INNOMATICS®
RESEARCH LABS

INNOVATION. AUTOMATION. ANALYTICS

PROJECT ON

ANALYSIS OF MOBILES

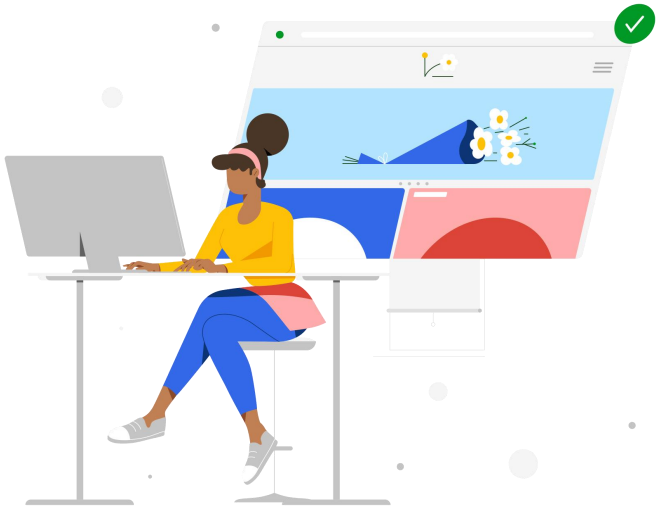
[Team members]
GEETHA SREE M
RAVALIKA K

AGENDA

- Introduction
- Project objective
- Problem statement
- Data Cleaning
- Data Visualization
- Conclusion

INTRODUCTION

- The phone market on e-commerce landscape is highly competitive with numerous brands vying for consumer attention.
- Flipkart is one of those prominent Indian e-commerce platforms.
- It offers a diverse range of products, including electronics, fashion, appliances, and more.
- The website provides a user-friendly interface:
  - Users click on product listings to view detailed information, including product specifications, prices, and customer support.
  - Users often read and contribute to product reviews and ratings. This information helps them make informed decisions.
  - Users often check for ongoing deals, discounts, and special offers on the platform.



PROJECT OBJECTIVE

- To perform a comprehensive analysis on Mobiles based on their specifications to gain insights which influence consumer preferences.
- To analyze the mobiles data using any relevant website.
- To analyze on aggregating information on product details such as prices, customer ratings, including some specifications .

PROBLEM STATEMENT

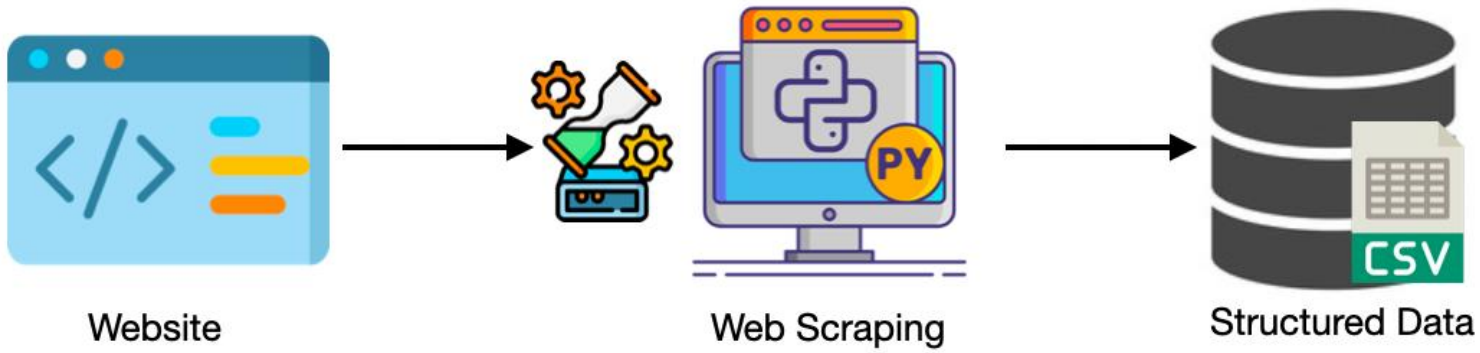
- How do the relationships between required features vary over different mobile devices?
- Identify key factors to recommend premium and budget-friendly options to consumers based on analysis.

WEB SCRAPING

Web scraping is a technique used to extract data from websites. It is also known as web harvesting or web data extraction. This can be done manually or, more commonly, using automated tools and scripts.

Steps for web scraping:

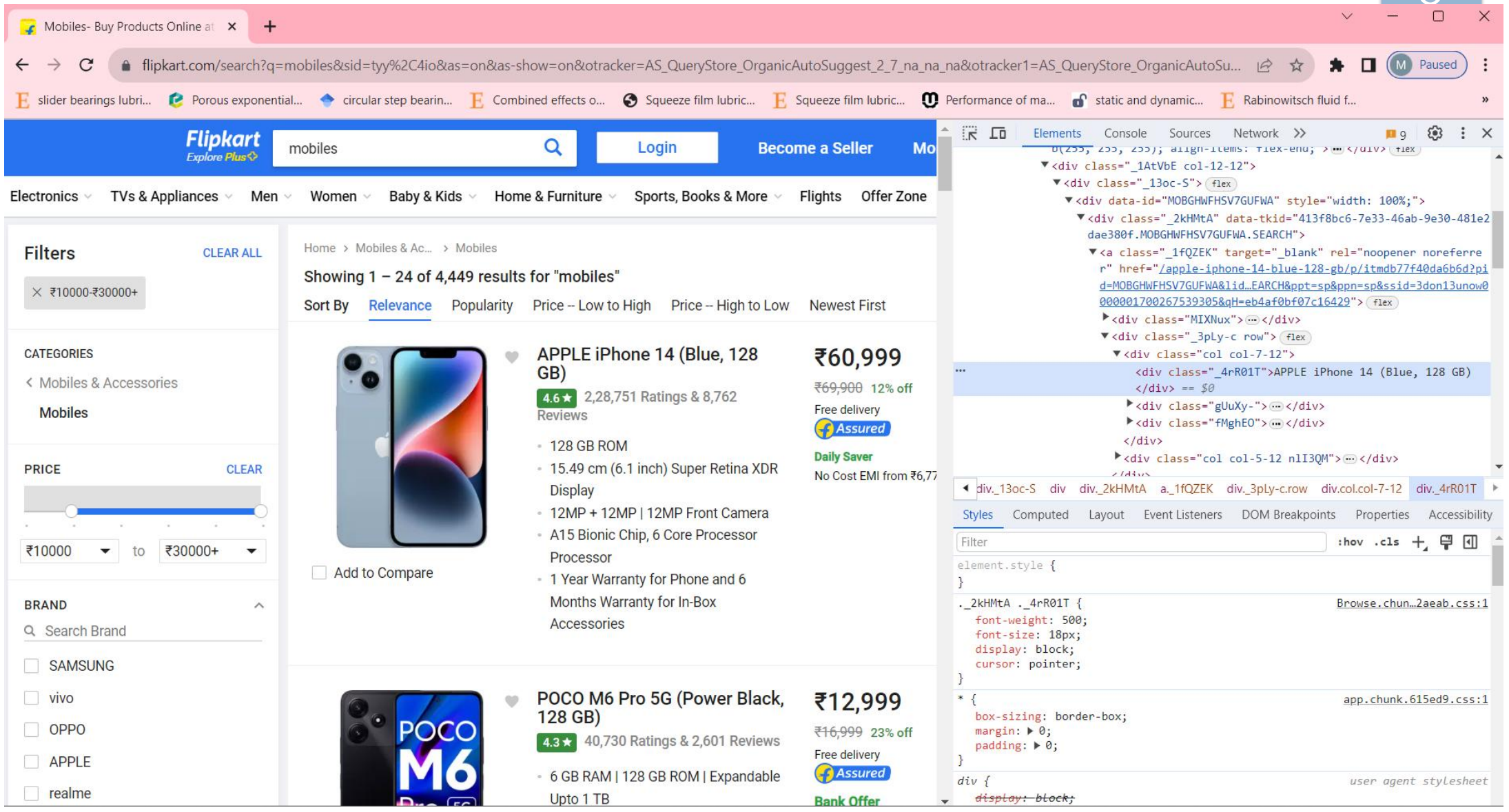
- Examine the structure of the website and inspect the page source.
- Install Necessary Libraries such as ‘Beautiful Soup’ and ‘requests’ using a package manager like pip.
- Retrieve the HTML content of the page using the library requests.
- Extract data from the HTML content by using a parsing library ‘Beautiful Soup’. This involves selecting specific HTML elements or classes.
- Finally, refine and store scraped data in a suitable format such as a CSV file depending on project requirements.



Web scraping details of the project:

- **Libraries/Techniques used**:
  - Data collection : Beautiful Soup, requests, Regex.
  - Data Cleaning : Pandas, NumPy.
  - Data Visualizaton : Matplotlib, Seaborn
- **Website scraped**: Flipkart Website
- **Link :** "https://www.flipkart.com/search?q=mobiles&sid=tyy%2C4io&as=on&as-show=on&otracker=AS_QueryStore_OrganicAutoSuggest_1_7_na_na_ps&otracker1=AS_QueryStore_OrganicAutoSuggest_1_7_na_na_ps&as-pos=1&as-type=HISTORY&suggestionId=mobiles%7CMobiles&requestId=d3382d16-c54a-4313-bdbe-040a559ac9d9"
- **Features extracted:** brand, price, discount, color, camera resolutions, memory, rating.

Interface of the website along with Inspect



Extracted data before cleaning


```
df
```

	Brand	Color	Memory	Cost	Discount	Rating	Front_cam	Main_cam1	Back_cam2
0	Cellecor	NaN	NaN	₹969	44.0	3.6	NaN	NaN	NaN
1	Good	NaN	NaN	₹1,199	40.0	3.8	NaN	NaN	NaN
2	vivo	Aurora Gold	8.0	₹14,999	28.0	4.3	8.0	50MP	2MP
3	vivo	Glimmer Black	8.0	₹14,999	28.0	4.3	8.0	50MP	2MP
4	vivo	Marine Blue	8.0	₹14,999	28.0	4.3	8.0	50MP	2MP
...	...	...	...	...	...	...	...	...	...
907	APPLE	Pink	NaN	₹1,19,900	NaN	4.7	12.0	48MP	12MP
908	APPLE	Graphite	NaN	₹1,79,900	NaN	4.6	12.0	12MP	12MP
909	Kechaoda	NaN	NaN	₹1,249	37.0	3.9	NaN	NaN	NaN
910	Xiaomi	Purple Mist	8.0	₹21,999	35.0	4.2	16.0	108MP	8MP
911	Tecno	Endless Black	64.0	₹7,999	11.0	4.4	NaN	NaN	NaN

912 rows × 9 columns

```
df.isnull().sum()
```

```
Brand          0
Color        410
Memory       460
Cost           0
Discount      68
Rating        12
Front_cam    502
Main_cam1    616
Back_cam2    616
dtype: int64
```

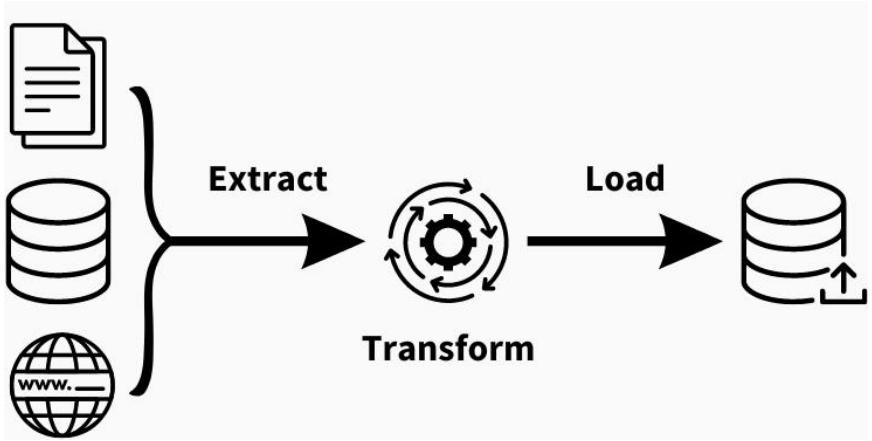
```
Percentage of NaN values before cleaning:
Brand         0.0
Color        45.0
Memory       50.4
Cost          0.0
Discount      7.5
Rating        1.3
Front_cam    55.0
Main_cam1    67.5
Back_cam2    67.5
dtype: float64
```

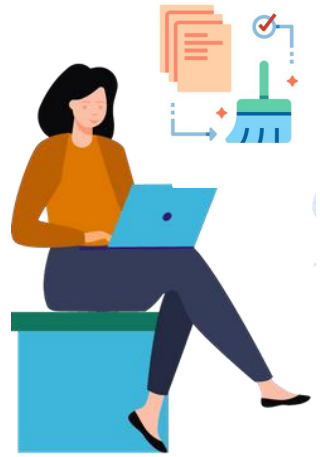
DATA CLEANING:

The goal of data cleaning is to enhance the accuracy and usefulness of the data for analysis, reporting, and decision-making.

Steps for data cleaning:

- Import necessary libraries.
- Remove special characters.
- Identify and handle missing values.
- Identify and remove required duplicate rows .
- Ensure that data types are appropriate for each variable.
- Handle Outliers and save the cleaned data to a new CSV file.





Extracted data after cleaning

df

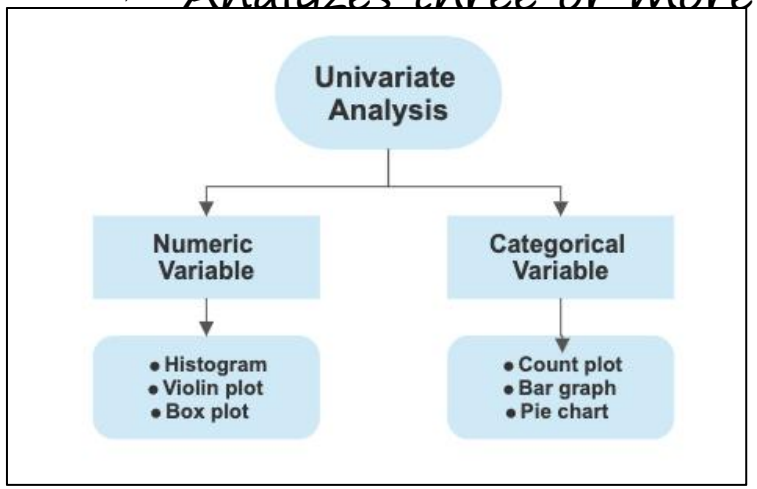
	Brand	Color	Memory	Cost	Discount	Rating	Front_cam	Main_cam1	Back_cam2
0	Cellecor	Blue	6	969	44	3.6	13	50MP	2MP
1	Good	Blue	6	1199	40	3.8	13	50MP	2MP
2	vivo	Aurora Gold	8	14999	28	4.3	8	50MP	2MP
3	vivo	Glimmer Black	8	14999	28	4.3	8	50MP	2MP
4	vivo	Marine Blue	8	14999	28	4.3	8	50MP	2MP
...	...	...	...	...	...	...	...	...	...
906	KARBONN	Blue	6	1367	23	3.9	13	50MP	2MP
908	APPLE	Graphite	6	179900	28	4.6	12	12MP	12MP
909	Kechaoda	Blue	6	1249	37	3.9	13	50MP	2MP
910	Xiaomi	Purple Mist	8	21999	35	4.2	16	108MP	8MP
911	Tecno	Endless Black	64	7999	11	4.4	13	50MP	2MP

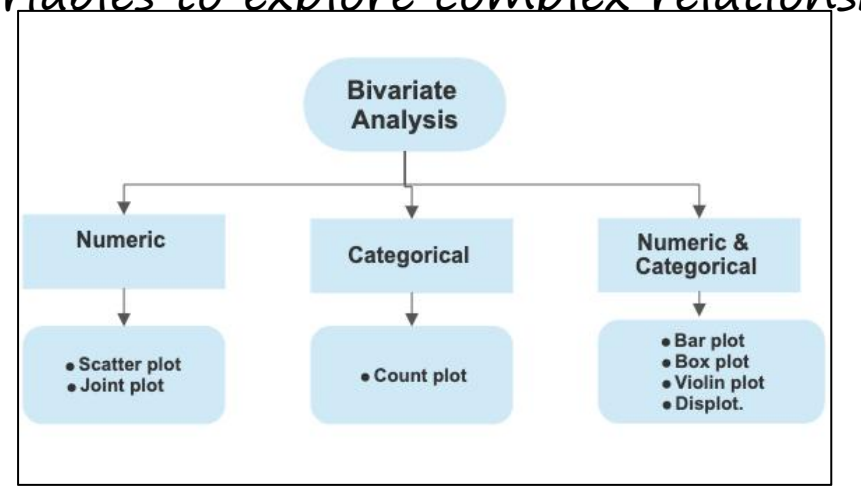
569 rows × 9 columns

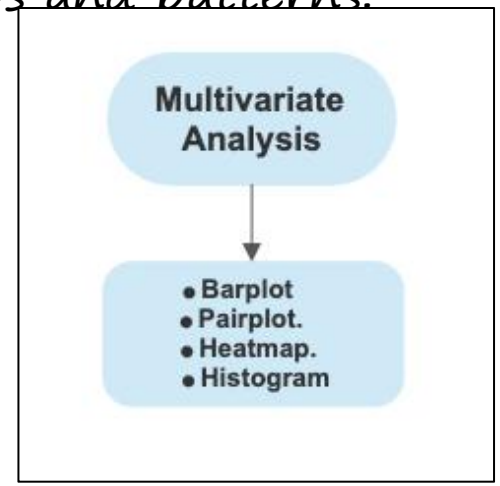
```
Percentage of NaN values in after cleaning:
Brand        0.0
Color        0.0
Memory       0.0
Cost         0.0
Discount     0.0
Rating       0.0
Front_cam    0.0
Main_cam1    0.0
Back_cam2    0.0
dtype: float64
```

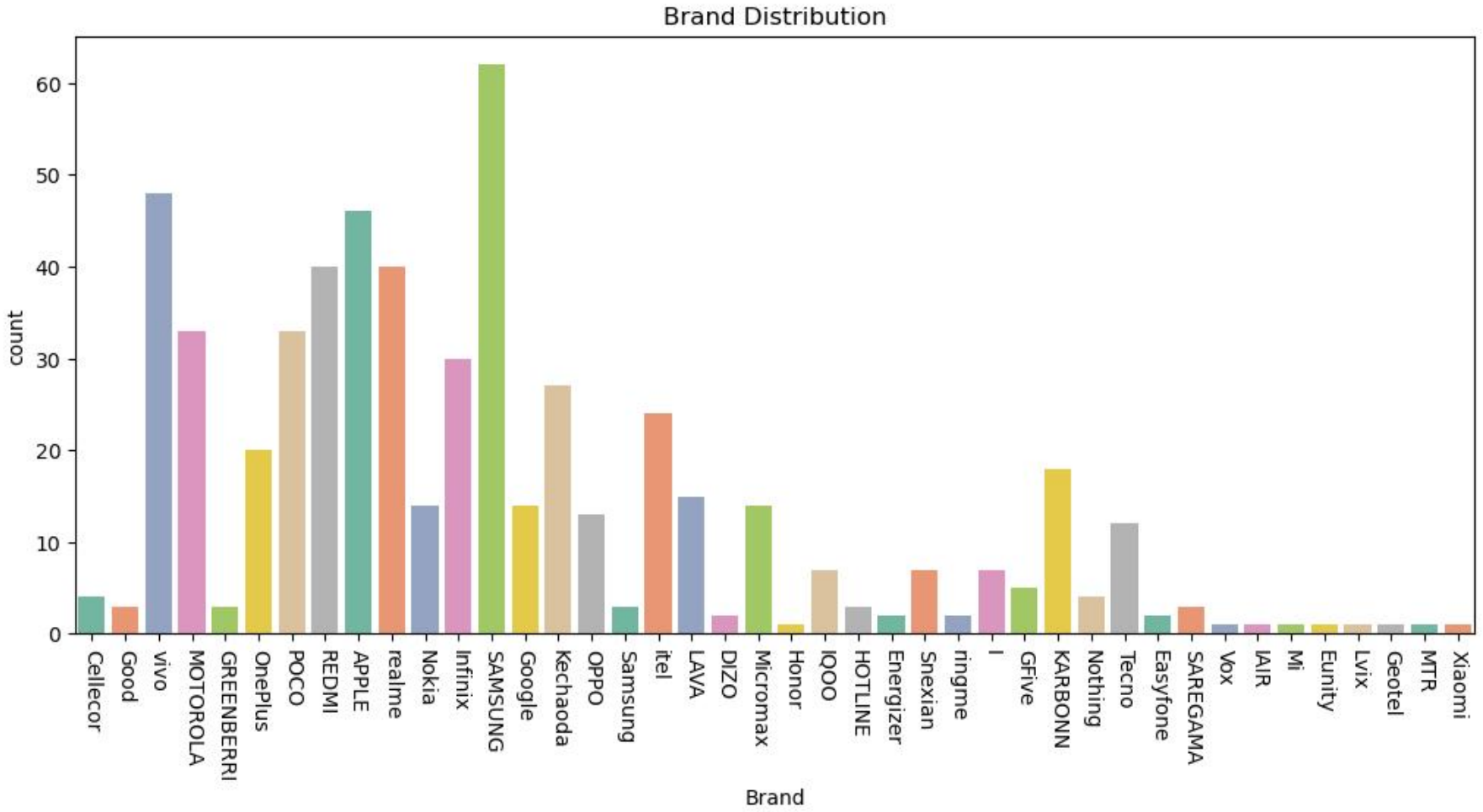
DATA VISUALIZATION:

- Data visualization is the representation of data in graphical or visual formats to facilitate better understanding.
- *Univariate analysis :*
  - Deals with a single variable by focusing on its characteristics.
- *Bivariate analysis :*
  - Involves the study of two variables to understand their relationship.
- *Multivariate analysis :*
  - Analyzes three or more variables to explore complex relationships and patterns.

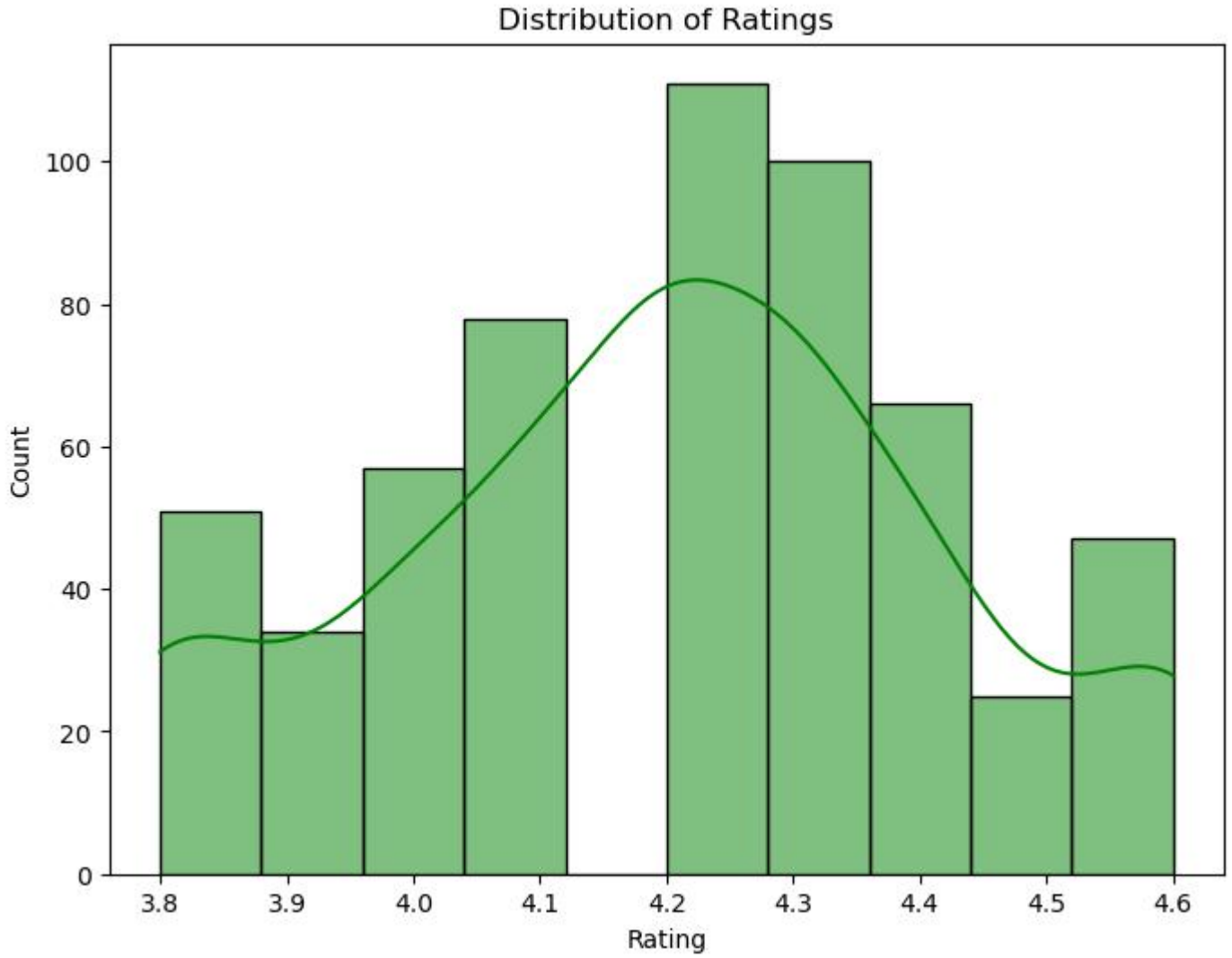




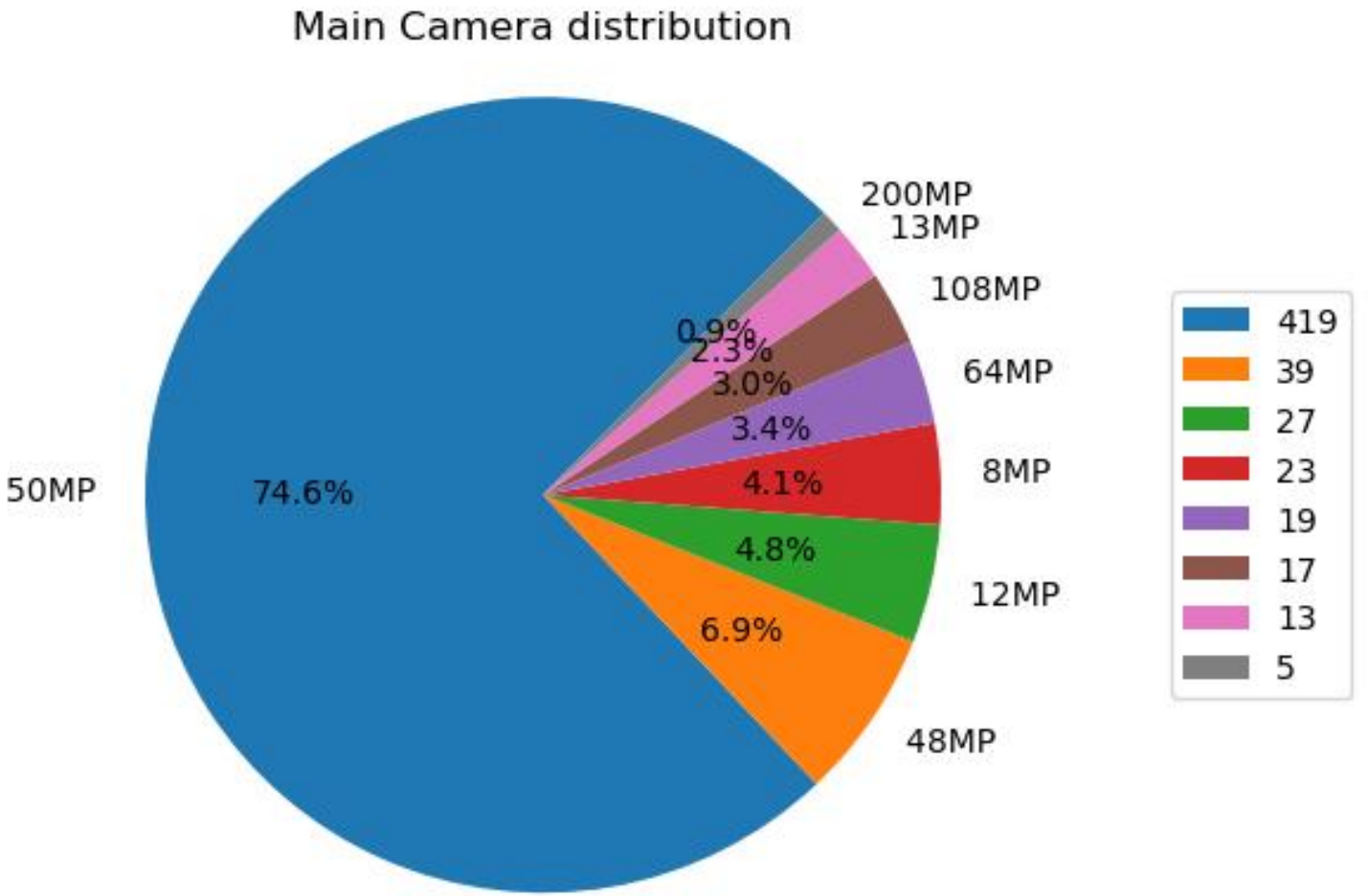




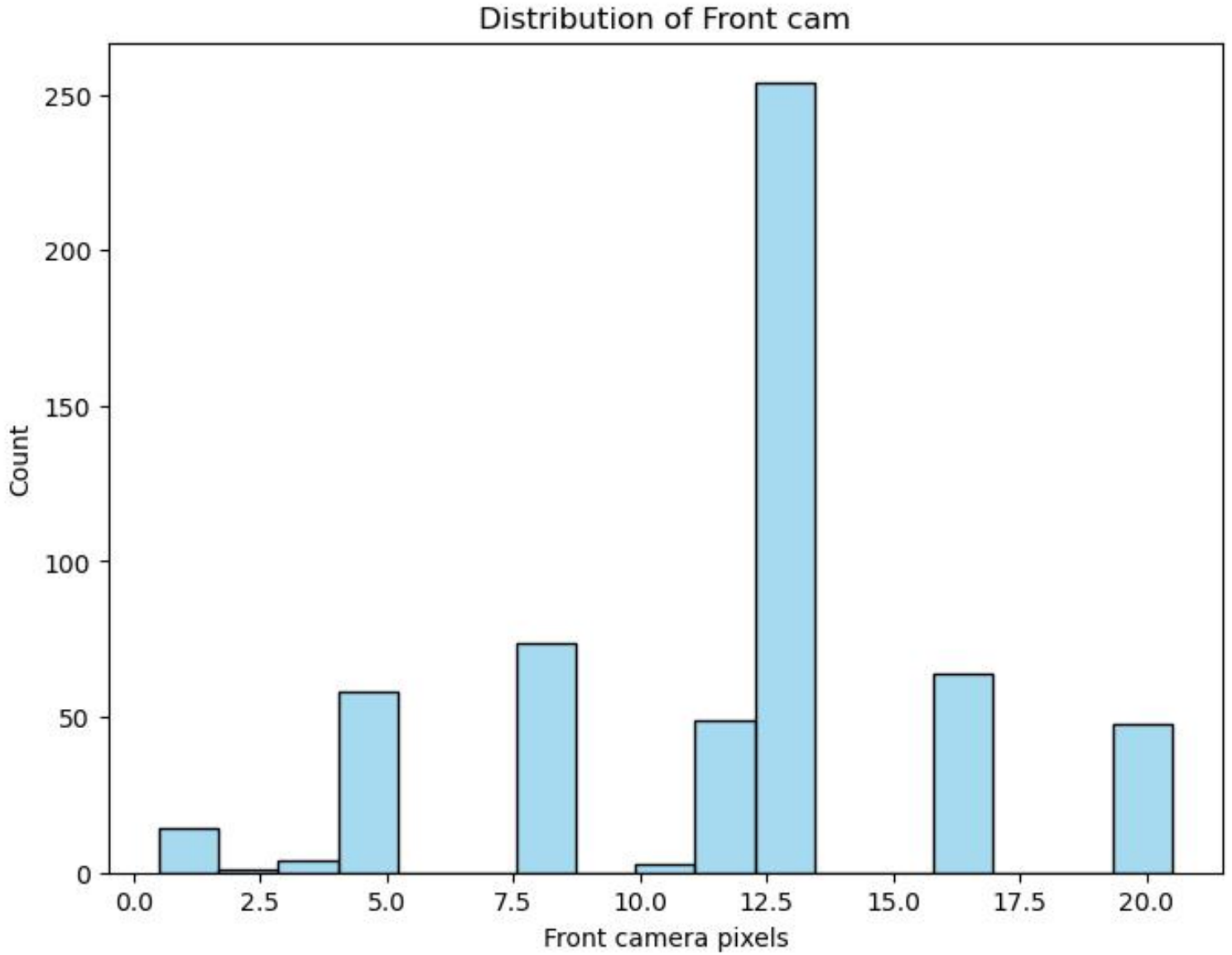
- ❑ It is observed that,
- ✓ It is visualizing the distribution of a single categorical variable named ‘Brand’.
- ✓ we can conclude that the most of the mobiles belongs to SAMSUNG brand compared to other brands from our dataset.



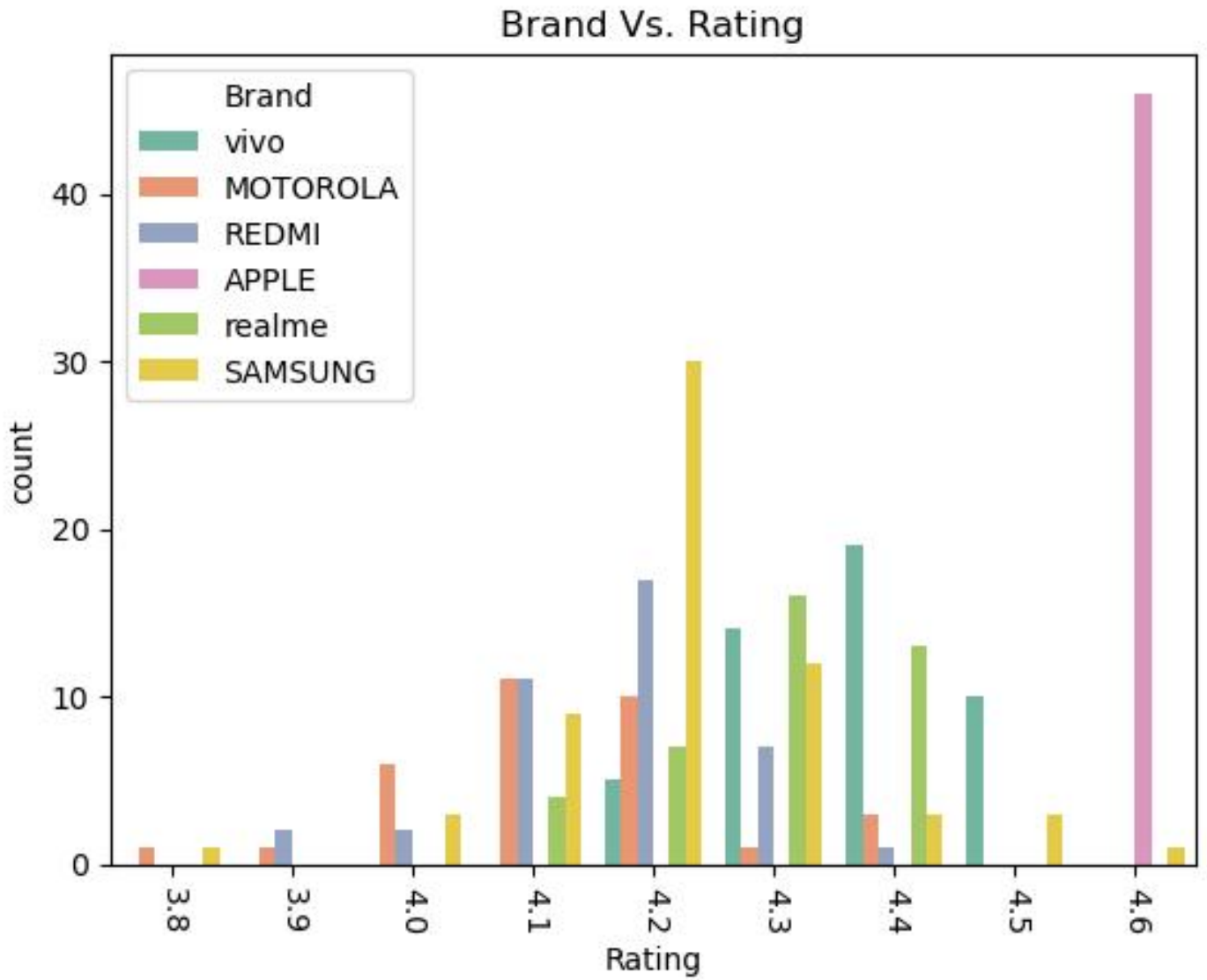
- ❑ It is observed that,
- ✓ It is visualizing the distribution of a single categorical variable named ‘Ratings’.
- ✓ It depicts the information about the highest ratings of the mobiles.
- ✓ we can conclude that the most of the mobiles have highest rating at ‘4.2’ and ‘4.3’ ratings as compared to others.



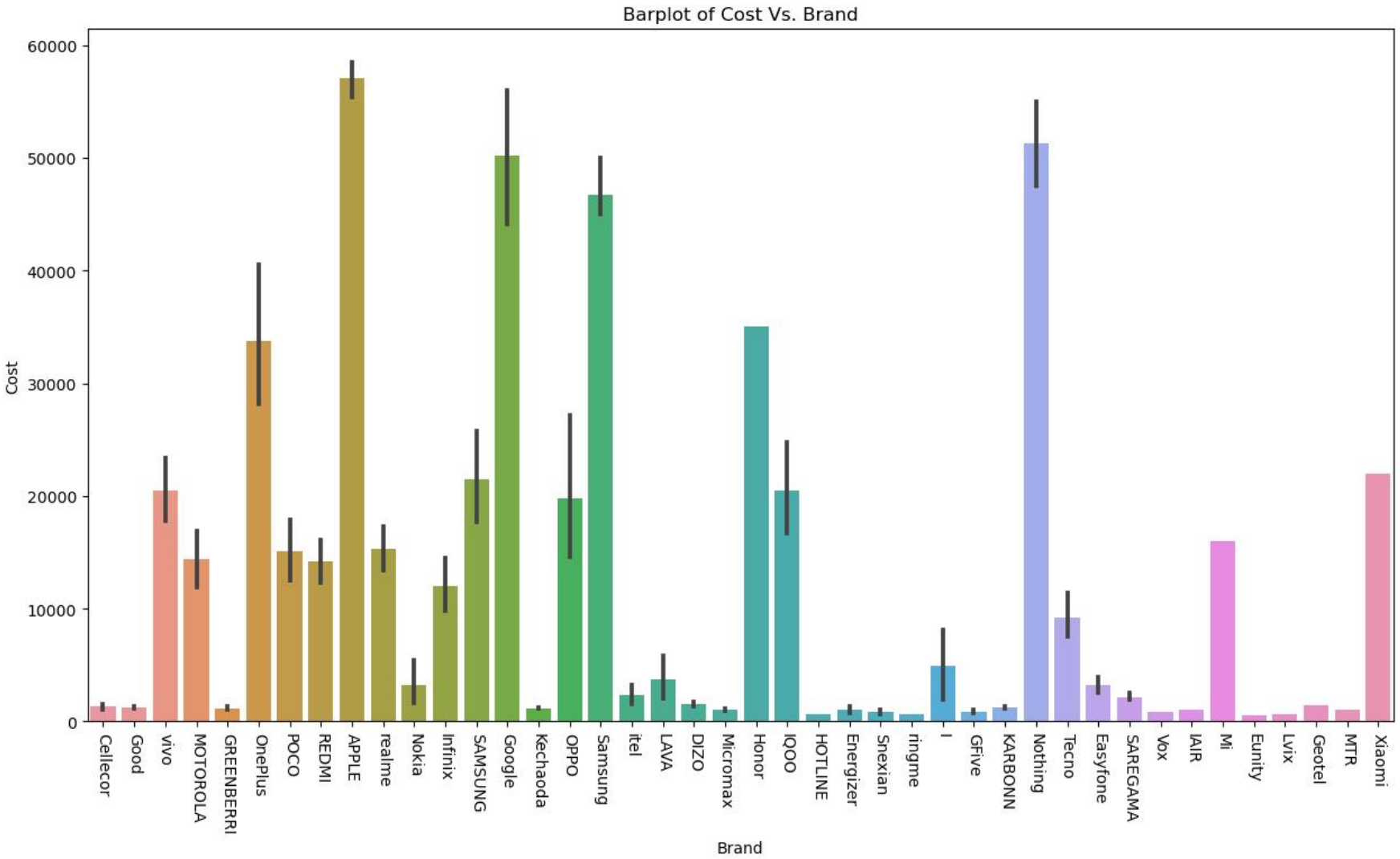
- ❑ It is observed that,
- ✓ It is visualizing the distribution of a single categorical variable named 'Main_cam1'.
- ✓ we can conclude that the mobile main *camera resolutions are most prevalent at 50MP based on frequency .*



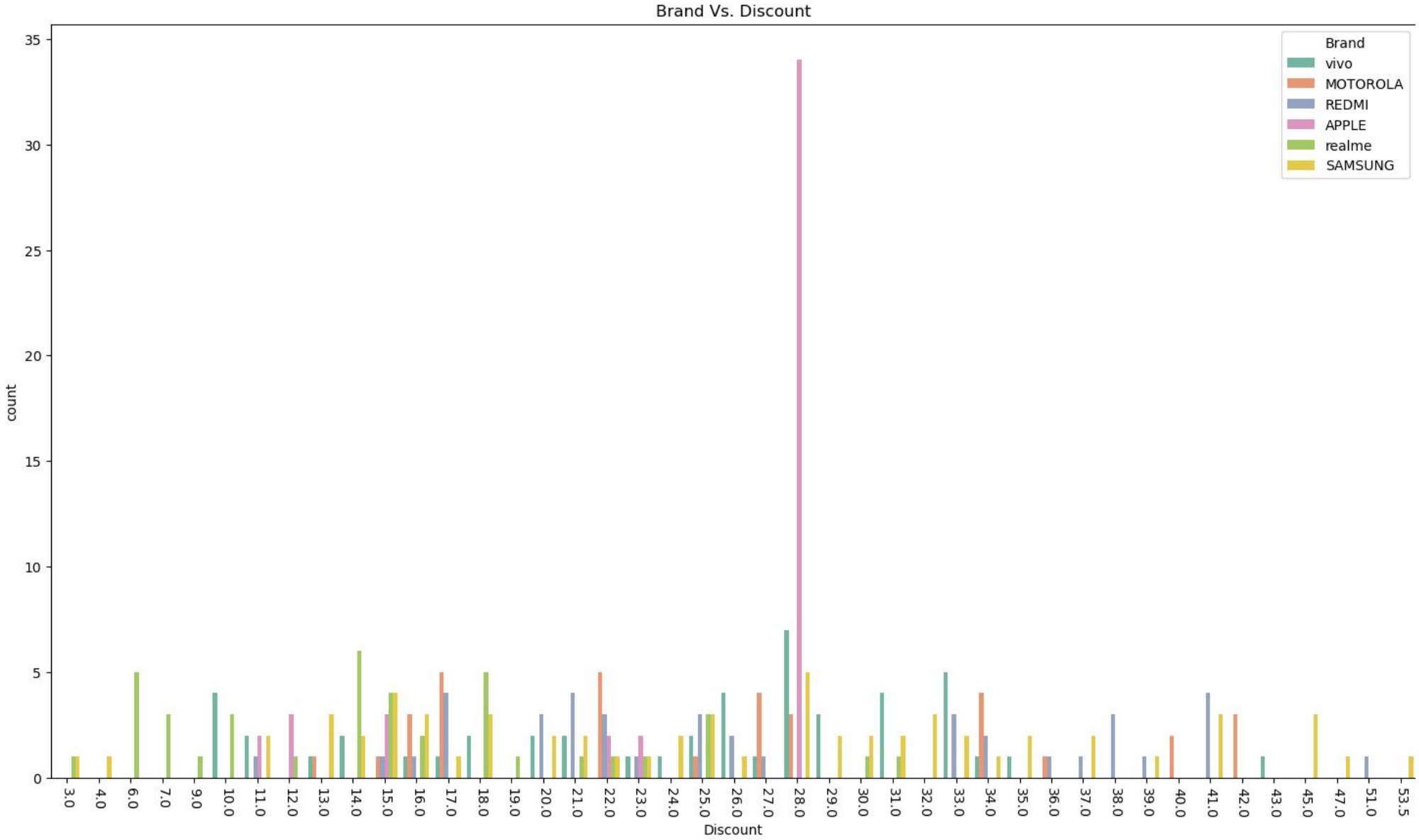
- ❑ It is observed that,
- ✓ It is visualizing the distribution of a single categorical variable named 'Front_cam'.
- ✓ we can conclude that the mobile *camera resolutions are most prevalent at 50MP.*



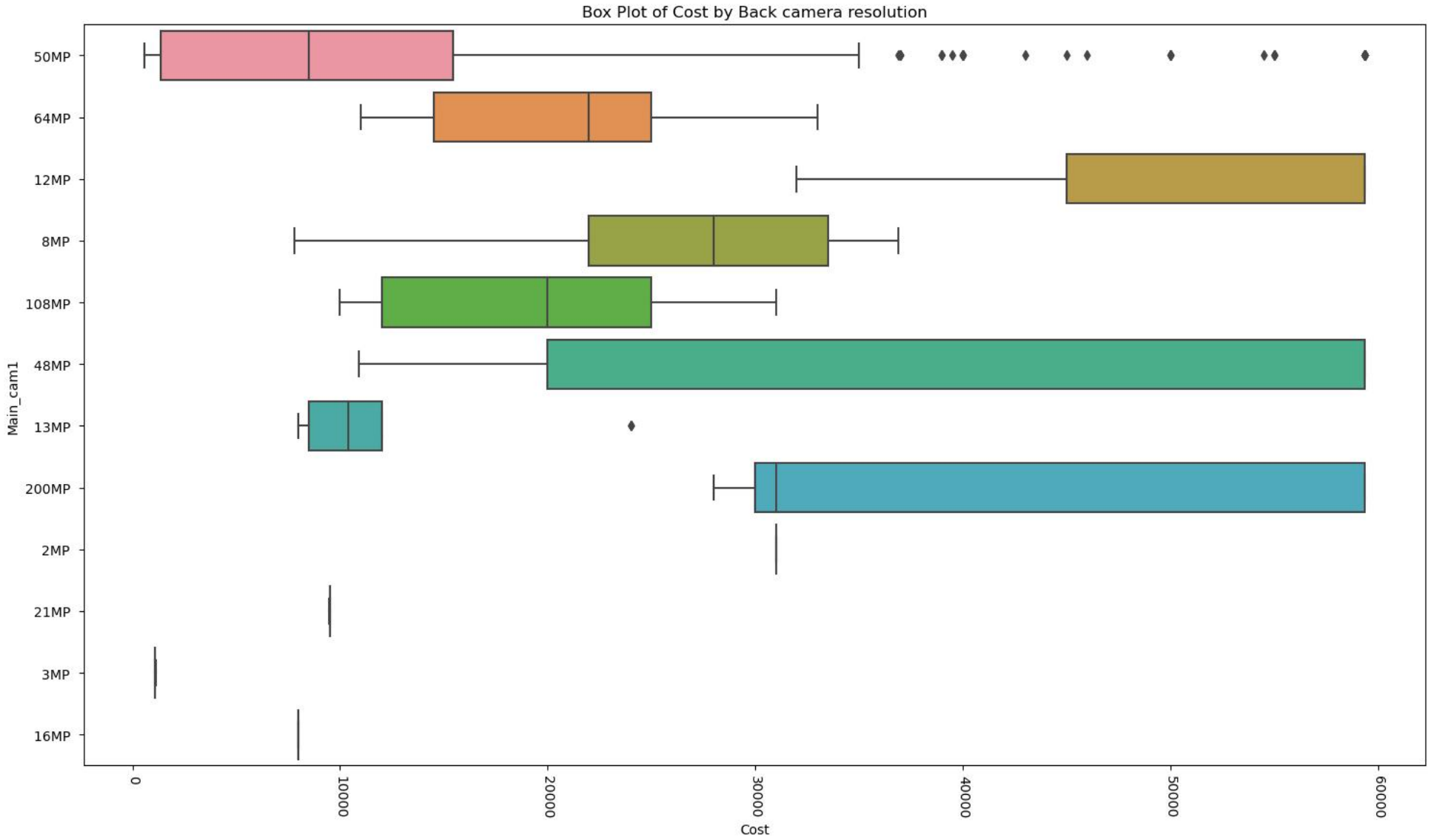
- ❑ It is observed that,
- ✓ It is visualizing the distribution between ‘Brand’ and ‘Rating’ .
- ✓ Different colors represent different brands.
- ✓ we can conclude that the brand APPLE has highest rating ‘4.6’ of all the brands.
- ✓ SAMSUNG has second highest rating at ‘4.2’.



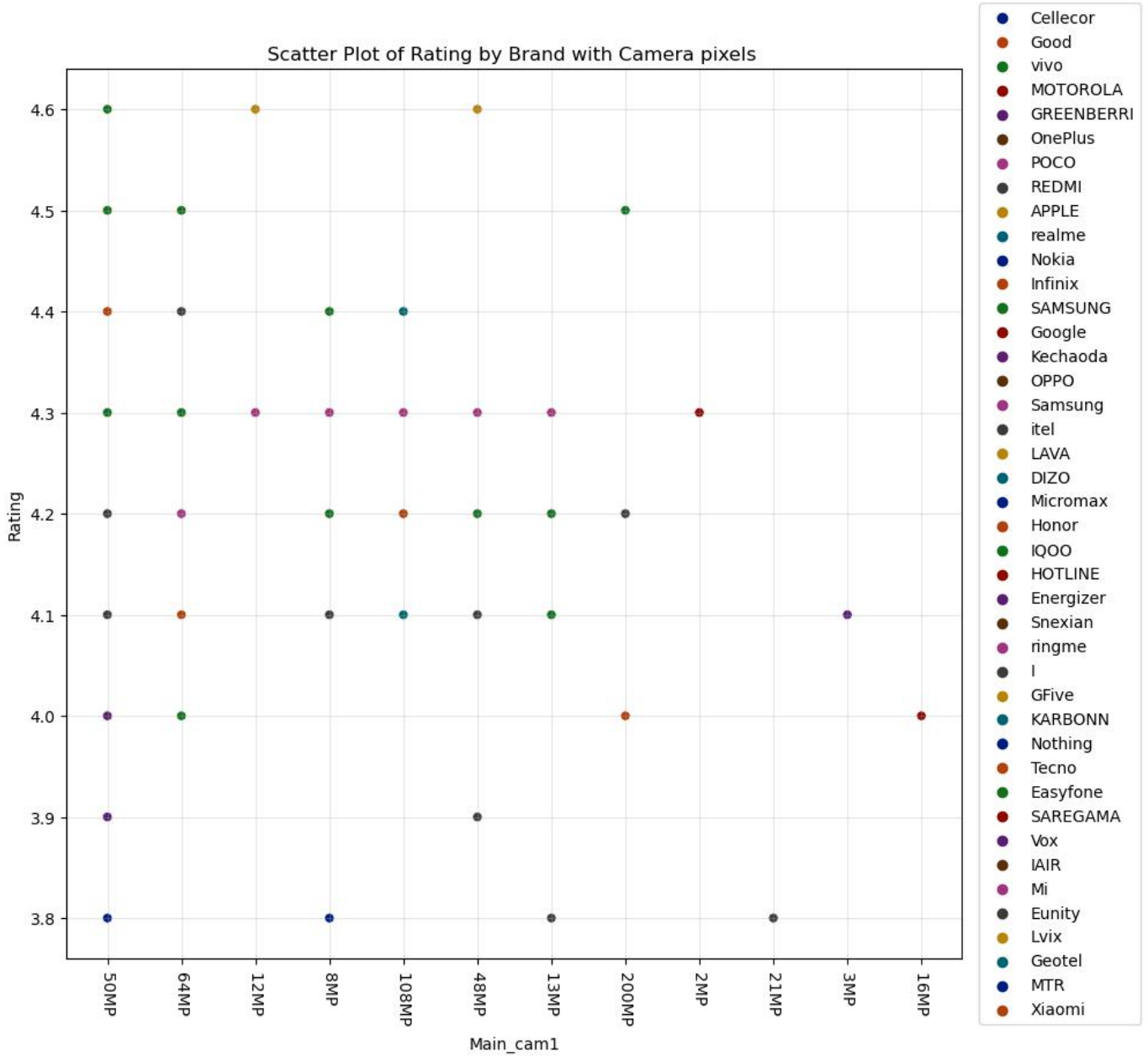
- ❑ It is observed that,
- ✓ It is visualizing the distribution between categories named ‘Brand’ and ‘Cost’.
- ✓ It depicts the information about the brand which is costing high.
- ✓ we can conclude that the brand ‘APPLE’ has highest price as compared to others.



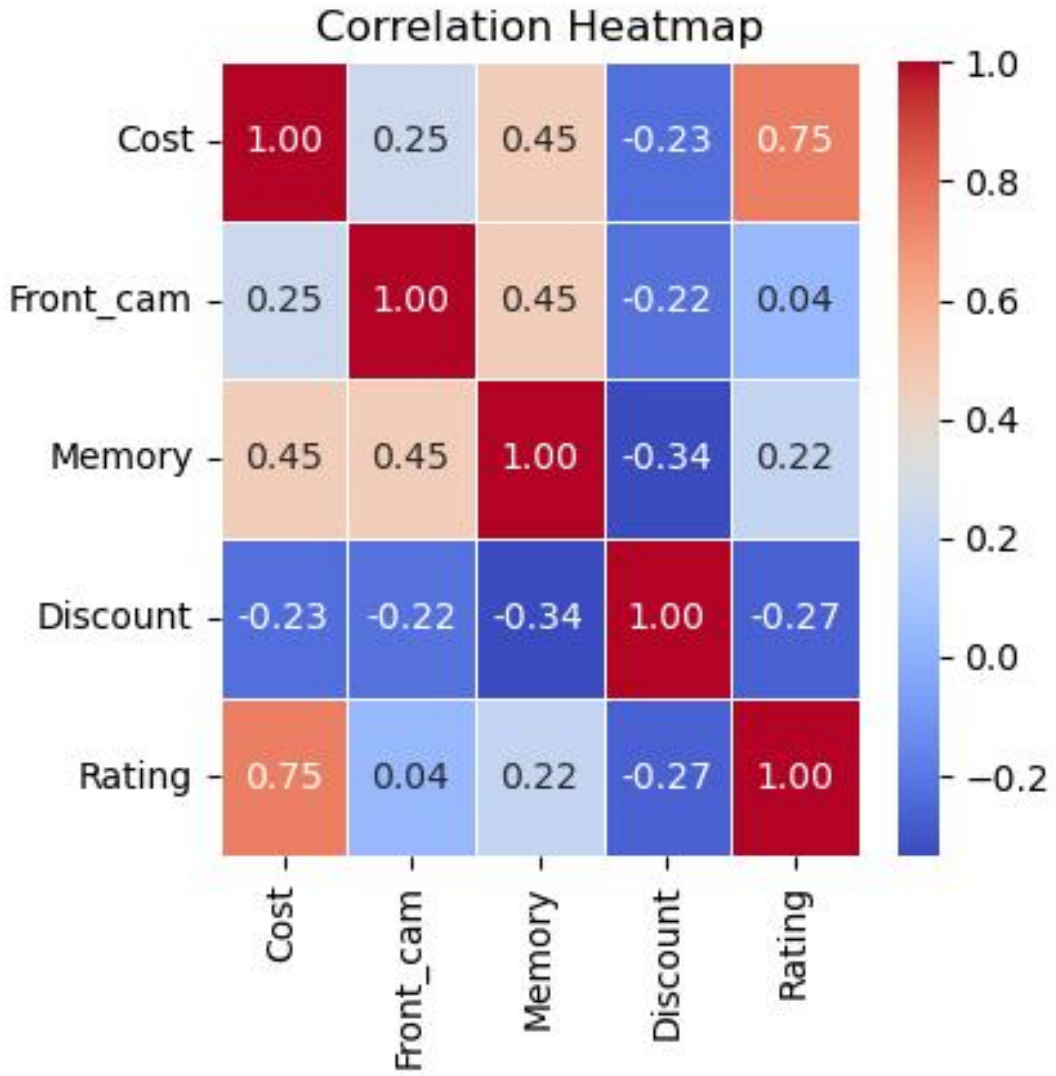
- ❑ It is observed that,
- ✓ It is visualizing the distribution between ‘Brand’ and ‘Discount’ .
- ✓ we can conclude that the brand SAMSUNG and REDMI offering highest discount.



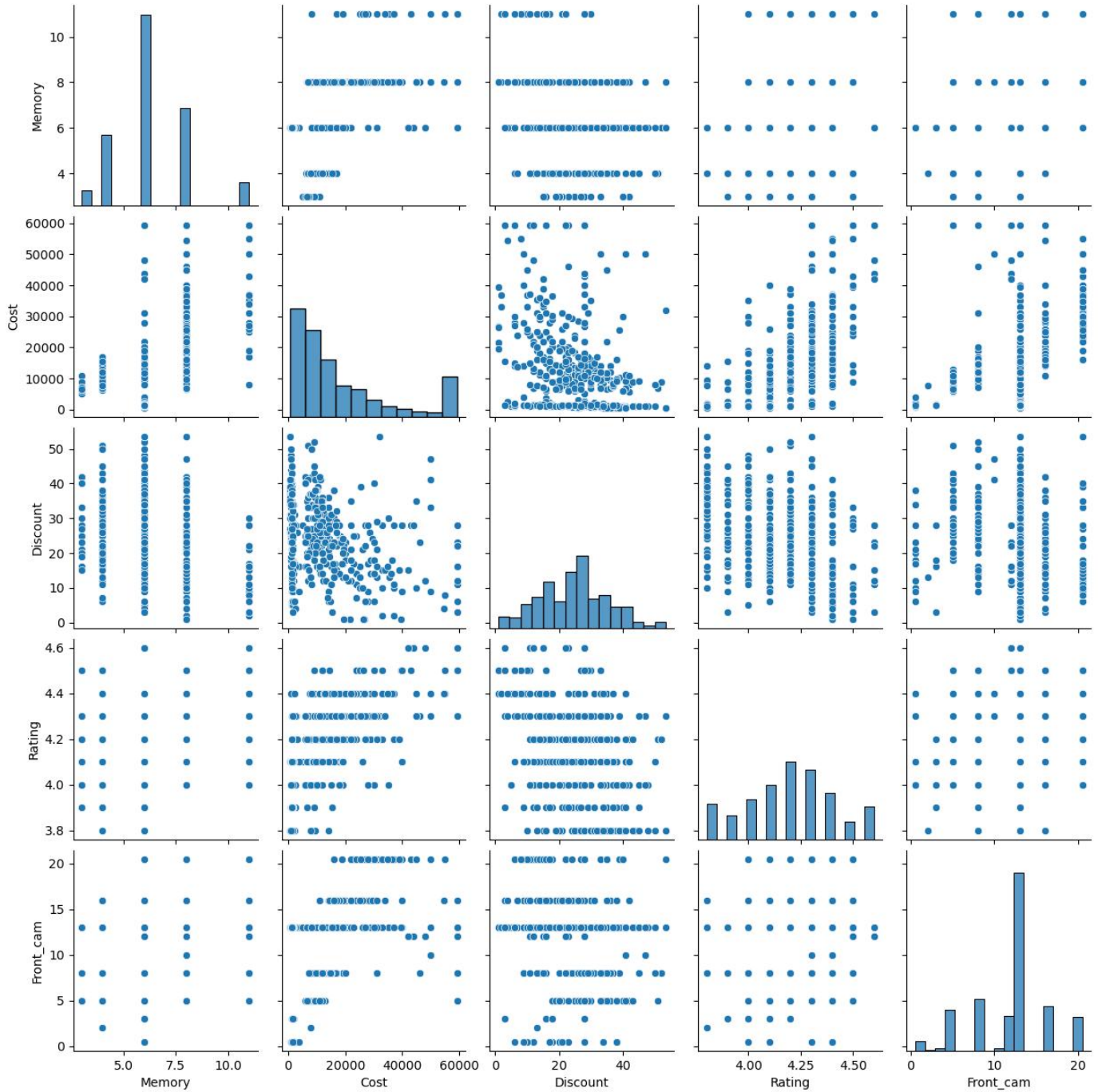
- ❑ It is observed that,
- ✓ This visualization provides how the cost of a phone correlates with the quality and capabilities of its back camera.
- ✓ The outliers could indicate exceptional value-for premium-priced devices with unique features.



- ❑ It is observed that,
- ✓ It is visualizing the relation between categories ‘Rating’ and ‘Back_cam’ and ‘Brand’.
- ✓ *Points on the scatterplot appear scattered randomly without forming a clear trend so we can conclude that there is no correlation.*



- It is observed that,
- ✓ The visualization of the correlation matrix reveals the relationship between each variable.
- ✓ The intensity of the colors indicates the strength of the correlation.
- ✓ We can conclude that correlation between ‘Rating’ and ‘Cost’ suggests a strong positive relationship. .



- It is observed that,
  - ✓ It is a pairplot visualizing scatterplots for all pairs of variables along with histograms on the diagonal.
  - ✓ The histograms on the diagonal show the univariate distribution of each variable.
  - ✓ Scatterplots in the pair plot shows bivariate relationships between pairs of variables.

CONCLUSION

- As of my analysis,
  - Popular brands include Apple, Samsung, Realme, Vivo, Redmi.
  - Prices vary widely, ranging from budget-friendly options around less than or equal to ₹10000 to almost ₹80000.
  - The most highest price preferred brands are ‘Apple’, ‘Samsung’, ‘Google’.
  - The brands ‘Vivo’ and ‘Redmi’ offering maximum discount but still consumers preferring the brand ‘Apple’ due to its brand reputation.
  - According to users ratings, the brands ‘Apple’ and ‘Samsung’ have highest rating.
  - Finally we can recommend that,
    - Premium brands: Apple, Samsung.
    - Budget-friendly brands: Vivo, Realme, Redmi.

THANK YOU