GPU Forecast and Analysis

Tyler Anderson

Bellevue University

DSC 680: Applied Data Science

Professor Catie Williams

October 14, 2020

Abstract

This project will discuss the Graphics Processing Unit (GPU) performance over time as well as predict what the performance of GPUs in the year 2030. For the past 10 years the growth of computer games has resulted in competitive development of Graphics Cards. I want to see if the competitive nature of the graphics card market pushes past Moors law or if Moors law is a limiting factor in the development of new Graphics Cards. To predict the performance of graphics cards in 2030 I will need to use regression models to get a good estimate. In order to focus on performance of GPUs I will need to specifically focus o 3 parameters. Core speed [MHz], Memory Bandwidth [GB/sec] and Memory size [MB or GB].

This dataset contains two CSV files: gpus.csv for Graphics Processing Units (GPUs), and cpus.csv for Central Processing Units (CPUs), for this project I will only be using the GPU dataset. Each table has its own list of unique entries, but the list of features includes clock speeds, maximum temperatures, display resolutions, power draws, number of threads, release dates, release prices, die size, virtualization support, and many other similar fields. With moors law stating the number of transistors in a dense integrated circuit (IC) doubles about every two years. I want to see how the memory size of these GPUs changes over time. I plan to see how moors law effects GPU releases and if in the year 2030 GPU memory capacity will change in line with moors law. I also want to analyze how performance over ratio evolve over time. How is the general computing power different? And are there any manufacturers that are known for some specific range of performance & price? I hope to see some interesting results here as NVidias new GPU line has had impressive performance boosts.

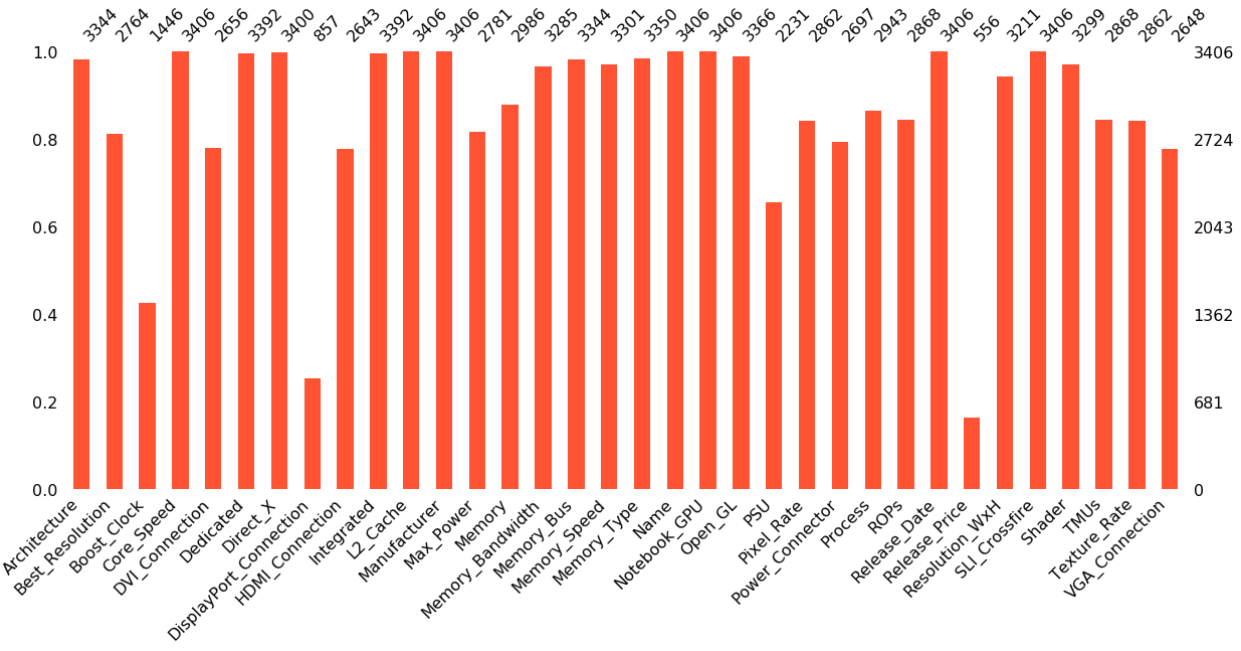
Meaning of Variables

For this section I will only focus on the 3 main variables I will be using: Core speed [MHz], Memory Bandwidth [GB/sec] and Memory size [MB or GB], Name, Manufacturer, Release Date and Best Resolution.

- **Core Speed [MHz]** - The frequency at which the GPU is running. This can be sort of compared to a CPU's operating frequency. "Speed" depends on numerous factors, architecture being one of them. It's not necessarily an apples-to-apples comparison to look at the core clock speed of an older GPU and a newer one (or cross-brand differences), but for sake of ease, greater core frequencies equates faster computing.
- **Memory Bandwidth [GB/sec]** - This is one of the single, most important aspects of graphics processors. Memory bandwidth determines your card's ability to utilize its onboard video RAM efficiently when under stress.
- **Memory Size [MB or GB]** - VRAM is the storage location for textures and 3D meshes. So, the more available, the more complex the scene can be, before it overflows to the standard RAM.
- **Name** – The name of the individual Graphics Cards.
- **Manufacturer –** The name of the company that makes the Graphics Cards**.**
- **Release Date –** The date the GPU was released.
- **Best Resolution –** The resolution the GPU was designed and marketed for.

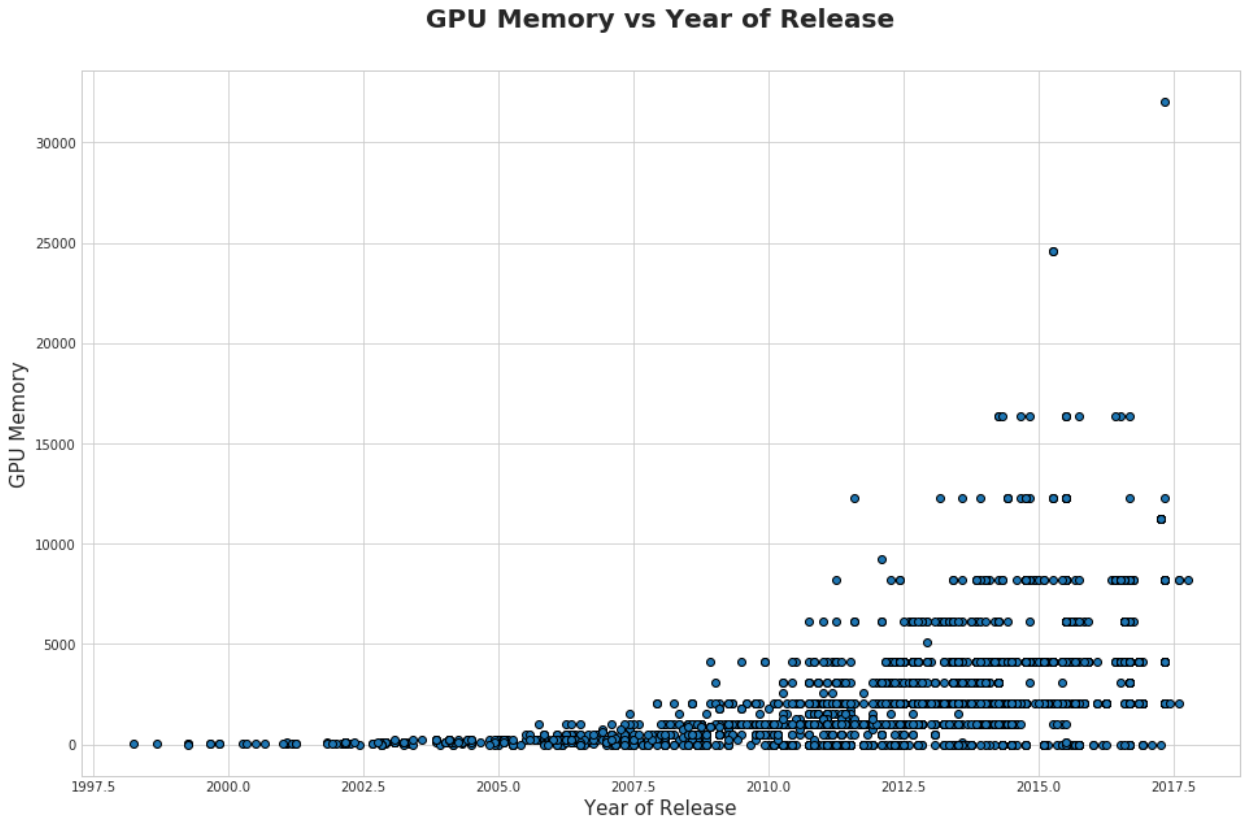
Analysis

My first step was to investigate the data and see of there was anything that stood out to me. Right away we can see that there is lots of undefined values. I used the “Missingno" package which is a useful and convenient tool in visualizing missing values in the dataset. There are only few columns that can by use as reliable source of information for our analysis. Most of columns are incomplete or in wrong format. We will perform data preprocessing to make some columns useful in our analysis.

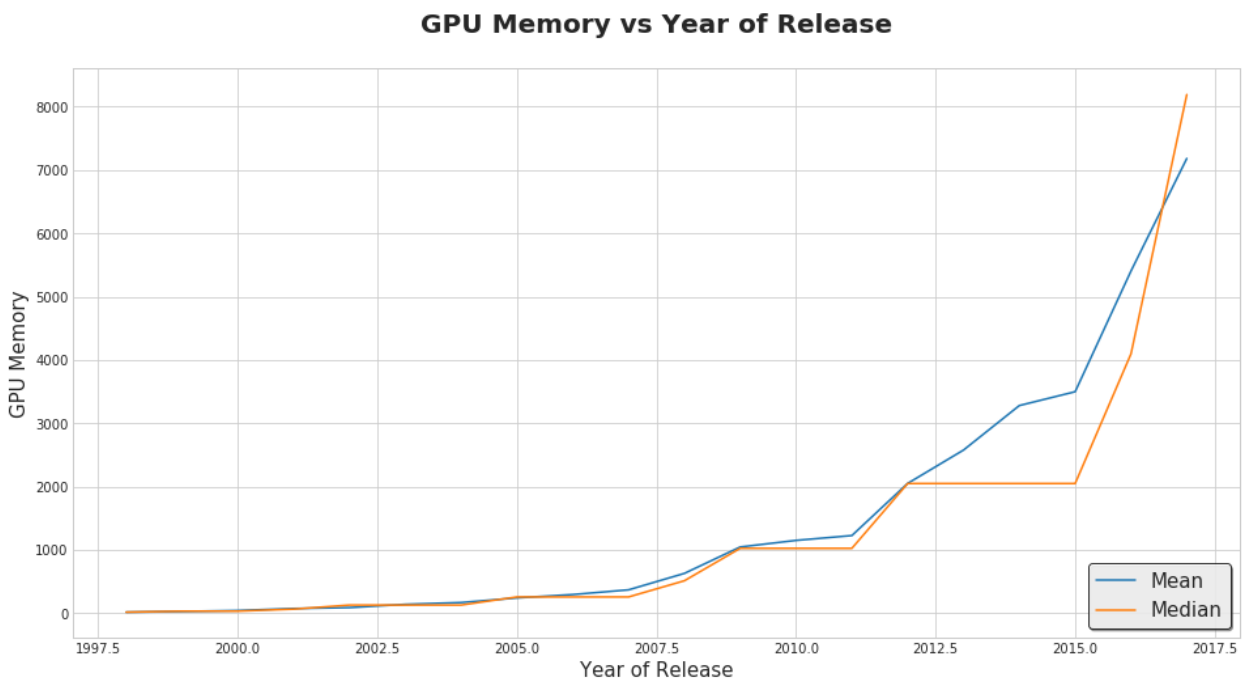


Based on our research and our main hypothesis for this project the columns we select will indicate the key data for our project. After selecting the data, we can use data preprocessing and feature engineering to process our columns into something we can use later on in the project.

We can also look at some visualizations to take a look at GPU memory to see how much the competition has increased over the last 10 years and how much improvement in GPU performance has been made since 2010.



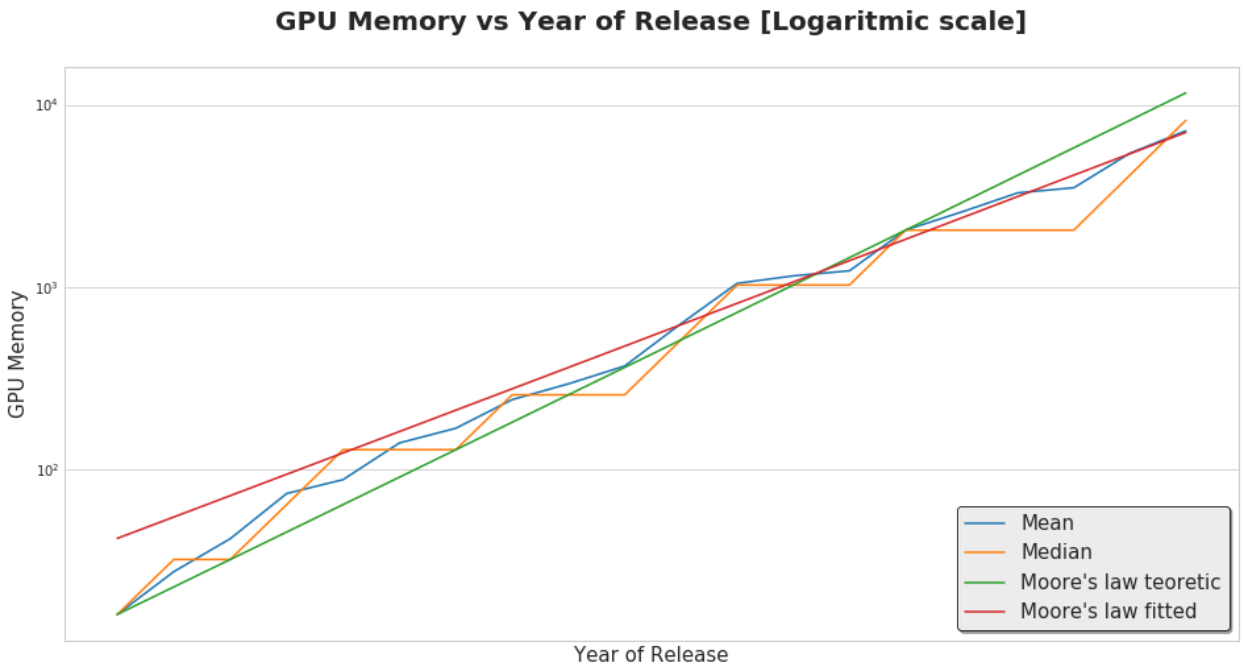
We can clearly see that since 2010 GPU performance has increased dramatically possibly due to increased competition or Moors law dictating the GPU performance. We can now calculate mean and median Memory Size of GPU for each year and create line plots to visualize the upward trend.



I wanted to see how the above plot would look like in a logarithmic scale. If Moors law is truly the driving force of the increased performance, we should see a relatively straight line. This function would represent memory size calculated using moors law considering the initial values would have this as the following formula. $f(x) = y_{min} 2^{\frac{x - x_{min}}{2}}$

Where:

- ymin - initial size of memory in MB,
- xmin - initial year of dataset.



We can see that the mean size of the GPUs memory tends to follow Moors law curve but not quite exactly. I used curve_fit to create a second curve from the same family (red line) that fits the dataset better.

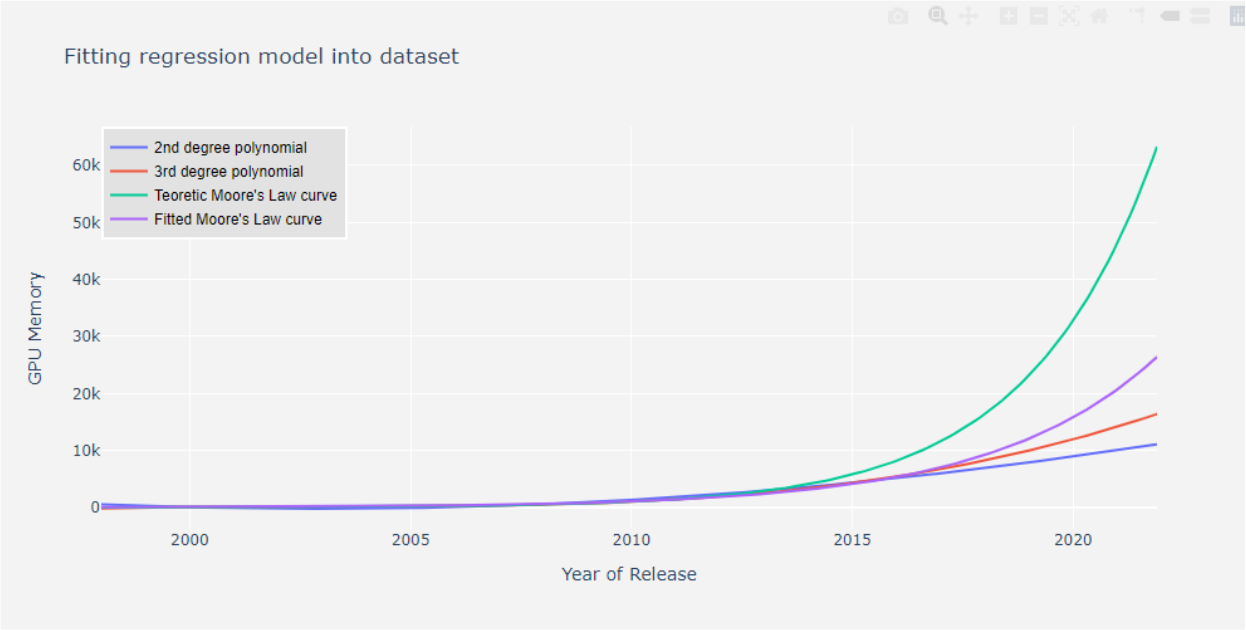
I used the Polynomial regression prediction model from sklearn and fitted it to the dataset. I then selected the best model against the moors law curve and got the following results:

```
from sklearn.metrics import r2_score

# 2nd degree curve
score = r2_score(y_pred_lin_reg_2, memory_arr_mean)
print("r2 of 2nd degree curve is equal " + str(round(score, 3)))
# 3rd degree curve
score = r2_score(y_pred_lin_reg_3, memory_arr_mean)
print("r2 of 3rd degree curve is equal " + str(round(score, 3)))
# Teoretic Moore's Law curve
score = r2_score(y_pred_moore_law_teoretic, memory_arr_mean)
print("r2 of Teoretic Moore's Law curve is equal " + str(round(score, 3)))
# Fitted Moore's Law curve
score = r2_score(y_pred_moore_law_fitted, memory_arr_mean)
print("r2 of Fitted Moore's Law curve is equal " + str(round(score, 3)))
```

```
r2 of 2nd degree curve is equal 0.947
r2 of 3rd degree curve is equal 0.985
r2 of Teoretic Moore's Law curve is equal 0.825
r2 of Fitted Moore's Law curve is equal 0.992
```

Based on the r2 scores I selected the Fitted moors law curve for predicting the GPUs mean memory size.



After selecting the model I can now predict the mean memory size of GPUs in 2030.

```
# Predicting GPUs mean memory size in 2030

memory_2030 = exponentialCurve(2030, *popt)
```

```
print("Predicted mean size of GPU memory in 2030 is " + str(round(int(memory_2030) / 1024, 2)) + " GB.")
```

```
Predicted mean size of GPU memory in 2030 is 229.25 GB.
```

Conclusion

In conclusion I can say I was wrong to think that GPU performance over time was the result of competition in the GPU market and was Moors law that was the main driver in the increased performance seen over the last 10 years. The competition between the two main companies in the GPU market, Nvidia and AMD, was actually the price not performance. The competition did not bring performance, but it did make GPUs more affordable for consumers. The main difference here is that the technology has changed in the GPU market but the raw memory of GPUs cannot escape Moors law so in essence GPUs in the future will see higher memory size but the technology of how to put pixels on your screen will change to be more efficient and that is where the GPU market competition will increase. I was surprised to see the predicted size for memory in 2030 be 229 GB that seems so large for todays standards however if I look back to the GPU sizes from 2010, I should not be as surprised.

Questions

1. Does Moors law also effect the CPU dataset?
2. What kind of technology makes a difference in GPU performance other than memory size?
3. How do the combination with other computer parts affect the performance of the GPU? RAM, Motherboard, CPU
4. Where there any other parameters that were better suited for this analysis than GPU memory size?
5. If there were no competition in the GPU market would we see a decrease in GPU performance compared to what we see now?
6. Will we expect to see GPU memory sizes reach the levels predicted in 2030?

7. Will GPUs depend on the current demand in games/software being developed in its performance?
8. How did you conclude to use Polynomial regression in your analysis?
9. Do you think GPU architecture will change dramatically in 2030 to accommodate the predicted GPU memory size?
10. What kind of resolutions do you think will be affordable and practical in 2030 to warrant the predicted size of GPUs?

Works Cited

Andronico, M., & Moore-Colyer, R. (2020, Oct). *Nvidia GeForce RTX 3080 release date, price, specs and pre-order.* Retrieved from tomsguide: https://www.tomsguide.com/news/nvidia-geforce-rtx-3080-revealed-release-date-price-specs-and-pre-order

askpython. (2020). *Polynomial Regression in Python – Complete Implementation in Python.* Retrieved from askpython: https://www.askpython.com/python/examples/polynomial-regression-in-python

Baldini, I., Fink, S. J., & Altman, E. R. (2014, Dec). *Predicting GPU Performance from CPU Runs Using Machine Learning.* Retrieved from researchgate: https://www.researchgate.net/publication/292845940_Predicting_GPU_Performance_from_CPU_Runs_Using_Machine_Learning

Bronshtein, A. (2017, May 8). *Simple and Multiple Linear Regression in Python.* Retrieved from towardsdatascience: https://towardsdatascience.com/simple-and-multiple-linear-regression-in-python-c928425168f9

Burke, S. (2012, Jan 27). *GPU Dictionary: Understanding GPU & Video Card Specs.* Retrieved from gamersnexus: https://www.gamersnexus.net/guides/717-gpu-dictionary-understanding-gpu-video-card-specs

Burke, S. (2020, Sept 16). *NVIDIA GeForce RTX 3080 Founders Edition Review & GPU Benchmarks.* Retrieved from gamersnexus: https://www.gamersnexus.net/hwreviews/3618-nvidia-rtx-3080-founders-edition-review-benchmarks

Skalski, P. (2017). *Using Regression to predict GPUs of the future.* Retrieved from kaggle: https://www.kaggle.com/skalskip/using-regression-to-predict-gpus-of-the-future

w3schools. (2020). *Machine Learning - Polynomial Regression.* Retrieved from w3schools: https://www.w3schools.com/Python/python_ml_polynomial_regression.asp

Walker, G. (2018). *GPU EDA.* Retrieved from kaggle: https://www.kaggle.com/gw00207/gpu-eda