ECSA PROJECT

Quality Assurance 344

Mieke Walters
23255366

Table of Contents

Part 1.1

Introduction

Part 1.1 provides a comprehensive analysis of the company's sales, customer, and product data for the years 2022 and 2023. As part of this analysis, data was consolidated, cleaned, and explored to identify key trends and insights. The goal is to transform raw data into actionable information that supports strategic decision-making by management.

The analysis covers descriptive statistics, revenue performance by product and customer segments, sales trends over time, and pricing consistency. The findings are presented with supporting tables and visualisations, followed by recommendations aimed at improving profitability, streamlining operations, and strengthening customer relationships.

Summary Statistics

Sales summary

The sales dataset has 100 000 rows and 9 columns.

Table 1.1

Variable type: character

skim_variable	n_missing	complete_rate	min	max	empty	n_unique	whitespace
customer_id	0	1	7	8	0	5000	0
product_id	0	1	6	6	0	60	0

Variable type: numeric

skim_variable	n_missing	complete_rate	mean	sd	p0	p25	p50	p75	p100	hist
quantity	0	1	13.50	13.76	1.00	3.00	6.00	23.00	50.00	
order_time	0	1	12.93	5.50	1.00	9.00	13.00	17.00	23.00	
order_day	0	1	15.50	8.65	1.00	8.00	15.00	23.00	30.00	
order_month	0	1	6.45	3.28	1.00	4.00	6.00	9.00	12.00	
order_year	0	1	2022.46	0.50	2022.00	2022.00	2022.00	2023.00	2023.00	
picking_hours	0	1	14.70	10.39	0.43	9.39	14.05	18.72	45.06	
delivery_hours	0	1	17.48	10.00	0.28	11.55	19.55	25.04	38.05	

Products summary

The product dataset has 60 rows and 5 columns.

Table 1.2

Variable type: character

skim_variable	n_missing	complete_rate	min	max	empty	n_unique	whitespace
product_id	0	1	6	6	0	60	0
category	0	1	5	18	0	6	0
description	0	1	9	21	0	35	0

Variable type: numeric

skim_variable	n_missing	complete_rate	mean	sd	p0	p25	p50	p75	p100	hist
selling_price	0	1	4493.59	6503.77	350.45	512.18	794.18	6416.66	19725.18	
markup	0	1	20.46	6.07	10.13	16.14	20.34	25.71	29.84	

Products head office summary

The head office has its own product data. It consists of 360 rows and 5 columns.

Table 1.3

Variable type: character

skim_variable	n_missing	complete_rate	min	max	empty	n_unique	whitespace
product_id	0	1	5	6	0	110	0
category	0	1	5	18	0	6	0
description	0	1	9	24	0	60	0

Variable type: numeric

skim_variable	n_missing	complete_rate	mean	sd	p0	p25	p50	p75	p100	hist
selling_price	0	1	4410.96	6463.82	290.52	495.94	797.22	5843.33	22420.14	
markup	0	1	20.39	5.67	10.06	15.84	20.58	24.84	30.00	

Customers summary

The customer dataset consists of 5000 rows and 5 columns.

Table 1.4

Variable type: character

skim_variable	n_missing	complete_rate	min	max	empty	n_unique	whitespace
customer_id	0	1	7	8	0	5000	0
gender	0	1	4	6	0	3	0
city	0	1	5	13	0	7	0

Variable type: numeric

skim_variable	n_missing	complete_rate	mean	sd	p0	p25	p50	p75	p100	hist
age	0	1	51.55	21.22	16	33	51	68	105	
income	0	1	80797.00	33150.11	5000	55000	85000	105000	140000	

Analysis

The sales dataset has 100 000 transactions across 2022 and 2023. The customer dataset has 5000 unique customers, which indicates a broad customer base. The product catalogue has around 60 unique products (branch) and more in the head office list (110).

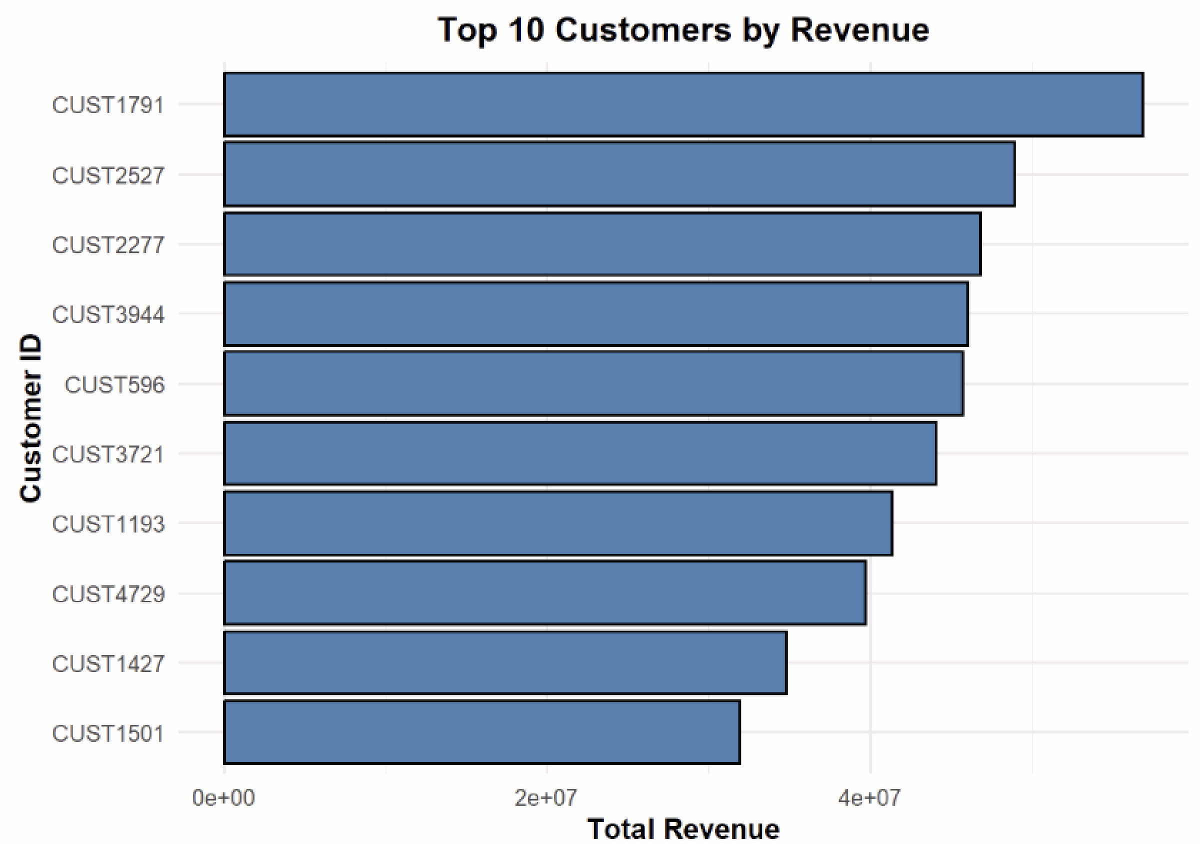
The company thus has a solid customer base and a manageable product portfolio.

There are no missing values in any of the data sets that need to be addressed.

Data Visualisation and Analysis

For this analysis, the products_data file was preferred over products_headoffice because it reflects the actual prices and product details recorded in sales transactions. This ensures that revenue and profitability measures are based on real store activity rather than head office reference values, which sometimes differ. Products_data is simply referred to as products from here on.

Top 10 customers by revenue



Graph 1.1

As seen on graph 1, the top 10 customers account for a large share of the company's revenue.

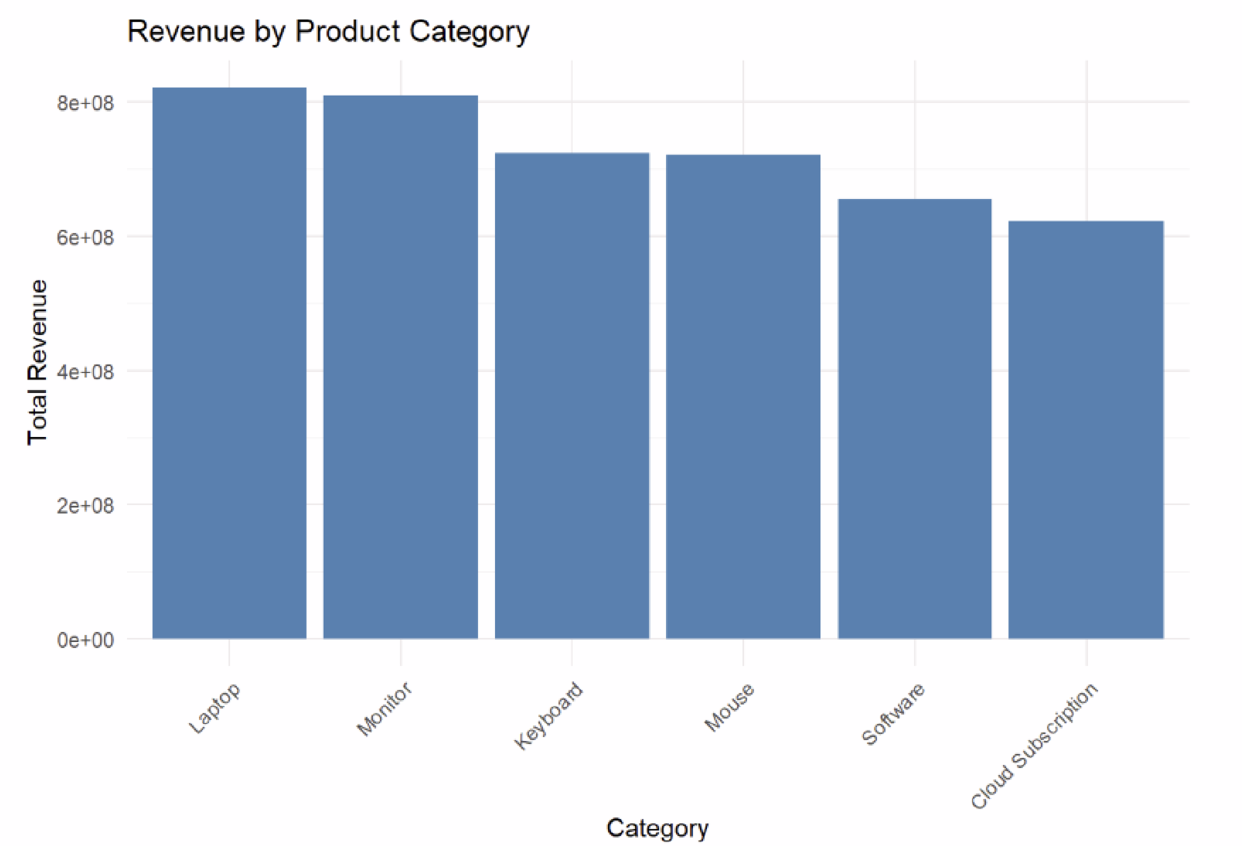
These customers are highly valuable to the company and losing them could significantly impact the company's revenue. It would be a good idea for management to consider introducing loyalty programs or priority service for the top customers to retain them.

Revenue by Product Category

Table 1.5

Revenue by Product Category

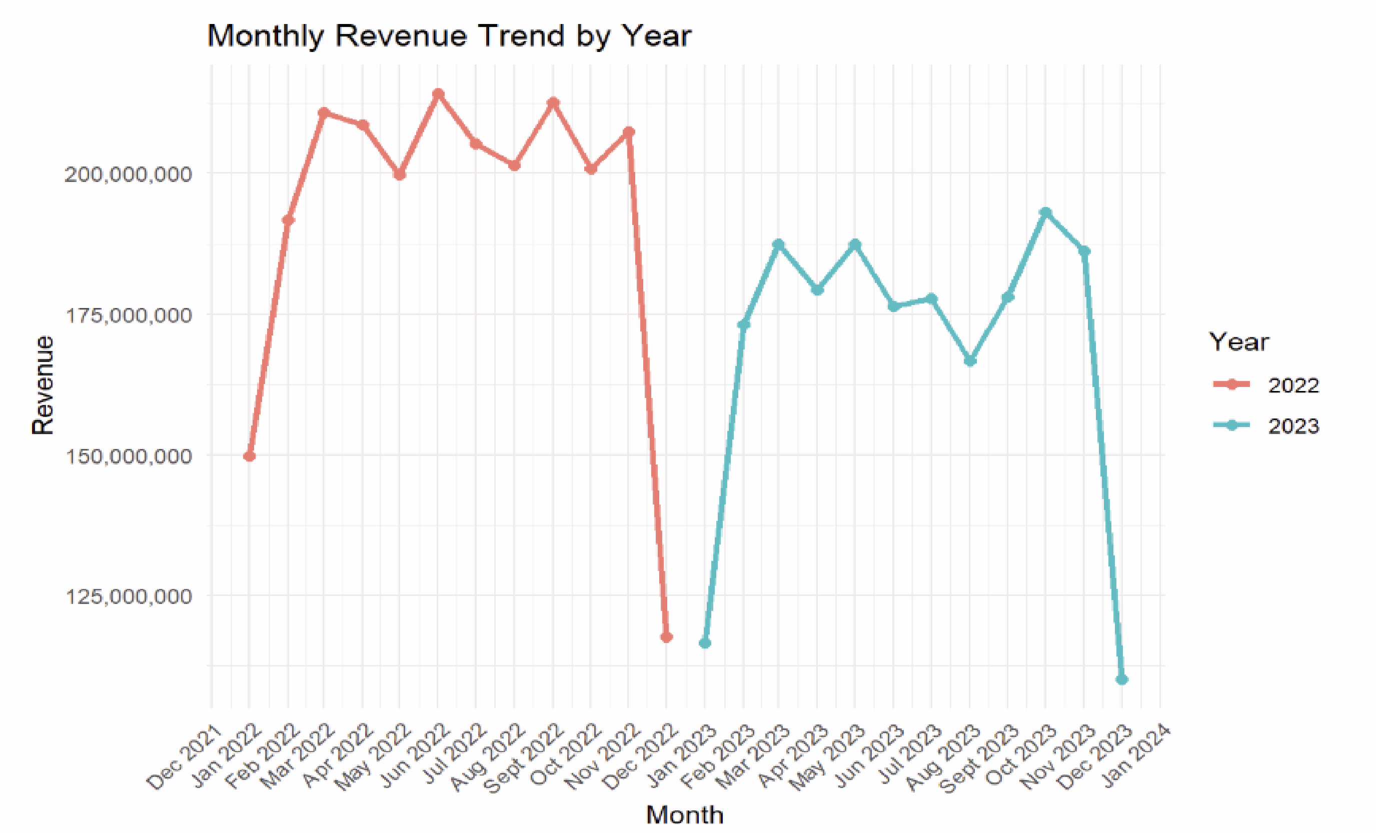
category	total_revenue	total_quantity
Laptop	821533851	220867
Monitor	809104952	231513
Keyboard	723693159	225067
Mouse	721090260	222350
Software	655365933	225805
Cloud Subscription	621799523	224745



Graph 1.2

Graph 1.2 shows that the highest revenue is obtained from laptops and monitors, but all the products are selling in large quantities (Table1.5) and are bringing in high revenues. Hardware drives the bulk of revenue. However, software and cloud subscriptions still represent recurring income streams that could be increased through promotions.

Revenue Trends



Graph 1.3

The year of 2022 consistently has higher revenues than 2023 (Graph 1.3). Both 2022 and 2023 do, however, follow similar seasonality. In both years, the revenues are significantly lower for the months of January and December.

Management should align marketing campaigns and stock planning with seasonal peaks. They should also consider special promotions in slower months (January and December) to stabilise cash flow.

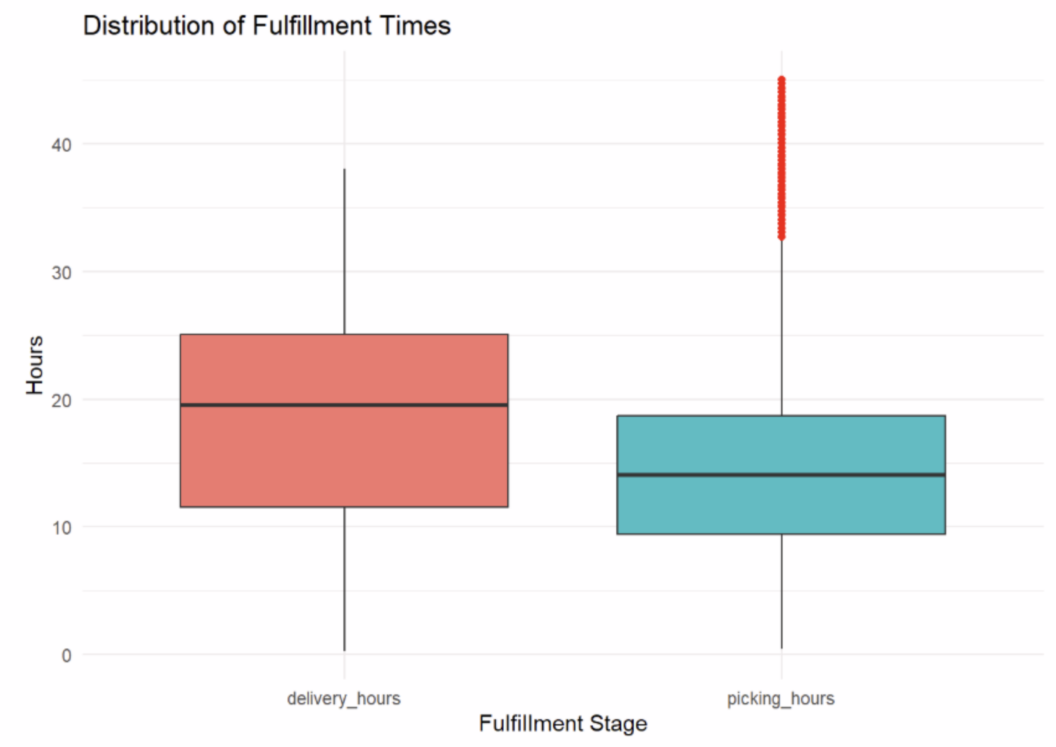
Fulfilment of orders

This boxplot of picking/delivery hours shows how long, on average, it takes the company to fulfil customer orders, and how variable that process is.

Table1.6

Summary Statistics for Picking vs Delivery Hours

stage	mean_hours	median_hours	sd_hours
delivery_hours	17.47646	19.546	9.999944
picking_hours	14.69547	14.055	10.387334



Graph 1.4

Median delivery time is about 19.546 hours whereas the median picking time is about 14.055 hours (Table 1.6). On average, delivery adds about 3 hours more than picking to order fulfilment.

When it comes to variation, delivery has a slightly tighter spread (sd ≈ 10.0 hours) compared to picking (sd ≈ 10.4 hours). Picking also has more extreme outliers, indicated in Graph 1.4 (in red), with some orders taking over 40 hours.

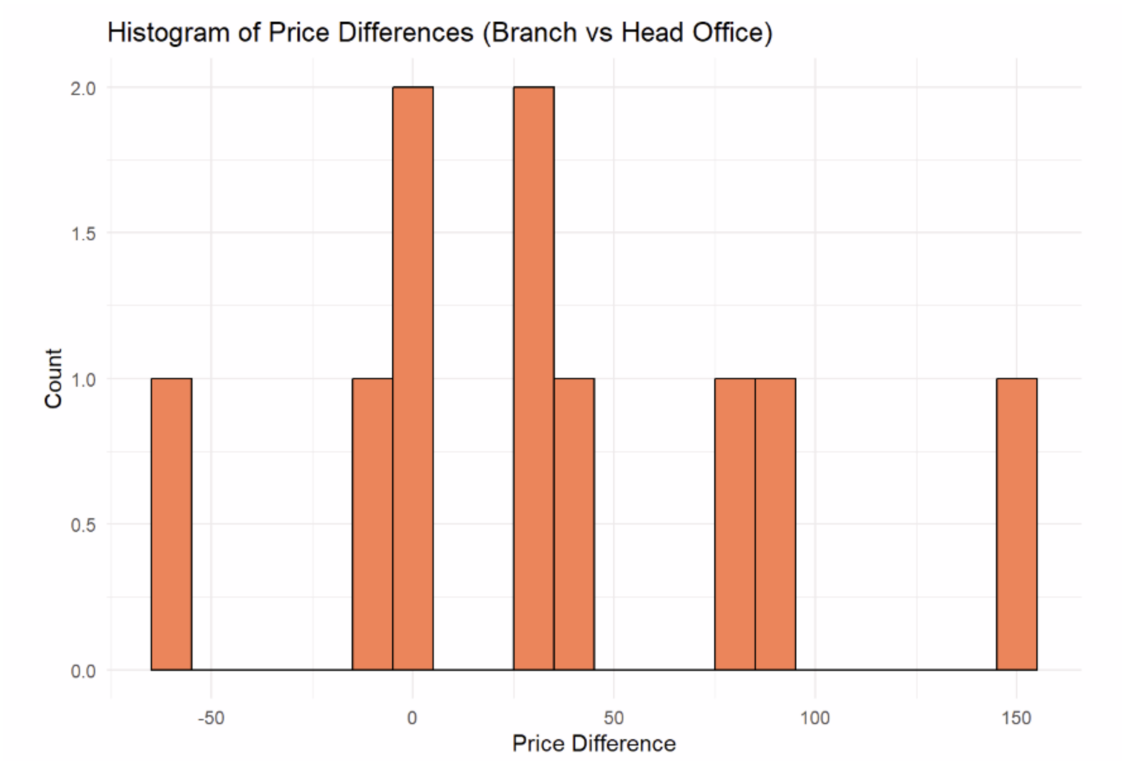
Delivery is the bottleneck, and management should consider investing in logistics or faster delivery options to decrease the time. Management should also attempt to reduce variability in picking by investigating the root causes of outliers.

Price Discrepancy Analysis

Table 1.7

Top 10 Price/Markup Discrepancies

product_id	category	description_branch	selling_price_branch	markup_branch	description_ho	selling_price_ho	markup_ho	price_diff	markup_diff
SOF004	Monitor	blue silk	542.56	17.19	black marble	389.33	17.25	153.23	-0.06
SOF009	Laptop	azure sandpaper	540.41	11.34	black bright	452.40	19.64	88.01	-8.30
SOF008	Cloud Subscription	burlywood silk	549.02	11.95	cornflowerblue marble	465.73	21.89	83.29	-9.94
SOF006	Mouse	black silk	478.93	16.99	cornflowerblue marble	539.33	25.57	-60.40	-8.58
SOF002	Cloud Subscription	cyan silk	505.26	10.43	black silk	466.95	28.42	38.31	-17.99
SOF005	Keyboard	aliceblue wood	516.15	11.01	chartreuse sandpaper	482.64	17.60	33.51	-6.59
SOF007	Software	black bright	527.56	16.79	blue marble	495.13	10.23	32.43	6.56
SOF001	Software	coral matt	511.53	25.05	coral silk	521.72	15.65	-10.19	9.40
SOF003	Laptop	burlywood marble	493.69	16.18	burlywood marble	496.43	20.07	-2.74	-3.89
SOF010	Monitor	chocolate sandpaper	396.72	23.47	cornflowerblue matt	399.43	17.08	-2.71	6.39



Graph 1.5

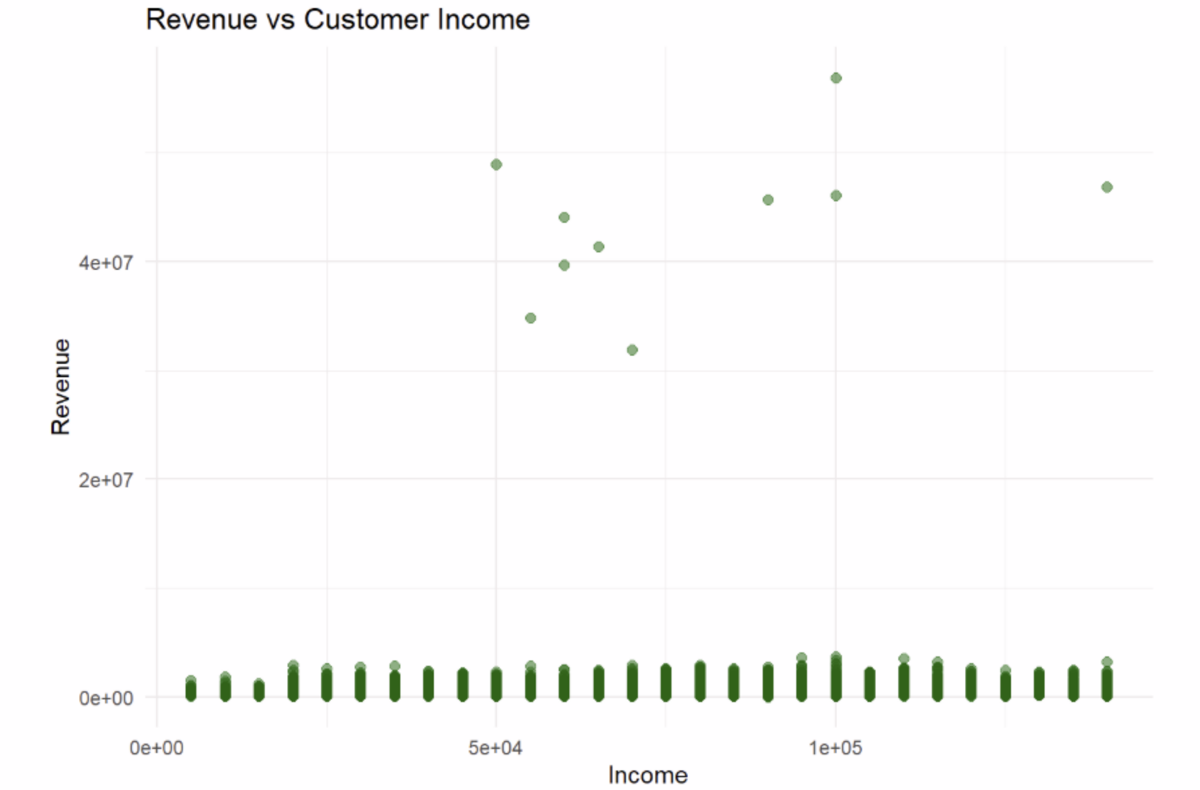
Graph 1.5 shows that there are indeed discrepancies in the prices of products between the branch and head office. According to Table 1.7, the 10 products with the largest differences in prices are all software (SOF) products. There are also differences in the markup percentages between the branch and head office.

Management needs to standardise pricing and markup policies as inconsistent pricing can confuse customers and employees, reduce trust, and lead to missed revenue opportunities. They also need to ensure that the information is up to date and correct any mistakes in the data.

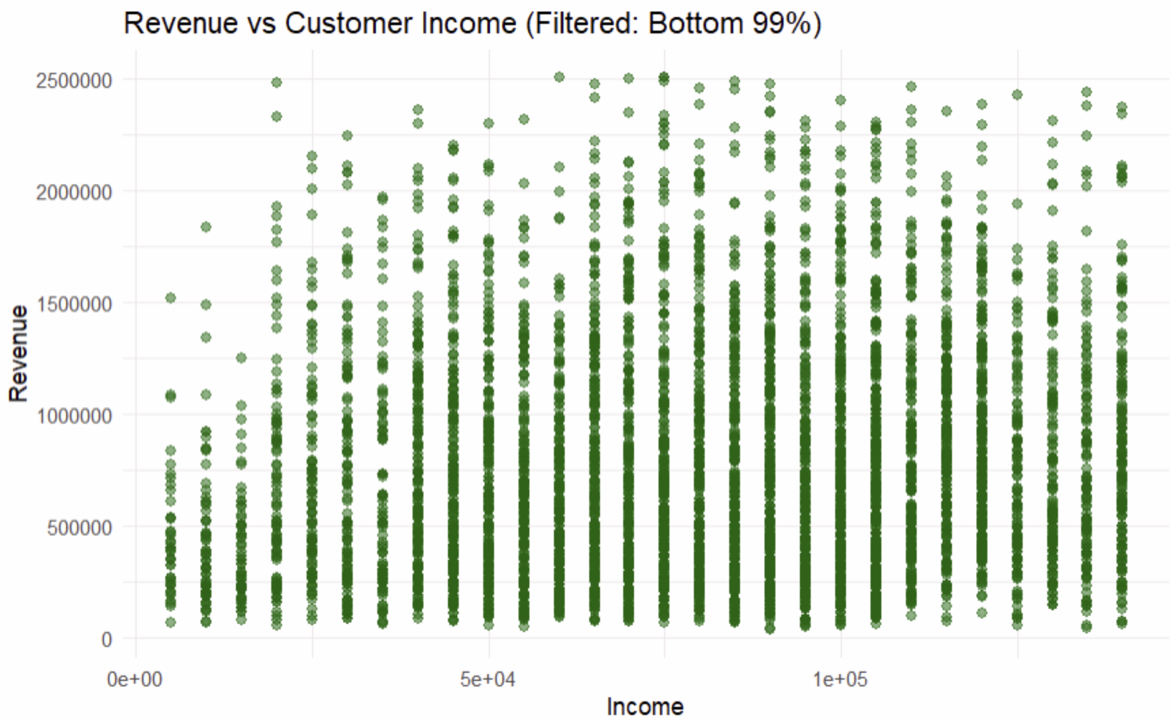
Customer Demographics

Graph 1.6

The revenue distribution by customer gender in Graph 1.6 shows that the revenue is relatively balanced across the different genders. The products are thus deemed accessible and desirable to all genders.



Graph 1.7

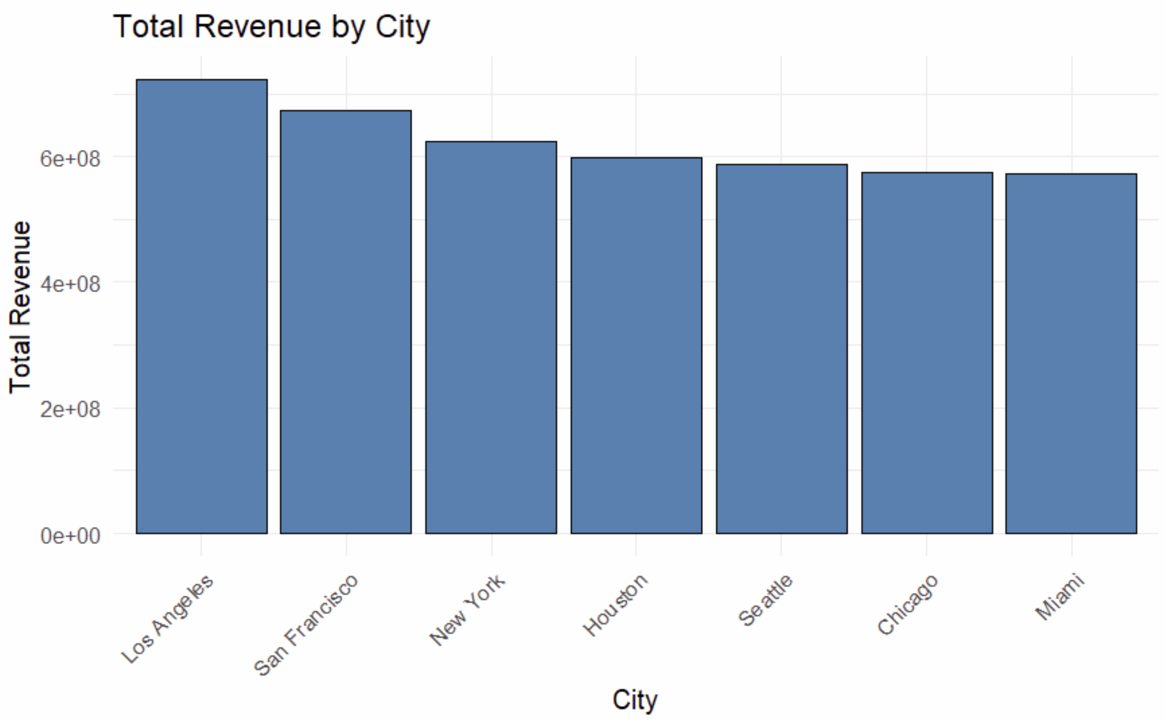


Graph 1.8

The revenue vs customer income in Graph 1.7 shows several outliers representing customers that spend substantially more money than the average customer spends.

In Graph 1.8, the outliers were filtered to improve visibility of the overall trend. This graph shows that there is not much difference between the amount of revenue from middle-income customers and high-income customers, but lower-income customers tend to spend less.

This graph shows management that the products are generally affordable to people with varying income levels.



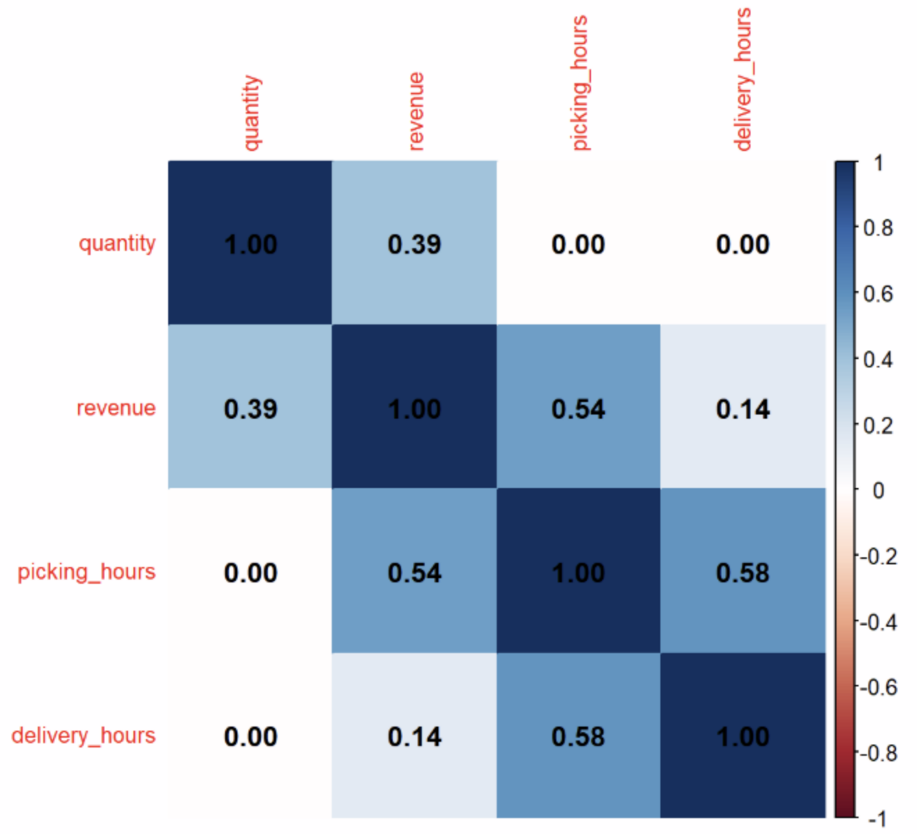
Graph 1.9

The revenue analysis by city (Graph 1.9) shows that Los Angeles and San Francisco are the company's strongest markets, generating the highest total revenues of approximately 722M and 674M respectively. This highlights them as critical regions for maintaining customer relationships and supply efficiency.

The other cities (New York, Houston, Seattle, Chicago, and Miami) also contribute significantly, though slightly below the top two, suggesting consistent sales activity.

This geographic distribution indicates that management should prioritise inventory planning, promotional campaigns, and customer engagement strategies in the top-performing cities, while exploring growth opportunities in mid-performing regions to balance overall market performance.

Correlation Analysis



Graph 1.10

The correlation matrix (Graph 1.9) shows a moderate positive correlation between quantity and revenue. This indicates that when more units are sold, revenue tends to increase, although the relationship is not perfectly proportional. This makes sense as there is price variation among the different products, as some items have higher value per unit.

The correlation analysis further reveals that revenue has a moderate positive relationship (r = 0.54) with picking hours, indicating that higher-value orders demand more preparation time. However, delivery hours show only a weak relationship with revenue (r = 0.14), suggesting that delivery efficiency remains relatively stable regardless of order size.

The number of items sold (quantity) shows no correlation with fulfilment times, implying that order complexity or product type, rather than quantity, drives resource usage.

These insights can guide process improvements, particularly in the picking stage for high-value orders.

Key Insights and Recommendations

The analysis highlights several important insights for management decision-making. Top products and customers contribute more to total revenue, suggesting that inventory planning and promotional efforts should focus on these key drivers. Revenue trends reveal seasonal fluctuations, providing an opportunity to align marketing campaigns with high-revenue months to maximise impact. In terms of fulfilment efficiency, while average picking and delivery times are within reasonable ranges, the presence of extreme outliers indicates potential bottlenecks that should be investigated to improve consistency and reduce delays. Price discrepancies between branch and head office catalogues were identified, and these should be reviewed and corrected to ensure consistency across the business.

Part 1.2 (4.3)

Introduction

Differences between the product data and head-office data were identified. The products head-office data was subsequently corrected according to the instructions sent by the head office. Part 1.2 will address how the outcomes of Part 1.1 has changed and what the new results mean for management.

Summary Statistics

All the summary statistics stayed the same except for the products head office summary, which is to be expected.

Table 1.8

Comparison of Head Office Product Data: Before vs After Corrections

Metric	Before	After
Total Rows	360.00	360.00
Unique Product IDs	110.00	360.00
Unique Categories	6.00	6.00
Unique Descriptions	60.00	60.00
Mean Selling Price	4410.96	4493.59
Std. Dev. Selling Price	6463.82	6458.32
Mean Markup	20.39	20.46
Std. Dev. Markup	5.67	6.03

Table 1.8 shows the differences in the summaries of the head office product data before and after After the Head Office product data was corrected, the number of unique product IDs increased from 110 to 360, confirming that missing or duplicated product entries were successfully expanded to represent all inventory records.

The selling_price and markup values were also restored. Previously, these fields contained incomplete or misaligned data; after correction, both now display realistic ranges and variability (average selling price R4 493.59 with a standard deviation of R6 458.32, and markups averaging 20%). These corrections substantially improved the reliability of subsequent revenue analyses.

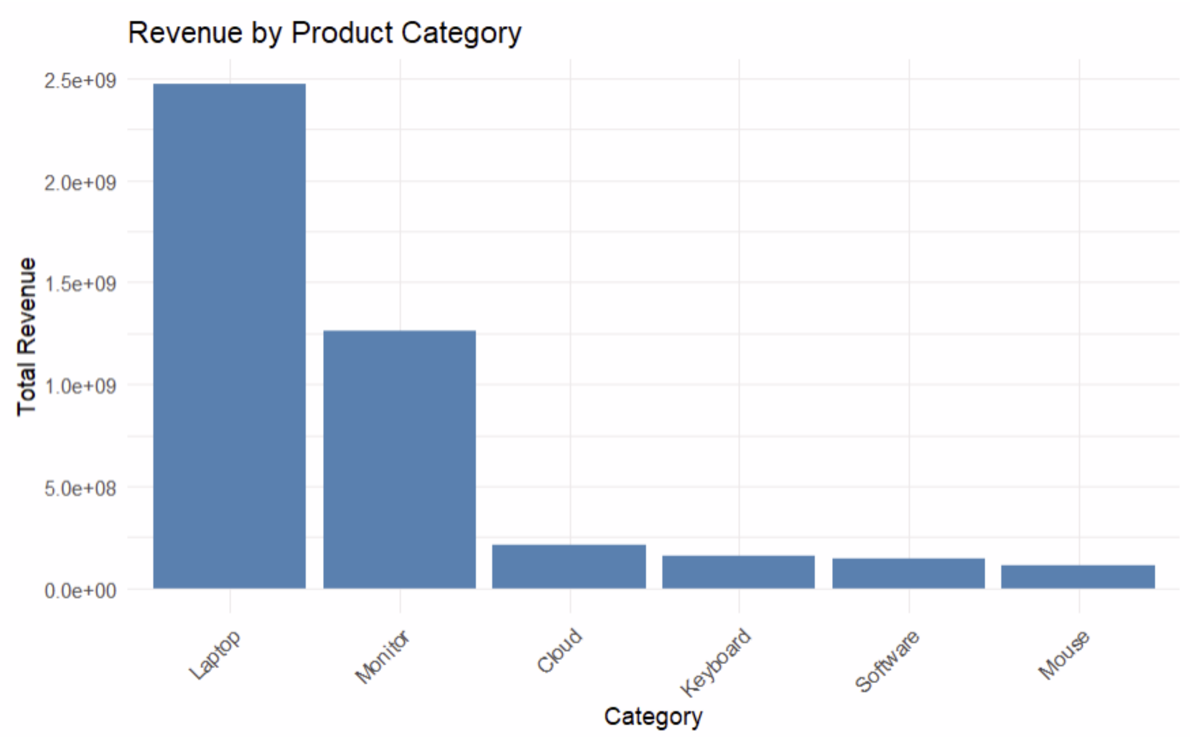
Data Visualisation and Analysis

Similar Results

The Top 10 customers by revenue remained unchanged after the correction. Revenue trends and the distribution of fulfilment times also did not change. All the customer demographics and the correlation matrix remained the same.

Different Results

The revenue by product category changed drastically. The new revenue by product category is shown in Graph 1.11. The differences between the revenue and quantity of units sold before and after the data correction are illustrated in Table 1.9 and Graph 1.12.

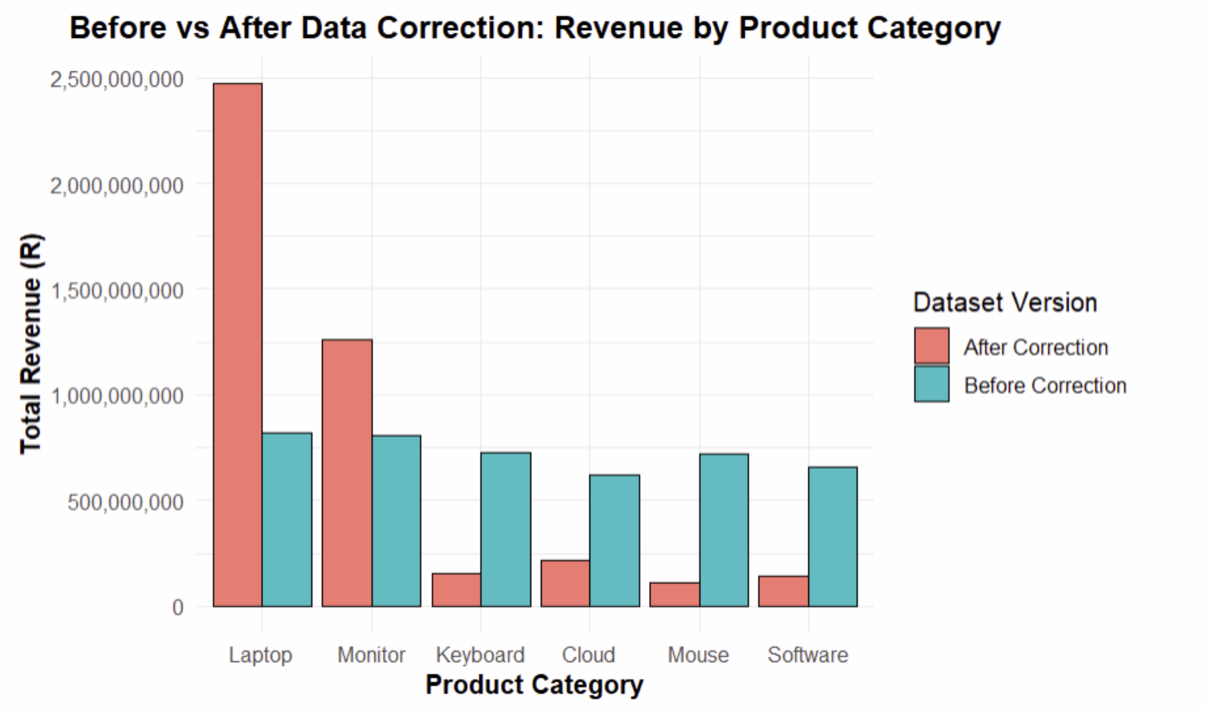


Graph 1.11

Table 1.9

Before vs After Data Correction: Revenue by Product Category

category	total_revenue_before	total_quantity.x	total_revenue_after	total_quantity.y	change_abs
Laptop	821533851	220867	2470814376	136721	1649280524
Monitor	809104952	231513	1258942847	199691	449837895
Cloud	621799523	224745	214110418	210007	-407689105
Keyboard	723693159	225067	155002210	240925	-568690949
Software	655365933	225805	142527355	281703	-512838578
Mouse	721090260	222350	111190471	281300	-609899789

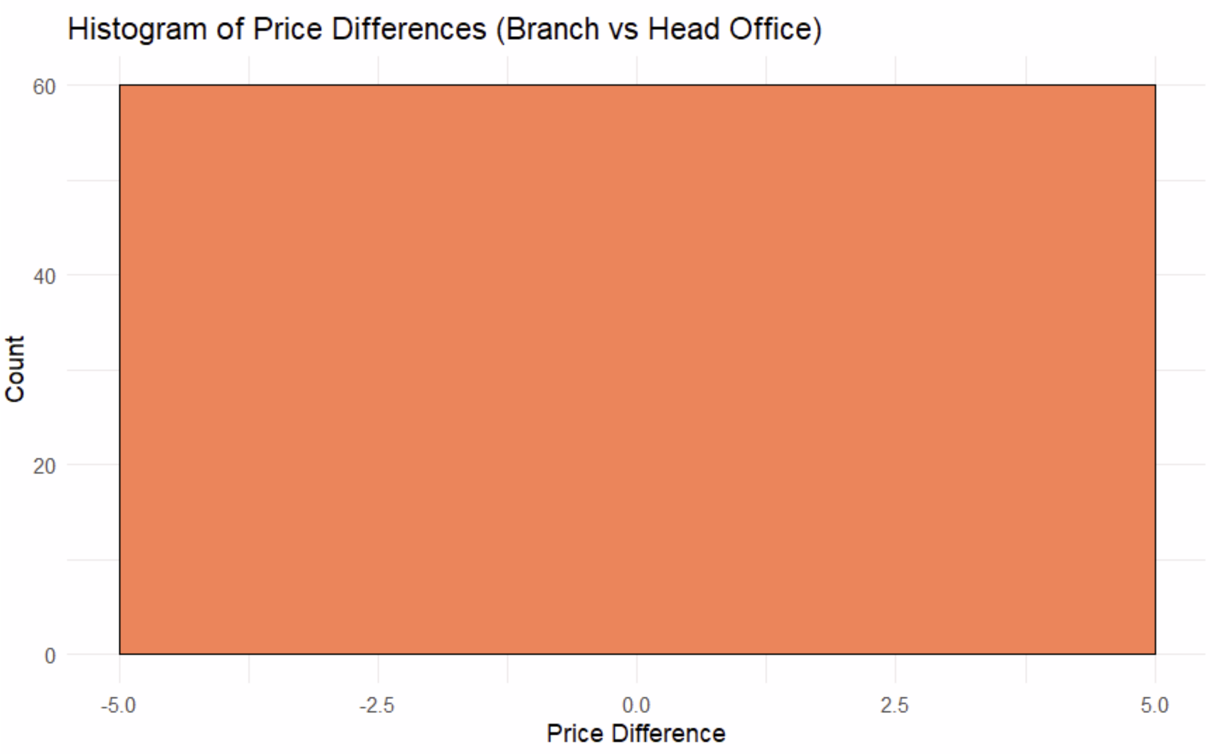


Graph 1.12

As Graph 1.12 shows, before the data correction, the revenue by category appeared evenly distributed across all products, suggesting inconsistencies in pricing and markup information. After the Head Office data correction, the revenue distribution presents a more realistic picture of product performance. These corrections significantly improve the accuracy of performance analysis and financial reporting, enabling better resource allocation and strategy decisions.

Graph 1.11 illustrates that Laptops (R2.47B) and Monitors (R1.26B) dominate the overall revenue, confirming that the business's profitability is primarily driven by high-value hardware. Table 1.9 shows that the accessories, such as Keyboards, Mice, and Software, generate high unit volumes but lower revenue, indicating their role as complementary rather than profit-driving products. Cloud services show promising steady adoption and represent a strategic opportunity for recurring income.

Management should focus on hardware profitability while expanding Cloud offerings to diversify and stabilise revenue streams. Management can try bundling cloud subscriptions with hardware purchases to strengthen long-term customer value.



Graph 1.13

The new price discrepancy analysis (Graph 1.13) confirms that the corrected head office products data now accurately corresponds to the branch products data and there are no more discrepancies between the two data sets.

Conclusion

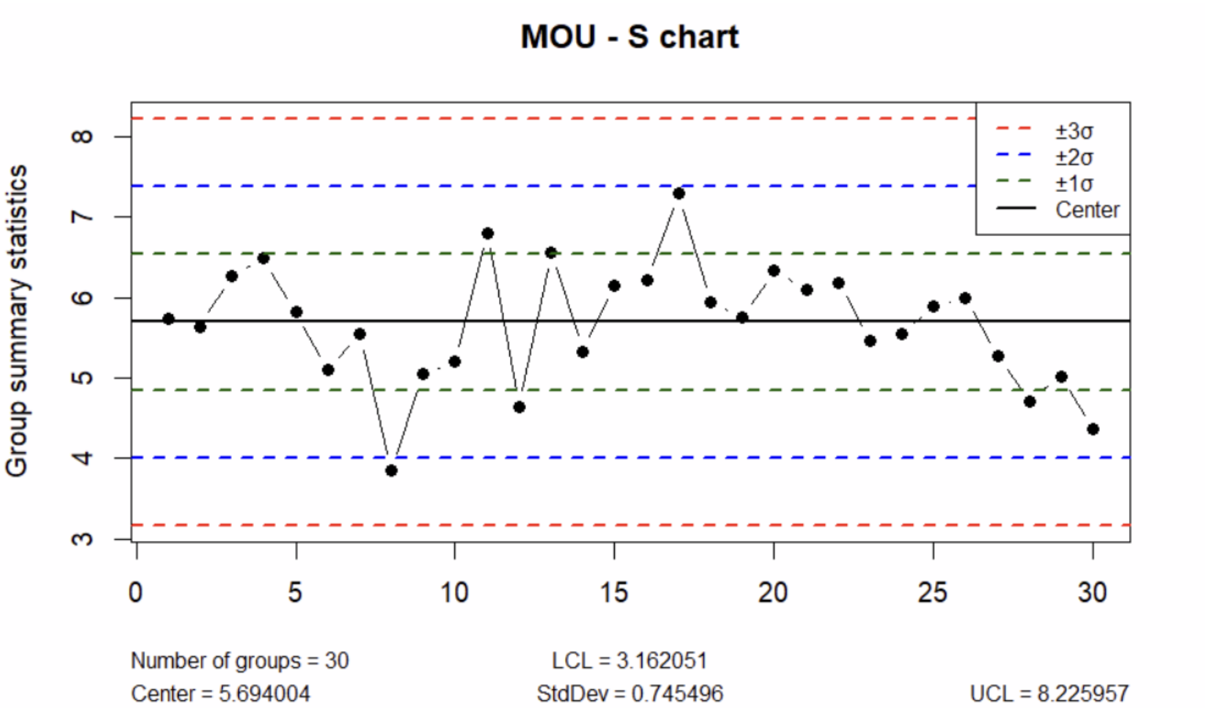
The corrected data provides a more accurate reflection of company performance, highlighting that laptops and monitors are the main revenue drivers, while accessories and software contribute smaller, supporting shares. Cloud services show steady growth and present an opportunity for recurring income. Overall, the corrected dataset enables more reliable insights to guide future pricing, product focus, and strategic decisions.

Part 2 and 3

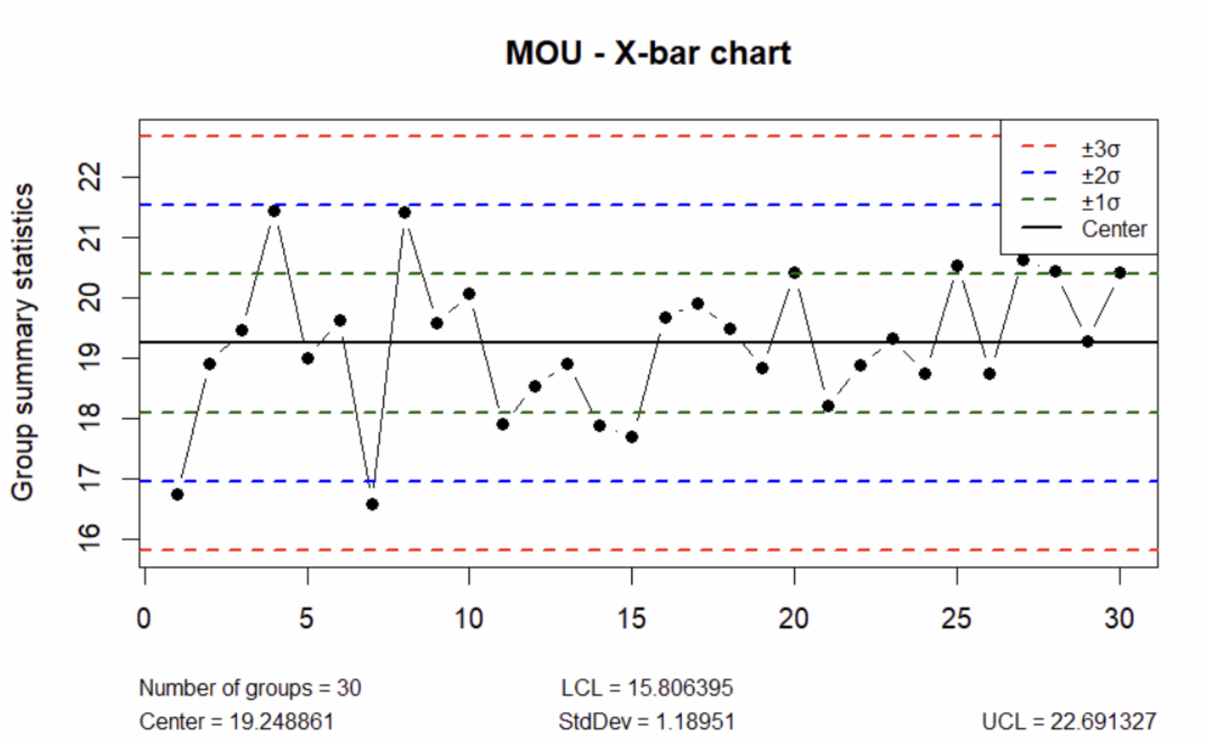
Control Charts

To monitor the stability of delivery time performance across all product types, Statistical Process Control (SPC) charts were constructed for both process variation (S-charts) and process means (X-bar charts). Each product type was analysed using 30 initial subgroups of 24 samples to establish baseline control limits and centre lines. The S-charts assess the stability of process variation, while the X-bar charts track shifts in process averages over time. Together, these charts provide a comprehensive view of whether each process operates under statistical control or shows signs of special-cause variation.

Mouse Product



Graph 2.1



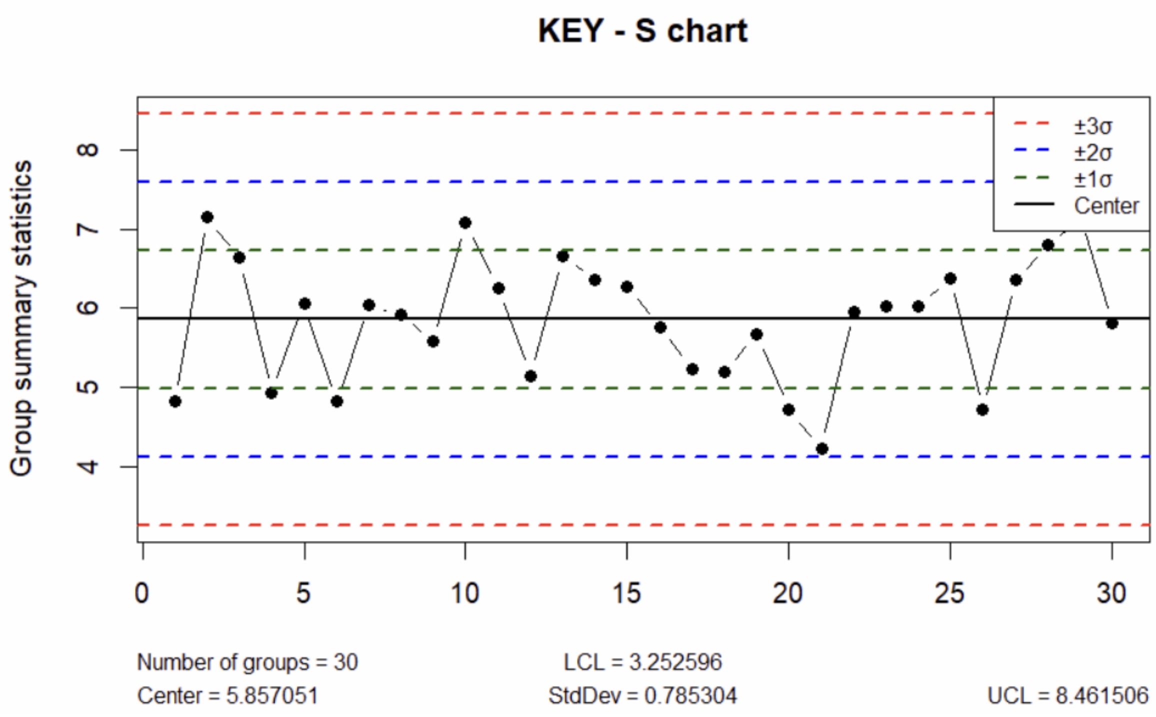
Graph 2.2

The MOU S-chart (Graph 2.1) shows that all 30 subgroup standard deviations lie within the 3-sigma control limits (LCL = 3.162, UCL = 8.226), and the points fluctuate randomly around the centre line (≈ 5.69). This indicates that process variation is stable.

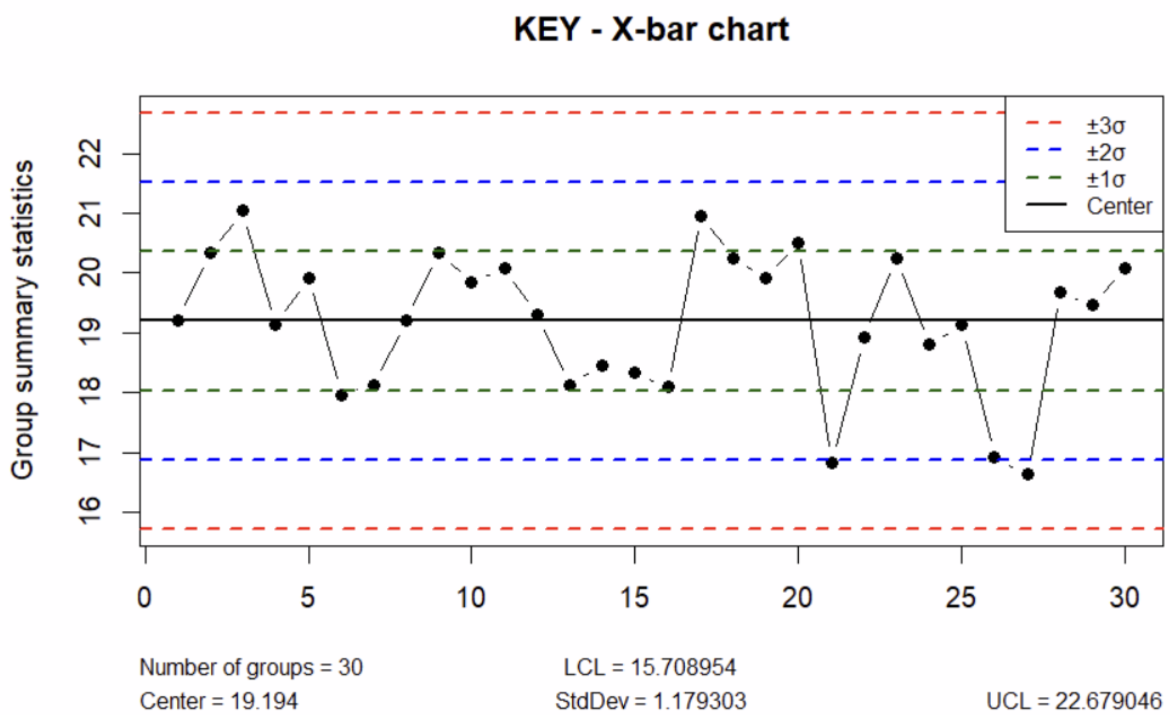
The X-bar chart (Graph 2.2) shows that no points appear to fall outside the ±3σ limits (LCL = 15.81, UCL = 22.69). However, the X-bar chart shows visible shifts and runs around the centre line (≈ 19.25), suggesting that the process average fluctuates over time (Graph 2.2).

These charts imply that while the process variation is under control, the process mean is not fully stable and may be influenced by special causes such as workload or scheduling changes

Keyboard Product



Graph 2.3



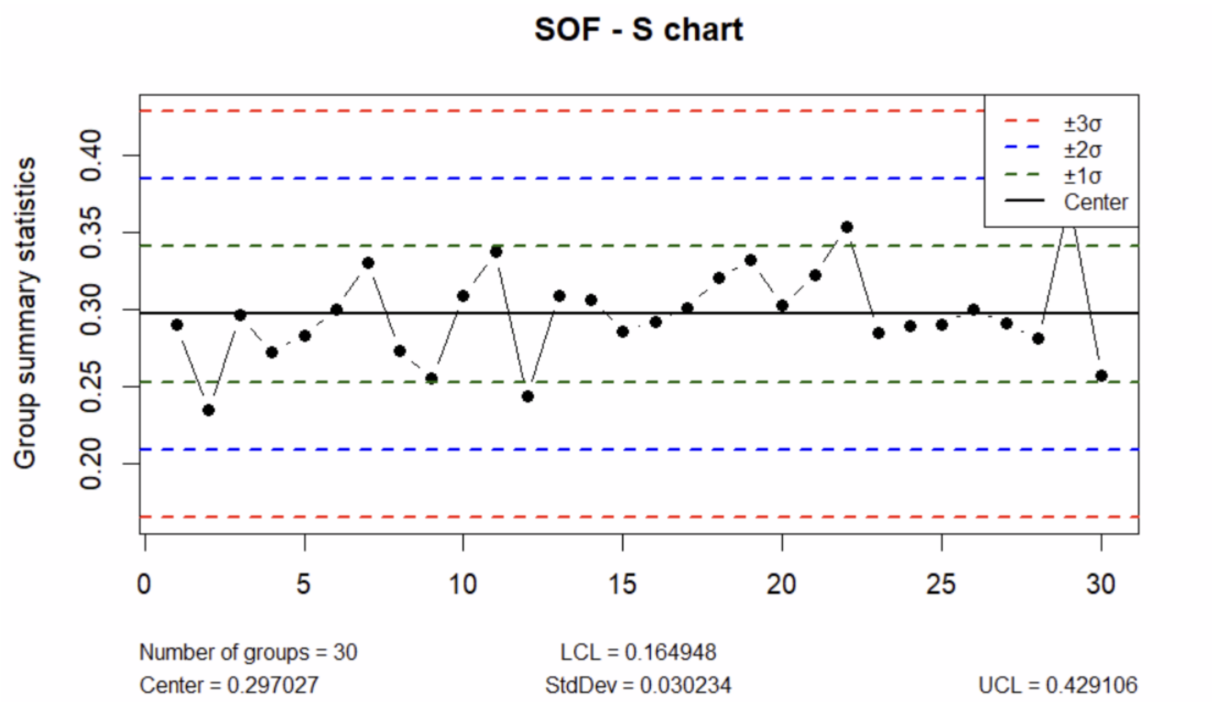
Graph 2.4

For the Keyboard (KEY) product, the S-chart (Graph 2.3) shows that all 30 points fall between the ±3σ limits, where LCL = 3.25 and UCL = 8.46. The points fluctuate randomly around the centre line (≈5.86).

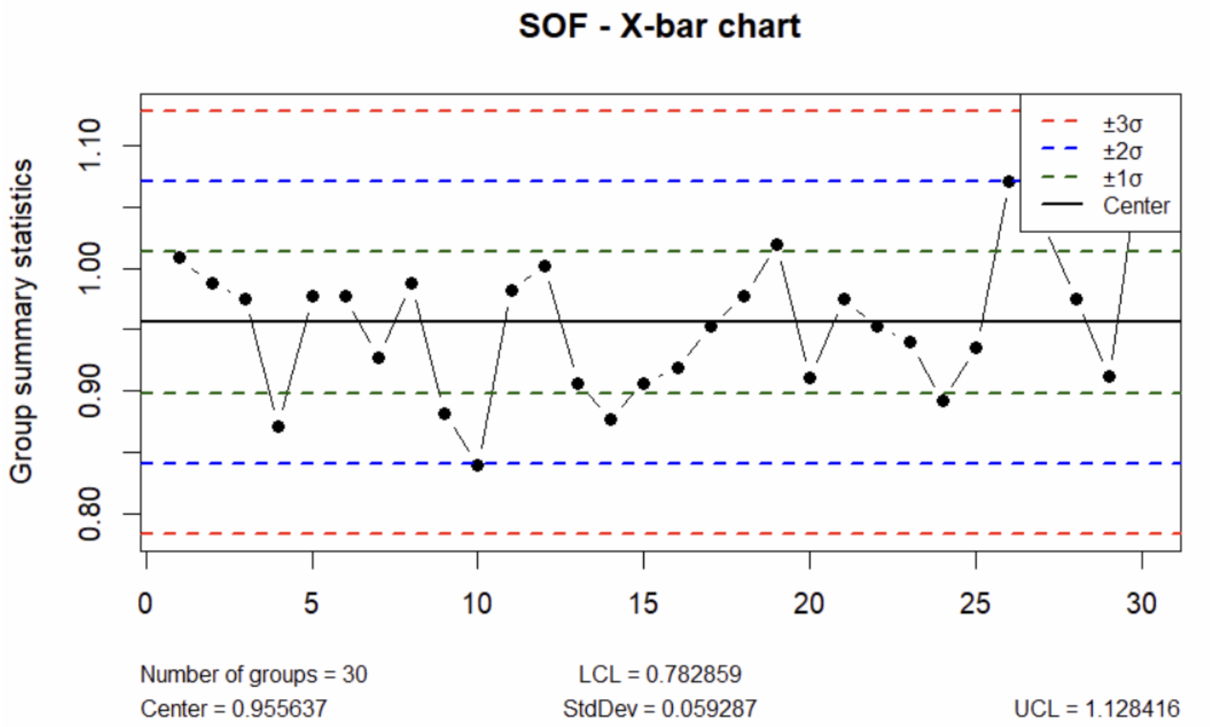
In the X-bar chart (Graph 2.4) all 30 points are also within the control limits (LCL = 15.71, UCL = 22.68). The mean values fluctuate around the centre line (≈ 19.19), with groups of consecutive points above or below the mean.

The S-chart thus indicates stable and consistent variation, with no out-of-control points. The X-bar chart shows small but noticeable runs and mean shifts around the centre line, suggesting that while the process variation is predictable, the average delivery time is not fully stable. This reflects the presence of special causes that occasionally affect the process mean.

Software Product



Graph 2.5



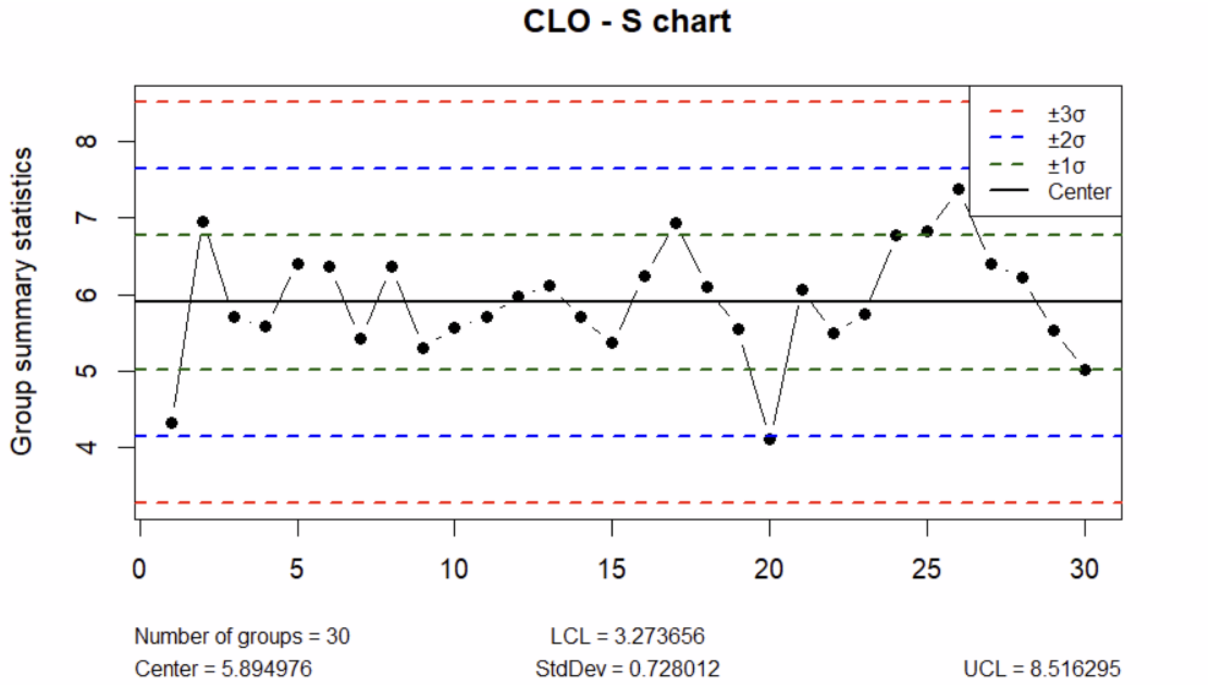
Graph 2.6

The Software (SOF) S-chart (Graph 2.5) shows that all 30 subgroup standard deviations lie well within control limits (LCL = 0.165, UCL = 0.429). The points fluctuate tightly around the centre line (≈ 0.297), with no visible patterns or trends. The process variation is thus stable and consistent.

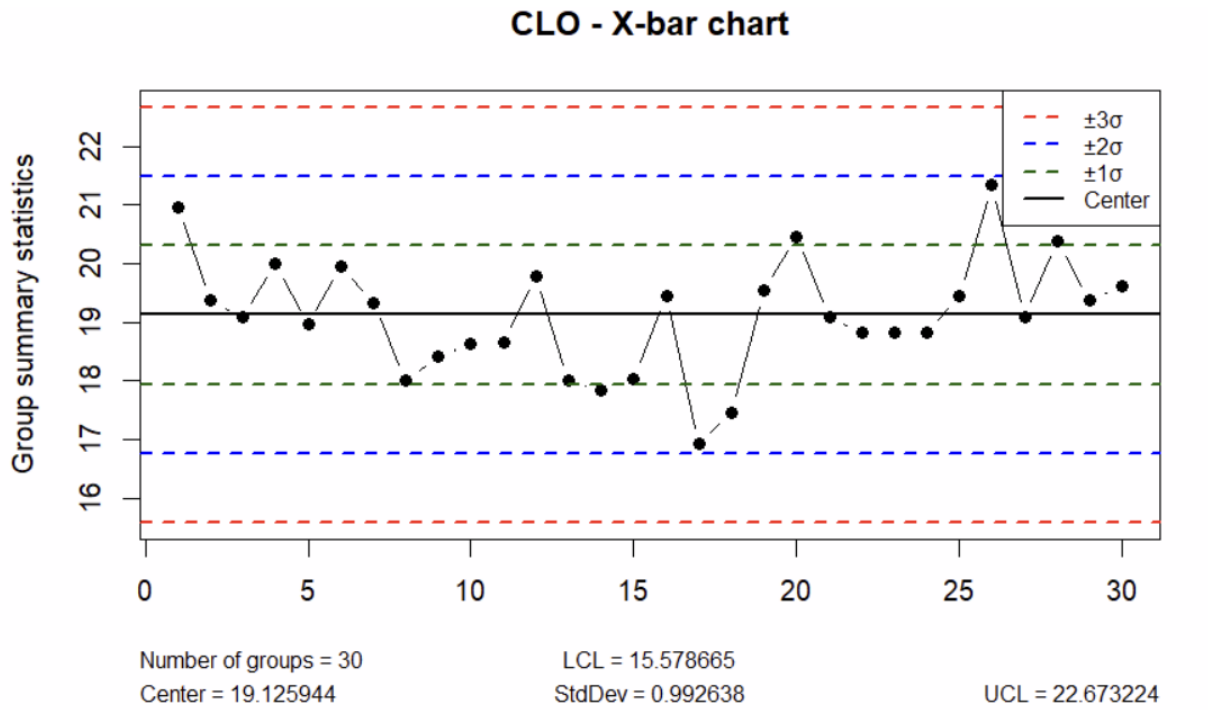
The X-bar chart (Graph 2.6) shows all 30 subgroup means within control limits (LCL = 0.783, UCL = 1.128). The subgroup means fluctuate randomly around the centre line (≈ 0.956), with no runs or trends. The process mean is thus stable and consistent.

The Software (SOF) process shows both variation stability and mean stability. All points lie well within control limits on both charts, and variation is minimal. This indicates that software delivery is predictable, consistent, and capable of meeting delivery-time requirements.

Cloud Product



Graph 2.7



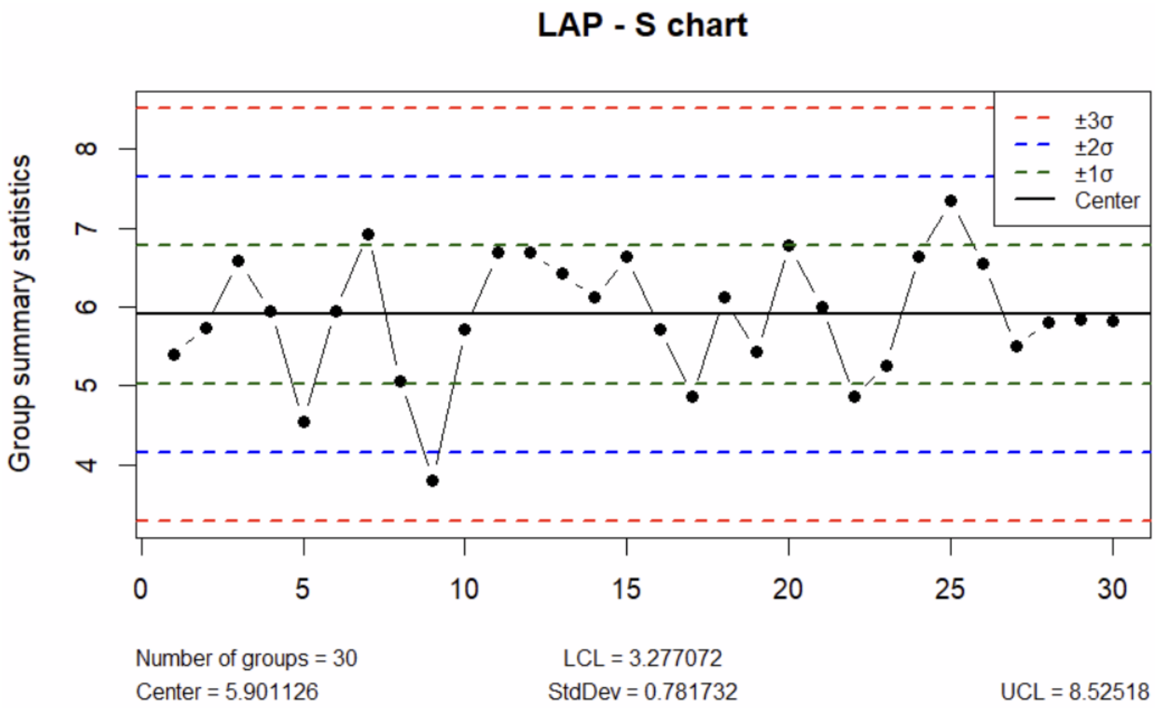
Graph 2.8

For the Cloud (CLO) product, the S-chart (Graph 2.7) shows that all 30 subgroup standard deviations are within control limits (LCL = 3.27, UCL = 8.52). The variation values fluctuate randomly around the centre line (≈ 5.89), without any points exceeding the ±3σ limits. The spread is thus stable and predictable, meaning there are no sudden increases or decreases in process variability.

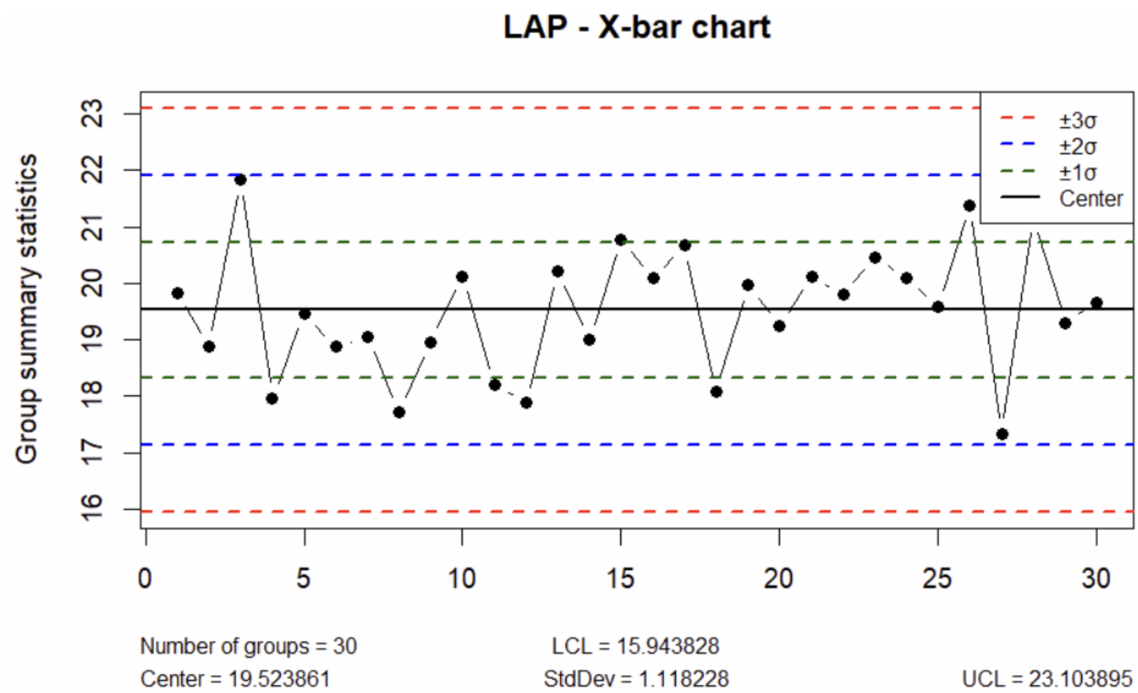
The X-bar chart (Graph 2.8) shows all 30 subgroup means within control limits (LCL = 15.58, UCL = 22.67) and the process mean fluctuates around the centre line (≈ 19.13). The process mean is thus stable.

Both S and X-bar charts indicate that the process is stable and in control. All points fall within control limits, showing consistent variation and no evidence of special-cause variation.

Laptop Product



Graph 2.9



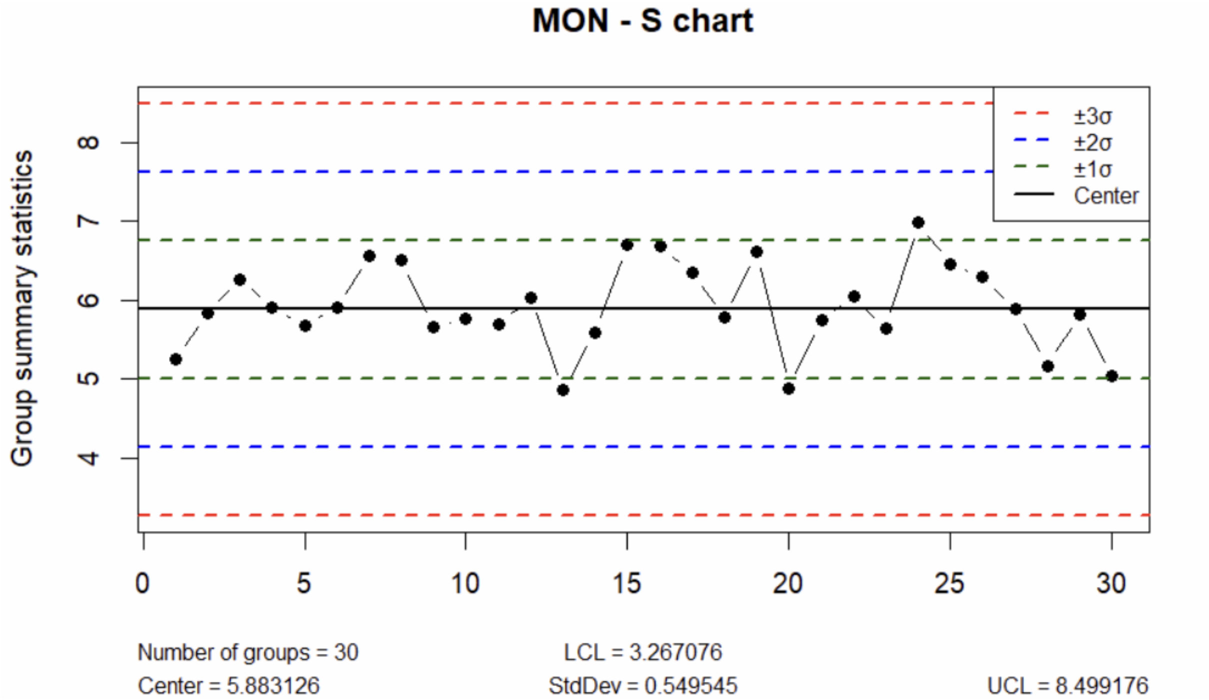
Graph 2.10

The Laptop (LAP) S-chart (Graph 2.9) shows that all 30 subgroup standard deviations lie within control limits (LCL = 3.277, UCL = 8.525). The points fluctuate around the centre line (≈ 5.901), with no points exceeding the control limits and no clear patterns or trends. The process variation is therefore stable and consistent over time.

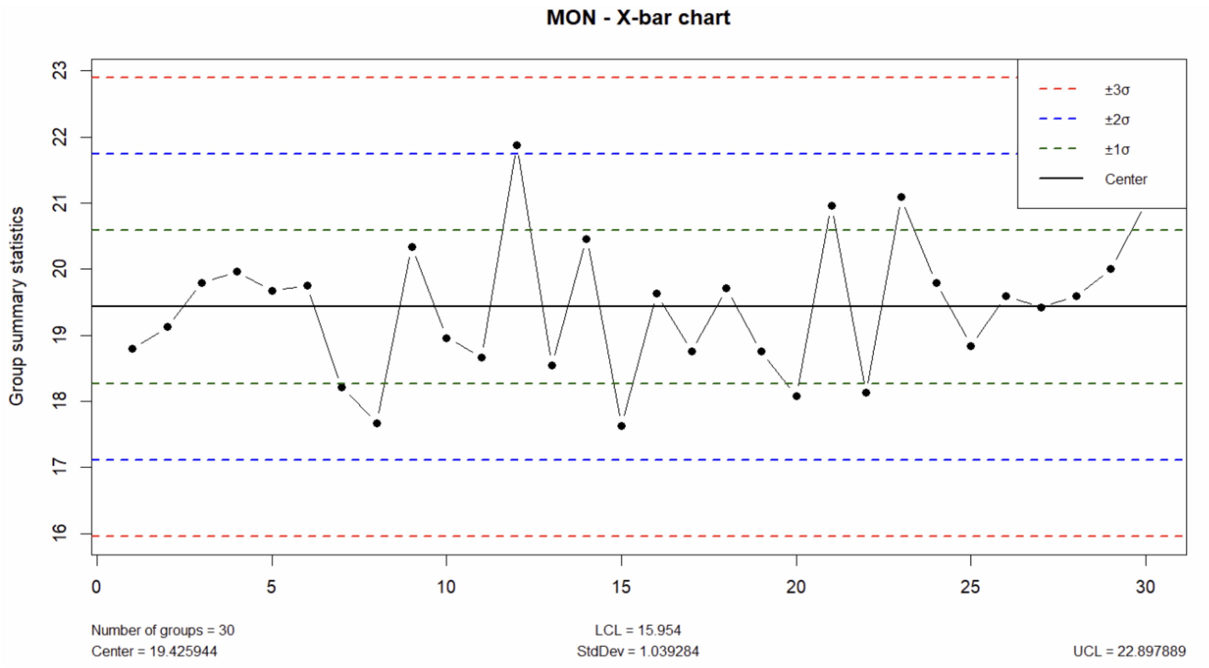
The X-bar chart (Graph 2.10) shows that all 30 subgroup means are within control limits (LCL = 15.944, UCL = 23.104). The subgroup means fluctuate around the centre line (≈ 19.524). No distinct runs or shifts are evident, indicating that the process mean remains mostly stable.

The Laptop (LAP) process thus demonstrates both variation and mean stability. All points remain within control limits on both charts, confirming that the process is statistically in control. Overall, the process is predictable and operates with consistent performance.

Monitor Product



Graph 2.11



Graph 2.12

The Monitor (MON) S-chart (Graph 2.11) shows that all 30 subgroup standard deviations lie within the control limits (LCL = 3.267, UCL = 8.499). The points fluctuate randomly around the centre line (≈ 5.883) with no points exceeding the limits or displaying visible patterns. This indicates that the process variation is stable and consistent over time.

The X-bar chart (Graph 2.12) shows that all 30 means fall within the control limits (LCL = 15.954, UCL = 22.898). The means fluctuate around the centre line (≈ 19.426). These variations remain within acceptable control boundaries and are likely caused by common-cause variation.

The Monitor (MON) process therefore demonstrates stable variation and a stable mean. All points remain within control limits on both charts, confirming that the process is statistically in control. Overall, the process operates predictably with consistent performance.

Combined Analysis

Table 2.1

Product <chr>	ChartType <chr>	Total_Out_of_Control <int>
SOF	X-bar	320
MOU	S	1
MOU	X-bar	295
MON	X-bar	190
LAP	S	1
LAP	X-bar	115
KEY	X-bar	264
CLO	X-bar	226

According to Table 2.1, all six products show a lot of out-of-control sample points in the X-bar charts, which means that for all the products the average performance of the processes keep changing over time. This indicates that the processes are not stable or consistent.

The S-charts generally do not have out-of-control points, except for MOU and LAP that each have 1 out-of-control point in the S-charts (Table 2.1). The fact that the points are mostly in-control indicates that the amount of variation in the processes is consistent and stable over time.

The systems do not have problems with variability, but rather with the process average shifting. The process is thus repeatable in its spread indicating that the measurements and day-to-day fluctuations are behaving consistently. However, the average delivery time frequently shifts up or down, which means that the processes are being affected by special causes (unusual, non-random events), such as bulk orders, staffing shortages, courier delays etc.

Management should investigate what changes occurred in the system whenever the out-of-control points appeared.

Process Capability Indices

To evaluate how well each product's delivery process meets customer requirements, Process Capability Indices (Cp, Cpu, Cpl, and Cpk) were calculated per product type. These indices measure how capable each process is of consistently meeting delivery times within the acceptable limits of 32 hours, with Cpk values above 1.0 generally indicating that the process can meet customer expectations (VOC).

Table 2.2

ProductType <chr>	Cp <dbl>	Cpu <dbl>	Cpl <dbl>	Cpk <dbl>	VOC_met <lgl>
MOU	0.915	0.727	1.104	0.727	FALSE
KEY	0.917	0.729	1.105	0.729	FALSE
SOF	18.135	35.188	1.083	1.083	TRUE
CLO	0.898	0.717	1.079	0.717	FALSE
LAP	0.899	0.696	1.101	0.696	FALSE
MON	0.889	0.700	1.079	0.700	FALSE

Table 2.2 shows that only software (SOF) meets the customer requirements (VOC_met), with a Cpk of 1.083, and are thus capable of consistently completing the deliveries in less than 32 hours. All other product delivery processes have Cpk values below 1, indicating that they are not capable of consistently meeting the 0–32 hour requirement.

Capability indices assume the process is stable, but since the X-bar charts showed instability, the Cpk results for these unstable processes cannot be trusted. The Cpk values determined are likely unreliable and the processes will need to be stabilised first.

Once the processes are stabilised, the Cpk values will likely increase (improve) and more product processes will be able to meet the VOC.

Process Control Issues

To assess process stability and detect potential control issues, three SPC rules were applied to all product types. Rule A identifies samples with points beyond the ±3σ control limits, indicating possible special-cause variation. Rule B evaluates long runs of points within ±1σ limits, reflecting good process consistency, while Rule C detects four or more consecutive X-bar points outside the ±2σ limits, signalling potential mean shifts.

Rule A and Rule C

Table 2.3

Product <chr>	RuleType <chr>	ChartType <chr>	Total <int>
CLO	Rule A	X-bar	226
CLO	Rule B	S	1
CLO	Rule C	X-bar	460
KEY	Rule A	X-bar	264
KEY	Rule B	S	1
KEY	Rule C	X-bar	452
LAP	Rule A	X-bar	115
LAP	Rule B	S	1
LAP	Rule C	X-bar	254
MON	Rule A	X-bar	190
MON	Rule B	S	1
MON	Rule C	X-bar	358
MOU	Rule A	S	1
MOU	Rule A	X-bar	295
MOU	Rule B	S	1
MOU	Rule C	X-bar	530
SOF	Rule A	X-bar	320
SOF	Rule B	S	1
SOF	Rule C	X-bar	556

Table 2.3 demonstrates that the S-charts show almost no rule violations for any of the products (only one violation each for a few products). This indicates that process variability is stable for all products. There are no signs of special-cause variation in process spread, confirming consistent variation over time.

The X-bar charts, however, show multiple violations for most products. Software (SOF) has the highest number of violations (Rule A: 320, Rule C: 556), followed by Mouse (MOU) (Rule A: 295, Rule C: 530) and Keyboard (KEY) (Rule A: 264 Rule C: 452). These violations suggest instability in process means, where subgroup averages fluctuate significantly, often trending or clustering. Such behaviour may be caused by special-cause variation (e.g., process shifts or workload spikes).

Products such as Laptop (LAP) and Monitor (MON) show fewer total violations compared to the others. These processes appear relatively more stable, although deviations still occur. This indicates some predictable, routine variation likely due to normal operating conditions.

Across all six products, the S-charts indicate stable process variation, while the X-bar charts reveal that process means are not yet fully stable.
The high counts of Rule A and Rule C violations on X-bar charts indicate that the processes frequently drift above or below target, possibly due to non-random influences.
In contrast, the low number of S-chart violations confirms that variability itself is stable and the main problem lies with process means and consistency.

Rule B

Rule B measures the most consecutive samples of subgroup standard deviations between the -1 and +1 sigma-control limits for the products. This signifies good process control.

The maximum consecutive s in $\pm 1\sigma$ for product MOU is 16. For product Key: 15, product SOF: 21, product CLO: 35, product LAP: 19, and the maximum consecutive points for product MON is 34.

The Cloud (CLO) and Monitor (MON) products showed the longest stable runs (35 and 34 points respectively), reflecting excellent process consistency. The Software (SOF) and Laptop (LAP) processes also displayed strong control with 21 and 19 points respectively, while MOU and KEY maintained moderate stability (16 and 15 points).

Combined Analysis

While variation (Rule A and B) is generally well-controlled across all products, several processes, especially SOF, MOU, CLO, and KEY, show repeated mean shifts (Rule C), indicating that their averages fluctuate more than expected even though variability remains stable. This suggests that while process consistency is maintained, external factors may be influencing delivery times.

Part 4

In quality control, Type I and Type II errors are used to evaluate the effectiveness of control charts in detecting process changes. A Type I error (manufacturer's risk) occurs when a process that is actually in control is incorrectly flagged as out of control, while a Type II error (consumer's risk) occurs when a real process shift goes undetected. Analysing both errors provides management with insight into the sensitivity and reliability of the SPC system, helping balance the trade-off between false alarms and missed detections.

Type I Error

A Type I error (manufacturer's error) is when the rule signals that the process is out-of-control even though the process is actually in control.

For a single subgroup in the SPC charts, the Rule A false-alarm probability is 0.001349898, meaning that there is roughly a 0.135% chance of a point exceeding the $+3\sigma$ control limit even when the process is perfectly in control. In practical terms, this equates to about one false alarm in every 740 subgroups purely due to normal random variation. Therefore, if only a single point breaches the $+3\sigma$ limit (as seen for the MOU product), it is likely to be the result of chance rather than an indication of a real process shift. However, repeated Rule A violations would suggest genuine special-cause variation affecting process consistency.

For Rule C, which detects four consecutive X-bar points beyond the +2σ control limit, the probability of this occurring by chance is extremely small ($\approx 2.68 \times 10^{-7}$). This means that, on average, such a pattern would occur only once in about 3.7 million opportunities if the process were truly in control. Therefore, when multiple Rule C sequences are observed (as in the SOF, MOU, and CLO processes), the likelihood that these are false alarms is effectively zero. Instead, they represent strong evidence of systematic mean shifts or non-random influences in the process that require investigation.

For Rule B (r consecutive S samples inside ±1σ control limits) the probability is p^r where $p \approx 0.6827$. That is the probability that a specific run of r consecutive samples will all be within ±1σ control limits.

Table 4.1

	Product <chr>	Max_Consecutive_Run <dbl>	Type1_error_B <dbl>
MOU	MOU	16	2.221573e-03
KEY	KEY	15	3.254575e-03
SOF	SOF	21	3.292244e-04
CLO	CLO	35	1.569712e-06
LAP	LAP	19	7.065768e-04
MON	MON	34	2.299607e-06

Table 4.1 shows, for each product, the maximum consecutive run length (r) that was observed and the corresponding Type I probability p^r. This is the probability that the maximum number of consecutive runs of S samples inside ±1σ was by chance and does not actually signify a stable process. The Type I errors for Rule B of all the products are small and thus the probability that the consecutive runs were purely by chance are very small. The results of Rule B are thus trustworthy and indicate stable variations in the processes of the products.

Type II Error

The calculated Type II error (β) for the X-bar chart is 0.841178, which means there is an 84.12% chance of failing to detect the process shift even though it has occurred. In other words, when the process mean shifts from 25.05 L to 25.028 L, and the standard deviation increases from 0.013 L to 0.017 L, the X-bar chart will still appear to be “in control” approximately 84% of the time. This happens because the sample means remain within the existing control limits by random chance, even though the process is no longer centred at its target value.

The statistical power (1 – β) of the chart is therefore only 15.88%, indicating that there is just a 15.88% probability of correctly detecting the shift when it actually occurs. A power value this low means that the chart is relatively insensitive to small process shifts, and such a change in the mean may go unnoticed for a while before an out-of-control signal is triggered.

From a quality assurance perspective, this represents a high consumer’s risk, since the process is producing off-centred or potentially nonconforming product while the control chart fails to issue an alarm. In real-world terms, the customer continues to receive product from a process that is no longer operating at its intended average fill volume.

To reduce the likelihood of missing such shifts, the manufacturer could increase the subgroup size (which reduces the standard error of the mean) or tighten the control limits (for example, using $\pm 2\sigma$ instead of $\pm 3\sigma$).

Conclusion

Together, the Type I and Type II results show that the current control system is imbalanced in sensitivity. The relatively low Type I error rates indicate that the charts rarely trigger false alarms, but the very high Type II error ($\beta \approx 0.84$) means the system is too insensitive to real process shifts. For management, this implies that while operations appear stable, true deviations in process performance may go undetected, risking product quality issues and customer dissatisfaction unless chart parameters or sampling practices are improved.

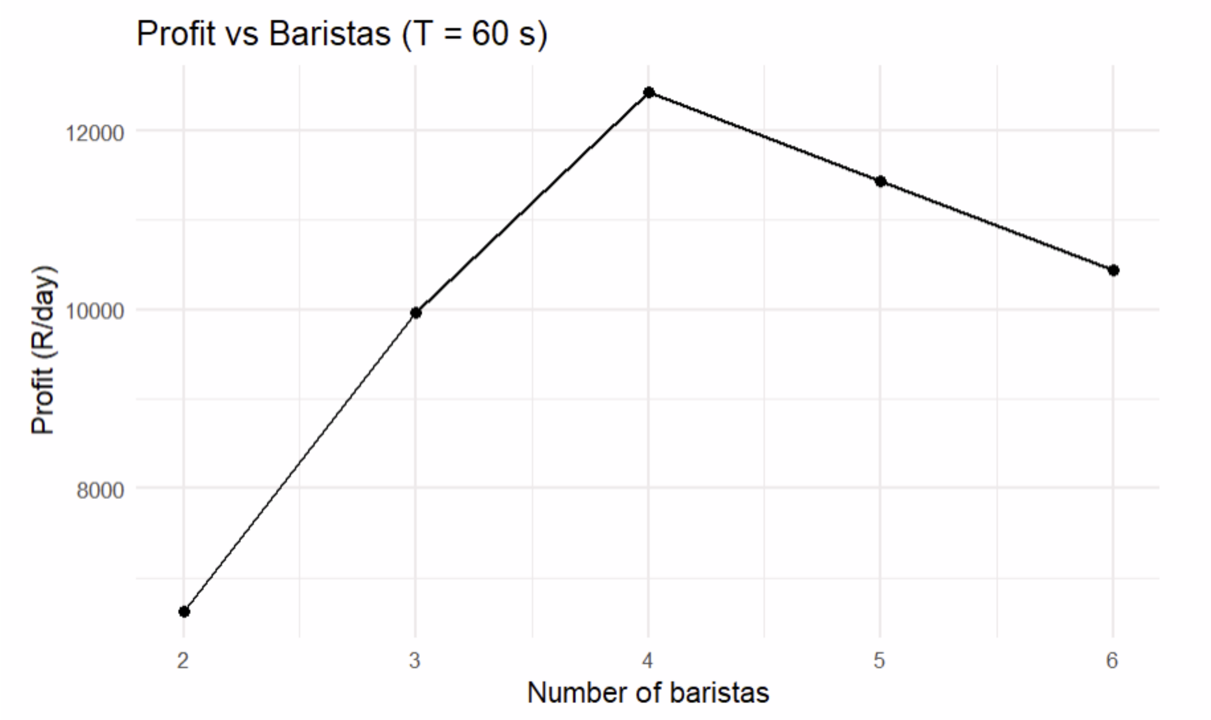
Part 5

This section focuses on optimising the operational performance and profitability of two coffee shops, Shop 1 and Shop 2, using the datasets timeToServe.csv and timeToServe2.csv. Each dataset represents one year of service records, detailing the number of baristas on duty and corresponding customer service times. The objective of this analysis is to determine the optimal staffing level that maximises daily profit while maintaining a high level of service reliability. Profitability is evaluated by balancing revenue per customer against daily staff costs, while reliability is defined as the proportion of orders completed within a target service time.

Assumptions

This optimisation analysis assumes that each dataset (timeToServe.csv and timeToServe2.csv) represents one year of service records for a coffee shop, with around 200,000 customer transactions. Each record contains the number of baristas on duty and the corresponding service time in seconds. Both shops operate eight hours per day (28,800 seconds) for 365 days per year, serving an estimated average of 548 customers per day. Service reliability is defined as the proportion of orders completed within a target time of 60 seconds, with alternative thresholds of 45 and 90 seconds considered. If all service times exceed these limits, the reliability threshold is dynamically adjusted using the 25th, 50th, and 75th percentiles of the dataset to allow meaningful comparison. The daily service capacity for each staffing level is estimated by dividing the total available seconds by the mean service time, while the actual number of customers served is limited by daily demand. Each sale contributes a profit of R30, and each barista costs R1,000 per day in wages and overhead. The model evaluates barista counts from two to six, calculating daily profit and reliability for each case. The optimal number of baristas is identified as the one that maximises daily profit, and two plots, Profit vs. Number of Baristas and Reliability vs. Number of Baristas, are generated to visualise performance trends.

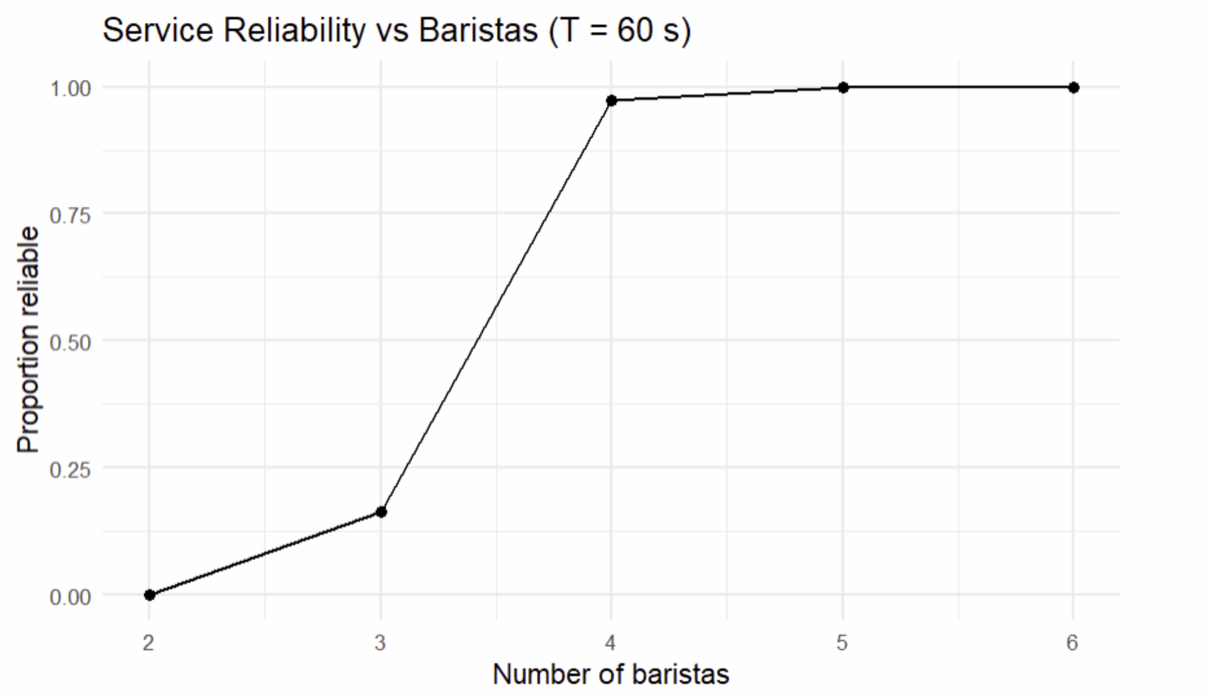
Shop 1



Graph 5.1

The profit vs baristas of Shop 1 (Graph 5.1) shows that the profit increases sharply up to four baristas, peaking around R12 500/day. After that, the profit starts to decline as staff costs outweigh the benefit of slightly faster service.

This pattern shows a classic cost–benefit trade-off where the operational efficiency improves with staffing, but profit has an upper limit once reliability is saturated.



Graph 5.2

The service reliability vs baristas (Graph 5.2) graph is analysed with a threshold of 60 seconds. This graph shows that with two baristas; reliability is effectively 0% because the service times almost always exceed 60 s. When there are three baristas, only about 10–15% of orders meet the reliability target. At four baristas, reliability jumps to about 99%, showing a nonlinear gain. Beyond four baristas (5–6), reliability plateaus near 100%, meaning that adding more baristas offers no further reliability benefit.

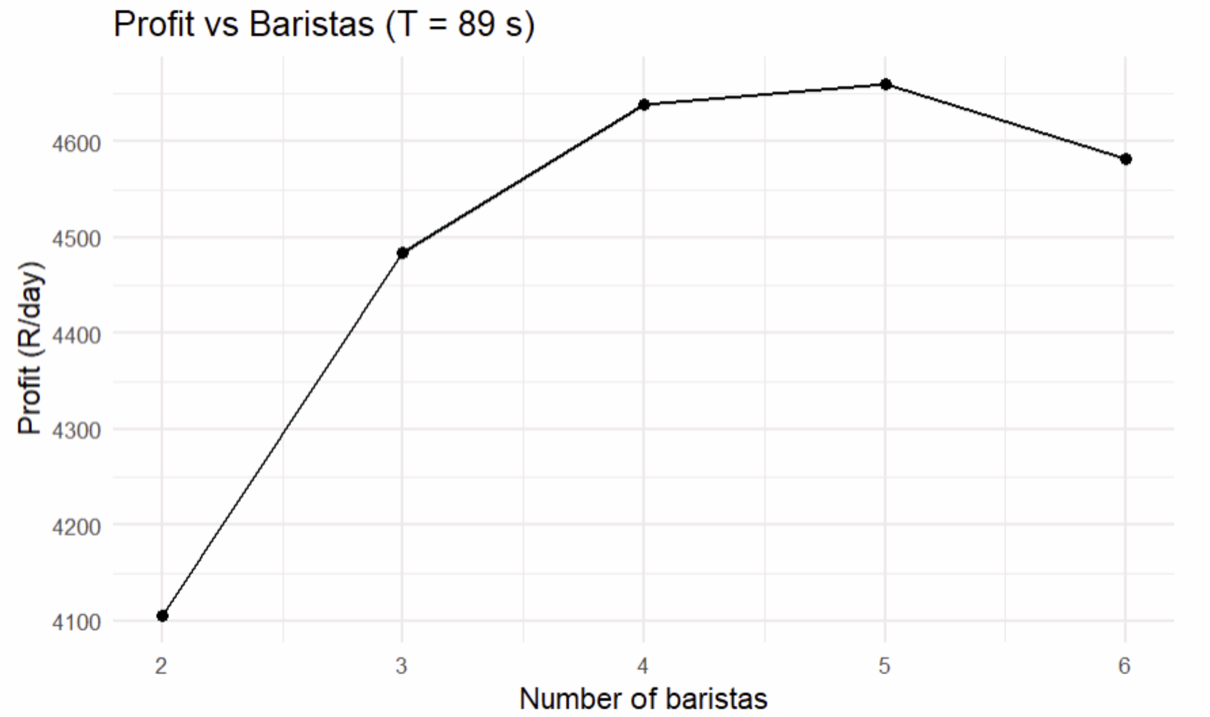
Four baristas are thus the minimum number needed to ensure almost all customers are served within 60 seconds.

Optimal Staffing Decision

The optimal number of baristas for Shop 1 is four, as this achieves near-perfect service reliability (≈99%) with a maximum daily profit (~R12 500/day).

Adding a 5th or 6th barista increases cost without improving reliability or speed significantly.

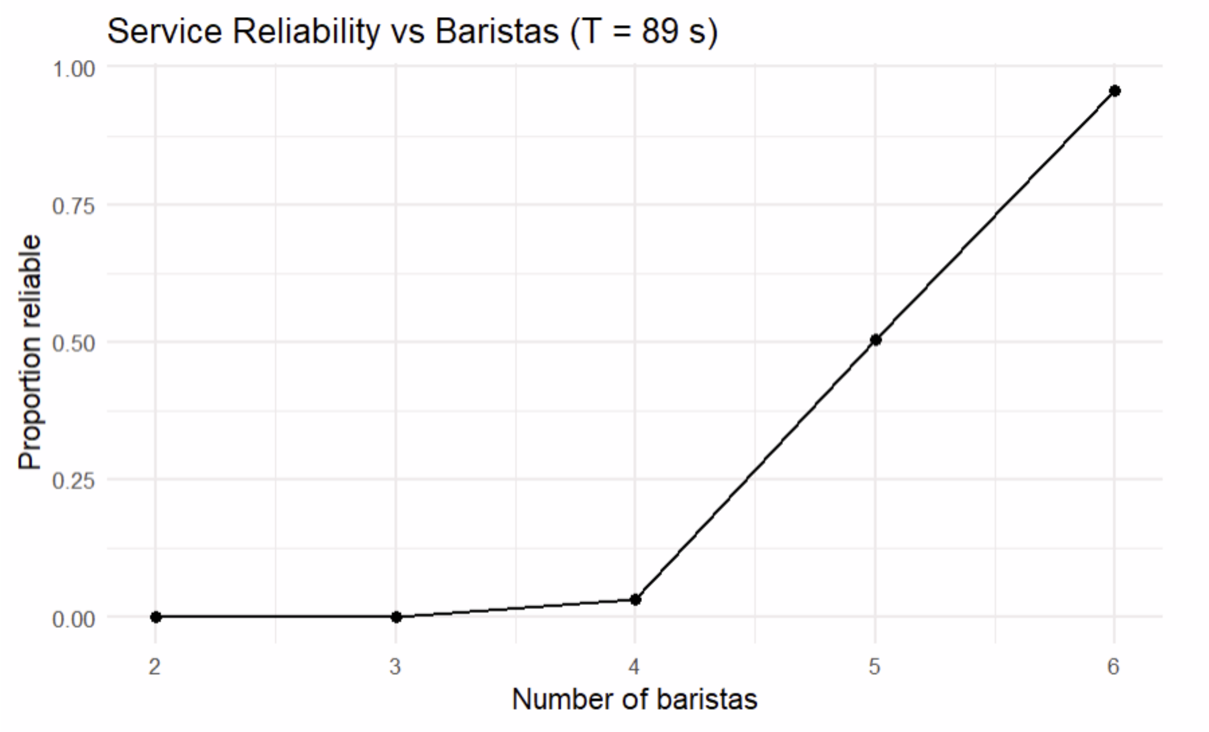
Shop 2



Graph 5.3

Graph 5.3 shows that in Shop 2 the profit rises steadily from two to five baristas, peaking around R4 650/day. At six baristas, profit starts declining slightly, since the additional wage cost outweighs any minor service improvement.

The peak profit thus occurs at five baristas, which provides a balance between cost and customer service.



Graph 5.4

The service reliability vs baristas (Graph 5.4) graph is analysed with a threshold of 89 seconds.

When there are only two to three baristas at Shop 2, the reliability is essentially 0% and very few customers are served under 89 seconds. From four to five baristas the reliability increases slowly. Only at six baristas does reliability approach 95–100%, meaning consistent on-time service is only possible with heavy staffing.

Optimal Staffing Decision

The optimal number of baristas for Shop 2 is five baristas. This amount of baristas produce the highest daily profit (~R4 650/day) with ~50–60% service reliability, a fair compromise given high labour cost.

Comparison:

Shop 2's operations are less efficient than Shop 1 as it needs more staff to achieve the same reliability level as Shop 1. This may be due to more complex menu or preparation steps, slower equipment or layout inefficiency, or lower staff productivity or training differences.

Shop 1 is also significantly more profitable, with an optimal profit of ≈12 500 R/day, than Shop 2 with an optimal profit of ≈4 650 R/day.

Conclusion

The optimisation results reveal that Shop 1 reaches maximum profit and service reliability with four baristas, whereas Shop 2 requires five baristas to achieve peak profitability and moderate reliability. Shop 2's slower performance suggests process inefficiencies or longer preparation times. Both shops demonstrate diminishing returns beyond their respective optimal staffing levels, where increased wages reduce profit despite marginal gains in speed. Quality assurance efforts should therefore focus on improving workflow efficiency in Shop 2 to reduce the service time gap and allow similar reliability to be achieved with fewer staff.

Part 6

In this section, an Analysis of Variance (ANOVA) was conducted to determine whether the mean delivery times for the SOF product differed significantly across months or years. This analysis helps identify whether observed variations in delivery performance are due to natural process variation or systematic differences linked to time-based factors such as workload or seasonal demand. Both yearly and monthly effects were examined, followed by a post-hoc Fisher's LSD test to identify which months differ significantly from one another.

ANOVA

The SOF product was selected for ANOVA analysis because it demonstrated the most stable process performance in earlier SPC and capability studies. Its S- and X-bar charts showed consistent control, and its Cpk value (1.083) indicated that it met the Voice of the Customer (VOC) requirements. This made SOF an ideal candidate for further statistical comparison, as any observed differences in delivery time could be attributed primarily to time-based factors (month or year) rather than underlying process instability.

The other products exhibited high process instability and did not meet VOC requirements, indicating that their processes were not yet in statistical control. Therefore, ANOVA was not conducted for these products, as any differences detected would likely reflect process variation rather than meaningful performance trends.

Table 6.1

One-way ANOVA for SOF Delivery Times by Year						
	DF	Sum Sq	Mean Sq	F value	Pr(>F)	Significance
Year	1	0	0.01695	0.179	0.672	Not significant
Residuals	20747	1966	0.09475	—	—	

Table 6.2

One-way ANOVA for SOF Delivery Times by Month						
	DF	Sum Sq	Mean Sq	F value	Pr(>F)	Significance
Year	11	138.2	12.563	142.5	<2e-16	Highly significant
Residuals	20737	1827.5	0.088	—	—	

The one-way ANOVA for Year (Table 6.1) shows no statistically significant difference in mean delivery times between 2022 and 2023 (F = 0.179, *p* = 0.672). This means that delivery performance for the SOF product remained stable over the two years, with no measurable year-to-year improvement or decline in efficiency.

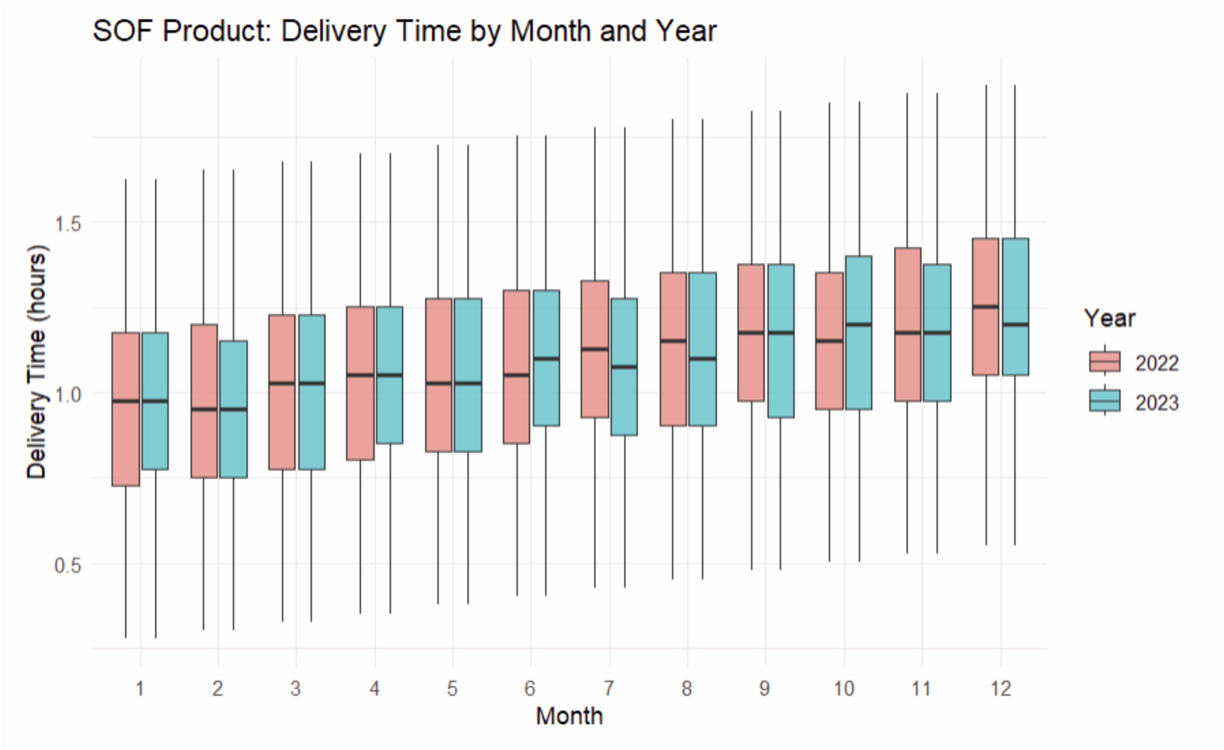
In contrast, the ANOVA for Month (Table 6.2) indicates a highly significant effect of month on delivery time (F = 142.5, *p* < 2e-16). This suggests that delivery times varied systematically throughout the year.

Table 6.3

Fisher’s LSD post-hoc test		
	deliveryHours <dbl>	**groups** <chr>
12	1.2393251	a
11	1.1928480	b
10	1.1781171	bc
9	1.1611674	c
8	1.1220673	d
7	1.0945059	e
6	1.0841766	e
5	1.0529084	f
4	1.0368634	f
3	1.0023405	g
2	0.9751329	h
1	0.9553307	h

The Fisher’s LSD post-hoc test further shows clear monthly groupings. The highest mean delivery times occurred in December (1.24 hours) and November (1.19 hours), while the lowest occurred in January (0.96 hours) and February (0.98 hours). These results imply that delivery performance tends to slow down toward the end of the year, possibly due to higher order volumes, capacity constraints, or seasonal workload fluctuations.

The boxplot (Graph 6.1) visually confirms these trends. The monthly medians increase gradually from early to late months, and the spread (interquartile range) also widens, indicating slightly more variability in later months.



Graph 6.1

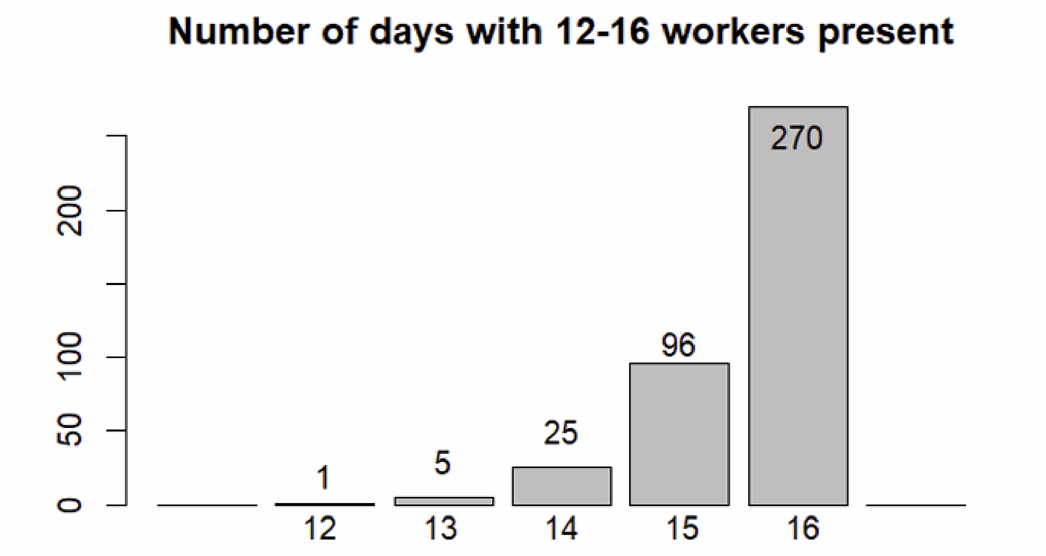
Conclusion

The ANOVA results suggest that year-to-year performance remained stable, but monthly variation in delivery times is statistically significant. This indicates that process efficiency fluctuates across the year, likely due to seasonal workload changes or resource constraints. Management should focus on balancing capacity and staffing levels during high-demand months (such as November and December) to maintain consistent delivery performance. Implementing seasonal scheduling adjustments or process optimisation strategies could help reduce variation and ensure more predictable delivery times throughout the year.

Part 7

This section investigates the relationship between staffing levels and service reliability at a car rental agency, using operational data collected over 397 days. The goal is to estimate the probability of providing reliable service based on the number of employees on duty and to determine the optimal workforce size that maximises annual profit. A binomial probability model is used to represent the likelihood of worker attendance and to quantify how variations in staffing impact reliability, daily revenue losses, and overall profitability.

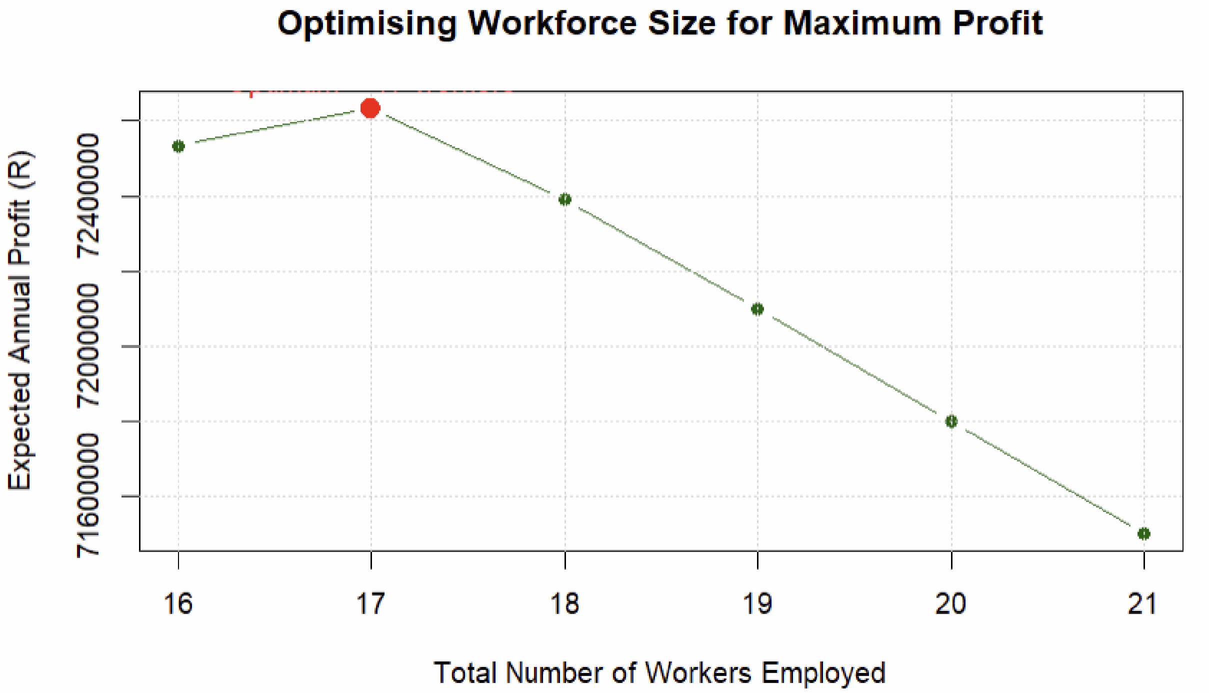
Reliability of service



Graph 7.1

By analysing Graph 7.1, the reliability of daily service at the car rental agency was estimated using the distribution of staff availability over 397 recorded days. Modelling the number of workers present as a binomial process with a weighted probability of attendance $p = 0.974$ and a threshold of 15 workers (the minimum required for reliable service), the probability of achieving reliable service on any given day was calculated as $P(X \geq 15) = 0.936$. This indicates that the agency can expect reliable service on approximately 93.6% of days, equivalent to about 342 reliable days per year. Conversely, around 23 days per year are expected to experience staffing shortages, which negatively impact service quality and sales performance.

Optimising Profit



Graph 7.2

The profit optimisation model (Graph 7.2), based on the binomial reliability framework, demonstrates that annual profit is maximised when 17 workers are employed. Increasing staff from 16 to 17 improves reliability sufficiently to prevent costly service disruptions, raising annual profit from approximately R72.53 million to R72.63 million. Beyond this point, the added salary expense outweighs any reliability gains, leading to a gradual decline in profit. Thus, the optimal balance between cost and reliability occurs at 17 employees, ensuring stable service quality and efficient use of personnel resources.

Conclusion

The analysis shows that the agency currently provides reliable service on approximately 93.6% of days, equivalent to around 342 days per year. Profit optimisation results indicate that hiring one additional worker, bringing the total staff to 17, maximises expected annual profit by balancing labour costs and reliability benefits. Beyond this point, further increases in staffing yield diminishing returns. Therefore, maintaining a workforce of 17 employees is recommended to ensure consistently reliable service while optimising profitability.