

Amazon:

Using ProPublic's data that examines Amazon's shopping algorithm, I will be analyzing the cost of products that were sold by Amazon versus vendors, the displayed price versus the total cost (plus shipping), and analyzing "Fulfilled by Amazon" to determine if it actually saves consumer's money.

Hypothesis: Amazon uses "dynamic pricing" by increasing the retail cost of their products (surge pricing), which in turn online buyers do not actually receive a cheaper price through the site. This is particularly true for "Fulfilled by Amazon" products.

Approach:

- Predicting "Fulfilled by Amazon" pricing by evaluating top products within a time series dataset
 - Price increases based on supply and demand, more popular items see an uptick in price
 - Test theory by randomly selecting 5 products from the best-selling products
- Predicting "Fulfilled by Amazon" products are more expensive than retailers
 - Do a train/test split to subsample the time series data to create a model and help accurately predict Amazon's algorithmic pricing
- Predicting a target variable
 - If we pass X through number of columns, are we able to predict the product ranking?

Combine data results to help determine if Amazon is actually boosting the price of their own products to prove the hypothesis true

Potential Problems:

- Overfitting the model
- There are various holes in the web scraped data