Jalen Randleman

DSC 140

Dr. butler

12/1/2024

The Analysis of GPUs Throughout History

In today's modern age of technology, virtually everyone on this planet has used a computer, played video games, or watched television. When using these pieces of technology, there are many key components that allow these to work. The most notable component is the graphics processing unit, more commonly known as the graphics card or GPU. According to Intel, a GPU is "Designed for parallel processing, the GPU is used in a wide range of applications, including graphics and video rendering" (Intel). Due to this, GPUS are one of the cornerstones of modern technology as there are many different types of GPUs made by multiple different companies. At the forefront of GPU manufacturing stands three companies: NVIDIA, AMD, and Intel. When manufacturing and choosing a GPU, there are many individual specs that one must make note of. Many of these include the memory size, bus width, and clock speed. As well as rendering output units (ROP) and texture mapping units (TMU), which are important components that perform tasks within the rendering pipeline. Due to this, many datasets have compiled every GPU in history from the three main companies. In this report, we are going to analyze one created by Tesla, Inc., which is found on the website known as Kaggle, as it takes note of NVIDIA, AMD, and Intel's GPUs full individual specs and how they evolved overtime. Throughout this analysis, many questions are going to be answered, such as how have specs changed overtime, the differences between companies, and much more. To do this, we are going to leverage the techniques we learned in data science.

First, what is Data Science and how does it help? Data science is a discipline that allows people to gather intel and find interesting insights that can be used as guides for people and companies to help create plans and make informed decisions (IBM). According to IBM, which is a company that provides, hardware, software, and more to clients, states that "data science combines math and statistics, specialized programming, advanced analytics, artificial intelligence (AI) and machine learning with specific subject matter expertise to uncover actionable insights hidden in an organization's data" (IDM). In data science, there are many ways to use conduct data analysis, the most notable ways are through programming languages, such as Python and R. Based on this; to conduct our GPU analysis, we are going to use the programming language of Python. The reason for this is because Python grants us the abilities to use numerous libraries to conduct our research, such as Pandas, Matplotlib, and NumPy, which allow us to efficiently create graphs and statistical tests for data analysis. In this analysis, we are going to use these tools to create statistical summaries, find correlations, and create machine learning algorithms to answer our questions.

When analyzing the dataset, we must first do a statistical analysis, which means finding the minimum, maximums, standard deviations, etc. of each column of data. However, in the GPU dataset, there were 16 columns of data, each column having between 2,400 and 2,900 rows of data. However, two of the 16 columns were mostly empty; to make the analysis smoother, it was decided to drop the columns entirely. These columns being pixel shader and vertex shader. Additionally, not every row was completely filled out, with some having blank space. Due to this, it was determined that dropping each row that was not completely filled out was necessary. After this process, the dataset was then left with 14 columns of data, which included: manufacturer, productName, releaseYear, memSize (in GB), memBusWidth (in bits), gpuClock (in MHz), memClock (in MHz), unifiedShader, tmu, rop, igp, bus, memType, gpuChip; all of which consisted of 1721 rows of data.

Now that the dataset has been "cleaned," we can now perform the statistical analysis. When conducting the analysis, there were two tables that were created, one for numerical columns (see Figure 1.1) and one for categorical columns (see Figure 1.2). In the numerical columns, the most notable findings were that the minimum release year was from 2005, while the maximum release year was 2023. Additionally, in the categorical columns, the most notable finding was that the top manufacturer was NVIDIA. The reason for this is because NVIDIA was shown to be the most frequent manufacture in the dataset, appearing 904 times.

	releaseYear	memSize	memBusWidth	gpuClock	memClock	unifiedShader	tmu	rop
count	1721.000000	1721.000000	1721.000000	1721.000000	1721.000000	1721.000000	1721.000000	1721.000000
mean	2013.571761	4.345259	322.584544	861.147008	1102.266124	1170.514817	71.453806	27.373620
std	4.130951	8.273479	761.177359	325.848376	407.848886	1769.078898	84.930007	28.780896
min	2005.000000	0.128000	32.000000	300.000000	266.000000	8.000000	4.000000	0.000000
25%	2010.000000	1.024000	128.000000	620.000000	800.000000	160.000000	20.000000	8.000000
50%	2013.000000	2.000000	128.000000	796.000000	1000.000000	480.000000	40.000000	16.000000
75%	2017.000000	4.000000	256.000000	1005.000000	1375.000000	1536.000000	96.000000	32.000000
max	2023.000000	128.000000	8192.000000	2331.000000	2257.000000	17408.000000	880.000000	256.000000

Figure 1.1: Statistical Data of Numerical Columns

	manufacturer	productName	igp	bus	memType	gpuChip
count	1721	1721	1721	1721	1721	1721
unique	4	1592	1	21	15	203
top	NVIDIA	GeForce GT 555M	No	PCIe 2.0 x16	GDDR5	GK104
freq	904	5	1721	539	712	51

Figure 1.2: Statistical Data of Categorical Columns

After conducting the statistical summary, we can find relationships between different columns of data. First, the numerical data. While there were many columns to correlate, there were two that stood out the most. Foremost, the year released, and memory clock speeds was correlated (see Figure 2.1) using a Pearson correlation. These were found to be a very positive correlation, with a correlation score of 0.741. What this means is that as the year increased, so did GPU memory clock speeds.

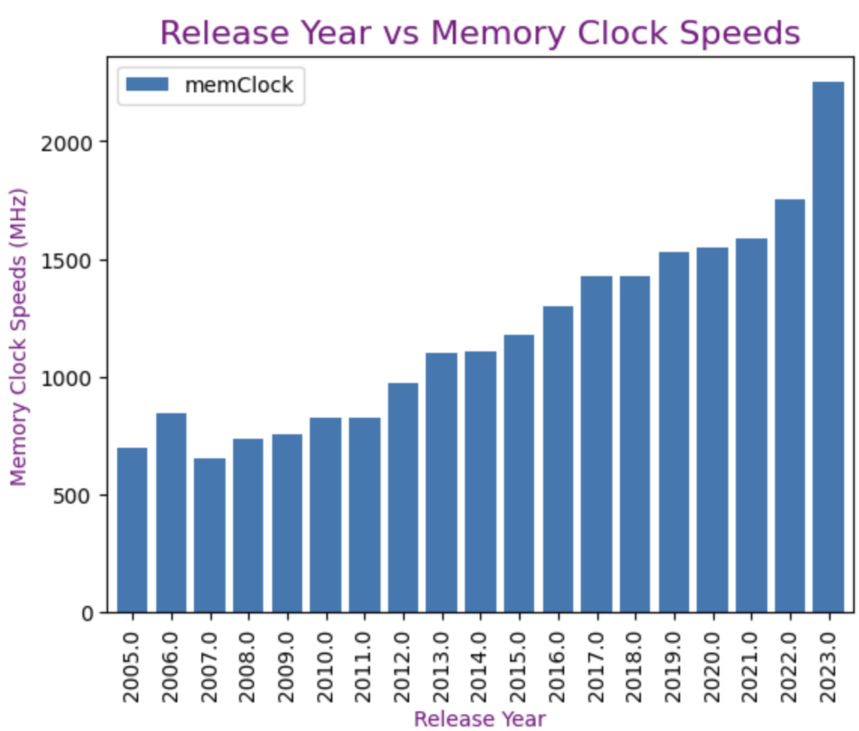

Figure 2: Release Year vs Memory Clock Speeds

In addition to the Pearson correlation, a Spearman correlation was conducted to explore the monotonic relationship between ROP and TMU as opposed to the linear relationship that a Pearson correlation would show. What was found, was that the correlation score was 0.924, meaning that as one increases, so does the other.

Alongside the correlations of numerical data, categorical data was next. In order to find relationships between these variables, a chi-squared test was in play, which is a test that determines how statistically significant a relationship is between two columns of data. In the analysis, the test was performed between the columns manufacturer, and bus, which is the pathway that allows transfer between GPU and VRAM. During this test, it was found that the p-value was 3.527e-110, which is a very small number. This means that the manufacturer and bus are very closely related and statistically significant.

On top of this analysis, we decided to implement a machine learning algorithm, known is k-means. What this algorithm does is that it takes the data from the dataset and compiles it into clusters based on a given k-value, which determines how many clusters to group the data in. The algorithm also bases these clusters off columns of data given to it. In this case, it was determined that the clusters would be based on GPU clock speeds and memory clock speeds. However, to determine the number of clusters needed in the analysis, an elbow plot was at play, which is a method that helps determine the optimal number of k values (see Figure 3.1).

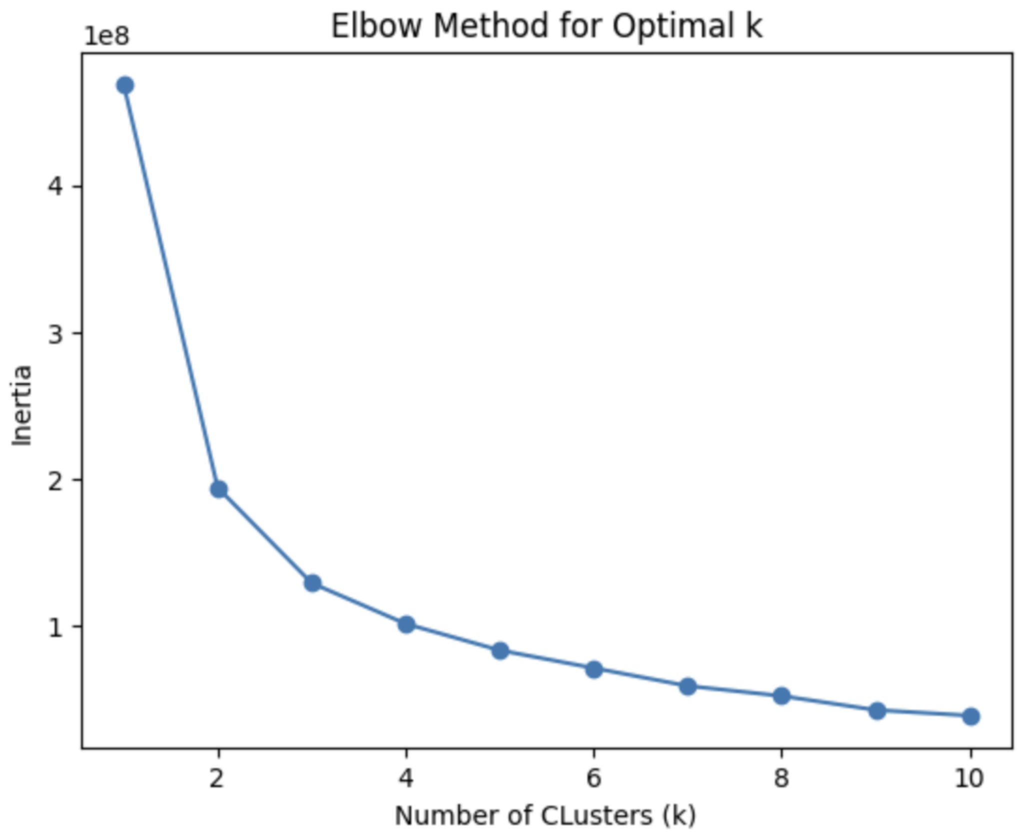

Figure 3.1: Elbow Method for Optimal k

In this case, the optimal number of k values was determined to be nine. The reason for this is because nine is where the elbow plot starts to even out, meaning that there is no longer any variability. Due to this, a scatter plot was able to be made using the number of clusters to visualize the data (see Figure 3.2). Based on the graph, it was determined that the lowest cluster's GPU clock mean was 571.55 MHz, with a median of 564.0 MHz; it also contained a memory clock mean of 472.53 MHz, with a median of 500.0 MHz. On the other hand, the highest cluster's GPU clock mean was 1999.67 MHz, with a median of 1946.5 MHz; it also contained a memory clock mean of 2046.89 MHz, with a median of 2000.0 MHz. As regards to what this data means, it can be implied that GPUs can be segmented based on their clock speeds, where lower clusters may be more suited for entry level tasks, while higher clusters may be more suited towards much heavier tasks, such as high-performance video editing or gaming.

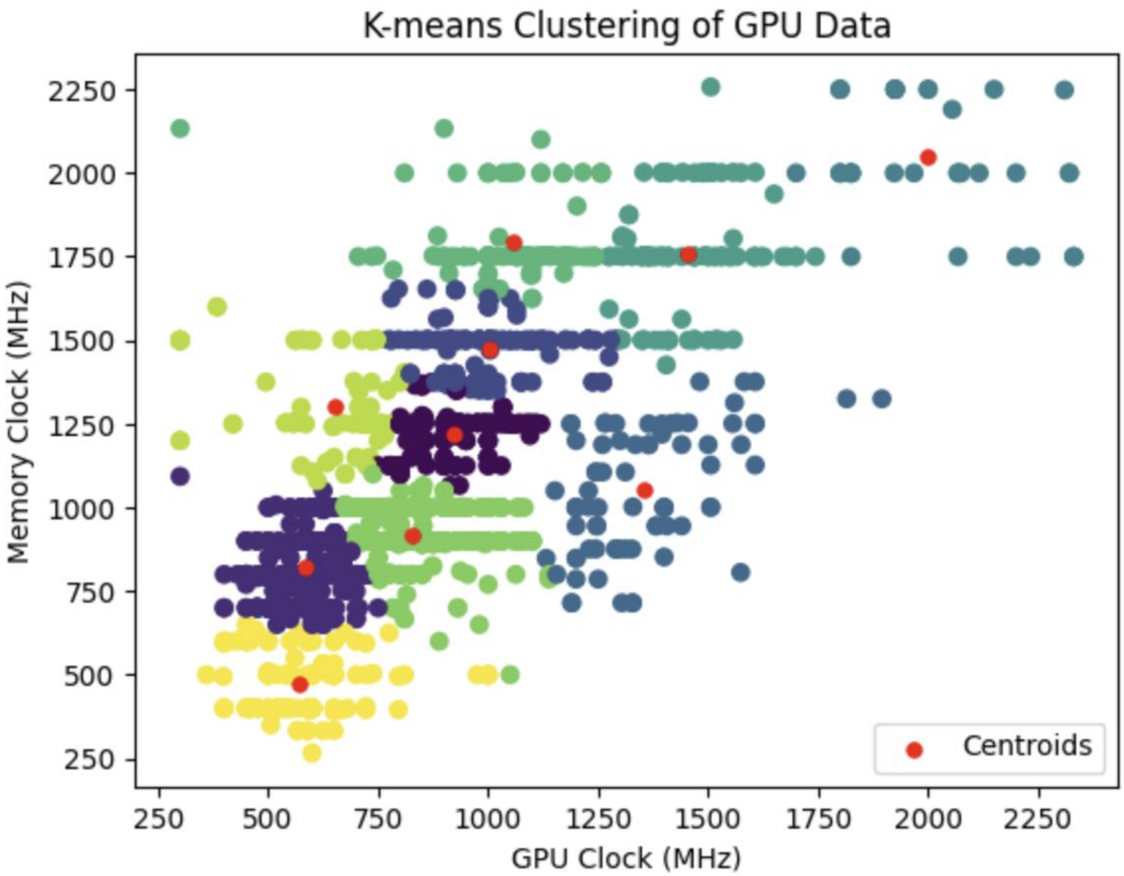

Figure 3.2: K-means Clustering of GPU Data

After conducting this analysis, we now have a better understanding of the GPU dataset. Due to this, we can now test and analyze to answer our previous questions. The first of our question is: How have GPU specs changed overtime? To answer this, over the course of many years, GPU specs have seen an overall increase in overall performances across the board (see Figure 4.1). It was found that GPU clock speeds had a 285% increase from 2005-2023; memory bus width had a 0% increase from 2005-2023, due to the drop in performance, but had a 262.5% increase from 2005-2022; memory clock speeds had a

221.43% increase from 2005-2023; and memory size had a 1,462.5% increase from 2005-2023.

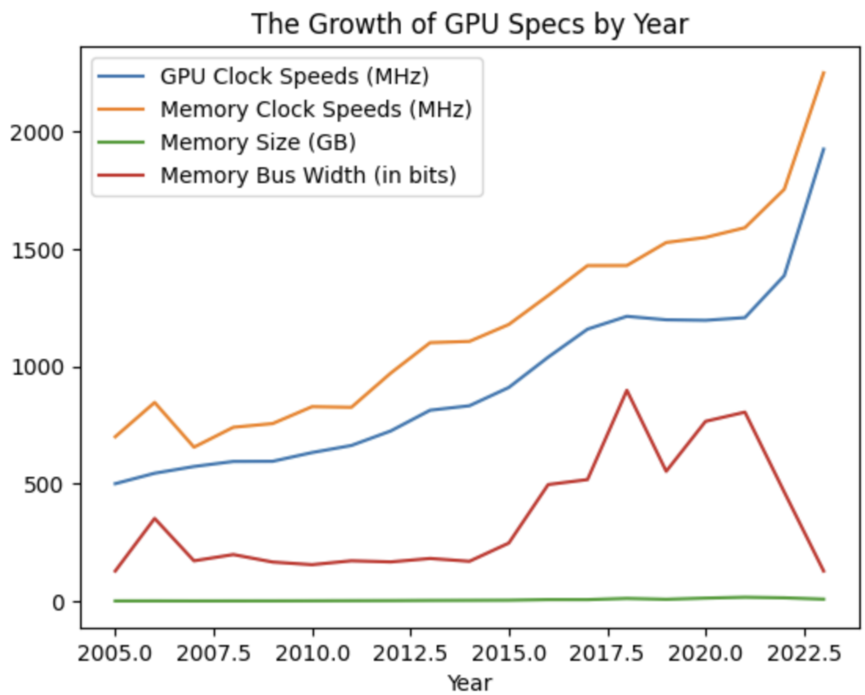

Figure 4.1: The Growth of GPU Specs by Year

The next question that was asked was: How do companies differ from one another; or in other words, which companies produce different specs better than others? To answer this question, four pivot tables were creating using the values of memory size, bus width, and clock speeds. GPU clock speeds were also used. In doing this, it was found that Intel was the top manufacturer, earning the top spots in memory size, memory bus width, and memory clock speeds; with averages of 10.2 GB, 851.2 bits, and 1426.7 MHz respectively. However, while this may be the case, you must first consider that Intel only started producing GPUs much later than AMD and NVIDIA, which likely causes them to have higher specs. Despite this, the highest performer when it came to GPU clock speeds was AMD with an average of 946.3 MHz, while Intel placed third out of the three.

Finally, the third question was how have rendering output units (ROP) and texture mapping units (TMU) changed overtime. To answer this, 2-sample Kolmogorov-Smirnov Test to determine if these two were a good match to begin with. Based on this test, they were as the K-S statistic was 0.339 and the p-value was 6.85e-88, which means that there was a moderate distribution between TMU and ROP based on the K-S statistic, but they were correlated based on the p-value. Due to this, TMU and ROP were then grouped by release year and plotted onto a graph to view how they have changed over time (see figure 4.2).

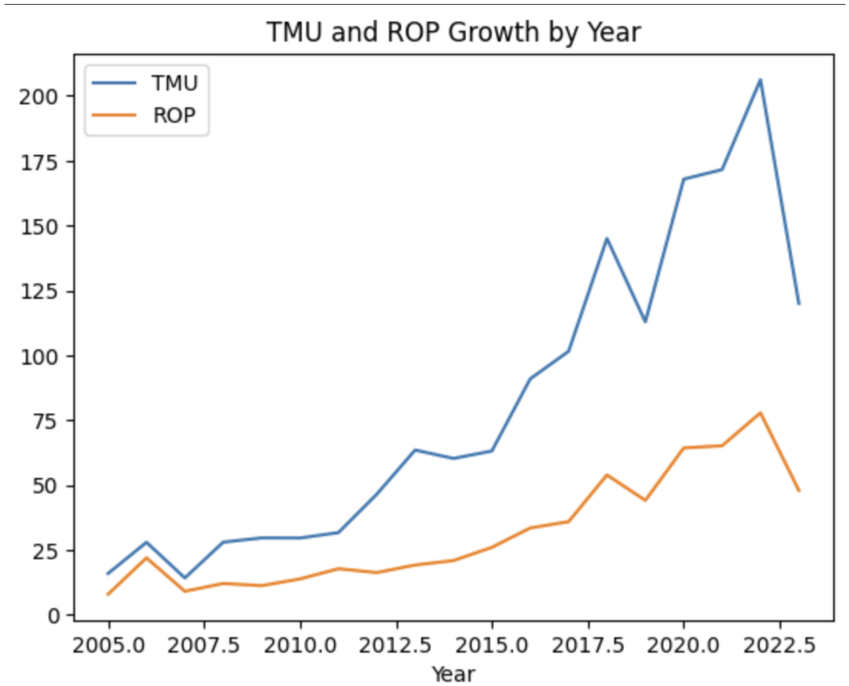

Figure 4.2: TMU and ROP Growth by Year

Based on this graph, it can be viewed that, will there is a drop for both TMU and ROP in 2023, there was an overall increase from 2005-2023. With this increase being 650% and 500% respectively.

So, what does this all mean and how does it help GPU manufacturers in the future. When it comes to selecting or manufacturing a GPU, these are all things that must be considered. For example, Figure 2.1 showed how memory clock speeds have increased over the years. By knowing this data, people can know just how fast memory clocks should be, how much they should be increasing each year. On the contrary, by using the data found in figure 4.2, manufactures can interpret that, while there has been an increase over the years, the drop in 2023 is something that should be looked at, and avoided in 2024.

As it stands, data science is a fundamental in the GPU industry, and even other industries as well. The reason for this is because data science allows people to apply tools to conduct analysis and research, mainly using python and R studio. By doing this, data scientists can make insights into what the future holds for a company. This applies to this analysis as well. By using the data here, companies can interpret how they should approach the future in the manufacturing of GPUs and how they should work. Through this analysis, we have uncovered how GPU specs have changed overtime, which companies produce the best specs, and how TMU and ROP have changed overtime. Overall, GPU technology has grown and improved over the course of many years and shows no signs of slowing down anytime soon.

Works Cited

"What Is a GPU? Graphics Processing Units Defined." *Intel*, www.intel.com/content/www/us/en/products/docs/processors/what-is-a-gpu.html. Accessed 1 Dec. 2024.

"What Is Data Science?" *IBM*, 1 Nov. 2024, www.ibm.com/topics/data-science.

Inc., Tesla. "Nvidia & AMD Gpus Full Specs." *Kaggle*, 27 Aug. 2024, www.kaggle.com/datasets/alanjo/graphics-card-full-specs.