2025

# ECSA PROJECT

## QUALITY ASSURANCE 344

MIA DU PLESSIS, [26869241]

# Table of Contents

# Introduction

This report takes an in-depth look at several key performance areas in the company, focusing on delivery stability, product profitability, and staffing efficiency. It draws on multiple datasets, including customer records, product catalogs, sales transactions, and staffing logs to uncover patterns, identify problem areas, and highlight opportunities for improvement.

A variety of analytical methods were used to get a full picture of the performance. Statistical Process Control (SPC) and process capability analysis (Cp and Cpk) assessed the reliability and consistency of delivery processes, while ANOVA tests helped pinpoint the factors that most affect performance. Data-cleaning steps were also applied to fix pricing errors in the product catalog and restore the accuracy of financial information. Finally, financial and staffing models were developed to identify the most cost-effective ways to allocate resources, balancing profitability with service reliability.

Overall, the analysis reveals both strengths and weaknesses in the company's operations. While some processes are well-controlled and efficient, others show instability or systemic issues, particularly in delivery times and data quality. The findings provide practical, data-driven insights that can help management improve service, boost profitability, and make better-informed decisions.

## Part 1.2: Descriptive statistics

### Customer data

**1. Data Loading and Inspection:**

The customer dataset was successfully loaded and contains 5 000 records. Each record includes a customer ID, gender, age, annual income, and city of residence.

All data is complete and correctly formatted, with no missing values. A quick review shows a wide range of ages, incomes, and locations, which aligns with expectations for a diverse customer base.

Two customers are listed as 101 years old. While not impossible, these ages are unusual and will be monitored to ensure they don't skew the results. They could be data entry errors or placeholder values, but for now, they remain in the dataset.

**2. Summary Statistics:**

The analysis of the 5 000 customers shows that the company serves a diverse group in terms of both age and income. Customer ages range from 16 to 105 years, with an average of around 51, representing a mix of younger, middle-aged, and older individuals. Annual incomes range from $5 000 to $140 000, with a median of $85 000, slightly higher than the average. This indicates that a significant portion of customers fall within the middle-to-upper income range.

The customer base is fairly balanced by gender, with 48.6% female, 47.0% male, and 4.4% identifying as another gender. Customers are spread across seven major cities, with San Francisco representing the largest market at 780 customers, followed by Los Angeles and New York. This demonstrates that the business is not reliant on just one or two locations.

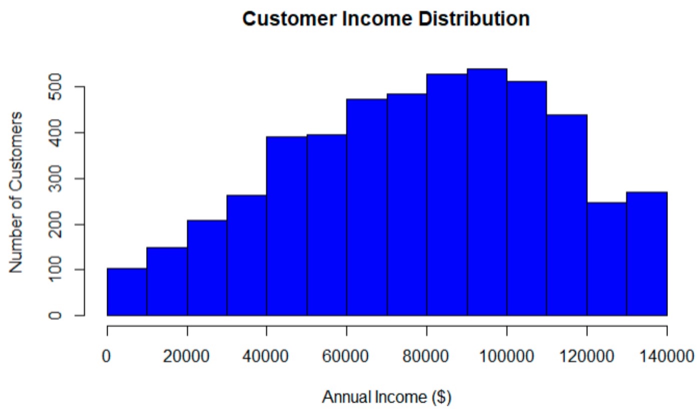
**3. Handling Missing Values**

A check of the customer dataset confirmed that the data is complete and clean. All 5,000 records include information for Customer ID, Gender, Age, Income, and City, with no missing values. Common issues such as blank entries or invalid values were also checked for and none were found. This indicates that the dataset is reliable and ready for analysis.

## 4. Data Filtering and Subsetting

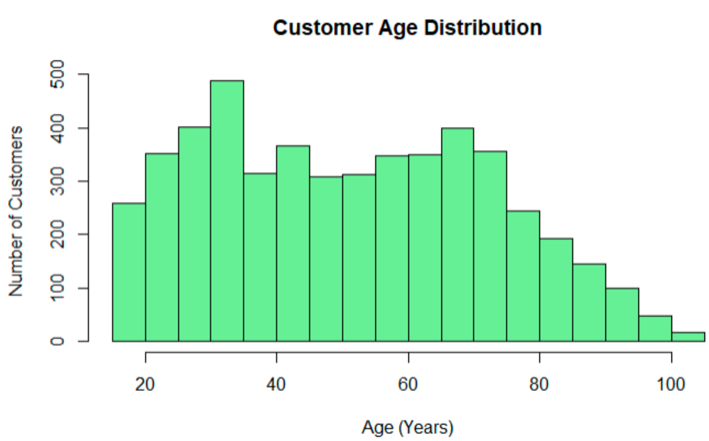
The data filtering step was skipped for this dataset because the initial review found it to be complete. The few customers with unusually high ages have been noted, but there is no reason to remove them.

## 5. Data Visualization:



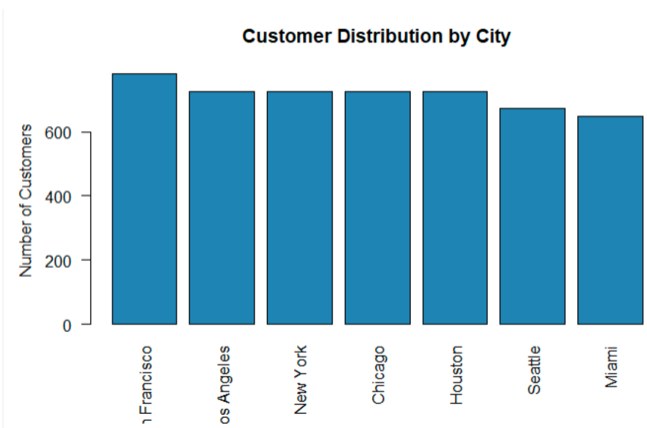

The income distribution is right-skewed, peaking in the $80 000 to $110 000 range. While a minority of customers report incomes as high as $140 000, the chart also shows a smaller, distinct cluster at the lower end of the scale.

The distribution highlights two key segments: a primary, affluent customer base with the spending power for premium products, and a secondary, more price-sensitive group. The presence of this second segment indicates that the company's products appeal beyond the initial target market. Marketing and pricing strategies can be refined to address these two groups effectively.

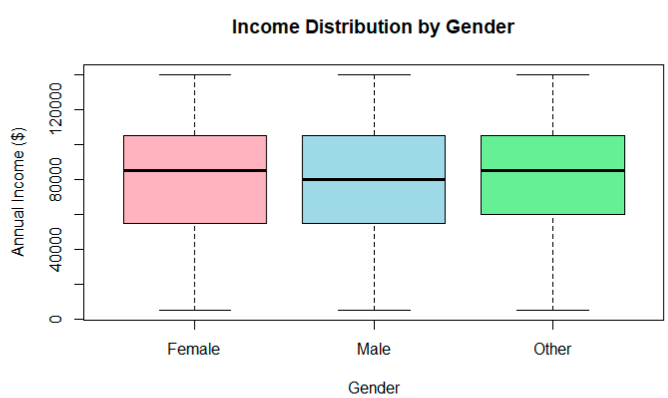


In contrast to the skewed income distribution, customer ages are spread evenly across the range. This broad appeal is a significant strength, providing long-term stability by avoiding reliance on a single age group.

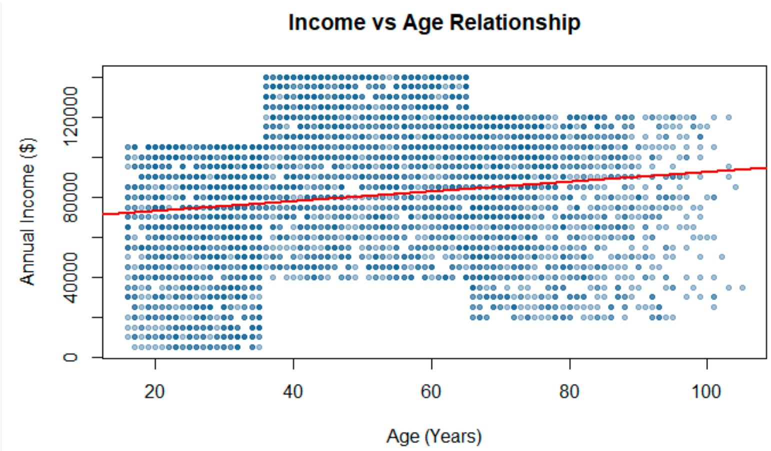
However, this diversity also requires more targeted outreach. A single marketing message is unlikely to resonate with all customers. Communication and product development strategies should therefore be tailored to address the specific needs and motivations of different age segments.




The customer base is distributed across seven major U.S. cities. San Francisco serves as the primary market, supported by Los Angeles, New York, Chicago, and Houston, which all have similar customer counts. Seattle and Miami, while smaller, represent important established markets. This distribution reduces operational risk by avoiding over-reliance on a single region.



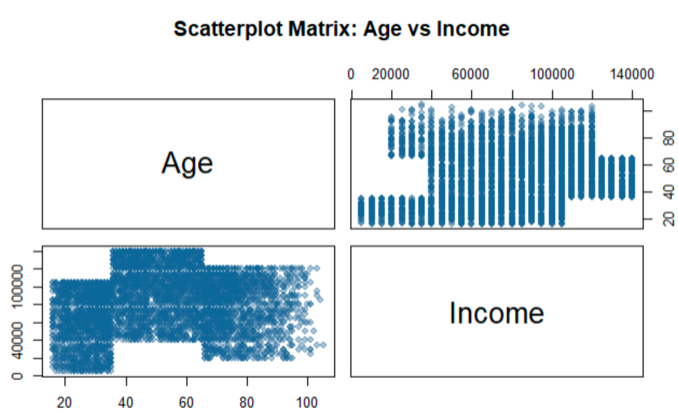

Analysis of income by gender reveals distinct distribution patterns, even though median incomes are similar across groups. The "Other" gender category displays a tightly clustered income range, while the "Male" and "Female" categories show a wider dispersion and a greater number of high-income earners. This suggests that, although average spending power is consistent, strategies for identifying high-value customers should differ, with broader targeting applied to the "Male" and "Female" segments compared to the "Other" segment.




The scatter plot of age against income shows no clear correlation, with data points scattered randomly. This indicates that customer age is not a reliable predictor of income level.

Marketing strategies for premium products should therefore not target specific age groups under the assumption of greater impact. A more effective approach is to identify high-value customers using direct income data and other spending metrics.

## 6. Exploring Relationships:



The analysis confirms the absence of a correlation between customer age and income, showing no discernible pattern. This establishes that age is not a reliable indicator of spending power.

This finding is critical, as it rules out targeting premium products solely to older demographics. High-value customers should instead be identified through direct income data and purchasing behavior.

## Product data

**1. Data Loading and Inspection:**

The product dataset comprises 60 distinct items, each defined by a unique ID, category, description, selling price, and markup percentage. The dataset is complete and well-structured, with products categorized as Software, Cloud Subscription, Laptop, Monitor, Keyboard, or Mouse. Descriptions follow a consistent format based on color and material.

A review of the dataset confirms consistency throughout. The inclusion of both selling price and markup percentage for each item provides a good foundation for analyzing individual product profitability.

**2. Summary Statistics:**

The product dataset is evenly distributed, with ten items in each of the six categories: Cloud Subscription, Keyboard, Laptop, Monitor, Mouse, and Software.

Pricing spans a wide range, from $350 to over $19 700. The mean price is $4 494, considerably higher than the median of $794. This indicates a large group of moderately priced items alongside a small selection of high-cost, premium products.

Despite these differences, the company applies a uniform pricing strategy. Markup percentages are consistent across all categories, averaging 20%, which suggests that selling prices are derived directly from cost using a standard multiplier rather than adjusting margins.

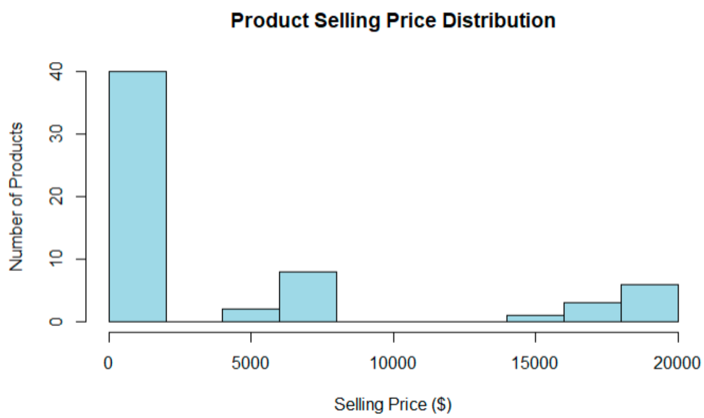
**3. Handling Missing Values:**

The dataset is complete, with all 60 products containing full information. Multiple checks confirmed that no data is missing. Since the price and markup information is fully intact, no data cleaning is required.

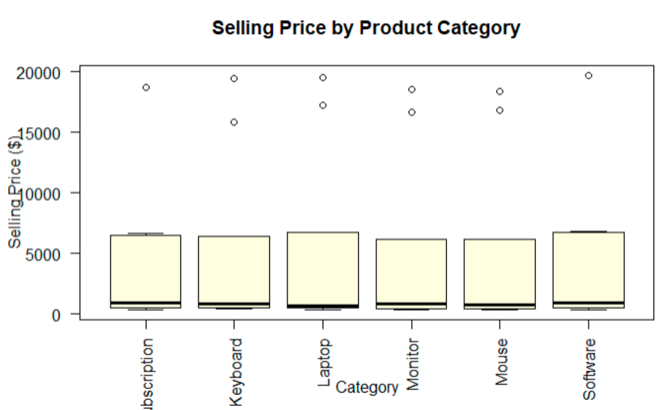
## 4. Data Filtering and Subsetting:

This step was skipped during the product analysis. This decision was made because the initial inspection showed the data was complete and well-structured, with no clear need to remove any products.

## 5. Data Visualization:

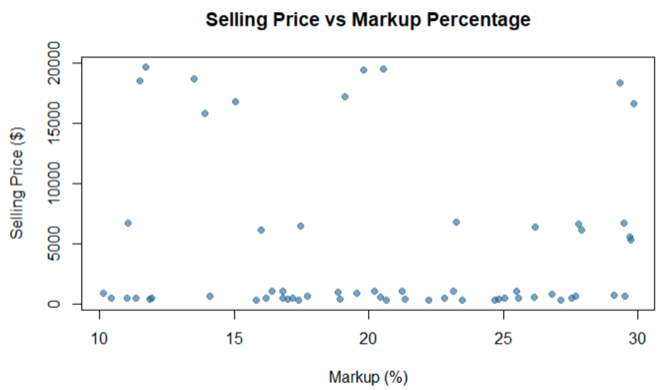



The price distribution histogram reveals a clear bimodal structure, reflecting the spread of products across the different price points. The primary mode consists of a large number of products priced below $5 000, representing a mass-market segment. A smaller cluster of premium products is positioned between $15 000 and $20 000. This distribution demonstrates positioning that targets both budget-conscious and high-end customers, while highlighting a distinct gap in the middle of the market.



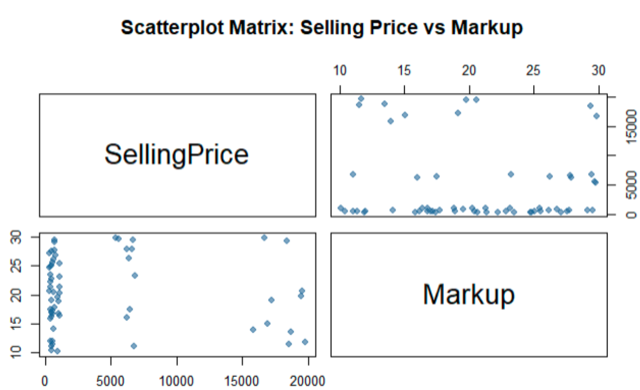
The product pricing structure reflects a clear two-tier strategy. The majority of products are priced under $5 000, while a smaller group of premium offerings is positioned between $15 000 and $20 000.

The gap between these tiers indicates a deliberate segmentation strategy targeting two distinct customer groups: the mass market and the high-end market. As a result, there is minimal representation of products in the middle-market price range.




The analysis of selling price against markup percentage reveals a key finding: there is no correlation between a product's price and its profit margin. The company applies a uniform markup strategy across the entire portfolio. As a result, profit per item is directly determined by its base cost, with premium products generating substantially higher profit per sale than lower-cost items, despite the identical percentage markup.

## <u>6. Exploring Relationships:</u>



The final analysis confirms that there is no relationship between a product's selling price and its profit margin percentage, with data points showing a random scatter. This supports the use of a standardized margin across the entire portfolio.

Total profit per unit is determined by the item's original cost. Applying a uniform margin percentage means that the sale of a high-cost item has a greater absolute profit than a low-cost item.

## Head office products

**1. Data Loading and Inspection:**

This analysis uses a larger dataset of 360 products and it represents the master product catalog. Each record follows the same format, including Product ID, Category, Description, Selling Price, and Markup, and the dataset is complete with no missing values. An initial review confirms consistency throughout the file.

**2. Summary Statistics:**

The head office product catalog is a comprehensive and well-balanced portfolio of 360 items, with 60 products in each category: Cloud Subscription, Keyboard, Laptop, Monitor, Mouse, and Software. Prices range widely, from $290 to over $22 400. The large difference between the mean price of $4 411 and the median of $797 highlights a two-tier strategy, with many affordable items alongside a smaller group of premium offerings.

Profit margins are consistent across the catalog, averaging 20%, reflecting a clear uniform pricing policy. The dataset is complete and reliable, providing a strong foundation for analyzing the company’s overall product strategy.

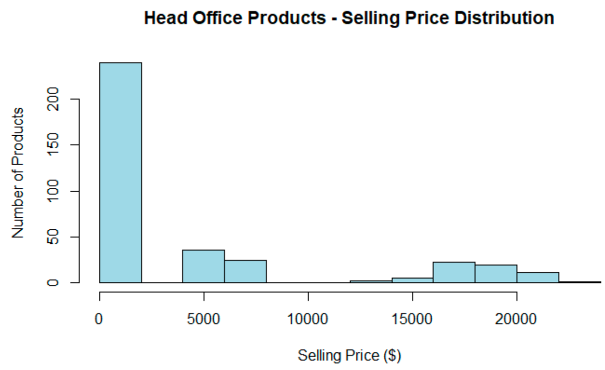
**3. Handling Missing Values:**

A check confirmed that the data is complete. All 360 product records are fully filled out, and there is no missing information for any product ID, category, description, price, or markup.

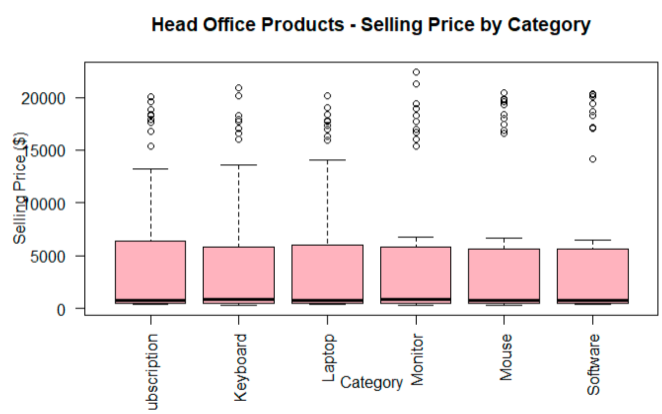
**4. Data Filtering and Subsetting:**

The data filtering step was also skipped for the head office product catalog. Since no data quality issues were found, all products were kept to provide an overall view.

## 5. Data Visualization:

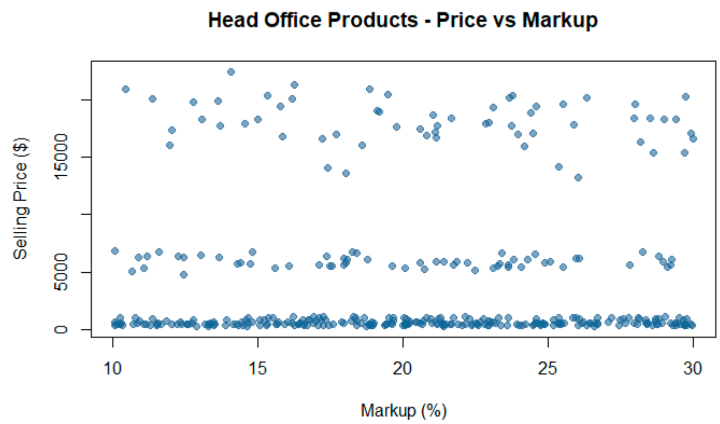



The price distribution shows a dense concentration of products under $5 000, representing the high-volume, accessible segment. A second cluster appears in the $15 000 to $20 000 premium range. The gap between $5 000 and $10 000 is largely unpopulated, indicating a deliberate strategy to avoid the mid-market and maintain a clear distinction between budget-friendly and premium offerings.



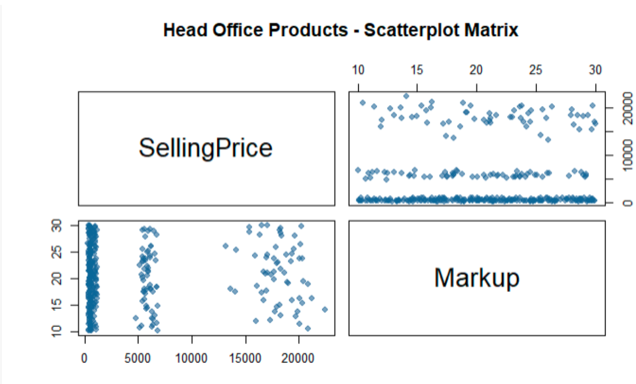

Examining prices by category reinforces the two-tier product strategy. Laptops represent the premium line, with the highest prices, while software products display the widest price range.

Cloud Subscriptions, Keyboards, Monitors, and Mice form the volume-driven categories, consistently priced lower. This structure aligns with customer expectations and supports targeted marketing efforts.




The relationship between price and markup confirms a consistent pricing rule. There is no connection between a product's selling price and its profit margin, which remains within a narrow range of 15% to 25%. This confirms the use of a cost-plus pricing model, where the selling price is calculated directly from the product's cost. This approach makes pricing decisions easier and ensures a uniform percentage of profits across all sales.

## 6. Exploring Relationships:




The final analysis confirms that no relationship exists between an item's price and its profit margin percentage. The random scatter of data supports the use of a uniform margin across all products, regardless of price tier.

## Sales from 2022 and 2023

**1. Data Loading and Inspection:**

This dataset contains 100 000 sales transactions from 2022 and 2023. Each transaction includes nine data points that track order for the entire order process, from placement to delivery. Key fields include CustomerID and ProductID, along with operational information such as order timestamps and fulfillment durations.

The dataset is complete and order sizes range from single units to large bulk purchases, reflecting service to both individual consumers and business clients. This dataset provides a good foundation for analyzing sales performance, customer behavior, and operational efficiency.

**2. Summary Statistics:**

Analysis of the 100 000 sales transactions from 2022 and 2023 reveals several key business patterns. Sales volume was higher in 2022, with 53 727 orders, compared to 46 273 in 2023.

The distribution of order sizes is bimodal. While the median order contains six items, the mean is 13.5, indicating a combination of small individual purchases and a substantial number of large bulk orders from business clients. Purchasing activity peaks around 1 PM, and sales volume remains relatively stable throughout the year, with a notable decrease in December.

The average order picking time is 14.7 hours, and delivery averages 17.5 hours. Delivery times are more consistent than picking times, although both processes exhibit high variability. The dataset is complete and reliable for this assessment.

**3. Handling Missing Values:**

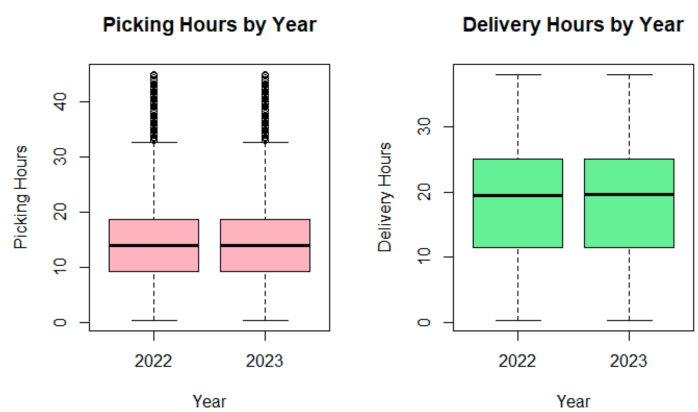
A data check confirmed that all transactions are complete. Each record contains the required information, including Customer ID, Product ID, quantity, timestamps, and operational hours. With no missing or invalid data, no cleaning or removal of records was necessary.

**4. Data Filtering and Subsetting:**

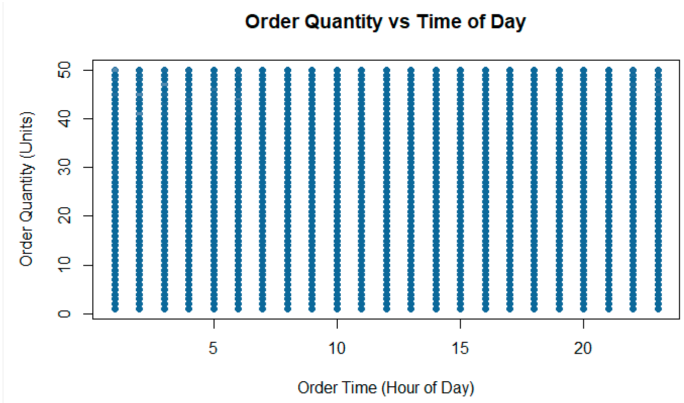
Since no data quality issues were found, all products were kept to provide a comprehensive view.

## 5. Data Visualization:


The order size chart provides insight into customer behavior. Most orders are small, containing between 1 and 10 items. However, there is also a significant number of very large orders, reaching up to 50 items. These bulk purchases raise the average order size and indicate that the company serves two distinct customer types: individual shoppers making small purchases and business clients buying in large volumes.

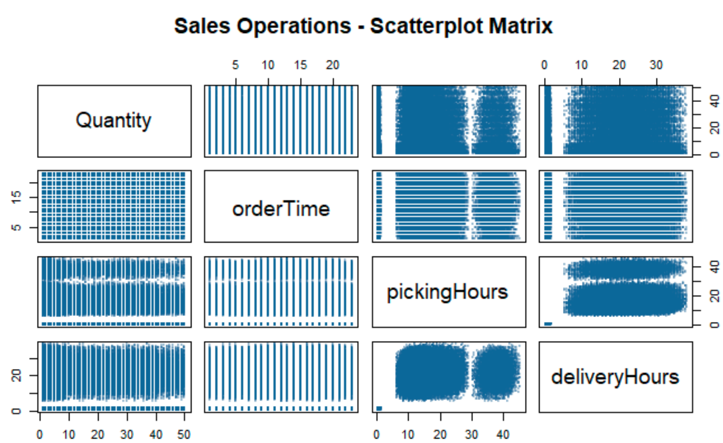



The comparison of 2022 and 2023 shows that efficiency has improved. Item picking times have decreased, and delivery performance has become faster and more consistent, with fewer delays. These improvements indicate that changes made to operations are having a positive effect, positively influencing both internal logistics and customer delivery performance.




The chart indicates that demand remains stable throughout the day, with order sizes generally consistent at all hours. A slight increase in larger orders occurs between 10 AM and 4 PM, likely reflecting business customers placing orders during standard working hours.

## **6. Exploring Relationships:**




The final analysis highlights several key connections between sales and operational performance. Order size is directly correlated with warehouse picking time, while having only a minor impact on the delivery phase. The time of day an order is placed does not significantly affect processing speed, indicating consistent performance throughout business hours.

A strong correlation between picking time and delivery time demonstrates that warehouse efficiency is a primary driver of overall order fulfillment speed.

# Part 3: SPA

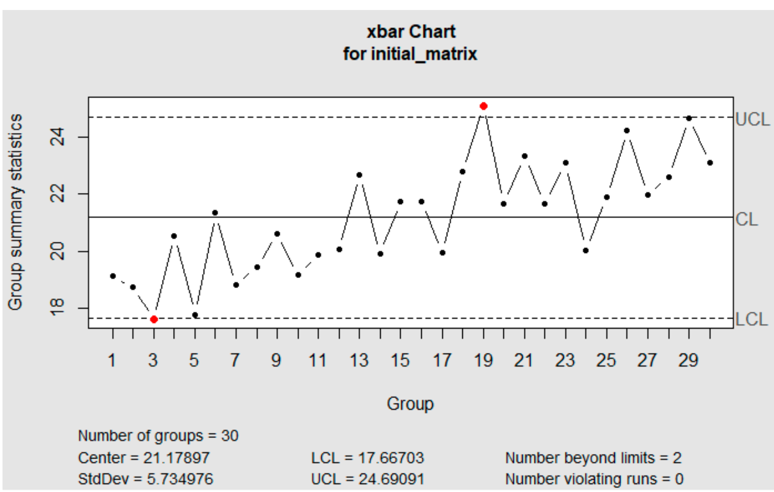
## 3.1 Initialization of Control Charts

A control limit analysis was conducted on delivery times for all 60 products. For each product, the first 720 deliveries were grouped into 30 batches of 24. Two types of control charts were prepared: an X-bar chart to monitor the average delivery time per batch, and an S-chart to evaluate consistency within each batch.

The S-charts indicated that delivery performance was stable and predictable across all products, suggesting strong control over process variation. However, the X-bar charts revealed a significant issue: average delivery times were unstable for many products, with several batches exceeding the three-sigma control limits.

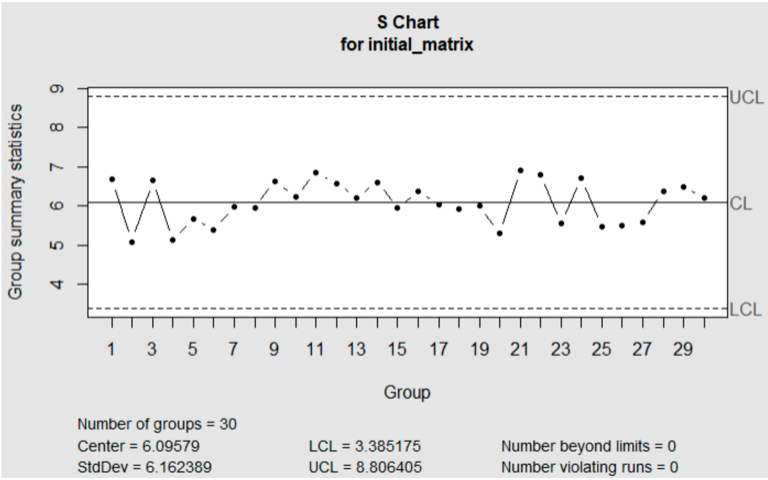
Overall, the findings show that while the company maintains consistent delivery performance, the average delivery speed itself is not under statistical control. Addressing this instability would improve process reliability and improve customer satisfaction.

**X-bar Chart for CLO011:**




The center line is 21.18 hours, and this is the average delivery time for the first 30 samples. The chart has a Lower Control Limit (LCL) of 17.67 hours and an Upper Control Limit (UCL) of 24.69 hours.

**S Chart for CLO011:**



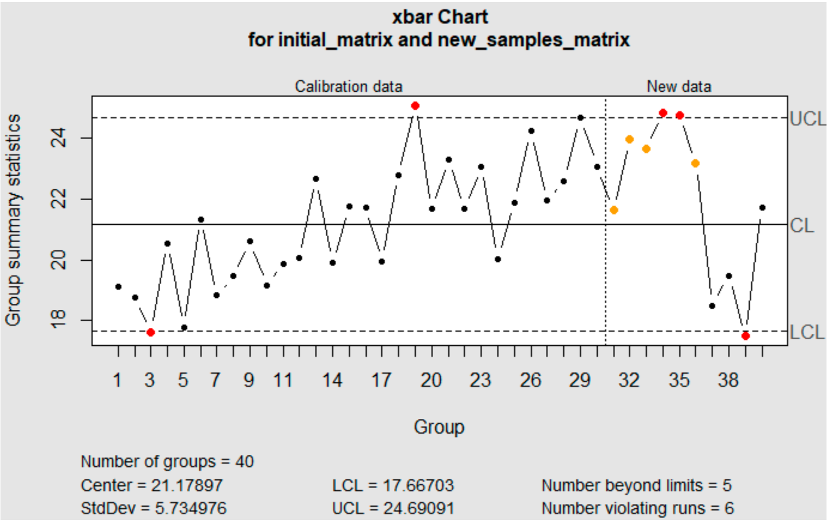

The center line for the S chart is 6.10 hours. This is the average variability between delivery times. There is also a LCL of 3.39 hours and an UCL of 8.81 hours.

The delivery process for CLO011 has an average delivery time of about 21 hours, but it varies with a standard deviation of 6 hours. The control limits indicate what range of variation is normal for this process.

## 3.2 Ongoing Process Monitoring

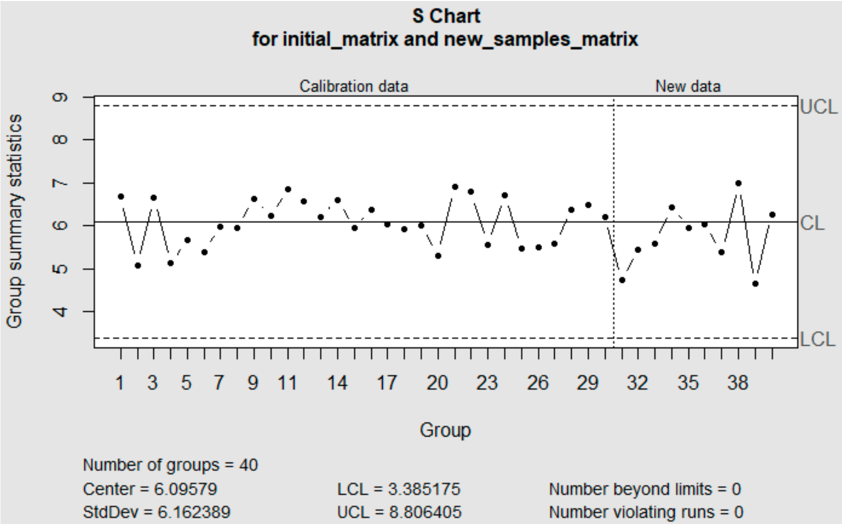
The monitoring process was continued by analyzing the next set of delivery samples, batches 31 to 40, to simulate ongoing real-time performance tracking. The updated charts for selected products confirmed the initial findings. The S-charts remained stable, with no points exceeding the control limits, confirming that process variability is not the root cause of the issue. In contrast, the X-bar charts continued to display instability in average delivery times. These results indicate that the fluctuations are persistent and show a systemic problem within the company's delivery process rather than a temporary problem.

## Analysis of X-bar Control Charts:



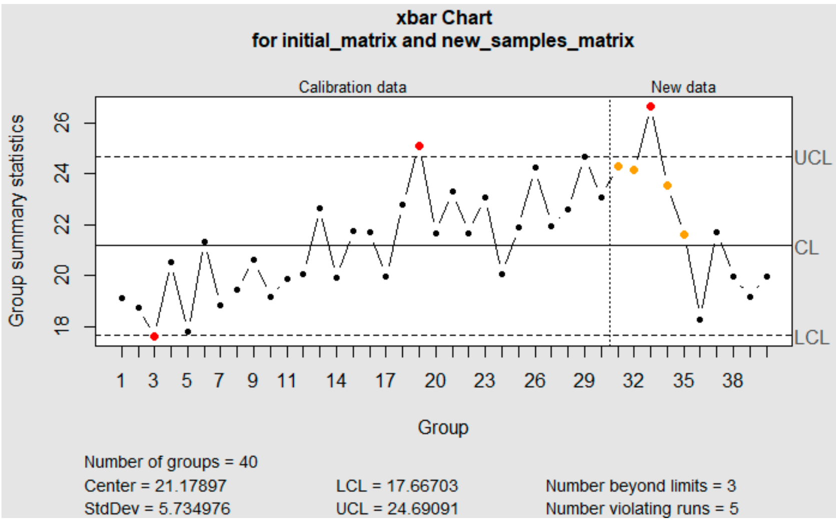

For CLO011, there are 5 points beyond control limits, but for other products, there are 3-4 points beyond limits. This means the process is out of control. The average delivery time is unstable and some samples have much higher or lower average delivery times than expected. Product managers need to investigate what's causing these extreme variations.

## Analysis of S Control Charts:



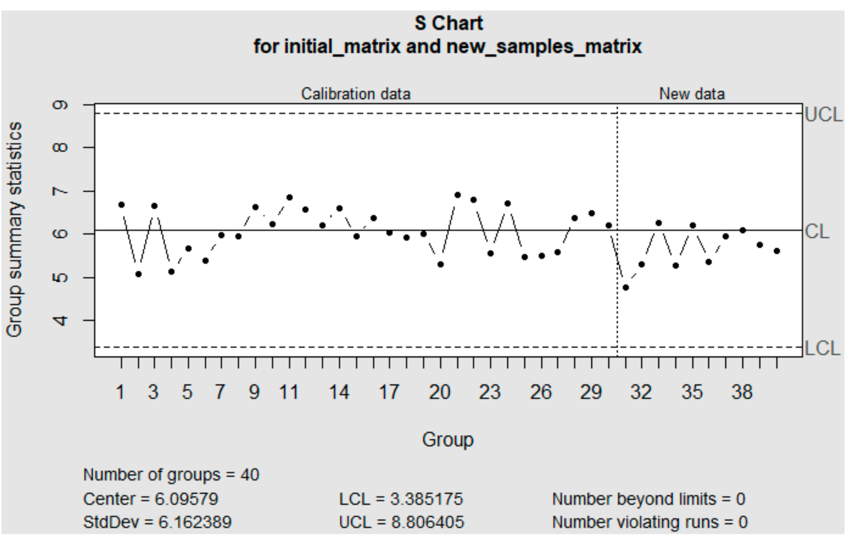
There are 0 points beyond control limits. This indicates that the process variability is in control. While the average delivery time jumps around, the variability between individual delivery times remains stable.

**Average delivery times for CLO012 (samples 1-40):**



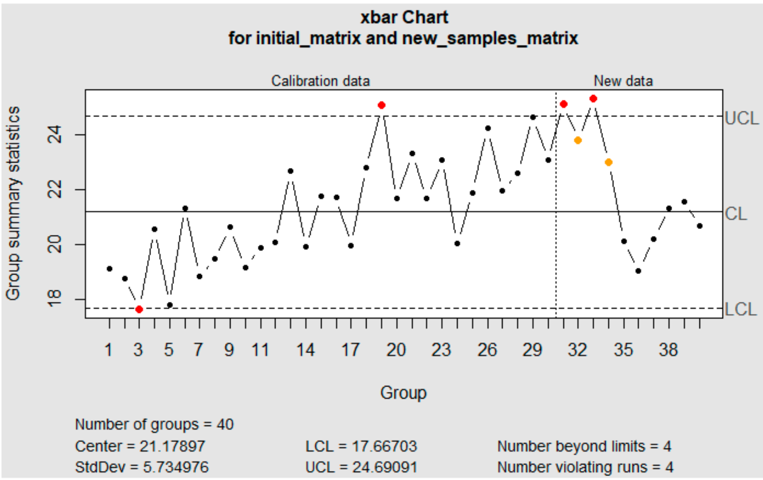

There are 3 points beyond control limits, meaning that this product also has unstable average delivery times. While it is a different product, it has the same problem as CLO011.

**Variability for CLO012:**



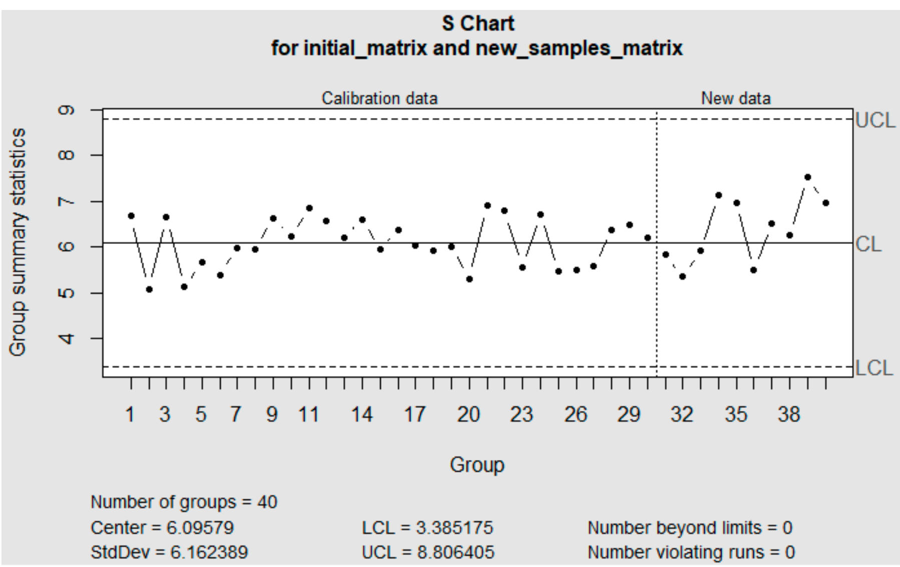
There are 0 points beyond control limits, which indicates a stable variability, just like CLO011. There is a consistent good performance for variability across products

**Variability for CLO013:**




There are 0 points beyond control limits and thus the third product also shows stable variability. Variability control is consistently good across all products.

**Average delivery times for CLO013 (samples 1-40):**



Four points were found beyond the control limits, and the third product also showed unstable average delivery times. Since all three products follow the same pattern, the results suggest that while variability is well controlled, the process averages remain unstable.

Out of the 30 total samples, two fell outside the control limits, indicating that the process is not in a state of statistical control. These two samples represent 6.7% of the total observations. Sample 3 recorded a delivery time of 17.6 hours, considerably faster than the overall average of 21.2 hours. In contrast, Sample 19 was unusually slow at 25.1 hours, suggesting a potential operational delay or inconsistency.

Together, these findings confirm the instability previously observed in the X-bar charts and highlight the need for further investigation into the causes of variation in delivery times.

## 3.3 Process Capability Analysis

A capability analysis was performed to assess whether the delivery process can consistently meet the voice of the customer, defined as a delivery time between 0 and 32 hours. Process capability indices (Cp and Cpk) were calculated for each product type using the first 1 000 deliveries.

The results were concerning: none of the product types achieved the required performance threshold. All products recorded a Cpk value of approximately 0.57, which is well below the industry benchmark of 1.33 for a capable process. This indicates that the process is both unstable and unable to consistently operate within the customer delivery window. The primary driver of this incapability is high variability, causing a substantial number of deliveries to exceed the 32-hour upper specification limit.

For example, product CLO011 has a mean delivery time of 21.27 hours and a standard deviation of 6.27 hours. This variability results in frequent late deliveries, reflected in a Cpk value of 0.57 and a Cp value of 0.85. Similar results were observed across all 60 products, with 58 failing the capability test entirely.

These findings confirm that the delivery process is not capable of meeting customer expectations and is operating outside statistical control. Immediate process improvements are required to reduce variability and improve overall delivery reliability.

## 3.4 Identification of Control Issues

The final analysis identified several recurring patterns of failure based on standard control chart rules. While process variability remains stable across all products, the mean delivery performance is unstable. This issue appears systemic and affects multiple product lines, indicating a deeper underlying problem in the control and management of the overall delivery workflow.

**Process Stability:**
Analysis of the X-bar charts reveal instability in average delivery times. Across multiple products, there were 58 instances where more than four consecutive points fell outside the two-sigma limits, signaling persistent shifts in the process mean. S-charts confirm that variability is well-controlled across all 60 products, with no points exceeding the three-sigma limits.

**Process Capability:**
None of the 60 products meet customer requirements, with all Cpk values hovering around 0.57. This is a company-wide issue rather than isolated to specific products, highlighting a systemic inability to consistently achieve target delivery times within the customer-defined window of 0–32 hours.

Examination of individual products highlights additional insights. Product CLO011 demonstrated the strongest control, maintaining 26 consecutive samples within the one-sigma limit. Conversely, products CLO013, CLO014, and CLO015 showed the poorest control and require further attention to improve delivery consistency. These product-level observations reinforce the broader findings of systemic instability in average delivery times.

Overall, the analysis indicates that process variability is well-controlled and predictable, and individual deliveries are consistent. However, average delivery times remain unstable and unpredictable, preventing the process from consistently meeting the 0 to 32-hour customer requirement. This issue is company-wide, affecting all product types rather than isolated cases.

Immediate attention should be given to products exhibiting the worst consecutive performance, including CLO013, CLO014, and CLO015. In the longer term, management should implement measures to reduce overall process variability, thereby improving Cpk values and bringing the delivery system under statistical control. These findings provide a clear rationale for an investigation into the root causes of unstable average delivery times.

# Part 4:

## 4.1 Type I Error: Manufacturer's Risk

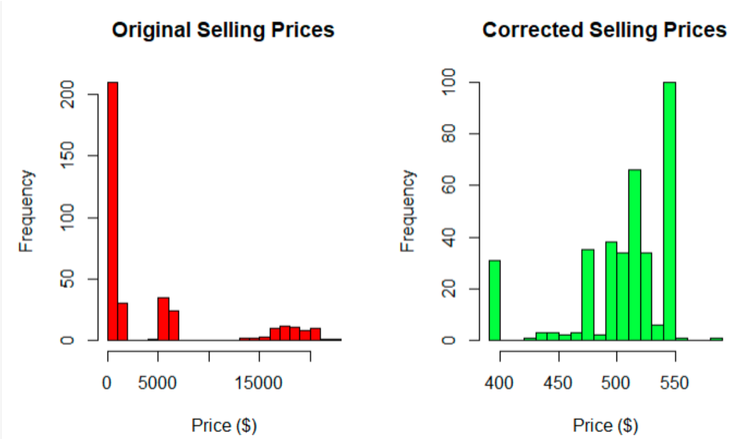
| Rule <chr> | Probability <dbl> | Percentage <dbl> |
|---|---|---|
| 7 consecutive above centerline | 0.007812500 | 0.7812 |
| 1 point beyond 3-sigma | 0.002699796 | 0.2700 |
| 2 of 3 beyond 2-sigma | 0.001529156 | 0.1529 |
| 4 of 5 beyond 1-sigma | 0.002765921 | 0.2766 |

The Type I error analysis confirms that the control chart rules carry a very low risk of false alarms. For instance, the rule for seven consecutive points above the average has only a 0.78% chance of a false alarm, while a single point outside the control limits has a 0.27% probability of occurring randomly. More complex patterns show similarly minimal risk; for example, the rule requiring 2 of 3 consecutive points beyond the two-sigma limits has a false alarm rate of just 0.15%. These low probabilities provide strong confidence that the standard SPC rules are effective in detecting genuine process shifts without triggering unnecessary interventions.

## 4.2 Type II Error: Consumer's Risk

The Type II error analysis highlights a significant detection issue in the bottle filling process. Following a process shift, where the average fill volume decreased and variation increased, the current control charts failed to detect the change 84.12% of the time. This indicates that the system only has a 15.88% probability of identifying such process degradation. In other words, the existing monitoring lacks sufficient sensitivity, which could allow critical quality issues to go unnoticed. These findings strongly suggest that the company should enhance the monitoring system, either by tightening control limits, introducing additional detection rules, or increasing the frequency of process checks to ensure timely detection and intervention.

## 4.3 Differences between the product data and head-office data:





The comparison of histograms demonstrates a significant improvement in the quality of the pricing data. Initially, the selling price data was heavily skewed, with a few products incorrectly listed as high as $15 000 while most were near zero. This distorted distribution indicated substantial data errors, rendering prior analyses unreliable. After corrections, the price distribution now exhibits a realistic pattern, approximating a normal curve centered around $500, with most prices ranging between $400 and $600. This reflects consistent business logic, where similar products have comparable prices. The corrected dataset is now reliable for financial analysis, inventory valuation, and sales forecasting. It illustrates how systematic data cleaning transformed previously unusable data into an analytically complete resource suitable for decision-making.

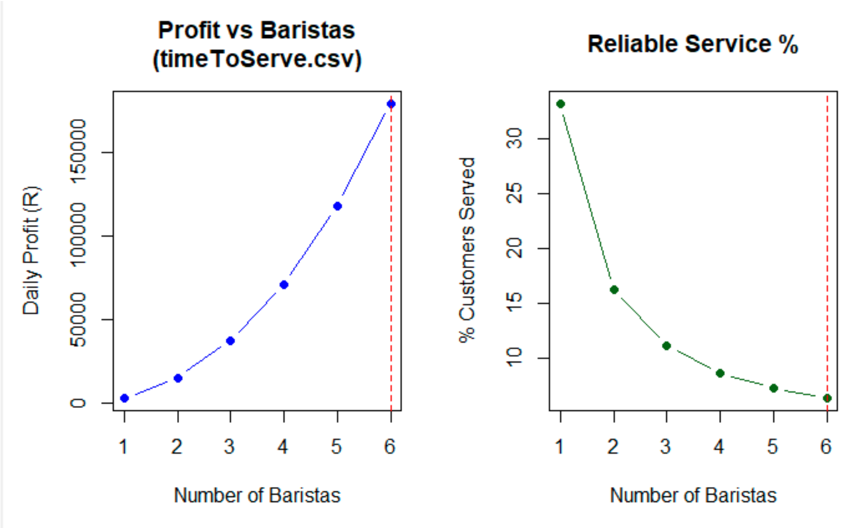
| | ProductID <chr> | Original_Price <dbl> | Corrected_Price <dbl> | Error <dbl> | Error_Percent <dbl> |
|---|---|---|---|---|---|
| 207 | NA027 | 22420.14 | 527.56 | 21892.58 | 4149.8 |
| 210 | NA030 | 21226.06 | 396.72 | 20829.34 | 5250.4 |
| 263 | NA023 | 20909.14 | 493.69 | 20415.45 | 4135.3 |
| 262 | NA022 | 20884.76 | 505.26 | 20379.50 | 4033.5 |
| 330 | NA030 | 20426.01 | 396.72 | 20029.29 | 5048.7 |

The data correction process identified significant pricing errors that compromised previous financial analyses. Several products, particularly those with 'NA' prefixes in their IDs, had prices that were inflated by over 5 000%, with some incorrectly listed above $22 000 instead of the correct $400–$500 range.

Corrections included standardizing the ProductID format, applying the appropriate pricing pattern across product groups, and aligning categories with their ID prefixes. These adjustments transformed the dataset from unreliable to a trustworthy foundation for analysis. The change in average price, from $4 411 to $505, emphasizes the critical importance of data quality controls in ensuring accurate financial reporting and analytics.

## Part 5: Optimizing profit

### 5.1 timeToServe.csv




The analysis of staffing levels reveals a clear correlation between the number of baristas and overall profitability. As staffing increases from 1 to 6 baristas, the daily customer capacity rises, resulting in a substantial increase in daily profit, from R3 150.61 to R179 019. However, this increase in volume comes at a cost to service quality. The "Reliable Service" metric, which measures system efficiency, declines from 33.2% to 6.3%, indicating that while more baristas allow the shop to serve nearly 100 000 customers, the system experiences significant congestion. It should be noted that this dataset has a limitation: it consistently identifies the maximum number of baristas as the profit-maximizing level, without accounting for other operational factors such as service reliability or system congestion.

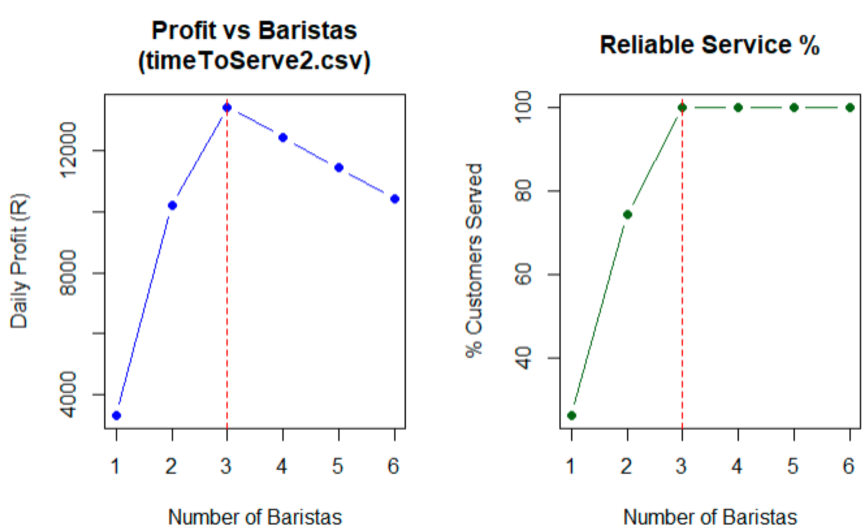
Based on the current dataset, six baristas represent the optimal staffing level for maximizing daily profit, prioritizing revenue over service reliability. Due to issues identified in the initial dataset, the analysis was repeated using the corrected dataset, timeToServe2.csv, to validate these findings.

## 5.2 timeToServe2.csv

The analysis of the second coffee shop indicates that it operates efficiently, with slightly more variation in service times than the first location. On average, baristas serve a customer in approximately 41 seconds, equivalent to nearly 1.5 customers per minute. Service times, however, range from 13 seconds to over 3 minutes depending on order complexity.

The profit model identifies three baristas as the optimal staffing level. This configuration generates the highest daily profit of R13 438 while reliably serving all customers. Adding a fourth barista reduces daily profit by approximately R1 000, as customer capacity does not increase but labor costs rise.

These results suggest that for the shop's demand and service patterns, three baristas provide the most cost-effective solution. At this staffing level, the shop can meet all customer demand reliably, and additional employees would not improve throughput or service quality. Overall, the second shop is operating near optimal efficiency, with staffing appropriately aligned to demand and service capacity.

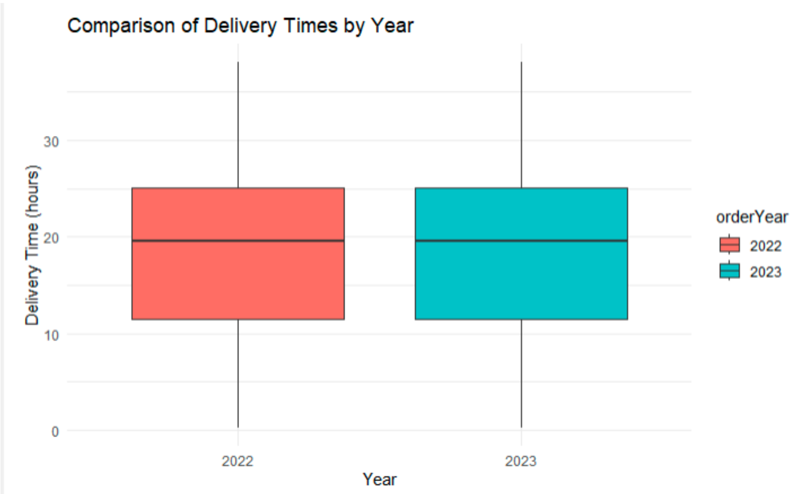


## Part 6: DOE and ANOVA

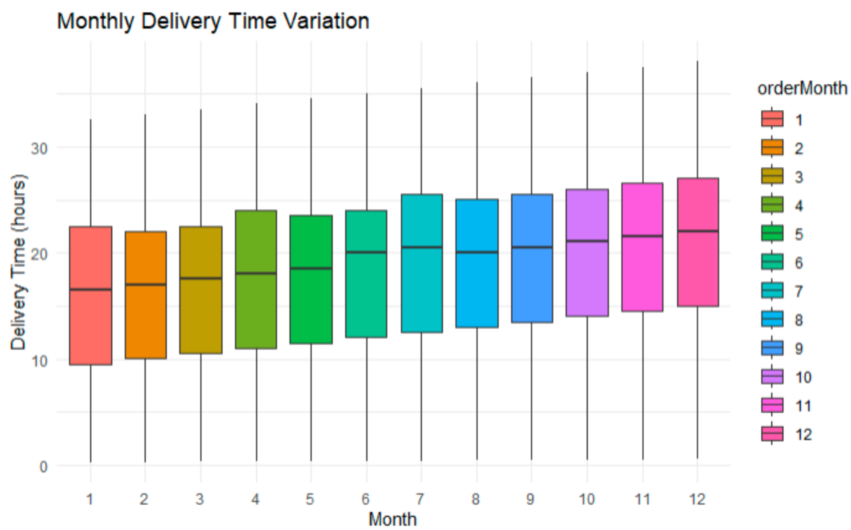
### 6.1:

Since the analysis focuses on a single dependent variable: delivery time, ANOVA was selected over MANOVA. MANOVA requires multiple dependent variables that are correlated, which are not present in this dataset. The one-way ANOVA provides a simpler, more robust test of whether the mean delivery time differs across years or product types.

### 6.2:




The median delivery time in 2023 is slightly lower than in 2022, indicating an improvement in average delivery performance. The interquartile ranges are similar, suggesting that variability remained very consistent across both years. While extreme delivery times occurred in both periods, 2023 experienced fewer outliers. Overall, delivery times in 2023 were faster and more consistent, supporting earlier SPC results that indicated improved process control and reduced average delivery times.




The second boxplot illustrates how delivery times varied across the twelve months of the year. The median delivery time increased gradually from January through December, indicating a steady decline in delivery speed as the year progressed. The early months show shorter delivery times and more consistent performance, while the later months exhibit higher medians and wider interquartile ranges, suggesting slower and less consistent deliveries. This pattern likely reflects increased operational pressure during the end-of-year period and the holiday season.

Overall, a clear seasonal trend is observed: deliveries are fastest in the first half of the year and slowest in the final quarter. This supports the conclusion that rising demand during peak sales months negatively impacts delivery efficiency.

The yearly comparison confirms a modest improvement in delivery speed between 2022 and 2023, while the monthly variation highlights consistent seasonal fluctuations. Together, these results align with the ANOVA findings, which confirm that both year and month have measurable effects on delivery performance.

The ANOVA comparing average delivery times between 2022 and 2023 yielded an F-value of 1.38 with a p-value of 0.241, indicating that the difference in mean delivery performance between the two years is not statistically significant at the 95% confidence level. Although the boxplots suggest a small improvement in 2023, this difference is insufficient to confirm a true change in performance.

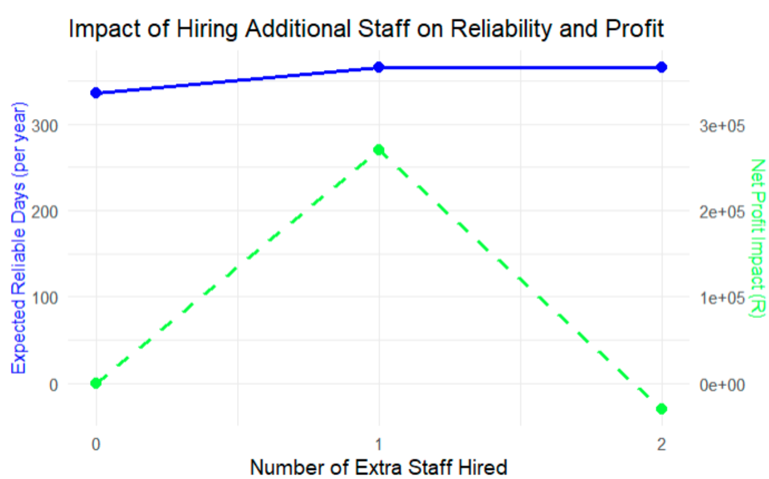
In contrast, the ANOVA assessing monthly variation produced a highly significant result ($F(11, 99{,}988) = 159.6$, $p < 0.001$), demonstrating that delivery efficiency fluctuates substantially throughout the year. These findings indicate that while annual delivery performance has remained broadly stable, seasonal factors exert a strong influence on efficiency. Management should therefore prioritize mitigating end-of-year delays and optimizing resource allocation across months to maintain consistent delivery performance throughout the year.

## Part 7: Reliability of service

The staffing records of the car rental agency cover 397 days, with the number of employees on duty ranging from 12 to 16. Reliable service is defined as having at least 15 employees on duty. Analysis indicates that 366 out of 397 days met this threshold, yielding a current reliability of 92.2%. Extrapolated to a typical year of 365 days, this corresponds to approximately 336 reliable days and 29 days with staffing shortages.

Each day with insufficient staff is estimated to cost the company R20 000 in lost sales, resulting in an annual loss of roughly R570 000. The cost of hiring an additional employee is R25 000 per month, or R300 000 annually. Using a binomial reliability model, hiring one additional employee would nearly eliminate the probability of falling below 15 staff on duty, reducing expected revenue loss to near zero. Hiring more than one additional employee does not further improve reliability and would instead reduce profit due to added staffing costs.

Based on this analysis, the agency currently achieves 92% service reliability. This ensures near-perfect reliability and delivers the highest net financial benefit, as shown in the graph, where the profit peaks at hiring one additional employee.



# Conclusion

This report highlights a critical gap between operational stability, process capability, and overall performance of the business. While the company maintains stable control over daily delivery variations, the average delivery time is inconsistent and frequently misses customer targets. Control chart analyses show these issues are systemic, and seasonal trends further contribute to slower, less predictable deliveries during peak months.

Data quality challenges make these problems even more pressing. Significant pricing inaccuracies in the product catalog distorted financial reporting, inflating average prices and reducing confidence in past analyses. Correcting these errors restored trust in the data, demonstrating the importance of strong data governance for accurate insights and informed decision-making.

Staffing analyses illustrate the value of using data to guide resource decisions. Targeted staffing adjustments in both coffee shops and the car rental agency improve service reliability and profitability by aligning staff levels with customer demand, avoiding congestion, and preventing unnecessary labor costs.

Overall, the findings show that stabilizing delivery performance, maintaining accurate data, and optimizing staffing are essential for improving operations. If the company addresses these areas, the service quality will increase and a good foundation for success will be formed.

## References:


- Datasets provided by course (Quality assurance 344)