Amazon Price War
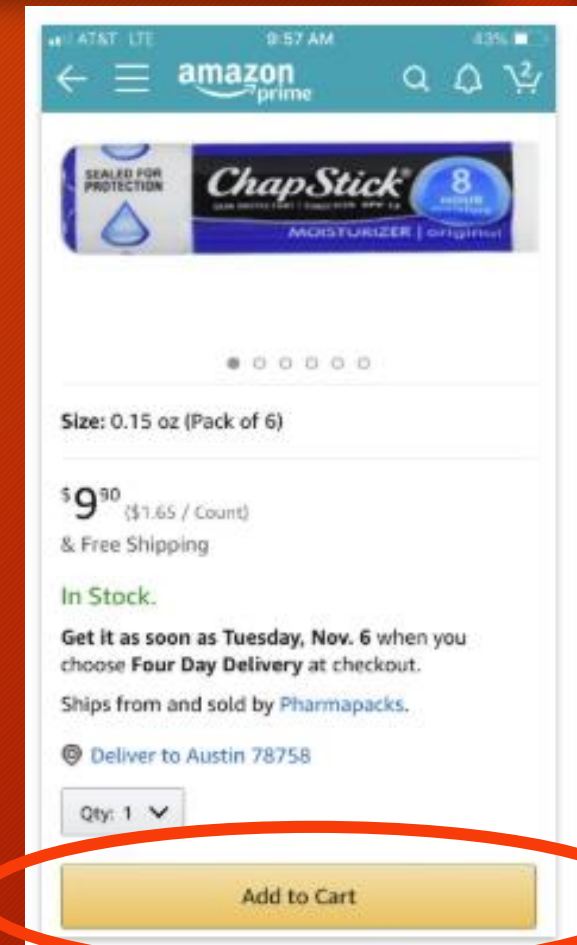
General Assemble: Data Science Remote

Hypothesis, Assumptions, and Goals

How to increase the chance to win the Buy Box among multiple type of Amazon Reseller?

- Lower the price?

- Higher the ranking?

- Use Amazon platform?

82% of Amazon sales go through the **Buy Box**, and the percentage is even higher for mobile purchases.

The Dataset

Data: Amazon Ranking Analysis during summer 2016

ProPublica reporters examined Amazon's shopping algorithm. We scraped data from the company's website to examine listings for 250 best-selling products across a wide range of categories, from electronics to household supplies, over a period of several weeks during summer 2016. We compared pricing and shipping costs for products offered by multiple vendors, including those sold by Amazon and sellers in the "Fulfilled by Amazon" program. In total, we examined 6,973 vendor listings.
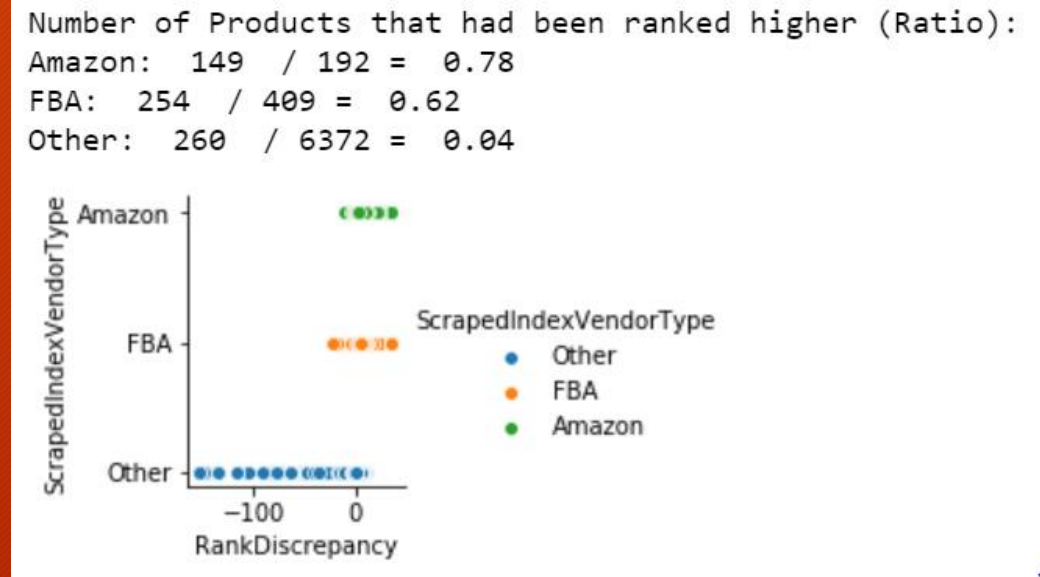
Approach

There is no linear relationship between the ranking and the pricing.
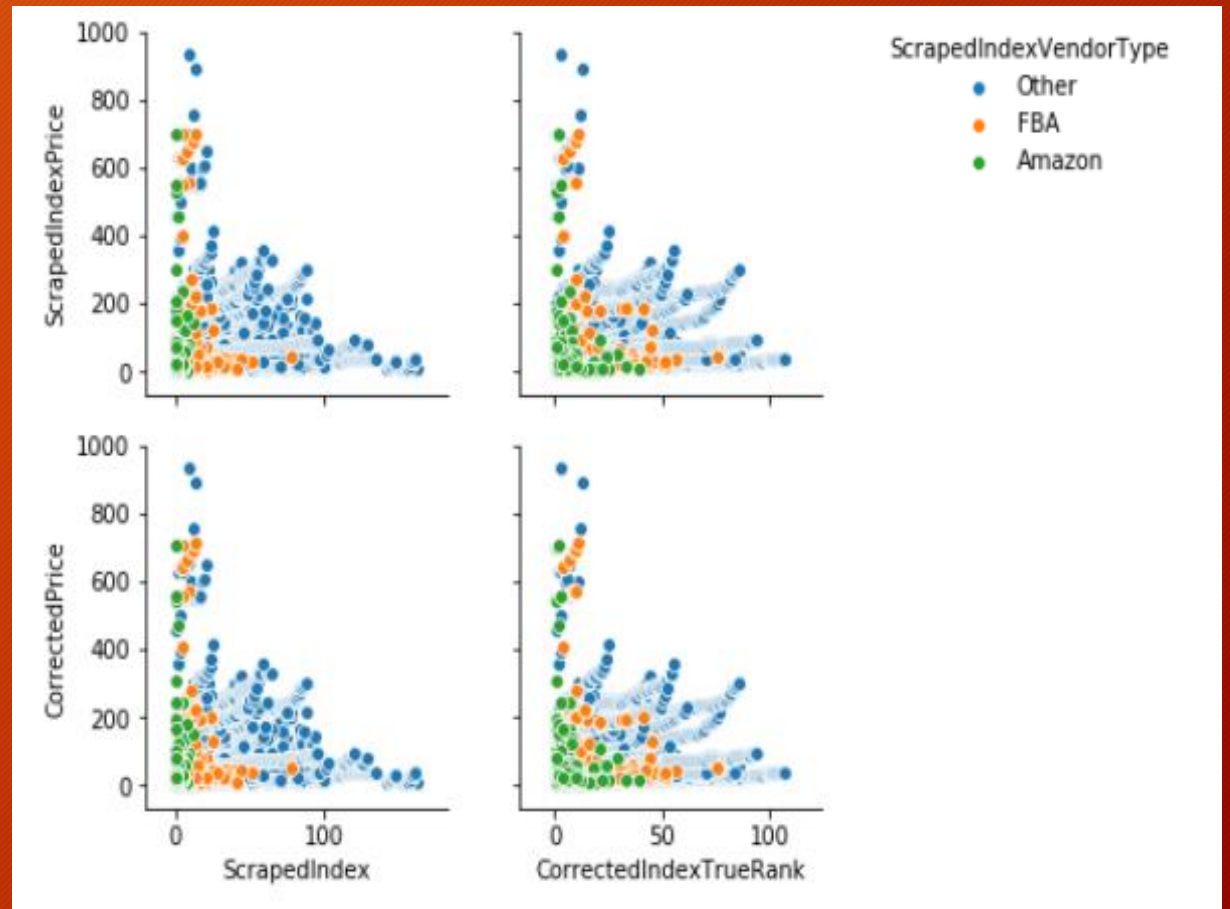
Approach

The seller who use Amazon Platform has a higher chance to win the buy box.

- The items that lacked a shipping price were sold by Amazon or sellers in the "Fultilled by Amazon" program

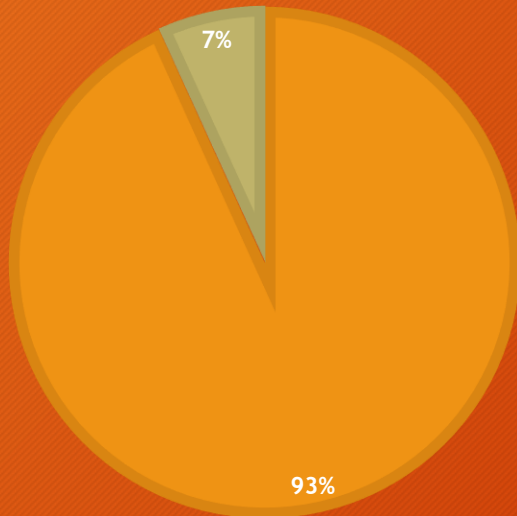- Amazon-linked products higher rankings in more than 80 percent of cases.

Number of Products that had been ranked higher (Ratio):
Amazon: 149 / 192 = 0.78
FBA: 254 / 409 = 0.62
Other: 260 / 6372 = 0.04

Approach

Amazon = 179/192 = 93%   FBA = 58 /409 = 14%   Others = 105 / 6372 = 1%

AMAZON

■ Win Buy Box   ■ Didn't Win Buy Box

7%

93%

FBA

■ Win Buy Box   ■ Didn't Win Buy Box

14%

86%
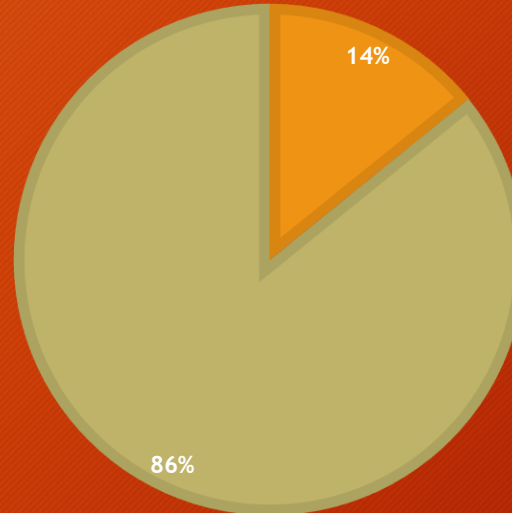
OTHERS

■ Win Buy Box   ■ Didn't Win Buy Box

2%

98%
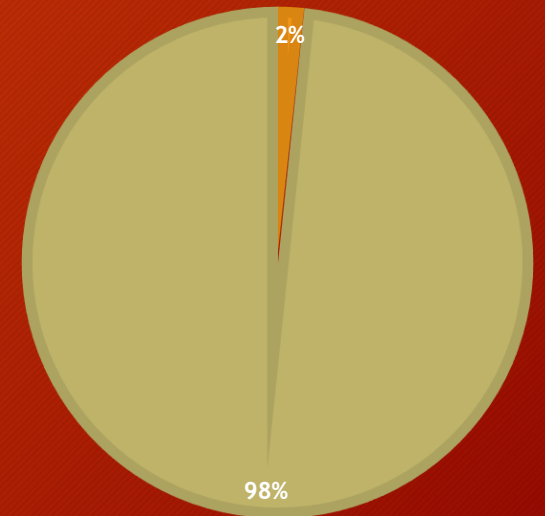
Feature Importance

	feature	importance
4	CorrectedPrice	0.007093
2	CorrectedIndexTrueRank	0.009125
6	BBCorrectedPrice	0.010269
3	ScrapedIndexPrice	0.012762
8	BBVendorNum	0.052566
5	BBScrapedPrice	0.055509
0	ScrapedIndex	0.080012
1	ScrapedIndexTrueRank	0.167252
7	VendorNum	0.605412

Models

- The data source has been split to training set (70%) and testing set (30%).

- Tried out four different models, including:
  - Logistic Regression
  - K-Nearest Neighbors
  - Decision Tree Classifier
  - Random Forest

Results

- Logistic Regression
  - Accuracy Score (Train): 98.07%
  - Accuracy Score (Test): 98.04%

- Decision Tree (depth=6)
  - Accuracy Score (Train): 98.05%
  - Accuracy Score (Test): 98.27%

- K-Nearest Neighbors (neighbors=7)
  - Accuracy Score (Train): 97.86%
  - Accuracy Score (Test): 96.41%

- Random Forest (features=3)
  - Mean Accuracy Score: 80.76%

Recommendations and Next Steps

Partner with Amazon

- Hands off the responsibilities to pick, package and ship every order
- Product can be deliver on time (within 2 days) and without defect

Other features to look at

- Consumers' comments / rating
- Shipping Time