**INT353 ASSIGNMENT CA-1**

**Name: K.Ranjith Kumar Reddy**

**Reg No: 12111200**

**Roll No: RK21UNA08**

MY DATASET: MOBILE PRICES

# INTRODUCTION:

The Mobile Prices Dataset is a collection of data related to mobile phones and their pricing. This dataset provides valuable information for analysis, research, and decision-making in the mobile industry. It can be used by researchers, analysts, and businesses to gain insights into the mobile phone market, consumer preferences, and pricing strategies. The dataset was compiled from various sources, including online marketplaces, retailers, and manufacturers. It represents a diverse range of mobile phones available in the market.

This dataset has each and every review for the customers who are in search to buy a new phone. This Data set contains large variety of Mobile phone models along with their prices and processors, which helps the customers to choose their new mobile phone wisely.

# WHY?

I have taken this dataset because, people who are considering to buy a mobile phone can get a clear view or information about the processor, price etc. Also, to leverage univariate and many other analyses and show a best model using statistically significant variables from the given data set.

# DOMAIN

A "Mobile Prices" Dataset file typically falls within the domain of data related to mobile devices and their pricing. The domain encompasses information about various aspects of mobile phones, including their specifications, features, and market prices. Pricing Data: Mobile phone prices, often listed in various currencies (e.g., USD, EUR), are a central aspect of this domain.

Brand and Model Information: The dataset typically contains data on the brand (manufacturer) and specific model names or numbers of mobile phones. This information helps in categorizing and identifying different devices.

Market Availability: Details about where and when a particular mobile phone model is available in various regions or markets can be included. This is essential for market analysis and availability tracking.

User Ratings and Reviews: Some datasets may include user-generated ratings and reviews for mobile phones, providing insights into user satisfaction and opinions.

**INFORMATION**

This Dataset consists of 1837 rows and 11 columns It contains columns:

**Phone name** – Name

**Rating** – Rating of the phone out of 5

**Number of Ratings** – Total number of people who rated the particular phone

**RAM** – Random Access Memory

**ROM/Storage** – Read Only Memory, Storage capacity of the mobile phone

**Back/Rare Camera** – Back Camera

**Front Camera –** Front Camera or Selfie Camera **Battery –** Battery capacity of the mobile phone **Processor** – Processor of the mobile phone

**Price in INR –** Price of the mobile phone in Indian currency

**Date of scrapping –** Date of scrapping

# QUESTIONS/PLANS:

I will try to perform EDA with the following questions:

1. What is the average rating of phones in the dataset?
2. What are the top 10 most common phone names in the dataset?
3. How does RAM size correlate with battery capacity?
4. Is there a relationship between front and back camera resolutions?
5. How does RAM size affect phone prices?
6. Is there a correlation between the number of ratings and ratings?
7. Is there a relationship between RAM and ROM sizes?
8. How does the number of ratings impact phone prices?
9. How does the Rom impact phone prices?
10. Which is most commonly used back/rear camera in phones?
11. Which is most commonly used front camera in phones?
12. Which is most commonly used Rom in phone?
13. Which is most commonly used battery in phones?
14. Which is most commonly used processor in phone?
15. Which is most commonly used Ram in phone?
16. What is the distribution of resolutions for the back (rare) cameras in the dataset, and how does it vary across different resolution values?
17. What is the distribution of resolutions for the front cameras in the dataset, and how does it vary across different resolution values?
18. What is the distribution of battery capacities among the phones in the dataset, and how common are different capacity levels?
19. What are the phones with highest price?
20. What are the phones with least price?
21. What are the names and data types of the columns?
22. What are the basic summary statistics ?
23. Are there any categorical variables and missing values? If so print it.
24. Are there any outliers in the data? if so use box

plots, Histograms and visualize.

**CA – 2**

**Libraries used in my EDA Project:**

**PANDAS**

Pandas is a Python library for data manipulation and analysis. It provides data structures like DataFrame and Series to work with structured data easily. Pandas is commonly used for tasks such as cleaning, transforming, and analyzing datasets, making it an essential tool in data science and analysis workflows.

**NUMPY**

NumPy is a powerful Python library for numerical computing. It provides support for large, multi-dimensional arrays and matrices, along with mathematical functions to operate on these arrays efficiently. NumPy is widely used in scientific and mathematical applications and serves as a fundamental building block for many other libraries in the Python ecosystem, especially in the fields of data science and machine learning.

**MATPLOTLYB**

Matplotlib is a versatile Python library for creating static, animated, and interactive visualizations. It provides a wide range of plotting functions to generate high-quality charts, graphs, and plots. Matplotlib is often used in data analysis, scientific computing, and machine learning to visualize data and results in a clear and effective manner.

**SEABORN**

Seaborn is a Python data visualization library built on top of Matplotlib. It provides a high-level interface for creating attractive and informative statistical graphics. Seaborn simplifies the process of generating complex visualizations with concise syntax and default themes, making it particularly useful for exploring and presenting data in a visually appealing way.

**Questions and Answers**

1. What is the average rating of phones in the dataset?

Ans: The average rating of phones in the dataset is 4.27

1. What are the top 10 most common phone names in the dataset?

Ans: Most common phone names:

Brand

APPLE iPhone 14 16

APPLE iPhone 14 Plus 14

APPLE iPhone 13 13

Realme GT Master Edition 12

APPLE iPhone 14 Pro 11

Realme C55 11

APPLE iPhone SE 3rd Gen 11

OnePlus 10R 5G 11

APPLE iPhone 12 10

vivo Y16 10

1. How does RAM size correlate with battery capacity?

Ans: It has a negative correlation Ram increases Battery power capacity decreases.

4.Is there a relationship between front and back camera resolutions?

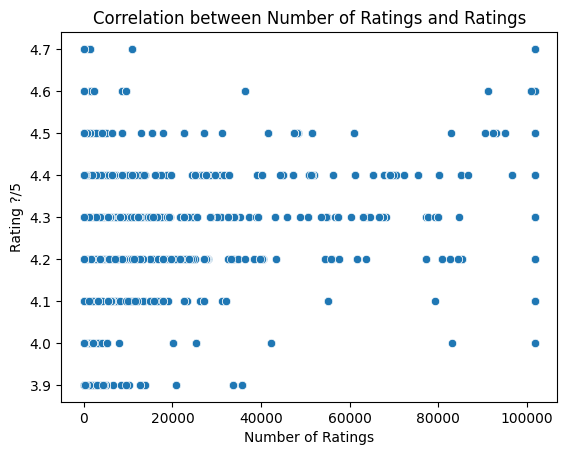
Ans: There is a weak positive correlation.

1. How does RAM size affect phone prices?

Ans: There is a high positive correlation, correlation coefficient nearing 1. As the value in one variable increases, value in another variable also increases

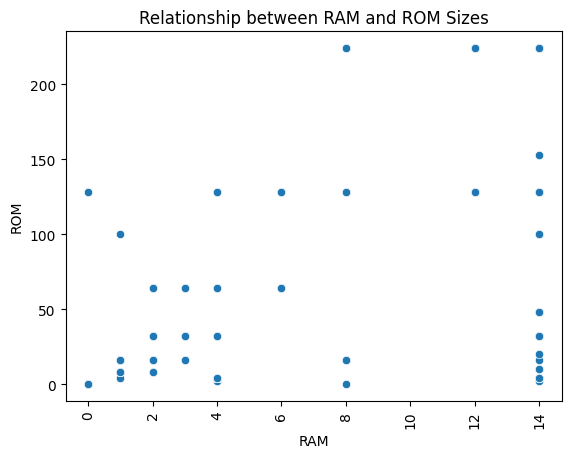
1. Is there a correlation between the number of ratings and ratings?

Ans: Most of the people rated there with 4 to 5 range.



1. Is there a relationship between RAM and ROM sizes?

Ans: There is weak positive correlation.



1. How does the number of ratings impact phone prices?

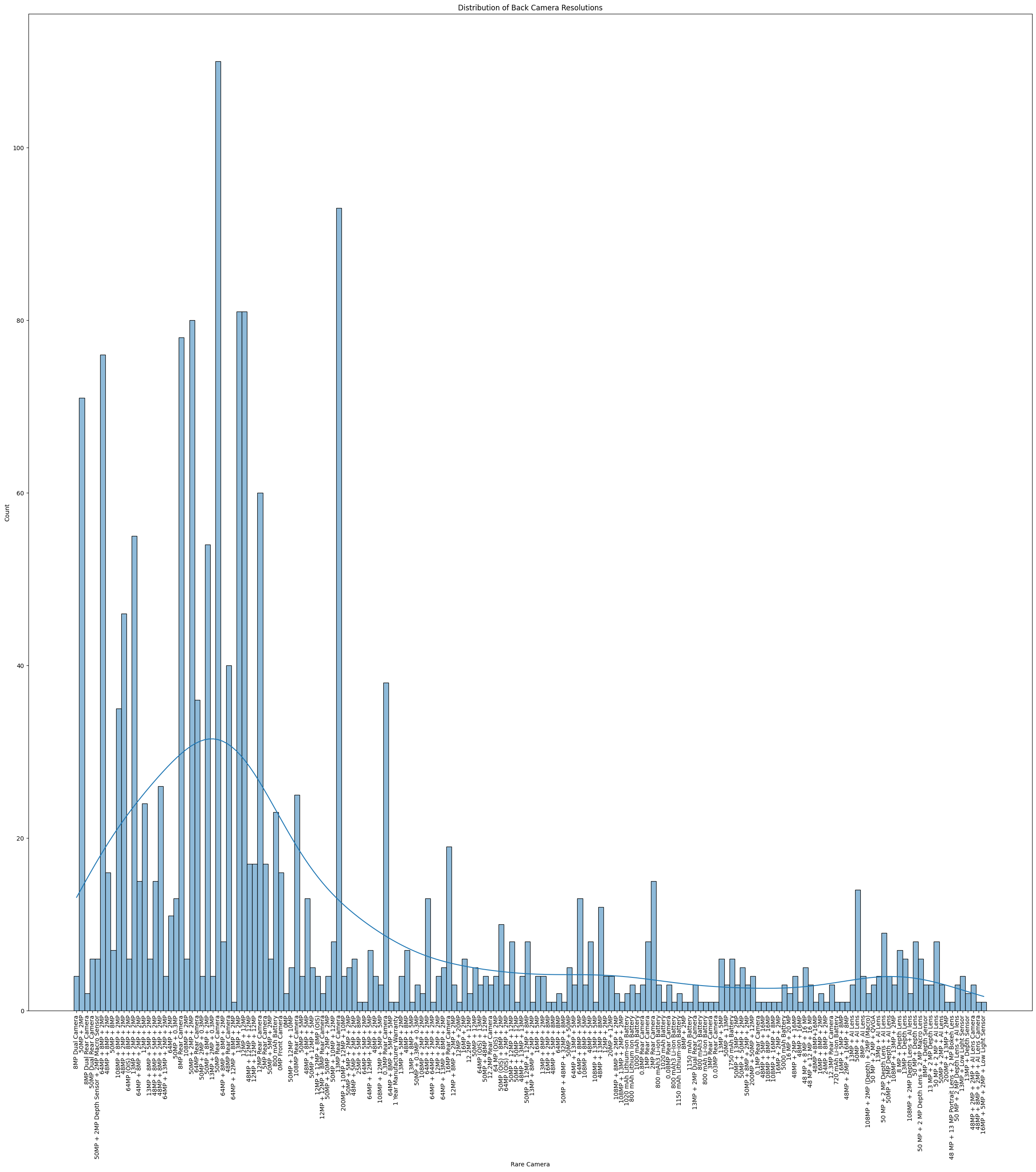
Ans: It has a positive correlation Higher the battery, Higher the price .

1. How does the Rom impact phone prices?

Ans: It has a positive correlation Higher the Rom, Higher the price.

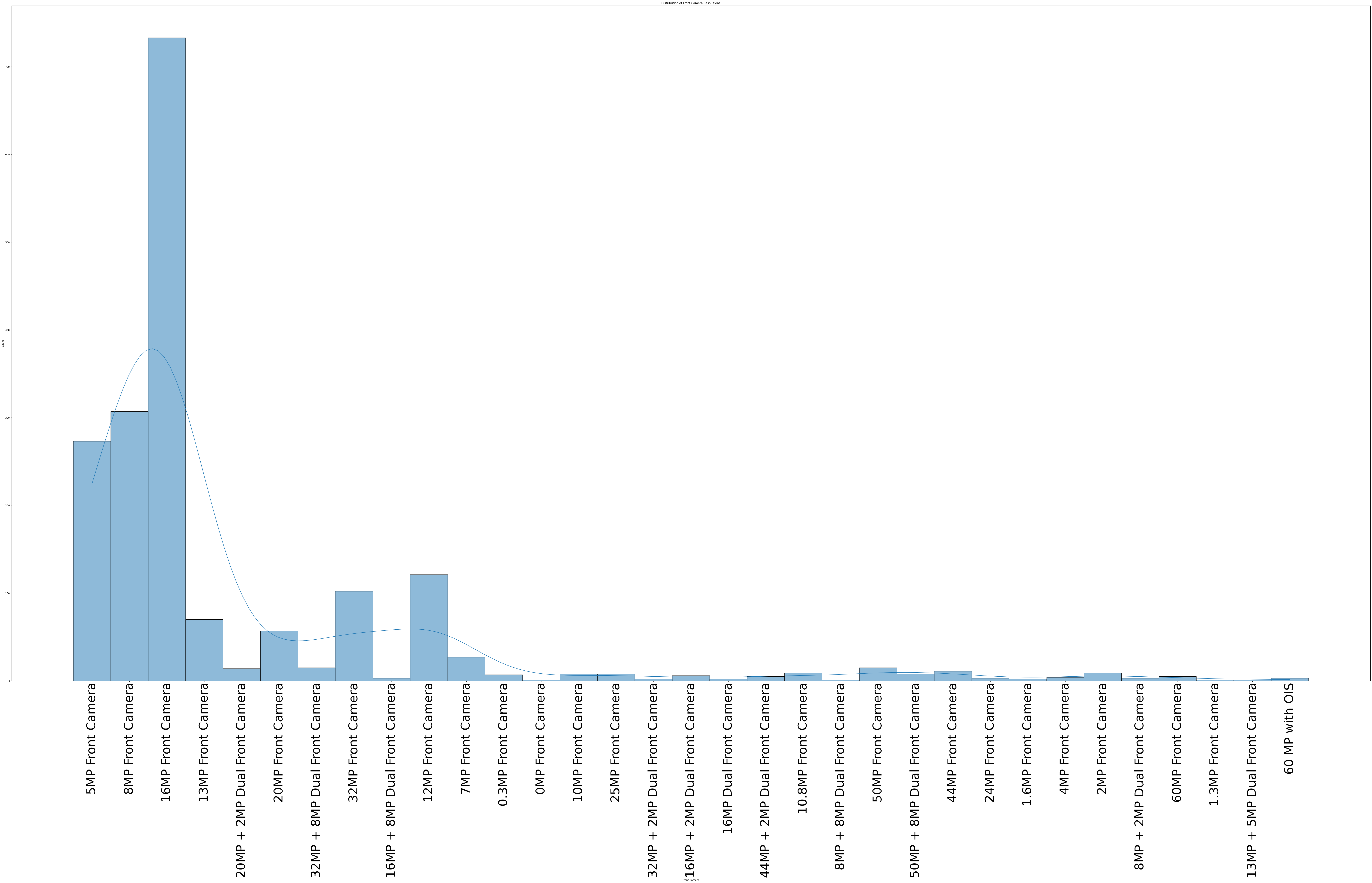
10.Which is most commonly used back/rear camera in phones?

Ans: The most commonly used back/rear camera in phones is 50 MP Rear camera.



11.Which is most commonly used front camera in phones?

Ans: The most commonly used front camera in phones is 16 MP camera.

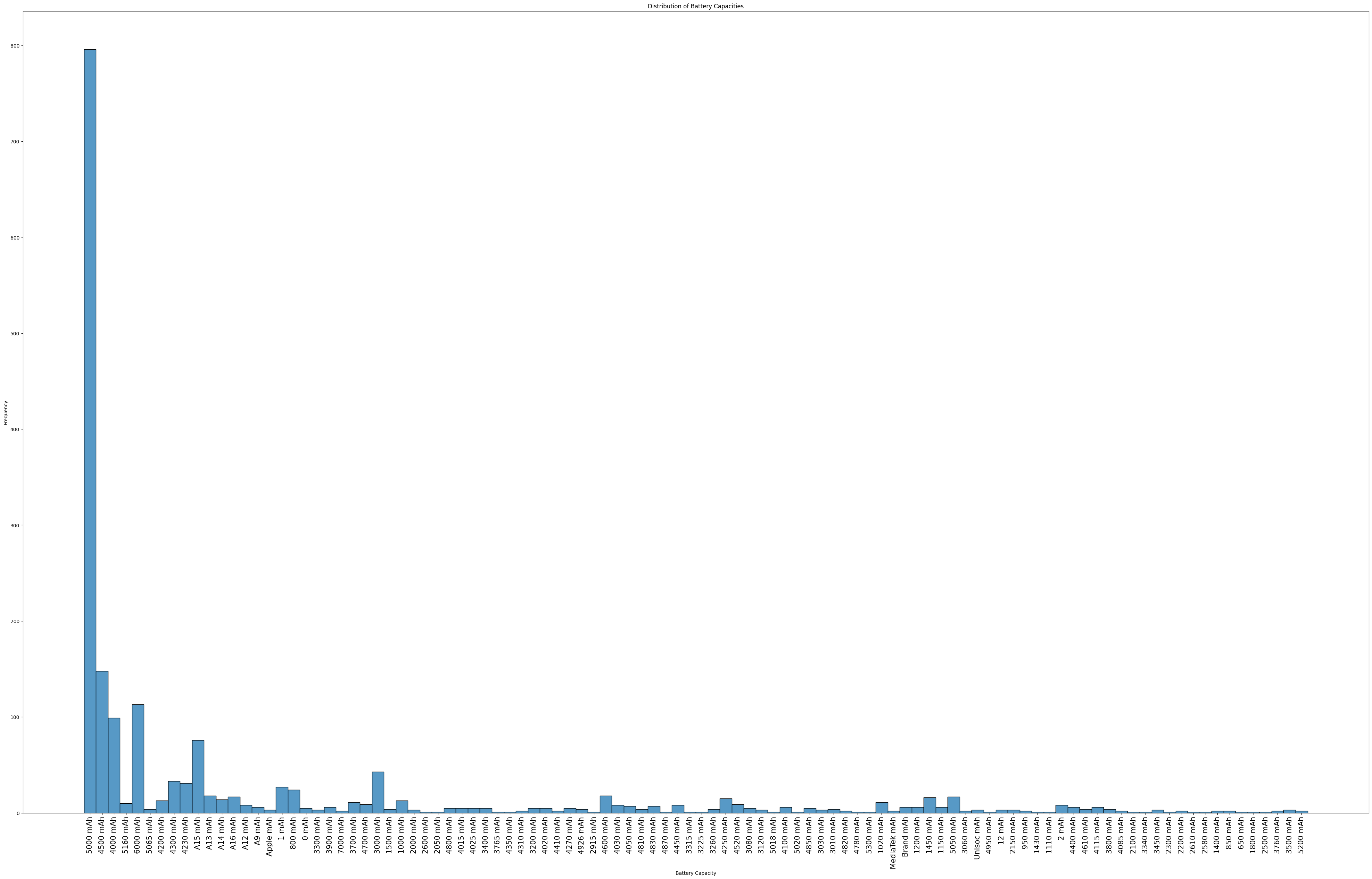


12. Which is most commonly used Rom in phone?

Ans: The most commonly used Rom in phones is 128 GB Rom.

13. Which is most commonly used battery in phones?

Ans: The most commonly used battery in phones is 5000 mah.

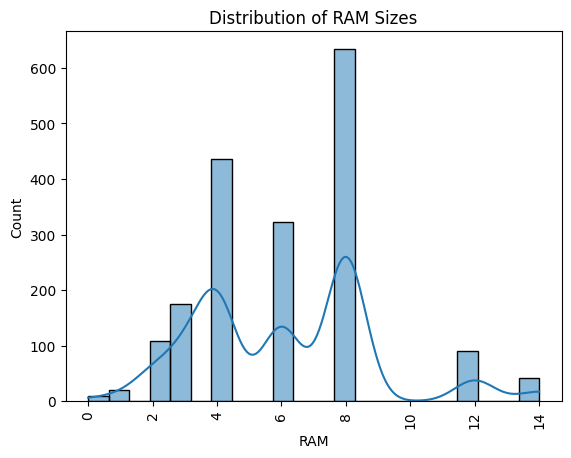


14. Which is most commonly used processor in phone?

Ans: The most commonly used processor in phones is Qualcomm Snapdragon 680 Processor.

15. Which is most commonly used Ram in phone?

Ans: The most commonly used Ram in phones is 8 Gb.



16. What is the distribution of resolutions for the back (rare) cameras in the dataset, and how does it vary across different resolution values?

Ans: The highest used back camera is 50 MP.

The lowest used back camera is 12 MP + 20 MP.

17. What is the distribution of resolutions for the front cameras in the dataset, and how does it vary across different resolution values?

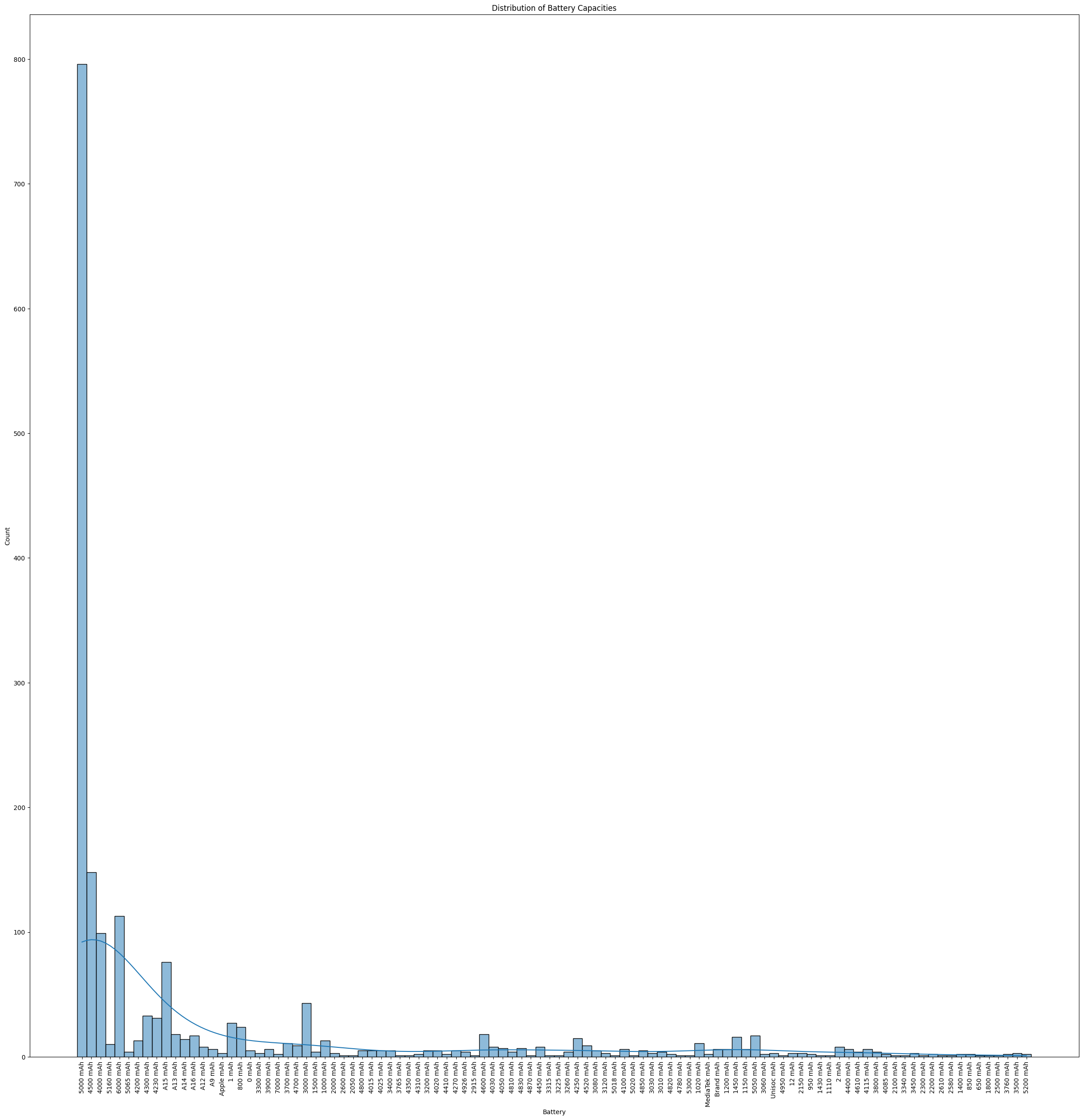
Ans: The highest used front camera is 16 MP.

The lowest used front camera is 8 MP + 8 MP.

18. What is the distribution of battery capacities among the phones in the dataset, and how common are different capacity levels?

Ans: The highest battery is 5000 MAH.

The lowest battery is 1500 MAH.



19. What are the phones with highest price?

Ans: The highest price of the phones is Rupees 1,69,999.

APPLE IPHONE 14 Pro (Silver, 1TB)

APPLE IPHONE 14 Pro (Space black, 1TB)

APPLE IPHONE 14 Pro (Gold, 1TB)

1. What are the phones with least price?

Ans: The least price of the phones is Rupees 1,199.

Nokia 105

Nokia Ta – 1010/105

21.What are the names and data types of the columns?

Ans: 0 Brand 1836 non-null object

1 Specifications 1836 non-null object

2 RAM 1836 non-null float64

3 ROM 1836 non-null float64

4 Front Camera 1836 non-null object

5 Rare Camera 1836 non-null object

6 Processor 1836 non-null object

7 Battery 1836 non-null object

8 Price 1836 non-null float64

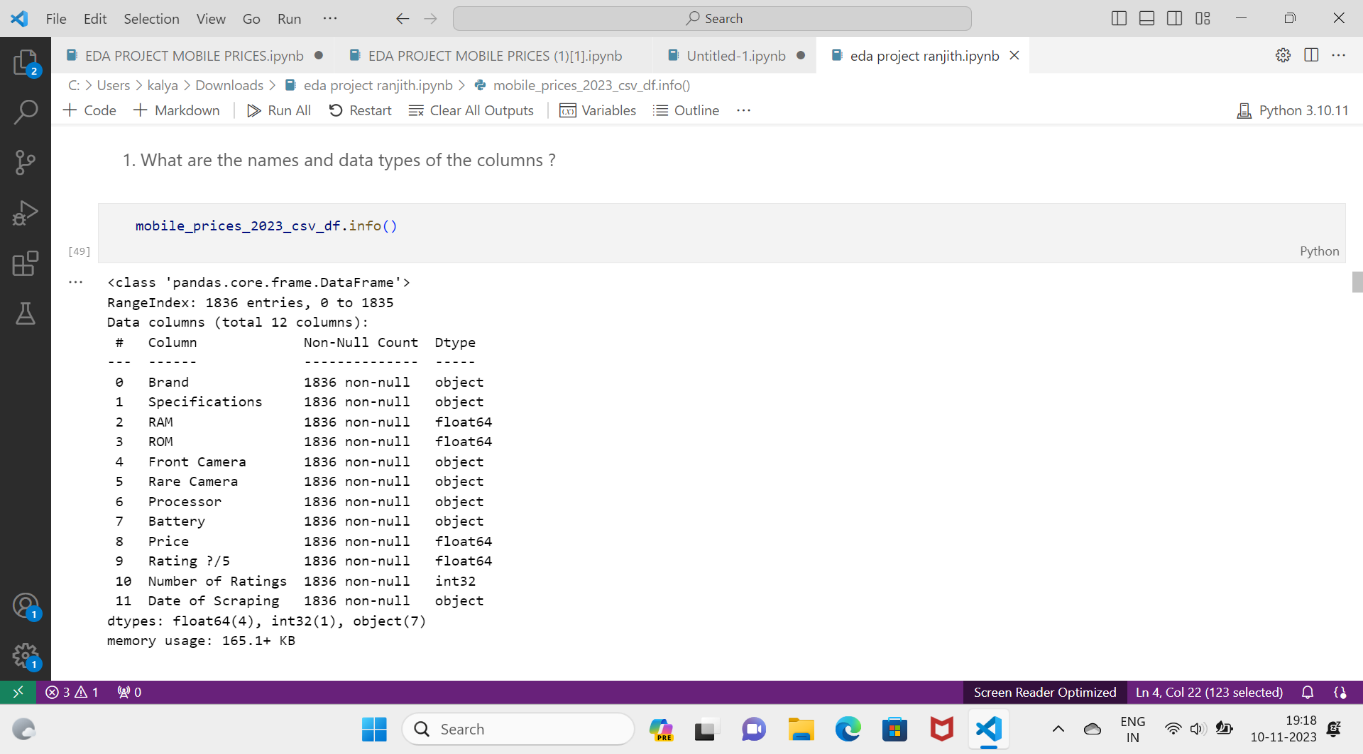
9 Rating ?/5 1836 non-null float64

10 Number of Ratings 1836 non-null int32

11 Date of Scraping 1836 non-null object

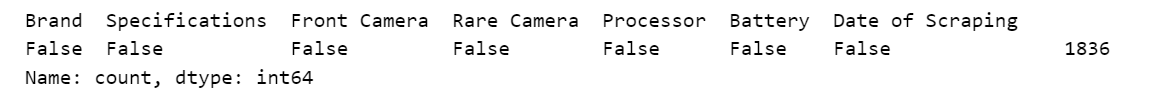
22.What are the basic summary statistics ?

Ans:



23. Are there any categorical variables and missing values?

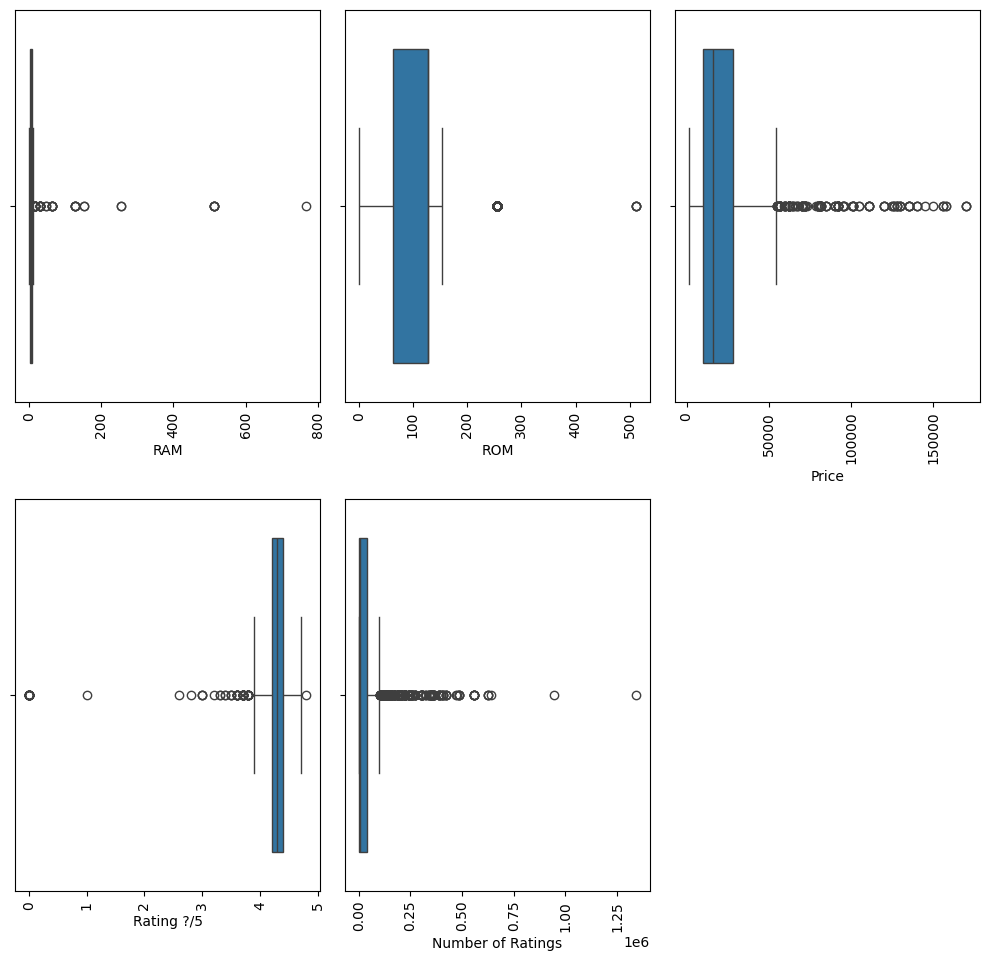
If so print it.

 Ans:

24.  Are there any outliers in the data? if so use box

plots,Histograms and visualize.

Ans:



**FINDING AND INSIGHTS :**

1. **Average Rating**:

- The average rating of phones in the dataset is 4.27.

**2. Most Common Phone Names:**

- The most common phone names and their counts are listed, with "APPLE iPhone 14" having the highest count (16).

1. **Negative Correlation:**

- There is a negative correlation between RAM and battery power capacity. As RAM increases, battery power capacity decreases.

**4. Weak Positive Correlation:**

- There is a weak positive correlation, but the variables involved are not specified.

**5. High Positive Correlation:**

- There is a high positive correlation, with a correlation coefficient nearing 1. When the value in one variable increases, the value in another variable also increases.

**6. Rating Distribution:**

- Most people rated their phones in the 4 to 5 range.

**7. Weak Positive Correlation (Another Instance):**

- There is a weak positive correlation between unspecified variables.

**8. Positive Correlation: Higher Battery, Higher Price:**

- There is a positive correlation between higher battery capacity and a higher price.

**9. Positive Correlation: Higher ROM, Higher Price:**

- There is a positive correlation between higher ROM (storage) and a higher price.

**10. Most Common Back/Rear Camera:**

- The most commonly used back/rear camera in phones is 50 MP.

**11. Most Common Front Camera:**

- The most commonly used front camera in phones is 16 MP.

**12. Most Common ROM:**

- The most commonly used ROM in phones is 128 GB.

**13. Most Common Battery Capacity:**

- The most commonly used battery capacity in phones is 5000 mAh.

**14. Most Common Processor:**

- The most commonly used processor in phones is Qualcomm Snapdragon 680 Processor.

**15. Most Common RAM:**

- The most commonly used RAM in phones is 8 GB.

**16. Camera Resolution:**

- The highest used back camera resolution is 50 MP, and the lowest is 12 MP + 20 MP.

- The highest used front camera resolution is 16 MP, and the lowest is 8 MP + 8 MP.

**17. Battery Capacity Range:**

- The highest battery capacity is 5000 mAh, and the lowest is 1500 mAh.

**18. Price Range:**

- The highest price of phones is Rupees 1,69,999, with specific models listed.

- The least expensive phones are priced at Rupees 1,199, including Nokia 105 and Nokia Ta – 1010/105.

**19. Dataset Information:**

- The dataset has 1836 entries with various features, including brand, specifications, RAM, ROM, camera details, processor, battery, price, rating, and the number of ratings.

**CONCLUSION :**

In conclusion, the Mobile Prices dataset provides valuable insights into the mobile phone market, consumer preferences, and pricing strategies. The dataset can be used by researchers, analysts, and businesses to gain insights into the mobile phone industry. The analysis of the dataset reveals interesting correlations between various features of mobile phones and their prices. The dataset can be further explored to gain more insights and make informed decisions in the mobile phone industry.

**REFERENCE :**

**1.** Kaggle

**2.** AI ASSISTANTS(BARD, CHATGPT)

**3.** DAILY DOSE OF DATA SCIENCE BLOG

**4.** AUTOMATED EDA LIBRARIES LIKE YDATA - PROFILING,AUTOVIZ

**5.** MATPLOTLIB FOR ERROR CORRECTIONS

**6.** ANALYTICS VIDHYA PLATFORM

**LINKS :**

**PPT LINK :**

[**https://docs.google.com/presentation/d/15\_yafBBToybxyrmoZML3K02rcnsBiMPO/edit?usp=sharing&ouid=102119487751632430428&rtpof=true&sd=true**](https://docs.google.com/presentation/d/15_yafBBToybxyrmoZML3K02rcnsBiMPO/edit?usp=sharing&ouid=102119487751632430428&rtpof=true&sd=true)

**DATA SET LINK :**

[**https://drive.google.com/file/d/19nDbofGg3mnRGVcdGFaq46UpNeC3fIdb/view?usp=sharing**](https://drive.google.com/file/d/19nDbofGg3mnRGVcdGFaq46UpNeC3fIdb/view?usp=sharing)

**IPYNB LINK :**

[**https://drive.google.com/file/d/1qfT7ckVMJywmxlSbakASmlGE9P0LtD4U/view?usp=sharing**](https://drive.google.com/file/d/1qfT7ckVMJywmxlSbakASmlGE9P0LtD4U/view?usp=sharing)