1 Introduction

This document contains two programming tasks to be finished in the next 48 hours. You can use the programming language and tools that are convenient for you. Please keep track of your progress and all your code inside a GitHub repository. After you finish, please share with me the URL of the GitHub repository. Note that these tasks are not about being the fastest to finish! You will get bonus points for creativity and the quality of your code and documentation.

2 Check documents link status

Create a script (in a language of your preference) that can find links from a directory of files (a sample zip file test.zip is given for testing purposes) and print their status—whether a link is still valid (still exist) or not. In the end, it should print something to this line:

Total link's STATUS
==========
42 Success links
58 Redirection links
26 Client error links
6 Server error links
Total link's validity: 79.00%

The above indicates how many of the extracted links are successful, redirections, client errors, and server errors based on the HTTP response. No need to go into details about the status of the links. Just simple categorize them. Note that we consider any link as valid if the HTTP response is from 200 to 399 (see more details below).

Successful responses (200–299)
Redirection messages (300–399)
Client error responses (400–499)
Server error responses (500–599)

3 Get GitHub Repository statistics

Create a script that can get statistics from all the GitHub repositories of Kaggle (https://github.com/Kaggle). Specifically, the script should generate a report (in a format that you prefer) with the following information for all the repositories found under the kaggle account.

- Total and median number of commits, stars, contributors, branches, tags, forks, releases, closed issues, and environments (please use the GitHub' API to perform this task).

- It should print out the total and median number of source code lines per programming languages used from all the repositories of Kaggle (you may use any tool or library to achieve this).

Please try as much as possible to make the script reusable even for other GitHub accounts.