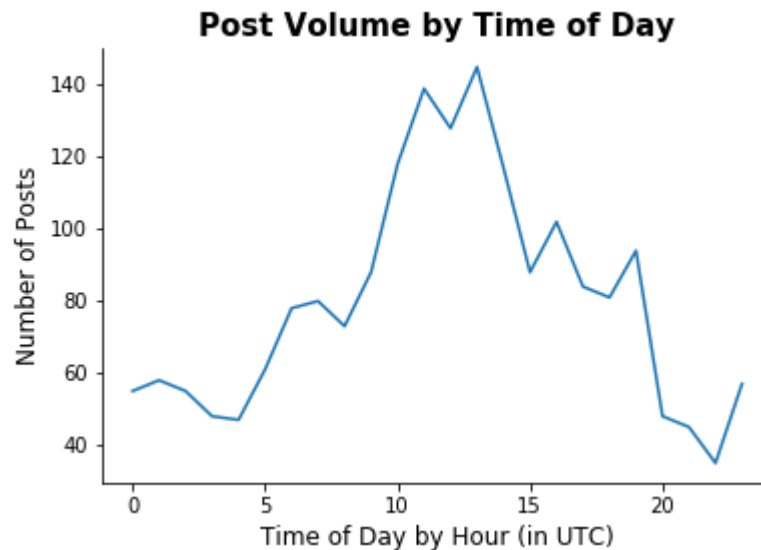


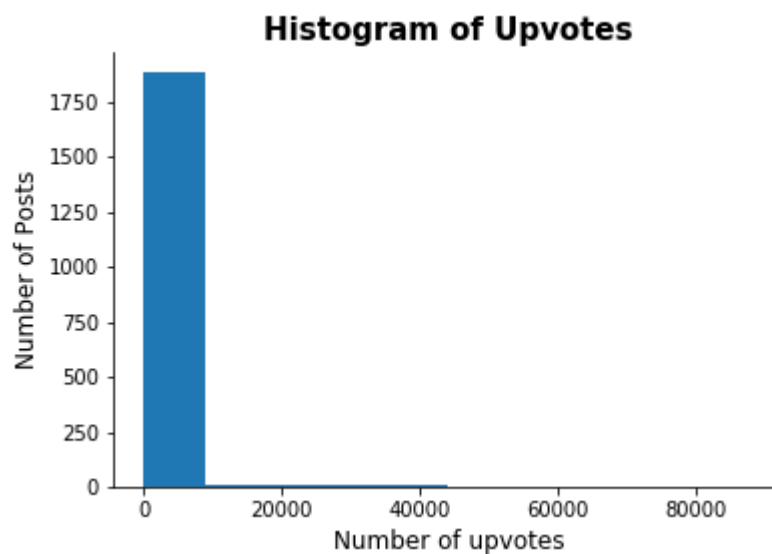
Reddit Post Popularity Predictors

A key factor that could predict the popularity of a post on Reddit is the time of day that it was posted. Most posts are made between 10:00 and 15:00 UTC, showing that this is a viable feature variable.



Another component of time that could be a factor is the age of the post, as newer posts might not have reached their upvote potential yet. Posts that are too new will be completely excluded for this reason.

In terms of the actual upvotes, a tiny amount of posts gain a disproportionately large amount of upvotes.



Once those values are logged, a skewed distribution is formed, showing that a log transformation will have to be performed on the upvotes, which is the target variable.

