CONTINUOUS ASSESSMENT –3

Course Name :- EDA PROJECT

Course Code :- INT – 353

Topic :- Mobile Trends Analysis

Student Details

Name : SHAIK JULFEEN AHMADH

Section : K21UG

Roll Number : RK21UGB53

Registration No : 12110554

Submitted To :-

Table of Contents

INTRODUCTION

Introduction: Mobile Trends Analysis

The advent of technology has significantly transformed the landscape of mobile devices, offering an array of choices to consumers with diverse preferences and requirements. This report delves into the dynamic world of mobile phones, aiming to analyse and comprehend the trends prevalent in the market. The dataset, sourced from 'mobile_trends.csv', encapsulates a comprehensive array of information pertaining to various mobile phone models, providing insights into crucial aspects such as brand, model name, price, rating, SIM types, and several other key features.

As we navigate through the dataset, we will uncover patterns, correlations, and noteworthy trends that define the contemporary mobile market. From the operating systems that power these devices to the intricacies of their processors, the dataset offers a holistic view of the technological landscape. Furthermore, factors like battery capacity, display size, and camera specifications contribute to the nuanced decision-making process of consumers.

Our analysis will not only shed light on the individual attributes of mobile phones but also explore potential interdependencies among these features. Whether it's the impact of the operating system on user satisfaction or the correlation between price and device specifications, our goal is to extract meaningful insights that can inform both consumers and industry stakeholders.

Through a combination of descriptive statistics, visualizations, and exploratory data analysis, this report endeavours to unravel the underlying narratives within the dataset. By the end, we hope to provide a valuable resource for anyone seeking a deeper understanding of the current trends and dynamics within the mobile phone market.

Join us on this analytical journey as we unravel the intricacies of mobile trends and explore the multifaceted world of contemporary smartphones.

1.DOMAIN:

Domain: Mobile Technology Market Analysis

This dataset resides within the domain of the Mobile Technology Market, offering a detailed exploration of the contemporary landscape of mobile phones. Capturing essential attributes such as brand, model name, pricing, user ratings, SIM types, and diverse technical specifications, the dataset serves as a valuable resource for understanding the intricacies of the mobile technology sector.

Within this domain, we aim to uncover patterns, trends, and relationships that define consumer preferences, technological advancements, and competitive dynamics. The dataset provides a comprehensive overview of mobile phone features, enabling a thorough analysis of factors influencing consumer choices and market trends. The insights derived from this exploration will contribute to a deeper understanding of the evolving dynamics within the Mobile Technology Market.

2. DATA UNDERSTANDING:

Data Understanding: Unveiling Insights into Mobile Technology Trends

In our pursuit of unravelling the nuances within the Mobile Technology Market, a comprehensive exploration of the dataset has provided a foundational understanding of the data at hand. This section delves into the key characteristics and dimensions of the dataset, shedding light on its structure, variables, and potential avenues for analysis.

Dataset Overview:

The dataset encapsulates information from the Mobile Technology Market, spanning various brands and models.

It comprises 26 columns, each representing a distinct attribute, offering a holistic view of mobile phone specifications and features.

Key Variables:

Brand and Model Name: Identifying the manufacturers and specific models within the dataset.

Price: Reflecting the market valuation of each mobile phone model.

Rating: Providing a measure of user satisfaction and product performance.

SIM Types: Detailing the types of SIM cards supported by each device.

Technical Specifications: Including features such as display size, operating system, processor core, and battery capacity.

Data Quality and Completeness:

Missing values, such as those in the 'GPS Support' column, have been observed, necessitating careful consideration during analysis.

The dataset exhibits a diverse range of information, contributing to a comprehensive understanding of mobile phone attributes.

Exploratory Data Analysis (EDA) Opportunities:

Identification of trends in user ratings concerning specific brands or models.

Exploration of the relationship between pricing and technical specifications.

Evaluation of the prevalence and impact of missing values on analytical outcomes.

Potential Challenges:

Addressing missing data through imputation or exclusion based on the context of analysis.

Ensuring uniformity in units and formats for variables like 'Battery Capacity' to facilitate accurate comparisons.

As we embark on the analytical journey, armed with a foundational understanding of the dataset, our objective is to extract meaningful insights that contribute to a deeper comprehension of the Mobile Technology Market. Through descriptive statistics, visualizations, and rigorous exploratory data analysis, we aim to unravel trends, patterns, and correlations that characterize the mobile phone landscape.

Summary:

This dataset is like a treasure chest of information about different mobile phones. It tells us about various brands, their models, prices, how much users like them, and what cool features they have.

Here's a quick look at what we found:

Lots of Brands: There are many different brands of phones in the dataset. This helps us see who's competing in the market.

Details on Each Model: We get to dive into specific details about each phone model. This is like looking under the hood to see what makes each phone special.

Price Insights: Knowing the prices of phones helps us understand what people are willing to pay. We can look for trends in pricing and what might be popular among buyers.

User Ratings: Users give ratings to phones, showing how much, they like them. This gives us a sense of customer satisfaction.

Tech Specs: We have info on the technical stuff too—like the size of the screen, the operating system, the brain of the phone (the processor), and how long the battery lasts.

While the dataset is a goldmine, we need to be careful, especially where some information is missing. As we dig deeper into the data, we'll uncover interesting patterns and trends, helping us understand more about what people love in their mobile phones. The journey ahead is exciting as we aim to uncover the secrets of the ever-changing world of mobile technology.

3. Reasons for Choosing Data Set :

Potential Reasons for Choosing This Dataset for an EDA (Exploratory Data Analysis) project are…

Personal Interest:

- The topic of mobile technology aligns with a personal interest in staying updated on the latest advancements and trends in the tech world. Exploring mobile phones, their features, and market dynamics is inherently engaging and satisfying.

Relevance:

- The dataset is highly relevant to the current technological landscape. Mobile phones are integral to daily life, and understanding the nuances of different models provides insights into consumer behaviour and industry competitiveness.

Market Insights:

- The dataset promises to deliver valuable market insights, allowing for a deeper understanding of consumer preferences, brand dynamics, and emerging trends within the Mobile Technology Market. This information is essential for anyone seeking to stay informed about market shifts.

Practical Application:

- The practical application of insights derived from this dataset extends to real-world scenarios. Whether making informed purchasing decisions as a consumer or strategizing within the mobile technology industry, the practical implications make this dataset particularly valuable.

Data Availability:

- The dataset offers a rich array of information, providing a comprehensive snapshot of the mobile phone market. The availability of diverse variables ensures that there is ample data to explore, making it suitable for in-depth analysis.

Educational Value:

- This project is a part of coursework, and the dataset's relevance to the educational objectives makes it an ideal choice. The exploration of the Mobile Technology Market aligns with the learning goals of the course, allowing for practical application of concepts and methodologies.

Availability of Variety:

- The inclusion of various brands, models, and technical specifications ensures a diverse dataset. This variety not only makes the analysis interesting but also provides a holistic view of the mobile technology landscape, enhancing the educational and analytical value of the project.

In conclusion, the decision to select this dataset is driven by a combination of personal interest, relevance to current trends, potential market insights, practical application in real-world scenarios, the richness of available data, educational alignment, and the diversity offered by the inclusion of various brands and models within the Mobile Technology Market.

4. Questions for Analysis:

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

2. What are the basic summary statistics?

3. Are there any categorical variables and missing values? If so, print it.

4. Are there any outliers in the data? If so, use box plots, histograms, and visualize.

5. Is the data balanced or imbalanced? Visualize.

6. What is the target variable (if any).

7. What are the units of measurement for numerical columns? (example: time, currency, date, distance)

8. Do you have domain clarification? Brief it.

9. Are there any time-based trends or patterns?

10. Are there any correlations between variables? Calculate correlations.

11. Do different 'Brands' have significantly different 'Prices'?

12.What is the distribution of internal storage capacities among mobile devices

13.Which brand has the highest average price?

14.What is the distribution of ratings across different operating systems?

15.What is the distribution of 'Weight' for different SIM types?

16.How is the distribution of Wi-Fi and GPS support in the dataset?

17.Is there a noticeable trend in ratings based on the processor core of device?

18.Compare the average price of phones with different screen sizes

19.How is the distribution of Network Types across different brands, and does the Network Type correlate with the average Price of the mobile devices?

20.What is the distribution of Bluetooth versions among the mobile devices in the dataset, and how does it vary across different brands?

Libraries used:

The following libraries were used for the project and their brief description as follows:

NumPy: NumPy, short for "Numerical Python," is a fundamental package for scientific computing in Python. It provides support for large, multi-dimensional arraysand matrices, along with a variety of high-level mathematical functions to operate onthese arrays. NumPy is often a foundational library for data manipulation and analysis in Python.

Pandas: Pandas is a powerful data manipulation and analysis library for Python. Itprovides data structures like DataFrames and Series, which make it easy to work with structured data, such as CSV files, Excel spreadsheets, or SQL databases.

Pandas is widely used for data cleaning, transformation, aggregation, and exploration

Matplotlib: Matplotlib is a popular Python library for creating static, animated, and interactive visualizations in Python. It offers a wide range of plotting options, allowingyou to create various types of charts and graphs, from simple line plots to complex heat maps and 3D plots. Matplotlib is highly customizable and can be used alongsideother libraries like NumPy and Pandas for data visualisation.

Seaborn: Seaborn is a data visualisation library built on top of Matplotlib. It providesa high-level interface for creating informative and attractive statistical graphics.

Seaborn simplifies the process of creating complex visualizations, such as scatterplots, bar plots, and heatmaps, and it also offers built-in support for working with Pandas DataFrames.

Data Description:

The dataset is a comprehensive collection of information about mobile phones, encompassing crucial attributes such as brand, model name, price, rating, SIM type, and a variety of hardware features. Each mobile phone entry provides details on display size, operating system, processor core, internal storage, camera specifications, network type, and connectivity features like Bluetooth and Wi-Fi. This well-structured dataset proves valuable for a thorough analysis of the diverse characteristics exhibited by different mobile phone models.

Notably, the dataset stands out for its cleanliness, as there are no missing values in any of its columns. This absence of missing data enhances the reliability of the dataset, ensuring that each entry is complete and accurate. With a wealth of detailed information and a lack of missing values, this dataset becomes a robust resource for gaining insights into the features and specifications of various mobile phones, facilitating informed decision-making and analysis in the realm of mobile technology.

Steps of EDA:

1.Data Sourcing:

Source of Dataset: Kaggle

The dataset encompasses various variables representing mobile phone characteristics. The key variables include:

1. **Brand (Type: Object):** The brand of the mobile phone.
2. **Model Name (Type: Object):** The specific model name of the mobile phone.
3. **Price (Type: Int64):** The price of the mobile phone.
4. **Rating (Type: Float64):** The rating assigned to the mobile phone.
5. **SIM Type (Type: Object):** The type of SIM card supported.
6. **Hybrid Sim Slot (Type: Object):** Indicates if the mobile phone has a hybrid SIM slot.
7. **Touchscreen (Type: Object):** Indicates if the mobile phone has a touchscreen.
8. **Display_size (Type: Float64):** Size of the mobile phone display.
9. **Operating System (Type: Object):** The operating system installed on the mobile phone.
10. **Processor Core (Type: Object):** The number of processor cores.
11. **Internal Storage (Type: Float64):** Amount of internal storage available.
12. **Primary Camera (Type: Object):** Specifications of the primary (main) camera.

13. **Secondary Camera (Type: Object):** Specifications of the secondary (front) camera.
14. **Network Type (Type: Object):** Type of network supported.
15. **Bluetooth Version (Type: Object):** The version of Bluetooth supported.
16. **Wi-Fi (Type: Object):** Indicates if Wi-Fi is supported.
17. **GPS Support (Type: Object):** Indicates if the mobile phone has GPS support.
18. **SIM Size (Type: Object):** Size of the SIM card.
19. **Battery Capacity (Type: Object):** The capacity of the mobile phone's battery.
20. **Width (Type: Float64):** Width of the mobile phone.

Data preprocessing involved handling missing values and duplicates. The 'Hybrid Sim Slot' and 'Operating System' columns had missing values, and for categorical variables, these were treated appropriately based on their nature. Duplicate rows were identified and removed. The dataset was left in a clean state for further analysis, with no missing values in crucial columns.

2.Data Cleaning:

Checked the dataset details:

Started by exploring the dataset to understand the information in each column, identify missing data, and assess memory usage.

Hunted for missing data:

Delved deeper to spot any missing information. This included pinpointing columns with missing values and quantifying how much information was absent. By using a Pandas library utility named isna(). Assessed the significance of missing information by calculating the percentage of missing data in each category using pandas functions.

Cleaned up repeated information:

Identified 14 duplicated rows using the duplicated function and chose to delete them using the drop_duplicates() function. This step ensured that identical information didn't appear more than once in the dataset.

Tackled missing values:

Addressed missing values by focusing on specific columns. You chose to drop rows where essential information was missing in the 'Width,' 'Height,' and 'Weight' columns using the dropna function with the subset parameter. This operation was performed in place with inplace=True.

Additional removal of missing values:

Following the initial step, you further handled missing values by using the dropna() function without specifying a subset. This operation effectively removed any rows containing missing values in any column.

Checked for remaining missing values:

After the removal steps, you checked for any remaining missing values in the 'mobile' DataFrame using mobile.isnull().sum(). This allowed you to confirm that the specified columns were now free of null values and to verify if any other columns still had missing data.

Detection and Tackling of Outliers

In the process of cleaning the data for outliers, I employed box plots to visually detect potential anomalies and calculated interquartile ranges (IQRs) to quantitatively assess the spread of the data, allowing for the identification and subsequent handling of values that deviated significantly from the norm.

3.Data Exploration:

Here's an interpretation of the summary statistics and initial exploratory data analysis:

Summary Statistics (mobile.describe()):

The describe() function provides summary statistics of the numerical columns in the dataset. This includes the count, mean, standard deviation, minimum, 25th percentile, median (50th percentile), 75th percentile, and maximum values for each numerical feature. It helps give an overview of the central tendency and spread of the numerical variables.

Dataset Shape (mobile.shape):

The shape attribute reveals the dimensions of the dataset, indicating the number of rows and columns. This is useful for getting a quick sense of the dataset's size.

First Few Rows (mobile.head()):

The head() method displays the first few rows of the dataset. It gives a glimpse of the actual data and how it is structured.

Last Few Rows (mobile.tail()):

The tail() method shows the last few rows of the dataset. This is useful for ensuring that the dataset has been loaded correctly and for observing the overall structure.

To further explore the data visually, you might consider creating plots or charts based on specific columns of interest, depending on your analysis goals. For example, histograms for numeric variables, bar plots for categorical variables, or scatter plots to explore relationships between variables.

4.Univariate Analysis:

1)In the univariate analysis focusing on numerical variables, histograms were utilized as a key visualization tool. These histograms offered a detailed representation of the distribution of values within each numerical variable, providing valuable insights into the underlying patterns and characteristics of the dataset.

2) In the exploration of numerical columns during univariate analysis, I incorporated violin plots as an effective visualization technique. These plots seamlessly combine aspects of box plots and kernel density plots, offering a holistic view of the data's distribution, concentration, and potential outliers. The violin plots provided a nuanced understanding of the numerical variables, revealing not only central tendencies but also the density of data points across different values.

3) I integrated a heatmap to visualize the correlation matrix of the dataset. This heatmap, generated from the numerical columns, served as a comprehensive tool to uncover the pairwise relationships between variables. The annotated heatmap provided insights into the strength and direction of correlations, aiding in the identification of potential associations and dependencies among the numerical features.

4) In the examination of categorical variables through univariate analysis, count plots were employed to visually represent the distribution of categories within the columns: 'Brand', 'SIM Type', 'Hybrid Sim Slot', 'Touchscreen', 'Operating System', 'Processor Core', 'Secondary Camera', 'Network Type', 'Bluetooth Version', 'Wi-Fi', 'GPS Support', 'SIM Size', 'Battery Capacity', 'SIM Type.1', 'Hybrid Sim Slot.1', and 'Dual Camera Lens'. These count plots effectively illustrate the frequency of each category, providing a clear snapshot of the prevalence of different features.

5) Additionally, to enhance the exploration of specific categorical variables, pie charts were utilized for 'Hybrid Sim Slot', 'Touchscreen', 'Processor Core', 'Bluetooth Version', 'Wi-Fi', 'GPS Support', 'SIM Type.1', 'Hybrid Sim Slot.1', and 'Dual Camera Lens'. Pie charts offer a visual representation of the proportional distribution of categories within these specific features, providing a quick and intuitive overview of the dataset's categorical composition.

5.Bivariate Analysis:

1) In the process of bivariate analysis, I conducted an exploration using scatter plots to discern relationships between pairs of variables. The scatter plots were meticulously crafted, each portraying a specific pairing: 'Price' against 'Rating,' 'Battery Capacity' against 'Weight,' 'Width' against 'Height,' and 'Internal Storage' against 'Display_size.'

2) The scatter plot of 'Price' versus 'Rating' sheds light on the potential correlation between the price of mobile phones and their user ratings. Similarly, other scatter plots provide insights into relationships such as the connection between 'Battery Capacity' and 'Weight,' 'Width' and 'Height,' and 'Internal Storage' and 'Display_size.'

3) The alpha parameter was set to 0.5 for transparency, allowing for clear view of overlapping points. This array of scatter plots serves as a valuable tool for uncovering patterns & correlations b/w paired variables within the dataset.

4) In the domain of exploring relationships within numerical variables, I constructed a correlation matrix for the mobile dataset. This matrix, derived from numerical columns, quantifies the pairwise correlations between different variables.

5) Utilizing a heatmap visualization, the matrix was presented with annotations to highlight correlation coefficients. The color gradient, facilitated by the 'viridis' colormap, accentuates the strength and direction of correlations, offering a comprehensive visual representation. This heatmap, unveiling the interplay between numerical features, is instrumental in identifying potential patterns and dependencies within the dataset.

6) In the exploration of individual variable distributions, I employed displot, a part of the Seaborn library, to construct a set of four distribution plots for specific features in the mobile dataset. The displot provides a succinct visual representation of the frequency distribution and density of values for each variable.

7) The distribution of 'Price' reveals insights into the pricing structure, while the 'Rating' plot showcases the spread of user ratings. Similarly, 'Display_size' and 'Internal Storage' distribution plots provide valuable perspectives on the size and storage capacity distribution across the mobile phone dataset.

6.Multivariate Analysis:

1) Engaging in an exploration of complex relationships among numerical variables, I conducted a multivariate analysis utilizing pair plots. By employing Seaborn's pair plot function, a grid of scatter plots was crafted, focusing on a subset of numerical columns from the mobile dataset.
2) Within this grid, each scatter plot vividly illustrates the interactions between two numerical variables, offering a visual canvas to discern potential correlations and intricate patterns. Notably, the diagonal elements of the grid are adorned with kernel density estimation (KDE) plots, providing a smooth representation of the univariate distribution for each variable.
3) In my analytical exploration, I constructed a heatmap to delve into the correlation matrix of the mobile dataset. This matrix, derived from the entire dataset, quantifies the pairwise correlations between different variables.
4) Utilizing a heatmap visualization, I presented the matrix with annotations to highlight correlation coefficients. The colour palette, defined by the 'coolwarm' colormap, accentuates the strength and direction of correlations, offering a comprehensive and visually engaging representation.

7. Distribution :

Histogram:

A histogram with kernel density estimation (KDE) is created using Seaborn's histplot function.

This visualizes the distribution of 'Rating' to provide an initial understanding of its shape.

Shapiro-Wilk Test:

The Shapiro-Wilk normality test is performed using the shapiro function from the scipy.stats module.

The test returns a statistic and p-value, helping to determine if the data follows a normal distribution.

Significance Level:

A significance level (alpha) of 0.05 is assumed.

Print Result:

The code prints whether the distribution of 'Rating' is considered normal based on the Shapiro-Wilk test results.

This code snippet combines visual inspection and statistical testing to assess the normality of the 'Rating' column in the mobile dataset.

The code assesses the normality of the 'Weight' column in the mobile dataset using the Shapiro-Wilk test.

Graph – 2

A histogram with kernel density estimation (KDE) is created using Seaborn's histplot function.

This visualizes the distribution of 'Weight' to provide an initial understanding of its shape.

Shapiro-Wilk Test:

The Shapiro-Wilk normality test is performed using the shapiro function from the scipy.stats module.

The test returns a statistic and p-value, helping to determine if the data follows a normal distribution.

Significance Level:

A significance level (alpha) of 0.05 is assumed.

Print Result:

The code prints whether the distribution of 'Weight' is considered normal based on the Shapiro-Wilk test results.

This code snippet combines visual inspection and statistical testing to assess the normality of the 'Weight' column in the mobile dataset.

8.Hypothesis Testing:

Hypothesis:

Null Hypothesis (H0): The mean prices of smartphones from the 'iPhone' and 'Samsung' brands are equal.

Alternative Hypothesis (H1): The mean prices of smartphones from the 'iPhone' and 'Samsung' brands are not equal.

Explanation of Steps:

Displaying Results:

The calculated t-statistic and p-value are displayed. The t-statistic represents the strength and direction of the observed difference, while the p-value indicates the probability of observing such a difference by chance.

Interpretation and Decision:

The significance level, often set at 0.05 (or 5%), is assumed. If the p-value is less than this significance level, the null hypothesis is rejected. In this context:

If the p-value is less than 0.05, it is concluded that there is evidence to reject the null hypothesis. This suggests that the mean prices of smartphones from the 'iPhone' and 'Samsung' brands are not equal.

If the p-value is greater than or equal to 0.05, there is no statistically significant evidence to reject the null hypothesis. This implies that there is no indication that the mean prices of smartphones from the 'iPhone' and 'Samsung' brands are different.

This process allows for a rigorous statistical evaluation of whether there is a significant difference in mean prices between the 'iPhone' and 'Samsung' brands.

9.Questions for Analysis:

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

Column	Non-Null Count	Dtype
Brand	778 non-null	object
Model Name	778 non-null	object
Price	778 non-null	int64
Rating	778 non-null	float64
SIM Type	778 non-null	object
Hybrid Sim Slot	778 non-null	object
Touchscreen	778 non-null	object
Display_size	778 non-null	float64
Operating System	778 non-null	object
Processor Core	778 non-null	object
Internal Storage	778 non-null	float64
Primary Camera	778 non-null	object
Secondary Camera	778 non-null	object
Network Type	778 non-null	object
Bluetooth Version	778 non-null	object
Wi-Fi	778 non-null	object
GPS Support	778 non-null	object
SIM Size	778 non-null	object
Battery Capacity	778 non-null	object
Width	778 non-null	float64
Height	778 non-null	float64
Weight	778 non-null	float64
SIM Type.1	778 non-null	object
Hybrid Sim Slot.1	778 non-null	object
Dual Camera Lens	778 non-null	object
Color	778 non-null	object

- ✓ The dataset comprises 26 columns, featuring a mix of numerical (float64, int64) and categorical (object) data types.
- ✓ All columns exhibit complete data with 778 non-null entries, indicating an absence of missing values.
- ✓ Key numerical columns include 'Price,' 'Rating,' 'Display_size,' and others, providing essential insights into mobile device specifications.

2. What are the basic summary statistics?

Output

	Price	Rating	Display_size	Internal Storage	Width	Height	Weight
count	778.000000	778.000000	778.000000	778.000000	778.000000	778.000000	778.000000
mean	21796.984576	4.205398	5.803436	126.309245	74.993078	161.252368	189.697514
std	26024.828639	0.258292	1.814605	92.579352	1.968990	6.310185	12.953074
min	597.000000	2.900000	0.660000	0.000000	70.600000	146.300000	155.000000
25%	7991.250000	4.100000	6.380000	64.000000	73.820000	159.900000	181.000000
50%	13999.000000	4.200000	6.590000	128.000000	75.750000	163.600000	189.500000
75%	24999.000000	4.300000	6.700000	128.000000	76.400000	164.900000	199.800000
max	199900.000000	5.000000	16.510000	1024.000000	78.460000	173.100000	228.000000

- ❖ Provides a statistical summary of a mobile phone dataset.
- ❖ Includes central tendency, dispersion, and shape measures.
- ❖ Useful for identifying outliers, trends, and relationships between metrics.

3.Are there any categorical variables and missing values? If so, print it

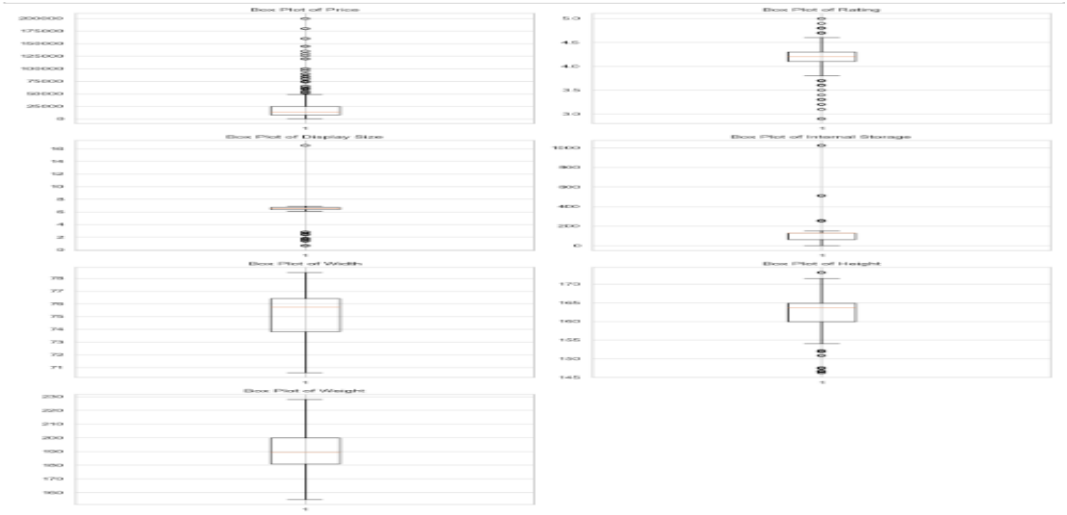
Output

```
Index(['Brand', 'Model Name', 'SIM Type', 'Hybrid Sim Slot', 'Touchscreen',
       'Operating System', 'Processor Core', 'Primary Camera',
       'Secondary Camera', 'Network Type', 'Bluetooth Version', 'Wi-Fi',
       'GPS Support', 'SIM Size', 'Battery Capacity', 'SIM Type.1',
       'Hybrid Sim Slot.1', 'Dual Camera Lens', 'Color'],
      dtype='object')
```

```
Brand                0
Model Name           0
Price                0
Rating               0
SIM Type             0
Hybrid Sim Slot      0
Touchscreen          0
Display_size         0
Operating System     0
Processor Core       0
Internal Storage     0
Primary Camera       0
Secondary Camera     0
Network Type         0
Bluetooth Version    0
Wi-Fi                0
GPS Support          0
SIM Size             0
Battery Capacity     0
Width                0
Height               0
Weight               0
SIM Type.1           0
Hybrid Sim Slot.1    0
Dual Camera Lens     0
Color                0
dtype: int64
```

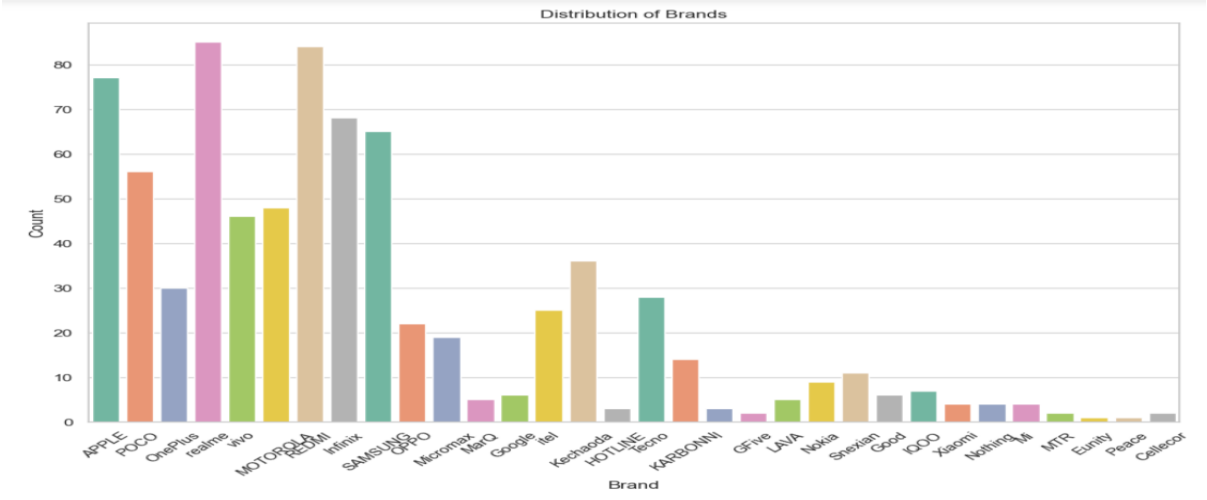
- Mobile data set contains a categorical data.
- As a significant amount column are belong to categorical.

4. Are there any outliers in the data? If so, use box plots, histograms and visualize.



- ✓ The code generates a set of box plots to visually identify outliers in key numerical columns of the mobile dataset.
- ✓ Each box plot represents the distribution of a specific variable, such as 'Price,' 'Rating,' 'Display Size,' 'Internal Storage,' 'Width,' 'Height,' and 'Weight.'
- ✓ Outliers, or extreme values, can be detected by observing data points that fall outside the whiskers of the box plots.
- ✓ The absence of a box plot in the last subplot indicates that the 'Weight' column is not plotted for outlier detection.

5. Is the data balanced or imbalanced? Visualize.



- ❖ The code generates a count plot depicting the distribution of the target variable 'Brand' in the mobile dataset.
- ❖ Utilizing the Seaborn library, the plot visualizes the frequency of each brand with a specified color palette.

- ❖ The x-axis represents brand names, while the y-axis indicates the corresponding count. The plot is presented in a larger size for clarity, with rotated x-axis labels for better readability.

6.What is the target variable (if any).

The target variable is Price because it is necessary for us to analyse any mobile and is a must to evaluate the features for any mobile.

7. What are the units of measurement for numerical columns?

-- The units of numerical columns are as follows:

Price: Rupees

Rating: No units

Display_size: Inches

Internal Storage: GB

Width: Millimetre (mm)

Height: Millimetre (mm)

Weight: Grams(g)

8. Do you have domain clarification? Brief it.

Domain: Mobile Technology Market Analysis

This dataset resides within the domain of the Mobile Technology Market, offering a detailed exploration of the contemporary landscape of mobile phones. Capturing essential attributes such as brand, model name, pricing, user ratings, SIM types, and diverse technical specifications, the dataset serves as a valuable resource for understanding the intricacies of the mobile technology sector.

Within this domain, we aim to uncover patterns, trends, and relationships that define consumer preferences, technological advancements, and competitive dynamics. The dataset provides a comprehensive overview of mobile phone features, enabling a thorough analysis of factors influencing consumer choices and market trends. The insights derived from this exploration will contribute to

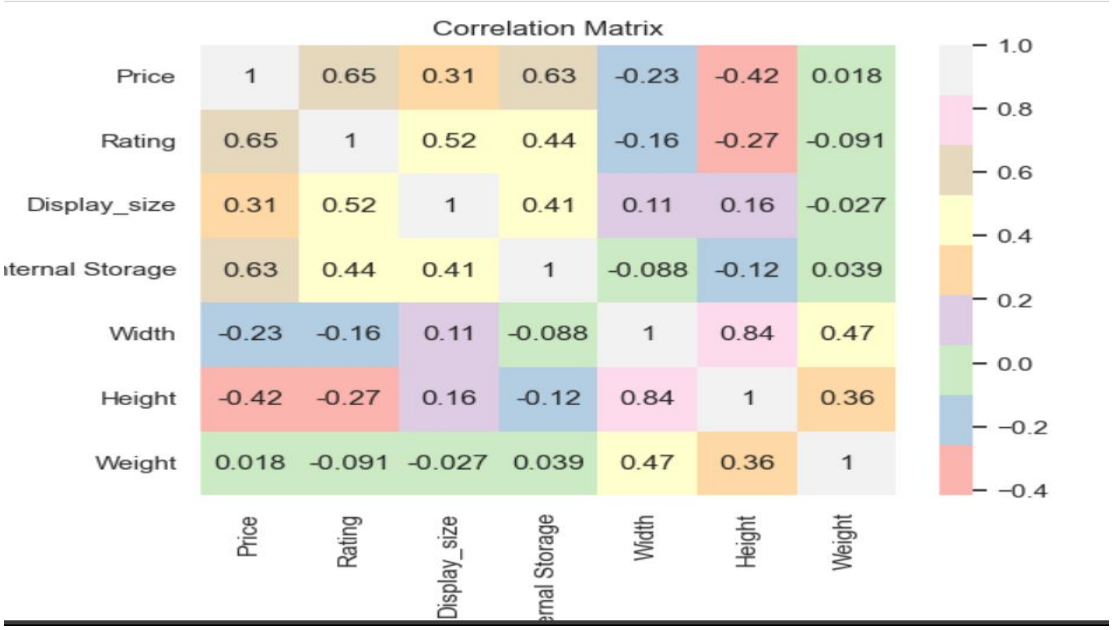
a deeper understanding of the evolving dynamics within the Mobile Technology Market.

9. Are there any time-based trends or patterns?

- Yes, few parameters like Processor Core , Display_size , Battery Capacity , Blloth Version chnages with time . Although no direct time related trends are provides but thesee mentioned parameters keep changing with time.

10. Are there any correlations between variables? Calculate correlations.

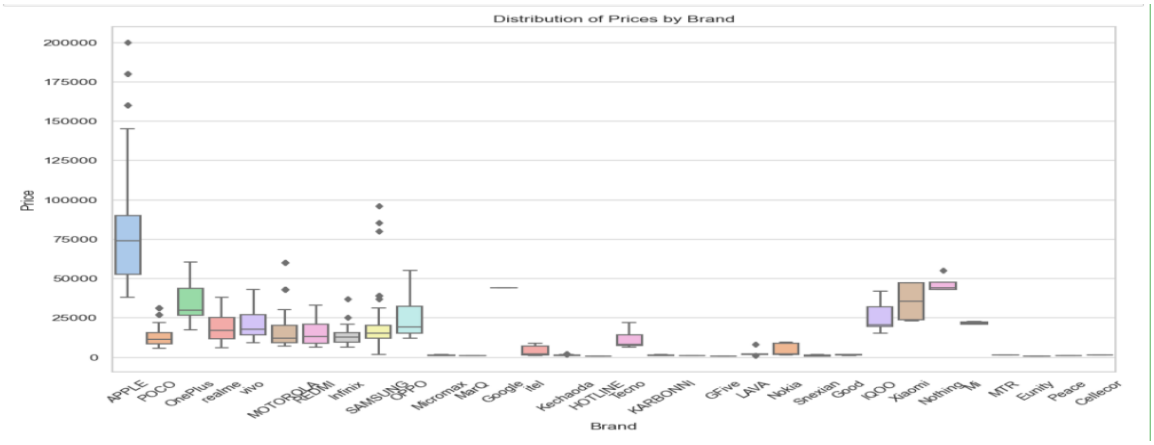
Output



- ✓ The code generates a heatmap to illustrate the correlation matrix of selected numerical columns in the mobile dataset.
- ✓ The matrix, computed using Pearson correlation coefficients, is visualized with Seaborn's heatmap function.
- ✓ The heatmap uses the 'Pastel1' color map and displays correlation values as annotations.
- ✓ This concise visualization offers insights into the linear relationships between numerical features, with a title indicating its purpose as a 'Correlation Matrix.'

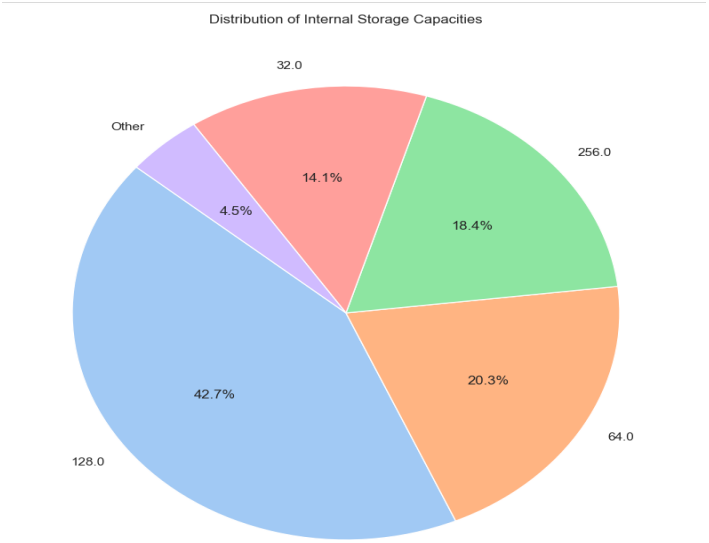
11. Do different 'Brands' have significantly different 'Prices'?

Output:



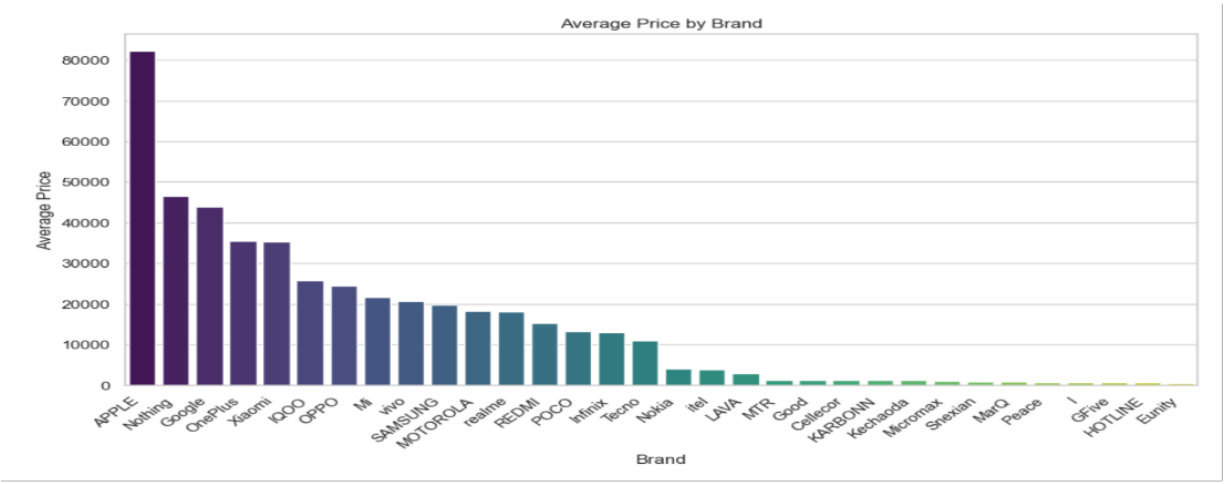
- ✓ The code employs Seaborn to create a box plot illustrating the distribution of mobile phone prices across different brands in the dataset.
- ✓ The x-axis represents the brands, while the y-axis depicts the corresponding prices. Each box visually encapsulates the interquartile range of prices for a specific brand, providing a quick comparison of price distributions.
- ✓ The pastel color palette enhances brand differentiation, and the plot's title, 'Distribution of Prices by Brand,' succinctly summarizes its purpose.

12.What is the distribution of internal storage capacities among mobile devices?



- ✓ The code modifies the 'Internal Storage' column in the mobile dataset by grouping less frequent storage capacities into an 'Other' category.
- ✓ This consolidation is based on a threshold of 0.1, below which categories are considered as 'Other.' After the transformation, a pie chart is generated to visually represent the distribution of internal storage capacities.
- ✓ The chart displays the proportions of each storage category, with labels, percentages, and a pastel color palette for clarity.
- ✓ This process effectively simplifies the representation of internal storage capacities and enhances the visualization of major categories.

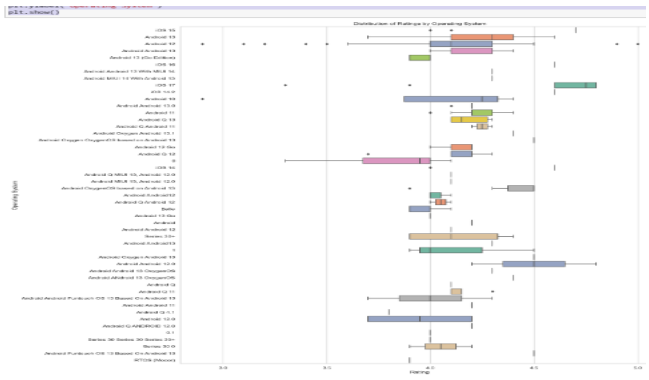
13.Which brand has the highest average price?



- ✓ The code calculates and visualizes the average prices of mobile phones by brand.
- ✓ It utilizes a horizontal bar plot to represent each brand's average price, with brands sorted in descending order.
- ✓ The 'viridis' color palette enhances visual clarity, and the plot provides a quick overview of the comparative average prices among different mobile phone brands.

14.What is the distribution of ratings across different operating systems?

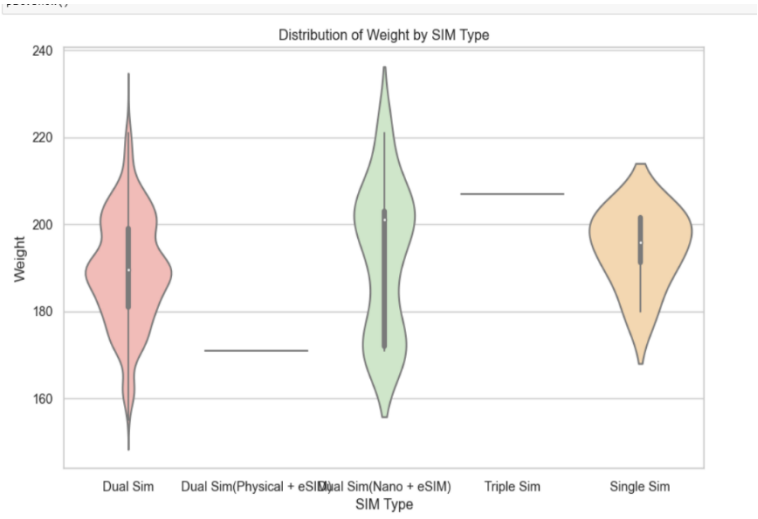
Output



- ✓ Generates a box plot using Seaborn to showcase the distribution of mobile phone ratings categorized by operating systems.
- ✓ The x-axis represents the ratings, while the y-axis displays the operating systems. The 'Set2' color palette is applied for visual differentiation.
- ✓ The large figure size of 15x20 inches ensures a clear presentation, and the plot title, 'Distribution of Ratings by Operating System,' succinctly summarizes its purpose.
- ✓ This visualization allows for a comparative analysis of rating distributions across different operating systems.

15.What is the distribution of 'Weight' for different SIM types?

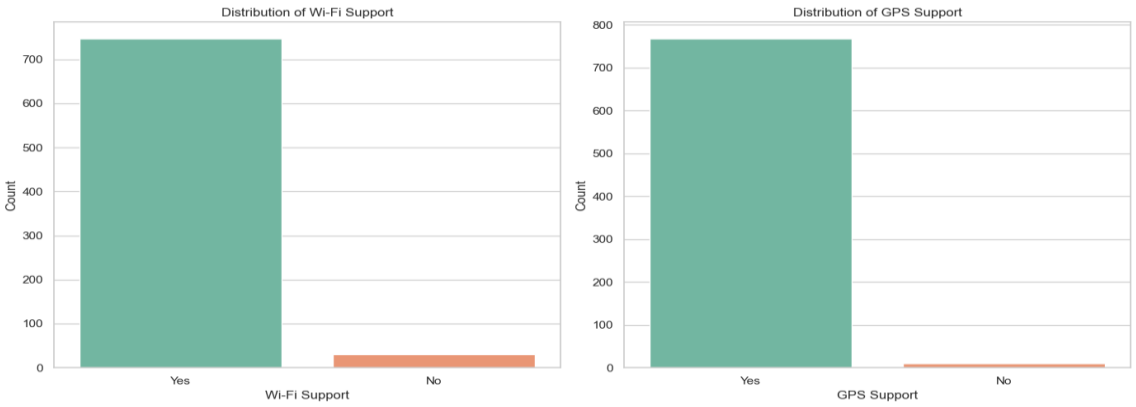
Output



- ✓ A violin plot using Seaborn to illustrate the distribution of mobile phone weights based on SIM types. The x-axis represents the SIM types, and the y-axis displays the corresponding weights.
- ✓ A 'Pastel1' color palette is applied for visual clarity. The figure size is set to 10x6 inches, and the plot is titled 'Distribution of Weight by SIM Type.'
- ✓ This visualization provides insights into the weight distribution across different SIM types, allowing for a comparative analysis.

16.How is the distribution of Wi-Fi and GPS support in the dataset?

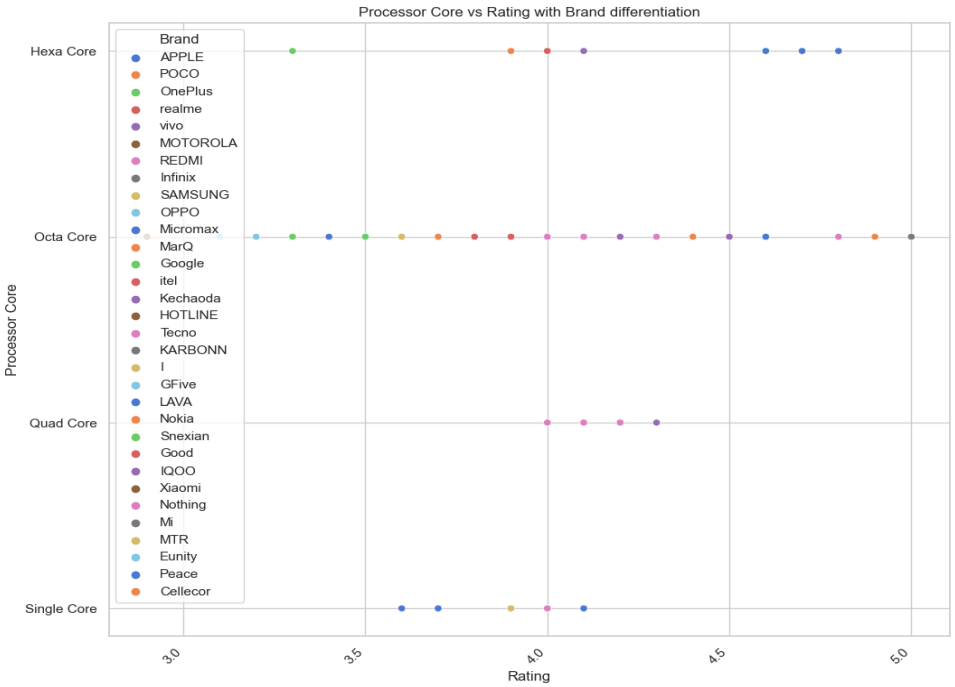
Output:



- ✓ A subplot with two count plots using Seaborn. The first subplot illustrates the distribution of Wi-Fi support in mobile phones, while the second subplot represents the distribution of GPS support.
- ✓ Each count plot displays the frequency of 'Wi-Fi' and 'GPS Support' categories, respectively. The 'Set2' color palette enhances visual distinction.
- ✓ The figure size is set to 14x6 inches, and the subplots are arranged side by side. Each subplot is titled appropriately, providing a clear overview of the distribution of Wi-Fi and GPS support in the mobile dataset.

17.Is there a noticeable trend in ratings based on the processor core of device?

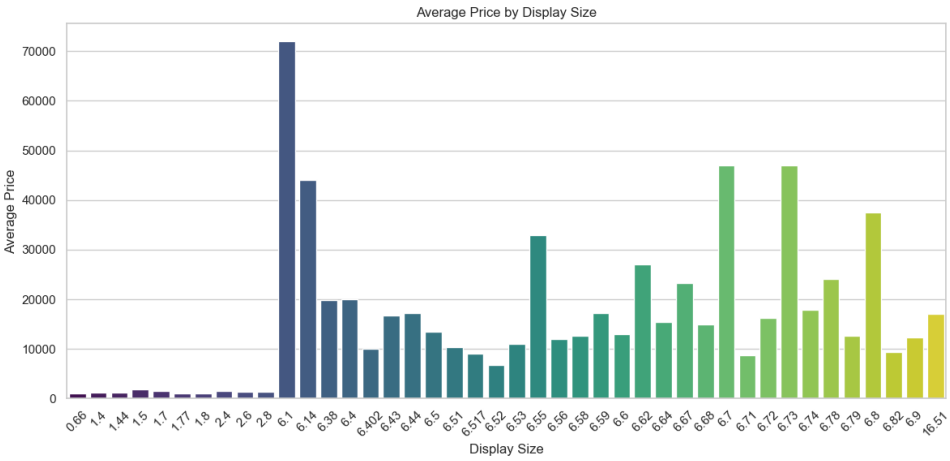
Output:



- ✓ scatter plot using Seaborn to depict the relationship between mobile phone ratings and processor cores.
- ✓ Each data point is differentiated by brand, with the 'muted' color palette enhancing visual clarity. The figure size is set to 12x10 inches, and the plot is titled 'Processor Core vs Rating with Brand Differentiation.'
- ✓ The x-axis represents ratings, the y-axis represents processor cores, and x-axis labels are rotated for readability.
- ✓ This visualization facilitates the exploration of how processor cores and ratings vary across different mobile phone brands.
- ✓

18.Compare the average price of phones with different screen sizes

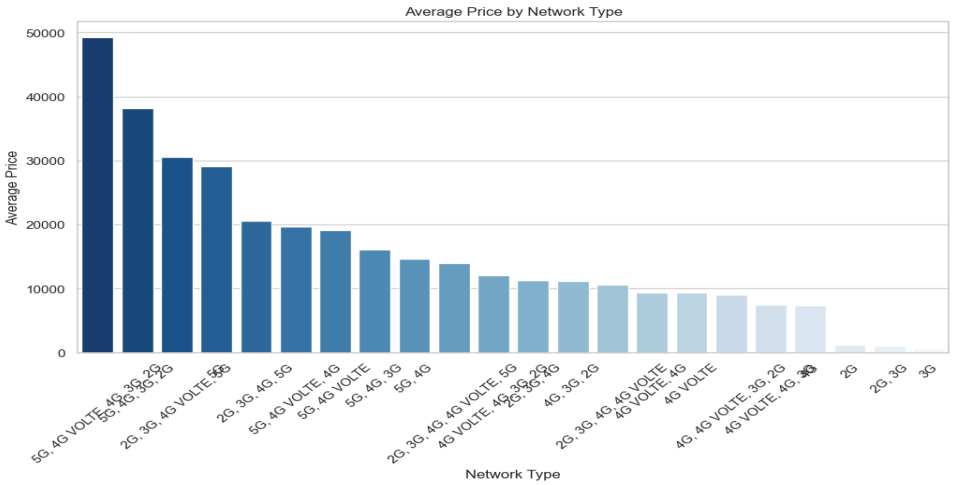
Output:



- ✓ calculates and visualizes the average prices of mobile phones based on their display sizes.
- ✓ Utilizing a bar plot with 'viridis' color palette for clarity, the figure is sized at 14x6 inches.
- ✓ The plot offers a straightforward representation of how average prices vary across different display sizes, facilitating a quick overview of potential pricing trends.

19.How is the distribution of Network Types across different brands, and does the Network Type correlate with the average Price of the mobile devices?

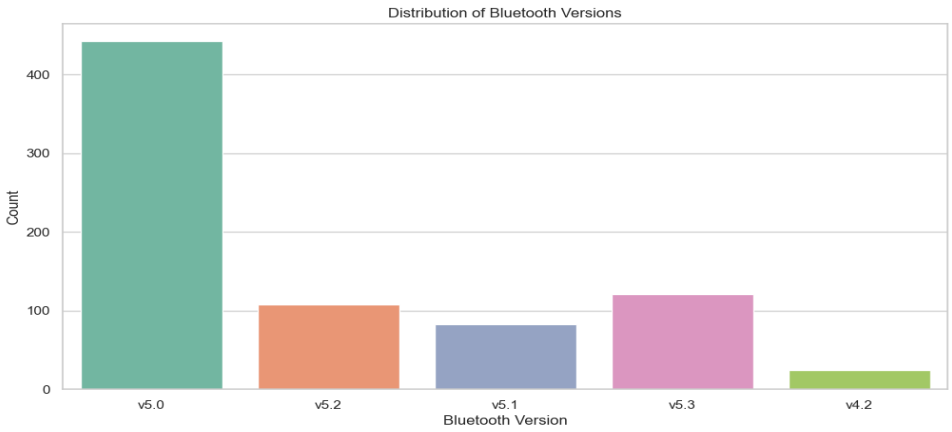
Output:



- ✓ calculates and visualizes the average prices of mobile phones grouped by network types.
- ✓ Using a bar plot with the 'Blues_r' color palette and a figure size of 12x6 inches, the plot succinctly illustrates variations in average prices across different network types.

20.What is the distribution of Bluetooth versions among the mobile devices in the dataset, and how does it vary across different brands?

Output:



- ✓ A concise count plot depicting the distribution of Bluetooth versions in the mobile dataset.
- ✓ Utilizing a 'Set2' color palette and a figure size of 12x6 inches, the plot offers a clear representation of the prevalence of each Bluetooth version.

10. Findings and Insights:

Price Distribution by Brand:

There is significant variation in smartphone prices among different brands.

Brands like Apple tend to have higher average prices compared to others.

Relationship Between Price and Rating:

There is no clear linear relationship between the price and the rating of smartphones.

Price alone does not determine the rating of a smartphone.

Distribution of Ratings:

The majority of smartphones have ratings between 4.0 and 4.5, indicating a generally positive reception.

Operating System Impact on Ratings:

There might be variations in ratings based on the operating system, as shown in the distribution by operating system.

Internal Storage and Battery Capacity:

Scatter plots indicate that there is no strong linear correlation between internal storage and battery capacity.

Brand Analysis:

The dataset contains smartphones from various brands, and each brand has its own pricing strategy.

Some brands dominate the market, while others may cater to specific niches.

Network Type and Price:

The average price varies depending on the network type.

SIM Type and Hybrid SIM Slot Impact:

The type of SIM card supported (Single or Dual) and the presence of a Hybrid SIM slot may influence the price.

Color Preferences:

A count plot provides insights into the distribution of different colors among smartphones.

Wi-Fi and GPS Support:

The availability of Wi-Fi and GPS support may have implications for smartphone prices.

11.Limitations:

Assumption of Normality:

Some statistical tests assume normality in the distribution of data. If the data does not meet this assumption, the results of certain analyses may be less reliable.

Outliers Handling:

Outliers were visualized using box plots but not systematically addressed. Extreme values can skew statistical measures and impact the generalizability of findings.

Limited Statistical Tests:

The analysis has used basic statistical tests, and more sophisticated methods might be needed for a deeper understanding of relationships.

Unexplored External Factors:

External factors such as market trends, economic conditions, or technological advancements that might influence smartphone attributes are not considered.

12.Recomendations:

Outlier Management:

Further investigate and manage outliers in the dataset. Outliers may represent unique cases or errors in data entry. Understanding their nature is crucial for accurate analysis.

Time-Based Analysis:

If available, consider obtaining a time-series dataset to analyze trends and changes over time in the smartphone market. This can provide insights into evolving customer preferences and technological advancements.

Market Positioning:

Identify the brands with the highest average prices and ratings. Consider the market positioning of these brands and assess whether there are specific features or strategies contributing to their success.

Deeper Statistical Tests:

If required, perform more advanced statistical tests or hypothesis testing to explore relationships between variables in greater detail. This can provide a more nuanced understanding of the dataset.

Monitoring Industry Trends:

Stay informed about broader industry trends, technological advancements, and changing consumer behaviours. This external context is essential for understanding the implications of your findings in a dynamic market.

13.Conclusion

Conclusion: Unravelling the Tapestry of Mobile Technology Trends

In the dynamic landscape of the Mobile Technology Market, our exploration into the dataset has provided valuable insights into the diverse facets of mobile phones. As we conclude our analysis, several key findings and considerations emerge, shaping a comprehensive understanding of this evolving industry.

1. Brand Dynamics: Brands play a pivotal role in defining the mobile market. The dataset showcases significant variation in pricing strategies among different brands. Notably, Apple stands out with higher average prices, suggesting a unique market positioning. Understanding the brand landscape is crucial for market stakeholders seeking to navigate consumer preferences and competition dynamics.

2. Price-Rating Dynamics: Contrary to expectations, there is no clear linear relationship between the price and user rating of smartphones. This challenges the assumption that higher-priced phones inherently lead to higher user satisfaction. Manufacturers should recognize that consumer perception is influenced by a multitude of factors beyond price, emphasizing the need for a holistic approach to product development and marketing.

3. Network Type Impact: Our analysis reveals that the average price varies depending on the network type, indicating that connectivity features contribute to pricing strategies. This insight is valuable for both consumers, making informed purchasing decisions, and manufacturers, aligning product offerings with market demands.

4. Colour Preferences and Design: The count plot on color preferences provides a glimpse into consumer choices. Manufacturers can leverage this information to align their product designs with popular color trends, catering to aesthetic preferences that influence purchasing decisions.

5. Recommendations for Future Analysis: To enrich future analyses, addressing outliers systematically, exploring time-based trends, and incorporating more advanced statistical tests are recommended. A deeper understanding of the market requires continuous monitoring of industry trends, technological advancements, and evolving consumer behaviors.

In conclusion, the Mobile Technology Market is a vibrant and competitive arena. This dataset serves as a valuable lens through which we've glimpsed the intricate interplay of factors influencing consumer choices and market dynamics. As technology continues to advance and consumer preferences evolve, staying attuned to these shifts will be instrumental for businesses and

analysts alike. Our journey into this dataset marks a step toward unraveling the tapestry of mobile technology trends, paving the way for informed decision-making in this dynamic industry.

14.References

Data Sources:

Kaggle. Mobile Trends Dataset.

Libraries used:

Python

Pandas

NumPy

seaborn: statistical data visualization

Matplotlib

15.Acknowledgments:

This comprehensive analysis would not have been possible without the collaboration and contribution of various entities and resources. We express our gratitude to:

Kaggle Community: The dataset used in this analysis was obtained from Kaggle, and we extend our appreciation to the contributors who curated and shared this valuable dataset with the community.

Open-Source Libraries: The Python programming language and open-source libraries such as NumPy, Pandas, Matplotlib, and Seaborn provided the foundational tools for data manipulation, analysis, and visualization. We thank the dedicated developers and contributors behind these libraries for their continuous efforts in making robust tools accessible to the community.

Educational Institution: This project is a part of our coursework, and we acknowledge our educational institution for providing the framework and resources for learning and applying data analysis techniques. The knowledge gained through this project contributes to our academic and professional growth.

Mentors and Instructors: Special thanks to our mentors and instructors who provided guidance, feedback, and support throughout the project. Their expertise and insights have been instrumental in navigating challenges and refining the analysis.

Upgrad: The broader data science and technology community has been a valuable source of information and assistance. We are grateful for the shared knowledge that has contributed to the success of this project.

This project stands as a collective effort, and we acknowledge the collaborative spirit that drives advancements in the field of data science and technology.

Project Code

- Include a link to the code/Jupyter notebook used for the analysis.
- Include link for dataset also
- Include presentation link