**Datasets**

* Covid cases of TSA officers by airport: <https://www.tsa.gov/coronavirus>
* Daily flights by day, this year vs. last: <https://www.tsa.gov/coronavirus/passenger-throughput>

**Outline**

Traveling During Covid? Consider Data Science.

* Introduction
  + It’s a difficult, scary, or uncomfortable time to travel
  + For many, this is a scary time to leave home. Unfortunately, for many, it’s the time of year when we want to travel most. The conjunction of a pandemic and the holiday season creates a difficult conundrum, often without a perfect answer.
  + Data science can help is make better informed decisions
  + The idea of applying this methodology to holiday travel was inspired by a close friend and aspiring data scientist, Sam.
* Higher exposure means higher risk
  + The more people you’re exposed to, the higher your risk of contracting Covid. In other words, busier airports are more dangerous.
  + Let’s determine when airports are least busy
* Solving the problem
  + TSA provides data on the number of passengers traveling every day since March for 2020 alongside the same data for 2019. Note that they adjust the data such that a side-by-side comparison of data points corresponds to the same day of the week, as opposed to the same day of the month, as it tends to be a more valuable comparison when considering how travel volume changed one year to the next.
  + Overall, travel volume in 2020 is much lower than in 2019
    - Provide overlay of two area plots showing 2019 vs. 2020 travel volume since March
  + Value comes in comparing *relative* risks. The risk of catching Covid while flying in 2019 was null, thus comparing travel volume directly to 2020 to determine *risk* isn’t valuable. Rather, the value comes in investigating when it is safest to travel, under the assumption that traveling via plane is necessary and the only option.
  + Thanks to the public data provided by TSA, let’s look at travel frequency by day of the week for the last \_\_\_ days
    - Provide horizontal bar graph of day of the week travel volume
    - Although at immediate glance, metrics look similar. However, at closer investigation, there are some key findings:
      * \_% fewer people travel on \_ compared to \_
* Proxies
  + Keep in mind that there are often valuable, more simplistic proxies when considering questions such as these.
  + As it turns out, the relative price of flights tends to be a good proxy for the busyness of airports.
  + The price as a proxy is one that I speculated about before conducting the analysis – a technique I have used in the past.
  + The power of data science is that we don’t always have to use proxies. We get to conduct the analyses that can later lead to the discovery of valuable proxies.
* Caveats worth considering
  + We are living in strange times. This means that historical patterns often break down
    - We have of course seen the volume of travel diminish dramatically [cite numbers from TSA]
    - Was the relative breakdown of travel volume by day of the week dramatically impacted? That’s hard to say.
    - The best thing we can do is work off the data we have and use data for adjustments as possible
  + The data considers only domestic flights. It also assumes that the flight will take place over the course of exclusively one day, as opposed to longer flights spanning two days.
  + Busier airports don’t necessarily correlate to busier flights, though they’re a good *proxy* – a more comprehensive analysis would consider additional datasets, if available, to address this issue.
* Conclusions
  + Ideally, don’t travel. However, if you do travel, these are the lessons we have gathered:
    - If possible, travel on [day]. If not, [] are also good options. Avoid [] if at all possible – you’ll encounter \_% more people than on [].
    - Take the shortest flight possible with the fewest layovers
    - Comparing 2019 to 2020, we can speculate that \_\_ days of the month will be the busiest. Avoid if possible.
  + Overall, when making decisions for which significant data is available, consider an analysis through data science techniques as a highly informed approach to decision making.