Gregory Walsh
Week 3 Submission

Define metadata. Prepare a response piece discussing the usefulness, importance and legal implications of metadata.

Metadata is data about data. It's useful for determining data thats hot and cold and also who was the last person to view specific data. Its argued that data about data is not a threat to privacy but as the second article explains it definitely is. Metadata can be useful in company settings like the first article mentions with IT. Using Metadata to predict where data flows will be and how often is very powerful and used at Stocktons IT Help Desk currently. We look at Metadata on shared drives to know how frequent people look at the info and which ones to keep on Solid State Drives so they are able to be accessed faster. However, when Metadata is used by the government specifically call logs this start to get really scary. The Freedom Act needs to be strengthened.

Read the following article and prepare a response outlining how you would design a data-driven solution avoiding confirmation bias.

The key to a data-driven solution is to set the decision criteria first. Ask questions about what we need to do with the data. If we are looking for cost reduction, How much of a reduction? Then look at the data. Do not change your question to match the data. Don’t let emotions get involved with the decisions. Ask the important questions and stick to them.

Prepare a one paragraph response outlining why this is going to be your philosophy. "A complex system (or code) that works is invariably found to have evolved from a simple system (or code] that worked. The inverse proposition also appears to be true: a complex system (or code) designed from scratch never works and cannot be made to work. You have to start over, beginning with a simple system (or code). "

This will be my philosophy because you must always follow the start simple approach. You can always add fancy buttons and features but if the implementation does not solve the basic needs of the client then it will never be useful no matter how nice it looks. Make it work then make it look good. Don’t worry about new colors or personalization you can make. Focus on primary features needed and start from there. It will be much harder to look at a complex design and figure out why things are not working then looking at your simple implementation and work out all the bugs to make it perfect. Then it will in itself be beautiful and you can add on to it.

1. Using only file, wc, head, tail explore the file. Document all you can about the file and the data structure.

2. Using only the tr command delete all double quotes (") redirecting the output to a new file called students1.csv

3. Using only the commands specified create two new csv files called students2.csv and students3.csv from students1.csv.

students2.csv should only have the columns gender and parents level of education and all should be capitalized (commands: cut, tr)

students3.csv should be a single column of all the writing scores with no header (commands: tail, cut)

4. Using the paste command with the -s and -d options, and the bc command find the sum of all the writing scores in students3.csv

For your submission, show the first ten lines of your constructed new data files but, more importantly, document the Linux commands you used to get from the original file to your new data file.

StudentsPerformance.csv: ASCII text
WC: 1001 3188 72036 StudentsPerformance.csv
columns: Gender, Race/Ethnicity, Parental Level of Education, Lunch, Test Preperation on Course, Math Score, Reading Score, Writing Score
Race/Ethnicity have Grouping A-D

```
cat StudentsPerformance.csv | tr -d "\"" > students1.csv
cut -d ',' -f 1,3 students1.csv | tr "[a-z]" "[A-Z]" > students2.csv
cut -d ',' -f 8 students1.csv | tail -n +2 > students3.csv
paste -s -d'+' students3.csv | bc
```

SUM = 68054

students3.csv
74
88
93
44
75
78
92
39
67
50

students2.csv

```
GENDER,PARENTAL LEVEL OF EDUCATION
FEMALE,BACHELOR'S DEGREE
FEMALE,SOME COLLEGE
FEMALE,MASTER'S DEGREE
MALE,ASSOCIATE'S DEGREE
MALE,SOME COLLEGE
FEMALE,ASSOCIATE'S DEGREE
FEMALE,SOME COLLEGE
MALE,SOME COLLEGE
MALE,HIGH SCHOOL
```

StudentsPerformance.csv: ASCII text

WC: 1001 3188 72036 StudentsPerformance.csv

columns: Gender, Race/Ethnicity, Parental Level of Education, Lunch, Test Preparation on Course, Math Score, Reading Score, Writing Score
Race/Ethnicity have Grouping A-D
cat StudentsPerformance.csv | tr -d "\"" > students1.csv
cut -d ',' -f 1,3 students1.csv | tr "[a-z]" "[A-Z]" > students2.csv
cut -d ',' -f 8 students1.csv | tail -n +2 > students3.csv
paste -s -d'+' students3.csv | bc
SUM = 68054

students2.csv
GENDER,PARENTAL LEVEL OF EDUCATION
FEMALE,BACHELOR'S DEGREE
FEMALE,SOME COLLEGE
FEMALE,MASTER'S DEGREE
MALE,ASSOCIATE'S DEGREE
MALE,SOME COLLEGE
FEMALE,ASSOCIATE'S DEGREE
FEMALE,SOME COLLEGE
MALE,SOME COLLEGE
MALE,HIGH SCHOOL

students3.csv
74
88
93
44
75
78
92
39
67
50