# ECSA Graduate

# Attributes Project

## Quality Assurance 344

**Louis Wilhelm Ernst Boshoff**

**23662972**

**Third Year of Study**

**17 October 2022**

## Abstract

In this report, various quality factors relating to an online sales organisation are being analysed. Utilizing data analytics, it incorporates quality management analysis to provide graphs, tables, and statistics that lead to control measures and recommendations for the company's quality features.

# Table of Contents

## Table of Figures

## Table of Tables

# Introduction

This report's objectives are to assess how well the company is currently operating and to provide critical recommendations regarding the outcomes. For the analysis to be accurate, all invalid data must first be removed through data wrangling. In order to better comprehend the features, the data is then examined for all statistical metrics. To provide advice to the business regarding the current technology process, the process capabilities of the technology class are computed. The data is subsequently subjected to statistical process control for additional analysis. The uncontrolled data is investigated. The best delivery period for the technology class is then chosen in order to minimise revenue loss. It is estimated and described how likely it is for the process to make type 1 and type 2 errors. The relationship between the data is explained using a MANOVA table. Lastly, it is determined how reliable other corporate operations are.

# 1 Data Wrangling

There are 180 000 instances of each of the 10 features in the dataset. The price feature contains all unusable data. Negative values and missing values are examples of invalid data. Errors in data recording, such as typos, might result in missing values. Typographical errors or incorrectly recording returns as sales can result in negative figures. The dataset has 5 cases of negative values and 17 instances of missing values. Separating valid data from invalid data is known as data wrangling. Next, the data was stored in three separate files. The tables below display an example of the clean data set and the faulty data. 179978 instances of 11 variables make up the valid dataset.

*Table 1: Sample of table with only valid data*

| | Index | X | ID | AGE | Class | Price | Year | Month | Day | Delivery.time | Why.Bought |
|---|---|---|---|---|---|---|---|---|---|---|---|
| 463 | 1 | 463 | 47101 | 50 | Clothing | 1030.86 | 2021 | 1 | 1 | 9.0 | Recommended |
| 2627 | 2 | 2627 | 88087 | 21 | Clothing | 428.03 | 2021 | 1 | 1 | 10.0 | Recommended |
| 3374 | 3 | 3374 | 25418 | 68 | Household | 13184.41 | 2021 | 1 | 1 | 48.5 | Website |
| 5288 | 4 | 5288 | 13566 | 94 | Household | 7021.90 | 2021 | 1 | 1 | 42.0 | Recommended |
| 8182 | 5 | 8182 | 84692 | 35 | Clothing | 475.18 | 2021 | 1 | 1 | 9.0 | Recommended |
| 9272 | 6 | 9272 | 46305 | 72 | Clothing | 580.98 | 2021 | 1 | 1 | 8.5 | Random |
| 9712 | 7 | 9712 | 92105 | 45 | Household | 6877.00 | 2021 | 1 | 1 | 43.0 | Recommended |
| 12163 | 8 | 12163 | 21614 | 27 | Clothing | 513.13 | 2021 | 1 | 1 | 9.5 | Recommended |
| 12195 | 9 | 12195 | 12174 | 56 | Household | 14538.64 | 2021 | 1 | 1 | 41.5 | EMail |
| 20004 | 10 | 20004 | 84558 | 74 | Food | 255.41 | 2021 | 1 | 1 | 2.0 | Recommended |
| 20509 | 11 | 20509 | 15630 | 32 | Clothing | 164.56 | 2021 | 1 | 1 | 9.0 | Recommended |
| 21970 | 12 | 21970 | 81216 | 87 | Clothing | 173.76 | 2021 | 1 | 1 | 10.0 | Recommended |

*Table 2: Table of all the invalid data removed from the data set*

| | X | ID | AGE | Class | Price | Year | Month | Day | Delivery.time | Why.Bought |
|---|---|---|---|---|---|---|---|---|---|---|
| 144443 | 144443 | 37737 | 81 | Food | -588.8 | 2022 | 12 | 10 | 2.5 | Recommended |
| 16320 | 16320 | 44142 | 82 | Household | -588.8 | 2023 | 10 | 2 | 48.0 | EMail |
| 19998 | 19998 | 68743 | 45 | Household | -588.8 | 2024 | 7 | 16 | 45.5 | Recommended |
| 155554 | 155554 | 36599 | 29 | Luxury | -588.8 | 2026 | 4 | 14 | 3.5 | Recommended |
| 19540 | 19540 | 65689 | 96 | Sweets | -588.8 | 2028 | 4 | 7 | 3.0 | Random |
| 98765 | 98765 | 64288 | 25 | Clothing | NA | 2021 | 1 | 24 | 8.5 | Browsing |
| 54321 | 54321 | 62209 | 34 | Clothing | NA | 2021 | 3 | 24 | 9.5 | Recommended |
| 34567 | 34567 | 18748 | 48 | Clothing | NA | 2021 | 4 | 9 | 8.0 | Recommended |
| 155555 | 155555 | 33583 | 56 | Gifts | NA | 2022 | 12 | 9 | 10.0 | Recommended |
| 177777 | 177777 | 68698 | 30 | Food | NA | 2023 | 8 | 14 | 2.5 | Recommended |
| 56789 | 56789 | 63849 | 51 | Gifts | NA | 2024 | 5 | 3 | 10.5 | Website |
| 87654 | 87654 | 40983 | 33 | Food | NA | 2024 | 8 | 27 | 2.0 | Recommended |
| 166666 | 166666 | 60188 | 37 | Technology | NA | 2024 | 10 | 9 | 21.5 | Website |
| 19541 | 19541 | 71169 | 42 | Technology | NA | 2025 | 1 | 19 | 20.5 | Recommended |
| 19999 | 19999 | 67228 | 89 | Gifts | NA | 2026 | 2 | 4 | 15.0 | Recommended |
| 12345 | 12345 | 18973 | 93 | Gifts | NA | 2026 | 6 | 11 | 15.5 | Website |
| 23456 | 23456 | 88622 | 71 | Food | NA | 2027 | 4 | 18 | 2.5 | Random |
| 65432 | 65432 | 51904 | 31 | Gifts | NA | 2027 | 7 | 24 | 14.5 | Recommended |
| 144444 | 144444 | 70761 | 70 | Food | NA | 2027 | 9 | 28 | 2.5 | Recommended |
| 76543 | 76543 | 79732 | 71 | Food | NA | 2028 | 9 | 24 | 2.5 | Recommended |
| 16321 | 16321 | 81959 | 43 | Technology | NA | 2029 | 9 | 6 | 22.0 | Recommended |
| 45678 | 45678 | 89095 | 65 | Sweets | NA | 2029 | 11 | 6 | 2.0 | Recommended |

# 2 Descriptive Statistics

## 2.1 Observational Descriptive Statistics

Number of observations in the valid data set: 179978

Number of variables in the valid data set: 11

### 2.1.1 Valid data analysis

We may get a rudimentary knowledge of the important aspects of the data from the data quality overview table (Table 3). Since they were eliminated from the valid dataset during data wrangling, there are no missing values in this dataset. There are two character features, nine numeric features, seven of which are integer features. Due to their low cardinalities, Class, Year, and Why.Bought are categorical attributes.

*Table 3: Data quality overview table*

| | variables | types | missing_count | missing_percent | unique_count | unique_rate |
|---|---|---|---|---|---|---|
| 1 | Index | integer | 0 | 0 | 179978 | 1.000000e+00 |
| 2 | X | integer | 0 | 0 | 179978 | 1.000000e+00 |
| 3 | ID | integer | 0 | 0 | 15000 | 8.334352e-02 |
| 4 | AGE | integer | 0 | 0 | 91 | 5.056174e-04 |
| 5 | Class | character | 0 | 0 | 7 | 3.889364e-05 |
| 6 | Price | numeric | 0 | 0 | 78832 | 4.380091e-01 |
| 7 | Year | integer | 0 | 0 | 9 | 5.000611e-05 |
| 8 | Month | integer | 0 | 0 | 12 | 6.667482e-05 |
| 9 | Day | integer | 0 | 0 | 30 | 1.666870e-04 |
| 10 | Delivery.time | numeric | 0 | 0 | 148 | 8.223227e-04 |
| 11 | Why.Bought | character | 0 | 0 | 6 | 3.333741e-05 |

### 2.1.2 Continuous feature analysis

This table provides statistical measurements of numerical features. There are no categorical features in this table.

*Table 4: Numerical Feature Diagnosis*

| | variables | min | Q1 | mean | median | Q3 | max | zero | minus | outlier |
|---|---|---|---|---|---|---|---|---|---|---|
| 1 | Index | 1.00 | 44995.25 | 89989.500000 | 89989.50 | 134983.75 | 179978 | 0 | 0 | 0 |
| 2 | X | 1.00 | 45004.25 | 90002.546378 | 90004.50 | 134999.75 | 180000 | 0 | 0 | 0 |
| 3 | ID | 11126.00 | 32700.00 | 55234.688668 | 55081.00 | 77637.00 | 99992 | 0 | 0 | 0 |
| 4 | AGE | 18.00 | 38.00 | 54.565519 | 53.00 | 70.00 | 108 | 0 | 0 | 0 |
| 5 | Price | 35.65 | 482.31 | 12294.098366 | 2259.63 | 15270.97 | 116619 | 0 | 0 | 22171 |
| 6 | Year | 2021.00 | 2022.00 | 2024.854643 | 2025.00 | 2027.00 | 2029 | 0 | 0 | 0 |
| 7 | Month | 1.00 | 4.00 | 6.521064 | 7.00 | 10.00 | 12 | 0 | 0 | 0 |
| 8 | Day | 1.00 | 8.00 | 15.538949 | 16.00 | 23.00 | 30 | 0 | 0 | 0 |
| 9 | Delivery.time | 0.50 | 3.00 | 14.500311 | 10.00 | 18.50 | 75 | 0 | 0 | 17516 |

Both “X” and “ID” have very high cardinality values that are nearly equivalent to the number of instances for that feature. They should also be eliminated for additional analysis and visualisation because they don't contribute any meaningful information to the sales data.The average sales price is R12 294.10, with the lowest sales price being R35.65 and the highest being R116619. There are quite a few outliers in this feature—22171 in all. This can be a result of some pricey products.The lowest and longest delivery times, respectively, are 0.5 days and 75 days, and the average delivery time is 14.5 days. This feature contains 17516 outliers, which may be related to the extended delivery times for some products."Age" has a range of 18 to 108, with a mean of 55. This indicates that adults around the age of 55 make up most of the customers.

### 2.1.2 Categorical feature analysis

The features Class and Why.Bought are character variables, making them categorical. According to the table, "Gifts" had the largest percentage of sales (21.8%) for the class. The second most common sales category, with a 20.20% share, is "Technology". Only 6.59% of sales are made in the "Luxury" category, which may be a result of the high prices.

We may learn useful information about why individuals purchase particular things via the feature "Why.Bought". " Recommended" is the most popular justification for purchasing a product (59.44%). The second most common reason consumers buy things is "Website"(16.36%), recommend, which falls far short of “Recommended".

| | variables | levels | N | freq | ratio | rank |
|---|---|---|---|---|---|---|
| 1 | Class | Gifts | 179978 | 39149 | 21.752103 | 1 |
| 2 | Class | Technology | 179978 | 36347 | 20.195246 | 2 |
| 3 | Class | Clothing | 179978 | 26403 | 14.670126 | 3 |
| 4 | Class | Food | 179978 | 24582 | 13.658336 | 4 |
| 5 | Class | Sweets | 179978 | 21564 | 11.981464 | 5 |
| 6 | Class | Household | 179978 | 20065 | 11.148585 | 6 |
| 7 | Class | Luxury | 179978 | 11868 | 6.594139 | 7 |
| 8 | Why.Bought | Recommended | 179978 | 106985 | 59.443376 | 1 |
| 9 | Why.Bought | Website | 179978 | 29447 | 16.361444 | 2 |
| 10 | Why.Bought | Browsing | 179978 | 18994 | 10.553512 | 3 |
| 11 | Why.Bought | Random | 179978 | 13120 | 7.289780 | 4 |
| 12 | Why.Bought | EMail | 179978 | 7224 | 4.013824 | 5 |
| 13 | Why.Bought | Spam | 179978 | 4208 | 2.338064 | 6 |

*Table 5: Categorical Data Diagnosis*

## 2.2 Visualization of variables

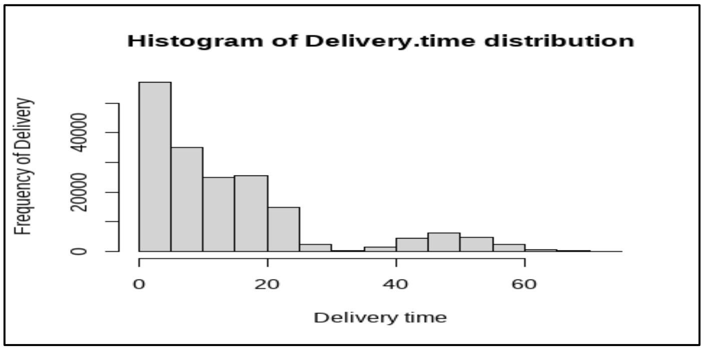


*Figure 1*

*Figure 1:* The most frequent delivery period is between 3 and 5 days. With more brief and fewer lengthy delivery times, the distribution is skewed to the right. Between 30 days and 65 days, there is a normal distribution

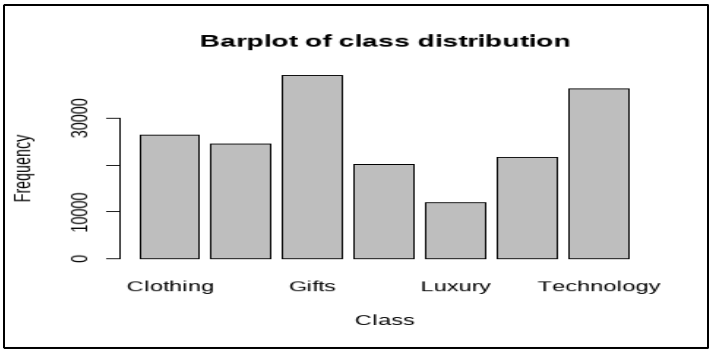


*Figure 2*

*Figure 2:* This graph demonstrates that the Gifts and Technology classes see significantly higher sales than the other classes, with luxury goods experiencing the lowest sales of any class.

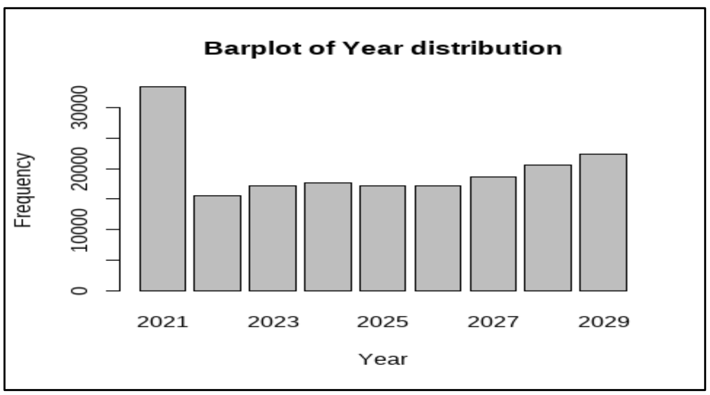


*Figure 3*

*Figure 3:* This graph shows that sales peaked in 2021 before abruptly declining in 2022. From 2022 to 2025, the sales then fluctuated, and from 2026 to 2029, they grew slowly but steadily.

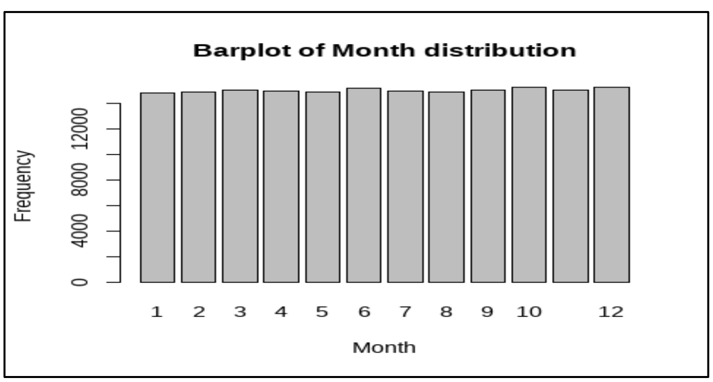


*Figure 4*

*Figure 4:* This graph demonstrates that from January through to December, the sales per month did not significantly fluctuate. The sales every month follow a uniform distribution.

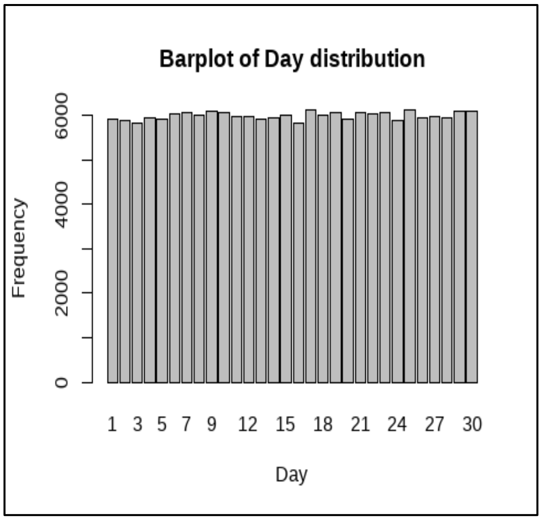


*Figure 5*

*Figure 5:* This graph demonstrates that throughout a given month, the sales per day did not significantly fluctuate. The sales every day follow a uniform distribution, similar to that of the sales of every month.

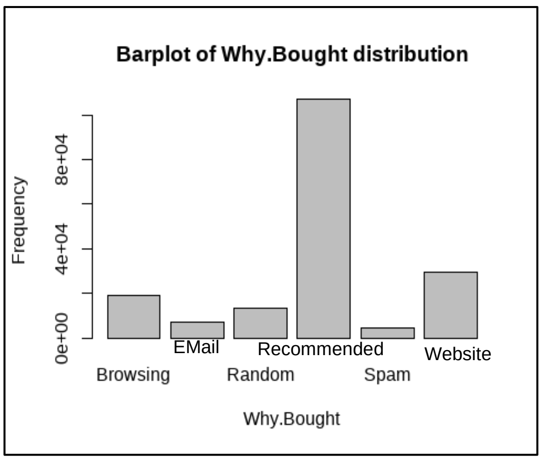


*Figure 6*

*Figure 6:* This graph clearly shows that recommendations are the most frequent justification for purchasing products of any type.

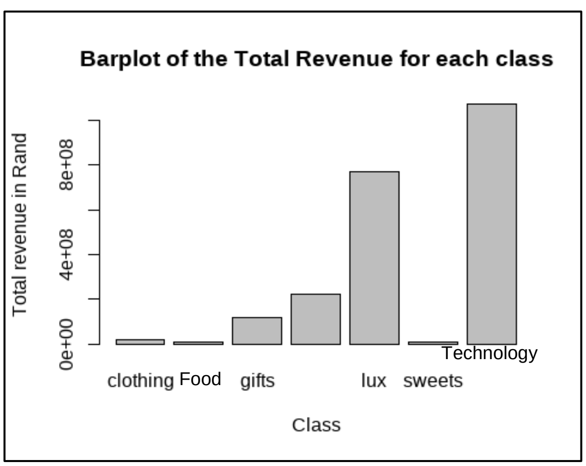


*Figure 7*

*Figure 7:* According to this graph, the Luxury and Technology classes have the highest overall sales revenue.

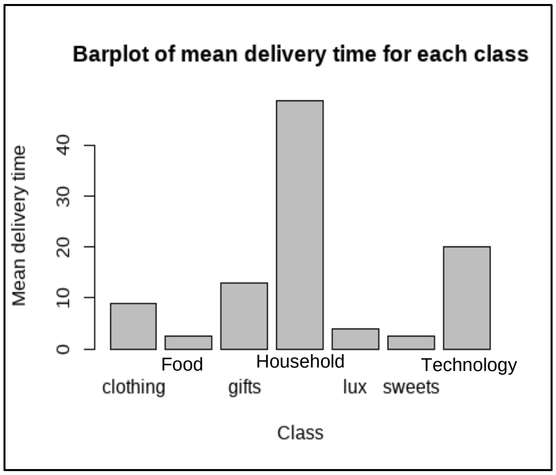


*Figure 8*

*Figure 8:* This Bar plot makes it obvious that the household sales class has the longest mean delivery time. Sales of the Technology class, which have a mean delivery time that is significantly less than that of the Household class, are discovered to have the second-longest mean delivery time.

## 2.5 Process Capabilities calculations

Process capability refers to a process' ability to produce within predetermined parameters. The process is capable if it can produce products that meet these requirements; otherwise, it is not.

Cp and Cpk are typically used in circumstances where a process is in control. Your Cp and Cpk values represent how consistently you execute in comparison to your typical performance.

Upper Specification Limit (USL) and Lower Specification Limit (LSL) are defined in the specifications as limits that establish whether the process is feasible or not. In our situation, the LSL is zero and the USL is 28 days.

```
USL <- 24
LSL <- 0
Std <- 3.50199
X <- 20.01095
```

```
Cp <- (USL - LSL)/(6*Std) <-1.14220
Cpu <- (USL - X)/(3*Std) <-0.37969
Cpl <- (X - LSL)/(3*Std) <-1.90472
Cpk <- min(Cpu, Cpl) <-0.37969
```

A LSL of zero makes sense because delivery time cannot be negative and 0 is the lowest limit, hence the lower specification limit (LSL) cannot be negative (less than 0). Process Capability Index (Cp) = 1.14220. This demonstrates that the process can fulfill requirements.

# 3. Statistical Process Control (SPC)

## 3.1 First 30 Samples

### 3.1.1 X-Chart Limits

The mean change of a process over time as measured by subgroup values is displayed using an x chart. A subgroup is represented by a sample's measurement at a certain time.

| Class | UCL | UCL2 | UCL1 | CL | LCL1 | LCL2 | LCL |
|---|---|---|---|---|---|---|---|
| Clothing | 9.4047 | 9.2598 | 9.1149 | 8.97 | 8.8251 | 8.6802 | 8.5353 |
| Household | 50.2462 | 49.0182 | 47.7902 | 46.5622 | 45.3342 | 44.1062 | 42.8783 |
| Food | 2.7119 | 2.6383 | 2.5647 | 2.4911 | 2.4175 | 2.3439 | 2.2704 |
| Technology | 22.9745 | 22.1138 | 21.253 | 20.3922 | 19.5315 | 18.6707 | 17.8099 |
| Sweets | 2.9007 | 2.7597 | 2.6188 | 2.4778 | 2.3368 | 2.1958 | 2.0548 |
| Gifts | 9.4879 | 9.1123 | 8.7367 | 8.3611 | 7.9855 | 7.6099 | 7.2343 |
| Luxury | 5.4847 | 5.2335 | 4.9823 | 4.7311 | 4.4799 | 4.2287 | 3.9776 |

*Table 6: X-Chart Limits*

### 3.1.2 S-Chart Limits

The standard deviation of a process over time from subgroup values is displayed using a s chart. It keeps track of the process's approximate standard deviation using the sample moving average.

| Class | UCL | UCL2 | UCL1 | CL | LCL1 | LCL2 | LCL |
|---|---|---|---|---|---|---|---|
| Clothing | 0.8664 | 0.7614 | 0.6563 | 0.5512 | 0.4462 | 0.3411 | 0.236 |
| Household | 7.3432 | 6.4528 | 5.5623 | 4.6719 | 3.7814 | 2.891 | 2.0005 |
| Food | 0.44 | 0.3867 | 0.3333 | 0.2799 | 0.2266 | 0.1732 | 0.1199 |
| Technology | 5.1473 | 4.5231 | 3.899 | 3.2748 | 2.6506 | 2.0264 | 1.4023 |
| Sweets | 0.843 | 0.7408 | 0.6386 | 0.5363 | 0.4341 | 0.3319 | 0.2297 |
| Gifts | 2.246 | 1.9737 | 1.7013 | 1.429 | 1.1566 | 0.8842 | 0.6119 |
| Luxury | 1.5021 | 1.3199 | 1.1378 | 0.9556 | 0.7735 | 0.5913 | 0.4092 |

*Table 7: S-Chart Limits*

### 3.1.3 S-Charts

S-Charts are always generated before X-Charts since they are used to assess the dependability of the accompanying X-Chart. A S-Chart is out of control and the related X-Chart is unreliable if its values are above the UCL or below the LCL.

The calculated S-Charts for the first 30 samples are shown below. All the S-Charts are under control after analysis and the X-Charts can be found as a result.

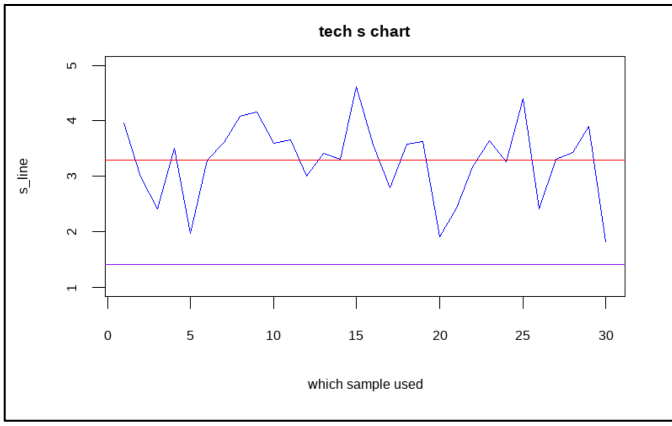


*Figure 9: Technology S-Chart*

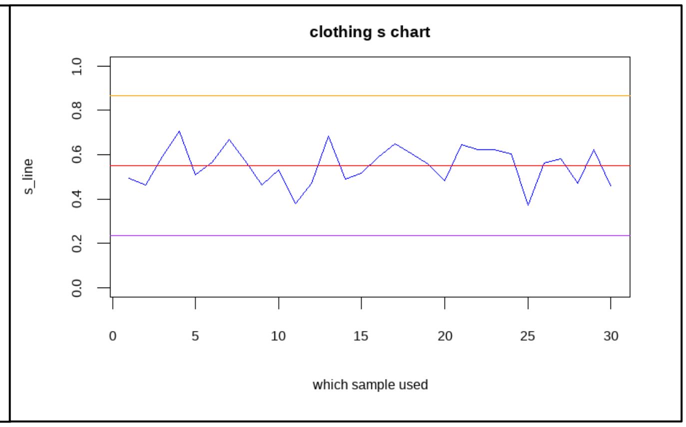


*Figure 10: Clothing S-Chart*

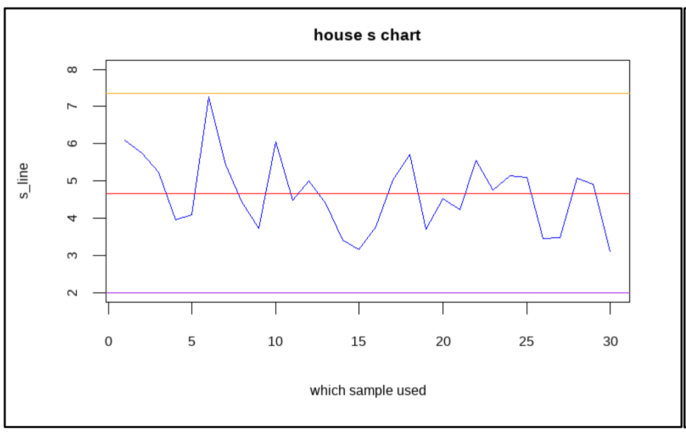


*Figure 11: House S-Chart*

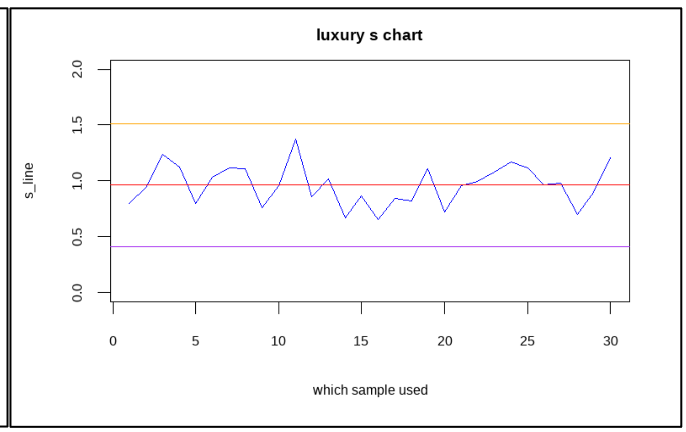


*Figure 12: Luxury S-Chart*

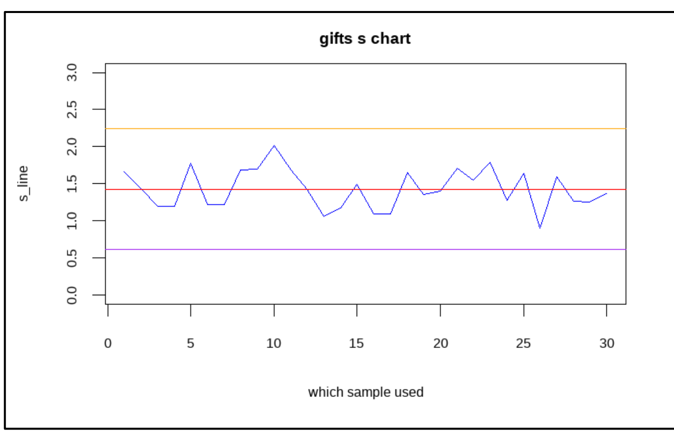


*Figure 13: Gifts S-Chart*

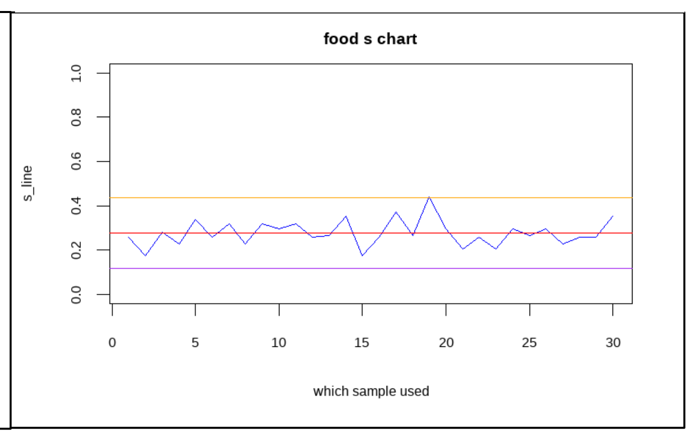


*Figure 14: Food S-Chart*

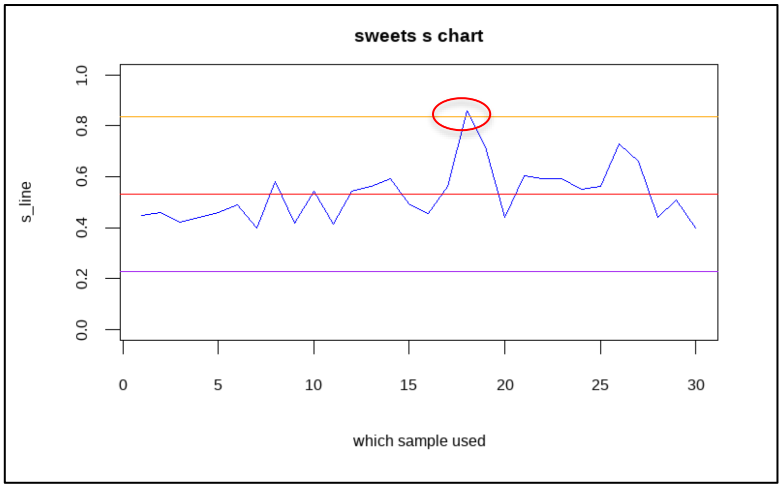


Figure 15: Sweets S-Chart (with out-of-control instance)

The S-Chart for the sweets class was out of control, as can be seen in the figure. The X-Charts will be unreliable since all the plots on the S-Chart are not under control. This could possibly be fixed by removing the instances that are out of bounds.

### 3.1.4 X-Charts

The statistical control of a process can be assessed using X-Charts. In this case, it would establish whether each class's delivery time complies with client expectations.

Below are the X-Charts for the first 30 samples for each class. All the X-Charts are in control, excluding the Class Sweets, because they are all above the LCL and below the UCL. Now it is possible to investigate the remaining occurrences in greater detail.

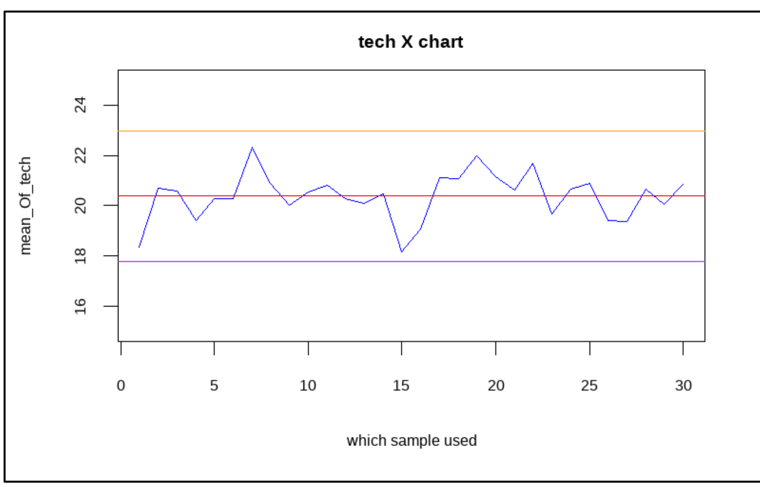


Figure 16: Technology X-Chart

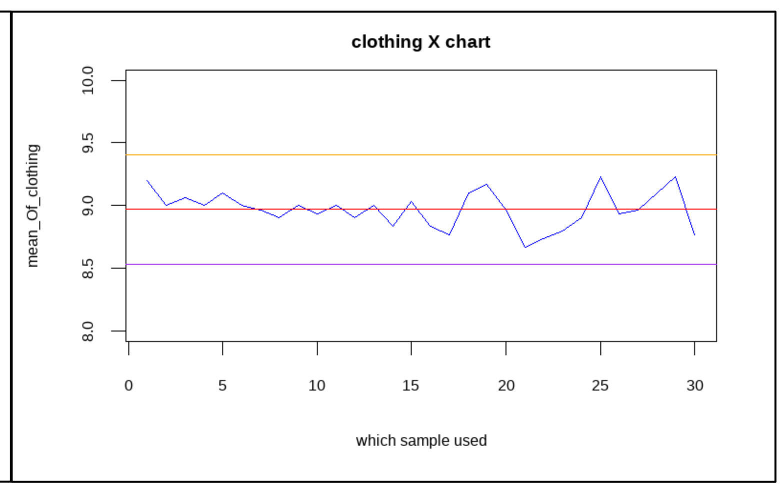


Figure 17: Clothing X-Chart

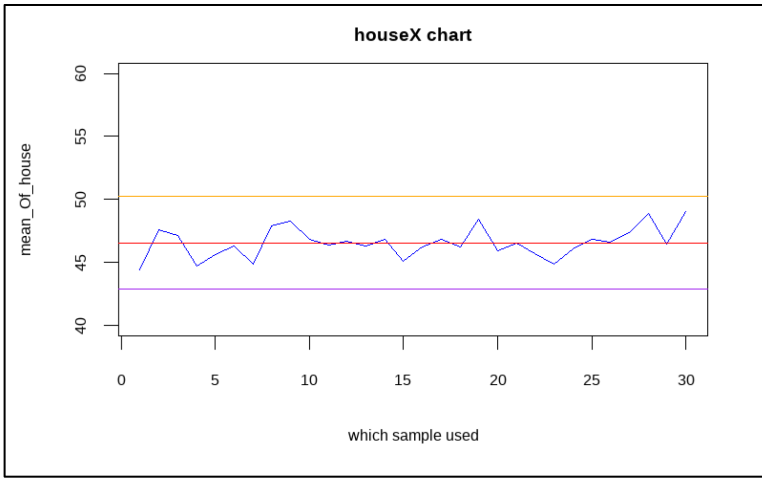


*Figure 18: House X-Chart*

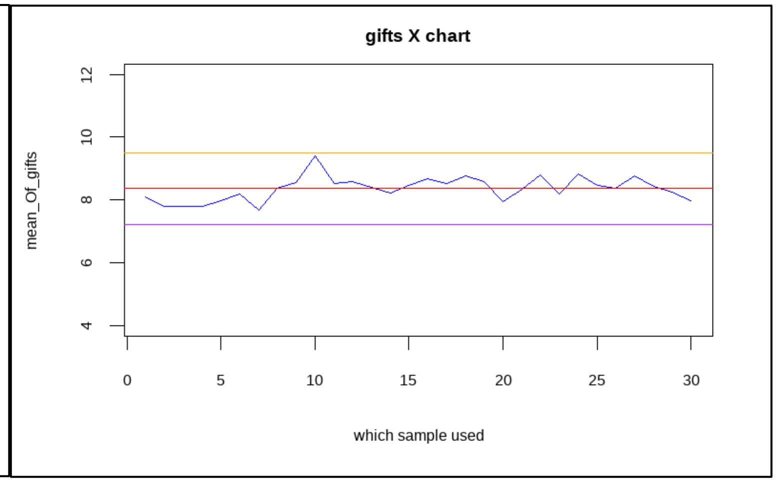


*Figure 19: Gifts X-Chart*

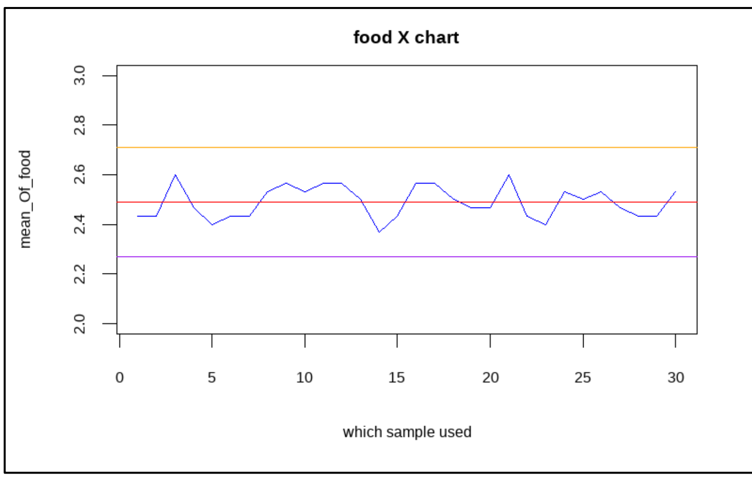


*Figure 20: Food X-Chart*

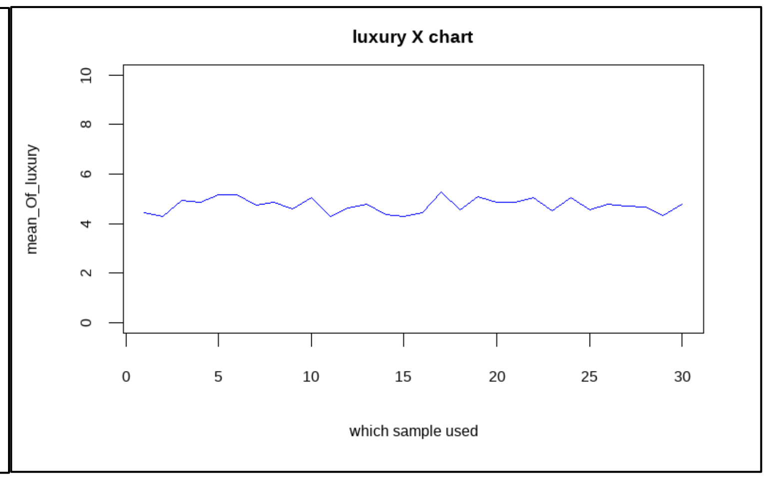


*Figure 21: Luxury X-Chart*

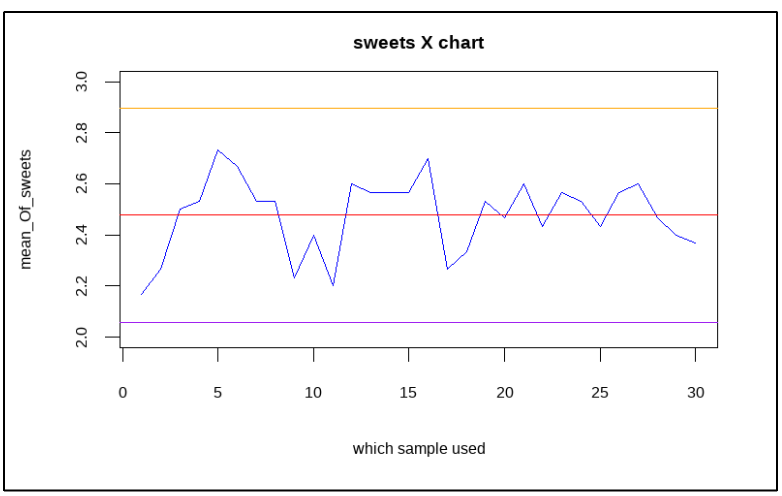


*Figure 22: Sweets X-Chart*

## 3.2 Process Control

### 3.2.1 Out-of-control process

The following figures demonstrate unequivocally that the Household, Luxury, and Gifts classes are unstable. These processes must be improved. Process performance should be tracked over time once the process is stabilised and the limitations are set.

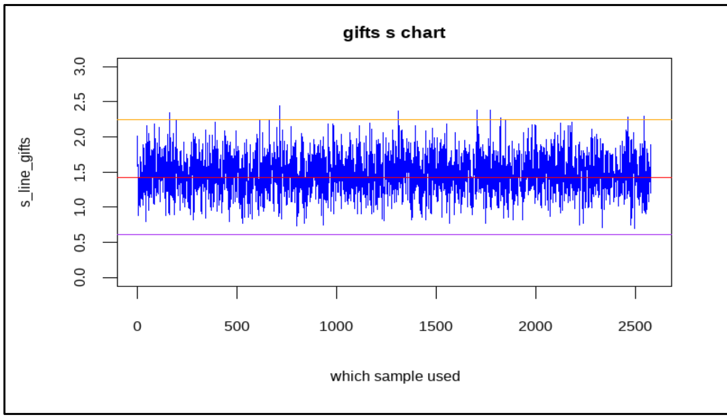


*Figure 23: Gifts S-Chart*

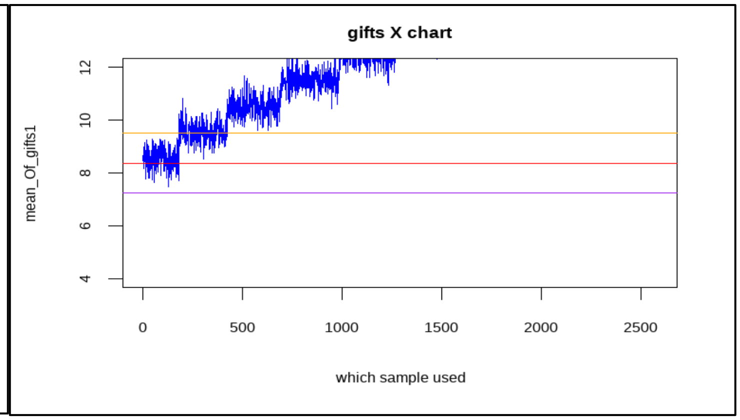


*Figure 24: Gifts X-Chart*

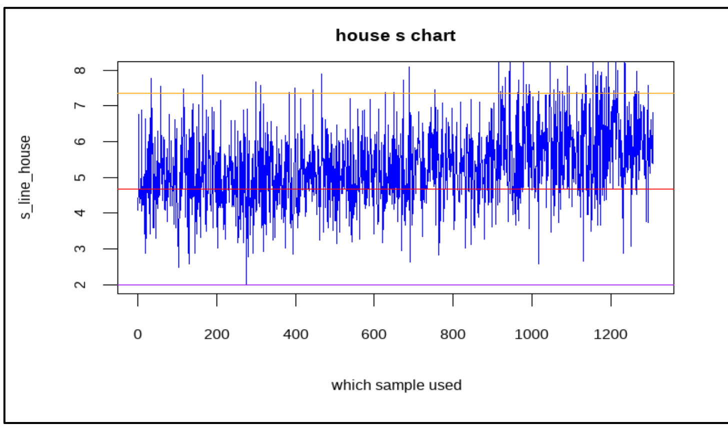


*Figure 25: House S-Chart*

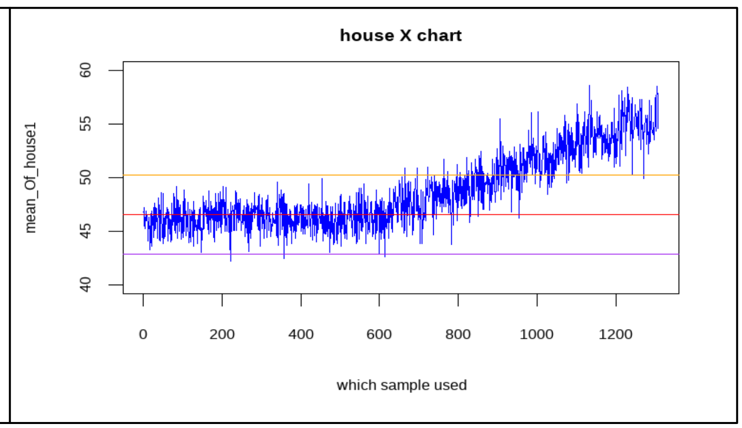


*Figure 26: House X-Chart*

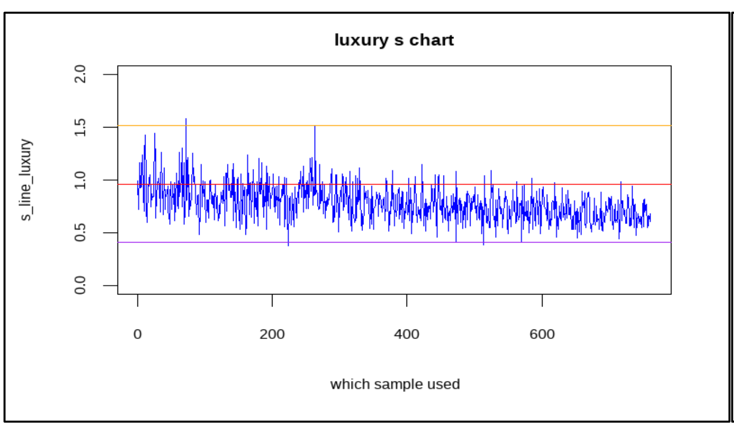


*Figure 27: Luxury S-Chart*

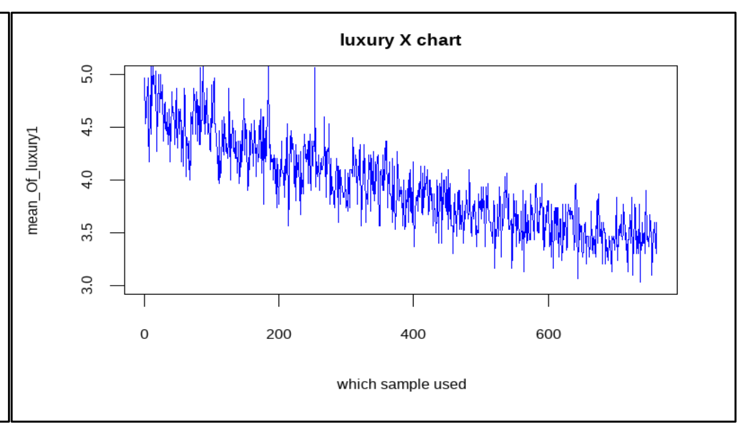


*Figure 28: Luxury X-Chart*

### 3.2.2 In-control process

Almost all the procedures for technology, clothing, food, and sweets are steady, albeit a few isolated incidents occasionally go outside of the acceptable range. On these out-of-control occurrences, root cause analysis can be performed.

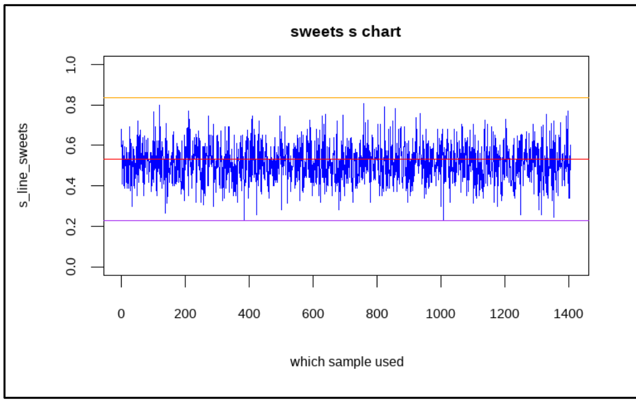


*Figure 29: Sweets S-Chart*

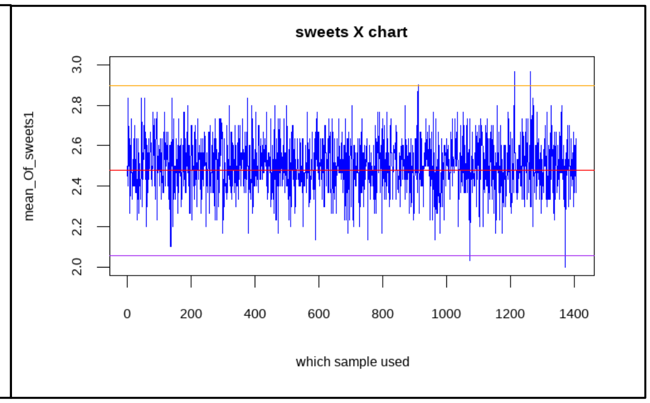


*Figure 30: Sweets X-Chart*

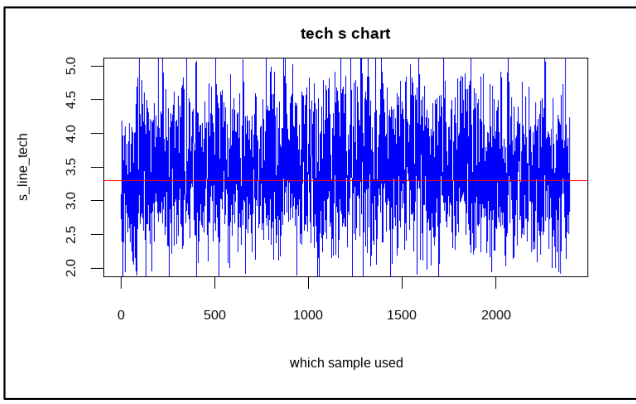


*Figure 31: Technology S-Chart*

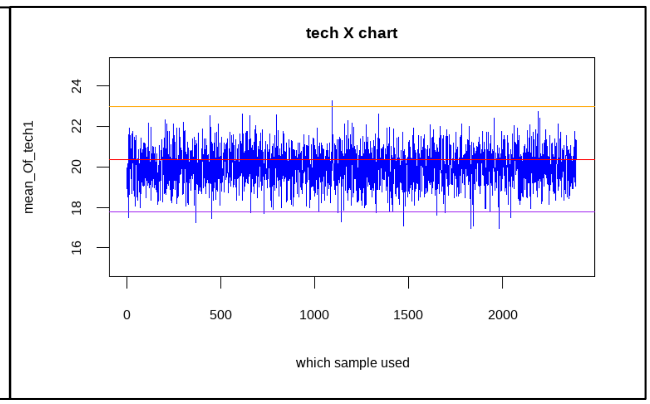


*Figure 32: Technology X-Chart*

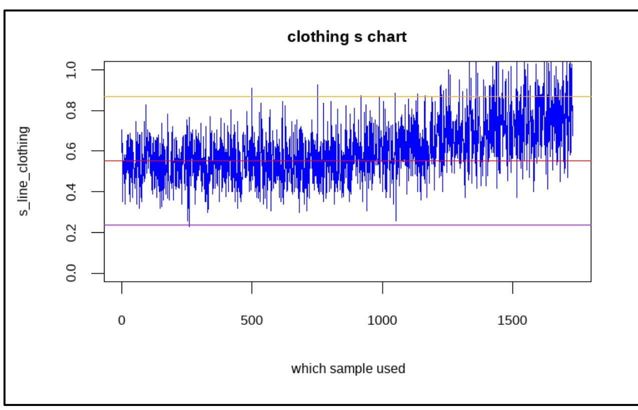


*Figure 33: Clothing S-Chart*

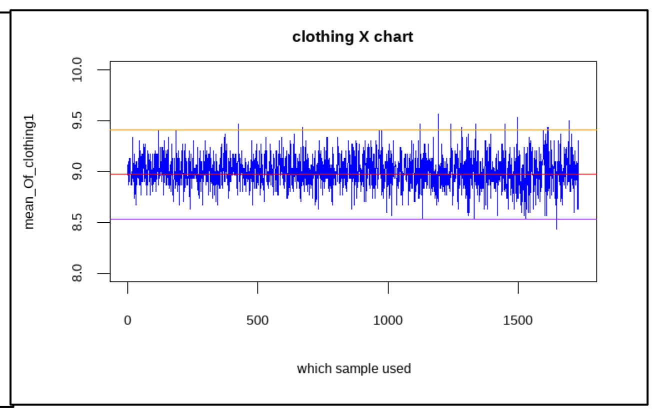


*Figure 34: Clothing X-Chart*

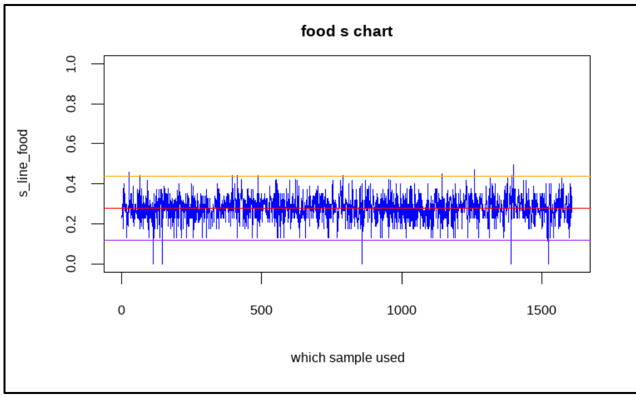

Figure 35: Food S-Chart

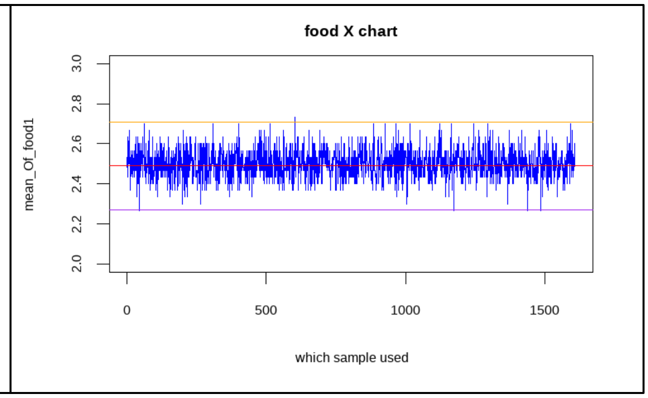

Figure 36: Food X-Chart

# 4. Optimising the delivery processes

## 4.1.1 Rule A

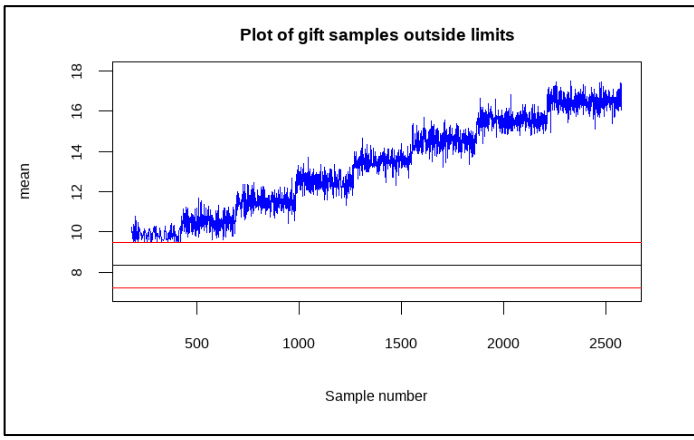

Figure 37: X-Bar chart of “Gift” samples outside limits

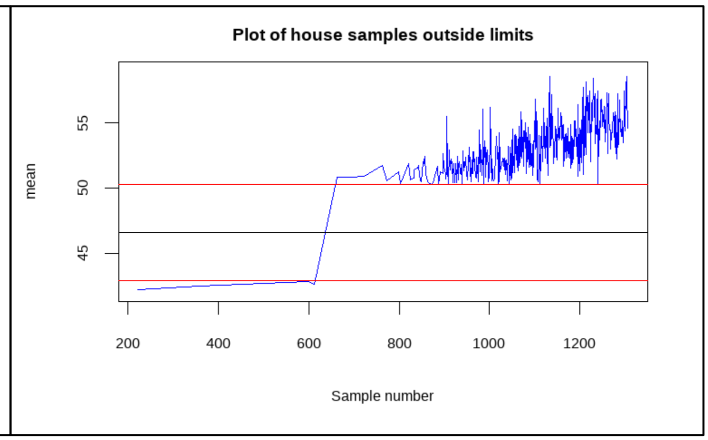

Figure 38: X-Bar chart of “House” samples outside limits

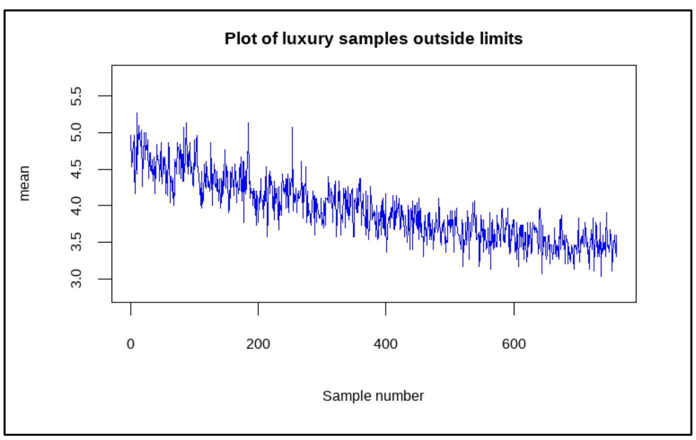

Figure 39: X-Bar chart of “Luxury” samples outside limits

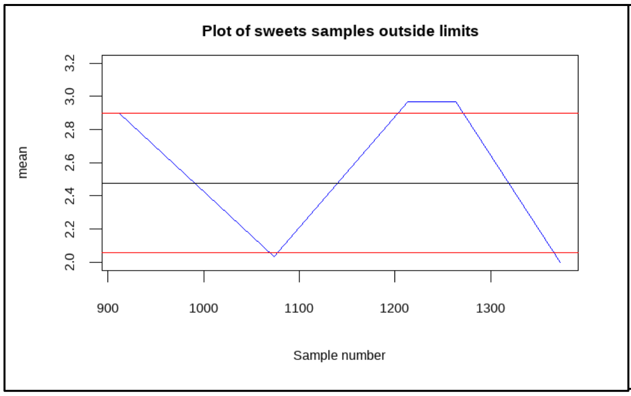


*Figure 40: X-Bar chart of "Sweets" samples outside limits*

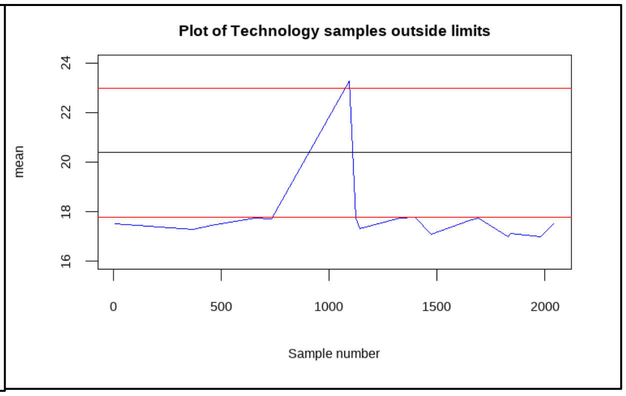


*Figure 41: X-Bar chart of "Technology" samples outside limits*

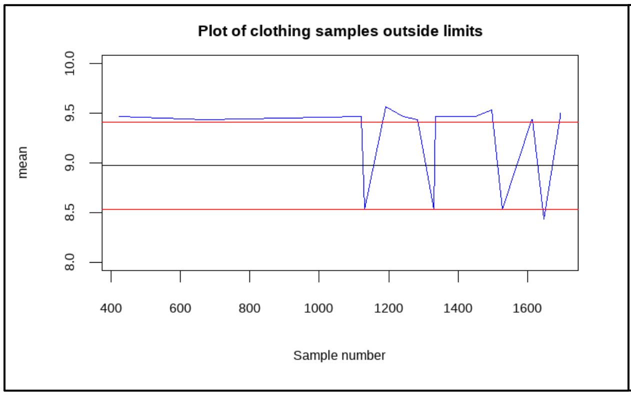


*Figure 42: X-Bar chart of "Clothing" samples outside limits*

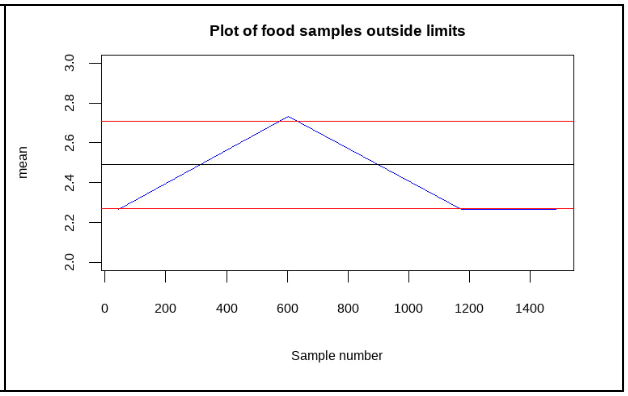


*Figure 43: X-Bar chart of "Food" samples outside limits*

*Table 8: Summary of samples outside the control limits*

| | Total ouliers | 1st | 2nd | 3rd | 3rd last | 2nd last | last |
|---|---|---|---|---|---|---|---|
| sweets | 5 | 912 | 1074 | 1213 | 1213 | 1264 | 1373 |
| luxury | 761 | 1 | 2 | 3 | 759 | 760 | 761 |
| technology | 17 | 7 | 368 | 453 | 1842 | 1979 | 2041 |
| gifts | 2290 | 183 | 186 | 188 | 2577 | 2578 | 2579 |
| house | 400 | 222 | 357 | 599 | 1305 | 1306 | 1307 |
| food | 5 | 45 | 603 | 1173 | 1173 | 1437 | 1485 |
| clothing | 17 | 425 | 672 | 1122 | 1647 | 1693 | 1694 |

As can be seen from the table above, there are very few samples from the classes that are statistically within control. Since "technology" is the class with the highest incomes, it is quite fortunate that there are so few occurrences outside of the control limits. Due to the huge number of samples beyond the control boundaries and the fact that "luxury" is the second highest income class, additional research is necessary. The significant number of occurrences involving "gifts" and the "house" sales are further investigated.

It is evident that most of the "luxury"-class average delivery time is getting shorter over time. Being outside of the control limits is not ideal, but it might be advantageous because shorter delivery times can make customers happier. *Figure 37* shows the samples outside of the "gift" control ranges. As previously said, the delivery time grew longer every year. This is unacceptable because, if the delivery time continues to rise at the same rate, it will eventually be much too high. The rapid increase depicted in *figure 38* is alarming since it indicates that the delivery times for items from the class "house" has worsened and is in dire need of attention from the company.

### 4.1.2 Rule B

*Table 9: Consecutive samples*

| **Class** | **Most consecutive samples of "s-bar or sample standard deviations between -0.3 and +0.4 sigma-control limits** | **Last sample to be in the given range (-0.3 -> 0.4)** |
|---|---|---|
| Technology | 6 | 2380 |
| House | 3 | 1290 |
| Gifts | 5 | 2574 |
| Sweets | 4 | 1407 |
| Luxury | 6 | 734 |
| Clothing | 4 | 1720 |
| Food | 7 | 1606 |

## 4.2 Type 1 Error

Rejecting a correct null hypothesis is a type 1 error, also referred as a manufacturing's error. In this instance, it means that a controlled process was mistakenly determined to be out of control. It leads to unneeded additional investigation.

### 4.2.1 Probability of a type 1 error for Rule A

Probability of making a type 1 error for *Rule A* is:

= pnorm(-3)*2

= 0.002699796
= 0.2699796%

The probability of making a type 1 error for *Rule A* is thus 0.2699796%.

### 4.2.2 Probability of a type 1 error for Rule B

Probability of making a type 1 error for *Rule B* is:

= 1-pnorm(0.4, lower.tail = FALSE) + pnorm(-0.3)
= 0.2733332
= 27.33332%

Thus, the likelihood of making a type 1 error for *Rule B* is 27.33332%. This suggests that the likelihood of discovering an in-control process to be out-of-control is 27.33332%.

## 4.3 Streamlining the distribution of technology by optimizing the delivery time

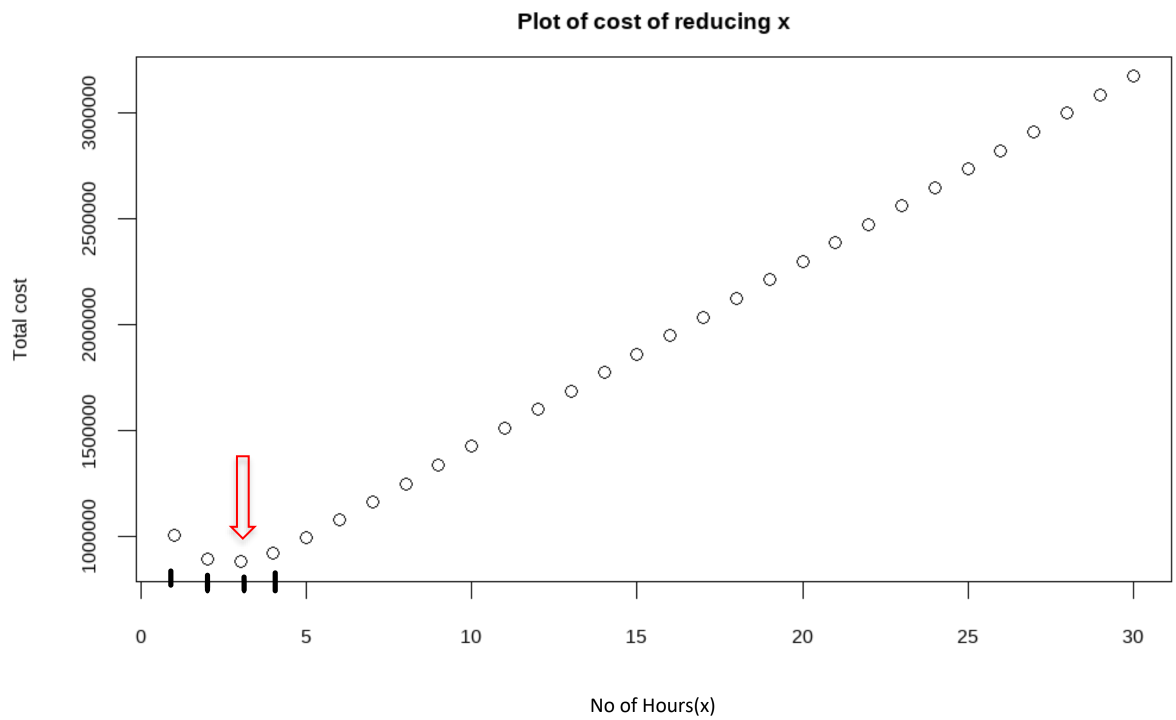


*Figure 44: Cost of reducing X*

Figure 44 demonstrates how changing the delivery time will affect the cost and produce the most profit attainable. It is obvious that a delivery period reduction of three will be ideal for goods in the "Technology" category, and it also demonstrates that delivering a few hours later will be less expensive than shipping goods a few hours earlier.

## 4.4 Type 2 Error

When a false null hypothesis is accepted, this is a type 2 error, also referred to as a consumer's error. In this instance, it signifies that a process that is out of control is mistakenly thought to be under control.

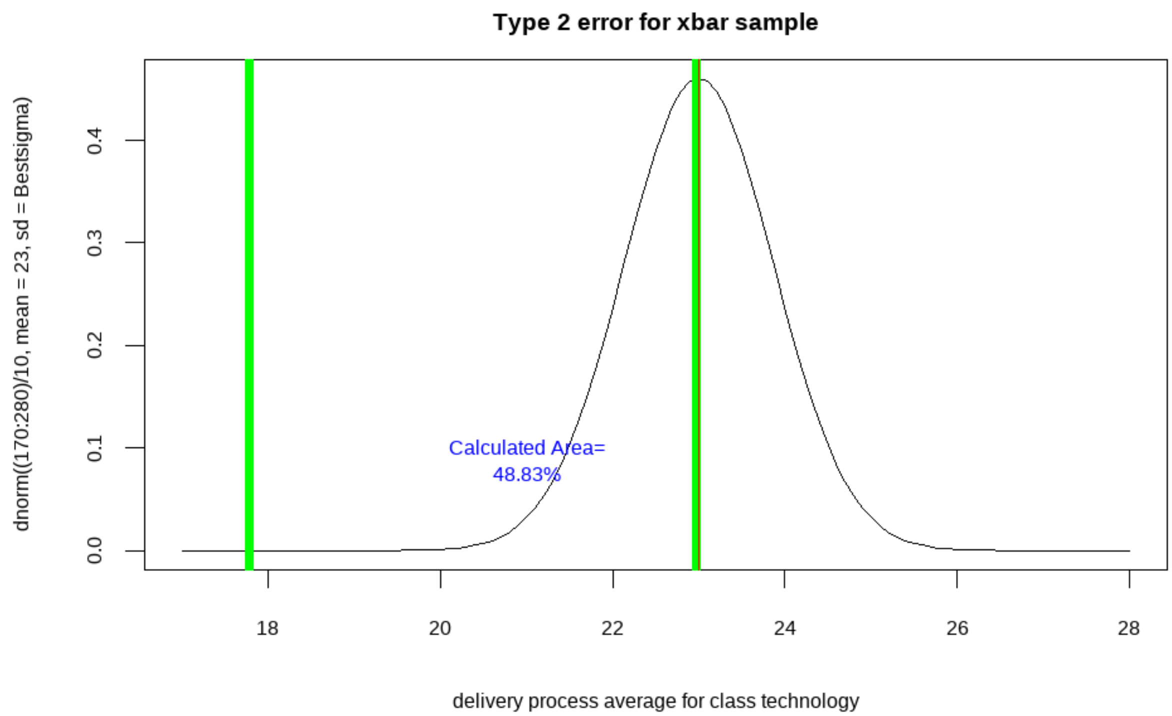


*Figure 45: Type 2 error for X bar sample*

Given that the delivery process mean shifts to 23, the likelihood of a type 2 error for A in Class=Technology is 48.83%

# 5 DOE and MANOVA

A well-known method for figuring out how different qualities affect one another is called MANOVA. Here, it is examined how delivery time and cost affect "Why. Bought." The following hypotheses are included in the produced MANOVA.

H0: The feature means are equal. This suggests that the features (delivery time and price) have no bearing on the motivation for purchasing.

H1: One or more mean variations exist. This suggests that at least one feature—possibly both—influences why a customer chooses a particular product (in this case, delivery time and pricing).

*Table 10: MANOVA for Price and Delivery.time vs Why.Bought*

```
 Response Price :
                Df     Sum Sq    Mean Sq F value    Pr(>F)
Why.Bought       5 1.5742e+12 3.1484e+11  736.26 < 2.2e-16 ***
Residuals   179972 7.6960e+13 4.2762e+08
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

 Response Delivery.time :
                Df   Sum Sq Mean Sq F value    Pr(>F)
Why.Bought       5   783320  156664  822.74 < 2.2e-16 ***
Residuals   179972 34269697     190
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1
```

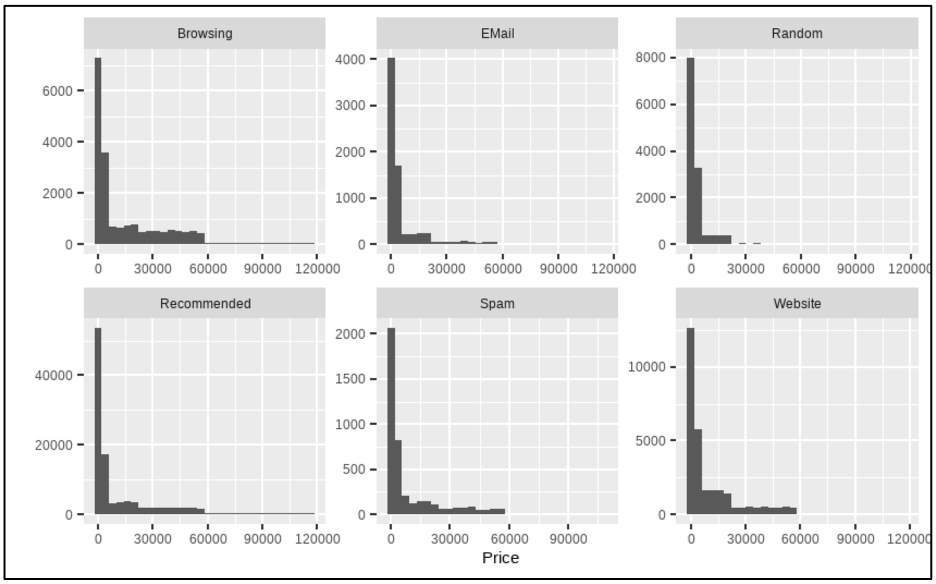


*Figure 46: Price distribution over reasons for Buying a product.*

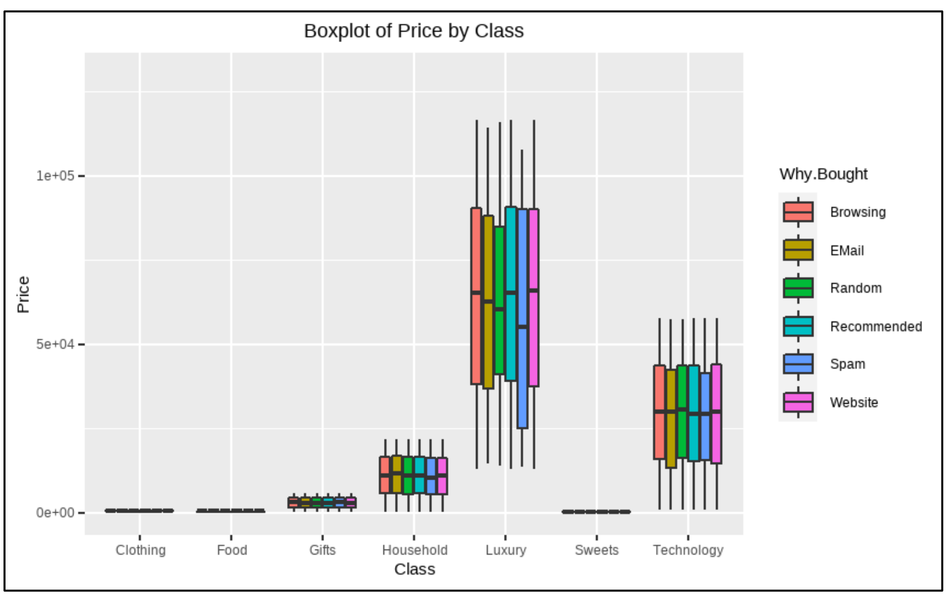


*Figure 47: Boxplot displaying relation between the reason for buying a product and the Price by class*

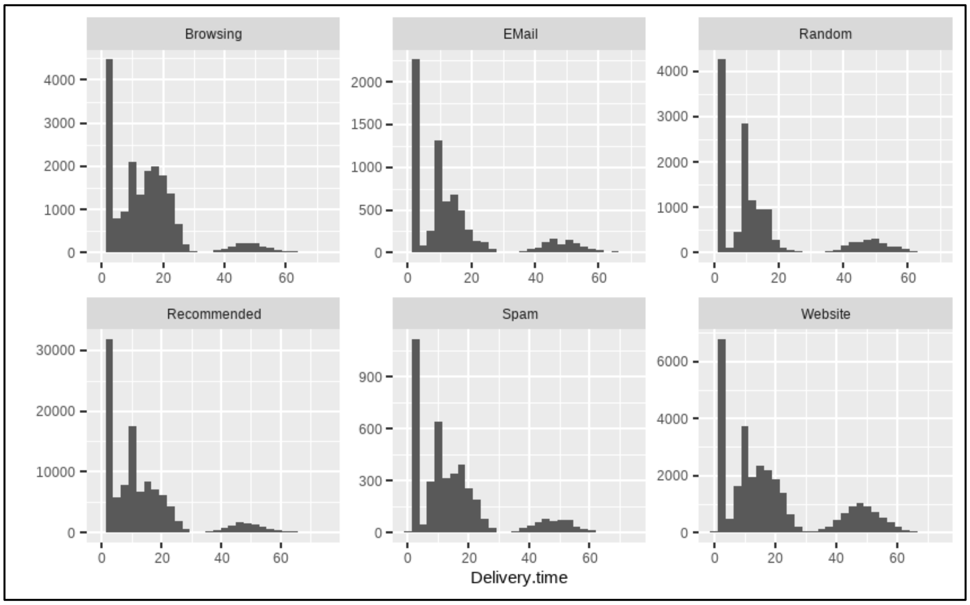


*Figure 48: Delivery.time distribution over reasons for Buying a product.*


*Figure 49: Boxplot displaying relation between the reason for buying a product and the Delivery.time by class*

The p-value for the MANOVA was 2.2e-16. Therefore, it is argued that at least one of the attributes (delivery time or price) influences "Why.Bought" and therefore the null hypothesis must be rejected. If the delivery time and price are compared across the various classes, they will be varied for various reasons for buying. The strongest link between "Why.Bought" and "Delivery Time" may be implied by the fact that delivery time has the biggest F value.

# 6 Reliability of the Service and Products

## 6.1 Lafrideradora

### 6.1.1 Problem 6 p.363

One of the company's subsidiaries, Lafrigeradora, manufactures parts for the cooling systems. A refrigeration part must meet the following standards for thickness: 0.06 ± 0.04 cm

The loss (L) = k(y–m)^2 = 45 when (y-m)=0.04,
Thus, k = 45/(0.04^2) = 28125,
Which makes the Taguchi loss function:
L = 28125(y-0.06)^2

In figure 50, it is plotted. The cost increases as the divergence grows. Therefore, minimising deviation should be the main objective since the loss is the smallest near 0 deviation.

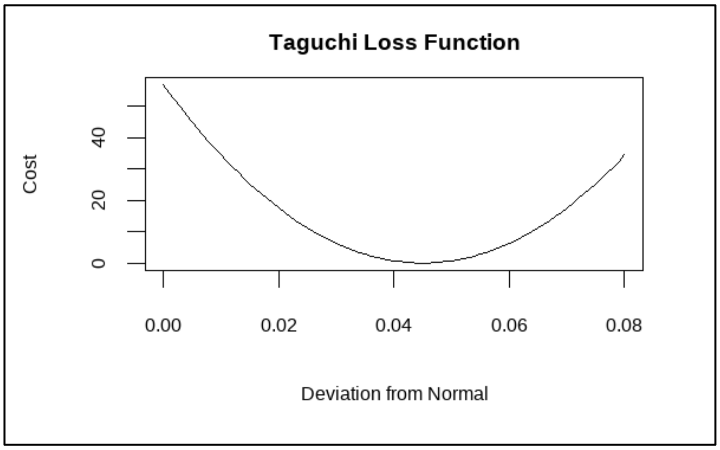


*Figure 50: Taguchi Loss Function @ original scrapping cost*

### 6.1.1 Problem 7 a p.363

The new scrapping cost is $35 per part:

Thus, k = 35/(0.04^2) = 28175,
Which makes the Taguchi loss function:
L = 28175(y-0.06)^2

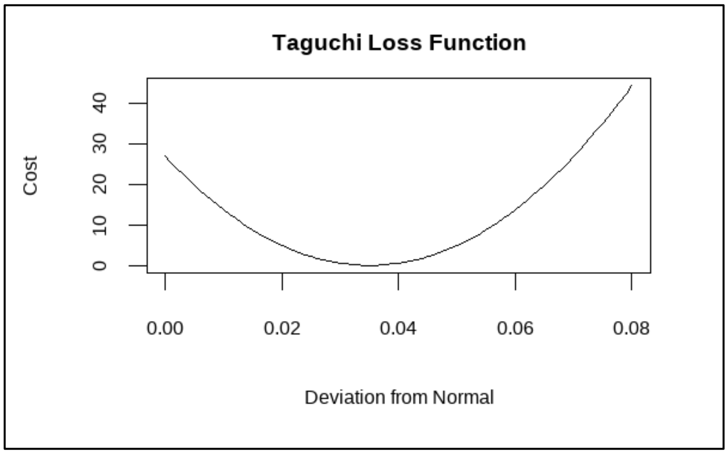


*Figure 51: Taguchi Loss Function @new scrapping cost*

In figure 51, it is plotted. The cost increases as the divergence grows. Therefore, minimising deviation should be the main objective since the loss is the smallest near 0 deviation.

### 6.1.1 Problem 7 b p.363

The process deviation can be reduced to 0.027 cm:

Which makes the Taguchi loss function:
L = 21875*0.027^2 = 15.95

## 6.2 Magnaplex

## 6.2.1 Problem 27 a:

If only one machine is operational in each of A, B, and C, the reliability is simply:

*Figure 51: Taguchi Loss Function @ new scrapping cost*

$$Rs = R1 \times R2 \times R3 = 0.85 \times 0.92 \times 0.90 = 0.7038$$

Thus, the system is only 70.38% reliable

## 6.2.1 Problem 27 b:

The first process will be more reliable if both of each machine, A, B, and C, are operational. The combined dependability of components A that are connected in parallel is 1 - probability (both fail).

When both A's fail, the combined reliability is RA = 1- 0.0225 = 0.9775, with a failure probability of (1-0.85)^2 = 0.0225.

The total reliability of both B's is RB = 1- 0.0064 = 0.9936, and both B's fail with a probability of (1-0.92)^2 = 0.0064.

The combined dependability of both C's is RC = 1- 0.01 = 0.99 when they fail with a probability of (1-0.9)^2 = 0.01.

The Reliability is simply:

$$Rs = R\mathrm{A} \times R\mathrm{B} \times R\mathrm{C} = 0.9775 \times 0.9936 \times 0.99 = 0.9615$$

Thus, the system's reliability has increased by 25.77%

## 6.3 Delivery process

reliability = ((190+22)/1560)*(1559/1560) = 0.1358103
dbinom(days,365, reliability,log=F) = 49

The process will be reliable for 49 days per year.

The approach should be reliable for the following if we raise the fleet size to 21 vehicles:

reliability = ((1560-3-1)/1560)*(1559/1560)= 0.9967965
dbinom = dbinom(days,365, reliability,log=F) = 364

If there are 21 vehicles, the procedure will be reliable for 364 days a year. Therefore, having more vehicles is better.

## Conclusion

After the sales data was analysed, the company functions were better understood. The company will considerably benefit from the recommendations that were made based on the analysis's findings.

The selling of the classes of sweets, food, and clothing needs to be examined because it was discovered that they generate very little income compared to the other classes. Removing these goods will also make it easier to concentrate on other significant categories, including technology and luxury, which have the highest income by a considerable margin. Technology delivery is under control, while luxury is out of control. But it is declining, which is a good thing. Household goods, and gift deliveries are all completely out of control. Further examination of the technology delivery process revealed that, in order to maximise profit, the average delivery time must be cut by three hours. The results of the MANOVA research showed that the factors that affect "Why.Bought" include delivery time, and price. The other business functions were examined, and it was found that overall reliability is quite high. The company's overall priorities should be technology- and luxury goods.

## References

MANOVA test in R: Multivariate analysis of variance STHDA. [Online]

Available at: http://www.sthda.com/english/wiki/manova-test-in-r-multivariate-analysis-of-variance

[Accessed 10 October 2022]