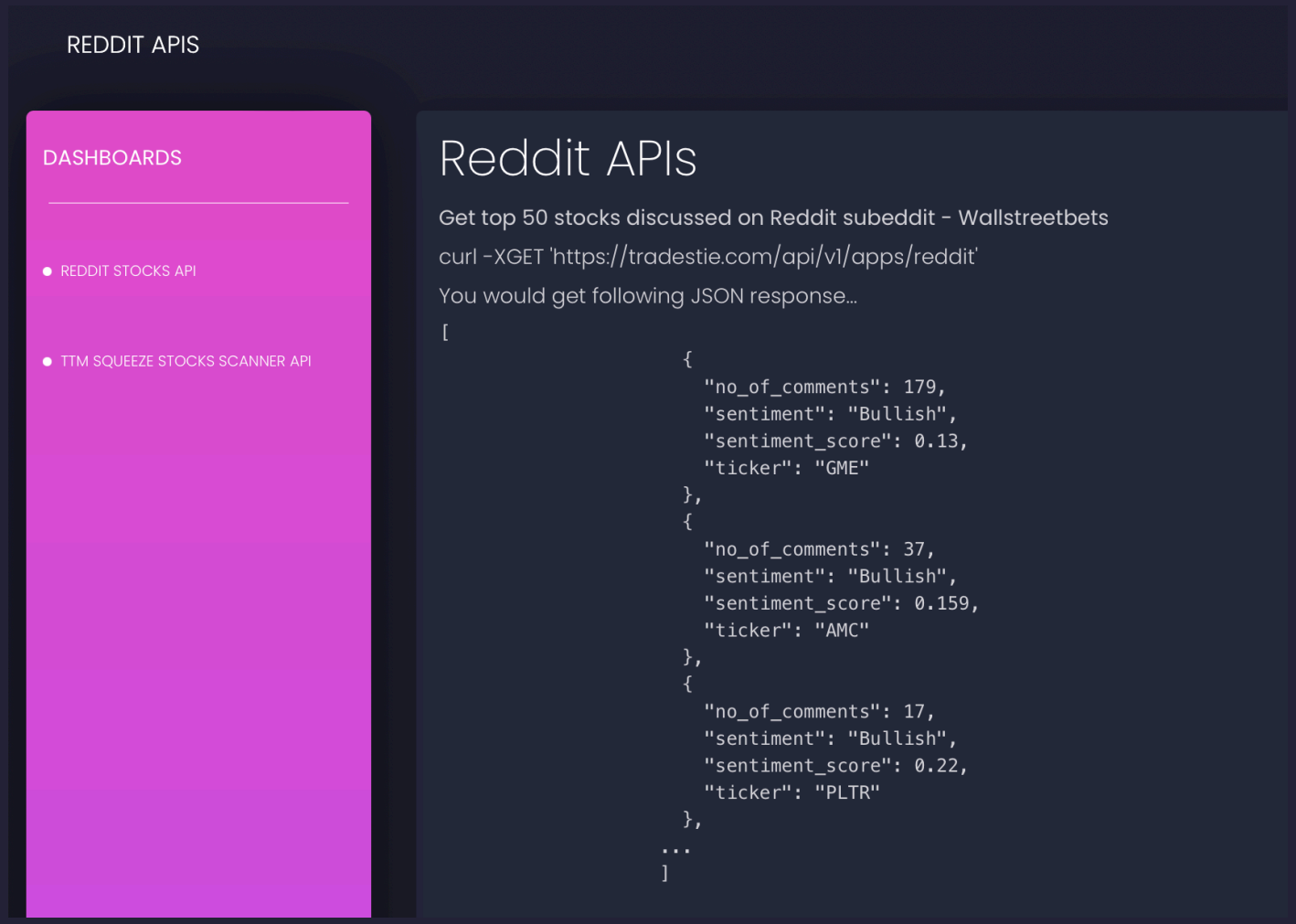
Assignment 3 Overview

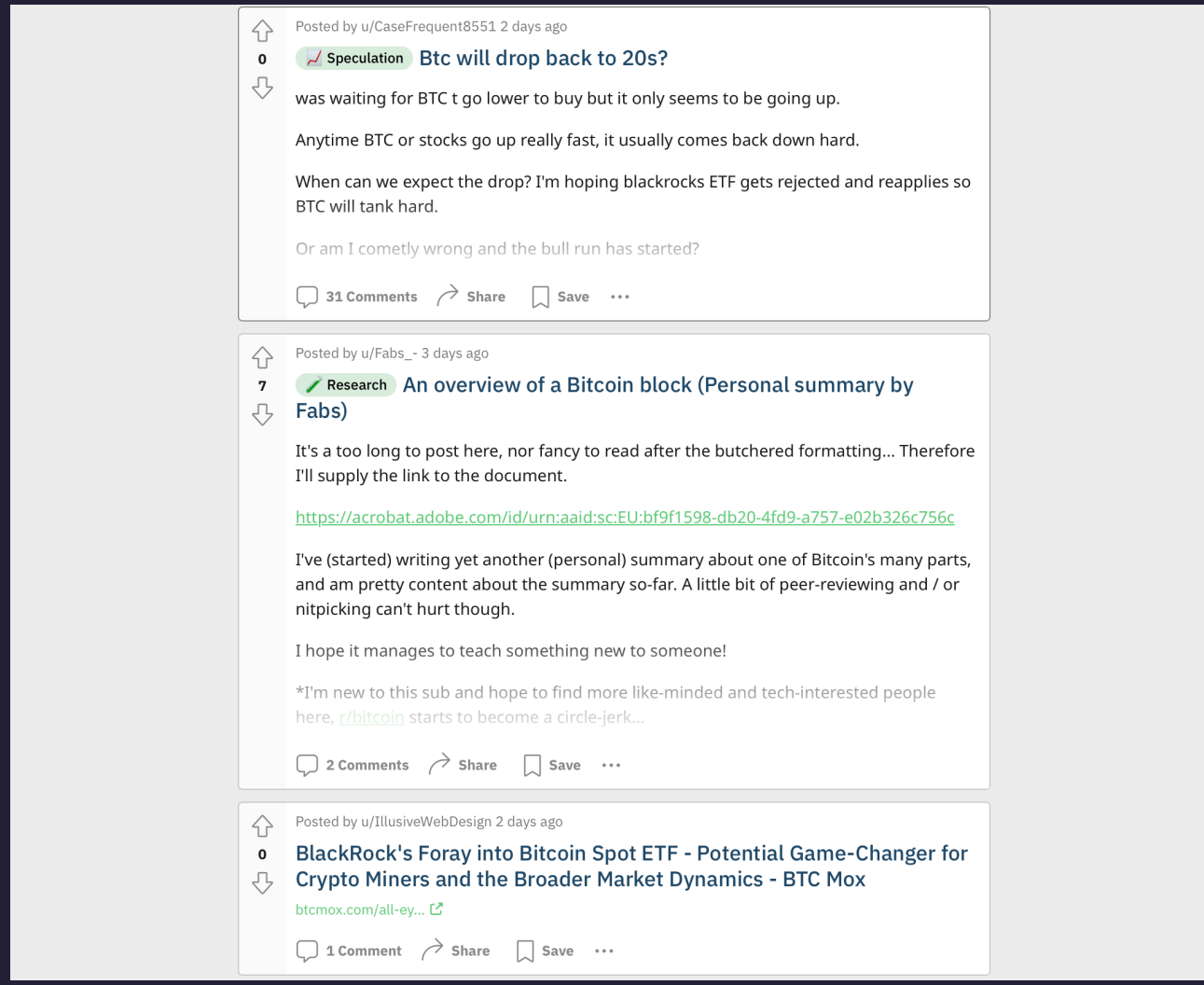
Social Media Analytics

Tradestie API



Reddit

- Online community for sharing content
- Links, text, and images
- Subreddits:
  communities for specific topics
  - `r/Politics`
  - `r/rbc`
  - `r/wallstreetbets`
  - `r/btc`



ASMT3 - Social Speculation: Harnessing Reddit to Forecast Bitcoin Fluctuations

1. Get Reddit posts from crypto-related subreddits
2. Use sentiment analysis to determine the polarity and subjectivity of the posts
3. Use them to predict the price of Bitcoin
4. Create your own sentiment index using information from the posts

Grading criteria

- **ETL functions (80%)**
  Functions must pass the testcases
- **Your sentiment index (20%)**
  Points for how well your sentiment index is correlated with the price of Bitcoin. Your rationale for why your sentiment index is better along with evidence (e.g., data, chart) must be included in the notebook.

Data Source

- Reddit API (https://www.reddit.com/dev/api/)
- Alpha Vantage API (https://www.alphavantage.co/documentation/)

Sentiment Analysis Package

- `TextBlob` for sentiment analysis (https://textblob.readthedocs.io/en/dev/)

💻 Setting Up Reddit API Authentication

1. **Go to https://old.reddit.com/prefs/apps**
    - Sign up for a Reddit account if you don't have one (Copy `username` and `password`)
2. **Click on "are you a developer? create an app..."**
3. **Fill in the form with random information**
    - name: random app name
    - type: select "script"
    - description: random description
    - about url, redirect url: random url (e.g., https://pyprogramming.net)
4. **Click on "create app"**
5. **Copy `client ID` and `client secret`**

💻 Reddit API Authentication

Four pieces of information are needed to generate authentication headers: `username`, `password`, `client ID`, `client secret`.

```
headers = get_authheaders(username, password, client_id, client_secret)
```

The authentication `headers` must be passed to the `requests.get()` when making a request to the Reddit API.

```
# headers from get_authheaders()
response = requests.get(url, headers=headers, params=params)
```

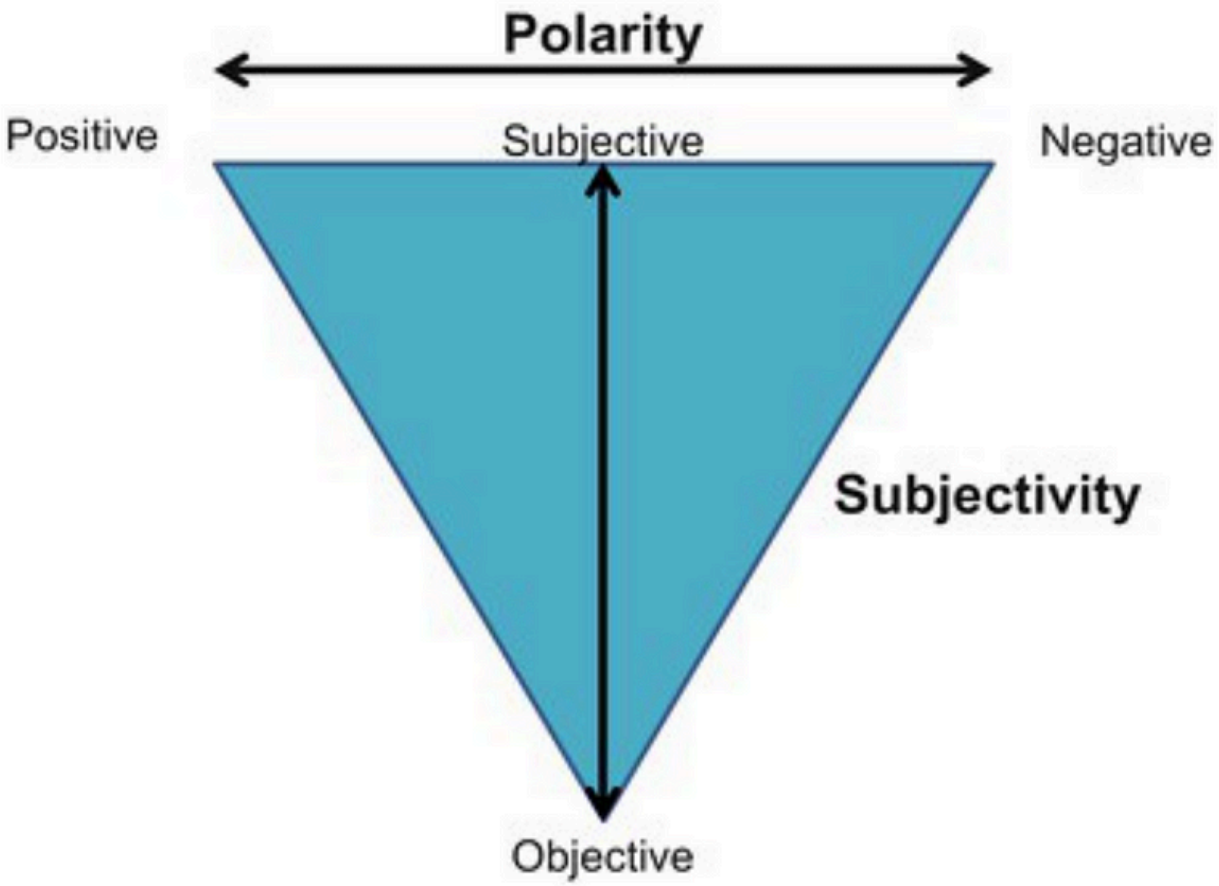
New post endpoint

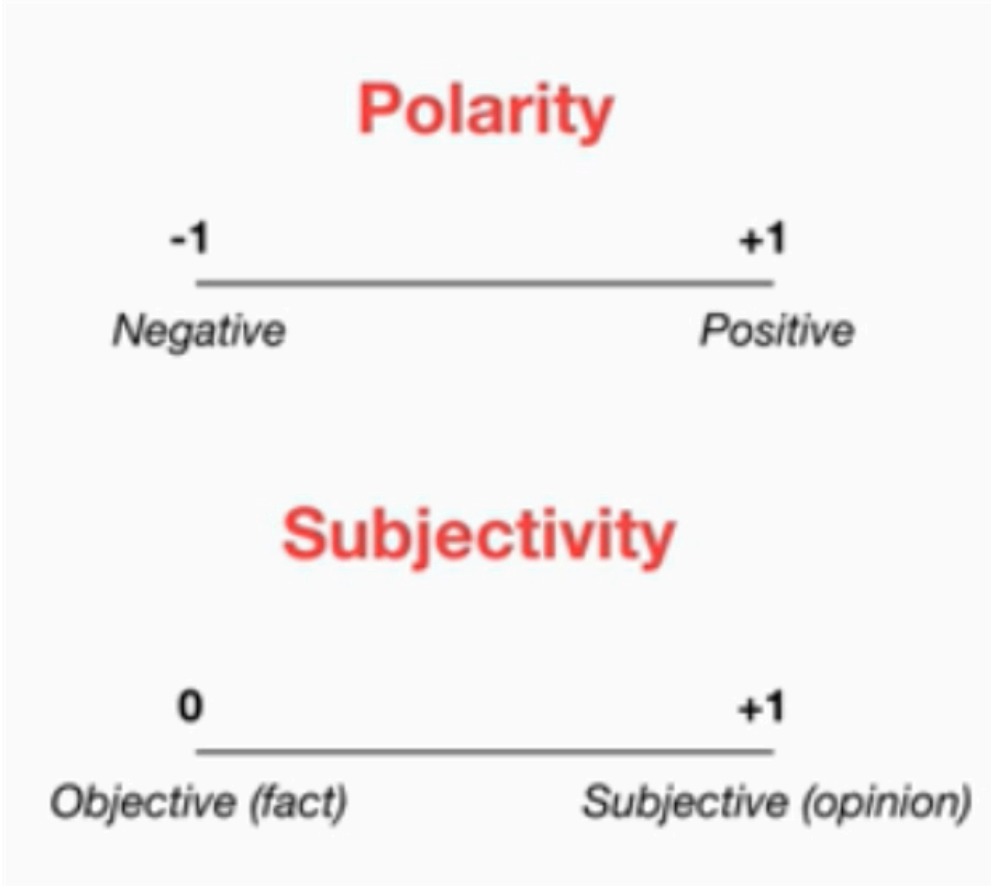
- Endpoint: `https://oauth.reddit.com`
- Path: `/r/{subreddit}/new`
- Query parameters: See here (https://www.reddit.com/dev/api/#GET_new)

Hint 1: Extracting posts through pagination

- Reddit API `limit`s the number of posts returned to 25 (default) or 100 (max)
- Pagination
  - `after` parameter: fullname of the last post in the current page
  - `before` parameter: fullname of the last post in the previous page

Hint 2: Sentiment analysis using `TextBlob` package





Hint 3: Better sentiment index

Use the polarity and/or subjectivity scores from TextBlob along with other Reddit data to create a sentiment index that is better correlated with the price of Bitcoin.

- assigning different weights to different posts based on the characteristics of the posts
- pulling data from a different subreddit
- assigning different weights to posts from different subreddits
- ...