* Do I need to get rid of common words
* What is problem with distribution
* None of categorical columns are suitable to numerical because they either have 1 category or too many category