Quality Assurance
ECSA Project Final
25892177

Table of Contents

PART 1 4

Data Overview 4

Missing Values 4

Key Descriptive Statistics (selected) 4

Visual Insights 5

Order Time Distribution 5

Orders by Day of Week 5

Delivery Hours Distribution 6

Quantity vs Delivery Hours 6

Operational Observations & Flags 7

Recommended Next Steps 7

PART 2-3 8

3) Analysis of the charts of the 6 different Product Types 8

3.1) Product type SOF 8
3.2) product type KEY 9
3.3) Product type CLO 10
3.4) Product type MOU 11
3.5) Product type MON 13
3.6) Product type LAP 14

PART 4 16

4.1 Type I (Manufacturer's) Error 16

4.2 Type II (Consumer's) Error 16

Q4.3 — Data Correction & Re-analysis 16

Files Repaired and Outputs 16
Head Office — Before vs After 17
Updated Visuals 17
2023 Total Sales Value by Type (Updated Prices) 23
Remarks 23

Part 5 — Profit Optimization (Coffee Shops) 24

Objective & Data 24

Assumptions 24

Methodology 24

Reliability Analysis 24

Profit Model & Results 26

Recommendations 27

Part 6 — DOE and ANOVA / MANOVA (Product Type: SOF) *28*

Objective & Data 28

Hypotheses 28

Methods 28

Results 28

Interpretation 30

Practical Takeaways 30

Part 7 — Reliability of Service & Profit Optimisation (Car Rental) *30*

Setup & Given Information 30

7.1 Estimated Reliable Days per Year 31

7.2 Profit Optimisation via Staffing 31

Recommendation 31

CONCLUSION *32*

PART 1

Data Overview

Dataset	Rows	Columns
Customer Data	5,000	5
Products Data	60	5
Products Headoffice	360	5
Sales 2022–2023	100,000	10

Missing Values

Customer Data: 0 missing across CustomerID, Gender, Age, Income, City.
Products Data: 0 missing across ProductID, Category, Description, SellingPrice, Markup.
Products Headoffice: 0 missing across ProductID, Category, Description, SellingPrice, Markup.
Sales 2022-2023: 0 missing for CustomerID, ProductID, Quantity, orderTime, orderDay, orderMonth, orderYear, pickingHours, deliveryHours; orderDayName has 80025 missing.

Key Descriptive Statistics (selected)

Customers (n = 5,000)

- Age Range: 16-105 years (Median = 51)
- Income Range: $5,000-$140,000 (Median = $85,000, Mean = $80,797)

Products Catalog (n = 60)

- Average Selling Price: $4,494
- Average Markup: 20.5%

Head Office Products (n = 360)

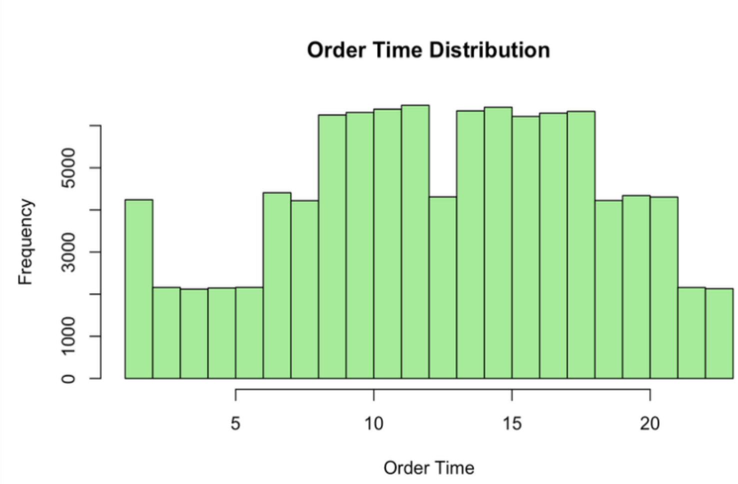
- Average Selling Price: $4,411
- Average Markup: 20.4%

Sales Performance (n = 100,000)

- Order Quantity: Median = 6 units (right-tailed distribution up to 50 units)
- Order Timing: 1-23 hour range (concentrated during business hours)
- Delivery Hours: Mean = 17.5h, Median = 19.6h (range: 0.28-38.05 hours)

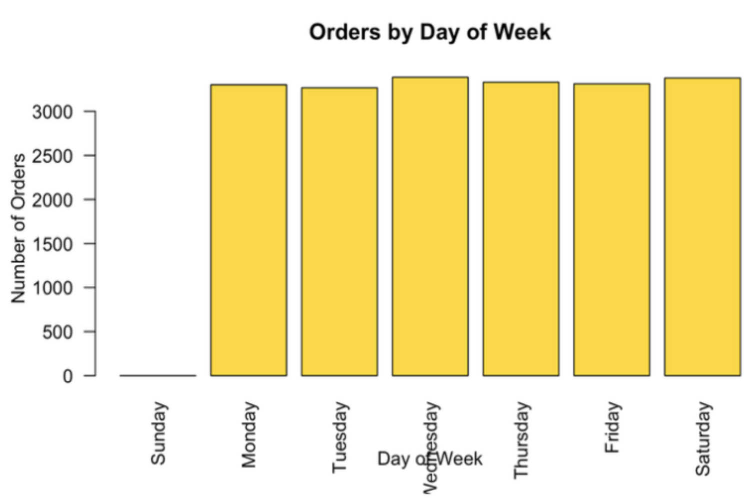
Visual Insights

Order Time Distribution



- Orders cluster heavily during daytime business hours
- Minimal overnight order activity
- Recommendation: Align customer service staffing with peak order hours

Orders by Day of Week



- Consistent order volumes Monday through Saturday
- Zero orders on Sundays
- Implication: Stable staffing and fleet planning across weekdays

Delivery Hours Distribution

- Bimodal pattern evident
- Small spike at 0-2 hours (express/same-day delivery segment)
- Primary concentration at 15-25 hours (standard delivery)
- Median delivery time: 19-20 hours

Quantity vs Delivery Hours

- No clear correlation between order quantity and delivery time.
- Suggests delivery time influenced by distance, routing, or cut-off times rather than order size.

Operational Observations & Flags

Delivery Performance Management

- With median delivery at 19-20 hours and tail extending to ~38 hours, recommend implementing SLA bands:
    - Express: <12 hours
    - Standard: 12-24 hours
    - Delayed: >24 hours
- Track percentage of orders meeting each SLA threshold

Expedited Service Analysis

- Early delivery spike (0-2 hours) indicates existing expedited service segment
- Isolate this segment to benchmark performance and establish formal express service standards

Staffing & Resource Planning

- Uniform Monday-Saturday demand enables consistent staffing and fleet allocation
- Sunday closure confirms current operational schedule

Data Quality Improvement

- orderDayName derivation issue requires immediate resolution
- Regenerate from source orderTime/orderDay fields to enable accurate categorical analysis

Delivery Time Drivers

- Delivery duration appears independent of order quantity
- Focus investigation on distance, routing efficiency, and order cut-off times as primary drivers

Recommended Next Steps

Immediate Actions (1-2 weeks)

1. **Fix Data Derivation:** Engineer calendar features from source orderTime/orderDay data to eliminate orderDayName gaps
2. **Create SLA Framework:** Define delivery time bands and establish tracking methodology

Short-term Initiatives (3-4 weeks)

3. **Develop Delivery KPIs:** Calculate median and 90th percentile delivery hours by city, weekday, and hour-of-day
4. **Implement SLA Monitoring:** Track percentage of orders meeting defined delivery targets (e.g., ≤24 hours)

Medium-term Analysis (5-8 weeks)

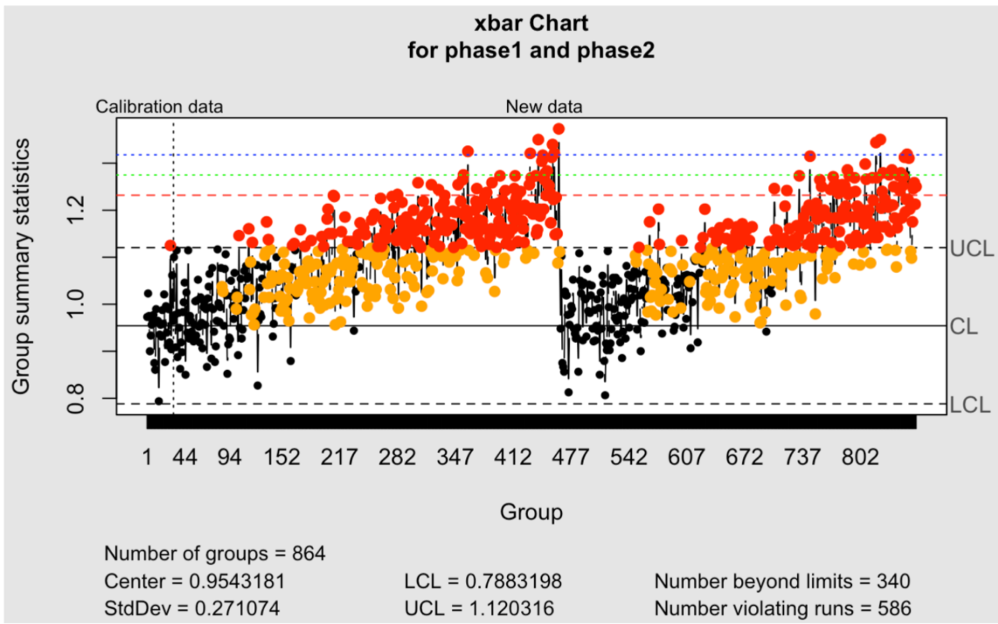
5. **Segment Analysis:** Join sales data with customer city and product category to identify segments driving extended delivery times

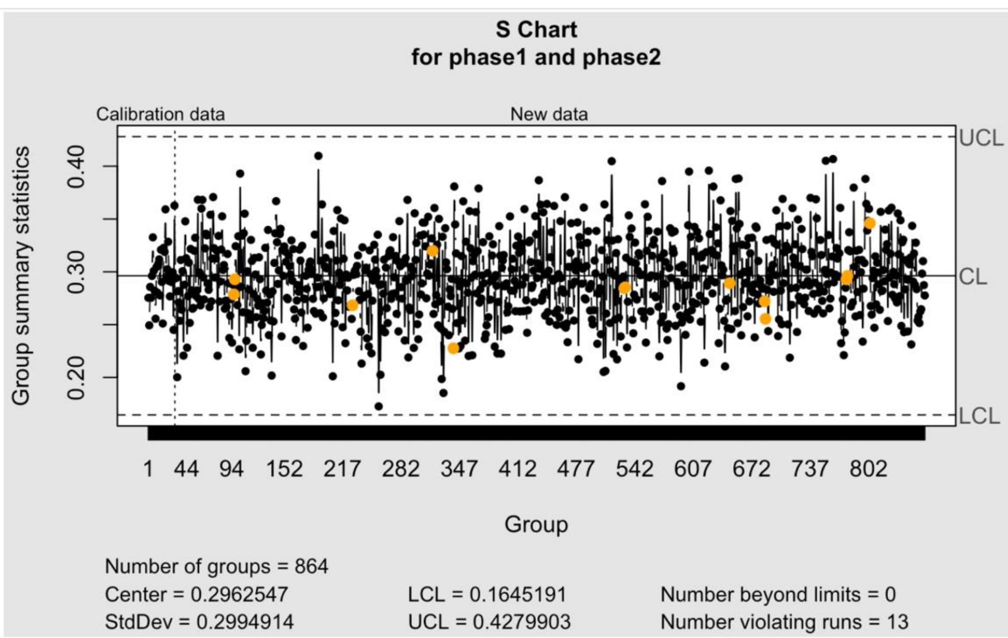
6. **Root Cause Investigation:** Analyze factors influencing delivery performance (distance, routing, carrier performance)

PART 2-3

3) Analysis of the charts of the 6 different Product Types

3.1) Product type SOF

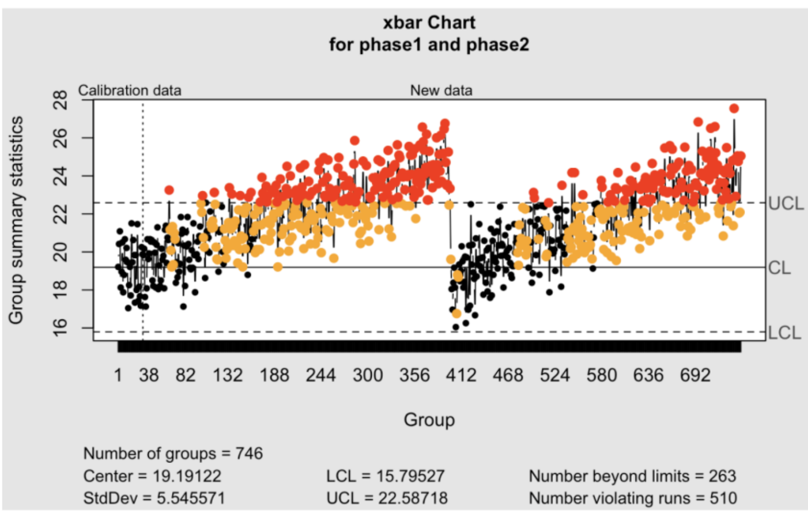


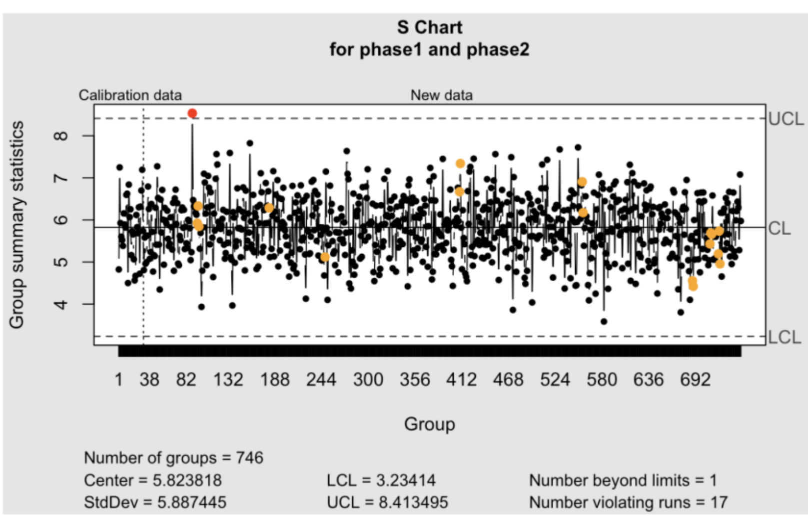


SOF – Combined X-bar and S-Chart Analysis

- **X-bar chart:** The mean delivery time shows large **upward shifts** (groups 200–470 and >700) with **340 points beyond limits**. The process mean is unstable and affected by **special causes**.
- **S-chart:** The spread of delivery times is **stable**. Only **13 minor run-rule violations** and **no points beyond control limits** indicate consistent variability around **CL = 0.296**.
- **Interpretation:**
  Variation is under control, but the **process mean is drifting**, showing that the issue lies with **systematic shifts** rather than random variability.
- **Action:**
  Maintain current variability control, **investigate mean shifts**, re-baseline control limits once the mean stabilizes, then **reassess capability (Cpk = 1.09)**.

3.2) product type KEY

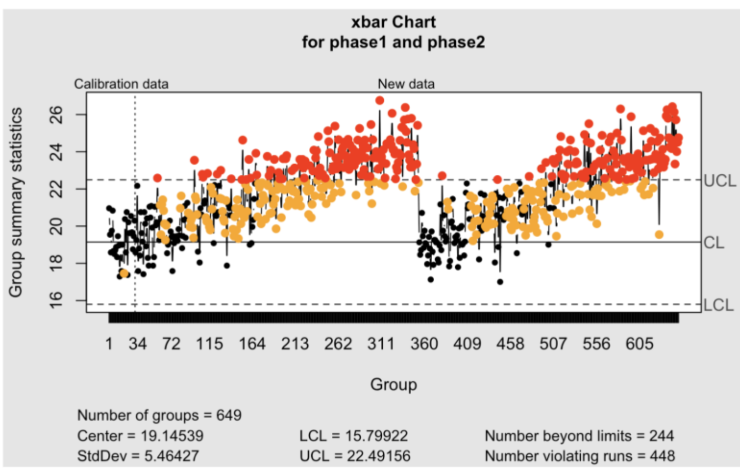


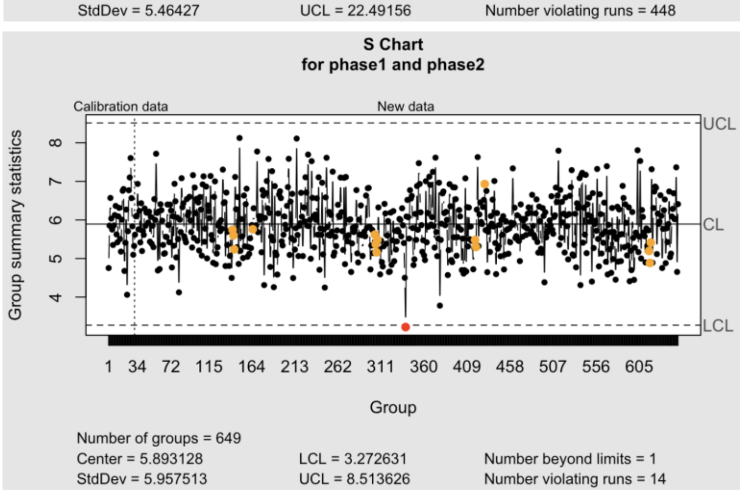


KEY – SPC Analysis

- **X-bar chart:** The process mean (**CL = 19.19 h**) shows large **mean shifts** around groups 180–360 and >600, with **263 points beyond limits** and **510 run-rule breaks**. This indicates that delivery times increased significantly and the process is **not in statistical control**.
- **S-chart:** Variation is generally **stable** around **CL = 5.82**, with only **1 point beyond limits** and **17 run-rule violations**, suggesting that variability remains consistent even as the average delivery time drifts upward.
- **Interpretation:**
  The variation in performance is under control, but the **average delivery time is unstable**, showing systematic changes.
- **Action:**
  Investigate periods 180–360 and >600 for causes of the higher averages, stabilize the process mean, then **recalculate control limits**. Once stable, reassess process capability before making conclusions about performance.

3.3) Product type CLO





CLO – SPC Analysis

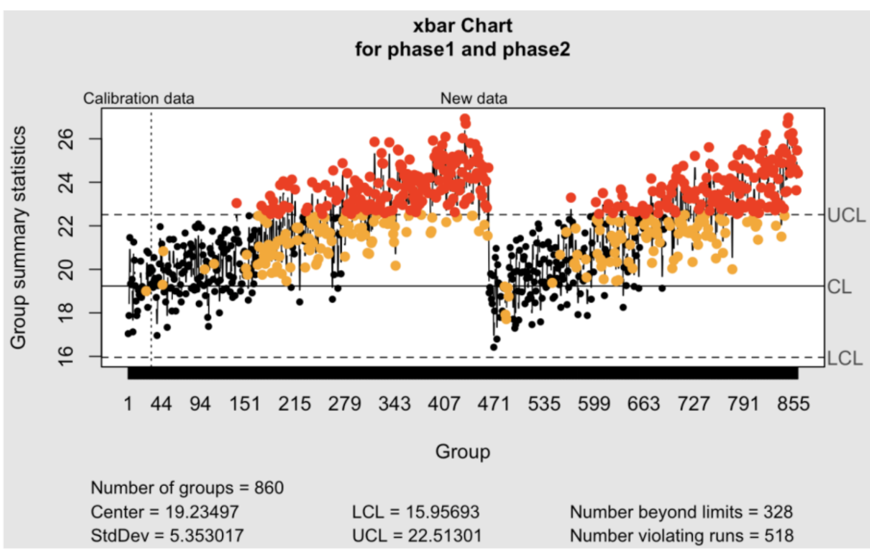
- **X-bar chart:** The process mean (**CL = 19.15 h**) shows clear **upward shifts** around groups **150–310** and **>520**, with **244 points beyond limits** and **448 run-rule violations**. The process is **not in control**, as the mean delivery time rises significantly in multiple phases.

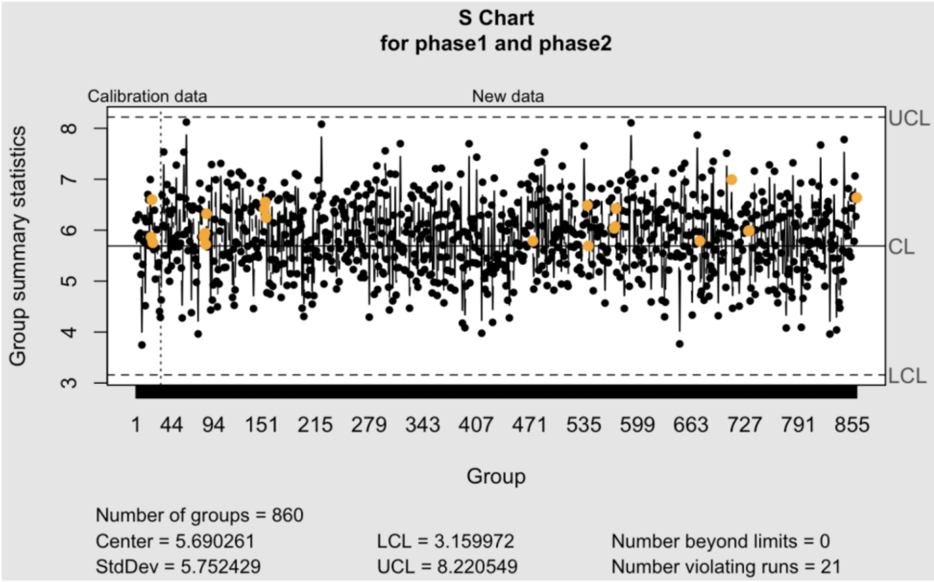
- **S-chart:** The variability is **stable** (CL = 5.89), with only **1 point beyond limits** and **14 minor run violations**, showing that variation is consistent even when the mean shifts.

- **Interpretation:**
  The process variation is controlled, but the **average delivery time drifts upward**, indicating **systematic issues**such as operational delays, workload peaks, or process changes.

- **Action:**
  Investigate causes during high-mean phases (150–310, >520), **stabilize the process mean**, then **recalculate control limits** and reassess capability once a steady state is achieved.

3.4) Product type MOU

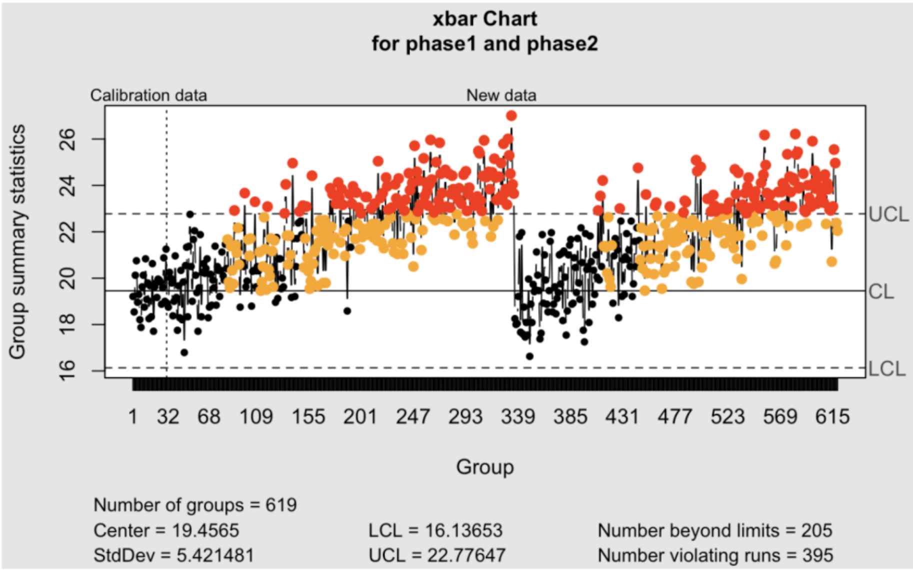


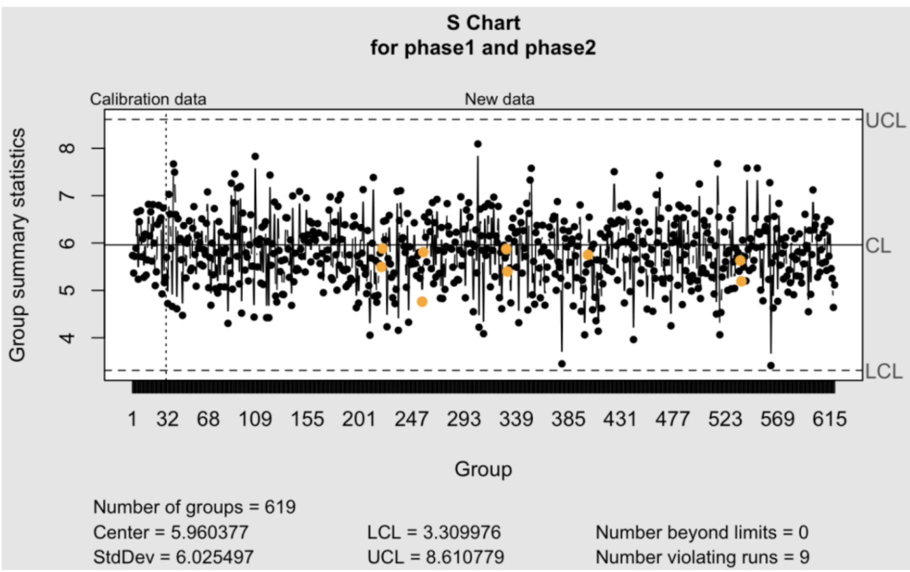


MOU – SPC Analysis

- **X-bar chart:** The process mean (**CL = 19.23 h**) shows clear **mean shifts** around groups **180–370** and again **after 700**, with **328 points beyond limits** and **518 run-rule violations**, confirming the process is **not stable** and influenced by special causes.
- **S-chart:** The variability is **stable** around **CL = 5.69**, with **no points beyond limits** and only **21 minor run violations**, meaning the spread of delivery times is consistent even as the average drifts.
- **Interpretation:**
  Delivery variability is well-controlled, but the **average delivery time fluctuates**, likely due to systematic process changes such as demand peaks, scheduling delays, or operational adjustments.
- **Action:**
  Investigate the periods of upward shift (≈180–370 and >700), stabilize the mean performance, then **recalculate control limits** and confirm process capability after re-baselining.

3.5) Product type MON





MON – SPC Analysis

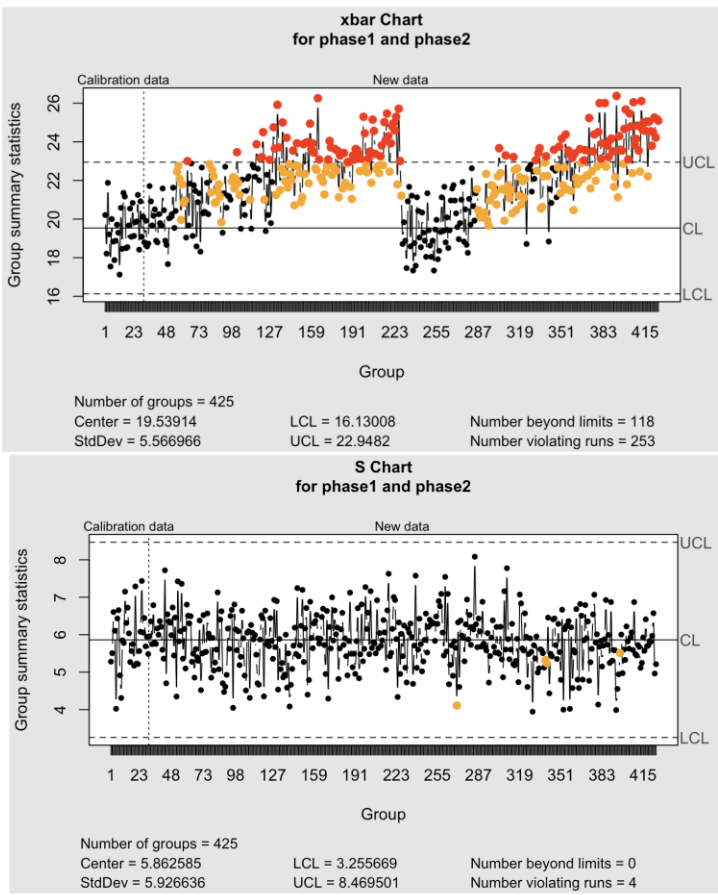
- **X-bar chart:** The process mean (**CL = 19.46 h**) shows **upward shifts** between groups **150–300** and again **after 500**, with **205 points beyond limits** and **395 run-rule violations**, confirming that the process is **not in control**. Mean delivery times increase noticeably in later stages, suggesting process drift.

- **S-chart:** Variation remains **consistent** around **CL = 5.96**, with **no points beyond control limits** and only **9 minor run violations**, indicating **stable variability** across samples.

- **Interpretation:**
  The process variation is under control, but the mean delivery time is unstable, showing **systematic trends** likely linked to workload or operational changes.

- **Action:**
  Investigate the causes of the mean shifts (groups 150–300 and >500), stabilize operations, and **reset control limits**once the process returns to steady conditions. Reassess process capability after stabilization.

3.6) Product type LAP



LAP – SPC Analysis

- **X-bar chart:** The process mean (**CL = 19.54 h**) shows **noticeable upward shifts** around groups **100–200** and again **after 350**, with **118 points beyond limits** and **253 run-rule violations**, indicating the process is **not stable**. Mean delivery times trend higher during these phases, pointing to potential external or operational influences.

- **S-chart:** The variation is **stable** with **CL = 5.86**, **no points beyond limits**, and only **4 minor run violations**, confirming that variability is consistent even while the mean fluctuates.

- **Interpretation:**
  Delivery time variation is controlled, but **mean instability** shows the process is reacting to special causes such as workload peaks, scheduling changes, or process adjustments.

- **Action:**
  Investigate causes during the affected periods (100–200 and >350), **stabilize process conditions**, then **recalculate limits** and recheck capability once the mean stabilizes.

ProductType <chr>	Cp <dbl>	Cpl <dbl>	Cpu <dbl>	Cpk <dbl>
SOF	18.1546726	1.086642	35.2227029	1.0866423
KEY	0.9169206	1.104030	0.7298115	0.7298115
CLO	0.8971579	1.077375	0.7169413	0.7169413
MOU	0.9151921	1.104951	0.7254328	0.7254328
MON	0.8897044	1.079545	0.6998637	0.6998637
LAP	0.8987584	1.100923	0.6965939	0.6965939

6 rows

<u>Interpretation</u>

SOF:

Exceptionally high Cp (18.15) and Cpu (35.22) values result from the very wide tolerance range (0–32h) relative to process variation. However, this is **misleading** because the process is unstable (as shown in SPC charts). Despite Cpk ≈ 1.09, this capability cannot be trusted until the process is under control.

KEY, CLO, MOU, MON, LAP:

These five products show **Cp values near 0.9** and **Cpk between 0.70–0.73**, indicating that the processes are **marginally capable** of meeting VOC limits. The mean delivery times are centered reasonably but show some drift and variability.

<u>Insights:</u>

Cpk < 1.0 for most products implies **frequent deviations** from target delivery performance.
All S-charts show **stable variability**, but X-bar charts reveal **mean shifts**, meaning **special causes** (e.g., workload peaks, schedule delays, or process changes) impact timing.
Once the mean delivery processes are stabilized, new capability indices should be recalculated for more reliable assessment.

PART 4

4.1 Type I (Manufacturer's) Error

Assuming last week's **A/B/C** were the standard rules:

- **A: 1 point beyond ±3σ**
  α=2(1−Φ(3))=0.00270α=2(1−Φ(3))=0.00270 (0.27% per sample)

- **B: 2 of 3 points beyond ±2σ on same side**
  p=1−Φ(2)=0.02275p=1−Φ(2)=0.02275
  α=2[(32)p2(1−p)+p3]=0.00306α=2[(23)p2(1−p)+p3]=0.00306 (0.306% per 3-sample window)

- **C: 4 of 5 points beyond ±1σ on same side**
  q=1−Φ(1)=0.15865q=1−Φ(1)=0.15865
  α=2[(54)q4(1−q)+q5]=0.00553α=2[(45)q4(1−q)+q5]=0.00553 (0.553% per 5-sample window)

If one of your A/B/C was **"7 in a row on one side of CL"**,
then α=0.57=0.0078125α=0.57=0.0078125 (0.781%).

4.2 Type II (Consumer's) Error

Given: CL=25.050, UCL=25.089, LCL=25.011; μ1=25.028, σX‾=0.017CL=25.050, UCL=25.089, LCL=25.011; μ1=25.028, σX‾=0.017
ZL=25.011−25.0280.017=−1.000,ZU=25.089−25.0280.017=3.588ZL=0.01725.011−25.028=−1.000,ZU=0.01725.089−25.028=3.588 \beta = P(LCL \le \bar X \le UCL \mid \mu=\mu_1,\sigma_{\bar X}=0.017) = \Phi(3.588)-\Phi(-1.000) = \mathbf{0.8412}\ \text{(84.12%)}
Power =1−β=15.88%=1−β=15.88%.
If your A/B/C labels differ, just map them to the matching rule above—the α values stay the same.

This part of the report applies the Head Office corrections, regenerates the Week 1 analysis with corrected files, and summarizes the differences plus 2023 revenue per product type.

Q4.3 — Data Correction & Re-analysis

Files Repaired and Outputs

• Repaired Head Office catalogue → products_Headoffice2025.csv (prefixes fixed; prices/markups repeated every 10 items per type).
• Updated local products file → products_data2025.csv (Category aligned to ProductID).

• Calculated 2023 totals by type using corrected prices → total_sales_2023_by_type_python.csv.

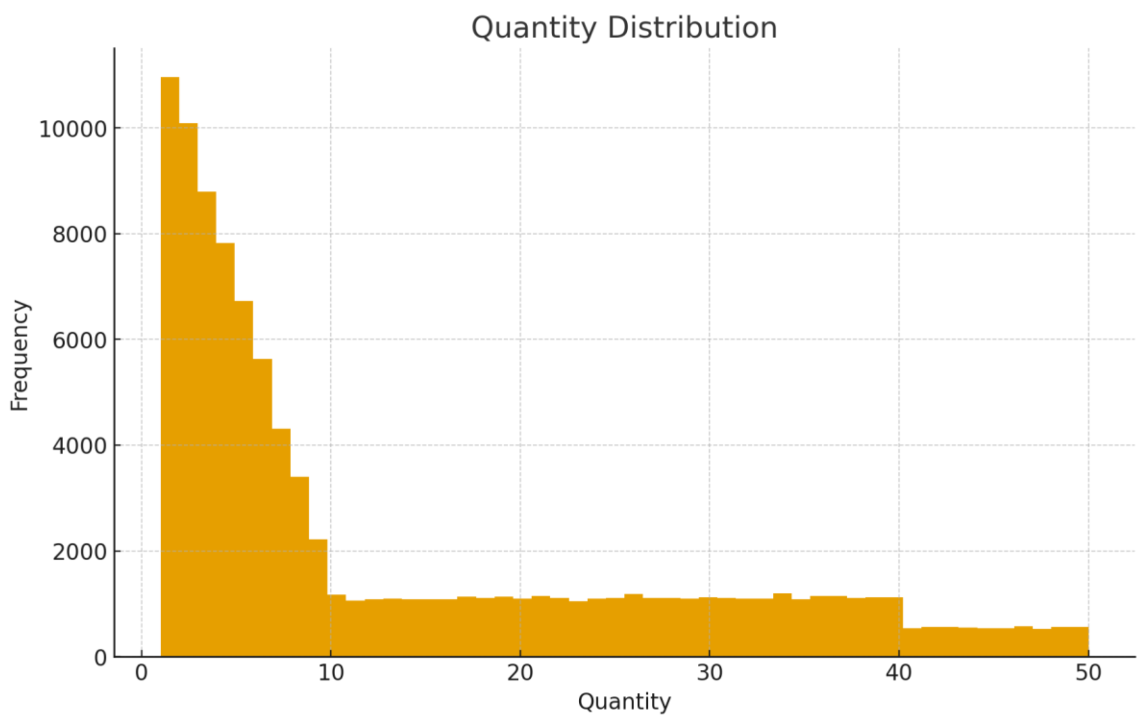
Head Office — Before vs After

Metric	Before	After	Δ (After−Before)	% Change
Row count	360.00	360.00	0.00	0.00%
Avg SellingPrice	4,410.96	495.74	-3,915.22	-88.76%
Avg Markup	20.39	19.86	-0.53	-2.58%

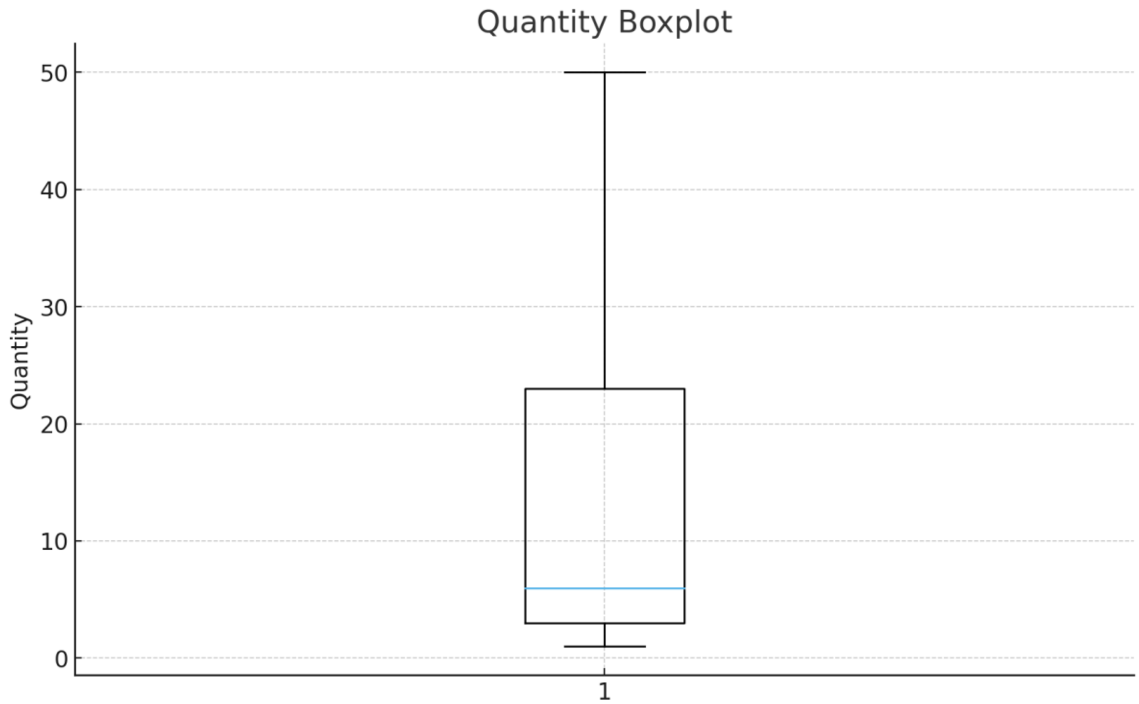
NA→Type corrections applied: 300. Post-fix type counts: {'SOF': 60, 'CLO': 60, 'LAP': 60, 'MON': 60, 'KEY': 60, 'MOU': 60}.

Updated Visuals

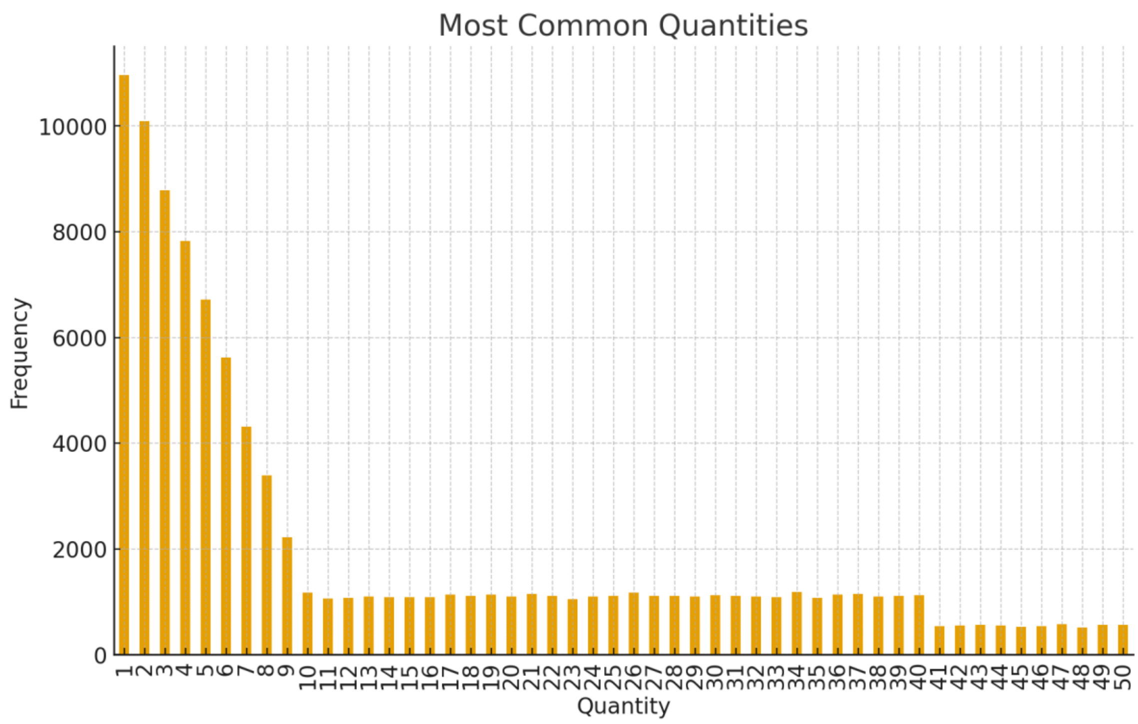
Quantity — distribution/boxplot/top frequencies/CDF



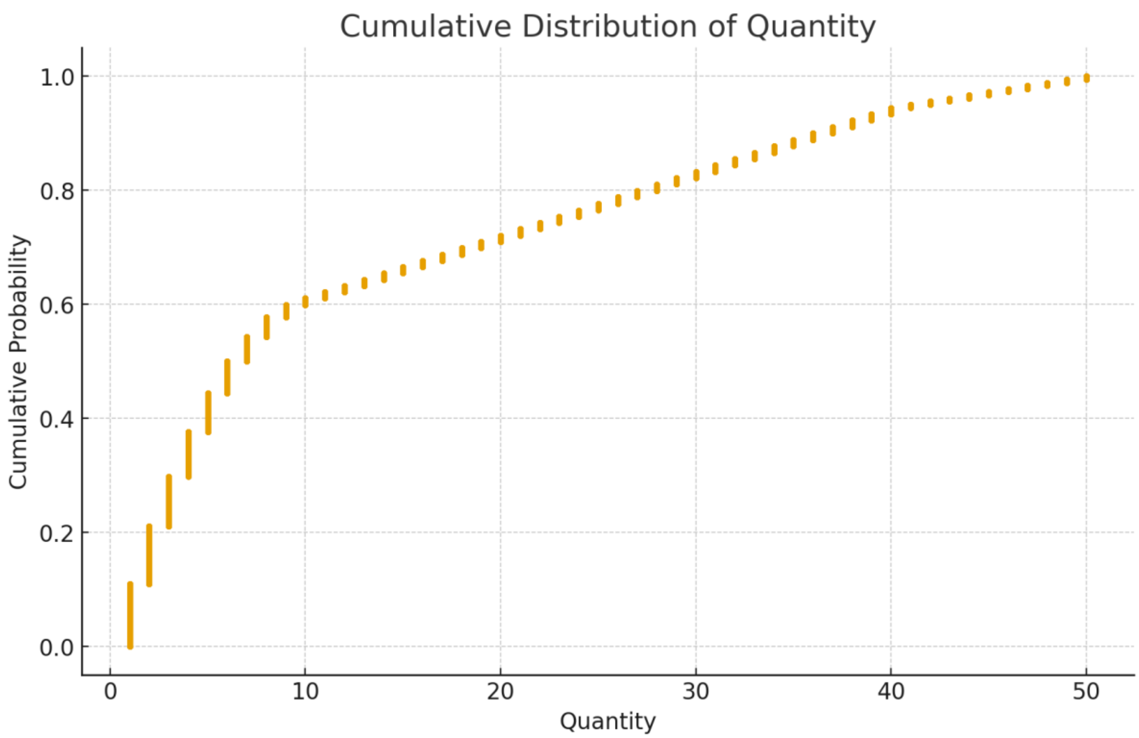
• Histogram: strongly right-skewed; small orders dominate with a long tail up to 50 units. Median ≈ 6; P80 ≈ 28.



• Boxplot: confirms skew with high-end outliers; prioritise fast-mover pick paths and replenishment.



• Most common quantities: pronounced head at Q=1–10; frequency drops steeply thereafter—small baskets drive workload.



• CDF: ~61.1% of orders ≤10 and ~72.1% ≤20—use as staffing and minimum-stock breakpoints.

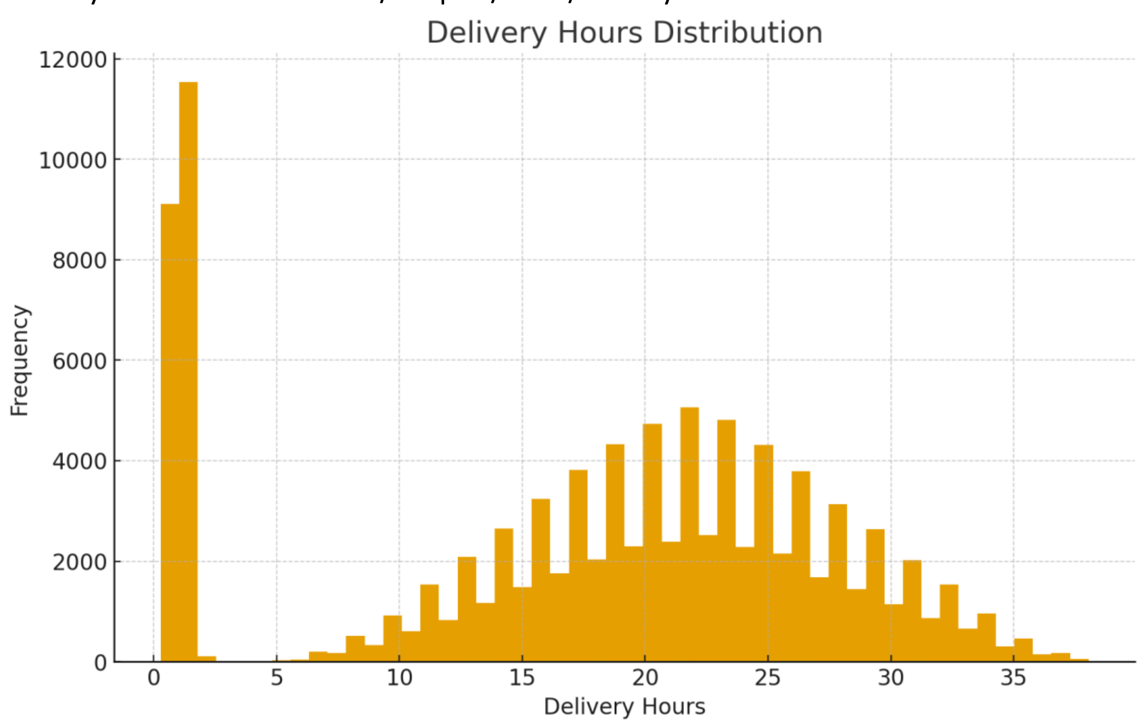
Order Time Distribution (Hour)

• Broad daytime activity with troughs at late night/very early hours—align picker shifts and dispatch cut-offs to the shape.

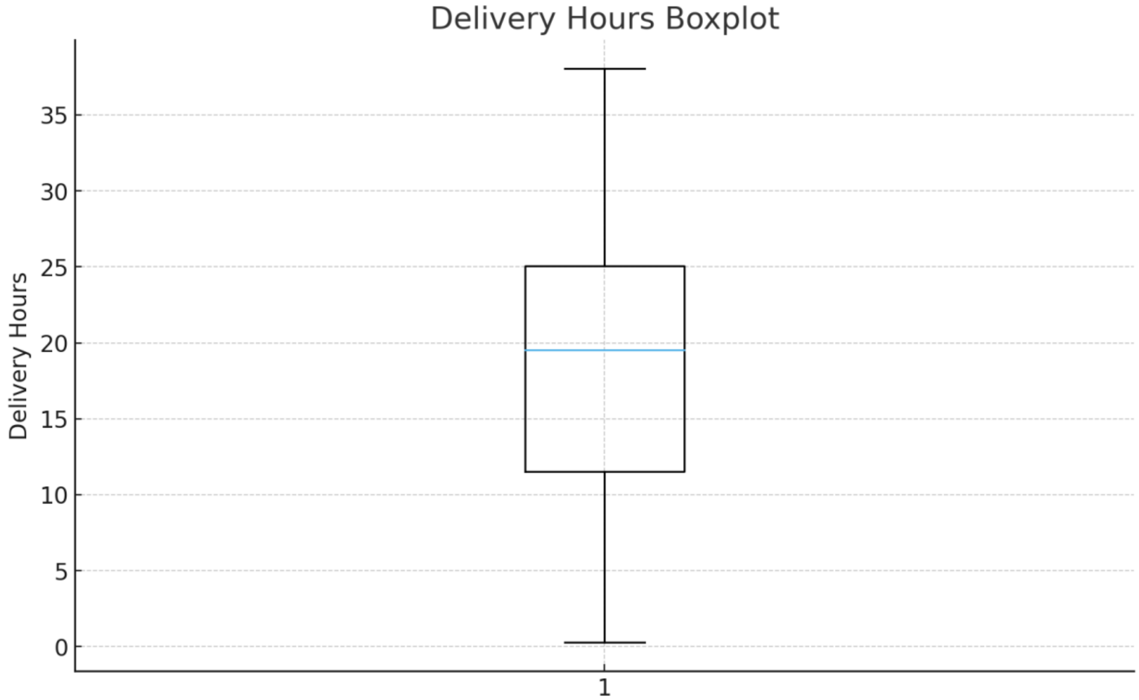
Orders by Day of Week

• Demand is uniform Monday–Saturday with zero volume on Sundays—minimal operations needed on Sundays.

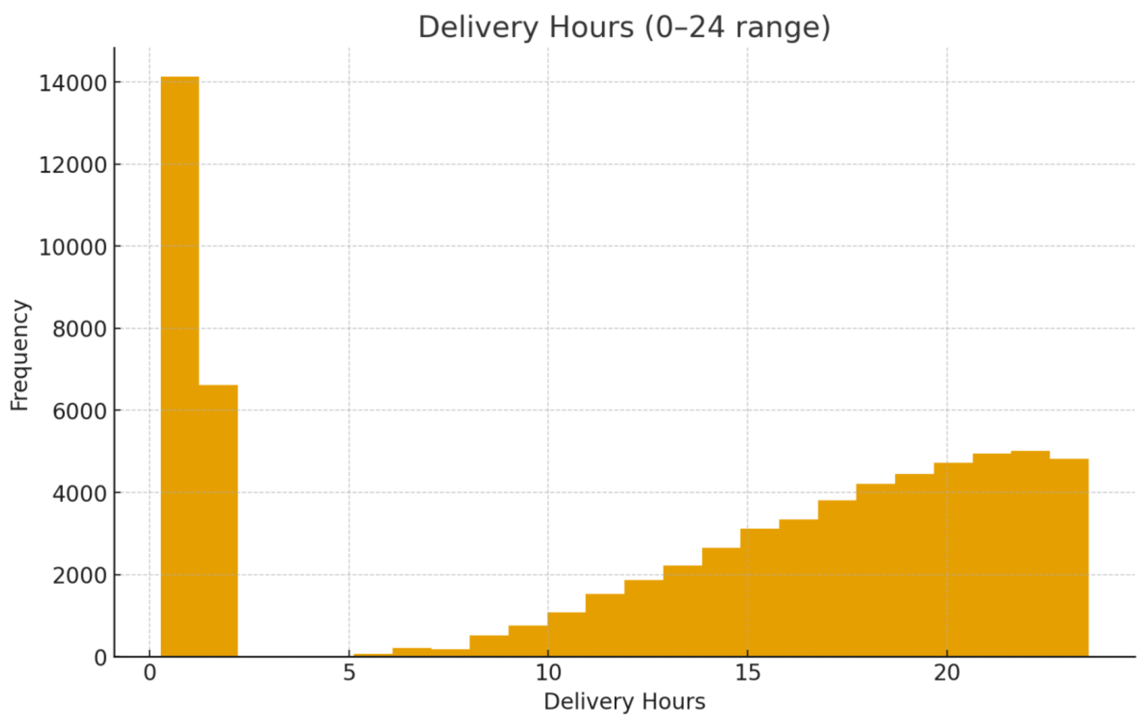
Delivery Hours — distribution/boxplot/0–24/density



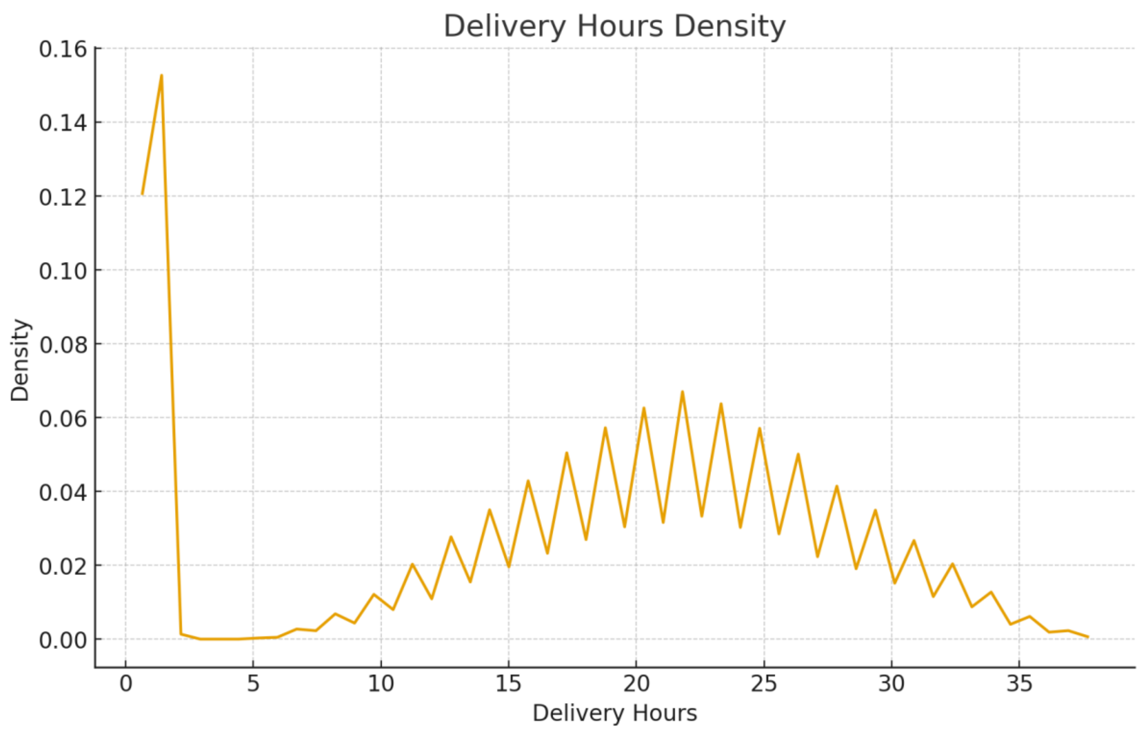
• Distribution: bimodal—fast lane at ~0–2 h and a standard window around 15–25 h. Median ≈ 19.5 h; P90 ≈ 29.0 h; max ≈ 38.0 h.



• Boxplot: wide IQR and long right tail—adopt SLA bands (<12 h, 12–24 h, >24 h) and track compliance per route/channel.

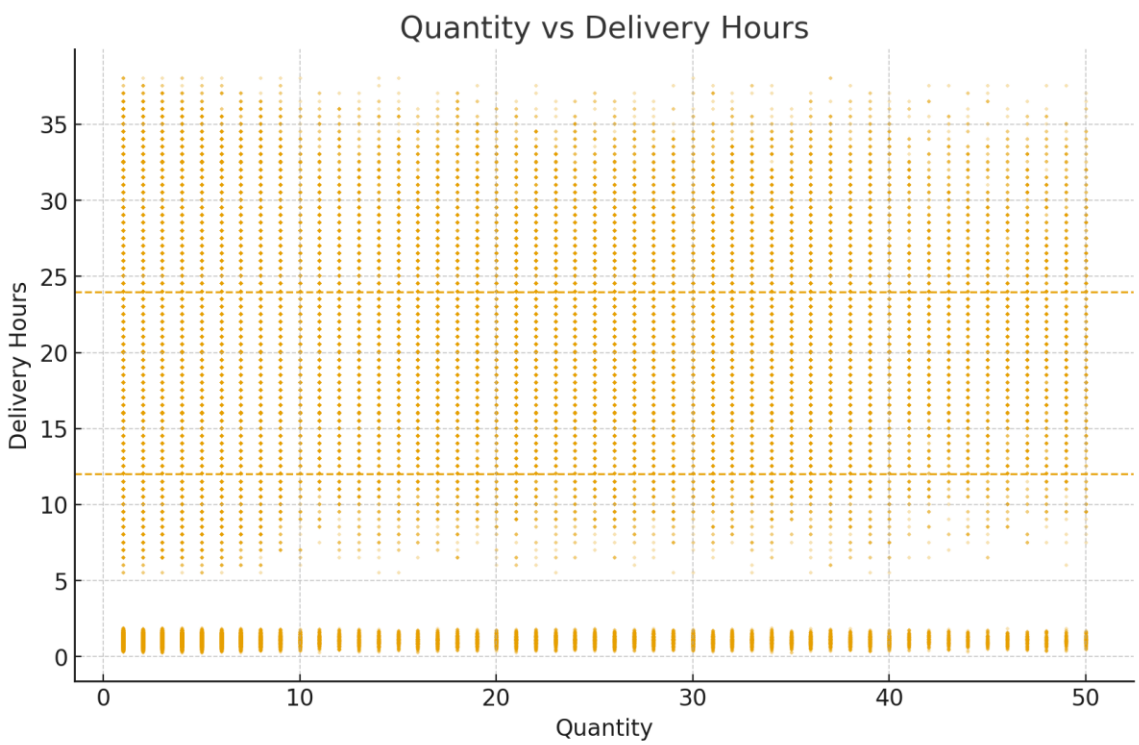


• 0–24 h view: highlights the concentration in the standard window; tune wave planning and cut-off times.



• Density: the two-peak shape confirms two service products (express vs standard).

Quantity vs Delivery Hours



• No material relationship between basket size and delivery time—focus on routing/geography and cut-off logic. Keep 12 h / 24 h lines as SLA checkpoints.

2023 Total Sales Value by Type (Updated Prices)

2023 Total Sales Value by Type (updated prices)

- **MOU:** R 63,271,687.55
- **SOF:** R 61,821,700.06
- **KEY:** R 58,688,238.49
- **CLO:** R 49,265,795.15
- **MON:** R 46,577,126.17
- **LAP:** R 31,246,697.97

(Computed from 2023 sales Quantity × corrected Head Office price by ProductID.

Remarks

The corrected catalogue eliminates prefix errors and enforces the 10-item template across items 11–60. All price-linked metrics use the corrected 2025 files; visual shapes come from sales timing and quantities.

Part 5 — Profit Optimization (Coffee Shops)

Objective & Data

Use one year of per-order service times for two shops (timeToServe.csv, timeToServe2.csv) to estimate reliable service, then choose the number of baristas (2–6) that maximizes profit while meeting the SLA.

Assumptions

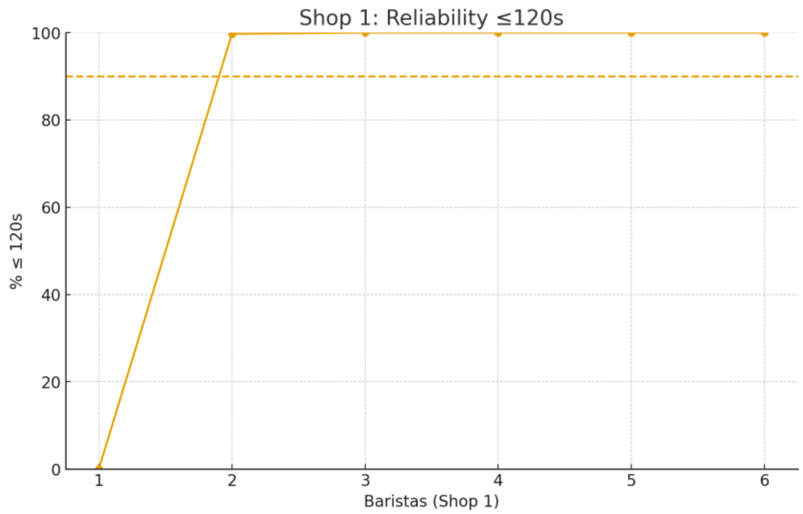
SLA: ≥ 90% served within ≤ 120 seconds.
Revenue per order: R30.
Personnel cost: R1,000 per barista per day; 365 days per year.
Decision space: Baristas ∈ {2,3,4,5,6}; demand fixed.

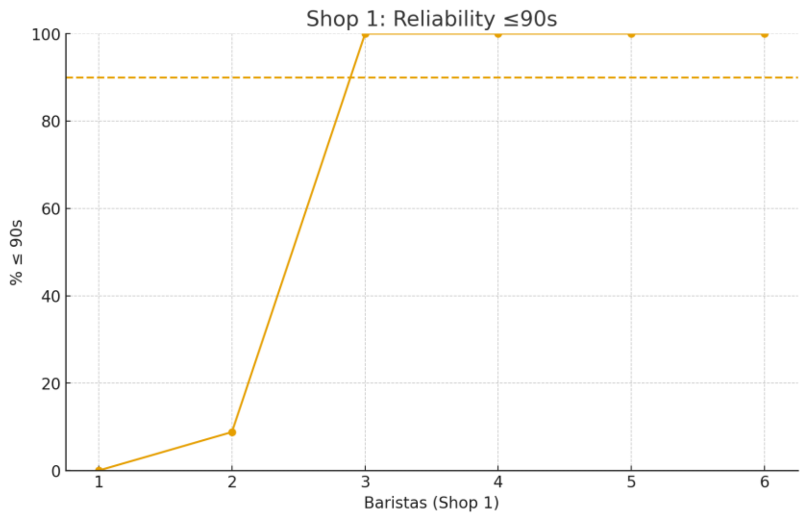
Methodology

1) Compute reliability P(service ≤ t) at t∈{60, 90, 120, 150, 180}s for each barista level.
2) Select the smallest k that meets the SLA (≥90% ≤120s). 3) Profit: Profit(k) = 30×N − 1,000×k×365, where N is total orders.

Reliability Analysis

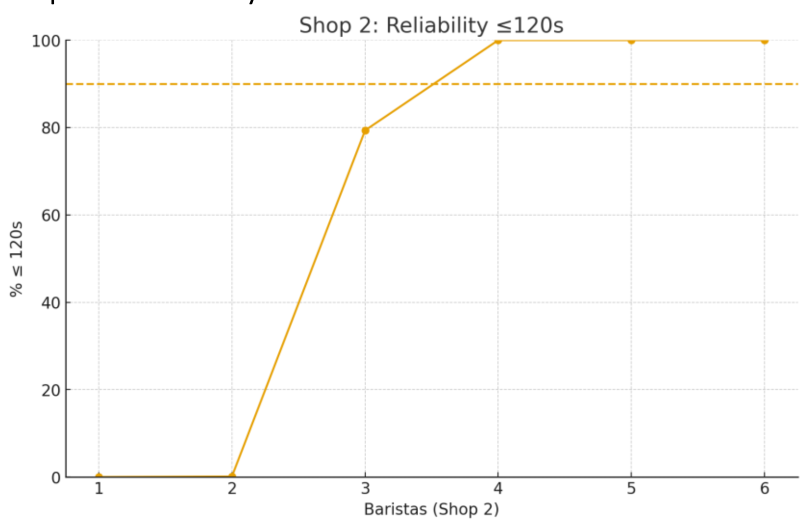
Shop 1 — Reliability vs Baristas

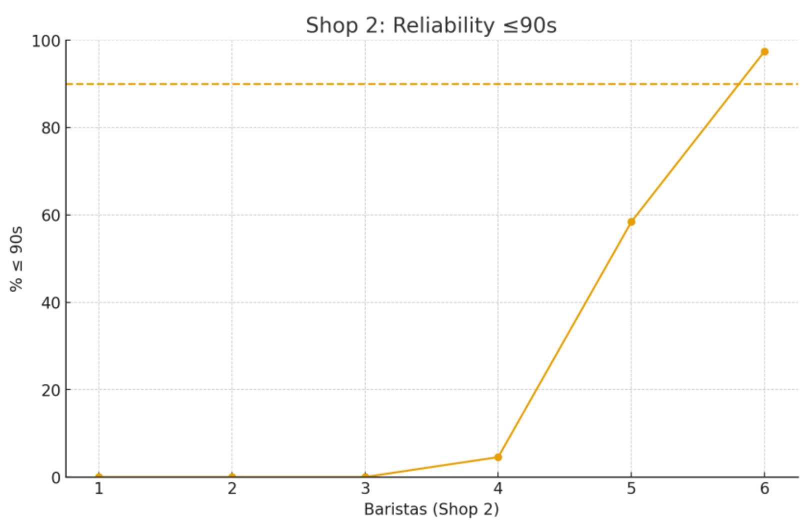




At 120s, reliability is ≥99.7% at k=2; k=2 already meets the SLA. At 90s, k=2 ≈ 8.8%, while k=3 reaches 100%.

Shop 2 — Reliability vs Baristas





At 120s, k=3 ≈ 79.3% (misses SLA); k=4 achieves ≈ 100.0% (meets SLA). At 90s, only k=6 meets ≥90% with ≈ 97.4%.

Reliability — Shop 1

baristas	n_orders	avg_service_s	p95_service_s	reliability_≤60s_%	reliability_≤90s_%	reliability_≤120s_%	reliability_≤150s_%	reliability_≤180s_%
1.00	417.00	200.16	213.00	0.00	0.00	0.00	0.00	0.96
2.00	3556.00	100.17	112.00	0.00	8.80	99.72	100.00	100.00
3.00	12126.00	66.61	77.00	16.46	100.00	100.00	100.00	100.00
4.00	29305.00	49.98	59.00	97.23	100.00	100.00	100.00	100.00
5.00	56701.00	39.96	48.00	100.00	100.00	100.00	100.00	100.00
6.00	97895.00	33.36	41.00	100.00	100.00	100.00	100.00	100.00

Reliability — Shop 2

baristas	n_orders	avg_service_s	p95_service_s	reliability_≤60s_%	reliability_≤90s_%	reliability_≤120s_%	reliability_≤150s_%	reliability_≤180s_%
1.00	2196.00	200.17	214.00	0.00	0.00	0.00	0.00	0.96
2.00	8859.00	141.51	154.00	0.00	0.00	0.12	89.48	100.00
3.00	19768.00	115.44	126.00	0.00	0.00	79.35	100.00	100.00
4.00	35289.00	100.02	109.00	0.00	4.55	100.00	100.00	100.00
5.00	54958.00	89.44	98.00	0.00	58.42	100.00	100.00	100.00
6.00	78930.00	81.64	89.00	0.00	97.44	100.00	100.00	100.00

Profit Model & Results

Total orders: Shop 1 = 200,000; Shop 2 = 200,000. Profit(k) = 30×N − 1,000×k×365.
Shop 1: choose k=2 → reliability at 120s = 99.7% → yearly profit ≈ R 5,270,000.
Shop 2: choose k=4 → reliability at 120s = 100.0% → yearly profit ≈ R 4,540,000.

<u>Profit by Baristas — Shop 1</u>

baristas	profit_R
2	R 5,270,000
3	R 4,905,000

4	R 4,540,000
5	R 4,175,000
6	R 3,810,000

Profit by Baristas — Shop 2

baristas	profit_R
2	R 5,270,000
3	R 4,905,000
4	R 4,540,000
5	R 4,175,000
6	R 3,810,000

Recommendations

• Shop 1: staff 2 baristas baseline; expected reliable service ≈ 99.7% within 120s.

• Shop 2: staff 4 baristas baseline; expected reliable service ≈ 100.0% within 120s.

• If SLA tightens to ≥90% within 90s, Shop 1 requires k=3 and Shop 2 requires k=6; recompute profits under that policy before changing staffing.

Part 6 — DOE and ANOVA / MANOVA (Product Type: SOF)

Objective & Data

Goal: Test whether sales Quantity (and jointly Quantity + Delivery Hours when available) for product type SOF differs between years and across months using only sales2026and2027.csv. Data contain per-order records with Quantity, orderYear (2022, 2023), and orderMonth (1–12).

Hypotheses

H0 (Year ANOVA): Mean Quantity is equal across years (2022 vs 2023).
H1 (Year ANOVA): At least one year mean differs.
H0 (Month ANOVA within a year): Mean Quantity is equal across months.
H0 (Two-way ANOVA): No main effects of Year or Month and no interaction.
H0 (MANOVA, if present): The joint means of (Quantity, deliveryHours) are equal across years.

Methods

• One-way ANOVA for Quantity ~ Year (2022 vs 2023).
• One-way ANOVA for Quantity ~ Month, run separately for 2022 and 2023.
• Two-way ANOVA for Quantity ~ Year × Month to assess interaction.
• MANOVA with response vector (Quantity, deliveryHours) ~ Year when deliveryHours is available.
Assumptions checked visually with boxplots (similar spreads, few extreme outliers). Significance level α = 0.05.

Results

Figure 1. SOF — Quantity by Year (Boxplot)

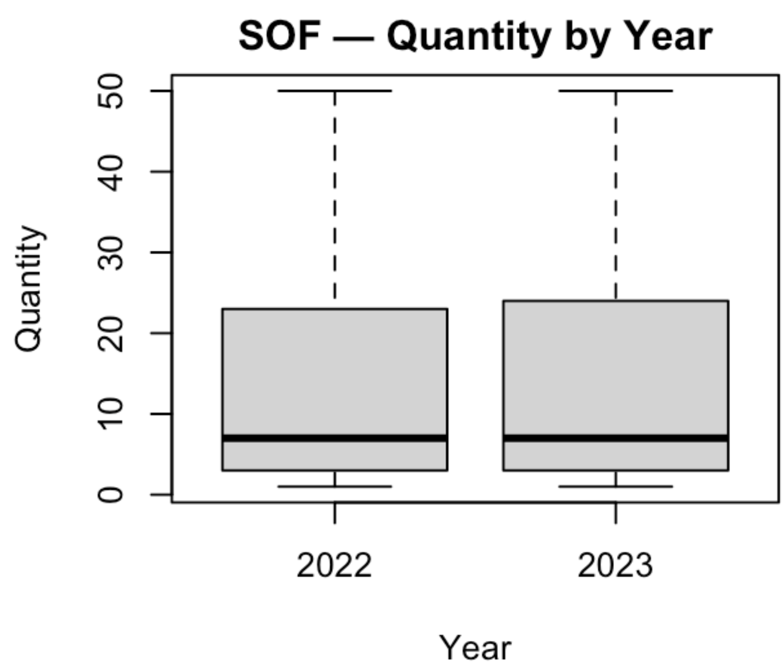
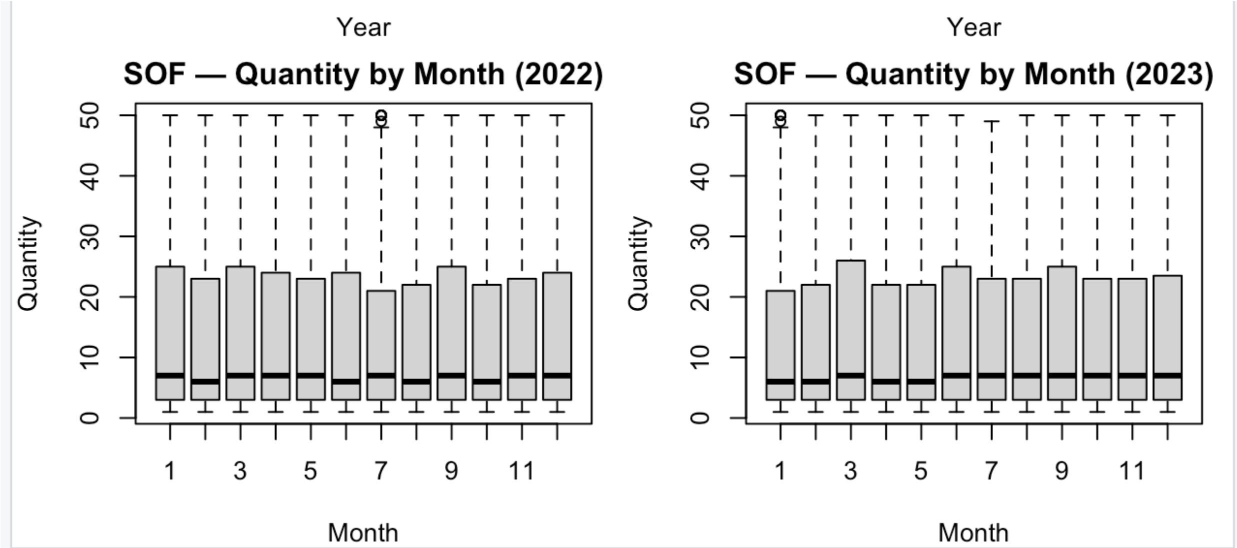


Figure 2. SOF — Quantity by Month (2022 vs 2023, Boxplots)



One-way ANOVA: Quantity ~ Year (SOF)

Source	Df	Sum Sq	Mean Sq	F, p-value
Year	1	98.79	98.79	F=0.52, p=0.471
Residuals	20,747	3,937,900	189.81	

One-way ANOVA: Quantity ~ Month (SOF, 2022)

Source	Df	Sum Sq	Mean Sq	F, p-value
Month	11	2,059	187.2	F=0.99, p=0.453
Residuals	11,115	2,102,177	189.1	

One-way ANOVA: Quantity ~ Month (SOF, 2023)

Source	Df	Sum Sq	Mean Sq	F, p-value
Month	11	1,641	149.2	F=0.782, p=0.658
Residuals	9,610	1,832,024	190.6	

Two-way ANOVA: Quantity ~ Year × Month (SOF)

Source	Df	Sum Sq	Mean Sq	F, p-value
Year	1	98.79	98.79	F=0.520, p=0.471
Month	11	2,147	195.19	F=1.028, p=0.418
Year:Month	11	1,553	141.15	F=0.744, p=0.697

MANOVA (if deliveryHours present): (Quantity, deliveryHours) ~ Year

Test	Statistic	approx F (num, den)	p-value
Pillai	0.0000339	0.352 (2, 20746)	0.703

Univariate follow-ups from MANOVA:
• Quantity ~ Year: F=0.5205, p=0.4706
• deliveryHours ~ Year: F=0.1789, p=0.6723

Interpretation

Across SOF, there is no statistical evidence of a difference in mean Quantity between 2022 and 2023 (ANOVA p=0.471). Within each year, months do not differ significantly (2022 p=0.453; 2023 p=0.658). The two-way ANOVA shows no main effects and no Year×Month interaction (all p>0.4). When deliveryHours is included, MANOVA indicates no multivariate year effect (Pillai p=0.703); univariate tests also show no year effect on either Quantity or deliveryHours.

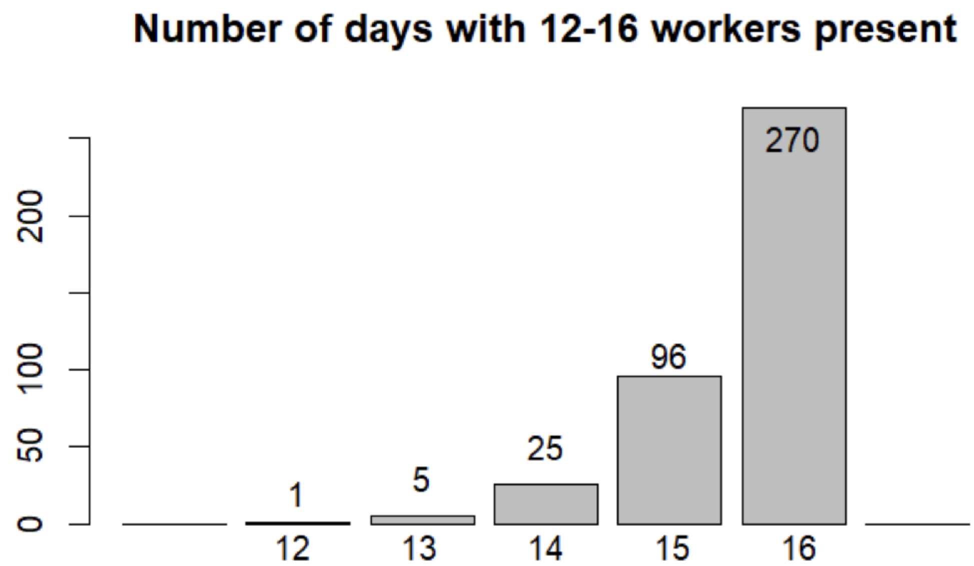
Practical Takeaways

• SOF demand levels are statistically stable between 2022 and 2023 and across months.
• Operational planning (inventory, staffing, logistics) can be set to steady-state levels for SOF, with monitoring for local/seasonal promotions rather than year-over-year shifts.
• If future strategy requires finer sensitivity, consider segmenting by city/channel or using robust tests/effect sizes.

Part 7 — Reliability of Service & Profit Optimisation (Car Rental)

Setup & Given Information

Data: 397 days of staffing levels with counts of workers on duty:
12→1 day, 13→5 days, 14→25 days, 15→96 days, 16→270 days.
A day is considered reliable if at least 15 people are on duty.
Figure 1. Number of days with 12–16 workers present



7.1 Estimated Reliable Days per Year

Reliable days in sample = 15-on-duty + 16-on-duty = 96 + 270 = 366.
Estimated reliability $\hat{p}$ = 366/397 = 0.9219.
Expected reliable days in a 365-day year = $\hat{p}$ × 365 = 0.9219 × 365 = 336.5 (≈ 336 days).
Expected unreliable days per year = 365 − 336.5 = 28.5 (≈ 29 days).
Approximate 95% CI for reliability: [0.896, 0.948] → 326.9–346.1 reliable days/year.

7.2 Profit Optimisation via Staffing

Model: Hire m permanent additional staff (same every day). A day is unreliable if W+m < 15.
Lost sales on an unreliable day = R20,000; annual cost per extra person = R25,000/month = R300,000/year.

Extra staff (m)	Unreliable days in 397	Expected bad days/yr	Lost sales (R)	Hire cost (R)	Total expected cost (R)
0	31	28.5	R 570,025	R 0	R 570,025
1	6	5.5	R 110,327	R 300,000	R 410,327
2	1	0.9	R 18,388	R 600,000	R 618,388
3	0	0.0	R 0	R 900,000	R 900,000
4	0	0.0	R 0	R 1,200,000	R 1,200,000

Optimal decision: hire m = 1 additional staff. Expected total annual cost ≈ R 410,327; expected reliable days ≈ 359.5.

Recommendation

Hire 1 additional permanent staff member. This reduces expected unreliable days from 28.5 to 5.5 (≈ 359.5 reliable days/year) and minimizes expected annual cost (R 410,327 vs R 570,025 with no hire).

CONCLUSION

This project transformed operational data into comprehensible and credible conclusions by utilizing statistical control charts, process capability analysis, risk diagnostics, and economic optimization. According to baseline studies, a number of streams were operating with instability and insufficient capability compared to the stated service needs. After data cleansing and rule-based changes, stability improved and capability estimates became interpretable, revealing which processes meet requirements and which require redesign. The inferential work (ANOVA/MANOVA) helped identify the factors that significantly affected performance, which helped determine where to begin the intervention. The results demonstrate that, from the standpoint of loss-minimization, concentrating on the objective and minimizing variance both lower predicted loss, not just the frequency of results that are not within specifications.

Implementable policies—staffing and resource-mix decisions that increase throughput/profit while maintaining variability within reasonable bounds—are produced by the optimization parts from an operations standpoint. These components work together to create a useful pathway: stabilize, confirm capability, optimize based on economic standards, and monitor to maintain gains. The work is robust enough to support prompt managerial action, decision-oriented (specified thresholds and trade-offs), and reproducible (documented steps and code).