May 1, 2024
TRINITY UNIVERSITY

ANALYTICS REPORT

Course: Business Analytics Consulting
Term: Spring 2024

1. Introduction

XYZ is an American nonprofit organization that provides employment placement services, job training, and sells second-hand items supplied to the store through donations. XYZ’s business model operates by selling donated goods across its retail stores. XYZ must effectively price and merchandise its diverse inventory to optimize sales revenue. This project analyzes XYZ’s data to provide insights into optimal pricing strategies. Specifically, we aimed to examine how price points correlated with sales volumes and pull rates by stock-keeping unit (SKU). We also evaluated whether production associates’ pricing aligned with ideal price points for maximizing sales. Finally, we planned to assess the impact of recent production process changes across XYZ’s retail locations. These findings will help XYZ refine its pricing policies and production practices to maximize profits and improve the sustainability of their program.

2. Background Information

XYZ retail stores have an organized process from when donations are dropped off at the store to when they are sold on the sales floor. After a donation, employees will pre-sort the items into five basic departments: shoes, accessories, electrical, wares, and books and media. After the pre-sort, a different set of employees will further sort them into specific categories, which is the next level of item grouping. Following that, the items are sorted into a subcategory and an SKU, which is the lowest grouping level. For this project, we only worked with data on wares. Wares do not have a subcategory, but the SKU is based on the type of material the item is made of, which includes paper, glass, wicker, and plastic.

Once all items are correctly sorted, prices are assigned. Pricing is subjectively determined by the employee responsible for producing the item. Each SKU has multiple possible price points for an employee to select from depending on the quality and value of the item. There is also an open pricing option for most SKUs, so if the employee feels that none of the preset price points properly represent the item’s value, they can create their own price point. After pricing, the item is transferred to the sales floor and available for customers to purchase.

This price assigned to an item can affect XYZ’s revenue and operations. Because prices are assigned subjectively, one employee might price a bundle of children’s toys differently than another. This is why we investigated to determine which price points sold best by SKU and which production associates effectively price items that yield optimal revenue.

Another essential aspect of XYZ’s production process is its “color of the week” system. There are five colors in the rotation. Every week, items are tagged with a specific color before being placed on the sales floor. All items tagged with that color stay on the sales floor for five weeks. During their fourth week, items with this color tag are sold at a discount of 50%. If these items do not sell by then, their price will revert to the original amount as they are slowly removed from the sales floor over the fifth week. This means XYZ tags one color each week, sells another at a discounted price, and pulls another color weekly. For example, one week, they might tag all received donations with yellow before transferring them to the sales floor, sell any items with a blue tag for 50% off, and start to pull all items with an orange tag that have been on the sales floor for five weeks. The table below visually represents this process.

	Week 1	Week 2	Week 3	Week 4	Week 5	Week 7	Week 8	Week 9	Week 10
Tagged full price									

Salesfloor full price									
Salesfloor full price									
Discounted									
Pulled									

Table 1. XYZ's Color of the Week

After the color is pulled from the sales floor, it is sent to XYZ's warehouses to be redistributed to another store. This system keeps XYZ's inventory fresh as it receives many donations.

3. Research Questions

This section lays out the six research questions that we plan to answer through our analysis.

RQ1. By the SKU, what are the sales rankings by different price points? How do different price points influence whether a product is sold or pulled?

RQ2. After determining the above question, how does the rate of sales reflect on the pricing associate? Is the associate with the most sales aligned with the most sold price point?

RQ3. How do pricing associates' items do when they reach the sales floor?

RQ4. How well and effectively do employees make pricing decisions and adhere to XYZ's pricing guidelines?

RQ5. How do associates' pricing decisions differentiate between stores?

RQ6. With the last two years of data, how did the change in XYZ's process, which was finalized in 2023, impact the production numbers of each location?

4. Data Overview

Our dataset initially consisted of six tables, each providing information on various aspects of XYZ's operations, including production, pulls, and color rotation schedules. The data covered two years, from January 2022 to January 2024, and offered a comprehensive view of XYZ's activities.

4.1. EmpProdPullThw (Employee Production Pull-Through)

The first table, EmpProdPullThw, displayed valuable information for products pulled or reconciled. The data was aggregated based on the date an item was removed from inventory. This table contained six numeric variables, ten categorical variables, and two date variables. Because pulls reflected the production agent's ability to price effectively, this table helped us determine which production associates need improvement in setting appropriate price points.

4.2. EmpProdSellThw (Employee Production Sell-Through)

The following table, EmpProdSellThw, was almost the same as the previous table but it focused on when the item was sold instead of pulled. This table was aggregated based on the date of sale, production agent, price point, and SKU. It contained five numeric variables, eight categorical variables, and two date variables. This table helped investigate trends between specific price points and successful sell-through rates.

4.3. ProductionSkuAgent (Production SKU Agent)

Table ProductionSkuAgent contained information about employees' production sessions. When an employee started a production session, a production ID was assigned. The same production ID was applied to every item produced in that period, including all types of SKUs and price points. The table provided data on the quantity of items sold, returned, or pulled from different production sessions. It contained ten numerical, seven categorical, and four date variables. This table also helped to detect trends between production associates, price points, and sell-through rates.

4.4. ColorSched (Color Schedule)

The fourth table, ColorSched, contained information about the color schedule used for specific XYZ stores. For every tuple in the table, we could see the sequence of the color schedule, when the color schedule for each store started and ended, which color bin was being produced at each store, which color bin was being pulled at each store, and the specific store ID where the schedule was taking place. The table contained seven variables: one numerical, four categorical, and two date variables. The value this table brought to the project came from its ability to connect the patterns of XYZ's color codes with the trends of when items were being sold. The table showed if the discount on the color of the week items played a role in whether they were sold or pulled. However, we ultimately decided to use this table as a reference for the schedule rather than incorporating it directly into the data analysis process.

4.5. Employee

The fifth table, Employee, provided additional information about each employee and supplemented the employee IDs present in other tables. It contained details such as the employee's first name and last names, which allowed for a more comprehensive understanding of the individuals associated with the employee IDs. The table consisted of ten variables in total, including seven categorical variables and three date variables. While this table did not directly contribute to the core analysis, it served as a valuable reference for linking employee IDs to their respective names, enhancing the interpretability of the results.

4.6. Store

The sixth table, Store, provided additional details about each XYZ store and complemented the store IDs found in other tables. It included information such as the store name and location, which enabled clearer identification of the stores mentioned throughout the dataset. The table contained five variables, including three categorical variables and two date variables. While this table did not play a direct role in answering the core research questions, it served as a valuable reference for linking store IDs to their corresponding names and locations and enhances the context and interpretability of the analysis.

5. Data Cleaning

Our analysis focused on five tables: PullThw, SellThw, ProductionSkuAgent, Employee, and Store. These data sets contained the most relevant information for addressing our research objectives. This section explained the essential steps taken in the data-cleaning process for these tables, including removing redundant variables, recoding variables, handling missing values and outliers, aggregating data, joining tables, and creating new variables. This process was crucial to ensuring data accuracy, data consistency, and enabling meaningful insights to be derived.

5.1. Removing variables

After examining all the tables provided, we identified and dropped irrelevant or redundant variables that did not contribute to the analysis. These included variables related to item IDs, categories, price point types, or date types. For example, we removed *DateLoaded* (date the item was loaded) and *Dateupdated* (date the item was updated) columns from the ProductionSkuAgent table since the *ProdDate* (production date) variable was sufficient for analyzing production. Similarly, *PullAmt* (pull amount), *ReconAmt* (reconciliation amount), and *SoldAmt* (sold amount) were also dropped, as we could retrieve this information from the *TotalQty* (total quantity) columns in SellThw and PullThw tables. Another example was the *PPType* (price point type) variable. This column indicated whether an item was set at an open or pre-set price point. Still, this information was already captured in the *PricePoint* column, where a value of 0 indicated an open price point and any other value indicated a pre-set price point. Therefore, *PPType* was removed from both PullThw and SellThw tables to avoid redundancy. In the Employee and Store tables, we also only kept the name-related variables while dropping all other columns. By doing so, we ensured that Employee and Store tables served their primary purpose of supplementing the employee IDs and store IDs once they were joined with other data sets.

5.2. Recoding variables

Since the *TTLDur* (total duration) variable represented the total sell and pull duration for the entire quantity of items, we determined that it did not provide a meaningful measure for analysis. Instead, calculating the average duration per item would bring more valuable insights. To address this, we renamed the column to *AvgTTLDur* (average total duration per item) and recalculated its values by dividing the original total duration by the total quantity. The calculation was as follows:

$$AvgTTLDur = TTLDur/Qty$$

In addition, we decided to recategorize the *SKU* variable in both PullThw and SellThw tables to enhance the clarity and usefulness of the data. Some categories had misleading names and insufficient observations to provide meaningful insights, so they were merged with the closest related category. Specifically, the “EASTER” and “EasterBasket” values were combined into the “Easter” category, while “bin” and “gaylord” were grouped under the “Wares” category.

5.3. Handling missing values and outliers

After the above steps, we proceeded to identify missing values in the dataset. The missing values only existed in the PullThw and SellThw data sets, specifically in the employee ID and total duration columns. However, these missing values accounted for a relatively small portion of the observations.

To handle the missing *ProdEmpID* (production employee ID) values, we decided to filter out the rows with unknown employees from these two tables. Since our research questions involve analyzing employee performance, rows with an unknown employee contribute little to our analysis. Fortunately, filtering out the missing *ProdEmpID* values also covered the missing *AvgTTLDur* values in the PullThw table. However, some of the missing *AvgTTLDur* values still existed in the SellThw dataset. Nonetheless, these missing values were minimal and accounted for only 1,794 out of approximately 2 million observations. We replaced the missing *AvgTTLDur* with the average value of the corresponding column.

In addition to handling missing values, we also identified and removed observations that were outliers resulting from data input errors. For example, we encountered a row in the PullThw dataset where the number of pulled items was approximately 9 million, significantly higher than the typical values. Another example was an observation in the ProductionSkuAgent table, with the total amount of items an employee produced per day equaling 33,401 items, which was also unrealistic. After careful consideration and consultation with stakeholders, it was determined that these observations were indeed mistakes in data entry. To maintain the integrity and accuracy of our dataset, we decided to remove these observations from those tables.

Besides that, in the PullThw and SellThw tables, we encountered negative values in the *Qty* (quantity), *SoldValue* (sold value), and *IfSoldValue* (value if the item is sold) columns, which led us to question those observations. In the SellThw table, upon closer examination, we discovered that the negative values in that table represented product returns and were not associated with any specific employee ID. After consulting with stakeholders and considering the relatively small proportion of these observations (12,510 out of approximately 2 million), we decided to filter them out of the dataset. In the PullThw table, the negative values indicated unusual cases where items were pulled before they were "produced." These occurrences were infrequent and accounted for a minor portion (only 9 cases out of 300,000). After seeking guidance from stakeholders, we determined that they could be removed from the data set. However, we took an additional step to maintain data accuracy before removing the negative duration cases from the PullThw table. Since the negative duration implied that the items would be produced in the future, we calculated the estimated production date by adding the absolute value of the negative duration to the corresponding *DatePull* (the date the item was pulled). We then located that calculated date in the *ProdDate* variable of the ProductionSkuAgent table and subtracted the total quantity of those pulled items from the *TotalProdAmount* (total production amount). This adjustment ensured that the production quantities in the ProductionSkuAgent table accurately reflected the removal of the items.

In the ProductionSkuAgent table, we also encountered some observations that represented mismatches. We identified an employee named "(not) Nathan (not) Tindel." Upon confirmation from stakeholders that his user ID was once entered incorrectly, leading to his name appearing in that format, we updated that employee ID to his correct ID and restored his name to "Nathan Tindel." Additionally, we found several observations named "OFLE1234", which represented a system login. We decided to rename this observation as "System Login." By carefully addressing these data inconsistencies and taking appropriate actions, we ensured data consistency and accuracy and enhanced our dataset's overall quality and reliability.

5.4. Aggregating tables

To effectively manage the large volume of data, we aggregated the critical tables in the dataset, including PullThw, SellThw, and ProductionSkuAgent. This process involved summarizing the data based on specific dimensions and measures to create a more manageable version of the dataset while preserving the essential information required for our analysis.

For the PullThw table, we aggregated the data using the dimensions of *PullStoreid* (the ID of the store that the item was pulled), *ProdEmpid* (employee ID), *DatePull*, *PricePoint*, *TransType* (transaction type, whether pulled or reconciled), and SKU. The measures aggregated in this process were *TotalQty*, which represents the total quantity of items pulled; *AvgIfSoldValue*, which calculates the average value of the items if they were sold; and *AvgDur*, which computes the average duration of the pull process per item.

Similarly, for the SellThw table, we performed aggregation based on the dimensions of *Storeid* (the store's ID that where the item was sold), *ProdEmpid*, *DateSold*, *SKU*, and *PricePoint*. The *measures* aggregated in this table were *TotalQty*, which represents the total quantity of items sold; *AvgSoldValue*, which calculates the average value of the sold items; and *AvgDur*, which determines the average duration of the selling process.

Due to its extensive size, the ProductionSkuAgent table also required aggregation. We calculated the measures of *TotalProdAmt* (total production amount) based on the dimensions of *Emplid* (employee ID), *OrgStoreid* (original store ID), *SKU*, and *ProdDate* (production date).

By performing these aggregations, we successfully reduced the dataset size while keeping the crucial information needed for our analysis. This process not only improved the data's manageability but also facilitated faster and more efficient querying and analysis.

5.5. Joining tables

We also performed a series of table joins to improve the interpretability and context of the data. Before joining, we created a new variable called *FullName* in the Employee table by concatenating the *FirstName* and *LastName* columns. This step allowed us to have a single, comprehensive variable that represents the complete name of each employee. After creating the *FullName* variable, we joined the Employee and Store tables with the three main tables in our dataset: PullThw, SellThw, and ProductionSkuAgent. The purpose of this join was to incorporate the *FullName* of the employees and the corresponding *StoreName* where they work into these tables.

5.6. Creating new variables and tables

To assess the pricing effectiveness of associates, we created a new variable called *PricingEffectiveness* in the SellThw table. This variable represented the ratio of an item's actual average selling price to its set price point, expressed as a percentage. It measured how well associates price their items and how closely the selling price aligns with the initial pricing. We calculated this by dividing *AvgSoldValue* (the actual value at which an item was sold) by the product of the corresponding set *PricePoint* and *TotalQty* (total quantity of items sold). We then multiplied the result by 100 to express it as a percentage. For example, if an item was sold at $9.5 and its *PricePoint* was $10, the *PricingEffectiveness* would be calculated as follows:

*PricingEffectiveness = ($9.5 / $10) * 100 = 95%*

So, if an employee has a *PricingEffectiveness* of 100%, it indicates that the items they priced were sold at the same price point, without any discounts. In other words, the selling price perfectly matched the initial pricing. On the other hand, if an employee has a *PricingEffectiveness* of 90%, it suggests that, on average, their items were sold at 90% of the initial price point. Which means the items were sold at a slightly discounted price compared to the original pricing. Therefore, creating this new variable allowed us to evaluate the effectiveness of pricing strategies and generate data-driven findings to optimize pricing for XYZ.

In addition, to evaluate how closely employees adhere to XYZ's pricing guidelines, we created a new table called EmpPricingVariance that calculated the pricing variances of each employee. The *Pricing_Variable* column in this table aims to capture the variability in price points set by employees for each SKU. We hypothesized that if employees have a low variance in price points by SKU, they might be defaulting to the same price point for items without thinking it through. On the other hand, high pricing variance by SKU indicates that an employee is putting much consideration into pricing items. To calculate *Pricing_Variance*, we grouped the data in the ProductionSkuAgent table by stores, employees, SKUs, and production dates, then determined each group's standard deviation of the price points. This quantified the spread of the price points assigned by employees for a specific SKU on a specific date. However, it should be noted that there were instances where an employee only produced 1 item for a particular SKU on a date. Therefore, we decided not to include those cases in our table. The resulting table contains the pricing variance and total quantity produced for each store, employee, date, and SKU. This information can be valuable for identifying training needs, evaluating employee performance, and optimizing pricing strategies.

6. Variable Descriptions

After completing the data cleaning process, we were left with the following tables for our analysis: PullThw, SellThw, ProductionSkuAgent, and EmpPricingVariance. These data sets contained the variables and data points necessary to address our research questions and derive meaningful insights. To provide a clear understanding of the data structure and contents, **Error! Reference source not found.** below presented a comprehensive description of each variable within these tables. Table 2 defined the variables and specified the respective table to which each variable belonged.

Variable Name	Description	Table
ProdDate	Date the item was produced.	ProductionSkuAgent
Emplid	Employee ID number.	ProductionSkuAgent
OrgStoreid	The store ID where the item was produced.	ProductionSkuAgent
TotalProdAmount	Total quantity of items produced in a production session.	ProductionSkuAgent
Storeid	The store ID that sold the item.	SellThw
ProdEmpid	Employee ID number.	SellThw, PullThw, EmpPricingVariance
DateSold	The date an item was sold.	SellThw, EmpPricingVariance
SKU	The stock-keeping unit category of an item.	SellThw, PullThw, EmpPricingVariance
PricePoint	The price an employee sets for an item.	SellThw, PullThw
StoreName	The name of the store.	SellThw, PullThw, ProductionSkuAgent, EmpPricingVariance
AvgSoldValue	The average sold value of an item.	SellThw
PullStoreid	The store ID of a pull item.	PullThw
TransType	The transaction type.	PullThw
DatePull	The date an item was pulled.	PullThw
TotalQty	The total quantity of an item.	SellThw, PullThw
AvgDur	The average duration of an item from production to pull or sale.	SellThw, PullThw
FullName	The full name of an employee.	SellThw, PullThw, ProductionSkuAgent, EmpPricingVariance
AvgIfSoldValue	The average sold value per item if it is sold at full price.	PullThw
PricingEffectiveness	Percent out of 100 that determines how effectively an employee prices an item.	SellThw
Pricing_Variance	The variability of price points produced within a SKU	EmpPricingVariance

Table 2. Variables Description

7. Descriptive Statistics

In this section, we present the key descriptive statistics for each of our tables' numeric and categorical variables. These statistics provide an overview of the data distribution and characteristics and help us gain insights into the variables' nature and their potential impact on the analysis.

7.1. Numeric Variables

We calculated descriptive statistics for numeric variables, including mean, standard deviation, maximum, and minimum values. Table 3 through Table 6 below illustrate these statistics for each numeric variable in our four tables: ProductionSkuAgent, PullThw, SellThw, and EmpPricingVariance. Upon examining the skewness of the numeric variables, we found that all variables had a right-skewed distribution. This means that most of the data points were concentrated on the left side of the distribution, with a long tail extending to the right. The right-skewed nature of these variables could be attributed to two main factors: XYZ's overall philosophy of encouraging affordable prices and the diverse range of time that items remain on the sales floor.

To illustrate this, let's consider the *AvgDur* variable in the PullThw table (Table 4). Most items followed XYZ's standard schedule and were pulled within the expected timeframe. However, there were many exceptional cases where items remained on the floor for extended periods, deviating from the typical pull schedule. Another potential explanation was that XYZ's affordable pricing strategy and the diverse range of item types in the Wares category led to most items being priced at low prices. However, there were instances where certain items were priced at thousands of dollars. Since these high-priced items were relatively uncommon, they created a right-skewed distribution. Given the reasonable explanation for the right-skewed nature of the numeric variables, we chose not to include the skewness score in each of the descriptive statistics table.

7.1.1. ProductionSkuAgent Table

Table 3 below provides insights into the productivity of XYZ's employees. The average production quantity of approximately 29 items per SKU for each employee on a day suggested that XYZ employees had a considerable level of output. However, the standard deviation of 63.47 items, along with the wide range of *TotalProdAmt* values, from a minimum of 1 to a maximum of 2,253 items, indicated notable variation in the number of items produced across different dates, different SKUs, and different employees. Some employees might only produce one item on a given day, possibly due to specialized tasks, low demand, or unique items. On the other hand, some employees had highly productive days where they generated a substantial volume of items.

Numeric Variable	Mean	Standard Deviation	Max	Min	Units
TotalProdAmt	29.04	63.47	2,253.00	1.00	Items

Table 3. Descriptive Statistics for Numeric Variables ProductionSkuAgent Table

7.1.2. PullThw Table

Table 4 displays summary statistics regarding the PullThw table. All items that were pulled from the sales floor in the past two years are represented in this table. The *TotalQty* variable represented the total amount of items of a specific SKU pulled on a particular date per employee. On average, 4.11 SKU items were pulled, with a standard deviation of 8.51 items. The maximum quantity pulled for a single SKU on a specific date per employee was 440 items, while the minimum was 1 item. These statistics provided insights into the pulling patterns and characteristics of items at XYZ. The average total quantity pulled (4.11 items) suggested a relatively small pulling volume per SKU on specific dates by each employee. However, the standard deviation of 8.51 items indicated some variability in the pulling quantities.

AvgIfSoldValue represented the average potential revenue of all items within a specific SKU on average on a particular date sold at full price. The mean value of this variable was $17.75. It implied that, on average, the items being pulled had a moderate value if sold at their full price. However, the significant standard deviation of $87.01 and the wide range of values (from $0.10 to $24,900.00) indicated the presence of both low-value and high-value items in the pulling mix.

AvgDur showed an item's average time on the sales floor in days. The average duration was 17.61 days (about 2 and a half weeks), and the standard deviation was 18.68 days. The maximum number of days an item stayed on the sales floor before being pulled was 1,406 days (almost 4 years), while the minimum duration was 0 days. It is important to note that there were two transaction types in the PullThw table: pulls and reconciliations. Pulls refer to items that had yet to be sold within the designated timeframe and were removed from the sales floor, typically after 30 days or more. On the other hand, reconciliations are items removed from the sales floor due to discrepancies or errors in the inventory, typically 3 days or less. This factor, along with a fire that occurred at one of XYZ's locations a while ago, explained the presence of both short-term and long-term durations in the *AvgDur* variable.

Numeric Variable	Mean	Standard Deviation	Max	Min	Units
TotalQty	4.11	8.51	440.00	1.00	Items
AvgIfSoldValue	17.75	87.01	24,900.00	0.10	Dollars
AvgDur	17.61	18.68	1,406.00	0.00	Days

Table 4. Descriptive Statistics for Numeric Variables in PullThw Table

7.1.3. SellThw Table

Table 5 summarizes key statistics for the SellThw table, which contained all items sold in the past two years. The *TotalQty* variable represented the total quantity sold within each SKU on a particular date by an employee. The average TotalQty of items sold was 4.20, with a standard deviation of 7.61. The maximum quantity sold was 263 items, and the minimum was 1. *AvgSoldValue* was the average revenue of sold items per employee within a specific SKU on a particular date. This variable differs from *AvgIfSoldValue* in the PullThw table, as *AvgSoldValue* represented the actual selling price. In contrast, *AvgIfSoldValue* represented the potential selling price if the item had been sold at full price. The average value of this variable was $15.35, and the standard deviation was $22.85.

Interestingly, while the means of *AvgSoldValue* and *AvgIfSoldValue* were similar, it is important to note that the standard deviation of *AvgSoldValue* in the SellThw table was significantly lower compared to *AvgIfSoldValue* in the PullThw table. In addition, the maximum value of *AvgSoldValue* was $906.63, and the minimum was $0.13, which indicated that the range of *AvgSoldValue* was not as wide as the range of *AvgIfSoldValue*. This suggested that the actual selling prices of items tend to have less variability than the potential selling prices.

The *AvgDur* variable in the SellThw table represented the average duration, in days, that an item remained on the sales floor before being sold. The mean *AvgDur* is 5.27 days with a standard deviation of 10.88 days (approximately 1 and a half weeks). Notably, the mean and standard deviation of *AvgDur* in the SellThw table were considerably smaller than the corresponding statistics in the PullThw table, which indicated that items that were sold typically spend less time on the sales floor than those that were pulled. However, the maximum number of days an item spent on the sales floor before being sold was 1,258 days (about 3 and a half years). This suggested exceptional cases where items remain on the sales floor for an extended period without being pulled or sold. Like *AvgDur* in the PullThw table, such instances were likely due to various factors, such as the fire that happened at one of the XYZ's stores a while ago, item uniqueness, or oversight in the inventory management process. On the other hand, the minimum *AvgDur* value of 0 days indicated that some items were sold on the same day they were put on the sales floor.

PricingEffectiveness represented the ratio of the actual selling price of an item to its set price point, expressed as a percentage. It measured how closely the selling price aligns with the intended pricing strategy. The average pricing effectiveness was 88.75%, with a standard deviation of 15.51%. This suggested that, on average, the selling prices were close to the set price points and indicated effective pricing strategies. The maximum pricing effectiveness of 100% indicated instances where items were sold at their exact price point without any discounts. On the other hand, the minimum pricing effectiveness of 20% suggested cases where items were sold at a significant discount compared to their set price point. Low pricing effectiveness metrics could signal potential issues in pricing decisions or employee adherence to pricing guidelines.

Numeric Variable	Mean	Standard Deviation	Max	Min	Units
TotalQty	4.20	7.61	263.00	1.00	Items
AvgSoldValue	15.35	22.85	906.63	0.13	Dollars
AvgDur	5.27	10.88	1,258.00	0.00	Days
PricingEffectiveness	88.75	15.51	100.00	20.00	Percentage (0-100)

Table 5. Descriptive Statistics for Numeric Variables in SellThw Table

7.1.4. EmpPricingVariance Table

Table 6 summarizes statistics about the *Pricing_Variance* variable in the EmpPricingVariance table. It represented the standard deviation of the price points assigned by employees for items within the same SKU. It served as an indicator of the level of consideration and thought put into pricing decisions by employees.

The average *Pricing_Variance* was $2.51. This indicated that, on average, the price points assigned by employees for items within the same SKU deviated by $2.51 from the mean price point. The maximum *Pricing_Variance* value of $69.41 represented cases where employees show high variability in their pricing decisions within a SKU. This could indicate that employees actively consider various factors and apply their judgment when pricing items rather than relying on a standard or default price point. Conversely, the minimum *Pricing_Variance* value of $0.06 suggested instances where employees show minimal variation in their pricing decisions within an SKU. This could result from employees consistently applying a set price point with little individual consideration.

Numeric Variable	Mean	Standard Deviation	Max	Min	Units
Pricing_Variance	2.51	1.89	69.41	0.06	Dollars

Table 6. Descriptive Statistics for Numeric Variables in EmpPricingVariance Table

7.2. Categorical Variables

We comprehensively analyzed categorical variables to gain insights into their distribution and characteristics. We started by calculating the total count of observations and the count of unique values for each categorical variable. This provided us with a high-level overview of the data and helps identify variables with many distinct categories. After obtaining the overall statistics, we delved deeper into the individual categories within each variable. We calculated the count and percentage of observations for each category. This allowed us to understand the relative frequency and importance of different levels within a variable. We decided to include all the categorical variables in our report except for *FullName* since *FullName* contains thousands of unique values that represent individual employee names. We chose to exclude it from the detailed analysis to maintain a focus on variables with more meaningful categories. In the tables below, we sorted the variables in descending order by total count. This created a better understanding of each variable’s most frequently occurring categories and highlighted the dominant level.

7.2.1. ProductionSKUAgent Table

This section summarizes the count and percentage of categories for both the overall table and each categorical variable in the ProductionSkuAgent table. Table 7 below summarizes the categorical variables for the ProductionSKUAgent table. There were two categorical variables. First, the table had *FullName*, the full name of the employees who created entries for the table. For the *FullName* variable, there were 800 unique employee names, meaning there were 800 unique employee entries. The second variable in the table is *StoreName*, which included the names of the stores where the items were produced. The *StoreName* variable had 24 unique names, which meant that the dataset accounts for production from 24 of XYZ's stores. In addition, produced items also fell within 17 SKUs in total.

	Total Count	Unique Categories
FullName	346,572	800
StoreName	346,572	24
SKU	346,572	17

Table 7. Descriptive Statistics for Categorical Variables in ProductionSKUAgent Table

Table 8 below provides a breakdown of the entries recorded from each store in the ProductionSKUAgent table. It is also worth noting that the percentages for each store were relatively evenly distributed, with most stores contributing between 3% to 5% of the total entries, which suggested a balanced distribution of production activities across most XYZ's store locations. The Austin Hwy store was the most represented location, accounting for 5.98% of the total entries. On the other hand, the Bandera store had the least representation, making up only 1.64% of the total entries. The remaining stores show varying levels of representation, ranging from 5.90% for the Bulverde store to 1.74% for the Bulverde North store. This distribution highlighted the diversity of production activities across different store locations and emphasizes the importance of considering store-specific factors when analyzing production patterns and performance.

StoreName	Count	Percentage
Austin Hwy	2,982	05.98
Bulverde	2,942	05.90
Blanco	2,846	05.71
Bitters	2,819	05.57
DeZavala	2,795	05.56
Blanco North	2,568	05.15
Seguin	2,503	05.02
Evans	2,437	04.89
Gateway	2,317	04.65
Cibolo	2,303	04.62
Culebra	2,282	04.58
Potranco	2,265	04.54
New Braunfels	2,190	04.39
Goliad	2,069	04.15
Marbach	2,027	04.07
Fredericksburg	2,010	04.03
South Park	1,874	03.76
Laredo	1,760	03.52
Summit	1,450	02.91
WW White	1,357	02.72
Kerrville	1,332	02.67
Commerce	1,032	02.07
Bulverde North	868	01.74
Bandera	817	01.64

Table 8. Descriptive Statistics for StoreName Variable in ProductionSKUAgent Table

Table 9 below presents the descriptive statistics for the SKU variable in the ProductionSKUAgent table. The data in this table revealed a relatively balanced distribution of entries across the different SKU categories. Metal had the highest representation, accounting for 6.90% of the total entries, closely followed by Toys at 6.79% and Plastic at 6.68%. The least represented SKU is Large Wares, making up 3.96% of the total entries. The remaining SKUs, such as Pictures/Frames, Wood, Vases/Figurines, Sports, Wares, Office, Dishes, Cups/Glass, Bed/Bath, Seasonal, Wicker, Pots/Pans, and Games, had percentages ranging from 4.44% to 6.63%, which indicated a fairly even distribution of production activities across these categories. This balanced distribution suggested that XYZ handled a diverse range of product categories in its production processes.

SKU	Count	Percentage
Metal	23,908	06.90
Toys	23,528	06.79
Plastic	23,155	06.68
Pictures/Frames	22,993	06.63
Wood	22,220	06.41
Vases/Figurines	21,732	06.27
Sports	21,523	06.21
Wares	20,973	06.05
Office	20,610	05.95
Dishes	20,115	05.80
Cups/Glass	20,112	05.80
Bed/Bath	19,661	05.67
Seasonal	19,453	05.61
Wicker	19,070	05.50
Pots/Pans	18,389	05.31
Games	15,396	04.44
Large Wares	13,734	03.96

Table 9. Descriptive Statistics for SKU Variable in ProductionSKUAgent Table

7.2.2. PullThw Table

This section summarizes the count and percentage of categories for both the overall table and each categorical variable in the PullThw table. Table 10 summarizes the categorical variables for the PullThw table. Each variable had a total count of 302,845 entries. The table contains the variables of *SKU*, *FullName*, *StoreName*, *TransType*, and *PricePoint*.

	Total Count	Unique Categories
SKU	302,845	17
FullName	302,845	692
StoreName	302,845	24
TransType	302,845	2
PricePoint	302,845	1,598

Table 10. Descriptive Statistics for Categorical Variables in PullThw Table

Table 11 below shows the descriptive statistics for the *SKU* variable. The data revealed that there is one dominant SKU value: Wares. With a count of 54,894 entries, Wares accounted for 18.13% of all the values in the table. This suggested that a significant portion of the pulled items fall under the Wares category. Cups/Glass was the second most common SKU value, making up 8.77% of the dataset. Vases/Figurines, Plastic, Dishes, and Toys each contributed between 6% to 7% of the total entries. The remaining SKU values, such as Pictures/Frames, Seasonal, Metal, Wood, Office, and Bed/Bath, had relatively similar percentages, ranging from 4.58% to 6.40%. These categories collectively made up a substantial portion of the pulled items. On the lower end, the SKU values of Sports, Pots/Pans, Games, Wicker, and Large Wares had smaller percentages, ranging from 1.30% to 2.84%. While these categories may be less frequently pulled, they still contributed to XYZ's the overall diversity of items handled.

SKU	Count	Percentage
Wares	54,894	18.13
Cups/Glass	26,566	08.77
Vases/Figurines	22,512	07.43
Plastic	21,409	07.07

Dishes	21,299	07.03
Toys	19,729	06.51
Pictures/Frames	19,376	06.40
Seasonal	18,122	05.98
Metal	17,607	05.81
Wood	17,025	05.62
Office	16,996	05.61
Bed/Bath	13,864	04.58
Sports	8,591	02.84
Pots/Pans	8,367	02.76
Games	6,292	02.08
Wicker	6,264	02.07
Large Wares	3,932	01.30

Table 11. Descriptive Statistics for SKU Variable in PullThw Table

Table 12 lists the descriptive statistics for the StoreName variable in the PullThw table. The data was spread out within this table, as no single value comprised more than 10% of the entire dataset. The store with the highest percentage of pulled items was Laredo, accounting for 7.88% of the total count. On the other hand, the stores with the lowest percentages of pulled items were Evans (1.70%), Bulverde North (1.78%), Fredericksburg (2.16%), Bandera (2.30%), and Commerce (2.44%). These stores had relatively less pulled items compared to the rest of the stores.

StoreName	Count	Percentage
Laredo	23,892	07.88
WW White	20,757	06.85
Gateway	18,144	05.99
Culebra	17,282	05.71
South Park	16,532	05.46
Bitters	16,283	05.37
Potranco	15,412	05.09
DeZavala	14,762	04.87
Blanco	14,000	04.62
Goliad	13,446	04.44
Marbach	13,198	04.36
Cibolo	12,786	04.22
Seguin	12,606	04.16
New Braunfels	11,695	03.86
Summit	10,507	03.47
Austin Hwy	10,443	03.45
Bulverde	10,344	03.42
Blanco North	9,710	03.21
Kerrville	9,560	03.16
Commerce	7,390	02.44
Bandera	6,980	02.30
Fredericksburg	6,545	02.16
Bulverde North	5,407	01.78
Evans	5,158	01.70

Table 12. Descriptive Statistics for StoreName Variable in PullThw Table

Table 13 shows the descriptive statistics for the *TransType* variable in the PullThw table. The statistics from this table show that most of the items in the PullThw table were pulled more than reconciled. Pulls made up 88% of all entries in the PullThw table.

TransType	Count	Percentage
Pull	266,526	88.00
Recon	36,382	12.00

Table 13. Descriptive Statistics for TransType Variable in PullThw TableTable 14 shows the descriptive statistics for the top 10 most frequent price points in the PullThw table. The *PricePoint* variable contains 1,598 unique values, so we only included the top 10 price points in frequency for the descriptive statistics. The most common price point shown in the table was $2.99, 23.66% of the dataset. $3.99 and $1.99 are the second and third most popular price points, accounting for 16.97% and 16.77% respectively.

PricePoint	Count	Percentage
$2.99	71,651	23.66
$3.99	51,403	16.97
$1.99	50,777	16.77
$4.99	31,708	10.47
$5.99	17,441	05.76
$7.99	9,693	03.20
$9.99	9,538	03.15
$6.99	8,607	02.84
$2.49	7,178	02.37
$14.99	4,485	01.48

Table 14. Descriptive Statistics for PricePoint Variable in PullThw Table (Top 10)

7.2.3. SellThw Table

This section summarizes the count and percentage of categories for both the overall table and each categorical variable in the SellThw table. Table 15 shows the overview of categorical variables in the SellThw table. There were four categorical variables: *SKU*, *StoreName*, *FullName*, and *PricePoint*. All variables had a total of 1,437,849 entries. All 17 *SKU* levels were represented in the table. For *StoreName*, all 24 of the stores were represented in the dataset. Like the PullThw table, 892 unique employees were represented in the *FullName* variable. The *PricePoint* variable contained 119 unique categories in the SellThw table.

	Total Count	Unique Categories
SKU	1,437,849	17
StoreName	1,437,849	24
FullName	1,437,849	892
PricePoint	1,437,849	119

Table 15. Descriptive Statistics for Categorical Variables in SellThw Table

Table 16 shows the descriptive statistics for the *SKU* variable in the SellThw table. Like the PullThw table, data revealed that Wares was the dominant SKU category, accounting for 20.34% of the total entries. This suggested that a significant portion of the items sold fell under the Wares category, indicating its popularity among customers. The remaining SKUs had a more balanced distribution, with percentages ranging from 2.38% to 7.13%. Plastic, Cups/Glass, Toys, Metal, Vases/Figurines, and Dishes each contributed between 6% to 7% of the total entries, which highlighted their substantial presence in the sold items. Other SKUs such as Wood, Pictures/Frames, Bed/Bath, Office, Seasonal, Pots/Pans, and Sports have percentages ranging from 3.77% to 5.35%, showing a relatively even distribution of sales across these categories. On the lower end, Large Wares, Wicker, and Games had percentages between 2.38% to 3.33%, suggesting a smaller contribution to the overall sales volume.

SKU	Count	Percentage
Wares	292,456	20.34

Plastic	102,562	07.13
Cups/Glass	97,147	06.76
Toys	97,067	06.75
Metal	95,930	06.67
Vases/Figurines	88,735	06.17
Dishes	87,794	06.11
Wood	76,899	05.35
Pictures/Frames	74,234	05.16
Bed/Bath	66,899	04.65
Office	65,926	04.59
Seasonal	59,660	04.15
Pots/Pans	58,258	04.05
Sports	54,171	03.77
Large Wares	47,914	03.33
Wicker	37,945	02.64
Games	34,252	02.38

Table 16. Descriptive Statistics for SKU Variable in SellThw Table

Table 17 below shows the descriptive statistics for *StoreName* in the SellThw table. The distribution of entries for the table was even, with no value making up more than 10% of total entries. The most represented store in the dataset was Gateway with 5.82% of all entries of a sold item. The stores with the least amount of representation were Bulverde North and Commerce. They made up only 1.45% and 1.47% of all entries in the SellThw table.

StoreName	Count	Percentage
Gateway	83,724	05.82
Fredericksburg	81,110	05.56
Potranco	76,376	05.31
Blanco North	74,532	05.18
Bulverde	74,166	05.16
Culebra	72,177	05.02
Austin Hwy	71,424	04.97
Blanco	71,036	04.94
Bitters	68,924	04.79
Cibolo	65,703	04.57
New Braunfels	62,590	04.35
Summit	61,472	04.28
Goliad	61,082	04.25
DeZavala	60,409	04.20
WW White	59,914	04.17
Marbach	59,713	04.15
Seguin	56,873	03.96
Evans	55,858	03.88
Laredo	52,417	03.65
Kerrville	49,386	03.43
South Park	47,979	03.33
Bandera	29,041	02.02
Commerce	21,152	01.47
Bulverde North	20,791	01.45

Table 17. Descriptive Statistics for StoreName Variable in SellThw Table

outlines the top 10 most common price points in the SellThw table. The PricePoint variable contained 119 unique categories. To be concise, we only showed the top 10 most frequent price points in this table. Like the PullThw table, $2.99 was the most common price point, making up approximately 22.12% of the dataset. $3.99 and $1.99 followed as the second and third most common price points.

PricePoint	Count	Percentage
$2.99	318,011	22.12

$3.99	268,348	18.66
$1.99	195,174	13.57
$4.99	189,423	13.17
$5.99	124,387	08.65
$7.99	88,422	06.15
$9.99	81,264	05.65
$6.99	59,653	04.15
$14.99	30,305	02.11
$8.99	20,317	01.41

Table 18. Descriptive Statistics for PricePoint Variable in SellThw Table (Top 10)

7.2.4. EmpPricingVariance Table

This section summarizes the descriptive statistics of each categorical variable in the EmpPricingVariance Table. Table 19 summarizes the total count and unique categories for the *SKU*, *FullName*, and *StoreName* categorical variables in the EmpPricingVariance Table. Each variable had a total count of 379,369, meaning there were 379,369 entries in this table. The SKU refers to the Stock Keeping Unit designation of each item. This dataset had 17 different choices for an item's *SKU*. *FullName* refers to the full name of an employee which indicateed those 733 unique employees made entries in this table. *StoreName* refers to the unique store names in the dataset. The table recorded entries from 24 stores.

Name	Total Count	Unique Categories
SKU	379,369	17
FullName	379,369	733
StoreName	379,369	24

Table 19. Descriptive Statistics for Categorical Variables in EmpPricingVariance Table

Table 20 below presents the count and percentage for all the different SKU types in the EmpPricingVariance Table. It had a very similar pattern to SellThw table's SKU. The data revealed that the dominant SKU category is Wares. Wares accounted for 13.48% of all SKUs in the dataset, making it the most prominent SKU. Meanwhile, Large Wares, Wicker, and Games had the least presentation, ranging from only 2.44% to 2.97%.

SKU	Count	Percentage
Wares	51,130	13.48
Plastic	28,466	07.50
Toys	28,237	07.44
Cups/Glass	28,087	07.40
Metal	27,389	07.22
Vases/Figurines	26,729	07.05
Dishes	25,083	06.61
Pictures/Frames	22,835	06.02
Wood	22,359	05.89
Bed/Bath	19,219	05.07
Office	18,950	05.00
Pots/Pans	17,627	04.65
Seasonal	17,620	04.64
Sports	14,855	03.92
Large Wares	11,268	02.97
Wicker	10,244	02.70
Games	9,271	02.44

Table 20. Descriptive Statistics for SKU in EmpPricingVariance Table

In the table below, Table 21, we observed that the number of entries by store varied with no store representing more than 10% of the total dataset. The most represented store was Fredericksburg with

6.07% of the total data and the least represented store was again Bulverde North and Commerce, which respectively made up only 1.48% and 1.28% of the total data.

StoreName	Count	Percentage
Fredericksburg	23,062	06.07
Gateway	21,824	05.75
Potranco	20,799	05.48
Culebra	20,025	05.27
Blanco North	19,767	05.20
Bulverde	19,366	05.09
Austin Hwy	18,655	04.91
Blanco	18,343	04.82
Bitters	18,059	04.75
Cibolo	17,412	04.58
DeZavala	16,902	04.46
Summit	16,876	04.45
Marbach	16,154	04.44
Goliad	15,927	04.25
New Braunfels	15,869	04.19
WW White	15,633	04.18
Seguin	14,457	04.12
Laredo	13,359	03.81
Evans	13,339	03.51
South Park	12,764	03.51
Kerrville	12,425	03.27
Bandera	8,336	02.19
Bulverde North	5,636	01.48
Commerce	4,845	01.28

Table 21. Descriptive Statistics for StoreName in EmpPricingVariance Table

8. Analytical Methods

This section describes our analytical methods used to answer each proposed research question. All the research questions are answered through findings in the dashboard we created: ***XYZ Performance Tracker.***

We approached the first two research questions from a price point perspective. For the first research question, we compared sold and pulled items by price point. To further analyze this question, we investigated top selling SKUs at the most sold price points. To answer the second research question, we determined top employees based on their sell through rate, which is a calculation that computes the percentage of an employee's produced items that were sold. From that, we examined the pricing pattern of the employee that had the highest sell-through rate to see if their top price points aligned with the overall pattern, and if their top price points also fell under top selling SKUs.

Next, the third and fourth research questions are related to employee metrics. We used the pricing effectiveness variable that we created to determine employees with the most effective pricing tactics for the third research question. The fourth research question required us to analyze the relationship between pricing variance and pricing effectiveness to determine if employees with higher pricing variance tend to have a higher pricing effectiveness.

Lastly, our methods for the fifth and sixth research question compare sales metrics between XYZ store locations. We compared average pricing effectiveness between employees at different store locations for the fifth research question. For the final research question, we compared the XYZ store's changes in production, sales, and pulls before and after the production process change that occurred around May 2023.

9. Analytical Results and Discussion

In this section, we present the key findings derived from our data analysis as well as our discussion of the findings. We discovered meaningful insights that address the core research questions of this project. To make these insights more accessible and actionable, we created a dashboard titled ***XYZ Performance Tracker.*** This dashboard serves as a practical tool that XYZ can utilize to monitor key metrics, track performance, and make data-driven decisions. By presenting the findings in a visually intuitive and interactive format, we hope that the dashboard empowers stakeholders to easily explore and analyze the data, as well as to identify areas of strength and opportunities for improvement.

Based on these findings and the dashboard, we also aim to provide a set of recommendations and solutions in the hope that XYZ can optimize pricing strategies, enhance employees' productivity, and improve stores' overall performance.

9.1. Research Question 1

By the SKU, what are the sales rankings by different price points? How do different price points influence whether a product is sold or pulled?

9.1.1. Findings

To address this research question, we created Figure 1 and Figure 2 to show the top 10 price points based on total items sold and total items pulled. In both graphs, the y-axis represents the total amount of items sold/pulled, while the x-axis represents the top 10 price points. These graphs allowed us to examine at which price points items were sold the most and at which price points items were pulled the most. Overall, there is not much difference between them. The price points that appear in both graphs are largely the same, with a few exceptions. The price point of $2.49 only exists in Figure 2 (pull data), while $1.09 appears only in Figure 2 (sell data). It is worth noting that both graphs are dominated by the price points of $1.99, $2.99, and $3.99. This suggests that these were the most popular price points for both sold and pulled items. This finding can be attributed to XYZ's nature of affordable prices, and the fact that a significant number of items are produced at these price points.

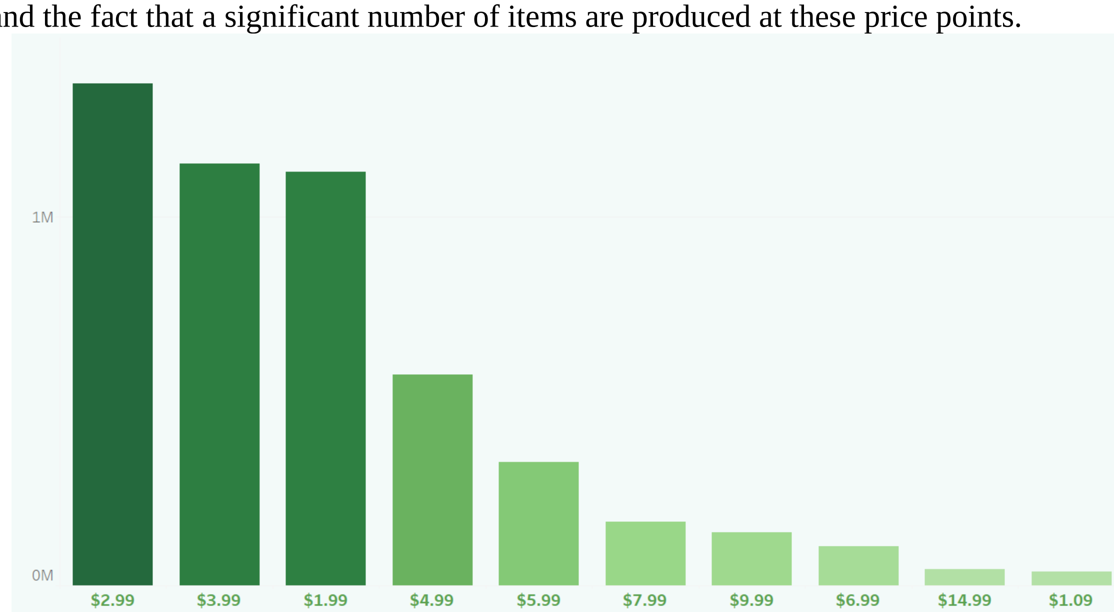


Figure 1. Top 10 Price Points with the Highest Number of Items Sold

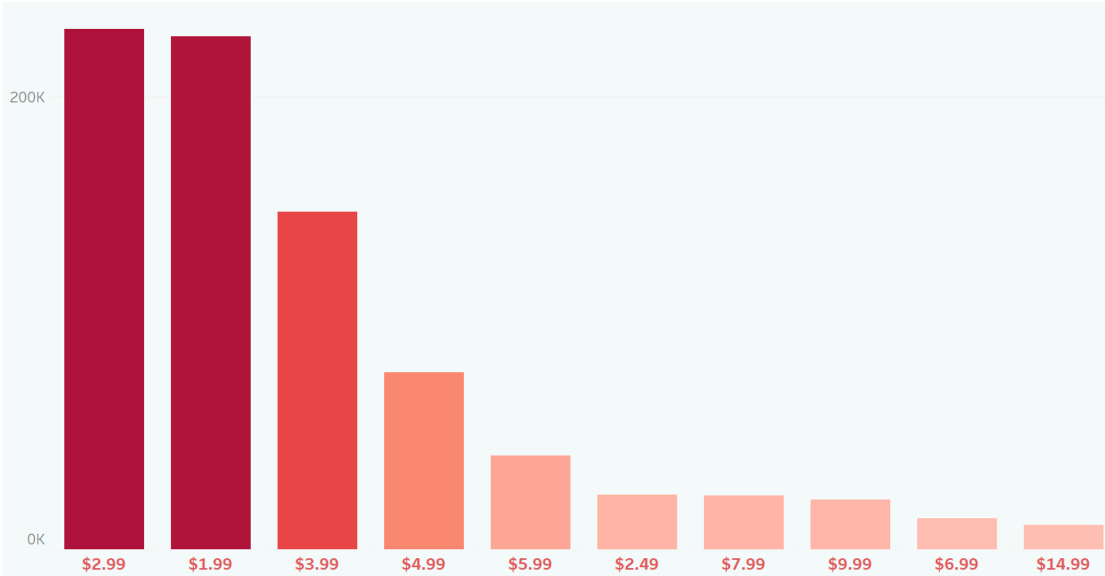


Figure 2. Top 10 Price Points with the Highest Amount Items Pulled

To further analyze the sales performance at these specific price points, we provided a breakdown of the top-selling SKUs and their corresponding quantities sold in Table 22 and Table 23:

Price Point	Details
$1.99	• **Wares**: 644,271 items sold • **Cups/Glass**: 187,646 items sold • **Plastic**: 111,614 items sold
$2.99	• **Wares**: 626,125 items sold • **Cups/Glass**: 128,356 items sold • **Plastic**: 114,085 items sold
$3.99	• **Wares**: 486,436 items sold • **Toys**: 166,153 items sold • **Vases/Figurines**: 155,399 items sold

Table 22. Top selling-SKUs at top 3 price points

Price Point	Details
$1.99	• **Cups/Glass**: 76,741 items pulled • **Wares**: 54,479 items pulled • **Plastic**: 30,479 items pulled
$2.99	• **Wares**: 55,187 items pulled • **Cups/Glass**: 32,933 items pulled • **Pictures/Frames**: 23,960 items pulled
$3.99	• **Wares**: 44,605 items pulled • **Toys**: 25,398 items pulled • **Vases/Figurines**: 24,787 items pulled

Table 23. Top pulling SKUs at top 3 price points

In general, the Wares SKU consistently dominated sales across all the top 3 price points. This could be because Wares made up the largest portion of XYZ's production, and items falling under Wares SKU possess qualities that consistently attract customers, which could be their affordability or aesthetic appeal.

Another noteworthy observation is the consistent sales performance of the Wares, Plastic, and Cups/Glass at $1.99 and $2.99 price point, which also suggests a stable and recurring demand for items within these SKUs. It highlights the appeal to customers that seek affordable yet functional items that are usually bought in quantities higher than one.

Furthermore, we went into the top-selling SKUs across the three primary price tiers to gain deeper insights into consumer behaviors and preferences. Notably, our analysis revealed recognizable patterns. For instance, at the $1.99 mark detailed in Table 22 (sales data), Wares led in terms of items sold, followed by Cups/Glass. However, when examining Table 23 (inventory data), these rankings were reversed. This suggests that, at $1.99, Wares outperformed Cups/Glass in revenue generation. Additionally, the $2.99 price tier highlighted Pictures/Frames as a top-pulled SKU, suggesting that this price point may not have been optimal for items in this category.

Moreover, the presence of Wares, Cups/Glass, and Plastics in Table 22 and Table 23 also highlights a clear trend: smaller items dominate lower price points, selling in higher quantities. Meanwhile, larger items sold less frequently and at higher prices

(

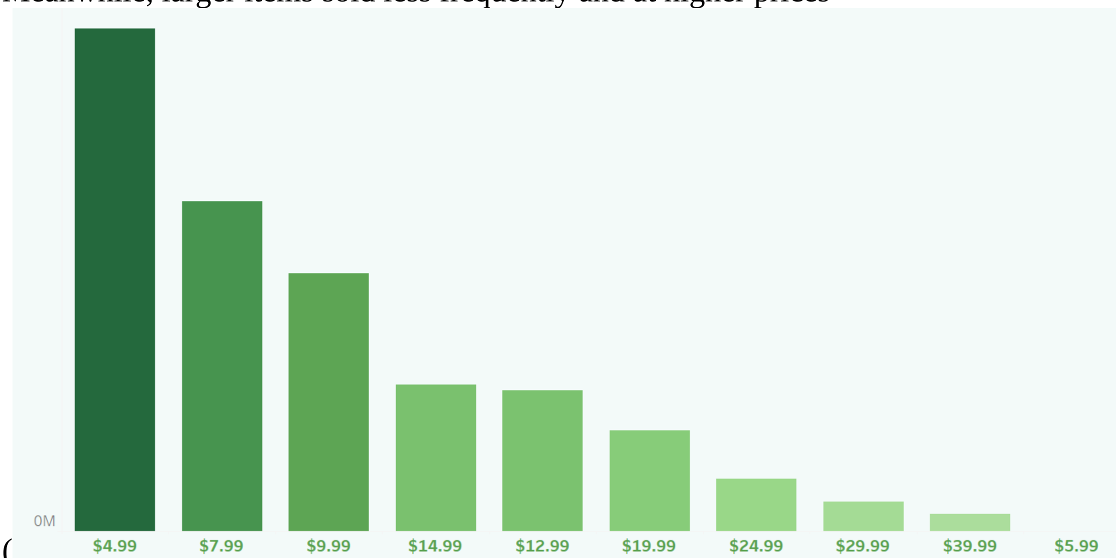


Figure *3. Top 10 selling price points in **Large Wares SKU***

Smaller objects, being more affordable, align with the demand for budget-friendly deals among XYZ’s clientele. This preference for lower-priced, easily replaceable items persists despite the availability of pricier, higher-quality alternatives within the same SKU category. Conversely, larger items such as Large Wares, which are harder to replace, more durable, and purchased less frequently, tend to have higher price points.

below for example).

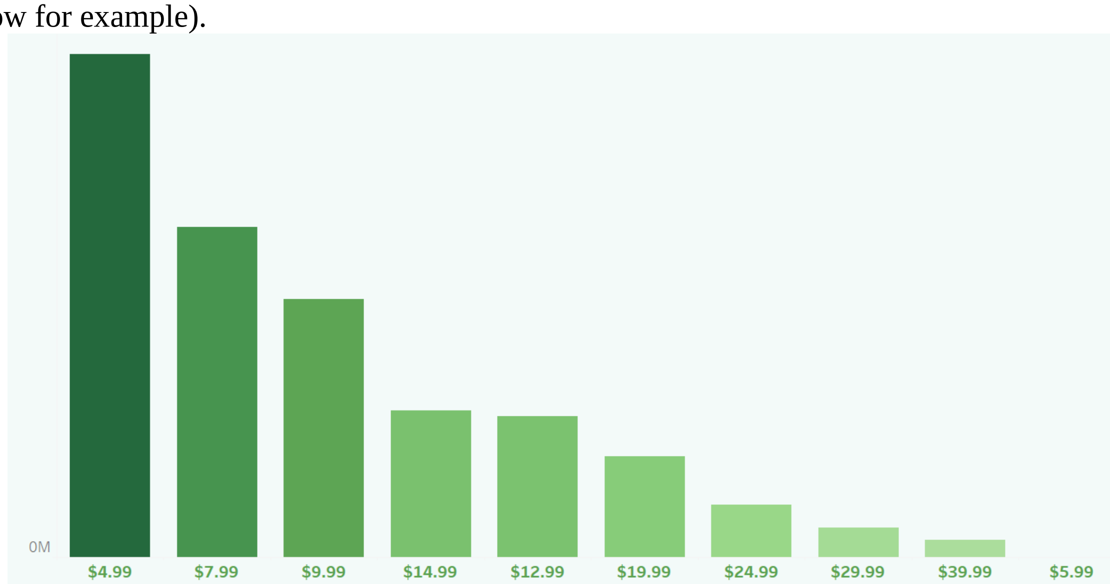


*Figure 3. Top 10 selling price points in **Large Wares SKU***

Smaller objects, being more affordable, align with the demand for budget-friendly deals among XYZ’s clientele. This preference for lower-priced, easily replaceable items persists despite the availability of pricier, higher-quality alternatives within the same SKU category. Conversely, larger

items such as Large Wares, which are harder to replace, more durable, and purchased less frequently, tend to have higher price points.

9.1.2. Discussion

XYZ should focus on maintaining a sufficient supply of items at the most popular price points, particularly $1.99, $2.99, and $3.99, which have shown to drive the highest sales volumes. In addition, XYZ can allocate more resources and shelf space to SKUs that consistently perform well at these price points, such as Wares, Cups/Glass, and Plastic.

Furthermore, XYZ can also develop targeted pricing strategies for specific SKUs based on their performance at different price points. Specifically, the Cups/Glass SKU shows strong performance at both $1.99 and $2.99, but there were more items pulled at $1.99. We had two assumptions: 1) the items did not have a good quality and 2) customers may view the $1.99 price point as too low for Cups/Glass and have a higher risk of damage or lower quality. As a result, they might gravitate towards $2.99 for perceived higher quality. However, $3.99 might appear to be too high for this SKU because Cups/Glass does not rank among the top-selling SKUs at that price point. Therefore, the first step that XYZ should take involves mandating employees to conduct a more careful investigation into the quality of the cups/glassware offered at this price point. If the quality of the items were ensured, XYZ then can consider testing slightly higher price points for $1.99 items to optimize revenue while still maintaining its popularity. For SKUs like Pictures/Frames that have a higher number of pulled items at $2.99 price point, XYZ can experiment with lower price points to encourage sales and reduce the need for pulling.

9.2. Research Question 2

After determining the above question, how does the rate of sales reflect on the pricing associate? Is the associate with the most sales aligned with the most sold price point?

9.2.1. Findings

To address this research, we built Figure 4 and Figure 5 graphs to find which employees have the highest sell-through rate. We calculated this measure in Tableau based on how many items out of the number an employee produced were sold out of their total production amount. The formula is as follows:

*Sell-through Rate = (Total Sold Items) / (Total Produced Items) * 100Figure* 4 showcases the top 10 performers by sell-through rate, while Figure 5 illustrates their production volumes. Moreover, we implemented a threshold, focusing only on employees who sold a minimum of 1,000 items. This was to ensure only employees with a robust sales history would be considered, and to facilitate a more reliable performance assessment. However, to accommodate different perspectives, we incorporated a filter in the ***XYZ Performance Tracker*** dashboard, which allows users to adjust the threshold and view the corresponding metrics.

By highlighting the employee with the highest sell-through rate, we aimed to compare their SKU-based pricing against the storewide average. From Figure 4, we found that over the two-year period spanning January 2022 to January 2024, Marie Flores had the highest sell-through rate.

Figure 4. Top 10 Employees with the Highest Sell-through Rate and their corresponding pricing effectiveness

Figure 5 below provides insights on the number of items produced by the top 10 employees with the highest sell-through rates. It can be seen from the graph that all of them had a robust production history. Overall, Marie Flores had a sell-through rate of 85.7% while also producing a substantial number of items.

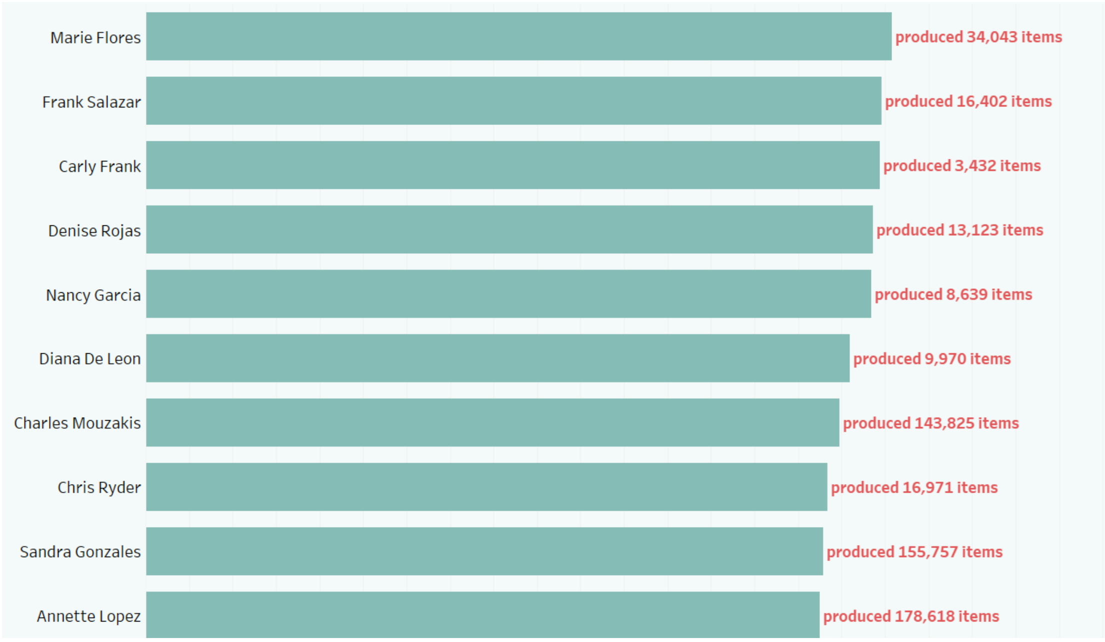


Figure 5. Top 10 Employees with the Highest Sell-through Rate and their corresponding production amount

Based on the findings, we were curious to explore the pricing patterns of Marie Flores. In *Figure 6*, we depicted the price distribution of Marie's sold items. Interestingly, Marie predominantly priced her items at $1.99, which did not exactly align with the trend observed from the storewide average of price points, where $2.99 stood out as the top 1 price point. However, Marie's top three price points ($1.99, $2.99, and $3.99) still aligned with the overall top three.

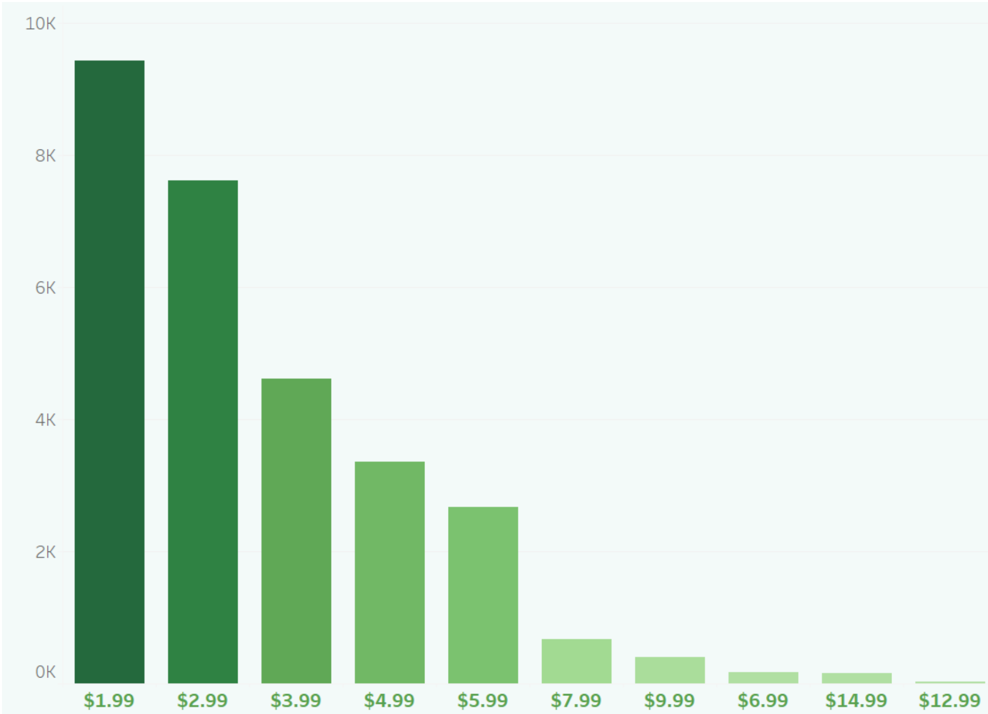


*Figure 6. Total Sold Items of **Marie Flores** by Price Point*

Table 24 below provided a closer examination of Marie's selling pattern. It illustrated Marie's total quantity sold for each SKU at the top 10 selling price points. Notably, Wares still emerged as the frontrunner in terms of items sold at this price point for Marie Flores, which aligned with the overall pricing pattern. This alignment is noteworthy since Figure 1 and Figure 6 both showed a trend where the highest number of sold items originates from the lowest price points, gradually decreasing in quantity with each dollar increase in price point.

	Price Point									
	$1.09	$1.99	$2.99	$3.99	$4.99	$5.99	$6.99	$7.99	$9.99	$14.99
Wares	16	9,154	6,598	3,601	2,995	2,320		556	298	153
Cups/Glass		98	261	133	34			3		
Toys		31		213		150		31	18	
Plastic		54	141	110	62	41				
Vases/Figurines		38		231			87		27	
Office		41		114		52		9	7	
Seasonal			138		53		21		7	1
Pictures/Frames			105		73				26	
Pots/Pans			80		79			43		
Bed/Bath			45	107		37		3		
Metal			53	52		45				2
Wood			64		40		36			
Dishes		11	52	30			23		7	
Sports			10	25		30		6		
Wicker			31		11			3		
Games		1	30		5		1		3	
Large Wares					10			11	2	5
Easter							5			

Table 24. Marie Flores's Sold Quantity by SKU and Price Point

9.2.2. Discussion

Marie's sell-through rate, along with her pricing pattern, once again reinforces the popularity of Wares items within lower price points: $1.99, $2.99, and $3.99. Her pattern aligns well with the overall trend addressed in section 9.1, which indicated that customers have a strong preference for affordable Wares items. Given this consistency, our recommendation for XYZ is to strategically allocate resources and focus on Wares within these price points to maximize sales and meet customer demand.

9.3. Research Question 3

How do pricing associates' items do when they reach the sales floor?

9.3.1. Findings

To address research question 3, we used the pricing effectiveness metrics. As previously explained in section 5.6, pricing effectiveness represents the ratio of the actual selling price of an item to the price point that the employee set.

From Figure 5, we were curious to see if the employees with the highest sell-through rates also priced their items effectively. Interestingly, while Marie Flores has the highest sell-through rate, only 86.9% of her items were sold at the full price point, which indicated that some of her items were sold at a discounted price. This suggests that although Marie Flores had a large quantity of items sold, there may be room for improvement in her pricing strategies to maximize revenue. On the other hand, employees such as Charles Mouzakis, Chris Ryder, Sandra Gonzales and Annette Lopez, despite having lower sell-through rates compared to Marie Flores, demonstrated higher pricing effectiveness, which meant that a large percentage of their sold items were sold at the full price point without discounts.

Figure 7. Pricing Effectiveness of Top 10 Employees with the Highest Sell-through Rate

We were also curious to see which employees excelled at pricing strategies. Figure 8 below shows the top 10 employees who demonstrated exceptional pricing effectiveness. To ensure a meaningful comparison, we also established a threshold of including only employees who have sold more than 1,000 items. This decision was made to address the potential dilemma where employees who have produced only a few items, all of which were sold at the set price, would exhibit 100% pricing effectiveness. Such a scenario could be misleading when comparing their performance to employees who have been with XYZ for a longer period and have a larger volume of sold items. While we settled on the 1,000-item threshold, we acknowledged that opinions may vary, and some might argue for a higher or lower threshold. As mentioned in section 9.2.1, to accommodate different perspectives, we incorporated a filter in the ***XYZ Performance Tracker*** dashboard that allows users to adjust the threshold and view the corresponding metrics.

The range of pricing effectiveness for the top employees in Figure 8 spans from 91.3% to 92.2%. Top employees with impressive pricing effectiveness include Thomas Wilson, Shannon Reynolds, Maria De Dena, and Maribel Romero. Their metrics indicated that the revenue from their items reached almost 92% of their expected revenues. Such a high percentage demonstrated the employees' skill in accurately pricing items based on their perceived value and market demand. By selling items at their full price, these employees significantly contributed to the overall revenue and sustainability of XYZ's operations.

Figure 8. Top 10 Employees by Pricing Effectiveness

9.3.2. Discussion

High sell-through rates do not always indicate high revenues generated. Based on our findings, we recommended that XYZ could develop targeted training and coaching programs to help employees improve their pricing strategies and effectiveness. For example, XYZ can provide guidance and support to employees like Marie Flores, who have high sell-through rates but lower pricing effectiveness, to help them optimize their pricing decisions. XYZ's training could focus on understanding market trends, competitor pricing, and effective pricing techniques to maximize revenue while maintaining high sell-through rates. In addition, XYZ should encourage a culture of knowledge sharing and collaboration among employees. For example, XYZ could hold pricing workshops where top employees in pricing effectiveness can share their experiences, techniques, and decision-making processes with their colleagues. This can help elevate the overall pricing effectiveness across the firm.

Furthermore, we believe that the ***XYZ Performance Tracker*** dashboard that we created would be a valuable tool for monitoring and analyzing employee performance metrics, including pricing effectiveness. XYZ should encourage managers and employees to actively utilize the dashboard to track their performance, identify areas for improvement, and set goals.

9.4. Research Question 4

How well and effectively do employees make pricing decisions and adhere to XYZ's pricing guidelines?

9.4.1. Findings

We approached this research question by considering whether the variability of prices within a SKU could be an indicator of an employee's adherence to XYZ's pricing guidelines and the thoughtfulness of their pricing decisions. One interesting finding from our visit to XYZ's production space was that many employees seemed to lack careful consideration when pricing products. We observed instances where employees would grab an item and automatically assign a price without thorough examination. For example, when one of our team members engaged in a conversation with an employee, she noticed that the employee consistently assigned a price of $2.99 to all items within a particular SKU. This observation led us to hypothesize that the variability of prices within an SKU could indicate the level of thoughtfulness and accuracy in an employee's pricing decisions. We hypothesized that if the price variability within a SKU was large, it would suggest that the employee took the time to carefully consider and assign accurate prices to individual items. Conversely, if employees exhibited low price variance within an SKU, it might indicate that they were defaulting to the same price point for items without thoroughly evaluating each piece.

To test this hypothesis, we calculated the correlation coefficients between pricing effectiveness and pricing variance within each SKU. The definition of pricing effectiveness and pricing variance can be found in section 6. Correlation coefficients measure the strength of the relationship between two variables. The strength of a relationship can be anywhere between −1 and +1. The stronger the correlation, the closer the correlation coefficient comes to ±1. For those unfamiliar with correlation coefficients, we provided an interpretation guide in **Error! Reference source not found.**.

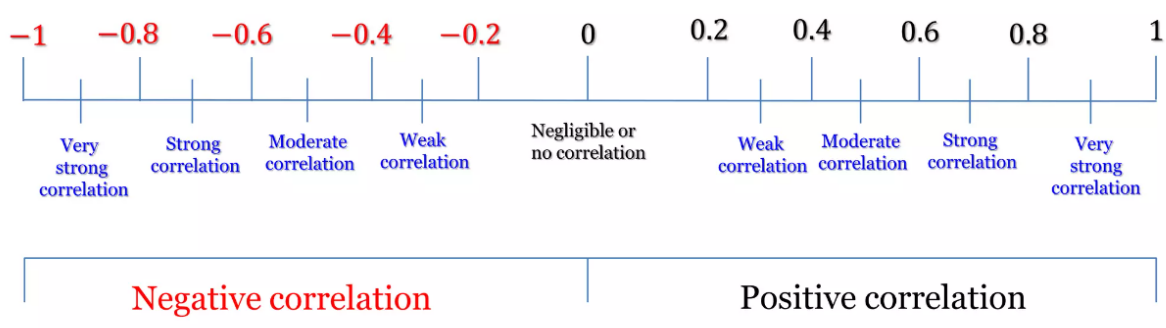


Figure 9. Interpretation of Correlation Coefficients

Surprisingly, all the correlation coefficients in Table 25 below only ranged from -0.03 to 0.01. That means there was almost no correlation between pricing effectiveness and pricing variance.

SKU	Correlation between Pricing Effectiveness and Pricing Variance
Large Wares	-0.03304
Wares	-0.02232
Dishes	-0.016184
Cups/Glass	-0.01025
Bed/Bath	-0.00584
Office	-0.00441
Games	-0.00364
Vases/Figurines	-0.00117
Sports	-0.00108
Plastic	0.00019
Pictures/Frames	0.00092
Metal	0.00122

Wood	0.00212
Pots/Pans	0.00495
Toys	0.00503
Seasonal	0.01004
Wicker	0.01048

Table 25. Correlation between Pricing Effectiveness and Pricing Variance within each SKU

Furthermore, to visualize the relationship between pricing effectiveness and pricing variance, we created scatterplots (Figure 10) for the four most popular SKUs identified in 7.2.3.2: Wares, Plastic, Cups/Glass, and Toys. As illustrated, none of the scatterplots revealed a significant relationship between pricing effectiveness and pricing variance within these SKUs either. This finding contradicted our initial hypothesis and suggested that the variability of prices within an SKU may not be a reliable indicator of an employee's pricing accuracy or effectiveness.

Figure 10. Relationship between Pricing Effectiveness and Pricing Variance within four SKUs: Wares, Cups/Glass, Plastic and Toys

9.4.2. Discussion

We assumed that several factors could contribute to this lack of correlation. The pricing of individual items within an SKU may depend on specific characteristics such as quality, condition, or unique features. Employees might assign different prices based on these factors, leading to price variability that is not necessarily indicative of their overall pricing effectiveness. Some SKUs might contain a wide range of items with varying attributes, while some SKU's price range might not vary that much. In such cases, price variability might be a result of the inherent complexity of the SKU rather than the employee's pricing decisions.

While our initial hypothesis was not supported by the data, this finding highlighted the complexity of pricing decisions. Therefore, there is a need for XYZ to maintain ongoing pricing training, and pricing monitoring to ensure effective and consistent pricing practices.

9.5. Research Question 5

How do associates' pricing decisions differentiate between stores?

9.5.1. Findings

The fifth research question deals with analyzing the pricing effectiveness of associates by store. For this question, we took the average pricing effectiveness score of all employees at each store. The average pricing effectiveness score for all stores is in the range of 86.8%-90.8%. The stores with the highest pricing effectiveness scores are DeZavala and New Braunfels. Figure 11 below outlines the top 10 stores by average pricing effectiveness.

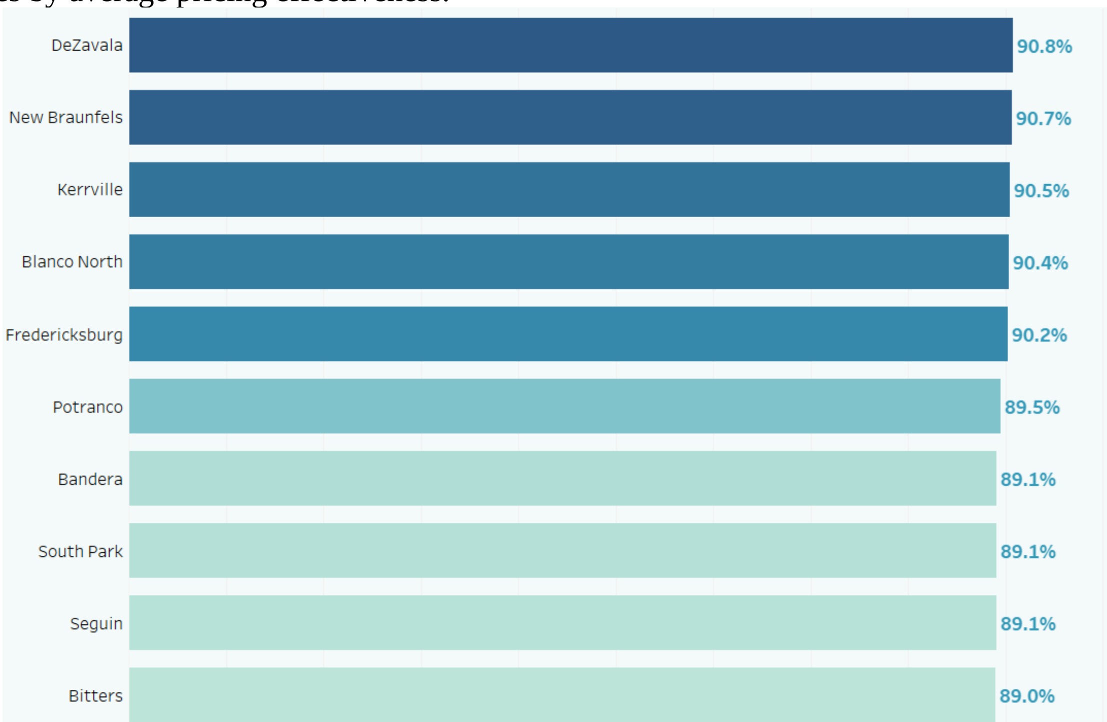


Figure 11. Top 10 Stores by Pricing Effectiveness

Figure 12 below outlines the bottom 10 stores by average pricing effectiveness. From this table, we observe that the least effective stores in terms of pricing are WW White and Gateway. These stores could benefit from employee pricing training, which is discussed in the next section.

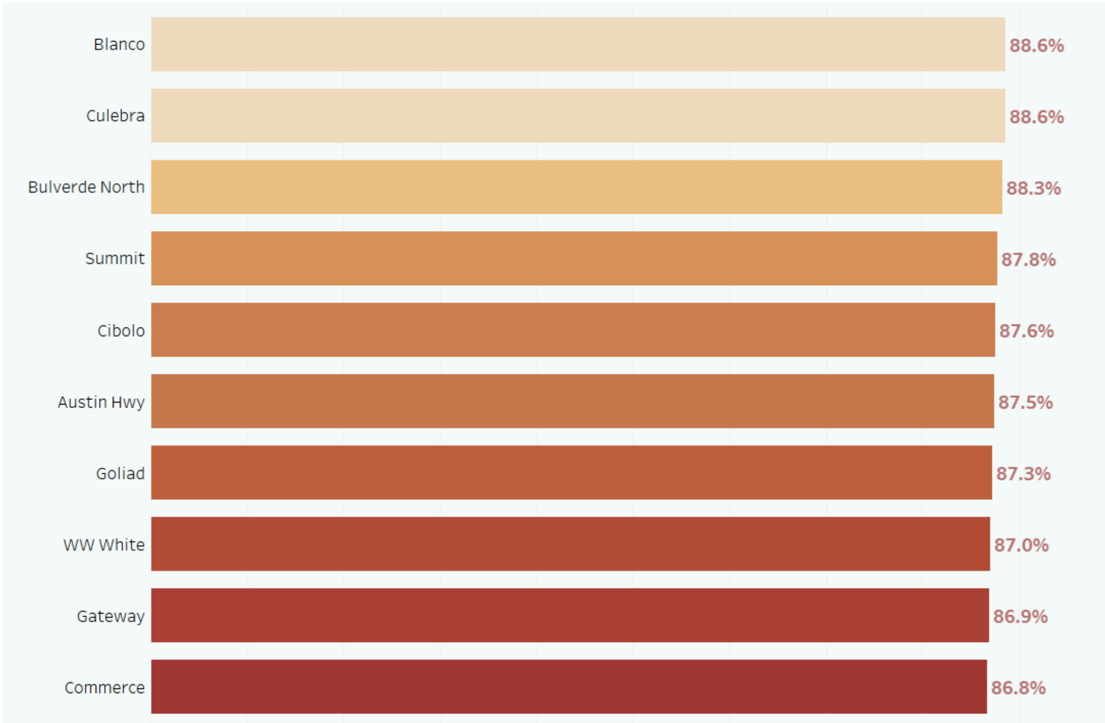


Figure 12. Bottom 10 Stores by Pricing Effectiveness

9.5.2. Discussion

The main suggestion we devised to improve XYZ's business practices is to focus on stores with lower pricing effectiveness according to the ***XYZ Performance Tracker*** dashboard. The lowest five stores in terms of pricing effectiveness are WW White, Gateway, Summit, Goliad, and Cibolo. These stores and other lower ranked stores could benefit from employee training regarding pricing effectiveness. Management could determine the top price points for each SKU and create a table or chart illustrating this that is easily accessible to all employees. Employees could benefit from this because they would have data to reference on which price points are most likely to be sold by SKU. This could be especially helpful in scenarios when employees are debating between two price points. This decision would become easier because they could select the most effective price point based on the data.

9.6. Research Question 6

With the last two years of data, how did the change in XYZ's process, which was finalized in 2023, impact the production numbers of each location?

9.6.1. Findings

To analyze the sixth research question, we built another dashboard comparing production, sell through, and pull through rates before and after the production process changes by store. While each store implemented the new production process on a slightly different date, the most common time frame for stores was May 2023. The dashboard shows rates both before and after May 2023 for a simple comparison.

Overall, there were not many significant changes in production rates after the new production process implementation. Production rates slightly increased in both items and monetary value produced. However, the most telling insight that this production change in stores was an effective decision is shown in the sell through graph. The visualization for combined store data shows that the number of items sold has decreased since the change, while the revenue from sold items has increased. This finding led us to infer that the new process has improved the quality of items on the shelves and decreased some of the clutter of low-quality items. The quantity of pulls has also decreased significantly since May 2023, which once again indicates that the shelves are being stocked with higher quality items that customers are interested in purchasing. Figure 13 below displays the decrease in total quantity sold and increase in revenue since May 2023 for all stores.

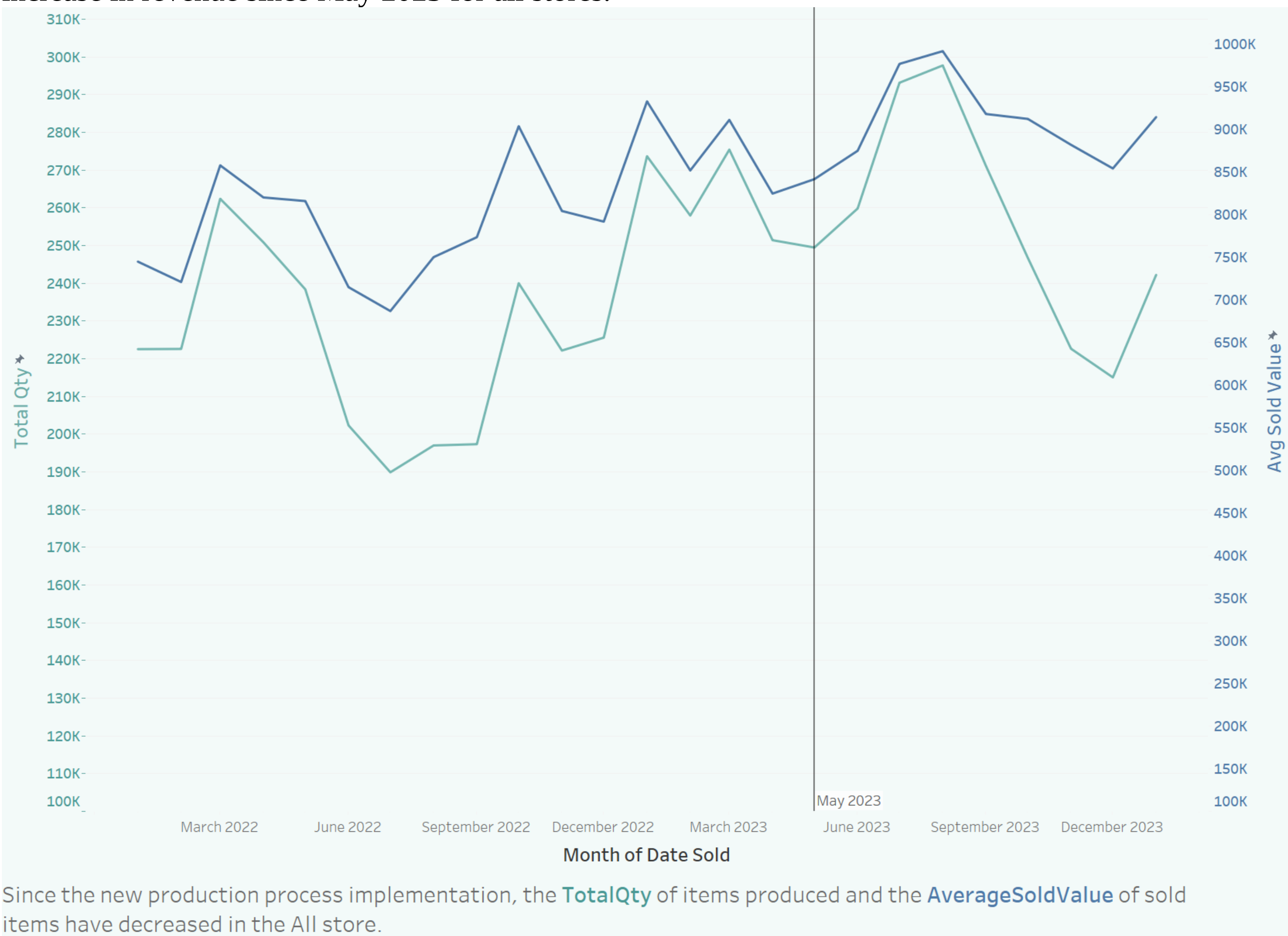
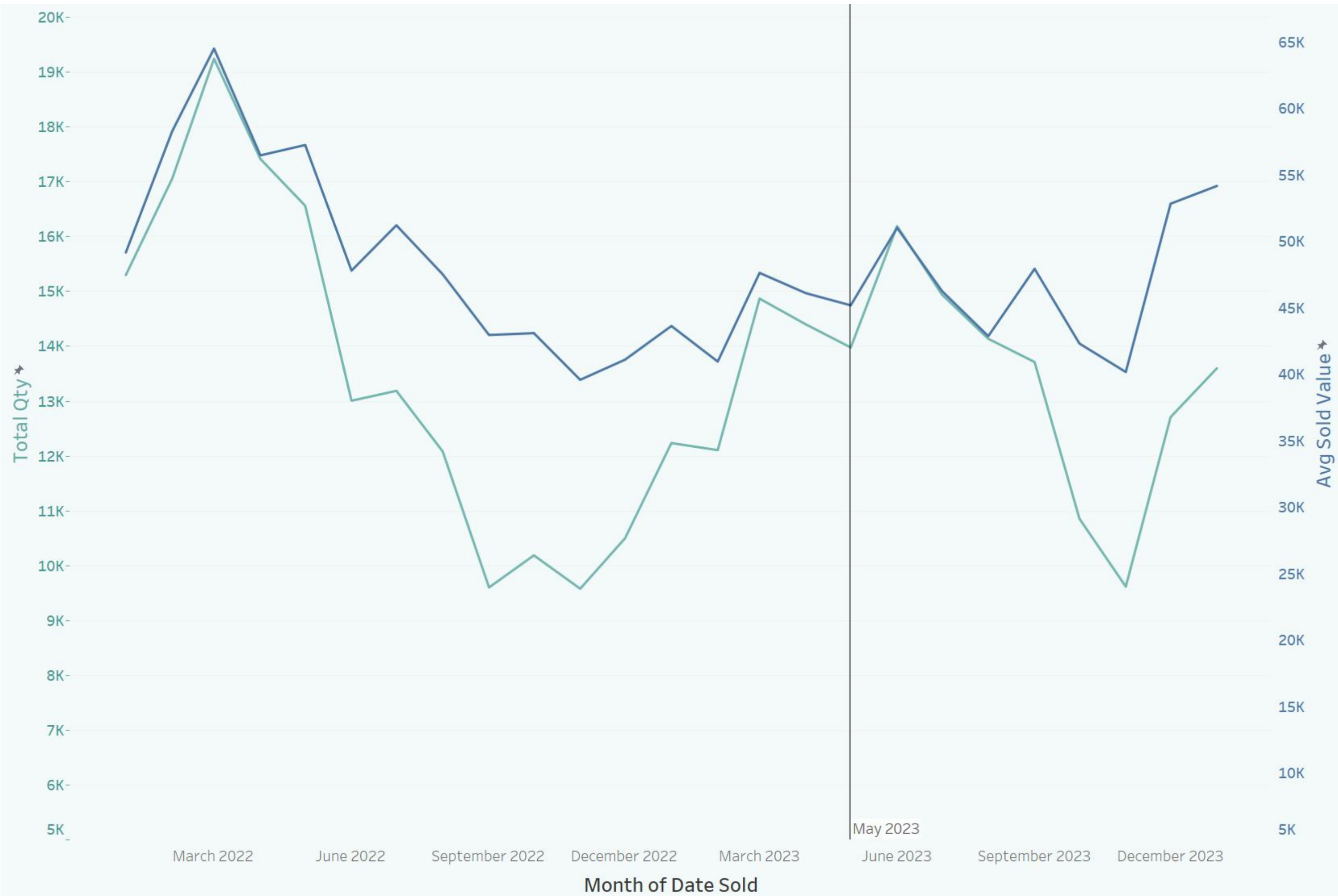


Figure 13. Sell-through Rates for All Stores Before and After Production Change

In analyzing the different stores, many of the findings went in different directions. While many stores were affected by this production change in a positive way, some have not performed as successfully since the production change. One example of a store performing very well since the production change is Culebra. Items sold in quantity have decreased and average revenue has increased approximately $10,000 since May 2023. The quantity of pulls has also decreased significantly. Production rate in both quantity and monetary value have increased slightly. Other stores that follow this trend in better quality production after the change include Blanco, Commerce, and Summit. Figure 14 below shows the Culebra location’s improvement in revenue since the production change.



Since the new production process implementation, the **TotalQty** of items produced and the **AverageSoldValue** of sold items have decreased in the Culebra store.

Figure 14. Sell-through Rates for Culebra store

While many stores like Culebra have performed well after the process change, others saw significant declines in sales and production. Laredo is the store with the most significant decline in sales. Average revenue for Laredo dropped from about $31,000 to only $99 since 2023. However, production quantity in items did decrease and production value in dollars increased. Below, Figure 15 shows the drastic decline in Laredo sales after May 2023.

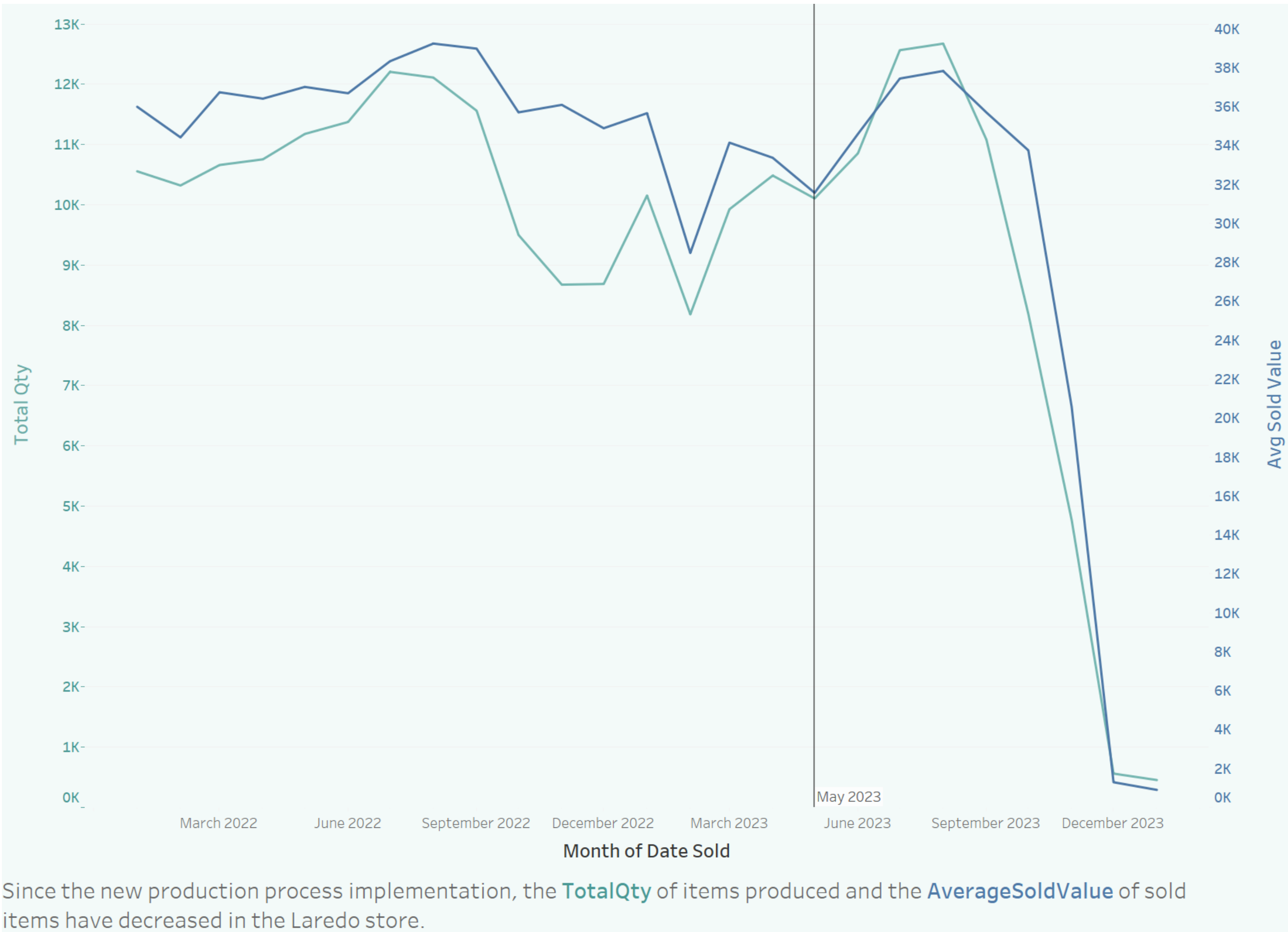


Figure 15. Sell-through Rates for Laredo store

Other stores that follow a similar trend with decreasing sales and increasing production are Fredericksburg, Goliad, Marbach, New Braunfels, Potranco, and Seguin. Evans experiences a significant decline in sales, but a decline in production as well. While these stores are decreasing, most other stores are increasing slightly in sales and production rates, indicating that this production change has created overall positive results.

9.6.2. Discussion

Our suggestions for improvement mainly include investigating further how to make production more efficient in these stores with declining sales. The stores with the most significant decline that require the most future attention are Laredo and New Braunfels. We also suggest being pickier about the quality of items put on the sales floor because it seems that succeeding stores are making more revenue by putting fewer items on the sales floor. This could include further vetting items in the pre-sort stage right after donations are received. Because lower quality items can waste space on the shelves for weeks, it is important to produce higher quality items, which makes customers more inclined to purchase them.

10. Conclusion

The analysis conducted throughout this project led to many key findings that can help optimize XYZ's business procedures and increase revenue for the company. One of the major findings to highlight is the success of lower price points in all SKUs, but specifically wares. As we have noted many times, $1.99, $2.99, and $3.99 are consistently the most sold price points for most SKUs.

Continued emphasis on these price points will help increase XYZ’s profit. Another major insight derived from the ***XYZ Performance Tracker*** dashboard is the pricing effectiveness metrics for employees and stores. We suggest that management continually monitors these metrics and provide training and mentorship programs for employees or overall stores with lower pricing effectiveness. The last key takeaway is that the quality standard for items should be improved to decrease the clutter of lower quality items on the shelves that customers are not inclined to purchase. Implementing these suggestions and others listed previously in this report, as well as monitoring the ***XYZ Performance Tracker*** dashboard will improve XYZ’s sales trends and allow for better understanding from internal management of the company’s areas for improvement.