

Done !

Out[31]:

Analysis Values	
Mean	104608.560347
Mode	[86399.0]
Median	86399.0
Skew	3.952582
Kurtosis	37.02118
Standard deviation	81037.08123
Variance	6567008534.276449

On avearge users xDr sessions were 104608.560347 milliseconds long(1046.0 seconds)

Most of the user spent 86399.0 milliseconds on their sessions

The column is highly positively skewed(LeptoKurtic) with a skew value of 3.9 . This depicts a longer right tail, this also indicates that most of the values lie in the left hand side of the median value.(More people spent less time than the median duration)

The column has excess kurtosis of 37. This illustrates heavy tails on both ends which implies high number of outliers.

Out[33]:

Dispersion Values	
Q1	574.420000
Q2	863.990000
Q3	1324.300000
Std deviation	810.370812
IQR	749.880000
Max Value	18593.360000
Min Value	71.420000

The Quatile Statistics for the variable show that:

- 25% of the users spent less than 574.4 seconds on the internet
- 50% of the users spent 863 seconds or more on the internet
- 75% of users spent less than 1324.3 seconds and only 25% spent more than 1324.3 seconds there

The standard Deviation from the mean is 810 hence most of the data lies between +-810 sconds from the mean

- The max duration sent on a session was 18593.36

Out[66]:

Analysis Values	
Mean	430333100.706107
Mode	[15151281.0]
Median	431614977.0
Skew	-0.003225
Kurtosis	-1.202256
Standard deviation	244019110.62485
Variance	59545326350142760.0

The avrage combined data volume on The Gaming platforms is 430333100.706107 Bytes (~430 Mb) this is the current highest data usage this can be accredited to the intensive nature of live/ online game streaming which reuqires lots of data.

Creating DataFrame

Out[48]:

Dispersion Values	
Q1	2.431072e+08
Q2	4.558409e+08
Q3	6.657051e+08
Std deviation	2.441412e+08
IQR	4.225980e+08
Max Value	9.029696e+08
Min Value	7.114041e+06

- Only 25% of user sessions had a total bytes downloaded being less than 2.431072e+08 (~243 Mb)
- 50% of the user sessions downloaded more than and less than 4.558409e+08 (~455 Mb)
- Only 25% of user sessions downloaded more than 6.657051e+08 Bytes (~665 Mb) of data packets

The highest session downloaded 9.029696e+08 Bytes(902 Mb) worth of data

Analysis Values	
Mean	454643430.078967
Mode	[74126119.0]
Median	455840911.0
Skew	-0.003153
Kurtosis	-1.198363
Standard deviation	244141246.762547
Variance	59604948370770632.0

Analysis

The feature Total DL in Bytes had a mean of 454643430 bytes (454.6Mb) hence on avearge each user's session downloaded 454 mb worth of packets. With a majority of them downloading 74126119 (74.1 Mb) of data.

The data has a negative skew and hence has a longer left tail and most of the values lie on the right hand side of the median. The negative kurtosis suggest it having a flat tail and also little outliers in this case

With the Standard deviation at 244141246, it interpretes to user sessions downloading +- 244141246 around the mean of 454643430 bytes

Descriptive Analysis

Out[51]:

Dispersion Values	
Q1	3.322203e+07
Q2	4.114324e+07
Q3	4.903424e+07
Std deviation	1.127631e+07
IQR	1.581221e+07
Max Value	7.833131e+07
Min Value	2.866892e+06

1. 25% of the user sessions uploaded less than 3.322203e+07 bytes (33 Mb) of data
2. 50% of the user sessions uploaded more than 4.114324e+07 bytes (41 Mb) of data
3. Only 25% of the user sessions uploaded more than 4.903424e+07 Bytes (~50Mb) of data

The std deviation suggest that most of the user sessions uplaoded $+1.127631\text{e}+07$ (~11Mb) around the mean

The largest upload size was 7.833131e+07 (~78 Mb) significantly smaller than the largest download at 902 Mb

• **Fluorescence**

	Analysis Values
Mean	41121206.292107
Mode	[19362459.0]
Median	41143242.0
Skew	-0.002456
Kurtosis	-0.327465
Standard deviation	11276311.33903
Variance	127155197414725.65625

The average Total Upload was 41121206 Bytes (41 Mb) significantly smaller than the average downloads. It's safe to say people do more of downloading of content(data packets) than they do upload them.

The feature is almost normally distributed though it does have slight negative skewness of -.002 the distribution will form a bell curve with median, mean and mode being almost alike.

Creating DataFrame

Out[53]:

Dispersion Values	
Q1	9.322180e+05
Q2	1.826471e+06
Q3	2.727487e+06
Std deviation	1.035643e+06
IQR	1.795269e+06
Max Value	3.650861e+06
Min Value	1.563000e+03

1. 25% of the user sessions used less than 9.322180e+05 Bytes (~.9 Mb) of data both upload and download
2. More than 50% of the users used more than 1.826471e+06 Bytes (~18 Mb) data volume
3. Only 25% of customer used more than 2.727487e+06 Bytes (~ 27 Mb) data volume during their session

The user who did use the most data volume was 3.650861e+06 Bytes (~36 Mb) of data during their sessions

Out[52]:

	Analysis Values
Mean	1828250.208027
Mode	[154418.0]
Median	1826471.0
Skew	-0.002121
Kurtosis	-1.202178
Standard deviation	1035642.685868
Variance	1072555772791.031494

The social media platform had an average total data volume of 1828250 bytes (1.8 Mb) The distribution of this is almost normally distributed and the negative kurtosis depicts a flat tail and few outliers

Out[54]:

	Analysis Values
Mean	7807294.545573
Mode	[5240092.0]
Median	7812835.0
Skew	-0.010602
Kurtosis	-0.963043
Standard deviation	3516408.516617
Variance	12365128855735.333984

The average combined data volume for Google platform was 7807294 Bytes (~7.8 Mb). This is higher than Social Media platform reflecting that many users used Google as compared to Social Media

The data also takes a normal distribution though slightly negatively skewed.

Out[55]:

	Dispersion Values
Q1	4.943599e+06
Q2	7.812835e+06
Q3	1.068280e+07
Std deviation	3.516409e+06
IQR	5.739197e+06
Max Value	1.552878e+07
Min Value	4.033000e+04

1. 25% of the users had lower than 4.943599e+06 (~4 Mb) of data volume usage on Google
2. 50% of user had more/ less than 7.812835e+06 (~7 Mb) of data volume usage on Google
3. Only 25% of users had more than 1.068280e+07 (10 Mb) o data volume usage on Google

The highest recorded data volume usage on google was 1.552878e+07 Bytes (15 Mb) with the lowest being 40330 Bytes (~.04Mb)

Out[58]:

Dispersion Values	
Q1	1.359344e+06
Q2	2.263567e+06
Q3	3.159818e+06
Std deviation	1.071105e+06
IQR	1.800474e+06
Max Value	4.518036e+06
Min Value	8.359000e+03

1. 25% of cutomers used less than 1.359344e+06 Bytes (~1.3 Mb) of data volume on Email
2. 50% of cusomer used more than/ less than 2.263567e+06 (~ 2.2 Mb) of data volume on Emails
3. Only 25% of cutomers had more than 3.159818e+06 Bytes (~3.5 Mb) of data volume used on Email platforms

With the standard deviation of 1.0 Mb it interperates to data usage being +/- 1.0Mb around the mean

The highest data volume usage on Emails was 4.5 Mb with the lowest being at .008Mb on Emails

Out[57]:

Analysis Values	
Mean	2259102.30996
Mode	[648993.0]
Median	2263567.0
Skew	-0.006033
Kurtosis	-1.061701
Standard deviation	1071105.490272
Variance	1147266971291.749023

The average combined data usage of Email was 2259102 Bytes (~2.2 Mb) this shows that fewer data requests were done on Email when to compared to google.

Out[63]:

Analysis Values	
Mean	22628606.53838
Mode	[3192097.0]
Median	22635545.0
Skew	-0.000409
Kurtosis	-0.608492
Standard deviation	9260789.216948
Variance	85762216920746.9375

The average combined data volume usage on Netflix is 22628606.53838 Bytes (~22Mb) which is also similar to that of Youtub as both as video streaming services and are rather demanding

Out[64]:

Dispersion Values	
Q1	1.597946e+07
Q2	2.263554e+07
Q3	2.929044e+07
Std deviation	9.260789e+06
IQR	1.331099e+07
Max Value	4.519815e+07
Min Value	9.843200e+04

1. 25% of the customers use less than 15 Mb of data while accessing Netflix
2. 50% of the customers use more than 22 Mb of data packets while on the Netflix platform
3. 25% of the customers used more than 29 Mb

The highest recorded combined data volume usage on Netflix was 45 Mb while the least was .09 Mb

Out[61]:

Dispersion Values	
Q1	1.599846e+07
Q2	2.266177e+07
Q3	2.929260e+07
Std deviation	9.246769e+06
IQR	1.329414e+07
Max Value	4.519008e+07
Min Value	7.890300e+04

1. Less than 25% of people used lower than 1.599846e+07 Bytes (~15 Mb) of data volume while on Youtube
2. 50% of the people used more than 2.266177e+07 Bytes of data on Youtube platform
3. Only 25% of the population used more than 2.929260e+07 (~29 Mb) of data on youtube

The highest data volume usage on youtbe was at 45 Mb while the lowest was .07Mb

Done !

Out[60]:

Analysis Values	
Mean	22643482.6386
Mode	[18836800.0]
Median	22661770.0
Skew	-0.005465
Kurtosis	-0.604644
Standard deviation	9246769.484043
Variance	85502745891037.921875

The average combined data volume usage on Youtube platform is 22643482.6386 Bytes (22 Mb) this is higher than the previous platforms. This can be accrdited to it being a video streaming platform and hence would be data demanding.

Out[69]:

Analysis Values	
Mean	429365343.61892
Mode	[54037059.0]
Median	429986492.0
Skew	0.001256
Kurtosis	-1.196015
Standard deviation	243267262.402083
Variance	59178960956604120.0

The combined data volume average of other pilatforms used by the customers stands at 429365343.61892 Bytes (~429 Mb) of data.

With a platkurtic kurtosis it depicts flat tails with minimal outliers

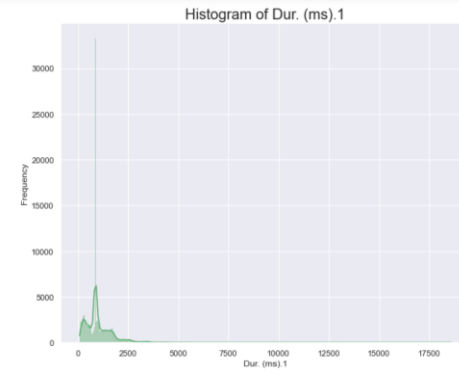
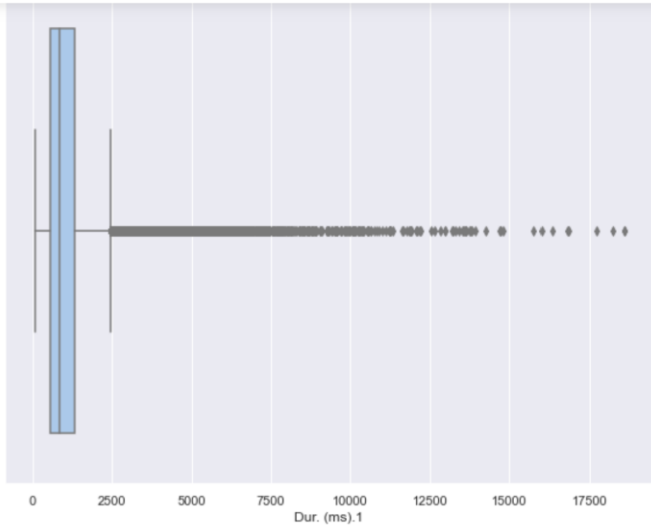
Creating DataFrame

Out[79]:

Dispersion Values	
Q1	2.185534e+08
Q2	4.299865e+08
Q3	6.399275e+08
Std deviation	2.432673e+08
IQR	4.213741e+08
Max Value	8.595200e+08
Min Value	1.490450e+05

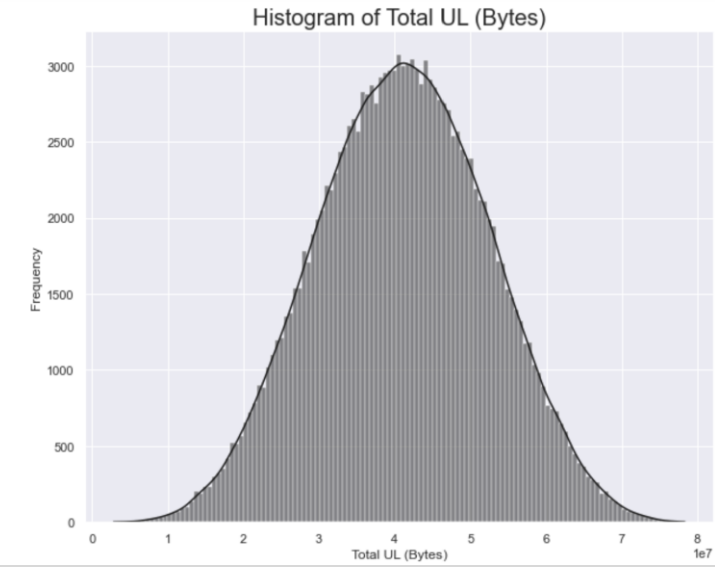
1. 25% of the customers used less than 218 Mb on other platforms
2. More than 50% of customers used more than 429 Mb on other platforms
3. 25% of users had more than 639 Mb of data volume on other platforms

The highest recorded data volume was 859 Mb with the least being .14Mb



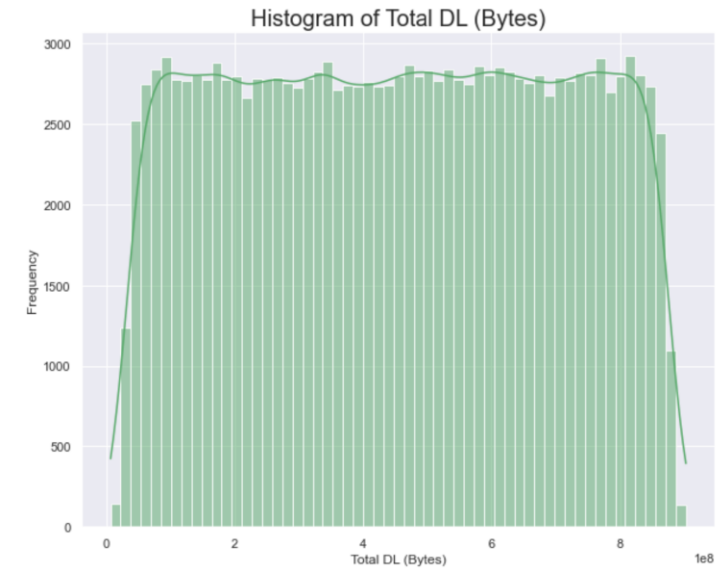
The visualization plot supports the figures in the histogram we see along right tail which is due to the positive skew nature of the column, this also indicates a heavy(long tail) which can also be seen in the presence of numerous outliers in the boxplot .

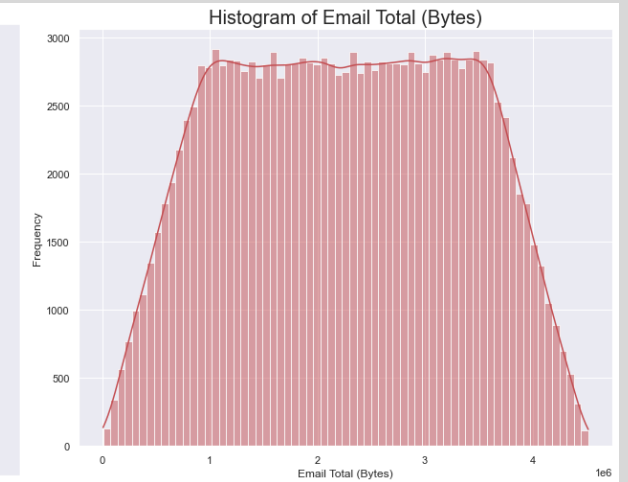
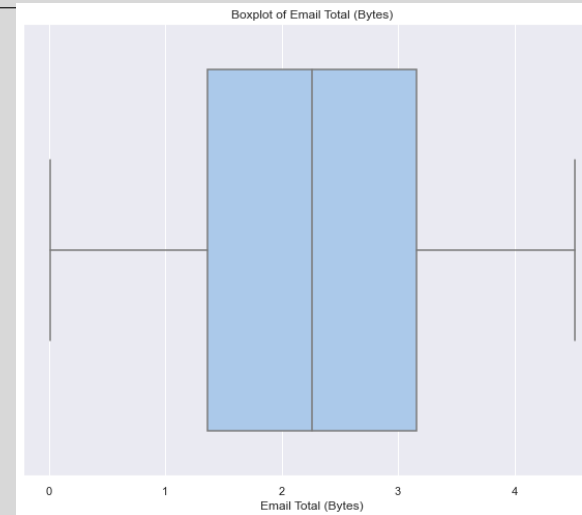
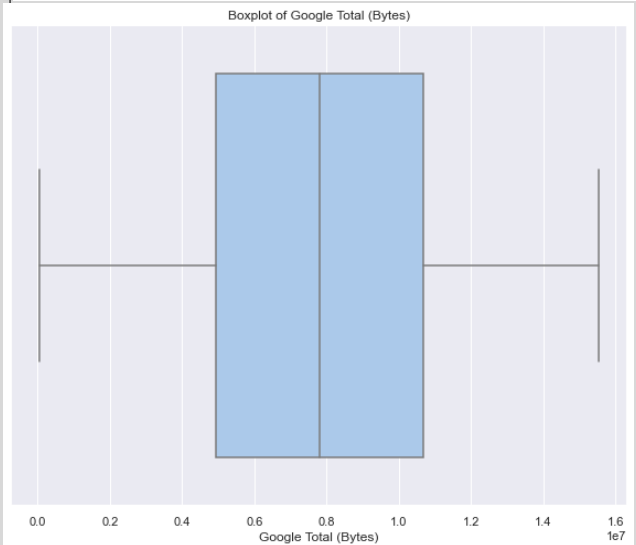
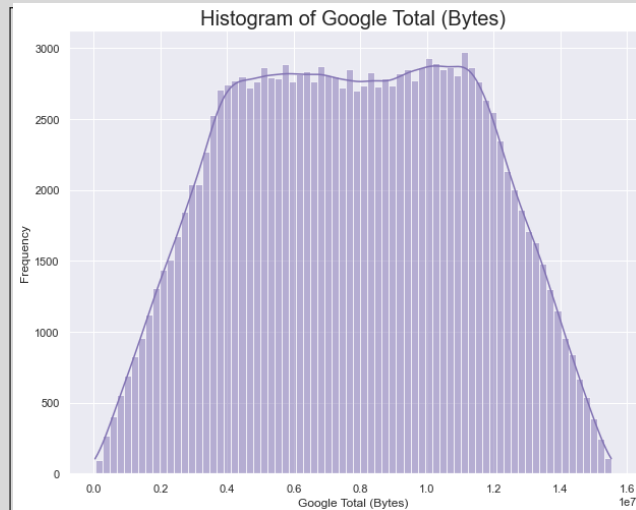
The outliers are mainly focused on the right hand side of the mean



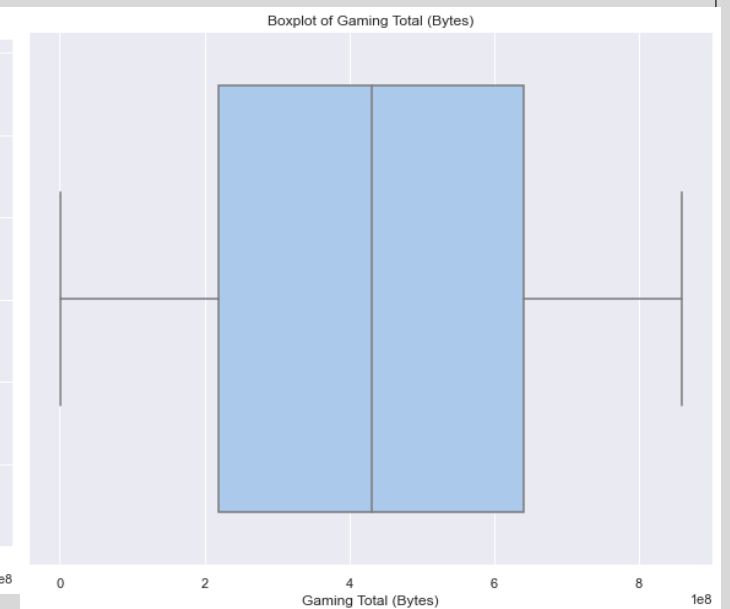
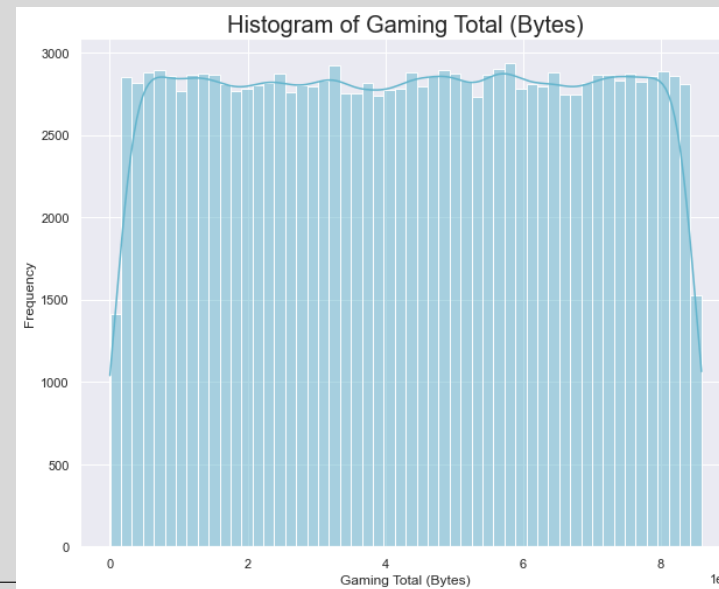
As seen with previous stats the distribution looks normal with a bell shape

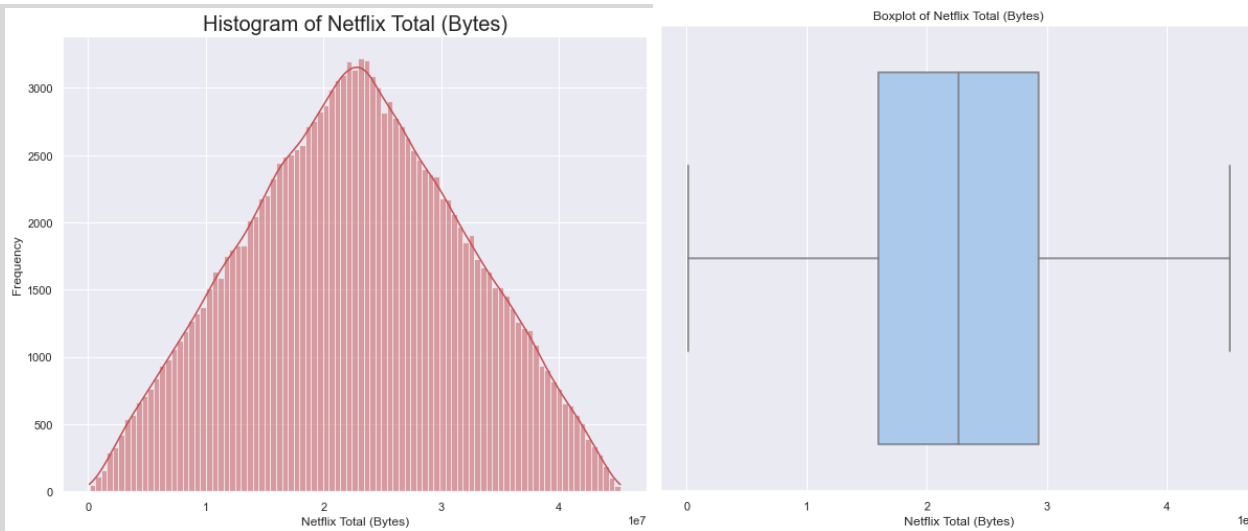
```
jupyterlab.analysis.visualization('Total DL (Bytes)', g)
```



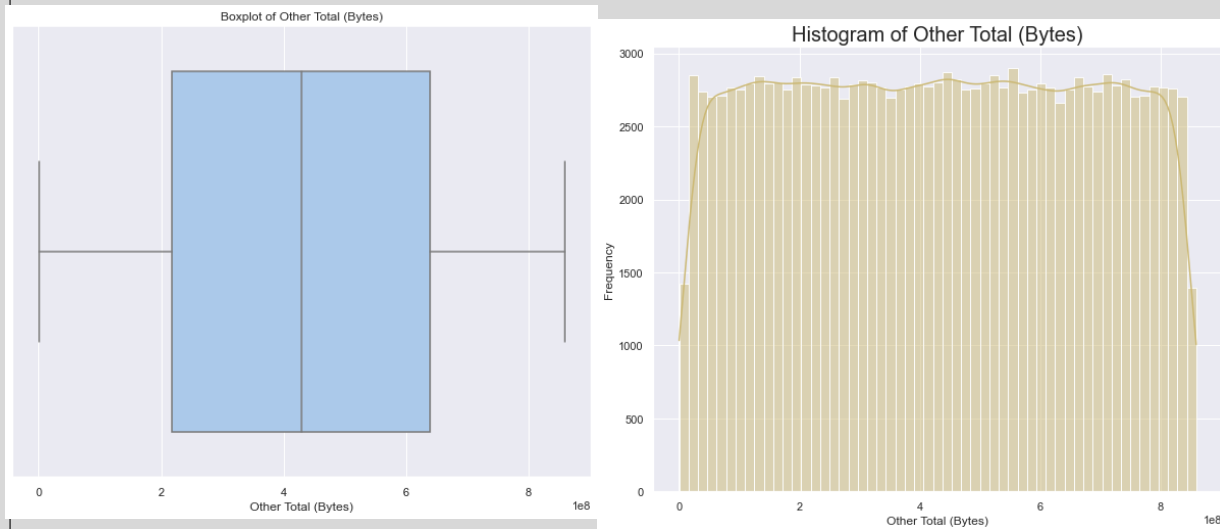


The distributions are almost normal they however do not take the conventional bell shape but due to their negative kurtosis have an almost flat top and end





The distribution on Netflix shows a spike around the median and shorter tails thus reduced outliers



This displays an even distribution where almost all values have equal number frequency

