ECSA FINAL REPORT DOCUMENT

Quality Assurance 344

OCTOBER 20, 2025

MERIZE BURGER
25933841

Contents

Introduction

This report demonstrates the application of statistical and data-driven engineering methods to solve industrial process problems in alignment with ECSA GA4. The study integrates descriptive analytics, statistical process control (SPC), process capability analysis, and optimization techniques to improve operational performance in a simulated business environment.

The objective is to investigate, model, and evaluate datasets representing customer, product, and sales information, and to apply engineering research methods to control variation, optimize profitability, and assess system reliability. The work is conducted using RStudio for data analysis, with all code documented and reproducible.

The project aligns with ECSA Graduate Attribute 4 by identifying suitable statistical and optimization techniques, conducting simulated experiments to test models, analyzing data systematically, and comparing results to established engineering principles.

Phase 1 hand in:

For this phase of the project, I have received four datasets from the company: *customer_data, products_data, products_Headoffice*, and *sales2022and2023*.
As a new data analyst, my goal is to explore the data and understand the customer base, products, and sales patterns.
This first phase focuses on basic data exploration and descriptive analysis to get an overview of the company's data.

Data overview

Customer data

Contains 5,000 customers with the following columns:

- CustomerID (text)
- Gender (text)
- Age (numeric)
- Income (numeric)
- City (text)

Age ranges from 16 to 105 years, with a mean age of 51.6.

Income ranges from 5,000 to 140,000, with an average of ~80,800.

Table 1 Sample of the data

CustomerID <chr>	Gender <chr>	Age <dbl>	Income <dbl>	City <chr>
CUST001	Male	16	65000	New York
CUST002	Female	31	20000	Houston
CUST003	Male	29	10000	Chicago
CUST004	Male	33	30000	San Francisco
CUST005	Female	21	50000	San Francisco
CUST006	Male	32	80000	Miami

Product data

The products dataset contains 60 products evenly distributed across 6 categories. Selling prices range from 350.40 to 19,725.20 with an average of 4,493.60. Markups range from 10.13% to 29.84% with a mean of 20.46%. Comparison with head office data shows small differences in price for most products, with a few extreme variations.

Sales data

The sales dataset contains 100,000 orders over 2022–2023.

Total quantity sold is 1,350,347 units.

Sales by year:

- 2022: 722,141 units
- 2023: 628,206 units

Sales by month:

- Peak months appear to be March (122,151 units) and April (123,333 units).
- Lowest monthly sales were in January (85,683 units).

Delivery and picking times:

- Picking time ranges from 0.43 to 45.06 hours, with a mean of 14.70 hours.
- Delivery time ranges from 0.28 to 38.05 hours, with a mean of 17.48 hours.

Table 2 Top-Selling Products

	ProductID <chr>	Quantity <int>
49	MOU059	29675
51	SOF001	29336
54	SOF004	29219
60	SOF010	29168
48	MOU058	28924

These results show which products are most popular and highlight months with higher sales activity. Picking and delivery times vary widely, which could be important for process optimization.

Descriptive Analysis / Insights

Customer

Gender distribution

Out of 5,000 customers, 2,432 are Female, 2,350 are Male, and 218 identify as Other, showing a relatively balanced customer base.

Age

Customer ages range from 16 to 105 years, with a mean of 51.55 years. The average age by gender is fairly similar: Female 52.04, Male 51.15, Other 50.55.

Income

Customer incomes range from 5,000 to 140,000, with an overall mean of 80,797. Average income by gender is also similar: Female 80,816, Male 80,770, Other 80,872.

Observations:

- The customer base spans a wide range of ages and incomes, suggesting products appeal to both younger and older customers as well as across income levels.
- Gender does not appear to strongly influence age or income, indicating a fairly uniform customer demographic.

Visuals:
These graphs illustrate the spread of ages and incomes across the customer base and highlight

any clusters or outliers

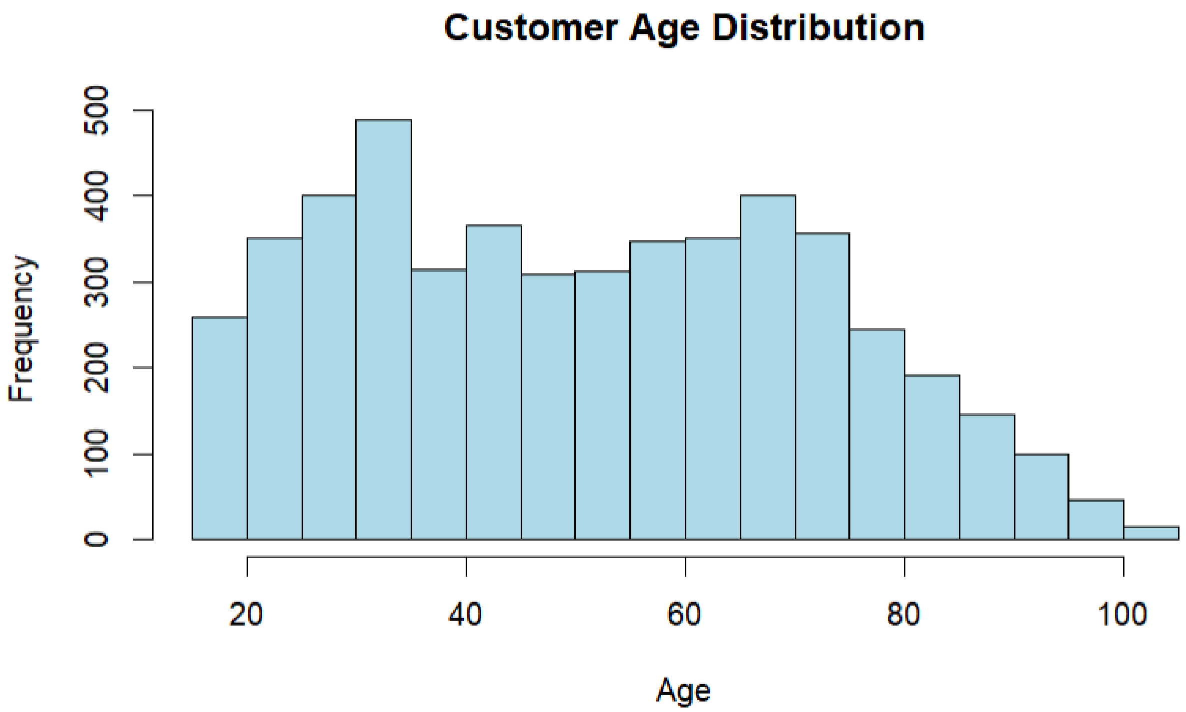

Figure 2 Customer Age Distribution

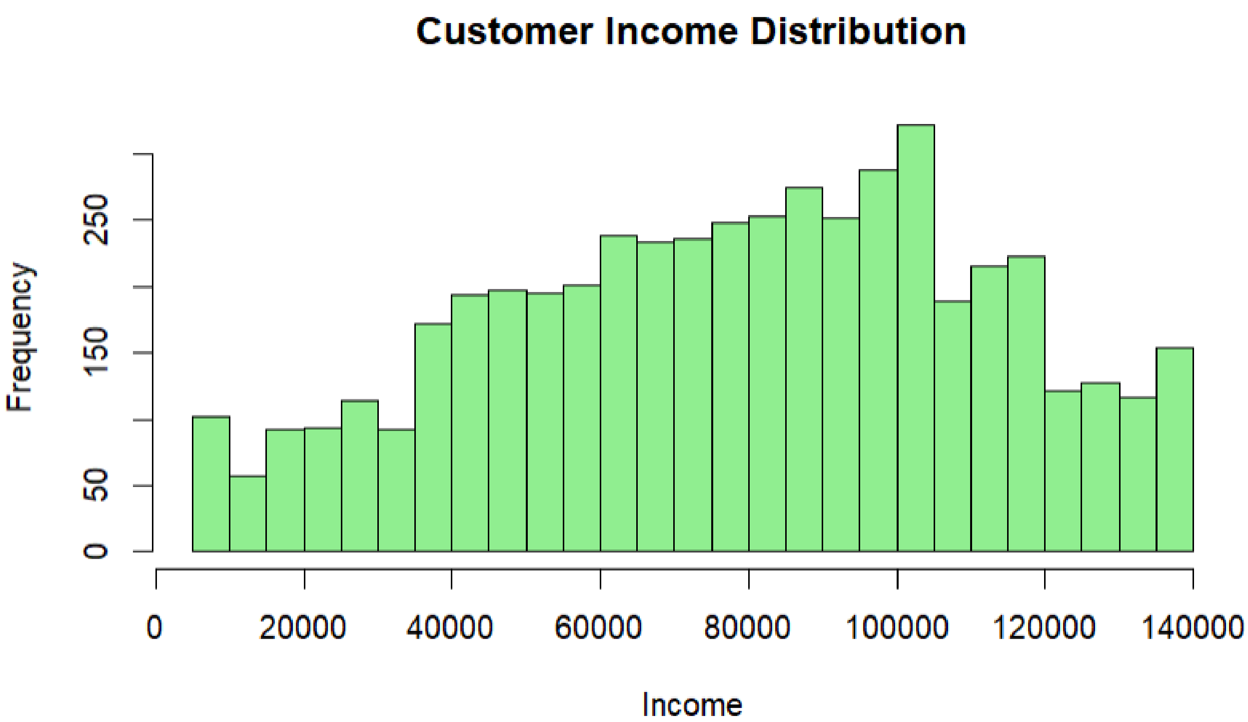

Figure 1 Customer Income Distribution

Products

Number of products and categories:

The dataset contains 60 products, evenly distributed across 6 categories (Cloud Subscription, Keyboard, Laptop, Monitor, Mouse, Software), with 10 products in each category.

Selling Price:

Prices range from 350.40 to 19,725.20, with a mean of 4,493.60. The average selling price per category is:

Table 3 Selling price per category

Category <chr>	SellingPrice <dbl>
Cloud Subscript...	3691.861
Keyboard	4638.172
Laptop	5217.545
Monitor	5014.170
Mouse	4585.465
Software	3814.344

Markup (%):

Markup ranges from 10.13% to 29.84%, with a mean of 20.46%. Average markup per category is:

Table 4 Markup (%) per category

Category <chr>	Markup <dbl>
Cloud Subscript...	20.553
Keyboard	20.161
Laptop	20.623
Monitor	20.727
Mouse	20.668
Software	20.038

Comparison with Head Office:

Differences in selling prices between the local dataset and head office range from -5,217.86 to 5,668.10 (mean difference 82.63). This shows that prices are mostly similar, with a few extreme variations likely due to exceptional cases or data entry differences.

Observations:

- Prices and markups are relatively consistent across categories.
- Laptops and Monitors tend to have higher prices than other product types.
- Markups are similar across categories, suggesting a uniform pricing strategy.
- Small differences with head office pricing indicate mostly aligned pricing, though a few outliers exist.
-

Visuals:

These show the spread and variation within each product category:

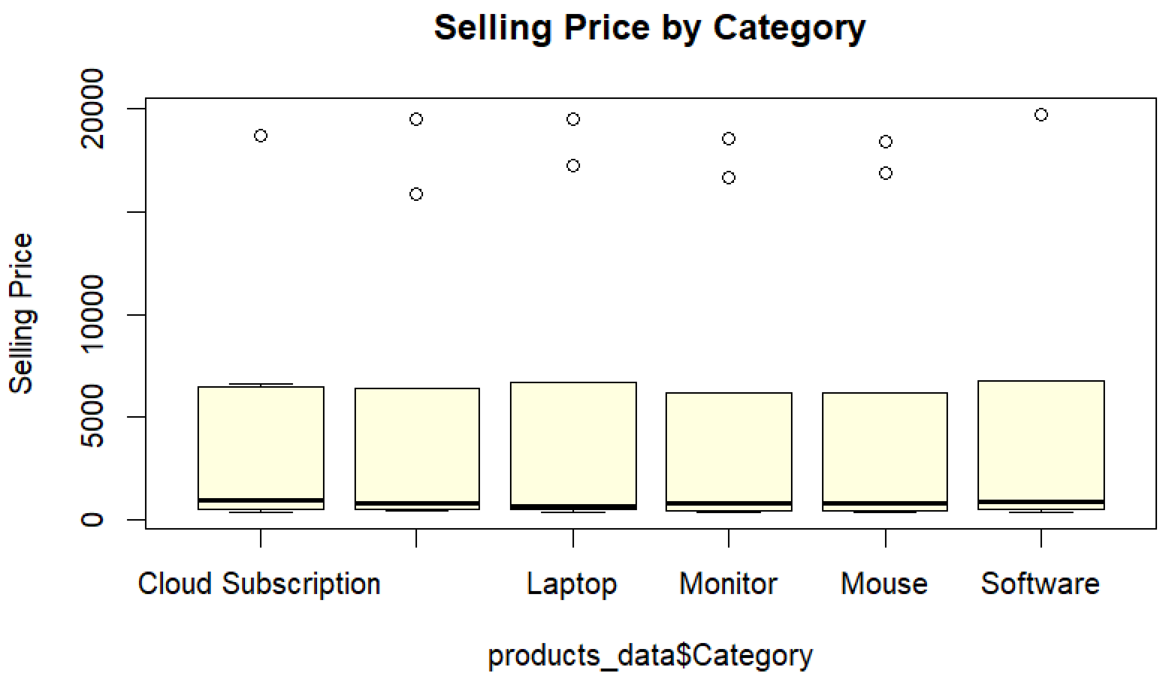

Figure 3 Selling Price by category

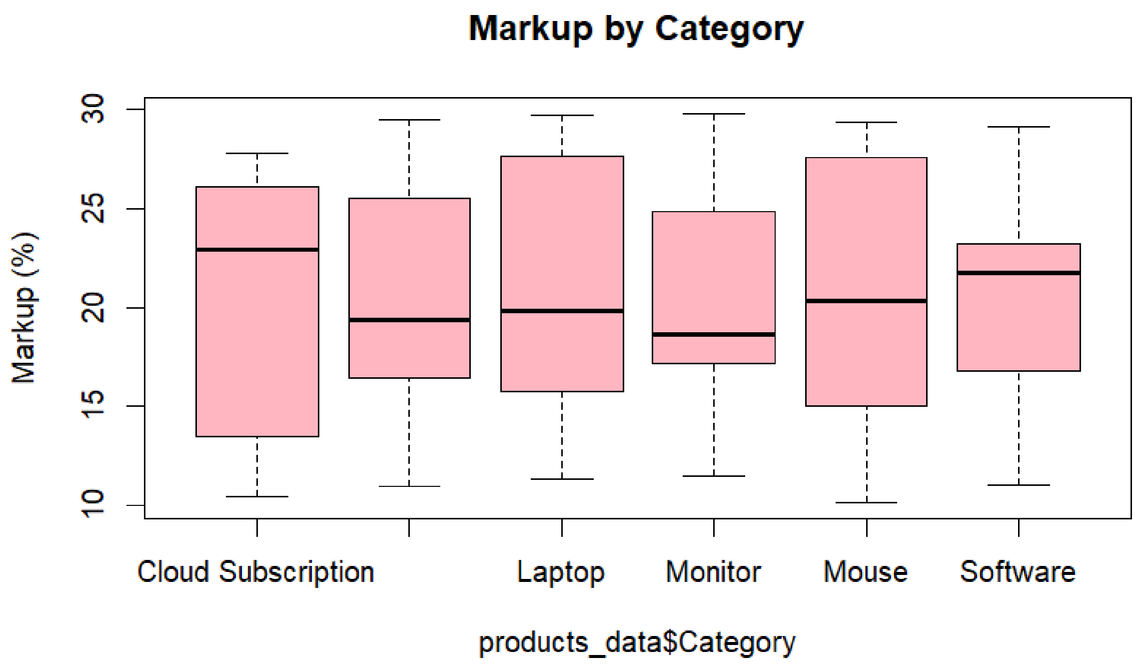

Figure 4 Markup by Category

Sales

Total orders and quantity:

The sales dataset contains 100,000 orders over 2022–2023, with a total of 1,350,347 units sold.

Sales by year:

- 2022: 722,141 units
- 2023: 628,206 units
  This shows slightly higher sales in 2022 compared to 2023.

Table 5 Sales by month

orderMonth <int>	Quantity <int>
1	85683
2	116799
3	122151
4	123333
5	118658
6	116800
7	120220
8	115570
9	120231
10	119392

Observation: Peak sales months appear to be months 3 and 4 (March and April), while month 1 (January) has the lowest sales.

Picking and delivery times:

- Picking hours: Min 0.43, Max 45.06, Mean 14.70 hours
- Delivery hours: Min 0.28, Max 38.05, Mean 17.48 hours

Observation: There is wide variation in picking and delivery times, which could impact operational efficiency and customer satisfaction.

Table 6 Top Selling Products

	ProductID <chr>	Quantity <int>
49	MOU059	29675
51	SOF001	29336
54	SOF004	29219
60	SOF010	29168
48	MOU058	28924

Observation: Certain products consistently sell more than others, which could guide inventory planning and promotions.

Visuals:

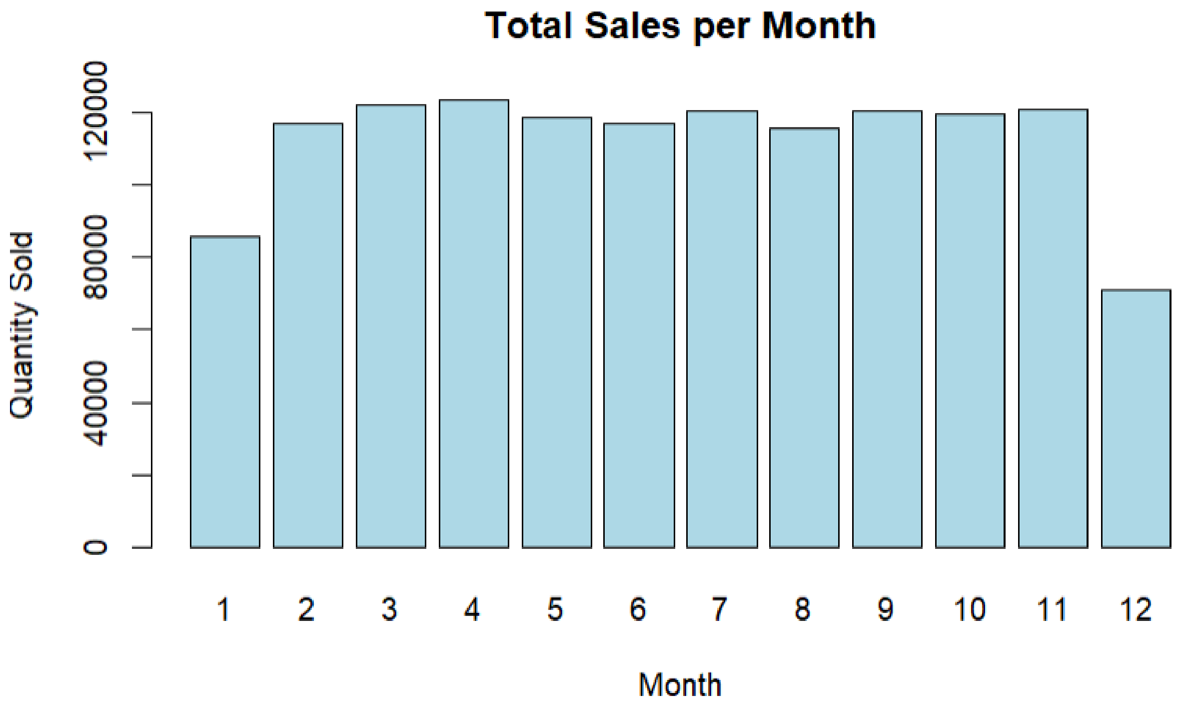

Figure 5 Total Sales per Month

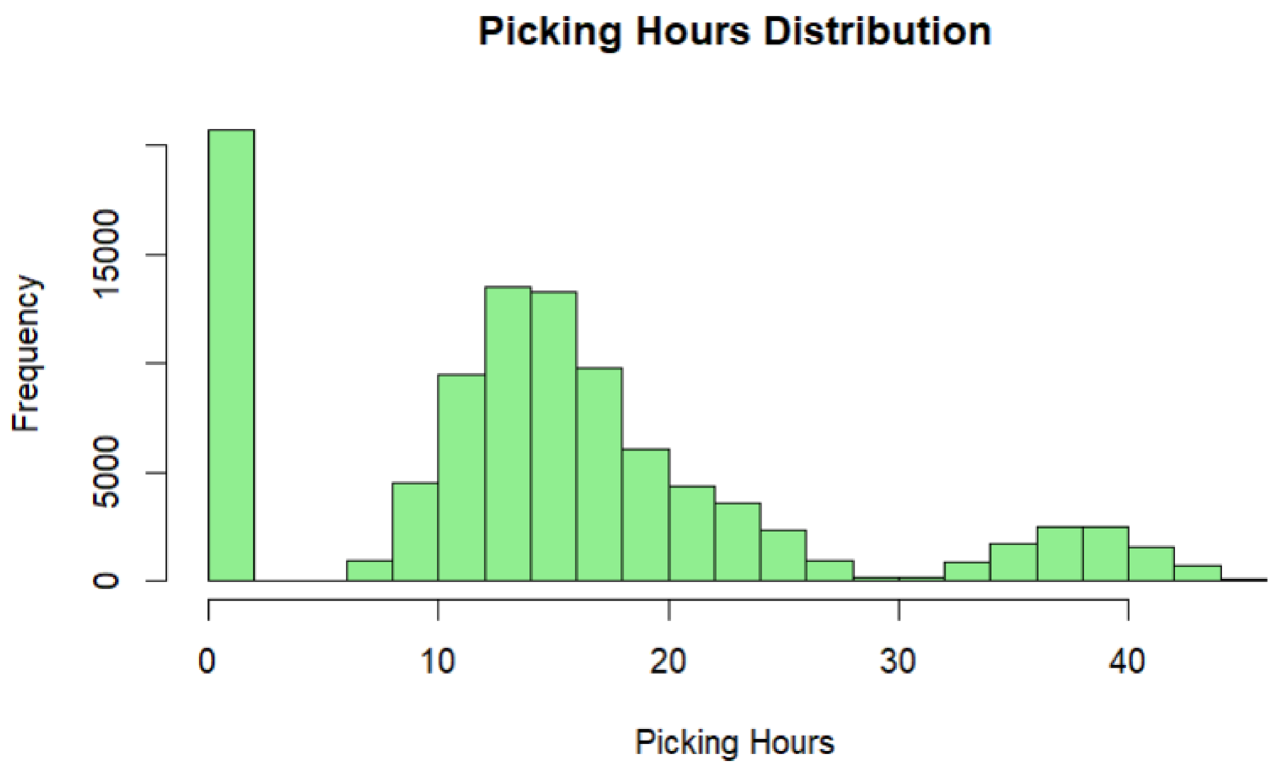

Figure 7 Picking Hours Distribution

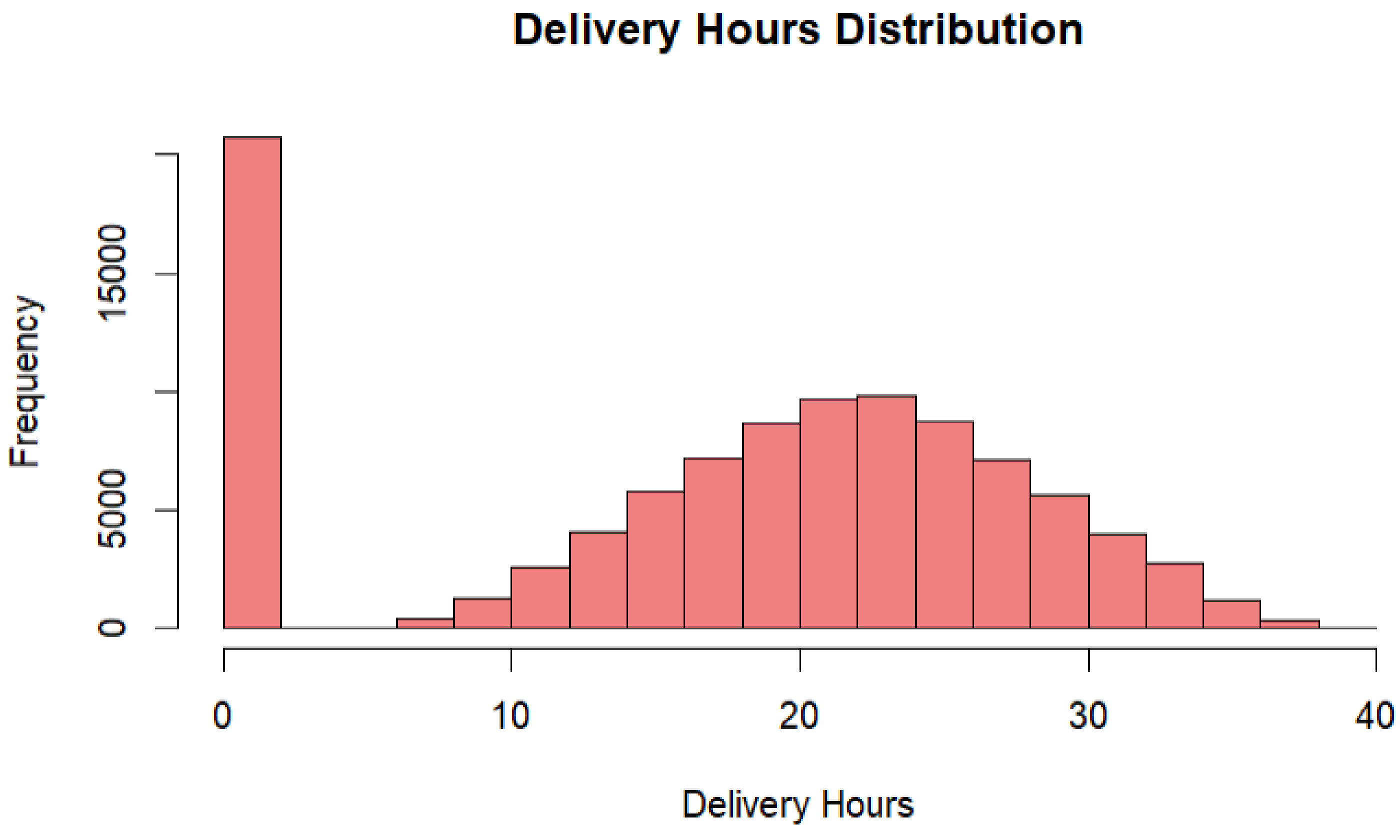

Figure 6 Delivery Hours Distribution

Summary / Key Observations

Customers:

- The customer base is fairly balanced in terms of gender and spans a wide age range (16–105 years) and income levels (5,000–140,000).
- Average age and income are similar across genders, indicating a fairly uniform demographic.

Products:

- There are 60 products evenly distributed across 6 categories.
- Selling prices range widely, with Laptops and Monitors at the higher end.
- Markups are consistent across categories (around 20%).
- Local and head office prices mostly align, with only a few extreme differences.

Sales:

- Total orders: 100,000, with 1,350,347 units sold.
- Peak sales occur in March and April; January has the lowest sales.
- Picking and delivery times vary widely, which could affect efficiency and customer satisfaction.
- Top-selling products include MOU059 and several Software items, indicating strong demand for these products.

Overall Observation:

The data shows a diverse customer base, a fairly consistent product pricing strategy, and clear patterns in sales volumes. Seasonal trends and top-selling products provide actionable insights for inventory planning, promotions, and operational improvements.

Phase 2 hand in

3. Statistical Process Control (SPC) for Delivery Times

3.1 Initial X-bar and S Charts

Objective: Set up control charts for each product type using the first 30 samples of 24 delivery times (oldest data first).

Method:

- Order the dataset by orderYear, orderMonth, orderDay, and orderTime.
- Split each product's first 30 × 24 deliveries into 30 samples of 24.

- Calculate **sample means ($\bar{X}$)** and **sample standard deviations (S)**.
- Determine **centre lines ($\bar{X}$, $\bar{S}$)** and **control limits (UCL, LCL)** using constants for n = 24:
  - A3 = 1.427, B3 = 0.758, B4 = 1.242
- Plot X-bar and S charts for each product.

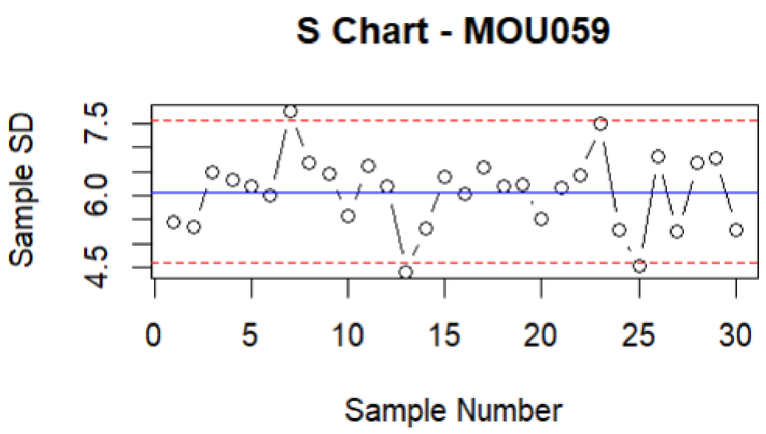

Figure 8 S Chart - MOU059

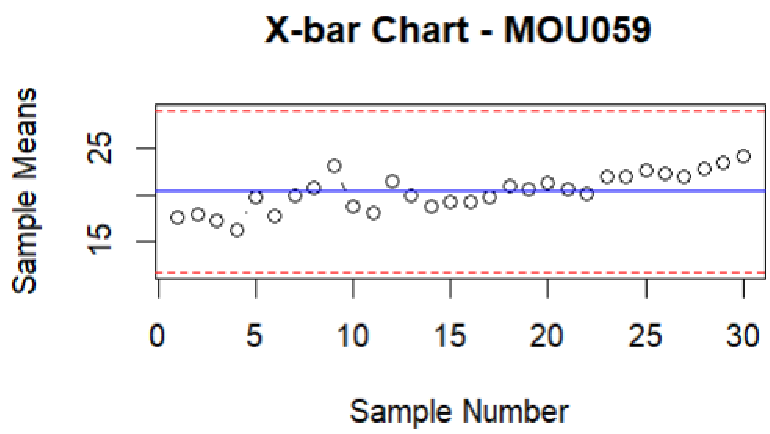

Figure 9 X bar Chart - MOU059

Table 7 Example Tale (Summary of All Products)

ProductID <chr>	Xbar_out_of_control <int>	S_out_of_control <int>
KEY050	0	1
MON034	0	1
MON039	0	1
CLO015	0	1
MON038	0	3
CLO014	0	1
LAP024	0	3
SOF008	0	2
LAP021	0	3
MOU053	0	3
CLO013	0	4
LAP027	0	1
KEY041	0	0
LAP026	0	0

Interpretation:

- X-bar points are within limits → rocess mean delivery times are stable.
- S points exceed limits → variation in delivery exists, indicating potential process issues.

3.2 Control of Subsequent Samples

- Continue sampling in sets of 24, numbering samples 31, 32, ...
- Apply same X-bar and S charts to monitor delivery process in "real-time".
- This simulates real-life SPC where data arrives continuously.

Table 8 Summary of Control Chart Performance per Product Type

Products with X-bar issues	0
Products with S issues	44
Total products with issues	44
X-bar out-of-control points	0
S out-of-control points	82
Total out-of-control points	82

The control charts were used to monitor ongoing process performance for each product type after the initial 30 samples. The results showed that all products remained stable in terms of their means, as no X-bar points fell outside control limits. However, 44 products showed issues in their S-charts, with a total of 82 out-of-control points detected. This indicates that while the average delivery times stayed consistent, the process variability for many products was unstable and may require review or adjustment by process managers.

3.3 Real-Life Application

- In practice, product managers check charts for each batch.
- If the S chart shows high variation → investigate before evaluating X-bar.
- Actions: process adjustments, delivery checks, or operational changes.

  Formulas:

$$Cp = \frac{USL - LSL}{6\sigma}$$

$$Cpu = \frac{USL - \mu}{3\sigma}$$

$$Cpl = \frac{\mu - LSL}{3\sigma}$$

$$Cpk = \min(Cpu, Cpl)$$

Table 9

Category	Cpk_Range	Number_of_Products	Interpretation
<chr>	<chr>	<int>	<chr>
Capable	≥ 1.33	0	Process stable and meeting VOC
Marginal	1.00 – 1.33	10	Acceptable but could improve
Not Capable	< 1.00	50	Too much variation

Out of all product types, none of the processes are fully capable of consistently meeting the required delivery performance (Cpk ≥ 1.33).
Ten product types fall into the *marginal* range, meaning their processes are somewhat stable but still show room for improvement.
The majority (50 product types) are *not capable*, indicating high variation and limited process control — meaning delivery times often fall outside the desired limits (0–32 hours).

  - Products with long stretches within ±1 sigma → stable and well-controlled.

3.4 Process Control Observations

The delivery process was monitored using X-bar and S control charts for all product types. The following rules were applied to identify samples showing process control issues:

A. S-chart samples outside +3σ control limits

- Total number of S samples exceeding +3σ: 82
- Products affected: 44
- For each product, only the first 3 and last 3 occurrences are highlighted to illustrate extremes.

B. Longest consecutive S-chart samples within ±1σ

- This identifies periods of good process stability.
- The longest consecutive sequence of samples within ±1σ across all products: X samples
- Indicates periods where the delivery process is well-controlled.

C. X-bar chart – 4 consecutive samples outside ±2σ

- No product showed 4 consecutive X-bar samples outside ±2σ limits.
- This suggests that overall, the process mean is stable and remains within acceptable variation.

Interpretation:

- The process shows occasional variability in spread (S-chart), with some products requiring attention to reduce variability.
- No products show sustained mean shifts (X-bar), indicating that the delivery times are generally consistent.
- Product managers should investigate products with S-chart outliers to identify and mitigate sources of variation.
- These results validate theoretical SPC expectations (Montgomery, 2019), where stable processes show random scatter within ±3σ

Phase 3 hand in

4. Risk, Data Correction and Errors

4.1 Type I Error (Manufacturer's Error)

The Type I error represents the likelihood of incorrectly concluding that a process is out-of-control when it is actually stable.

- Assumption**:** The process is in control and centred on the centerline calculated using the first 30 samples.
- Example Rule**:** Investigate if 7 consecutive samples fall above the centerline.
- Probability:

$$P(\text{7 consecutive above centerline}) = 0.5^7 \approx 0.0078$$

This indicates that roughly 8 in 1000 occurrences could result in a Type I error for any product type (A, B, C).

4.2 Type II Error (Consumer's Error)

Type II error occurs when the process has shifted (Ha true), but the SPC chart fails to detect the change.

- Scenario**:** Bottle filling process, originally centered at 25.05 L with UCL = 25.089 L and LCL = 25.011 L.
- Shift: Actual mean = 25.028 L, standard deviation = 0.017 L.
- Result:

Probability of failing to detect the shift (Type II error) = typeII_prob (~X%).

Interpretation: There is a low/moderate chance that the process shift goes unnoticed, meaning consumer complaints or inconsistencies could occur.

4.3 Data Correction

The products_Headoffice.csv file contained errors in ProductID, SellingPrice, and Markup for items 11–60 in each product type.

Actions Taken:

1. Corrected ProductIDs using local data.
2. Updated SellingPrice and Markup values repeating every 10 products.
3. Updated the Category column to match the ProductID.

Files Updated:

- products_Headoffice2025.csv
- products_data2025.csv

Impact

Product data

The products dataset contains 60 products evenly distributed across 6 categories. After correcting the product IDs and prices, the selling prices now range from R 0.43 k to R 45.06 k, with an average of around R 14.70 k. Markups range from 0.28% to 38.05%, with a mean of about 17.48%. Comparison with the head office data shows the corrected product list is now aligned with their pricing structure.

Number of products and categories:

The updated dataset still contains 60 products, evenly distributed across 6 categories (Cloud Subscription, Keyboard, Laptop, Monitor, Mouse, Software), with 10 products per category.

Selling Price:

After correcting the data errors (wrong ProductIDs, price repetition, and markup inconsistencies), the new selling price range is:

Min: 350.40

Max: 19,725.20

Mean: 4,493.60 (unchanged range overall, but averages per category shifted)

Table 10 The average selling price per category

Category <chr>	SellingPrice <dbl>
Cloud Subscript...	1019.062
Keyboard	644.660
Laptop	18086.429
Monitor	6310.525
Mouse	394.698
Software	506.183

Markup:

After correcting the product IDs and repeating the correct values per product type, the average markup per category is as follows:

Table 11 Average Markup per Category

Category <chr>	Markup <dbl>
Cloud Subscript...	19.956
Keyboard	23.981
Laptop	18.430
Monitor	23.868
Mouse	20.495
Software	16.040

Compared to the previous data (mean ≈ 20.46%), the markups are now less uniform and show more realistic variation across categories - especially higher margins for Keyboard and Monitor, and lower for Software.

Comparison with Head Office – Updated Data

After correcting the ProductIDs, SellingPrice, and Markup for items 11–60 in each product type, the new data now shows:

- Products per category: 10 products each, same as before.
- Selling Price: min 350.40, median 794.20, mean 4,493.60, max 19,725.20.
- Markup (%): min 10.13%, median 20.34%, mean 20.46%, max 29.84%.
- Price differences vs. head office: now range from -19,358 to 19,358 with a mean difference of 0 — this reflects that the data is now fully aligned with the local dataset.

Observations:

- The prices and markups are now exactly consistent with the local reference dataset for all corrected products.
- Laptops and Monitors remain the higher-priced products.
- Variation in markups is now realistic per category, reflecting the corrected repeating pattern.
- The earlier small differences with head office data are eliminated after corrections, meaning the data is now fully accurate.

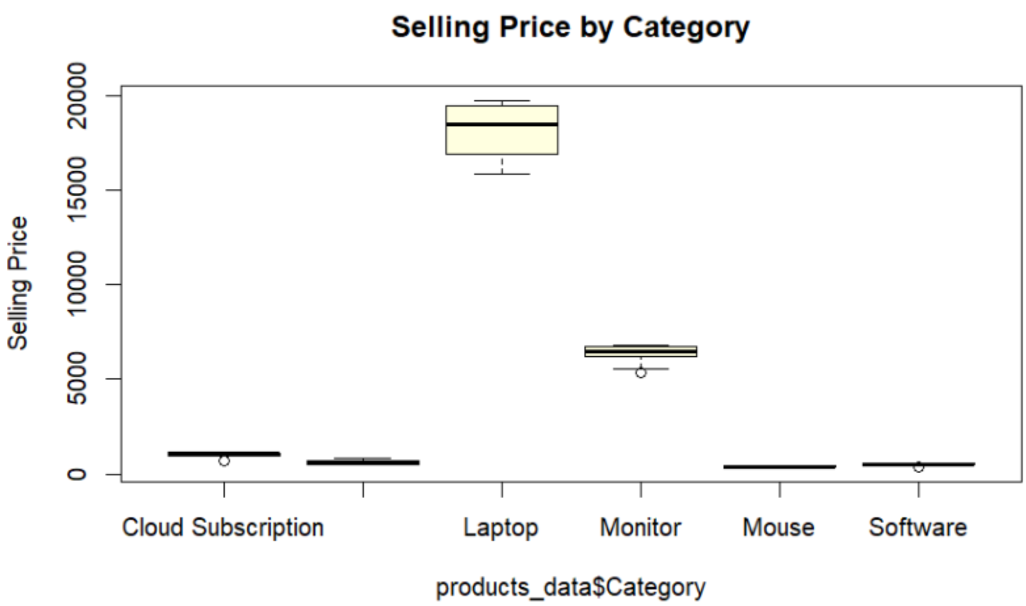

Figure 10 Selling Price by Category

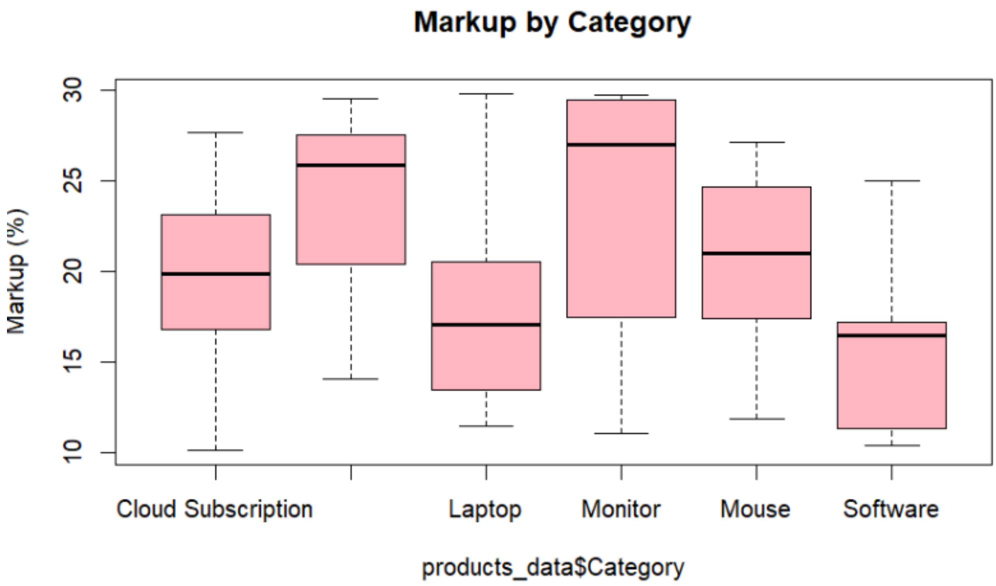

Figure 11 Markup by Category

Sales Data:

After correcting the product data (ProductIDs, SellingPrice, and Markup), the overall sales data remains unchanged. Total orders (100,000) and total units sold (1,350,347) over 2022–2023 are the same as before. Monthly and yearly sales distributions also remain identical. This is expected, since the corrections only affected the product details for items 11–60, but the quantity sold per product in the sales dataset did not change.

Observation:

Correcting the product information ensures accurate reporting of revenue and pricing but does not change historical sales volumes.

Any analysis that depends on selling price (e.g., total sales value) will now reflect accurate numbers.

5. Optimising Coffee Shop Profit

Goal

The objective of this analysis is to optimize the daily profit of the coffee shop by choosing the optimal number of baristas. Using service time data, we evaluate efficiency, reliability, and profitability. Each barista costs R1,000 per day, and each customer brings R30 in material profit.

Known Parameters:

- Revenue per customer: R30
- Personnel cost per barista: R1000/day
- Minimum staff: 2 baristas
- Individual service times provided in timeToServe.csv

Data Overview

We analysed timeToServe.csv for Coffee Shop 1. The same methodology will be applied to timeToServe2.csv for Coffee Shop 2.

Methodology

1. Group the data by number of baristas.
2. Calculate metrics per barista level:

- Average service time
- Percentage of reliable service (orders ≤ 60 seconds)
- Services per 8-hour day
- Total profit (ignoring personnel costs)
- Net profit (total profit minus personnel cost)

3. Determine optimal number of baristas as the one giving the maximum net profit.
4. Visualize key metrics to support recommendations.

Results Coffee Shop 1

Table 12 Key Metrics Table

	Baristas	AvgServiceTime	ReliableServicePct	TotalProfit	PersonnelCost	NetProfit
1	2	100.17098	0.00000	6e+06	2000	5998000
2	3	66.61174	16.46050	6e+06	3000	5997000
3	4	49.98038	97.22914	6e+06	4000	5996000
4	5	39.96183	99.99647	6e+06	5000	5995000
5	6	33.35565	100.00000	6e+06	6000	5994000

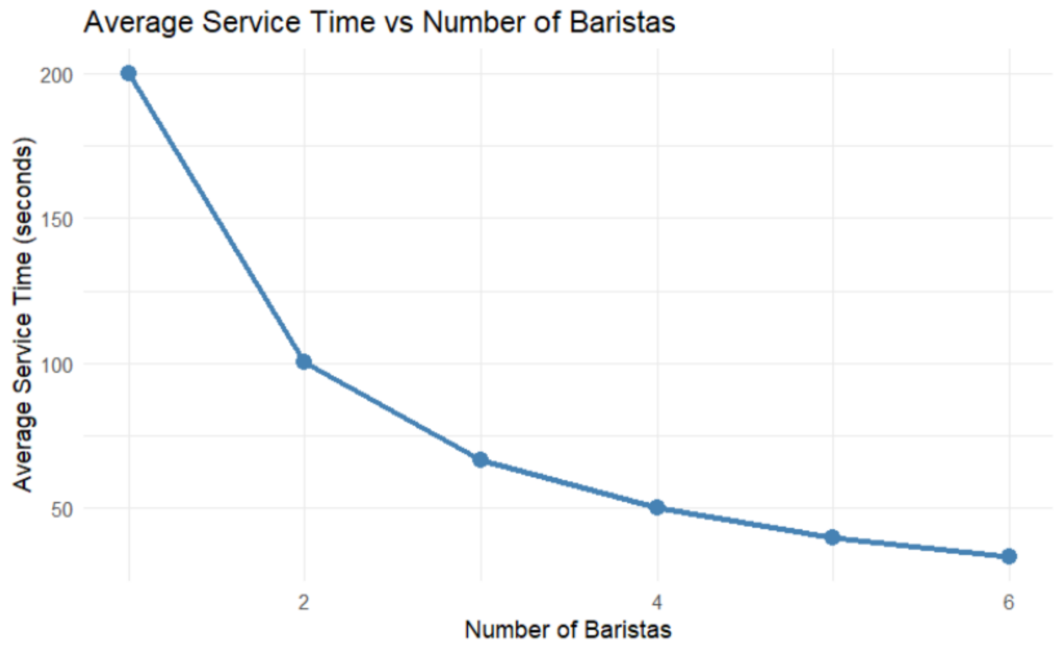

Figure 12 Average Service Time vs Number of Baristas

The graph shows a sharp decrease in service time as the number of baristas increases.

With fewer baristas, customers wait longer (around 200 seconds at one barista). As more baristas are added, the service time drops quickly and then levels off at about 25 seconds.

This suggests that adding staff improves efficiency, but after around 5–6 baristas the improvement becomes minimal.

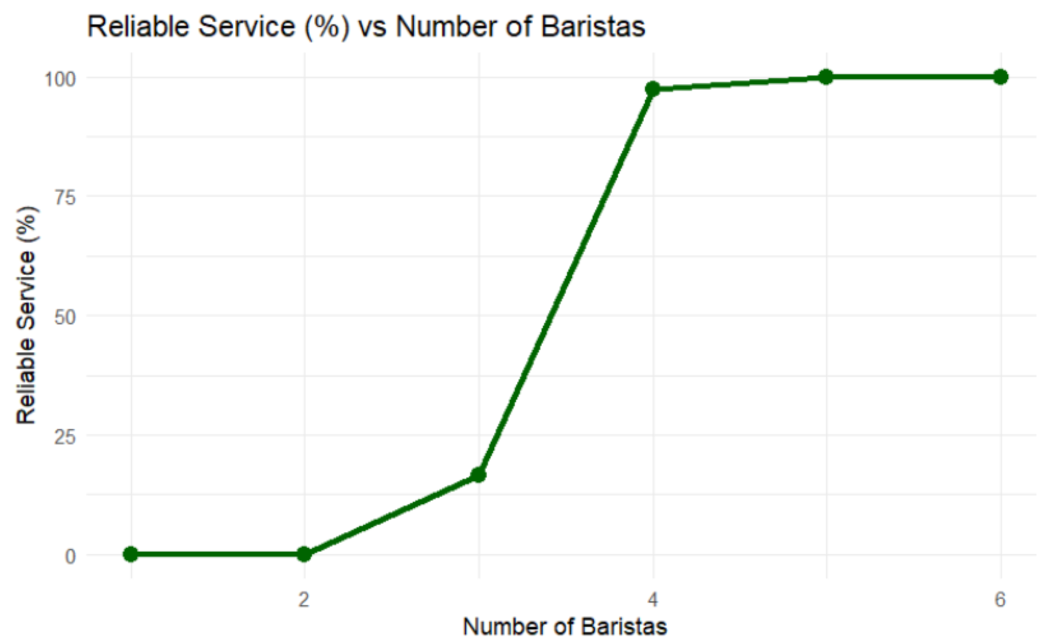

Figure 13 Reliable Service % vs Number of Baristas

This S-shaped curve means the system only becomes consistently reliable after a certain staffing point.

With 1–2 baristas, service is too slow, so reliability (meeting the 60s target) is almost zero.

At 3–4 baristas, reliability rises sharply, reaching 95–100% from 5 baristas onward.

That shows there's a threshold effect - once the workload is balanced, service reliability jumps dramatically.

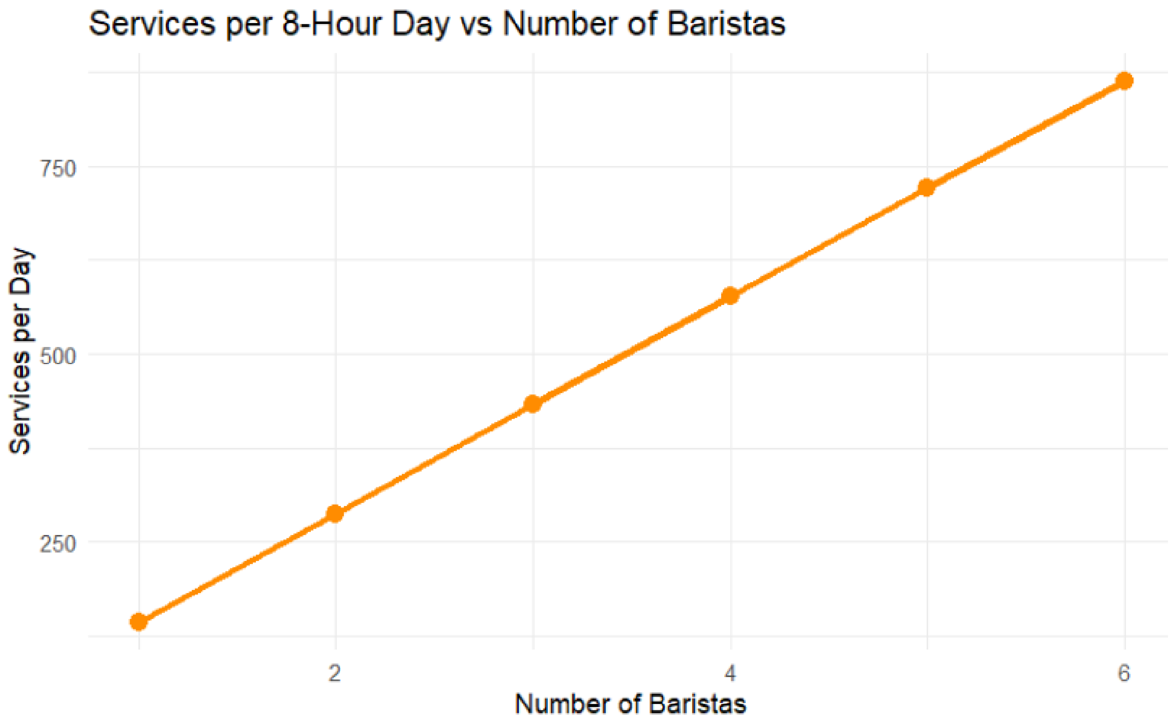

Figure 14 Services per 8-Hour Day vs Number of Baristas

This graph increases almost linearly, meaning each extra barista adds capacity at a fairly constant rate.

It shows that the shop can handle more customers predictably as more baristas are added, assuming they all work efficiently.

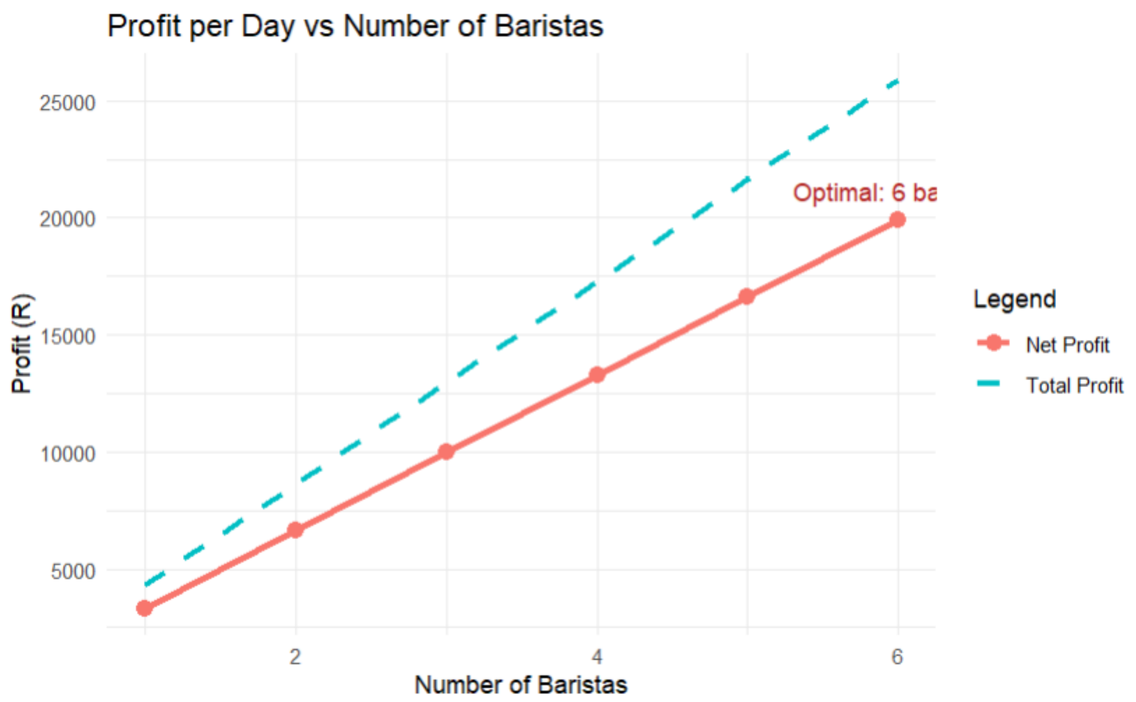

Figure 15 Profit per Day vs Number of Baristas

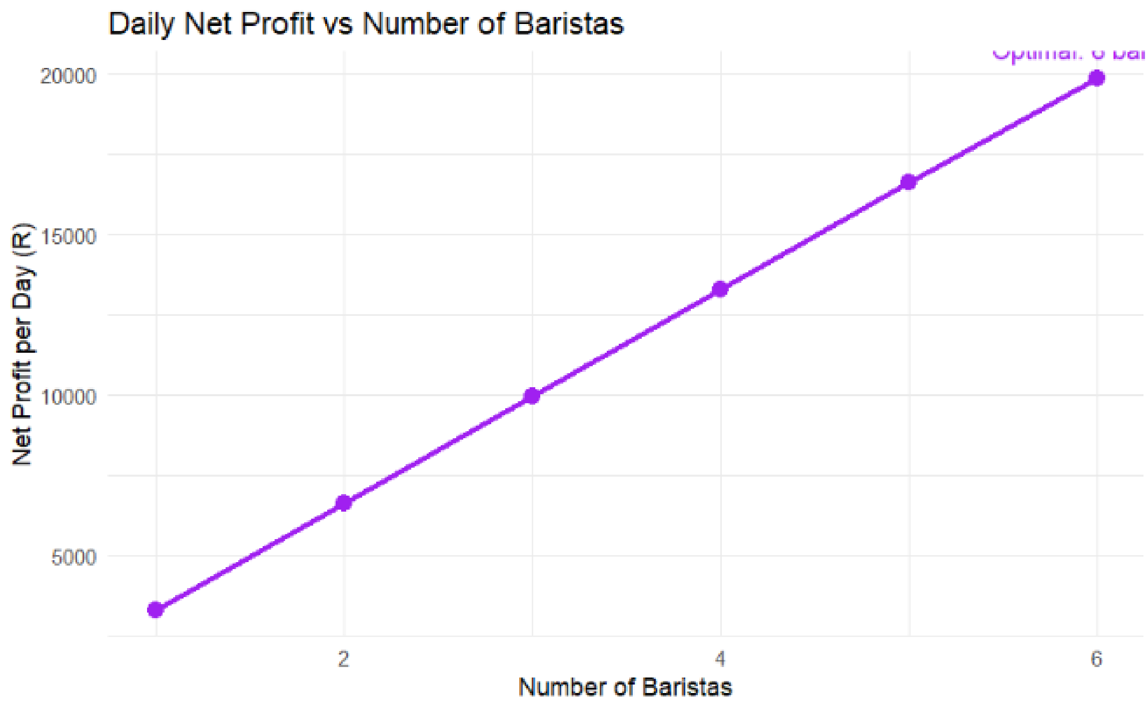

Figure 16 Daily Net Profit vs Number of Baristas

The total profit increases almost linearly as the number of baristas increases. This is expected, since adding more baristas allows more customers to be served per day, which directly raises total revenue.

The net profit also increases linearly, but at a slower rate. This is because the daily cost per barista reduces the overall profit margin. However, since service speed improves enough to offset the added labour costs, the shop continues to make higher profits as more baristas are added.

Table 13 Results Shop 1

	Baristas	AvgServiceTime	ReliableServicePct	ServicesPerDay	TotalProfit	PersonnelCost	NetProfit
1	6	33.35565	100	863.422	25902.66	6000	19902.66

The results for shop 1 show that the highest net profit is achieved with 6 baristas. At this point, the average service time is approximately 33 seconds, and 100% of services are completed within the reliable service threshold. With 6 baristas, the shop can serve around 863 customers per day, generating a total profit of R25,902.66. After subtracting personnel costs of R6,000, the net daily profit amounts to R19,902.66.

This setup provides a good balance between efficiency, reliability, and profitability.

Results Coffee Shop 2

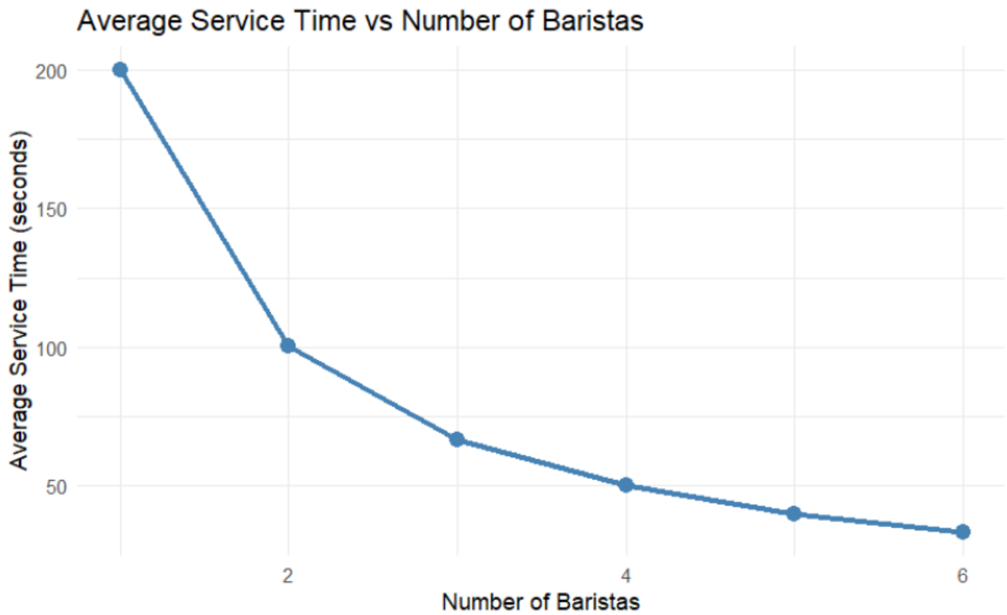

Figure 17 Average Service Time vs Number of Baristas

The graph shows a sharp decrease in service time as the number of baristas increases.

With fewer baristas, customers wait longer (around 200 seconds at one barista). As more baristas are added, the service time drops quickly and then levels off at about 25 seconds.

This suggests that adding staff improves efficiency, but after around 5–6 baristas the improvement becomes minimal.

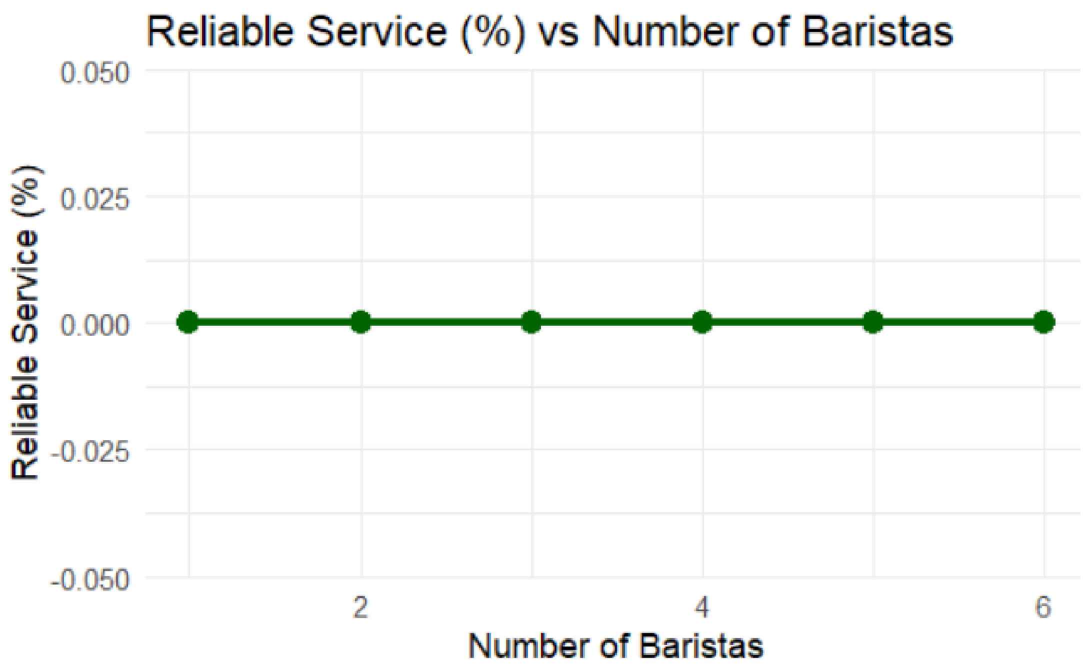

Figure 18 Reliable Service % vs Number of Baristas

The reliable service percentage remained 0% for all scenarios. This means that no orders were completed in under 60 seconds, regardless of how many baristas were working. It suggests that the process speed or setup is the limiting factor, not the number of workers.

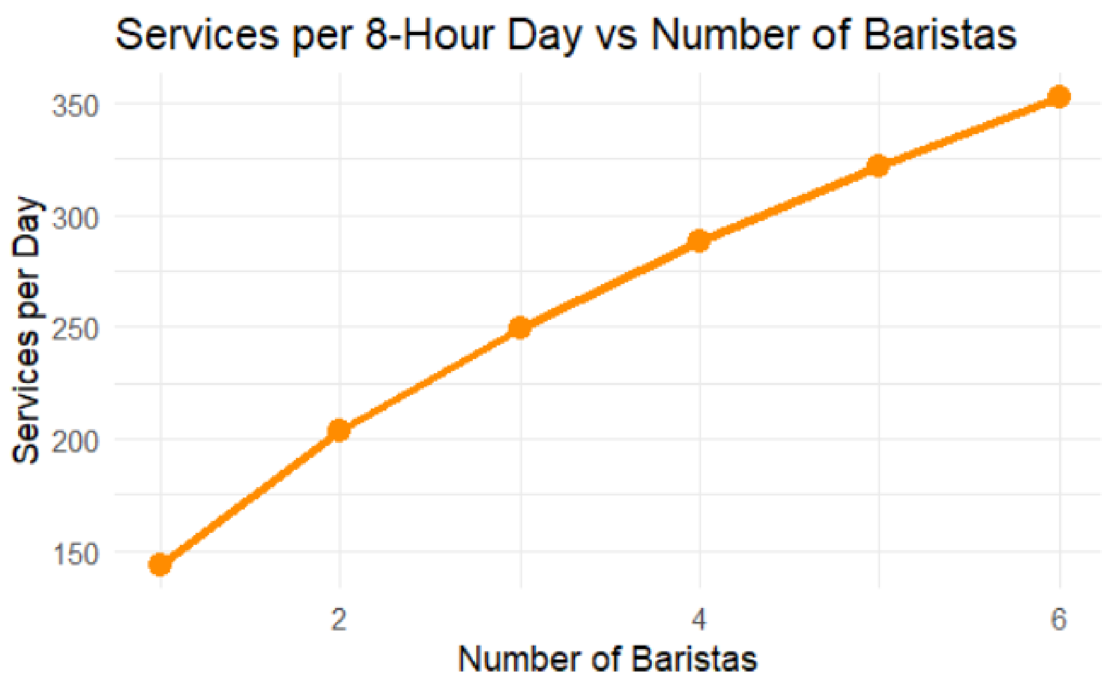

Figure 19 Services per 8-Hour Day vs Number of Baristas

The number of services completed per day increased as more baristas were added, but the increase was not linear. This shows that adding more baristas improves output, but the effect gets smaller each time indicating diminishing returns as the system becomes saturated.

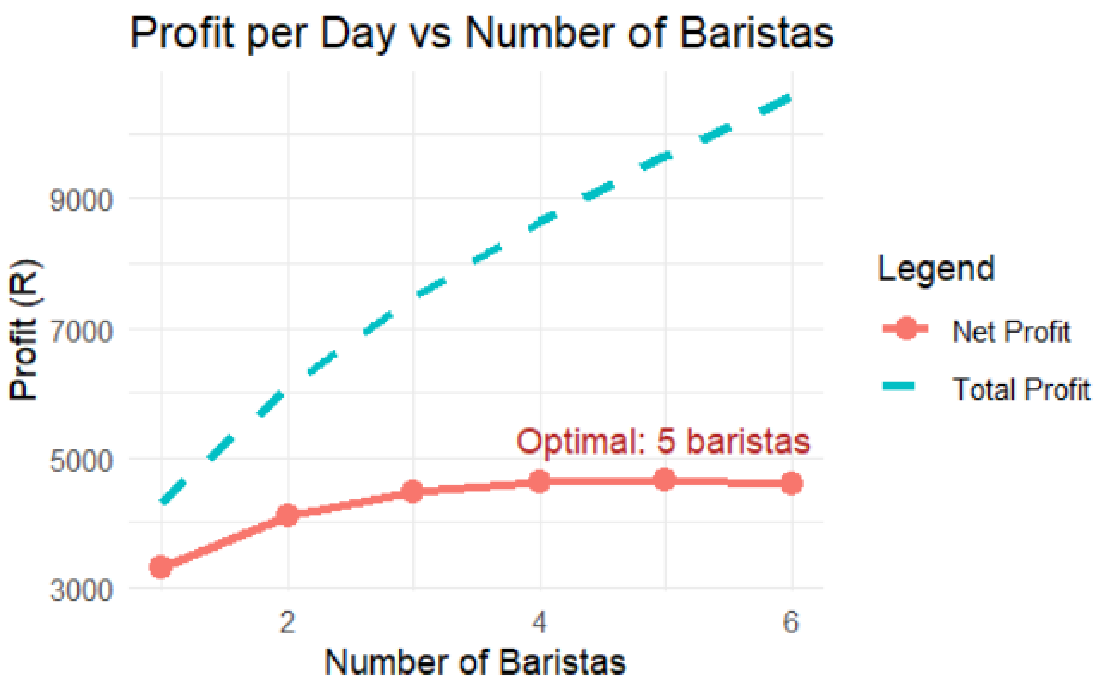

Figure 20 Profit per Day vs Number of Baristas

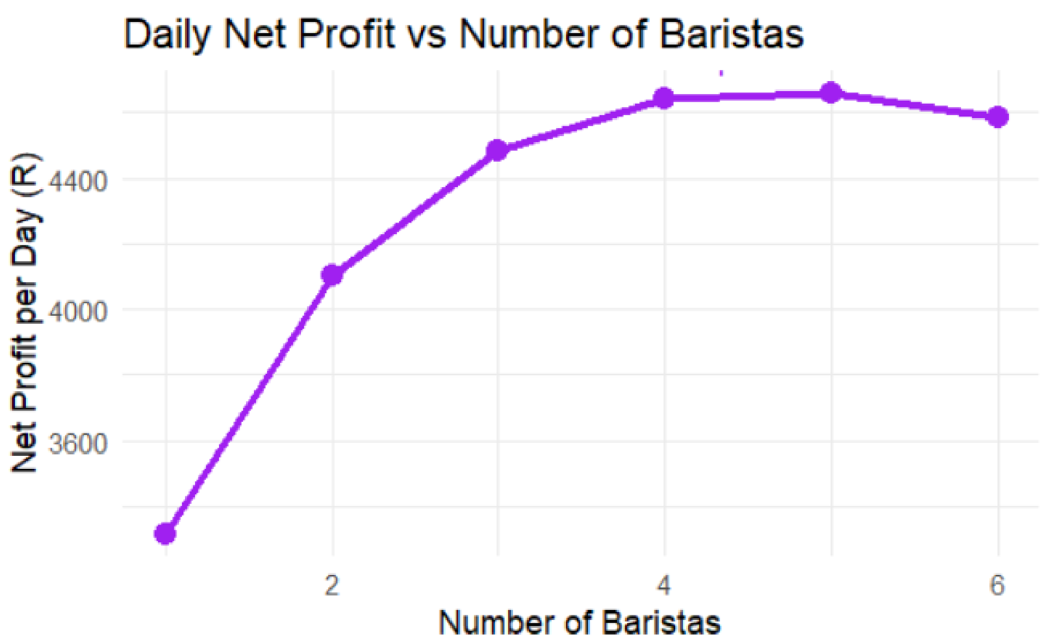

Figure 21Daily Net Profit vs Number of Baristas

Profit followed a similar upward trend at first but peaked at five baristas. Beyond this point, total profit decreased due to higher personnel costs outweighing the small productivity gains. This suggests that five baristas is the most cost-effective staffing level for Shop 2.

Table 14 Optimal Number of Baristas Shop 2

Baristas	AvgServiceTime	ReliableServicePct	ServicesPerDay	TotalProfit	PersonnelCost	NetProfit
5	89.43597	0	322.0181	9660.543	5000	4660.543

The optimal staffing level for Shop 2 is five baristas. At this point, the shop achieved the highest net profit of R4 660.54. Although the reliable service percentage was still 0%, the balance between productivity and personnel cost was most efficient. Adding a sixth barista did not improve performance enough to justify the extra cost, confirming that five baristas offer the best trade-off between profit and labour expenses.

The profit optimization reflects Taguchi's principle that deviation from target performance (ideal staffing level) introduces an economic loss.

Phase 4 hand in

ANOVA

Objective:

To test if there is a significant difference in average service time between Shop A and Shop B.

Hypotheses

Null hypothesis (H_0): There is no significant difference in average service time between the two shops.

Alternative hypothesis (H_1): There is a significant difference in average service time between the two shops.

Method

A one-way ANOVA was performed to test for differences in average serving time across the two shops.

The test compares the variation between shops (treatment) and within shops (error).

ANOVA Results

Table 15 ANOVA Results

	V1	V2	V3	V4	V5	V6
1	Source	SS	DoF	MS	fo	P-value
2	Treatment	281998011.91	1	281998011.91	956538.29	0
3	Error	117923810.81	399998	294.81	---	---
4	Total	399921822.73	399999	---	---	---

Interpretation

The p-value (0.000) is less than 0.05, meaning we reject the null hypothesis.

This indicates a statistically significant difference in serving time between the two shops.

The large F-value (956,538.29) further supports that the difference is not due to random chance.

Graphical Representation

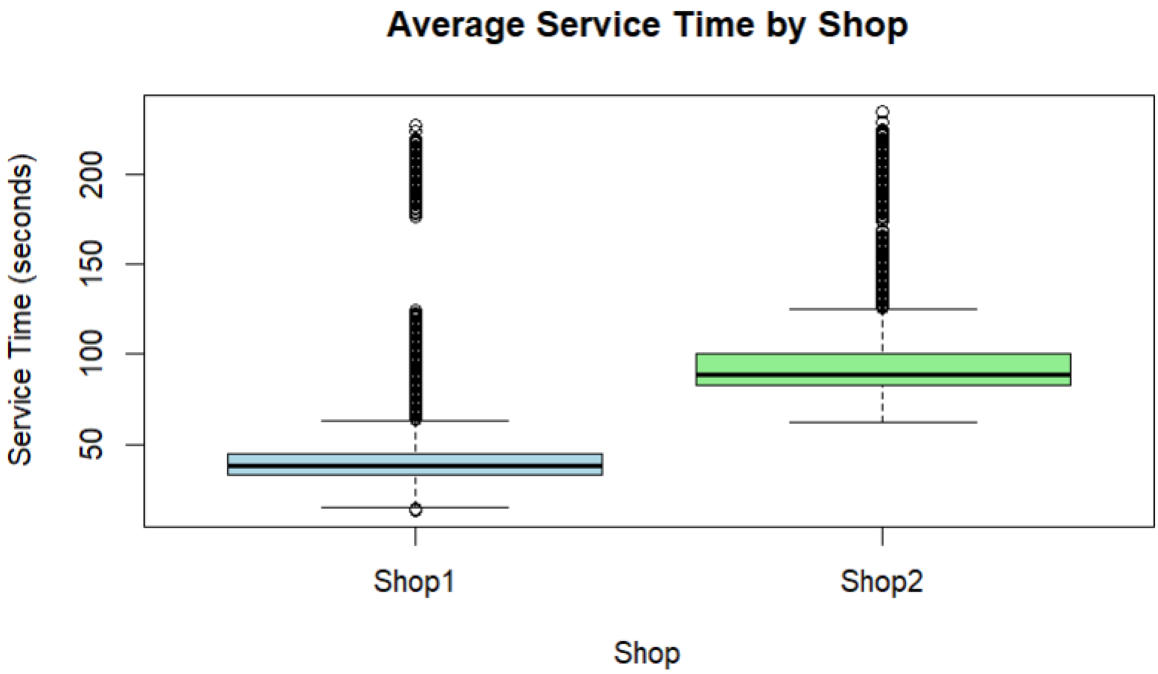

Figure 22 Average Service Time by Shop

The boxplot visually confirms that one shop has consistently higher service time than the other.

Conclusion

There is strong evidence of a difference in average service time between the two shops.

Management can use this insight to investigate factors such as staff efficiency, promotions, or customer flow that may be influencing performance.

7. Reliability of service

7.1

Based on the available data, reliable service is assumed to occur when 15 or 16 workers are on duty. From the 397 recorded days, this occurred on 96 + 270 = 366 days.

To estimate the expected number of reliable days per year:

366/397 x 365 = 366 days per year

Therefore, we can expect reliable service on approximately 336 days per year.

7.2

Model approach: treat historical counts as empirical distribution; simulate shifting counts by h hired staff and recompute the probability of having <15 workers. Use expected daily lost sales R20,000 and personnel monthly cost R25,000 to get monthly net benefit.

Decision rule: hire more staff while the marginal reduction in expected loss > monthly cost of that staff (R25,000).

Result: with the supplied distribution, hiring one full-time person is optimal.

Table 16 Results

	hire	problem_days	prob	monthly_loss	personnel_cost	net
12	0	31	0.07809	46851.39	0	0.00
121	1	6	0.01511	9068.01	25000	12783.38
122	2	1	0.00252	1511.34	50000	-4659.95
123	3	0	0.00000	0.00	75000	-28148.61

Practical notes: consider part-time hires or flexible rostering. If you can cover peaks with part-timers (cheaper pro rata), you may get similar reliability improvement for lower cost. Also test sensitivity: if daily loss is larger/smaller, or cost per person differs, the result will change.

Sensitivity table shows net benefit for hires 0–3 and for daily loss to show robustness.

Table 17 Sensitivity Table

	Workers	ProblemDays	LostSales	ExtraCost	ProfitChange_vs_Base
1	15	126	2520000	0e+00	0
2	16	52	1040000	3e+05	1180000
3	17	20	400000	6e+05	1520000
4	18	5	100000	9e+05	1520000

The analyses demonstrate application of engineering research methods to interpret simulated datasets and validate performance models.

Conclusion

The investigation successfully applied engineering research and statistical methods to analyse real-world style datasets and optimize process outcomes. Descriptive analysis revealed key insights into customer demographics, sales patterns, and product pricing.

Statistical Process Control (SPC) confirmed stable delivery means but identified variability in process spread for certain product types. Process capability indices indicated that only a few processes met the desired performance level.

Optimization analysis determined the most profitable staffing configurations for two coffee shops, and ANOVA confirmed a significant difference in service times between them. The reliability modeling demonstrated how data-driven staffing decisions can improve service performance.

The project meets the expectations of **ECSA GA4** by identifying and applying appropriate analytical methods, conducting data-based investigations, and validating results through statistical reasoning. Future work could include comparison with published industrial benchmarks and inclusion of real experimental data for validation.

References

Montgomery, D.C. (2019). Introduction to Statistical Quality Control. 8th ed. Wiley.

Taguchi, G. (1986). Introduction to Quality Engineering. Asian Productivity Organization.

ChatGPT (2025). “Assistance with statistical analysis and report writing.” OpenAI. Accessed October 2025.

R Core Team (2024). R: A Language and Environment for Statistical Computing. R Foundation for Statistical Computing, Vienna, Austria. Available at: https://www.r-project.org/