Posts with the most (comments)

In the “climatechange” subreddit

Lara Strachan

What gets a post commented on?

topic | title | main text | link to external website

Why does it matter?

Getting as much engagement as possible is critical for involvement and education around climate change.

How and What

Reddit Scraping

10k+ submissions collected from the r/climatechange subreddit

Process and Model

Scraped data was cleaned, explored, engineered, and run through multiple models

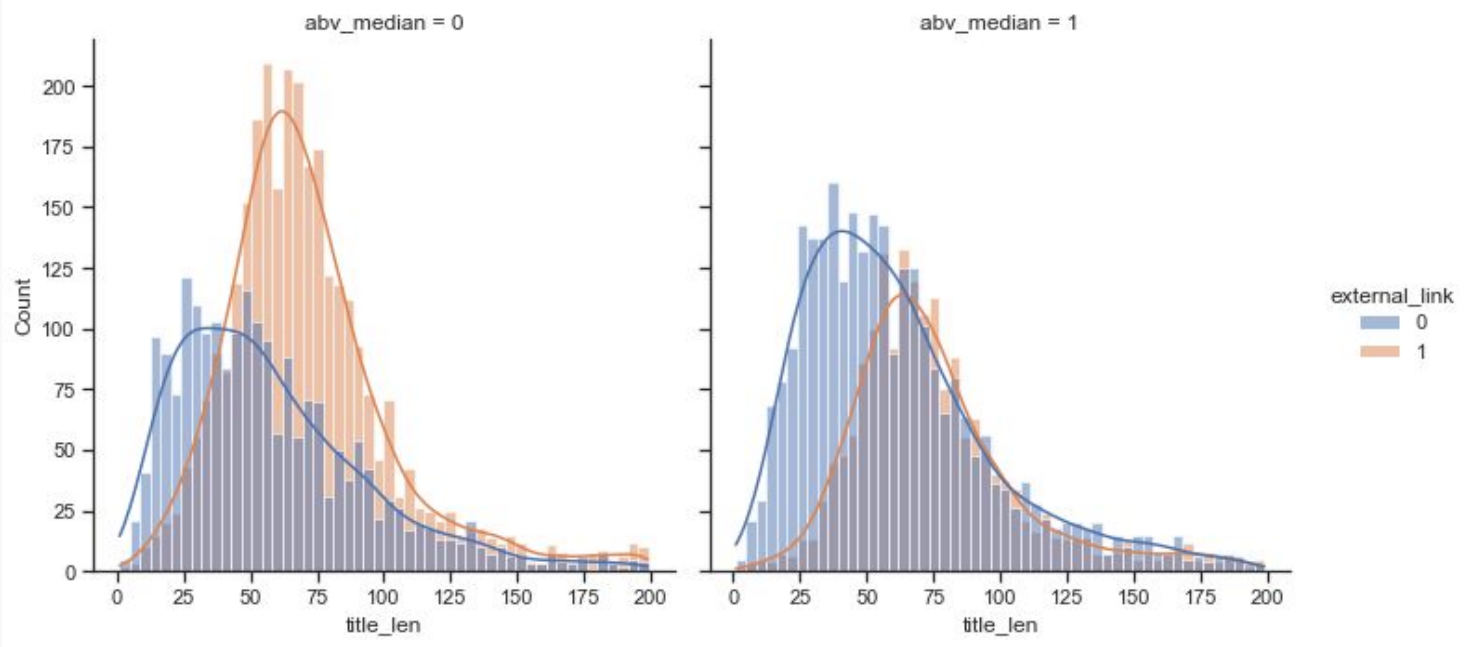
www.wordcloud.com

Initial trends

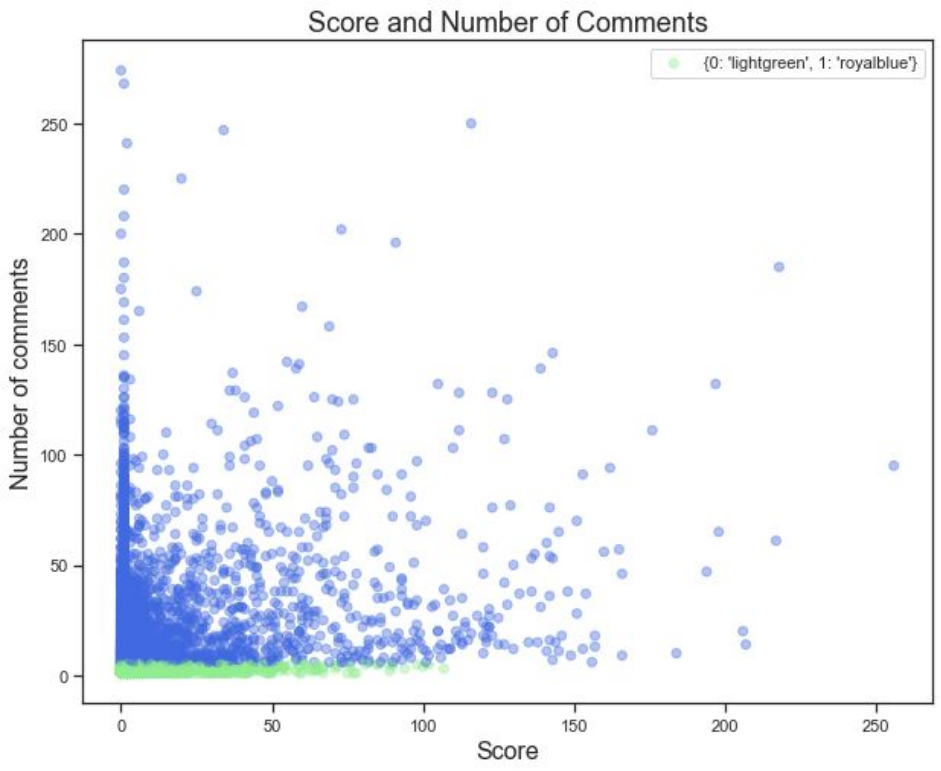
* Words taken from 25 post titles with the most comments

Feature exploration

Title Length



This visualization suggests that posts that are NOT just external links are more frequently receiving more than the median of 5 comments.



The majority of posts have both a score and total number of comments less than 50.

Baseline

52.5%

Posts with 5 or fewer comments

Title Text

Normalize

Remove non-alphanumeric characters

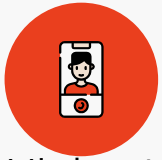
Tokenize

Separate words

Stem

Isolate word root

Model



Highest accuracy score was 59%

Words with largest feature importance from RandomForest:
'what', 'co2', 'trump', 'question', 'are we', 'skeptic'

Numeric Features

- Title length
- Body length
- Score
- External link y/n

	score	title_len	body_len	external_link
0	0.681998	0.098498	0.167752	-0.435614

Highest accuracy score was 63%

Best Model

Used both title text and numeric features.

- Count Vectorizer and TfidfTransformer to prepare text data
- Standard scaled the numeric data

Logistic Regression

Accuracy score of 66.1%

- (text only was 59% and numeric only was 63%)

Conclusions

Using text AND additional features gave highest accuracy

Better at predicting when a post would *not* get more than 5 comments

Ask away! Questions get answers

Next steps:

- Look at timing - day of week, time of year, around natural disaster
- Original poster engagement