Non graphical Univariate analysis

- This preliminary data analysis step focuses on four points, i.e. for mechanisms that you will want to examine. These include: measures of central tendency, i.e. the mean, the media and mode, measures of spread, i.e. variability, variants and standard deviation, the shape of the distribution, and the existence of outliers.

- Outliers were detected and treated using the Inter Quartile range of each features

Non graphical Univariate analysis contd

- It can be observed that average duration users spent per session is about 83s, the values ranges from 149.748s to as low as 22.33 s

- The average user spends about **474.52 MB** data

- Most of the data spent was on Gaming

Non graphical Univariate analysis contd

Measure	Dur. (s)	Total data (MBytes)	Total Google data (MBytes)	Total Social Media data (MBytes)	Total Email data (MBytes)	Total Youtube data (MBytes)	Total Netflix data (MBytes)	Total Gaming data (MBytes)	Total Other data (MBytes)
mean	83.13848	474.5219	7.441799	1.744182	2.151528	21.58043	21.58435	412.1221	409.3537
min	22.33	29.65293	0.051401	0.001491	0.020191	0.075248	0.093872	0.780932	0.867578
25%	86.169	377.3234	6.121752	1.333488	1.738368	18.79974	18.83158	315.5267	311.0325
50%	86.399	475.1882	7.454777	1.740233	2.151704	21.56043	21.59459	412.8209	409.8029
75%	86.399	572.7384	8.764079	2.158618	2.568119	24.34952	24.37053	510.6865	505.6405
max	149.748	908.5096	14.79847	3.481732	4.304484	43.0664	43.02059	819.3996	819.2574

Table 1: Univariate description of the dataset

Graphical Univariate Analysis

- These analysis involves the graphical presentation of the distribution of each features in the dataset

- Figure 1 shows the distribution plot the Total duration of session, the mean and the percentiles are clustered around a point

- Figure 2 shows the distribution plot of the Total Email data, the distribution has more data greater than the mean
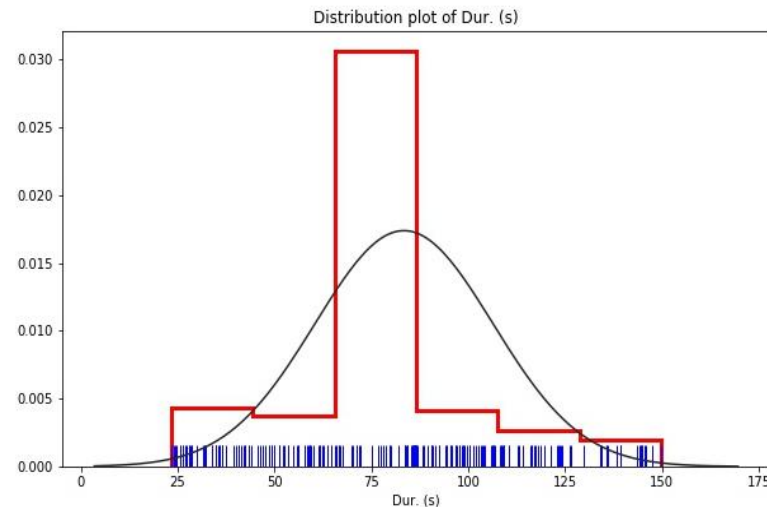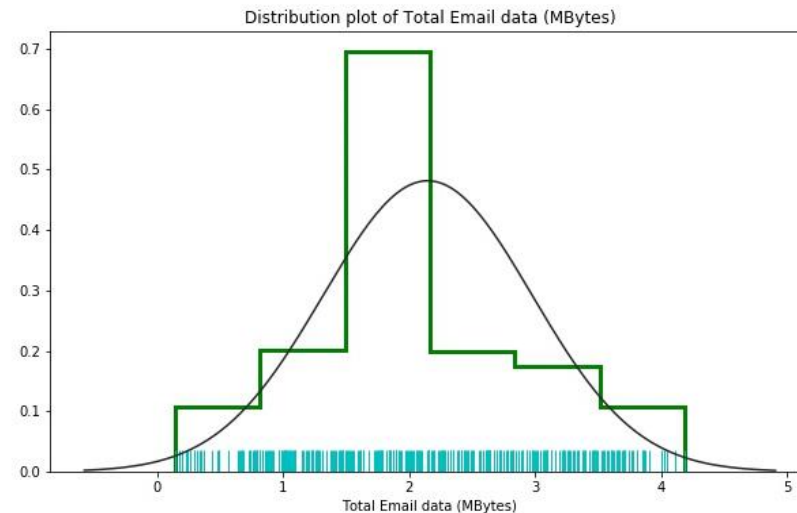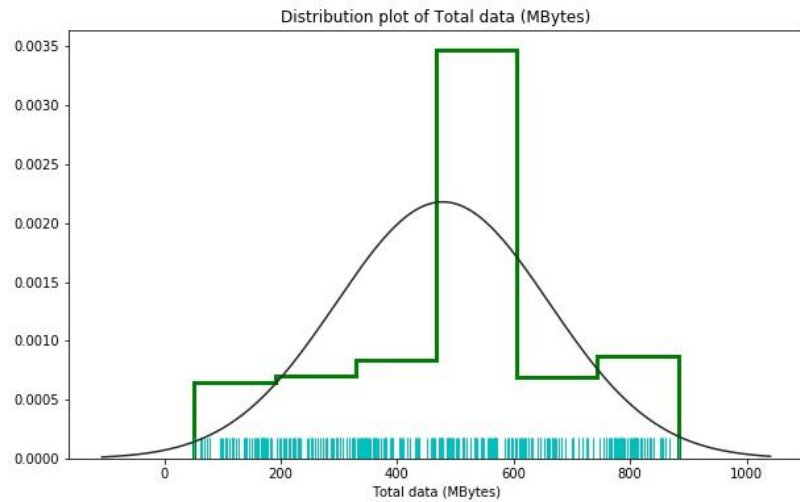

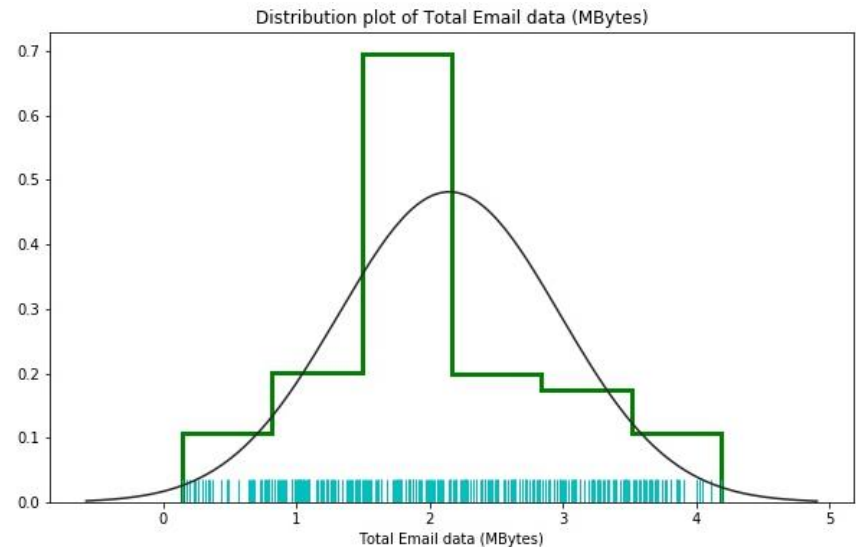Figure1: Distribution plot of total duration(s)

Figure1: Distribution plot of total email data (MB)

Graphical Univariate Analysis cont.

Figure 3: Distribution plot of the Total Data usage

The plot is not normally distributed, the plot shows a wide spread amount of data usage over similar ranges

Figure 3: Distribution plot of the Total Email data usage

The plot is not normally distributed, the plot shows that, it has wide spread.

Graphical Univariate Analysis cont.
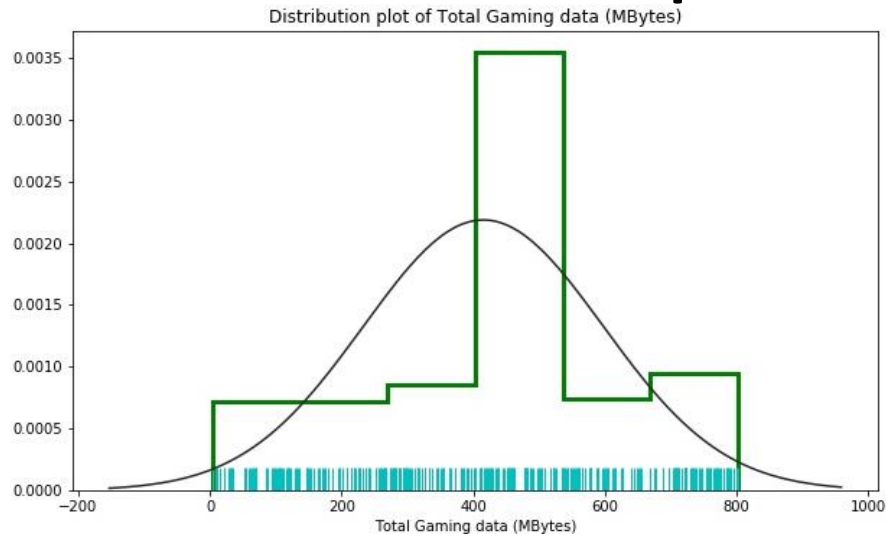
Figure 5:Distribution plot of the Total Gaming Data usage

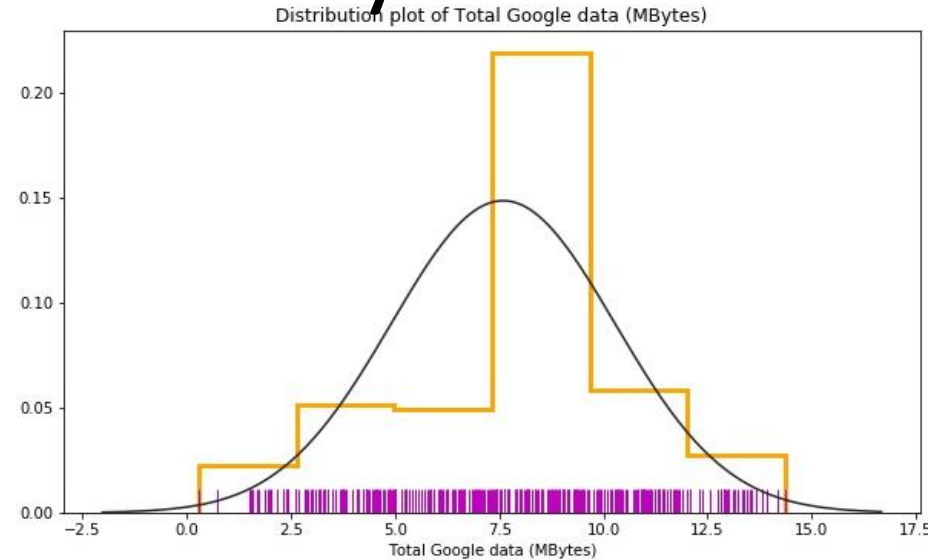The plot shows deviation from the normal fitted curve, it shows a spread away from the mean towards left

Figure 6:Distribution plot of the Total Google Data usage

The above plot shows deviation from the normal fitted curve, it shows a spread around a data point that is skewed to the left

The plot shows deviation from the normal fitted curve, it shows there are more data less than the mean

Figure 7:Distribution plot of the Total Netflix Data usage

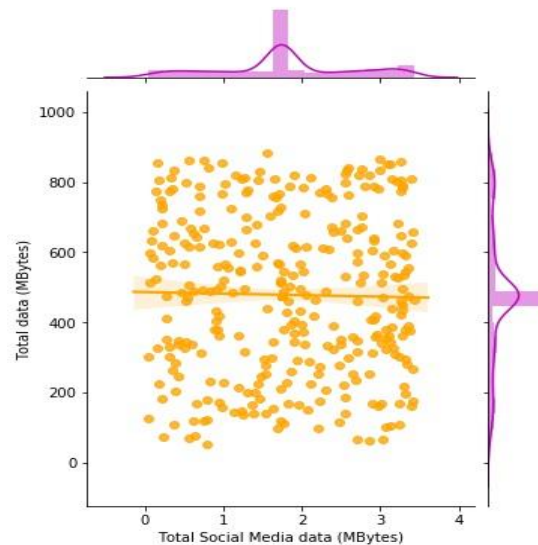# Graphical Bivariate Analysis

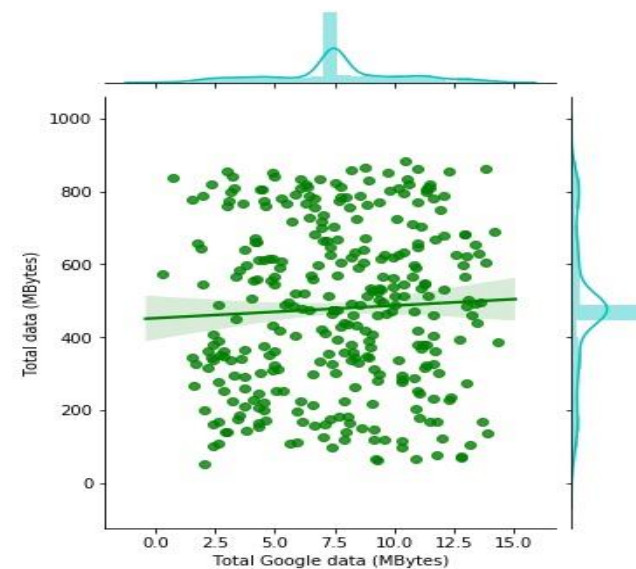
Plot of Total Data against Total Other data (MBytes)
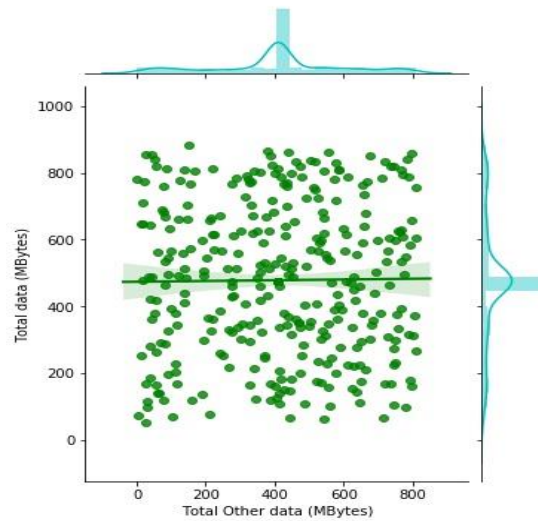
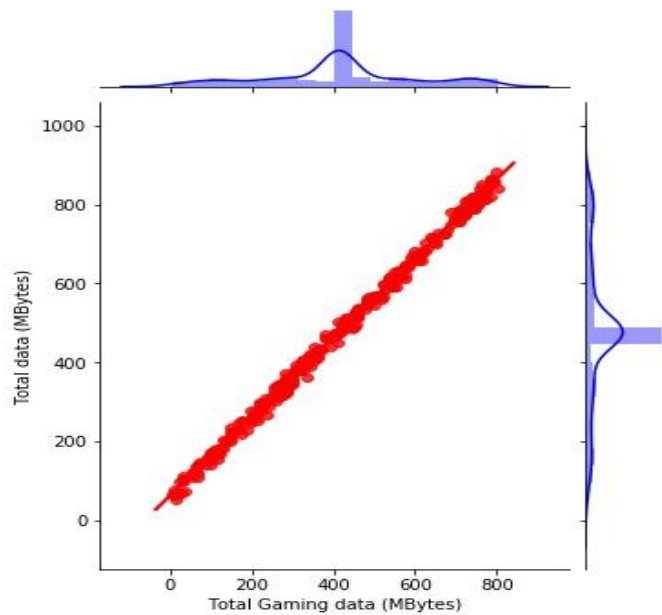Plot of Total Data against Total Netflix data (MBytes)
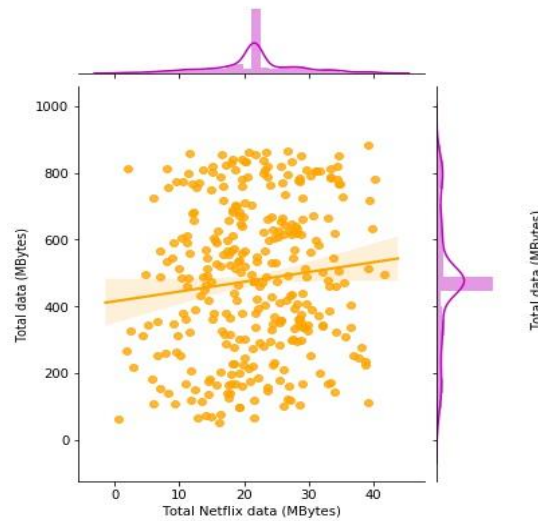
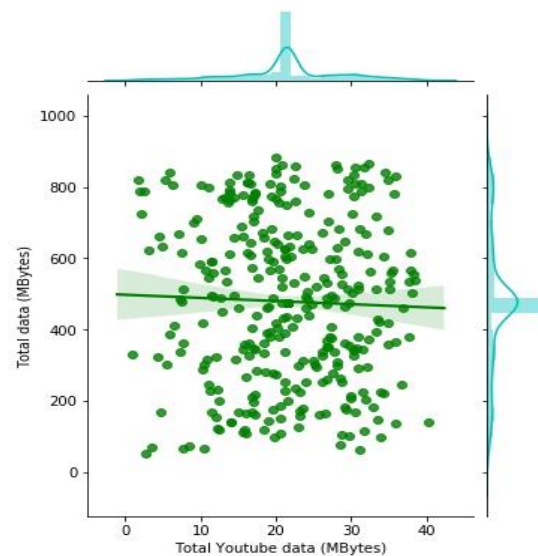Plot of Total Data against Total Social Media data (MBytes)

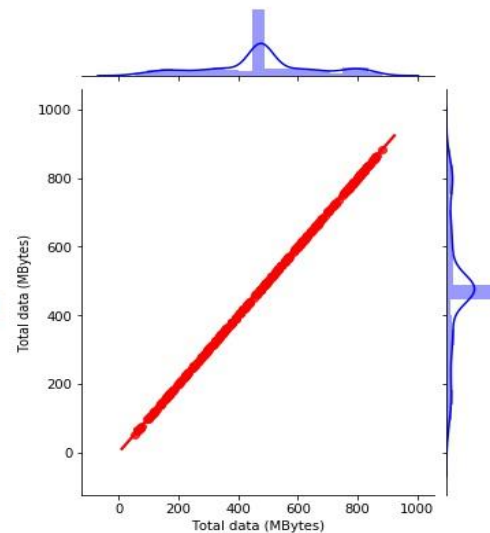Plot of Total Data against Total Google data (MBytes)

Plot of Total Data against Total Gaming data (MBytes)

Plot of Total Data against Total Youtube data (MBytes)

Plot of Total Data against Total data (MBytes)

Plot of Total Data against Total Email data (MBytes)