Homework 5

FROM: Jonas Zhonghan Xie
TO: Raj
SUBJECT: RE: Thank you and a few more questions

Hi Raj,

Hope my responses are helpful to you. I am happy to help with your follow up questions.

Part 1

1. The first column shows the tasks PIDs running `prime.py`

```
jonasxie@ip-172-31-78-96:~/week5$ ps -ax | grep 'prime.py'
   1886 pts/1    S      0:00 python3 prime.py 300 300000000
   1887 pts/1    S      0:00 python3 prime.py 300 30000
   1888 pts/1    S      0:00 python3 prime.py 300 30000
   1889 pts/1    S      0:00 python3 prime.py 300 300000
   1890 pts/1    S      0:00 python3 prime.py 300 30000
   1891 pts/1    S      0:00 python3 prime.py 300 30000
   1892 pts/1    S      0:00 python3 prime.py 300 300000
   1898 pts/1    S      0:00 python3 prime.py 300 300000000
   1899 pts/1    S      0:00 python3 prime.py 300 30000
   1900 pts/1    S      0:00 python3 prime.py 300 30000
   1901 pts/1    S      0:00 python3 prime.py 300 300000
   1902 pts/1    S      0:00 python3 prime.py 300 30000
   1903 pts/1    S      0:00 python3 prime.py 300 30000
   1904 pts/1    S      0:00 python3 prime.py 300 300000
```

2. The system load is shown below. The system load in the recent 1 min is 0.02, in the recent 5 mins is 0.52, and in the recent 15 mins is 0.37. As we have 1 processor, the system load is less than 1, which means the system is not overloaded.

```
jonasxie@ip-172-31-78-96:~/week5$ w
 20:55:16 up 57 min,  2 users,  load average: 0.02, 0.52, 0.37
USER     TTY      FROM             LOGIN@   IDLE   JCPU   PCPU WHAT
jonasxie pts/0    69.136.155.115   19:57    0.00s  0.22s  0.00s w
jonasxie pts/1    69.136.155.115   20:47    6:18   0.88s  0.03s -bash
```

3. It seems none of them stopped.

```
jonasxie@ip-172-31-78-96:~/week5$ ps -ax | grep 'prime.py'
   1886 pts/1    S      0:00 python3 prime.py 300 300000000
   1887 pts/1    S      0:00 python3 prime.py 300 30000
   1888 pts/1    S      0:00 python3 prime.py 300 30000
   1889 pts/1    S      0:00 python3 prime.py 300 300000
   1890 pts/1    S      0:00 python3 prime.py 300 30000
   1891 pts/1    S      0:00 python3 prime.py 300 30000
   1892 pts/1    S      0:00 python3 prime.py 300 300000
   1898 pts/1    S      0:00 python3 prime.py 300 300000000
   1899 pts/1    S      0:00 python3 prime.py 300 30000
   1900 pts/1    S      0:00 python3 prime.py 300 30000
   1901 pts/1    S      0:00 python3 prime.py 300 300000
   1902 pts/1    S      0:00 python3 prime.py 300 30000
   1903 pts/1    S      0:00 python3 prime.py 300 30000
   1904 pts/1    S      0:00 python3 prime.py 300 300000
```

4. I used `grep` to locate the python processes and then used `awk` to locate the processIDs, finally used `xargs kill` to kill the processes. I used `grep 'pts/1'` to locate only the python3 processes.

```
jonasxie@ip-172-31-78-96:~/week5$ ps -ax | grep 'prime.py' | grep 'pts/1' | awk -F " " '{print
 $1}' | xargs kill
```

Part 2

1&2. I used `wget` and `unzip` to download and unzip the NYC inspection file.

```
jonasxie@ip-172-31-78-96:~/week5$ wget "https://github.com/SI504/TextParse1/raw/main/nyc.zip"
--2025-02-23 21:06:21--  https://github.com/SI504/TextParse1/raw/main/nyc.zip
Resolving github.com (github.com)... 140.82.112.3
Connecting to github.com (github.com)|140.82.112.3|:443... connected.
HTTP request sent, awaiting response... 302 Found
Location: https://raw.githubusercontent.com/SI504/TextParse1/main/nyc.zip [following]
--2025-02-23 21:06:21--  https://raw.githubusercontent.com/SI504/TextParse1/main/nyc.zip
Resolving raw.githubusercontent.com (raw.githubusercontent.com)... 185.199.110.133, 185.199.11
1.133, 185.199.108.133, ...
Connecting to raw.githubusercontent.com (raw.githubusercontent.com)|185.199.110.133|:443... co
nnected.
HTTP request sent, awaiting response... 200 OK
Length: 22184392 (21M) [application/zip]
Saving to: ‘nyc.zip’

nyc.zip                 100%[=============================>]  21.16M   112MB/s    in 0.2s

2025-02-23 21:06:22 (112 MB/s) - ‘nyc.zip’ saved [22184392/22184392]

jonasxie@ip-172-31-78-96:~/week5$ unzip nyc.zip
Archive:  nyc.zip
  inflating: nyc.csv
```

3. I used `wc` to count the number of lines, words and characters in the file. There are 399,919 lines, more than 1.3M words and more than 14.6M characters in the file.

```
jonasxie@ip-172-31-78-96:~/week5$ wc nyc.csv > counts.txt
jonasxie@ip-172-31-78-96:~/week5$ cat counts.txt
   399919  13816290 146142203 nyc.csv
```

4. I used `grep` to filter out the column row in the csv file. Then I used `awk` to locate the fifth column. I used `sort` and `uniq` to filter the unique streets in the column. And finally saved it to the txt file.

```
jonasxie@ip-172-31-78-96:~/week5$ cat nyc.csv | grep -v "|STREET|" | awk -F "," '{print $5}' |
 sort | uniq > "unique_streets.txt"
```

5. The first (5) lines of the files are shown below.

```
jonasxie@ip-172-31-78-96:~/week5$ head -n 5 counts.txt
   399919  13816290 146142203 nyc.csv
jonasxie@ip-172-31-78-96:~/week5$ head -n 5 unique_streets.txt
 TOSSED"
"5TH AVENUE
"AMERICAN AIR
"HIGHLINE PARK
"W 15th Street @ 10th Ave
```

Part 3

1. I used `curl -s` to fetch the file without showing the transfer status. I used `grep '#||'` to filter out the first row. Then I used `awk` to locate the column of "Type 1" and `sort` and `uniq` to identify the unique values. Finally I saved it to the `unique_type1.txt` file.

```
jonasxie@ip-172-31-78-96:~/week5$ curl -s https://raw.githubusercontent.com/SI504/TextParse1/main/Pokemon.txt  | grep -v '#||' | awk -F '|' '{print $5}' | sort -n | sort | uniq > "unique_type1.txt"
jonasxie@ip-172-31-78-96:~/week5$ cat unique_type1.txt
Bug
Dark
Dragon
Electric
Fairy
Fighting
Fire
Flying
Ghost
Grass
Ground
Ice
Normal
Poison
Psychic
Rock
Steel
Water
```

2. Using the similar command as 3.1, I used `sort -n` to sort the values in the attack column.

```
jonasxie@ip-172-31-78-96:~/week5$ curl -s https://raw.githubusercontent.com/SI504/TextParse1/main/Pokemon.txt | grep -v '#||' | awk -F '|' '{print $13}' | sort -n > attack.txt
```

3. I used "||True" to identify the legendary pokemon. The command is shown below and the top 5 lines are shown below.

```
jonasxie@ip-172-31-78-96:~/week5$ curl -s "https://raw.githubusercontent.com/SI504/TextParse1/main/Pokemon.txt" | grep '||True' > "legendary.txt"
```

4. I used `awk` to locate the column of HP and then used `sort -n` to sort the values in the HP column. Then I used `grep` to filter the row of 45 and then passed it to `wc` to count the lines. There are 38 pokemons with HP of 45. Then I saved the result to `hp_45_count.txt`

```
jonasxie@ip-172-31-78-96:~/week5$ curl -s https://raw.githubusercontent.com/SI504/TextParse1/main/Pokemon.txt  | awk -F '|' '{print $11}' | grep '45' | wc -l
38
```

```
jonasxie@ip-172-31-78-96:~/week5$ curl -s https://raw.githubusercontent.com/SI504/TextParse1/main/Pokemon.txt  | awk -F '|' '{print $11}' | grep '45' | wc -l > "hp_45_count.txt"
```

5. The first 7 lines of the txt files are shown below.

```
jonasxie@ip-172-31-78-96:~/week5$ head -n 7 unique_type1.txt
Bug
Dark
Dragon
Electric
Fairy
Fighting
Fire
jonasxie@ip-172-31-78-96:~/week5$ head -n 7 attack.txt
5
5
10
10
10
15
20
jonasxie@ip-172-31-78-96:~/week5$ head -n  7 legendary.txt
144||Articuno||Ice||Flying||580||90||85||100||95||125||85||1||True
145||Zapdos||Electric||Flying||580||90||90||85||125||90||100||1||True
146||Moltres||Fire||Flying||580||90||100||90||125||85||90||1||True
150||Mewtwo||Psychic||||680||106||110||90||154||90||130||1||True
150||MewtwoMega Mewtwo X||Psychic||Fighting||780||106||190||100||154||100||130||1||True
150||MewtwoMega Mewtwo Y||Psychic||||780||106||150||70||194||120||140||1||True
243||Raikou||Electric||||580||90||85||75||115||100||115||2||True
jonasxie@ip-172-31-78-96:~/week5$ head -n 7 hp_45_count.txt
38
```

Part 4

1. I created the zip file `week5.tar.gz`. It is about 43M in size.

```
jonasxie@ip-172-31-78-96:~$ tar -zcvf week5.tar.gz week5/
week5/
week5/attack.txt
week5/prime.sh
week5/hp_45_count.txt
week5/nyc.zip
week5/prime.py
week5/legendary.txt
week5/unique_streets.txt
week5/counts.txt
week5/unique_type1.txt
week5/nyc.csv
```

```
jonasxie@ip-172-31-78-96:~$ du ~/week5.tar.gz
43352   /home/jonasxie/week5.tar.gz
```

Let me know if you have any follow up questions.

Best,
Jonas