Quality Assurance ECSA GA4 Report

Daniela Nicole Marques Fernandes

Student No.:26911582

The Department of Industrial Engineering,

Stellenbosch University

24 October 2025

Abstract

This report contains a comprehensive data analysis report completed for the graduate attribute requirements (GA4) by the Engineering Council of South Africa (ECSA). The project entails a detailed data analysis of a company's product sales, statistical process control (SPC), process capability analysis, and optimization. The results of this report are found using the R programming language, using sales and product data. SPC charts, process capability indices, and type I and type II errors were used to assess the business's performance. Additionally, the report contains optimized models used for determining ideal staffing levels for maximizing profit and concludes with a multivariate analysis of variance to find any significant factors that may be affecting performance. The application of statistics and engineering principles has been used to solve real-world business problems, ensuring capable and optimized processes.

Table of Contents

List of Figures

List of Tables

Introduction

The main objective of this project is to demonstrate the use of statistical methods in problem-solving complex datasets, often found in real-world scenarios. It begins with a foundational data analysis of sales and product data using statistics to understand the underlying processes of the company. Statistical Process Control is applied to delivery times, establishing control limits and monitoring out-of-control conditions to increase stability, followed by a process capability study to determine whether customer requirements are met.

The subsequent sections involve type I and type II error calculations and data correction to ensure accurate distributions of the data. The project also discusses 2 separate models focusing on the optimization of staffing levels and their effect on profitability. A statistical analysis using MANOVA/ANOVA tests for process differences over time.

Part 1: Data Analysis & Descriptive Statistics

This section contains the data analysis of a company's sales, products, and customer data for the years 2022-2023. The primary objective of this analysis is to understand the business's performance, identify customer patterns that can improve the company's growth, and establish a basis for data-driven decision-making.

Key findings:

- The datasets are mostly complete, which provides a strong foundation for analysis.
- A large portion of revenue comes from a small number of high-value products, particularly in the laptop category.
- Although customer demographics are broad, high-value customers have been identified through an RFM analysis as prime targets for retention campaigns.
- There are several outliers in order fulfilment times, with some orders taking extremely long in the picking and delivery phases, indicating potential bottlenecks.

Method

The first step was to do an exploratory data analysis using the datasets provided, which contained sales, products, and customer data. The analysis was done as follows:

1. Data Loading and Joining: Merging all the data information to create a fully comprehensive sales dataset.
2. Data Quality Assessment: Checking for missing values and inconsistencies.
3. Descriptive Analysis: Summarising the tendencies and distributions of the data.
4. Visualisation: Creating graphs to visualise patterns, trends, and relationships.
5. Segmentation: Implementing RFM analysis to segment the customer base.

Data Overview

Sales Data: 100,000 transactions from 2022-2023

Product Data: 60 unique products across 6 categories (Cloud Subscription, Keyboard, Laptop, Monitor, Mouse, Software), with 10 products per category.

Customer Data: 5000 unique customer demographics instances

Completeness: No missing values were found in any of the data

Detailed Analysis

Sales and Revenue

Revenue by Product Category

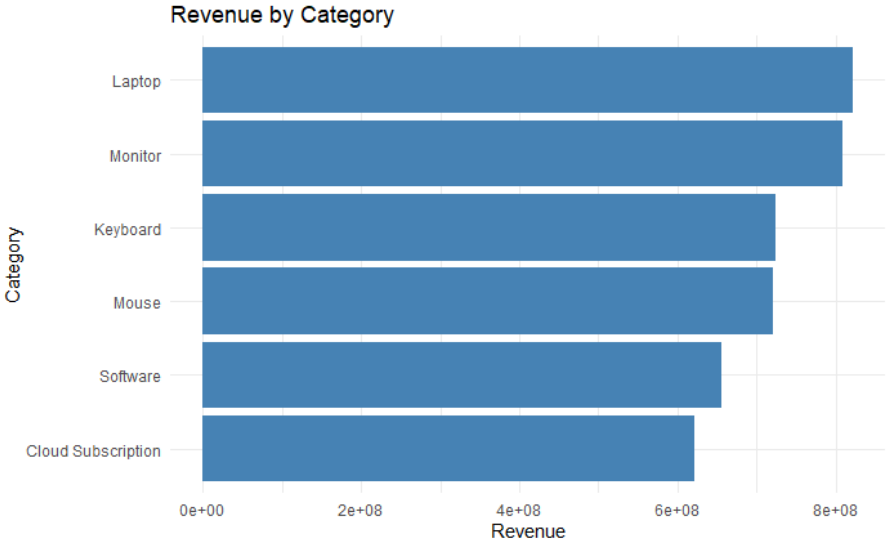


Figure 1 Revenue by Category

The revenue is not evenly distributed across the 6 categories. There are key market leaders, such as laptops and monitors, while other products, like software and cloud subscription, contribute less to the total revenue. This information is further supported by the Pareto analysis below, which shows the top 10 revenue-generating products, all of which are laptops.

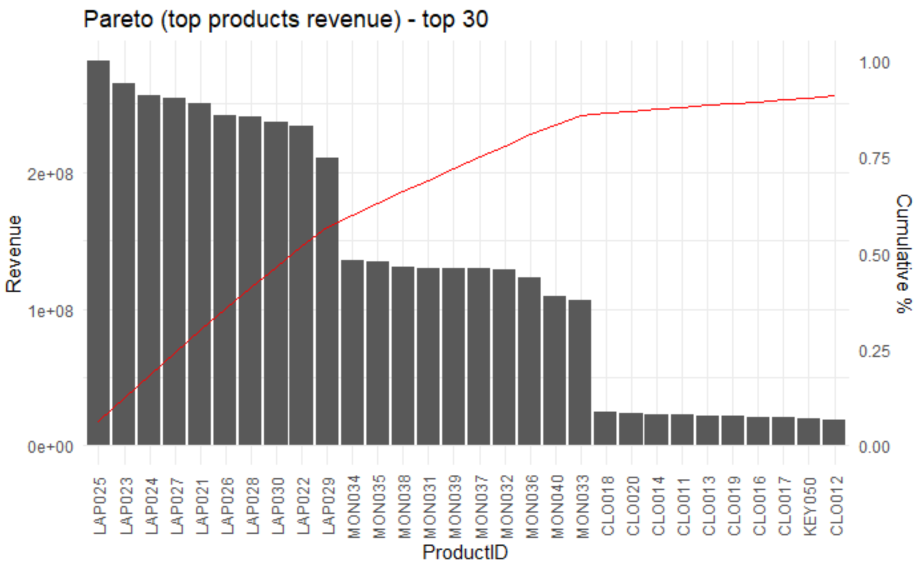


Figure 2 Pareto Analysis Plot (Top Revenue-Generating Products)

While these graphs indicate winning products, it means that a very small subset of the products is responsible for sustaining a large portion of the company's finances. The business should put a significant emphasis on protecting and maintaining a smooth supply chain, especially for laptops and monitors, as a small disruption in the supply chain could cause a huge impact on the business's revenue.

The analysis also reveals a need to promote categories such as software and cloud subscriptions to increase their sales and revenue. These categories are clearly not performing as well as others and can be used to diversify the company's revenue streams.

Top-Performing Product Descriptions

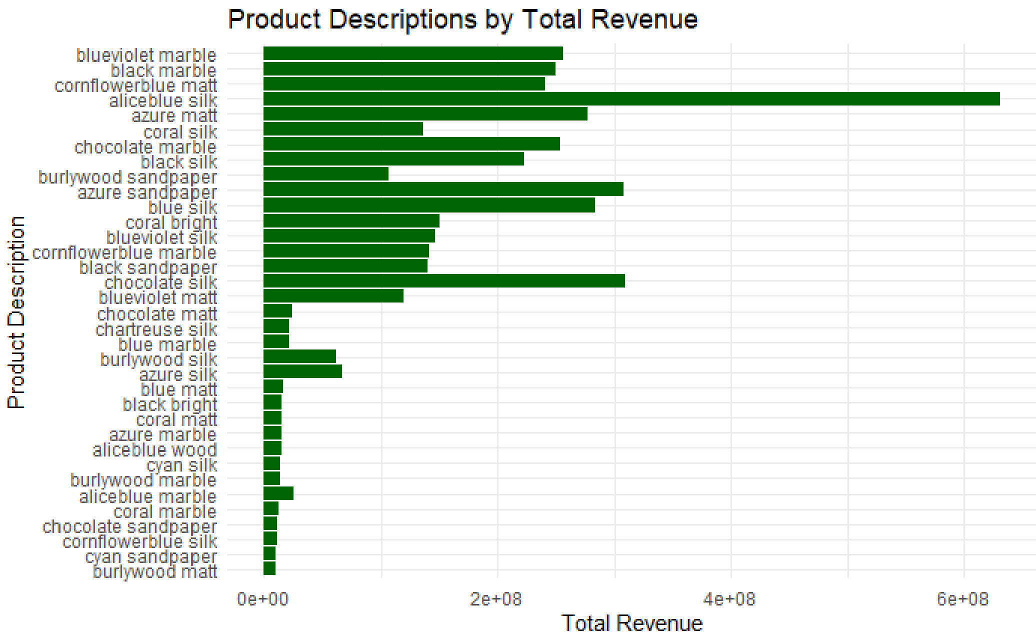


Figure 3 Product Descriptions by Total Revenue

Figure 3 displays product descriptions, indicating the ones that perform the best according to the revenue they contribute. There is a clear outlier, with aliceblue silk contributing a significantly larger amount than any other product descriptions. There are also a number of product descriptions that contribute very little to the overall revenue.

The chart distribution is uneven and displays a hyper-concentrated shape. The company is likely spending a disproportionate amount of time, money, warehouse space, and management on a large number of low-revenue products.

Instead of treating all products equally, the company should focus on a tiered strategy. This will prioritise and maximise high-performing products and grow middle-performing products. For low-performing products, an analysis should be done to determine whether it is reasonable to keep these items in stock or whether they should be removed, as they are incurring more costs than producing sales.

Sales Trends Over Time

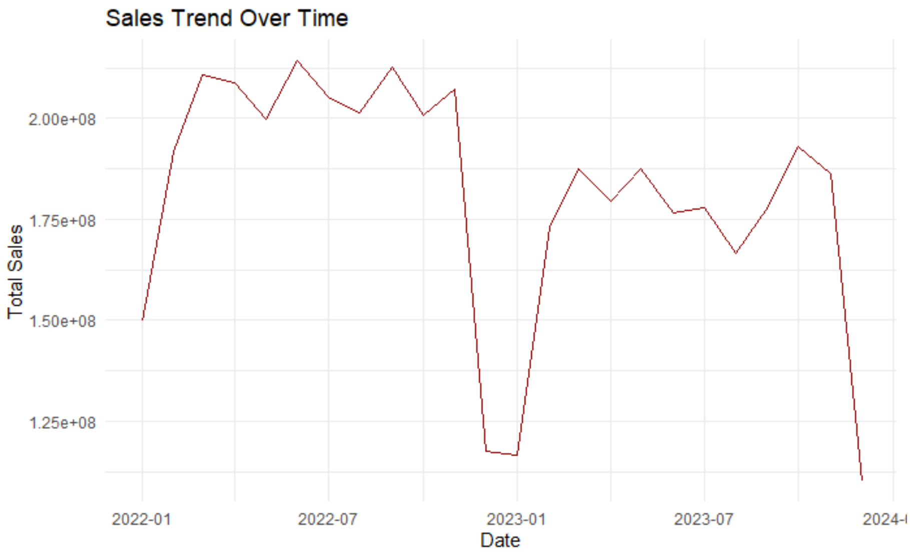


Figure 4 Sales Trend Over Time

The sales trend from the start of 2022 to the end of 2023 reveals a slight downward trajectory, with a huge decline in sales at the end of 2022. This may indicate that the business is either significantly impacted by seasonal fluctuations or that it closes down at the end of the year. The latter is, however, unlikely; according to their sales, the company is too large to close down during the end-of-year season. The business may therefore be facing challenges in dealing with overwhelming seasonal boosts.

The company must identify the root cause of large declines in sales at the end of both years to improve their overall performance and stabilise and increase the sales trajectory.

Transaction Profile

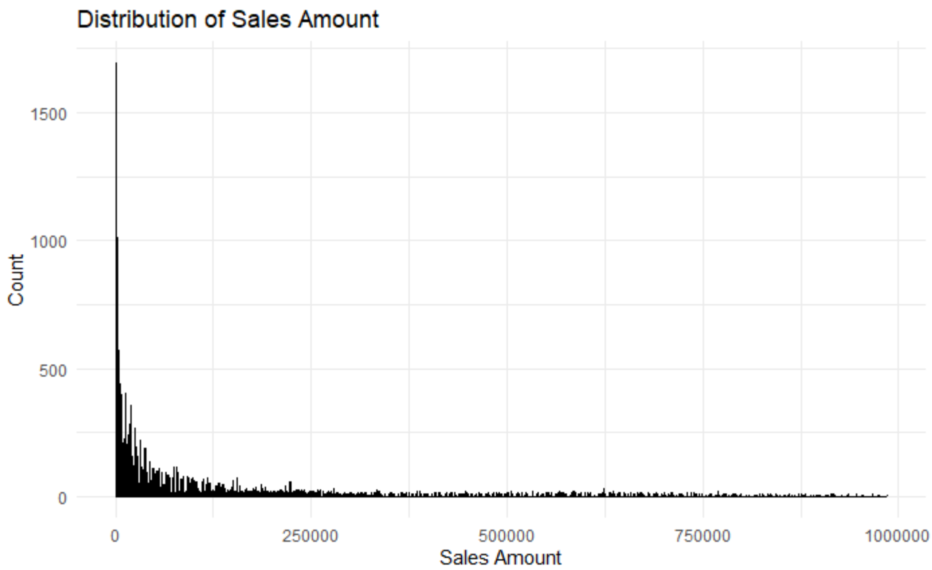


Figure 5 Distribution of Sales Amount

The distribution of sales provides important insight into the business's transaction profile. The graph is highly right-skewed, with peaks indicating that a large number of transactions are of low value, while the tail shows that a small number of transactions are of high value.

This means that revenue is stable in terms of frequent, low-value purchases, which contribute towards cash flow, as well as in less-frequent, but very high-value purchases, which drive total revenue.

Order Times

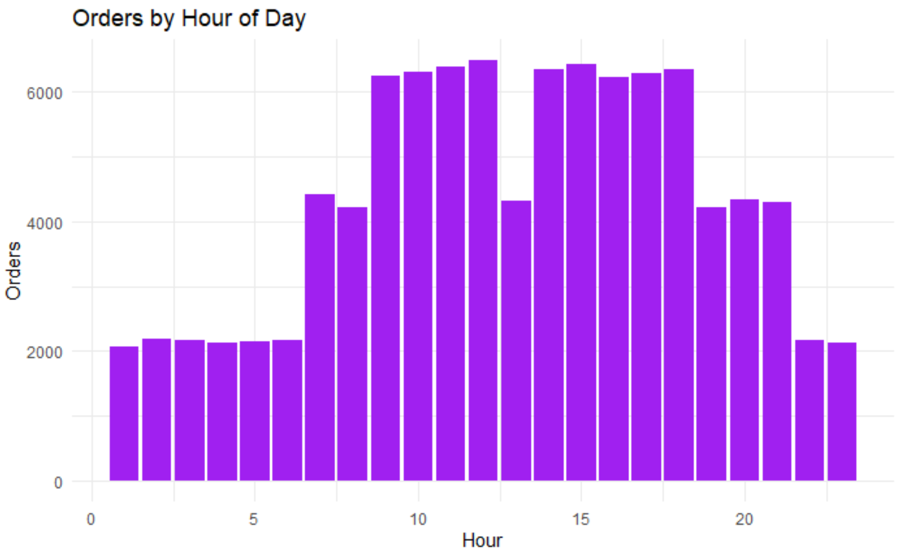


Figure 6 Orders by Hour of Day

The graph above shows orders by hour of the day. Orders peak around the hours of 7 am to 9 pm. This requires an increased need for staff during these times to maintain demand. These fluctuations can help the business plan staff numbers better, which would decrease costs and improve employee satisfaction.

Customer Analysis

Customer Demographics

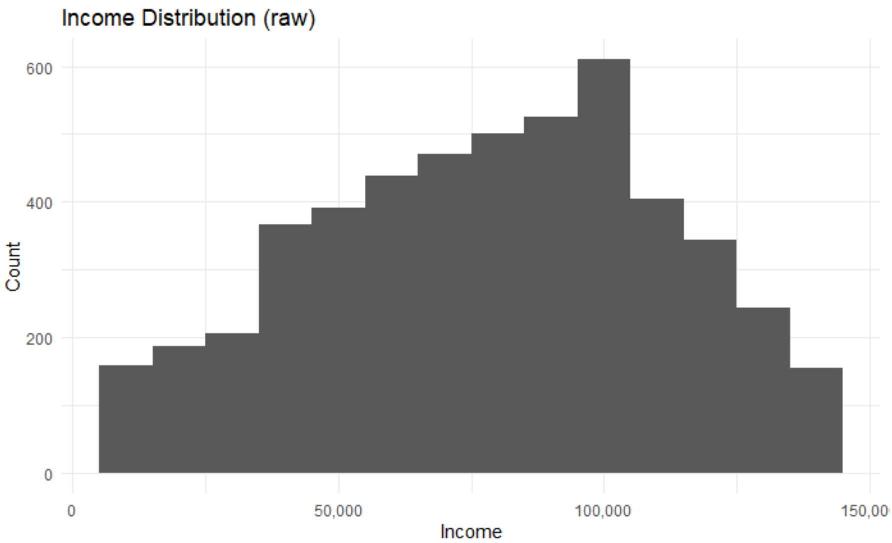


Figure 7 Income Distribution (raw)

The customer income graph above is slightly left-skewed, showing that a larger portion of customers have higher incomes. However, this view is misleadingly compressed. It is better represented on the log scale below.

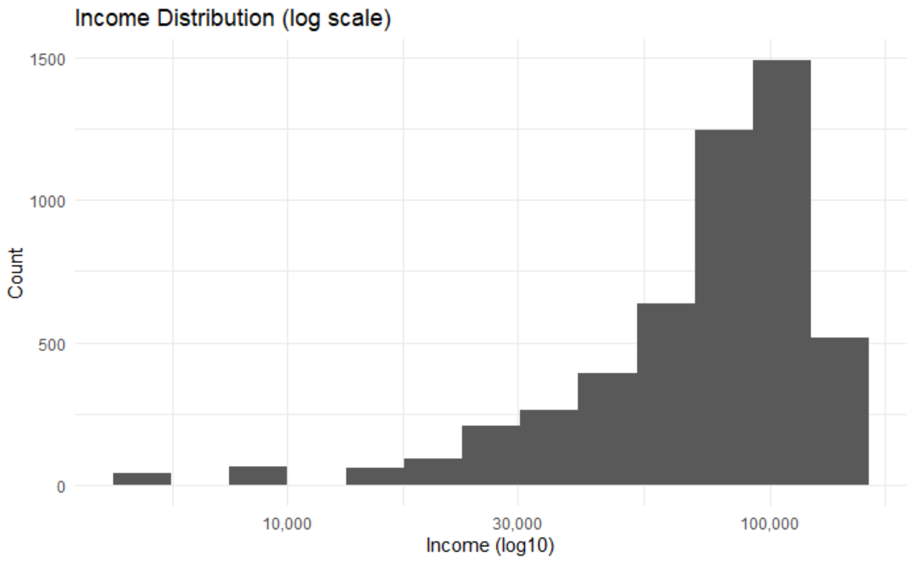


Figure 8 Income Distribution (log scale)

This graph better illustrates the income distribution. For the business, it validates why high-value products are the main source of income. Marketing teams should continue to strategize and advertise to this demographic, as they bring in high levels of revenue.

Average Income by City

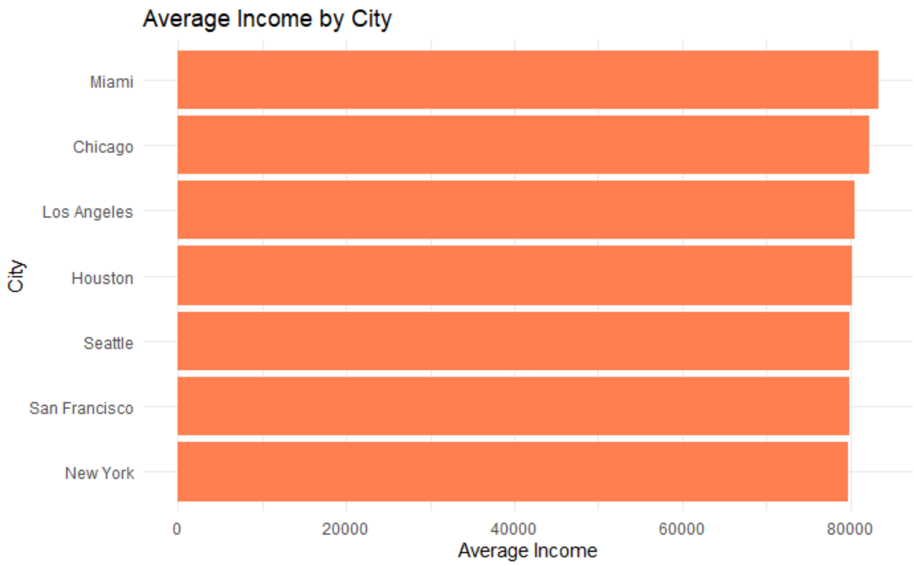


Figure 9 Average Income by City

There is only a slight difference in average income between cities, and so although that doesn't provide a lot of insight into a high-income market, it does show a consistently wealthy customer base. Hence, there is no need for the business to drastically change its strategy according to different cities.

Customer Value (RFM Analysis)

Using the RFM analysis (recency, frequency, and monetary), customers can be segmented based on their purchasing behaviour.

- Recency: How recently the customer made a purchase

- Frequency: How often customers make a purchase
- Monetary: How much customers tend to spend

The segments revealed through the RFM analysis in Figure 10 help build a clear strategic roadmap for customer management. The customers on the top-right are the most valuable customers, as they frequently make high-value purchases. Retention campaigns should be held for these customers.

Contrastingly, the top-left corner signifies customers who make high-value orders, but at a low frequency. This presents a revenue recovery opportunity to “win back” these customers to make high-value purchases again.

The bottom-left represents the majority of the customer base, which keeps cash flow up.

This data should help the company make tailored decisions according to its customer base, to improve sales and revenue.

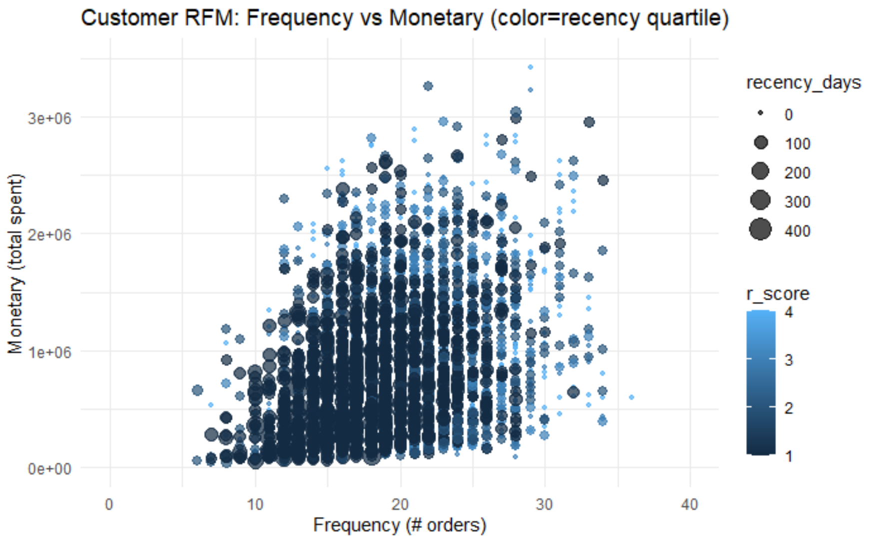


Figure 10 Customer RFM

Operational Efficiency

Root of Fulfilment Problems

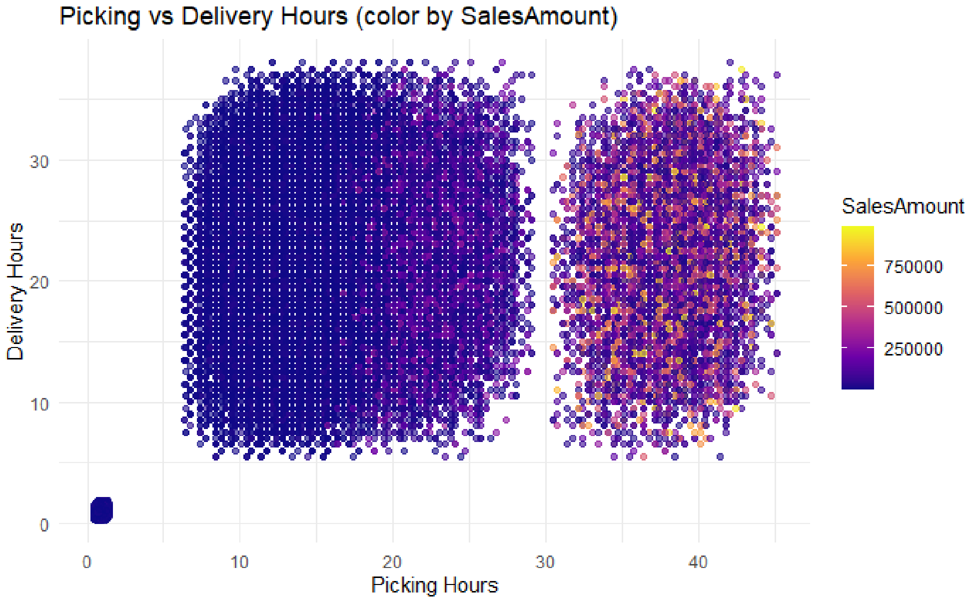


Figure 11 Picking vs Delivery Hours

The scatter plot above that compares picking vs delivery hours reveals a concerning operational inefficiency. It indicates that a large number of orders experience slow picking and delivery times, which are creating a long delay for customers. According to the colour coding, most of the high-value transactions are not being prioritised. This is a huge issue as bottlenecks within the system are the direct cause of losses. Addressing these delays is vital to improve the overall order speed and customer satisfaction. The large gap found around the 32-hour mark is likely due to the maximum allowed delivery time that is being enforced, which further demonstrates that many orders are only being delivered after the promised threshold.

This is further supported by the picking and delivery hour distribution graphs below. They are both concentrated around acceptable time frames, but both have long tails, indicating that a significant number of orders are processed very slowly.

A specific query was run to identify orders with picking times exceeding 48 hours or deliveries exceeding 72 hours. The *slow_orders* data frame in Table 1 contains these problematic cases. These instances require immediate investigation into the root cause of the excess time, which could be due to warehouse inefficiencies, supplier issues, or problems with delivery partners.

CustomerID	ProductID	Quantity	order_date	pickingHours	deliveryHours	total_process_hours	SalesAmount
CUST4759	LAP025	7	44896	45.055	37.044	82.099	138076.26
CUST3721	LAP023	49	45231	42.72417	37.546	80.27017	953183.28
CUST3737	LAP026	8	44866	43.72167	36.544	80.26567	149693.76
CUST2709	LAP027	4	45261	44.0575	36.046	80.1035	68809.12
CUST2694	LAP022	9	44896	43.055	37.044	80.099	149797.89
CUST3519	LAP029	4	44896	42.055	38.044	80.099	63406.96
CUST158	LAP025	2	45231	42.72417	36.546	79.27017	39450.36
CUST3367	LAP028	13	45261	44.0575	35.046	79.1035	241205.64
CUST3367	LAP025	10	44896	43.055	36.044	79.099	197251.8
CUST2033	LAP023	4	44805	42.055	36.544	78.599	77810.88

Table 1 Processing Hours

The graph below reveals the direct link between internal and external delays. Orders with longer picking times also suffer from long delivery times. Bottlenecks in the warehouse are creating a domino effect by stalling orders before they reach the delivery partner. The business must therefore focus on fixing the internal picking process to resolve the overall order fulfilment issue.

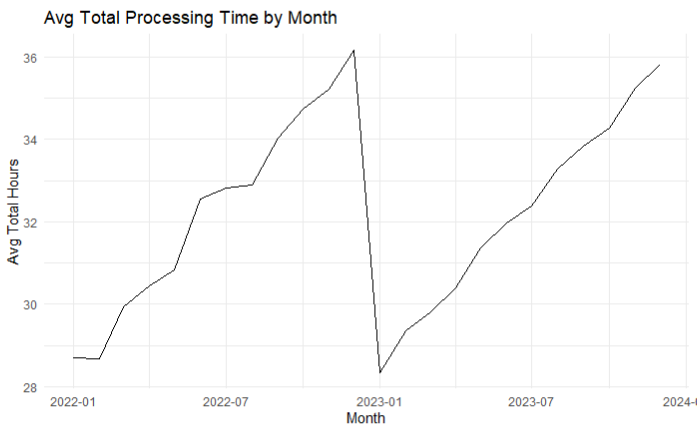


Figure 12 Total Average Processing Time by Month

Upon further investigation of the data, one would expect the ratio of laptops and monitors to be slightly higher compared to the other 4 products in figure 1. This may be an issue with the data provided and should be investigated.

Recommendations

The company should prioritise an inventory strategy for top products. They should classify the top 10 revenue-generating products and apply a tiered inventory management process. Increase safety stock for these items and lower lead times. This is critical as a large portion of the company's income comes from a small subset of its high-value products. Ensure little to no disruptions in the supply chain, as any interferences would severely impact the overall revenue.

Develop marketing campaigns to increase the sales of lower-performing products such as software and cloud subscriptions. This could be done by selling the products in bundles, along with well-performing products like laptops. This would build a more diverse revenue profile and reduce vulnerability if the hardware market shifts.

The root cause for the major sales decline at the end of 2022 should be investigated, and a strategic plan must be developed to mitigate the effect in future years. This will stabilise the company's revenue and ensure consistent performance throughout the year, counteracting seasonal sales declines.

Launching a marketing campaign that is tailored to the customer-specific segments identified in the RFM analysis will ensure that marketing resources are invested in the areas that will generate the highest return.

Warehouse fulfilment bottlenecks need to be addressed immediately by diagnosing and resolving the root cause of slow picking times. Fixing this bottleneck will reduce delivery times because of the domino effect taking place from picking to delivery. These operational inefficiencies must be fixed urgently as they are causing poor customer satisfaction and sales loss.

Part 3: Statistical Process Control

The objective of this section is to perform a statistical process control on the future sales data provided for the years 2026 and 2027.

3.1. The sales datasets for 2026 and 2027 were ordered chronologically by year, month, date, and time. Each sample consisted of 24 deliveries, with the first 30 samples being used to set the control limits. The control limits were calculated using constants for n=24 as follows:

- A3 = 0.619
- B3 = 0.555
- B4 = 1.445

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

The control limits calculated for each product type were as follows:

X-bar Chart:

- Centre Line: $\bar{\bar{x}} = \frac{\sum_{i=1}^{k} \bar{x}_i}{k}$
- Upper Control Limit: $UCL_{\bar{x}} = \bar{x} + A_3 \cdot \bar{s}$
- Lower Control Limit: $LCL_{\bar{x}} = \bar{x} - A_3 \cdot \bar{s}$
- Upper σ_1 limits: $U1L_{\bar{x}} = \bar{x} + \frac{1}{3}(UCL_{\bar{x}} - \bar{x})$
- Lower σ_1 limits: $L1L_{\bar{x}} = \bar{x} - \frac{1}{3}(UCL_{\bar{x}} - \bar{x})$
- Upper σ_2 limits: $U2L_{\bar{x}} = \bar{x} + \frac{2}{3}(UCL_{\bar{x}} - \bar{x})$
- Lower σ_2 limits: $L2L_{\bar{x}} = \bar{x} - \frac{2}{3}(UCL_{\bar{x}} - \bar{x})$

s-Chart:

- Centre Line: $\bar{s} = \frac{\sum_{i=1}^{k} s_i}{k}$
- Upper Control Limit: $UCL_s = B_4 \cdot \bar{s}$
- Lower Control Limit: $LCL_s = B_3 \cdot \bar{s}$

- Upper σ_1 limits: $U1L_s = \bar{s} + \frac{1}{3}(UCL_s - \bar{s})$
- Lower σ_1 limits: $L1L_s = \bar{s} - \frac{1}{3}(UCL_s - \bar{s})$
- Upper σ_2 limits: $U2L_s = \bar{s} + \frac{2}{3}(UCL_s - \bar{s})$
- Lower σ_2 limits: $L2L_s = \bar{s} - \frac{2}{3}(UCL_s - \bar{s})$

Analysis of X-Bar and s-Charts

The first 30 samples were used to set the centre lines, upper control limits, and lower control limits. Each s-Chart displays the standard deviation of delivery times per product type for all samples, and each X-bar chart displays the mean delivery time per product type for all samples.

Product Type: Mouse

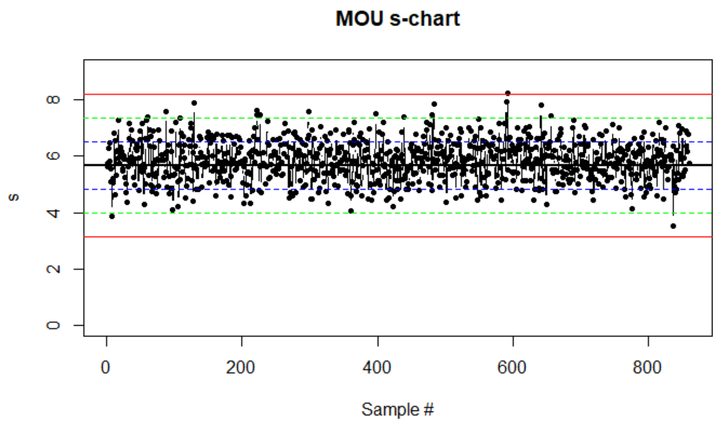


Figure 13 Mouse s-Chart

- Centre Lines: 5.9
- Upper Control Limit: 8.2
- Lower Control Limit: 3.4

The s-Chart above displays the spread of Mouse product sample delivery times. The dataset remains mostly within bounds, except for 1 sample around the 600-sample mark, which reaches outside of the upper 3σ limit. This may have been a cause for investigation, which would be the reason why the samples performed a lot better afterwards. Several other samples come close to the upper and lower 3σ limit.

Those near or past the upper control limit indicate that the process has special cause variation, suggesting that a specific condition may be contributing to the high variability in delivery times. Those near the lower control limit also indicate special cause variation; however, it displays samples that have lower delivery times than usual, which is a positive result.

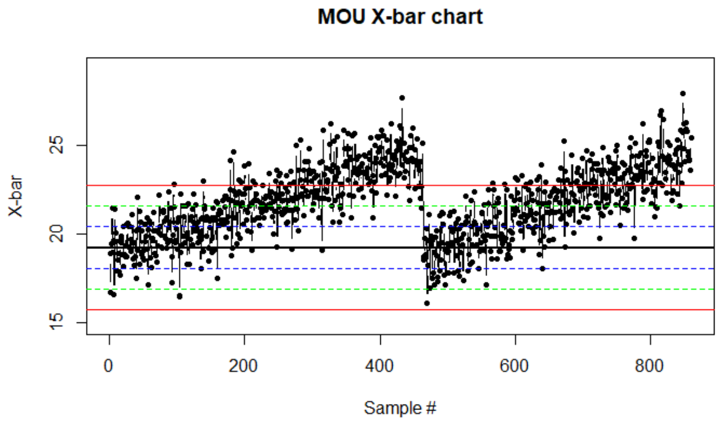


Figure 14 Mouse X-bar Chart

- Centre Lines: 19.2
- Upper Control Limit: 20.7
- Lower Control Limit: 15.7

The X-bar chart above reveals the mean delivery times per sample, and it is clear that the process is statistically unstable. Delivery times remain within bounds for about the first quarter of each year, but then begin to increase.

The X-bar chart gives reason to believe that managers may not be monitoring delivery processes, which counteracts the idea that the sample discussed above exceeded the upper control limit. The sample may have been a statistical outlier, as all the other sample standard deviations remained within bounds.

Between the years of 2026 and 2027, delivery times are rectified; however, over time, the company falls back into patterns of slow delivery times. There is a clear issue with the delivery process, as the company is unable to maintain its delivery times within bounds throughout the year. This needs to be addressed to improve the company's delivery issues.

Product Type: Keyboards

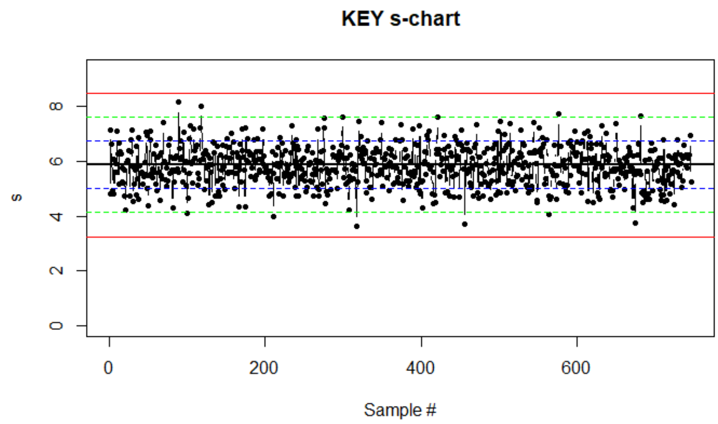


Figure 15 Keyboard s-Chart

- Centre Lines: 5.9
- Upper Control Limit: 8.5

- Lower Control Limit: 3.3

The s-Chart above displays the spread of Keyboard product sample delivery times. The dataset remains within bounds, with only a few samples nearing the upper and lower 3σ control limits.

Those near the upper control limit suggest that a specific condition may be contributing to the high variability in delivery times, while those near the lower 3σ control limit also indicate special cause variation for lower delivery times.

The spread is mostly even, and there are no real issues with the standard deviation in delivery times for the keyboard data.

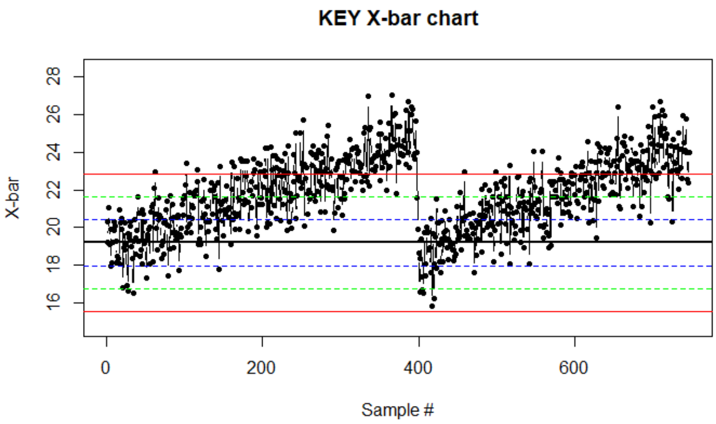


Figure 16 Keyboard X-bar Chart

- Centre Lines: 19.2
- Upper Control Limit: 22.8
- Lower Control Limit: 15.5

The X-bar chart above reveals the mean delivery times per sample, and it is clear that the process is statistically unstable. Delivery times remain within bounds for about the first 6 months of 2026 and first 7 months of 2027 (slight improvement from the previous year), but then they begin to increase.

It is clear that between the years of 2026 and 2027, delivery times are rectified; however, over time, the company falls back into patterns of slow delivery times. There is a clear issue with the delivery process, as the company is unable to maintain its delivery times within bounds. This needs to be addressed to improve the company's delivery issues.

Product Type: Software

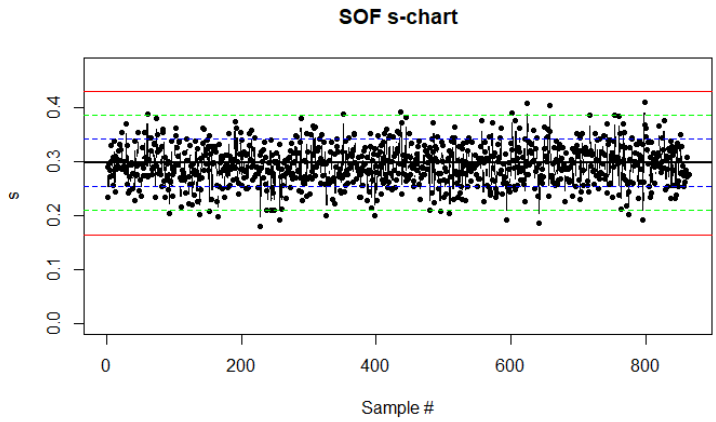


Figure 17 Software s-Chart

- Centre Lines: 0.3
- Upper Control Limit: 0.43
- Lower Control Limit: 0.16

The s-Chart above displays the spread of Software product sample delivery times. The dataset remains within bounds, with only a few samples nearing the upper and lower 3σ control limits.

Those near the upper control limit suggest that a specific condition may be contributing to the high variability in delivery times, while those near the lower 3σ control limit also indicate special cause variation for lower delivery times.

The spread is mostly even, and there are no real issues with the standard deviation in delivery times for the software data.

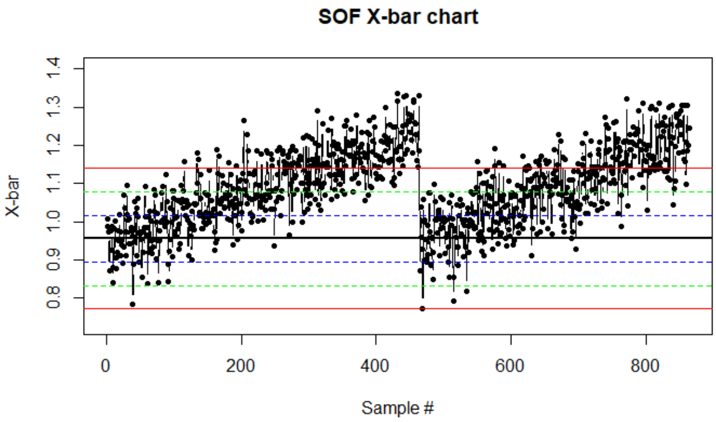


Figure 18 Software X-bar Chart

- Centre Lines: 0.3
- Upper Control Limit: 0.43
- Lower Control Limit: 0.16

The X-bar chart above reveals the mean delivery times per sample, and it is clear that the process is statistically unstable. Delivery times remain within bounds for about 6-7 months of each year, with only a slight improvement from 2026 to 2027, but then begin to increase.

It is clear that between the years of 2026 and 2027, delivery times are rectified; however, over time, the company falls back into patterns of slow delivery times. Although there are a few samples at the start of each year that fall at the lower control limit, indicating lower delivery times, these may just be outliers, as most of the other samples do not meet the required delivery times. There is a clear issue with the delivery process, as the company is unable to maintain its delivery times within bounds. This needs to be addressed to improve the company's delivery issues.

Product Type: Cloud Subscription

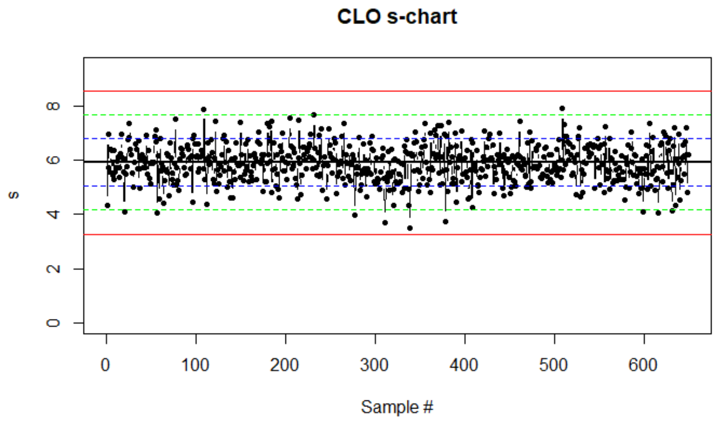


Figure 19 Cloud Subscription s-Chart

- Centre Lines: 5.95
- Upper Control Limit: 8.6
- Lower Control Limit: 3.25

The s-Chart above displays the spread of Cloud Subscription product sample delivery times. The dataset remains within bounds, with some samples dipping near the lower control limits between samples 270-340 and 560-630. This displays a decrease in delivery time variability, which shows that they are becoming more predictable and stable. Samples near the lower control limit indicate special cause variation for the lower delivery times.

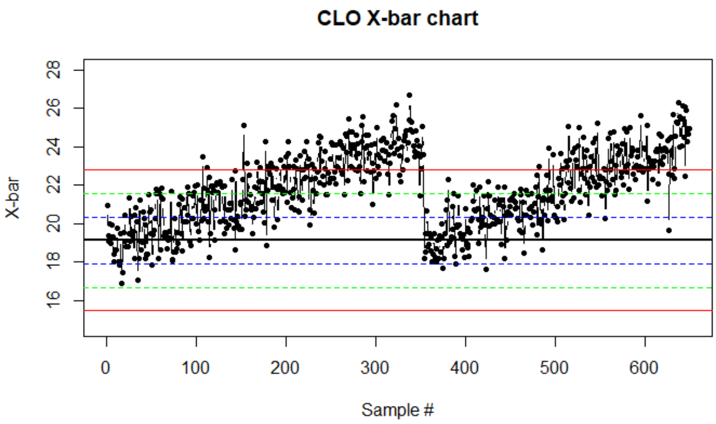


Figure 20 Cloud Subscription X-bar Chart

- Centre Lines: 19.25
- Upper Control Limit: 22.8
- Lower Control Limit: 15.5

The X-bar chart above reveals the mean delivery times per sample, and it is clear that the process is statistically unstable. Delivery times remain within bounds for about the first half of each year, but, then begin to increase. This will affect revenue and lead to customer dissatisfaction.

It is clear that between the years of 2026 and 2027, delivery times are rectified; however, over time, the company falls back into patterns of slow delivery times. There are no samples that reach below the 2σ control limit, hence most samples have a much higher mean delivery time. This means that, on average, most deliveries take too long, which is a concern. There is a clear issue with the delivery process, as the company is unable to maintain its delivery times within bounds. This needs to be addressed to improve the company's delivery issues.

Product Type: Laptop

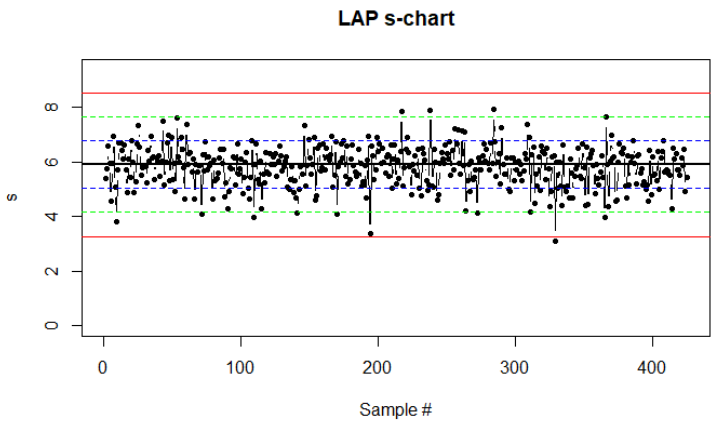


Figure 21 Laptop s-Chart

- Centre Lines: 5.9
- Upper Control Limit: 8.5
- Lower Control Limit: 3.2

The s-Chart above displays the spread of Laptop product sample delivery times. The dataset remains mostly within bounds, except for a few samples which reach the lower 3σ control limit.

Those near the lower 3σ control limits indicate lower variability in delivery times; hence, they are more stable and predictable.

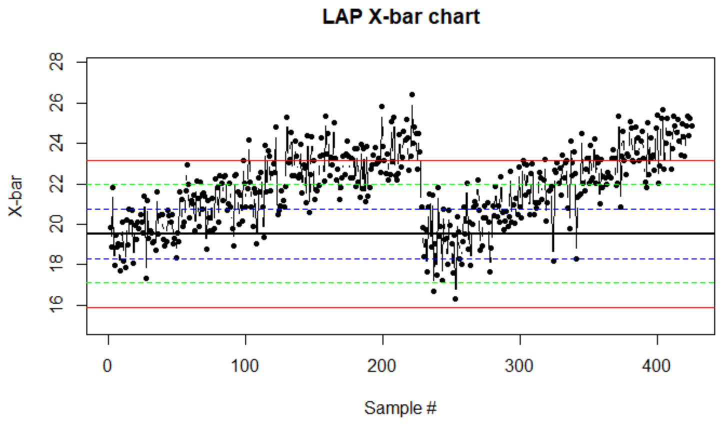


Figure 22 Laptop X-bar Chart

- Centre Lines: 19.6
- Upper Control Limit: 23.1
- Lower Control Limit: 15.95

The X-bar chart above reveals the mean delivery times per sample, and it is clear that the process is statistically unstable. Delivery times remain within bounds for about the first two-thirds of each year, which is much better than other products, but then they begin to increase.

With laptops being the company's major source of income, it is good to know that delivery times remain within bounds for over half of the year; however, there is still a need for major improvements.

It is clear that between the years of 2026 and 2027, delivery times are rectified; however, over time, the company falls back into patterns of slow delivery times. There is a clear issue with the delivery process, as the company is unable to maintain its delivery times within bounds. This needs to be addressed to improve the company's delivery issues.

Product Type: Monitors

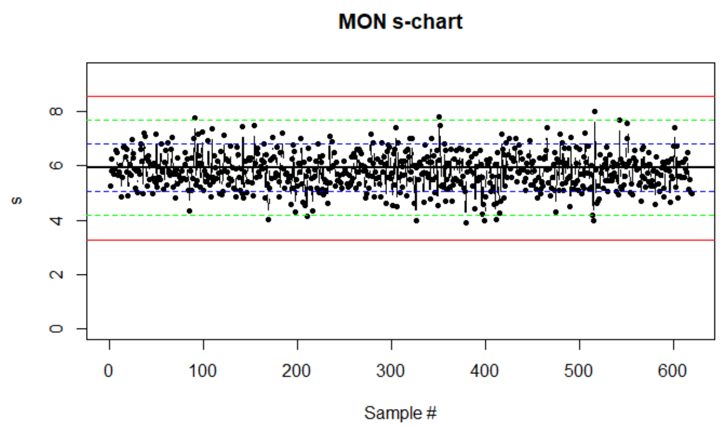


Figure 23 Monitors s-Chart

- Centre Lines: 5.9
- Upper Control Limit: 8.55
- Lower Control Limit: 3.2

The s-Chart above displays the spread of Monitor product sample delivery times. The dataset remains well within bounds, with few samples reaching over the 2σ control limit. This is an indication of a stable and predictable delivery time spread.

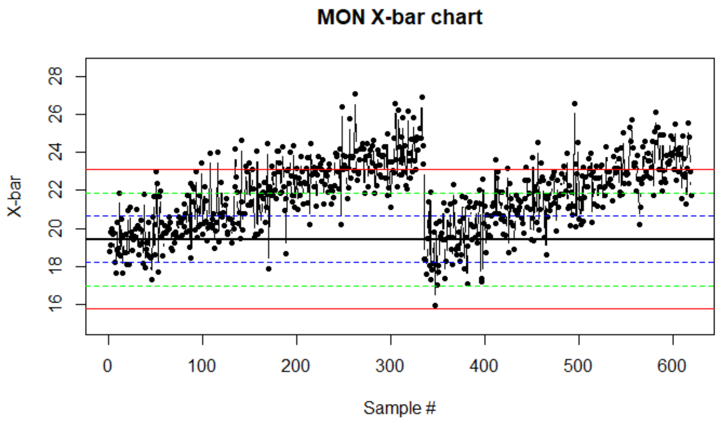


Figure 24 Monitors X-bar Chart

- Centre Lines: 19.4
- Upper Control Limit: 23
- Lower Control Limit: 15.8

The X-bar chart above reveals the mean delivery times per sample, and it is clear that the process is statistically unstable. Delivery times remain within bounds for about the first half of each year, which is slightly better compared to other products, but then they begin to increase.

Although monitors have one of the more stable delivery time means, it is clear that between the years of 2026 and 2027, delivery times are rectified; however, over time, the company falls back into patterns of slow delivery times. There is a clear issue with the delivery process, as the company is unable to maintain its delivery times within bounds. This needs to be addressed to improve the company's delivery issues.

Summary of Results

After conducting a comprehensive statistical process control analysis, the performance of the company's 6 product types was revealed. All products maintained a stable spread/ standard deviation of delivery times with little variation, however, in regard to mean delivery time, they all exceeded the acceptable bounds over time.

This is extremely concerning for the company's revenue and customer satisfaction. Towards the end of the year, they fail to meet delivery time requirements. Management may be monitoring the delivery process at the start of each year, but is not observing the process year-round, which may be the cause of increasing delivery times. It could also be linked to seasonality, which was discussed in the data analysis portion of this report.

It is crucial that these issues are addressed to avoid any further implications to revenue and customer satisfaction, focusing especially on seasonal changes and how those affect the company's delivery times. The company should study samples with low mean delivery times and low variability and model their systems to have the same conditions as these samples. This could help them in the process of improving their delivery process.

Taguchi Loss

The observed process loss, where delivery times increasingly deviate from the target, correlates with the Taguchi Loss Function concept. In both cases, any deviation from the ideal state, even within specification limits, results in a gradual loss, such as customer dissatisfaction or hidden costs. However, unlike the classic Taguchi loss, the loss in this delivery process is asymmetric, as delays cause far larger issues than early deliveries provide benefit.

3.2. The sampling continued beyond the 30 samples for all product types. They were numbered sequentially (30, 31, 32, ...) for ongoing process monitoring.

Total samples available per product type:

Product Type	Deliveries	Samples
Mouse	20662	861
Keyboard	17920	747
Software	20749	865
Cloud Subscription	15598	650
Laptop	10207	425
Monitors	14864	619

Table 2 Total Samples Available per Product

3.3. Specification Limits:

- Lower specification limit (LSL) = 0 hours
- Upper specification limit (USL) = 32 hours

Capabilities based on the first 1000 deliveries:

Product Type	Cp	Cpu	Cpl	Cpk	Capable (Cpk ≥ 1)
Mouse	0.915	0.727	1.104	0.727	No
Keyboard	0.917	0.729	1.105	0.729	No
Software	18.135	35.188	1.083	1.083	Yes
Cloud Subscription	0.898	0.717	1.079	0.717	No
Laptop	0.899	0.696	1.101	0.696	No
Monitors	0.889	0.700	1.079	0.700	No

Table 3 Product Capabilities

Although the X-bar and s-Chart analysis show that no products meet delivery time capabilities, the upper limit for the above calculation has been set to 32 hours, and because software delivery times fall far below the 32-hour upper specification limit, they meet customer requirements. The remaining product types demonstrate insufficient capability.

The primary issue is that most products in the Cp and Cpu values are too low, indicating that the process mean is too close to the upper specification limit.

Product managers should actively be involved in the processes that are statistically out of control, especially those exceeding the 3σ control limit, when non-random patterns emerge, or when process capabilities such as Cpk fall below 1.0, all indicating an inability to meet customer requirements.

For the s-charts, immediate intervention is needed when points exceed the upper control limit, signalling an increase in delivery time and variability, like the data found in laptops.

For X-bar charts, intervention is needed to respond to points outside of the control limits, shifts in the mean, or patterns suggesting a specific cause for variation. These processes should be reviewed during periods of change, and in regular intervals, even during stable periods. The focus should be on end of year periods where the company begins to fail.

3.4.

	s Samples found above +3σ		Most consecutive s samples between -1σ and +1σ	4 consecutive X-bar samples above +2σ		
Product Type	**Rule A**	**Sample**	**Rule B**	**Rule C**	**First 3 Samples**	**Last 3 Samples**
MOU	1	592	16	324	194, 195, 196	858, 859, 860
KEY	0	N/A	15	294	112, 113, 114	744, 745, 746
SOF	0	N/A	21	334	202, 203, 204	862, 863, 864
CLO	0	N/A	35	263	122, 123, 124	647, 648, 649
LAP	0	N/A	19	159	119, 120, 121	423, 424, 425
MON	0	N/A	34	226	134, 135, 136	616, 617, 618

Table 4 Samples found for Rules A, B, and C

According to Rule A, only the mouse product showed a process control issue, with one sample (sample 592) falling outside the upper +3σ control limit. All other product types were within control limits, indicating stable variability.

For Rule B, which identifies the most consecutive samples within the ±1σ limits (a sign of good process control), the results show that cloud software and monitors demonstrated the most consistent performance with 35 and 34 consecutive samples, respectively. Software followed with 21, mouse with 16, keyboard with 15, and laptop with 19 consecutive samples within this range.

For Rule C, sequences of more than 4 consecutive $\bar{X}$ samples outside the upper second control limits were found for all product types. The first three and last three identified sequences for each are displayed in the table above.

Part 4: Risk and Data Correction

4.1. A type I error occurs when we incorrectly identify a process as unstable, but it was stable all along. This is known as an α-error. In statistical process control, H_0 assumes that the process is in control and centred around a calculated centre line. The type I error incorrectly rejects the null hypothesis (H_0) when it is actually true.

The type I error varies according to each detection rule:

Rule A: Any Point Beyond 3σ Control Limit (s-Charts)

Type I error probability (α) = 0.0027 ≈ 0.270%

The probability that a point is detected to fall outside of the ±3σ control limits when the process is actually in control are 0.270%.

Rule B: Most Consecutive Samples Between ±1σ Limits

The maximum number of 12 consecutive samples within the ±1σ control limits were used to calculate the type I error.

Longest consecutive run within ±1σ: 12 samples

Type I error probability (α) = 0.01024893 ≈ 1.025%

Rule C: Four consecutive points outside the upper 2σ limit

Type I error probability (α) = 0.00000027 ≈ 0.000027%

The probability that 4 consecutive points fall within the upper 2σ limit when no actual process shift occurred is 0.000027%.

4.2. A type II error occurs when we incorrectly identify a process as stable, but it was unstable all along. This is known as a β-error. In statistical process control, H_a assumes that the process is out of control and has shifted from the centre line. The type II error incorrectly fails to reject the null hypothesis (H_0) when it is actually false.

β = P(LCL < X̄ < UCL | μ = 25.028, $\sigma_{\bar{x}}$ = 0.017) = 0.841178 ≈ 84.12%

Power to detect shift = 1 - β= 0.158822 ≈ 15.88%

4.3. The key changes made to the head office and product data involved standardising product categories from descriptive names to uniform product codes. Rerunning the data analysis with the corrected dataset resulted in notable changes to two key diagrams.

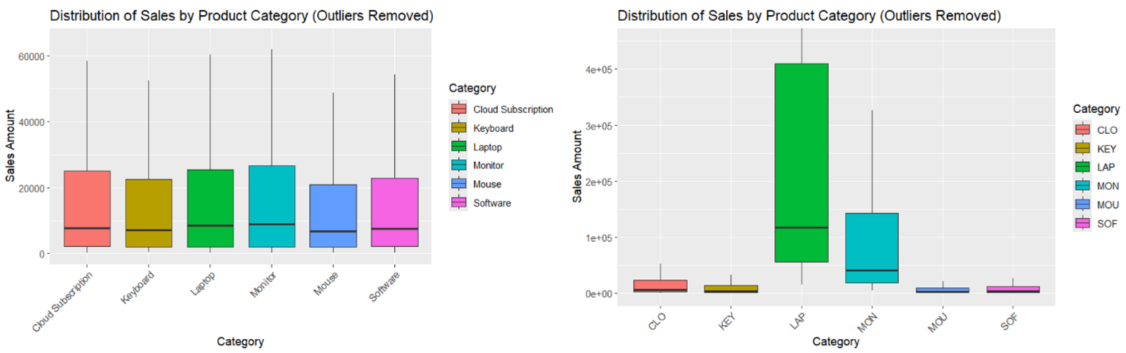


Figure 25 Comparison of the Distribution of Sales by Product Category

The standardisation of category labels significantly improved data consistency, eliminating the potential for misinterpretations that came from previously mixed naming conventions. The new distribution of sales amounts across product categories changed dramatically, now more accurately portraying the reality that laptops generate the largest proportion of sales. The bar-and-whiskers chart also confirms that monitors represent another significant contributor to the overall sales volume.

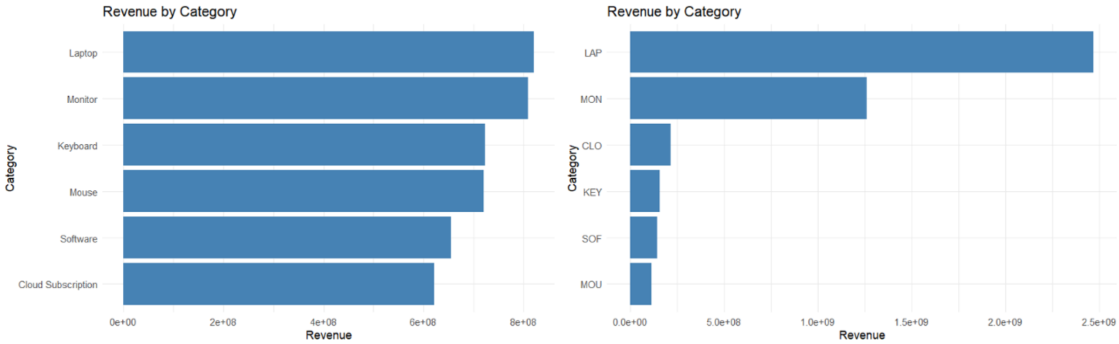


Figure 26 Comparison of the Revenue by Category

There was also a significant discrepancy between the previous and corrected *Revenue by Category* analysis. The new graph now correctly represents the proportion of total revenue provided by each category, highlighting the huge contribution of laptops and monitors compared to the other four categories.

The standardisation of the data was crucial, as it transformed the data to truthfully depict how revenue is actually distributed across each product category. This fixes the error identified in the initial data analysis, where it was initially commented that one would expect the ratio of laptops and monitors to be slightly higher compared to the other 4 products. It is clear that the fixed data has now allowed for a clearer depiction of revenue distribution according to different products.

Total Sales Value of 2023 per Type

Product Type	Total Sales Original	Total Sales Updated
CLO	98715482	98715482
KEY	73499067	73499067
LAP	1163889479	1163889479
MON	578385570	578385570
MOU	51219577	51219577
SOF	66468485	66468485

Table 5 Total Sales Value of 2023 per Type

Part 5: Optimising Profit

Objective

The objective of this analysis is to address the following business questions:

1. Estimate the percentage of clients who should expect reliable service.
2. Build a model to optimise the number of baristas per weekday for 2 coffee shops.

Analysis and Model Building

The focus of this analysis was to optimise the profitability of the coffee shop by determining the ideal number of baristas that would balance service efficiency against employee costs. The *Time to Serve* datasets contain the individual service times across different staffing levels over the period of one year for both shops.

All the individual service times were used for the analysis, using key metrics such as:

- Average service time per barista configuration
- Customers served per day (8-hour operating window)
- Material profit: R30 per customer (excluding employee costs)
- Employee Costs: R1000 per barista per day
- Reliability Threshold: Service time ≤ 300 seconds is considered acceptable

The daily profit was calculated as follows:

$$Daily\ Profit = (Customers\ Serves \times R30) - (Baristas \times R1000)$$

Results

All clients should expect 100% reliable service. Average service time was calculated according to different configurations of baristas from 1-6. On average, coffee shops should take around 1.5-2 minutes to make a coffee. The highest average service time is 200 seconds, slightly higher than this threshold. Employing more than 1 barista would ensure that all client services are met on time for both shops.

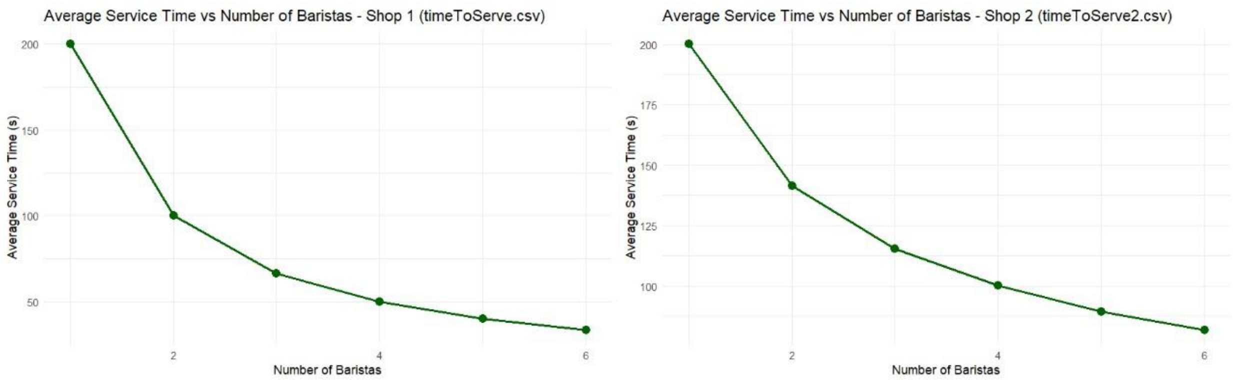
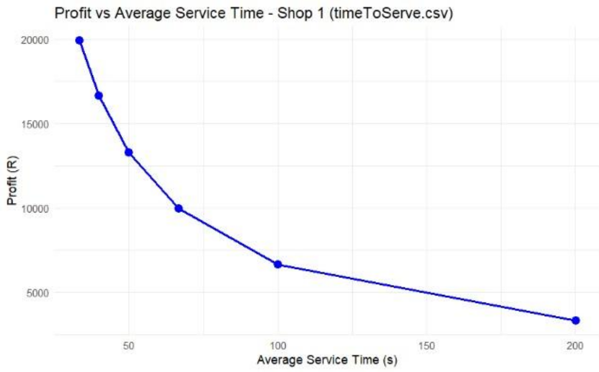


Figure 27 Average Service Times vs Number of Baristas for Coffee Shops 1 and 2

For shop 1, having shorter service times results in a higher profit. With more baristas working in the shop, the work is split evenly, which lowers average service time. Orders are met more quickly, bringing in income much faster than it would if service times were slower.

For shop 2, however, if service times are too fast, profit actually decreases. This could be due to under-utilisation, where baristas are spending time idly, or it could be due to constrained resources, like having to share 1 grinder between 6 people. These factors could be the reason for the highest profitability being found at slightly higher average service times.

The graphs below illustrate the concepts discussed above.



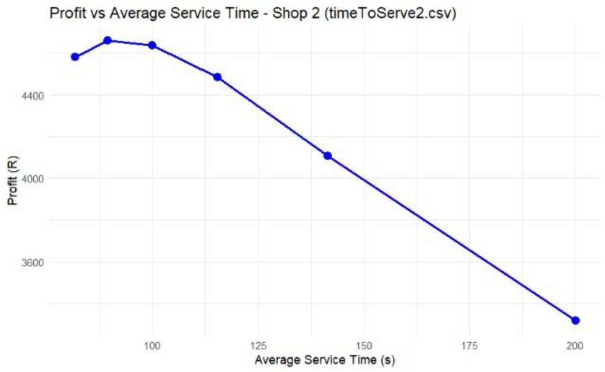
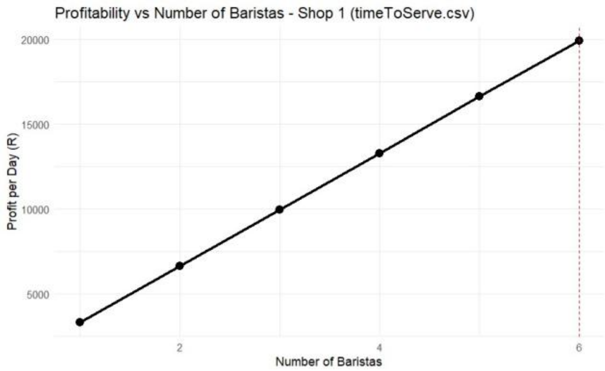


Figure 28 Profit vs Average Service Time for Coffee Shops 1 and 2

Profitability was graphed against the number of baristas working per day. Because more baristas improve service speeds for shop 1, more sales are able to take place, increasing the profit. According to the graph below, the ideal number of baristas per day is 6, resulting in a maximum profit of R19,902.66 per day.

However, for shop 2, the ideal number is 5, resulting in the highest profit of R4,660.54 per day. More than 5 baristas could be causing congestion in the coffee-making station, or wages could be lowering overall profit due to higher employment expenses.



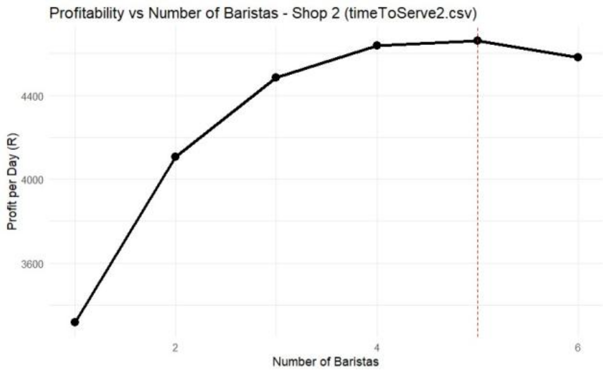


Figure 29 Profitability vs Number of Baristas for Coffee Shops 1 and 2

One must consider shop space, location, average customers, and wage expenses when deciding on the ideal number of baristas. Several factors could influence profit, so it is important to consider the shop's specific business model in order to make decisions.

Part 6: DOE and MANOVA/ ANOVA

This section focuses on determining whether there are significant differences in sales and operational performance across various time periods using ANOVA and MANOVA techniques. Variables such as total quantity, picking hours, and delivery hours were tested to see whether they differ significantly between years.

The sales data was aggregated by year and month to calculate the total quantity sold for each time period. A one-way ANOVA was performed to test whether there were significant differences between treatment groups. The Least Significant Difference (LSD) method was applied to identify which specific groups differed, and thereafter, a MANOVA was performed to assess whether the combined effect of a year on the 3 variables was significant. The Pillai's trace statistic was used at the multivariate test criterion.

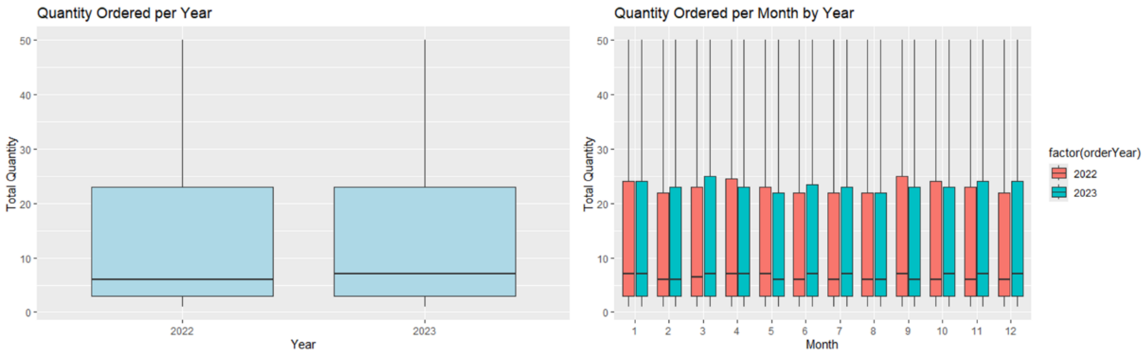


Figure 30 Quantity Ordered per Year and per Month by Year

According to the graphs above, comparing order quantities between the 2 years, the ANOVA and MANOVA should not display a significant difference.

ANOVA/ MANOVA Results

The results from both the ANOVA and MANOVA tests indicate that there is no statistically significant difference between the years 2022 and 2023 in terms of order quantities and picking and delivery hours. The MANOVA test using Pillai's trace resulted in a p-value of 0.1235 > 0.05 significance level. This suggests that the observed variations are likely due to random chance rather than actual differences in performance between the 2 years.

Similarly, the Least Significant Differences (LSD) value was calculated as 11,972.62, meaning that only differences in mean that are greater than this threshold will be considered statistically significant. In the case of the years 2022 and 2023, the difference is just -0.0301, which is much smaller than the LSD value.

The ANOVA pairwise comparison results from the estimated marginal means (*emmeans*) also support this result, with a p-value of 0.5311, confirming that the difference in means between the 2 years is not statistically significant.

From a service and reliability perspective, no significance indicates that the level of service delivery in terms of order processing speed and fulfilment timeliness remained stable and predictable between 2022 and 2023. Customers could therefore expect a similar standard of reliability from one year to the next, with no statistically significant changes in service.

Part 7: Reliability of Service

The aim of this analysis is to estimate how often reliable service can be expected from a car rental agency, and to determine how staffing levels affect profit. Reliable service is defined as having at least 15 employees on duty. When staffing falls below this threshold, the company loses an average of R20,000 in sales per day. Each employee costs R25,000 per month.

7.1. $Total\ Days = 397$

$$Reliable\ Days = 96 + 270 = 366$$

$$Reliability\ Rate = \frac{366}{397} \times 100 = 92.19\%$$

$$Reliable\ Days\ per\ Year = 365 \times 92.19\% = 336.499 \approx 336 \text{ days}$$

This means that approximately 92.19% of the days, the company offers reliable service, with around 336 reliable days per year.

7.2. The company loses R20,000 per day when there are fewer than 15 employees.

$$Unreliable\ days = 1 + 5 + 25 = 31$$

$$Expected\ Loss = 31 \times 20000 = R620\ 000$$

The other 29 days of the year are likely to experience service problems due to insufficient staffing. Hiring more employees could solve the problem, even though it would cost more in terms of wage payments. The scenario was modelled to evaluate the financial trade-off between expected savings from reducing unreliable days and additional annual employee costs.

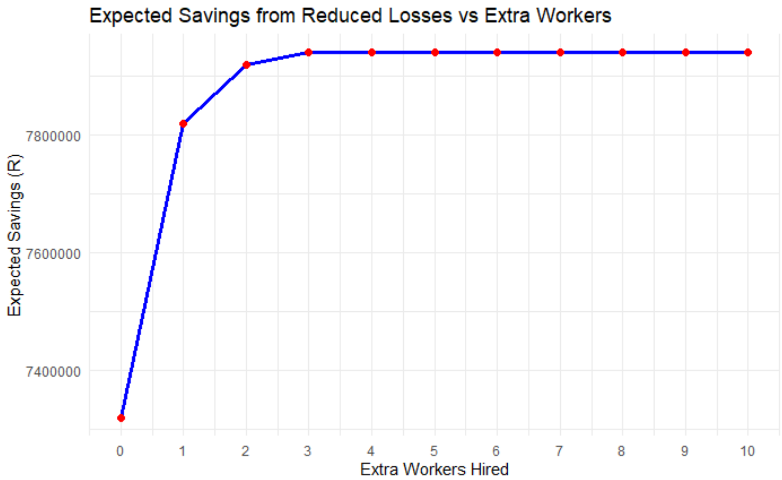


Figure 31 Expected Savings from Reduced Losses vs Extra Workings

Expected savings increase from hiring 1-3 new employees and then remain constant despite having more staff. Therefore, the company should hire no more than 3 more employees.

However, the figure below shows that the net gain is maximised at 1 extra worker. With one additional employee, reliable days increase to 391, which means that 98.49% of days will be reliable. Although there are not 100% reliable days, the goal is to maximise profit, and if more than one employee is to be hired, then revenue with decrease due to the additional personnel costs.

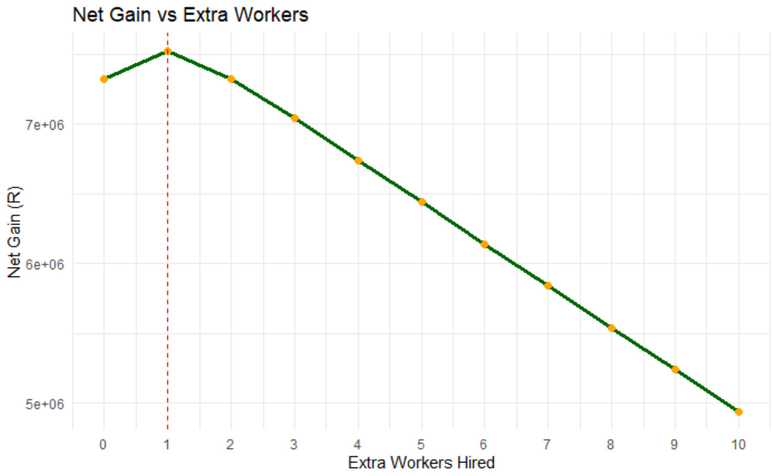


Figure 32 Net Gain vs Extra Workers

Therefore, hiring one more person will optimise the company's profit.

Conclusion

This report demonstrated the application of statistical processes to solve real-world business problems. Through a structured analysis of sales, operational, and delivery data, several key insights into the business and recommendations were made to improve business performance, process stability, and profitability.

The data analysis revealed that the company's revenue is heavily dependent on a small number of high-value products, namely laptops and monitors, while other products like software and cloud subscriptions are slightly underperforming. Though an RFM analysis, customers were segmented into high-value customers for retention programs, and bottlenecks in order fulfilment were identified as critical areas for improvement.

The SPC analysis of delivery times for the years 2026–2027 indicated that while process variability (s-charts) was mostly stable but mean delivery times (X-bar charts) consistently exceeded control limits, especially toward the end of each year. This points to a systemic issue in maintaining delivery performance, likely due to seasonal demand or inadequate year-round process monitoring. Process capability analysis further confirmed that only software deliveries met customer specifications, with all other product types failing to achieve a Cpk ≥ 1.

Risk analysis calculated type I and type II errors, highlighting the trade-offs in process monitoring decisions. Data correction standardized product categories, leading to more accurate revenue distributions, confirming the much larger proportion of laptops and monitors in sales compared to other products.

An optimization model for coffee shop staffing identified ideal barista levels to maximize profit - 6 baristas for Shop 1 and 5 baristas for Shop 2 - balancing average service level against labour costs. Similarly, for the car rental agency, hiring one additional employee would maximize net gain by increasing reliable service days to 98.49%, reducing the losses from understaffing.

MANOVA/ANOVA tests confirmed that there was no significant difference in order quantities or operational performance between the years of 2022 and 2023, indicating consistent performance.

In summary, this project displays the importance of data-driven decision-making in industrial engineering. By implementing the recommended strategies, the company can improve delivery reliability, diversify revenue sources, and enhance overall profitability. The company should focus on addressing seasonal delivery fluctuations and subsequent inefficiencies, and further refine operational models based on continuous data monitoring.

References

Chun, S. (2021). *How Long Should It Take To Help a Customer at Your Coffee Shop?* [online] Start My Coffee Shop. Available at: https://startmycoffeeshop.com/how-long-should-it-take-to-help-a-customer-at-your-coffee-shop/

Grammarly (2025). *Grammarly*. [online] Grammarly.com. Available at: https://www.grammarly.com

OpenAI (2025). *ChatGPT*. [online] ChatGPT. Available at: https://chatgpt.com/

van Schalkwyk, T.D. (2025). *QA344Statistics*. [online] STEMLearn. Available at: https://stemlearn.sun.ac.za/pluginfile.php/65516/mod_resource/content/5/QA344%20Statistics.pdf