CP-367 A (Winter 2026) Assignment 3 due date: March 1st , 2026

Description: a take-home , individual assignment.

Objectives: demonstrating your understanding and implementation of text processing using Linux utility services (sed, awk).

Task: Use **sed** and **awk** to clean, transform, and analyze datasets.

Submission (read carefully): Copy your complete scripts with outputs in a single Word document that contain screenshots of all the outputs. Make sure you include your name and student ID# in the report. Redirect your outputs to new files (whenever applicable) and submit these files as well. You zip everything and upload the zipped file to Assignment 3 drop box in Dropbox section on MyLS.

Note (1): The use of generative AI is not permitted in this course in terms of generating partial or complete solutions – it is a violation of the Academic Code of Conduct. However, the students can search and learn about the assignment topics using AI tools.

Note (2): It is the student's responsibility to read, understand and follow the Laurie's Academic Code of Conduct.

Note (3): Students with approved accommodation by WLU must indicate in the comment box in the drop box that they have an approved accommodation with details of what they are entitled to, so it matches my record.

General rubrics:

- Functionality: 50% (scripts meet all the specified requirements)

- Code Quality: 20% (Readability, style, comments, proper variable naming)

- Correct Outputs: 20% (Outputs are correct and accurate- whenever applicable Implementation of basic error handling if required)

- Testing: 10% (Evidence of thorough testing)

Question 0: [10 marks]

Create a text file that contains the following 12 lines:

2024-11-01 10:15:32 , 192.168.1.10 , GET , /index.html , 200 , 532

2024-11-01 10:15:35,192.168.1.11,POST,/login,401,721

2024-11-01 10:16:01 ,192.168.1.10 ,get,/products,200,1024

2024-11-01 10:16:45,192.168.1.12,GET,/index.html,500,0

2024-11-01 10:17:02 , 192.168.1.13 , POST , /cart , 200 , 2048

2024-11-01 10:17:30,192.168.1.11,POST,/login,200,850

2024-11-01 10:17:30,192.168.1.11,POST,/login,200,850

2024-11-01 10:18:05 ,192.168.1.14 , GET , /checkout , 200 ,

2024-11-01 10:18:20, ,GET,/index.html,200,532

2024-11-01 10:19:01,192.168.1.15,DELETE,/admin,403,245

2024-11-01 10:19:45,192.168.1.10,GET,/products,200,1100

2024-11-01 10:20:10,192.168.1.16,PUT,/api/update,201,999

Below is a description to the fields you see in the dataset:

Field	Description
Date	request date
Time	request time
IP	client IP
Method	GET, POST, etc
URL	resource
Status	HTTP status
Size	response size

PART I: In the following 4 question, use (sed) to clean the data
[40 marks]

Question 1 : Remove extra spaces around commas

Question 2 : Convert HTTP methods to uppercase

Question 3 : Remove duplicate records

Question 4 : Replace missing size values with 0

Question 5: Replace All Status Code 200 with the Word SUCCESS

PART II: In the following 10 question, use (awk) to analyze the data
[50 marks]

Question 6: Print only IP and URL

Question 7: Count number of ERROR requests(Error = status ≥ 400)

Question 8: Calculate average response size

Question 9: Count requests per IP

Question 10: Find IP that generated the most traffic (Traffic = total response size)

Question 11: Find the Largest Response Size and Its IP address

Question 12: Count requests per HTTP method

Question 13: Detect Suspicious IP (Any IP with more than 3 requests)

Question 14: Find the Most Frequently Accessed URL (the URL that was accessed the most times)