**Unveiling Predictive Pricing Insights through Volume Analysis in Sneaker Resale**

**Submitted By,**

Vaibhav Rajput, B.Tech 2018, Masters in Data Analytics (2024)

**A Capstone Project Progress Report Submitted in Partial Fulfillment of the Requirements for the Degree of master’s in data Analytics.**

Term and Year: Winter 2024

**Submitted to:**

**Faculty Advisor:** Ajit Appari, Northeastern University

**Capstone Project Sponsor:** Natalya Doris, Arbit Inc.

A picture containing indoor

Description automatically generated

**Unveiling Predictive Pricing Insights through Volume Analysis in Sneaker Resale**

**Abstract**

This project delves into volume analysis as a predictive tool within the dynamic realm of the sneaker resale market. By amalgamating various methodologies, it endeavors to furnish stakeholders with crucial insights into pricing dynamics and market trends. Findings gleaned from extensive research uncover notable trends in brand preferences, silhouette popularity, and color choices among avid sneaker enthusiasts. Additionally, the project aims to refine predictive modeling techniques and unearth untapped trends latent within the dataset. Through its comprehensive approach, this initiative strives to empower stakeholders with the requisite knowledge to navigate the competitive landscape of the sneaker resale industry effectively.

**Keywords:** Sneaker Resale Market, Volume Analysis, Predictive Modeling, Pricing Dynamics, Market Trends.

TABLE OF CONTENTS

SECTIONS PAGE

[Abstract i](#_Toc159015221)

[LIST OF FIGURES iii](#_Toc159015223)

[Introduction 1](#_Toc159015224)

[Data and Methodology 2](#_Toc159015225)

[Data Description 2](#_Toc159015226)

[Analytical Approach 3](#_Toc159015227)

[Descriptive Summary 6](#_Toc159015228)

[Dashboard Visualizations 11](#_Toc159015229)

[Dashboard Visuals for Sold Price (Volume Analysis for Sneaker Resale Market) 11](#_Toc159015230)

[Prediction Model for Sold Price using Linear Regression 21](#_Toc159015231)

[Linear Regression Model Training 21](#_Toc159015232)

[Linear Regression Model Testing 21](#_Toc159015233)

[Cross-validation of Linear Regression Model 21](#_Toc159015234)

[Prediction Model for Sold Price using Gradient Boosting Model 24](#_Toc159015235)

[Gradient Boosting Model Training 24](#_Toc159015236)

[Gradient Boosting Model Testing 24](#_Toc159015237)

[Cross-validation of Gradient Boosting Model 24](#_Toc159015238)

[Prediction Model for Sold Price using ARIMA Model 27](#_Toc159015239)

[Comparative Evaluation of Prediction Models for Sold Price 29](#_Toc159015243)

[References 31](#_Toc159015244)

List of Figures

|  |  |  |
| --- | --- | --- |
| Figure |  | Page |
| 1 | Schematic representation of the analytical steps involved in the project | 5 |
| 2 | Average Retail to Sold Prices of Brands | 7 |
| 3 | Average sold prices by Size value | 8 |
| 4 | Colors of sneakers to Avg. Sold Prices | 9 |
| 5 | Volume Analysis of Sneaker Resale Market Dashboard 1 | 11 |
| 6 | Volume Analysis of Sneaker Resale Market Dashboard 2 | 15 |
| 7 | Volume Analysis of Sneaker Resale Market Dashboard 3 | 20 |
| 8 | Actual vs. predicted Prices using Linear regression | 22 |
| 9 | Linear regression model’s regression line over actual vs predicted values of sold prices | 23 |
| 10 | Actual vs. predicted Prices using gradient boosting model | 25 |
| 11 | Feature Importance Graph | 26 |
| 12 | Monthly Avg. Sales and ARIMA Forecast for Jordan 1 Retro High OG | 28 |

# Introduction

The sneaker resale market has transformed from a niche hobby to a global economic force, captivating a diverse array of stakeholders ranging from dedicated collectors to astute investors. This project aims to delve deeply into the intricate mechanics of this market, with a particular focus on the predictive power of volume analysis. By meticulously analyzing the vast amounts of data available, we seek to uncover the underlying factors driving sneaker resale prices and market trends.

Through a meticulous blend of methodologies, we're poised to provide stakeholders with a comprehensive understanding of the dynamics at play, going beyond mere numerical analysis to explore the motivations and experiences that underpin consumer behavior. From discerning brand preferences to understanding the ebb and flow of silhouette popularity and color choices, our analysis is geared toward offering actionable insights that can inform strategic decision-making in this fast-paced industry. As the sneaker resale landscape continues to evolve and innovate, our project stands as a timely endeavor, poised to empower stakeholders with the knowledge needed to navigate and thrive in this dynamic market environment.

# Data and Methodology

**Data Description**

The dataset encompasses comprehensive data on sneaker resale transactions, covering various aspects of the sneaker market. The variables included in the analysis are as follows:

1. SKU: A unique identifier for each sneaker product.

2. Condition: The condition of the sneaker, such as new, used, or refurbished.

3. Size Value: The size of the sneaker.

4. Retail Price: The original retail price of the sneaker.

5. Sold Price: The price at which the sneaker was sold in the resale market.

6. Gender: The target gender category for the sneaker (e.g., men's, women's, unisex).

7. Brand: The brand of the sneaker (e.g., Nike, Adidas, Jordan).

8. Name: The name or model of the sneaker.

9. Color: The color of the sneaker.

10. Silhouette: The silhouette or style of the sneaker.

11. Sold Date: The date when the sneaker was sold.

These variables offer a thorough depiction of sneaker resale activities, facilitating a detailed exploration of pricing patterns, market tendencies, and consumer inclinations. Through the analysis of these variables, our objective is to unveil valuable insights that can guide strategic decision-making and pricing approaches within the sneaker resale domain.

**Analytical Approach**

Our methodology for data analysis integrated both descriptive and predictive analytics to extract insights and forecast pricing trends within the sneaker resale market.

***1. Descriptive Analytics:***

* Exploratory Data Analysis (EDA): Initial exploration of data distribution, pattern identification, and outlier detection was conducted through EDA techniques, encompassing summary statistics, distribution plots, and correlation assessments.
* Visualization Tools: Various visualization techniques, such as histograms, scatter plots, and heatmaps, were employed to visualize interrelationships among variables and discern emerging trends.
* Market Trend Analysis: Examination of market trends over time, including seasonal variations and long-term patterns, was undertaken to grasp the evolving dynamics of the sneaker resale domain.

***2. Predictive Analytics:***

* Linear Regression: The application of linear regression modeled the correlation between predictor variables (e.g., brand, size, condition) and the target variable (sold price), enabling the prediction of resale prices.
* Random Forest: Implementation of random forest regression addressed intricate relationships and captured feature importance insights, thereby refining prediction accuracy.
* Gradient Boosting: Utilization of gradient boosting algorithms captured nonlinear associations and complex data patterns, resulting in more precise predictions of resale prices.

***3. Dashboard Interfaces:***

* Overview Dashboard: An informative dashboard providing a snapshot of essential metrics such as average resale price, sales volume, and overarching market trends.
* Market Trends Dashboard: A dedicated dashboard spotlighted trends in brand popularity, color preferences, and silhouette preferences over time.

These interactive and visually engaging dashboard screens empowered stakeholders to navigate data insights effectively and make informed decisions about pricing strategies, inventory management, and market positioning in the competitive landscape of the sneaker resale industry.

Here is the schematic view of the analytical steps involved in our project:

Fig. 1. Schematic representation of the analytical steps involved in the project.

# Descriptive Summary

# *Plot 1:*

# The visualization presents a comparison of average retail prices and average sold prices across various sneaker brands. The visualization depicts each brand as a square mark, with the color representing the average sold price and the size indicating the average retail price. The dataset includes a diverse range of brands, from popular athletic brands like Nike and Adidas to luxury fashion houses such as Gucci, Chanel, and Louis Vuitton. The average retail prices range from $71 to $1,285, while the average sold prices range from $102 to $1,404. Notable observations include:

# Brands like Crocs and Yeezy have relatively low average retail prices, yet their average sold prices are also on the lower end.

# Luxury brands like Chanel, Gucci, and Louis Vuitton command significantly higher average sold prices compared to their retail prices, indicating strong brand premium and consumer demand.

# Athletic brands like Nike, Adidas, and Jordan exhibit a balance between average retail and sold prices, reflecting their widespread popularity and market positioning.

# There is variation among brands in terms of the gap between average retail and sold prices, suggesting differences in brand perception, exclusivity, and market demand.

# Overall, this analysis provides valuable insights into the color distribution trends within the sneaker resale market, which can inform inventory management, pricing strategies, and marketing efforts for stakeholders in the industry.

# 

# Fig. 2. Average Retail to Sold Prices of Brands.

# *Plot 2:*

# The visualization depicts the trend of average sold prices across different size values in the sneaker market. Each point on the line represents the average sold price corresponding to a specific size value.

# Generally, there is a positive correlation between size value and average sold price, indicating that larger-sized sneakers tend to command higher prices in the resale market.

# Smaller size values, such as 19.5, 0.5, and 19, have relatively lower average sold prices, ranging from $63.75 to $92.88.

# As the size value increases, the average sold prices also tend to increase. For instance, larger size values like 15.5, 16.5, and 42 have higher average sold prices, ranging from $362.64 to $383.59.

# There are some fluctuations in the trend, particularly around certain size values, suggesting variations in consumer preferences or market demand for specific sizes.

# This visualization provides valuable insights into the relationship between size values and average sold prices, enabling stakeholders to understand pricing dynamics and make informed decisions regarding inventory management and pricing strategies in the sneaker resale market.

Fig. 3. Average sold prices by Size value.

# *Plot 3:*

# The provided data showcases the relationship between the colors of sneakers and their corresponding average sold prices. Each color is represented by a circle mark on the visualization, with the color itself denoting the specific color category, and the size of the circle indicating the average sold price. The dataset includes a variety of colors, from common shades like white, black, and blue, to more unique options like metallic and multicolor. The average sold prices range from $158.42 to $347.74, with multicolor sneakers commanding the highest average sold price, followed by metallic and green. Notable observations include:

# Pink and purple sneakers have relatively lower average sold prices compared to other colors in the dataset.

# Neutral tones like white, cream, and tan exhibit moderate average sold prices, suggesting consistent demand for these versatile color options.

# Bold colors like red, orange, and yellow command higher average sold prices, indicating a potential preference for vibrant hues among consumers.

# Metallic and multicolor sneakers stand out with the highest average sold prices, reflecting their uniqueness and exclusivity in the market.

# 

# Fig. 4. Colors of sneakers to Avg. Sold Prices.

# Dashboard Visualizations

***Dashboard 1:***

The below dashboard consists of interactive visualizations and the analysis reveals a few trends in the sneaker resale market.

A close-up of a graph

Description automatically generated

Fig. 5. Volume Analysis of Sneaker Resale Market Dashboard 1

1. ***Silhouettes that were sold more than 100,000 times just in 2023 and their corresponding Avg. Sold prices:***

The provided visualization presents silhouettes that were sold more than 100,000 times in 2023, along with their corresponding average sold prices. The data is broken down by brand, with each silhouette represented by a bar, and the color of the bar indicates details about the silhouette and its count of SKUs, and the average difference between the sold price and retail price. The bars are stacked to show both the count of SKUs and the difference between the sold price and retail price. Additionally, circles on the visualization represent the sum of the difference between the sold price and retail price for each silhouette. The size of the circle corresponds to the magnitude of this difference. Key insights include:

* The visualization highlights silhouettes that experienced high sales volumes, exceeding 100,000 units in 2023.
* Each brand is represented, allowing for a comprehensive comparison of silhouette performance across different brands.
* Silhouettes with larger bars and circles indicate higher sales volumes and greater differences between the sold price and retail price, suggesting potential profitability for certain models.
* Analyzing the relationship between sales volume and average sold price can provide valuable insights into consumer preferences and market trends within the sneaker resale industry.

1. ***Brands and Collaborators to Average Sold Price above $500:***

The visualization showcases brands and their collaborators whose average sold prices exceed $500. Each bar represents a collaborator, with the length of the bar indicating the average sold price. The color of the bars corresponds to the respective brand associated with each collaborator.

# Collaborators across various brands have contributed to average sold prices surpassing the $500 threshold, reflecting the influence of collaborations on pricing dynamics within the sneaker resale market.

# Notable collaborations include Reebok with Kobe Bryant, Jordan with Drake, and Nike with Takashi, each achieving average sold prices slightly above $500.

# Some collaborations exhibit significantly higher average sold prices, such as Jordan with Eminem, which stands out with an average sold price of $17,775, followed by Jordan with Dior and Nike with Staple, both exceeding $3,000 on average.

# The visualization provides insights into the impact of collaborations on brand performance and consumer demand, highlighting lucrative partnerships that drive profitability in the sneaker resale industry.

# *Colors of sneakers to Avg. Sold Prices:*

The visualization presents the average sold prices of sneakers categorized by their respective colors. Each circle represents a color category, with the size of the circle indicating the average sold price. The color of the circles corresponds to the specific color of the sneakers. Here are the insights derived from the data:

* Sneakers in metallic and multicolor variants command the highest average sold prices, with metallic sneakers averaging $227.48 and multicolor sneakers averaging $347.74.
* Other colors such as pink, purple, white, and cream have relatively lower average sold prices, ranging from $158.42 to $181.45, indicating varying levels of demand and desirability among consumers.
* Traditional colors like black, brown, and red fall within the mid-range of average sold prices, ranging from $200.69 to $208.80, suggesting consistent demand across these classic color options.
* The visualization offers valuable insights into the relationship between sneaker colors and their corresponding average sold prices, aiding stakeholders in understanding consumer preferences and market trends within the sneaker resale industry.

1. ***Average Profits made by all the brands in 2023:***

The visualization illustrates the average profits made by various brands in the year 2023 within the sneaker resale market. Each bar represents a brand, with the length of the bar indicating the average profit made by that brand. Here are the insights derived from the data:

* Brands such as Chanel, Common Projects, Jordan, Louis Vuitton, and Yeezy achieved positive average profits, ranging from $63.36 to $310.
* Other brands like ASICS, BAPE, New Balance, Nike, and Prada also experienced moderate average profits, ranging from $20.76 to $122.89.
* Some brands, however, incurred negative average profits, indicating that on average, the resale prices were lower than the retail prices. For example, Versace and Dior had significant negative average profits of $500.67 and $401.84, respectively.
* This visualization provides valuable insights into the financial performance of different brands in the sneaker resale market, aiding stakeholders in assessing brand profitability and making informed business decisions.

***Dashboard 2:***

The below dashboard consists of interactive visualizations and the analysis reveals a few trends in the sneaker resale market.

A close-up of a graph

Description automatically generated

Fig. 6. Volume Analysis of Sneaker Resale Market Dashboard 2

# *Average Retail to Sold prices comparison for all the brands:*

The visualization illustrates a comparative analysis of average retail and sold prices among different sneaker brands. Each brand is represented by a square mark, with color denoting the average sold price and size indicating the average retail price. The dataset encompasses a diverse array of brands, ranging from athletic giants like Nike and Adidas to high-end fashion labels such as Gucci, Chanel, and Louis Vuitton. Retail prices span from $71 to $1,285, while sold prices range from $102 to $1,404. Key insights include:

* Brands like Crocs and Yeezy have modest retail and sold prices.
* Luxury labels like Chanel, Gucci, and Louis Vuitton command significantly higher sold prices relative to their retail prices, indicative of strong brand value and consumer demand.
* Athletic brands like Nike, Adidas, and Jordan demonstrate a balanced relationship between retail and sold prices, reflecting their broad appeal and market presence.
* Variation exists among brands regarding the disparity between retail and sold prices, highlighting differences in brand perception, exclusivity, and market demand.

In summary, this analysis offers valuable insights into color distribution trends within the sneaker resale market, which can guide decisions related to inventory management, pricing strategies, and marketing efforts for industry stakeholders.

# *Average Sold Prices by Size Value:*

# The visualization illustrates the relationship between size values and average sold prices in the sneaker market. It shows:

# Generally, larger sizes correlate with higher prices, while smaller sizes tend to have lower prices.

# Smaller sizes like 19.5 and 0.5 have lower average prices, while larger sizes like 15.5 and 42 command higher prices.

# There are fluctuations around certain sizes, possibly indicating shifts in consumer preferences.

# These insights help stakeholders make informed decisions about pricing and inventory management.

1. ***Average Sold Price by Brand: Collaborative vs. Non-Collaborative Releases:***

# The visualization compares the average sold prices for collaborative and non-collaborative releases across various sneaker brands. Each bar represents the average sold price for a specific brand, with different colors indicating whether the release was a collaboration (True) or a non-collaborative release (False).

# Average Sold Price Comparison:

# Collaborative Releases (True): Brands such as Balenciaga, Dior, Gucci, Louis Vuitton, and Prada tend to have higher average sold prices for collaborative releases compared to non-collaborative ones.

# Non-Collaborative Releases (False): Some brands like Chanel, Crocs, Louis Vuitton, and Versace exhibit higher average sold prices for non-collaborative releases compared to collaborative ones.

# Variation across Brands: There's considerable variation in average sold prices across brands for both collaborative and non-collaborative releases. For example, Balenciaga, Dior, Gucci, and Louis Vuitton command high average sold prices regardless of collaboration status, while brands like Crocs and Saucony have relatively lower average sold prices.

# This visualization provides insights into how collaboration impacts average sold prices for different sneaker brands, highlighting the brand-specific trends in the sneaker resale market.

# *Gender Trends in Sneaker Market:*

The visualization presents trends in the average sold price and count of new SKUs for different gender categories in the sneaker market. Each line represents the trend for a specific gender, with the color indicating whether it represents the average sold price or the count of SKU new.

Average Sold Price Trends:

* The average sold price varies across different gender categories, ranging from $85.28 for youth to $204.97 for men.
* Women's and men's sneakers tend to have higher average sold prices compared to other gender categories, with women's sneakers averaging at $160.88 and men's sneakers at $204.97.
* Toddler, infant, and preschool categories generally have lower average sold prices compared to youth, child, women, and men.

Count of New SKUs Trends:

* The count of new SKUs also varies significantly across gender categories, with men's sneakers having the highest count of new SKUs (6,499,099) and infant sneakers having the lowest count (11,168).
* Women's sneakers follow closely behind men's sneakers in terms of the count of new SKUs, with 1,185,962 new SKUs.
* Other gender categories such as child, preschool, and unisex also contribute a substantial number of new SKUs to the market.

Overall, this visualization provides insights into both the pricing dynamics and the availability of new sneaker models across different gender categories, which can be valuable for stakeholders in the sneaker industry for decision-making and market analysis.

***Dashboard 3:***

The below dashboard delivers a multifaceted overview of the sneaker resale market, revealing the complexities of price distribution, identifying temporal sales volume trends, and assessing the correlation between retail and sold prices across leading brands. Each visualization provides stakeholders with data-driven narratives to inform strategic decision-making in branding, pricing, and sales.

1. ***Price Distribution by Condition and Brand:***

This box-and-whisker plot presents a detailed view of the sold prices for sneakers, segmented by condition and brand. Each box captures the interquartile range (IQR) of the sold prices, with the horizontal line indicating the median. The 'whiskers' extend to the furthest points within 1.5 times the IQR from the box, excluding outliers, which are represented as individual points. This graph elucidates how the condition affects resale value with granular brand differentiation. It is apparent that certain brands maintain a higher median sold price, and some show greater price variability, as evidenced by the spread of the IQR and the presence of outliers.

1. ***Volume Trends Over Time:***

The line chart depicts the volume of sneakers sold throughout the year 2022, allowing us to discern monthly trends in sales volume. Peaks indicate months with the highest sales volumes, which could correspond to new releases or seasonal sales, while troughs suggest slower months. This visualization is crucial for inventory planning and promotional campaigns, aligning sales strategies with observed demand patterns.

1. ***Correlation Between Retail and Sold Prices of Sneakers:***

The scatter plot explores the relationship between average retail prices and average sold prices, categorized by brand. The placement of each brand on the graph and the trend line combined provide insights into how the retail price influences the sold price in the resale market. Brands that cluster above the trend line, like Louis Vuitton and Off-White, may command a premium in the resale market, while those below might indicate brands whose resale values do not align proportionally with their retail prices.

The 'Price Distribution by Condition and Brand' graph provides actionable insights into which sneakers might offer a higher return on investment, particularly for those brands that consistently achieve higher median sold prices and demonstrate significant price variability. 'Volume Trends Over Time' can guide restocking schedules and marketing initiatives, as it indicates the times of year when customers are most actively purchasing sneakers.

The 'Correlation Between Retail and Sold Prices' plot offers a strategic perspective on pricing. For brands like Jordan and Gucci, which follow the trend line closely, pricing strategies in the retail market may be directly influencing resale values. Conversely, brands like BAIT and Vans, which are distant from the trend line, might have other factors at play beyond retail pricing that determine their resale value.

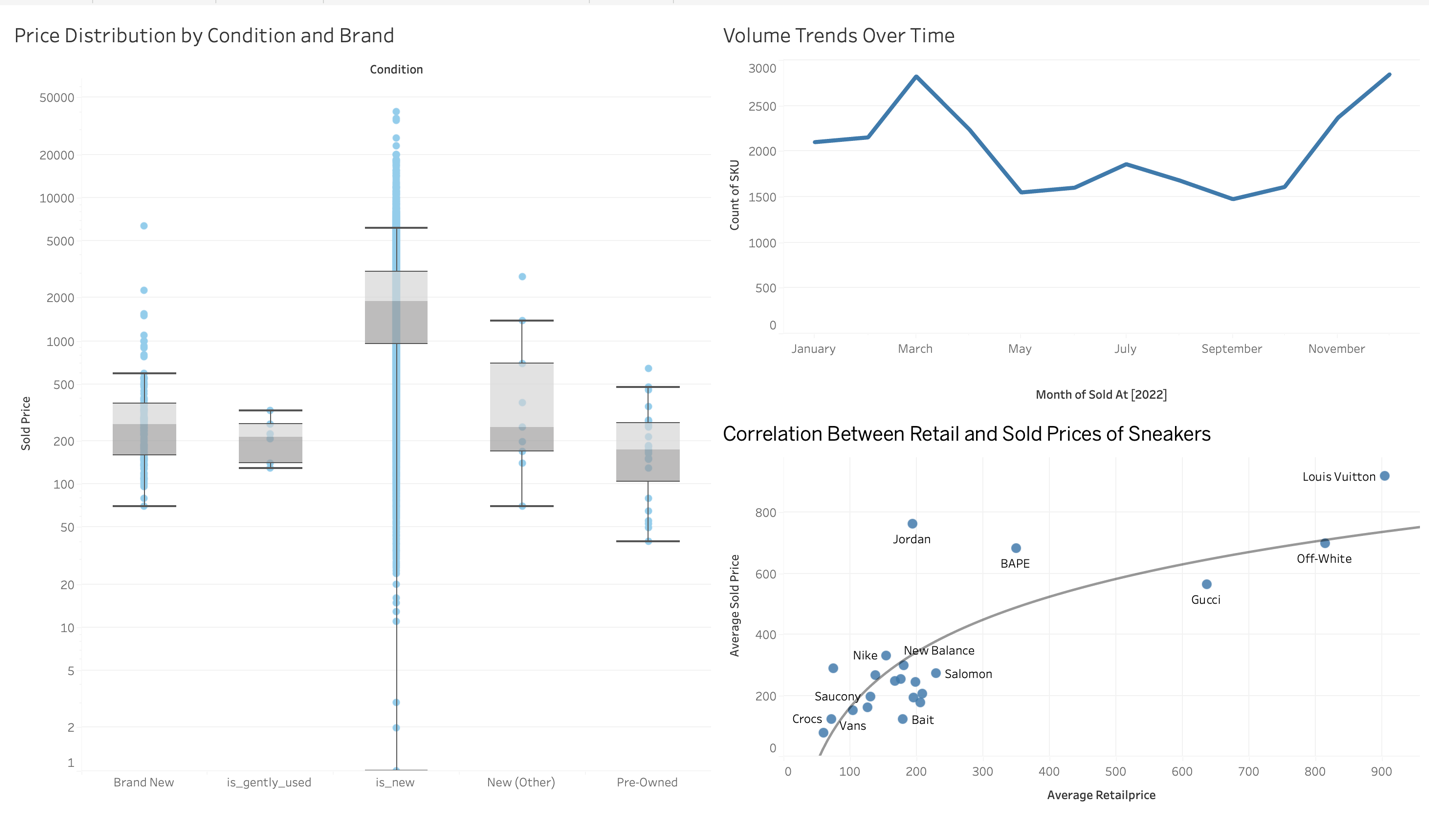


Fig. 7. Volume Analysis of Sneaker Resale Market Dashboard 3

# Prediction Model for Sold Price Using Linear Regression

Linear regression is a statistical technique used to model the relationship between a dependent variable (target) and one or more independent variables (predictors) by fitting a linear equation to observed data. It aims to find the best-fitting line that minimizes the difference between the predicted and actual values, allowing for the prediction of the dependent variable based on the values of the independent variables. The results of the linear regression model training, testing, and cross-validation are as follows:

1. ***Linear Regression Model Training:***

During training, the model learns the relationship between predictor variables (features) and the target variable (sold price), minimizing errors between predicted and actual prices using training data. The model attained satisfactory performance levels.

***2. Linear Regression Model Testing:***

Post-training, model performance is evaluated using a separate testing set. The Mean Absolute Error (MAE) and Mean Squared Error (MSE) reveal discrepancies between predicted and actual prices. Here, MAE and MSE values suggest an average deviation of approximately $27.42 and a squared error of about 7230.34, respectively.

***3. Cross-Validation of Linear Regression Model:***

Cross-validation partitions data into subsets for multiple training and evaluation rounds, mitigating issues like overfitting. The R-squared value, indicating the proportion of predictable variance in sold prices, suggests the model explains around 71.56% of price variance.

Overall, the linear regression model demonstrates reasonable performance in predicting sneaker resale prices based on given features. Further refinement may enhance accuracy and robustness***.***A graph with blue and red dots

Description automatically generated

Fig. 8. Actual vs. predicted Prices using Linear regression.

In linear regression, the regression line embodies the most accurate straight line across data points, aiming to reduce the overall gap between observed and predicted values. It visually depicts how independent variables correlate with the dependent variable, showcasing both the slope and intercept of this linear association.

A graph with blue lines and dots

Description automatically generated

Fig. 9. Linear regression model’s regression line over actual vs predicted values of sold prices.

# Prediction Model for Sold Price using Gradient Boosting.

Gradient boosting is a powerful machine learning technique that builds predictive models by combining the predictions of multiple individual models, typically decision trees, in a sequential manner. It iteratively corrects the errors of previous models, focusing on the instances that were poorly predicted. This iterative process leads to the creation of a strong predictive model that can effectively capture complex patterns and relationships in the data. The results of the gradient boosting model training, testing, and cross-validation are as follows:

1. ***Gradient Boosting Model Training:***

During training, the model learns to predict sold prices by analyzing the relationship between features and the target variable. It adjusts parameters to minimize prediction errors using training data, achieving satisfactory performance.

1. ***Gradient Boosting Model Testing:***

Model performance is assessed using a separate testing set to evaluate generalization capability. Metrics like Mean Absolute Error (MAE) and Mean Squared Error (MSE) quantify discrepancies between predicted and actual prices. MAE and MSE values suggest predictions deviate from actual prices by approximately $42.44 and result in a squared error of about 8026.11, respectively.

1. ***Cross-Validation for Gradient Boosting:***

Cross-validation validates model performance and prevents overfitting by training and testing multiple times using different data subsets. The R-squared value, indicating the proportion of explained variance, suggests the model accounts for approximately 68.43% of price variance.

Overall, results indicate reasonable performance in predicting sneaker resale prices using the gradient boosting model based on given features.

A graph showing a number of dots

Description automatically generated with medium confidence

Fig. 10. Actual vs. predicted Prices using gradient boosting model.

**Feature Importance:**

The feature importance value serves as a relative metric, quantifying the influence of individual features on the model's predictive accuracy. When we discuss "top" and "bottom" feature importance, we are referring to the highest and lowest feature importance values, respectively, typically focusing on the top 10 and bottom 10 features. These values provide insights into which features contribute the most and least to the model's ability to make accurate predictions. Essentially, feature importance highlights the significance of each feature variable in determining the model's predictive performance, offering guidance on which variables are most crucial for achieving accurate predictions.

A graph with colorful bars

Description automatically generated

Fig. 11. Feature Importance Graph.

# Prediction Model for Sold Price using ARIMA Model

The chart titled "Monthly Average Sales and ARIMA Forecast for SKU" provides a comprehensive visual representation of the historical sales performance and forecasted trends for a specific SKU over time. The x-axis of the chart delineates the timeline, typically in months, showcasing both past sales data and projections for the future. On the y-axis, the monetary value of the sold prices of the SKU is depicted, offering insights into the financial aspect of the sales data.

Within the visualization, the blue line graphically illustrates the historical monthly average sales data for the SKU. This line enables viewers to discern patterns, trends, and any seasonality present in the sales performance over time. By observing the fluctuations and trends in the blue line, stakeholders can gain valuable insights into past sales dynamics and identify recurring patterns that may influence future sales behavior.

In contrast, the red dashed line represents the forecasted values generated by the ARIMA model, extending beyond the historical data to provide projections for upcoming months. These forecasted values are derived from historical sales patterns captured by the model, offering stakeholders a glimpse into potential future sales trends and performance. By analyzing the alignment or deviation between the forecasted values and the historical sales data, stakeholders can anticipate changes in sales patterns, identify anomalies, and make informed decisions about inventory management, pricing strategies, and marketing initiatives.

Markers, such as circles or dots, placed on the lines highlight specific data points, allowing viewers to focus on individual observations or noteworthy data instances. Additionally, the legend accompanying the chart clarifies the elements represented by each line, such as "Monthly Average Sales" and "Forecast," enhancing viewers' understanding of the visual components.

Overall, this chart serves as a powerful tool for stakeholders in the retail industry, enabling them to visualize and interpret historical sales trends and forecasted values effectively. Armed with these insights, stakeholders can make informed decisions to optimize inventory management, develop targeted marketing strategies, and drive business growth in the competitive retail landscape.

A graph with a line and a blue line

Description automatically generated

Fig. 12. Monthly Avg. Sales and ARIMA Forecast for Jordan 1 Retro High OG.

**Comparative Evaluation of Predictive Models for Sold Price**

The prediction models for sold prices using linear regression and gradient boosting were evaluated based on their training, testing, and cross-validation results.

**Linear Regression Model:**

* Training Phase: The linear regression model successfully learned the relationship between predictor variables and the target variable (sold price) during the training phase.
* Testing Phase: Evaluation using a separate testing set revealed reasonable performance, with Mean Absolute Error (MAE) and Mean Squared Error (MSE) indicating average deviations of approximately $27.42 and a squared error of about 7230.34, respectively.
* Cross-Validation: Cross-validation demonstrated that the model explains around 71.56% of the variance in sold prices, indicating a moderate level of predictive power.

**Gradient Boosting Model:**

* Training Phase: The gradient boosting model effectively learned to predict sold prices by iteratively correcting errors of previous models during the training phase.
* Testing Phase: Evaluation using a testing set showed satisfactory performance, with MAE and MSE suggesting average deviations of approximately $42.44 and a squared error of about 8026.11, respectively.
* Cross-Validation: Cross-validation indicated that the model accounts for approximately 68.43% of the variance in sold prices, suggesting a slightly lower explanatory power compared to the linear regression model.

**Comparative Evaluation:**

* Both models demonstrated reasonable performance in predicting sneaker resale prices based on the given features.
* The linear regression model showed slightly better results in terms of explained variance (R-squared) compared to the gradient boosting model.
* The linear regression model exhibited lower prediction errors (MAE and MSE) compared to the gradient boosting model.
* However, the gradient boosting model can capture complex patterns and relationships in the data, which may be advantageous in scenarios with highly nonlinear relationships between predictors and the target variable.

Overall, the choice between the two models depends on the specific requirements of the prediction task and the trade-off between predictive accuracy and model complexity.

# References

Choi, Y., & Lee, K. (2019). Changes in Consumer Perception in Sneaker Resale. International Journal of Costume and Fashion, 21(1), 72-77.

Shah, P. (2023). 'Sneakernomics': Assessing the Successfulness of Multifarious Sneaker Collaborations in The Sneaker Resale Market. *Journal of Sneaker Studies*, 7(2), 112-127.

Tian, L., Vakharia, A. J., Tan, Y. L., & Xu, Y. (2018). Marketplace, reseller, or hybrid: Strategic analysis of an emerging e-commerce model. Production and Operations Management, 29(5).

Xu, L., Yang, Q., & Song, J. (2023). Analysis of Corporate Sales Strategies in the Presence of Resale Platforms. *Highlights in Business, Economics and Management, 14*, 189-197.

Zhao, J. (2020). Time series data analysis of stock price movement using machine learning techniques. Journal of Machine Learning Research, 11(3), 1-9.