Name: Muhibullah Noorzad

Roll/No: 19/19044

Name: Edris Azizi

Roll/No: 19/19054

**2. Problem Statement**

1. We will predict the price attribute according to all other attributes for example someone enters some property of phones like RAM, Storage, and Camera. We predict the price of the phones according to this characteristic.
2. We will predict the brand attribute according to all other attributes for example someone enters some property of phones like RAM, Storage, and Camera. We predict the price of the phones according to this characteristic.

**3. Data set, its source, and the description**

The data was retrieved from the Flipkart website. Data was obtained on September 14, 2021.

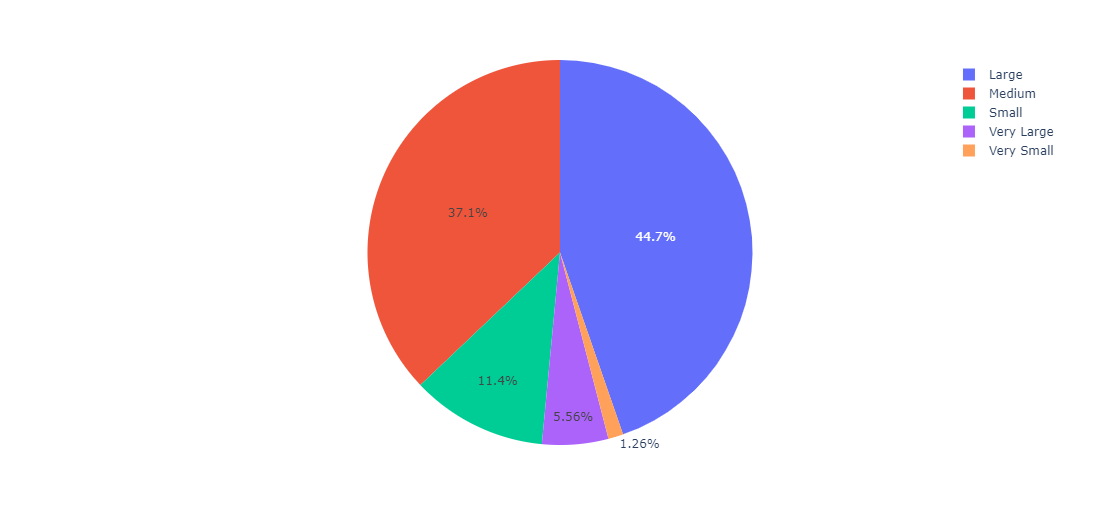
The dataset includes data on mobile phones from the top five most popular brands in India. Information like RAM, ROM, Display Size. etc. are present which distinguishes one product from another. At least one attribute distinguishes each product. Dataset has no null value. Columns: There are 16 columns each having a title that is self-explanatory. Rows: There are 430 rows each having a mobile with at least a distinct feature.

**4. IPNB File showing the summary of the dataset.**

1. brand: Brand Name (Categorical)
2. model: Model Name (Categorical)
3. base\_color: Phone Color (Categorical)
4. processor: Processor brand used (Categorical)
5. screen\_size: Categorical screen size (Categorical)
6. ROM: ROM in gigabytes (Numeric – Discrete)
7. RAM: RAM in gigabyte (Numeric – Discrete)
8. display\_size: Actual display size in inches (Numeric – Continuous)
9. num\_rear\_camera: No. of cameras on back (Numeric – Discrete)
10. num\_front\_camera: No. of cameras onthe front (Numeric – Discrete)
11. battery\_size: Battery in mAH (Numeric – Continuous)
12. ratings: Customer rating for the product (Numeric – Continuous)
13. num\_of\_ratings: No. of people rating the product, also the equivalent no. of unit sold for our problem (Numeric – Continuous)
14. sales\_price: The selling price of the unit after discount (Numeric – Continuous)
15. discount\_percent: Discount in percentage offered (Numeric – Continuous)
16. sales: Sales of product in crore rupees (Numeric – Continuous)

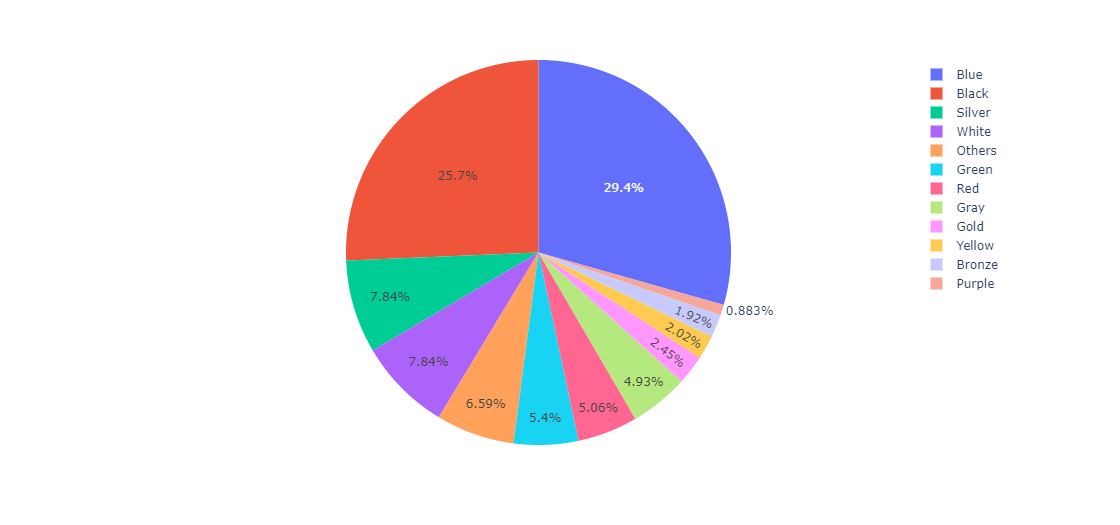
**Histogram of Attributes**

* ROM
* RAM
* Display size
* Number of rear cameras
* Number of front cameras
* Battery capacity
* Ratings
* Number of ratings
* Sales price
* Discount percent
* Sales

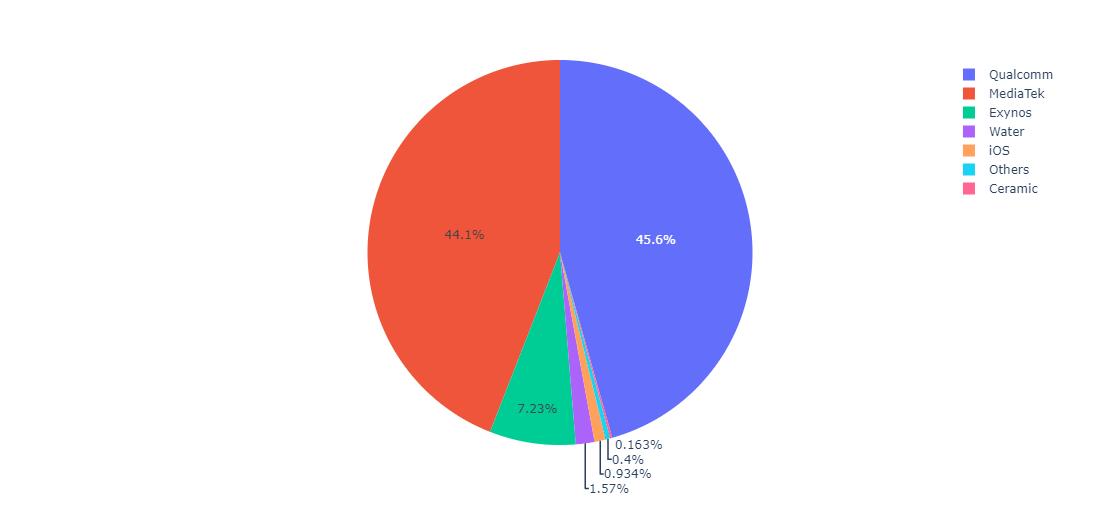
\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_

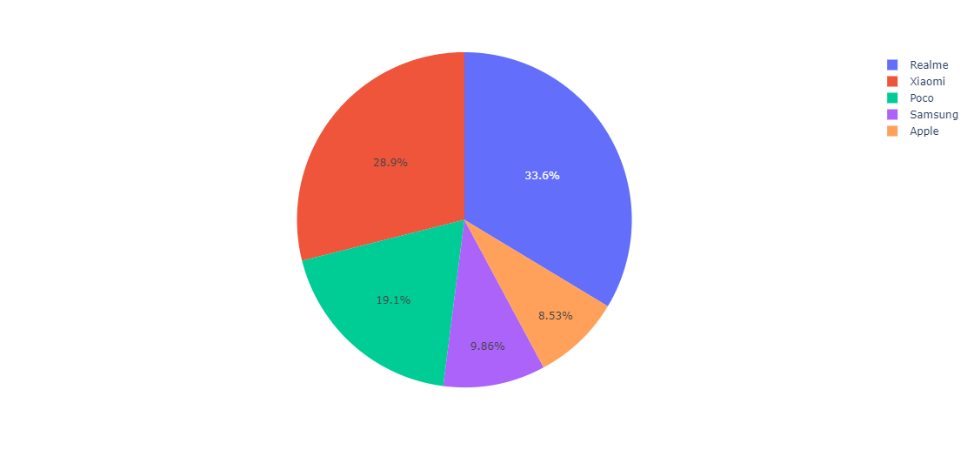
Half of the mobile products are in the large category, meaning they are larger than 6.35 inches. Because of the wide selection of products available in big and medium sizes, other display sizes are inconsistent.

\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_

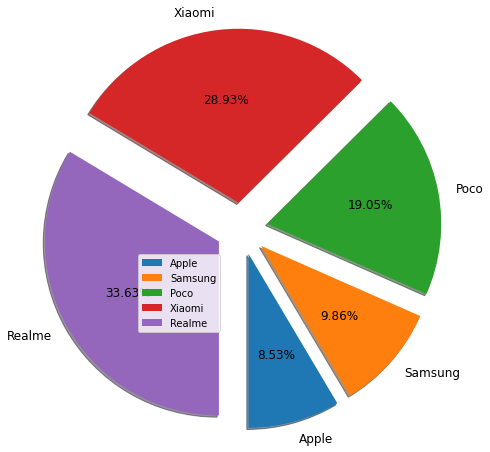
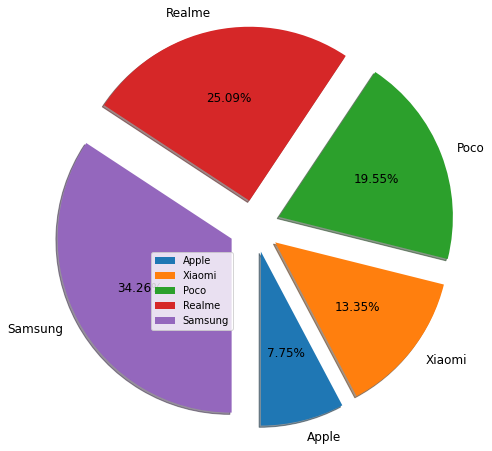
Blue is the most common color, followed by Black (Orange) and White (Green) 

\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_

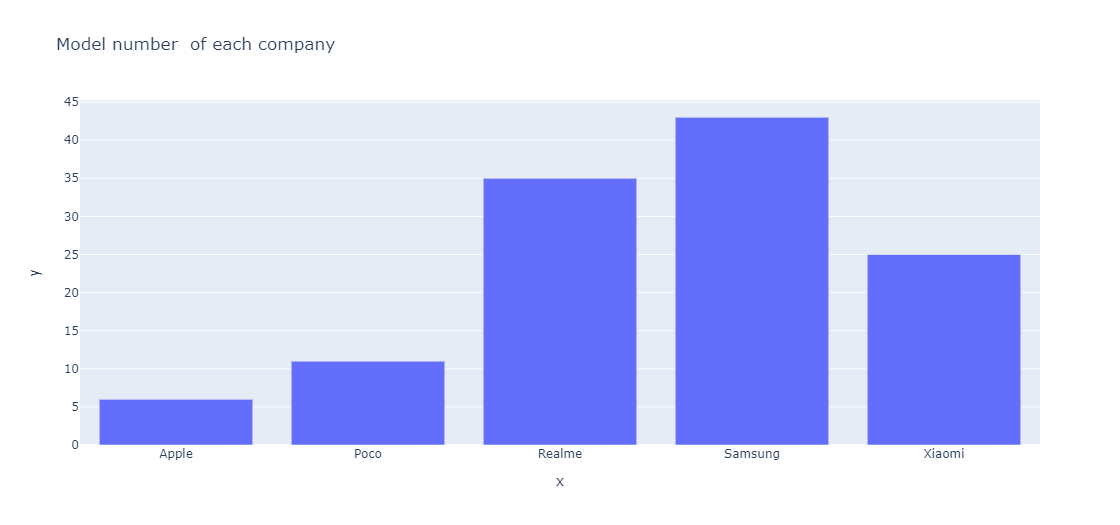
Qualcomm is the most prevalent CPU brand, accounting for 168 of the 430 mobile phones. Together, MediaTek and Qualcomm offer processors for more than half of all mobile phones.\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_

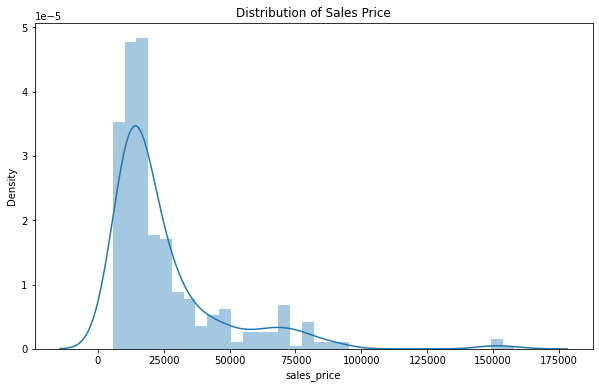
Realme offers the most options, as seen by the pie charts. In terms of specs, they have a broad range of phones to select from. Poco, on the other hand, has the fewest alternatives. This might be since Poco is a relatively new brand.

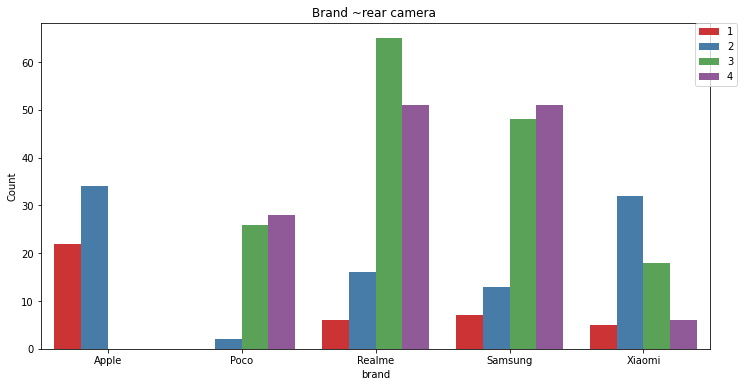
**According to selling price According to discount**



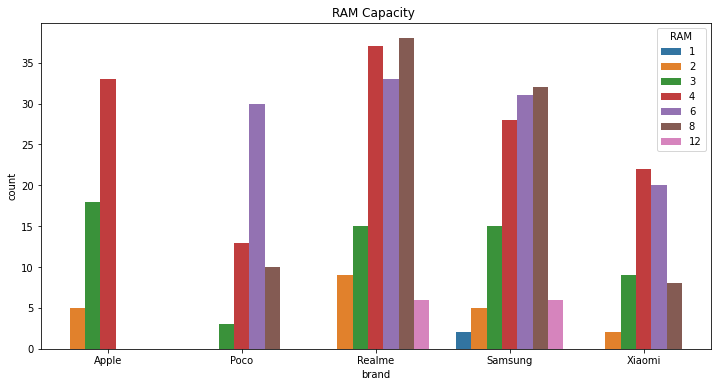
Samsung has a maximum model than other companies.



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.

\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_

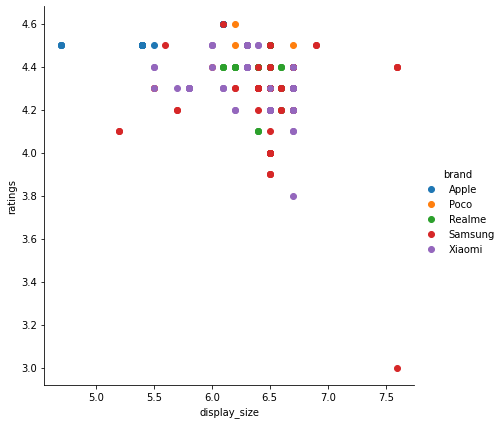
From the bar chart, we can see Apple have the most 2 and 3 rear camera. Realme have the most 3 rear camera. Samsung have the most 4 rear cameras then others companies.

\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_

As we see in bar chart, the most of phone sold the RAM capacity of 4GB and 8GB.

Samsung and Realme have the market share of 8GB RAM of phone Apple and Xiaomi have the most market share of 4GB RAM and Poco have sold most phones with 6 GB RAM.

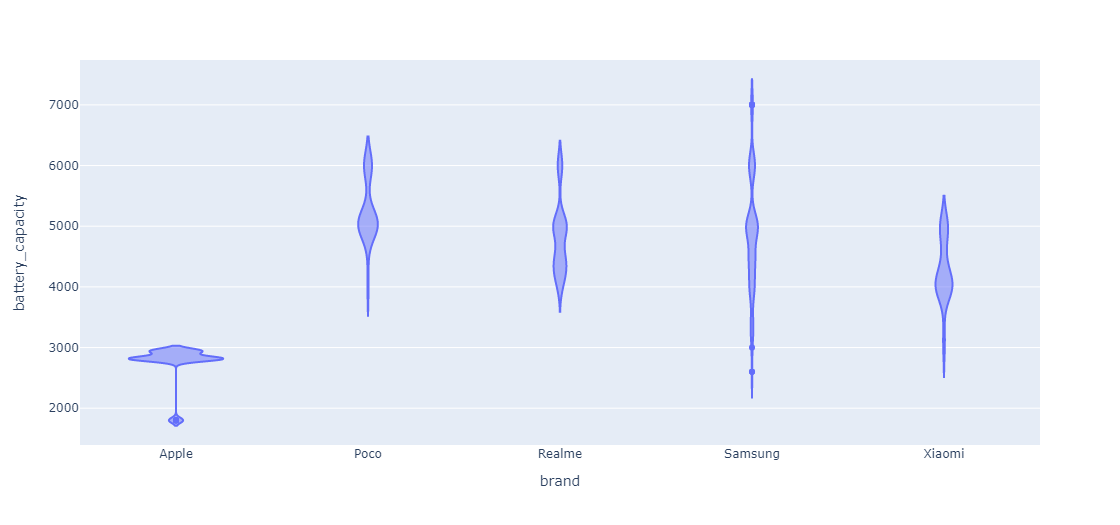
\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_

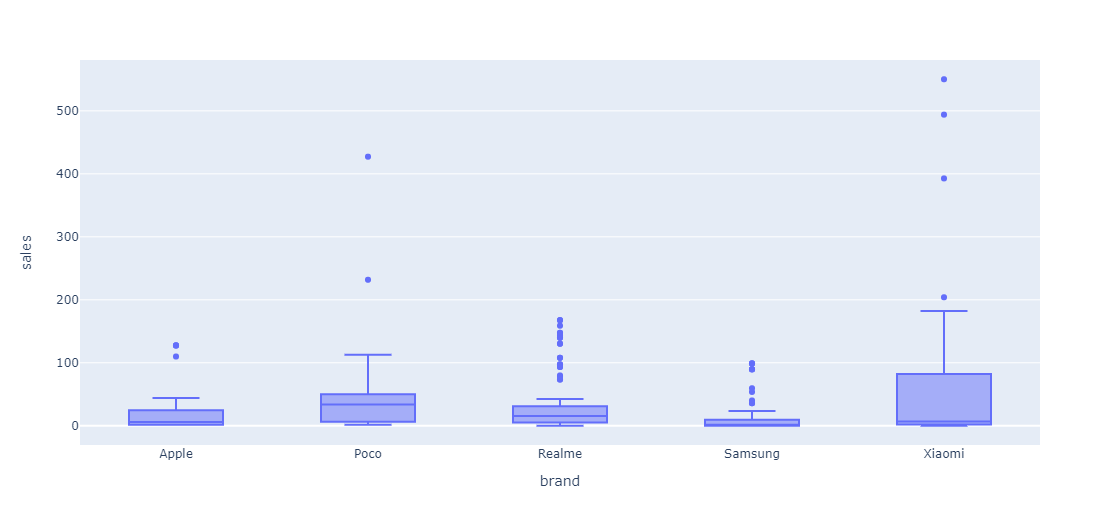
Display the size vs rating of all brands, which brand have the most 

Ratings with which display size.

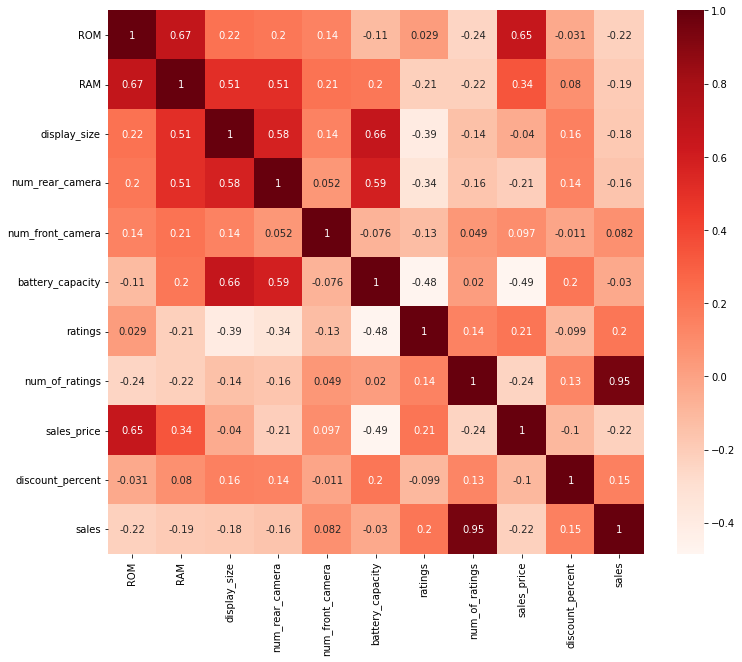
Most brands have the same display size between 6.0 and

7.0.

According this violin chart we can see each 

Brand have different battery capacity and the most of brand have battery capacity between 4000 to 6000, except apple company the most of battery of this company came with capacity of lesser then 3000. \_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_

Because of growing of market share of cellphone none of the outlier is wrong data input.

\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_ 

Here we have the correlation between all attributes.

As we can see the most correlation between RAM and ROM, the price and RAM, the number of rating,s, and price.

**Preprocessing Part**

We will map nominal values to numeric values like (brand, base color, … ) to a numeric value

We don’t need to remove any rows and columns because we don’t have any Null values and duplicate values.

We removed the model columns because it was not useful.

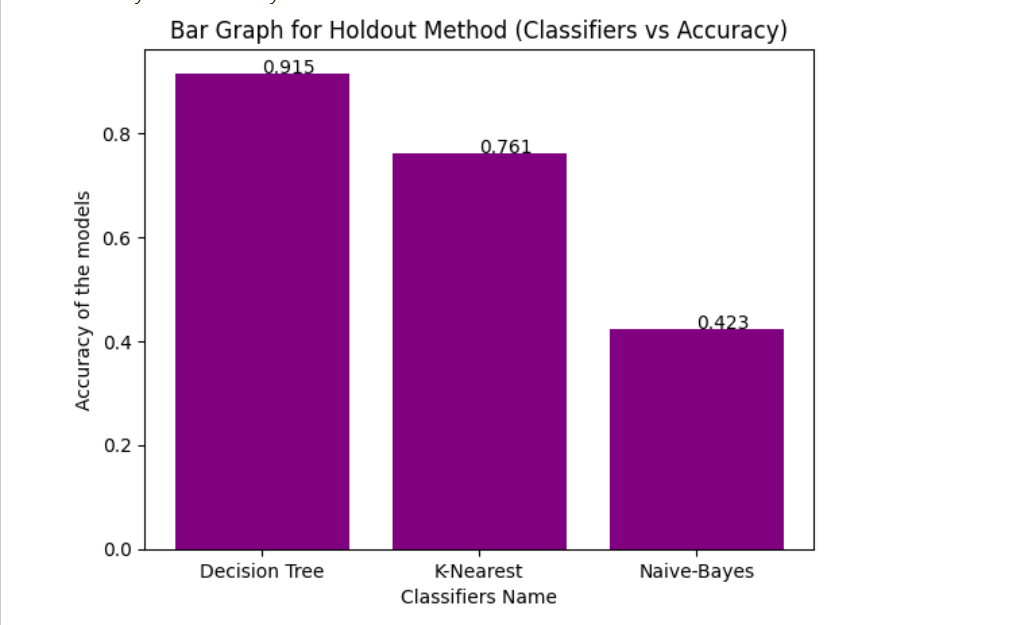
**Values Predicting part**

1. We predict according to all features of phones price and we use the holdout method and decision tree method and K nearest classifier method and Naïve Bayes and we got the most accuracy from the decision tree.
2. We predict according to all features of phones price and we use the holdout method and decision tree method and K nearest classifier method and Naïve Bayes.

We got accuracy for each method.

* Decision tree “ 0.915 ”
* K nearest ‘’ 0.761 ’’
* Naïve Bayes “ 0.57 ”

**Holdout Method Accuracy of each Algorithm**



**Random Subsampling Accuracy of each Algorithm**

