Detailed Report on Hotel Booking Visualization Research

Overview:

The research primarily focuses on the analysis of hotel booking data to visualize patterns, trends, and key insights related to hotel reservations. Using various visualization techniques, the study aims to understand the factors affecting hotel bookings, including cancellation rates, pricing trends, and guest behavior. The data used for this research is sourced from a hotel booking dataset.

Research Objectives:

- **Cancellation Analysis:** To examine the percentage of reservations that were canceled versus those that were not.
- **Price Distribution:** To analyze the distribution of rates paid by guests, distinguishing between those who canceled their bookings and those who didn't.
- **Rate Trends Over Time:** To observe how the rates paid by guests have varied over the years and how these rates correlate with cancellations.
- **Special Requests:** To understand the frequency and distribution of special requests made by guests and how these vary by hotel type.
- **Booking Patterns:** To explore the relationship between booking patterns and the day of the month or year.

Methodology:

1. **Data Cleaning:**
   - The dataset was initially cleaned by addressing missing values, particularly in the 'children' and 'agent' columns. The mean was used to fill these missing values.
   - Irrelevant columns such as personal information and reservation status were dropped to focus on significant attributes.
2. **Visualization Techniques:**
   - **Pie Charts** were used to visualize the proportion of canceled versus non-canceled reservations.
   - **Distribution Plots** to show the spread of the rates paid by guests.
   - **Rug Plots** and **KDE (Kernel Density Estimation) Plots** for understanding the distribution of rates with respect to cancellation status.
   - **Violin Plots** and **Box Plots** were employed to show the distribution of rates over the years and to compare different types of room reservations.
   - **Heatmaps** and **Strip Plots** were used to observe how rates vary over different days of the month and across months.

Key Findings:

1. **Cancellation Rates:**
   - **Pie Chart Analysis** shows that approximately 37.04% of the reservations were canceled, while 62.96% were not canceled.
2. **Rates Paid by Guests:**

- **Distribution Plot** reveals a wide range of prices paid by guests, with most falling within a certain range. The analysis shows the rates paid by guests who did not cancel their bookings.
- **Rug Plot** further supports the distribution analysis, showing concentration at specific price points.
- **KDE Plot** comparing rates for canceled and non-canceled bookings indicates that canceled reservations often have different rate distributions compared to non-canceled ones.

3. **Rate Trends Over Time:**
   - **Violin Plot** illustrates that the average daily rate (ADR) has varied over the years, with noticeable differences between canceled and non-canceled bookings.
   - **Joint Plot** provides a visualization of rate changes over the weeks of the year and highlights seasonal pricing trends.

4. **Room Type and Rate:**
   - **Box Plot Analysis** shows the relationship between reserved room types and rates across different years. Certain room types consistently have higher rates, and this trend varies annually.
   - **Bar Plot** emphasizes how different hotel types (e.g., resort vs. city hotels) have distinct pricing trends over the years.

5. **Special Requests:**
   - **Count Plot** demonstrates the distribution of special requests by guests and how these requests vary by hotel type, with some hotels receiving more special requests than others.

6. **Rate Variations by Date:**
   - **Strip Plot and Heatmap** reveal that rates fluctuate significantly depending on the day of the month and the month itself, suggesting a seasonal or time-based pricing strategy by hotels.

Explanation of Images:

1. **Figure 1: Percent of Canceled vs. Non-Canceled Reservations (Pie Chart)**
   - This pie chart illustrates the distribution of reservations, showing that around 37.04% were canceled while 62.96% were not. The chart uses different colors and slight separation (explode) to emphasize the distinction between the two categories.

2. **Figure 2: Distribution of Rates Paid by Non-Canceled Guests (Distribution Plot)**
   - This plot shows the distribution of ADR (Average Daily Rate) for guests who did not cancel their reservations. It highlights the concentration of bookings around certain price points, with a blue color indicating a higher frequency of specific rates.

3. **Figure 3: Rug Plot of Rates for Non-Canceled Bookings**
   - The rug plot shows tiny vertical lines along the x-axis, each representing a booking. The concentration of lines at specific points indicates that many bookings occurred at those price levels.

4. **Figure 4: KDE Plot for Canceled vs. Non-Canceled Reservations**
   - The KDE plot overlays the distribution of rates for both canceled and non-canceled bookings, illustrating the differences in pricing patterns between

these two categories. The graph shows that canceled bookings have a distinct distribution compared to non-canceled ones.

5. **Figure 5: Violin Plot of Rate Changes Over the Years (Based on Cancellation Status)**
   - This violin plot combines aspects of a box plot and a KDE plot to show the distribution of rates over different years. The plot splits the data by cancellation status, showing how rates have varied for both canceled and non-canceled bookings.

6. **Figure 6: Joint Plot of Rate Changes Over the Weeks of the Year**
   - The joint plot shows the relationship between the week number of the year and the ADR. The plot indicates that certain weeks tend to have higher rates, reflecting possible seasonal pricing strategies.

7. **Figure 7: Box Plot of Reserved Room Types vs. Rate Across Years**
   - This box plot compares the rates of different room types across various years. The plot shows that some room types consistently have higher rates and that these rates can vary significantly depending on the year.

8. **Figure 8: Bar Plot of Rates Based on Hotel Type and Year**
   - The bar plot highlights how ADR differs between different hotel types (e.g., resort vs. city hotel) over the years, providing insights into how pricing strategies differ between these hotel categories.

9. **Figure 9: Count Plot of Special Requests by Hotel Type**
   - This count plot shows the number of special requests made by guests, broken down by hotel type. The plot reveals which hotels tend to receive more special requests, which may indicate higher guest expectations or different service offerings.

10. **Figure 10: Strip Plot of Day of the Month vs. Rate**
    - The strip plot displays individual bookings as points on a graph, showing how ADR varies depending on the day of the month. The plot reveals a pattern of pricing fluctuations, possibly indicating strategic pricing on certain days.

11. **Figure 11: Heatmap of Rates by Day and Month**
    - The heatmap shows the variation in ADR across different days of the month and months of the year. The color gradient indicates higher or lower rates, providing a visual representation of pricing trends over time.

Conclusion:

This research provides valuable insights into hotel booking behaviors, cancellation rates, and pricing trends. The visualizations help in understanding how different factors influence hotel bookings and can be instrumental in informing pricing strategies, customer relationship management, and operational planning for hotels.