Exploratory Data Analysis (EDA)

Data preparation

As the first step Mobile prices data checked for the missing values and none were found. In the Mobile prices data set, Mobile phones without a front camera were given 0 value for the amount of megapixels in the front camera.

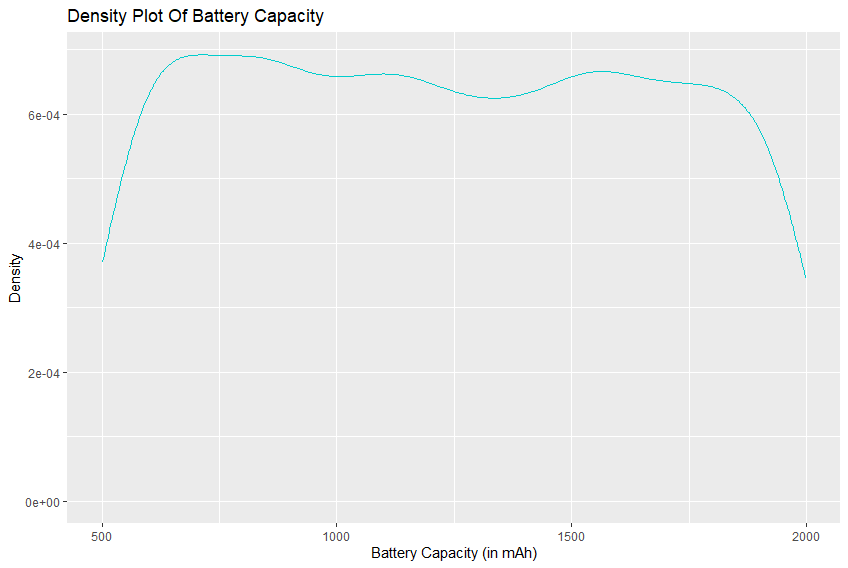
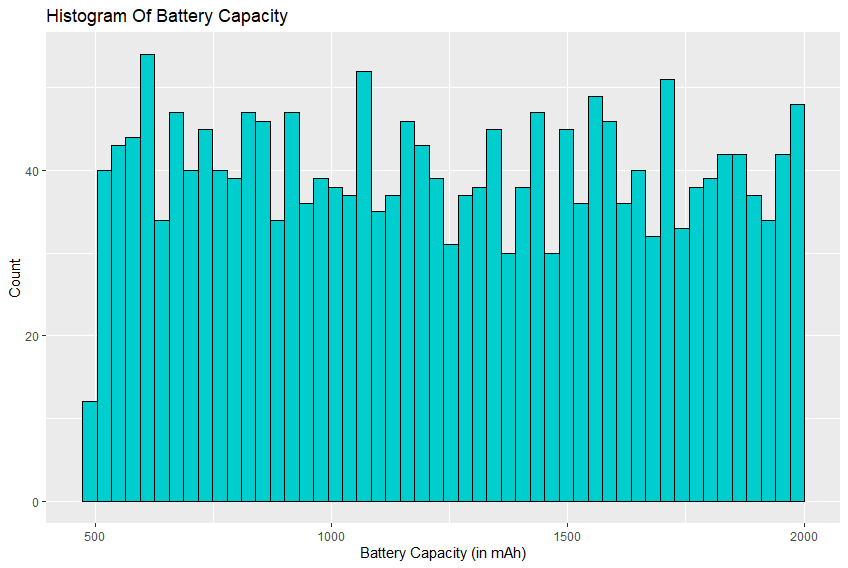
When processing the dataset a new column was added to represent the availability of the front camera if the front camera was available coded 1,else coded 0. Similarly in the Mobile prices data set, Mobile phones without a primary camera were given 0 value for the amount of megapixels in the primary camera. When processing the dataset a new column was added to represent the availability of the primary camera if the primary camera was available coded 1,else coded 0.

When calculating the statistics of the amount of megapixels of the camera of mobile phones, the mobile phones without a camera were filtered out instead of taking, not having a camera as ‘0 megapixel camera’ in the original data set.

The data set was tested for any duplicated values mobile phones with all equal characteristics, there were none. The coded categorical values are decoded and price ranges are ordered by factoring.

Univariate Data Analysis

**Quantitative data.**

1.Univariate analysis for battery power

The image above shows the histogram of battery capacity and the density plot of battery capacity; these two graphs show that the distribution is uniform. Battery power is uniformly distributed.

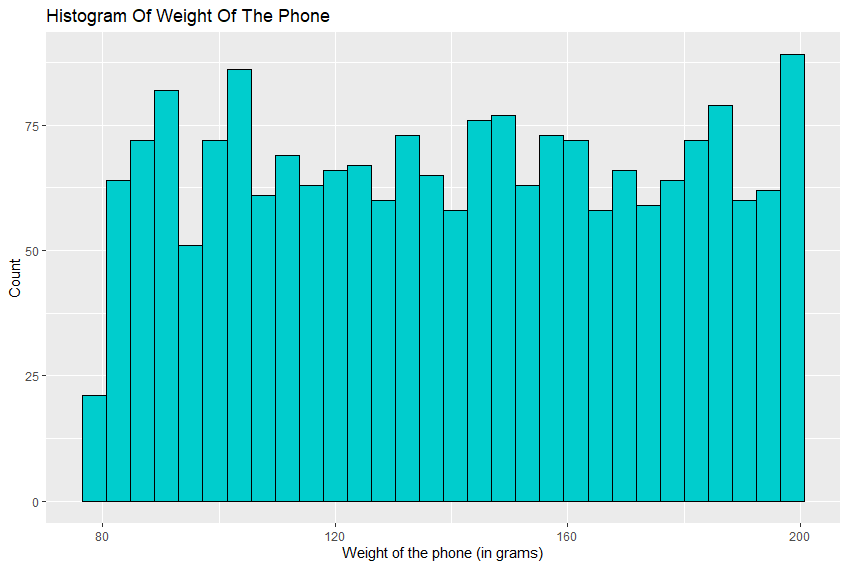
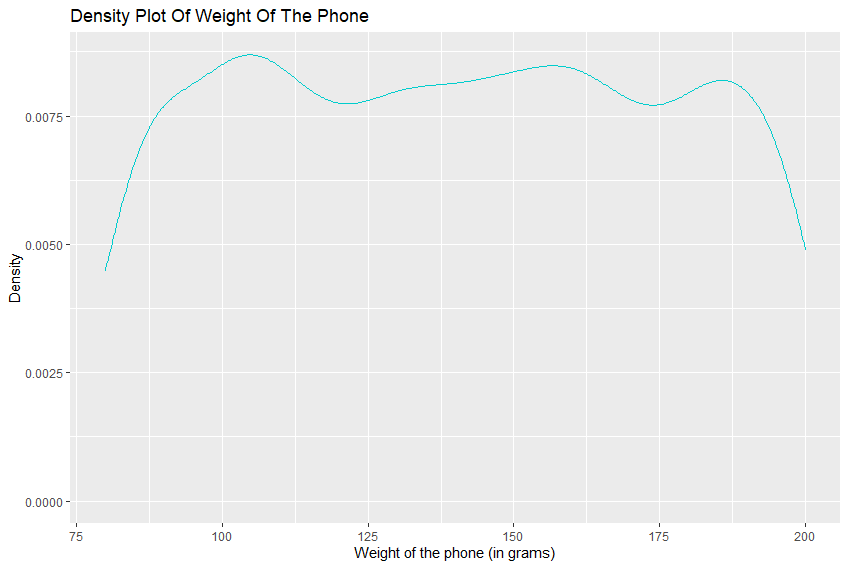
Summary Statistics: Data : Battery power (N = 2,000)

minimum : 501.00

median (IQR) : 1,226.00 (851.75, 1,615.25)

mean (sd) : 1,238.52 ± 439.42

maximum : 1,998.00

2.Univariate analysis for the weight of the phone.

The image above shows the histogram of the weight of the phone and density plot of battery capacity; these two graphs show that the distribution is uniform. Weight of the phone is uniformly distributed.

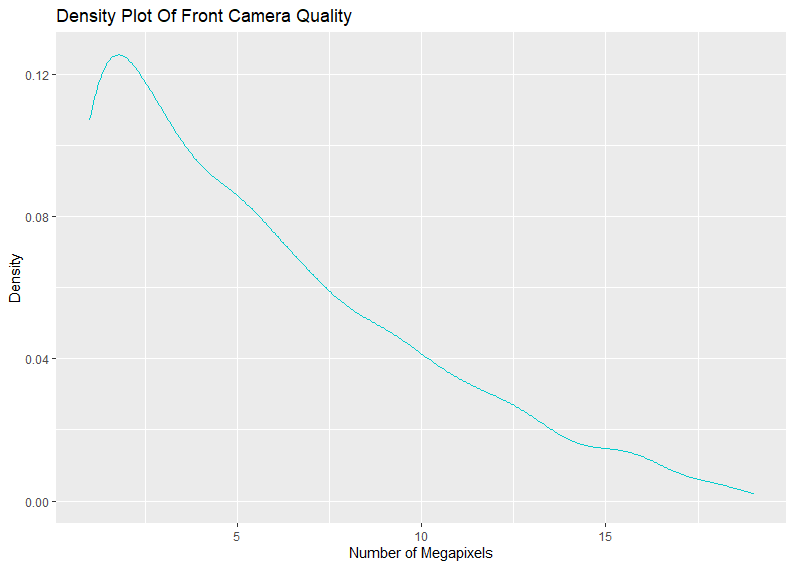
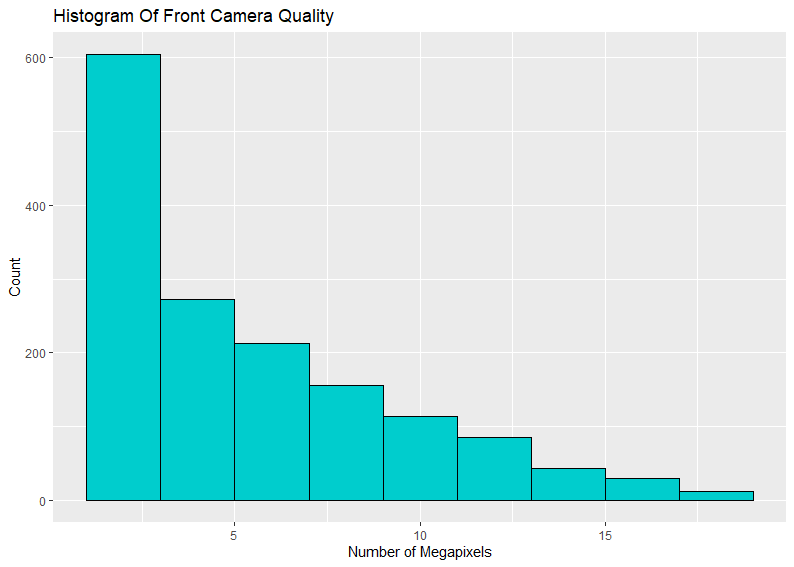
Summary Statistics: Weight of the mobile(N=2000)

minimum : 80.00

median (IQR) : 141.00 (109.00, 170.00)

mean (sd) : 140.25 ± 35.40

maximum : 200.00

3.Univariate Analysis for the Front Camera Quality (Megapixcel amount).

The image above shows the histogram and the density plot of the front camera quality (amount of megapixels). These two graphs show that the distribution is positively skewed (skewed to the right). Most of the phones have relatively low megapixel value front cameras and very few phones have high megapixel front cameras.

Summary statistics: Front Camera Quality(N = 1,526)

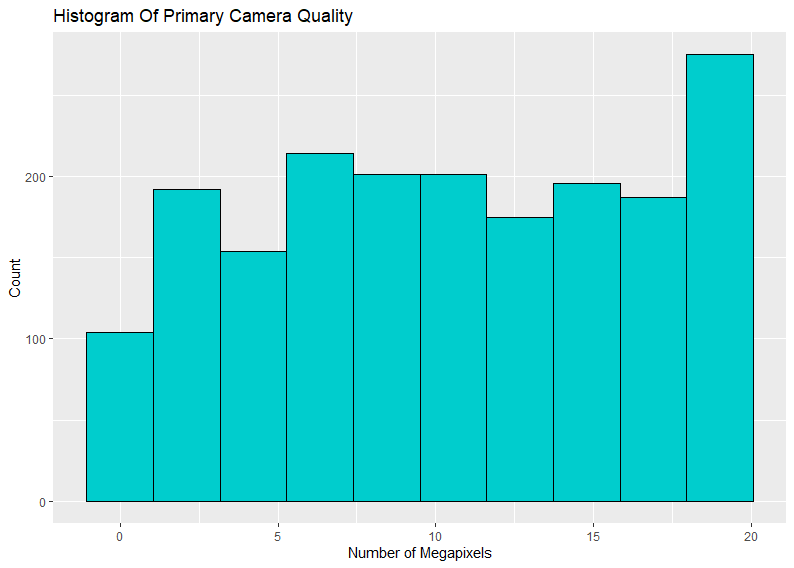
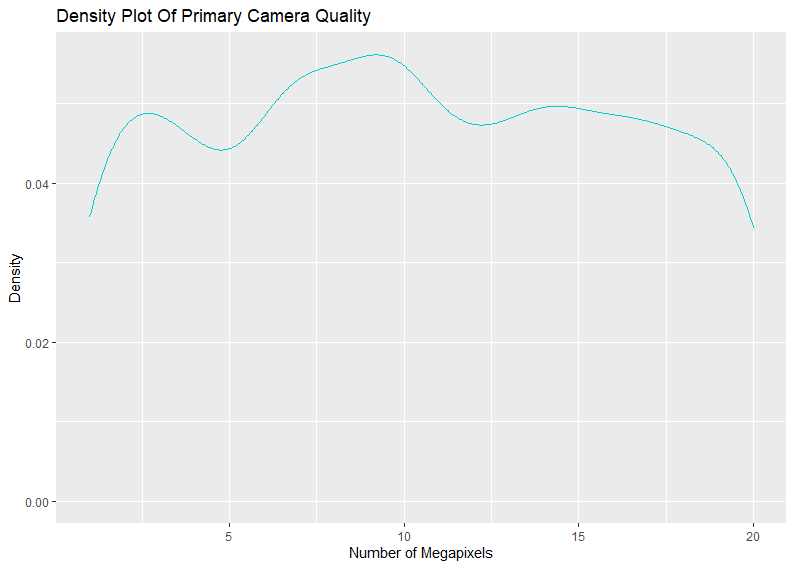
minimum : 1.00

median (IQR) : 5.00 (2.00, 8.00)

mean (sd) : 5.65 ± 4.14

maximum : 19.00

4.Univariate Analysis for Primary camera quality

The image above shows the histogram and the density plot of the Primary camera quality (amount of megapixels). These two graphs show that the camera quality is uniformly distributed.

Summary statistics: Primary Camera Quality(N = 1,899)

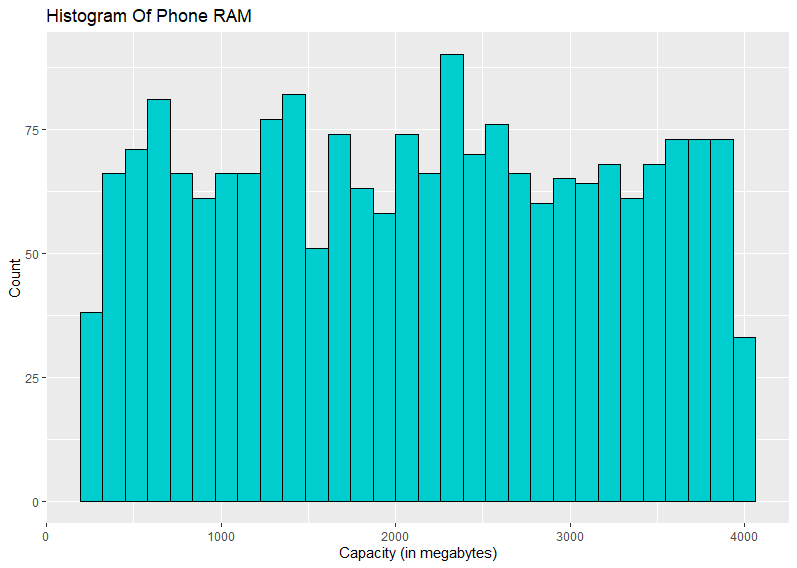
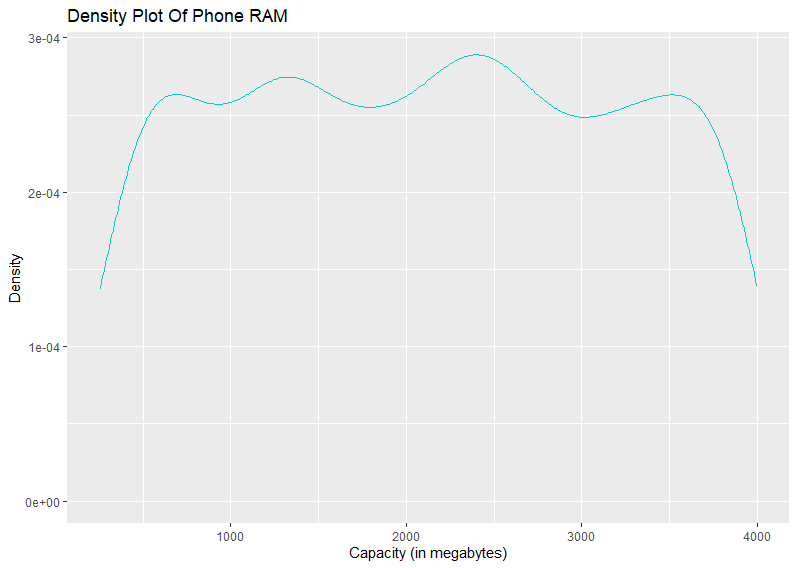
minimum : 1.00

median (IQR) : 10.00 (6.00, 15.00)

mean (sd) : 10.44 ± 5.76

maximum : 20.00

5.Univariate Analysis for RAM Capacity

The image above shows the histogram and the density plot of the RAM Capacity. These two grapes show that the distribution is uniform.

Summary Statistics: RAM(N=2000)[in megabytes]

minimum : 256.00

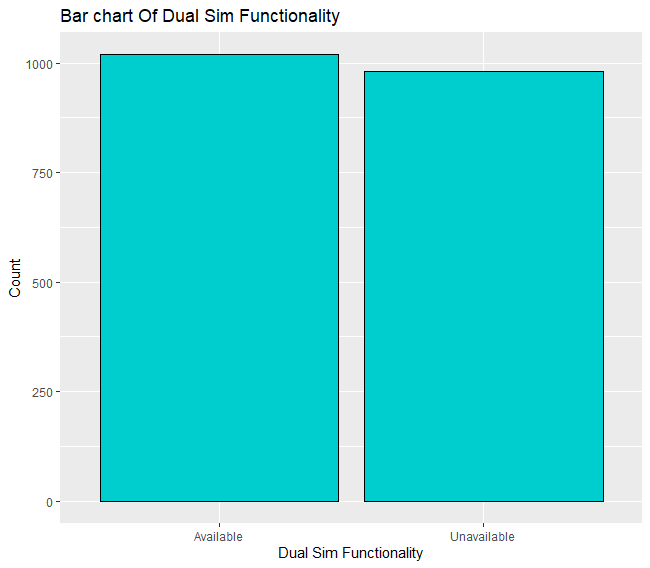
median (IQR) : 2,146.50 (1,207.50, 3,064.50)

mean (sd) : 2,124.21 ± 1,084.73

maximum : 3,998.00

**Categorical Data**

6.Univariate Analysis for Dual Sim Function

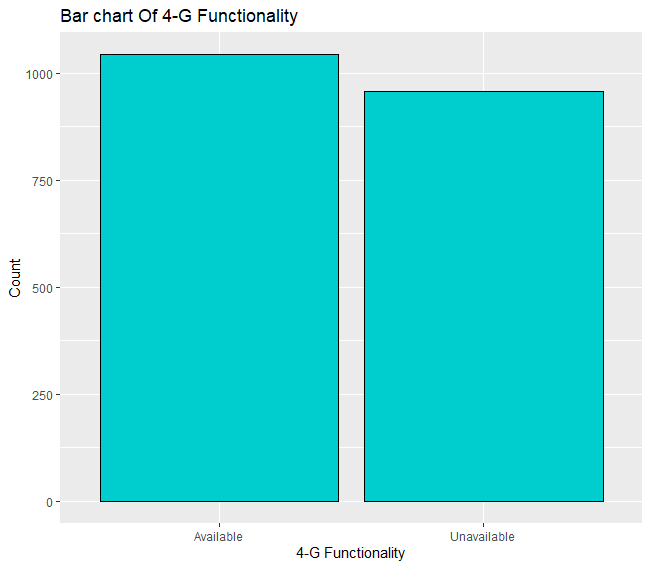


The image above shows the bar chart of dual sim functionality

Summary: With dual sim : 1019

Without dual sim : 981

7.Univariate Analysis for 4-G Function.

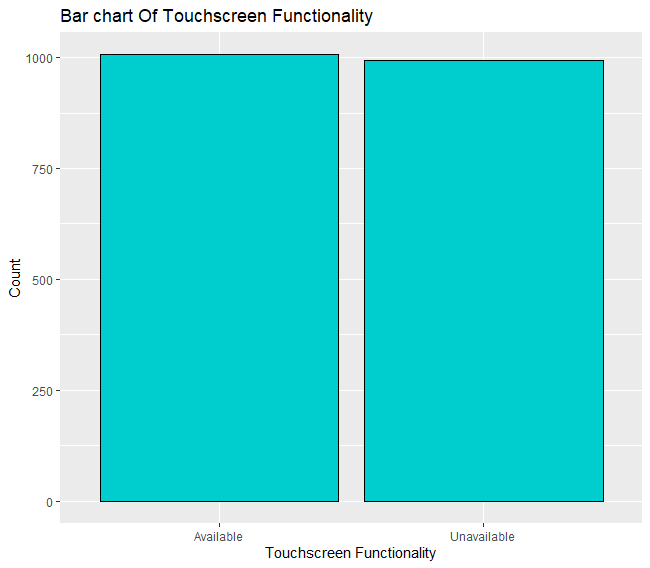


The image above shows the chart of 4-G functionality

Summary: With 4-G functionality : 1043

Without 4-G functionality : 957

8.Univariate Analysis for Touchscreen Function.

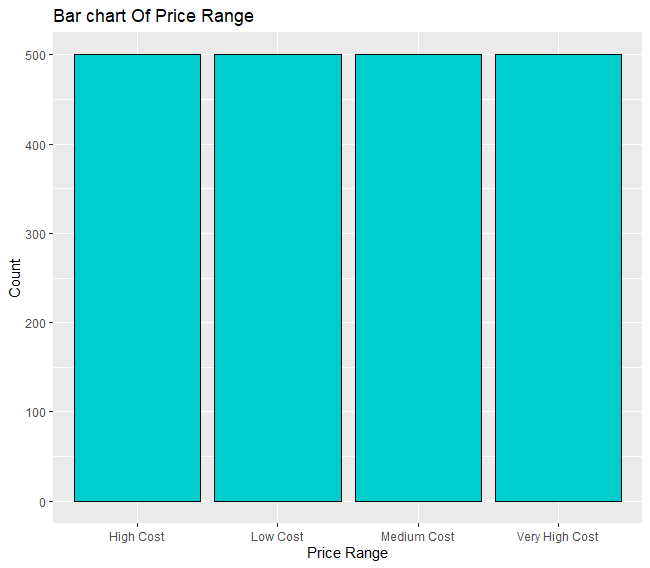


The image above shows the chart of Touchscreen Function

Summary: With Touchscreen Function functionality : 1006

Without Touchscreen Function functionality : 994

9.Univariate Analysis for Price Range.



The image above shows the bar chart of Price Range

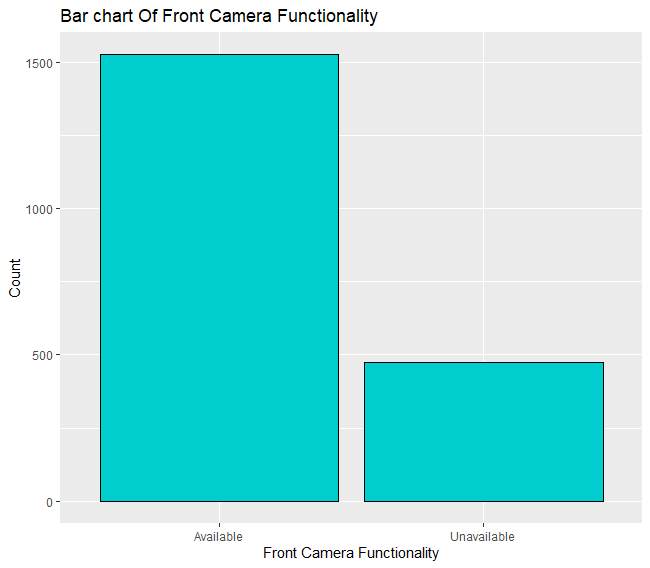
Summary: Low cost:500

Medium cost:500

High cost:500

Very High cost:500

10.Univariate Analysis for Front Camera Functionality.

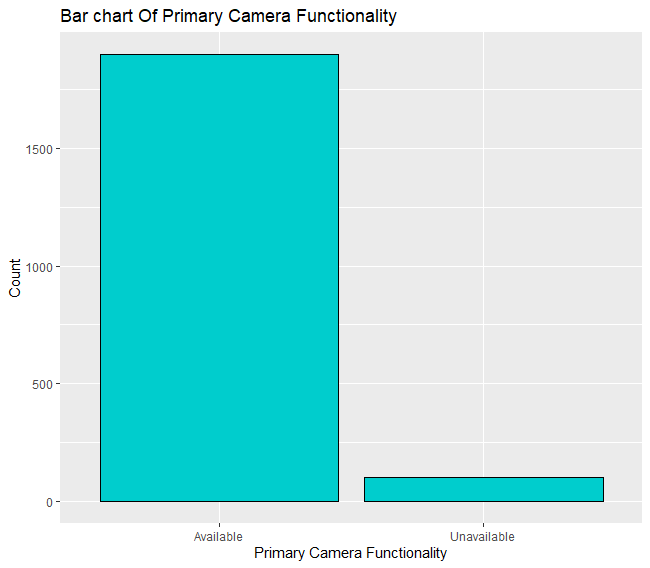


The image above shows the bar chart of Front Camera Functionality.

Summary: Available:1526

Unavailable:474

11.Univariate Analysis for Primary Camera Functionality.



The image above shows the bar chart of Front Camera

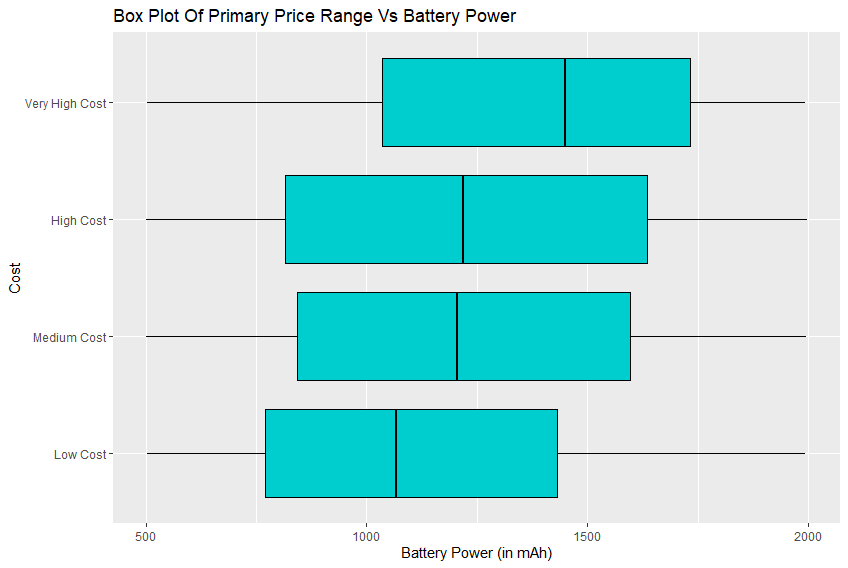
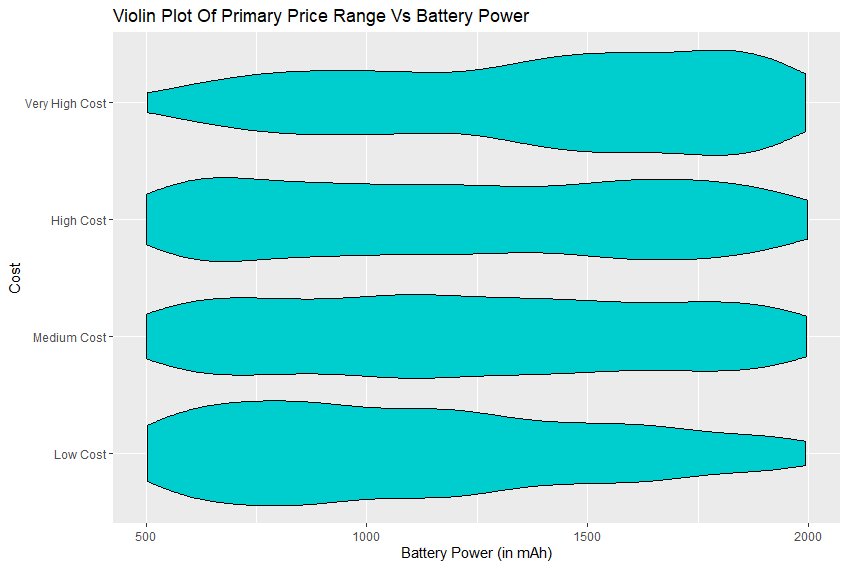
Functionality.

Summary: Available:1899

Unavailable:101

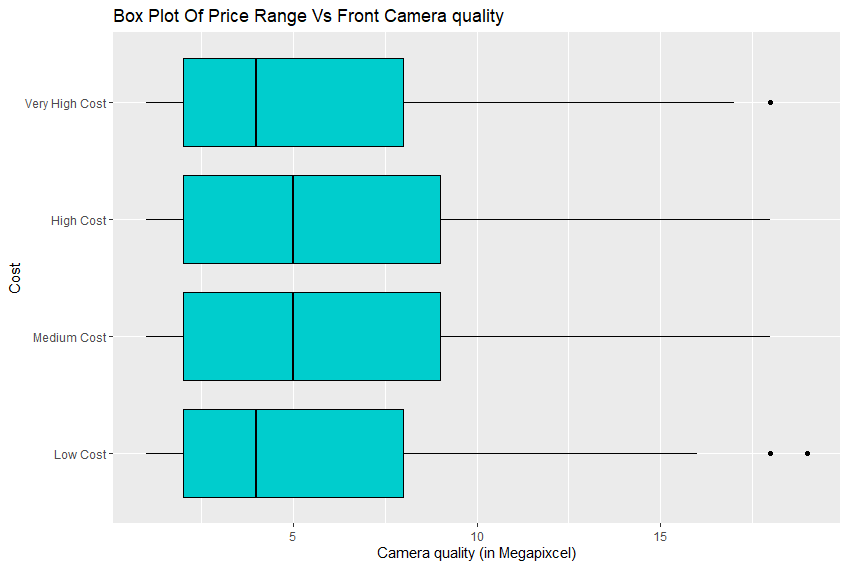
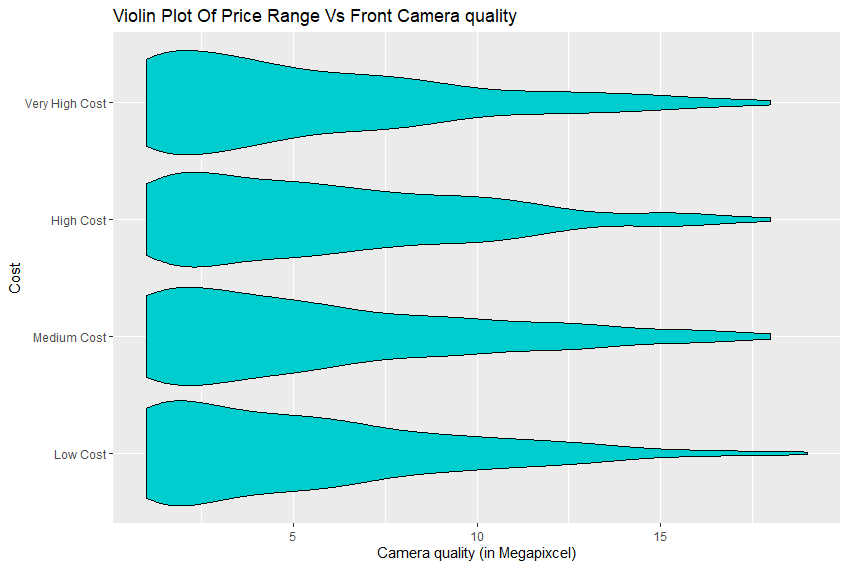
Bivariate Data Analysis

01.Price Range Vs Battery Power.



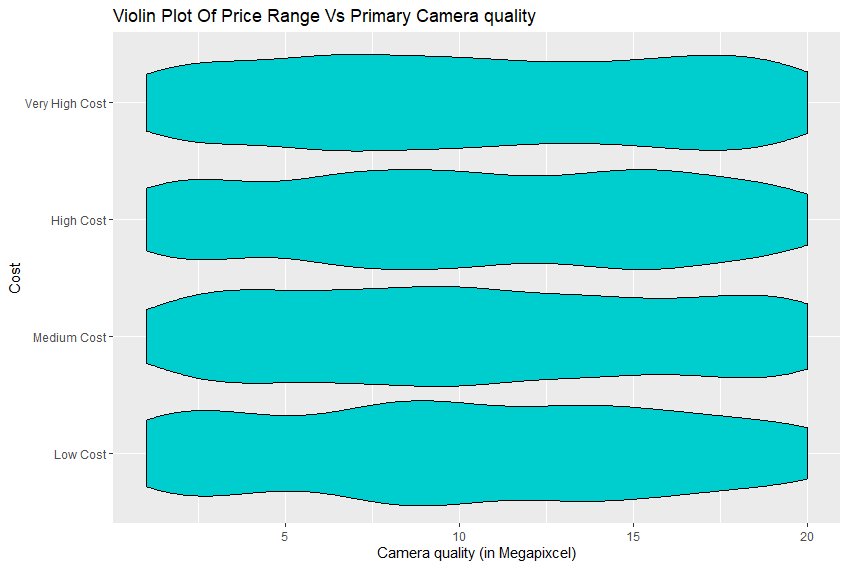
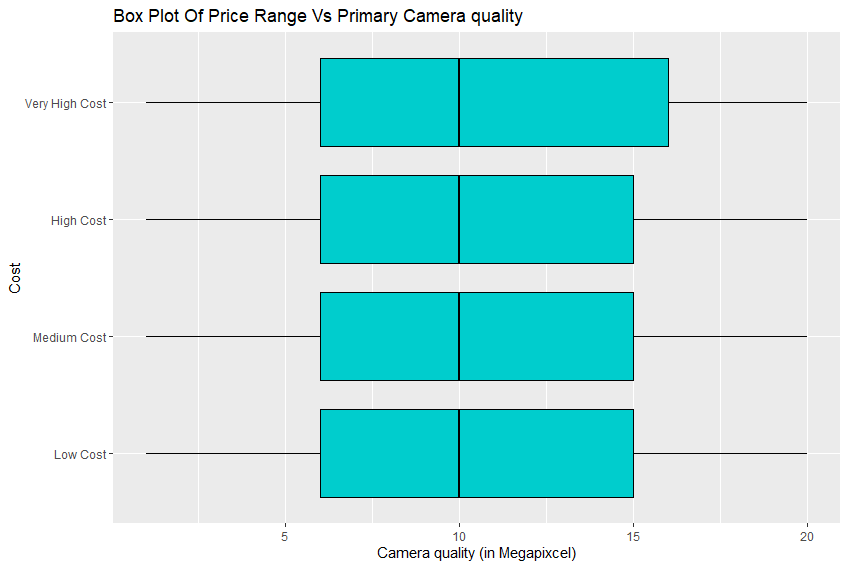
The image above shows the boxplot and the violin plot of the price range vs battery power. These two grapes show that there is a clear increase of battery power as cost increases. Spread of the distribution is higher in high cost and medium cost than that of low cost and very high cost.

02.Price Range Vs Front Camera Quality.

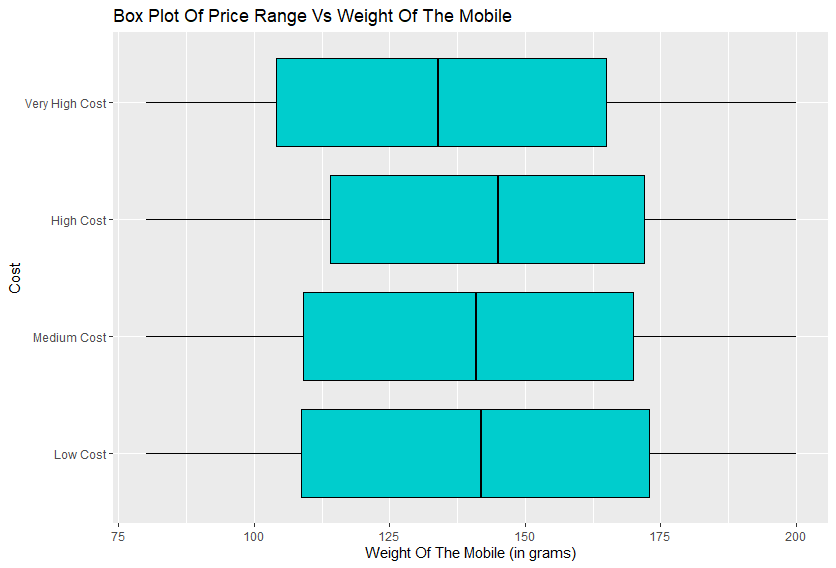
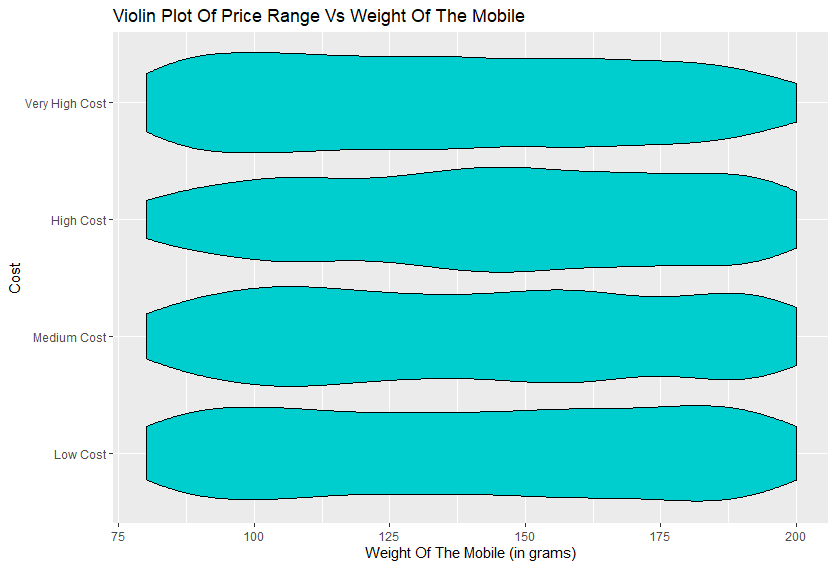


The image above shows the boxplot and the violin plot of the price range vs front camera quality. These two graphs don't show that there is an increase of front camera quality with the price. All four price ranges show a similar spread -a positively skewed (skewed to the right) distribution- , which is similar to the aggregate distribution of all price ranges shown earlier. There are 3 outliers in the graph which are likely to be caused by mobile phones built for photography.

03.Price Range Vs Primary Camera Quality.

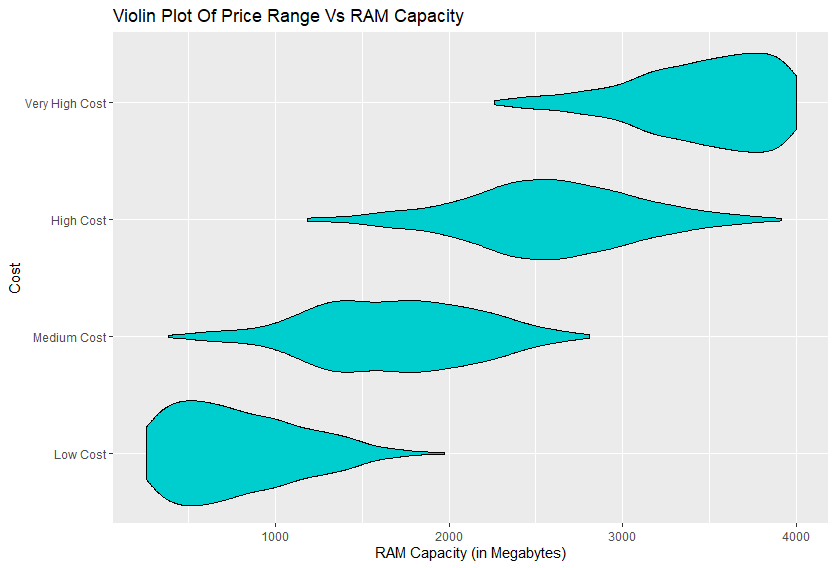
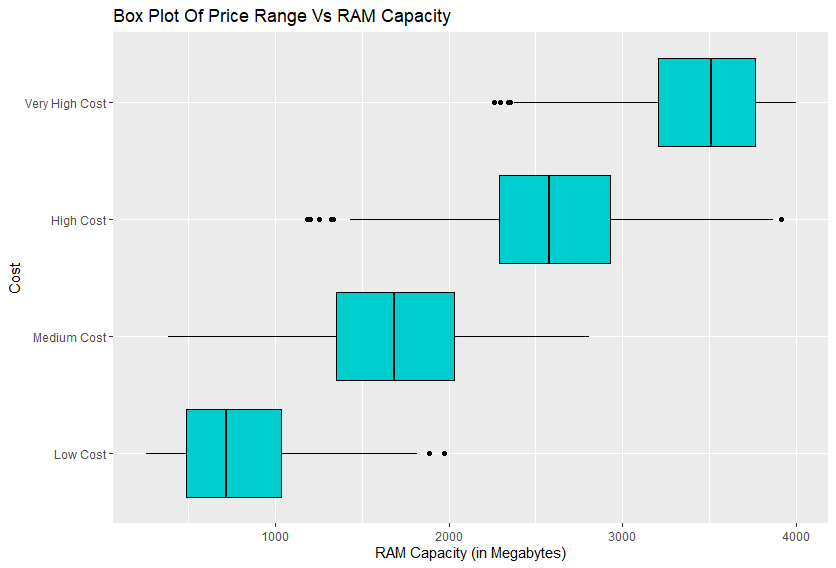


The image above shows the boxplot and the violin plot of the price range vs primary camera quality. These two graphs don't show that there is an increase of primary camera quality with the price. All four price ranges show a similar spread -a uniform distribution- , which is similar to the aggregate distribution of all price ranges shown earlier in the univariate analysis.

04.Price Range Vs Weight of the phone.

The image above shows the boxplot and the violin plot of the price range vs weight of the phone. These two graphs don't show that there is a significant increase or decrease of weight with the price. However the median of the very high cost price range is lower than other price ranges which might be an indication of higher prices associated with low weight mobile phones. All four price ranges show a similar spread -a uniform distribution- , which is similar to the aggregate distribution of all price ranges shown earlier in the univariate analysis.

05.Pricer Range Vs RAM Capacity.



The image above shows the boxplot and the violin plot of the price range vs weight of the phone. These two graphs clearly show that there is a relationship between price range and RAM. The RAM capacity of a mobile phone increases with the price according to this graph. The low cost mobile phones show positively skewed distributions indicating most of the low cost mobile phones have less RAM capacity. Medium cost and high cost price range shows more elongated distribution compared to the other two groups indicating that medium cost and high cost price ranges have much more variety of choices for the ram capacity than that of low cost and high cost. The very high cost mobile phones show negatively skewed distribution indicating that most of the high cost mobile phones have a high ram capacity.

06.Price Range Vs Dual Sim Function.

|  |  |  |  |
| --- | --- | --- | --- |
|  | available | unavailable | total |
| Low Price | 250 | 250 | 500 |
| Medium Price | 255 | 245 | 500 |
| High Price | 249 | 251 | 500 |
| Very High Price | 265 | 235 | 500 |
| Total | 1019 | 981 | 2000 |

The contingency table above shows how price ranges differ with dual sim function. Values in the table are uniformly distributed among the cells. Therefore there’s no apparent relationship between price and dual sim function by inspecting the table. However a chi-square test can be done to test this.

07.Price Range Vs Touchscreen Function.

|  |  |  |  |
| --- | --- | --- | --- |
|  | available | unavailable | total |
| Low Price | 262 | 238 | 500 |
| Medium Price | 261 | 239 | 500 |
| High Price | 235 | 265 | 500 |
| Very High Price | 248 | 252 | 500 |
| Total | 1006 | 994 | 2000 |

The contingency table above shows how price ranges differ with touchscreen function. Values in the table are uniformly distributed among the cells. Therefore there’s no apparent relationship between price and touchscreen function by inspecting the table. However a chi-square test can be done to test this.

08.Price Range Vs 4-G Capability.

|  |  |  |  |
| --- | --- | --- | --- |
|  | available | unavailable | total |
| Low Price | 259 | 241 | 500 |
| Medium Price | 262 | 238 | 500 |
| High Price | 247 | 253 | 500 |
| Very High Price | 275 | 225 | 500 |
| Total | 1043 | 957 | 2000 |

The contingency table above shows how price ranges differ with 4G capability. Values in the table are uniformly distributed among the cells. Therefore there’s no apparent relationship between price and 4G capability by inspecting the table. However a chi-square test can be done to test this.

09.Price Range Vs Front Camera Function.

|  |  |  |  |
| --- | --- | --- | --- |
|  | available | unavailable | total |
| Low Price | 379 | 121 | 500 |
| Medium Price | 376 | 124 | 500 |
| High Price | 379 | 121 | 500 |
| Very High Price | 392 | 108 | 500 |
| Total | 1526 | 474 | 2000 |

The contingency table above shows how price ranges differ with front camera function. Values of the table shows that most of the phones (1526) regardless of price range (as mentioned in univariate analysis) do have a front camera but the table doesn’t show much difference between values of different price ranges. Therefore there’s no apparent relationship between price and front camera function by inspecting the table. However a chi-square test can be done to test this.

10.Price Range Vs Primary Camera Function.

|  |  |  |  |
| --- | --- | --- | --- |
|  | available | unavailable | total |
| Low Price | 467 | 33 | 500 |
| Medium Price | 473 | 27 | 500 |
| High Price | 480 | 20 | 500 |
| Very High Price | 479 | 21 | 500 |
| Total | 1899 | 101 | 2000 |

The contingency table above shows how price ranges differ with primary camera function. Values of the table shows that most of the phones (1899) regardless of price range (as mentioned in univariate analysis) do have a front camera and the table shows a slight relation between values of different price ranges as the price increases the availability of primary camera increases. However a chi-square test can be done to test this.