Green University of Bangladesh

Department of Computer Science and Engineering (CSE)
Semester: (Spring, Year: 2023), B.Sc. in CSE (Day)

Hotel Reservation Cancellations Data Analysis and Prediction

Course Title: Data Mining Lab
Course Code: CSE 424
Section: 193 D5

Students Details

Name	ID
Md Sajadul Islam (Group Leader)	193002138
Tanvir Sheikh	191002006
Md Rumman Hossain Juwel	193002151

Submission Date: 16/06/2023
Course Teacher's Name: Rabea Khatun

[For teachers use only: Don't write anything inside this box]

Lab Project Status	
Marks:	**Signature:**
Comments:	**Date:**

Contents

Chapter 1

Introduction

1.1 Overview

Hotel Reservation Cancellations Data Analysis and Prediction is a project that aims to identify the factors that contribute to reservation cancellations for a City Hotel and a Resort Hotel. This project involves analyzing a dataset of booking information, testing hypotheses related to the pricing and the day of the week, and developing predictive models to make predictions about future cancellations. The primary objective of this project is to provide insights into the factors that lead to reservation cancellations and to develop strategies that can help both hotels reduce the number of cancellations. By analyzing the dataset and identifying patterns and trends in the data, this project can help both hotels to improve their revenue generation and operational efficiency. To achieve these goals, the project will involve a thorough analysis of the dataset, including the use of statistical techniques and data visualization tools. Additionally, the project will involve the development of predictive models, which can help both hotels to anticipate cancellations and take proactive measures to prevent them.

Overall, this project represents a comprehensive analysis of hotel reservation cancellations and aims to provide valuable insights and recommendations to help both City Hotel and Resort Hotel reduce cancellations, improve their revenue generation, and operate more efficiently.

1.2 Motivation

There are several compelling reasons why the Hotel Reservation Cancellations Data Analysis and Prediction project is important and worth pursuing:

- The project aims to help City Hotel and Resort Hotel to reduce cancellations, improve their revenue generation, and operate more efficiently.
- By identifying the factors that contribute to cancellations, this project can provide valuable insights that can inform pricing, promotions, and other business decisions.

- The project also involves developing predictive models to anticipate cancellations, which can help the hotels to better manage their resources and staffing.
- Reducing cancellations can lead to increased customer satisfaction and loyalty, as well as positive word-of-mouth recommendations.
- The insights and recommendations generated by this project can have broader applications in the hospitality industry, making it a valuable contribution to the field.

1.3 Problem Definition

1.3.1 Problem Definition

The Hotel Reservation Cancellations Data Analysis and Prediction project aims to address the problem of high cancellation rates in the hospitality industry. The project involves analyzing data from two hotels to identify factors that contribute to cancellations and developing predictive models to anticipate cancellations, thereby reducing revenue losses and improving operational efficiency.

1.3.2 Problem Statement

The hospitality industry faces a significant challenge with high rates of hotel reservation cancellations, resulting in lost revenue and operational inefficiencies. The Hotel Reservation Cancellations Data Analysis and Prediction project aims to address this problem by analyzing reservation data from City Hotel and Resort Hotel to identify the factors that contribute to cancellations and developing predictive models to anticipate cancellations. The project aims to provide valuable insights and recommendations to help the hotels reduce cancellations, improve their revenue generation, and operate more efficiently. elaborate this and make it professional

1.4 Design Goals/Objectives

The Design Goals and Objectives of the Hotel Reservation Cancellations Data Analysis and Prediction project are

1. To identify the factors that contribute to cancellations in City Hotel and Resort Hotel reservation data.
2. To develop predictive models to anticipate cancellations and provide early warning to hotel management.
3. To provide valuable insights and recommendations that can inform pricing, promotions, and other business decisions, helping the hotels to reduce cancellations, improve their revenue generation, and operate more efficiently.

4. To enhance customer satisfaction and loyalty by reducing cancellations and providing a better overall hotel experience.

5. To create a valuable contribution to the hospitality industry by providing insights and recommendations that can be applied more broadly to other hotels and businesses.

The project design goals/objectives align with the problem definition and will guide the project team in developing and implementing effective solutions.

1.5 Application

The Hotel Reservation Cancellations Data Analysis and Prediction project has a variety of applications in the hospitality industry, including:

1. **Revenue Optimization:** By identifying the factors that contribute to cancellations, hotels can make more informed decisions about pricing and promotions, ultimately leading to increased revenue and profitability.

2. **Resource Management:** Predictive models developed through this project can help hotels better manage their resources and staffing by anticipating cancellations and adjusting accordingly. This can improve operational efficiency and reduce waste.

3. **Customer Satisfaction:** Reducing cancellations can lead to increased customer satisfaction and loyalty, as well as positive word-of-mouth recommendations. By creating a more seamless and enjoyable hotel experience, hotels can enhance their reputation and attract more business.

4. **Industry-Wide Impact:** The insights and recommendations generated by this project can have broader applications in the hospitality industry, making it a valuable contribution to the field. Other hotels and businesses can benefit from the knowledge gained through this project, ultimately leading to a more competitive and innovative industry as a whole.

Overall, the Hotel Reservation Cancellations Data Analysis and Prediction project has the potential to significantly improve the efficiency and profitability of hotels, while also enhancing the customer experience and contributing to the advancement of the hospitality industry.

Chapter 2

Implementation of the Project

2.1 Introduction

In this project, there will be two parts. In part one we will focus on the Data Analysis part and try to find out some Insights. Generally in this part one we will focus on data cleaning, Data Analyst, and Data Visualization for finding some valuable insights that will help the Hotel Manager to understand why people basically cancel hotel reservations. By following our valuable insights, assumption, and suggestions the authority can take necessary steps to improve their activities so that the hotel reservation cancellations can reduce.

Furthermore, in part two of our project, we will focus on prediction-based tasks. Specifically, we will concentrate on data preprocessing, feature selection techniques, and model building using ML algorithms that we have gained from Data Mining courses.

2.2 Project Details

This project aims to analyze a dataset obtained from Kaggle, which contains booking information for a city hotel and a resort hotel. The dataset includes various details about the bookings, such as when they were made, the length of stay, the number of adults, children, and/or babies, and the number of available parking spaces, among other things. The goal is to gain insights into why people cancel hotel reservations and provide valuable suggestions to the hotel manager for reducing cancellations.

The project is divided into two parts.

2.2.1 Part One: Data Analysis

Part One focuses on data analysis and includes the following steps:

1. **Data cleaning:** Prepare the dataset by handling missing values, outliers, and any inconsistencies in the data.

2. **Data analysis:** Explore the dataset to identify patterns, trends, and correlations related to hotel reservation cancellations.

3. **Data visualization:** Use visualizations to present the findings and make it easier to interpret the data.

2.2.2 Part Two: Prediction-Based Tasks

Part Two focuses on prediction-based tasks and involves the following steps:

1. **Data preprocessing:** Prepare the data for model training by handling categorical variables, scaling numerical features, and splitting the dataset into training and testing sets.

2. **Feature selection:** Identify the most relevant features that have a significant impact on predicting hotel reservation cancellations.

3. **Model building:** Utilize machine learning algorithms, learned from Data Mining courses, to develop prediction models that can accurately predict whether a booking will be canceled or not.

2.2.3 Data Set Description

The dataset used in this project contains booking information for a city hotel and a resort hotel. It includes various columns that provide valuable information about the bookings. Here is a description of some important columns:

- **hotel**: This column represents the name of the hotel. It distinguishes between the city hotel and the resort hotel.

- **is_canceled**: This column indicates whether the booking was canceled or not. It has a binary value of 1 for canceled bookings and 0 for non-canceled bookings.

- **lead_time**: This column represents the number of days that elapsed between the entering date of the booking into the property management system (PMS) and the arrival date. It gives an idea of how far in advance the bookings were made.

- **arrival_date_year**: This column represents the year of the arrival date. It helps in analyzing trends and patterns over different years.

These are just a few examples of the columns present in the dataset. There are additional columns such as **arrival_date_month**, **arrival_date_week_number**, **stays_in_weekend_nights**, **stays_in_week_nights**, **adults**, **children**, **babies**, and many others that provide more detailed information about the bookings.

By analyzing and exploring these columns, valuable insights can be gained regarding the booking patterns, lead times, cancellation rates, and other factors that can help the hotel management understand why people cancel hotel reservations. These insights can then be used to make improvements and take necessary steps to reduce hotel reservation cancellations.

Figure 2.1 represents the Dataset summary of our application

```
df=pd.read_csv('/content/hotel_bookings.csv')
```

```
[ ] df.head(10)
```

	hotel	is_canceled	lead_time	arrival_date_year	arrival_date_month	arrival_date_week_number	arrival_date_day_of_month	stays_in_weekend_nights
0	Resort Hotel	0	342	2015	July	27	1	0
1	Resort Hotel	0	737	2015	July	27	1	0
2	Resort Hotel	0	7	2015	July	27	1	0
3	Resort Hotel	0	13	2015	July	27	1	0
4	Resort Hotel	0	14	2015	July	27	1	0
5	Resort Hotel	0	14	2015	July	27	1	0

Figure 2.1: Data Set Description

2.3 Implementation

In this chapter, we will focus o the implementation part of our project. Basically, we have represented the implementation into several sub-sections.

2.3.1 Dataset Load

Figure 2.2 represents libraries that we have used for part one of our application

Importing the libraries

```
[ ] import pandas as pd
    import matplotlib.pyplot as plt
    import seaborn as sns
    import warnings
    warnings.filterwarnings('ignore')
```

Figure 2.2: Libraries that we have used for part one

Figure 2.3 represents how we have loaded our Dataset using Pandas in collab

```
Loading the dataset

[ ]  df=pd.read_csv('/content/hotel_bookings.csv')

[ ]  df.head(10)
```

Figure 2.3: Loaded our Dataset using Pandas in collab

Figure 2.4 represents making a copy of a dataset before data cleaning is done to preserve the original data in its raw form. This ensures that any changes made during the data-cleaning process do not affect the original dataset

```
[ ]  df_copy=df.copy()

     df.columns

     Index(['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', 'children', 'babies', 'meal',
            'country', 'market_segment', 'distribution_channel',
            'is_repeated_guest', 'previous_cancellations',
            'previous_bookings_not_canceled', 'reserved_room_type',
            'assigned_room_type', 'booking_changes', 'deposit_type', 'agent',
            'company', 'days_in_waiting_list', 'customer_type', 'adr',
            'required_car_parking_spaces', 'total_of_special_requests',
            'reservation_status', 'reservation_status_date'],
           dtype='object')
```

Figure 2.4: Dataset copy and represents the columns

2.3.2 Exploratory data analysis:

In this section, we will perform Exploratory Data Analysis (EDA) on the dataset to gain valuable insights that can help the hotel authority reduce the cancellation rate. By analyzing factors such as lead time, booking channel, previous cancellations, and special requests, we can identify patterns and trends that contribute to cancellations. This information can guide the hotel authority in making improvements, such as optimizing booking processes, offering incentives for non-cancellation, or addressing specific issues that lead to higher cancellation rates. By leveraging these insights, the hotel can take necessary steps to improve their activities and reduce hotel reservation cancellations.

Figure 2.5 represent the amount of reservation that is canceled. we will focus on finding some valuable insights

Let's visualize this for a better understanding Figure 2.6 represent the amount of reservation that is canceled on visualization

amount of reservation that is cancelled

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

     0    74745
     1    44152
     Name: is_canceled, dtype: int64
```

Figure 2.5: amount of reservation that is canceled

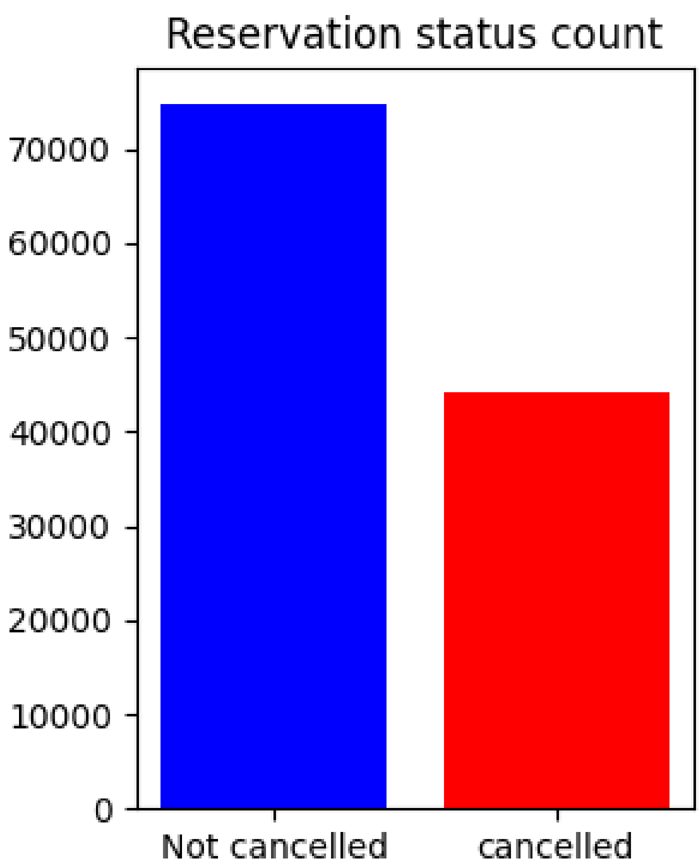

Figure 2.6: amount of reservation that is canceled

Figure 2.7 represents the reservation cancellation status for different hotel **Some**

Figure 2.7: Each Hotel Cancellation Asumption

Valuable insights from Figure 2.7:

1. City hotel has a large number of cancelation compared to Resort hotel.

Assumption:

1. City hotel has more booking because it is cheaper than a resort hotel
2. Price or maintenance may be the cause of high cancelation rates

2.3.3 Data Visualization:

Figure 2.8 represents the resort hotel cancellation status

Figure 2.9 represents the city hotel cancellation status

Some Insights from the above Visualization:

1) around 42% of reservations are getting canceled in City Hotel which is a very high number

2) Cancellations are Extremely high for city hotels compared with the Resort hotel

Figure 2.10 Represents the month wise cancellation status

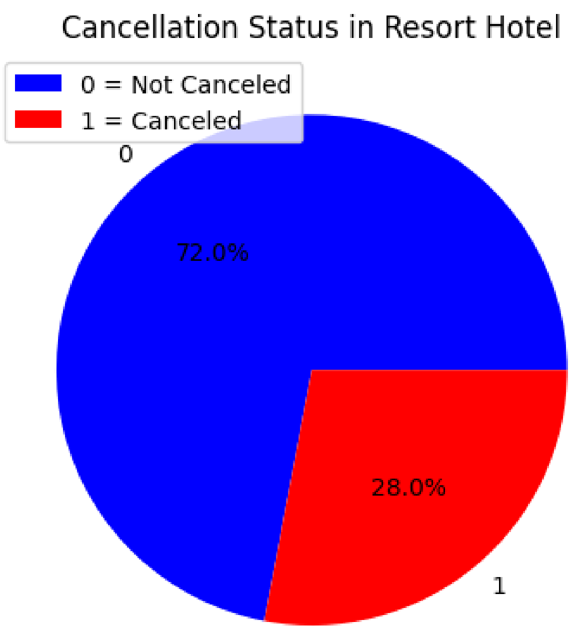

Figure 2.8: Resort hotel cancellation status

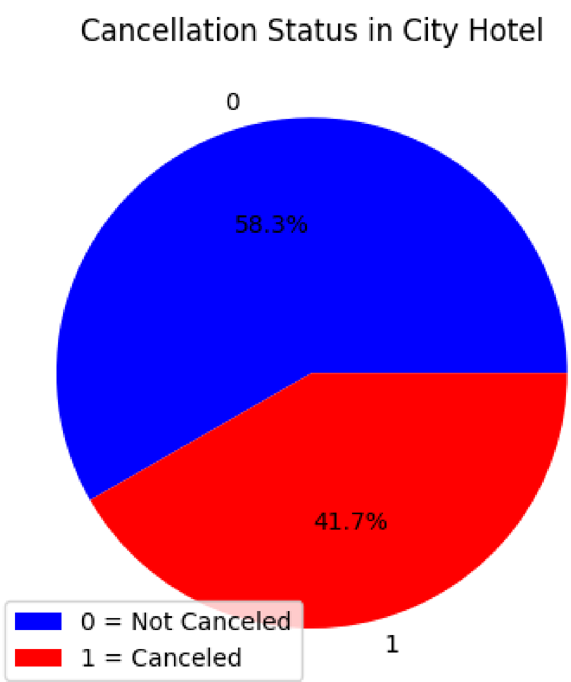

Figure 2.9: city hotel cancellation status

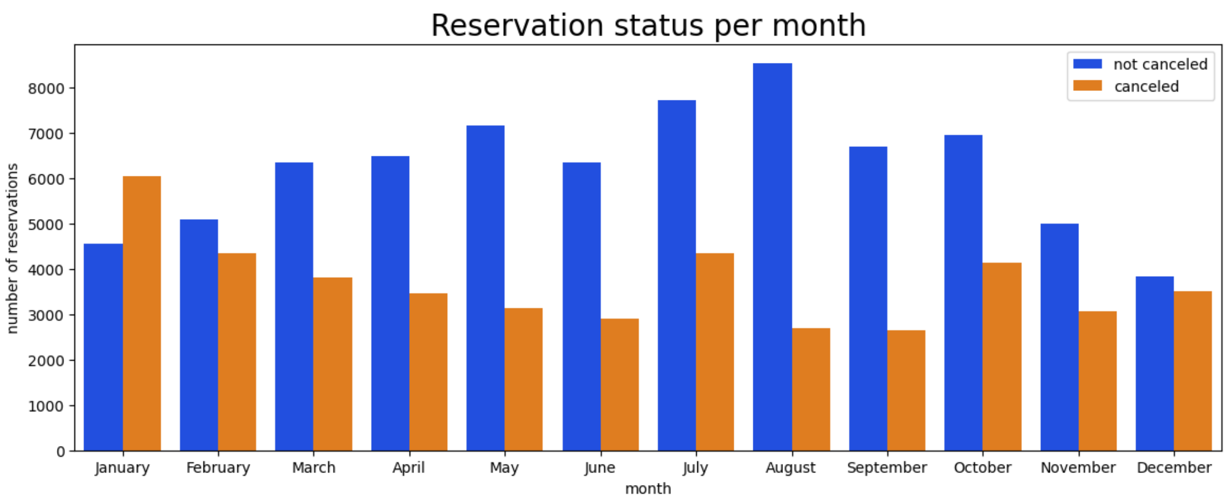

Figure 2.10: Month-wise cancellation status

Some Insights from the above Visualization:

1) the maximum number of cancellations in the month of January

2) And the lowest rate of cancelations is done in the month of September and August

Figure 2.11 Represents the day wise reservation cancellation

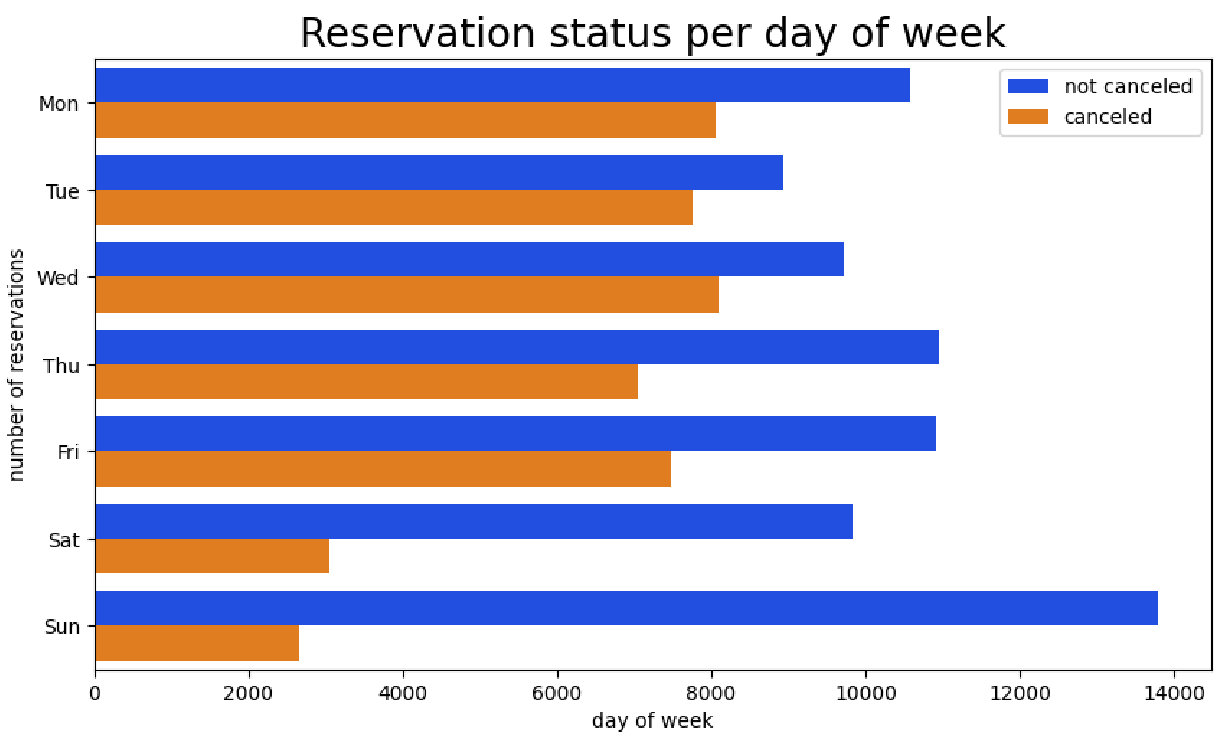

Figure 2.11: Day-wise reservation cancellation

Insights From the Visualization:
As shown in the above bar graph cancelation was highest on weekdays and very minimal on weekends.

Figure 3.3 Represents the ADR per month

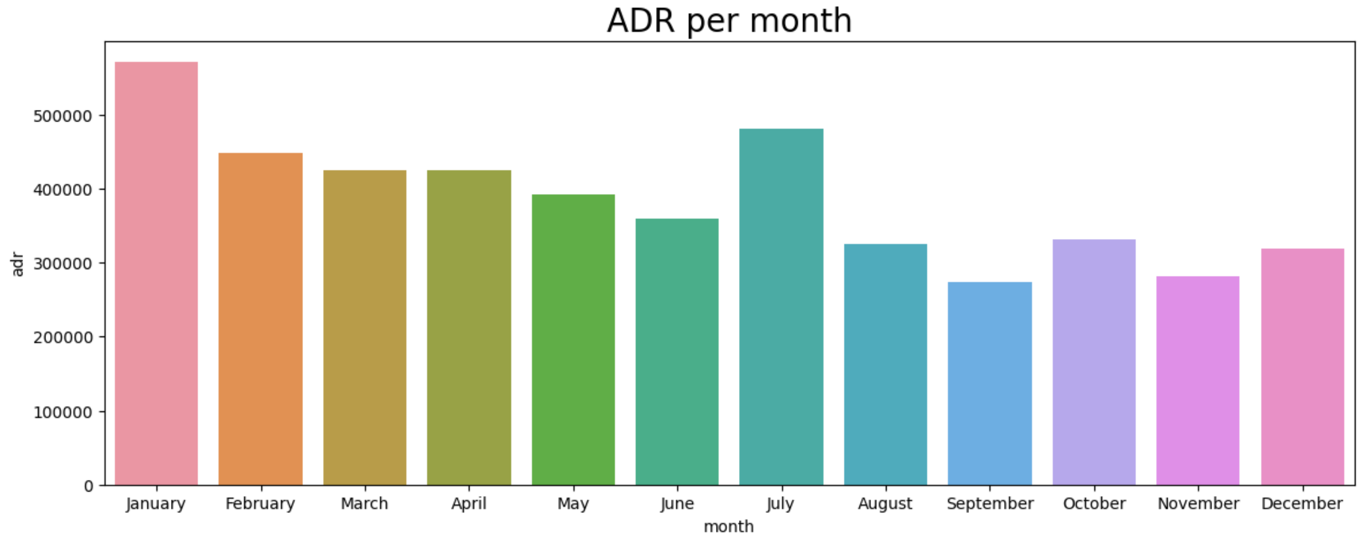

Figure 2.12: ADR per month

Valuable Insights:
1. months with the lowest cancelation rates have the lowest prices and months with the highest cancellation has the highest prices

January: highest prices also have the highest cancellations

September: lowest prices also have the weakest cancellations

Figure 2.13 cancelation rate based on Country. Here we have checked only the top 10 countries as we have data from 100s countries which will be difficult to analyze.

Insights:
Portugal has the maximum cancelation out of all countries and It is Extremely high 70%

Solving This Problem:
Increase facilities in hotels situated in Protugle, lower the Prices, give promotional discounts, run campines, increase advertisements to decrease the cancellation

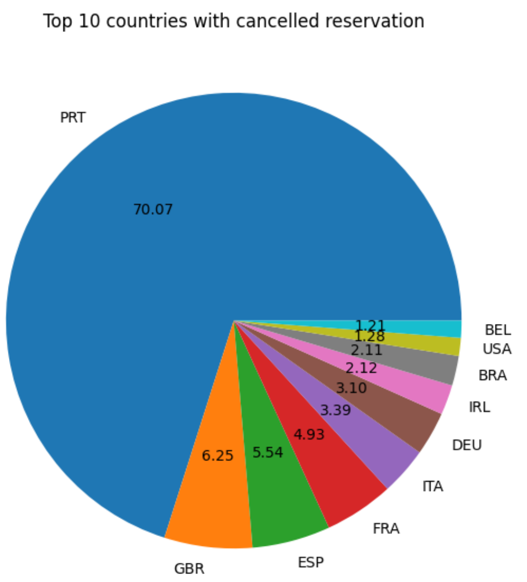

Figure 2.13: cancelation rate based on Country

Figure 2.14 represents the Market Segment Distribution for Booking Hotel

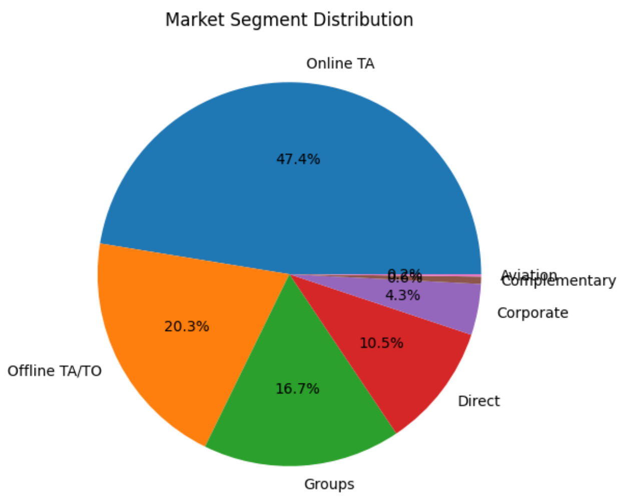

Figure 2.14: Market Segment Distribution for Booking Hotel

Insights:

From the above visualization, we can see that 47% of booking is coming from customers who book online

Figure 2.15 represents Market Segment Distribution for Cancellation Mediuml

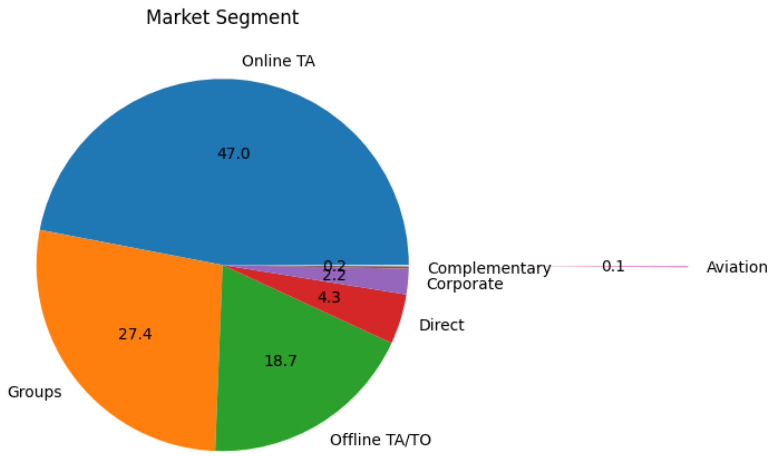

Figure 2.15: Market Segment Distribution for Cancellation Medium

Valuable Insights:
From the above visualization, we can see that around 47% around 4cancelation is done by customers who book online

Chapter 3

Performance Evaluation

3.1 Corelation with Heatmap:

Figure 3.3 Represents the finding correlation with heatmap

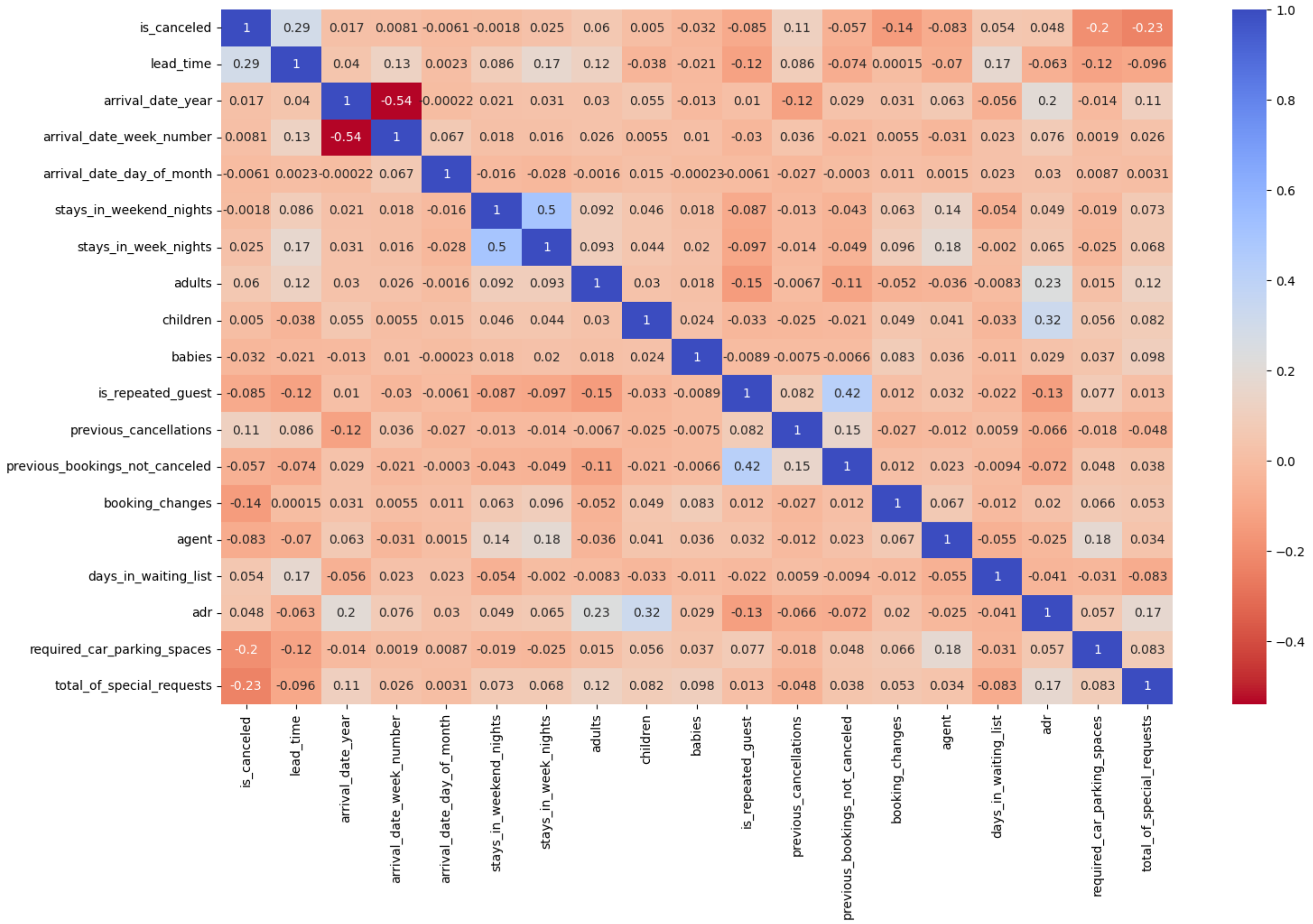

Figure 3.1: Finding correlation with heatmap

3.2 Feature Selection Technique:

Figure 3.3 Represents the feature selection technique. Higher the correlation higher the important feature.

```
[ ]  df.corr()['is_canceled'].abs().sort_values()

     stays_in_weekend_nights           0.001791
     children                          0.005048
     arrival_date_day_of_month         0.006130
     arrival_date_week_number          0.008148
     arrival_date_year                 0.016660
     stays_in_week_nights              0.024765
     babies                            0.032491
     adr                               0.047557
     days_in_waiting_list              0.054186
     previous_bookings_not_canceled    0.057358
     adults                            0.060017
     agent                             0.083114
     is_repeated_guest                 0.084793
     previous_cancellations            0.110133
     booking_changes                   0.144381
     required_car_parking_spaces       0.195498
     total_of_special_requests         0.234658
     lead_time                         0.293123
     is_canceled                       1.000000
     Name: is_canceled, dtype: float64
```

Figure 3.2: Finding importance correlation

3.3 Encoding Categorical Feature:

Figure 3.3 Represents the Label Encoding. AS ML model does not understand string values so to understand ML model string values have to convert into numerical values using Label Encoding Technique

Encoding Categorical Feature

```
[ ]  encoder=LabelEncoder()
     dict_cat={}
     for feature in df_cat.columns:
         dict_cat[feature]=encoder.fit_transform(df_cat[feature])
     #converting back the encoded feature into dataframe
     df_cat=pd.DataFrame(dict_cat)
```

```
[ ]  df_num.head(3)
```

	is_canceled	lead_time	stays_in_weekend_nights	stays_in_week_nights	adults
0	0	342	0	0	2
1	0	737	0	0	2
2	0	7	0	1	1

Figure 3.3: Label Encoding

3.4 Results Analysis

In this section, we will analyze the result of our application

3.4.1 Logistic Regression Classifier:

For our project, we have utilized the Logistic Regression classifier as the machine learning model of choice. This model is specifically employed for predicting whether a hotel booking will be canceled or not. By training the model on our dataset, we can generate predictions and insights regarding booking cancellations based on the input features. The Logistic Regression classifier enables us to understand the relationship between the features and the probability of cancellation, providing valuable information for the hotel authority to make informed decisions and take proactive measures to reduce the cancellation rate.

Logistic Regression Classifier

```
log_reg = LogisticRegression()
log_reg.fit(X_train, y_train)
y_pred = log_reg.predict(X_test)

accuracy = accuracy_score(y_test, y_pred)
f1 = f1_score(y_test, y_pred)
precision = precision_score(y_test, y_pred)
recall = recall_score(y_test, y_pred)

print('Accuracy Score: {:.2f}'.format(accuracy))
print('F1 Score: {:.2f}'.format(f1))
print('Precision: {:.2f}'.format(precision))
print('Recall: {:.2f}'.format(recall))
```

```
Accuracy Score: 0.78
F1 Score: 0.62
Precision: 0.85
Recall: 0.49
```

Figure 3.4: Logistic Regression Classifier Accuracy Score

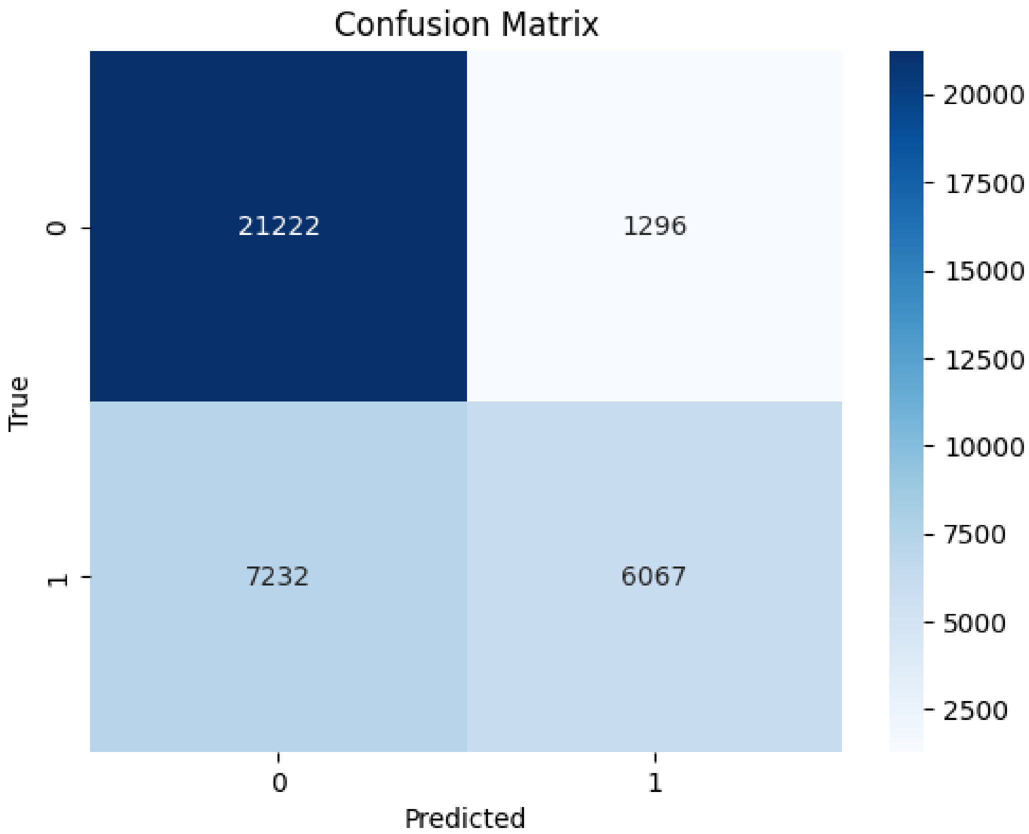

Figure 3.5: Logistic Regression Confusion Matrix

3.4.2 KNN Result Analysis:

For our project, we have employed the K-Nearest Neighbors (KNN) model as part of our machine-learning approach. The KNN algorithm is utilized to predict the likelihood of hotel booking cancellations. By training the model on our dataset, we can make predictions based on the nearest neighbors in the feature space. The KNN model allows us to identify patterns and similarities between bookings, enabling the hotel authority to gain insights into the factors influencing cancellations. This information can be used to implement targeted strategies and interventions to reduce the cancellation rate and improve overall booking satisfaction.

KNN

```
[ ] KNN = KNeighborsClassifier()
    KNN.fit(X_train, y_train)
    y_pred = KNN.predict(X_test)

    accuracy = accuracy_score(y_test, y_pred)
    f1 = f1_score(y_test, y_pred)
    precision = precision_score(y_test, y_pred)
    recall = recall_score(y_test, y_pred)

    print('Accuracy Score: {:.2f}'.format(accuracy))
    print('F1 Score: {:.2f}'.format(f1))
    print('Precision: {:.2f}'.format(precision))
    print('Recall: {:.2f}'.format(recall))
```

```
Accuracy Score: 0.90
F1 Score: 0.85
Precision: 0.92
Recall: 0.79
```

Figure 3.6: KNN Accuracy Score

3.4.3 Decision Tree Classifier:

For our project, we have utilized the Decision Tree Classifier as part of our machine learning methodology. The Decision Tree algorithm is employed to predict hotel booking cancellations based on various features and their corresponding decision rules. By constructing a tree-like model, the algorithm partitions the data based on different attribute values and recursively builds a decision tree. This allows us to identify important features and their thresholds that contribute to booking cancellations. The Decision Tree Classifier provides interpretability and can help the hotel authority understand the key

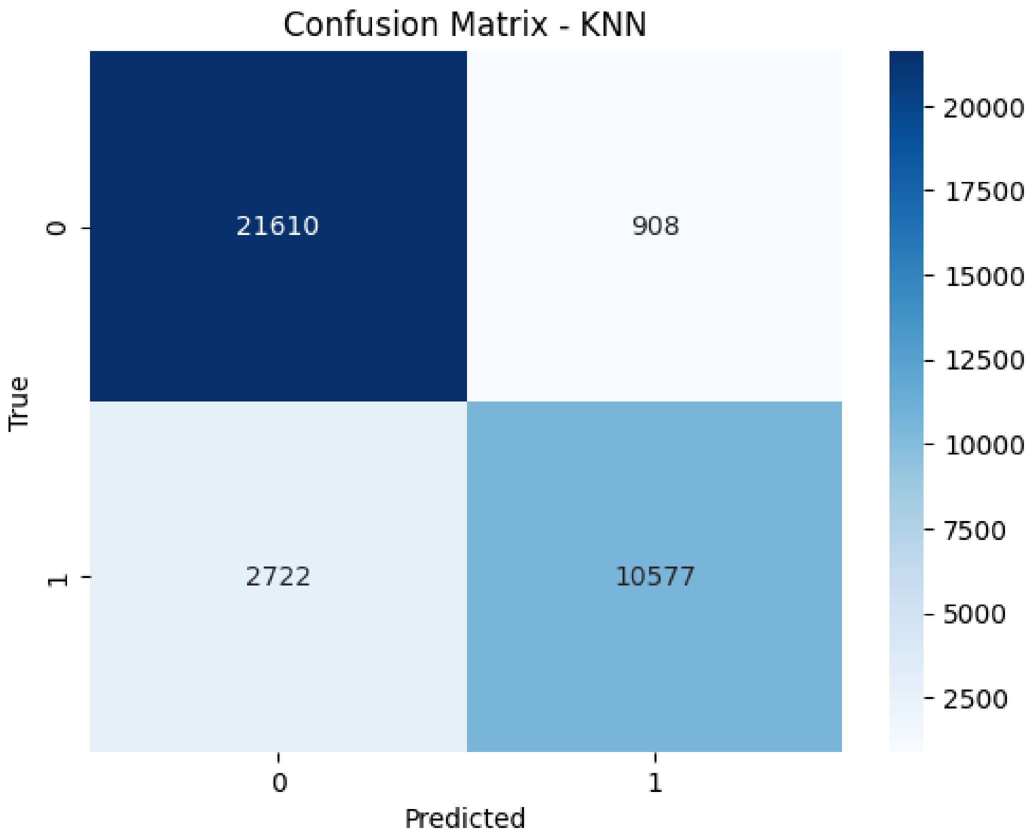

Figure 3.7: KNN Confusion Matrix

factors driving cancellations. This understanding can facilitate targeted interventions and strategies to mitigate cancellation rates and enhance the overall booking experience.

Decision Tree Classifier

```
[ ]  tree = DecisionTreeClassifier()
     tree.fit(X_train, y_train)
     y_pred = tree.predict(X_test)

     accuracy = accuracy_score(y_test, y_pred)
     f1 = f1_score(y_test, y_pred)
     precision = precision_score(y_test, y_pred)
     recall = recall_score(y_test, y_pred)

     print('Accuracy Score: {:.2f}'.format(accuracy))
     print('F1 Score: {:.2f}'.format(f1))
     print('Precision: {:.2f}'.format(precision))
     print('Recall: {:.2f}'.format(recall))
```

```
Accuracy Score: 0.91
F1 Score: 0.88
Precision: 0.87
Recall: 0.88
```

Figure 3.8: Decision Tree Classifier Accuracy Score

3.4.4 Random Forest Classifier:

The Random Forest model has shown high accuracy in our project, making it a valuable tool for the hotel industry. By analyzing various factors such as lead time, number of adults, previous cancellations, and other relevant features, the Random Forest Classifier can provide insights into the likelihood of booking cancellations. With its ability to handle complex interactions and capture non-linear relationships, this model helps the hotel authority make informed decisions and take proactive measures to reduce cancellation rates.

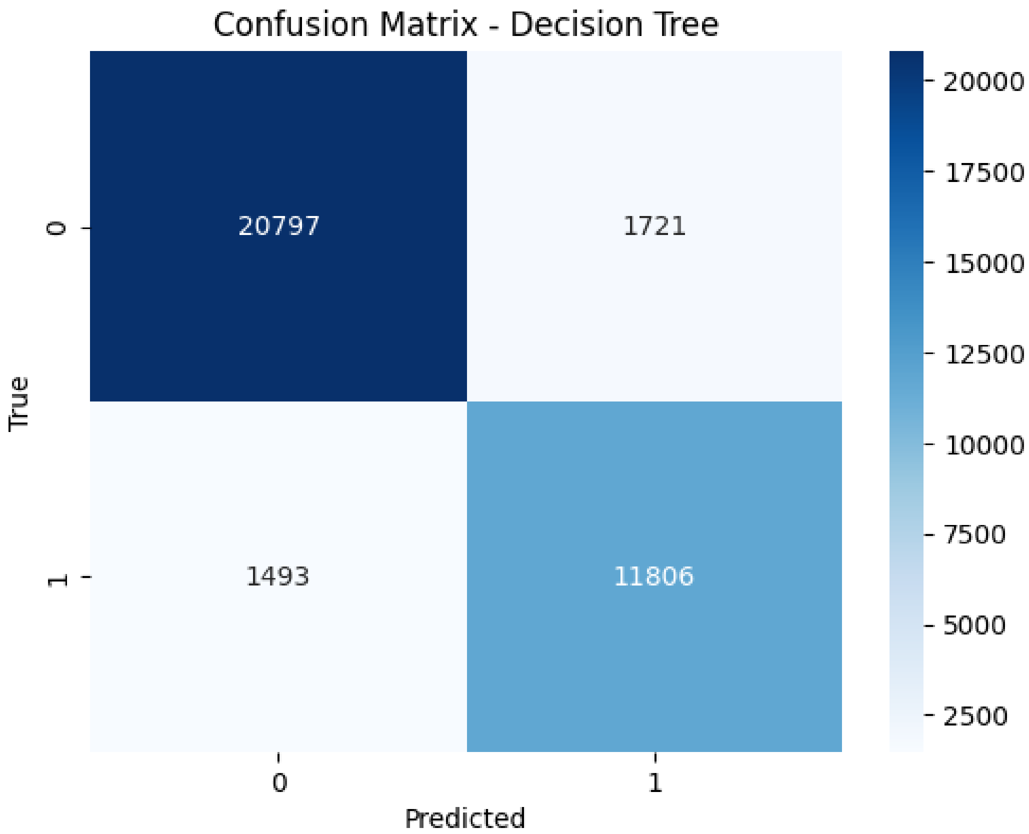

Figure 3.9: Decision Tree Classifier Confusion Matrix

Random Forest Classifier

```
[ ]  forest = RandomForestClassifier()
     forest.fit(X_train, y_train)
     y_pred = forest.predict(X_test)

     accuracy = accuracy_score(y_test, y_pred)
     f1 = f1_score(y_test, y_pred)
     precision = precision_score(y_test, y_pred)
     recall = recall_score(y_test, y_pred)

     print('Accuracy Score: {:.2f}'.format(accuracy))
     print('F1 Score: {:.2f}'.format(f1))
     print('Precision: {:.2f}'.format(precision))
     print('Recall: {:.2f}'.format(recall))
```

```
Accuracy Score: 0.93
F1 Score: 0.91
Precision: 0.95
Recall: 0.87
```

Figure 3.10: Random Forest Classifier Accuracy Score

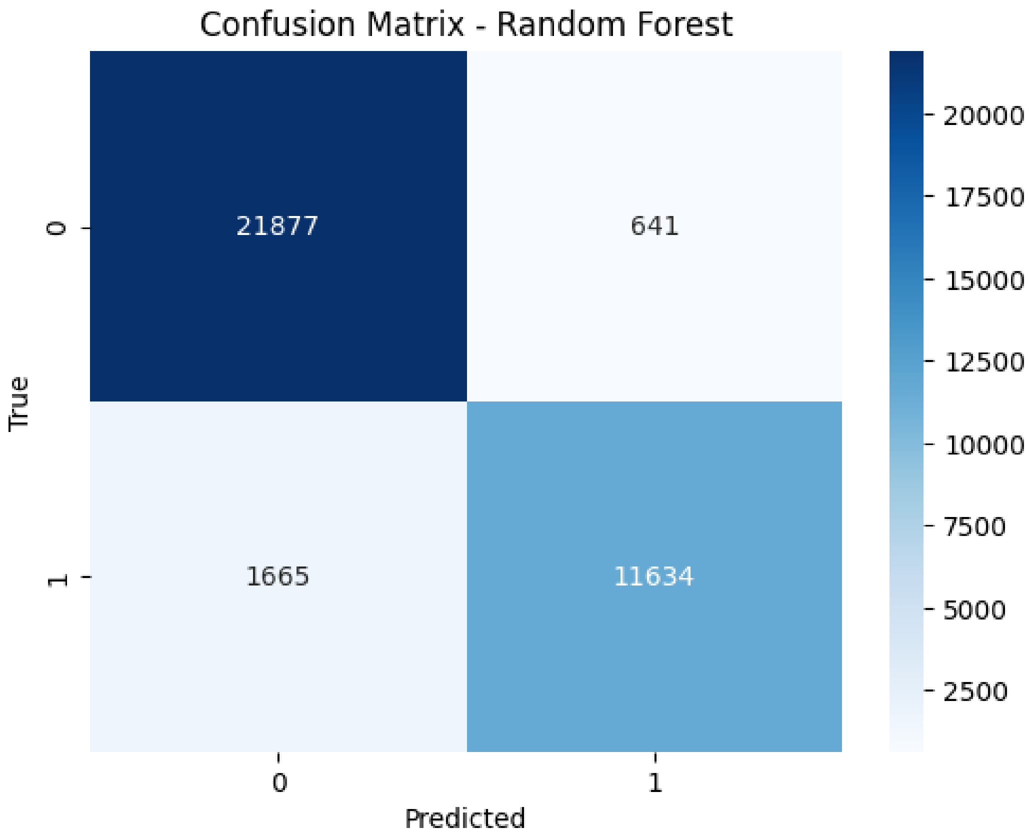

Figure 3.11: Random Forest Classifier Confusion Matrix

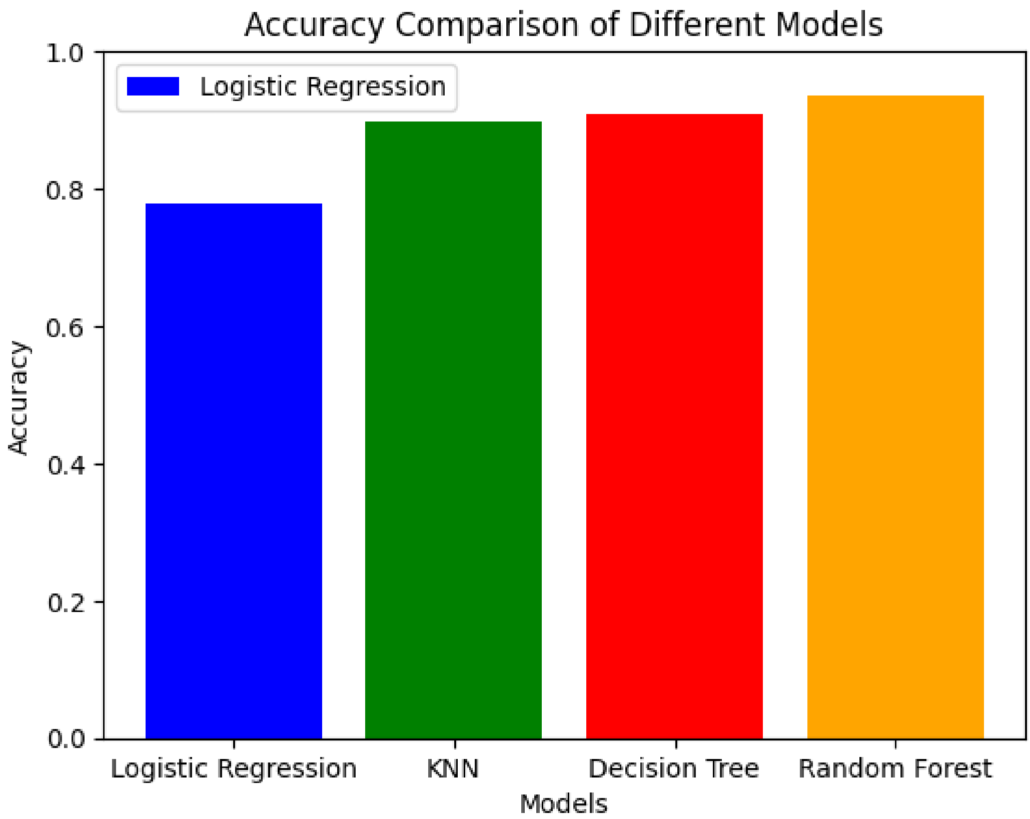

Figure 3.12: Accuracy Comparision of Different Model that we have used in our project

3.5 Accuracy Comparision of Different Model

From the above comparison, we can see that Random Forest has provided the highest accuracy of 93 %, indicating its effectiveness in predicting hotel booking cancellations. Both Decision Tree and K-Nearest Neighbors (KNN) algorithms have shown similar accuracy to Random Forest. However, Logistic Regression has yielded lower accuracy compared to the other models. This suggests that Logistic Regression may not be as suitable for this particular prediction task as the other algorithms.

Chapter 4

Conclusion

4.1 Discussion

In conclusion, our project focused on analyzing hotel booking data to gain insights into the factors affecting reservation cancellations and developing predictive models to reduce cancellation rates. Through exploratory data analysis, we discovered that the price, month, and hotel type were significant factors influencing cancellation rates.

We utilized various machine learning algorithms, including Logistic Regression, Decision Tree, K-Nearest Neighbors (KNN), and Random Forest, to predict booking cancellations. Among these models, Random Forest emerged as the most accurate, achieving a 93% accuracy rate. This algorithm's ability to combine multiple decision trees and leverage ensemble learning techniques contributed to its superior performance.

Our findings provide valuable insights for hotel authorities to understand the reasons behind booking cancellations and take proactive measures to mitigate them. By focusing on improving factors such as lead time management, customer communication, and service quality, hotels can work towards reducing cancellation rates and optimizing their revenue.

It is important to note that the choice of the most appropriate model depends on the specific requirements and characteristics of the dataset. Further research and experimentation with other machine learning algorithms may yield additional insights and potentially improve predictive accuracy.

Overall, our project demonstrates the value of data analysis and machine learning techniques in understanding and addressing challenges related to hotel reservation cancellations. By leveraging these insights, hotels can enhance their operations, customer satisfaction, and revenue generation.

4.2 Limitations

The project has a few limitations to consider. Firstly, the scope of the analysis is limited to a specific city hotel and resort hotel, which may restrict the generalizability of the findings to other types of hotels or locations. Secondly, the presence of missing or erroneous data could potentially impact the accuracy and reliability of the analysis and models. Additionally, the project focused on a selected set of machine learning algorithms, and there may be other algorithms that could yield better results. It's important to note that external factors, such as economic conditions or travel advisories, were not fully captured in the dataset, which could influence hotel booking cancellations. Lastly, while correlations were observed in the data, establishing causal relationships would require further research and analysis.

4.3 Scope of Future Work

The scope of future work includes expanding data collection, exploring advanced analysis techniques, incorporating natural language processing, developing real-time prediction systems, and implementing personalized recommendations to reduce hotel booking cancellations.

References

1. Putro, N. A., Septian, R., Widiastuti, M., & Pardede, H. F. (2021). *Penelitian Dosen Yayasan*.

2. Afrianto, M. A., & Wasesa, M. (2020). Booking Prediction Models for Peer-to-peer Accommodation Listings using Logistics Regression, Decision Tree, K-Nearest Neighbor, and Random Forest Classifiers. *Journal of Information Systems Engineering and Business Intelligence*, 6(2), 123–132. doi:10.20473/jisebi.6.2.123-132.

3. Aggarwal, R., & Ranganathan, P. (2017). Common pitfalls in statistical analysis: Linear regression analysis. *Perspectives in Clinical Research*, 8(2), 100–102. doi:10.4103/2229-3485.203040.

4. Al-Smadi, M., Talafha, B., Al-Ayyoub, M., & Jararweh, Y. (2018). Using long short-term memory deep neural networks for aspect-based sentiment analysis. *International Journal of Machine Learning and Cybernetics*, 9(3), 489–502. doi:10.1007/s13042-018-0799-4.

5. Alotaibi, E. (2020). Application of Machine Learning in the Hotel Industry: A Critical Review. *Journal of Association of Arab Universities for Tourism and Hospitality*, 14(1), 55–67. doi:10.21608/jaauth.2020.3878.4.1060.