SMonson 3.6 Task

1.

Graphical user interface, application, Word

Description automatically generated

Graphical user interface, application, Word

Description automatically generated

There is no duplicate data. If necessary, cleaning may depend on what authority the analyst has-Creating a view of unique values or deleting duplicates (if able) or the analyst may also use group by and distinct commands for analyzing data.

Non-Uniform Data-Film

Rectangle

Description automatically generated with medium confidence

To prevent non-uniform data entry, constraints on data type could be put in place so that only certain values could be entered. To fix non-uniform data, if the analyst has the authority, an UPDATE would be used.

Missing Data-Film

Graphical user interface, text, application

Description automatically generated

Missing Data-Customer

Graphical user interface, text, application

Description automatically generated

If large amounts of data are missing from a column, it may be necessary to ignore that column. If smaller amounts are missing, it may be possible to impute the missing data. (the analyst would need the authority to UPDATE to impute missing data)

2. Customer table-numeric columns

Graphical user interface, application, Word

Description automatically generated

Customer non-numeric columns

Graphical user interface, text, application, Word

Description automatically generated

Film non-numeric columns

Graphical user interface, application

Description automatically generated

Film numeric columns

Graphical user interface, text, application, Word

Description automatically generated

While both Excel and SQL can be used for data profiling, Excel is more suited to smaller data sets. There is a lot of repetitive steps with Excel to get the same results as SQL. SQL is faster and more efficient and works better with larger data sets.