canceled['arrival_date_month'].value_counts().reindex(new_order)

         x = sorted_months.index
         y = sorted_months/sorted_months.sum()*100


         #sns.lineplot(x, y.values)
         plot(x, y.values, x_label='Months', y_label='Booking (%)', title='Booking Trend (Monthly)', type='line', figsize=(18,6))
```

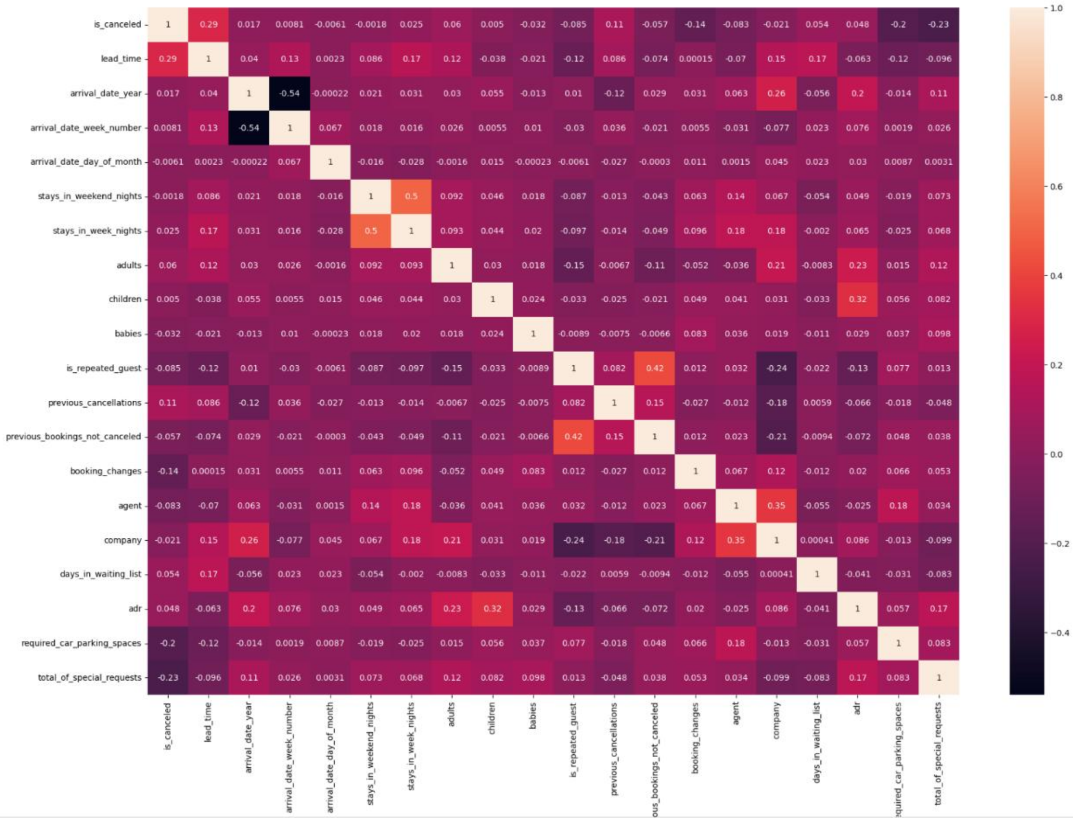
we will select the arrival_date_month feature and get its value count. Now the resulting data will not be sorted according to month order so we have to sort it. We will make the new list with the names of months in order to sort our data according to this list.



We can see that the most bookings were held in the month of July to August. And the least bookings were made at the end of the year.

Figure 4.1.4.12: Correlation matrix

```
In [30]: fig, ax = plt.subplots(figsize=(22,15))
         sns.heatmap(df_subset.corr(), annot=True, ax=ax);
```



We can see our new features, Room and ner_canceled have a higher correlation with is_cancelled than the most of the other columns.

Figure 4.1.4.13: Converting Categorical variables to Numerical and train test split

```
In [31]: def transform(dataframe):


             ## Import LabelEncoder from sklearn
             from sklearn.preprocessing import LabelEncoder

             le = LabelEncoder()


             ## Select all categorcial features
             categorical_features = list(dataframe.columns[dataframe.dtypes == object])


             ## Apply Label Encoding on all categorical features
             return dataframe[categorical_features].apply(lambda x: le.fit_transform(x))

         df = transform(df)
```

```
In [32]: def data_split(df, label):

             from sklearn.model_selection import train_test_split

             X = df.drop(label, axis=1)
             Y = df[label]

             x_train, x_test, y_train, y_test = train_test_split(X,Y,random_state=0)

             return x_train, x_test, y_train, y_test

         7

         x_train, x_test, y_train, y_test = data_split(df_subset, 'is_canceled')
```

Let’s convert categorical values into numerical form.We will use LabelEncoder from Sklearn to encode in an ordinal fashion.

Now let’s split the dataset into train and test. The default size of the split ratio is 3:1

```
========================================
Training Accuracy of our model is: 0.9956043710224032
Test Accuracy of our model is: 0.9956043710224032
========================================
```

4.2 Features :

The hospitality industry is an ever-evolving landscape, with hotels constantly seeking ways to maximize revenue, enhance customer satisfaction, and streamline their operations. In this digital age, hotel booking analysis has emerged as a crucial tool that empowers hoteliers to make data-driven decisions. This essay explores the key features of hotel booking analysis and how it revolutionizes the hotel industry. Hotel booking analysis enables businesses to delve into historical data to identify booking patterns and trends. This insight helps hotels understand when their properties are in high demand, allowing them to adjust pricing, marketing efforts, and staff scheduling accordingly. For instance, they can offer discounts during low-demand periods to boost occupancy. Hotels utilize various booking channels, including their official websites, online travel agencies (OTAs), and direct phone reservations. Booking analysis helps hotels track which channels are the most lucrative and how guests prefer to make reservations. This knowledge informs marketing strategies and budget allocation. Understanding the lead time for bookings and the average length of guest stays is pivotal for optimizing hotel operations. By analyzing these metrics, hotels can manage inventory effectively and tailor their customer service based on guests' anticipated needs. Hoteliers often grapple with reservation cancellations, which can significantly impact revenue and occupancy rates. Hotel booking analysis can reveal patterns in cancellations, enabling hotels to implement more effective cancellation policies and minimize revenue loss. One of the most critical aspects of hotel booking analysis is revenue management. By scrutinizing demand, competitor pricing, and historical data, hotels can set optimal room rates to maximize revenue. This dynamic pricing strategy ensures that room rates align with market conditions.

CHAPTER 5

CONCLUSION

In conclusion, hotel booking analysis is an indispensable tool for the modern hospitality industry. It empowers hotels to unlock valuable insights from data, make informed decisions, and ultimately enhance their performance in various critical areas. By delving into booking patterns, understanding guest behavior, optimizing pricing strategies, and maintaining a strong online presence, hotels can improve their revenue, occupancy rates, and customer satisfaction. Furthermore, hotel booking analysis goes beyond the quantitative aspects; it enables hotels to harness qualitative data from guest reviews and feedback, fostering continuous improvement and guest-centric services. This analytical approach not only boosts financial performance but also contributes to a positive brand reputation in an era where online reviews and ratings carry significant weight. In a dynamic and competitive market, the ability to adapt to changing trends and guest preferences is essential for a hotel's long-term success. Hotel booking analysis provides the means to adapt, innovate, and stay ahead of the curve, making it an invaluable asset for any hotelier aiming to thrive in today's hospitality landscape. As technology continues to advance and data becomes increasingly abundant, the role of hotel booking analysis in shaping the future of the industry is set to grow, enabling hotels to provide even more memorable experiences to their guests.

REFERENCES:

Booking.com for Partners. 2021. *How to better understand, prevent, and reduce cancellations*. [online] Available at: <https://partner.booking.com/en-us/help/guides/how-better-understand prevent-and-reduce-cancellations> [Accessed 16 June 2022]. Verot, B., 2021.

8 Tips to Reduce Last Minute Hotel Cancellations and No Shows. [online] Hotelminder.com. Available at: <https://www.hotelminder.com/8-tips-to-reduce-last-minute-hotel cancellations-and-no-shows> [Accessed 16 June 2022]. Antonio, N., de Almeida, A. and Nunes, L., 2019. Hotel booking demand datasets. *Data in Brief*, [online] 22, pp.41-49.

Available at: <https://www.sciencedirect.com/science/article/pii/S2352340918315191> [Accessed 15 June 2022]. [Accessed 15 June 2022]. Mostipak, J., 2020. *Hotel booking demand*. [online] Kaggle.com. Available at: <https://www.kaggle.com/datasets/jessemostipak/hotel-booking demand> [Accessed 16 June 2022]

SORCE CODE :

```
#Importing Libraries

import pandas as pd

import matplotlib.pyplot as plt

import seaborn as sns

import warnings

import seaborn as sns

warnings.filterwarnings('ignore')

#Loading the Dataset

df = pd.read_csv(r"C:\Users\karan\Downloads\hotel_bookings.csv")

#Exploratory Data Analysis and Data Cleaning

df.head()

df.tail()

df.shape

df.info()

df['reservation_status_date'] = pd.to_datetime(df['reservation_status_date'])

df.describe(include = 'object')

for col in df.describe(include = 'object').columns:

    print(col)

    print(df[col].unique())

    print('-'*50)

df.isnull().sum()

df.drop(['company','agent'],axis = 1, inplace = True)

df.dropna(inplace = True)

df.describe()
```

```
cancelled_perc = df['is_canceled'].value_counts(normalize = True)

print(cancelled_perc)


plt.figure(figsize = (5,4))

plt.title('Reservation status count')

plt.bar(['Not canceled','canceled'],df['is_canceled'].value_counts(), edgecolor = 'k',width
= 0.7)

plt.show()

plt.figure(figsize = (8,4))

ax1= sns.countplot(x = 'hotel', hue = 'is_canceled',data = df, palette = 'Blues')

legend_labels,_ = ax1. get_legend_handles_labels()

plt.title('Reservation status in different hotels',size = 20)

plt.xlabel('hotel')

plt.ylabel('number of reservations')

resort_hotel = df[df['hotel'] == 'Resort Hotel']

resort_hotel['is_canceled'].value_counts(normalize = True)

city_hotel = df[df['hotel'] == 'City Hotel']

city_hotel['is_canceled'].value_counts(normalize = True)

resort_hotel = resort_hotel.groupby('reservation_status_date')[['adr']].mean()

city_hotel = city_hotel.groupby('reservation_status_date')[['adr']].mean()

plt.figure(figsize = (20,8))

plt.title('Average Daily Rate in City and Resort Hotel', fontsize = 30)

plt.plot(resort_hotel.index,resort_hotel['adr'], label = 'Resort Hotel')

plt.plot(city_hotel.index,city_hotel['adr'], label = 'City Hotel')

plt.legend(fontsize = 20)

plt.show()
```

```
df['month'] =df['reservation_status_date'].dt.month

plt.figure(figsize = (16,8))

ax1 = sns.countplot(x = 'month',hue = 'is_canceled',data = df,palette = 'bright')

legend_labels,_ = ax1. get_legend_handles_labels()

plt.title('Reservation status per month',size = 20)

plt.xlabel('month')

plt.ylabel('number of reservations')

plt.legend(['not canceled','canceled'])

plt.show()

plt.figure(figsize=(15, 8))

plt.title('ADR per month', fontsize=30)

sns.barplot(x='month', y='adr', data=df[df['is_canceled'] ==
1].groupby('month')[['adr']].sum().reset_index())

plt.legend(fontsize=20)

plt.show()

cancelled_data = df[df['is_canceled'] == 1]

top_10_country = cancelled_data['country'].value_counts()[:10]

plt.figure(figsize = (8,8))

plt.title('Top 10 countries with reservation canceled')

plt.pie(top_10_country,autopct = '%.2f',labels = top_10_country.index)

plt.show()

df['market_segment'].value_counts()

df['market_segment'].value_counts(normalize = True)

cancelled_data['market_segment'].value_counts(normalize = True)

sns.barplot(x= 'market_segment', y= 'adr', hue= 'reserved_room_type', data= df)

df_subset = df.copy()
```

```
fig, ax = plt.subplots(figsize=(22,15))
sns.heatmap(df_subset.corr(), annot=True, ax=ax);
def transform(dataframe):


    ## Import LabelEncoder from sklearn
    from sklearn.preprocessing import LabelEncoder

    le = LabelEncoder()


    ## Select all categorcial features
    categorical_features = list(dataframe.columns[dataframe.dtypes == object])


    ## Apply Label Encoding on all categorical features
    return dataframe[categorical_features].apply(lambda x: le.fit_transform(x))

df = transform(df)
def data_split(df, label):

    from sklearn.model_selection import train_test_split

    X = df.drop(label, axis=1)
    Y = df[label]
```

```
    x_train, x_test, y_train, y_test = train_test_split(X,Y,random_state=0)


    return x_train, x_test, y_train, y_test
```

7

```
x_train, x_test, y_train, y_test = data_split(df_subset, 'is_canceled')
## Getting Prediciton of 10th record of x_train
prediction = clf.predict(x_train.iloc[10].values.reshape(1,-1))

## Actual Value of 10th record of x_train from y_train
actual_value = y_train.iloc[10]

print(f'Predicted Value \t: {prediction[0]}')
print(f'Actual Value\t\t: {actual_value}')
```