**Exploratory Data Analysis (EDA):**

Dataset Overview: The dataset consisted of hotel booking information, including features like room type, meal plan, booking status, and more.

Data Cleaning and Preprocessing: Included handling missing values, removing duplicates, and correcting anomalies.

Visualization and Insights: Histograms, box plots, and scatter plots provided insights into distributions and relationships. Key findings were the skewness in certain variables and the lack of strong linear relationships.

**Feature Engineering:**

New Features: Created features such as total stay duration and total guests for potential predictive power.

Feature Selection and Scaling:

Dimensionality Reduction: Applied techniques like correlation analysis and feature importance using a Random Forest model, identifying key features like lead time, average room price, and number of special requests.

Feature Scaling: Standardized the selected features to have a mean of 0 and a standard deviation of 1, making the data suitable for algorithms sensitive to feature scale.

**Model Selection and Training:**

Baseline Models: Started with simple models like Decision Trees, achieving an accuracy of approximately 84.65%.

Additional Model: Trained a KNN model, which showed a slightly lower accuracy of around 84.34%.

**Evaluating and Fine-Tuning Models:**

Hyperparameter Tuning: Utilized Grid Search CV to optimize the Decision Tree and KNN models.

Performance Assessment: Post-tuning, the models were evaluated based on accuracy, with a slight improvement observed.

**Handling Class Imbalance with SMOTE:**

Class Imbalance Analysis: Identified a significant imbalance in the target variable (booking status).

Application of SMOTE: Balanced the dataset using SMOTE, creating synthetic samples for the minority class.

Model Retraining and Comparison: Retrained models on the balanced dataset showed a trade-off between improved class balance and a slight decrease in overall accuracy.

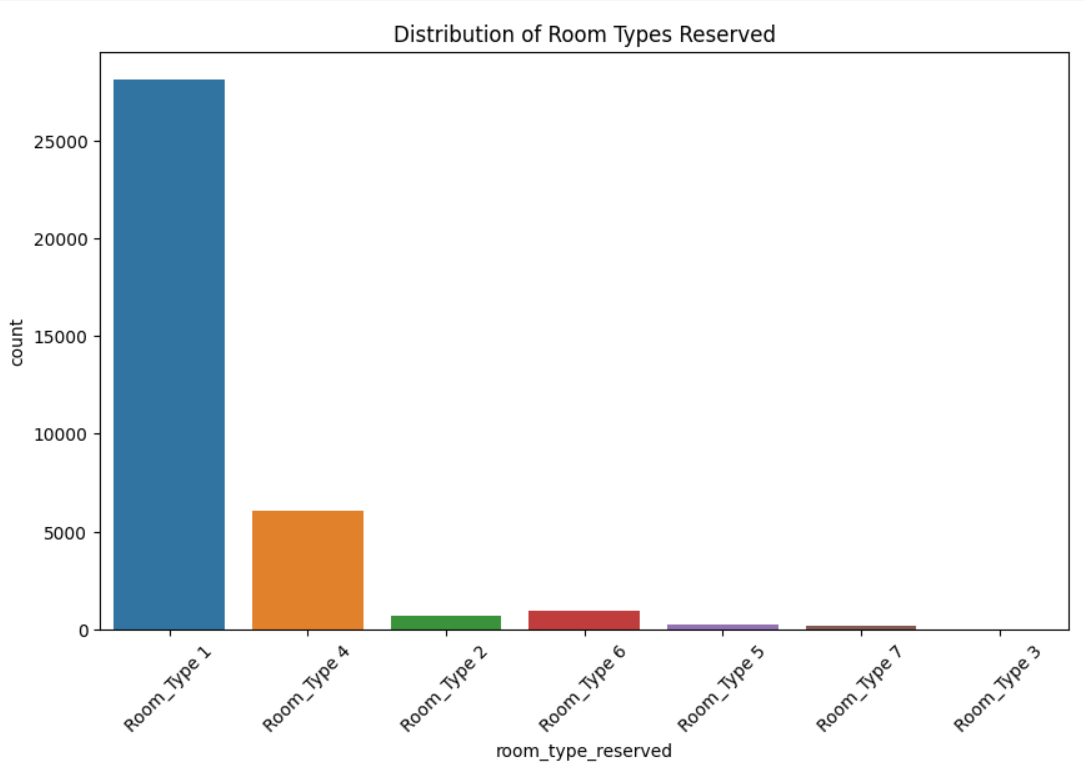
**Deliverables:**

**Comprehensive Report**: This summary encapsulates the key aspects of the analysis, including EDA, feature engineering, model training, evaluation, and insights.

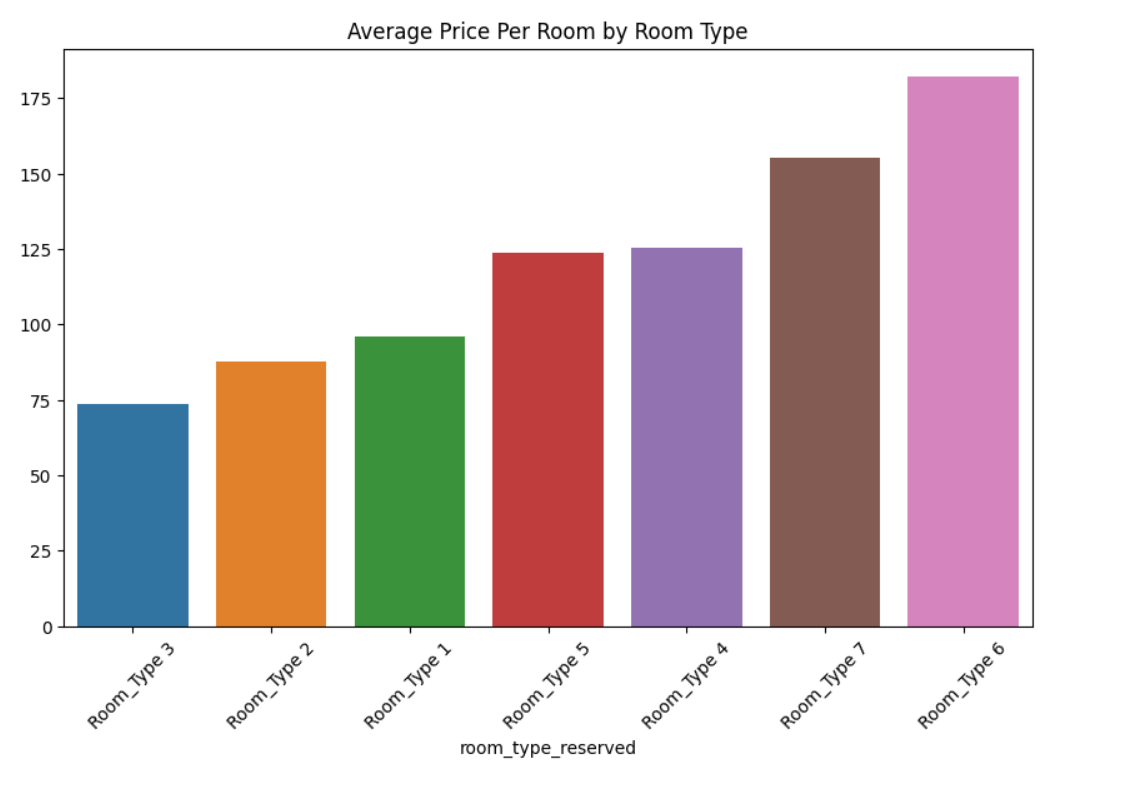
**Insights and Recommendations:** The analysis highlighted the importance of addressing class imbalance and the impact of feature selection on model performance. Future work could explore additional models and further fine-tuning to enhance model accuracy, especially in predicting the minority class.

This report provides a holistic view of the data analysis process, from initial exploration to model optimization, offering valuable insights into the hotel booking dataset.

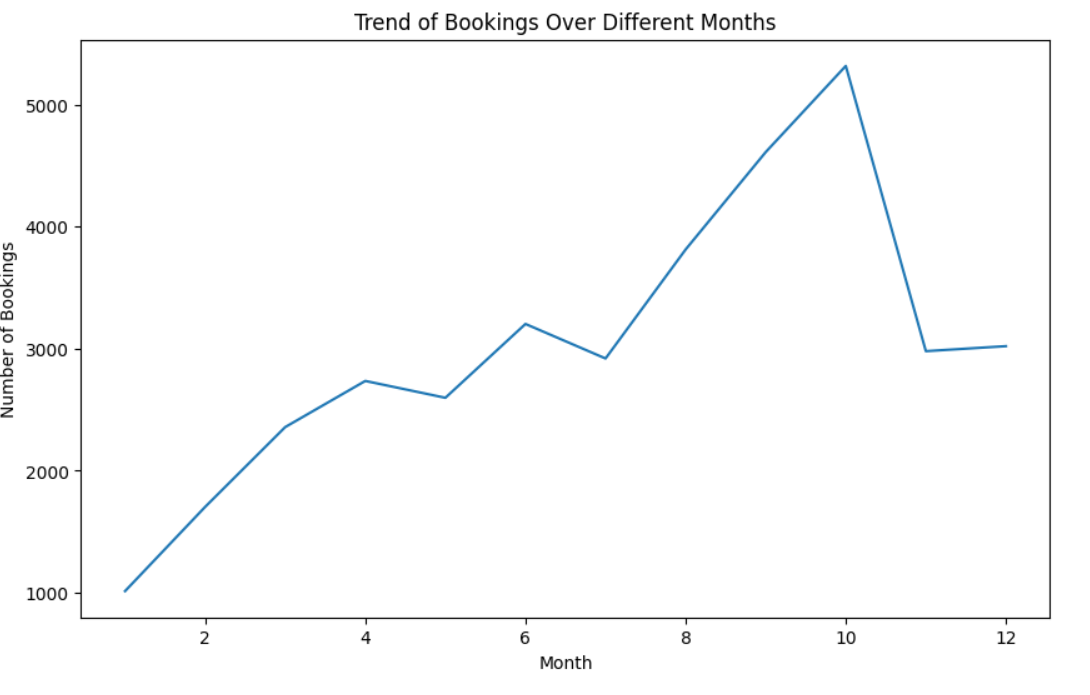
**Question 1: What is the distribution of room types reserved?**



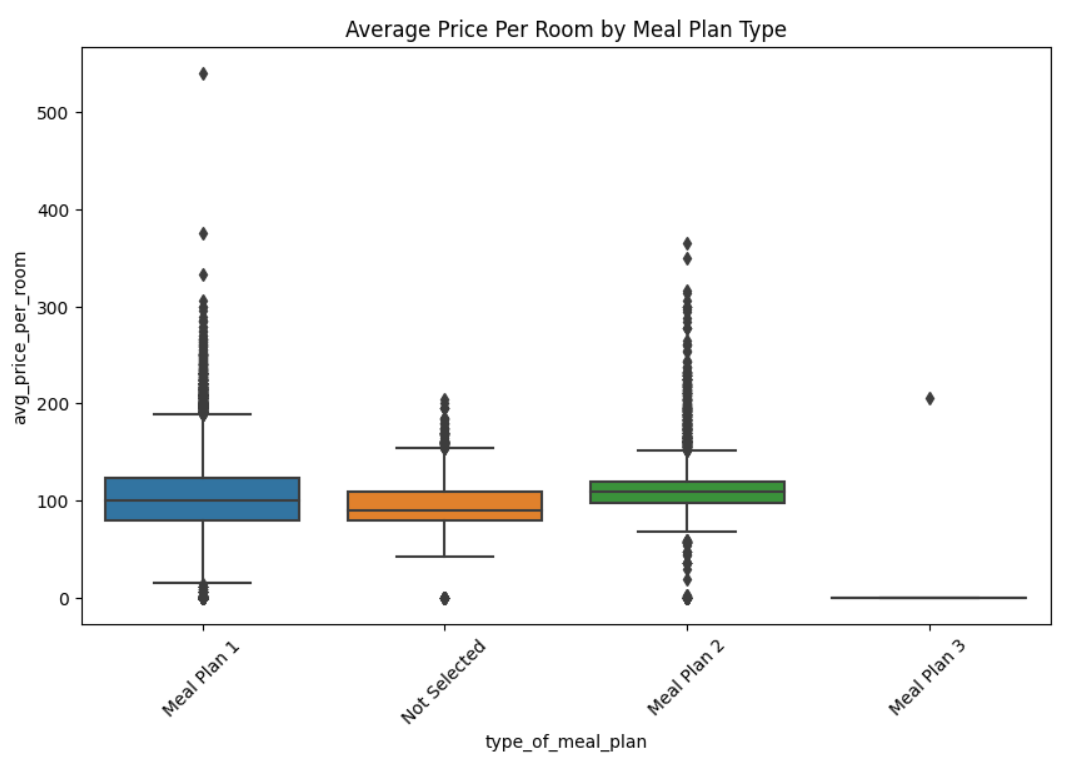
**Question 2: How does the average price per room vary by room type?**



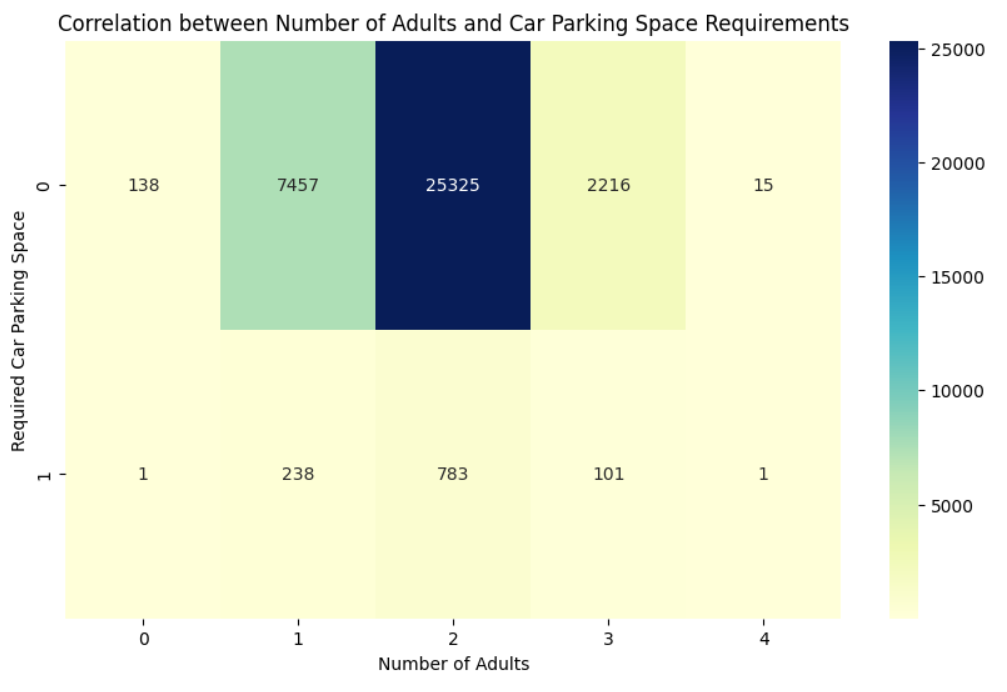
**Question 3: What is the trend of bookings over different months of the year?**



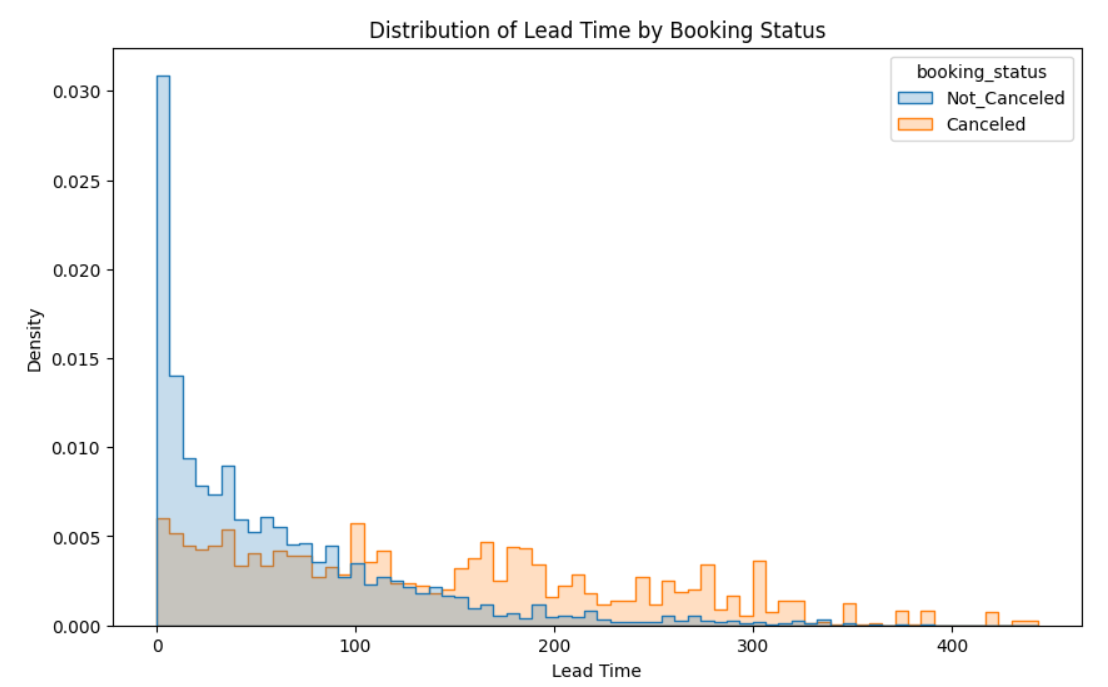
**Question 4: What is the relationship between the type of meal plan and the average price per room?**



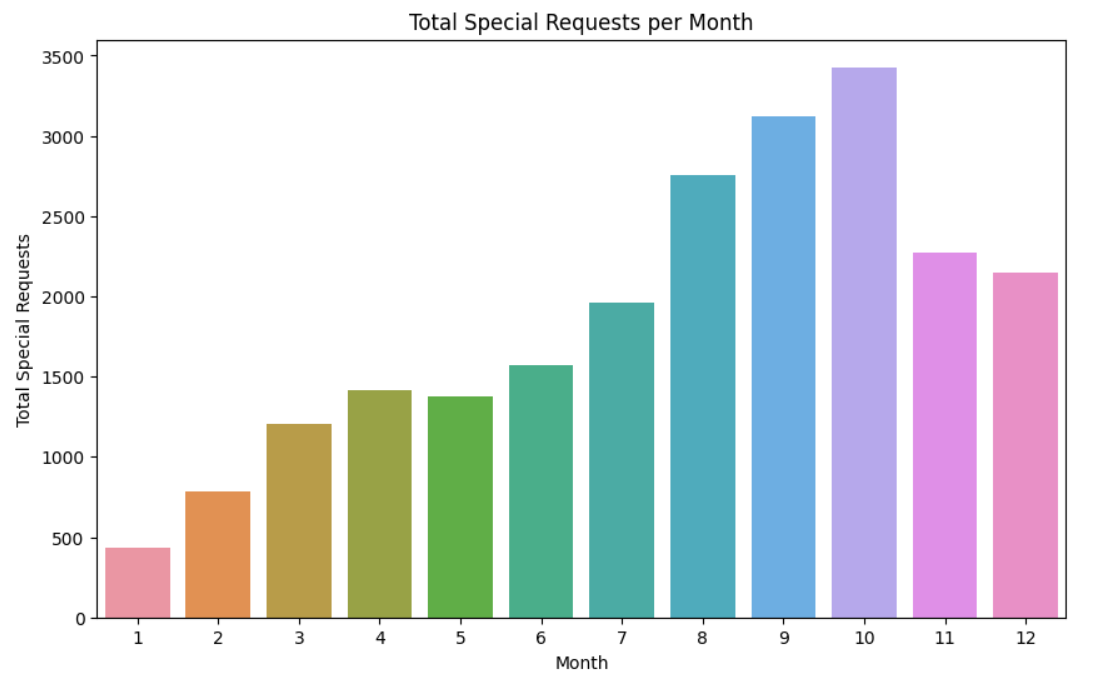
**Question 5: How does the requirement for car parking space relate to the number of adults in a booking?**



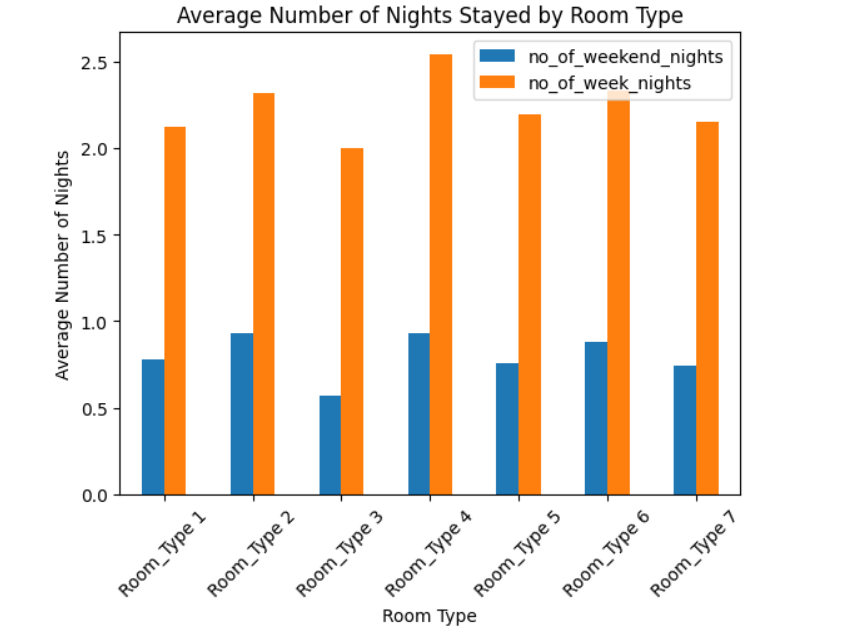
**Question6: What is the distribution of lead time for bookings, and how does it vary by booking status (canceled vs. not canceled)?**

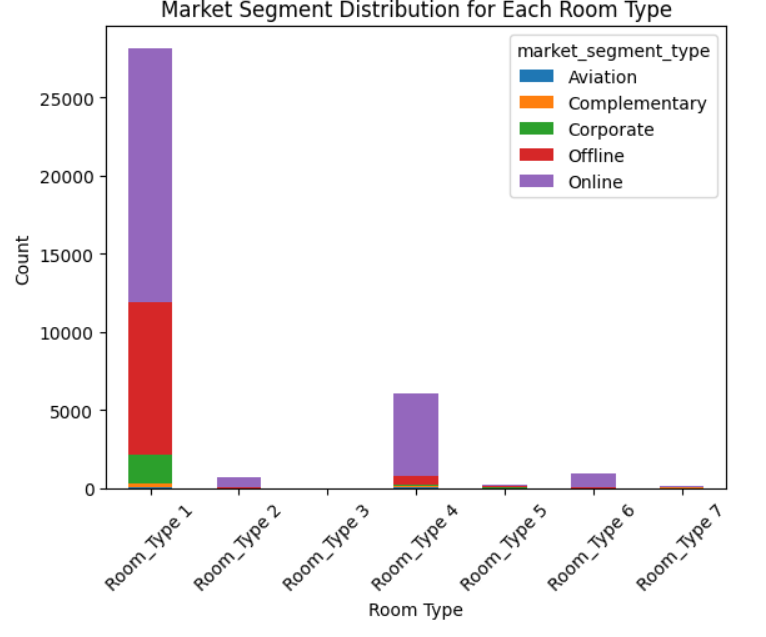


**Question7: Do certain months of the year have a higher number of special requests?**

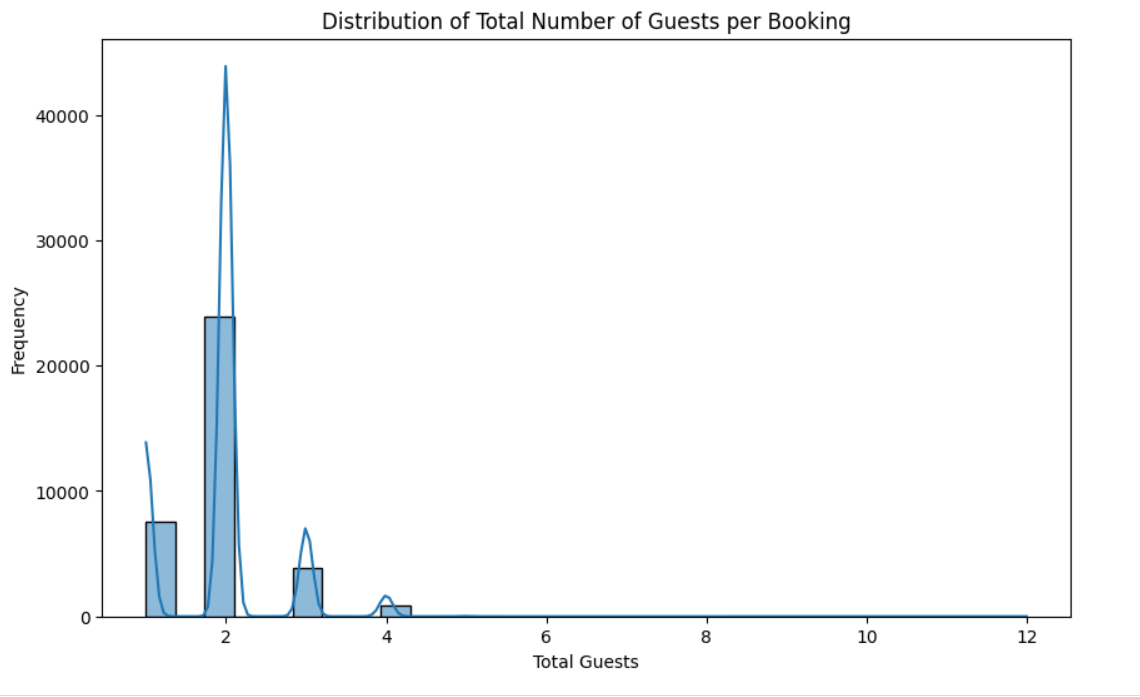


**Question8: How does the average number of nights stayed (weekdays vs. weekends) vary across different room types?**

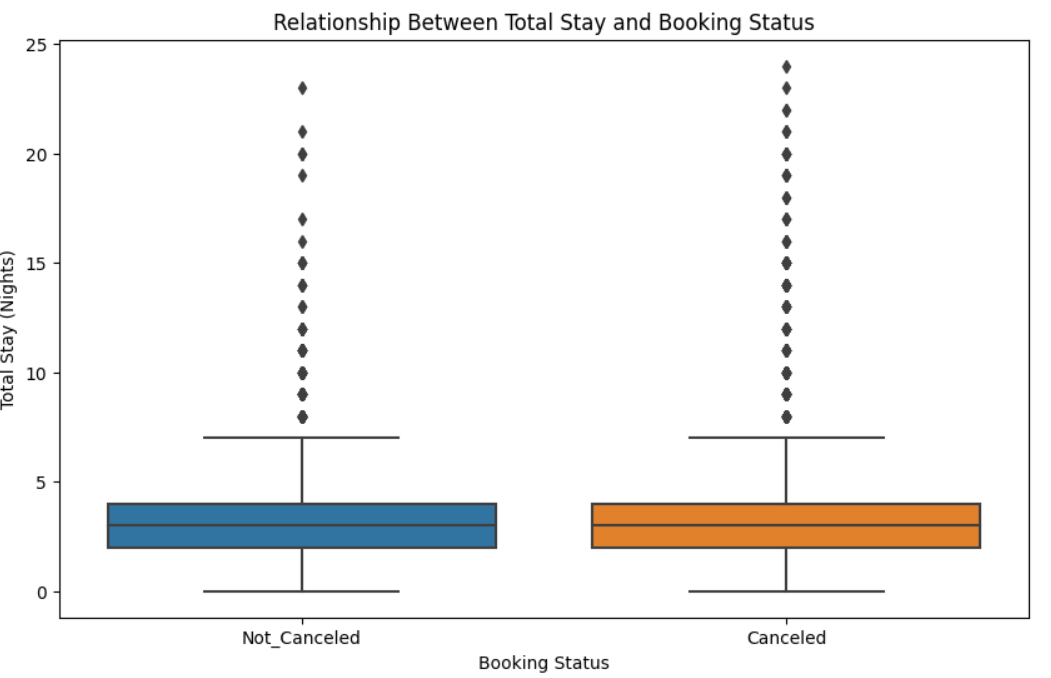


**Question9: Is there a predominant market segment type for different room types?**

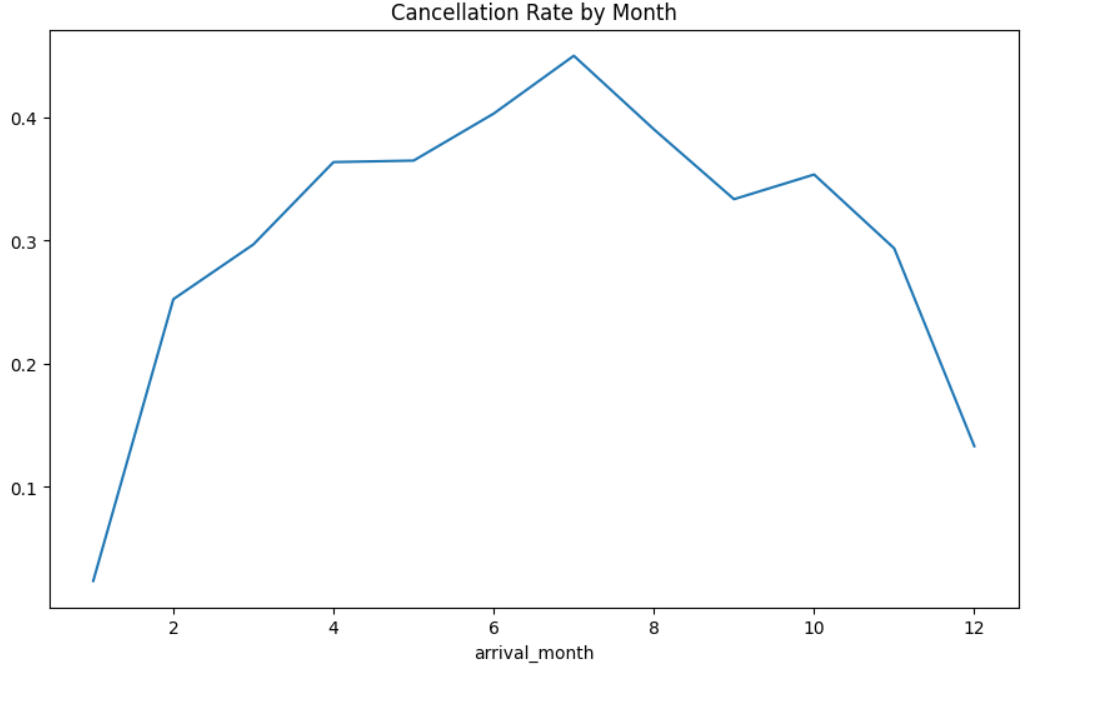
**Question 10: What is the distribution of the total number of guests (adults + children) in bookings?**



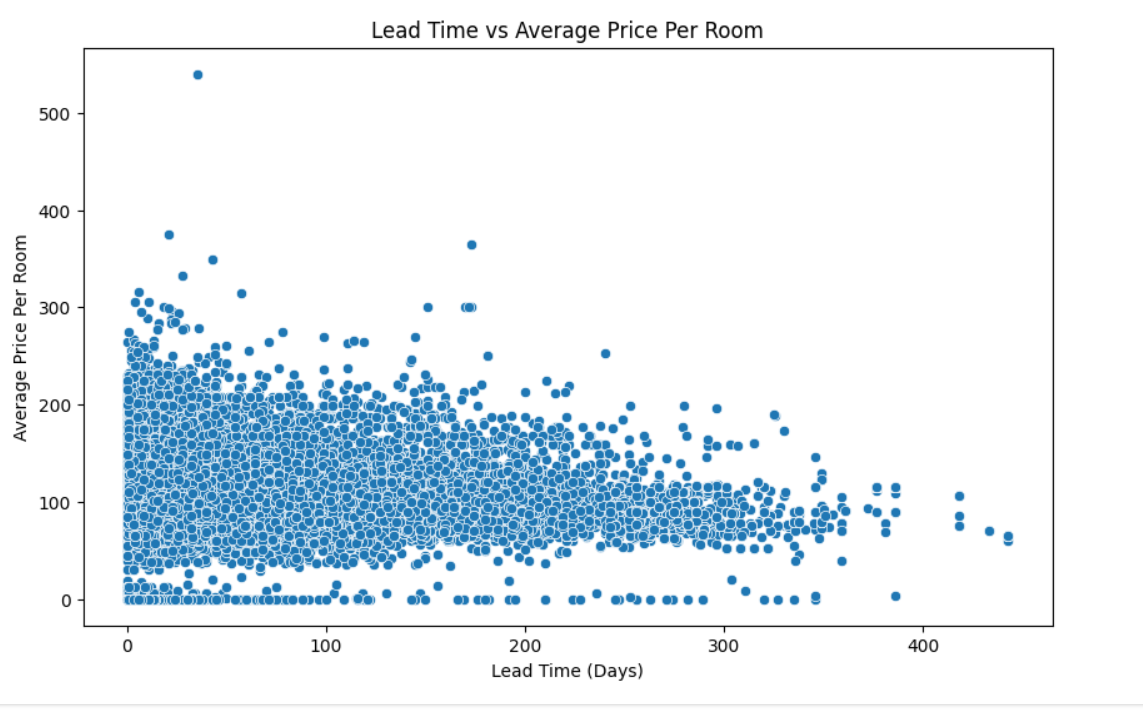
**Question 11: What is the relationship between the length of stay and booking cancellations?**



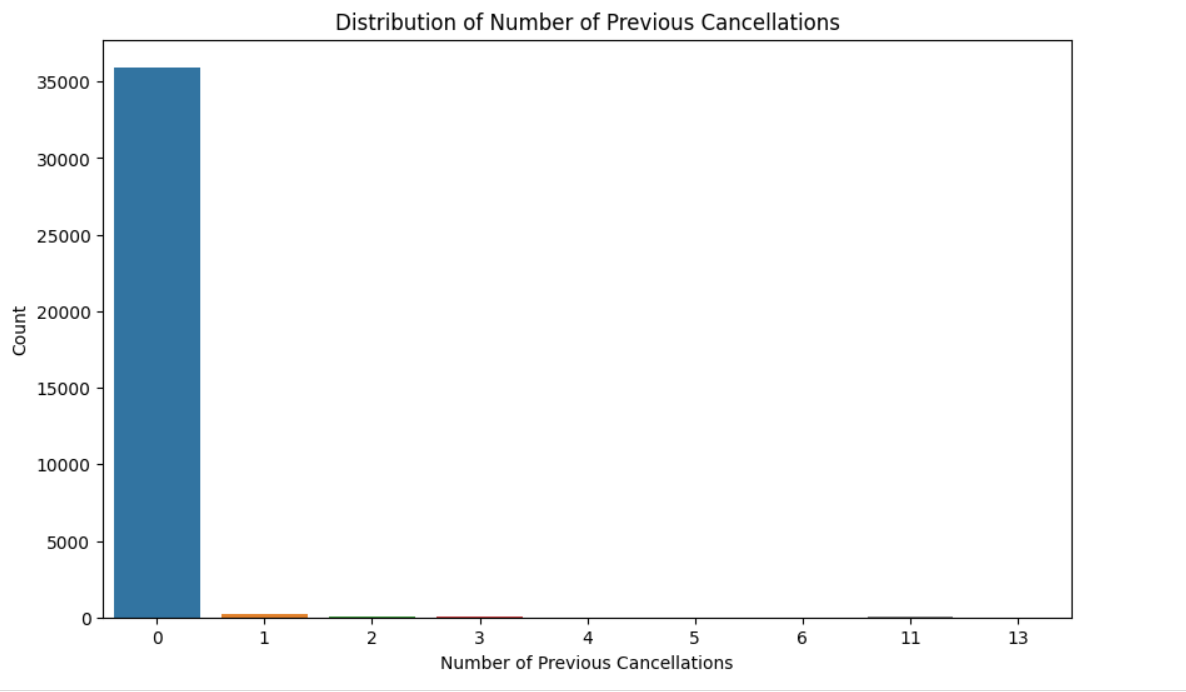
**Question 12: Are there specific months with higher cancellation rates?**



**Question 13: How does the lead time affect the average price per room?**



**Question 14: What is the distribution of the number of previous cancellations among all bookings?**



**Question 15: Does the average price per room vary across different market segments?**

