Predicting the price of a gaming PC

This project will scrape PC component data from Amazon.com. The price of a gaming PC is heavily dependent on its components, but certain components will inevitably influence the price more than others. My linear regression model will aim to quantify how much the market values changes in the technical specifications of components.

This model would be useful to someone who wants to price their PC optimally on Amazon.com based on its components. Ideally, this model will give someone an idea of how much they should list their PC for so it doesn't sit unsold for months, and simultaneously doesn't get sold for under market value.

An **MVP** would provide price prediction based on a few key variables, including processor speed, RAM capacity, and whether the graphics card is an Nvidia or AMD branded card.

Variable	**Variable type**
Processor brand (AMD or Intel)	Categorical binary
Processor speed	Continuous
Processor count	Quantitative (discrete)
RAM capacity	Quantitative (discrete)
Graphics card type (onboard or dedicated)	Categorical binary
Graphics card brand (AMD or Nvidia)	Categorical binary
Graphics RAM	Quantitative (discrete)
Number USB 2.0 ports	Quantitative (discrete)
Number USB 3.0 ports	Quantitative (discrete)
Hard disk (solid state or HDD)	Categorical binary
Hard disk capacity	Continuous
Hard drive RPM (if HDD)	Continuous
Number customer reviews	Quantitative (discrete)
Average customer review	Continuous
Price	**Continuous**

Anticipated hurdles

- Getting Selenium to play nice with Amazon's website
- Some variables might be present in some computers, but not others. For example, one computer might have only USB 2.0 while others might have only USB 3.0. What to do with nulls will be an important question in this project.
- Amazon has a reputation for being difficult to scrape. I am able to make single requests using a proxy no problem, but I have concerns about scaling to 1000+ queries.