Quality Assurance 344 ESCA 2025

SU number: 26875829

Name: NJ van Blerk

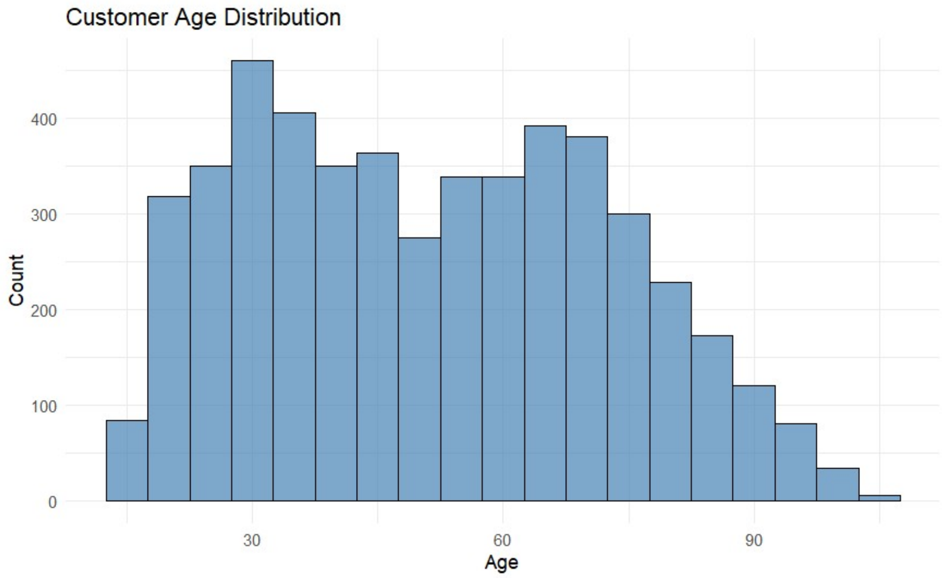
Table of Contents:

1. Introduction

2. Part 1.2: Descriptive Statistics

The R code included a check for missing values, which confirmed that the dataset was complete.

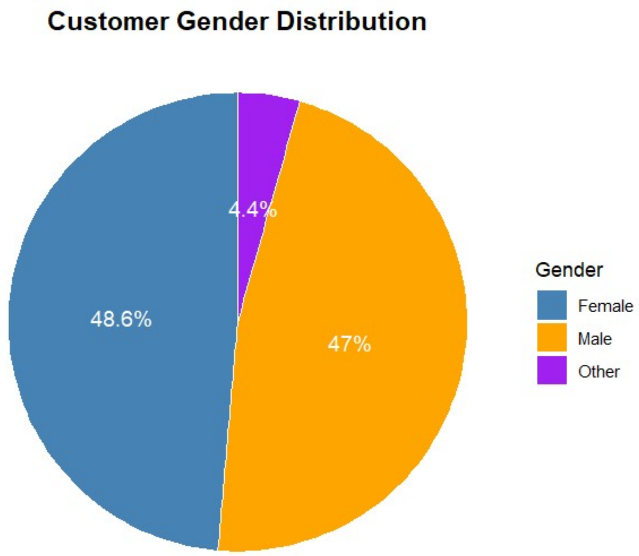
Figure 1: Customer age distribution



Min = 16, Max = 105, Average = 51.5538

This graph shows that customer ages range broadly from 16 to 105 years, with the majority between 25 and 70.
The distribution is slightly right skewed, indicating fewer elderly customers.
The mean age of 51.55 years suggests that the customer base mainly consists of middle-aged adults, implying a mature demographic with stable purchasing power.

Figure 2: Customer gender distribution



There is a nearly equal gender split reflects an inclusive and balanced customer profile, no dominant demographic segment.

The “other” percentage can be companies that bought products, or it can be people that don’t want to divulge their gender.

Figure 3: Number of customers per city

Average income = 80797

The graph reveals seven cities contribute comparably to the customer base, though San Francisco has the highest representation and Miami the lowest. This even geographic distribution suggests consistent market reach across regions. The average income suggests a financially capable customer base.

Figure 4: Products Selling Prices per Category

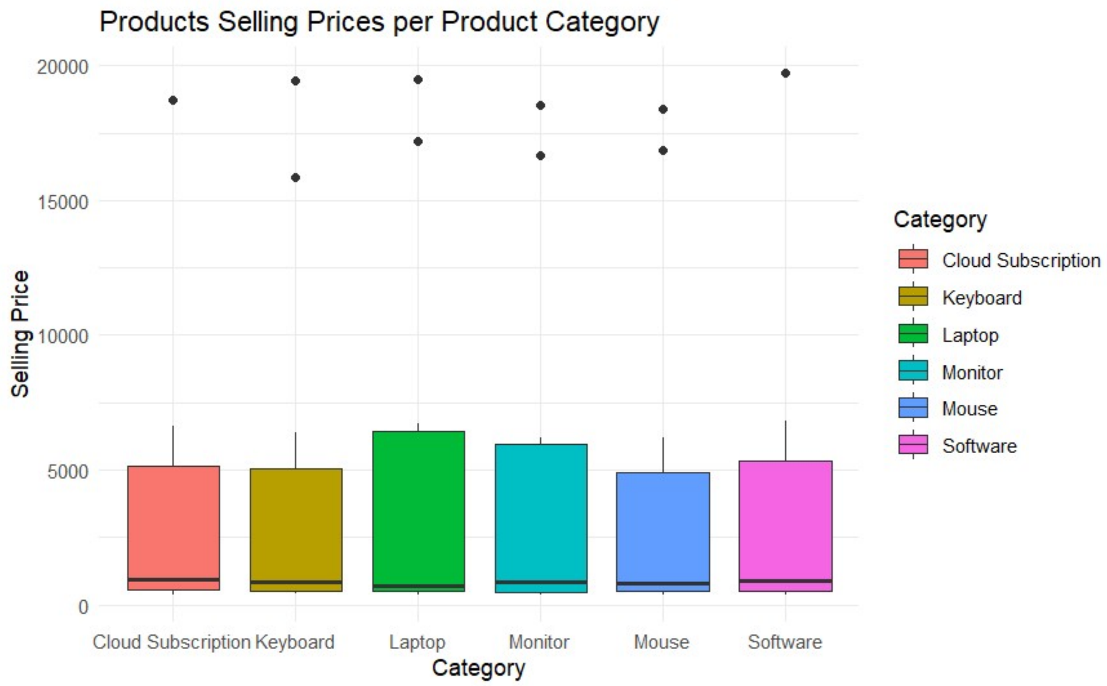


Figure 5: Table for Products data

Total products	Unique Categories	Average Selling price	Average Markup
60	6	4493.59	20.4617

The graph displays median selling prices ranging from about $4000 to $6000.
Laptops exhibit the highest inter-quartile range, reflecting greater price variability, while peripherals such as mice and keyboards show tighter price clusters.
A few high-priced outliers above $20000 appear in most categories, indicating the presence of premium product lines.

Figure 6: Head Office selling prices per product category

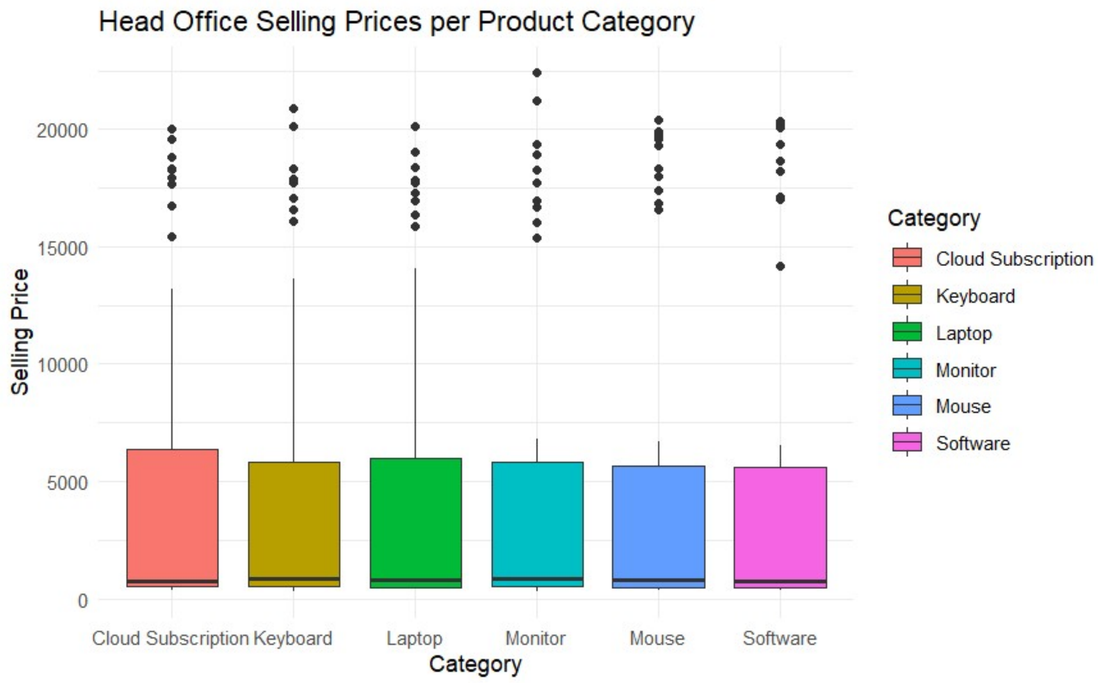
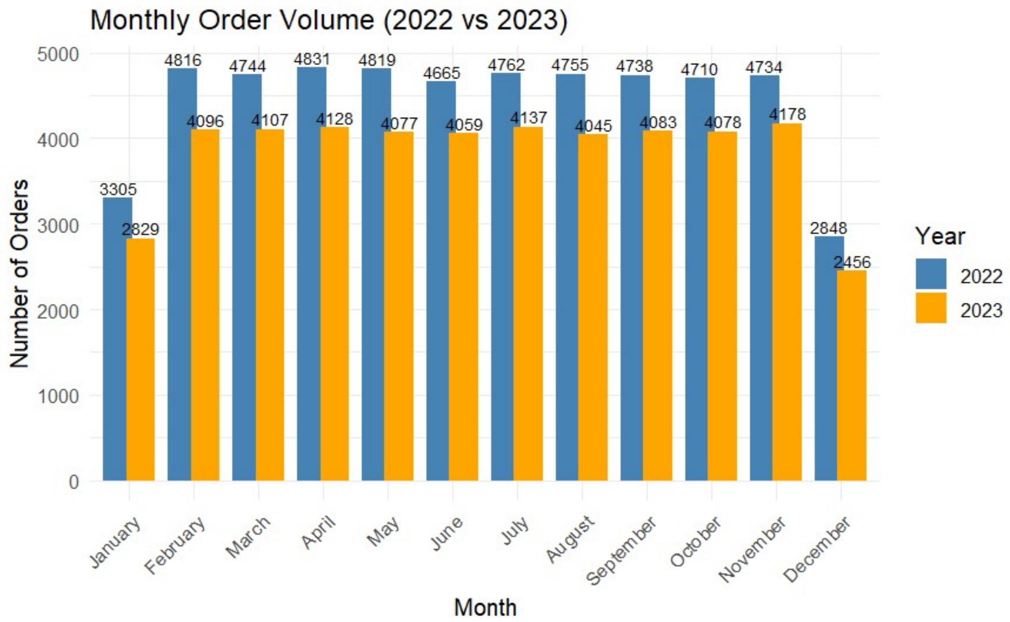


Figure 7: Table for Head office product data

Total products	Unique Categories	Average Selling price	Average Markup
360	6	4410.96	20.3855

The graph displays median selling prices ranging from about $4000 and $6000. Cloud Subscription exhibit the highest inter-quartile range, while mouse and software show lower inter-quartile range. The Head office product data also shows more products being sold at a premium price, if you compare it with the product data.

Figure 8: Monthly Order Volume

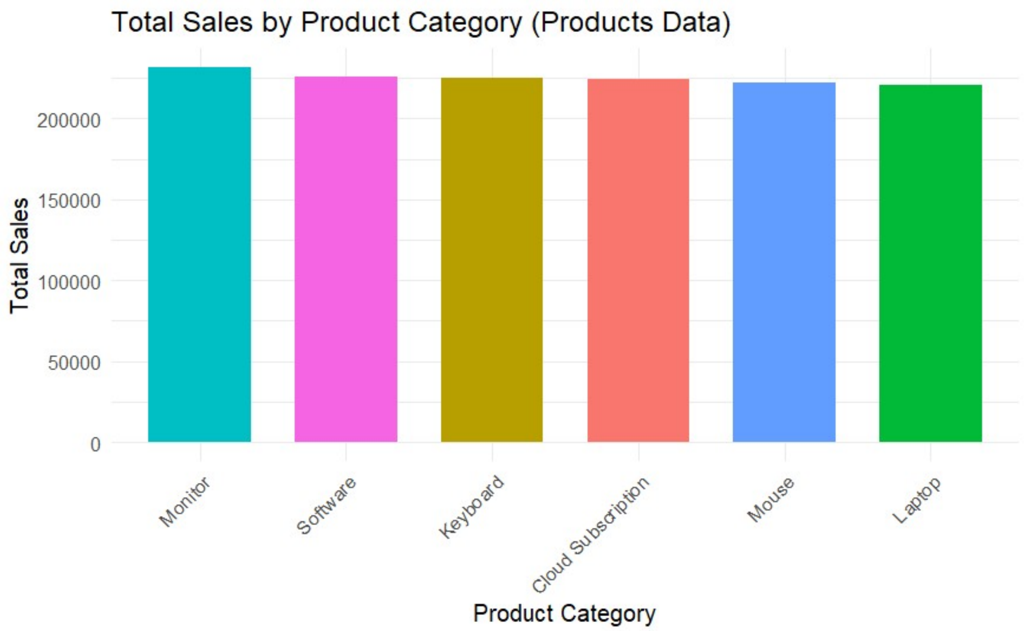


Both years show a steady pattern, with 2023 volumes slightly below 2022, particularly in early and late months. This could be because of several reasons. The highest order activity occurred between February and October 2022, averaging around 4 700 orders per month. Seasonality is visible, but there is no dramatic decline, suggesting consistent demand year over year.

Figure 9: Relationship Between Picking and Delivery Time

The wide rectangular clusters suggest operational batching or process scheduling rather than a simple linear relationship. There is no clear correlation, implying that delivery speed is not directly dependent on picking time but may instead be constrained by logistics or routing.

Figure 10: Total Sales by product category



All categories achieve nearly identical total sales. This indicates a balanced product portfolio with minimal dependence on a single category — a sign of healthy diversification.

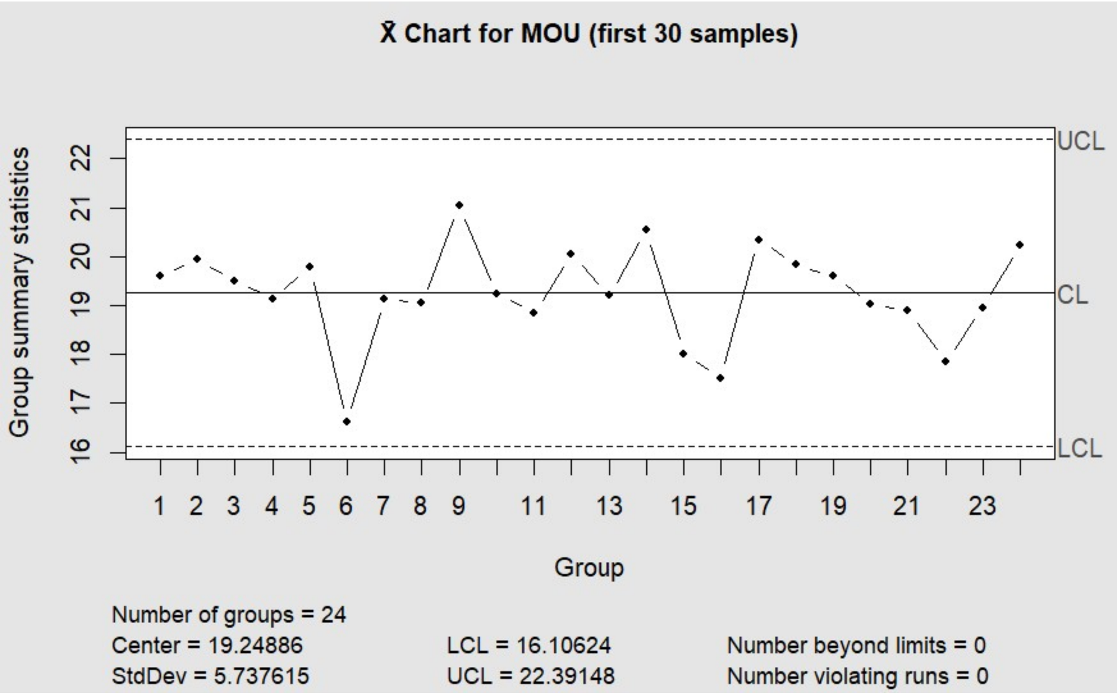
3. Part 3: SPC X-s Charts and Process Capability

Part 3.1

The data was ordered by year, month, day and order time. See code to understand how the ordering was done. For every product type, the first 30 × 24 = 720 observations were used to initialise control limits, forming the historical "in-control" baseline from which process variation could be assessed. There are 6 product families: "MOU" "KEY" "SOF" "CLO" "LAP" "MON".

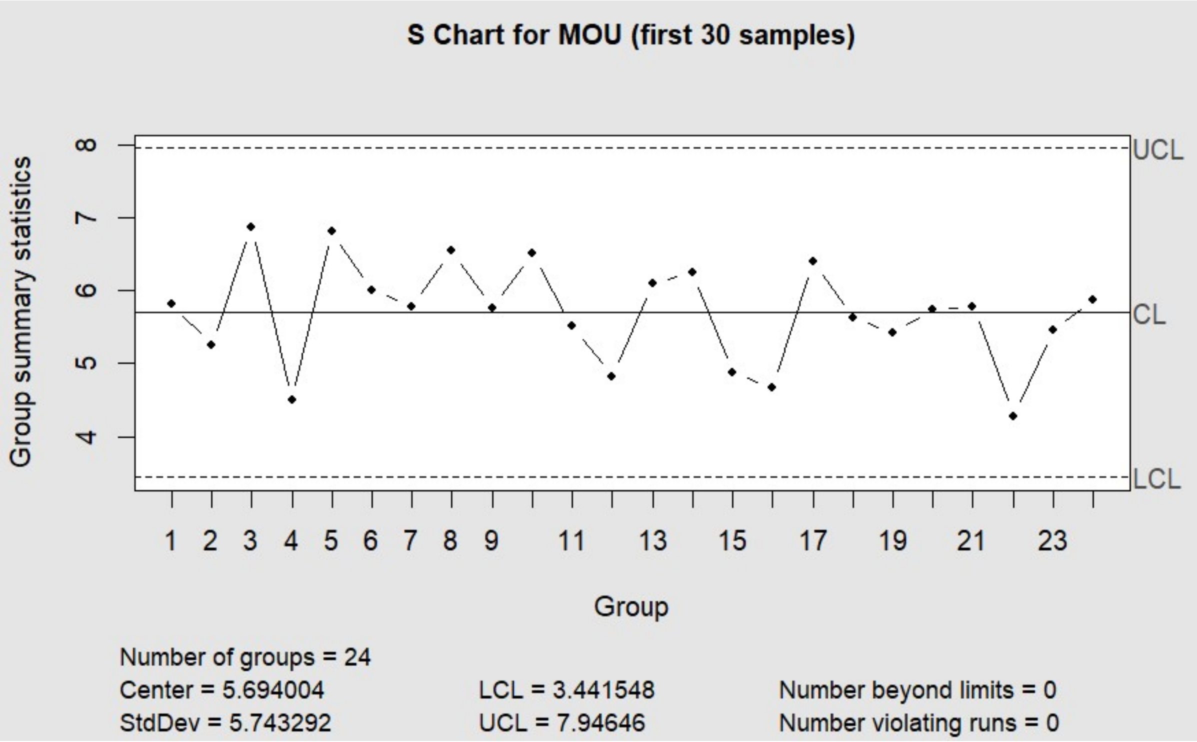
We will focus on one family to explain what the charts mean.

Figure 11: $\bar{X}$ chart for MOU



The centre line (CL = 19.25) represents the average delivery-time mean, while the UCL = 22.39 and LCL = 16.11 define the 3-σ limits.
All points lie within these boundaries, and no abnormal patterns or run violations are observed. The chart therefore confirms that the mean process level for delivery time is in statistical control with stable central tendency and no evidence of assignable variation.

Figure 12: S chart for MOU

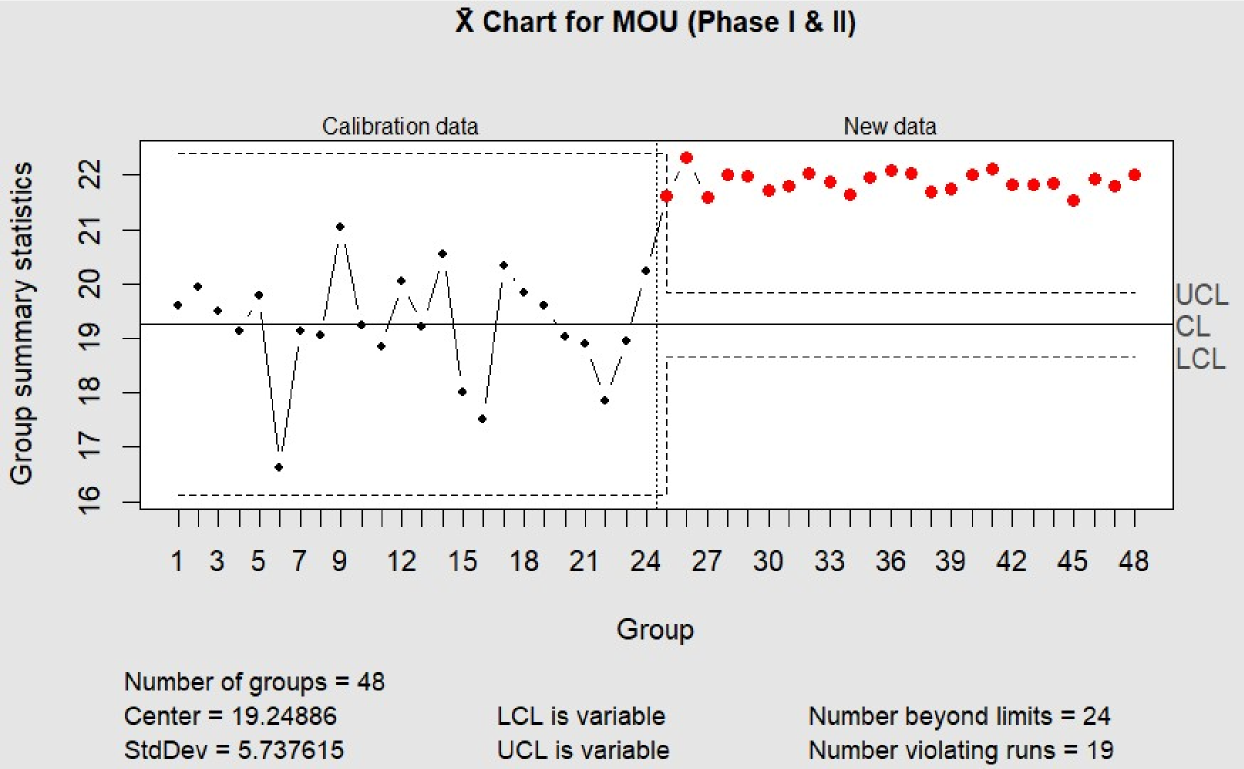


The process variability is centred at CL = 5.69, with UCL = 7.95 and LCL = 3.44.
No subgroup exceeds these limits, and the pattern fluctuates randomly around the centre line, showing consistent process dispersion.
This indicates that the inherent variability of the delivery-time process is stable and not influenced by systematic shifts or out-of-control causes.

Part 3.2

Following the initial calibration using the first 30 samples (Phase I), continuous process monitoring was implemented in Phase II using subsequent samples (31, 32, ...).

Figure 13: X̄ chart for MOU (Phase I & II)



The graph shows both the initial calibration data (Groups 1–24) and the subsequent new monitoring data (Groups 25–48). The calibration portion represents a stable baseline, while the new data — highlighted in red — depict the results of ongoing production monitoring. Notably, 24 points lie beyond the upper control limit, and 19 consecutive points violate standard run rules, indicating non-random patterns and sustained upward shifts in the process mean. This behaviour is characteristic of an “out-of-control” process, where the mean has systematically increased above its historical level.

Phase II monitoring confirms that while the system began in a state of control, subsequent samples revealed a process drift.
The upper-limit violations and run-rule breaches clearly indicate a change in process performance, meaning the delivery-time process can no longer be assumed stable.

Part 3.3

Figure 14: Capability for all product families

ProductFamily	mean_delivery	sd_delivery	Cp	Cpu	Cpl	Cpk	Capable
CLO	19.226	5.9408054	0.8977458	0.7167378	1.078754	0.7167378	Not Capable
KEY	19.276	5.815195	0.9171375	0.7293536	1.104921	0.7293536	Not Capable
LAP	19.606	5.9339589	0.8987816	0.6962187	1.101345	0.6962187	Not Capable
MON	19.41	5.9989192	0.889049	0.6995705	1.078528	0.6995705	Not Capable
MOU	19.2975	5.8276023	0.9151848	0.726571	1.103799	0.726571	Not Capable
SOF	0.955375	0.2940868	18.1352369	35.1876018	1.082872	1.082872	Not Capable

The capability analysis confirms that all product families fall below the threshold of statistical capability.
While processes remain stable (as verified in earlier control charts), they do not meet the desired performance relative to the 0–32 hour specification limits.
To meet the VOC, product managers must therefore implement targeted improvements to reduce delivery-time variation and increase process precision.

Part 3.4

Figure 15: SPC rule summary for all product types

Product Family	RuleA_Count	RuleB_Longest_Run	RuleC_Count
MOU	1	16	369
KEY	0	10	326
SOF	0	13	384
CLO	0	29	283
LAP	0	18	172
MON	0	34	263

Only MOU showed a single 3σ outlier (Rule A), suggesting one temporary increase in variation.
MON and CLO had the longest stable runs (Rule B), indicating good control.
All product types had many Rule C violations, meaning the average delivery times have shifted upwards.

In practice, managers for MOU, KEY and SOF should check their delivery operations for systematic delays, while MON shows the most consistent process performance.

4. Part 4: Type I and II Errors

5. Part 5: Optimisation of Coffee Shop Profit

To optimize our business strategy, we analysed data from two coffee shops correlating the number of baristas on shift with customer service times (in seconds). Using this data, we developed a model to determine the ideal number of baristas that minimizes wait times while maintaining operational efficiency.

The model used only 1-6 baristas for the model, because 6 baristas is the limit.

Profit: R30 × customers served − R1000 × baristas/day

Figure 16: Coffee shop 1

baristas	mean_service_sec	reliability_lt60	reliability_lt120	profit_per_day
1	200.15588	0	0	3316.636
2	100.17098	0	99.71879	6625.253
3	66.61174	16.4605	100	9970.686
4	49.98038	97.22914	100	13286.784
5	39.96183	99.99647	100	16620.629
6	33.35565	100	100	19902.661

As we can see from the table that the best choice for this coffee shop will be 6 baristas. The reliability is 100% for under 1 and 2 minutes. It also has the best profit compared to the other barista numbers. The Taguchi loss explains why the lower number baristas are not chosen, the reliability is worse, and this angers customers. Then they don't pay for their coffee.

Figure 17: Coffee shop 2

baristas	mean_service_sec	reliability_lt60	reliability_lt120	profit_per_day
1	200.16894	0	0	3316.354
2	141.51462	0	0.1241675	4105.376
3	115.44091	0	79.3454067	4484.348
4	100.01527	0	99.9971663	4638.681
5	89.43597	0	100	4660.543
6	81.64272	0	100	4582.695

As we can see from the table that the best number of baristas is 5. The profit lowers when you have 6 baristas. This coffee shop is also a lot slower in their service as the first coffee shop. It could mean they make more complex coffee products or other reasons.

6. Part 6: ANOVA

This one of the product families where an analysis is done on.

Hypothesis

Research question:
Is there a significant difference in CLO delivery hours across years and months?

Null hypothesis (H_0): Mean delivery hours do not differ significantly between years or months.
Alternative hypothesis (H_1): Mean delivery hours differ significantly between years or months.

Figure 18: Delivery hours by Year for CLO

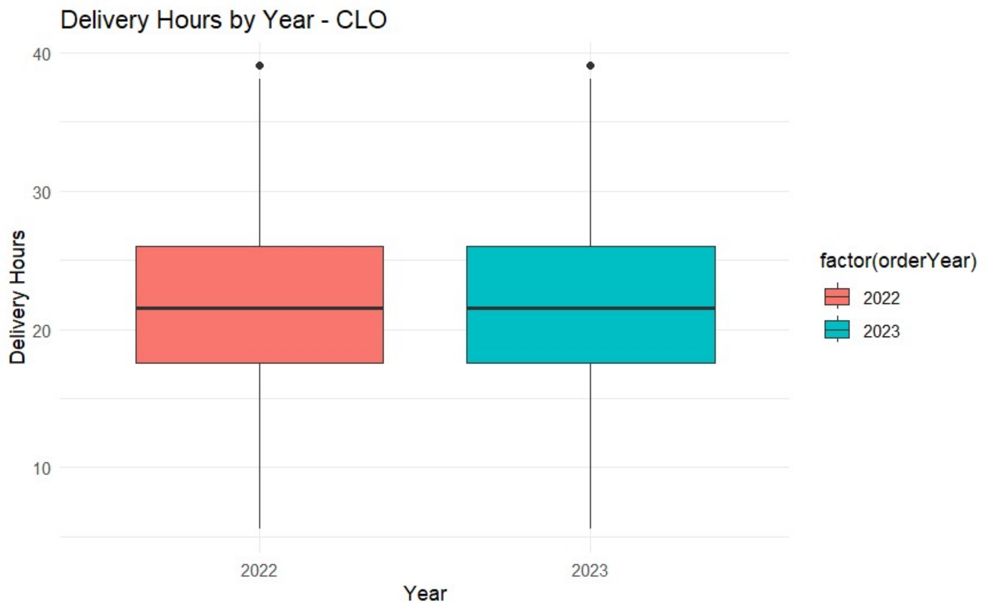


Figure 19: Monthly delivery variation for CLO

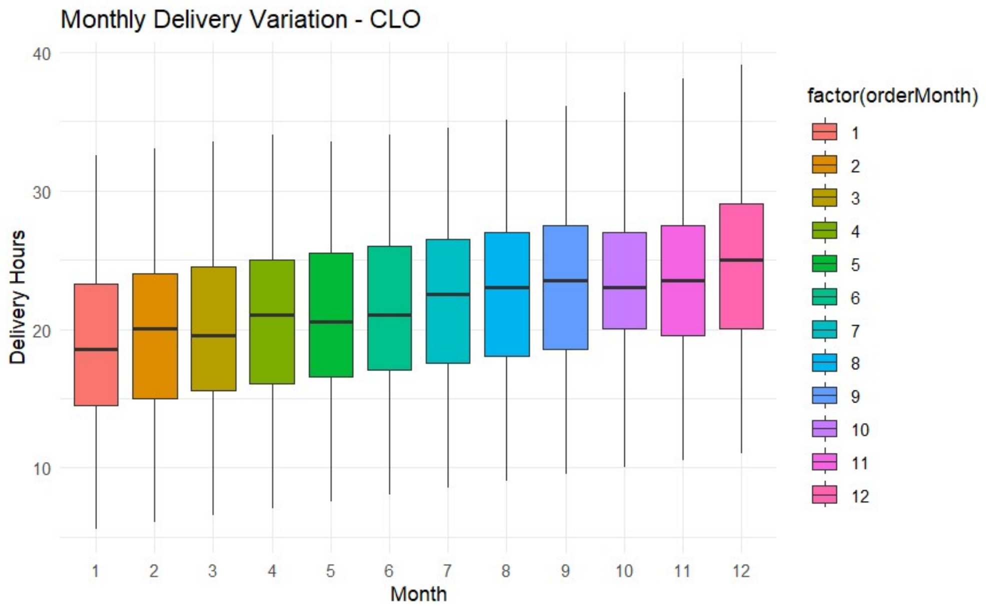


Figure 20: Delivery hours by year statistics for CLO

Source	Df	Sum Sq	Mean Sq	F value	Pr(>F)
factor(orderYear)	1	1	1.17	0.031	0.86
Residuals	15 596	583 187	37.39		

Figure 21: Delivery hours by month statistics for CLO

Source	Df	Sum Sq	Mean Sq	F value	Pr(>F)
factor(orderMonth)	11	41 559	3 778	108.7	<2e-16 ***
Residuals	15 586	541 629	35		

The p-value for year (p = 0.86) shows no significant difference in average delivery times between 2022 and 2023.
However, the p-value for month (< 0.001) indicates a highly significant difference between months.

Figure 20 (Delivery Hours by Year) confirms that yearly medians and spread are nearly identical, with a few mild outliers around 40 hours.
Figure 21 (Monthly Delivery Variation) shows a steady upward trend from early to late months, meaning deliveries generally take longer toward year-end — likely due to seasonal workload or demand spikes.

There is no significant change across years, but clear monthly variation in CLO delivery times. This suggests the delivery process is stable overall yet affected by seasonal demand cycles rather than long-term performance drift.
Managers should focus on improving delivery efficiency during peak months (November–December) to maintain consistent performance throughout the year.

7. Part 7: Reliability of Service

Figure 22: Table of staff reliability and relative profit

staff	reliability	problem_days	lost_revenue	labour_cost	total_cost	relative_profit
14	0	365	7300000	4200000	11500000	-6340359
15	0.6840207	115.3324	2306648.995	4500000	6806649	-1647008
16	0.9405284	21.70712	434142.369	4800000	5234142	-74501.33
17	0.99183	2.982052	59641.043	5100000	5159641	0
18	0.9990977	0.329334	6586.689	5400000	5406587	-246945.7

Figure 23: Total annual cost vs staff level

The best number of staff to have for the company is 17, based on the relative profit. It has the least amount of cost. It has 3 problem days; 362 days will have reliable service.