# Hotel Bookings

Since my initial proposal, I learned the dataset I found does include pricing information in the form of “adr” or average daily rate. First, that shows it’s important to fully understand your data before attempting to formulate questions or work with it. Second, that changed my actual research to when’s the best time to travel and which booking method allows you to do so cheapest. The dataset I used was found on [Kaggle](https://www.kaggle.com/datasets/jessemostipak/hotel-booking-demand), contained 32 features, and 119,390 entries from 2015-2017.

My approach to determining when the best time is to travel is based on when the fewest people are booking. Like many others I enjoy being around people, however, when I’m on vacation I don’t want to wait in lines or feel like I’m right on top of someone else. Additionally, to find the cheapest rates I explored the various “market segment” types within the dataset. Most bookings were from “TA/TO” (travel agent or travel organizations), “Direct”, or “Corporate”. There were a few others like “Aviation”, “Compensatory”, etc. However, I wouldn’t be using these booking methods, so I eliminated them from my visuals.

The data shows what most preconceived notions would have you believe. Here’s a few key takeaways:

1. The fewest people are booking (and staying) between November -- March.
2. August is the busiest and most expensive month to stay.
3. Prices start rising in April and begin coming down in September.
4. Booking through a site like Hotels.com typically yields cheaper rates.

A picture containing line, plot, diagram, slope

Description automatically generated

A picture containing plot, line, slope, diagram

Description automatically generated

A picture containing diagram, plot, line

Description automatically generated

Also, the data shows people are less likely to cancel bookings at a Resort Hotel when compared to a City Hotel. A surprising takeaway is cancellations are at roughly 37% although when looking at solely the United States it’s roughly 25% cancellations. Attendance of kids is proportionally greater at resort hotels vs city hotels but not by much. Across both hotel types, kids only make up ~5% of visitors.

When creating visuals, my goal was to keep as much information digestible solely based on the plots title. By using the same color schemes on each of my pages, it allows for ease of scanning the page to understand the results. For example, on the “Bookings” page it shows Completed Bookings vs Cancellations by Month and Percent Cancellations by Hotel Type. In both of those plots it’s using green for completed bookings and red for cancellations. Naturally we (Americans) translate green as go and red as stop so it’s a natural perception of the data without the need for a legend (although it’s included for both). Lastly, my goal was to limit overloading data in graphs. In almost all my plots there’s a maximum of three things to analyze, which aids in the ease of digestion.

Based on my findings, the best months to travel are November – March. The cheapest method to stay is always from a corporate booking site (perks of some companies) and second best is through a travel agent site like hotels.com.

The dashboard I created can be viewed at <https://campb223.shinyapps.io/hotel-bookings/>

The full code and analysis can be found at <https://github.com/campb223/hotel_bookings>