Shell Scripting (Bash) - In Class Lab

1. Learn to find files and merge CSV data safely.

Given

- A folder containing many .csv files in subfolders.
- a name for the combined file

Task

Write commands or a small script that:

- Finds all .csv files in a folder
- Extracts data rows only (ignore headers).
- Combines them into one file.

```
OUTPUT_FILE="$1"
first=1
rm $OUTPUT_FILE
for file in *.csv
do
    echo $file
    echo $first
    if [ $first -eq 1 ]; then
        cat $file > $OUTPUT_FILE      # keep header from first file
        first=0
    else
        tail -n +2 $file >> $OUTPUT_FILE   # skip header
    fi
done
```

2. Validate CSV Columns

Given

1. Give a csv file as input with two columns
2. Column `date` in YYYYMMDD
3. Column `student_email` like abc@example.com format

- Like containts `@` and a `.` character

Task

1. Read each row
2. Validate date, student_email if doesnt match then show Row number and error

```
file="$1"
row=1
tail -n +2 "$file" | while IFS=, read -r date student_email
do
    row=$((row+1))
    # check date (8 digits)
    if echo "$date" | grep -Eq '^[0-9]{8}$'; then
        : # do nothing
    else
        echo "Row $row: date error"
    fi
    # check email
    if echo "$student_email" | grep -Eq '@.*\.'; then
        : # do nothing
    else
        echo "Row $row: student_email error"
    fi
done
```

Steps

- tail -n +2 skips the header.
- IFS=, read splits the CSV row into two columns.
- Regex checks:
  - `^[0-9]{8}$` → date format.
  - `^[^@]+@[^@]+\.[^@]+$` → simple email format.

3. Practice grouping and aggregation.

Given

1. Give a csv file as input with two columns
2. `date,student_name,student_email,course_code,attendance`

Task

1. Read course_code and attendance
2. Calculate attendance percentage for each course_code
3. print

```
file="$1"
courses=$(tail -n +2 "$file" | cut -d, -f4 | sort | uniq)
data=$(tail -n +2 "$file" | cut -d, -f4,5 )
for course in $courses
do
    total=0
    present=0
    while IFS=, read -r course_code attendance
    do
        if [[ $course_code == $course ]]; then
            total=$((total+1))
            if [[ $attendance == "1" ]]; then
                present=$((present+1))
            fi
        fi
    done <<< $data
    percent=$((100 * present / total))
    echo "$course $percent%"
done
```