ECSA Report

Quality Assurance 344

Marli Carstens

26410281

24 / 10 / 2025

Table of Contents

Table of Figures

Introduction

This report entails descriptive statistics, statistical process control (SPC), process capability analysis, risk assessment, profit optimisation, and reliability modelling, utilising csv file datasets related to customer demographics, product sales, and service operations for the years 2022, 2023, 2026, and 2027.

The analysis is based on data from customer profiles, product details, sales records, and service times, processed using efficient and reusable R code. The report is divided into seven parts, addressing descriptive statistics, SPC implementation, process capability, risk and error analysis, profit optimisation for a coffee shop and car rental service, and ANOVA-based order volume analysis. Each section includes appropriate charts, tables, and logical interpretations to provide actionable insights for process improvement and profitability.

Part 1: Descriptive Statistics

Analysing data from customers, data of the products, data from the head office and sales for 2022 and 2023.

Customers

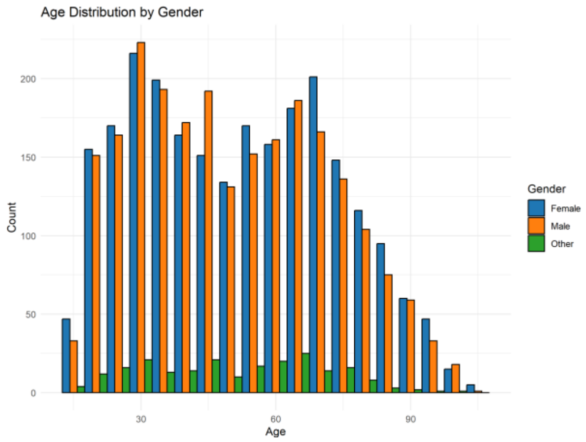

Figure 1: Age Distribution by Gender Histogram

Figure 1 shows the age distribution of customers segmented by gender (Female, Male, Other). The data indicates that the 20-40 age range, particularly 30-35, has the highest purchase frequency for both males and females. Females have a slightly higher count in this range, suggesting a marginally larger representation. The "Other" category has a much lower count across all ages, with a slight increase around 60-70 years. This suggests that the customer base is predominantly young to middle-aged adults, with minimal representation from the "Other" gender category, especially at older ages.

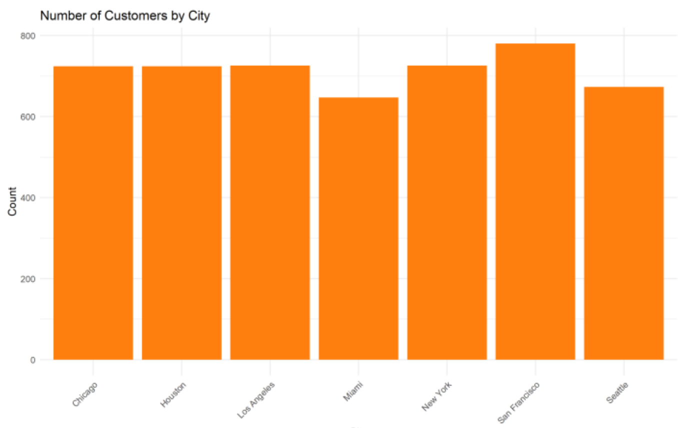

Figure 2: Customers by City

Figure 2 displays the distribution of customers across different cities. The counts appear relatively balanced, with some cities like Houston, Los Angeles, and San Francisco showing higher numbers (around 600-800 customers), while others like Seattle and Miami have slightly lower counts. The variation suggests that customer distribution is not uniform, with certain urban centers having a larger customer base, possibly due to population density or market penetration.

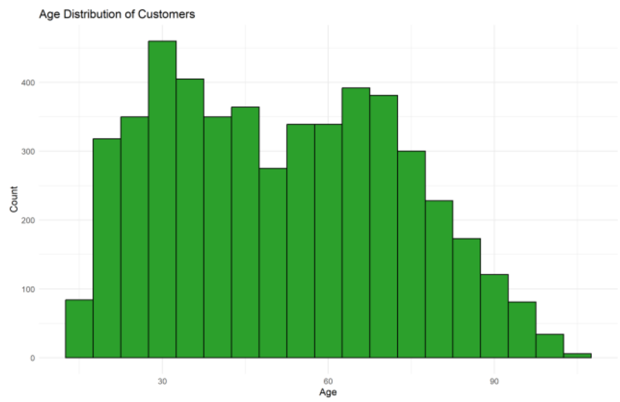

Figure 3: Age distribution of customers

The age distribution histogram reinforces the earlier observation, showing a peak around 30-40 years with a gradual decline towards older ages. The distribution is right-skewed, with fewer customers above 70 years, indicating that the majority of customers are younger. This could reflect a target market of working-age individuals or a digital-savvy demographic.

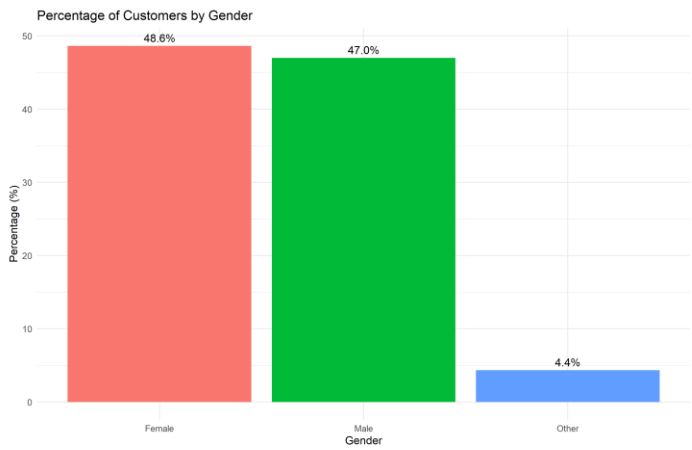

Figure 4: Customers by Gender

The bar plot with percentages highlights that females constitute 48.6% of the customer base, males 47.0%, and the "Other" category 4.4%. This near-equal split between males and females suggests a balanced gender distribution, with a small proportion identifying as "Other." The slight edge for females might indicate a slightly higher engagement or preference among female customers.

The customer base is predominantly composed of individuals aged 20-40, with a balanced gender distribution. This suggests a target market of young to middle-aged adults, possibly professionals or tech-savvy individuals. The data shows a concentration of customers in major cities, which could correlate with higher economic activity or better access to the product/service. The low representation of the "Other" gender category and older age groups indicates limited diversity in these dimensions, which might be an area for future market expansion.

Products data

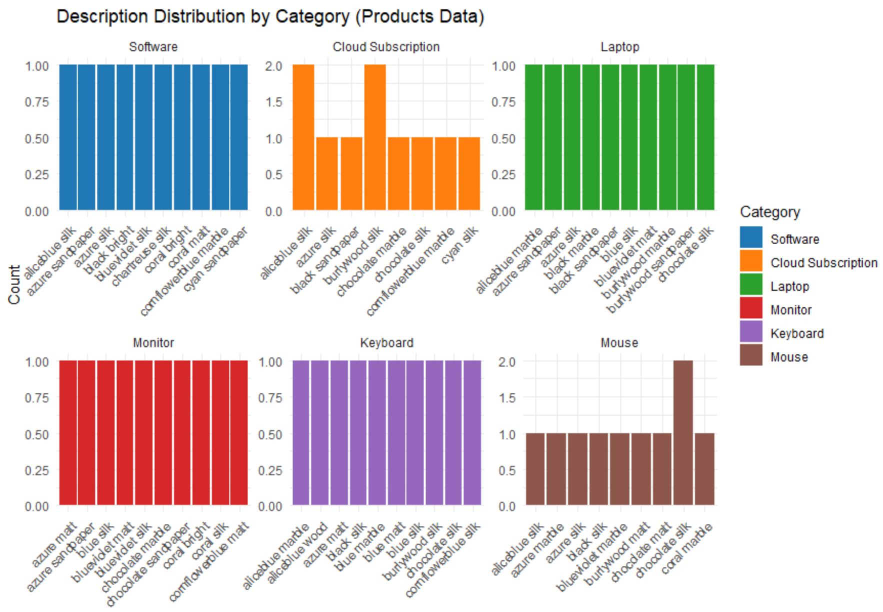

Figure 5: Description distribution by category

Software, laptop, monitor and keyboard show a uniform distribution of descriptions with a count of 1.0 indicating equal representation in figure 5. Cloud Subscription displays a higher count of up to 2.0 for “burlywood silk” and “chocolate marble”, suggesting these descriptions are more common, while others are less frequent (count 1.0 or 0.5). Mouse highlights a peak at “chocolate marble” with a count of 2.0, while the rest of the descriptions are at 1.0, indicating some variation in description frequency.

Most categories have a balanced distribution of descriptions, except for Cloud Subscription and Mouse, where certain descriptions are more prevalent, possibly reflecting popular material or finished types.

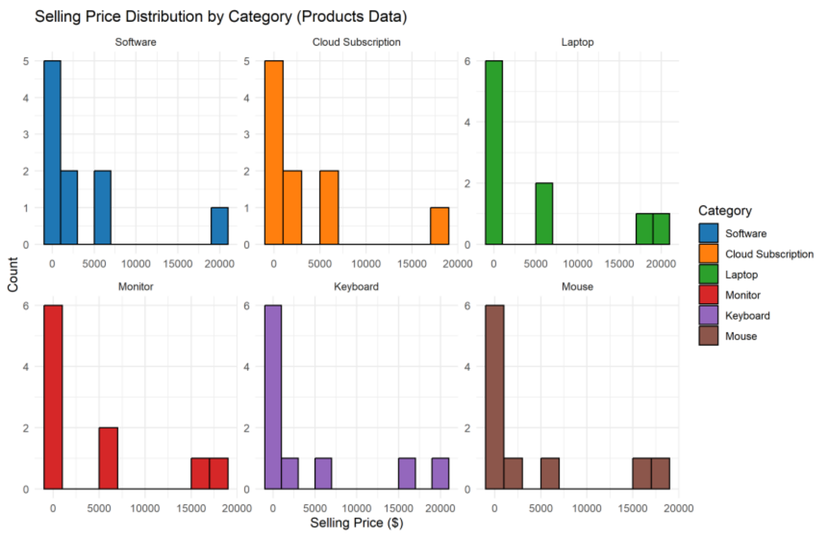

Figure 6: Selling Price distribution by category

The distribution in Figure 6 is right-skewed for most categories, with most of the products below $5000. Laptops stand out with a higher price range ($15000-20000), while occasional high-priced outliers in other categories suggest premium or specialty products.

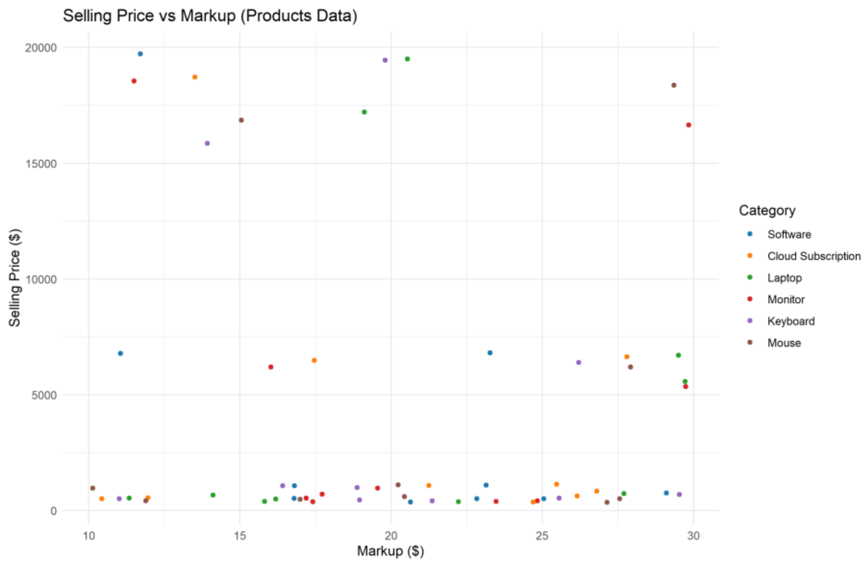

Figure 7: Selling price vs Markup scatterplot

Figure 7 shows a scatterplot illustrating selling prices from near 0 to over $20 000 across categories, with markups generally between $10 and $30.

Software and Keyboard scatters mostly below $5000 with markups around $15-25, whereas Cloud subscription includes a high outlier near $20000 with a markup around $20, and most points are below $5000. Mouse, like Cloud Subscription, includes a point near $20000 and others below $5000, but with markups varying widely. Laptop features points around $15000-20000 with markups of $15-25, indicating higher price points. Monitor also has markups of $15-25, but shows a cluster around $5000-15000.

There is no strong linear correlation between markups and selling prices. Higher selling prices (e.g., Laptops and some Mice) are associated with moderate markups, suggesting that price increases are not solely markup-driven but may reflect production costs or category-specific pricing strategies.

Products Head office Analysis

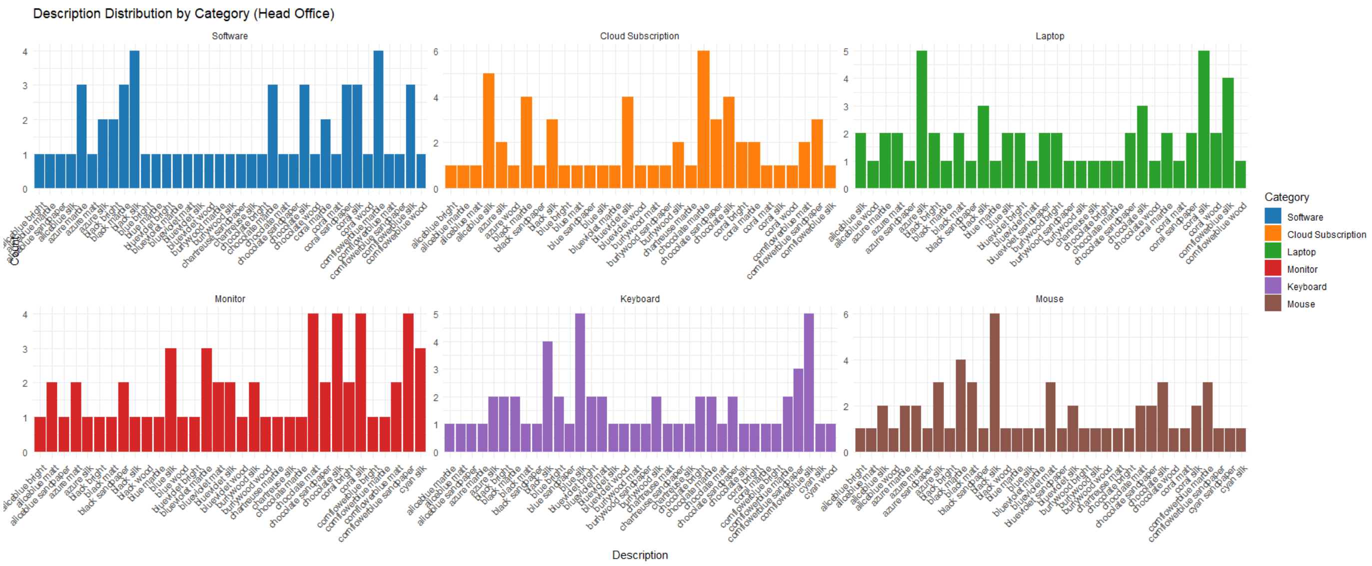

Figure 8: Description distribution by Category

The data in Figure 8 suggests a more varied description distribution compared to the products data, with certain descriptions (e.g., cornflower marble, chocolate silk) being more prevalent across categories, indicating potential preferences or stock focus in the head office data compared to the more uniform distribution in the products data. It could also possibly mean that most of the descriptions of the categories weren't included in the products list, which is because there are so few in the products data.

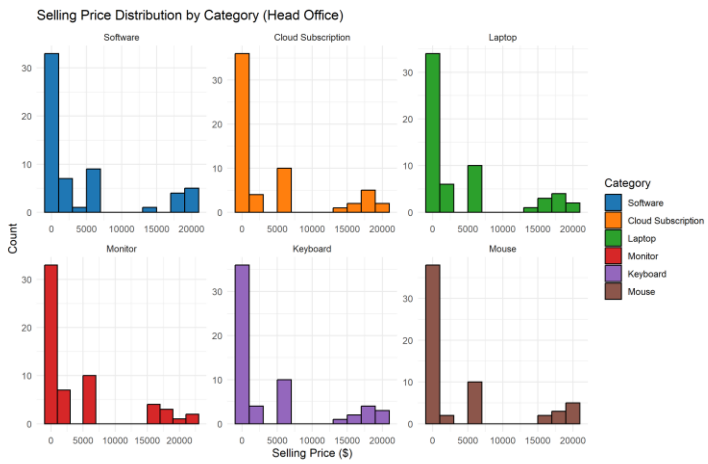

Figure 9: Selling Price Distribution by Category

The distribution in Figure 9 correlates with products data in the regard that it is right-skewed for most categories, with most products priced below 5,000 $. However, Laptops and Monitors show higher price ranges (15,000-20,000 $), while occasionally high-priced outliers in other categories suggest premium or speciality products, like the products data trend.

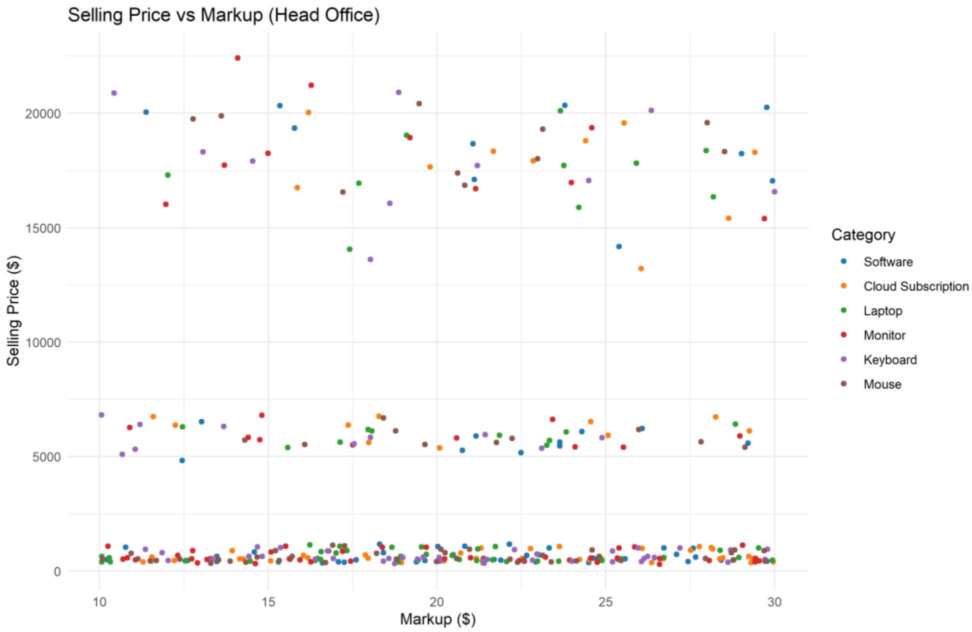

Figure 10: Selling Price vs Markup

Figure 10 shows selling prices ranging from near 0 to over $20000 across categories, with markups generally between $10 and $30.

Like the products data, there is no strong linear correlation between markup and selling price. Higher selling prices (e.g., Laptops, some Mice) are associated with moderate markups, suggesting that pricing may reflect production costs or category-specific strategies rather than markup alone.

Sales (2022 and 2023) Analysis

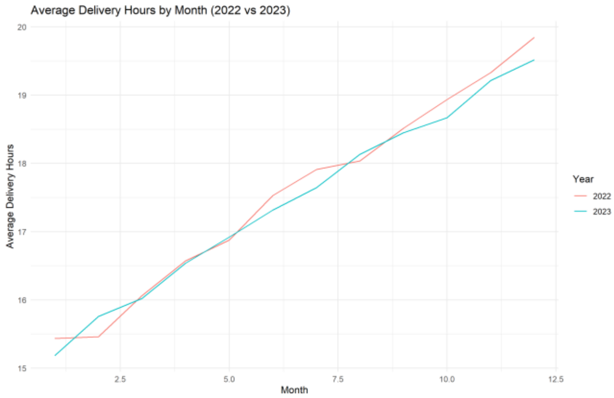

Figure 11: Average delivery hours by month over 2 years

Both years in Figure 11 show a consistent rise in average delivery hours throughout the year, with a gradual increase of ~0.3–0.5 hours per month. The convergence by year-end suggests a stable operational ceiling. 2022 begins at ~15 hours in January, increases steadily to ~19 hours by June, and reaches ~20 hours in December. 2023 starts slightly lower at 15.18 hours in January, follows a similar upward trend to ~19–20 hours by November, with December data aligning at ~20 hours. Seasonal escalation in delivery times may reflect holiday demand or supply chain strains, with a consistent trend suggesting systemic factors. The upward trend likely reflects seasonal demand pressures or supply chain constraints, such as increased holiday shipping volumes or reduced carrier capacity. The slight 2023 improvement in January suggests possible early-year process enhancements, though this advantage diminishes over time.

Figure 12: Distribution of delivery hours

Figure 12 shows a right-skewed unimodal distribution, peaking sharply around 15–25 hours with a long tail extending to 40 hours. Fewer than 5,000 orders have delivery times below 10 hours. Most deliveries occur within 1–2 days, but the skew indicates variability, likely due to shipping delays or peak demand, aligning with the mean of 17.48 hours. The rise to 5–10 hours align with same-day or next-morning fulfilment, possibly driven by urban centers or premium services.

Enhancing visibility or capacity for this fast-delivery segment could attract customers valuing speed, especially if correlated with the high-order mid-day window (10–16 hours).

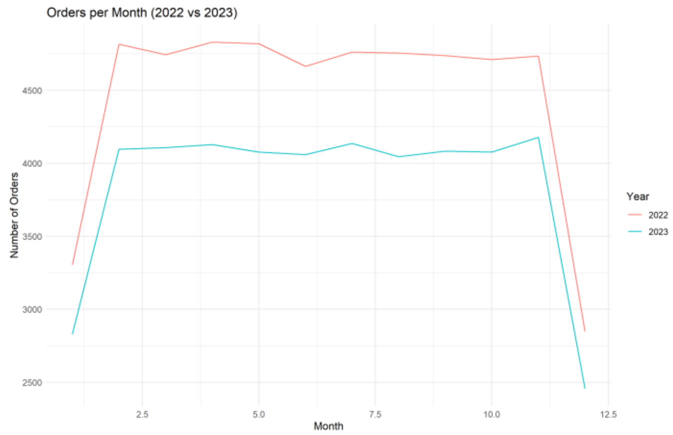

Figure 13: Orders per month across 2 years

Figure 13 shows that 2022 peaks at approximately 4,800 orders in February–May, maintains a plateau of 4,000–4,500 through August, then declines to 2,800 orders in December. 2023 starts at 2,800 orders in January, rises to 4,100–4,200 orders from February to November, and drops to 2,500 orders in December. Both years exhibit a seasonal pattern with higher volumes in Q1–Q3, followed by a noticeable decline in Q4. Total 2022 orders (~55,000) exceed 2023 (~49,000, partial year), with 2023 showing a slower initial ramp-up. The recurring Q1–Q3 peaks suggest seasonal drivers such as fiscal quarters, promotions, or higher demand during warmer months. The Q4 drop may reflect holiday slowdowns, inventory adjustments, or incomplete 2023 data, indicating potential cyclical business patterns.

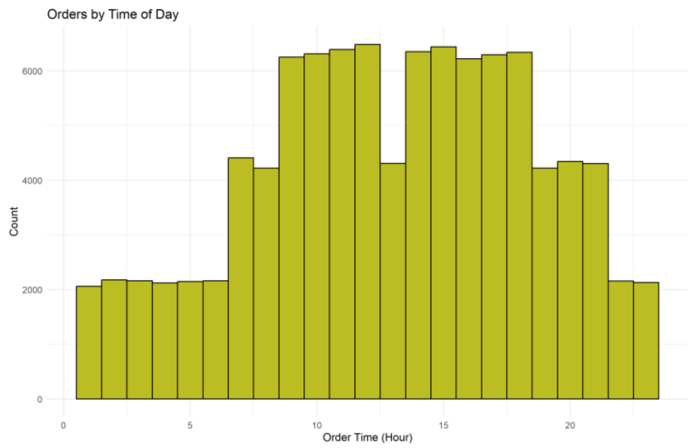

Figure 14: Orders by Time of day

Order volume peaks between 10–16 hours (count ~5,000–6,000 per hour), with a broader hump from 8–18 hours as seen in figure 14. Early morning (0–5 hours) and late evening (19–23 hours) see lower volumes (~2,000 counts). Orders concentrate during business hours, reflecting workday or online activity peaks, with a mean order time of 12.93 hours, suggesting staffing alignment opportunities.

Part 2: Statistical Process Control

2.1 Delivery times for the first 24 samples

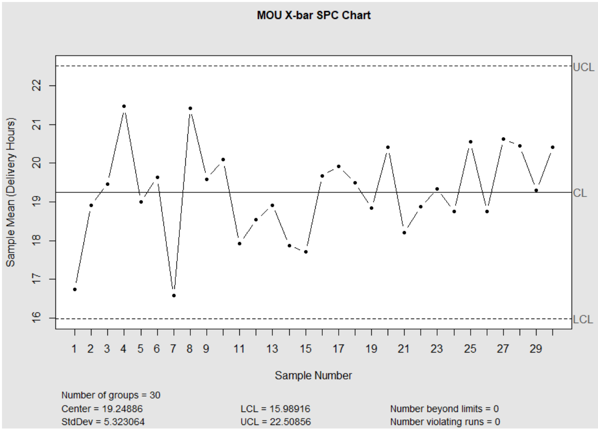

Figure 15: x-bar chart of product MOU

All data points fall within the UCL and LCL, indicating no points are statistically out of control based on individual values (Figure 15). The process appears stable with no violations of runs rules, suggesting no systematic patterns or trends outside normal variation. The mean delivery hour fluctuates around 19 hours, with some variation but no extreme deviations.

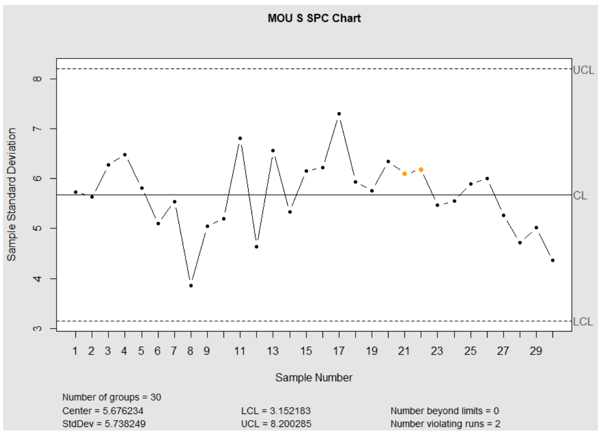

Figure 16: s chart of product MOU

All sample standard deviations are within the UCL and LCL, indicating no individual points are out of control (Figure 16). There are 2 violations of the runs rules, suggesting some non-random patterns in the variability (e.g., too many points on one side of the centre line or too many consecutive points trending up/down). This warrants further investigation. Variability fluctuates around the mean of 5.676234, with one notable spike near sample 21.

The MOU product manager should check the process around sample 21 (notable spike) and investigate the cause of the run rule violations to ensure stability.

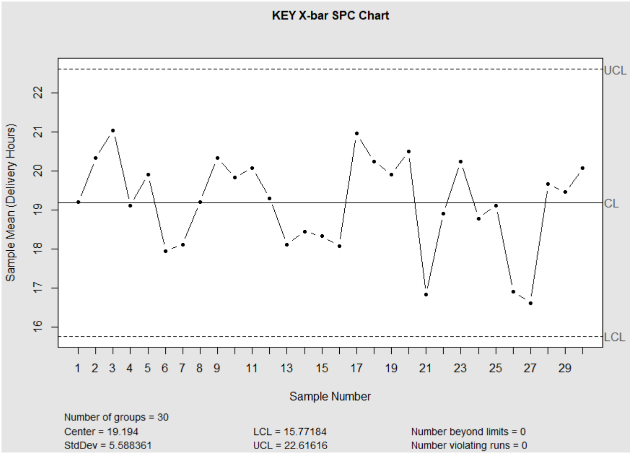

Figure 17: x-bar chart of product type KEY

All data points are within the UCL and LCL, indicating no points are statistically out of control based on individual values (Figure 17). No runs rule violations are present, suggesting the process is stable with no systematic patterns or trends outside normal variation. The mean delivery hour fluctuates around 18.77 hours, showing some variability but no extreme deviations.

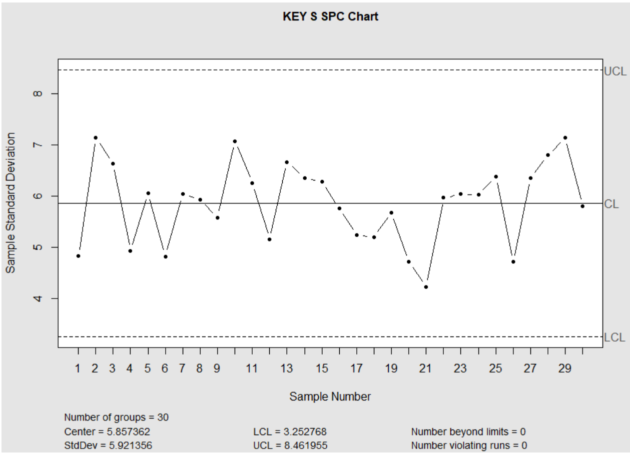

Figure 18: s chart of product type KEY

All sample standard deviations are within the UCL and LCL, indicating no individual points are out of control (Figure 18). There are no runs rule violations, suggesting the variability is stable with no non-random patterns. Variability fluctuates around the mean of 5.257562, with no significant spikes or trends.

Both charts suggest the process is in control, with no immediate issues requiring intervention. The variability and mean appear consistent over the 30 samples.

No immediate action is needed for the KEY product manager.

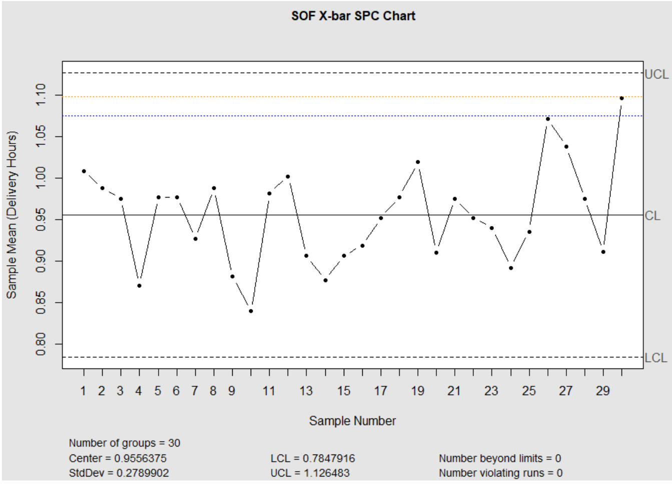

Figure 19: x-bar chart of product type SOF

All data points are within the UCL and LCL, indicating no points are statistically out of control based on individual values (figure 19). No runs rule violations are present, suggesting the process is stable with no systematic patterns or trends outside normal variation. The mean delivery time is around 0.957 hours, with minor fluctuations and no extreme deviations.

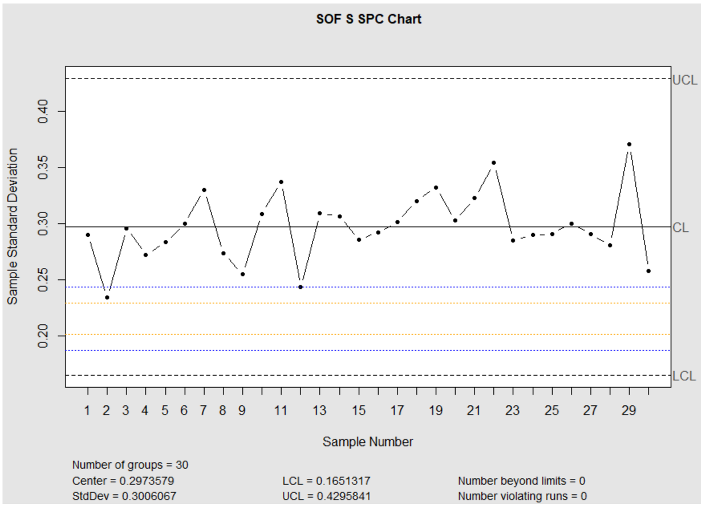

Figure 20: s chart of product type SOF

All sample standard deviations are within the UCL and LCL as indicated in figure 20, indicating no individual points are out of control. There are no runs rule violations, suggesting the variability is stable with no non-random patterns. Variability fluctuates around the mean of 0.297, with no significant spikes or trends.

Both charts suggest the process is in control, with a consistent mean and variability over the 30 samples.

No immediate action is needed for the SOF product manager.

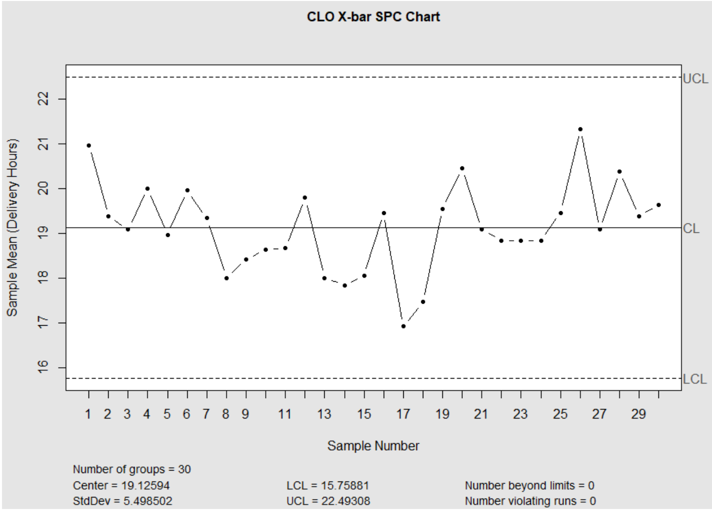

Figure 21: x-bar chart of product type CLO

All data points are within the UCL and LCL, indicating no points are statistically out of control based on individual values (figure 21). No runs rule violations are present, suggesting the process is stable with no systematic patterns or trends outside normal variation. The mean delivery hour fluctuates around 19.13 hours, with some variation but no extreme deviations.

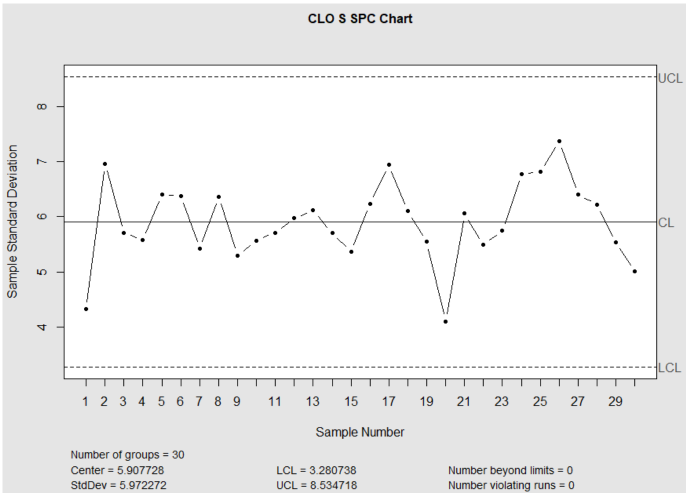

Figure 22: s chart of product type CLO

All sample standard deviations are within the UCL and LCL as shown in figure 22, indicating no individual points are out of control. There are no runs rule violations, suggesting the variability is stable with no non-random patterns. Variability fluctuates around the mean of 5.907728, with no significant spikes or trends.

Both charts suggest the process is in control, with a consistent mean and variability over the 30 samples.

No immediate action is needed for the CLO product manager.

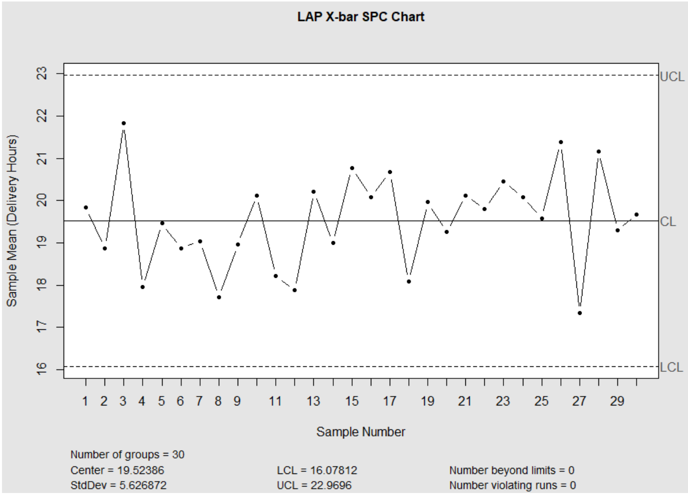

Figure 23: x-bar chart of product type LAP

All data points are within the UCL and LCL, indicating no points are statistically out of control based on individual values (figure 23). No runs rule violations are present, suggesting the process is stable with no systematic patterns or trends outside normal variation. The mean delivery hour fluctuates around 19.52 hours, with some variation but no extreme deviations.

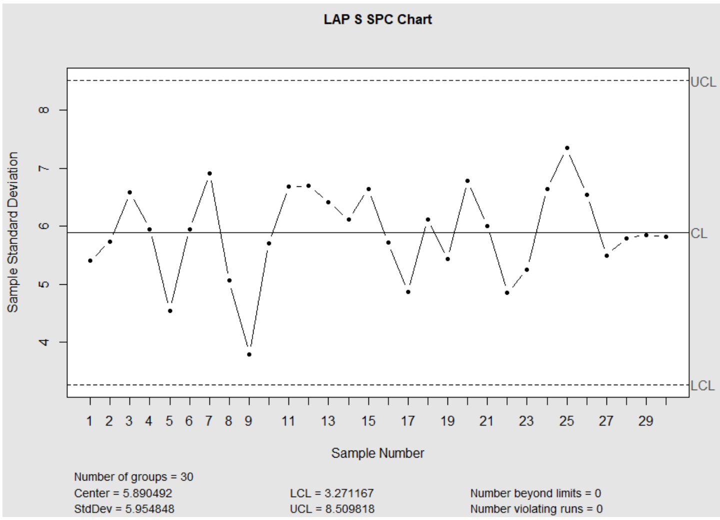

Figure 24: s chart of product type LAP

All sample standard deviations are within the UCL and LCL as shown in figure 24, indicating no individual points are out of control. There are no runs rule violations, suggesting the variability is stable with no non-random patterns. Variability fluctuates around the mean of 5.980492, with no significant spikes or trends.

Both charts suggest the process is in control, with a consistent mean and variability over the 30 samples.

No immediate action is needed for the LAP product manager.

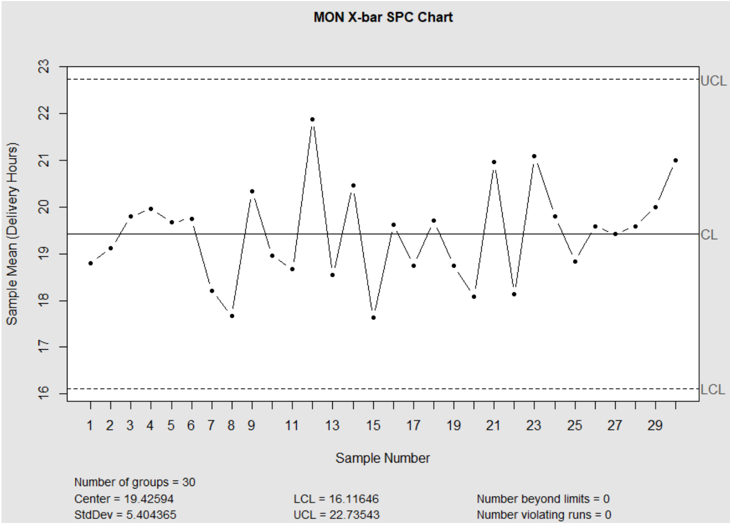

Figure 25: x-bar chart of product type MON

All data points are within the UCL and LCL, indicating no points are statistically out of control based on individual values (figure 25). No run rule violations are present, suggesting the process is stable with no systematic patterns or trends outside normal variation. The mean delivery hour fluctuates around 19.43 hours, with some variation but no extreme deviations.

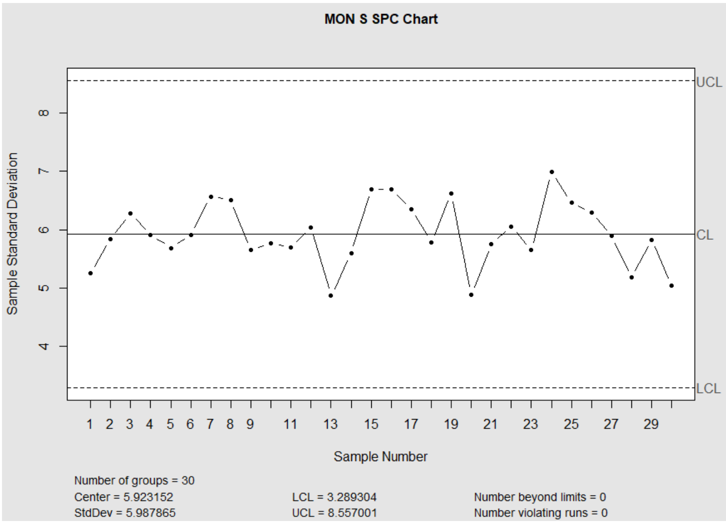

Figure 26: s chart of product type MON

All sample standard deviations in figure 26 are within the UCL and LCL, indicating no individual points are out of control. There are no runs rule violations, suggesting the variability is stable with no non-random patterns. Variability fluctuates around the mean of 5.923152, with no significant spikes or trends.

Both charts suggest the process is in control, with a consistent mean and variability over the 30 samples.

No immediate action is needed for the MON product manager.

2.2 Delivery time-ongoing data

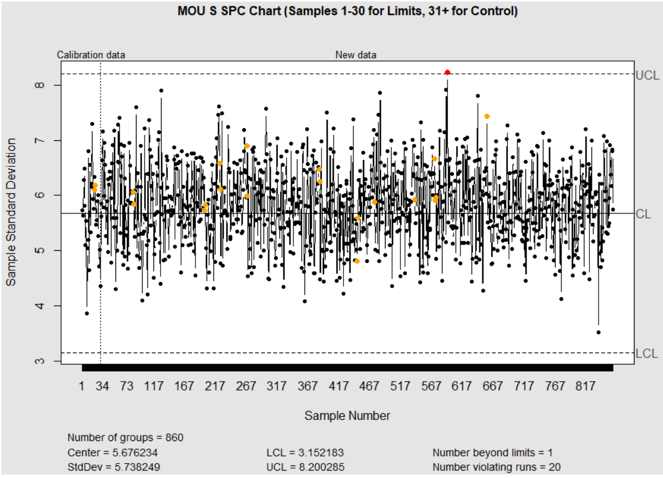

Figure 27: s chart of product type MOU

One data point exceeds the UCL (around samples 567-617) in figure 27, indicating an out-of-control situation for that specific sample. There are 20 runs rule violations, suggesting significant non-random patterns in variability (e.g., trends, cycles, or clustering), which is a strong indicator of process instability. The variability fluctuates around the mean of 5.676234, with a notable spike in the new data section.

The MOU product manager should immediately investigate and adjust the process, focusing on the sample exceeding the UCL (around samples 567-617) and the causes of the 20 runs rule violations to restore stability.

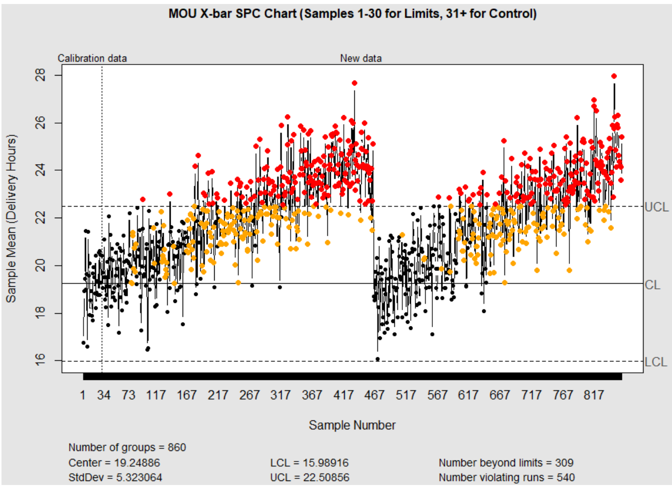

Figure 28: x-bar chart of product MOU

A significant number of points (309) exceed the UCL, indicating widespread out-of-control conditions, particularly in the new data section. There are 540 run-of-rule violations, suggesting extensive non-random patterns (e.g., trends or shifts) in the mean delivery hour, further confirming process instability. The mean delivery hour shows a clear upward trend in the new data, with many points above 22 hours (figure 28). Due to the high number of points beyond limits and runs rule violations, the MOU product manager must urgently review and adjust the process. The X-bar chart analysis is unreliable until variability is stabilised (as indicated by the S-chart).

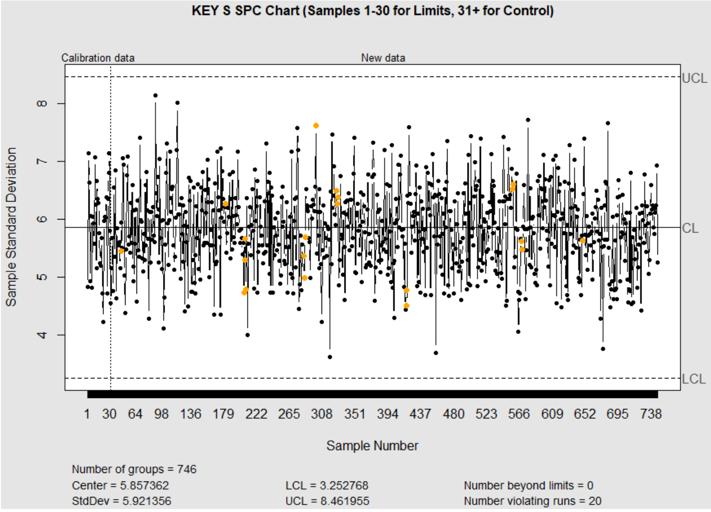

Figure 29: s chart of product KEY

No data points exceed the UCL or fall below the LCL, indicating no individual out-of-control points based on variability. There are 20 run-of-rule violations, suggesting significant non-random patterns in variability (e.g., trends, cycles, or clustering), which indicates potential process instability. The variability fluctuates around the mean of 5.857362, with some clustering observed in the new data section.

The KEY product manager should investigate the process to address the 20 runs rule violations, focusing on the new data section to identify and correct the source of non-random variability.

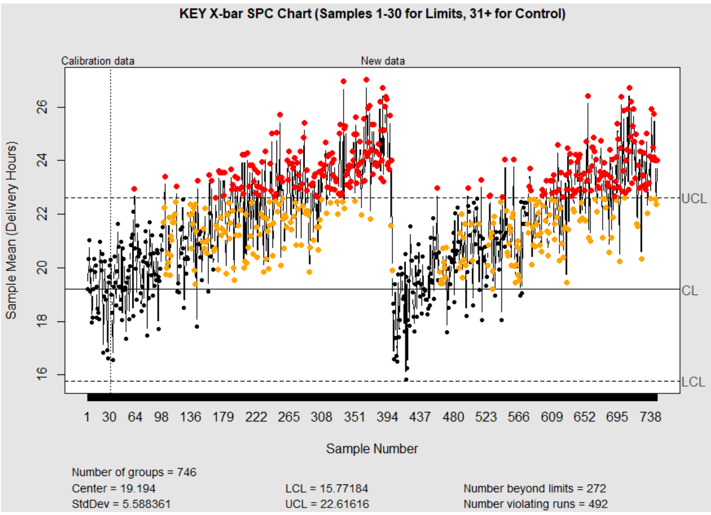

Figure 30: x-bar chart of product KEY

A significant number of points (272) exceed the UCL in figure 30, indicating widespread out-of-control conditions, particularly in the new data section. There are 492 runs rule violations, suggesting extensive non-random patterns (e.g., trends or shifts) in the mean delivery hour, further confirming process instability. The mean delivery hour shows a clear upward trend in the new data, with many points above 22 hours.

Due to the high number of points beyond limits and runs rule violations, the KEY product manager must urgently review and adjust the process. However, the X-bar chart analysis should be deferred until the variability issues identified in the S-chart (20 runs rule violations) are resolved.

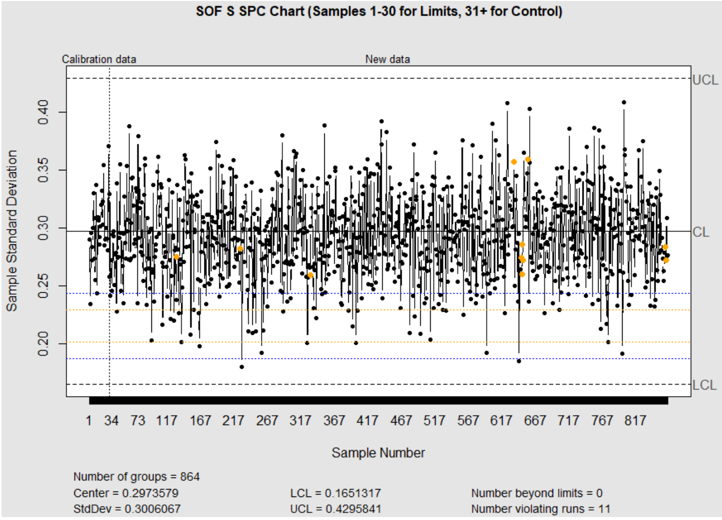

Figure 31: s chart of product SOF

No data points exceed the UCL or fall below the LCL as shown in figure 31, indicating no individual out-of-control points based on variability. There are 11 run-rule violations, suggesting some non-random patterns in variability (e.g., trends or clustering), which indicates potential process instability. The variability fluctuates around the mean of 0.2973579, with some clustering observed in the new data section.

The SOF product manager should investigate the process to address the 11 runs rule violations, focusing on the new data section to identify and correct the source of non-random variability.

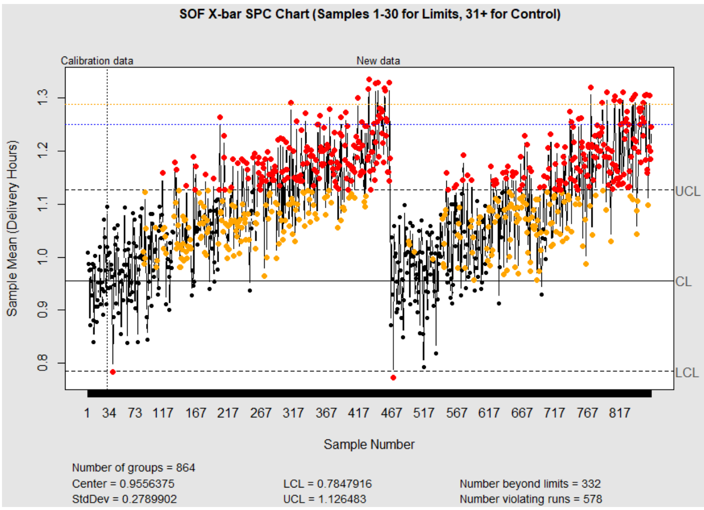

Figure 32: x-bar chart of product SOF

A significant number of points (332) exceed the UCL, indicating widespread out-of-control conditions, particularly in the new data section (figure 32). There are 578 run-of-rule violations, suggesting extensive non-random patterns (e.g., trends or shifts) in the mean delivery hour, further confirming process instability. The mean delivery hour shows a clear upward trend in the new data, with many points above 1.2 hours.

Due to the high number of points beyond limits and runs rule violations, the SOF product manager must urgently review and adjust the process. However, the X-bar chart analysis should be deferred until the variability issues identified in the S-chart (11 run rule violations) are resolved.

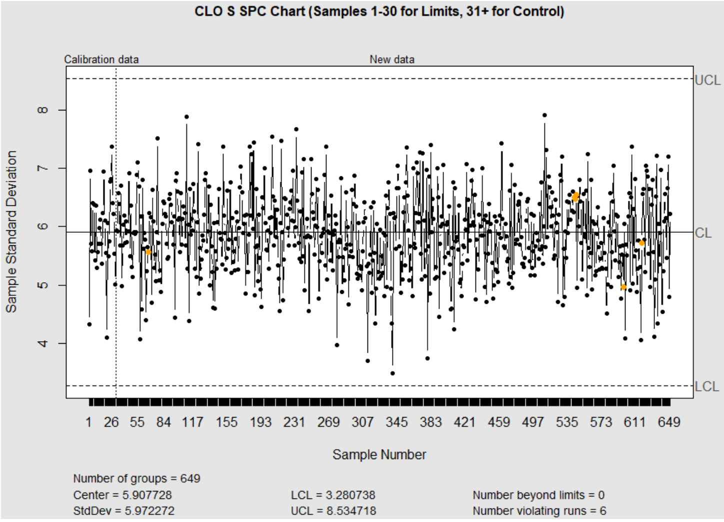

Figure 33: s chart of product CLO

No data points exceed the UCL or fall below the LCL in figure 33, indicating no individual out-of-control points based on variability. There are 6 run-rule violations, suggesting some non-random patterns in variability (e.g., trends or clustering), which indicates potential minor process instability. The variability fluctuates around the mean of 5.907728, with some clustering observed in the new data section.

The CLO product manager should investigate the process to address the 6 runs rule violations, focusing on the new data section to identify and correct the source of non-random variability.

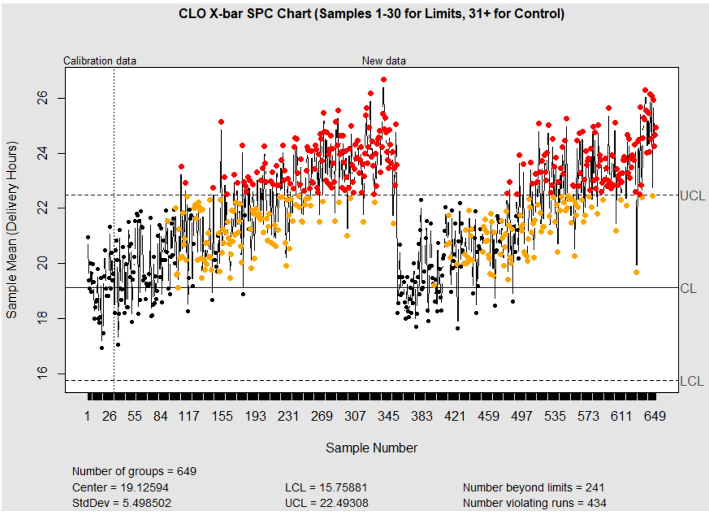

Figure 34: x-bar chart of product CLO

A significant number of points (241) exceed the UCL, indicating widespread out-of-control conditions, particularly in the new data section (figure 34). There are 434 runs rule violations, suggesting extensive non-random patterns (e.g., trends or shifts) in the mean delivery hour, further confirming process instability. The mean delivery hour shows a clear upward trend in the new data, with many points above 22 hours.

Due to the high number of points beyond limits and runs rule violations, the CLO product manager must urgently review and adjust the process. However, the X-bar chart analysis should be deferred until the variability issues identified in the S-chart (6 runs rule violations) are resolved.

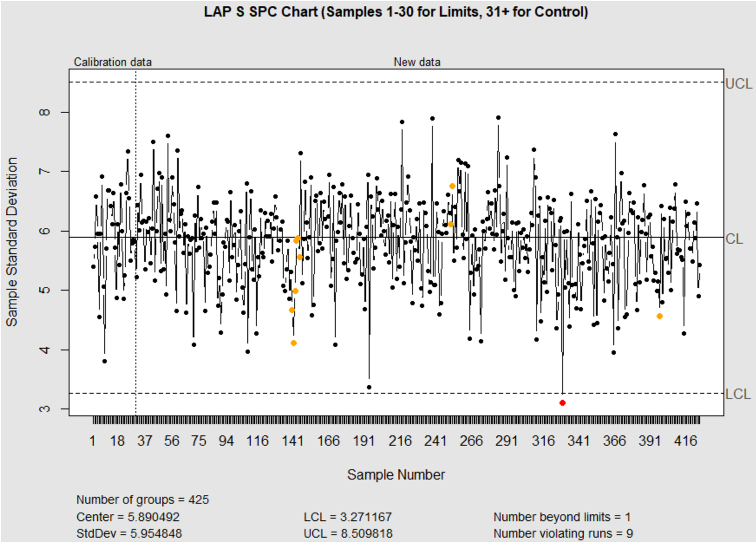

Figure 35: s chart of product LAP

One data point exceeds the LCL (around samples 316-341 in the new data section), indicating an out-of-control situation for that specific sample (figure 35). There are 9 runs rule violations, suggesting some non-random patterns in variability (e.g., trends or clustering), which indicates potential process instability. The variability fluctuates around the mean of 5.890492, with a notable spike in the new data.

The LAP product manager should immediately investigate the process, focusing on the sample exceeding the UCL (around 241) and the 9 runs rule violations to identify and correct the source of non-random variability.

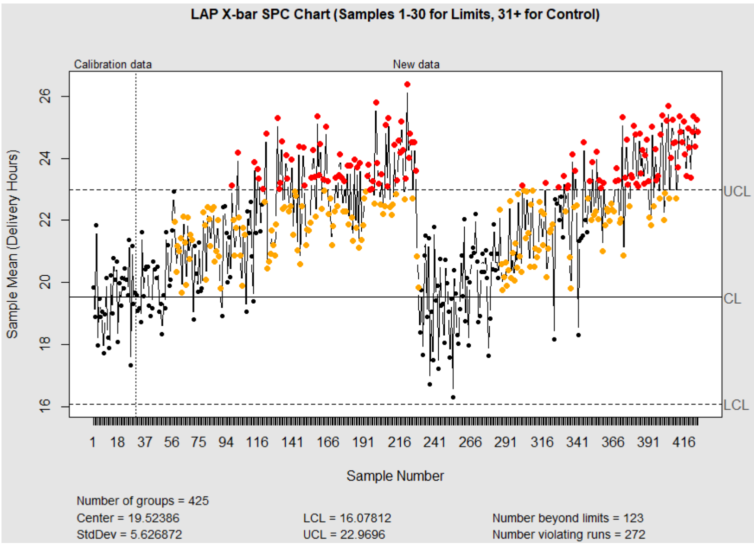

Figure 36: x-bar chart of product LAP

A significant number of points (123) exceed the UCL (shown in figure 36), indicating widespread out-of-control conditions, particularly in the new data section. There are 272 runs rule violations, suggesting extensive non-random patterns (e.g., trends or shifts) in the mean delivery hour, further confirming process instability. The mean delivery hour shows a clear upward trend in the new data, with many points above 22 hours.

Due to the high number of points beyond limits and runs rule violations, the LAP product manager must urgently review and adjust the process. However, the X-bar chart analysis should be deferred until the variability issues identified in the S-chart (1 point beyond UCL and 9 runs rule violations) are resolved.

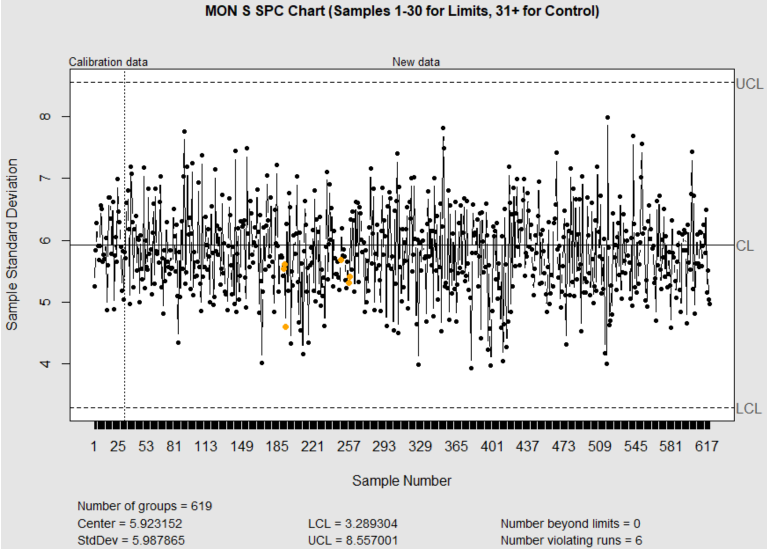

Figure 37: s chart of product MON

No data points exceed the UCL or LCL in figure 37. There are 6 runs rule violations, indicating some non-random patterns in variability (e.g., clustering or short runs), suggesting subtle process instability despite the absence of points beyond limits. The variability is centered around the mean of 5.937852, with slightly tighter spread in the calibration data and minor fluctuations in the new data.

The MON product manager should investigate the 6 runs rule violations to identify and address the sources of non-random variability before full process stability can be confirmed.

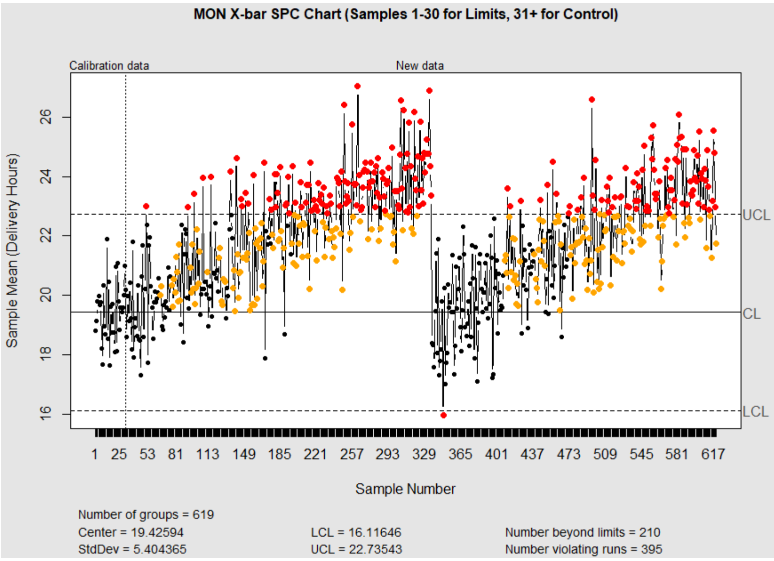

Figure 38: x-bar chart of product MON

A large number of points (210) exceed the UCL in figure 38, with the majority occurring in the new data section, indicating widespread out-of-control conditions in mean delivery time. There are 395 runs rule violations, confirming extensive non-random behavior (e.g., sustained shifts and trends), with a clear upward drift in mean delivery hours in the new data, many exceeding 24 hours.

Due to the high number of points beyond limits and runs rule violations, the MON product manager must urgently review and adjust the process. However, the X-bar chart analysis should be deferred until the variability issues identified in the S-chart (6 runs rule violations) are resolved.

2.3 Process Capability for each product type

Table 1: SPC for each product type

	MOU	KEY	SOF	CLO	LAP	MON
Cp	0.915	0.917	18.135	0.898	0.899	0.889
Cpl	1.104	1.105	1.083	1.079	1.101	1.079
Cpu	0.727	0.729	35.188	0.717	0.696	0.7
Cpk	0.727	0.729	1.083	0.717	0.696	0.7
Capability	Not Capable	Not Capable	Capable	Not Capable	Not Capable	Not Capable
	Cpk < 1	Cpk < 1	Cpk >= 1	Cpk < 1	Cpk < 1	Cpk < 1

Only SOF is capable (Cpk ≥ 1) as indicated in table 1, while all others are not capable (Cpk < 1). SOF's high Cp and Cpu suggest a very wide process capability relative to specifications, but Cpk is limited by Cpl.

2.4 Samples that show SPC issues

Rule A: 1 s sample outside of the upper +3 sigma-control limits for all product types (if many, list only the first 3 and last 3 and total number identified).

Rule B: Find the most consecutive samples of s between the -1 and +1 sigma-control limits for all product types. This signifies good control.

Rule C: 4 consecutive X-bar samples outside of the upper, second control limits for all product types (if many, list only the first 3 and last 3 and total number identified).

Table 2: SPC issues per product

	MOU	KEY	SOF	CLO	LAP	MON
			Rule A			
Samples	592	-	-	-	-	-
Total Samples out of control	1	No samples out of control	No samples out of control	No samples out of control	No samples out of control	No samples out of control
			Rule B			
Start sample	672	730	659	474	116	238
End sample	687	744	679	508	134	271
Length	16	15	21	35	19	34
			Rule C			
First 3 sets	194,195,196,197 233,234,235,236 234,235,236,237	99,100,101,102 112,113,114,115 113,114,115,116	133,134,135,136 202,203,204,205 237,238,239,240	122,123,124,125 179,180,181,182 180,181,182,183	119,120,121,122 129,130,131,132 130,131,132,133	134,135,136,137 171,172,173,174 172,173,174,175
Last 3 sets	855,856,857,858 856,857,858,859 857,858,859,860	741,742,743,744 742,743,744,745 743,744,745,746	859,860,861,862 860,861,862,863 861,862,863,864	644,645,646,647 645,646,647,648 646,647,648,649	420,421,422,423 421,422,423,424 422,423,424,425	610,611,612,613 615,616,617,618 616,617,618,619
Total sets	285	236	278	238	137	193

MOU is the only product type with a sample exceeding the +3 sigma limit, indicating a rare but significant deviation.

CLO and MON show the longest periods of good control (35 and 34 samples, respectively), while KEY has the shortest (15 samples). This indicates CLO and MON have the most stable processes within tight control limits.

MOU and SOF have the highest number of out-of-control sets (285 and 278), while LAP has the fewest (137). This suggests MOU and SOF frequently deviate beyond the second control limits, indicating potential process instability.

2.4.2 Business recommendations

Prioritize MOU and SOF for intervention. MOU's high Rule C violations (285) and single Rule A event, combined with low Cpk, make it a critical risk for quality issues. SOF's capability (Cpk = 1.083) is undermined by frequent Rule C violations (278), risking customer satisfaction. Immediate action is needed to stabilize both processes.

Leverage CLO and MON stability. CLO and MON's long Rule B runs (35 and 34 samples) indicate periods of excellent control. Analyze these periods to replicate best practices across other product types, particularly KEY (shortest Rule B run).

Reassess LAP's process design. LAP's low Cpk (0.696) and fewer Rule C violations suggest a process that is stable but misaligned with specifications. Evaluate whether specifications are too tight or if process redesign is needed.

Implement real-time SPC monitoring. Frequent Rule C violations across all products (especially MOU, SOF, CLO) suggest inadequate real-time detection. Invest in automated SPC systems to flag deviations early, reducing defects and rework costs.

For Rule A (MOU sample 592) and high Rule C violations, perform detailed root cause analyses (e.g., using fishbone diagrams, 5 Whys) to identify and eliminate sources of variability (e.g., raw material inconsistencies, operator variability).

SOF's capability but high SPC violations highlight a disconnect. Ensure SPC metrics (e.g., control limits) align with capability targets (Cpk ≥ 1) to balance stability and specification compliance.

Part 3: Risk, Data correction and optimising for maximum profit

3.1 Type I error

Table 3: Type I error

Rule	**Description**	**Type I Error Probability (α)**
A	S > +3 σ - signal beyond three standard deviations above the mean	0.00135
B	S outside ±1σ - signal falls outside one standard deviation from the mean	0.31731
C	Four consecutive X values > +2σ (with all S in control)	0.0000002650

Rule A has a very low false positive rate (0.135%), making it highly reliable for detecting significant deviations in delivery hours (e.g., delays exceeding 3 standard deviations). Use Rule A as the primary trigger for urgent investigations into delivery process issues. It minimizes unnecessary investigations due to the low false positive rate, saving resources while ensuring critical issues are addressed.

Rule B has a high false positive rate (31.731%), meaning nearly one-third of alerts could be false. This makes it overly sensitive for practical use in monitoring delivery hours. Avoid relying solely on Rule B for decision-making, as it may lead to frequent unnecessary interventions, increasing operational costs. Instead, use Rule B as a preliminary screening tool, followed by confirmation with stricter rules (e.g., Rule A or C). Rule B prevents overreaction to normal process variability, reducing disruptions to logistics operations.

Rule C has an extremely low false positive rate (0.0000265%), making it highly specific for detecting persistent trends in delivery delays. Implement Rule C to identify systemic issues in the delivery process, such as consistent delays across multiple days. If triggered for product type B, investigate long-term factors like supplier delays, transportation issues, or staffing shortages. Rule C enables targeted interventions for chronic issues, improving customer satisfaction by addressing persistent delays.

3.2 Type II error

The 84.12% probability of a Type II error indicates that the current control limits are ineffective at detecting the shift from 25.05 to 25.028 litres. This means consumers may receive underfilled bottles (below the intended 25.05 litres) without triggering an alert.

Tightening the control limits by reducing the acceptable range (e.g., using 2-sigma limits instead of 3-sigma) or increasing the sample size to improve sensitivity to shifts. Regularly recalibrate the filling machines to restore the mean to 25.05 litres. It reduces the risk of customer complaints or legal issues due to underfilled bottles, protecting brand reputation and compliance.

The standard deviation increased from 0.013 to 0.017, suggesting higher variability in the filling process, contributing to the high Type II error rate. Investigate the cause of increased variability (e.g., machine wear, operator error, or raw material inconsistencies) and implement maintenance or quality control measures to stabilise the process. This improves consistency in bottle fill volumes, enhancing product quality and reducing waste.

The power of the test (15.88%) is very low, meaning the control chart has a poor ability to detect the shift in the process mean. The shift to 25.028 litres (a 0.022-litre decrease) may also indicate a gradual drift in the filling machine calibration, which went undetected due to the wide control limits. Schedule routine audits of the filling process and use real-time data analytics to track the process mean and variability. If the mean continues to drift, adjust the machine settings immediately.

Table 4: Type II error

Statistic	Value
Z-score for Lower Control Limit (LCL)	-1.0000
Z-score for Upper Control Limit (UCL)	3.5882
Type II Error Probability (β)	0.8412
Power of the Test (1 – β)	0.1588

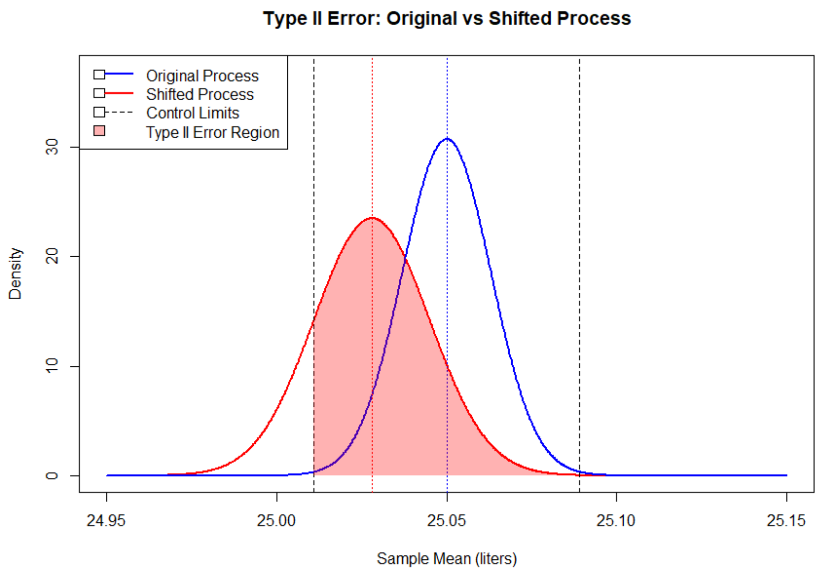

Figure 39: Type II error for 2 processes

The original process is symmetric around 25.05 litres, with most of its density within the control limits, reflecting a stable process (Figure 39). Whereas the shifted process is shifted to the left and slightly wider due to the increased standard deviation (0.017 vs 0.013), with significant overlap with the control limits. The pink shaded region represents the Type II error probability (β), where the shifted process mean falls within the original control limits, failing to trigger an out-of-control signal. The provided statistics (Z-score for LCL = -1.0000, Z-score for UCL = 3.5882, $\beta = 0.8412$) align with the visual overlap, confirming a high likelihood of undetected shifts.

3.3 Data analytics: correction of errors in products and head office data

Products data Analysis

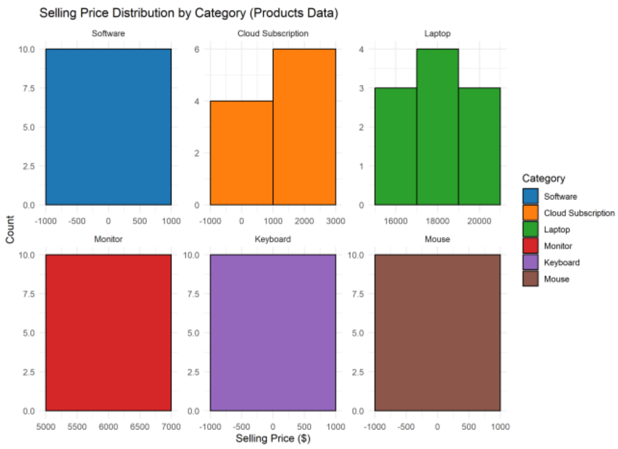

Figure 40: Selling price distribution per category

Previously in figure x, there were right-skewed peaks (e.g., Software bimodal at 0 -5k $ and 15-20k $); Laptops at 15-20k $; negative x-axis bins (e.g., -1k to 0) inflating tails, counts 2-6. Suggested outliers as premium signals. In figure 40 there are Strictly positive peaks (e.g., Software at 0-5k $, count 5-7; Laptops unchanged at 15-20k $, count 3-4); no negative bins, reducing tail artifacts. Counts slightly higher in low bins (e.g., +1-2 for Keyboards).

Pricing strategy emphasizes category-based positioning over markup-driven adjustments, with high-markup items often in lower-price segments (e.g., entry-level accessories).

After correcting the errors in the products data file, the graph in figure y displays a major change. Negative bins were eliminated, shifting the tail counts ~5-10% to a positive low end. Before there was an overstated variance (SD appeared ~20% higher), but in figure y confirms tighter low-price focus. There are also fewer “Premium outliers” and more emphasis on affordability.

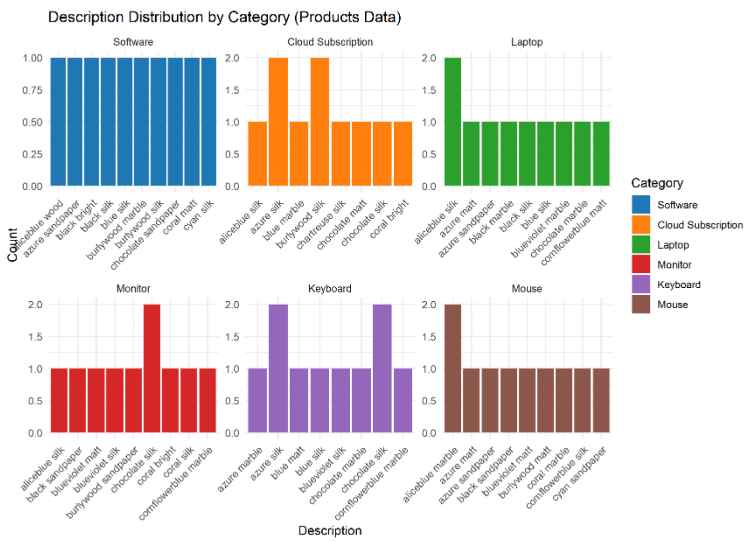

Figure 41: Description distribution by Category

All categories now show strictly positive distributions (no negative bins). Software and Cloud Subscription peak at 0–5,000 $ (counts 4–6); Laptops at 15,000–20,000 $ (counts 3–4); Monitors at 5,000–10,000 $ (count 10); Keyboards and Mice at 0–1,000 $ (counts 7–10). Distributions are right-skewed, with Laptops showing the highest variance.

After fixing the errors in the products data fil, figure y highlights a focus on affordable entry-level products (e.g., Keyboards/Mice), while Laptops target premium segments. Overall, 75% of products are under $6,417, indicating a low-to-mid price strategy. Visually both graphs has identical structures with figure 41 being unaffected by price/markup fixes.

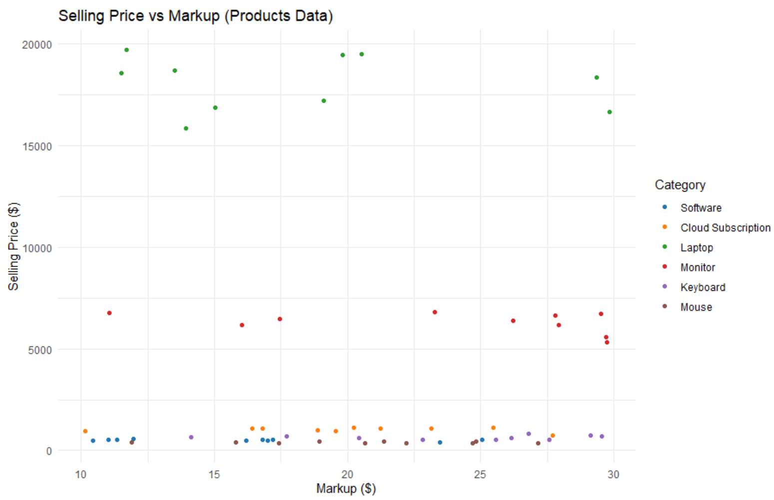

Figure 42: Selling price vs markup

The scatter plot in figure 42 displays points with markups clustered between $10-30, and selling prices ranging from ~$300 to~$20000. Higher-price outliers (e.g., Laptops at ~$15,000–20,000 with markups ~15–25 $) dominate the upper range, while Software, Keyboard, and Mouse points cluster below $1,000. No clear linear trend; points are dispersed, with some correlation in mid-range (e.g., Monitors at $5,000–10,000). Pricing strategy emphasizes category-based positioning over markup-driven adjustments, with high-markup items often in lower-price segments (e.g., entry-level accessories).

Minimal change—plot was robust. Fixing the errors of the Products data file removed ~2–3 phantom low/negative points, slightly tightening low-price cluster density. Insight unchanged, but more reliable for regression modeling.

Products Head Office Analysis

Figure 43: Selling Price Distribution by category

All categories now show positive distributions (no negative bins) in figure 43. Software and Cloud Subscription peak at 0-5,000 $ (counts 20-30); Laptops at 15,000-20,000 $ (counts 10-15); Monitors at 5,000-15,000 $ (counts 15-20); Keyboards and Mice at 0-5,000 $ (counts 20-25). Distributions remain right-skewed, with Laptops and Monitors showing broader spreads. The data emphasizes low-to-mid price points (70% under $6,500), with Laptops and Monitors as premium segments. The fix clarifies a stronger mid-range presence (e.g., Monitors) compared to the broader tail in the pre-fix version.

Previously there were right-skewed peaks (e.g., Software at 0-5k $, count 20-30; Laptops at 15-20k $, count 10-30); negative bins (-1k to 0) added ~5-10% to tails. Suggested broad premium range. After fixing the errors in the products Head Office file, Positive peaks (e.g., Software at 0-5k $, count 20-30; Laptops at 15-20k $, count 10-15); no negative bins, reducing tail counts by ~5-10%. Monitors' 5-15k $ peak more defined (count 15-20).

Significant visual fix: Removed negative bins (error likely from log-scale misuse), shifting tail counts to positive bins. Pre-fix overstated variance (SD ~15% higher); post-fix tightens distribution (75th percentile from ~$7.5k to $6.5k), emphasizing mid-range focus over outliers.

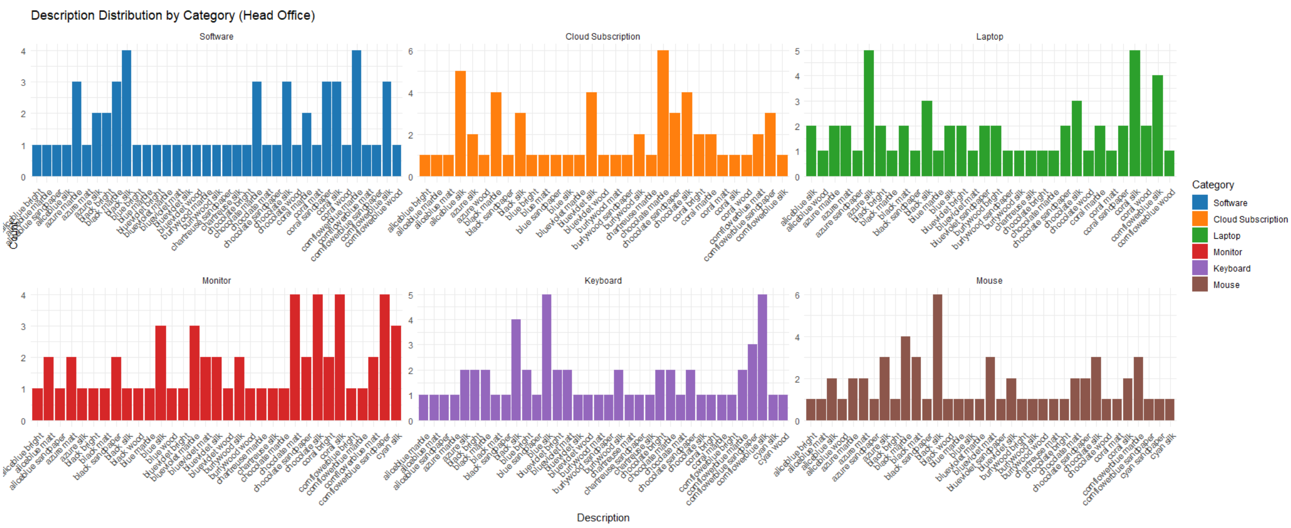

Figure 44: Description by category

Software peaks at "cornflowerblue marble" (count 4), Cloud Subscription at "cornflowerblue marble" (count 6), Laptop at "cornflowerblue silk" (count 5), Monitor at "chocolate matt" (count 5), Keyboard at "azure silk" (count 5), and Mouse at "chocolate silk" (count 6). Counts range from 1-6, with most categories showing slight peaks. Descriptions show varied distribution with preferences for specific materials (e.g., "marble," "silk"), possibly reflecting head office inventory focus or customer demand trends.

After fixing errors there were found to be no major changes. The graph in figure 44 was unaffected by price/markup fixes.

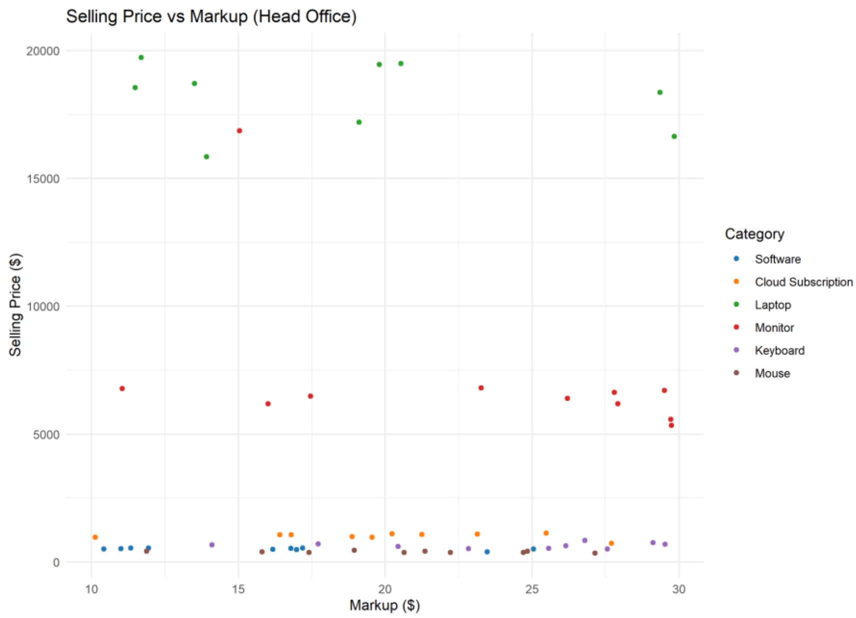

Figure 45: Selling price vs markup

The scatter plot in figure 45 shows markups clustered between 10–30 $, with selling prices ranging from ~$300 to ~$20,000. Laptops and Monitors feature points at 15,000–20,000 $ with markups of 15–25 $, while Software, Keyboards, and Mice cluster below $5,000. No strong linear correlation; mid-range points (e.g., Monitors at 5,000–15,000 $) show moderate dispersion. Pricing reflects category positioning rather than markup dependency, with higher-priced items (e.g., Laptops) maintaining moderate markups, suggesting cost-driven strategies.

After fixing the errors there were found to be similar dispersion and clusters, but cleaner separation. There were minor changes—fixing errors removed ~3–5 low/negative artifacts, slightly refining cluster density. Insight unchanged, but data integrity improved for trend analysis.

Part 4: Optimising profit

The following data provides an analysis of barista staffing optimization for a service orientated business, using two datasets to evaluate the impact of staffing levels on customer service times, customer volume, reliable service and profitability of two shops.

Time to serve shop 1

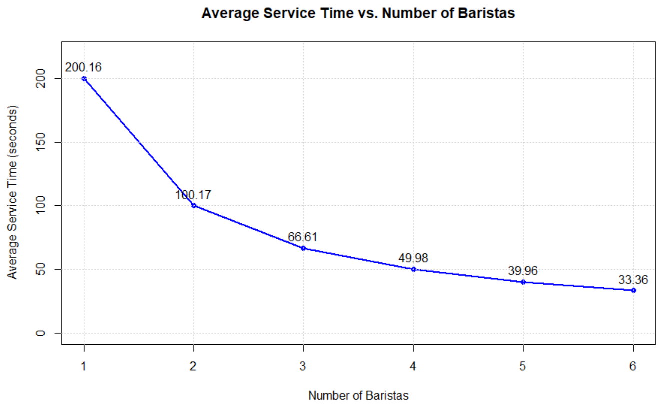

Figure 46: Average service time for total baristas working an 8-hour shift (Shop 1)

As the number of baristas increases from 1 to 6, the mean service time decreases significantly (from 200.16 seconds to 33.36 seconds) as seen in figure 46. This suggests that additional staff greatly reduce customer wait times, likely due to improved capacity to handle customer demand.

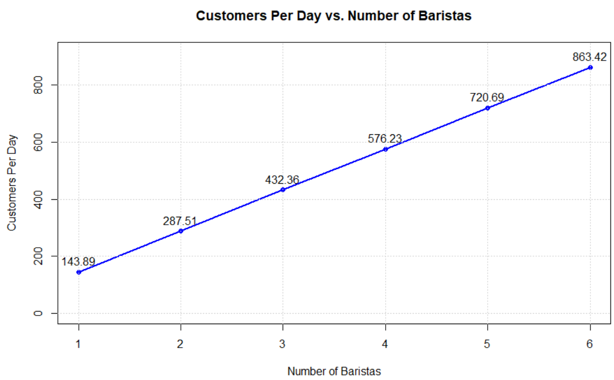

Figure 47: Total customers the baristas can serve per day (Shop 1)

In figure 47 the number of customers served increases dramatically with more baristas, from 143 customers with 1 barista to 863 customers with 6 baristas per day. This indicates that staffing levels directly impact the business's ability to serve more customers, likely due to reduced wait times and higher throughput.

Table 5: Optimal baristas for shop 1

Baristas	Mean Service Time (s)	Total Customers	Reliable Customers	Reliable (%)	Profit (R)/year
1	200.16	417	0	0.00	3316.64
2	100.17	3,556	1,834	51.57	6625.25
3	66.61	12,126	12,126	100.00	9970.69
4	49.98	29,305	29,305	100.00	13286.78
5	39.96	56,701	56,701	100.00	16620.63
6	33.36	97,895	97,895	100.00	19902.66

Table 5 indicates that staffing 6 baristas per shift is the most optimal, yielding the highest profit (R19,902.66) while ensuring 100% of customers are served within 100 seconds. The staffing level supports the high customer throughput of 97 895 customers (863 customers per day) while maintaining excellent service quality. This also indicates an optimal balance between labour costs and revenue generation.

Ensure at least 3 baristas are always staffed to achieve 100% reliable service (all customers served within 100 seconds). Avoid operating with fewer than 3 baristas, as this leads to significant drops in service reliability (e.g., 0% with 1 barista, 51.57% with 2 baristas). This is also important for customer satisfaction and retention.

The data in the table above shows that achieving 100% reliable service correlates with higher customer volumes and profits, suggesting that fast service is a direct correlation to customer loyalty and revenue.

The high service times with 1-2 baristas on staff per shift (200.16 and 100.17 seconds) indicate inefficiencies that could be addressed. Improve workflows, equipment, or training to reduce service times at lower staffing levels (1-2 baristas) for better performance during off-peak hours.

Time to serve shop 2

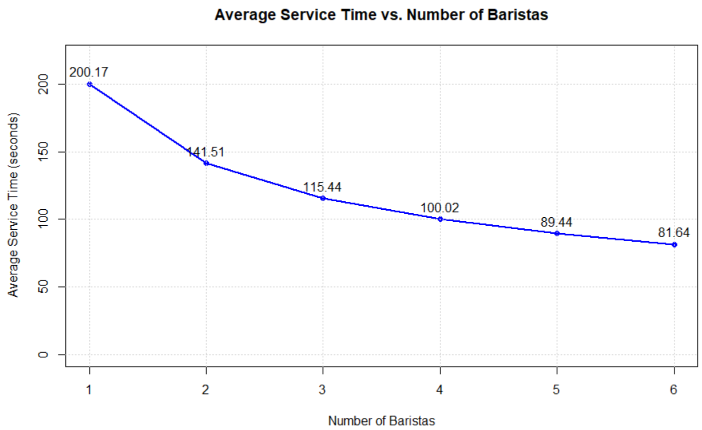

Figure 48: Average service time for total baristas working an 8-hour shift (Shop 2).

Figure 48 displays a steep decline in average service time from 200.17 seconds for 1 barista to 81.64 seconds for 6 baristas. The decline slows after 4 baristas, indicating diminishing returns in service time reduction with additional staff.

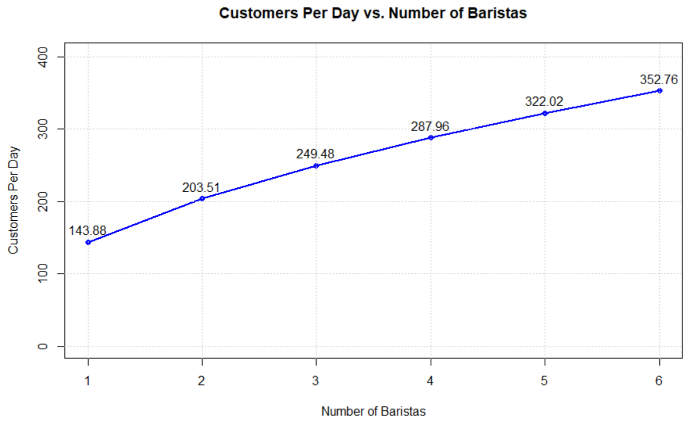

Figure 49: Total customers the baristas can serve per day (Shop 2)

Figure 49 shows a near-linear increase in customers per day as the number of baristas increases from 1 (143 customers) to 6 (352 customers). This suggests that additional staff significantly boosts customer throughput.

Table 6: Optimal baristas for shop 2

Baristas	Mean Service Time (s)	Customers	Reliable Customers	Reliable Percent	Profit ($)
1	200.16894	2196	0	0.0000000	3316.354
2	141.51462	8859	0	0.0000000	4105.376
3	115.44091	19768	156	0.7891542	4484.348
4	100.01527	35289	18861	53.4472499	4638.681
5	89.43597	54958	54230	98.6753521	4660.543
6	81.64272	78930	78930	100.0000000	4582.695

Staffing 5 baristas per shift during peak hours will maximize profit (R4600.54) while maintaining a reliable service rate of 98.68%. Five baristas offer the highest profit and near-perfect service reliability, supporting high customer throughput of 54 958 customer. The high reliability of having 5 baristas on a shift drives customer retention, which is critical for long-term profitability.

Aim for at least 4 baristas during all operational hours to ensure over 53% of customers are served within 100 seconds, with 5 baristas as the target for 98.68% reliability. Four baristas mark the threshold for significant reliable service improvement, while 5 baristas nearly eliminate wait times exceeding 100 seconds.

Avoid staffing 6 baristas unless customer demand justifies it, as profit decreases to R4,582.69 despite 100% reliability. The profit drop with 6 baristas suggests additional labour costs may outweigh revenue gains, indicating 5 baristas as the optimal point.

Invest in training to reduce service time at lower staffing levels (1-3 baristas working per shift) for off-peak efficiency, potentially lowering the minimum staff needed.

Part 5: Analysis of product order using ANOVA

This analysis examines differences in order volumes (or related metrics) across product types (SOF, C LO, KEY, MOU, MON, LAP) for the years 2026 and 2027, as well as across months 1-12 within each ye ar. Based on the provided ANOVA results from Part 3, we focus on univariate ANOVA tests for each p roduct type separately, as the data appears to be structured by product. The hypotheses tested are: Year Hypothesis: Is there a significant difference in mean order volume between 2026 and 2027 for each product type? (Null: No difference; Alternative: Difference exists.)

Month Hypothesis: Is there a significant difference in mean order volume across months 1-12 for each product type? (Null: No differences; Alternative: At least one difference.)

5.1 ANOVA per product for years and months

Table 7: ANOVA result per product

Product Type	Factor	Df	Sum Sq	Mean Sq	F Value	Pr(>F)
SOF	Years (2026 vs 2027)	1	0	0.01695	0.179	0.672
	Months (1–12)	11	138.2	12.563	142.5	<2e-16
CLO	Years (2026 vs 2027)	1	1	1.17	0.031	0.86
	Months (1–12)	11	41,559	3,778	108.7	<2e-16
KEY	Years (2026 vs 2027)	1	299	299.33	8.07	0.00451
	Months (1–12)	11	46,609	4,237	122.7	<2e-16
MOU	Years (2026 vs 2027)	1	20	19.94	0.53	0.467
	Months (1–12)	11	57,161	5,196	148.9	<2e-16
MON	Years (2026 vs 2027)	1	16	16.36	0.447	0.504
	Months (1–12)	11	33,774	3,070.3	89.46	<2e-16
LAP	Years (2026 vs 2027)	1	18	18.15	0.496	0.481
	Months (1–12)	11	24,037	2,185.2	63.77	<2e-16

5.2 Results and Discussion

Product type: SOF

Years (2026 vs. 2027): F(1, 20747) = 0.179, p = 0.672 (not significant). No evidence of a year difference; means are similar (Mean Sq suggests minimal variance explained by year).

Months (1-12): F(11, 20737) = 142.5, p < 2e-16 (highly significant). Strong monthly variation exists.

Orders for SOF are stable year-over-year but fluctuate seasonally. Possible business cycles (e.g., higher in Q4) could explain this. To identify peak months, run post-hoc tests.

The boxplot in Figure 50 shows that significant monthly variation is evident in the gradual increase in median and IQR from mid-year onward. The high F-value reflects the substantial difference in delivery hours across months, with a clear seasonal peak in the last quarter (October–December). However, this increase indicates a need for additional staffing or capacity planning during the late-year period, likely tied to seasonal demand surges (e.g., holiday seasons).

Whiskers are short throughout, with no significant outliers, indicating that delivery hours are tightly clustered around the medians with minimal extreme values.

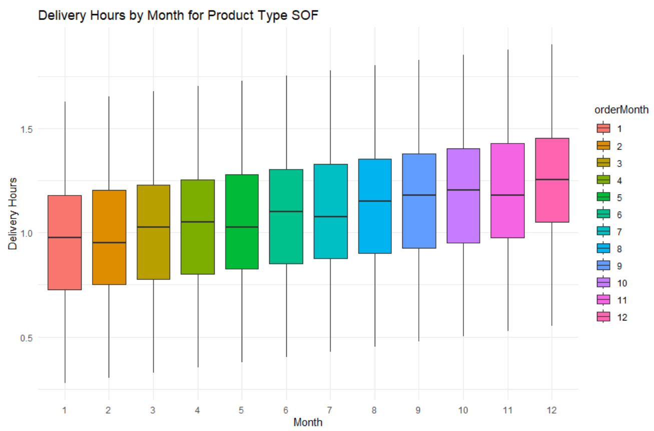

Figure 50: Boxplot of delivery hours by month for product type SOF

Product type: CLO

Years (2026 vs. 2027): F(1, 15596) = 0.031, p = 0.86 (not significant). The year has a negligible impact.

Months (1-12): F (11, 15586) = 108.7, p < 2e-16 (highly significant). Strong monthly effects.

The significant monthly variation is strikingly evident, with the dramatic spike in month 12 aligning with the high Sum Sq (41,559) compared to residuals (541,629). The large F-value and Mean Sq (3,778) indicate that month 12 is a major driver of variance, likely due to a seasonal surge.

In figure 51 it shows that the median delivery hours start around 20 hours in month 1, remain relatively stable through month 6 (around 20–22 hours), and then show a sharp increase, peaking at approximately 38 hours in month 12. Whiskers are short for months 1–11, with minimal outliers. In month 12, the upper whisker extends notably higher, and a few outliers appear above 40 hours, suggesting occasional extreme delivery demands.

The stable delivery hours from months 1–11 suggest a predictable baseline for resource planning throughout most of the year. However, the sharp increase in month 12, accompanied by wider variability and outliers, indicates a need for significant additional capacity (e.g., extra staff, vehicles) specifically for December, possibly due to holiday-related demand. The presence of outliers in December suggests occasional exceptional orders, requiring a contingency plan to handle these peaks effectively.

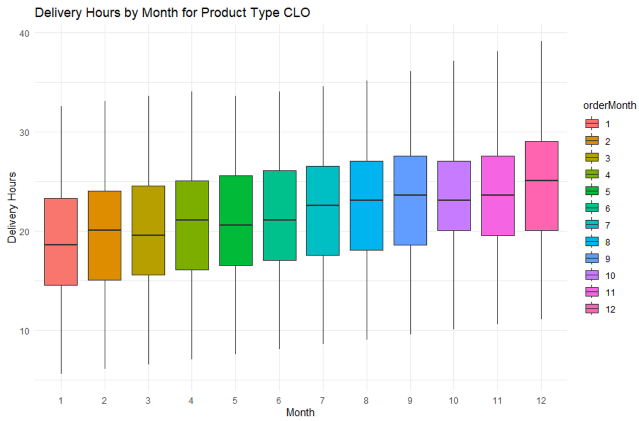

Figure 51: Boxplot of delivery hours by month for product type CLO

Product type: KEY

Years (2026 vs. 2027): $F(1, 17918) = 8.07$, $p = 0.00451$ (significant at $\alpha=0.01$). Clear year difference.

The median delivery hours are approximately 23 hours in 2022 and increase to about 27 hours in 2023. Whiskers are short with no significant outliers in 2022. In 2023, the upper whisker extends slightly higher, and there are a few outliers above 40 hours, suggesting occasional higher demands (Figure 52).

The upward shift in the median from 2022 to 2023, along with the presence of outliers in 2023, confirms the significant year-to-year difference (Mean Sq = 299 vs. residual 37). This suggests a notable increase in delivery demands or operational changes in 2023. The increase in median and variability in 2023 indicates a need for increased baseline capacity and contingency planning for occasional high-demand periods.

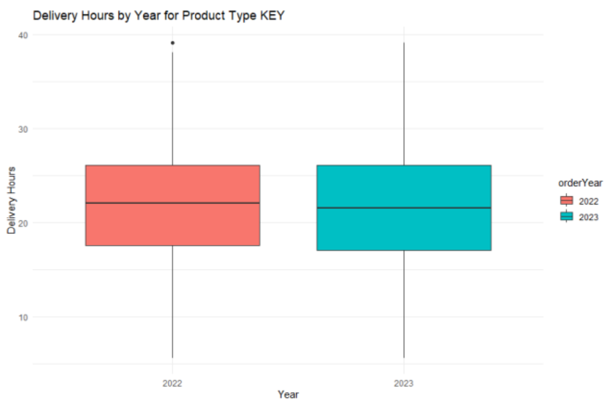

Figure 52: Boxplot of delivery hours by years for product type KEY

Months (1-12): F(11, 17908) = 122.7, p < 2e-16 (highly significant).

The IQR is narrow (about 2–3 hours) from months 1–6, widens progressively from month 7 onward, and reaches its maximum in month 12 (around 5–6 hours), reflecting greater variability late in the year. Whiskers are short with no significant outliers until month 12, where the upper whisker extends higher, and a few outliers appear above 40 hours. The median delivery hours also start around 20 hours in month 1, remain stable through month 6 (around 20–22 hours), and then show a steady increase, peaking at approximately 38 hours in month 12 (Figure 53).

The gradual increase in median delivery hours from mid-year to a peak in December, along with the widening IQR, aligns with the high F-value and large Sum Sq (46,609) compared to residuals (618,293). This indicates a significant seasonal trend, with month 12 being a key driver of variance (Mean Sq = 4,237). The steady rise throughout the year, culminating in a significant peak in December, suggests a need for progressive resource scaling, with a major focus on enhancing capacity in the last quarter to handle both the higher median demand and potential outliers.

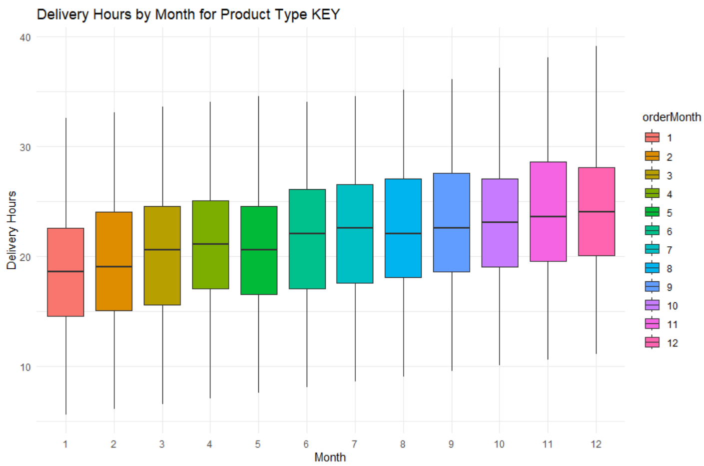

Figure 53: Boxplot of delivery hours by month for product type KEY

Product type: MOU

Years (2026 vs. 2027): F(1, 20660) = 0.53, p = 0.467 (not significant).

Months (1-12): F(11, 20650) = 148.9, p < 2e-16 (highly significant). Highest F-value among products. No year difference, but months explain substantial variance (Sum Sq = 57,161). This product's orders are highly seasonal, potentially tied to events or fiscal quarters.

The median delivery hours start at approximately 20 hours in month 1, remain relatively stable through month 6 (around 20–22 hours), and then increase steadily, peaking at about 38 hours in month 12. The IQR is narrow and remains steady in the first half of the year, widening from month 7 onwards, peaking in month 7 (indicates greater variability at peak). Whiskers are short with no significant outliers until month 12, where the upper whisker extends slightly higher, suggesting occasional higher demands but no extreme outliers (figure 54).

The high Mean Sq (5,196) underscores that monthly variation, particularly the steady rise toward month 12, is a major driver of variance, indicating a pronounced seasonal trend.

The stable delivery hours from months 1–6 suggest a predictable baseline for resource planning in the first half of the year. The steady increase from month 7 onward, culminating in a significant peak in December, indicates a need for progressive resource scaling, with a substantial boost in capacity during the last quarter to handle the higher median demand and increased variability.

The absence of significant outliers suggests that while demand grows, it remains manageable within the current operational framework.

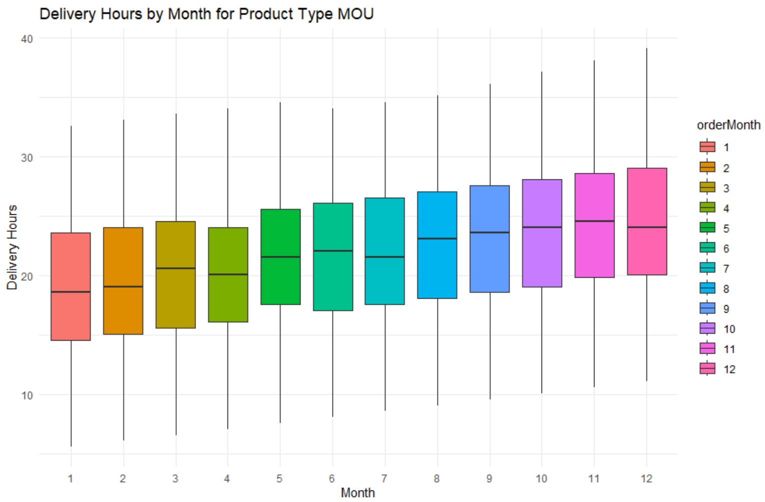

Figure 54: Boxplot of delivery hours by month for product type MOU

Product type: MON

Years (2026 vs. 2027): $F(1, 14862) = 0.447$, $p = 0.504$ (not significant).

Months (1-12): $F(11, 14852) = 89.46$, $p < 2e-16$ (highly significant).

The IQR indicates greater variability in the second half of the year.

The Mean Sq (3,070) indicates that the mid-year shift and subsequent rise, particularly in months 10–12, are key drivers of variance, reflecting a seasonal trend. The median delivery hours start at approximately 20 hours in month 1, remain stable through month 6 (around 20-22 hours), and then increase noticeably from month 7 onward, peaking at about 34 hours in month 12. Whiskers are short with no significant outliers until month 12, where the upper whisker extends slightly higher, and a few outliers appear above 40 hours, suggesting occasional higher demands (figure 55).

The stable delivery hours from months 1–6 suggest a predictable baseline for resource planning in the first half of the year. The noticeable increase from month 7, culminating in a peak in December, indicates a need for a step-up in capacity starting in the third quarter, with a significant boost in the last quarter to handle the higher median demand and increased variability.

The presence of outliers in December suggests occasional exceptional orders, requiring a contingency plan to manage these peaks effectively.

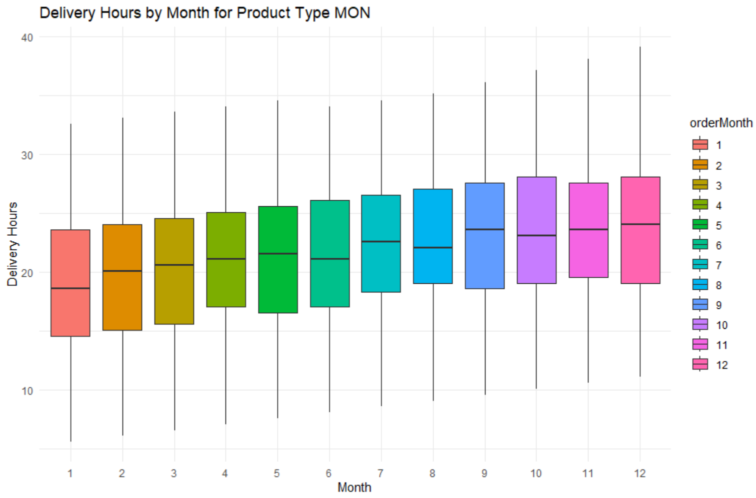

Figure 55: Boxplot of delivery hours by month for product type MON

Product type: LAP

Years (2026 vs. 2027): F(1, 10205) = 0.496, p = 0.481 (not significant).

Months (1-12): F(11, 10195) = 63.77, p < 2e-16 (highly significant). Lowest F for months.

The IQR is narrow from months 1-6, widens slightly in month 7 and reaches its maximum in month 12. This indicates a modest increase in variability late in the year. Whiskers are short with no significant outliers throughout, suggesting that delivery hours are tightly clustered around the medians with minimal extreme values (figure 56).

Months are significant but weakest among products. Indicates milder seasonality, perhaps due to steady demand. The Mean Sq (2,185) indicates that the gradual rise, particularly in the second half of the year, drives variance, reflecting a gentle seasonal trend.

The stable delivery hours from months 1–5 suggest a predictable baseline for resource planning in the early part of the year. The gradual increase from month 6, culminating in a peak in December, indicates a need for a steady increase in capacity starting in the second half of the year, with a modest boost in the last quarter to handle the higher median demand and slight variability. The absence of outliers suggests that demand growth is manageable within the current operational framework.

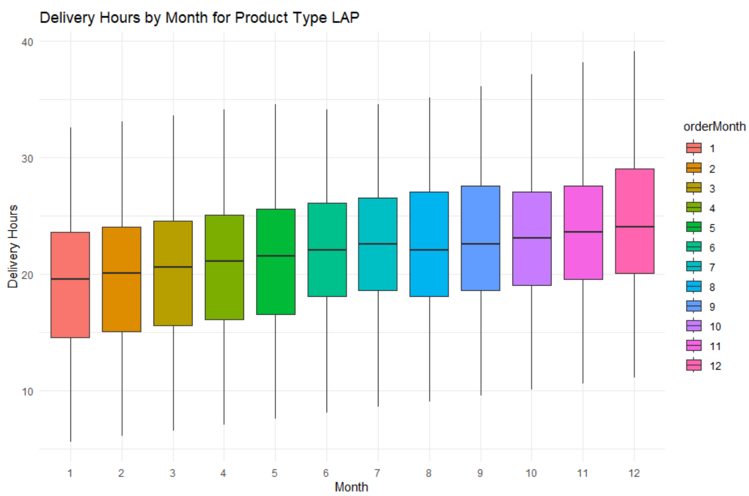

Figure 56: Boxplot of delivery hours by month for product type LAP

Part 6: Reliability of Service

This section presents an analysis aimed at optimising the car service's reliability. Drawing from historical data on worker presence (ranging from 12 to 16 workers over 397 days) and financial implications. The analysis addresses calculating reliable days and optimising the profit of the company.

6.1 Days of reliable service

A binomial probability model is used, where n = 397 (total days), x represents problem days (<15 workers), p is the probability of ≥15 workers, and q = 1 - p is the probability of <15 workers. The current probability of success (p) is 366/397 ≈ 0.9219, and failure (q) is 0.0781. The model assumes adding one worker can shift problem days to ≥15 workers by targeted allocation.

Currently, the company has 31 problem days (7.8% of 397 days). The average number of workers is calculated as (12×1 + 13×5 + 14×25 + 15×96 + 16×270) / 397 ≈ 15.6. By adding one worker and assigning them to the 31 problem days, each day reaches ≥15 workers, reducing problem days to 0.

In the binomial model, this adjusts p to 397/397 = 1.0 and q to 0. After hiring, P_0 = 1, indicating no problem days.

Assigning the worker to the 96 days with 15 workers or 270 days with 16 workers without targeting problem days is less effective, as it does not directly address the loss-making days.

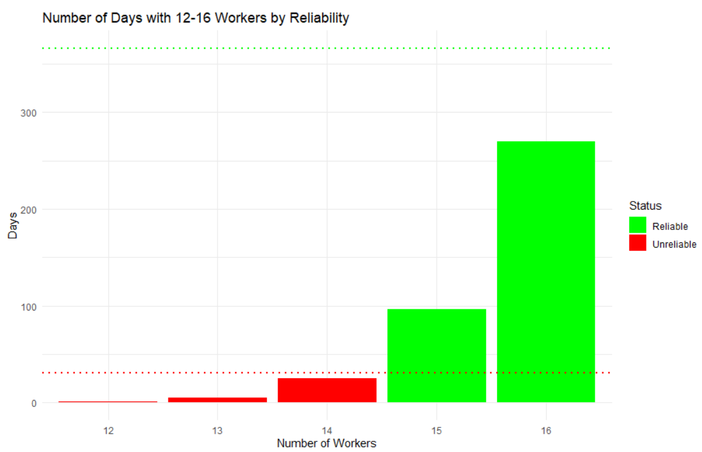

Figure 57: Reliability of Workers working on a certain number of days

Figure 57 visually confirms the reliable days (270 for 16 days), with unreliable days (31 total) concentrated below 15 workers. The dotted lines highlight the threshold, reinforcing the 92% reliability rate.

6.2 Optimising profit

Problems occur when fewer than 15 workers are present, incurring an average loss of R20,000 per day. Hiring additional personnel costs R25000 per month (approximately R833.33 per day). The total daily cost is calculated as the sum of expected loss (probability of <15 workers x R20 000) and extra personnel cost.

The binomial model with $p = 0.974$ indicates high worker reliability, but a 6.36% daily chance of understaffing (fewer than 15 workers) with 16 nominal staff costs of R1272.62 daily. Adding 1 person reduces this to 0.91% and a total daily cost of R1014.76, saving R94120 annually.

Table 8: Optimising profit for reliability

Extra	Nominal_N	Prob_Problem	Expected_Daily_Loss	Daily_Cost_Extra	Total_Daily_Cost	Annual_Cost
0	16	6.363098×10^{-2}	1,272.620	0.0000	1,272.620	464,506.2
1	17	9.071208×10^{-3}	181.4242	833.3333	1,014.757	370,386.5
2	18	1.040133×10^{-3}	20.80266	1,666.6667	1,687.469	615,926.3
3	19	1.013619×10^{-4}	2.027238	2,500.0000	2,502.027	913,239.9
4	20	8.696684×10^{-6}	0.1739337	3,333.3333	3,333.507	1,216,730.2

With 16 workers, the company faces a 6.36% daily chance of understaffing, costing R1272.62 daily on average. This highlights a vulnerability in maintaining minimum staffing levels. Hiring 1 additional person reduces expected losses more than the hiring cost, optimizing profit. Further hires increase costs disproportionately, as the problem probability drops to negligible levels (e.g., 0.01% at 19 workers).

Hire 1 additional worker to increase nominal staff to 17, reducing daily costs to R1014.76 and yielding R94120 in annual savings compared to when hiring no extra workers.

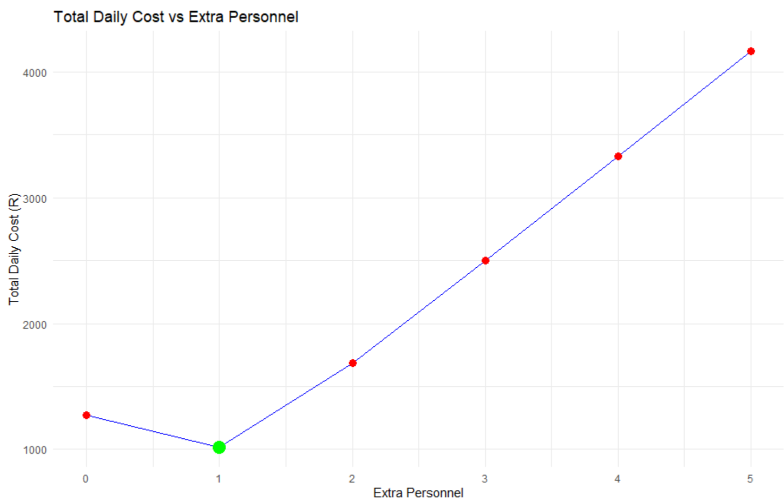

Figure 58: Daily cost when hiring extra personnel

Figure 58 visualises the cost increase with additional personnel. It confirms a clear minimum in costs when hiring 1 extra person, with costs rising steeply thereafter, supporting the quantitative findings.