7COM1079-0901-2024 - Team Research and Development Project

Final report title: (Analysis of AMD Processor Data*)*

Group ID: A 168

Dataset number: DS059

Prepared by: Hemil Parmar ID- 23078371,

Kartik Mistry ID- 23073391,

Andappa B Chandoor ID- 23060670,

Poorna Sujampathi ID- 23102780,

Abhijit Anilkumar and ID- 23070329

University of Hertfordshire

Hatfield, 2024

Table of Contents

[1. Introduction 4](#_Toc187059863)

[1.1 Problem statement and research motivation 4](#_Toc187059864)

[1.2 The data set 4](#_Toc187059865)

[1.3 Research Question 4](#_Toc187059866)

[1.4 Null hypothesis and alternative hypothesis (H0/H1) 4](#_Toc187059867)

[2. Background research 5](#_Toc187059868)

[2.1 Research papers (at least 3 relevant to your topic / DS) 5](#_Toc187059869)

[2.2 Why RQ is of interest (research gap and future directions according to the literature) 5](#_Toc187059870)

[3. Visualisation 6](#_Toc187059871)

[3.1 Appropriate plot for the RQ output of an R script 6](#_Toc187059872)

[3.2 Additional information relating to understanding the data (optional) 6](#_Toc187059873)

[3.3 Useful information for the data understanding 7](#_Toc187059874)

[4. Analysis 8](#_Toc187059875)

[4.1 Statistical test used to test the hypotheses and output 8](#_Toc187059876)

[4.2 The null hypothesis is rejected /not rejected based on the p-value 8](#_Toc187059877)

[5. Evaluation 9](#_Toc187059878)

[5.1 What Went Well 9](#_Toc187059879)

[5.2 Points for improvement 9](#_Toc187059880)

[5.3 Group’s time management 9](#_Toc187059881)

[5.4 Project’s Overall Judgment 9](#_Toc187059882)

[6. Conclusions 10](#_Toc187059883)

[6.1 Results explained 10](#_Toc187059884)

[6.2 Interpretation of the results 10](#_Toc187059885)

[6.3 Reasons and/or implications for future work, limitations of your study 10](#_Toc187059886)

[7. Reference list 11](#_Toc187059887)

[8. Appendices 12](#_Toc187059888)

1. [R code is used for analysis and visualization analysis. R code with the appropriate statistics to test the hypotheses. 12](#_Toc187059889)

B. [GitHub log output](#_A._R_code)

**List of Figures**

[Figure 1: Histogram of CPU Cores in AMD Processors 6](#_Toc187059890)

[Figure 2: Box Plot of CPU Cores by Processor Family 6](#_Toc187059891)

[Figure 3: Results 8](#_Toc187059892)

# 1. Introduction

## 1.1 Problem statement and research motivation

A constant evolution of processors, especially in the processors manufactured by AMD, has led to different families like the AMD A-Series and Ryzen™ Processors which meet different performance thresholds. While they are widely used, understanding differences in core specifications is a relatively un-investigated area, thus constraining the ability to determine the best use of these processors. Some require a significant number of cores because central processing units’ speed or multitasking and computations mainly depend on this count (Mohammed and Sulaiman, 2023). It is possible to learn whether the differences between the mean or median of the core numbers in these families amount to something meaningful knowing that would help users and industries choose a processor.

## 1.2 The data set

The dataset is 395 entries and 36 columns in size and provides quick information regarding the different specs of AMD processors. It has characteristics such as model, family, platform, and launch date, as well as benchmark figures of CPU cores/threads, GPU cores, and clocks. It also gives information about cache levels (L1, L2, L3), graphics features (model, core number, and supported displays), and physical characteristics of a package, such as package type and the number of sockets. It is a mixture of categorical and numerical types of data.

## 1.3 Research Question

This study explores whether the average or median number of cores of the CPU is significantly different between AMD A-Series and Ryzen™ Processors. Toward this end, a comparison will be made based on descriptive and inferential statistics, using a dataset that contains specifications of AMD processors to compare core counts in the two families.

## 1.4 Null hypothesis and alternative hypothesis (H0/H1)

The null hypothesis (H0) would be that there is no significant difference in the mean or median number of CPU cores between AMD A-Series Processors and AMD Ryzen™ Processors, implying that core counts are statistically similar across these families. The alternative hypothesis (H1) would then be that there is a significant difference, meaning the two families have different distributions of core counts. This hypothesis will be tested using t-tests or Mann-Whitney U tests based on the normality of data. The conclusions would help determine whether it is variability in core counts observed because of inherent differences or some sort of random variation, thus illuminating some aspects of processor choice.

# 2. Background research

## 2.1 Research papers (at least 3 relevant to your topic / DS)

All three research papers that were reviewed include datasets in the analysis although the studies work on different facets of the processors and their design. The first paper titled “Analysis of the Differences Between Intel and AMD Processor Designs” considers and contrasts the features of Intel and AMD processors. Yet, it does not describe a dataset used for comparing processor performance (Lim et al., 2024). Partly due to the methodology, there is much more focus on design aspects like cores and clock rates with a fairly comparative rather than statistical approach.

The second paper identified is titled, “Assessing the Effect of ARM and AMD Processors on Computer Performance” and based on the usage of a variety of performance benchmarks from tasks like gaming and computational, the paper seems to have gathered actual research data to compare the two types of processors Sinaga et al., 2024). In this case, the dataset presumably contains metrics of performance that are effective in various scenarios, but their specifics are not described in the paper in enough detail. Likely the dataset of benchmarks and performance statistics, the third paper is titled “Comparative Analysis of an Intel and AMD Processor” The authors analyze other tags like processing power, energy efficiency, cycle per instruction as well as core logarithm though they do imply the use of some performance-based data but again, the set is not described in detail (Anthony, 2022).

## 2.2 Why RQ is of interest (research gap and future directions according to the literature)

The research question focuses on a gap within the literature in that no one has directly compared core counts between AMD A-Series and Ryzen™ Processors. Previous studies discuss general performance and architectural differences between various processors, but specific information regarding differences in core count variations remains limited. These are key differences because core count highly influences multitasking and computing efficiency. This research contributes to bridging this gap by providing statistically backed findings and guiding users and industries in processor selection. Future studies could expand on this by exploring other performance metrics, such as clock speed and energy efficiency, across various processor families.

# 3. Visualisation

## 3.1 Appropriate plot for the RQ output of an R script

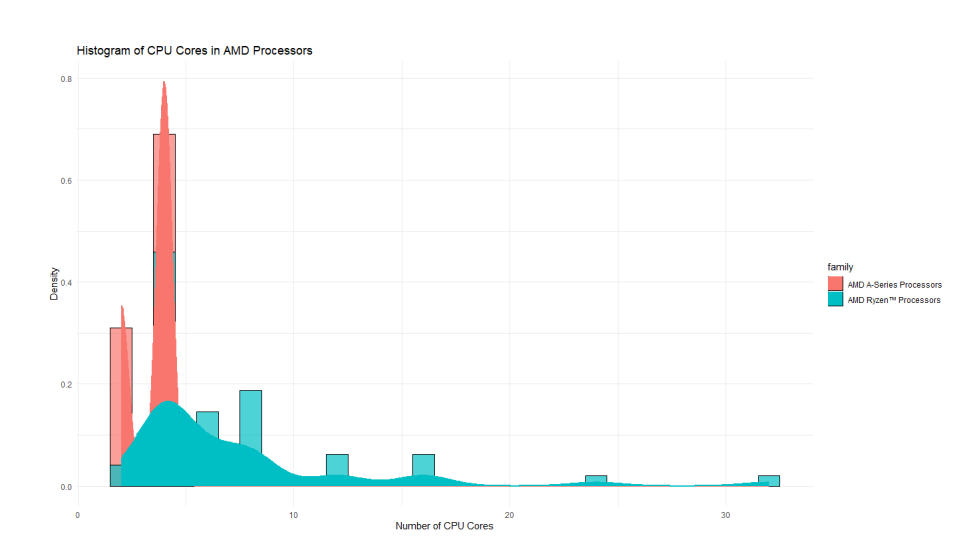


Figure 1: Histogram of CPU Cores in AMD Processors

The above graph image shows that the AMD A-Series processors have a peak of 4 cores, while Ryzen processors show a broader distribution, peaking at around 6 cores, with a range extending up to 32 cores. It shows the distribution of the number of cores in both types of processors by density and number of CPU Cores.

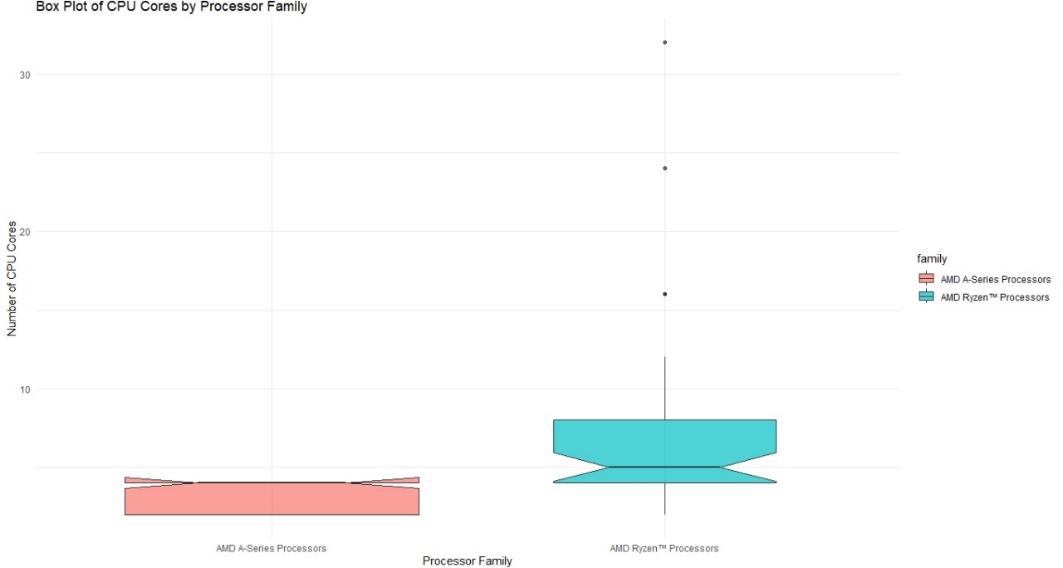


Figure 2: Box Plot of CPU Cores by Processor Family

This graph shows the differences between medians and ranges of the core counts. The box plot shows that the AMD A-Series processors with a median of 4 cores, while AMD Ryzen processors have a median of 8 cores, with outliers reaching up to 32 cores.

## 3.2 Additional information relating to understanding the data (optional)

The histogram shows the distribution patterns of CPU core counts for AMD A-Series and Ryzen™ Processors, which shows that Ryzen™ Processors have a wider range and higher frequency of larger core counts. The boxplot emphasizes the differences in medians, variability, and outliers, thus providing a visual summary of core count disparities between the two processor families.

## 3.3 Useful information for the data understanding

The plots indicate that AMD Ryzen™ Processors possess a larger median and greater variability within CPU core counts than do AMD A-Series Processors. Ryzen™ shows a wider range with more extreme core configurations, whereas A-Series cores congregate around a smaller set of lower counts. As such, there are perhaps more significant differences between the distribution of core counts for the two families.

# 4. Analysis

## 4.1 Statistical test used to test the hypotheses and output

To compare the mean number of CPU cores for 'AMD A-Series Processors with 'AMD Ryzen™ Processors, the T-test for independent samples (ttest\_ind) was used. This test is appropriate because the RQ seeks out if there are differences in the mean number of CPU cores between the two families (Takano, 2021). The data are independent samples, and by density plots, there is a reasonable basis for normality, at which point the t-test is appropriate for comparing means.

## 4.2 The null hypothesis is rejected /not rejected based on the p-value

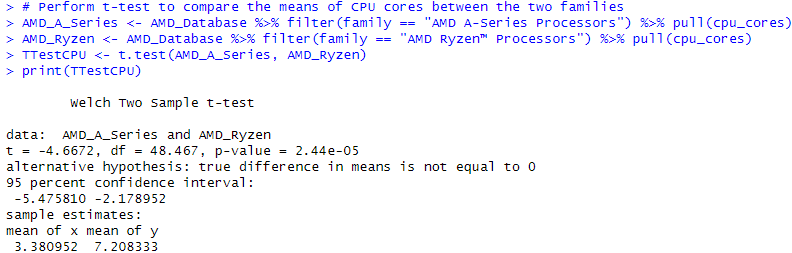


Figure 3: Results

The results of the Welch Two Sample t-test reject the null hypothesis. The p-value of 2.44e-05 is way lower than the typical significance level of 0.05, so there is a circumspect difference in the mean number of CPU cores in AMD A-Series and AMD Ryzen processors. This test also demonstrates that on average, the AMD Ryzen processors have more CPU cores than the AMD AM processors, namely, AMD Ryzen has 7.21 cores and the AMD A-Series processor has 3.38 cores (Fasiku et al., 2014). It is confirmed that the difference is significant and not due to random chance with a 95% confidence interval for the mean difference of -5.48 to -2.18.

# 5. Evaluation

## 5.1 What Went Well

The group efficiently managed data cleaning and transformation tasks such that the dataset was ready for analysis. The dataset was made clearer by renaming columns for better clarity and handling missing values. Histograms and box plots showed how useful precise visualizations can be, and also pinpointed interesting facts, like differences in CPU cores between processor families. Finally, the statistical analysis was nicely performed and the results were nicely reported along with their interpretations.

## 5.2 Points for improvement

Documentation and communication play an important role in the process, and certainly, this area can be improved. The tasks were done efficiently but better effort could've been put together on each step had been done so that everybody was on the same page. Very few assumptions were assumed to have been met (normality assumed), so it would be helpful to explore some other statistical tests, and possibly, if necessary, non-parametric tests (such as the Mann-Whitney U test).

## 5.3 Group’s time management

All the members were able to manage their time well, preparing, visualizing, and analyzing their data during the given time. Some delays in communication and decision-making occurred but, these could have been minimized by working out roles at the beginning of the project. All the members spent time well across tasks.

## 5.4 Project’s Overall Judgment

Overall, this project was able to answer the research question and provide clear and interpretable results. Visualizations and statistical testing were used and the findings were well supported. The collaborative process was thought that refining it would also help the group work efficiently in future projects.

# 6. Conclusions

## 6.1 Results explained

Testing using the Welch Two-Sample t-test showed a large difference in the mean number of CPU cores between AMD A-Series Processors and AMD Ryzen Processors. The result is that the difference is statistically significant, with a p-value of 2.44e-05. AMD Ryzen processors have 7.21 cores on average vs. AMD A-Series cores on average: 3.38 cores. Analysts confirm the disparity in core counts by reporting the 95% confidence interval for the mean difference to be anywhere from -5.48 to -2.18.

## 6.2 Interpretation of the results

From the results, it might be implied that the CPU cores of the AMD Ryzen processors are relatively more than those of the AMD A-Series processors, implying that they are perhaps performing much more optimally in parallel computing tasks. Since Ryzen processors could be more suitable for gaming and content creation, which are high-performance type apps, this difference could affect consumer choices. However, in the context of the broader market, this means that Ryzen can no longer be competed with just by having a lower price point.

## 6.3 Reasons and/or implications for future work, limitations of your study

Future work could be made to include a wider range of AMD processor families and another set of performance metrics (clock speed, TDP), and non-parametric tests when the data may not pass normality assumptions. This study is limited because only one metric (CPU cores) is used in evaluating the performance difference.

# 7. Reference list

Anthony, K. (2022). *Comparative analysis of an Intel and AMD Processor*. [online] Data Science Central. Available at: https://www.datasciencecentral.com/comparative-analysis-of-an-intel-and-amd-processor/ [Accessed 30 Dec. 2024].

‌Fasiku, A., Oyinloye, O.E., O, F.S. and S, A.O. (2014). *Performance Evaluation of Multicore Processors*. [online] Available at: https://www.researchgate.net/publication/350950064\_Performance\_Evaluation\_of\_Multicore\_Processors.

‌Lim, R., Lubis, I.F., Handika, J. and Sitompul, D. (2024). *ANALYSIS OF THE DIFFERENCES BETWEEN INTEL AND AMD PROCESSOR DESIGNS*. [online] ResearchGate. Available at: https://www.researchgate.net/publication/381427331\_ANALYSIS\_OF\_THE\_DIFFERENCES\_BETWEEN\_INTEL\_AND\_AMD\_PROCESSOR\_DESIGNS [Accessed 30 Dec. 2024].

‌Mohammed, J. and Sulaiman, D.R. (2023). Performance analysis of multicore processors using multi-scaling techniques. *International Journal of Power Electronics and Drive Systems/International Journal of Electrical and Computer Engineering*, 13(3), pp.3079–3079. doi:https://doi.org/10.11591/ijece.v13i3.pp3079-3087.

‌Sinaga, Y.M., Situmorang, P., Sinaga, N.A. and Sitompul, D. (2024). *ASSESSING THE EFFECT OF ARM AND AMD PROCESSORS ON COMPUTER PERFORMANCE*. [online] ResearchGate. Available at: https://www.researchgate.net/publication/387166169\_ASSESSING\_THE\_EFFECT\_OF\_ARM\_AND\_AMD\_PROCESSORS\_ON\_COMPUTER\_PERFORMANCE [Accessed 30 Dec. 2024].

‌Takano, S. (2021). Traditional microarchitectures. *Elsevier eBooks*, [online] pp.19–47. doi:https://doi.org/10.1016/b978-0-12-818279-6.00012-8.

# 8. Appendices

## A. R code is used for analysis and visualization analysis. R code with the appropriate statistics to test the hypotheses.

**Code**:

# Load necessary libraries

library(tidyverse)

library(ggplot2)

# Read the dataset

AMD\_Database <- read\_csv("amd(database).csv", show\_col\_types = FALSE)

# Print Column Names

colnames(AMD\_Database)

# Renaming columns for easy analysis

AMD\_Database <- AMD\_Database %>% rename(

  model = Model,

  family = Family,

  line = Line,

  platform = Platform,

  launch\_date = `Launch Date`,

  cpu\_cores = `# of CPU Cores`,

  threads = `# of Threads`,

  gpu\_cores = `# of GPU Cores`,

  compute\_cores = `Compute Cores`,

  base\_clock = `Base Clock`,

  max\_boost\_clock = `Max Boost Clock`,

  all\_core\_boost\_speed = `All Core Boost Speed`,

  total\_l1\_cache = `Total L1 Cache`,

  total\_l2\_cache = `Total L2 Cache`,

  total\_l3\_cache = `Total L3 Cache`,

  unlocked = Unlocked,

  cmos = CMOS,

  package = Package,

  socket\_count = `Socket Count`,

  pci\_express\_version = `PCI Express® Version AMD EPYC™ 7002 Series`,

  thermal\_solution\_pib = `Thermal Solution PIB`,

  default\_tdp = `Default TDP / TDP AMD EPYC™ 7002 Series`,

  ctdp = cTDP,

  max\_temps = `Max Temps`,

  system\_memory\_spec = `System Memory Specification AMD EPYC™ 7002 Series`,

  system\_memory\_type = `System Memory Type`,

  memory\_channels = `Memory Channels`,

  per\_socket\_mem\_bw = `Per Socket Mem BW`,

  graphics\_frequency = `Graphics Frequency`,

  gpu\_base = `GPU Base`,

  graphics\_model = `Graphics Model`,

  graphics\_core\_count = `Graphics Core Count`,

  displayport = DisplayPort,

  hdmi = HDMI,

  dvi = DVI,

  vga = VGA)

# Print head of the AMD\_Database

head(AMD\_Database)

# Print tail of the AMD\_Database

tail(AMD\_Database)

# Check missing values in the AMD\_Database

colSums(is.na(AMD\_Database))

# Print total missing values in the AMD\_Database

sum(is.na(AMD\_Database))

# Remove columns with more than 50% missing values

threshold <- 0.5 \* nrow(AMD\_Database)

AMD\_Database <- AMD\_Database %>% select(where(~ sum(is.na(.)) < threshold))

# Print column names after removing columns with too many missing values

colnames(AMD\_Database)

# Fill remaining missing values with specific values ensuring to match column types

AMD\_Database <- AMD\_Database %>% replace\_na(list(

  platform = "Unknown",

  launch\_date = "Unknown",

  cpu\_cores = 0,

  threads = 0,

  gpu\_cores = 0,

  compute\_cores = "Unknown",

  base\_clock = "0",

  max\_boost\_clock = "0",

  all\_core\_boost\_speed = "0",

  total\_l1\_cache = "0",

  total\_l2\_cache = "0",

  total\_l3\_cache = "0",

  unlocked = "No",

  cmos = "Unknown",

  package = "Unknown",

  socket\_count = "0",

  pci\_express\_version = "Unknown",

  thermal\_solution\_pib = "Unknown",

  default\_tdp = "0",

  ctdp = "0",

  max\_temps = "0",

  system\_memory\_spec = "Unknown",

  system\_memory\_type = "Unknown",

  memory\_channels = 0,

  per\_socket\_mem\_bw = "0",

  graphics\_frequency = "0",

  gpu\_base = "0",

  graphics\_model = "Unknown",

  graphics\_core\_count = 0,

  displayport = "Unknown",

  hdmi = "Unknown"

))

# Check if all missing values are filled

colSums(is.na(AMD\_Database))

# Print structure of the AMD\_Database

str(AMD\_Database)

# Converting the family and cpu\_cores columns in the correct format

AMD\_Database <- AMD\_Database %>%

  mutate(cpu\_cores = as.numeric(cpu\_cores))

# Combined histogram diagram for AMD A-Series and AMD Ryzen Processors

ggplot(AMD\_Database %>% filter(family %in% c("AMD A-Series Processors", "AMD Ryzen™ Processors")), aes(x = cpu\_cores, fill = family)) +

  geom\_histogram(aes(y = ..density..), binwidth = 1, color = "black", alpha = 0.7, position = "identity") +

  geom\_density(aes(y = ..density.., color = family), size = 1) +

  labs(title = "Histogram of CPU Cores in AMD Processors", x = "Number of CPU Cores", y = "Density") +

  theme\_minimal()

# Box plot diagram comparing CPU cores between AMD A-Series and AMD Ryzen Processors

ggplot(AMD\_Database %>% filter(family %in% c("AMD A-Series Processors", "AMD Ryzen™ Processors")), aes(x = family, y = cpu\_cores, fill = family)) +

  geom\_boxplot(alpha = 0.7, notch = TRUE, varwidth = TRUE) +

  labs(title = "Box Plot of CPU Cores by Processor Family", x = "Processor Family", y = "Number of CPU Cores") +

  theme\_minimal()

# Perform t-test to compare the means of CPU cores between the two families

AMD\_A\_Series <- AMD\_Database %>% filter(family == "AMD A-Series Processors") %>% pull(cpu\_cores)

AMD\_Ryzen <- AMD\_Database %>% filter(family == "AMD Ryzen™ Processors") %>% pull(cpu\_cores)

TTestCPU <- t.test(AMD\_A\_Series, AMD\_Ryzen)

print(TTestCPU)

B. GitHub log output

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
| '7fb55b4 | 2025-01-07T00:53:07Z | Hemil Parmar | CPU Specifications - AMD Data Set' | | | |
| 'a568a32 | 2025-01-07T00:44:20Z | Hemil59 | Delete AMD\_Database.csv' | | |  |
| '39e595d | 2025-01-07T00:42:28Z | Hemil Parmar | CPU Specifications - AMD Data Set' | | | |
| '7f7ac1b | 2025-01-07T00:34:11Z | Hemil Parmar | Delete' |  |  |  |
| '9dcd551 | 2025-01-07T00:27:03Z | Hemil Parmar | My First' |  |  |  |
| 'f03fdc9 | 2025-01-06T13:53:58Z | xpernza | Add files via upload' | |  |  |
| 'cc42ad4 | 2025-01-06T13:47:07Z | Hemil59 | Delete Team-Research-2024.Rproj' | | | |
| '306fed0 | 2025-01-06T13:46:53Z | Hemil59 | Delete .gitignore' | |  |  |
| '25eed38 | 2025-01-06T13:45:35Z | Kartik Mistry | analysis and evaluation file' | | |  |
| 'b06a348 | 2025-01-06T13:17:47Z | Hemil59 | Delete amd1.csv' | |  |  |
| '7136932 | 2025-01-06T12:59:54Z | Hemil Parmar | histogram image and box plot image' | | | |
| 'fd0d528 | 2025-01-06T12:53:36Z | Hemil59 | Delete Team-Research-2024.Rproj' | | | |
| '4345d93 | 2025-01-06T12:52:55Z | Hemil59 | Delete .Rproj.user directory' | | |  |
| '648e0d4 | 2025-01-06T12:51:29Z | Hemil Parmar | Visualisation and Appendices text file' | | | |
| '6e2b702 | 2025-01-06T12:45:29Z | Hemil59 | Delete histogram\_cpu\_cores.png' | | | |
| '840e828 | 2025-01-06T12:45:18Z | Hemil59 | Delete boxplot\_cpu\_cores.png' | | |  |
| '04e3cd9 | 2025-01-05T19:34:06Z | Andappa Chandoor | Research' |  |  |  |
| '3dc2c46 | 2025-01-05T19:14:39Z | Poorna443 | Create Conclusions & Reference' | | | |
| 'f27d2df | 2024-11-24T23:21:25Z | Hemil59 | Add files via upload' | |  |  |
| 'a6cf538 | 2024-11-24T23:19:44Z | Hemil59 | Update README.md' | |  |  |
| '223f282 | 2024-11-24T23:19:24Z | Hemil59 | Delete Screenshot 2024-11-22 095251.png' | | | |
| '619d3ff | 2024-11-24T23:19:05Z | Hemil59 | Delete Histogram' | |  |  |
| '6fd1747 | 2024-11-24T22:24:54Z | Hemil59 | Add files via upload' | |  |  |
| 'd07d390 | 2024-11-22T09:56:32Z | Poorna443 | Create Histogram' | |  |  |
| 'f758a08 | 2024-11-22T09:54:39Z | Poorna443 | Update README.md' | |  |  |
| 'a1899ef | 2024-11-22T09:53:52Z | Hemil59 | Add files via upload' | |  |  |
| 'b3efe80 | 2024-11-19T20:50:58Z | Hemil59 | Create README.md' | |  |  |
| '9bf9b97 | 2024-11-17T15:25:57Z | Poorna443 | CPU Specifications - AMD Data Set' | | | |
| '4a7c8aa | 2024-11-17T15:23:27Z | Poorna443 | CPU Specifications' | |  |  |