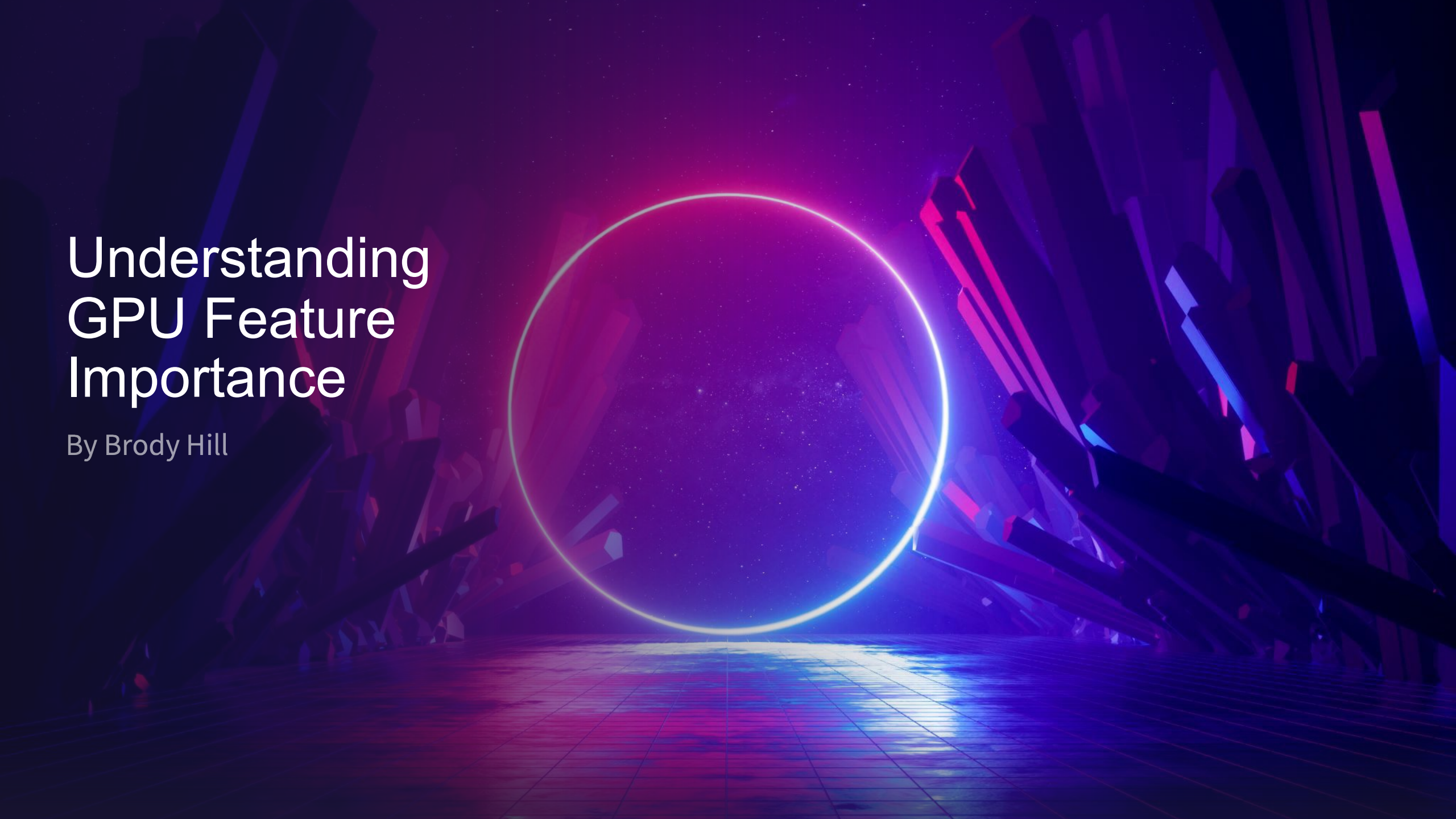# Understanding GPU Feature Importance

By Brody Hill

The Problem:

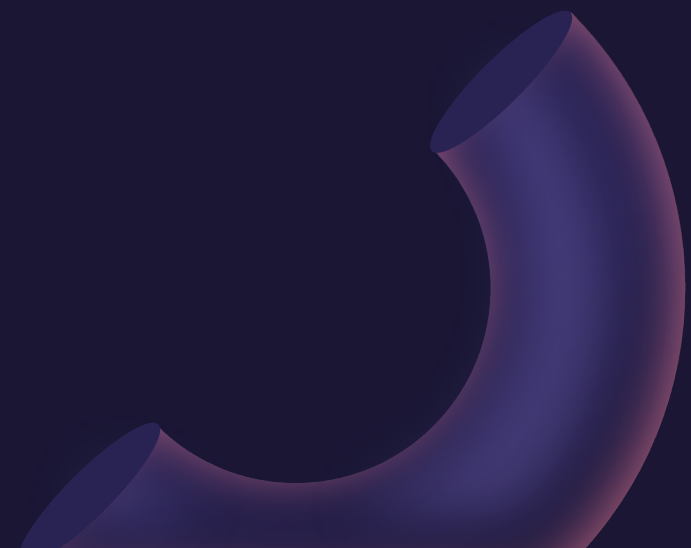You're intel and you've just released your first line of GPUs.

AMD and NVIDIAs cards are crushing your best cards in rasterization performance. You need to understand which GPU features most impact performance so you can invest accordingly.

The Solution:

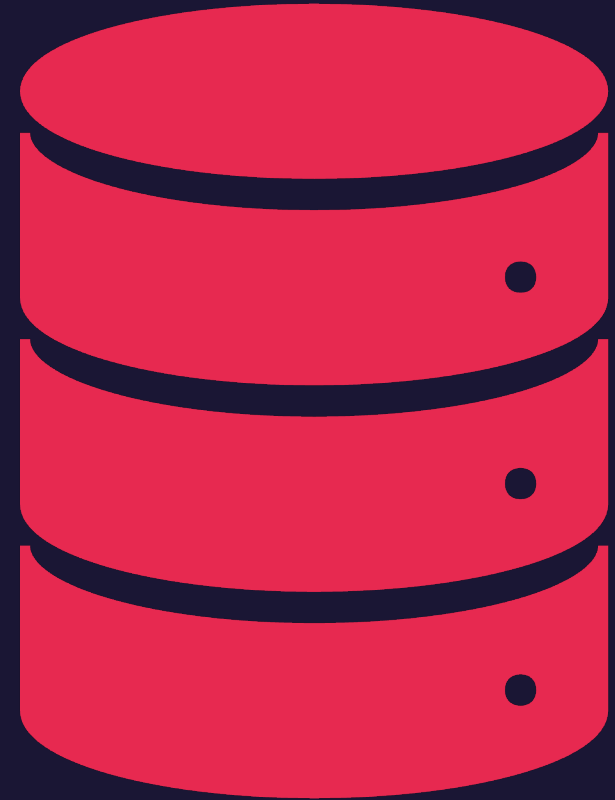Invest heavily into cards with more texture mapping units and render output units.

The Data:

I used two datasets here.

Dataset 1: GPU Hardware Features

Dataset 2: GPU 3DMark Performance

Data Wrangling:

After merging my two datasets together on the GPU column, I had 1088 rows and 24 columns.

After examining and cleaning I had 786 rows and 17 columns.

What actions did I take to get there?

- ✓ Dropped all duplicate GPUs, there were 134.

- ✓ Dropped columns that had over 20% of their data missing, there were seven.

- ✓ Created histograms and checked for outliers, there were not any suspicious values.

- ✓ Dropped any remaining rows that still contained missing values

Exploratory Data Analysis Part 1:
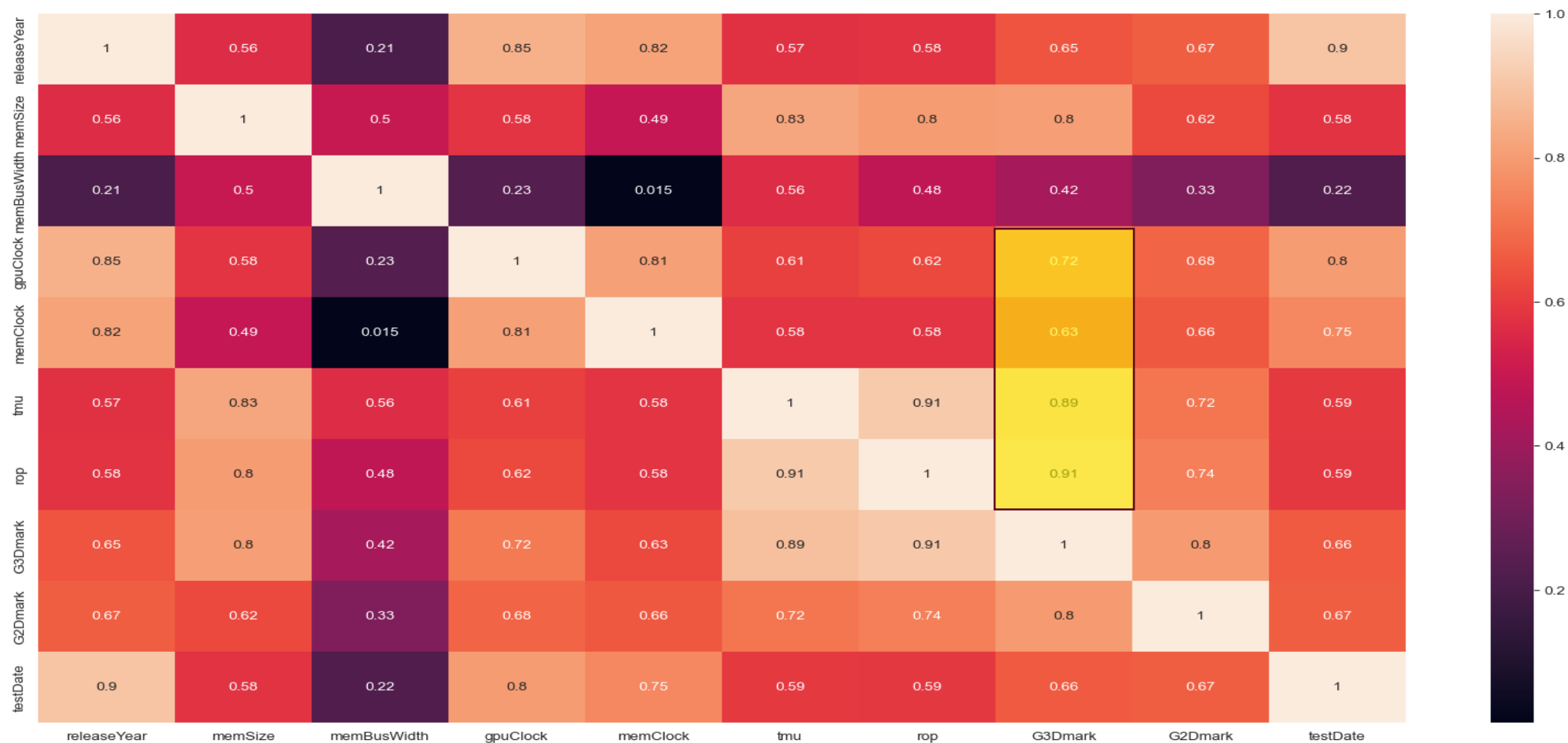
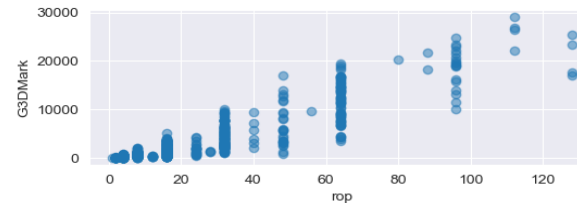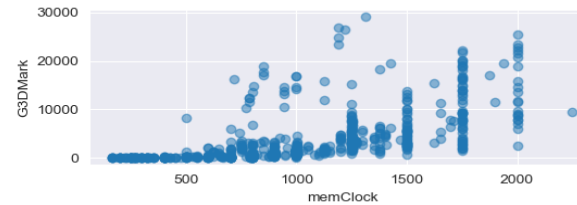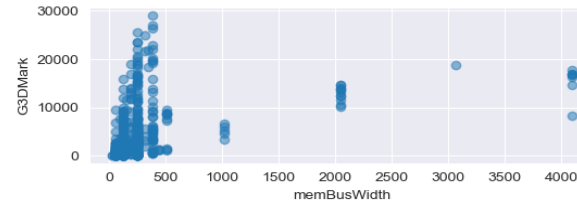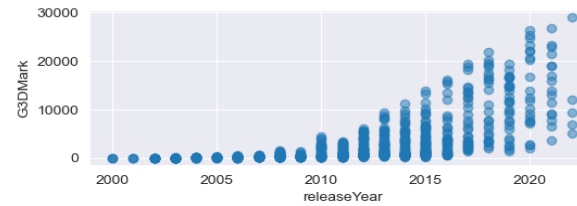# Exploratory Data Analysis Part 2:
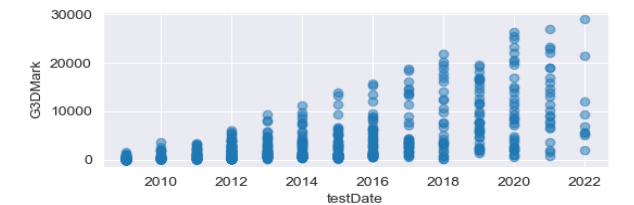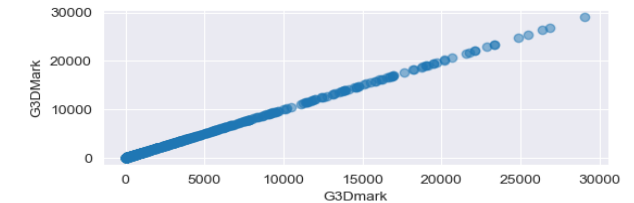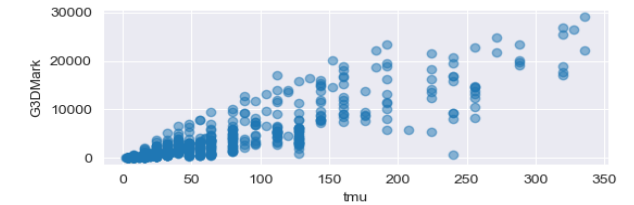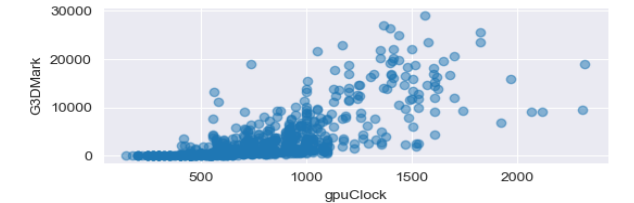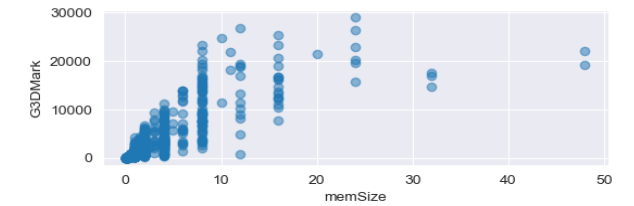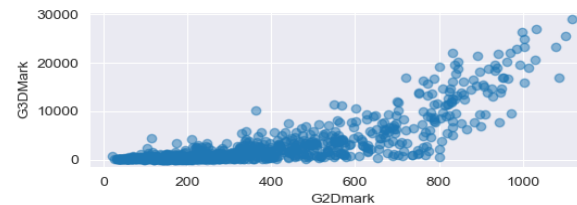
Model Results:

	Algo	RMSE	MAX ERROR
0	Basic RF - Dummy Encoded - Not Scaled	775.443139	4287.890000
1	Basic RF Random Search CV - Dummy Enc - Not Sc...	817.014045	5142.237000
2	Basic RF - Dummy Encoded - Standard Scaled	754.970010	4636.120000
3	Basic RF - No Cat Features - Not Scaled	791.746025	4703.260000
4	Basic RF - No Cat Features - Standard Scaled	724.610854	3210.330000
5	Basic RF - Count Encoded - Not Scaled	756.819373	3771.060000
6	Basic RF - Count Encoded - Standard Scaled	778.336998	5102.500000
7	Basic GB - Count Encoded- Not Scaled	757.826355	3702.212734
8	Tuned GB - Count Encoded - Not Scaled	709.018359	3463.126447

Feature Importance:
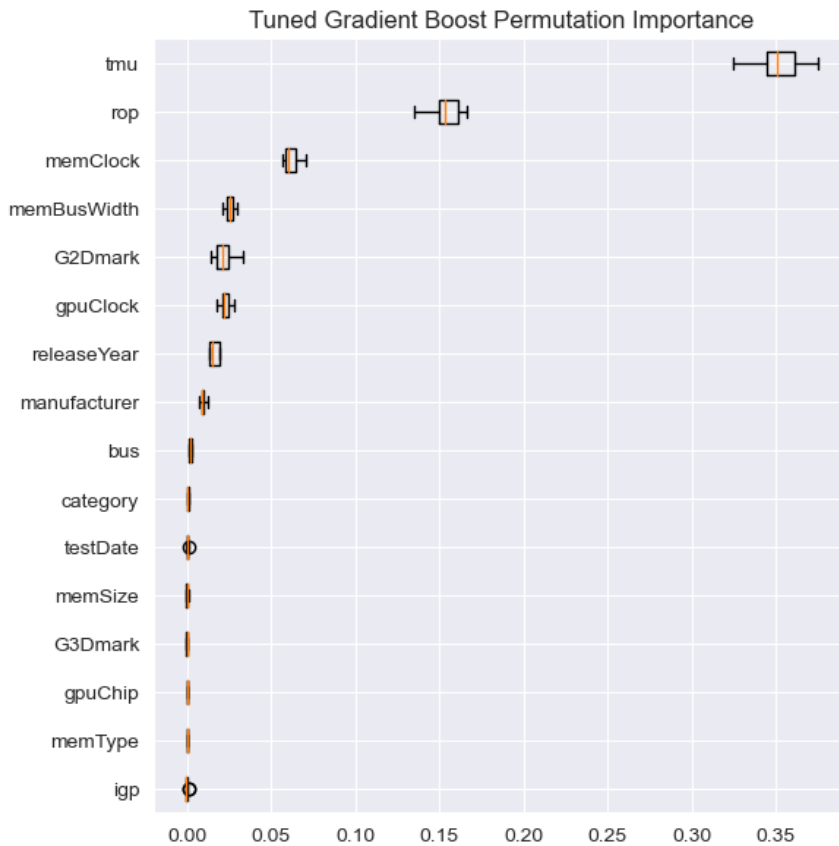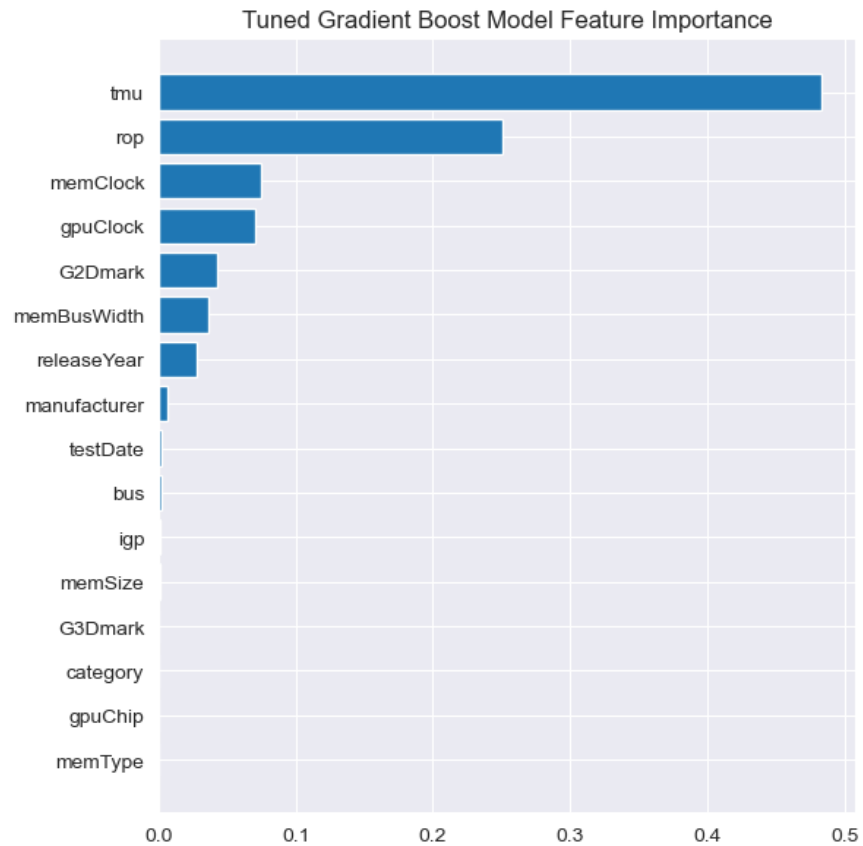
Takeaways:

✓ Gradient Boosted model tuned on the learning rate and number of estimators parameters provided the best results.

✓ Feature importance aligned well with initial heatmaps. TMUs and ROPs are the best predictors of 3DMark performance.

✓ Invest in cards with more TMUs and ROPs, while keeping clock speeds in mind.

Future Research:

✓Gather more GPU data and repeat.

✓Find CPU data and combine with GPU data for a more comprehensive view.

✓Focus only on newer hardware. (2020+)

Questions?