Dale Campbell
DSSA-5103 – Data Visualization
Topic Submission

Hotel Bookings

I've found a dataset on hotel bookings from 2015 – 2017 on Kaggle. Wildly enough it was cleaned and featured on TidyTuesday on February 11th, 2020. It doesn't have pricing and it appears to mostly be international hotels, but it contains enough data to draw some insights from. I love to travel but hate spending more money than I think something should cost. I hope to see the patterns of when the most popular booking months, weeks, and days are as well as when cancellations are most common. Based off the findings, in my personal life I plan to use that to evaluate hotel pricing during those months to see if the trends hold true, although that won't really be showcased in this project.

I'm most comfortable with Python so I plan to do my initial data analysis in Python. However, to build a dashboard that tells the story of my findings, I plan to use a R Shiny Web App. Having the calculations in R & Python should give some reassurance the calculations are correct, or at least consistent. My measure of success is more on how helpful the dashboard is in telling the story. At this point, I haven't deep dived into the data so I'm not sure what that story is. I'm comfortable with EDA so I'd like to use this as an opportunity to focus on building my visualization skills.

I plan to analyze and visualize the following but of course these could change if the data points me new directions:

1. Average lead time to stay
2. How often reservations are cancelled
3. Number of repeat guests
4. Average length of stay
5. Typical number of adults, children, and babies
6. Deposit types
7. Resort types

The dataset contains 32 features with 119,390 entries:
https://www.kaggle.com/datasets/jessemostipak/hotel-booking-demand