EXPLORATORY DATA ANALYSIS

EDA is the process of exploring, visualizing, and summarizing data to gain insights and aiding decision-making.

4 STEP PROCESS

DATA COMPREHENSION

Understand dataset structure, features, and meanings to identify relevant information for analysis.

DATA CLEANING

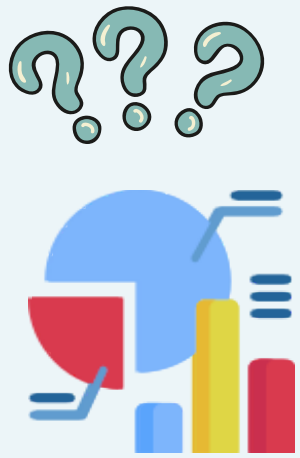
Address missing values, outliers, and inconsistencies for reliable and accurate data preparation.

DATA VISUALIZATION

Present data graphically to reveal patterns, trends, and relationships for intuitive understanding.

DATA INSIGHTS

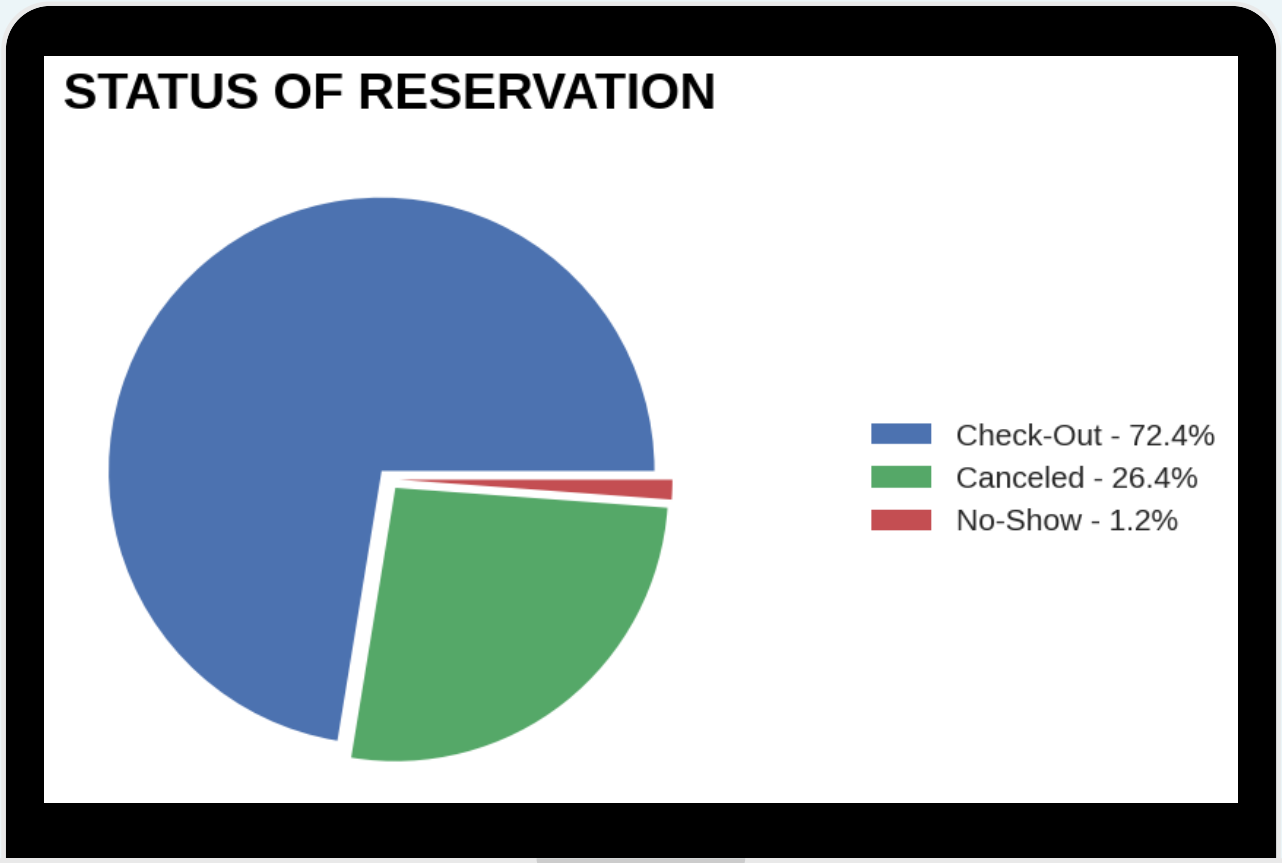
Extract valuable information from visualizations to make informed decisions and draw meaningful conclusions.



Understanding different charts is essenential for EDA to effectively visualize, interpret, and communicate patterns and insights succintctly.

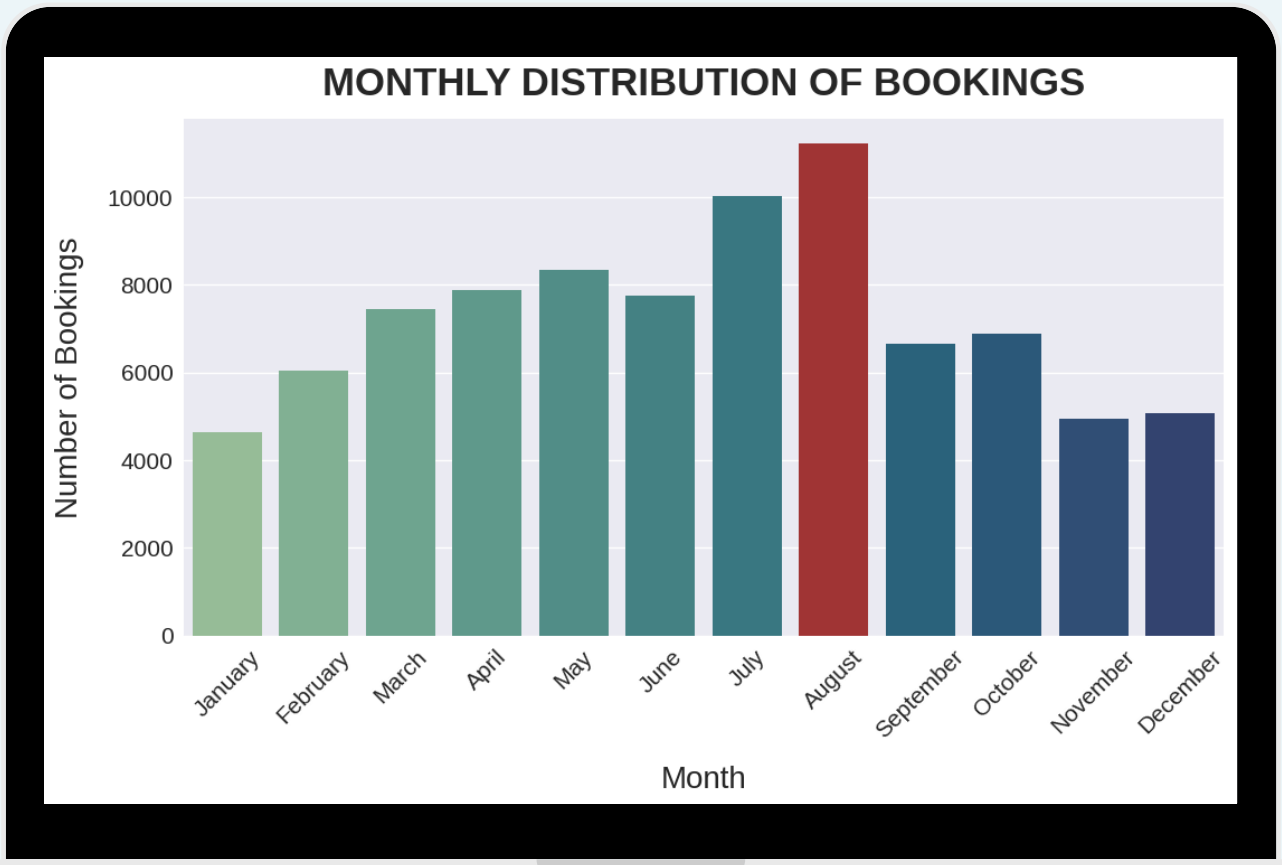
Let's discover the essential visual representations used in EDA

PIE CHARTS



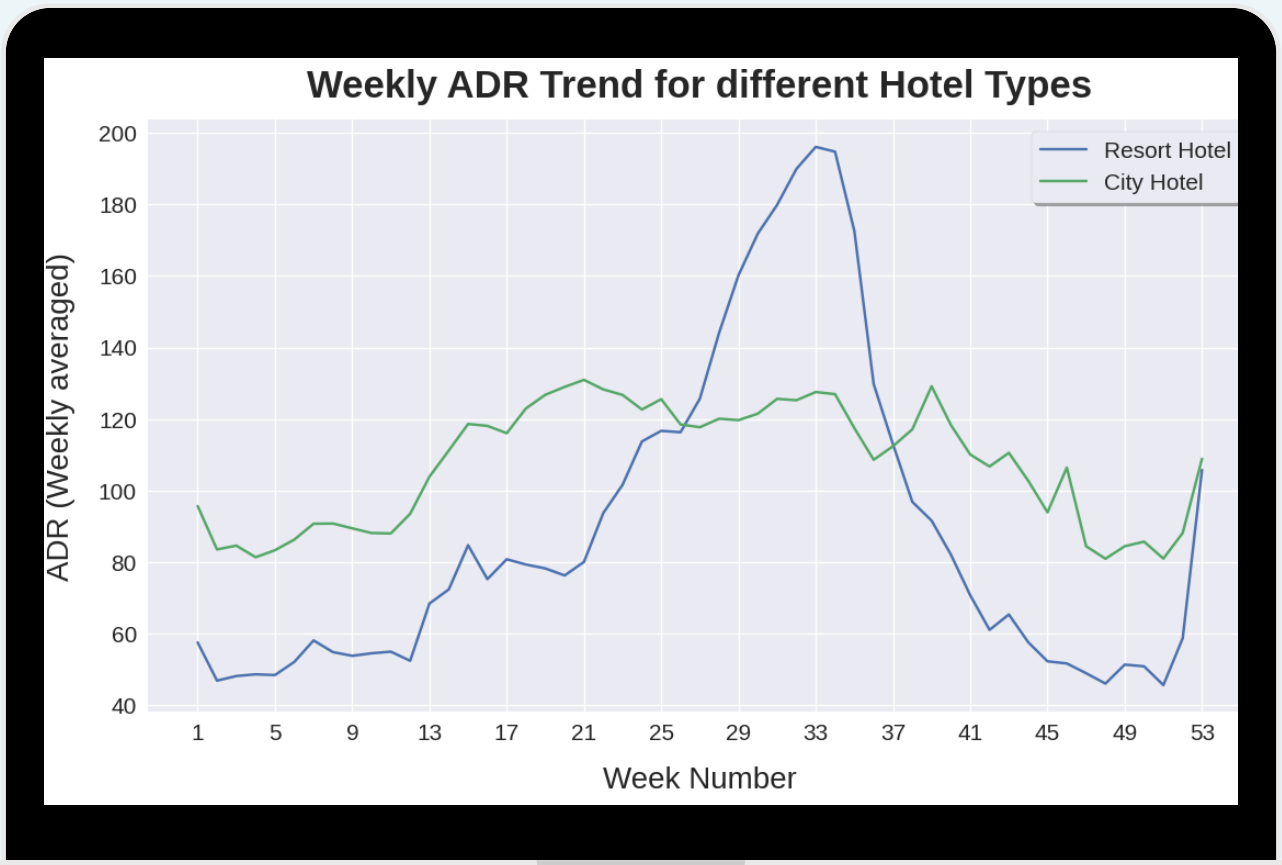
Concise data visualization; illustrates proportionate parts in a circular graph.

BAR GRAPHS



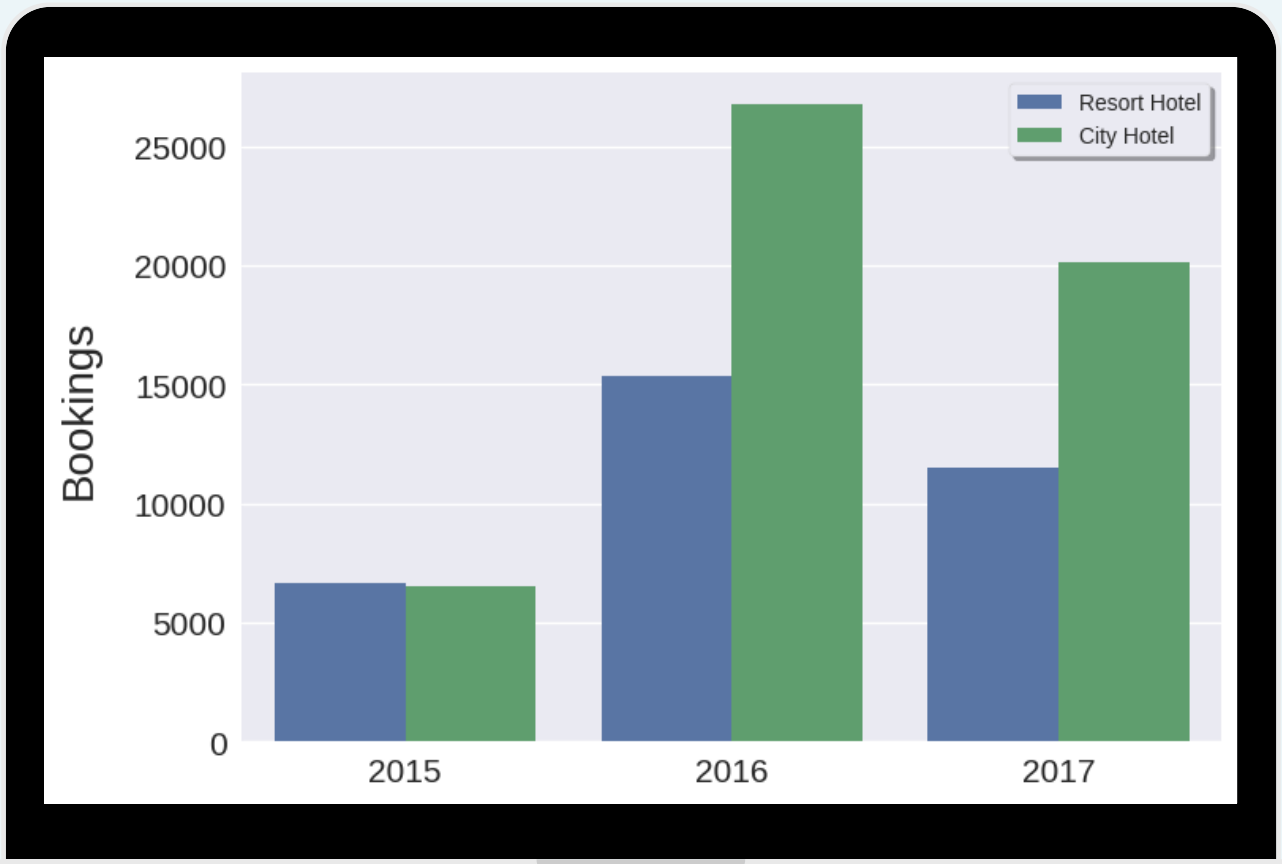
Clear data presentation; compares different categories using rectangular bars for quick insights.

LINE GRAPHS



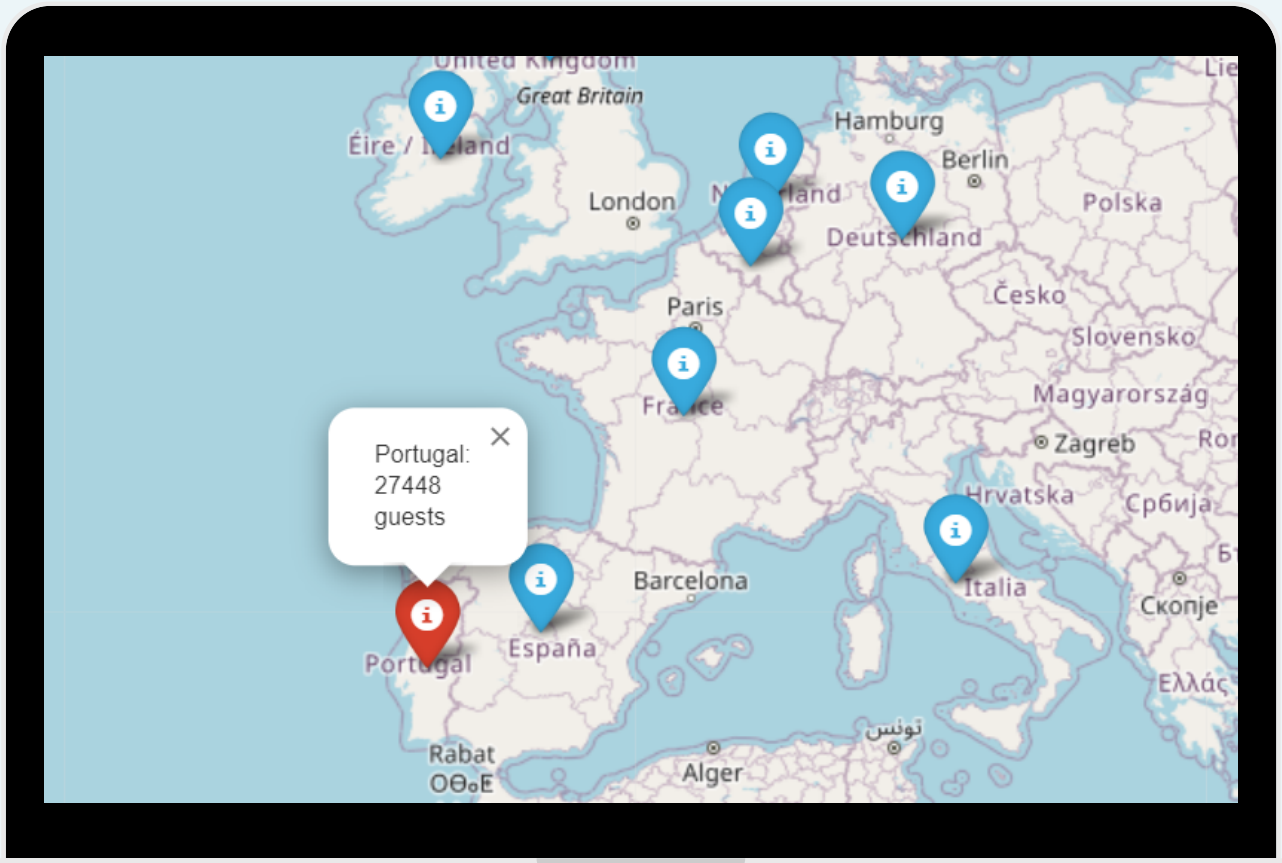
Graphical representation of data trends over time, using points connected by lines for visual analysis.

GROUPED BAR CHARTS



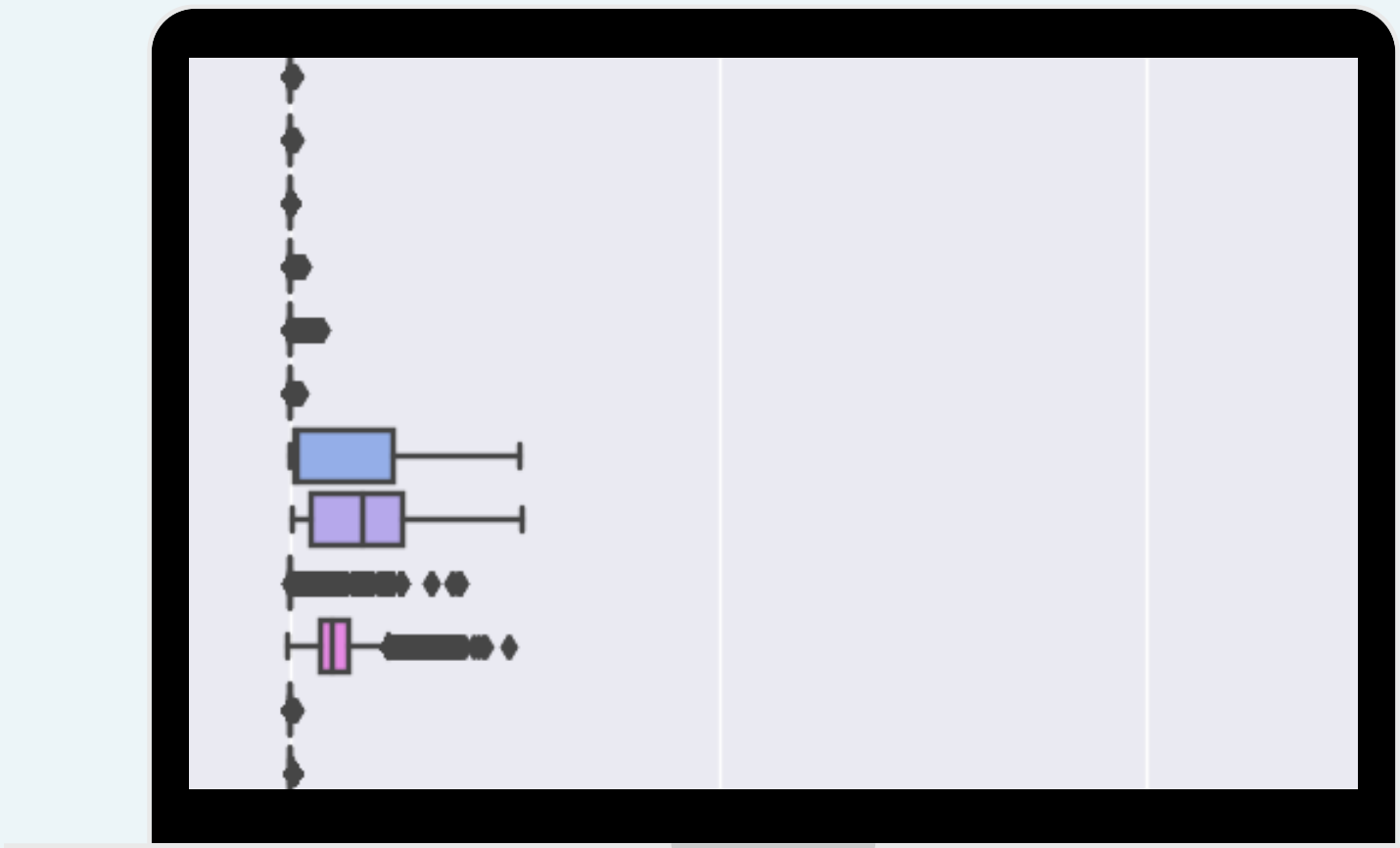
Compare data between multiple groups, using bars side by side for easy visual comparison

GEOGRAPHICAL PLOTS



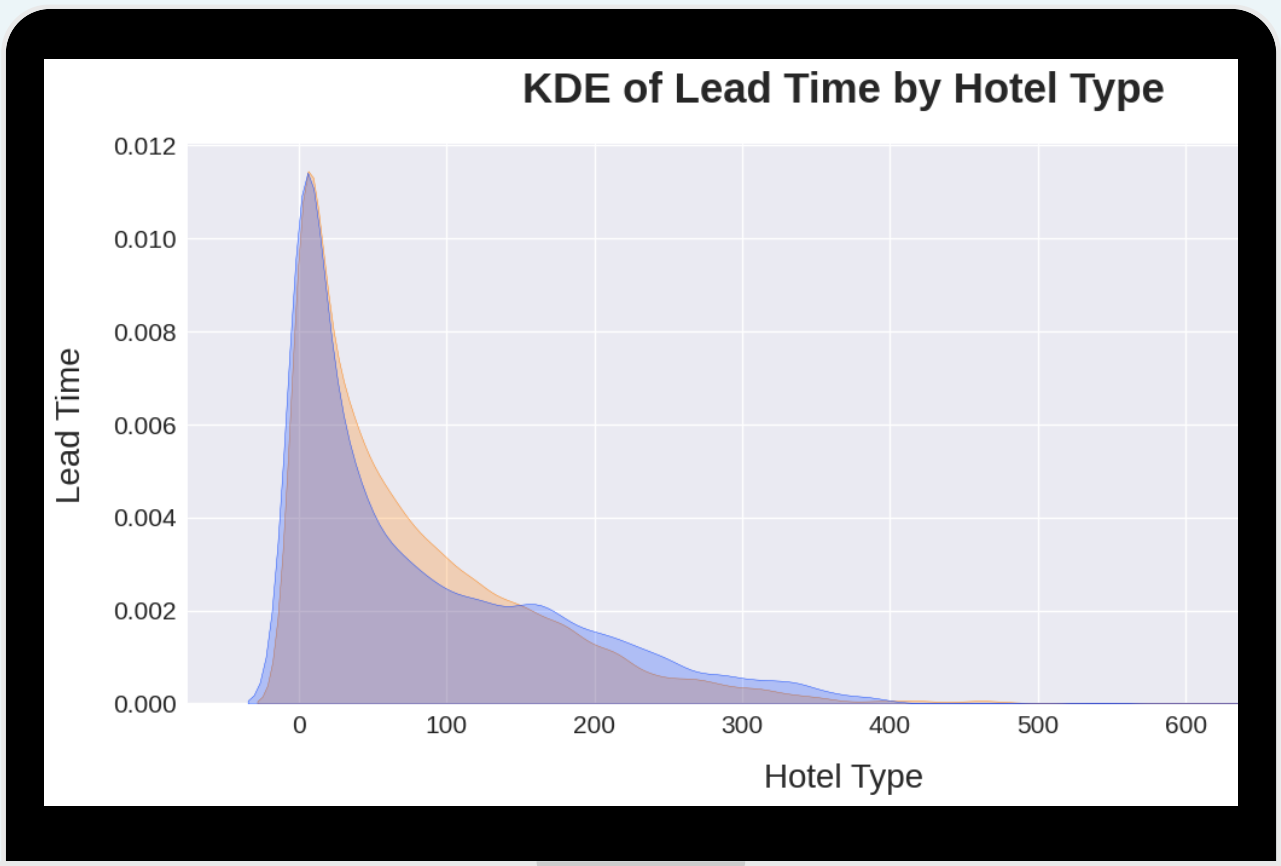
Represent data on a map, indicating spatial distribution and patterns for location-based insights.

BOX PLOTS



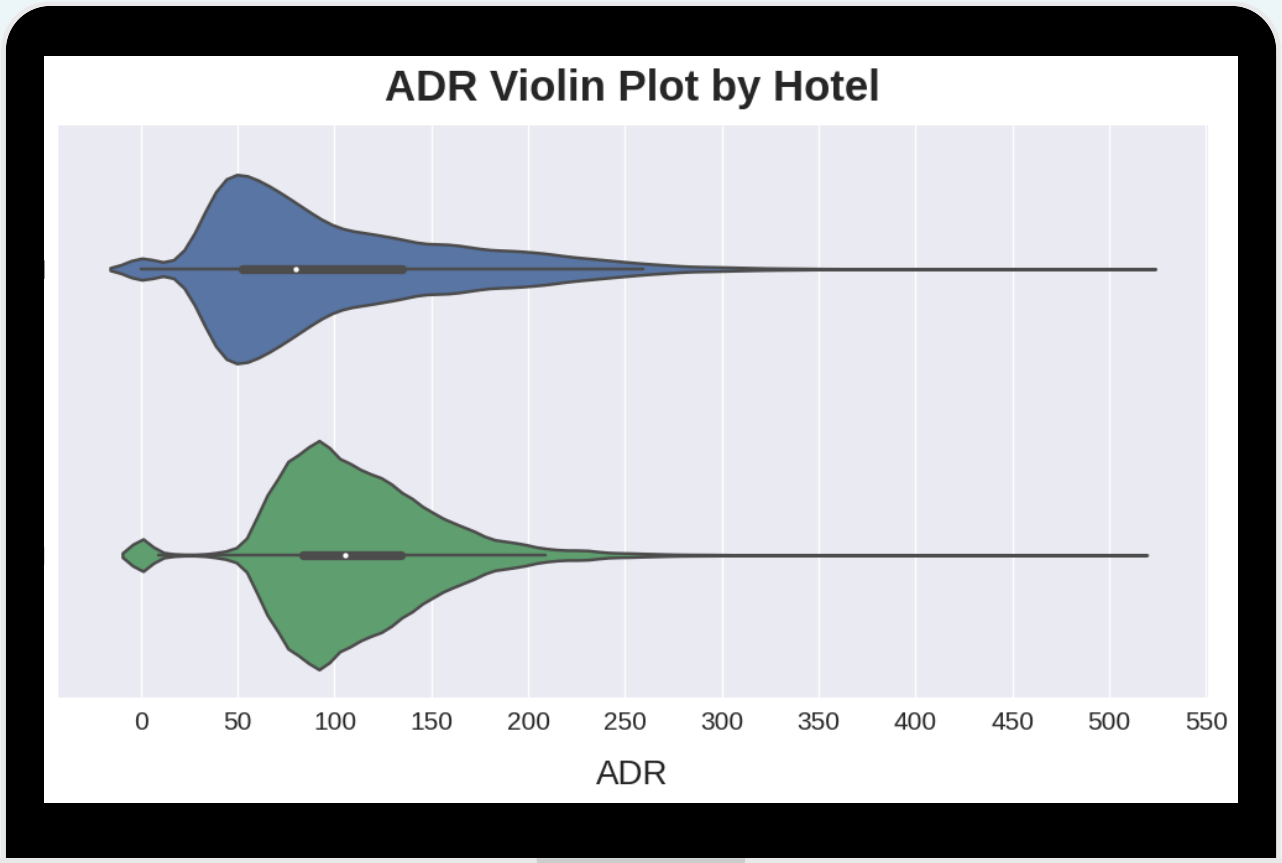
Visualize data distribution and identify outliers using quartiles, median, and whiskers in a concise graph.

KDE PLOTS



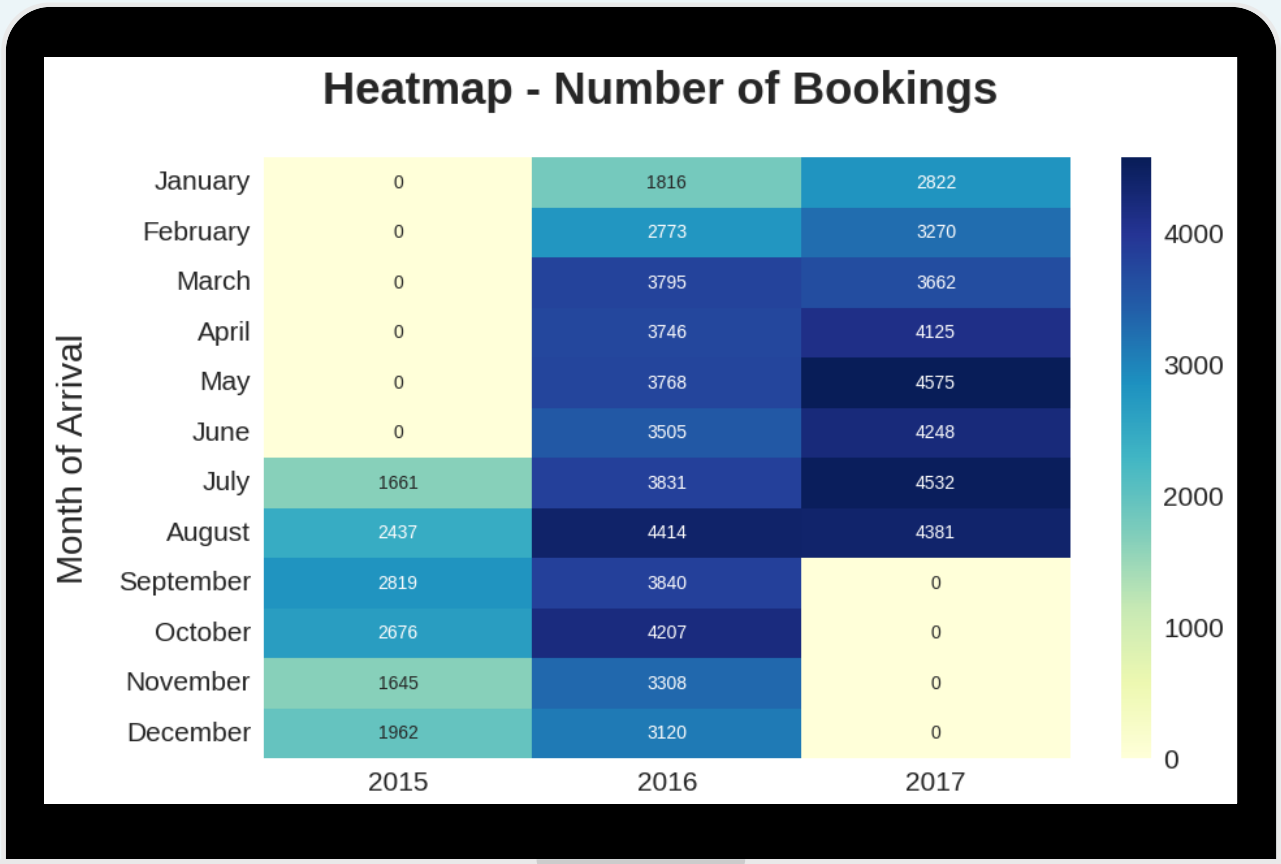
Visualize data distribution using smoothed kernel density estimation for a continuous representation of probability density.

VIOLIN PLOTS



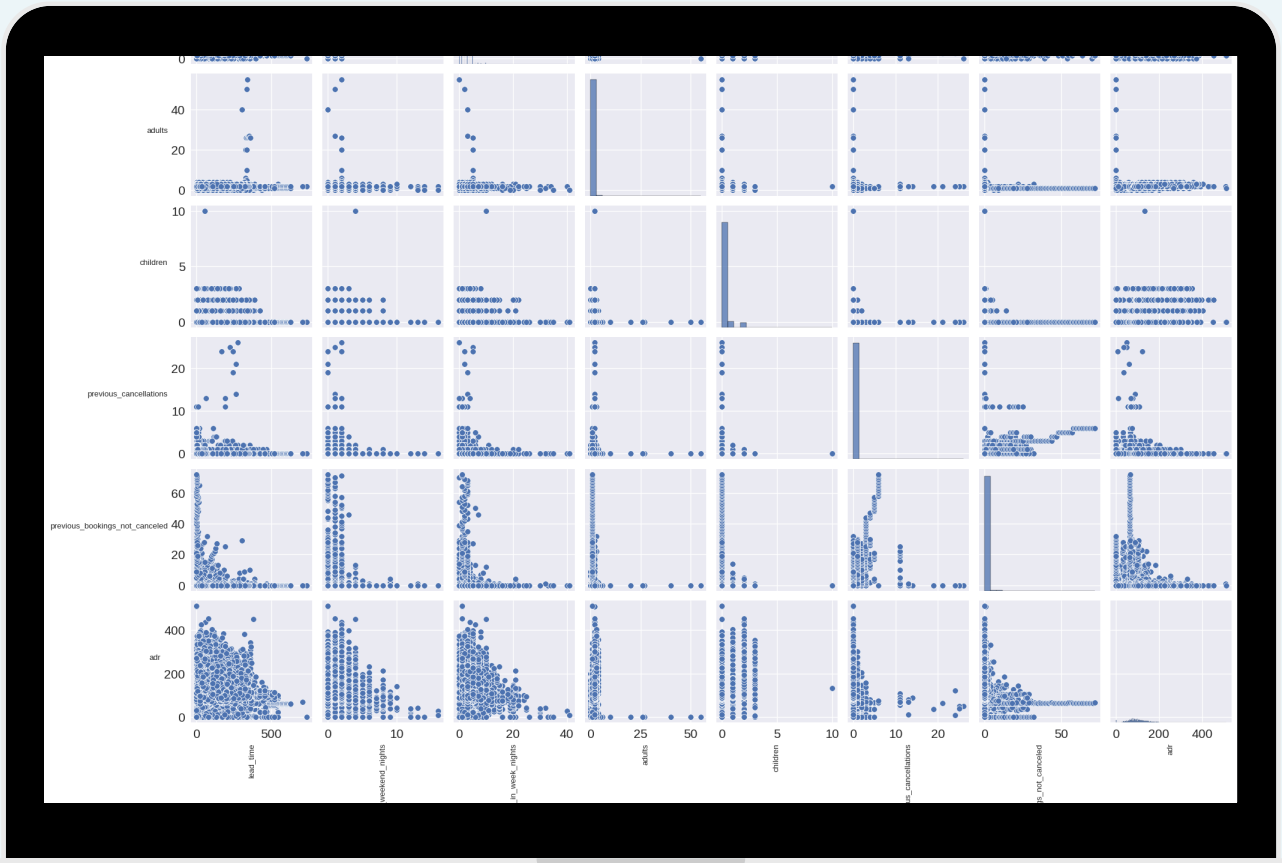
Combines box plot with kernel density plot; displays data distribution and density estimation graphically.

HEATMAPS



Graphical representation of data with colors; shows values on a 2D matrix for pattern analysis.

PAIR PLOTS



Visualize relationships between multiple variables in a dataset using scatter plots and histograms simultaneously.

Data analysis must align with objectives

for instance, improving a business requires identifying insights impacting it positively and negatively.

- **Implement** actions to increase profit
- **Mitigate** negative impacts based on trends

The graphs in this document are extracted from my EDA project focusing on hotel bookings.

Exploratory Data Analysis

Name - Harsh Verma

Project Summary -

This EDA project aims to explore a hotel booking dataset, encompassing data from both a city hotel and
identifying information removed. The primary objectives include determining the best time of year to bo
optimal length of stay for obtaining the best daily rate, and predicting whether a hotel is likely to receive
number of special requests. By analyzing the data, we seek to discover crucial factors that influence hotel b
optimization of operations and enhancing overall customer experience.

GitHub Link -

https://github.com/hkv-code/EDA-Hotel_Booking_Analysis

Business Objective

The business objective is to increase the amount of bookings and optimize operations.