A
Project Report
on

Big Data Analytics

ANALYZING COMPUTER HARDWARE AND SOFTWARE TRENDS

Submitted by-

Student Name(s)

Saksham Pandla ***Ishika Kochhar***

220382 ***220435***

Under the guidance of

Dr. Yogesh Gupta
Professor

Department of Computer Science and Engineering
SCHOOL OF ENGINEERING AND TECHNOLOGY
BML MUNJAL UNIVERSITY GURGAON-122413, INDIA

Dec, 2024

Acknowledgement

We are deeply grateful to everyone who has helped and mentored us throughout this effort. We are extremely grateful to Prof. Yogesh Gupta, Professor at BML Munjal University, for his helpful assistance, insightful input, and continual support, which contributed significantly to the successful completion of our work.

We also want to thank the faculty members of the Computer Science department of BML Munjal University for providing the knowledge, resources, and support required for this project. Furthermore, we thank our families and friends for their ongoing motivation and understanding, which helped us stay focused and devoted.

The project is a reminder of the joint efforts of everyone involved, and we really appreciate their contributions.

Thanking You
Saksham Pandla (220382)
Ishika Kochhar (220435)

INDEX

LIST OF FIGURES

ABSTRACT

In this project, the trends in computer hardware and software are studied by performing an analysis of data from the Steam Hardware and Software Survey. It uses Python for data extraction and visualization in an effort to expose information about gaming preferences (e.g., which operating system, GPU, CPU are used, and which hardware is used).

The analysis process included cleaning and organizing the dataset to address missing values and maintain consistency. The main trends were graphically illustrated using the libraries Matplotlib and Seaborn that illustrate the evolution of user preferences. The research finds significant trends in hardware acceptance, the growing popularity of some operating systems, and changes in technology trends in the gaming industry.

Through sophisticated application of data preprocessing and visualization methods, this work provides an overall perception of the hardware and software ecosystem, with important implications for gamers, developers, and hardware manufacturers.

MOTIVATION

The explosive growth of the gaming market has led to amazing developments in both hardware and software. Since the Steam Hardware and Software Survey data offer meaningful information on what users want, this work is designed to use that data to identify technology adoption patterns that can inform both producers and consumers on how to make good decisions. The motivations for carrying out this research are:

(1) Understanding Gaming Preferences: This work is based on the analysis of data from the Steam Hardware and Software Survey for the purpose of understanding how gamers' decisions about hardware and software are changing. With visual analysis of those evolving trends, we are able to understand better the gaming tech world and its role on future of gaming.

(2) Driving Innovation and Development: Based on trends of hardware and software adoption, this analysis provides, critical information about what should be optimized by the hardware manufacturers and what should be optimized by the software developers for gaining better compatibility and performance. Additionally, it provides gamers with a comprehensive view of the latest technological advancements, helping them make informed decisions about their gaming setups.

(3) Practical Application of Data Analytics: The project provides a hands-on approach to learn how to apply data analytics methods, such as cleaning, preprocessing, and visualization of data. It demonstrates the ability of turning big datasets into usable information, which may have an impact on industry trends and decision process.

In this work, the results not only shed light on emerging gaming trends but also demonstrate the real-world application of data analysis. It highlights how data-driven insights can shape the gaming industry and drive technological progress, reinforcing the significance of translating theoretical knowledge into practical, impactful outcomes.

1. INTRODUCTION

The gaming market has experienced exponential growth in the past twenty years and is fueled, in part, by ITS and in part by the growing awareness of interactive entertainment. Based on a community of millions of players worldwide, the technical decisions made of hardware and software by this community cannot be random, but at the heart of the trend of technical advances. The tastes are not restricted to the design and function of game devices and can provide us with some information of the trends of the adoption of new technologies.

Being one of the leading gaming platforms worldwide, Steam continuously collects data through a Hardware and Software Survey, which is rich with information regarding the hardware, operating system, and software arrangements of its users. This data set provides an unprecedented ability to explore the changing gaming world as it speaks to individual decisions and the macro-directional industry trends. The Steam Hardware and Software Survey dataset contains a wealth of information that researchers have available to study patterns of GPU and CPU usage, storage usage, OS usage, and other major components.

This work uses the Steam Hardware and Software Survey dataset to study and plot technology trends in games. By analyzing this data, this work aims to close the gaps of what is changing in the game world by highlighting trends that can be used to understand the transformation process of the gamers throughout the technological development. For instance, rates of change in GPU utilization, operating system selection, etc. can help us to understand the speed at which athletes, or gamers, delete and hide new technologies from an old technology.

Contrary to trend analysis, motivation for this work stems from its possible application of data analytic approaches. It is also a platform—in the sense of presenting an example of the application of the available tools and pipelines (data cleaning, processing and visualisation), making use of ensemble data sets, and turning them into usable, actionable information. Python-based toolboxes (e.g., Pandas, Seaborn and Matplotlib) are employed for pattern recognition, outlier identification, and for producing easily interpretable visual information to help reduce the sheer volume of data.

The findings of the present study are of concern to many stakeholders. For hardware developers, it will be possible to best understand user preferences for product design and feature sets. However, information on the typical configurations can be used to increase the level of service and experience of the users for software engineers. Gamers can also be equipped with the means to use a conceptual model by understanding the role of their own individual choices within larger trends, which, again, may help them to do better than before when upgrading their own setups.

At a time when methods based on information derived from data are altering business fields, the current study provides an example of how such real-world data sets can be used to extract meaningful information. In this way, it highlights the role of analytics in the dismantling of the consumer environment around which analytics are being leveraged to fuel innovation within the gaming sector.

2. **PROBLEM STATEMENT**

2.1 Motivation for the Problem

There are very dynamic markets in the technology world, like gaming, where hardware and software is always increasing their capabilities. It's important for manufacturers to innovate, developers to enhance their software, and gamers to decide whether to upgrade intelligently, that user preferences and technological advancements be understood. On the other hand, the raw data from the Steam survey is currently underexplored since it is cumbersome with significant size. The aim of this project is to prove the utility of data analysis in a practical sense, transposing this data rich resource into practical, actionable wisdom, with implications for all parties in the gaming business.

2.2 Specific Problems Addressed by the Project

This project addresses the following problems:

1. **Identifying Key Trends:** Monitoring hardware preferences (with GPUs, CPUs and storage solutions) is a challenging task because of the amount and heterogeneity of data

available. Also, more structured analysis to uncover patterns and knowledge in a software application adoption rate, such as operating systems or gaming APIs, is necessary for informing future gaming technology advances.

2. **Bridging the Knowledge Gap:** The raw Steam survey data lacks structure and accessibility, making it difficult for users to extract meaningful information. Without effective tools for data visualization and organization, the valuable insights hidden within the dataset remain underutilized.

3. **Supporting Decision-Making:** It is a difficult task for producers and developers to personalize their products depending on the latest trends, and as a gamer, it is a rare thing to find a suitable hardware or software combination. As this data is analyzed, the project offers practical realization of the insights, which allows informed action from all industry stakeholders.

2.3 How the Project Solves These Problems

This project solves the problem in the following ways:

1. **Comprehensive Data Analysis:** In the project, the data from the Steam Hardware and Software Survey is preprocessed and cleaned from bias, missing values, and unnecessary information, making it accurate.

2. **Trend Identification:** Visualizations and trend analysis demonstrate the changes in user preferences over time as it relates to the hardware and software elements thereof.

3. **Actionable Insights:** Manufacturers can benefit from the information by organizing the results and presenting them in a clear matter. In order to optimize their designs, developers can gain from better compatibility, and the gamers can make better decisions.

3. **LITERATURE REVIEW**

Machine learning (ML) and data analysis in the context of gaming hardware trends have been gaining significant attention in recent years. Many studies have explored the use of various analytical tools and techniques to understand and predict user preferences and system requirements based on trends observed through platforms like Steam.

3.1 Existing State-of-the-Art

- In the study by Tom's Hardware (2024), the author discusses the ongoing trends in gaming hardware, particularly the significant increase in Steam users upgrading their systems to 32GB of RAM. This shift is attributed to the growing demand for gaming performance, as modern titles require more memory to handle complex tasks such as large open-world environments and detailed textures. These hardware trends are critical in understanding how user preferences are evolving in response to game complexity and performance demands

  Reference: A Study on Ram comparison and Linux

- The Steam Hardware Survey (2024) reflects the trends toward more advanced hardware setups. The survey's findings reveal an increasing adoption of 32GB RAM configurations, as well as more powerful CPUs and GPUs, as gamers seek systems capable of handling cutting-edge games and applications. This research aligns with the growing demand for systems that can support both gaming and content creation tasks, highlighting the trend toward versatility in gaming hardware

  Reference: A study on Steam Hardware and software survey

- CheeseTalks (2024) also offers an in-depth exploration of Steam's hardware trends, focusing on the adoption of high-end CPUs and GPUs to meet the increasing demands of modern gaming. Their analysis reveals that as games become more demanding, users are prioritizing processors and graphics cards that can handle complex workloads. This trend is crucial for understanding how gaming systems are evolving in line with the requirements of both high-definition and virtual reality gaming

  Reference: Steam Survey Insights: Gaming Hardware Trends and System Evolution

- BetaBreakers (2023) delves deeper into the latest hardware survey results, showing a shift toward more powerful GPUs among Steam users. The demand for high-end graphics cards is growing as more gamers play graphically intensive games that require advanced GPU capabilities. This trend highlights the increasing importance of graphics in modern gaming setups, and the authors suggest that this pattern will likely continue with the rise of virtual reality and ray tracing technologies

Reference: Steam Hardware survey of May 2023

3.2 Patents and Prior Art in Gaming Hardware and Software Optimization

Several patents exist in the field of gaming hardware and software optimization, particularly focusing on enhancing the performance of gaming systems like Steam. These patents explore innovations in graphics processing, CPU optimizations, memory management, and rendering techniques, all crucial in addressing the increasing demands of modern gaming systems. Below is a table summarizing key patents that are relevant to the Steam hardware survey analysis:

Table 1. Data of Analysis of *Existing Patents*

S. No.	Existing state of art	Drawbacks in existing state of art	Overcome
1	**UUS11816954 –** Gaming System and Method for Graphics Processing	Limited scalability for high-performance gaming systems when handling real-time rendering tasks.	Provides solutions for improving real-time graphical rendering, optimizing the use of GPUs in high-performance gaming.
2	**US10500498B2 –** Memory Management for Gaming Systems	Static memory allocation that may not efficiently adapt to rapidly changing game requirements.	Implements dynamic memory allocation to better manage resource usage in memory-heavy games.
3	**US20010016517A1 –** Game Processing Using Distributed Computing Systems	High latency and processing delays in cloud-based gaming environments.	Distributes game processing across multiple machines to reduce delays and improve overall performance.

1. US11816954 – Gaming System and Method for Graphics Processing

This patent introduces an innovative method to optimize graphics processing in gaming systems, specifically addressing real-time rendering tasks. It enhances the GPU performance by improving the rendering pipeline, allowing for more fluid and high-quality graphics. This technology is particularly relevant to improving gaming experiences for Steam users who frequently push for higher graphical fidelity and smooth gameplay in resource-intensive games.

2. US10500498B2 – Memory Management for Gaming Systems

This patent focuses on efficient memory management within gaming systems, which is essential for gaming platforms such as Steam. The technology proposes dynamic memory allocation, which ensures that resources are distributed optimally across multiple tasks. As gamers upgrade their systems for higher RAM capacities, this patent becomes especially relevant by allowing better resource management for memory-intensive games.

3. US20010016517A1 – Game Processing Using Distributed Computing Systems

This patent suggests a distributed computing model for gaming systems, where tasks such as game rendering and processing are spread across multiple machines. This patent can significantly improve the performance of cloud gaming platforms, a trend becoming increasingly popular in the gaming community, particularly on platforms like Steam. It tackles the issue of latency and performance bottlenecks commonly encountered in distributed gaming environments.

3.3 Additional Notes on Prior Art

The ongoing innovations in gaming hardware and software, driven by trends in user demands for high-performance gaming experiences, continue to fuel further research in the field. Many of these innovations, as demonstrated by the patents reviewed above, focus on enhancing the capability of GPUs and memory systems, as well as distributing workloads more effectively across multiple computing nodes. These advancements, combined with platforms such as Steam, are enabling a new generation of gaming experiences that prioritize performance, stability, and quality.

4. METHODOLOGY

4.1. Explaining Methodology

The project follows a systematic approach for handling and processing large-scale datasets, particularly for multi-class classification tasks. Here's a breakdown of the methodology:

4.1 Dataset Overview and Challenges

The dataset under analysis was loaded using Python from a **Parquet file format**, a choice made for its efficiency in handling large-scale datasets. It consists of **334,677 rows** and **5 columns**, covering categories such as **index**, **perc**, **category**, **date**, and **platform**.

- **Data Format and Schema:**
    - The Parquet format enables efficient storage and access to data, particularly suitable for large datasets. However, schema inference during loading using libraries like Pandas or PyArrow can sometimes slow down the process.
    - Defining a predefined schema during data loading may optimize performance and reduce computational overhead.
- **Index Column:**
    - The **index** column acts as a descriptor, containing values such as "Windows 7" and specific types like "Intel HD Graphics 530".
    - It includes **125 missing values**, which could represent an absence of classification or unrecorded data.
    - Handling these missing values is critical, as they could influence the interpretation of related columns like **category**. Imputation strategies or exclusion during analysis may be necessary.
- **Categorical Features:**
    - The **category** and **platform** column includes key groupings such as **Processor Vendor** and **PC** respectively, and other classifications, making it crucial for trend analysis in the hardware and software sectors. Ensuring consistency in this column is important for meaningful insights.

- **Numerical Data:**
    - The **perc** column represents numerical data such as the market share or prevalence percentage, which is fundamental for analyzing trends over time. Proper scaling or normalization may be required depending on the analytical approach.
- **Date Handling:**
    - The **date** column is in an ISO-8601 string format, requiring conversion to a proper **datetime** type for efficient time-based operations and analysis.

Classification Strategy:

The goal of this project is to analyze trends in the **computer hardware and software industry** over previous years. The strategy involves:

1. **Categorical Analysis:**
    - Use the **category** column to group data (e.g., **Processor Vendor**) and analyze trends across vendors such as AMD, Intel, and others.
2. **Temporal Analysis:**
    - Leverage the **date** column to track changes in the prevalence (**perc**) of specific indices (e.g., "Authentic AMD") over time. This will help identify patterns in the adoption of hardware and software technologies.
3. **Trend Insights:**
    - By combining the **perc** values with categories and indices, the model aims to highlight key trends, such as shifts in processor dominance, platform usage, or network preferences.

4.2 Data Collection

- Source: The dataset utilized in this project is derived from the Steam Hardware Survey, which is conducted by Valve Corporation to gather information about the hardware configurations of users who play games on the Steam platform. This survey collects data on various hardware components, including CPU, GPU, RAM, and operating systems.

- Data Format: The data is originally available in Parquet format, which is a columnar storage file format optimized for use with big data processing frameworks. Parquet files are efficient for both storage and retrieval, making them suitable for handling large datasets.
- Data Retrieval: The data was retrieved using a web scraping approach, where archived snapshots of the survey results were accessed through the Internet Archive's Wayback Machine. The scraping process involved:
    - Building Metadata: A metadata file was created to track the URLs of archived snapshots for each month from 2004 to 2025. This was done using the **build_metadata()** function, which checks for existing data and retrieves snapshot links from the **Internet Archive's Wayback Machine**.
    - Downloading Content: The HTML content of the archived pages was downloaded using the **download_web_content()** function. This function saved the webpage content locally for faster access and analysis.
- Data Parsing: The downloaded HTML content was parsed using BeautifulSoup. Two parsing functions, old_parser() **and** modern_parser()**, were implemented to extract relevant data based on the structure of the HTML pages. The parsed data was then saved in a JSON format for further analysis.**
- Data Integrity: Throughout the data collection process, checks were implemented to ensure the integrity and completeness of the data. This included verifying the availability of snapshots for each month and handling any missing or incomplete data entries.

4.3 Data Preparation

- **Loading Data**: The JSON files containing the parsed data were loaded into a pandas DataFrame for analysis.
- **Data Cleaning:** The dataset was cleaned by handling missing values and ensuring that data types were appropriate for analysis. This included converting date strings to datetime objects and ensuring numerical values were in the correct format.

4.4 Data Exploration and Analysis

- Descriptive Statistics: Summary statistics were generated to understand the distribution of hardware specifications across different survey periods.
- Data Visualization: Various visualizations were created using libraries like Matplotlib and Seaborn to illustrate trends in hardware usage over time. This included plotting the prevalence of different hardware configurations and analyzing changes in user preferences.

4.5. Statistical Analysis

- Statistical tests were conducted to identify significant differences in hardware usage across different time periods. Correlations between hardware specifications and user demographics were also explored where applicable.

4.6 Interpretation of Results

- The results of the analysis were interpreted to draw conclusions about trends in hardware usage among Steam users. Key findings were highlighted, such as the rise in specific hardware types or configurations over time.

4.7 Components

1. Hardware (Local machine for the first level of search)

- AMD Ryzen 9 6900HS
- Nvidia RTX 3050
- 16 GB of RAM

. 2. Software:

- Python 3.10.
- Jupyter NoteBook.
- Aaconda
- Libraries: Pandas, Matplotlib, Scikit, Beautifulsoup4, certifi

4.7. Status of the Project

The project has been **implemented and tested**:

- Built and executed by the team at **BML Munjal University** during the **BDA Course Project**.
- First successful build: **November 2024**.
- Evidence:
  - Dataset used: Scraped using BeautifulSoup4 from **Internet Archive's Wayback Machine**.
  - All the results and transformations documented in notebooks.

5. RESULTS AND DISCUSSION

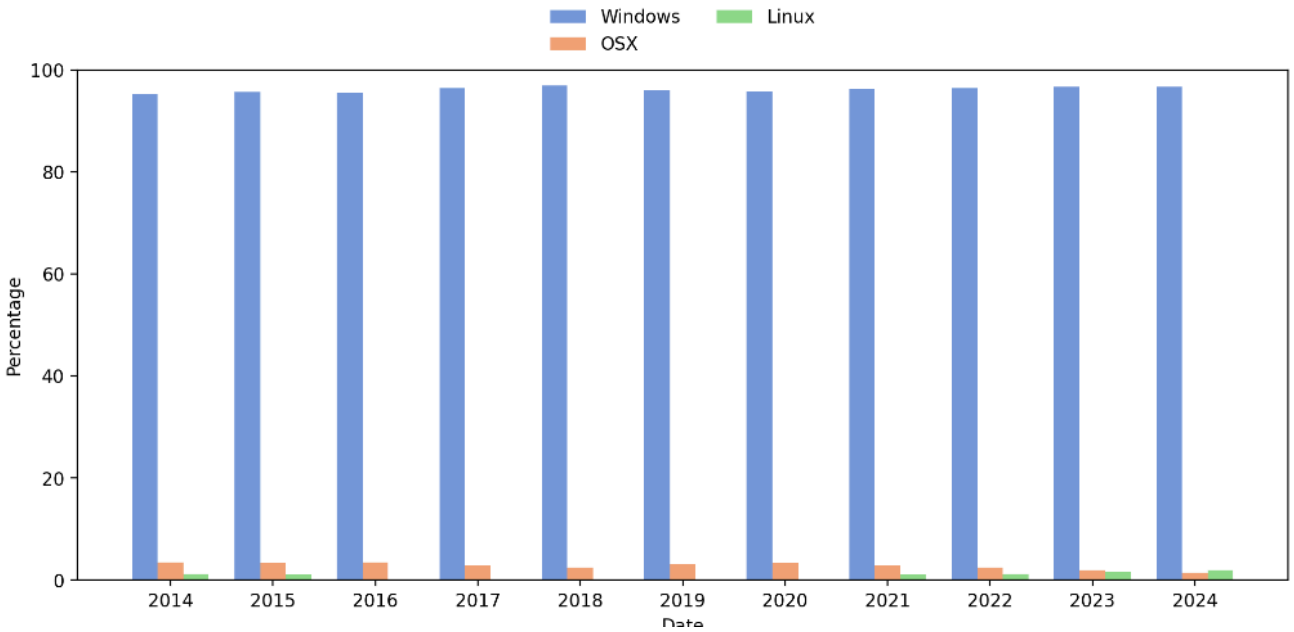

Fig 1: *Operating System Distribution on Steam (2014–2024)*

The bar chart depicts the distribution of operating systems (OS) of Steam users among years from 2014 to 2024. The findings also clearly show that Windows has continued to show a clear and absolute command of the user base of Steam for the past 10 years. Windows almost exclusively dominates the platform market with a tiny fluctuation from one year to another. On the other hand, both macOS (OSX) and Linux have kept very low adoption percentages over the same period. Though this is sometimes a matter of variation, Linux is the OS with least steam on Steam and OSX has a minimal share of the market. This ongoing trend implies that, although some other OSes such as Linux or OSX have gradually spread within other domains, there has been little spread in the gaming community on Steam. These results underscore the continued relevance of Windows as the major OS used for gaming, due to its position in the gaming market and overall computing market

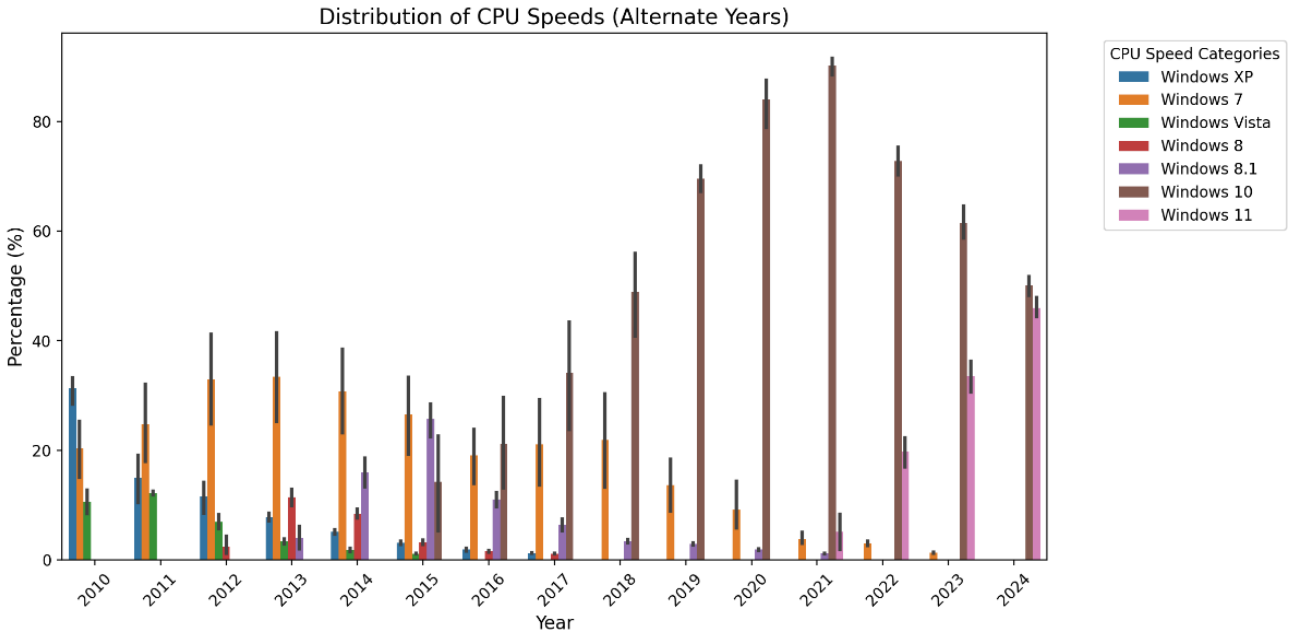

Fig 2: Distribution of CPU Speeds (2010-2024)

The chart shows the evolution of the use of operating systems over time. From 2010 to 2015, Windows XP and Windows 7 were the two most popular OSs. Although Windows XP usage decreased slowly across this timespan, Windows 7 usage continued to increase over the years, with Windows 7 usage peaking in 2013 and 2014. [M1] Other versions, e.g., as Windows Vista and 7, found limited reception and have been long out of circulation by 2015.

From 2016, the Windows 10 operating system became the market leader, with rapid growth in the number of users. By 2019, Windows 10 usage exceeded 80% and remained the dominant OS through 2020 and 2021. But, the scene changed in 2022, when Windows 11 then appeared and started to be embraced. Between 2023 and 2024, Windows 11 experienced a serious spurt in the number of installations capturing more than 40% of the distribution, which suggests a switch from Windows 10 to Windows 11. This trend reveals a clear pattern of replacement, in which one new version successively replaces older systems.
In general, the chart illustrates the changes in OS usage patterns over time, corresponding to user expectations and technological developments. Past versions (e.g., Windows XP, Windows Vista, Windows 8) have been retired, while today's dominant versions (e.g., Windows 10, Windows 11) take the lead.

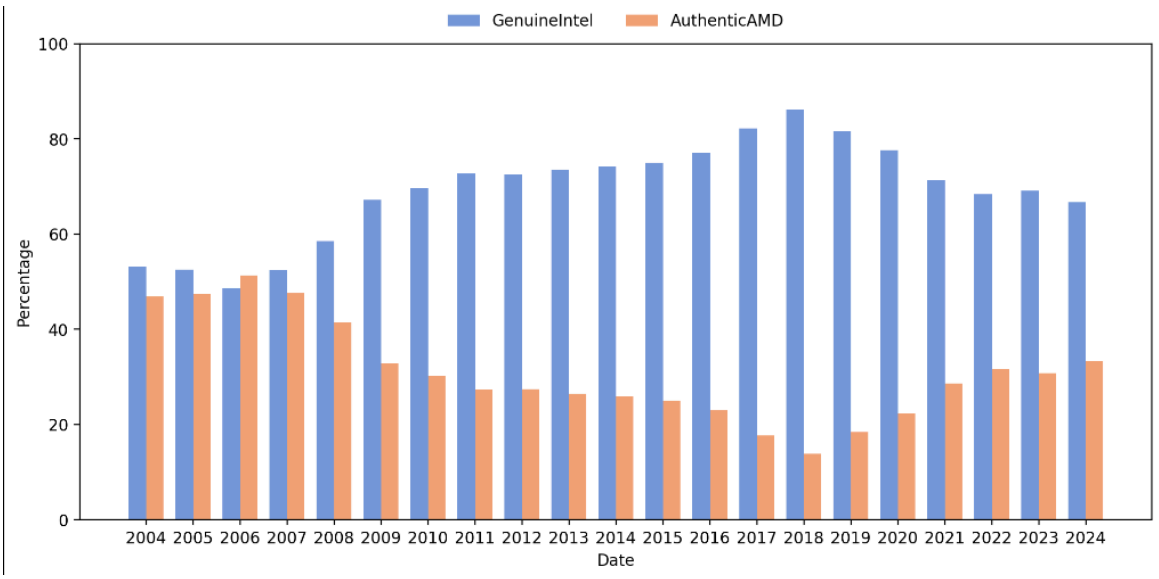

Fig 3: *CPU Brand Distribution (2004–2024): GenuineIntel vs. AuthenticAMD*

The chart depicts the trend of CPU brand distribution (GenuineIntel and AuthenticAMD) from 2004 to 2024 according to their percentage of use. In all years, GenuineIntel (blue bars) consistently outperforms the market, with a greater proportion of usage than AuthenticAMD (orange bars).

During the period between the years 2004 and 2007, the two brands had very similar market shares, GenuineIntel fluctuated from 50% to 55% and AuthenticAMD from 45% to 50%. However, from 2008 onwards, a noticeable gap emerges as GenuineIntel's market share increases, peaking between 2010 and 2018 at around 70% to 85%, while AuthenticAMD steadily declines to a low of around 15% during this period. The lowest bottom for AuthenticAMD seems to fall between 2017 and 2018 when Share goes below 20%.
From 2019, AuthenticAMD begins to recover slightly, and its market share increases progressively until about 30% to 35% in 2024. While, at the same time, GenuineIntel's market share drops slightly but still leads above 65% in 2024.

Taken together, the chart shows the enduring position of GenuineIntel at the top of the CPU market over the period of the two decades, with AuthenticAMD showing some periods of decline that have been met with a small upswing in recent years. All this reflects changes in market competition and in technology and in consumer behaviour.

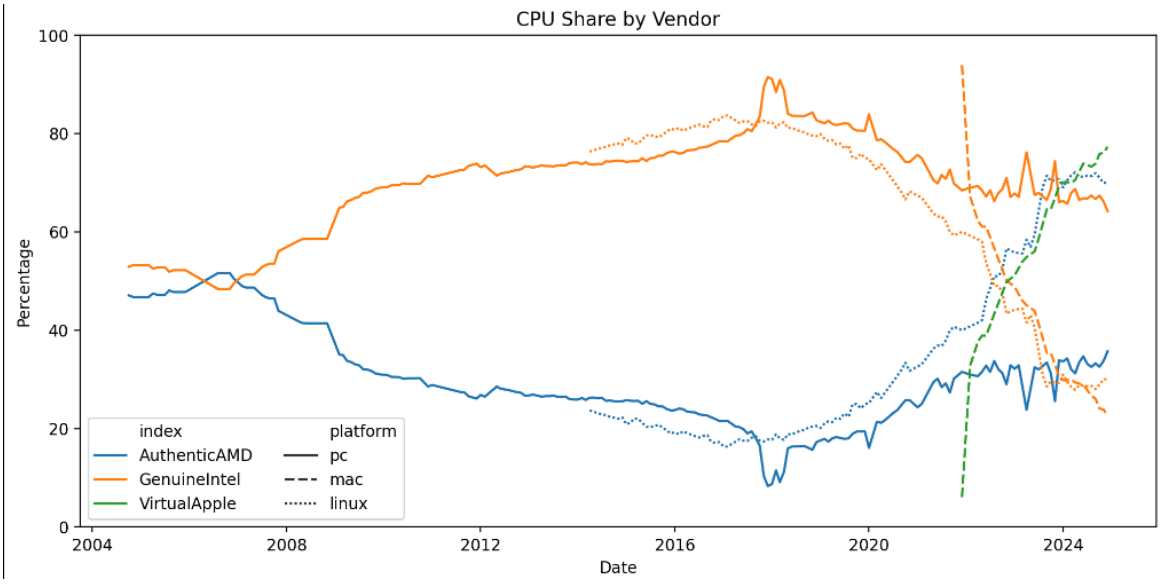

Fig 4: *CPU share by vendors*

The graph illustrates the CPU market share trends by vendor (AuthenticAMD, GenuineIntel, and VirtualApple) across different platforms (PC, Mac, and Linux) from 2004 to 2024. Intel initially dominated the market, particularly on PC platforms, maintaining a stronghold until 2016.

However, its share began to decline significantly, reflecting increased competition from AMD and Apple. AMD's market share, after declining post-2007, started to recover steadily around 2016, likely due to the success of its Ryzen series, leading to a resurgence in the PC market.

Apple's introduction of in-house processors (VirtualApple) for Mac devices in 2020 marked a sharp rise in its CPU share, highlighting the impact of its transition away from Intel.

On Linux platforms, the graph shows a more varied distribution of market share among vendors, reflecting the diverse hardware preferences in the open-source community. This analysis reveals the evolving landscape of the CPU market, driven by competition, innovation, and strategic shifts in platform-specific adoption.

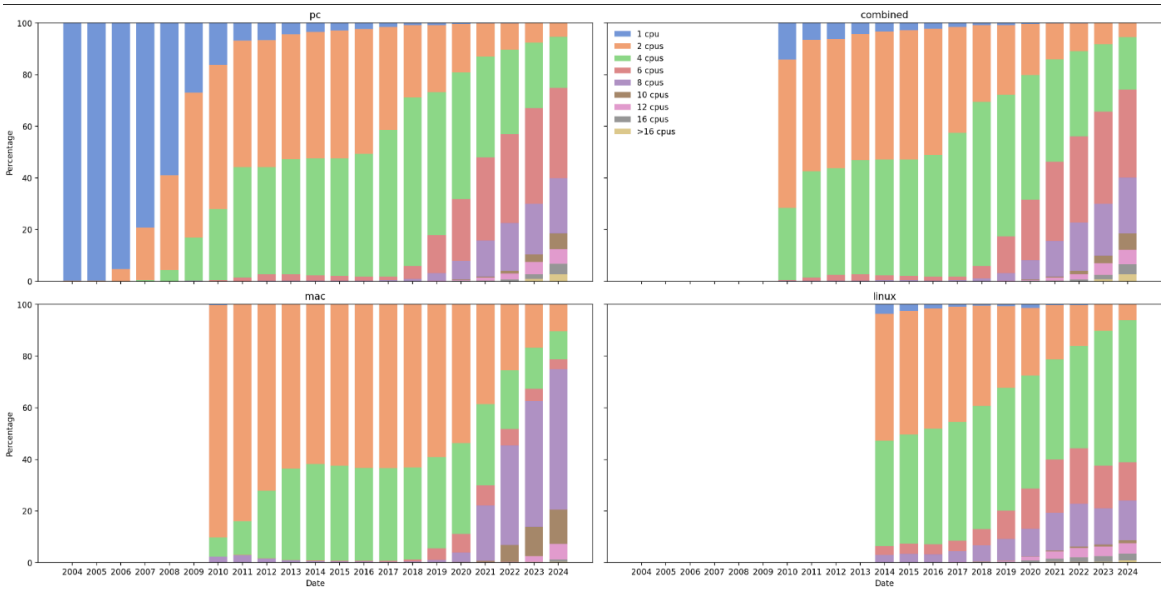

Fig 5: *Distribution of CPU Core Counts Across Platforms (2004–2024)*

The chart provides a comprehensive overview of the evolution of CPU core distributions across various platforms (PC, Mac, Linux, and Combined) from 2004 to 2024. It highlights a clear transition from single-core processors to multi-core architectures over the years, reflecting advancements in technology and the growing demand for higher computational efficiency.

Initially, single-core CPUs dominated the market until around 2008 across all platforms. However, as the need for multitasking and improved processing power grew, dual-core CPUs started gaining traction. By 2010, dual-core processors had become the leading configuration across most platforms. Subsequently, four-core CPUs saw a sharp rise in adoption, especially after 2010, driven by the increasing demands of modern applications and gaming.

From 2015 onward, higher core counts, including 6-core, 8-core, and 10-core CPUs, began to gain significant market share, particularly in PC and Linux platforms. This shift was likely fueled by the need for more powerful systems for gaming, productivity, and professional workloads. By 2020, configurations with more than 8 cores became more prevalent, reflecting the rise of high-performance computing and server-grade hardware.

Platform-specific trends indicate that PCs have shown a steady progression in adopting multi-core CPUs, while Macs focused predominantly on 2-core and 4-core processors until the mid-2010s, transitioning to higher core counts only after 2020 as Apple introduced advanced hardware solutions. Linux platforms, commonly used in servers and professional environments, displayed the most diverse core distribution, with an earlier adoption of CPUs with higher core counts. The combined data aligns with the overall industry trend of moving toward multi-core architectures.

In conclusion, the chart underscores the CPU industry's rapid evolution toward multi-core architectures over the past two decades. This shift highlights the growing demand for enhanced multitasking capabilities, efficiency, and processing power across various platforms and computing environments, driven by technological advancements and changing user requirements.

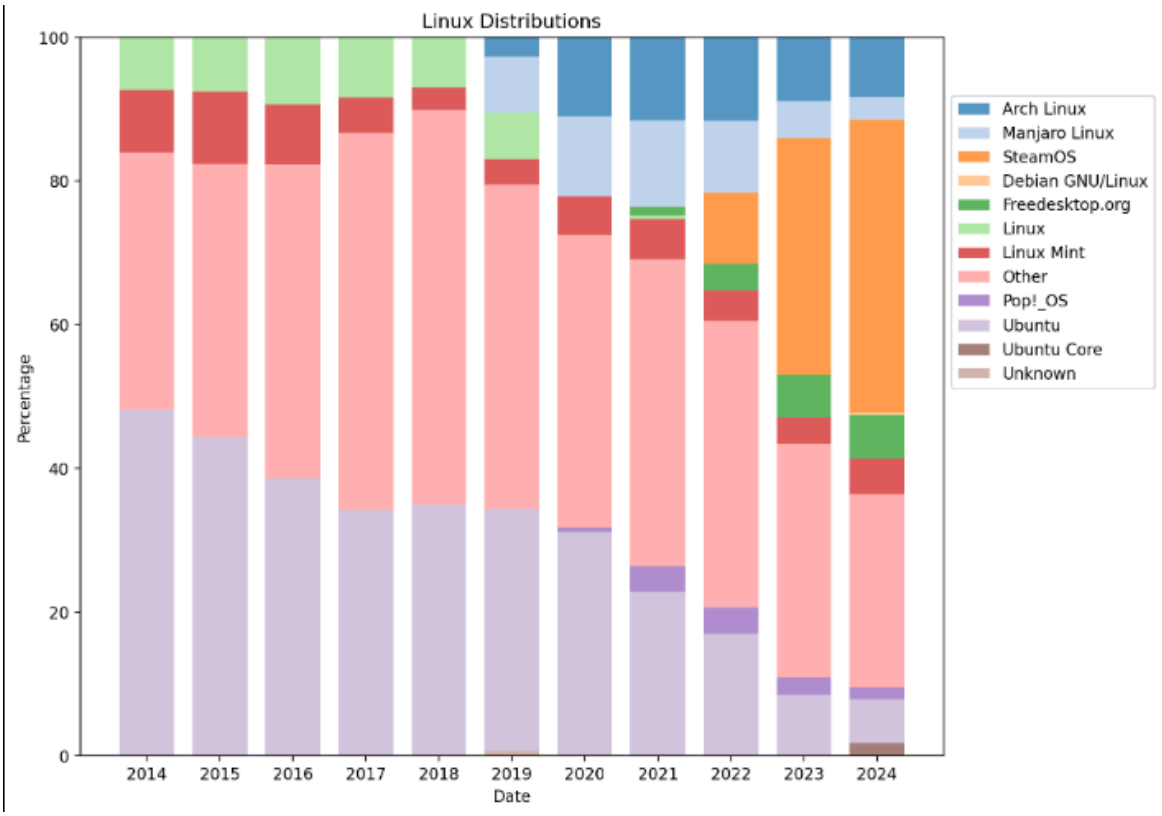

Fig 6: *Trends in Linux Distribution Usage*

This chart depicts the relative popularity of various Linux distributions over time, from 2014 to 2024, based on their percentage usage. The chart includes major distributions like Ubuntu, Debian GNU/Linux, Arch Linux, and others, along with a category for unknown distributions.

From the visualization, Ubuntu consistently dominated the market share until around 2020, after which its prominence declined slightly. Starting in 2020, there is noticeable growth in the use of distributions like Arch Linux, Manjaro Linux, and Pop!_OS, reflecting a diversification in user preferences. The "Unknown" category also saw a significant reduction over time, indicating better reporting or categorization of distributions in recent years.

Another key trend is the gradual decline of Debian GNU/Linux's share, suggesting a shift in preference from foundational distributions to newer, user-friendly derivatives like Pop!_OS and Linux Mint. SteamOS emerges significantly post-2021, which could be attributed to the rise in gaming-related usage of Linux platforms.

In summary, the chart highlights the increasing fragmentation and diversity in the Linux ecosystem, with users exploring newer distributions while traditional ones like Ubuntu and Debian maintain a notable but reduced presence.

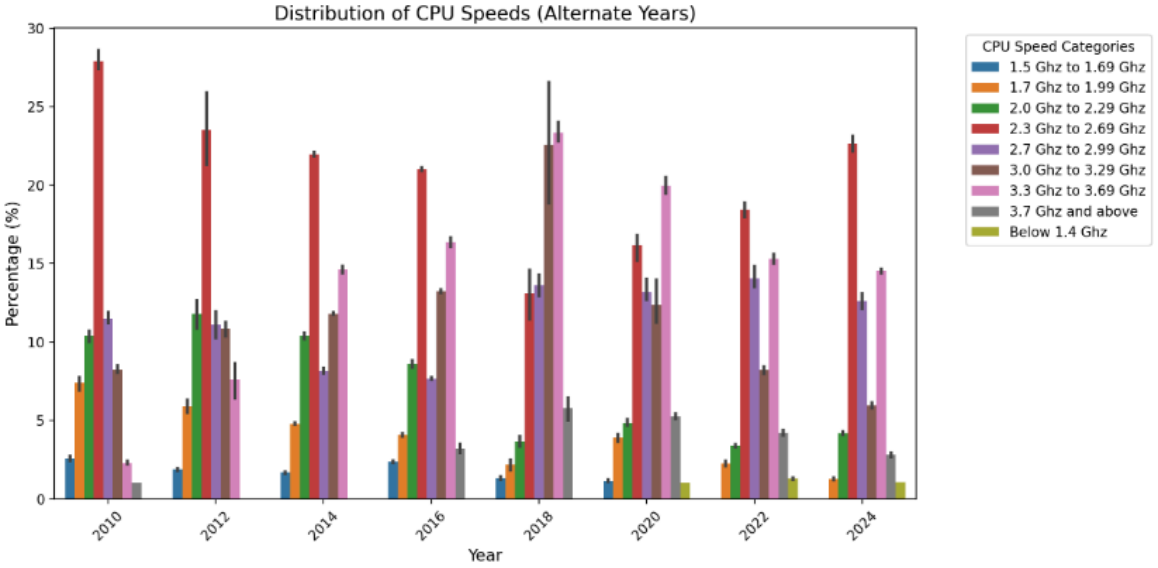

Fig 7: *Trends in CPU Speed Distribution Across Alternate Years*

The bar chart titled "Distribution of CPU Speeds (Alternate Years)" illustrates the percentage distribution of CPUs across various speed categories from 2010 to 2024, focusing on alternate years. It provides insights into the evolution of CPU speed preferences and technological advancements over time.

From the chart, a clear trend emerges where mid-range CPU speeds (e.g., 2.3 GHz to 2.99 GHz) have consistently held a significant share, particularly from 2014 onwards, reflecting their popularity for general-purpose computing. In earlier years, there was a more balanced distribution across lower-speed CPUs, such as those in the 1.7 GHz to 1.99 GHz range, indicating earlier hardware limitations or a broader range of older machines in use.

As technology advanced, higher-speed CPUs, especially those in the 3.0 GHz to 3.69 GHz and above categories, gained prominence, particularly after 2018. This shift aligns with the increasing demand for more powerful processors due to advancements in gaming, professional applications, and data-intensive workloads.

Conversely, the share of CPUs below 1.4 GHz and those in the 1.5 GHz to 1.69 GHz range has steadily declined, almost vanishing by 2024, signaling the phasing out of older or less capable processors. This reflects a general industry trend toward higher baseline performance in consumer and enterprise markets.

In summary, the chart highlights a gradual shift toward higher-speed CPUs, driven by advancements in technology and changing performance requirements, while lower-speed CPUs have become obsolete over time.

6. CONCLUSION

This work has yielded a deep understanding of the Steam Hardwares and Software Survey data from the point of view of gamers, on which the distribution of gaming hardware, software, and platform preference have undergone profound changes. Through processing and visualizing the survey data, we have found the trends which enable the manufacturers and developers to know the current situation of the gaming market. These observations can be used to gain useful insight into adoption of GPUs, CPUs, storage and such OSs, enabling stakeholders to take an informed decision about product development, and software optimization.

By applying appropriate data cleaning, preprocessing and visualisation techniques, the project has revealed the potential of data analytics for the interpretation of large-scale datasets. It has illustrated the potential of seemingly complex, unstructured data to be interpreted into useful knowledge for real-life decisions. Not only does this project help the gaming industry in this specific way, it also draws attention to the wider application of data analytics in the search for patterns and trends across any other rapidly advancing industry.

The results of the project provide direct utility to multiple stakeholders, including the hardware manufacturers, game developers, and gamers. Manufacturers can utilize this information to better design products to meet gamers' taste and requirements, and program developers can achieve better software compatibility and performance. Gamers, in turn, are able to take advantage of the information being offered and make informed decisions about the hardware/software they are going to be using, and as a consequence improve their experience.

In addition, this work has also demonstrated the applicability of data-driven decision making within the gaming industry, demonstrating how data-driven analyses can be used to identify and respond to changes in the market through predictive adaptations. Based on a study of historical and current trends, it offers a pathway to the development of gaming technology in the future, with impact on product design and user experience.

7. FUTURE WORK

Although this work has already offered some insights into the trends in the gaming industry now, there are still many topics where further research could capitalize on the results obtained and contribute to the overall impact of this project. A promising area for future research is machine learning model integration. Using algorithms such as regression or clustering, the project could be used for future trend prediction of hardware and software technology adoption. Such predictive models would provide manufacturers and developers with a more reactive method for what consumer tastes they may see changing and they can use this in order to change the design of the product or of the software.

Additionally, the project could benefit from real-time data analysis. When continued collection and analysis of new survey data stream to the project as it becomes available, the project could monitor changing patterns more accurately and quickly. This could aid stakeholders to stay ahead of continuously evolving technologies and gamesplayed preferences, thus speeding up their decision-making and product iterations. Incorporating real-time information would also enable detection of developing gaming trends or dramatic changes in user behavior, thus maintaining the relevance of the insights and their timeliness.

The data expansion is also an important possibility for the next works. The Steam as well Hardware and Software survey is useful information however it only represents part of the global gaming market. Integrating data from other platforms (for example, Epic Games, PlayStation, Xbox, or from mobile gaming survey) the work has the potential to provide a more holistic perspective of the gaming industry. A larger and more diverse dataset would provide a deeper understanding of global gaming trends, enabling more accurate analyses of regional preferences and gaming behaviors.

In addition, the creation of interactive dashboards or web-based application for data visualization may increase the accessibility and usability of the project. These tools would allow stakeholders to explore the data dynamically, giving them the ability to drill down into specific trends and compare variables in real time. Interactive visualizations would make it

easier for users to gain insights from the data, empowering them to make more informed decisions based on their unique needs and interests.

Most recently, future research could investigate cross-industry trend analysis by comparing gaming hardware and software choices with choices in other tech industries, e.g., mobile computing or cloud gaming. Such a wider analysis would lead to a more general perspective of the technological scene and could also allow the reader to understand how gaming technologies work and how they affect other industries. Through an enlarged field of view, the stakeholders may be better able to appreciate the role of gaming within the broader technology landscape and consider more strategic, sustainable choices.

These avenues for future directions would not only extend the generality of information produced by this project, but will also help to make the distributed information continue to be of value and interest in a dynamic gaming world.

8. REFERENCES

1. Steam. (2024). Steam Hardware & Software Survey
https://store.steampowered.com/hwsurvey

2. Analytics Insights, "How to Preprocess and Analyze Data with Python," YouTube, 2023. Available: https://youtu.be/_C8kWso4ne4?si=j5pCZI4q1OfNTCqZ.

3.] Reddit Databricks, “Effective ways to preview large datasets,” Online discussion, 2020. Available:
https://www.reddit.com/r/databricks/comments/1avta5s/whats-the-fastestmost-effective-way-to_look_at_a/

4. A. Cook, "Understanding Data Types," Kaggle Notebook, available at: https://www.kaggle.com/code/alexisbcook/data-types.

5. Tom’s Hardware - Steam Gamers Upgrade to 32GB RAM
https://www.tomshardware.com/video-games/pc-gaming/steam-gamers-hurry-to-upgrade-to-32gb-ram-and-linux-breaks-above-2-user-share-milestone-in-latest-survey

6. Implementing Web Scraping in Python with BeautifulSoup
https://www.geeksforgeeks.org/implementing-web-scraping-python-beautiful-soup/

7.Beta Breakers - Steam Hardware Survey Statistics
https://www.betabreakers.com/stats-sources-steam-hardware-survey/

8. Cheese Talks - Steam Hardware Survey Insights
https://cheesetalks.net/steamsurvey.php