

Case Study: Reddit Data Analysis using PRAW

By Sanika Dhanwate

What is PRAW?



Python Reddit API Wrapper

PRAW is a comprehensive Python package that provides simple, official access to Reddit's powerful API functionality.



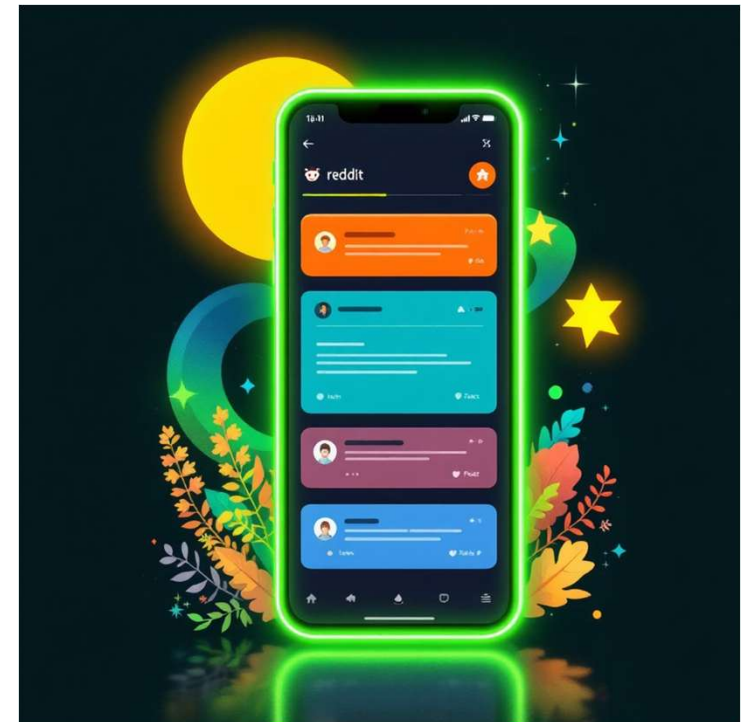
Compliant & Easy

Designed to be user-friendly whilst maintaining full compliance with Reddit's API rules and rate limiting.



Full Functionality

Enables both reading data from Reddit and performing actions like posting, commenting and moderating.





Goals of the case study!?



Understand concerns and discussions among data science practitioners and learners.

Explore posts that gained high engagement in year 2025

Subreddit Analyzed: **r/datascience**

Easier to collect structured data using PRAW.



Used extensively for academic research, automated moderation bots, content analysis and large-scale data collection projects.

Setting Up PRAW: Getting Reddit API Credentials

01

Create Reddit Account

Register a Reddit account if you haven't already. This will be used for API authentication.

02

Visit App Preferences

Navigate to <https://www.reddit.com/prefs/apps> to access your application settings.

03

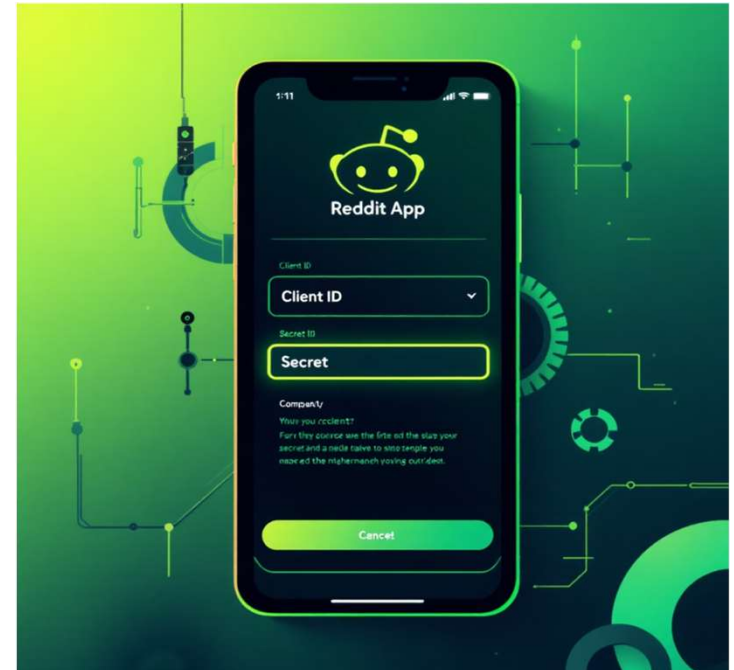
Create Script Application

Click "Create App" and select "script" as the application type for personal use projects.

04

Secure Your Credentials

Obtain your `client_id`, `client_secret` and set a descriptive `user_agent` string for authentication.



Methodology

Tool Used: Python Reddit API Wrapper (PRAW)

Steps & Filters Applied in this:

- 1 Collected top 100 posts from r/datascience (2025 only).
- 2 Filtered posts:
 - Score > 50 (high engagement)
 - Title \geq 5 words (informative posts)
- 3 Converted timestamps \rightarrow human-readable dates.
- 4 Collected: title, body, top 3 comments.
- 5 Saved all data into a CSV file for analysis.

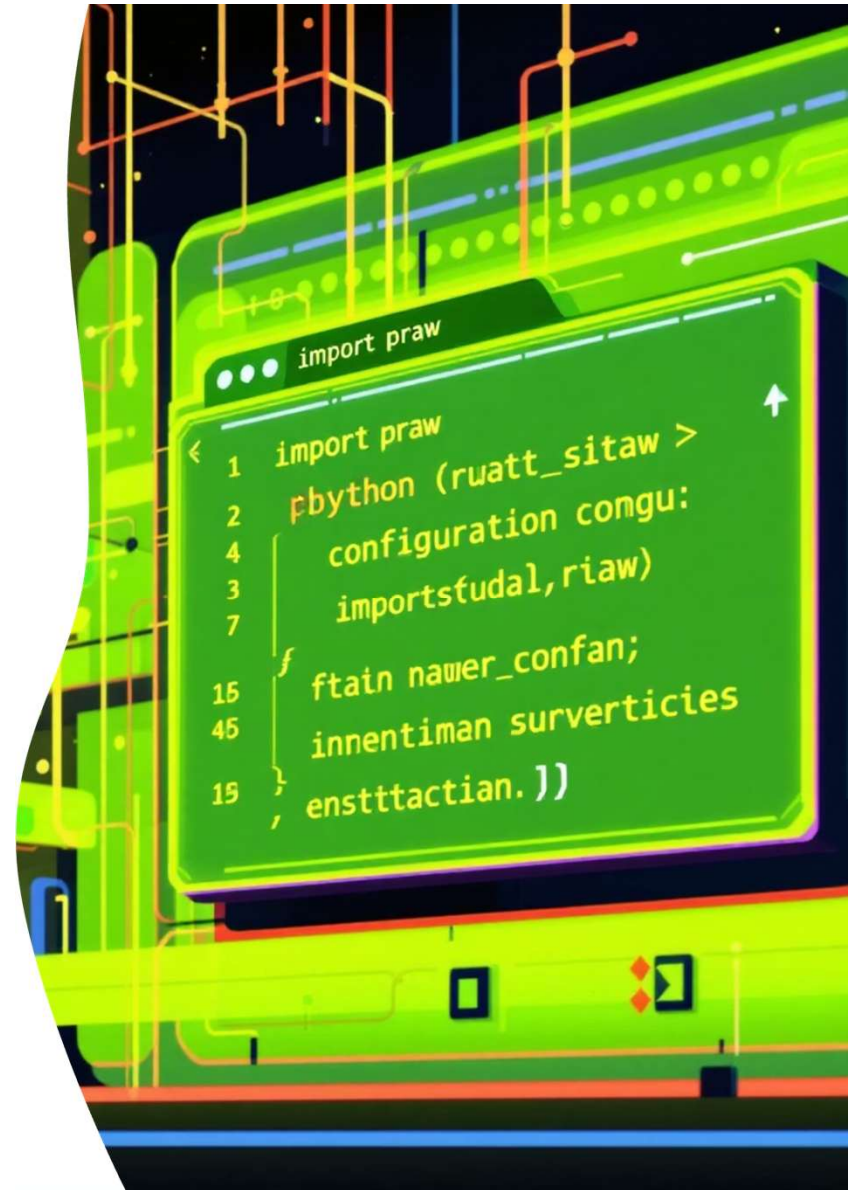
Dataset Summary:

- Total posts collected: 100
- Each post contains:


- Date of posting
- Score
- Title
- Body
- Top 3 comments

Saved as:

Sanika_Dhanwate_Casestudy_1.csv



Results Observed:



Most high engagement posts had descriptive & clear titles.

Posts often focused on career advice, technical challenges or best practices.

Example Post I Found Interesting:

How do data scientists add value to LLMs?

The discussion highlighted a few key points:

- **Domain knowledge:** Data Scientist have deep knowledge of certain industry which helps them choose the right datasets and make sure the data used in LLMs is high quality.
- **Data quality:** Data Scientist clean and check data to make sure models work well.
- **Evaluation:** Data Scientist create metrics to see how well models work making sure they achieve business goals and satisfy users.

Key Insights & Conclusion:

Clear and detailed posts attract more discussion.

Community focuses on practical guidance: career, technical tips and best practices.

Filtering by score and title length helps remove low value or meme style posts.