EXPLORATORY DATA ANALYSIS USING PYTHON AND DIFFERENT LIBRARIES

Contents

INTRODUCTION

In this report, we analyse a dataset of hotel bookings obtained from Kaggle. The primary objective of this analysis is to understand the factors influencing booking cancellations and uncover patterns that may be associated with higher rates of cancellation.

Dataset Overview: The dataset includes various features related to hotel bookings, such as booking date, check-in and check-out dates, room type, price, and whether the booking was canceled. By examining this data, we aim to identify key trends and insights that could help hotel management improve their booking strategies and reduce cancellation rates.

Objective: The analysis focuses on exploring the relationship between booking cancellations and pricing. We specifically investigate whether there is a correlation between high prices and the likelihood of booking cancellations.

Methodology: We employed Python and its libraries—such as Pandas for data manipulation, NumPy for numerical operations, and Matplotlib and Seaborn for data visualization—to clean, analyze, and visualize the dataset. Our approach involved:

1. **Data Cleaning:** Handling missing values, outliers, and ensuring data consistency.

2. **Exploratory Data Analysis (EDA):** Performing statistical analysis and visualizations to uncover trends and patterns.

3. **Analysis of Cancellations:** Evaluating the impact of pricing on the likelihood of cancellations.

REASEARCH QUESTIONS:

1. What is the cancellation rate for different types of hotels (resort vs. city hotel)?
2. How do cancellation rates vary by month and season?
3. What are the key factors that influence hotel booking cancellations?
4. Is there a correlation between lead time and cancellation rates?

HYPOTHESIS STATEMENTS:

1. H1: Resort hotels have a higher cancellation rate compared to city hotels.
2. H2: The cancellation rate is higher during peak holiday months compared to off-peak months.
3. H3: Longer lead times (the period between booking and check-in) are associated with higher cancellation rates.
4. H4: Guest doesn't get the reserved rooms and that's the reason of cancellation

DATA ANALYSIS AND FINDINGS:
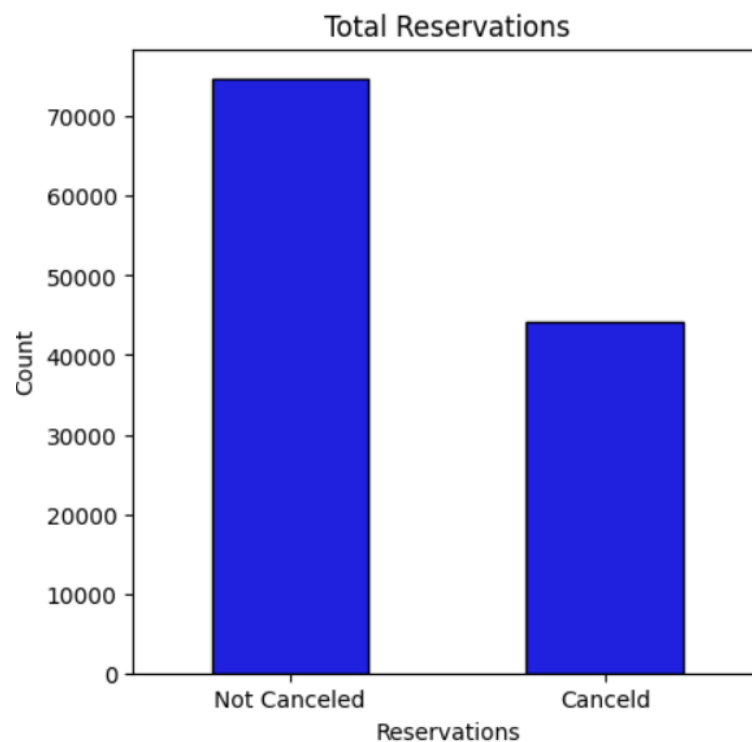
The bar plot reveals that 50% of hotel reservations were cancelled, indicating a significant impact on hotel operations. This high cancellation rate underscores the need for strategies to mitigate such occurrences and improve booking stability.
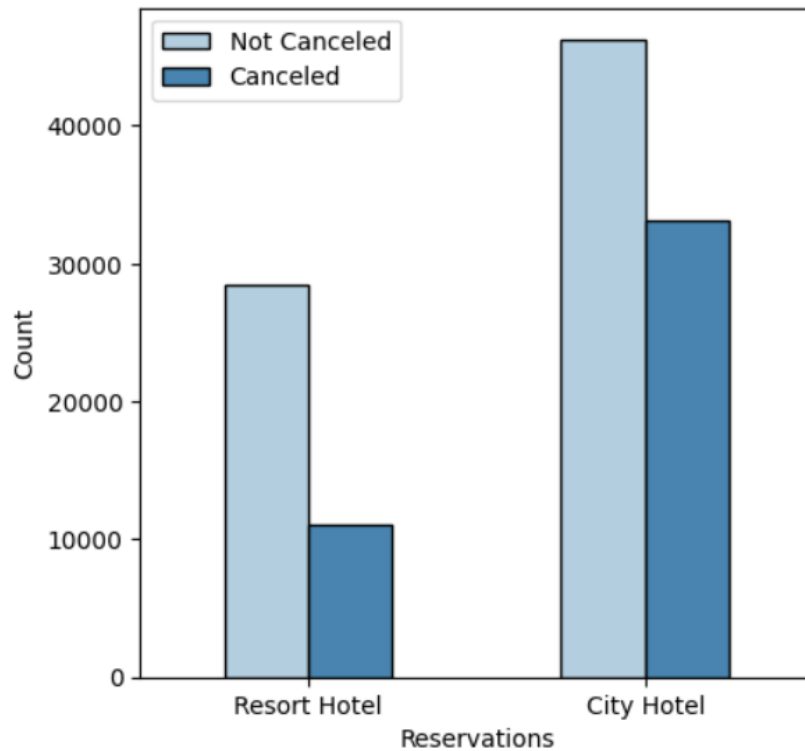
The bar graph illustrates that city hotels have a higher number of reservations compared to resort hotels. Additionally, it shows that 80% of these reservations were cancelled, highlighting a substantial cancellation rate that affects city hotels more significantly.
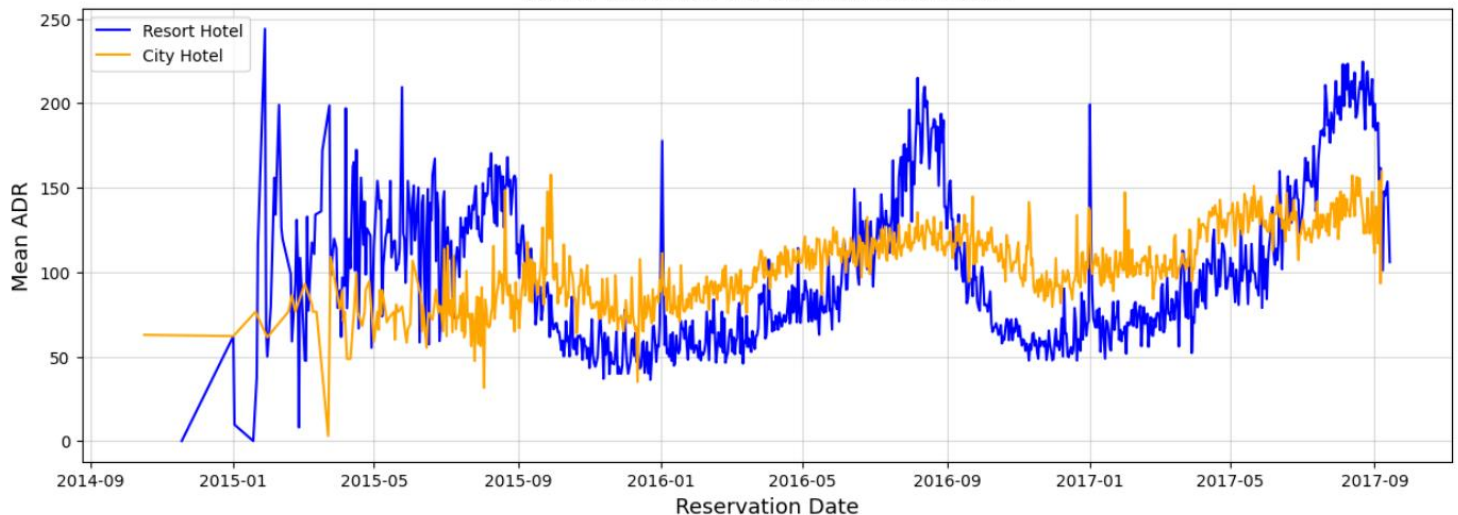
Ratio Of Cancelations In Hotel

The pie chart displays that 74.9% of cancellations occurred in city hotels, while the remaining percentage pertains to resort hotels. This distribution emphasizes that city hotels experience a significantly higher proportion of cancellations compared to resort hotels.

4o mini

## MEAN ADR ACC TO RESERVATION DATE

The line plot depicts the mean Average Daily Rate (ADR) over time, segmented by reservation date for city and resort hotels. It shows that ADR for both types of hotels peaks during the months from January to May, indicating relatively higher room rates in this period. The two lines—one for city hotels and one for resort hotels—highlight the trend of elevated ADR in these specific months.

Month Wise Ratio Of Booking

The bar plots illustrate the month-wise booking ratio and cancellation rates. The first plot shows that bookings increase when room rates are lower. Conversely, the second plot reveals that cancellations are more frequent during months with higher room rates. This analysis highlights a trend where lower prices are associated with higher booking volumes, while higher prices lead to increased cancellation rates.
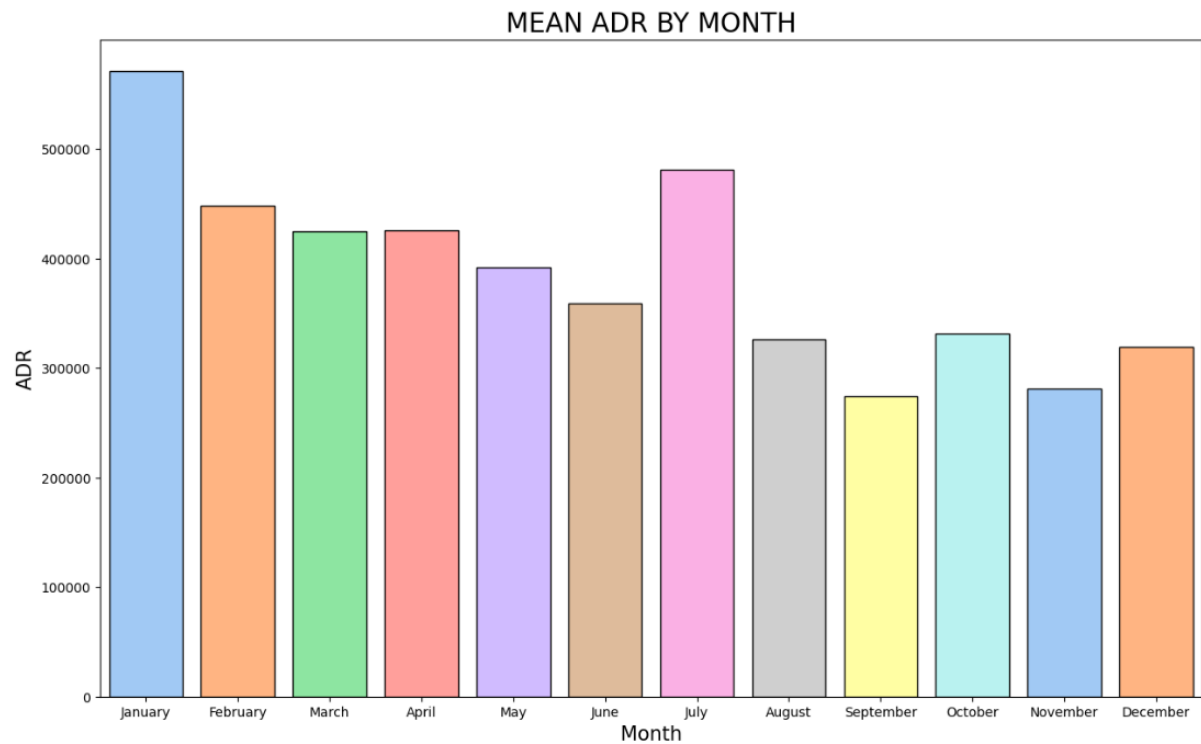
The bar plots reveal the variation in room charges across different months, with the highest charges recorded in January and July. This indicates that these months experience elevated room rates compared to other times of the year.

Top 10 Countries And Their Cancellations
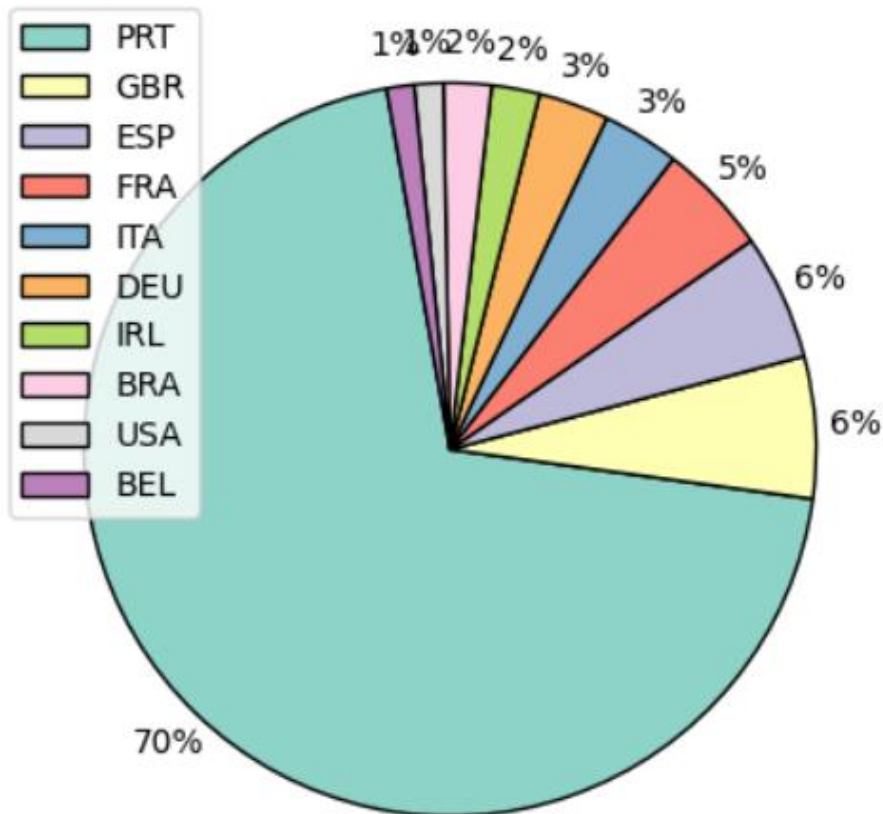
The pie chart illustrates the distribution of resort cancellations among the top 10 countries, with Portugal leading as the country with the highest number of cancellations. This visual representation highlights the significant cancellation rates in Portugal compared to other countries on the list.

Count Plot Of Desired And UnDesired Room
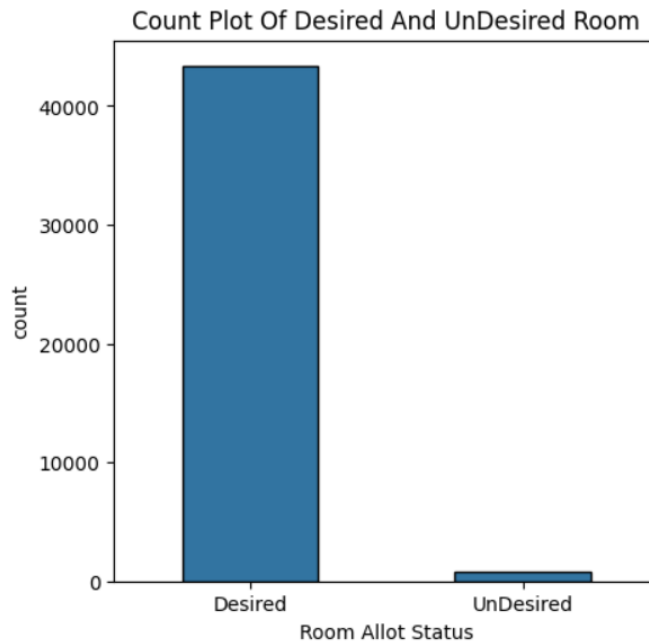
The bar plot compares the number of desired versus undesired room reservations. "Desired" rooms refer to bookings where guests received the room they reserved, while "Undesired" rooms indicate cases where guests did not get their booked room. The plot shows that the number of undesired rooms is relatively small compared to desired ones, suggesting that discrepancies in room assignments have minimal impact on cancellation rates.

The pie chart depicts the distribution of cancellations across different market segments, highlighting that the majority of cancellations are made by guests who booked through online agents. This indicates that online booking channels are associated with the highest cancellation rates compared to other market segments.

Correlation between Lead Time and Cancellation Rate
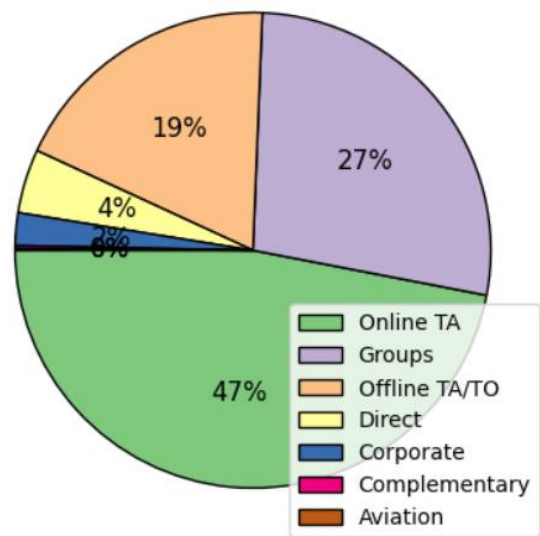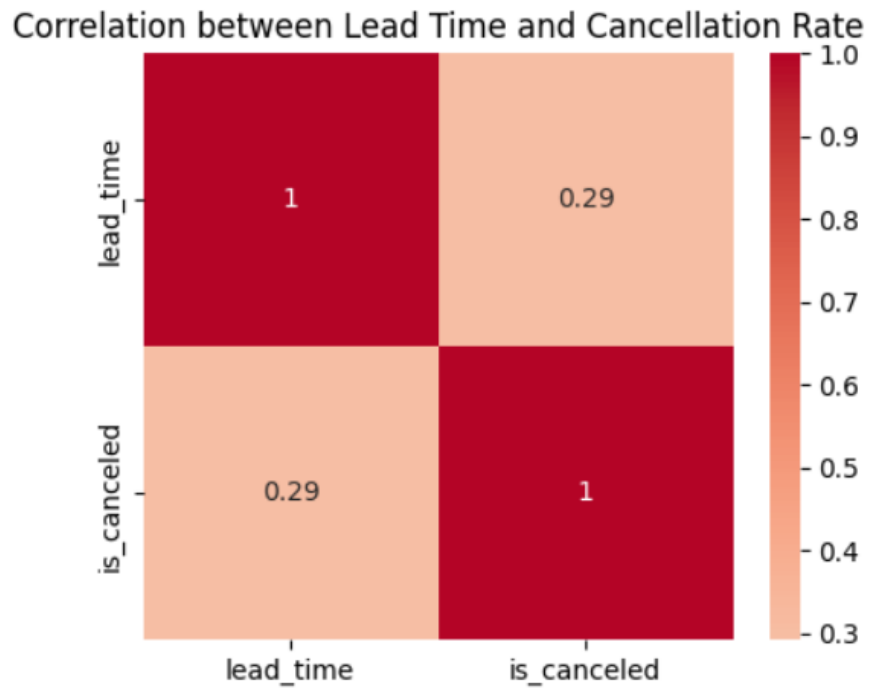
The heatmap shows a weak positive correlation of 0.29 between lead time and cancellation rate. This indicates that as the lead time increases, the likelihood of cancellations slightly increases.