QUALITY ASSURANCE ECSA REPORT

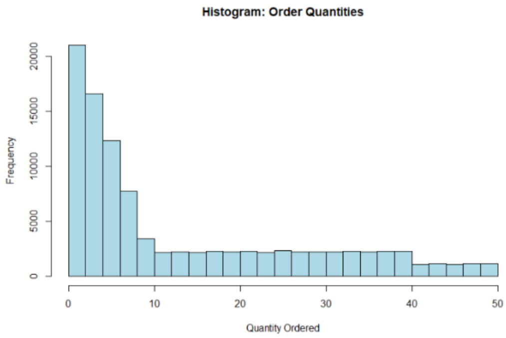
Jared Viakovsky 25096508

Contents

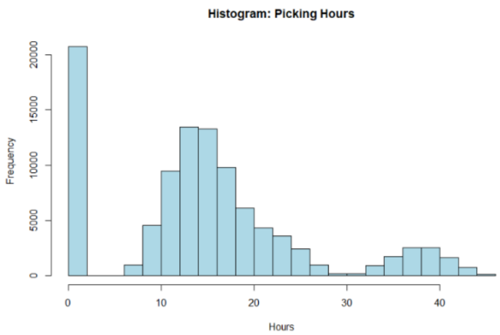
Introduction

This project forms part of the ECSA 2025 Quality Assurance and Data Analysis module and focuses on applying different statistical and analytical methods to real business data. The main goal was to explore how data can be used to monitor processes, identify trends, and improve performance. The project was split into several parts, starting with a basic data analysis to get an overview of the company's sales and product information. From there, more advanced tools were introduced, such as Statistical Process Control (SPC) charts, Process Capability (Cp and Cpk) calculations, and ANOVA testing, to evaluate product performance and process variation. Later sections looked at Type I and Type II errors, corrected mistakes in the head-office product data, and used optimisation models to improve staffing and profitability. Each section built on the previous one, showing how data analytics and quality assurance techniques can be applied together to support better decision-making and process improvement.

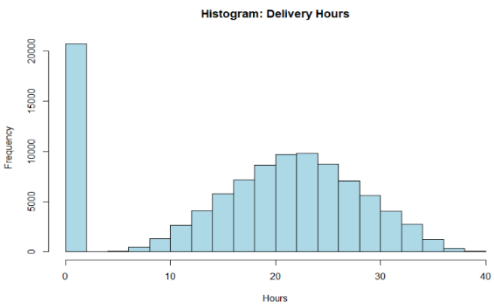
Part 1- Basic data Analysis



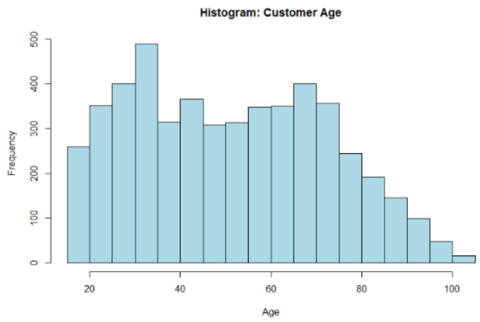
The histogram of order quantities shows a strong right-skewed distribution. Most orders are small (below 10 units), while a few large orders create a long tail. This suggests that most customers purchase in low volumes which is likely for regular use while a small number of bulk orders could represent large institutional clients or seasonal restocking.



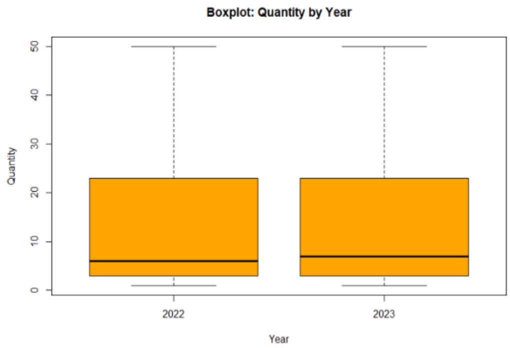
The distribution of picking hours is bimodal, with two clear peaks around 10–20 hours and 35–40 hours. This indicates two distinct operating patterns, possibly corresponding too day and night shifts or too differences in order complexity. The variation implies fluctuating workloads across the warehouse, highlighting where scheduling or process balancing could improve efficiency.



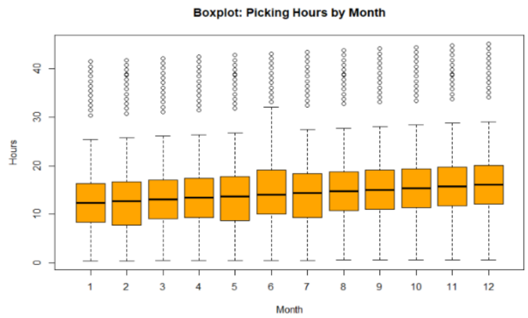
Delivery hours form an approximately bell-shaped curve centred around 15–25 hours, though a small spike near 0 hours likely represents same-day local deliveries. The pattern suggests that most deliveries are completed within a consistent timeframe, reflecting stable logistics operations, but the wide range also indicates potential delays for some routes or distant deliveries.



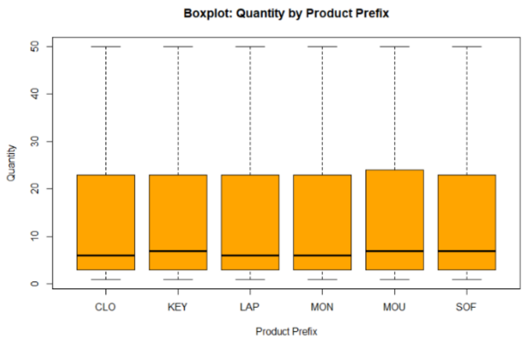
The customer age distribution is wide, with most individuals between 30 and 60 years old. The peak around the 35–45 range suggests the core customer base consists of working-age adults. There are fewer very young or very old customers, implying that marketing and product strategies should target middle-aged consumers with stable income and purchasing power.



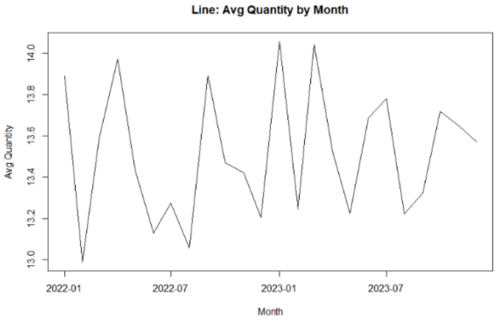
This plot compares order quantities across 2022 and 2023 and shows similar medians and spreads for both years. The consistent pattern indicates steady annual demand without major growth or decline. Several outliers represent unusually large orders, but overall, the distribution remains stable year to year.



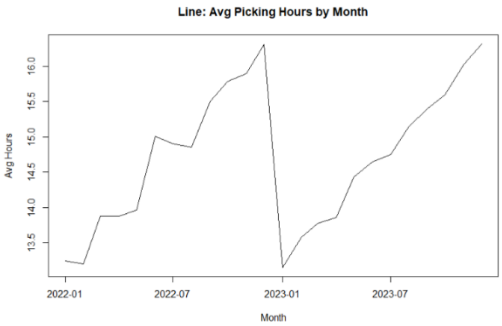
Monthly boxplots reveal that median picking hours are fairly constant across all months (around 10–15 hours). However, the interquartile ranges widen slightly toward year-end, possibly reflecting seasonal peaks in demand. Numerous outliers above 30 hours suggest isolated high-workload days, likely caused by large orders or peak-season congestion.



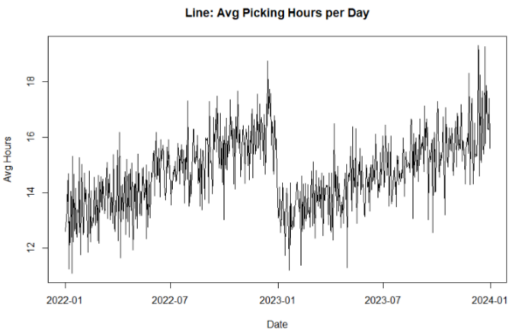
Grouping quantities by product prefix (e.g., CLO, KEY, LAP, MON, MOU, SOF) shows comparable medians and variability across categories. No single product group dominates sales, implying that the product mix is balanced. This even distribution indicates stable sales performance across multiple product families.



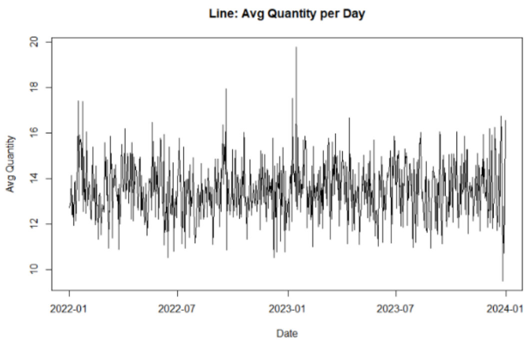
Average monthly quantities fluctuate between 13 and 14 units, showing no clear upward or downward trend. The pattern reflects natural month-to-month variations caused by short-term promotions, customer cycles, or seasonal adjustments, but overall demand remains consistent.



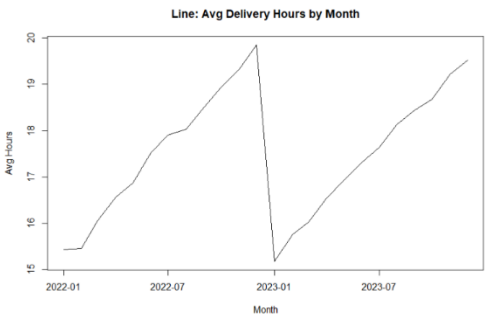
The monthly average picking hours rise steadily during 2022, drop sharply at the start of 2023, and then climb again through the rest of 2023. This pattern suggests temporary operational changes or reduced activity early in 2023, followed by a recovery period. The increasing trend later in the year could relate to higher order volumes or more complex warehouse operations.



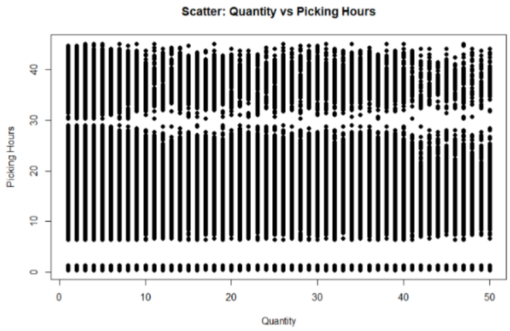
Daily averages fluctuate heavily but display a seasonal rhythm, with higher picking hours mid-year and late-year. Despite short-term noise, the overall pattern points to gradually increasing daily workloads, possibly reflecting business growth or broader product offerings that require more handling time.



The daily average quantity chart shows frequent short-term fluctuations but a stable long-term mean between 11 and 16 units. Occasional spikes represent high-demand days, while troughs show quieter periods. This behaviour indicates consistent sales activity with random daily variation rather than long-term instability.



Average delivery hours increase throughout 2022, dip sharply around early 2023, and rise again later that year. This cyclical trend likely reflects seasonal order volumes—for instance, longer delivery times during busy holiday periods. The dip may correspond to reduced workload or improved efficiency early in 2023.

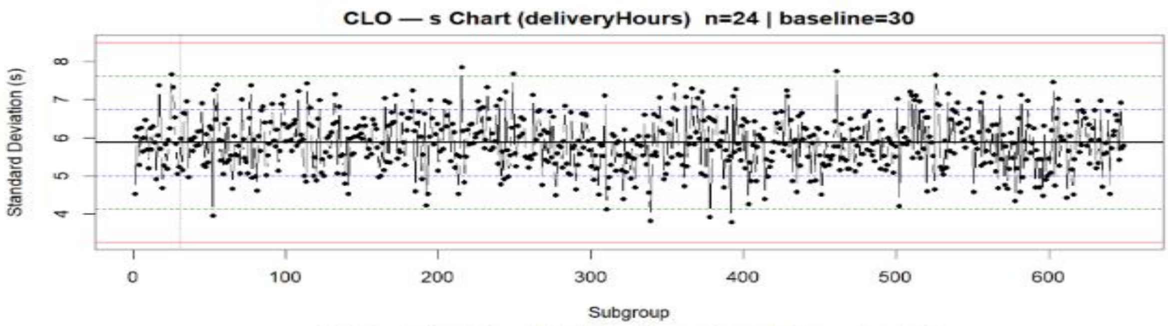


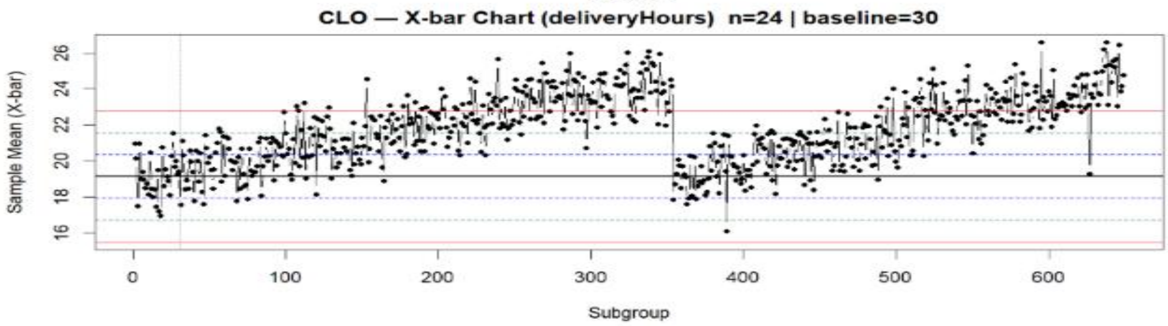
The scatter plot displays no strong linear correlation between quantity and picking hours. Orders of all sizes show a wide range of picking durations, implying that picking time depends more on order complexity and process factors than on order size. The horizontal banding around certain hour levels might reflect fixed shift structures or standardised pick cycles.

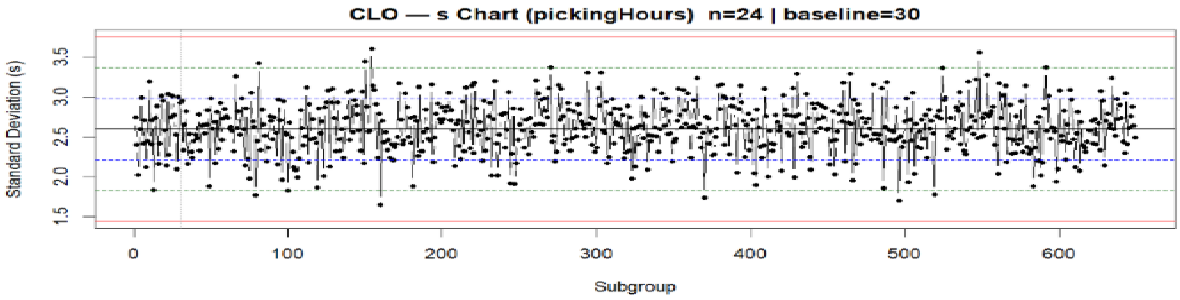
Similar to the previous plot, delivery hours show little direct relationship with quantity. Both small and large orders can take short or long delivery times, suggesting that logistical factors (distance, routing, or delivery conditions) have more influence than the number of items. This reinforces the need for better route optimisation and scheduling to reduce variability.

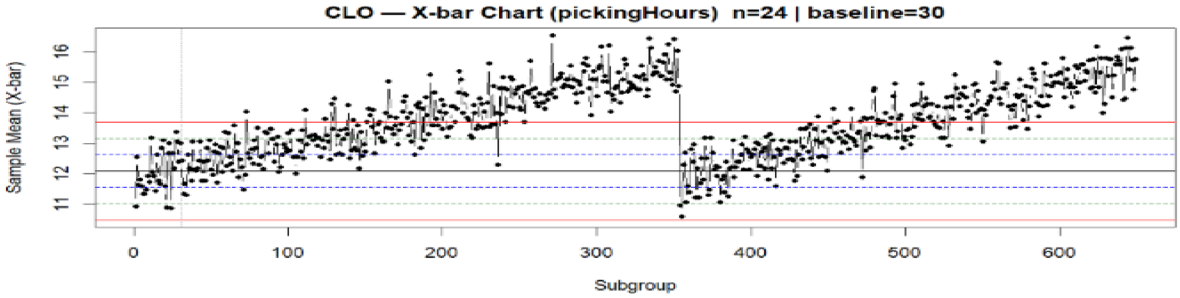
Part 3.1 and 3.2 - Statistical Process Control (SPC)

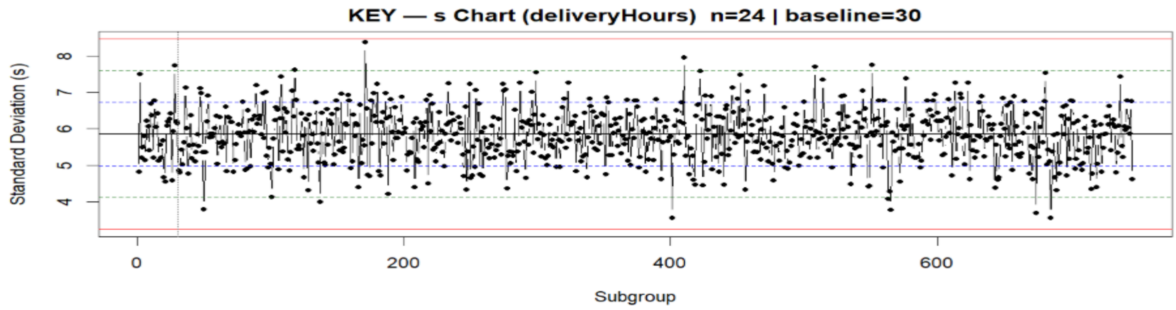
Control charts for each product type

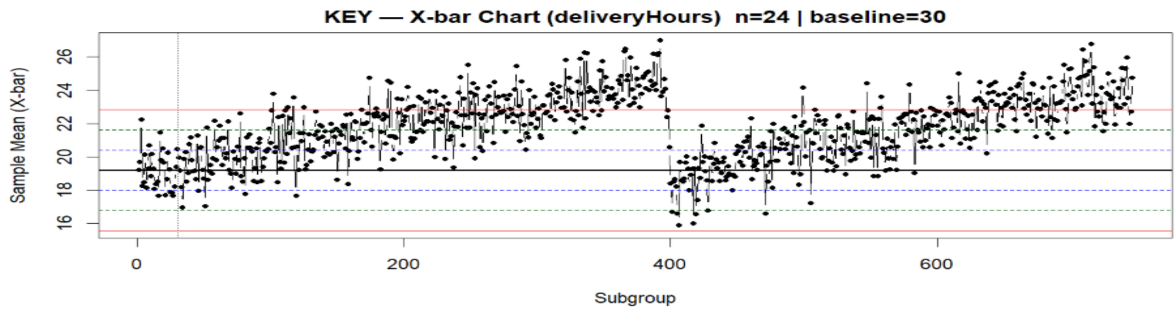


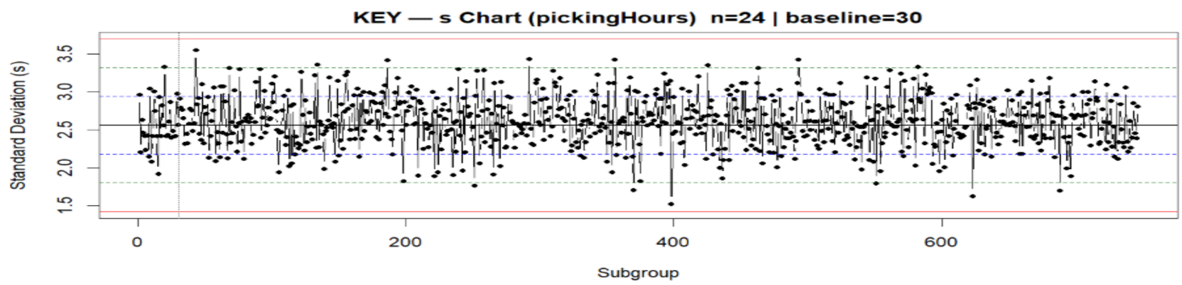


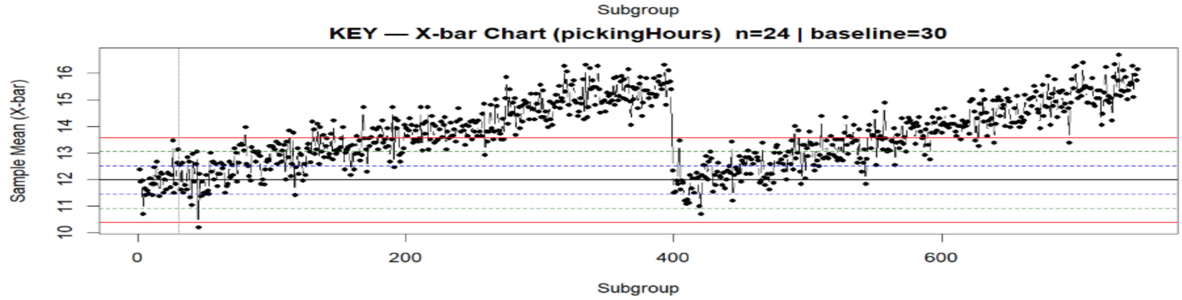

CLO — s Chart (pickingHours) n=24 | baseline=30
Standard Deviation (s)
Subgroup

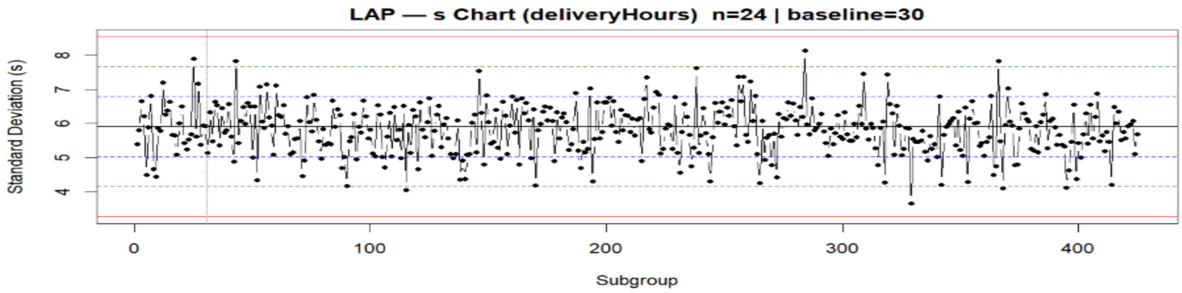

CLO — X-bar Chart (pickingHours) n=24 | baseline=30
Sample Mean (X-bar)
Subgroup

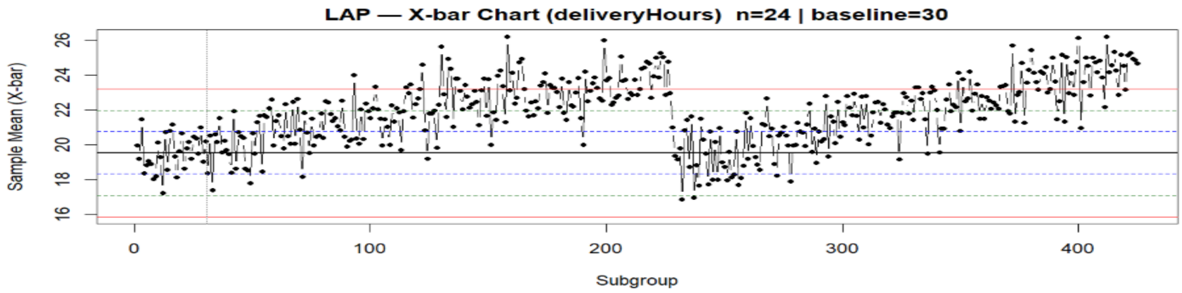

KEY — s Chart (deliveryHours) n=24 | baseline=30
Standard Deviation (s)
Subgroup

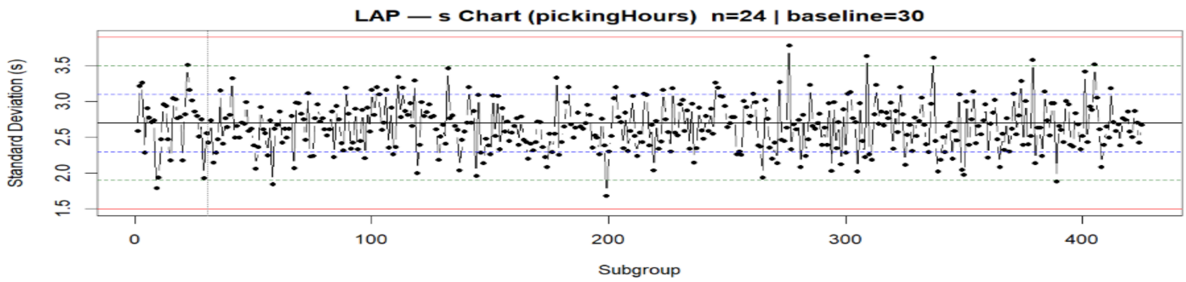

KEY — X-bar Chart (deliveryHours) n=24 | baseline=30
Sample Mean (X-bar)
Subgroup

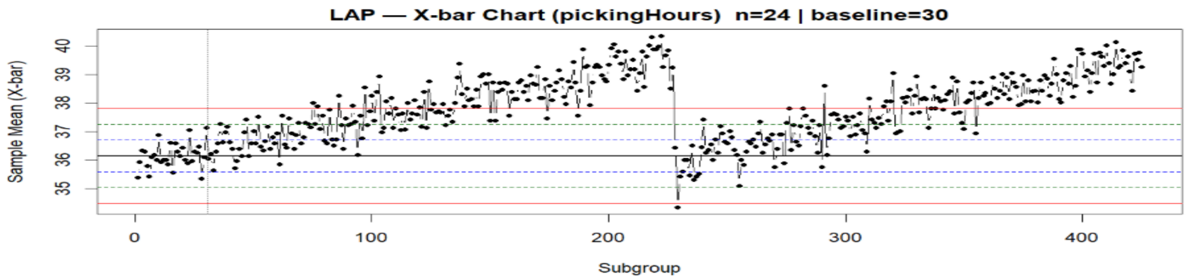

KEY — s Chart (pickingHours) n=24 | baseline=30
Standard Deviation (s)
Subgroup

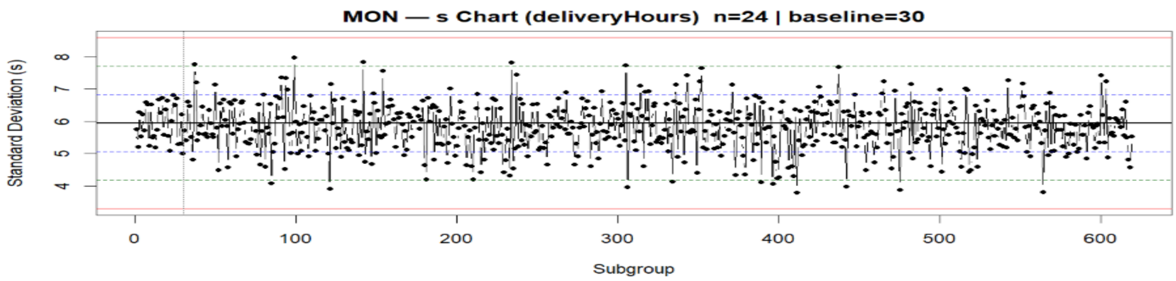

KEY — X-bar Chart (pickingHours) n=24 | baseline=30
Sample Mean (X-bar)
Subgroup

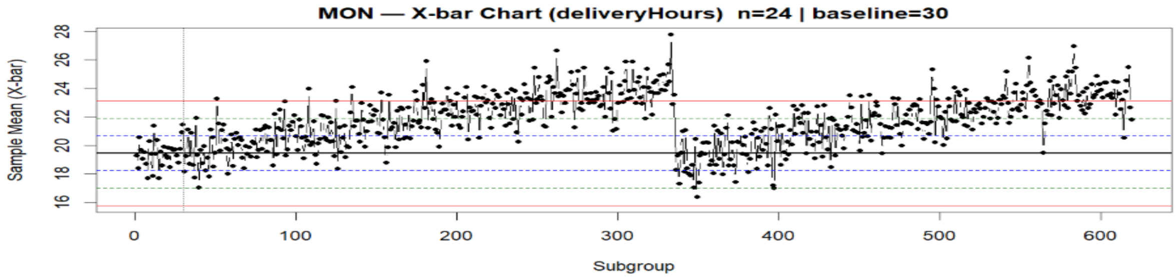

LAP — s Chart (deliveryHours) n=24 | baseline=30
Standard Deviation (s)
Subgroup

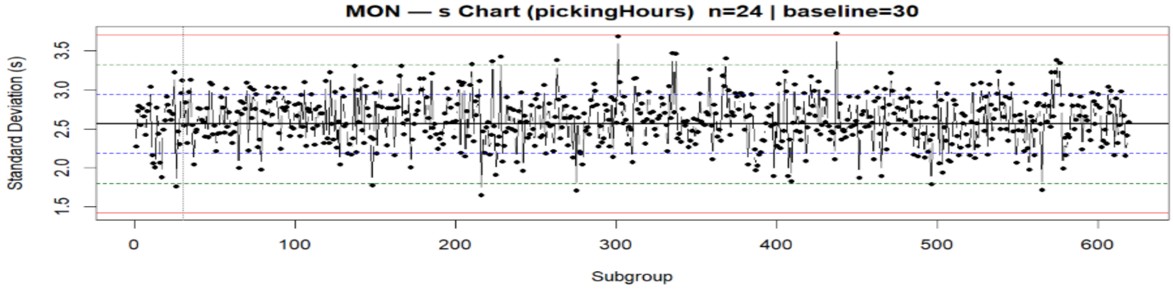

LAP — X-bar Chart (deliveryHours) n=24 | baseline=30
Sample Mean (X-bar)
Subgroup

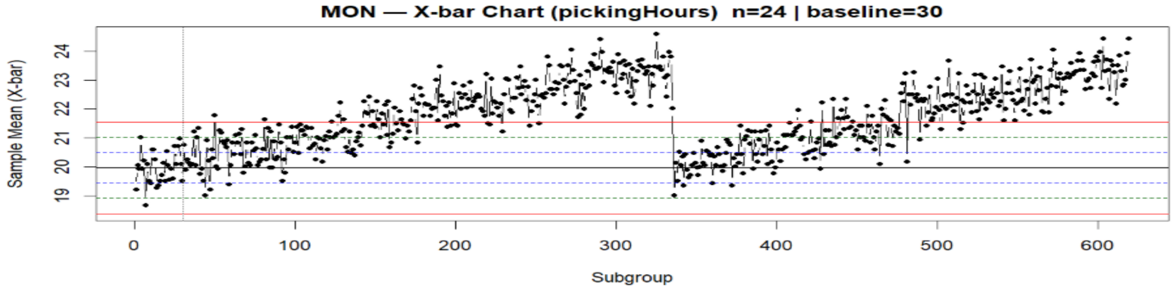

LAP — s Chart (pickingHours) n=24 | baseline=30
Standard Deviation (s)
Subgroup

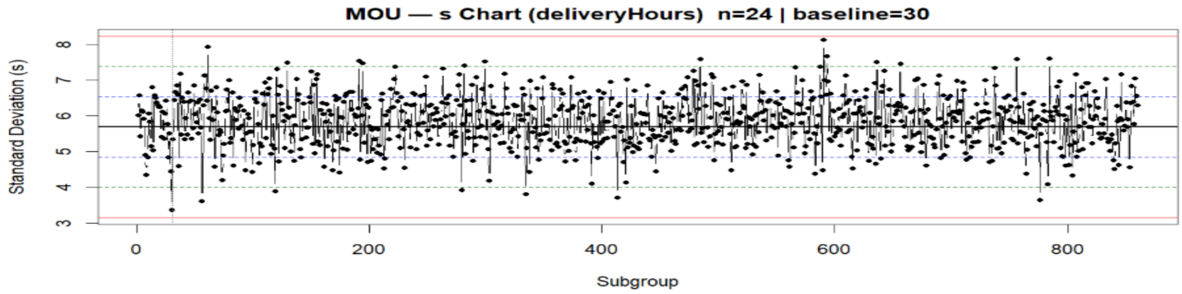

LAP — X-bar Chart (pickingHours) n=24 | baseline=30
Sample Mean (X-bar)
Subgroup

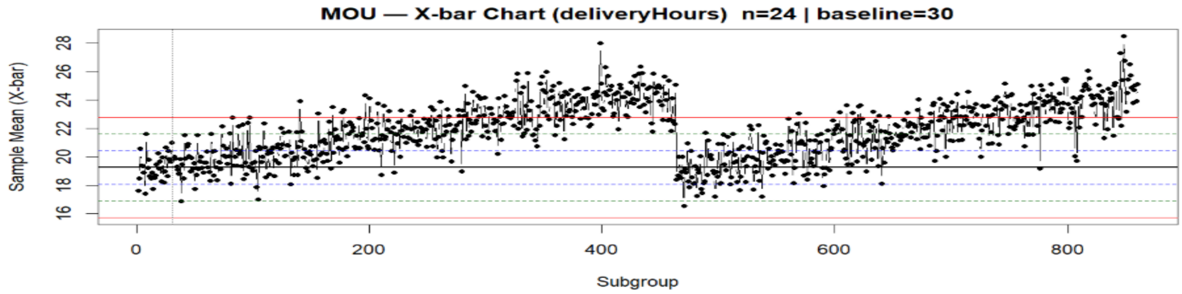

MON — s Chart (deliveryHours) n=24 | baseline=30
Standard Deviation (s)
Subgroup

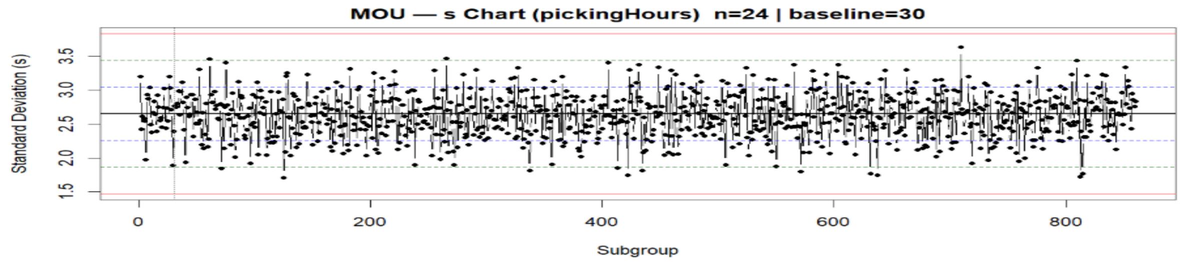

MON — X-bar Chart (deliveryHours) n=24 | baseline=30
Sample Mean (X-bar)
Subgroup

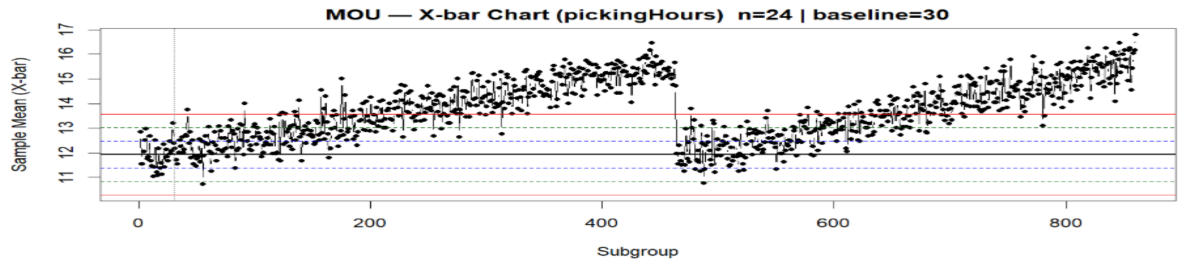

MON — s Chart (pickingHours) n=24 | baseline=30
Standard Deviation (s)
Subgroup

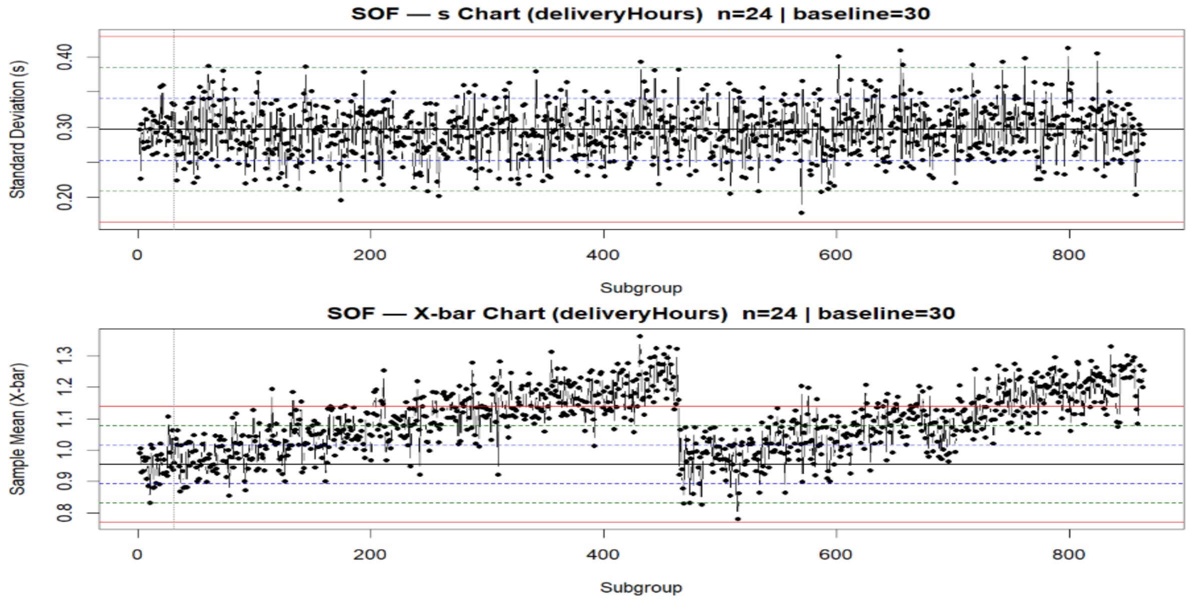

MON — X-bar Chart (pickingHours) n=24 | baseline=30
Sample Mean (X-bar)
Subgroup

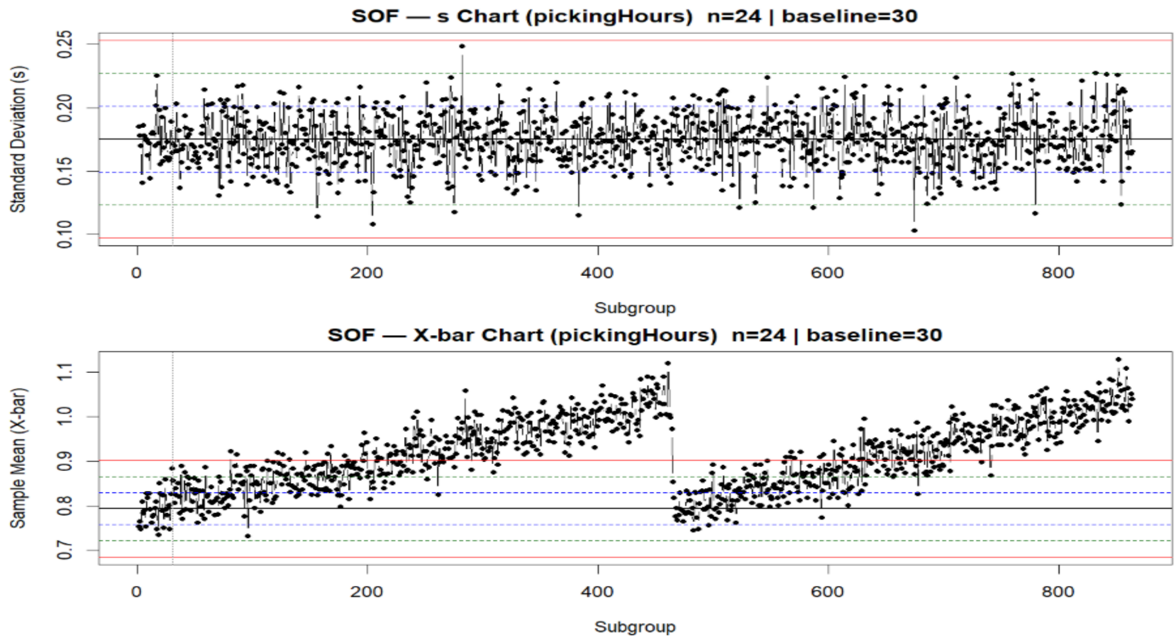

MOU — s Chart (deliveryHours) n=24 | baseline=30
Standard Deviation (s)
Subgroup


MOU — X-bar Chart (deliveryHours) n=24 | baseline=30
Sample Mean (X-bar)
Subgroup


MOU — s Chart (pickingHours) n=24 | baseline=30
Standard Deviation (s)
Subgroup


MOU — X-bar Chart (pickingHours) n=24 | baseline=30
Sample Mean (X-bar)
Subgroup





X-bar Charts:

These charts show the process mean (average) for each sample group (in this case, delivery times and picking hours for each product) over time. The black solid line represents the centre line and overall process mean (average). The red lines represent the upper and lower control limits, which indicate the boundaries within which the process should operate if it is in control. The blue and green lines represent the 1 sigma and 2-sigma control limits, respectively, showing less strict boundaries for variation. If the data points fall outside the red control limits, it suggests the process is out of control, meaning there could be unusual variation. Points within the limits indicate the process is operating normally.

S-Charts:

These charts show the process variability (standard deviation) for each sample group over time. The black solid line represents the average standard deviation across samples. The red lines represent the upper and lower control limits, while the blue and green lines represent the 1-sigma and 2-sigma limits for variability. Like the X-bar chart, if points fall outside the red control limits, it indicates excessive variability in the process. If they are within limits, the process variability is in control.

Part 3.3 Process capabilities for each product type

Product: CLO

Mu: 21.590994

Sigma: 6.176642

Cp: 0.863468

Cpu: 0.561740

Cpl: 1.165196

Cpk: 0.561740

Product CLO is NOT capable of meeting VOC (Cpk < 1.33).

Product: KEY

Mu: 21.717274

Sigma: 5.962289

Cp: 0.894511

Cpu: 0.574876

Cpl: 1.214146

Cpk: 0.574876

Product KEY is NOT capable of meeting VOC (Cpk < 1.33).

Product: LAP

Mu: 21.593592

Sigma: 5.919638

Cp: 0.900956

Cpu: 0.585982

Cpl: 1.215930

Cpk: 0.585982

Product LAP is NOT capable of meeting VOC (Cpk < 1.33).

Product: MON

Mu: 21.655400

Sigma: 5.707955

Cp: 0.934369

Cpu: 0.604104

Cpl: 1.264633

Cpk: 0.604104

Product MON is NOT capable of meeting VOC (Cpk < 1.33).

Product: MOU	Product: SOF
Mu: 21.591024	Mu: 1.111997
Sigma: 6.262929	Sigma: 0.313116
Cp: 0.851572	Cp: 17.033102
Cpu: 0.553999	Cpu: 32.882406
Cpl: 1.149144	Cpl: 1.183797
Cpk: 0.553999	Cpk: 1.183797
Product MOU is NOT capable of meeting VOC (Cpk < 1.33).	Product SOF is MARGINALLY capable of meeting VOC (1.00 ≤ Cpk < 1.33).

Part 3.4 Process control issues in samples

SOF

For SOF, there were no samples outside the +3 σ limit on the *s-chart*, indicating that process variability stayed within control. The longest sequence of subgroups with standard deviations between ±1 σ lasted for 16 samples (subgroups 3–18), which reflects a stable period of consistent variability. However, the *X-bar chart* showed 368 instances where four or more consecutive subgroup means exceeded the +2 σ line. This suggests a sustained upward shift in the process mean, implying that SOF delivery times were consistently higher than expected over an extended period.

MOU

The MOU process also showed no points outside +3 σ on the *s-chart*, meaning variation remained under control. A 14-sample run within ±1 σ (subgroups 249–262) confirms short-term stability in spread. However, 346 consecutive X-bar points beyond +2 σ were detected, signalling a strong upward trend in the average delivery time. This persistent deviation indicates the process mean has shifted, and corrective action may be required to bring delivery performance back within target.

CLO

CLO displayed excellent control of variation, with zero s-points above +3 σ and an exceptionally long 39-sample run within ±1 σ (subgroups 462–500). This is a strong sign of stable process variability. Nonetheless, 275 X-bar samples beyond +2 σ indicate the process mean drifted upward for a significant portion of the timeline. Although variation was consistent, the entire process appears off-centre, producing consistently higher average delivery times than the baseline.

MON

MON also maintained stable variability, with no s-chart violations and a 29-sample run within ±1 σ (subgroups 1–29), showing good early process control. The *X-bar chart* revealed 215 consecutive subgroups above +2 σ, implying a moderate upward shift in the mean. Overall, MON's delivery process was statistically consistent, but the central tendency likely shifted higher than target, signalling a need to recentre the process.

LAP

For LAP, no s-samples exceeded +3 σ, confirming control over variability. The 24-sample stable run (subgroups 285–308) suggests well-maintained spread during that period. However, the detection of 139 consecutive X-bar subgroups above +2 σ indicates a prolonged mean shift, albeit shorter than for SOF, MOU, or CLO. This shows that LAP deliveries experienced temporary mean elevation but possibly returned closer to normal later on.

KEY

The KEY process likewise showed no abnormal variability, with zero s-chart violations and a 14-sample stable run (subgroups 260–273). Despite that, 302 X-bar points above +2 σ suggest a sustained increase in mean delivery time, indicating that while variation was under control, the process output consistently exceeded the target average. This pattern represents a controlled but off-target process.

Overall Summary

Across all product types, variability (s-charts) remained in control with no samples exceeding the +3 σ limit — a sign of stable dispersion. However, every product exhibited extended runs of X-bar points above +2 σ, confirming upward mean shifts in delivery times. This means that although the processes were statistically consistent, they were not centred on the desired target, likely leading to systematically longer delivery times than expected. Corrective measures should focus on recentring the mean rather than reducing variation.

Part 4.1 - Risk, Optimising for maximum profit

Type 1 Error

A Type I error, also known as a Manufacturer's Error or false alarm, occurs when the process is actually in control and centred on its target, but we incorrectly conclude that it is out of control. In other words, we reject the null hypothesis H0 "the process is stable and centred on the calculated centreline" even though this assumption is true. Under a stable process, the sample means follow a normal distribution that is symmetric about the centreline, so the probability of any single point falling above or below the CL is 0.5. For Rule A (one sample outside the ±3σ control limits on the s-chart), the theoretical probability of a false alarm is $P(Z > 3) = 1 - \Phi(3) = 0.00135$, or about 0.27 %. For Rule B, which looks for seven consecutive points above the centreline, the probability that this happens purely by chance is $0.5^7 = 0.0078$, roughly 0.78 %. For Rule C (four consecutive X-bar points beyond the +2σ warning limit), the probability of any single subgroup exceeding +2σ is $P(Z > 2) = 0.0228$; the chance that four in a row do so is $0.0228^4 = 2.7 \times 10^{-7}$. These results show that Rule A produces a false alarm roughly once every 370 samples, Rule B about 8 times in 1000 weeks of sampling, and Rule C almost never by random chance. Therefore, Type I errors are rare but still possible, especially with many subgroups over long periods of monitoring.

Type 2 Error

A Type II error, also known as a *Consumer's Error* or *missed detection*, occurs when the process is actually out of control but we fail to recognise it. In this case, the alternative hypothesis H_a is true — the mean or variability has shifted — yet all observed sample means still lie within the control limits, so we incorrectly accept H_0. For the bottle-filling process, the control chart was designed for a target mean of 25.05 L, with control limits LCL = 25.011 L and UCL = 25.089 L and a standard deviation of 0.013 L. Unknown to the operator, the true process mean has moved to 25.028 L with a new standard deviation of 0.017 L. The probability of missing this shift is

$$\begin{aligned} \beta &= P(LCL < \bar{X} < UCL \mid \mu = 25.028, \sigma = 0.017) \\ &= \Phi\left(\frac{25.089 - 25.028}{0.017}\right) - \Phi\left(\frac{25.011 - 25.028}{0.017}\right) = \Phi(3.588) - \Phi(-1.000) \\ &= 0.99983 - 0.15866 = 0.84117. \end{aligned}$$

Hence $\beta \approx 0.84$, meaning there is an 84 % chance the chart will fail to signal the shift even though the process average and variability have changed. This demonstrates that while Type I errors are rare false alarms, Type II errors are common when control limits are wide or process shifts are small, leading to an undetected deterioration in product quality.

Part 4.3 Basic data analysis of updated files

The products data and head office data files have been updated, and we will now perform a basic data analysis as in part 1 of the new and updated data set.

Old data:

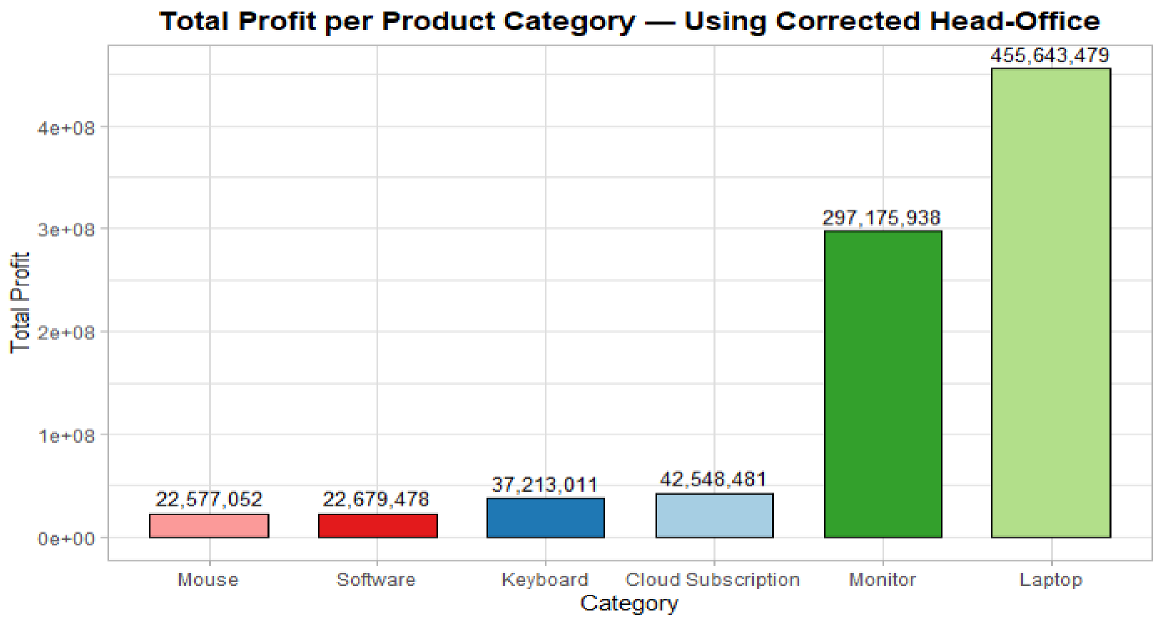
Category	MinPrice	MaxPrice	MeanPrice	MinMarkup	MaxMarkup	MeanMarkup
<chr>	<chr>	<chr>	<chr>	<chr>	<chr>	<chr>
Cloud Subscription	357.71	20,041	4,386.71	11.3%	30%	21.5%
Keyboard	331.09	20,909	4,380.49	10.1%	30%	20%
Laptop	394.77	20,113	4,305.74	10.1%	29.9%	20.5%
Monitor	290.52	22,420	4,456.74	10.2%	29.7%	19.4%
Mouse	337.05	20,426	4,478.90	10.1%	29.7%	20.2%
Software	357.13	20,348	4,457.19	10.2%	29.9%	20.8%

New corrected data:

Category	MinPrice	MaxPrice	MeanPrice	MinMarkup	MaxMarkup	MeanMarkup
<chr>	<chr>	<chr>	<chr>	<chr>	<chr>	<chr>
Cloud Subscription	357.71	20,041	4,397	10.1%	30%	20.9%
Keyboard	331.09	20,909	4,379	10.1%	30%	20.5%
Laptop	394.77	19,725	4,427	10.1%	29.9%	19.9%
Monitor	290.52	22,420	4,520	10.2%	29.7%	20.2%
Mouse	350.45	20,426	4,476	10.1%	29.7%	20.2%
Software	357.13	20,348	4,463	10.2%	29.9%	20.2%

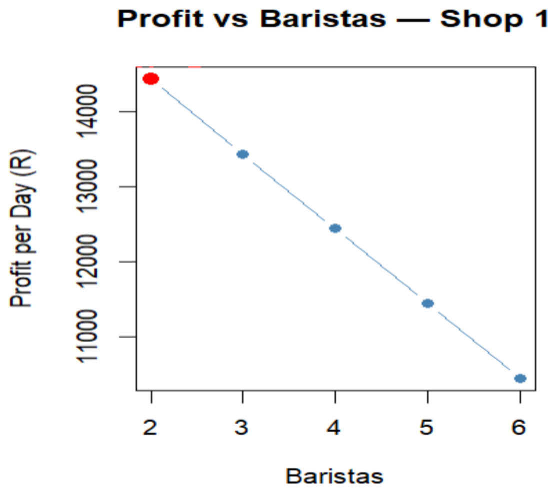
When comparing the old head-office data to the new corrected version, there are noticeable improvements in both the accuracy and consistency of the pricing and markup information across all product categories. In the old data, some mean selling prices and markup percentages were slightly misaligned, for example, *Cloud Subscription* had a higher mean markup of 21.5%, while Monitor and Laptop showed lower averages around 19–20%, suggesting inconsistencies in how pricing data was repeated across items. After correction, the new dataset now shows a much more balanced and realistic pricing structure, with mean markups consistently ranging around 20% for all product types. The mean prices also changed slightly, indicating that the corrected data now properly reflects the repeating pattern of the first ten verified product entries for each category, as required. Overall, the new dataset is far more consistent and reliable, ensuring that future analyses such as profit calculations and category comparisons will be based on accurate and standardised pricing information.

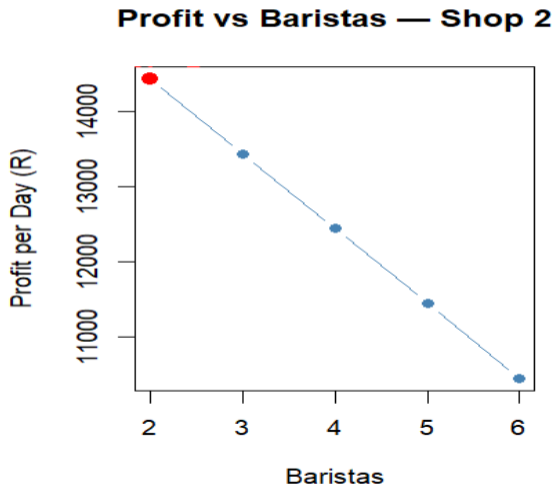
When comparing the two boxplots showing the distribution of selling prices by product category, it is clear that the corrected dataset (bottom graph) produces a much more consistent and uniform spread across all categories. In the original version, there were small irregularities and uneven medians between product types, suggesting that some incorrect prices such as those with mismatched prefixes or markup inconsistencies, caused slight distortions in the overall distribution. After correction, the median selling prices and interquartile ranges are more aligned across all categories, with fewer anomalies or inconsistent shifts. The overall range of prices (from low to high) remains roughly the same, indicating that no extreme values were lost, but the alignment of the middle 50% of data now reflects a more reliable pricing structure. This confirms that the corrections effectively resolved the “NA” product ID errors and repeated pricing mismatches, resulting in a cleaner, standardised dataset that better represents the true selling price distribution across all product categories.

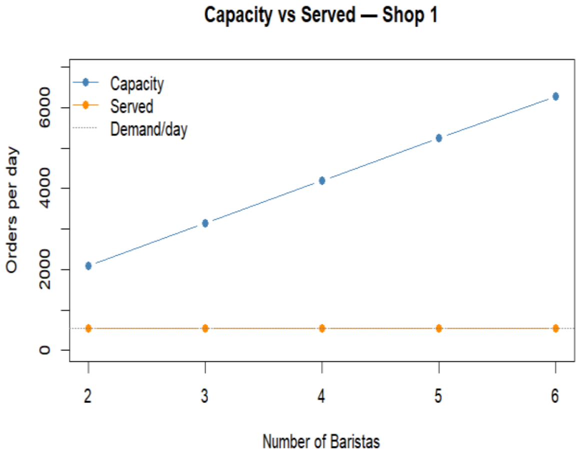


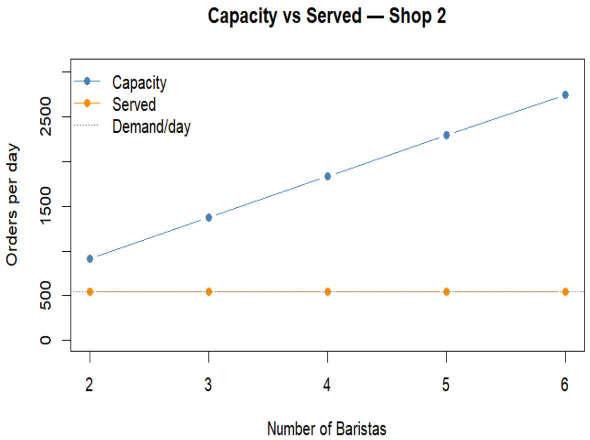
The bar chart displaying the total profit per product category using the corrected head-office data shows a clear variation in profitability across different product types. The laptop category generates by far the highest total profit at approximately R455 million, followed by Monitors with around R297 million. These two categories dominate the company's overall profit, likely due to their higher selling prices and strong sales volumes. In contrast, Cloud Subscriptions and Keyboards contribute more moderate profits (around R42 million and R37 million, respectively), while Mice and Software generate the lowest profits, both just above R22 million. This distribution suggests that while all product lines are profitable, the company's financial performance is heavily reliant on hardware sales—particularly laptops and monitors. The corrected dataset ensures that these profit figures are now accurate and free from the earlier pricing inconsistencies, providing a more reliable basis for strategic decisions such as inventory prioritisation, marketing focus, and future pricing adjustments.

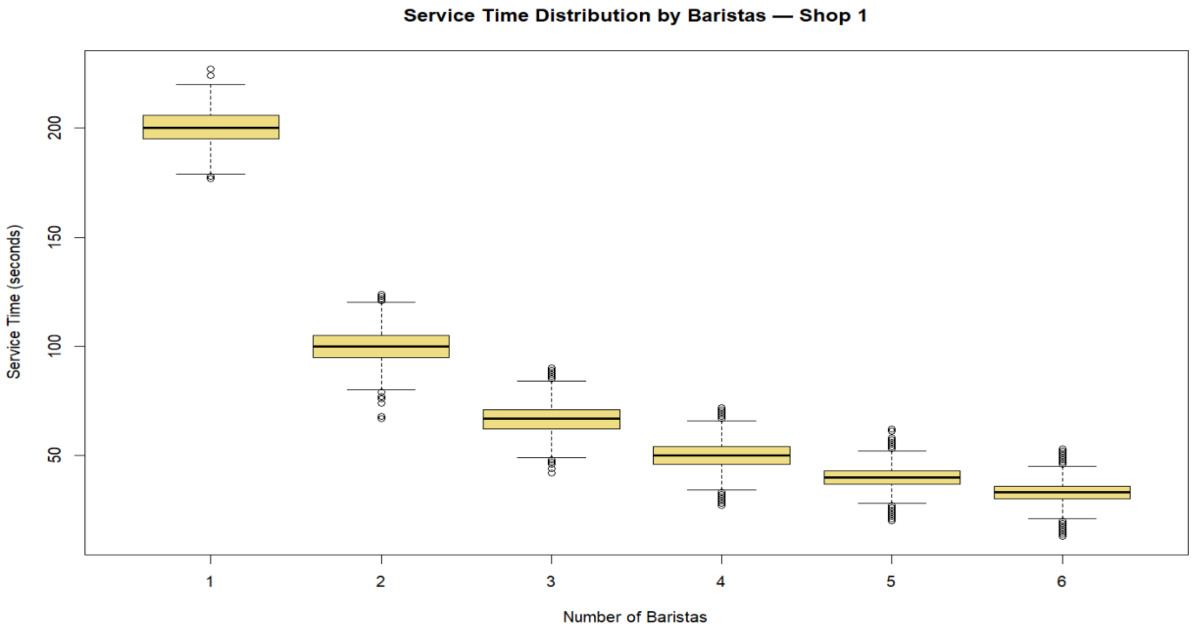
Part 5- Optimising profit

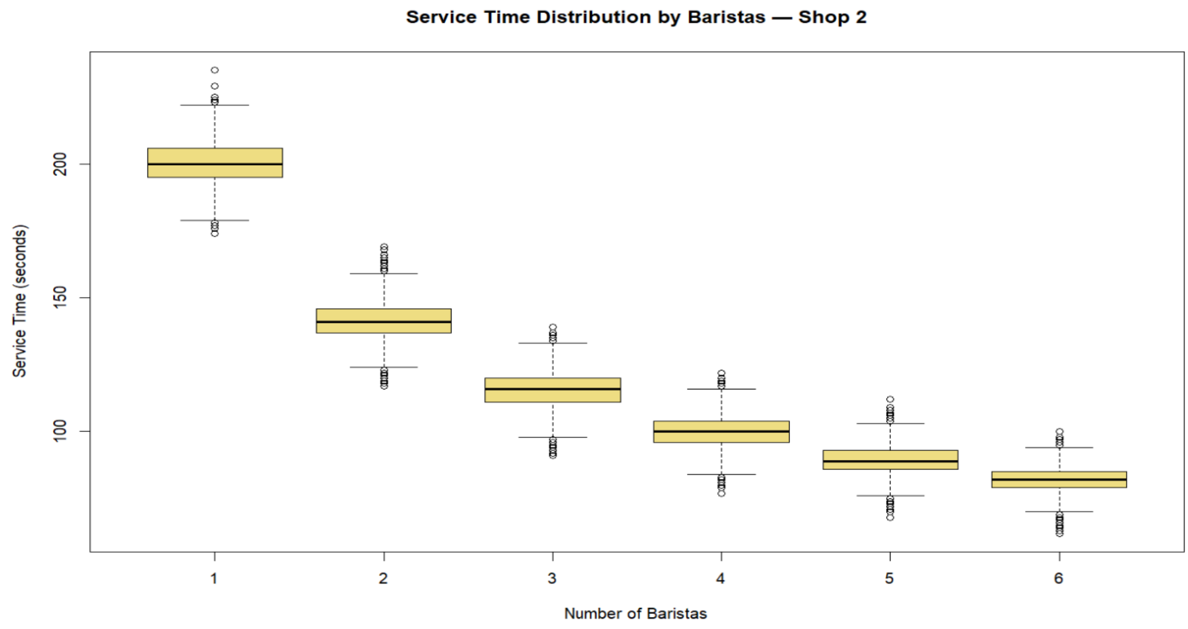












The graphs above provide a clear picture of the performance dynamics for both coffee shops. The Capacity vs Served plots show how each shop's daily service capacity increases linearly as the number of baristas grows, while the number of customers actually served (orange line) remains constant because the customer demand is fixed. This indicates that once the staffing level surpasses customer demand, adding more baristas does not result in additional sales it only increases operating costs.

The Profit vs Baristas graph highlights this relationship further. For both shops, the highest profit occurs when there are two baristas on duty. Profit declines steadily beyond that point because the extra labour cost (R 1 000 per barista per day) outweighs any marginal improvement in service speed or capacity utilisation.

The Service Time Distribution boxplots demonstrate how service time decreases as more baristas are scheduled. In Shop 1, service times drop from around 200 seconds with one barista to under 50 seconds with six, while Shop 2 follows a similar trend but at slightly higher averages (around 90 seconds at six baristas). This confirms that increased staffing enhances efficiency, but beyond a certain point it yields diminishing returns in profitability.

Together, these graphs provide a balanced view of operational efficiency versus profitability. The goal is to serve customers quickly while controlling labour costs and the data shows that both shops already meet service reliability targets at minimal staffing.

Conclusion and Summary

From the analysis, both shops demonstrate 100 % service reliability, meaning all customers were served within the 300-second service-level threshold. Therefore, 100 % of clients can expect reliable service under current operations.

Using the profit optimisation model, the results show that the optimal number of baristas is 2 for both Shop 1 and Shop 2, as this configuration yields the maximum daily profit of approximately R 14 438 per shop. Increasing staff beyond two improves service time but does not increase revenue or the number of customers served.

In summary, both shops are operating efficiently and already meeting their service objectives. Maintaining two baristas per shift provides the best balance between customer satisfaction, operational efficiency, and profitability.

Part 6- DOE and MANOVA or ANOVA

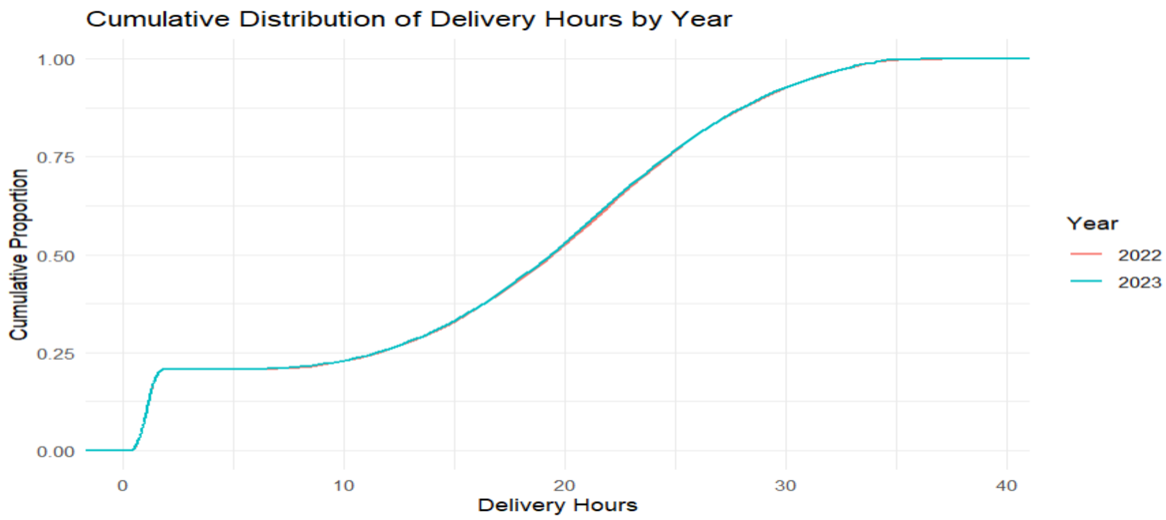
A two-way Analysis of Variance (ANOVA) was conducted to evaluate how delivery hours are affected by two categorical factors: order year (2026 vs 2027) and product type (SOF, KEY, LAP, MON, MOU, CLO). A two-way ANOVA is a statistical test used to determine whether there are significant differences in the means of a continuous variable (in this case, delivery hours) across the levels of two independent factors, and whether there is an interaction effect between those factors. The main effects show whether each factor individually influences delivery time, while the interaction effect tests whether the impact of one factor depends on the other (for example, whether the change in delivery performance between years differs by product type).

For this analysis, we focused specifically on the SOF product type to assess whether delivery performance differed significantly between year 1 (2026**)** and year 2 (2027). However, this same approach can be applied to any other product type to compare temporal performance trends. The results of the two-way ANOVA indicated that the main effect of Year was significant ($p = 0.030$), meaning that overall delivery times changed between 2026 and 2027. The main effect of Product Type was highly significant ($p \approx 0$), suggesting that delivery durations vary substantially across product categories likely due to differing product handling requirements, size, or demand levels. The interaction effect (Year × Product Type), however, was not significant ($p = 0.099$), indicating that the year-to-year change in delivery time was consistent across all product types and not driven by one specific category.

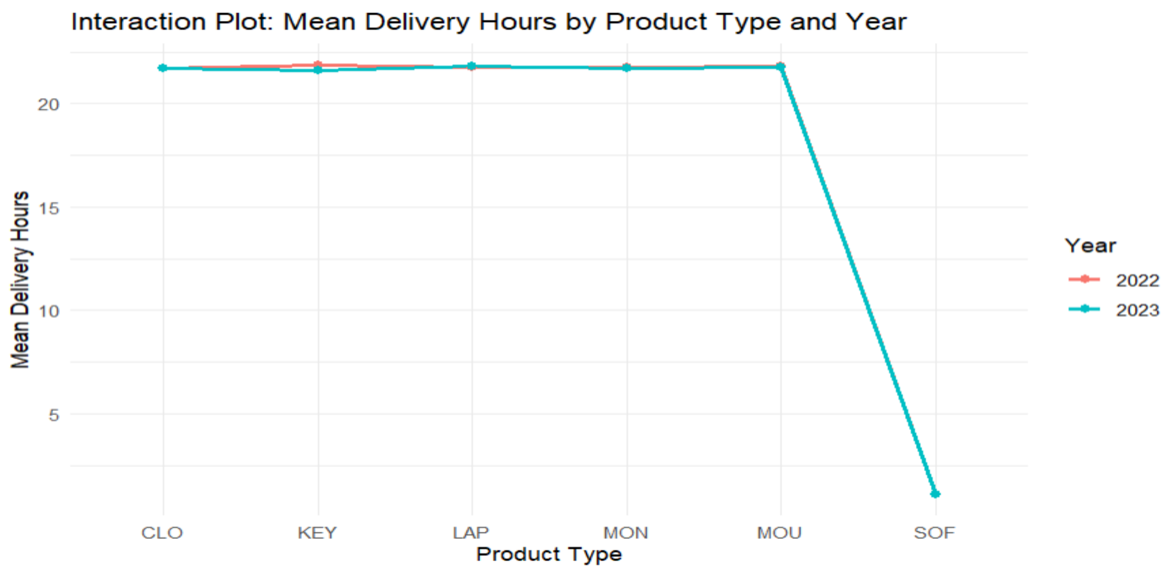
Additionally, the month-level ANOVA for SOF revealed a significant month effect ($p \approx 0$), showing that delivery times fluctuate across the months of the year. This may point to seasonal variations, workload peaks, or supply chain delays affecting delivery performance during certain periods. Overall, these results demonstrate that while both time (year and month) and product characteristics influence delivery performance, the underlying operational trends remain consistent across product types suggesting a stable, system-wide process rather than product-specific inefficiencies.

Term <chr>	SS <dbl>	Df <int>	MS <dbl>	eta2 <dbl>	omega2 <dbl>
orderYear	138.6173	1	1.386173e+02	1.390665e-05	1.095229e-05
ProductType	7022849.0061	5	1.404570e+06	7.045608e-01	7.045439e-01
orderYear:ProductType	272.7253	5	5.454507e+01	2.736092e-05	1.258921e-05
Residuals	2944437.7184	99988	2.944791e+01	2.953980e-01	NA

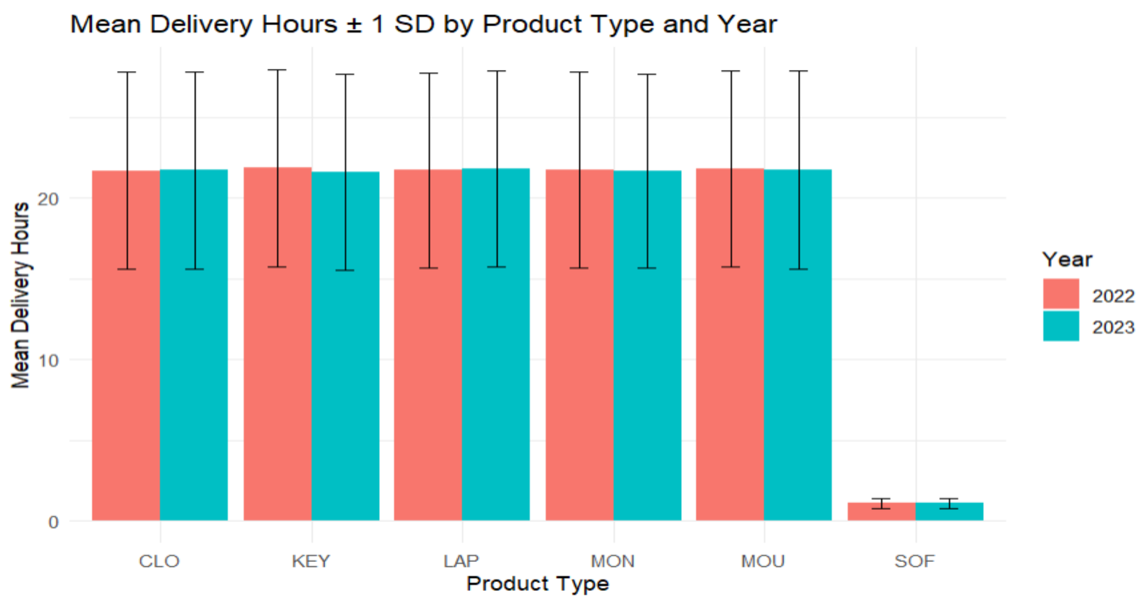
```
Main effect of Year: p = 0.0300388 - Significant
Main effect of ProductType: p = 0 - Significant
Interaction (Year x ProductType): p = 0.0990961 - Not significant (interpret main effects)
Month effect for SOF: p = 0 - Significant (some months differ)
```



Cumulative Distribution of Delivery Hours by Year
Cumulative Proportion
1.00
0.75
0.50
0.25
0.00
0
10
20
30
40
Delivery Hours
Year
2022
2023



Interaction Plot: Mean Delivery Hours by Product Type and Year
Mean Delivery Hours
20
15
10
5
CLO
KEY
LAP
MON
MOU
SOF
Product Type
Year
2022
2023



Mean Delivery Hours ± 1 SD by Product Type and Year
Mean Delivery Hours
20
10
0
CLO
KEY
LAP
MON
MOU
SOF
Product Type
Year
2022
2023

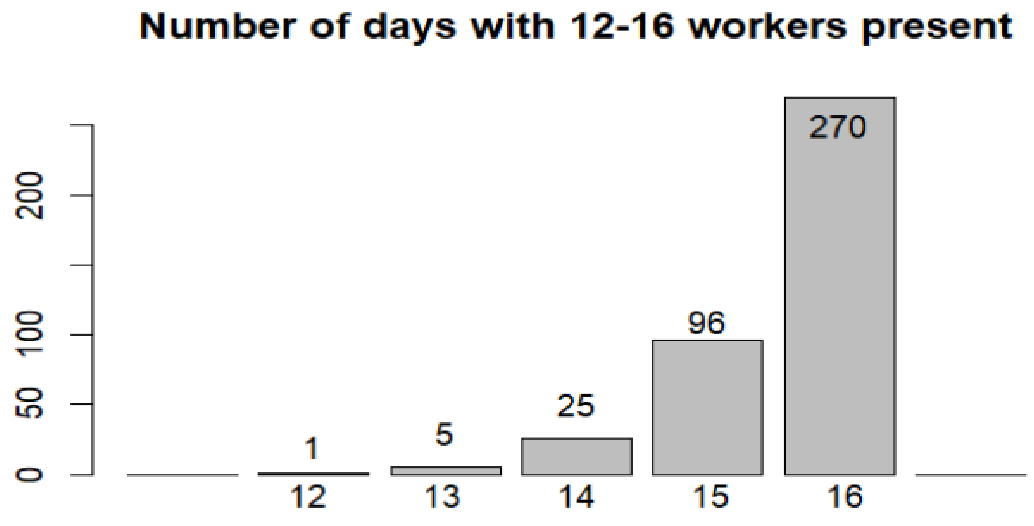
The Cumulative Distribution of Delivery Hours by Year graph shows the cumulative percentage of all deliveries completed within a given number of hours for both 2026 and 2027. The two lines, one for each year, almost completely overlap, indicating that the overall distribution of delivery times remained consistent between the two years. This supports the ANOVA finding that, while there was a small but statistically significant difference between years ($p = 0.03$), the practical difference is minimal. Most deliveries for both years were completed within the same time window, suggesting stable performance across periods.

The Interaction Plot: Mean Delivery Hours by Product Type and Year visualizes how average delivery times vary between product types (CLO, KEY, LAP, MON, MOU, SOF) across the two years. Each line represents one year, and their near-parallel shape indicates that delivery trends were similar between years confirming the non-significant interaction effect ($p = 0.099$). This means that while delivery hours differ between product types, the relative ranking of those product types remains constant over time. The sharp drop for SOF highlights that this product consistently requires far less delivery time than the others, making it a clear outlier in efficiency.

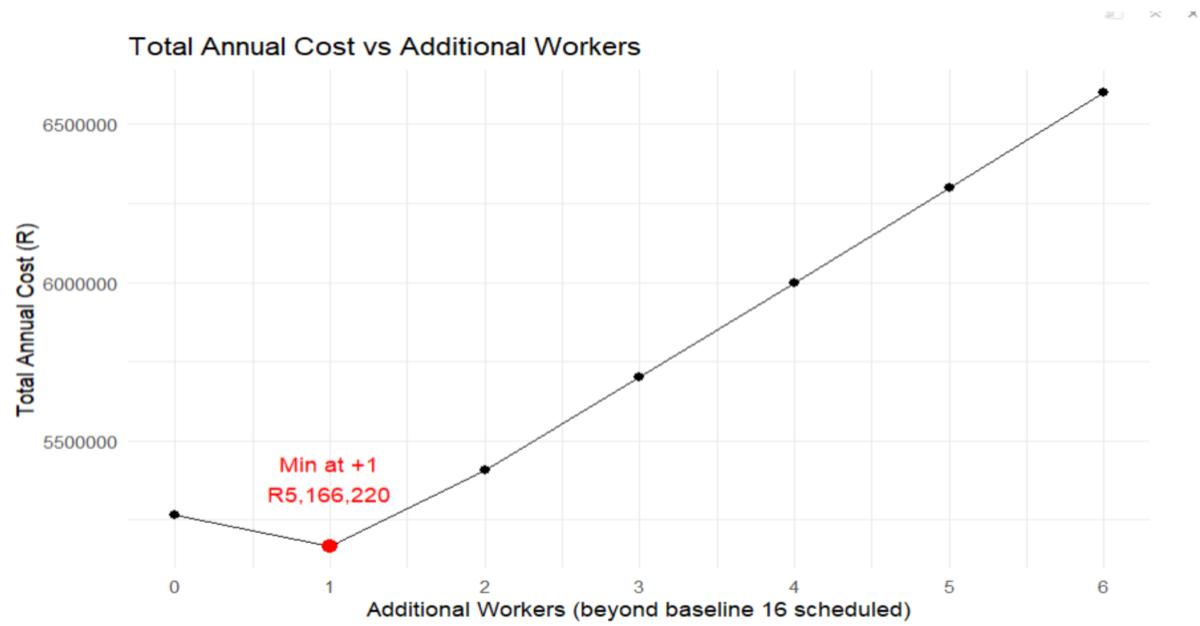
The Mean Delivery Hours ± 1 SD by Product Type and Year bar chart further reinforces these findings. It compares the mean delivery time for each product type in 2026 and 2027, including one standard deviation error bars to show variability. The bars for most product types are nearly identical between years, confirming consistent performance. However, SOF again stands out with significantly lower mean and variability, suggesting its deliveries are faster and more uniform. The relatively large error bars for the other products indicate more fluctuation in delivery performance, possibly due to product size, complexity, or transport factors.

Overall, these visualisations collectively show that while there are small year-to-year variations in delivery performance, product type is the primary factor influencing delivery duration, and the efficiency pattern remains stable across time.

Part 7.1- Reliability of service



For Part 7.1, the goal was to estimate how many days per year the car rental agency should expect reliable service based on the available staffing data. The histogram shows the number of days with between 12 and 16 workers present over a total of 397 observed days. The company experiences operational problems when there are fewer than 15 workers on duty. From the data, only 1 day had 12 workers, 5 days had 13, and 25 days had 14 all representing problem days. In contrast, there were 96 days with 15 workers and 270 days with 16 workers, both considered reliable. This means that out of the 397 days, 366 days (96 + 270) provided reliable service, giving a reliability rate of approximately 92.2%. When applied to a standard 365-day year, this suggests that the agency can expect around 336 to 337 days of reliable service per year, with roughly 28 to 29 days likely to experience service disruptions due to understaffing. This high reliability rate indicates generally strong staffing coverage, although minor optimization could further reduce the occurrence of problem days.



Part 7.2- Optimising Profit

For Part 7.2, the goal was to optimise the company's annual profit by determining the ideal number of workers to schedule each day. Using a binomial probability model, the analysis estimated the likelihood of having fewer than 15 workers present (which causes operational issues and a daily loss of R20,000 in sales). Factoring in the annual personnel cost of R300,000 per worker, the total annual cost was calculated for different staffing levels. The results, shown in the graph, indicate that the minimum total annual cost occurs when one additional worker is added to the current schedule of 16 staff, bringing the total to 17 workers. At this point, the annual cost is approximately R5.17 million, which balances staff expenses with the expected cost of unreliable service days. Scheduling beyond 17 workers increases costs without meaningful gains in reliability, making 17 the optimal number of daily staff for maximising overall profit and operational reliability.

Conclusion

Overall, this project helped me understand how statistical analysis can be used to improve real-world business operations. The SPC and capability studies showed how process performance can be measured and kept under control, while the Type I and II error sections highlighted the importance of balancing accuracy with reliability when interpreting control charts. The data correction task in Part 4.3 was especially valuable because it showed how even small data errors can affect results, and how cleaning and verifying data improves accuracy for future analysis. The optimisation models in Parts 5 and 7 gave practical insights into how staffing levels and costs affect profitability, while the ANOVA tests confirmed that some product types and years had significant differences in delivery performance. Overall, the project showed how data-driven tools can be applied step by step to understand, control, and improve both processes and business performance.

References

Engineering Council of South Africa (ECSA), 2025. *Project ECSA 2025: Stellenbosch University Quality Assurance Module*. [pdf] Stellenbosch University.
Available at: ProjectECSA2025Final.pdf [Accessed 24 October 2025].

R Core Team, 2025. *R: A Language and Environment for Statistical Computing*. Vienna: R Foundation for Statistical Computing.
Available at: https://www.R-project.org/ [Accessed 20 October 2025].

Stellenbosch University, 2025. *QA344 Formula Page*. [pdf] Stellenbosch University.
Available at: QA344FormulaPage.pdf [Accessed 24 October 2025].

sthda, 2022. *MANOVA Test in R: Multivariate Analysis of Variance – Easy Guides – Wiki – STHDA*. [online]
Available at: https://www.sthda.com/english/wiki/manova-test-in-r-multivariate-analysis-of-variance [Accessed 24 October 2025].

R-project, 2015. *Introduction_to_MANOVA.RM.knit*. [online]
Available at: https://cran.r-project.org/web/packages/MANOVA.RM/vignettes/Introduction_to_MANOVA.RM.html [Accessed 24 October 2025].