Quality Assurance 344

ECSA Project Report 2025

By: Nicola Shakerley – 26023490

Date: October 2025

Table of Contents

1. Introduction 3
2. Descriptive Statistics 4
   2.1. Data Loading and Inspection 4
   2.2. Summary Statistics 5
   2.3. Handling Missing Values 6
   2.4. Data Filtering and Sub setting 6
   2.5. Data Visualisation and Exploring Relationships 6
3. Statistical Process Control 13
   3.1. Initialisation of X-bar and S charts for Delivery Times 13
   3.2. Continuation of X-bar and S charts 13
   3.3. Calculation of Process Capability Indices 13
   3.4. Process Control Issues for Rules A – C 15
4. Type I and II Errors and Data Correction 17
   4.1. Type I Errors for Rules A – C 17
   4.2. Type II Error for a Bottling Process 17
   4.3. Correction of Head Office Data 17
5. Profit Optimisation 18
   5.1. Analysis of Data Set timeToServe.csv 18
   5.2. Analysis of Data Set timeToServe2.csv 19
6. DOE and ANOVA 20
   6.1. 20
   6.2. 20
7. Reliability of Service 21
   7.1. Estimated Days of Reliable Service 21
   7.2. Optimisation of Profit 22
8. Conclusion 23
9. References 24
10. Appendices 25
   Appendix A 25
   Appendix B 31

1. Introduction

The following report focuses on analysing and optimising the delivery processes and overall service of a company selling technological devices and accessories. The products offered by this company include laptops (LAP), monitors (MON), keyboards (KEY), software (SOF), mice (MOU), and cloud subscriptions (CLO). The primary goal of this project was to demonstrate the ability to use data analysis techniques and statistical process control (SPC) methods to interpret and multiple data sets and make decisions based on the evidence found. Multiple data sets were given for the project, these included information on products, sales, and customers. The Statistical Process Control Methods included process capability indices such as Cp, Cpk, Cpu, and Cpl values in order to assess the capability of each product type in satisfying the VOC. X-bar and S charts were also created in order to analyse the variability of the delivery processes for each product type. The report then follows on to identify the risks of making Type I and Type II errors with regards to certain rules on the variability of the delivery process data. An ANOVA test is also done to analyse particular differences between certain product groups. The report also includes a section on the optimisation of the staffing level required in a coffee shop in order to maximize the overall profits of the company. This section focuses on minimizing to company's operational costs while at the same time ensuring that the service reliability is kept at a high level. The results acquired allow for a detailed analysis and identification of problem areas in which recommendations are needed in order to improve the success rate of the company.

2. Descriptive Statistics

2.1. Data Loading and Inspection

After the four data sets were read into the file, the following information was gathered on each data set:

products_data:

- **Dimensions**: 60 rows, and 5 columns
- **Structure**: The ProductID, Category, and Description columns are characters; and SellingPrice, and Markup are numerical values
- **Column Names**: ProductID, Category, Description, SellingPrice, and Markup

customer_data:

- **Dimensions**: 5000 rows, and 5 columns
- **Column Names**: CustomerID, Gender, Age, Income, and City

products_HeadOffice:

- **Dimensions**: 360 rows, and 5 columns
- **Column Names**: ProductID, Category, Description, SellingPrice, and Markup

sales2022and2023:

- **Dimensions**: 100000 rows, and 9 columns
- **Column Names**: CustomerID, ProductID, Quantity, orderTime, orderDay, orderMonth, orderYear, pickingHours, deliveryHours

2.2. Summary Statistics

A general summary of each dataset was performed and gave the following results:

products_data:

```
  ProductID            Category          Description          SellingPrice        Markup      
 Length:60          Length:60          Length:60          Min.   :  350.4   Min.   :10.13  
 Class :character   Class :character   Class :character   1st Qu.:  512.2   1st Qu.:16.14  
 Mode  :character   Mode  :character   Mode  :character   Median :  794.2   Median :20.34  
                                                          Mean   : 4493.6   Mean   :20.46  
                                                          3rd Qu.: 6416.7   3rd Qu.:25.71  
                                                          Max.   :19725.2   Max.   :29.84  
```

customer_data:

```
  CustomerID           Gender               Age             Income          City          
 Length:5000        Length:5000        Min.   : 16.00   Min.   :  5000   Length:5000       
 Class :character   Class :character   1st Qu.: 33.00   1st Qu.: 55000   Class :character  
 Mode  :character   Mode  :character   Median : 51.00   Median : 85000   Mode  :character  
                                       Mean   : 51.55   Mean   : 80797                     
                                       3rd Qu.: 68.00   3rd Qu.:105000                     
                                       Max.   :105.00   Max.   :140000                     
```

products_HeadOffice:

```
  ProductID            Category          Description          SellingPrice        Markup      
 Length:360         Length:360         Length:360         Min.   :  290.5   Min.   :10.06  
 Class :character   Class :character   Class :character   1st Qu.:  495.9   1st Qu.:15.84  
 Mode  :character   Mode  :character   Mode  :character   Median :  797.2   Median :20.58  
                                                          Mean   : 4411.0   Mean   :20.39  
                                                          3rd Qu.: 5843.3   3rd Qu.:24.84  
                                                          Max.   :22420.1   Max.   :30.00  
```

sales2022and2023:

```
  CustomerID         ProductID            Quantity       orderTime        orderDay       orderMonth       orderYear   
 Length:100000      Length:100000      Min.   : 1.0   Min.   : 1.00   Min.   : 1.0   Min.   : 1.000   Min.   :2022  
 Class :character   Class :character   1st Qu.: 3.0   1st Qu.: 9.00   1st Qu.: 8.0   1st Qu.: 4.000   1st Qu.:2022  
 Mode  :character   Mode  :character   Median : 6.0   Median :13.00   Median :15.0   Median : 6.000   Median :2022  
                                       Mean   :13.5   Mean   :12.93   Mean   :15.5   Mean   : 6.448   Mean   :2022  
                                       3rd Qu.:23.0   3rd Qu.:17.00   3rd Qu.:23.0   3rd Qu.: 9.000   3rd Qu.:2023  
                                       Max.   :50.0   Max.   :23.00   Max.   :30.0   Max.   :12.000   Max.   :2023  

  pickingHours      deliveryHours    
 Min.   : 0.4259   Min.   : 0.2772  
 1st Qu.: 9.3908   1st Qu.:11.5460  
 Median :14.0550   Median :19.5460  
 Mean   :14.6955   Mean   :17.4765  
 3rd Qu.:18.7217   3rd Qu.:25.0440  
 Max.   :45.0575   Max.   :38.0460  
```

2.3. Handling Missing Values

The following code was run to identify if there were any missing values in the given datasets:

```
colSums(is.na(products_data))
colSums(is.na(customer_data))
colSums(is.na(products_HeadOffice))
colSums(is.na(sales2022and2023))
```

This code indicated that there were no missing values in any of the columns of the given datasets and that all the datasets are complete. This means that the datasets are ready for analysis without the need for cleaning of the data.

2.4. Data Filtering and Sub setting

In order to better understand the data, the following subsets were created:

- Sales were separated by year (2022 and 2023) so that a comparison between the two could be done.
- Customers were filtered by age (in age bands of 10 years) to see the different demographics among the company customers.
- Products with a price of over 500 were extracted and considered to be expensive products.
- The sales during the month of December were isolated to identify if there were any holiday season trends in the data.
- Sales were also filtered by category, such as laptops, to evaluate the different categories of product sales.

The subsets that were created are useful for identifying and exploring trends within different groups or categories and for identifying areas in which a more detailed study should be considered.

2.5. Data Visualisation and Exploring Relationships

Before any of the data was visualised, it was noticed that not all of the products in the product_data csv file were correct. Many of them had the incorrect category compared to the product code in the ProductID column. This was corrected before any analysis was done on the data to ensure the results were accurate.

Table 1 below shows us the average income of customers per city. We can see from the table that Chicago and Miami have the highest average income levels, with values of 82244.48 and 83346.21 respectively. This information is useful for the company when it comes to pricing strategies and marketing methods in the different cities.

City <chr>	average_inc <dbl>
Chicago	82244.48
Houston	80248.62
Los Angeles	80475.21
Miami	83346.21
New York	79752.07
San Francisco	79852.56
Seattle	79947.99

Table 1: Average Income per City

The Pearson Correlation Coefficient between age and income was calculated to be 0.16. This correlation coefficient tells us that as the age increases, income also begins to gently increase. This is highlighted in Figure 1 below which is a scatter plot of Ave versus Income. The gently increasing blue line indicated a positive correlation between the two variables. However, the line is not particularly steep which tells us that although the correlation is positive, it is not an extremely strong positive correlation.

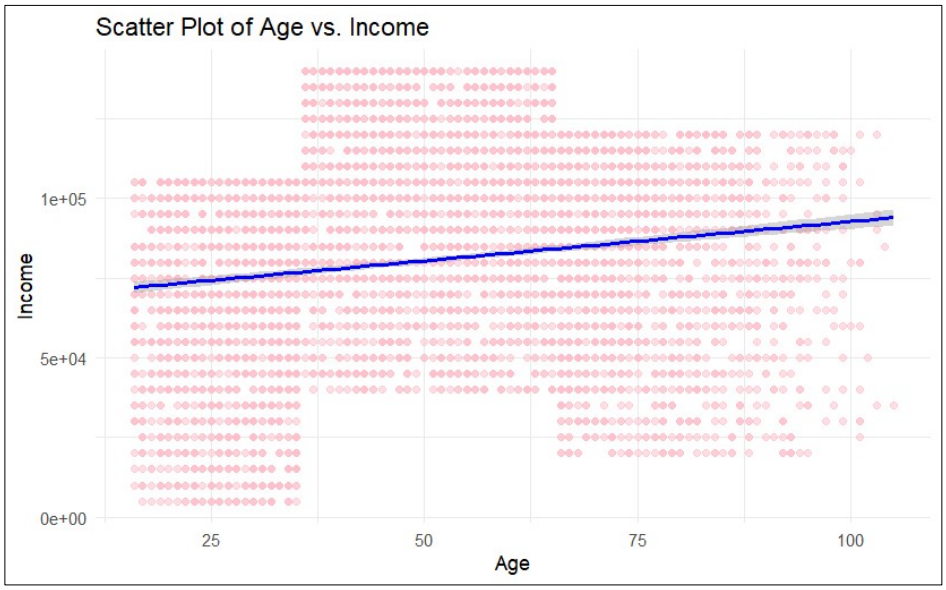


Figure 1: Scatter Plot of Age vs Income

Next, the relationship between purchasing power and gender was investigated. An ANOVA test was conducted, and the results are shown in table 2 below. The P-value for this test is 0.998 which tells us that there is not a significant variation between the purchasing power of the different gender groups.

Table 3 gives us the raw data on the average income for the different gender groups. We can see that the slight difference between male and female is so small that it is insignificant, and we can thus conclude that gender has no influence on the purchasing power of the customers. Figures 2 and 3 below also highlight the similarity between the gender groups. In Figure 3, it can be seen that the median income level is very similar in all three groups, and the interquartile ranges are also very similar.

	Df	Sum Sq	Mean Sq	F value	Pr(>F)
Gender	2	3.795e+06	1.897e+06	0.002	0.998
Residuals	4997	5.494e+12	1.099e+09		

Table 2: ANOVA Results

Gender <chr>	Average_Income <dbl>
Female	80816.20
Male	80770.21
Other	80871.56

Table 3: Average Income per Gender group

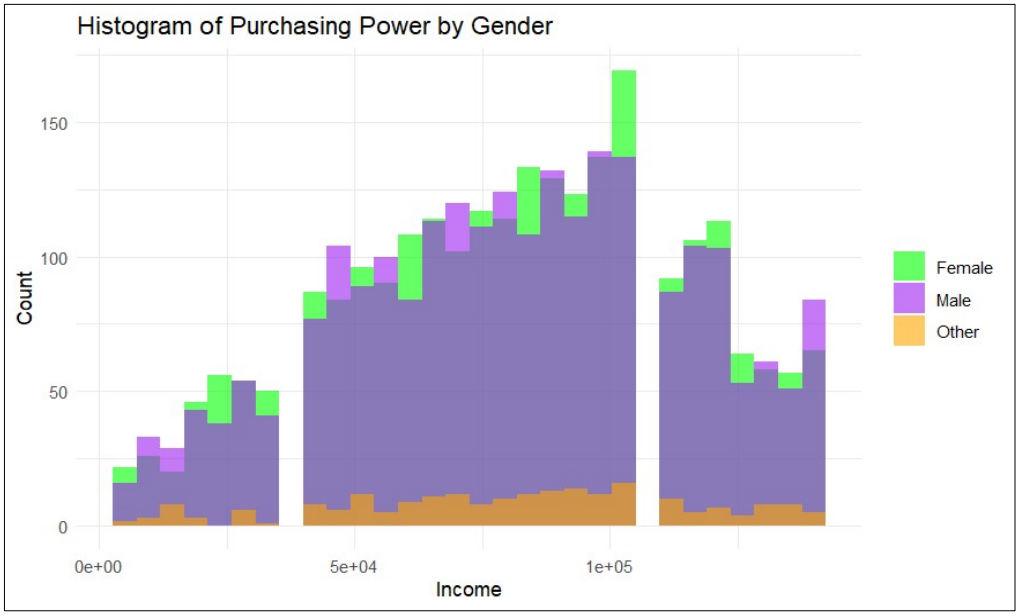


Figure 2: Histogram of Purchasing Power by Gender

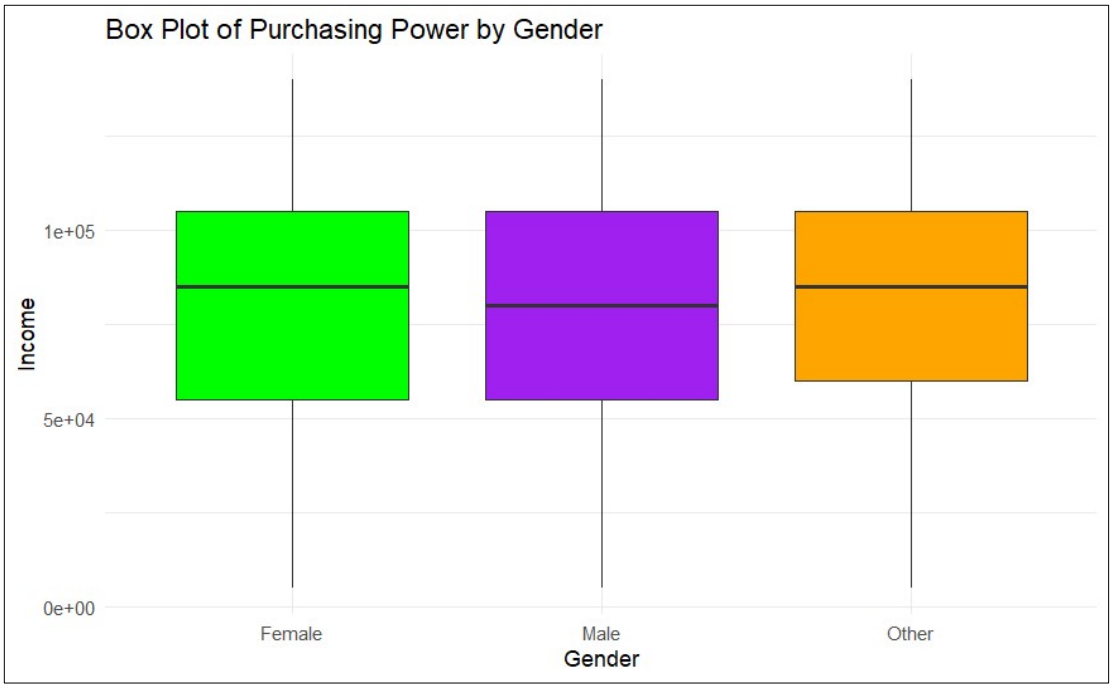


Figure 3: Box Plot of Purchasing Power by Gender

The distribution of customers across different age products is extremely useful when designing marketing strategies and product deals. Table 4 below shows the number of customers in each age bracket. We can see from this table that the age bracket >65 has the highest number of customers, with 1484 customers in total.

age_brackets <fctr>	Count <int>
<18	128
18-25	482
26-35	890
36-45	681
46-55	621
56-65	698
>65	1484
NA	16

Table 4: Number of Customers per Age Bracket

Figures 4 and 5 below show the age distribution of customers across the different age brackets. These figures support the earlier statement that most of the customers are over the age of 65, as we can see the tallest bars in both figures are those of the “>65” age bracket.

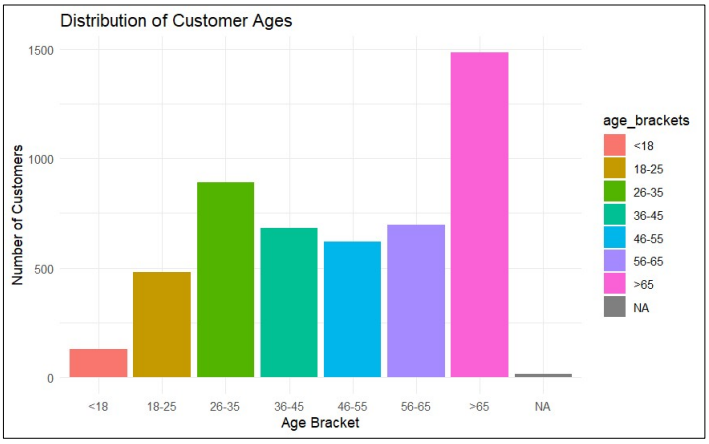


Figure 4: Histogram of Distribution of Customer Ages

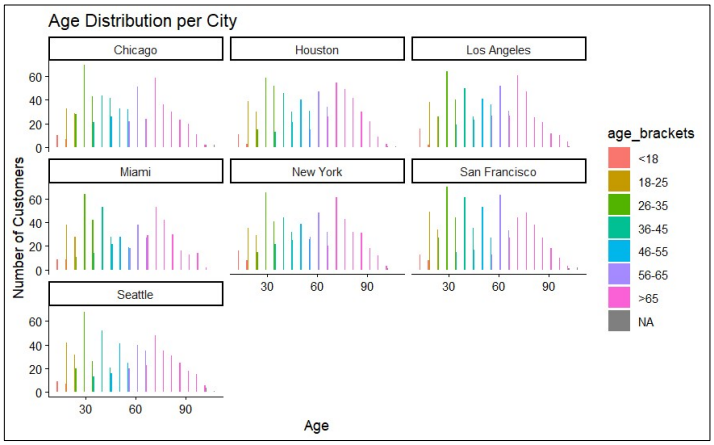


Figure 4: Customer Age Distribution per City

The boxplot in Figure 6 below illustrates the distribution of the selling price across the different product categories, highlighting a significant difference between certain categories. The Laptop (LAP) category has the highest selling price, with the median being around R18 500 and the interquartile range is relatively small which indicates that the pricing between different laptop models is relatively consistent. The Monitor (MON) product category is the next highest, with a median selling price of around R6 000. There are a couple of outliers in this category which could possible represent older or discounted products that were sold. The other four categories CLO, KEY, MOU, and SOF have much lower average selling prices, with all their median selling prices being with the range of R500 – R1 500. All four categories also have very small interquartile ranges indicating that that the pricing of these products is standardised. Overall, the plot shows a big differentiation between the higher value products such as laptops and monitors and the lower priced items.

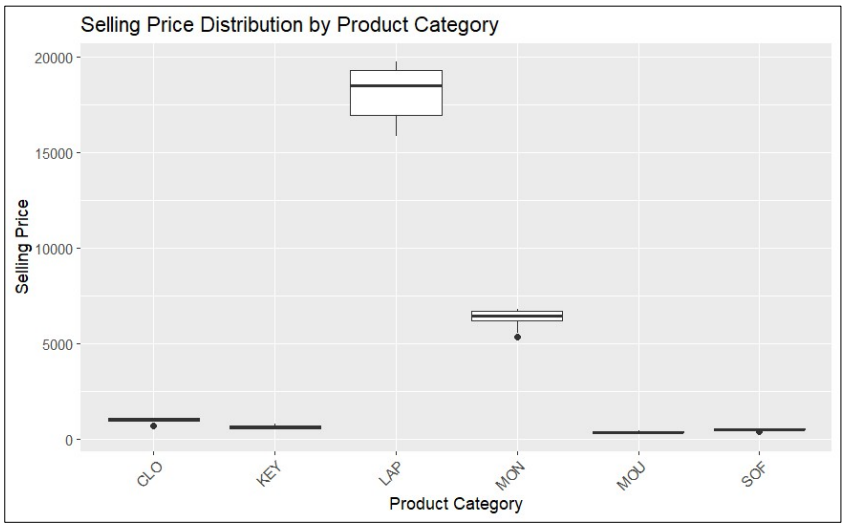


Figure 5: Distribution of Selling Price per Product Category

The bar chart in Figure 7 below illustrates the total revenue generated by each of the product categories. As seen in the graph, the overall revenue generation is clearly dominated by the higher valued products such as laptops and monitors. The Laptop (LAP) category contributes the highest amount of revenue, bringing in just under R2.5 billion, which tells us that laptops are the company's primary driver of revenue. Monitors (MON) are the next highest, brining in just over R1.25 billion. The CLO, KEY, SOF, and MOU categories on the other hand produce a significantly lower amount of revenue to the company, contribution a very small fraction compared to that of laptops and monitors. This indicates that even though the smaller items and accessories may sell in higher quantities, it results in only a smaller amount of revenue due to the much lower selling prices. Overall, this graph highlights the heavy reliance of the company on the sale of laptops and monitors to generate the majority of its revenue.

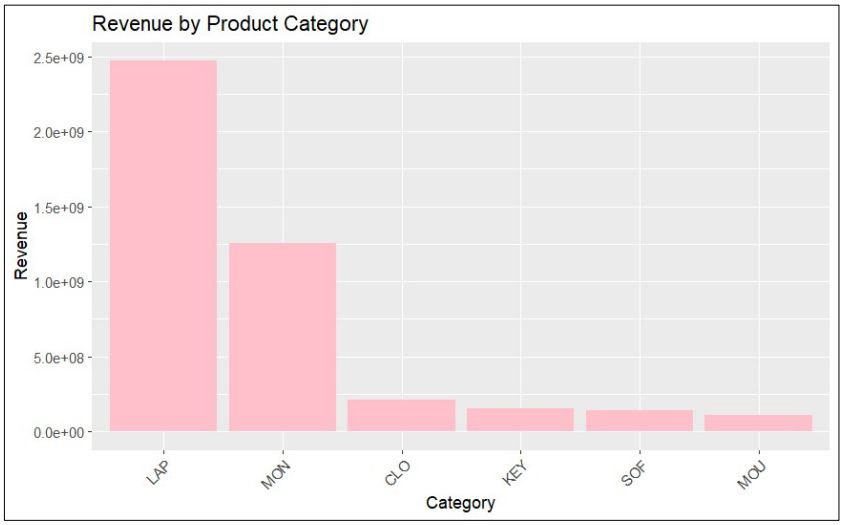


Figure 6: Bar Chart of the Revenue Generated by each Product Category

The line plot in Figure 8 below compares the monthly sales trends across the years of 2022 and 2023. Both years clearly indicate seasonal fluctuations, with peaks in the early to mid-year range, and again towards the end of the year. However, the graph clearly indicates that the sales in 2023 were much less than in 2022, confirming the overall decrease in sales and revenue from 2022 to 2023.

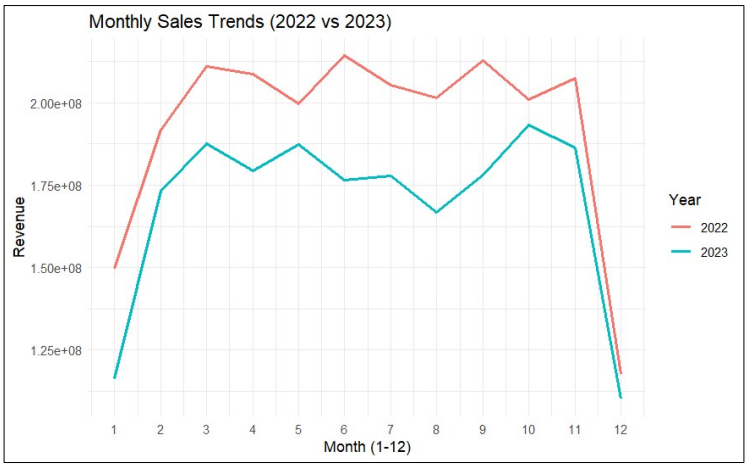


Figure 7: Line plot of the Monthly Sales Trends per Year

The bar chart in Figure 9 further confirms the drop in sales between 2022 and 2023. The volume of sales in 2023 is slightly lower than that of 2022. This decline in sales indicates a possible issue that higher management may need to investigate further into.

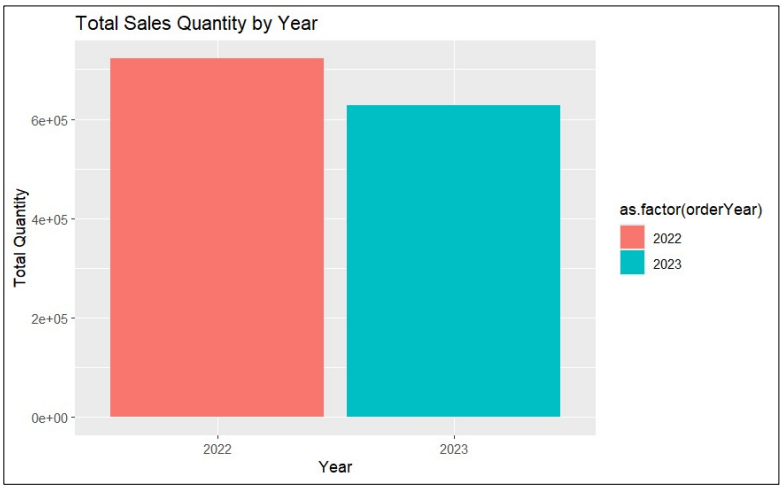


Figure 8: Bar Chart of the Total Sales Quantity per Year

3. Statistical Process Control

3.1. Initialisation of X-bar and S charts for Delivery Times

The X-bar and S charts for all product groups can be seen in Appendix A. Sample sizes of 24 were used and only the first 30 samples were used to draw these graphs. In all of the X-bar charts have a mean delivery time of around 19.2 hours and no points deviate from the control limits indicated by the coloured dashed lines. In all the S charts, the standard deviation sits at around 6, and again, none of the points deviate from the control limits. The centre line is represented by the red dashed line, the outer control limits are represented by the blue dashed line, the 2-sigma control limits are represented by the pink line, and the 1-sigma control limits are represented by the green line.

3.2. Continuation of X-bar and S charts

Following on from 3.1, we have continued to draw samples, still with a sample size of 24, beyond the first 30 samples. The X-bar and S charts for these samples can be seen in Appendix B. It can be seen in all the X-bar charts that the mean delivery time remains around 19.2 hours, and in the S charts the standard deviation also remains at around 6 hours for all the samples. The red dots in the X-bar charts indicate samples that exceed the control limits of the sample. And the blue dots in the S charts indicate samples that exceed the control limits of the sample.

3.3. Calculation of Process Capability Indices

Process capability indices are extremely important is the determination of whether the delivery process is meeting the needs of the customer requirements. The Cp value is the measure of potential capability of a process, and it is measured when it is centred around the Lower Specification Limit (LSL) and the Upper Specification Limit (USL). If the Cp value is 1.33 or higher, it indicated that the process has a good potential to meet the specific requirements. The Cpk value is a measure of the actual capability of a process, and it looks into how well the process is centred between the LSL and the USL.

The product group MOU has a mead delivery time of 19.3 hours and a standard deviation of 5.83 hours, which is indicative of a relatively high variation in the delivery times of this product group. The Cp value is 0.92 and the Cpk value is 0.73. Both of these values are less that 1 which tells us that this process is not capable of meting the expectations of our customers. The low Cpk value on 0.73

suggests to us that the mean of the process is not centred between the specification limits, and thus a certain portion of deliveries falls outside of the desired range of acceptance. Overall, the deliveries of the Mou product group are relatively inconsistent and improvement of the process is needed in order for it to meet the VOC requirements.

For the product group KEY, the mean delivery time is 19.28 hours, and the standard deviation is 5.82. These values are very similar to those of the MOU group and suggest that there is a large variance in the delivery times of these products. The Cp value is also 0.92 and the Cpk value is also 0.73. This indicated, as said above, that the process is not capable of meeting the requirement. The low Cpk value and the high variability suggests that the process is not stable or reliable enough to be consistently delivering products within the specified limits. From all given information it can be deduced that the product group KEY is also not capable of satisfying the VOC.

The SOF category performs much better. It has a mean of 0.96 and a much lower standard deviation of 0.29, which indicated that there is minimal variation in the delivery times of this category. The Cp value is 18.14 and the Cpk value is 1.08. Both these values are very high and indicate that this process is highly capable of meeting the requirements and it is well centred within the specification limits. These high values tell us that SOF is consistently meeting and possibly exceeding the VOC expectations for the delivery process.

The product group CLO has a mean delivery time of 19.23 hours and a standard deviation of 5.94 hours, which indicates that the variability between the delivery performance is quite high. The Cp value is 0.9 and the Cpk value is 0.72. These values suggest that the process is not capable of meeting the VOC requirements. Although the process is relatively consistent, it is too variable and it is not centred as well as it should be, meaning that it fails to meet the requirements and some improvements in the control of the process is needed in order to reduce the variability and meet the necessary requirements.

The product group LAP has a mean of 19.61 hours and a standard deviation of 5.93, which also indicates a high variability in the performance of the delivery process. The Cp value is 0.9 and the Cpk value is 0.7. Both these values are below 1, which indicated that the process is not capable of

meeting the VOC requirements. These numbers tell us that the process mean is not well aligned withing the required specifications. The deliveries of the LAP products are inconsistent, and improvements are needed in order to meet the VOC requirements.

The product category MON has a mean delivery time of 19.41 hours and a standard deviation of 6 hours, which makes it the most highly variable category out of all the products. The Cp value is 0.89 and the Cpk value is 0.7, which indicate that this process too is not capable of meeting the VOC requirements. Similar to the other product groups, MON also shows poor centring within the specification and high variability between the delivery performance, which tells us that the delivery times are often greater than the specification limits and a large improvement is needed.

```
Product Category: MOU
Mean Delivery Hours: 19.3 hours
Standard Deviation: 5.83 hours
Cp: 0.92
Cpu: 0.73
Cpl: 1.1
Cpk: 0.73
Product Category: MOU is not capable of satisfying the VOC.

Product Category: KEY
Mean Delivery Hours: 19.28 hours
Standard Deviation: 5.82 hours
Cp: 0.92
Cpu: 0.73
Cpl: 1.1
Cpk: 0.73
Product Category: KEY is not capable of satisfying the VOC.

Product Category: SOF
Mean Delivery Hours: 0.96 hours
Standard Deviation: 0.29 hours
Cp: 18.14
Cpu: 35.19
Cpl: 1.08
Cpk: 1.08
Product Category: SOF is capable of satisfying the VOC.
```

```
Product Category: CLO
Mean Delivery Hours: 19.23 hours
Standard Deviation: 5.94 hours
Cp: 0.9
Cpu: 0.72
Cpl: 1.08
Cpk: 0.72
Product Category: CLO is not capable of satisfying the VOC.

Product Category: LAP
Mean Delivery Hours: 19.61 hours
Standard Deviation: 5.93 hours
Cp: 0.9
Cpu: 0.7
Cpl: 1.1
Cpk: 0.7
Product Category: LAP is not capable of satisfying the VOC.

Product Category: MON
Mean Delivery Hours: 19.41 hours
Standard Deviation: 6 hours
Cp: 0.89
Cpu: 0.7
Cpl: 1.08
Cpk: 0.7
Product Category: MON is not capable of satisfying the VOC.
```

3.4. Process Control Issues for Rules A – C

In this question the data was analysed in order to identify any process control issues according to three different rules. The rules were as follows:

Rule A: 1 s sample outside of the upper +3 sigma control limits for all types of products. Any sample that sits outside of the +3-sigma limit is indicative of extreme variation, which means the process may be out of control.

Rule B: Most consecutive samples of s between the – 1 and +1 sigma control limits for all types of products. If there is a long string of consecutive s values, it suggests that the variability of the process is consistent. This is an indicator of good process control.

Rule C: 4 consecutive X-bar samples outside of the upper, second control limit for all types of products. An upward trend in the process mean is indicated by 4 or more X-bar samples fall above the upper 2-sigma line.

The results of this investigation are shown below:

```
Product Type: MOU
Rule A (S > UCL_3σ): total = 1
  First 3 / Last 3 sample numbers: 592
Rule B (longest run with S in ±1σ band): length = 16
  Run from sample 672 to 687
Rule C (≥4 consecutive X̄ above +2σ): total runs = 25
  First 3 / Last 3 runs (start-end, length):
    • 194–197 (length=4)
    • 233–240 (length=8)
    • 249–252 (length=4)
    • 768–775 (length=8)
    • 777–805 (length=29)
    • 807–860 (length=54)

Product Type: KEY
Rule A (S > UCL_3σ): total = 0
Rule B (longest run with S in ±1σ band): length = 15
  Run from sample 730 to 744
Rule C (≥4 consecutive X̄ above +2σ): total runs = 27
  First 3 / Last 3 runs (start-end, length):
    • 99–102 (length=4)
    • 112–117 (length=6)
    • 172–175 (length=4)
    • 687–696 (length=10)
    • 698–724 (length=27)
    • 726–746 (length=21)
```

```
Product Type: SOF
Rule A (S > UCL_3σ): total = 0
Rule B (longest run with S in ±1σ band): length = 21
  Run from sample 659 to 679
Rule C (≥4 consecutive X̄ above +2σ): total runs = 27
  First 3 / Last 3 runs (start-end, length):
    • 133–136 (length=4)
    • 202–205 (length=4)
    • 237–241 (length=5)
    • 774–801 (length=28)
    • 803–840 (length=38)
    • 842–864 (length=23)

Product Type: CLO
Rule A (S > UCL_3σ): total = 0
Rule B (longest run with S in ±1σ band): length = 35
  Run from sample 474 to 508
Rule C (≥4 consecutive X̄ above +2σ): total runs = 14
  First 3 / Last 3 runs (start-end, length):
    • 122–125 (length=4)
    • 179–183 (length=5)
    • 192–200 (length=9)
    • 557–602 (length=46)
    • 604–626 (length=23)
    • 628–649 (length=22)
```

```
Product Type: LAP
Rule A (S > UCL_3σ): total = 0
Rule B (longest run with S in ±1σ band): length = 19
  Run from sample 116 to 134
Rule C (≥4 consecutive X̄ above +2σ): total runs = 11
  First 3 / Last 3 runs (start-end, length):
    • 119–122 (length=4)
    • 129–140 (length=12)
    • 153–167 (length=15)
    • 348–357 (length=10)
    • 359–372 (length=14)
    • 374–425 (length=52)

Product Type: MON
Rule A (S > UCL_3σ): total = 0
Rule B (longest run with S in ±1σ band): length = 34
  Run from sample 238 to 271
Rule C (≥4 consecutive X̄ above +2σ): total runs = 22
  First 3 / Last 3 runs (start-end, length):
    • 134–137 (length=4)
    • 171–177 (length=7)
    • 179–186 (length=8)
    • 566–608 (length=43)
    • 610–613 (length=4)
    • 615–619 (length=5)
```

4. Type I and II Errors and Data Correction

4.1. Type I Errors for Rules A – C

The likelihood of a Type I error occurring for each of the three rules in 3.4. were calculated and the results obtained were as follows:

Rule <chr>	Description <chr>	Type_I_Error <chr>	Percentage <chr>	Frequency <chr>
A	1 sample > UCL_3σ (s-chart)	1.35e-03	0.135%	1 in 741
B	Consecutive samples within ±1σ (good control)	N/A (indicates good control)	N/A	More = Better
C	4 consecutive samples > UCL_2σ (X-bar)	2.679e-07	2.679e-05%	1 in 3,733,054

For **Rule A**, the likelihood of a Type I error is 0.135%, which means that a "false alarm" is expected every once in 741 samples. From this it can be deduced that the rule has a very small chance of falsely identifying a stable process as out-of-control.

For **Rule B**, the likelihood of a type I error occurring is 0 because the pattern doesn't represent and out-of-control condition.

For **Rule C**, the likelihood of a Type I error occurring is 0.00002679% which suggests that this rule will very rarely, if ever, incorrectly identify a stable process as out-of-control.

4.2. Type II Error for a Bottling Process

Given the following information on a bottle filling process: a mean and CL of 25.05 litres, a UCL of 25.089 litres, an LCL of 25.011 litres, an average fill volume of 25.028 litres, and a standard deviation of 0.017 litres, the likelihood of making a Type II error for this scenario was calculated. The probability of making a Type II error was calculated to be 84.118%. This is an extremely high value for a type II error, and it indicates that the test is not good at identifying true positives and thus, majority of meaningful results are overlooked. The test needs adjustments such as increases in the sample size to reduce the likelihood of missing differences or adjustment of the significance level.

4.3. Correction of Head Office Data

It was previously noticed in the data analysis section of the project, that the products_data file had the incorrect category names and did not align with the ProductID column of the data. This was corrected before the data analysis was done to ensure accurate visualisation of the trends in the data.

In this section of the project, the products_HeadOffice data was also corrected. All of the incorrectly labelled “NA” ProductID’s were corrected, and the selling price and markup values were standardized using the first 10 values as a reference. The new data files were rewritten as csv files and saved as "products_data2025.csv" and "products_Headoffice2025.csv" respectively.

Table 5 below shows us the first six lines of the new products_HeadOffice data file. We can see here that all the ProductID codes have been aligned with the correct name in the category column. The selling price has also been updated, and the markup value standardised throughout the data file. Since the products_data file was already updated before the data analysis was completed, none of the changes made to the products_HeadOffice file will have any effect on the sales or revenue metrics as the sales data file was correct when it was first analysed.

	ProductID <chr>	Category <chr>	Description <chr>	SellingPrice <dbl>	Markup <dbl>
1	SOF001	Software	coral matt	511.53	25.05
2	SOF002	Software	cyan silk	505.26	10.43
3	SOF003	Software	burlywood marble	493.69	16.18
4	SOF004	Software	blue silk	542.56	17.19
5	SOF005	Software	aliceblue wood	516.15	11.01
6	SOF006	Software	black silk	478.93	16.99

Table 5: Updated Products Data

5. Profit Optimisation

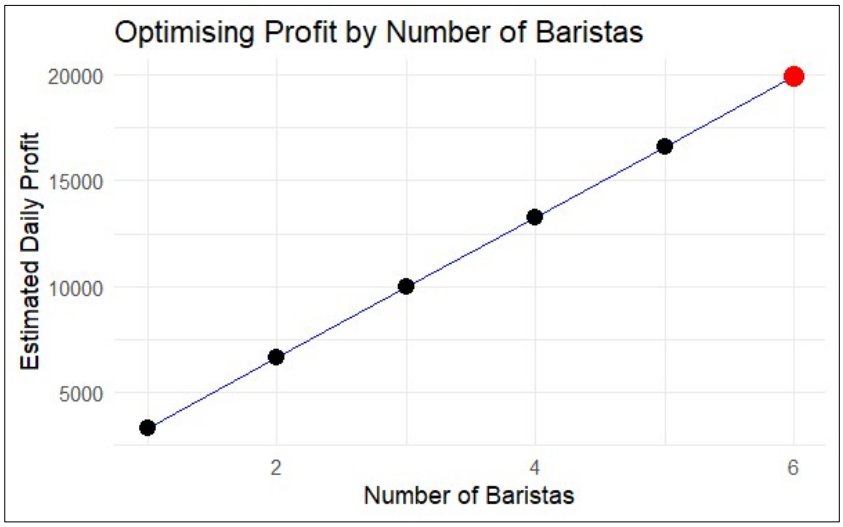
5.1. Analysis of Data Set timeToServe.csv

This question required us to determine the optimal number of baristas required in a coffee shop in order to maximise the net profit for the company, while simultaneously ensuring that the reliability of service in the shop is still very high. To calculate the net profit for each scenario, the mean service time was calculated for each different number of baristas, and an estimate of 8 hours per working day was used to estimate the number of customers served per day. Using the given information of R30 material profit per customer served and R1000 cost per barista hired, the total revenue, cost, and net profit was calculated for each scenario. The following table is a summary of all these values.

Baristas <int>	Customers_Served <dbl>	Revenue <dbl>	Cost <dbl>	Net_Profit <dbl>
1	143.8879	4316.636	1000	3316.636
2	287.5084	8625.253	2000	6625.253
3	432.3562	12970.686	3000	9970.686
4	576.2261	17286.784	4000	13286.784
5	720.6876	21620.629	5000	16620.629
6	863.4220	25902.661	6000	19902.661

For the first data set “timeToServe.csv” was analysed using the model, and the graph below was drawn showing the estimated daily profit against the number of baristas working in the shop. There is a clear trend show in this graph, the estimated daily profit increases linearly as the number of baristas working increases. This trend suggests that the more baristas hired, this higher the daily profit will be. In this scenario, a maximum daily profit of R19902.66 is achieved when there are 6 baristas working in the shop. This point is highlighted on the graph with the red dot.

From the model, we can deduce that 6 baristas in this case provided a good balance between the operational costs of staffing and the service reliability of the coffee shop. Having 6 baristas working at the same time is crucial for the business to ensure that their reliability levels are high, especially during peak service hours when there is a large volume of customers.



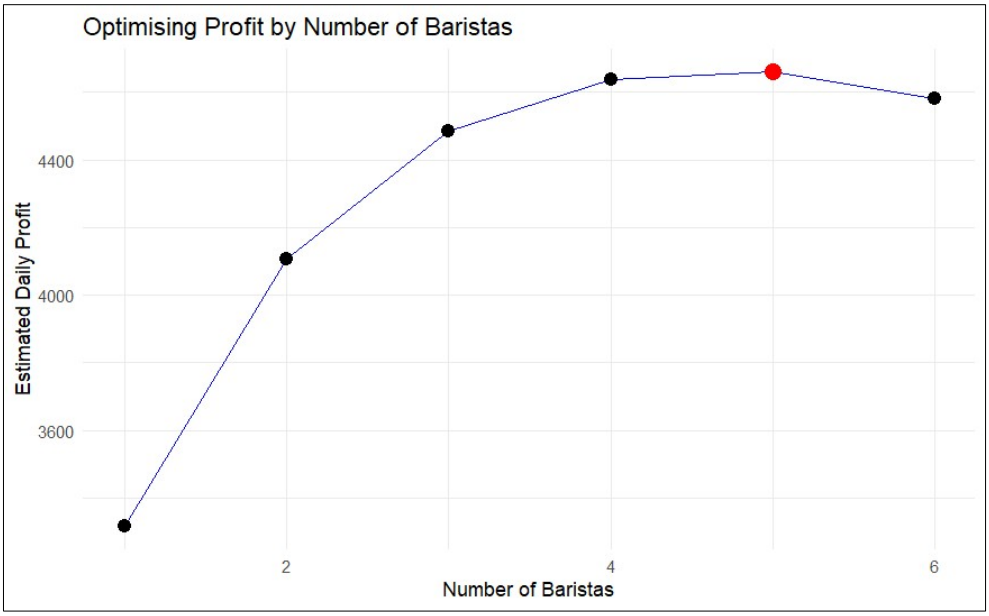
5.2. Analysis of Data Set timeToServe2.csv

The same model was then used to analyse another set of data, “timeToServe2.csv”. The summary table of this new data set is as follows:

Baristas <int>	Customers_Served <dbl>	Revenue <dbl>	Cost <dbl>	Net_Profit <dbl>
1	143.8785	4316.354	1000	3316.354
2	203.5125	6105.376	2000	4105.376
3	249.4783	7484.348	3000	4484.348
4	287.9560	8638.681	4000	4638.681
5	322.0181	9660.543	5000	4660.543
6	352.7565	10582.695	6000	4582.695

We can see from this table that this data set has a different optimisation scenario compared to the first data set. This coffee shop’s optimal solution would be to hire 5 baristas instead of 6, ultimately

resulting in a net profit of R4660.54. The graph below highlights the relationship between the number of baristas and the estimated daily profit, and again the optimal point is highlighted by the red dot.



6. DOE and ANOVA

6.1.

The following question was analysed by the ANOVA test:

Do the delivery hours of the product groups “MOU” (mouse) and “KEY” (keyboards) show a significant difference?

- Null Hypothesis (H0): The average delivery times between the product groups “MOU” and “KEY” show no significant difference.
- Alternative Hypothesis (H1): The average delivery times between the product groups “MOU” and “KEY” show a significant difference.

6.2.

The results obtained from the ANOVA test were as follows:

product_group <chr>	mean_deliveryHours <dbl>	sd_deliveryHours <dbl>
KEY	21.74372	6.091467
MOU	21.79001	6.135815

	Df	Sum Sq	Mean Sq	F value	Pr(>F)
product_group	1	21	20.57	0.55	0.458
Residuals	38580	1442752	37.40		

It can be seen that the mean delivery hours and the standard deviation of both categories “KEY” and “MOU” are extremely similar to each other. From this, we can deduce that the variation between the delivery hours in both groups are relatively similar.

The F-value for this ANOVA test was 0.55. The F-value is the ratio of comparison between the variance within each product group and the variance between the two different product groups. The F-vale calculated is relatively small which indicates that there is little difference between the delivery times between the two different product types.

The P-value for this particular test was given as 0.458, which is much higher than the usual significance level of 0.05. Because this P-value is so high, we can conclude that we failed to reject the null hypothesis and thus there is no evidently significant difference in the average delivery times of the product groups “KEY” and “MOU”.

Overall, the results gathered from this ANOVA test tell us that the delivery process for both product groups is consistent and is well controlled. Neither one of the product groups has significantly different delivery times which indicated that both delivery processes are similar and consistent with each other.

7. Reliability of Service

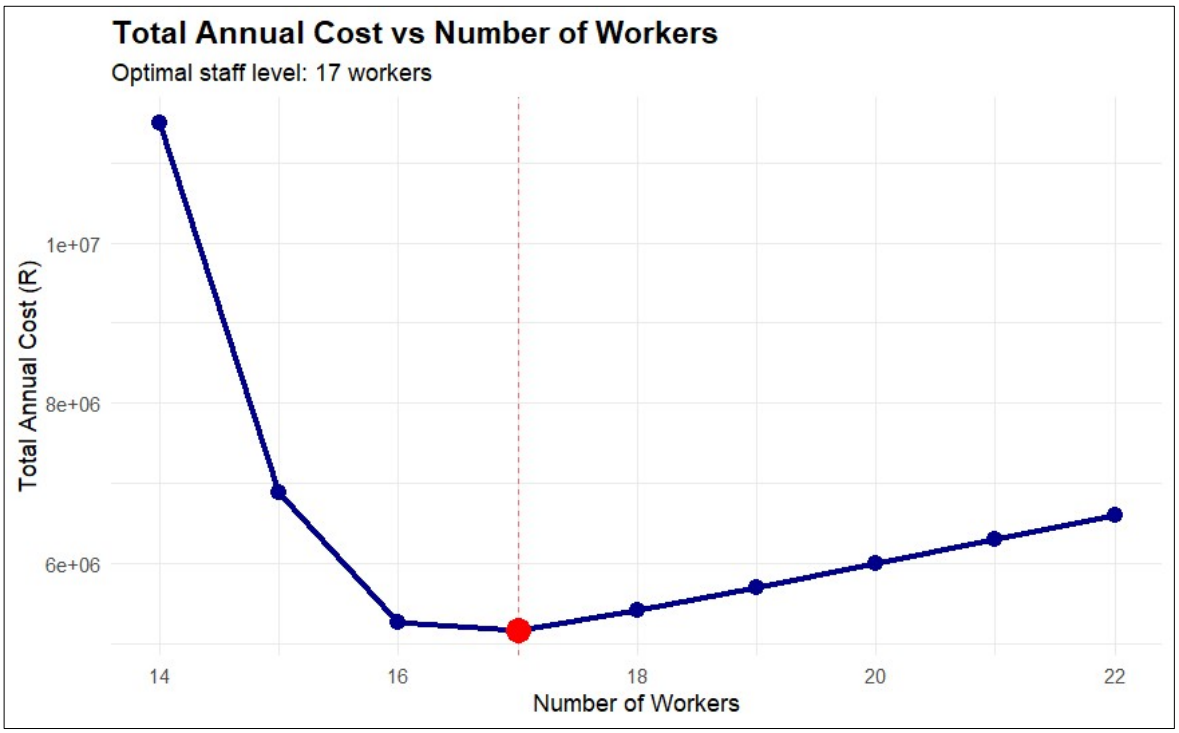
7.1. Estimated Days of Reliable Service

Given the information and the bar chart of number of days with a certain number of workers present, it was calculated that there would be 366 days of reliable service per year. We were told that we would experience problems if there were less than 15 workers on duty, which meant that the 96 days with 15 workers and the 270 days with 16 workers were the number of days with reliable service out of the total 397 days. The sum of reliable days was then divided by the total days and then multiplied by 365 days per year $[\frac{366}{397} \times 365]$. This gave us a value of 366.4987, which was then rounded down to give us the final value of 366 days of reliable service per year.

7.2. Optimisation of Profit

To optimise the profitability of the company, the data given on worker attendance was modelled as a binomial distribution problem, and the parameters were estimated from the given data. The result given was that the optimal number of workers needed is 17 workers in a day. The total cost at this optimal level was calculated to be R5 166 220.

The graph below shows the change in Total Annual Cost for the different number of workers present per day. As highlighted by the red point and dashed line, the optimal number of workers present per day is 17, this is where the total annual cost is at its minimum. The outcome aligns with the Taguchi Loss Principle which states that even the smallest deviations from the target value will cause a loss. This is highly evident in the graph below as we can see on either side of the red optimal line that the cost immediately starts to increase. When there are too few workers present the total cost increases due to the increased loss in sales experience, and when there are too many workers present the cost also increases due to the extra salaries needing to be paid. Thus, by having 17 workers present per day, the cost between the loss of sales and the salaries to be paid is balanced and the total profit for the company is at a maximum.




```
Optimal number of workers: 17
Total annual cost at optimal level: R 5,166,220
```

8. Conclusion

This report gave an all-inclusive of the technology company's data through the use of descriptive statistics, process capability indices, SPC analysis, and optimisation methods. The SPC analysis method indicated that the software product group (SOF) was the only group that was capable of meeting the VOC specifications, it also showed a very consistent delivery process which is essential for meeting customer expectations. The five other product groups, MON, MOU, CLO, KEY, and LAP, however, were not capable of meeting the VOC specifications, which suggested that the delivery process control for these product groups needs to be improved in order to enhance the consistency of delivery. In part four, the evaluation of the likelihood of Type I and Type II errors showed a good understanding of process reliability, while at the same time minimising risks and false negatives that may cause important results to be overlooked. The data correction tasks ensured that the data used by the company and for the analysis was consistent and accurate. The profit optimisation models further highlighted the importance of balancing operational costs with service reliability in order to generate the highest possible revenue for a company. Overall, the project highlighted the importance of effective resource allocation and continuous improvement in maintaining a high-quality service performance and enhancing profitability of a business, while also ensuring the requirements of customers are met with accuracy and consistency.

9. References

Ahluwalia, S., 2025. *Type 1 And Type 2 Errors In A/B Testing And How To Avoid Them.* [Online]
Available at: https://vwo.com/blog/errors-in-ab-testing/
[Accessed 21 October 2025].

eLeaP Editorial Team, 2025. *Analysis of Variance (ANOVA) in Quality Management Systems: A Complete Guide.* [Online]
Available at: https://quality.eleapsoftware.com/glossary/analysis-of-variance-anova-in-quality-management-systems-a-complete-guide/
[Accessed 18 October 2025].

Lean Six Sigma Definition, n.d. *Taguchi Loss Function.* [Online]
Available at: https://www.leansixsigmadefinition.com/glossary/taguchi-loss-function/#google_vignette
[Accessed 14 October 2025].

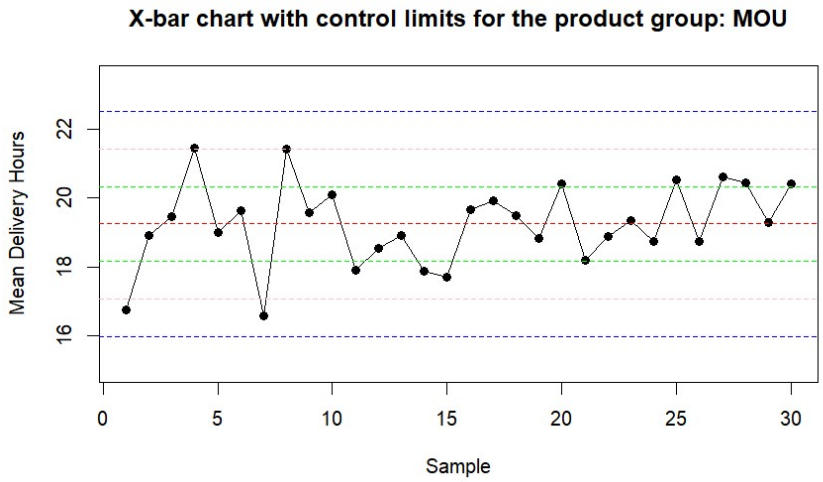
Six Sigma, 2024. *Six Sigma Principles. Process Capability Index (Cpk) for Business Success.* [Online]
Available at: https://www.6sigma.us/process-improvement/process-capability-index-cpk/
[Accessed 16 October 2025].

SixSigma, 2024. *SPC Charts: The Ultimate Guide to Statistical Process Control & Quality.* [Online]
Available at: https://www.6sigma.us/six-sigma-in-focus/spc-charts/
[Accessed 19 October 2025].

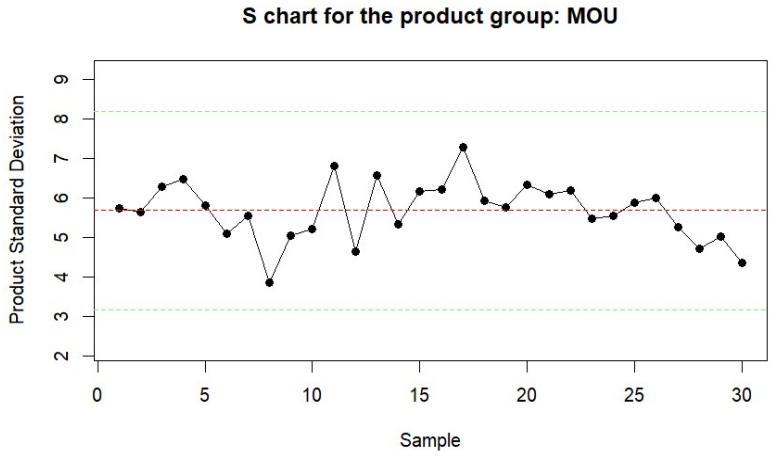
10. Appendices

Appendix A

Product Group: MON

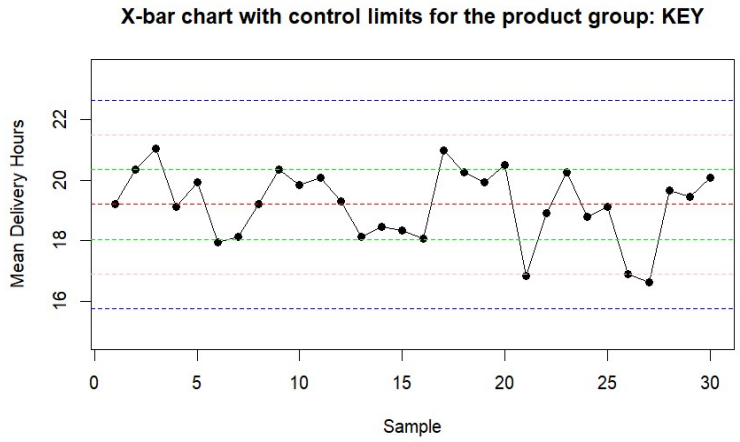


Graph A1

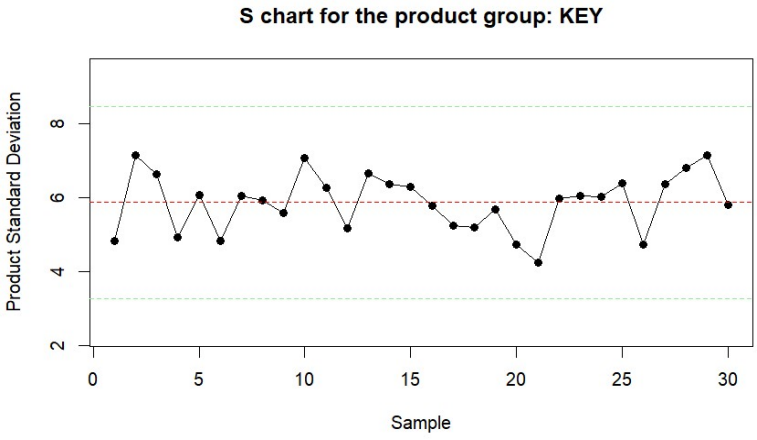


Graph A2

Product Group: KEY

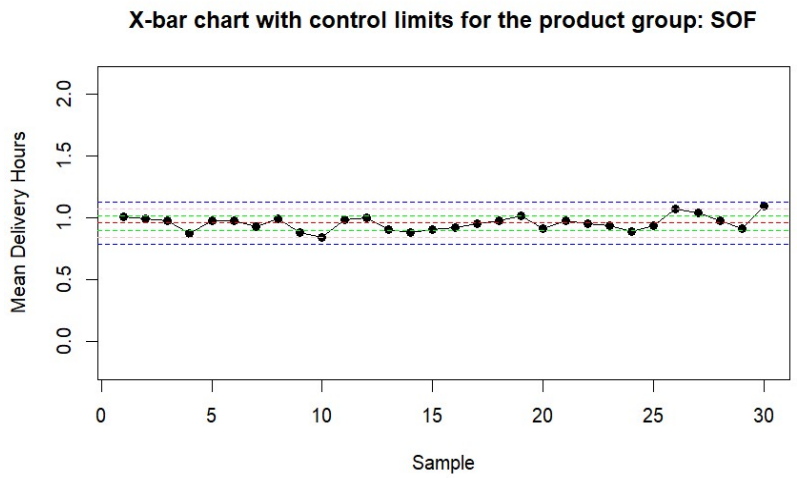


Graph A3

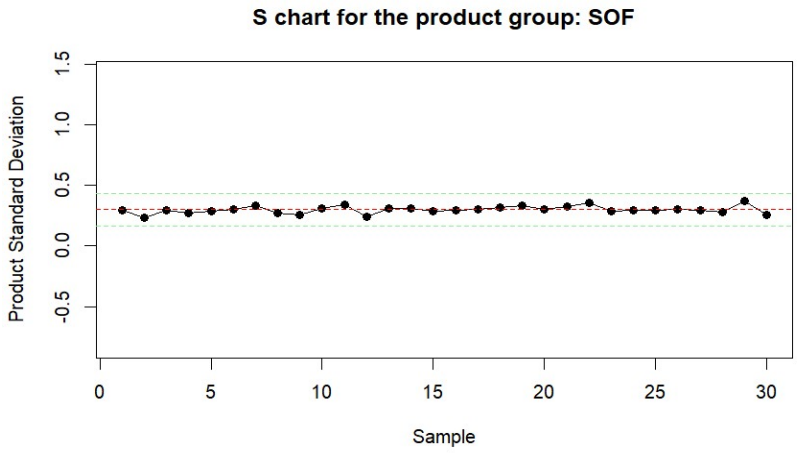


Graph A4

Product Group: SOF

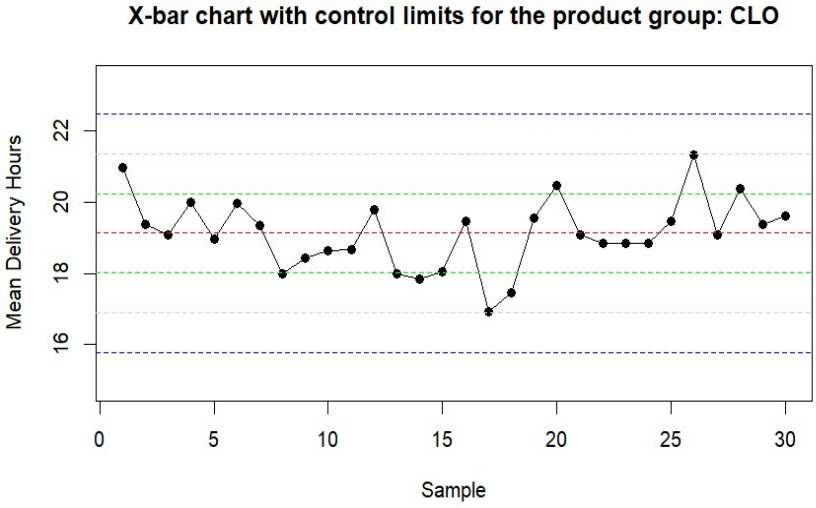


Graph A5

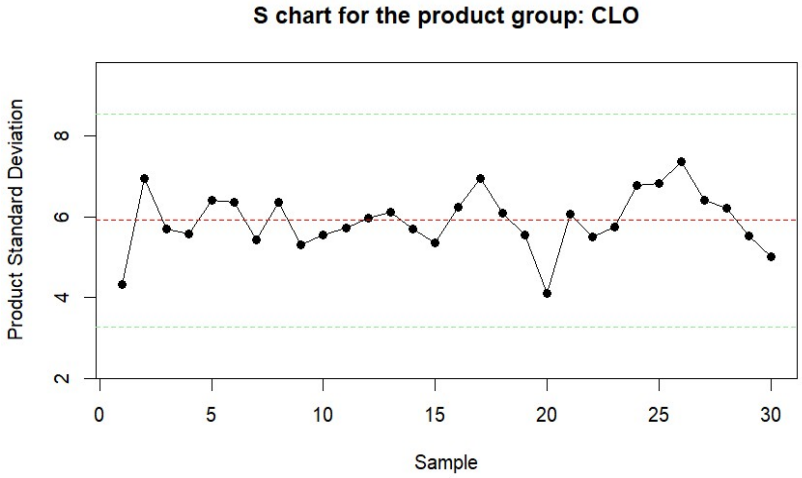


Graph A6

Product Group: CLO

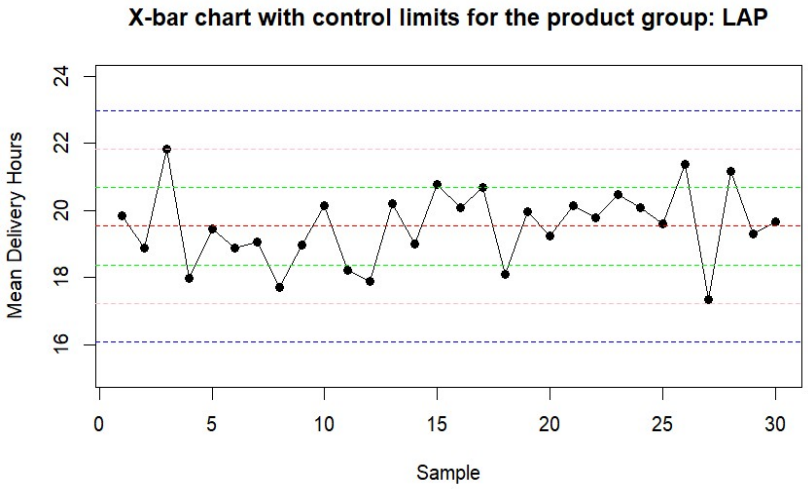


Graph A7

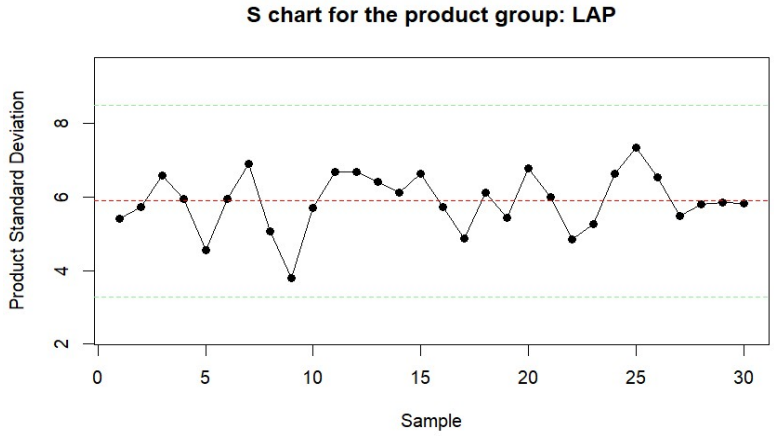


Graph A8

Product Group: LAP

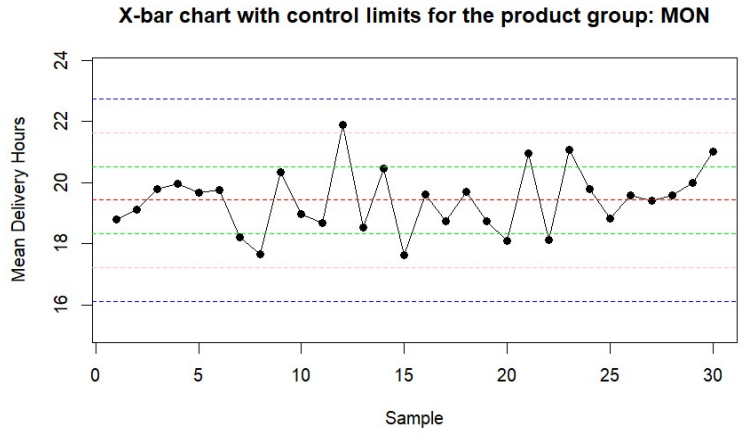


Graph A9

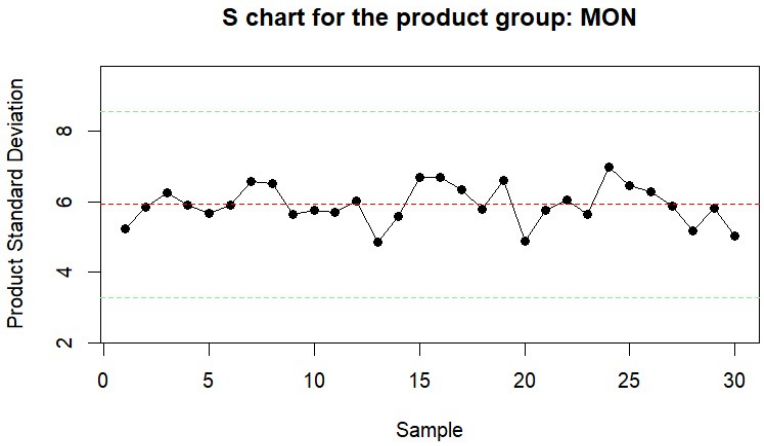


Graph A10

Product Group: MON



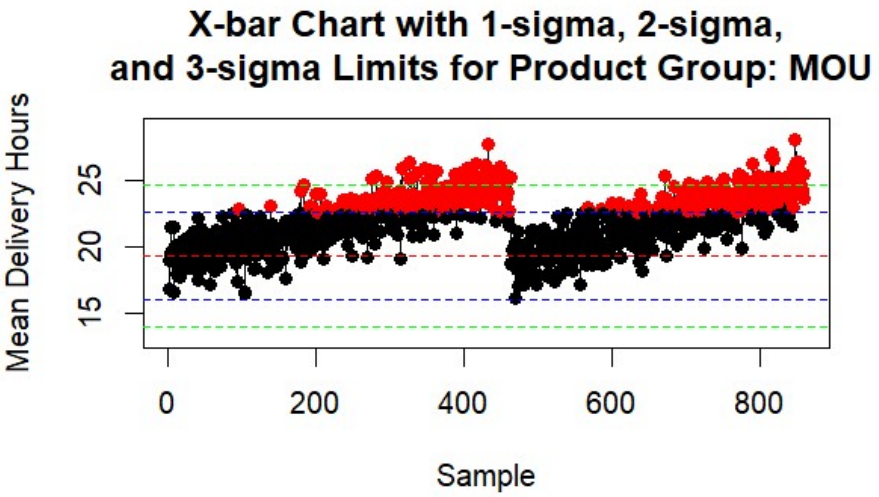
Graph A11



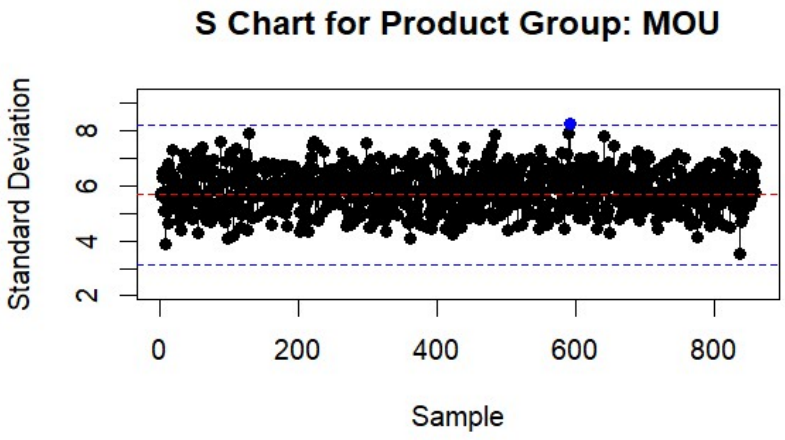
Graph A12

Appendix B

Product Group: MON

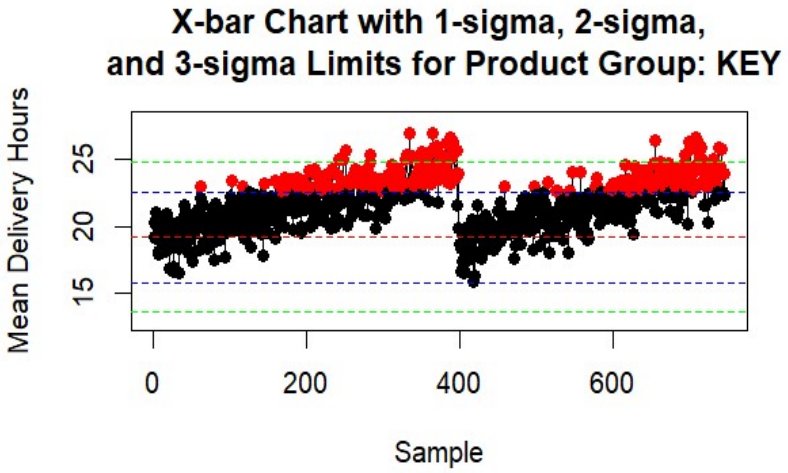


Graph B1

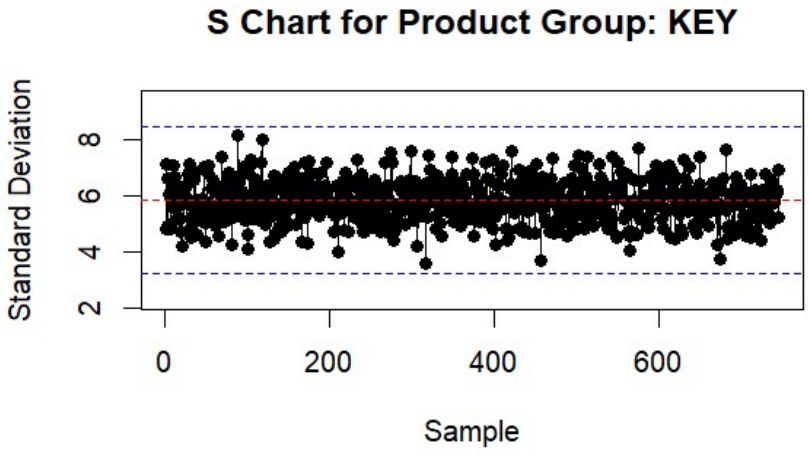


Graph B2

Product Group: KEY

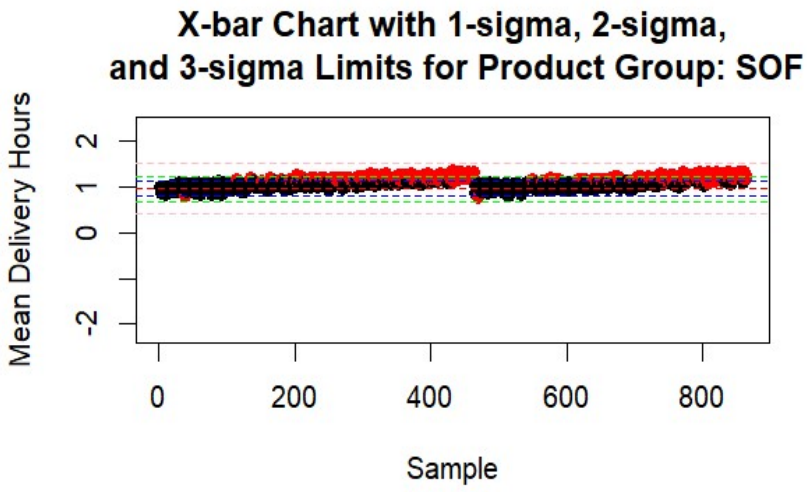


Graph B3

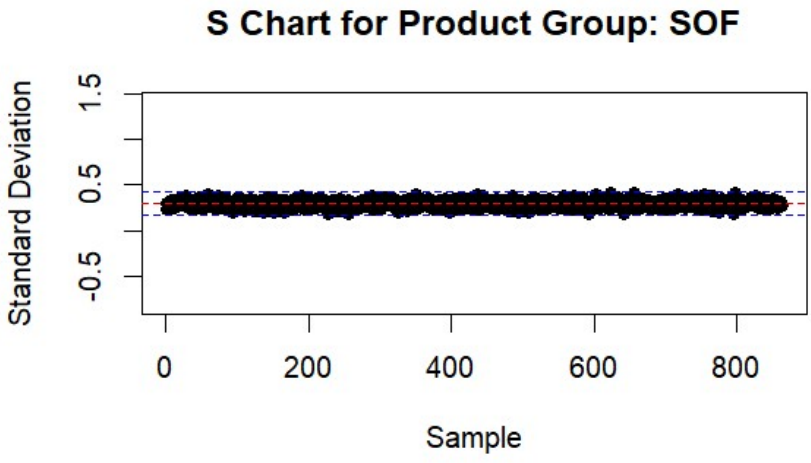


Graph B4

Product Group: SOF

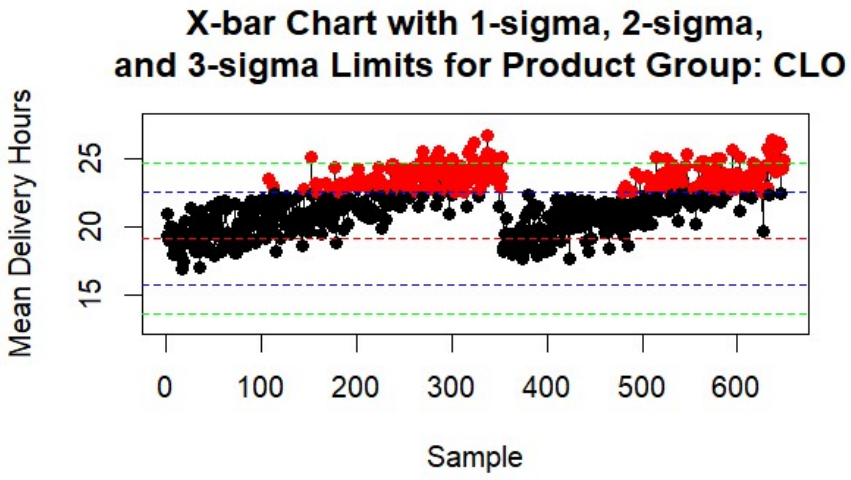


Graph B5

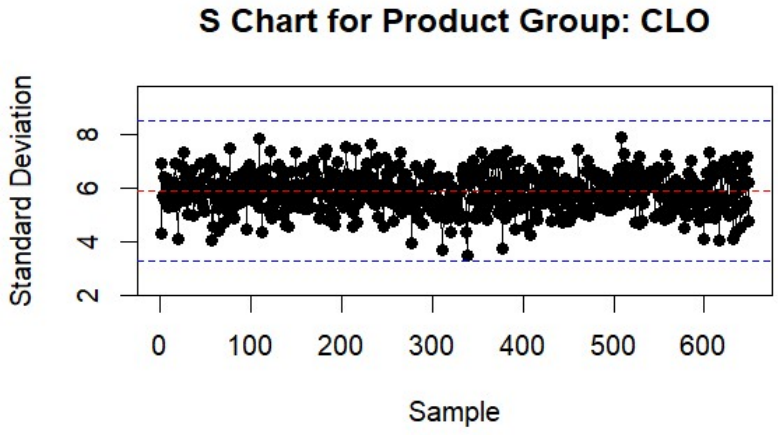


Graph B6

Product Group: CLO

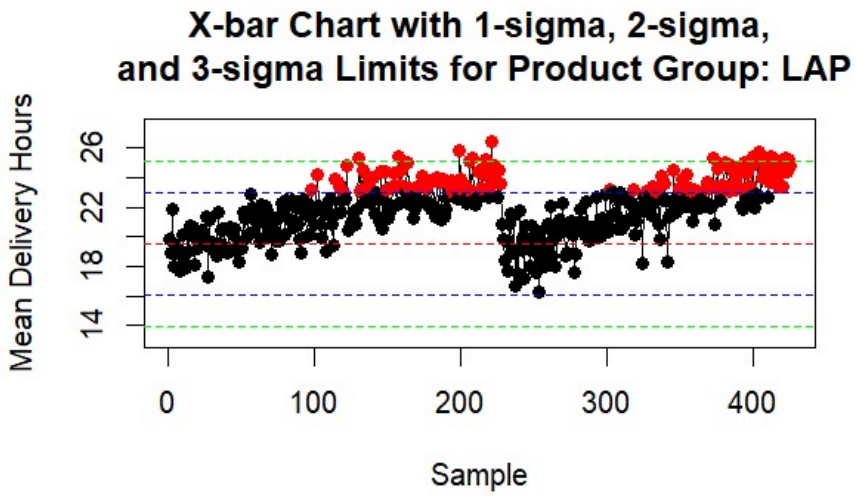


Graph B7

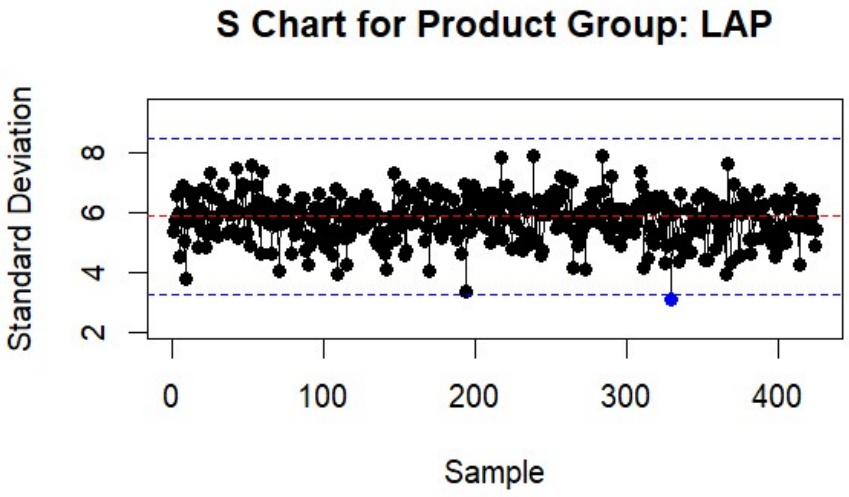


Graph B8

Product Group: LAP

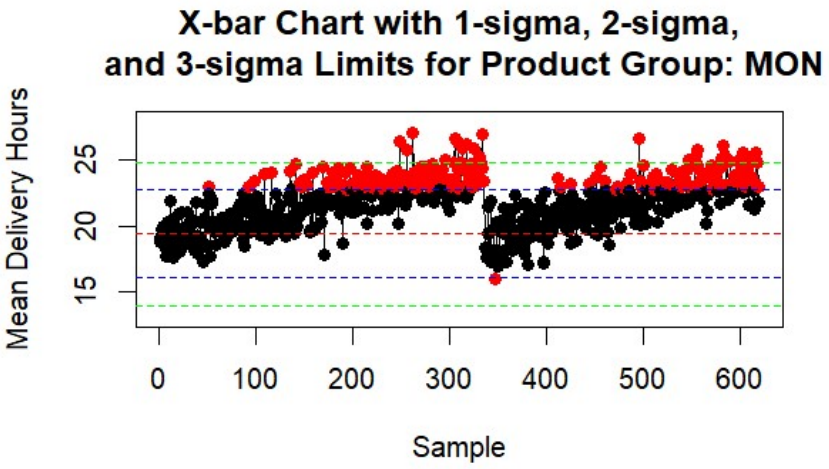


Graph B9

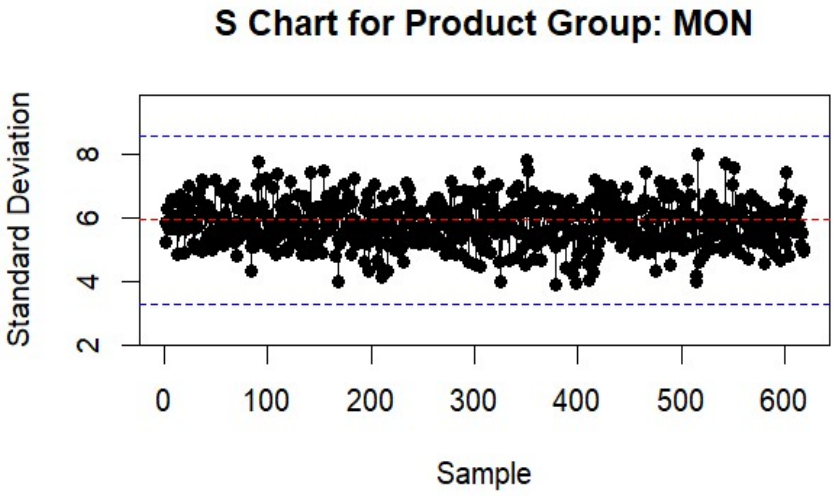


Graph B10

Product Group: MON



Graph B11



Graph B12