1. *Investigate the missing data in this dataset. Specifically, for each of the following variables that have missing data, decide if any imputation is possible. Give your reasoning and code if you decide to impute missing values.*
    a. I imputed CARRIER using UNIQUE_CARRIER, since they both represented the same thing. If CARRIER is missing, it makes sense to just fill it in with UNIQUE_CARRIER.
        i. I used this same logic for CARRIER_NAME/UNIQUE_CARRIER_NAME
    b. I imputed NUMBER_OF_SEATS using the median seat count per MODEL
        i. Aircrafts of the same model typically have the same (or at least similar) number of seats, so I felt that using the groupwise median here would be a way to logically and consistently avoid bias from outliers.
            1. I used this same logic for CAPACITY_IN_POUNDS
    c. I chose not to impute AIRLINE_ID, as this is a unique identifier and imputing with an incorrect value could misrepresent an airline.
2. *Inspect the columns MANUFACTURER, MODEL, AIRCRAFT_STATUS, and OPERATING_STATUS. Decide, for each column, if transformation or standardization of data are required. Give your reasoning and code if you decide to transform the data.*
    a. For each of the columns, I standardized the values by converting them to uppercase and removing all whitespace. This ensured consistency across the variables, preventing formatting inconsistencies from affecting any data manipulation or analysis.
3. *Remove data rows that still have missing values. Report the amount of remaining data you obtained.*
    a. Amount of remaining data: 101275
4. *Transformation and derivative variables - describe what you observe before and after transformation.*
    a. NUMBER_OF_SEATS:
        i. Before Box-Cox: right-skewed, with a high concentration of aircrafts with 50-200 seats
        ii. After Box-Cox: more symmetric/bell-shaped distribution, suggesting that the transformation normalized the variable
    b. CAPACITY_IN_POUNDS:
        i. Before Box-Cox: very right-skewed, with a high concentration of aircrafts under 100,000 pounds
        ii. After Box-Cox: more centered distribution, demonstrating that the transformation normalized the variable and reduced skewness
5. *Feature engineering - Provide a written summary of your findings.*
    a. Operating Status by Size:
        i. The vast majority of aircrafts are actively operating for each size group
        ii. Small and medium-sized aircrafts have slightly higher proportions of non-operating aircrafts compared to larger aircrafts
    b. Aircraft Status by Size:
        i. Smaller aircrafts have higher proportions of status B
        ii. Larger aircrafts have higher proportions of status O