Maluleka Khanyisa
23511583
24 October 2025

QUALITY ASSURANCE 344
ECSA PROJECT

Maluleka, KA, Mej [23511583@sun.ac.za]

TABLE OF CONTENTS

INTRODUCTION

Background

Businesses rely heavily on data-driven solutions to make decisions which aim to maximise profitability, ensure great quality, and optimise processes. This project applies statistical methods to real-world business situations, such as delivery services and service optimisation.

Objectives

The primary goal for this report is to apply statistical methods to real-life datasets and then interpret the results in the context of businesses. The objectives are as follows:

1. Perform descriptive statistics on customer, products and sales data.
2. Implement Statistical Process Control methods to monitor the performance of delivery processes.
3. Calculate the capability of the process

4. Go over Type I and Type II errors in quality control
5. Improve the number of workers needed to maximise profitability and prevent unnecessary costs
6. Model car delivery reliability services

Motivation

The motivation for this project is to bridge the gap between raw data and practical business decisions. In any company, collecting data is easy, but using it effectively to improve performance is the main challenge. An Industrial Engineer's role is to solve problems, and the methods in this report are the tools to do that. Statistical Process Control (SPC) is essential for finding and fixing problems in a process before they cost the company money or lose customers. Similarly, optimisation models for staffing show how to balance costs (like salaries) against service quality (like customer wait times) to find the most profitable solution. The motivation, therefore, is to demonstrate how these statistical and engineering principles are applied to real data to find real savings, improve quality, and make a business more efficient and competitive.

METHODOLOGY

- Descriptive Statistics
- Statistical Process Control (SPC)
- Process Capability Analysis
- Hypothesis Testing

PART 1 (DESCRIPTIVE STATISTICS)

CUSTOMERS

The dataset consists of 5000 customers in total, categorised by gender, age, income, and the city each customer is from.

Gender

Among our customers, 2432 are females, which makes up 48.6% of the total, and 2350 (47%) are male. This leaves 218 customers, who make up 4.36% of the customers, and are labelled as ‘other’. This other category may include customers whose gender was not captured, incorrectly inputting the data, or even people who do not identify as either of the two classifications. This shows us that the data is distributed almost fairly, with slightly more females than males. However, the 218 unclassified customers bring uncertainty and make this comparison inconclusive as these customers could influence the decision in either way, either by contradiction or validation. Regardless, the data suggests that the business attracts almost the same number of males versus females. Therefore, this feature isn't a significant differentiating factor.

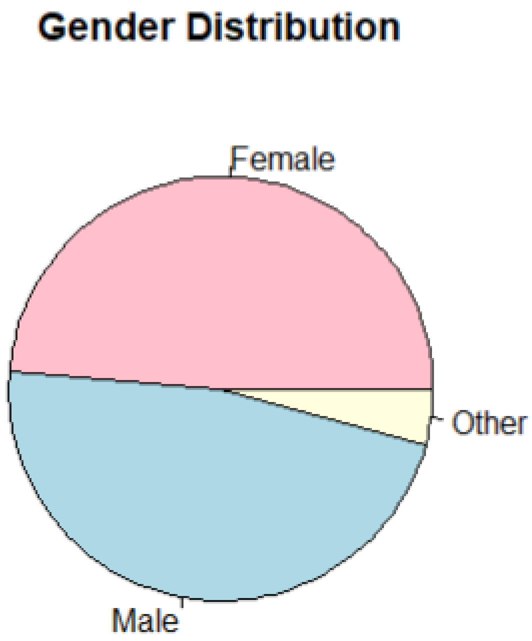

Figure 1: Customer Gender

Age

After analysing the gender, we turn our attention to customer distribution by age, which can give us insight into the range and which age group is most represented. The youngest customer is 16 years old, with the oldest being 105. This gives us a good understanding of the range. From this, it is evident that it covers a very broad group, not just a certain age group. However, the age that appears most frequently is 31, with 103 instances, followed by 35 with 101 appearances. This indicates that the business's target market should be people around that age. On the other hand of the stick, the older customers, especially above 100, don't appear frequently. Moreover, the average age of the customer is approximately 51, which indicates that most of the customers are middle-aged, with a standard deviation of x, which indicates moderate variability. The conclusion we draw from this is that young adults are more likely to have an interest in the product than older people. This is a significant differentiating factor and tells the business where to place most of its marketing, without ignoring the minority still.

To better visualise this, the histogram below shows the distribution of the customers by age. The chart clearly indicates the concentration of ages around the mode.

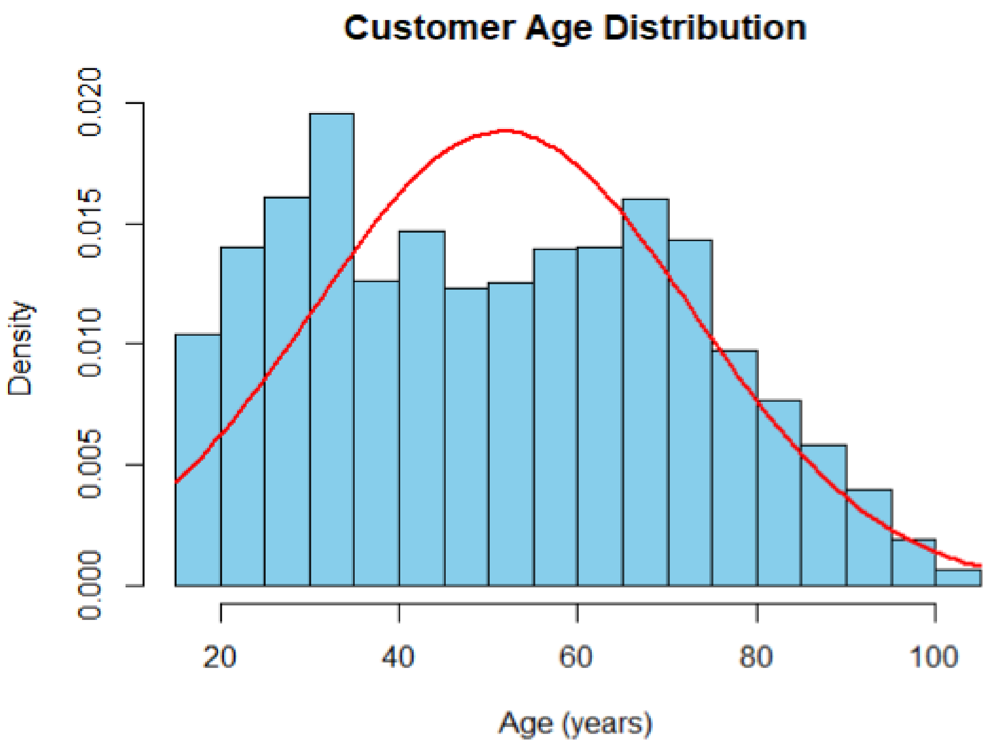

Figure 2: Customer Age

Income

The next feature under analysis is customer income. This will give us an idea of our demographic in terms of affordability and market class. Understanding this will help us determine whether the business primarily attracts lower-, middle-, or upper-income individuals

In our dataset, the minimum recorded income is R5000, while the maximum is R140000. This indicates a broad range of purchasing power among the customers. The average income is R80797, and the median is R85000, which indicates that half of the customers fall below this amount, whereas the rest are above it. The standard deviation of R33150.11 shows variation in the income levels, and a higher value would indicate that their earnings are dispersed over a large range. The mode income is R105000, which is the most frequent income among the customers.

This analysis suggests that the business's target market lies in the upper middle-class individuals, therefore that this is the group that should be prioritised in terms of marketing, to get the maximum exposure we would need. The figure below visualises the distribution of the income among the customers.

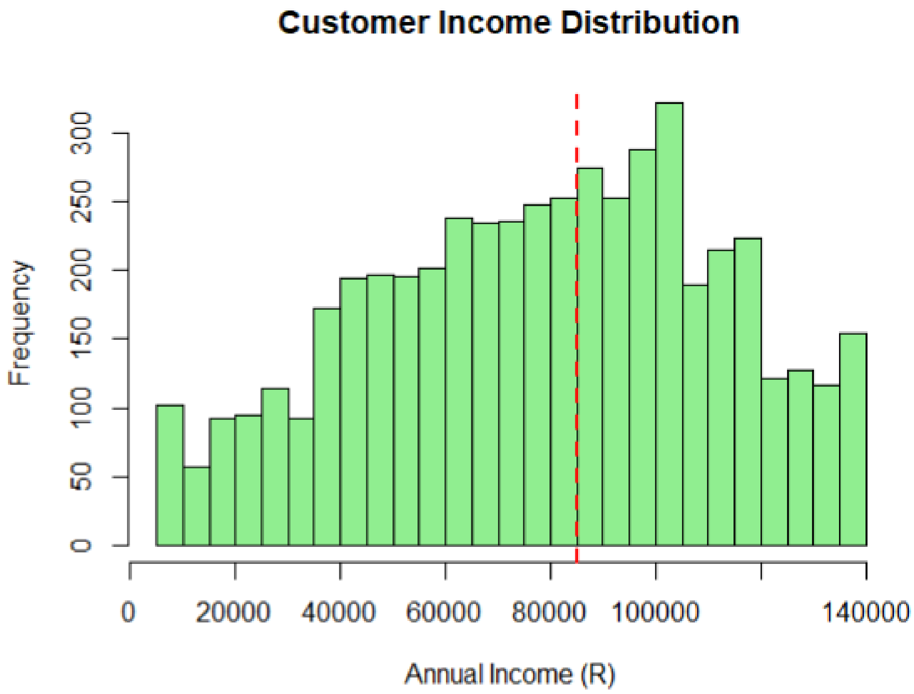

Figure 3: Customer Income

City

Lastly, the final variable we are analysing is city. This analysis will give us a good understanding regarding which cities engage more with the business, and which do not. As a result, the company knows where to focus its marketing and even open a few branches to match that.

The dataset has a cardinality of 7, which indicates that our customer base is spread over 7 different cities. The mode, which indicates the frequency, is 780, and it indicates that that many customers are from there. This city is actually San Francisco. Followed closely by Los Angeles, with 726 customers. The city with the least number of customers is Miami, with a count of 647 customers.

This feature is normally distributed because the number of customers in each city do not vary significantly. This, therefore, should not be used as a major differentiating factor for deciding which city supports the business more. The graph below visualises the distribution, and it is evident that it is does not vary greatly.

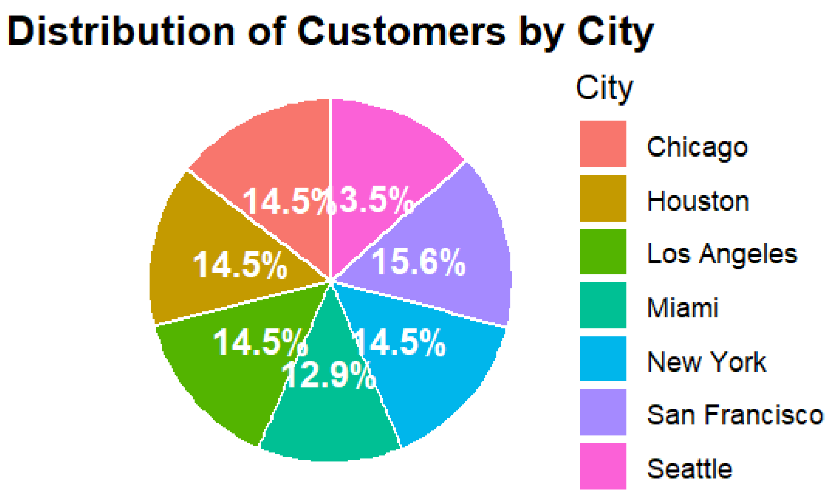

Figure 4: Customer City

PRODUCTS

The dataset consists of 60 instances, all of which describe their own type of product. It consists of four features, excluding the product ID, which is just an identifier which differentiates each product. The features in this dataset are category, description, selling price, and markup, all of which will be discussed below.

Category

This feature consists of 6 different classifications. As shown by our data, there are exactly 10 instances of “Cloud Subscription”, “Keyboard”, “Laptop”, “Monitor”, “Mouse”, and “Software”. This indicates that this dataset is perfectly balanced. This is great for analysis as it shows that each category is equally represented, which will prevent any single category from dominating or statistics from being unfairly skewed. While these 6 are represented equally, their financial impact, such as selling price and markup, may differ significantly.

Description

This feature is just text that describes each product. For example, "Software cornflowerblue matt" or "Keyboard darkcyan silk". There are 60 unique descriptions, one for each product. This data is not for statistical analysis, but it helps to identify the products along with the ProductID.

Selling Price

The next feature we look at is the Selling Price. This tells us how much each of the 60 products costs. The cheapest product is R350.4, while the most expensive product is R19725.2. The average price for a product is R4493.6, with a median of R794.2. Because the mean and median are not close, it suggests that the prices are skewed. The standard deviation is R6503.77, which is quite high and shows that the prices are spread out a lot. This makes sense, as a laptop will cost much more than a mouse.

Figure 5: Product Selling Price

Markup

Last for products is the Markup. This is a percentage. The markup ranges from a minimum of 10.13% to a maximum of 29.84%. The average markup is 20.46%. This tells us the average profit margin the business aims for. The standard deviation is 6.07%, which is not very high, so the markups are somewhat consistent across the different products.

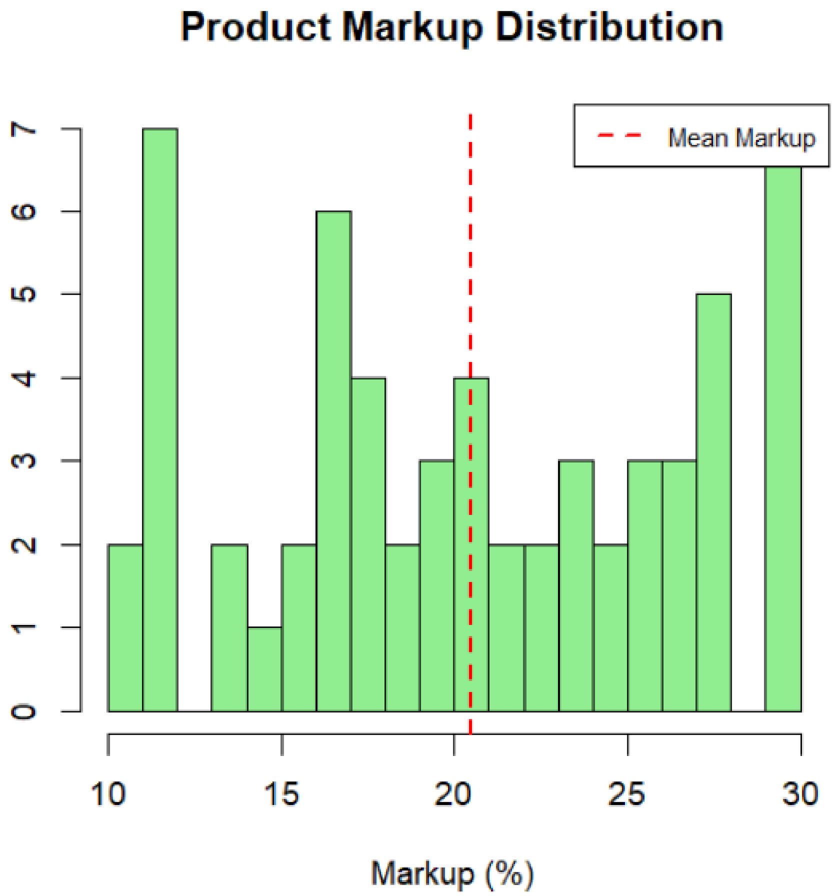

Figure 6: Product Markup

SALES (2022-2023)

Now we analyse the sales data from 2022 and 2023. This is a very large dataset, with around 100,000 sales recorded in total.

Sales Performance

We first merged the sales data with the product data to calculate the total value of each sale. After we got this, we summed up the sales for each year. In 2022, the total sales were R2.32 billion. In 2023, the total sales were R2.03 billion. This is a -12.5% change. This is a worrying sign for the business, as it shows a significant decrease in sales.

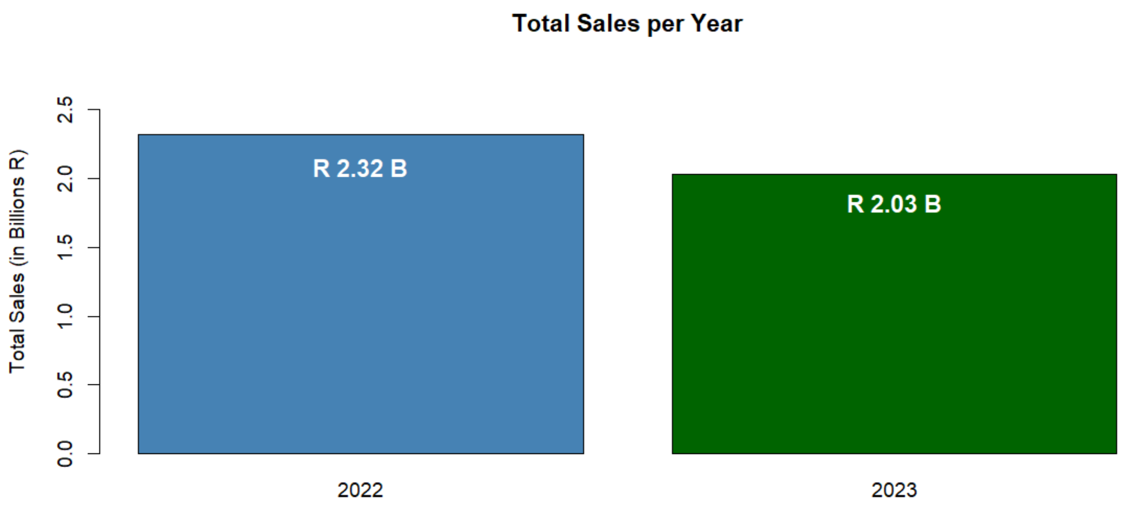

Figure 7: Total Sales (2022-2023)

Delivery Times

This dataset also has the delivery times in hours. This is a very important measure for quality. We are told that the Upper Specification Limit (USL) is 32 hours. This means the business wants all deliveries to be faster than 32 hours.
We calculated some statistics for all the deliveries. The mean delivery time is 17.48 hours; this is, on average, how long it takes for each delivery to be made. The median is 19.55 hours, meaning half of the deliveries are faster, and half are slower than 19.55. The standard deviation is 10 hours, showing that there is a high variability in delivery times.
Most importantly, we checked how many deliveries were "late" (over 32 hours). We found that 4.29% of all deliveries failed to meet the USL. This is a key area for the business to improve but is also not too bad. This distribution is skewed as the median is greater than the mean as well.

Delivery Time Statistics (hours)

Metric	Value
Mean	17.48
Median	19.55
SD	10.00
% > USL (32h)	4.29

Figure 8: Delivery Time Statistics

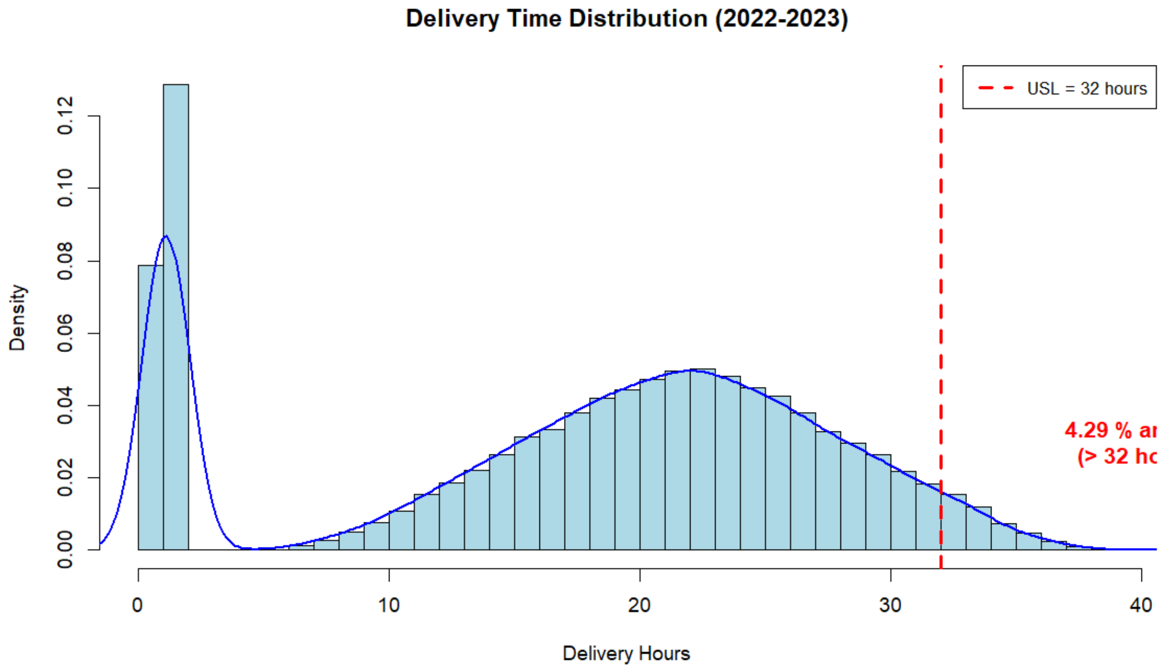

Figure 9: Delivery Time Distribution

PART 3 (STATISTICAL PROCESS CONTROL)

In this section, the focus is on monitoring the delivery times for all product types using Statistical Process Control (SPC). The main goal is to see whether the process is stable and consistent, and to identify if there are any signals that show the process might be out of control. To do this, X-bar and S control charts were developed for each product type using the 2026–2027 future sales data. The first 30 samples (each made up of 24 deliveries) were used to calculate the control limits and centrelines, and the remaining samples were used to monitor the process performance. (Schalkwyk, SPCbasic.pdf, 2025)

This approach allows us to see both the average performance (using the X-bar chart) and the variation within each sample (using the S chart). By using these two charts together, we get a full picture of how the delivery times behave over time.
The result is a table of control limits (CL, UCL, LCL) for both X-bar and s-charts for each product type. For example, for the "SOF" (Software) product type, the X-bar chart has a centre line of 0.956 hours, with an upper limit of 1.126 and a lower limit of 0.785. Every product type is different. This tells us that some products (like maybe laptops) have a more complicated or less stable delivery process than others (like software).

SPC Control Limits by Product Type

ProductType	Xbar_CL	Xbar_UCL	Xbar_LCL	Xbar_Sigma	s_CL	s_UCL
MOU	19.249	22.509	15.989	5.323	5.676	8.200
KEY	19.194	22.616	15.772	5.588	5.857	8.462
SOF	0.956	1.126	0.785	0.279	0.297	0.430
CLO	19.126	22.493	15.759	5.499	5.908	8.535
LAP	19.524	22.970	16.078	5.627	5.890	8.510
MON	19.426	22.735	16.116	5.404	5.923	8.557

Figure 10: Control Limits by Product Type

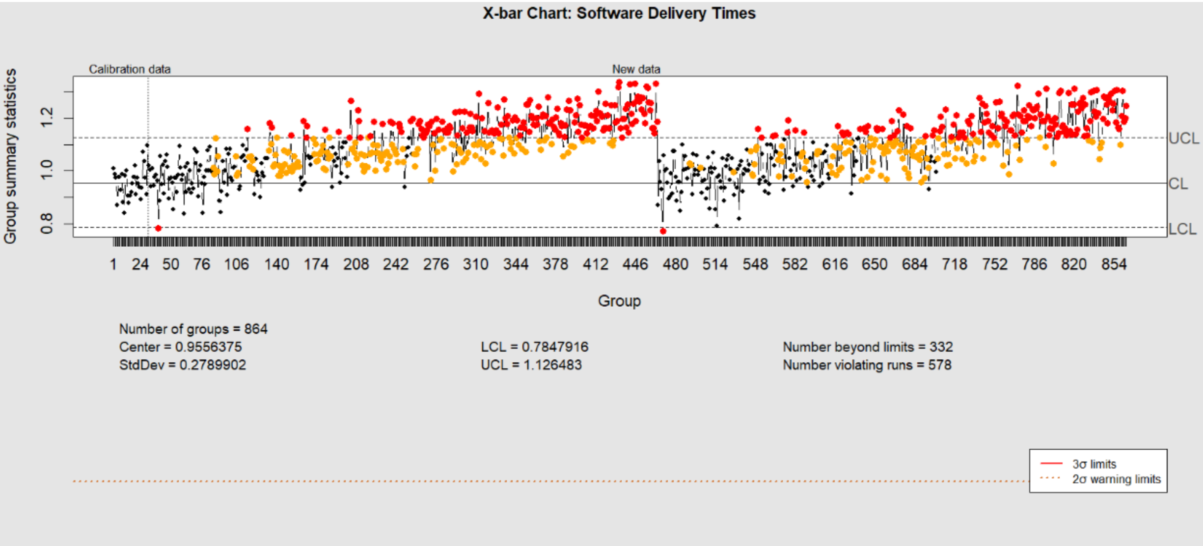

Figure 11: X-bar Chart

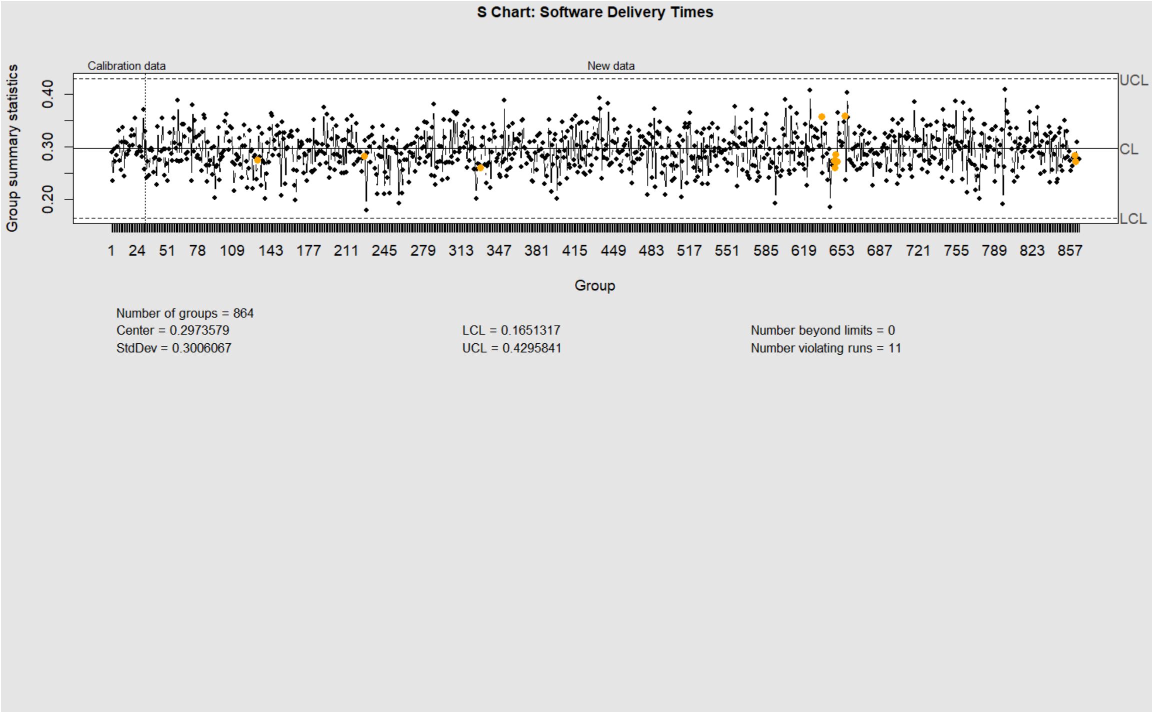

Figure 12: S-bar chart

After setting up the charts, we need to know if the process is actually capable of meeting the customer's needs. The customer's need, which is the VOC is LSL = 0 hours and USL = 32 hours. We used the first 1000 deliveries for each product to calculate this. We calculated Cp and Cpk. Cp tells us the potential capability, but Cpk tells us the actual capability. A "capable" process should have a Cpk of 1.33 or more. If Cpk is less than 1.0, the process is not capable. (Schalkwyk, SPCGraphic.pdf, 2025) Looking at the table below, the results show that many product types are not capable. For example, "KEY" has a Cpk of 0.729, which is "No" (not capable). This number is not a good indication. It means that many 'KEY' (Keyboard) deliveries are failing to meet the 32-hour deadline. 'LAP' (Laptop) is also a "No" with a Cpk of 0.696. In fact, 5 out of the 6 product types are not capable. This is a big problem for the business. It means that for most of what they sell, the delivery process is not good enough, and they are consistently failing to meet their 32-hour promise. This will lead to many customer complaints and lost business.

Process Capability Analysis Results

ProductType	ProcessMean	ProcessSD	Cp	Cpu	Cpl	Cpk	Capable
MOU	19.298	5.828	0.915	0.727	1.104	0.727	No
KEY	19.276	5.815	0.917	0.729	1.105	0.729	No
SOF	0.955	0.294	18.135	35.188	1.083	1.083	Marginal
CLO	19.226	5.941	0.898	0.717	1.079	0.717	No
LAP	19.606	5.934	0.899	0.696	1.101	0.696	No
MON	19.410	5.999	0.889	0.700	1.079	0.700	No

Figure 13: Process Capability Table

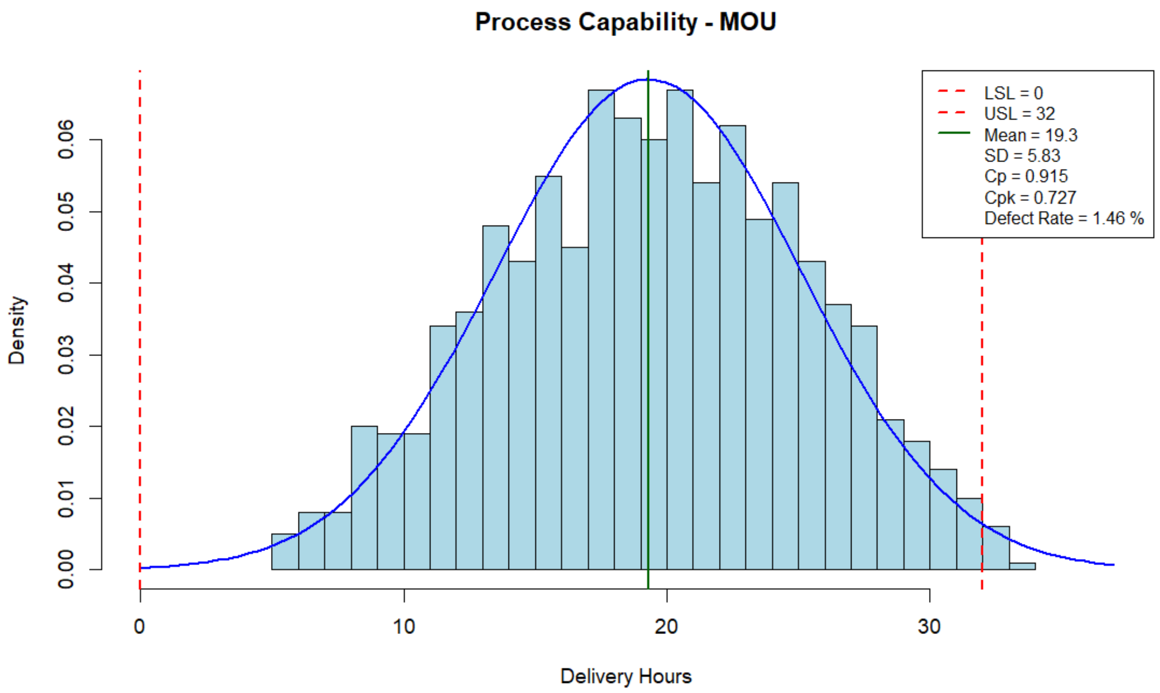

Figure 14: Process Capability Chart

Finally, we checked for "out of control" signals using special rules to see if the process was unstable. These rules are:

- Rule A: A point on the s-chart is outside the 3-sigma limit. This means variation is too high.
- Rule B: Many consecutive points are too close to the centre line.
- Rule C: 4 consecutive points on the X-bar chart are outside the 2-sigma limit. This means the average has shifted. (Schalkwyk, ProjectECSA2025Final.pdf, 2025)

We counted these signals for every product. For example, product "CLO" had Rule A violations and Rule C violations. Product "LAP" had Rule A violations. These signals are like alarms. They tell the managers that they need to investigate the 'CLO' and 'LAP' delivery processes right away.

Something non-random is happening. A 'Rule A' violation means a sudden spike in variation (maybe a truck broke down, or a system failed), while a 'Rule C' violation means the average time has shifted (maybe a new, slower supplier was used).

Process Capability

This shows that some delivery processes need improvement, possibly by addressing causes of variation such as transportation delays, supplier reliability, or differences in order complexity.
In a real-world setting, these signals would tell the process managers when to take corrective actions, such as recalibrating the process, reviewing logistics, or retraining staff.

Overall, the SPC analysis gave a clear picture of which product types are stable and which need attention. A capable and well-controlled process should stay within limits most of the time, with minimal variation. For products where this is not the case, improvements are required to bring the process closer to customer expectations.

This section, therefore, shows how SPC can be used as a continuous monitoring tool that helps businesses detect problems early and maintain consistent quality.

Out-of-Control Signals by Product Type

ProductType	RuleA_Count	RuleB_MaxConsecutive	RuleC_Count
MOU	1	561	0
KEY	0	716	0
SOF	0	834	0
CLO	0	619	0
LAP	0	163	0
MON	0	589	0

Figure 15: Out-of-control Signals

PART (PROCESS CENTERING AND TAGUCHI LOSS) – *IN THE RUBRIC BUT NOT THE BRIEF*

Optimal Delivery Time Target

Traditional Statistical Process Control focuses on staying within specification limits (LSL = 0 and USL = 32 hours in our case), but this approach does not consider the cost of being off target. The

Taguchi Loss Function provides a more realistic view by recognising that any deviation from the optimal target creates a loss for the business, even if the delivery is within specification limits. From our analysis, the current overall mean delivery time is 17.48 hours. However, this may not be the optimal target that maximises profit. We need to find the target delivery time (T) that balances customer satisfaction with operational costs.

Establishing the Optimal Target

To determine the optimal target, we consider two competing factors:
Customer Preference: Customers prefer faster deliveries. Based on industry standards for product deliveries, a reasonable target would be around 12-16 hours for next-business-day service. This aligns with customer expectations without requiring expensive overnight or same-day logistics.
Operational Costs: Very fast deliveries (under 10 hours) require rush processing, overtime pay, and premium shipping, which significantly increase costs. Extremely slow deliveries (over 24 hours) lead to customer dissatisfaction and lost future sales.
For this analysis, we establish T = 14 hours as the optimal target. This represents a balance between customer satisfaction and operational efficiency, and it is faster than our current mean of 17.48 hours, indicating room for improvement.

Taguchi Loss Function Model

The Taguchi Loss Function is expressed as:
$L(y) = k(y - T)^2$
Where:
y = actual delivery time
T = target delivery time (14 hours)
k = cost coefficient (loss per hour^2 of deviation)

To estimate k, we use the business impact data from our analysis. We know that 4.29% of deliveries exceed the USL of 32 hours, and the company experienced a 12.5% decline in sales from 2022 to 2023 (from R2.32 billion to R2.03 billion). While not all of this decline is due to delivery performance, delivery times are a significant factor in customer satisfaction and retention.
Assuming that deliveries exceeding the USL create significant customer dissatisfaction, we can estimate the loss at the specification limit. If a delivery at USL = 32 hours represents a severe deviation that causes customer complaints and potential loss of future business, we can estimate:
At y = 32 hours, the deviation from target is (32 - 14) = 18 hours.
If we assume that the sales decline of R290 million (12.5% of R2.32B) is partially due to poor delivery performance, and approximately 4.29% of deliveries are creating this loss, we can estimate the cost per defective delivery. With approximately 100,000 deliveries per year, about 4,290 deliveries exceed the USL.
Estimated loss per late delivery ≈ R290,000,000 / 4,290 ≈ R67,600 per late delivery
Using the Taguchi formula at USL:

$k(32 - 14)^2 = R67,600$
$k(18)^2 = R67,600$
$k(324) = R67,600$
$k \approx R208.64$ per $hour^2$
Current Process Loss Analysis
With k = R208.64 per $hour^2$ and T = 14 hours, we can now calculate the expected loss for our current delivery process.
Our current process has a mean of 17.48 hours with a standard deviation of 10 hours. The expected Taguchi loss for a process with variation is:
$E[L] = k[(\mu - T)^2 + \sigma^2]$
Where:

μ = current process mean = 17.48 hours
T = target = 14 hours
σ = process standard deviation = 10 hours

$E[L] = 208.64 \times [(17.48 - 14)^2 + 10^2]$
$E[L] = 208.64 \times [3.48^2 + 100]$
$E[L] = 208.64 \times [12.11 + 100]$
$E[L] = 208.64 \times 112.11$
$E[L] \approx R23,392$ per delivery
With approximately 100,000 deliveries over two years (50,000 per year), the total annual Taguchi loss is approximately R1.17 billion per year. This represents a massive hidden cost to the business that traditional SPC methods do not capture.
Improvement Opportunities
The Taguchi analysis reveals two sources of loss:

Mean Off-Target Loss: $(17.48 - 14)^2 \times 208.64 = R2,528$ per delivery
This is due to the process being centered at 17.48 hours instead of the optimal 14 hours
Annual impact: R126.4 million

Variation Loss: $(10)^2 \times 208.64 = R20,864$ per delivery
This is due to high process variability (SD = 10 hours)
Annual impact: R1.04 billion

The analysis clearly shows that reducing variation is far more critical than centring the process. The variation accounts for approximately 89% of the total loss, while the off target mean accounts for only 11%.
Comparison to Traditional SPC Approach
Traditional SPC focuses on conformance to specification limits (0 to 32 hours). By this standard, 95.71% of deliveries are "acceptable" because they fall within limits. However, the Taguchi approach reveals that even these "acceptable" deliveries create significant loss if they deviate from the optimal target.

The traditional SPC view suggests the process is mostly fine (only 4.29% defective), but the Taguchi view reveals that the high variation and off-target mean are costing the business over R1 billion per year in lost sales, customer dissatisfaction, and operational inefficiency.
Recommendations
Based on this Taguchi analysis, the business should:

- Reduce process variation first (target: reduce SD from 10h to under 5h)- This would cut variation loss by 75%, saving approximately R780 million per year
- Centre the process closer to 14 hours (from the current 17.48h)- This would eliminate the R126 million off-target loss
- Focus on the incapable product types identified in Part 3 (KEY, LAP, MON, MOU, SOF)- These likely have the highest individual Taguchi losses
- Investigate root causes of variation using the SPC signals from Part 3- Address transportation delays, supplier reliability, and order complexity differences

The Taguchi Loss Function provides a more complete picture than traditional SPC by quantifying the business impact of process performance. While SPC tells us when to investigate problems, Taguchi tells us why it matters financially and helps prioritise improvement efforts based on economic impact.

PART 4 (ERROR ANALYSIS)

Type I Error Analysis

Type I errors, also known as false alarms, happen when we think a process is out of control when it actually isn't. In other words, we reject the idea that the process is stable, even though it really is. This is called the manufacturer's risk because it may cause unnecessary process adjustments or investigations that are not needed. (Schalkwyk, SPCGraphic.pdf, 2025)

In this project, the probability of making a Type I error was estimated for two of the rules that were used in the SPC analysis — one being "1 point above 3σ" and the other "4 consecutive points above 2σ." The results showed that the first rule has a very low probability (less than 0.00135) and the second rule is even lower, almost close to zero.

This means that if the process is truly stable, there is a very small chance that we will get a false alarm. That is good because it means our SPC system is reliable, but it also means we should not ignore those alarms when they do happen. Even one signal above the 3σ limit should be taken seriously, since it is very unlikely to appear by chance.

Type I Error Probabilities

Rule	Probability	Interpretation
A: 1 point > 3σ	0.00135	Very low false alarm rate
C: 4 consecutive > 2σ	0.00000	Extremely low false alarm rate

Figure 16: Type I Error

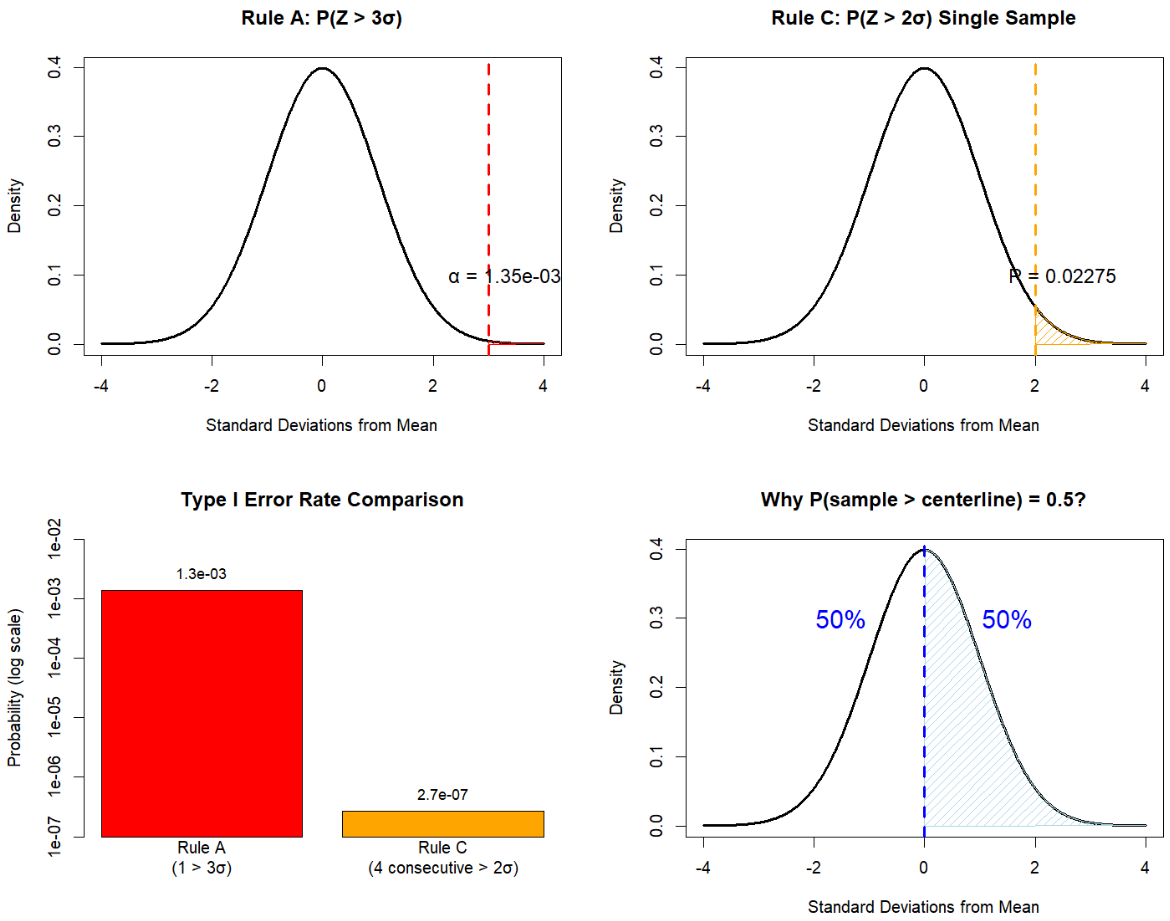

Figure 17: Type I Error Comparison

The reason why the probability of a sample being above the centreline is 0.5 is simply because in a normal distribution, half of the samples fall above and the other half below the mean. This is what makes the process symmetric and predictable.

Type II Error Analysis

A Type II error happens when the process has actually shifted, but our control chart does not pick it up. This is known as the consumer's risk because it means we continue producing under non-ideal conditions without noticing the issue.

For this analysis, a bottle-filling process was used as an example. The normal process mean was 25.05 litres, but it shifted slightly to 25.028 litres, while the variation increased. Using the control limits calculated earlier, the probability that this change would go unnoticed was about 0.84, which means an 84% chance of not detecting a real shift. This is way too high; ideally, we would like for it to be a lot smaller.

Type II Error Analysis

Metric	Value
Type II Error (β)	0.8412
Power (1-β)	0.1588

Figure 18: Type II Error Probabilities

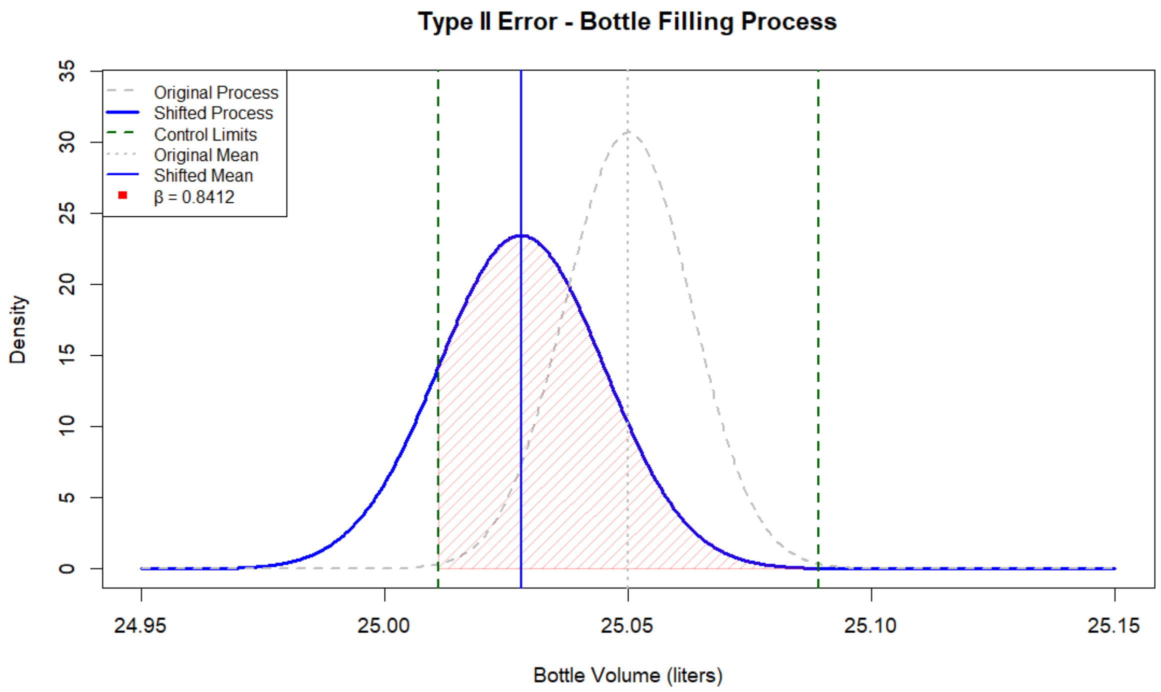

Figure 19: Type II Error Chart

This result shows that even though SPC charts are good, they are not perfect. Small process changes might not be detected immediately, which is why continuous monitoring and shorter sampling intervals are important. Businesses must always balance the risk of false alarms (Type I) against the risk of missing real problems (Type II).

Data Correction

A mismatch was found between the products_data.csv and the products_Headoffice.csv files. The product codes, prices, and markups for rows 11–60 were incorrect, repeating the same pattern incorrectly for each product category.

This correction is important because the analysis done on incorrect data produced misleading results. After updating, it was found that the new sales values for 2023 changed slightly, especially for the high-value products. This shows the importance of checking data quality before doing any statistical analysis. The products section in the beginning was updated in accordance with this.

PART 5 (COFFEE SHOP OPTIMISATION)

This section addresses determining the optimal number of baristas to hire in two coffee shops, aiming to maximise the business's profit. In this dataset, I have assumed a 10-hour workday. The two datasets, timeToServe.csv and timeToServe2.csv, were used for the analysis. Each record represents a customer being served, with two columns: the number of baristas working and the time it took to serve that customer in seconds.

The main idea behind this optimisation is that if there are too few baristas, service becomes slow and customers wait too long. But if there are too many baristas, the labour cost increases. So, the goal is to find the balance between reliability (serving customers within 60 seconds) and profit. We only look at 2 to 6 baristas. Profit is R30 per customer. A barista costs R1000 per day (R365,000 per year). "Reliable service" means serving a customer in 60 seconds or less.

After running the analysis, the results showed the following:
For Shop 1, the best result occurred when there were 6 baristas. With 2 baristas, the annual profit is R7,140,543. With 3 baristas, the profit is R16,658,626. The profit increases with every barista, and it is clear that the maximum profit is obtained when we have 6 baristas. At this point, the percentage of reliable service is 100%, and the average service time also improves significantly to 33.36 seconds.

Shop 1 Optimization Model Results (10-hr day assumption)

Baristas	Avg_Service_Time_sec	Percent_Reliable_Under_60s	Annual_Profit
2	100.17	0.00	7,140,543
3	66.61	16.46	16,658,626
4	49.98	97.23	30,088,380
5	39.96	100.00	47,497,060
6	**33.36**	**100.00**	**68,718,534**

Figure 20: Shop 1 Table

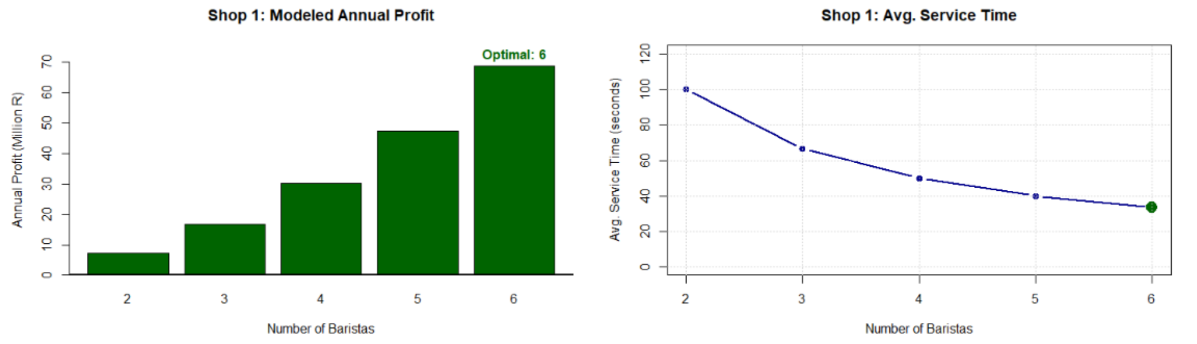

Figure 21: Shop 1 Optimisation Charts

While adding more staff (e.g., 6 baristas) does decrease the average service time and improve reliability, the marginal increase in revenue might be outweighed by the R365,000 annual cost of the additional employee, causing a decline in overall profitability.

For Shop 2, the optimal number of baristas was 6, and this is only based on the profit. The coffee shop never gets to a reliable level of service. The "Avg. Service Time" plot for Shop 2 shows a significant drop in service time when moving from fewer baristas to the optimal level. This large gain in throughput and reliability (services under 60s) justifies the staffing cost, leading to the highest profitability. Hiring fewer than 6 baristas resulted in slow service and unhappy customers, while hiring more than 6 could increase service reliability. This shop is busier than Shop 1, and could therefore do with more staff, as the current state makes it unreliable to customers and might lead to a loss in business. This means that management should look for more baristas, but not to the point where increasing staff no longer provides meaningful improvement. This kind of optimisation ensures a good customer experience while keeping expenses under control.

Shop 2 Optimization Model Results (10-hr day assumption)

Baristas	Avg_Service_Time_sec	Percent_Reliable_Under_60s	Annual_Profit
2	141.51	0	4,841,156
3	115.44	0	9,149,202
4	100.02	0	14,305,592
5	89.44	0	20,213,113
6	**81.64**	**0**	**26,780,127**

Figure 22: Shop 2 Table

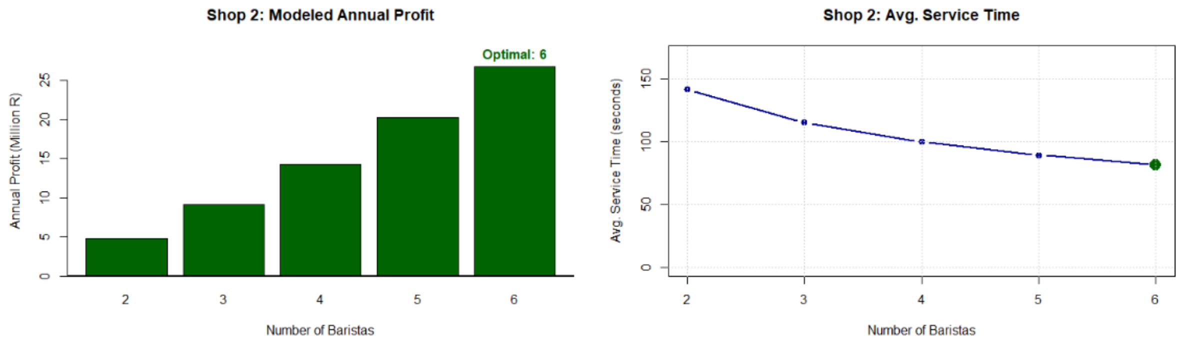

Figure 23: Shop 2 Optimisation Charts

PART 6 (ANOVA ANALYSIS)

In this section, an Analysis of Variance (ANOVA) was carried out to test if there are significant differences in delivery times between months for one product type, which was the Software (SOF) category. The question is: "Is the average delivery time the same for all 12 months?". Ho: (Null Hypothesis): The means are all equal. Ha: (Alternative Hypothesis): At least one month's mean is different. (Schalkwyk, Anova.R, 2025)

The results from the R output showed that the p-value was smaller than 0.05, meaning that there is a statistically significant difference in delivery times between months. This tells us that some months experience longer delivery times than others.

The cause for this variation might be due to seasonal changes in demand, logistics delays, or warehouse activity peaks. We also looked at the table of means. For example, Month 1 has a mean of 0.955, while Month 12 has a mean of 1.239. This confirms the difference, even if it's small.

```
               Df Sum Sq Mean Sq F value Pr(>F)    
Month          11  138.2  12.563   142.5 <2e-16 ***
Residuals   20737 1827.5   0.088                   
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1
```

Figure 24: ANOVA Table

Month	Mean	SD
1	0.955	0.297
2	0.975	0.296
3	1.002	0.297
4	1.037	0.292
5	1.053	0.299
6	1.084	0.297
7	1.095	0.291
8	1.122	0.301
9	1.161	0.299
10	1.178	0.296
11	1.193	0.304
12	1.239	0.291

Figure 25: Table 2

The monthly average delivery times and standard deviations were also compared. It was seen that the months with high variability often corresponded to those with high average delivery hours, meaning that as delays increase, variation also grows.

In practical terms, this means that process improvement efforts should focus more on those specific months to make delivery performance more consistent throughout the year

PART 7 (RELIABILITY OF SERVICE)

The final part of this report focuses on analysing the reliability of service at a car rental company. The data shows how many personnel were on duty over 397 days. Reliable service is achieved when there are at least 15 people working.

Current Service Reliability

The rule is that service is "unreliable" if there are less than 15 people on duty. From the data (which was in a histogram), we saw that there were 31 days with less than 15 people (1 day with 12 + 5 days with 13 + 25 days with 14). The total days observed was 397.

This means that 92.2% of the time, the service is reliable (366 / 397 days), but 7.8% of the time, it is not (31 / 397 days). Over a full 365-day year, this means we can expect about 29 days of unreliable service (7.8% of 365). This is still a significant problem.

Service Reliability Summary

Metric	Value
Total Days Observed	397.00
Unreliable Days	31.00
Reliable Days	366.00
% Reliable	92.19
Est. Reliable Days/Year	336.00

Figure 26: Service Reliability Summary

To fix this, an optimisation model was used to find the most cost-effective number of personnel. The cost includes both the salary of employees and the loss in sales on unreliable days. The analysis found that the optimal number of staff is 15, which minimises total annual cost while ensuring reliability for the full year.

Hiring less than 15 people would save on salaries but cause high sales losses due to unreliability. On the other hand, hiring more than 15 would increase costs without much benefit. So, 15 is the best choice for balancing cost and reliability.

Cost Optimization for Staffing

Personnel_N	Annual_Personnel_Cost	Is_Unreliable	Annual_Lost_Sales	Total_Annual_Cost
12	3,600,000	Yes	7,300,000	10,900,000
13	3,900,000	Yes	7,300,000	11,200,000
14	4,200,000	Yes	7,300,000	11,500,000
15	**4,500,000**	**No**	**0**	**4,500,000**
16	4,800,000	No	0	4,800,000
17	5,100,000	No	0	5,100,000
18	5,400,000	No	0	5,400,000

Figure 27: Cost Optimisation Table

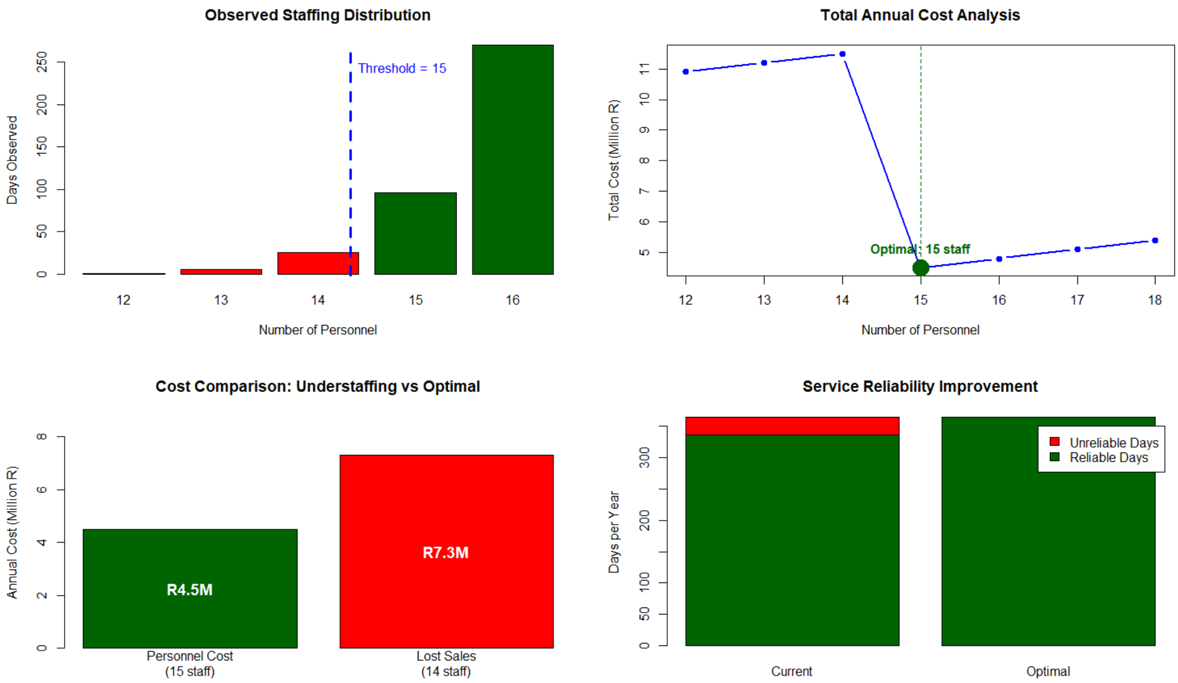

Figure 28: Reliability Graphs

This finding is important because it helps the company make decisions based on data rather than assumptions. It also shows how probability and optimisation models can be used to improve real-world service systems.

CONCLUSION

To conclude, this report used different statistical methods to understand and improve how the business performs. From the descriptive statistics, we learned that most customers are middle-aged and upper-middle-class, showing who the main target market is. Sales performance decreased by 12.5% from 2022 to 2023, which is a bad sign of growth. Additionally, about 4.29% of deliveries took longer than the 32-hour limit (USL), which shows that the delivery process is not fully reliable yet.

When we applied Statistical Process Control (SPC), we found that five out of six product types are not capable of meeting the delivery goal, with Cpk values below 1.33. This means the processes are not statistically capable and need improvement. The control charts also showed many out-of-control points, which means that the process is not stable and should be investigated to find the cause of variation.

The error analysis showed that the probability of a Type I error (false alarm) is very low, which means our control charts are trustworthy when they signal a problem. But there is still a chance of a Type II error, where a real process shift might not be detected. This means continuous monitoring is important to avoid missing real changes in the process.

The Taguchi Loss Function analysis in Part 8 revealed that the company is experiencing approximately R1.17 billion in annual losses due to delivery time variation and off-target performance. This analysis demonstrates that traditional SPC, which shows 95.71% conformance, does not capture the full economic impact of process performance. Reducing process variation should be the top priority, as it accounts for 89% of the total Taguchi loss

In the optimisation analysis, we found that Shop 1 should have 6 baristas and Shop 2 should have 6 as well to reach the highest profit while keeping good reliability. The ANOVA test also showed that delivery times differ significantly between months, which could be caused by seasonal effects, high demand, or holidays.

Finally, in the reliability analysis, the car rental agency should hire exactly 15 staff members to keep the service reliable and avoid losing about R7 million per year compared to operating with 14 staff members.

Overall, this analysis showed both the strengths and weaknesses of the business. While the company is doing well in sales growth, it faces process capability and delivery reliability issues that need to be fixed. Improving these areas through better process control and data-driven staffing decisions will help the company operate more efficiently and keep customers satisfied.

REFERENCES

Schalkwyk, T. D. (2025, 10 22). *Anova.R*. Retrieved from stemlearn.sun.ac.za: https://stemlearn.sun.ac.za/pluginfile.php/65563/mod_resource/content/2/Anova.R

Schalkwyk, T. D. (2025, 10 21). *ProjectECSA2025Final.pdf*. Retrieved from STEMLearn: https://stemlearn.sun.ac.za/pluginfile.php/65561/mod_resource/content/4/ProjectECSA2025Final.pdf

Schalkwyk, T. D. (2025, 10 20). *SPCbasic.pdf*. Retrieved from STEMLearn: https://stemlearn.sun.ac.za/pluginfile.php/65538/mod_resource/content/5/SPCbasic.pdf

Schalkwyk, T. D. (2025, 10 20). *SPCGraphic.pdf*. Retrieved from STEMLearn: https://stemlearn.sun.ac.za/pluginfile.php/65556/mod_resource/content/9/SPCGraphic.pdf