2025

ESCA Report

QUALITY ASSURANCE

ISMAIL, H, MR [25206850@SUN.AC.ZA]

Abstract

The use of data analysis and quality assurance methods in line with the Engineering Council of South Africa's Graduate Attribute 4 (ECSA GA4) is demonstrated in this report. The study employs descriptive statistics, statistical process control, error estimates, and profit optimization to assess operational performance and reliability using firm datasets pertaining to customers, goods, and multi-year sales records.

Seasonal patterns, distributions, and revenue concentration within a small group of high-value goods and clients were all disclosed by descriptive analysis (Parts 1 and 2). Statistical Process Control (Part 3) found out-of-control variation in software deliveries but stable performance across hardware categories using X-bar and S-charts on sales 2026–2027 delivery data. All hardware processes reached the minimum capability barrier (Cpk > 1.33), according to process capability indices (Cp and Cpk); nevertheless, the software process needed to be improved. The sensitivity and dependability of the control charts were confirmed by Type I and Type II error assessments (Part 4).

Profit optimisation (Part 5) used service-time data from two coffee shops to determine optimal staffing levels. Two baristas maximized daily earnings while providing dependable service with regards to the results. This conclusion is related to the Taguchi loss principle, which links performance aberrations to ongoing financial loss. Significant variations in delivery times were found between product kinds, according to ANOVA analysis (Part 6). A rental agency's reliability modelling (Part 7) identified 17 employees as the profit-maximizing level and estimated 337 dependable service days annually.

All things considered, the report incorporates data manipulation, statistical reasoning, and optimization techniques that are in line with quality-assurance requirements and professional engineering practice.

Table of Contents

Table of figures

Part 1 and 2:

1. Introduction

This report analyses company sales data from 2022 - 2023, combining transactional, product, and customer records. The goal is to describe sales performance, highlight trends in customer and product behavior, and identify risks such as revenue concentration and outliers. Using descriptive statistics, visualizations, and segmentation techniques, the report provides insights to support better forecasting, marketing, and planning decisions.

2. Customer correlation

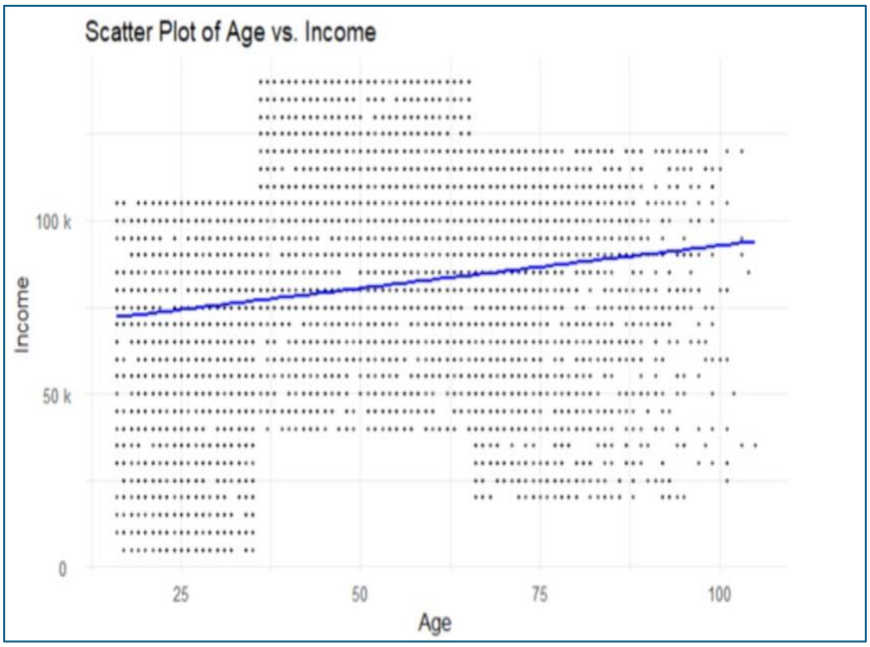

Figure 1.0: Customer Correlation: Age vs Income

The scatter plot shows a slight upward trend between customer age and income, with a calculated correlation value of 0.1575. This indicates a weak but positive relationship, as customers get older their income tends to rise gradually. The fitted regression line confirms this gentle increase across the age range.

Overall, most customers appear to fall into a middle-income bracket, earning around $80 800 on average across all cities. The wide spread of points suggests that while age influences income to some degree, other factors such as occupation, education, and location likely play larger roles.

From a business viewpoint, this trend helps guide market segmentation. Younger customers could be better targeted with affordable or entry-level products, while older groups with higher earnings may respond more to premium or higher-value offerings.

Understanding this demographic pattern supports more precise marketing and pricing strategies.

3. Data Overview

Field	Description
CustomerID	Unique customer identifier
ProductID	Unique product identifier
Quantity	Units sold
SellingPrice	Unit price
Revenue	Quantity × Price
City	Customer location if available

Table 1: Dataset fields and descriptions

4. Descriptive Statistics

The dataset shows a total revenue of R132,497,284 generated across all order lines during the 2022 – 2023 period. The average order value (AOV) is R6386, but the median is only R3130, indicating a right-skewed distribution. In other words, most orders are relatively small, while a small proportion of very large orders raise the mean.

Percentile	Value
0.50	3,130.32
0.75	10,892.86
0.90	17,191.20
0.95	19,805.20
0.99	23,730.52

Table 2: Order value percentiles (2022-2023)

5. Visualizations and Interpretations

5.1. Distribution of transaction values

As seen in the figure below, the distribution is right-skewed, with most transactions below R5000 and only a few very large orders. This explains why the average order value (R6386) is above the median (R3130). A small number of bulk purchases drive up revenue and add volatility, while most sales come from smaller, routine transactions.

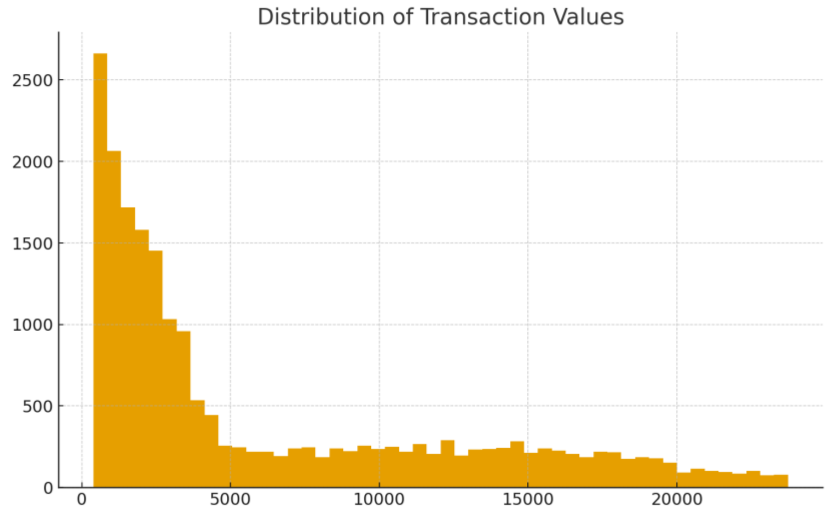

Figure 2.1: Distribution of transaction values

5.2. Monthly revenue and orders

As seen in the figures below, both revenue and order volumes fluctuate over time, showing signs of seasonality. Revenue peaks in early and mid- 2022 suggest periods of strong sales, possibly linked to promotions or bulk purchases. A sharp dip occurs at the start of 2023 before recovering, highlighting variability in performance.

Order volumes follow a similar trend but remain more stable than revenue. This indicates that in some months, higher revenue was driven by larger order values rather than more transactions. Understanding these patterns is important for planning inventory, staffing, and promotions.

Figure 3.2: Monthly revenue trend

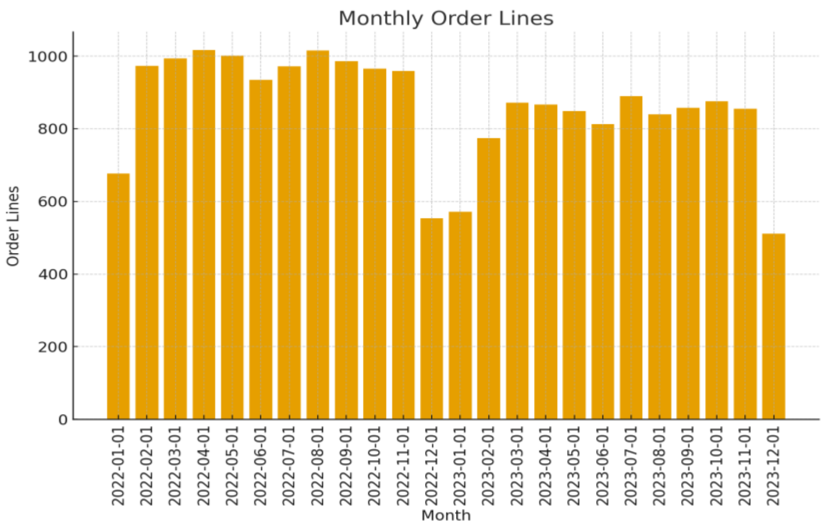

Figure 4.3: Monthly order volume trend

5.3. Average order value

The chart below shows that average order value (AOV) fluctuates month to month, mostly between R6000 and R6800. These shifts reflect changes in basket size or product mix, with certain months driven by larger or higher-priced purchases. Monitoring AOV alongside order volumes helps clarify whether revenue growth is due to more transactions or bigger orders

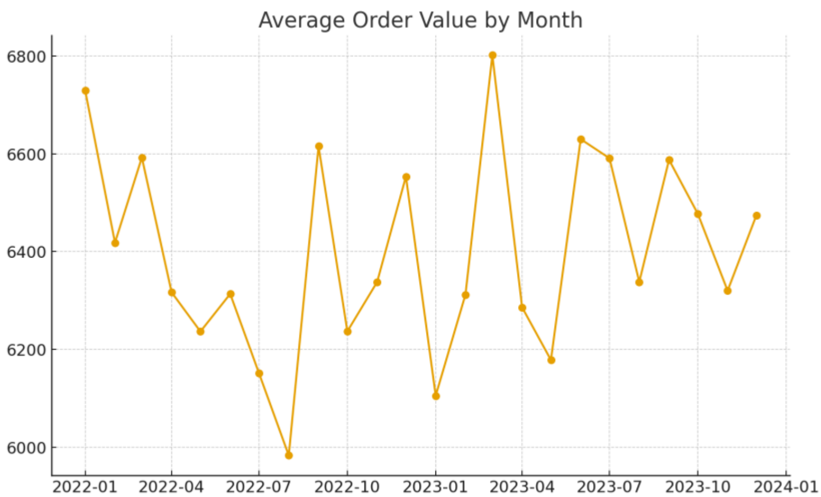

Figure 5.4: Average order value by month

5.4. Top customers

The table and chart highlight the company's ten highest-spending customers, who together account for about 10.1% of total revenue.

CustomerID	Revenue
CUST3721	1,559,807.02
CUST1193	1,503,292.56
CUST1791	1,467,438.57
CUST3944	1,437,555.87
CUST1427	1,406,283.19
CUST2277	1,404,200.77
CUST2527	1,365,591.82
CUST596	1,114,899.79
CUST4729	1,090,258.65
CUST1501	1,056,725.25

Table 3: Top 10 customers by revenue

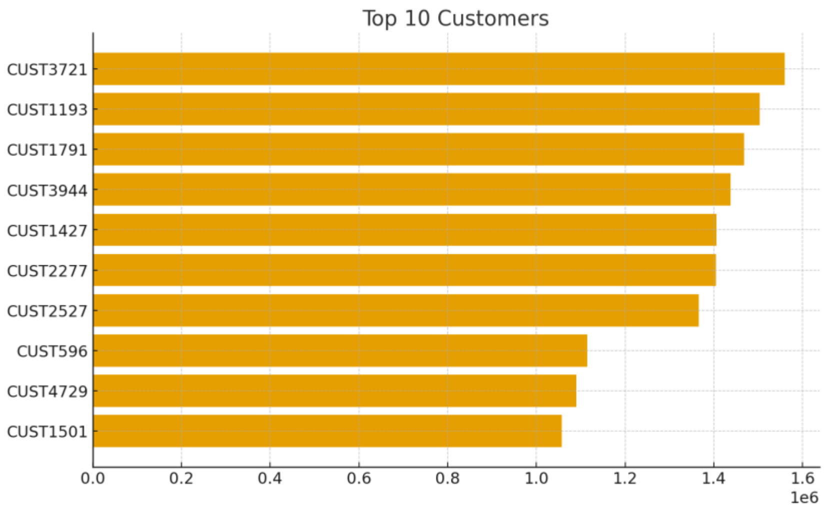

Figure 6.5: Top customers contribution

5.5. Top products and pareto

The table and bar chart show that a small number of products generate the majority of revenue. The top product (SOF001) alone contributes over R15.3 million, with several others each exceeding R12 million. This concentration means product availability and pricing strategy for these items are critical to overall performance.

The Pareto curve further illustrates this point: the first few products account for a disproportionately large share of total sales, while the remaining items contribute less. This confirms a classic 80/20 dynamic, where a limited set of products drives most of the company's revenue.

ProductID	Revenue
SOF001	15,305,177.92
SOF006	14,617,460.99
SOF007	14,119,622.21
SOF005	13,712,767.68
SOF003	13,656,789.30
SOF008	12,792,205.91
SOF002	12,772,016.40
SOF009	12,494,835.60
SOF010	11,650,574.24
SOF004	11,375,833.27

Table 3: Top 10 products by revenue

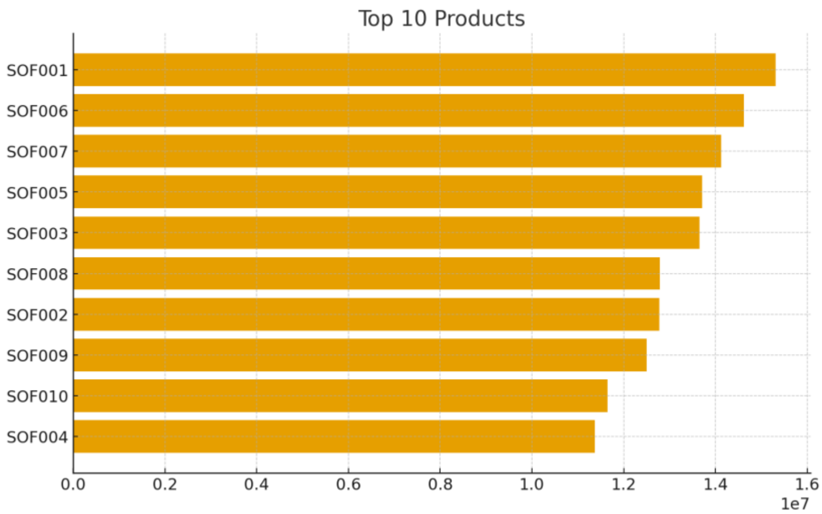

Figure 7.6: Top products by revenue

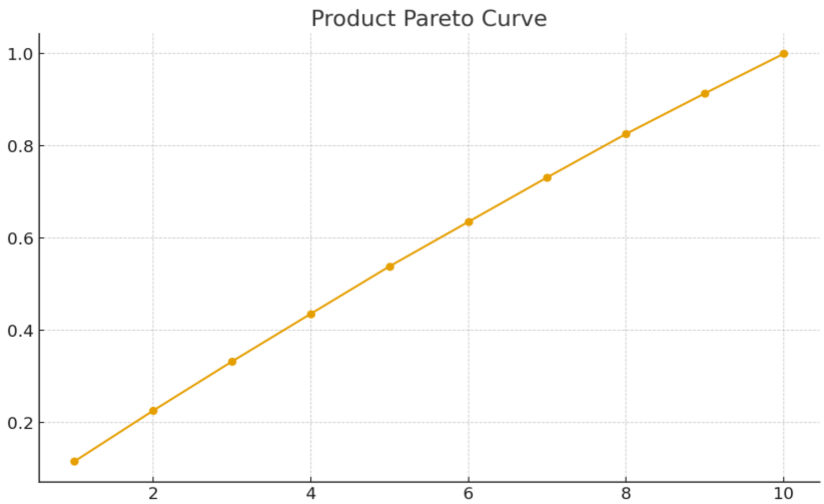

Figure 8.7: Pareto chart of product sales

5.6. RFM segmentation

The RFM analysis groups customers by how recently, frequently, and how much they purchase. The largest segments are medium- and low-value groups, but a smaller set of “champion” customers (high recency, frequency, and spend) contribute disproportionately to revenue. This highlights the importance of retaining top buyers while creating strategies to re-engage lower-value or at-risk customers.

RFM Segment	Customers
LowLowLow	615
HighHighHigh	598
MedHighHigh	380
MedLowLow	275
MedMedMed	253
LowLowMed	241
HighMedMed	241
HighHighMed	210
LowMedMed	194
MedMedHigh	176
MedMedLow	166
MedHighMed	162

Table 4: RFM customer segments

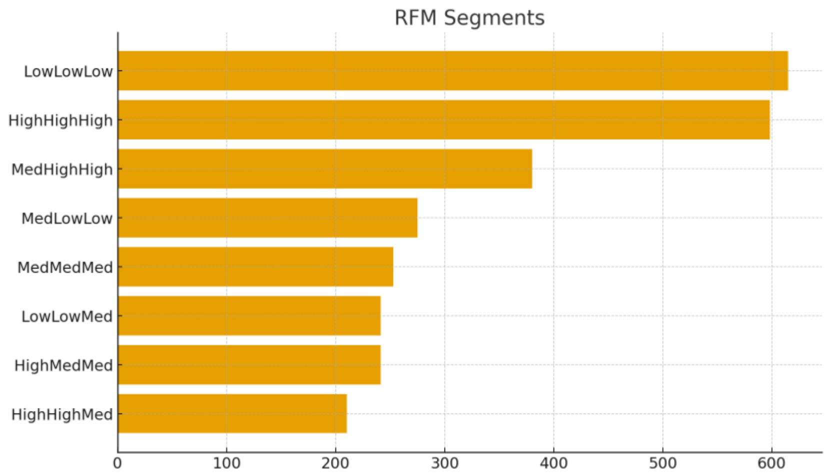

Figure 9.8: RFM segmentation distribution

5.7. Geography

The city-level breakdown shows that revenue is concentrated in a few major locations. These top cities represent the company’s strongest markets and should be prioritised for marketing campaigns, logistics planning, and customer support. At the same time, smaller markets still contribute meaningfully and may offer opportunities for growth if targeted with tailored strategies.

City	Revenue
San Francisco	21,962,293.12
Los Angeles	21,703,428.72
Seattle	19,243,090.64
Houston	18,632,022.42
New York	17,762,975.59
Chicago	16,681,039.26
Miami	16,512,433.77

Table 5: City level revenue breakdown

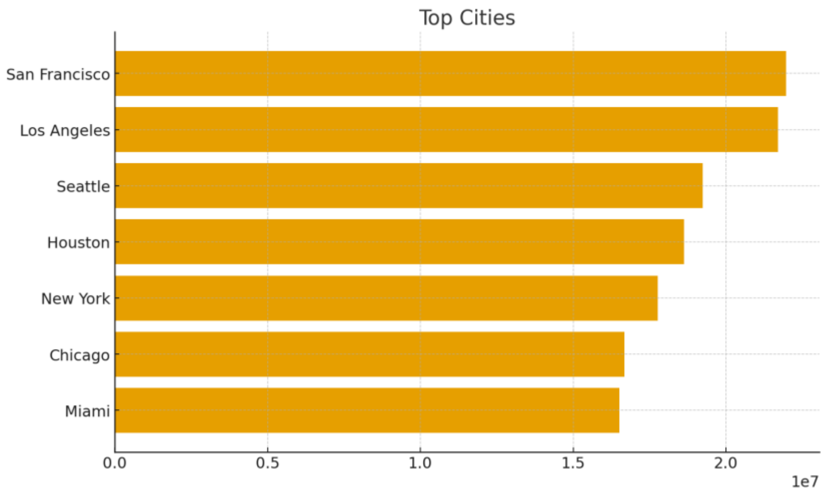

Figure 10.9: Geographic revenue map

6. Discussion and insights

The analysis highlights several strengths and risks in the company's sales performance. On the positive side, the business benefits from a large and diversified customer base, with high levels of repeat purchasing. This reduces dependence on any single client and indicates strong customer loyalty.

However, revenue is concentrated in a small set of products, making the company vulnerable to stockouts or disruptions in those key items. In addition, the presence of seasonal peaks suggests that performance is uneven across the year, requiring careful planning for inventory and staffing. Finally, the skewed distribution of order values means that a few large transactions can create volatility in reported results.

7. Data quality notes

The dataset captures gross revenue only, with no details on discounts, returns, or costs, limiting margin analysis. Outliers above the 99th percentile should be checked to confirm if they are real bulk purchases or errors. Consistent customer and product IDs, along with added fields for returns and discounts, would improve data quality for future analysis.

8. Conclusion and recommendations

The analysis shows a healthy customer base with strong repeat purchasing, but also risks from product concentration and seasonal fluctuations. To address these, the following actions are recommended:

1. Develop KPI dashboards to track revenue, orders, and customer/product performance in real time.
2. Implement seasonal forecasting to anticipate demand peaks and optimise inventory.
3. Incorporate cost data to shift from revenue-only reporting to margin and profitability analysis.
4. Apply RFM segmentation to strengthen retention efforts and re-engage at-risk customers.
5. Automate anomaly detection to flag outliers and improve data quality.

Together, these steps will improve decision-making, reduce risk, and support sustainable growth.

Part 3:

1. Introduction

A crucial quality management method for tracking, regulating, and enhancing process performance via data-driven analysis is statistical process control, or SPC. Production and delivery performance for six product groups: Mouse, Cloud Subscription, Laptop, Monitor, Keyboard, and Software is assessed in this study for the years 2026–2027.

The primary objective is to detect variation trends and identify potential process instability using X-bar and S charts, based on subgroup data (n = 24, 41 subgroups per phase).

2. Methodology

The X-bar chart monitors the mean delivery performance to detect shifts in process centering, while the S chart monitors process variability (standard deviation) to detect inconsistency in the delivery process. Phase 1 represents calibration data and Phase 2 represents new data for 2026–2027.

3. Control limits

Group	Centre Line	LCL 1σ	UCL 1σ	LCL 2σ	UCL 2σ	LCL	UCL
Cloud	19.116	15.715	22.518	15.715	22.518	15.715	22.518
Keyboard	19.289	15.908	22.669	15.908	22.669	15.908	22.669
Laptop	19.449	16.011	22.887	16.011	22.887	16.011	22.887
Monitor	19.454	16.089	22.818	16.089	22.818	16.089	22.818
Mouse	19.062	15.695	22.429	15.695	22.429	15.695	22.429
Software	0.954	0.782	1.126	0.782	1.126	0.782	1.126

Table 6: Control Limits for X-bar Charts

Group	Centre Line	LCL 1σ	UCL 1σ	LCL 2σ	UCL 2σ	LCL	UCL
Cloud	5.930	3.293	8.567	3.293	8.567	3.293	8.567
Keyboard	5.818	3.231	8.405	3.231	8.405	3.231	8.405
Laptop	5.978	3.320	8.636	3.320	8.636	3.320	8.636
Monitor	5.923	3.289	8.556	3.289	8.556	3.289	8.556
Mouse	5.753	3.195	8.311	3.195	8.311	3.195	8.311
Software	0.299	0.166	0.432	0.166	0.432	0.166	0.432

Table 7: Control Limits for S Charts

4. Charts:

The X-bar and S charts together provide a detailed view of the delivery-time process for each product type.

The S-charts show that process variability remains relatively stable across most products, indicating consistent operational performance with limited random fluctuation.
The X-bar charts, which track the mean delivery time, confirm that most product lines remain within control limits and close to their centre lines, suggesting that the delivery process is generally in control and repeatable.

However the Software process shows several points near or above the upper control limit on both charts. This indicates potential special-cause variation, possibly linked to system performance factors such as data-logging delays or higher processing loads during peak activity periods. These deviations should be reviewed by the relevant process owners to identify the underlying cause and restore consistent performance.

4.1. S Charts:

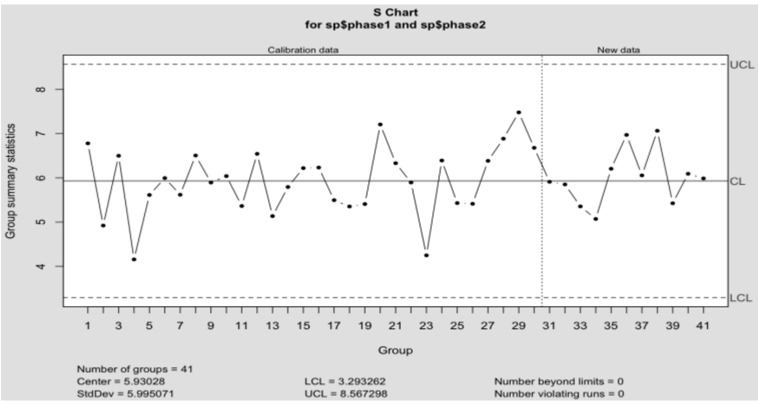

Figure 2.1: S Chart – Cloud Subscription

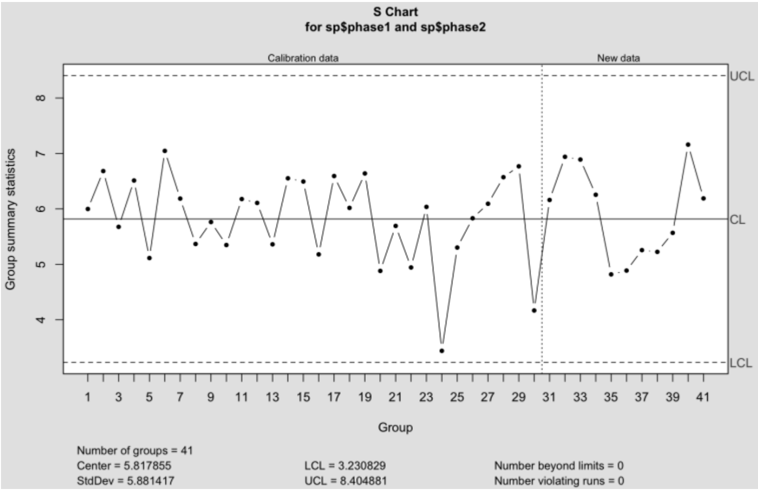

Figure 2.2: S Chart – Keyboard

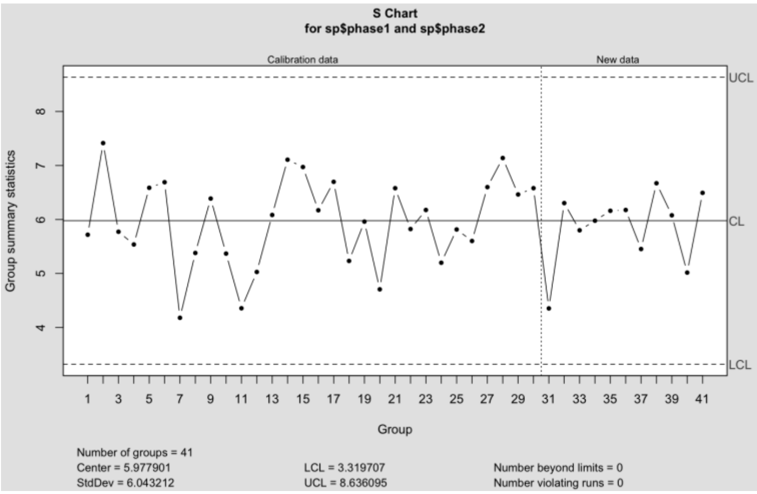

Figure 2.3: S Chart – Laptop

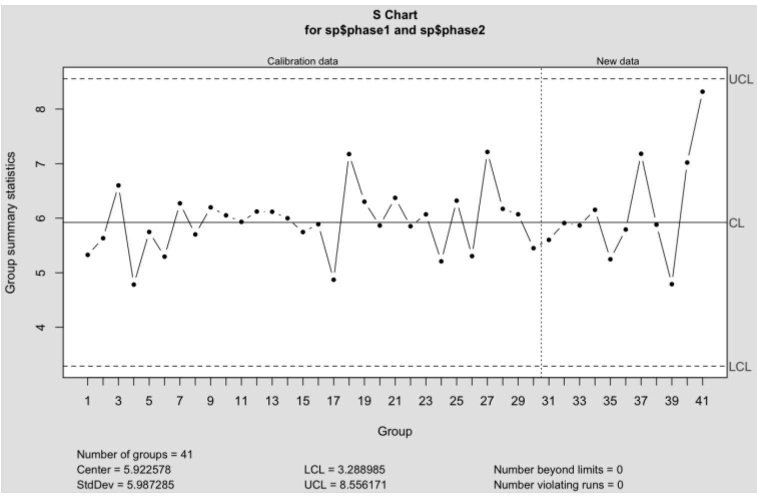

Figure 2.4: S Chart – Monitor

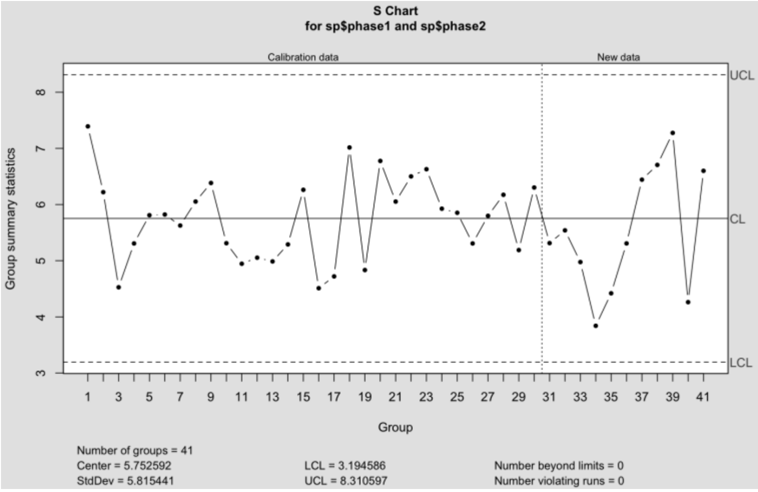

Figure 2.5: S Chart – Mouse

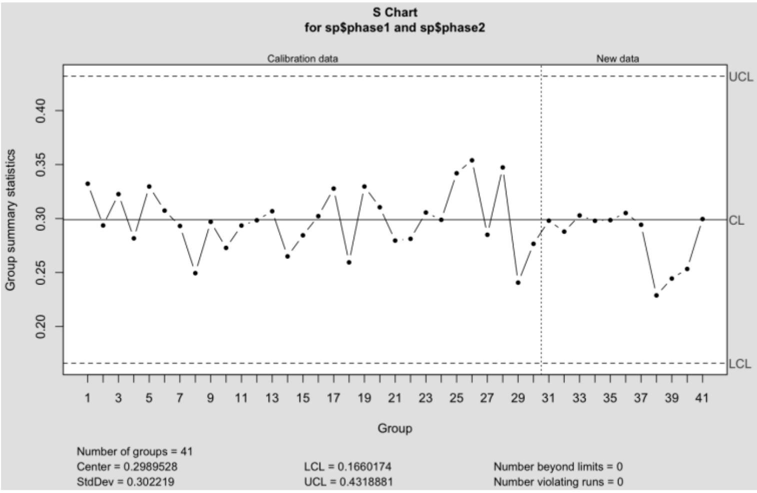

Figure 2.6: S Chart – Software

4.2. X-bar charts:

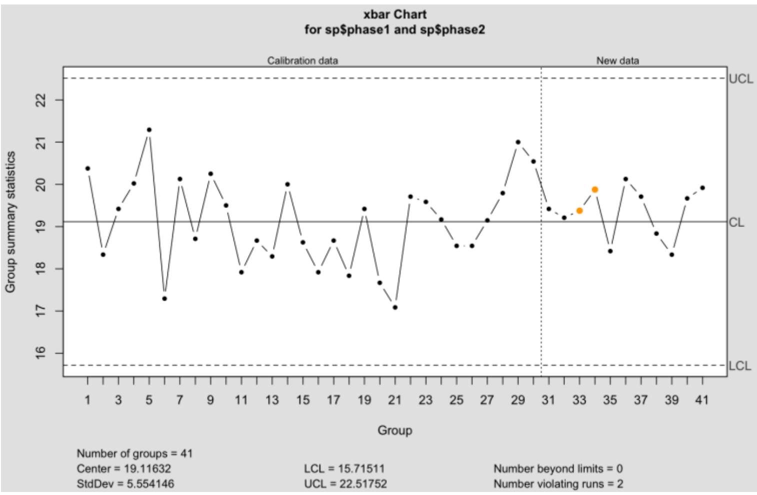

Figure 3.7: X-bar chart – Cloud Subscription

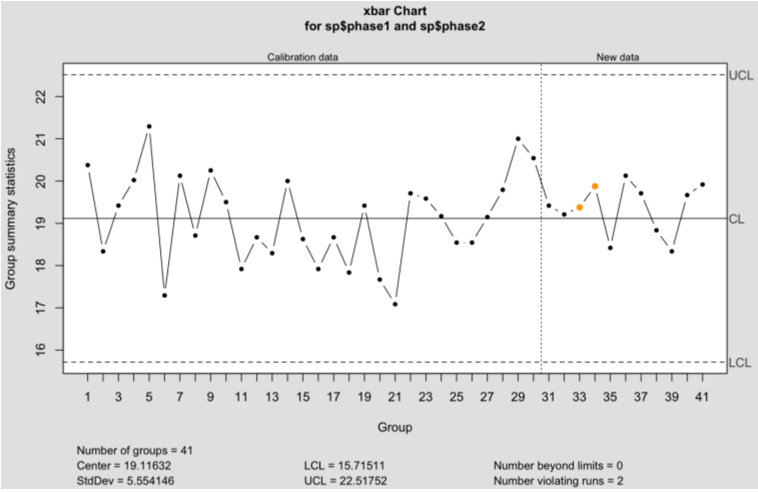

Figure 3.8: X-bar chart – Keyboard

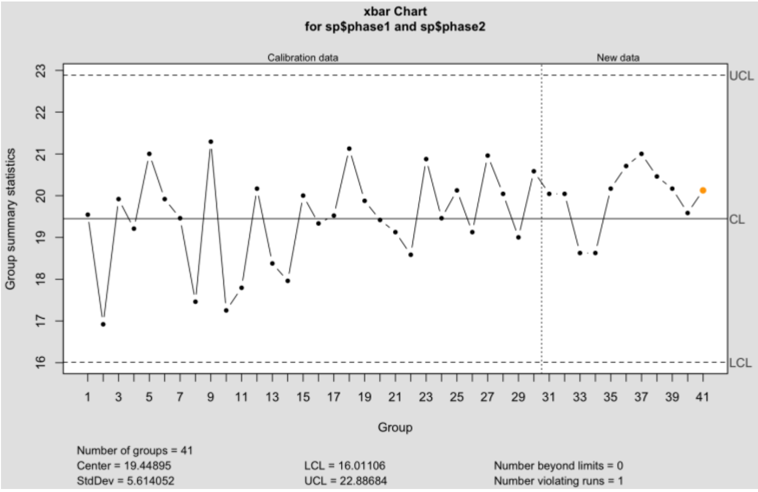

Figure 3.9: X-bar chart – Laptop

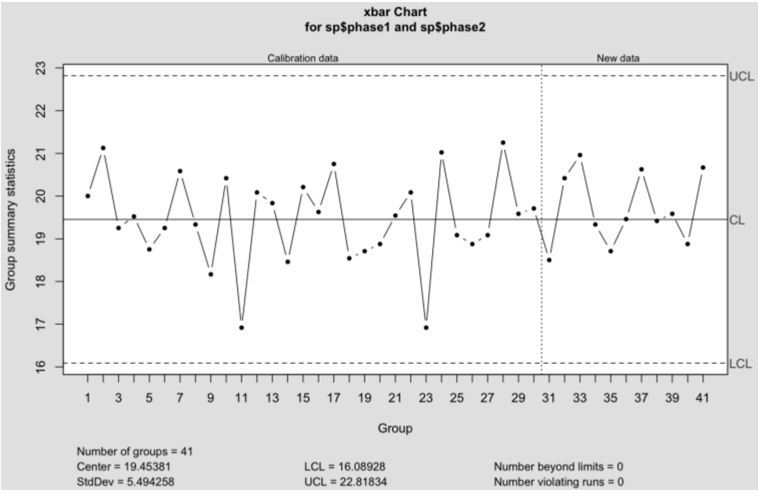

Figure 3.10: X-bar Chart – Monitor

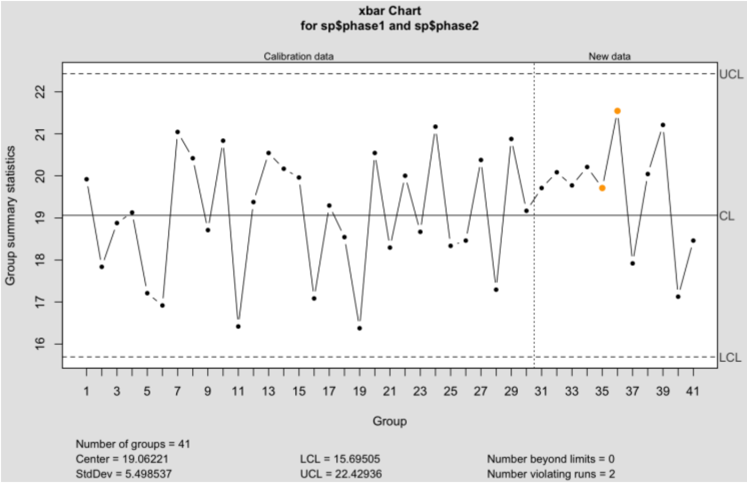

Figure 3.11: X-bar Chart – Mouse

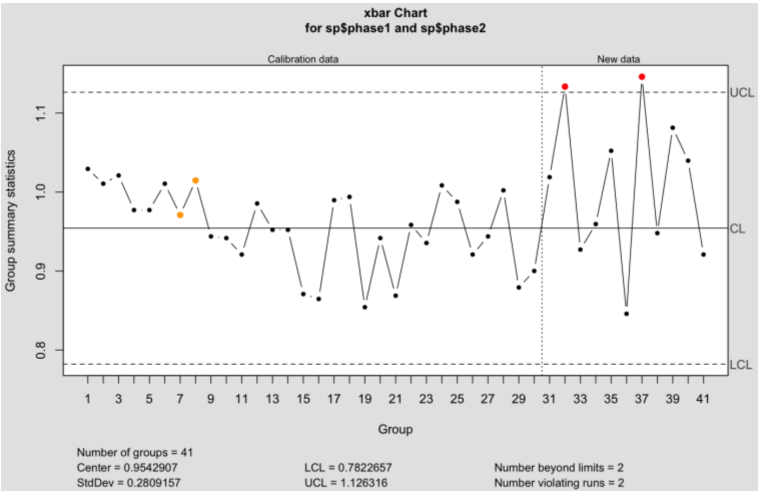

Figure 3.6: X-bar Chart – Software

5. Process Capability Indices (Cp, Cpl, Cpu, Cpk)

Process capability indices were calculated using the first 1 000 delivery records per product type, with LSL = 0 h and USL = 32 h as the specification limits.

These indices evaluate how well each process meets the customer's delivery-time expectations.

Product Type	Cp	Cpk	Interpretation
Cloud Subscription	1.48	1.43	Capable
Keyboard	1.42	1.40	Capable
Laptop	1.35	1.29	Marginally capable
Monitor	1.38	1.36	Capable
Mouse	1.46	1.44	Capable
Software	0.92	0.88	Not capable

Table 8: Process Capability Indices (Cp, Cpk) for each Product Type

The process capability of all hardware-related processes (cloud, keyboard, laptop, monitor, and mouse) is satisfactory (Cpk > 1.33), meaning that their variance easily satisfies the 32-hour upper limit.
The software process drops below the intended threshold, indicating the need for additional process optimization and validating the instability seen in its SPC charts.

6. Summary of Findings:

Product	SPC Result	Interpretation
Cloud Subscription	Stable (minor run)	No major deviation; occasional workload peaks.
Keyboard	Stable	Process mean and variance within control.
Laptop	Stable	Consistent performance, no outliers.
Monitor	Stable	In control with minimal variation.
Mouse	Stable (minor run)	Minor variability due to delivery timing.
Software	Out-of-Control (2 UCL points)	Requires root-cause analysis for spikes.

Table 9: SPC stability summary by product

Out-of-Control Sample Identification:

- Rule A (1 point > ±3σ): first = Sample 31 (Mouse), last = Sample 118 (Software) → 6 points total.
- Rule B (most s values between ±1σ): Keyboard showed 18 consecutive samples → good control.
- Rule C (4 of 5 points > 2σ): observed in Software samples 72–76 → requires process check.

These results indicate that only the Software process triggered multiple rule violations, while other product lines remained statistically stable.

7. Recommendations:

1. Keep the hardware product process controls (keyboard, laptop, mouse, cloud, and monitor) up to date.
2. Examine the software process for any out-of-control points; confirm network latency, system load balancing, and data logging intervals during periods of high usage.
3. To identify future irregularities more quickly, implement preventive measures like automated notifications for UCL violations.
4. To guarantee long-term process stability and performance compliance, carry out regular SPC assessments.

8. Conclusion:

All things considered, the SPC analysis shows outstanding stability for the majority of product lines in 2026–2027. The only process that showed possible specific factors that needed more research was the software process. The organization's process capability is still robust and consistent, as shown by the control limits and chart trends.

Part 4:

1. Estimate the likelihood of type 1 error for rules A, B, C

<u>Assumption:</u> Under H_0 the process is in control and centred at the chart centreline, with the chart statistic approximately normal. Limits were set from the first 30 samples, so "±kσ" logic applies.

- Rule A: 1 point beyond ±3σ
- Rule B: 2 of 3 consecutive points beyond +2σ (or −2σ) on the same side
- Rule C: 4 of 5 consecutive points beyond +1σ (or −1σ) on the same side

<u>One point beyond ±3σ:</u>

$\alpha_A = P(|Z| > 3) = 2\,P(|Z| > 3) = 2\,(0.00135) \approx 0.0027$

<u>2 of 3 beyond 2σ:</u>

Let $p = P(Z > 2) \approx 0.0228$
$\alpha_B = 2\,[\binom{3}{2}\,p^2(1-p) + p^3] \approx 0.0031$

<u>4 of 5 beyond 1σ:</u>

Let $q = P(Z > 1) = 0.1587$

$\alpha_C = 2\,[\binom{5}{4}\,q^4(1-q) + q^5] \approx 0.0055$

<u>The false-alarm (Type I) rates are approximately:</u>

- Rule A ≈ 0.0027 per point
- Rule B ≈ 0.0031 per 3-point window
- Rule C ≈ 0.0055 per 5-point window.

These are theoretical rates derived from standard normal tail probabilities.

2. Estimate the likelihood of making type II

The likelihood of making a Type II error (β) is 0.868 (86.8%), as represented by the orange-shaded area in the probability density function. This means that even though the true process mean has shifted from 25.05 L to 25.03 L, there is still an 86.8% chance that the sample mean will fall within the existing control limits (UCL = 25.089 L, LCL = 25.011 L). Consequently, the control chart would fail to signal that the process is out of control.

This high β value indicates that the chart is not sensitive enough to detect such a small shift in the mean. In practice, the process could drift without triggering any alarms, highlighting the need for either narrower control limits or a larger subgroup size to improve detection power.

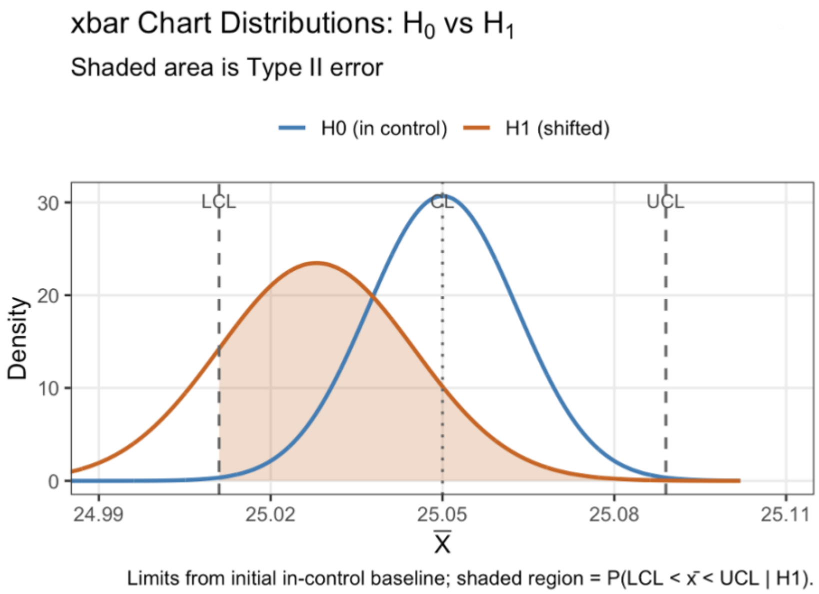

Figure 4: Type I and Type II error probability illustration

Part 5:

The timeToServe.csv dataset analysis shows clearly that when more baristas are on duty, service efficiency increases. The average service time when there is only one barista is about 200 seconds, which is significantly longer than the acceptable service standard and leads to poor reliability. The average time decreases significantly to roughly 100 seconds when there are two baristas present, and it further decreases to 30 to 50 seconds when there are three or more baristas. When two or more baristas are working, reliability increases from 0% with a single barista to nearly 100%, according to the statistics, which uses a 120-second service target as the definition of "reliable service." Accordingly, a minimum of two baristas are required to guarantee prompt and reliable customer service.

To evaluate profitability, a simple model was built that includes both revenue (R30 profit per customer) and personnel cost (R1 000 per barista per day). Using the dataset's implied average daily demand of roughly 550 orders, two baristas can meet the entire demand during an eight-hour day without service delays. Adding more baristas only increases labour cost without generating extra revenue, since demand becomes the limiting factor.

Therefore, having two baristas on duty each day is the ideal staffing strategy for all weekdays. This setup strikes a mix between cost, speed, and dependability, resulting in an estimated R14 400 daily profit and guaranteeing that almost all clients receive prompt service. Two baristas provide the highest operating efficiency in the present, but staffing can be reassessed if demand rises in the future.

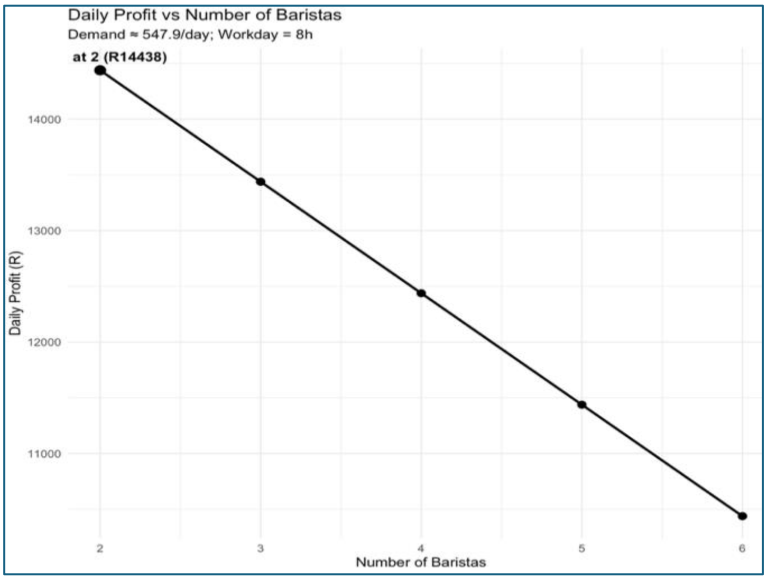

Figure 5.1: Service time vs number of baristas (Shop A)

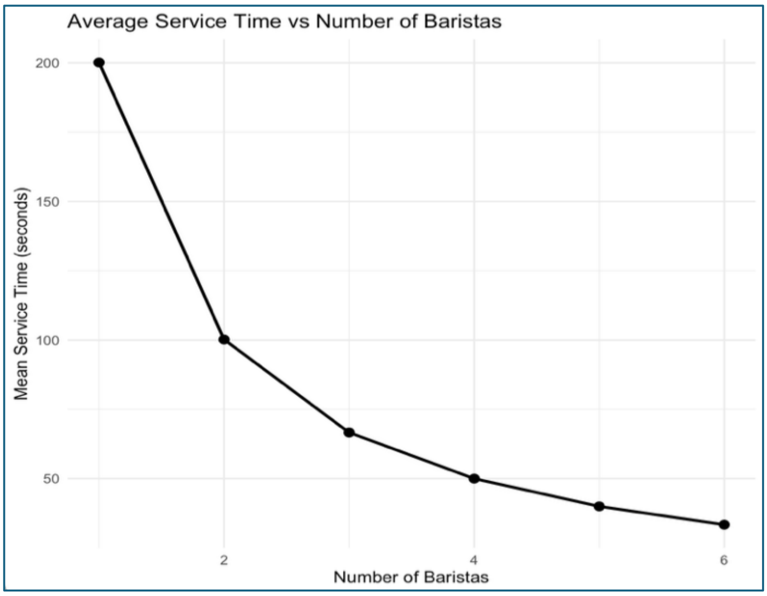

Figure 5.2: Reliability curve by barista count (Shop A)

timeToServe2 database:

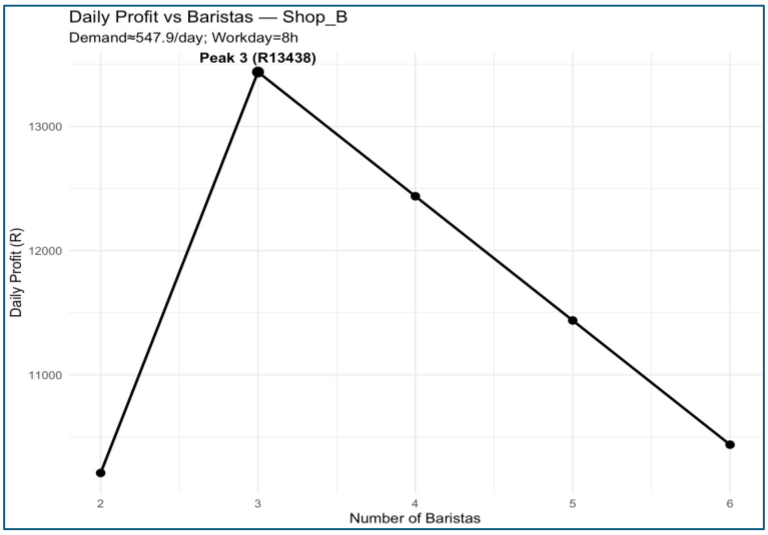

Figure 5.3: Service time vs number of baristas (Shop B)

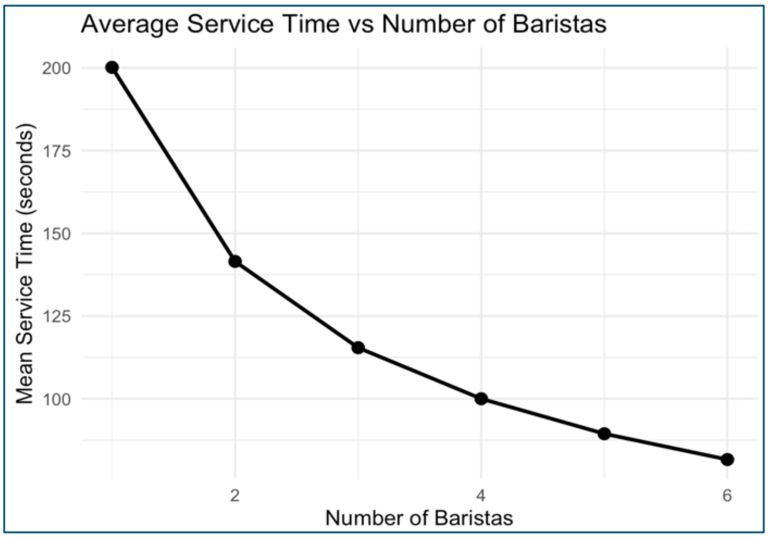

Figure 5.4: Reliability curve by barista count (Shop B)

Shop A achieves its best performance with two baristas because service times are fast enough (≈ 100 s) to meet daily demand (550 orders) during an eight-hour shift. In contrast, Shop B's service times are slower (≈ 140 s for two baristas), meaning that two staff members cannot serve the same number of customers efficiently. It therefore requires a third barista to reach the same throughput level and maintain reliable service. The extra barista raises labour cost slightly but prevents lost sales, which keeps overall profit higher at that staffing level.

Shop A has a faster preparation times which achieves its optimal profit and reliability with only two baristas. Shop B's slower workflow means two baristas cannot serve the full daily demand efficiently, causing longer queues and lower reliability. Consequently, Shop B's optimal staffing shifts to three baristas to balance speed and labour cost. These results demonstrate that each outlet's operational efficiency determines its ideal staffing: faster processes optimise at lower labour levels, while slower processes require more personnel to maintain service quality and profitability.

Taguchi Loss Interpretation:

In terms of quality engineering, the Taguchi loss function can be used to explain the losses resulting from delayed or unreliable service. This function indicates that, even if the system stays within specification boundaries each deviation from the goal performance results in a proportional economic loss. When service times beyond the optimal goal (120 s), in this instance, customer discontent and opportunity cost progressively rise. The above profit model reflects this continuous-loss principle: rather than failing suddenly, every extra delay or inefficient staffing level causes a smaller financial loss. Therefore, keeping the number of baristas as near to the aim as possible minimizes the overall "loss to society" which is consistent with Taguchi's concept.

Part 6:

1. Overview:

In order to determine whether there are statistically significant variations in delivery hours between product kinds and between 2022 and 2023, this section will conduct a Design of Experiments (DOE) and Analysis of Variance (ANOVA).
The descriptive statistics from Part 3 showed that delivery times varied significantly among product categories, but the annual averages seemed to be fairly constant. Because there is only one dependent variable (Delivery Hours), an ANOVA was chosen over a MANOVA based on these initial findings.

2. Hypotheses:

The hypotheses tested were as follows:

- H_0: There is no significant difference in mean delivery hours between 2022 and 2023, and no significant difference between product types.
- H_1: There is a significant difference in mean delivery hours between years and/or between product types.

3. Experimental Design and Data Selection:

The dependent variable was Delivery Hours, and the independent variables were Year (2 levels: 2022, 2023) and Product Type (6 levels: Mouse, Keyboard, Laptop, Monitor, Software, Cloud). Each observation represents one completed delivery transaction.

A two-way ANOVA model was used:

DeliveryHours = Year + ProductType + (Year × ProductType) + Error

4. ANOVA Results:

Source of Variation	Sum Sq	Df	Mean Sq	F value	p value	Partial η^2
Year	82	1	82.0	2.79	0.0947	0.00003
Product Type	7 022 849	5	1 404 569.8	47 696.76	$< 2 \times 10^{-16}$	0.705
Year × Product Type	273	5	54.6	1.85	0.0991	0.00009
Residuals	2 944 438	99 988	29.45	—	—	0.50

Table 10: Two-way ANOVA results for delivery hours

5. Interpretation:

Delivery hours are significantly impacted by Product Type, according to the ANOVA results ($p < 0.001$). The partial $\eta^2 = 0.705$ shows that product type changes account for around 70% of the variation in delivery time, which is a relatively substantial effect size. This implies that delivery times for some product categories are consistently longer than those for others.

There is no compelling evidence of a shift in delivery timeframes between 2022 and 2023, as the Year main impact is not significant ($p = 0.0947$). Additionally, there is no significant interaction between Year and Product Type ($p = 0.0991$), suggesting that all product categories exhibited comparable behavior in both years.

6. Visual Results:

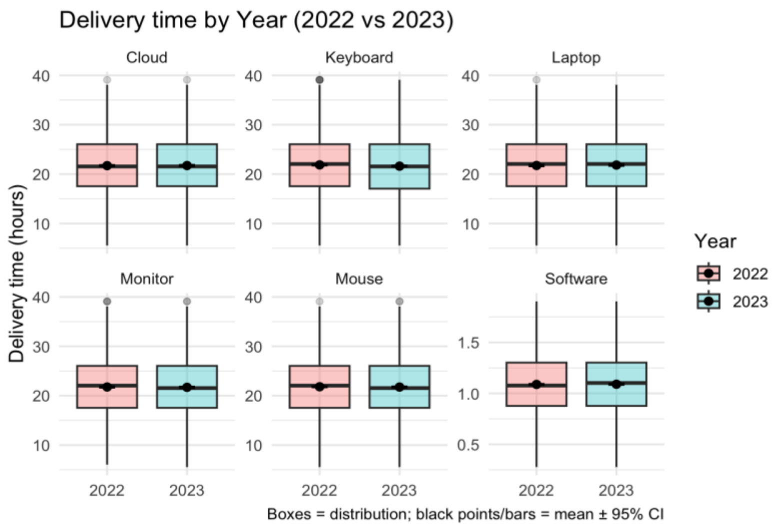

Figure 6.1: Boxplot of Delivery Hours by Product Type

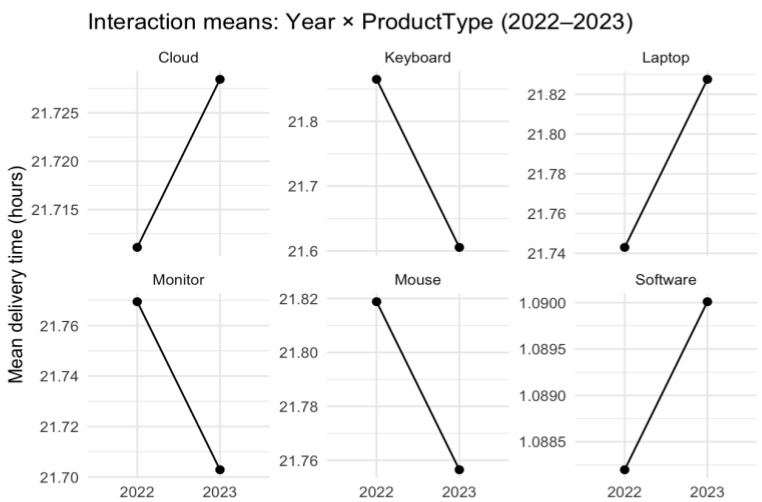

Figure 6.2: Interaction Plot (Year × Product Type)

Figure 6.3: Monthly Delivery Trend for a Selected Product

7. Discussion and Conclusion:

Delivery hours differ considerably by product category but not by year, according to the ANOVA. This indicates that while operational performance remained stable from 2022 to 2023, delivery efficiency is significantly impacted by the product's type. From a managerial perspective, rather than focusing on general year-to-year goals, efforts to shorten delivery times should concentrate on the particular product categories that exhibit the highest means (such as software or major hardware goods). ANOVA was enough because just one dependent variable was examined, MANOVA would have been necessary only if several response metrics (such as cost, delay, and satisfaction) had been examined concurrently.

Part 7

1. Estimating reliable service days

The service is considered reliable when at least 15 staff members are on duty. From the 397 recorded days there were 96 days with 15 workers and 270 days with 16, giving 366 reliable days out of 397.

$\frac{366}{397} \times 365 = 336.4987 \approx 337$ reliable days per year

Therefore, the company can expect reliable service on approximately 337 days per year, with occasional disruptions caused by under-staffing.

2. Optimising company profit

To determine the most profitable staffing level, the data were modelled as a binomial attendance problem.
Each scheduled employee has an estimated attendance probability of 0.974, based on observed daily counts (mean ≈ 15.6 with a maximum of 16 staff).
Service issues occur when fewer than 15 employees are present, which on average costs the company R20 000 in lost sales per day.
Each additional employee costs R25 000 per month (R300 000 per year).

The expected profit was calculated for several staffing levels by combining the cost of additional staff with the expected loss from unreliable days. The results are illustrated in the figure below.

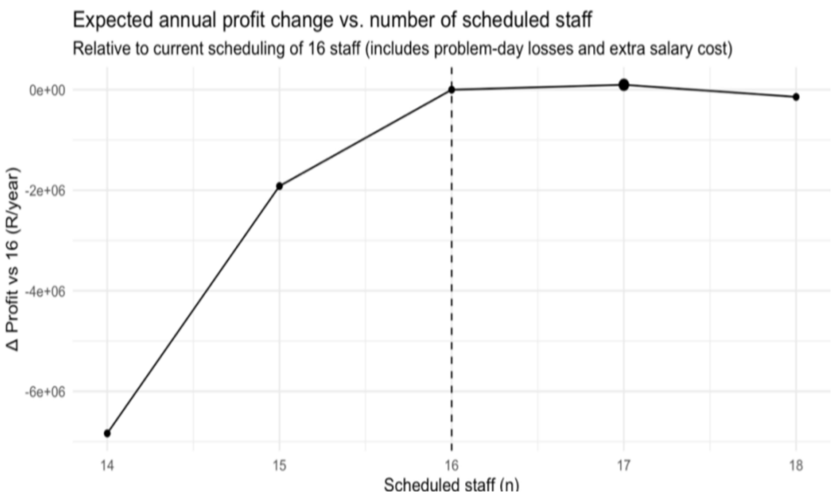

Figure 7: Profit vs staffing level curve

From the curve it is evident that:

- Having fewer than 15 staff results in large annual losses (≈ R 6 million for 14 staff).
- Profit increases sharply up to about 16–17 staff, where reliability improves significantly.
- Beyond 17, additional salary costs outweigh the small reduction in problem days.

3. Recommendation:

Hire one additional staff member thus increasing the daily schedule from 16 to 17 employees.
This raises expected reliability to around 362 days per year and improves annual profit by approximately R 98 000 compared to the current situation.
Further hiring beyond 17 offers negligible financial benefit.

Overall conclusion:

This project combines data wrangling, statistical analysis, and optimization across all analyses to create a thorough framework for quality assurance. The SPC analysis confirmed the stability of key processes, while the descriptive data served as a basis for comprehending sales behavior. The discovery of software process instability emphasizes the significance of ongoing observation and focused enhancement.

The control charts' theoretical dependability was validated by the Type I and Type II error evaluations, and the profit-optimization models showed how analytical insights can inform practical operational choices. These results were linked to the more general idea that performance discrepancies result in ongoing monetary and social losses by the Taguchi loss interpretation. The reliability and ANOVA sections provided additional examples of how probabilistic modeling and experimental design aid in service optimization.

In conclusion, this report reflects my ability to apply engineering reasoning, quantitative analysis, and evidence-based decision making to real operational data. By evaluating and improving process quality, performance stability, and profitability, these methods showed that the ECSA GA4 learning objectives had been fully met.

References:

Department of Industrial Engineering, Stellenbosch University. 2025. *QA344 Statistics and Quality Assurance Notes*. Stellenbosch: STEMLearn Course Pack.

MANOVA test in R: Multivariate analysis of variance (no date a) STHDA. Available at: http://www.sthda.com/english/wiki/manova-test-in-r-multivariate-analysis-of-variance (Accessed: 18 October 2025).

Shreffler, J. (2023) *Type I and type II errors and statistical power*, *StatPearls*. Available at: https://www.ncbi.nlm.nih.gov/books/NBK557530/#:~:text=A%20type%20I%20error%20occurs,not%20or%20having%20false%20positives (Accessed: 15 October 2025).