15. Linux Basics for Data Science

Title:

Perform Basic Data Operations using Linux Commands.

Problem Statement:

You receive a compressed CSV file of 100MB that you must explore without using a GUI.

Tasks:

• Unzip the file using Linux commands.

• List the top 5 lines using command-line utilities.

• Search for rows containing a particular keyword.

• Move the processed file to a specific directory and change file permissions.

Deliverables:

• Commands used with brief explanation

• Snapshot of command outputs (can be terminal output pasted in file)

Unzipping a zip file

```
srijan@srijan:/mnt/c/csv$ ls -la
total 164
drwxrwxrwx 1 srijan srijan   4096 Jun  7 23:54 .
drwxrwxrwx 1 srijan srijan   4096 Jun  7 23:53 ..
-rwxrwxrwx 1 srijan srijan 166098 Jun  7 23:54 Dataset_Customer_Churn.zip
srijan@srijan:/mnt/c/csv$ unzip Dataset_Customer_Churn.zip
Archive:  Dataset_Customer_Churn.zip
  inflating: Dataset_Customer_Churn.csv
srijan@srijan:/mnt/c/csv$ ls -la
total 1120
drwxrwxrwx 1 srijan srijan   4096 Jun  7 23:55 .
drwxrwxrwx 1 srijan srijan   4096 Jun  7 23:53 ..
-rwxrwxrwx 1 srijan srijan 977501 Apr  6 01:34 Dataset_Customer_Churn.csv
-rwxrwxrwx 1 srijan srijan 166098 Jun  7 23:54 Dataset_Customer_Churn.zip
```

Show top 5 and bottom 5 lines

```
srijan@srijan:/mnt/c/csv$ head -n 5 Dataset_Customer_Churn.csv
customerID,gender,SeniorCitizen,Partner,Dependents,tenure,PhoneService,MultipleLines,InternetService,OnlineSecurity,OnlineBa
ckup,DeviceProtection,TechSupport,StreamingTV,StreamingMovies,Contract,PaperlessBilling,PaymentMethod,MonthlyCharges,TotalCh
arges,Churn
7590-VHVEG,Female,0,Yes,No,1,No,No phone service,DSL,No,Yes,No,No,No,No,Month-to-month,Yes,Electronic check,29.85,29.85,No
5575-GNVDE,Male,0,No,No,34,Yes,No,DSL,Yes,No,Yes,No,No,No,One year,No,Mailed check,56.95,1889.5,No
3668-QPYBK,Male,0,No,No,2,Yes,No,DSL,Yes,Yes,No,No,No,No,Month-to-month,Yes,Mailed check,53.85,108.15,Yes
7795-CFOCW,Male,0,No,No,45,No,No phone service,DSL,Yes,No,Yes,Yes,No,No,One year,No,Bank transfer (automatic),42.3,1840.75,N
o
```

```
srijan@srijan:/mnt/c/csv$ tail -n 5 Dataset_Customer_Churn.csv
6840-RESVB,Male,0,Yes,Yes,24,Yes,Yes,DSL,Yes,No,Yes,Yes,Yes,Yes,One year,Yes,Mailed check,84.8,1990.5,No
2234-XADUH,Female,0,Yes,Yes,72,Yes,Yes,Fiber optic,No,Yes,Yes,No,Yes,Yes,One year,Yes,Credit card (automatic),103.2,7362.9,N
o
4801-JZAZL,Female,0,Yes,Yes,11,No,No phone service,DSL,Yes,No,No,No,No,No,Month-to-month,Yes,Electronic check,29.6,346.45,No
8361-LTMKD,Male,1,Yes,No,4,Yes,Yes,Fiber optic,No,No,No,No,No,No,Month-to-month,Yes,Mailed check,74.4,306.6,Yes
3186-AJIEK,Male,0,No,No,66,Yes,No,Fiber optic,Yes,No,Yes,Yes,Yes,Yes,Two year,Yes,Bank transfer (automatic),105.65,6844.5,No
```

Search for word "Electronic check"

```
srijan@srijan:/mnt/c/csv$ grep "Electronic check" Dataset_Customer_Churn.csv
7590-VHVEG,Female,0,Yes,No,1,No,No phone service,DSL,No,Yes,No,No,No,No,Month-to-month,Yes,Electronic check,29.85,29.85,No
9237-HQITU,Female,0,No,No,2,Yes,No,Fiber optic,No,No,No,No,No,No,Month-to-month,Yes,Electronic check,70.7,151.65,Yes
9305-CDSKC,Female,0,No,No,8,Yes,Yes,Fiber optic,No,No,Yes,No,Yes,Yes,Month-to-month,Yes,Electronic check,99.65,820.5,Yes
7892-POOKP,Female,0,Yes,No,28,Yes,Yes,Fiber optic,No,No,Yes,Yes,Yes,Yes,Month-to-month,Yes,Electronic check,104.8,3046.05,Ye
s
5129-JLPIS,Male,0,No,No,25,Yes,No,Fiber optic,Yes,No,Yes,Yes,Yes,Yes,Month-to-month,Yes,Electronic check,105.5,2686.05,No
4183-MYFRB,Female,0,No,No,21,Yes,No,Fiber optic,No,Yes,Yes,No,No,Yes,Month-to-month,Yes,Electronic check,90.05,1862.9,No
8779-QRDMV,Male,1,No,No,1,No,No phone service,DSL,No,No,Yes,No,No,Yes,Month-to-month,Yes,Electronic check,39.65,39.65,Yes
6467-CHFZW,Male,0,Yes,Yes,47,Yes,Yes,Fiber optic,No,Yes,No,No,Yes,Yes,Month-to-month,Yes,Electronic check,99.35,4749.15,Yes
8665-UTDHZ,Male,0,Yes,Yes,1,No,No phone service,DSL,No,Yes,No,No,No,No,Month-to-month,No,Electronic check,30.2,30.2,Yes
6047-YHPVI,Male,0,No,No,5,Yes,No,Fiber optic,No,No,No,No,No,No,Month-to-month,Yes,Electronic check,69.7,316.9,Yes
5380-WJKOV,Male,0,No,No,34,Yes,Yes,Fiber optic,No,Yes,Yes,No,Yes,Yes,Month-to-month,Yes,Electronic check,106.35,3549.25,Yes
4080-IIARD,Female,0,Yes,No,13,Yes,Yes,DSL,Yes,Yes,No,Yes,No,No,Month-to-month,Yes,Electronic check,76.2,981.45,No
3714-NTNFO,Female,0,No,No,49,Yes,Yes,Fiber optic,No,No,No,No,No,Yes,Month-to-month,Yes,Electronic check,84.5,3906.7,No
7760-OYPDY,Female,0,No,No,2,Yes,No,Fiber optic,No,No,No,No,Yes,No,Month-to-month,Yes,Electronic check,80.65,144.15,Yes
8012-SOUDQ,Female,1,No,No,43,Yes,Yes,Fiber optic,No,Yes,No,No,Yes,No,Month-to-month,Yes,Electronic check,90.25,3838.75,No
1658-BYGOY,Male,1,No,No,18,Yes,Yes,Fiber optic,No,No,No,Yes,Yes,Month-to-month,Yes,Electronic check,95.45,1752.55,Yes
5067-XJQFU,Male,1,Yes,Yes,66,Yes,Yes,Fiber optic,No,Yes,Yes,Yes,Yes,Yes,One year,Yes,Electronic check,108.45,7076.35,No
0434-CSFON,Female,0,Yes,No,47,Yes,Yes,Fiber optic,No,No,Yes,No,Yes,Yes,Month-to-month,Yes,Electronic check,100.5,4707.1,No
5698-BQJOH,Female,0,No,No,9,Yes,Yes,Fiber optic,No,No,No,No,Yes,Yes,Month-to-month,No,Electronic check,94.4,857.25,Yes
5122-CYFXA,Female,0,No,No,3,Yes,No,DSL,No,Yes,No,Yes,Yes,Yes,Month-to-month,Yes,Electronic check,75.3,244.1,No
8627-ZYGSZ,Male,0,Yes,No,47,Yes,Yes,Fiber optic,No,Yes,No,No,No,No,One year,Yes,Electronic check,78.9,3650.35,No
0731-EBJQB,Female,0,Yes,Yes,52,Yes,No,No internet service,No internet service,No internet service,No internet service,No
internet service,No internet service,One year,Yes,Electronic check,20.4,1090.65,No
2673-CXQEU,Female,1,No,No,56,Yes,Yes,Fiber optic,Yes,Yes,Yes,No,Yes,Yes,One year,No,Electronic check,110.5,6139.5,No
0191-ZHSKZ,Male,1,No,No,30,Yes,No,DSL,Yes,Yes,No,No,Yes,No,Month-to-month,Yes,Electronic check,74.75,2111.3,No
5919-TMRGD,Female,0,No,Yes,1,Yes,No,Fiber optic,No,No,No,No,Yes,No,Month-to-month,Yes,Electronic check,79.35,79.35,Yes
8108-UXRQN,Female,0,Yes,Yes,11,No,No phone service,DSL,Yes,No,No,No,Yes,Yes,Month-to-month,No,Electronic check,50.55,565.35,
No
```

Move the file to a new directory and change permissions

```
srijan@srijan:/mnt/c/csv$ sudo mv Dataset_Customer_Churn.csv /mnt/c/csv1
srijan@srijan:/mnt/c/csv$ ls -la
total 164
drwxrwxrwx 1 srijan srijan    4096 Jun  8 00:03 .
drwxrwxrwx 1 srijan srijan    4096 Jun  8 00:03 ..
-rwxrwxrwx 1 srijan srijan 166098 Jun  7 23:54 Dataset_Customer_Churn.zip
srijan@srijan:/mnt/c/csv$ cd ..
srijan@srijan:/mnt/c$ cd csv1
srijan@srijan:/mnt/c/csv1$ ls -la
total 956
drwxrwxrwx 1 srijan srijan    4096 Jun  8 00:03 .
drwxrwxrwx 1 srijan srijan    4096 Jun  8 00:03 ..
-rwxrwxrwx 1 srijan srijan 977501 Apr  6 01:34 Dataset_Customer_Churn.csv
```

```
srijan@srijan:/mnt/c/csv1$ chmod 640 /mnt/c/csv1/Dataset_Customer_Churn.csv
srijan@srijan:/mnt/c/csv1$ ls -la
total 956
drwxrwxrwx 1 srijan srijan    4096 Jun  8 00:03 .
drwxrwxrwx 1 srijan srijan    4096 Jun  8 00:03 ..
-rwxrwxrwx 1 srijan srijan 977501 Apr  6 01:34 Dataset_Customer_Churn.csv
```