**Hotel Booking Data Analysis**

**Submitted by: Lalit Bhamare**

**1. Summary**

This report examines hotel booking data to uncover trends that help improve **pricing, customer experience, and operational efficiency**. It focuses on booking behaviors, cancellations, pricing strategies, and customer preferences.

**2. Data Overview**

The dataset includes **guest bookings, room assignments, pricing, cancellations, and special requests**, providing insights into hotel operations.

**2.1 Column Descriptions**

Below is a summary of key dataset columns:

| **Column Name** | **Description** |
| --- | --- |
| **hotel** | Type of hotel: "Resort Hotel" or "City Hotel" |
| **is\_canceled** | Whether the booking was canceled (1) or not (0) |
| **lead\_time** | Days between booking and arrival date |
| **adults** | Number of adults in booking |
| **children** | Number of children in booking |
| **babies** | Number of babies in booking |
| **meal** | Meal plan booked (SC, BB, HB, FB, Undefined) |
| **market\_segment** | Booking source (e.g., Direct, Corporate, Online TA) |
| **distribution\_channel** | How booking was made (Direct, TA/TO) |
| **reserved\_room\_type** | Room type initially reserved |
| **assigned\_room\_type** | Room type actually assigned |
| **booking\_changes** | Number of booking modifications before check-in |
| **deposit\_type** | Type of deposit made: No Deposit, Non Refund, Refundable |
| **adr** | Average Daily Rate (lodging revenue / total nights stayed) |
| **total\_of\_special\_requests** | Number of special requests made |
| **reservation\_status** | Final booking status: Canceled, Check-Out, No-Show |
| **reservation\_status\_date** | Date of last update to booking status |

**3. Data Cleaning & Processing**

* **Handled missing values** for children, country, agent, and company.
* **Formatted dates** for better tracking and analysis.
* **Removed duplicate records** to maintain data accuracy.

**4. Exploratory Data Analysis (EDA)**

* **ADR (Average Daily Rate):** Shows pricing trends, seasonal fluctuations, and differences across booking sources.
* **Market Segment Analysis:** Highlights variations in ADR across different customer types and booking channels.

**5. Correlation Analysis**

* **Positive correlations:** Families (adults & children) tend to **book longer stays**.
* **Negative correlations:** Lower-priced bookings have a **higher chance of cancellation**.
* **Multicollinearity check:** Helps refine predictive models.

**6. Hypothesis Testing**

**6.1 ADR Differences by Booking Channel**

* **Null Hypothesis (H0):** significant difference in ADR between **Online TA** and **Direct** bookings.
* **Alternative Hypothesis (H1):** No is a significant difference.
* **Findings:**
  + **T-statistic:** -0.1822089487704939
  + **P-value:** 0.8554209429620393 (Fail to reject H0)
* **Insight:** **Online TA bookings may have different pricing strategies than Direct reservations** due to discounts or package deals.

**6.2 Room Upgrades and Booking Lead Time**

* **Null Hypothesis (H0):** Room upgrades are **independent** of lead time.
* **Alternative Hypothesis (H1):** There is a significant difference.
* **Findings:**
  + **F-statistic:** 2482.1374551581357
  + **P-value:** 0.0 (very low, rejecting H0)
* **Insight:** **Shorter lead-time bookings may get more upgrades** due to last-minute availability or hotel policies.

**6.3 Stay Duration by Customer Type**

* **Null Hypothesis (H0):** No significant difference in **stay duration** between **Transient** and **Group** customers.
* **Alternative Hypothesis (H1):** There is a difference.
* **Findings:**
  + **F-statistic:** 943.7598907118576
  + **P-value:** 0.0 (rejecting H0)
* **Insight:** **Group customers tend to stay longer than transient ones**, affecting pricing and revenue strategies.

**7. Key Takeaways**

* **Optimize pricing** across booking channels.
* **Refine upgrade policies** for last-minute bookings.
* **Plan better for long-stay customers, especially group bookings.**