A Report On: -
Mobile Price Classification

Submitted By: -

Rajan Ghimire

Submitted To: -

Victoria Shtern

Introduction

Mobile devices are widely used in everyday life, and the characteristics of these devices are the essential factor that may affect the overall performance and usability. This dataset contains all the features which define the capabilities of the mobile device such as battery, processor speed, memory, camera, network and physical size of the device. The goal of this report is to perform Variable Analysis and Exploratory Data Analysis to find out hidden insights, trends and patterns form data. We want to find out the general characteristics of the distribution and dependencies between these parameters that will help to make the correct decisions in the framework of product development, pricing and marketing.

Problem Statement

The purpose of this exploratory data analysis is to identify trends within a comprehensive dataset of mobile-device specifications for company managers to effectively develop, price and sell their products within the mobile market niche.

About the Dataset

Bob has started his own mobile company. He wants to give tough fight to big companies like Apple, Samsung etc. He does not know how to estimate the price of mobiles his company creates. In this competitive mobile phone market, you cannot simply assume things. To solve this problem, he collects sales data of mobile phones of various companies. Bob wants to find out some relation between features of a mobile phone (eg:- RAM,Internal Memory etc.) and its selling price

Feature Analysis

	Name	isKey	Type (int / float)	Description	Range of Values	Is wrong (all no)	Any outlier (all no)	Categorical or Continuous
0	battery_power	no	int	Total energy a battery can store in one time m...	501 - 1998	no	no	Continuous
1	blue	no	int	Has bluetooth or not	0 - 1	no	no	Categorical
2	clock_speed	no	float	Speed at which microprocessor executes instruc...	0.5 - 3.0	no	no	Continuous
3	dual_sim	no	int	Has dual sim support or not	0 - 1	no	no	Categorical
4	fc	no	int	Front Camera mega pixels	0 - 19	no	no	Continuous
5	four_g	no	int	Has 4G or not	0 - 1	no	no	Categorical
6	int_memory	no	int	Internal Memory in Gigabytes	2 - 64	no	no	Continuous
7	m_dep	no	float	Mobile Depth in cm	0.1 - 1.0	no	no	Continuous
8	mobile_wt	no	int	Weight of mobile phone	80 - 200	no	no	Continuous
9	n_cores	no	int	Number of cores of processor	1 - 8	no	no	Categorical
10	pc	no	int	Primary Camera mega pixels	0 - 20	no	no	Continuous
11	px_height	no	int	Pixel Resolution Height	0 - 1960	no	no	Continuous
12	px_width	no	int	Pixel Resolution Width	500 - 1998	no	no	Continuous
13	ram	no	int	Random Access Memory in Megabytes	256 - 3998	no	no	Continuous
14	sc_h	no	int	Screen Height of mobile in cm	5 - 19	no	no	Continuous
15	sc_w	no	int	Screen Width of mobile in cm	0 - 18	no	no	Continuous
16	talk_time	no	int	Longest time that a single battery charge will...	2 - 20	no	no	Continuous
17	three_g	no	int	Has 3G or not	0 - 1	no	no	Categorical
18	touch_screen	no	int	Has touch screen or not	0 - 1	no	no	Categorical
19	wifi	no	int	Has wifi or not	0 - 1	no	no	Categorical
20	price_range	no	int	Price range of the phone	0 - 3	no	no	Categorical

Exploratory Data Analysis

Missing Values in Data:

The dataset contains no such missing values.

	Column	Missing_Percentage
0	battery_power	0
1	blue	0
2	clock_speed	0
3	dual_sim	0
4	fc	0
5	four_g	0
6	int_memory	0
7	m_dep	0
8	mobile_wt	0
9	n_cores	0
10	pc	0
11	px_height	0
12	px_width	0
13	ram	0
14	sc_h	0
15	sc_w	0
16	talk_time	0
17	three_g	0
18	touch_screen	0
19	wifi	0
20	price_range	0

Analysis of class variable 'price range':

price_range	Count of price_range
3	500
2	500
0	500
1	500
Grand Total	**2000**

The price range is equally distributed among 4 classes. 0 means low cost, 1 means Medium Cost, 2 means high cost and 3 means very high cost.

Count of price_range by price_range
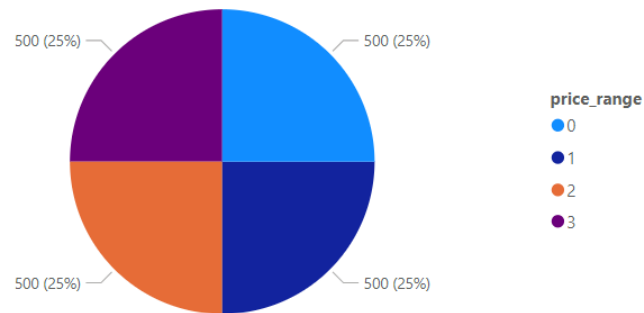
Analysis of battery power:
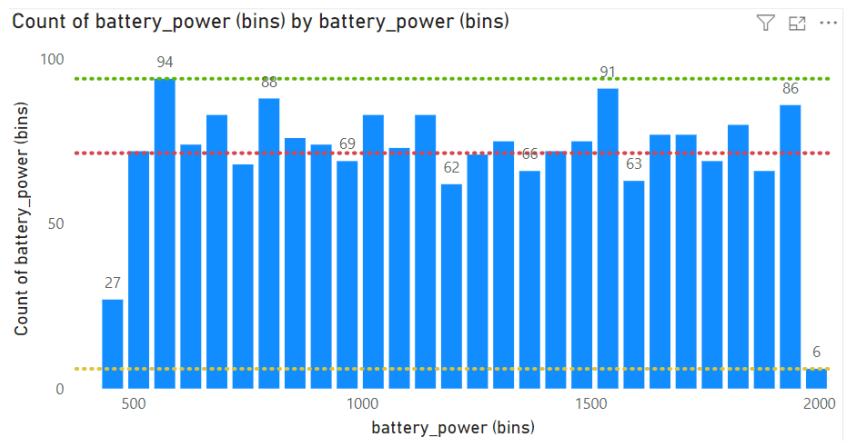
The graph shows the distribution of battery power. Here, the yellow line represents the minimum values if the distribution, red line depicts the average of the distribution and green line shows the maximum values of the distribution.
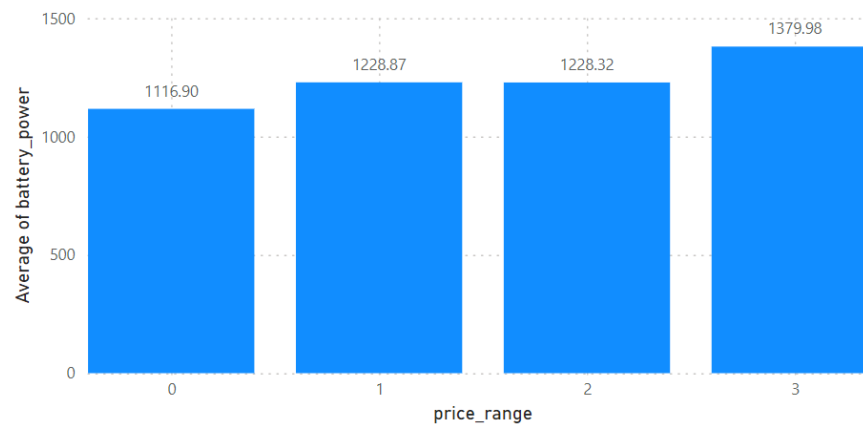
Here, after looking at average battery power per each price range we can interpret that: There is positive co-relation between price range and battery power. As the battery power increases, the price of the mobile also increases.

Analysis of ram:
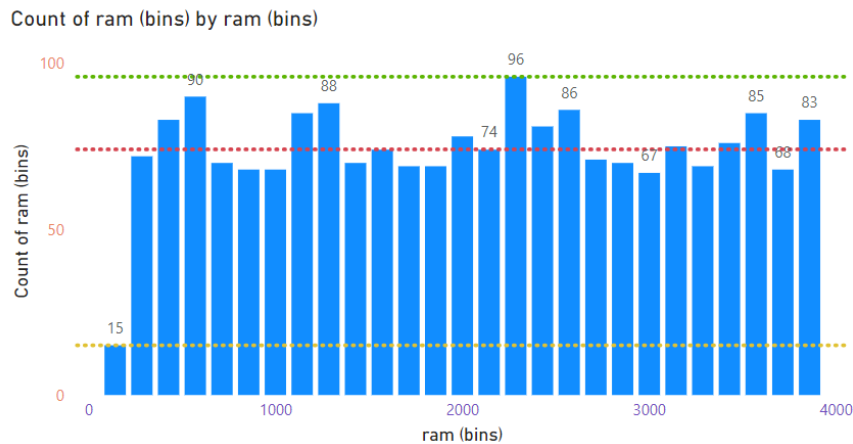
The histogram shows the distribution of ram. Here, the yellow line represents the minimum values if the distribution, red line depicts the average of the distribution and green line shows the maximum values of the distribution.
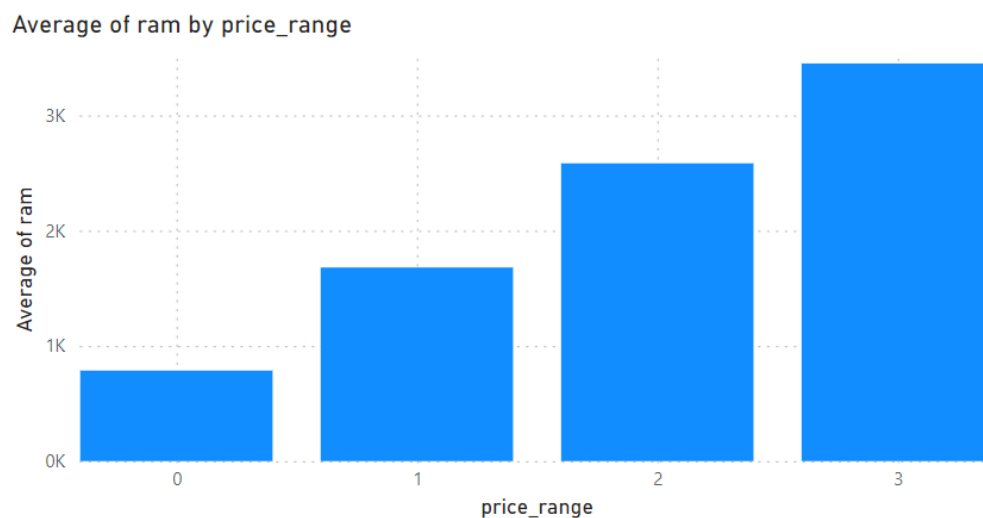
Here, after looking at the average ram per each price range we can interpret that: There is positive co-relation between price range and ram. As the size of ram increases, the price of the mobile also increases.

Analysis of 3G:

The feature three_g is a categorical variable with two classes 1 and 0. 1 represents presence of 3G feature in phone and 0 represents absence of meeting.

three_g	Count of three_g
1	1523
0	477
Grand Total	2000

Count of three_g by three_g
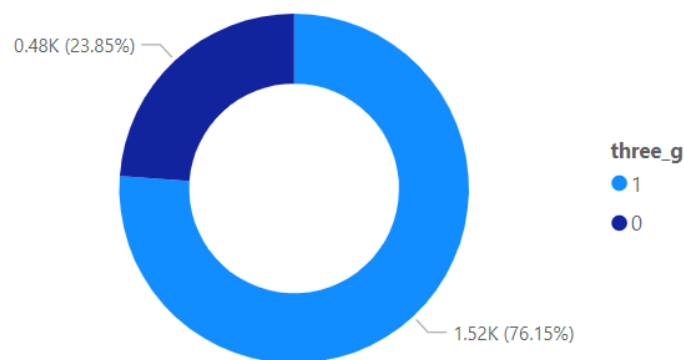
From the given donut plot, we can understand that most of the phones have 3G.
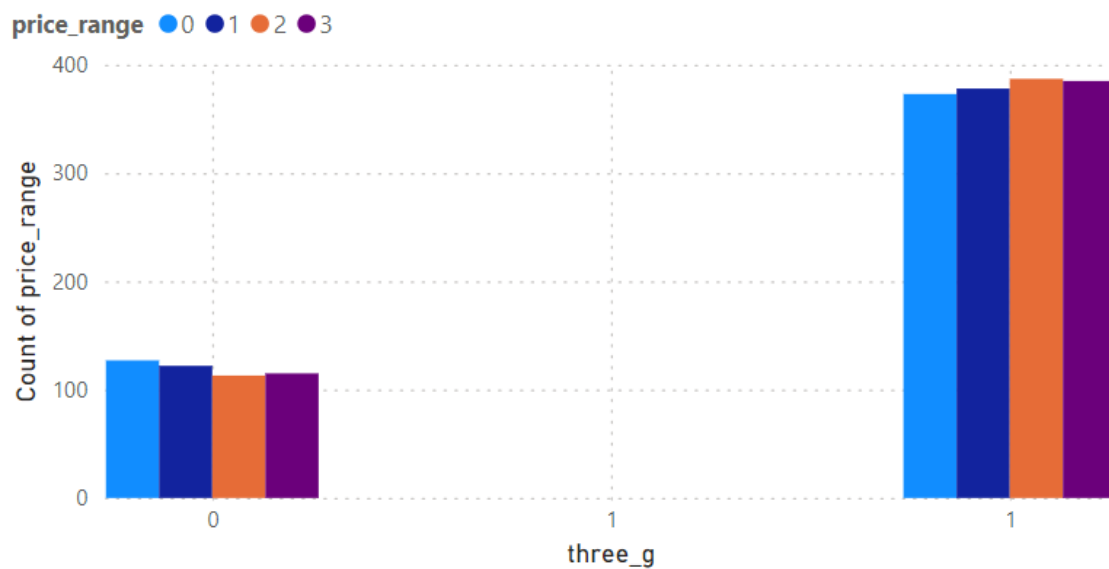
Now, when we look at the count of 3G per price category, we can interpret that the price goes up when the phone has 3G.

Analysis of 4G:

The feature four_g is a categorical variable with two classes 1 and 0. 1 represents presence of 4G feature in phone and 0 represents absence of meeting.

four_g	Count of four_g
1	1043
0	957
Grand Total	**2000**

Count of four_g by four_g
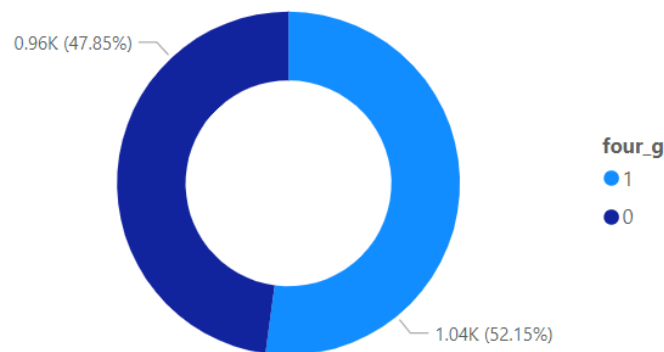
From the given donut plot, we can understand that there is equal distribution between phone with and without 4G.
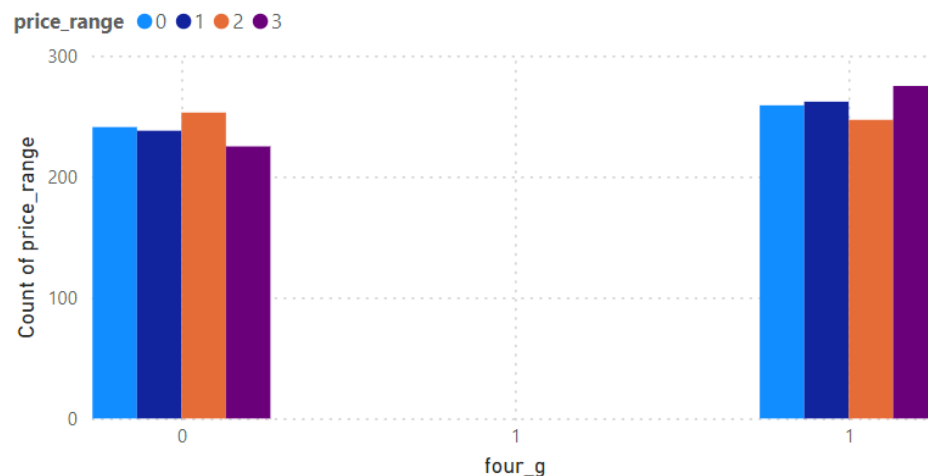
Now, when we look at the count of 4G per price category, we can interpret that 4G feature does not have that much of difference in price of the mobile phone.

Analysis of n_cores:

The n_cores features represent the total number of processing cores within the system. The higher the number of processing cores, the better the performance.

n_cores	Count of n_cores
4	274
7	259
8	256
2	247
3	246
5	246
1	242
6	230
Grand Total	**2000**
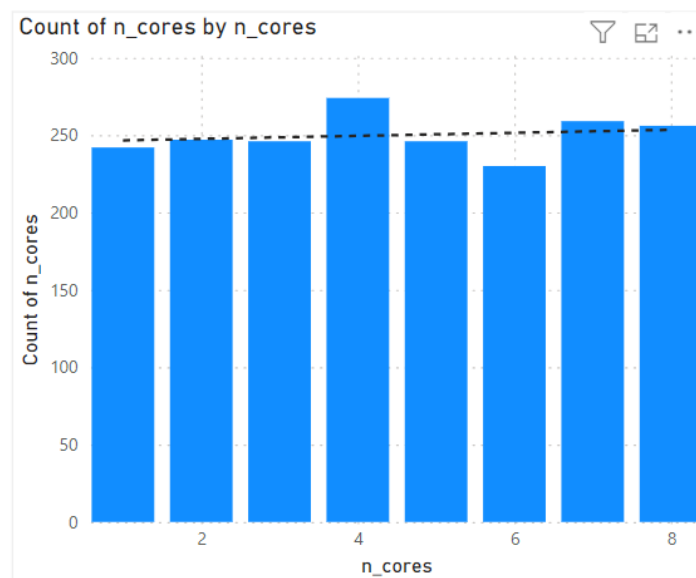
The histogram shows the distribution of n_cores. The black dotted line represents overall trend of the distribution.

Now, when we look at the count of n_cores per price category, we can interpret that n_cores feature does not have that much of difference in price of the mobile phone. This means, mobile phones of different price range have same number of cores.

Count of price_range by n_cores and price_range

price_range ●0 ●1 ●2 ●3
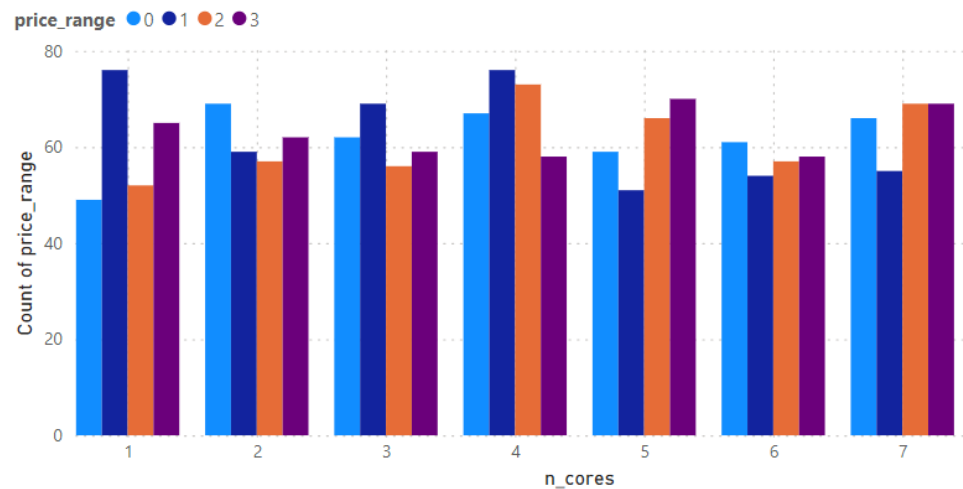
Analysis of talk_time vs battery_power:

In this analysis, we are trying to find out if mobile phones with longer talk time have longer battery life. This is a confirmatory data analysis where we want to confirm this hypothesis.

talk_time	Count of talk_time
7	124
4	123
16	116
15	115
19	113
6	111
10	105
8	104
11	103
20	102
14	101
9	100
13	100
18	100
12	99
2	99
17	98
3	94
5	93
Grand Total	**2000**
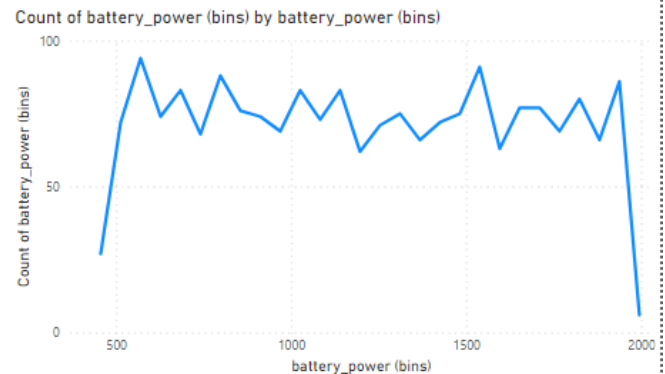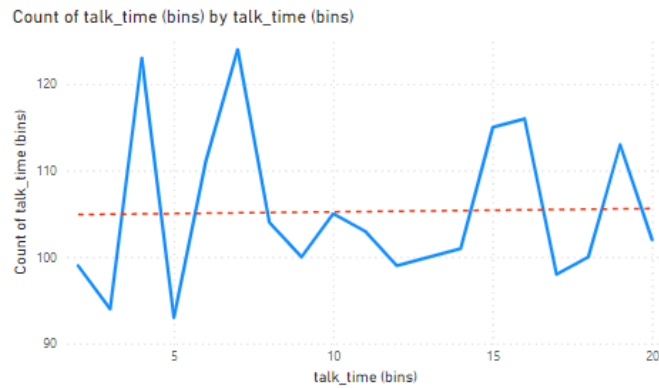
Here we have two-line charts that show count of talk time and battery for different bins. The red line is to indicate overall trend in the talk time.
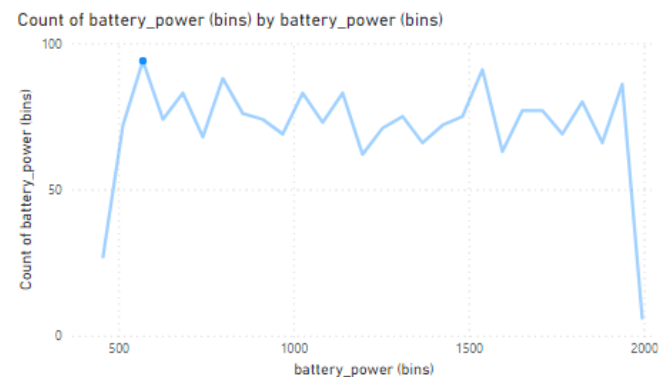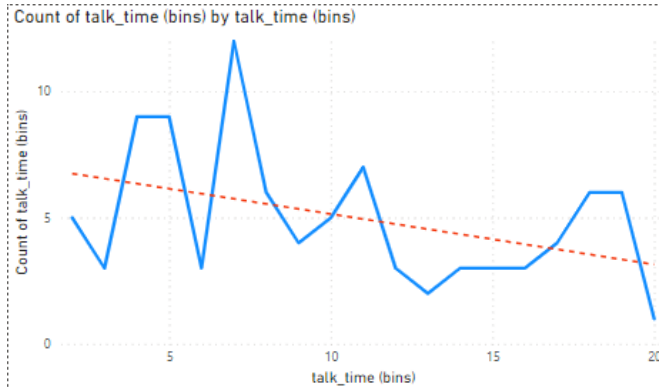
Here If we select lower battery power, the overall talk time is decreasing shown by the red trend line in talk time.

Count of talk_time (bins) by talk_time (bins)

Count of battery_power (bins) by battery_power (bins)
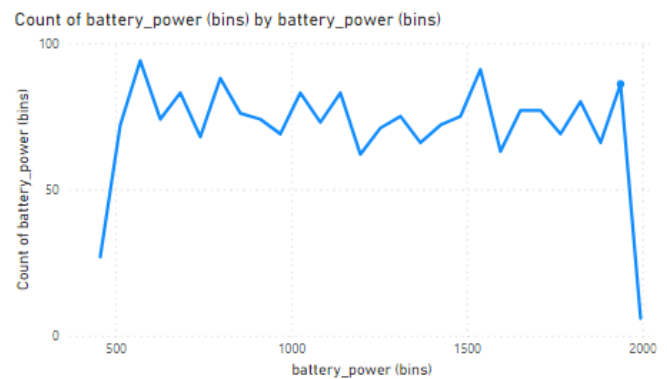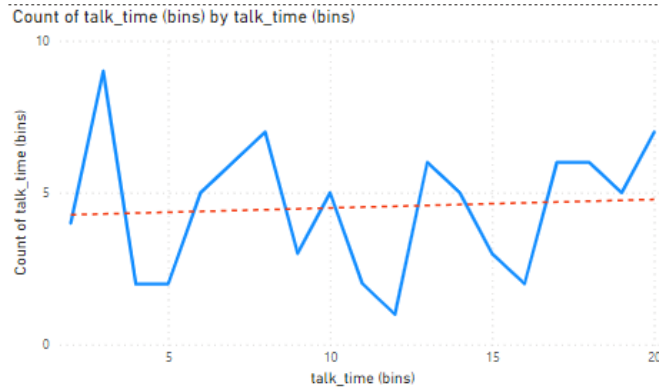
Similarly, when we select higher talk time, we can see that overall talk time is increasing shown by the red trend line which proves our hypothesis.

Conclusion

The exploratory data analysis on the mobile device dataset has yielded valuable insights into the relationships between various features and the price range of the devices. The analysis revealed a positive correlation between battery power, RAM capacity, and the device's price range, suggesting that higher-end devices tend to offer larger batteries and more RAM. Interestingly, the presence of 3G connectivity appeared to influence pricing, while 4G connectivity and the number of processor cores did not significantly impact pricing across different price categories. Additionally, the analysis confirmed the expected behavior that lower battery power corresponds to shorter talk times, and vice versa. These findings can inform product development strategies, pricing decisions, and marketing efforts, enabling manufacturers and retailers to better align their offerings with consumer preferences and market dynamics within the highly competitive mobile device.

References

Iabhishekofficial. (n.d.). Mobile Price Classification. Kaggle. Retrieved June 5, 2024, from https://www.kaggle.com/datasets/iabhishekofficial/mobile-price-classification

Dashboard

Average of battery_power by price_range
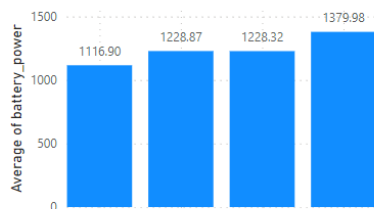
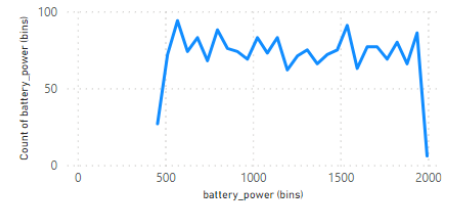**Count of price_range by three_g and price_range**
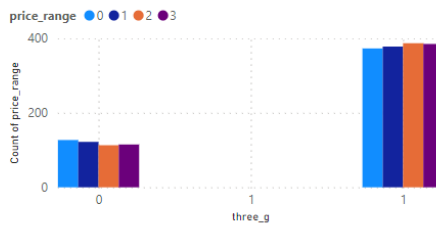
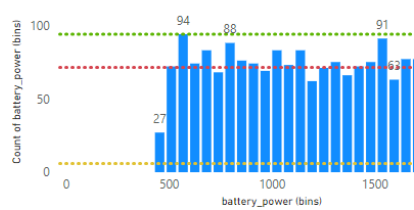**Count of talk_time (bins) by talk_time (bins)**
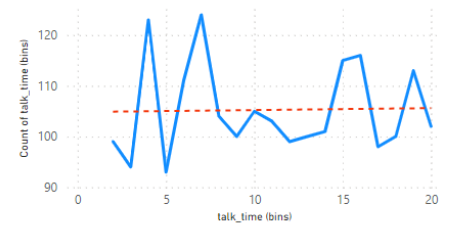
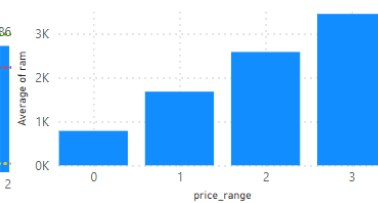**Count of price_range by four_g and price_range**
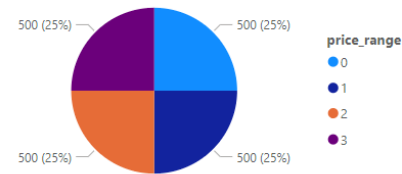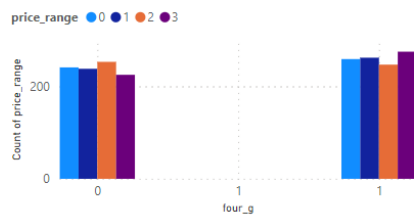
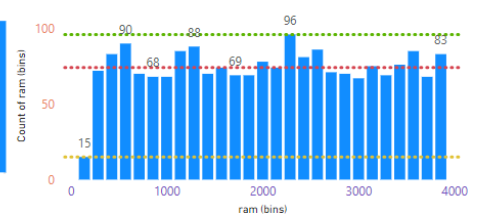**Count of battery_power (bins) by battery_power (bins)**

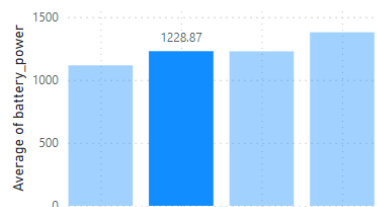**Count of battery_power (bins) by battery_power (bins)**

Average of ram by price_range

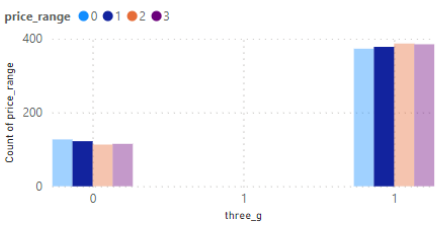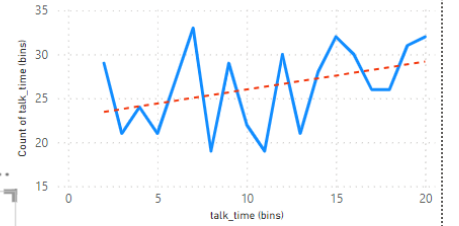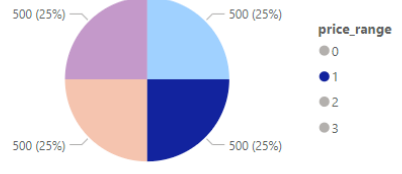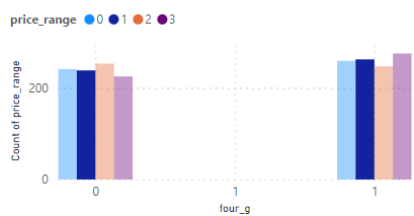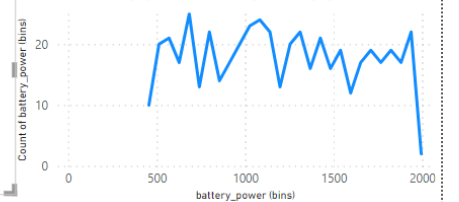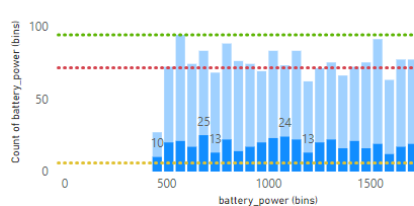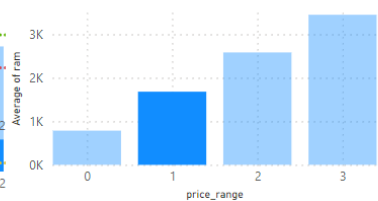**Count of ram (bins) by ram (bins)**