Hotel Reservation Analysis

Presented By

Shih-Chieh Ku

Shireesha Thyaranahalli Narayana

Introduction

Hotel booking has changed how people reserve rooms and act when making reservations. Many hotel bookings get canceled or guests don’t show up.

Common reasons for this are changes in travel plans or scheduling conflicts. This often happens because guests can cancel their reservations for free or at a low cost. While this is good for guests, it can hurt the hotel's revenue.

The main objective of this project is to predict whether a customer will honor the reservation or cancel it. This prediction can be valuable for hotels to manage their bookings more efficiently and reduce revenue loss due to cancellations.

Dataset Description

Data Description

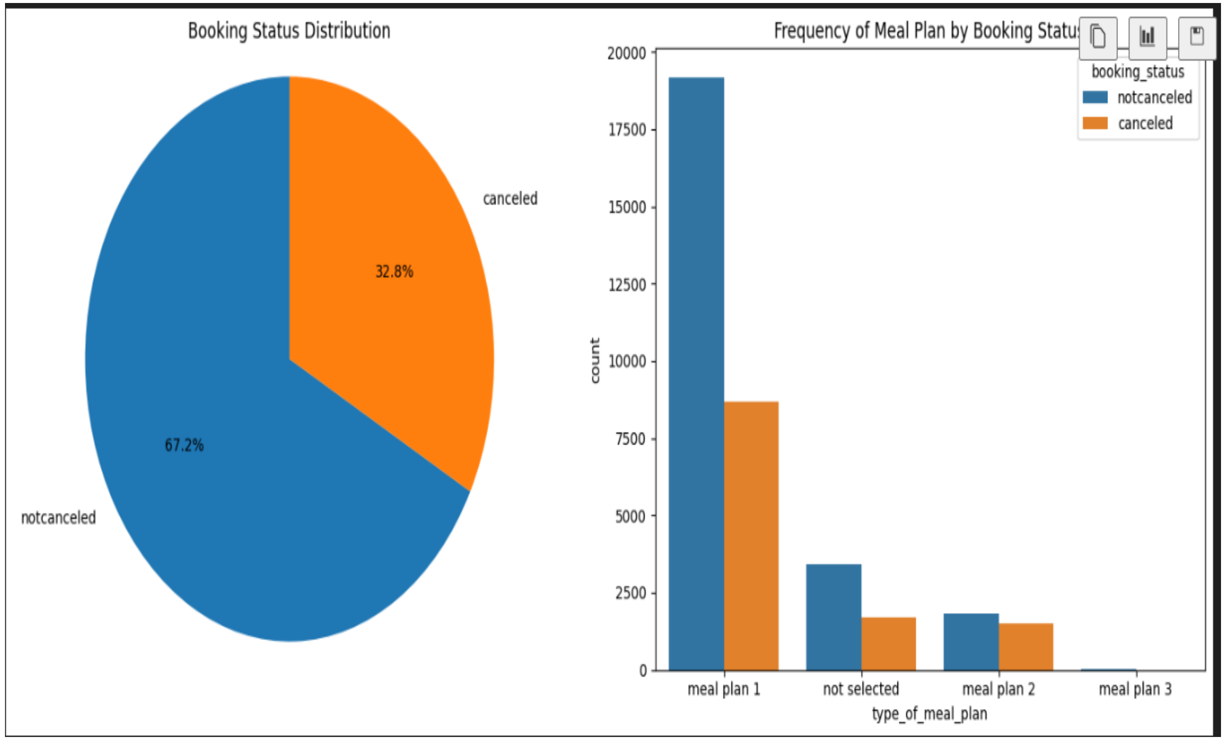
- **no_of_adults**: The number of adults in the reservation (categorical: 5 values)
- **no_of_children**: The number of children in the reservation (categorical: 6 values)
- **no_of_weekend_nights**: The number of weekend nights included in the reservation (categorical: 8 values)
- **no_of_week_nights**: The number of weeknights included in the reservation (categorical: 18 values)
- **type_of_meal_plan**: The type of meal plan chosen (categorical: 4 values)
- **required_car_parking_space**: Whether a car parking space is required (binary: 2 values)
- **room_type_reserved**: The type of room reserved (categorical: 7 values)
- **lead_time**: The number of days between booking and arrival (numerical: range up to 352 days)
- **arrival_year**: The year of arrival (categorical: 2 values)
- **arrival_month**: The month of arrival (categorical: 12 values)
- **arrival_date**: The day of arrival (categorical: 31 values)
- **market_segment_type**: The type of market segment (categorical: 5 values)
- **repeated_guest**: Whether the guest is a repeated customer (binary: 2 values)
- **no_of_previous_cancellations**: The number of previous cancellations by the guest (categorical: 9 values)
- **no_of_previous_bookings_not_canceled**: The number of previous bookings not canceled by the guest (numerical: range up to 59)
- **avg_price_per_room**: The average price per room (numerical: range up to 3930)
- **no_of_special_requests**: The number of special requests made by the guest (categorical: 6 values)
- **booking_status**: The target variable, indicating whether the reservation was canceled or not (binary: 2 values)

- Dataset is collected by Kaggle.
- The dataset contains valuable information about hotel bookings, including various features such as customer details, booking history, and reservation status.
- The dataset's target variable is booking_status, which has two categories: "Not Canceled" and "Canceled."
- The dataset includes various features such as the number of adults and children, lead time, room type, and more. Initial exploration shows that "Canceled" reservations account for a significant portion of the dataset.

Data Exploration

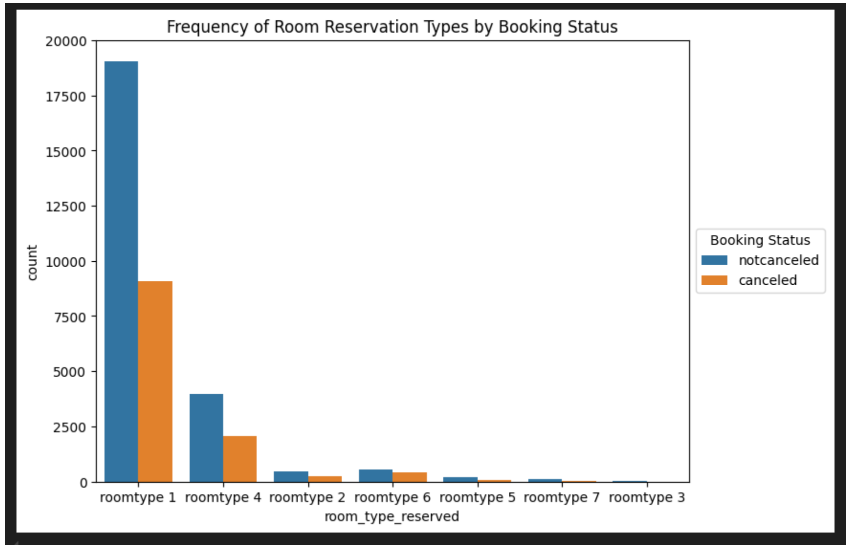
We are focus on understanding the relationships between key factors and booking outcomes. Some of the aspects central to our analysis

- **Type of Meal Plan:** We will investigate how the choice of meal plan, selected by guests, is associated with booking status (whether a reservation was canceled or not).
- By the analysis, we can tell 67.2% has not canceled the stay and 32.8% has canceled the stay.



Data Exploration

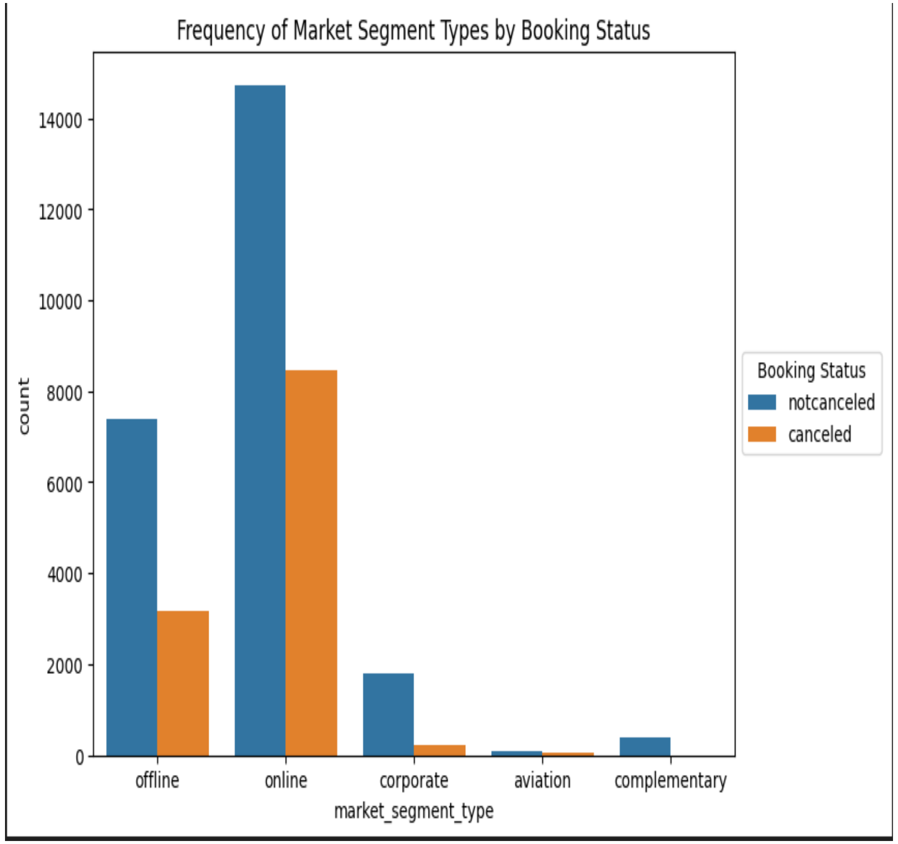
- **Room Type Reserved:** Selection of room types, such as "Single," "Double," or "Suite," influences booking outcomes.
- Room type 1 has the highest percentage in both not canceling and canceled the booking.
- Room Type 1 is very popular. On average, people book it 89.6 days in advance, showing they prefer to book early. The middle booking time is 59 days. However, the wide range in booking times (with a standard deviation of 90.8 days) shows that guests have different planning habits.



	Room Type Reserved	Average Lead Time	Median Lead Time	Lead Time Std Dev
0	roomtype 1	89.642943	59.000000	90.860265
1	roomtype 2	101.021676	75.000000	85.377079
2	roomtype 3	69.571429	66.000000	70.056780
3	roomtype 4	69.205712	55.000000	59.718577
4	roomtype 5	59.547170	32.000000	70.581182
5	roomtype 6	61.015528	38.000000	64.116238
6	roomtype 7	37.094937	20.500000	43.630292

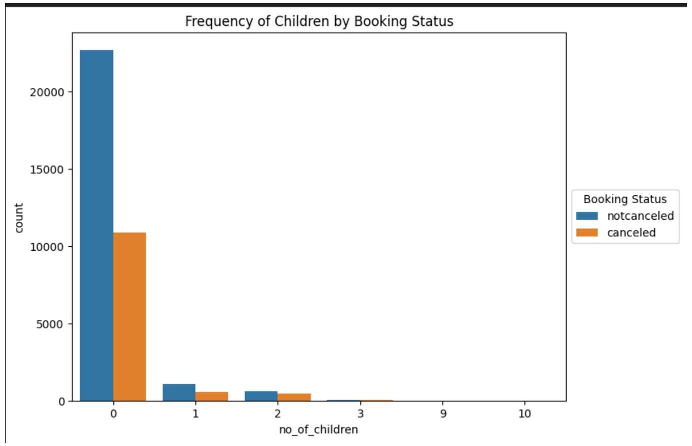
Data Exploration

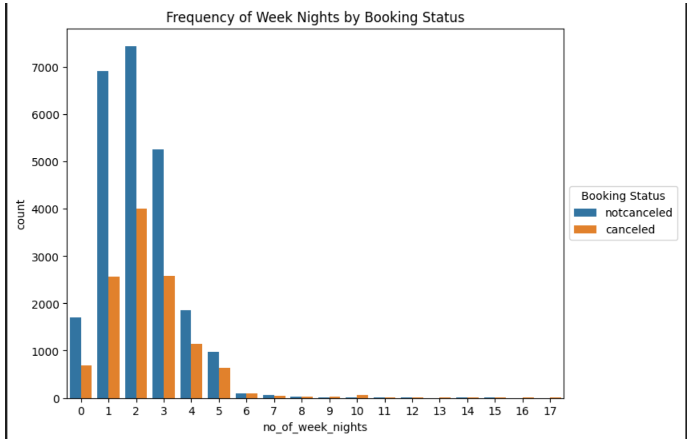
- **Market Segment Type:** Explore how different market segments, including Online Travel Agents, Corporate, Direct, and others, impact booking status.
- Online booking Platform has the highest number of not canceled.
- Complementary has no canceled of booking.

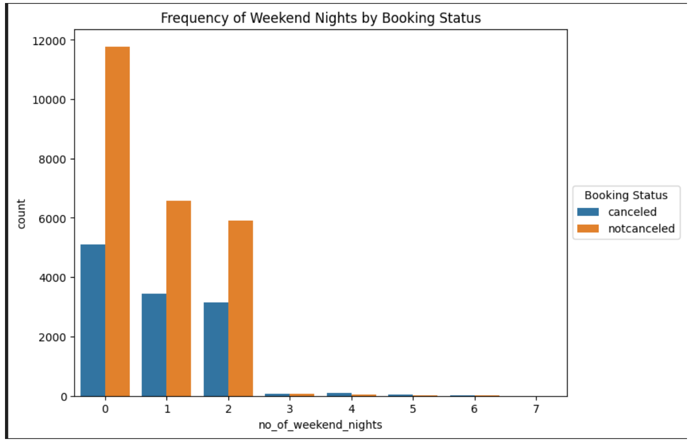


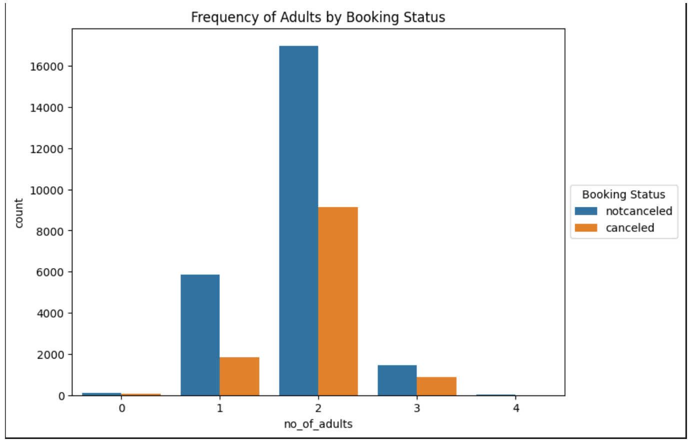
Data Exploration

- **Other Key Factors to consider**
- Number of Adults
- Repeated Guest
- Number of Children
- Required Car Parking Space
- Weekend Night bookings









Modeling

- Logistic Regression
- Random Forest Classification
- Decision Tree Classification:
- K-Nearest Neighbors (KNN)

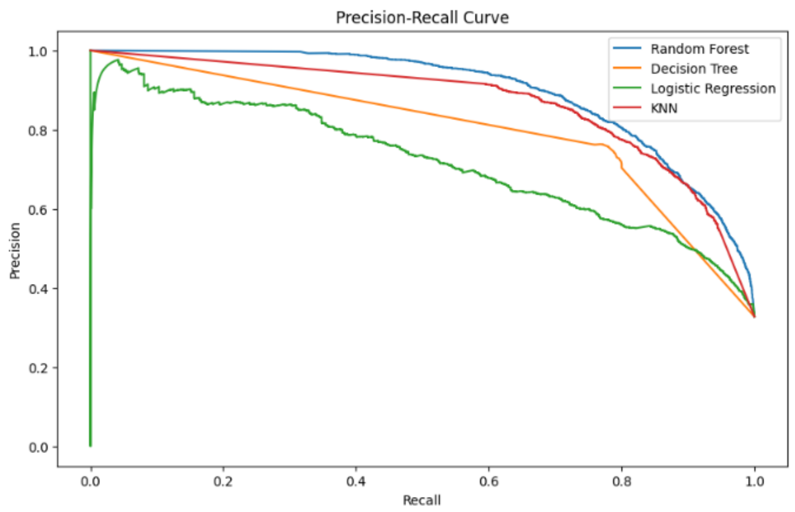
```
# Define models with best parameters
rf_model = RandomForestClassifier(bootstrap=False, max_depth=None, min_sam
dt_model = DecisionTreeClassifier(criterion='gini', max_depth=None, max_fe
logistic_model = LogisticRegression(C=0.01, penalty='l2')
knn_model = KNeighborsClassifier(algorithm='brute', n_neighbors=9, p=1, we

# Train models
rf_model.fit(X_resampled, y_resampled)
dt_model.fit(X_resampled, y_resampled)
logistic_model.fit(X_resampled, y_resampled)
knn_model.fit(X_resampled, y_resampled)
```

Model Evaluation

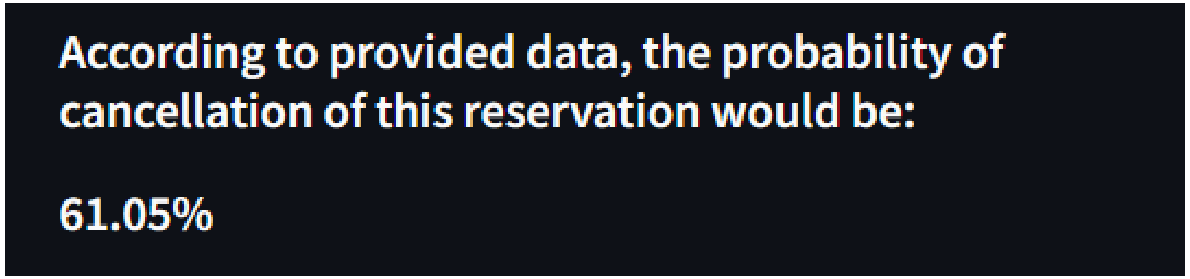
- The Random Forest model exhibits the highest performance across all metrics. The Random Forest model stood out with an F1 Score of 0.868632, reflecting a strong balance between precision and recall.

- Logistic Regression has the lowest performance metrics among the models listed. Its F1 score and recall are significantly lower.



	Model	F1 Score	Recall	Accuracy	Precision
0	Random Forest	0.867933	0.867571	0.867571	0.868390
1	Decision Tree	0.846682	0.846069	0.846069	0.847497
2	Logistic Regression	0.758917	0.753556	0.753556	0.772623
3	KNN	0.858460	0.857426	0.857426	0.860218

User Interface



Conclusion

- Increasing prices are associated with a higher rate of cancellations. To mitigate reservation cancellations, hotels could refine their pricing strategies by offering reduced rates for specific locations and providing discounts to customers.
- During the month of January, hotels can launch marketing campaigns with attractive offers to boost their revenue, especially since cancellations tend to peak during this period.

Business Decision and Marketing Strategies

Overbooking: Slightly overbook based on predicted cancellations to ensure maximum occupancy.

Dynamic Pricing: Adjust room rates based on predicted demand and cancellation rates.

Targeted Offers: Send special offers to customers likely to cancel, encouraging them to keep their bookings.

Segment Targeting: Focus on market segments with lower cancellation rates.

Advance Booking Incentives: Offer discounts for early bookings to secure reservations.

Pre-arrival Engagement: Send reminders and personalized messages to keep guests committed to their bookings.

Real-time Monitoring: Integrate the predictive model with the booking system for real-time insights.

Thank You