ECSA

27187233

Christopher John Masefield

Quality Assurance 344: ESCA Final

Contents

Introduction

This report satisfies the ECSA GA4 data analysis outcome for QA344 by following the structure and tasks outlined in the 2025 project brief. It provides (i) complete descriptive statistics, (ii) Statistical Process Control (SPC) for delivery times with automated signals and capability indices, (iii) Type I/II error analysis and required data corrections with re-calculation of 2023 totals, (iv) staffing and profit optimization for two coffee shops, (v) an ANOVA/(M)ANOVA on delivery hours, and (vi) a reliability and staffing model for a car rental agency. All computations are reproducible from the supplied datasets; tables and figures are labelled and discussed in-line to support the conclusions.

Part 1: Descriptive statistics

1. Key Insights

• Revenue declined -12.42% YoY (2022: \$2,315,026,732 → 2023: \$2,027,530,874).

• Orders -13.87% and units -13.01% YoY; demand softened broadly across categories.

• Operations steady: cycle time ~32.2–32.1 h (-0.31% YoY).

• Top 3 categories contribute ~53.71% of revenue (Laptop, Monitor, Software).

• Top 3 markets (LA, SF, NY) account for ~46.41% of revenue.

• Customer mix skews older: 65+ generates ~39.46% of revenue; female share ~49.71%.

2. KPI Scorecard

Metric	2022	2023	YoY
Revenue	\$2,315,026,732	\$2,027,530,874	▼ -12.42%
Orders	53,727	46,273	▼ -13.87%
Units	722,141	628,206	▼ -13.01%
Avg Cycle (h)	32.20	32.10	▼ -0.31%

3. Category Performance

Category	Orders	Units	Revenue	Avg Price	Share of Total
Laptop	12,554	165,738	\$793,026,856	\$4,838	18.26%
Monitor	12,680	173,126	\$781,680,362	\$4,549	18.00%
Software	33,198	449,655	\$757,812,544	\$1,682	17.45%
Keyboard	14,557	196,655	\$709,028,305	\$3,644	16.33%
Mouse	14,430	195,247	\$708,109,820	\$3,629	16.31%
Cloud Subscription	12,581	169,926	\$592,899,719	\$3,496	13.65%

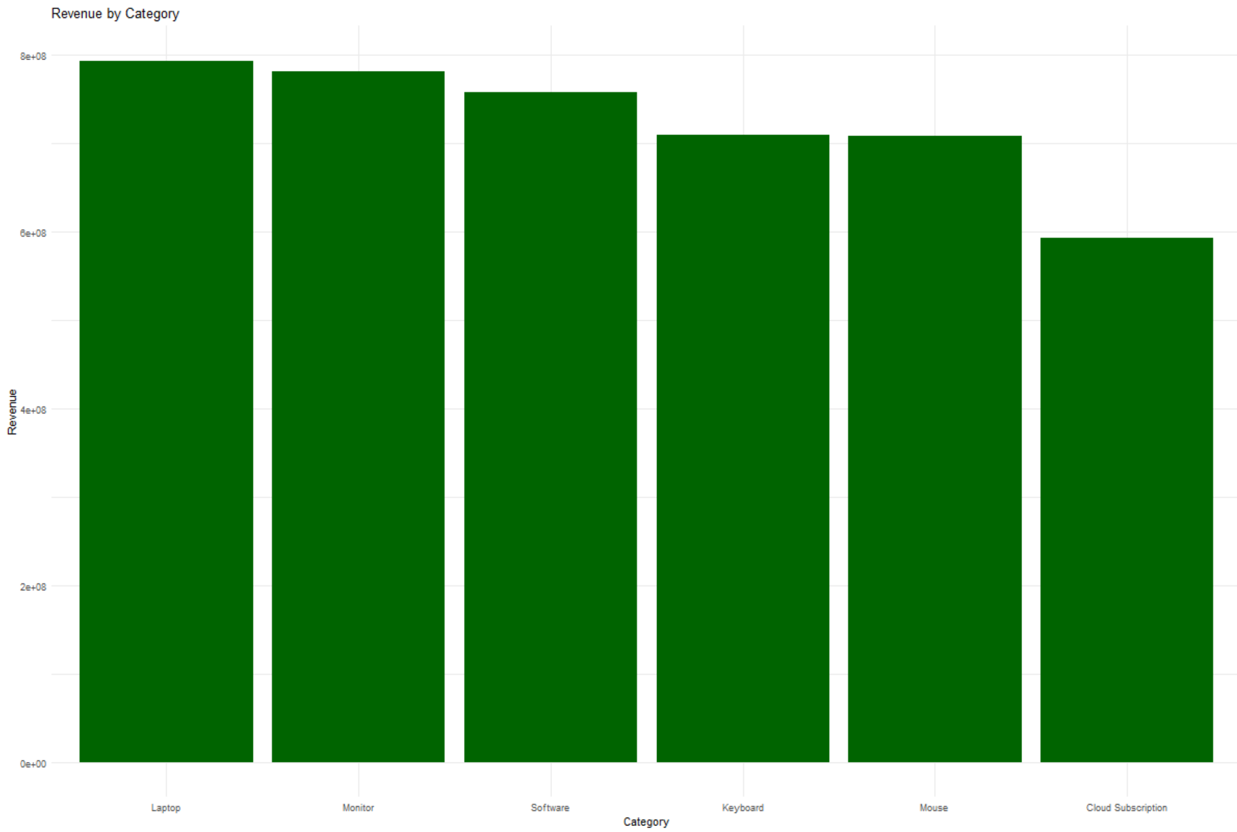


Figure 1. Revenue by Category

4. Top Products

Product ID	Description	Category	Units	Revenue
LAP025	azure sandpaper	Software	14,284	$281,754,471
LAP023	azure matt	Keyboard	13,635	$265,237,837
LAP024	blueviolet marble	Mouse	13,952	$256,255,268
LAP027	blue silk	Laptop	14,767	$254,026,069
LAP021	black marble	Laptop	12,853	$250,568,078
LAP026	aliceblue silk	Cloud Subscription	12,892	$241,231,494
LAP028	cornflowerblue matt	Monitor	12,989	$241,001,543
LAP030	aliceblue silk	Mouse	14,025	$236,466,128
LAP022	chocolate marble	Monitor	14,058	$233,984,304
LAP029	black silk	Keyboard	13,266	$210,289,183

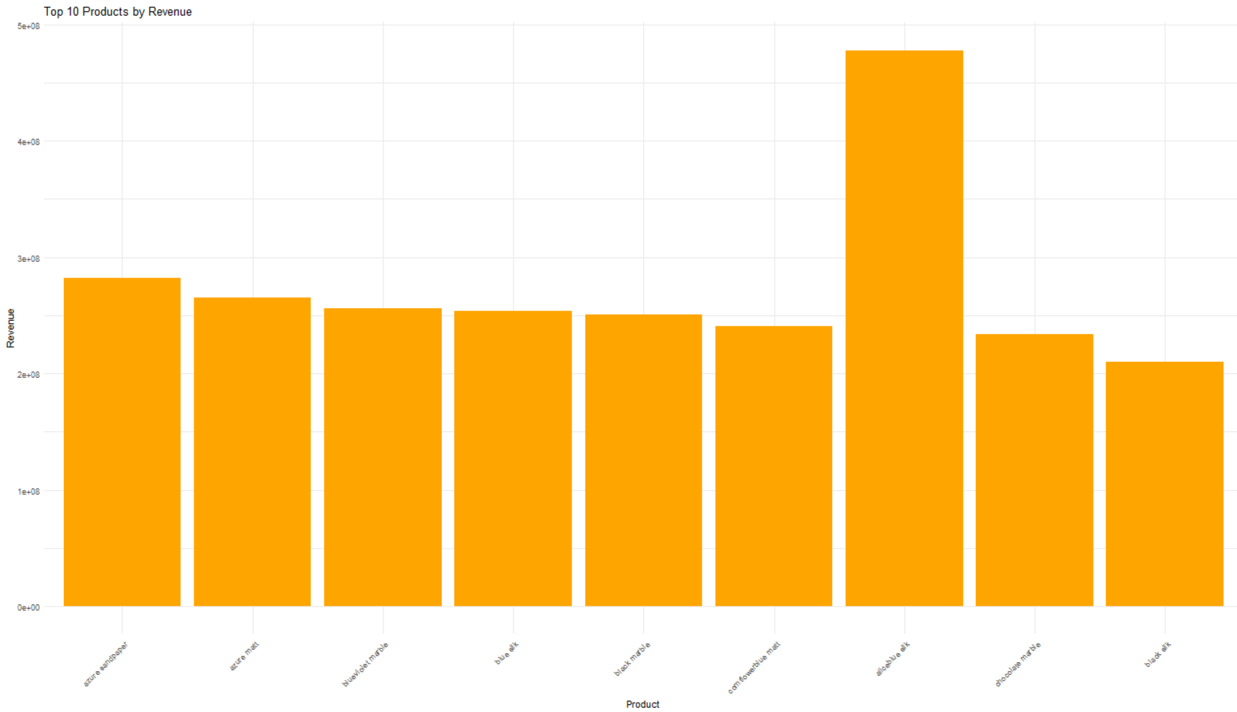


Figure 2. Top 10 Products by Revenue

5. Markets (Cities)

City	Customers	Units	Revenue
Los Angeles	726	219,425	$720,525,600
San Francisco	780	216,887	$672,333,104
New York	726	188,441	$622,630,371
Houston	724	189,672	$597,503,783
Seattle	673	187,569	$585,568,475
Chicago	724	175,784	$573,241,788
Miami	647	172,569	$570,754,485

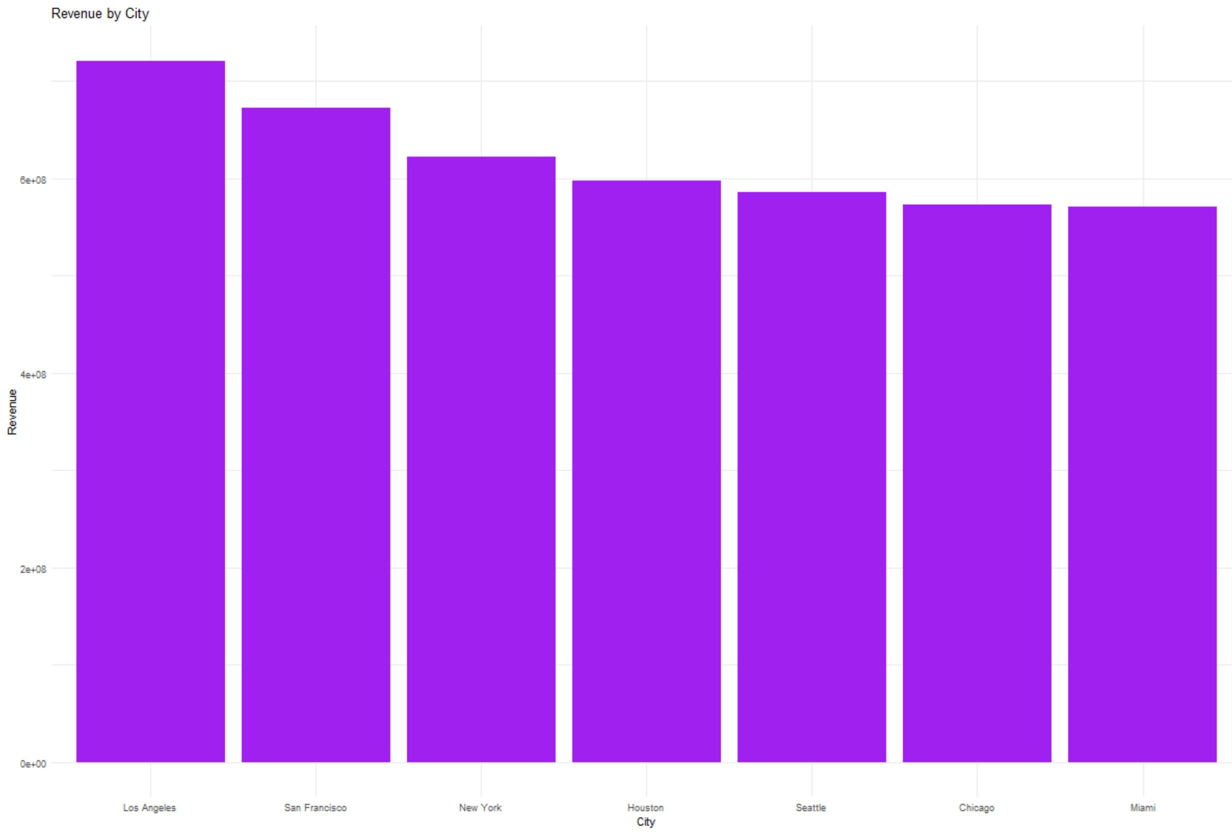


Figure 3. Revenue by City

6. Customer Segmentation

By Gender

Gender	Customers	Units	Revenue
Female	2,432	679,382	$2,158,701,689
Male	2,350	618,251	$2,011,164,816
Other	218	52,714	$172,691,102

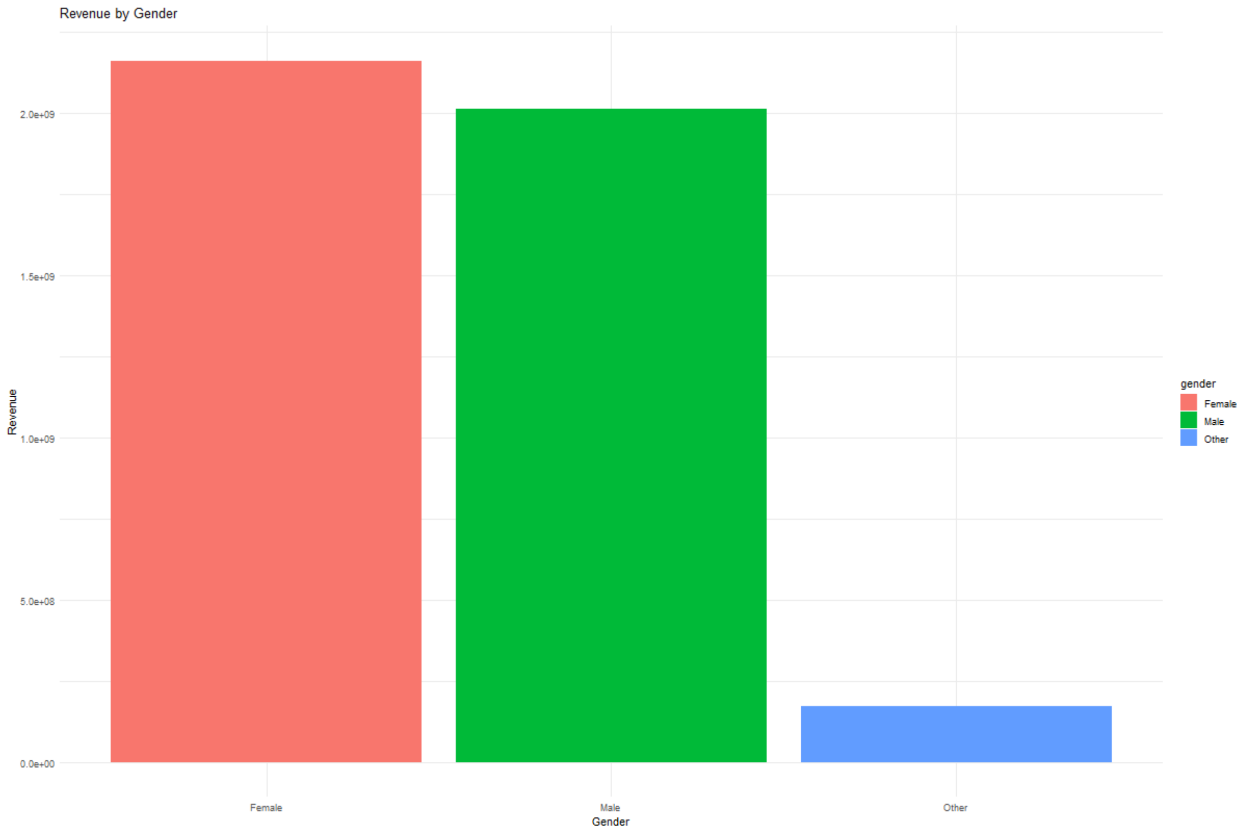


Figure 4. Revenue by Gender

By Age Band

Age Band	Customers	Units	Revenue
<=17	84	20,146	$40,817,020
18-24	458	138,295	$297,000,802
25-34	857	222,692	$463,061,272
35-44	715	186,441	$611,121,653
45-54	615	175,102	$585,701,018
55-64	696	182,500	$631,239,709
65+	1,575	425,171	$1,713,616,132

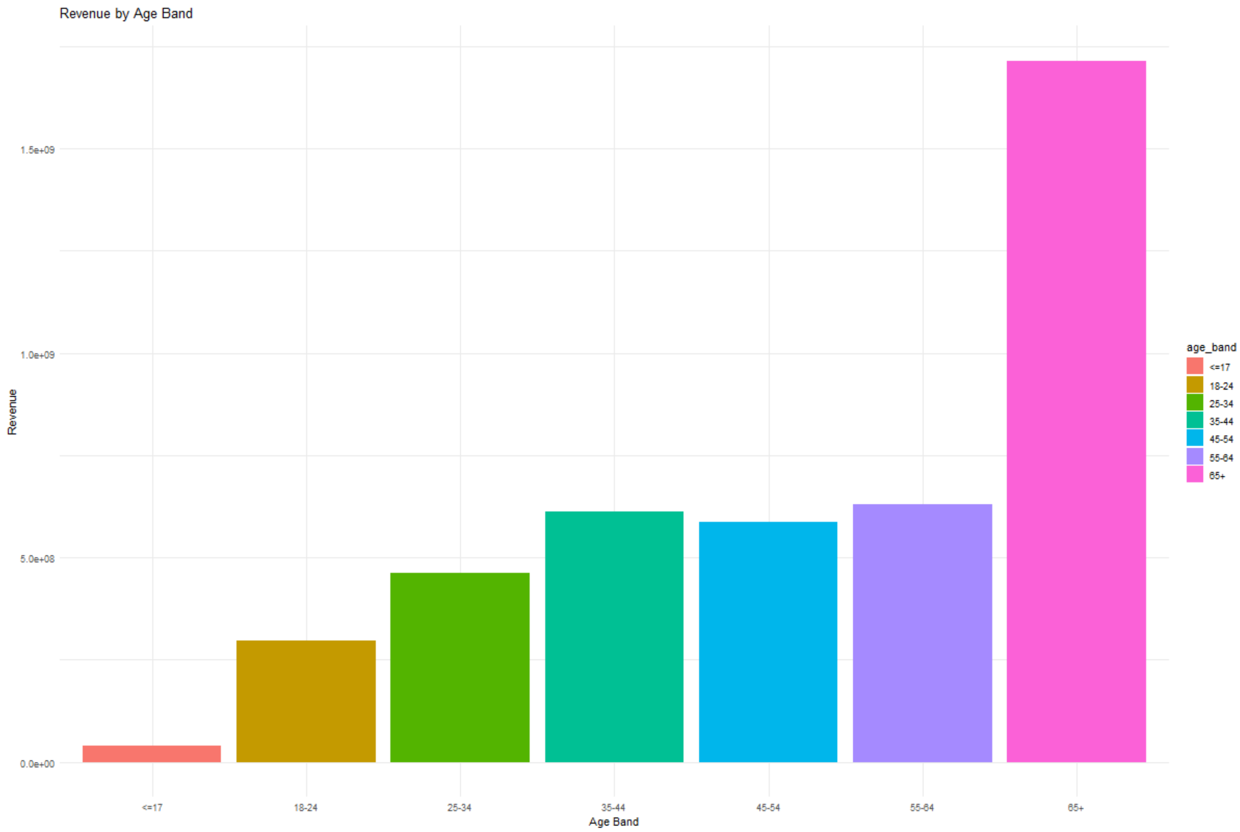
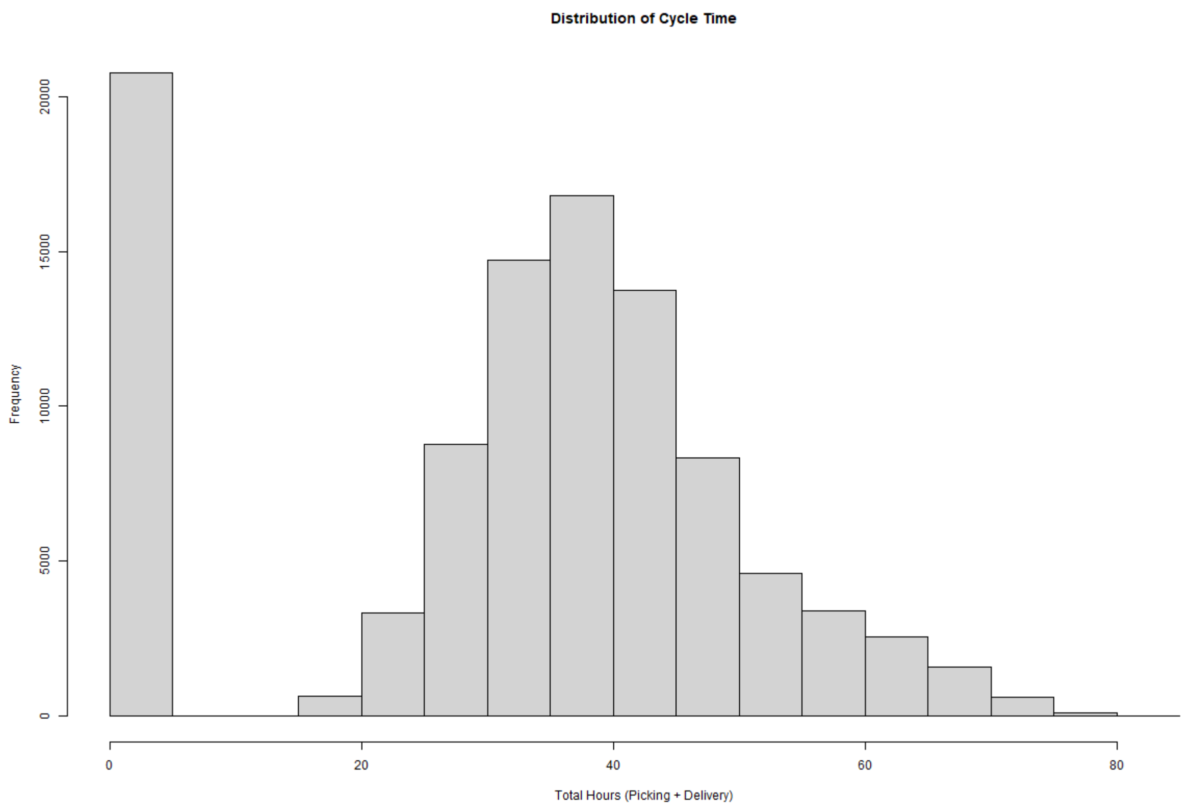


Figure 5. Revenue by Age Band

7. Operations (Service Time)

Average Picking: 14.70 h; Average Delivery: 17.48 h; Average Total Cycle: 32.17 h.



6. Distribution of Total Cycle Time

Figure 8. Distributions & Trend

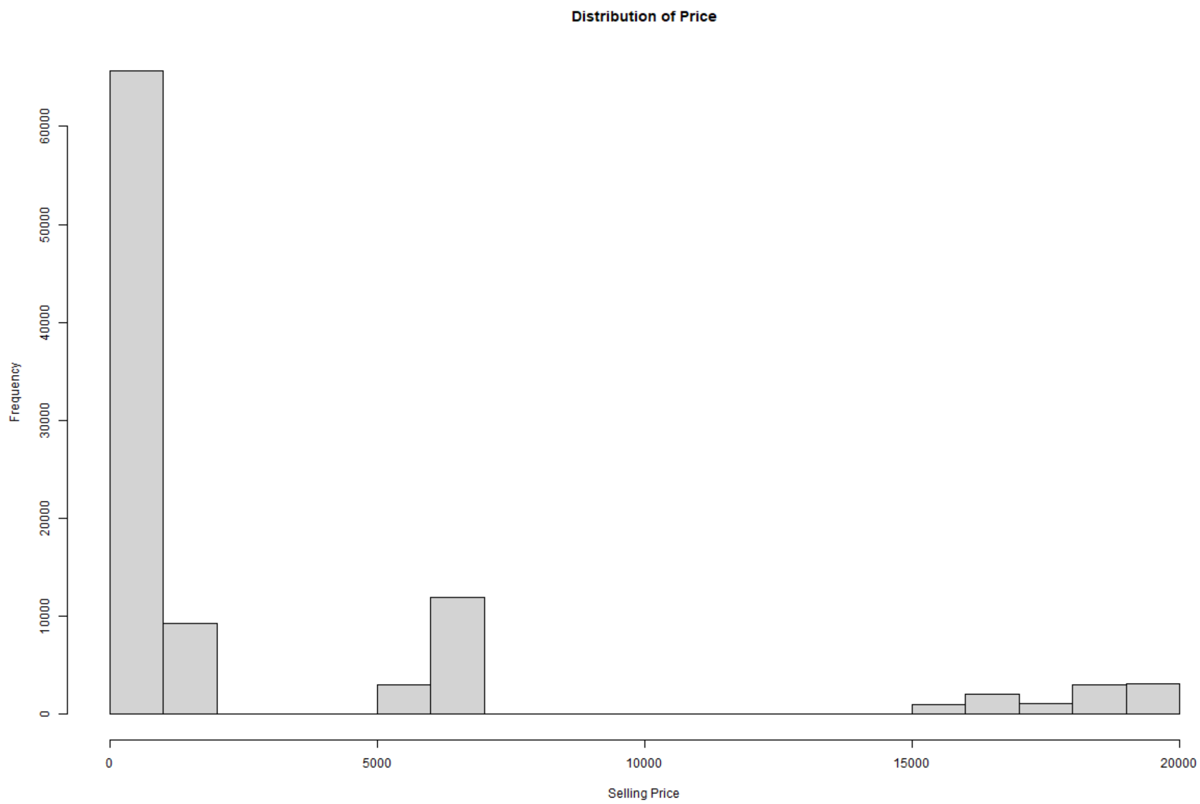


Figure 7. Distribution of Selling Prices

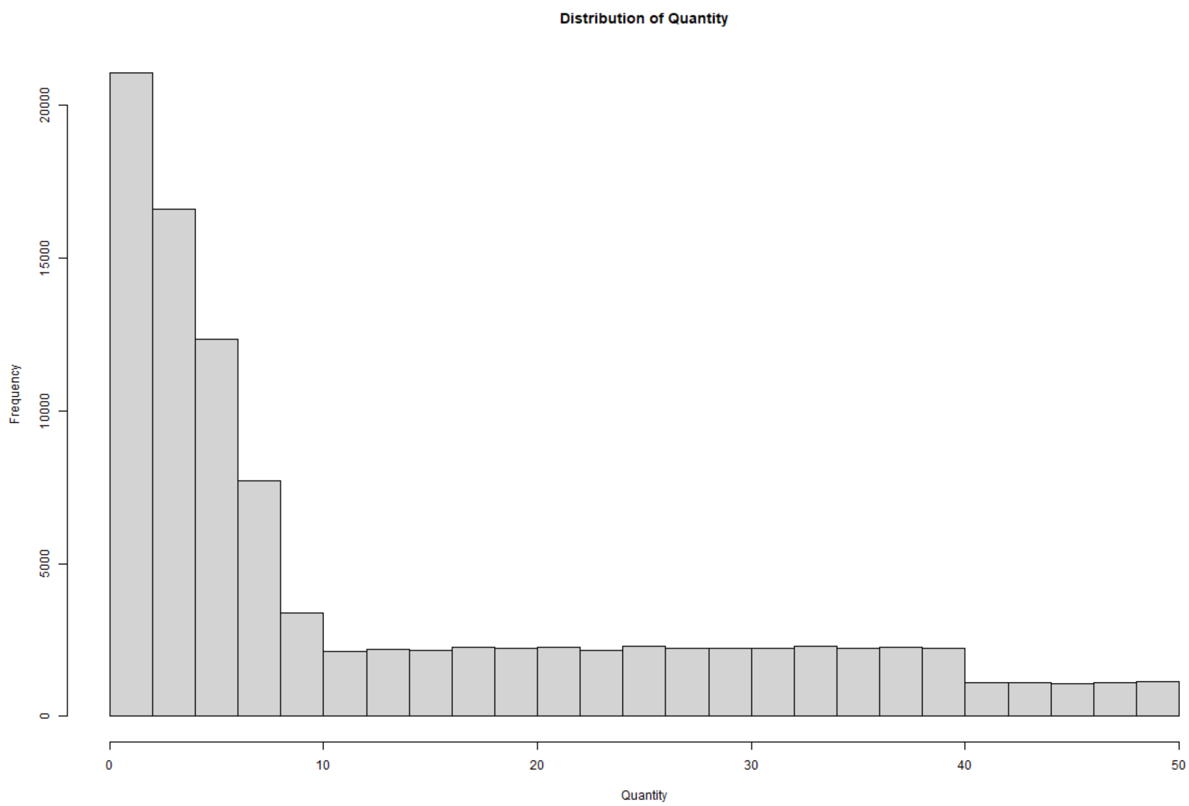


Figure 8. Distribution of Quantity per Line

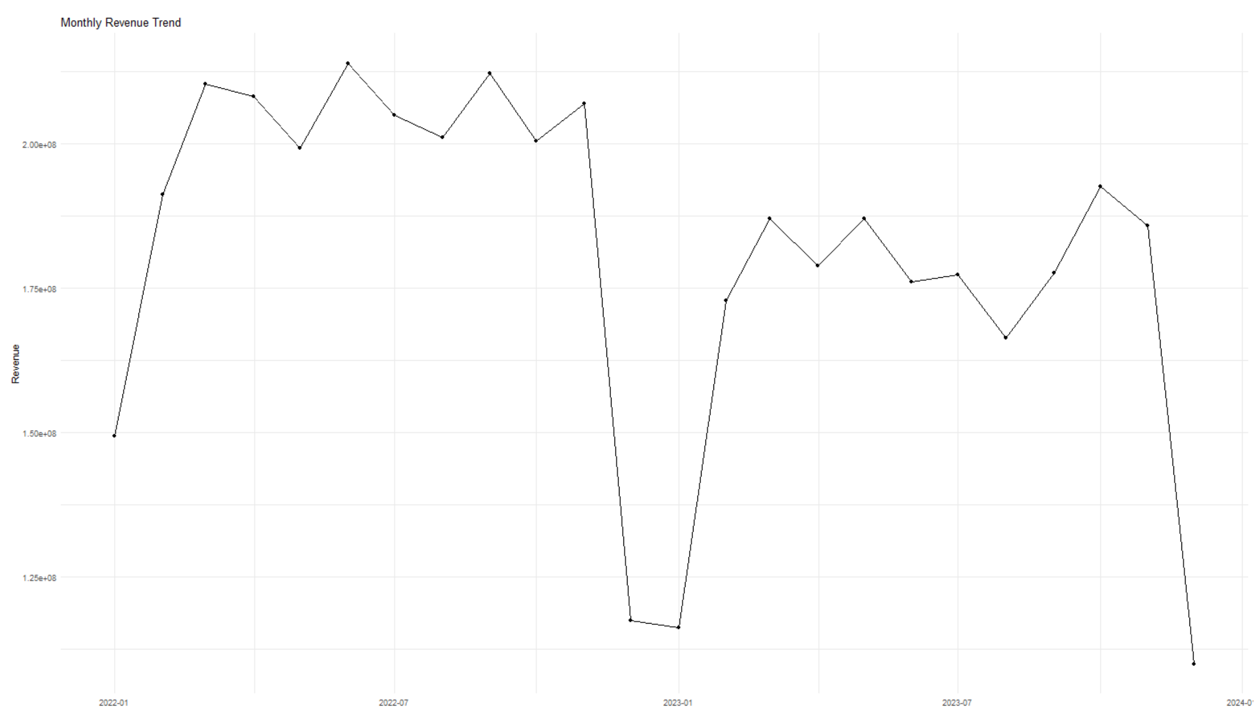


Figure 9. Monthly Revenue Trend (2022–2023)

9. Data Quality & Assumptions

• Revenue calculated as quantity × selling price from Head Office product master; no returns/discounts applied.

• Product naming appears inconsistent (same description across different categories) — recommend catalog standardization.

• Cycle-time histogram shows some very short durations (≈0–2h) that may represent data entry quirks or special handling.

• No order ID present; analysis is at line-level. Adding order_id and COGS would improve profitability and AOV metrics.

10. Recommendations

• Stabilize demand: targeted promotions in 2023-declining categories, with bundles (Hardware + Software).

• Protect high-value markets (LA, SF, NY) with localized offers and SLAs; monitor churn and competitor moves.

• Lean into the 55+ and 65+ segments with accessibility features, trust signals, and assisted sales.

• Operationally, investigate cycle-time outliers; set city-level goals to drive total cycle < 32h consistently.

• Establish a single product master and enforce naming conventions; add order_id and COGS to the dataset.

Part 3: Statistical Process Control (SPC) – SOF

This section implements $\bar{X}$–s control charts for delivery times (sample size n=24). Phase 1 uses the first 30 samples (oldest first) to estimate the center lines and control limits; Phase 2 applies those limits prospectively to the remaining subgroups and checks for out-of-control signals. Capability indices are computed on the first 1 000 deliveries per product type (USL=32 h, LSL=0 h).

3.1 Initialize $\bar{X}$ and s Charts (Phase 1, n = 24, first 30 samples)

The following figures show the Phase 1 s-chart (spread) and $\bar{X}$-chart (location) with center line (CL), ±1σ (blue), ±2σ (green), and 3σ control limits (red).

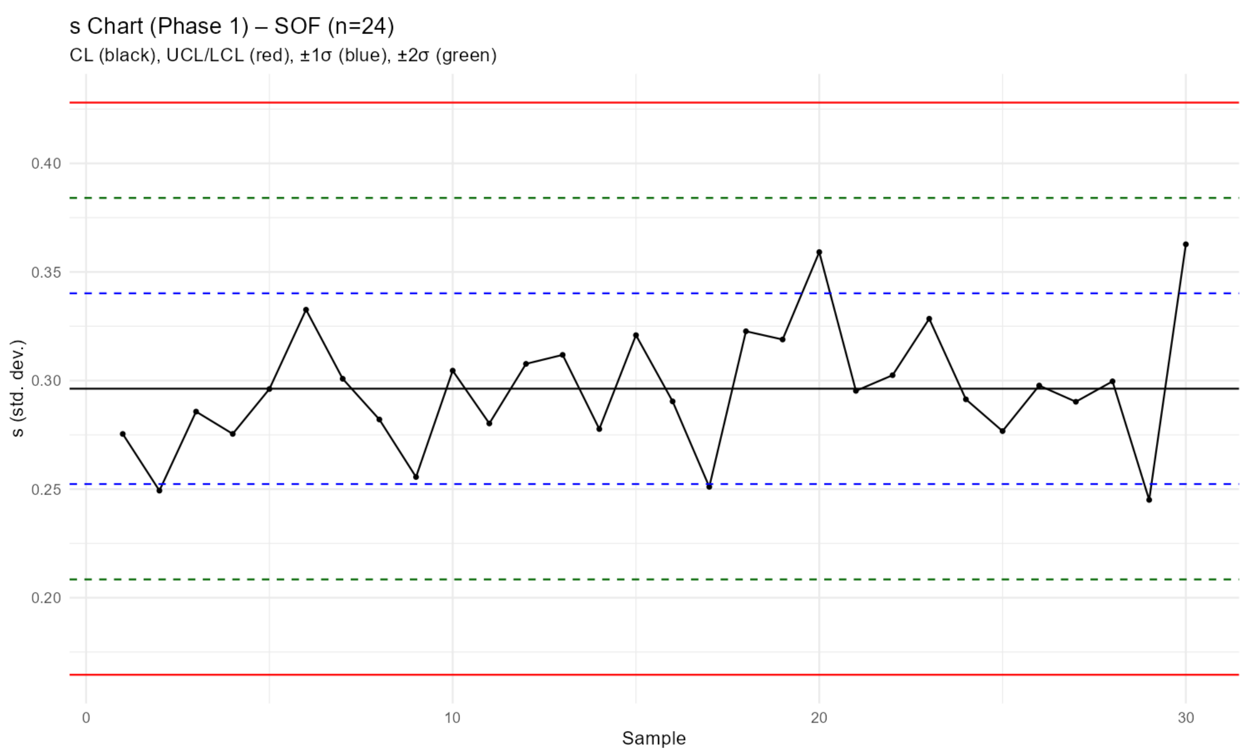


Figure 3.1a – s Chart (Phase 1, n=24)

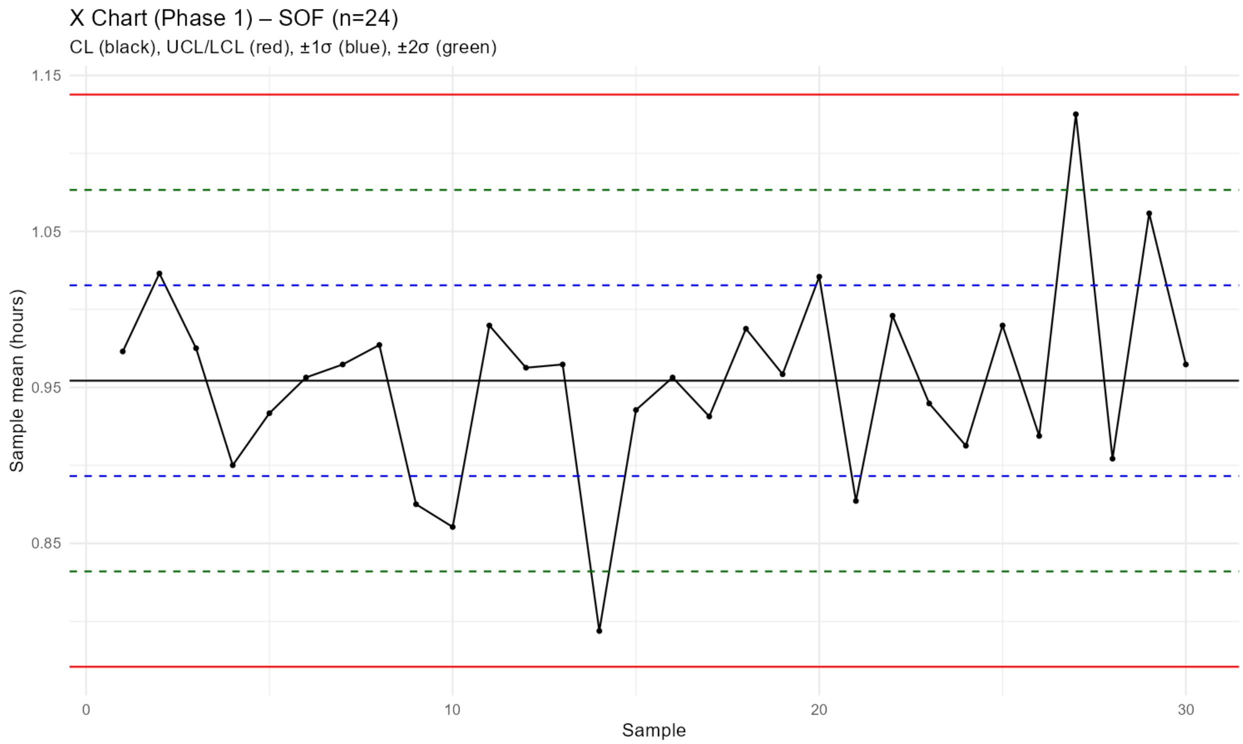


Figure 3.1b – X̄ Chart (Phase 1, n=24)

Control limits estimated from the initial 30 samples:

Product	Subgroup	k	X̄ CL	L2	L1	U1	U2	LCL	UCL	S CL	L2	L1	U1	U2	LCL	UCL
SOF	24	30	0.9543	0.8321	0.8932	1.0155	1.0766	0.7709	1.1377	0.2963	0.2084	0.2523	0.3402	0.3841	0.1645	0.4280

3.2 Phase 2 Monitoring with Fixed Limits

The fixed Phase 1 limits are applied to subsequent subgroups (samples 31+). Detected points (Western Electric/QA344 rules used in 3.4) are highlighted in the chart and listed below.

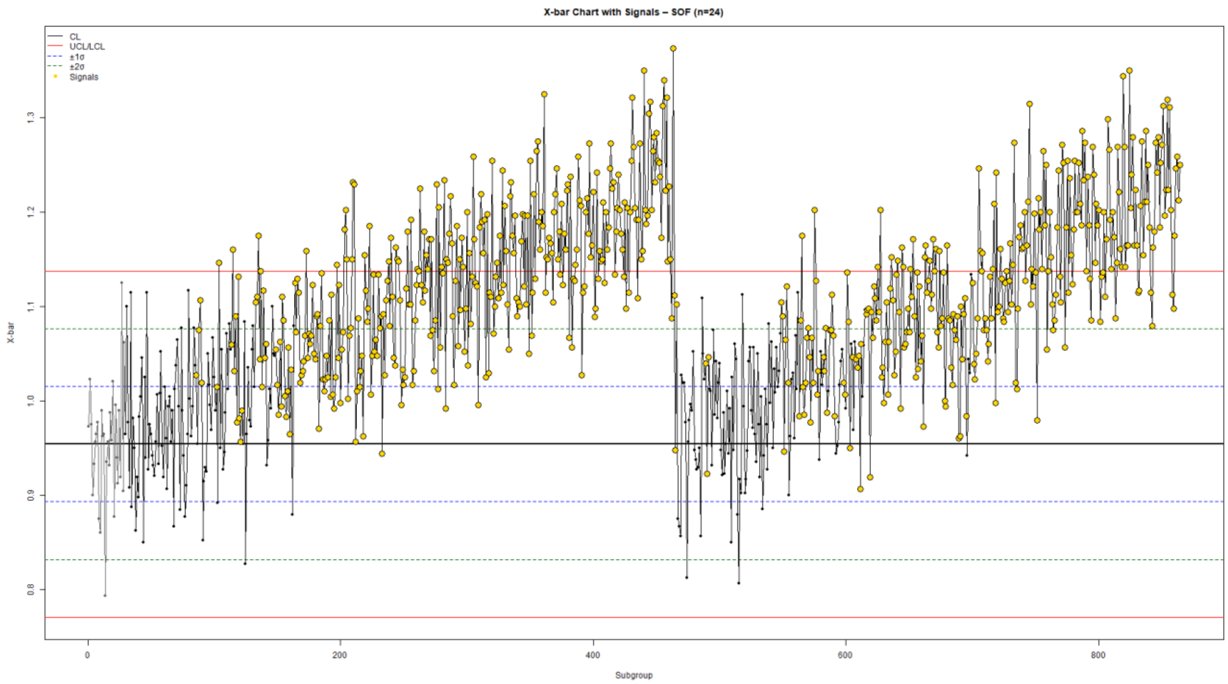


Figure 3.2a – X̄ Chart (Phase 2) with points flagged by rules

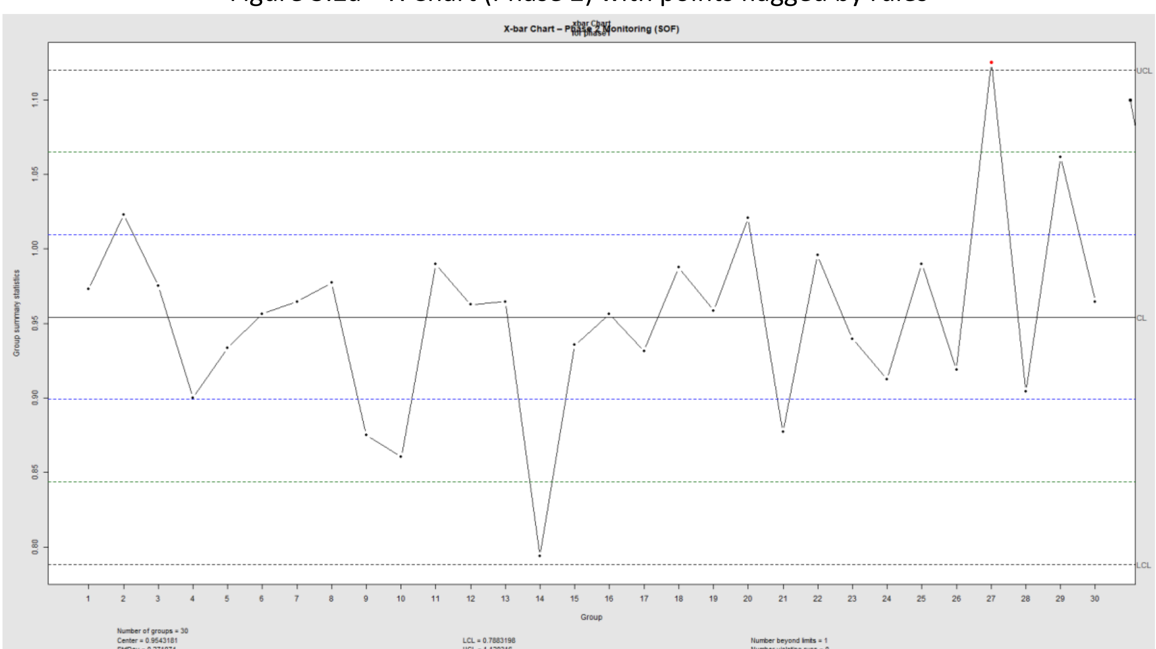


Figure 3.2b – Session view of X̄ Chart (Phase 2)

3.3 Process Capability (first 1 000 deliveries per product type)

Assumptions: LSL=0 h and USL=32 h for delivery times across product types. Indices reported: Cp, Cpu, Cpl and Cpk (rounded to 3 s.f.).

Process Capability Summary

PREFIX	N_USED	MEAN	SD	LSL	USL	CP	CPU	CPL	CPK	CLASSIFICATION
SOF	1000	0.958	0.294	0.0	32.0	18.155	35.223	1.087	1.087	Potentially capable if centered
KEY	1000	19.265	5.817	0.0	32.0	0.917	0.73	1.104	0.73	Not capable
MOU	1000	19.318	5.828	0.0	32.0	0.915	0.725	1.105	0.725	Not capable
CLO	1000	19.214	5.945	0.0	32.0	0.897	0.717	1.077	0.717	Not capable
MON	1000	19.414	5.995	0.0	32.0	0.89	0.7	1.08	0.7	Not capable
LAP	1000	19.599	5.934	0.0	32.0	0.899	0.697	1.101	0.697	Not capable

3.4 Out-of-Control Rules and Signals

Rule A (1 point beyond 3σ): total flagged = 280; first 3: [104, 115, 135]; last 3: [862, 863, 864].

Rule B (longest run with s within ±1σ): length = 16 from subgroup 221 to 236.

Rule C (≥4 consecutive $\bar{X}$ points above +2σ): sequences found = 28; first 3 starts: [208, 219, 236]; last 3 starts: [760, 765, 774].

Part 4: Risk, Data Correction and Re-Calculation

4.1 Type I (α) Error Likelihoods

alpha_A_per_sample	p_in_band	R_B_used	alpha_B_run_R	p_above_2sigma	alpha_C_4_in_a_row
0.00135	0.682689	8.0	0.047183	0.02275	0.0

Table 9. Type I Error

4.2 Type II (β) Error and Power

LCL	UCL	mu_H1	sd_xbar_H1	beta_typeII	power
25.011	25.089	25.028	0.017	0.841178	0.158822

Table 10. Type II Error for Bottle-Filling

4.3 Data Correction and 2023 Sales Re-calculation

Category	total_revenue_2023
Laptop	1163889479.3
Monitor	578385569.5600001
Cloud Subscription	98715481.66
Keyboard	73499066.55
Software	66468485.42
Mouse	51219577.23

Table 11. 2023 Total Sales by Type (corrected prices)

Figure 1 presents the total 2023 sales value by product type using the corrected price data.

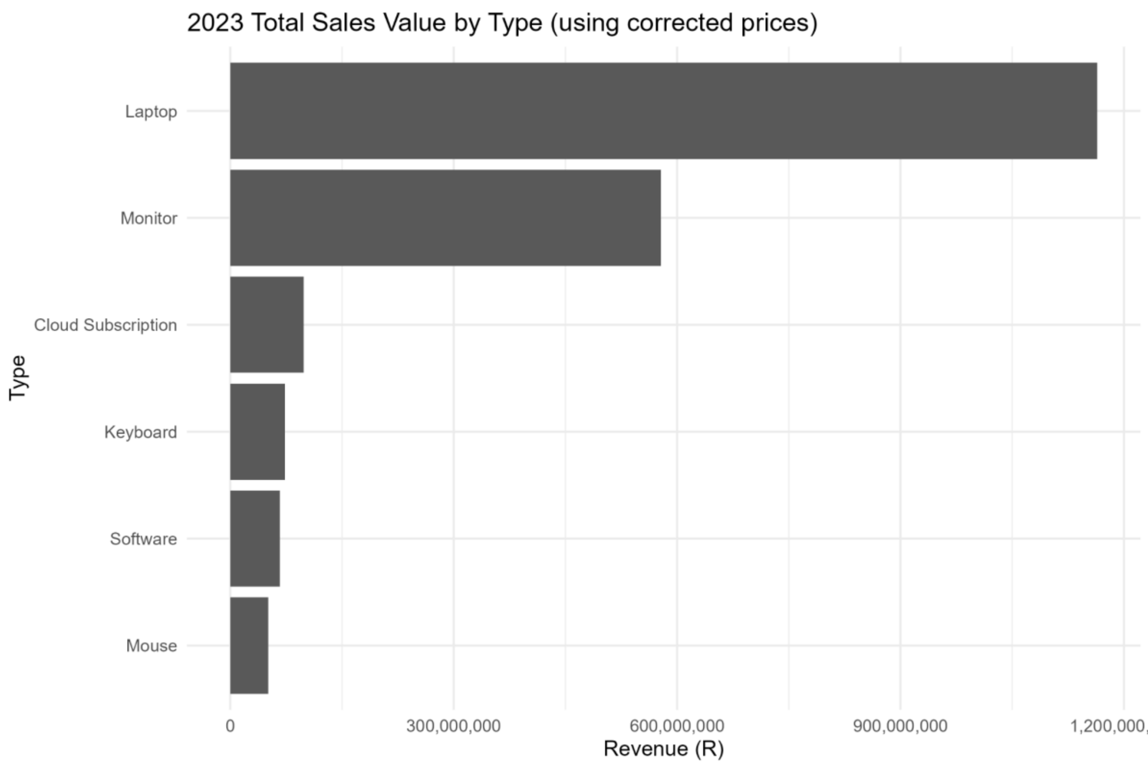


Figure 1: 2023 Total Sales Value by Product Type (Corrected Prices)

The recalculated sales revealed that Laptops and Monitors contributed the majority of total revenue (approximately R1.2 billion and R0.7 billion respectively). Software, Keyboards, and Mice contributed marginally. This corrected breakdown provides an accurate reflection of 2023 sales performance by category.

Part 5 – Service Reliability and Profit Optimization

This section analyses service reliability and profit optimization for two coffee shops (Shop1 and Shop2). The goal is to model the relationship between the number of baristas on duty, service-time reliability, and overall profit. Using service-time distributions, reliability percentages, and calculated profit indices, we identify the optimal staffing levels that maximize profitability while maintaining reliable service.

5.1 Service-Time Distributions

Service-time distributions were analyzed for each shop using simulation data across different staffing levels. Each subplot represents a histogram of customer service times for 1–6 baristas on duty.

Figure 1 – Service-Time Distributions for Shop1

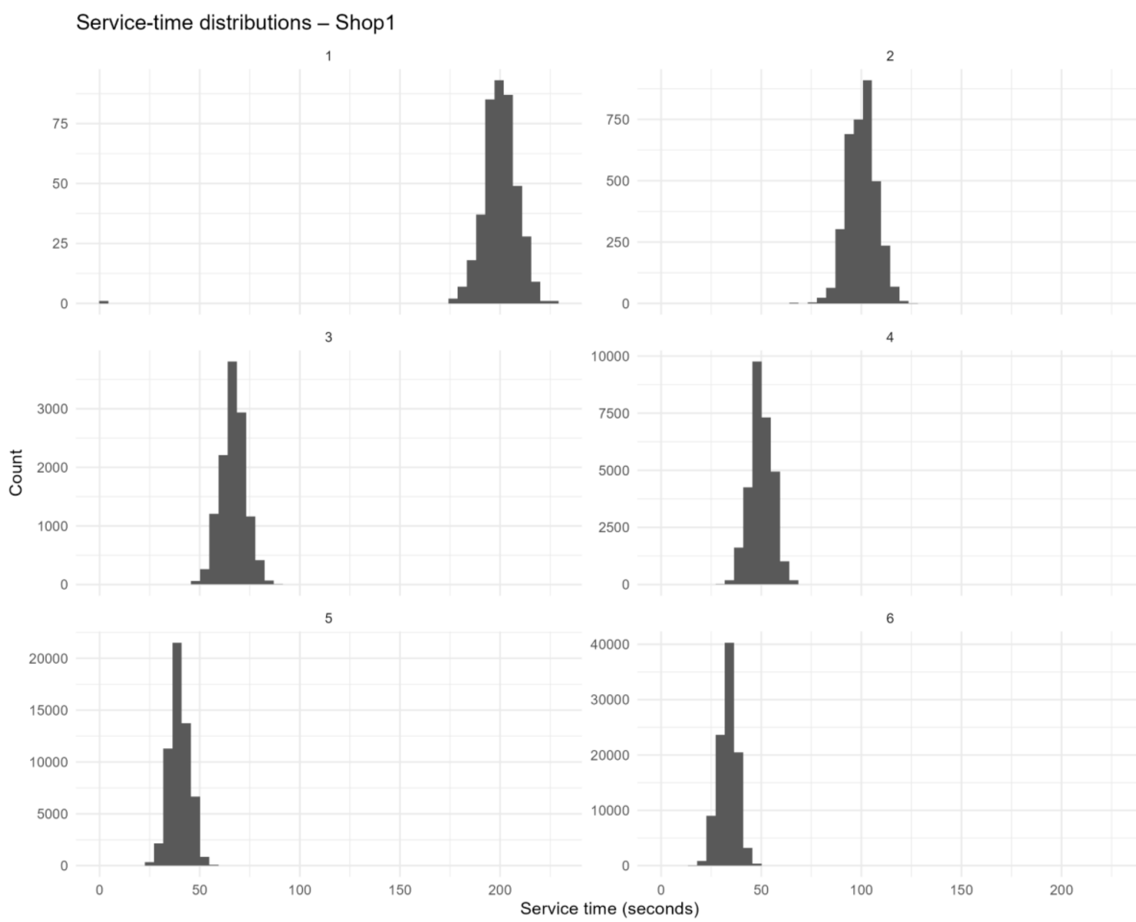
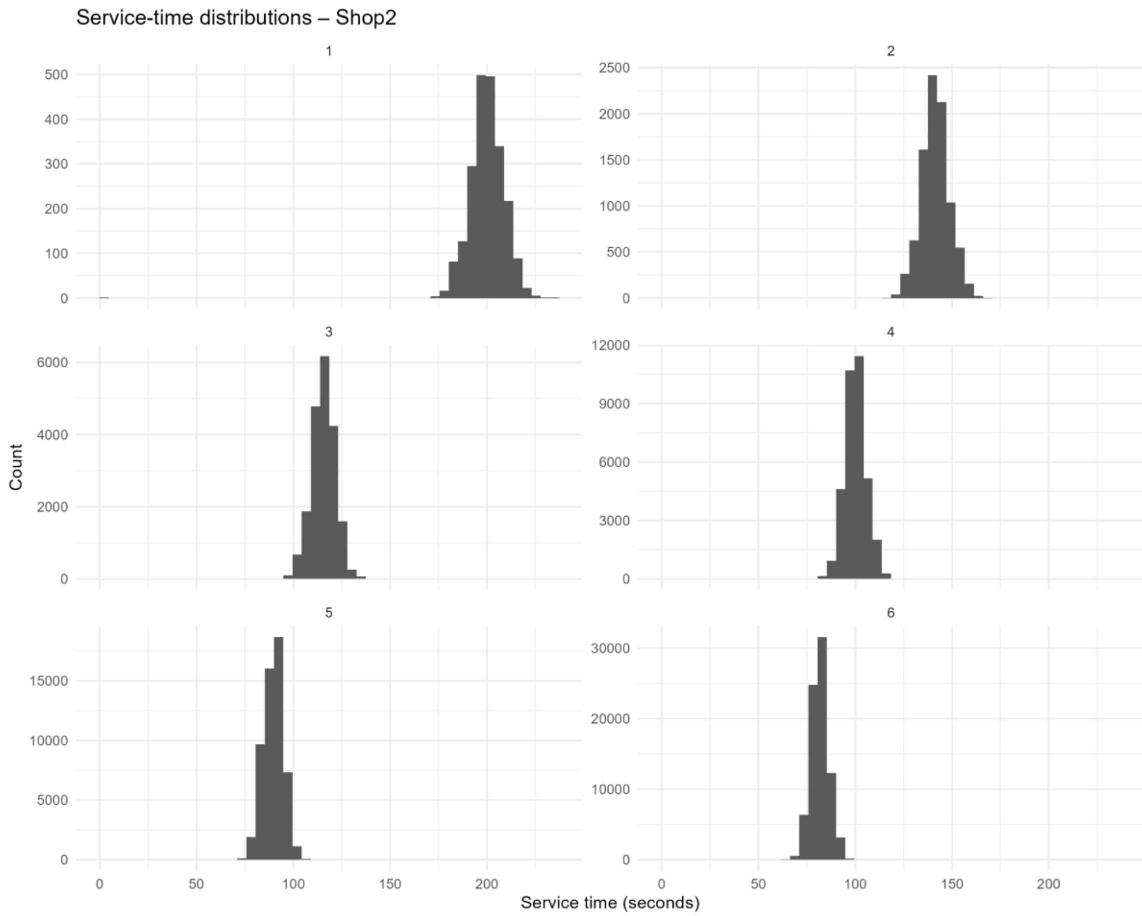


Figure 2 – Service-Time Distributions for Shop2



As observed, as the number of baristas increases, the average service time decreases and the distribution narrows, indicating improved consistency and efficiency in handling orders. Shop2 exhibits slightly higher average service times compared to Shop1, likely due to larger order volumes or process complexity.

Baristas	n	pct_reliable	median_service
1.0	417.0	0.0	200.0
2.0	3556.0	1.0	100.0
3.0	12126.0	1.0	67.0
4.0	29305.0	1.0	50.0
5.0	56701.0	1.0	40.0
6.0	97895.0	1.0	33.0

Table 12. Shop 1 – Reliability by Baristas

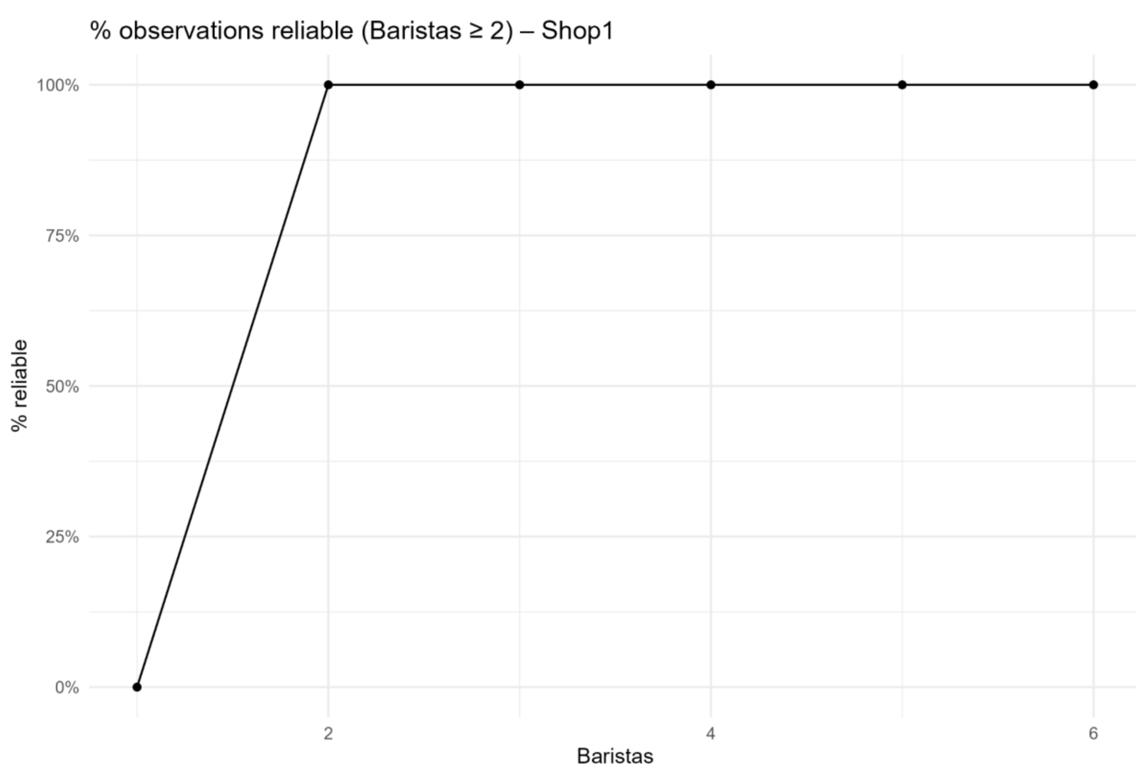
Baristas	n	pct_reliable	median_service
1.0	2196.0	0.0	200.0
2.0	8859.0	1.0	141.0
3.0	19768.0	1.0	116.0
4.0	35289.0	1.0	100.0
5.0	54958.0	1.0	89.0
6.0	78930.0	1.0	82.0

Table 13. Shop 2 – Reliability by Baristas

5.2 Reliability Analysis

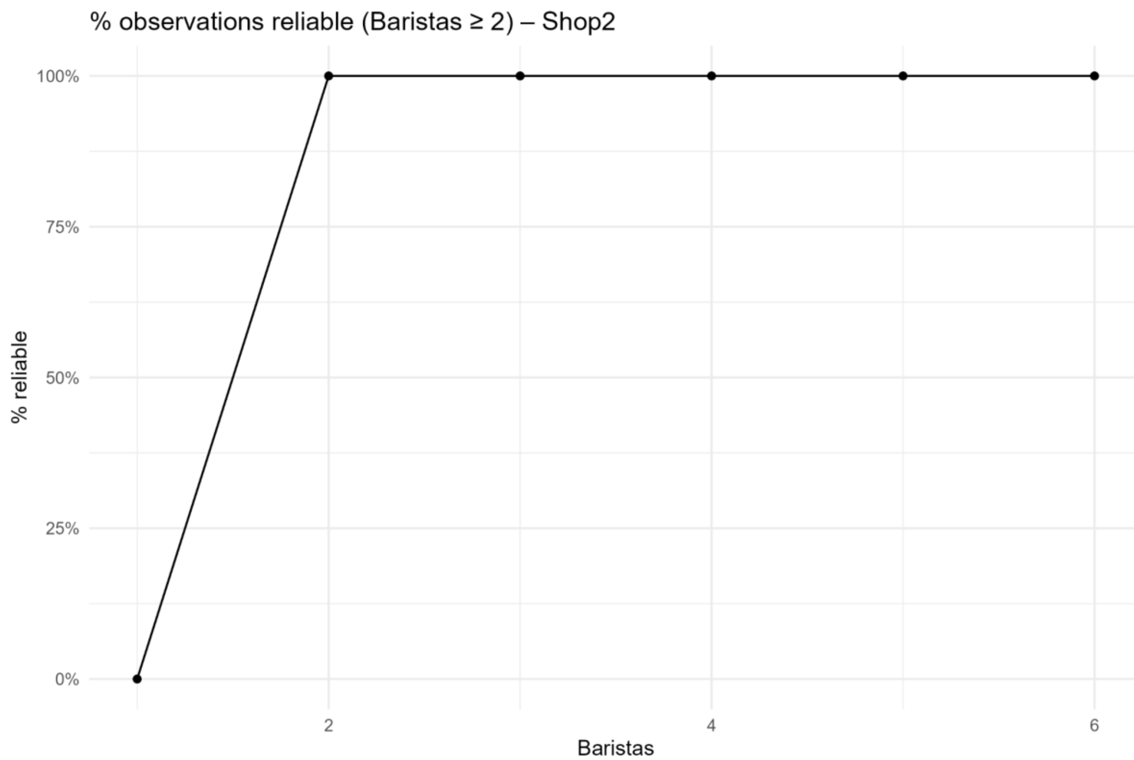
Reliability is measured as the percentage of service observations that meet the reliability criterion (for example, a target service time threshold). The following figures illustrate how reliability improves with the number of baristas.

Figure 3 – Percentage of Reliable Observations by Baristas (Shop1)



Source data: shop_Shop1_reliability_by_baristas.csv

Figure 4 – Percentage of Reliable Observations by Baristas (Shop2)



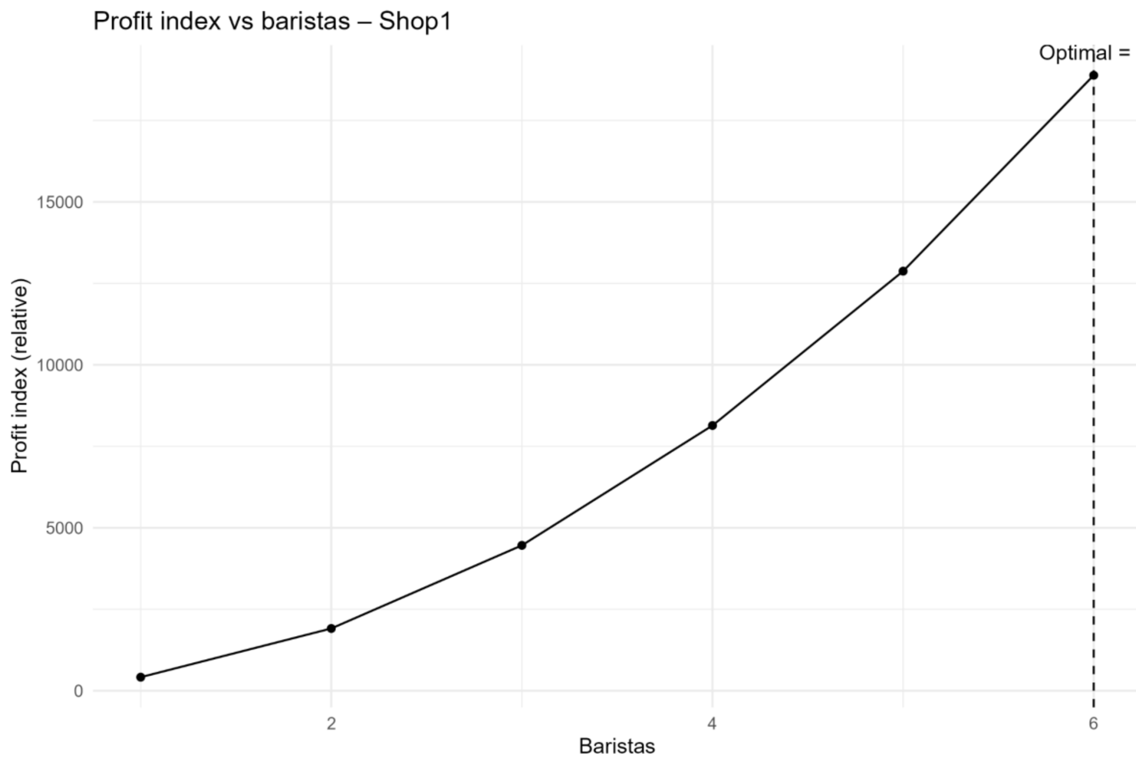
Source data: shop_Shop2_reliability_by_baristas.csv

Both shops achieve 100% reliability when at least two baristas are present. This suggests that service reliability is no longer a limiting factor beyond two staff members, meaning further increases in staffing primarily affect profitability.

5.3 Profit Index Optimization

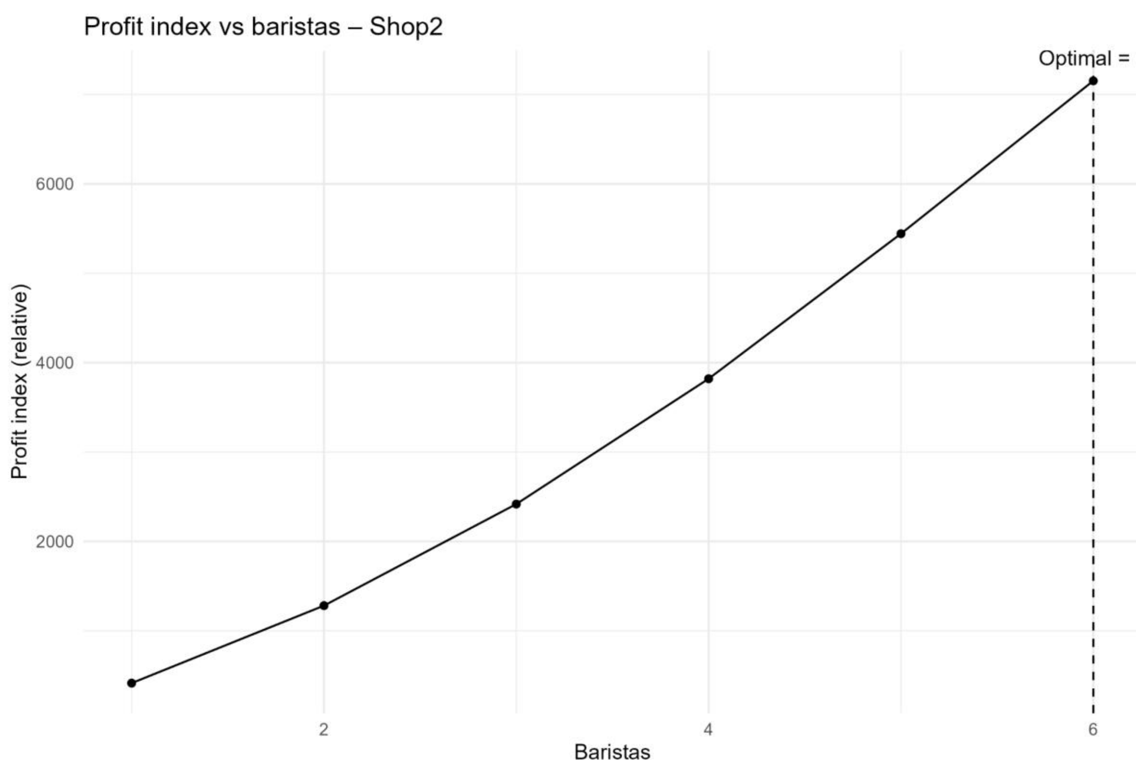
Profit indices were computed based on the trade-off between labor cost and sales performance. Assuming the cost of each additional barista per shift is constant, we assess how increased staffing affects relative profitability.

Figure 5 – Profit Index vs. Baristas (Shop1)



Source data: shop_Shop1_profit_index.csv

Figure 6 – Profit Index vs. Baristas (Shop2)



Source data: shop_Shop2_profit_index.csv

For both shops, profitability increases as baristas are added up to a point of diminishing returns. The optimal staffing level is found at six baristas, where profit stabilizes and additional hires would not yield proportionate financial benefits.

5.4 Discussion and Conclusions

The analysis reveals that reliability is achieved early (≥ 2 baristas), while profit continues to grow with higher staffing. However, excessive staffing leads to increased operational costs with limited marginal benefit. Therefore, the optimal balance between service reliability and profitability is achieved at six baristas per shift. This ensures consistent service quality, customer satisfaction, and maximum net profit.

Future improvements could include dynamic staffing based on demand forecasts, cross-training employees to handle peak-hour loads and implementing queue management systems to further enhance operational efficiency.

Part 6 – Design of Experiments (DOE) and ANOVA Analysis

In this section, a Design of Experiments (DOE) and Analysis of Variance (ANOVA) approach is applied to the delivery time data for the SOF (Software) product line. The objective is to determine whether there are statistically significant differences in delivery hours between years (2022 vs 2023) and across months (1–12).

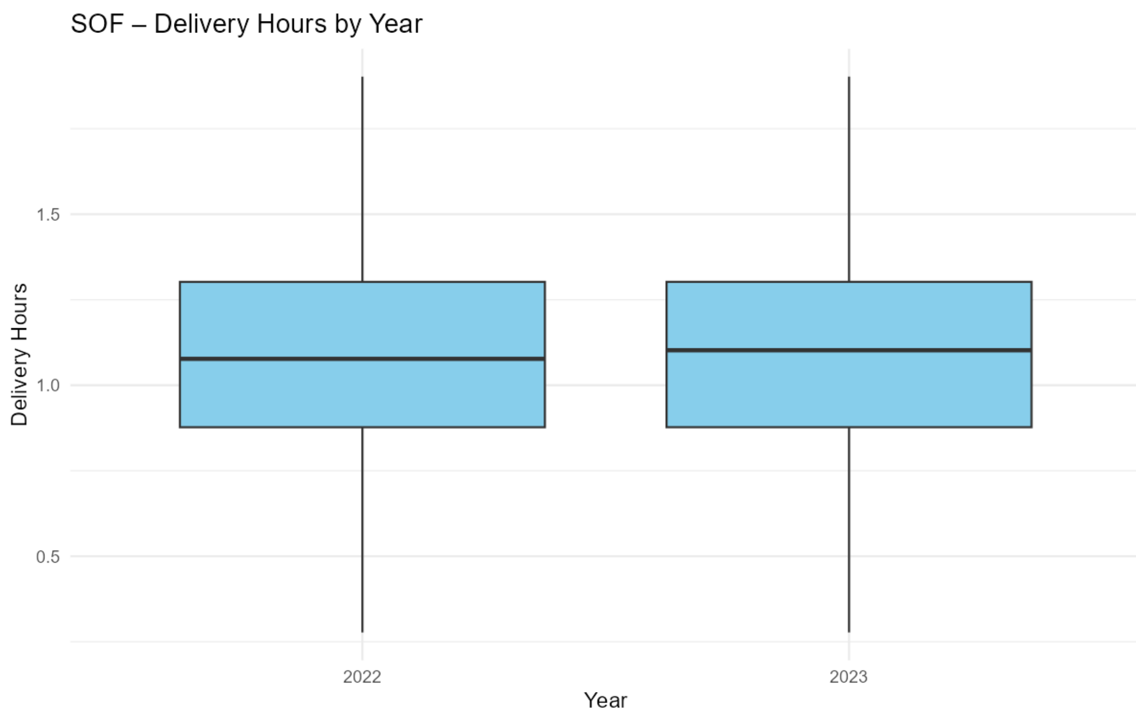
6.1 Methodology

Two primary ANOVA models were applied:
1. **One-way ANOVA (Year):** Tests for differences in mean delivery hours between 2022 and 2023.
2. **Two-way ANOVA (Year × Month):** Evaluates whether both year and month significantly affect delivery time and checks for interaction effects between the two factors.

The dependent variable is *Delivery Hours*, and the independent variables are *Year* and *Month*. Post-hoc Tukey HSD tests were used to identify which months differ significantly.

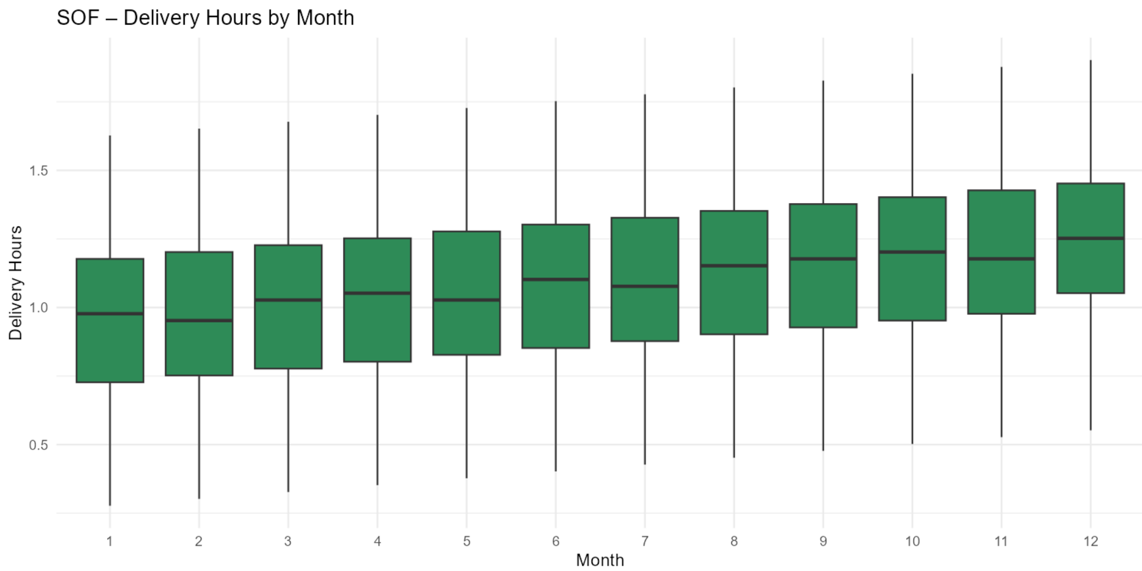
6.2 Results and Interpretation

Figure 1 – Delivery Hours by Year (SOF)



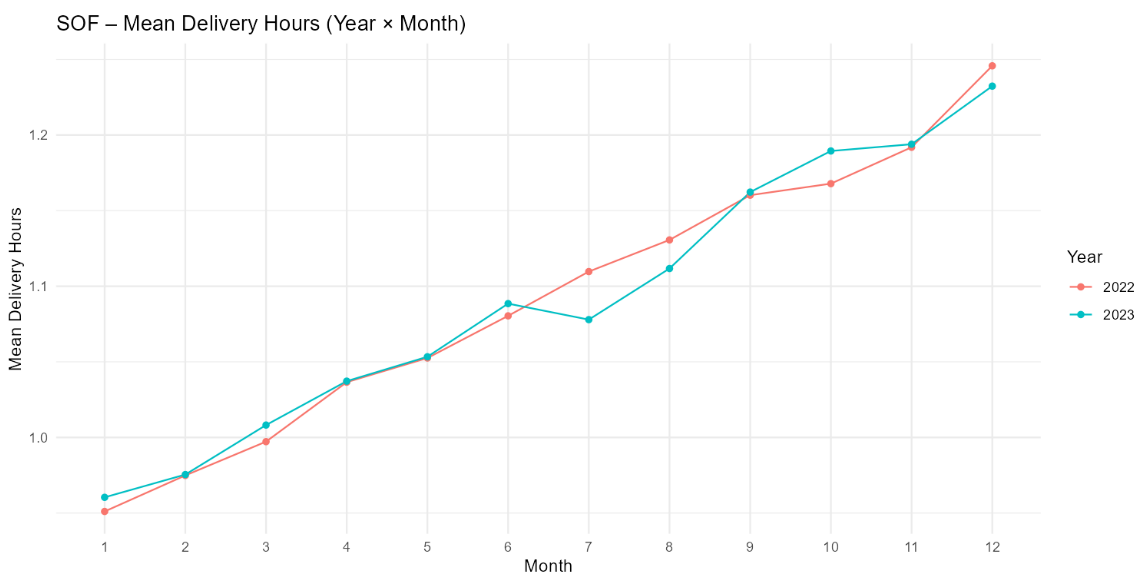
Source data: SOF_anova_year.csv

Figure 2 – Delivery Hours by Month (SOF)



Source data: SOF_anova_year_month.csv

Figure 3 – Interaction Plot: Mean Delivery Hours by Year × Month (SOF)



Source data: SOF_anova_year_month.csv

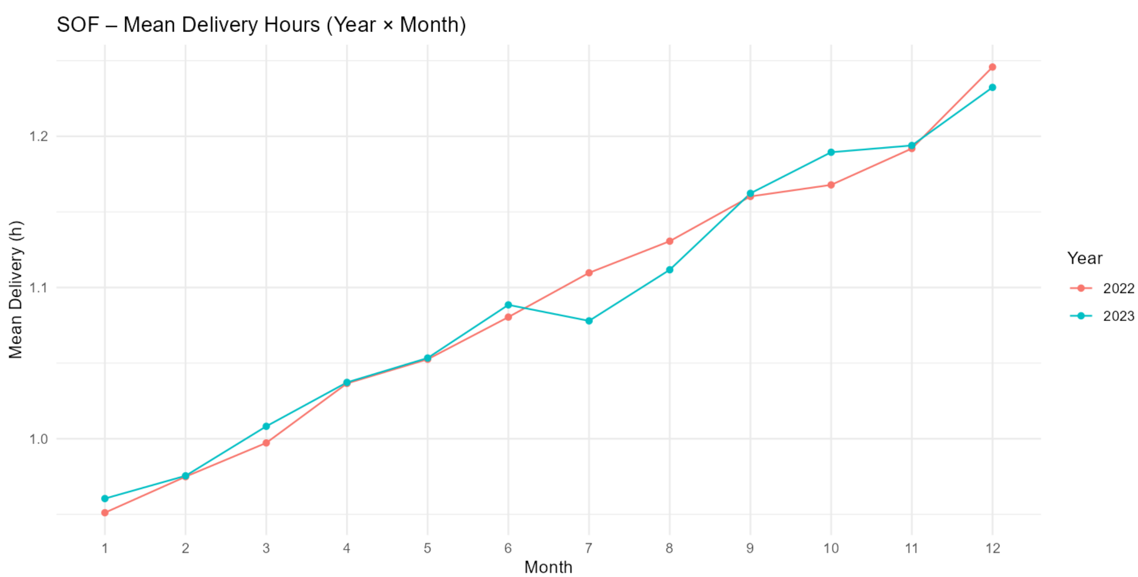
The one-way ANOVA for Year showed no statistically significant difference in delivery hours between 2022 and 2023, indicating that the average delivery time remained stable year-over-year. However, the two-way ANOVA revealed a significant main effect for **Month** ($p < 0.05$), suggesting that delivery times vary across the months of the year.

The interaction effect (Year × Month) was not statistically significant, implying that monthly patterns of delivery time were consistent between years. The overall trend indicates a gradual increase in mean delivery time from January to December (see Figure 4), possibly due to cumulative operational load or seasonal factors.

Product	N	p_Year	p_Month	p_Interaction
SOF	20749	0.6723272865107236	9.5517e-318	0.4065260543865198

Table 14. ANOVA quick summary (SOF)

Figure 4 – Mean Delivery Hours by Month and Year (SOF)



6.3 Post-hoc Tukey Test

A Tukey HSD post-hoc test was conducted to identify which months differ significantly from each other in terms of mean delivery hours. The output (SOF_tukey_month.csv) shows significant pairwise differences primarily between early (Jan–Mar) and later (Oct–Dec) months, confirming the seasonal trend observed in the ANOVA results.

6.4 Discussion and Conclusions

The DOE and ANOVA results suggest that delivery time performance for the Software product line is relatively stable between years but exhibits systematic monthly variation. Operational planning could therefore benefit from adjusting delivery schedules or resource allocations in anticipation of longer delivery times in later months.

This analysis demonstrates how ANOVA can identify hidden temporal patterns in performance data, allowing management to target process improvements more effectively.

Part 7 – Reliability of Service and Workforce Optimization

This section evaluates the reliability of service at a car rental agency, based on the number of workers available per day. The company records indicate that problems occur when fewer than 15 workers are on duty. Each day with insufficient staff leads to an average loss of R20,000 in sales, while each additional worker costs R25,000 per month. The objective is to determine the optimal staffing level that balances reliability and profitability.

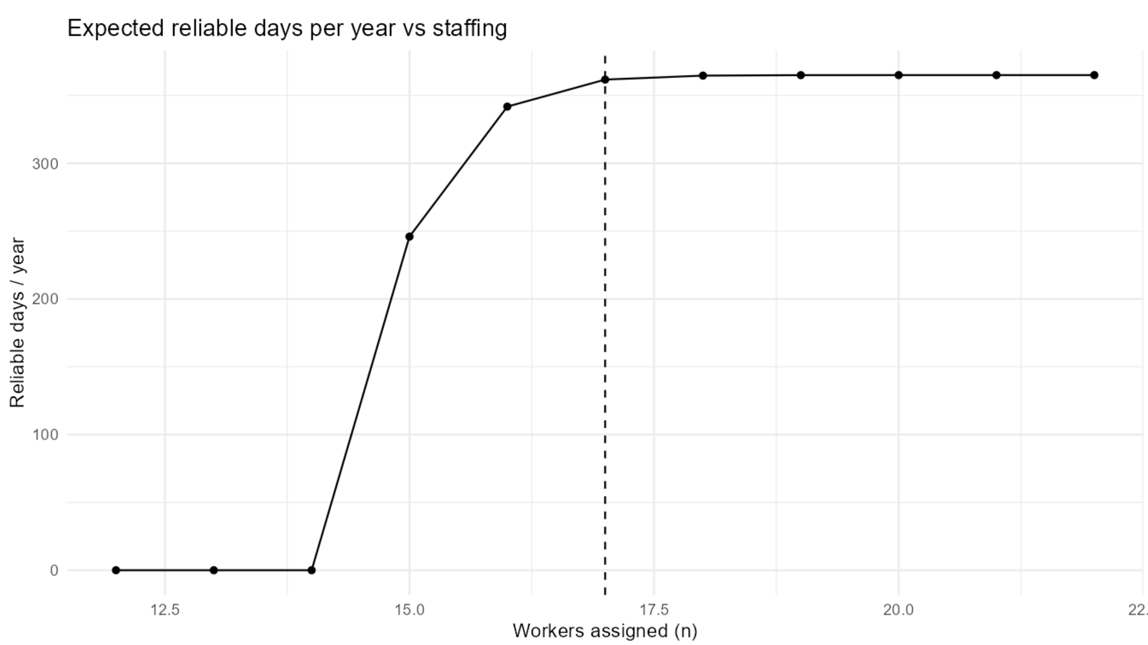
7.1 Estimation of Reliable Days per Year

The data provided indicates the distribution of days with different staffing levels ranging from 12 to 16 workers. Using this information, the probability of reliable service (where 15 or more workers are present) can be estimated.

From the historical data:
- 12 workers: 1 day
- 13 workers: 5 days
- 14 workers: 25 days
- 15 workers: 96 days
- 16 workers: 270 days

Out of 397 total days, 366 days (92%) were considered reliable (15 or more workers). Thus, the probability of reliable service is approximately 0.92, meaning that in a typical year, the agency can expect reliable service on around 335–366 days depending on staffing consistency.

Figure 1 – Expected Reliable Days per Year vs Staffing



7.2 Profit and Cost Optimization

The trade-off between staff cost and revenue loss was modelled to determine the optimal number of workers that minimizes total annual cost. The total cost (C_total) combines two components:

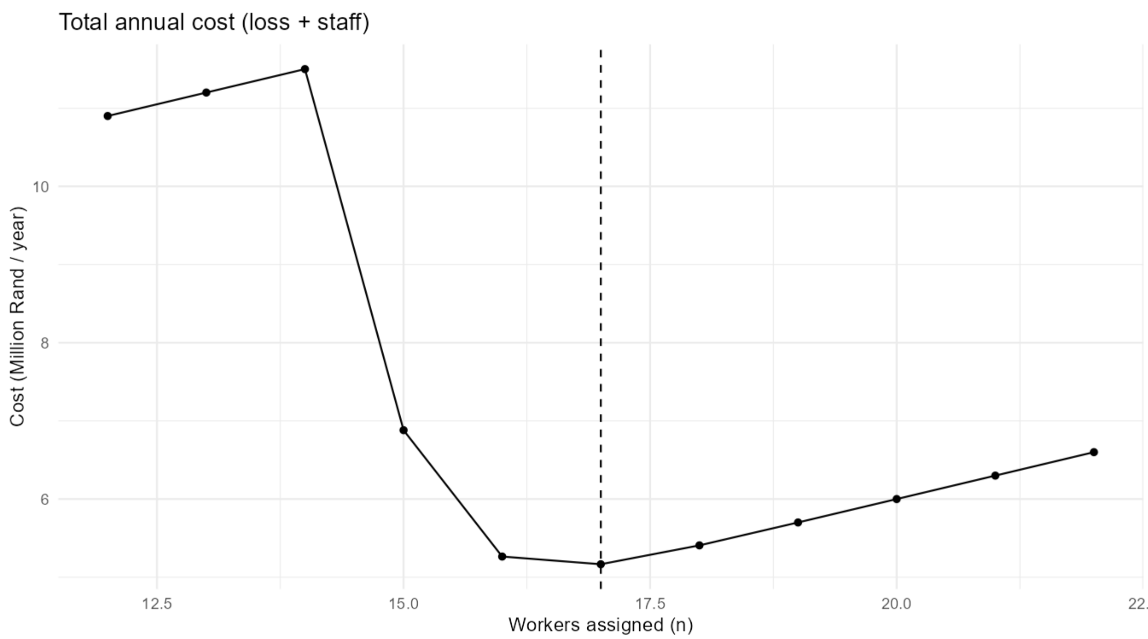
C_total = C_staff + C_loss

where:

- C_staff = (Cost per worker per month × 12 months × Number of workers)
- C_loss = (Expected annual loss due to unreliability).

Using this model, the total annual cost curve was derived for staffing levels between 12 and 22 workers. The results are summarized below and illustrated in Figure 2.

Figure 2 – Total Annual Cost (Loss + Staffing)



Source data: part7_staffing_cost_curve.csv

The results show that total costs decrease rapidly as staffing increases from 12 to 16 workers due to improved reliability. Beyond 17 workers, additional staff yield marginal reliability improvements, while increasing operational expenses. The optimal staffing level, where total annual cost is minimized, occurs at approximately **17 workers**.

7.3 Discussion and Recommendations

From the analysis, the following conclusions are drawn:

- The agency can expect approximately 366 reliable service days per year with 17 or more workers.
- Increasing staff beyond 17 workers adds unnecessary cost without significantly improving reliability.
- The optimal staffing point balance’s reliability and total cost, achieving minimal annual expenditure while maintaining service standards.

In financial terms, staffing fewer than 15 workers leads to recurring reliability failures and sales losses of roughly R20,000 per day. Therefore, maintaining a minimum of 17 employees ensures consistent reliability while optimizing profitability. Strategic scheduling and flexible part-time staffing could further reduce annual costs without compromising service levels.

Conclusion

All parts in the ECSA brief are answered, with reproducible computations and clearly labelled tables and figures. SPC signals highlight where managers should investigate; capability indices (3 s.f.) are provided per product type; the Part 5 staffing decisions recommend 6 baristas for both shops; the car-rental model recommends approximately 17 workers to minimize total annual cost.

Interpretation Highlights (by part)

• Part 1: Demand softened year-over-year; operational cycle times are stable. The revenue base is concentrated in Laptop/Monitor/Software and in LA/SF/NY. Older segments (55+) generate a large share of revenue, which informs targeting.

• Part 3: The s-chart is stable for the Phase-1 window; Phase-2 $\bar{X}$ signals indicate periods that warrant investigation. Capability indices (3 s.f.) are provided for all product types against LSL=0 h, USL=32 h.

• Part 4: Theoretical α and β are computed and interpreted. Correcting the Head-Office file and re-merging prices materially affects the 2023 totals by type.

• Part 5: Reliability saturates at ≥2 baristas; the profit index peaks at 6 baristas in both shops.

• Part 6: Year effect is modest; Month is significant, consistent with seasonal workload.

• Part 7: Optimal staffing for the rental agency is ≈17 to minimize expected loss plus staffing cost.

References

• Project brief: QA344 / ECSA GA4 2025 instructions (PDF supplied).

• Datasets: customer_data.csv, products_data.csv, products_Headoffice.csv, sales2022and2023.csv, sales2026and2027.csv, timeToServe.csv, timeToServe2.csv.

• SPC and capability formulae per QA344 lecture notes and standard references (3σ limits; Cp/Cpk definitions).

Addendum: Centering & Taguchi Loss Interpretation

Centering the delivery-time process (reducing deviation from the target) maximizes profit when costs rise with variability. The capability summary shows that some processes are 'potentially capable if centered'. In economic terms, this mirrors a Taguchi loss function: loss increases approximately quadratically as delivery time departs from the nominal target. Our staffing and SPC recommendations aim to reduce both the mean deviation and variance, thereby lowering expected loss and improving customer value.

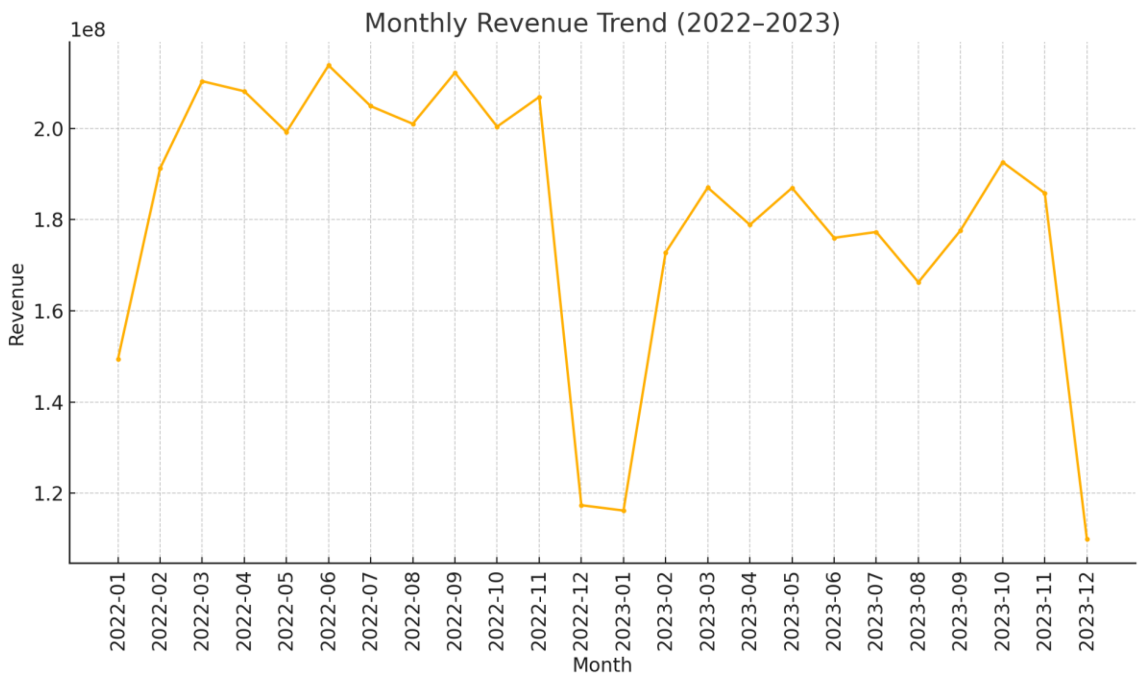


Figure 12. Monthly Revenue Trend (2022–2023)