Hotel Booking Data Analysis Project

By: Alishba Naeem

Date: July 2025

Tools: Python, Pandas, Matplotlib, Seaborn

ABOUT THIS PROJECT

In this project, I worked on analyzing hotel booking data using Python.

The main goal was to understand customer behavior, booking trends, and cancellation patterns.

I used libraries like Pandas for data handling and Seaborn/Matplotlib for visualizations.

By creating different types of charts, I was able to identify useful insights such as which months had more cancellations, which customer type books more often, and how lead time affects bookings.

This project helped me improve my data analysis and visualization skills, and also gave me a better understanding of how real-world hotel data can be explored using Python.

Dataset Description

The dataset used in this project contains information about hotel bookings made over a period of time.

It includes both city hotels and resort hotels.

There are many columns in the dataset, such as:

- Booking dates
- Customer type
- Number of adults and children
- Lead time (days between booking and arrival)
- Special requests
- Whether the booking was canceled or not

This data helps us understand how customers behave, when they book more, why some bookings get canceled, and what factors might affect those bookings.

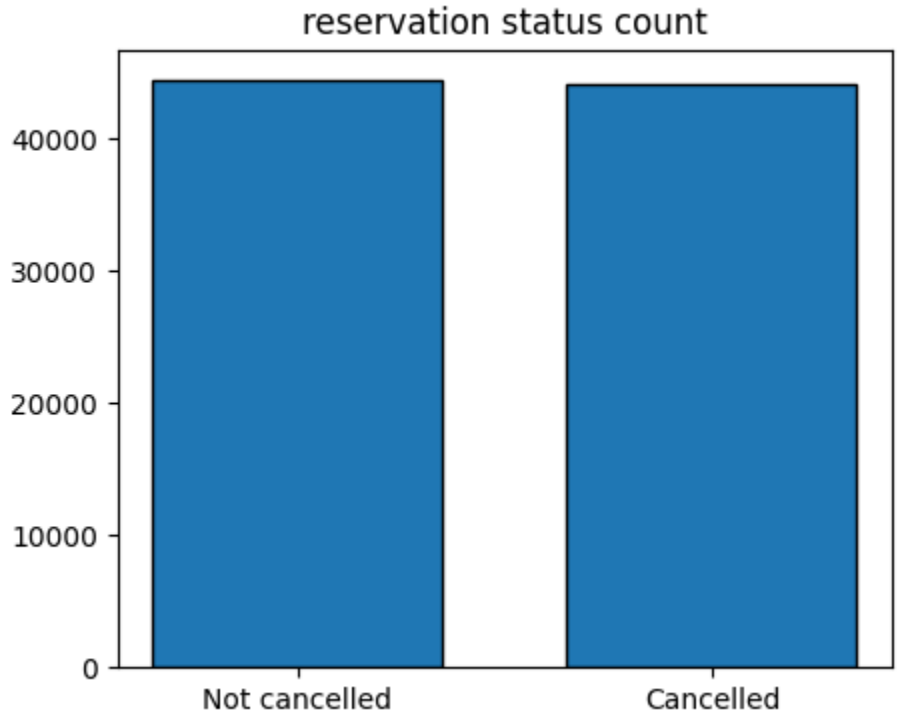
The dataset is suitable for performing both data cleaning and visual analysis.

Analysis & Graphs Explanation

Reservation Status Count

This bar chart shows the number of reservations that were either canceled or not canceled.

It gives us an overview of how many bookings were successful versus how many got canceled.



We can see that the number of canceled and not canceled bookings is almost equal.

This means a large portion of people booked the hotel but later canceled their reservation.

Reservation Status in Different Hotel Types

This chart compares the number of canceled and not canceled bookings in Resort Hotels and City Hotels.

We can observe that:

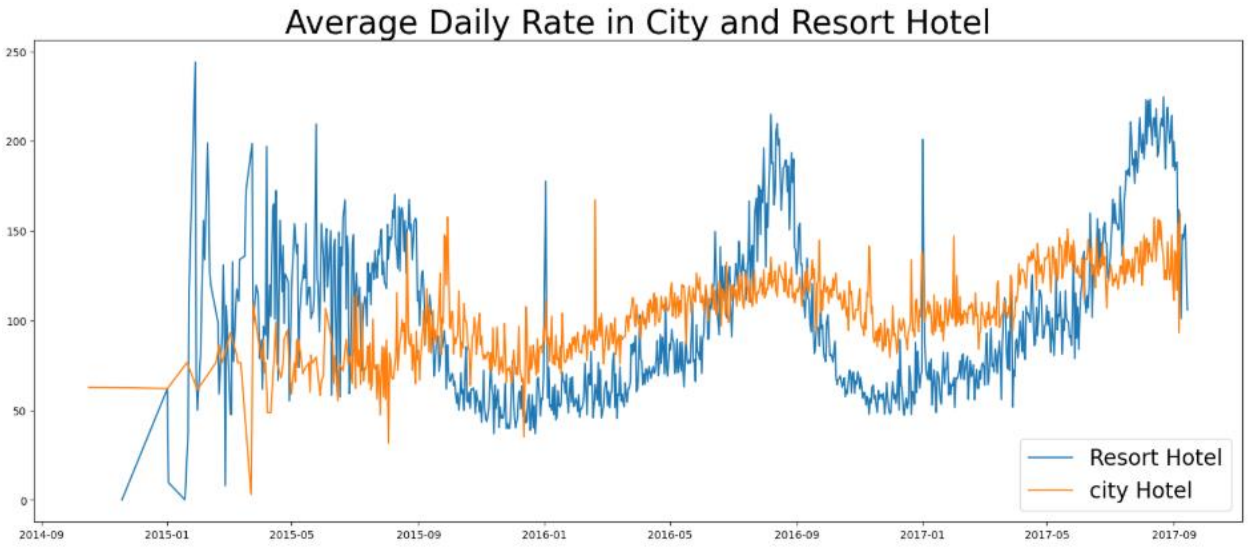
- In Resort Hotels, the number of not canceled bookings is higher than canceled ones.

- In City Hotels, canceled bookings are more than not canceled ones.

This shows that cancellation behavior varies based on hotel type.

Average Daily Rate in City and Resort Hotel:

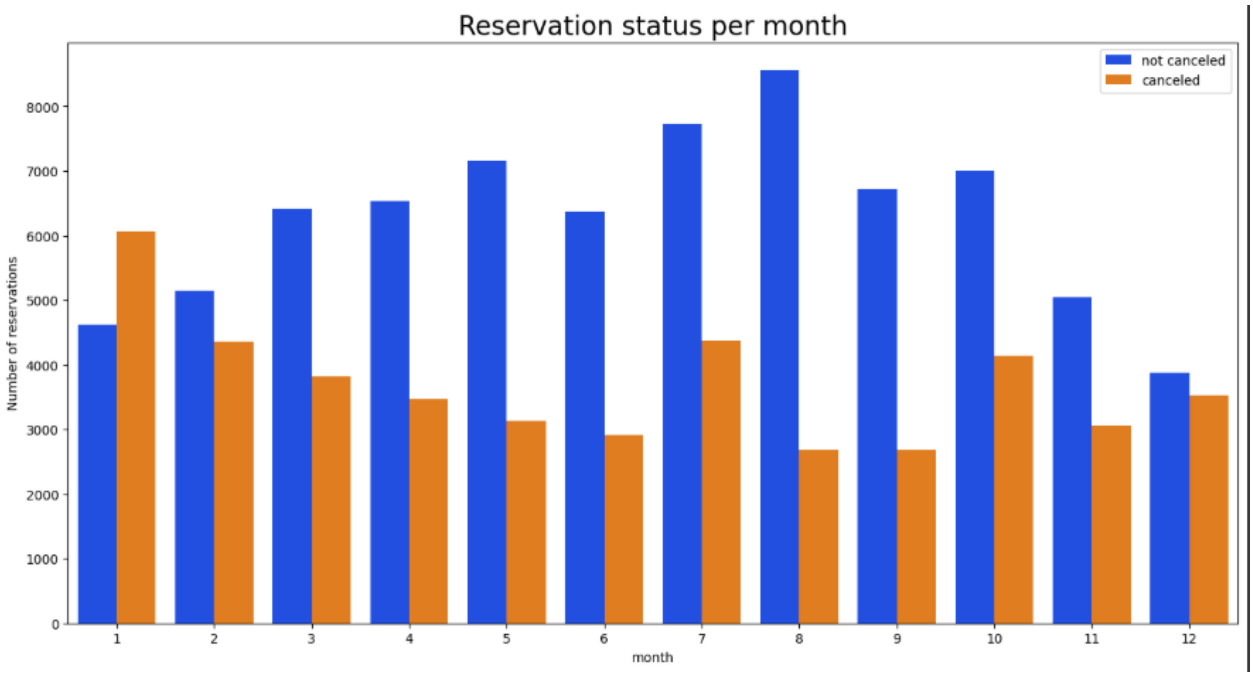
This line graph represents the trend in Average Daily Rates (ADR) for City Hotels and Resort Hotels from late 2014 to mid-2017. It compares how pricing differs between the two types of hotels over time.



- **Resort Hotels** (blue line) show frequent price fluctuations, mainly due to seasonal demand like holidays, school breaks, and weather changes. Since they attract leisure travelers, their pricing responds to vacation seasons — especially in summer.
- **City Hotels** (orange line) maintain more consistent prices, with minor ups and downs. These hotels mostly serve business travelers and professionals, leading to steady year-round demand.
- From mid-2016 to mid-2017, Resort Hotels recorded significantly higher ADRs, indicating that people are willing to pay more for leisure stays during peak times.
- The graph also shows that Resort Hotels are more sensitive to market conditions, while City Hotels follow a more stable and competitive pricing strategy.

<u>Reservation Status Per Month</u>:

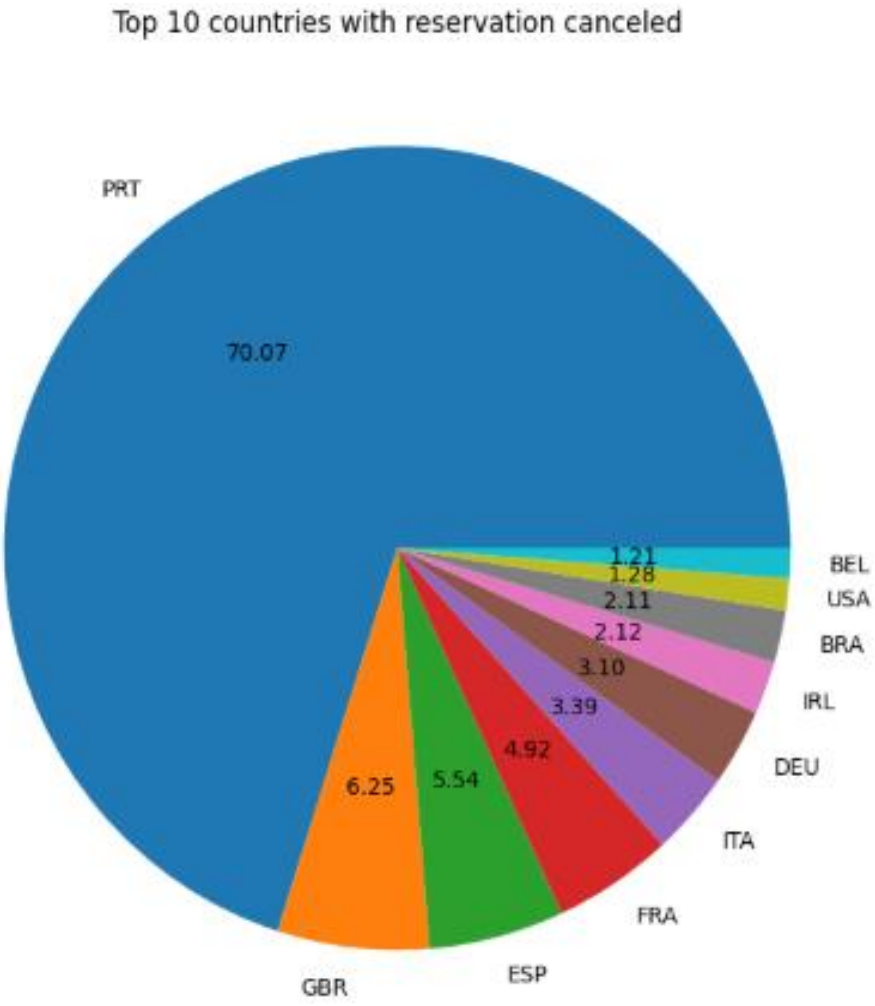
This bar chart displays the monthly comparison of hotel reservations that were canceled (orange bars) versus those that were not canceled (blue bars) throughout the calendar year.



- The **blue bars**, showing non-canceled bookings, are higher in most months, indicating that successful reservations are more common overall.
- **July and August** have the highest number of confirmed bookings, with August crossing 8500, highlighting peak tourist season likely due to summer vacations.
- **January** is unique in showing more cancellations than bookings, possibly due to post-holiday exhaustion, bad weather, or limited travel budgets after December.
- From **March to October**, there's a consistent trend of strong bookings with fewer cancellations, indicating a reliable business season for hotels.
- **December** shows average booking numbers with moderate cancellations, possibly due to mixed holiday plans.

<u>Top 10 countries with the highest number of hotel reservation canceled:</u>

This pie chart presents the top 10 countries with the highest hotel reservation cancellations. Each country is labeled using its ISO code, along with its share (%) of total cancellations.



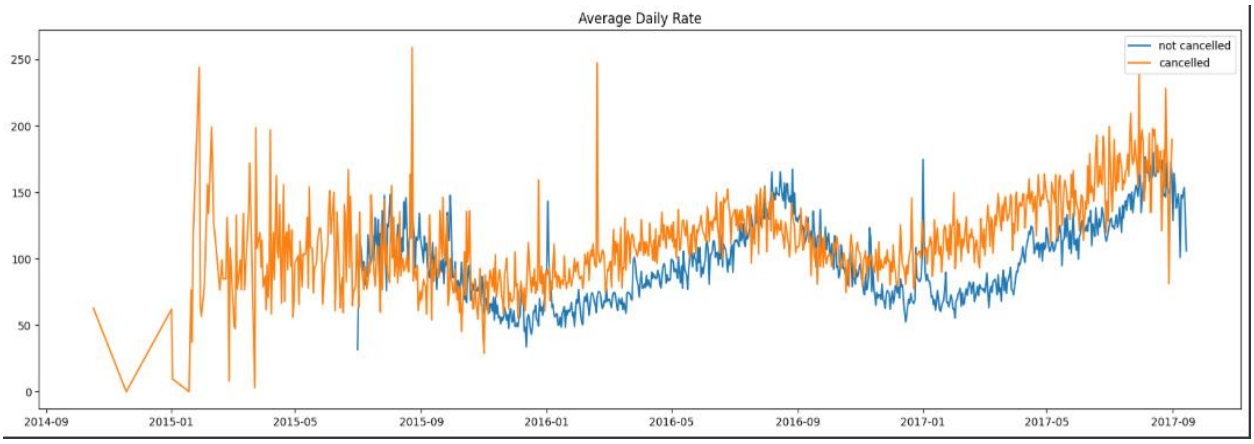
- **Portugal (PRT)** leads by a large margin with **70.07%** of total cancellations. This suggests either a high volume of bookings from Portugal or that domestic guests tend to cancel more often.
- Other major contributors include:

    - **Great Britain (GBR):** 6.25%
    - **Spain (ESP):** 5.54%
    - **France (FRA):** 4.92%
    - **Italy (ITA):** 3.39%
    - **Germany (DEU):** 3.10%
- Countries like **Ireland (IRL)**, **Brazil (BRA)**, **USA**, and **Belgium (BEL)** make up the remaining share, each contributing between 1.2% and 2.1%.
- The large number of cancellations from Portugal could indicate flexible or spontaneous local travel behavior, where guests might cancel easily due to short-distance or last-minute changes.
- The rest are mostly **European countries**, which aligns with the likely location of the hotels in the dataset (probably within Europe).
- This information can help hotel managers **identify high-cancellation regions** and adapt their booking policies, such as adding stricter cancellation rules or requiring upfront payments for certain markets.

Analyzing which countries show the highest cancellation rates helps hotels with **risk assessment**, **marketing decisions**, and **revenue strategies**. By focusing on regions with higher cancellation tendencies, hotels can reduce losses and better manage their booking flow.

Average Daily Rate:

This line graph compares the **Average Daily Rate (ADR)** for hotel bookings that were **cancelled** (orange line) versus **not cancelled** (blue line), across a timeline from late 2014 to mid-2017.



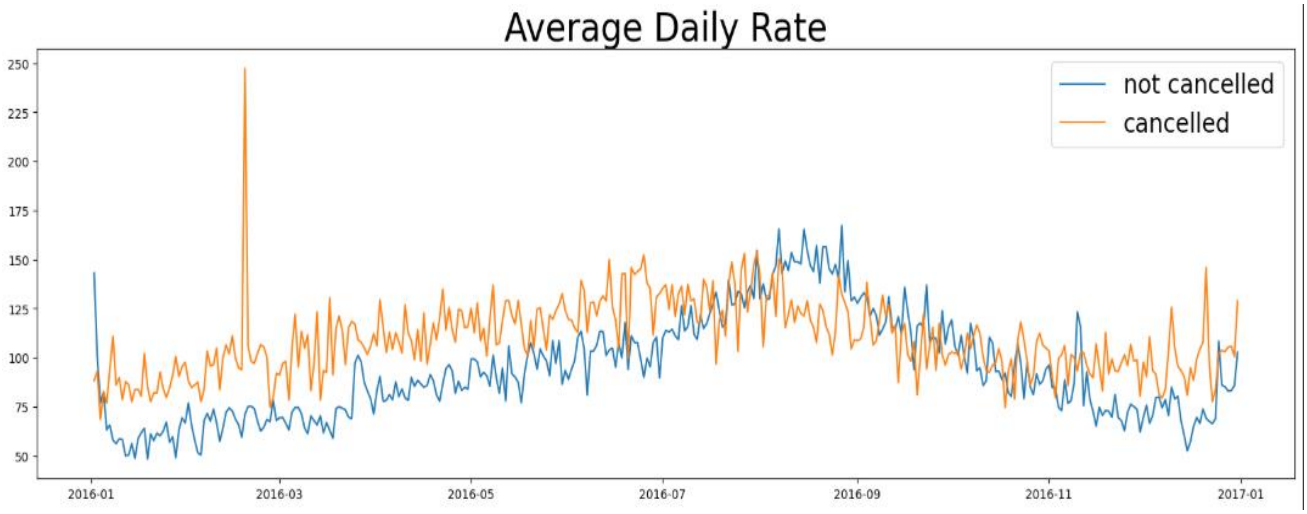
- Overall, **cancelled bookings** tend to have a **higher ADR** than non-cancelled ones. This suggests that **guests who pay more are also more likely to cancel**, possibly due to luxury or optional leisure travel behavior.
- During peak seasons (especially mid-2016 to mid-2017), both lines show an upward trend, but **ADR for cancelled bookings remains more volatile**, showing sudden spikes and dips.
- The **non-cancelled ADR** follows a more **stable and gradual rise**, indicating consistent pricing and possibly more reliable customer segments.
- In early 2015, the cancelled ADR shows **sharp irregular spikes**, possibly due to outlier events, pricing experiments, or inconsistent booking behavior.
- The **gap between cancelled and non-cancelled ADRs** becomes clearer from mid-2016 onwards, where the orange line (cancelled) stays mostly above the blue line (not cancelled).

This graph reveals that **higher room rates are often linked to cancellations**, possibly due to flexible plans or less commitment from high-paying guests. In

contrast, guests who book at **moderate prices are more likely to follow through with their stay**.

Average Daily Rate:

This line graph shows a comparison of the **Average Daily Rate (ADR)** for hotel bookings that were **cancelled** (orange line) and **not cancelled** (blue line) during the period from **January 2016 to January 2017**.



- For most of the year, **cancelled bookings (orange line)** have slightly **higher ADRs** compared to non-cancelled ones, suggesting that higher-priced bookings are more likely to be canceled.
- The **blue line** (not cancelled) shows more **stable pricing**, while the **orange line** has occasional spikes, with one noticeable **peak above 240** in early 2016.
- From **May to October**, the ADR for both cancelled and non-cancelled bookings trends upward, showing increased pricing during what appears to be a **busy or peak season**.
- From **November onwards**, the ADRs for both categories begin to decline slightly, showing the **post-peak season drop** in hotel rates.
- The **gap between the two lines** fluctuates but remains visible throughout the year, reinforcing the pattern of higher ADRs for bookings that eventually get canceled.