Megan Heppell

Quality Assurance ECSA Report

25928945

Stellenbosch
UNIVERSITY
IYUNIVESITHI
UNIVERSITEIT

Contents

Executive Summary

This report aims to present a comprehensive analysis of customer, product, head office, and sales data in order to gather business insights that can lead the organisation into an even more profitable future. The findings inform the reader on core customer profiles, income trends, age groups and genders whilst relating high-income customers with specific regions. Product analysis revealed some minor inconsistencies within those data sets, which were quickly rectified. Sales data shows clear seasonal trends, with stronger performance in 2022 compared to 2023.

Statistical Process Control (SPC) was implemented using X-bar and S charts to evaluate delivery-time stability. It was found that most processes are statistically in control, with software deliveries demonstrating exceptional consistency. Capability indices confirmed this; other product types require improvement to meet customer delivery expectations. Type I and Type II error analyses further confirmed the reliability of the SPC approach.

Optimisation processes were applied in two coffee shops and a car rental company to model operational performance and maximise profit. In both contexts, a model was developed to balance service reliability with cost efficiency. The near-optimal staffing levels were determined and discussed.

Finally, ANOVA testing confirmed process consistency between years, with predictable monthly variations indicating stable operations. Overall, the findings demonstrate how data analytics, SPC, and optimisation techniques can be used together to improve performance, reliability, and profitability. Successfully addressing the ECSA GA4 requirements by integrating data analysis, statistical reasoning, and engineering optimisation to support evidence-based decision-making and continuous improvement.

Introduction

In order to maintain a competitive advantage in today's dynamic business environment, it is essential to be able to extract and understand actionable insights from organisational data.

By thoroughly exploring, analysing, and investigating this data, one can quickly identify valuable insights into customer behaviour. These insights enable businesses to proactively adapt their strategies to better meet customer needs and achieve sustainable success. In this report, the analysis of multiple data sets will be covered, and analytical methods applied include descriptive statistics, control chart analysis, process capability indices, Type I and II error evaluation, and profit optimisation modelling. Additionally, ANOVA techniques are used to evaluate process variation over time. The aim is to extract actionable insights, identify improvement opportunities, and propose optimal operational solutions that enhance profitability and reliability across business units.

Part 1: Basic Data Analysis

Customer Data

Start by building a profile of the given customers. Figure 1 successfully identifies the distribution of the organisation's customers by their gender and income bracket. From the figure, it is clear that both male and female customer incomes follow a normal distribution, peaking as they approach the high-income bracket. The other gender has many fewer instances than male and female.

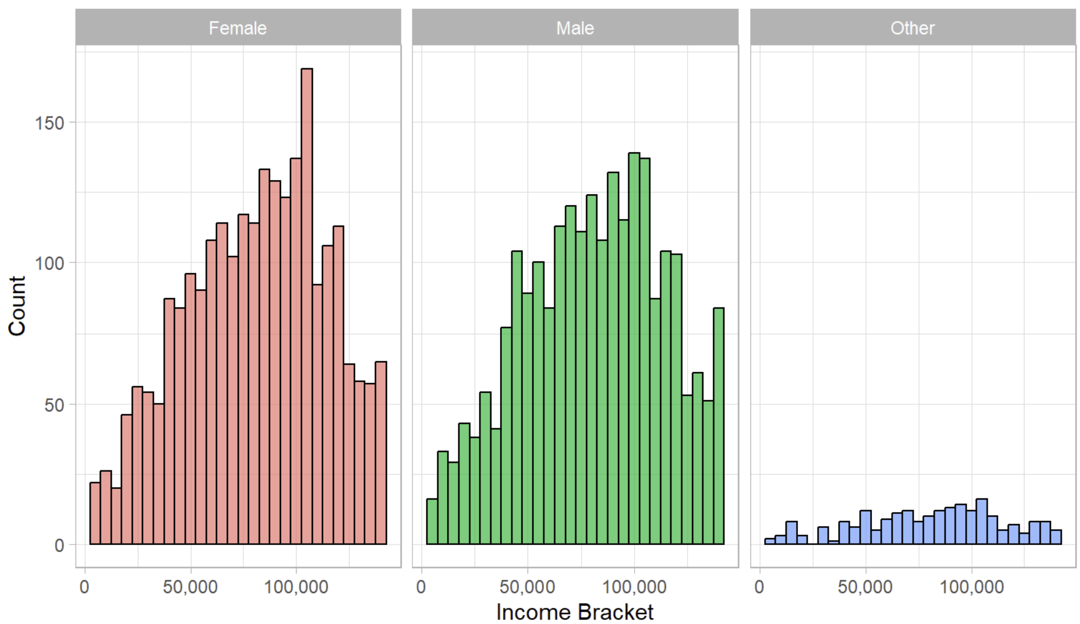


Figure 1: Income by Gender

Important to note that the gender profiles of the customers are somewhat well balanced, with males and females representing 47% and 48.64% of the total customers, therefore ensure that the products offered meet the needs of the respective gender groups.

By exploring the general trend of the customers' income categories by city, it is found that Miami has the highest average income per customer of $83 346.21. The range between the cities stretches from New York with the lowest avg income at $79 752.07 up to that of Miami. This is seen in the bar plot of

Figure 2, presenting the cities in descending order.

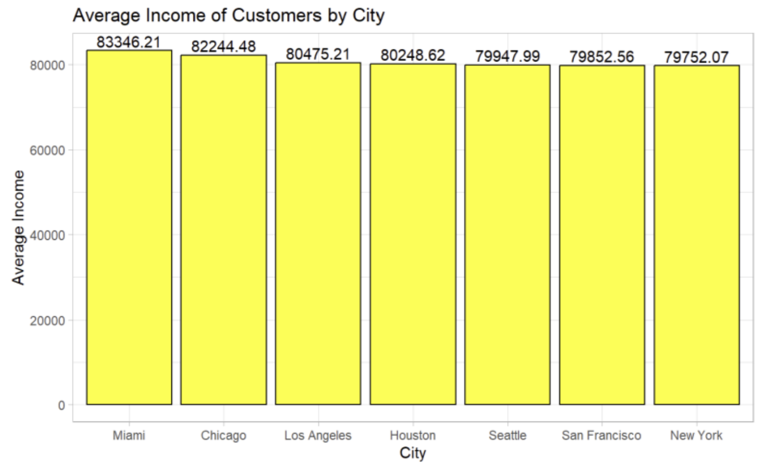


Figure 2: Avg Income by City

Though the averages of the cities are fairly similar, it would still be a good strategy to focus on promoting higher-priced products among the higher-income customers. Looking at Figure 3 Houston has the highest percentage of high-income customers, yet its average income per customer is ranked 4^{th}.

To determine which customers are classified as high-income customers, a threshold income value was chosen. It was calculated by obtaining the maximum income value from the customers' data and then multiplying it by 80%, resulting in a threshold income value of $ 112,000.

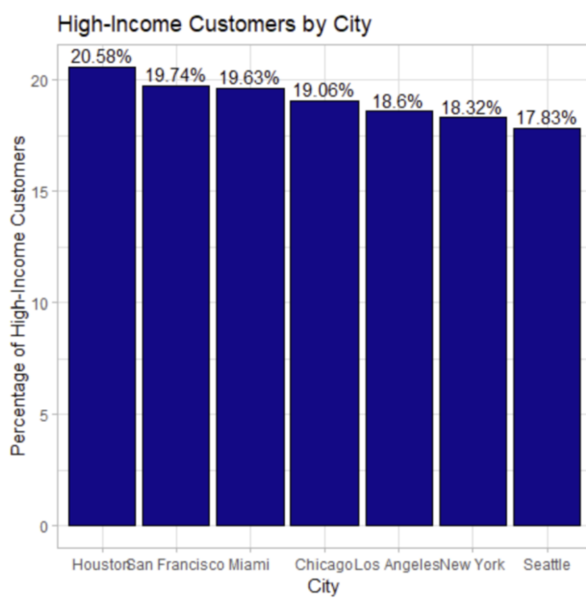


Figure 3: Percentage of High-Income customers

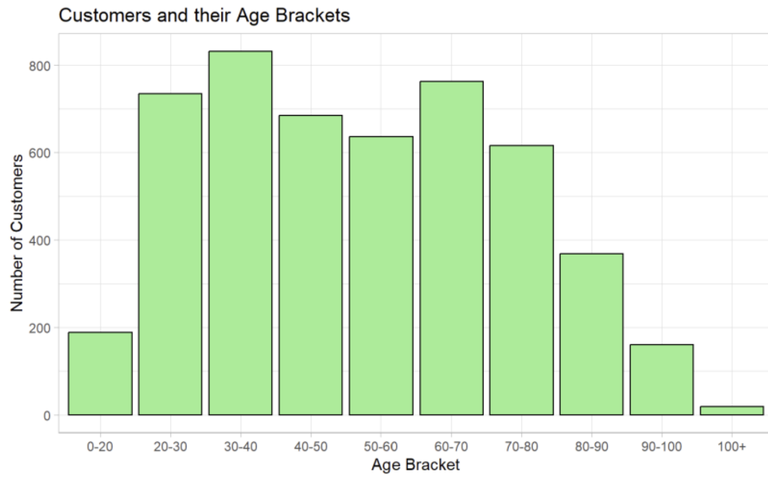


Figure 4: Customers by their Age Bracket

Figure 4 shows that the age bracket with the highest number of customers is between 30 and 40 years old, with approximately 830 customers. This is closely followed by the 20-30 age group and the 60-70 age group. The general distribution of the customers and their age brackets is bimodal, with a peak at 30-40 and again at 60-70. The core customer base consists primarily of individuals aged 30 to 70, with very few young adults under 20 or elderly individuals over 90 actively engaged.

The box plots in Figure 5 identifies that income tends to rise steadily from young adulthood, peaking in the 40-50 age bracket, where the median income is at its highest, around 100,000. After the age of 50, the median income begins to decline gradually, likely due to retirement.

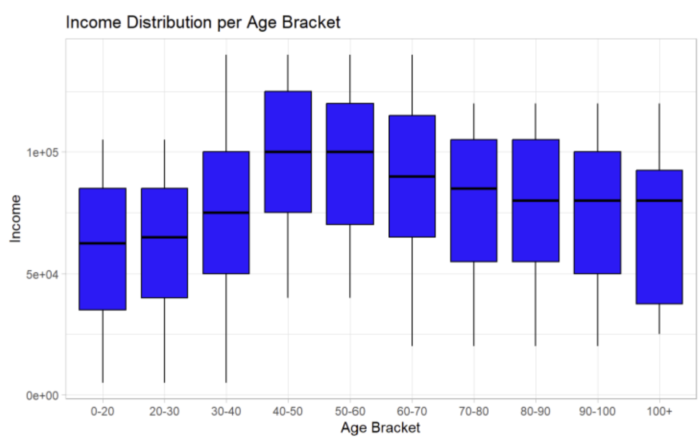


Figure 5: Income distribution per Age bracket

Together, these charts suggest that the 40–60 age range is the most valuable segment, as it combines both high customer numbers and strong spending power. Businesses should focus on this group while also working to attract younger customers for long-term growth and tailor products to meet the needs of older customers as their numbers increase.

While calculating the correlation between the customer age and the income, it was found that there is a very weak positive linear relationship with a correlation value of 0.1575221. Thus, it is safe to assume that one should not rely too heavily on age when making customer or marketing decisions.

Figure 6: Customer Age vs Income Correlation

Product Data

When analysing product data, start by grouping the products appropriately. By inspecting the first three letters of the productID it is possible to accurately identify the product type. This business streamlines six different products, each with slight variations within the product type. The summary of the products can be seen in Figure 7.

Category <chr>	Markup_Avg <dbl>	AvgPrice <dbl>	MinPrice <dbl>	MaxPrice <dbl>
Cloud Subscription	19.956	1019.062	728.26	1128.98
Keyboard	23.981	644.660	512.40	835.62
Laptop	18.430	18086.429	15851.74	19725.18
Monitor	23.868	6310.525	5346.14	6806.08
Mouse	20.495	394.698	350.45	454.04
Software	16.040	506.183	396.72	549.02

Figure 7: Product Summary Table

In order to visualise the summary table, the bar plots of the Avg markup per category are seen in *Figure 8* and the Avg selling price in each category in Figure 9. In examining these figures it is clear that although laptops sell for far higher than any other product type their average markup is not as strong as the others eg. Monitors. Taking both figures into consideration, a business strategy should be to research the profitability of the most expensive products and gather insight on how to raise the markup whilst still ensuring competitive prices. The markup of 23.87% and the selling price of $ 6,310.52 on the Monitor appear to be a good fit for this market research initiative.

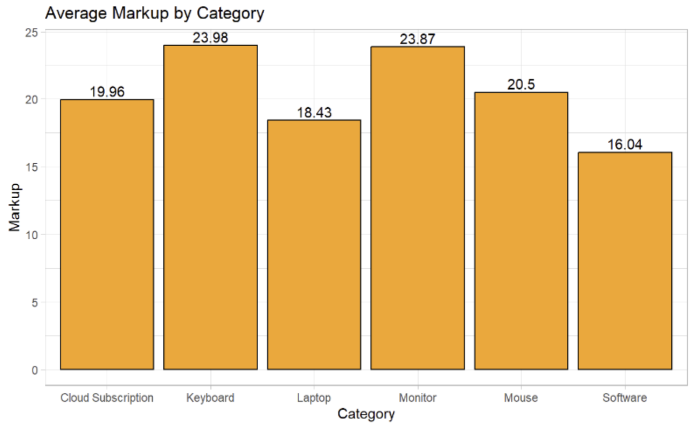


Figure 8: Avg Markup (Product Data)

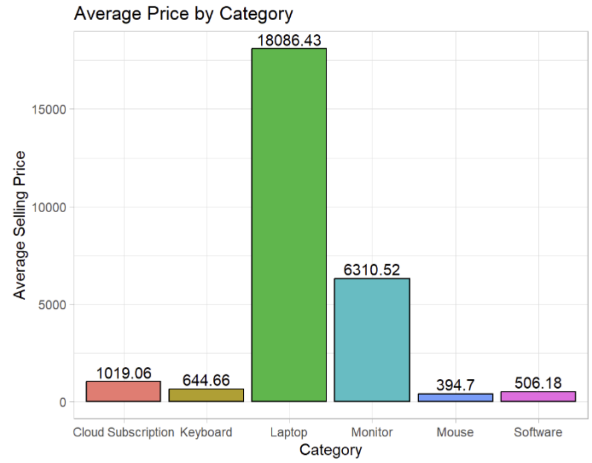


Figure 9: Avg Price (Product Data)

The assumption can be made that the product data analysed in this section came from a specific business branch within the organization; hence, there might be minor discrepancies between this section and the Head Office product data.

Head Office Data

The same approach to product categorisation was applied to the Head Office dataset as it was to the product dataset. Minor differences are noted between these data sets, see Figure 10 as an example the average markup of each category in this figure differs from those in Figure 8. This is lightly due to the location, demographic and customer profile of the branch investigated in the previous section. The Head Office data would be a more accurate representation of the entire organisation's standard, but further investigation is required.

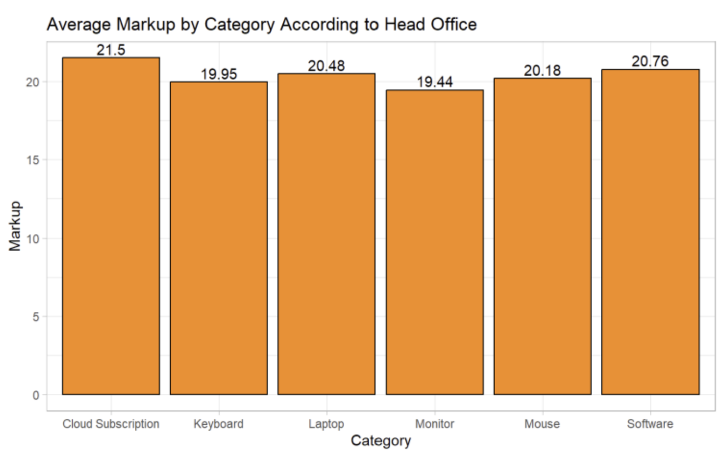


Figure 10: Average Markup (Head Office Data)

4.3 Data wrangling

After carefully examining the provided data, some questionable inconsistencies were identified. It was therefore decided to re-evaluate the data analysis of the Head Office data.This correction step ensures future analyses reflect true organisational standards and reduces data-quality risk in decision-making.

Data issues that were addressed included:

- Product types 11–60 were replaced with updated IDs from *products_data*.
- Selling price and markup values were repeated correctly
- Product Data columns were aligned, *Category* with *ProductID*

Figure 11 clearly illustrates the comparative data affected by quality issues.

	Category	Markup_Avg	AvgPrice
1	Cloud Subscription	21.50000	4386.710
2	Keyboard	19.95417	4380.485
3	Laptop	20.47517	4305.739
4	Monitor	19.44250	4456.745
5	Mouse	20.17967	4478.900
6	Software	20.76150	4457.193

	Category	Markup_Avg	AvgPrice
1	Cloud Subscription	20.553	3691.861
2	Keyboard	20.161	4638.172
3	Laptop	20.623	5217.545
4	Monitor	20.727	5014.170
5	Mouse	20.668	4585.465
6	Software	20.038	3814.344

Figure 11: Head Office Data summary before and after addressing data issues

After these corrections, the descriptive analyses produced the following observations. Average markups between branch-level and head-office datasets now align as seen in Figure 11: Head Office Data summary before and after addressing data issues. It is assumed that sales values per product category will most likely be adjusted upward by 1-2% due to the corrected pricing.

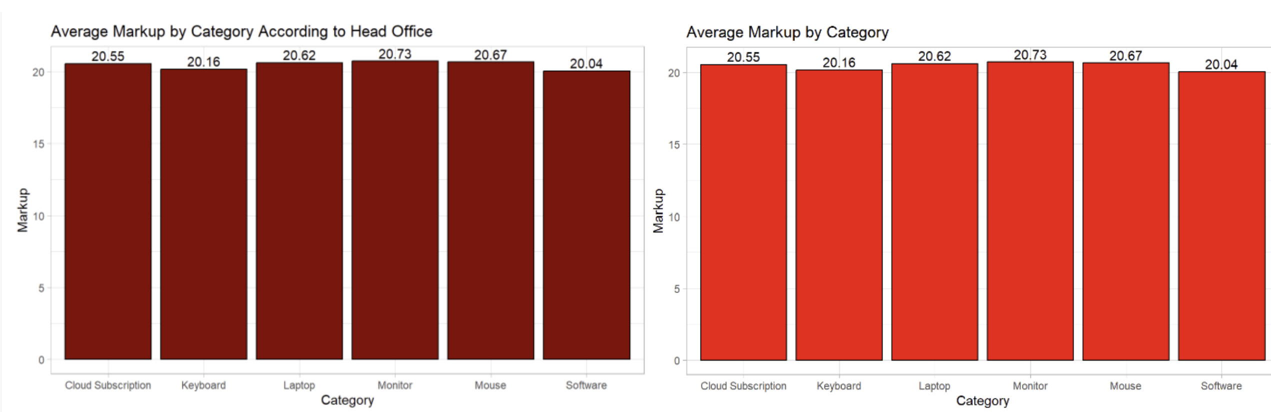


Figure 12: Correct Average Markup of Head Office and Products data

After correcting Figure 13, prices were normalised across categories, with laptops and monitors averaging 5 217.54 and 5 014.17, and lower-priced items adjusting upward to realistic levels. The revised data now accurately reflects market segmentation and resolves prior misclassifications.

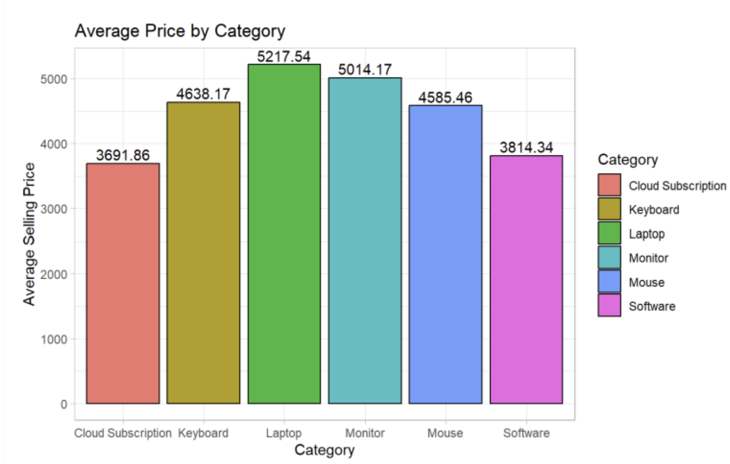


Figure 13: Realistic Avg Price per category

Sales data for 2022 and 2023

To delve into business development strategies, examining income streams is undoubtedly important. To make informed business decisions, insights into sales are a valuable guide and an effective problem identifier.

Figure 14 succesfully indicates the sales of each month over the years 2022 and 2023. This can be used to identify seasonal trends, which are evident in the drastic decrease in sales counts during the "holiday months" of December and January, likely because people are not working during these festive seasons. A notable remark is that the sales of 2022 performed a lot better across all months than those of 2023. Need further investigation.

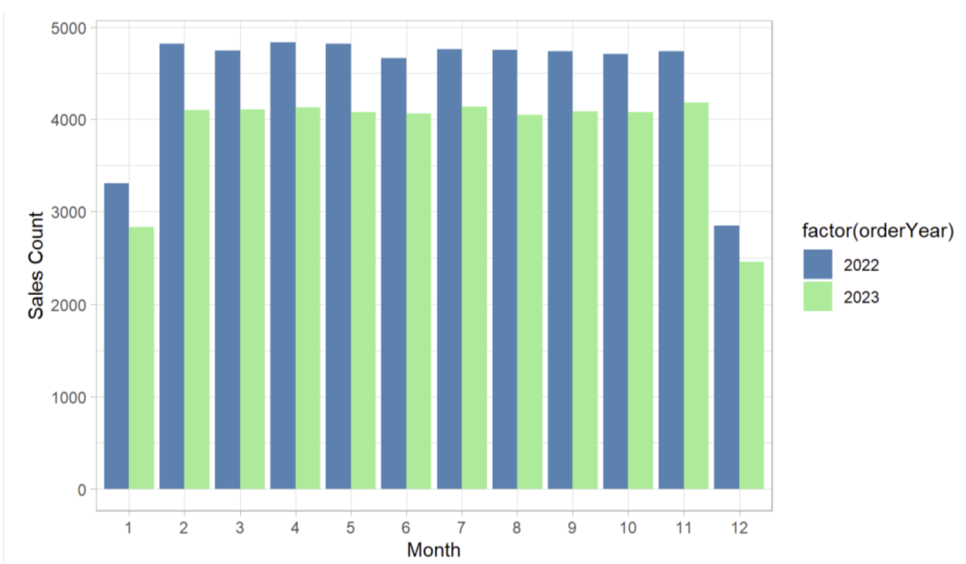


Figure 14: Monthly Sales per year

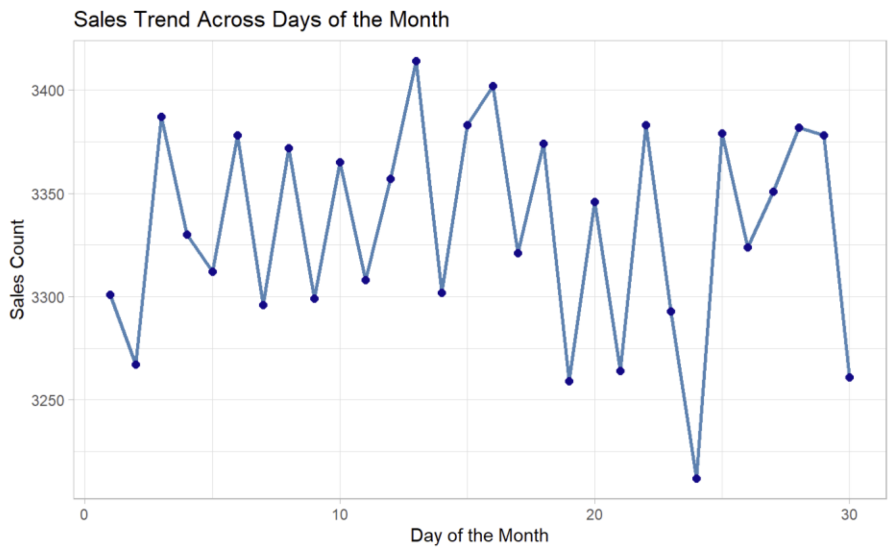


Figure 15: Sales trend over the month

Figure 15 does not lead one to believe that there is a specific trend related to the days of the month and the sales count. Notably, there is a slight dip between the 20th and 30th of the month.

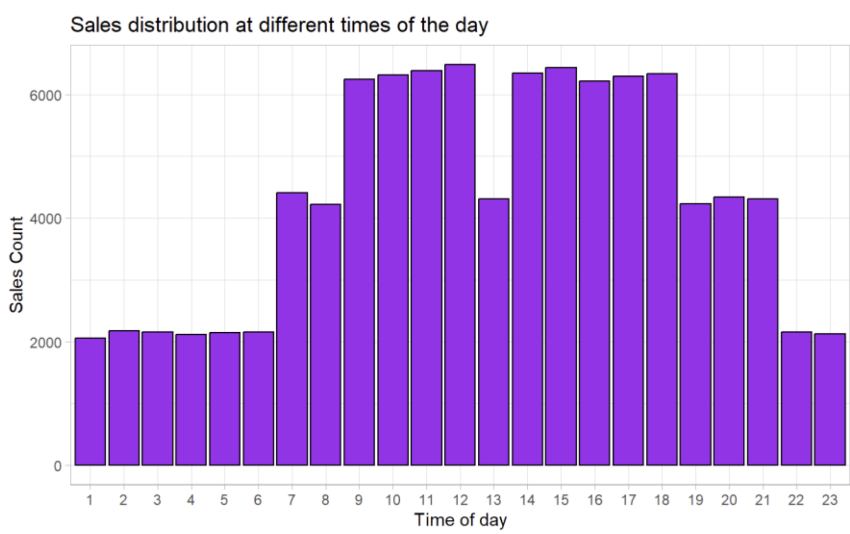


Figure 16: Sales trend at time of day

Figure 16 identifies what time of the day, on average, tends to be most popular among our customer bases. Seems like a binomial distribution with the peak times being 12:00 and then again at 15:00, essentially before and after lunch hour.

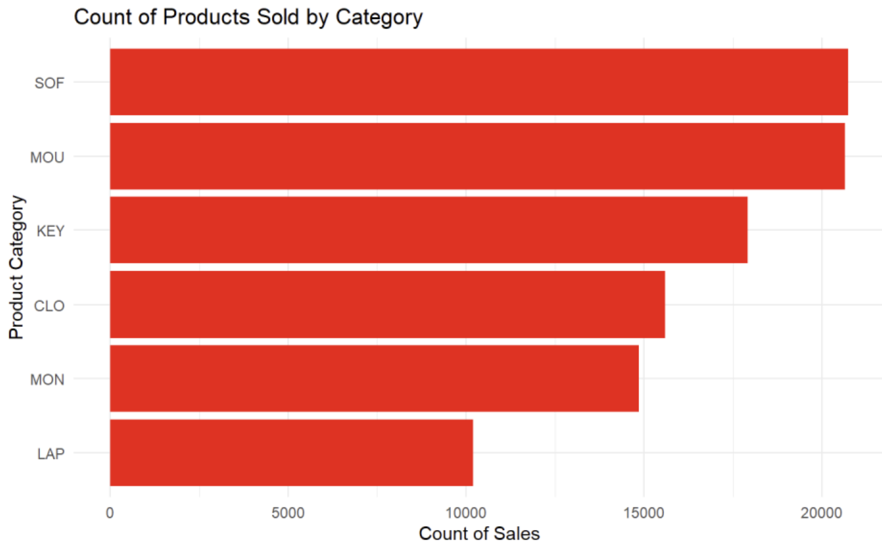


Figure 17: Sales distribution per category

Understanding the organisations that drive profitable products will help inform business decisions. From Figure 17 a conclusion is drawn that Software and Mouses are the most popular products. Realistically, this aligns with what is expected, as those are the less expensive products. Laptops perform the worst in terms of sales, which indicates that the organisation should consider running promotions, etc., since laptops are the most expensive. Increasing their sales, combined with a competitive markup, can help improve the business's profitability

Statistical Process Control (SPC)

This section applies Statistical Process Control (SPC) to monitor delivery times for each product type. X and S charts are used to assess process stability, while capability indices (Cp, Cpk) measure performance against delivery limits. Samples showing out-of-control conditions are identified and discussed.

3.1. Control charts

The dataset was sorted by year, month, day, and order time to ensure correct sequencing, which is similar to the analysis done in Figure 14 to Figure 16 . For each product type, the first 30 samples of 24 delivery times were used to calculate the initial process statistics. The centre lines and control limits for the X and S charts were determined by using the sample means and standard deviations.

In each chart, the red lines indicate the UCL and the LCL. The blue lines, positioned just one step inward from the red lines, represent the 2-sigma control limits for the chart and the purple lines depict the 1-sigma control limits for the chart.

The resulting charts provide a baseline for process performance. Most samples fall within the control limits, indicating that the process is generally stable with only normal variation. With further inspection, some key elements can be highlighted.

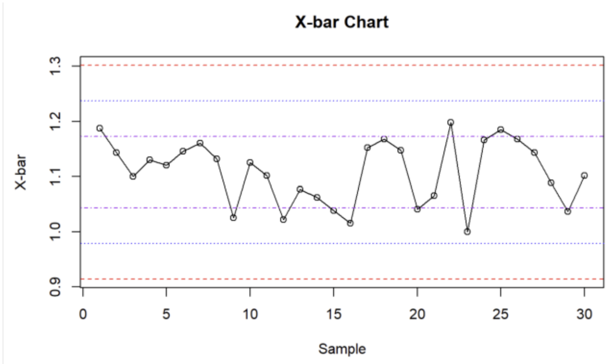


Figure 18: Software X-bar and S charts

SPC's charts of software, as seen in Figure 18, clearly deviate from the other product categories. The control limits are significantly lower, indicating that software delivery times are generally shorter and more consistent, resulting in less variability within the process. This is largely due to the nature of the software delivery process, as software is a digital asset; therefore, shipment time is not included, which drastically speeds up the process.

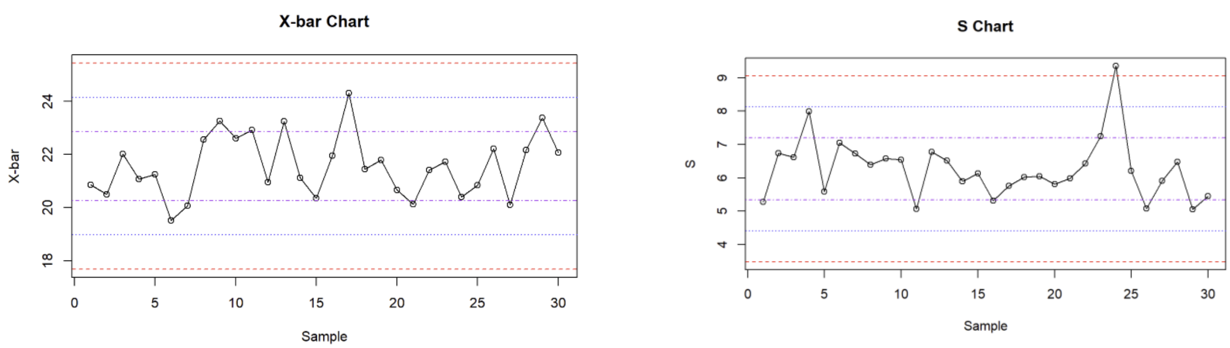


Figure 19: Mouse X-bar and S charts

Apart from the deviation in control limits of the Software charts, all other product categories have similar limits, ranging between 18-25 for the X-bar and 3-9 for S values.

This indicates that all category groups except software have a similar range of delivery times. It is assumed that this is because all other categories are physical products that must be shipped to the customer. Upon examining the X-bar and S charts for all categories, no dramatic spikes or changes in distribution were observed. The largest spike examined was the one seen in Figure 19, with a single peak reaching just past the 3-sigma upper control limit in the S chart around sample 24. Since it did not show a recurring pattern, it is likely the result of common-cause variation rather than a specific process issue. The absence of patterns or shifts indicates statistical control.

3.2 Monitoring the Process

Using the control limits established in Section 3.1, additional samples (from sample 31 onward) were analysed to monitor ongoing process stability. The updated X bar and S charts reflect the new data that has been added, simulating real-time process control. These observations also remained within the control limits, confirming consistent performance; however, a few samples showed minor variation that may warrant further observation in future monitoring

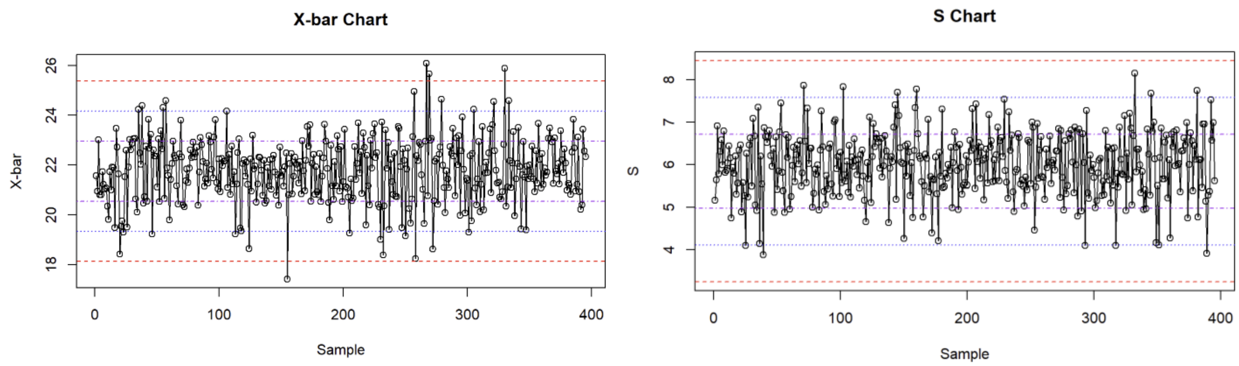
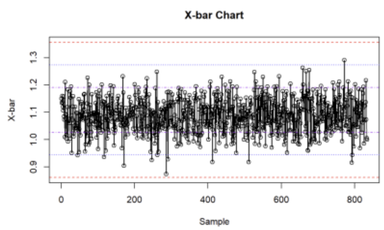
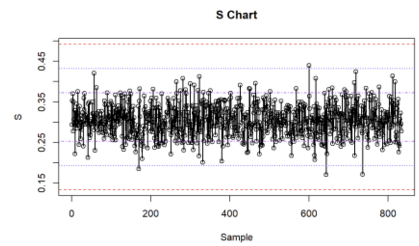
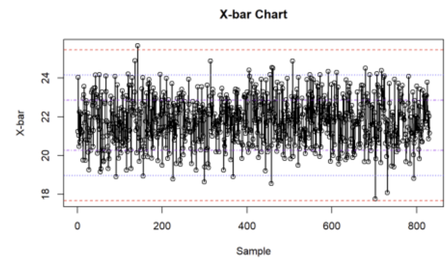
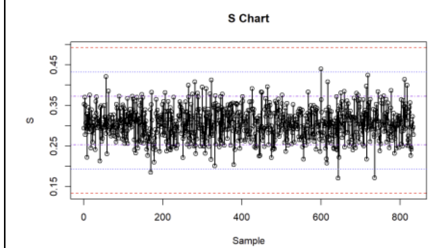
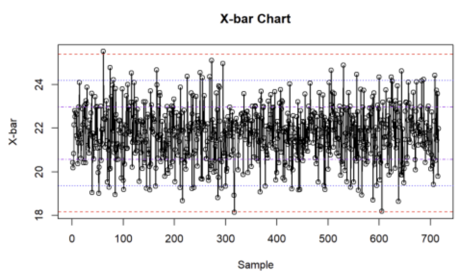
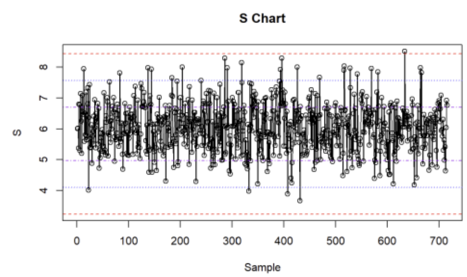
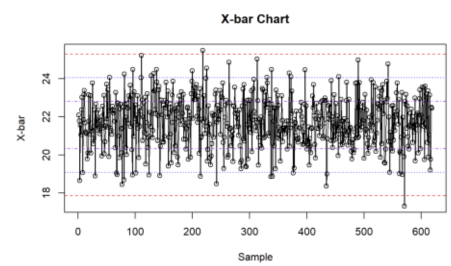
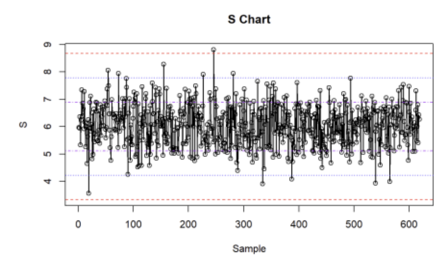
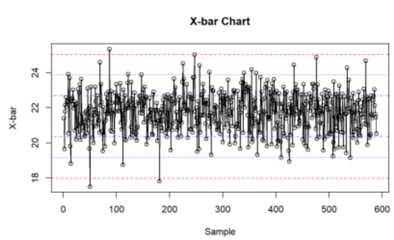
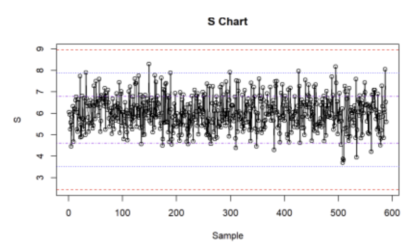
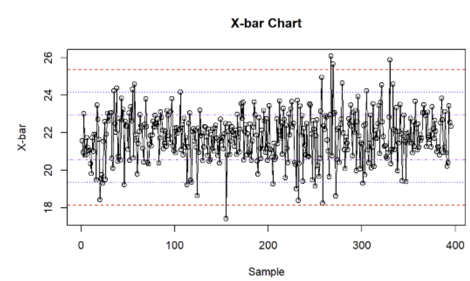
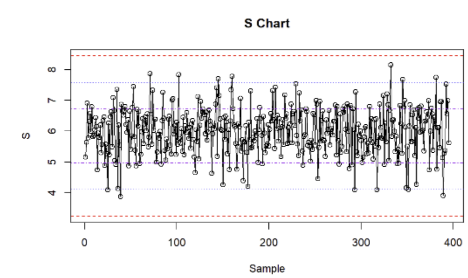


Figure 20: Laptop extended X-bar and S charts

Although no significant out-of-control findings were observed in monitoring the output of all product categories, data fluctuations increased across the board, but none exceeded the 3-sigma control limits. The X-bar and S chart of laptops that can be seen in Figure 20, shows the most notable variations, with a few instances reaching for the control limits yet still within bounds. Table 1 presents all outputs clearly, and thus one can see the general similarities among the different categories. Samples range between 400 and 800, while the control limits remain the same as defined in 3.1. Software again demonstrated the best control and lowest spread, while Monitors exhibited slightly higher S values.

Table 1: Summary of extended X-bar and S charts of all categories

<table>
<tr><td colspan="2">Software</td><td colspan="2">Mouse</td></tr>
<tr>
<td>
</td>
<td>
</td>
<td>
</td>
<td>
</td>
</tr>
<tr><td colspan="2">Keyboard</td><td colspan="2">Cloud Subscription</td></tr>
<tr>
<td>
</td>
<td>
</td>
<td>
</td>
<td>
</td>
</tr>
<tr><td colspan="2">Monitor</td><td colspan="2">Laptop</td></tr>
<tr>
<td>
</td>
<td>
</td>
<td>
</td>
<td>
</td>
</tr>
</table>

3.3 Process Capability Analysis

Process capability indices (Cp, Cpk, Cpl, and Cpu) were calculated for each product category using the first 1,000 delivery records. The lower specification limit (LSL) was set to 0 hours, and the upper specification limit (USL) to 32 hours. Cp (Capability Potential) indicates how well spread the process is, while the Cpk (Capability Performance) measures the ability based on the current centre and distribution. Categories with Cp and Cpk values above 1 generally indicate that the process is capable of meeting customer delivery requirements. If a product category has a Cpk value lower than 1, it suggests a need for process improvement or tighter control.

	Product	Cp	Cpu	Cpl	Cpk
1	Software	17.0331017	32.8824061	1.183797	1.1837973
2	Mouse	0.8515717	0.5539994	1.149144	0.5539994
3	Keyboard	0.8945111	0.5748758	1.214146	0.5748758
4	Cloud Subscription	0.8634681	0.5617403	1.165196	0.5617403
5	Monitor	0.9343685	0.6041043	1.264633	0.6041043
6	Laptop	0.9009559	0.5859822	1.215930	0.5859822

Figure 21: Process Capability of all categories

Capability Potential (Cp)

If a category were perfectly centered within specification limits, a Cp ≥ 1.33 indicates a capable process, in Figure 21 It is found that all categories have a Cp value < 1, with most of them ranging from 0.85 to 0.95, implying their delivery times vary too widely relative to the specification range.. Software proves to be the exception among these, with a Cp value of 17.033, which far exceeds the requirement. This indicates an extremely low variation and outstanding potential capability. The process is more than capable of meeting VOC if centered.

Capability Performance (Cpk)

Cpk accounts for both process variability and centering; thus, this characteristic is the determining factor in investigating which categories will be able to meet the VOC. The "rule" is that a Cpk ≥ 1.33 indicates that the process actually meets VOC consistently. In considering this with the output, none of the categories meet the requirement, as all categories have values ranging from 0.55 to 0.6, except for Software, with a Cpk value of 1.18. This means the Software has a marginal capability; the process is slightly off-center, despite having a very high Cp. With centering adjustments, it could meet VOC.

Software process is effectively capable, while physical product logistics require improvement, perhaps in shipping consistency or order batching.

3.4 Identification of Process Control Issues

Using the established control charts, samples were identified as showing potential process issues based on SPC rules. These included points outside the 3-sigma limits, consecutive points near the centre line indicating good control, and runs of four or more points beyond the 2-sigma limits. Most product types remained within control, though a few isolated samples exceeded the limits, suggesting temporary process disturbances that should be investigated.

A)

To determine which points lie outside the control limits, an OutofControlSamples function was created and applied. This function highlights points that fall outside the 3-sigma limits. Since there are only a few instances within each category, this indicates that the majority of the processes are statistically stable and operating within control limits. Figure 22 Indicates that there are few data points outside the upper control limits when examining the S-values.

	Product	Total_Out_Of_Control
1	Software	0
2	Mouse	0
3	Keyboard	1
4	Cloud Subscription	1
5	Monitor	1
6	Laptop	0

Figure 22: Data points outside of upper control limits for S values

B)

In an attempt to find the most consecutive samples of s between the -1 and +1 sigma-control limits for all product types. The s-charts indicate that process variability is largely under

	Product	Longest_Consecutive_In_Control
1	Software	18
2	Mouse	16
3	Keyboard	15
4	Cloud Subscription	20
5	Monitor	11
6	Laptop	17

good control.

Figure 23 shows that the Cloud Subscription process demonstrates the highest stability, whereas Monitor shows comparatively more variation and may benefit from review or adjustment.

	Product	Longest_Consecutive_In_Control
1	Software	18
2	Mouse	16
3	Keyboard	15
4	Cloud Subscription	20
5	Monitor	11
6	Laptop	17

Figure 23: Consecutive s-values between -1 and +1 sigma control limits

C)

Investigating the x-bar values, there were no 4 consecutive data points outside the 2-sigma control limits. This indicates that the process means are stable and there is no evidence of sustained shifts or trends in the central tendency of the processes. The absence of such patterns suggests that the system is operating under statistical control with respect to its mean values.

	Product	Total_4Consec_Outside_2Sigma
1	Software	0
2	Mouse	0
3	Keyboard	0
4	Cloud Subscription	0
5	Monitor	0
6	Laptop	0

Figure 24: X-bar values consecutively outside the 2-sigma control limits

In conclusion, all delivery processes are statistically in control, with only common cause variation.

Risk, Data correction and Optimising

4.1 Type I Error

To calculate the Type I error for different categories, it is assumed that the process is normally distributed around the mean and that the process is in control. Type I errors represent false alarm instances where the control chart signals a problem even though the process remains stable. The subgroup size is n = 24

Ho: the process is in statistical control (only random variation).

RULE A

For a single point on an *s*-chart, the probability of exceeding the upper 3sigma limit under Ho is obtained from the upper tail of the standard normal distribution. Rule A represents a typical Shewhart 3σ false alarm (Montgomery, D. C., 2024) .

A_type1_error = 0.001349898

The probability of making a type I error is thus 0.00135. This indicates that there is a 0.135% chance that the process is in-control when the control chart indicates that it is outside the outer control limits.

RULE B

When investigating rule B it is important to understand it conceptually. This rule signifies good control, not a violation. Unlike traditional control rules that flag points outside of control limits, this rule highlights unusually consistent behavior as a sign of good control. Thus under the Ho, the probability that a single point lies within 1sigma limits are 0.6827. Further, consecutive points all remaining within the 1sigma limit are determined by $P_{Longest\_run} = (0.6827)^{Longest\_run}$

To calculate the Type1 error of that is to determine the probability that at least one point in such a sequence lies outside 1sigma limits. Figure 25 indicates a list of high probabilities for type 1 errors.

These values show that such long stable runs are statistically rare, confirming that the process was under excellent control.

	Category	Longest_Run	Type1_Error_B
1	Software	18	0.998962
2	Mouse	16	0.997774
3	Keyboard	15	0.996739
4	Cloud Subscription	20	0.999516
5	Monitor	11	0.984987
6	Laptop	17	0.998480

Figure 25: B Type 1 errors per category

Rule B does not produce a Type I error in the traditional sense.
It identifies periods of high process stability rather than an out-of-control signal.

RULE C

The probability that four consecutive points all exceed the upper 2-sigma limit is extremely low.

C_type1_error = 2.678772e07

It provides strong protection against false alarms and only triggers under highly unlikely random variation. When such a pattern appears, it is almost certainly due to a real shift rather than chance. No such runs were detected in this dataset, confirming that the means of all processes remained stable over time.

4.2 Type II error

A type 2 error is the failure to detect that the mean of the process has shifted. To determine this, it was necessary to calculate the Z-scores for the given limits. The probability between the limits was thus obtained.

Type 2 Error = 0.8412

This suggests a 84.12% likelihood to assume that the bottling process is stable and centred around a mean of 25.05 litres with a standard deviation of 0.013 litres, when in fact, it is centred around a mean of 25.028 litres with a standard deviation of 0.017 litres. There is therefore only a 15.88% probability that the shift in standard deviation and mean will be detected. The process therefore has low sensitivity to small mean shifts, suggesting that tighter limits or larger sample sizes would improve detection ability.

Service Reliability, Optimisation of Profit

5.1 Coffee Shop 1- Barista Optimality

A model was developed to balance service efficiency and barista cost. Each customer contributes R 30 profit, while each barista costs R 1 000 per day.The coffee shops struggle to operate with less than 2 baristas, Figure 26 verifies this. Service time follows a decreasing exponential distribution as seen below. Notably the daily profit vs no of baristas has a positive linear relationship. This indicates that as baristas increase, the average service time decreases and profit increases.

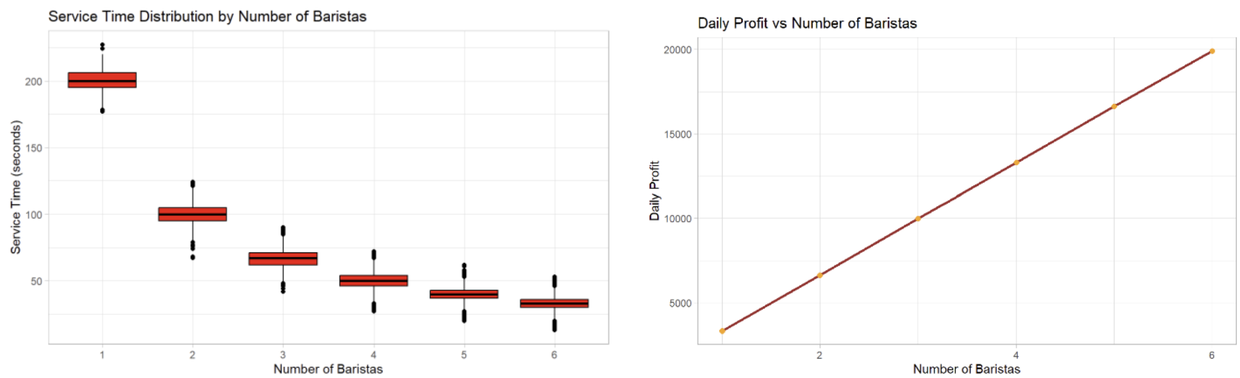


Figure 26: Coffee Shop 1 relationships

After developing a model to maximise profit and determine the near-optimal number of baristas, it is found that 6 baristas prove to be the most optimal. See Figure 27

	Baristas	mean_service_time	customers_served_per_day	daily_profit
1	1	200.15588	143.8879	3316.636
2	2	100.17098	287.5084	6625.253
3	3	66.61174	432.3562	9970.686
4	4	49.98038	576.2261	13286.784
5	5	39.96183	720.6876	16620.629
6	6	33.35565	863.4220	19902.661

Figure 27: Summary of Coffee Shop 1

Model output:

Optimal number of baristas: 6

Maximum profit (R): 19902,66

5.2 Service Reliability

When discussing reliability, it was assessed against the threshold that an acceptable service time is 50 seconds. With that said, there are 97878 instances where service time was equal to or less than 50 seconds when six Baristas were on duty. Six Baristas thus ensure service reliability of 99.98%. This is nearly optimal and ideal as one can confidently ensure customer satisfaction.

count	total	reliability
97878	97895	99.98263

Figure 28: Service reliability

5.1 Bonus - Coffee Shop 2 barista optimality

Coffee Shop 2 is evaluated under similar conditions. This shop exhibits longer baseline service times and higher customer-arrival variability, so the marginal improvement from extra baristas is stronger at first but levels off sooner.

In investigating the data set, seen in Figure 30, the observation is that the profit increases as the baristas increase up until a certain point and the start to decrease. Thus, the optimal number of baristas for Shop 2 is five, giving a maximum profit of R 4660.53 per day. Adding a sixth barista increases wage costs faster than throughput gains, reducing total profit, this is clearly displayed in *Figure 30*.

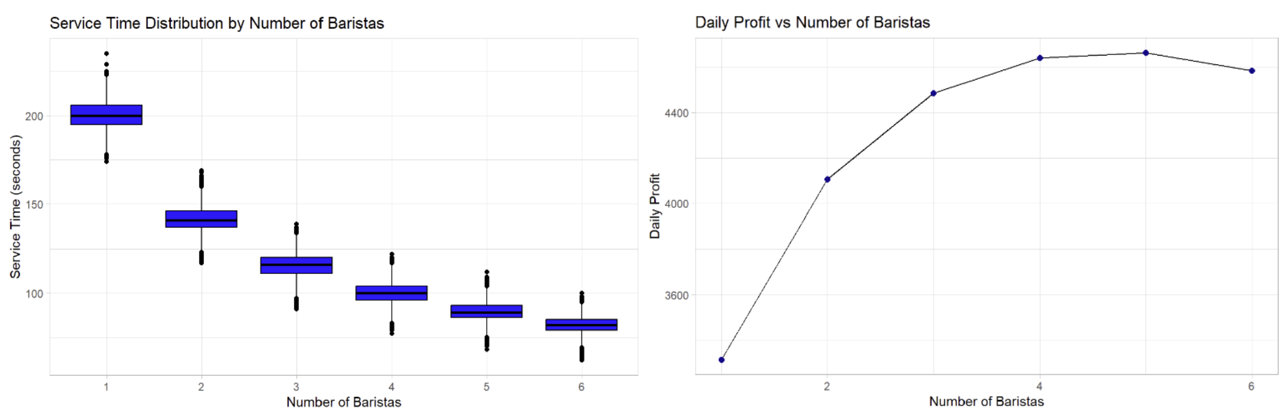


Figure 29: Coffee Shop 2 relationships

	Baristas2	mean_service_time2	customers_served_per_day2	daily_profit2
1	1	200.16894	143.8785	3316.354
2	2	141.51462	203.5125	4105.376
3	3	115.44091	249.4783	4484.348
4	4	100.01527	287.9560	4638.681
5	5	89.43597	322.0181	4660.543
6	6	81.64272	352.7565	4582.695

Figure 30: Coffee Shop 2 Summary

This optimisation directly demonstrates how engineering data analysis can inform operational decision-making to maximise profit while guaranteeing customer satisfaction.

DOE and ANOVA.

Two separate one-way ANOVAs were conducted to determine whether delivery and picking hours differed between years.

6.1 ANOVA R

In both cases, the p-values exceeded 0.05, indicating that there were no statistically significant year-on-year differences in mean delivery or picking times, displayed in Figure 31and Figure 32. These results suggest that process performance remained consistent throughout the evaluated years, implying stable operational control and reliability in both delivery and picking activities.

	Df	Sum Sq	Mean Sq	F value	Pr(>F)
orderYear	1	138.6173	138.61733	1.390657	0.238297
Residuals	99998	9967559.4499	99.67759	NA	NA

Figure 31: Anova Delivery Hours

	Df	Sum Sq	Mean Sq	F value	Pr(>F)
orderYear	1	2.200594e+01	22.00594	0.2039521	0.6515505
Residuals	99998	1.078954e+07	107.89758	NA	NA

Figure 32: Anova Picking Hours

These findings are further validated visually in Figure 33. From a service and reliability perspective, this consistency reflects a well-controlled process without significant external disruptions or process drift.

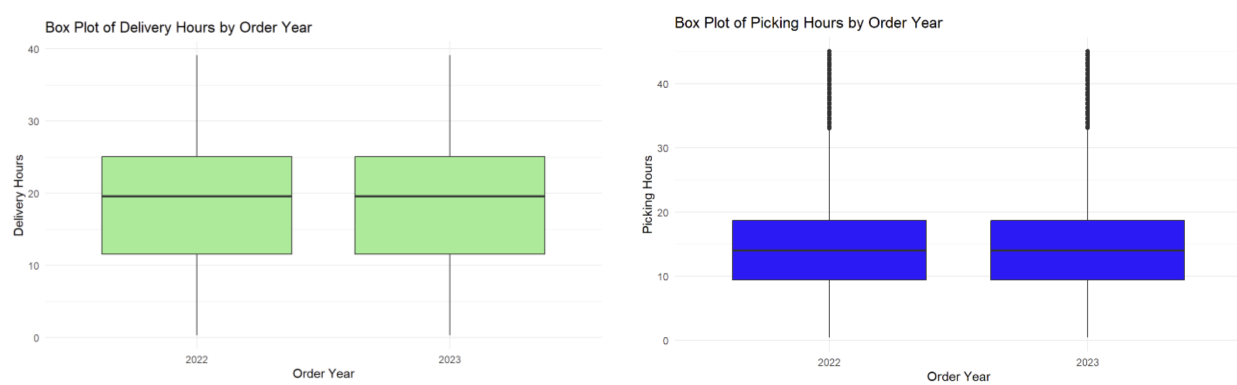


Figure 33: Illustration of statistical difference

A two-way ANOVA was conducted to evaluate the effects of year and month on delivery hours as demonstrated in *Figure 34*. The main effect of year was discussed above, no difference. The main effect of month, however, was highly significant. Since the p< 0.05 the months are statistically significantly different, demonstrating substantial monthly fluctuations in delivery times. The interaction between year and month was not significant, p>0.05 suggesting that the month-to-month pattern remained consistent across both years. No statistically significant difference. .

	Df <dbl>	Sum Sq <dbl>	Mean Sq <dbl>	F value <dbl>	Pr(>F) <dbl>
orderYear	1	138.6173	138.61733	1.414619	0.2342939
orderMonth	11	170246.9420	15476.99473	157.946043	0.0000000
orderYear:orderMonth	11	751.6210	68.32919	0.697314	0.7424801
Residuals	99976	9796560.8869	97.98913	NA	NA

Figure 34: Anova of Delivery Hours per order month

These findings indicate that the delivery process is reliable at the annual scale but experiences predictable seasonal variability, validated by *Figure 35*.

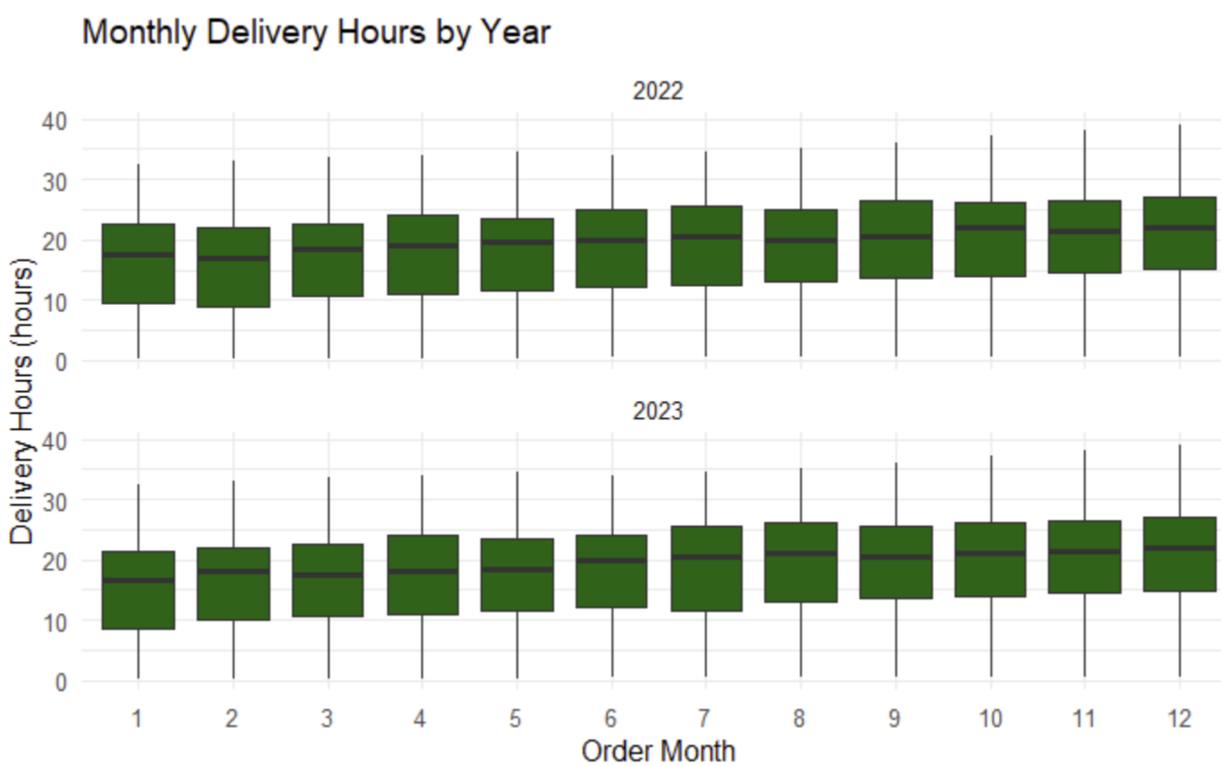


Figure 35: Monthly delivery hours difference by year

Operationally, this means that the company should anticipate higher delivery times in certain months and plan staffing or logistics accordingly to maintain service reliability. Note that the data set refers to the sales years 2022 and 2023, yet we are considering 2026 and 2027.

Reliability

This section investigates the reliability of service at a car-rental agency based on daily staffing levels. Reliability is defined as the likelihood that at least 15 staff members are available to serve customers on any given day. Whenever fewer than 15 employees are present, the agency faces operational issues that diminish customer satisfaction and result in a measurable reduction in sales.

7.1

The dataset of staffing levels and number of reliable service days was summarised in Figure 36

	workers	days	observed_p	predicted_p	reliable_days_per_year	expected_loss	annual_cost	profit
1	12	1	0.0025	0.0019	0.7	7286130	3600000	-10886130
2	13	5	0.0126	0.0106	3.9	7222620	3900000	-11122620
3	14	25	0.0630	0.0579	21.1	6877330	4200000	-11077330
4	15	96	0.2418	0.2606	95.1	5397620	4500000	-9897620
5	16	270	0.6801	0.6691	244.2	2415570	4800000	-7215570

Figure 36: Summary table of car rental company

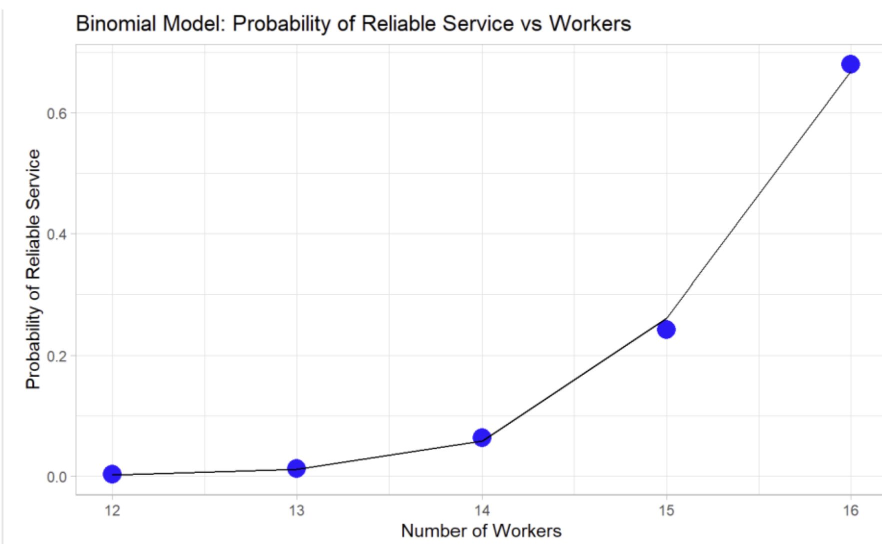


Figure 37: Probability of reliability

Figure 37 shows the relationship between workers and service reliability. The trend follows the expected binomial behaviour, where reliability increases sharply with staffing levels.In operational terms, this analysis suggests that the company requires at least 15 staff members on duty to achieve a moderately reliable service level, supporting the information given in the brief, while 16 or more staff would provide consistent reliability and minimise the risk of lost sales due to poor service.

The expected number of reliable days during the year is thus between 95-244 days, depending on the number of staff on duty. Assumed that 244 days is the best option

7.2

To model the probability of reliable service and optimise staffing levels, a binomial generalised linear model was fitted, see Figure 38. The numeric values are visible in the summary table presented in Figure 36. Both these figures verify that 16 workers is the most near optimal solution.

Figure 38: Profit to worker relationship

The trend reveals that profit increases sharply as the number of workers rises. At low staffing levels (12–14 workers), profits remain strongly negative, reflecting high daily losses due to frequent service failures. At 16 workers, the model predicts the highest expected profit, indicating that the additional salary expense is outweighed by the revenue retained from more consistent service.

Conclusion

The analyses in this report show how to effectively use statistical and optimisation techniques to assess and boost business performance. Customer and product evaluations provided valuable insights into profitability, emphasising the importance of ensuring accurate data sets. SPC and capability studies assure us that most processes are well within control, with software delivery demonstrating outstanding consistency and capability. Type 1 and Type 2 error analyses confirm that false alarms and missed detections stay within acceptable ranges.

Profit optimisation models for service operations illustrated how quantitative analysis can guide operational staffing decisions. ANOVA findings further supported consistent year-to-year operational reliability, with predictable seasonal variation.

Collectively, these findings confirm that data-driven quality assurance and management are powerful tools for continuous improvement.

References

Montgomery, D. C. (2024) ‘The 100th anniversary of the control chart’, *Journal of Quality Technology*, 56(1), pp. 2–4. doi: 10.1080/00224065.2023.2282926.

Dirkse-van Schalkwyk, T. (2025). QA344 Statistics. Stellenbosch: University Stellenbosch.