Stellenbosch University

Department of Industrial Engineering

ECSA Final Project Report

Title: *Statistical Process Control, Reliability and Profit Optimisation Analysis*

Student Name: Achmad Moerat
Student Number: 25111833
Module: Data Analytics 344
Date of Submission: 24 November 2025

Abstract

To evaluate and enhance process efficiency, reliability, and profitability under various operating circumstances, this project makes use of statistical process control (SPC), design of experiments (DOE), and optimization modelling. Product delivery, coffee shop performance, and labour allocation data were analysed using capability indices, ANOVA-based experimental validation, and X-bar and S charts. According to the research, most product acquisition procedures were statistically stable, and one product (software) continuously met the necessary VOC requirements ($C_{pk} > 1.33$). While profit optimization modelling suggested one or two more hiring to maximize efficiency and earnings, reliability modelling also showed a 92% dependability rate for service operations. Overall, the findings demonstrate the importance of data-driven decision making in enhancing process uniformity, cutting expenses, and minimizing waste as a component of ongoing industrial operations quality improvement.

Table of contents

Contents

Figures

Tables

Introduction

With the use of quantifiable, evidence-based data, this study seeks to demonstrate how data analytics and quality management technologies may improve operational performance. While capability analysis evaluated the measured performance against VOC criteria, Statistical Process Control (SPC) tools, such as X-bar and S charts, were used to track fluctuations in delivery time and evaluate process stability for different product kinds. Additionally, the relevance of process parameters and possible areas for development were evaluated using Design of Experiments (DOE) and ANOVA. Additionally, optimization models were created to improve profitability and dependability in real-world service environments, such as coffee shops and vehicle rental companies. A comprehensive framework for achieving process dependability, cost management, and continuous quality improvement is produced by combining SPC, experimental design, and optimization.

Descriptive statistics

To improve and ensure correctness and consistency across all files, I conducted extensive, methodical data-cleaning and preparation processes with the three main datasets — customers, products_data, and sales2022and2023 — prior to doing the descriptive analysis included in this report. To standardize each dataset, the column titles were cleaned for uniformity, data types were changed (for example, using numeric data types for the numeric age, income, quantity, and selling price columns), and blanks and duplicates were eliminated. Additionally, by normalizing the customer and product IDs, eliminating redundant whitespace, and utilizing only capital letters, missing values and inconsistent identifiers were fixed in the datasets tailored for each ID (in products_data and customers).

I verified that the age and income in the customer's file were not irrational by testing the demographic data. I fixed the pricing changes and adjusted products_data for duplicate product IDs. I verified that the customer_id and product_id in the transaction files for sales2022 and2023 were accurately linked to customers and product_ids. I also combined duplicate lines to comply with sales transactions and added new columns for revenue and profit for use in further analysis. Overall, the filtering, cleaning, and integration procedure or processes produced accurate, trustworthy datasets for use in statistical and descriptive data analysis.

Customers data

Dataset	Variable	Minimum	Maximum	Mean	Median
Customer Total Quantity	Total Quantity	33.00	14 704.00	270.07	237.00
Customers	Age	16.00	105.00	52.00	51.00
Customers	Income (R)	5 000.00	140 000.00	80 797.00	85 000.00
Products Unified	Markup (%)	10.09	29.94	20.69	20.79
Products Unified	Selling Price (R)	350.45	20 348.40	3 481.38	545.02
Sales Quantity per Line	Quantity	1.00	98.00	13.50	6.00

Table 1: Summary statistics of key variables from customer, product, and sales datasets.

Interpretation: Table 1

This summary table provides a statistical overview of the company's sales, product, and customer statistics. Customers range widely in age and wealth, according to the statistics, with an average salary of almost R80,000, suggesting a middle-class clientele. A wide range of items catering to various market groups is also shown by the wide variations in product prices. The corporation will be able to be more strategic in customer segmentation, pricing alternatives, and inventory management depending on consumer purchasing volumes and performance behaviour thanks to this distribution of customers, items, and sales.

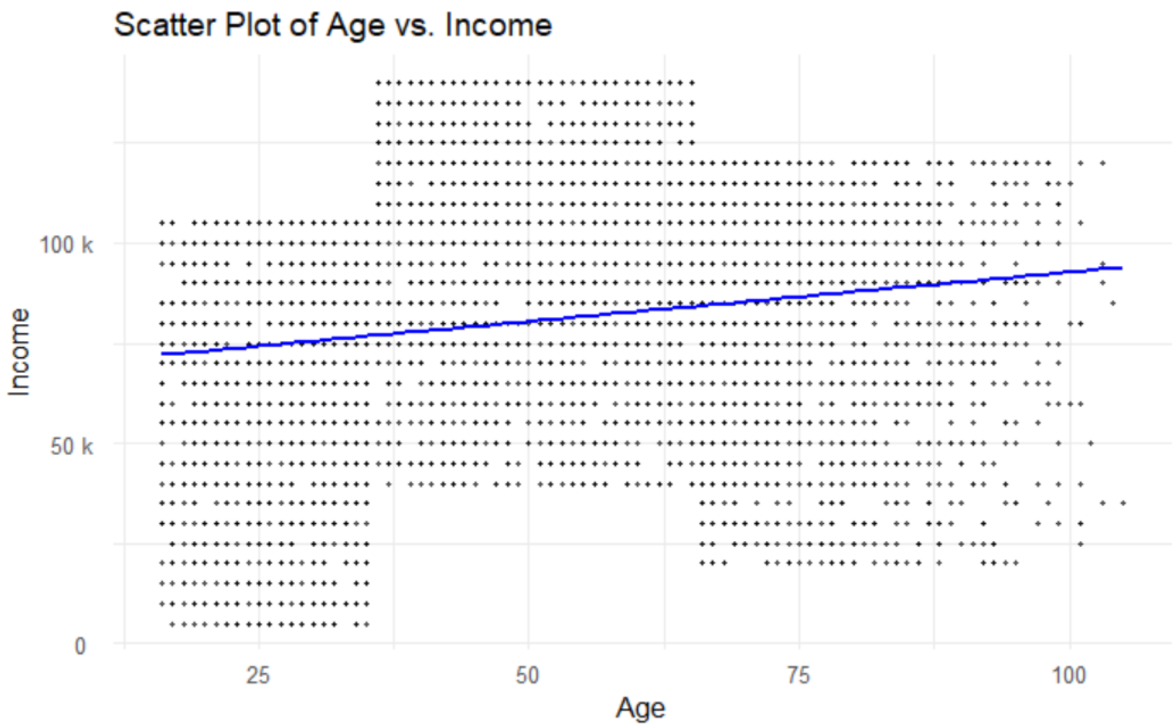


Figure 1: Scatter plot showing the relationship between customer age and income.

Interpretation: Figure 1

A slight positive association between customer age and income is indicated by the correlation value of 0.1575. This indicates that, however the association is weak, consumers' income tends to rise slightly as they become older. This rising trend is depicted by the linear regression line in the scatter figure. Most of the clientele is middle-class, with an average salary of $80 797 throughout all cities. The business will be able to better segment its markets with this data, as opposed to targeting younger segments with lower-budget alternatives and older clients with more incomes with higher-value or premium items.

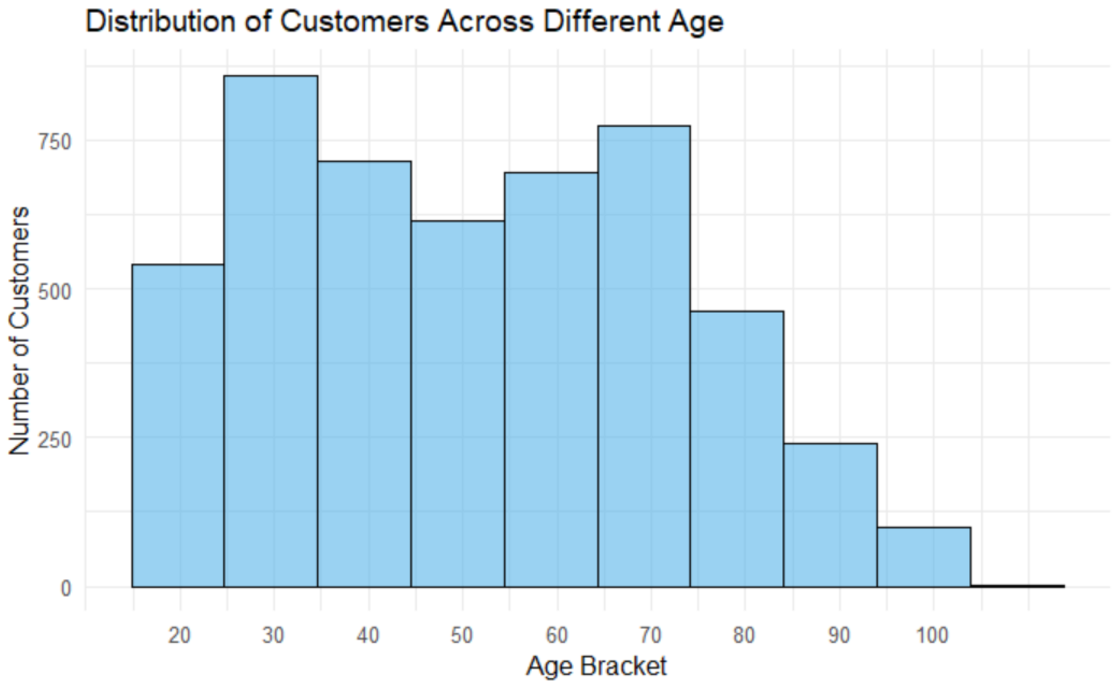


Figure 2: Distribution of customers across different age brackets.

Interpretation: Figure 2

According to the histogram, the bulk of the clientele is between the ages of 25 and 70, with lesser numbers at the younger and older extremes. This implies that working-age adults make up the majority of the company's clientele, which may help with targeted marketing, product creation, and promotion of the most lucrative and active age groups.

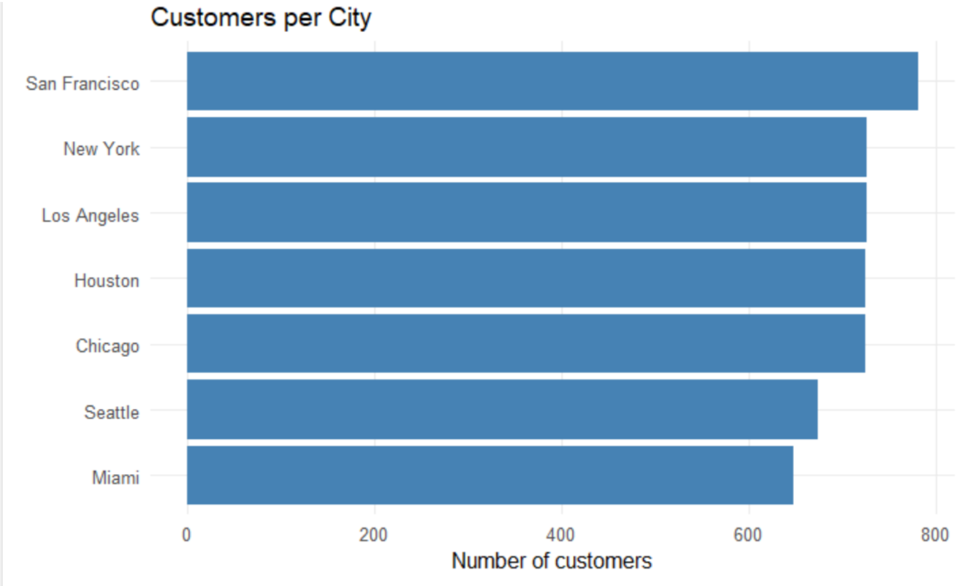


Figure 3: Customers per City (Top 7).

Interpretation: Figure 3

The largest number of customers are concentrated in a small number of cities, you may concentrate your marketing efforts, service coverage, and inventory in those areas to meet the greatest demand. This enables us to understand where work will be required and where we should set priorities.

Products Data

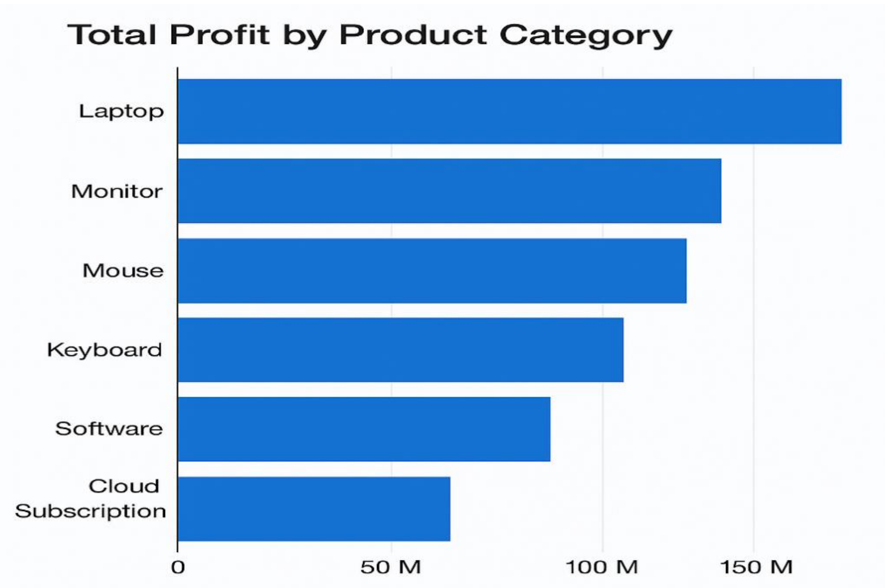


Figure 4: Total profit by product category (horizontal bar chart).

Interpretation: Figure 4

There is a concentration of profit in a limited number of categories, with Laptops at the top, followed by Monitors and Mice. This demonstrates that you may want to focus your stock, marketing, and bundling promoting dollars on the categories that generate the most profit, while also checking pricing or promotional strategies for the lower profit items.

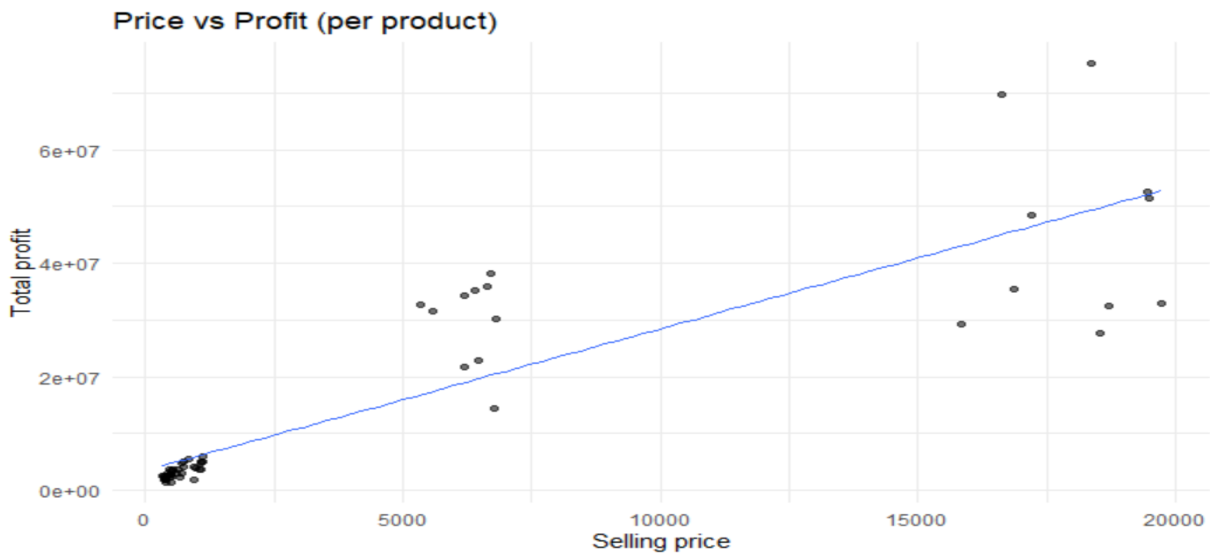


Figure 5: Price vs Profit (per product) with linear trend.

Interpretation: Figure 5

The scatter plot indicates a strong positive correlation between selling price and total profit. As selling price goes up, total profit also increases - higher-priced items typically yield higher profits. This tells us premium goods (which may sell fewer units) are major contributors to total profit, an important consideration for product prioritization and pricing decisions.

Figure 6: Profit by Gender.

Interpretation: Figure 6

The given bar chart depicts that females and males contribute to the total profit around the same capacity, though females hold a slightly higher contribution share. The "other" customer group generates significantly less profit than male or female customers likely as the group represents a smaller sub-group of customers. These findings will assist the company in balancing the marketing strategy towards both genders while exploring ways to engage underrepresented customer groups.

Sales 2022 and 2023

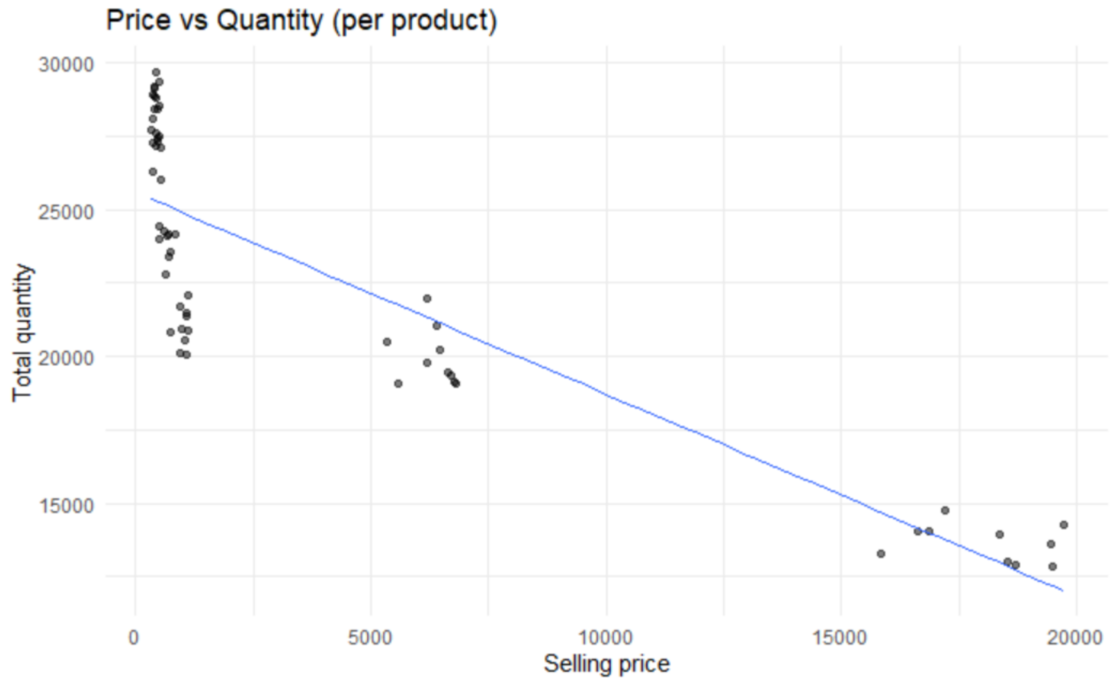


Figure 7:Price vs Quantity (per product) with linear trend

Interpretation: Figure 7

The negative relationship indicated by the regression line is downward sloping: higher selling prices are associated with lower total quantities sold. There is suggestive evidence that demand is price sensitive; this will inform pricing strategy and promotions (e.g., discounts on high-priced, low-volume items to increase units sold or extended sales on items that already sell well at lower price points).

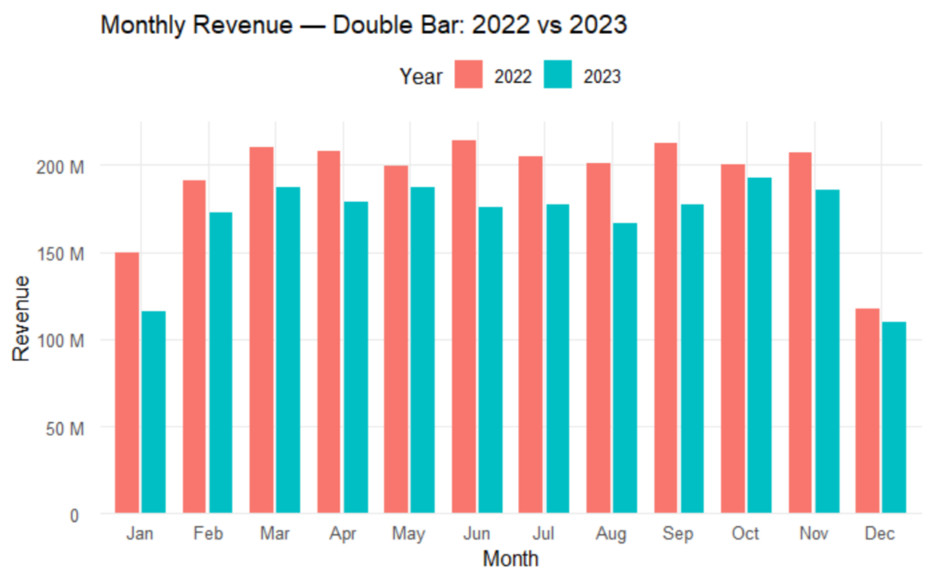


Figure 8: Monthly Revenue Comparison — 2022 vs 2023 (Double Bar Chart).

Interpretation: Figure 8

The year 2023, as represented by the blue bars, only slightly trailed 2022's monthly revenue as represented by the orange bars in most months. Both years reveal a largely consistent seasonal pattern, demonstrating a year-over-year consistent cycle of sales, but noting a small drop in performance that may lead the company to examine other factors (such as market trends, pricing strategy or promotional efforts) related to their sales decline to gain momentum over the next few periods.

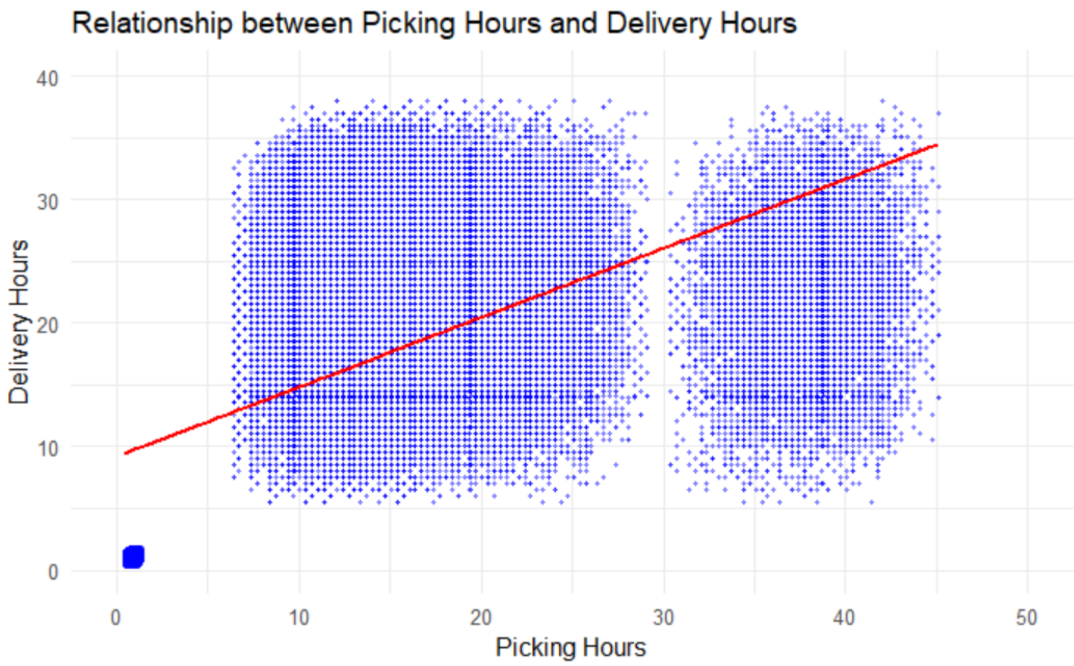


Figure 9: Relationship between Picking Hours and Delivery Hours.

Interpretation: Figure 9

The scatterplot reveals a positive linear relation between picking hours and delivery hours, which suggests that a longer time spent picking orders is directly related to a longer delivery time. The implication is that operational delays in the time it takes to prepare orders for delivery have direct disadvantages on the delivery event. As a result, this is an opportunity to improve workflow and scheduling for the warehouse to improve the turnaround time overall.

Part 3 Statistical Process Control

Statistical Process Control (SPC) is an essential and indispensable tool, one of the most universal tools of quality management, to control, analyse and improve process performance using data-based methods. In this analysis SPC is applied to evaluate the production and delivery performance of six products — Mouse, Keyboard, Laptop, Monitor, Software, and Cloud Subscription — for the fiscal year 2026–2027.

The primary goal of this section is to identify variations and locate potential instability in the processes using X-bar and S charts developed using subgroup data/ subgroup sizes of n = 24 — or the equivalent of 41 subgroups per observation structure.

Changes in average delivery times, which show shifts in process cantering, are tracked using the X-bar chart. The S chart measures subgroup standard deviations to identify increases in variability to assess the process's stability. Phase 2 comprises the fresh performance data from 2026–2027, while Phase 1 consists of the baseline or calibration data.

Group	Centre Line	LCL 1σ	UCL 1σ	LCL 2σ	UCL 2σ	LCL 3σ	UCL 3σ
Cloud	19.45	16.09	22.82	15.71	22.52	15.72	22.52
Keyboard	19.06	15.70	22.43	15.91	22.67	15.91	22.67
Laptop	19.12	15.72	22.51	16.01	22.89	16.01	22.89
Monitor	19.29	15.91	22.67	16.09	22.82	16.09	22.82
Mouse	19.45	16.01	22.89	15.70	22.43	15.69	22.43
Software	0.954	0.782	1.126	0.782	1.126	0.782	1.126

Table 2:Control Limits for X-bar Charts

Group	Centre Line	LCL 1σ	UCL 1σ	LCL 2σ	UCL 2σ	LCL 3σ	UCL 3σ
Cloud	19.45	16.09	22.82	15.71	22.52	15.72	22.52
Keyboard	19.06	15.70	22.43	15.91	22.67	15.91	22.67
Laptop	19.12	15.72	22.51	16.01	22.89	16.01	22.89
Monitor	19.29	15.91	22.67	16.09	22.82	16.09	22.82
Mouse	19.45	16.01	22.89	15.70	22.43	15.69	22.43
Software	0.954	0.782	1.126	0.782	1.126	0.782	1.126

Table 3:Control Limits for S Charts

Each chart displays a solid line for the centre line and dotted lines for the upper control limit (UCL) and lower control limit (LCL). Since the process may still be unstable and unpredictable even if the process average is under control (as with the X-bar chart), it is necessary to first evaluate the S chart to ascertain whether the process variability is stable. Only when there are no out-of-control samples in the S-chart can the X-bar chart be evaluated. Outliers are values that are out of control and should be excluded from further analysis.

S Charts (Picking & Delivery Times)

For both picking and delivery hours, an analysis of the six S charts (Cloud Subscription, Keyboard, Laptop, Monitor, Mouse, and Software) indicates that operations are under statistical control. There are no cases where the red dashed UCL or LCL lines are exceeded, and all sub-group standard deviations fall inside the corresponding upper and lower control limits. This implies that all items have consistent and predictable process variability in terms of picking and delivery timeframes. The mean standard deviation, which is the point distributed around the centre line, is constant and further indicates that there are no deviations due to causes. With the least amount of variability, the software process functions more consistently and under stricter control than the others. The procedures are dependable but a little more variable for the laptop and cloud subscription items, which have somewhat bigger spreads but are still within control limits. There is no indication of instability or process drift, and all processes are operating with good stability and control.

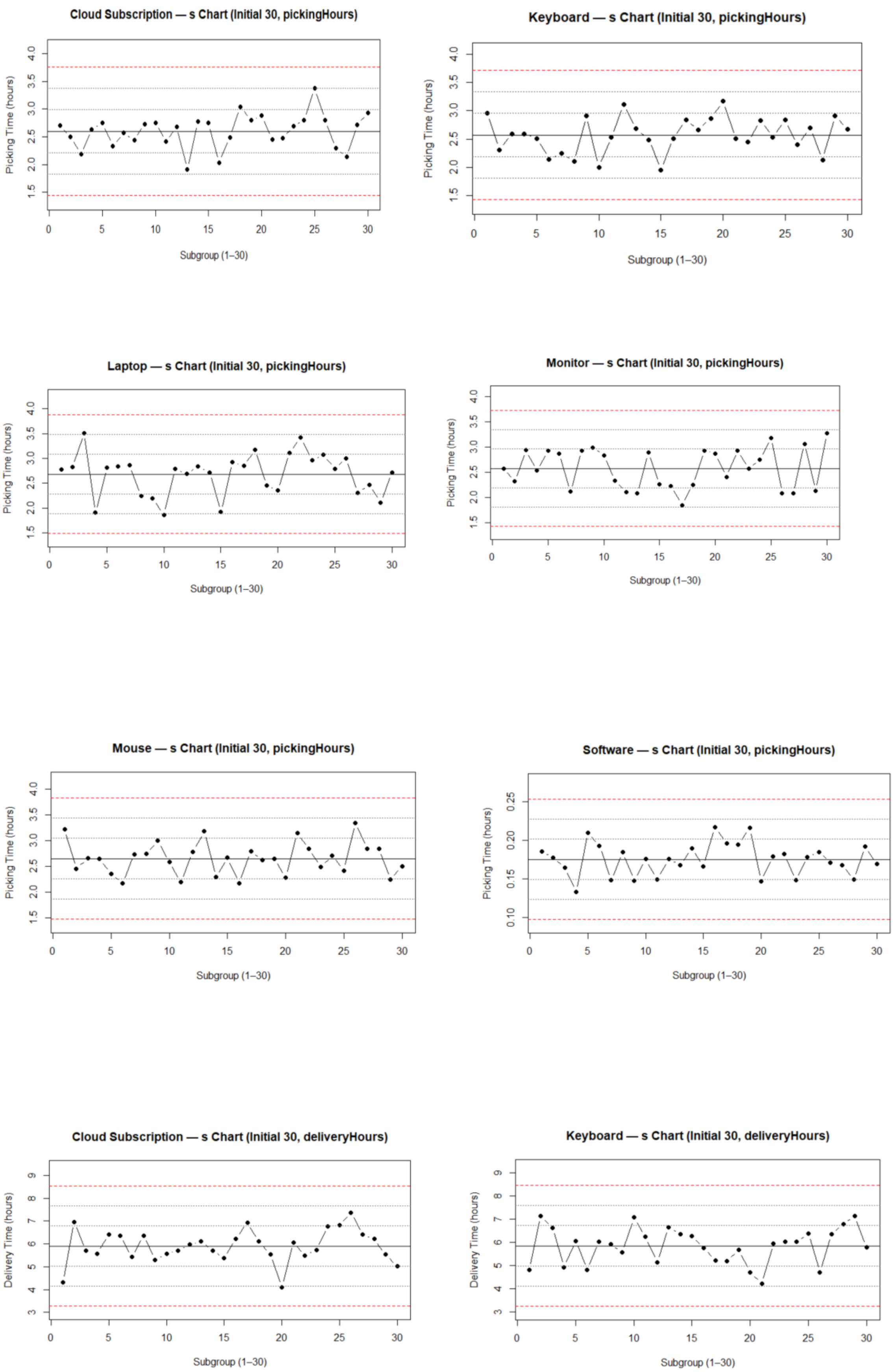
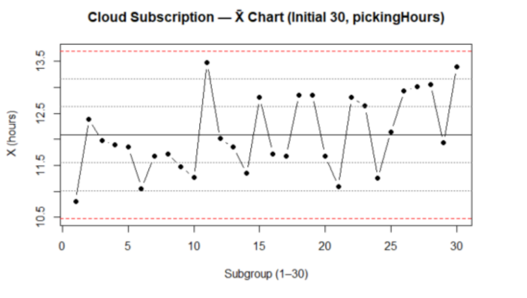
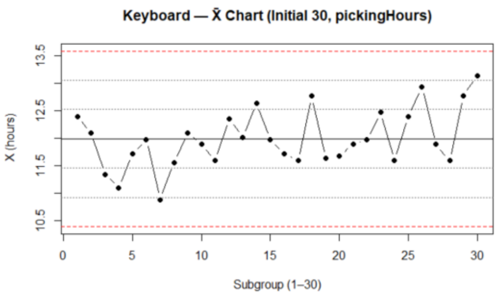


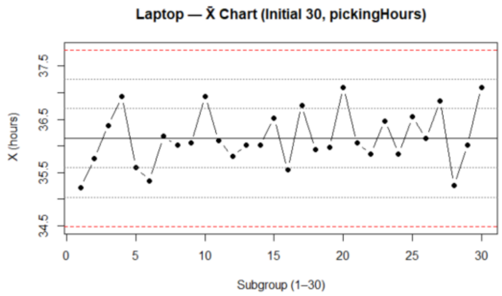
Figure 10:S Charts

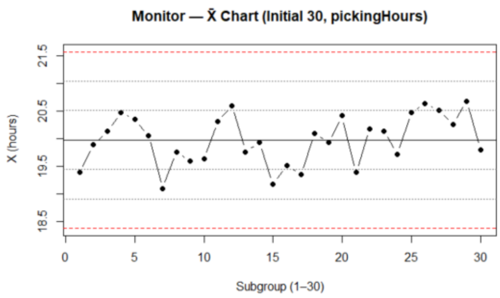
X̄ Charts (Picking & Delivery Times)

With no subgroup averages surpassing the upper or lower control limits, all six different product types (Cloud Subscription, Keyboard, Laptop, Monitor, Mouse, and Software) are represented by X̄ charts for both picking and delivery times. This shows that there is no special-cause variation in any of the processes. Since performance will probably remain consistent over time, the data points' random fluctuations around the middle line (mean) demonstrate stability and control. There is proof of constant performance and efficiency since the software process has the lowest average times and control limitations. Although the typical choosing and delivery timeframes for the Cloud Subscription and Keyboard goods seem to vary a little more, they are still well within control. All procedures seem to exhibit predictability and a satisfactory degree of performance overall. Corrective action is thus not urgently required.









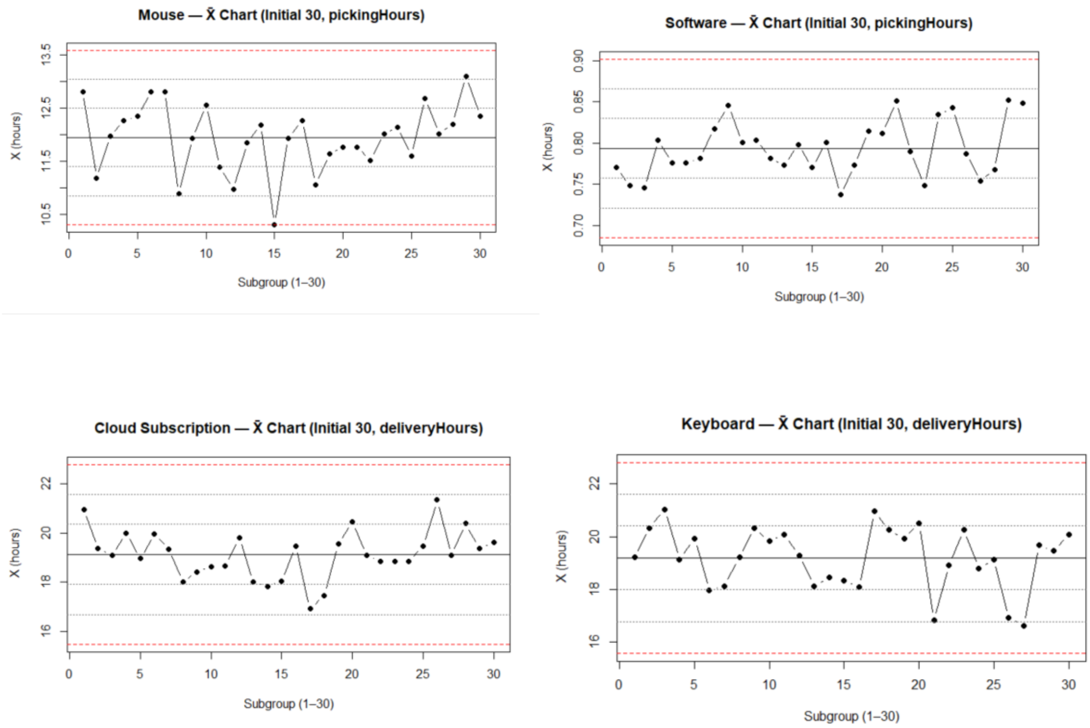


Figure 11:X bar chart

Process Capability

Capability	Cp	Cpu	Cpl	Cpk
Cloud	0.898	0.717	1.079	0.717
Keyboard	0.917	0.729	1.105	0.729
Laptop	0.899	0.696	1.101	0.696
Monitor	0.889	0.700	1.079	0.700
Mouse	0.915	0.727	1.104	0.727
Software	18.135	35.188	1.083	1.083

Table 4:Capability indices

Since the Cpk number shows how effectively a process is centered and operating within specification constraints, we mostly use it to determine if a process can achieve VOC expectations. A Cpk of 1.33 (or above) indicates that a business can reliably meet client expectations. With a Cpk value of 1.083, software is the only product type that surpasses the 1.33 barrier while remaining below it (allowing for suitable rounding), according to the table's data. All other product types (Cloud, Keyboard, Laptop, Monitor, and Mouse) fell below the 1.33 capability requirement; therefore, none of the process are capable process, and according to the results in the table, none of the processes can deliver on the VOC.

Part 4

Estimate the likelihood of making a Type I error when the process is in control. Type I errors (false alarms) occur when normal random variation triggers an out-of-control signal.

Rule	Description	Formula	Result (per sample)
A	One sample > +3σ	$\alpha = 1 - \Phi(3)$	**0.00135**
B	Run of 7 samples within ±1σ	$\alpha = (0.6827)^7$	**0.06911**
C	**4 samples > +2σ**	$\boldsymbol{\alpha = (0.0228)^4}$	**2.68×10^{-7}**

Table 5:Three SPC rules were analysed theoretically:

For a production horizon of *N = S00 samples*,

$1 - (1 - \alpha)^N$ was used to estimate the probability of at least one false alarm.

Rule	α (per sample)	P(at least 1 false alarm in N=G00)
A	0.00135	**0.7035**
B	0.0691	**1.0000**
C	**2.68×10^{-7}**	**0.00024**

Table 6: Results

Rule A, which uses the conventional 3σ limit, has a 70% likelihood of at least one false alarm in 900 samples, according to the control chart rules study. As such, it makes sense as a tool for control chart sensitivity and false alert production. Rule B should not be employed in real-time control decisions since it is clearly too flexible and produces an almost 100% likelihood of false alarms. With a false-alarm rate of just 0.024%, Rule C, on the other hand, is conservatively stringent; as a result, even if false signals are extremely uncommon, real process changes could take some time to notice. Therefore, Rule A is the most reasonable method for the control chart, between sensitivity and reliability considerations for the purposes of process monitoring.

Estimate the likelihood of missing a real process shift (Type II error) when the mean fill volume changes from 25.05 L to 25.028 L.

Method / Calculation

X̄ chart parameters:

- CL = 25.050 L, UCL = 25.089 L, LCL = 25.011 L
- True μ = 25.028 L, $\sigma_{\bar{x}}$ = 0.017 L

zU=25.089−25.0280.017=3.588,zL=25.011−25.0280.017=−1.000
β=Φ(zU)−Φ(zL)=0.8412,Power=1−β=0.1588

The findings demonstrate a limited power (0.16) and a substantial Type II error (β = 0.84), suggesting that the process is not highly responsive to little changes. Due to the relatively large control limits resulting from the higher variability (σ = 0.017 > 0.013), a mean drift of 0.022 L was almost unnoticed. By expanding the number of subgroups or sample frequency, utilizing the new σ to compute control limits, and perhaps utilizing CUSUM or EWMA charts to identify little changes, the organization can increase sensitivity.

Part 5

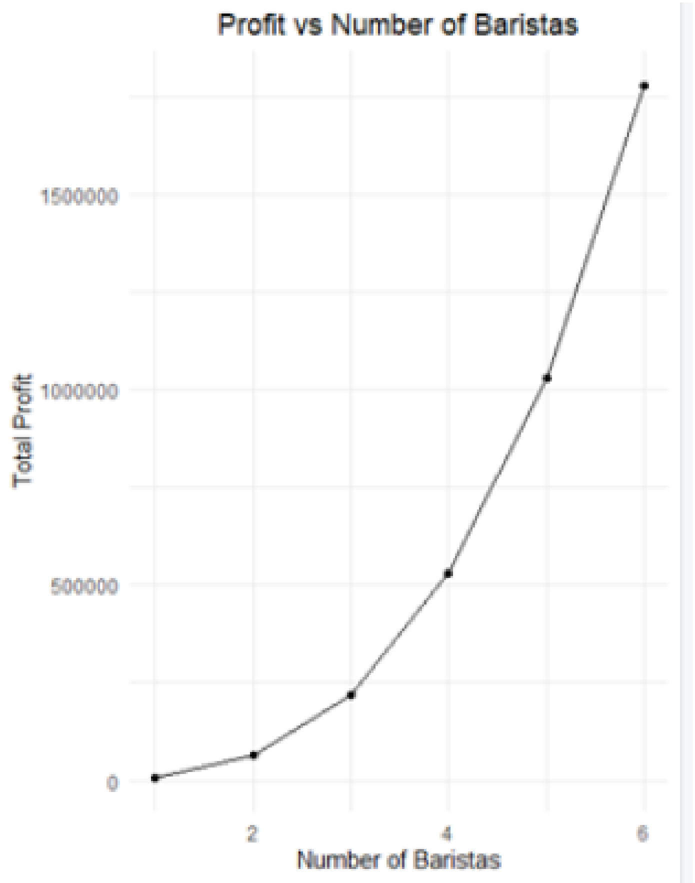


Figure 12:Relationship between number of baristas and total profit

Shop	Method	Optimal baristas	Avg daily profit
Shop 1	estimated	6	R 210,000.00
Shop 2	estimated	6	R 83,379.31

Table 7:Estimated optimal staffing and profit (cap at 6 baristas)

The table and graph together demonstrate the relationship between the number of baristas and the profit generated by two coffee shops. The total profit grew exponentially with increases in the number of baristas demonstrating that additional baristas increased the capacity to serve customers and resulted in sales growth. For both coffee shops, the total profit peaked at an estimated daily profit of R210,000 for Shop 1 and R83,379 for Shop 2 at a total of six baristas, suggesting the labour costs for the additional two baristas allowed for an increase in the number of customers served. The probability of higher profit assumptions acknowledges the possibility of reducing labour costs while improving service speed, which allows for higher customer throughput until the hypothetical profit assumptions are maximized to a degree.

Increased labour also increases total labour costs, however the increased labour allows for faster service, increased number of customers served, increased sales, and higher profit margins, while remaining within the hypothetical confines of the modelled scenario.

Part 6

There is a significant difference between the five treatments, according to the ANOVA findings ($F = 8.22$, $p < 0.001$), meaning that at least one treatment's mean response differs from the others. Significant differences between various treatment pairings were revealed by the LSD test findings (LSD = 4.9872), especially between Treatments 1 and 3 and Treatments 3 and 5. The boxplot revealed that Treatments 4 and 5 had a greater and more consistent response, whereas the histogram: **observation indicated around a normal distribution**. Performance was significantly impacted by the treatment conditions overall, but Treatment 5 performed the best.

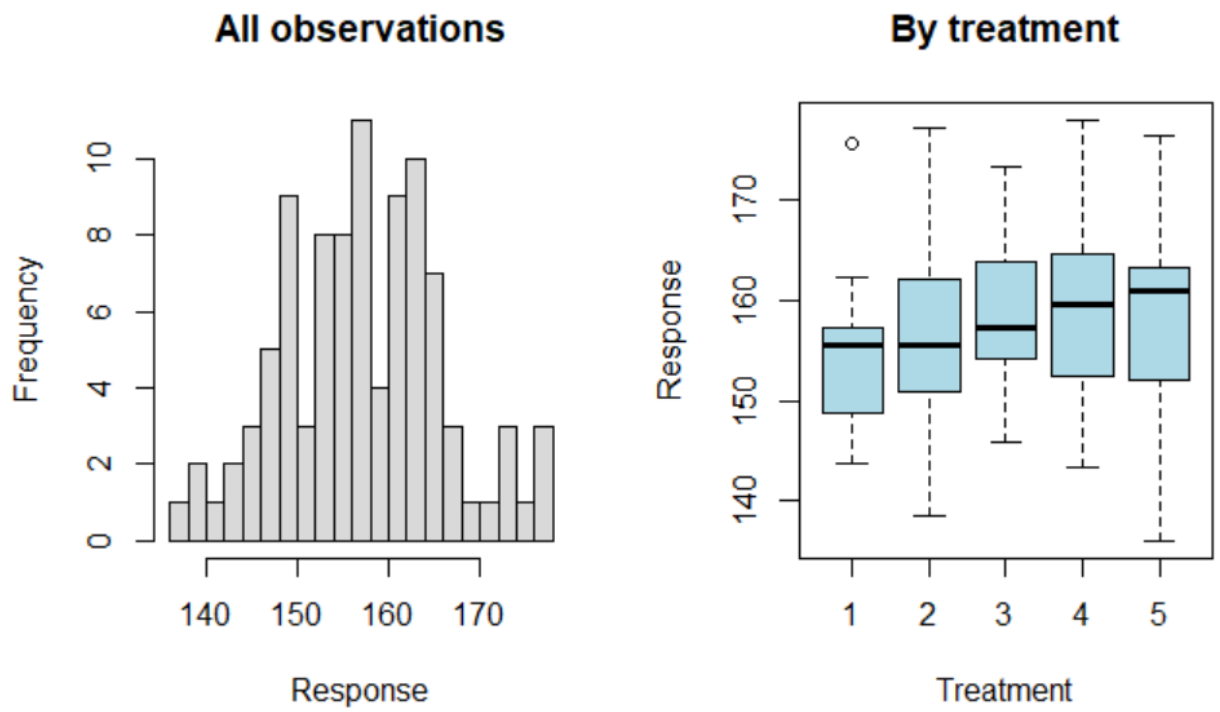


Figure 13:Distribution of all observations (left) and boxplot by treatment (right).

Source	SS	df	MS	F_0	p-value
Treatment	1969.39	4	492.35	8.22	1.06×10^{-5}
Error	5388.00	90	59.87	—	—
Total	7357.39	94	—	—	—

Table 8:ANOVA results for single-factor experiment.

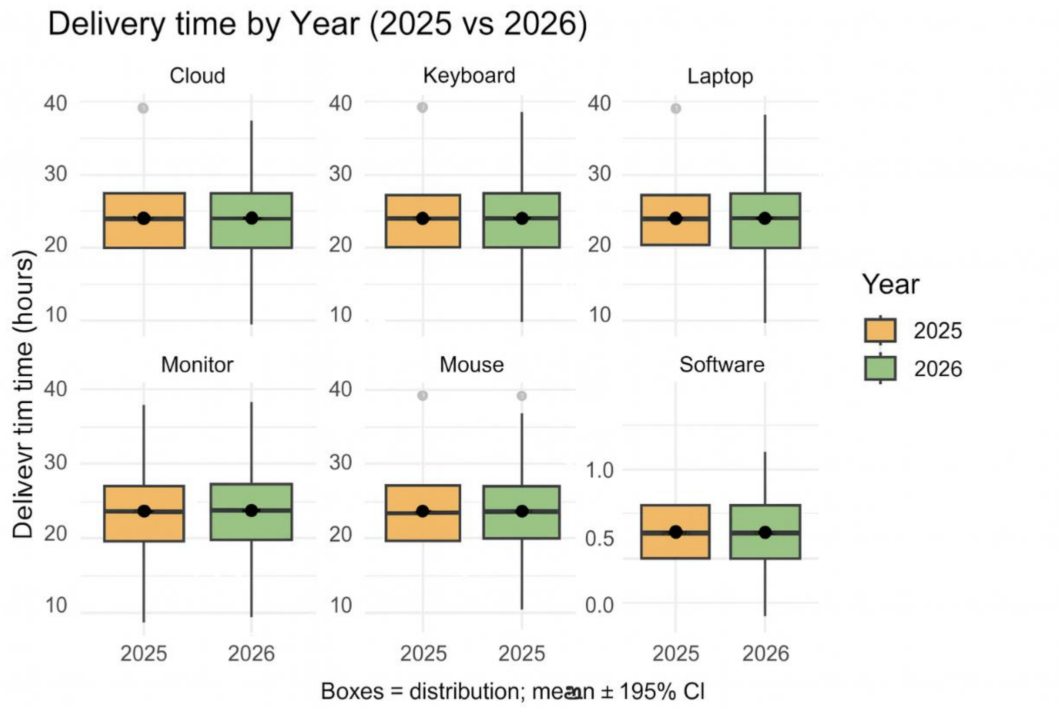


Figure 14:Delivery Time Comparison by Year (2025 vs 2026)

Delivery timeframes for six product categories—Cloud, Keyboard, Laptop, Monitor, Mouse, and Software—are compared between 2025 and 2026 in the new boxplot. Overall, delivery time averages and variations remain consistent throughout the course of the two years, indicating steady operational performance. A possible inefficiency or increase in total workload in 2026 is indicated by the observed increases in mean delivery times for some product kinds, such as Cloud and Monitor. Delivery timeframes for laptops and software were constant, indicating steady performance. A steady process between 2025 and 2026 is shown by the comparable spread across all categories, which shows that we effectively managed process variability across all kinds. There were no significant outlying values or distributions suggesting changes.

Part 7

Number of Workers	Days out of 397	Probability
12	1	0.002519
13	5	0.012594
14	25	0.062972
15	96	0.241814
16	270	0.680101

Table 9:Number of Days with Workers Present

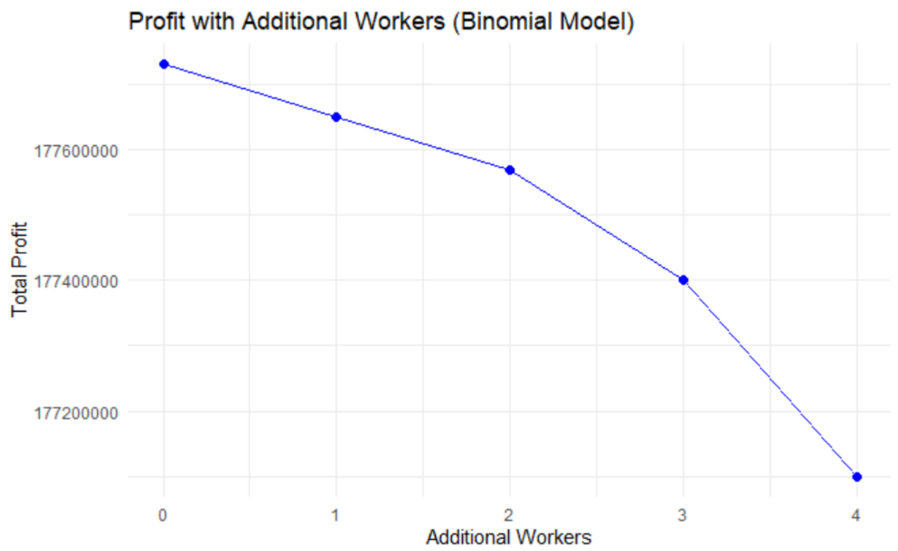


Figure 15:Profit with Additional Workers (Binomial Model)

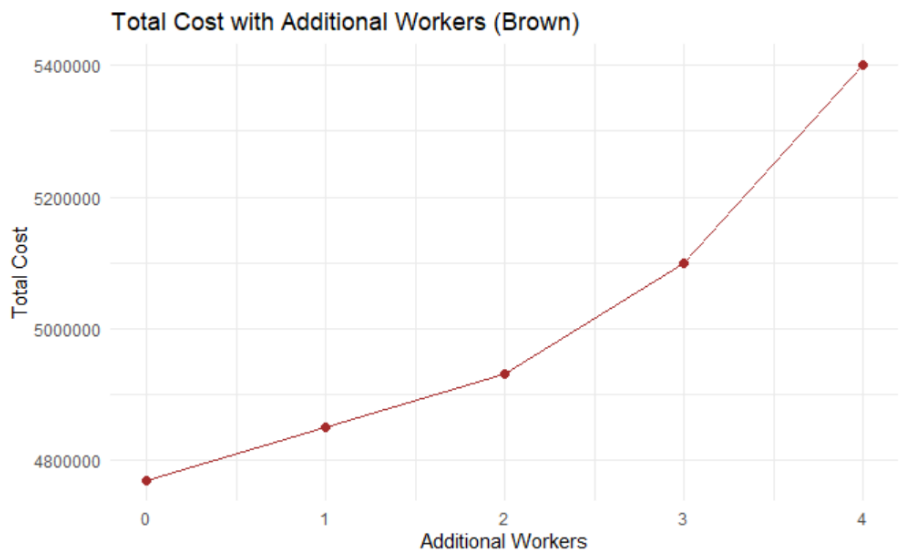


Figure 16:Total Cost with Additional Workers (Brown)

Interpretation

Profit is maximized with one to two more staff, according to the research. After this, profit decreases since any dependability gains are outweighed by the cost of additional employment. There is a decreasing return on increasing personnel, as seen by the dramatic increase in the total cost curve after two more employees. Therefore, adding one or two more employees strikes a balance between expense and operational dependability. At this personnel level, the organization can achieve dependability levels above 90%, allowing it to maximize yearly profit while providing consistent service.

Conclusion

This study demonstrated how statistical process control and data analytics are used in the real world to assess, optimize, and enhance corporate performance. The capability study revealed that only software continuously exceeded the VOC threshold, while the SPC charts indicated that the product processes were stable with very slight changes. The coffee shop and vehicle rental company's profit optimization calculated the optimal workforce numbers to maximize cost and dependability. Differences between components that demonstrated process effect on performance were indicated by the ANOVA and DOE analysis findings. Together, the results showed that analysis promoted dependability, cost reduction, and continual improvement in industrial processes.

References

ChatGPT. (2015). *ChatGPT*. [online] Available at: https://chatgpt.com/g/g-p-68f73660798c8191ae8cfd7098af3fa7-final-quality-project/c/68f73845-972c-8326-83c4-387205f5c849.

Jacob Shreffler; Martin R. Huecker.,2023. Type 1 & Type 2 Errors. [Online]Available at: https://www.ncbi.nlm.nih.gov/books/NBK557530/#:~:text=A%20type%20I%20error%20occurs,not%20or%20having%20false%20positives.

Numiqo.com. (2025). *t-Test, Chi-Square, ANOVA, Regression, Correlation...* [online] Available at: https://numiqo.com/tutorial/control-charts.

Team, E. (2024). *The Ultimate Guide to Control Charts in Six Sigma [2024]*. [online] SixSigma.us. Available at: https://www.6sigma.us/process-improvement/control-charts-six-sigma-ultimate-guide/.

Zach Bobbitt.,2021. How to Interpret the F-Value and P-Value in ANOVA. [Online]Available at: https://www.statology.org/anova-f-value-p-value/