

## DE300 Homework 1 - Written Component by Ryan Beam

### 1. Investigate Missing Data pt 1

- a. The Carrier column had an issue where NA for North American Airlines registered as nan with pandas. This was resolved.
- b. The Carrier Name column had some missing values for a couple of regional US Airlines. This was resolved by looking up the IATA codes and matching with non-missing values in the data.
- c. The Manufacture Year column had a few missing values that were resolved by looking up aircraft tail numbers and finding the correct data.
- d. The Number of Seats column had null data for several cargo planes, which was fixed by setting them to zero.
- e. The Capacity column had missing data for several aircraft. This was fixed by matching the aircraft with missing data to the instances of those aircraft in a similar timespan, and then taking the average capacity for that aircraft variant.
- f. The Airline ID column was resolved by using other instances of the carrier matching null data.

### 2. Missing Data pt. 2

- a. Manufacturer was not null.
  - b. For Model, each missing model was found by matching aircraft types and manufacturers.
  - c. Aircraft Type was resolved in the same way, by matching Models with null rows.
3. The remaining null data was then dropped, resulting in a loss of around 21,000 rows.
  4. Using a boxcox transformation, the Number of Seats and Capacity columns were normalized. However, there was not much observable change from this transformation, which indicates the columns were already close to normalized. This is somewhat surprising considering the spikes present in their distributions.
  5. Aside from Medium aircraft, the vast majority of aircraft are still in operation according to the data. This is likely due to the mass retirement of Boeing 757s and Boeing 767s by American and Delta during the pandemic

to save costs. However, this move has set both carriers back massively in competition with United for international travel, and has helped United become one of the best performing stocks on the market in the past 12 months.