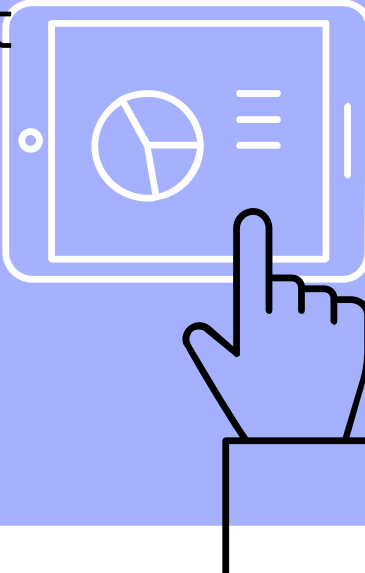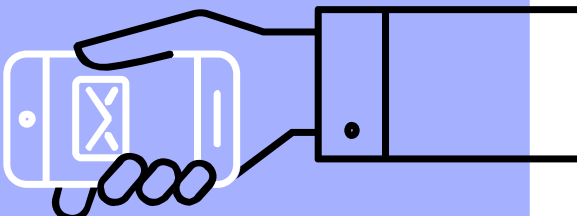# ENGAGE!

A predictive tool for
maximizing reader engagement
on Reddit

Austin Lasseter
February 23, 2018

ENGAGE!

What fires readers up?

Reader engagement on Reddit:

- Gets the message out
- Persuades public opinion
- Can lead to profit

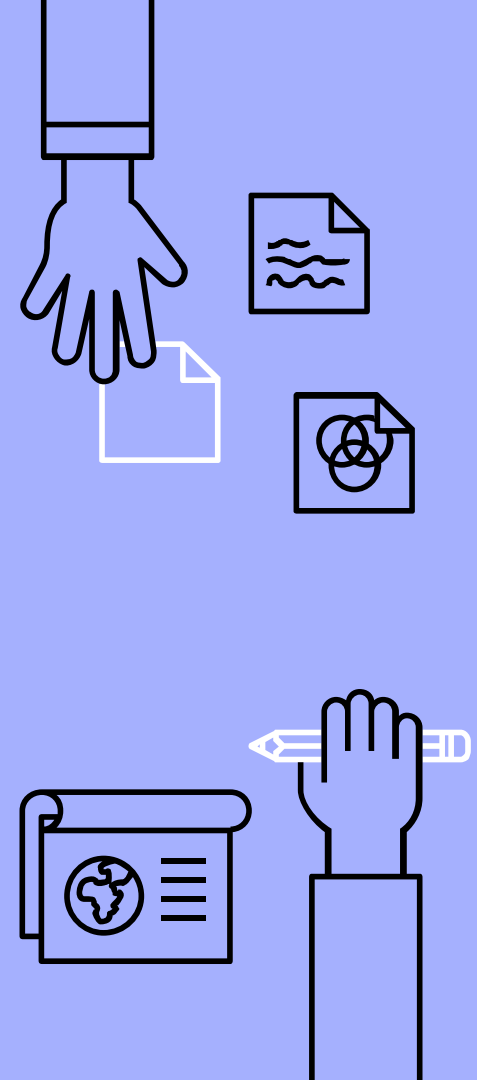# OUR TOOL CAN PREDICT READER ENGAGEMENT FOR YOUR POST

DATA SOURCE

- Hosted scraper on AWS EC2
- 1,000 posts every 6 hours
- Ran for 5 days on 'hot' reddit
- Collected 37,000 posts
- Reduced to 5,300 unique

DATA COMPONENTS

- Subreddit
- Length of Time
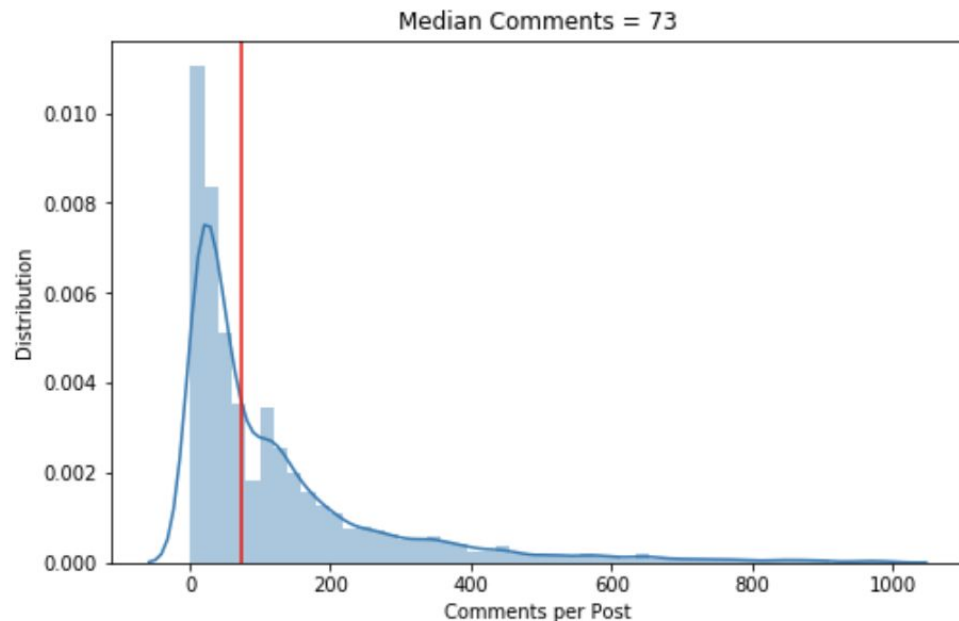- Number of Comments
- Title of Post

34.3k

Billboard removal in Toronto reveals original Spaceballs advertisement
submitted 8 hours ago by CrispyMiner to r/mildlyinteresting
416 comments share save hide report

More info about this analysis is available at:
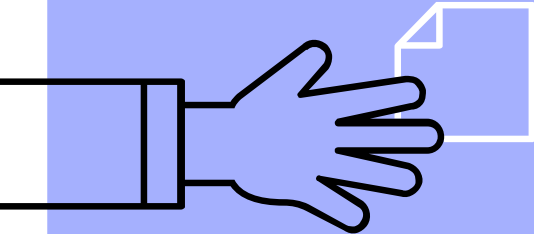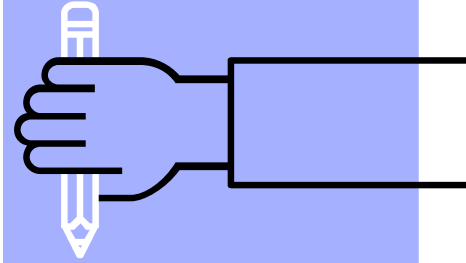https://git.generalassemb.ly/austinlasseter/project-3

73 Comments

That's all it takes to hit the top half of reader-engaged posts.

Median Comments = 73

Distribution

Comments per Post

What's the average length of popular Reddit posts?

Most data posts are about 50 characters long

Popular posts tend to be a little longer
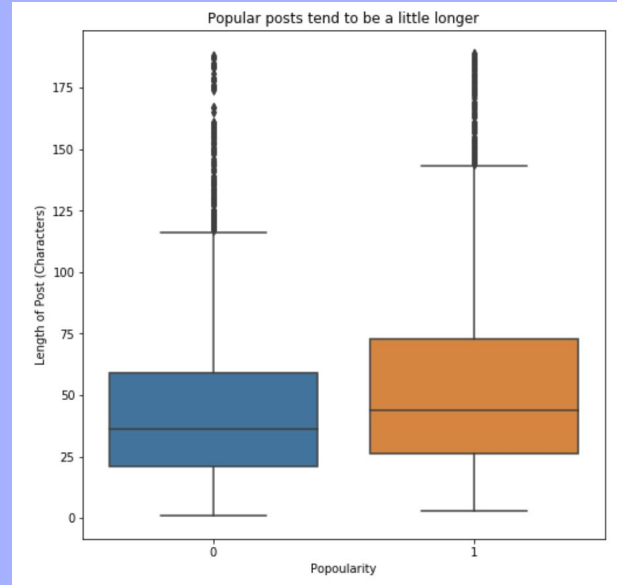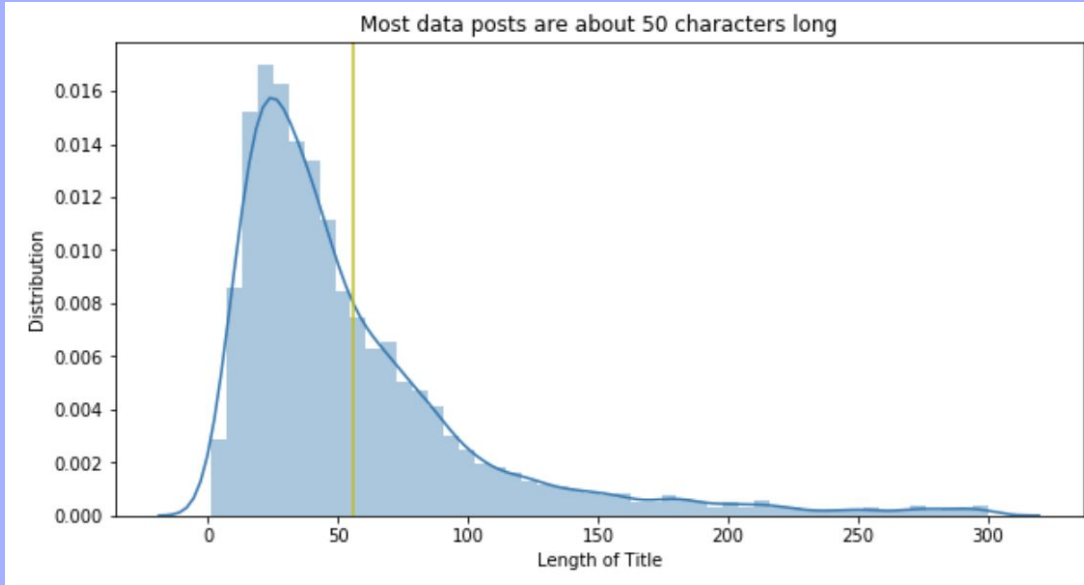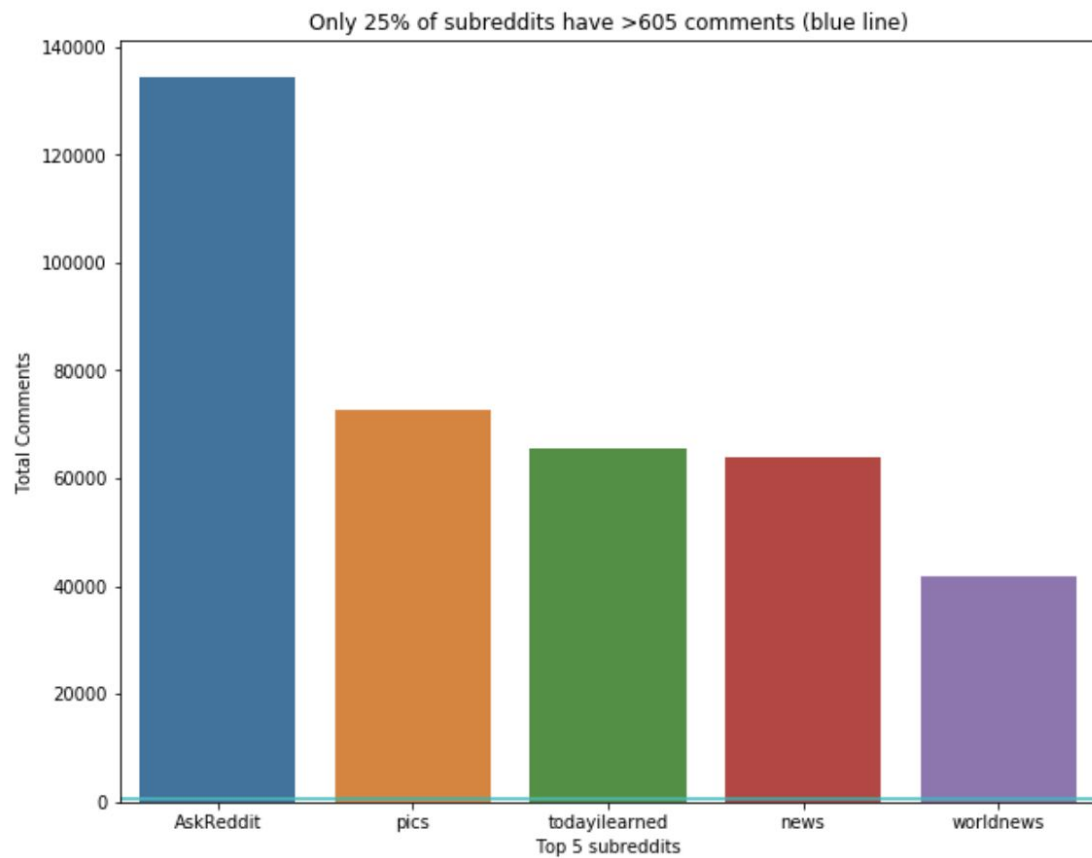
It takes some time to achieve reader engagement

Most posts have been up for about 9 hours

Popular posts have been longer than 9 hours

Only 25% of subreddits have >605 comments (blue line)
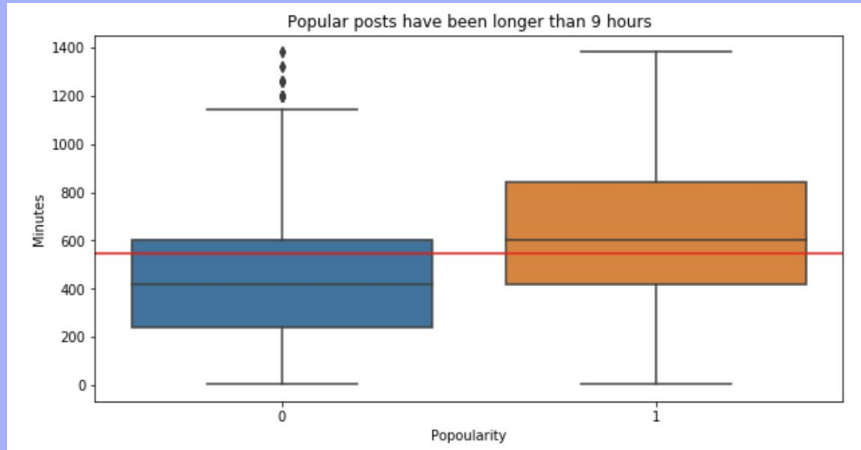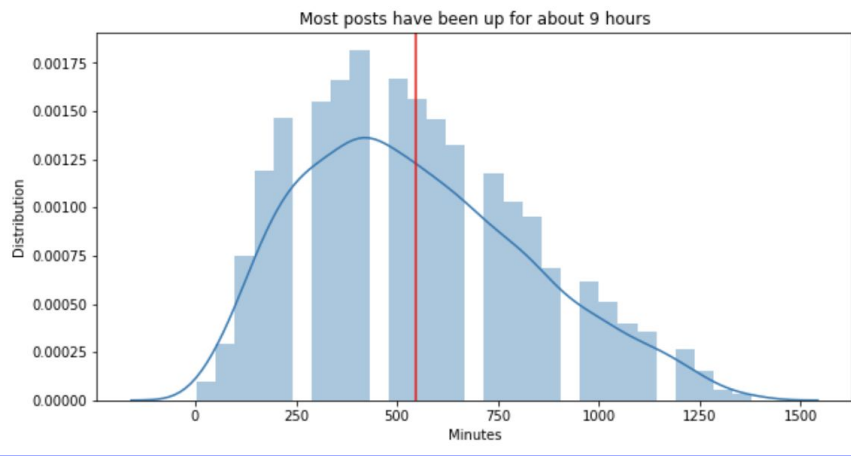
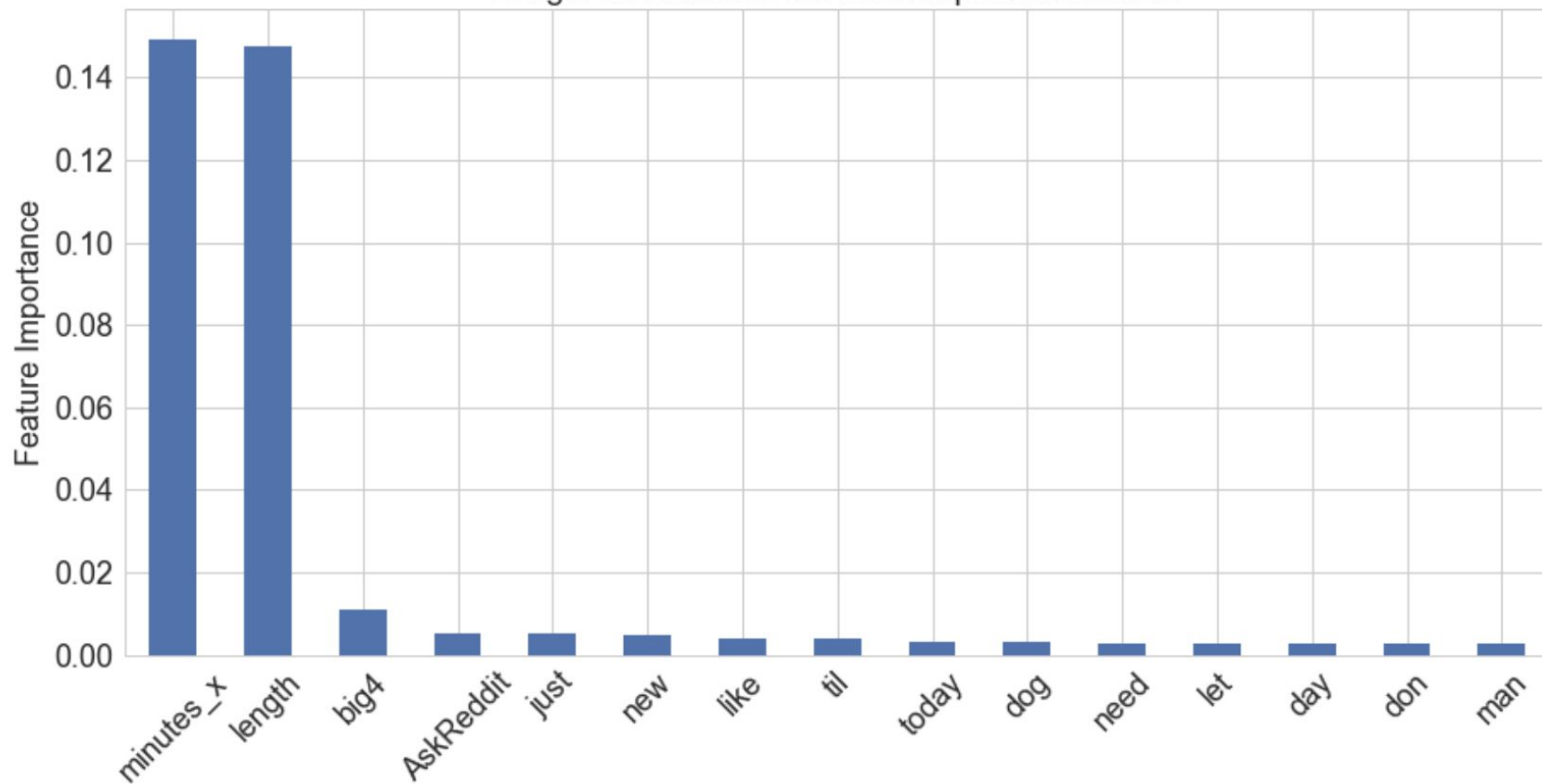25% of subreddits account for almost 95% of all comments!

"

But what should it say?
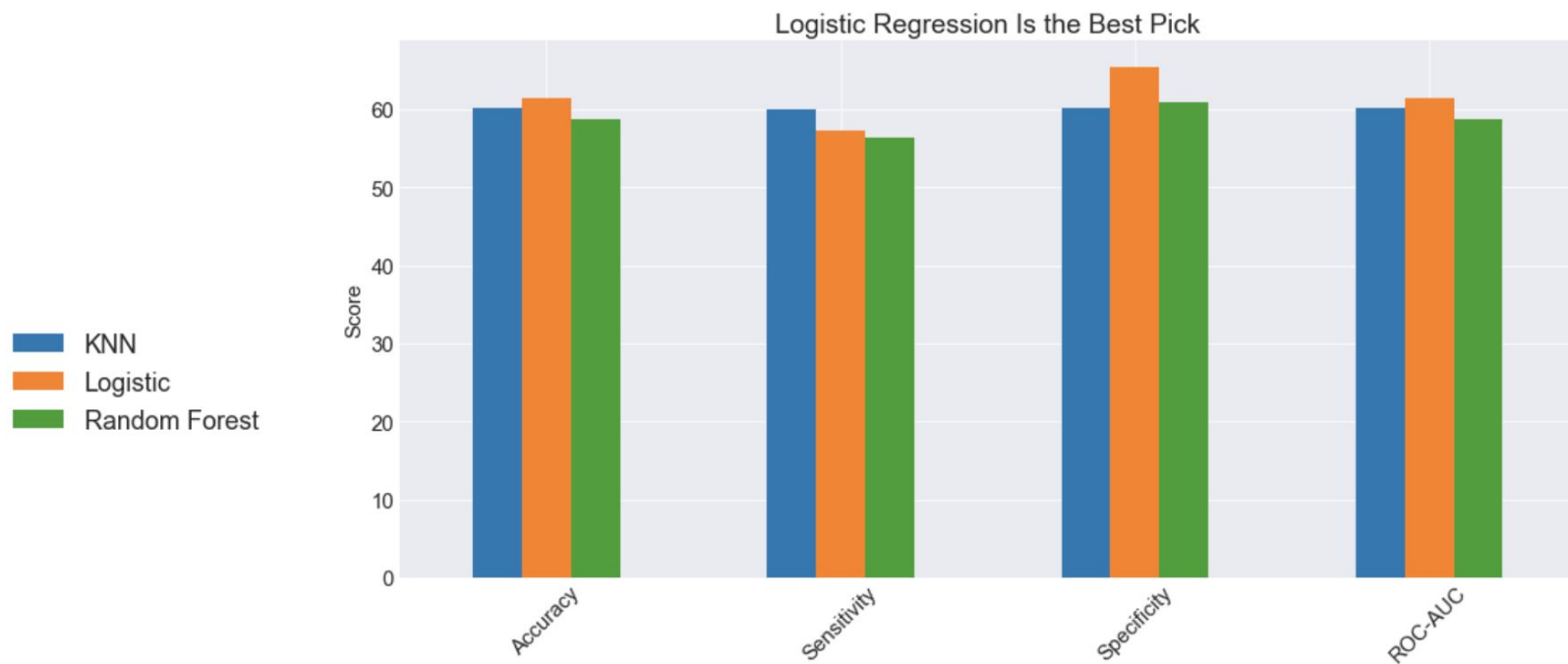

Almost all words < 10 times.
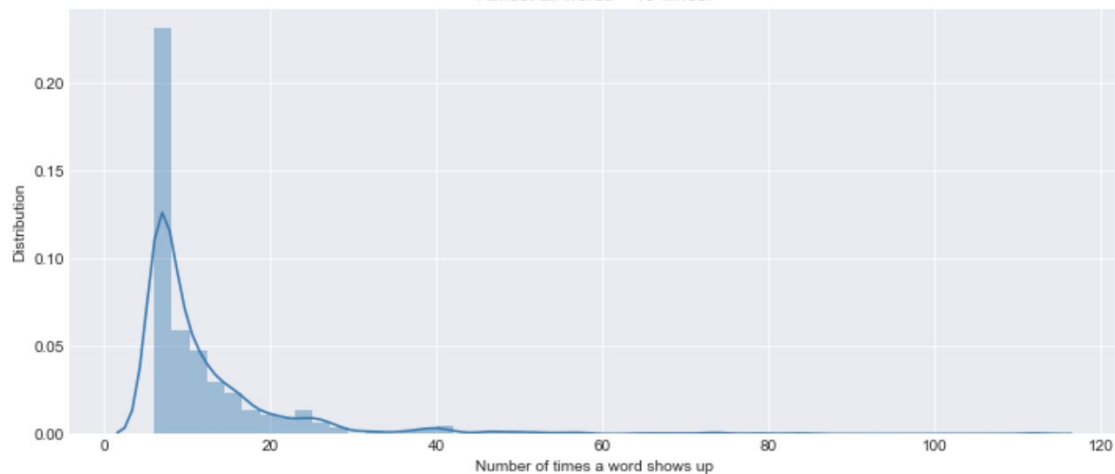
We call "vectorization" the process of turning text into numbers so that we can predict an outcome

Length and Minutes are most important features

Using these features, we tried
PREDICTING ENGAGEMENT 3 DIFFERENT WAYS

WHY YOU SHOULD INVEST IN OUR PREDICTIVE TOOL

▷ Our tool outperforms competitors:
  ■ Accuracy (61.4%)
  ■ ROC/AUC Score (61.4%)
▷ For each post, the tool produces a probability score based on:
  ■ length, subreddit, timing
  ■ content of your post
▷ Can be enhanced with funding