

# Hot or Not: Creating Popular Reddit Posts

Caitlin Streamer

# As the 6th most visited website in the world, Reddit has the power to spread content to the masses

## Structure:

- Popular posts default homepage
- Subreddits for subject specific content

## Engagement:

- Posts
- Comments
- Up or down vote
- Share

The screenshot shows a Reddit post from the subreddit r/movies. The post title is "Box Office Week - Solo: A Star Wars Story debuts at #1 with a worrisome \$83.3M domestic on an estimated budget of \$250M-\$300M. Worldwide it's even worse as the film debuted to a disastrous \$65M international, less than what Deadpool 2 made internationally on its second weekend." The post is by user u/mi-16evil, posted 3 hours ago, and has 1 gold medal. The post has 14.9k upvotes and 5.0k comments. Below the post is a table with box office data for "Solo: A Star Wars Story".

Annotations in the image:

- Vote:** Points to the upvote/downvote arrows and the "14.9k" vote count.
- Subreddit:** Points to the "r/movies" link.
- Time posted:** Points to the "3 hours ago" text.
- Post title:** Points to the main title of the post.
- Comment:** Points to the "5.0k Comments" link.
- Share:** Points to the "Share" button.

Rank	Title	Domestic Gross (Weekend)	Worldwide Gross (Cume)	Week #	Percentage Change	Budget
1	Solo: A Star	\$83,325,000	\$148,325,000	1	N/A	\$250M



What makes a post popular?

# Using Reddit's open source API, 13k popular posts were scraped over a 6 day period

## Data Science Problem Statement

What characteristics of a Reddit post are most predictive of the overall interaction on a thread (as measured by number of comments)?

### Features Selected

- Post title
- Subreddit
- Post age (hours)
- Number of comments

### Target Classification

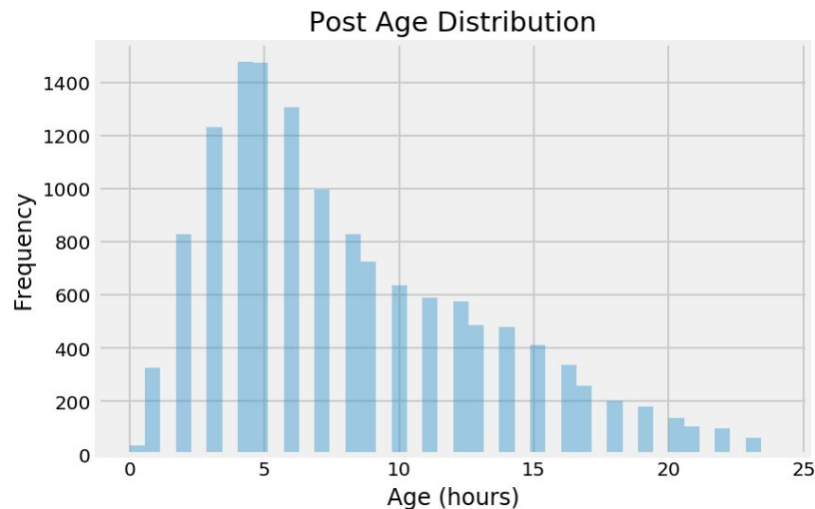
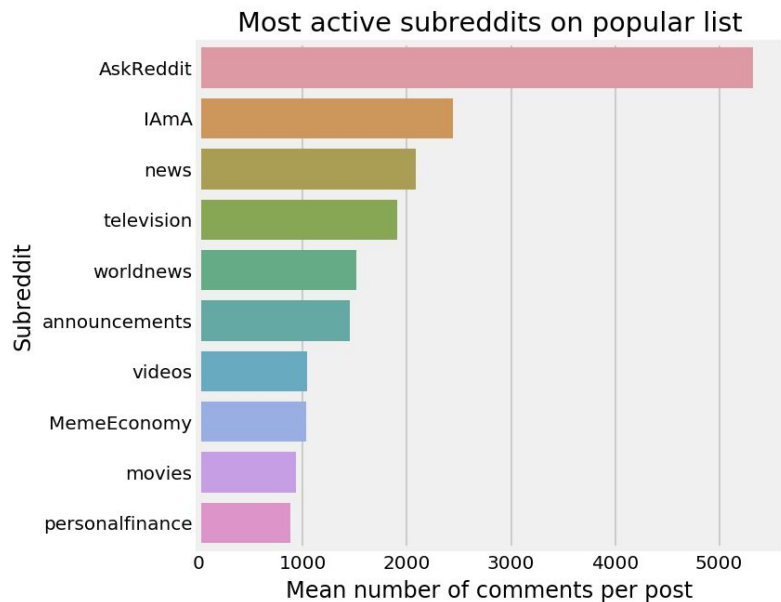
- High number of posts: above median
- Low number of posts: below median

### Machine Learning Model

- Identify which features are indicative of a high number of comments based on scraped posts
- Will lead to recommendations for creating a new highly commented post



# Data analysis revealed highly active subreddits and a shelf life for time spent on the popular list



# Natural language processing is required to convert text features into numerics for use in modeling

## Utilized Bag of Words NLP methods

- Frequency of words
- Common words are penalized
- Rare words have more influence



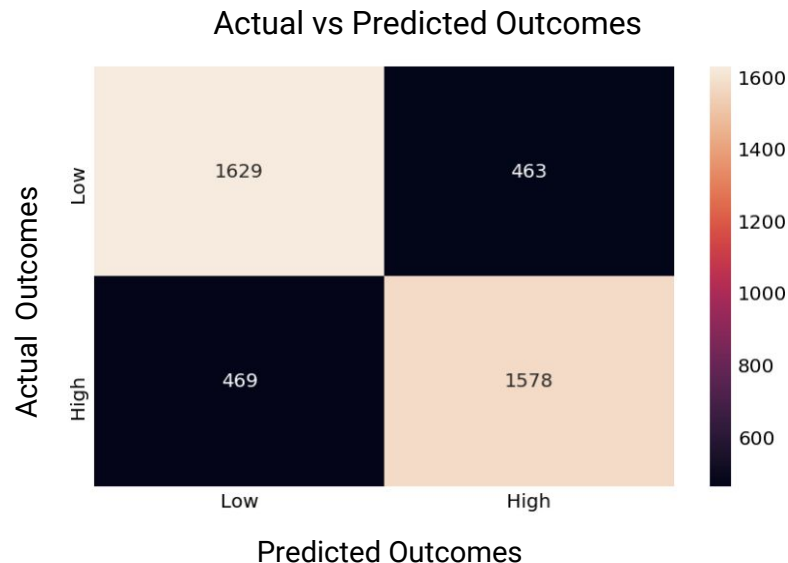
Reddit is a very popular website

1	0	1	1	0	1	1	0	1
very	dog	website	is	hot	Reddit	popular	not	a

# Final model has 77% accuracy rate using subreddit, title, and post age as predictors

## Key Takeaways

- Subreddit is the strongest indicator for how popular a post will be
- Some subreddits have more followers and thus a higher activity than others
- The odds of having a large number of comments increase the longer the post is on Reddit



# Post to funny, FortNiteBR, and gaming subreddits to increase odds of creating a highly commented post

## Top predictive post features

1. Funny subreddit
2. FortNiteBR subreddit
3. Gaming subreddit
4. Gifs subreddit
5. Pics subreddit

All subreddits, except FortNiteBR, are in the **top 20 most subscribed** to list

## Next steps

- Create a post based on recommendations to test outcome
- Look at additional interaction features
  - Voting
  - Shares
  - Comment content
  - Post date and time
- Change target classification thresholds