QA344 ECSA Final Report 2025:

By Louis de Beer (SU 26892901)

Department of Industrial Engineering

Stellenbosch University

Stellenbosch UNIVERSITY IYUNIVESITHI UNIVERSITEIT

SU Language Centre
IZiko leeLwimi lase-SU
US Taalsentrum
STELLENBOSCH UNIVERSITY

Voorletters en van **Initials and surname**	LJ De Beer
Studentenommer **Student number**	26892901
Modulenaam **Module name**	Quality Assurance 344
Titel van opdrag **Title of assignment**	ECSA Final Report 2025

Afrikaans: Plagiaatverklaring

- Ek het die Stellenbosch Universiteit Plagiaatbeleid gelees en verstaan, sowel as die definisies van plagiaat en selfplagiaat soos beskryf in die Beleid [Plagiaat: Die gebruik van ander se idees of materiaal sonder erkenning, of die hergebruik van 'n persoon se eie voorheen geëvalueerde of gepubliseerde materiaal sonder erkenning of 'n aanduiding daarvan (selfplagiaat of tekshergebruik)].
- Ek verstaan ook dat direkte vertalings plagiaat is.
- Dienooreenkomstig is alle aanhalings en bydraes vanuit enige bron (ingesluit die internet) volledig verwys (erken). Ek verstaan dat die woordelikse aanhaal van teks sonder aanhalingstekens (selfs al word die bron volledig erken) plagiaat is.
- Ek verklaar dat die werk in hierdie skryfstuk vervat my eie oorspronklike werk is en dat ek dit nie vantevore (in die geheel of gedeeltelik) ingehandig het vir bepunting in hierdie module/werkstuk of 'n ander module/werkstuk nie.

English: Plagiarism declaration

- I have read and understand the Stellenbosch University Policy on Plagiarism and the definitions of plagiarism and self-plagiarism contained in the Policy [Plagiarism: The use of the ideas or material of others without acknowledgement, or the re-use of one's own previously evaluated or published material without acknowledgement or indication thereof (self- plagiarism or text-recycling)].
- I also understand that direct translations are plagiarism.
- Accordingly, all quotations and contributions from any source whatsoever (including the internet) have been cited fully. I understand that the reproduction of text without quotation marks (even when the source is cited) is plagiarism.
- I declare that the work contained in this assignment is my own work and that I have not previously (in its entirety or in part) submitted it for grading in this module/assignment or another module/assignment.

22/10/2025

--- ----------------------

Handtekening / Signature Datum / Date

Table of Contents

Table of Figures

Table of Tables

1. Introduction

In an increasingly globalised market of cutthroat competition and razor-thin margins, educated decisions have not only become critical to daily survival, but remain one of the few avenues left for business to truly distinguish themselves. However, the quality of a decision is often determined by the bounds of information circumscribing the context in which that decision is made – a limitation that has since given way to the practice of data-driven analytics as a guidepost of corporate governance. It is against this backdrop of demand for greater statistical insights as a competitive advantage, that this report on various small business enterprises transpires. The purpose of this report is to review sets of provided historical data and then to process its contents through established methods of statistical analysis, with the aim of extracting useful insights into trends and relationships that can inform future decision making. Recorded details pertaining to sales, products, and customers will be investigated - with special attention paid to market leaders, outliers, statistical process control, process optimisation and recommended areas of improvement – in the hope of undergirding executive management functions.

2. Descriptive Statistical Analysis

The bulk of this report will focus on scrutinising the provided information on Company X's product details, customer characteristics, and historical sales data – with the express purpose of identifying consequential trends of interest to corporate management. As will be seen below, the investigative scope focuses on demographic and profit-related indicators, to generate actionable insights that are of use to the company's marketing and sales operations.

2.1 Customer Overview

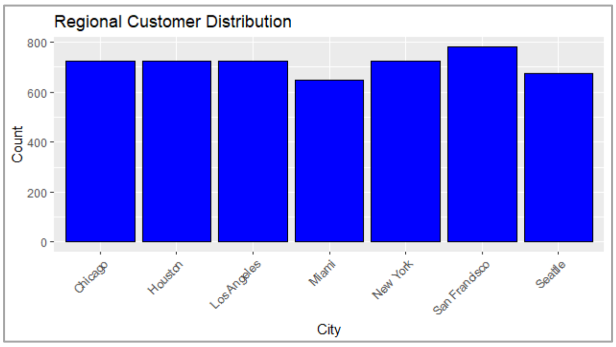


Figure 1: Regional Customer Distribution

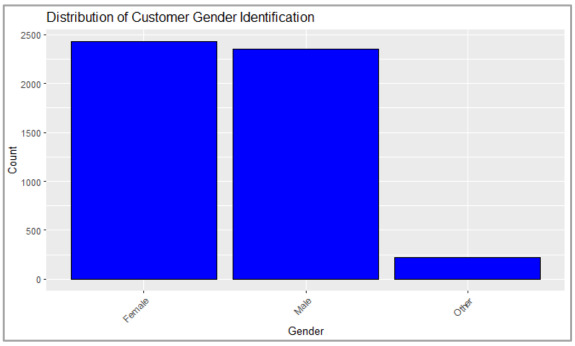


Figure 2: Distribution of Customer Gender Identification

Company X's typical customer profile can be approximated as a **51.6**-year-old **female**, living in **San Francisco** with a median income of **$85 000**. However, these traits should be correctly interpreted when formulating corporate strategy. As seen above, the total number of female customers only slightly outweighs the males, which is relationally consistent with US demographic trends (Statista, 2024) and therefore not suggestive of any particularly strong gender connotation with our product. Thus, it would be unadvisable to attempt a strongly gendered marketing strategy of any sort, since the marginal gains could be greatly outweighed by large losses in competing categories. Similar caution should be exercised when considering a regional strategy. As shown above, the top city (San Francisco) only slightly outranks the next four

(New York, Los Angeles, Houston, Chicago) - which are all very evenly matched - thereby demonstrating an even geographical distribution in our customer base. However, an even spread

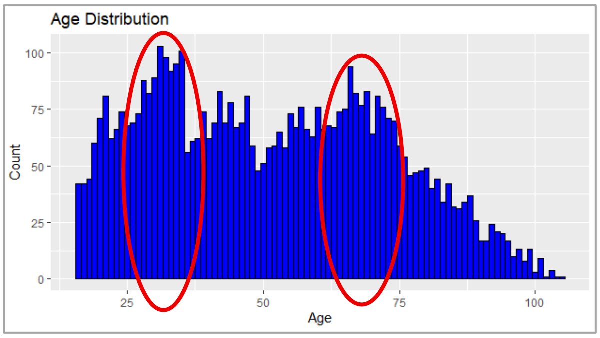


Figure 3: Customer Age Distribution

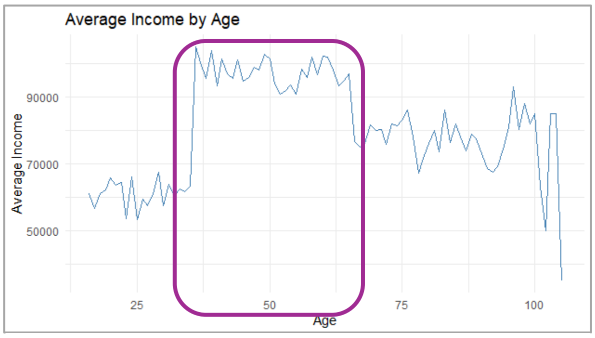


Figure 4: Customer Average Income by Age

does not mean that we are maximising our market reach. By consulting the following diagrams, we can determine the following underutilised market opportunities:

Upon evaluation of customer age distribution, it becomes clear that the company's clients are predominantly grouped into two categories, namely between ages 25-35 and 55-75 (indicated in red). However, when we consider the average income of these customers, we see that these groups are not consistent with the age bracket that boasts the highest disposable income. In fact, the group between these two clusters which appears to hold the highest average income (aged 43-56 and indicated in purple). This suggests that Company X is critically underserving the age demographic with the highest levels of disposable income. This is a costly mistake, since this vital portion of the market presides over the greatest share of excess money and should therefore be targeted first. After filtering for the city with the greatest number of individuals in this target range, it was determined that Chicago is home to the most individuals between the ages of 43 – 56 and therefore holds the most customers with average incomes in the top range. Therefore, it is strongly advised that promotional efforts aimed at middle-aged Chicagoans are increased to leverage this underutilised segment of yet unrealised revenue streams.

2.2 Company Profitability

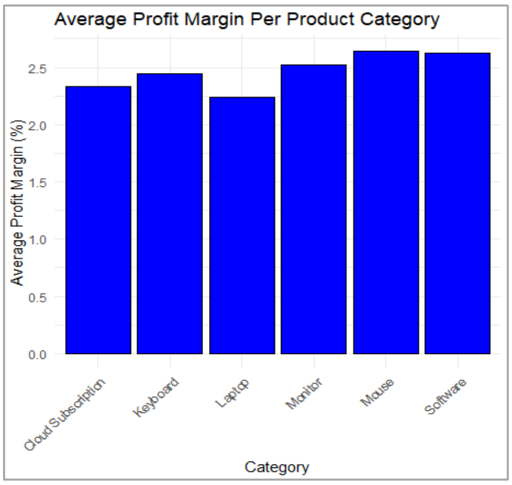


Figure 5: Average Profit Margin per Product Category

The most profitable product range, in terms of average profit margin, was the "*Mouse*" category, and "*Laptops*" the poorest, with figures spread between a low of 2.24% and 2.65% across all categories. These figures are worrying, especially when compared to the industry benchmark, which reported an average profit margin of 3.49% for the electronics sector (Blokhin, 2022) – suggesting that the company is not reaching its expected potential, even in its top category. The following investigation of the relationship between product selling price and its markup might provide reasons for this underperformance.

As seen in the figure alongside, a clear linear relationship exists between product selling price and markup for items priced below $5000 (orange). However, above this threshold, the markup allocations can often appear random and illogical, with extremely expensive items carrying low markups and moderately priced items reporting outlandishly high markups. The most unreasonable instances have been identified for review. Items in critical need of higher markups (indicated in the green square) include the "*black silk keyboard*", "*aliceblue silk mouse*", "*aliceblue silk Cloud subscription*", "*cornflower blue matt monitor*" and "*azure sandpaper software*". Items in dire need of reduced markups (indicated in the red square) include the "*blueviolet matt monitor*", "*burlywood sandpaper laptop*", "*black sandpaper laptop*", "chocolate silk mouse", "*cornflowerblue marble software subscription*", "*chocolate silk keyboard*" and "*blueviolet silk software*".

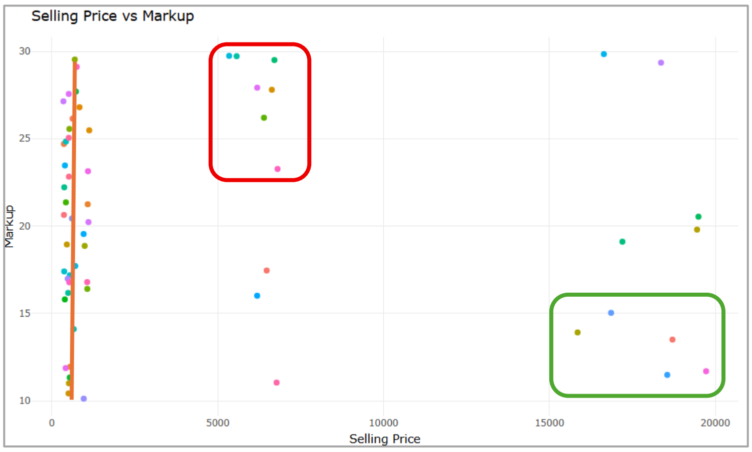


Figure 6: Product Selling Price vs Markup

When comparing the number of sales across product selling prices, we see consistent anomalies in the previously identified groups, with low markup expensive items (green) and mid-tier high markup products (red) recording lower than desirable sales figures, thereby confirming the presence of a price inconsistency. Applying the previously mentioned recommendations to these categories should boost sales, since slightly lower prices on the mid-tier items would make them more accessible to lower-income clientele, thereby boosting sales. The marketing department could pitch these discounts as limited promotional campaigns to drive demand and test the effects on sales. Concomitantly, slightly higher prices on the expensive items – which are already unattainable for most low to medium income consumers – would take advantage of the well-documented "snob effect" phenomena, thus boosting profits. Here, the marketing department could brand these items as "limited editions", rare, or only available to select customers with store-spending above a certain limit etc. to promote the effect. Summarily, it is projected that reasonable adjustments to these identified outlier products should result in increased sales for the overpriced items and greater profits within the underpriced ones.

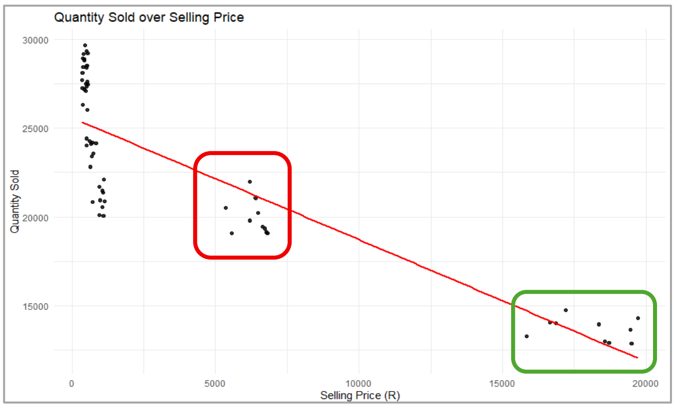


Figure 7: Product Quantity Sold vs Selling Price

2.3 Retail Performance

When looking at the sales breakdown per product category, it is seen that monitors are the most popular product, and the mouse is the least demanded (i.e. lowest sales). This is not ideal, especially since it has previously been shown that the "*Mouse*" category is the company's most profitable, which means that its sales should be prioritised if overall profitability is to be raised. By comparison, monitors are only the third most profitable category yet reports the highest sales. It is therefore recommended that the company strategically review their monitor pricing to take advantage of this high demand and thereby boost profitability. This trend is shown to hold in terms of brute revenue as well, as shown below in Figure 9 and Table 1, where the Monitor category consistently ranks 3rd in terms of garnered sales value.

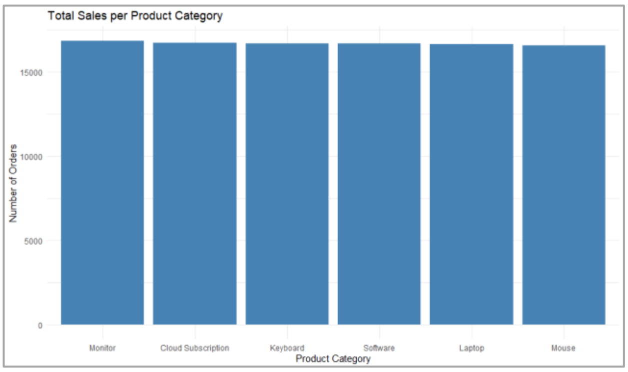


Figure 8: Total Sales per Product Category

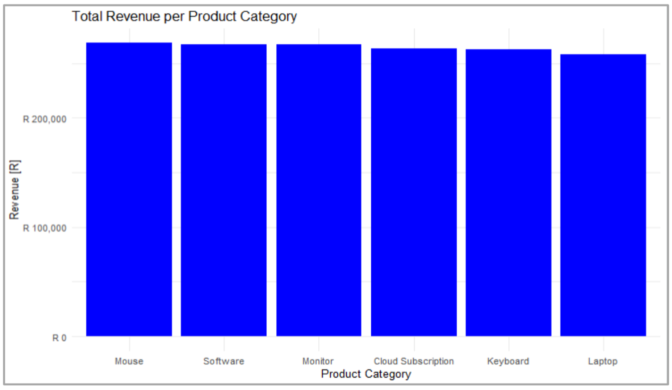


Figure 9: Total Revenue per Product Category

Table 1: Total Revenue per Product Category

Category	Revenue_in_Rands
Cloud Subscription	263,202.6
Keyboard	262,829.1
Laptop	258,344.3
Monitor	267,404.7
Mouse	268,734.0
Software	267,431.6

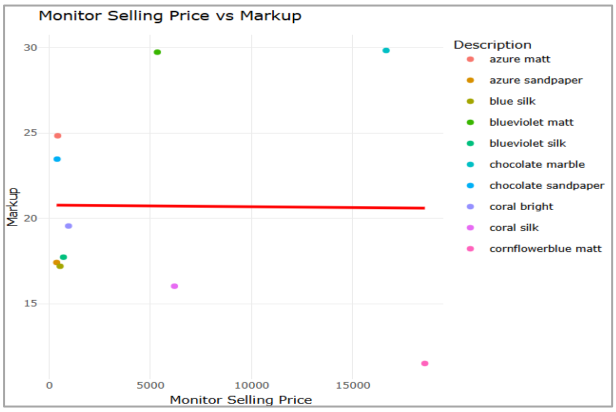


Figure 10: Monitor Selling Price vs Markup

In the graph alongside, a negative trend between selling price and markup for each type of monitor can be observed. This means that, in general, more expensive monitors are being priced lower than they could be in relation to their cheaper counterparts. Unless an aggressive pricing policy is intentionally being adopted to subvert competition, this trend would otherwise represent an illogical and inconsistent pricing scheme, since the company should be making at least as much, if not more, profit on premium items. When looking closer, it becomes clear that the higher priced monitors often seem randomly priced, with instances that are severely over or under marked and thus out of relation with established trends. There are a couple of specific instances that can be immediately reviewed to address this inconsistency. It is recommended that the "*cornflowerblue matt*" and "*coral silk*" monitor markups be increased by at least 10 and 50 points respectively and that the "*chocolate marble*" and "*blueviolet matt*" monitors be marked down by at least 5 points. It is projected that more consistent pricing principles in the best performing product category would not only lead to more consistent sales, but also greater profitability.

2.4 Sales Seasonality

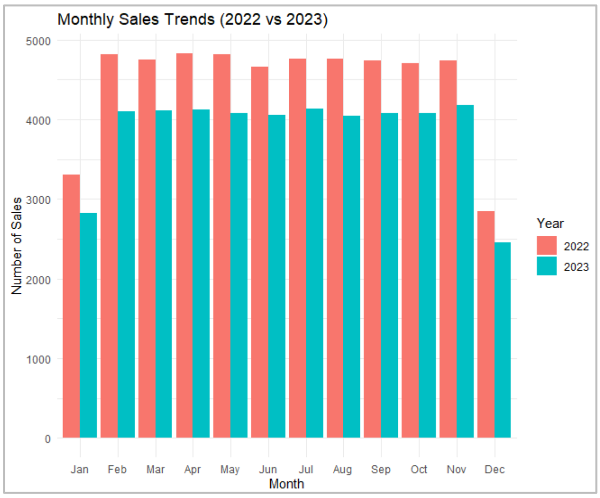


Figure 11: Aggregated Monthly Sales Trends (2022 - 2023)

When tracking the monthly number of sales recorded throughout 2022 and 2023, distinct trends emerge. For example, sales figures across all products remained fairly stable during the middle 10 months (Feb – Nov) but display vastly decreased numbers at the start (Jan) and end (Dec) of the year. This seems counterintuitive, since one would expect these two months to be some of the best performing, because December aligns with Christmas which is typically a large driving force of consumerism through gifting, and January coincides with the start of the academic year, when students typically need to make electronic purchases in preparation for their studies. However, it should be noted that purchases are not always made in the month of use, but rather when it is convenient to do so. For example, in the case of Christmas, where most sales would tend to be gifts, items would be purchased in advance so that they are ready to give as presents on the day. This is supported by the month-over-month increase observed in November, which coincides with Black Friday sales that are held in preparation for Christmas. Similarly, even though schools typically start in January, it is likely that students would postpone expensive electronic purchases until they have experienced the class demands of the new year, after which they will be able to make a more informed decision based on their needs. This is supported by the large month-over-month increase seen in February, which could be explained by student purchases after experiencing a month of class.

Furthermore, the year 2022 reported its highest sales in April, and 2023 in November, which is strange since it would be expected that the two should coincide. Sales were also notably diminished in 2023 across all product categories, compared to the previous year. This is in line with larger industry trends for the time period, which reflects a general sales decrease over the same period (Statista, 2022). There is a slew of economic factors that can account for this decrease. These two years experienced a post-pandemic global economic recession (Global-imi.com, 2023) which tends to raise prices through inflation while lowering disposable income, translating into less sales – particularly for consumer goods deemed less essential by the market. This period also witnessed the escalation of conflict between Russia and Ukraine which are both key sources of neon, palladium, nickel and platinum recession (Global-

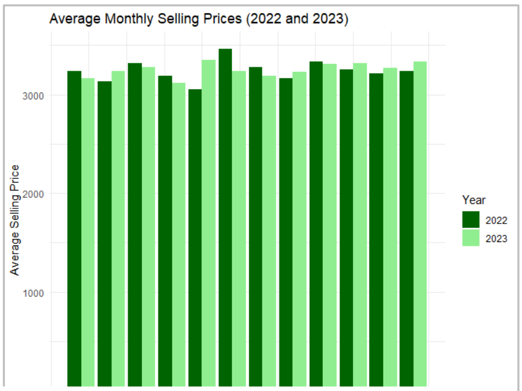


Figure 12: Average Monthly Selling Prices (All Categories - 2022 to 2023)

imi.com, 2023) – all of which are vital for semiconductor manufacturing. Scarcity of critical raw materials lead to increased prices that lower sales. This is confirmed by the graph alongside, as well as by independent calculation, which shows that 2023 did in fact have a higher average selling price across all products and months, at R3251.71, compared to R3237.47 in 2022. It is quite likely that this marginal difference could be explained by raised manufacturing costs due to more expensive components. Further developments include the Biden administration's changes to US regulation in an attempted to throttle Chinese semi-conductor access to the American market (Global-imi.com, 2023), which could have contributed to raised prices and diminished sales.

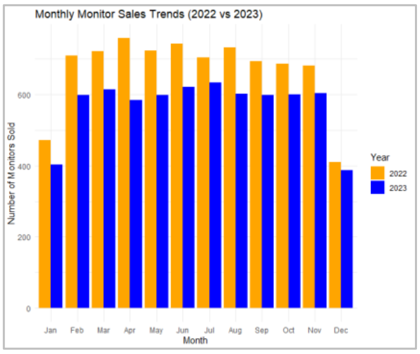


Figure 18: Monthly Monitor Sales (2022 - 2023)

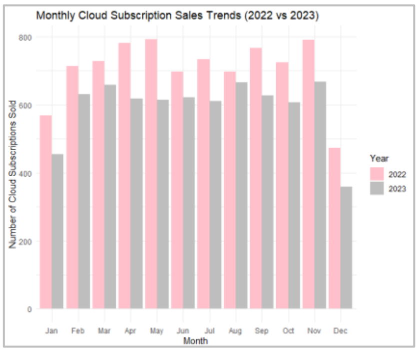


Figure 17: Monthly Cloud Sales (2022 - 2023)

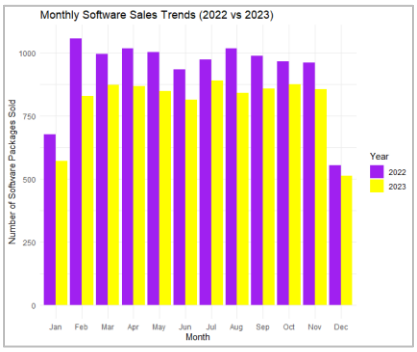


Figure 16: Monthly Software Sales (2022 - 2023)

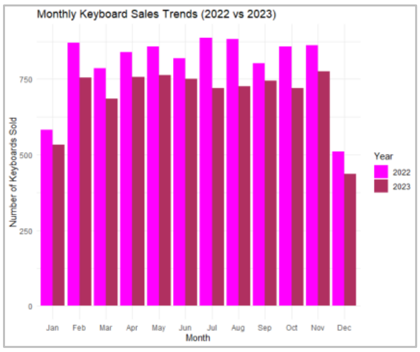


Figure 15: Monthly Keyboard Sales (2022 - 2023)

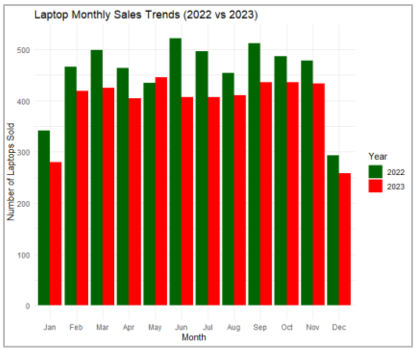


Figure 14: Monthly Laptop Sales (2022 - 2023)

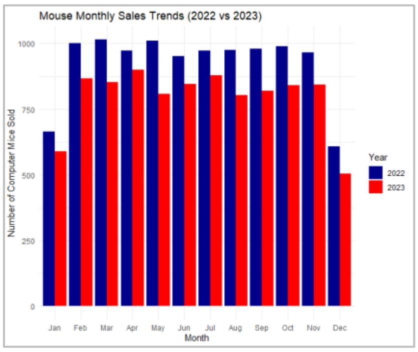


Figure 13: Monthly Mouse Sales (2022 - 2023)

Table 2: Average Selling Price Change per Category from 2022 to 2023

	Laptop	Mouse	Monitor	Software	Cloud	Keyboard
Change in Average Selling Price from 2022 to 2023	-R24.63	-R0.181	R3.56	- R0.05	R0.86	-R2.41

However, when investigating monthly sales trends by product, it is observed that although previously discussed distribution trends seem to hold true on an individual level, more detailed is needed to contextualise true behaviour. Evaluating each product separately reveals that most actually experienced a decrease in selling price from 2022 to 2023. Monitors and cloud subscriptions were the only two products whose average selling prices increased during that time. Since these items don't contain silicone computer chips, it appears as if the discussed macroeconomic events were less impactful than assumed, since the product with the greatest dependence on computer chips (laptop) experienced the greatest average year-over-year decrease. It then needs to be explained why the change in average selling price of all products previously rendered an increase. This is because of the classic statistical phenomenon of weighted averages. The mean of all products selling prices is influenced by volumes in the dataset, since the product with the most entries will skew the central tendency. As discussed in the "*Retail Performance*" section, the monitor was the best performing product with the most sales, but also the greatest single increase in average price, as seen above. Therefore, it's weighted impact would pull the average upwards, even though most individual product prices dropped over this period.

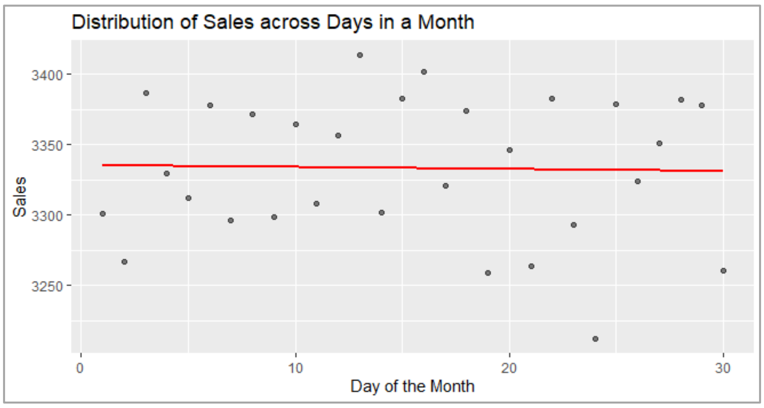


Figure 20: Sales Distribution Across Days in a Month

Interesting insights can also be gleaned by looking at the number of sales that occur on different days of a month to see if any days are more popular than others. What is revealed is a slight negative trend, meaning that most sales will typically occur at the start of the month, with less towards the end. This makes sense when considering that a great number of consumers will be salaried individuals who receive their income at the end of the previous month, therefore able to make large purchases at the start of the next. Similarly interesting trends emerge from the hourly behaviour of sales. There appears to be a strong positive trend of sale figures across daily hours, with the number of sales climbing steadily as the hours pass. There also appears to be a distinct clustering of large sales values in the midday hours (purple), suggesting that most of Company X's customers are "daylight" buyers, or prefer making purchases during the peak workhours of the day. When considering these trends in conjunction, useful adjustments to advertising campaigns, staffing arrangements and order scheduling can be made to best facilitate client rhythms. For example, employee staffing should reach its maximum during the start of the month, particularly during regular business hours, with less system resources allocated during early morning and late-night hours and tapering down as the month progresses. Promotions should also be planned for the end of the month in particular, to encourage buying during slower periods.

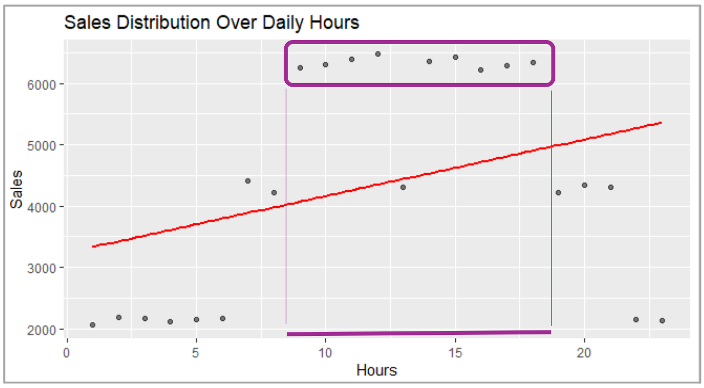
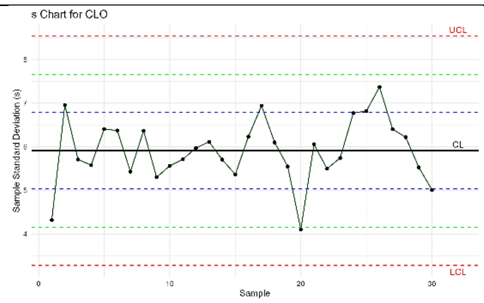
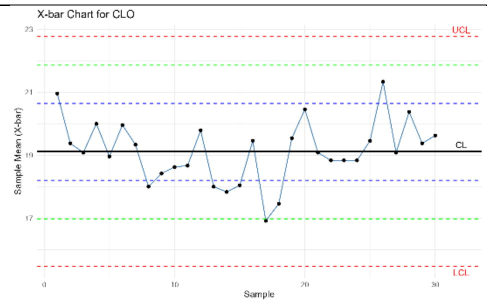
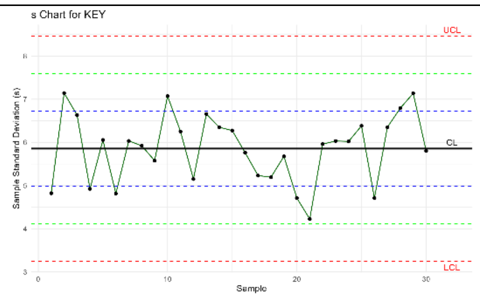
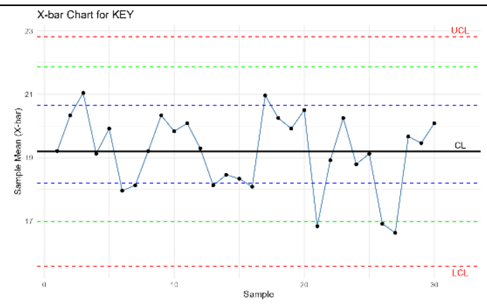
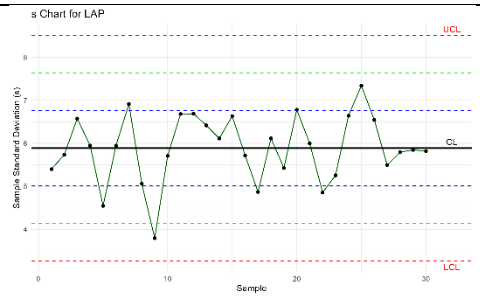
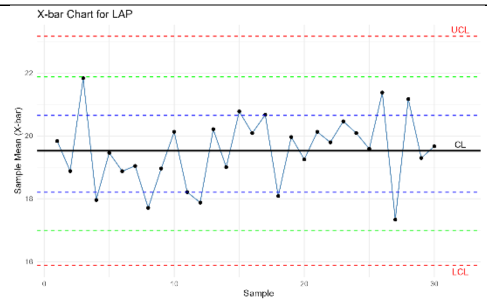
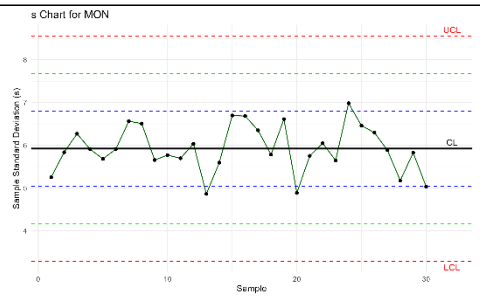
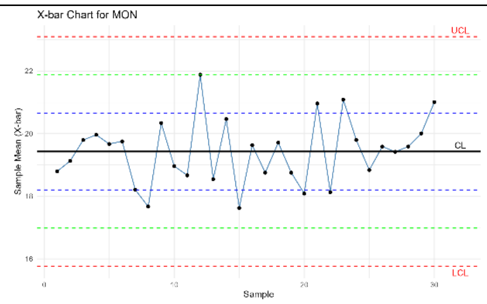


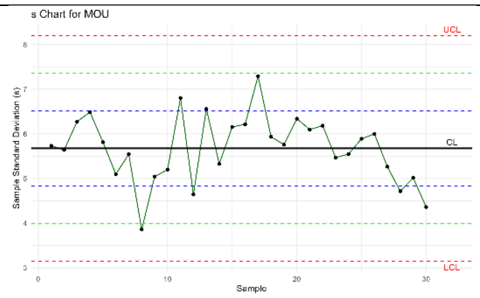
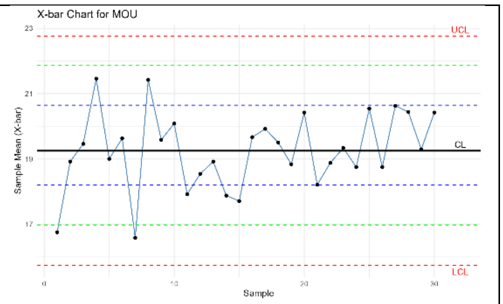
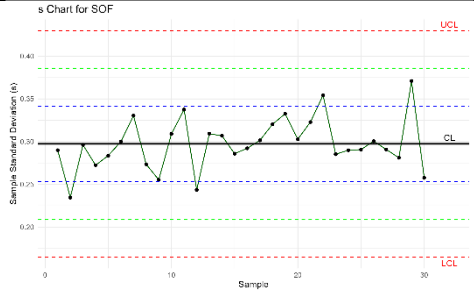
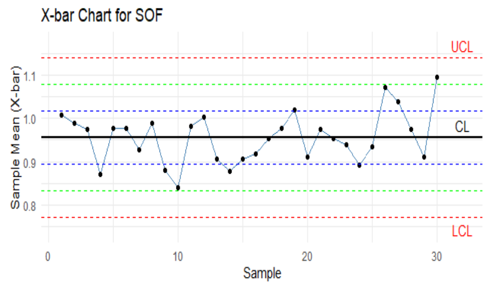
Figure 19: Sales Distribution Over Daily Hours

3. Statistical Process Control (SPC) Analysis

The next aim is to apply the principles of Statistical Process Control (SPC) to the company's sales data from years 2026 and 2027, particularly with regards to the recorded delivery times of each product. SPC analysis will help us to determine the stability of the company's delivery performance, as well as its capabilities (ability to satisfy the customer requirements or "Voice of the Customer" (VOC)), so that variation can be reduced and outliers eliminated through such remedies as root cause analysis (sixsigmamoneybelt, 2010).

3.1 Control Charts

Product Category	**S – Chart**	**$\bar{x}$ Chart**
Cloud Subscription	 *Figure 21: Cloud Subscription S-Chart*	 *Figure 22: Cloud Subscription $\bar{x}$ Chart*
Keyboard	 *Figure 23: Keyboard S-Chart*	 *Figure 24: Keyboard $\bar{x}$ Chart*
Laptop	 *Figure 25: Laptop S-Chart*	 *Figure 26: Laptop $\bar{x}$ Chart*
Monitor	 *Figure 27: Monitor S-Chart*	 *Figure 28: Monitor $\bar{x}$ Chart*

Mouse	 *Figure 29: Mouse S-Chart*	 *Figure 30: Mouse x̄ Chart*
Software Package	 *Figure 31: Software Package S-Chart*	 *Figure 32: Software Package x̄ Chart*

At first glance the s (standard deviation) and x̄ (sample mean) charts depicted above for each product category, seem to display perfectly normal and expected behaviour. Most points seem to fall between the ±1σ (blue) limits, and no obviously discernible trends are immediately apparent. However, upon closer inspection, it is seen that the Mouse category displays a concerning amount of sample points outside of the 2σ (green) limits, suggesting a high tendency towards variability. Also, when looking at changes in the sample means of Software Packages and Cloud Subscriptions, it appears as if a slight parabolic trend exists, with values in later samples moving slightly upwards as the sample number increases. We can also deduce from the y-domain of theses charts that the Software category has the tightest spread of values, making it the most precise and accurate, whereas the Mouse and Laptop categories seem to report the widest spread throughout its samples, suggesting high tendencies towards variability and therefore instability.

3.2 Test Continued Sampling Against SPC Charts

By using the produced S and x̄ charts, it is possible to determine the distribution or "spread" of specific samples, do identify cases that exceed acceptable limits. This is important, because unusually large spreads would render insights from the x̄ charts unusable. By proper evaluation of R-and-S-chart data, it is possible to identify specific product cases that require physical adjustment or alteration by a manager. Such instances have been identified and tabulated below:

Table 3: Identifying "Out-of-Control" Products

ProductID	sample_number	s	xbar	action_needed
CLO011	39	4.138017	16.96267	⚠️ S out of control. Investigate variation first.
CLO013	44	3.306504	19.58767	⚠️ S out of control. Investigate variation first.
CLO014	65	2.192780	18.18857	⚠️ S out of control. Investigate variation first.
CLO015	66	5.697872	26.17675	⚠️ X-bar out of control. Investigate mean shift.
CLO019	64	3.287718	19.46233	⚠️ S out of control. Investigate variation first.

It is clear from the results depicted above, that the only product category with out-of-control delivery times, relative to its central limit tendency, is the cloud subscription items. Four instances where recorded where the item's standard deviations registered as "out of control" when compared to its sample characteristics. This signals a high degree of variability within the cloud subscription delivery process, meaning some subprocess is likely talking far longer or happening much faster than intended, resulting in unpredictable behaviour and a larger degree of variability than is acceptable. In this case, it is recommended that the specific product and its corresponding delivery scheme be physically inspected by a manager for the presence of inconsistencies and errors. In contrast, the one instance where the product's $\bar{x}$-value was out of control, represents a case where the sample mean has been shifted beyond acceptable limits ($\pm 3\sigma$ from centre line). This indicates that some unknown influence is significantly skewing the average tendency of the process and warrants physical inspection by a manager to determine the root cause of the bias.

3.3 Identify VOC-Capable Products

Table 4: Satisfactory Process Capability Indices

ProductID	Mean	SD	Cp	Cpu	Cpl	Cpk	Capable
SOF009	1.09	0.31	17.473	33.759	1.188	1.188	Yes
SOF010	1.07	0.30	17.938	34.676	1.201	1.201	Yes
SOF007	1.09	0.30	17.566	33.938	1.194	1.194	Yes
SOF001	1.07	0.31	17.268	33.381	1.155	1.155	Yes
SOF002	1.07	0.31	17.132	33.119	1.144	1.144	Yes
SOF005	1.08	0.31	17.296	33.426	1.166	1.166	Yes
SOF003	1.07	0.30	18.027	34.850	1.205	1.205	Yes
SOF004	1.07	0.30	17.634	34.088	1.180	1.180	Yes
SOF006	1.06	0.30	17.554	33.944	1.164	1.164	Yes
SOF008	1.08	0.29	18.255	35.283	1.227	1.227	Yes

After calculating the process capability indices for the process delivery times of all product types and filtering out all items with Cpk values between 1 and 1.3, it was found that the only product category capable of satisfying the voice of the customer (VOC) was software, as shown in the table alongside. By this logic, the most capable item was the *SOF009* software package. In theory, a high process capability performance (Cpk) value would also indicate a tight spread within specification limits (minimal variation), which is corroborated by the previously displayed s and $\bar{x}$ charts for this category, which record far smaller distributions than others. In other words, the Cpk signals how close the process mean is to the process target value. Therefore, Software is the only product category that, on average, met the customer expectation targets.

3.4 Identify Samples with Process Control Issues

One of the benefits of SPC analysis is its ability to spot outliers amidst samples that contribute to unacceptable levels of variability in the system. Below, three heuristic criteria (A-C) have been applied to identify products that are problematic or out of acceptable performance bounds.

A. ***Samples Exceeding Upper +3σ Control Limit by 1 S:***

The table alongside displays the first and last three instances of identified products that exhibited standard deviations of at least one standard deviation above the upper $+3\sigma$

Table 5: SPC Test A Results

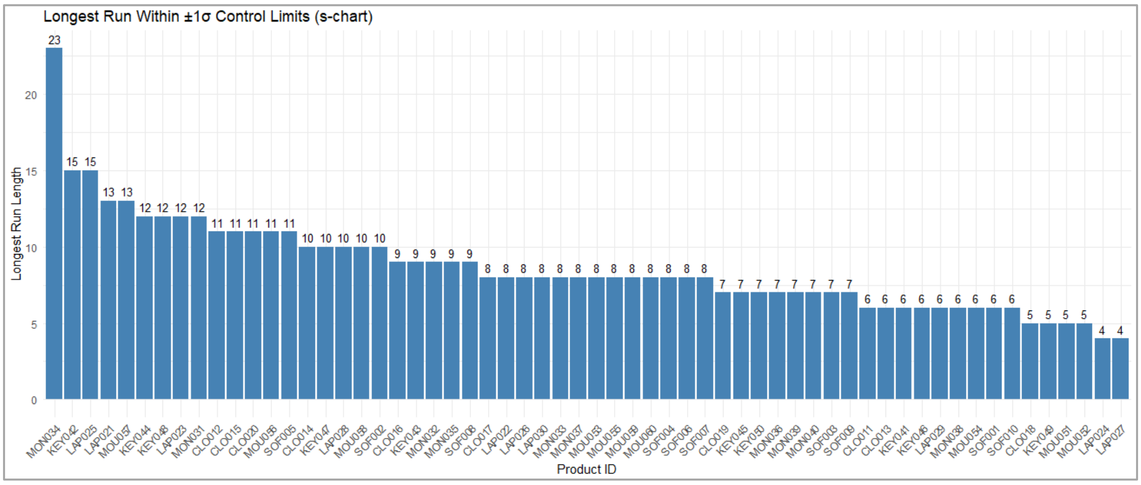
ProductID	sample_number	s	UCL_s_3sigma
LAP021	32	8.0223589	7.8748354
LAP026	41	9.1280521	7.8758874
MON032	61	7.7328016	7.4756885
MOU060	85	8.3667855	8.1683628
SOF003	66	0.3985879	0.3963741
SOF004	86	0.4593387	0.4301924

control limit of their sample. These cases represent extreme deviations from acceptable limits, likely indicating the presence of some external influence, like environmental factors or operator error etc. These cases should be individually investigated to determine their causes via root cause analysis (6-Sigma, 5-Why's etc.) Additionally, there does not seem to be much of a discernible trend in affected product categories, with laptops, monitors, mice and software all represented (the table only depicts the first and last three spotted instances, as per instructions, making it likely that all product categories are represented in the full list). This suggests that whatever factor(s) are responsible for these variations, likely affects all categories and therefore system-wide processes that affect all delivery types (such as employee training, delivery methods, customer service etc.) should be investigated first.

B. *Greatest Number of Consecutive Samples with -1≥S≤1 Sigma Control Limits:*

This measure tested for the greatest number of consecutive samples that had standard deviations between -1 σ and 1 σ from the CL. This range typically represents statistically insignificant deviations and is therefore an indication of near-perfect (or ideal) performance. The graph below therefore displays the greatest number of instances where a certain product recorded consecutive samples of a (near) perfect delivery time. As shown, the product with the greatest number of consecutive delivery times within this range, is MON034 monitor, with 23 recordings. The lowest performing products that managed to list were the LAP024 and LAP027 laptops, at 4 instances each.

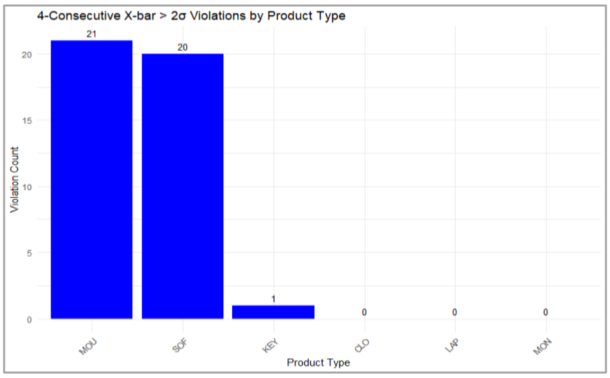
Table 6: SPC Test B Results



Although it might seem desirable to have as many cases of invariance as possible, this is not always the case. If a sample's standard deviation is routinely registered this close to the centre line, it might suggest an unnatural degree of consistency - or consistency beyond what would be seen from random variation – likely caused by an undue external influence or methodological fault. There are many possible causes, including measurement defects (such as instrumentation failure or calibration flaws etc.), incorrect sampling, data tampering etc. each of which needs to be investigated. The product could be performing much worse than displayed, because of a disconnect between the readings and reality. A sign that this might be the case, is values that repeatedly show up as too good to be true. These need to be scrutinized for hidden methodological errors, that could potentially affect other unrelated processes as well.

C. *Samples with Four Consecutive instances where $\bar{x}$ exceeds 2σ:*

Table 7: SPC Test C Results



This test identified samples per product category that displayed four consecutive instances where the item mean exceeded two standard deviations from the sample centre line. To put this in perspective, under normal circumstances there is about a 2.3% chance that a single point would have a standard deviation beyond +2σ. In our case, we tested for four consecutive points **all** above this limit, which by the laws of cumulative probability is exponentially less likely to occur in a natural dataset. This means that an unintentional external influence is positively skewing the data, causing it to drift upwards away from the CL. Only three product categories displayed this tendency, namely the Mouse, Software and Keyboard groups. The most concerning case is the Mouse group, which recorded 21 such instances, thereby indicating a significant departure from its target centre, with the Software category not far behind at 20 readings. It is recommended that the delivery process of these categories be immediately investigated for undue influences that could contribute to this phenomenon.

4. Risk Analysis and Data Correction

The following section seeks to evaluate the risks associated with making various errors in the previous SPC Sample groups. The likelihood of both Type I and II errors will be estimated, and their associated implications explained. The author of this report was also notified of erroneous entries in the previously provided "*products_Headoffice.csv*" dataset, which was to be amended.

4.1 Estimation of Type I (Manufacturer's) Error Likelihood

Type I errors, also known as "false positives", typically refer to the mistake of incorrectly rejecting a null hypothesis (H_0). This essentially registers a change or phenomena which is not present in reality, thereby leading to a rejection of H_0 without true cause (Kameleoon, 2024). For the purposes of this report, it was assumed that the null hypothesis represents a state where the process is both in control and centred on the centreline, calculated using the first 30 samples. From this assumption, a normal distribution with $\sigma = \frac{UCL-LCL}{6}$ could be expected, where the samples are symmetrically balanced around the CL, meaning that the probability of a sample lying above or below it is exactly 50% either way. The likelihood of detecting a false-positive reading was then estimated for the three SPC case (A – C) previously investigated in Section 3.4. The results of each are tabulated below:

Table 8: Type I Error Estimation Results

Rule	Probability_of_Type_I_Error	Interpretation
A: s > +3σ	0.001349900000	0.135% chance that a single sample exceeds UCL (false alarm)
B: within ±1σ	0.682689000000	68.2689% chance that a sample falls within ±1σ (good control)
C: 4 x̄ > +2σ	0.000000267877	2.679e-05% chance that four consecutive samples fall above +2σ (false alarm)

Applying Shewhart's rule for condition A (where $sd_{\text{sample}} > 3\ \sigma$), returned a Type I error probability of **0.135%** - meaning, under stable conditions (H_0), test A should only return a false positive 0.135% of the time, making it a very reliable metric for gauging the performance of the delivery process. Test B however, returned a probability of **68.2689%**, meaning that testing whether a sample's standard deviation falls within the bounds of $\pm\ 1\sigma$, will return a falsely positive reading about 68.3% (or the vast majority) of the time, thereby making it a heavily unreliable reading which should not be interpreted as an indication of process control in isolation. Test C returns the lowest probability score, indicating that it is the most reliable metric of the three and the least likely to return a false-positive reading, because the chance of 4 consecutive sample means being greater than $+2\sigma$ is so small as to be nearly statistically impossible.

4.2 Estimation of Type II (Consumer's) Error Likelihood

It was requested that the likelihood of a Type II error be estimated for an entirely unrelated process, that deals with the filling of bottles. A type II error is almost the inverse of a false positive, because it occurs when changes brought on by a test are not detected and therefore the incorrect conclusion that these changes had not effects is made (Kameleoon, 2024). The process underwent an undetected mean shift in bottling volumes and the probability that this change is not recognised was calculated to be approximately **84.1178%**, as shown in the table below:

Table 9: Type II Error Estimation Results

Probability_of_Type_II_Error	Interpretation
0.841178	There is a 84.1178% chance that the process mean shift from 25.05 L to 25.028 L goes undetected.

This value is concerning, because it suggests an incredibly high chance of accepting the null hypothesis (H_0) despite fundamental changes in the mean from 25.05 L to 25.028 L. Although this shift seems relatively small, it is still statistically significant and failure to detect it could have critical ramifications for the stability of the process. There are multiple factors that could account for this large probability of missing the shift, such as the sample size, frequency of testing and sensitivity of the measurement equipment. It is recommended that the testing methodology be independently reviewed to ensure that all components function as intended. Sample sizes could be enlarged to gain representational diversity. Additionally, an investment in more precise machinery should be considered to further refine the analysis.

4.3 Data Correction and Associated Analysis Impact

The data file titled "*products_Headoffice.csv*" was reformatted to remove certain errors such as price and markup discrepancies between its content and those values recorded in "*products_data.csv*". Instances of incorrect *ProductID* identifiers were also updated to reflect the correct prefixes. The Descriptive Statistical Analysis performed in Section 2 was then redone to check whether or not the updated data had any significant impact on the previously obtained results. For the most part, it was found that the erroneous data had no statistically significant impact on most of the metrics that were presented in Section 2. This is because the analysis in Section 2 primarily focused on information from other datasets, such as the customer, product and sales information, which was correctly recorded from the start. It is also important to note that the information contained in the faulty "*products_Headoffice.csv*" file is essentially a redundant repetition of records that are correctly dispersed across the other datasets. Therefore, little attention was paid to this set of information to begin with, because it was possible to draw conclusions on its contents through the other files – particularly the *products* dataset. The only glaring difference in results, was with respect to the revenue visualisations, which relied on the Head Office product data. The revenue amassed per product category, differed wildly when using the updated "*products_Headoffice.csv*" dataset. As is evident by comparing the following figures with those previously displayed in Section 2, it is clear that the previous Head Office figures were overinflated to present much larger revenue than was truly recorded by Company X. These corrected figures are far more realistic, with the Laptop, Software and Monitor categories contributing the bulk of the revenue, which is to be expected, since these are the most expensive items.

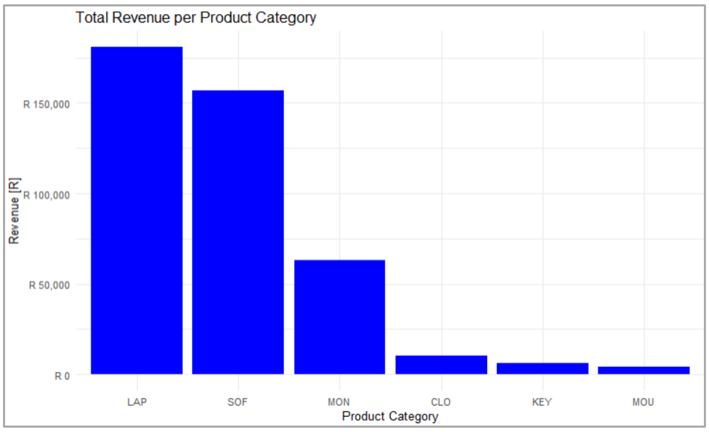


Figure 33: Total Revenue per Product Category (Corrected Head Office Data)

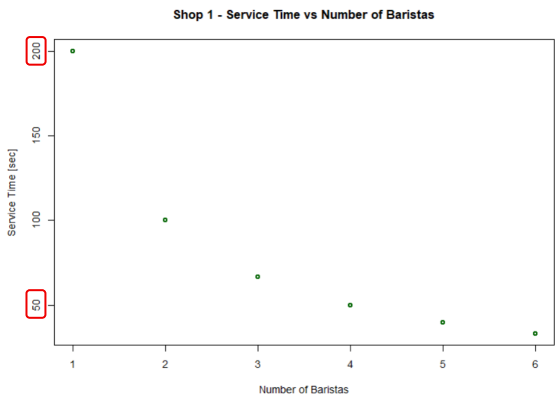
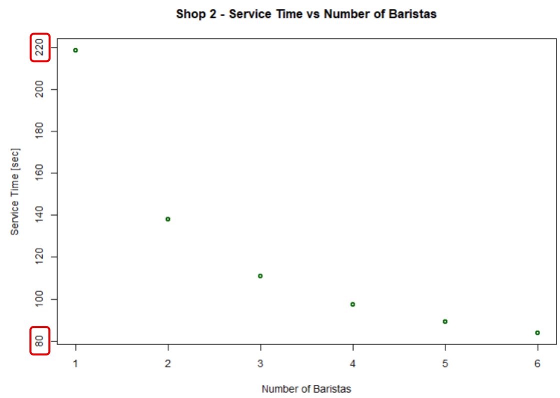
Table 10: Total Revenue per Product Category (Corrected Head Office Data)

Category	Revenue_in_Rands
CLO	10,190.62
KEY	6,446.60
LAP	180,864.29
MON	63,105.25
MOU	3,946.98
SOF	156,916.73

(To see the process of formatting and updating the flawed "*products_Headoffice.csv*" to the corrected "*products_data2025.csv*", please see the relevant code portion in the submitted "*QA344 ECSA Final Report 2025 Code (26892901).Rmd*" companion file.)

5. Optimising for Maximum Profit

Service times and their corresponding employee figures were provided for two separate coffee shops with the purpose of identifying the optimal number of baristas that should be employed to maximise profit. For the purposes of the exercise, it was assumed that the coffee shops could not hire more than six baristas each, due to space and availability constraints. With the number of baristas and service time as the independent and dependent variables respectively, it was possible to approximate a function for each shop's data by essentially fitting a polynomial to the data points through non-linear regression analysis. These models could then be used to estimate service delivery times, by seeding the function with a range of values corresponding to hypothetical numbers of baristas. By interrogating these derived relationships, it was found that an exponential disproportionality exists between these two variables, as depicted in the graphs below. This makes logical sense, because a greater number of baristas would be able to make more orders (or make the same number of orders faster), thereby decreasing the service time. This logic is visually depicted below in figures 33 and 34:

Shop 1	**Shop 2**
 *Figure 34: Shop 1 - Service Time vs No. Baristas*	 *Figure 35: Shop 2 - Service Times vs No. Baristas*

It is however interesting to note the difference in service time ranges between *Shop 1* and *2*, given the same number of employees. Given that baristas have the same level of training and work at comparable speeds, there is little immediately obvious reason to expect vast discrepancies between service performance. However, a single barista operating in *Shop 2* takes approximately 20 seconds longer than their peer in *Shop 1* and similar gaps seem to hold at every level, where the same number of baristas deliver vastly slower service times in *Shop 2* compared to *Shop 1*. This strongly suggests a difference in operational procedure between these two locations. There are various reasons that could explain *Shop 1*'s greater efficiency, ranging from pre-preparation protocols, a superior store (workstation) layout or the integration of modernised Point of Sale (POS) systems etc. It is recommended that the management of *Shop 2* study these divergent tactics to improve its output and align its service times with those achieved by *Shop 1*. These projected service time values could then be used to compose a function that represents the expected profits for each coffee shop. Net profit was calculated as *Revenue – Staff Costs*, where revenue and staff costs were determined assuming R30 spent per customer and a salary of R1000 per barista, per day. It was also assumed that a full workday is comprised of a standard 8 hours. Anticipated profit could then be plotted against the number of baristas to visually inspect the effect of barista numbers on projected profit for an ideal value, as tabulated below in figures 35 and 36:

Shop 1

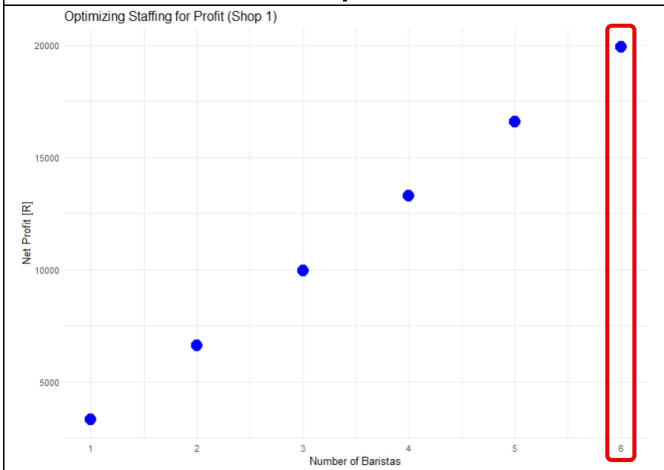


Figure 36: Shop 1 - Net Profit vs No. Baristas

Interpretation:

For *Shop 1*, the charted relationship between number of baristas and net profit displays a positive, linearly proportional trend upwards. If this pattern were to be extrapolated indefintely, it would seem as if profits will continue to rise continuously, as more baristas are added. Therefore, the maximum profit corresponds to the greatest number of baristas available, which the assumed limit stipulated as six. In reality, this will likely not be the case, since the total revenue that can be potentially extracted from customers is limited by factors such as the number of customers in the vicinity and the size of the shop, which could not accommodate either infinite customers or baristas. As the amount of customers served approaches its natural limit, the benefits accrued from hiring additional baristas would taper off into diminishing returns. It is believed that this logic is not apparent in this graph due to defective data – perhaps selective or insufficient sampling etc.

Shop 2

Figure 37: Shop 2 - Net Profit vs No. Baristas

Interpretation:

For *Shop 2*, the relationship between net profit and number of baristas seems to follow a somewhat negative parabolic trend, increasing steadily until an apex is reached, at which point the costs of additional labour surpasses the marginal benefits incurred, whereafter further hiring will only decrease profit. This inflection point signals four baristas as the optimal number to hire, and is likely caused by environmental factors that limit the scaling of further growth, such as the shop dimensions, the population size and its share of coffee enthusiasts etc. The behaviour observed in *Shop 2*'s chart is far more realistic than that shown by *Shop 1*, which suggests superior sampling and/or information handling in the collection of *Shop 2* data.

To confirm the conclusions drawn above through visual inspection, the built-in "optimize" R function was used to inductively determine the optimal balance of barista employees through brute force. The results concur that for *Shop 1*, an ideal number of six employees would result in the maximum profit of R19 923.26, and for *Shop 2*, four employees would maximise expected net profit at R4 881.55, as seen below:

Shop 1

Table 11: Shop 1 - Confirmation of Optimal Baristas & Max Profit

Metric	Value
Optimal Number of Baristas	6.00
Maximum Profit [R]	19,923.36

Shop 2

Table 12: Shop 2 - Confirmation of Optimal Baristas & Max Profit

Metric	Value
Optimal Number of Baristas	4.00
Maximum Profit [R]	4,881.55

Additionally, it was requested that the percentage of clients who should expect reliable service from each shop be estimated. Assuming that a reliable order is delivered within 90 sec, it was determined that *Shop 1* would reliably serve **85.72%** of customers whereas *Shop 2* would only manage to reliably serve **42.62%** of customers within the same timeframe. This conclusion is independently affirmed by figures 34 and 35, which previously demonstrated similarly glaring discrepancies in service times between the two shops.

The Taguchi Loss Function, is defined as a "mathematical representation that quantifies the relationship between product quality and the financial loss (loss to society) associated with deviations from a target value." (Quality Gurus, n.d.) Conceptually, *Shop 1*'s setup differs from the Taguchi loss function, which is represented by a quadratic curve where financial losses exponentially increase with changes in the target value variable (no. baristas) (Quality Gurus, n.d.). Since *Shop 1*'s net profit function is a straight-line graph, changes in the number of baristas does not have an exponential effect on the net profit. However, *Shop 2*'s behaviour strongly correlates with the parabolic behaviour of the Taguchi function, where even small changes in barista number seem to induce exponential changes in net profit. However, the Service Time vs Barista curves for both shops exhibit decreasing exponential functions, which suggests that any deviation from their ideal staffing target values is likely to increase quality-related costs (increase service times thereby decreasing profits). It can therefore be said that both shops display similarities to the Taguchi loss function but *Shop 2* displays a much stronger correlation due to the exponential nature of its profit function, which is very sensitive to variation in staffing quantities, thereby exhibiting strong correlations between product quality and financial success.

6. DOE and MANOVA

When evaluating the results from the previous SPC analysis in Part 3, it is clear that the *Cloud Subscription* products proved to be the most problematic category, constituting all of the identified "out-of-control" processes. It was also previously established that *Cloud* products experienced a sizable decline in sales from 2022 to 2023. For the purposes of this experiment, the *Cloud* products category will be investigated using a multivariant analysis of variance (MANOVA), to determine whether explicit changes in delivery and picking times transpired during that time period. Insight into this behaviour could help identify the underlying causes behind the categorical instability and hopefully shed light on possible remedies. The design for the experiment parameters is as follows:

H_0: there exists no consequential change in the combination of average delivery and picking hours of cloud products, between the years 2022 and 2023.
H_1: a multivariate discrepancy exists for either or both the average delivery time or picking time of cloud products between 2022 and 2023.
$\boldsymbol{\alpha} = 0.05$

After importing the "sales2026and2027 (1).csv" dataset and filtering for relevant *Cloud* product entries (as well as removing any empty points), the MANOVA protocol could commence through use of the relevant built-in R functionalities. The procedural test results are summarised in the table below:

Delivery Hours

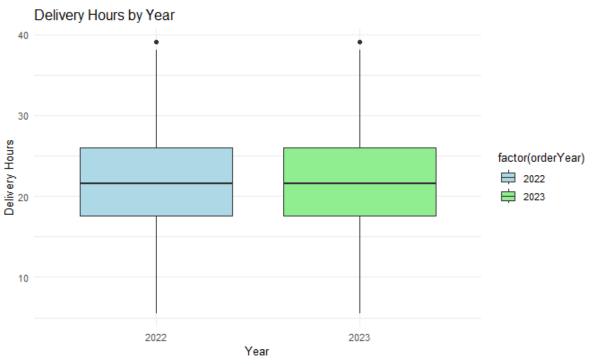


Figure 38: MANOVA Delivery Hours by Year (2022 vs 2023)

Table 13: MANOVA deliveryHours Results

```
 Response deliveryHours :
                Df Sum Sq Mean Sq F value Pr(>F)
orderYear        1      1   1.170  0.0313 0.8596
Residuals    15596 583187  37.393
```

Interpretation:
The p-value for the *Cloud* products' delivery hours was 0.8596, which is much larger than 0.05, indicating that there is no statistically significant difference between recorded delivery hour values for 2022 and 2023. Additionally, the small F-value of 0.0313 is extremely close to 0, which suggests that the changes between years is small compared to those that occur within a year, meaning that the delivery hours fluctuate considerably more within each year than across the two. These findings suggests that no drastic change in delivery hours took place for *Cloud* products between 2022 and 2023. This conclusion is useful since it signals that none of the managerial or operational decisions that might have been made during this transition could have caused the instability within the *Cloud* category delivery times. This deduction is also supported by the box-and-whisker plot above, which shows virtually identical figures for both years, thereby indicating predictability in *Cloud* product delivery times across 2022 and 2023.

Picking Hours

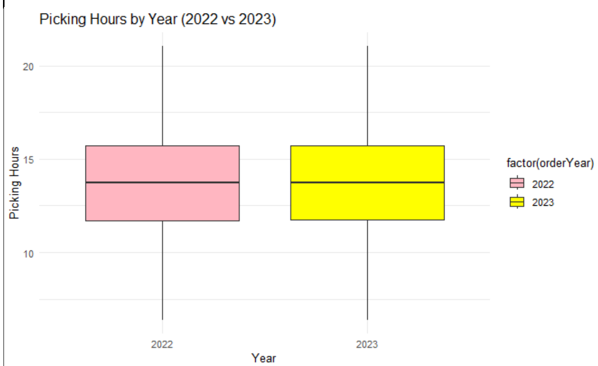


Figure 39: MANOVA Picking Hours by Year (2022 vs 2023)

Table 14: MANOVA pickingHours Results

```
 Response pickingHours :
                Df Sum Sq Mean Sq F value Pr(>F)
orderYear        1      1  0.5196   0.064 0.8003
Residuals    15596 126699  8.1238
```

Interpretation:
The p-value for the *Cloud* products' picking hours was 0.8003, which is also much larger than 0.05, thereby indicating the absence of any noticeable difference between values across 2022 and 2023. Furthermore, its F-value of 0.064 was also close to 0, which once again suggests that the changes within the year were greater than those between the two years, as confirmed by the large mean sq value. This suggests that no significant change occurred in picking times between 2022 and 2023, meaning that picking procedure likely did not change and is therefore not responsible for the previously observed decrease in year-over-year *Cloud* product sales. These conclusions are also substantiated by the accompanying box-and-whisker plot, which displays nearly identical charts for picking hours in both years, thereby confirming the stability of picking metrics for 2022 compared to 2023. However, this does not mean that these values are stable within the years themselves, just that the degree or tendency of instability did not shift noticeably over this time span.

When taken in unison, the results of the MANOVA test suggests that there was no significant change in either delivery or picking hours for *Cloud* products, between 2022 and 2023 - thereby suggesting that no major changes or reforms were made by Company X to either process during this time. As a result, the alternative hypothesis (H_1) failed to unseat the null hypothesis (H_0), as neither picking or delivery teams underwent consequential year-over-year changes. In fact, these procedures remained surprisingly consistent over this period, which rules out changes made between the two years as possible contributors to the *Cloud* category's delivery time instability (identified in Part 3). However, both delivery and picking times recorded unusually large residual mean sq values, which confirms the conclusion drawn from the *Cloud* SPC chart that large internal instability still persists within these measures. Therefore, it should be noted that the identical nature of the box-and-whisker plots does not mean that the process is stable within each year – such notions are contradicted by the previous *Cloud* SPC charts and high mean sq values. On the contrary, the similarity only signals that the internal tendency for instability, or the degree to which readings fluctuated, did not change noticeably from 2022 to 2023. In essence, the process is as unstable in 2023 as it was in 2022. This is useful to know, because it affords two critical insights. Firstly, it dispels the idea that the decreased sales figures previously observed in the *Cloud* category, were brought on by changes in either delivery or picking times over 2022 to 2023. As seen in the previous interpretation, this is clearly not the case. Therefore, other culprits such as market sentiment, economic behaviour and product popularity etc. should be considered for their influence. Secondly, it proves that any changes in operational protocol or strategy which the company implemented in 2022 to try and decreases delivery and picking times for 2023 were essentially inconsequential. The management team should reconsider their efforts to try and generate new approaches that will reduce picking and delivery times. Options to evaluate include warehouse automation, digital stock-keeping systems, lean and/or 6 Sigma methodologies etc.

7. Reliability of Service

A histogram displaying the number of "on-duty" employees of a car rental agency over 397 days was provided, with the aim of estimating how many days per year a customer should expect reliable service. It was assumed that the establishment could provide reliable service with a minimum of 15 employees present. Using the provided information, it was possible to define two variables in R, namely *absence_ratios* for the ratio of each instance where a certain number of employees are absent and *weights* for the weighting of each case in *absence_ratios* out of the total cases. This information was then used to compute results of a scalar variable, *p_weighted*, which represents the weighted means of *absence_ratios* according to the formula, $p_{weighted} = \frac{\sum(absence\_ratios \times weights)}{\sum weights}$. After which the "*pbinom(...)*" R function could be used to determine the probability of days where 15 workers or more were present, or P(x ≤1) out of the given set. This probability could then be scaled down to the length of a year, which returned the result that **342 days** of reliable service can be expected per year, assuming that a minimum of 15 workers is required to provide reliable service.

It was then necessary to optimise the agency's profit. Every day where there is less than 15 employees on-duty, costs the agency R20 000. Additional personnel can be hired at the rate of R25 000 per month per person. Since no information on the company's revenue or sales was provided, it was assumed that the net profit would be a maximised when the money lost on unreliable sales due to staffing shortages are minimised. Lost revenue was approximated as $(365 - no.reliable\ days) \times 20\,000$ and the staffing costs as $(extra\ employees) \times (no.months\ short) \times (25000)$, where the number of months that additional help will need to be hired for is found rounding the number of unreliable service days up to the closest month.

The total losses are the sum of both quantities. All these values were defined as characteristics of a loss function, which is then minimised using built-in R functions to determine the optimal number of employees that correspond to the lowest possible loss, which conversely correlates to the greatest profit. The results are depicted below:

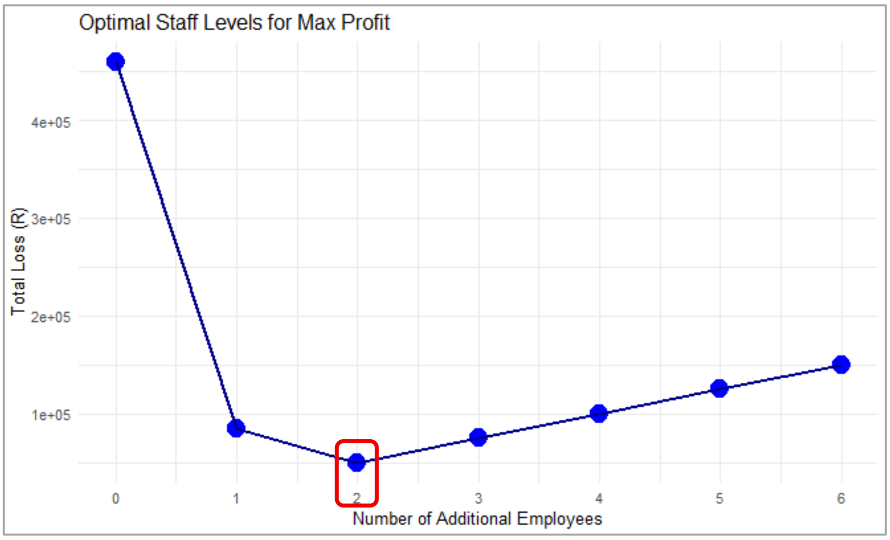


Figure 40: ROS - Optimal Staffing for Maximal Profit

Metric	Value
Optimal Number of Additional Employees	2
Minimum Total Loss	50,000

Table 15: ROS - Optimal Staffing for Maximal Profit

As is clearly shown in Figure 40, the total losses decline steadily as additional personnel is hired, until a global minimum is reached at two additional employees. After this point, the cost of hiring of further employees would start to outweigh the benefits previously incurred from saving lost business by increasing reliability. The minimum total "loss" is therefore the amount paid to retain two additional employees for one month at R25 000 each, since this counteracts the losses that would have accrued due to lower reliability. Since this is the lowest projected amount of money that could be forfeited in this setup, it is also assumed that this is the arrangement in which maximum profitability can be extracted by the agency.

8. Conclusion

In conclusion, the performed statistical analysis has delivered interesting and useful insights into the nature of data yielded by various small business ventures. Descriptive Statistical Analysis was applied to identify key underutilised demographics and opportunities, to focus the operations of and maximise value extraction for Company X. Retail performance and sales trends were also evaluated to help position the company's future promotional strategy and resource allocation with the aim of increased efficiency. Furthermore, Statistical Process Control methods where utilised to analyse the delivery time data for sales in 2026 and 2027 and illuminated areas of success and future improvement. S and $\bar{x}$ control charts were implemented alongside a set of rigorous heuristic tests, to identify products and categories that performed below expectation. The *Software* category of products established itself as the best-delivered item, solely able to meet the VOC. Specific products that require detailed attention were identified, particularly within the *Mouse* and *Laptop* categories. Item-specific recommendations were made where applicable. A risk meta-analysis was then conducted on the methodological validity of the applied SPC methods to establish the likelihood of Type I and II errors, in order to gauge the trustworthiness of the previously attained results. This process revealed that out of the three tested heuristic devices, Test C was the most reliable. Additionally, erroneous entries in past data sets were also corrected and the effects of these changes on previous results from Section 2 were discussed. Attentions then turned to the optimisation of financial and staffing arrangements for two coffee shop locations. By evaluating historical data on staff numbers and order delivery times, it was possible to determine the optimal number of employees to obtain at each store, as well as the share of orders that each location would reliably service. Appropriate recommendations to improve shortfalls were made where applicable. Part 3's SPC results were further interrogated by designing an experiment to gain further insights into the *Cloud Subscription* product category, which had previously been identified as problematically out of control. A multivariant analysis of variance (MANOVA) test was applied to the picking and delivery times of this product class to establish whether significant chances in either occurred between 2022 and 2023. It was discovered that the worrisome trends were not caused by changes in either of these variables across that timeframe. More likely contributors were discussed along with recommendations in the subsequent interpretation of findings. Finally, the reliability of a car rental agency's service was investigated in similar fashion to the coffee shop scenario, with the aim of determining the number of days in a year that the agency could provide reliable service with its current staffing arrangement, before using an optimisation procedure to find the optimal number of additional employees to hire, which would minimise losses and thereby maximise net profit. All detailed results for each procedure are included and discussed in its relevant section throughout the document. Recommendations and interpretations have been appropriately made where applicable. It is the final conclusion of this report that statistical analysis is absolutely essential to the smooth operation of any commercial enterprise – especially where quality is concerned. The ability to interpret historical data opens up an entirely new world of insights into opportunities and solutions that are otherwise unattainable. And in a world where the scale of competition has steadily grown with the expansive reach of global markets, the benefits of statistics in the pursuit of quality assurance is only likely to grow ever more indispensable.

9. References

Blokhin, A. 2022. *What profit margin is average for a company in the electronics sector?* [Online]. Investopedia. Available at: https://www.investopedia.com/ask/answers/051215/what-profit-margin-average-company-electronics-sector.asp [Accessed 25 Sep. 2025].

OpenAI. 2023. ChatGPT (GPT-4) [large language model]. Available at: www.chat.openai.com [Accessed: 23 October 2025].

Global-imi.com. 2023. *IMI: The Cause of the Digital Downturn*. [Online]. Available at: https://www.global-imi.com/blog/cause-digital-downturn [Accessed 1 Oct. 2025].

Quality Gurus. n.d. Quality Gurus: *Taguchi Loss Function*. [Online]. Available at: https://www.qualitygurus.com/taguchi-loss-function/ [Accessed 24 Oct. 2025].

sixsigmamoneybelt. 2010. *YouTube: What is SPC (Statistical Process Control)?*. [Online]. Available at: https://www.youtube.com/watch?v=0GfBSuwhUwI.

Statista. 2022. *Market Forecast: Consumer Electronics - South Africa*. [Online]. Available at: https://www.statista.com/outlook/cmo/consumer-electronics/south-africa?srsltid=AfmBOopz2cu5WBO-Apmmedhn6JEGi3ZxgTBymwK6ubbw0W9BcSJoglgR [Accessed 1 Oct. 2025].

Statista. 2024. *Statista: US population by gender 2027*. [Online]. Available at: https://www.statista.com/statistics/737923/us-population-by-gender/?srsltid=AfmBOoqk6S87-KiHXuiJZl81iwaHxqINTJQpB0obdJ8DQs3mqkIgqk2w [Accessed 25 Sep. 2025].

Kameleoon. 2024. *Kameleoon: What are Type I and Type II errors?*. [Online]. Available at: https://www.kameleoon.com/blog/what-are-type-i-and-type-ii-errors.