

EDA-Exploratory Data Analysis By Anusha N

1. Importing and loading data

First, we need to import the necessary libraries and load the data into a Pandas dataframe. Since the dataset is already cleaned there are no missing values.

1. Dataset Overview

- **Rows:** 3,271
 - **Columns:** 14
 - **Date Range:** April 1, 2018 – June 30, 2018 (3 months)
-

2. Data Quality

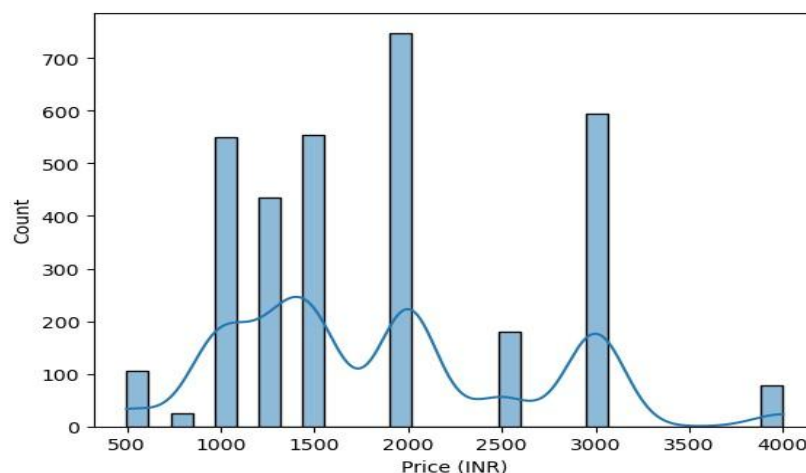
- **Clean dataset:** No missing values as previously cleaned.
 - **Consistent date formats:** All booking dates fall within the 3-month window.
-

3. Booking Trends

- **Average Capacity (MaxBookees):** ~30 seats per event
 - **Average Actual Bookings:** ~16–17 seats per event (~55% utilization)
 - **Key Insight:** Underutilization of event capacity is common.
-

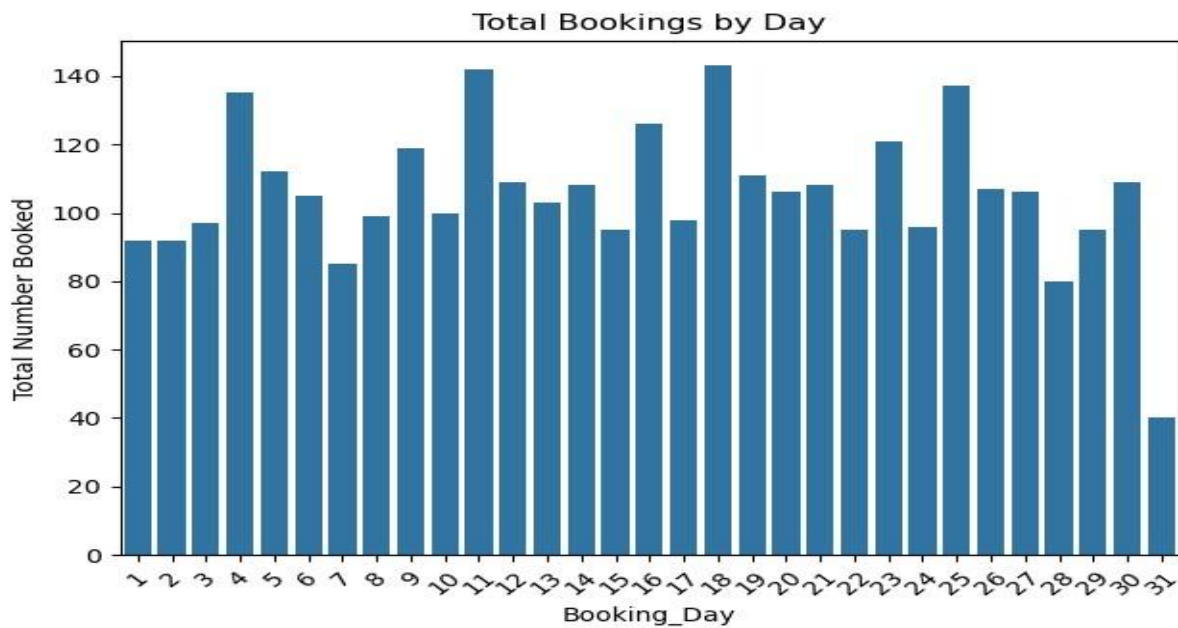
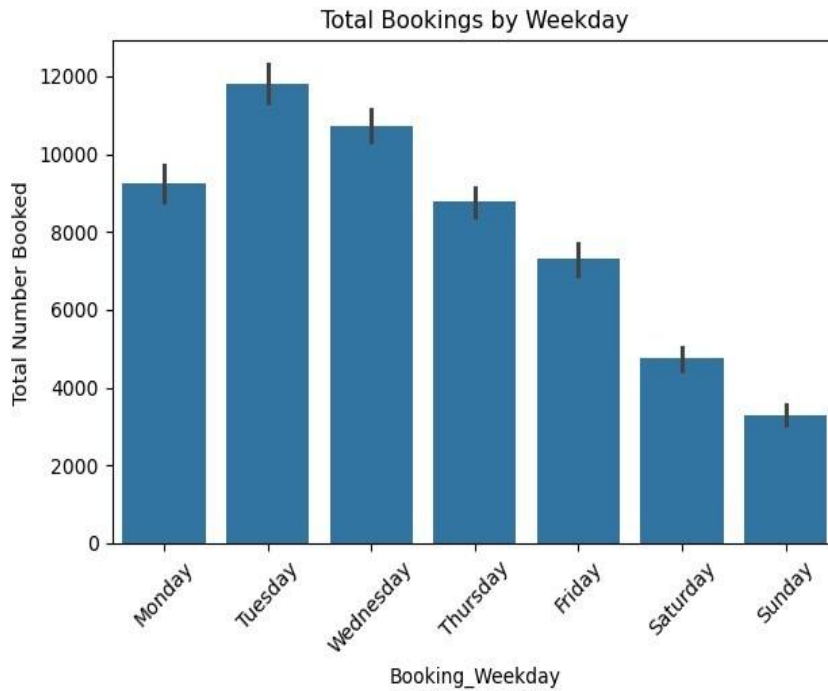
4. Pricing Analysis

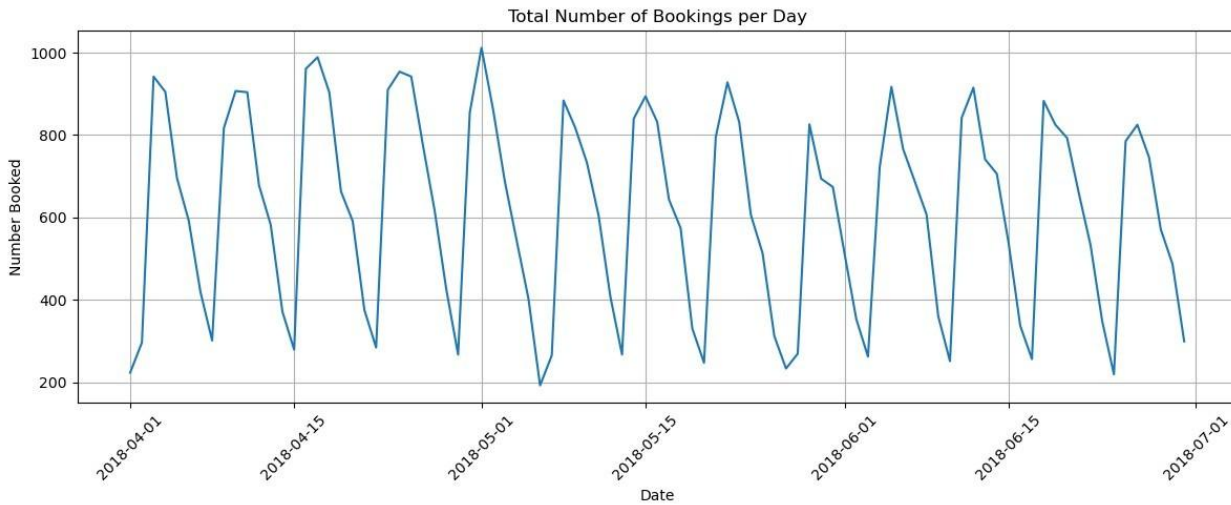
- **Average Price Range:** ₹1499 – ₹1852
- **Maximum Price:** ₹3300–₹3999 (*outliers*)
- **Price per Seat:** Mostly below ₹100 → Budget-friendly strategy



5. Temporal Patterns

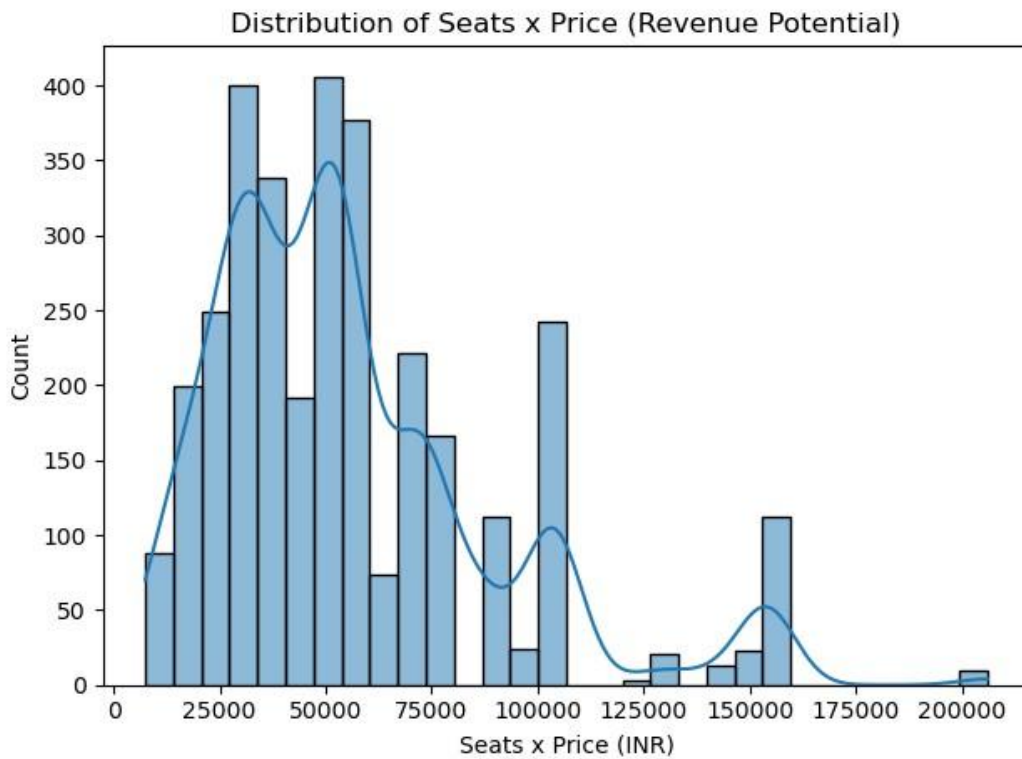
- **Peak Booking Month: May**
- **Least Bookings:** Occur towards end of each month
- **Booking Frequency:**
 - **Weekdays:** 200–300 bookings/day
 - **Weekends/Spikes:** 800–1000 bookings/day

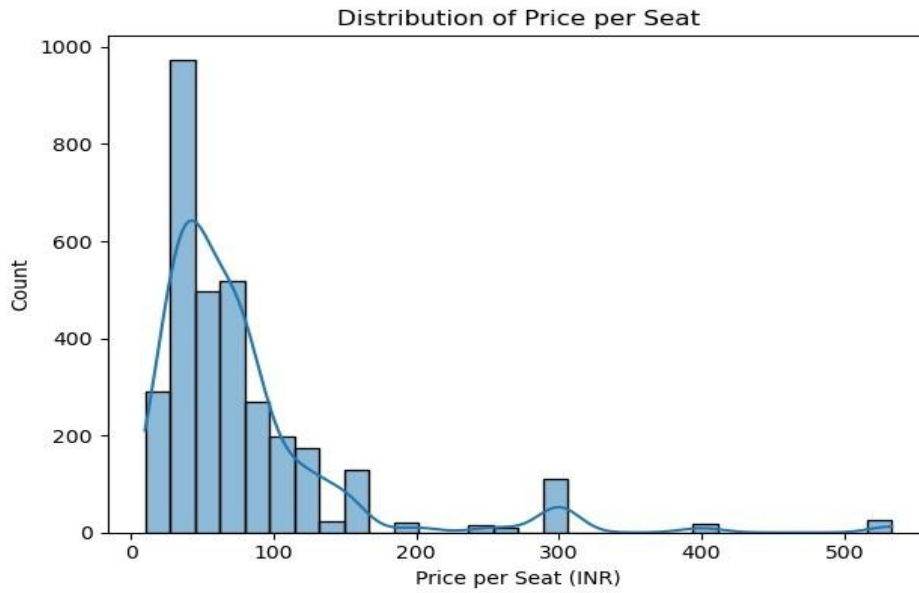




6. Revenue Insights

- **Typical Revenue Range:** ₹25,000 – ₹100,000
- **High-Value Events:** Up to ₹2,00,000 (*outliers*)
- **Insight:** Majority of events are mid-tier in revenue → Ideal for tiered pricing/marketing strategies

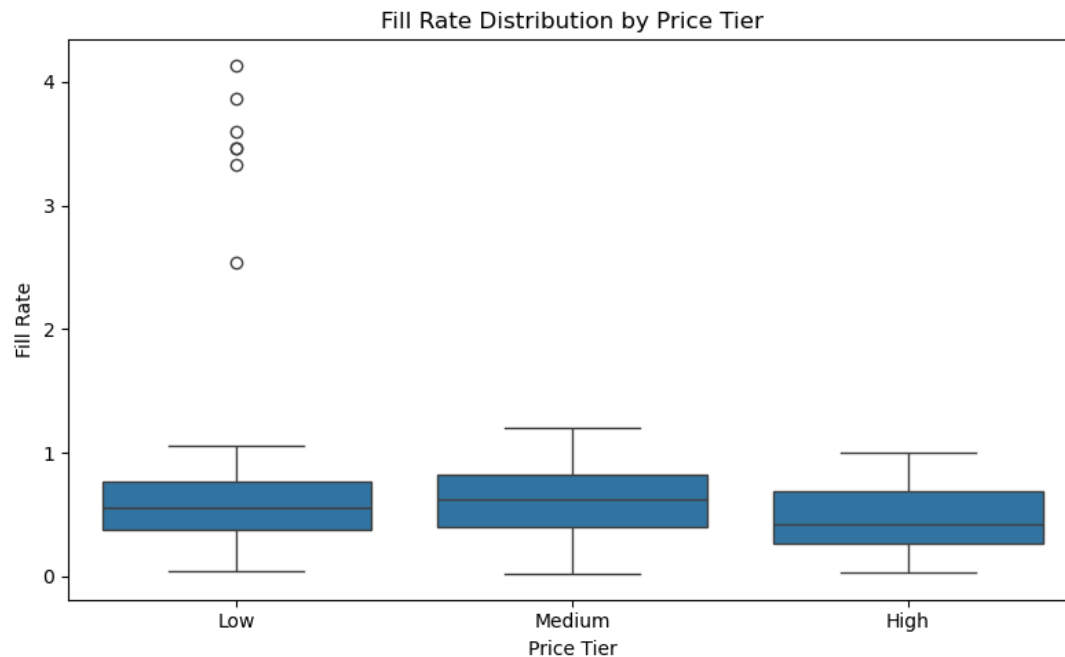




Fill Rate vs Price Tier

Interpretation:

- Median fill rate remains similar across **Low**, **Medium**, and **High** price tiers.
- **Low-priced events** show more **outliers** with very high fill rates.
- Suggests budget events are more likely to oversell or reach max capacity.



B. Pair Plot of All Numeric Features

Interpretation:

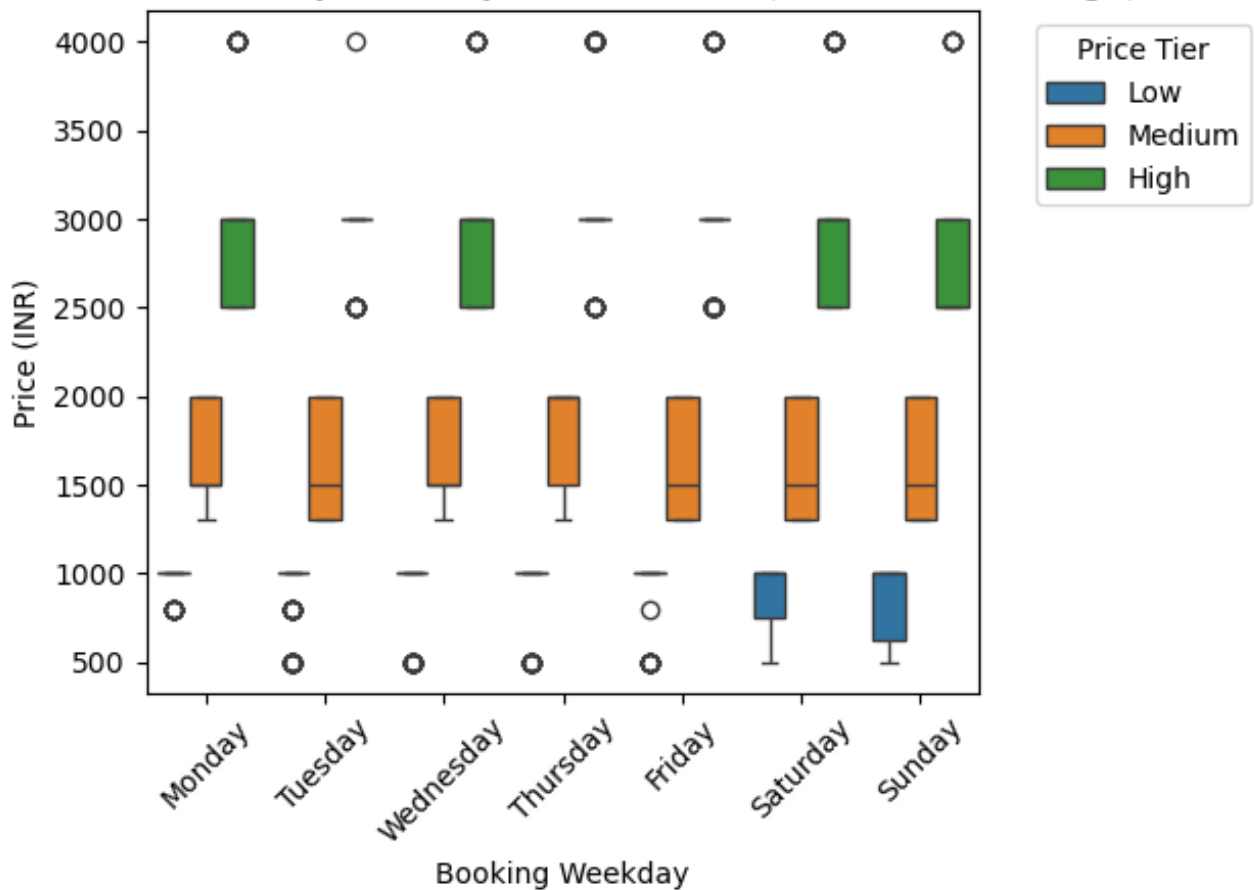
- Some variables show **nonlinear relationships** (e.g., bookings vs price, bookings vs capacity)
- `MaxBookees` and `Number Booked` show a clear **positive correlation**
- **Price** shows no strong direct trend with most features, indicating room for segmentation

C. Price Distribution by Weekday and Tier

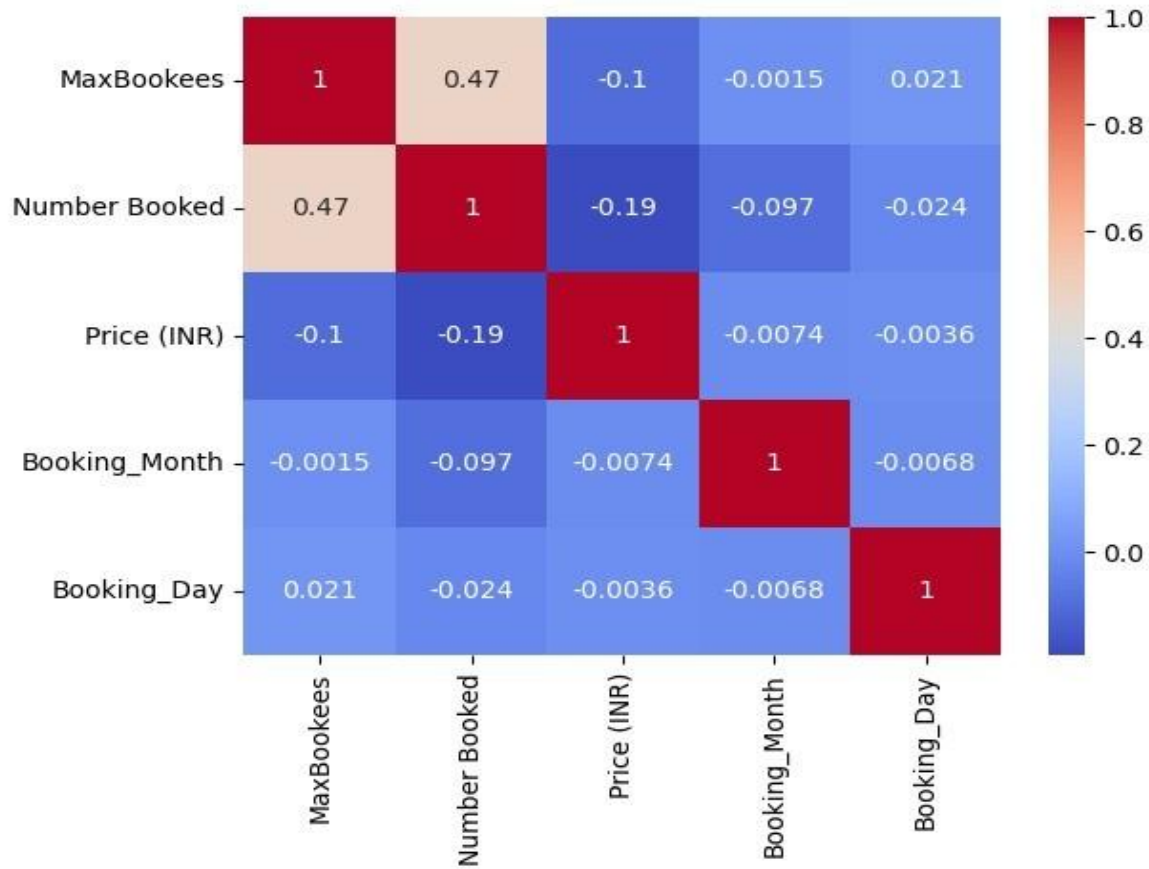
Interpretation:

- **Low-priced events** are mostly on weekends (Saturday/Sunday)
- **Medium and High-priced events** are more uniformly distributed throughout the week
- Could indicate **weekend promotions or budget targeting**

Price Distribution by Weekday with Price Tier (Low / Medium / High)



7. Feature Correlations



Feature Pair	Correlation	Interpretation
MaxBookees ↔ Number Booked	+0.47	Moderate positive — bigger capacity, more bookings
Price (INR) ↔ Number Booked	-0.19	Weak negative — slightly fewer bookings with higher prices
MaxBookees ↔ Price (INR)	-0.10	Very weak negative — higher-capacity events often lower-priced (bulk pricing)
Booking Month/Day ↔ Bookings	~0	No meaningful linear relationship with bookings

8. Key Business Insights

- Events are **underbooked on average**, pointing to possible issues in marketing or overestimation of demand.
- May** is a hot month for event engagement → Focus promotions during this time.
- Low price per seat** dominance suggests events are **budget-focused**, appealing to mass market.
- Revenue clustering** between ₹25K–₹100K enables **targeted marketing strategies** for mid-tier events.

