Transformations:

board\_type & purchase\_type: change to categorial numbering

handle missing values - !!maybe we should leave it as missing!!

price & lead\_time: handle missing values

lead time is a continious number so we'll fill the missing values with generated numbers from its distribtuion. We look at the histogram and it seems like a exp distribution

we will do the same the the 'price' attribute in 2 different way

option 1

-> we can see that we maintained the distribution

option 2 - we will use the correlation (0.47) between price and room\_type

In class 1 and 4 there are too many missing values hence we will handle them differently.

-> we can see that price\_2 preserved the correlation between the features in a better way then price\_1

deleting the price outliers

**discretezaition**

Normalization