

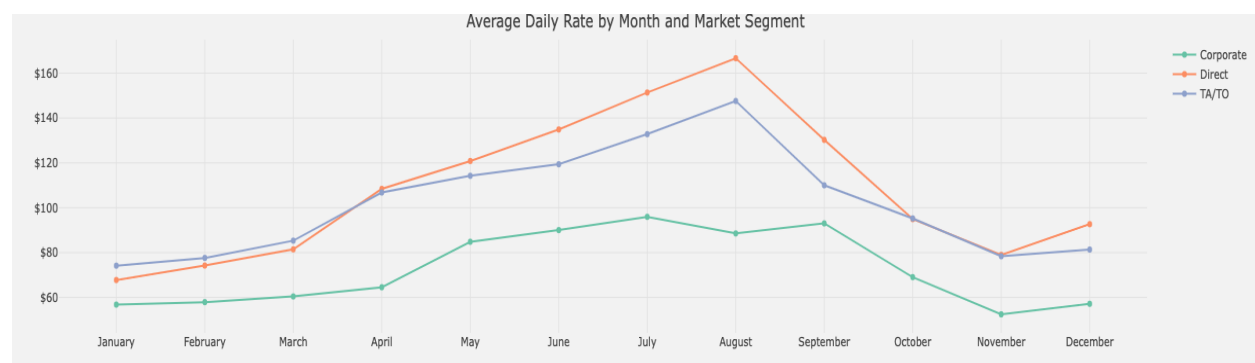
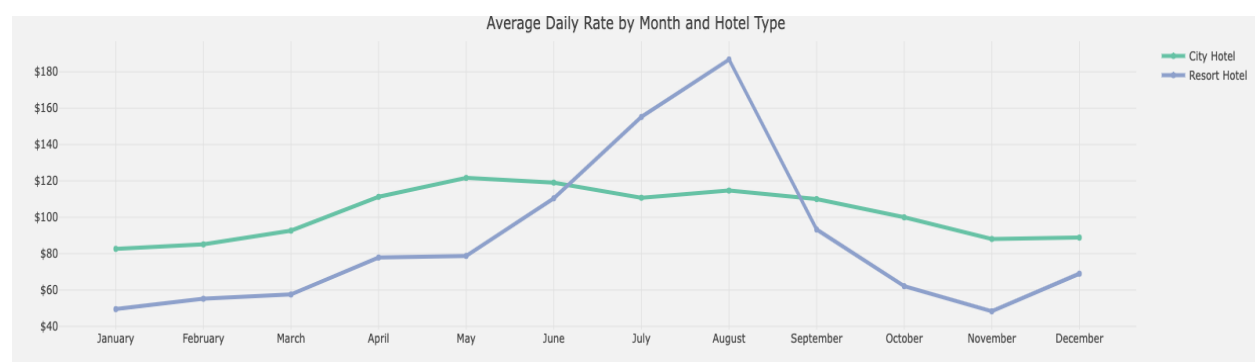
Hotel Bookings

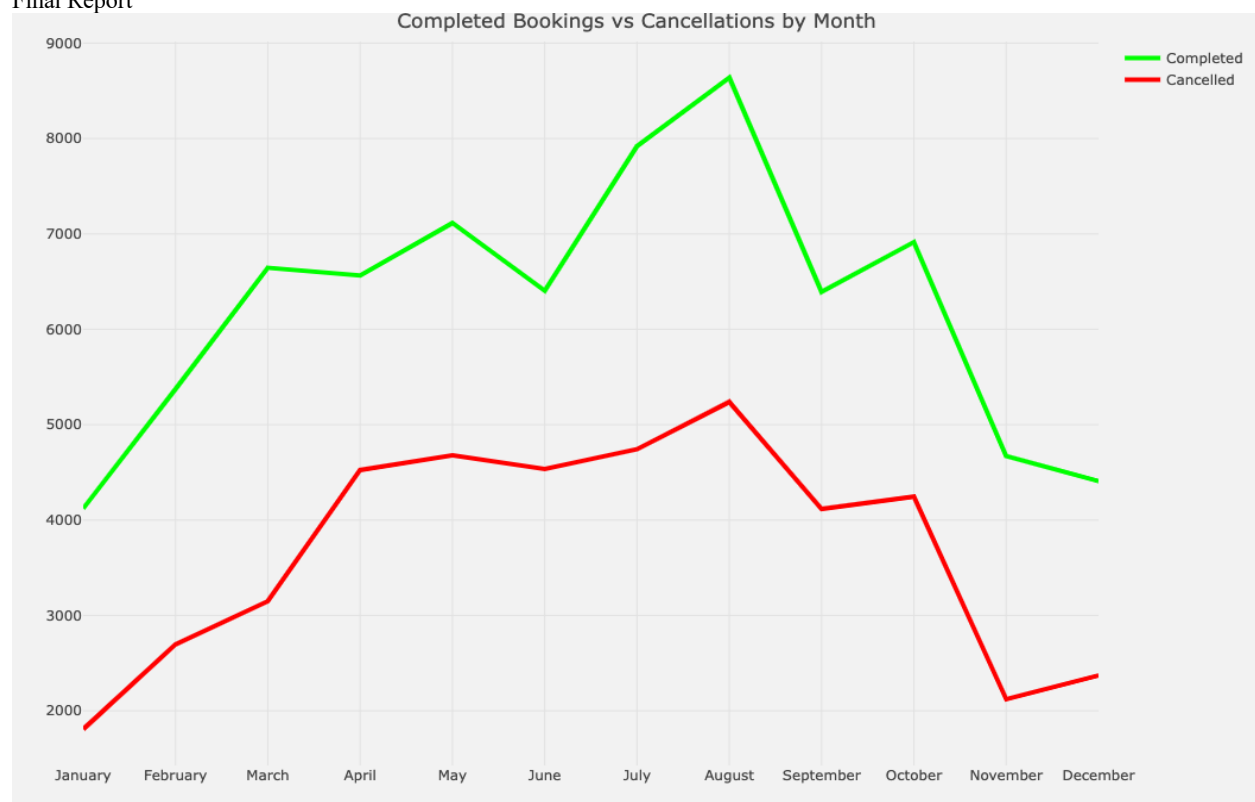
Since my initial proposal, I learned the dataset I found does include pricing information in the form of “adr” or average daily rate. First, that shows it’s important to fully understand your data before attempting to formulate questions or work with it. Second, that changed my actual research to when’s the best time to travel and which booking method allows you to do so cheapest. The dataset I’m using to answer these questions can be found on [Kaggle](#). It contains 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 (and staying). 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, I explored the various “market segment” types within the dataset to find the cheapest rates. Most bookings were from *TA/TO* (travel agent or travel organizations), *Direct*, or *Corporate*. There were a few others like *Aviation*, *Compensatory*, etc. However, few people will have the ability to use these booking methods, so I didn’t include them in my results.

The data shows what most preconceived notions would have you believe. Here are 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 your corporate (or company portal) will be the cheapest option.
5. Booking through a site like Hotels.com typically yields cheaper rates vs booking direct.





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 roughly 37% of bookings are cancelled although when looking solely at 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 guests.

When creating visuals, my goal was to keep information digestible solely based on the title of the plot. By using similar or 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. By using these colors to plot data, the viewer could easily make the mental leaps to understand the data without the need for a legend (although it's included in most plots). Lastly, my goal was to limit overloading data in graphs. Virtually all my plots have a maximum of three things to see or analyze, which aids in that ease of digestion.

All in all, based on my interpretation of the questions proposed, 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 the next best option is typically through a travel agent site like hotels.com.

Dashboard: <https://campb223.shinyapps.io/hotel-bookings/>

Full Code: https://github.com/campb223/hotel_bookings