Scytale Junior Data Engineer Home Assignment summary report

Endpoints

For this assignment I used the GitHub REST API, where the endpoints I used where,

- `https://api.github.com/orgs/{org}/repos?per_page=100&page={page}`
  - Get the organization and fetch repos associated to that organisation page by page.
- `https://api.github.com/repos/{repo_owner}/{repo_name}/pulls?state=closed&per_page=100&page={page}`
  - Get each repo for the organisation and get pull requests per repo where the state of the pull request is closed.
- `https://api.github.com/repos/{repo_owner}/{repo_name}/pulls/{pr['number']}/reviews`
  - Get each pull requests reviews to see if they had any reviewers to later check if there were any approvals.
- `https://api.github.com/repos/{repo_owner}/{repo_name}/commits/{commit_sha}/status`
  - Get the status of the pull request to check the statuses to determine if there were any checks and if they were passed

Authentication method

The authentication method used was a personal access token, in which i generated a fine grained token to authenticate in order for most endpoints to allow requests. I used a fine grained PAT instead of a classic as they are recommended since they have better scope control, IP restrictions and lower security risks when leaked.

Although PATs are better than using a password and username, they are not always the optimal choice as they are manually generated and updated, in some cases have no expiration (but this can be set by user) and can be leaked more easily than others. However for this assignment it was just simpler to use a PAT than to set up an alternative authentication method.