Reddit Investment Data Sink Documentation

My project is a web scraper for reddit using the Reddit API through the Python library PRAW. I want to get data for my data sink from the main, largest subreddits for investing, and sort the posts into a database that could help with comparing trends in the stock market to trending topics in social platforms. The code creates a SQLite database with the 100 hottest posts from the top 100 subreddits that have invest in their title or description. One of the biggest ideas I had for data visualization is looking for buzzwords and trends on specific topics and catagorizing how popular they are discussed.

All data was obtained using the Reddit API https://www.reddit.com/dev/api/

The code requires a user's Reddit credentials to Access the Reddit API. It creates a SQLite database locally where the script is ran, and then creates a table with the columns title, score, id, subreddit, url, num_comments, body, and created.

Then, the script makes a request to the Reddit API to get the first hundred subreddits that match the word invest in their namer or description. The script then takes the hottest one hundred posts from each of these subreddits, iterating through them and placing them in the table created in the database before.

title	Title of subreddit post
score	Number of upvotes
id	ID of post
subreddit	Subreddit it was posted in

url	URL to post
Num_comments	Number of comments in post
body	Body text of post (first 30,000 characters)
created	Date post was created