Basic Plan (Will fill out these details for the assignment due Tuesday)

Feature “creation”/ manipulation (one hot encoding etc)

Pipeline?

Determining best mix of train/dev/test splits

* Aspect of request is to only use training data for periods earlier in the year compared to what is being predicted
  + IE Use January and Feb data for March.

Try out various methods/techniques

* Standard regression
* Trees
* Bayes etc

Split data into casual and professional writers