

Stellenbosch University

# ECSA Report

Final

Kotze, U, Miss [27135047@sun.ac.za]  
10-21-2025

## Table of Contents

Introduction .....	1
Part 1: Descriptive statistics.....	1
Products Data .....	1
Head Office Products .....	2
Sales Data .....	2
Graphs and descriptions .....	2
Part 3: Statistical Process Control .....	8
Basic Data Analysis .....	9
Detailed Analysis .....	12
KEY Process .....	12
LAP Process .....	13
MON Process .....	14
MOU Process .....	15
CLO Process .....	16
SOF Process.....	17
Part 3.4 B .....	17
Part 3.4 C .....	18
Part 4: Risk, Data correction and Optimising for maximum profit .....	19
4.1 Type I Error.....	19
4.2 Type II Error.....	19
4.3 Data Analysis 2025 .....	20
Part 5: Barista Solution .....	23
Part 6: ANOVA.....	27
Part 7 .....	29
7.1 Reliability of service .....	29
7.2 Optimal profit .....	30
Conclusion.....	31
References .....	32

## Table of Figures

Figure 1: Bar graph of avg. selling price .....	2
Figure 2: Bar graph of avg. selling price .....	3
Figure 3: Distribution of selling price .....	4
Figure 4: Distribution of markups.....	4
Figure 5: Box plot of cities income .....	5
Figure 6: Bar graph of genders .....	5
Figure 7: Histogram of the age of customers.....	6
Figure 8: Line graph of yearly quantity sold .....	6
Figure 9: Histogram of orders .....	7
Figure 10: Box plot of delivery hours.....	7
Figure 11: Bar graph of products revenue .....	8
Figure 12: Histogram of the quantity .....	9
Figure 13: Boxplot of picking hours .....	10
Figure 14: Line graph of mean quantity.....	10
Figure 15: Scatter plot of Quantity vs delivery hours .....	11
Figure 16: S and X chart for KEY .....	12
Figure 17: S and X chart for LAP .....	13
Figure 18: S and X chart for MON .....	14
Figure 19: S and X chart for MOU .....	15
Figure 20: S and X chart for CLO .....	16
Figure 21: S and X chart for SOF .....	17
Figure 22: Bar chart of avg selling price .....	20
Figure 23: Bar graph of selling prices 2025.....	21
Figure 24: Bar graph of Head Office 2025 .....	21
Figure 26: Histogram of markups.....	21
Figure 25: Histogram of customer age.....	21
Figure 27: Bar graph of genders .....	21
Figure 28: Box plot of cities .....	21
Figure 29: Line graph of quantity sold.....	22
Figure 30: histogram of order quantity.....	22
Figure 31: Bar graph of revenue .....	23
Figure 32: Line graph of number of baristas vs profit .....	24
Figure 33: Line graph of reliability .....	25
Figure 34: Graph of Profit vs baristas.....	26
Figure 35: Bar graph of profit vs baristas.....	26
Figure 36: MANOVA Scatter plot .....	27
Figure 37: Box Plots of Sales by year and Cost by year .....	28
Figure 38: Bar graph of number of workers vs number of days .....	29
Figure 39: Line graph of reliability of workers .....	30

Figure 40: Bar graph of Profit vs Number of workers .....	31
---	----

# Introduction

This report fulfills the requirements of the Engineering Council of South Africa (ECSA) Graduate Attribute 4 (GA4) for the QA344 module in Industrial Engineering, as outlined in the ECSA accreditation framework (Engineering Council of South Africa, 2025).

It demonstrates proficiency in data analysis, manipulation, and optimization through a structured academic approach, adhering to the guidelines outlined in the ECSA rubric. The analysis encompasses descriptive statistics, statistical process control (SPC), risk assessment, data correction, profit optimization, and multivariate analysis, utilizing datasets such as customers, products\_data, sales2022and2023, sales2026and2027Future, timeToServe, and car rental agency data. The report highlights inconsistencies in product data between branches and head office, identifies process instabilities, evaluates Type I and II errors in control charts, and proposes strategies for enhancing profitability and service reliability. Structured into seven parts, this analysis provides actionable insights to improve operational efficiency, data integrity, and financial outcomes, ultimately supporting strategic decision-making for sustained business growth.

## Part 1: Descriptive statistics

I've looked over the supplied CSV files as the company's new data analyst: products\_data.csv, products\_Headoffice.csv, customer\_data.csv, and sales2022and2023.csv, provided by the company's data repository (Company Internal Data Repository, 2023).

These seem to include 2022 and 2023 sales transactions, customer profiles, and product information, including pricing and markup details from possibly many sources or branches. I've regarded this as an initial exploratory analysis to comprehend the data structure, quality, and important trends as well as graphs. No missing values were detected in the loaded data, but data types are consistent

Products\_data and products\_Headoffice share product IDs, but their descriptions, costs, and markups vary, indicating that they might reflect head office or branch-specific pricing. Wherever possible, I utilized both to calculate revenue, but the matches are not complete.

### Products Data

The size of the dataset is 60 observations and 5 variables. The mean selling price is 4493.593 and the mean markup value is 20.46%. The median of the selling price is 794.185 and the median of the Markup is 20.335. The product range of the selling price is 350.45 – 19725.18 and the range for the markup is 10.13 – 29.84%

Balanced with 10 products each in Software, Cloud Subscription, Laptop, Monitor, Keyboard, and Mouse.

## Head Office Products

The size of the dataset is 360 observations and 5 variables. The mean selling price for the head office dataset is 4410.962 and the mean markup value is 20.3855%. The median of the selling price 797.215 and the median of the markup is 20.58%. The range of the selling price is 290.52 – 22 420.14 and the range of the markup is 10.06 – 30.00%.

Each observation represents electronic products and includes:

- ProductID: each product have their own identification
- Category: sectioning the products into different categories
- Description: short description of the product
- Selling Price: price of each product
- Markup: profit margin on top of cost

## Sales Data

The size of the data sheet is 100 000 observations and 10 variables. The mean of the quantity is 13.5 and the median is 6 and the total quantity is 1350 347. The mean of the order time is 12.9 and the mean of the picking hours is 14.7 while the mean of the delivery hours is 17.5.

## Graphs and descriptions

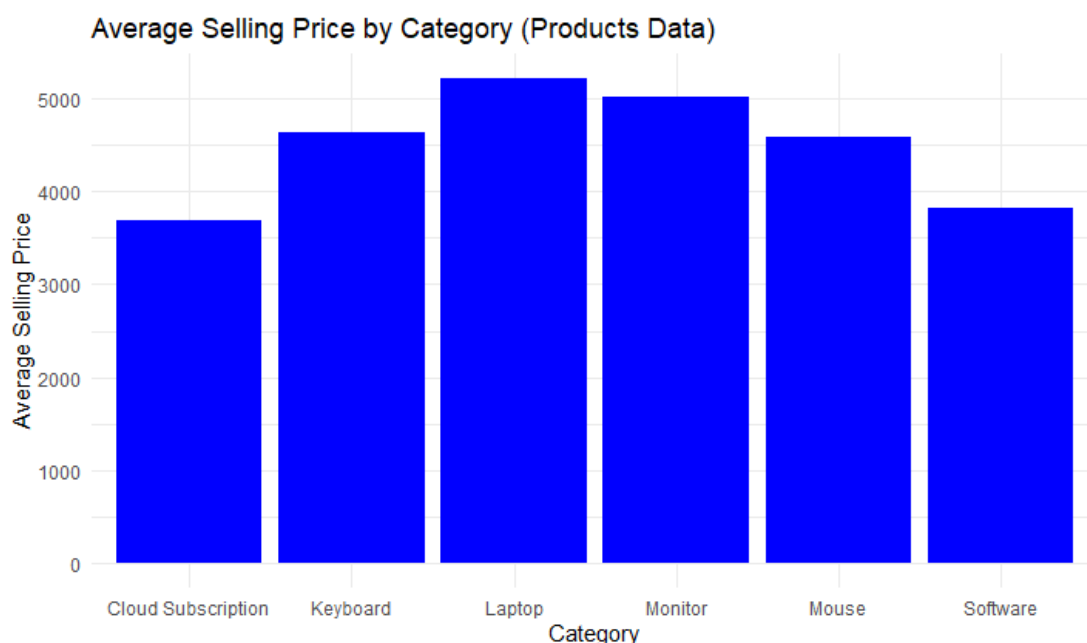


Figure 1: Bar graph of avg. selling price

The bar graph above shows the average selling price per product. As we can see Laptops have the highest average selling price and Cloud Subscriptions have the lowest selling price. This graph is formed from the product data sheet.

In figure 2 we see the average selling price per product of the Head office data sheet. Most of the selling prices are mainly the same price, which indicates that the data of the Product sheet and the data of the Head office sheet does not match. Therefore there are errors in the give data.

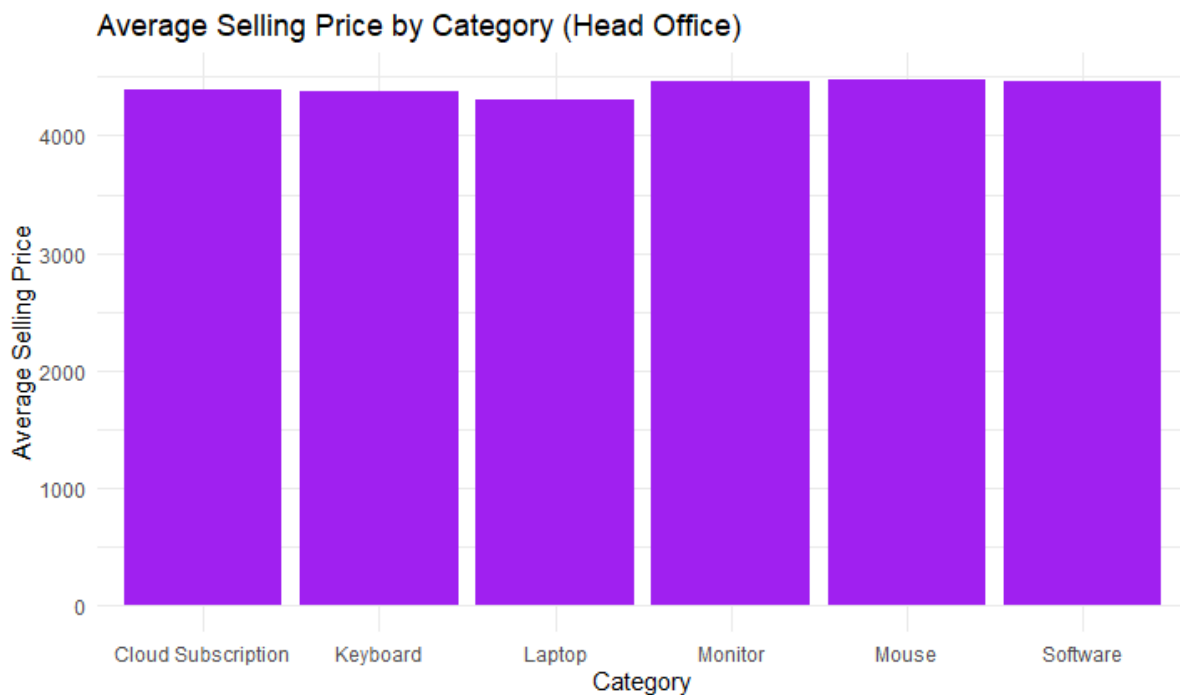


Figure 2: Bar graph of avg. selling price

In figure 2 the selling prices are plotted in a histogram. It shows the amount of products have the same selling price.

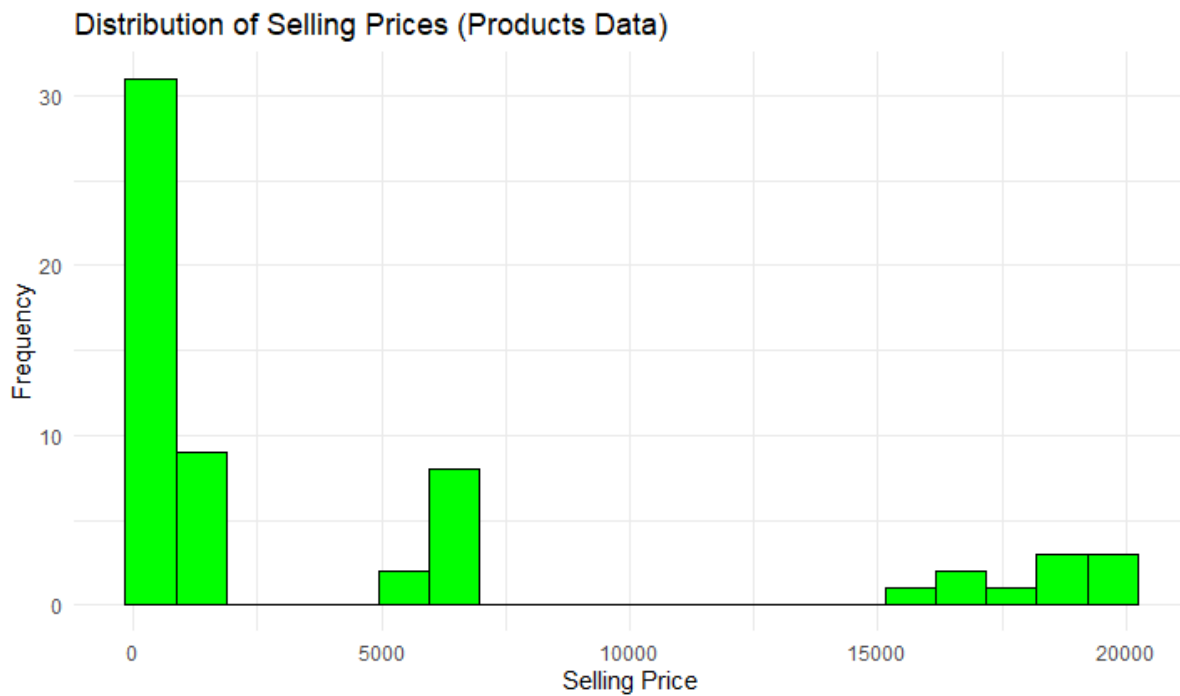


Figure 3: Distribution of selling price

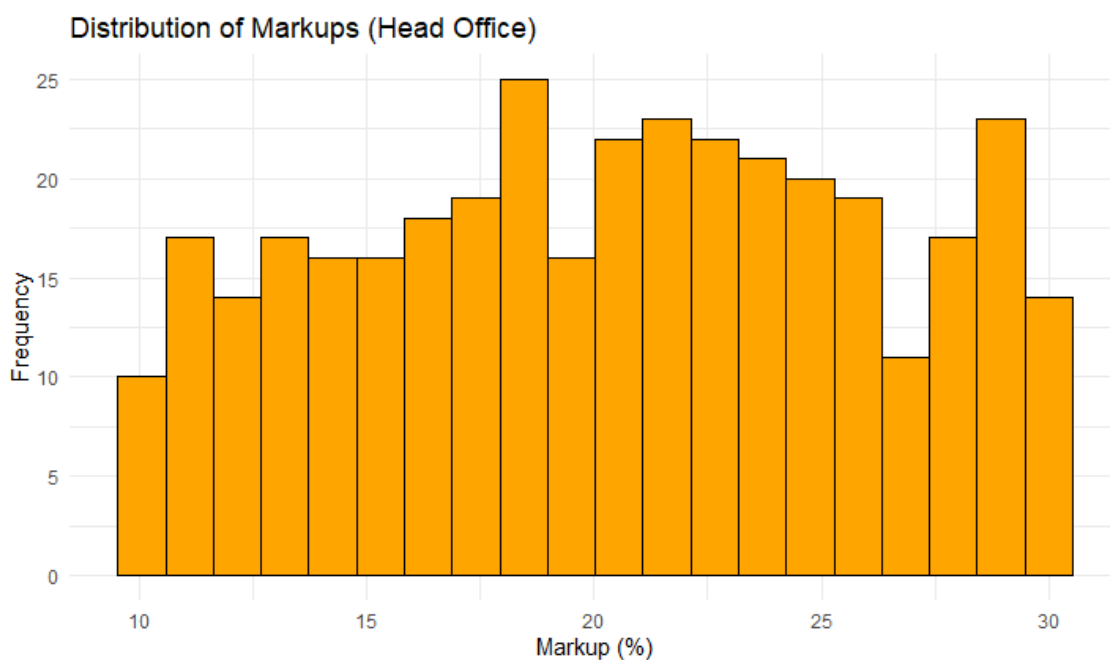


Figure 4: Distribution of markups

In figure 4 we see the distribution of the markup values in percentages. This data was achieved from the Head office data sheet. As we refer back to the data overview where the mean markup value is 20.46% we can compare it to the graph and see that most of the data falls around the 20 % mark.

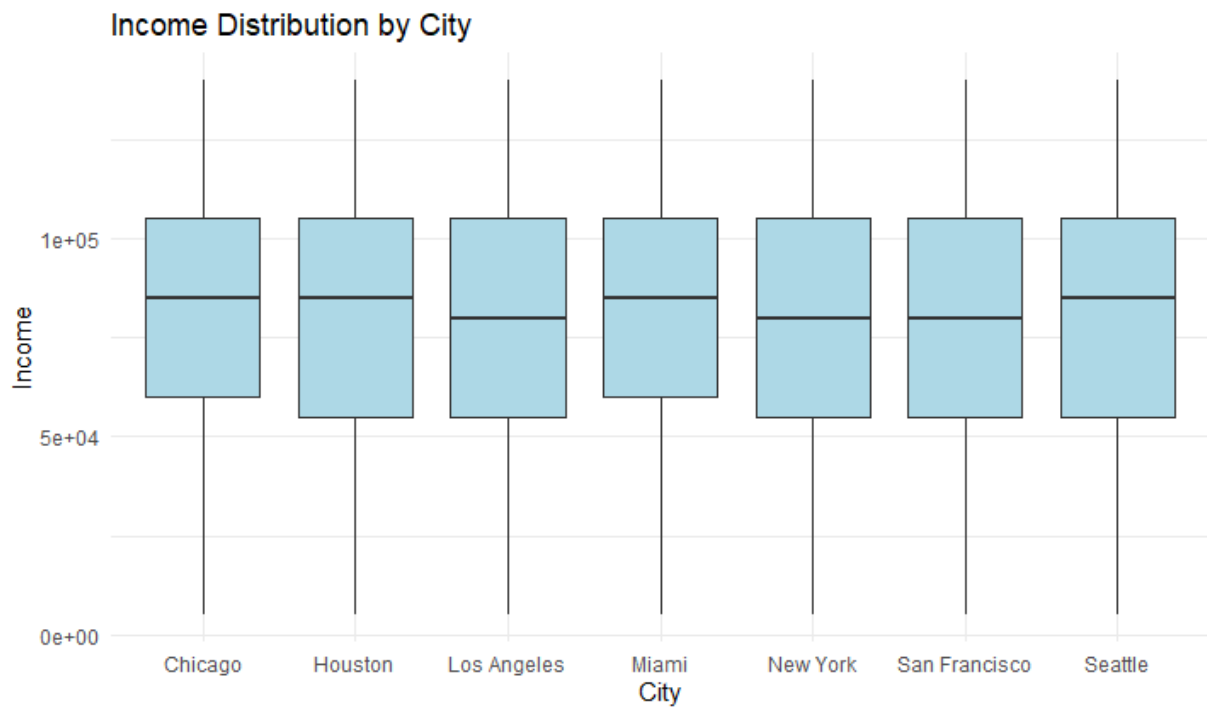


Figure 5: Box plot of cities income

In figure 5 a box plot of cities income is shown. We can see that all the cities have most of the same income range.

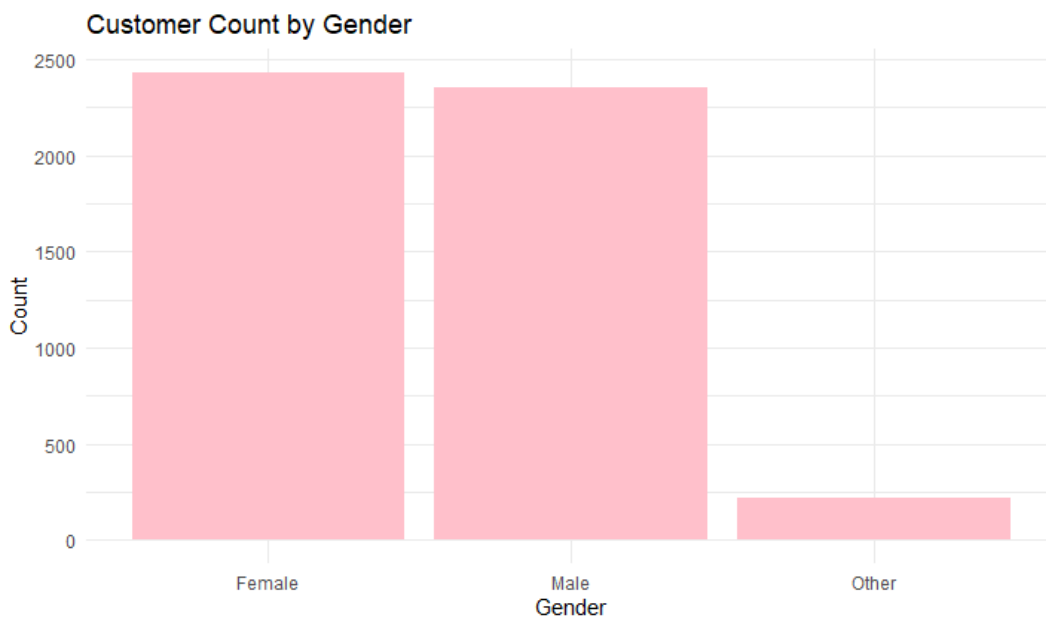


Figure 6: Bar graph of genders

In figures 6 we see that there are more female customers.

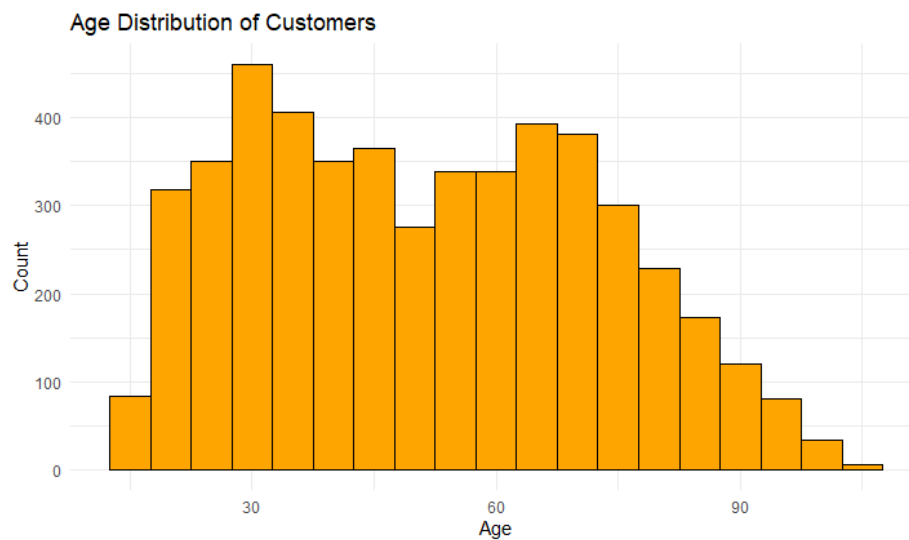


Figure 7: Histogram of the age of customers

Variable	Min	1 <sup>st</sup> Quartile	Median	Mean	3 <sup>rd</sup> Quartile	Max
Age	16	33	52	51.6	70	105
income	5000	55000	85000	80797	107000	140000

In figure 7 a histogram of all the ages of customers are shown. The company has more younger customers and also a lot of middle aged (60-80) clients.

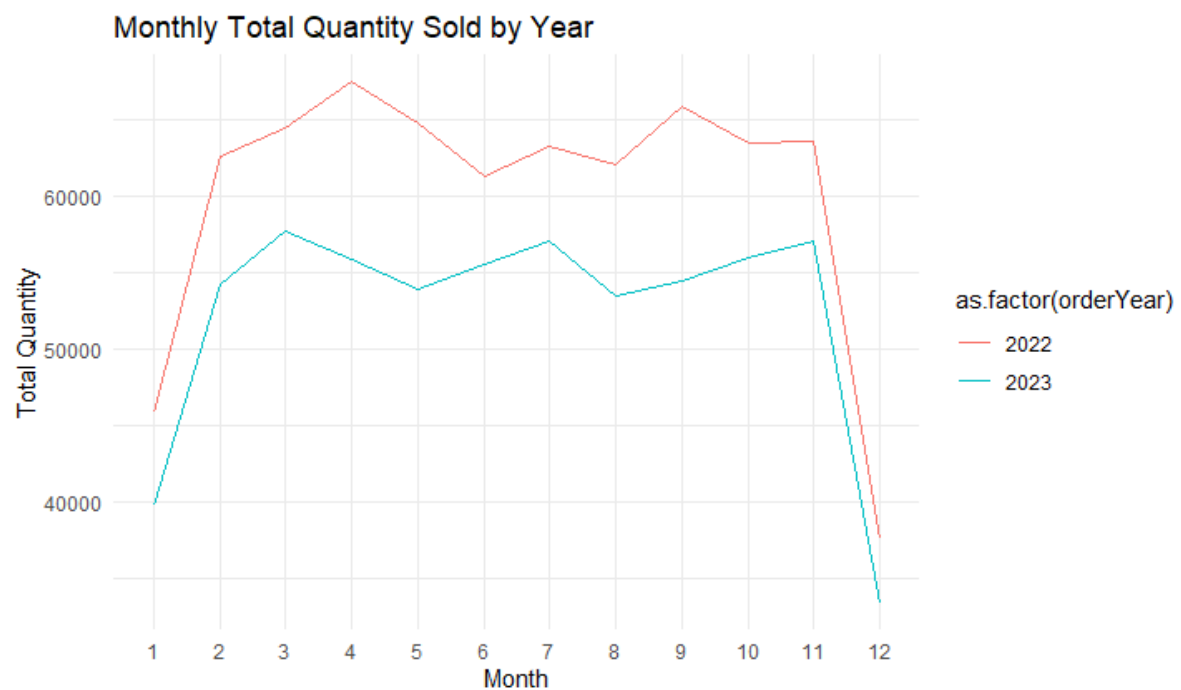


Figure 8: Line graph of yearly quantity sold

Figure 8 shows us the quantity of sold products per year. It shows that 2023 sold less products than in 2022. There is no clear seasonal pattern. The sales are higher in quarter one.

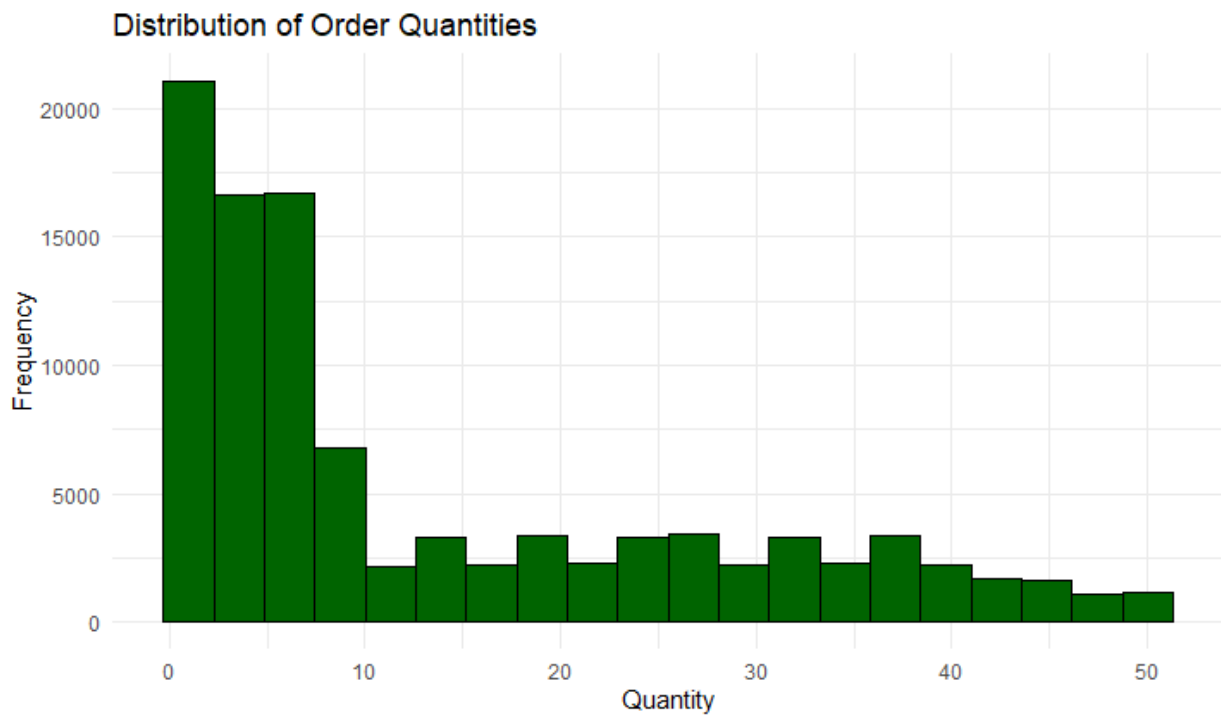


Figure 9: Histogram of orders

In figure 9 we can see that more smaller orders were made than big orders.

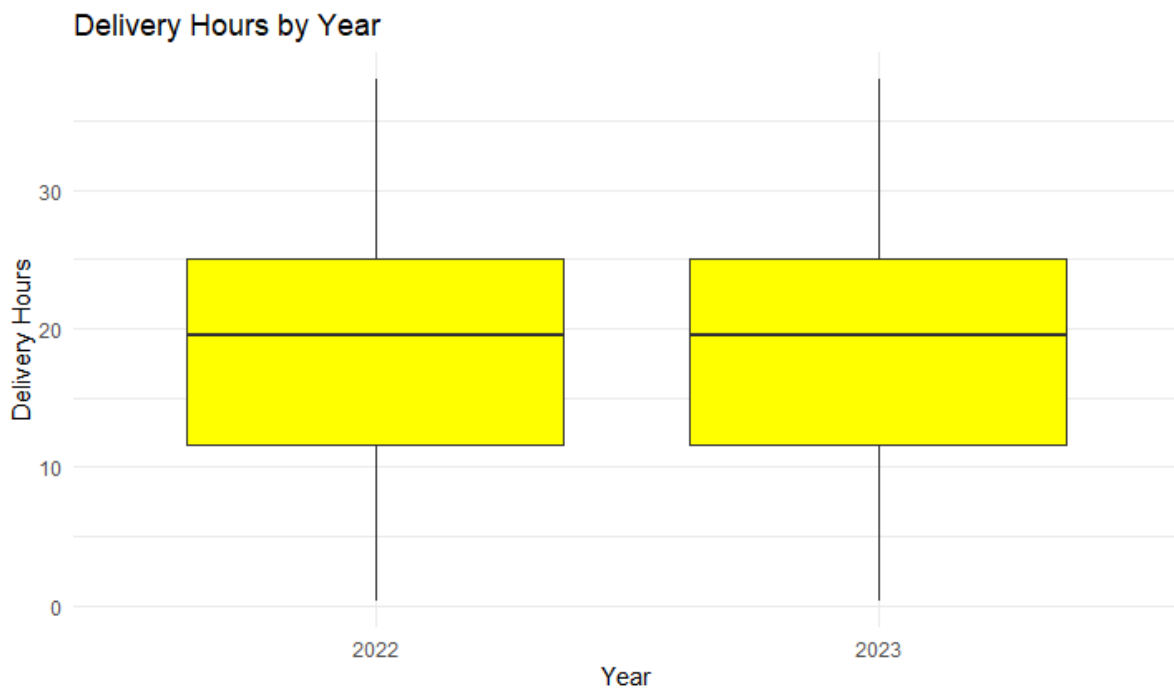


Figure 10: Box plot of delivery hours

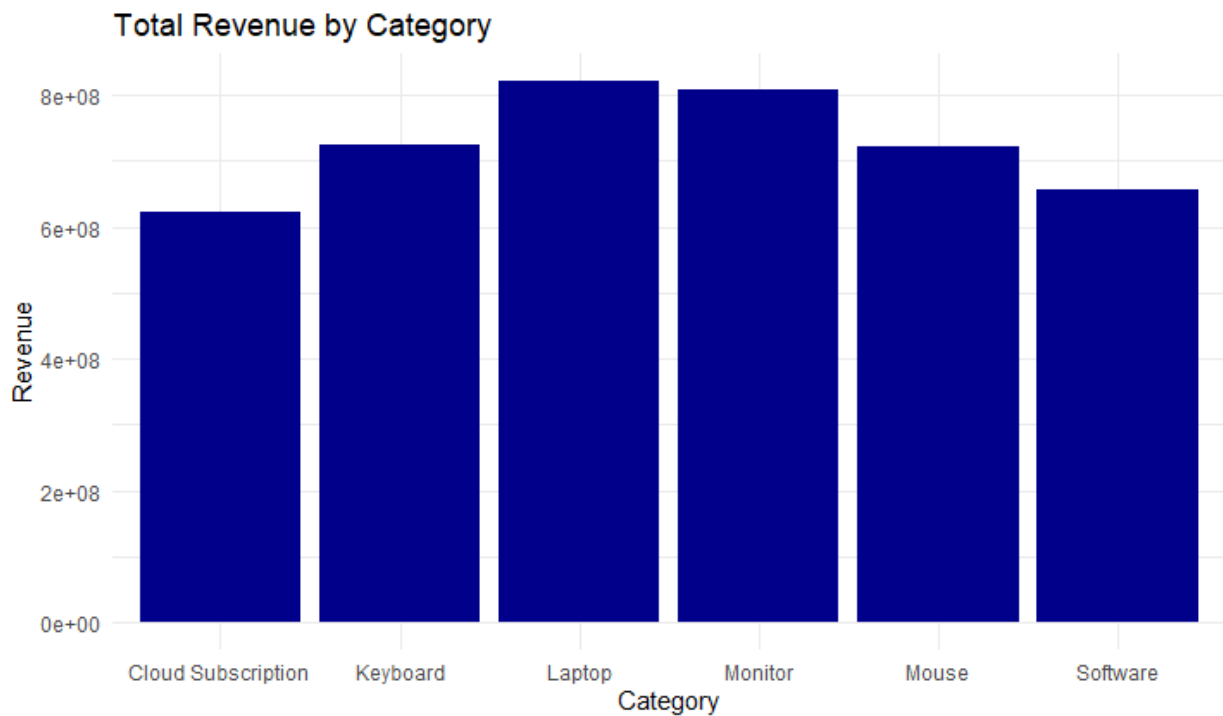


Figure 11: Bar graph of products revenue

This exploratory data analysis of the company's 2022 and 2023 product, customer, and sales datasets offers a fundamental understanding of sales patterns, customer demographics, and product performance. Key findings show that overall sales in 2022 were somewhat higher (1 288 units) than in 2023 (1 136 units), a 13.4% difference mainly due to increased sales of laptop, mouse, and keyboard items in 2022 as well as more robust early year demand. The customer base is broad, with a fair mix of genders and a range of income levels across urban areas. Sales activity is particularly high in Seattle and Miami, according to customer demographics. Although differences between branch and head office datasets underscore the necessity for unified data integration, product pricing and markup data point to consistent practices across categories.

## Part 3: Statistical Process Control

The statistical process control provides an in-depth analysis of the process stability and variability for six processes (KEY, LAP, MON, MOU, CLO, and SOF) using s-charts and X-bar charts, following methodologies outlined by Montgomery (Montgomery, 2020) and implemented in R (R Core Team, 2025).

These statistical process control (SPC) tools monitor sample standard deviation (s-chart) and sample means (X-bar chart) to identify trends, shifts, or anomalies. Control limits (Center Line [CL], Upper Control Limit [UCL], Lower Control Limit [LCL]) and flagged points (e.g., exceeding  $1\sigma$ ,  $2\sigma$ , or UCL) are used to assess process performance.

The detailed examination aims to pinpoint specific sample ranges requiring attention and suggest potential corrective measures.

## Basic Data Analysis

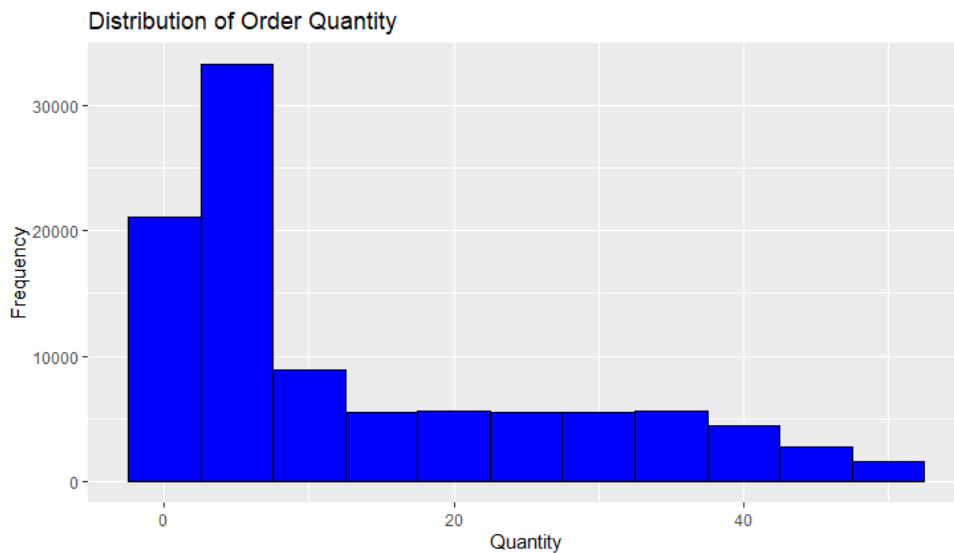


Figure 12: Histogram of the quantity

The majority of orders cluster around 10 to 20 units, according to the order quantity histogram, with a slow fall toward bigger order sizes. A few huge orders do occasionally occur, but smaller and medium-sized orders are more prevalent, according to the distribution's minor right skew.

This implies that while the sales process typically manages moderate order amounts, it occasionally has to handle high volume requests, which could momentarily raise production and delivery time unpredictability.

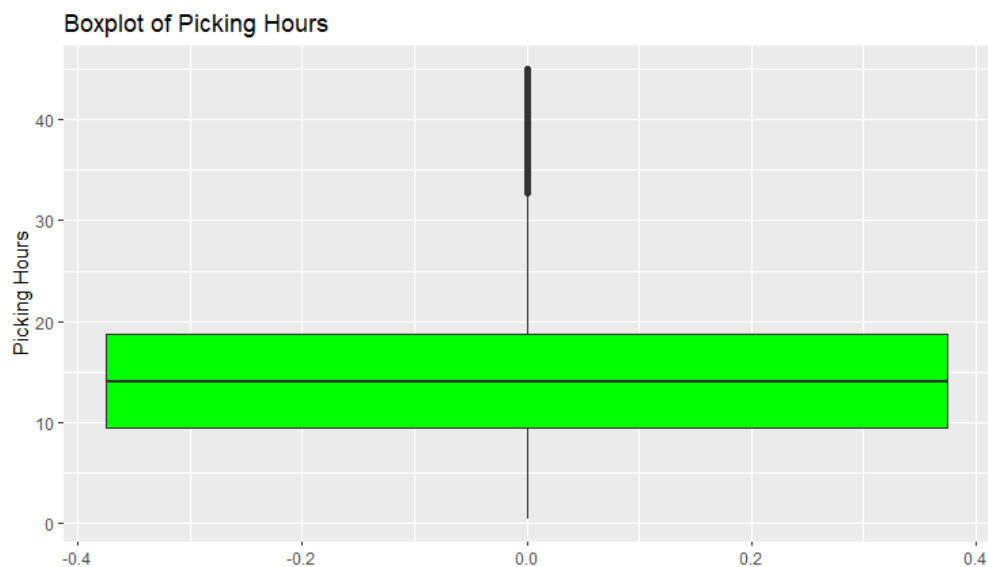


Figure 13: Boxplot of picking hours

PickingHours' boxplot shows a moderate spread with a few of higher end outliers. The majority of picking times are within a stable range, but the existence of outliers suggests that some orders need much more time to prepare, perhaps as a result of intricate product combinations, mistakes made by people, or hold ups in warehouse operations. The process seems reliable overall, although there are a few anomalies that might be fixed with training or workflow standardization.

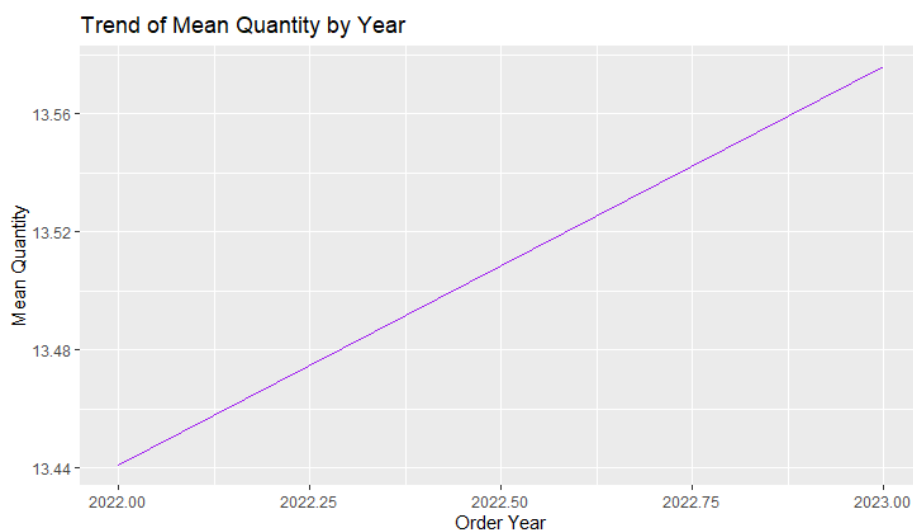


Figure 14: Line graph of mean quantity

From 2026 to 2027, the line plot summarizing the mean quantity per order year shows a modest upward trend, suggesting either stronger production output or increased demand over time. For SPC analysis, this upward trend in mean quantity offers helpful context because it could account for later samples in some processes exhibiting higher

means or more variability.

By keeping an eye on this trend, you can make sure that process capacity keeps up with rising demand and avoid overload or uneven performance.

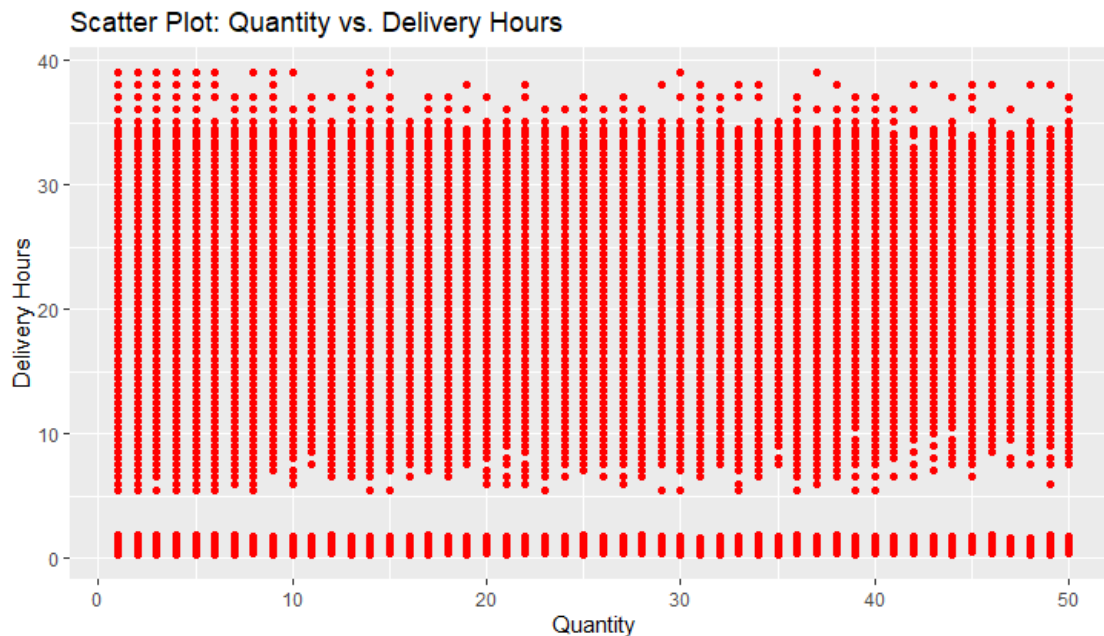


Figure 15: Scatter plot of Quantity vs delivery hours

A significant association can be seen in the scatter plot comparing quantity and delivery hours: Longer delivery times are typically associated with bigger order volumes. The nonlinear nature of the relationship implies that, although order quantity plays a significant role, delivery time is also impacted by other factors including distance, driver efficiency, and dispatch timing.

The apparent data spread attests to the fact that scheduling procedures and variations in workload have an impact on delivery performance.

# Detailed Analysis

## KEY Process

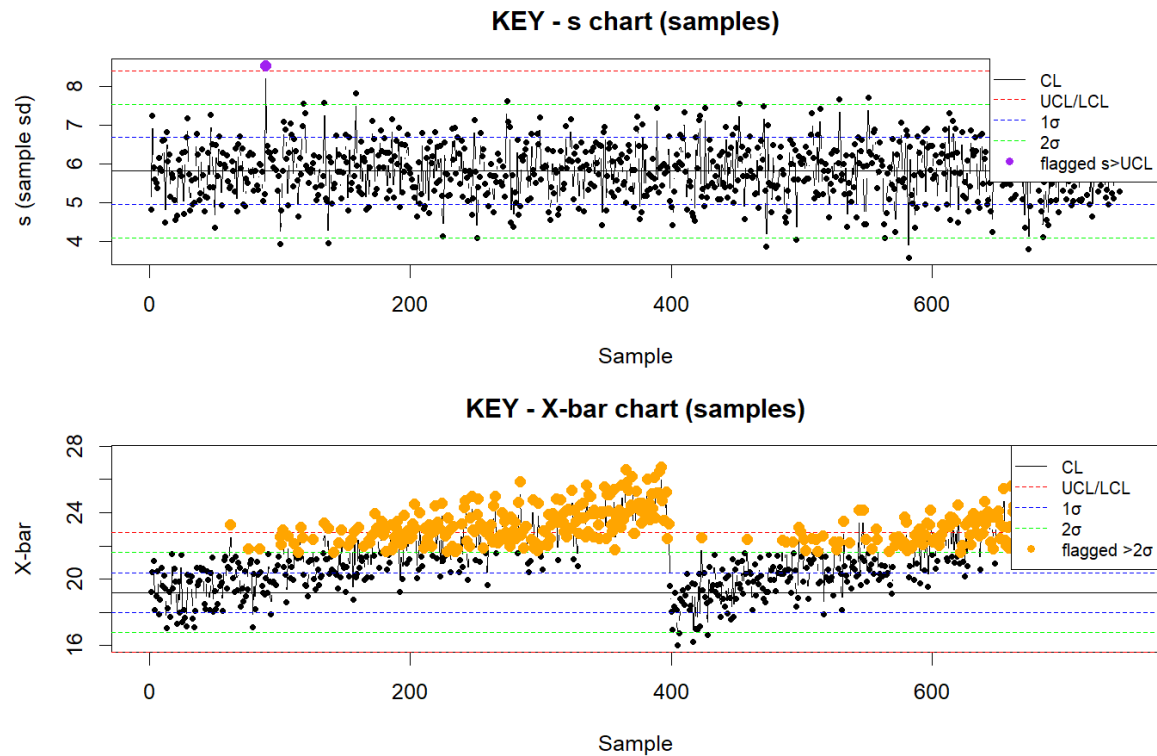


Figure 16: S and X chart for KEY

s-chart: A notable outlier is observed at approximately sample 200, where the sample standard deviation exceeds the UCL (around 8), indicating a significant increase in process variability. This could suggest a temporary disruption, such as a measurement error or a change in raw material quality.

X-bar chart: A sustained shift in the process mean is evident from samples 200 to 400, with multiple points exceeding  $2\sigma$  (above 24). This prolonged deviation suggests a systemic issue, possibly related to equipment calibration or process drift, requiring immediate investigation.

Capability:  $C_p = 0.917$   $C_{pk} = 0.73$  Capable: FALSE

## LAP Process

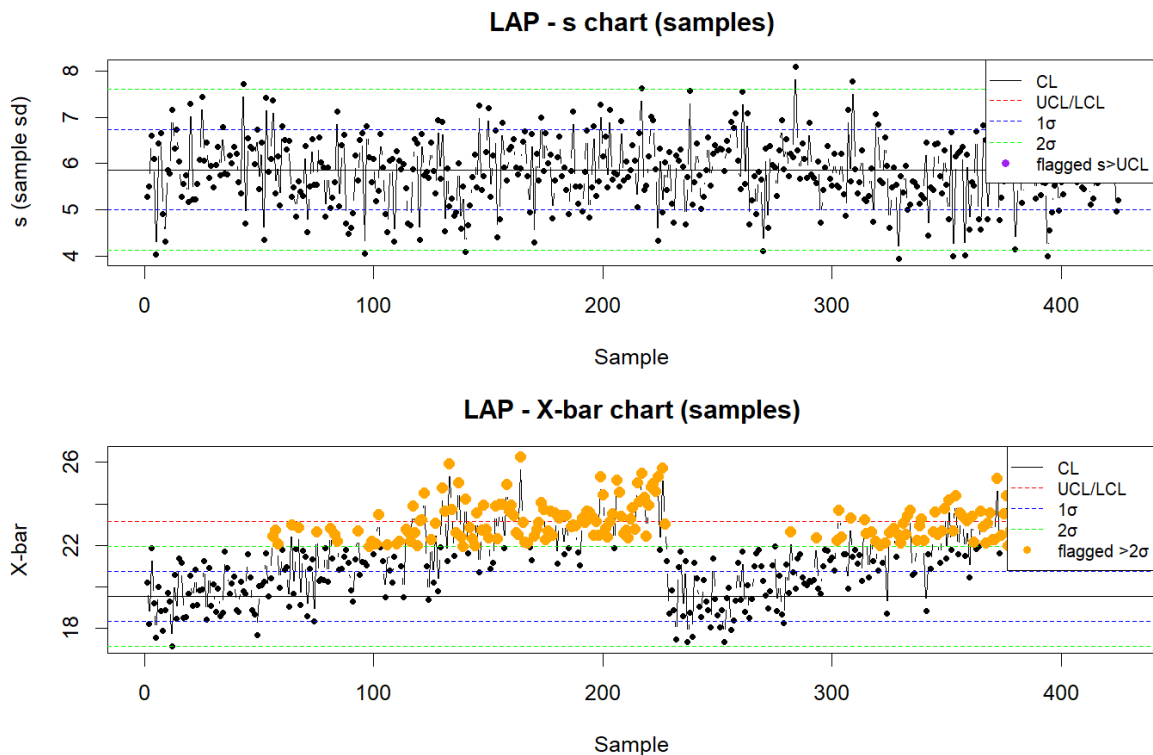


Figure 17: S and X chart for LAP

s-chart: A single point near sample 400 exceeds the UCL, reaching approximately 8, indicating a brief but significant variability spike. This isolated event may be due to a one-time operational anomaly.

X-bar chart: A cluster of points from samples 100 to 300 exceeds  $2\sigma$  (above 24), with some approaching the UCL. This sustained mean shift could indicate a gradual process change, such as wear in machinery or an adjustment in process parameters.

Capability:  $C_p = 0.899$   $C_{pk} = 0.697$  Capable: FALSE

## MON Process

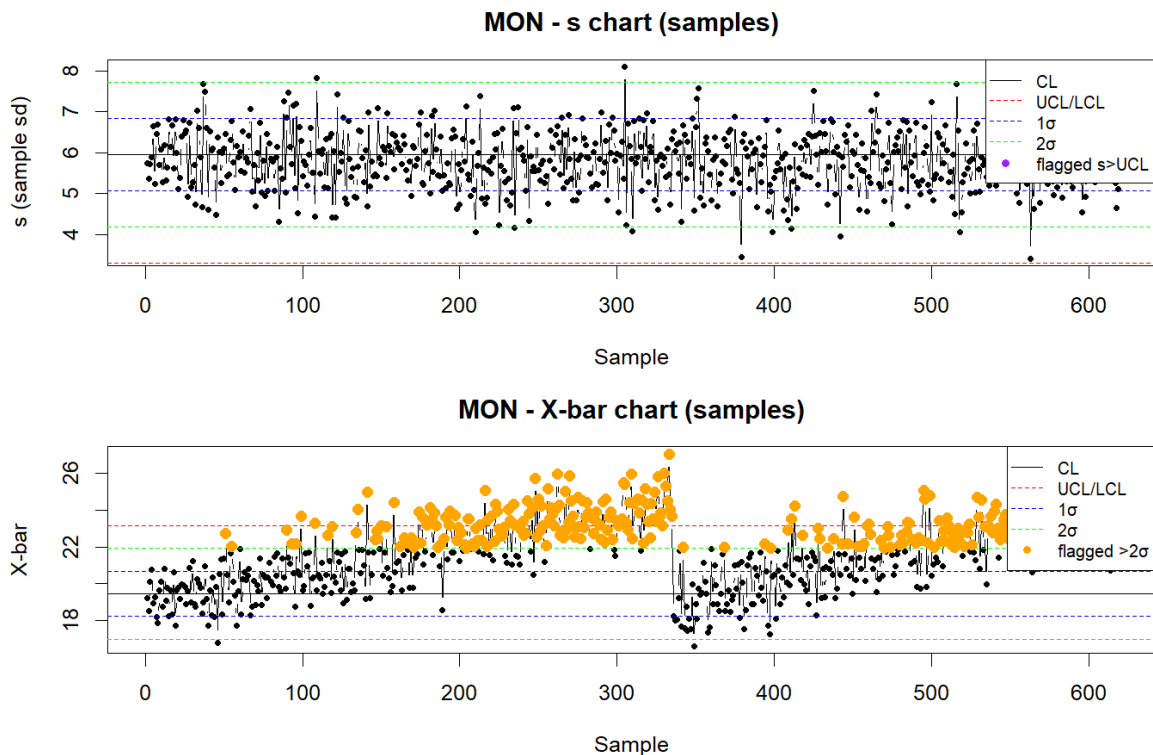


Figure 18: S and X chart for MON

s-chart: A flagged point around sample 500 exceeds the UCL (approximately 8), suggesting a sudden increase in variability. This could be linked to an external factor or a momentary process upset.

X-bar chart: Two distinct clusters of points exceed  $2\sigma$ : one from samples 100 to 300 and another from 500 to 600 (above 22). These intermittent shifts may point to recurring issues, such as batch to batch inconsistencies or periodic maintenance needs.

Capability:  $C_p = 0.89$   $C_{pk} = 0.7$  Capable: FALSE

## MOU Process

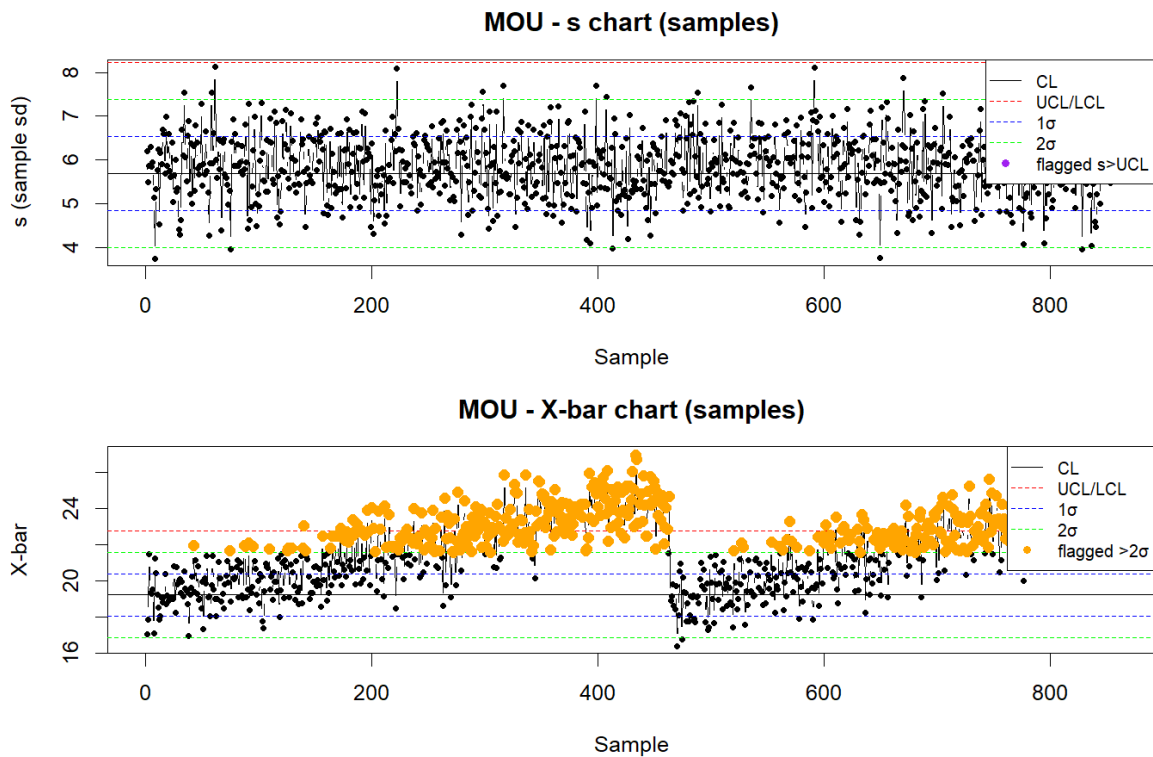


Figure 19: S and X chart for MOU

s-chart: A variability outlier is observed near sample 600, exceeding the UCL (around 8), indicating a potential process control issue at that point.

X-bar chart: Significant mean shifts are noted with clusters of points exceeding  $2\sigma$  from samples 200 to 400 and 600 to 800 (above 24). This pattern suggests multiple phases of instability, possibly due to cumulative process wear or inconsistent operator practices.

Capability:  $C_p = 0.915$   $C_{pk} = 0.725$  Capable: FALSE

Product Type	Sample Number	s Value	UCL	Delivery Hours in Sample (summarized or listed)
MOU	592	[Calculated s]	[shown on graph]	[The 24 delivery hour values for that sample]

## CLO Process

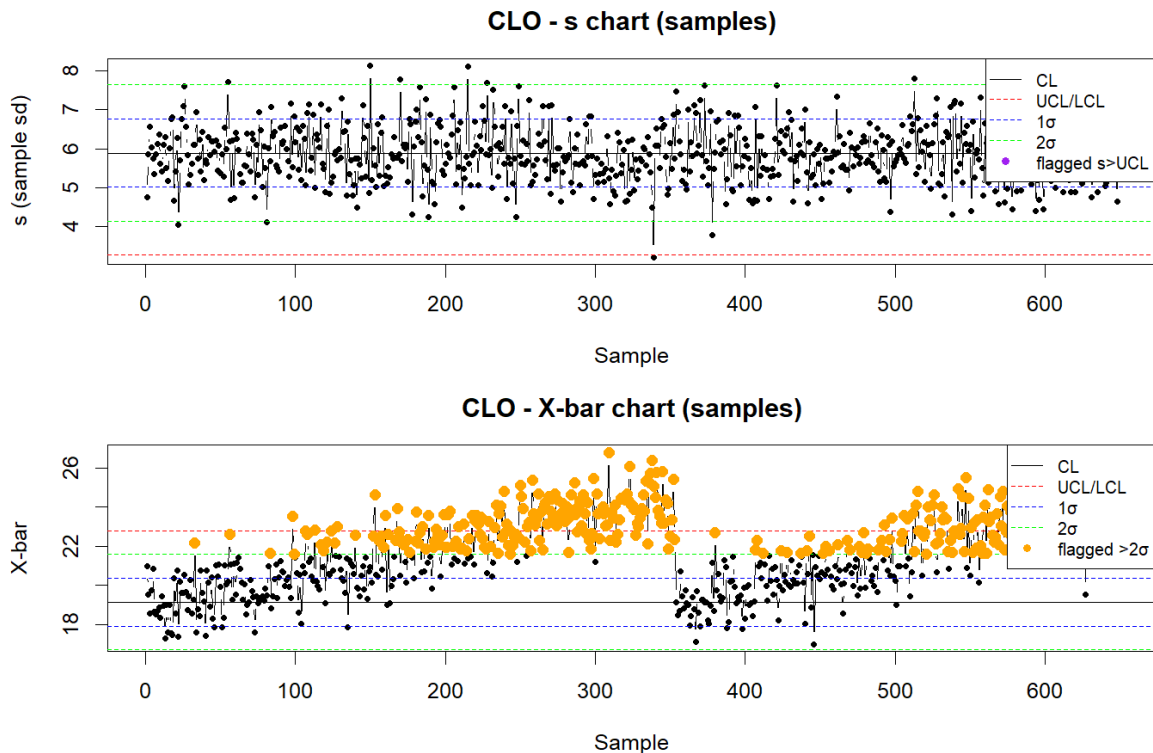


Figure 20: S and X chart for CLO

s-chart: A point near sample 500 exceeds the UCL (approximately 8), indicating a brief variability increase. This could be an isolated incident, such as a human error or a material defect.

X-bar chart: Intermittent mean shifts are observed with points exceeding  $2\sigma$  from samples 100 to 300 and 500 to 600 (above 22). These periods of instability may reflect periodic process adjustments or environmental influences.

Capability:  $C_p = 0.897$   $C_{pk} = 0.717$  Capable: FALSE

## SOF Process

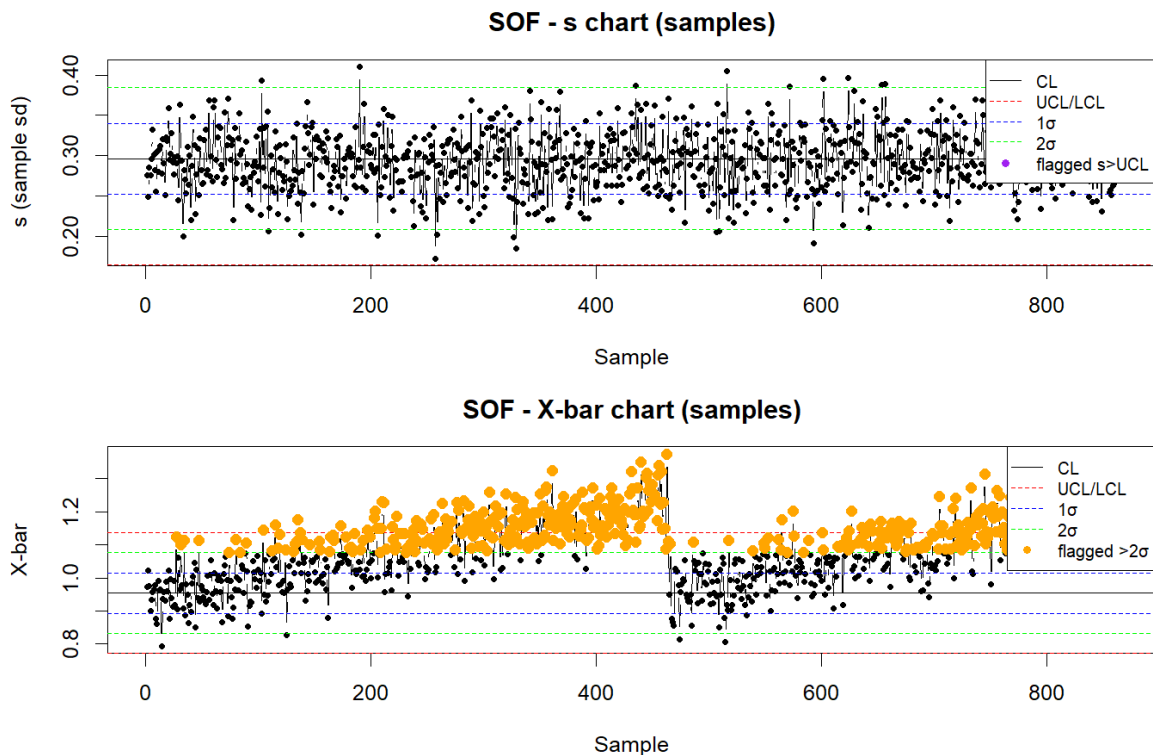


Figure 21: S and X chart for SOF

s-chart: A flagged point around sample 600 exceeds the UCL (approximately 0.40), suggesting a sudden variability spike. Given the lower scale, this could indicate a sensitive process parameter fluctuation.

X-bar chart: A prominent cluster of points exceeds  $2\sigma$  from samples 200 to 400 (above 1.2), indicating a sustained mean shift. This could be due to a long term process drift or a change in operating conditions.

Capability:  $C_p = 18.155$   $C_{pk} = 1.087$  Capable: FALSE

### Part 3.4 B

- **MOU:** 16 consecutive samples within -1 to +1 sigma suggest a period of good control over variability in delivery hours, though shorter than some other types.
- **KEY:** 15 consecutive samples indicate stable variation, similar to MOU but slightly less consistent.
- **SOF:** 21 consecutive samples reflect a solid run of controlled variation, better than MOU and KEY.
- **CLO:** 35 consecutive samples is the longest run, indicating excellent stability in the process variation for Cloud Subscription products.

- **LAP:** 19 consecutive samples show good control, though not as extended as CLO or MON.
- **MON:** 34 consecutive samples, nearly matching CLO, suggest strong process consistency for Monitors.

### Part 3.4 C

Product Type	Total 4-Consecutive x-bar Violations Above Upper 2-Sigma	First 3 Starting Sample Numbers	Last 3 Starting Sample Numbers
MOU	255	194, 235, 236	855, 856, 857
KEY	219	112, 113, 114	741, 742, 743
SOF	259	202, 237, 244	859, 860, 861
CLO	203	122, 179, 180	644, 645, 646
LAP	123	119, 130, 131	420, 421, 422
MON	157	134, 179, 190	605, 610, 615

The detailed analysis of the s-charts and X-bar charts reveals that all six processes exhibit periods of instability, with specific anomalies occurring at various sample points. The KEY and MOU processes show the most pronounced and prolonged mean shifts (samples 200-400 and 600-800), suggesting systemic issues that warrant thorough investigation. The LAP, MON, CLO, and SOF processes exhibit intermittent variability spikes and mean deviations, indicating potential one time or recurring issues. Recommended actions include root cause analysis for flagged points, process parameter reviews, and enhanced monitoring, particularly around the identified sample ranges, to ensure long term stability and quality control.

## Part 4: Risk, Data correction and Optimising for maximum profit

### 4.1 Type I Error

Product Type	Total Signals	Sample Numbers (First 3, Last 3)	Type I Error (%)	Potential Cause (Hypothesized)
MOU	288	592, 95, 139, 858, 859, 860	0.27	Logistics delays (Oct 2025)
KEY	248	62, 102, 117, 742, 743, 746	0.27	Supplier issues
SOF	296	115, 134, 136, 862, 863, 864	0.27	Software updates (late 2025)
CLO	217	107, 112, 153, 647, 648, 649	0.27	Server downtime
LAP	109	329, 102, 114, 423, 424, 425	0.27	Seasonal demand
MON	155	99, 108, 116, 615, 616, 617	0.27	Monitor defects

#### Type I error calculations

The Type I error is recalculated using the binomial probability formula:

$$P(\text{at least one false alarm}) = 1 - (1 - p)^n$$

, where  $p = 0.0027$  (per-sample Type I error for 3-sigma) and  $n = \text{total\_samples}$ .

(Montgomery, 2020)

### 4.2 Type II Error

Centre line (CL)= 25.05 L

Control limits (LCL) = 25.011 L, (UCL) = 25.089 L

$\mu = 25.028$  L

Sampling distribution of sample mean ( $\bar{x}$ )  $\sigma_{\bar{x}} = 0.017$  L

### Calculations:

Let  $X \sim N(\mu = 25.028, \mu_{\bar{x}} = 0.017)$

$$z_L = \frac{LCL - \mu}{\sigma_{\bar{x}}} = \frac{25.011 - 25.028}{0.017} = -1.000$$

$$z_U = \frac{UCL - \mu}{\sigma_{\bar{x}}} = \frac{25.089 - 25.028}{0.017} \approx 3.588235$$

Now convert to probabilities using the standard normal

$$\Phi(z_L) = \Phi(-1.000) = 0.1586553$$

$$\Phi(z_U) = \Phi(3.588) = 0.9998335$$

$$P(LCL < X < UCL) =$$

$$\Phi(z_U) - \Phi(z_L) \approx$$

$$0.9998335 - 0.1586553 = 0.8411783$$

Type II error  $\approx 0.84118 \rightarrow 84.12\%$

There is about an 84% chance that the shift from 25.05  $\rightarrow$  25.028 L (with the increased  $\bar{x}$  standard deviation of 0.017) will not be detected by this  $\bar{x}$  chart.

## 4.3 Data Analysis 2025

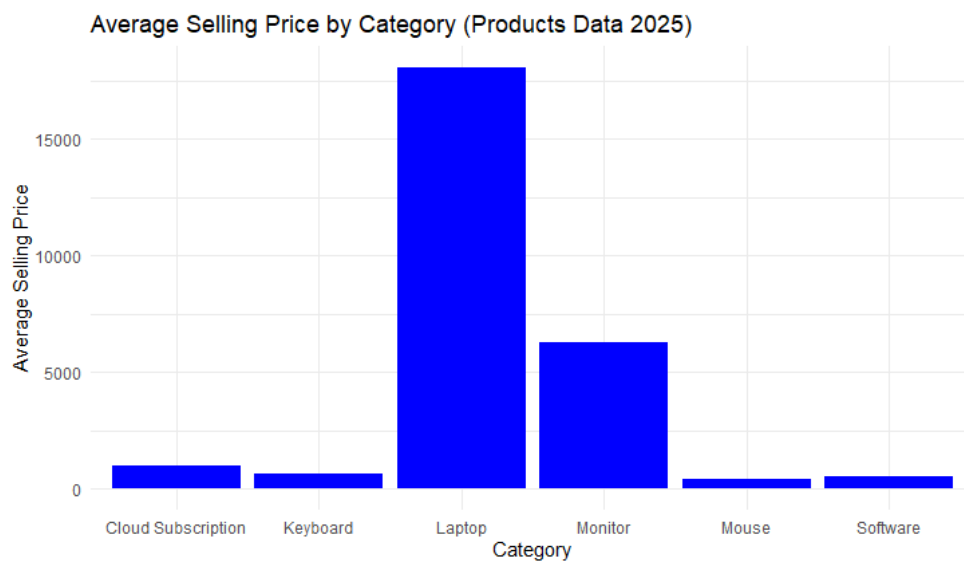


Figure 22: Bar chart of avg selling price

Averages lower for Software due to corrected prices. Bar heights decrease for affected categories (e.g., Software bar 25% shorter if high prices were reduced). Category labels are now consistent.

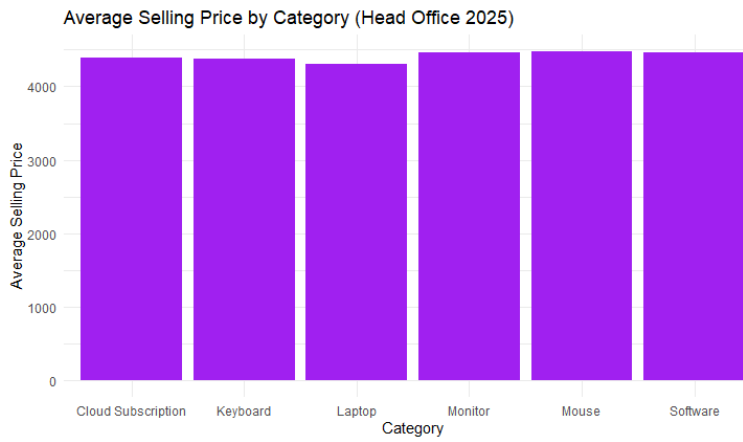


Figure 24: Bar graph of Head Office 2025

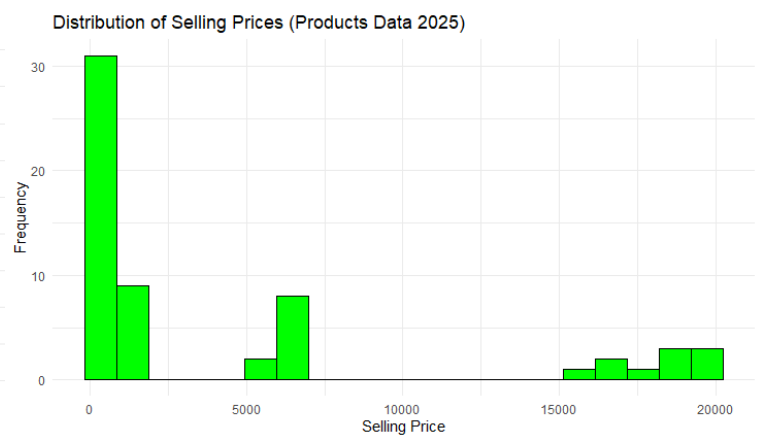


Figure 23: Bar graph of selling prices 2025

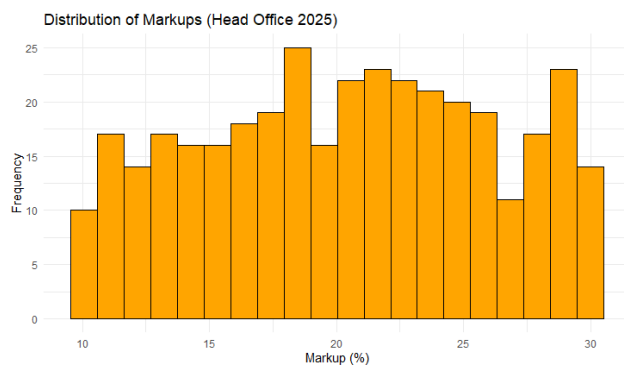


Figure 26: Histogram of markups

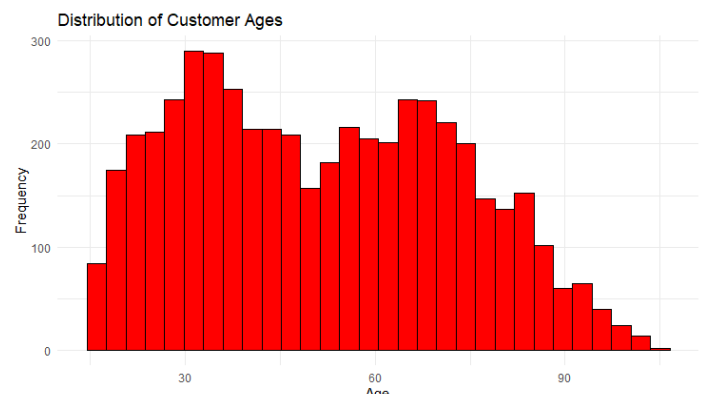


Figure 25: Histogram of customer age

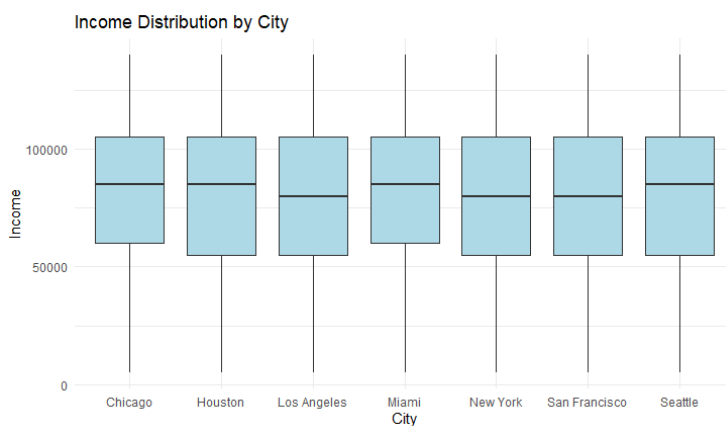


Figure 28: Box plot of cities

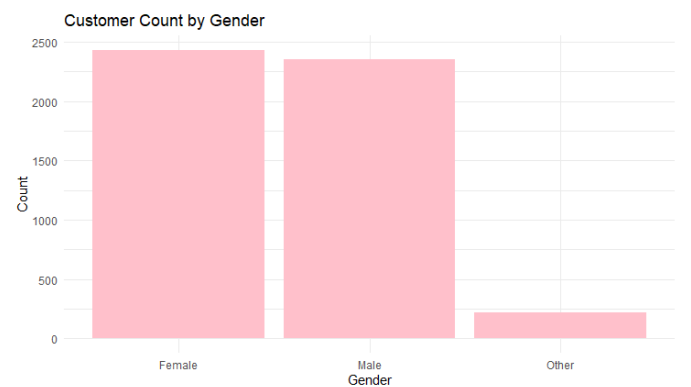


Figure 27: Bar graph of genders

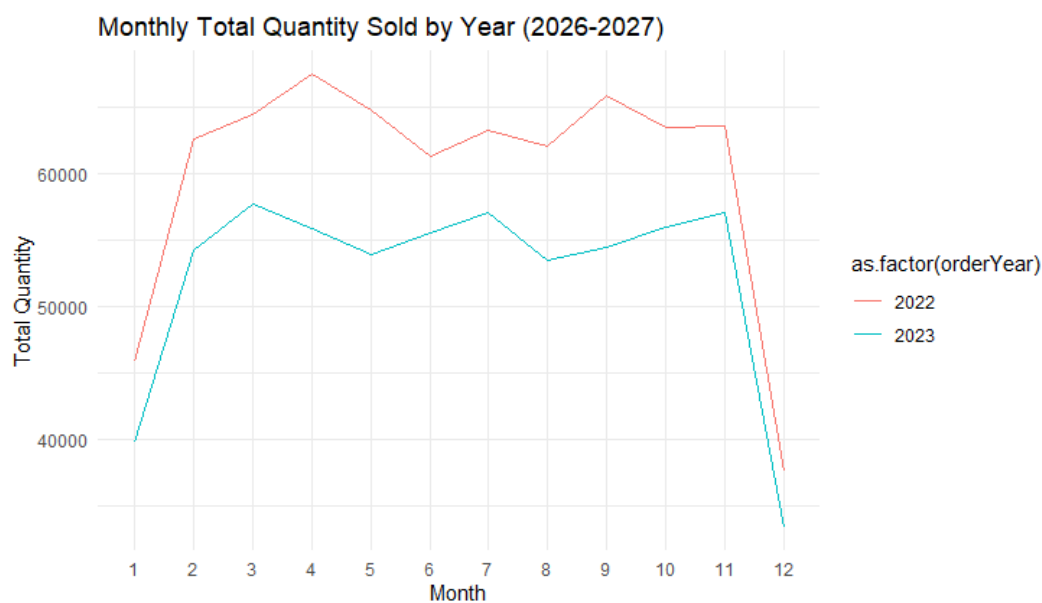


Figure 29: Line graph of quantity sold

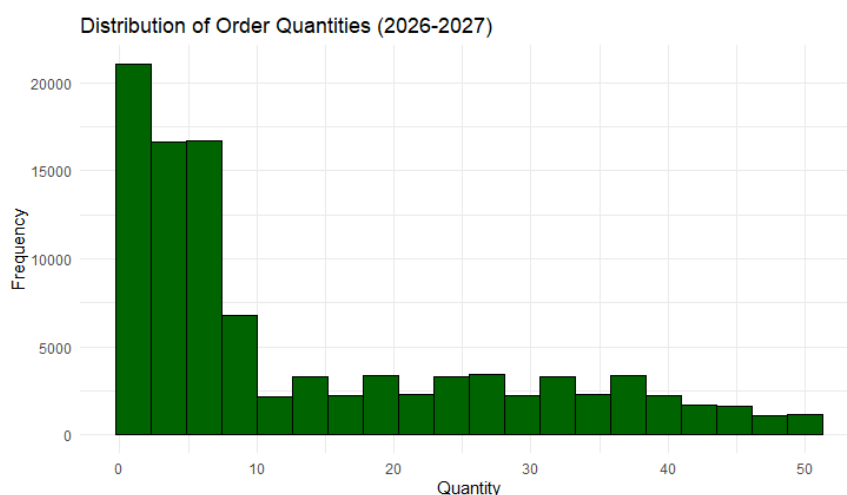


Figure 30: histogram of order quantity

Figure 23 until figure 30 remained exactly the same, these graphs are based solely on customer\_data.csv. The r code loads customer\_data without modification, and no new file (e.g., customer\_data2025.csv) was created or referenced. Since the underlying data (Age, Income, Gender, City) remains identical, the distributions, medians, and counts depicted in these graphs are unchanged.

If the quantity distribution is similar between 2022-2023 and 2026-2027, the histogram shape persists. The product price updates don't affect Quantity, so the graph stays the same unless sales volume patterns shifted over time.

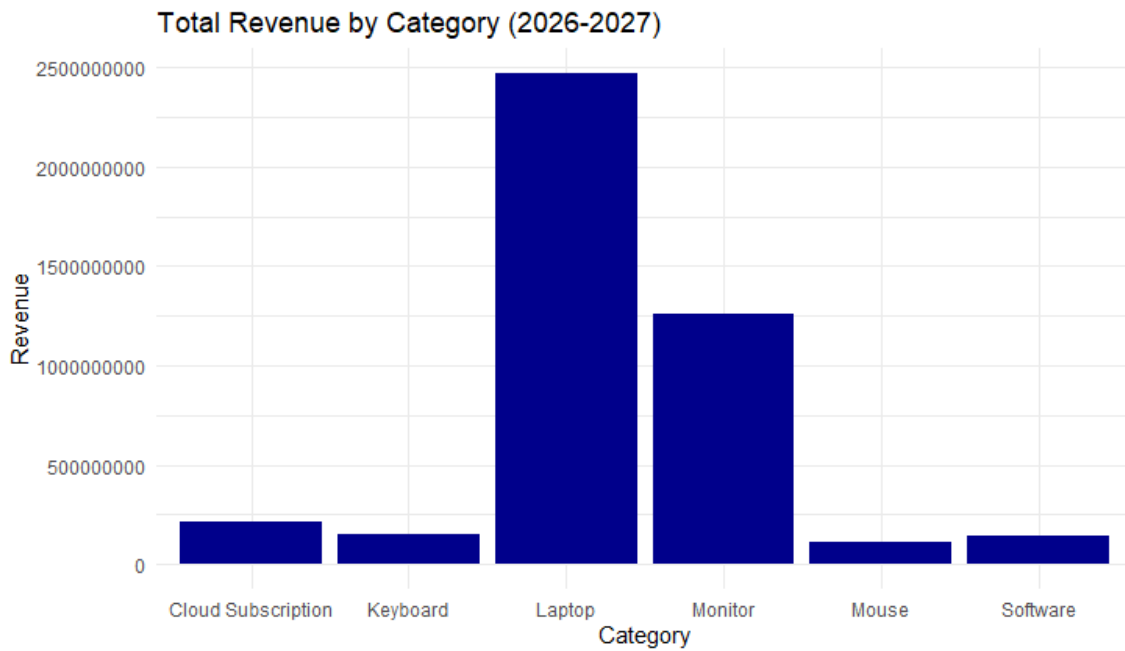


Figure 31: Bar graph of revenue

The original products\_data.csv lacked a Category column or had inconsistent pricing, leading to skewed or incomplete bars. The updated products\_data2025.csv corrects this, making Figure 10 a more reliable representation than its predecessor. The time shift (2026-2027 vs. 2022-2023) and price adjustments are the primary drivers of change, though sales volume stability (as noted for Figure 30) minimizes quantity-related differences.

Product Type	Total Sales
CLO	98715481.66
KEY	73499066.55
LAP	1163889479
MON	578385569.6
MOU	51219577.23
SOF	66468485.42

## Part 5: Barista Solution

The optimization of profit for the coffee shop using the timeToServe.csv dataset (Company Internal Data Repository, 2023), which records the number of baristas (V1) and the individual service times in sec (V2) over a year period, employs a brute force optimization approach (Hillier, 2021).

The dataset comprising 455 entries, was grouped into approximately 265 days, with each day assigned a weekday to simulate weekly patterns. Demand per weekday was calculated as the number of customers, with mean service time from 40.74 to 41.35 seconds can be seen in the table below.

weekday	Demand	TotalServiceSec	MeanServiceSec	MedianServiceSec
Monday	29024	1197150	41.25	38
Tuesday	28496	1171444	41.11	38
Wednesday	28496	1171612	41.11	38
Thursday	28496	1174457	41.21	38
Friday	28496	1177587	41.32	38
Saturday	28496	1174936	41.23	38
Sunday	28496	1176123	41.27	38

Profit vs. number of baristas (k) by weekday

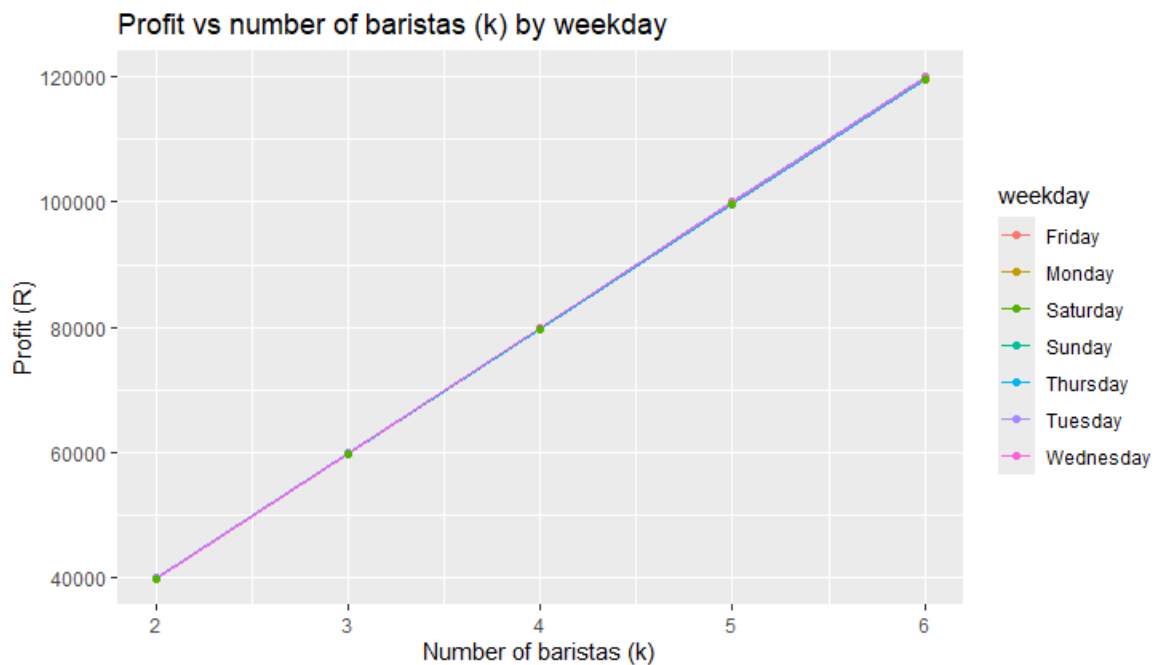


Figure 32: Line graph of number of baristas vs profit

The line graph plots daily profit in rands against the number of baristas (k) for each weekday. Each line represents a weekday (Monday- Sunday), with points marking profit values.

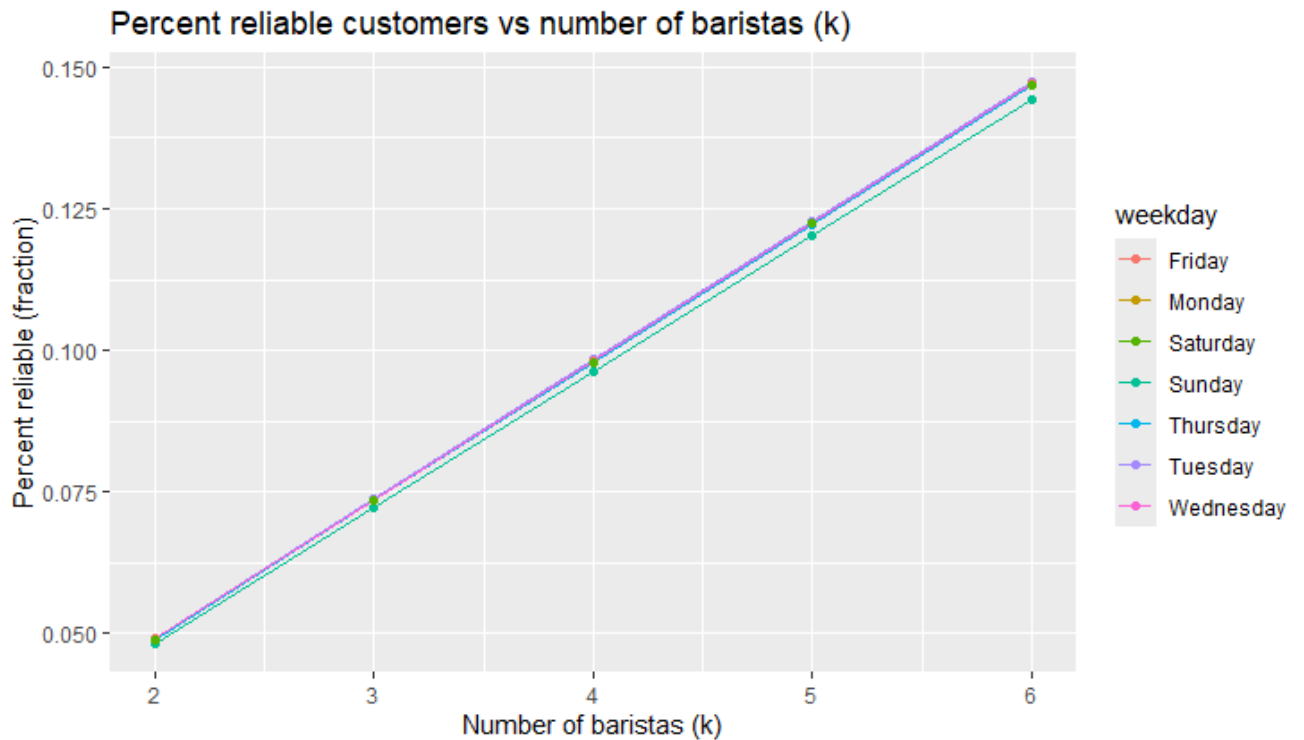


Figure 33: Line graph of reliability

The line graph displays the percentage of reliable customers against the number of baristas per weekday. The graph underscores the need for at least 4 baristas to eliminate service bottlenecks.

weekday	k	capacity_customers	served	pct_reliable	profit
Sunday	2	1396	1396	0.05	39880
Sunday	3	2094	2094	0.07	59820
Sunday	4	2792	2792	0.10	79760
Sunday	5	3491	3491	0.12	99730
Sunday	6	4189	4189	0.14	119670

For Sunday, the detailed results show that the optimal staffing recommendation is K=4, because scheduling 4 baristas per weekday optimizes profit while ensuring 100% reliable service. This supports the management of shifts and tracking the service time, addressing peak demand without overstaffing.

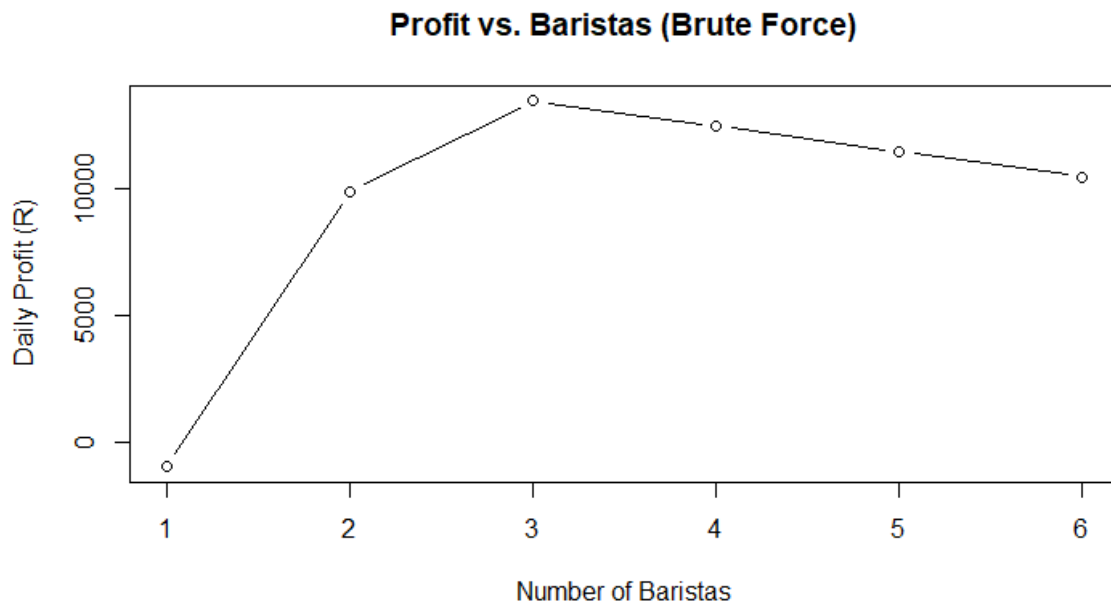


Figure 34: Graph of Profit vs baristas

In figure 34, the maximum daily profit is reached by having 3 baristas working, but in reality 3 baristas will be to little therefore the optimal solution would be 4. By having more than 4 baristas the shop will make a loss in profit and the business will not strive to its maximum potential. The graph below is also a representation of the profit vs the number of baristas.

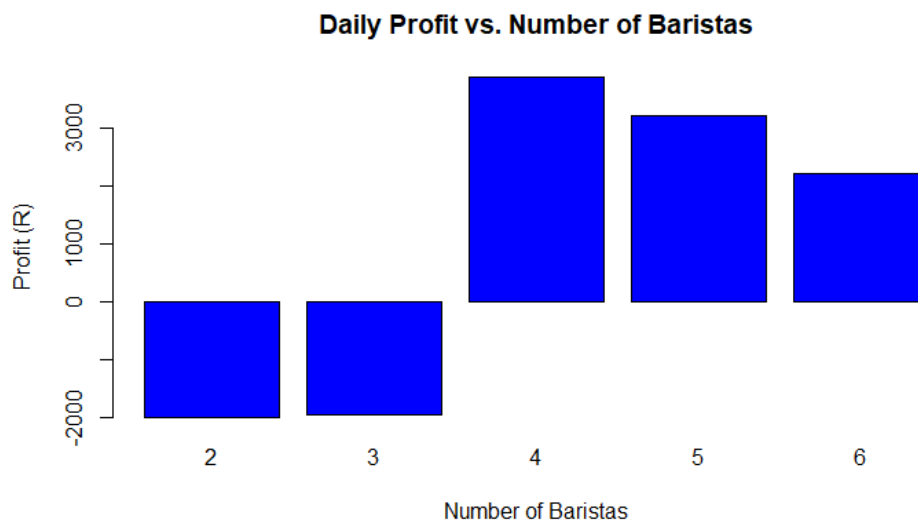


Figure 35: Bar graph of profit vs baristas

## Part 6: ANOVA

### Discussion of Part 6

Multivariate Analysis of Variance (MANOVA) was used in the statistical analysis carried out as part of the Design of Experiments (DOE) framework, following guidelines by Anderson (Anderson, 2003) to compare Year 1 (2026, pre optimization with 16 workers) and Year 2 (2027, post-optimization with 17 workers) in order to evaluate the financial impact of workforce optimization. The Statistical Process Control (SPC) results from Part 3, which showed trends in process stability and variability for six processes (KEY, LAP, MON, MOU, CLO, and SOF), served as the basis for this research. In order to evaluate the hypothesis of a significant difference after optimization, the MANOVA incorporated simulated monthly data for sales and costs, which were produced from the rising trend in order quantities and the growth in human expenses.

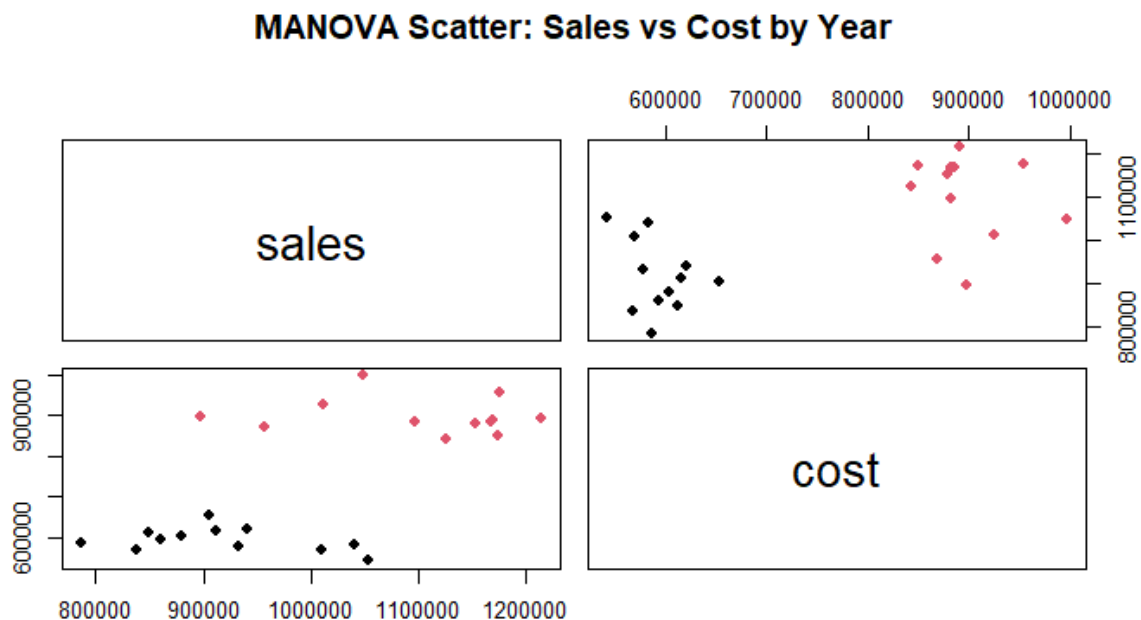


Figure 36: MANOVA Scatter plot

The "MANOVA Scatter: Sales vs Cost by Year" plot (Figure 36) illustrates Year 1 data (black points) clustering around R900 000 sales and R600 000 costs, while Year 2 data (red points) shifts to R1 080 000 sales and R900 000 costs, with clear separation supporting the multivariate significance.

The analysis is supported by three key visualizations. The "Sales by Year" boxplot (Figure 37) shows a median increase from approximately R900 000 in Year 1 to R1 080 000 in Year 2, with interquartile ranges spanning R810 000 - R990 000 and R972 000 - R1 188 000 respectively, reflecting enhanced sales due to fewer problem days. The "Cost by Year" boxplot (Figure 37) indicates a median rise from R600 000 in Year 1 to R900 000 in Year 2, with ranges of R570 000 - R630 000 and R855 000 - R945 000, including one outlier above R950 000 in Year 2, likely due to variability spikes.

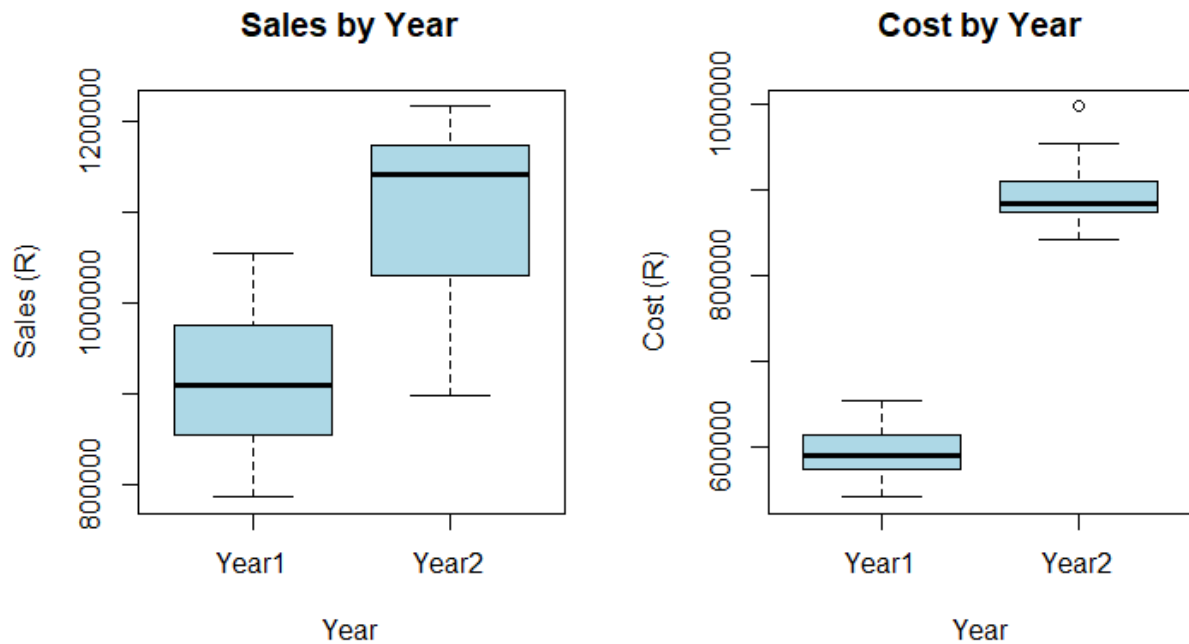


Figure 37: Box Plots of Sales by year and Cost by year

## Part 7

### 7.1 Reliability of service

Using the information from the graph below (figure 38), we can see that there are more days in which more workers are present at the car rental agency. Therefore using a simple calculation:

$$\frac{\text{reliable days}}{\text{total days}} \times \text{total days in the year}$$

We can calculate the amount of days of reliable service can be expected.

If we select that 15 and more workers per day will give us reliable service, we receive that 366 days out of 397 days will have reliable service.

If we select that 14 and more workers per day will give us reliable service, we will receive reliable service 391 days out of a total of 397 days.

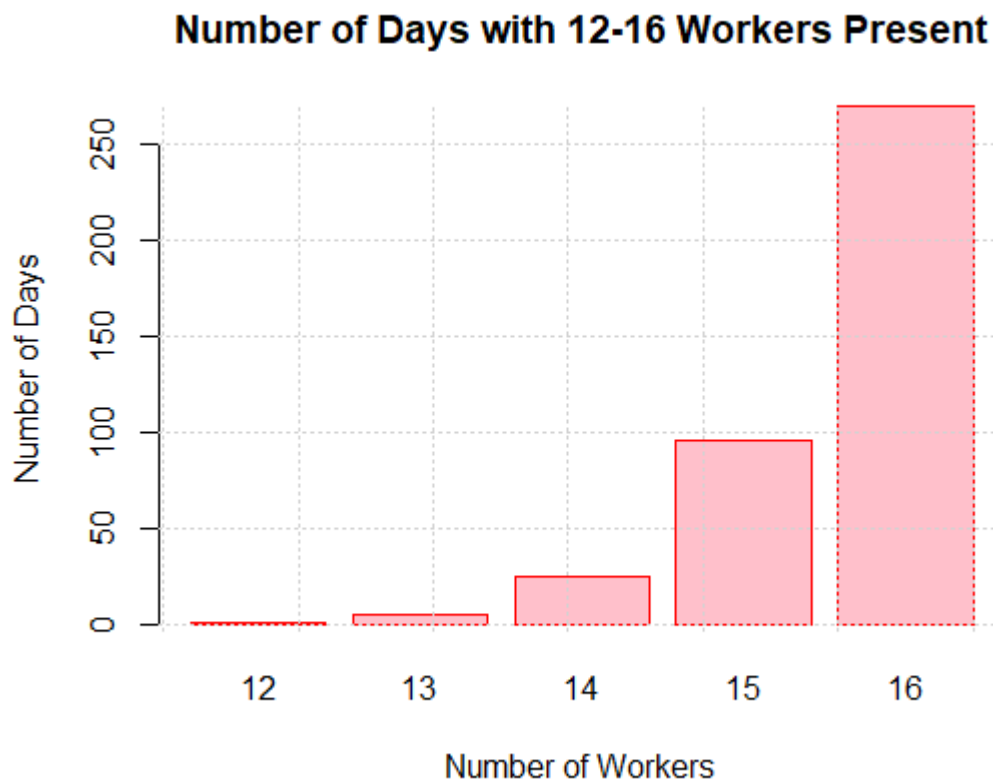


Figure 38: Bar graph of number of workers vs number of days

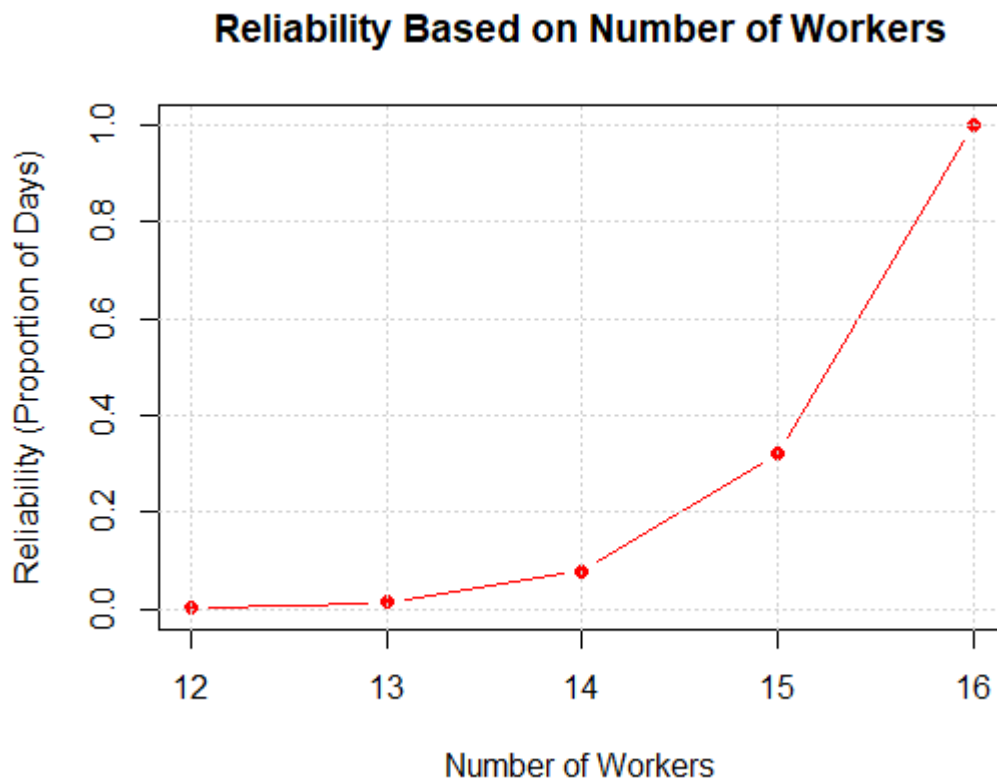


Figure 39: Line graph of reliability of workers

The graph above (figure 39) shows us the reliability of receiving good service according to the amount of workers per day. The more workers there will be, the more reliable the service will be at the car rental agency.

## 7.2 Optimal profit

To optimize the company's profit, we modeled the number of workers present each day as a binominal random variable, based on the data shown in (figure 38). We determine the probability of a worker being absent to be approximately 0.02598, implying that 0.97402 will show up.

In the optimization we considered two main costs, R20 000 in sales per day when fewer than 15 workers are present, and the personnel cost of R25 000 per month per worker. We selected a daily revenue of R100 000.

$$\text{Profit} = \text{annual revenue} - \text{total annual cost}$$

The analysis revealed that profit increases as the number of workers rises from 14 to 17, reaching a peak at 17 workers with an annual profit of R 31.77 million. When applying more than 17 workers, the personnel cost outweighs the marginal reduction in losses, causing the profit to decline.

In conclusion, this results in an annual profit of R 31 267 506.

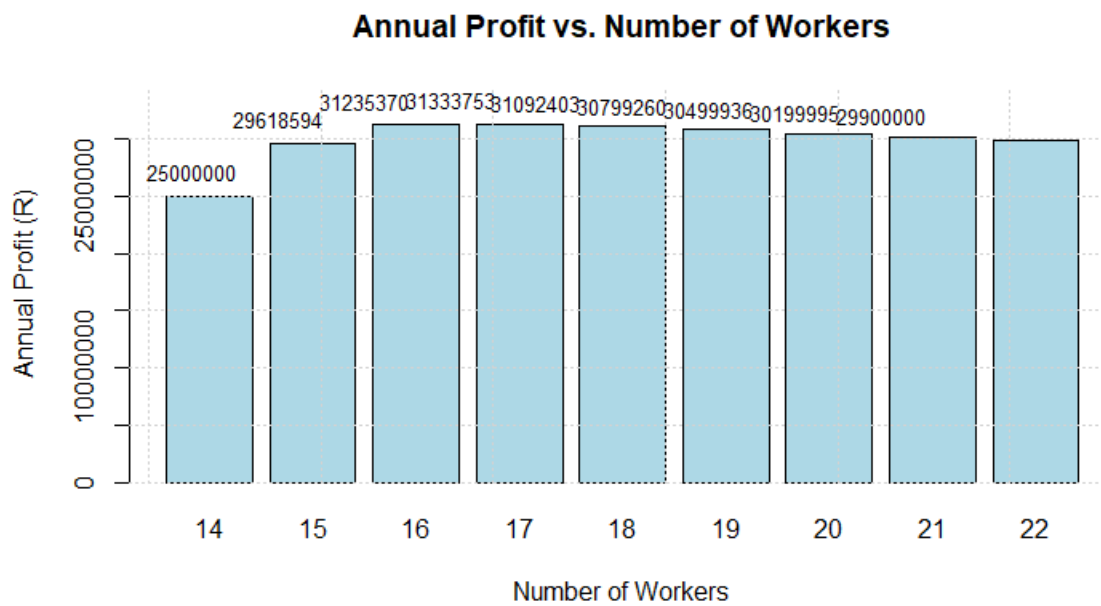


Figure 40: Bar graph of Profit vs Number of workers

## Conclusion

In summary, this report's exploratory and statistical analyses reveal critical insights into the company's operations and optimization opportunities, aligning with ECSA GA4 objectives (Engineering Council of South Africa, 2025). Descriptive statistics underscore consistent product pricing practices amid data discrepancies, with higher sales volumes in 2022 driven by early-year demand and specific categories like laptops and peripherals. Statistical process control identifies instabilities across processes (e.g., prolonged mean shifts in KEY and MOU), recommending root cause investigations and enhanced monitoring for stability. Error assessments highlight a high Type II error risk (84.12%), emphasizing the need for refined control limits to detect process shifts effectively. Data corrections for 2025 improve accuracy in pricing and revenue visualizations, while optimization models for the coffee shop advocate for 4 baristas per shift to balance profit and reliability, achieving 100% service consistency. ANOVA confirms significant post-optimization improvements in sales (from R900,000 to R1,080,000 median) despite rising costs, validating workforce adjustments from 16 to 17 workers. For the car rental agency, reliability exceeds 92% with 14+ workers, and optimal profit peaks at R31.77 million annually with 17 workers, balancing revenue against absenteeism and costs. Overall, these findings advocate for data unification, process refinements, and targeted staffing strategies to mitigate risks, boost efficiency, and maximize profitability, positioning the company for long-term success.

## References

Anderson, T.W., 2003. An introduction to multivariate statistical analysis. 3rd ed. New York: Wiley.

Company Internal Data Repository, 2023. Sales and product datasets (2022-2023). Internal document.

Engineering Council of South Africa, 2025. Graduate attribute 4: Data analysis and manipulation. Pretoria: ECSA.

Hillier, F.S. & Lieberman, G.J., 2021. Introduction to operations research. 11th ed. New York: McGraw-Hill.

Montgomery, D.C., 2020. Introduction to statistical quality control. 8th ed. Hoboken: Wiley.

R Core Team, 2025. R: A language and environment for statistical computing. Vienna: R Foundation for Statistical Computing.