Quality Assurance 344

ECSA Project Report

T Barnard – 21733929

Executive Summary

Key findings are as follows:
ECSA Part 1: An improvement in picking hours of products can improve overall delivery time. Quantity ordered by customers has decreased from 2022 to 2023 despite pricing being in line with customer income. 50% of customers are not at risk of becoming inactive, however due to the top products being purchased (keyboards, monitors and software) it would be advisable to invest in attracting new customers.

ECSA Part 3: The SPC Charts revealed that delivery times/ hours were predominantly within spec, however picking times for SOF, MOU & KEY show excessive variation and as such require a process review.

ECSA Part 4: The chance of a Type I error is 0.78% for the SPC Charts, the chance of a Type II error is 0.84% for the bottle filling process. The erroneous data was corrected and it was found that the average selling prices has significantly changed for the laptop, monitor and software.

ECSA Part 5: It is recommended based on reliable service of 90% (<120s) that coffee shop 1 employ two baristas and coffee shop 2 employ 4 baristas.

ECSA Part 6: The results obtained from the M/ANOVA aligned with those found in Section 2 of this report (SPC). It would therefore be recommended to use these findings to target specific products or processes to facilitate improvement and refinement.

ECSA Part 7: The optimal number of new staff members to employ is 1 with reliable service days being: Empirical = 359.4836 & Binomial = 361.6890.

Investigation

A data analysis was conducted on the following datasets: customer_data.csv, products_data.csv, products_Headoffice.csv, sales2022and2023.csv, sales2026and2027, timeToServe.csv and timeToServe2.csv as provided by the company.

The information/ data was assessed for completeness and further assessed for any correlations/ insights which can be considered of interest to the company, such as customer behaviour, pricing vs customer income and product performance. Furthermore, Statistical Process Control (SPC) Charts were drawn up for picking and delivery times/ hours and additional tasks such as profit optimisation were carried out and documented.

All data was processed and analysed using R Studio.

1.) Key Findings in Part 1

1.1) Customer Data:

It was found that the dataset provided was complete and as such no data cleaning was required. The table below shows the average age, income and gender split for the following cities: Chicago, Houston, Los Angeles, Miami, New York, San Francisco and Seattle.

Table 1

City	Av.Age	SD	Av.Income	SD	%Male	%Female	%Other
Chicago	51.1	21.0	82244.	32405.	50.4	45.0	4.56
Houston	52.9	21.3	80249.	32810.	46.5	49.7	3.73
L.A.	51.2	20.5	80475.	33136.	45.5	50.7	3.86
Miami	51.0	21.1	83346.	31870.	46.8	48.2	4.95
N.Y.	52.0	21.2	79752.	33658.	47.1	48.9	3.99
SanFran.	50.7	21.3	79853.	34290.	46.5	48.6	4.87
Seattle	52.0	22.2	79948.	33598.	46.1	49.3	4.61

The age of customers was then contrasted against their income, this was later further grouped into the above-mentioned cities, and further divided by gender as seen bellow:

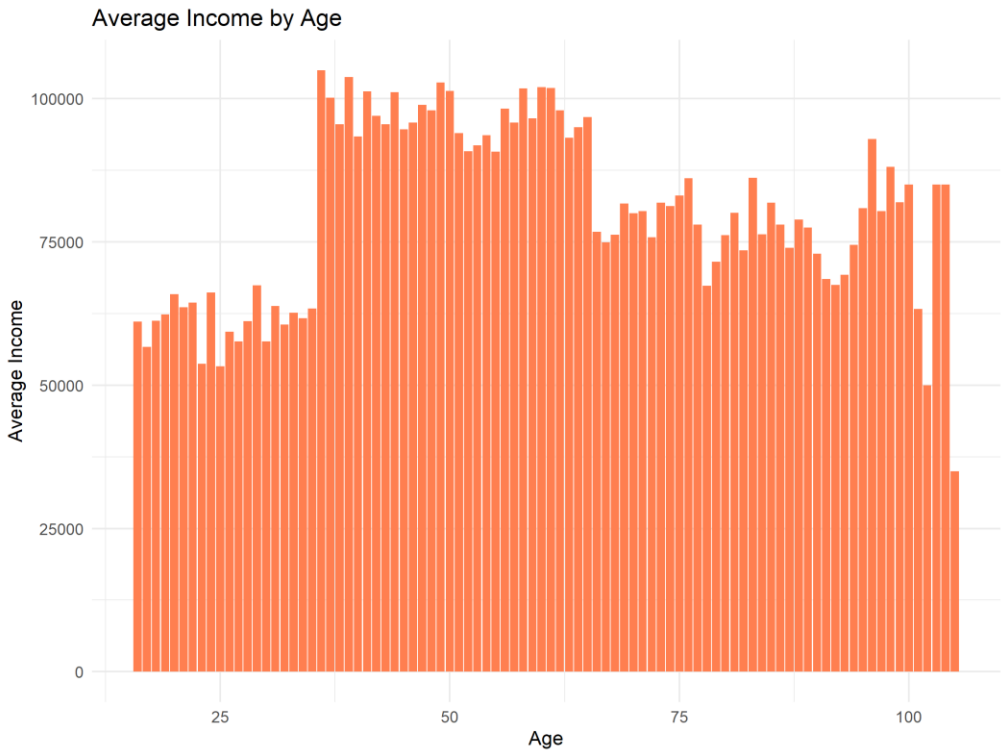


Figure 1.1

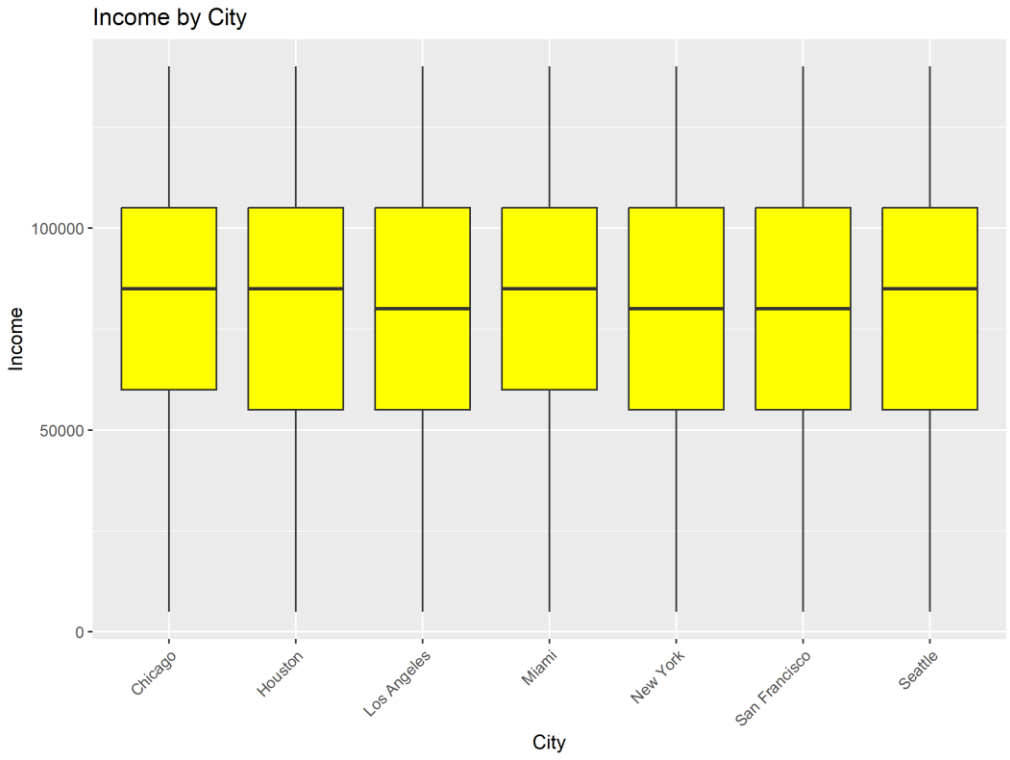


Figure 1.2

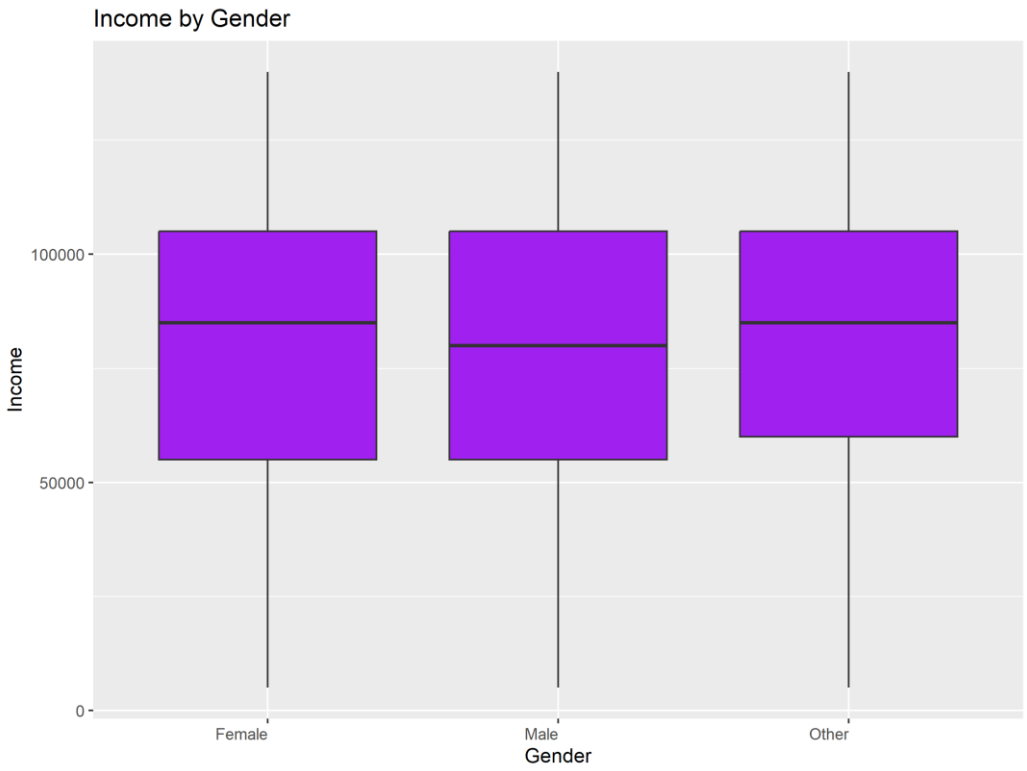


Figure 1.3

A correlation test done was then carried out on age vs income and the following results were obtained:

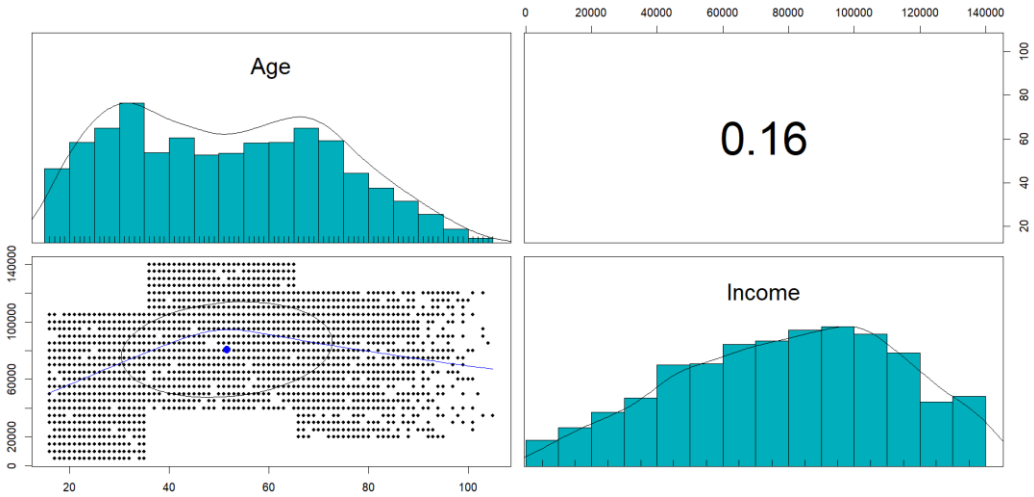


Figure 1.4

There appears to be a weak positive correlation between age and income with an R correlation coefficient value of 0.16 or 16% of the income gain can be explained by age. Furthermore, the age group that was found to possess the largest income was age groups 40 – 60.

Lastly, there does not seem to be any significant differences in income between the genders and or cities supplied.

1.2) Product and Head office Data:

The data supplied by the company for the product and head office was also found to be complete. The selling prices of products per category was compared in each dataset along with mark up as seen below.

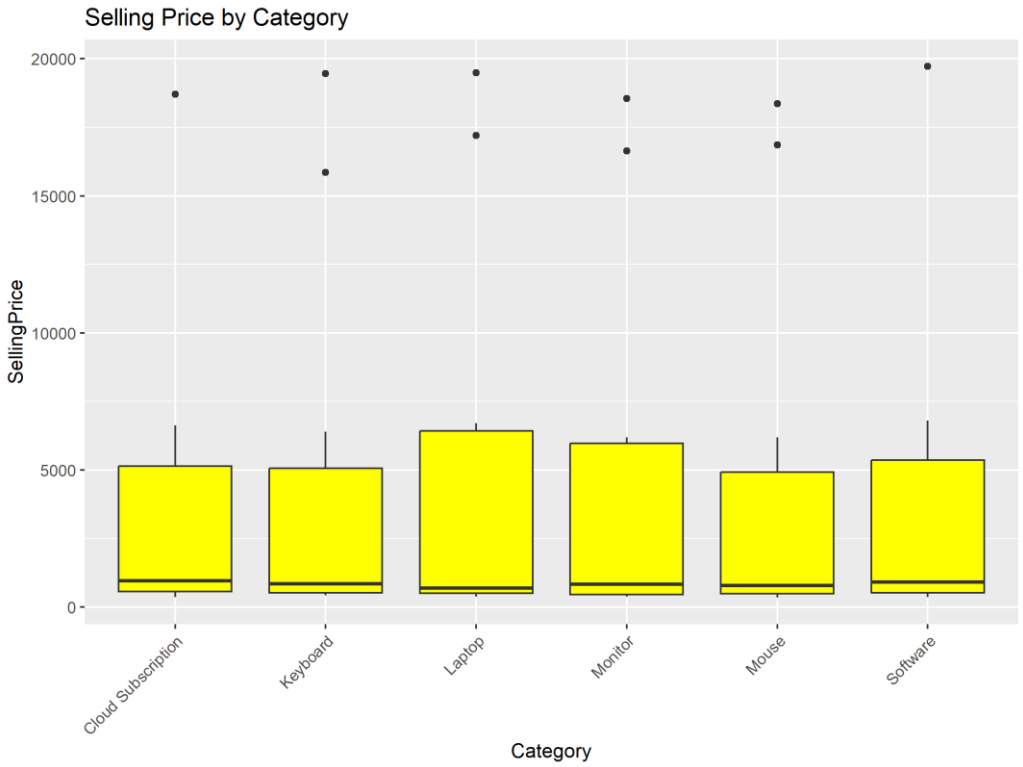


Figure 2.1 Products Data

It can be seen within figure 2.1 that the data is skewed due to some products selling for far more than other products as well as including

further outliers. It can also be seen below in figure 2.2 that sales price is not exclusively tied to markup.

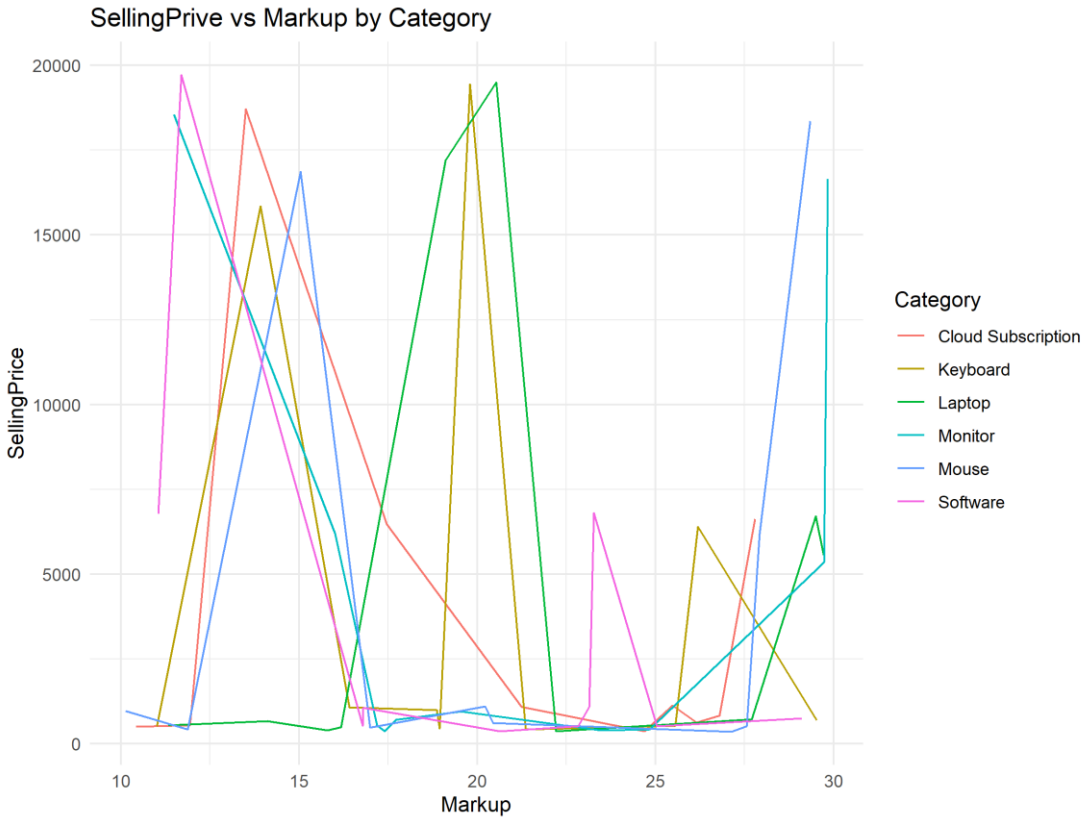


Figure 2.2 Products Data 1

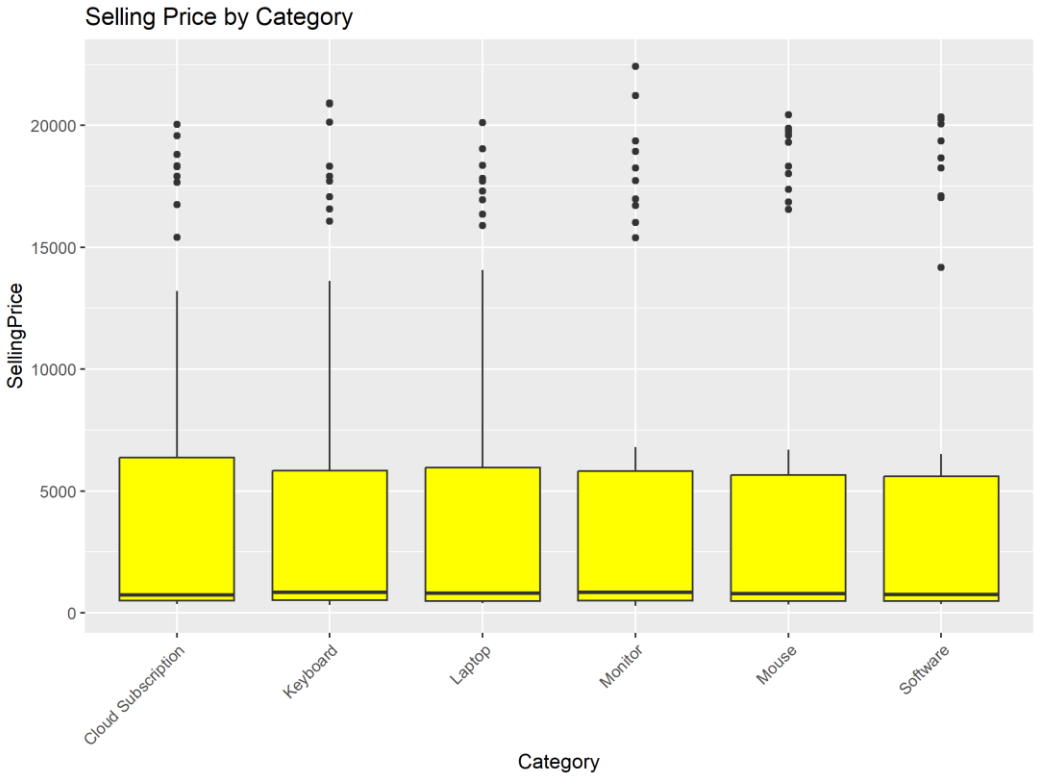


Figure 2.3 Head office Data

Figure 2.4

It can also be seen that the above-mentioned trend is repeated within figure 2.3. Finally, the selling price density of the products were compared to the income density of the clients in figure 2.4 and it was found that all clients are easily able to afford the current pricing.

1.3) Sales Data:

The data supplied contained no missing information in addition to customer purchasing quantities as well as delivery hour information. Firstly, the picking hours and delivery hours were contrasted against quantity ordered and order time for correlations, as seen below.

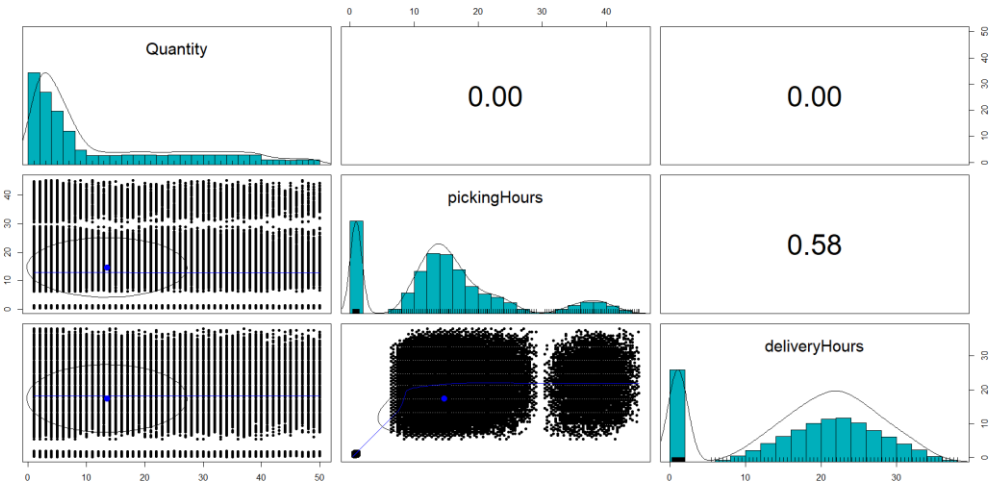


Figure 3.1

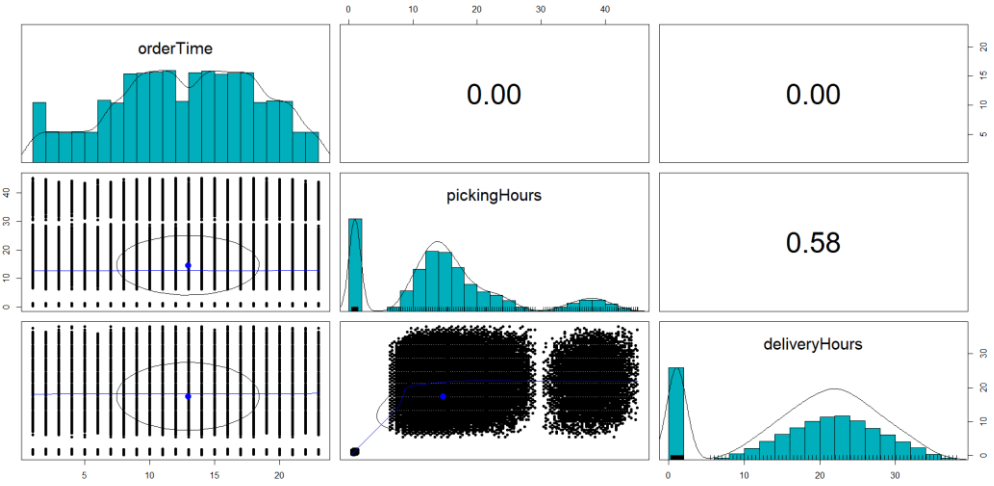


Figure 3.2

It was found that there was no correlation between order time and quantity with picking hours and delivery hours, however there was a relatively strong correlation of an R value of 0.58 between picking hours and delivery hours. It was therefore found that picking hours explained 58% of the delivery hours data.

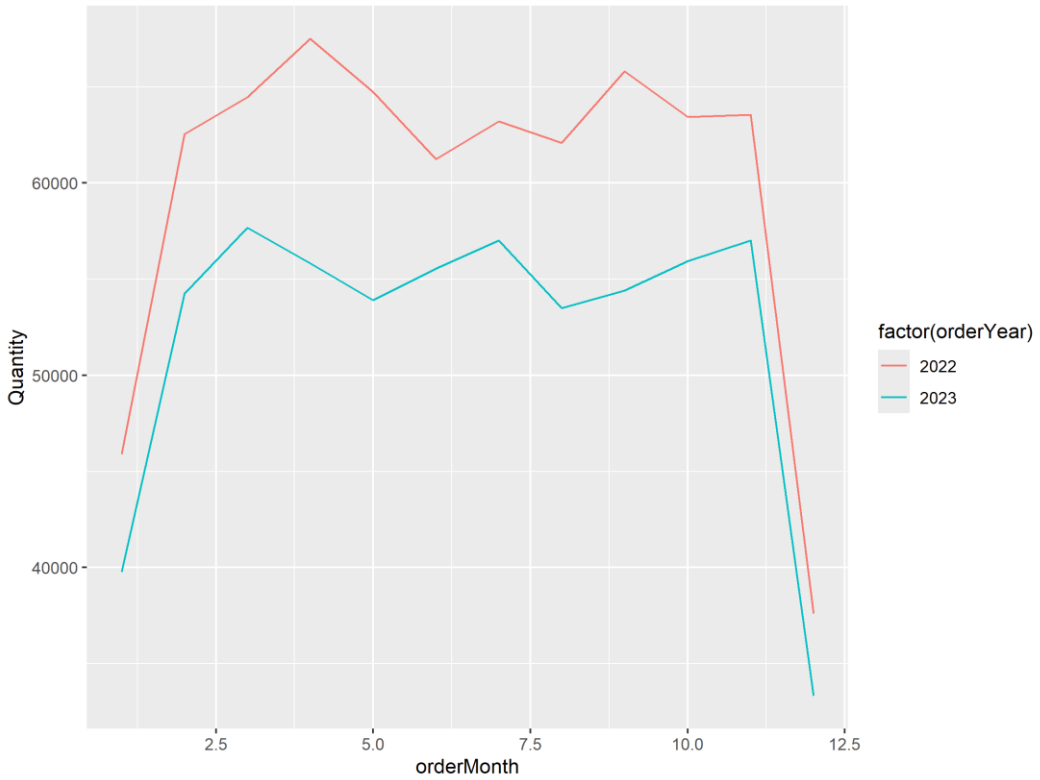


Figure 3.3

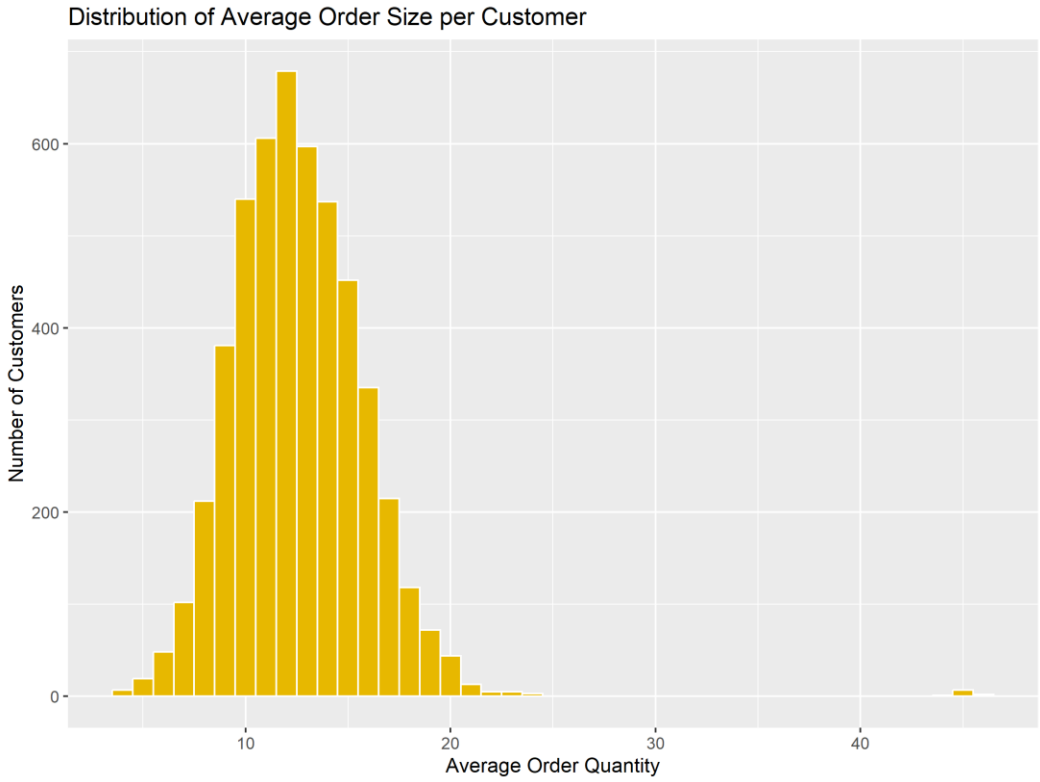


Figure 3.4

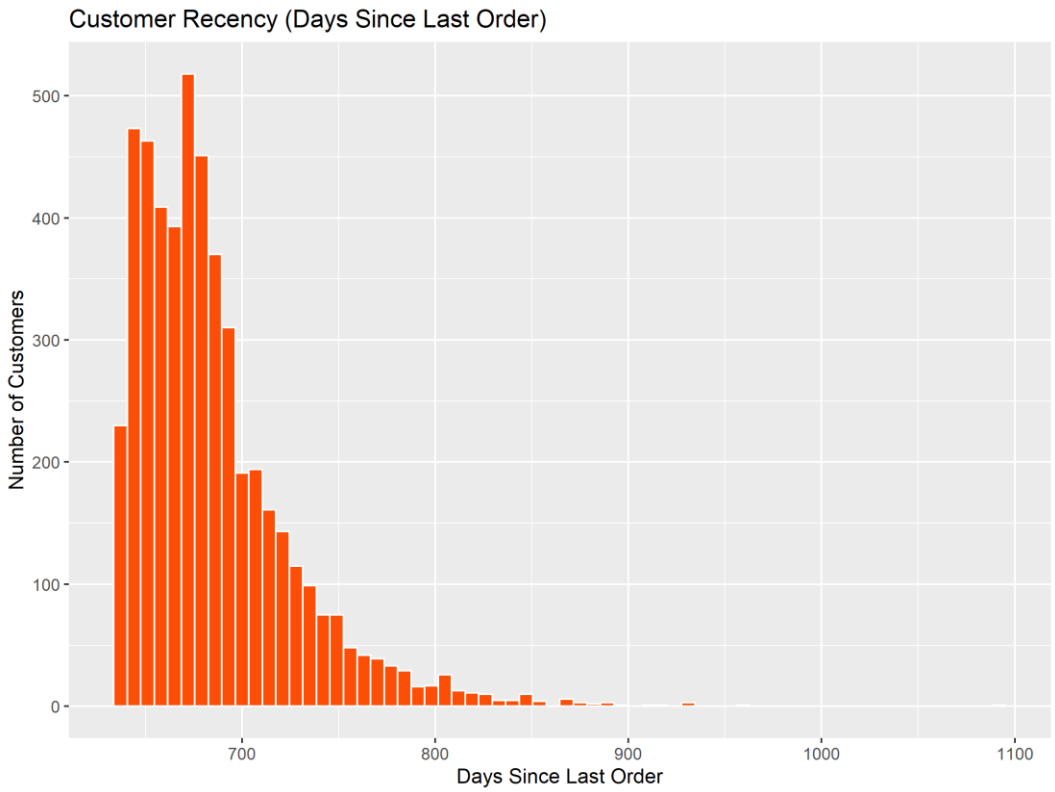


Figure 3.5

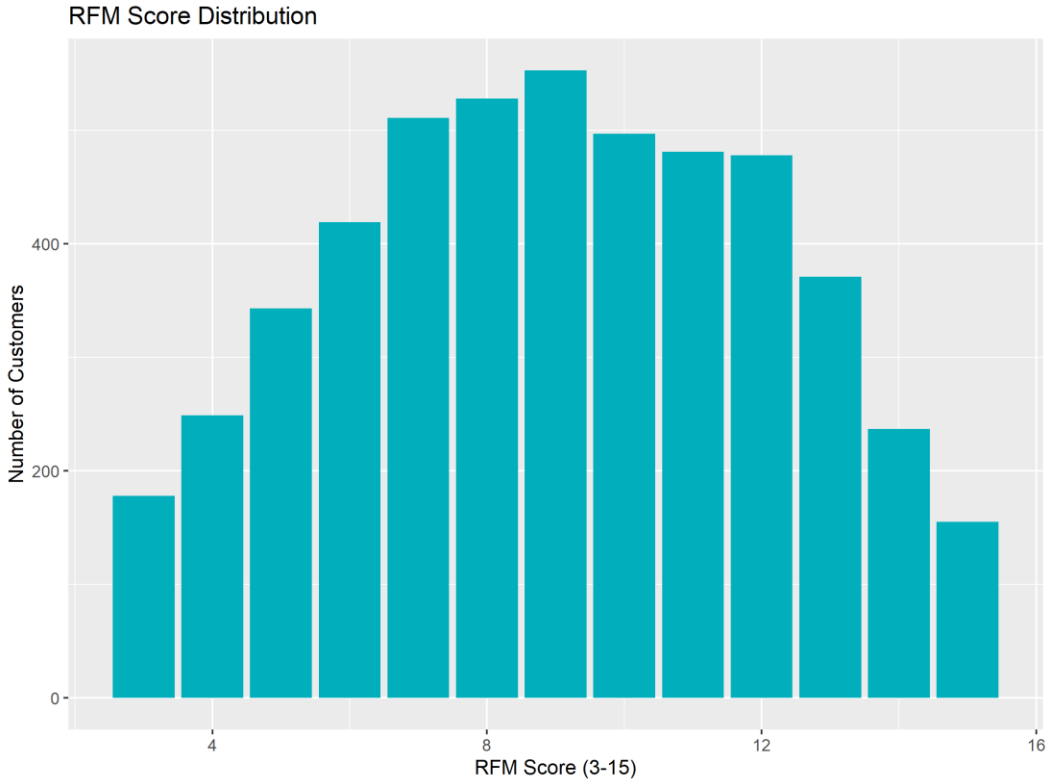


Figure 3.6

It was found that the order quantity has decreased from 2022 to 2023 with most customers having days between orders ranging to 700 days as seen in figures 3.3 and 3.5, while figure 3.4 shows that most clients on

average had product order quantities of roughly 10 items. Lastly, figure 3.6 shows an RFM score which is a ranked score based on recency, frequency and monetary spending, low scores indicate recent activity while high scores indicate inactivity. It can therefore be seen that 50% of customers are relatively active.

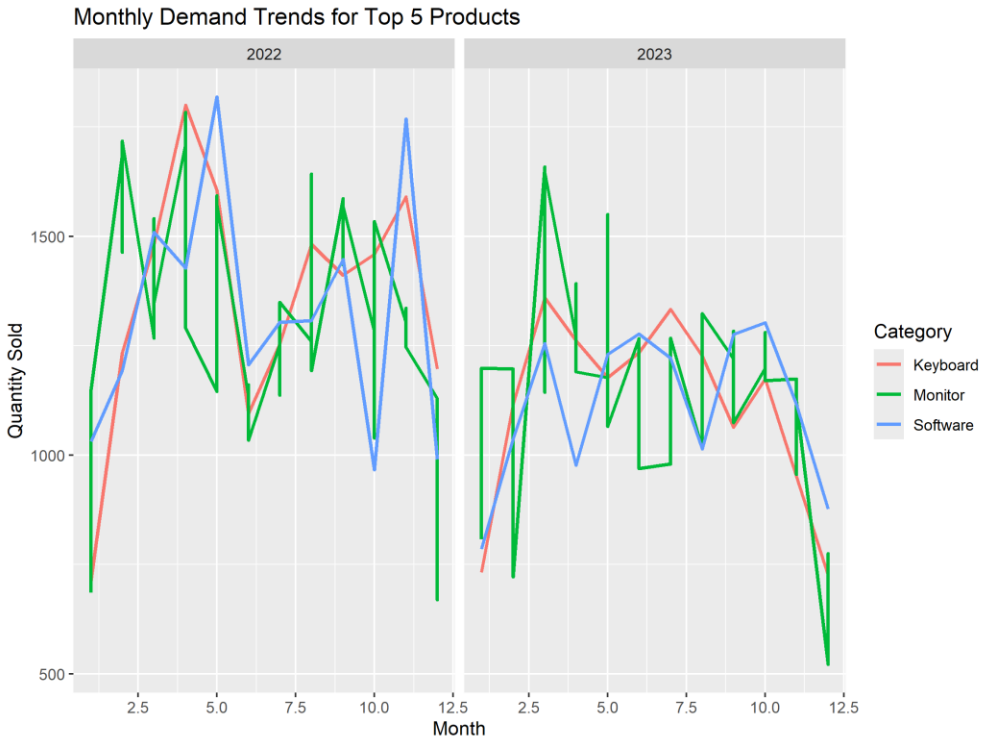


Figure 3.7

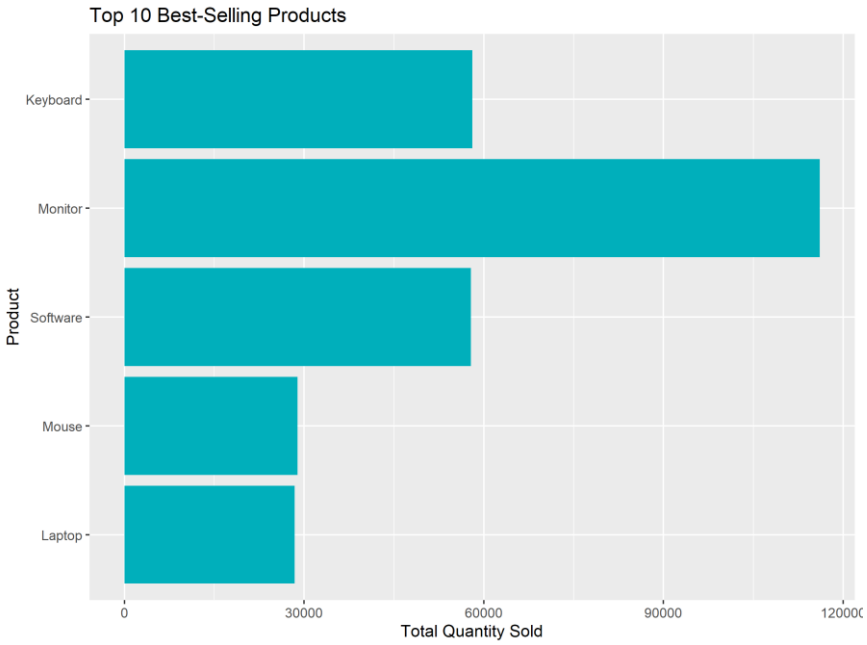


Figure 3.8

It was also found that the most purchased and demanded products were keyboards, software and monitors, where monitors were the highest preforming item across both 2022 and 2023 as seen in figures 3.7 and 3.8.

1.4) Recommendations & Conclusion:

In conclusion, it was found that if picking hours were improved it would improve the subsequent delivery hours' time therefore speeding up the process of delivery to customers. It was further found that quantity purchased has decreased from 2022 to 2023 while pricing has been in line with customer income. Despite this, only 50% of customers are not at risk of becoming inactive. When this is contrasted against the most purchased products being software, keyboards and monitors which have longer service lives and therefore requiring less frequent replacement. This would therefore imply that customer inactivity is no surprise, but new customers are required to bridge the gap in inactivity.

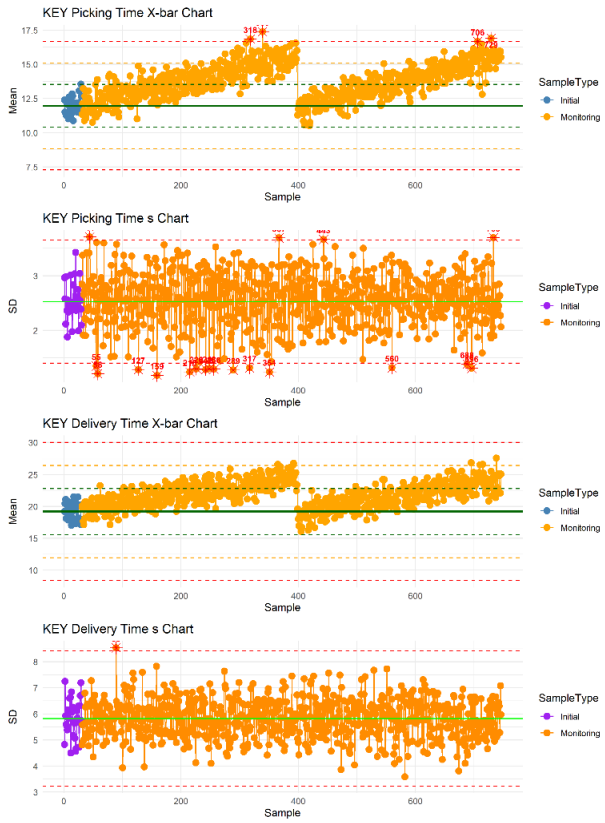
2.) Part 3 SPC

The following Statistical Process Control (SPC) charts were generated in R Studio for the picking and delivery process for the following products: Software, Mouse, Keyboard, Monitor, Cloud Services and Laptops. The Product ID was used to group the products and compile the SPC Charts using the first three digits of each product code.

Metric	Value
Rule A ±3σ Violations	0
Rule B Longest ±1σ Run	36
Rule C 4+ above +2σ	Yes
Cp	0.87
Cpk	0.56

Metric	Value
Rule A ±3σ Violations	0
Rule B Longest ±1σ Run	140
Rule C 4+ above +2σ	No
Cp	0.87
Cpk	0.56

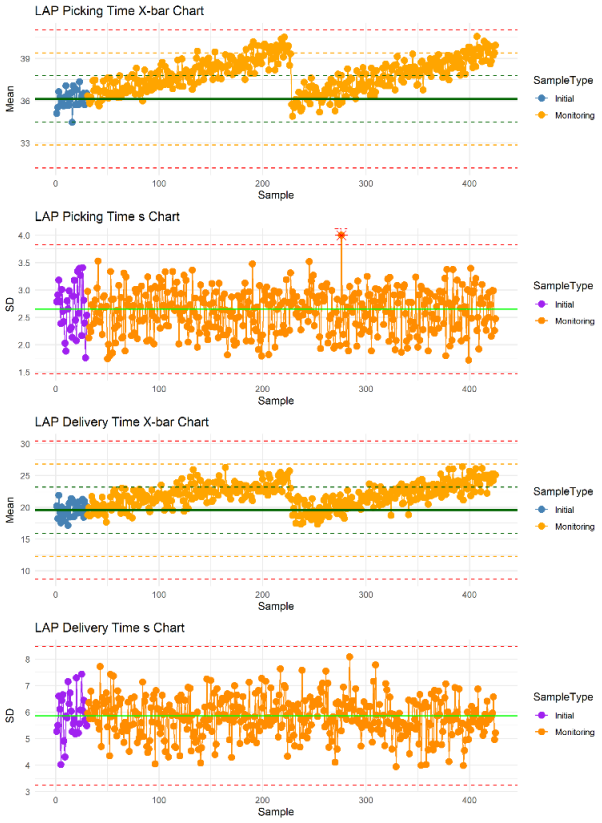
Figure 4.1 CLO



Metric	Value
Rule A ±3σ Violations	4
Rule B Longest ±1σ Run	78
Rule C 4+ above +2σ	Yes
Cp	0.88
Cpk	0.56

Metric	Value
Rule A ±3σ Violations	0
Rule B Longest ±1σ Run	99
Rule C 4+ above +2σ	No
Cp	0.88
Cpk	0.56

Figure 4.2 KEY



Metric	Value
Rule A ±3σ Violations	0
Rule B Longest ±1σ Run	57
Rule C 4+ above +2σ	Yes
Cp	0.88
Cpk	0.56

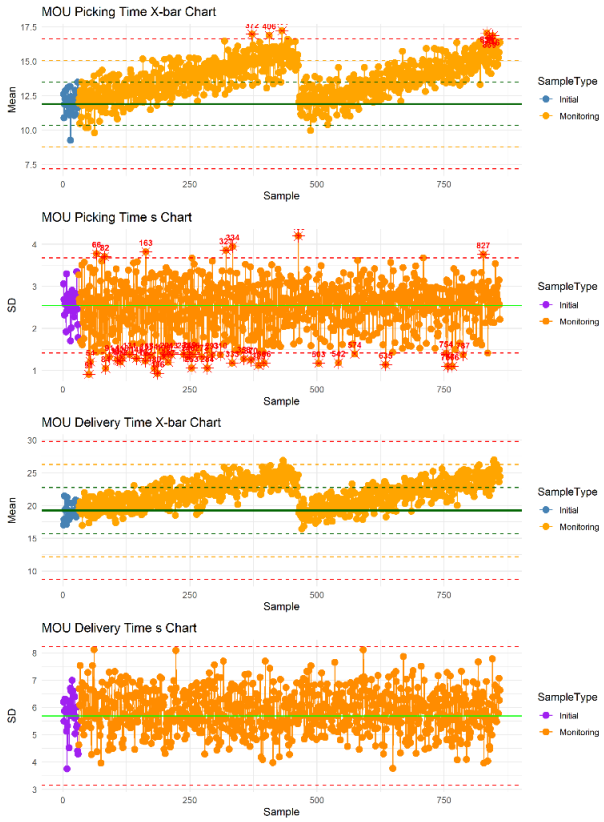
Metric	Value
Rule A ±3σ Violations	0
Rule B Longest ±1σ Run	101
Rule C 4+ above +2σ	No
Cp	0.88
Cpk	0.56

Figure 4.3 LAP

Metric	Value
Rule A ±3σ Violations	2
Rule B Longest ±1σ Run	46
Rule C 4+ above +2σ	Yes
Cp	0.88
Cpk	0.57

Metric	Value
Rule A ±3σ Violations	0
Rule B Longest ±1σ Run	98
Rule C 4+ above +2σ	No
Cp	0.88
Cpk	0.57

Figure 4.4 MON



Metric	Value
Rule A ±3σ Violations	6
Rule B Longest ±1σ Run	29
Rule C 4+ above +2σ	Yes
Cp	0.87
Cpk	0.55

Metric	Value
Rule A ±3σ Violations	0
Rule B Longest ±1σ Run	139
Rule C 4+ above +2σ	No
Cp	0.87
Cpk	0.55

Figure 4.5 MOU

Metric	Value
Rule A ±3σ Violations	9
Rule B Longest ±1σ Run	32
Rule C 4+ above +2σ	Yes
Cp	17.33
Cpk	1.18

Metric	Value
Rule A ±3σ Violations	0
Rule B Longest ±1σ Run	103
Rule C 4+ above +2σ	No
Cp	17.33
Cpk	1.18

Figure 4.6 SOF

The SPC Charts show that the delivery times are stable however the picking times of SOF, MOU & KEY show excessive variation indicating that the warehousing system needs to be reviewed in order to stabilise these times and bring them into spec.

2.1) Recommendations & Conclusion:

The SPC Charts have indicated that the warehousing operations that are responsible for the pickingTimes data are in need of review so as to garner further insight into potential process deficits and improvements.

3.) Part 4 Risk, Data Correction & Optimisation for maximum profit

4.1) It was found that the chance of making a Type I error on the SPC charts is 0.007812 or 0.78 %. This is still relatively high as in 1 million repetitions that will be equivalent to 7 812 false alarms and as such potential lost profit in stoppages. It would be recommended to review this finding and determine whether or not this is inline with the company objectives.

4.2) It was found that the bottle filling process as outlined in the project brief had a 0.008412 or 0.84 % chance of being a Type II error with a 0.1588 or 15.88 % detection chance and as such the chart is relatively insensitive to small changes. This might be of no consequence if small shifts are inconsequential however these small shifts might indicate the start of gradually moving out of spec.

4.3) It was found that product ID mismatches were present in the data during the analysis of products_data.csv and products_Headoffice.csv. Head office recommended correcting the file products_Headoffice.csv and generating new sales data. The following comparisons were drawn between the corrected data (new) and the old data:

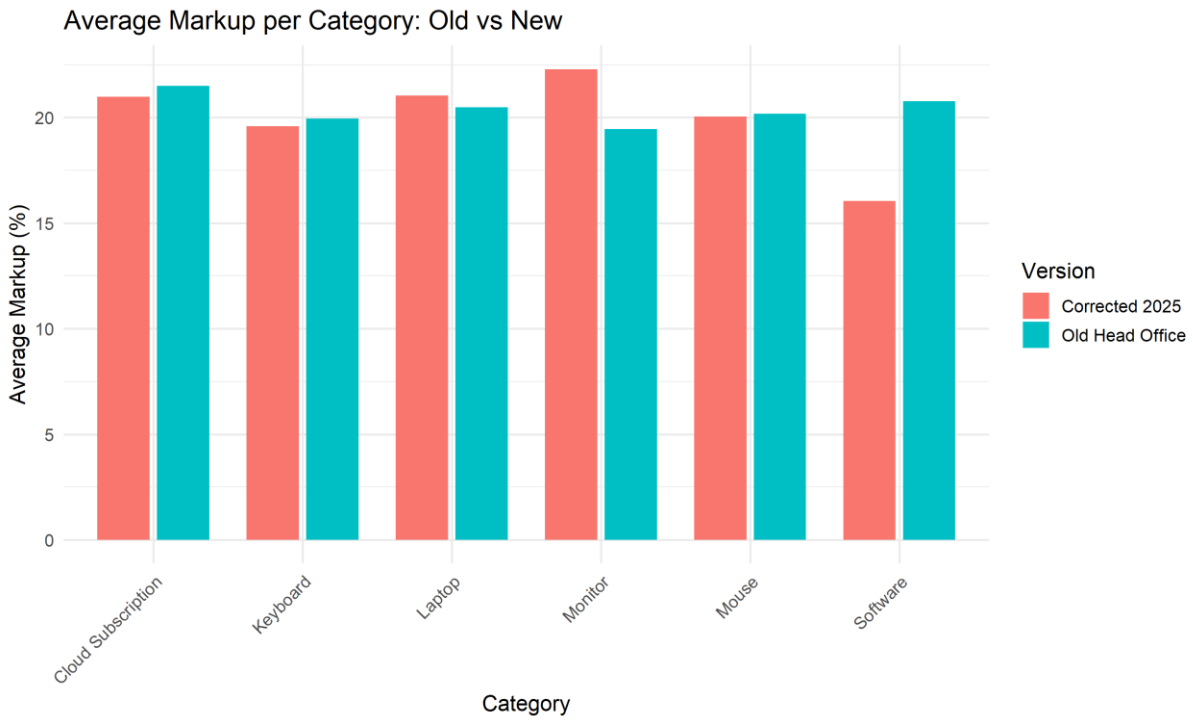


Figure 5.1 Average Markup (New vs Old)

Figure 5.2 Average Selling Price (New vs Old)

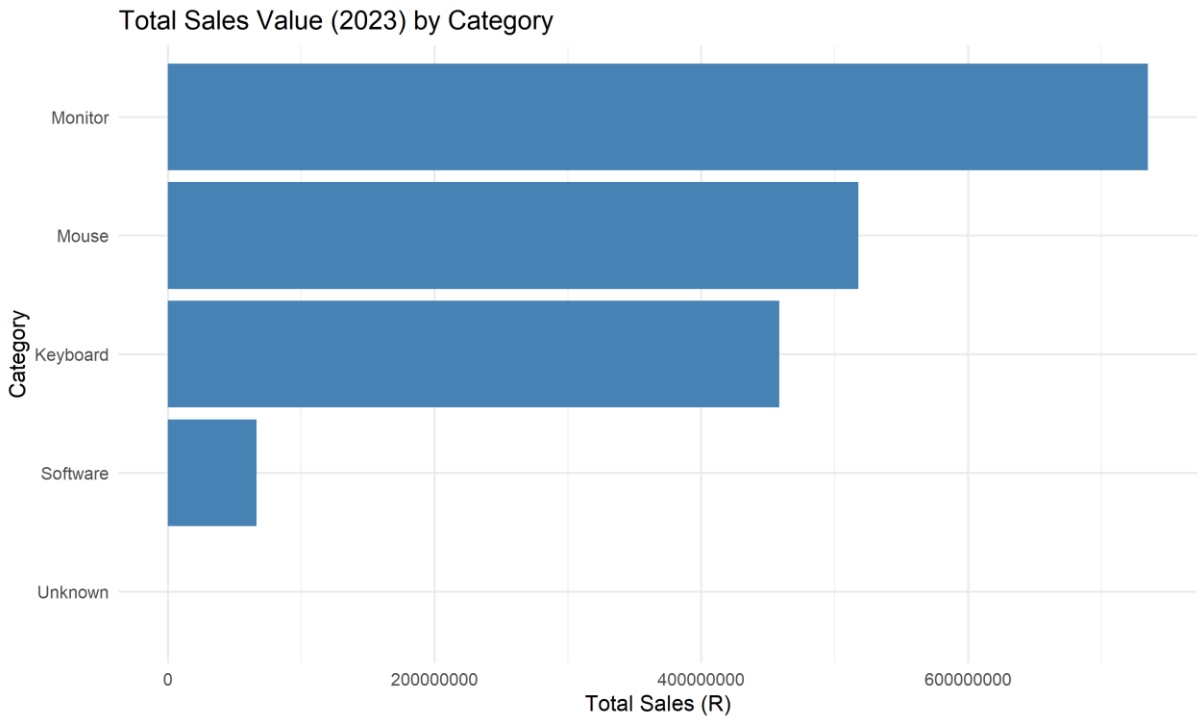


Figure 5.3 Total Sales by Category (New)

The restructured data when compared to the old data shows that the markups are consistent with only the monitor and software being significantly different (figure 5.1). Finally, the average selling prices have significantly changed for the laptop, monitor and software (figure 5.2) which has therefore altered their contribution to the total sales figures as seen in (figure 5.3) accordingly.

4.) Part 5 Profit Optimisation

Two sets of data were provided by two alternat coffee shops, namely datasets: timeToServe.csv and timeToServe2.csv. Profit gain per employee and reliability was modelled along with a simulated week showing a heat map of profits. The following graphs were obtained with thresholds of a maximum wait time of 120s and 90% reliability.

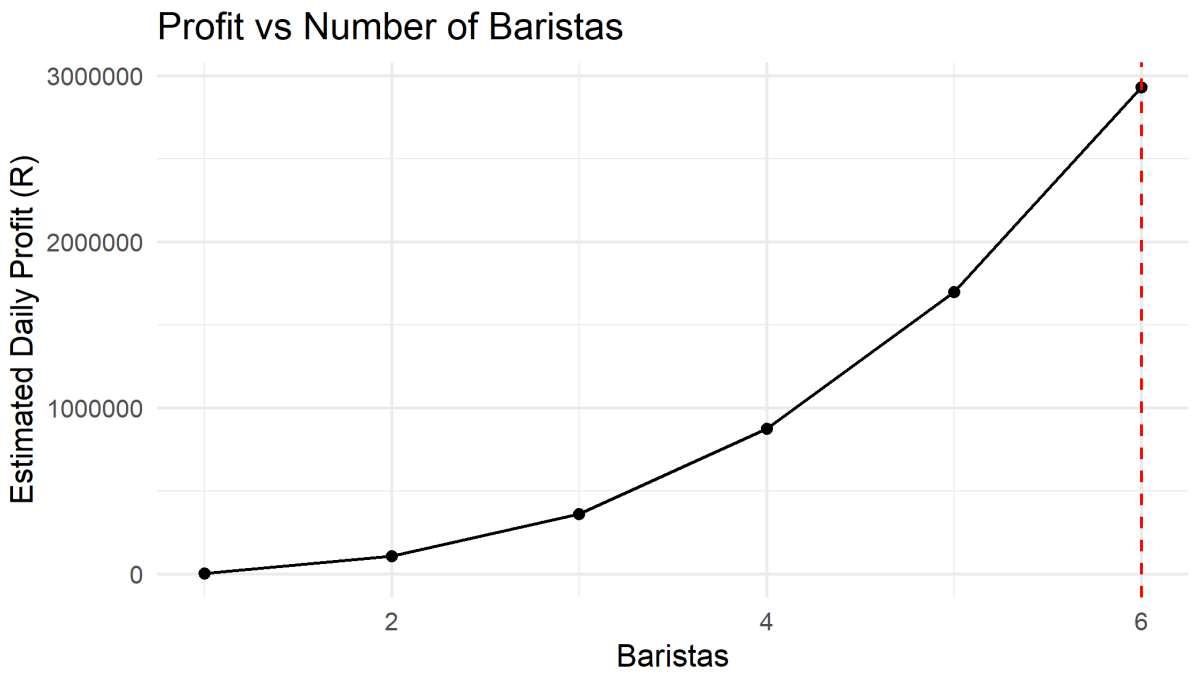


Figure 6.1 Profit vs Number of Baristas (Shop 1)

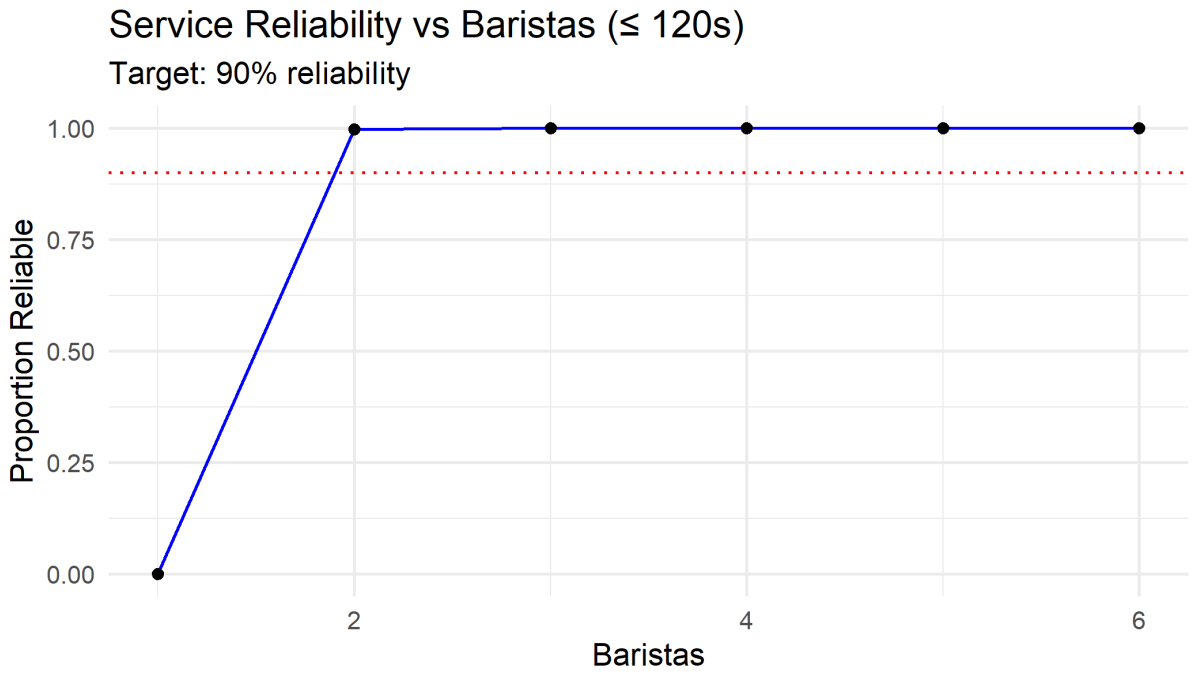


Figure 6.2 Service Reliability vs Baristas (Shop 1)

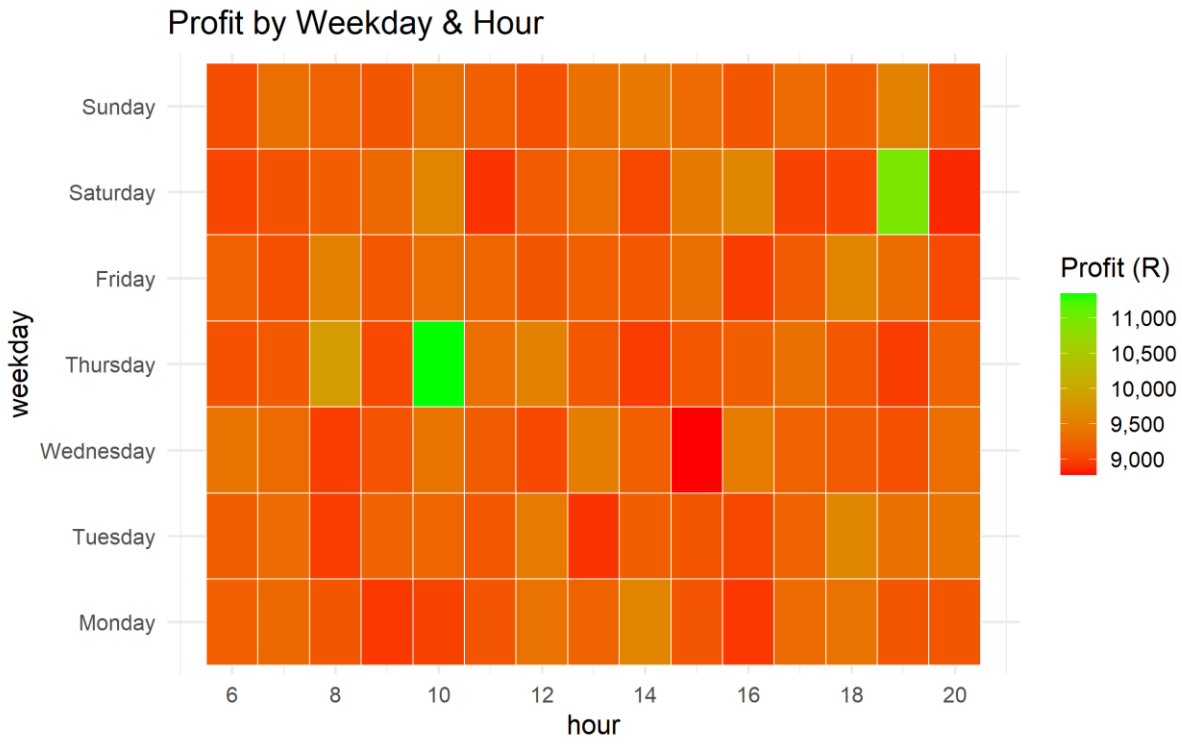


Figure 6.3 Profit by Weekday & Hour (Shop 1)

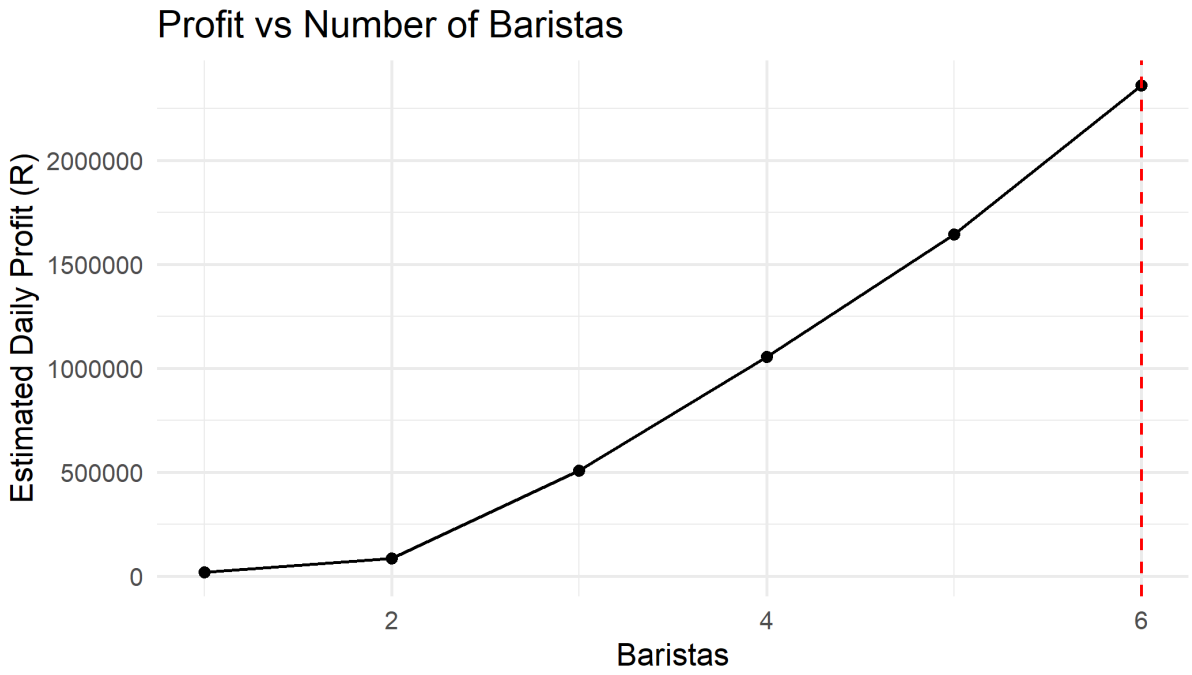


Figure 6.4 Profit vs Number of Baristas (Shop 2)

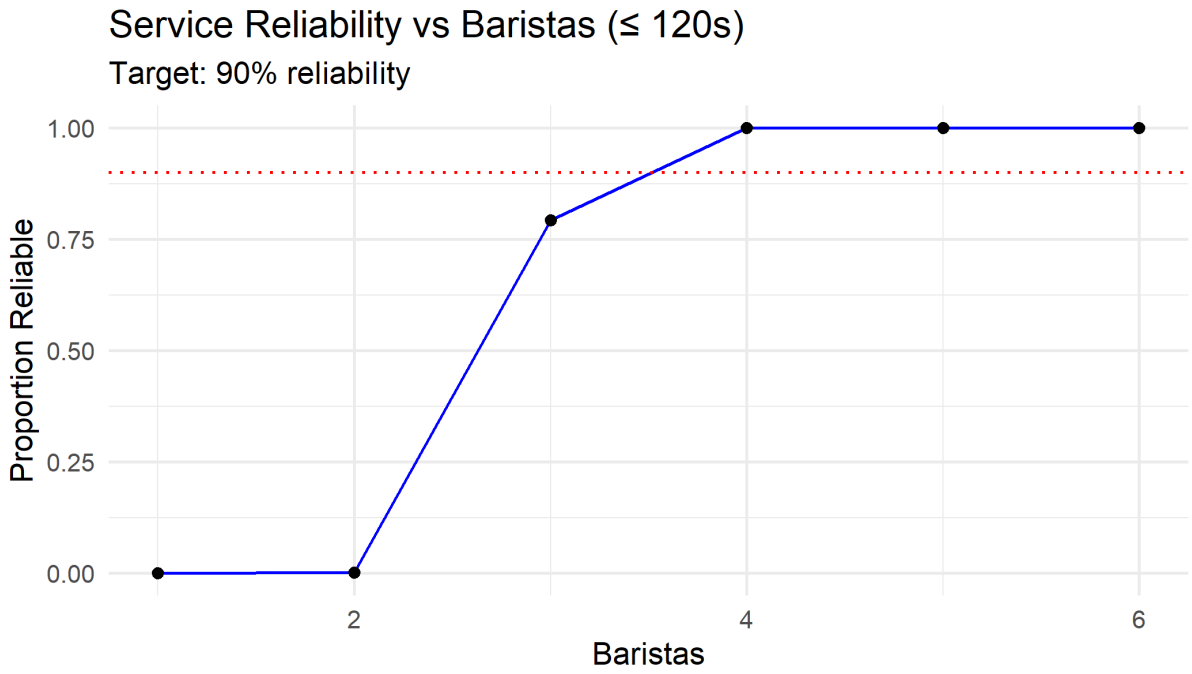


Figure 6.5 Service Reliability vs Baristas (Shop 2)

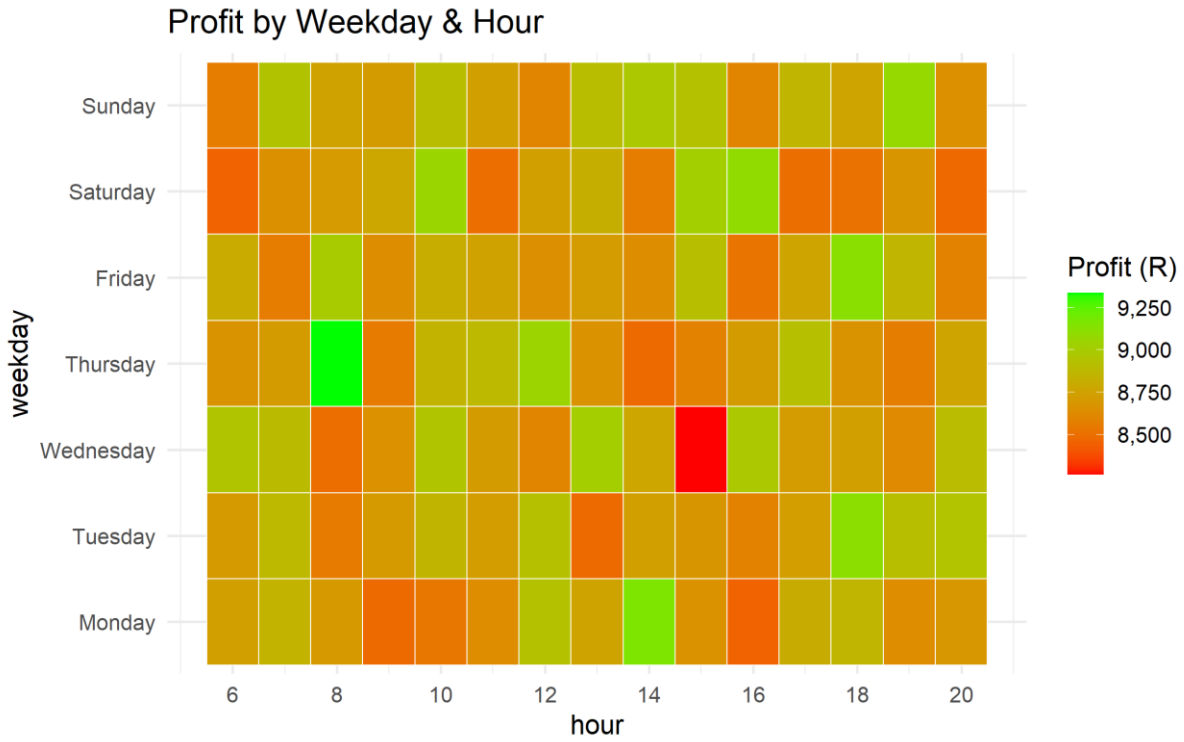


Figure 6.6 Profit by Weekday & Hour (Shop 2)

It can be seen in figures 6.1 and 6.4 that both shops' data trends towards higher profits with each added barista, however the economical viability of this strategy can be considered questionable. Furthermore, it was found that shop 1 exceeded 90% service reliability with two baristas (figure 6.2) and shop 2 with four baristas (figure 6.5).

4.1) Recommendations & Conclusion:

The minimum baristas needed to achieve a service reliability of 90% (<120s) is two for shop 1 and four for shop 2. The data however would indicate that higher profits can be achieved with more baristas however this data did not include the real-world operational costs. It is therefore a questionable conclusion to presume that profits will continue to increase beyond the graph's threshold of 6 baristas and as such require further study. It would be far more appropriate to base the appointment of baristas in each shop to the number required to maintain an acceptable level of service which was two and four for shops 1 and 2 respectively. The number of baristas can then be further increased at a later date provided new evidence indicates that the appointment of said new baristas would prove beneficial.

5.) Part 6 DOE, ANOVA/ MANOVA

This section of the report deals with a Design of Experiment (DOE) that includes a Multivariance Analysis of Variance (MANOVA) for the data analysed in Section 2 of this report or more specifically section 3 of the ECSA Brief.

The purpose of this MANOVA was to investigate whether year and product type impacted picking times and delivery times. The data was prepared by factorising year, month with the product type discerned from productID for >200 observations. An ANOVA was preformed on delivery hours and assumptions were checked using Levene's Test for variance homogeneity, Box's M Test for equality of covariance matrices. Furthermore, to explore pairwise product differences, a Tukey HSD post-hoc test was applied, and a MANOVA jointly assessed both response variables using Pillai's Trace. The results can be found below:

Test	p.value	Significant
ANOVA: Year	0.03003881	Yes
ANOVA: ProductType	0.00000000	Yes
ANOVA: Year*ProductType	0.09909614	No
Levene's Test	0.00000000	Yes
Box's M Test	0.00000000	Yes
MANOVA (Pillai Trace)	0.01116182	Yes

Table 2 M/ANOVA Summary Table

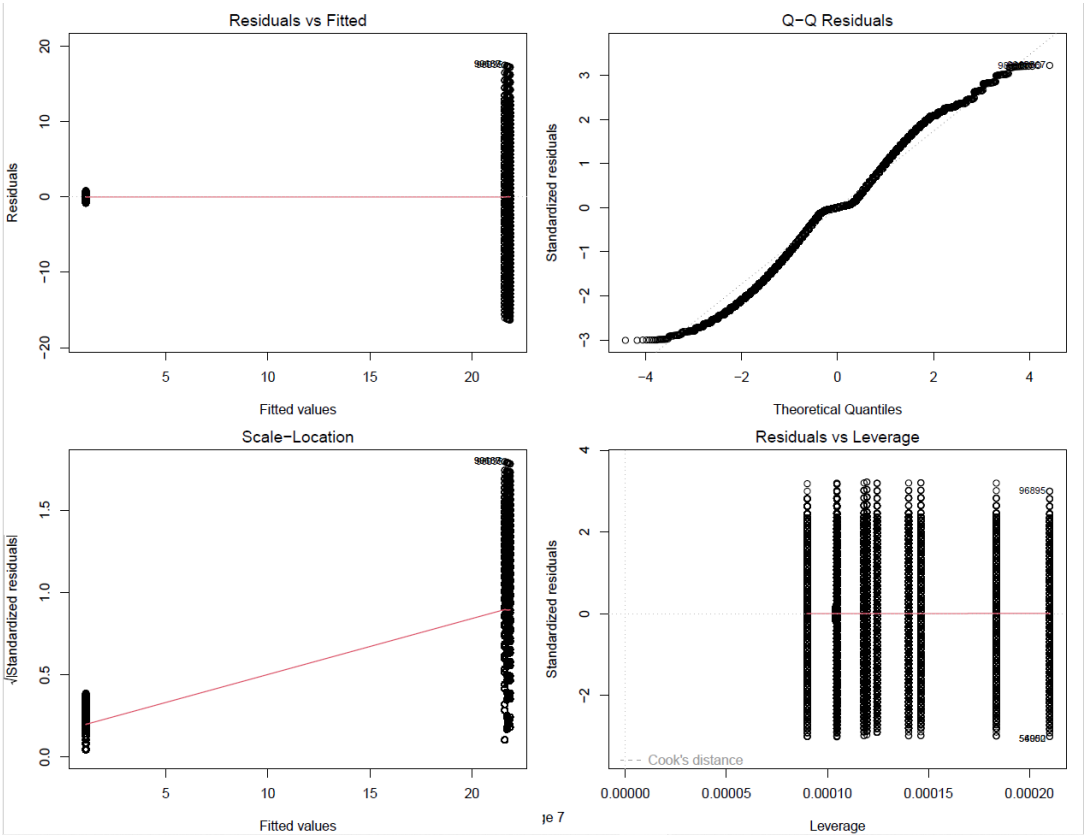


Figure 7.2 ANOVA Plots

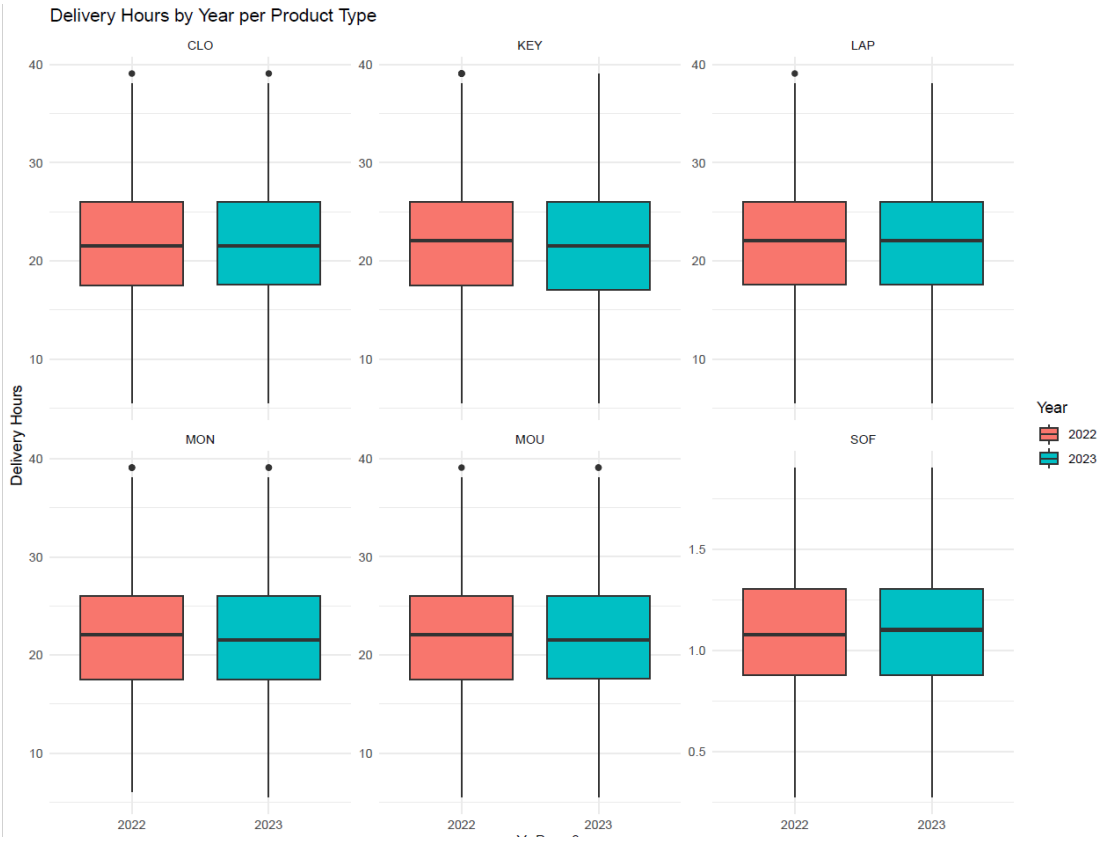


Figure 7.3 Mean Picking vs Delivery Hours

It can be seen in table 2 that both Year ($p = 0.03$) and Product Type ($p < 0.000$) significantly affect delivery performance, while their interaction is not significant ($p = 0.099$). On the other hand, Levene's Test ($p < 0.000$) and Box's M Test ($p < 0.000$) suggest some inequality in variances and covariances, warranting caution in further interpretation. The MANOVA confirmed significant effects of both Year ($p = 0.011$) and Product Type ($p < 0.000$) on the joint outcome of picking and delivery hours.

5.1) Recommendations & Conclusion:

The above results mirror the SPC findings from Section 2 of this report, indicating that systematic product-level and year-to-year differences (not random variation) drive process shifts. It would therefore be recommended to use these findings to target specific products or processes to facilitate improvement and refinement.

6.) Part 7 Reliability of Service

Information was provided by a car rental company where number of service days were contrasted against number of employees working. The employee count was 12-16 with a cost of R25 000 per employee and a loss of R20 000 per ineffective work day. The problem was modelled using an empirical method and binomial method to obtain the following results and recommendations:

Method	ExtraStaff	ProbReliable	ReliableDays	AnnualBenefit	AnnualCost	NetProfit
Empirical	1	0.9848866	359.4836	459697.7	300000	159697.73
Binomial	1	0.9909288	361.6890	398286.3	300000	98286.34

Table 3 Reliability Summary Table

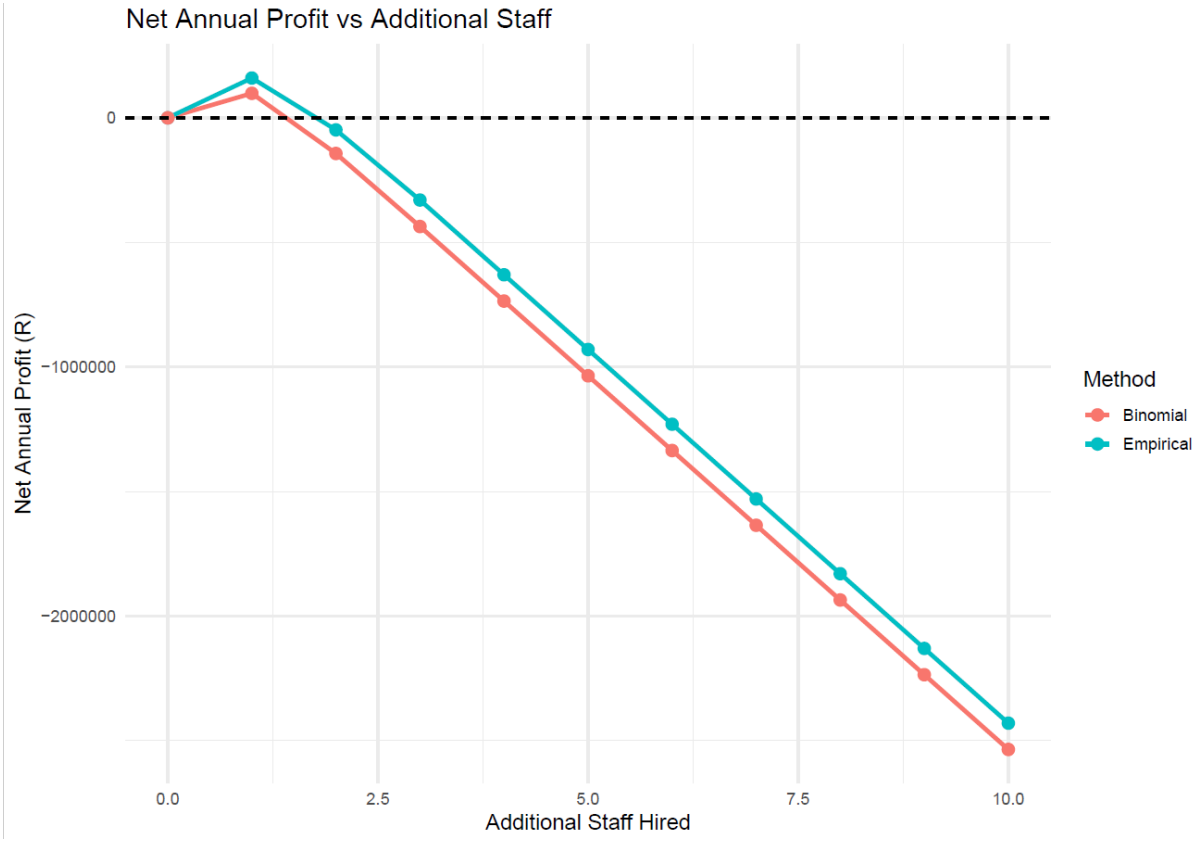


Figure 8.1 Net Annual Profit vs Additional Staff

Both methods yielded the result that an additional employee should be hired thus raising the total staff count to 17 (table 3).

6.1) Recommendations & Conclusion:

It was therefore found that an additional staff member should be hired to ensure that reliable service is maintained while incurring the lest loss in earnings.