Task 1:
CARRIER and CARRIER_NAME can partially be imputed by relating them to one another. The only 'missing' carrier is NA for North American Airlines so we can manually impute that. For carrier names, the ones that are missing have carrier columns of L4 and OH. From the dataset OH has been used for two carrier names so we cannot assume either. On the other hand, L4 is only used for one carrier name so we can insert Lynx Aviation for the missing L4 carrier names. Similarly we can assume the AIRLINE_ID for the missing L4 rows is 21217.

I do not believe we can impute the few missing MANUFACTURE_YEAR entries because that can vary a lot even with the data we have in the other columns. For CAPACITY_IN_POUNDS and NUMBER_OF_SEATS, we should be able to use predictive imputation based on the model of the plane which determines the seat and weight capacities.

Task 2:
For manufacturing, I will transform the inputs to be more consistent since some manufacturers are entered under differnet names (ex. BOEING vs boeing vs THEBOEINGCOMPANY)
For aircraft/operating status and model, I will simply remove inconsistent punctuation and make them uppercase

Task 3:
N/A

Task 4:
After the treansformation, the values are much closer to a normal distribution but number of seats has a large amount of 0 entries which results in a spike at 0 but the nonzero entries are properly transformed to be more normal.

Task 5:
Based on my graphs, the SIZE of the aircrafts does not affect the OPERATING_STATUS because every size seems to have about 97% operating.

On the othe hand, the AIRCRAFT_STATUS distribution does seem to change with the SIZE of the aircraft. It appears that larger aircrafts are slightly more active in general based on how the proportion increases as you go from small to large and only slightly dips for xlarge. It also appears that the middle 50% (Medium and Large) are not as commonly out of service based on how Small and XLarge are out about 70% of the time. We can alos see that Medium aircraft have a Higher proportion in status B while the others are fairly even. Finally, all sizes have a very low proportion in L but Small planes appear to have none.