Quality Assurance

ECSA Project

Sybrand de Waal
US No 27271684
24 October 2025

Contents

Table of Figures

Introduction

This report aims to answer the objectives set out in the ECSA project brief. The project consists of three main sections each corresponding to the weekly tasks set out in the project brief. R studio was used for data processing and plotting throughout the project.

Four data sources were provided for initial data analysis. The data was customer -, product -, product_Head_office - and sales data. The data was analysed to attempt to identify any useful information contained within the data set.

Client Information

The business provided the details of 5 000 clients with each clients customer ID, gender, age, income and city. A sample of the data is shown below.

	Customer ID	Gender	Age	Income	City
1	CUST001	Male	16	65000	New York
2	CUST002	Female	31	20000	Houston
3	CUST003	Male	29	10000	Chicago

Client Gender

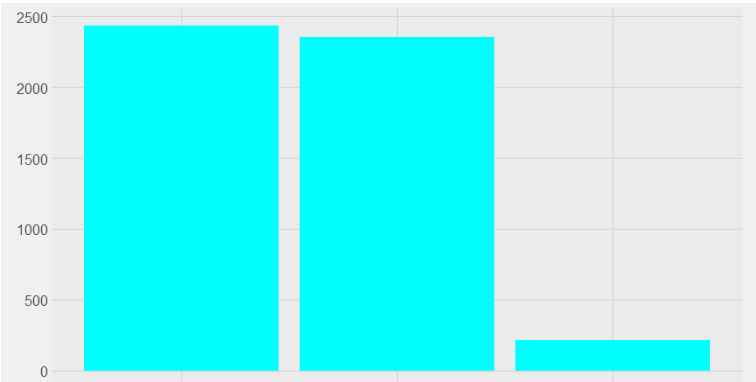


Figure 1: Gender distribution

The distribution of registered clients can be seen in the bar graph on the right. (Figure 1) From the data it can be seen that the distribution between male and female clients are very similar, with there being slightly more females registrations.

Three are a notable number of customers registered as other, this could be due to data collection errors or customers not included in the main categories.

Due to the relatively even split between male and females and low presence of other relative to male and female the Client gender does not provide any insight into the data.

Income distribution by gender

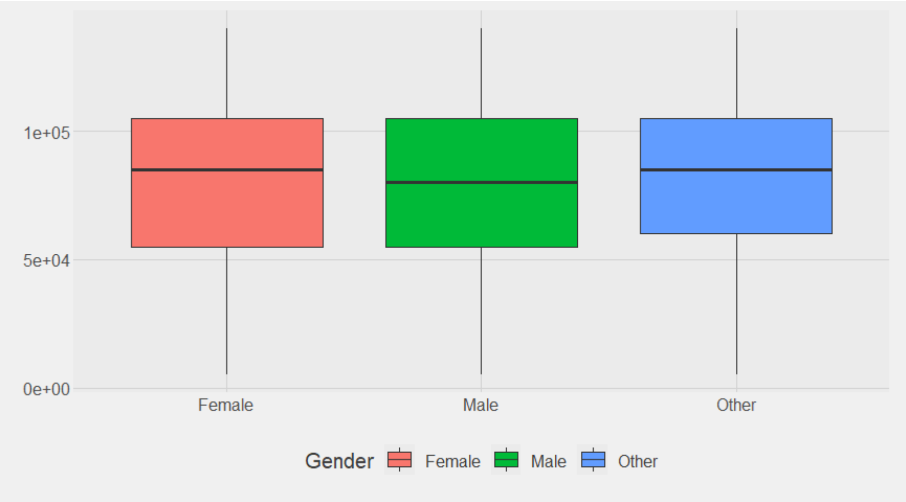


Figure 2: Income Distribution (Gender)

Figure 2 shows the income spread of registered clients at the company split according to the gender that each customer had indicated.

It is important to take note that the gender choice "other" is significantly smaller than that of "Male" and "Female".

From the figure it can be seen that the average income for female customers is slightly higher than that of male customers. The first and third quartiles and 1.5 IQR spread for all three categories are similar to each other, suggesting that the low and high tail for both data sets are similar with the only difference in the mean of the data set.

This is supported by Figure 3 which shows a similar income distribution for each gender. The peak income is higher for Female clients compared to that of other or male clients. Male clients have the lowest income distribution.

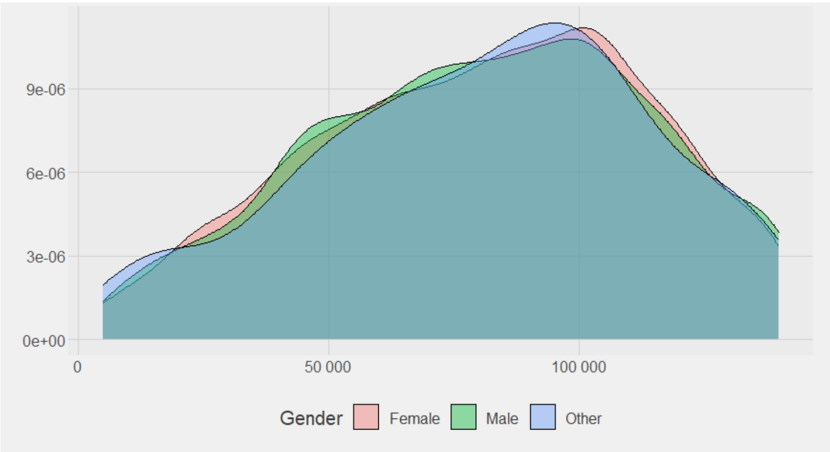


Figure 3: Gendered Income distribution

From this it can be concluded that gender does not affect the income of the registered clients. The income distribution of female, male and other is similar enough that there is no statistically significant difference.

Client Age

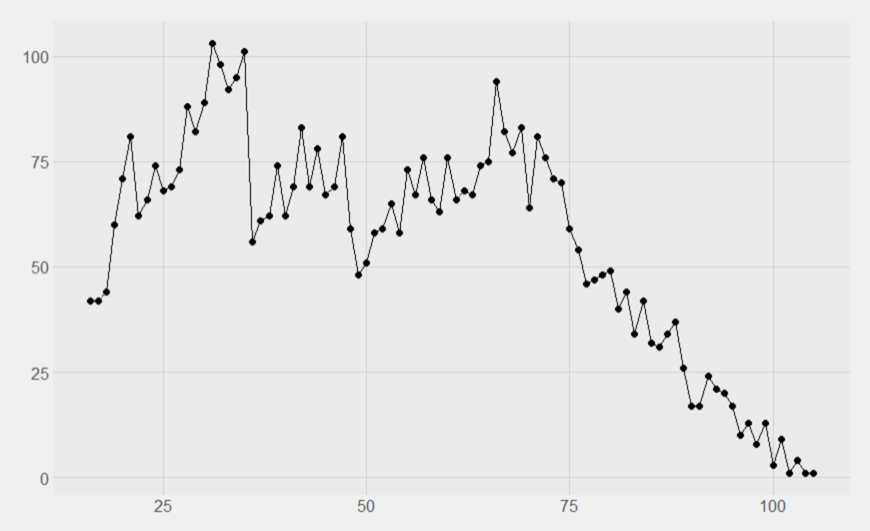


Figure 4: Distribution of Client Age

From Figure 4 it can be seen that the majority of clients are young adults with the peak age for registered clients being 30-35 years. From 75 year and up the number of customers decreases linearly with the oldest registered customer having an age of just over 100 years.

It can be gathered from the data that a majority of users are young to middle aged.

Figure 5 shows the average income for each age. There is a notable increase in average income between young and middle aged (40-60 years) customers.

The average income decreases again after 60 years. This can be attributed to the retirement age of 60 causing a decrease in income for customers.

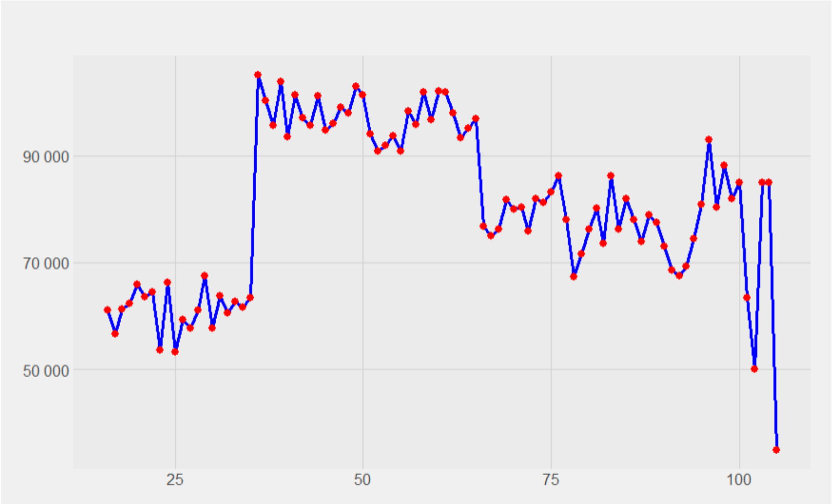


Figure 5: Average Income over Age

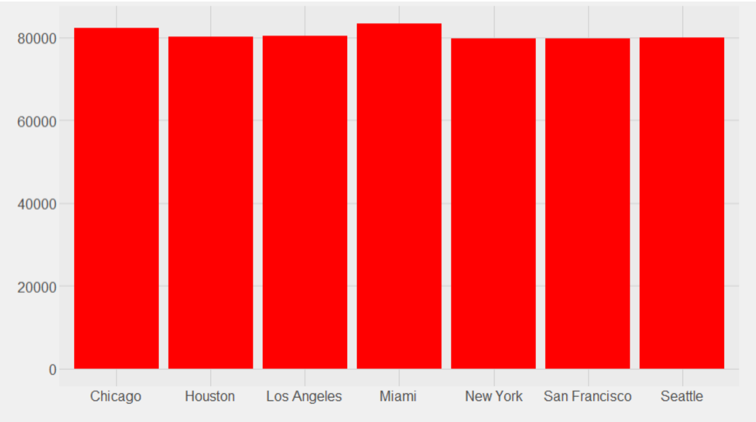


Figure 6: Average Income per city

The average income for each cities customers is shown in Figure 6.

The average income for towns is very similar to each other. Miami and Chicago are outlier cities with slightly higher average incomes relative to the other cities.

This demonstrates that the clients city does not play a meaningful role in in predicting income and can thus be disregarded.

Product Information

Two data sets were provided for product information, that of products_data and products_Headoffice with 60 and 360 entries respectively. Each entry has a Product ID, category, description, selling price and markup. The profit made on each product was further calculated and added to the data set.

Selling price per category

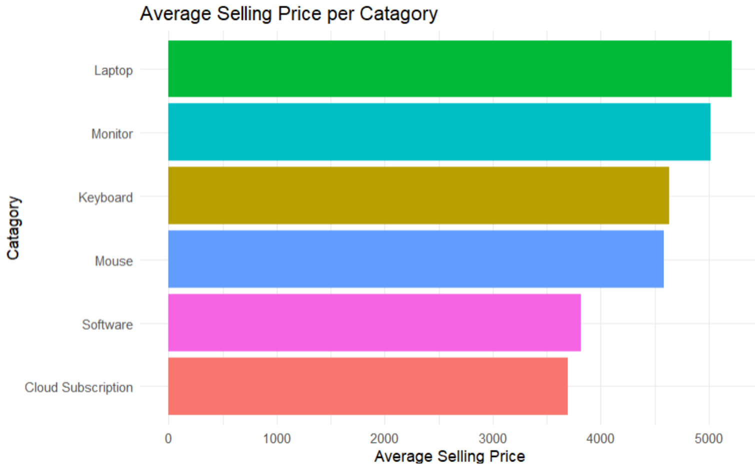


Figure 7: Average Selling Price

Figure 7 shows the average selling price of each product category. The prices of each product category vary from 3600 to 5200 with an even distribution between each extreme.

Cloud Subscriptions and Software have the lowest and 2nd lowest average selling price.

In Figure 8 the average selling price per product category can be seen for the Head Office data set.

The average price per product category varies significantly less for the head office data set with all categories being between 4300 and 4500.

The order of selling price categories has changed and laptops has the lowest selling price in Figure 8 compared to the highest in Figure 7.

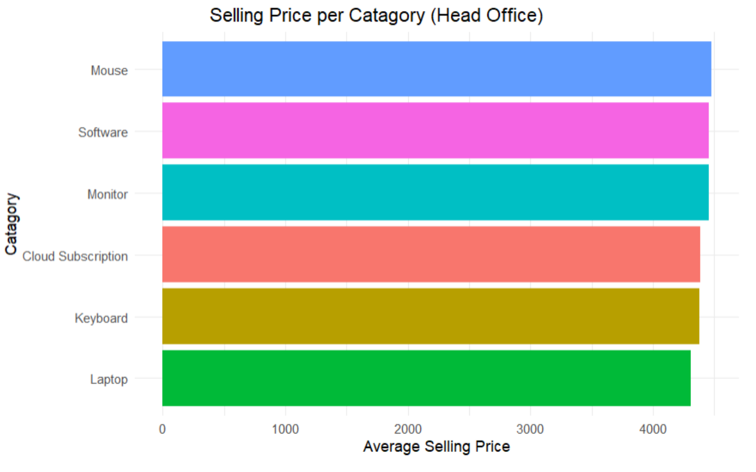


Figure 8: Average Selling Price (Head Office)

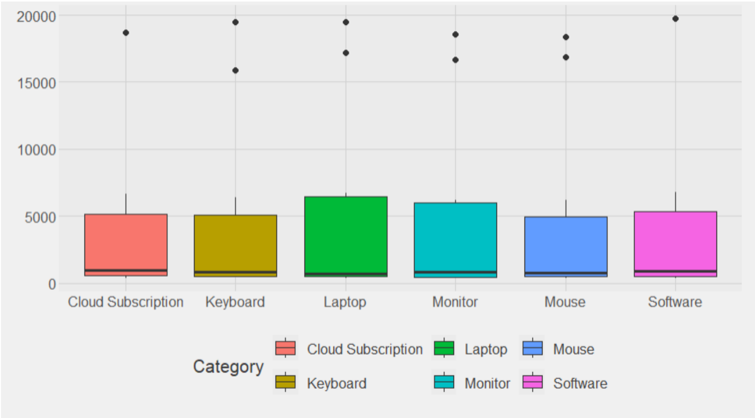


Figure 9: Selling Price per Category

From Figure 9 and Figure 10 the difference discussed can be seen with the upper quartile varying significantly in Figure 9 whilst the spread in Figure 10 is significantly less spread out.

There are however many more extreme outliers is Figure 10 with head office product prices. This could be due to higher price products only being sold though the head office and not standard stores.

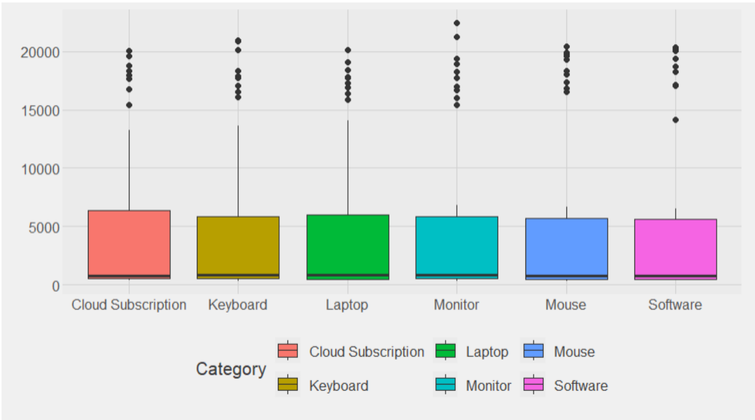
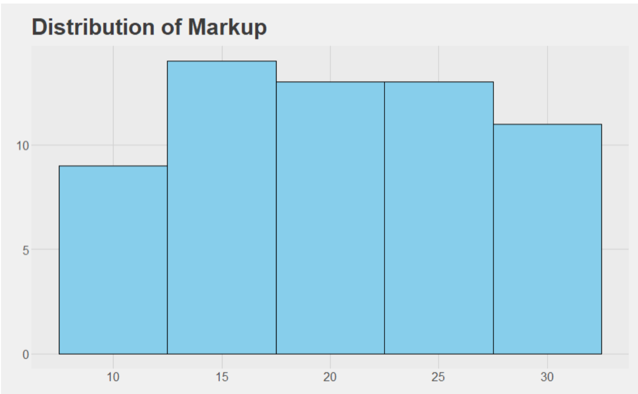


Figure 10: Selling Price per Category (Head Office)

Markup Distribution



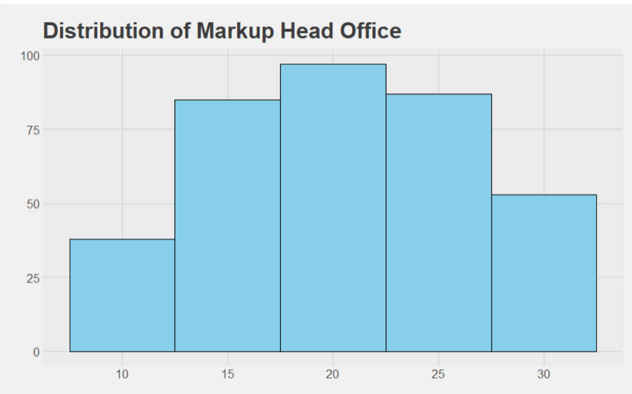


Figure 11: Markup Distribution

Figure 12: Markup Distribution (Head Office)

From the two figures above it can be seen that the markup distribution at the head office (Figure 8) is normally distributed with relatively equal high- and low-end distributions. The standard markup distribution (Figure 11) is significantly more uniformly distributed compared to the head office markup

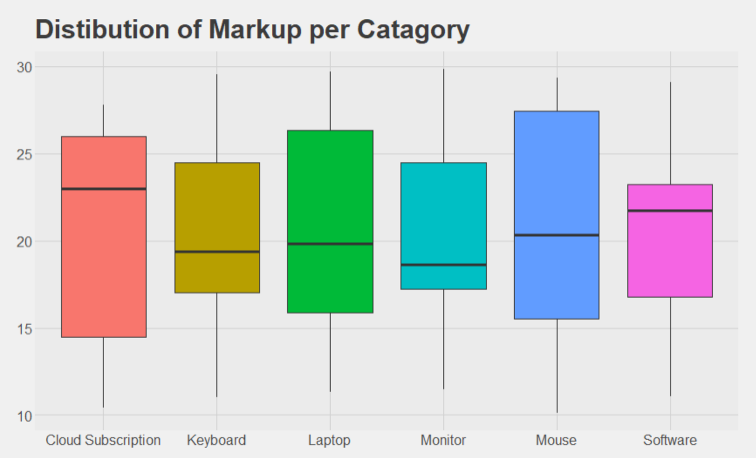


Figure 13: Markup per Category

In Figure 13 the markup distribution can be seen per category. There is significant variation in the mean as well as inter quartile range and total spread between categories.

This would suggest that there is no clear correlation between category and markup.

The distribution of markups for categories in the head office shown in Figure 14 is more evenly distributed. The mean markup as well as total spread is similar for the different categories.

The first and third quartile does still however vary between categories.

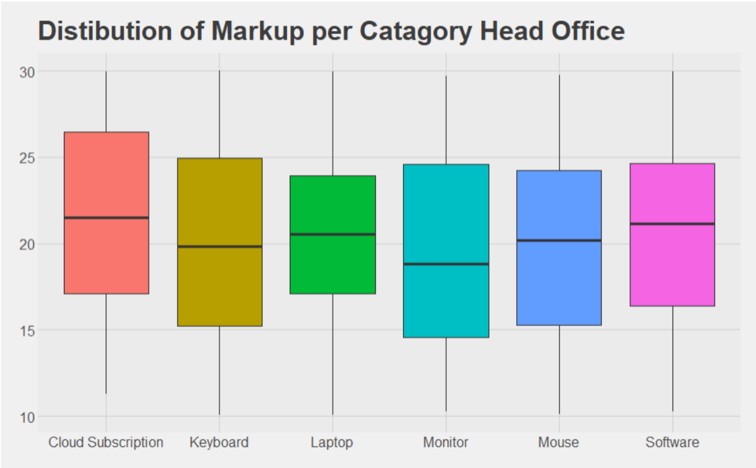


Figure 14: Markup per Category

Profit on Products

The profit the company makes per product was calculated by taking the selling price and multiplying it with the markup in a percentage form. Profit can offer a unique insight into what makes the company money, as a high sales price and low markup or a low sales price and high markup can result in similar profits to the company.

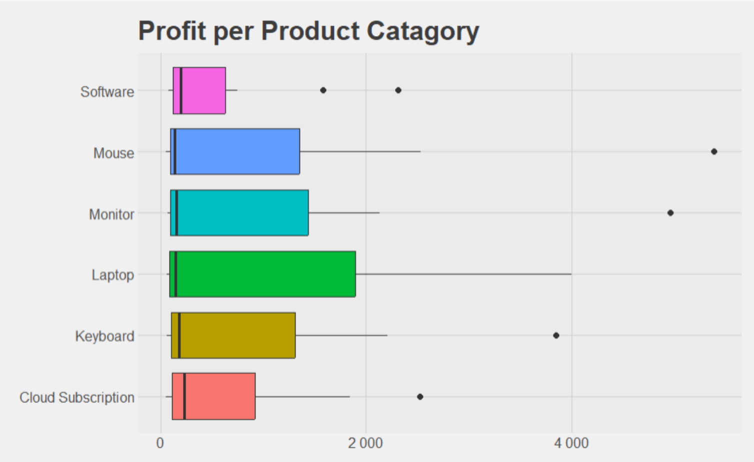


Figure 15: Profit per Category

From Figure 15 it can be seen that the profit that is made on products in the various categories varies dramatically with software and customer subscriptions having significantly lower 3rd quartiles compared to the other products.

Furthermore, there are also low amounts of outliers.

When the profit is compared to the head office products it can be seen that the average profit is significantly higher in the head office.

Both the 3rd quartile and spread are significantly higher in the head office. Furthermore there are also significant outliers in the head office profit compared to the normal products profit.

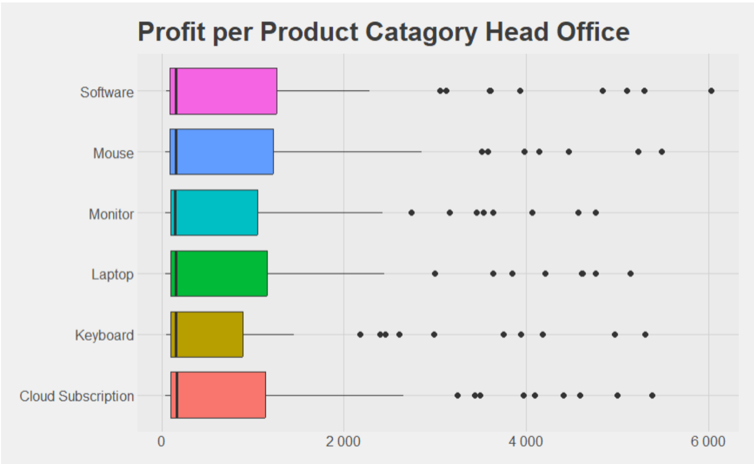


Figure 16: Profit per Category (Head Office)

Statistical process control limits

Future sales data was provided for 2026 and 2027 in the form of a csv file. This file contains sales data for the coming two years. Notable data is that of sales quantity, order time, - day, -month, -year, picking and delivery hours.

The order time information was used to create the absolute day when orders was placed. This is the day which the order was placed with the 1st of January of year 1 being day 1 and the hours expressed as a decimal.

The data was further divided into its product categories and sampled with a sample size of 24. The average (x bar) and standard deviation (s) was calculated for each of the product category samples.

Initial Control Charts

The initial control chart limits were calculated using the first 30 samples for each of the product categories. The one, two and three sigma upper and lower control levels (UCL & LCL) was calculated for each product category. The charts are illustrated below

Keyboard initial control charts

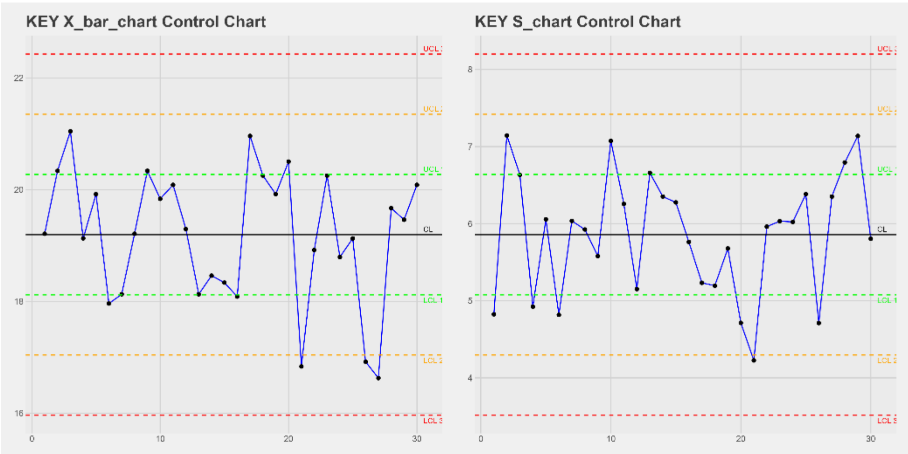


Figure 17: Keyboard Initial Control Charts

Laptop initial control charts

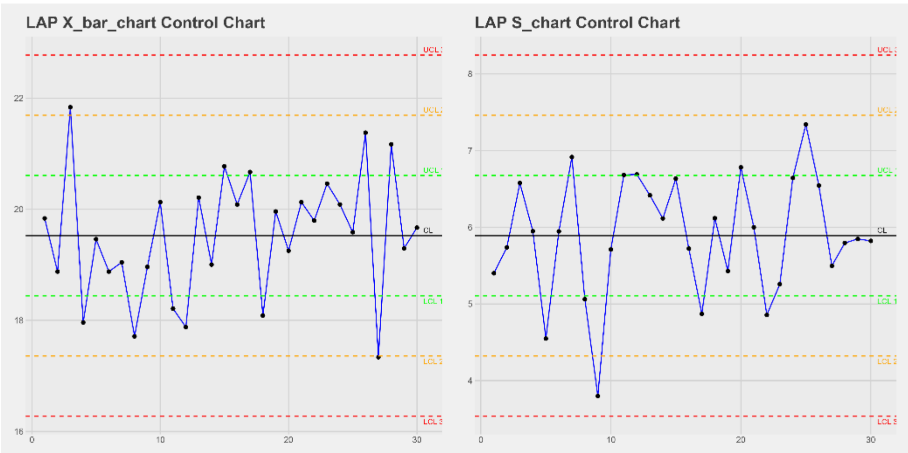


Figure 18: Laptop Initial Control Charts

Monitor initial control charts

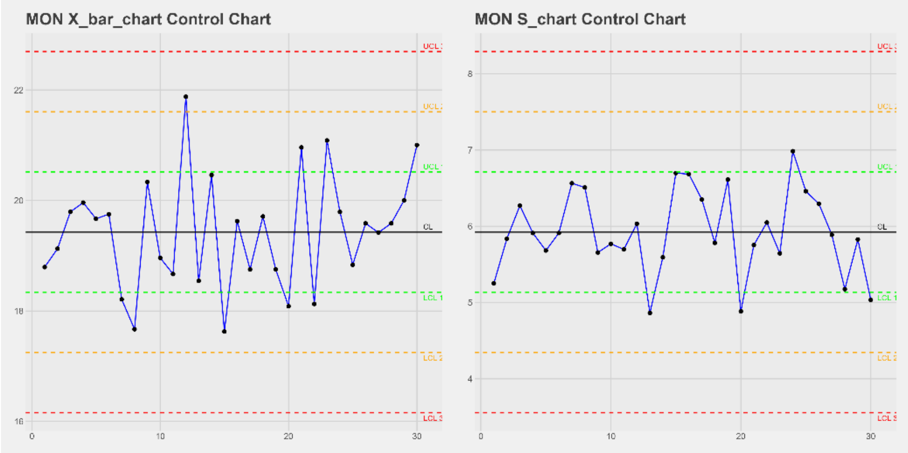


Figure 19: Monitor Initial Control Charts

Software initial control charts

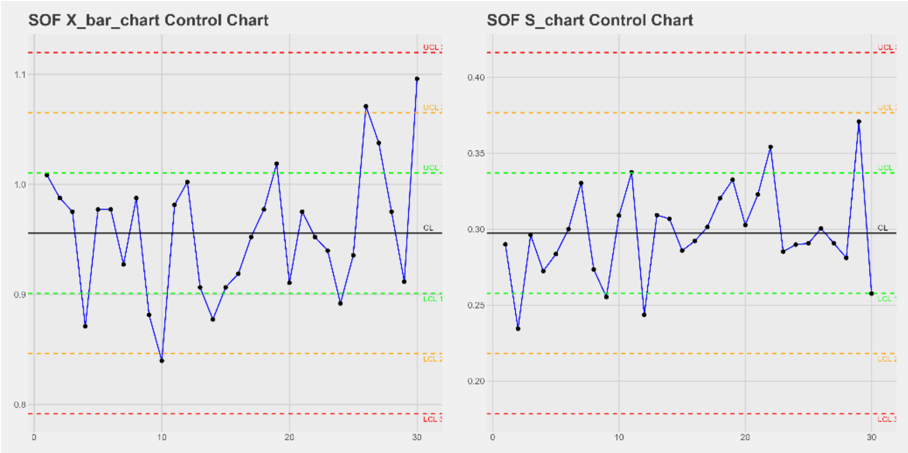


Figure 20: Software Initial Control Charts

Mouse initial control charts

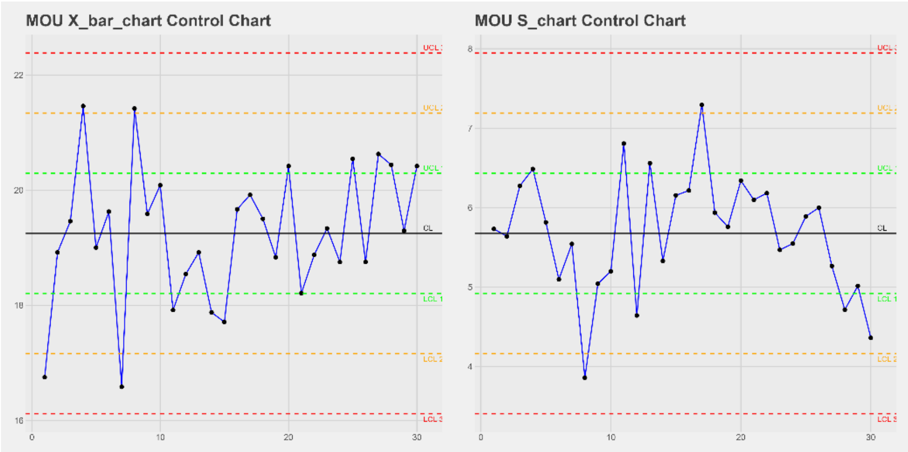


Figure 21: Mouse Initial Control Charts

Cloud subscription initial control charts

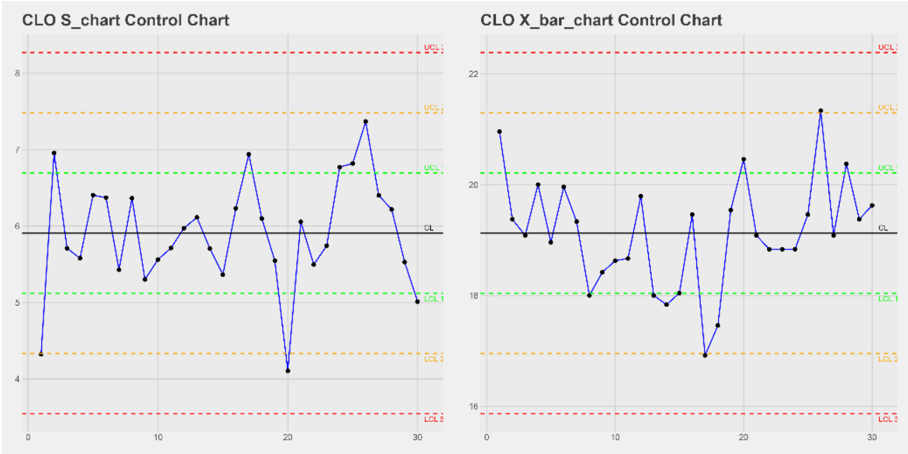


Figure 22: Cloud Subscription Initial Control Charts

Continued Control Charts

Groups of twenty-four (24) samples was added at a time to the control charts until all data was included in the control charts. The control charts with all data added was added below. From the graphs it can be seen that there are dramatic upward trends in the X_bar graphs for all products.

Keyboard continued control charts

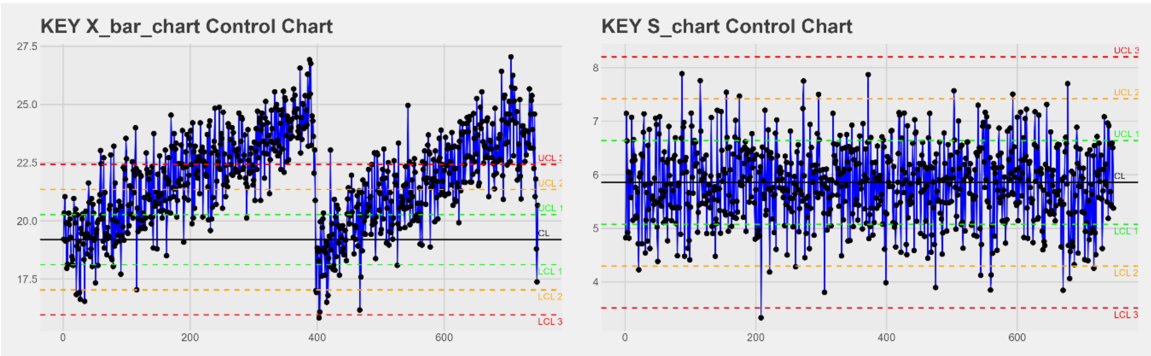


Figure 23: Keyboard Continued Control Charts

Laptop continued control charts

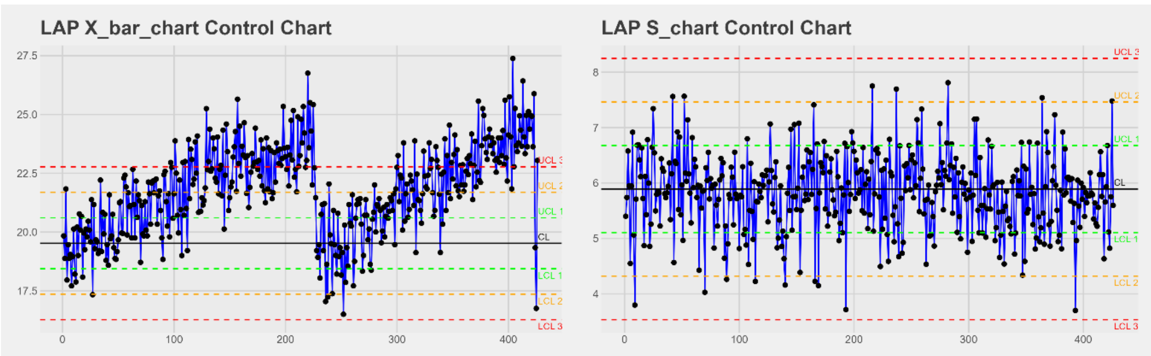


Figure 24: Laptop Continued Control Charts

Monitor continued control charts

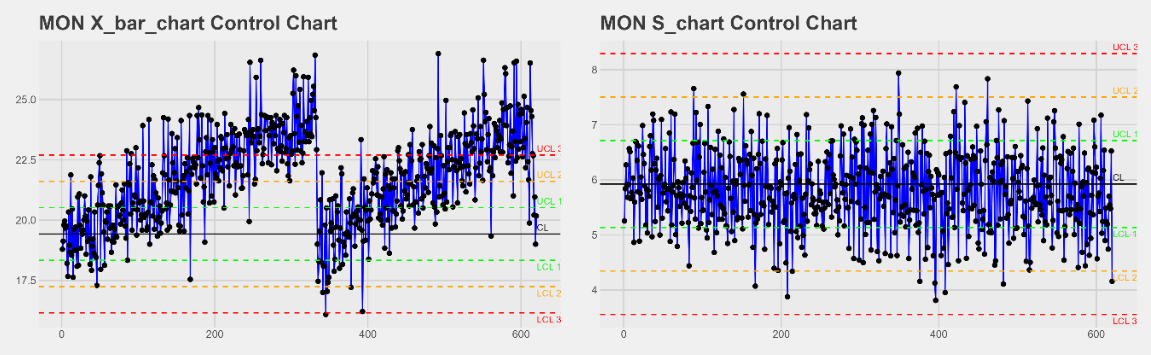


Figure 25: Monitor Continued Control Charts

Software continued control charts

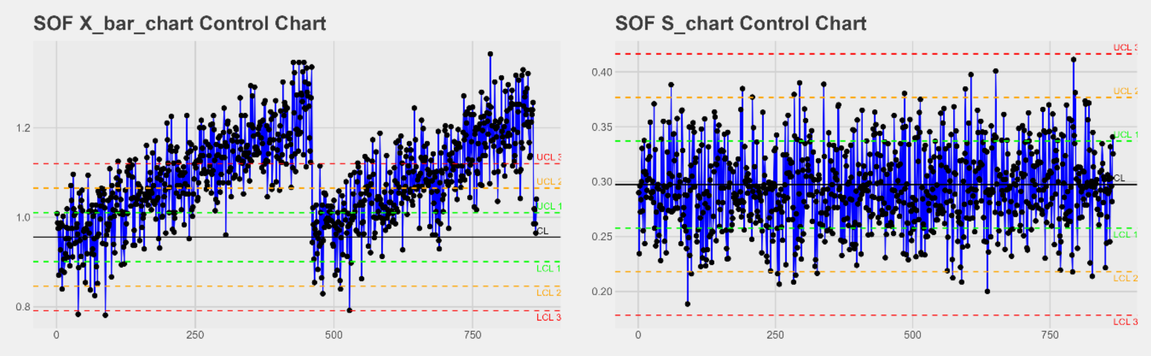


Figure 26: Software Continued Control Charts

Mouse continued control charts

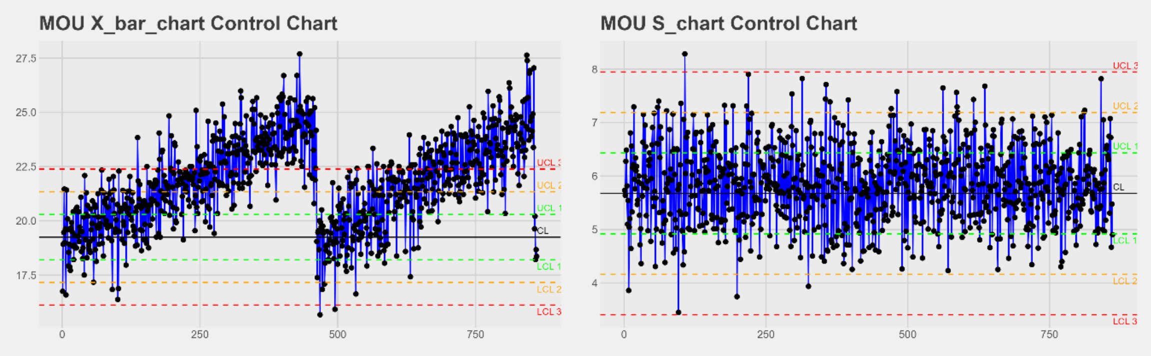


Figure 27: Mouse Continued Control Charts

Cloud subscription continued control charts

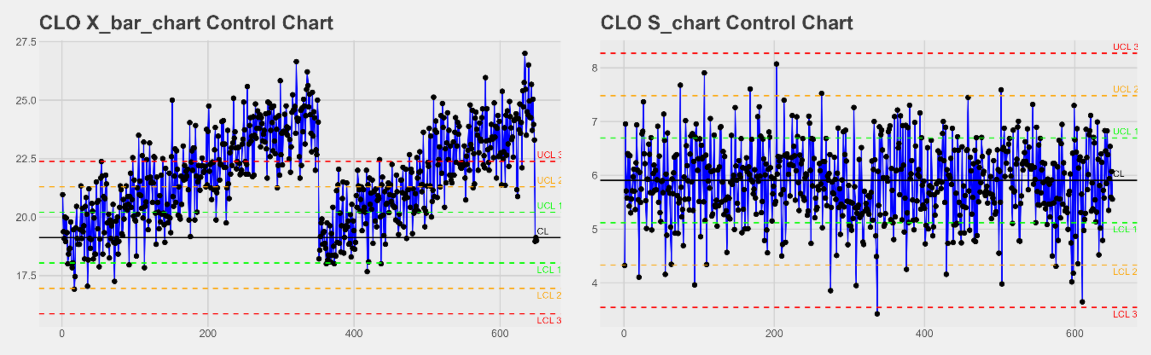


Figure 28: Cloud Subscription Continued Control Charts

Process checking

Throughout the production process there was trends and outliers in the sample averages and standard deviations. At times during the production process there were significant outliers, and these were identified.

Due to the large quantity of outliers, especially on the sample averages the trends and the start and end of the outliers were identified but not all outliers were mentioned.

Cloud subscription

The cloud subscription process drifted upwards from the initial starting point. The following is the first three samples that were outside of the upper control level.

	X_bar	S	Out Of Control Reason
105	23.50233	6.338147	X_bar > UCL3
112	23.21067	5.895220	X_bar > UCL3
120	23.16900	6.748993	X_bar > UCL3

The process was reinitiated after sample 351 and returned to the normal range. The process did however drift upwards again until it went out of bounds again. The first three instances are shown below.

	X_bar	S	Out Of Control Reason
490	22.79600	5.643388	X_bar > UCL3
495	23.62933	5.323587	X_bar > UCL3
503	23.79600	3.980933	X_bar > UCL3

There was high variation in the sample data for cloud subscriptions but mostly remained within the boundaries of the 3 sigma control limits. The upward trend in the delivery hours could be attributed to various factors. High sales volumes or client onboarding issues can cause a gradual increase in delivery hours.

Laptop

The laptop delivery hours samples drifted upward form its initial starting position. Its started exceeding its three-sigma upper control limit at sample 101.

	X_bar	S	Out Of Control Reason
101	23.87917	5.779055	X_bar > UCL3
113	23.93983	6.023467	X_bar > UCL3
115	23.52500	5.665177	X_bar > UCL3

This continued until the process was recalibrated, the upward trend did however continue after the recalibration. The last out of specification sample was sample 225 and the samples stayed withing the UCL3 until sample 301.

The standard deviation of the samples remained within the 3 sigma boundaries throughout. The standard deviation shows that there is random variation between samples but that there is no trend in standard deviation in the data.

Keyboard

The standard deviation of delivery hours for the keyboard samples mostly remained within the three sigma control levels with a few outliers. Many of the samples standard deviation did however exceed the single sigma control levels.

Similar to the other products keyboards experienced an upward trend in delivery hours. The first outlier above the three-sigma control line was sample 59 and continued to increase from there.

	X_bar	S	Out Of Control Reason
59	23.02317	5.106431	X_bar > UCL3
75	22.83567	5.384997	X_bar > UCL3
80	23.21067	4.488310	X_bar > UCL3

The process was readjusted after sample 396 where the average returned to the control line. The upward trend did however continue, and the first outlier occurred at sample 421 with repeating outlier occurring more frequently from sample 490.

	X_bar	S	Out Of Control Reason
394	23.62917	5.076972	X_bar > UCL3
396	23.04817	6.519774	X_bar > UCL3
421	23.04600	5.485158	X_bar > UCL3
490	22.92100	5.479706	X_bar > UCL3

Monitor

The monitor product category group's samples have lower variation within its standard deviation across all samples. The sample average has a positive trend from where it starts within the control limits. The first sample that is out of bounds is sample 106 and samples are out of bound more frequently after sample 106 as demonstrated in the table below.

	X_bar	S	Out Of Control Reason
106	23.91900	4.945903	X_bar > UCL3
114	24.16900	5.777110	X_bar > UCL3
133	24.25233	5.849408	X_bar > UCL3
139	24.18983	5.416318	X_bar > UCL3

Midway through the samples the process was readjusted. This can be seen after sample 332 where many samples were inbounds where the first out of bound samples were sample 391 and 455. The strong upward trend did however continue, and the samples were soon out of bound again.

	X_bar	S	Out Of Control Reason
331	26.84117	5.740604	X_bar > UCL3
332	24.25233	4.662330	X_bar > UCL3
391	23.33767	5.853124	X_bar > UCL3
455	23.12933	5.098309	X_bar > UCL3

Mouse

The mouse product categories sample standard deviation was low with most samples within the three sigma control boundaries and few outliers. The product category does however demonstrate the same strong upward positive trend within its samples that the other product categories demonstrate. The first out of bound sample can be seen below was sample 137 after which sample 176 and 177 were out of bounds as the trend continued upward.

	X_bar	S	Out Of Control Reason
137	23.83567	5.717853	X_bar > UCL3
176	24.04400	6.136137	X_bar > UCL3
177	22.83567	7.321672	X_bar > UCL3

The process was adjusted after sample 461 with the averages returning to the control line. Sample 468 was the first out of bound sample but was below the lower control limit. The first sample above the upper control limit was sample 564 with repeating frequency after sample 630.

	X_bar	S	Out Of Control Reason
461	24.16983	5.266878	X_bar > UCL3
468	15.67100	5.870431	X_bar < LCL3
564	23.33767	5.583198	X_bar > UCL3
630	23.96267	5.315414	X_bar > UCL3
648	23.04600	6.527534	X_bar > UCL3

Software

Software is a digital product and as such a lower standard deviation would be expected, but the standard deviation is similar to other products. Software has a similar upward trend that the other product categories have. The first outlier is sample 162 with the following outliers shown below.

	X_bar	S	Out Of Control Reason
162	1.160533	0.2296816	X_bar > UCL3
173	1.143867	0.3091597	X_bar > UCL3
178	1.170950	0.3156472	X_bar > UCL3
199	1.210533	0.2895524	X_bar > UCL3

The process continued with an upward trend and was recalibrated after sample 460. As such the sample average returned to the initial value. The upward trend did however continue and as such the samples deviated again after the adjustment with time. The first out of bound sample was sample 570 with 576 following after.

	X_bar	S	Out Of Control Reason
459	1.266783	0.3115456	X_bar > UCL3
460	1.335533	0.2627888	X_bar > UCL3
570	1.173133	0.3429148	X_bar > UCL3
576	1.146050	0.3079782	X_bar > UCL3

Process capability

The process capably was calculated for each of the six product categories. This compromised of the Cp, Cpu, Cpl and Cpk statistics for each product category. It is however important to note that the Cpl is of less importance due to the one-sided nature of delivery hours.

The process has the hard limit that delivery cannot occur before a customer has ordered the product and as such the Cpl is of less importance than the Cpu.

	Cloud	Laptop	Keyboard	Monitor	Mouse	Software
Cp	0.8977458	0.8950269	0.9171375	0.8890490	0.9151848	18.135237
Cpu	0.7167378	0.6928906	0.7293536	0.6995705	0.7265710	35.187602
Cpl	1.0787538	1.0971631	1.1049214	1.0785276	1.1037987	1.082872
Cpk	0.7167378	0.6928906	0.7293536	0.6995705	0.7265710	1.082872

All physical products (Cloud, Laptop, Keyboard, Monitor & Mouse) has low Cpu, Cpl and Cpk values, especially when compared to software which is the only digital product category.

A Cpk value below one (1) indicates that the process is barely capable of meeting the constraints imposed of LSL = 0 hours and USL = 32 hours (Schalkwyk, 2025). All physical products Cpk is significantly lower than one which indicates the processes is barely capable of adhering to the constraints.

There is low variation in the Cpk of the physical products, which would indicate that the low process capability is not inherent to a specific physical product but to all physical products. The Cpu is lower for all physical products than their respective Cpl figures and as such, the one-sided nature of the delivery hours does not impact the Cpk figures of the various physical products.

The software product category has a low Cpk, but it is important to note its high Cpu. Due to the one-sided nature of delivery hours we can place more emphasis on the Cpu than the Cpk. The high Cpu indicates that the software delivery is very capable of adhering to the LSL and USL set out.

Process control issues

Standard deviation outside 3 sigma control lines

The outliers were identified for each product group. Outliers are samples that has an standard deviation outside of the upper or lower three sigma control levels.

Cloud

No standard deviation outliers in data set.

Keyboard

No standard deviation outliers in data set.

Mouse

Sample Position	Value (S)	Fault
107	8.285551	S above UCL3 (+3σ)

Monitor

No standard deviation outliers in data set.

Software

No standard deviation outliers in data set

Laptop

No standard deviation outliers in data set

Most consecutive samples

Product Category	**Start**	**End**	**Length**
Cloud	477	498	22
Monitor	236	253	18
Mouse	455	466	12
Software	829	850	22
Laptop	395	417	23
Keyboard	652	666	15

4 consecutive X-bar samples

Cloud

The cloud product category has 15 instances of at least 4 consecutive X_bar samples above the 2-sigma upper control line.

Index	**Starting Position**	**Ending Position**
1	120	123
2	165	172
3	177	181
13	554	562
14	564	622
15	625	647

Laptop

The laptop product category has 10 instances of at least 4 consecutive X_bar samples above the 2-sigma upper control line.

Index	**Starting Position**	**Ending Position**
1	115	121
2	128	139
3	151	164
8	349	355
9	357	367
10	369	423

Software

The software product category has 30 instances of at least 4 consecutive X_bar samples above the 2-sigma upper control line.

Index	Starting Position	Ending Position
1	129	133
2	198	202
3	204	208
28	758	794
29	796	835
30	837	859

Monitor

The monitor product category has 23 instances of at least 4 consecutive X_bar samples above the 2-sigma upper control line.

Index	Starting Position	Ending Position
1	132	135
2	169	175
3	177	184
21	562	589
22	591	610
23	612	616

Mouse

The mouse product category has 29 instances of at least 4 consecutive X_bar samples above the 2-sigma upper control line.

Index	Starting Position	Ending Position
1	197	200
2	209	216
3	221	225
27	784	797
28	799	803
29	805	856

Keyboard

The keyboard product category has 27 instances of at least 4 consecutive X_bar samples above the 2-sigma upper control line.

Index	Starting Position	Ending Position
1	97	100
2	170	173
3	175	178
25	671	692
26	694	720
27	722	742

Risk, Data correction & optimising for maximum profit

The second deliverables work is contained in this chapter. It entails the risks associated with type I and II errors, data correction and cafe employee optimisation for maximum profit.

Type I (Manufacturer's) Error

A type I error or false positive is when the null hypothesis is rejected for a population for which the null hypothesis is true (Banerjee, 2025). The probability of making a type I error for each of the rules were calculated and is shown in the table below.

Process	**Chance of Error**
A	0.001349898
B	0.6826895
C	2.678772e-07

Type II (Consumer's) Errors

Type II errors or false negatives occurs when it is failed to reject the null hypothesis for a sample for which the null hypothesis is actually false (Banerjee, 2025).

Given that the filling processes average fill volume (x_bar) and its standard deviation has changed without our knowledge the chance of making a type II error was calculated and is given in the table below (Schalkwyk, 2025).

Chance of Error
0.8411783

Data Correction & Interoperation

The product data was corrected for its incorrect product categorisation. The head office data was also corrected with its product ID and price being corrected.

Product Data

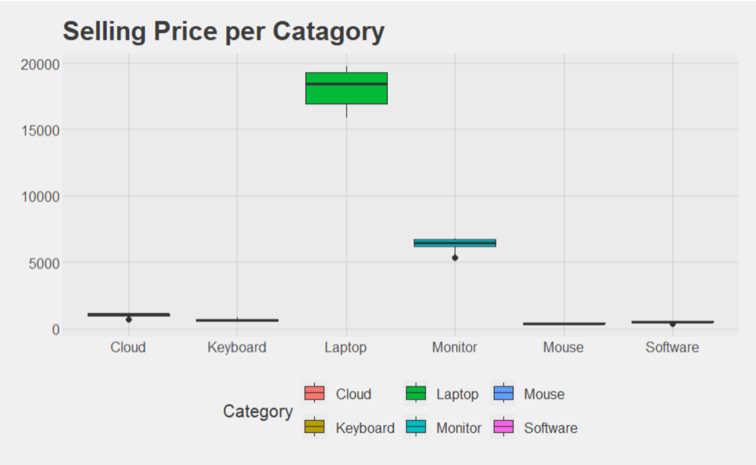


Figure 29: Corrected Selling Price per Category

After the correction of the product category it can be seen that there are dramatic differences between the selling prices of various products.

Both Figure 29 and Figure 31 clearly demonstrates the selling price difference between categories after the corrections has been made.

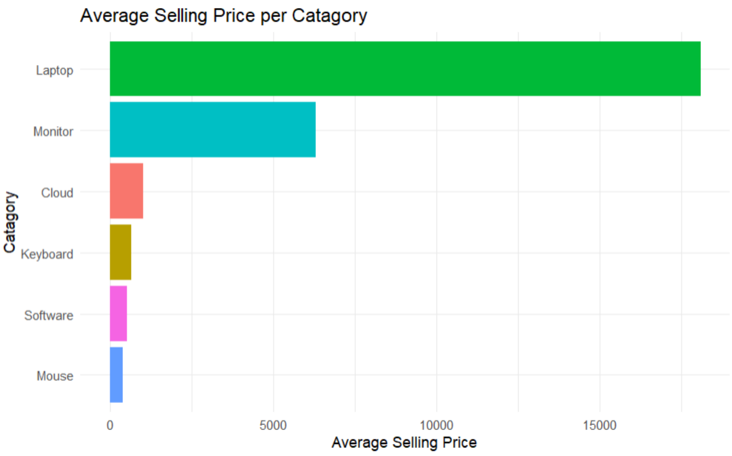


Figure 31: Corrected Average Selling price per Category

Laptops have the highest product selling price and the largest value spread. Monitors have the second highest selling price out of the six categories.

The average selling price before correction was a uniform distribution whilst after correction it is much closer to an exponential distribution.

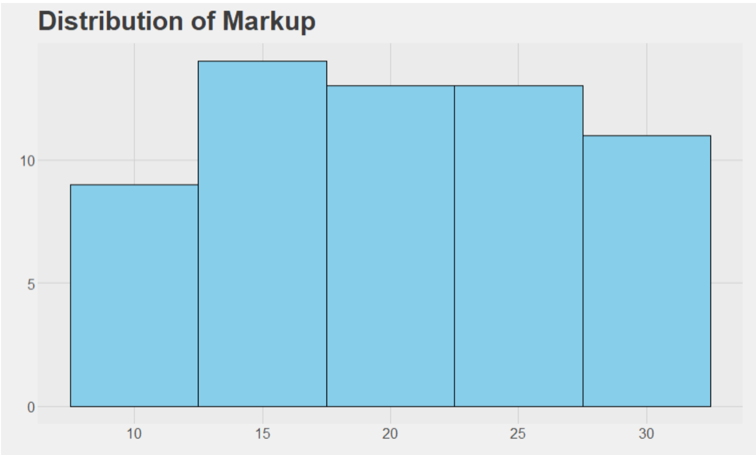


Figure 30: Corrected Distribution of Markup

It can be seen that the correction of product categories did not have an impact on the distribution of markups. As the markups were not adjusted but only the product categories.

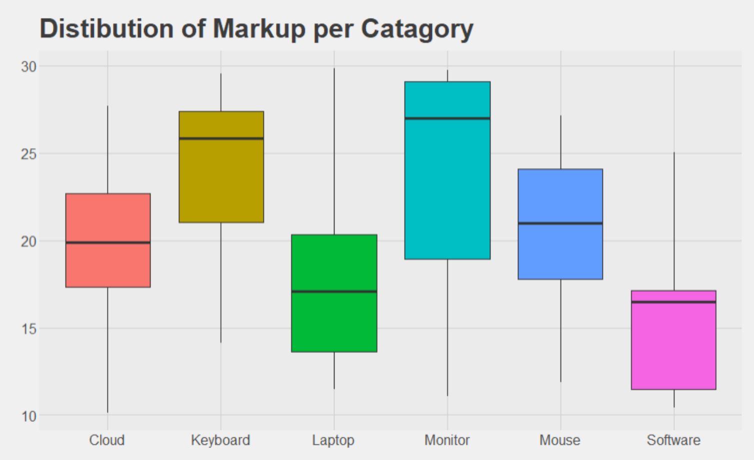


Figure 33: Corrected Distribution of Markup per Category

The distribution of markups per product category has significantly changed after the data correction. Previously the markup distribution was similar for all product categories, with more variation in markup after correction.

All product categories have a large spread of markup. Monitors have the highest average markup followed by keyboard. Software has the lowest markup and the lowest 3rd quartile.

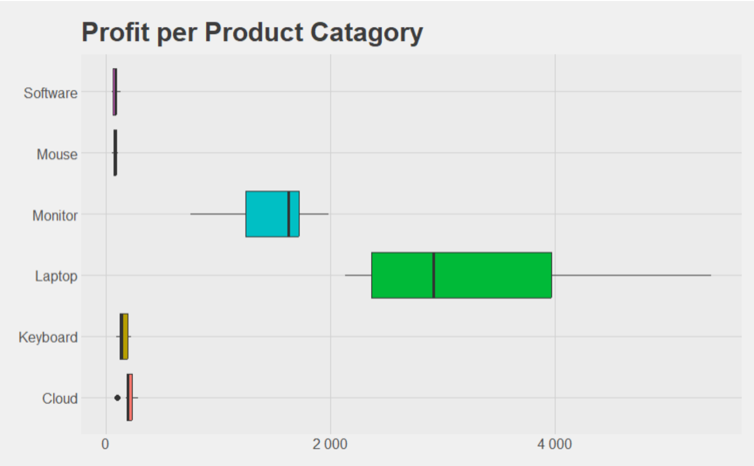


Figure 32: Profit per Product Category

The profit made on the various product categories vary dramatically with Laptops experiencing the highest average and absolute profit out of all categories. This can be attributed to their high selling price as it counteracts the lower markup compared to other product categories.

Monitors have the second highest profit with all other product categories having similar profits.

Head office product Data

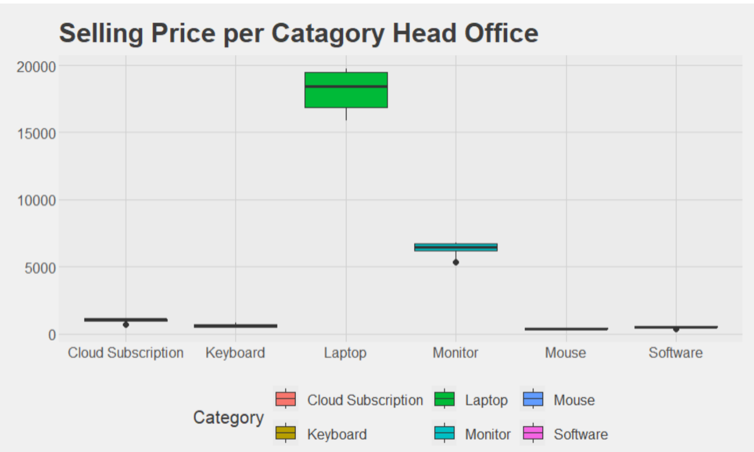


Figure 34: Corrected Selling Price (Head office)

After the corrections were made to the data both Figure 34 and Figure 36 shows the differences in selling price for products in a product category and the category average.

The selling price distribution for head office looks similar to that of product data with similar spreads and averages.

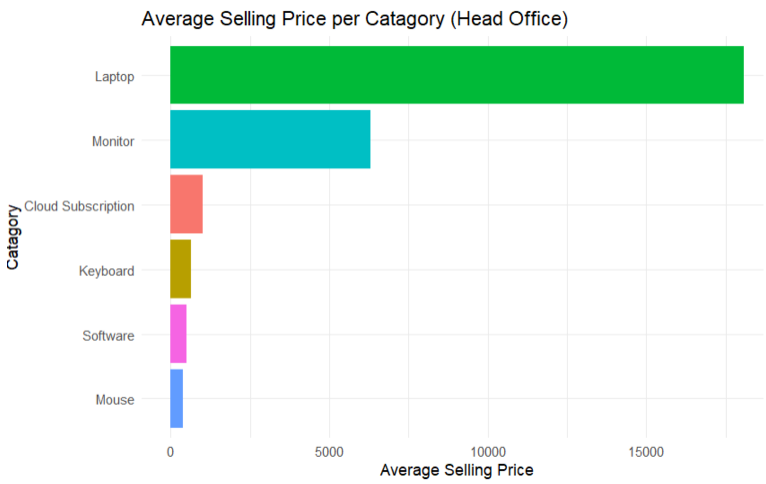


Figure 36: Corrected Average Selling Price (Head Office)

The head office data before correction was similarly evenly spread with the correction resulting in an exponential distribution of average selling price.

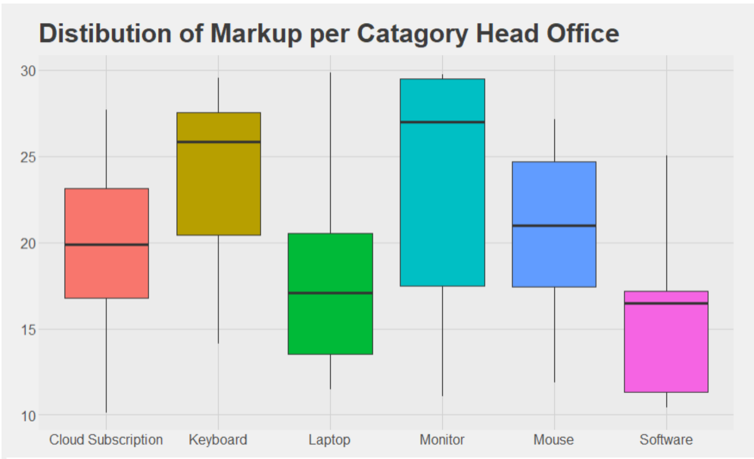


Figure 35: Corrected Distribution of Markup (Head Office)

After the correction of the head office data the distribution of markup per category is now similar between the normal and head office.

Monitors have the highest average markup per product category followed by Keyboards and Mouses.

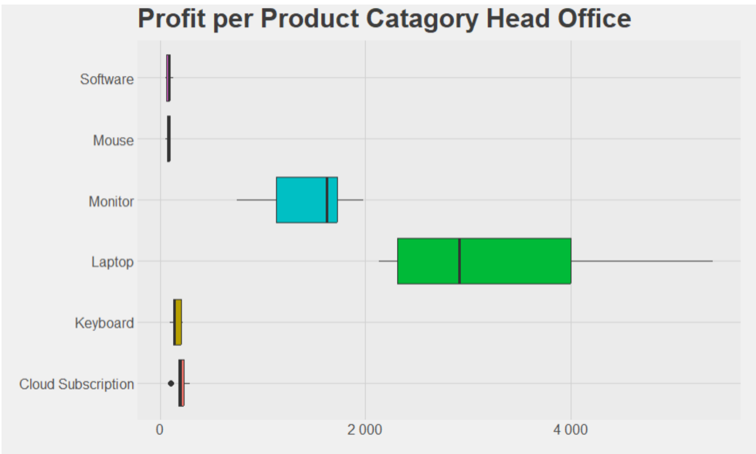


Figure 37: Corrected Profit per Category (Head Office)

The profit per product category is also similar to that of the product information after the correction of the data.

Laptop's has the highest profit per product category with monitors as the second most profitable product category.

Cloud subscriptions are the only product category with an outlier.

Both profit and markup distribution has significantly changed after the correction in data.

Sales Value

Using the updated sales figures the 2023 sales data was analysed. The sales data was analysed per product category to return the average sales value and the total sales volume.

Product type	**Average sales value**	**Total Sales volume**
Keyboard	R 644.66	R 5 378 598.87
Monitor	R 6 310.53	R 43 126 707.9
Mouse	R 394.70	R 3 773 413.87
Laptop	R 18 086.43	R 86 027 413.33
Software	R 506.18	R 4 867 780.65
Cloud	R 1 019.06	R 7 261 887.10

The total sales volume given in the table above is plotted in the figure below to better demonstrate the contribution of each product category to the total sales volume.

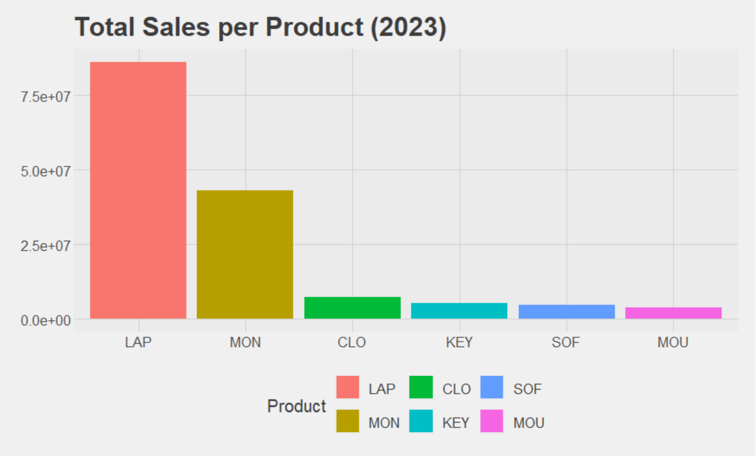


Figure 38: Ordered Total Sales per product (2023)

From the data it can be seen that Laptop sales contribute the most to the total sales of the company followed by Monitor sales. The total sales per product category is similar to an exponential distribution with the contribution of the keyboard, software and mouse product categories being much smaller than that of the laptop and monitor product categories.

Coffee Shop Profit Optimisation

For the optimisation of the number of baristas a coffee shop should employ there were a few assumptions that needed to be made. It was assumed that the coffee shop is open from 8am to 5pm without a lunch break. Furthermore it was assumed that there is at the scale of the data, virtually unlimited demand with demand not decreasing as coffees sold increased.

Coffee Shop 1

The variation of customer waiting time for varying numbers of baristas working has been plotted below. From this it can be seen that there is an exponential decrease in the time gained by adding a barista.

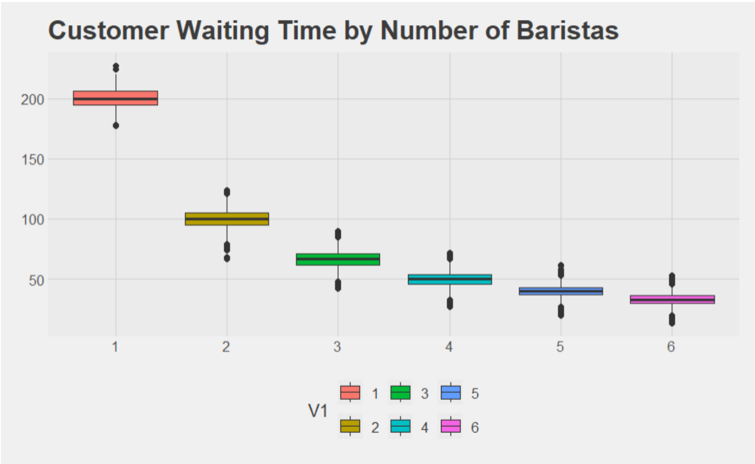


Figure 39: Boxplot of waiting time with varying barista quantities

The quantity of people serviced per day was shown below. It can be seen from the graph that it exhibits a linear increase in customers serviced compared to the number of baristas.

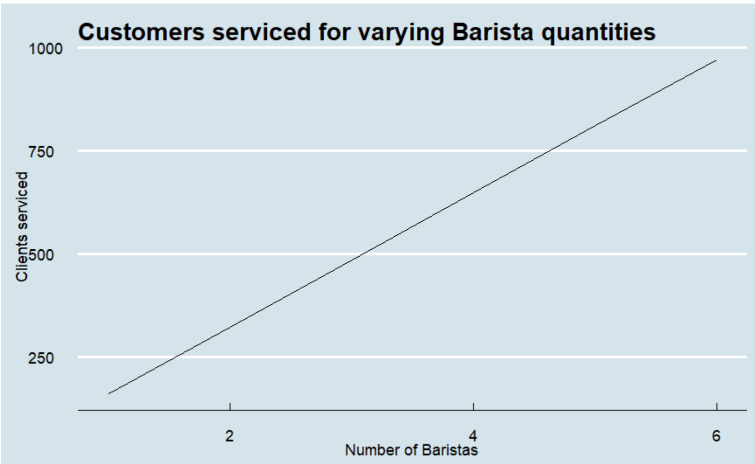


Figure 40: Customers serviced

The profit made for each of the barista is shown it the graph below. From this it can be seen that there is a linear increase in the profit made as more baristas are added. This would make sense as both the profit and barista costs are linear graphs, and the profit graph is the subtraction of the two graphs.

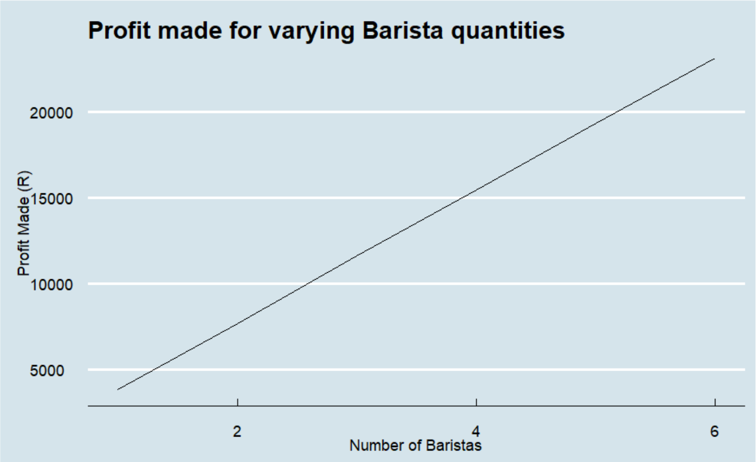


Figure 41: Profit made

As such the ideal number of baristas are six with a profit of R23 140.49 per day.

Baristas	Probability
1	0.0000000
2	0.0000000
3	0.1646050
4	0.9722914
5	0.9999647
6	1.0000000

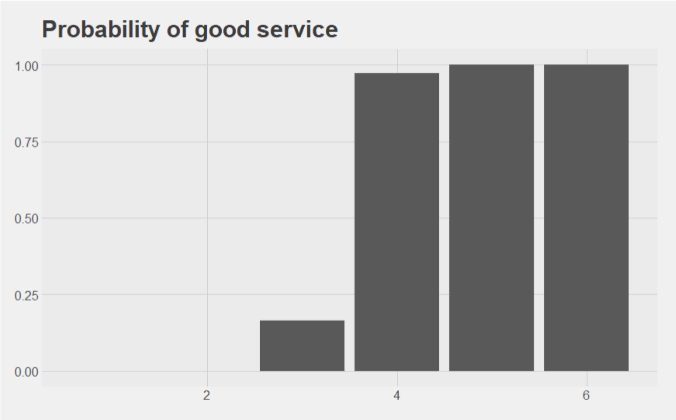


Figure 42: Probability of good service

It was assumed that good service is when the customer receives their coffee up to 60 seconds after placing their order. The probability of good service is shown in the table above. For one or two baristas the probability of good service is zero. The probability of good service increases with the number of baristas from three onward with it reaching 100% at 6 baristas.

Taguchi loss

The Taguchi loss function could be applied to account for the gradual loss in customer satisfaction as service time deviates from the 60 second target. Instead of a pass/fail definition of good service, the loss increases faster with longer waiting times, reflecting reduced future sales and reputational effects. Incorporating this loss into the model shows that slightly more baristas may be optimal, as better service consistency offsets higher staffing costs.

Coffee Shop 2

The same assumption that was made for coffee shop 1 is assumed to hold for coffee shop 2.

The variation of customer waiting time for varying numbers of baristas working has been plotted below. From this it can be seen that there is an exponential decrease in the time gained by adding a barista. The limit the exponential decrease approaches are however different than that of the first coffee shop.

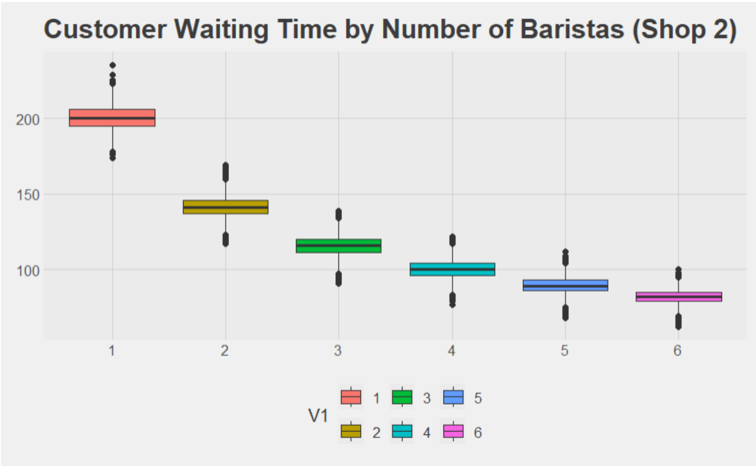


Figure 43: Boxplot of waiting time with varying barista quantities (Shop 2)

The quantity of people serviced per day was shown below. It can be seen from the graph that it exhibits an exponentially decreasing increase in customers serviced compared to the number of baristas. This would suggest that there is a maximum limit of customers serviced. This could be due to inefficiencies with baristas sharing the work area in the café.

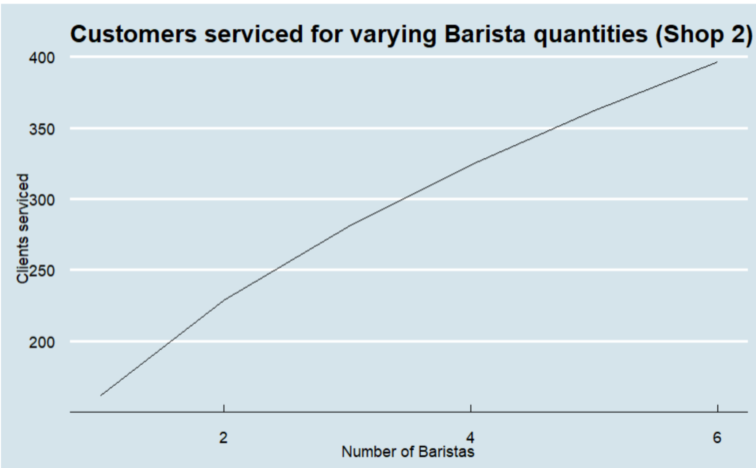


Figure 44: Customers serviced (Shop 2)

The profit made per day for the various numbers of baristas were calculated and is shown below. It can be seen that each additional barista has a decreasing impact on the amount of extra profit made. The profit made approaches a limit as the graph below starts to plateau at higher barista quantities.

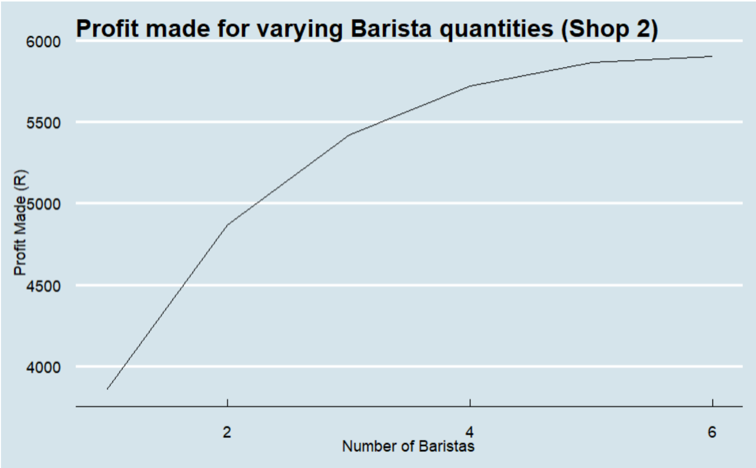


Figure 45: Profit made (Shop 2)

As such the ideal number of baristas are six with a profit of R5 905.53 per day. It would however not be advisable to increase the number of baristas. The graph seems to have reached a plateau at 6 baristas.

If it is still assumed that good service is delivered when delivery is under 60 seconds, no number of baristas would lead to any chance of good service.

Baristas	Probability of good service
1	0
2	0
3	0
4	0
5	0
6	0

If due to the increased waiting time the expectation of good service is adjusted to service rendered at or faster than 90 seconds, the probability of good service looks as follows.

Baristas	Probability of good service
1	0
2	0
3	0
4	0.04550993
5	0.58420976
6	0.97440770

Figure 46: Chance of good service (90 seconds)

From this it can be seen that six baristas would result in a 97% chance of good service. At 5 baristas the chance of good service drops dramatically to 58% and would not be acceptable in business.

ANOVA

One-way ANOVA was calculated for each product category in relation to mean delivery hours between the two years. We reject the null hypothesis if p is smaller than 0.05. The results are discussed below for each product category.

Null hypothesis (H_0): There is no difference in the mean delivery time between the two years.

Alternative hypothesis (H_1): There is a difference in the mean delivery time between the two years.

Cloud

```
=== ANOVA Results ===
               Df Sum Sq Mean Sq F value Pr(>F)
orderYear       1      1    1.17   0.031   0.86
Residuals   15596 583187   37.39
```

The p value for the cloud product category is 0.86, thus we fail to reject the null hypothesis. Therefore there is no statistically significant difference between mean delivery time for the years.

Software

```
=== ANOVA Results ===
               Df Sum Sq Mean Sq F value Pr(>F)
orderYear       1      0 0.01695   0.179  0.672
Residuals   20747   1966 0.09475
```

The p value for the software product category is 0.672, thus we fail to reject the null hypothesis. Therefore there is no statistically significant difference between mean delivery time for the years.

Keyboard

```
=== ANOVA Results ===
               Df Sum Sq Mean Sq F value  Pr(>F)
orderYear       1    299  299.33    8.07 0.00451 **
Residuals   17918 664603   37.09
```

The p value for the keyboard product category is 0.00451, thus we reject the null hypothesis. Therefore there is a statistically significant difference between mean delivery time for the years.

Mouse

```
=== ANOVA Results ===
               Df Sum Sq Mean Sq F value Pr(>F)
orderYear       1     20   19.94    0.53  0.467
Residuals   20660 777830   37.65
```

The p value for the mouse product category is 0.467, thus we fail to reject the null hypothesis. Therefore there is no statistically significant difference between mean delivery time for the years.

Monitor

```
=== ANOVA Results ===
               Df Sum Sq Mean Sq F value Pr(>F)
orderYear       1     16   16.36   0.447  0.504
Residuals   14862 543499   36.57
```

The p value for the monitor product category is 0.504, thus we fail to reject the null hypothesis. Therefore there is no statistically significant difference between mean delivery time for the years.

Laptop

```
=== ANOVA Results ===
               Df Sum Sq Mean Sq F value Pr(>F)
orderYear       1     18   18.15   0.496  0.481
Residuals   10205 373353   36.59
```

The p value for the laptop product category is 0.481, thus we fail to reject the null hypothesis. Therefore there is no statistically significant difference between mean delivery time for the years.

Reliability of service

From the graph that was provided (shown below) it can be seen that there is an exponential distribution in the number of days where a specific number of workers were present. It is furthermore assumed that 16 people are employed at the car rental agency.

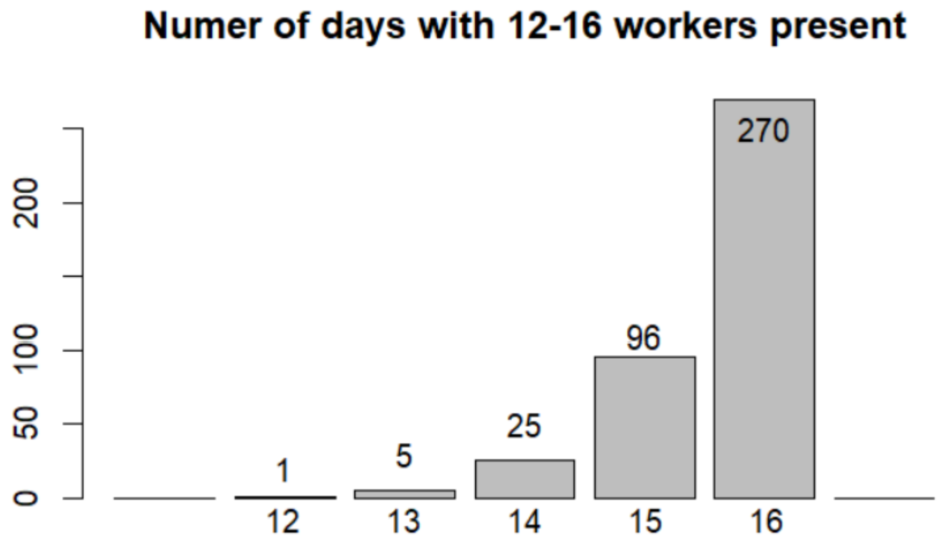


Figure 47: Number of days for various workers present

Days of reliable service

The probability of an employee coming to work was calculated by way of the total number of employee days worked (number of days for specific employees multiplied by number of employees summed together) over the total number of employee work days if all 16 employees worked each of the 397 days.

The probability was calculated to be 0.9740239, which is within what would have been expected from an employee. This means an employee is on average not at work for 9.5 days per year (365 days)

If less than 15 employees work any given day the service rendered by the business is not reliable. This was calculated using the binomial formula for good service (15 or 16 employees present) to obtain percentage of good service days. The percentage of good service days was multiplied by 365 days in a year to obtain the number of good service days per year.

The percentage of good service days was **0.936369** resulting in **341.7747** good service days per year.

Optimise the profit for the company

It was given that for each bad service day that the company experiences it loses R20 000 in sales for that given day. Furthermore appointing an additional worker costs R25 000 per month.

The given costs combined with the probability of an employee coming to work was used to calculate the cost in lost sales, the employee costs as well as the total cost.

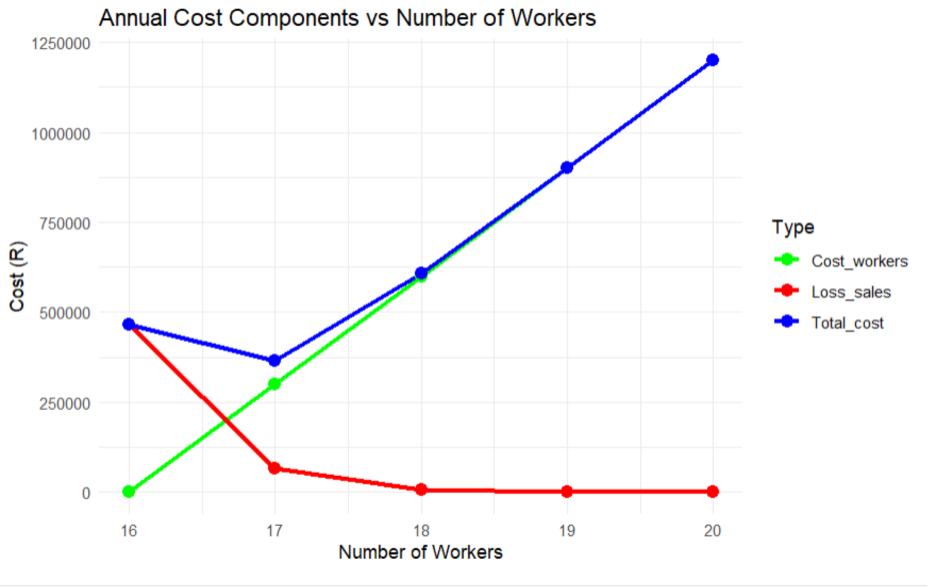


Figure 48: Labour and lost sales costs

From Figure 48 it can be seen that the cost of workers linearly increases from R0 for 16 employees by R300 000 per added employee. The loss of sales exponentially decreases from 16 employees and approaches R0 as employees are added.

Workers	Loss_sales	Cost_workers	Total_cost
16	464506.15934	0	464506.2
17	66219.81652	300000	366219.8
18	7592.96978	600000	607593.0
19	739.94180	900000	900739.9
20	63.48579	1200000	1200063.5

The Total cost to the company from sales losses and extra employees is shown in the blue line in Figure 48 as well as in the table above.

The minimum cost to the company is at 17 employees with a total cost of R366 219.80.

Conclusion

From the data analysis carried out it can be seen that the gender distribution of clients is equaly spread. Furthermore middle-aged clients has the highest average income and forms a high proportion of the businesses clients. As such focus should be placed on them.

From the statistical process control that was carried out on sales data in the second phase of the process SPC charts and control limits for each product category was gathered. The SPC charts indicated an upward trend in mean delivery time for all product categories from the beginning of each year. Outliers were identified to enable alerts to be given to each product categories manager for corrective actions.

Trends in the data was analysed and described to enable better understanding by process managers. Each product categories process capability was further calculated and interpreted. All processes except for those related to Software are not capable to meet the delivery hour bounds that was set. Only the software product category was capable of performing within the set bounds. Lastly the three different types of errors were identified for each product category as outlined in the project brief.

The third section of the project regarded Risk, Data correction & optimising for maximum profit. Type I and II errors for the three different types of errors outlined in the second phase was calculated. Data correction and reinterpretation formed a large part of the third phase. Both head office and product data had to be corrected and reinterpreted.

Profit optimisation was carried out on two coffee shops with the time to serve being provided for varying numbers of baristas. An optimal amount of baristas was calculated for each coffee shop to maximise the profit of each shop. It was found that six baristas were optimal for both coffee shops.

In the last phase of the project ANOVA tables and reliability of service were calculated. An ANOVA table was calculated for each product category regarding if the average delivery time was different between the two years of data. It was found that only Keyboards mean delivery had a statistically significant variation between the two years, with all other product categories not having a statistically significant variation.

Profit maximization was carried out on a car rental agency to find an ideal amount of workers. It was concluded that the business requires 17 employees. This was the ideal trade-off between loss of sales and personnel costs.

References

Banerjee, A. (2025, 10 22). *Hypothesis testing, type I and type II errors*. Retrieved from National Library of Medicine: https://pmc.ncbi.nlm.nih.gov/articles/PMC2996198/

Schalkwyk, T. D. (2025). *QA344 Statistics.* Stellenbosch.