

About the creative process

Approach

Based on the Data Source API Analyst Homework document, I decided to focus my efforts on extracting commits, putting special attention on handling pagination, error control, and rate limiting.

The queries for public repositories, organization repositories, and repository contents were developed with less detail to focus on the commit extraction feature. Still, they're set up with enough parameters to work as a tool for exploring repositories and their contents for anyone who wants to extract commits.

Motivation for features

To get familiar with how the API works, I read the docs to pull out useful info for testing. Here are some key points I found in the docs or while testing, and then implemented or kept in mind during development:

Authentication

- You can use the API without authentication, but the rate limits are lower and some endpoints are either restricted or limited.
- If you want to make requests without authentication, just don't add the "Authorization" header.

Rate Limits

- You can check your current rate limit at any time.
- There are different categories of responses or endpoints, each with their own rate limits. In practice, I noticed the endpoints used in the notebook fall under the "core" category.
- Using the rate limit endpoint, I saw that authenticated users get a limit of 5000 requests per hour, while unauthenticated users only get 60 per hour.

- To avoid hitting the rate limit, you can filter directly in the API using query parameters like “path”, “author”, “since”, or “until”.

Pagination

- Endpoints have a max limit of 100 items per request, which you can adjust with the “per_page” query parameter.
- You move between pages using the “page” query parameter.

Error Handling

- While testing, I was able to spot the API’s response when the rate limit was reached, which helped me handle rate limiting efficiently to make sure I could get all the data I needed. This shows up as a 403 status code and the message “API rate limit exceeded”.
- I also handled errors caused by failed searches for certain parameter values, which showed up as a 404 status code and the message “Not Found”.
- The documentation specifies the formats and parameters the API accepts, so I used that to set up validation for the values users provide to the functions I created.

Final Notes

I used AI to help me create the code, but I take the learning and understanding process of the code I write very seriously. That is, I understand each of the features I implemented and learned about them by consulting the documentation and the AI. I consider it a very powerful tool for learning and improving my productivity, and I always try to use it responsibly to get the most out of it.

During this write-up, I tried to reflect a bit of my creative process, and I aimed to leave the code with useful but not excessive comments—just enough to understand the structure.