JENNIFER **TRAN** github.com/jenniettrann c. 2025

"SOLE"

ARBITRAGE

The Business of Hype Kicks

TABLE OF CONTENTS

THE OVERVIEW

The emergence of the sneaker resale market as a high-yield asset class reflects broader shifts in consumer behavior and alternative investing. This study aims to shed a light on the resale performances of Nike and Adidas sneakers using transaction data from StockX data contest in 2019, made publicly available through Kaggle (Stuck, 2019). The data sample

Set against a macroeconomic backdrop of low interest rates (U.S. Treasury, 2019), the dataset reveals how limited-edition sneakers gained traction as alternative investment vehicles. Through statistical modeling and clustering, this analysis analysis seeks to answer:

1. What are the defining statistical properties of successful resale models?
2. Risk-adjusted return characteristics
3. Indicators of market inefficiency or arbitrage

METHODOLOGY

Our analysis utilized Python data science libraries to process and analyze StockX transaction data from 2019. The methodology included:

- Descriptive statistical analysis of sales data
- Brand performance comparisons
- Clustering analysis to identify similar performing models
- Risk-adjusted return calculations (Sharpe ratio)
- Efficiency score to assess investment value relative to volatility
- Liquidity score to measure how quickly and frequently models traded
- ROI heatmap visualization to explore patterns in profitability across brands

DATA OVERVIEW

This dataset contains U.S.-based sneaker resale transactions sourced from StockX data contest in 2019 and shared via Kaggle. The full dataset includes 99,956 random sample for all Off-White x Nike and Yeezy 350 sales from September 1, 2017, including: 27,794 Off-White x Nike and 72,162 Yeezy 350. All entries in the dataset are complete (no missing values), and contains U.S. sales only.
The dataset was generated by sampling a fixed random percentage (X%) of daily transactions per sneaker colorway, ensuring representative coverage across time and product variants (Einhorn, 2019).

Column	Description	Example
Order Date	Date the transaction was completed on StockX.	2017-09-01
Brand	Sneaker brand involved in the sale; categorical feature.	Yeezy
Sneaker Name	Full product name including model and colorway.	Adidas-Yeezy-Boost-350-Low-Moonrock
Sale Price	Final resale price in USD.	$1097
Retail Price	Original retail price in USD.	$220
Release Date	Official sneaker release date.	2016-09-24
Shoe Size	US shoe size sold.	11.0
Buyer Region	U.S. region of the buyer.	California

Table 1: Input Variables and Data Characteristics

BRAND PERFORMANCE ANALYSIS

Off-White sneakers consistently outperformed Yeezy models across every key metric. They delivered over 4x the ROI, with an average resale price nearly double that of Yeezys. On a per-item basis, Off-White profits averaged $492.10, more than three times Yeezy's $140.16. Additionally, Off-White had a broader presence in the dataset, with 30 unique models compared to 20 for Yeezy.

Brand	ROI	Sale Price	Profit	Unique Models
Off-White	282.67%	$671.48	$492.10	30
Yeezy	63.95%	$360.03	$140.16	20

Table 2: Brand-Level Resale Performance Summary

These findings highlight the premium position of Off-White collaborations in the resale market, with both higher returns and higher absolute profit margins.

[1] Air-Jordan-1-Retro-High-Off-White-White, [2] Air-Jordan-1-Retro-High-Off-White-Chicago, [3] Adidas-Yeezy-Boost-350-Low-Turtledove, [4] Nike-Air-Presto-Off-White, [5] Nike-Air-Force-1-Low-Virgil-Abloh-Off-White-AF10, [6] Nike-Blazer-Mid-Off-White-Wolf-Grey, [7] Nike-Blazer-Mid-Off-White, [8] Adidas-Yeezy-Boost-350-Low-Oxford-Tan, [9] Adidas-Yeezy-Boost-350-Low-Moonrock, [10] Adidas-Yeezy-Boost-350-Low-Pirate-Black-2016

MARKET METRICS

Metric	Value
Average ROI	124.82%
Median ROI	70.45%
Total Profit	$23,791,640.00
Total Sales Volume	99,956

Table 3: Key Resale Market Metrics

The large gap between the mean ROI (124.82%) and median ROI (70.45%) suggests a right-skewed distribution, where a few highly profitable models inflate the average. Total profit exceeding $23 million highlights the significant scale of this segment within the sneaker resale market. With only 50 unique models contributing to this profit, the data reveal a highly concentrated value distribution where most resale gains are driven by a relatively small subset of products.

MODEL CLUSTERING

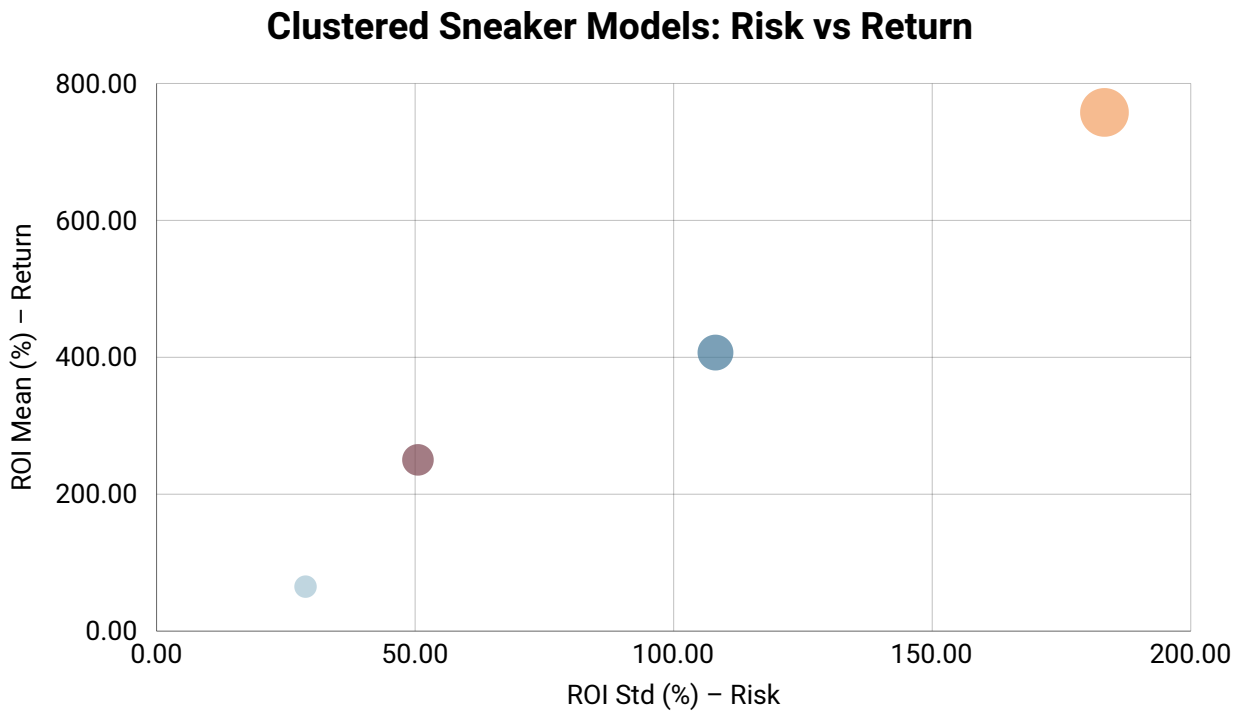
To better understand patterns in sneaker resale performance, we applied KMeans clustering to segment models based on their return and risk characteristics. The resulting four clusters reveal distinct market behaviors:

- **Cluster 3** represents the most attractive group for investors, with n average ROI of 250%, moderate risk (std: 50.5%), and the highest Sharpe Ratio (5.33), this group offers the most efficient returns per unit of risk. Models in this cluster are Nike Air Presto Off-White and Adidas Yeezy Boost 350 V2 Zebra.

- **Cluster 2** represents the high-risk, high-reward category, boasting the highest ROI (758%) and the highest average sale price ($1,590), but also substantial volatility (std: 183%). It includes premium, hype-driven releases like the Air Jordan 1 Retro High Off-White Chicago.

- **Cluster 1** delivers a strong balance of return and risk, with an average ROI of 407%, moderate-to-high volatility (std: 108%), and a healthy Sharpe Ratio (4.02). This group may include models like the Nike Dunk Low SB Travis Scott.

- **Cluster 0** reflects the lower-risk, lower-return segment, with an average ROI of 65% and the lowest volatility (std: 28.8%). While not flashy, models like the Adidas Yeezy Boost 350 V2 Cream White fall into this group.



KEY FINDINGS

Brand Performance Disparity: Off-White collaborations dramatically outperformed other brands, generating 282.87% ROI compared to Yeezy's 63.96%.

High-Performance Market Cluster: Cluster analysis identified a standout group of sneaker models with high ROI (250.19%), moderate risk (std: 50.55%), and the highest Sharpe Ratio (5.24). This segment, which includes the Nike Air Presto Off-White, offers the best balance between return and risk, making it the most efficient part of the market.

Premium Segment Advantage: The most profitable cluster also had a high average sale price ($668.96), suggesting that sneakers in the premium price range tend to generate the strongest investment returns.

Investment Timing: The best performing models showed strong performance within a specific time window (127-421 days after release), indicating an optimal holding period.

Risk-Adjusted Returns: The high Sharpe Ratio (5.33) of the top cluster demonstrates that these sneakers not only provided high returns but did so with relatively lower volatility compared to their return level.

Market Concentration: With just 50 unique models generating over $23M in profit across 99,956 transactions, the analysis reveals significant concentration of value in a small number of releases.

INVESTMENT IMPLICATIONS

Branding Matters: The dramatic outperformance of Off-White collaborations suggests that brand success thrived in the fusion of luxury fashion and sportswear credibility, whereas Yeezy relied more heavily on celebrity branding.

Collab Payoff: The scarcity of Off-White and Yeezy at launch, which retail only at maximum $220, delivered phenomenal resale gains reaching several thousand dollars.

Holding Period Optimization: The data suggests an optimal holding window, with peak performance occurring between 4–14 months post-release.

Portfolio Concentration: Given the high performance of a small cluster of models, a focused investment strategy targeting similar characteristics may outperform diversification.

Risk Management: The high Sharpe Ratio indicates that despite higher prices, premium sneakers may actually represent lower risk when returns are taken into account.

BEHAVIORAL AND TACTICAL CONSIDERATIONS

In addition to quantitative metrics, successful sneaker market strategies may benefit from behavioral insights and tactical frameworks:

Price Anchoring for Arbitrage: Understanding the base (retail) and ceiling (peak resale) prices of a model can highlight price inefficiencies, allowing for arbitrage by purchasing near the base and selling near observed ceilings.

Sizing Intelligence: Sizes 10–11 for men and 6–7 for women consistently show higher sales velocity and pricing power. Focusing acquisitions on these sizes may improve inventory turnover and return on investment.

Behavioral Edge: Other market participants often exhibit overconfidence or act on incomplete information. Recognizing this allows for contrarian strategies of buying undervalued pairs overlooked by less data-driven sellers.

Probabilistic Pricing Strategies: In illiquid or tightly contested markets, a “coin flip” pricing tactic where buyer and seller each propose a price (floor and ceiling) then flip a coin to determine the final price. This method can enable fair and faster trade closure while introducing game-theoretic dynamics.

CONCLUSION

This analysis of the 2019 StockX sneaker transaction dataset aims to understand the sneaker resale market's potential as a high-yield, data-driven investment space. Off-White collaborations emerged as the undisputed success with strong absolute profits and exceptional risk-adjusted returns. Yeezy models, although popular, lagged behind the Nike collaboration in performance. Clustering revealed that premium-priced, hype-driven releases can deliver outsized returns when acquired within optimal time windows and managed with disciplined holding strategies.

The market's concentration, where a small fraction of models generated the majority of profit, points toward the importance of selectivity over broad diversification. Behavioral and tactical insights, such as size targeting, price anchoring, and strategic release timing, further enhance returns and reduce inefficiencies.

Ultimately, the sneaker resale ecosystem mirrors traditional asset markets in its mix of volatility, inefficiency, and concentrated value, yet offers unique arbitrage opportunities for informed participants. With the right blend of quantitative analysis, market timing, and behavioral strategy, sneakers can serve not just as cultural icons, but as a legitimate alternative investment vehicle.

APPENDIX

Data Source

Stuck, H. (n.d.). Off-White x Nike and Yeezy 350 sales data [Data set]. Kaggle. https://www.kaggle.com/datasets/hstuck/offwhitexnike-and-yeezy350-sales

Reference

Einhorn, J. (2019, February 16). Attention Data Nerds: The StockX Data Contest Is Back. StockX News; StockX. https://stockx.com/news/the-2019-data-contest/

Resource Center. (2019). U.S. Department of the Treasury. https://home.treasury.gov/resource-center/data-chart-center/interest-rates/TextView?type=daily_treasury_yield_curve&field_tdr_date_value=2019

Python libraries:

- Pandas for data wrangling and transformation.
- NumPy for numerical operations and ROI calculations.
- Scikit-learn for unsupervised clustering via K-Means.
- Matplotlib and Seaborn for data visualization.

Key analytical methods included:

- K-Means clustering to identify performance-based market segments.
- Sharpe Ratio computation to evaluate risk-adjusted returns.
- Descriptive statistics to explore profitability and volume metrics.
- Time-based analysis to assess price trends relative to release dates.