Christine Luongo

General Assembly: Data Science

Project Problem Statement

Transportation has always been a personal interest of mine and fortunately through my consultant work and a little of my classwork I have been able to come into some contact with transportation analyses. Commercial flights are used occasionally for removal of illegal aliens, however the capabilities of our Analysis team at Immigration and Customs Enforcement are limited. Furthermore, of the mode of transportation air transportation has better tractability. Due to the Law Enforcement Sensitivity our data, I cannot use any data from my consultant work.

The Research and Innovative Technology Administration (RITA) provides detailed data, both on summary statistics and down to the granularity of every US domestic flight.[[1]](#footnote-1) Information is provided pertaining to average fare costs,[[2]](#footnote-2) carrier market share, and on-time performance[[3]](#footnote-3). This data goes back several years but due to the size, only focusing on a single year should suffice.

Due to the size and scope of this data, an exploratory analysis and the tools developed in the course may result in answering the following:

* What factors determine or influence the on-time performance of a flight?
  + Potential variates include, but are not limited to, carriers, planes, planned duration of flight, airports, airport-carrier combinations, weather, time-of-year, time-of-day etc.
  + Using Tail Numbers and sorting by date, the data can be manipulated to see the effect of delay over an entire mission (flight loops). A delay in the first leg of a mission could have a cascading effect on the remainder of the legs. This will also be a good exercise in data manipulation and reshaping.
  + Do commuter routes have better on-time performance?
  + Is there any relationship between average fare and on-time performance?
  + Is there any way to cluster flight paths? Can we see if any characteristics are shared amount flights with similar outcomes?
  + Can we cluster airports or airport-path combinations and see which airports are most similar in characteristics and in on-time performance?
* Due to the sheer size of the dataset, it might be wise to subset the scope to only a few major airports: can we compare the operations of certain hub airports (perhaps the busiest ones) and determine which ones have smoothest operations and why?
* For my own personal interest, I’d like to explore some other questions for my travel. When’s the best time to fly around Christmas or Thanksgiving – in terms of best on-time performance.

1. transtats.bts.gov [↑](#footnote-ref-1)
2. Tables of summary statistics are available at <http://www.rita.dot.gov/bts/airfares> along with more detailed tables: http://www.dot.gov/policy/aviation-policy/domestic-airline-fares-consumer-report [↑](#footnote-ref-2)
3. RITA’s website offers the ability to pull queries from their online database: http://www.transtats.bts.gov/DL\_SelectFields.asp?Table\_ID=236&DB\_Short\_Name=On-Time [↑](#footnote-ref-3)