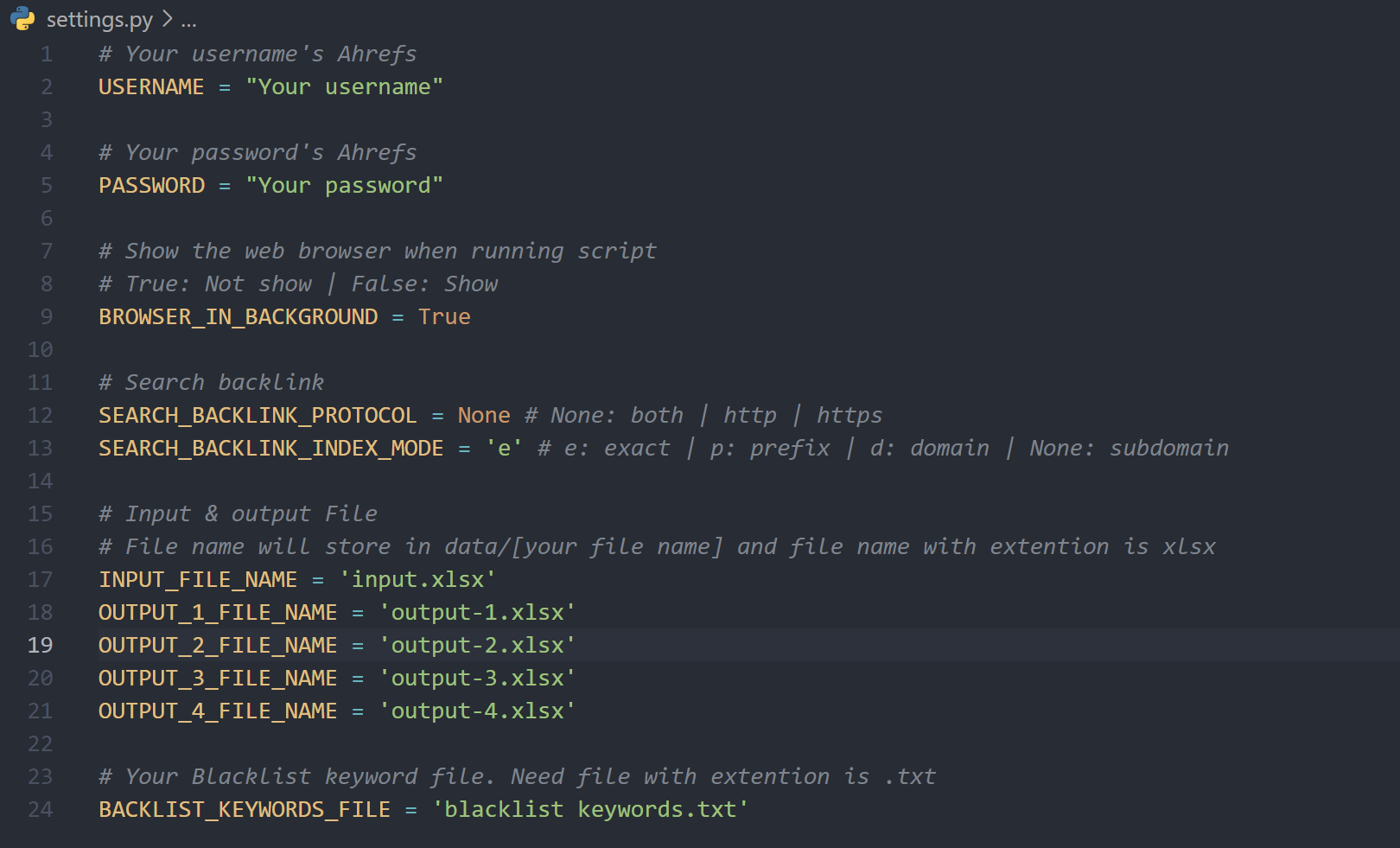
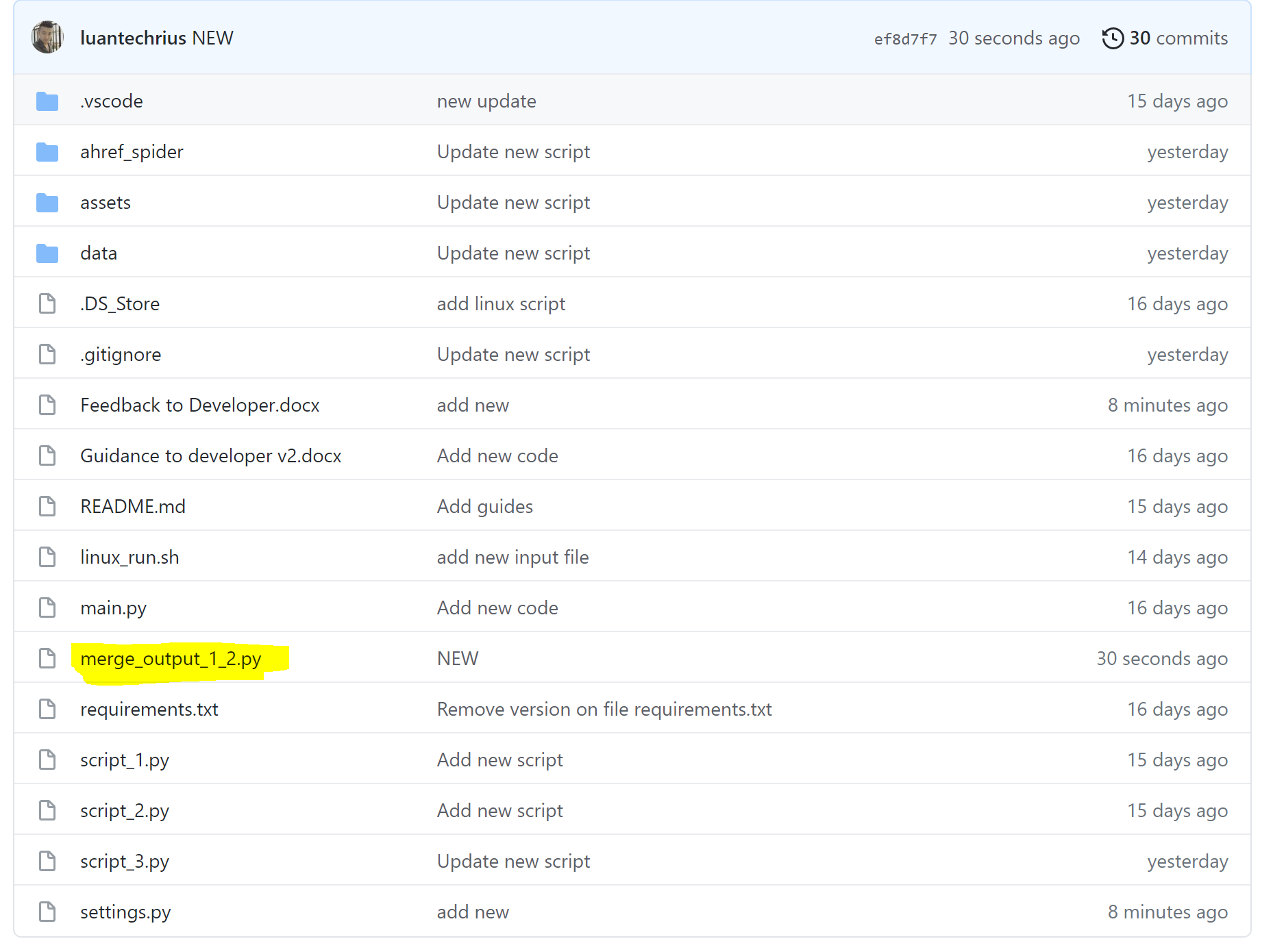
1. General questions –
   1. running the script depends on credentials to AHREFS. Right now you are using the login and password for [luan@techrius.com](mailto:luan@techrius.com) , but when I will run in the future I will need to use my login and password. Is there a way for me to insert that so the script uses that?

**Yes, You can use your username and password to run script. Go to the settings.py file in root folder of my script and fill your email and password.**



* 1. The step “Then merge output-1.xls and output-2.xls to one file >> output-3.xls” – is that something that I do manually?

**Yes, you can merge it by manual. I have added new merge\_output\_1\_2.py file in my scripts. You can download it on** <https://github.com/luantechrius/ahrefs_scraper>



**Notes: Make sure the files output-1.xls and output-2.xls is exist in data folder**

* 1. Is there any limitations that I need to be aware of when I run the scripts?
     1. Is there a limit to the number of rows that I can have in file input.xls?

**No, it’s not limit to the number of rows.**

* + 1. How long will it take to generate output-1.xls on an input.xls file with 1000 rows? With 10,000 rows?

**It will depend the URLs(rows) in your input file. If a URL in your input file have more 1 million backlink, my script need time to crawl a lot of pages. Maybe is one day.**

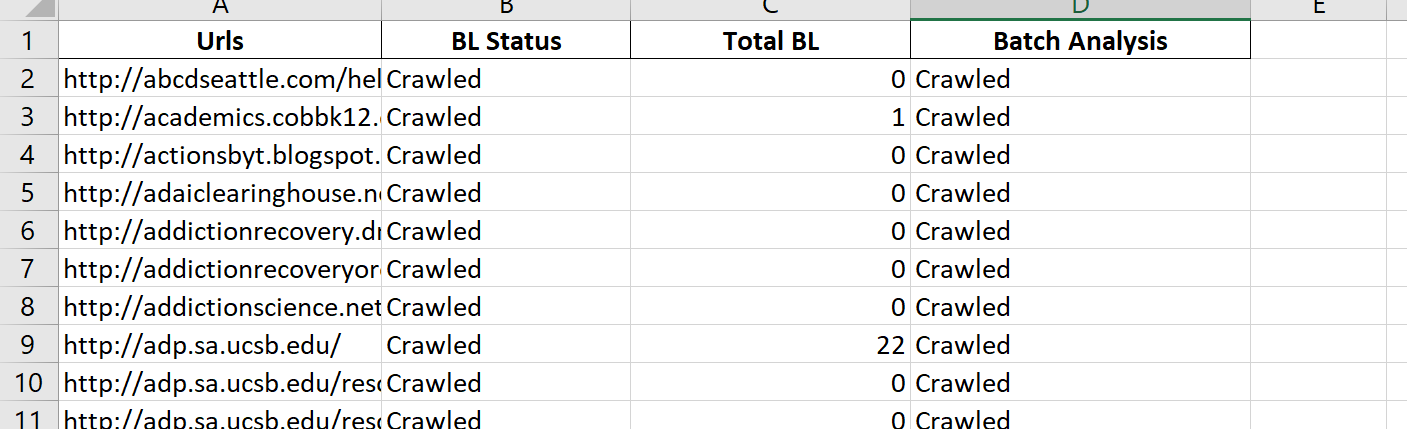
**Note: I got a lot of URLs in your input file is the root domain. It is the reason why my script takes more time to crawl backlinks. I suggest you remove the root domain, only crawl URLs.**

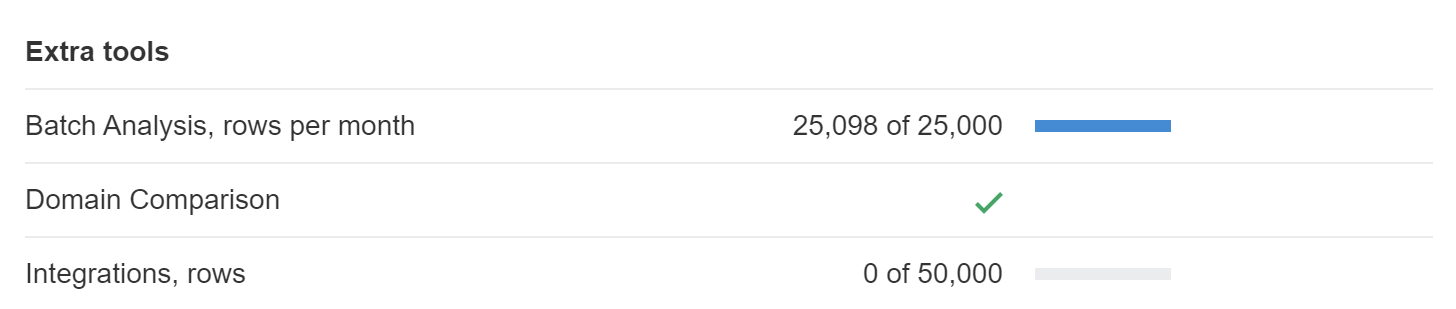
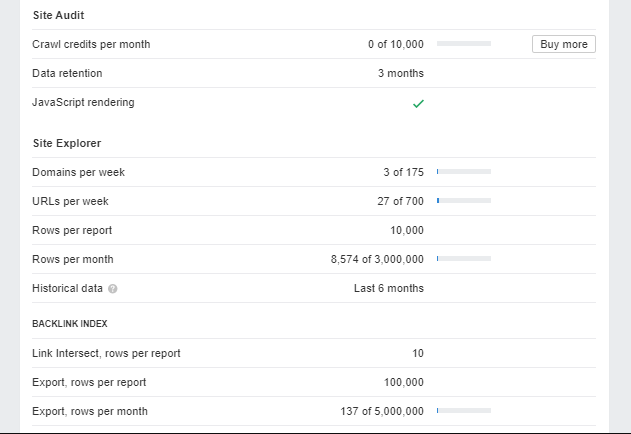
**Example**: <https://www.oyez.org/cases/2020/19-251> => **Keep** | <https://www.oyez.org> => **remove**

* + 1. Similarly – how long will it take to run the other scripts?

**When it crawls per 100 URLs, my script automatically calculates URLs to export backlink data to the output file. I have added the crawl status.**

**My script will add 3 columns "BL Status", "Total BL", "Batch Analysis" to identify crawl later when you stop it...and rut it again.**



* + 1. Are there any limitations because of AHREFS that I need to be aware of, as I’m running the script? Like any of these that the scripts depends on?
       - **You will get unlimited when running script-1, because I crawl data from Ahrefs API**
       - **You will get limited from “Batch Analysis, rows per month”when running script-2**
    2. 
    3. 

1. I realized that one reason for this being difficult is that there are root domains in input.xls, and these pages (the domains) have too many backlinks. **Before we even get started generating output-1.xls**, is there a way to go over **input.xls** and then **find and remove all URLs that are root domains**? for example, I would like to remove URLs like:
   1. [*http://ascendrehabinc.org/*](http://ascendrehabinc.org/)
   2. [*http://brynawel.org/*](http://brynawel.org/)
   3. [*http://dfsp.org/*](http://dfsp.org/)
   4. [*http://kool.corrections.ky.gov/*](http://kool.corrections.ky.gov/)
   5. [*www.hhs.gov/*](http://www.hhs.gov/)
   6. [*www.hhs.gov*](http://www.hhs.gov)

**It’s a hard code. I need time to write a script to remove it. If you want a script, please add a milestone with $20 for a script to remove root domains.**

1. Right now, the file output-1.xls does not include some of the URLs that were needed to be put into AHREFS. For example: <http://abcdseattle.com/helpful-resources/>
   1. Is that only because the missing URLs do not have any backlinks to that URL (for example – the URL above gives me this: <https://ahrefs.com/site-explorer/backlinks/v7/external-similar-links/exact/live/all/all/dofollow/1/domain_rank_desc?target=abcdseattle.com%2Fhelpful-resources%2F> )

**My script auto removes URLs not have a backlink to export to output files, but it has added "Total BL" columns in the input files. You check it when you complete the script-1.py**

1. In the file output-1.xls there are some URLs where you have empty lines. You can see those under the column “Target” (column E in output1.mls) Usually these are root domains:
   1. [www.mass.gov/](http://www.mass.gov/)
   2. [www.samhsa.gov/](http://www.samhsa.gov/)

**This is prolem when the script crawl big backlink data from a root domain via API. That is the reason why I suggest you remove the root domain.**

**OR: My script will download file export on Ahrefs, but you will get limit in site explorer “Rows per month”**