

# Data Wrangling Completed

- Converting all words to lowercase
- Removing punctuation
- Counting the number of times a word appears and adding it as a feature to each post
- Removing rare words
- Adding an indicator to clarify the difference between “subreddit” as the feature and “subreddit” in the post
- Joining the training and test data
- Adding the word columns to the original data

# Top Words By Category

