Research Project:

Final Paper

Dalton Murray

College of Business & I.T., Lawrence Technological University

INT 6303, Introduction to Social Media Data Analytics

Dr. Mazin Al Hamando

July 18, 2023

Introduction

Technology, especially computer hardware, is constantly changing, with these changes there are constant releases of new hardware and new requirements to be able to do basic things such as accessing applications and running games. These new requirements require newer hardware which has incredibly high performance, as a result of this there is an ever-growing price for this new hardware. The performance to cost ratio is becoming astonishing, making it less worth the money to buy the new hardware but it is becoming necessary to purchase because of the requirements of applications and games. There is also a growing monopoly which further allows companies such as Nvidia not to worry about their prices because there are limited options available.

In order to figure out how true computer hardware prices are and to understand what people think about computer hardware and their price to performance, research must be conducted. However, we not only need to figure out what people think about computer hardware price to performance we have to figure out what needs to be done in order to fix this if anything is even considered wrong by the consumers of this market. Due to how large this question is, Graphics Processing Units, or GPUs, will be the only topic of focus rather than computer hardware in general. This should also allow for a more precise analysis as well as more fine results.

Data Aggregation and Analysis

In order to properly perform any sort of data aggregation you have to know exactly what you are looking for, or at least have a good idea. Since we have already redefined the scope to be more precise, it should be easier to be able to do this. Then we have to determine the source of the data we want to use, I chose Reddit because it is not only the largest platform in the world, but it will allow us to be able to get data which is relevant to the topic. Initially I found around 20 posts and cut it down to 15, unfortunately, however, due to Reddit's API request limitations I had to cut it down even further, leaving me with only 8 posts. I had initially planned for roughly 10,000 comments but I am now at a total of 2,364. Despite this being significantly less than initially planned, it will still work perfectly and allow us to get a more precise idea of the market.

Following up with finding the posts I wanted to use, I then had to find a tool to be able to gather, or scrape, the comments. I used something called URS, or Universal Reddit Scraper. Unfortunately, its documentation isn't very good, and it took a couple of weeks to make it work properly, however, in the end, it was a great tool to use. After figuring it out, I was able to properly tell it how I want everything scraped. I then had to do some manual filtering and organization to make the data work with the programming language R, which is what was used for the data analysis outside of manual review.

Once everything was properly formatted the analysis could start. To start off, I ran the data through every basic function we learned during this course, which allowed me to be able to get a good idea of the trends of the market as well as what everyone's feelings were. It also gave me really good information from all of the data. After running it through basic functions I started to create graphs and plots of the data to more easily

visualize it and understand it, which proved to be incredibly useful. An example of a graph of the data can be found at Figure 1 in this paper.

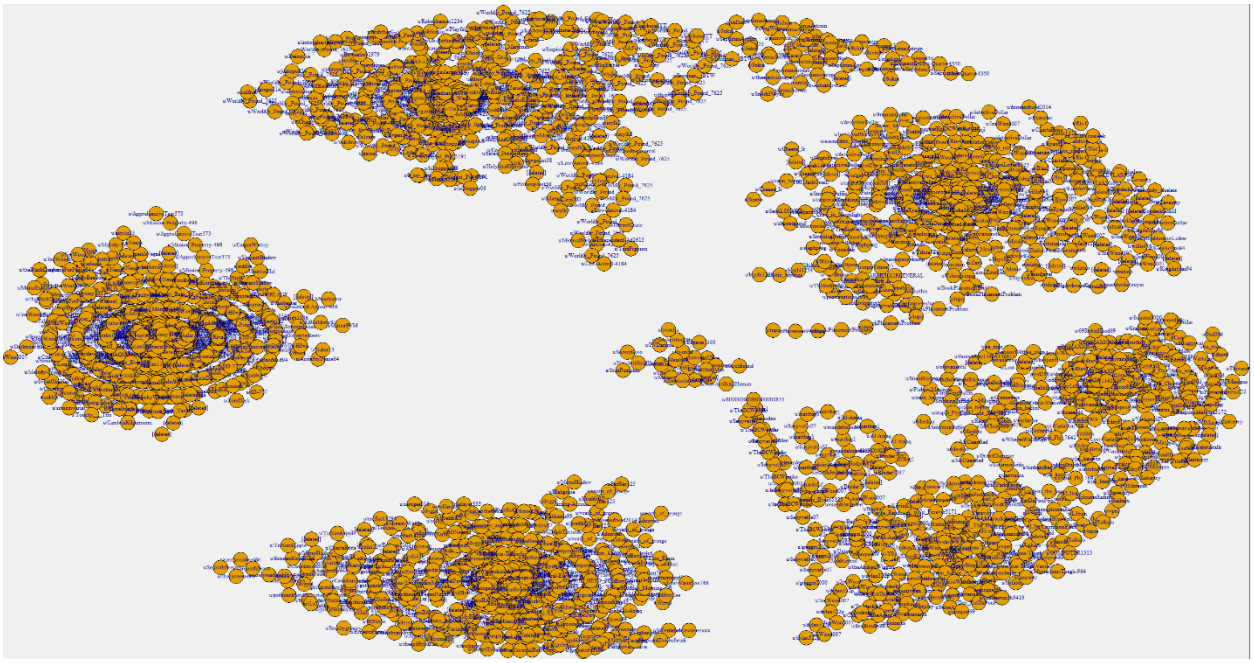

Figure 1 General clustered plot (*Murray, D.*).

Figure 1 shows the 8 different posts which were used, their center points or the original post, and the surrounding comments. This is useful as we are able to see how large a single post is, and how much that post should have sway, or influence, over final results as generally a larger post should have more influence than a smaller post. We also need to consider the score of each post and comment to determine sway, or influence. The score, or karma, is calculated by taking the total amount of upvotes and subtracting it by the total number of downvotes per each post and individual comment. In general, a higher score means more influence as more people agree with what the comment or post is saying.

Continuing, in order for me to also understand the data better, I had to conduct further outside research on data analysis. This is where I read the following articles: The Influence of Power Shifts in Data Collection and Analysis Stages: A Focus on Qualitative Research Interview, Studying Reddit: A Systematic Overview of Disciplines, Approaches, Methods, and Ethics and The anatomy of Reddit: An overview of academic research (Aryan, F; Medvedev, A., Lambiotte, R., & Delvenne, J.-C.; Proferes, N., Jones, N., Gilbert, S., Fiesler, C., & Zimmer, M.). These three articles helped me greatly in understanding how the data in Reddit should be used to influence my final conclusion as well as how to properly conduct my own research which I used in the initial stages in my Technical Documentation for this project (Murray, D. Research Project Technical Documentation.).

Brands

Intel

When reviewing the analysis specifically for the brand Intel as well as manually reviewing the highest scoring comments, there are a few trends and results which are extremely important to take into consideration. To start off, Intel has only recently started making graphics cards with their Arc GPU's (Intel® ARCTM graphics overview).

The pricing of their graphics cards is an amazing deal for most people, they provide a great budget GPU which performs equal to most mid to high-end graphics cards (Murray, D. Reddit Graphics Card Scraped Data.).

The only downside of Intel's graphics cards currently is the lack of driver support. Since their GPU's are so new, they have a lot of firmware and software issues, but this is reportedly getting better, which also has the upside of increasing performance (Murray, D. Reddit Graphics Card Scraped Data.).

In order to improve Intel's graphics card, I propose the following: keep working on the firmware/drivers as much as you can, this is the current largest problem, following this then work on increasing performance and including more VRAM. Once Intel has done these things, they should be able to compete with Nvidia for the number one spot.

AMD

When reviewing the analysis of the data as well manually looking at the data for AMD there are a few clear tends. Firstly, a vast majority of people agree that AMD is the current best budget graphics card (Murray, D. Reddit Graphics Card Scraped Data.). Their performance also matches the price well, meaning that most people think that they are not underpaying or overpaying for Amd's graphics cards (Murray, D. Reddit Graphics Card Scraped Data.). A lot of people also think that AMD is still quite far behind comparing their top performance graphics card to Nvidia's top performance graphics card (Murray, D. Reddit Graphics Card Scraped Data.).

For AMD I propose the following solution: create a truly higher end card which is able to compete with Nvidia's high end card, a lot of people would be

willing to purchase these cards due to the current reputation of Nvidia, however, also balance this new high-end card with its price.

Nvidia

While reviewing Nvidia's analysis of data as well as manually reviewing it there are a large number of trends. It is clear that Nvidia is currently the number one seller, however, they are also currently the most despised brand (Murray, D. Reddit Graphics Card Scraped Data.). Nvidia has the current highest performance graphics card, but a wide majority of people believe that this graphics card is too expensive for the performance (3DMark Time Spy benchmark scores; Murray, D. Reddit Graphics Card Scraped Data.). Nvidia also has arguably less VRAM than Amd's graphics card, and in today's times VRAM is very much needed (Murray, D. Reddit Graphics Card Scraped Data.).

In order for Nvidia to become better, I propose the following: lower your prices to match your competitors but keep it high enough for the performance not to lower, as well as increase VRAM capacity as a lot of the cards from Nvidia do not offer more than 8GB which is barely touching what some games and applications require (Murray, D. Reddit Graphics Card Scraped Data.).

Steam/Valve

Since the Internet is a large place, and there are many people who say things, but their actions do not back up what they say, it is a good idea to verify this data again by looking outside of what people are saying on Reddit. In order to do so, we must look

at the facts. The facts which we can look at here are very easy to find, since we are seeing people complaining about Nvidia's prices, as well as performance of the Intel GPUs as well as AMD's GPUs we could reasonably assume that there will be a larger percentage of people who buy Nvidia's lower end cars because their price is lower than the end high cards and the performance is still relatively decent compared to the other brands.

Let's now take a look at the statistics provided by Steam, which is the largest online gaming marketplace and launcher which allows people to purchase and play games from, it has millions of active users daily (Steam Hardware & Software Survey.).

The Nvidia GeForce GTX 1650 has currently the largest amount of people using it, at an incredibly high 5.67% for June 2023, the next closest is Nvidia's 3060 at 4.75% June 2023, then its 1060 graphics card at 4.58% June 2023 (Steam Hardware & Software Survey.). Further looking at Nvidia's high end, only a very low amount of people uses their absolute highest end card, the 4090, with 0.56% of people using one (Steam Hardware & Software Survey.). This data clearly supports the results of the data analysis from the Reddit scrape as the lower end cards have a much higher percentage of users because people simply cannot afford the high end or believe it has a good enough performance to justify the price. It is not until we get to 2.01% of people using a graphics card in June 2023 where we even see an AMD graphics card, and at 1.70% of people using an Intel graphics card which demonstrates the severe need for AMD and Intel to boost their performance to compete with Nvidia.

Conclusion

The results of the Reddit scraped data are very clear on what people believe. They believe that Nvidia needs to lower their prices despite them having the highest performance, as well as increase their VRAM capacity, which is also what I propose as a solution for them (Murray, D. Reddit Graphics Card Scraped Data.). AMD needs to create a high-end card which can compete with Nvidia and have a price to undercut the market away from Nvidia, which I also propose as a solution for them (Murray, D. Reddit Graphics Card Scraped Data.). Intel needs to work on their drivers and create higher end cards, which I also propose as a solution (Murray, D. Reddit Graphics Card Scraped Data.). We can see that each brand has their own problems they need to worry about and fix, however, it is clear that Nvidia has substantially the largest userbase (Steam Hardware & Software Survey.).

References

3DMark Time Spy benchmark scores. 3DMark.com. (n.d.). https://www.3dmark.com/search#advanced?test=spy%20P&cpuId=&gpuId=1509&gpuCount=0&gpuType=ALL&deviceType=ALL&storageModel=ALL&memoryChannels=0&country=&scoreType=overallScore&hofMode=false&showInvalidResults=false&freeParams=&minGpuCoreClock=&maxGpuCoreClock=&minGpuMemClock=&maxGpuMemClock=&minCpuClock=&maxCpuClock=

Anyan, F. (2013). The Influence of Power Shifts in Data Collection and Analysis Stages: A Focus on Qualitative Research Interview. https://ntnuopen.ntnu.no/ntnu-xmlui/bitstream/handle/11250/2624458/Anyan.pdf?sequence=2

Burnes, A. (n.d.). GeForce RTX 4060 & RTX 4060 TI announced: Available from May 24th, starting at $299. NVIDIA. https://www.nvidia.com/en-us/geforce/news/geforce-rtx-4060-4060ti/

GPU Specs Database. TechPowerUp. (2023, June 16). https://www.techpowerup.com/gpu-specs/?sort=name

Graphics Cards Benchmarks - MSRP Price. Best Graphics Cards. (2023, July 13). https://benchmarks.ul.com/compare/best-gpus

Intel® ARCTM graphics overview. Intel. (n.d.). https://www.intel.com/content/www/us/en/products/details/discrete-gpus/arc.html

Klotz, A. (2023, May 24). Where to buy Nvidia RTX 4060 ti 8GB gpus: Links and prices, all custom cards. Tom's Hardware. https://www.tomshardware.com/news/where-to-buy-nvidia-geforce-rtx-4060-ti-8gb-links-prices-custom-cards-newegg-amazon-best-buy#:~:text=Here%20we've%20compiled%20a,at%20the%20official%20%24399%20MSRP

Lai, J. (n.d.). Josephlai241/urs: Universal reddit scraper - a comprehensive reddit scraping/archival command-line tool. GitHub. https://github.com/JosephLai241/URS

Medvedev, A., Lambiotte, R., & Delvenne, J.-C. (2020, June). The anatomy of Reddit: An overview of academic research. https://arxiv.org/pdf/1810.10881.pdf

Meikleham, D. (2023, April 12). Nvidia GeForce RTX 4070 unveiled for $599, and it could hit the sweet spot of Price and performance. Tom's Guide. https://www.tomsguide.com/news/nvidia-geforce-rtx-4070-unveiled-for-dollar599-and-it-could-hit-the-sweet-spot-of-price-and-performance

Murray, D. (2023, July). Reddit Graphics Card Scraped Data. "Formatted_comments_combined".; ""Formatted_primaryposts_combined". Database.

Murray, D. (2023, July). Research Project Technical Documentation.

Proferes, N., Jones, N., Gilbert, S., Fiesler, C., & Zimmer, M. (2021). Studying Reddit: A Systematic Overview of Disciplines, Approaches, Methods, and Ethic. https://journals.sagepub.com/doi/pdf/10.1177/20563051211019004

Reddit - GPU Pricing Update: Hardware Still Overpriced?. Reddit. (n.d.-a). https://www.reddit.com/r/hardware/comments/10jd9xf/gpu_pricing_update_hardware_still_overpriced/

Reddit - GPUs are very expensive nowadays, what would be a reasonable price for a low, mid and high end GPU for you?. Reddit. (n.d.-a). https://www.reddit.com/r/gaming/comments/12vl57c/gpus_are_very_expensive_nowadays_what_would_be_a/

Reddit - Most PC hardware pricing is back to normal so why not GPUs?. Reddit. (n.d.-c). https://www.reddit.com/r/pcgaming/comments/128r3j5/most_pc_hardware_pricing_is_back_to_normal_so_why/

Reddit - price of GPU are insane. Reddit. (n.d.). https://www.reddit.com/r/pcgaming/comments/105lgwm/price_of_gpu_are_insane/

Reddit - What are your thoughts on value for money on current gen cards (RTX 40/RX 7000) for RT gaming? Which is better?. Reddit. (n.d.-f). https://www.reddit.com/r/buildapc/comments/124tz2b/what_are_your_thoughts_on_value_for_money_on/

Reddit - What's going on with Nvidia GPU prices? Reddit. (n.d.-c). https://www.reddit.com/r/buildapc/comments/126rjun/whats_going_on_with_nvidia_gpu_prices/

Reddit - Why are graphic card prices still so high if they are not selling?. Reddit. (n.d.-e). https://www.reddit.com/r/hardware/comments/1134ddq/why_are_graphic_card_prices_still_so_high_if_they/

Reddit - [HUB] Nvidia and AMD Keep Dropping GPU Prices - June GPU Pricing Update. Reddit. (n.d.-a). https://www.reddit.com/r/hardware/comments/14ec8zx/hub_nvidia_and_amd_keep_dropping_gpu_prices_june/

Steam Hardware & Software Survey. (2023, June). https://store.steampowered.com/hwsurvey/videocard/

I pledge that on all academic work that I submit, I will neither give nor receive unauthorized aid, nor will I present another person's work as my own.

Dalton Murray