IDS-702 Final Project: StockX Sneaker Analysis

Rashaad Ratliff-Brown

12/12/2021

Github Repository: https://github.com/rmratliffbrown/final-stats-sneakers

Summary:

Inferential questions evaluted in this analysis all surround the factors that impact the resale (after retail) value of a sneaker. Multiple Linear Regression was used to evalute a retrieved dataset to seek insight, support, and answers to our inferential questions. During the process of the analysis it was clear that sneaker style and colorway had a statistically significant impact on a sneaker's resale value.

Introduction:

In recent years sneaker culture has become more popular around the world. Sneaker buyers (sneakerheads) take to the most popular mobile applications and websites to obtain the newly released and rare sneakers. Every year it becomes harder to obtain the most conveted sneakers due to the increasing the demand and hype. Sneaker retailers also only offer a limited supply of the sneaker making them that much harder to obtain. Usually only the most dedicated are able to walk away successful when attempting to purchase from popular sneakers stores and platforms. Many of the purchasers immediately choose to resell their pair(s) while others desire to simply add them to a collection. The resale market allows the buyer to obtain a profit usually with the goal of purchasing more sneakers.

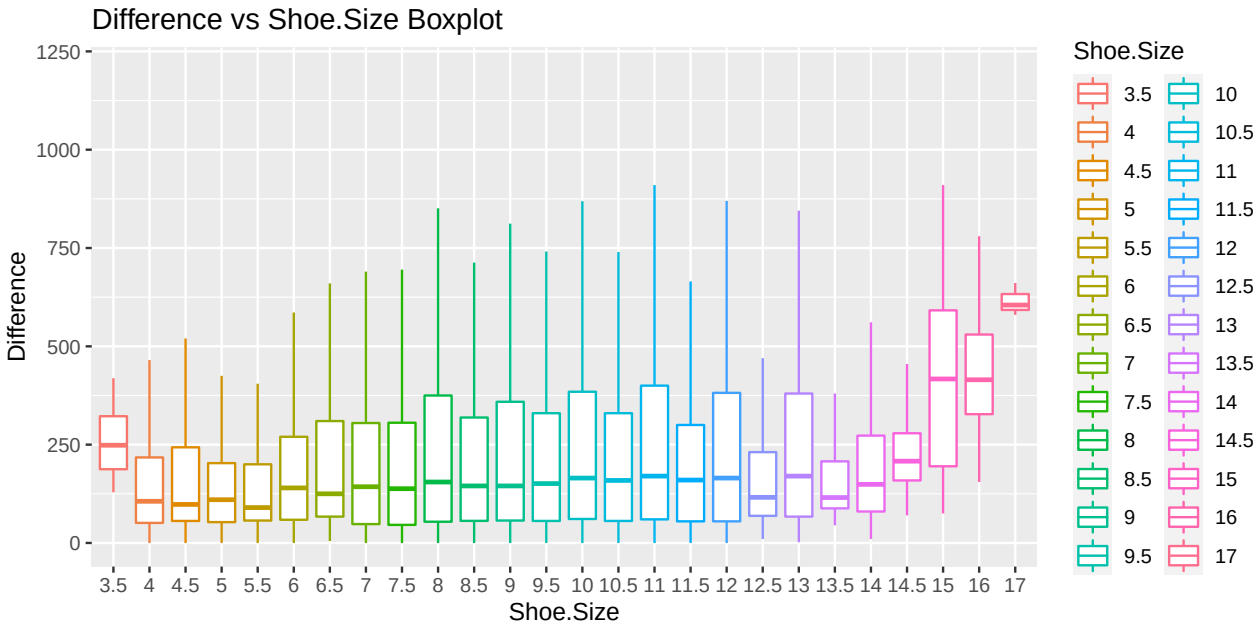
Websites like StockX, FlightClub, GOAT, and other small sneaker businesses all safer transactions for sneakerheads and simply resellers. They function as a middleman to help keep both ends safe, honest, and satisfied. Their job is to also confirm that each pair of sneakers are within proper conditions to sell or their state is properly indicated for buyers. StockX is arguably the most popular platform due to its process of verifying the authenticity of products sold on their marketplace. This avoids purchasers obtaining damaged or falsified pairs of sneakers. Each item is tagged with a signature pendant after several evalutions both certified StockX employees. StockX allows customers to post the sneakers they desire to sale at their desired price and buyers are able to bid or make an offer for the reseller to consider.

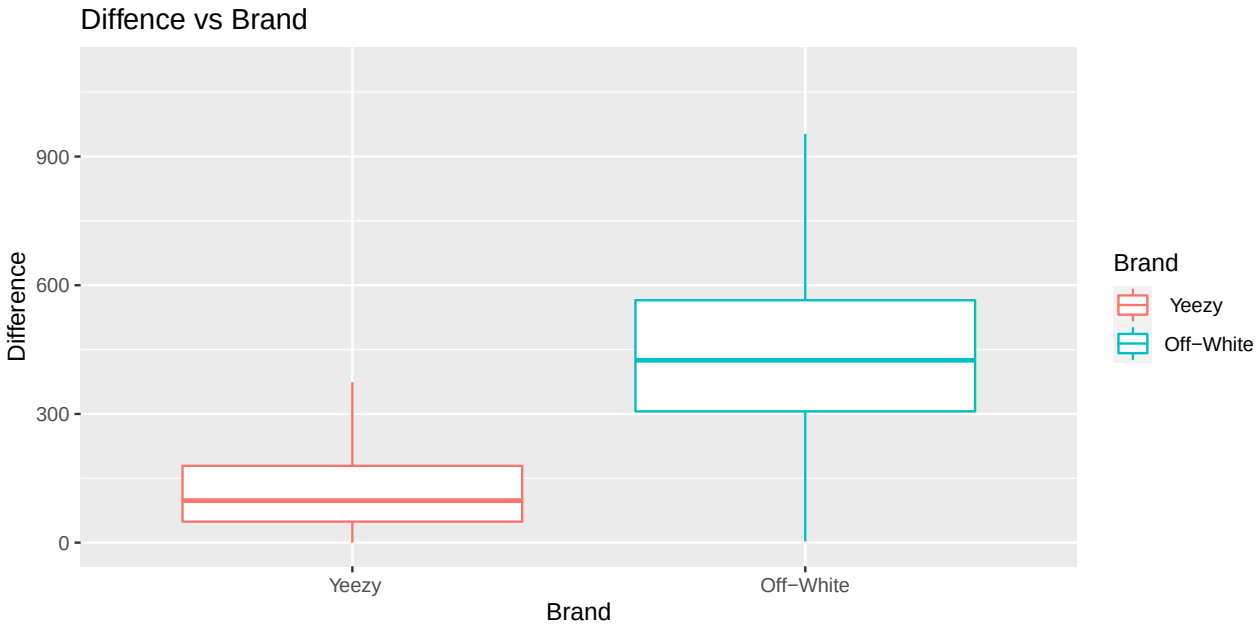
In 2019, StockX released the dataset as a part of a 2019 Data Challenge. The dataset consists of both Yeezy and Off-White sneakers released between Sept 2017 and Feb 2019, which will be explored during this analysis to find answers and support for the following inferential questions:

- What factors have a major influence on the Sale Price (resale) of a sneaker?
- Which of Brand yielded the highest grossing sneakers during resale?
- Which Buyer Regions seem to obtain the largest volume of Yeezy/Off-White sneakers?

Data:

An exploratory data analysis (EDA) began by developing numerous plots of predictor variables versus the response variable. Our response variable is a variable called Difference. It is the result of subtracting the retail price (Retail.Price) from the sale price (Sale.Price), which is the transaction price when obtained from StockX. Predictor variables used during EDA were: **Type**, **Brand**, **Sneaker.Name**, **Colorway**, **Shoe.Size**, and the variable indicating the number of days since the initial release **Since.Release**, which is the result of subtracting the order date (Order.Date) from the release date (Release.Date).





Two of the most insightful plots during EDA happened to be boxplots of Difference vs Brand and Difference vs Shoe.Size. These can clearly be seen above. They are insightful as they shed light on information and substance to support answers to our inferential questions. From the Difference vs Brand chart we can clearly see that sizes 15 to 17 had on average a higher Difference than smaller sizes. We can conclude that this may be due to a smaller supply of these sizes produced as these sizes had a lower transaction count. During EDA it was also discovered that California and New York are the two buyer regions with the highest number of transctions for Yeezy/Off-White sneakers. New York with 16,525 (16.5%) transactions and California with 19,349 (19.4%). These two states happen to also contain the most popular sneaker stores, which plays a factor in sneaker culture reach. In the Difference vs Brand plot above, we can clearly see that Off-White managed to produce sneakers with an average higher resale value than that of Yeezy sneakers.

Model:

	Estimate	Std. Error	t value	Pr(>|t|)

	Estimate	Std. Error	t value	Pr(>|t|)
(Intercept)	829.4543375	2.3625561	351.08345	0
TypeNike-Air-Force-1-Low	-495.5307527	4.2801796	-115.77336	0
TypeNike-Air-Max-90	-399.5942022	4.6295464	-86.31390	0
TypeNike-Air-Max-97	-304.4997623	5.3238615	-57.19528	0
TypeNike-Air-Presto	-238.6212533	3.5817717	-66.62101	0
TypeNike-Air-VaporMax	-443.2865608	3.8490956	-115.16642	0
TypeNike-Blazer-Mid	-362.3944372	3.7836674	-95.77862	0
TypeNike-React-Hyperdunk-2017-Flyknit	-548.0163354	8.4323277	-64.98992	0
TypeNike-Zoom-Fly	-677.9959521	4.0093797	-169.10245	0
TypeNike-Zoom-Fly-Mercurial	-715.2014539	5.4713301	-130.71802	0
TypeYeezy-Boost-350-V2	-716.2502638	2.4905292	-287.58959	0
Since.Release	0.1163864	0.0025719	45.25341	0

The multiple linear regression model can be translated as: a sneaker's hype value (difference) is 829.45 dollars when all of the predictor variables are set equal to zero. Price decreases by 495.53 dollars given the type is Nike Air Force 1 Low. Price decreases by 399.59 dollars given the type is Nike Air Max 90. Price decreases by 304.49 dollars given the type is Nike Air Max 97. Price decreases by 238.62 dollars given the type is Nike Air Presto. Price decreases by 443.28 dollars given the type is Nike Vapor Max. Price decreases by 362.39 dollars given the type is Nike Blazer Mid. Price decreases by 548.01 dollars given the type is React Hyperdunk 2017 Flyknit. Price decreases by 677.99 dollars given the type is Nike Zoom Fly. Price decreases by 715.20 dollars given the type is Nike Zoom Fly Mercurial. Price decreases by 716.25 dollars given the type is Yeezy Boost 350 V2. Price increases by 0.12 dollars given each day after the release the sneakers are purchased.

The model was obtained utilizing both a forward and a both-direction step-wise function. The forward step-wise function began with an intercept only model and a model with all predictors, then both were evaluated only stepping forward. An anova table allowed for evaluting the AIC with each step. Once the forward step-wise function was complete, the same steps occurred to obtain the both-direction step-wise function. The AIC for both models ended up being the same. The BIC was also evaluated for both models proving that a selection of either model would be okay. This allowed me to come to the conclusion that both models work, so a summary was used to evaluate the coefficients from the both-direction step-wise and the most significant predictors were selected. Those predictors happened to be the 10 sneaker types and since release (Since.Release).

The model was assessed for linearity assumptions and normality. There were no clear violations of the assumptions indicating that the model is indeed insightful. The desired R^2 value was not as has as desired, but is fine for the case of this analysis. The model contained outliers, leverage, and influential points that were evaluted to determine removal. A f-test was used to help assess the models for the best model selection. There is no clear presence of multicollinearity after evaluting the VIF. Aside from changing the listed variables above there were no forms of transformation to the model that took place. An attempt was made via evaluting the log() of the final model but it did not yield any improved model quality or insights. The model is already well laid along a Normal Q-Q plot so this supports there not being a need to implement a log() as a there is no clear skewness of the data. The overall distribution is normal and passes all normality assumptions.

	2.5 %	97.5 %
(Intercept)	824.8237565	834.0849185
TypeNike-Air-Force-1-Low	-503.9198521	-487.1416533
TypeNike-Air-Max-90	-408.6680564	-390.5203481
TypeNike-Air-Max-97	-314.9344654	-294.0650592
TypeNike-Air-Presto	-245.6414817	-231.6010248

	2.5 %	97.5 %
TypeNike-Air-VaporMax	-450.8307410	-435.7423807
TypeNike-Blazer-Mid	-369.8103788	-354.9784956
TypeNike-React-Hyperdunk-2017-Flyknit	-564.5435942	-531.4890766
TypeNike-Zoom-Fly	-685.8542872	-670.1376171
TypeNike-Zoom-Fly-Mercurial	-725.9251937	-704.4777141
TypeYeezy-Boost-350-V2	-721.1316704	-711.3688572
Since.Release	0.1113456	0.1214273

A confidence interval yields resale value is 95% likely to be between (824.82, 834.08) dollars based on sneaker type and number of days since between release and transaction.

Conclusion:

A multiple linear regression model was used for an observational look at factors influencing sneaker resale value. Positive associations were identified for number of days since release (Since.Release) and type (Type). This analysis also allows us to conclude that it is possible to evaluate the major factors that impact the resale value of a sneaker. The largest limitation appears to be access to other types/styles of sneakers. Also information regarding the impact of sneaker prices when an artist passes away. Virgil Abloh is no longer with us and prices for sneakers on stockx have tripled of quadrupled since the news. Other factors to be evaluted are if/when a celebrity is scene wearing a specific sneaker in public. In my experience this tends to drive up prices, but I cannot speak on it being statistically significant without having access to data to evaluate. This is clearly just an assumption regarding other factors to evalute. Further studies may help to identify the causal factors determining sneaker resale value, and the role other factors play in the sneaker resale market.