

Detailed Analysis Report

1. Introduction

The study is aimed at understanding the structure of the dataset 'Mobiles Dataset (2025)', the distribution of numerical variables including how these variables are correlated with other categorical factors. Many visualization methods have been adopted to study the patterns and trends.

2. Data Loading and Preprocessing

- **Libraries Used:**

The notebook starts with the standard import of data science libraries. These have been pandas, numpy, matplotlib.pyplot, and seaborn. These are important data manipulation and visualization libraries.

- **Dataset Loading:**

The dataset 'Mobiles Dataset (2025)' is read into a pandas DataFrame.

- **Data Cleaning:**

Filling of missing values, deletion of duplicates, treatment of outliers and relevant column types (numerical vs. categorical) are set.

3. Key Findings

- **Exploratory Data Analysis (EDA):** This has involved various visualization methods in investigating the dataset as follows:

a. Univariate Analysis for Numerical Columns

Histograms

- **Observation:**

RAM (in GB): There is to be a central concentration with a long right tail, implying that most devices have moderate RAM while a few high-in models are extending this range.

Battery Capacity: Mechanism appears to be normal distribution type but has a slight right-skew. It signifies that almost all devices are at about the same capacity with a few having very high values.

Launched Price (India): It is positively skewed whereby most devices lie in the lower tier and only a few occupy the extreme high-quality end of the spectrum.

- **Meaning:**

It tells that the market is flooded with standard models and few high-performance devices.

Box Plots

- **Observation:**

- **Battery Capacity:** The box plot shows a clear median with several outliers above the upper whisker; this indicates that a few models have abnormally high battery capacities.

- **Launched Price (India):** Features heavily spread with lots of high-price outliers, suggesting very distinct premium products.

b. Univariate Analysis for Categorical Columns

Count/Bar Charts

- **Observation:**
 - **Processor Processed:** The count plot indicates one dominant processor type (Snapdragon) while other processors such as MediaTek, Exynos, and Kirin appear in much lower frequencies.
- **Meaning:**

This imbalance reflects a dominant processor in the market.

Pie Charts

- **Observation:**
 - The pie chart shows a large percentage for the main processor and several smaller slices for the remaining types.

c. Bivariate Analysis

Numerical vs. Numerical

Scatter Plots and Pair Plots

- **Observation:**
 - **RAM vs. Launched Price (India):** Scatter plots show a clear positive linear trend; devices with higher RAM generally command higher prices.
 - **Battery Capacity vs. Mobile Weight:** Shows a weak linear relationship with most points clustered centrally and a few notable outliers.
- **Meaning:**

The strong trend between RAM and price implies that higher memory is valued, while the weak trend between battery capacity and weight suggests these features are influenced by other design factors.

Correlation Heatmap

- **Observation:**
 - The heatmap highlights strong positive correlations (e.g., between RAM and Launched Price) and moderate associations for others like Mobile Weight and Battery Capacity.
- **Meaning:**

This helps in identifying key for further model development.

Categorical vs. Numerical

Mobile Weight by Company

- **Observation:**
 - **Box Plot:** Shows that some companies have a narrow weight range while others exhibit high variability and outliers.
 - **Violin Plot:** Reveals bimodal distributions in certain companies, suggesting both lightweight and heavy models are offered.
 - **Bar Plot (Mean Mobile Weight):** Shows clear differences in average weights among companies.
- **Meaning:**

These plots indicate varying design standards across companies, where some brands maintain consistency while others offer a diverse product mix.

Mobile Weight by Processor

- **Observation:**
Box/Violin Plots: Devices with a dominant processor (e.g., Snapdragon) appear generally heavier compared to those with alternatives (e.g., MediaTek).
- **Meaning:**
This suggests that the choice of processor might be linked to additional design factors, such as the need for cooling or structural support.

Battery Capacity by Processor

- **Observation:**
Box/Violin Plots: Indicate that certain processor groups have a higher median battery capacity with varied spreads.
Bar Plot: Shows one processor category has a significantly higher average battery capacity.
- **Meaning:**
Higher battery capacities paired with specific processors may reflect a strategy to support power-intensive features.

Battery Capacity by Company

- **Observation:**
Box/Violin Plots: Show differences in median battery capacities and dispersion among companies.
Bar Plot: Highlights that some companies consistently deliver devices with higher battery capacities.

Launched Price (India) by Company

- **Observation:**
Box/Violin Plots: Display wide price ranges and even multimodal distributions for certain companies, indicating both budget and premium offerings.
Bar Plot: Clearly differentiates average price levels across companies.
- **Meaning:**
These differences underline market segmentation, with some companies targeting both entry-level and high-end consumer segments.

Launched Price (India) by Processor

- **Observation:**
Box/Violin Plots: Devices with specific processors show higher medians and long tails in the price distribution.
Bar Plot: Confirms significant average price differences between processor types.
- **Meaning:**
The association guides the disclosure that the processor technology occupies a role of significant importance in defining the device's marketplace value.

RAM by Processor

- **Observation:**
Box/Violin Plots: Reveal distinct median differences in RAM across various processor types.

- **Meaning:**
This suggests that performance-oriented devices require higher memory configurations with advanced processors.

RAM by Company

- **Observation:**
Box/Violin Plots: Indicate that some companies consistently offer higher RAM, while others show a multimodal spread hinting at varied product lines.
Bar Plot: Clearly differentiates companies based on average RAM.

Company vs. Processor (Categorical vs. Categorical)

- **Observation:**
 - **Heatmap:** Demonstrates that certain companies predominantly use one type of processor.
 - **Stacked Bar Chart:** Illustrates the processor mix within each company.
- **Meaning:**
These findings highlight technology choices that differentiate companies in the competitive market.

d. Multivariate Analysis

Grouped Bar Plots (Company & Processor vs. Launched Price, RAM, Battery Capacity, Mobile Weight)

- **Observation:**
 - Grouped bar plots show that within a company, devices equipped with premium processors tend to have higher prices, more RAM, larger batteries, and sometimes heavier builds.
- **Meaning:**
This suggests that processor choice is a key driver of overall device specifications and market positioning.

3D Scatter Plots

- **Observation:**
 - **Mobile Weight vs. RAM vs. Launched Price (India):** Devices with higher RAM and heavier builds lie in the premium price range.
 - **RAM vs. Battery Capacity vs. Launched Price (USA):** There is an identifiable trend showing that higher levels of RAM or higher capacity of battery will subsequently raise the prices.

4. Conclusion:

As was seen in our analysis, majority of the mobile device specifications are seen to follow standard specifications, with a distinct segment belonging toward the ultra-high-end models. The prevalence of certain types of processors in manufacturers stipulates critical characteristics such as RAM, battery capacity, and price.