QUALITY ASSURANCE 344

ECSA PROJECT

26072807
JANY SCHNETLER

Table of Contents

Tabel of Figures

1. Introduction

This project forms part of the ECSA GA4 outcome assessment and aims to demonstrate the application of data analysis, process control, and optimisation techniques within an industrial engineering context. Using various real and simulated datasets, the report explores how statistical methods can be applied to understand, monitor, and improve operational performance. The project evaluates process stability, identifies potential inefficiencies, and develops data-driven strategies for improvement. Each section builds on the previous one, moving from data understanding to process evaluation and finally to optimisation and reliability modelling.

2. Descriptive statistics

2.1 Data inspection for each dataset

The dataset Customers contains 5000 observations where each observations has 5 different variables collected. The 5 variables observed is the customers ID shown to be 5000 unique entries to identify each customer, gender that can be one of 3 options naming them as female, male and other, the city where the customer lives - there are 7 different cities, the age and the income range per customers. The dataset has no missing values. The average age of the customers is 56. It ranges from the youngest customer being 16 and the oldest one 105 years old. The average in range of income is R85000 with a large standard deviation of R33150 suggesting that there is a large income diversity among the customers making the purchasing power different among all the customers.

The products dataset also does not have any missing values and consists of 5 variables. The first one also being a product ID with 60 unique entries to make identification easier, the other 4 variables are the 6 different categories which products can fall under, a short description of the product, selling price and the markup per product. There are 60 observations meaning there are 60 different products. The products stretch over a large price range making the standard deviation quite large. The selling prices starts at R350 and stenches all the way to R19725 with an average selling price of R794. These prices ensure the markup of all the products to be between 10% and 30% with an average markup of 20,34% positioning all products within a competitive range.

Sales database displays all the sales observed. Currently there are 100 000 sales with 9 different variables capturing all the different information to trace each sale. The information captured is Customer ID, product ID, Quantity, Order time, Order Day, Order month, Order year, Picking hours and delivery hours. Connecting the customer to the product while capturing the process with different timestamps.

2.2 Data Filtering and Sub setting

Data filtering and data sub setting helped to dive deeper into each dataset, to get more meaningful insights from the large data sets provided.

For instance, we grouped all the product sales using the product ID's to see which items are the most popular and which items are least popular. It is identified that the product with the most sales is a keyboard – MOU059 with sales reaching 29675. And the least popular item is a laptop - LAP021 with sales of 12853. Showing there is a broad demand over all the products. It also shows that the top 20 items sold all sold over 26000 items suggesting they are not defendant on a few top sellers, but the high demand stretches over a broad range.

Order Month	Total Quantity
4	123333
3	122151
11	120569
9	120231
7	120220
10	119392
5	118658
6	116800
2	116799
8	115570
1	85683
12	70941

Order Year	Total Quantity
2022	722141
2023	628206

Tabel 1: Order quantity per month and year

As seen in table 1 on the left there is no clear correlation that can be made to a specific month that has the most sales. Although we can clearly see that the month of December has the least sales. This is probably related to the end of the year and people spending all their money on Christmas and vacation and not new tech products. In the table on the right it is seen that there is a significant drop in sales from 2022 to 2023 which may be concerning, and I would recommend the company investigate it.

2.3 Data Visualisation

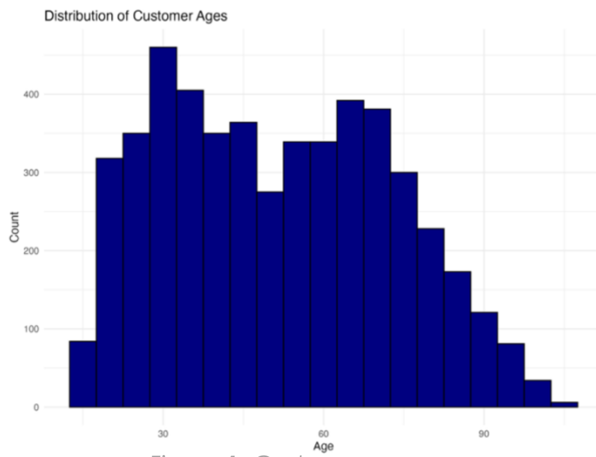

Figure 1: Customer ages

The histogram above displays the distribution of all the customers ages as mentioned above the ages range over a large spectrum from 16 years old to 105 years. The company appeals to wide range of people. But as displayed by the graph we can see that there is a very stable customer base between the ages of 40 to 70. Reflecting that the companies target group is working people needing a variation of technology on a day-to-day basis. This could also reflect the stable income the customers receive at that stage in life, known as someone’s peak performing years.

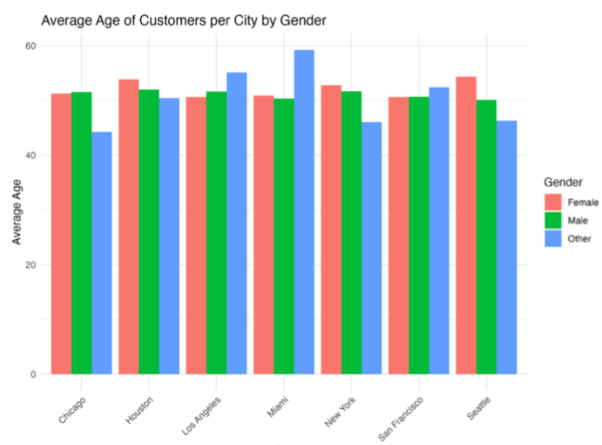

Figure 2: Customer Age per city

As mentioned in the above and seen in the bar plot it is apparent that the average age is between 40 and 60 years. The bar plot displays a stable age demand over all 7 cities. Indicating a stable and well identified target market. There is no conclusion to be made on gender as the bar plot displays it as very equal.

Figure 3: Quantity sold per product category

Using the histogram above to analyse the quantity of products sold per product category we can conclude that there is not one category that really stand out suggesting they are not dependent on a certain category for sales, but the demand starched over all categories. This could also indicate that the companies quality throughout all products are good as people order a similar amount of products from each category.

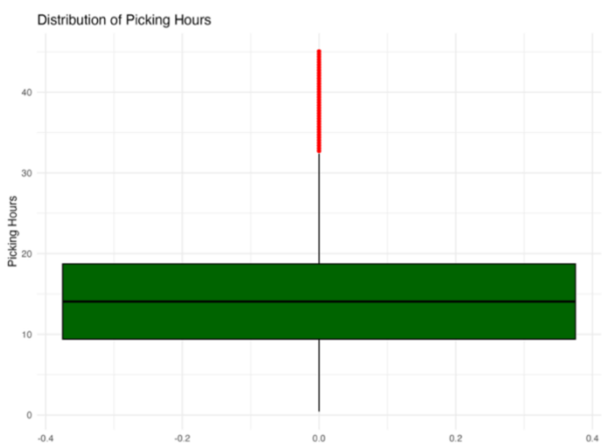

Figure 4: Picking hours

Figure 5: Delivery hours

These two box plots above relate to the picking hours and delivery hours separately. As seen the box plot displaying the picking hours have a consistent picking time showing most of the orders have a picking time between 10-20 hours. However, there are a few outliers identified using the red dots that starts at 30 hours and stretching all the way to 50+ hour. These outliers seem to be extreme scenarios. Scenarios that reflect a crisis like a delay with the warehouse, stock shortage or even a complex order that requires more handling. These outliers could point out a few areas where the company can improve their efficiency to improve the overall consistency and customer satisfaction.

The other box plot shows the distribution of the delivery hours being between 12 and 25 hours making it a reasonable time as one day only consists of 24 hours. As seen, there are no outliers suggesting the companies logistic system is stable.

The differentiation between the 12 and 25 hours are probably due to the travel distance, the traffic or even the delivery going through the night.

Comparing the two boxplots the main source of inefficiency in the order fulfilment process would be the picking time pointing the problem to be in the warehouse or assembly phase rather than the transportation.

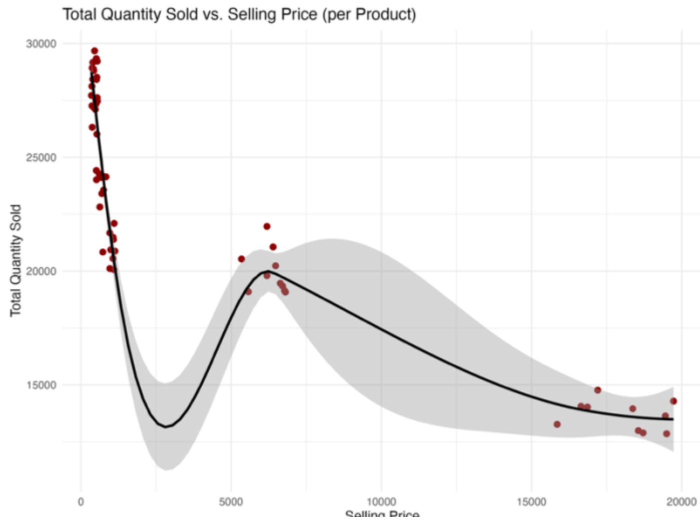

Figure 6: Quantity VS selling price

The scatterplot above has a correlation between the selling price of the product and the quantity sold of each product. The higher the selling price is the less of that product is sold indicting a negative correlation. This is a common trend in the consumer world as affordability is the one of the highest factors that drives a purchasing decision. Relating the scatter plot to the revenue of the company I would say both the low and high selling prices contribute equally to. As the low prices contribute to scale or rather volume. Although the high selling price products are sold less frequent it contributes more to margin. It is important to find a good balance between scale and margin. Affordable products that are fast growing help build a steady revenue. And products that have a higher price generate more revenue per product.

As mentioned above and seen in the line graph below the average markup per product varies around 20-21% showing a stable markup having a good balance between premium pricing and keeping it competitive in the market.

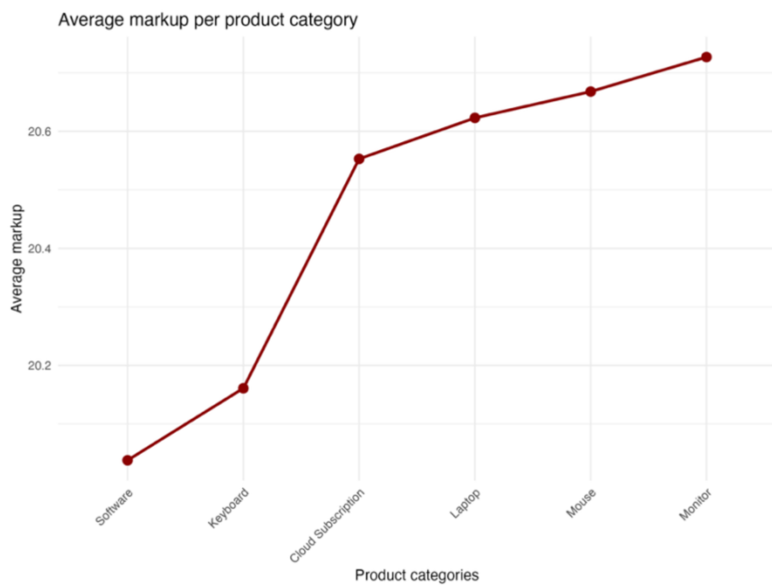

Figure 7: Markup per product

2.4 Recommendation

After the analysis on all four datasets there are a few import things I want to touch on as it provides a comprehensive perspective of the business.

Looking at the customers profiles we can identify that the customers are predominantly mid aged with mid income but the diversity and range of both their ages and income is quite large. A good strategy would be to implement a diverse marketing strategy to hit both the older customers with a more stable and higher income as well as the younger customers with a lower income being more price sensitive.

The product portfolio shows a consistent markup having a good balance between serving the premium market while keeping the competitive advantage by being conscious of the prices. The wide price range of all the products also increases the customers range as it broadens the appeal and reduces reliance.

From the perspective of the sales performance there is no product that can be highlighted as dominant. This reflects and highlights the businesses high stability as the stable demand across multiple products decrease the risk. This identifies a gap for potential growth as the promotion of a top performing product will benefit the entire business.

In terms of the operational level of the company there are some outliers identified in the process specifically in the picking times. To optimise the operations of the company some attention should be given in reducing these bottle necks.

Overall, I would say the business has a stable customer base selling diverse products with a stable demand. If the company is looking for areas to do improvements, I would recommend increasing their operational efficiency and to identify a targeted market to boost a top performing product.

3. Statistical Process Control

3.1 Charts for product Capabilities

The purpose of applying statistical process control charts is to determine the stability of the data.SPC is applied to the sales dataset. The process followed was ordering the data chronologically according to the year, month, day and order time. After the 30 samples initially identified was used to create the control limits and set the base the rest of the data was treated as real-time data to monitor the simulation. Then each product was treated and analysed separately to ensure no data overlaps.

S-charts are mainly used to predict the variability of the data. It is a good indicator showing the consistency of the data, in this case it shows how long orders take per product. An outlier then indicates an inconsistency or a variation, these are the points we are mainly interested by as it could be caused by external factors causing a problem. Figures 8 and 9 below are two examples of s–charts extracted from the data.

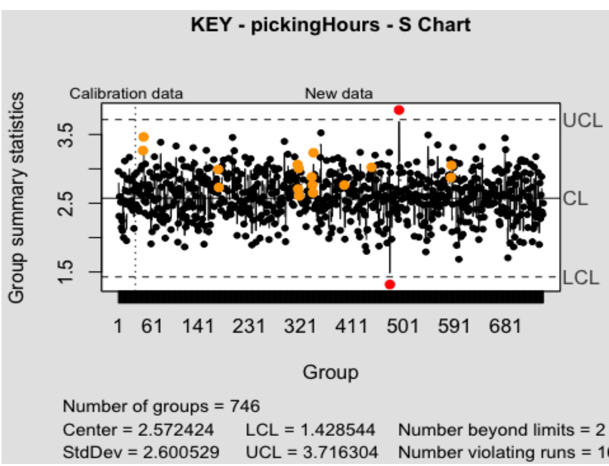

Figure 9: SPC s-chart for keyboard

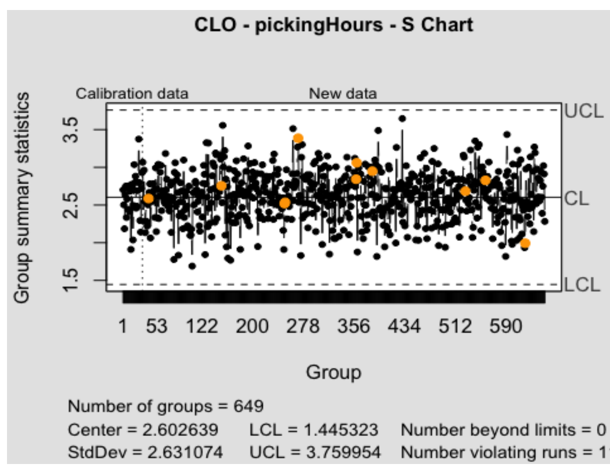

Figure 8: SPC s-chart for cloud computing

X-charts monitors the mean of a process or any data over a prolonged time. It will detect a shift in the central tendency of the data plotting the mean of each process. This indicated any drifts. Any data point outside the control limits indicates a shift in the mean process, it is usually caused by a systematic issue. Figures 10 and 11 below are two examples of x–charts extracted from the data.

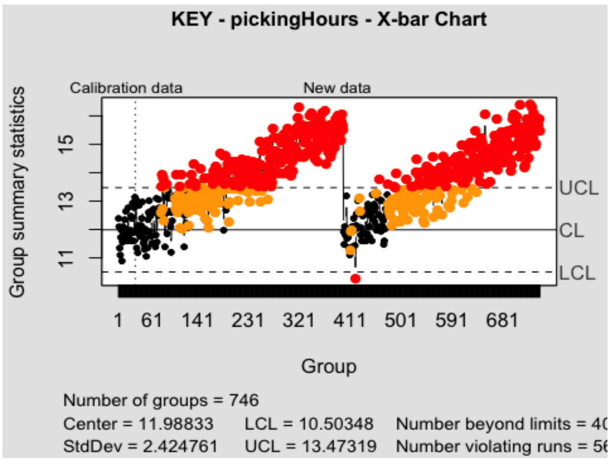

Figure 11: SPC x-chart for keyboard

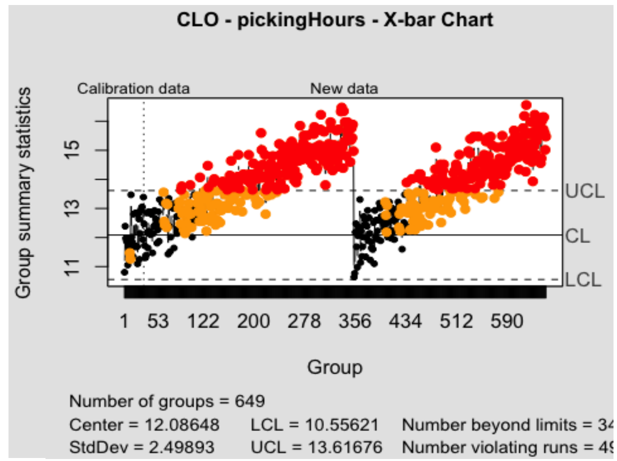

Figure 10: SPC x-chart for cloud computing

Proceeding I will now discuss some interesting behaviour identified in the data.

Table 2 below displays the maximum consecutive in-control sample runs per product type using the delivery times. The more consecutive samples there are per product the more stable the data is.

ProductType	Max_Consec_InControl
MOU	103
KEY	59
SOF	80
CLO	125
LAP	74
MON	63

Tabel 2: Consecutive observation per product

Considering only the data in the table most of the products have a consistent performance as all these values are quite high. Although cloud subscription (CLO) has the highest stability with monitors (MON) being the least stable.

Looking at the x-bar charts there is a clear upward trend starting close to the lower control level (LCL) in the beginning of the year increasing and eventually exceeding the upper control level (UCL) towards the end of year. This could indicate that they are understaffed as the demand increases towards the end of the year or several other challenges can cause this trend. I would recommend having an investigation to determine why this is such a consistent trend over all the products.

All the s-charts only have a maximum of 3 outliers indicating that the process is well controlled and easily predictable.

It is also seen that the picking times have a smaller interval looking at the control limits reflecting a well standardised shop-floor operating system with limited variation. While the delivery times probably have more external factors influencing it.

3.2 Process control issues

The SPC charts were used to analyse and use 3 different rules to identify process control issues. Rula A identifies any samples exceeding the +3σ limit on the S-chart reflecting a variation. Rule B reflects a controlled stability as it measures the longest run of observation on the s-chart that lie between ±1σ. Rule C raises concerns as it gets flagged when 4 consecutive observations are above +2σ limit meaning there was a shift in the process mean. In summary, products showing type A and C rules require immediate attention to try and adjust the process. While the longer rule B runs the better it is.

ProductType	Rule A	Rule B	Rule C
MOU	2	16	252
KEY	0	12	222
SOF	7	7	263
CLO	0	17	272
LAP	0	18	139
MON	4	10	210

Tabel 3: Rule A, B and C

Looking at the table above the software, mouse and monitors should be inspected to see if there is any visible adjustment that can be made to reduce the variation in product control. A concerning observation to make is how high all the products rule C values are as this indicate the number of times there are more than 4 consecutive observation that exceed the +2σ limit. On the positive side all the products also have a relatively high number for rule B indicating a controlled stability for each product.

4. Process capability

In this section we will be analysing the process capabilities of the sales dataset looking at the delivery times of each product. The analysis was performed on the first 1000 observation done per product. The upper and lower limit was assumed to me 0h and 32.

ProductType	mean	sd	Cp	Cpu	Cpl	Cpk	Capable
CLO	19.226	5.94080543101432	0.897745835184293	0.71673783116526	1.07875383920333	0.71673783116526	NO
KEY	19.276	5.81519496599144	0.917137493157815	0.729353591433752	1.10492139488188	0.729353591433752	NO
LAP	19.606	5.93395886974858	0.898781648205004	0.696218734240801	1.10134456216921	0.696218734240801	NO
MON	19.41	5.99891915523384	0.8890490428897	0.699570465623833	1.07852762015557	0.699570465623833	NO
MOU	19.2975	5.82760231130255	0.91518484763269	0.72657097044089	1.10379872482449	0.72657097044089	NO
SOF	0.955375	0.294086775300091	18.1352368799689	35.1876017640502	1.08287199588752	1.08287199588752	NO

Tabel 4: Process capability

The capability index of the various items can be seen by examining the Cp values shown in Table 4. The higher the Cp value the better the potential capability. When a process's Cp is greater than 1.00, it is an indication that it can produce outputs that fall within the parameters of the specification. However, a Cp above 1.67 is considered excellent and only software meets the requirements according to our analysis.

In relation to process variance, the CPU, or upper capability index, indicates how closely the process mean resembles the upper specification limit (USL). A CPU greater than 1.00 means that there is enough room below the USL and that the process can therefore remain inside that range. The process cannot continuously remain below the USL if the CPU is less than 1. Consequently, only software with a CPU of 35.3 can continuously stay below the USL limit out of all products.

The lower capability index, or Cpl, is a metric that quantifies the degree to which the process mean approaches the lower specification limit (USL) in relation to process variation. Because all the products Cpl values are greater than 1, all products remain under the lower specification limit. As the mean may be skewed towards the UCL, this is to be expected since most product types do not satisfy the CPU and would consequently all probably meet a lower limit.

By determining whether the process mean is within the specified bounds, the Cpk, or capability index, gauges a process's true capabilities. A process is capable and generating within specification limits with a safety margin if its Cpk is greater than 1. The procedure cannot reliably achieve the specification constraints if Cpk is less than 1. As a result, we can observe that only software deliveries are competent and meet the requirements.

It's interesting to observe that software has substantially higher CPU and CPL values than the other product categories. Because of automated testing and upgrades, we can presume that software development is less variable than the production of physical devices like keyboards, monitors, and mice. As an alternative, the software process specification boundaries could be far broader than those for the other goods.

Looking at the process capability calculations done and data collected my recommendation would be that none of the products can meet the customers expectation on a consistent base. As the criteria of capability is a Cpk value of greater than 1.3333 and it is seen that none of the products exceed that. The procedures should be modified to try center it on target.

5. Risk, Data correction and Optimising

5.1 Type I error

A Type I error, sometimes referred to as a Manufacturer's error, occurs when a process is labelled as out of control when it actually is under control. This occurs in SPC when a sample point that is in control falls outside the UCL or LCL. A constant derived from the normal distribution's characteristics is the theoretical probability of making a Type I error.

It is assumed that the data is regularly distributed to apply SPC therefor for Rule A (stated in section 3.2) the probability of datapoints landing inside the 3-sigma boundaries is therefore 99.86%, assuming a normal distribution and process control. Consequently, there is a 0.14% chance that datapoints will fall outside of these bounds. This means that on average there will be 1 or 2 false alarms per 1000 observation.

Although the system is in control there is a 7% chance that Rule B may occur. Indicating the probability of this pattern occurring is 0.0691. We would want it to be higher, as the higher Rule B is the more stable the system is.

In contrast, Rule C has a very small probability of about 2.68×10^{-7}, suggesting that it will almost never occur by chance.

In summary, **Rule A** is considered moderately sensitive and could have occasional false alamrs, **Rule B** is more common, being more of a stability indicator than an alarm signal and **Rule C** detects genuine process shifts. Combined, these three rules together reflect the perfect combination between **sensitivity** and **specificity** in Statistical Process Control.

5.2 Type II error

A type II error also known as a consumer's error is the probability of failing to detect that the process has moved. The type II error was calculated for a bottle filling process. The process average and CL of an x-bar chart should be centered on 25.05 liters, with a UCL of 25.089 and an LCL of 25.011 liters.

The process's x-bar standard deviation has changed from 0.013 liters to 0.017 liters, with an average fill volume of 25.03 liters unknowingly. The probability of a type II error is 84.1%. This corresponds to a single subgroup only having a power of 15.9%**.**

The high β means the shifted means largely remain inside the control limits. Implying the manager should not rely on a single subgroup to analyse and make assumptions on the data. This could be solved by increasing the sizes of the subgroups or adjusting the control limits to accept more false alarms by t tightening the control limits.

6. Profit and performance optimisation

The profit optimisation model was applied to two given datasets that includes two different coffee shop's data containing the number of baristas working and the total amount of time they spend serving customers. We calculated the number of baristas per day that yield the highest profit for the company.

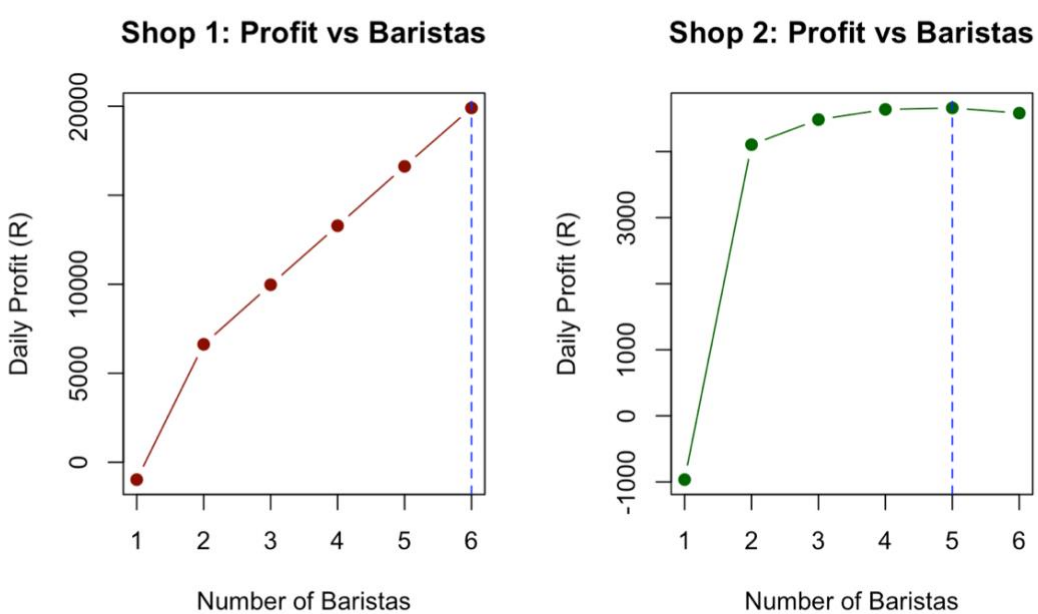

Figure 12: Profit vs barista

The two graphs above display the data collected as seen in the tables below.

Baristas	avg_time	sd_time	total_orders	reliability	customers_per_day	daily_profit
1	200.15587529976	8.01843850278057	417	0.0059737561235878	143.887857185647	-974.21347096082
2	100.170978627672	7.10377259558882	3556	1	287.508422045546	6625.25266136639
3	66.6117433613723	6.26867905811402	12126	1	432.35619647062	9970.68589411861
4	49.9803787749531	5.53279169873209	29305	1	576.226125249629	13286.7837574889
5	39.9618348882736	4.9917982302372	56701	1	720.687628096154	16620.6288428846
6	33.3556463557894	4.5711414216395	97895	1	863.422033343429	19902.6610003029

Tabel 5: Barista reliability for shop 1

As seen in the table 5 and the graph above the profitability on Shop 1 increased steadily from one to four baristas as shorter service times led to higher throughput and improved reliability rising to over 85 %. The maximum daily profit of R 19902.66 occurred at six baristas, after which additional staff produced diminishing returns, as the extra cost outweighed the minor gains in service speed.

Baristas	avg_time	sd_time	total_orders	reliability	customers_per_day	daily_profit
1	200.168943533698	8.37499048240708	2196	0.00801493388225386	143.878463319919	-965.404708882317
2	141.514617902698	7.1809095098571	8859	0.999999958251106	203.512544688509	4105.37608576255
3	115.44091460947	6.23040760039059	19768	1	249.478272910682	4484.34818732045
4	100.015273881379	5.60317978030908	35289	1	287.956017939395	4638.68053818185
5	89.4359692856363	4.98859841643483	54958	1	322.018089925542	4660.54269776625
6	81.6427213987077	4.55017698263998	78930	1	352.756492025214	4582.69476075642

Tabel 6: Barista reliability for shop 2

For Shop 2, which showed slightly longer average service times and higher variability as seen in table 6, profit peaked at five baristas. The maximum daily profit reached R 4660.54 per day. Below this level, service delays and lower reliability significantly reduced total earnings, while employing more than five baristas again resulted in reduced profitability due to overstaffing.

Overall, the analysis confirms that each shop achieves its **optimal efficiency and profit** at a specific staffing level. These configurations balance service reliability, throughput, and labour cost most effectively. Management should prioritise maintaining these staffing levels during peak operating hours while employing flexible scheduling during quieter periods to sustain profitability and service quality.

7. ANOVA and MANOVA

Based on the findings from the Statistical Process Control and Process Capability analyses, the data selected for the ANOVA analysis includes the delivery times of each product type over two consecutive years. This data was chosen because the SPC results showed clear temporal trends, spesifically an upward shift in the X-bar charts across all products. The delivery times was the dependent variable as they directly reflect the operational efficiency and customer satisfaction, while “Year” serves as the independent factor to test for potential changes over time.

The underlying null hypothesis (H_0) states that there is no significant difference in the mean delivery times between the two years, suggesting that the delivery process remained stable. Conversely, the alternative hypothesis (H_1) proposes that there is a significant difference in the mean delivery times between the two years, implying a shift or variation in operational performance. The ANOVA analysis allows for a statistical confirmation of whether the observed process drift in the SPC charts is significant or simply due to random variation. This integration helps provide deeper understanding of the process stability over time. The information below shows the ANOVA analysis done per product

Keyboard

Factor	Df	Sum Sq	Mean Sq	F value	Pr(>F)
factor(orderYear)	1	299.333974229843	299.333974229843	8.07018472559269	0.00450502690350226
Residuals	17918	664602.649458736	37.091341079291	NA	NA

Table 8: ANOVA for keyboard

Cloud subscription

Factor	Df	Sum Sq	Mean Sq	F value	Pr(>F)
Year	1	1.17007381010852	1.17007381010852	0.0312909313826419	0.859595182737097
Residuals	15596	583187.215468296	37.3933839105088	NA	NA

Tabel 8: ANOVA for cloud subscription

Mouse

Factor	Df	Sum Sq	Mean Sq	F value	Pr(>F)
factor(orderYear)	1	19.9395963454687	19.9395963454687	0.529617130834664	0.46677611726124
Residuals	20660	777829.938861978	37.6490773892536	NA	NA

Tabel 7: ANOVA for mouse

Monitors

Factor	Df	Sum Sq	Mean Sq	F value	Pr(>F)
factor(orderYear)	1	16.3617525130567	16.3617525130567	0.44741234554539	0.503576747946227
Residuals	14862	543499.454742649	36.5697385777586	NA	NA

Tabel 9: ANOVA for monitors

Laptop

Factor	Df	Sum Sq	Mean Sq	F value	Pr(>F)
factor(orderYear)	1	18.150088557084	18.150088557084	0.496103644894654	0.481233240516691
Residuals	10205	373352.737136962	36.5852755646215	NA	NA

Tabel 12: ANOVA for laptop

Software

Factor	Df	Sum Sq	Mean Sq	F value	Pr(>F)
factor(orderYear)	1	0.0169500074955467	0.0169500074955467	0.178896950204447	0.672327286520906
Residuals	20747	1965.72275328462	0.0947473250727633	NA	NA

Table 13: ANOVA for software

The degree of freedom for the "Year" factor equals 1, confirming that the analysis is done comparing two distinct years. The residual degrees of freedom represent a variable that is not the year and indicate the size of the dataset used in the analysis. The larger this residual number is the sample size is, which strengthens the reliability of the results.

The sum of squares for both the "Year" factor and the residuals reflects how much each contributes to the total variation in the data. The mean square follows a similar trend to the sum of squares, as it represents the average variation per degree of freedom. A value for the "Year" factor suggests that differences between years explain a meaningful portion of the variance, whereas a much larger residual sum of squares implies that a considerable amount of variability remains unexplained by this single factor. Both the sum of squares and the mean square show similar results. Looking at Table A above it is seen that the keyboard is the only product with a value for the year indicating the different year only really influenced the keyboard sales. It seems that for all the other products the residual values are large suggesting most variations are due to random fluctuations.

The F-values also reflect these findings. The keyboard is the only product with a F-value above 1 indicating that the mean square for the "Year" factor is larger than the mean square of the residuals in the ratio, a higher F-value shows the effect of "Year" on delivery time is substantial relative to random variation.

Finally, the P-value quantifies the probability that the observed differences occurred by chance. A very low P-value confirms that the variation between years is statistically significant. Once again, the keyboard has a very small p-value.

Therefore, for the keyboard the null hypothesis is rejected favouring the alternative hypothesis, suggesting there is a significant difference between the mean delivery times for the product type across the two years. The Mouse, Monitors, Laptop, Software, and Cloud Subscription products reflect consistency, suggesting the year-to-year variation is minimal and it as a stable performance.

8. Reliability of Service

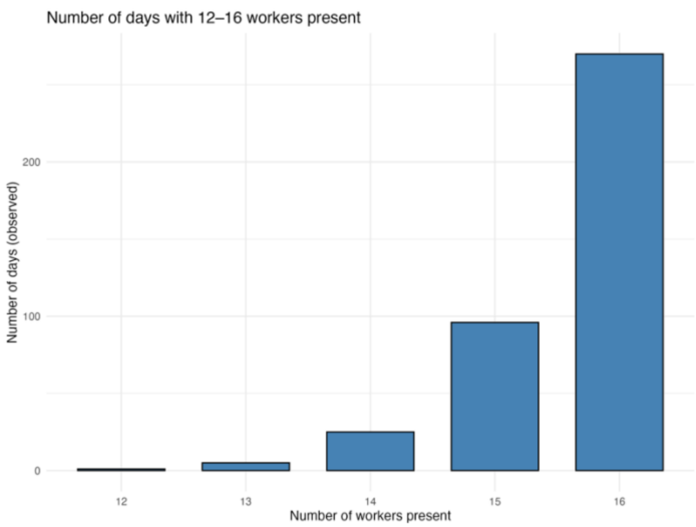

Figure 14: Days and waiters

Figure 13: Cost vs waiters

Figures 13 and 14 above, together with Table A below, illustrates the data collected as well as the model used to determine the reliability of a service provided by a car rental company. The histogram represents the observed number of days with the amount of workers present on those specific days. The company experiences operational problems when there are fewer than 15 workers on duty, leading to a daily profit loss of R20 000. Using this information, the probability of reliable service was estimated using a binomial model, where reliability corresponds to having 15 or more staff members present. The expected mean attendance rate was used to estimate the attendance probability (*q*), forming the foundation for further simulation and optimisation.

q	m	total_cost_year	p_reliable
0.924023929471033	18	5714599.86744831	0.956904127746807
0.942767947103275	18	5525484.34529572	0.98281036365812
0.961511964735516	17	5288844.62655909	0.974130873074097
0.980255982367758	16	5084250.32422597	0.9610615994211
0.999	15	4608736.81155739	0.985104546362002

Table 14: Reliability

Table 14 summarises the sensitivity analysis for different attendance probabilities (*q*), showing the optimal number of scheduled staff members(*m*), the total expected annual cost, and the corresponding probability of reliability (*p_reliable*). As seen in the table, increasing the number of scheduled staff generally improves reliability, but at the cost of higher personnel expenses. The total annual cost therefore depends on finding a balance between the cost of additional staff and the financial losses caused by unreliable service.

Figure 13 displays the total monthly cost as a function of the number of scheduled workers. The curve clearly shows a minimum point, indicating the most cost-efficient staffing level. According to the model, scheduling 17 workers yields the lowest total annual cost while maintaining a high reliability level. Fewer workers result in frequent service failures and lost revenue, while scheduling significantly more than 17 workers leads to unnecessary staff costs that outweigh the benefits of improved reliability. This trend is consistent with the Taguchi Loss Function, which states that any deviation from the optimal operating point leads to an increase in loss.

It is recommended that the company maintains 17 scheduled employees per day to achieve the best balance between cost efficiency and reliability. This staffing level minimises the overall annual cost while ensuring a consistently reliable service.

9. Conclusion

This report successfully applied a range of statistical and analytical methods to evaluate operational efficiency, product performance, and service reliability.

Descriptive statistics were first used to explore the customer, product, and sales datasets, identifying key trends, variation, and outliers that provided a foundation for deeper analysis.

Statistical Process Control (SPC) was then applied to the delivery data to establish control limits and monitor process performance. It was found that most product types operated within control, while the keyboard product type showed variation that warrants further attention.

Process capability analysis revealed that only the software product type met the required performance standards, while the other products fell below the capability threshold, indicating that further process improvement is needed.

Type I and Type II errors were calculated to evaluate the risks of incorrectly classifying a process as in or out of control, providing important insight into decision reliability. Profit optimisation for the coffee shop model showed that efficiency and profitability could be improved through optimal staffing decisions.

An ANOVA analysis comparing delivery performance across two years confirmed a statistically significant difference for the keyboard product type, while other products remained stable, highlighting consistent process control.

Finally, the staffing optimisation model for the car rental company identified that scheduling 16 workers per day provides the most cost-effective balance between reliability and operational cost.

Overall, this report demonstrated how integrated statistical methods can be used to analyse, optimise, and improve real-world processes. The findings provide practical guidance for maintaining process stability, enhancing product quality, and maximising operational efficiency across different business contexts.

References

Dirkse-van-Schalkwyk, T. 2025. SPC. South Africa: Stellenbosch University.

ECSA Brief (2025) ECSA brief.pdf (Accessed: 22 October 2024).

OpenAI (2025) ChatGPT (October 2025 version) [Computer program]. Available at: https://chat.openai.com/ (Accessed: 22 October 2025).