

# H.W. 4 : Data Wrangling

## • Output: task1.sh

```
hw4 > $ task1.sh
1  #!/bin/bash
2
3  pid=$(ps aux | grep "[b]ash infinite.sh" | awk '{print $2}') && { kill -9 $pid && echo "Process infinite.sh with pid: $pid is killed."; } || echo "Process infinite.sh not running."
4
```

PROBLEMS OUTPUT DEBUG CONSOLE TERMINAL PORTS COMMENTS

● @Jk42github → /workspaces/HW1 (main) \$ cd hw4  
● @Jk42github → /workspaces/HW1/hw4 (main) \$ bash infinite.sh  
● @Jk42github → /workspaces/HW1/hw4 (main) \$ pid=\$(ps aux | grep "[b]ash infinite.sh" | awk '{print \$2}') && { kill -9 \$pid && echo "Process infinite.sh with pid: \$pid is killed."; } || echo "Process infinite.sh not running."  
Process infinite.sh with pid: 4928 is killed.  
○ @Jk42github → /workspaces/HW1/hw4 (main) \$

## • Output: task2.sh

```
[Preview] README.md $ task2.sh x
```

```
hw4 > $ task2.sh
1  #!/bin/bash
2  cd dataset1/
3  grep -l "sample" file* | xargs grep -o "CSC510" | uniq -c |
4  cut -d: -f1 | grep -E -v "[0-2]\s+file" | sed 's/^ *//' |
5  gawk '{count=$1;file=$2;"ls -l " file | getline size; print count, size}' |
6  sort -k1,1nr -k6,6nr | awk '{print $10}' | sed 's/file_/filtered_/'
7
```

PROBLEMS OUTPUT DEBUG CONSOLE TERMINAL PORTS COMMENTS

● @Parth291102 → /workspaces/HW1/hw4 (main) \$ ./task2.sh  
filtered\_20  
filtered\_10  
filtered\_4  
filtered\_14  
filtered\_24  
filtered\_2  
filtered\_22  
filtered\_30  
filtered\_3  
filtered\_23  
filtered\_12  
filtered\_13  
○ @Parth291102 → /workspaces/HW1/hw4 (main) \$

- **Output: task3.sh**



```
hw4 > $ task3.sh
1  #!/bin/bash
2
3  sed 's/^,/NULL,/: :a;s/,/,NULL,/g;ta' titanic.csv > updated_titanic.csv
4
5  gawk -F',' ' $13 ~ /S\s*$/ && $3 == 2 { print }' updated_titanic.csv | sed 's/male/M/g; s/female/F/g' | awk -F',' ' $7 != "NULL" { sum += $7;
ULL" { sum += $7; count++ } END { if (count > 0) print sum / count;}'
30.3867
@Mrudanihada → /workspaces/HW1/hw4 (main) $
```

- **URL of the Github repo:** <https://github.com/PMJ-SE-Fall-24/HW1/tree/main/hw4>

**-Submitted by: Group 25**  
**Parth Ripal Parikh** ([pparikh2@ncsu.edu](mailto:pparikh2@ncsu.edu))  
**Mrudani Vishal Hada** ([mhada2@ncsu.edu](mailto:mhada2@ncsu.edu))  
**Jamnesh Patel** ([jpatel42@ncsu.edu](mailto:jpatel42@ncsu.edu))