Hotel Booking Analysis Report

Executive Summary

This report summarizes the analysis performed on the uploaded dataset (cleaned_data.csv). The dataset contains
118,897 rows and 31 columns. The overall booking cancellation rate is 37.1%. Average daily rate (ADR) is 101.96. Average lead time is 104.3 days. Detailed findings, visualizations, and recommendations follow in the report.

Research Question

1. What are the variables that affect hotel reservation cancellations?
2. How can we make hotel reservations cancellations better?
3. How will hotels be assisted in making pricing and promotional decisions?

Hypothesis

1. More cancellations occur when prices are higher.
2. When there is a longer waiting list, customers tend to cancel more frequently.
3. The majority of clients are coming from offline travel agents to make their reservations.

1. Dataset Overview

File: cleaned_data.csv
Rows: 118,897
Columns: 31
Columns and types:

Column	Type	Missing values
hotel	object	0
is_canceled	int64	0
lead_time	int64	0
arrival_date_year	int64	0
arrival_date_month	object	0
arrival_date_week_number	int64	0
arrival_date_day_of_month	int64	0
stays_in_weekend_nights	int64	0
stays_in_week_nights	int64	0
adults	int64	0
children	float64	0
babies	int64	0
meal	object	0
country	object	0
market_segment	object	0
distribution_channel	object	0
is_repeated_guest	int64	0
previous_cancellations	int64	0
previous_bookings_not_canceled	int64	0
reserved_room_type	object	0
assigned_room_type	object	0
booking_changes	int64	0
deposit_type	object	0
days_in_waiting_list	int64	0
customer_type	object	0

adr	float64	0
required_car_parking_spaces	int64	0
total_of_special_requests	int64	0
reservation_status	object	0
reservation_status_date	object	0
month	int64	0

2.Methods and Data Cleaning

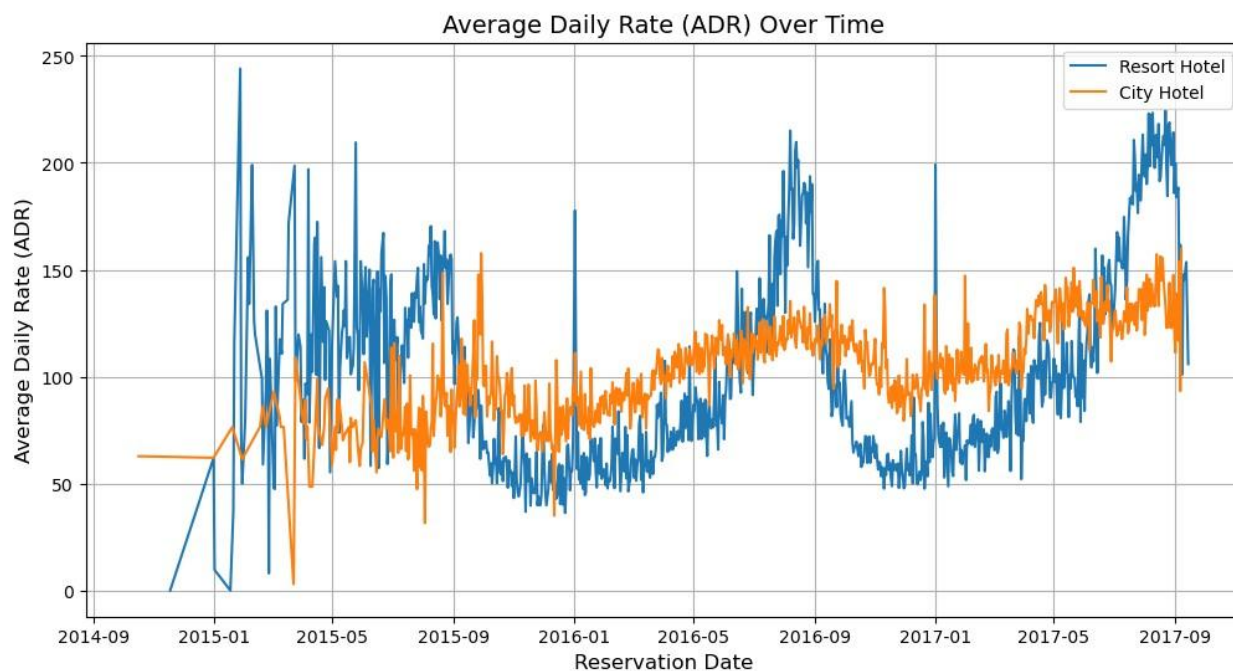# Importing Libraries

Loading Dataset

EDA and Data Cleaning

Data Analysis And Visualization

2. Visualizations (extracted from notebook)

Reservation Status Count

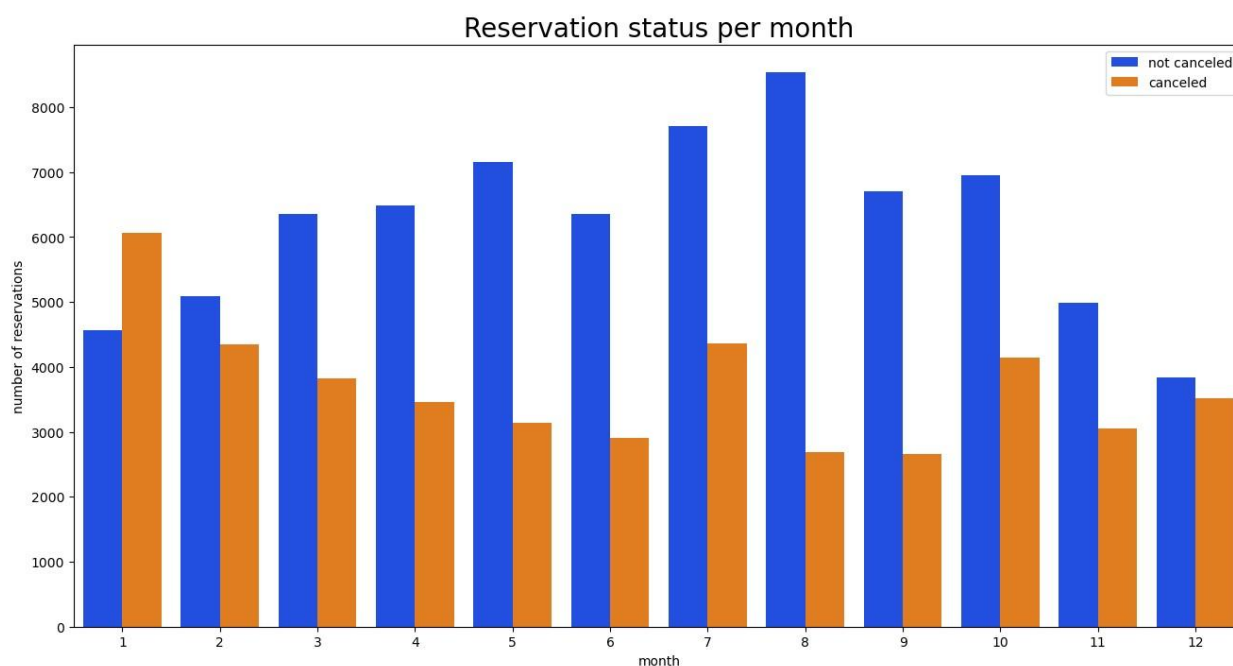Reservation status in differnet hotels

The accompanying bar graph shows the percentage of reservations that are canceled and those that are not. It is obvious that there are still a significant number of reservations that have not been canceled. There are still 37% of clients who canceled their reservation, which has a significant impact on the hotels' earnings.

In comparison to resort hotels, city hotels have more bookings. It's possible that resort hotels are more expensive than those in cities.

Average Daily Rate (ADR) Over Time

The line graph above shows that, on certain days, the average daily rate for a city hotel is less than that of a resort hotel, and on other days, it is even less. It goes without saying that weekends and holidays may see a rise in resort hotel rates.

Reservation status per month

We have developed the grouped bar graph to analyze the months with the highest and lowest reservation levels according to reservation status. As can be seen, both the number of confirmed reservations and the number of canceled reservations are largest in the month of August. whereas January is the month with the most canceled reservations.
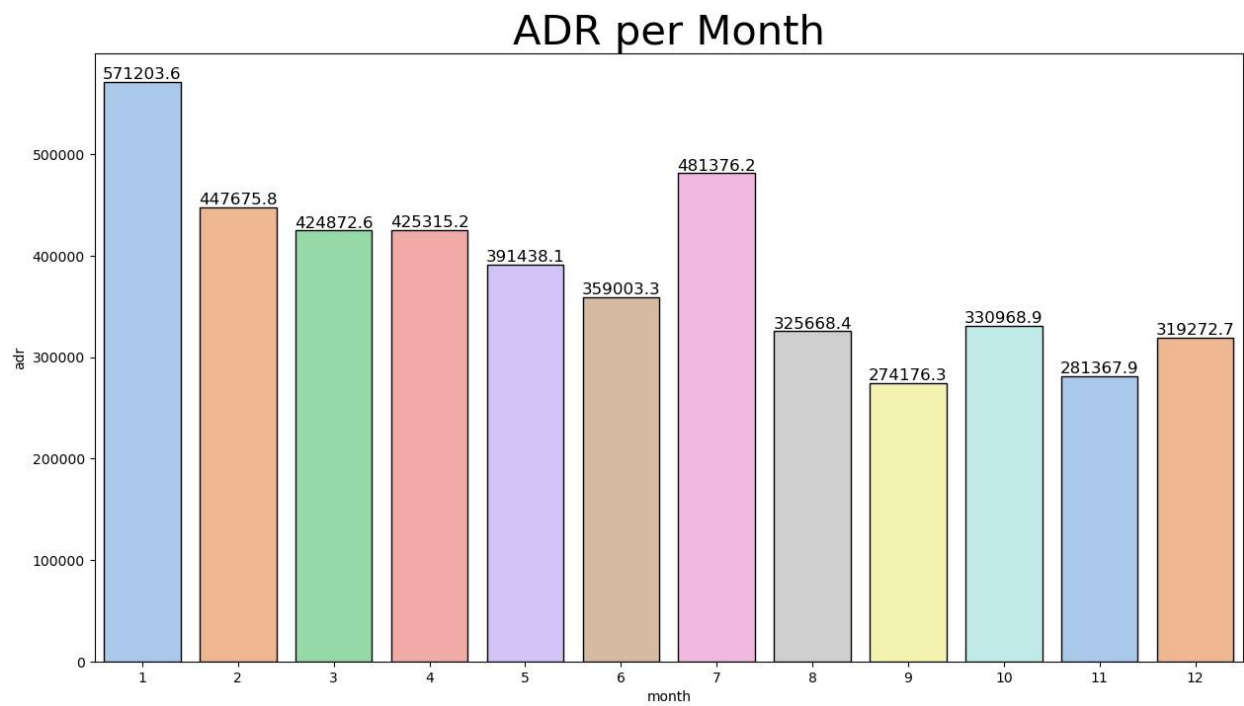
Figure: extracted from notebook output

This bar graph demonstrates that cancellations are most common when prices are greatest and are least common when they are lowest. Therefore, the cost of the accommodation is solely responsible for the cancellation.

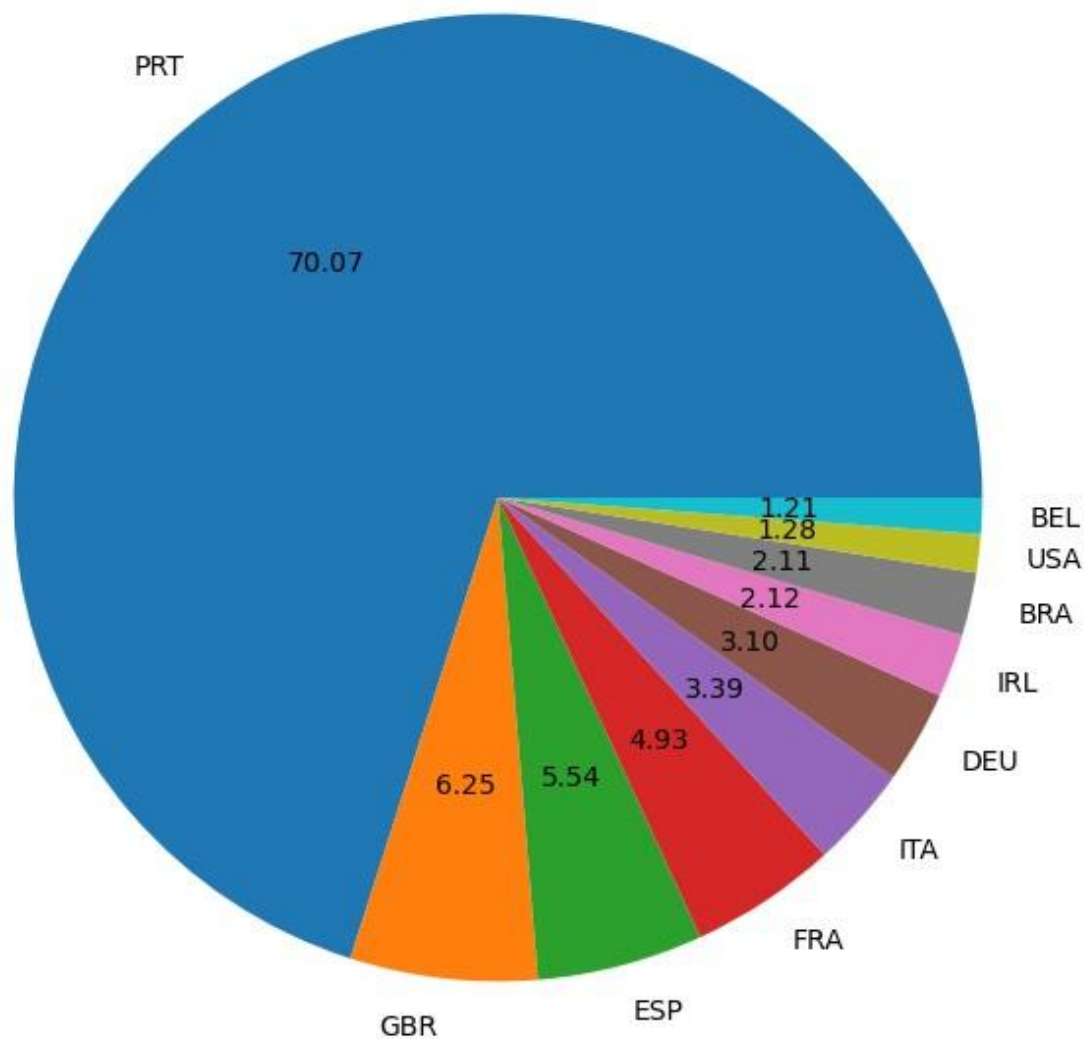Top 10 countries with reservation canceled

Figure: extracted from notebook output

Now, let's see which country has the highest reservation canceled. The top country is Portugal with the highest number of cancellations.

Let's check the area from where guests are visiting the hotels and making reservations. Is it coming from Direct or Groups, Online or Offline Travel Agents? Around 46% of the clients come from online travel agencies, whereas 27% come from groups. Only 4% of clients book hotels directly by visiting them and making reservations.
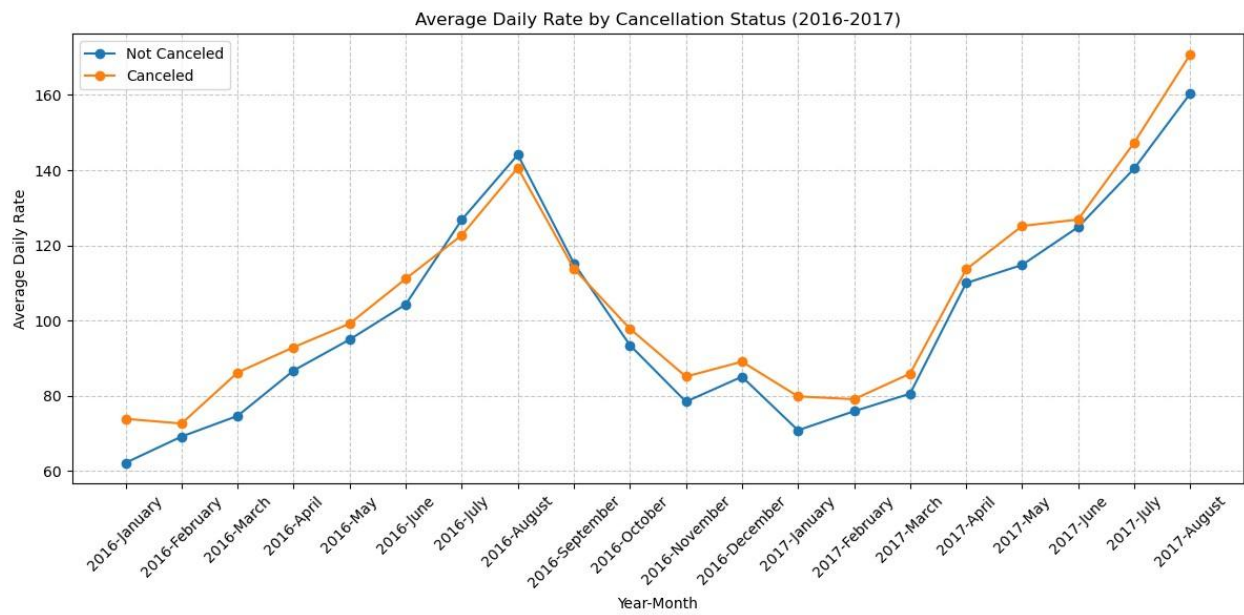
Average Daily Rate by Cancellation Status (2016-2017)

Figure: extracted from notebook output

As seen in the graph, reservations are canceled when the average daily rate is higher than when it is not canceled. It clearly proves all the above analysis, that the higher price leads to higher cancellation.

3. Key Findings

- Dataset contains 118,897 bookings and 31 columns.
- Cancellation rate: 37.1%.
- Average ADR: 101.96 (median 95.00).
- Average lead time: 104.3 days.
- Top countries by bookings: PRT (48585), GBR (12129), FRA (10415), ESP (8568), DEU (7287), ITA (3766), IRL(3375), BEL (2342), BRA (2224), NLD (2104).
- Peak booking months include: August, July, May, October, April.

4. Suggestion

1. Cancellation rates rise as the price does. In order to prevent cancellations of reservations, hotels could work on their pricing strategies and try to lower the rates for specific hotels based on locations. They can also provide some discounts to the consumers.

2. As the ratio of the cancellation and not cancellation of the resort hotel is higher in the resort hotel than the city hotels. So the hotels should provide a reasonable discount on the room prices on weekends or on holidays.

3. In the month of January, hotels can start campaigns or marketing with a reasonable amount to increase their revenue as the cancellation is the highest in this month.

4. They can also increase the quality of their hotels and their services mainly in Portugal to reduce the cancellation rate.

5. Appendix

Metric	Mean
is_canceled	0.37
lead_time	104.31
arrival_date_year	2016.16
arrival_date_week_number	27.17
arrival_date_day_of_month	15.8
stays_in_weekend_nights	0.93
stays_in_week_nights	2.5
adults	1.86
children	0.1
babies	0.01
is_repeated_guest	0.03
previous_cancellations	0.09
previous_bookings_not_canceled	0.13
booking_changes	0.22
days_in_waiting_list	2.33
adr	101.96
required_car_parking_spaces	0.06
total_of_special_requests	0.57
month	6.33