**Project Pitch - Hsin**

**Dataset**

I would like to use Amazon Pricing data. The dataset contains 6973 rows of primarily product + vendor with pricing information. The metadata includes the Buy Box vendor for given product which shows you which vendor was recommended at the time the data was collected.

**Problem Statements**

1. Amazon today is without a doubt the most influential and dominant online shopping site in the world. Its recommender system almost drives today’s customers’ buying behavior. I am interested in knowing that when Amazon collects fees from the vendors who sell their products on Amazon, how does that play in terms of Amazon promoting their products? I would like to examine those products which Amazon sells and compete with its vendors and understand Amazon’s pricing strategy. The dataset contains the information for a given product, what is the Amazon’s recommended vendor, aka the buy box. The reality is just because the price is cheaper or the rank is better, a vendor does not necessarily get to be in the Buy Box.
2. Today people do a lot of online shopping, and therefore shipping cost becomes an important factor when it comes to making the purchase decision. I would like to also examine how sipping cost affect the rank - How frequent do we see that when shipping is taken into consideration, the rank gets changed due to higher total price (unit price + shipping)?
3. Is the number of vendors affecting how Amazon prices its products? Where in terms of rank is Amazon normally positioned if there are multiple sellers? How frequent do we see Amazon has best price among the vendors? If it’s fulfilled and shipped by Amazon, do we normally see less competitive Amazon.com offer because Amazon gets paid more by those vendors?
4. Given what we know about the Amazon algorithm, I am interested in building a model that will predict how likely a vendor can show up in the buy box in Amazon’s recommender system given the features and thus how we can potentially increase the chance to make to the buy box.