

Presented by Adham Roushdy

# EXPLORATORY DATA ANALYSIS (EDA)

**Objective:**  
Analyze hotel booking data to identify trends,  
customer behavior, and factors influencing  
cancellations.

---



# Dataset Overview

---

## Data Structure:

- 36,285 bookings
- 17 columns, including:
  - Booking details (adults, children, stay duration)
  - Room type, meal plan, parking space
  - Lead time, market segment, booking status

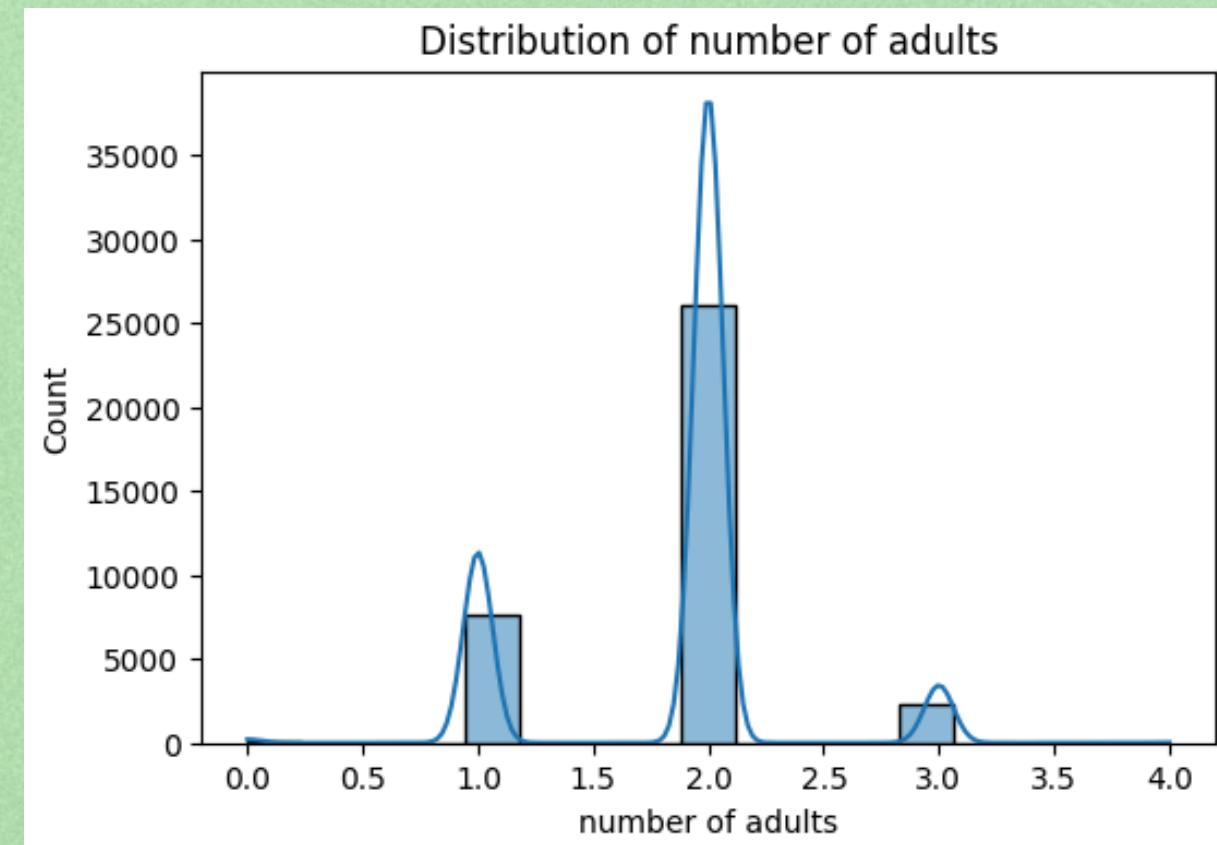
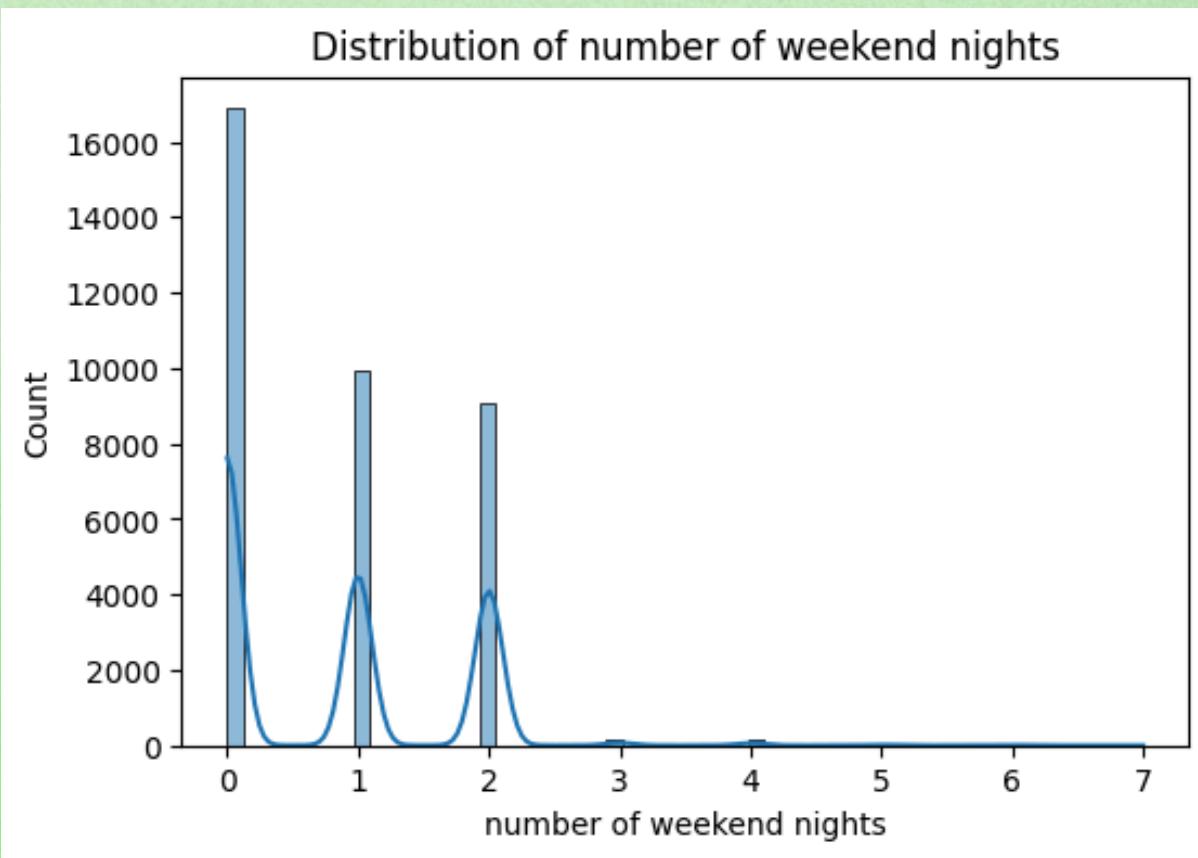
## Data Quality:

- No missing values
- Clean and structured for analysis

# Distribution of Numerical Variables

For each numeric variable, I've created a histogram with KDE line to show:

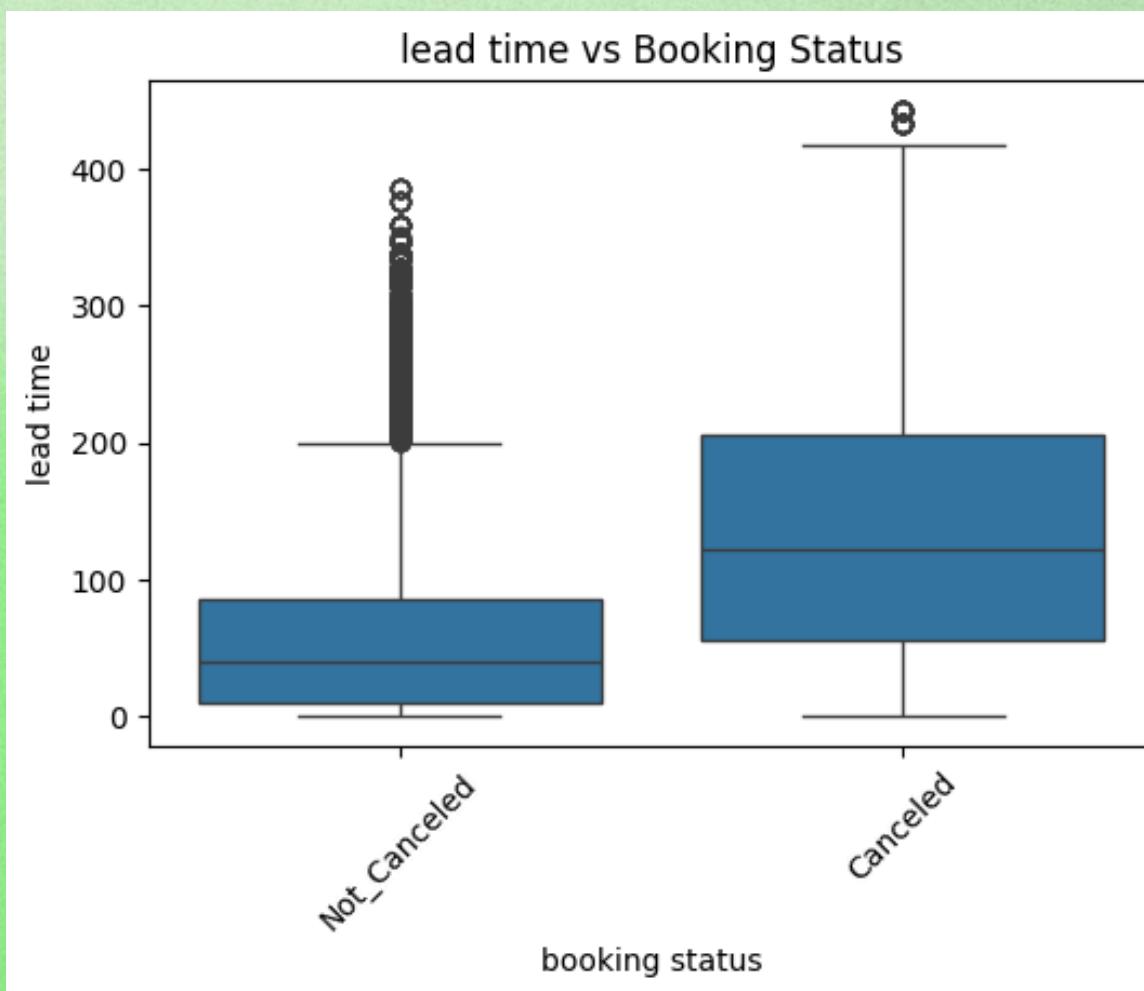
- The spread of values
- Outliers/skewness



# Numerical Features vs Booking Status

## Visualization Approach:

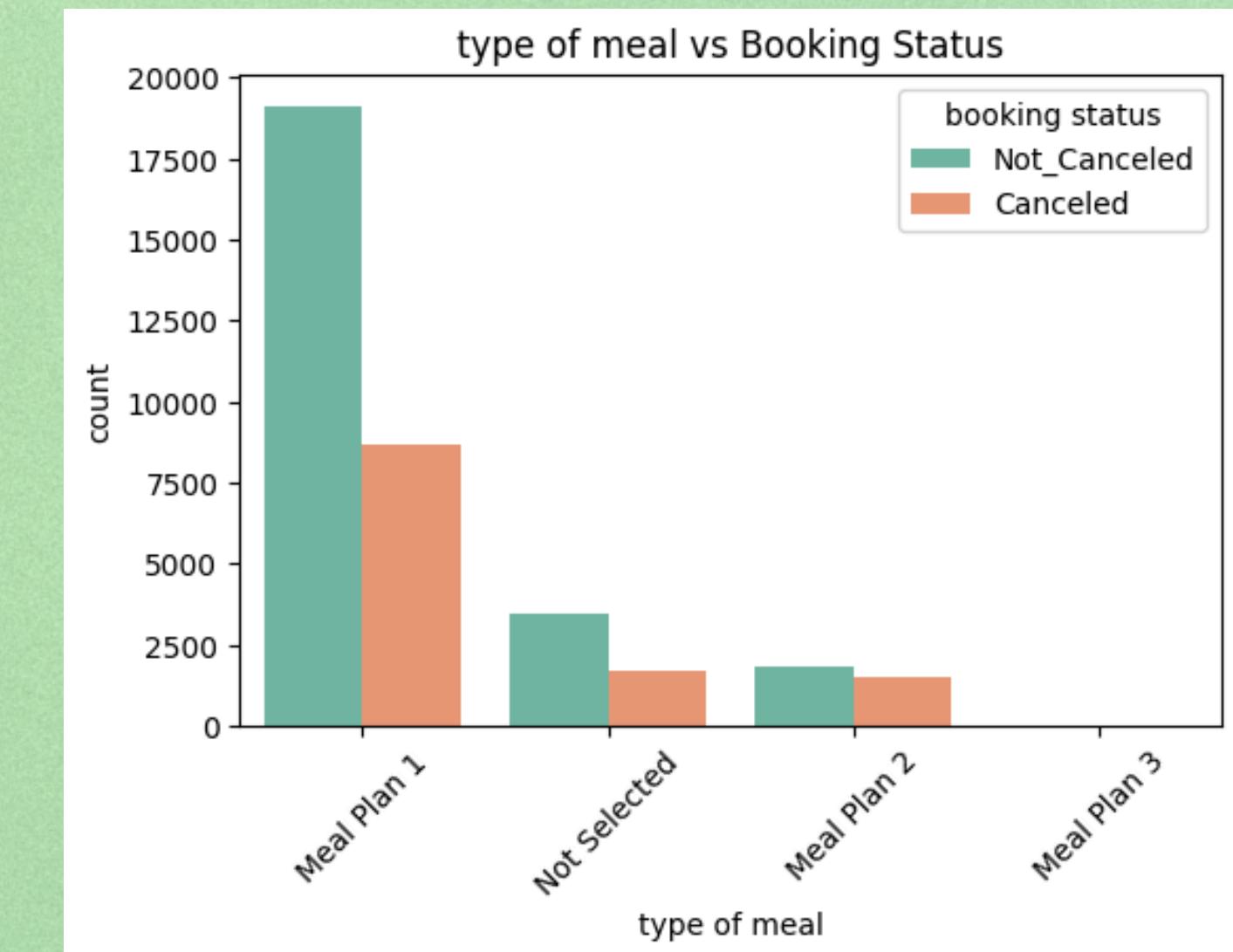
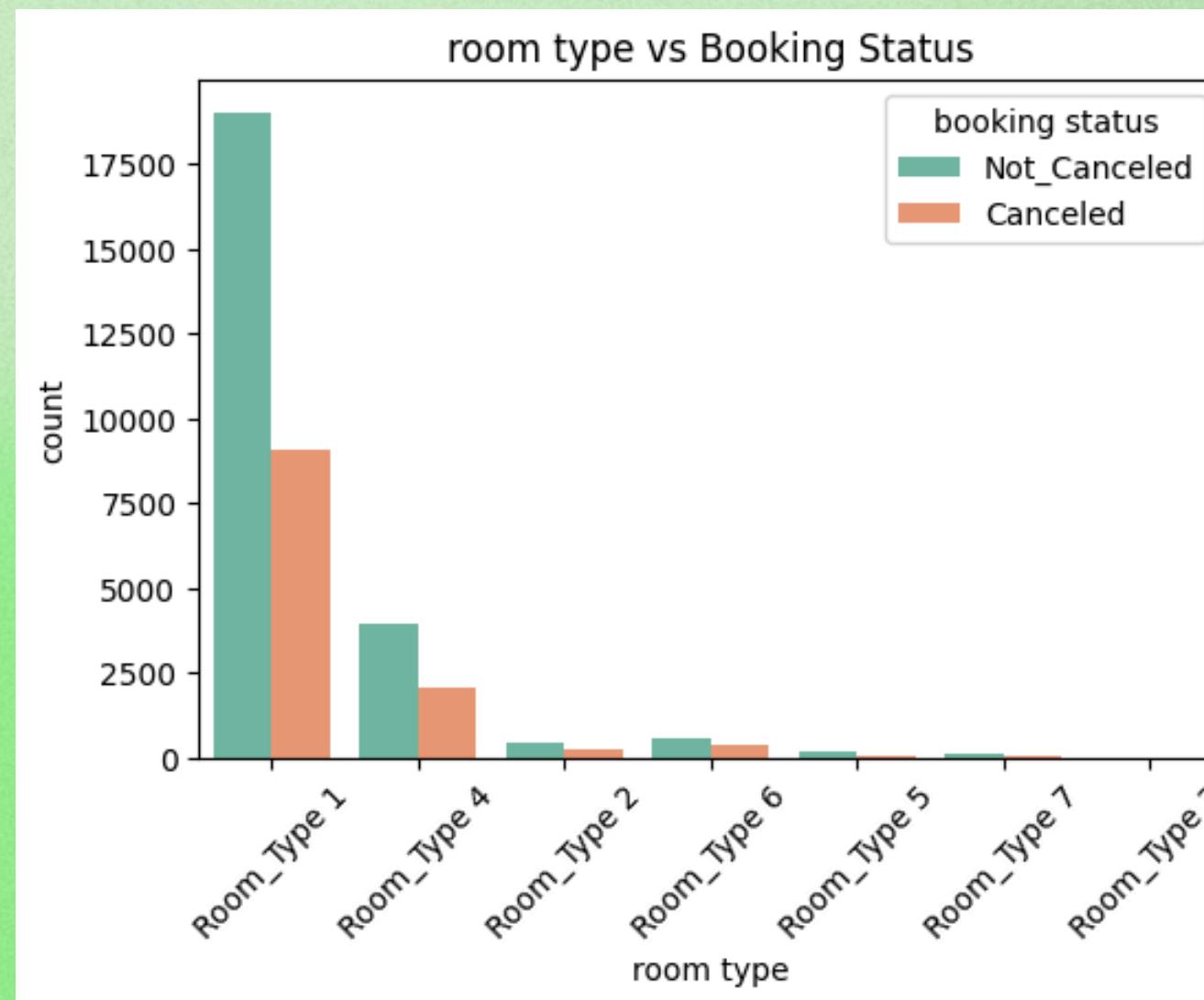
- **Boxplots** comparing each numeric variable between:
  - Canceled bookings
  - Not canceled bookings
- Shows median, quartiles, and outliers for each group



# Categorical Features vs Booking Status

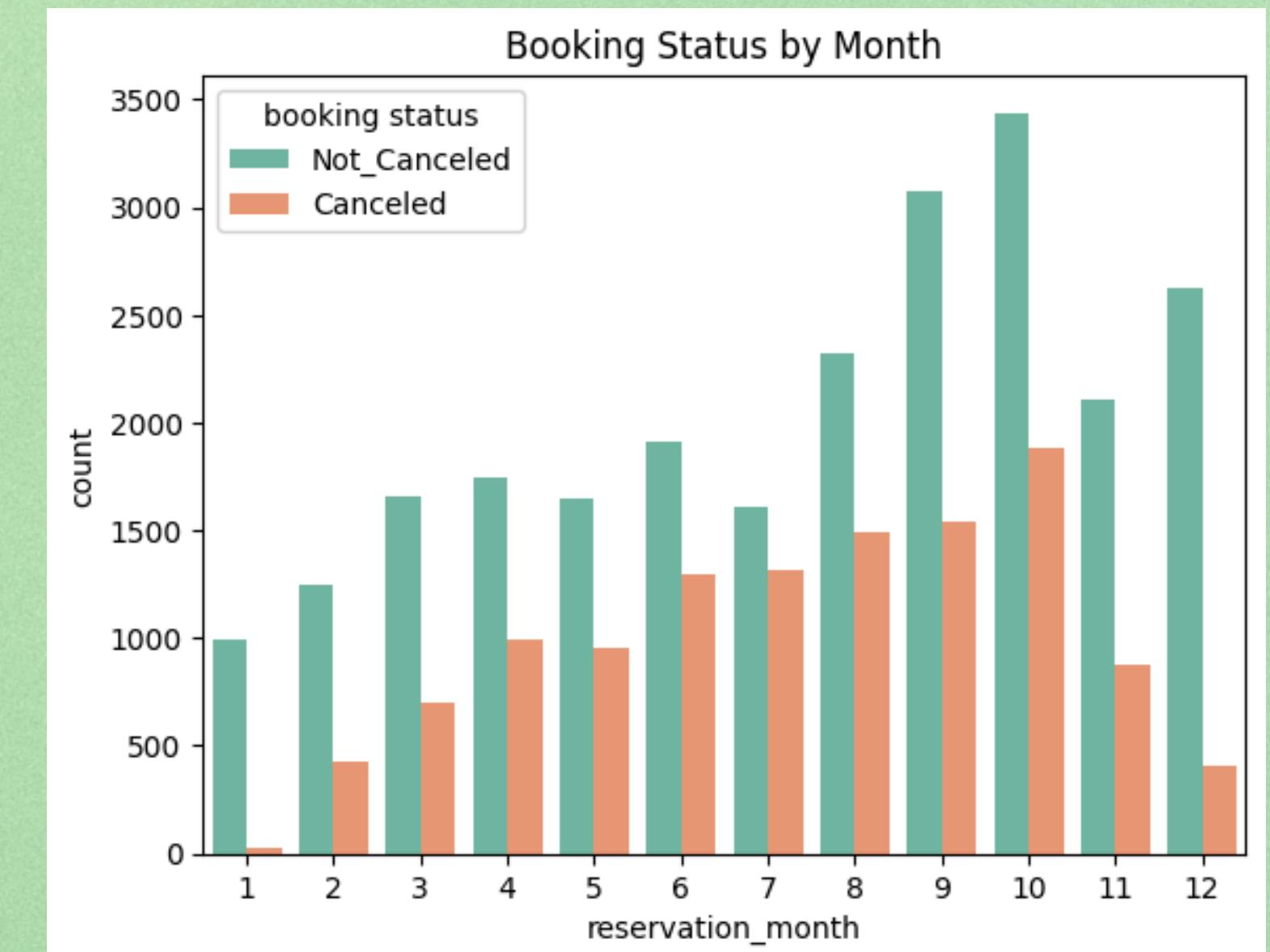
## Visualization Approach

- **Countplots** showing distribution of each category split by booking status
- **Color-coded** (Canceled vs Not Canceled) for easy comparison



# Add new column

- **Extracting** Reservation Month from Date
- **Create** a new column showing the month number (1 to 12) of each reservation.
- **countplot** to count the number of reservations for each month.

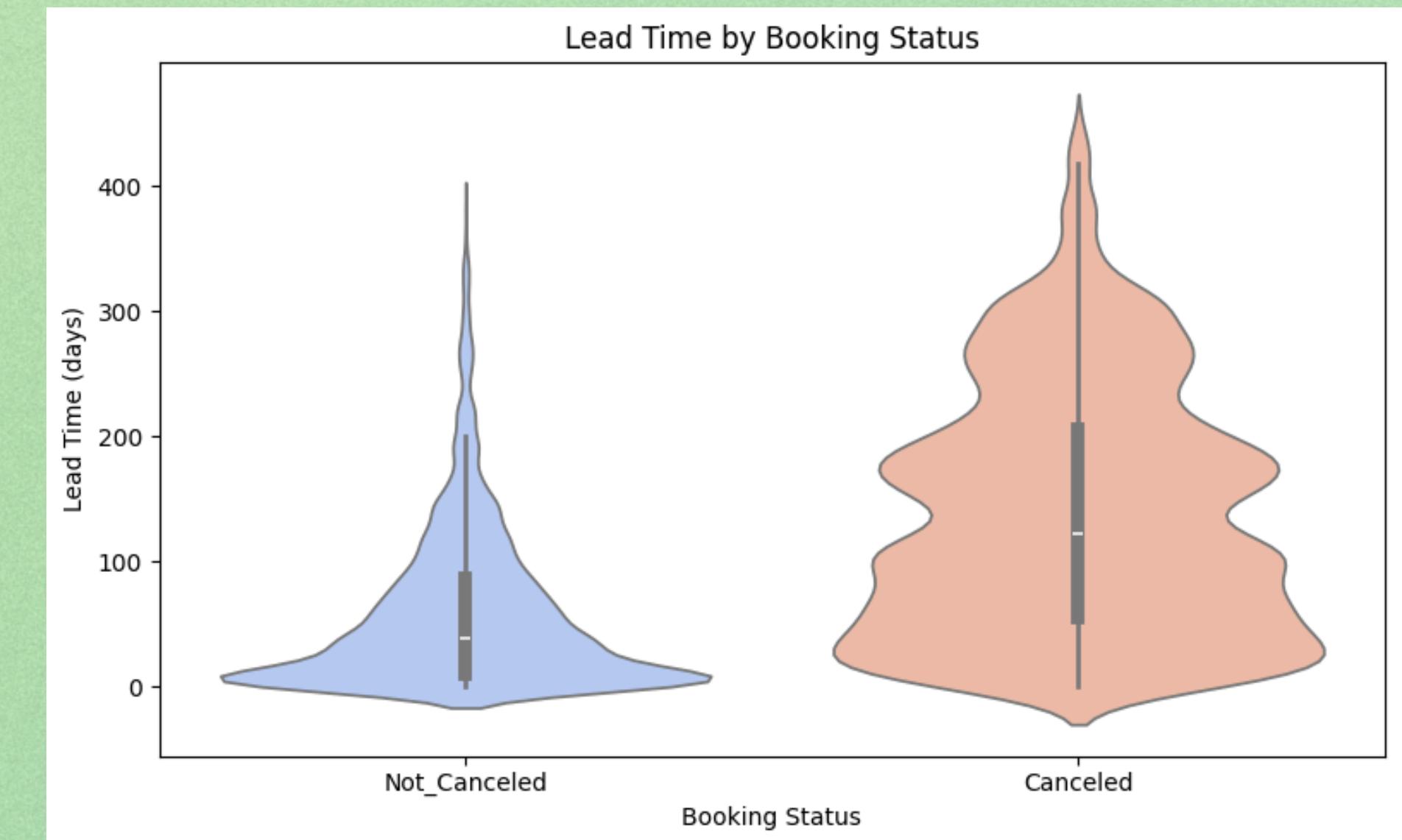


# lead time's impact on booking status

**Draws a violin plot** to show the distribution of lead time for each booking status.

Why a Violin Plot?

- Shows both distribution shape (like KDE) and
- summary statistics (like boxplot).



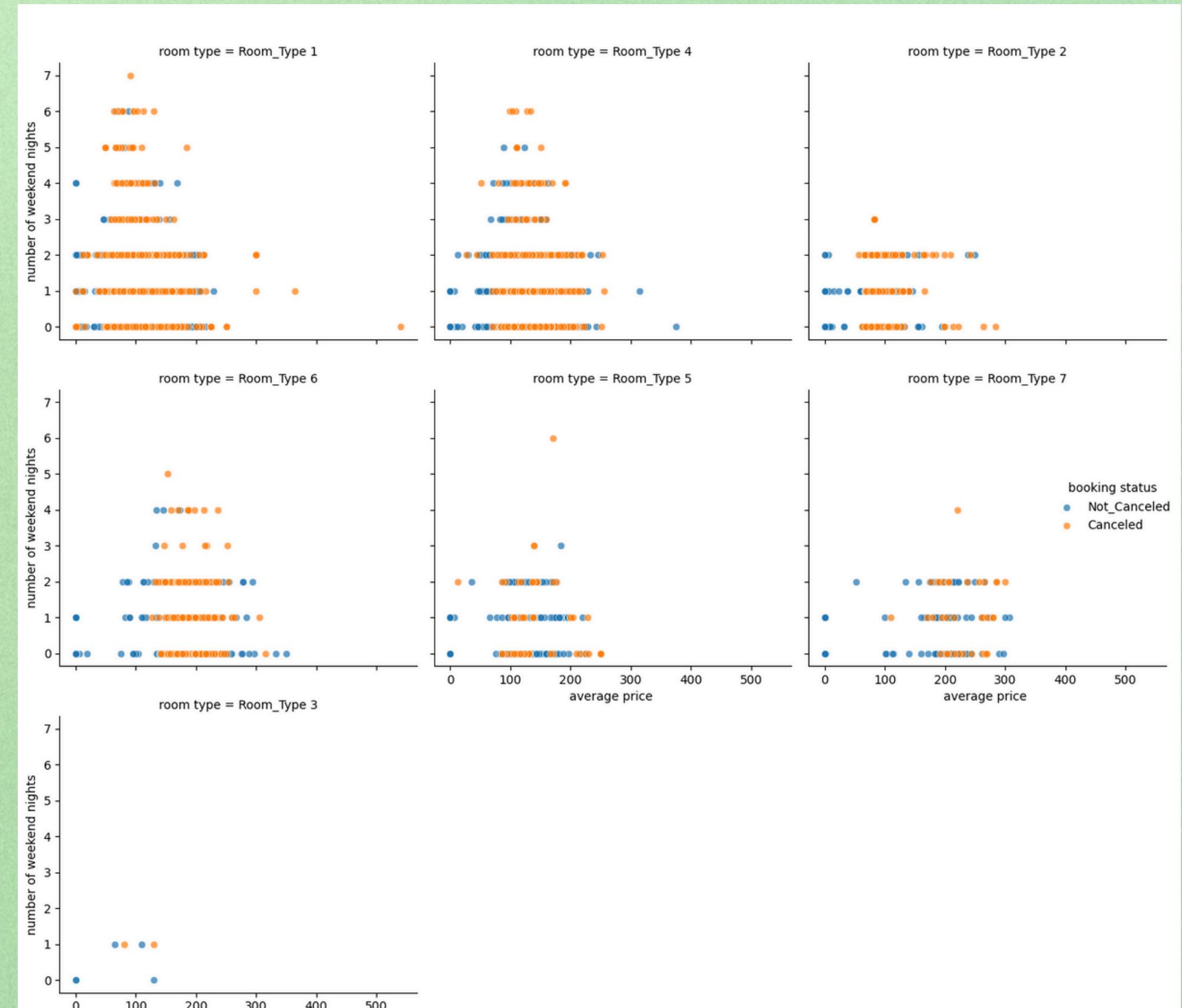
# Booking Behavior: Weekend Nights vs. Price by Room Type"

- Creates a grid of **scatter plots**, one for each room type.
- Colors the points based on booking status.

Why This is Useful:

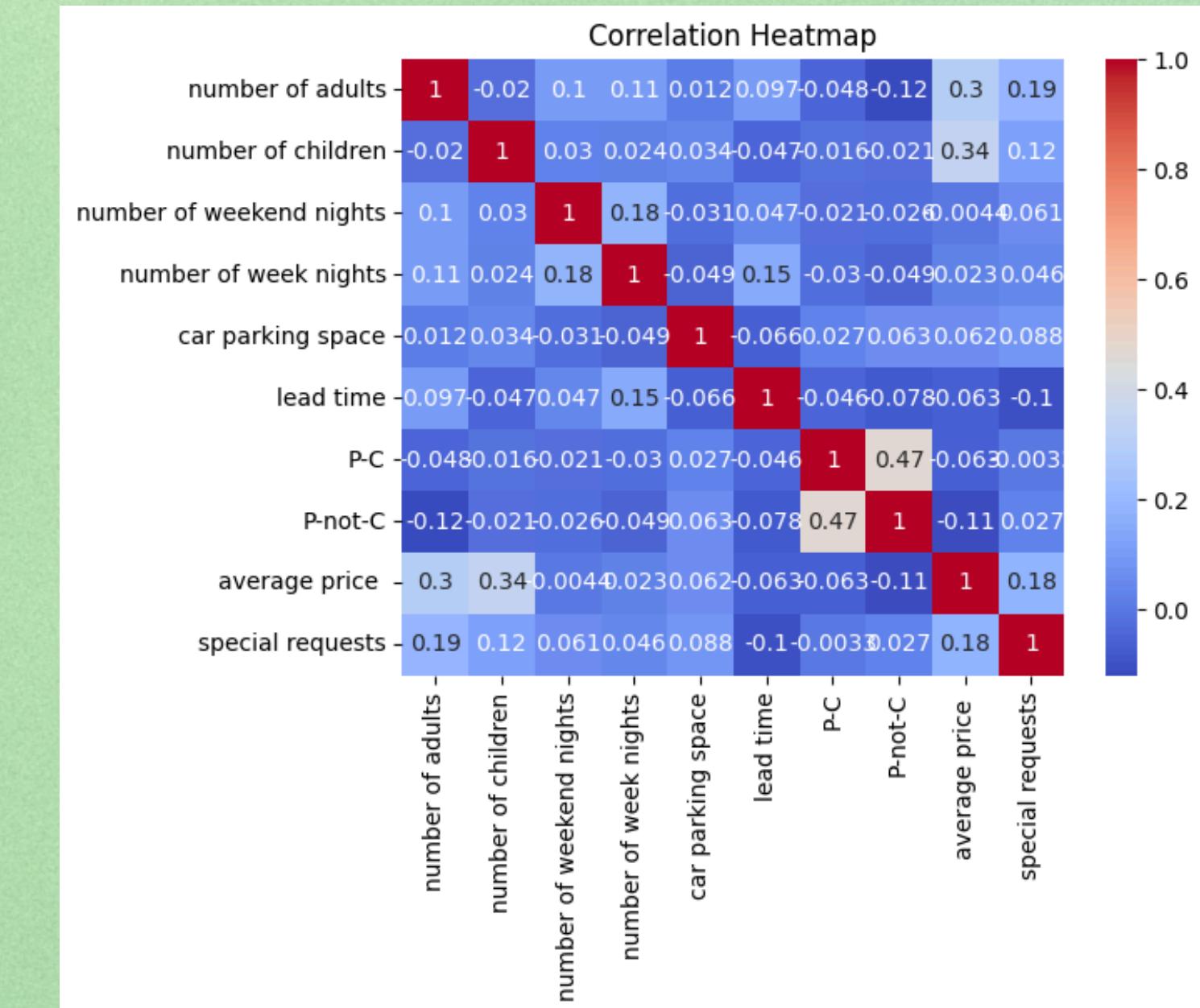
- Helps identify price patterns related to weekend stay duration.
- Allows us to compare booking behaviors for each room type.

# Booking Behavior: Weekend Nights vs. Price by Room Type"



# Correlation Heatmap of Numeric Features

- Calculates correlations between all numeric columns using `.corr()`.
- Displays a heatmap where:
- Red = Strong positive correlation
- Blue = Strong negative correlation



**THANK YOU**