Hotel Booking Cancelation Analysis

Data Analysis & Visualization in Python

By Kareem Elhawary

Project Overview

- **Objective**: Analyze hotel booking data to understand cancelation behavior.
- **Dataset**: 36,285 records of hotel bookings.
- **Tools Used**: Python, Pandas, Seaborn, Matplotlib.
- **Goal**: Identify which factors impact booking cancelation.

Dataset Description

- 17 columns including:
- Booking details (lead time, room nights, price)
- Customer info (adults, children, repeat guest)
- Preferences (meal, room type, segment)
- **Target column**: booking status (Canceled / Not Canceled)

Firstly, The Dataset Was Uploaded To Get The Perfect Features From it

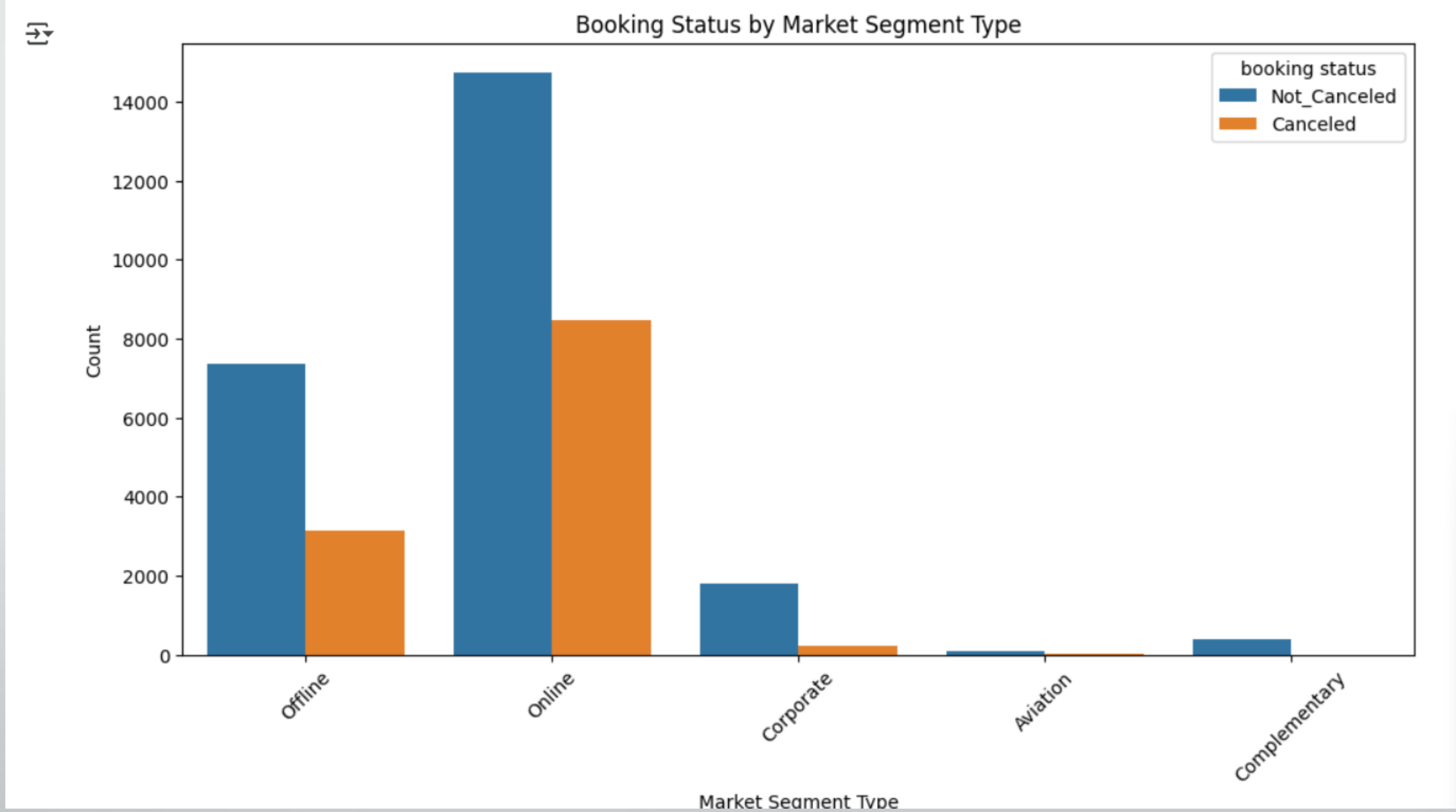
```
import pandas as pd

df = pd.read_csv('first inten project.csv')
df.head()
```

	Booking_ID	number of adults	number of children	number of weekend nights	number of week nights	type of meal	car parking space	room type	lead time	market segment type	repeated	P-C	P-not-C	average price	special requests	date of reservation	booking status
0	INN00001	1	1	2	5	Meal Plan 1	0	Room_Type 1	224	Offline	0	0	0	88.00	0	10/2/2015	Not_Canceled
1	INN00002	1	0	1	3	Not Selected	0	Room_Type 1	5	Online	0	0	0	106.68	1	11/6/2018	Not_Canceled
2	INN00003	2	1	1	3	Meal Plan 1	0	Room_Type 1	1	Online	0	0	0	50.00	0	2/28/2018	Canceled
3	INN00004	1	0	0	2	Meal Plan 1	0	Room_Type 1	211	Online	0	0	0	100.00	1	5/20/2017	Canceled
4	INN00005	1	0	1	2	Not Selected	0	Room_Type 1	48	Online	0	0	0	77.00	0	4/11/2018	Canceled

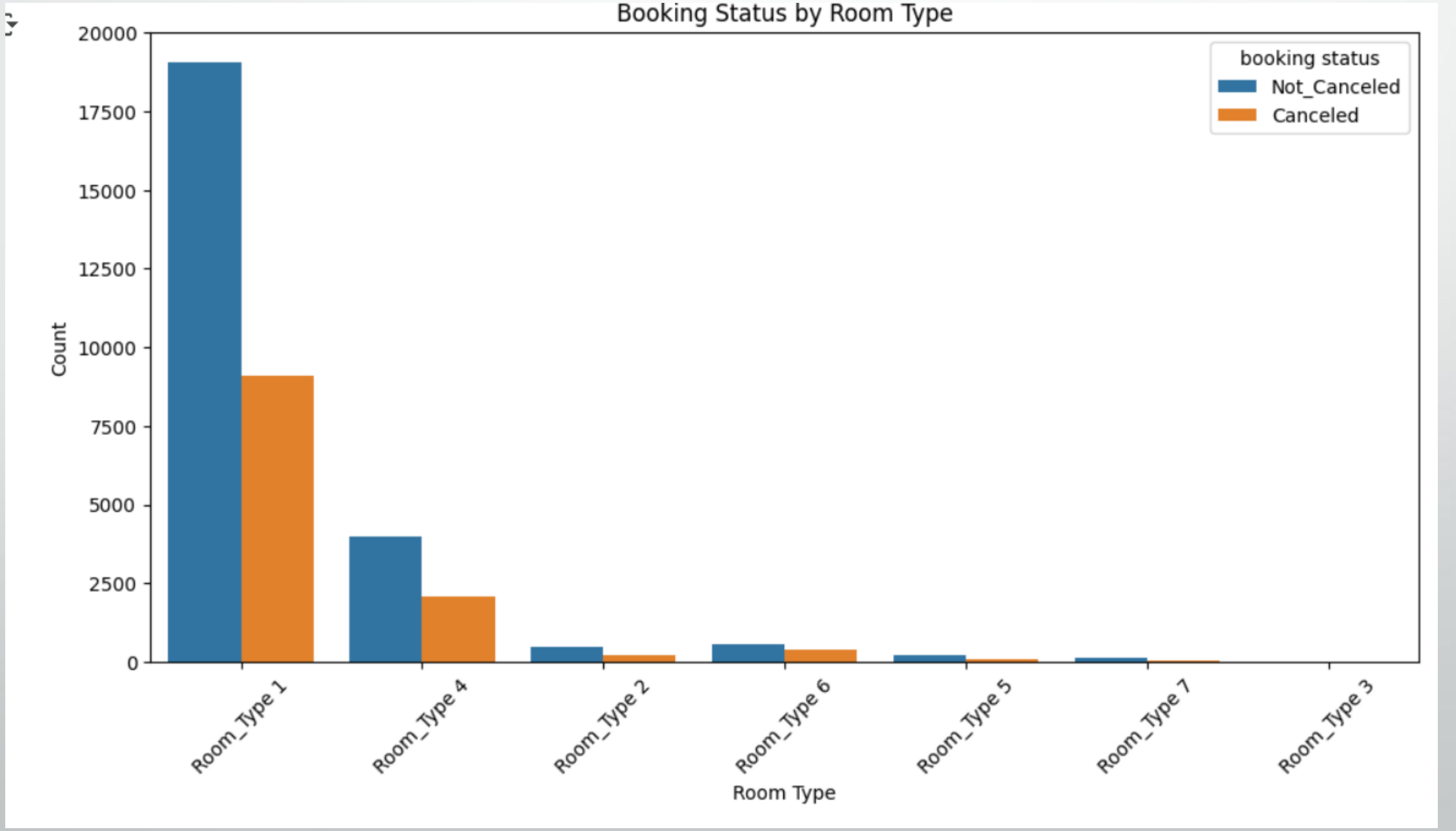
Market Segment vs Booking Status :

The "Online" segment has the highest number of bookings and the highest cancelation count
Offline and Corporate bookings show much lower cancelation rates, indicating more reliable channels.



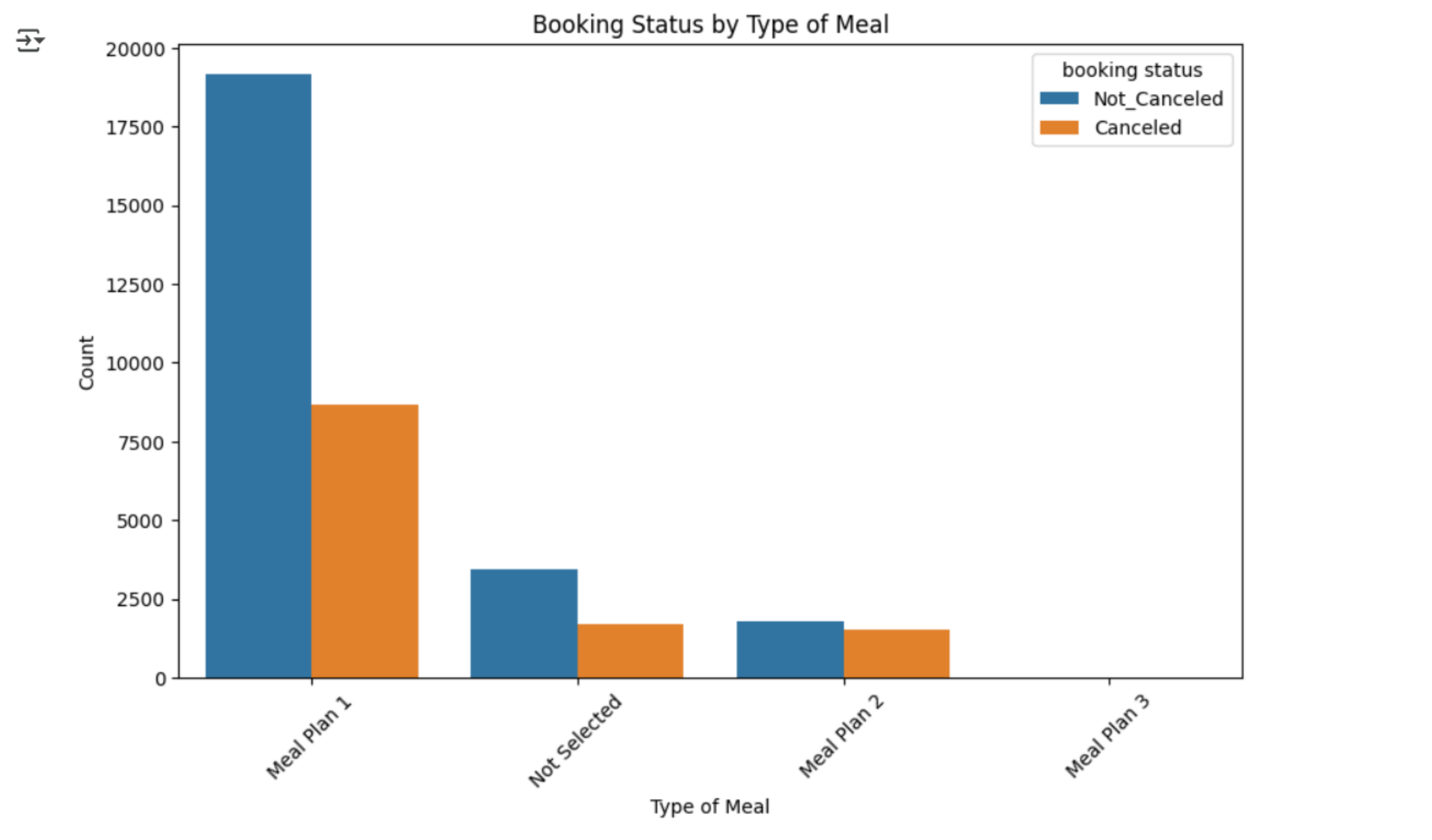
Room Type vs Booking Status :

Room Type 1 is the most booked and most canceled, likely because it's the default or cheapest option. Other room types show fewer cancelations, possibly reflecting more serious or business travelers.



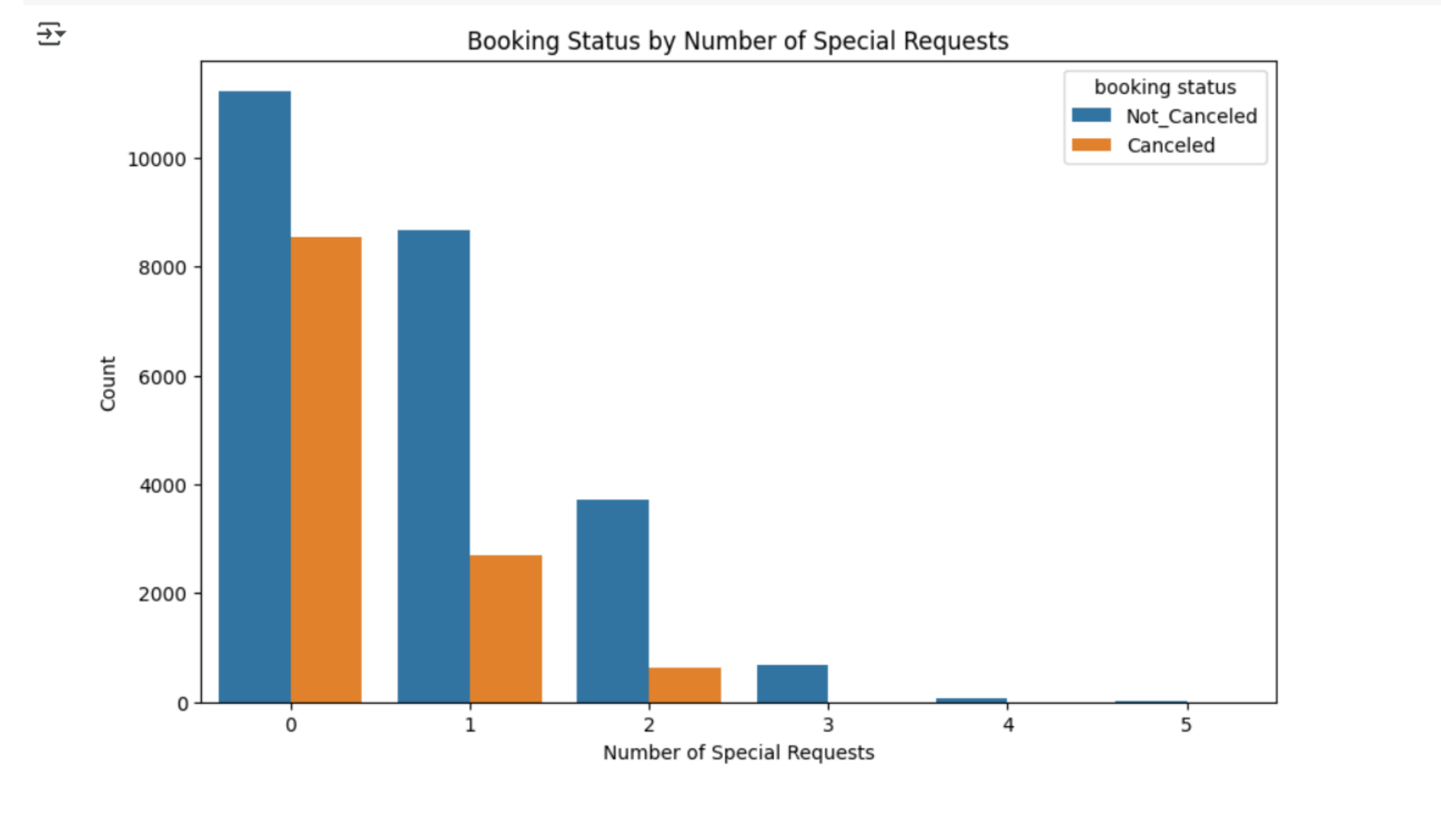
Meal Plan vs Booking Status :

The majority of guests chose "Meal Plan 1", followed by "Not Selected ".
Canceled bookings are slightly more common among guests who did not select a meal plan, possibly indicating lower commitment.



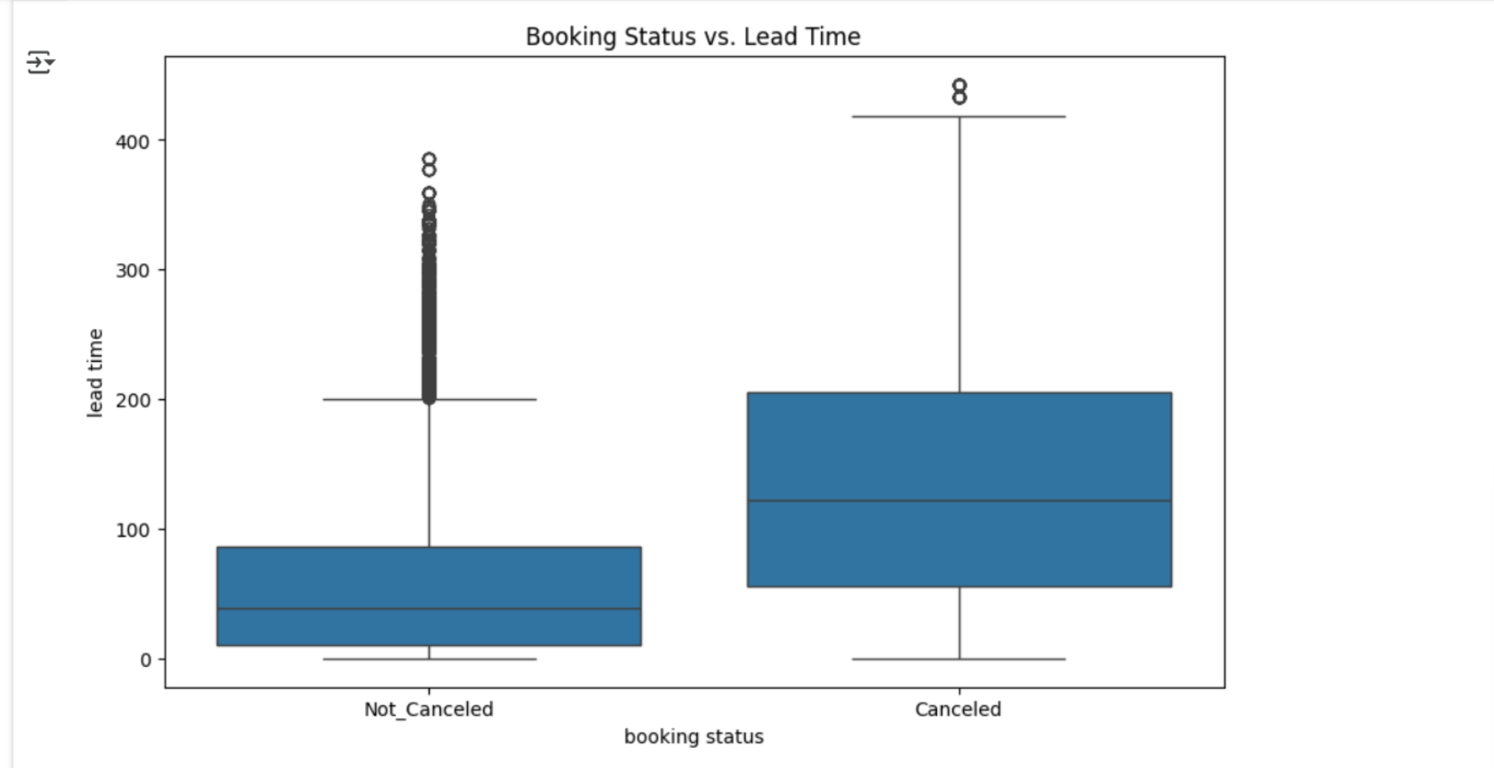
Special Requests vs Booking Status :

Guests with more special requests are less likely to cancel ,This could indicate stronger commitment or personalization needs that reduce cancelation likelihood.



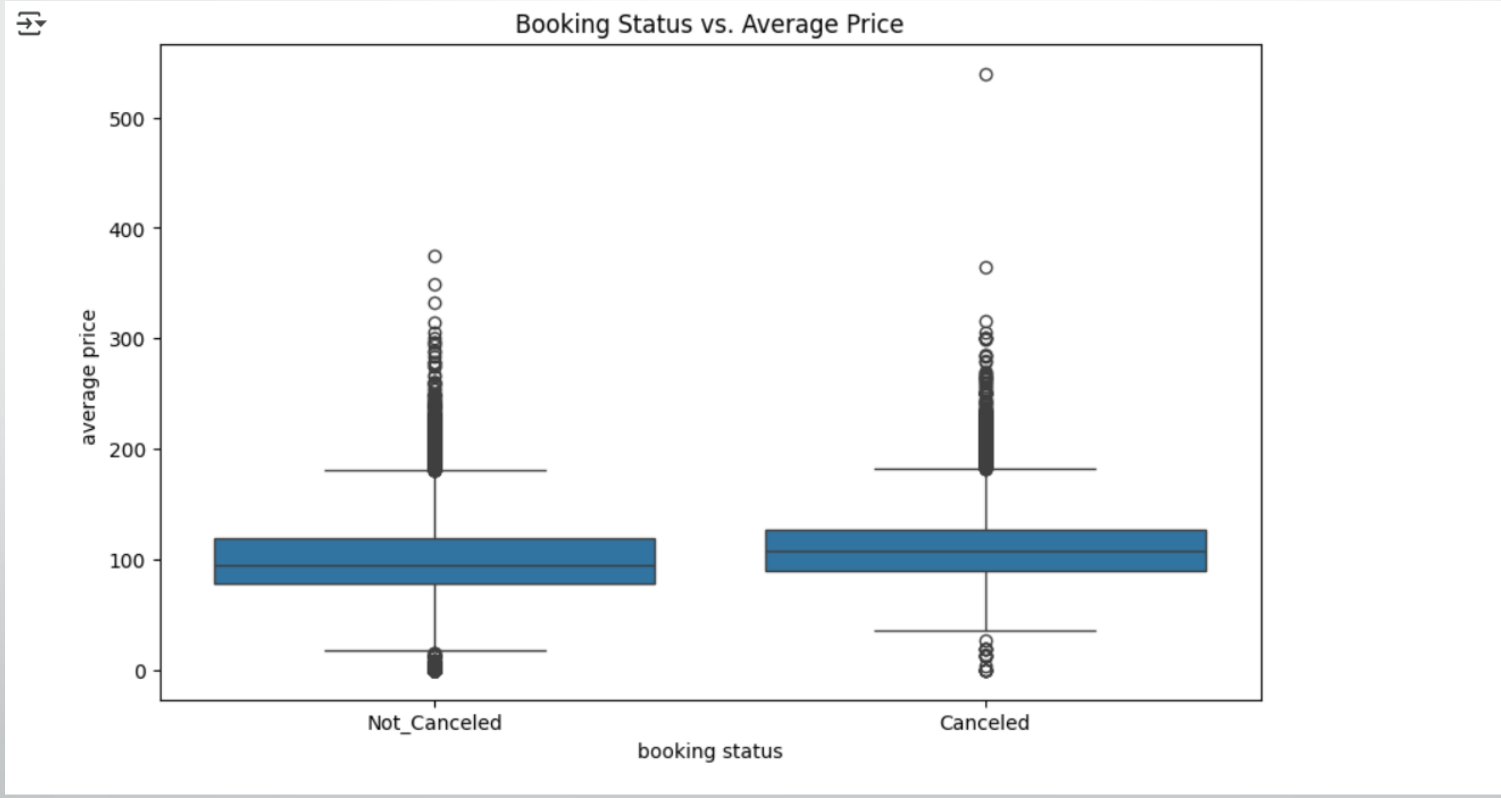
Lead Time vs Booking Status :

Canceled bookings tend to have much longer lead times (days between booking and arrival).This suggests guests who book far in advance are more likely to cancel later.



Average Price vs Booking Status :

Most bookings are priced between $50 and $150, with a few expensive outliers,There's no clear direct link between high price and cancelation, but price could interact with other features.



Executive Summary

- A total of 36,285 bookings were analyzed from a hotel dataset.
- The cancelation rate is ~28%, mainly among online bookings.
- Key features examined: lead time, special requests, meal plan, market segment, and room type.
- Goal: Understand patterns that lead to cancelation and identify actionable insights for reducing it.

Conclusions

- Longer lead time significantly increases cancelation risk.
- Guests with more special requests are less likely to cancel.
- Online bookings have the highest cancelation rate.
- Certain room types and meal plans are associated with higher cancelation behavior.
- These insights can help the hotel tailor pricing, marketing, and booking policies to minimize loss.

Predictive Modeling

- A machine learning model (e.g., Logistic Regression, Decision Tree) can be built to predict cancelation.
- Input features: lead time, room type, number of nights, special requests, market segment, etc.
- Output: Booking Status (Canceled or Not_Canceled)Benefits:
- Predict cancelation before arrival.
- Take proactive actions: follow-up calls, confirmations, or flexible rebooking offers.