Midterm Assignment

Your goal for this assignment is to predict the prices of Magic The Gathering cards. Whereas in previous assignments you had features and needed predictions, here you are not provided features directly. For this assignment I require the following.

- Note: There are many "missing values." Many of these are not missing but rather not applicable. For instance, many spell/land cards do not have a power and toughness. So sometimes a data cell being empty dose not convey missing information and should be handled accordingly.
- NULL Assignment: Since this data requires a good amount of feature manipulation, your first goal is to get on the leaderboard and beat the null benchmark which just guesses the average. To beat this null benchmark, use linear regression (not necessary but for simplicity). Focus on getting the feature manipulation done and later worry about the modeling. Using lm is relatively straight forward as compared to lasso and ridge, so do this first and latter worry about glmnet. That is, the goal of this part of the assignment is to tackle 1 problem at a time.
- You don't have a train a test set directly, so you must generate them with the given start files.
- The "types" feature is an important feature which may take more than one value. The way it is given is not model ready, so it must be engineered in a way such that each row can convey if and what types a card is.
- **Create at least 3 novel features.** While it needs to be novel, it need not be complicated. For instance, the square of the converted mana cost may be of interest. That is, instead of cmc you could use cmc^2. Now that I have used this as an example it cannot be used as a novel feature (or any cmc^z).
- For modeling I wish for you to use both LASSO and ridge regression. Have a script for each and include both in your run.R file with comments on which is which. Note that the difference between these two files is changing 2ish lines of code, so focus on one of the two then copy paste.