**Hotel Booking Analysis Report**

**Table of Contents**

1. [Introduction](#introduction)

2. [Methodology](#methodology)

3. [Data Preprocessing & Cleaning](#data-preprocessing--cleaning)

    - [Duplicates and Errors](#duplicates-and-errors)

    - [Missing Values](#missing-values)

    - [Standardization](#standardization)

    - [Data Validation](#data-validation)

4. [Exploratory Data Analysis (EDA)](#exploratory-data-analysis-eda)

    - [Summary Statistics](#summary-statistics)

    - [Visualizations](#visualizations)

    - [Correlations](#correlations)

    - [Outliers](#outliers)

    - [Patterns and Trends](#patterns-and-trends)

5. [Statistical Analysis](#statistical-analysis)

    - [Hypothesis Testing](#hypothesis-testing)

    - [Regression Analysis](#regression-analysis)

6. [Key Insights & Recommendations](#key-insights--recommendations)

    - [Actionable Insights](#actionable-insights)

    - [Measurable Results](#measurable-results)

    - [Alignment with Business Goals](#alignment-with-business-goals)

    - [Cost-Effectiveness](#cost-effectiveness)

7. [Conclusion](#conclusion)

8. [Appendix](#appendix)

**Introduction**

This report presents an analysis of hotel booking data to provide actionable insights for improving hotel operations and marketing strategies. The dataset includes various attributes related to hotel bookings, such as booking dates, customer demographics, and booking channels.

**Methodology**

The analysis involved several steps:

1. Data preprocessing and cleaning to ensure data quality.

2. Exploratory Data Analysis (EDA) to uncover patterns and trends.

3. Statistical analysis to validate findings and derive insights.

4. Formulating recommendations based on the analysis.

**Data Preprocessing & Cleaning**

**Duplicates and Errors**

- **Duplicates**: Identified and removed 0 duplicate records.

- **Errors**: No significant errors were found in the dataset.

**Missing Values**

- **Numerical Data**: Imputed missing values using mean for `children` and median for `agent` and `company`.

- **Categorical Data**: Imputed missing values using mode for `country` and filled missing values in `meal` with 'Missing'.

**Standardization**

- **Date Formats**: Standardized `reservation\_status\_date` to datetime format.

- **Text Case Consistency**: Standardized text case for columns like `hotel`, `meal`, `country`, etc.

- **Column Naming**: Converted all column names to lowercase.

**Data Validation**

- **Data Type Mismatches**: Ensured all columns have appropriate data types.

- **Unexpected Categories**: Checked for unexpected values in categorical columns.

- **Consistency Checks**: Verified consistency between related columns, e.g., `reserved\_room\_type` and `assigned\_room\_type`.

**Exploratory Data Analysis (EDA)**

**Summary Statistics**

- **Numerical Summary**:

    - Mean, median, standard deviation, and percentiles for key numerical columns.

    - Example: `lead\_time` mean = 104.01, median = 69, std = 106.86.

**Visualizations**

- **Histograms**: Distribution of `lead\_time`, `adr`.

- **Box Plots**: Outliers in `adr`, `lead\_time`.

- **Scatter Plots**: Relationship between `lead\_time` and `adr`.

- **Bar Charts**: Number of customers per year, month, and day of the week.

**Correlations**

- **Correlation Matrix**: Highlighted significant correlations, e.g., `adr` positively correlated with `adults` and `children`.

**Outliers**

- **Detection**: Identified outliers using Z-score and IQR methods.

- **Handling**: Removed outliers to improve model accuracy.

**Patterns and Trends**

- **Seasonal Trends**: Higher customer visits during summer months and weekends.

- **Booking Channels**: Online travel agents are the most popular booking channel.

**Statistical Analysis**

**Hypothesis Testing**

- **Tests**: Conducted t-tests to compare means between groups.

- **Results**: Significant differences found in `adr` between different booking channels.

**Regression Analysis**

- **Models**: Built regression models to predict `adr` based on `lead\_time`, `total\_guests`, etc.

- **Findings**: Key coefficients indicate that longer lead times and more guests lead to higher `adr`.

**Key Insights & Recommendations**

**Actionable Insights**

1. **Marketing Focus**: Target marketing efforts during summer months and weekends.

2. **Booking Channels**: Optimize partnerships with online travel agents.

**Measurable Results**

- **KPIs**: Track metrics like `adr`, occupancy rates, and cancellation rates.

**Alignment with Business Goals**

- **Revenue Optimization**: Insights align with goals to maximize revenue and occupancy.

**Cost-Effectiveness**

- **Implementation**: Focus on cost-effective strategies like targeted marketing and optimizing booking channels.

**Conclusion**

The analysis provides valuable insights into customer behavior and booking patterns. Implementing the recommendations can help improve hotel operations and increase revenue.

**Appendix**

- **Raw Data**: Available upon request.

- **Technical Details**: Additional analysis and code snippets.

- **Visualizations**: Links to detailed charts and graphs.