

Homework0

Due Date: 4th Feb, 11:59 PM

Name:

Student ID:

Submission Instructions: This PDF contains fields where you can type in your answers. Please save your work regularly so you don't lose any of your answers. In the answer box, you need to write the shell script that will answers the questions below. For all the questions, write the appropriate command.

Example: Print the second column of a given file example.txt.

```
awk '{ print $2 }' example.txt
```

Useful Links:

- awk, sed - [https://emb.carnegiescience.edu/sites/default/files/140602-sedawkbash.key .pdf](https://emb.carnegiescience.edu/sites/default/files/140602-sedawkbash.key.pdf)
- shell: <http://swcarpentry.github.io/shell-novice/>

After you complete the assignment, upload a copy to Gradescope. A link to Gradescope can be found on the left side of the course Canvas page.

1. To access the data, you need to clone the data from the given github repository link:

https://github.com/oudalab/russian-troll-tweets/blob/master/IRAhandle_tweets_1.csv. Write the command to access the file.

2. How many records, including header are there in the file retrieved? Provide the command and output.

3. What command extracts the column of the author names from the file?

4. How many tweets in the file are by author 10_GOP? Provide the command and output.

5. Write command to calculate the percentage of tweets that are by the author 10_GOP?

Note: use a one-line awk command, do not do the division from the previous questions except to check your answer.

6. How many tweets were published in the year 2017? Provide the command and the output.

7. How many tweets were published between the months of March and September of 2017? Provide the command and output.

8. How many tweets were retweets or quotes rather than original tweets? Provide the command and the output.

9. What was the change in number of followers for author 10_GOP from Jan to Dec 2016? Provide command and output.

10. How many different "account type" labels are there? What are they? Provide the command and output.

11. Which of the following commands will print the path of the file myfile.txt:

- a) `find 'myfile*'`
- b) `find "$(pwd)" -name 'myfile*'`
- c) `find "$pwd" -name 'myfile*'`
- d) `find . -path 'myfile*'`

12. Which of the following commands will print the square of numbers in range 7- 11 inclusive:

- a) `awk 'BEGIN { for x in range(7, 12) print x*x; }'`
- b) `awk 'BEGIN { for (x = 7; x < 12; x++) print x*x; }'`
- c) `awk 'BEGIN { for x in [7:12] print x*x; }'`
- d) none of the above

13. In a file abstract.txt, which of the following command will print all lines with more than 180 characters:

- a) `awk 'length($0) > 180' abstract.txt`
- b) `awk 'length(0) > 180'`
- c) `awk 'length() > 180' abstract.txt`
- d) None of the above

14. Which of the following commands is used to check if \$a is greater than 5 or not:

- a) `-gt`
- b) `-ge`
- c) `-lt`
- d) None of the above