

With the given columns in the dataset, perform the Exploratory Data Analysis (EDA) using appropriate figures and techniques. Use the necessary libraries to read, manipulate and visualize the data.

1. Data Inspection and Cleaning:

- Check for missing values and handle them appropriately (e.g., imputation, dropping rows/columns).
- Check for duplicates and remove them if necessary.
- Identify and handle outliers, if any.
- Check for inconsistencies in the data (e.g., invalid values, typos).

2. Descriptive Statistics:

- Calculate summary statistics (mean, median, mode, standard deviation, etc.) for numerical columns like `lead_time`, `stays_in_weekend_nights`, `stays_in_week_nights`, `adults`, `children`, `babies`, `previous_cancellations`, `previous_bookings_not_canceled`, `days_in_waiting_list`, `adr`, `required_car_parking_spaces`, and `total_of_special_requests`.
- Display value counts and frequencies for categorical columns like `hotel`, `country`, `market_segment`, `distribution_channel`, `is_repeated_guest`, `reserved_room_type`, `assigned_room_type`, `deposit_type`, `agent`, `company`, `customer_type`, and `reservation_status`.

3. Data Visualization:

- Create histograms or box plots for numerical columns to visualize the distribution and identify potential outliers.
- Use bar plots or pie charts to visualize the distribution of categorical columns.
- Create scatter plots or heatmaps to explore relationships between numerical columns.
- Use line plots to visualize trends over time for columns like `arrival_date_year`, `arrival_date_month`, `arrival_date_week_number`, and `arrival_date_day_of_month`.

4. Correlation Analysis:

- Calculate the correlation matrix to identify potential relationships between numerical columns. (use `pandas df.corr()`)
- Visualize the correlation matrix using a heatmap.

5. Categorical Data Analysis:

- Perform one-hot encoding or label encoding for categorical columns if required for further analysis.
- Analyze categorical columns like `hotel`, `country`, `market_segment`, `distribution_channel`, `reserved_room_type`, `assigned_room_type`, `deposit_type`, `agent`, `company`, `customer_type`, and `reservation_status` to identify patterns or trends.
- Create contingency tables or mosaic plots to analyze the relationship between categorical columns.

6. Time Series Analysis:

- Analyze the `arrival_date_year`, `arrival_date_month`, `arrival_date_week_number`, and `arrival_date_day_of_month` columns to identify seasonality or trends in bookings over time.
- Create time series plots or decompose the time series to understand the trend, seasonality, and residuals.

7. Feature Engineering:

- Create new features based on existing columns, such as calculating the duration of stay from stays_in_weekend_nights and stays_in_week_nights.
- Derive new features from date columns like arrival_date_year, arrival_date_month, arrival_date_week_number, and arrival_date_day_of_month (e.g., season, month name, day of the week).

8. Handling Datetime Columns:

- Convert the reservation_status_date column to datetime format if necessary.
- Extract additional features from the datetime column, such as day of the week, month, or hour.