

General Assembly

Introduction to Data Science

Jonathan Franks

<https://github.com/RobotGhost>

Final Project Brief

Summary

This project will look at over 1.7 billion JSON objects that comprise Reddit's entire publicly available comment dataset. An example of one object from the "reddit-comments" dataset:

```
{ "gilded":0, "author_flair_text":"Male", "author_flair_css_class":"male", "retrieved_on":1425124228, "ups":3, "subreddit_id":"t5_2s30g", "edited":false, "controversiality":0, "parent_id":"t1_cnapn0k", "subreddit":"AskMen", "body":"I can't agree with passing the blame, but I'm glad to hear it's at least helping you with the anxiety. I went the other direction and started taking responsibility for everything. I had to realize that people make mistakes including myself and it's gonna be alright. I don't have to be shackled to my mistakes and I don't have to be afraid of making them.", "created_utc":"1420070668", "downs":0, "score":3, "author":"TheDukeofEtown", "archived":false, "distinguished":null, "id":"cnasd6x", "score_hidden":false, "name":"t1_cnasd6x", "link_id":"t3_2qyhmp" }
```

Preliminary Questions

1. What can be discerned about communities from the markers on communities?
 - Upvotes ("ups")
 - How many "positive" votes comments received.
 - Downvotes ("downs")
 - How many "negative" votes the comments received.
 - Karma ("score")
 - The amount of positive votes after subtracting the negative votes.
 - Contraversiality ("contraversiality")
 - How close the difference is between "positive" and "negative" comments.
 - Flair ("author_flair_text": True/False)
 - Whether or not the subreddit allows for Flair on the commenters username.
2. What are the marker averages for the most contraversial comments?
3. What are the marker averages for the most upvoted comments?

Primary Question

- What kinds of comments can be found in an active subreddit community?
 - Are they more or less controversial, positive, or negative?

Progress

I have collected the primary dataset and started exploring the data, but have received errors and will be looking for assistance in solving those issues.