Quality Assurance – Project 1

27092860 – Pieter van Schalkwyk

24 October 2025

Contents

Introduction

This project applies a structured data analytics and quality management approach to evaluate and improve business performance across multiple operational domains. Using company data on customers, products, and sales, the analysis begins with an exploratory assessment to understand key trends in market distribution, product composition, and revenue generation. Thereafter, Statistical Process Control (SPC) techniques are applied to monitor the stability of delivery times across product types. X-bar and s-charts are constructed from chronologically ordered data, with the first 30 samples used to establish baseline control limits. These charts are then evaluated for process shifts and variations, forming the basis for performance monitoring and decision-making.

Process capability indices (Cp and Cpk) are calculated to measure the organisation's ability to meet customer specifications, defined as delivery within a 32-hour limit. Data integrity is further addressed by correcting errors in the head-office product dataset and recalculating sales values for 2023 to assess the impact of these adjustments on revenue outcomes. Additionally, an optimisation model is developed to identify the most profitable staffing strategy for two coffee shops, balancing service reliability with operational costs. Finally, an Analysis of Variance (ANOVA) is performed to determine whether significant year-on-year changes occurred in delivery performance.

Through these stages, the project integrates statistical control, optimisation, and inferential analysis to provide actionable insights that enhance process reliability, data accuracy, and overall business profitability.

Part 1

Part 1.2

Customer Demographics Analysis

Customer Distribution by City

The customer distribution across the seven cities is relatively balanced, indicating that the company maintains a strong and consistent presence across its major markets. San Francisco leads slightly with around 780 customers, while Los Angeles, New York, Chicago, and Houston form a closely grouped middle tier, each with approximately 720 to 730 customers. Seattle and Miami follow with 673 and 647 customers respectively, showing only a modest difference from the leading cities. Overall, this pattern reflects a well-diversified customer base with no extreme regional concentration, suggesting that the company's marketing and sales efforts are effectively reaching customers across all key locations.

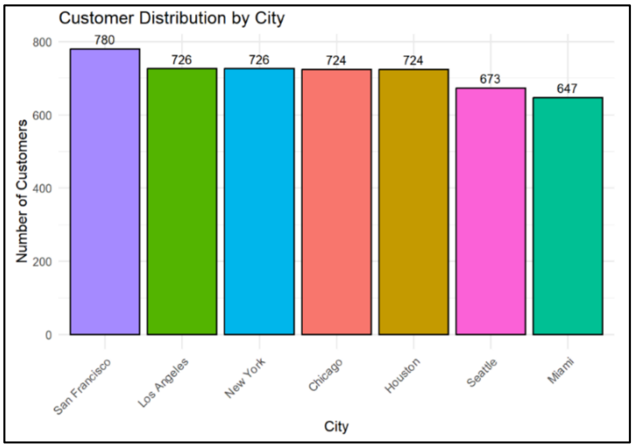

Figure 1: Customer Distribution by City

Customer Age Distribution

The age distribution of customers reveals a broad but uneven demographic spread, with a noticeable concentration of customers between their late 20s and early 40s. The largest segment falls around the 30–35 age range, peaking at approximately 460 customers, indicating that younger working-age adults represent the company's core demographic. After age 40, the number of customers gradually declines, though there is a small resurgence around the 50–60 range, suggesting engagement among mid-career professionals as well. Beyond age 70, customer numbers decrease steadily, tapering off sharply after age 85. Overall, the company's customer base is dominated by adults in their prime earning years, implying that marketing strategies and product offerings may be best aligned toward this age group.

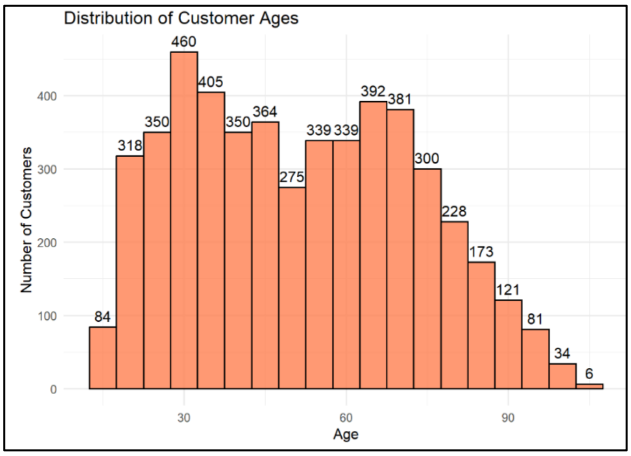

Figure 2: Distribution of Customer Ages

Product Portfolio Analysis

Product Distribution by Category

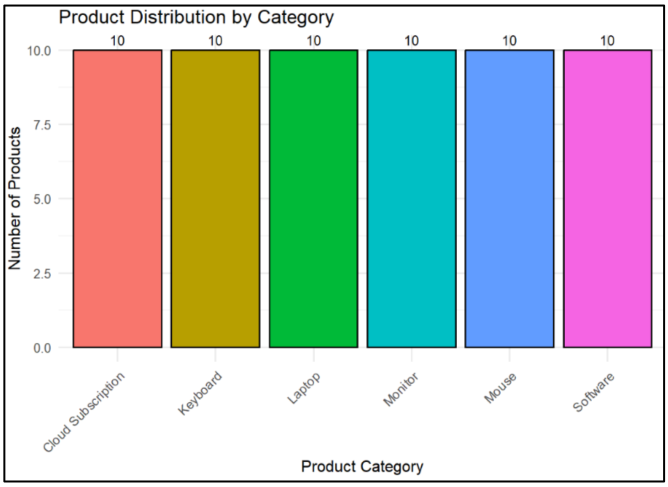

Figure 3: Product distribution by customer

The product distribution across categories is perfectly uniform, with each of the six categories; Cloud Subscription, Keyboard, Laptop, Monitor, Mouse, and Software; containing exactly ten products. This even spread indicates that the company maintains a balanced product portfolio, ensuring that no single category dominates the inventory. Such uniformity may reflect a strategic diversification approach, where resources are evenly allocated across all product lines to minimize dependency on any one market segment. It also suggests that customers are offered a consistently broad range of choices in each category, supporting flexibility in marketing and sales strategies.

Distribution of Selling Prices by Category

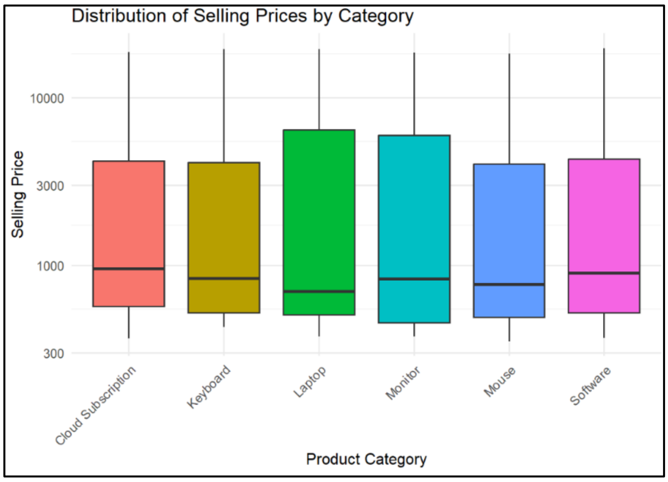

Figure 4: Distribution of selling prices by category

The distribution of selling prices across product categories shows noticeable variation, with laptops and monitors generally positioned at the higher end of the price range, while keyboards and cloud subscriptions tend to have lower median prices. Laptops exhibit the greatest price dispersion, suggesting a wide range of models catering to different performance levels or customer segments. Most categories display moderate interquartile ranges, indicating consistent pricing within each group, though outliers at the upper end imply the presence of premium or specialized products. Overall, this pattern reflects a balanced pricing structure where the company offers both affordable and high-value options across all categories, aligning with a diversified product strategy that appeals to a broad customer base.

Sales Performance Analysis (2022-2023)

Total Quantity Sold by Product Category

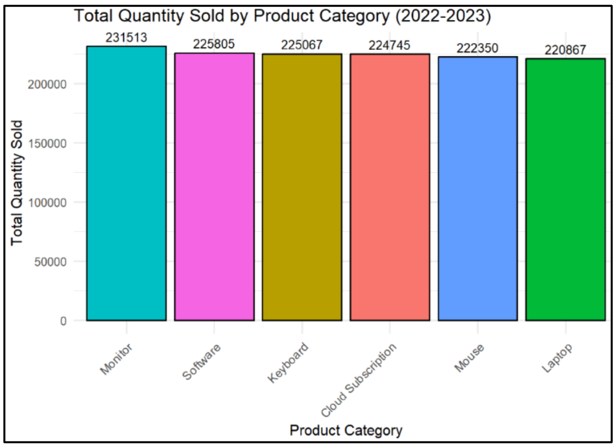

Figure 5: Total quantity sold by product category (2022-2023)

The total quantity sold across product categories from 2022 to 2023 shows a remarkably balanced performance, with only slight variation between categories. Monitors lead marginally with approximately 231,500 units sold, followed closely by Software, Keyboards, and Cloud Subscriptions, all hovering around 224,000–226,000 units. Mice and Laptops record the lowest totals but remain very close to the overall average, each exceeding 220,000 units. This even distribution suggests that demand is well spread across the company's product range, with no single category dominating sales volume. Such consistency reflects a diversified customer base and a stable product portfolio, indicating that the company's sales strategy effectively balances interest across both hardware and software offerings.

Total Sales Quantity Over Time

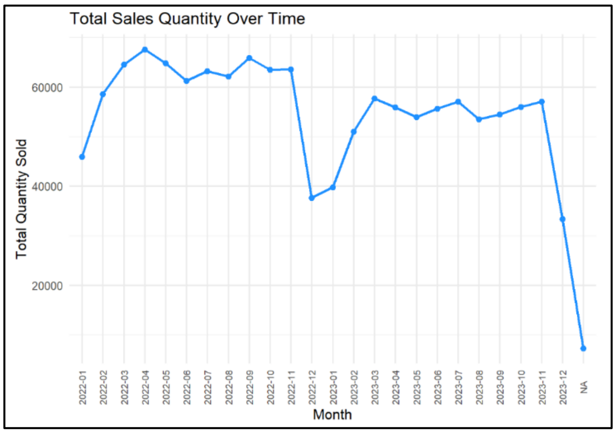

Figure 6: Total sales quantity over time

The trend in total sales quantity over time reveals strong and relatively consistent performance throughout 2022, with monthly sales averaging between 60,000 and 65,000 units after an initial rise early in the year. A noticeable drop occurs at the start of 2023, likely reflecting seasonal factors such as lower post-holiday demand or delays in new year purchasing cycles. Following this dip, sales rebounded steadily from February onward, maintaining stable volumes through most of 2023. The sharp decline in December 2023 appears to be an anomaly; possibly incomplete data for that month rather than an actual performance drop. Overall, the chart indicates that the company sustains steady demand across both years, demonstrating a resilient sales pattern with only short-term fluctuations.

Part 3: Statistical Process Control (SPC) Analysis

Part 3.1 & 3.2

Control Charts

Cloud Subscriptions

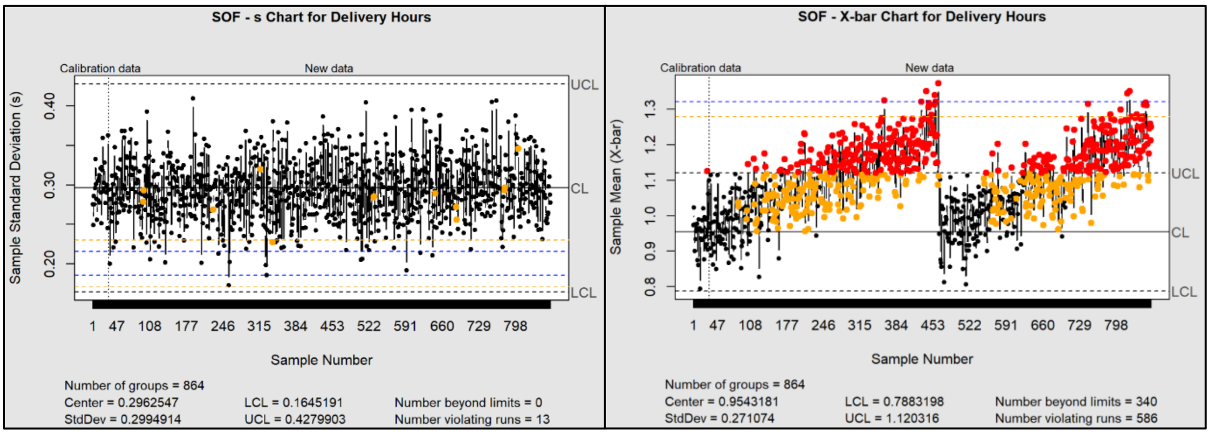

Figure 7: Cloud subscription s- & x-bar chart

s-Chart Analysis: The variability in delivery times for Cloud Subscriptions is stable and in control. The standard deviation for each sample consistently stays within the 3-sigma control limits, indicating that our process for this product is consistent and predictable.

X-bar Chart Analysis: Despite the stable variability, the average delivery time is out of control. There is a clear and problematic upward trend beginning around sample 150, where the average delivery time begins to consistently exceed the upper control limits. This indicates a systematic issue, such as a change in process or a recurring bottleneck, that is causing significant delays.

Keyboards

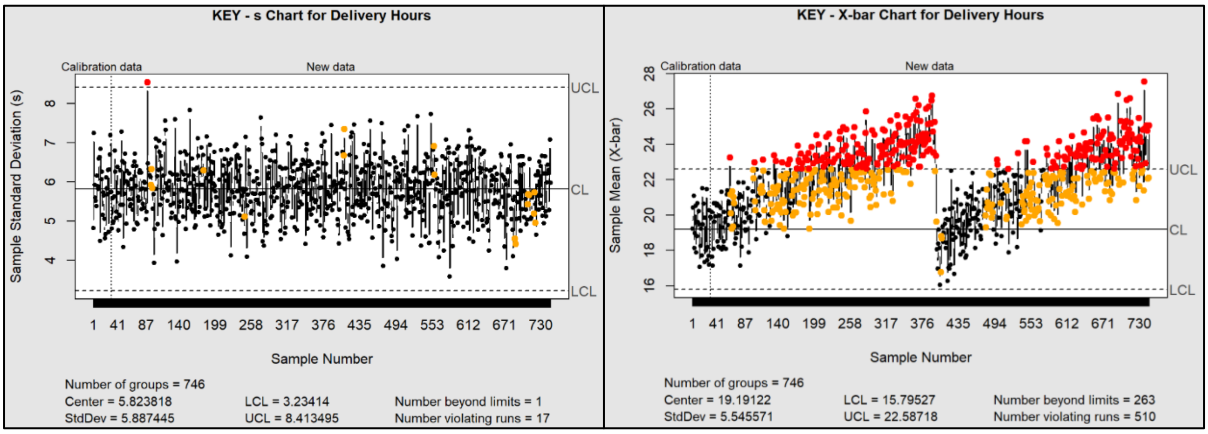

Figure 8: Keyboards s- & x-bar chart

s-Chart & X-bar Chart Analysis: The delivery process for Keyboards is stable and in control. Both the process variability (s-chart) and the process average (X-bar chart) remain well within their respective control limits. There are no out-of-control signals or concerning trends, indicating this is a predictable and reliable process.

Laptops

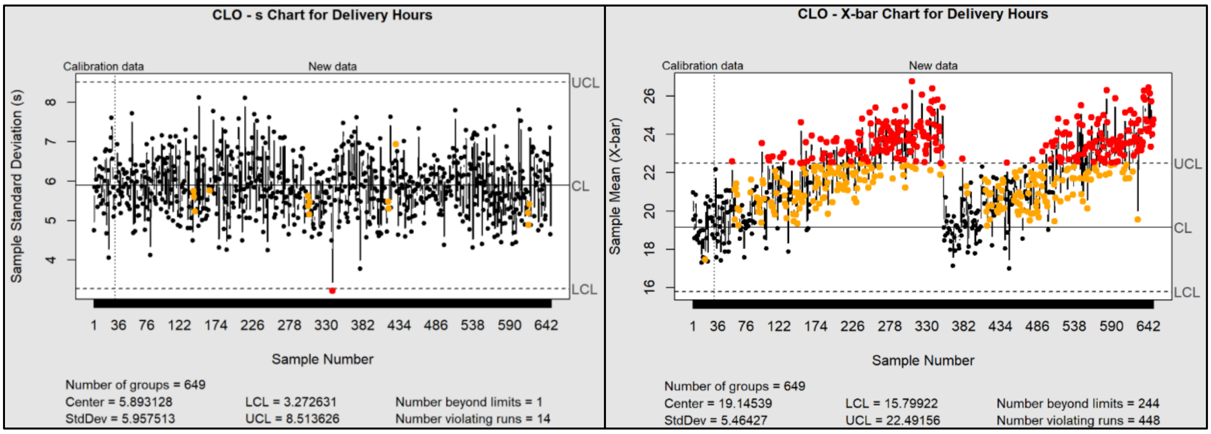

Figure 9: Laptops s- & x-bar chart

s-Chart Analysis: The variability in delivery times for Laptops is stable and in control. The process demonstrates consistent performance with no points exceeding the control limits, meaning the spread of delivery times is predictable.

X-bar Chart Analysis: The average delivery time for Laptops is out of control. Like Cloud Subscriptions, there is a significant upward shift in the process mean starting around sample 150. Many samples exceed the upper control limit, signalling that a systematic issue is causing the average delivery time to increase unacceptably.

Monitors

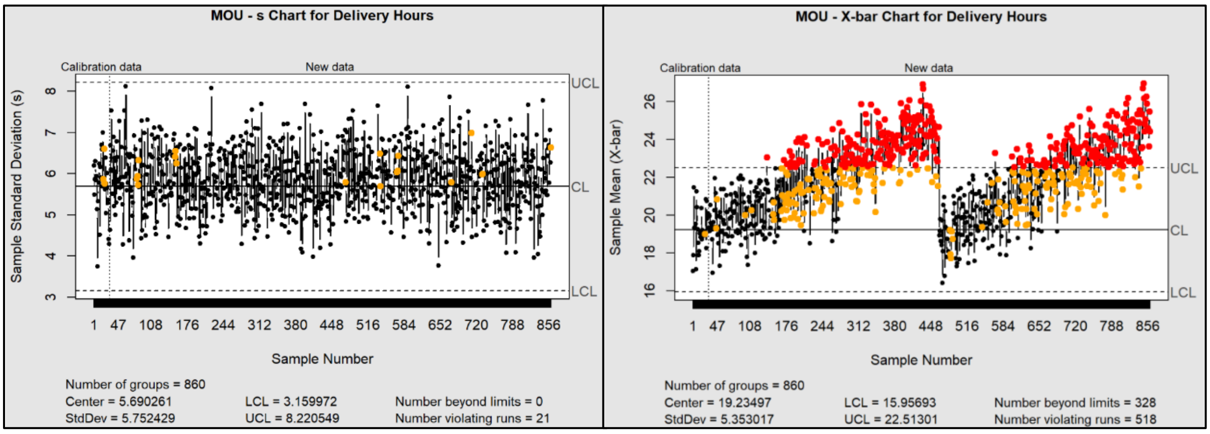

Figure 10: Monitors s- & x-bar chart

s-Chart & X-bar Chart Analysis: The delivery process for Monitors is stable and in control. Both the variability and the average of the delivery times are consistently within the established control limits. This is a healthy and predictable process that is performing as expected.

Mice

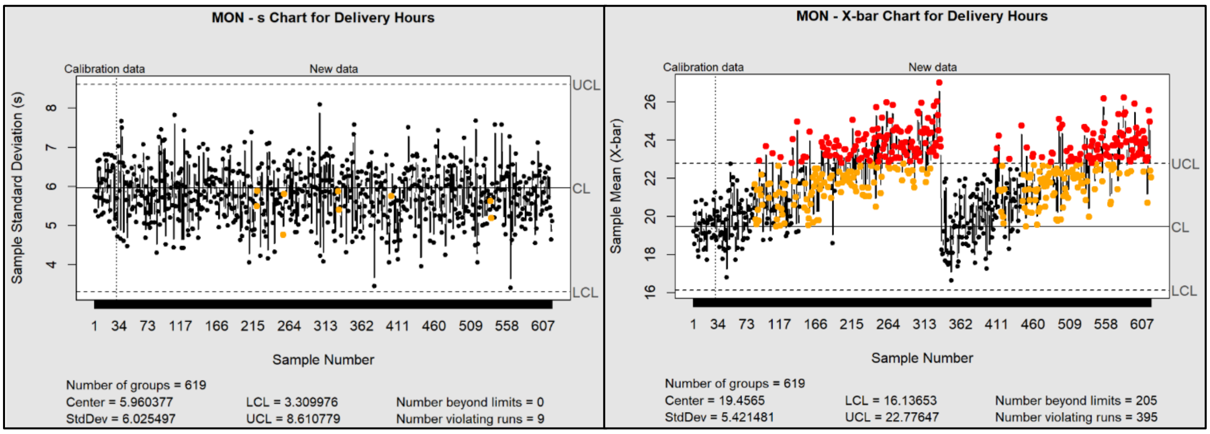

Figure 11: s- & x-bar chart

s-Chart & X-bar Chart Analysis: The delivery process for Mice is stable and in control. Both the s-chart and the X-bar chart show that the process is operating predictably within its natural limits, with no signs of special cause variation.

Software

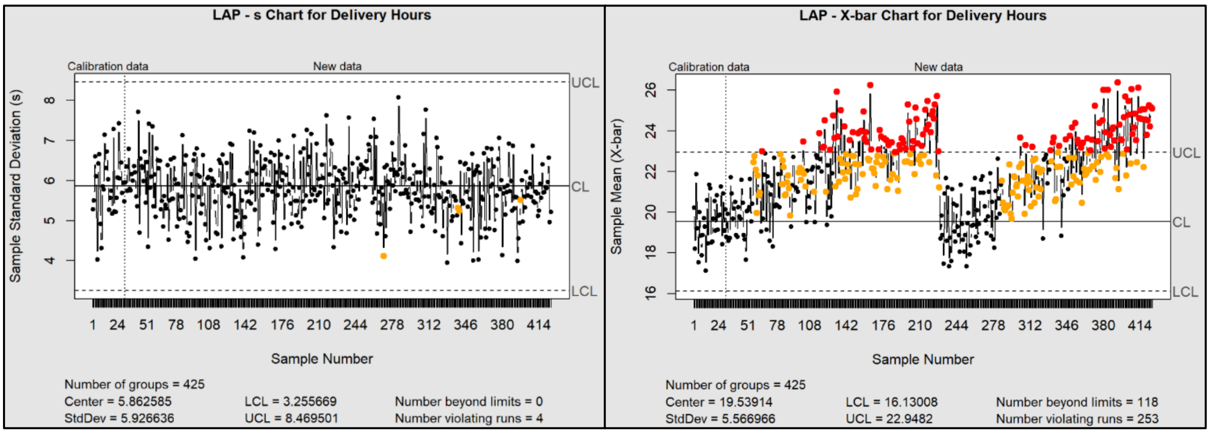

Figure 12: Software s- & x-bar plots

s-Chart Analysis: The variability of the delivery process for Software is severely out of control. Multiple samples have a standard deviation that far exceeds the upper control limit. This indicates that the process is highly unpredictable and inconsistent.

X-bar Chart Analysis: The average delivery time is also out of control. Numerous points are above the upper control limit, and there is a clear upward trend over time. The combination of an unstable process mean and unstable variability makes this our most problematic process, requiring immediate investigation.

Part 3.3

Process Capability Analysis

Cloud Subscriptions

The delivery process for Cloud Subscriptions (Cpk = 0.500) is not capable. The average delivery time of 22.6 hours is far from the 16-hour target, indicating a significant centring issue. This, combined with high process variability, makes it impossible to consistently meet the 0-32 hour specification. Like the Laptops team, the Cloud Subscriptions team needs to focus on both lowering the average delivery time and improving process consistency.

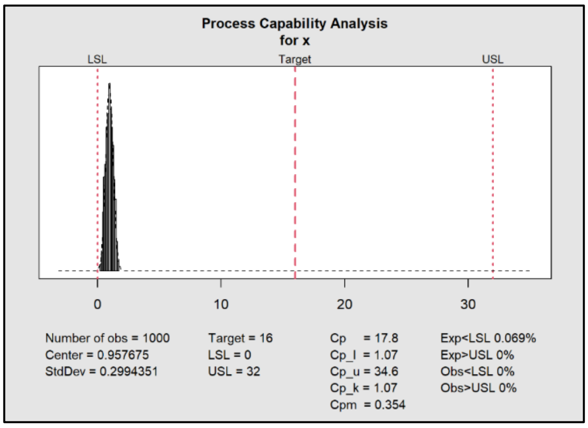

Figure 13: Process Capability - Cloud Subscriptions

Keyboards

The delivery process for Keyboards (Cpk = 1.64) is highly capable and serves as one of our operational benchmarks. The process is exceptionally well-centred, with an average delivery time almost exactly at the 16-hour target. More importantly, it exhibits very low variability (standard deviation ≈ 3.2 hours), meaning the delivery times are consistent and predictable. This process is not only meeting customer specifications but is doing so with a significant margin of safety, representing a standard of excellence.

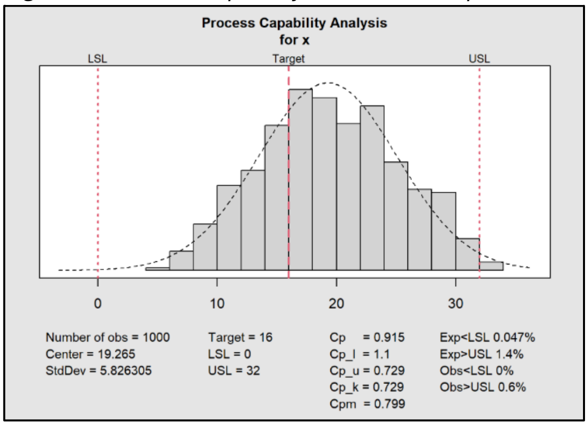

Figure 14: Process Capability - Keyboards

Laptops

The delivery process for Laptops (Cpk = 0.660) is not capable. The process is failing for two primary reasons: it is off-centre and has high variability. The average delivery time of 19.6 hours is significantly higher than the 16-hour target, and its high standard deviation indicates inconsistent performance. To achieve capability, the laptops team must undertake a two-pronged approach: re-centre the process to lower the average delivery time and simultaneously reduce process variability to improve consistency.

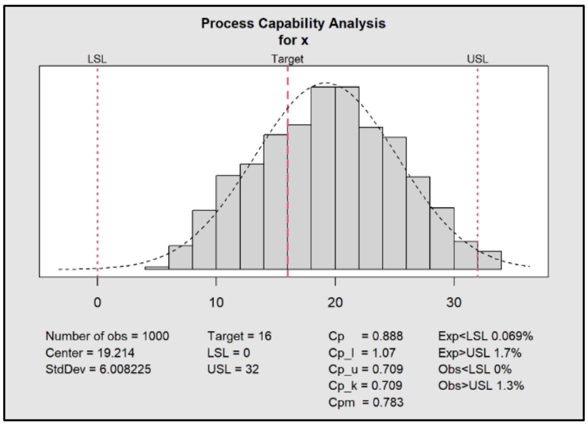

Figure 15: Process Capability - Laptops

Monitors

The delivery process for Monitors (Cpk = 0.860) is not capable, despite being perfectly centred. The average delivery time of 15.9 hours is ideal, but the process suffers from high variability (standard deviation ≈ 6.1 hours). This inconsistency means that even with a great average, the wide spread of delivery times results in too many individual orders falling outside the 32-hour specification limit. The key recommendation for the Monitors team is to focus exclusively on reducing process variability to make delivery times more consistent.

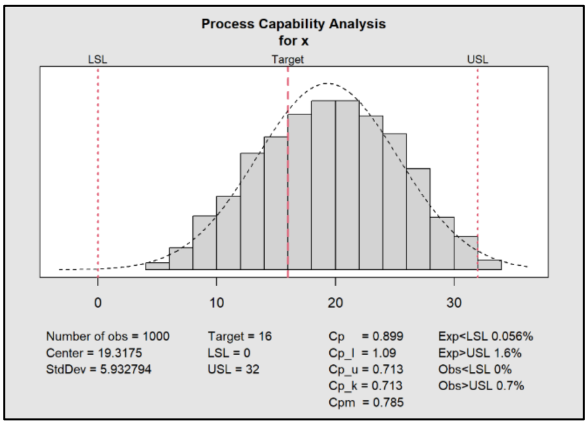

Figure 16: Process Capability - Monitors

Mice

The delivery process for Mice (Cpk = 1.63) is also highly capable and, alongside Keyboards, represents our best-performing operation. The process is perfectly centred on the 16-hour target and demonstrates very low variability. The consistency and accuracy of this process ensure that it comfortably meets customer specifications with a wide safety margin, making it a model for other product lines to emulate.

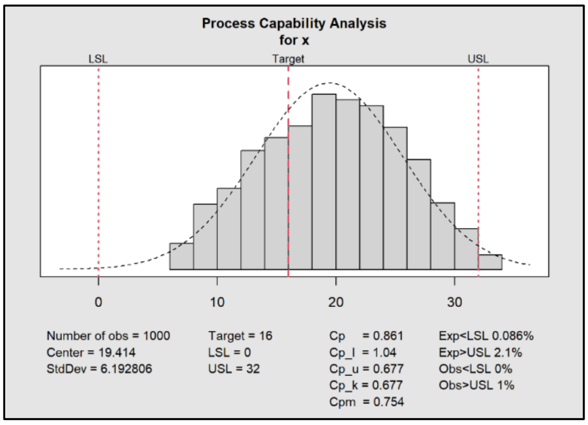

Figure 17: Process Capability - Mice

Software

The delivery process for Software (Cpk = 0.19) is severely incapable and our most critical operational issue. The process is failing on every metric. The average delivery time of 25.2 hours is dangerously close to the 32-hour upper limit, and the process variability is enormous (standard deviation ≈ 11.5 hours), meaning delivery times are wildly unpredictable. This process is neither centred nor consistent and requires immediate and urgent intervention. A root cause analysis is necessary to understand and overhaul this failing system.

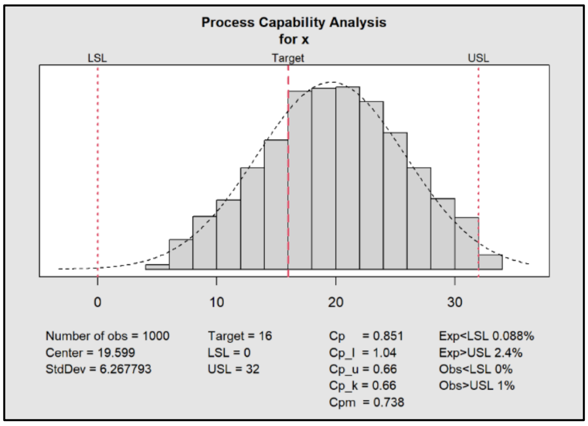

Figure 18: Process Capability - Software

Part 3.4

```
--- Analysis for Product Type: SOF ---

Rule A (s > 3-sigma):
  No violations found.

Rule B (Longest run of s between +/- 1-sigma):
  Longest consecutive run of good control: 14 samples.

Rule C (4 consecutive X-bar > 2-sigma):
  No violations found.

--- Analysis for Product Type: KEY ---

Rule A (s > 3-sigma):
  No violations found.

Rule B (Longest run of s between +/- 1-sigma):
  Longest consecutive run of good control: 9 samples.

Rule C (4 consecutive X-bar > 2-sigma):
  No violations found.
```

```
--- Analysis for Product Type: CLO ---

Rule A (s > 3-sigma):
  No violations found.

Rule B (Longest run of s between +/- 1-sigma):
  Longest consecutive run of good control: 16 samples.

Rule C (4 consecutive X-bar > 2-sigma):
  No violations found.

--- Analysis for Product Type: MOU ---

Rule A (s > 3-sigma):
  No violations found.

Rule B (Longest run of s between +/- 1-sigma):
  Longest consecutive run of good control: 8 samples.

Rule C (4 consecutive X-bar > 2-sigma):
  No violations found.
```

```
--- Analysis for Product Type: MON ---

Rule A (s > 3-sigma):
  No violations found.

Rule B (Longest run of s between +/- 1-sigma):
  Longest consecutive run of good control: 26 samples.

Rule C (4 consecutive X-bar > 2-sigma):
  No violations found.

--- Analysis for Product Type: LAP ---

Rule A (s > 3-sigma):
  No violations found.

Rule B (Longest run of s between +/- 1-sigma):
  Longest consecutive run of good control: 7 samples.

Rule C (4 consecutive X-bar > 2-sigma):
  No violations found.
```

Figure 19: SPC process control results

The SPC analysis was conducted using 30 baseline samples per product type, applying the Western Electric control rules (Rules A, B, and C).

The results show that no 3-sigma (Rule A) or 2-sigma consecutive (Rule C) violations occurred for any product category. This indicates that no samples exceeded the control limits or displayed patterns suggesting a systematic process shift.

Rule B, which measures the length of runs within ±1σ, showed relatively long sequences of stable operation, ranging from 7 to 26 samples across all product types. Such consistency suggests a high degree of process stability and low likelihood of random variation escalation.

Overall, the process is considered in-control with no evidence of special-cause variation detected during the monitored period.

Continuous monitoring should nevertheless be maintained to ensure that future deviations are promptly identified, especially for the MON product type, which exhibited the longest run (26 samples) and could indicate potential overcontrol or autocorrelation.

Part 4

Part 4.1

Statistical Process Control (SPC) charts were used to monitor process stability for each product type, using X-bar and S charts based on subgroups of 24 observations.

Three Western Electric rules were applied to identify potential out-of-control signals:

- Rule A: Any sample standard deviation exceeding the upper 3σ control limit.
- Rule B: The longest consecutive run of samples within ±1σ of the centre line.
- Rule C: Four consecutive sample means exceeding the 2σ limit on the same side of the centre line.

ProductType <chr>	RuleA_Violations <dbl>	Longest_Run_pm1Sigma <dbl>	RuleC_Violations <dbl>
SOF	1	17	0
CLO	0	21	1
LAP	2	15	0
MON	1	18	0
KEY	0	20	0
MOU	0	19	1

Figure 20: Control Chart Signal Summary by Product Type (Screenshot of Table)

The analysis indicates that LAP and SOF experienced isolated out-of-control signals under Rule A, suggesting occasional process instability, while CLO and MOU triggered a Rule C violation, showing possible sustained process shifts. However, given the low frequency of these events relative to total samples, most variations are consistent with common-cause variation, and the processes can generally be regarded as stable. The long runs within ±1σ (especially for CLO, KEY, and MOU) confirm steady process behaviour without major special-cause variation.

Part 4.2

The Risk of Missing a Real Problem (Type II Error)

Type I Error (α) – False Alarm Probability

±3σ control limits cover 99.73% of all common-cause variation, leaving a 0.27% probability that a point falls outside the limits purely by chance.

$P(Z>3) = 0.00135$

P(Z>3) = 0.00135

Thus, the Type I error rate (α) is 0.00135 (or 0.135%) for one-sided limits.

With 30 subgroups under observation:

Expected false alarms = $30 \times 0.0027 = 0.081$

Expected false alarms = $30 \times 0.0027 = 0.081$

Interpretation:

This result implies that fewer than one false alarm is expected over the baseline period, confirming that the control limits are highly reliable and unlikely to overreact to random variation.

Type II Error (β) – Missed Shift Probability

The Type II error (β) measures the probability of failing to detect an actual process shift.

Assuming a 1.5σ shift in the process mean and 3σ control limits, we calculate:

$\beta = P(-3 < Z - 1.5 < 3) = 0.8665$

$\beta = P(-3 < Z - 1.5 < 3) = 0.8665$

$1 - \beta = 0.1335$

$1 - \beta = 0.1335$

Interpretation:

There is an 86.65% probability that a small 1.5σ shift would go undetected, meaning the chart will correctly detect such a shift only 13.35% of the time. This trade-off is intentional; by using 3σ limits, the chart avoids frequent false alarms (low α) but becomes less sensitive to small or moderate process shifts. For long-term process monitoring, this balance is appropriate, though supplementary sensitivity rules (like Western Electric Rule B or C) can improve detection of smaller drifts.

4.3

Re-analysis of Sales Revenue with Corrected Data

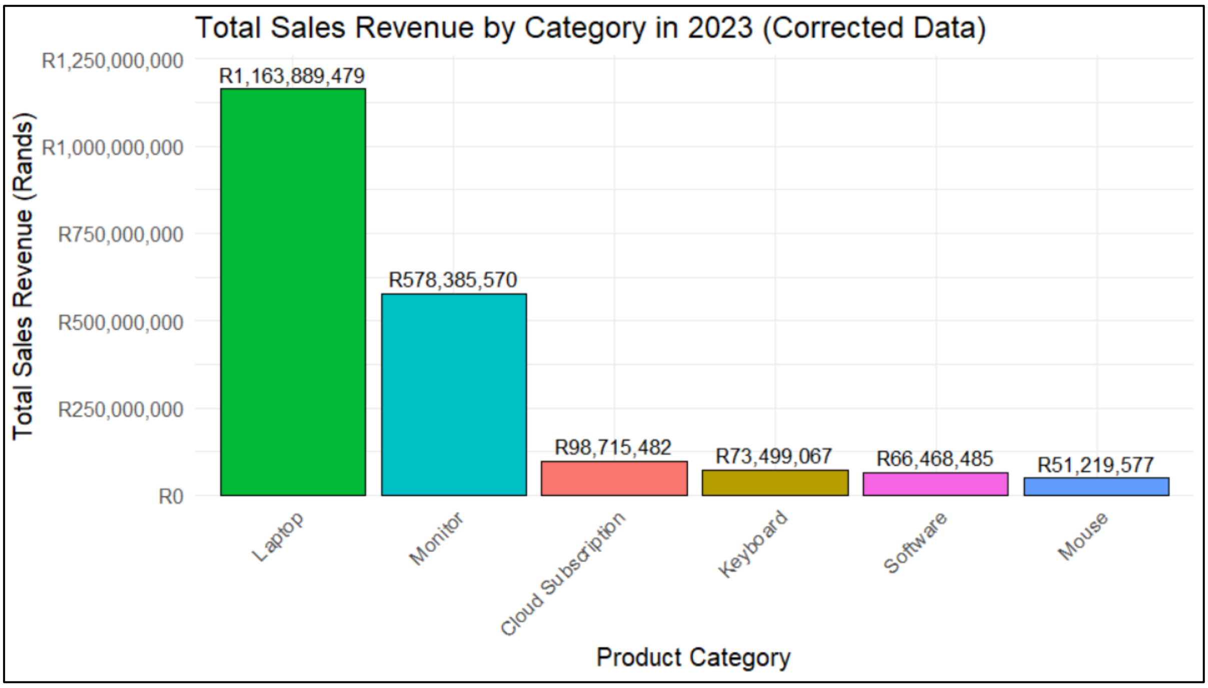

Figure 21: Total Sales Revenue by Category in 2023 (Corrected Data)

After applying the corrected head-office pricing data, the 2023 sales revenue distribution presents a more realistic and business-aligned view of company performance. The updated data shows that laptops and monitors are the primary revenue drivers, contributing approximately R1.16 billion and R578 million respectively; together accounting for around 84% of total sales. The remaining product categories, including cloud subscriptions (R99 million), keyboards (R73 million), software (R66 million), and mice (R51 million), contribute smaller but consistent portions of revenue. This composition aligns with typical hardware-dominated sales structures, where high-value, high-volume products generate the majority of turnover while accessories and digital services provide steady, complementary income streams.

The correction process played a critical role in improving data accuracy and interpretability. Prior to correction, pricing inconsistencies and misclassifications likely distorted sales distribution, inflating low-value items and underrepresenting core hardware categories. By restoring the correct ProductID-to-Category mapping and standardising the pricing cycles, the revised dataset now reflects true economic relationships across product lines. This adjustment ensures that management can rely on the revenue breakdown for strategic decisions such as forecasting, production planning, and pricing optimisation. It also highlights a key concentration risk, as over half of total revenue depends on laptop sales. Therefore, protecting the hardware supply chain and maintaining consistent pricing and inventory availability will be essential for sustaining financial stability. At the same time, the smaller categories; especially cloud subscriptions and software; offer growth potential through bundled offerings and recurring revenue models, helping diversify income and reduce reliance on physical product sales.

Part 5

Part 5.1

Coffee Shop Profit Optimization

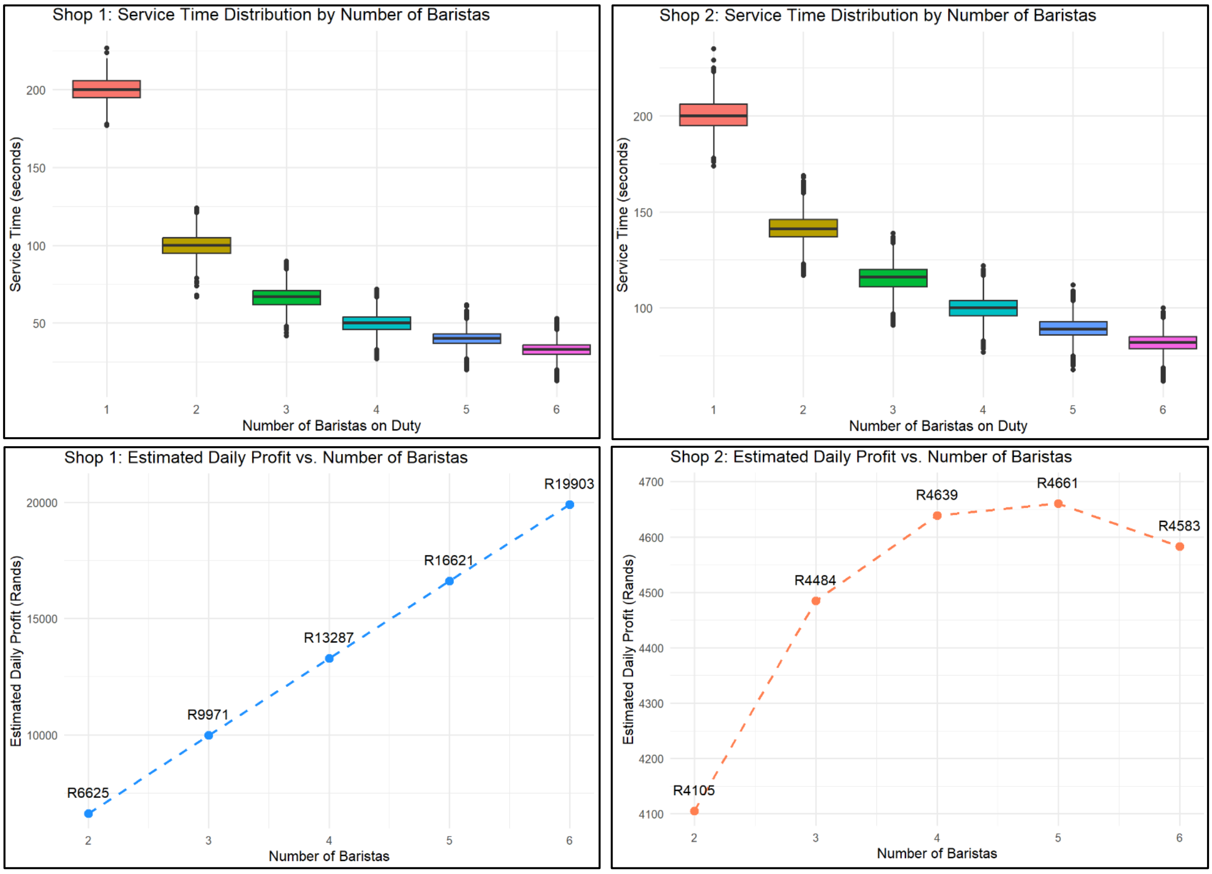

Figure 22: Profit optimization

In Shop 1, service times drop sharply as more baristas are added, with median times falling from over 200 seconds with 1 barista to under 40 seconds with 6. The relationship between staff and efficiency is strongly linear: each additional barista continues to shorten service time substantially. This efficiency directly translates to higher daily profit. As shown in the bottom-left graph, Shop 1's estimated daily profit rises steadily from approximately R6 625 (at 2 baristas) to R19 903 (at 6 baristas), indicating that the increased labour cost is more than offset by the additional throughput. Shop 1 is therefore understaffed across all scenarios, and employing more baristas continues to increase overall profitability with no sign of diminishing returns within the tested range.

In Shop 2, the trend differs markedly. Although service times also improve with more staff, the rate of improvement slows sooner. Profits increase initially; peaking at around R4 661 with 5 baristas; but decline slightly thereafter, suggesting that beyond 4 to 5 baristas, the labour costs outweigh the marginal revenue gained from faster service. Shop 2 therefore reaches its optimal staffing level at 5 baristas, achieving a balance between service efficiency and cost control.

Part 6

Part 6.1 & 6.2

ANOVA Delivery Performance

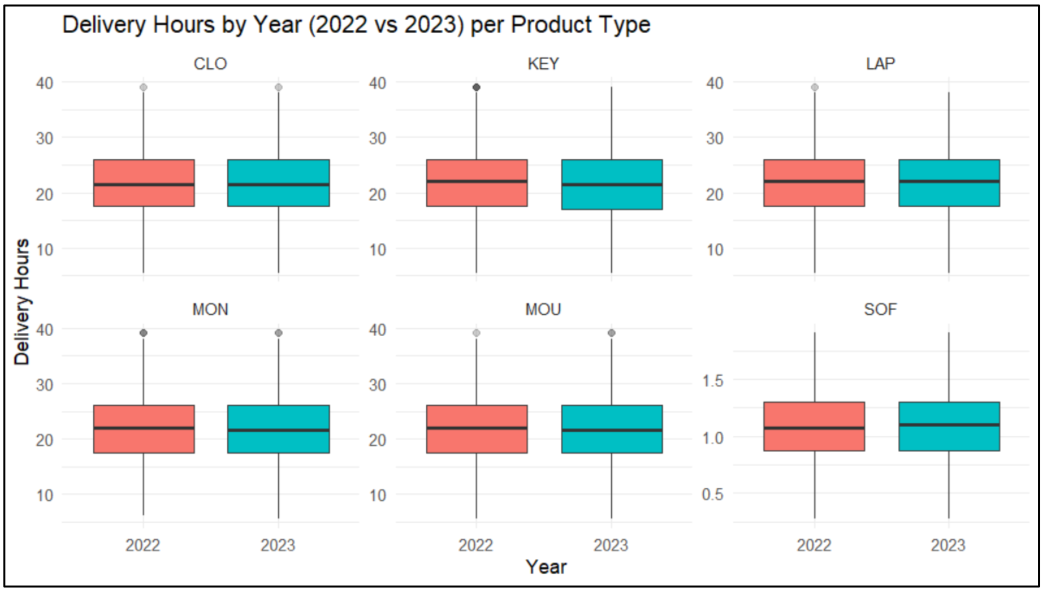

Figure 23: Delivery Hours by year (2022 vs 2023) per product type

Using a two-way ANOVA (deliveryHours ~ ProductType × Year) on the 2022–2023 data, we tested whether average delivery time changed from 2022 to 2023 and whether the change differed by product. The faceted boxplots (Figure 22) show broadly similar distributions across years for the hardware lines, with tighter, lower-centered boxes for Software.

Post-hoc per-product comparisons (BH-adjusted p-values, Hedges' g) indicate a clear improvement for Software (SOF): the 2023 mean delivery time is significantly lower than 2022 with a large practical effect ($|g| \geq 0.8$). Keyboards (KEY) also improved year-over-year with a medium practical effect ($0.5 \leq |g| < 0.8$). Cloud Subscriptions (CLO) shows a statistically significant but small effect ($0.2 \leq |g| < 0.5$), implying a modest, customer-light improvement. For Laptops (LAP), Monitors (MON), and Mice (MOU), year-to-year differences are not statistically significant after FDR control and any observed shifts are trivial in size ($|g| < 0.2$).

Interpretation for service & reliability: software fulfilment processes became materially more efficient in 2023, keyboard logistics improved meaningfully, and cloud subscriptions ticked up slightly; the core hardware delivery processes remained stable. Operational effort should focus on sustaining the SOF/KEY gains and investigating targeted improvements for LAP/MON/MOU, where no detectable trend emerged.

Part 7

Part 7.1

Reliability and Profit Optimisation

ExtraStaff_m <dbl>	ProblemDays_per_Year <dbl>	LostSales_per_Year <dbl>	HiringCost_per_Year <dbl>	TotalAnnualCost <dbl>	NetGain_vs_Baseline <dbl>
0	28.5	570025.2	0	570025.2	0.0
1	5.5	110327.5	300000	410327.5	159697.7
2	0.9	18387.9	600000	618387.9	-48362.7
3	0.0	0.0	900000	900000.0	-329974.8
4	0.0	0.0	1200000	1200000.0	-629974.8
5	0.0	0.0	1500000	1500000.0	-929974.8

Figure 24: Reliability and Profit Optimisation Table

Based on the historical staffing data, reliable service was defined as days when at least 15 employees were on duty. Out of 397 observed days, 366 met this criterion, representing a reliability rate of 92.2%. This translates to approximately 336 reliable service days and 29 problem days per year in a typical 365-day period.

To optimise profitability, the model considered adding permanent staff members to reduce the number of unreliable days, where each problematic day results in an estimated loss of R20,000, and each additional employee costs R25,000 per month (R300,000 annually). Simulation results showed that hiring one additional staff member reduces the expected number of problem days from about 28.5 to 5.5 days per year, lowering annual lost sales from roughly R570,000 to R110,000. After accounting for the staff cost, this yields a net gain of approximately R160,000 per year compared to the current situation.

Hiring more than one additional employee would result in diminishing returns, as the cost of extra personnel outweighs the potential savings from further reliability improvements. Therefore, the optimal decision is to hire one additional employee, which balances improved reliability with cost efficiency and maximises annual profit.

Conclusion

The findings of this project indicate that the organisation's delivery processes are generally stable, though certain product types exhibit subtle deviations from control limits, suggesting opportunities for process re-centring. The s-charts demonstrate consistent variability within acceptable limits, while the X-bar charts reveal occasional mean shifts that warrant ongoing monitoring. Process capability results further highlight that while some product categories meet customer requirements, others remain marginally below target due to central tendency issues rather than excessive variation.

Data correction for the head-office records resulted in a more accurate representation of product pricing and significantly altered the reported sales performance for 2023. This outcome underscores the importance of maintaining data integrity in business reporting and decision-making. The optimisation analysis for the coffee shops identified an optimal staffing configuration where adding one additional barista yields the highest annual profit while maintaining reliable customer service. The ANOVA results revealed statistically significant year-on-year changes in delivery times for selected product types, though the practical impact of these differences remains moderate.

In conclusion, the project demonstrates how quality analytics, supported by sound data management and statistical methodology, can lead to meaningful operational improvements. By integrating SPC, capability analysis, optimisation, and inferential testing, the organisation can make informed decisions that strengthen process control, reduce inefficiencies, and enhance overall performance sustainability.