

## Data Scientist / Researcher – Python or R Experience

The National Whistleblower Center is currently involved in a project where the goal is to research the concerns of employees who have witnessed misconduct and to better understand potential barriers to reporting misconduct.

The project will use open source intelligence gathering, digital marketing, data mining and analysis.

### What We're Looking For

The team is seeking a data scientist with Python or R experience who can help us connect to these websites using their API, to mine *and* analyze conversation or comment data to develop further insights. This data scientist should also be able to illustrate how to use these data mining and analysis techniques to the team for ongoing data mining processes.

In specific the ideal candidate will be able to effectively engage in:

- **Reddit Subreddit Mining** – Use Python or R to connect with Reddit API (PRAW) and pull all comments in a specific subreddit containing a specific keyword into a CSV file for further manual and machine-learning analysis.
- **Twitter Comment Mining** – Use Python or R to connect with the Twitter API and pull comments containing specific keywords into a CSV file for further manual or machine-learning analysis. We'd also like to visualize our findings in various ways.
- **Glassdoor Review Mining** – Use Python or R to connect with Glassdoor API to pull comments, company ratings over time, job descriptions, and other review data so that we can correlate this with our research findings.
- **TheLayoff.com Scraping** – Scrape the website for comments containing keywords that we've identified to be indicators of fraud. Using tools such as Octaparse, Parsehub or any tool that can accurately locate key conversations.

If you are interested in this, please get in touch with Carly Fabian at the National Whistleblower Center for more details at [carly.fabian@whistleblowers.org](mailto:carly.fabian@whistleblowers.org).

Below are some example videos and URLs with similar tasks we're trying to accomplish:

<https://www.youtube.com/watch?v=kkMBMxwxTEg>

<https://medium.com/@xyng17/scraping-twitter-and-sentiment-analysis-using-python-c5a44b9288ab>

<https://www.storybench.org/how-to-scrape-reddit-with-python/>

<https://towardsdatascience.com/scraping-reddit-data-1c0af3040768>

## **Data Scientist / Researcher – Python or R Experience**

<https://towardsdatascience.com/scraping-reddit-with-praw-76efc1d1e1d9>

<https://youtu.be/wpCoHEieOxw>