Problem

Air travel is an essential mode of transportation that many people rely on for personal and business purposes. U.S. airlines carried an estimated 849 million passengers in 2017 alone. Despite impressive advancements in air travel systems, commercial flight delays are still a fairly common occurrence. Delays can negatively impact passengers, airlines, and airports in significant ways. They often arise unexpectedly, making it difficult to schedule actual departure and arrival times effectively in advance. This study aims to investigate variables that influence flight delays at the 10 busiest airports in the United States.

Client(s) and How They Benefit

This project primarily serves the Airport Authorities and Major Airline Carriers. The analysis aims to identify variables that influence flight delays, such as weather conditions, airport location, and airline carrier. After the significant variables and their association with delays are identified, airport authorities could potentially better anticipate delays for future flights under certain conditions, which would help them plan airport operations. More accurate scheduling of departures and arrivals would directly improve efficiency of airport operations and management of resources, which may indirectly create other benefits such as improved passenger satisfaction and increased profits.

In addition, if time permits, this analysis aims to focus on the airline carriers and airports that have the worst performance in terms of delays. The analysis will examine these subjects more closely to determine whether there are any specific areas that can be improved upon, or at least investigated further. The airport authorities and airline carriers involved would benefit by identifying specific areas they may wish to evaluate and improve.

Data

Flight data is available from the Bureau of Transportation Statistics. This website provides downloadable On-Time performance data for flights dating back to 1987.
The flight data includes but is not limited to:

- Departure Delay
- Arrival Delay
- Airport ID (Origin and Destination)
- Airline Carrier Code

Weather information is available from NOAA Local Climatological Data. This website provides files that include METARs data. METARs are typically generated hourly, though at some locations they may be generated every 20 minutes.
The weather data includes but is not limited to:

- Wind Speed and Direction
- Precipitation Type
- Visibility

Analysis Strategy

A basic analysis framework is proposed below. Note that the actual scope of the final analysis will depend on a variety of factors which will become apparent while the data is collected and explored. The goal of the basic analysis framework is to identify the significant contributing variables to delays, and to evaluate their association with delay duration. This analysis will not provide a predictive model for delay duration. Such a task would likely exceed the scope of the project due to the amount of variables involved and the complexity of their relationships. Even so, the analysis results could potentially be used to anticipate increased delays under certain conditions.

The first step of this analysis is to collect flight data for the top 10 busiest airports in the U.S. as well as weather data for those airports. The data will only include the current 8 largest U.S. airline carriers, in terms of passengers carried. Due to constraints in data size and collection methods, flight data from the past 2 years will be used in this analysis. Next, exploratory data analysis will identify the significant variables that influence departure delay durations and arrival delay durations. For example, box plots could potentially compare delay distributions between airports or airlines, and time series plots could reveal seasonal or daily trends in delay durations. Correlation matrices could help identify weather variables of interest.

Next, the association between each variable and delay duration will be evaluated more closely. For example, simple regression models could potentially indicate the relationship between weather variables and delay durations. In addition, the analysis will test whether the airline experiencing the worst delays has a different distribution of delays than the distribution for all airlines combined, and similarly for airports.

This basic analysis framework can be expanded upon, if time permits. For example, the airport and airline with the worst performance in terms of delays can each be investigated more thoroughly. The analysis may explore the characteristics of the delays associated with these subjects more closely by including variables such as taxi time, air time, and the broadly defined “cause of delay” category from the flight data. The goal of the expanded analysis is to identify areas that clients may wish to investigate and improve, in order to mitigate delays in the future.

Deliverables

The results of this analysis will be presented in a report.