



Fig. 1. Above are categorical variables in the original dataset, and then after data preprocessing (Cleaned Data) and the dropping of all null values for the LoFtool target variable (After Dropping Nulls). Of course, the skew of the data after dropping nulls and encoding values of categorical variables is going to be the same as encoding does nothing to the overall distribution of a variable. If a variable is not included here, it is because it contained too many categories for a meaningful plot. As can be seen, distributions remained similar after every step of data processing.