Analysis of the popularity of Reddit posts

By: Eva Tarr and Kyle Krawec

Summary

We will be analyzing the popularity of reddit posts through a variety of methods, mainly ones that come from graphing and correlation techniques. We plan to try to answer questions related to the popularity of a post:

* Does the sentiment of a comment result in greater popularity (does negativity gain popularity more than positivity)?
* Does the number of words in a comment result in greater popularity (does user attention span play a role in popularity)?
* Do more popular posts have more comments (do users engage in more popular things)?
* Does the time of creation of a post effect its popularity (posting at a certain time of day or posting after the initial “hype” of a topic)?

Hypothesis

The main social computing goal for our research is to better understand the factors that contribute to popularity on social sites. Do users tend to engage in more negative posts? We hypothesize that this is the case and that the negativity on social sites is more welcomed than positive ones. What makes users engage in a post, is it the number of words in a post? We hypothesis that posts with less words will have a greater popularity. Do users engage more with already well-liked posts? We hypothesis that user will engage in more popular posts because it would demonstrate social conformity. Time of creation, we hypothesis, would also play a large role, we believe that the data will show that based on the time of day a post would be more likely to have lots of popularity as opposed to others. This would show the general human habits of users and when they would be using these social sites. We also hypothesize that as a topic becomes “old” it would decrease in popularity.

Don’t bury the lede

Out of all the hypothesis we had about the popularity of Reddit posts we found that X was the biggest contributor and YZ were incorrect, and we found different findings contrary to what we hypothesised.

Technology, Techniques and Methods

The main language used for data analysis and graphing was Python. Python libraries pandas and matplotlib were used to achieve basic analysis and graph creation. Data used for this analysis came from the open-source data library SocialGrep. The main data was presented in csv format making it easily usable…..

Analysis

Popularity and Sentiment:

The data used for these tests were from SocialGrap and they are formatted in csv. The tests were performed twice, with different data sets. The overall conclusion is that there is no correlation with the sentiment attached to a comment and its popularity. This is contrary to our hypothesis. The correlation coefficient for both tests came out to be 0.

Trucker Bitcoin

A graph with red lines and numbers

Description automatically generated A graph of a graph with red lines

Description automatically generated with medium confidence

Popularity and Word Count:

The data used for these tests were from SocialGrep and they are formatted din csv. The tests were performed on 2 different data sets. We theorized that the correlation to word count would be that popularity would coincide with shorter comments. Alth

Doge Trucker

**A graph with red lines and a point

Description automatically generatedA graph with red lines

Description automatically generated**

Popularity and Creation Time (Month):

Doge No New Normal

A graph of a bar

Description automatically generated with medium confidence A graph of a bar graph

Description automatically generated with medium confidence

Popularity and Creation Time (Daily):

No New Normal Doge

**A graph of a graph

Description automatically generated with medium confidence** **A graph of a graph

Description automatically generated with medium confidence**

Trucker

**A graph of a graph

Description automatically generated with medium confidence**

Conclusion:

(Same and bury the lead but with more detail and reference the results)

Analysis

What went wrong and how to do better next time.