

IS457_HW9_127

1.Code

```
#!/bin/bash

# HW 9 - Due Monday Dec 3, 2018 in moodle.

# Upload .sh file to Moodle with filename: HW9_457IDS_YOURCLASSID.sh

# Please make sure all the commends work well in WholeTale, we will test your script.

# In your hard copy report, please include the UNIX / Linux script, input arguments, and results.

# There are multiple solutions for this homework. The grading will be based on the successful
running of your code and the correct output as we specified. We will grade your homework on
WholeTale.

# For this assignment we will use some basic commends of UNIX / Linux.
# The text Hw_9.txt & adult.csv are uploaded to Moodle.
# You can use "text file" editor to edit HW9_457IDS_YOURCLASSID.sh and run in "Terminal".

# You can use the following commands to run the script (for example on google cloud):
# chmod +x HW9_457IDS_YOURCLASSID.sh
# ./HW9_457IDS_YOURCLASSID.sh Argument_1 Argument_2 Argument_3 Argument_4

# Here is a list of your input arguments:
# Argument_1: a positive number
# Argument_2: a lowercase word
# Argument_3: text file ( .txt)
# Argument_4: a positive integer which is less than 15

# Q1 (2pts). Check whether your input integer(Argument_1) is even or odd
#   and print your result. (5 points)

echo "***** Q1 *****"

# Your answer here:

if [ (($1 %2)) -eq 0 ];then
    echo "The number is even!"
else
    echo "The number is odd!"
fi

# Q2 (2pts). Input a lowercase letter(Argument_2) and convert it to uppercase and print your
result. (5 points)
#   (Hint: tr)
echo "***** Q2 *****"
echo $2 | tr "[a-z]" "[A-Z]"

# Q3 (2pts). Convert the following phrase "CS 398/IS 457:INTRODUCTION
#   TO DATA SCIENCE" into separate words, and put each word on its own
#   line (ignoring space, '/' and ':'). (5 points)
```

The output looks like:

```
# CS
# 398
# IS
# 457
# INTRODUCTION
# TO
# DATA
# SCIENCE
```

```
echo "***** Q3 *****"
echo "CS 398/IS 457:INTRODUCTION TO DATA SCIENCE" | tr '[:punct:]' '\n'
```

Q4 (2pts). Sort the answer in Q3 by descending order. (5 points)

The output would look like:

```
# TO
# SCIENCE
# IS
# INTRODUCTION
# DATA
# CS
# 457
# 398
echo "***** Q4 *****"
echo "CS 398/IS 457:INTRODUCTION TO DATA SCIENCE" | tr '[:punct:]' '\n' | sort -r
```

Q5 (2pts). Count the lines of your text file(Argument_3). (5 points)
(Hint: wc)

```
echo "***** Q5 *****"
echo "The number of lines in $3 is:"
wc -l $3
```

Q6 (2pts). Count the frequency of a input word(Argument_2) in a text
file(Argument_3), and print "The frequency of word ____ is ____".
(5 points)
(Hint: grep)

```
echo "***** Q6 *****"
echo "The frequency of word $2 is:"
```

```
tr ' ' '\n' < $3 | grep -i -w $2 | wc -l
```

Q7 (2pts). Print the number of unique words in the text file(Argument_3).
(5 points)
(Hint: uniq, sort)

```
echo "***** Q7 *****"
echo "The number of unique words in the text file:"
```

```
tr '[:punct:]' '\n' < $3 | sed '/^$/d' | sort -f|uniq -c | wc -l
```

Q8 (2pts). Print the number of words that begin with the letter 'b' in the

```

# text file(Argument_3) (5 points).
echo "***** Q8 *****"
echo "The number of words that begin with letter 'b':"
grep -o '\bb\w*' $3|wc -l

# Q9 (2pts). Print top-k(Argument_4) and find the most frequent word and their frequencies.
# (5 points).
# (Hint: head)
echo "***** Q9 *****"
echo "Top-$4 words are:"
tr '[[ ][:punct:]]' '\n'< $3 |sed '/^$/d'| sort | uniq -c |sort -n -r| head -$4

# Q10 (4pts). The dataset adult-income.csv provides some clean records of adults to predict
whether income exceeds $50K/yr. For details, visit the UCI repository.
# Calculate how many categories are there in "workclass" (2nd column) and print your result. (5
points)
# (Hint: awk)
echo "***** Q10 *****"
awk -F "\"*,\"" '{print $2}' adult-income.csv | tr '?' ' '|sed '/^$/d'| sort | uniq -c

# Q11 (4pts). For your output in Q10, change the format of categories. Replace "-" with "_".
# (Hint: sed)
echo "***** Q11 *****"

awk -F "\"*,\"" '{print $2}' adult-income.csv | tr '?' ' '|sed '/^$/d'| sort | uniq -c |sed 's/-/_/g'

```

2.Output

```

jovyan@31268883bb89:~/work/unixstff$ ./hw9_FA18_127.sh 4 we Hw_9.txt 5
***** Q1 *****
The number is even!
***** Q2 *****
WE
***** Q3 *****
CS
398
IS
457
INTRODUCTION
TO
DATA
SCIENCE
***** Q4 *****
TO
SCIENCE
IS
INTRODUCTION
DATA
CS
457
398
***** Q5 *****
The number of lines in Hw_9.txt is:
19 Hw_9.txt
***** Q6 *****
The frequency of word we is:
2
***** Q7 *****
The number of unique words in the text file:
206
***** Q8 *****
The number of words that begin with letter 'b':
7

```

***** Q9 *****

Top-5 words are:

48 the
20 and
14 shell
13 of
10 The

***** Q10 *****

960 Federal-gov
2093 Local-gov
7 Never-worked
22696 Private
1116 Self-emp-inc
2541 Self-emp-not-inc
1298 State-gov
14 Without-pay

***** Q11 *****

960 Federal_gov
2093 Local_gov
7 Never_worked
22696 Private
1116 Self_emp_inc
2541 Self_emp_not_inc
1298 State_gov
14 Without_pay