KATHMANDU UNIVERSITY

DHULIKHEL, KAVRE

COMP-482

Mini Project Report On

"Dashboard using PLotly"

Date: 2023/12/18

Submitted By:

Saugat Poudel

Roshan Jaishi

Group: CE

Submitted To:

Dr Rajani Chylyadyo

Department of Computer Science and Engineering

TABLE OF CONTENTS

Keywords

GPU - Graphics Processing Unit

Wh - Watt hour

KPI - Key Peformance Indicator

PSU - Power Supply Unit

Abstract

This project introduces an interactive dashboard crafted with Plotly's Dash framework, designed to empower users with the ability to explore and compare diverse components of notebooks. Through intuitive visualizations, users can dynamically analyze and compare essential features such as pricing, GPU types, and other components with the help of a dropdown that contains the processor type for various notebooks. The dashboard provides a user-friendly interface for gaining insights into the intricate relationships and variations across different notebooks within the dataset. This report details the design, implementation, and significance of the Notebook Dataset Dashboard, highlighting its role in facilitating comprehensive comparative analyses in the realm of notebooks, ultimately contributing to informed decision-making for users across various domains.

Chapter 1: Introduction

1.1 Background

In the dynamic landscape of digital technology, notebooks have become indispensable tools, serving as the crucibles of creativity, collaboration, and knowledge dissemination across diverse domains. As the prevalence of notebooks continues to rise, so does the need for robust analytical frameworks to comprehend their multifaceted nature. This project endeavors to address this need through the development of a sophisticated dashboard, powered by Plotly's Dash framework, that enables users to explore and dissect the nuances of a rich dataset encompassing a myriad of notebooks. The Notebook Dataset Dashboard serves as a gateway to unlock valuable insights into the composition and characteristics of notebooks. With a focus on interactive visualizations, including pie charts, bar graphs, box plots, and more, users gain the ability to navigate through programming language distributions, revision trends, and the comparative analysis of essential components such as pricing and GPU types. This project aims to bridge the gap between data and understanding, providing a user-friendly interface for researchers, educators, and developers to glean actionable insights from the expansive world of notebooks. As notebooks continue to evolve in complexity and utility, the Notebook Dataset Dashboard positions itself as an invaluable tool for those seeking to comprehend and harness the potential embedded in digital documentation and collaboration. This report delves into the intricacies of the dashboard's design and implementation, shedding light on the significance of employing Plotly's Dash to create an engaging platform for exploring the diverse facets of notebook data. Through this project, we embark on a journey to demystify the landscape of notebooks, fostering a deeper understanding of their components and empowering users with the tools to make informed decisions in an ever-evolving technological ecosystem.

1.2 Objectives

- Create a user-friendly and interactive dashboard using Plotly's Dash framework to visualize and explore the dataset of notebooks
- Implement visualizations, such as pie charts and bar graphs, to illustrate the distribution of attributes used across the dataset.
- Enable users to perform comparative analyses by incorporating features for exploring and contrasting different components of notebooks, including price, GPU types, and other relevant parameters.

1.3 Tools Used

The project utilizes Dash for web framework functionality, Plotly Express for interactive data visualization, and NumPy for numerical operations. Additionally, custom CSS styling is applied to enhance the visual presentation of the web application

Chapter: 2 Dataset Description

Dataset Link: https://www.kaggle.com/datasets/emilrueh/laptops-and-notebooks-from-2020-2023

The dataset utilized in this project comprises detailed information about various notebooks, capturing essential features that play a pivotal role in user decision-making. The dataset includes the following attributes:

- **Name:** The name or model identifier of the notebook.
- **Price (EUR):** The cost of the notebook in Euros.
- **Display (Inch/CM/Resolution/Ratio):** Specifications related to the notebook display, including size, resolution, and aspect ratio.
- **Weight (kg):** The weight of the notebook.

- **Dimensions (Height/Width/Depth in mm):** Physical dimensions of the notebook. Operating System: The operating system installed on the notebook.
- **CPU Processor:** Information about the central processing unit (CPU) of the notebook.
- **RAM Memory:** The random-access memory (RAM) capacity of the notebook.
- **GPU (Integrated/Extra):** Details about the graphics processing unit (GPU), both integrated and additional.
- **Internal Storage (GB/Type):** Information on the internal storage capacity and type (e.g., SSD).
- **Battery Life (h)/Capacity (Wh):** Battery-related metrics, including life in hours and capacity in watt-hours.
- **PSU Watts:** Power supply unit wattage.
- **Audio System/Speakers Count:** Specifications related to the audio system and the number of speakers.
- **Touchscreen/Backlit/Keyboard/Numpad/Webcam/Bluetooth/Wi-Fi:**Boolean values indicating the presence of these features.
- **Bluetooth Version/Wi-Fi Standard:** Versions of Bluetooth and Wi-Fi standards supported.
- **Product EAN/SKU/Release Year/Category:** Additional identifiers, release year, and the category of the notebook.

The dataset encompasses a diverse range of notebooks, including those from Apple, MSI, and Lenovo, each characterized by distinct specifications and configurations.

2.1 Pre-processing Steps:

Before analysis, the dataset underwent pre-processing to ensure its cleanliness and suitability for exploration. This involved handling missing or erroneous values, converting data types where necessary, and addressing any anomalies that could impede accurate analysis. The cleaned dataset is now poised for comprehensive exploration and visualization through the development of the Notebook Dataset Dashboard.

Chapter 3 : Dashboard Design and Features

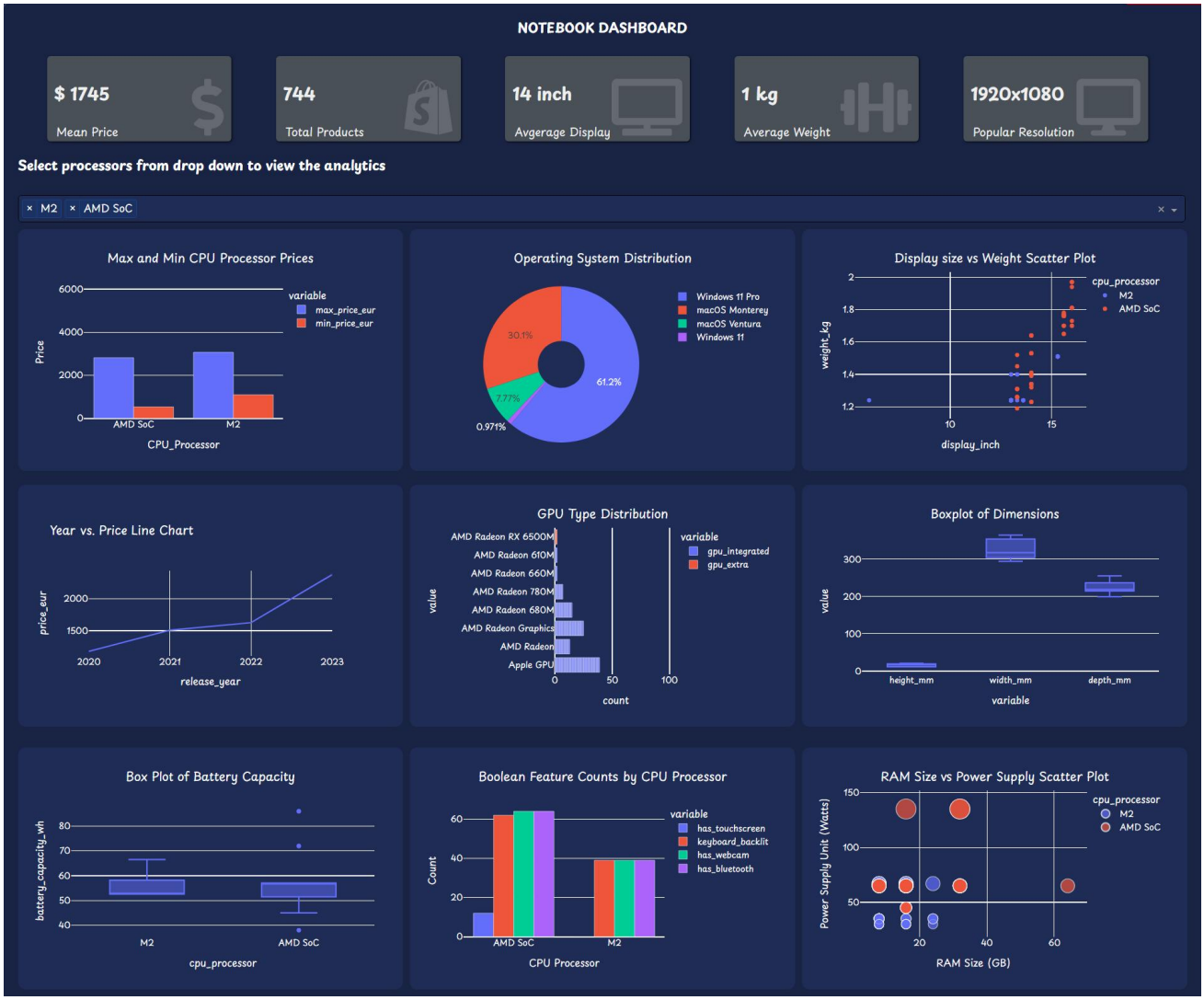

fig 1 (Dashboard Design)

3.1 Features:

1. KPI Indicator:

- Offers users a quick overview of essential notebook characteristics.
- Users can leverage these indicators to make decisions aligned with their preferences and requirements.

2. Dropdown Selector:

- Empowers users with a dynamic dropdown menu for selecting processors in real-time.
- Users can choose processor from the dropdown and view the analytics.

3. Analytics Sections:

- Organizes the dashboard into specialized sections for analytics tailored to the chosen processor.
- Utilizes various visualizations like line charts and pie charts for clear insights.

4. Real-time Data Updates:

- Provides real-time updates across visualizations based on the selected processor.
- Enables seamless exploration and comparison without manual refreshes.

5. Data Drill-Down:

- Allows users to drill down into specific data points, enhancing analysis depth.
- Interactive elements like tooltips provide additional context for detailed exploration.

6. Navigational Support:

- Incorporates information panels and tooltips for user guidance.

Chapter 4: Visualization and Analysis

4.1 Grouped Bargraph for Price Comparision

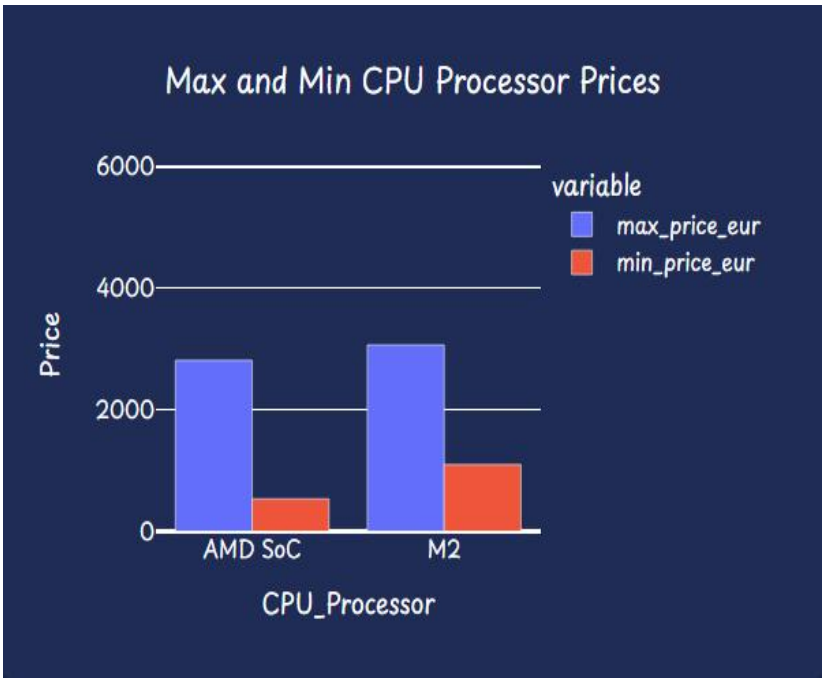

Fig 4.1 (Grouped Bargraph)

This bar graph shows the maximum and minimum prices of the notebooks based on the processor selected from the dropdown. This grouped bar graph also contains a hover information that contains max price for the max bar and min price for min bar respectively and also displays the name of those notebooks. This is shown in *fig (4.2 and 4.3)*

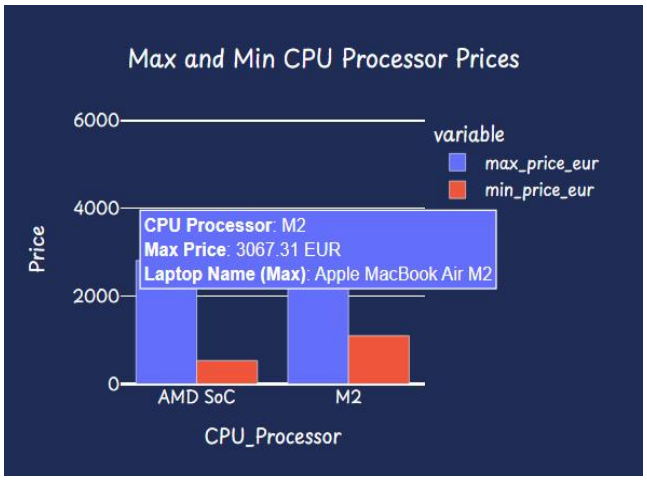

fig 4.2 (Max Price Bar Graph)

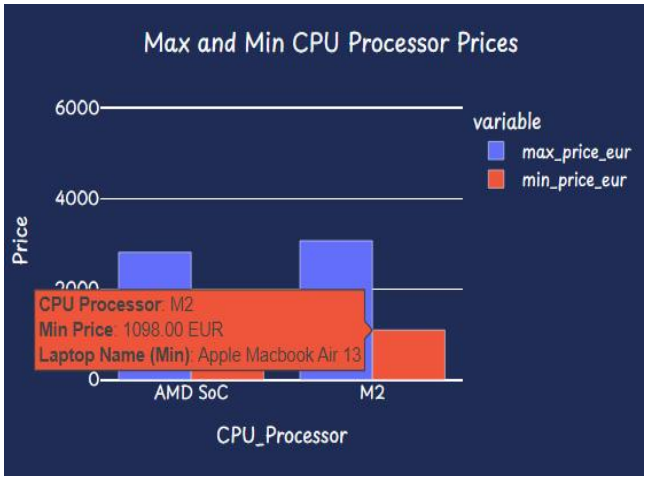

fig4.3 (Min Price Bargraph)

4.1.1 Key insights derived from visualization

- **Visualizing Price Distribution:**

  The bar graph visually represents the distribution of both maximum and minimum prices for different CPU processors. This allows stakeholders to quickly grasp the range of prices associated with each processor, providing insights into the variability within the dataset.

- **Identification of Outliers:**

  By presenting both maximum and minimum prices, the graph facilitates the identification of outliers or extreme values associated with specific CPU processors.

Example:

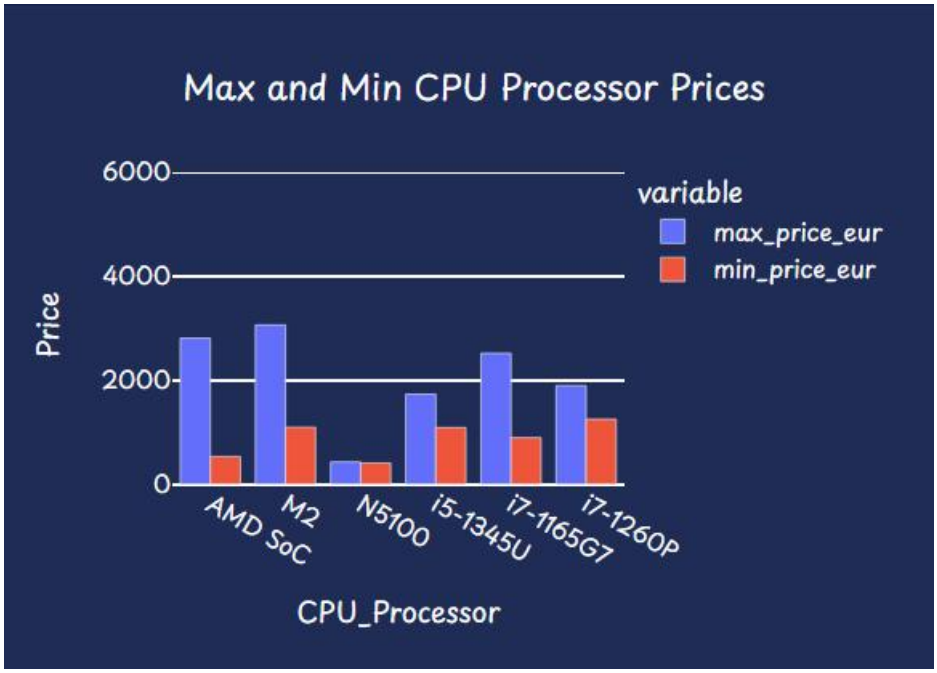

fig 4.4(Cpu processor Price comparison)

In the above figure(4.1) when six processors where selected we can clearly see from the max_price_eur bar that the notebook having M2 processor has the highest price as compared to others, whereras looking at the min_price_eur bar we can see that the notebook having N5100 processor has the least price as compared to others.

4.2. Piechart for operating system distribution

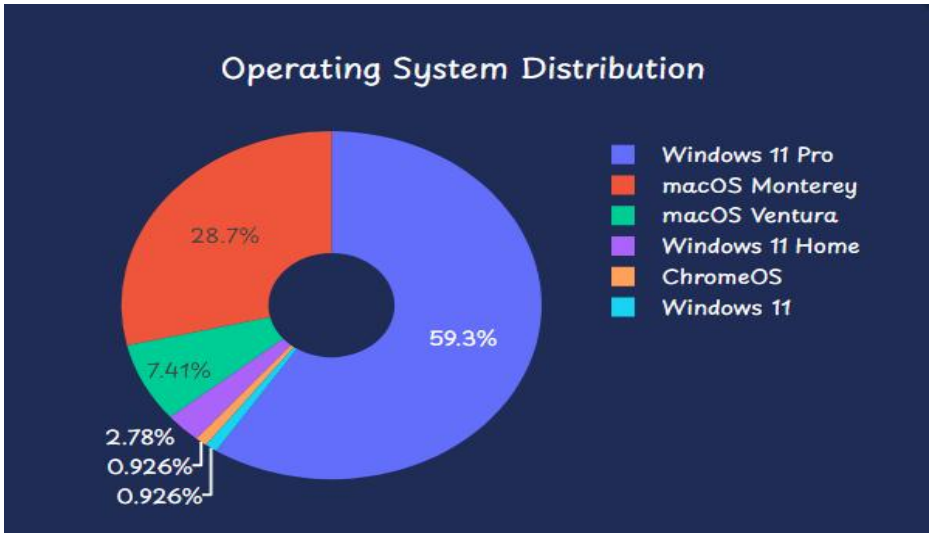

fig4.4 (Piechart for OS)

Fig(4.4) shows the operating system distribution for the selected processors.

4.2.1 Key insights derived from visualization:

- **Operating System Market Share:**

  The pie chart provides a clear representation of the market share of different operating systems within the selected CPU processors. The size of each slice corresponds to the proportion of laptops using a particular operating system.

- **Dominant Operating Systems:**

  The most dominant operating systems can be easily identified by the larger slices in the pie chart. In fig (4.4) we can clearly see that the dominant operating system is Windows 11 Pro. This insight is valuable for understanding the prevalent choices among users in the dataset.

- **Diversity of Operating Systems:**

  The distribution of smaller slices in the pie chart indicates the diversity of operating systems present in the selected CPU processors. In fig(4.4) we can clearly see that (windows 11 home, Chrome0S and Windows 11) contain smaller distribution as compared to others. This diversity may have implications for software compatibility, user preferences, or market trends

4.3. Scatter Plot

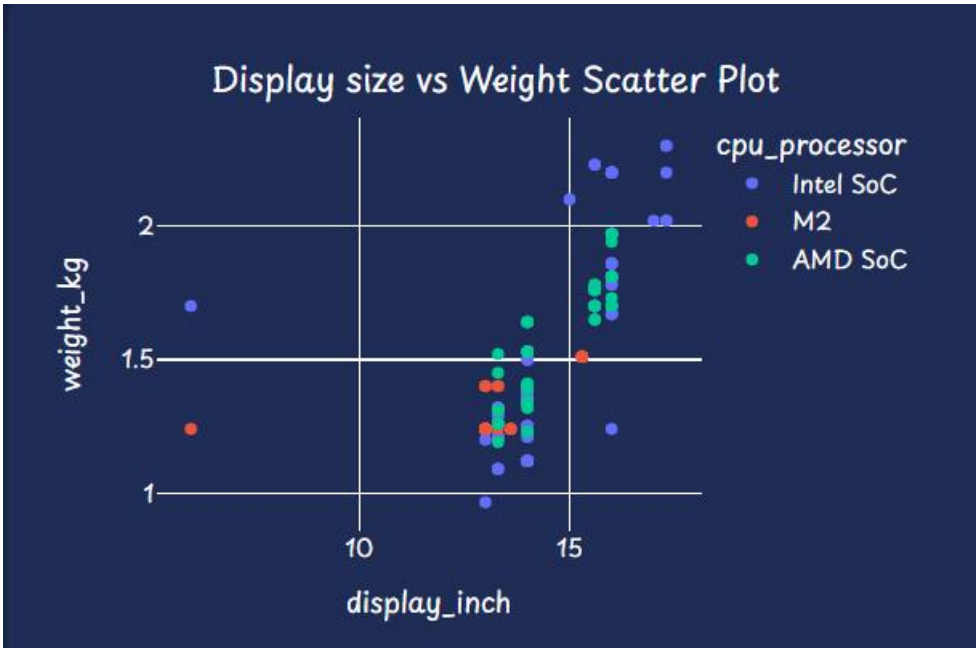

Fig4.5 (Scatter Plot)

4.3.1 Key insights derived from visualization

- **Display Size vs. Weight Relationship:**

  The scatter plot allows users to visually examine the relationship between display size (x-axis) and weight (y-axis). Insights into whether there is a correlation, pattern, or any outliers can be gained.

- **Identifying Clusters or Groups:**

  Clusters or groups of points in the scatter plot may indicate certain categories or classes of laptops. In *fig(4.5)* there is a cluster formation of lightweight notebooks with small displays and another cluster of heavier laptops with larger displays.

- **Outlier Detection:**

  Outliers in the scatter plot can be easily identified, indicating notebooks that deviate significantly from the general trend.

Fig4.6 (Outlier Scatter Plot)

From *fig(4.6)* we can clearly see that the Apple Macbook Air 13 with display_inch=6 and weight_kg=1.24 is the outlier as compared to the display and weight of other notebooks.

4.4. Line Chart for price comparison

fig4.7 (Line Chart for the year 2021)

fig 4.8 (Line chart for the year 2022)

4.4.1 Key insights derived from visualization:

- **Trend in Notebooks Prices:**

  The line chart provides a visual representation of how average notebook prices have changed over different release years. Users can observe trends, patterns, or fluctuations in pricing over time.

- **Identifying Price Fluctuations:**

  The chart allows users to identify periods of price fluctuations or stability. Sudden peaks or troughs in the line indicate significant changes in the average laptop prices during certain release years. In the above *fig(4.7)* and *fig(4.8)* we can clearly observe a drastic change in the price of notebooks, increasing from 1508 to 1626 in the years 2020 and 2021.

- **Release Year Impact on Price:**

  This line chart helps users assess the impact of the release year on laptop prices. Users can observe whether newer releases are associated with higher or lower average prices.

4.5 GPU Distribution Bar Chart:

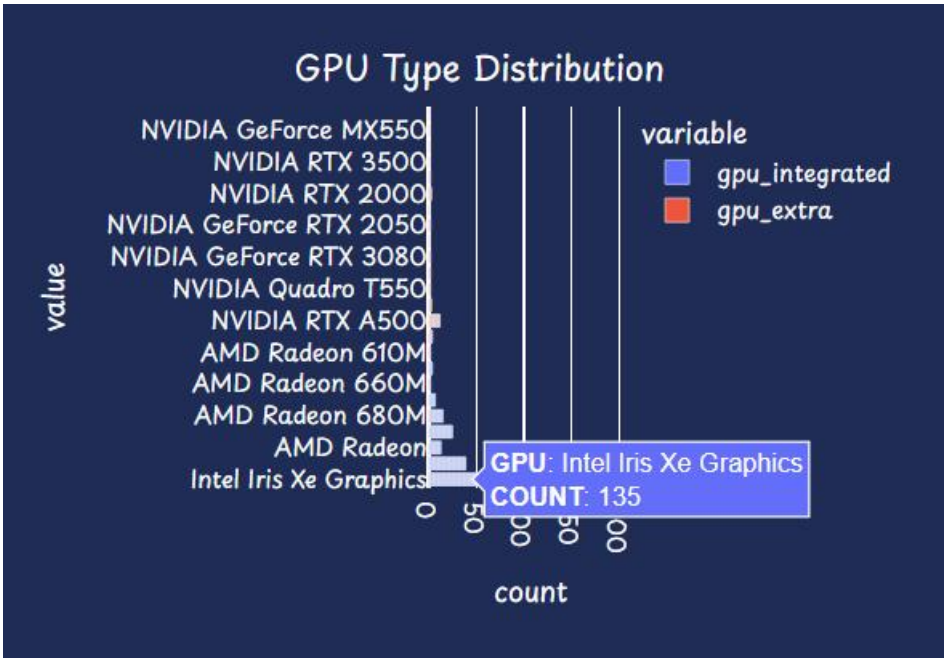

Fig4.9 (Integrated GPU count)

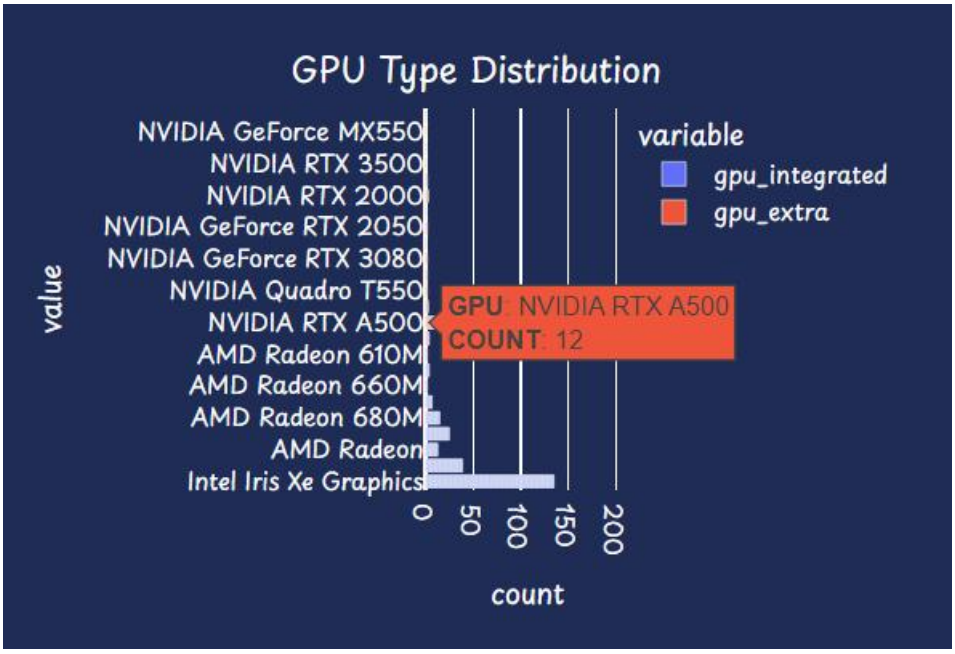

Fig4.10 (Dedicated GPU Count)

4.5.1 Key insights derived from visualization:

- **GPU Type Comparison:**

  The bar chart allows users to compare the distribution of two GPU types: 'gpu_integrated' and 'gpu_extra.' The lengths of the bars represent the count of

laptops with each GPU type, providing a visual comparison as shown in *fig(4.9*) and *fig(4.10).*

- **Dominant GPU Type:**

  Users can easily identify the dominant GPU type by observing the taller bar in the chart. The GPU type with a higher count is more prevalent among the selected CPU processors. In the fig() Intel Iris Xe Graphics is the dominant integrated graphics and Nvidia RTX A500 is the dominant dedicated graphics with overall count of 135 and 12 respectively in the dataset.

4.6 Dimensions Boxplot:

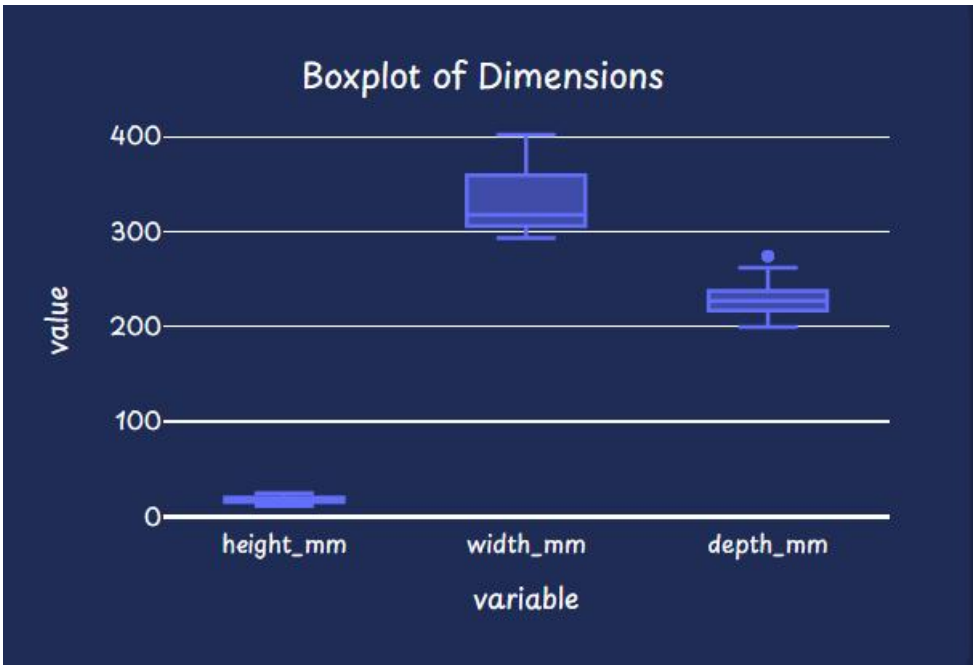

Fig4.11 (Boxplot of Dimensions)

4.6.1 Components of box plot:

- **Box (Interquartile Range, IQR):**

  The box in the middle of the plot represents the interquartile range (IQR), which is the range between the first quartile (Q1) and the third quartile (Q3). The width of the box corresponds to the spread of the middle 50% of the data.

- **Whiskers**:

  The whiskers extend from the box to indicate the range of the data. They typically represent 1.5 times the IQR. However, in some implementations, whiskers can extend to the minimum and maximum non-outlier data points.

- **Outliers:**

  Observations beyond the whiskers are considered potential outliers. Outliers are individual data points that fall outside a specified range.

- **Upper Fence:**

  The upper fence is a boundary beyond which data points are considered potential outliers. It is calculated as Q3 + 1.5 * IQR. Any data point above the upper fence is a candidate for being an outlier.

- **Lower Fence:**

  The lower fence is another boundary used to identify potential outliers. It is calculated as Q1 - 1.5 * IQR. Any data point below the lower fence is a candidate for being an outlier.

4.6.2 Key insights derived from visualization:

Selected Processor For Comparision:

Fig4.11 (Selected Processors)

4.6.2.1 For height:

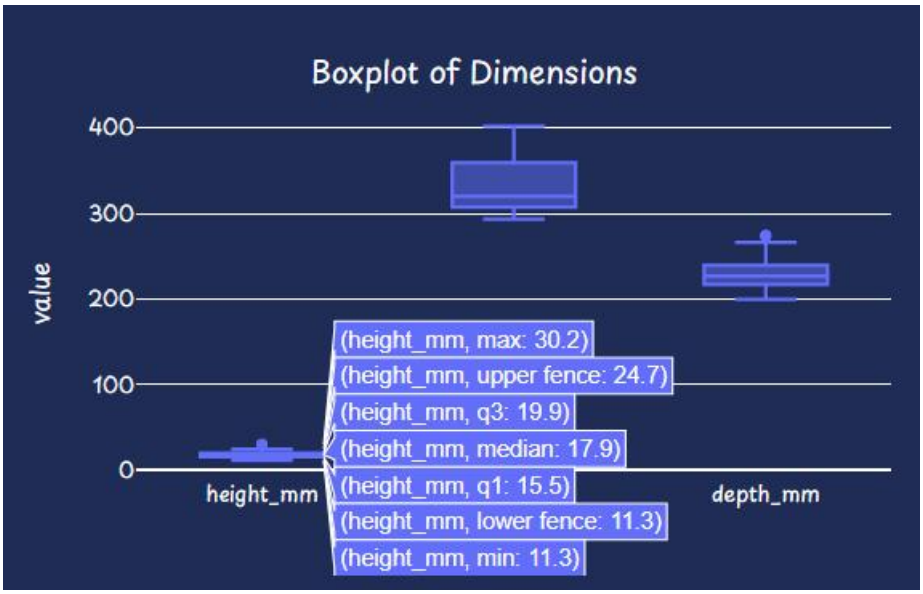

Fig4.12 (Height Boxplot)

From the above *fig(4.12)* we can see the the visualization of all the components of a boxplot for selected processor.

4.6.2.1.1 Identification of Outliers:

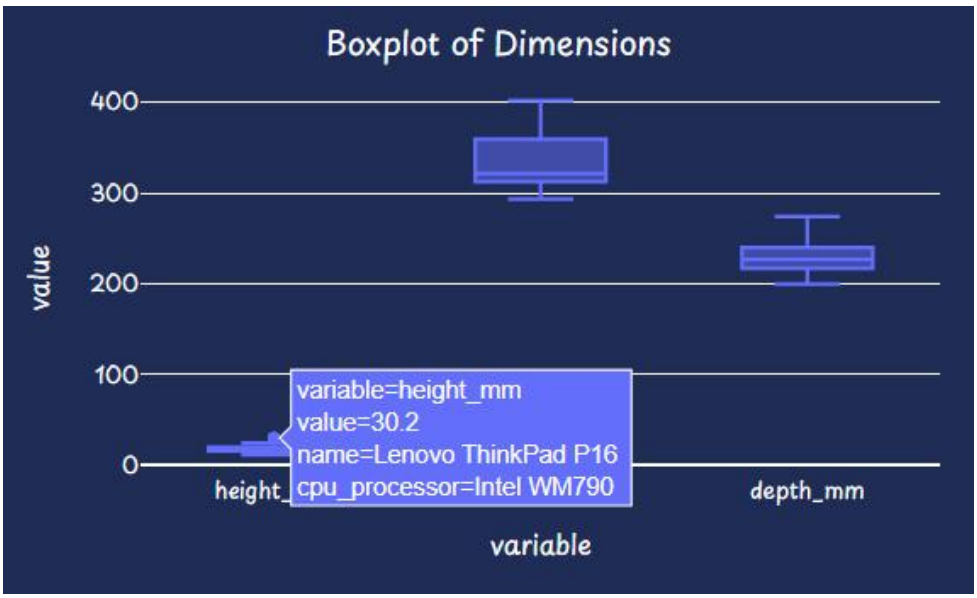

Fig4.13 (Height Outlier)

From the plot, it is evident that the outlier value is 30.2mm and the name of the notebook is Lenovo ThinkPad P16, whereas the upper fence for the selected processors was 24.7mm.

4.6.2.2 For width:

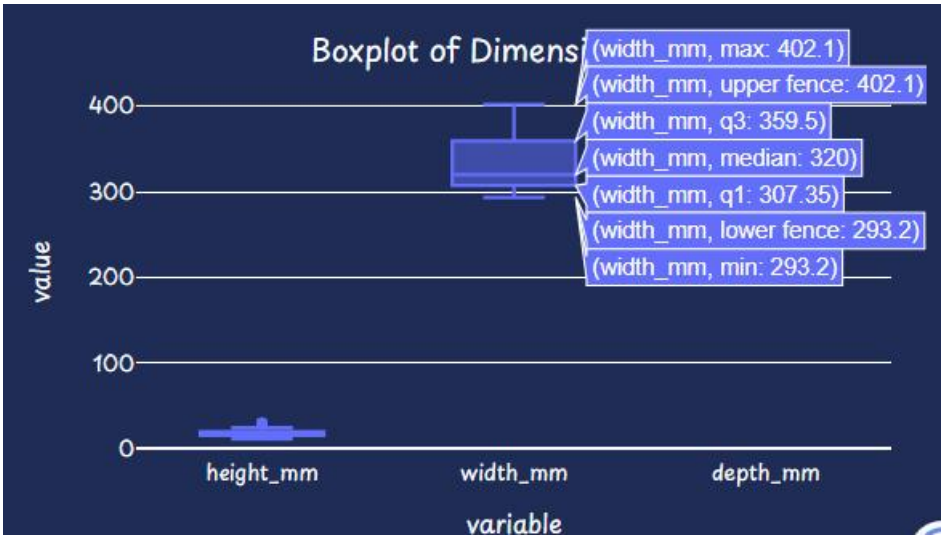

Fig4.14 (Width Boxplot)

The above fig(4.14) shows all the components of box plot for width dimension of our dataset for the selected processors shown in fig(4.11).

4.6.2.2.1 Identification for Outliers:

No outliers were identified for the dimension width.

4.6.2.3 For depth:

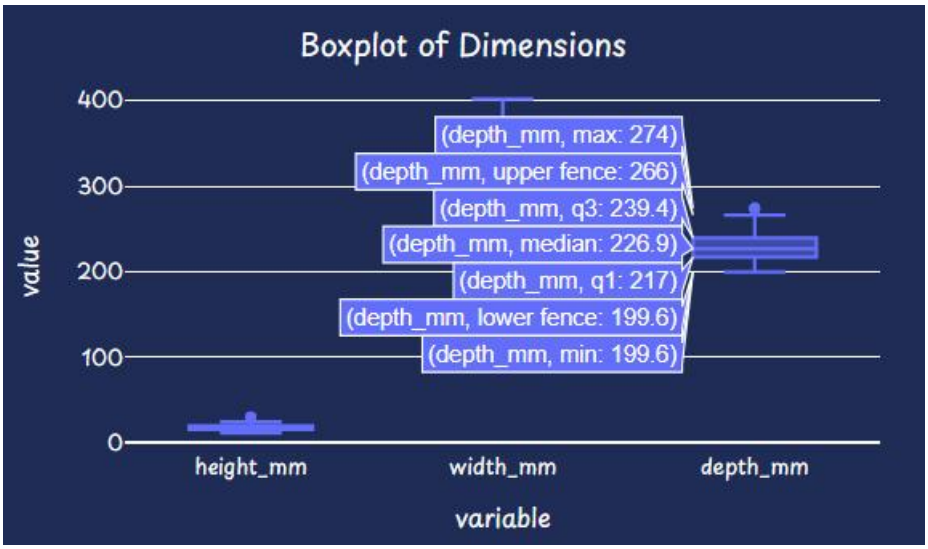

Fig4.15 (Depth Boxplot)

From above *fig(4.15)* we can see the visualization of all components of box plot for the selected processor as shown in *fig(4.11)*

4.6.2.3.1 Identification of outliers:

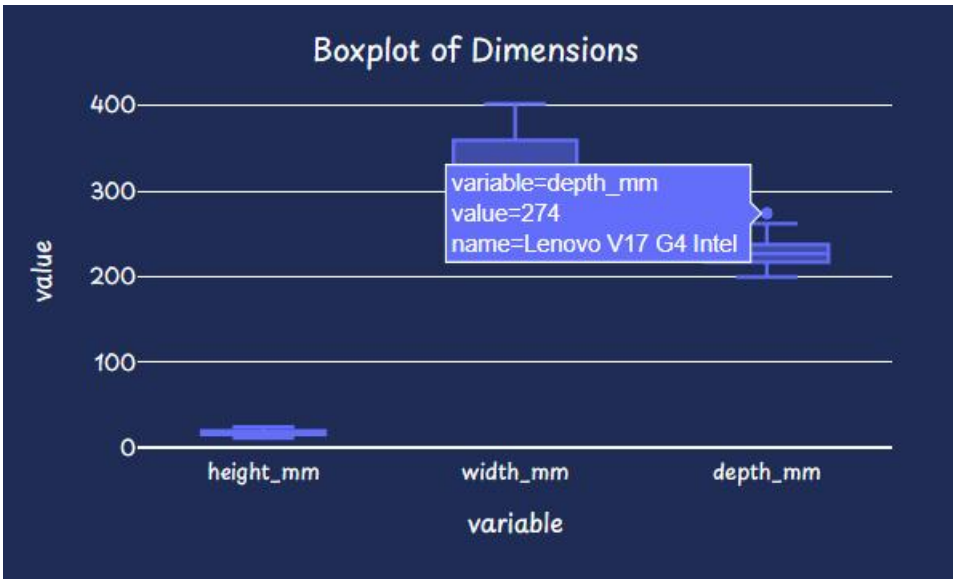

Fig4.16 (Depth Outlier)

From above fig(4.16) it is evident that the outlier is 274mm and the name of outlier notebook is Lenovo V17 G4 Intel, as the upper fence value is 266.

4.7 Battery Capacity Boxplot

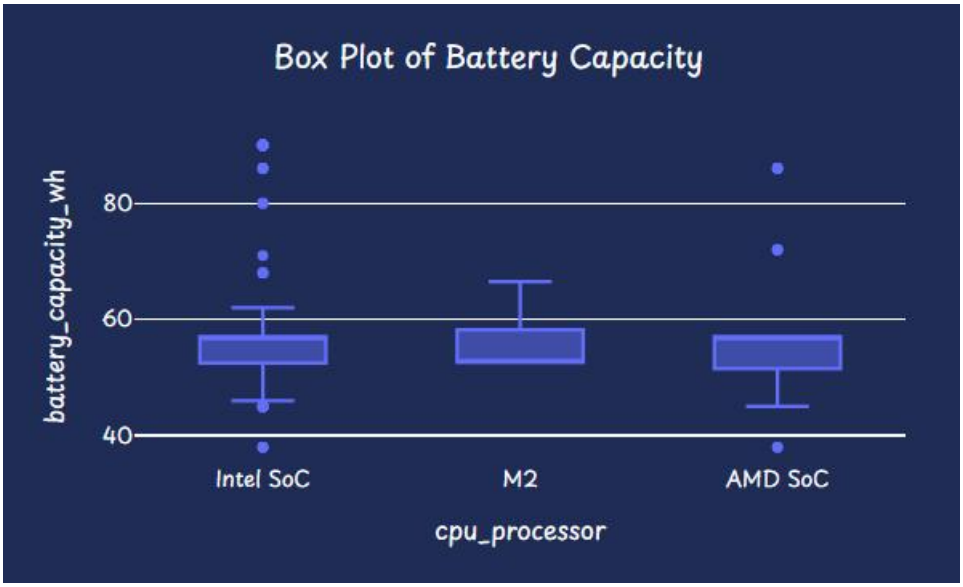

fig4.17 (Battery Capacity box plot)

4.7.1 key insights derived from visualization

- **Central Tendency:**

  The boxplot allows users to identify the central tendency of battery capacity for the selected CPU processors. This is represented by the median line within each box

- **Variability in Battery Capacity:**

  The spread of the box indicates the variability in battery capacity among the selected CPU processors. A wider box suggests greater variability, while a narrower box indicates more consistent battery capacity.

- **Outlier Detection:**

  Individual points outside the whiskers of the boxplot represent potential outliers in terms of battery capacity. These outliers may be CPUs with unusually high or low battery capacities compared to the majority.

  Example:

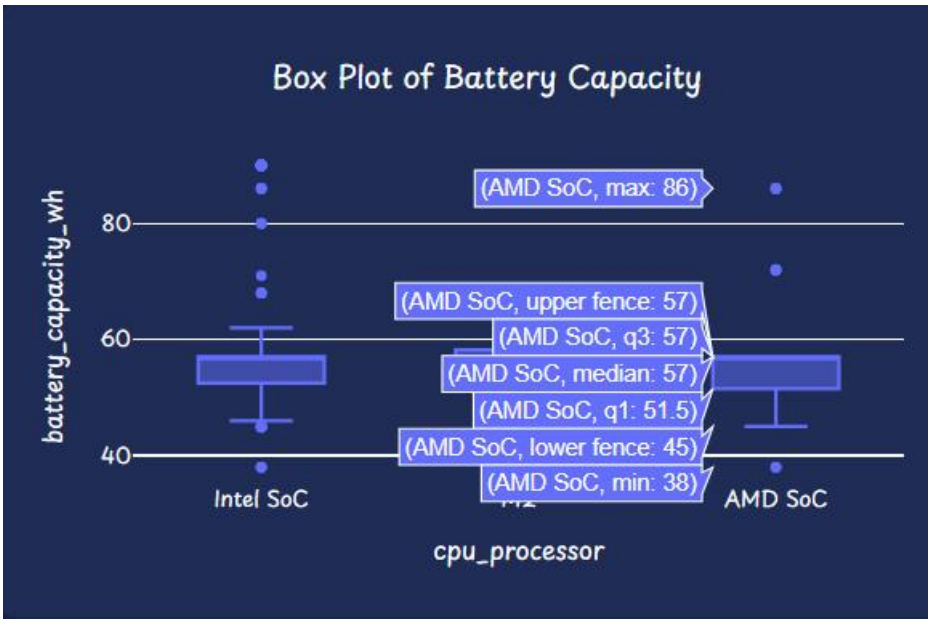

Fig4.18 (Batterycapacity Box plot)

from the above fig(4.18) we can see that any value beyond 57 and below 45 are the outliers. for the selected processor AMD SoC notebook Lenovo ThinkPad P16s is the outlier with max battery capacity= 86 as shown in fig(4.18)

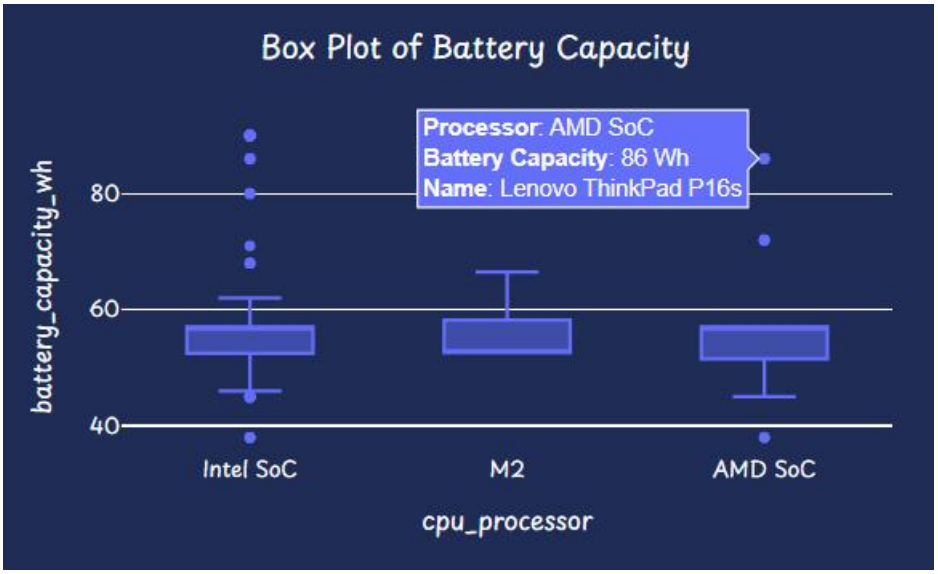

fig4.19 (Battery Capacity Outlier)

- **Comparison Across Processors:**

  Users can visually compare the distribution of battery capacity across different CPU processors. This comparison helps in identifying which processors tend to have higher or lower battery capacities.

4.8 Touchscreen,keyboard backlit,webcam and bluetooth grouped bar plot:

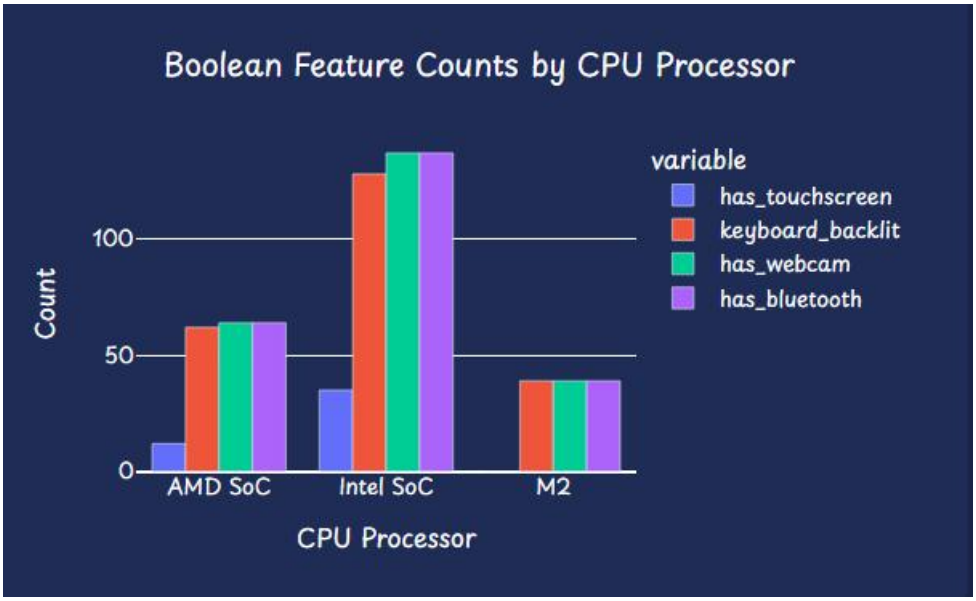

fig4.20 (Feature Bar Plot)

4.8.1 key insights derived from visualization:

- **Comparison of Boolean Features:**

  The bar graph allows users to compare the counts of different boolean features (has_touchscreen, keyboard_backlit, has_webcam, has_bluetooth) across selected CPU processors. This facilitates the identification of patterns and preferences in the inclusion of these features.

- **Identification of Dominant Features:**

  Users can quickly identify which boolean features are more prevalent or dominant among the selected CPU processors. This insight is valuable for understanding the common characteristics of notebooks with specific CPU processors.

- **Variability in Feature Counts:**

  The variability in the height of the bars indicates how much the counts of boolean features vary among the selected CPU processors. Higher variability suggests diversity in feature inclusion.

Example:

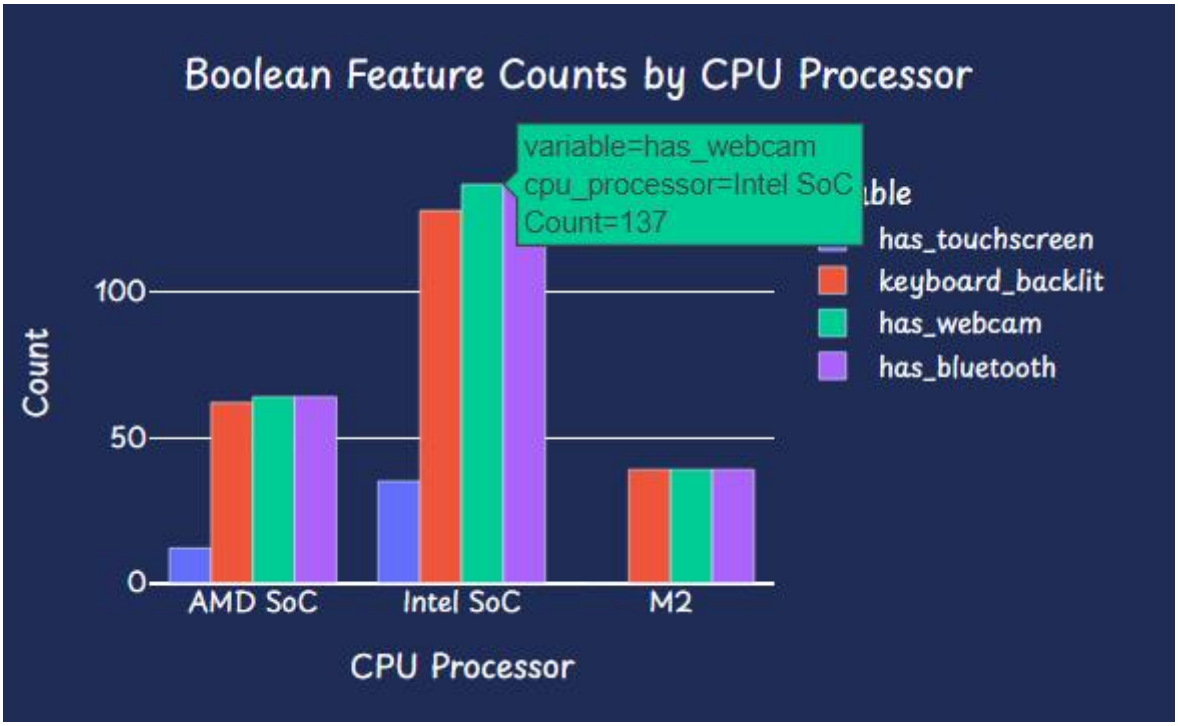

Fig4.21(variability in features)

From the above fig(4.21), we can conclude that notebooks with Intel SoC processors have a higher probability of possessing all the features. This inference is based on the observation that the bar for each attribute associated with Intel SoC processors is notably higher compared to the bars corresponding to the other two processors. The elevated counts across multiple attributes suggest a greater likelihood of Intel SoC-powered notebooks being equipped with a variety of features such as touchscreen, backlit keyboard, webcam, and Bluetooth.

4.9 Ram Size and Power Supply Unit Scatter Plot:

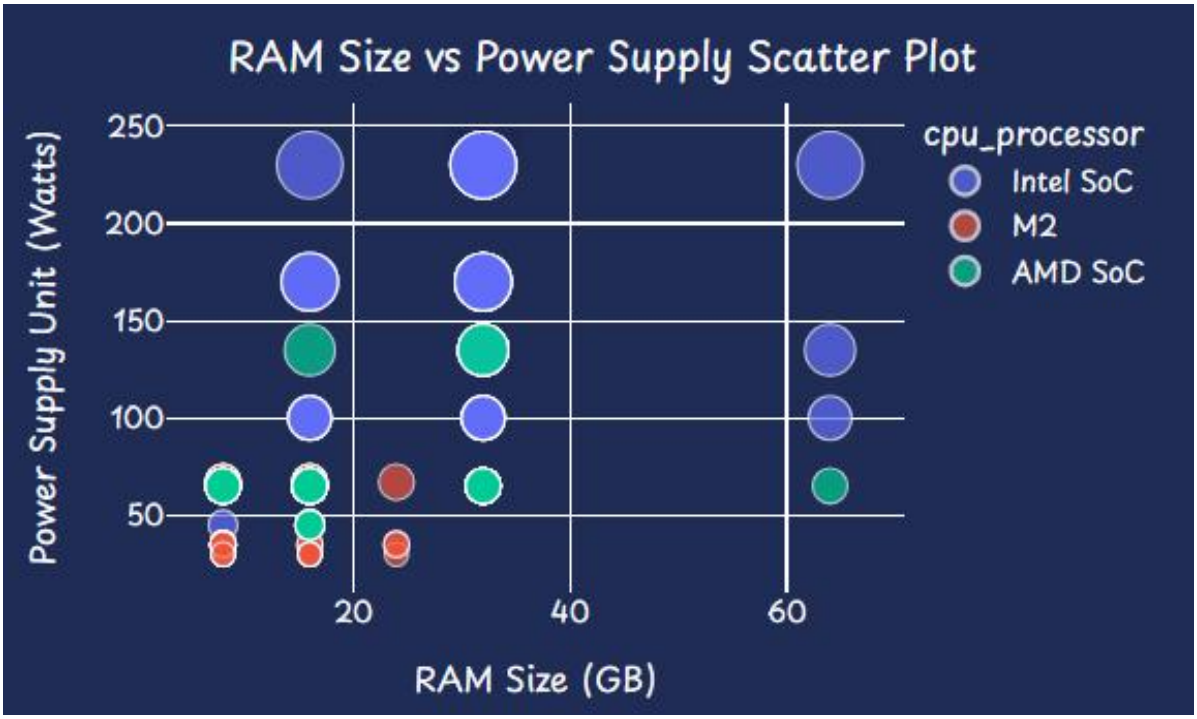

Fig4.22 (Ram and PSU scatter plot)

4.9.1 Key insights derived from visualization:

- **Correlation between RAM Size and Power Supply:**

  The scatter plot allows users to observe the correlation between RAM size (in gigabytes) and power supply (in watts) for the selected CPU processors. Users can

identify trends such as whether higher RAM size correlates with increased power supply.

Example:

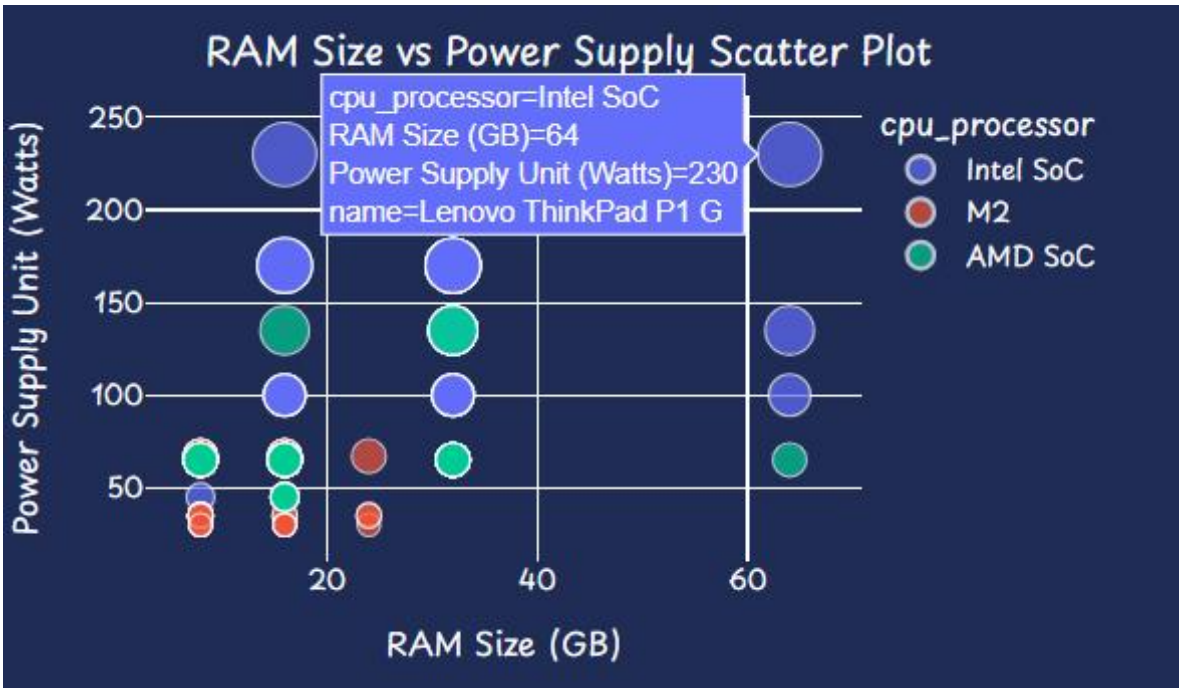

Fig4.23 (Outlier for Ram size and PSU)

From the above fig(4.23), we can conclude that there is a correlation between RAM size and power supply, suggesting that higher RAM sizes are associated with an increased demand for power supply. The scatter plot visually indicates that notebooks with larger RAM sizes tend to require a higher power supply, as evidenced by the positive relationship between RAM size (in gigabytes) and power consumption (in watts).

- **Comparison Across CPU Processors:**

  Different CPU processors are differentiated by colour in the scatter plot, enabling users to compare the RAM size and power supply patterns across various processors. This facilitates the identification of processor-specific trends.

- **Size Representation of Power Supply Unit:**

  The size of each data point in the scatter plot is proportional to the power supply unit (watts). Larger data points indicate higher power consumption. This visual

representation helps in highlighting outliers with significantly higher power requirements.

- **Outlier Detection**

  Example:

  In the fig(4.23), it is evident that the Lenovo ThinkPad P1 G notebook stands out as an outlier. This is observed by its significantly higher values in both power supply (watts) and RAM size (gigabytes) compared to the other two processors. The outlier status of the Lenovo ThinkPad P1 G indicates that it has notably higher power consumption and RAM capacity compared to the typical trend observed among the selected processors.

Chapter 5: Comparative Analysis

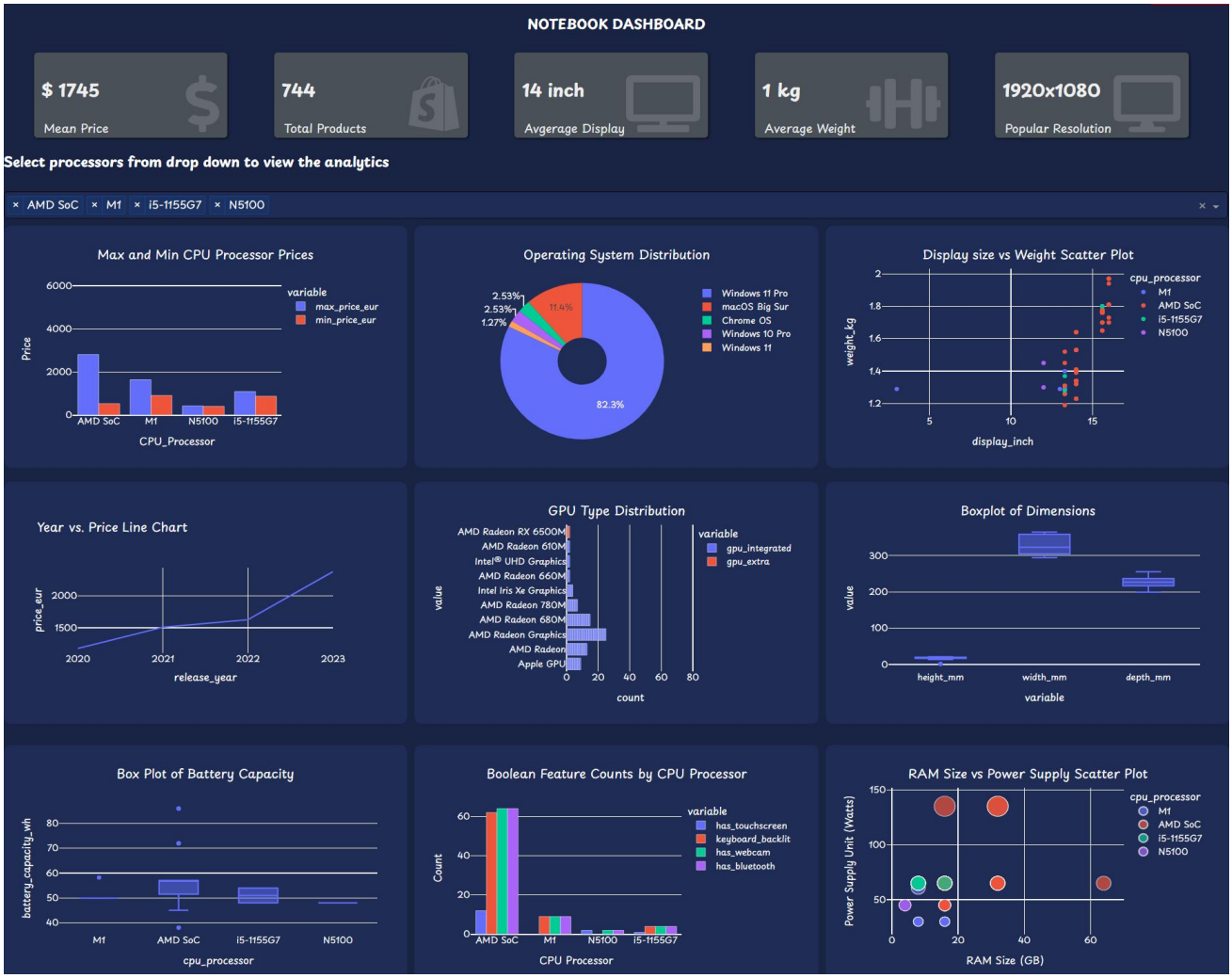

fig5.1 (Comparative Analysis)

When the processors, as illustrated in the figure (5.1), were selected, the following analytics were derived:

- **From the Bar Plot:**

  In the initial bar plot, it's evident that the AMD SoC processor exhibits the highest price and N5100 exhibits lower price among the selected processors, showcasing a broader price range compared to its counterparts.

- **From the Pie Chart:**

  Analyzing the Pie Chart, Windows 11 Pro emerges as the predominant operating system among the selected processors, overshadowing other operating systems in terms of usage.

- **From the Scatter Plot:**

  The scatter plot illustrates a consistent display size and weight trend among all processors, excluding the M1 processor, which exhibits distinct characteristics in these aspects.

- **From the Line Chart:**

  Insights from the line chart indicate a consistent annual increase in the price of notebooks, revealing a notable upward trajectory over time.

- **From the GPU Distribution Chart:**

  Examining the GPU distribution chart, it's apparent that AMD Radeon (integrated graphics) holds a substantial count in our dataset, surpassing other integrated GPUs. Additionally, the AMD Radeon RX 6500M (dedicated GPU) boasts the highest count.

- **From the Box Plot of Dimensions:**

  The box plot reveals a solitary outlier for the height dimension among the selected processors, suggesting a unique outlier in this specific dimension.

- **From the Box Plot of Battery Capacity:**

  The presence of a single line for processors M1 and N5100 implies a lower representation of notebooks containing these processors in our dataset. In contrast, AMD SoC and i5-1155G7 processors exhibit a more detailed box plot, indicative of

their higher prevalence, allowing for a comprehensive visualization of various box plot attributes.

- **Boolean Features Count by CPU Processor:**

  This bar plot distinctly highlights that notebooks with an AMD SoC processor boast the highest count for features like keyboard backlight, touchscreen, webcam, and Bluetooth, surpassing other processors.

- **Ram Size and Power Supply Scatter Plot:**

  The scatter plot clearly indicates that notebooks equipped with AMD SoC processors feature the largest RAM sizes and power supplies compared to other processors.

Chapter 6: Findings and Recommendations

Based on the comparative analysis done in chapter 5 The decision a user makes based on these findings would depend on their specific needs, preferences, and priorities when choosing a notebook.

Here are some potential considerations and decisions a user might make:

- **Processor Selection:**

  AMD SoC:

  Users looking for a diverse range of laptops with varying prices might find the AMD SoC processor appealing. However, they should be prepared for a wider price range and feature set within this category.

  Other Processors:

  Users with specific preferences for operating systems, such as Windows 11 Pro, or those looking for consistent correlations between display size and weight, might prefer notebooks with processors other than AMD SoC.

- **Operating System Preference:**

  Users who value or require Windows 11 Pro may prioritize notebooks with this operating system, as indicated by its dominance in the dataset.

- **GPU Preferences:**

  Users with a preference for AMD graphics solutions might be inclined towards notebooks featuring AMD Radeon GPUs, considering their higher count in the dataset.

- **Feature-Rich Laptops:**

  Users prioritizing features like keyboard backlight, touchscreen, webcam, and Bluetooth may lean towards laptops with AMD SoC processors, as they consistently show higher counts for these features in the dataset.

- **Price Considerations:**

  Users on a budget may want to explore notebooks with AMD SoC processors, considering the wide price variation within this category. However, they should carefully weigh the trade-offs between price and features.

- **Dimensions and Outliers:**

  Users concerned about the dimensions of their notebooks might want to investigate further into the outlier identified in the height dimension, as it suggests a unique characteristic that may or may not align with their preferences.

- **RAM Size and Power Consumption:**

  Users with specific requirements for RAM size and power supply might consider notebooks with AMD SoC processors, as this category exhibits higher RAM sizes and power supply needs.

- **Yearly Price Trends:**

  Users anticipating a potential increase in notebook prices over the years may want to make their purchasing decisions sooner rather than later, considering the upward trajectory observed in the yearly price trends.

Ultimately, the user's decision should align with their individual needs, budget constraints, and specific preferences regarding notebook features, performance, and operating systems.

Chapter 7: Conclusion

In conclusion, the notebook dashboard provides a comprehensive view of various aspects, enabling users to explore and analyze diverse data dimensions related to notebook specifications. With interactive visualizations, users can easily navigate through processor-specific analytics, gaining insights into pricing, dimensions, features, and more. The user-friendly interface enhances the understanding of the dataset, making it a valuable tool for individuals seeking detailed information on notebooks. The dashboard's versatility and depth make it a powerful resource for anyone interested in making informed decisions based on detailed notebook analytics.

References

Plotly. (2023). Dash User Guide. Retrieved from https://dash.plotly.com/

Plotly. (2023). Plotly Python Graphing Library Examples. Retrieved from https://plotly.com/examples/

Plotly. (2023). Plotly Python Open Source Graphing Library. Retrieved from https://plotly.com/python/

Mozilla Developer Network. (n.d.). CSS (Cascading Style Sheets). Retrieved from https://developer.mozilla.org/en-US/docs/Web/CSS

Kaggle. (n.d.). Kaggle Datasets. Retrieved from https://www.kaggle.com/datasets