EXPLARATORY DATA ANALYSIS

FLIPKART SELLER

STATISTICS ASSIGNMENT (STS-I)

PRESENTED BY,
SHUBHAM BATHWAL

INTRODUCTION

About Flipkart:

Flipkart is an Indian e-commerce company, headquartered in Bangalore, Karnataka, India. It is the largest e-commerce company in India and was founded by Sachin and Binny Bansal. The company has wide variety of products electronics like laptops, tablets, smartphones, and mobile accessories to in-vogue fashion staples like shoes, clothing and lifestyle accessories; from modern furniture like sofa sets, dining tables, and wardrobes to appliances that make your life easy like washing machines, TVs, ACs, mixer grinder juicers and other time-saving kitchen and small appliances; from home furnishings like cushion covers, mattresses and bedsheets to toys and musical instruments.

Mobile Phones

Mobile phones are one of the most rapidly rising industries, as well as one of the most prominent industries in the technology sector. The rate of increase has been exponential, with the number of mobile phone customers increasing fivefold in the last decade. Globally, the number of smartphones sold to end users climbed from 300 million in 2010 to 1.5 billion by 2020.

Flipkart and Mobile Phones

As previously stated, mobile phones are in high demand and are one of the ideal products for a novice to sell. Flipkart will be the ideal spot for a vendor to market their stuff because its reach.

BUSINESS PROBLEM

The objective is to address a hypothetical business problem for a Flipkart Authorized Seller. According to the problem the individual is looking to sell mobile phones on Flipkart. For this, the individual is looking for the best product, brand, specification and deals that can generate the most revenue with the least amount of investment and budget constraints.

Questions to be answered:

1. Should he simply sell products for one brand, or should he try to sell models from various brands?
2. Using EDA and Data Visualization find out insights and relation between different features.
3. Perform detailed analysis of each brand.

DATA COLLECTION

Using a web scraping technique, the data was retrieved from the Flipkart website. Data was obtained on September 14, 2021; thus, it may differ somewhat from latest revisions. The information in the dataset is genuine.

ASSUMPTION

We don't have a direct sales record that shows how many units of a certain mobile model were sold.

In most cases, the number of people who rate a product is proportional to the number of units sold. As a result, we are considering the number of persons rating the product as the equivalent units sold in the solution.

ABOUT DATASET

The dataset includes data on mobile phones from the top five most popular brands in India: Apple, Poco, Realme, Samsung, and Xiaomi. Information like RAM, ROM, Display Size. etc are present which distinguishes one product from another. At least one attribute distinguishes each product. Dataset has no null value.

Columns: There are 16 columns each having a title which is self-explanatory.
Rows: There are 430 rows each having a mobile with at least a distinct feature.

DESCRIPTION OF ATTRIBUTES

I. **brand**: Brand Name (Categorical)
II. **model**: Model Name (Categorical)
III. **base_color**: Phone Color (Categorical)
IV. **processor**: Processor brand used (Categorical)
V. **screen_size**: Categorical screen size (Categorical)
VI. **ROM**: ROM in gigabyte (Numeric – Discrete)
VII. **RAM**: RAM in gigabyte (Numeric – Discrete)
VIII. **display_size**: Actual display size in inches (Numeric – Continuous)
IX. **num_rear_camera**: No. of cameras on back (Numeric – Discrete)
X. **num_front_camera**: No. of cameras on front (Numeric – Discrete)
XI. **battery_size**: Battery in mAH (Numeric – Continuous)
XII. **ratings**: Customer rating for the product (Numeric – Continuous)
XIII. **num_of_ratings**: No. of people rating the product, also the equivalent no. of unit sold for our problem (Numeric – Continuous)
XIV. **sales_price**: Selling price of the unit after discount (Numeric – Continuous)
XV. **discount_percent**: Discount in percentage offered (Numeric – Continuous)
XVI. **sales**: Sales of product in crore rupees (Numeric – Continuous)

EXPLORATORY DATA ANALYSIS

EDA is one of the most important phases in data science since it helps us to obtain critical insights and statistical metrics. In general, EDA can be categorised in two ways.

The first distinction is that each method is either non-graphical or graphical.

Second, each method is univariate or multivariate in nature (usually just bivariate).

Non-graphical approaches typically include the computation of summary statistics, but graphical methods clearly summarize the data in a diagrammatic or pictorial manner.

Let's look at each type individually.

INSTANCE OF THE DATASET

	brand	model	base_color	processor	screen_size	ROM	RAM	display_size	num_rear_camera	num_front_camera
0	Apple	iPhone SE	Black	Water	Very Small	64	2	4.7	1	1
1	Apple	iPhone 12 Mini	Red	Ceramic	Small	64	4	5.4	2	1
2	Apple	iPhone SE	Red	Water	Very Small	64	2	4.7	1	1
3	Apple	iPhone XR	Others	iOS	Medium	64	3	6.1	1	1
4	Apple	iPhone 12	Red	Ceramic	Medium	128	4	6.1	2	1

battery_capacity	ratings	num_of_ratings	sales_price	discount_percent	sales
1800	4.5	38645	32999	0.17	127.52
2815	4.5	244	57149	0.04	1.39
1800	4.5	38645	32999	0.17	127.52
2942	4.6	5366	42999	0.10	23.07
2815	4.6	745	69149	0.02	5.15

DESCRIPTIVE STATISTICS (NON-GRAPHICAL)

In this section, we will look at the Measures of Central Tendency (Mean, Median, Mode) and Measures of Dispersion (Standard Deviation, Range and Quartiles).

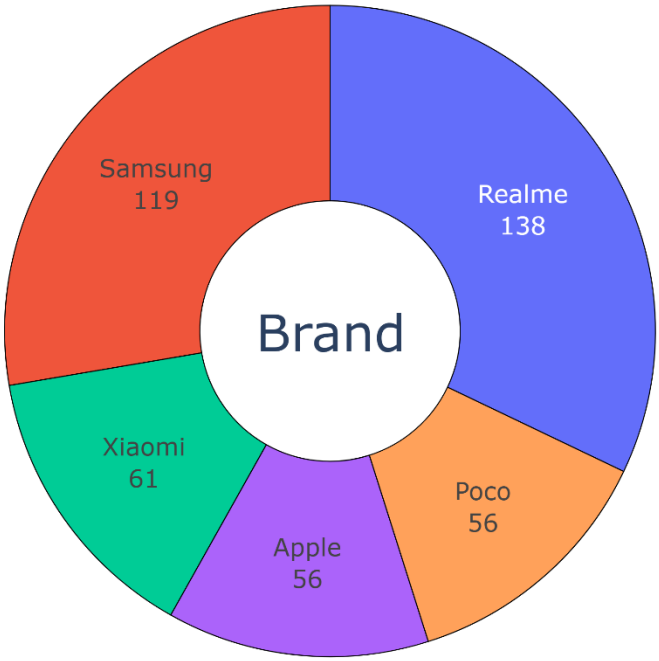
Numerical Features

	display_size	battery_capacity	ratings	num_of_ratings	sales_price	discount_percent	sales
count	430.000000	430.000000	430.000000	430.000000	430.000000	430.000000	430.000000
mean	6.369767	4529.397674	4.339302	23567.944186	25433.234884	0.108000	29.752326
std	0.369549	986.907252	0.151494	56096.277784	22471.926588	0.073432	58.399588
min	4.700000	1800.000000	3.000000	4.000000	5742.000000	0.010000	0.000000
25%	6.300000	4000.000000	4.300000	745.000000	11999.000000	0.060000	1.640000
50%	6.500000	4500.000000	4.300000	5197.500000	16989.500000	0.090000	9.655000
75%	6.500000	5000.000000	4.400000	21089.250000	28999.000000	0.160000	29.717500
max	7.600000	7000.000000	4.600000	642373.000000	157999.000000	0.440000	550.190000

Categorical Features

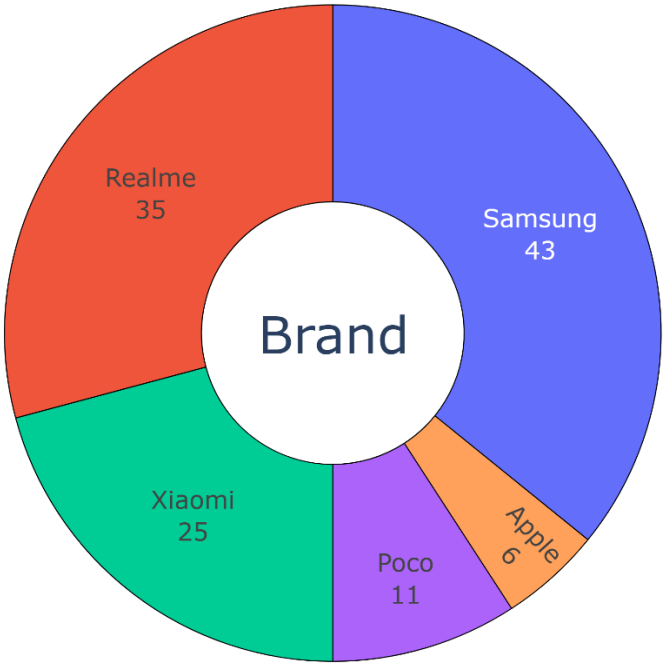
	brand	model	base_color	processor	screen_size	ROM	RAM	num_rear_camera	num_front_camera
count	430	430	430	430	430	430	430	430	430
unique	5	119	12	7	5	7	7	4	3
top	Realme	iPhone XR	Blue	Qualcomm	Large	128	4	3	1
freq	138	18	117	168	242	192	133	157	413

BRAND-WISE DISTINCT PRODUCT



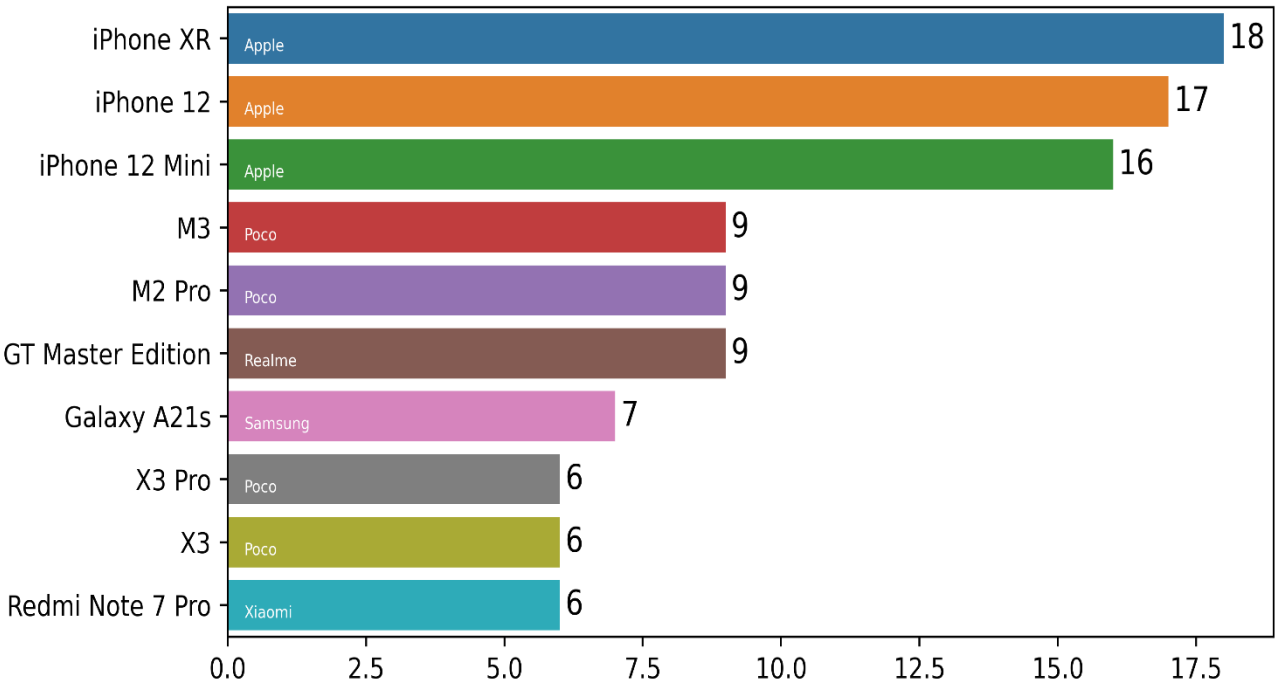
Realme offers the most options, as seen by pie charts. In terms of specs, they have a broad range of phones to select from. Poco, on the other hand, has the fewest alternatives. This might be since Poco is a relatively new brand.

BRAND-WISE MODEL COUNT



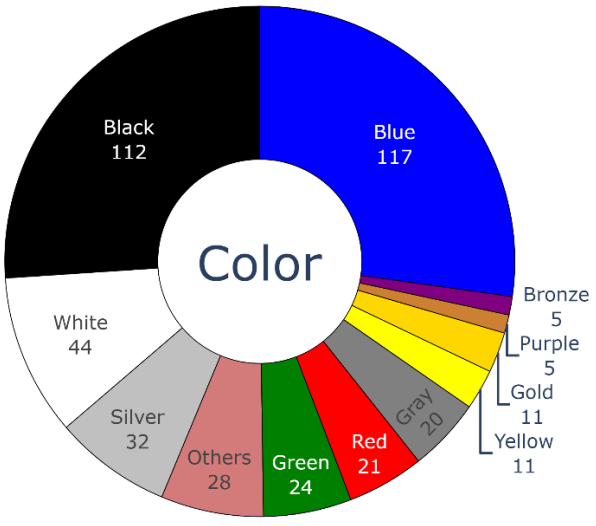
Even though Realme has the most product possibilities, Samsung has a lot more distinct mobile phone model. Apple, while being a well-known brand, has the smallest amount of model possibilities.

MODEL WISE PRODUCT COUNT



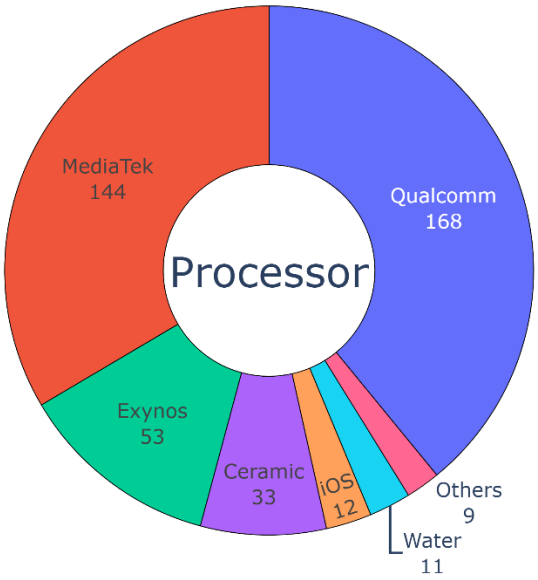
The model iphone XR, iphone 12, and iphone 12 mini have the largest variance within the model, as shown in the above bar chart. There might be differences in the mobile's specifications, pricing, and colour. This also explains why Apple and Poco have a lower model count because they have the most versions with different specifications.

COLOUR WISE PRODUCT COUNT



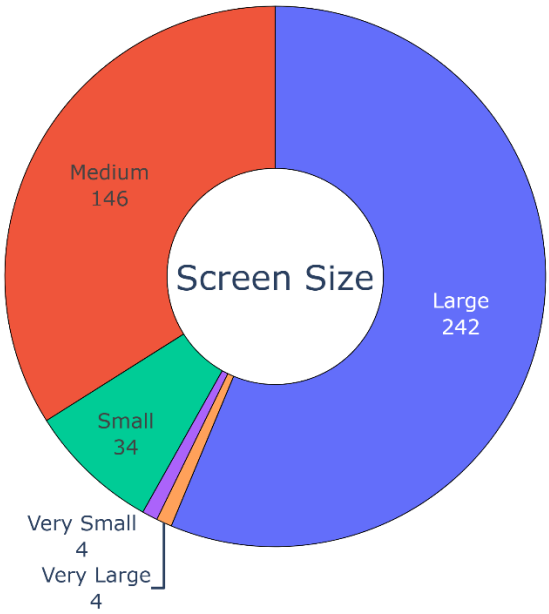
Blue is the most common colour, followed by black and white.

PROCESSOR WISE PRODUCT COUNT



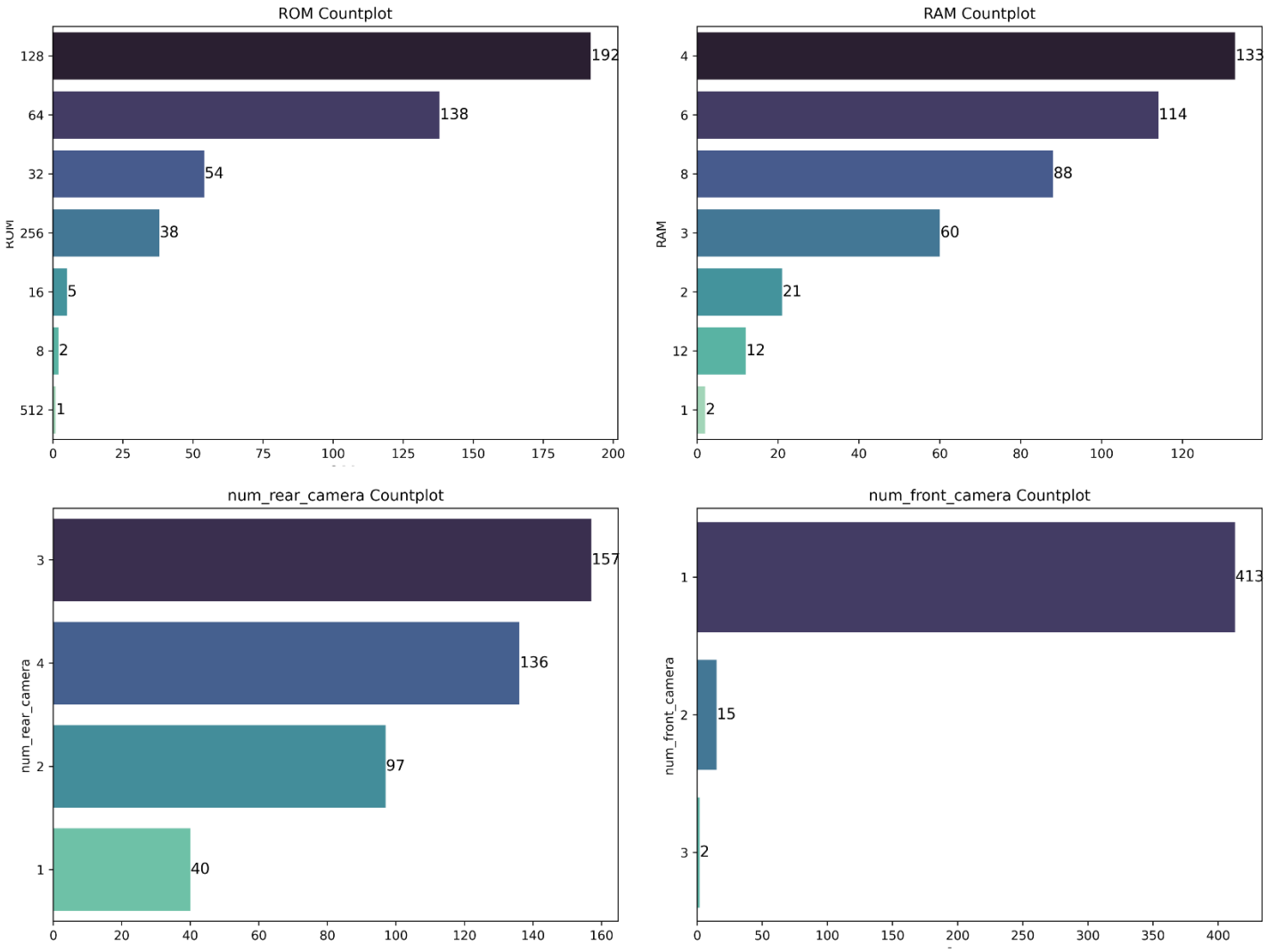
Qualcomm is the most prevalent CPU brand, accounting for 168 of the 430 mobile phones. Together, MediaTek and Qualcomm offer processors for more than half of all mobile phones.

SREENSIZE WISE PRODUCT COUNT



Half of the mobile products are in the large category, meaning they are larger than 6.35 inches.
Because of the wide selection of products available in big and medium sizes, other display sizes are inconsistent.

COUNTPLOT

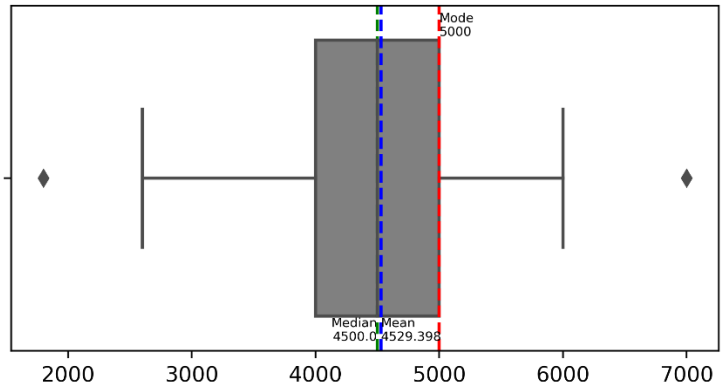


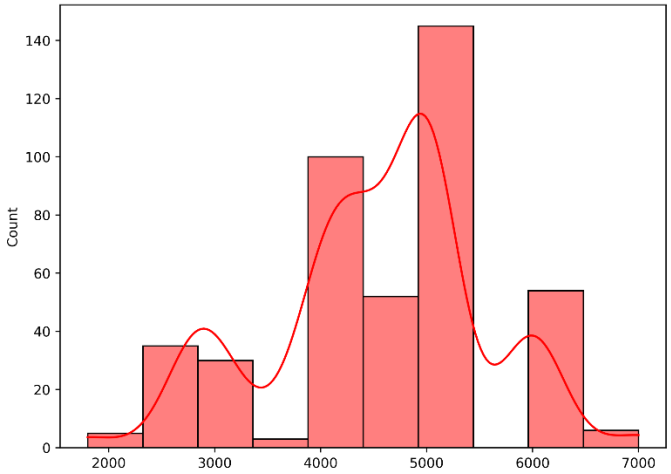
The bar graphs above indicate the number of different products based on the features - RAM, ROM, number of front cameras, and number of back cameras.

There are around 192 mobile phones with 128 GB of ROM and 133 mobile phones with 4 GB of RAM.

Mobile phones with multiple front cameras are rare, but phones with multiple rear cameras are widespread.

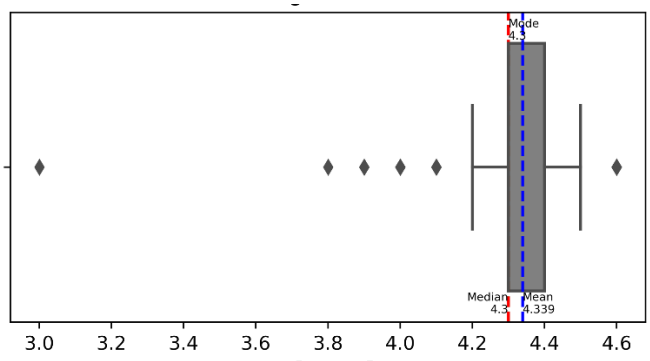
BATTERY CAPACITY BOXPLOT AND HISTOGRAM

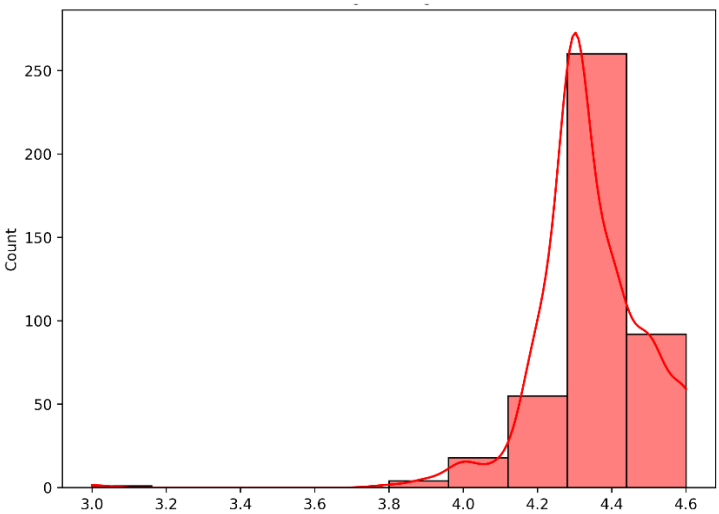




Most mobile phones have a battery backup of 4000 to 5000 mAH. But there is option to go higher or lower in terms of mAH.

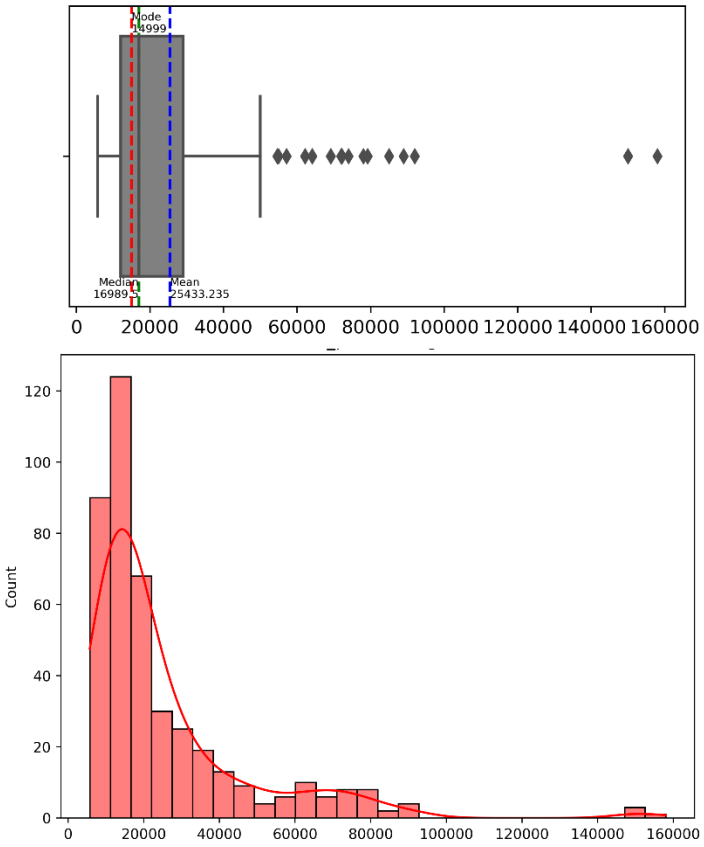
RATINGS BOXPLOT AND HISTOGRAM





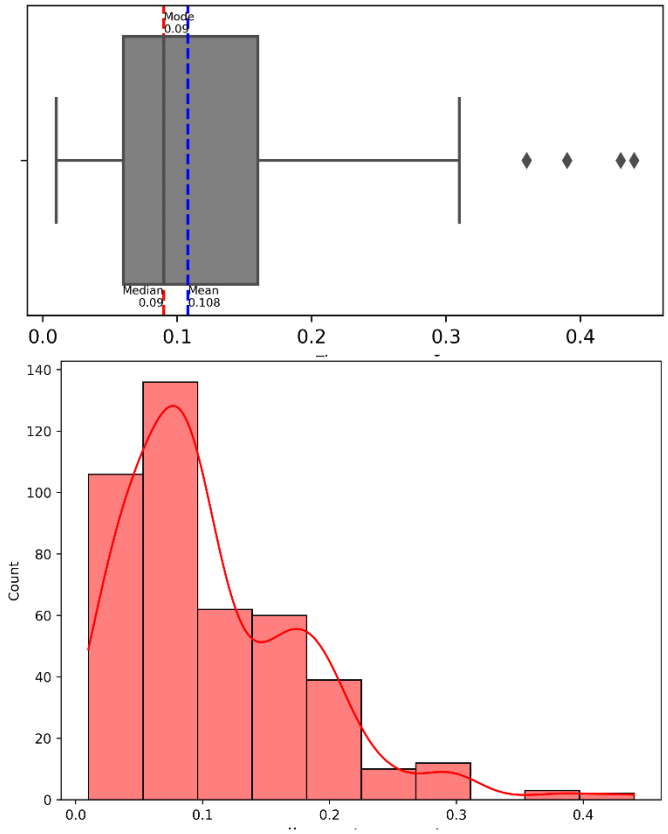
Most of the products get excellent reviews. With a mean of 4.4 and a median of 4.3. Due to some mobiles with lower ratings, the distribution is skewed.

SELLING PRICE BOXPLOT AND HISTOGRAM



As expected, most of the products sold are under Rs 20000. There are several outliers in the higher range, therefore the distribution is right skewed.

DISCOUNT PERCENTAGE BOXPLOT AND HISTOGRAM

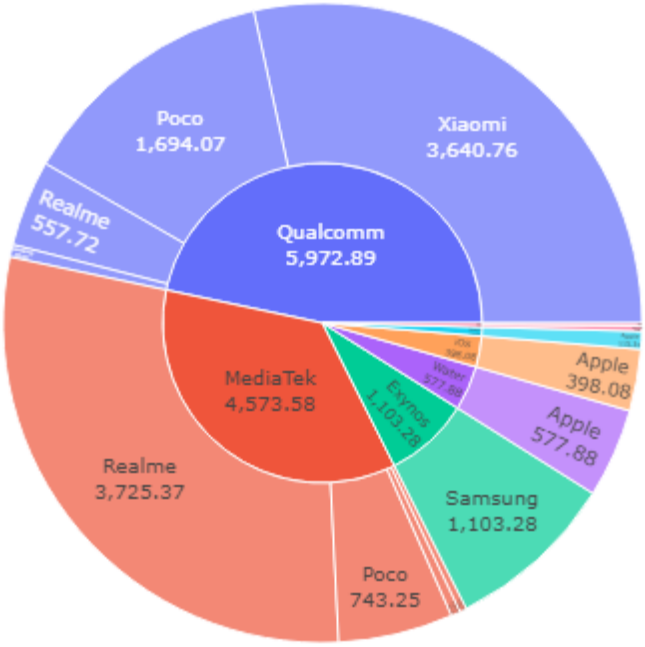


All brands offer some kind of deals. Mean discount offered by brand is 10 percent.

BRANDWISE NO. OF UNITS SOLD

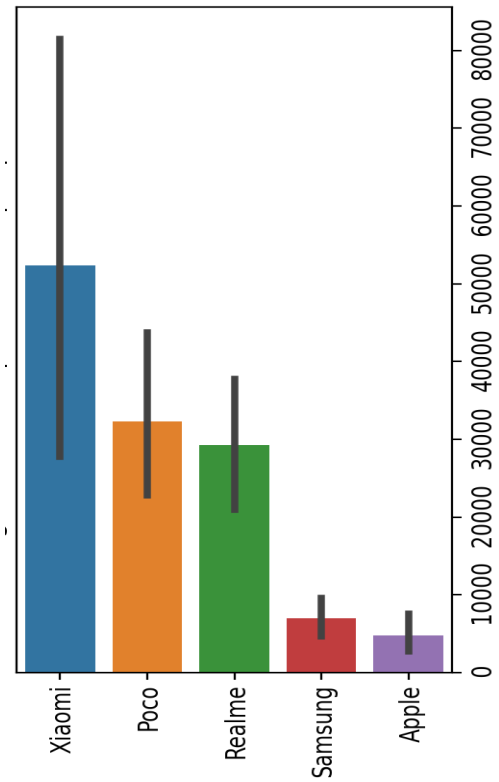
Realme appears to be doing well in terms of sales, with more than 4 million smartphones sold on Flipkart. While premium mobile brands such as Apple and Samsung are lagging in sales, this might be attributed to their expensive pricing.

Processor and Brand Wise Sales

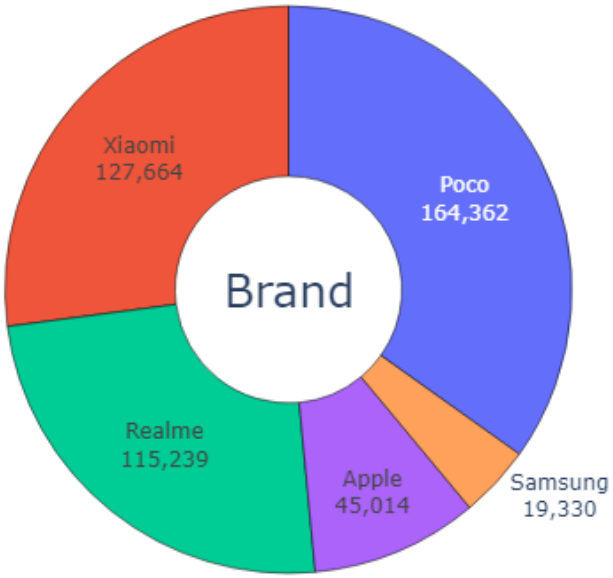


For Indian customers, Qualcomm and MediaTek are the preferred CPU brands. They are responsible for more than 70% of overall sales.
Xiaomi is a big backer of Qualcomm and Realme of MediaTek.

Average Number of Products sold per product

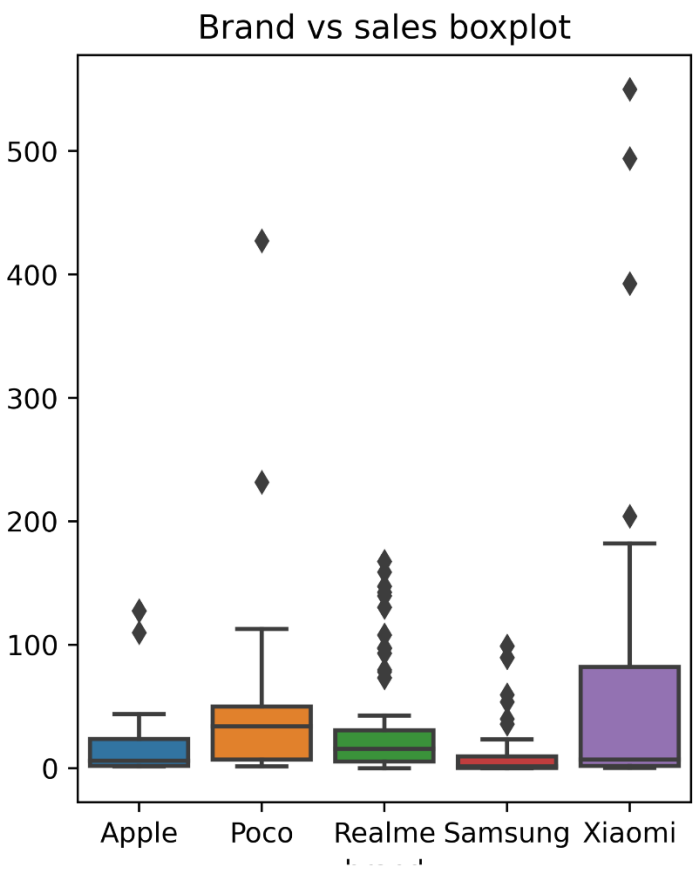


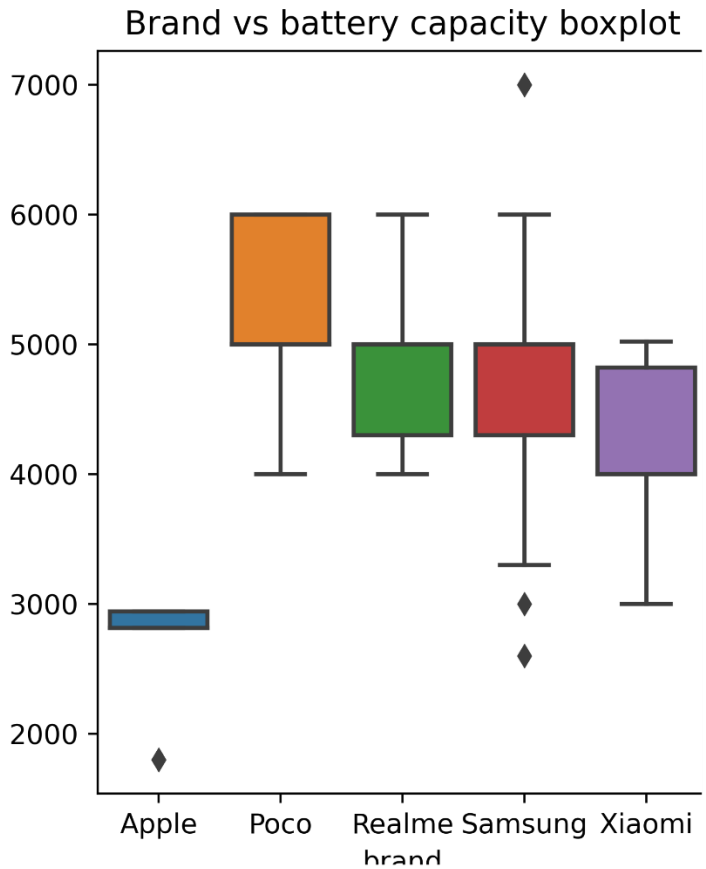
Average Number of Products sold per model

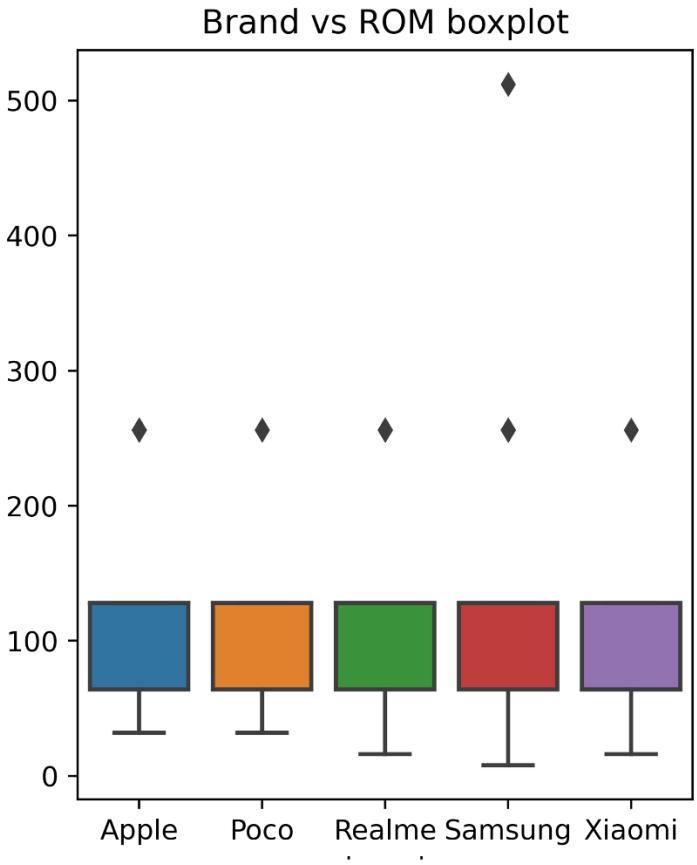


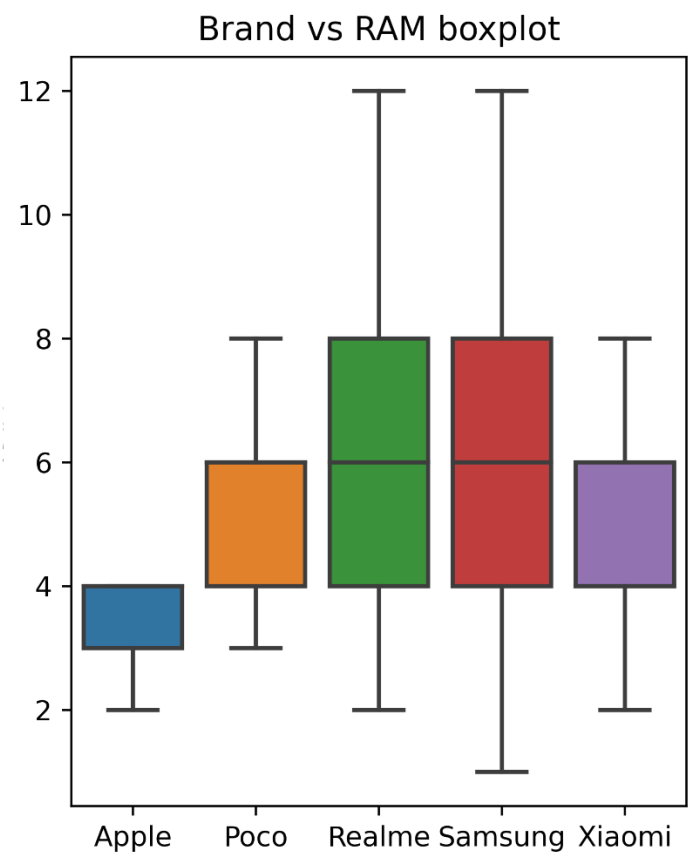
We can see that Xiaomi and Poco are the most popular brands in India, with Realme following closely after. According to the bar graph, each Xiaomi device with any specification and colour will sell at least 28000 units, with an average of 48000 units. While the pie chart shows that Poco is the more popular brand in terms of models, this discrepancy might be due to Xiaomi customers having more options per model than Poco customers.

BRANDWISE BOXPLOTS



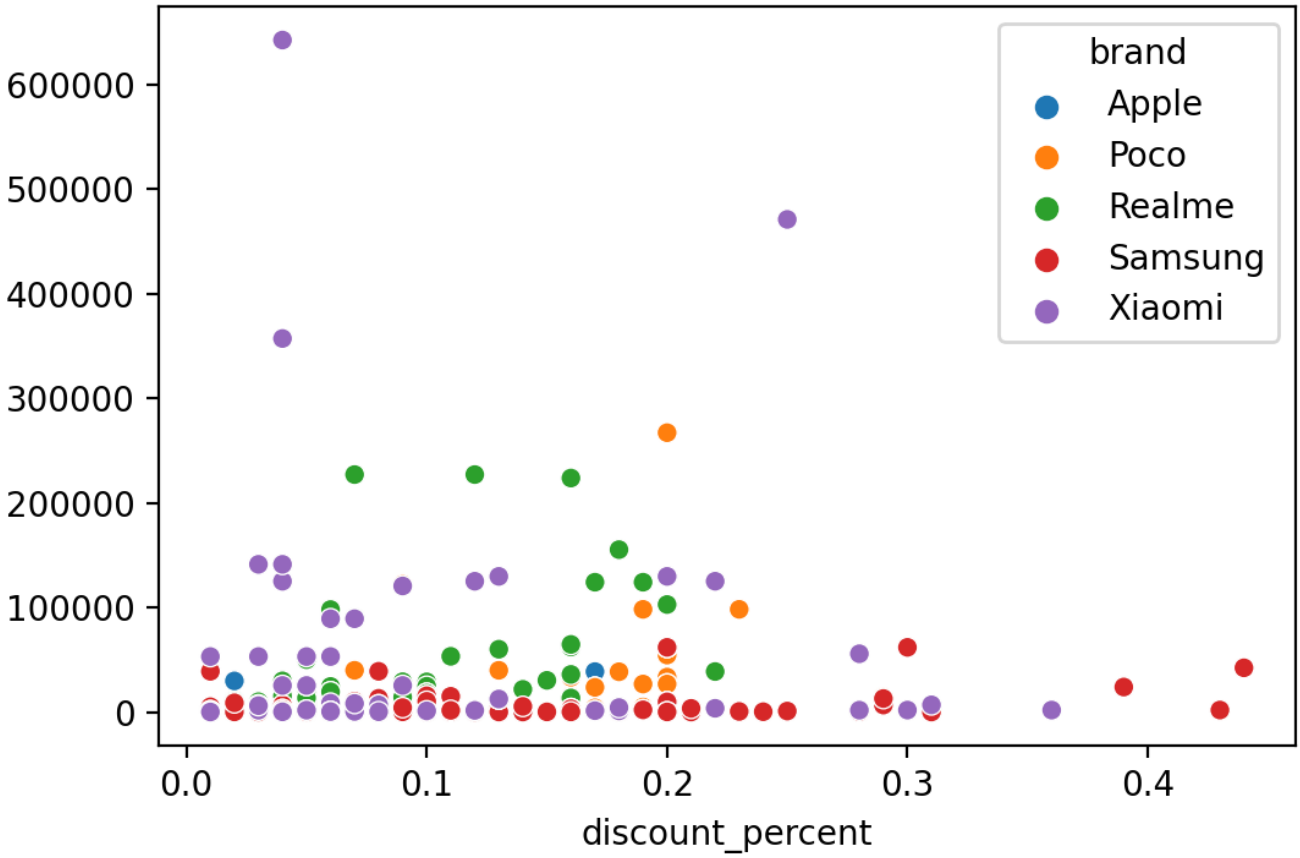




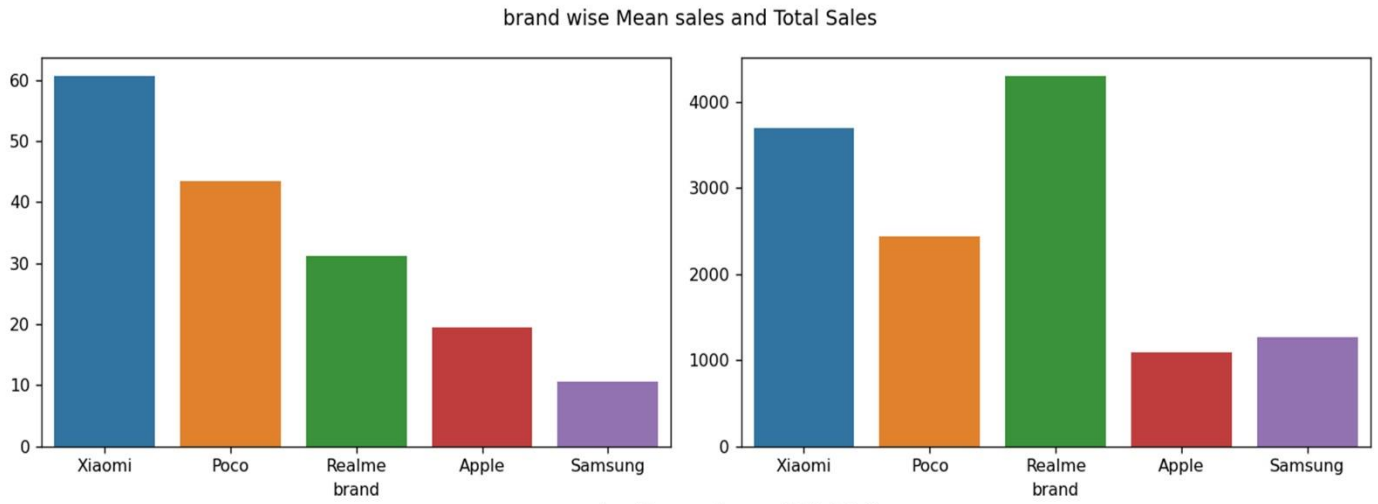


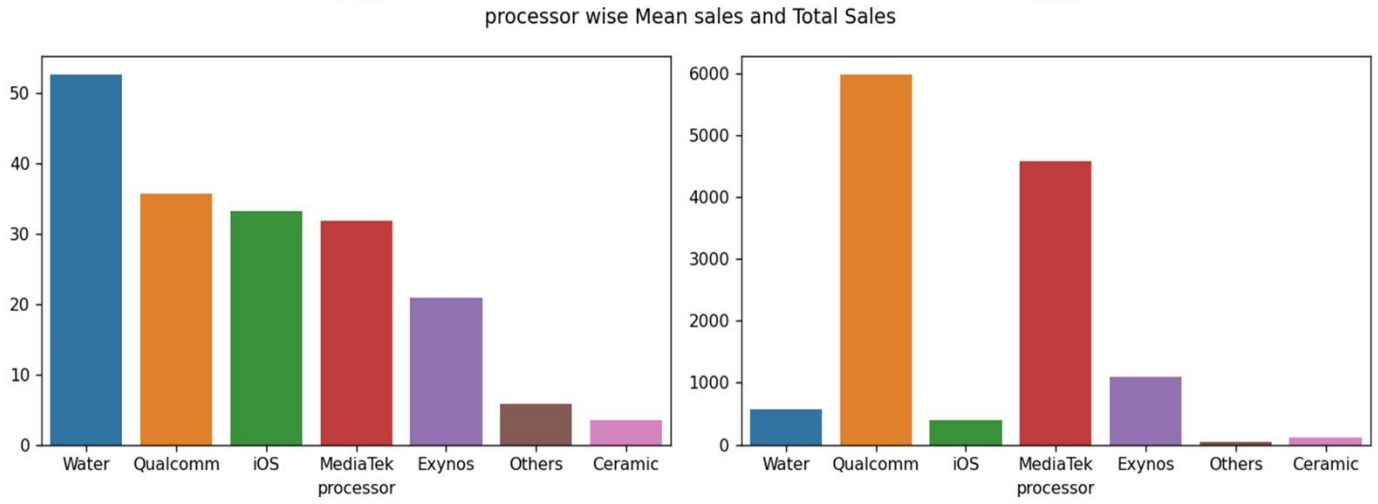
1. The 1st boxplot shows that Apple and Samsung have nearly identical sales. In comparison to Realme, Apple, and Samsung; Poco has a larger variance in mobile product sales. While Xiaomi's sales are phenomenal, the longest upper whisker demonstrates that Xiaomi items sell well. Outliers demonstrate that Xiaomi has some devices that do exceptionally well in terms of sales.
2. The 2nd boxplot shows that Poco phones have the best battery capacity, while Apple phones have the poorest.
3. The 3rd boxplot indicates that the ROM size of all brands is nearly same.
4. The 4th boxplot shows that Realme and Samsung have phones with a wider variety of RAM choices. Poco and Xiaomi have the most devices of 4 and 6 GB RAM size.

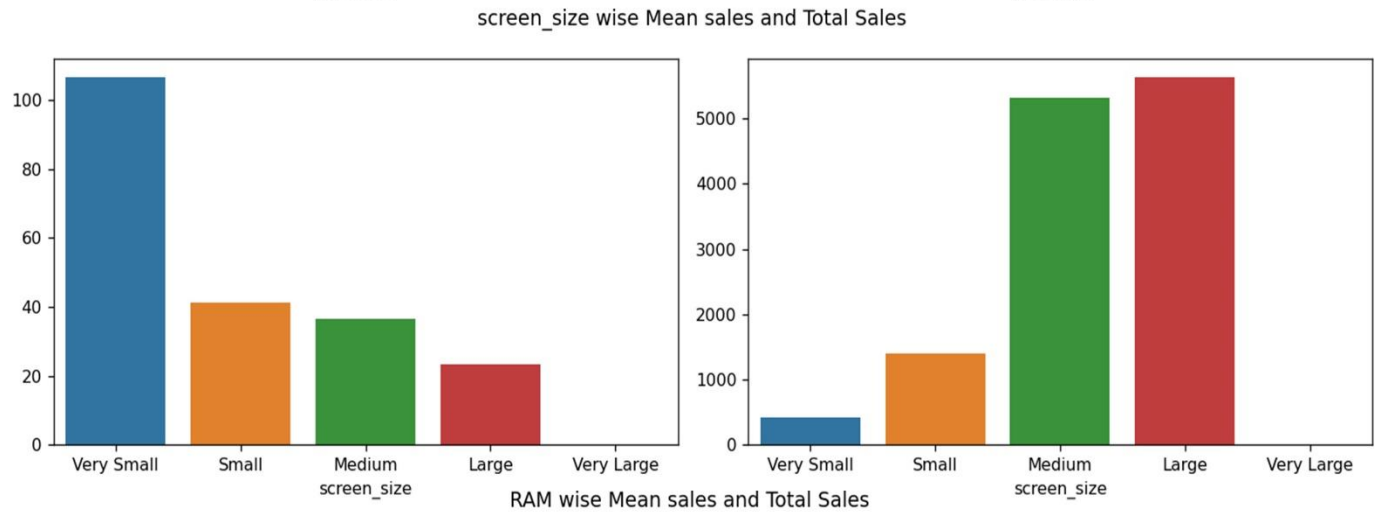
UNITS SOLD VS DISCOUNT OFFERED

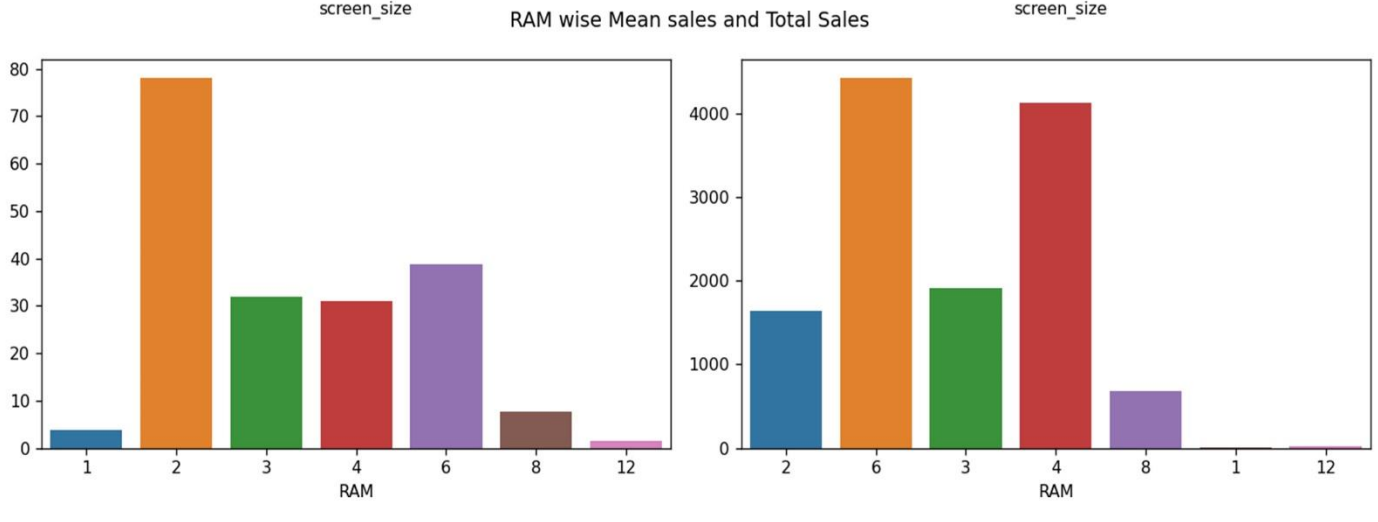


The scatterplot clearly shows that there is no link between sales and the discount provided. This demonstrates that Indian customers are more concerned with the actual worth of phones than with promotional deals. Samsung is a company that consistently provides the most enticing deals, with discounts as high as 40%.

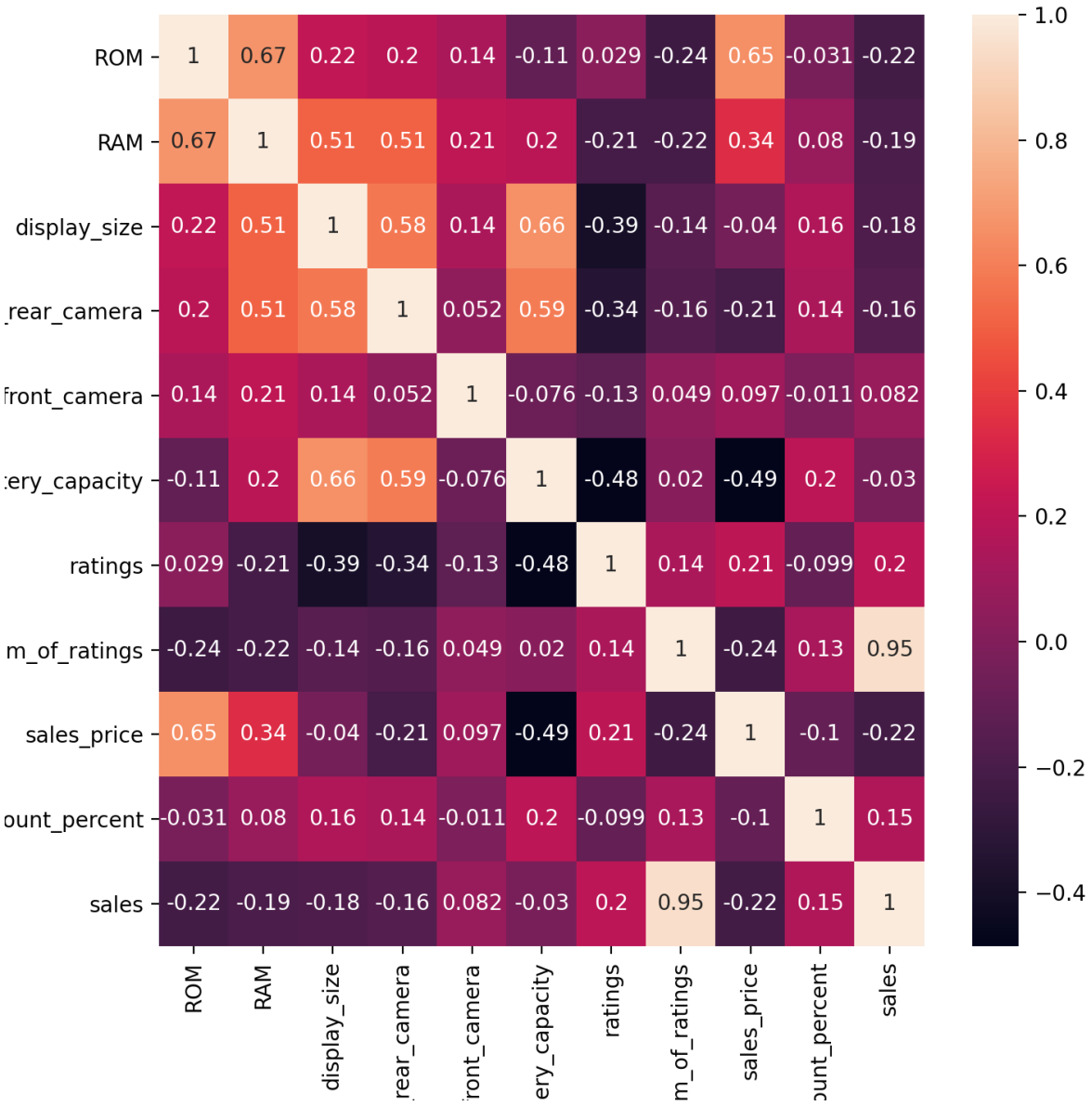

brand wise Mean sales and Total Sales
Xiaomi
Poco
Realme
Apple
Samsung
brand


processor wise Mean sales and Total Sales
Water
Qualcomm
iOS
MediaTek
Exynos
Others
Ceramic
processor


screen_size wise Mean sales and Total Sales
Very Small
Small
Medium
Large
Very Large
screen_size


RAM wise Mean sales and Total Sales
RAM

CORRELATION HEATMAP



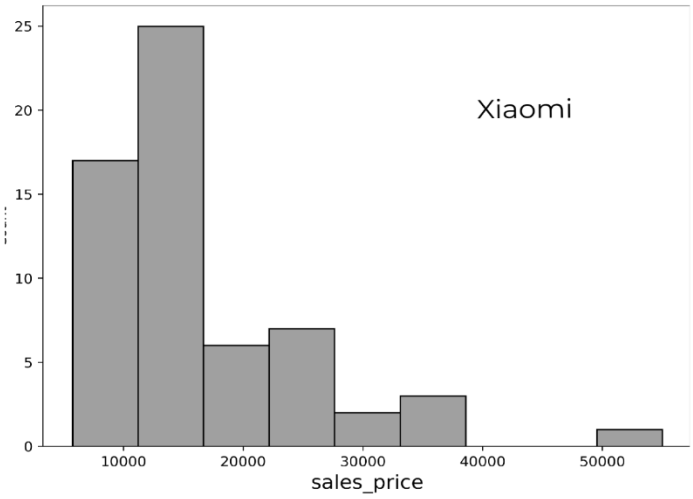
Correlation between different features:

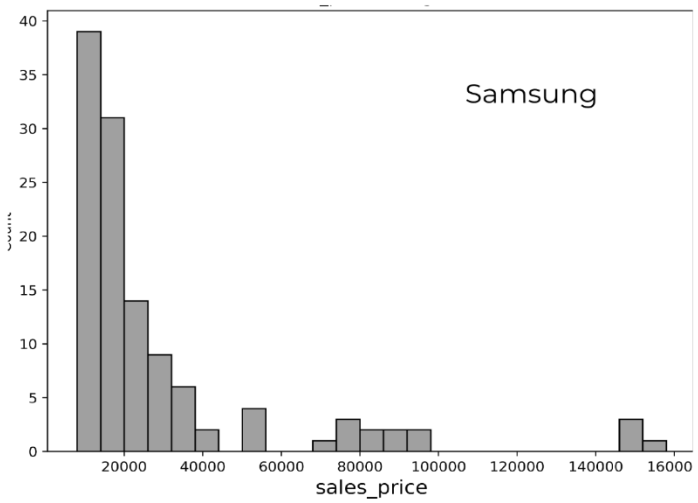
- ROM - ROM is moderately correlated with ROM and sales_price, As RAM size increase sales price and RAM size also increases
- RAM - RAM is positively related with display size and number of cameras. It is relatively less correlated with sales price.
- Display Size - As Size of phone increase number of rear camera and battery capacity also increases.
- Battery Capacity - Battery capacity opposite to what one may think, it is negatively correlated with sales price and ratings.
- Number of ratings - As sales is a derived quantity of number of ratings that's why it has very high correlation with it.

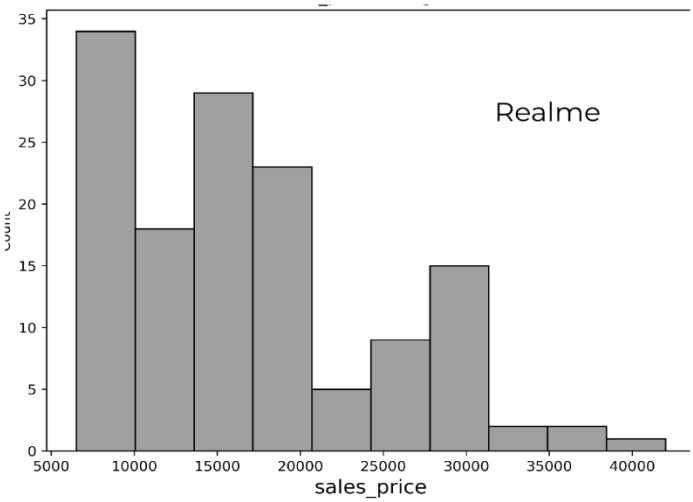
DESCRIPTIVE STATISTICS FOR SALES PRICE FOR DIFFERENT BRANDS

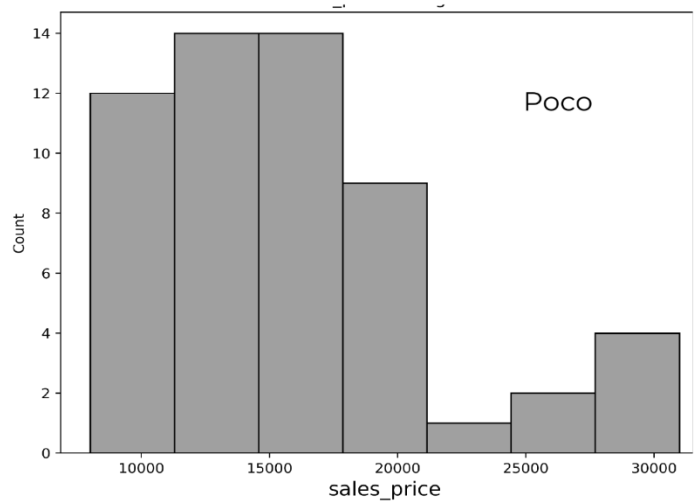
Descriptive Statistics	**Apple**	**Poco**	**Realme**	**Samsung**	**Xiaomi**
Minimum	32999	7999	6499	7990	5742
Maximum	79149	30999	41999	157999	54999
Range	46150	23000	35500	150009	49257
Count	61	56	140	120	61
Mean	56105.52	15918.94	17060.25	28616.24	16711.31
1st Quartile	43555.81	11704.56	10133.52	11058.36	10770.32
Median	57227.75	15154.56	15976.67	16490.51	13400
3rd Quartile	68733.20	19051.78	20580.67	25907.74	21545.29
Mode(1st)	43382.75	15857.33	15115.15	12365.26	12747.44
Standard Deviation	14056.74	5719.24	7718.90	30429.76	9015.50
Variance	197591930.90	32709711.50	59581444.37	925970481.46	81279334.97

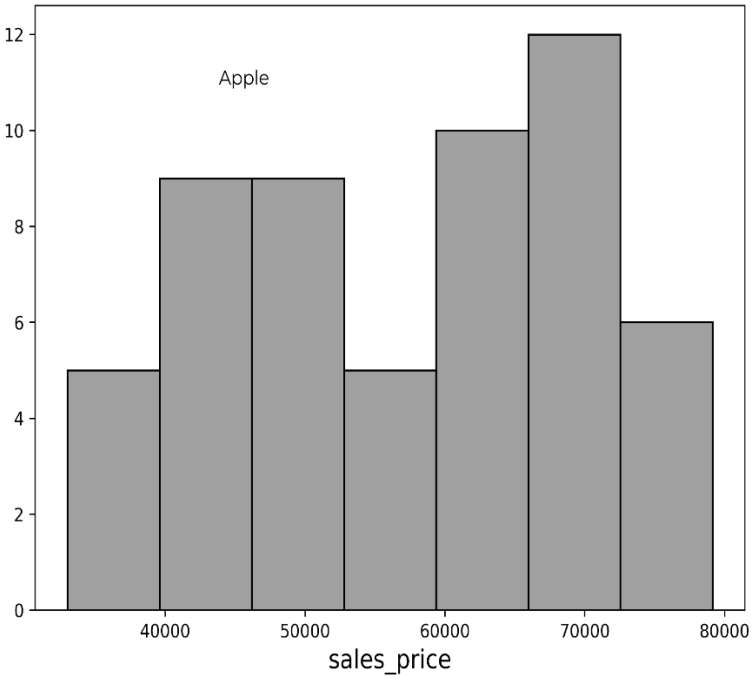
GRAPHICAL STATISTICS FOR SALES PRICE FOR DIFFERENT BRANDS

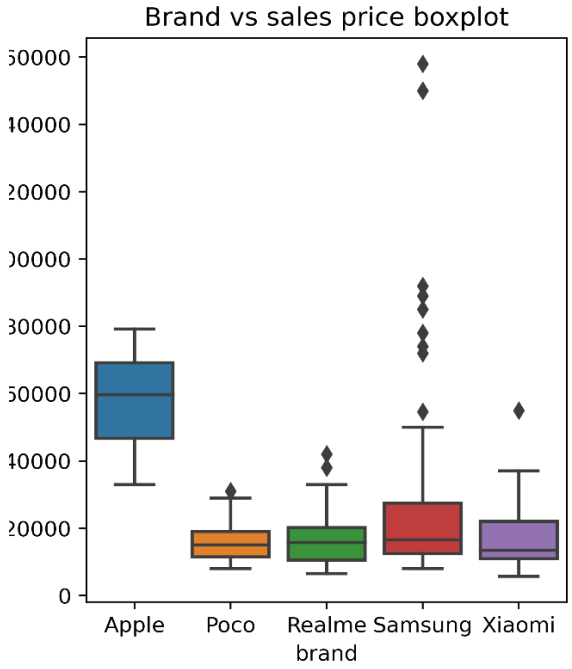












- From the table, we can see most of the mobile companies are doing good in selling budget phones (sales price less than 20000). Poco, Realme and Xiaomi operate only in low price and mid price segments and the median values suggest that 50% of their customers prefer buying phones worth less than Rs 16000.
- Poco is a relatively new entrant in the market but it has managed to capture market share with sales price varying from 7999 to 30999. Price range is only 23000 but still they have managed to sell 56 units.
- Realme is the real winner having sold the maximum number of phones. Phones are available at attractive prices with price ranging from 6499-41999.
- Samsung is the only brand which sells low-priced, medium priced and high priced mobiles with sales price varying from 7990 to 157999. Some of the Galaxy models are doing good in the higher range.
- Apple's iphone brand enjoys the most loyal customers and is the market leader in the 30-75K price segment with 92% of the existing iphone users plan to stick to the brand when they upgrade to a new phone. For Apple, only some selected models are doing good.
- For Xiaomi, 75% of the phones sold are below Rs. 21550.

CONCLUSION

According to the data, most phones sold are in the price range of $15,000 to $20,000. Poco, Xiaomi, and Realme have done well to capture the market in this segment. It is recommended that the vendor have phones from all these companies in this price range. In terms of colour, RAM, and ROM, all these brands provide a wide range of models. They've done their homework and are concentrating on the processor, display size, and colour that Indian consumers desire.

However, if getting a franchise of multiple brands is difficult and costly, then the seller should choose Poco. Because Poco is a new brand, the seller may be able to negotiate better terms with the vendor, resulting in more revenue and profit.

People in India is changing their buying patterns, and they now want to buy branded and high-end products. So, if the seller decides to sell Apple or Samsung Mobile, he may see an increase in sales and revenue over time, but it will be difficult to gain customers quickly in the current situation. The risk will also be substantial because the profit margin is unknown and the investment to be made is high.