PROJECT 7

LINUX SYSTEM

A PROPOSAL BY

SAMAHER ALOQAYLI

SUPERVISOR
MS.SHEFA

Project folder

Access to the project folder via (cd) .And verify its contents by (ls)

```
samah@LAPTOP-IICCEV1N: /mnt/c/users/samah/desktop/project7
samah@LAPTOP-IICCEV1N:~$ cd /
samah@LAPTOP-IICCEV1N:/$
samah@LAPTOP-IICCEV1N:/$ cd mnt/c/users/samah/desktop/project7
samah@LAPTOP-IICCEV1N:/mnt/c/users/samah/desktop/project7$ ls
dataset.csv
samah@LAPTOP-IICCEV1N:/mnt/c/users/samah/desktop/project7$
```

create a file

Create an analysis file with (Cat). And make sure it is in the folder

```
samah@LAPTOP-IICCEV1N:/mnt/c/users/samah/desktop/project7$ cat>analysis.txt
samah@LAPTOP-IICCEV1N:/mnt/c/users/samah/desktop/project7$ ls
analysis.txt   dataset.csv
samah@LAPTOP-IICCEV1N:/mnt/c/users/samah/desktop/project7$
```

Positive and negative

I used Nf to find the second column and calculate the negative and positive cells

```
awk  -F "," '$NF=="negative"{count++}$NF=="positive"{num++};END{print "positive
count="num "\n""negative count="count;}' dataset.csv >>analysis.txt
```

unique words

have used the (awk)command.Initially I searched by unique words and then added a way to delete punctuation. Then save the result in the analysis file.

```
samah@LAPTOP-IICCEV1N:/mnt/c/users/samah/desktop/project7$ awk '{for(i = 1; i <= NF; i++) {{ gsub(/[[:punct:]]/, "", $0) } 1;a[$i]++}} END {for(k in a) if(a[k] == 1){counter++} print "Uni
NAME " are: "counter}' dataset.csv >>analysis.txt
samah@LAPTOP-IICCEV1N:/mnt/c/users/samah/desktop/project7$ head analysis.txt
positive count=1073
negative count=870
Unique words in file dataset.csv are: 125580
samah@LAPTOP-IICCEV1N:/mnt/c/users/samah/desktop/project7$
```

awk '{for(i = 1; i <= NF; i++) {{ gsub(/[[:punct:]]/, "", $0) } 1;a[$i]++}} END {for(k in a) if(a[k] == 1){counter++} print "Unique words in file " FILENAME " are: "counter}' dataset.csv >>analysis.txt
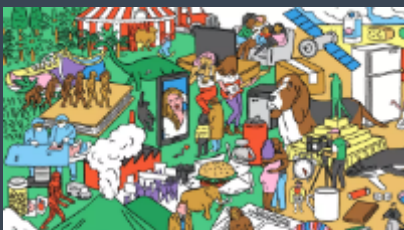
Sources

Count number of rows with certain columns equal to specific value

unix.stackexchange.com

Removing punctuation from txt file in Linux terminal using 'tr' and 'awk' commands

stackoverflow.com

How do I count the number of unique words in a file in Unix?

Answer (1 of 3): See this example - > [code]cat file.txt HI this is vipin kumar and testing for unique word in a file which contain vipin kumar two times....