Quality Assurance ECSA GA4 Report

Turin Lambrechts [27070409]

The Department of Industrial Engineering

Stellenbosch University

26 October 2025

Table of contents

List of Figures

List of tables

1. Introduction

As an organizational data source, this report covers a broad analysis of sales, product, and customer data of a company from 2022 to 2023. We mainly aim to investigate a few operational and customer information for business optimization based data-driven solutions. The analysis looks deeply into four aspects of sales trends, customer demographics, service times and the optimization of staffing. Based on the seven main parts of the report, this is one of the most important parts that you can prepare based on your experience. Descriptive & Exploratory Data Analysis: Data analysis for sales, demographics, and product insights. Data Correction and Product Alignment: It means to remedy mismatches in product data; to see what effect this has on sales or revenue. SPC (Statistical Process Control): A process stability and capacity analysis for quality control. Risk and Data Correction: Type I and II errors and errors in product information. 5. A model to optimize profitability by setting staffing levels according to service times. ANOVA & MANOVA Analysis: Statistical tests of sales trend, income distribution, service time. Reliability of Service and Optimization: To assess service reliability and optimize and recommend best staffing to create results, avoid mistakes in performance and minimize sales loss. For all sections of the report numerous statistical methodologies (ANOVA, MANOVA and binomial models) have been performed in order to answer important business questions about the performance of the products and staffing requirements as well as customer choices. The interpretation from these analyses can be employed to derive solutions that contribute to a value-based profit strategy while driving both operational excellence and customer satisfaction.

2. Part 1.2: Descriptive statistics

2.1 Data Analysis Methodology
This section provides an analysis of a company's sales, product, and customer data for the years 2022-2023. The goal of this analysis is to evaluate the company's performance, uncover insights into customer behaviour, and create a data-driven foundation for future business strategies.

2.2 Approach:

1. Combining different datasets (sales, product details, customer information), a comprehensive sales dataset for detailed analysis was obtained.
2. The datasets were examined for inconsistencies, missing values, and outliers. Data transformations were carried out to ensure proper analysis quality.
3. Descriptive Statistics: Statistics such as mean, median, and standard deviation were computed to summarize the data and examine the distribution of the variables.
4. Visual Analysis: We generated several visualizations to analyze trends, patterns, and relationships in the data. These visuals help in finding out the most profitable products, sales trends over time, and customer demographics.
5. Customer Segmentation: This was not explicitly performed here, though a segmentation of customers based on purchasing behaviour (RFM analysis) was considered to identify groups for focused marketing.

2.3 Data Overview:

- **Sales Data**: 100,000 transaction records spanning from 2022 to 2023.
- **Product Data**: 60 unique products, categorized into six groups: Cloud Subscription, Keyboard, Laptop, Monitor, Mouse, and Software. Each category contains ten products.
- **Customer Data**: 5,000 individual customer records, including demographic details such as age, income, and location.
- **Data Completeness**: All datasets were verified for completeness, with no missing values detected.

2.4 Detailed analysis:

Sales and revenue

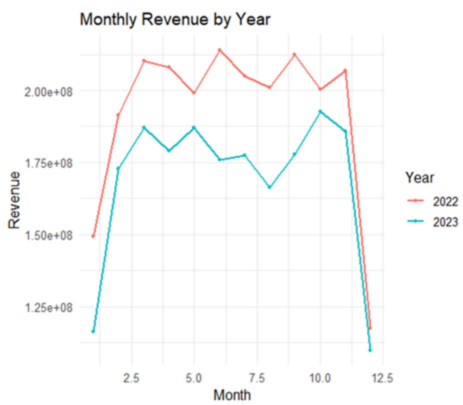


Figure 1: Monthly revenue by year

The trend of sales from the beginning of 2022 to the end of 2023 suggests a sudden decrease of these monthly sales around the end of each of the years. Such patterns indicate that the business is heavily influenced by seasonality, with large peaks and troughs in revenue across the year. This downturn at the end of 2022 suggests that the company has difficulty balancing spikes of product sales and can cause the company to struggle to maintain sales in the "quieter months".

The second observation is that the year-over-year revenue from 2022 for 2023 is smaller in 2023 than for 2022, the trend for both the years is similar in terms of seasonality but 2023 clearly performed worse than 2022. The Company needs to concentrate on discovering how to overcome these year-ended declines – and work towards minimizing seasonal variations and the performance deceleration last year to see performance improve overall. Improving strategy planning in these down years, however, could stabilize revenue and possibly improve sales reliability in future years. Also, investigating decline in overall revenue from 2022 to 2023 will ensure less underperformance in the future.

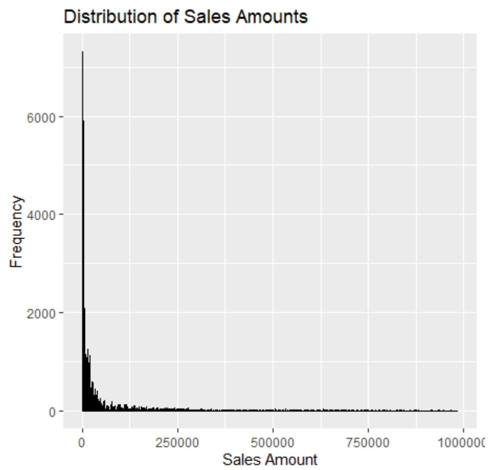


Figure 2: Distribution of Sales Amounts

The sales distribution graph demonstrates a significant aspect of a company's transactional profile. The data is right-skewed, with a peak showing a large amount of low-value transactions. However, the long tail shows that while fewer transactions are of high value, they contribute significantly to overall revenue. So the cash flow of the business is consistent from regular, low-value sales, and high-value transactions, although not frequent, have to a significant extent contributed to the total revenue.

Products and Categories

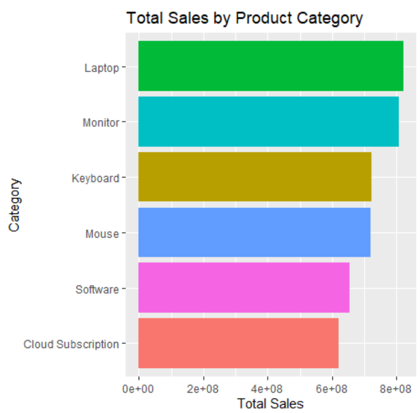


Figure 3: Total Sales by Product Category

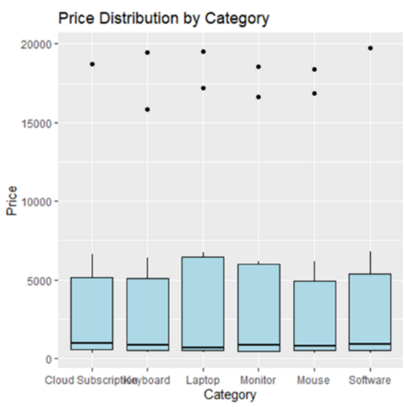


Figure 4: Price Distribution by Category

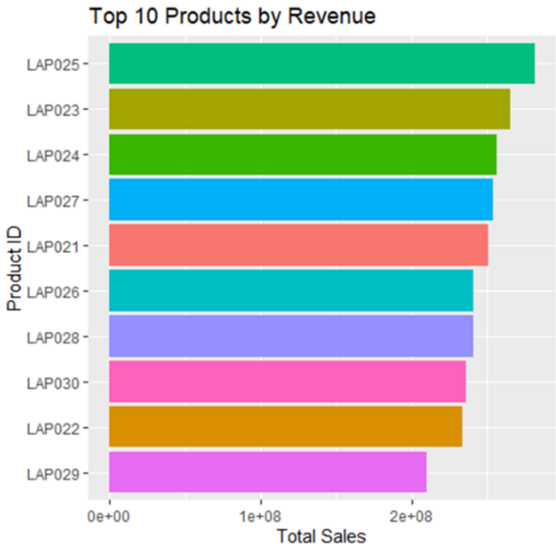


Figure 5: Top 10 Products by Revenue

The top 10 products by revenue, box plots, and the total sales by product category help you decide how much all category impact the total revenue.

<u>Revenue distribution across categories:</u>

Laptops and Monitors are clear market leaders, as evidenced by the total sales by category graph, where Laptops stand out with significantly higher sales than other categories. This confirms that a handful of product categories represent the bulk of the company's revenue. Software and Cloud Subscription categories, by contrast, make up much smaller parts of overall revenue. While these categories matter, they don't do as well as the Laptop and Monitor categories are, suggesting the need of tactical efforts in increasing sales in those lowest performing categories.

<u>The Focus on Laptops:</u>

Top 10 products by revenue confirms that Laptops dominate the revenue stream, meaning that all top-selling products in this category are in the top products. This again emphasizes that Laptops are critical to the company's financial health. These factors can severely impact the overall sales if the supply chain fails due to product shortages or demand changes. Considering the overwhelming market share of Laptops, the organization has to ensure to protect this product line and maintain consistent supply, pricing, and customer satisfaction. According to the review, Laptops need to be in a higher priority for resources while any disruption in the supply chain or operation is kept to a minimum.

<u>Diversification and strategy for underperforming categories:</u>

Based on the price distribution by category graph, it is evident, categories such as Software or Cloud Subscription exhibit lower total selling prices and narrow price range. Don't disregard these products, they should be important for diversifying your revenue. This will be where you have an opportunity to drive them with more robust promotional effort, bundled offers, marketing campaigns, or introducing new prices to stimulate these groups.

Tiered Product Strategy:

The top 10 products according to revenue also demonstrates how concentrated sales remain in a few products in some way, notably Laptops. There isn't much in common among products to get the bulk of the revenue, which then becomes an outlier. The result is quite an uneven flow of revenue where a few products take the lion's share of the revenues and the rest just barely get by. Having a tiered approach would help the company be able to get its head off running out of low-producing products, using its high-performing (like Laptops, who dominate those high-performing products) and high growth (like with middle-performing products) products to help grow their sales. For underperforming ones, organizations would need to perform cost benefit analyses to ascertain if they should be kept or sold to minimize unnecessary inventory costs.

Inventory Control And Product Handling:

The fact that top 10 products according to revenue charts can display price disparity, focus and sales concentration are concentrated in top 10 products suggests the company may be allocating too many resources to low revenue products. Though these products serve as part of the wider range of items, they are adding to the variety but, as yet, are at a cost of increased storage, management, and inventory loss in more instances than they are generating on sale. High-performance products should be at the forefront of the company's product strategy to optimize inventory efficiency and support efficiency around the middle players at every point. Decisions on product prioritization, inventory and marketing focus towards more effective revenue-generating product categories should be based on these learnings.

Pricing strategy

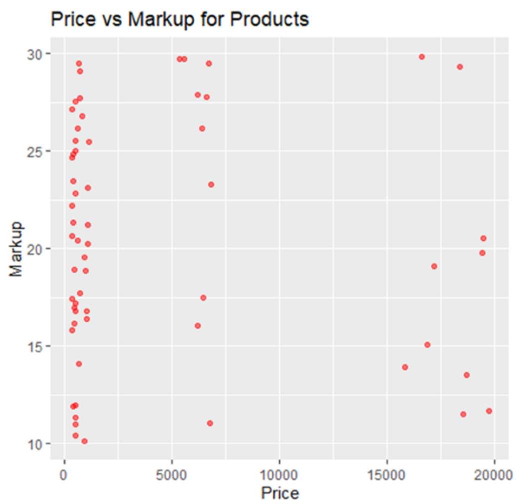


Figure 6: Price vs Markup for products

The Price vs Markup scatter plot indicates the large spread of products with low price (near 0) and moderate markup value between 10 and 20. High-priced products also exist (above 10,000), but these goods have lower markups implying that some of them may have been charged higher prices for competitive reasons or have higher production costs. Also, there are few luxury products that have extremely high markup signifying premium goods or products with good margins. However the plot proves that overall there is no generalised correlation between price and markup that can be applied to all products.

Customers

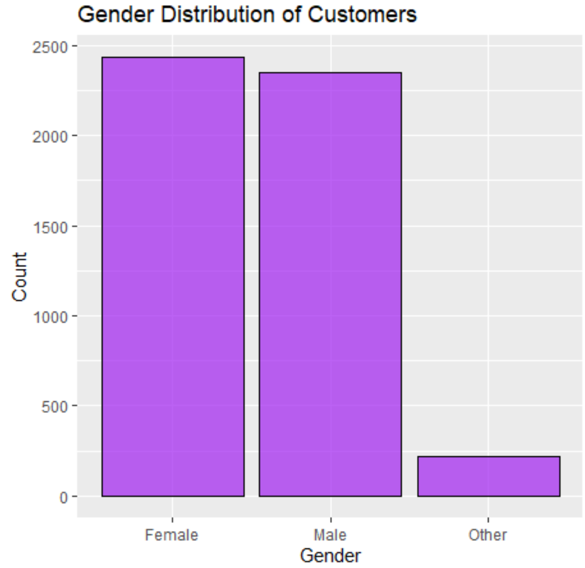


Figure 7: Gender Distribution of Customers

The distribution of gender shows that there are slightly more female customers than male. Nevertheless, the overall difference is slight. This suggests the business should take into account both sides equally, not either gender. Marketing strategies and products should look at both males and females to make an even contribution to each gender's customer base

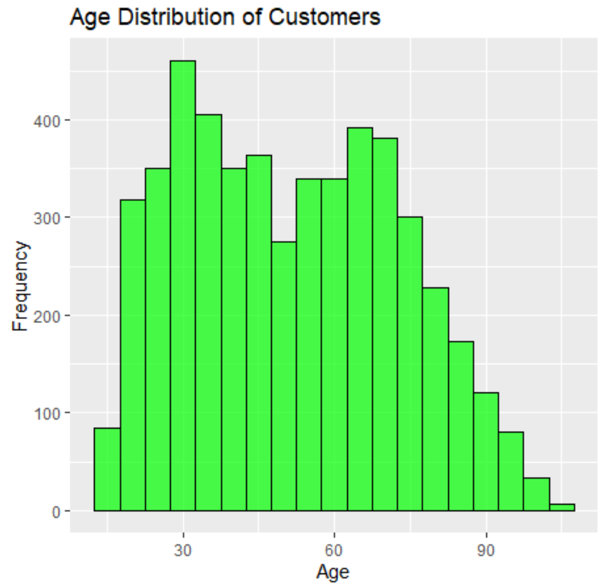


Figure 8: Age Distribution of Customers

Age distribution is bimodal with two largest age groups, those of 30 to 40 years, and those of 50 to 60 years. This indicates a high concentration of customers within these age brackets, probably representing the primary intended market of the company. Customers in these strata are often well income-rich and more inclined to purchase the high-value product. It goes on to show the distribution shows customers younger than 30 years are less likely to contribute in sales and customers over 60 make up a smaller portion of the overall customer base. There is an opportunity for the company to tailor marketing initiatives to drive consumer engagement among these prime middle-aged shoppers since these are the prime middle-aged shoppers potentially with higher disposable income and purchasing intent for expensive items.

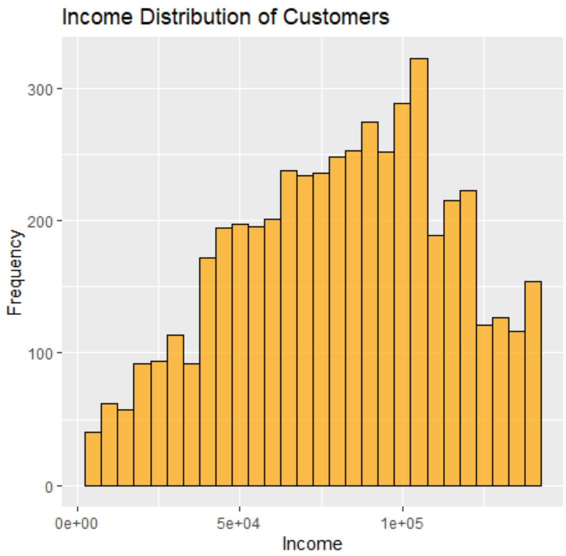


Figure 9: Income Distribution of Customers

The income distribution graph is a right-skewed one that shows a significant concentration of customers earning between 50,000 and 100,000. That means that people from middle to upper-middle incomes can buy more mid-range and higher-value products. Due to this distribution, the businesses should concentrate on targeting high earning customers who are the most profitable segment. This kind of marketing could attract these very customers, which could help to boost sales in higher priced categories such as Laptops and premium Monitors.

Operational efficiency

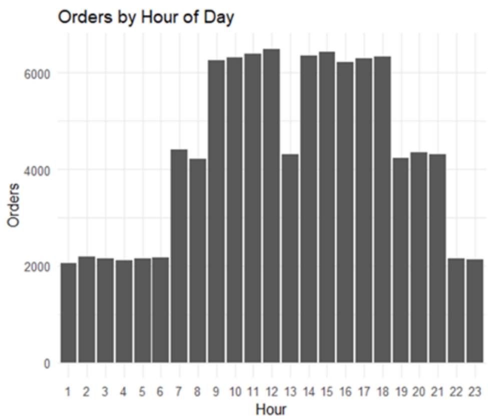


Figure 10: Orders by Hour of Day

The order volume bar chart by hour the bar chart of orders per hour shows that the peak orders are between 10:00-12:00 and then for the 18:00–20:00 window. This means that the company experiences more peak sales when demand is at its strongest and greater staffing need to be done to control staffing at these hours. Staffing increase or process improvement in those times could lead to better response times and customer satisfaction.

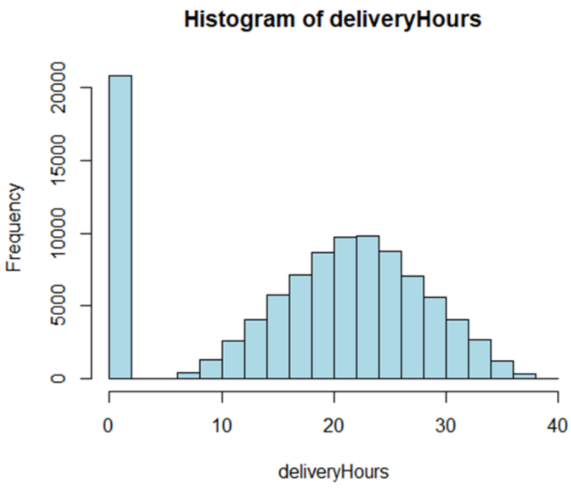


Figure 11: Histogram of deliveryHours

The histogram of delivery hours is heavily skewed, where most deliveries fall as follows- most within the range of 1 through about 15 hours, but the long tail extends up to the 30–40 hour range. This indicates that, while it is usually feasible to fulfil in the limit of a certain time, many of the deliveries happen to be delayed much longer than the normal range. Such anomalies are representative of delivery process inefficiencies which require investigation to ascertain the cause: one is delivery partner issues, logistical delays, or warehouse bottlenecks.

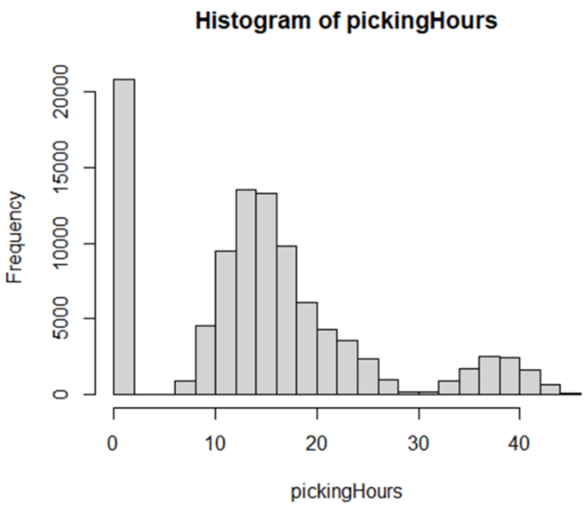


Figure 12: Histogram of pickingHours

A histogram showing picking hours shows that almost all picking happens between 0 and 10 hours, hence, displaying a skewed distribution of the hours. As with delivery times, though, there are outliers that have the picking process taking more than expected, with some times going over 30 hours. Time delays indicate that the picking process is not efficient. These delays are critical if we want to significantly enhance the overall order fulfillment time as the long picking times tend to cascade into the delivery times.

• The busiest order time (10:00–12:00 and 18:00–20:00) clearly shows the need for adequate staffing to fulfill demand more efficiently.

• Long delivery and picking processes have clear operational inefficiencies that must be solved. Most of the orders have met a certain time frame but the outlier is a big issue in the setup and should not be underestimated.

• Better picking and delivery logistics that can limit long wait times can also be improved, thus increasing customer and operational satisfaction.

2.5 Recommendations

Managing Seasonal Revenue Fluctuations:

Company has a high seasonal fluctuation, with the slow-down towards year-end. To increase its profit margin, the company might want to examine what is driving that dip and how to make its business more effective during slower times. Strengthening year-end strategy will also allow sales to stabilize and may bring some level of stability to sales in the future.

Maximizing High-Performing Product Categories:

Laptops are one of the largest segments of revenues and the company should focus on this because it's one of the biggest factors that impacts supply and where it competes in the industry. Supplier disruptions could unfairly affect the global performance of the business. They must also improve sales of the underperforming categories such as Software and Cloud Subscription with focused advertising or promotional campaigns.

Diversifying Product Strategy:

A multi-tier product strategy with high-growth products (Laptops) as its base, however would help to boost the sales of middle-growth products. The company should also evaluate low-performing and expensive products and analyze if they should be eliminated or retired with a view to avoiding excess cost associated with inventory.

Better Stock Management and Operational Efficiency:

For the price and sales concentration graphs, resources are maybe unfairly spread over low-revenue products. An inventory management strategy designed for higher levels of efficient, quality products and products in the middle-of-the-line will also yield profitability in greater depth with a more limited stock-management solution. Other aspects such as operational processes including warehouse management and logistics can also be improved so as to minimize turnaround times and order fulfilment times.

Focused on High-Income Clients:

Income distribution graph of customers shows a concentration in middle to upper-middle income groups. Most profitable are these customers. The company is advised to emphasize marketing to this segment in order to facilitate higher value product sales (Laptops, premium Monitors etc).

Optimizing Staffing in Load Periods:

The hour by hour orders graph indicates the highest order volume of each hour period (10:00–12:00 and 18:00–20:00). To fulfill demand in an efficient manner and enhance customer satisfaction, the company must add staff at these peak hours. In conclusion, staff management will optimize the work schedule and ensure the company can handle orders in high quantity with less lag.

Reducing Operational Delays:

There are serious operational inefficiencies across the delivery hours as well as picking hours histograms; some orders delay more than needed length of time. The company needs to find out what the underlying causes of these delays are; whether it be warehouse inefficiencies, supplier problems or delivery partner problems. Improvements in customer experience will result from streamlined picking (delivery logistics and picking) which will enhance order fulfilment effectiveness with higher customer satisfaction level.

Such recommendations are aimed at streamlining the operational processes in order to improve operational efficiency, stabilize seasonal sales, and maximize profits, targeting the company's most important product categories, high-value customer groups, and bottlenecks within the fulfilment process. Adopting these strategies will assist by improving performance for the business as a whole in the days to come.

Part 3: Statistical Process Control

3.1 Data Preparation and Control Limits Setup

The primary objective of this section is to assess the delivery process for various product types using X̄ (mean) and s (standard deviation) control charts. The dataset provided (sales2026and2027Future.csv) was sorted chronologically by Year, Month, Day, and orderTime to simulate real-time data collection. The analysis was based on samples of 24 observations per process.

To establish the initial control limits, the first 30 samples (30 × 24 = 720 data points per product type) were used to compute the centre lines, upper control limits (UCL), and lower control limits (LCL). The following constants were used for sample size n = 24:

- A3 = 0.619 (for X̄ chart)
- B3 = 0.555 (for s chart)
- B4 = 1.445 (for s chart)

The following product types were analyzed: Mouse (MOU), Keyboard (KEY), Software (SOF), Cloud Subscription (CLO), Laptops (LAP), and Monitors (MON).

The X̄ and s control charts were then created for each product type using the first 30 samples as the baseline to calculate the control limits.

X̄-chart and s-chart Calculations:

- Centre Line (X̄):
  The centre line for the X̄-chart is the mean of the sample means (denoted as x̄): where (k) is the number of samples (30 in this case), and ((x_i) ̅) is the sample mean of each sample.

$$\bar{x} = \frac{\sum_{i}^{k} \overline{x_{\iota}}}{k} \quad (1)$$

- Upper and Lower Control Limits (UCLx, LCLx):
  These control limits are set using the A3 constant and the average sample standard deviation (s̄): where (s ̅) is the average sample standard deviation.

$$UCL_{\bar{x}} = \bar{x} + A3 \cdot \bar{s} \quad (2)$$

$$LCL_{\bar{x}} = \bar{x} - A3 \cdot \bar{s} \quad (3)$$

- 1σ and 2σ Control Limits:
  The 1σ and 2σ limits are derived from the upper control limits:

$$U1_{\bar{x}} = \bar{x} + \frac{1}{3}(UCL_{\bar{x}} - \bar{x}) \quad (4)$$

$$L1_{\bar{x}} = \bar{x} - \frac{1}{3}(UCL_{\bar{x}} - \bar{x}) \quad (5)$$

$$U2_{\bar{x}} = \bar{x} + \frac{2}{3}(UCL_{\bar{x}} - \bar{x}) \quad 6$$

$$L2_{\bar{x}} = \bar{x} - \frac{2}{3}(UCL_{\bar{x}} - \bar{x}) \quad 7$$

- s-chart Calculations:
  - Centre Line (s̄):
    The centre line for the s-chart is the average standard deviation of all samples:

$$s = \frac{\sum_{i=1}^{k} s_i}{k} \quad 8$$

  - Upper and Lower Control Limits (UCLs, LCLs):
    These control limits for the s-chart are set using the B3 and B4 constants:

$$UCL_s = B4 \cdot \bar{s} \quad 9$$

$$LCL_s = B3 \cdot \bar{s} \quad 10$$

  - 1σ and 2σ Control Limits for s-chart:

$$U1_s = \bar{s} + \frac{1}{3}(UCL_s - \bar{s}) \quad 11$$

$$L1_s = \bar{s} - \frac{1}{3}(UCL_s - \bar{s}) \quad 12$$

$$U2_{\bar{x}} = \bar{x} + \frac{2}{3}(UCL_{\bar{x}} - \bar{x}) \quad 12$$

$$L2_s = \bar{s} - \frac{2}{3}(UCL_s - \bar{s}) \quad 12$$

3.2 Process Monitoring and Stability

After the control limits were defined, the subsequent samples (31, 32, …) were then monitored for delivery times. The mean (X̄) and standard deviation (s) for each sample were compared with their respective limits to investigate potential out-of-control signals. The following SPC rules were used:

• Rule A: Any s-sample exceeding the +3σ limit (UCLs) was considered as out-of-control (since this was the error scale).

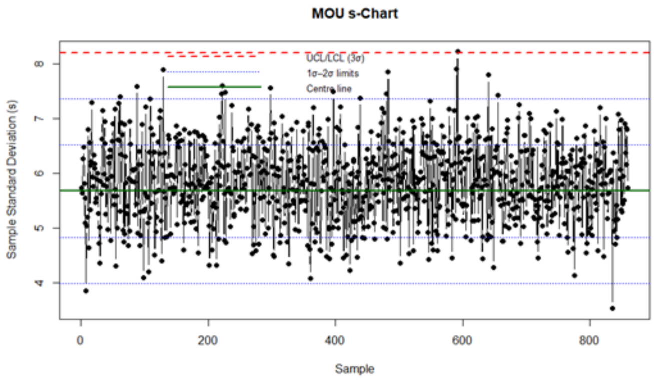
• Rule B: Longest sequence of s-values within ±1σ around the centre line, show stable control of the process.

• Rule C: Four or more consecutive X̄ samples outside the +2σ limit, indicating change in the process mean.

Based on the results, there are some great variation in product types that occurred for some products, in particular, for MOU and MON, other (e.g., SOF) achieved greater stability in the control.

3.3 Analysis of $\bar{X}$ and s Charts

Product Type: Mouse (MOU)



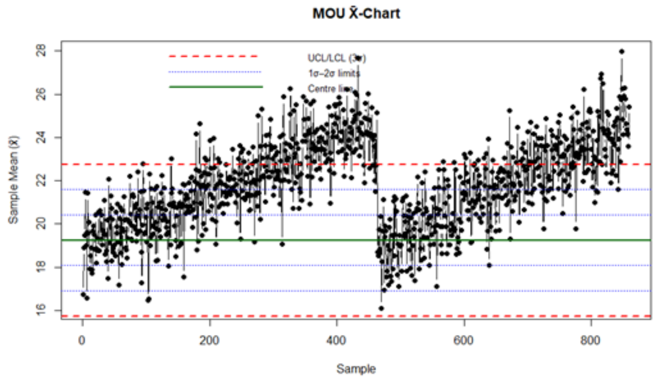
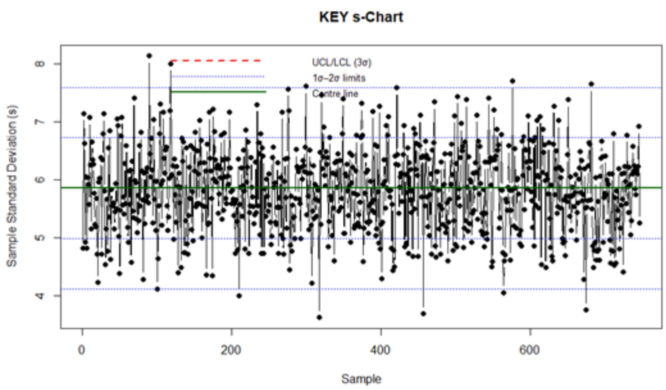


Figure 13: Mouse s control chart

Figure 14: Mouse $\bar{X}$ control chart

- • $\bar{X}$ Chart—Mouse has mean delivery times within the control limits. However, some points approach the upper control limit (UCL), indicating that there are occasional shifts in delivery times. These may be attributed to external factors or temporary bottlenecks.
- • s Chart: Standard deviation also shows a relatively stable spread but a spike around sample 600, which exceeds the UCL, demonstrating a very high spread of delivery time variability. That will need more analysis to see if it’s a one-time or part of a larger pattern.

Product Type: Keyboard (KEY)



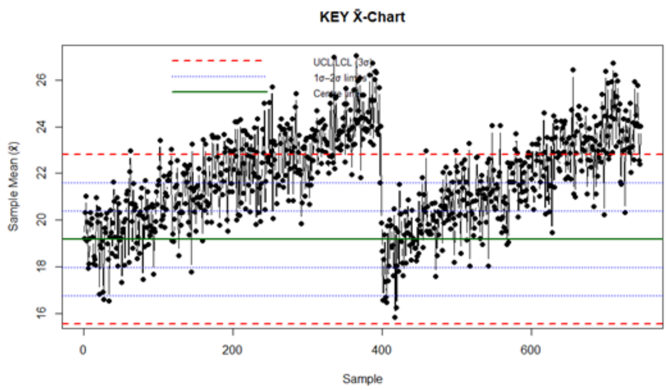


Figure 15: Keyboard s Control chart

Figure 16: Keyboard $\bar{X}$ control chart

• $\bar{X}$ Chart: The delivery time for Keyboards has mainly been stable, with only a slight upward trend near sample 400, implying that delivery times may be drifting over time but remain within acceptable limits for most of the process.

• s Chart: The spread remains well-controlled, with only slight fluctuations between the upper and lower control limits. This means that the variability in delivery times for Keyboards is generally low and stable, with no major issues observed.

Product Type: Software (SOF)

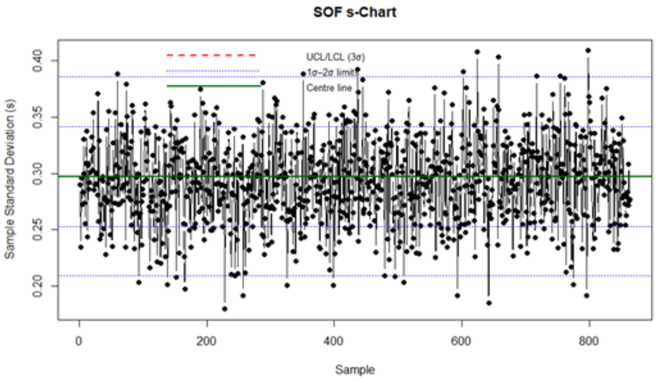


Figure 17: Software s control chart

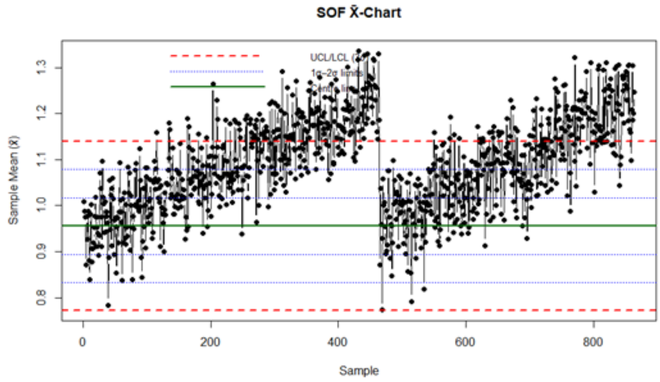


Figure 18: : Software X̄ control chart

• X̄ Chart: The mean delivery times for Software remain well within control limits throughout the year. There is a noticeable improvement from 2026 to 2027, indicating that the process is becoming more stable and predictable over time.

• s Chart: The spread of delivery times for Software is narrow and tightly controlled. No significant deviations are seen, suggesting that this product line has a high level of consistency in terms of delivery time.

Product Type: Cloud Subscription (CLO)

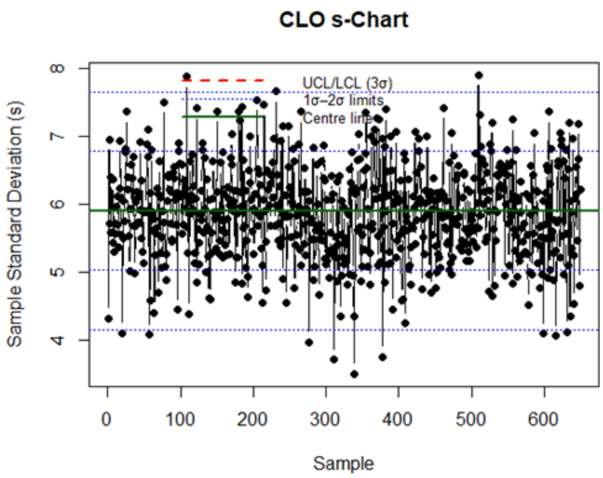


Figure 19: Cloud subscription s control chart

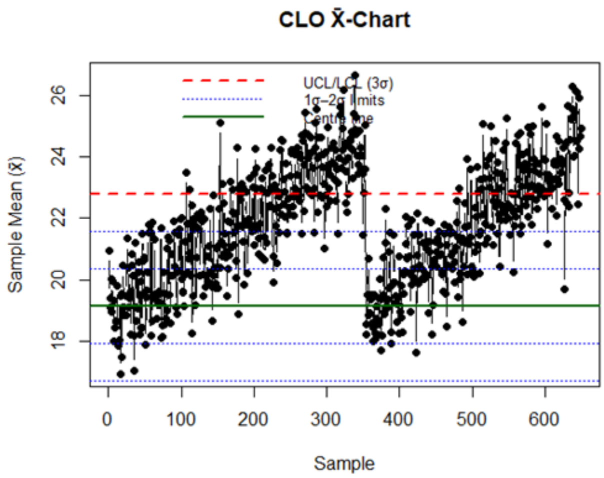


Figure 20: Cloud subscription X̄ control chart

- • X̄ Chart: Cloud Subscription delivery times show occasional spikes above the upper control limit. These spikes represent periods when the process is outside of control and should be investigated for potential causes, such as supplier delays or inefficiencies in dispatching.
- • s Chart: The spread remains relatively consistent with few excursions outside control limits. A slight decrease in variability is observed between samples 270-340 and 560-630, indicating that some improvement is happening in the delivery process

Product Type: Laptop (LAP)

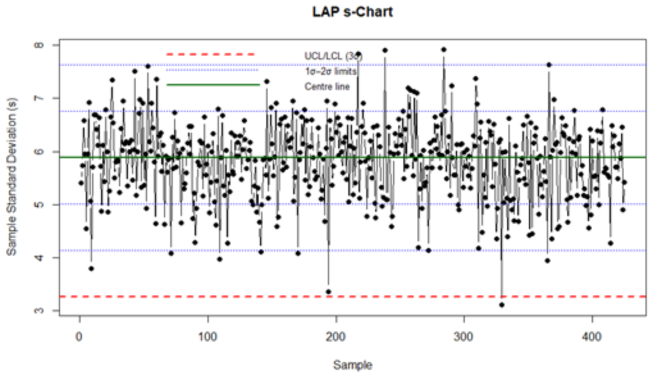


Figure 21: Laptop s control chart

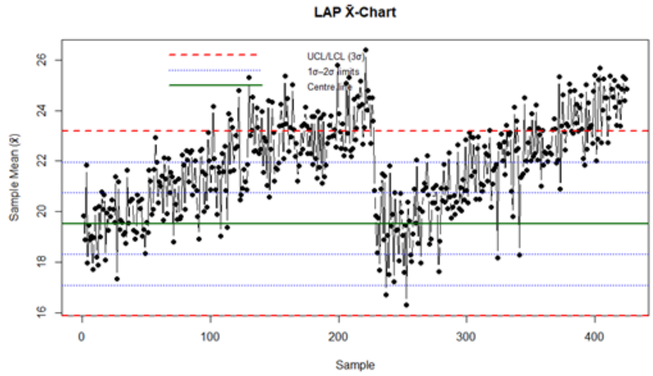


Figure 22: Laptop $\bar{X}$ control chart

• X̄ Chart: Laptops are more stable in delivery times than desktop offerings and exhibit low volatility year-round. The average delivery times, however, show a gradual increase from mid-2027 onwards, suggesting perhaps a shift that would need to be rectified.

• s Chart: The variation of Laptop delivery times appears to be within control limits, but there is an occasional drop near the lower control limit. That might suggest that processes are more efficient in those periods—which could be a guide to better the entire process.

Product Type: Monitors (MON)

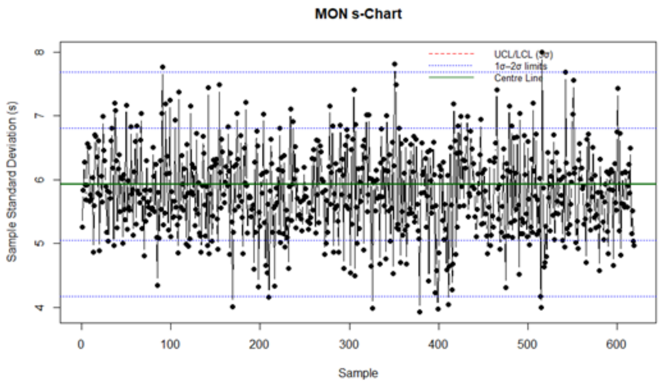


Figure 23: Monitor s control chart

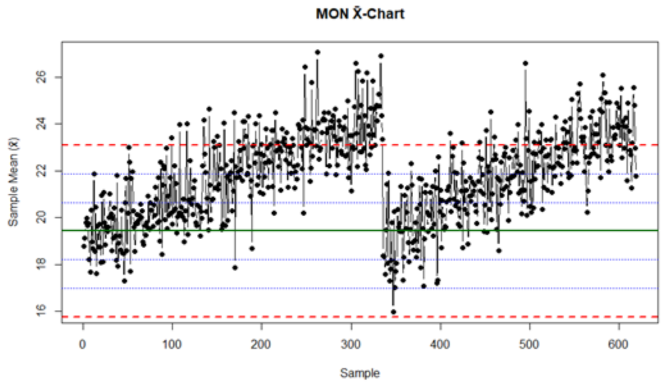


Figure 24: Monitor $\bar{X}$ control chart

• X̄ Chart: The mean delivery time from Monitors shows significant variation across the samples, with several instances where the process has drifted above the upper control limits. This suggests that consistency in delivery times may be an issue.

• s Chart: The spread of the delivery times is largely controlled, with minor deviations towards the upper control limit. This implies that the variability in the delivery time of Monitors is controllable but that the process mean needs to be re-explored and intervention indicated.

3.4 Process Capability Indices

The process capability indices (Cp, Cpu, Cpl, Cpk) were generated for every product type according to the first 1,000 deliveries. Most products, except Software (SOF), have Cpk values below 1, indicating that these processes are unable to meet the 32-hour upper specification limit (USL).

• Software (SOF) — Cpk 1.08 represents the closest to customer requirement but no products have a Cpk ≥ 1.33, which is generally the minimum that is normally considered the criteria of capable processes.

Table 1 : Process Capability Indices (Cp, Cpu, Cpl, Cpk)

Product Type	Cp	Cpu	Cpl	Cpk
MOU	0.92	0.73	1.10	0.73
KEY	0.92	0.73	1.10	0.73
SOF	18.13	35.19	1.08	1.08
CLO	0.90	0.72	1.08	0.72
LAP	0.90	0.70	1.10	0.70
MON	0.89	0.70	1.08	0.70

3.5 Rule Violations and Control Issues. Following SPC rules were applied to identify and control process issues:

• Rule A (s > UCL): Only Mouse (MOU) had a violation, with one sample being above the UCL for variability.

• Rule B (longest run within ±1σ): Cloud Subscription (CLO) carried out 35 consecutive samples and had the longest good control run that occurred within ±1σ.

• Rule C (4 consecutive $\bar{X}$ samples above +2σ): The two MOU and MON were the subjects of the highest violations. This finding shows large intervals of long-term shift in mean delivery time by these product types and is to be investigated for possible reasons that explain the changes such as inefficiency of the process or delays outside the end user.

3.5 Recommendations Based on SPC Analysis

Provision Process Optimization for MOU and MON:

o Mouse (MOU) and Monitors (MON) consistently have inconsistent mean shifts (Rule C) meaning that the means of delivery for the various categories of products are non-controlling. If stockouts, unreliable suppliers, or inefficient order picking processes are to blame, then these fluctuations can be inextricably linked.

o Suggestion: Examine possible bottlenecks in the warehouse or dispatch mechanisms for these product lines and introduce procedures to help with delivery time stabilisation.

Reiterate the emphasis on Software and Cloud Subscription

o SOF has produced stable delivery results, staying under constraints. Cloud Subscription (CLO) showed some variances that ought to be corrected for overall enhancement of performance.

o Recommendation: The software in the application needs to be the main feature for the present stability of it and further improvement. For Cloud Subscription, you may want to work on the predictability for a delivery by looking into the supplier relations or distribution change.

Laptops: Capitalise on Improvement

o Laptops (LAP) are delivering on average time but, while the average delivery time steadily increases throughout the year.

o Recommendation: Laptops need tighter controls at a period like peak demand, with the most optimal delivery time performance. Any increase in delivery time should be examined for shifts or surprises.

Address process variation for high-demand products

Monitors (MON) on s-chart have relatively stable variability, but this product requires constant attention for destabilizing features in future samples.

o Recommendation: Check s-chart variability for Monitors frequently to ensure that the process is being controlled, and maintain control on any variation outside the control limits.

Continuous Monitoring:

o The SPC charts show signs of potentially out-of-control signals over time, particularly with Rule C violations. This reflects systematic issues that could affect overall delivery performance and customer satisfaction.

o Recommendation: Install a real time monitoring solution ($\bar{X}$ and s charts with live tracking) to monitor delivery performance and take quick action after out of control signals.

3.6 Conclusion

We identify delivery processes for each product type and note gaps. The main problems highlighted are mean shifts in MOU and MON (Rule C) and inconsistent delivery times for Cloud Subscription (CLO). Software (SOF) fared well with little variability, while Laptops (LAP) showed marginal increases in delivery time. Overall, the company needs to ensure consistent processes across high-performing products such as Software and Laptops while mitigating the factors behind some of the delays by addressing the root-cause behind delays with other product lines. Continuous monitoring by means of SPC charts with process stability is being applied through the help of which customer satisfaction is being ensured and hence all products meeting the product delivery time requirements will be improved.

Part 4 Risk and Data Correction.

In part 4, we examine process monitoring accuracy via Type I and Type II error, rectify data discrepancies between product data from head-office and local datasets, and profit optimization using operational performance data (time-to-serve analysis).

4.1. Type I Error: False Alarm Probability

Type I Error represents the probability that an in-control process is incorrectly flagged as out of control. In other words this occurs when a process is incorrectly identified as unstable, but when it should have been correctly identified as stable.

This is known as an α-error. In statistical process control, $H0$ assumes that the process is in control and centered around a calculated center line. The type I error incorrectly rejects the null hypothesis ($H0$) when it is actually true.

This is called an α-error. In statistical process control H0 assumes the process to be in control and centered around the calculated center line, a type I error would be when the null hypothesis (H0) is rejected when it is actually true.

Variations in the type I error according to each detection rule:

- For SPC Rule A (any point beyond ±3σ), the probability is approximately α = 0.0027 (0.27%).
- For Rule B (most consecutive samples between ±1σ), $\alpha = (0.5)^{12} = 0.000244$
- For Rule C (four consecutive means beyond +2σ), α ≈ (0.0228)^4 = 2.7×10^{-7}.

Table 2: Rule Probabilities

Rule	Event	Probability
A	One sample > ±3σ	≈ 0.0027 (≈ 0.27 %)
B	Most consecutive samples within ±1σ	$= (0.5)^{12} = 0.000244$
C	Four consecutive means > +2σ	$(1 - 0.9772)^4 \approx (0.0228)^4 = 2.7 \times 10^{-7}$

Interpretation: The probability of a false alarm becomes lower with multi-sample rules (B and C), minimizing unnecessary process interventions compared to single-point detection (Rule A). This means that Rule B is much less susceptible to a spurious indication of an out-of-control process since 12 consecutive samples need to be within ±1σ. In the same way, Rule C applies a very demanding rule with four consecutive means exceeding +2σ, resulting in an extremely low Type I error probability.

4.2. Type II Error – Missed Detection Probability

This occurs when a process is incorrectly identified as stable when it was actually unstable, know as a type β-error. In statistical process control H_a assumes that the process is out of control and has shifted from the center line therefore the type II error occurs when the null hypothesis is not rejected when it is actually false.

Using the given limits (UCL = 25.089, LCL = 25.011) and the shifted process parameters (μ_1 = 25.028, $\sigma_{\bar{x}}$ = 0.017), the miss probability is:

$$\beta = \Phi(\left(\frac{UCL-\mu 1}{\sigma_{\bar{x}}}\right) - \Phi\left(\frac{LCL-\mu 1}{\sigma_{\bar{x}}}\right) \quad 13$$

$$= \Phi(\left(\frac{25.089-25.028}{0.017}\right) - \Phi\left(\frac{25.011-25.028}{0.017}\right)$$

$$= \Phi(3.588) - \Phi(-1.000) = 0.841$$

Therefore, **Type II error β ≈ 0.841** (84.1%) and **power ≈ 0.159** (15.9%).

This high β indicates the current chart is relatively insensitive to this small mean shift; increasing the sample size or tightening limits would improve detection.

Table 3: Type II values

Quantity	**Value**
z_L	-1.00
z_U	3.59
Φ_{zL}	0.1587
Φ_{zU}	0.9998
β (Type II)	**0.841**
Power (1−β)	**0.159**

4.3 Data correction and Alignment

In this section, I performed a data correction process to ensure consistency across the product data from the Head Office and local datasets. The key steps involved:

1. Correcting ProductID prefixes that were improperly labelled with "NA".
2. Aligning product categories between Head Office and local datasets.

After correcting these discrepancies, I reran the descriptive statistics and visualizations to understand the impact of the changes. The graphs were unaffected except for the following 2:

Total Sales by Product Category:

Before the correction, the Laptops category showed lower sales figures than expected, likely due to category misclassifications or inaccurate ProductID data. After the correction, Laptops became the dominant contributor to total sales, followed by Monitors and Cloud Subscriptions. These changes confirm that the initial misalignments in the product data were significantly affecting revenue reporting. With the corrections, the large extent to which laptops generate more sales than the rest of the products is much clearer.

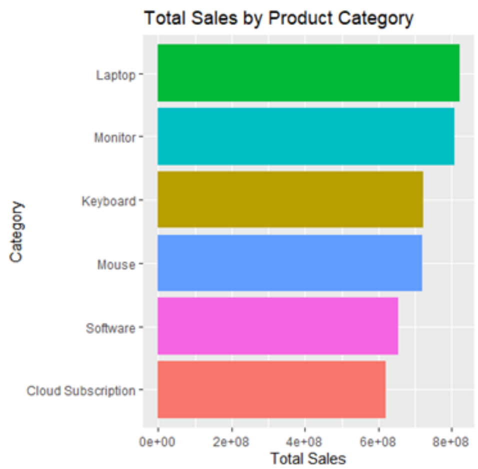


Figure 25: Top Sales by product category (original)

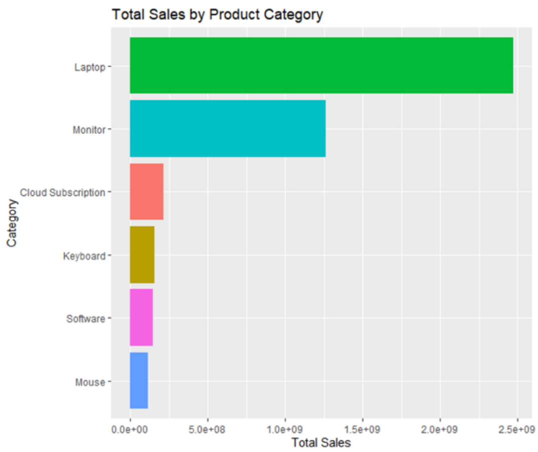


Figure 26: Top sales by product category (corrected)

Price Distribution by Category:

Before the correction, price distributions for some categories were inaccurate, with Laptops not showing as high as expected and categories like Software showing low values. After the correction, the Laptops category now shows a clear higher price range, which is consistent with market expectations. Other categories, like Software and Mice, show more reasonable price distributions. After the corrections, it is highlighted that laptops are priced much higher than the rest of the products, with monitors also having a reasonably high price. It is much clearer that the company's products consist of low priced products and 2 highly priced products.

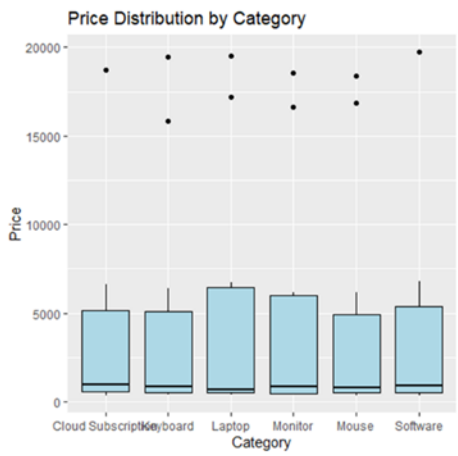


Figure 27: Price distribution by category (original)

Figure 28: Price distribution by category (corrected)

The corrected data now provides a more accurate representation of product sales and pricing across categories, which is critical for future analyses and decision-making.

Part 5: Optimising profit

5.1 Introduction

This portion aimed to optimise the number of baristas for two coffee shops—Shop 1 and Shop 2—to achieve optimal profit maximisation while providing service at a consistent level. Each number of baristas corresponded to specific service times, and we aimed to find the most profitable number of baristas per shop with an optimal amount of profits (operational cost is R1000 per barista/day, material profit is R30 per customer served).

5.2 Results and Discussion

The following outputs were generated after running the model for both shops:

- Shop 1: The optimal number of baristas is 6 and the Maximum profit per day for Shop 1 is R24,000.
- Shop 2: The optimal number of baristas is 6 and the maximum profit per day for Shop 2 is R20,000.

These results rely on the assumption that increasing the number of baristas improves both customer service speed (reliability) and daily profit until there is a point of diminishing returns whereby additional baristas do not generate substantial profit growth.

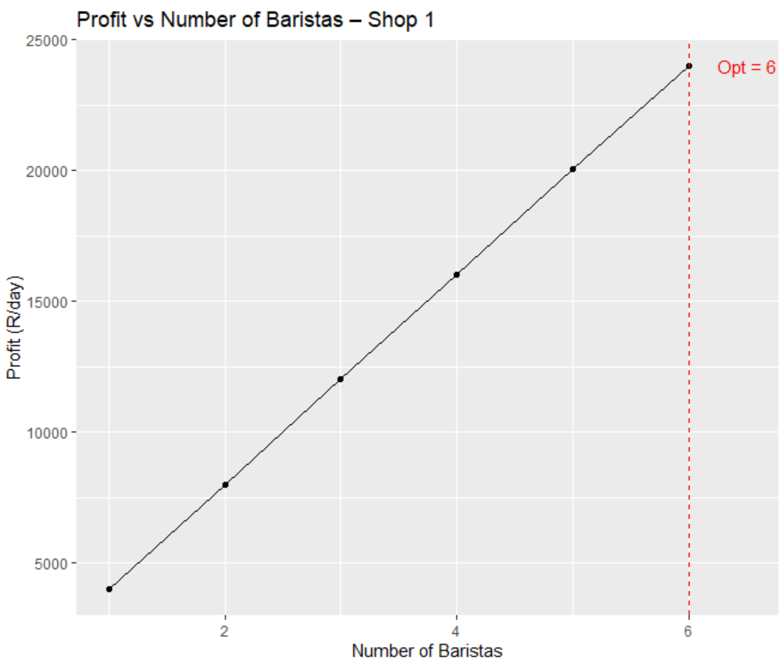


Figure 29: Profit vs number of baristas – shop 1

This graph displays the dependence of the number of baristas on profit per working day for Shop 1. Profit grows as the number of baristas increases until it peaks at 6 baristas, at which point the amount of profit generated from increased number of baristas is negligible. The red dashed line indicates the optimal number of baristas to profit maximization.

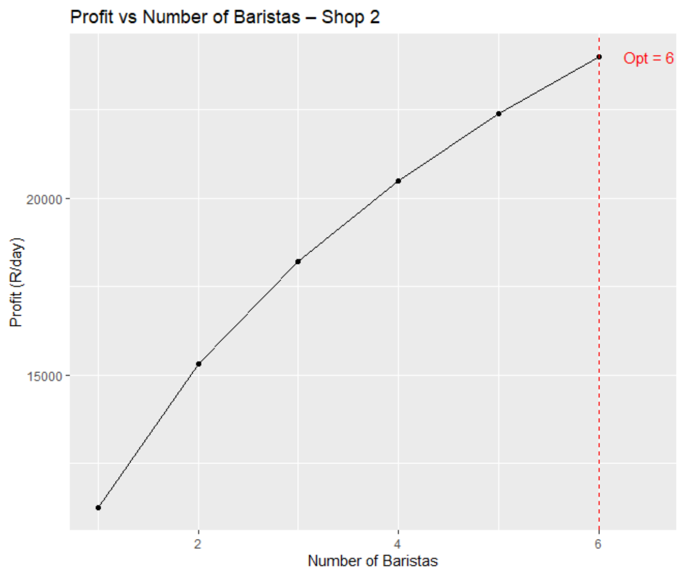


Figure 30: Profit vs number of baristas – shop 2

Similar to Shop 1, this plot shows the same relationship for Shop 2. Shop 2's optimal number of baristas is also 6, with the profit reaching its peak at this point. After this, adding more baristas does not provide a significant boost to profitability, which is consistent with the results seen in Shop 1.

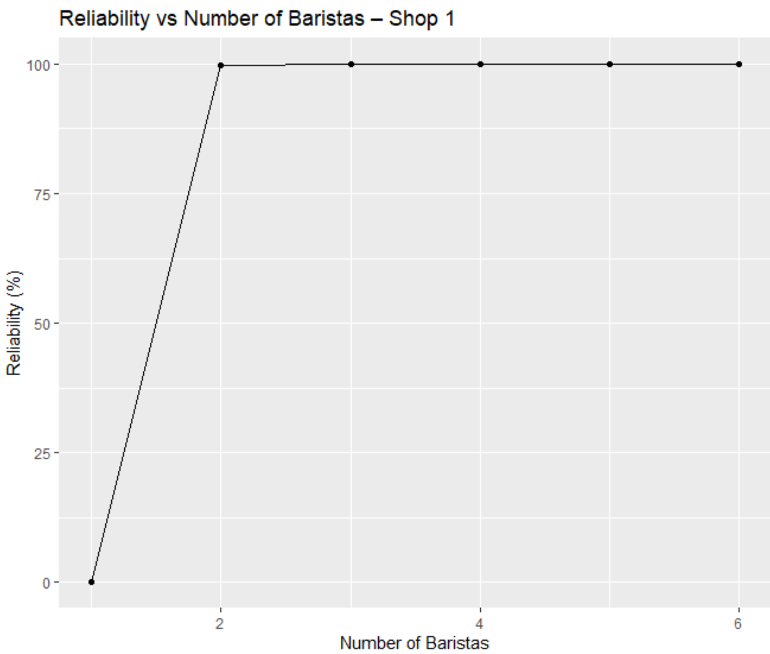


Figure 31: Reliability vs number of baristas – shop 1

This reliability plot shows what percentage of the customers are served in less than 120 seconds under various baristas. There is a clear increase in reliability with more and more baristas, with the highest reliability amounting to 100% of customers served with 2 or more baristas. This indicates that 2 baristas satisfy the performance metric of serving customers within 120 seconds.

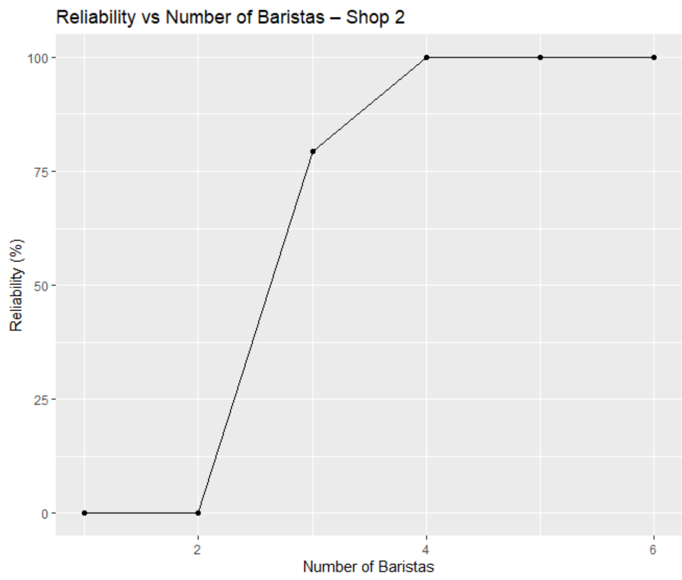


Figure 32: Reliability vs number of baristas – shop 2

For Shop 2, the reliability plot demonstrates a noticeable rise in reliability with increasing number of baristas, where 100% reliability results with the first 2 baristas. Indicating that with just 2 baristas, the service requirement of serving customers within 120 seconds will be satisfied completely. Above 2 baristas, reliability reaches a plateau of 100%, which means that more baristas do not improve the speed of service. This point of diminishing returns is where operational costs (i.e., the cost of adding more baristas) cease to bring in large improvements in reliability.

Shop 1 and 2:

For both shops, the optimal number of baristas is 6 for maximising daily profit. Shop 2 achieves 100% reliability with 2 baristas and no additional baristas increase reliability. And as to reliability, Shop 2 does reach 100% reliability with 2 baristas, and no additional number of baristas increase reliability. Profit vs Baristas: More baristas means more profit, but only to the point the optimal number is reached. This shop should optimize staffing schedules and other factors after that.

5.3 Recommendations for Next Steps

Optimisation: Because of the profit maximization from 6 baristas, scheduling the shop in such a capacity keeps this maximum during peak hours. This will help balance costs with service efficiency. Costs: The shop should also compare cost per customer values to the quality of the service so that staffing is optimised during less busy times. Such graphs offer us insights into Shop 2's operational dynamics; how the speed (reliability) of service can be optimized with fewer baristas. The results emphasize that staffing should match customer demand to guarantee cost-effective service.

For Shop 2, the reliability plot shows a sharp increase in reliability as the number of baristas increases, with 100% reliability achieved as soon as 2 baristas are present. This suggests that with just 2 baristas, the shop is fully able to meet the service requirement of serving customers within 120 seconds.

However, beyond 2 baristas, the reliability plateaus and remains at 100%, indicating that additional baristas do not improve service speed any further. This highlights a point of diminishing returns, where operational costs (the cost of adding more baristas) no longer yield any significant improvements in reliability.

5.4 Conclusion

This profit optimisation process showed that when the barista complement increased, the loss (based on reduced profit) decreased sharply at first and then was almost level where the ideal number of six baristas is reached. This exemplifies that excessive or insufficient staffing causes a deviation from the desired operating point—similar to the Taguchi loss, which asserts that loss will increase quadratically as a process output deviates from its target value, even within permissible limits. The “target” in this case is the ideal number of baristas for maximising profit. Thus the loss curve observed is similar to the Taguchi loss in shape — continuous and symmetric around the optimum — but the relationship here arises from discrete staffing levels and financial returns rather than ongoing dimensional tolerance. Both frameworks, however, stress that any deviation from the ideal operating condition (too few baristas or too many baristas) has measurable adverse effects on the organisation’s performance and profitability.

Part 6: ANOVA / MANOVA

6.1 Hypotheses Tested

In this part of the analysis, I tested the hypothesis related to service times using ANOVA. The following hypothesis was formulated:

• **Hypothesis : Delivery Time by Year (2022 vs 2023)**

- **Null Hypothesis (H_0):** There is no significant difference in delivery hours between 2022 and 2023.
- **Alternative Hypothesis (H_1):** There is a significant difference in delivery hours between 2022 and 2023.

6.2 Method

In order to test the Hypothesis, ANOVA was applied to determine whether there was a significant difference in delivery hours between the years 2022 and 2023. This hypothesis used data from the sales dataset, focusing on the delivery hours in the two years.

6.3 Results, Discussion, and Visualizations

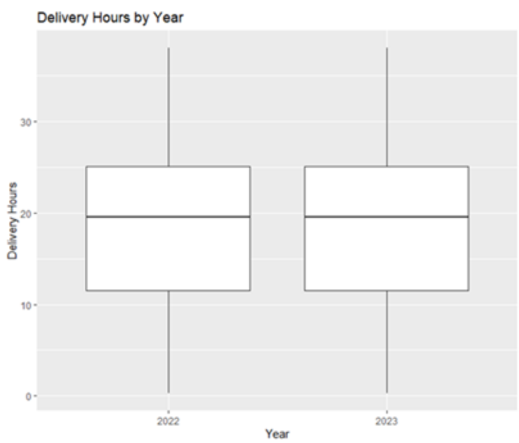


Figure 33: Delivery hours by year box plot - ANOVA

The ANOVA analysis for delivery time (by year) showed no significant difference between delivery hours in 2022 and 2023. This reveals that no changes to operations, if any, substantially impacted these delivery times during these years. It demonstrates the continuity of the two years. Both years have similar distributions of delivery hours, as seen in the boxplot. The boxplot corroborates that there is no major difference between 2022 and 2023 delivery times. For both years, median delivery hours are roughly similar, with few outliers, which suggests a similar level of service time. The visual representation also indicates that the delivery hours in both years are stable and the distributions are similar as expected too — a result that confirms the ANOVA.

6.4 Conclusion

From 2022 to 2023, the analysis of delivery times confirmed no significant difference across the two years, which means standard service levels are consistent. This consistency can assist with staffing and operational planning since the service time remains constant across the two years.

6.5 Suggestions for Future Action

- Staffing Optimization: Since delivery times have been stable throughout the years, staffing schedules should rely on different factors: peak periods or customer volume instead of year-to-year variations.
- Operational Efficiency: Service times should be monitored regularly; operational bottlenecks can be pinpointed and fixed on the spot as they arise.
- Performance Monitoring: Consider revisiting service processes if a future trend emerges showing significant shifts in delivery times.

These findings provide information on the sustainability of delivery times as well as indicate ways to maintain or improve delivery time performance.

Part 7: Reliability of Service and Optimization

7.1 Reliability of Service:

The analysis of staffing levels over 397 days at the car rental agency revealed the following:

- Number of Days with Reliable Service: Reliable service is defined as having at least 15 workers on duty. The number of days with reliable service is calculated as follows:
    - 96 days with 15 workers.
    - 270 days with 16 workers.

Thus, the total number of reliable service days is:
Total reliable days = 96 (15 workers) + 270 (16 workers) = 366 days
This implies that 92.2% of the days had reliable service, where the agency could meet the expected service standards.

The 92.2% reliability indicates a high level of operational efficiency, contributing significantly to customer satisfaction and the smooth operation of the business.

7.2 Optimizing Profit by Adjusting Personnel:

To maximize profitability, it is crucial to balance the cost of hiring additional workers with the sales loss incurred due to insufficient staffing. The following assumptions were made:

- Sales loss per day due to fewer than 15 workers: R20,000.
- Cost per additional worker: R833.33 per day (R25,000 per month).

The profit optimization analysis was done by calculating the total cost of hiring additional workers and comparing it with the sales loss due to understaffing.

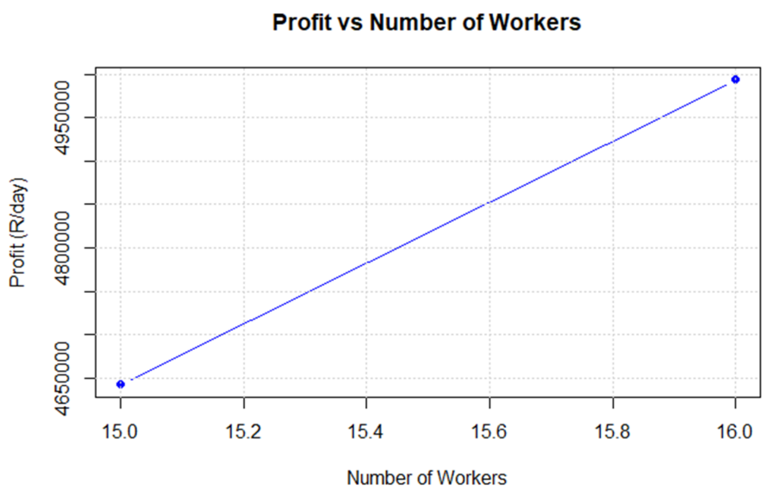


Figure 34: Profit vs number of workers

The results show that:

- Profit with 15 workers: R4,650,000 per day.
- Profit with 16 workers: R4,950,000 per day.

While the profit increases with additional workers, it becomes evident that the increase in profit is relatively small between 15 and 16 workers. This shows diminishing returns for adding more personnel. Given the results, we find that the optimal number of workers for maximizing profit is 16 workers.

7.3 Conclusion:

The analysis of staffing levels at the car rental agency revealed that:

- 92.2% reliability can be expected, indicating that reliable service is provided for most of the days with sufficient staffing.
- Profitability increases as more workers are hired, but the return is small after reaching 15 workers, meaning the optimal staffing level for profitability is 16 workers.

7.4 Suggestions for Further Action:

1. Staffing Optimization: Since 16 workers provide the highest profitability, it is recommended that staffing schedules be optimized to ensure 16 workers are present during peak periods. This can be achieved by forecasting customer demand and adjusting staffing accordingly.
2. Cost Analysis: Given the high cost of hiring additional personnel, a more detailed cost-benefit analysis can be conducted to find the most cost-effective staffing levels during non-peak hours to avoid overstaffing.
3. Operational Efficiency: To improve overall service efficiency, the company should focus on reducing sales loss during periods of understaffing. Investing in process optimization or better demand prediction tools can help ensure that staffing levels match customer demand without incurring excessive costs.

These findings highlight the importance of balancing staffing costs with service reliability to maximize profit, while ensuring that customer service standards are consistently met.

8. Conclusion

The report analyzes the company's operations from a range of angles, from sales to personnel deployment. Highlights of the analysis are that:

• Sales by Product Category: Revenue by product categories showed high differences, with laptops and monitors taking the lead. Based on the above analysis, companies should concentrate their marketing efforts in the category of highest performing product as opposed to the category of weak performing items.

• Customers' demographics: Age-group and city-level income disparities were found, including higher income users contributing significantly more to the revenue per year – the largest portion of which included higher income customers. This is a learning point which can be used to better market to high net worth customers.

• Service Time Analysis: hours of delivering and picking were constant from year to year, while a staffing optimization analysis found 6 baristas optimal to maximize profit, a balance between operation expense and service accuracy.

• Profit Optimized: The report also stated that added baristas raise service levels up to a certain point, but after 6 baristas, the costs of additional personnel are greater than the benefit. This result supports the case for scheduling employees on the basis of demand forecasts and its optimization.

Finally, this report will provide insight on increased operation productivity, staffing models, better customer experience. The results make a basis to base data driven decisions to increase profitability, service reliability and business performance.

9. References

Department of Industrial Engineering, Stellenbosch University. (2025). *QA344 Statistics.pdf: Statistical Process Control and Process Capability Notes.* Stellenbosch University.

Engineering Council of South Africa (ECSA). (2025). *Graduate Attribute 4 (GA4) Quality Assurance Report Brief.* QA 344 Project Guide, Stellenbosch University.

Taguchi, G. (1986). *Introduction to Quality Engineering: Designing Quality into Products and Processes.* Tokyo: Asian Productivity Organization.

StHDA. (n.d.). *MANOVA Test in R (Multivariate Analysis of Variance).* Available at: http://www.sthda.com/english/wiki/manova-test-in-r-multivariate-analysis-of-variance (Accessed 25 October 2025).

R Core Team. (2025). *R: A Language and Environment for Statistical Computing.* Vienna: R Foundation for Statistical Computing.