Grace Samuel

DSE6211

Analytic Plan – ABC Hotels

Business Need:

- o Identify bookings with a high risk of cancellation.
- o Determine which features of a customer are most predictive of cancellation.

Target Variable:

- o booking_status (0 = Not Canceled, 1 = Canceled)

Data Processing:

- o **Data Cleaning:**
  - Create an initial plan for variables; exclude, alter, leave as is (pg.2).
  - Handle unknown and missing values.
- o **Data Transformation (pg. 2):**
  - One-hot encode categorical variables and delete original data after transformation.
  - Standardize numerical variables for scaling.
    - $Xscaled = \frac{X - mean(X)}{std(X)}$
  - Extract information from date data and create new variables.

Initial Feature Inclusion:

- o At this phase, all features will be included except Booking_ID which is a unique identifier that can be deleted.
- o Demographic variables, booking details, preferences, and other predictors are included.

Expected Analytic and Informational Outcomes:

- o **Machine Learning Model:**
  - A model that outputs probability of cancellation.

- o **Feature Importance:**
  - Identify the most influential features contributing to cancellations.

- o **Risk Scores:**
  - Assign a cancellation probability to each booking.

Model Used in Practice:

- o **Integration:**
  - Integrate the model into ABC Hotel's booking system to calculate cancellation probabilities with real-time data.

- o **Targeted Interventions:**
  - Use probabilities to identify high-risk bookings and target them with personalized offers or incentives to reduce risk of cancellation.
- o **Monitoring:**
  - Continuously monitor model performance and adjust as necessary.

Next Steps:

- o Exploratory Data Analysis
- o Model Building
- o Implementation

Initial Plan for Variables:

Variable	Plan	Notes
Booking_ID	Exclude	This is a unique identifier. There are no duplicate identifiers, so this variable can be excluded from analysis.
no_of_adults	Standardize	Numeric variable representing the number of adults in the booking. Standardize by scaling to have a mean of 0 and a standard deviation of 1. This ensures consistent contribution across features and compatibility with models sensitive to scale (e.g., gradient descent-based or distance-based models)
no_of_children	Standardize	Numeric variable representing the number of children in the booking. Standardize by scaling to have a mean of 0 and a standard deviation of 1.
no_of_weekend_nights	Standardize	Numeric variable representing the number of weekend nights booked in the booking. Standardize by scaling to have a mean of 0 and a standard deviation of 1.
no_of_week_nights	Standardize	Numeric variable representing the number of week nights booked in the booking. Standardize by scaling to have a mean of 0 and a standard deviation of 1.
type_of_meal_plan	Alter	Categorical variable representing type of meal plan. One-hot encode to represent each meal plan as a binary variable. Drop original type_of_meal_plan column after transformation.
required_car_parking_space	Leave as is	Binary variable (0 or 1) indicating whether a parking space is required. No change needed.
room_type_reserved	Alter	Categorical variable representing the type of room reserved. One-hot encode to represent each type of room reserved as a binary variable. Drop original room_type_reserved column after transformation.
lead_time	Standardize	Numeric variable indicating the number of days between booking and arrival. Standardize by scaling to have a mean of 0 and a standard deviation of 1.

		This ensures consistent contribution across features and compatibility with models sensitive to scale (e.g., gradient descent-based or distance-based models).
arrival_date	Alter	Date variable representing arrival date. Create three new columns using arrival_date: • **Day of the week (0=Monday, 6=Sunday):** Captures weekly patterns. • **Month (1 = January, 12 = December):** Captures seasonal trends. • **Year :** Captures changes over multiple years. Next, one-hot encode for days of the week and month. Year can remain numeric but should be standardized to make it compatible with the model. Delete original arrival_date column.
market_segment_type	Alter	Categorical variable representing market segment types. One-hot encode to represent market segment as a binary variable. Drop original market_segment_type column after transformation.
repeated_guest	Leave as is	Binary variable (0 or 1) indicating whether a customer is a repeat guest. No change needed.
no_of_previous_cancellations	Standardize	Numeric variable representing the number of past cancellations. Standardize by scaling to have a mean of 0 and a standard deviation of 1.
no_of_previous_bookings _not_cancelled	Standardize	Numeric variable indicating the number of previous bookings that were not cancelled. Standardize by scaling to have a mean of 0 and a standard deviation of 1.
avg_price_per_room	Standardize	Numeric variable representing the average price per room. Standardize by scaling to have a mean of 0 and a standard deviation of 1.
no_of_special_requests	Standardize	Numeric variable indicating the number of special requests made by the customer. Standardize by scaling to have a mean of 0 and a standard deviation of 1.
booking_status	Alter	This is the **target variable**. Transform into a binary variable (0 = Not Canceled, 1 = Canceled).