**Data Science Final Project : Getting to the Top!**

Problem Statement

Our company products are not ranked highest on Amazon’s retail website despite offering the most competitive price. This will eventually result in the company’s revenue decreasing over time.

Hypothesis/Assumptions

Amazon uses various metrics uses to award a vendor the coveted ‘Buy’ box, and price is known to be a major factor. Background research showed that sales rank and customer reviews are known contributing factors to a Vendor’s ranking on the website.

We will have to narrow down to a few key features to simply the analytical process.

Goal/Success Metrics

Identify which are the major factors affecting a vendor’s product ranking on amazon using Python. This will help us better understand how Amazon’s A9 algorithm works in real life.

Do cost-benefit analysis to determine whether it is worth joining Amazon’s ‘Fulfilled by Amazon’ program to improve our company’s product rankings.

Create a model that will predict a product’s ranking based on a certain feature set.

Risk/Limitations

There may be hidden features that Amazon has kept hidden in their proprietary A9 algorithm.

The dataset may be too small to be representative of the entire ranked population. More data needs to be collected on a particular product subcategory.

Datasets Required

Propublica’s scrapped Amazon pricing data.

In general, we would need these data to create the model:

* Vendor’s pricing
* Product Ranking
* Product sales numbers (if available)
* Date of purchase
* Vendor Name