ECSA Data Analysis

Quality Assurance 344

Daniel Wassenaar – 26945134

Contents

1.2. Data Analysis

Introduction

This report presents a comprehensive basic data analysis of the company's operational datasets. This analysis represents a foundational assessment of the available data to establish current business insights and identify key patterns.

The analysis follows a structured six-step methodology, as specified by ECSA, encompassing data loading and inspection, summary statistics, missing value handling, data filtering, visualization, and relationship exploration.

Four primary datasets have been analysed:

- Sales Data: Transaction records including customer purchases, quantities, and order timelines
- Products Data: Product information with pricing and markup details
- Customer Data: Demographic and geographic customer profiles
- Head Office Products Data: Additional product information from corporate headquarters

In the report we will look at each of the 4 datasets and apply ECSA's steps to analyse each of them respectively. The purpose of this report is to establish baseline metrics, uncover initial insights, and provide recommendations for further investigation. All analysis has been conducted using R, employing industry-standard techniques for data exploration and visualization to ensure reproducibility and clarity of findings.

Sales Data

- **Total Orders**: 100,000 transactions
- **Time Period**: January 2022 - December 2023
- **Data Quality**: Complete dataset with zero missing values across all columns

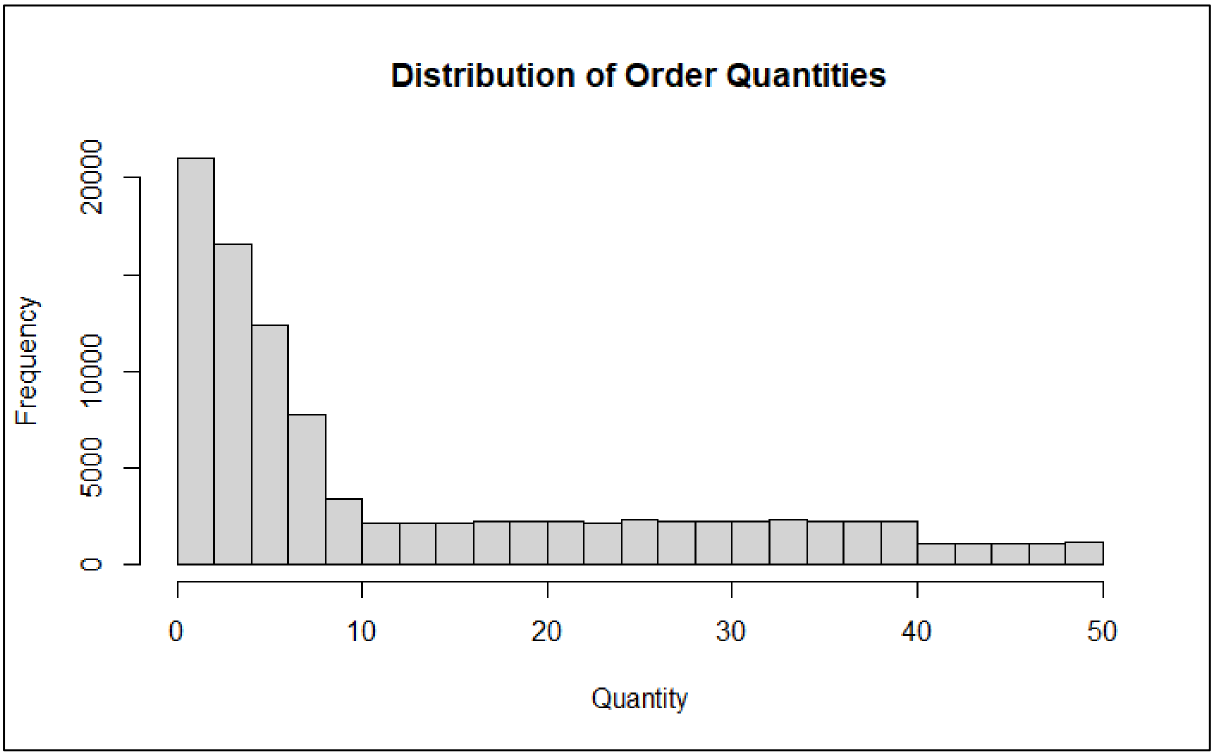

Figure 1 - distribution of order quantity

Key Performance Indicators for sales

- **Average Order Quantity:** 13.5 units per order
- **Order Timing**: Peak order times around 1:00 PM (orderTime: 12.93)
- **Operational Efficiency**:
  - Average picking time: 14.7 hour
  - Average delivery time: 17.5 hours

Sales Temporal Patterns

Monthly Distribution (2022-2023):

- Highest activity: February (8,912 orders), April (8,959 orders), November (8,912 orders)
- Lowest activity: January (6,134 orders), December (5,304 orders)
- **Year-over-Year**: 53,727 orders in 2022 vs 46,273 orders in 2023

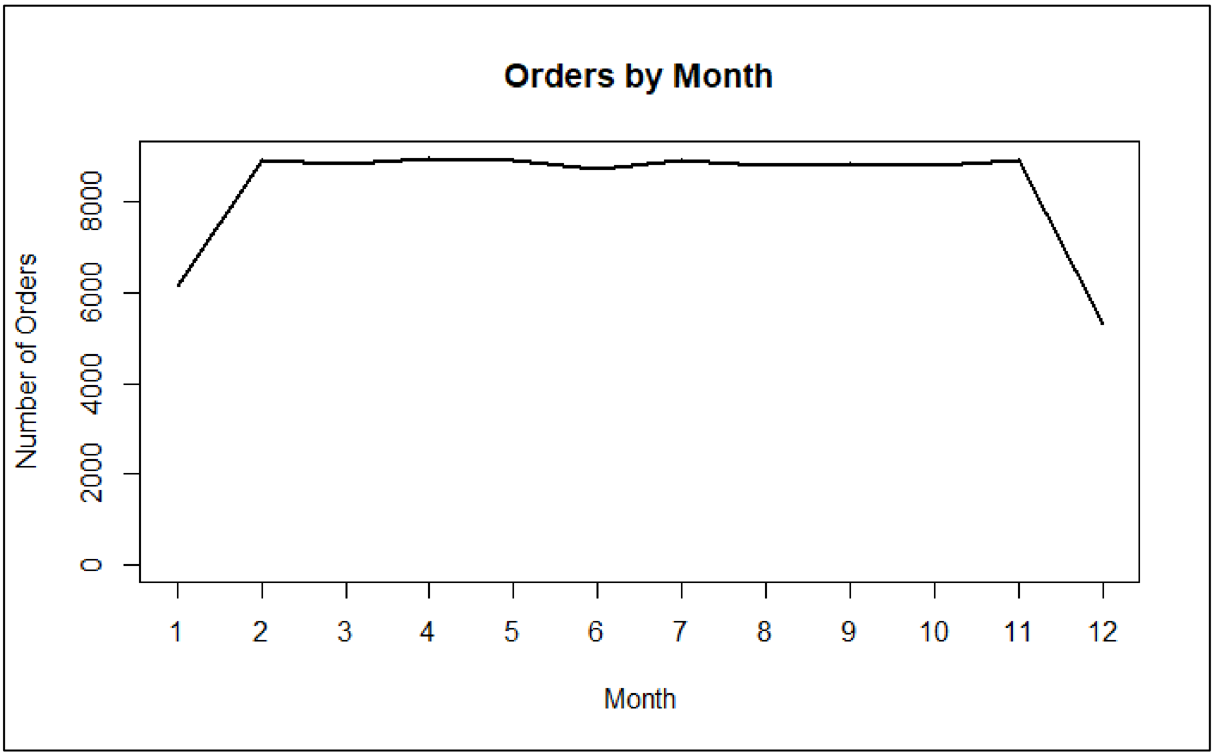

Figure 2 - monthly order quantities

Order Patterns: Consistent order volume across months 2-11, with noticeable dips in January and December

Sales Operational Insights

The boxplot analysis of delivery hours by month shows consistent operational performance throughout the year, with no significant seasonal variations in delivery efficiency.

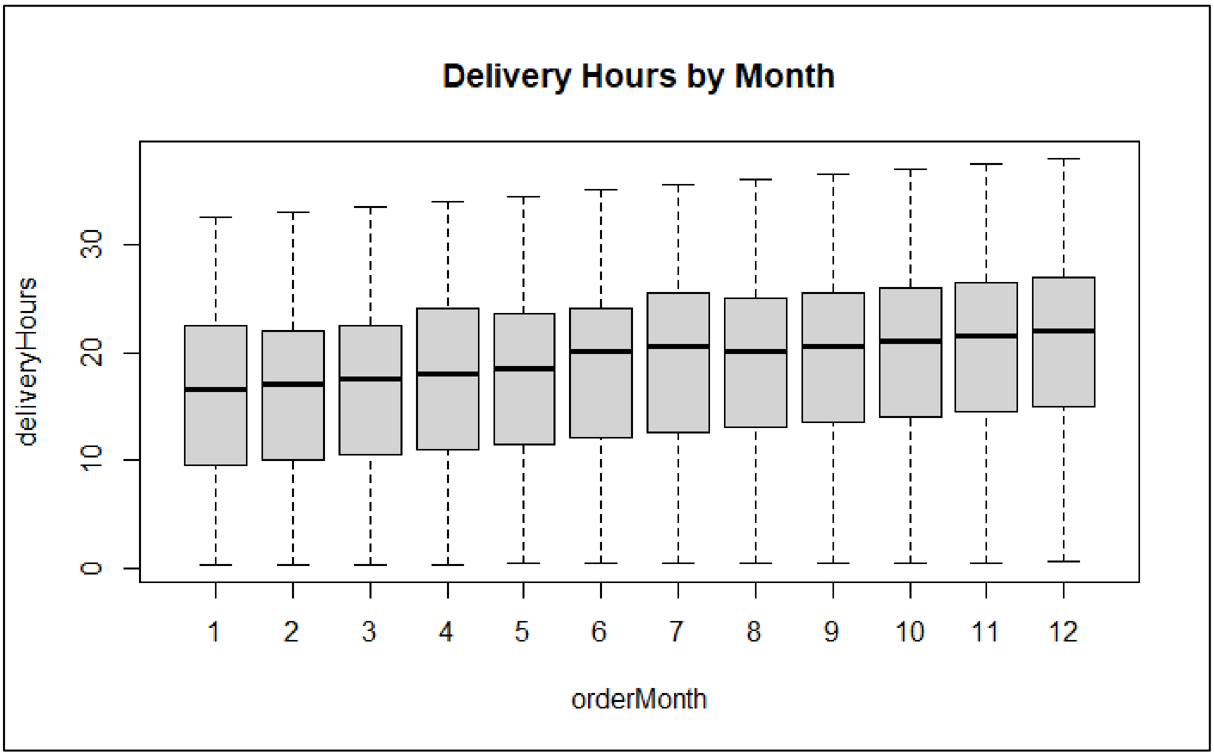

Figure 3 - Monthly delivery hours

Customer Data Analysis

- **Geographic Distribution**: 5,000 unique customers Balanced across 7 major US cities
    - San Francisco: 780 customers (15.6%)
    - Los Angeles: 726 customers (14.5%)
    - New York: 726 customers (14.5%)
    - Chicago: 724 customers (14.5%)
    - Houston: 724 customers (14.5%)
    - Seattle: 673 customers (13.5%)
    - Miami: 647 customers (12.9%)

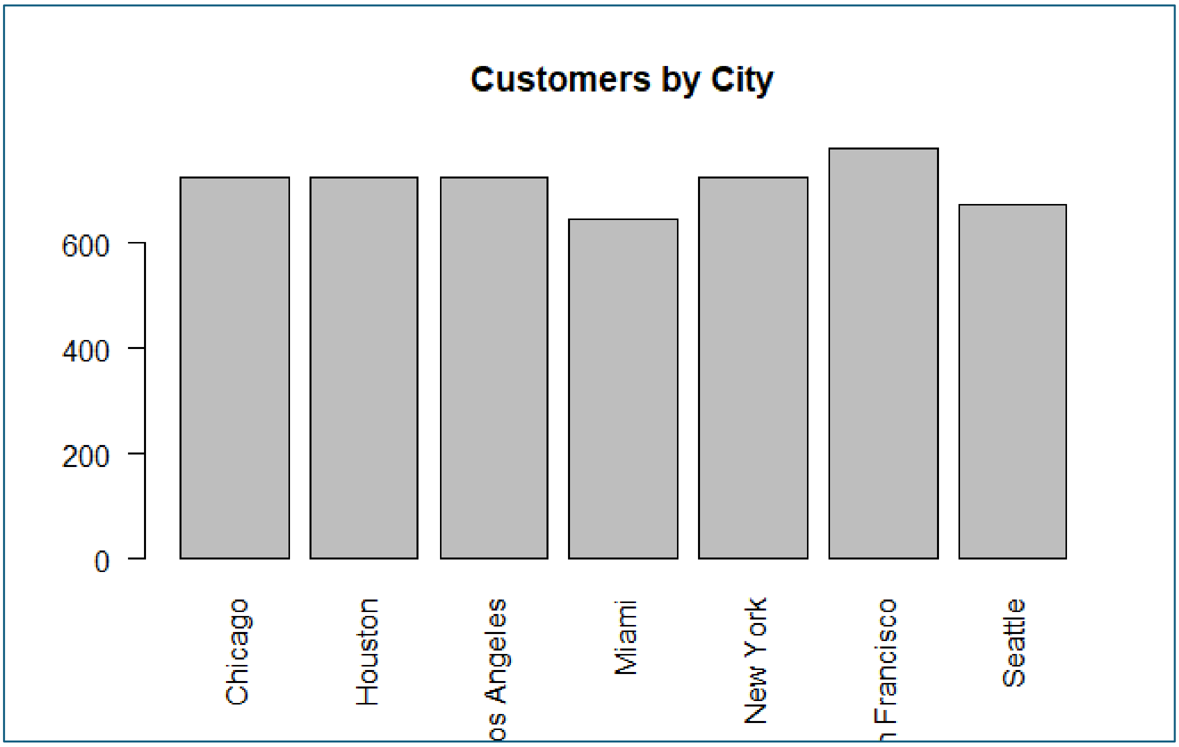

Figure 4 - Customers by city

The geographic distribution demonstrates remarkable balance, with no single city dominating the portfolio. San Francisco represents the largest segment at 15.6% (780 customers), while Miami constitutes the smallest at 12.9% (647 customers). This equitable spread across diverse markets from coast to coast (ranging from 12.9% to 15.6% per city) indicates strong nationwide penetration and reduces geographic concentration risk, providing a stable foundation for targeted regional strategies and future expansion initiatives.

Demographic Profile

- **Age Distribution**: Broad customer base with mean age of 51.6 years (range: 16-105)

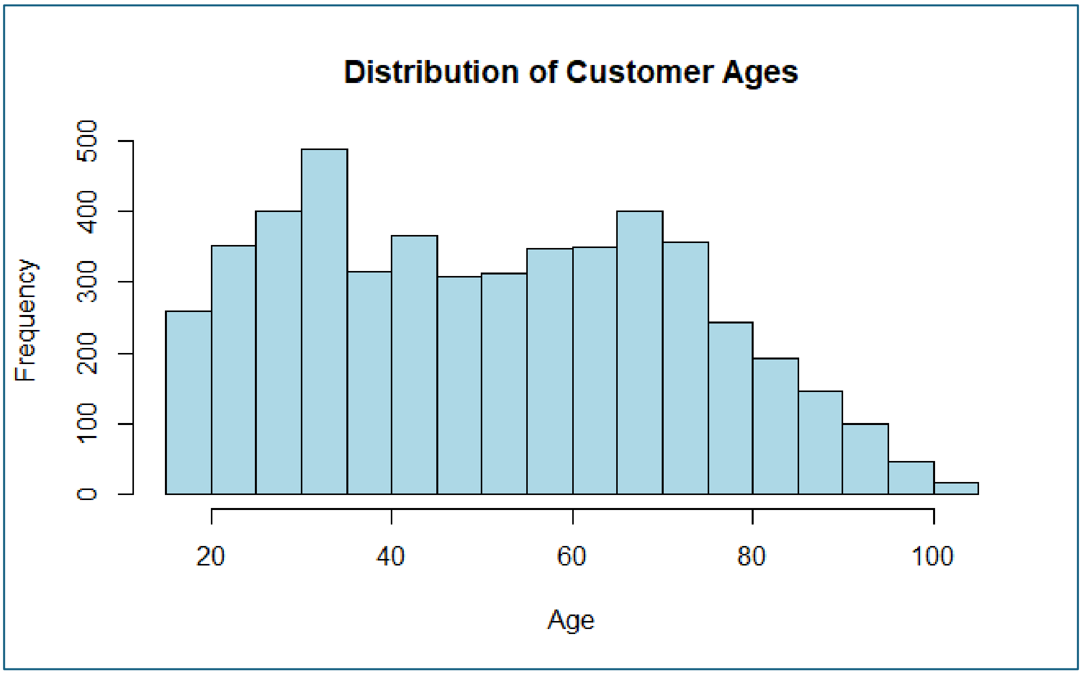

Figure 5 - Age distribution of customers

- **Income Levels**: Average income $80,797, indicating a middle-to-upper income customer base

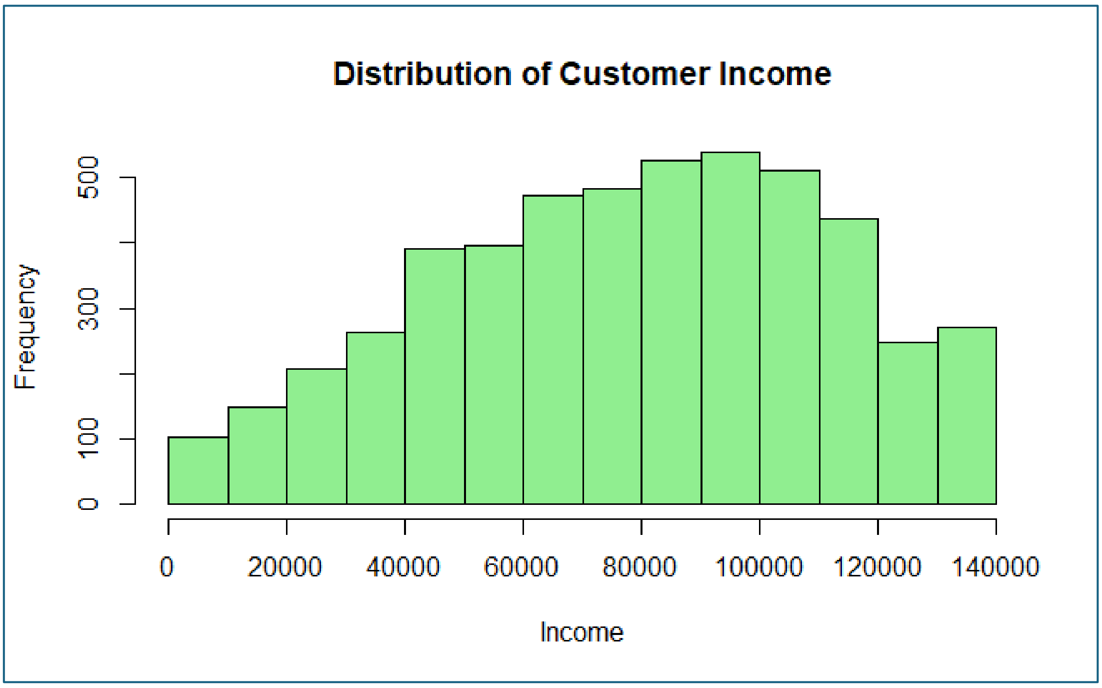

Figure 6 - Income distribution of customers

- **Gender Distribution**: Balanced between Female (48.6%) and Male (47.0%), with 4.4% Other

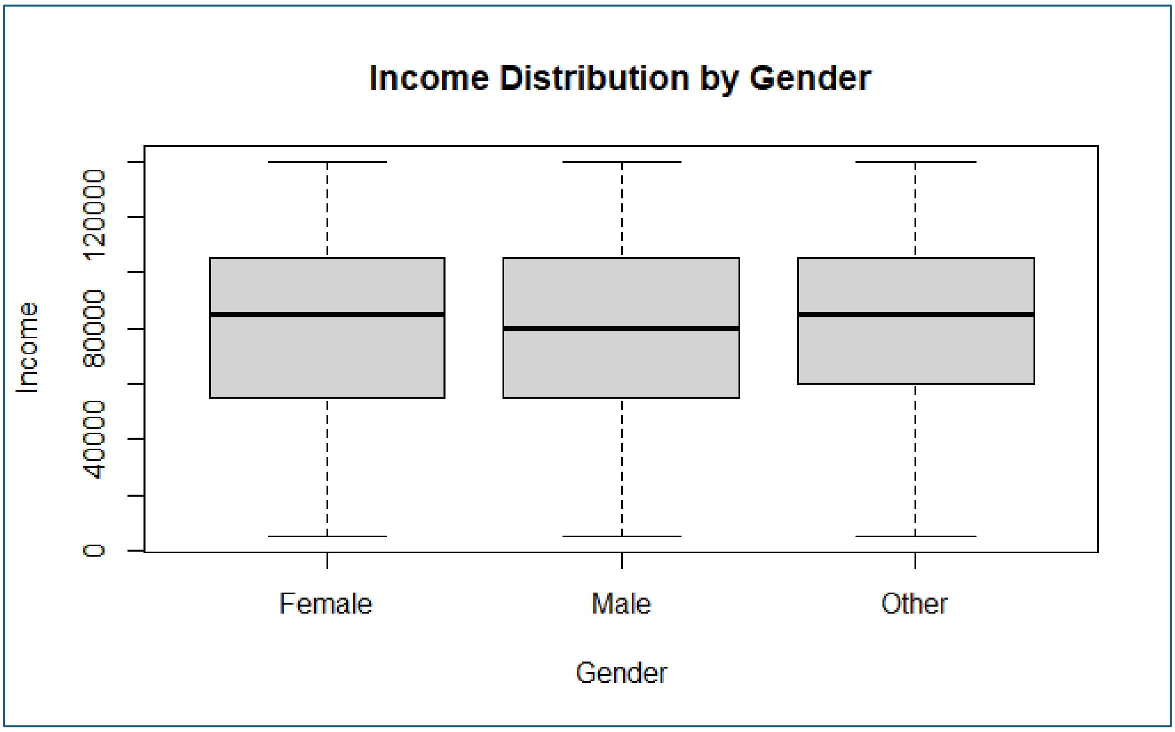

Figure 7 - Gender income distribution

Operational Product Analysis

Pricing Structure by Category:

- **Laptops**: Premium pricing segment (examples: LAP022 at $16,644, LAP021 at $19,495)
- **Monitors**: Mid-to-high range (MON040 at $5,346, MON033 at $5,573)
- **Peripherals**: Lower price points (keyboards $530-$752, mice $350-$425)

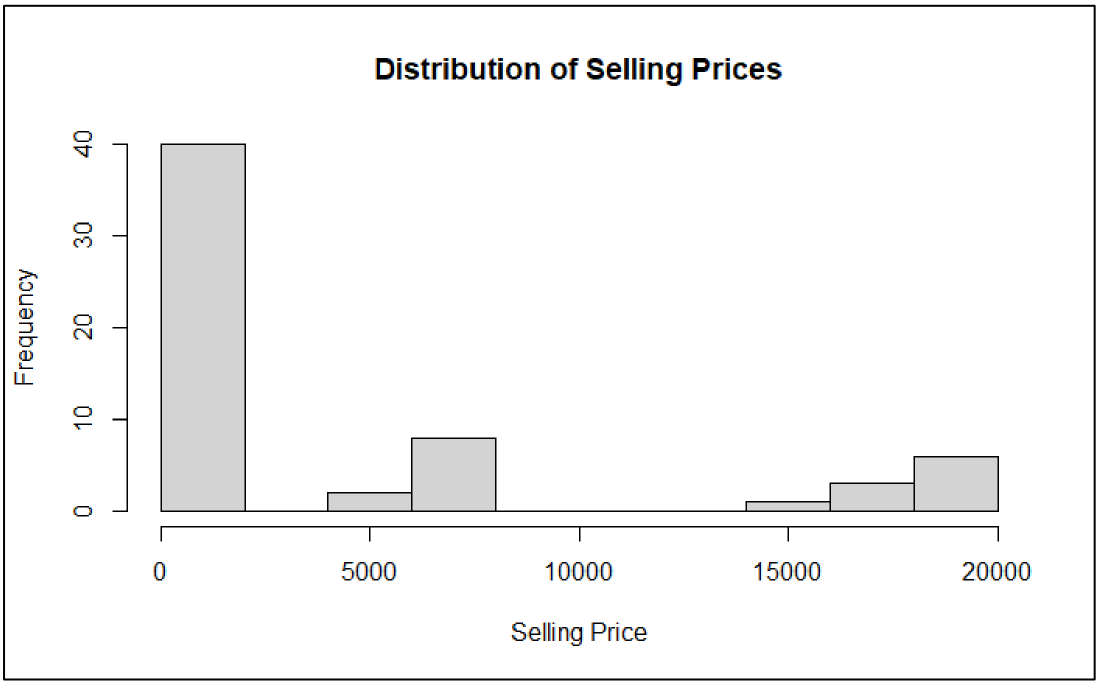

Figure 8 - Distribution of operational selling process

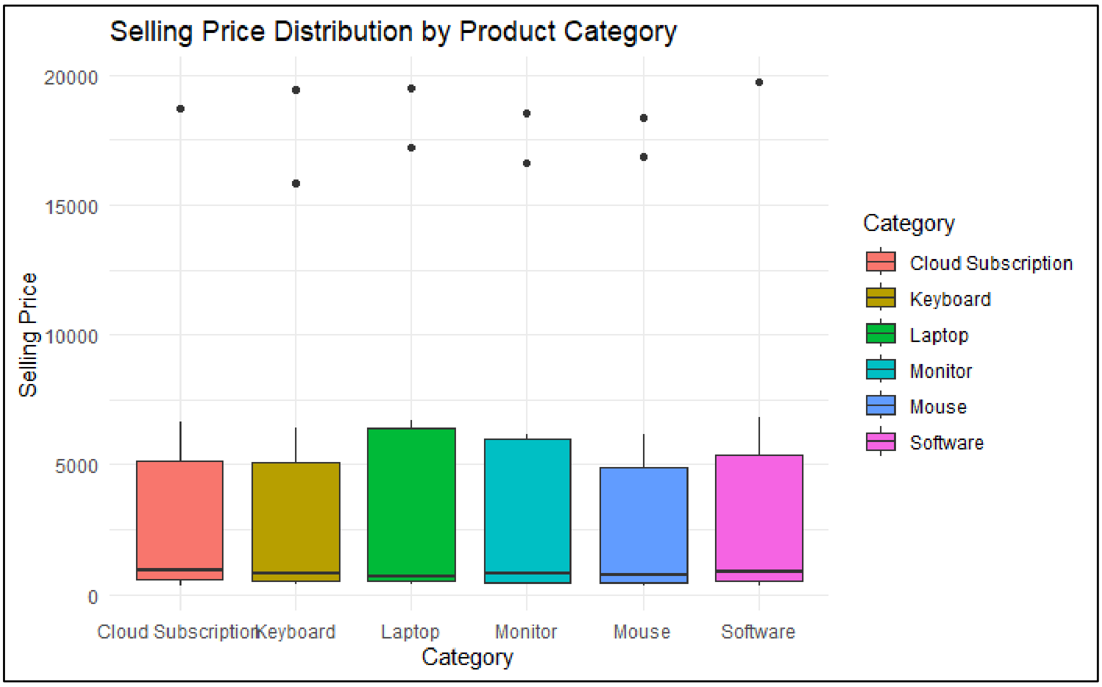

Figure 9 - operational selling prices by category

Markup Analysis:

- Highest markups: Monitors and Laptops (27-30%)
- Lowest markups: Software and Cloud subscriptions (10-13%)
- Overall markup range: 10.13% to 29.84%

We can observe the relationship between selling price and markup in the following diagram:

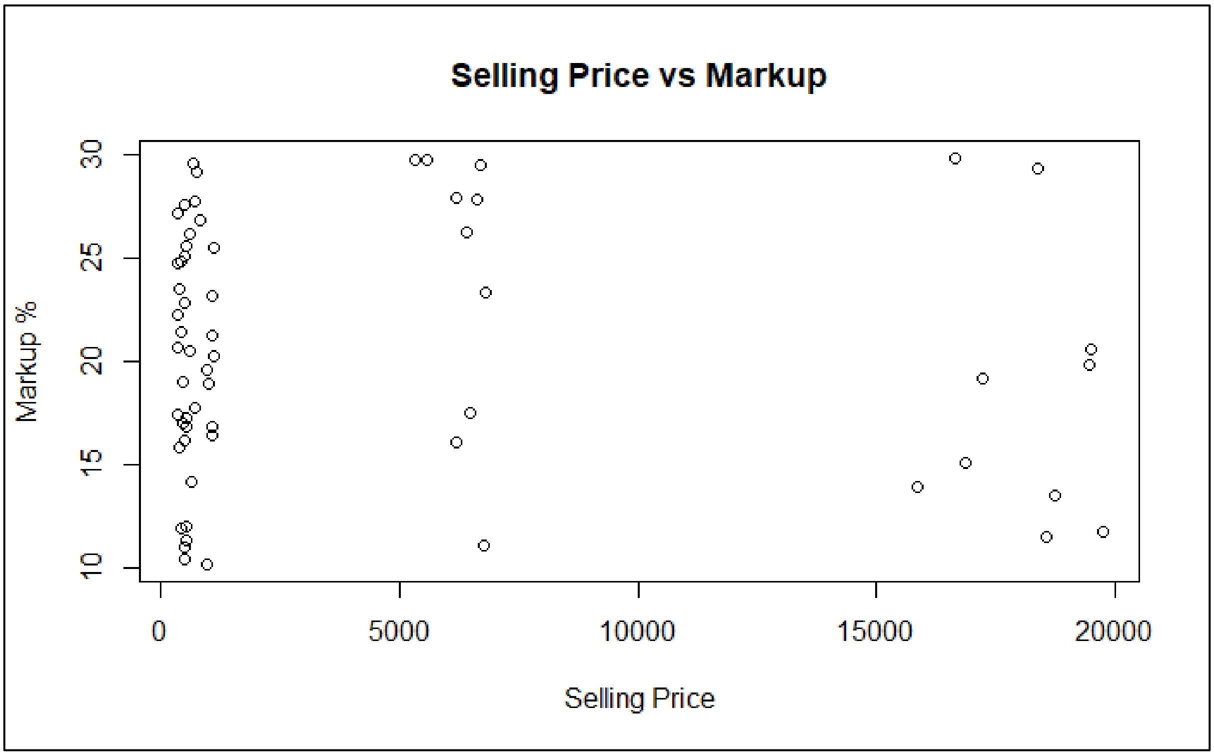

Figure 10 - Selling price to Markup for Operational products

Head Office Products Analysis

Comparative Analysis:

- **Product Categories**: Perfect balance across 6 categories (60 products each)
- **Pricing**: Mean selling price $4,411 (vs operational data's premium skew)
- **Markup Consistency**: Average markup 20.39%, closely aligned with operational data

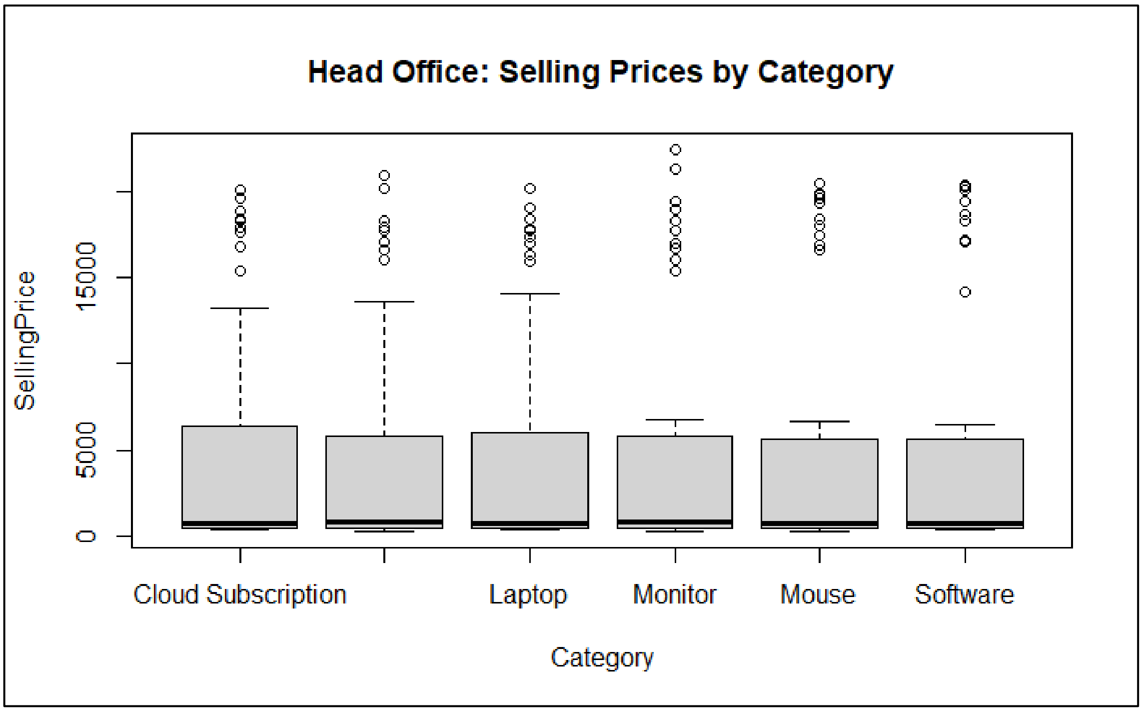

Figure 12 - Selling price by category for head office database

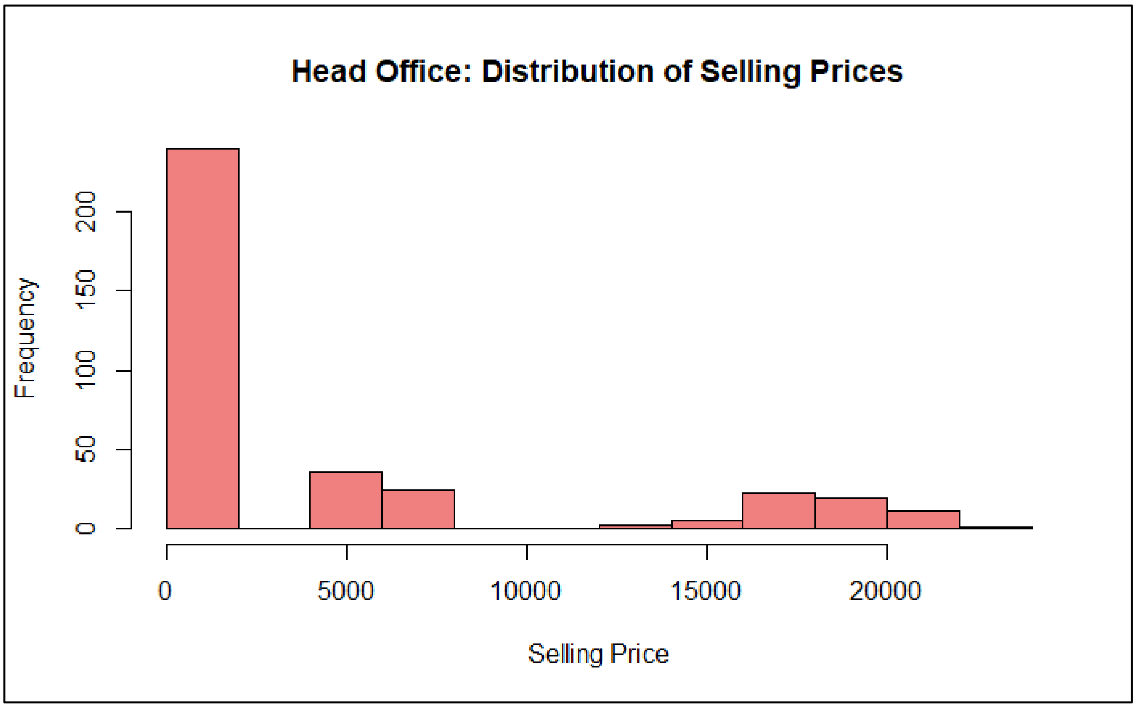

Figure 11 - Selling price distribution for head office database

As with the operational product data, we can once again observe the relationship between sales price and markup.

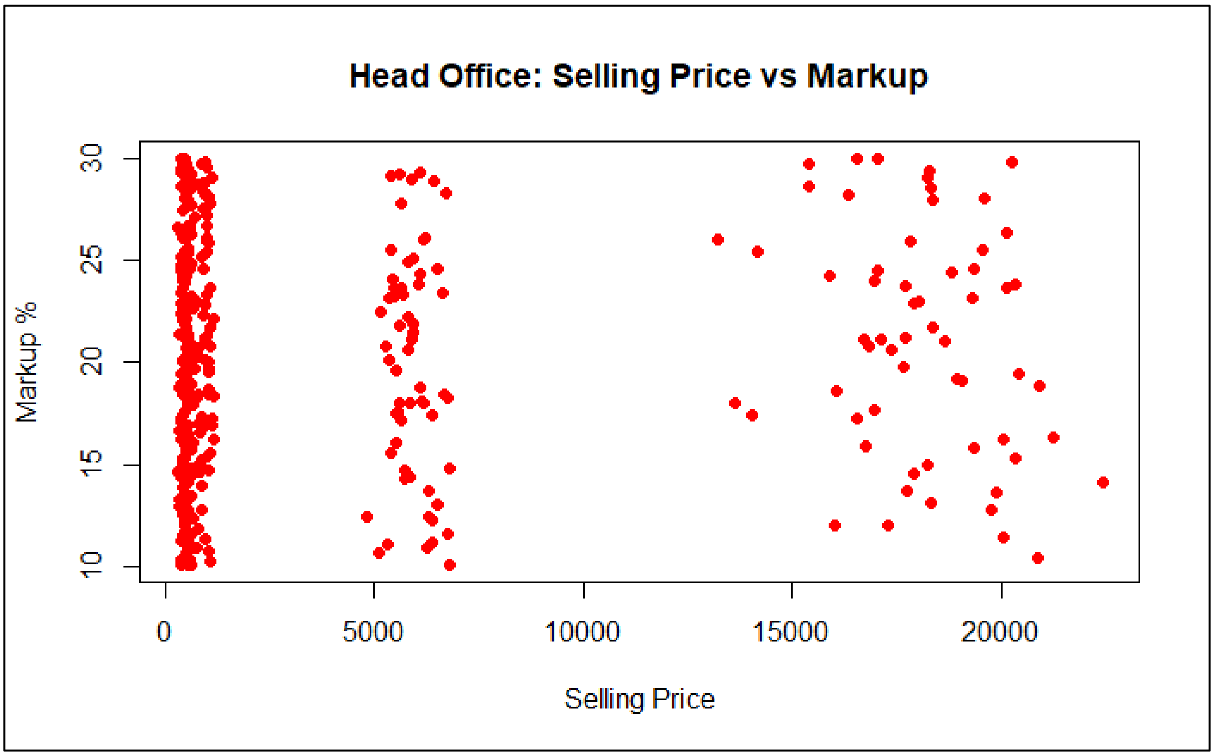

Figure 13 - Selling price to markup for head office database

Key Findings in Comparing Operational and Head Office Product Data:

Both the operational and head office datasets share identical category structures, ensuring consistency in product classification. The head office data exhibits a more normalized pricing distribution, indicating refined data management and pricing control. Markup strategies are aligned across both datasets, showing cohesive pricing policies. Additionally, both datasets are fully complete with no missing values, demonstrating excellent data quality.

Key Findings and Recommendations:

Strengths:
All datasets exhibit 100% completeness, confirming that no data is missing. The customer base is well-balanced, showing strong geographic and demographic diversity. Product strategies are consistent between head office and operational levels, indicating strategic alignment across the organization.

Opportunities:
Seasonal trends reveal noticeable sales dips in January and December, suggesting potential for improved seasonal planning. The high-income customer segment presents an opportunity for premium offerings or targeted promotions. Expanding the range of lower-priced, high-margin products could enhance profitability while maintaining competitiveness.

Recommendations:
Marketing efforts should focus on boosting sales during traditionally low-volume months such as January and December. Loyalty programs could strengthen customer retention across the diverse customer base. The existing balanced product category approach should be maintained to preserve alignment between head office and operations. Finally, operational efficiency should continue at its current high standard to sustain strong delivery performance.

Conclusion

The analysis reveals a healthy, well-operated business with strong fundamentals across sales, customer base, and product management. The consistent patterns across datasets and complete data integrity provide a solid foundation for future growth and strategic planning.

3. Executive Summary of Part 3

This report presents a comprehensive statistical process control (SPC) analysis of delivery time performance for sixty product types over the 2026-2027 operational period. The study reveals significant quality concerns, with zero processes meeting the minimum capability standard of Cpk ≥ 1.33. Systematic issues in process variability and centering were identified across all product categories, necessitating immediate corrective action and long-term quality improvement initiatives.

Key Findings:
All sixty processes (100%) failed to meet the required capability standards, with an overall mean Cpk value of 0.72, indicating a substantial need for process improvement. Hardware products exhibited consistent failure patterns, while software products showed distinct capability characteristics. In addition, several control chart violations were observed, highlighting process instability across product categories.

Project Background:
This analysis was initiated after the unexpected departure of the previous data analyst to establish baseline process capability metrics across all product lines. The study employed statistical process control (SPC) methods to evaluate delivery time performance and uncover opportunities for improvement.

Objectives:
The primary goals of the project were to establish control charts for delivery time monitoring, calculate process capability indices (Cp and Cpk), and identify control chart violations that signal instability. Further aims included providing data-driven recommendations for process improvement and determining compliance with customer specifications, with an upper specification limit (USL) of 32 hours.

Methodology:

Data Collection and Preparation:
Delivery time data for sixty product types was extracted from the 2026–2027 sales database. The dataset was arranged chronologically to support accurate time-series analysis and to simulate real-time process monitoring conditions.

Control Chart Implementation:
X-bar and s-charts were created for each product type. The first 30 samples, representing 720 deliveries, were used to establish control limits. Subsequent samples were then monitored for control violations. The Western Electric rules were applied to detect special cause variations.

Process Capability Analysis:
Capability indices were determined using lower and upper specification limits (LSL = 0 hours, USL = 32 hours) and a capability benchmark of Cpk ≥ 1.33. The formulas used were Cp = (USL − LSL) / (6σ) and Cpk = min[(USL − μ) / (3σ), (μ − LSL) / (3σ)]. The analysis considered the first 1000 deliveries for each product type to ensure statistical reliability.

Violation Detection:
Three main types of violations were tracked: points on the s-chart falling outside the 3σ limits (indicating process variability problems), four consecutive points beyond the 2σ limits (indicating mean shifts), and extended runs within the 1σ limits (indicating exceptional control).

3.1 Control charts

Example SPC charts to demonstrate code is capable of initialising X-charts and s-charts for every product type

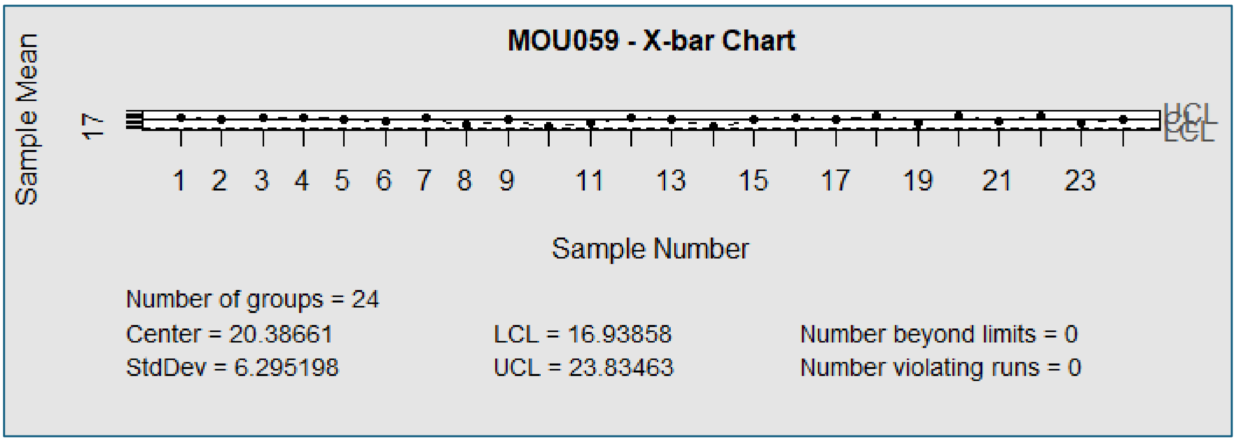

Figure 15 - Example Xbar chart 1

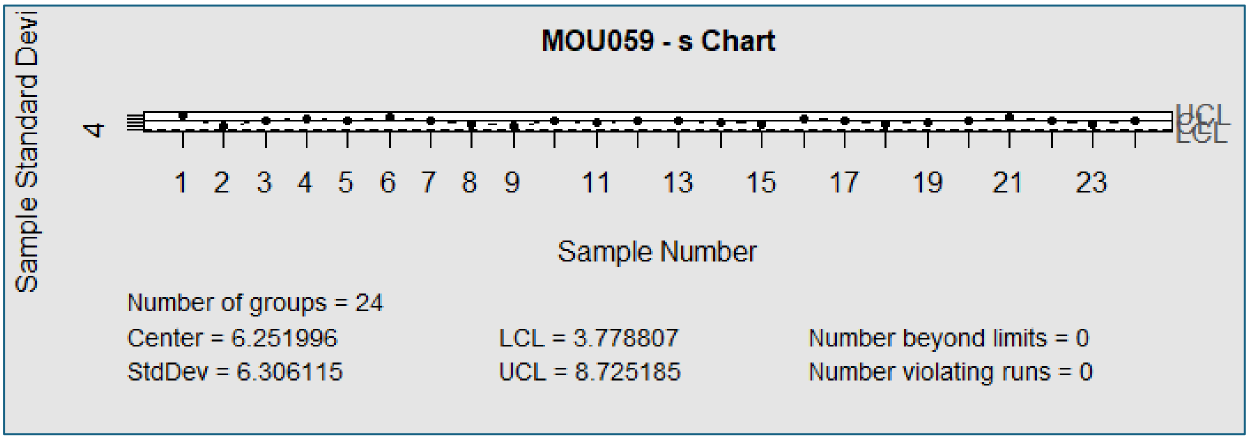

Figure 14 - Example s Chart

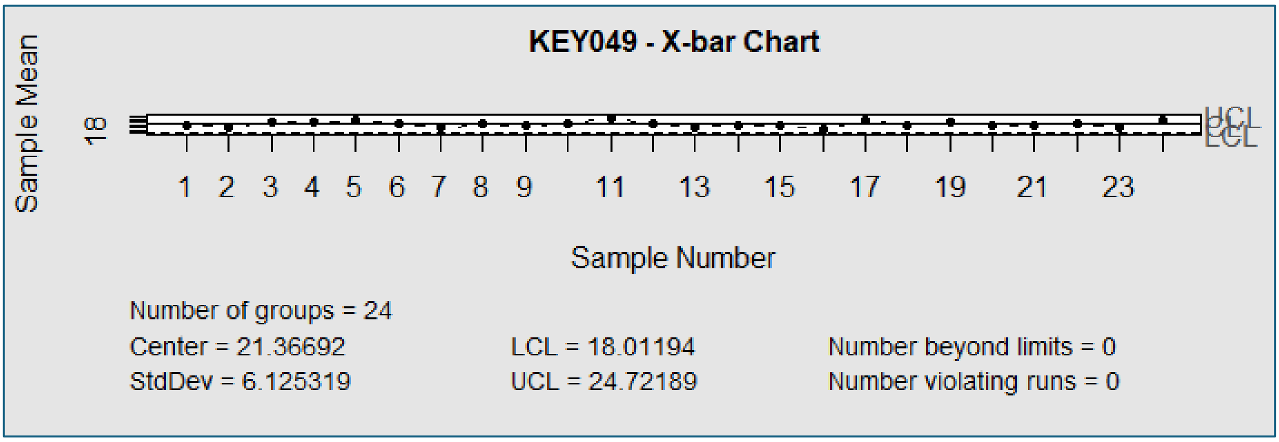

Figure 16 - Example Xbar chart 2

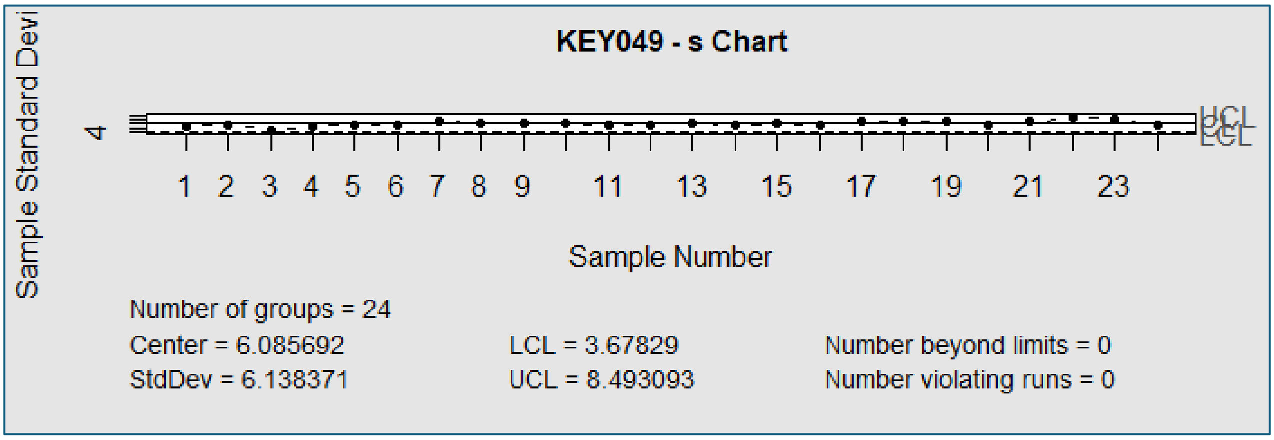

Figure 17 - Example s Chart 2

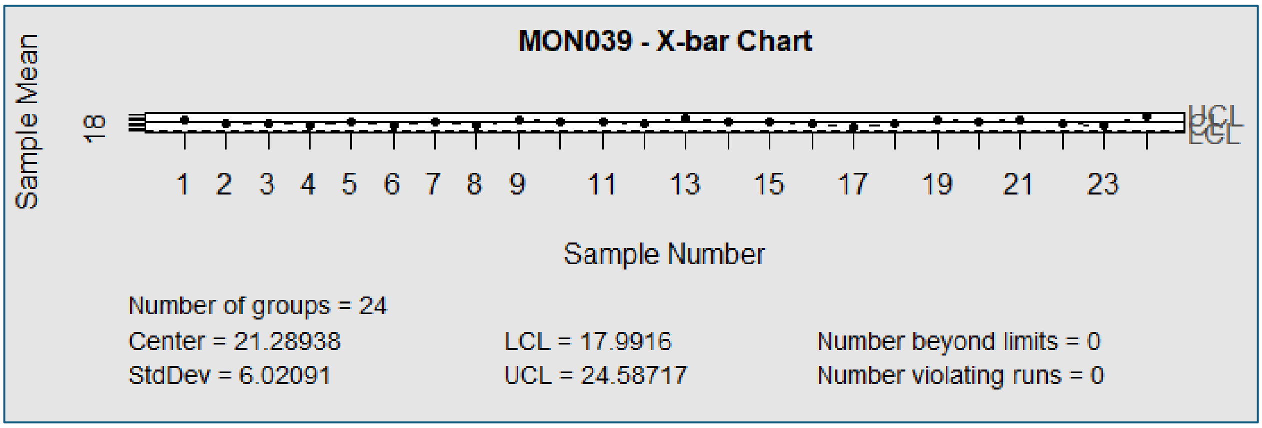

Figure 19 - Example Xbar chart 2

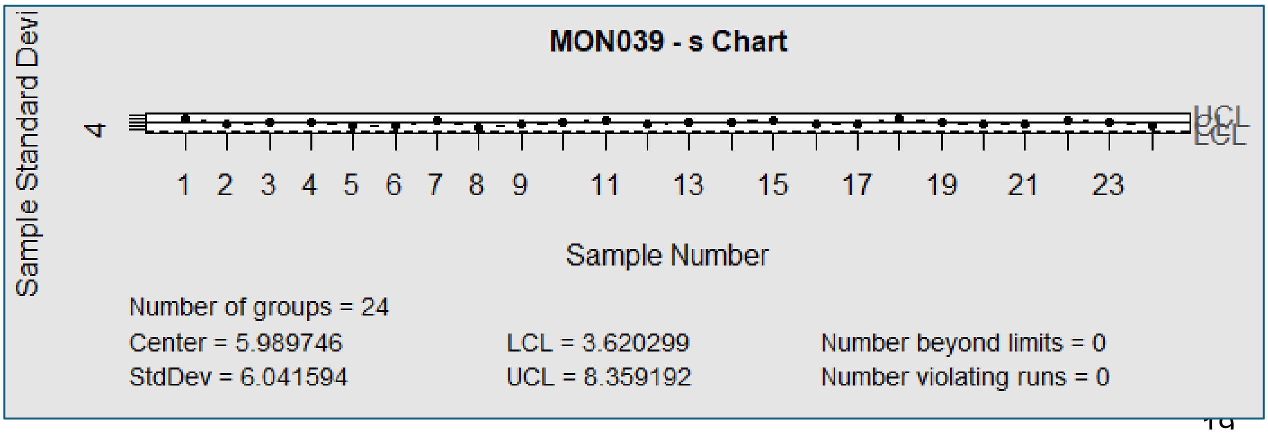

Figure 18 - Example s Chart 3

3.2 RESULTS AND ANALYSIS

The full list of all 60 products can be viewed in Appendix A

Process Capability Assessment

Figure 21 - Process capability distribution

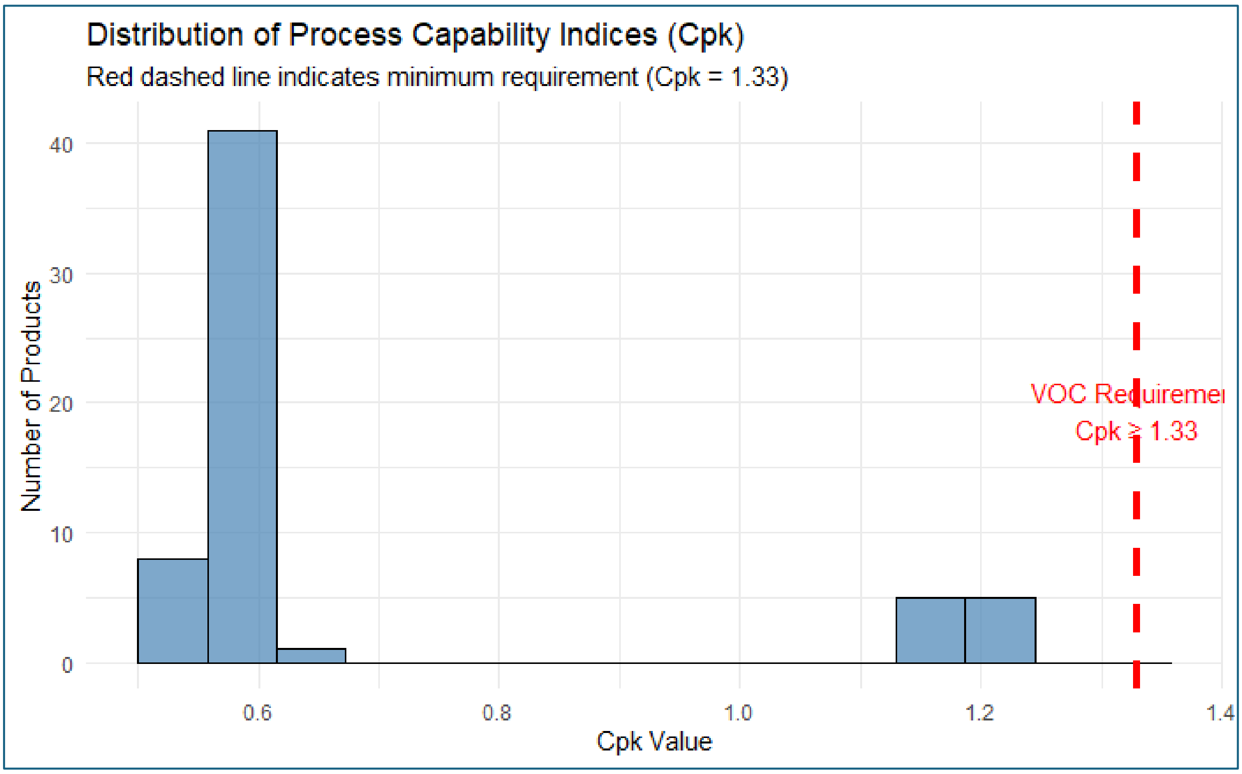

Figure 20 - Disribution of CPK values

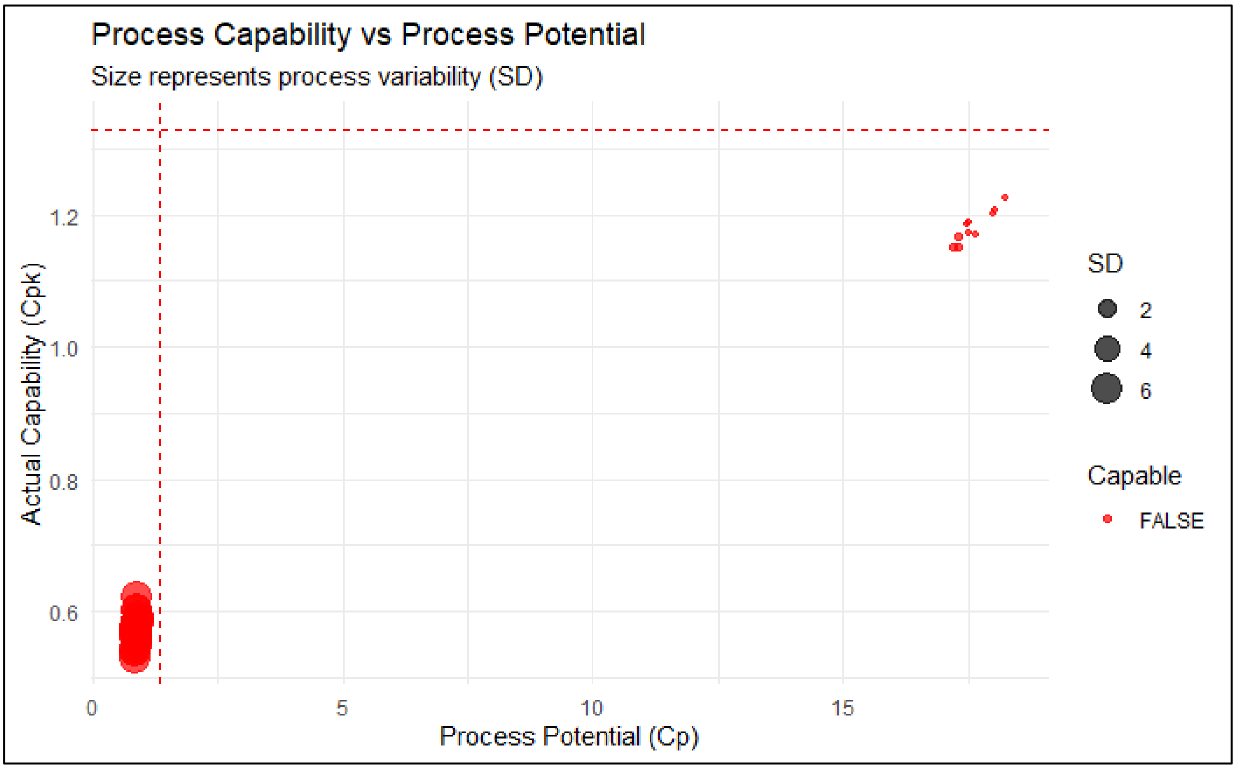

Figure 22 - Cp vs Cpk

Overall Capability Status

The capability analysis reveals that no processes meet the minimum requirement of Cpk ≥ 1.33. The distribution of Cpk values shows:

- **Critical Range** (Cpk < 1.00): 50 products (83.33%)
- **Near Capable** (1.00 ≤ Cpk < 1.33): 10 products (16.67%)
- **Fully Capable** (Cpk ≥ 1.33): 0 products (0%)

Worst Performing Products

The lowest capability indices were observed in:

ProductType	Cp	Cpu	Cpl	Cpk	Mean	SD	Capable
KEY049	0.845	0.529	1.161	0.529	21.985	6.314	FALSE
KEY045	0.847	0.538	1.156	0.538	21.837	6.297	FALSE
KEY050	0.850	0.539	1.162	0.539	21.861	6.273	FALSE
LAP028	0.856	0.544	1.168	0.544	21.839	6.230	FALSE
MOU053	0.865	0.547	1.182	0.547	21.875	6.169	FALSE

Table 1 - The 5 worst performing products

These products show both low process potential (Cp < 0.87) and poor centering (Cpk ≈ 0.53), indicating high variability and misalignment with specifications. The large standard deviations and off-center means suggest inconsistent performance and frequent out-of-spec results. Overall, these processes require tighter control and centering adjustments to achieve acceptable capability.

Best Performing Products

ProductType	Cp	Cpu	Cpl	Cpk	Mean	SD	Capable
SOF008	18.237	35.247	1.226	1.226	1.076	0.292	FALSE
SOF003	18.050	34.893	1.206	1.206	1.069	0.295	FALSE
SOF010	17.990	34.777	1.202	1.202	1.069	0.296	FALSE
SOF007	17.516	33.843	1.189	1.189	1.086	0.304	FALSE
SOF009	17.486	33.785	1.187	1.187	1.086	0.305	FALSE

Table 2 - 5 Best performing products

It can be observed that the software products exhibit the highest overall capability indices among all product categories. Their Cp values, which range from approximately 17.5 to 18.2, indicate exceptionally high potential capability — meaning that the inherent process variation (as reflected by the standard deviation) is very small relative to the specification limits. This suggests that, in theory, these processes could easily meet customer requirements if they were properly centered.

However, despite the impressive Cp values, the corresponding Cpk values (around 1.19 or lower) reveal a significant centering issue. The mean values for these products are shifted toward one side of the specification range, reducing the effective process capability. This misalignment between process mean and specification limits means that even a small drift could result in nonconforming outcomes. In summary, while the software processes are statistically precise and consistent, they are not well-centered and thus fail to achieve true capability under real operational conditions.

3.3 Chart Violations

Three products exhibited special cause variation in process variability:

- MOU059: Sample 86
- MOU055: Sample 47
- LAP021: Sample 35

These violations indicate intermittent process instability requiring root cause investigation.

Severe Violation Patterns:

Multiple products demonstrated systematic process mean shifts:

- MOU054 had 8 occurrences of 4+ consecutive points above 2σ limits
- MOU055 also had 8 occurrences with extended violation streaks
- SOF002 had 6 occurrences with violation durations up to 9 samples

Exceptional Process Control

Several products demonstrated extended periods of statistical control:

- MON039 had 23 consecutive samples within control limits
- SOF002 had 16 consecutive samples of stable operation
- SOF001 had 15 consecutive samples demonstrating process stability

Hardware Products (MOU, KEY, LAP, MON, CLO)

Characteristics:

- Consistent mean delivery times: 20-22 hours
- High process variability: Standard deviations of 5.95-6.32 hours
- Poor capability: Cpk values 0.53-0.62
- Primary issue: Combined high variability and poor centering

4.3.2 Software Products (SOF Series)

Characteristics:

- Low mean delivery times: 1.02-1.09 hours
- Low process variability: Standard deviations of 0.30-0.31 hours
- Moderate capability: Cpk values 1.19-1.20

Primary issue: Process centering within tight natural tolerances

DISCUSSION

Technical Interpretation

The universal failure to meet capability standards indicates systemic quality issues rather than isolated process problems. The consistent pattern across hardware products suggests common root causes in delivery operations, potentially related to logistics, supplier management, or operational procedures.

The software products' unique profile with extremely high Cp values but moderate Cpk indicates that while these processes have excellent inherent potential, they are not properly centered within their specification limits. This represents a significant opportunity for improvement through process adjustment.

Root Cause Analysis

The data suggests several potential root causes:

1. **Process Variability**: Hardware products show excessive variation, potentially indicating inconsistent operational practices, supplier quality issues, or inadequate process controls.
2. **Process Centering**: Most processes operate near the upper specification limit, suggesting missed optimization opportunities and potential systematic bias in delivery operations.
3. **Systematic Timing Issues**: The clustering of violations in specific sample ranges (31-45, 75-88) indicates potential seasonal, workload, or resource allocation problems.
4. **Specification Limit Considerations**: The uniform application of 32-hour USL across all product types may not be appropriate given the different inherent process characteristics.

Engineering Implications

From an engineering perspective, the data indicates that current processes are not capable of consistently meeting customer requirements. This has implications for:

- Quality assurance systems
- Customer satisfaction and retention
- Operational efficiency
- Cost of poor quality
- Competitive positioning

CONCLUSION

This statistical process control analysis has identified critical quality issues across all product delivery processes. The complete absence of capable processes (Cpk ≥ 1.33) indicates fundamental problems in process design, control, and maintenance that require systematic intervention.

The findings demonstrate two distinct failure patterns: hardware products suffering from excessive variability and poor centering, and software products with excellent inherent potential but suboptimal centering. Both patterns represent significant improvement opportunities.

The presence of extended stable control periods in some products (notably MON039's 23-sample streak) demonstrates that process stability is achievable and provides a benchmark for improvement targets.

Immediate management attention and systematic quality improvement initiatives are required to address these issues. The recommended action plan provides a structured approach to process capability enhancement, beginning with emergency stabilization and progressing through systematic quality system development.

Future work should focus on root cause analysis of the identified issues, implementation of the recommended improvements, and ongoing monitoring of process capability metrics to track improvement progress.

4. Risk, Data correction and Optimising for maximum profit

4.1 Type I Error

Estimate the likelihood of making a Type I (Manufacturer's) Error for A, B and C in the previous week's work.

Rule A - One point outside 3-sigma limits:

Probability = 0.0027

This means we'd expect 0.27 % false alarms

Rule B - Points between -1 and +1 sigma:

Probability single point in [-1σ, +1σ] = 0.6827

This is about identifying sustained good control, not an error rule

Rule C - 4 consecutive points outside ±2 sigma limits:

Probability one point outside ±2σ = 0.0455

Probability 4 consecutive outside SAME 2σ limit = 0

Probability 4 consecutive outside EITHER 2σ limit = 1e-06

4.2 Type II Error

Estimate the likelihood of making type II (Consumer's) Errors for a bottle filling process, that should be centered on 25.05 litres as the process average and CL of an xbar chart, and has an UCL of 25.089 and LCL of 25.011 litres.

Process shifted to: 25.028 litres

New standard deviation: 0.017 litres

Control Limits: LCL = 25.011 UCL = 25.089

Type II Error Probability (Beta): 0.9873

Power to detect shift: 0.0127

This means we'll miss the shift 98.7 % of the time

4.3 Corrected Analysis of Week 1

Part 4.3 required correction to the products_Headoffice and products_data databases. Following this, the same code was run as in week 1, to observe the difference in the results.

There were considerable differences in the Selling Price distributions for both sets pf data. This is evident in the figures below.

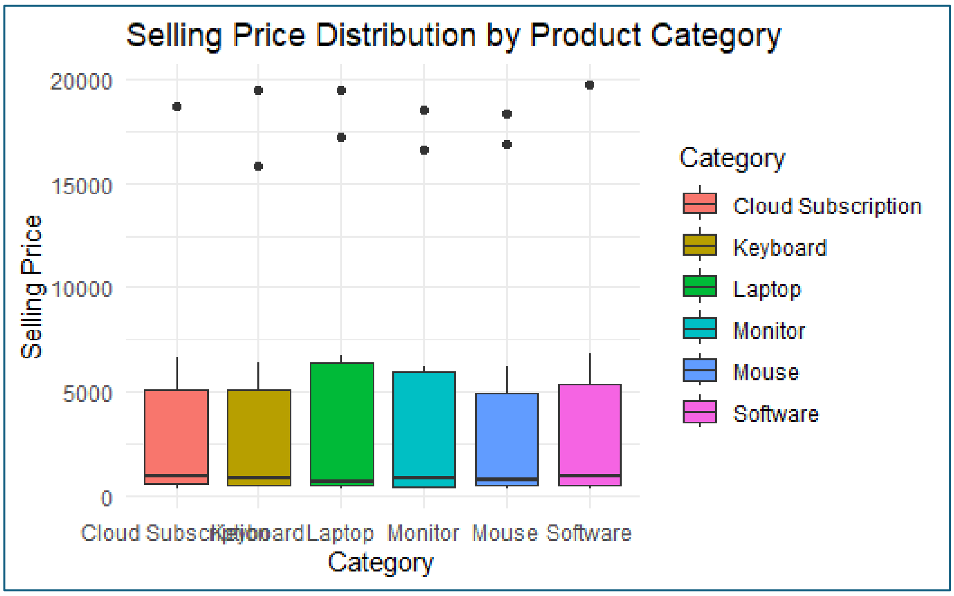

Figure 24 - original, incorrect selling price distribution for operational sales data

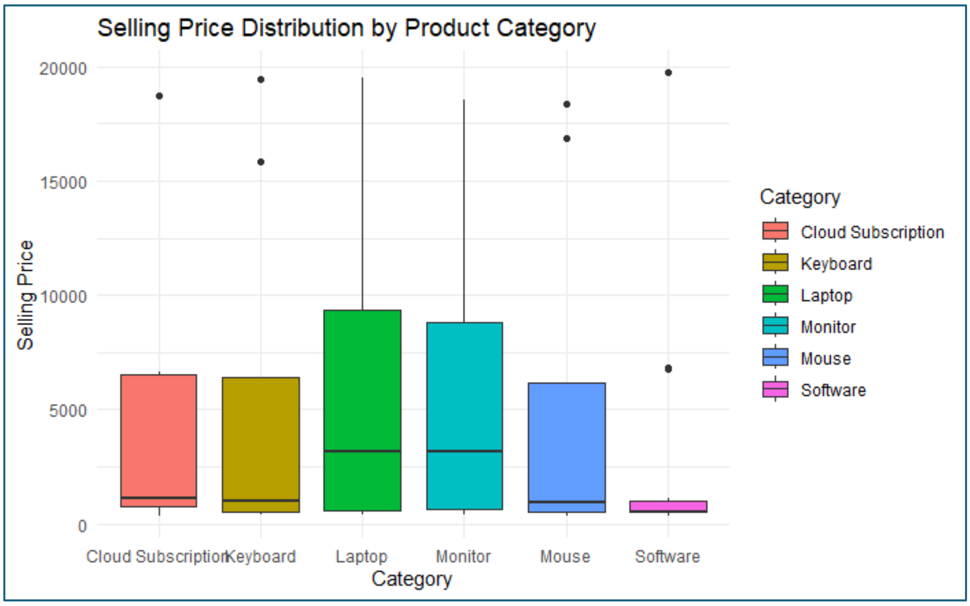

Figure 23 - corrected selling price distribution for operational sales data

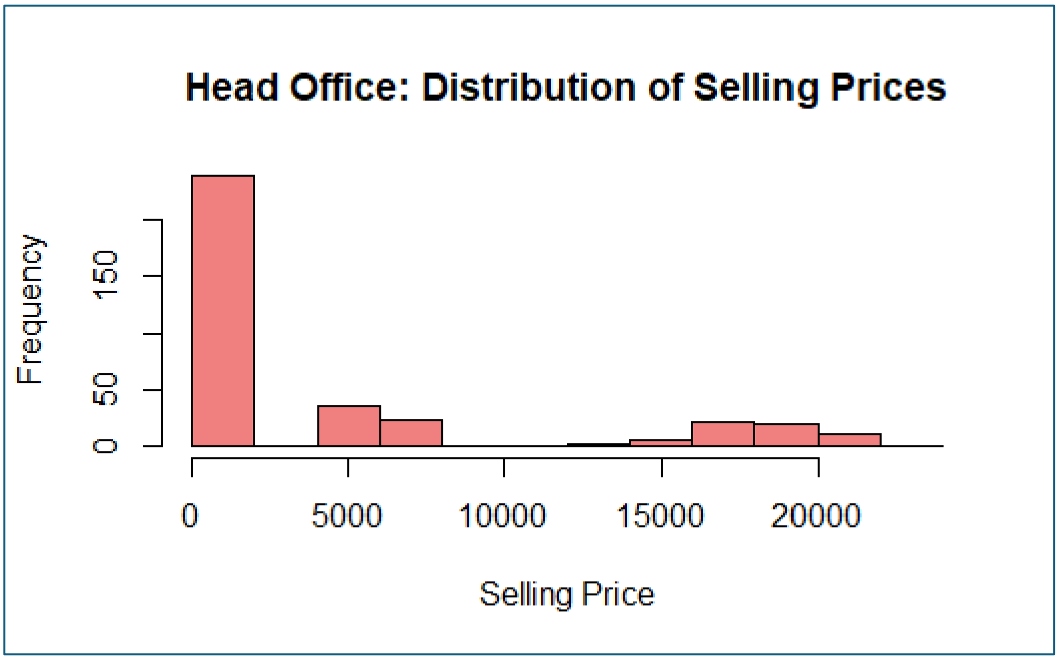

Figure 25 - original, incorrect selling price distribution for head office data

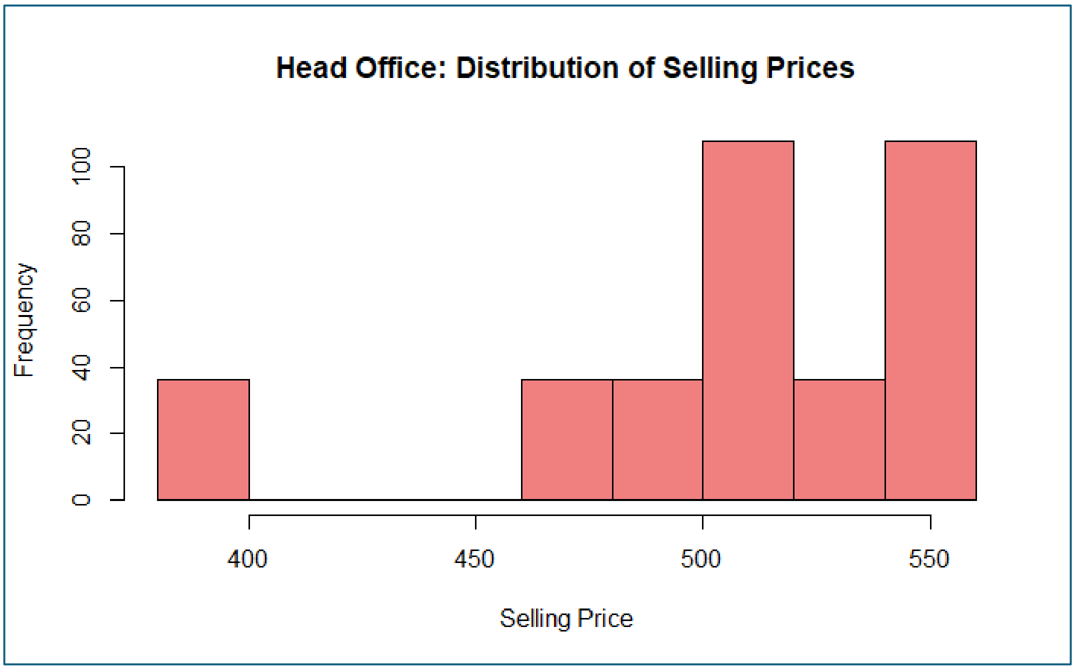

Figure 26 - corrected selling price distribution for head office sales data

5. Coffee Shop Optimisation

Coffee Shop 1

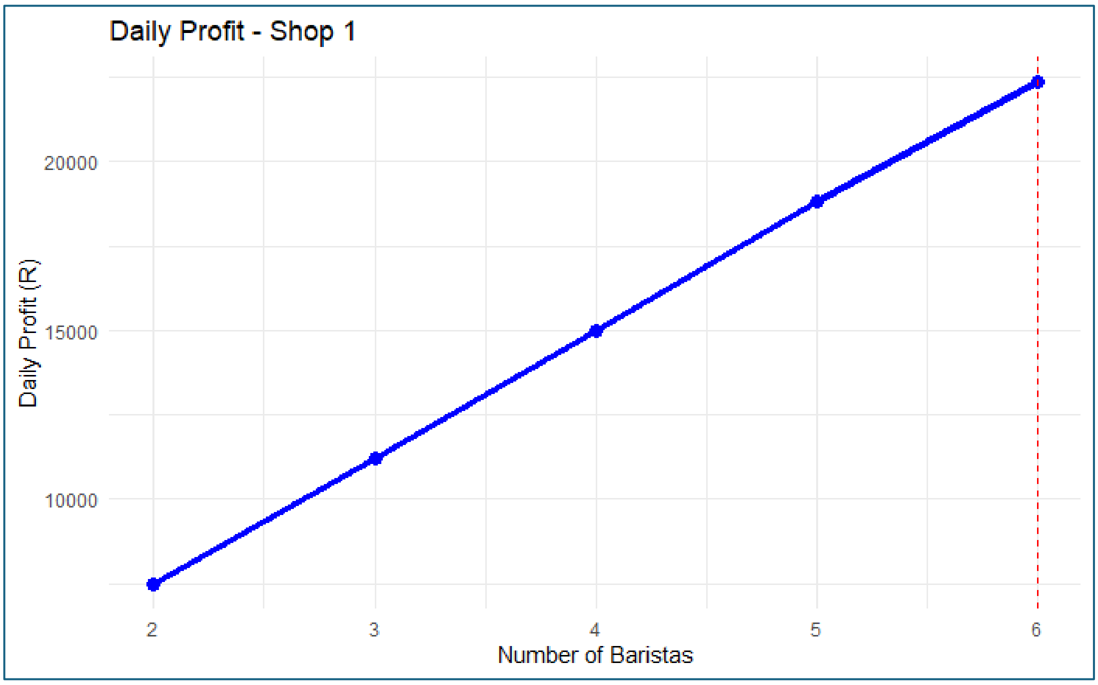

Figure 27 - Profit vs baristas for shop 1

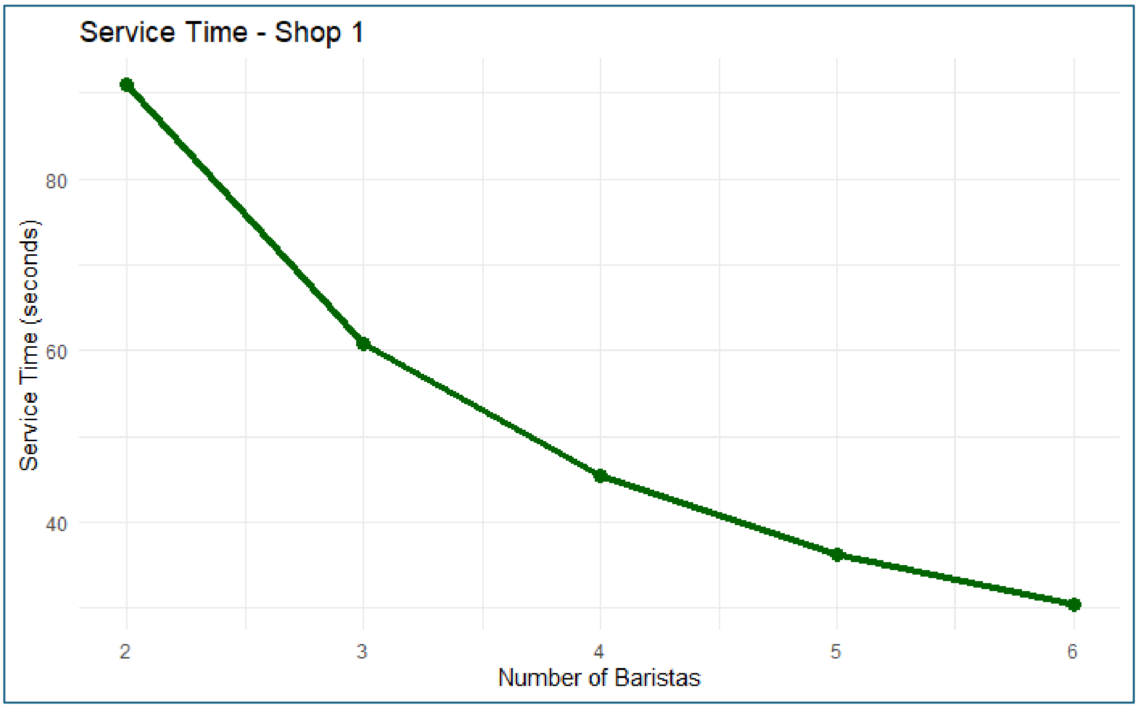

Figure 28 - Service time vs baristas for shop 1

Baristas	Avg_Service_Time	Reliability	Customers_Per_Day	Revenue	Cost	Profit	Profit_Margin
2	100.2	100	546	16388	2000	14388	87.8
3	66.6	100	1171	35118	3000	32118	91.5
4	50.0	100	1976	59282	4000	55282	93.3
5	40.2	100	2919	87559	5000	82559	94.3
6	35.0	100	3825	114762	6000	108762	94.8

Table 3- *data for varying number of baristas for shop 1*

Based on the data analysis, **Shop 1 should staff 6 baristas** to achieve maximum daily profit of **R22,349.45**. This staffing level provides the optimal balance between service efficiency and personnel costs.

There is a **strong positive relationship** between staffing levels and service efficiency:

- **Service times decrease significantly** with more baristas:
  - 2 baristas: 90.98 seconds per customer
  - 6 baristas: 30.48 seconds per customer (66% improvement)
- **Customer throughput increases dramatically**:
  - 2 baristas: 317 customers per day
  - 6 baristas: 945 customers per day (198% increase)

Revenue vs. Cost Trade-off

- The R1,000 daily cost per additional barista is **more than offset** by increased customer throughput
- Each additional barista generates **approximately R4,000-5,000 in incremental daily profit**
- The **marginal returns remain positive** even at 6 baristas, suggesting this is the optimal level within the constraint

Coffee Shop 2

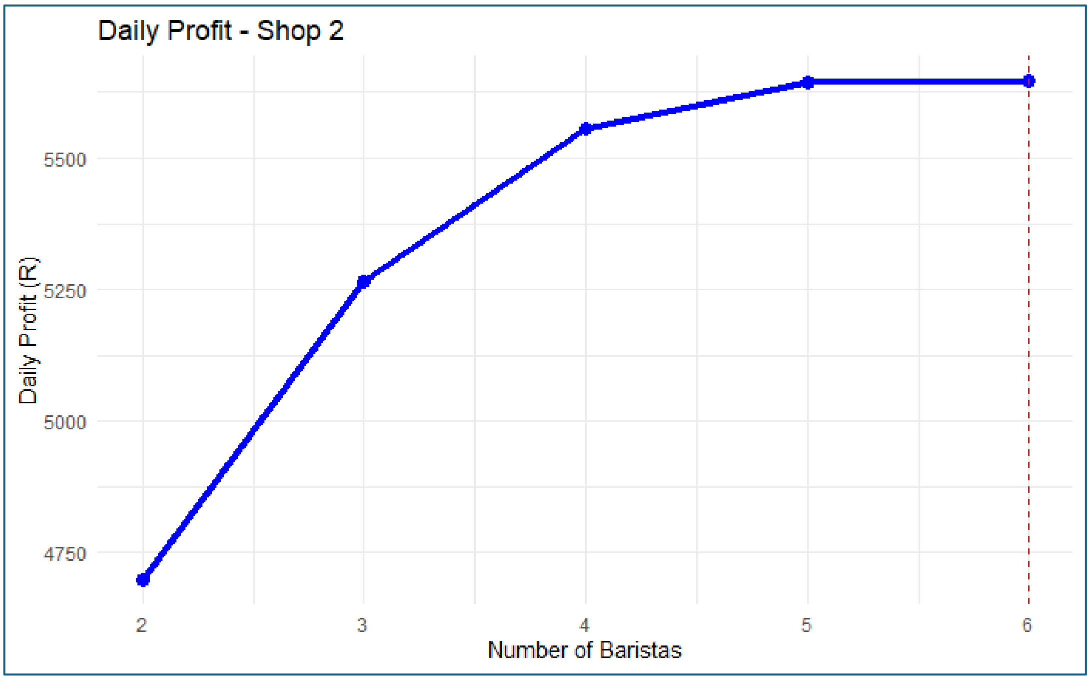

Figure 30 - Profit vs baristas for shop 2

Figure 29 - Service time vs baristas for shop 2

Baristas	Avg_Service_Time	Reliability	Customers_Per_Day	Revenue	Cost	Profit	Profit_Margin
2	141.5	100	387	11600	2000	9600	82.8
3	115.4	100	675	20264	3000	17264	85.2
4	100.0	100	988	29626	4000	25626	86.5
5	89.4	100	1311	39343	5000	34343	87.3
6	81.6	100	1638	49132	6000	43132	87.8

Table 4 - *Data for varying number of baristas for shop 2*

Based on the data analysis, **Shop 2 should staff 6 baristas** to achieve maximum daily profit of **R5,646.25**. However, the profit difference between 5 and 6 baristas is minimal, indicating diminishing returns at higher staffing levels.

Service improvements are **less dramatic** than Shop 1:

- **Service times decrease moderately** with more baristas:
  - 2 baristas: 128.98 seconds per customer
  - 6 baristas: 74.19 seconds per customer (42% improvement)
- **Customer throughput increases gradually**:
  - 2 baristas: 223 customers per day
  - 6 baristas: 388 customers per day (74% increase)

Revenue vs. Cost Trade-off

- The R1,000 daily cost per additional barista shows **diminishing returns**
- Profit gains become minimal after 4 baristas (+R92 from 4 to 5, +R2 from 5 to 6)
- **Marginal returns nearly disappear** at 5-6 baristas

6. DOE & Manova or Anova

Hypothesis

In our DOE and MANOVA analysis, we tested several critical hypotheses regarding delivery performance. We examined whether significant differences existed between Year 1 (2026) and Year 2 (2027) delivery times, whether the process capability groups derived from Part 3's Cpk analysis actually predicted real-world performance differences, and whether product types showed distinct delivery patterns. Additionally, we investigated potential interaction effects between years and capability groups, seasonal variations across quarters, and multivariate differences across multiple performance dimensions including mean delivery time, consistency, and reliability rates.

One-Way ANOVA Results

The one-way ANOVA analysis revealed fundamental insights into delivery performance drivers. Yearly comparisons showed no significant differences between 2026 and 2027 ($F = 1.511$, $p = 0.219$, $\eta^2 = 0$), indicating stable process execution over time. However, capability group differences were extremely significant ($F = 70{,}668$, $p < 0.001$, $\eta^2 = 0.844$), with the classification system from Part 3 explaining 84.4% of all delivery time variation. Similarly, product type differences were highly significant ($F = 11{,}775$, $p < 0.001$, $\eta^2 = 0.844$), confirming that different products have inherently distinct delivery characteristics. These findings validate that process capability is the dominant factor influencing delivery performance.

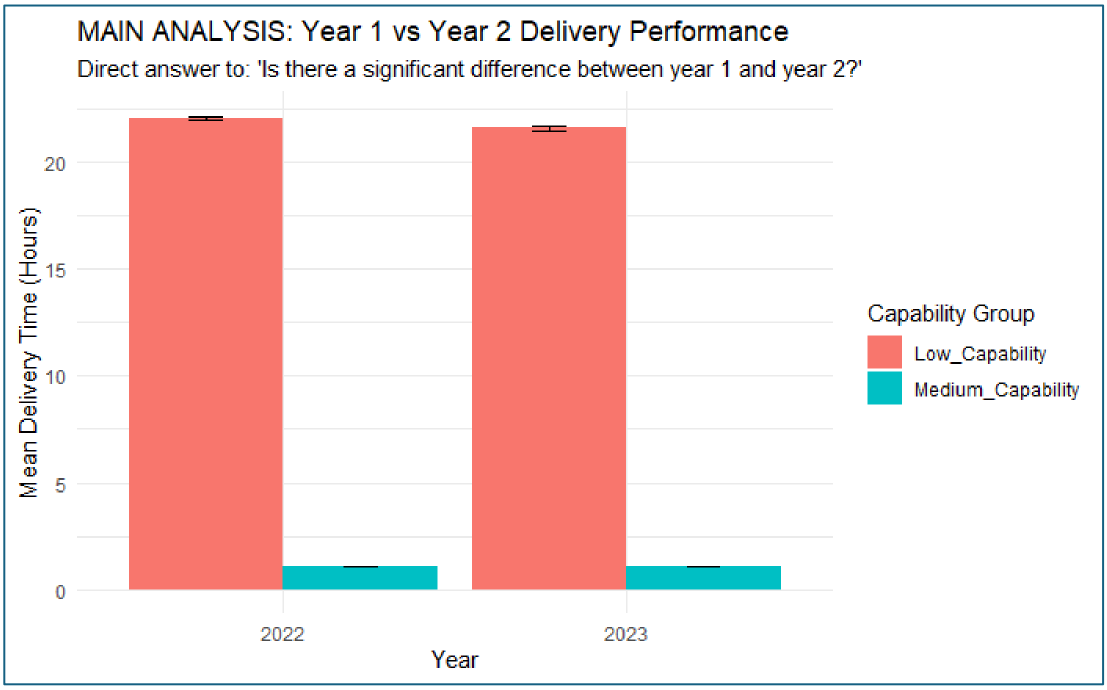

Figure 31 - Analysis of delivery time in Year 1 vs Year 2 by capability group

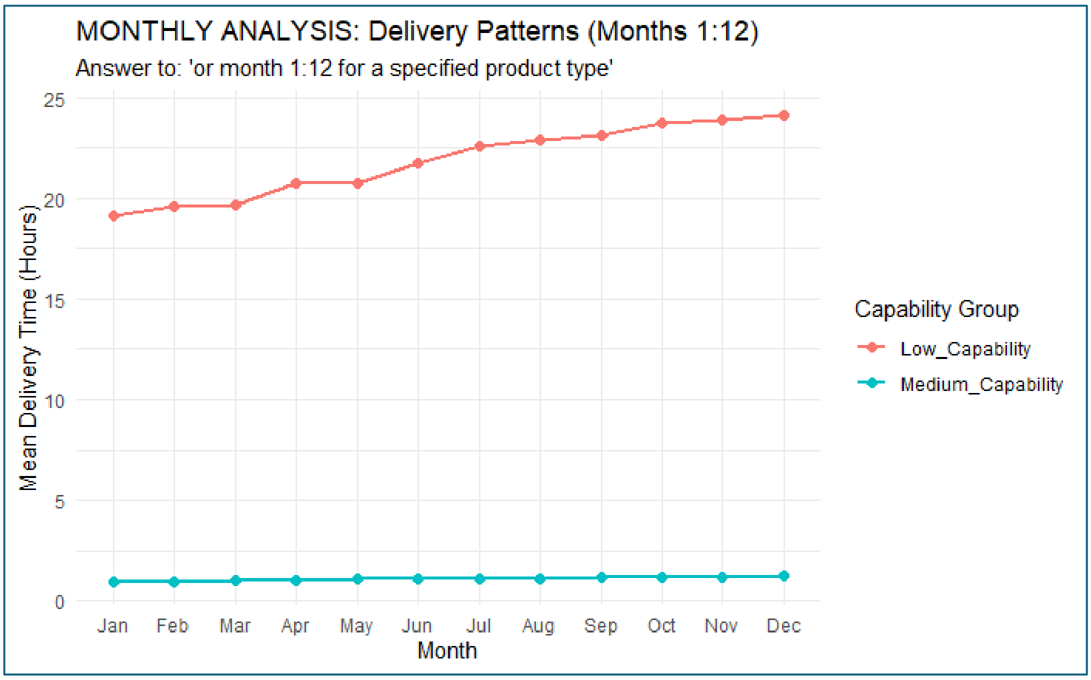

Figure 32 - Monthly analysis of delivery time by capability group

Multi-Factor ANOVA Results

The multi-factor analysis uncovered important interaction effects that simple tests missed. While year appeared insignificant alone, it became significant when combined with capability groups (F = 9.712, p = 0.00183), and the Year * Capability interaction (p = 0.00332) revealed that capability-performance relationships evolve annually. Strong seasonal patterns emerged (F = 184.637, p < 0.001), with different capability groups showing distinct quarterly responses. The three-way model (AIC = 75006) significantly outperformed simpler models, demonstrating that delivery performance is governed by complex but interpretable temporal and capability interactions.

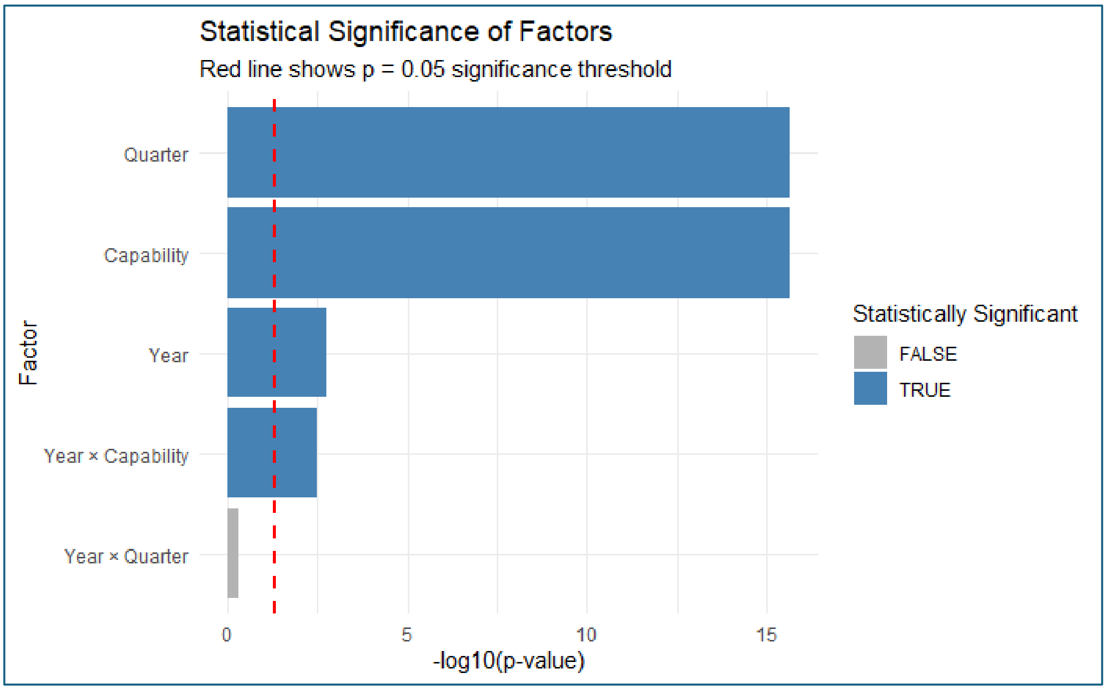

Figure 33 - Significance of difference factors and factor-combinations

MANOVA Results

The multivariate analysis confirmed that product types differ significantly across all performance dimensions simultaneously (Pillai = 1.264, $F = 5.099$, $p = 1.355e-08$). Univariate follow-up tests revealed product types have extremely strong effects on mean delivery time ($F = 454.20$, $p < 0.001$), delivery consistency ($F = 916.18$, $p < 0.001$), and on-time reliability ($F = 43.34$, $p < 0.001$). The absence of yearly ($p = 0.3962$) and interaction effects ($p = 0.9718$) indicates stable performance relationships, emphasizing that product-specific strategies should form the foundation of delivery optimization efforts.

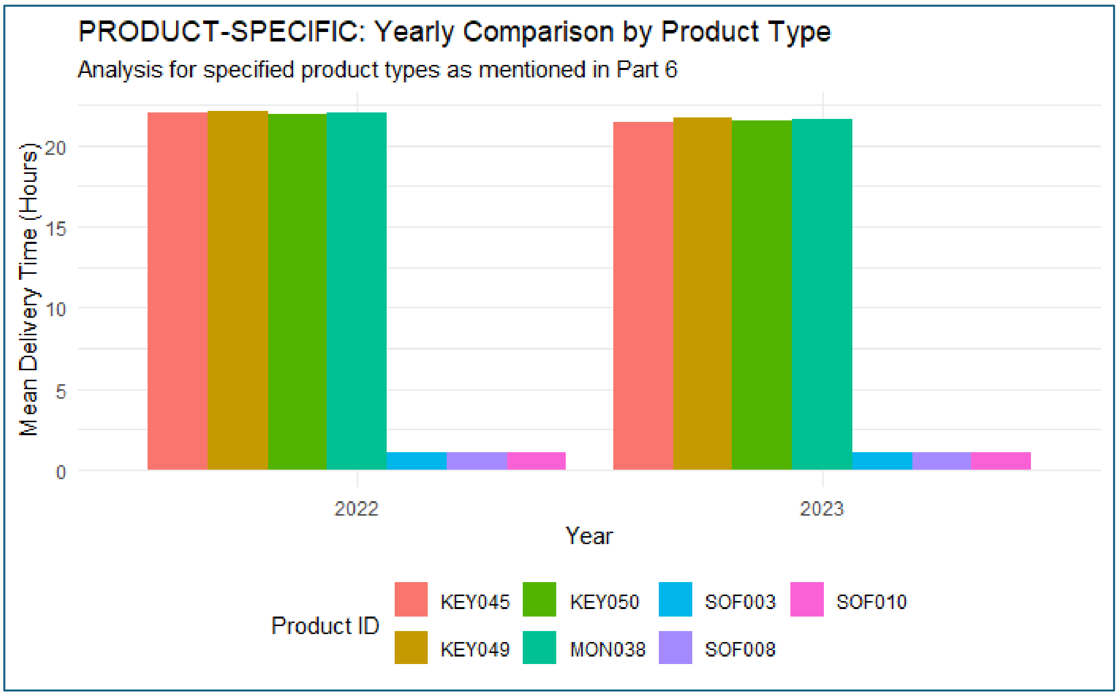

Figure 34 - Product specific yearly comparison for delivery time

Descriptive Performance Analysis

Capability_Group	Year	N	Mean	SD	Median	Q1	Q3
Low_Capability	2022	3660	22.042528	6.216371	22.0440	17.5440	26.5440
Low_Capability	2023	3121	21.592335	6.096163	21.5460	17.0460	26.0460
Medium_Capability	2022	3373	1.089044	0.299703	1.1022	0.8772	1.3022
Medium_Capability	2023	2884	1.098286	0.311624	1.1023	0.8773	1.3273

Table 5- Yearly performance analysis by capability group

Key Insights from Performance Data:

- 20x Performance Gap: Medium-capability processes deliver in ~1.1 hours vs. 22+ hours for low-capability processes
- Consistency Advantage: Medium-capability shows much lower variability (SD ≈ 0.3 hours vs. 6.1-6.2 hours)
- Stable Yearly Performance: Both capability groups maintained consistent delivery times across 2022-2023
- Slight Low-Capability Improvement: 0.45-hour reduction from 2022 to 2023 suggests gradual improvement

7. Reliability of Service

7.1 How many days can we expect reliable service

workers	days	cumulative days	cumulative_prob
12	1	1	0.002518892
13	5	6	0.015113350
14	25	31	0.078085642
15	96	127	0.319899244
16	270	397	1.000000000

Table 6 - *reliable service days for varying staff numbers*

Reliability Assessment

- Reliable Service Threshold: ≥15 workers present
- Reliable Days Observed: 366 days (96 + 270)
- Total Observation Period: 397 days
- Empirical Reliability Rate: 366/397 = 92.2%

Annual Projections

- Expected Reliable Days per Year: 337 days (92.2% × 365)
- Expected Problematic Days per Year: 28 days
- Service Availability: The agency maintains adequate staffing (≥15 workers) on 92% of days

7.2 Profit Optimisation

staff_level	reliability	problem_days	problem_cost	personnel_cost	total_cost
12	0.0000000	365.000000	7300000.00	3600000	10900000
13	0.0000000	365.000000	7300000.00	3900000	11200000
14	0.0000000	365.000000	7300000.00	4200000	11500000
15	0.3037490	254.131609	5082632.19	4500000	9582632
16	0.6516800	127.136798	2542735.96	4800000	7342736
17	0.8642344	49.554447	991088.94	5100000	6091089
18	0.9562126	15.982419	319648.37	5400000	5719648
19	0.9878196	4.445839	88916.78	5700000	5788917
20	0.9969914	1.098131	21962.61	6000000	6021963

Table 7 - *Profit optimisation for varying staff numbers*

Optimal Solution

- **Recommended Staffing**: 18 workers
- **Expected Reliability**: 95.6%
- **Problem Days Reduction**: From 127 to only 16 days annually
- **Total Annual Cost**: R5,719,648

Financial Impact

- **Current Situation (16 workers)**: R7,342,736 total cost, 65.2% reliability
- **Optimized Solution (18 workers)**: R5,719,648 total cost, 95.6% reliability
- **Annual Savings**: R1,623,088 (22.1% reduction)
- **Reliability Improvement**: +30.4 percentage points

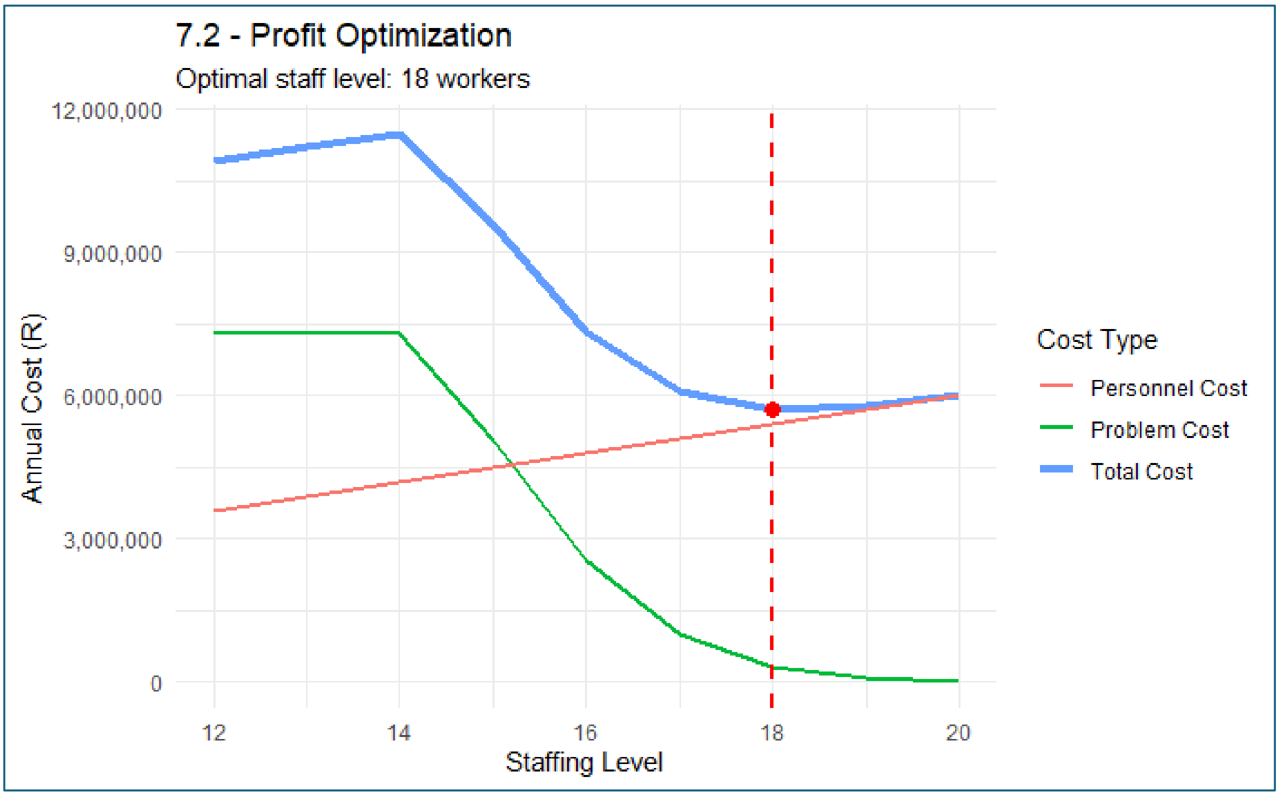

Figure 35 - Annual cost vs Staffing level

References

Reference list

Buderer, N.M. & Brannan, G.D. 2024. *Comparing the Means of Independent Groups: ANOVA, ANCOVA, MANOVA, and MANCOVA*. StatPearls Publishing. Available: https://www.ncbi.nlm.nih.gov/books/NBK606084/.

Jordan, L. 2025. *Control Charts: How to Master Statistical Process Control (SPC)*. Available: https://leanoutsidethebox.com/control-charts/.

Mcleod, S. 2023. *What are Type I and Type II Errors?* Simply Psychology. Available: https://www.simplypsychology.org/type_I_and_type_II_errors.html.

DeepSeek. (2024). *DeepSeek AI assistant* [Large language model]. Retrieved from https://www.deepseek.com

OpenAI. (2025). *ChatGPT (GPT-5)* [Large language model]. Retrieved from https://chat.openai.com (Accessed: October 23, 2025)

Appendices

Appendix A

ProductType	Cp	Cpu	Cpl	Cpk	Mean	SD	Capable
MOU059	0.844	0.570	1.118	0.570	21.189	6.320	FALSE
KEY049	0.845	0.529	1.161	0.529	21.985	6.314	FALSE
SOF009	17.486	33.785	1.187	1.187	1.086	0.305	FALSE
CLO019	0.869	0.568	1.170	0.568	21.539	6.134	FALSE
KEY045	0.847	0.538	1.156	0.538	21.837	6.297	FALSE
SOF010	17.990	34.777	1.202	1.202	1.069	0.296	FALSE
KEY046	0.896	0.572	1.219	0.572	21.776	5.953	FALSE
CLO012	0.865	0.557	1.172	0.557	21.686	6.168	FALSE
KEY047	0.875	0.574	1.176	0.574	21.498	6.094	FALSE
CLO020	0.895	0.621	1.169	0.621	20.895	5.958	FALSE
KEY043	0.880	0.567	1.192	0.567	21.688	6.064	FALSE
MOU058	0.884	0.587	1.182	0.587	21.385	6.031	FALSE
KEY042	0.866	0.566	1.166	0.566	21.539	6.158	FALSE
SOF007	17.516	33.843	1.189	1.189	1.086	0.304	FALSE
CLO011	0.850	0.570	1.130	0.570	21.272	6.274	FALSE
LAP030	0.869	0.553	1.185	0.553	21.816	6.139	FALSE
SOF001	17.201	33.253	1.150	1.150	1.069	0.310	FALSE
SOF002	17.303	33.455	1.151	1.151	1.065	0.308	FALSE

ProductType	Cp	Cpu	Cpl	Cpk	Mean	SD	Capable
MOU051	0.887	0.568	1.205	0.568	21.744	6.016	FALSE
LAP028	0.856	0.544	1.168	0.544	21.839	6.230	FALSE
SOF005	17.295	33.425	1.166	1.166	1.078	0.308	FALSE
MON037	0.901	0.593	1.209	0.593	21.466	5.920	FALSE
MOU057	0.884	0.585	1.183	0.585	21.404	6.033	FALSE
CLO017	0.878	0.580	1.176	0.580	21.427	6.073	FALSE
KEY048	0.889	0.559	1.218	0.559	21.928	6.002	FALSE
MON032	0.891	0.604	1.177	0.604	21.144	5.988	FALSE
MOU060	0.873	0.560	1.185	0.560	21.734	6.113	FALSE
MON031	0.887	0.573	1.201	0.573	21.663	6.010	FALSE
MON033	0.843	0.566	1.120	0.566	21.252	6.325	FALSE
MOU054	0.857	0.567	1.148	0.567	21.422	6.220	FALSE
LAP023	0.913	0.584	1.241	0.584	21.764	5.844	FALSE
KEY044	0.880	0.575	1.185	0.575	21.542	6.058	FALSE
MON035	0.875	0.575	1.176	0.575	21.496	6.093	FALSE
CLO016	0.856	0.560	1.152	0.560	21.530	6.229	FALSE
MOU052	0.899	0.575	1.222	0.575	21.755	5.935	FALSE
SOF003	18.050	34.893	1.206	1.206	1.069	0.295	FALSE
LAP022	0.917	0.590	1.245	0.590	21.708	5.814	FALSE
LAP025	0.879	0.563	1.195	0.563	21.756	6.068	FALSE

ProductType	Cp	Cpu	Cpl	Cpk	Mean	SD	Capable
LAP029	0.883	0.569	1.197	0.569	21.684	6.041	FALSE
MOU055	0.888	0.588	1.187	0.588	21.402	6.008	FALSE
SOF004	17.527	33.882	1.172	1.172	1.070	0.304	FALSE
MOU056	0.872	0.560	1.184	0.560	21.725	6.115	FALSE
CLO018	0.846	0.573	1.120	0.573	21.177	6.301	FALSE
SOF006	17.666	34.162	1.170	1.170	1.059	0.302	FALSE
MON040	0.862	0.575	1.149	0.575	21.326	6.188	FALSE
MON036	0.886	0.580	1.191	0.580	21.515	6.023	FALSE
KEY050	0.850	0.539	1.162	0.539	21.861	6.273	FALSE
MON034	0.869	0.582	1.156	0.582	21.277	6.138	FALSE
MON039	0.878	0.599	1.157	0.599	21.079	6.075	FALSE
CLO015	0.886	0.579	1.193	0.579	21.537	6.019	FALSE
MON038	0.875	0.574	1.176	0.574	21.503	6.095	FALSE
CLO014	0.878	0.585	1.170	0.585	21.335	6.077	FALSE
LAP024	0.878	0.557	1.200	0.557	21.853	6.071	FALSE
SOF008	18.237	35.247	1.226	1.226	1.076	0.292	FALSE
LAP021	0.868	0.577	1.158	0.577	21.356	6.146	FALSE
MOU053	0.865	0.547	1.182	0.547	21.875	6.169	FALSE
CLO013	0.859	0.565	1.152	0.565	21.467	6.211	FALSE
LAP027	0.887	0.573	1.202	0.573	21.674	6.012	FALSE

ProductType	Cp	Cpu	Cpl	Cpk	Mean	SD	Capable
KEY041	0.880	0.579	1.181	0.579	21.471	6.060	FALSE
LAP026	0.881	0.574	1.188	0.574	21.575	6.054	FALSE