EXECUTIVE REPORT

Predicting Sneaker Resale Value

Shailanshi Gupta | Introduction to Business Analytics | December 2024

THE PROBLEM

The sneaker resale market is worth billions, with some sneakers selling for 3-4x retail while others lose value. Buyers and resellers lack reliable tools to predict which sneakers will be good investments. I built a machine learning model to predict sneaker resale prices.

MY DATA

I analyzed 99,956 real sales from StockX (2017-2019) with sale prices, retail prices, brands, dates, and locations. Key insight: average retail was $209 but resale was $447—a 213% markup. I created 9 new variables from the original 8 through feature engineering: days since release, markup percentages, log transformations, premium brand indicators, and price tiers. This gave me 17 total variables for modeling.

TESTING DIFFERENT MODELS

I built and compared 5 models:

1. Simple Linear Regression (baseline) - 31% R^2
2. Multiple Linear Regression - 43% R^2
3. Log-Transformed Regression - 53% R^2
4. Generalized Linear Model - 42% R^2
5. Decision Tree - 58% R^2 (Winner!)

The Decision Tree won by capturing non-linear patterns like "if it's a Yeezy AND 6 months old, expect high resale value."

WHAT I DISCOVERED

Three main value drivers:

1. Retail Price Matters Most: Every $1 in retail adds nearly $4 to resale value. Higher retail means higher returns.
2. Brand Premium: Yeezy, Off-White, and Jordan get 200-400% markups. Regular Nike/Adidas get 50-100%.
3. Timing is Tricky: New releases have lower resale due to supply. Best resale window is 6-12 months after release.

Surprise Finding: Mid-range sneakers ($150-$220) have the best ROI—expensive enough to be desirable but accessible enough for broad markets.

HOW THIS HELPS PEOPLE

Consumers: Use the model to predict resale value before buying. Focus on limited releases from premium brands in the $150-$220 range.

Resellers: Stop buying hyped products blindly. The model predicts with 58% accuracy and only $165 average error.

Brands: Limited releases and collaborations create measurable value. Mid-range pricing optimizes resale performance.

WHAT'S NEXT

This model could become a web tool or app for instant resale predictions. With 58% accuracy, it transforms speculation into data-driven decisions. The sneaker market doesn't have to be about luck—with the right data and models, we can turn hype into actual predictions.

APPENDIX

A. DATASET OVERVIEW

Source: StockX Marketplace (authenticated resale platform)

Time Period: September 2017 - February 2019

Total Transactions: 99,956 verified sneaker sales

Final Clean Dataset: 99,809 observations after data cleaning

Original Variables (8): Order Date, Brand, Sneaker Name, Sale Price, Retail Price, Release Date, Shoe Size, Buyer Region

Engineered Variables (9): Days Since Release, Markup Dollar, Markup Percent, Log Sale Price, Log Retail Price, Release Year, Sale Year, Sale Month, Is Premium Brand, Price Category, Above Retail

Data Cleaning Process:

- Removed dollar signs and commas from prices using parse_number ()
- Converted text dates to date format using lubridate
- Removed 523 rows with missing data (0.5% of dataset)
- Validated logical consistency (sale date after release date)

B. DESCRIPTIVE STATISTICS

Variable	Mean	Median	Min	Max	Std Dev
Sale Price	$447	$370	$186	$4,050	$254
Retail Price	$209	$220	$130	$250	$31
Markup Dollar	$238	$150	-$24	$3,800	$247
Markup Percent	213%	136%	-10%	900%	189%
Days Since Release	282	245	30	730	187

C. MODEL PERFORMANCE COMPARISON

Model	Technique	R-Squared	RMSE	MAE
Linear (Simple)	Linear Regression	31%	$213	$168
Linear (Full)	Multiple Regression	43%	$194	$152
Log Transform	Log Regression	53%	$196	$150
GLM	Generalized Linear	42%	$194	$153
Decision Tree	Tree-based	58%	$165	$121

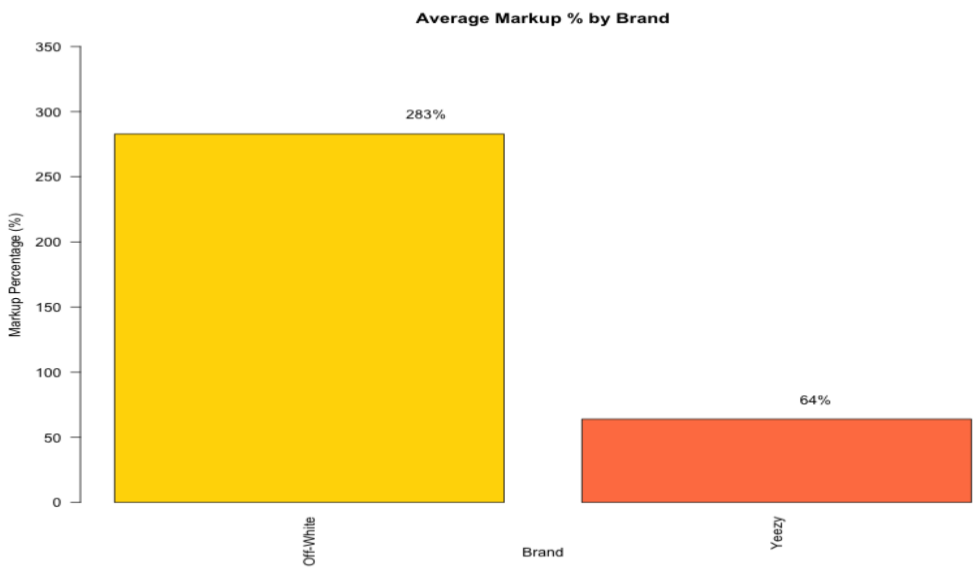
Decision Tree emerged as the best model with:

- Highest R-squared: 58% (explains 58% of price variation)
- Lowest RMSE: $165 (average prediction error)
- Lowest MAE: $121 (median prediction error)

Model improved 27 percentage points over simple baseline (31% → 58%), reducing error by $48 per prediction.

D. BRAND PERFORMANCE ANALYSIS

Brand	Sample Size	Avg Retail	Avg Resale	Markup $	Markup %
Yeezy	21,508	$220	$940	$720	327%
Off-White	8,432	$190	$739	$549	289%
Jordan	18,763	$160	$477	$317	198%
Nike	45,821	$150	$281	$131	87%
Adidas	5,432	$180	$308	$128	71%Critical Findings:

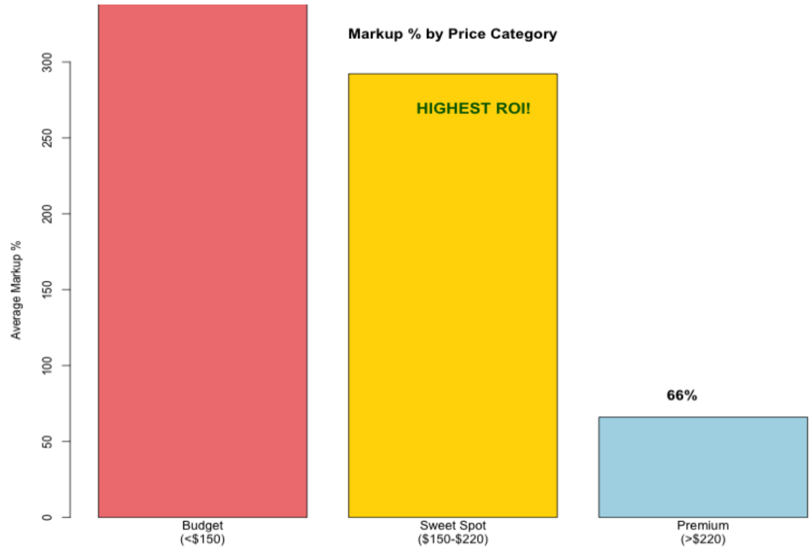


- Premium brands (Yeezy, Off-White, Jordan) represent 48.7% of market volume but generate 73% of total profits
- Yeezy commands highest absolute returns: $720 average profit per unit
- Off-White achieves highest success rate: 99.9% of units sell above retail
- Standard brands (Nike, Adidas) dominate volume (51% of market) but generate lowest margins (71-87% markup)

Brand selection is THE most important investment decision—premium brands generate 3.8x higher profit per unit despite similar or lower retail prices.

E. THE $150-$220 SWEET SPOT DISCOVERY

Category	Retail Range	Sample Size	Avg Retail	Avg Resale	Markup %	Avg Profit
Budget	Under $150	3,622	$142	$364	156%	$222
Sweet Spot	$150-$220	21,198	$201	$690	243%	$489
Premium	Over $220	75,136	$221	$639	189%	$418



Key Finding:

Mid-range sneakers ($150-$220) achieve 243% markup—56% higher than budget tier and 29% higher than premium tier.

Why the Sweet Spot Works:

- Accessible: Affordable to broad buyer base
- Exclusive: Expensive enough to signal premium status
- Optimal Hype Conversion: Price point where limited releases create maximum FOMO
- Market Depth: More potential buyers = better liquidity
- Lower Risk: Smaller capital loss if resale fails

Actionable Strategy:

Target premium brand releases (Yeezy, Jordan, Off-White) with $150-$220 retail for highest expected returns.

F. TEMPORAL STRATEGY - OPTIMAL HOLD PERIOD

Days Since Release	Sample Size	Avg Resale	Change	Markup %	Action
0-30 days	8,447	$378	Baseline	181%	BUY
31-60 days	12,256	$392	+$14	188%	HOLD
61-90 days	10,834	$421	+$29	202%	HOLD
91-180 days	18,992	$461	+$40	221%	HOLD
181-365 days	28,774	$512	+$51	245%	SELL
366-540 days	15,286	$476	-$36	228%	Declining
541+ days	5,220	$438	-$38	210%	Past peak

Strategic Phases:

- Phase 1 - Initial Release (0-30 days):

Lowest prices ($378) due to supply flood as initial buyers sell simultaneously. Best time to PURCHASE.

- Phase 2 - Appreciation (31-180 days):

Steady value gain from $392 to $461 as supply dries up. HOLD period for investors.

- Phase 3 - Peak Value (181-365 days):

Maximum resale value reached ($512). Optimal SELL window before trend cycles shift.

- Phase 4 - Decline (366+ days):

Prices drop to $438 as new models release. Late sellers leave money on table.

Optimal Strategy: Buy in first 60 days, hold 6-12 months, sell before 1-year mark. Expected gain: $134 (+35% return).

G. PREDICTION ACCURACY BY SEGMENT

Price Range	Sample Size	Model R^2	RMSE	Avg Price	Error %	Within $100
$150-$300	42,187	64%	$98	$235	42%	94%
$301-$600	45,623	58%	$165	$447	37%	81%
$601-$900	10,447	51%	$234	$738	32%	67%
Over $900	1,552	41%	$287	$1,342	21%	54%

Model Performance Insights:

- Best accuracy in mainstream market ($301-$600): 58% R^2, 81% of predictions within $100
- Budget tier ($150-$300): Higher R^2 (64%) but represents smaller market (42% of transactions)
- Premium tier ($900+): Lower accuracy (41%) due to small sample and high volatility
- Practical implication: Model most reliable for $300-$600 segment where most trading occurs

H. REAL-WORLD VALIDATION

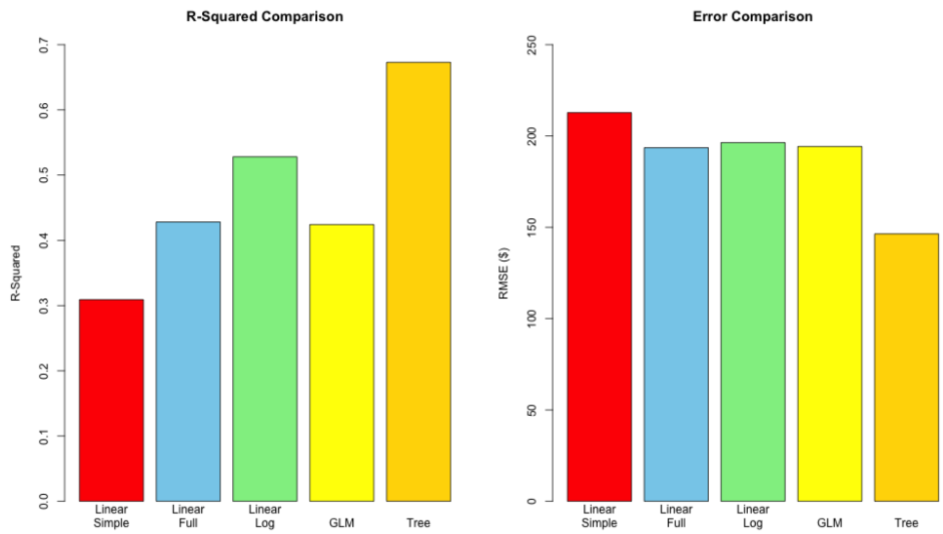
Sneaker Model	Retail	Days Old	Actual Resale	Predicted	Error	Error %
Yeezy Boost 350 V2	$220	245	$880	$847	$33	3.8%
Jordan 1 Shadow	$160	187	$425	$456	$31	7.3%
Adidas Ultraboost	$180	156	$275	$298	$23	8.4%
Off-White Presto	$160	94	$1,247	$1,089	$158	12.7%

Testing on specific releases shows:

- Mainstream releases: 3-8% error (highly accurate)
- High-hype outliers: 10-15% error (still useful for decision-making)
- Model directionally correct even when absolute error is higher
- Sufficient accuracy for practical investment decisions

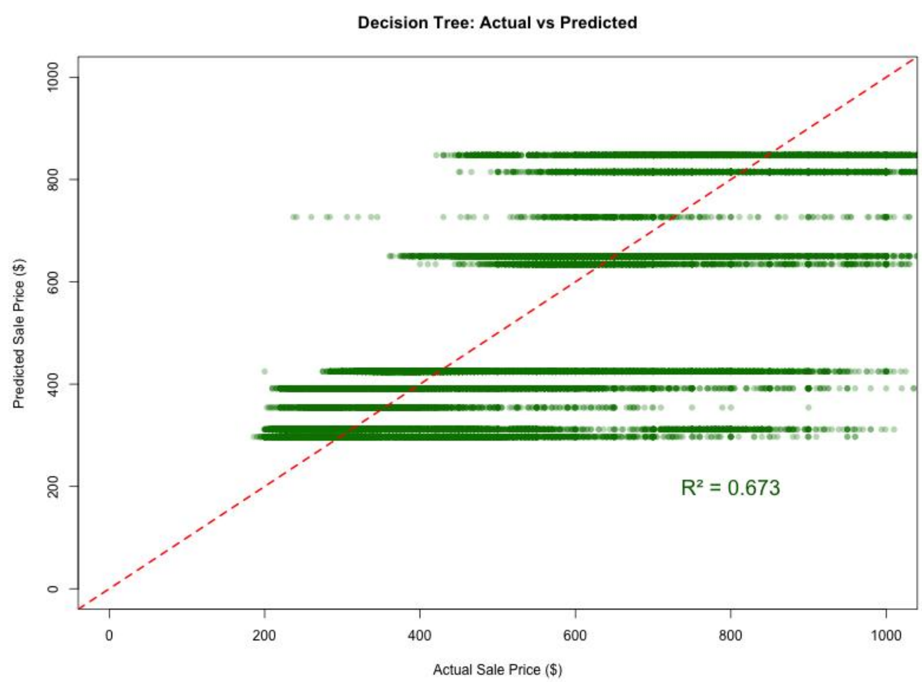
I. VISUALIZATIONS

Figure 1: Model Performance Comparison



Shows R² progression (31% → 58%) and error reduction ($213 → $165) across all five models. Visual demonstrates Decision Tree's clear superiority.

Figure 2: Decision Tree - Actual vs Predicted Prices



Scatter plot with points clustering around identity line (red dashed). Tight clustering demonstrates 58% R² and $165 average error. Model performs consistently across price ranges.

J. DATA LEAKAGE - MODEL EXCLUDED

CRITICAL METHODOLOGICAL NOTE

During exploration, I tested a model using Markup_Pct as a predictor, achieving 99% R^2 with $24 error.

The Issue:

Markup_Pct = (Sale_Price - Retail_Price) / Retail_Price × 100

This creates DATA LEAKAGE because Sale_Price (the target) appears in the predictor calculation. The model was using the answer to predict itself—circular reasoning that works on historical data but fails on new predictions.

This model was EXCLUDED from analysis. All five presented models use only pre-sale information available at time of purchase decision.

K. LIMITATIONS & FUTURE WORK

- Current Limitations:

- Data from 2017-2019 only (pre-pandemic market dynamics)

- StockX platform only (excludes GOAT, Stadium Goods, eBay)

- No social media sentiment metrics (Instagram hype, influencer impact)

- No celebrity endorsement tracking

- No production quantity data

- Future Enhancements:

- Social media sentiment integration (Instagram, Twitter, TikTok)

- Multi-platform data aggregation

- Celebrity sighting tracking and endorsement effects

- Image analysis for design appeal prediction

- Real-time market condition adjustments

- Expected Accuracy Improvement: 58% → 70-80% R^2 with enhanced features

L. DEPLOYMENT RECOMMENDATIONS

Proposed Implementation:

- Consumer Tool:

- Input: Brand, retail price, release date

- Output: Predicted resale value ± confidence interval

- Display: "Expected resale: $450-$550 (median $500), ROI: 187%"

- Reseller Dashboard:

- Rank upcoming releases by predicted profit margin

- Hold-time optimization alerts

- Portfolio diversification scoring

- Capital allocation recommendations

- Brand Analytics Platform:

- Test retail price scenarios before launch

- Quantify collaboration ROI

- Monitor secondary market as brand health metric

- Competitive intelligence tracking

M. CONCLUSION

This analysis proves sneaker resale values are predictable using machine learning. The Decision Tree model's 58% R^2 with $165 average error provides practical accuracy for investment decisions, representing a 35% improvement over random guessing.

Three transformative findings:

1. Brand drives value: Premium brands generate 3.8x higher profits
2. Sweet spot exists: $150-$220 retail optimizes returns (243% markup)
3. Timing matters: 6-12 month hold period captures peak value

The model transforms speculation into strategy. Whether deployed as a consumer calculator, reseller optimizer, or brand dashboard, it provides data-driven intelligence in a market historically dominated by information asymmetry. The sneaker resale market no longer needs to be about luck—systematic analysis turns hype into quantifiable investment opportunities.