

Stat 292 - Recitation 4

Unix Terminal - wget - Pipelines - AWK

Orçun Oltulu

7 / 4 / 2021

Exercise 1:

Part A:

Create a directory Recitation4. Inside that directory create a employee.txt file with the following information;

```
ajay manager account 45000
sunil clerk account 25000
varun manager sales 50000
amit manager account 47000
tarun intern sales 15000
deepak clerk sales 23000
sunil intern sales 13000
satvik director purchase 80000
```

```
# cat > employee.txt

# ajay manager account 45000
# sunil clerk account 25000
# varun manager sales 50000
# amit manager account 47000
# tarun intern sales 15000
# deepak clerk sales 23000
# sunil intern sales 13000
# satvik director purchase 80000
```

Part B:

Sort the employee.txt file alphabetically by employee name.

```
# sort employee.txt
```

Part C:

Sort the employee.txt file in reverse order.

```
# sort -r employee.txt
```

Part D:

Assume that 4th column represents their Salaries. Now, sort the employee.txt file by Salary. (**Hint:** Use -k option)

```
# sort -k 4n employee.txt
```

Part E:

Inside Recitation4 create a carbrands.txt file with the following information;

Audi
Cadillac
BMW
Dodge

Then, check if the file given is already sorted or not. (**Hint:** Use -c option)

```
# sort -c carbrands.txt
```

Exercise 2:

Part A:

Using previous employee.txt file; print only the Salary column.

```
# cut -d " " -f4 employee.txt
```

Part B:

Now get first 3 letters of each item in carbrands.txt

```
# cut -c 1,2,3 carbrands.txt
```

Part C:

Inside Recitation4 create a employee.csv file with the following information;

```
ajay,manager,account,45000
sunil,clerk,account,25000
varun,manager,sales,50000
amit,manager,account,47000
tarun,intern,sales,15000
deepak,clerk,sales,23000
sunil,intern,sales,13000
satvik,director,purchase,80000
```

Part D:

Print only the second and forth columns of employee.csv file. (Be careful about the delimiter)

```
# cut -d ',' -f 2,4 employee.csv
# alternatively;
# cut -d, -f 2,4 employee.csv
```

Exercise 3:

Part A:

Sort Salary column of the employee.csv file.

```
# cut -d, -f4 employee.csv | sort
```

Part B:

Obtain the ‘unique’ positions (2nd column) of the observations in employee.csv file. (**Hint:** Before using ‘uniq’ command, sort the column of interest.)

```
# cut -d, -f2 employee.csv | sort | uniq
```

Part C:

Obtain the number of people working in each department (3rd column) in employee.csv file.

```
# cut -d, -f3 employee.csv | sort | uniq -c
```

Part D:

What is the highest salary in employee.csv file.

```
# cut -d, -f4 employee.csv | sort -r | head -n1
```

Exercise 4:

Refer to the employee.csv file.

Part A:

Using AWK, print the observations whose title is ‘manager’.

```
# awk -F, '{if ($2=="manager") {print}}' employee.csv
```

Part B:

Print the observations whose Salary is less than 40000.

```
# awk -F, '{if ($4<40000) {print}}' employee.csv
```

Part C:

How many of the employees are working in ‘sales’ department?

```
# awk -F, '{if ($3=="sales") {print}}' employee.csv | wc -l
```

Part D:

How many of the employees are working in ‘sales’ department and having salaries less than 40000?

```
# awk -F, '{if ($3=="sales" && $4<40000) {print}}' employee.csv | wc -l
```

Part E:

What are the names of the employees who satisfy the condition in Part D?

```
# awk -F, '{if ($3=="sales" && $4<40000) {print $1}}' employee.csv
```

Exercise 5:

Part A:

Download the data set that Istanbul meteorology department has published for February, 2021.

Use this [link](#) to see the web page.

Use the following link to download the data set using ‘wget’ command.

https://data.ibb.gov.tr/dataset/94197af2-67f4-428a-a031-432e1848cc59/resource/db1b0480-5410-44f1-8df6-c13fbf159519/download/meteorology_observation_202102.csv

Part B:

Print the first 25 rows.

```
# head -n25 meteorology_observation_202102.csv
```

Part C:

Find how many different SENSOR_TYPE.

```
# cut -d, -f3 meteorology_observation_202102.csv | sort | uniq
```

Part D:

Now, obtain the frequencies of each SENSOR_TYPE.

```
# cut -d, -f3 meteorology_observation_202102.csv | sort | uniq -c
```

Part E:

Create a kadikoy.csv file in Recitation4 directory, consisting only the variable names of meteorology_observation_202102.csv file.

```
# head -n1 meteorology_observation_202102.csv | cat > kadikoy.csv
```

Part F:

Create a sub-sample of the dataset with the following condition; 'OBSERVATORY_NAME = KADIKOY_GOZTEPE_MGM'. **Append** this sub-sample to kadikoy.csv.

```
# awk -F, '{if ($2=="KADIKOY_GOZTEPE_MGM")  
# > {print}}' meteorology_observation_202102.csv | cat >> kadikoy.csv
```

*# Command continues, due to markdown problems, had to use a new command line.
You can copy paste this directly or combine the command in a single line.*

Part G:

Get the list of AVERAGE_TEMPERATURE (f6) and AVERAGE_FELT_TEMPERATURE (f24) for Kadikoy, but print only the first 20 rows.

```
# awk -F, '{print $6 "," $24}' kadikoy.csv | head -n20  
# Alternatively,  
# cut -d, -f6,24 kadikoy.csv | head -n20
```

Part H:

Get the list of AVERAGE_TEMPERATURE (f6) where AVERAGE_WIND (f12) is bigger than 4 for kadikoy.csv.

```
# awk -F, '{if ($12>4) {print $6}}' kadikoy.csv
```