# The booming sneaker resale market of the United States

Using linear regression analysis to explore drivers of sneaker resale prices by Benjamin Annor-Adjaye, FRM

# 1.0 INTRODUCTION

## 1.1 Background:

My love for sneakers started at the age of nine, while watching a Nike commercial for the 1998 Fifa World Cup competition held in France. The entirety of the commercial centered around the Brazilian national football team kicking the football and doing tricks all around a very busy airport. Though watching them perform some of the most daunting yet amazing skills with the football, there was one particular scene that caught my eyes. In that scene, Roberto Carlos wore a pair of gray and white Nike trainers, that was the moment I fell in love with the simplicity, beauty and story telling of sneakers. I spent the rest of the year asking my parents to get me a pair, but unfortunately there were more important financial obligations to take care of so that could not happen. As disappointed as I was for not being able to get those sneakers I made a promise to myself that I was always going to work hard and make a good living in order to not only take care of myself and family, but also buy as many sneakers as I wanted.  About five years later I found myself watching a movie by the legendary Spike Lee called "Do the right thing" those white Jordan 4 cements he wore in the movie reignited my love of sneakers. That moment right there set me on a path of collecting and trading sneakers with the goal of building a collection that reminds me of significant moments of my past and present.

After years of collecting some of my beloved sneakers from Nike, Adidas, Puma, Filas, New Balances, to name a few, I could never have imagined that there would ever come a day where getting a pair of your favorite sneakers could be sometimes as expensive as making a monthly mortgage or rent payment in a city like Chicago, New York or Los Angeles. Now don't get me wrong, there has always been sneaker resale from back when I was a kid, but never at this level.

## 1.2 Business Task:

The Global sneaker resale market was estimated at about $10 billion in February 2021 by the research firm Piper Sandler which was a $4 billion increase from 2019. The recent pandemic saw a boom in the market due to individuals looking for a side hustle to make some extra money to keep food on the table and a roof over their heads.  The resale market is proving to be a profitable source of income and I'm curious to find out if it is worth leveraging my passion for just collecting sneakers and wearing them, to possibly expand to reselling to make some extra income. I will be performing some analysis on sneaker resale data from Stock X, a major platform for buying and selling a variety of items with sneakers being the primary item of interest. I will use the results of my analysis to answer the following  questions which will hopefully help me make a decision at the end:

- What brand of sneakers is more profitable and a good investment
- What shoe size or sizes are most profitable
- Which state of the United States has most profitability in the resale.
- Perform a regression analysis to determine which factors greatly influence the resale value of sneakers

## 2.0 DATA AND ANALYSIS

### 2.1 Description of Data and Data Source:

Dataset consists of the single file of a random sample of all Nike (Off-White) and Adidas (Yeezy 350) sales from between 9/1/2017 (the month that Off-White first debuted "The Ten" collection) and 2/13/2019 from StockX. There are 99,956 total sales in the data set; 27,794 Off-White sales, and 72,162 Yeezy sales. The sample consists of U.S. sales only.

Data source: https://www.kaggle.com/hudsonstuck/stockx-data-contest

Initial data assessment was performed using pivot tables in Google Sheet to identify the column profile of each attribute. The assessment revealed that there were no data duplicates, no non null data values and all columns were valid. A second data assertion using SQL in BigQuery to verify my initial data profiling assessment was performed which confirmed the profiling results from my initial assessment. See details below about each column.

| Column Name | Description | Column Profile |
|---|---|---|
| Order Date | Represents the date the sneakers were ordered for purchase from StockX | 531 distinct values, 0 unique values |
| Brand | Represents the sneaker designer name which are Yeezy and Off-White. | 2 distinct values, 0 unique values |
| Sneaker Name | Represents the sneaker silhouettes from the brand designers Yeezy and Off-White | 50 distinct values, 0 unique values |
| Sale Price | Represents the resale prices of the sneakers on StockX in US Dollars | 1158 distinct values, 331 unique values |
| Retail Price | Represents the original price at which the sneakers were sold on release date in US Dollars | 8 distinct values, 0 unique values |
| Release Date | Represents the date the sneaker was first released to the public | 35 distinct values, 0 unique values |
| Shoe Size | Represents the various sneaker shoe sizes. Shoe size is not differentiated into Women's or Men's. | 26 distinct values, 0 unique value |
| Buyer Region | Represents the different states in the US where the resale purchase orders occured | 51 distinct values, 0 unique values |

### 2.2 Data Cleansing and Processing:

The first step in the data cleansing and processing was done using Google sheets to create a new data table called "Sneaker". Below are the steps highlighted as follows:

- Renamed the column "Brand" in the original dataset to "Designer" since the column contained records related to the designer of the sneaker silhouettes.
- Created a new column called "Brand" which captures the records of the actual sneaker brands Nike and Adidas.
- Created a new column "Year" capturing the record of each year in which sneaker resale orders were made via StockX using the "YEAR" function to extract the year of each order date from the "Order_Date" column.
- Created a new column "Calendar_Quarter" which captures a record of the quarter in which each resale order was made in each year. To do this I used a combination of the Concatenate, Roundup and Month functions.
- Created a new column "Order_Period" which represents a quarter and year date format for each order date record. To do this I used Concatenate to combine the records from the column "Calendar_Quarter" and "Year" to create a column with the date format QQ-YYYY.

The second step was done in RStudio to create a data frame called "Sneaker2.0". Below are the steps highlighted as follows:

- Converted all the column names to lowercase using the "rename_with" function.
- Renamed the columns Resale_Price and Retail_Price columns to resale_price_usd and retail_price_usd to indicate that the prices are in US Dollars.
- Removed the columns "Year" and "Calendar_Quarter" from the original data frame Sneaker using the select function.

The details of the resulting data frame "Sneaker2.0" is shown in the table below.

| Column Name | Description | Column Profile |
|---|---|---|
| order_date | Represents the date the sneakers were ordered for purchase from StockX | 531 distinct values, 0 unique values |
| order_period | Represents the date format that captures the calendar quarter and year in which a resale order was made. | 7 distinct values, 0 unique values |
| brand | Represents the sneaker brands Nike and Adidas | 2 distinct values, 0 unique values |
| designer | Represents the designers Kanye West and Virgel Abhlor brands Yeezy and Off-White respectively | 2 distinct values, 0 unique values |
| sneaker_name | Represents the sneaker silhouettes from the brand designers Yeezy and Off-White | 50 distinct values, 0 unique values |
| resale_price_usd | Represents the resale prices of the sneakers on StockX | 1158 distinct values, 331 unique values |
| retail_price_usd | Represents the original price at which the sneakers were sold on release date. | 8 distinct values, 0 unique values |

| | | |
|---|---|---|
| release_date | Represents the date the sneaker was first released to the public | 35 distinct values, 0 unique values |
| shoe_size | Represents the various sneaker shoe sizes. Shoe size is not differentiated into Women's or Men's, however all Men's shoes size translate to Women's sizes and vice versa | 26 distinct values, 0 unique value |
| buyer_region | Represents the different states in the US where the resale purchase orders occured | 51 distinct values, 0 unique values |

## 2.3 Data Analysis:

We began the analysis processes by identifying patterns, anomalies, biases and or outliers utilizing statistical and graphical representation, a concept known as Exploratory Data Analysis. The process will be broken into Estimation of location, Estimation of variability and Exploration of data distribution. The following are the highlights from the analysis. Detailed information can be found in the python notebook included in the appendix.

- Data has only floats, integers, datetime, objects across the 14 columns.

Summary statistics

| | Shoe_Size | Retail_Price | Resale_Price |
|---|---|---|---|
| Mean | 9.344181 | 208.61359 | 446.634719 |
| Trimmed Mean (removing top and bottom 10%) | 9.451241 | 213.34292 | 401.36177 |
| Standard Deviation | 2.329588 | 25.20001 | 255.982969 |
| Minimum | 3.500000 | 130.00000 | 186.000000 |
| 25 percentile | 8.000000 | 220.00000 | 275.000000 |
| 50 percentile | 9.500000 | 220.00000 | 370.000000 |
| 75 percentile | 11.000000 | 220.00000 | 540.000000 |
| Maximum | 17.000000 | 250.00000 | 4050.000000 |

- We notice that the mean value is less than the median for each of the columns
- There is a noticeable difference between the 75 percentile and the maximum value for the Shoe Size and Resale Price
- The preceding bullet points suggest the existence of possible outliers and skewness of data in Shoe Size and Resale Price which warrants additional investigation
- Due to the sensitivity of the mean to outliers we calculated the trimmed mean for the three columns to assess the impact of outliers on the mean. The results of the trimmed mean confirms the existence of outliers. Left skewness in Retail Prices since the trimmed mean and median are higher than the mean

retail price while Resale Price shows a right skewness since the trimmed mean and median are lower than the mean resale price
- The high standard deviation of the Resale Price indicates that resale prices are spread out over a wider range
- An analysis of the boxplots, violin plot and histograms confirms a slight left skewness in the shoe size, a strong left skewness in retail price and a strong right skewness in resale price confirm the existence of outliers and possible bias in the data we should look out for during our analysis
- An analysis of the distribution plot Nike and Adidas retail prices indicate that the majority of Nike retail prices are below the median retail price while Adidas retail prices are at the median price indicating that the Nike retail prices are responsible for the left skewness seen in the retail price. Resale price distribution plot of Nike and Adidas indicate that Nike resale price contributes mostly to the right skewness observed in resale prices.

We continued the analysis process with data visualization to identify significant patterns, outliers, differences and uniqueness in the data set. Since there are only two distinct observations for Brand, each associated with only one of the two distinct observations for Designer, we will use the name **Nike** interchangeably with **Off-White** and **Adidas** interchangeably with **Yeezy**. That is, all Off-White sneakers are Nike branded and all Yeezy sneakers are Adidas branded. The following visualizations provide us with an overview of the key observations from the data set.

**Total number of sneakers sold**
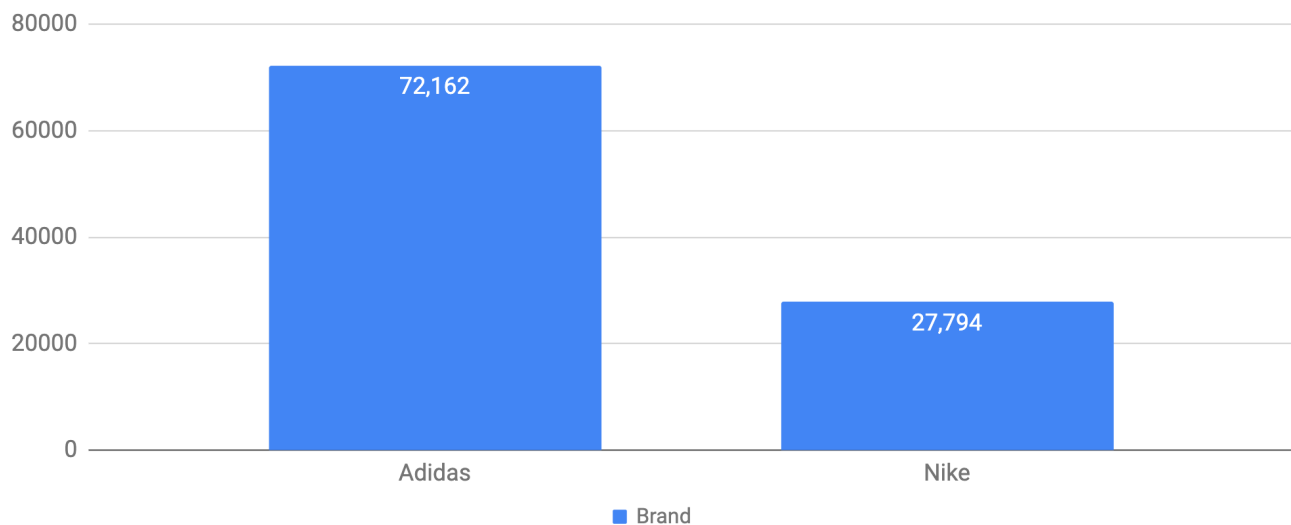


In this visualization we can see that there were 44,368 more Adidas sneakers sold as Nike sneakers over the period from September 2017 to 2019. There could be several reasons why that is, with a few major ones being price difference (i.e. Adidas on average being cheaper than Nike) and variability around the quantity of sneakers available for each brand at and post the sneaker release date as a way for these sneaker brands to create exclusivity. As we explore additional analysis, we will be able to check these potential factors which will hopefully aid to answer the question of why more Adidas sneakers were sold than Nike. To see how the number of

sneakers sold translated to profit we plotted a chart below which shows that though Adidas sold more quantity of sneakers than Nike, Nike made over $3MM more in profit than Adidas.

**Total Profit**
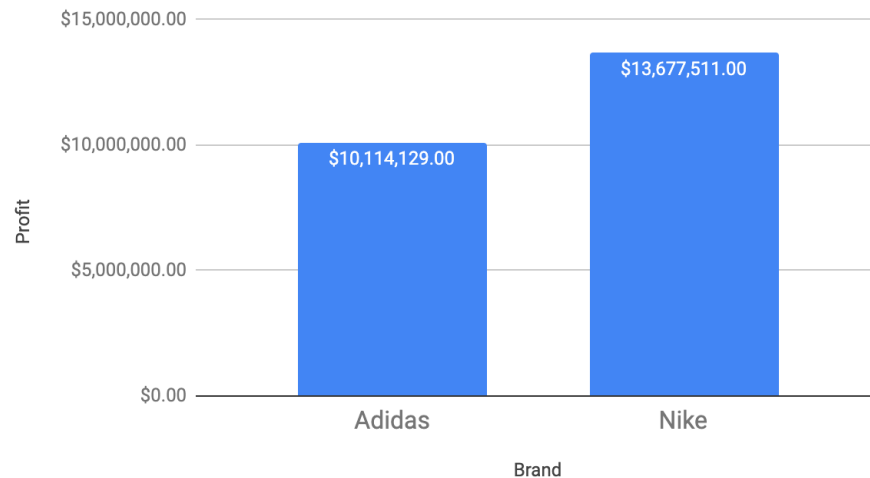


To determine what drivers are contributing to why sales quantity did not translate directly to profit we looked at a bar chart showing the distribution of the quantity of sneaker sales and average sale price per brand for each quarter from 9/1/2017 to 2/13/2019.
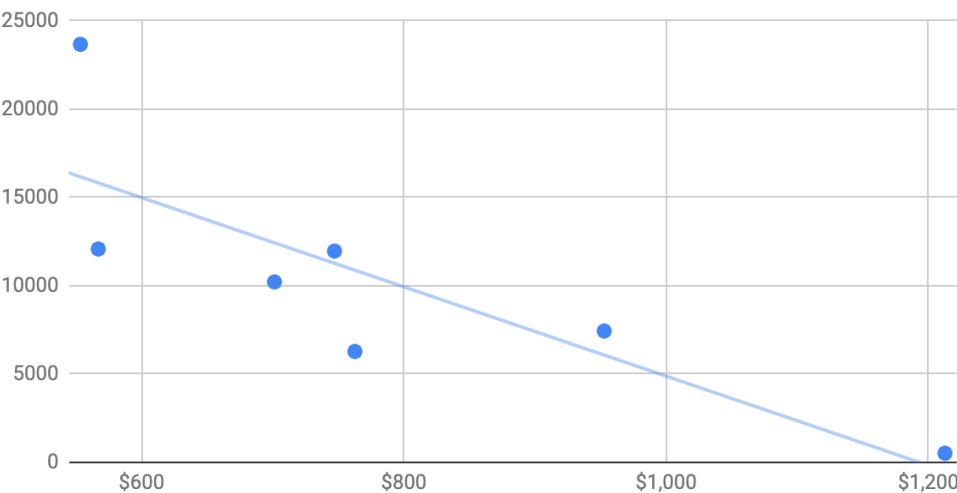
**Total sneaker sales**
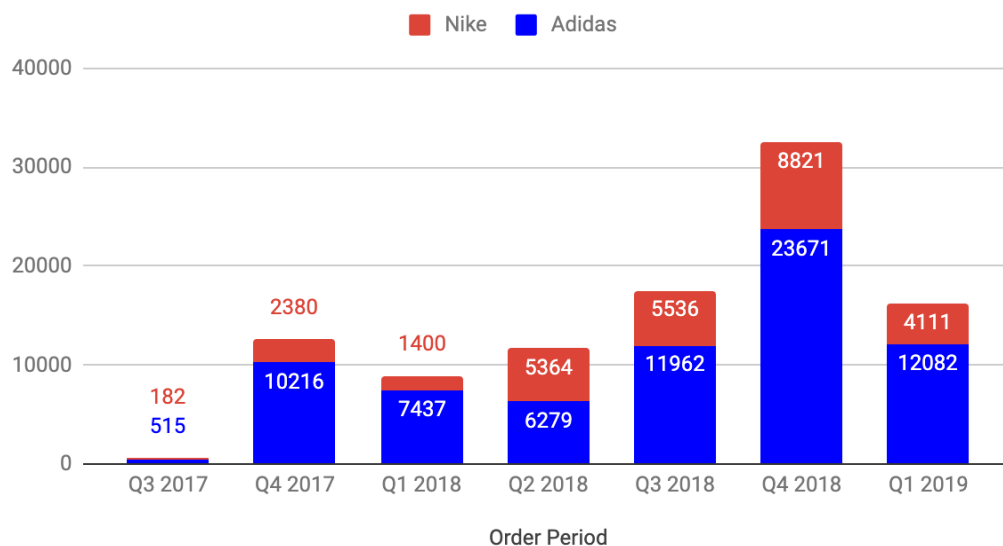
## Quarterly average sneaker resale price



We observe from the first chart that for both brands, quarterly average resale prices drop precipitously quarter-over-quarter and from the second chart that the quantity of sneaker sales increased precipitously quarter-over-quarter. Taking a closer look at both charts, we noticed that in every quarter where the quarterly average resale price was low, there was a significant increase in sneaker sales. An example is in Q4 2018 which had the lowest quarterly average resale prices and corresponding highest quantity of sneaker sales whereas Q3 2017 had the highest average resale prices and corresponding lowest sneaker sales. A scatter plot of average quarterly resale prices against total quarterly sales revealed a negative relationship between average resale price and quantity of sneaker sales supporting the observation made in the bar charts above, that the lower the resale price the higher the quantity of sneakers sold.

**Average quarterly Nike resale prices against total quarterly sneaker sales**



**Average quarterly Adidas resale prices against total quarterly sneaker sales**
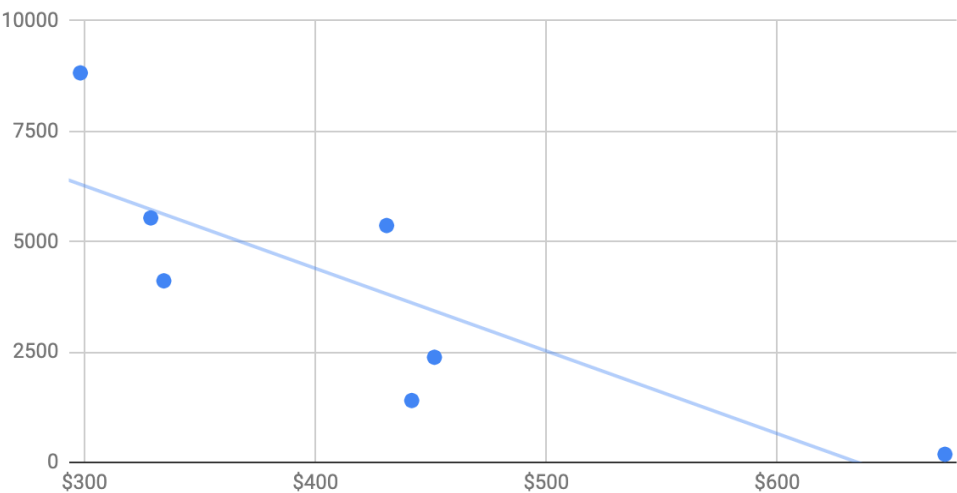


Analysis of the distribution of shoe sizes shows that the top five high selling shoe sizes in descending order are 10, 9, 11, 10.5 and 9.5 across both sneaker brands and that the top five selling sneakers make up 50% of all the profit from sneaker sales with shoe 10 being the highest contributor to profitability from sneaker resale, where profit is defined as the difference between resale price and retail price.
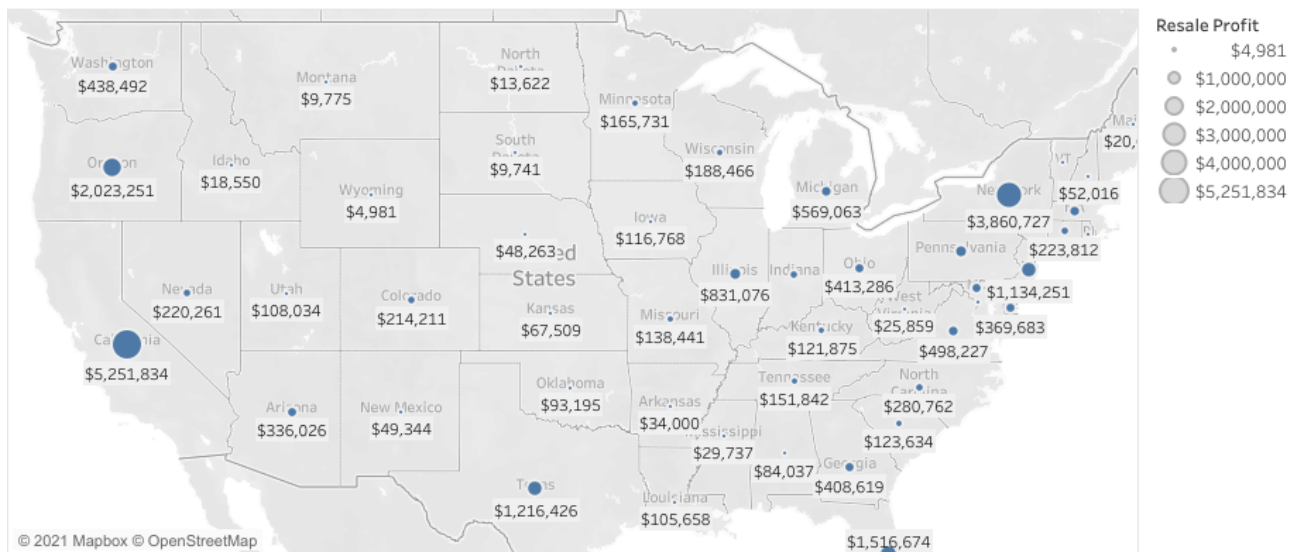
## Quantity of sneaker sale by shoe size
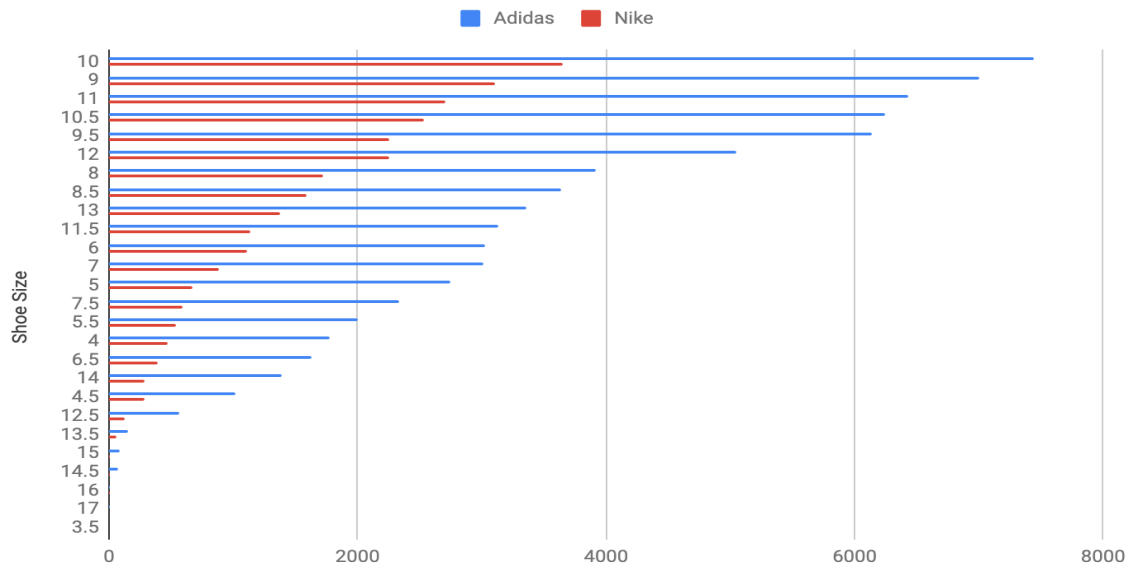


## Total profit by shoe size
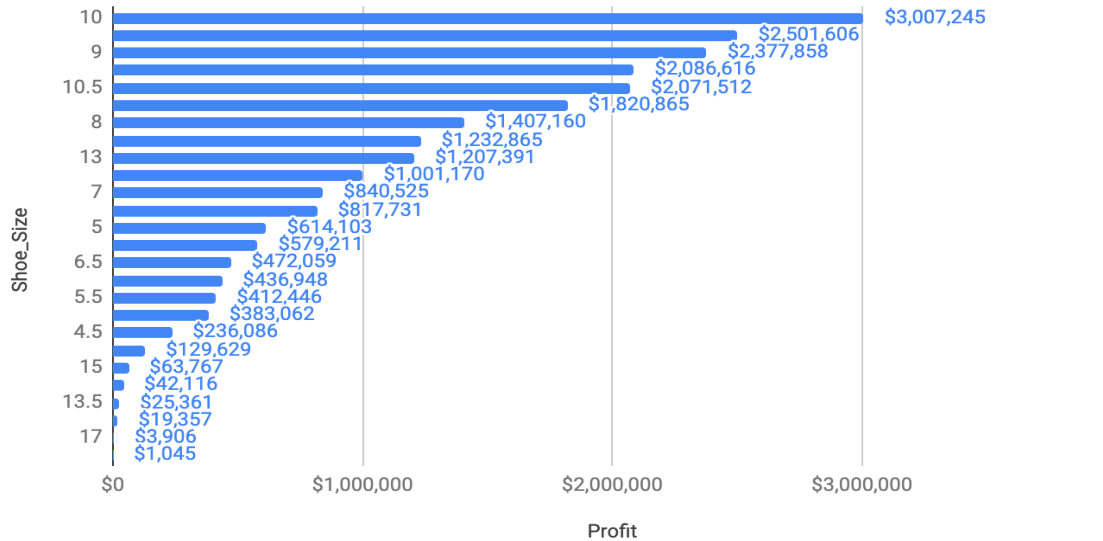


Another very important insight worth investigating is the distribution of sneaker sales across the US Region to explore which US states have the highest sneaker sale activity in regards to quantity, brand choice and profitability.

Resale Profit



We observed that the top five states with the highest generating profit in descending order are California, New York, Oregon, Florida and Texas. Additionally we observe from the tree map below that across the top five states in resale profitability, there were more Adidas sneakers sold relative to Nike sneakers at an average price of $300 less than Nike sneaker resale price.

## Quantity and average price of sneaker sales
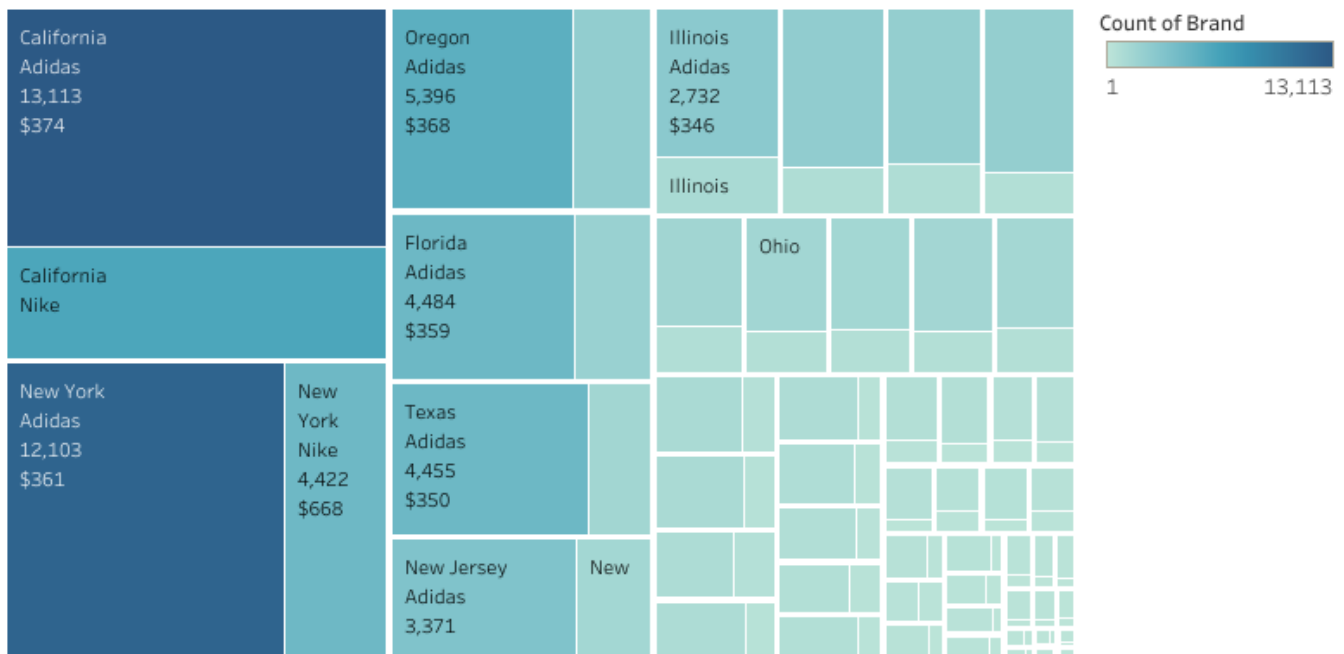


The final analysis worth exploring is determining if there are any attributes that influence the average resale price of sneakers. The following charts and analysis were explored in the process.

## Relationship between shoe size and average price
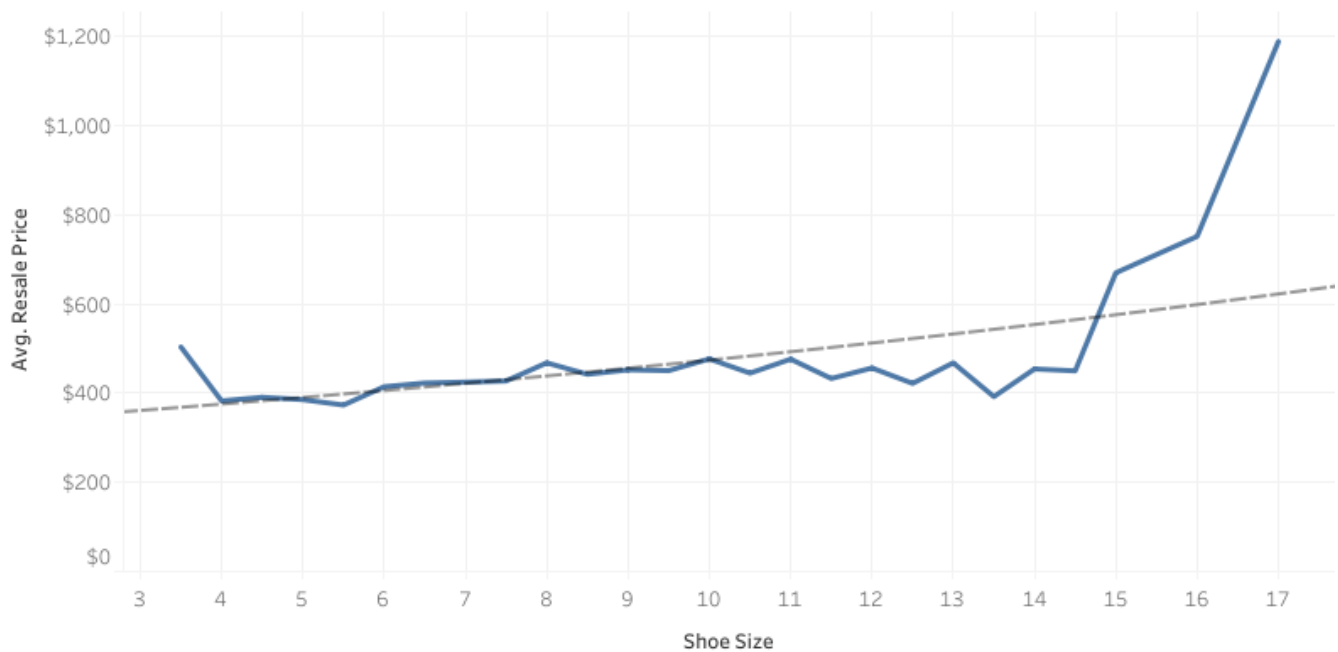


We observed from the line chart that there is some existence of a relationship between shoe sizes and average resale price with the most distinct relationship evident above shoe size 14. This is due to the fact that very few pairs of sneakers with shoe size above 14 are made compared to the other shoe sizes and as a result contributing

to the higher prices. To investigate what type of relationship exists between shoe size and average resale price we calculate the correlation coefficient which was 0.58 indicating a strong positive relationship between shoe size and average resale price. A correlation coefficient between shoe size and resale price was 0.082 indicating a low positive correlation. These two results suggest a need for further analysis using regression analysis.



Hold period against average resale price
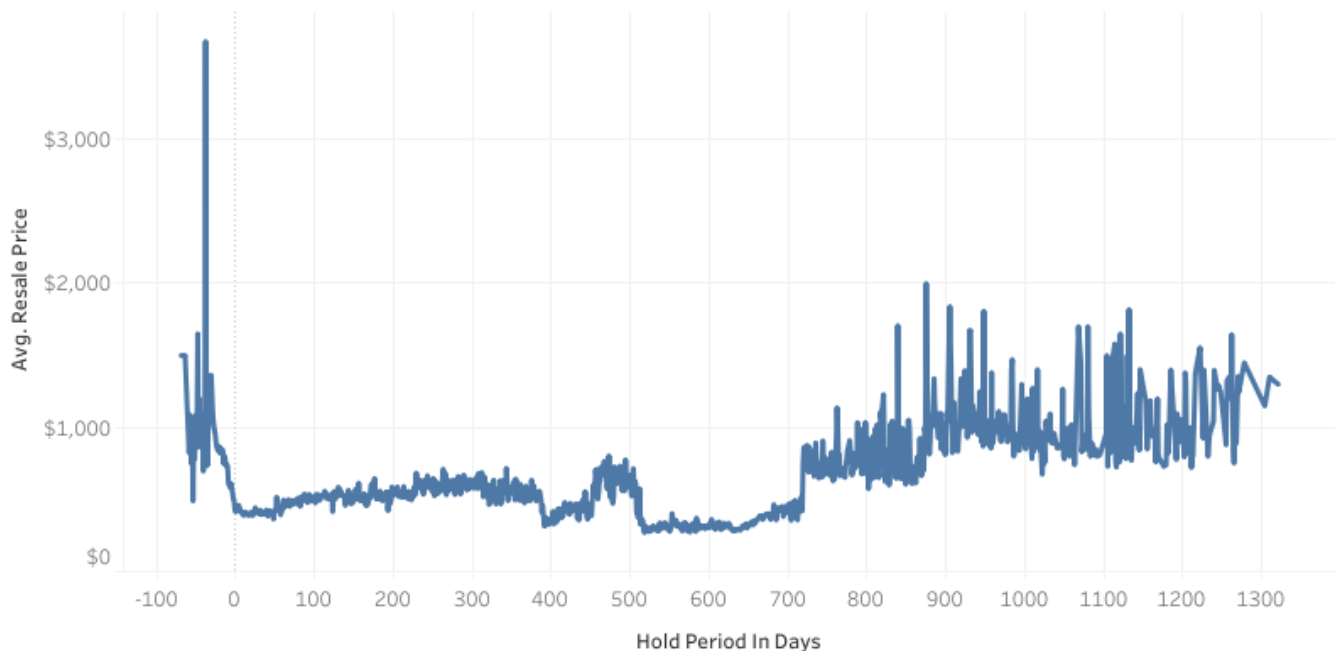
From the above chart we see that negative hold periods i.e when sneakers are resold before they are officially released, have an average resale price of $926 with 5,601 sneakers sold. Sneakers resold between the 0 to 700 days hold period have an average resale price of $473 with 93,074 sneakers sold. Sneakers sold over a 700 day hold period have an average resale price of $913 with 1,281 sneakers sold.. The take away from these observations is that it is more profitable to resell a sneaker before the official release or at least 750 days after release date since quantities are very limited during those periods and as a result driving the resale prices upward.

## 2.4 Regression analysis:
For full details of the regression analysis please refer to the R-Markdown link below

https://0a21e5d7a5644d6da479b6132c9f8c1a.app.rstudio.cloud/?view=rmarkdown

**Summary**

*Linear Regression formula:*

*Model 1*

**Resale_price_usd = 283.144+ Shoe _Size+Quarter + US_Region + Brand + Silhouette**

| Variable | Type | Description |
|---|---|---|
| $283.144 | Intercept | This is the expected average resale price if all the independent variables are 0. |
| Shoe Size | Independent | Refers to the different categories of shoe size in which sneakers are sold. Based on the regression model only shoe sizes 16 and 17 are significant. The remaining shoe sizes are not significant variables for predicting resale price since they all have p-value greater than or close to 0.05. |
| Quarter | Independent | Refers to the quarterly cycle of the year ; Q1, Q2, Q3, Q4. Based on the regression analysis Q2 is not significant since it has a p-value of 0.52537. |
| Silhouette | Independent | Refers to the categorization of the sneaker_names into sneaker models. Based on the regression analysis the silhouette Nike React is not significant since it has a p-value of 0.06308 Additionally the Yeezy 350 silhouette was not included in the model due to singularities. |
| US Region | Independent | Refers to the categorization of the states where sneaker purchase orders were made into the four US Regions i.e West, Northeast, Midwest and South. |
| Resale Price (USD) | Dependent | Refers to the estimated resale price of a sneaker as determined by the independent variables |

The **Multiple R-squared** and **Adjusted R-squared** of **0.5207** and **0.5205** respectively with a **p-value of < 2.2e-16** and a residual standard error of **177.2**. This indicates that about **52%** of the observed variation in resale price can be explained by the model's inputs. The regression model revealed that shoe size was not a significant variable for predicting resale price. To determine if the regression model is improved without the variable "Shoe_SIze" we run a second regression model.

*Model 2*

**Resale_price_usd = 377.457+ Quarter + US_Region + Brand + Silhouette**

The **Multiple R-squared** and **Adjusted R-squared** of **0.5139** and **0.5139** respectively with a **p-value of < 2.2e-16**. The standard error of **178.4** This indicates that about **51%** of the observed variation in resale price can be explained by the model's inputs.

## 2.5 Conclusion:
- What shoe size or sizes are most profitable - The most profitable shoe sizes in descending order are 10, 9, 11, 10.5 and 9.5 with shoe size 10 recording a total profit of about $3MM.
- What brand or designer is most profitable - The most profitable brand or designer is Nike (Off-white) with over $3MM more in profit than Adidas (Yeezy).

- Which region of the United States has most profitability in the resale - The state of California has the highest profit across the US Region of about $5.2MM.
- We observe that the output from both regression models (i.e. including shoe size and not including shoe size) both can explain over half the variation in resale price by the model's input with minimal change in the residual standard error from model 1 to model 2. Meaning, both models can be used to predict resale prices with about 50% accuracy. This in my opinion is not sufficient and would recommend exploring other factors or performing different combinations of factors to achieve a coefficient of determination of at least 70%. It could also be that there is no specific factor or combination of factors that could predict ,with high reliability resale prices.

The output from my analysis indicates that there is profitability in the sneaker resale market, with sneakers reselling on average twice as much as the retail price. Additionally we observed that the quantity of sneakers reselling has increased over time quarter-over-quarter by an average of about 3% indicating signs of growth and increasing demand for sneakers. These insights by no means provide concrete evidence that the sneaker market is worth investing in as a serious investor, however for the casual investor, the sneakerhead or the individual looking to make a quick profit without a huge capital and strenuous process, sneaker reselling might be worth looking into. I am already invested in the sneaker culture and with the insights from my analysis I have decided to test out the reselling market. My focus is however going to be on Nike sneakers since they have a higher resale value on average than Adidas and sell out quicker.

## Potential future analysis:

The following are questions worth exploring in future analysis related to the sneaker resale market which could provide additional insights:

- Are Nike and Adidas the only sneaker brands on the sneaker resale markets? Which sneaker brands make up the largest  proportion of the sneaker resale market and why?
- What does the distribution of sneaker resellers look like by gender and age?
- What does the distribution of sneaker resale prices look like by different designers under the same brand or designers across different brands. Is a designer a more relevant factor in determining sneaker resale price than the sneaker brand?