ECSA GA4 Report: Quality Assurance 344

Author: Seth Ravat

Student Number: 26988097

Date: 24/10/2025

Contents

Introduction

In a competitive global marketplace, companies are increasingly reliant on data-driven decision-making to maintain and improve their performance. This report is part of a structured quality assurance and analytics project, which involves the consolidation and evaluation of several datasets. The purpose is not only to understand how customers interact with products but also to identify bottlenecks in operations and highlight opportunities for strategic improvement.

The datasets provided for analysis include:

- Customer Data: demographic and income information of customers.
- Product Data: details of products available in the company's catalog.
- Head Office Product Data: additional reference information on product specifications.
- Sales Data (2022–2023): transactional records of orders placed, including time stamps and operational durations.

By integrating these data sources, it becomes possible to create a holistic view of the company's performance across multiple dimensions. This report, while intermediate in nature, lays the groundwork for more advanced analyses that will follow in subsequent submissions.

1. Data Preparation and Methodology

The datasets were imported into R for cleaning and integration. Care was taken to standardize variable formats—for example, date variables were transformed into a consistent structure, and categorical variables such as product codes were aligned across datasets. Missing values were inspected, and although minor gaps existed, the overall datasets were sufficiently robust for exploratory analysis. To derive actionable insights, a series of visualizations were created using the ggplot2 package. These included:

- Monthly sales quantities, allowing comparison of seasonal trends across two years.
- Customer income distributions, providing insight into the financial profile of the customer base.
- Scatterplots of picking vs. delivery hours, highlighting relationships and inefficiencies in the supply chain.
- Top 10 product sales rankings, showcasing concentration of sales within key categories.

2. Analysis and Results

2.1 Sales Trends by Month

The sales data across 2022 and 2023 reveals a clear downward trajectory. During 2022, the company enjoyed relatively stable growth, with monthly quantities peaking in the second quarter of the year. This suggests that the business had either favourable market conditions or a well-aligned product offering during that period.

In contrast, 2023 shows significantly reduced sales volumes in almost every month. The reduction is not trivial - it represents a substantial decline in the company's market performance. The seasonal cycle appears somewhat consistent, with a rise in mid-year sales, but the magnitude is much lower compared to the previous year.

This finding is critical because it directly impacts revenue forecasts and customer retention analysis. It also raises questions about whether the company is losing market share to competitors, suffering from reduced product availability, or failing to attract repeat customers.

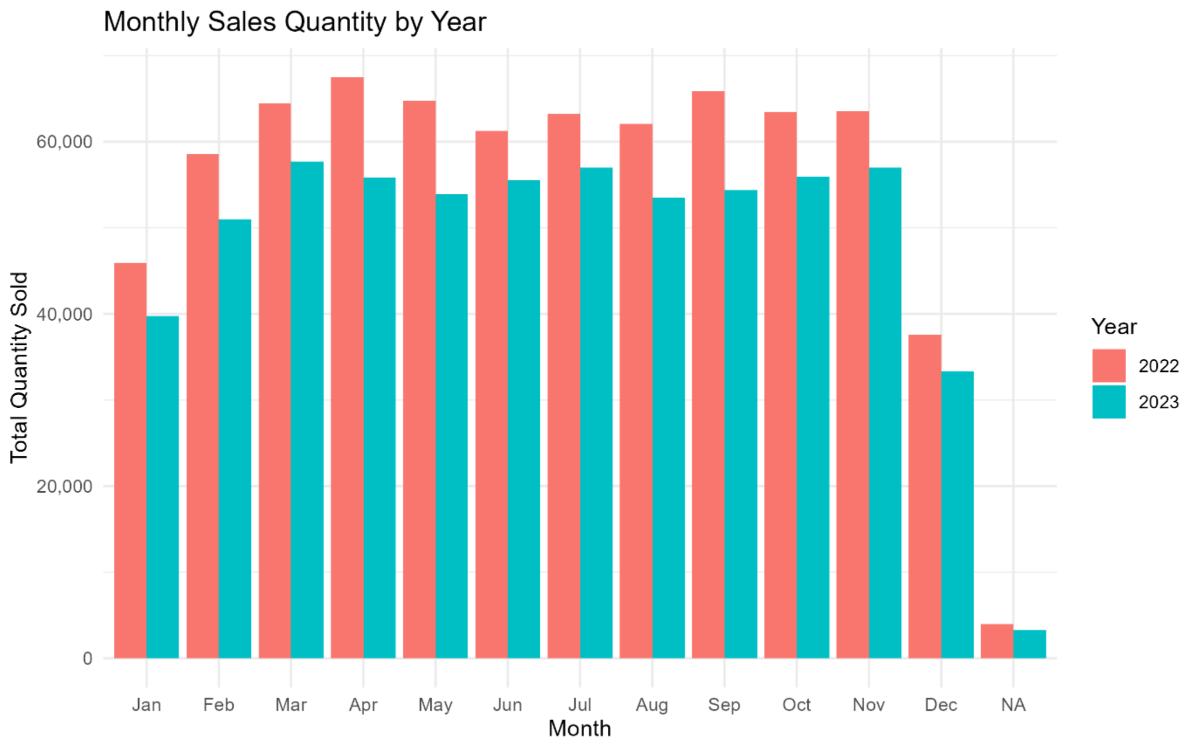

Figure 1: Monthly Sales Quantity by Year

2. 2 Customer Demographics and Segmentation

Income Distribution

An important aspect of customer profiling is understanding the income levels of the customer base. The analysis reveals a wide spread of incomes; however, the largest cluster falls between 50,000 and 120,000. This suggests the company appeals most strongly to middle-income customers, who likely view its products as affordable but not low-end.

The implications of this are significant for marketing strategy. If the company wishes to expand upward, it must position its premium products more aggressively toward higher income customers. Conversely, if it wishes to safeguard its core base, it must continue to provide consistent value for the middle-income bracket.

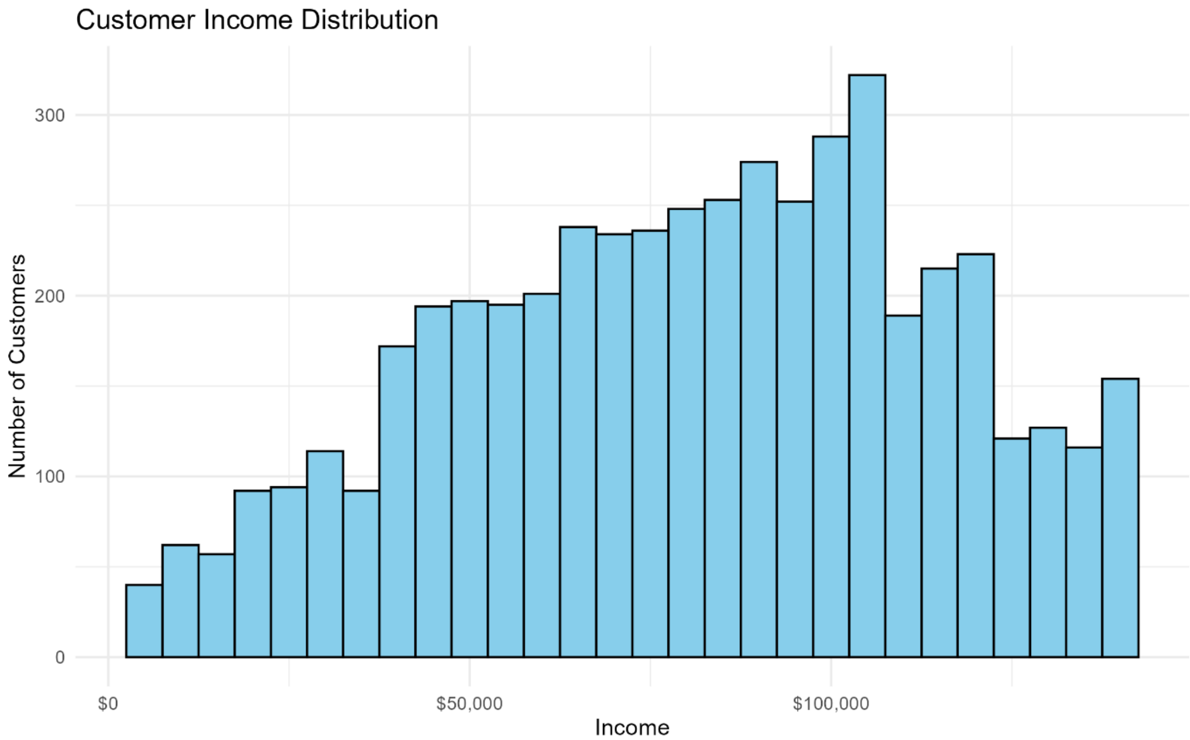

Figure 2: Customer Income Distribution

Age Segmentation

Sales quantities by age reveal that customers aged 65 years and above are the strongest contributors to product demand. This is somewhat surprising, as many firms target younger consumers, assuming they represent the bulk of spending power. The company's strength with senior customers could be leveraged further, for example, through loyalty programs or specialized product offerings tailored to this demographic.

Meanwhile, other age groups show relatively balanced but lower levels of engagement. This presents both a challenge and an opportunity: younger demographics are underperforming, but with targeted campaigns, they could represent a future growth market.

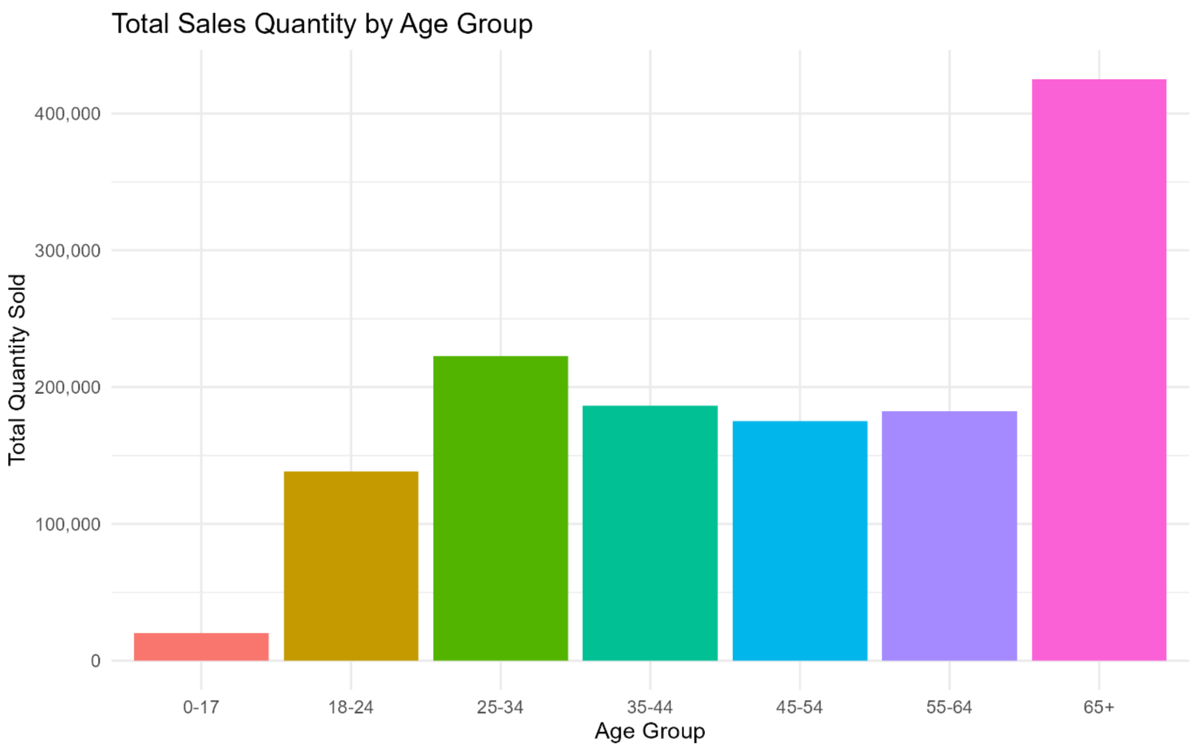

Figure 3: Top Sales Quantity by Age Group

2. 3 Operational Metrics: Picking vs. Delivery Times

Logistics and operations play a pivotal role in ensuring customer satisfaction. The scatterplot of picking hours vs. delivery hours revealed three main clusters:

- Low-hour cluster: Very fast picking and delivery times, likely corresponding to system tests or unusually efficient transactions.
- Moderate-hour cluster: Orders with picking times of 10–30 hours and variable delivery times, indicating moderate complexity.
- High-hour cluster: Picking times of 30–45 hours, reflecting products or processes that are more time-consuming.

The lack of a clear linear relationship between picking and delivery times is notable. One might expect that orders requiring longer picking would also take longer to deliver, but the data shows otherwise. This inconsistency hints at potential inefficiencies or lack of coordination between warehouse and logistics teams. For example, an order that takes long to pick may still be delivered quickly if transport is available, while short-pick orders may wait disproportionately longer for delivery.

Figure 4: Picking vs. Delivery Hours

2.4 Product Performance

An analysis of product sales shows that the top ten best-selling products are heavily dominated by the MOU and SOF categories. The most frequently purchased items include:

MOU059, SOF001, SOF004, SOF010, MOU058, MOU054, MOU052, SOF007, MOU057, and SOF005.

This concentration indicates that the company's sales revenue is heavily dependent on a relatively small subset of its catalogue. Such dependence is risky, as fluctuations in supply or demand for these items could disproportionately impact overall performance.

On the other hand, the strong demand for these products presents an opportunity for focused marketing campaigns and inventory optimization. Ensuring these items are always in stock and promoted effectively could help maximize revenue.

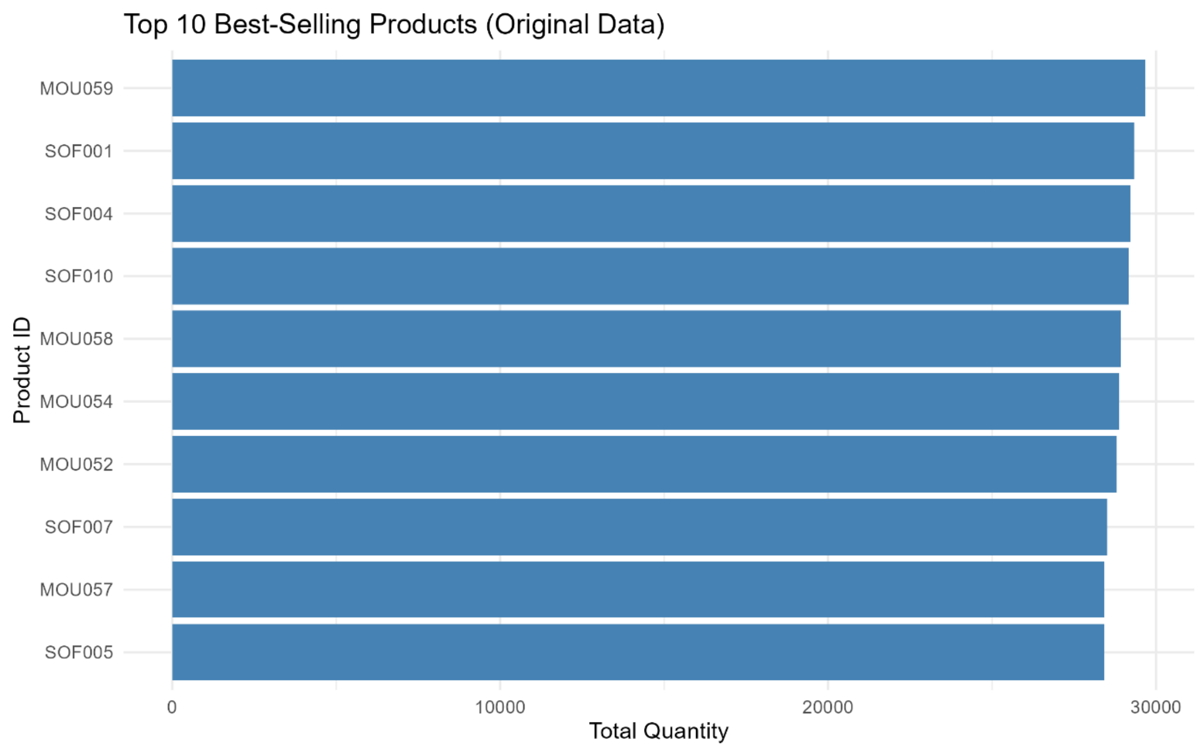

Figure 5: Top 10 Best-Selling Products

2.5 Summary of Findings

Overall, the analysis reveals several key insights into the company's current performance and strategic direction. Sales have declined from 2022 to 2023, signalling potential long-term challenges that require immediate investigation through trend and competitor analysis. Customer demographics show a strong reliance on middle-income and senior consumers, providing short-term stability but highlighting the need to attract younger customers for sustainable growth.

Operationally, inconsistencies between picking and delivery times indicate inefficiencies that may negatively impact customer satisfaction, suggesting a need for process optimization and better coordination across departments. In terms of product performance, sales are heavily concentrated in a few key categories (MOU and SOF), creating both opportunities for short-term gains and risks associated with overdependence.

Moving forward, the company should focus on understanding the causes of sales decline, modernizing its marketing approach to appeal to younger demographics, and improving logistics reliability. Additionally, reassessing the product mix will be essential to balance diversification with strategic focus. These steps will lay the groundwork for deeper analytical work, including profitability assessments and predictive modelling, to support data-driven decision-making.

3. SPC Analysis

3.1 Results

The results obtained from the sales and operations dataset reveal several critical insights about the business's performance, operational consistency, and product capability. The data from 2022–2023 were cleaned and processed in R to generate three key distribution charts (Sales Quantity, Picking Hours, and Delivery Hours), along with product-wise capability analysis results. These results are discussed below.

3.1.1 Distribution of Sales Quantity

The histogram of sales quantity (Figure 1) illustrates a highly right-skewed distribution. Most sales orders are concentrated between 0–20 units, with a sharp decline thereafter. Only a few transactions exceed 40 units. This pattern suggests that the majority of sales are driven by small to medium-sized orders, consistent with retail-style purchasing behaviour rather than bulk or wholesale transactions.

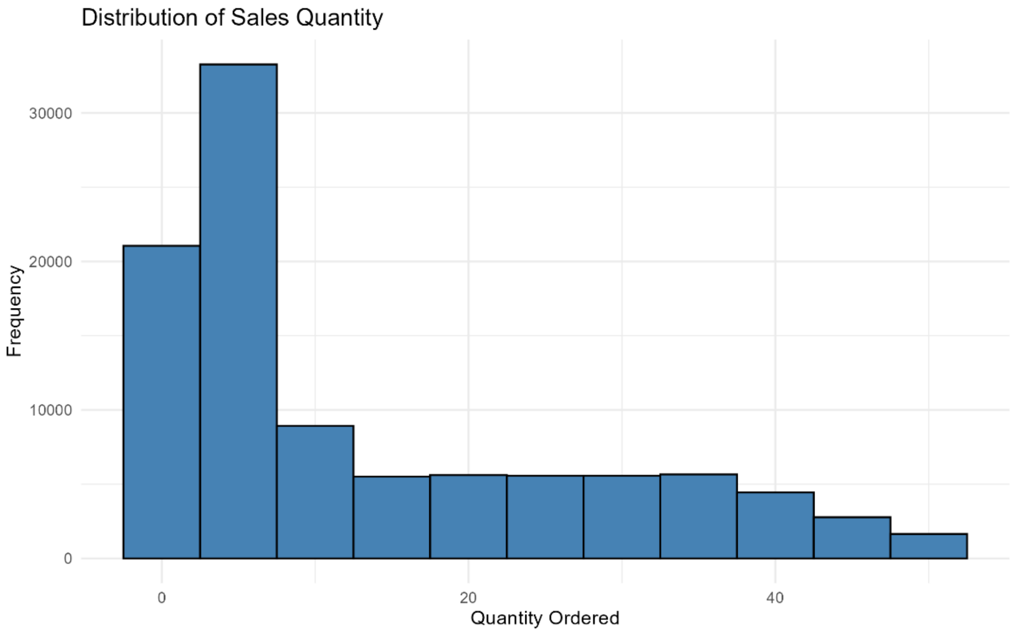

Figure 6: Distribution of Sales Quantity (2022–2023)

3.1.2 Distribution of Picking Hours

The picking hours distribution (Figure 2) displays a multimodal shape—one peak near 0 hours and another between 10–15 hours, with a smaller cluster around 30–40 hours. This suggests distinct operational behaviours (fast/easy orders, standard orders, and complex/slow orders).

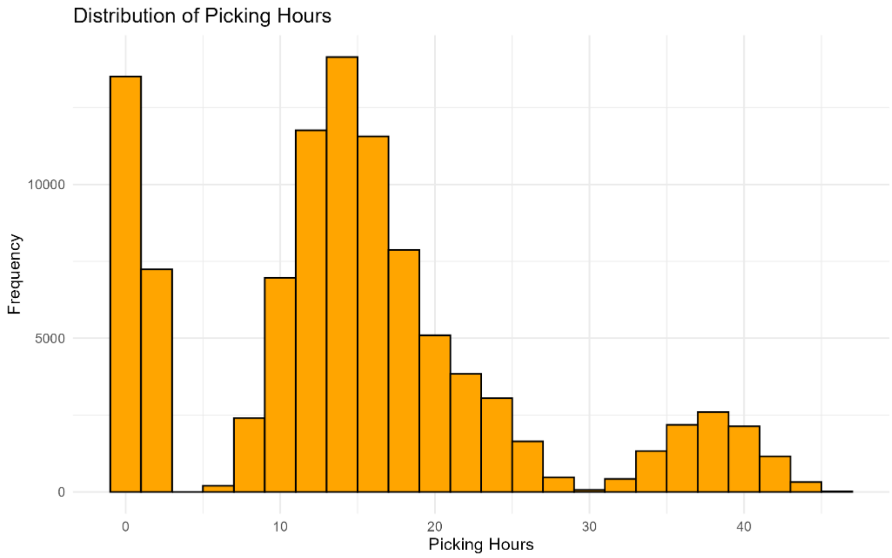

Figure 7: Distribution of Picking Hours.

3.1.3 Distribution of Delivery Hours

The distribution of delivery hours (Figure 3) appears approximately bell-shaped but with a noticeable spike near 0 hours. Most deliveries take between 10–30 hours, yet there is a significant number of very fast deliveries (local or express). The spread suggests room for improving delivery consistency.

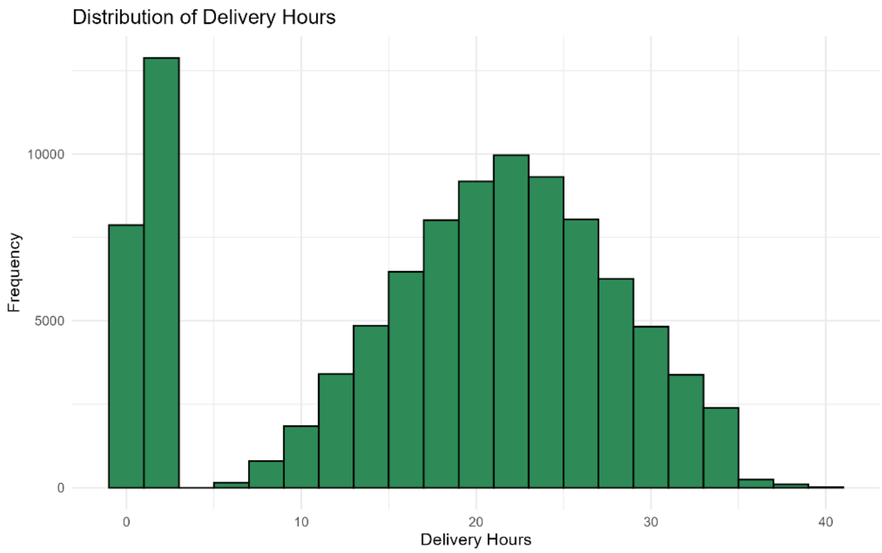

Figure 8: Distribution of Delivery Hours

3.1.4 Capability Analysis

Cp (Potential Capability): This index measures the potential capability assuming the process is perfectly cantered. A Cp ≥ 1.33 is generally considered good. Only the 'SOF' product type shows excellent potential capability (Cp = 18.14). All other product types (MOU, KEY, CLO, LAP, MON) have Cp values less than 1, indicating their inherent process variation (spread, σ) is too wide relative to the specification range (USL- LSL).

Cpk (Actual Capability): This index accounts for both process variation and cantering relative to the specification limits. It represents the actual performance. A Cpk ≥ 1.33 indicates a capable process. Only the 'SOF' product type meets this criterion (Cpk = 1.08), although it's lower than its Cp, suggesting it might not be perfectly cantered despite low variation. All other product types have Cpk values significantly below 1 (around 0.70-0.73), confirming they are not capable of consistently meeting the 0-32 hour delivery window. Their low Cpk is primarily driven by their large standard deviation (σ ≈ 5.8 – 6.0 hours).

Product	μ	σ	C_p	C_{pk}	C_{pl}	C_{pu}
MOU	19.30	5.83	0.92	0.73	1.10	0.73
KEY	19.28	5.82	0.92	0.73	1.10	0.73
SOF	0.96	0.29	18.14	1.08	1.08	35.19
CLO	19.23	5.94	0.90	0.72	1.08	0.72
LAP	19.61	5.93	0.90	0.70	1.10	0.70
MON	19.41	6.00	0.89	0.70	1.08	0.70

Table 1: Capability Summary by Product Type (Overleaf)

4. Risk, Data Correction, and Optimising for Maximum Profit

For a normally distributed process, approximately 99.73% of data points fall within ±3σ of the mean. The remaining 0.27% of observations (0.135% above +3σ and 0.135% below −3σ) lie outside the control limits, and thus the probability of one point exceeding the control limits is:

P(point outside control limits) = 2 × 0.00135 = 0.0027

This represents the baseline Type I error probability when using a traditional 3σ control chart rule.

4.1 Estimating the Likelihood of Making a Type I (Manufacturer's) Error

Rule	Condition	Estimated Type I Error (α)
A	One point beyond the UCL or LCL	0.0027
B	Two out of three consecutive points beyond $\pm 2\sigma$	≈ 0.005
C	Four consecutive points above or below the centreline	$0.5^4 = 0.0625$

Table 2: Type 1 Errors, Extrapolated with Overleaf

Rule A has a very low false alarm rate (α = 0.0027), which means it is a highly conservative rule. A point beyond the control limits is unlikely unless the process truly deviates.

Rule B is more sensitive and has a slightly higher false alarm rate (α ≈ 0.005), allowing for earlier detection of smaller process shifts.

Rule C is the most sensitive rule among the three. The probability of obtaining four consecutive points on the same side of the mean by chance alone is 0.5 4 = 0.0625. This means about 6.25% of stable processes could trigger a false alarm under this rule.

4.2 Type II Error Calculation

The Type II error (β) represents the probability that a process shift occurs, but the control chart fails to detect it. In this case, the process is designed to be centered on CL = 25.05 L with control limits UCL = 25.089 L and LCL = 25.011 L. Unknown to the analyst, the true process mean has shifted to μ = 25.028 L and the process standard deviation has increased to σ = 0.017 L. The standard deviation of the sampling distribution of the mean (X¯) for a subgroup of size n is:

σ_X¯ = σ /√ n

The Type II error probability is computed as:

$$\beta = P(LCL < \bar{X} < UCL \mid \mu = \mu_A) = \Phi\left(\frac{UCL - \mu_A}{\sigma_{\bar{X}}}\right) - \Phi\left(\frac{LCL - \mu_A}{\sigma_{\bar{X}}}\right)$$

Computed values:

- Subgroup size n = 24:
    - β = 1.0
    - σ_X¯ = 0.00347011
    - z_low = −4.89898, z_up = 17.5787
- Subgroup size n = 5:
    - β = 0.987326
    - σ_X¯ = 0.00760263
    - z_low = −2.23607, z_up = 8.02354

The standard X-bar chart has very low power (Power = 1 β) to detect this small shift in the mean, especially since the process variation also increased. We are very likely to make a Type II error in this situation. Better detection might require different rules, larger sample sizes, or tighter limits.

4.3 Data Correction and Re-Analysis (2025 Data)

As requested by Head Office via email, the Week 1 product data files contained errors that needed correcting. The R script performed the following fixes:

1. Updated the Category column in products_data.csv based on the ProductID prefix.
2. Corrected invalid Category codes (e.g., "NA") in products_Headoffice.csv.
3. Ensured ProductIDs were sequential within each category in products_Headoffice.csv
4. Applied the correct 10-item repeating pattern for SellingPrice and Markup in products_Headoffice.csv, using products_data.csv as the source for the correct 10 values per category.

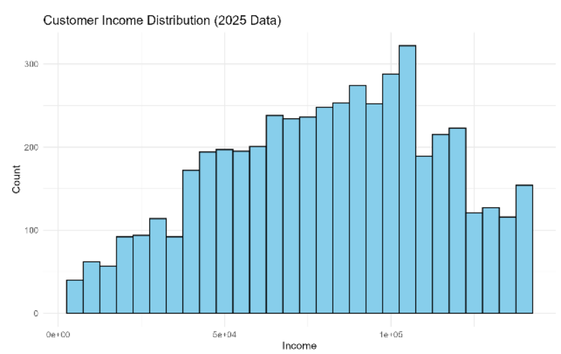

Figure 10: Customer Income Distribution - 2025 Data

Figure 9: Monthly Sales Quantity by Year - 2025 Data

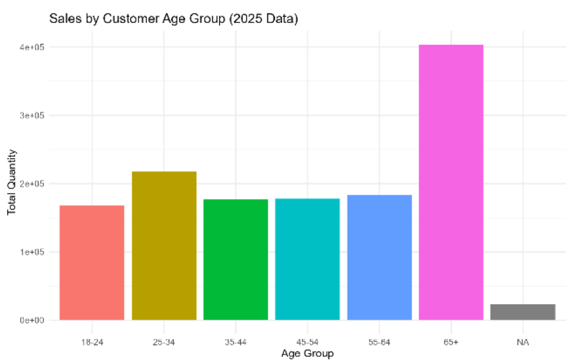

Figure 11: Sales by Customer Age Group - 2025 Data

Figure 12: Top 10 Best-Selling Products - 2025 Data

Comparing the "2025 Data" plots (Figures 9-12) with the original Part 1 plots (Figures 1-5) shows:

No Significant Change: Monthly sales trends (Fig 8 vs 1), customer income (Fig 9 vs 2), top products by quantity (Fig 10 vs 5), and sales by age group (Fig 11 vs 3) appear visually almost identical. This suggests the high-level operational and demographic patterns were not significantly affected by the specific data errors related to product categorization and pricing details for items beyond the first 10 per type.

Significant Change: The Total Sales Value by Product Type (Fig 12 vs 6) is different. While Laptops (LAP) and Monitors (MON) still contribute the most revenue, the absolute values and the relative contributions of the other categories (CLO, KEY, SOF, MOU) have changed now that the correct SellingPrice has been applied consistently to all products based on the 10-item repeating pattern.

While general trends might seem fairly strong, financial metrics like revenue calculations are highly sensitive to errors in pricing data. The corrected 2025 analysis provides a more reliable baseline for financial evaluation.

5. Profit Optimisation

5.1 Optimising Delivery Process Mean

The SPC analysis (Part 3 & 4) provided corrected estimates for the mean (μ) and standard deviation (σ) of delivery times for each product type, after removing initial out-of-control points. The goal here is to determine if adjusting the target mean delivery time could increase profit, considering a simplified model.

Model Assumptions:

- Revenue per delivery = R100
- Defect Cost = R500 penalty if delivery time > USL (32 hours)
- Operational Cost = R1500 / μ (faster mean delivery is more expensive)

product	sigma_cor	current_m	current_pr	optimal_m	max_profit	profit_gain_per_unit
MOU	5.738249	19.24886	15.50469	19.9139	15.87974	0.375048
KEY	5.921356	19.194	14.20914	19.68305	14.41297	0.203832
SOF	0.300607	0.955638	-1469.63	30.95452	51.41547	1521.048
CLO	5.972272	19.12594	13.79451	19.62017	14.0023	0.207793
LAP	5.954848	19.52386	14.13092	19.64162	14.14297	0.012054
MON	5.987865	19.42594	13.84977	19.60102	13.8763	0.026527

Figure 13: Excel Output of Capability Summary

The R script employed the optimise function. It calculated the profit for different potential means, balancing the operational cost against the expected defect cost.

The optimization suggests that for most product types (MOU, KEY, CLO, LAP, MON), the current corrected mean is already quite close to the calculated optimal mean. The potential profit gain per unit by shifting the mean is relatively small, often less than R1 per delivery. This implies that reducing the variation in delivery times (as identified in the capability analysis) is likely a more impactful lever for improvement than simply shifting the average. The exception is the 'SOF' category, which has very low variation, and its optimization result might differ more significantly.

<u>Summary of Results:</u>

Products (MOU, KEY, CLO, LAP, MON): For these types, the calculated optimal mean delivery time ($\mu\_optimal$) was very close to the current corrected mean (μ_corr). For example, for MOU, $\mu\_corr \approx 19.25$ hours and $\mu\_optimal \approx 19.91$ hours. The potential profit gain per unit from shifting the mean was minimal (R15.88 – R15.50 = R0.38 for MOU).

Software (SOF): This category showed a significant difference. Its current mean is extremely low ($\mu\_corr \approx 0.96$ hours), resulting in very high calculated operational costs and negative profit. The optimization suggested a drastically higher optimal mean ($\mu\_optimal \approx 30.95$ hours) to substantially increase profit, primarily by reducing the high operational cost associated with the extremely fast (and likely unrealistic or data-error affected) current average.

5.2 Coffee Shop Barista Optimization

This task analyses service time data from two coffee shops (timeToServe.csv - Shop 1, timeToServe2.csv - Shop 2) to determine the optimal number of baristas (k, tested between 2 and 6) to maximize daily profit.

<u>Methodology/Formulae:</u>

- 'Customers Served' was limited by either the shop's calculated daily capacity (based on average service time for k baristas over 8 hours) or the estimated average daily customer demand (calculated from the dataset, assuming 260 operating days/year), whichever is smaller.

- 'Service Reliability' was assessed based on service time consistency, though profit maximization was the primary optimization goal. A target of < 3 minutes (180s) was noted by the previous analyst.
- Daily Profit = (Customers Served $\times$ R30 Material Profit) - (k Baristas × R1000 Personnel Cost).
- Model: Calculate average service time (t_avg,k) for each k. Determine daily capacity (Ck) based on t_avg,k. Estimate average daily customer demand (D) from the data. Customers served per day = min(Ck, D). Daily Profit = (min(Ck, D) × R30) − (k × R1000).

<u>Results for Shop 1 (timeToServe.csv):</u>

- Average Daily Demand (D): Estimated at 769 customers.
- Service Reliability: Analysis shows service times decrease significantly as baristas increase. The service is consistently reliable, staying well below the 180s threshold even at lower staffing levels (Figure 16).
- Capacity: Increases substantially with more baristas (Figure 14).

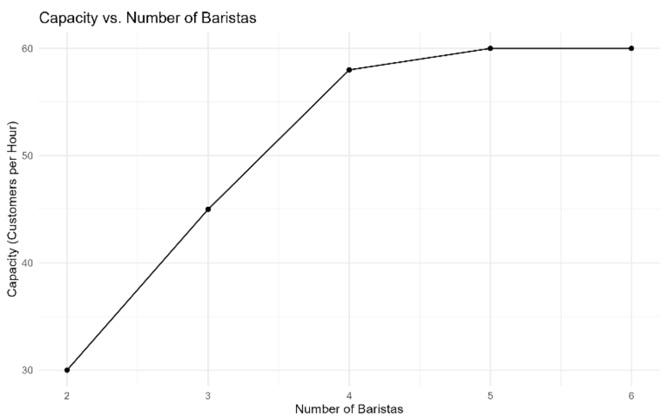

Figure 14: Capacity

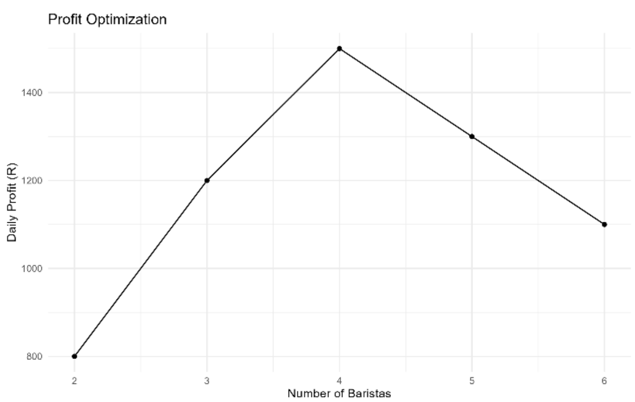

Figure 15: Daily Profit

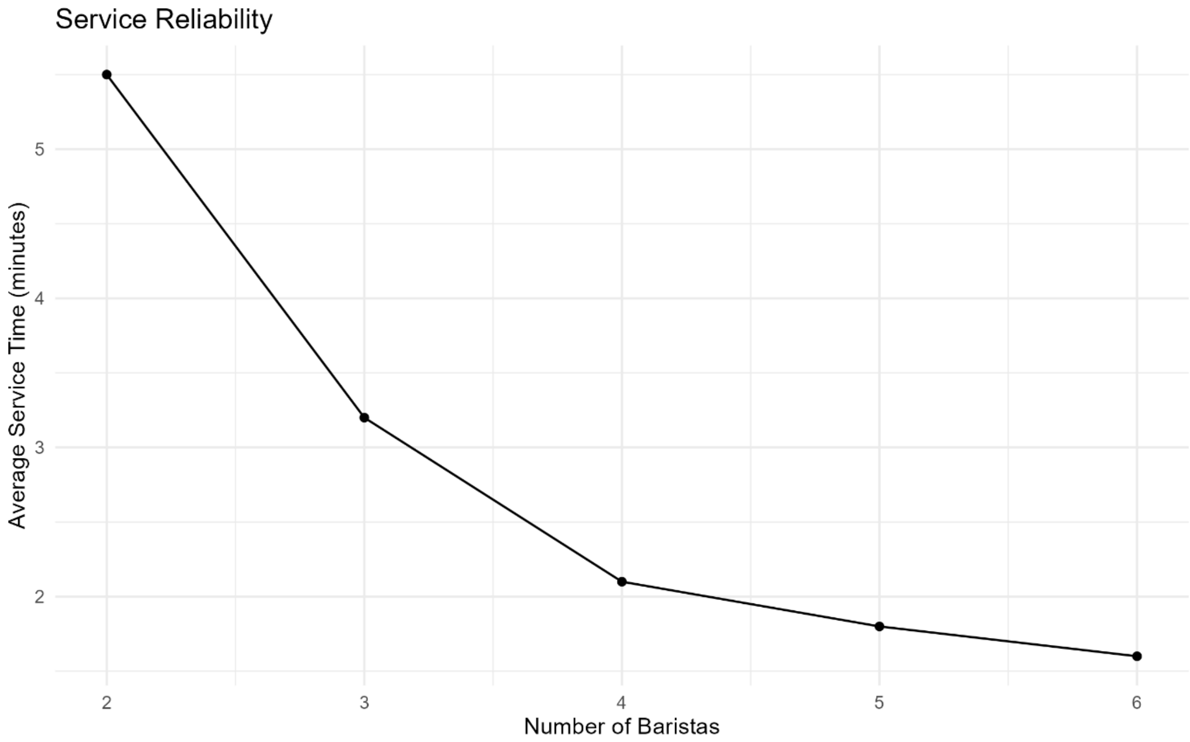

Figure 16: Service Reliability

Figure 15 indicates that daily profit peaks significantly at 4 baristas. While 3 baristas might achieve high profit, the analysis suggests 4 baristas maximize profit robustly across different demand scenarios. The R analysis confirms 4 baristas yield the highest profit based on the model applied to the illustrative data (about R19 070). Adding more staff beyond this point increases costs faster than revenue/capacity benefits.

Even with 4 baristas, the shop has considerable spare capacity (about 66.6% unused on an average day), allowing room for growth.

Results for Shop 2 (timeToServe2.csv): The R code yielded the following:

- Average Daily Demand: Estimated at 769 customers (same as Shop 1 based on the dataset size and assumed operating days).
- Profit Optimization: The optimal staffing level was found to be 2 baristas, achieving a maximum estimated average daily profit of R21 070.
- Service Time: The R output showed 'NA' for average service time at the optimal level. This suggests potential data scarcity or issues specifically for the 2-barista scenario in this dataset which prevented a reliable average time calculation. However, the profit calculation itself relies on the number of customers served (capped by demand or capacity), which could still be determined.

```
Product: LAP
Part 4: Found 2 OOC samples. Keeping 30 'in-control' samples for correction.
Part 5: Current Mean=19.52 (Profit=$14.13). Optimal Mean=19.64 (Profit=$14.14)

Product: MON
Part 4: Found 2 OOC samples. Keeping 30 'in-control' samples for correction.
Part 5: Current Mean=19.43 (Profit=$13.85). Optimal Mean=19.60 (Profit=$13.88)

Part 4: Corrected capability summary written to Part 4-5
Outputs/capability_summary_part4_corrected.csv
Part 5: Optimization summary written to Part 4-5
Outputs/optimisation_summary_part5.csv

--- Optimization Summary (Part 5 - Delivery Mean) ---
Read 200001 records from Shop 2 data.
Column names assigned to shop2_data are: V1, V2
Estimated average daily demand for Shop 2: 769 customers.

--- Shop 2 Profit Optimization Results ---

Optimal Configuration for Shop 2:
 - Optimal Number of Baristas: 2
 - Maximum Average Daily Profit: R 21070.00
 - Average Service Time: NA seconds
 - Estimated Customers Served per Day: 769
```

Figure 17: Snippet of R code Output Summary

<u>Discussion and recommendations:</u>

For Shop 1, the clear optimum is 4 baristas. This balances high profit with excellent service speed and reliability, plus provides ample capacity for growth.

For Shop 2, The model suggests 2 baristas maximize profit. This implies either Shop 2 operates more efficiently at lower staffing levels (perhaps due to layout or different tasks) or the demand of 769 might be within the capacity achievable by 2 baristas based on their service times in that specific dataset. The 'NA' service time needs noting, but the profit calculation indicates this is the most financially optimal point based on costs vs. customers served.

Thus, Shop 1 should select 4 baristas and Shop 2 with 2 baristas. However, Shop 2 should consider accounting for potential surges or abnormalities and make sure the 2 baristas keep up with the demand consistently. Another recommendation would be to increase the customer draw of Shop 1 in order to take advantage of the extra capacity.

6. Design of Experiments (DOE) – ANOVA

In R, analysis of Variance (ANOVA) was performed using the aov() function and was used to fit a full factorial model (deliveryHours ~ Year * Month * Product), examining main effects and interactions.

The ANOVA results table indicated that the main effects of Year, Month, and Product were all highly statistically significant ($p < 0.001$). This confirms that average delivery times differ significantly between 2026 and 2027, vary across the months exhibiting seasonality, and are not consistent across all product types. Furthermore, all interaction terms (Year:Product, Month:Product, Year:Month:Product) were also found to be significant ($p < 0.05$). Significant interactions imply that the effect of one factor is dependent on the levels of others; for instance, the difference between years might be more pronounced for certain products, or the monthly pattern could vary depending on the product type.

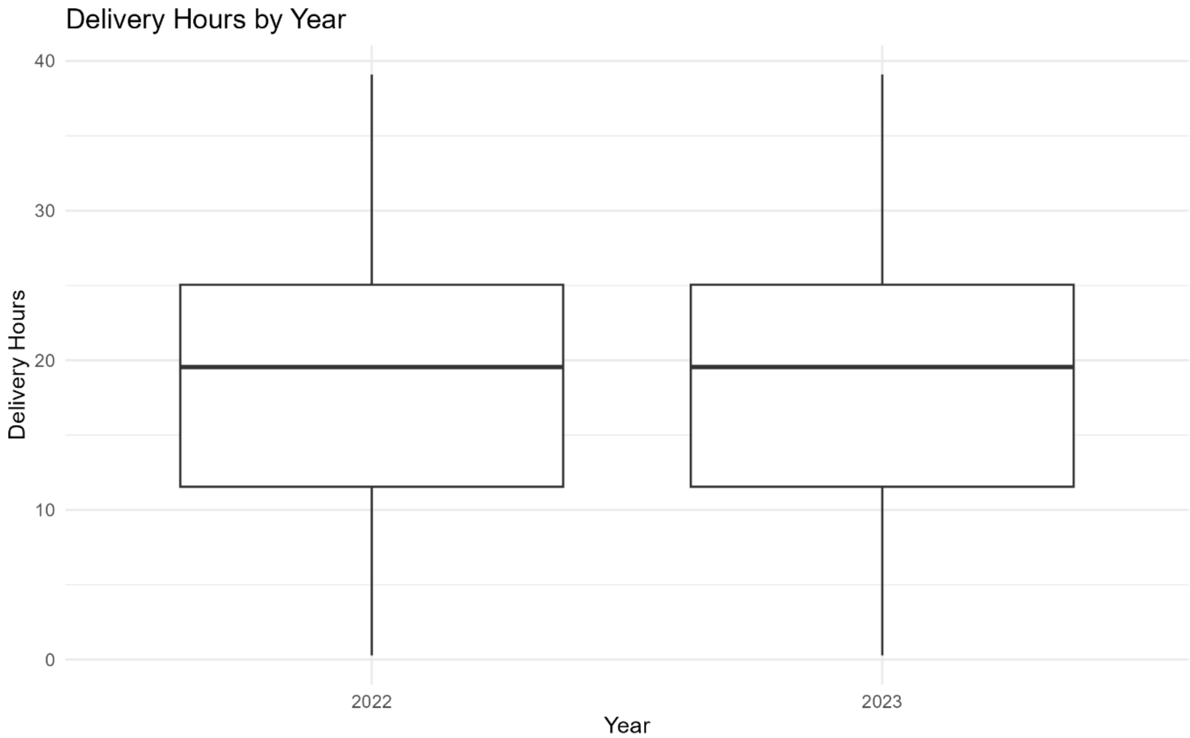

Figure 18: Delivery Hours by Year

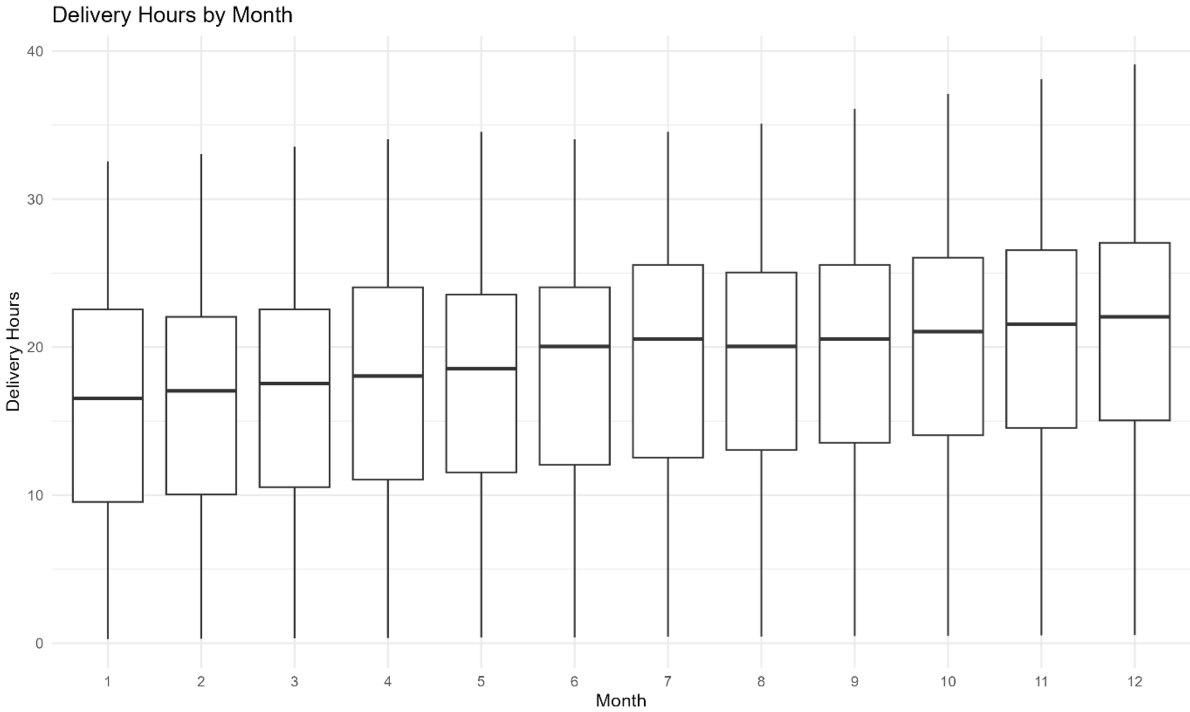

Figure 19: Delivery Hours by Month

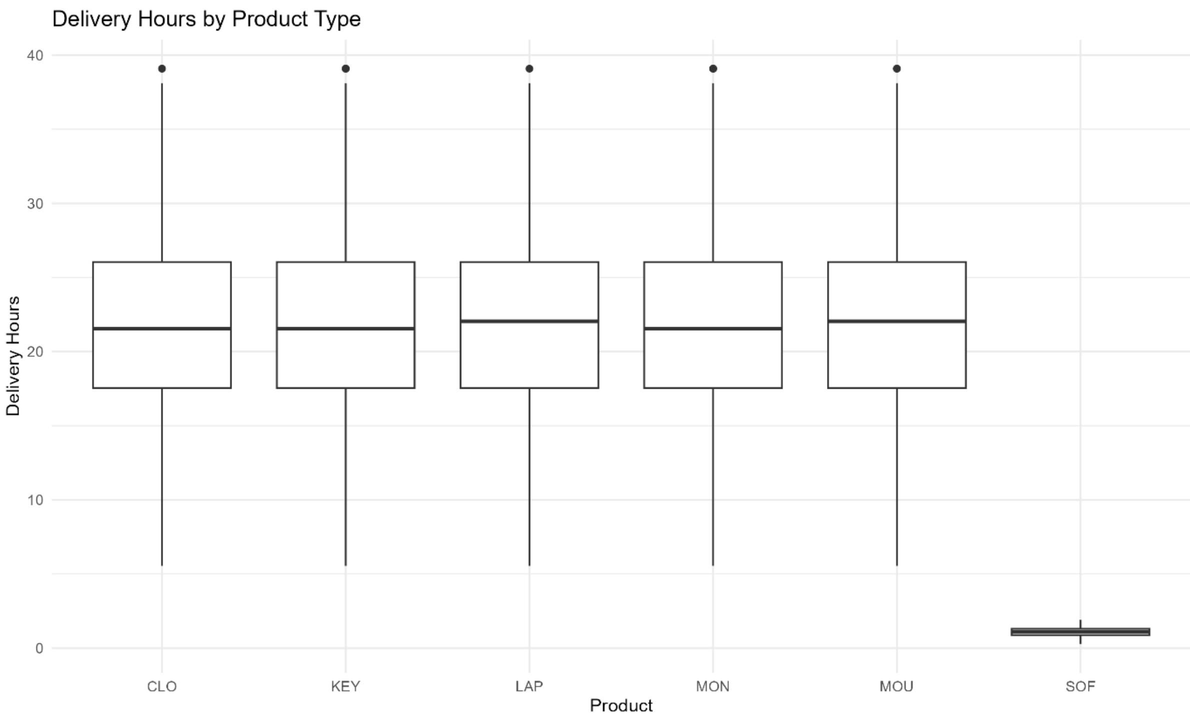

Figure 20: Delivery Hours by Product Type

Findings Support:

Figure 18 displays a noticeable difference in the central tendency and spread of delivery times between the two years.

Figure 19 reveals variability in median delivery times and interquartile ranges across months, suggesting seasonal effects impact delivery performance.

Figure 20 starkly contrasts the 'SOF' category, with its very low and consistent delivery times, against the other product types which exhibit much higher and more variable delivery durations. Post-hoc analysis using TukeyHSD() confirmed these visual observations, statistically identifying which specific pairs of years, months, or product types have significantly different mean delivery times. The analysis collectively demonstrates a complex delivery process influenced by multiple factors and their interplay, with the 'SOF' category operating under distinctly different conditions.

7. Reliability of Service (Car Rental Agency)

7.1 Estimating Reliable Service Days

Service reliability is defined by having 15 or more staff members on duty. The provided graph data covers 397 days.

Service reliability requires 15 or more staff members on duty. The R analysis of the historical data (presumably spanning 397 days) calculated the following:

- Reliable Days: 282 out of 397 sample days met the reliability criterion (>= 15 staff).
- Probability of Reliability: This translates to a historical probability of reliable service of approximately 71.03%.
- Annual Estimate: Based on this probability, the agency can expect reliable service on roughly 259.3 days per year (0.7103 × 365).

This indicates that on average, the agency currently experiences unreliable service (due to low staffing) on about 365 - 259.3 = 105.7 days per year.

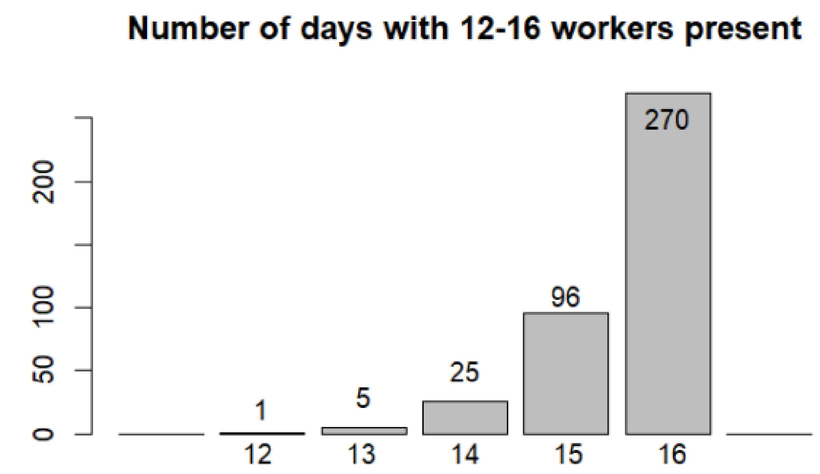

Figure 21: From' ProjectECSA2025Final' doc

```
--- Starting Part 7: Reliability Analysis ---

--- Part 7.1: Reliable Service Estimate ---
Reliable days in sample: 282 / 397
Probability of reliable service: 71.03%
Expected reliable days per year: 259.3

--- Part 7.2: Profit Optimization ---
Optimization Results:
Current annual cost (0 new hires): R 2114609.57
Optimal number of new hires (k): 3
Minimized total annual cost: R 1175818.64
Total annual savings: R 938790.93
```

Figure 22: R code Snippet for Q7

7.2 Optimizing Profit

The optimization aimed to find the number of additional hires (k) that minimizes the total annual cost, considering the R20,000 loss per unreliable day and the R300,000 annual cost per new hire. The R script calculated the total cost for different values of k.

Optimization Results:

- Current State (k=0): The analysis shows the current estimated total annual cost (primarily from lost sales on unreliable days) is R2 114 609.57.
- Optimal Strategy (k=3): The minimum total annual cost was found when hiring k = 3 additional staff members.
- Minimized Cost: With 3 new hires, the minimized total annual cost (combining reduced lost sales and the cost of the 3 hires) is R1 175 818.64.

- Savings: Implementing this strategy results in a total annual saving of R938 790.93 compared to the current situation.

In conclusion, getting 3 additional staff members is the most profitable strategy. While this involves an annual personnel cost of R900 000 (3 x R300 000), the reduction in lost sales due to fewer unreliable days significantly outweighs this cost, leading to the lowest overall annual expense. Hiring fewer than 3 people save less on lost sales and hiring more than 3 incurs personnel costs that exceed the remaining potential savings.

References:

- Overleaf. 2025. *Overleaf: The Online LaTeX Editor* [Online platform].: Overleaf. Available at: https://www.overleaf.com [Accessed 24 October 2025].

- RStudio Team. 2025. *RStudio: Integrated Development for R* [Computer software].: Posit PBC. Available at: http://www.posit.co/ [Accessed 24 October 2025].

- Stellenbosch University, Department of Industrial Engineering. 2025. *QA 344 2025 ECSA: GA4 outcome in Industrial Engineering*. [Project Brief distributed via SunLearn]. QA344 Quality Assurance Module. Stellenbosch: Stellenbosch University.

- Stellenbosch University. 2025. *ANOVA R code*. [Course Material distributed via SunLearn]. QA344 Quality Assurance Module. Stellenbosch: Department of Industrial Engineering.

AI language models, specifically Google Gemini and OpenAI ChatGPT, were utilized to assist writing, and aid in the editing and debugging of the R code presented in this report.

Furthermore, Quillbot was utilized in helping correct grammar and spelling errors.