When I Work®

Web Traffic Transformation

Objective

Write a python script/program to transform web traffic data stored in time-record format where each row is a page view into a per-user format where each row is a different user and the columns represent the time spent on each of the pages.

Source Data

The data set consists of 26 CSV files in an AWS S3 bucket. The files are named with lowercase ascii letters (a.csv, b.csv. c.csv, … z.csv). They can be accessed from the public root URL:

https://public.wiwdata.com/engineering-challenge/data/

such that a.csv is available at:
https://public.wiwdata.com/engineering-challenge/data/a.csv

And b.csv is available at:
https://public.wiwdata.com/engineering-challenge/data/b.csv

and so on. Each CSV file has a header row labeling the included columns, which are:

- **drop:** Whether or not this was the last page the user visited before leaving the site.
- **length:** How long the user spent on the page in seconds.
- **path:** The page within the website that the user visited.
- **user_agent:** The browser identifier of the user visiting the page.
- **user_id:** The unique identifier for the user visiting the page.

Output Data

Write a program that converts the web traffic from these 26 CSV files into a single CSV file that contains one row for each user_id and has columns populated with the ***length*** of time each user spent on each ***path***. The resulting CSV file should look something like:

user_id	/	/features/desktop	/signup	/tutorial/step-one	...
1	12	2	4	6	...
2	24	6	2	4	...
...	...	...	...	...	...

Success Criteria

1. The program should be designed so that the root URL could be changed later and the program re-run on new data. That means downloading the data must be done within the program.
2. The program should write out to a standard CSV file that can be opened in Excel for review.
3. The program should be written in python3.
4. The program should be available in a public GitHub or GitLab account with documentation on how to install and run it.

Submission Details:

- Please complete this challenge within 3 days of receipt.
- Challenges that use a code generator to complete the challenge will not be accepted.

Feel free to reach out to the When I Work team with questions.