Hotel Data Analysis and Cancelation Prediction

Introduction

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.

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

Objectives

The Design Goals and Objectives of the Hotel 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.

Implementation of the Project

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.

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 , 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.

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.

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.

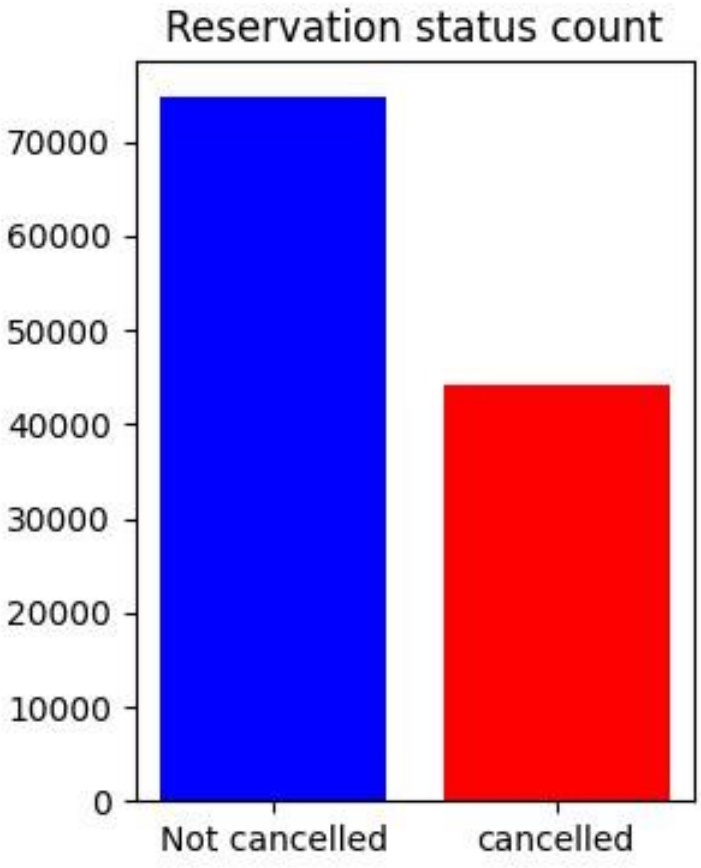

Figure 1 amount of reservation that is canceled

Figure 2 represents the reservation cancellation status for different hotel Some

Valuable insights from fig

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 3 : Resort hotel cancellation status

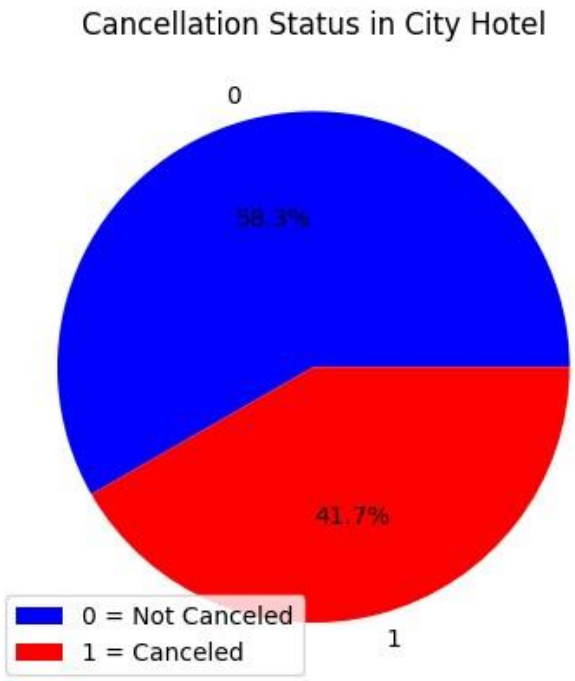

Figure 2.9: city hotel cancellation status

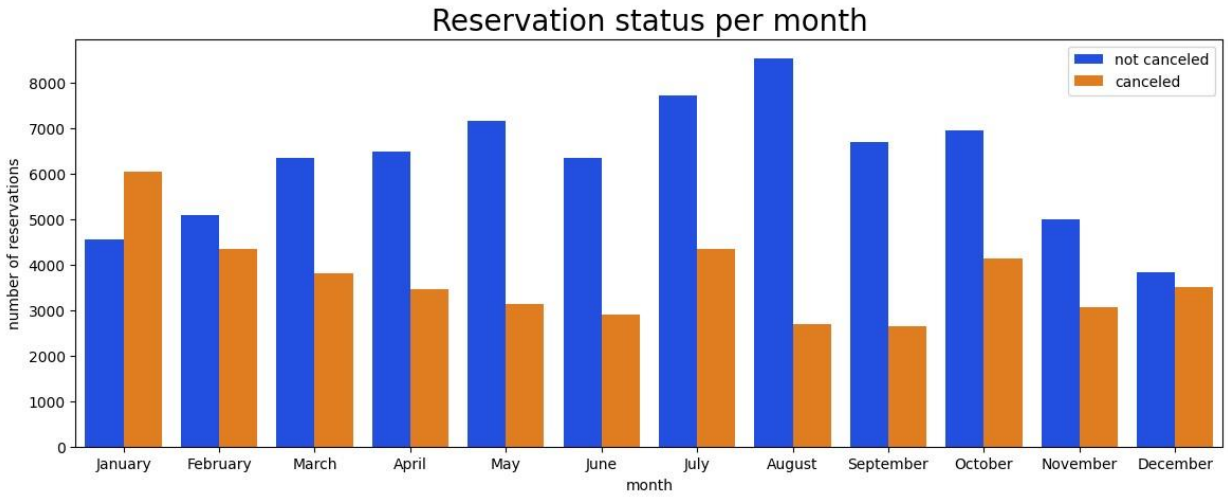

Figure 4 : Month-wise cancellation status

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

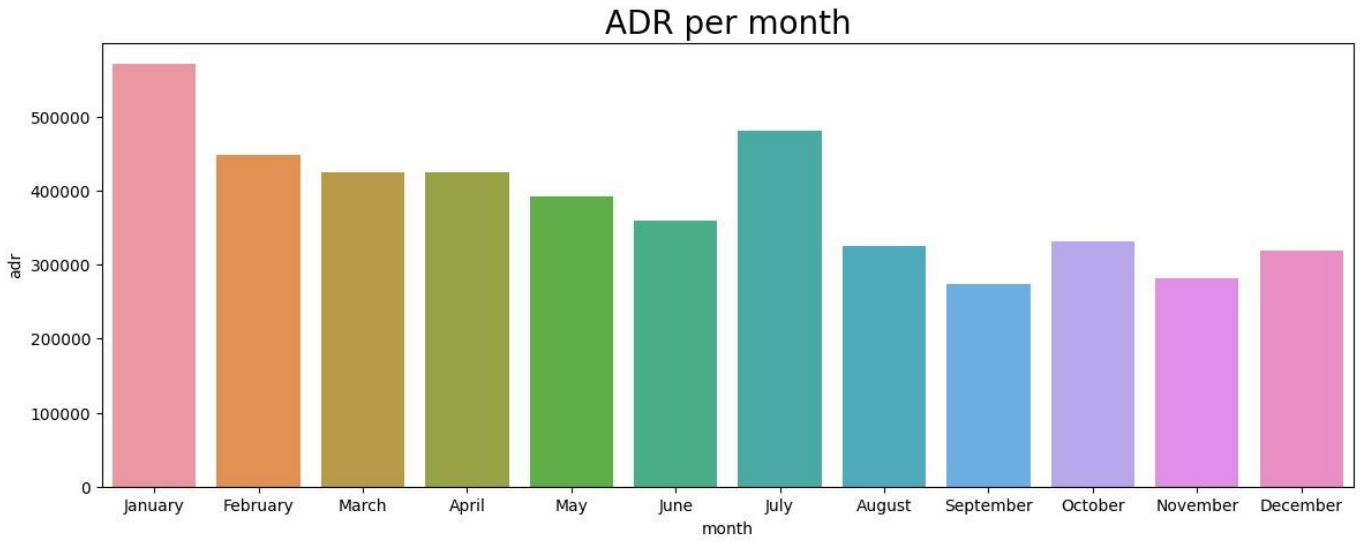

Figure 5 ADR per month

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 cancellation

Performance Evaluation

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.

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.

Accuracy Comparision of Different Model

Random Forest has provided the highest accuracy of 93 %, indicating its effectiveness in predicting hotel booking cancellations. 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.

Conclusion

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,), and Random Forest, to predict booking cancellations. Among these models, Random Forest emerged as the most accurate, achieving a 80% 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.