ECSA Report

ECSA GA4 Project 2022

Matthew Ferrant - 22639268

10/10/2022

Table of Contents

Introduction

The following report involves an in-depth analysis of the client data for an online business. The dataset is cleaned and stored into useful subsets in which analysis is done. The analysis of this data is done by means of descriptive statistics, where key findings of the data are visually and analytically interpreted via a number of various plots and comments. Statistical Process Control (SPC) for the item class delivery times is done via a series of X- and S-Charts in order to get a better understanding of the variation and fluctuations of the respective delivery times. These item class delivery times are then analyzed with respect to the SPC in an attempt to optimize these delivery processes. A MANOVA Test is then carried out in order to gain a more significant statistical understanding of the relationships within the data – with reference to price and delivery time across all item classes. Hypothetical example problems are then done with regards to subsidiaries of the online business. The report concludes with a number of key insights taken from the analysis of the data, as well as recommendations with respect to service delivery times and process control.

Part 1 – Data Wrangling

The csv file of ‘salesTable2022’ with 180 000 instances was analyzed. Several NA (missing) values and negative values were found under ‘Price’ feature. All rows containing the above-mentioned invalid data values were removed and stored in a separate csv file named ‘IncompleteData.csv’. The remaining valid data was stored in a csv file named ‘CompleteData.csv’.

- ‘IncompleteData.csv’ = 22 instances
- ‘CompleteData.csv’ = 179978 instances

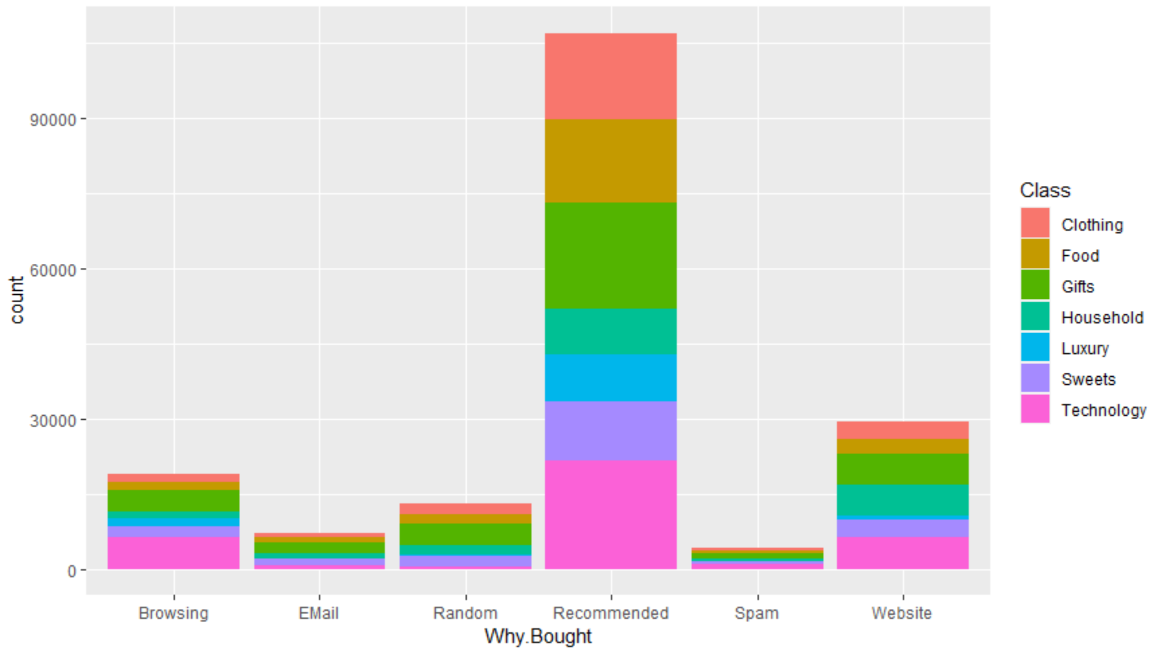
The respective csv files ‘IncompleteData.csv’ and ‘CompleteData.csv’ were then filtered in descending order with reference to the ‘X’ feature.

Part 2 – Descriptive Statistics

The following section will highlight and describe the key statistical insights found from analyzing the new valid dataset, ‘CompleteData.csv’.

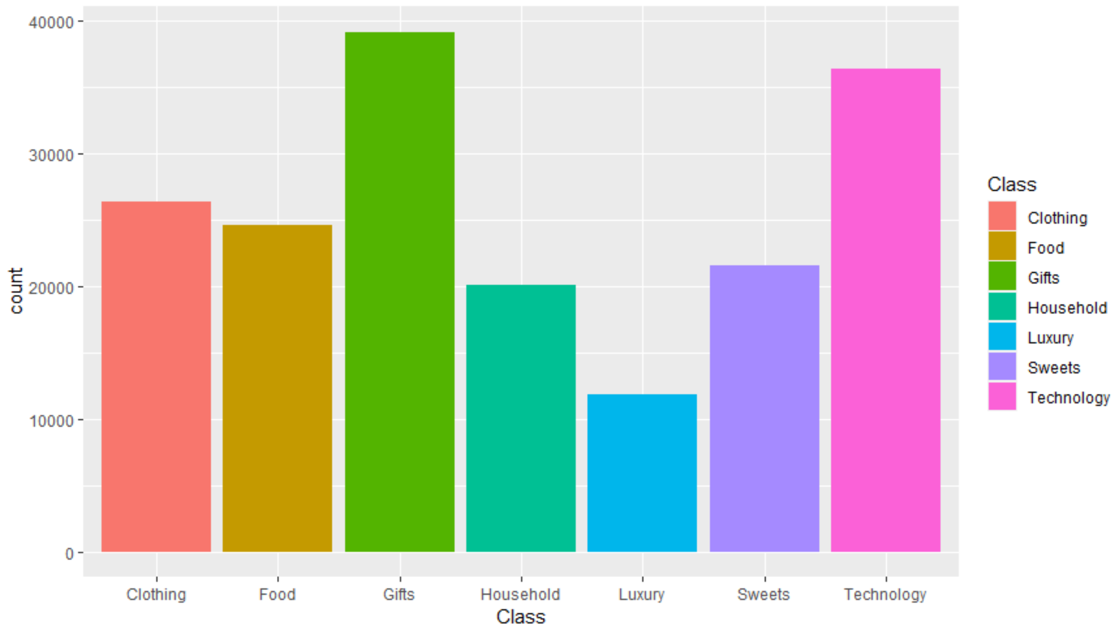
A summary of the valid data gives a rough insight into the patterns and correlations that could be relevant with regards to the analysis of the data set. The following insights were thus found and analyzed:

Why.Bought vs Class



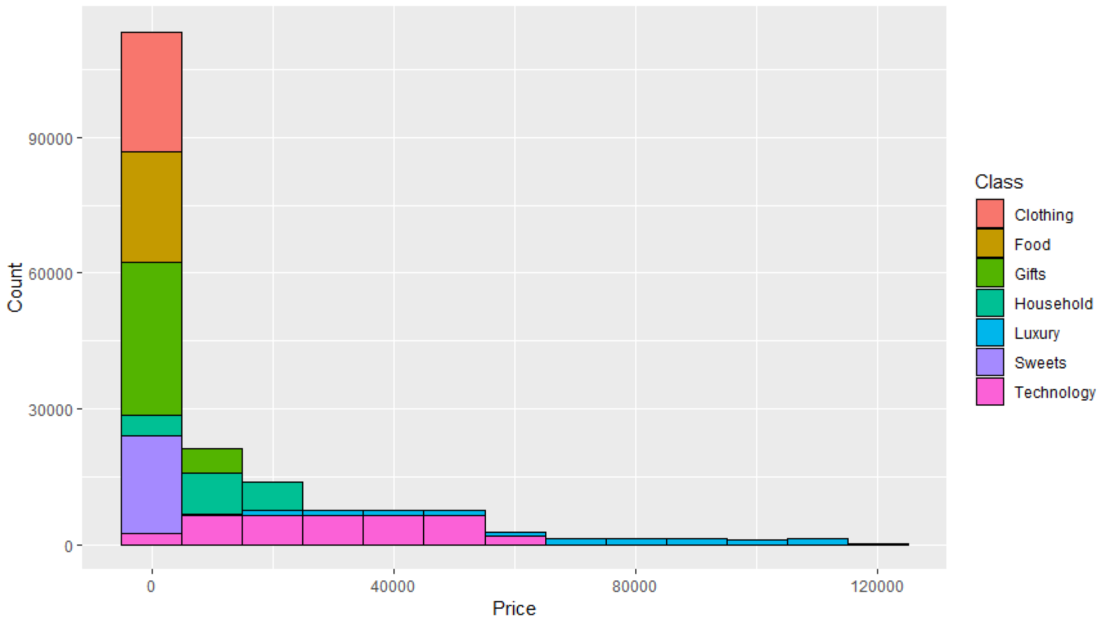
From the above bar graph we can see that the majority of purchases are made due to ‘Recommended’. This is also true for all purchase classes – Clothing, Food, Gifts, Household, Luxury, Sweets, and Technology. This is positive as it shows that many sales are generated based on customers giving positive recommendations to other potential customers. The ‘Website’ is also responsible for just over 30 000 sales – this is again positive as it could show that the company’s website is aesthetically pleasing and easy to navigate. The minority of the purchases made are results from ‘Email’ and ‘Spam’. Although these classes still result in sales, potential customers may get annoyed with the overload of emails and spam and as a result may refrain from purchasing products from the company in the future. This is useful information for the company as it gives a valuable insight as to what does and doesn’t attract customers into buying their products.

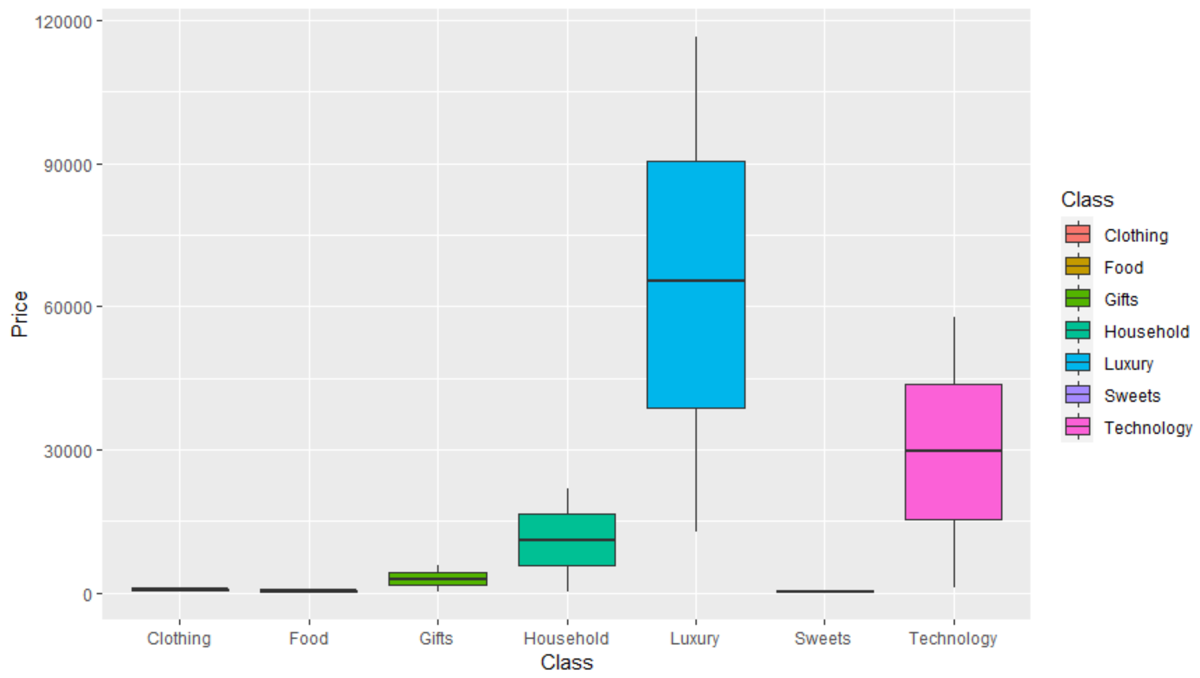
Class



From the bar graph above we can see that the majority of items sold are ‘Gifts’ and ‘Technology’. The least sold items are ‘Luxury’ goods.

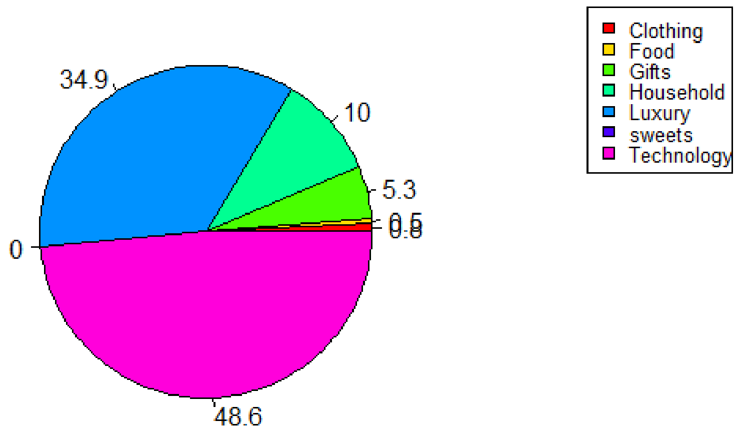
Price vs Class





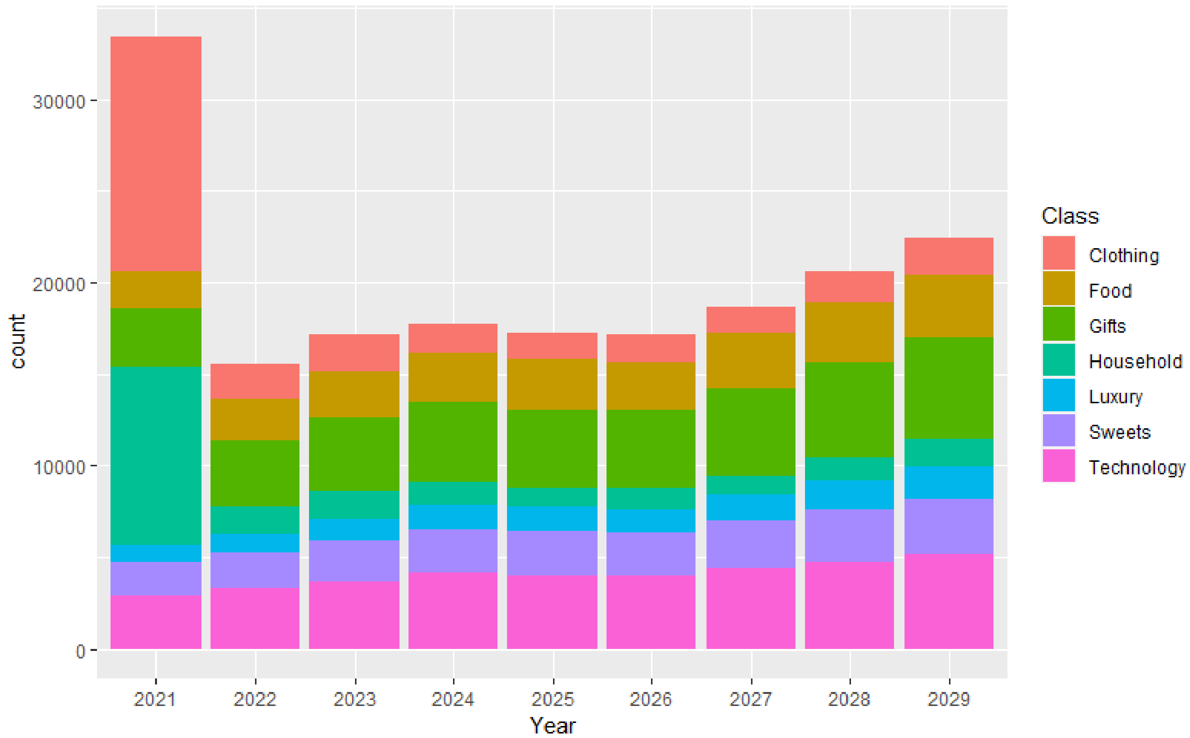
However, from the above histogram and boxplot we can see that the ‘Price’ is the highest for ‘Luxury’ items, shortly followed by ‘Technology’ and ‘Household’ items respectfully. The ‘Price’ of the other product classes (‘Clothing’, ‘Food’, ‘Gifts’ and ‘Luxury’) are all relatively low. The boxplot above gives a good representation of the distribution of each of the product class prices - highlighting their mean, inter-quartile ranges as well as their maximum and minimum values.

Revenue vs Class



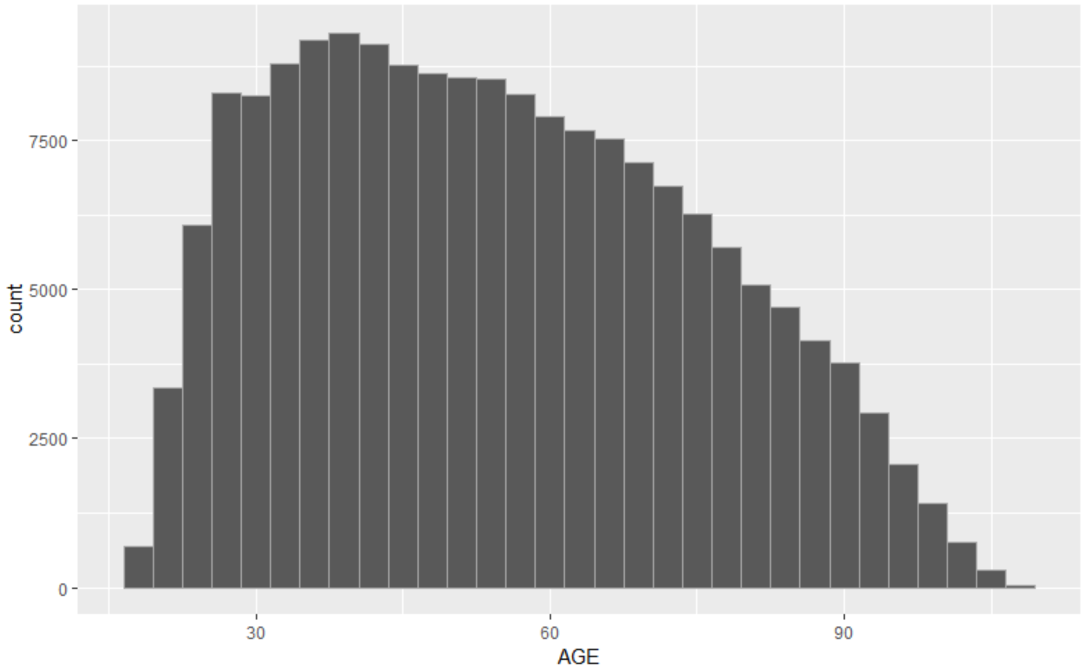
Pre-processing of the dataset was done in order to generate the ‘Revenue’ feature (number of sales with respect to each class multiplied by the price of each of those sales). The above pie chart gives a representation of the proportion of the revenue generated by the company with respect to each of the item classes. It is clear to see that ‘Technology’, ‘Luxury’ and ‘Household’ items generate the majority of the company's revenue (93.5%). The remaining 6.5% of the company's revenue is generated from 'Clothing', 'Food', 'Gifts' and 'Sweets' items. It is therefore recommended that the company's primary focus should be on the item class that bring in the majority of revenue; or to alternatively upscale the minority class departments (if logistically and financially possible) in order to diversify the company's revenue stream in times of market uncertainty or collapse of those majority classes.

Class vs Year



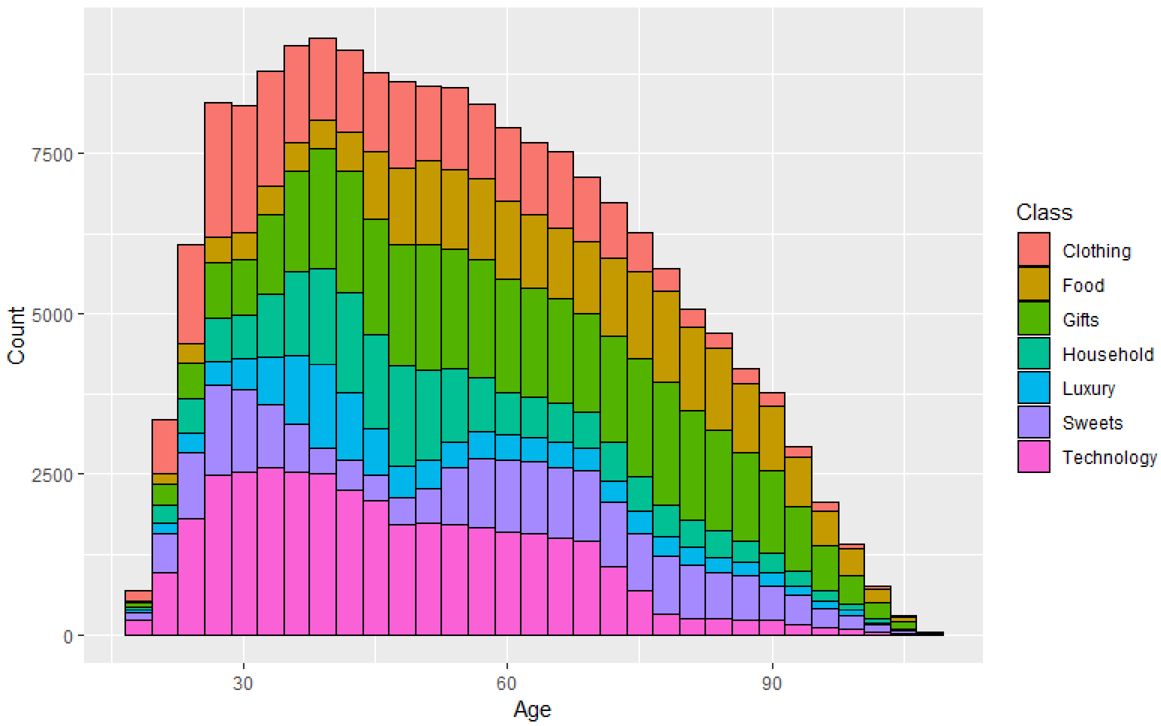
The above bar graph shows the amount of each item class sold with respect to each year. The number of sales for 'Clothing' and 'Household' items was extremely high in 2021 and has since decreased and stabilized over years 2022-2029. This is most likely due to the effects of Covid-19 and lockdown as the average consumer was unable to do their weekly household and clothing shopping at physical stores and subsequently did the majority of their shopping online, benefitting online companies such as yours. In general, the number of sales across all classes combined has been increasing steadily since 2022, with special reference to 'Technology' class items.

Age



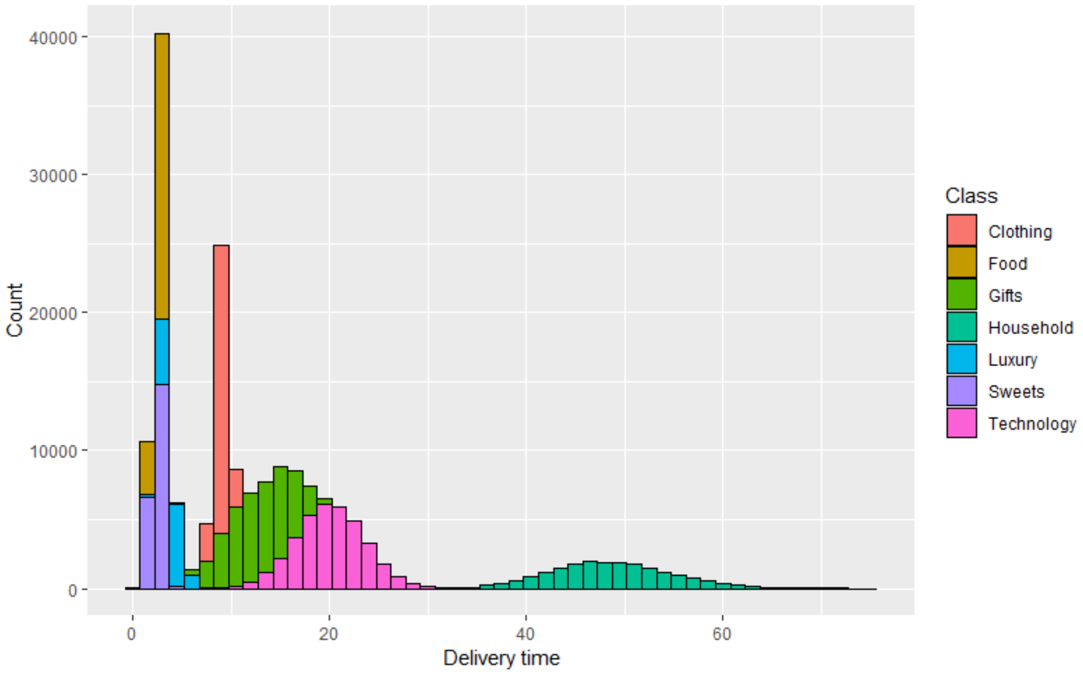
The ages of the company's customers are of unimodal (normal) distribution and skewed to the left. This means that the majority of customers are relatively young (>60 years old).

Sales vs Age

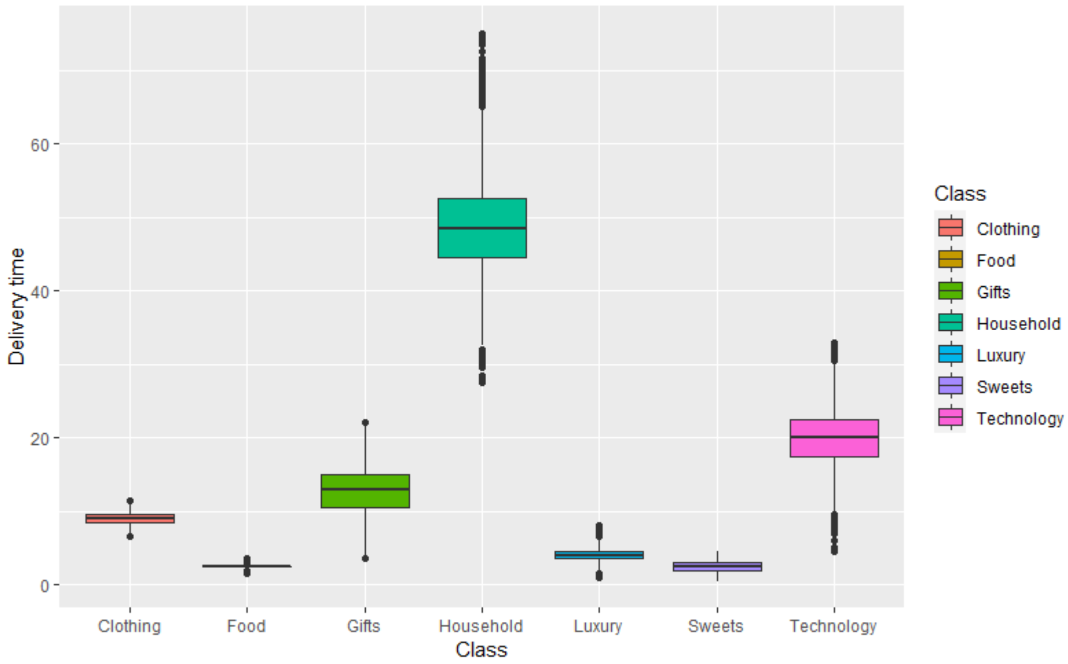


The number of item sales is also of the same unimodal (normal) distribution skewed to the left. However, the above histogram shows that the younger ages (<45 years old) are more likely to be purchasing ‘Technology’ and ‘Clothing’ class items while the older ages (>45 years old) are more likely to be purchasing ‘Food’ and to some degree and surprisingly ‘Sweet’ class items. The above histogram can therefore be used to get a better understanding of the target market of each of the product classes with respect to customer’s age and could be beneficial for potential advertising campaigns.

Delivery Time vs Class



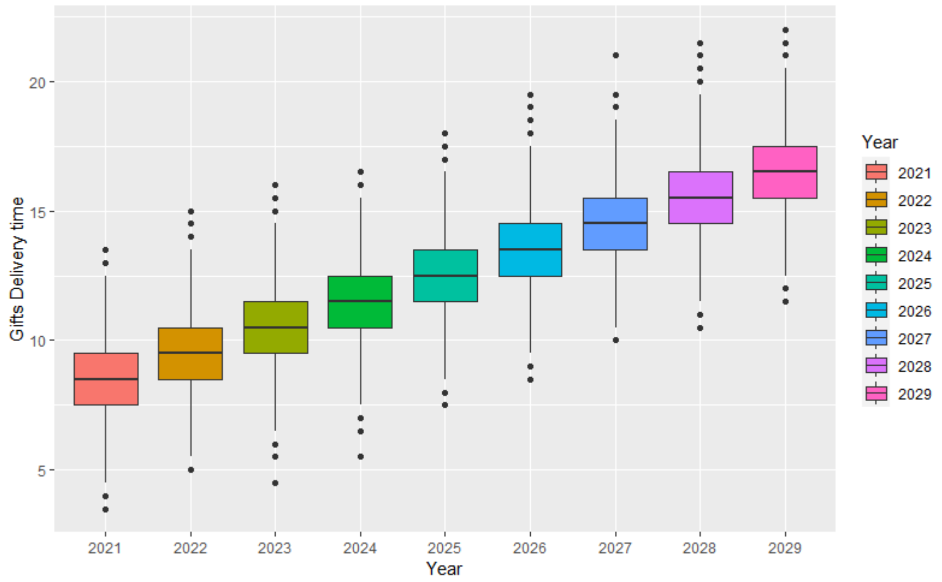
In the histogram above we can see the distribution of the delivery time of each of the item classes. All delivery times are of approximately unimodal (normal) distribution with varying mean and standard deviation values. This will be better depicted using the following boxplot:



In the boxplot above we can clearly see the distribution of the delivery times of the different item classes. 'Household' item classes seem to have largest delivery times. This is most likely due to the fact that household items are usually bulky and may be logistically harder to achieve lower delivery lead times. However, the upper 25% of delivery times are extremely high and customers may be dissatisfied by this long delivery time – a goal of the company should therefore be to try and reduce the upper 25% of delivery lead times for 'Household' class items in order to get a more consistent (less variant) and lower delivery time. 'Food' and 'Sweet' item classes have short delivery times – this is important to maintain as these items are perishable and customers will not be willing to wait any longer for these essentials. The delivery time of 'Gifts' class items is also relatively high and inconsistent – from the customers point of view, these gifts normally have important deadlines (birthdays, etc.) and therefore having such variance in delivery lead times of these "Gifts" item classes may lead to dissatisfied customers.

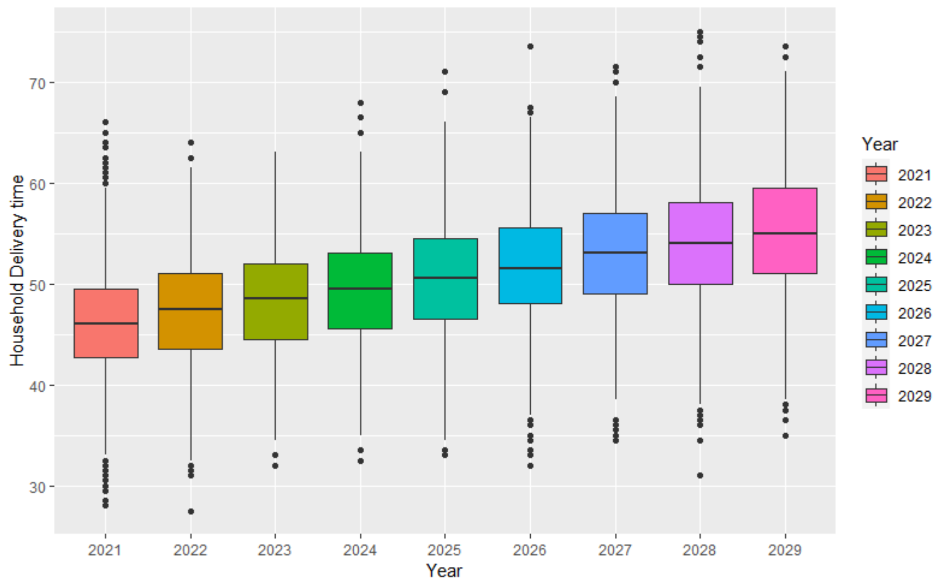
A more detailed and necessary analysis of the 'Gifts' and 'Household' item classes will follow.

Delivery Time vs Gifts



From the boxplot above, it is clear to see that the delivery times of ‘Gifts’ class items have been steadily increasing every year. This is a major problem and should be investigated to find the cause of this, as well as to implement possible solutions.

Delivery Time vs Household



Similar to the delivery times of the ‘Gifts’ item class, the ‘Household’ item class’s delivery lead time has also been increasing steadily every year. Again, this should be investigated and possible solutions should be implemented.

Process Capability Indices Calculations (Technology Class)

Process capability indices Cp, Cpu, Cpl and Cpk of the process delivery times of technology class items were calculated. USL = 24 hours and LSL = 0 is assumed. The value of the Lower Specification Limit (LSL) = 0 is intuitive as it is the smallest amount of time possible for a process delivery time, hence LSL = 0.

Calculations were done and resulted in the following: (all rounded to 3 significant figures)

- **Capability Potential (Cp) = 1.142**

- **Capabilty Performance (Cpk) = 0.380**
    - Cpu = 0.380
    - Cpl = 1.905

The Capability Potential (Cp) gives an indication of the spread of the process. A Cp of 1.142 indicates that the spread (distribution) of the process fits within the upper and lower service levels 1.142 times – this is relatively small; resulting in a relatively small margin of error with regards to centering the process.

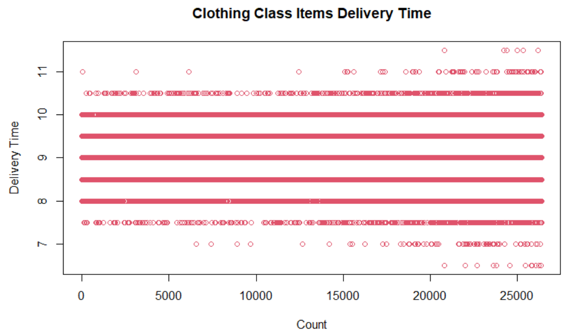
The Capability Performance (Cpk) gives an indication as to how close the process mean is to the target value of the specification. In this case, a Cpk value of 0.380 is observed; indicating a relatively poor fit to the process (a number of delivery times will be above the upper specification limit).

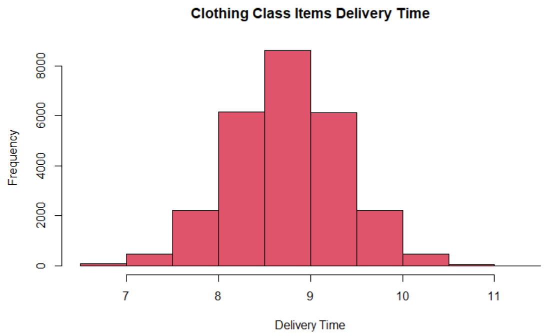
Based on the given USL of 24hours, the process is observed to be out of control and a number of control measures need to be implemented.

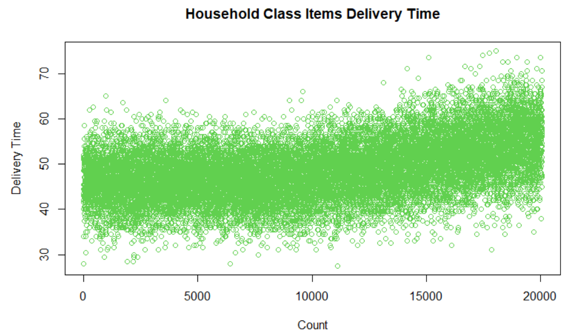
It is recommended to either re-center the process around the target value by reducing delivery time means (i.e. increase Cpk) or to reduce the variation of the spread of the process by redesigning the process itself (i.e. increase Cp). (QA344 Statistics.pdf, 2022)

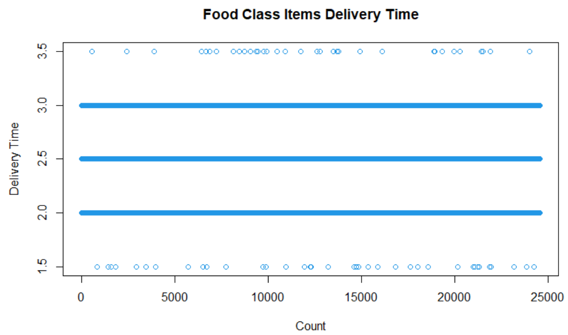
Part 3 - Statistical Process Control (SPC)

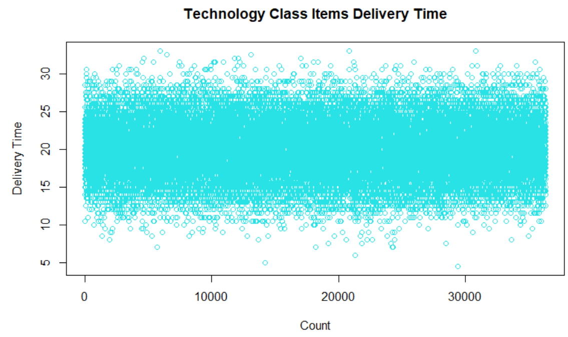
We are now going to do a more in-depth analysis of the distribution of each of the item class's delivery times. As mentioned previously, all distributions are similar to that of a unimodal (normal) distribution with varying means and standard deviations, while some distributions are also slightly skewed to the right or left. A visual summary of the distributions of the above-mentioned item class are given below:

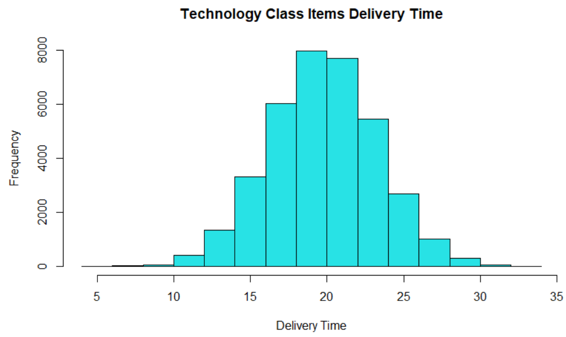






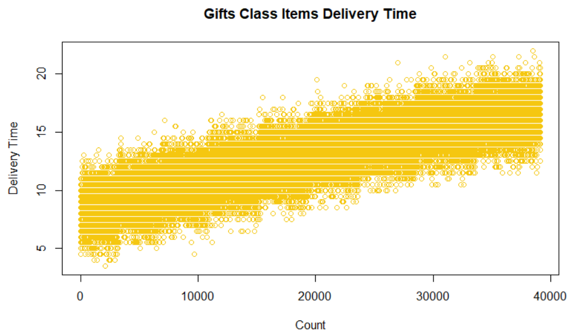



Technology Class Items Delivery Time
Delivery Time
Count
0 10000 20000 30000
5 10 15 20 25 30


Technology Class Items Delivery Time
Frequency
Delivery Time
5 10 15 20 25 30 35
0 2000 4000 6000 8000


Sweets Class Items Delivery Time
Delivery Time
Count
0 5000 10000 15000 20000
1 2 3 4


Sweets Class Items Delivery Time
Frequency
Delivery Time
1 2 3 4
0 2000 4000 6000 8000


Gifts Class Items Delivery Time
Delivery Time
Count
0 10000 20000 30000 40000
5 10 15 20


Gifts Class Items Delivery Time
Frequency
Delivery Time
5 10 15 20
0 1000 2000 3000 4000


Luxury Class Items Delivery Time
Delivery Time
Count
0 2000 4000 6000 8000 10000 12000
1 2 3 4 5 6 7 8


Luxury Class Items Delivery Time
Frequency
Delivery Time
1 2 3 4 5 6 7 8
0 500 1000 1500 2000 2500

3.1)

Iteration 1:

X-Chart

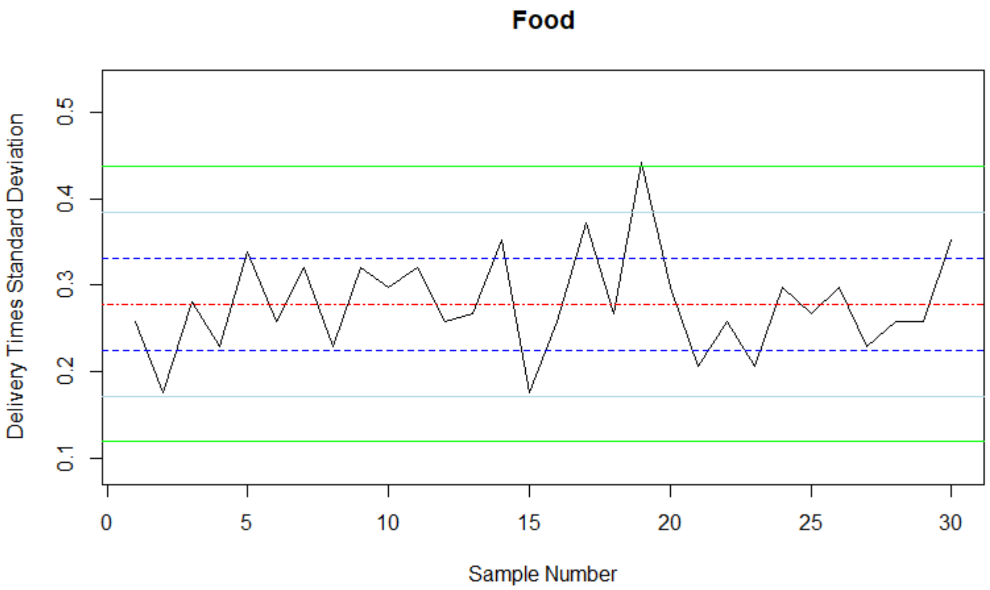
Class	X.UCL	X.U2sigma	X.U1sigma	X.CL	X.L1sigma	X.L2sigma	X.LCL
Clothing	***9.404934***	*9.25059*	*9.110295*	***8.97***	*8.825022*	*8.680044*	***8.535066***
Household	***50.248328***	*48.940248*	*47.751235*	***46.562222***	*45.33352*	*44.104818*	***42.876117***
Food	***2.709458***	*2.631579*	*2.56079*	***2.49***	*2.416847*	*2.343695*	***2.270542***
Technology	***22.974616***	*22.051899*	*21.213172*	***20.374444***	*19.507721*	*18.640997*	***17.774273***
Sweets	***2.897042***	*2.748258*	*2.613018*	***2.477778***	*2.338023*	*2.198269*	***2.058514***
Gifts	***9.488565***	*9.088468*	*8.72479*	***8.361111***	*7.985293*	*7.609475*	***7.233658***
Luxury	***5.493965***	*5.22483*	*4.980193*	***4.735556***	*4.482752*	*4.229949*	***3.977146***

S-Chart

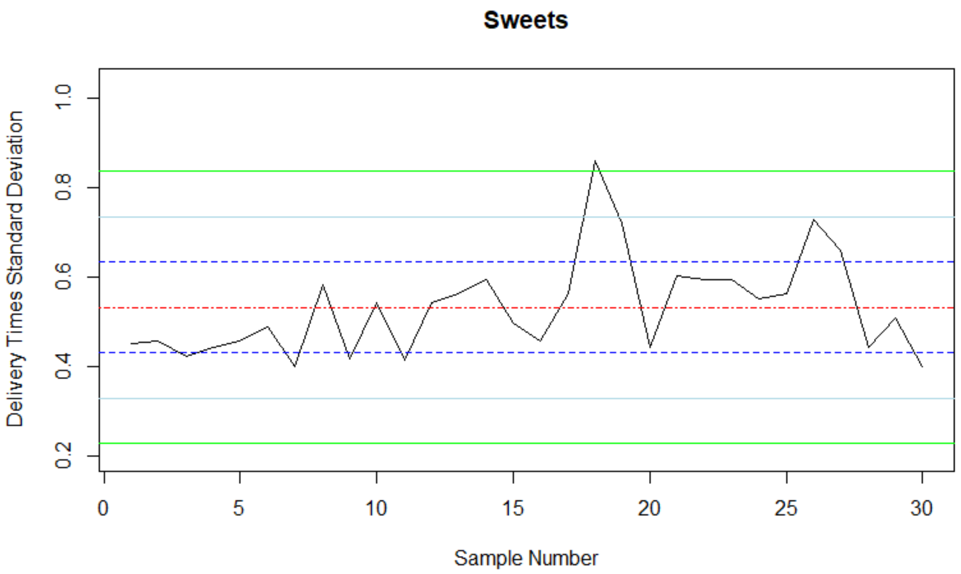
Class	S.UCL	S.U2sigma	S.U1sigma	S.CL	S.L1sigma	S.L2sigma	S.LCL
Clothing	***0.8665596***	*0.7614552*	*0.6563509*	***0.5512465***	*0.4461422*	*0.3410379*	***0.2359335***
Household	***7.3441801***	*6.4534101*	*5.5626402*	***4.6718703***	*3.7811003*	*2.8903304*	***1.9995605***
Food	***0.4372466***	*0.3842133*	*0.33118*	***0.2781467***	*0.2251134*	*0.1720801*	***0.1190468***
Technology	***5.1805697***	*4.5522224*	*3.9238751*	***3.2955278***	*2.6671805*	*2.0388332*	***1.4104859***
Sweets	***0.8353391***	*0.7340215*	*0.6327039*	***0.5313862***	*0.4300686*	*0.3287509*	***0.2274333***
Gifts	***2.2463333***	*1.9738773*	*1.7014213*	***1.4289652***	*1.1565092*	*0.8840532*	***0.6115971***
Luxury	***1.5110518***	*1.3277775*	*1.1445032*	***0.9612289***	*0.7779546*	*0.5946803*	***0.411406***

After detailed analysis of the respective X-Charts and S-Charts above, it was noted that the standard deviation delivery times of both ‘Food’ and ‘Sweets” item classes rose above their respective UCL.

(All other sample instances were found to fit adequately within the LCL and UCL of both X- and S-Charts)



From the above S-Chart (Delivery Times Standard Deviation vs Sample Number) for the ‘Food’ item class, sample number 19 falls just above the UCL. We therefore remove sample 19 from the sample space.



From the above S-Chart (Delivery Times Standard Deviation vs Sample Number) for the ‘Sweets’ item class, sample number 18 falls just above the UCL. We therefore remove sample 18 (as well as sample 19) from the sample space and recalculate the respective X- and S-Charts for all item classes.

Iteration 2: (after removing sample number 18 and 19)

X-Chart

Class	X.UCL	X.U2sigma	X.U1sigma	X.CL	X.L1sigma	X.L2sigma	X.LCL
Clothing	***9.391649***	*9.24721*	*9.102772*	***8.958333***	*8.825561*	*8.681123*	***8.525018***
Household	***50.193653***	*48.96561*	*47.737567*	***46.509524***	*45.334179*	*44.106136*	***42.825395***
Food	***2.705628***	*2.63391*	*2.562193*	***2.490476***	*2.418283*	*2.346566*	***2.275325***
Technology	***22.87439***	*22.013482*	*21.152574*	***20.291667***	*19.513537*	*18.652629*	***17.708944***
Sweets	***2.8857***	*2.750784*	*2.615868*	***2.480952***	*2.342862*	*2.207946*	***2.076205***
Gifts	***9.461593***	*9.087094*	*8.712595*	***8.338095***	*7.986612*	*7.612113*	***7.214597***
Luxury	***5.48688***	*5.23411*	*4.981341*	***4.728571***	*4.482786*	*4.230017*	***3.970263***

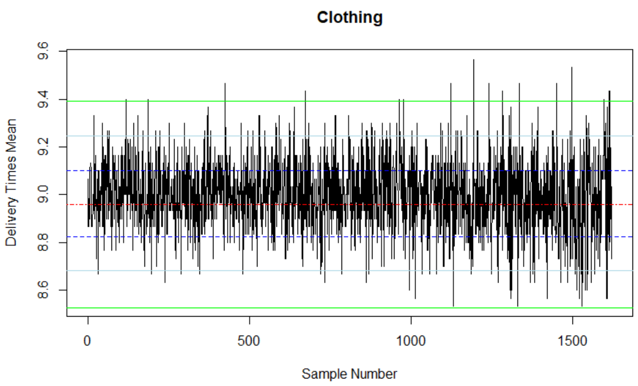
S-Chart

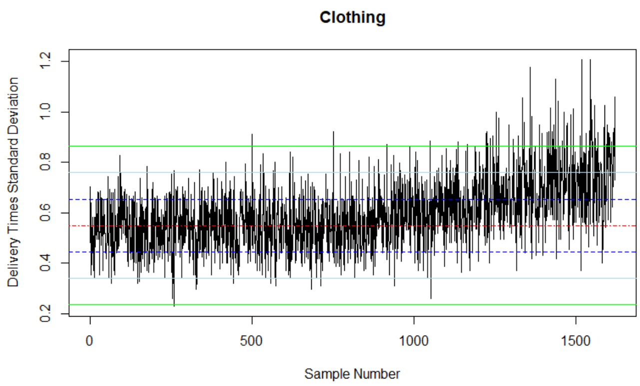
Class	S.UCL	S.U2sigma	S.U1sigma	S.CL	S.L1sigma	S.L2sigma	S.LCL
Clothing	***0.863336***	*0.7586227*	*0.6539093*	***0.549196***	*0.4444826*	*0.3397692*	***0.2350559***
Household	***7.3402419***	*6.4499496*	*5.5596574*	***4.6693651***	*3.7790728*	*2.8887805*	***1.9984883***
Food	***0.4286665***	*0.3766739*	*0.3246813*	***0.2726886***	*0.220696*	*0.1687034*	***0.1167107***
Technology	***5.1458054***	*4.5216746*	*3.8975439*	***3.2734131***	*2.6492823*	*2.0251516*	***1.4010208***
Sweets	***0.8064173***	*0.7086076*	*0.6107978*	***0.5129881***	*0.4151784*	*0.3173686*	***0.2195589***
Gifts	***2.238452***	*1.9669519*	*1.6954518*	***1.4239517***	*1.1524516*	*0.8809514*	***0.6094513***
Luxury	***1.5108506***	*1.3276007*	*1.1443508*	***0.9611009***	*0.777851*	*0.5946011*	***0.4113512***

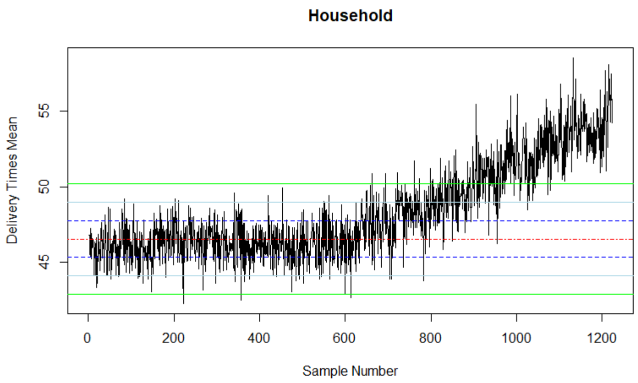
After further analysis of the above updated X- and S-Charts (iteration 2), it was found that no item classes exceeded the UCL or LCL of their respective X- and S-Charts. The above X- and S-Charts will now be used in the analysis of all sample numbers.

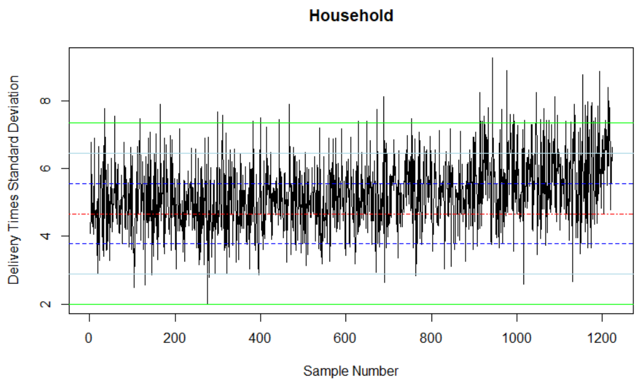
3.2)

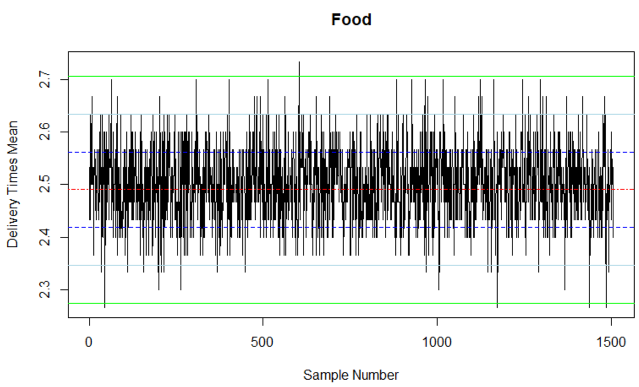
The respective X- and S-Charts of all item classes are visually displayed below – comments will be made appropriately:

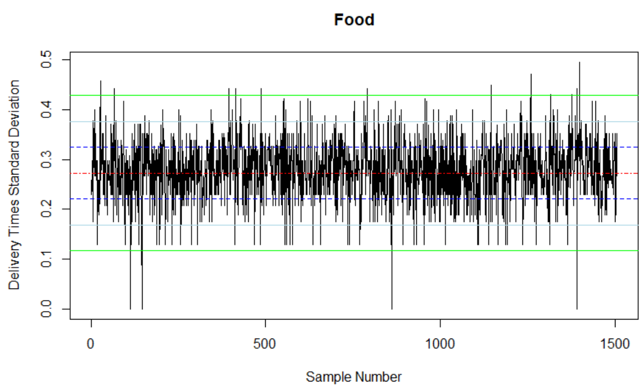


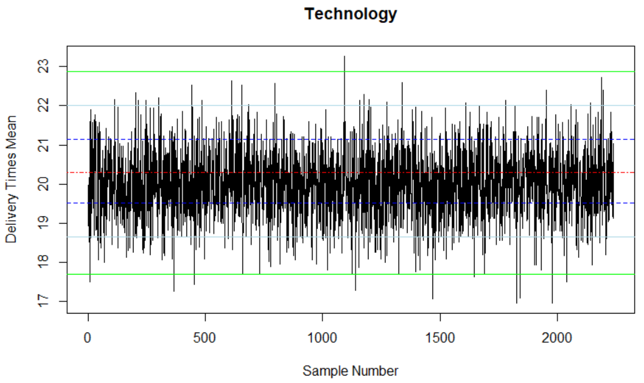


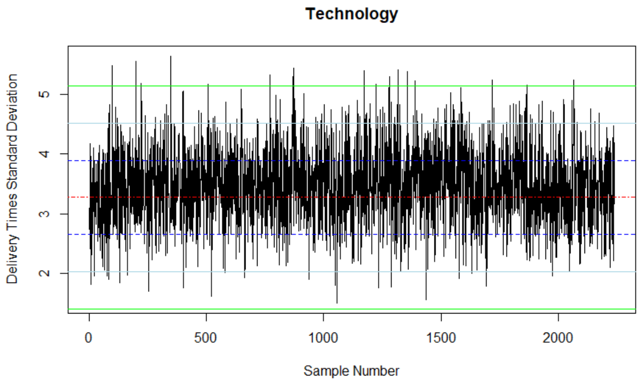


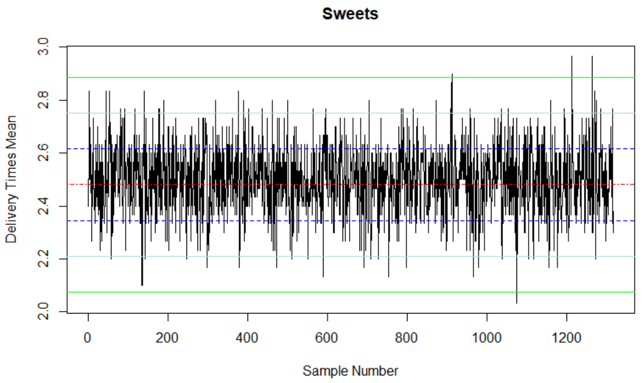


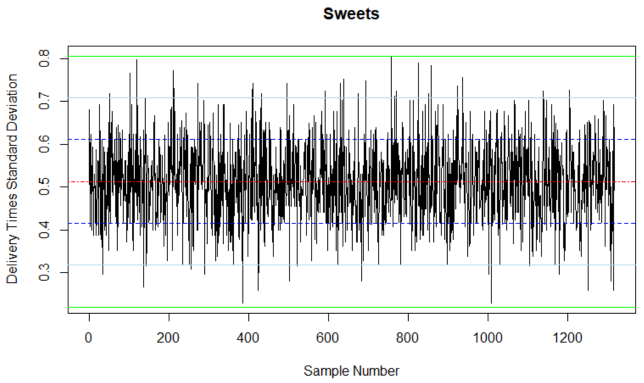


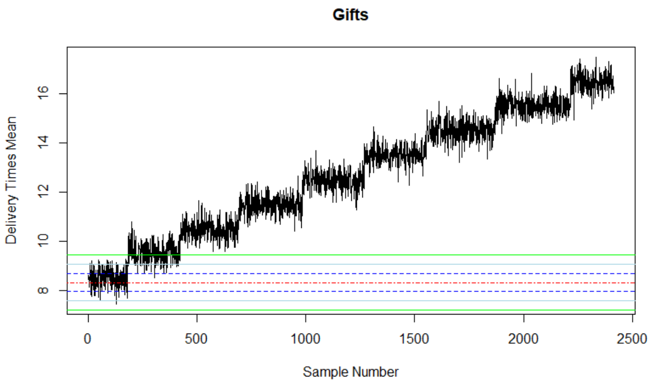


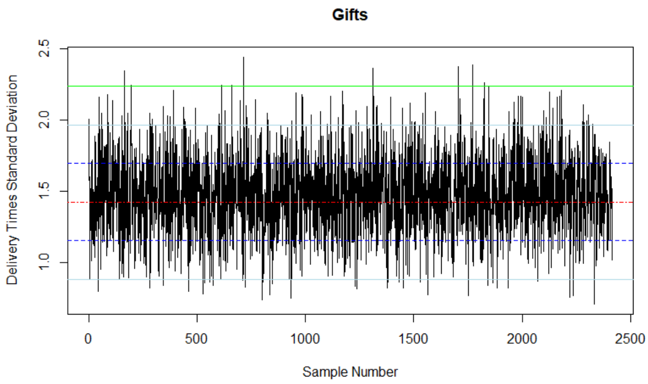

Technology
Delivery Times Mean
Sample Number
17 18 19 20 21 22 23
0 500 1000 1500 2000

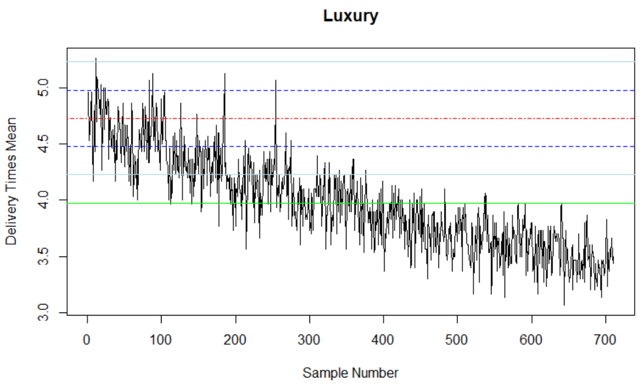

Technology
Delivery Times Standard Deviation
Sample Number
2 3 4 5
0 500 1000 1500 2000

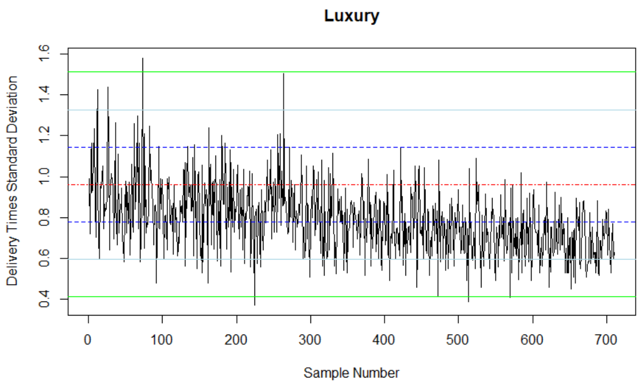

Sweets
Delivery Times Mean
Sample Number
2.0 2.2 2.4 2.6 2.8 3.0
0 200 400 600 800 1000 1200


Sweets
Delivery Times Standard Deviation
Sample Number
0.3 0.4 0.5 0.6 0.7 0.8
0 200 400 600 800 1000 1200


Gifts
Delivery Times Mean
Sample Number
8 10 12 14 16
0 500 1000 1500 2000 2500


Gifts
Delivery Times Standard Deviation
Sample Number
1.0 1.5 2.0 2.5
0 500 1000 1500 2000 2500


Luxury
Delivery Times Mean
Sample Number
3.0 3.5 4.0 4.5 5.0
0 100 200 300 400 500 600 700


Luxury
Delivery Times Standard Deviation
Sample Number
0.4 0.6 0.8 1.0 1.2 1.4 1.6
0 100 200 300 400 500 600 700

With regards to the technology, sweets and food item classes, the high majority of the delivery mean and standard deviation times fits within the upper and lower control limits; thus these processes are said to be in control.

With regards to the clothing and household item classes, there has been a steady increase in the mean and variation of the delivery times over most recent years – they are both above their respective upper control limits. This indicates inconsistent variation within the process and should be addressed and re-centered around their respective center lines.

The gifts item class has been subject to a drastic, although consistent increase in delivery mean time over the years and falls well above the upper control limit. The standard deviation has remained relatively consistent within the upper and lower limits. This is a major problem and should be addressed by means of re-centering the process means accordingly - the process is highly out of control.

Contradicting to the gift item class, the luxury item class is subject to consistent decrease in mean delivery times and falls well below the lower control limit. This, however, is not a bad thing as it shows that the mean delivery times are decreasing, although it is recommended to re-center the process around these new lower mean times in order to maintain control of the process. The standard deviation also falls well within the lower half of the upper and lower control limits, indicating a decrease in variance.

Part 4 – Optimizing the Delivery Processes

4.1)

A)

Analysis was done on the respective X-Charts (shown in 3.2). The following x-bar or sample means were found to be outside the outer control limits of their respective item class delivery times:

Clothing:

- **Number of Samples: 20**
    - First 3 Samples: 148 ; 217 ; 455
    - Last 3 Samples: 1723 ; 1724 ; 1761

Household:

- **Number of Samples: 404**
    - First 3 Samples: 252 ; 387 ; 643
    - Last 3 Samples: 1336 ; 1337 ; 1338

Food:

- **Number of Samples: 6**
    - First 3 Samples: 75 ; 633 ; 1203
    - Last 3 Samples: 1467 ; 1515 ; 1639

Technology:

- **Number of Samples: 13**
    - First 3 Samples: 37 ; 398 ; 483
    - Last 3 Samples: 2009 ; 2071 ; 2424

Sweets:

- **Number of Samples: 6**
    - First 3 Samples: 942 ; 1104 ; 1243
    - Last 3 Samples: 1294 ; 1403 ; 1438

Gifts:

- **Number of Samples: 2296**
    - First 3 Samples: 213 ; 216 ; 218
    - Last 3 Samples: 2607 ; 2608 ; 2609

Luxury:

- **Number of Samples: 435**
    - First 3 Samples: 142 ; 171 ; 184
    - Last 3 Samples: 790 ; 791 ; 792

B)

Analysis was done on the respective S-Charts (shown in 3.2). The most consecutive s-bar or sample standard deviations found between -0.3 and 0.4 sigma-control limits were as follows:

Clothing:

- **Most Consecutive Samples: 4**

Household:

- **Most Consecutive Samples: 3**

Food:

- **Most Consecutive Samples: 5**

Technology:

- **Most Consecutive Samples: 6**

Sweets:

- **Most Consecutive Samples: 5**

Gifts:

- **Most Consecutive Samples: 5**

Luxury:

- **Most Consecutive Samples: 4**

4.2)

Assumption: all processes are perfectly of normal distribution (QA344 Statistics.pdf, 2022)

- The theoretical probability of making a type I error is:

A) P(Z<-3) + P(Z>3) = 2*P(Z<-3) = 2*0.001350 = 0.0027 = 0.27%

- **P(type I error) = 0.0027 = 0.27%**

The above Type I error refers to the probability that an unstable/incorrect x-bar or sample mean (in 4.2 A) was identified, but was found to be stable/correct all along.

B) P(Z<-0.3)+P(Z>0.4) = 0.4236072

- **P(type I error) = 0.4236072 = 42.36072%**

The above Type I error refers to the probability that an unstable/incorrect s-bar or sample standard deviation (in 4.2 B) was identified, but was found to be stable/correct all along.

4.3)

With reference to the delivery process of 'Technology' class items, the following optimal solution (in terms of maximizing value) was found:

- The delivery process should be centered around 23 hours:
  - 26 – i = **23 hours** (where i=3 from calculations done in R)
  - This will cost the company **R324 913.50** to implement

In conclusion, this analysis is similar to that of a Tuguchi Loss as products within specification may still result in a loss – highlighting that customers are most satisfied when a product is perfectly on target (and that any deviation from this target may result in a loss). The delivery process should therefore be centered around a mean of 23 hours (instead of 26 hours). This will cost R324 913.50 to implement, but will optimize the value generated from this process. (QA344 Statistics.pdf, 2022)

4.4)

With reference to the delivery process of 'Technology' class items, the likelihood of making a type II error (given delivery process average = 23 hours) is as follows:

- pnorm(3) - pnorm(-3)
    - **P(type II error) = 0.9973002 = 99.73002%**

The above Type II error refers to the probability that a stable/correct x-bar or mean (in 4.2 A – technology class) was identified, but was found out to be unstable/incorrect all along. This is extremely high at 99.73002%, indicating that most stable/correct predictions are in fact unstable/incorrect. (QA344 Statistics.pdf, 2022)

Part 5 – DOE and MANOVA

The following MANOVA test include the analysis of the statistical significance between the price and delivery time of products across the different item classes (ie. whether the delivery time is influenced by the price of the item or not)

Assumption: all prices and delivery times are of normal distribution (QA344 Statistics.pdf, 2022)

- According to the data, the distributions of price and delivery times are not perfectly normally distributed (although they are similar), as shown in Part 2 and 3; but for the case of the below MANOVA Test, they are assumed to be so.

MANOVA Test:

- **Ho: Null Hypothesis**

Delivery Time is statistically dependent on Price across all item classes

  - $\mu_{Price,Clothing} = \mu_{Price,Household} = \mu_{Price,Food} = \mu_{Price,Technology} = \mu_{Price,Sweets} = \mu_{Price,Gifts} = \mu_{Price,Luxury}$
  - $\mu_{DeliveryTIme,Clothing} = \mu_{DeliveryTime,Household} = \mu_{DeliveryTime,Food} = \mu_{DeliveryTime,Technology} = \mu_{DeliveryTime,Sweets} = \mu_{DeliveryTime,Gifts} = \mu_{DeliveryTime,Luxury}$

- **Ha: Alternative Hypothesis**

Delivery Time is not statistically dependent on Price all item classes

  - $\mu_{Price,Clothing} \neq \mu_{Price,Household} \neq \mu_{Price,Food} \neq \mu_{Price,Technology} \neq \mu_{Price,Sweets} \neq \mu_{Price,Gifts} \neq \mu_{Price,Luxury}$
  - $\mu_{DeliveryTIme,Clothing} \neq \mu_{DeliveryTime,Household} \neq \mu_{DeliveryTime,Food} \neq \mu_{DeliveryTime,Technology} \neq \mu_{DeliveryTime,Sweets} \neq \mu_{DeliveryTime,Gifts} \neq \mu_{DeliveryTime,Luxury}$

The above MANOVA Test was conducted and resulted in a p-value of 0. 00000000000000022 (≈0). This is significantly less than the alpha value of 0.05 - therefore the Ho (null hypothesis) is rejected; that all delivery times are statistically dependent on the price of the item within that class. The Ha (alternative hypothesis) is thus true; indicating that for at least one item class there is no significant statistical dependency on the price of the item with respect to its delivery time. In other words, the delivery time of an item across at least one class is not dependent on the price of the item.

This is an interesting insight as it reveals that a more expensive item within a certain class will most likely not result in a quicker or slower delivery time, but rather that the delivery time of any item within a class is sampled independently and randomly from the unimodal (normal) distributions indicated in Part 3.

Being very critical, this could be seen as a slight problem as customers who buy very expensive items (and thus generate large amounts of revenue for the company) are subject to a random normal distribution with regards to delivery time, which could result in customer dissatisfaction and potential loss of future sales from such big spending customers – this highlights the analogy of "vital few versus the trivial many".

Systematic measures regarding the priority of the delivery times for these so-called "vital few" could be implemented in order to gain greater customer satisfaction, and thus greater potential sale opportunities. However, this implementation should not be at the absolute detriment of the "trivial-many" as although they don't bring in as much revenue, they are still a key part of the company's consistent revenue stream as well as form the basis for recommendations to potential future customers – which, as highlighted in Part 2, is responsible for the vast majority of the 'Why.Bought' feature.

Part 6 – Reliability of the Service and Products

6.1)

Problem 6:

- L = k(y–m)^2 = 45 when (y-m)=0.04
- k = 45/(0.042) = 28125

<u>Taguchi Loss Function:</u> **L = 28125*(y-0.06)^2**

The above Taguchi Loss Function indicates that any slight deviation of the actual size of the part (indicated by y) with respect to the target value (0.06) will result in a 'loss' even if the part itself is within specification.

Problem 7:

a)

- L=k(y-m)^2 = 35 when (y-m) = 0.04
- k = 35/(0.042) = 21875

<u>Taguchi Loss Function:</u> **L = 21875*(y-0.06)^2**

Similar to the previous Taguchi Loss Function, the above Taguchi Loss Function indicates again that any slight deviation of the actual size of the part (indicated by y) with respect to the target value (0.06) will result in a 'loss' even if the part itself is within specification. The scrap cost of an out of specification part is reduced in Problem 7a (compared to that of Problem 6) and thus the loss function is also reduced.

b)

- L= 21875*(y-m)^2 where (y-m) = 0.027

<u>Taguchi Loss Function:</u> **L = 21875*(0.027)^2 = 15.95**

The above loss can be expected if the process deviation from the target value is reduced to 0.027cm (i.e. y-m = 0.027cm).

6.2)

Problem 27:

a)

- Reliability = Ra*Rb*Rc where Ra=0.85,Rb=0.92,Rc=0.9

Reliability = 0.7038 = 70.38%

The above reliability relates to machines in series at stations A, B and C. Each of the individual machines' reliability is multiplied together to get an overall system reliability of **70.38%.**

b)

- Reliability = Raa*Rbb*Rcc where Raa=1-(1-0.85)^2,Rbb=1-(1-0.92)^2,Rcc=1-(1-0.9)^2
- Reliability = 0.9615316 = 96.15316%

Reliability Improvement = 0.9615316 - 0.7038 = 0.2577316 = 25.77316%

The above reliability measure takes into account adding a 2^{nd} machine at each station A,B and C. The probability of both machines failing is now taken into account when determining the overall system reliability – this resulted in a system reliability of 96.15316% (a **25.77316% improvement** compared to only having 1 machine at each station).

6.3)

Via trial and error, the probability that a truck will be defective (i.e. not working) was found to be p = 0.0077. This value of p=0.0077 is not entirely accurate but corresponds relatively accurately with the information given.

Again, via trial and error, the probability that a truck driver will be defective (i.e. not working) was found to be p = 0.004. This value of p=0.004 is not entirely accurate but corresponds relatively accurately with the information given.

Assumptions:

- p_{truck} = 0.0077
- p_{driver} = 0.004

Therefore, for the delivery process to be reliable, there needs to be at least 19 (>=19) trucks and truck drivers available on any given day.

- For 20 trucks; 21 drivers:
    - Based on the assumptions above, we should expect an average of **361.221 (361) reliable delivery days per year**.
- For 21 trucks; 21 drivers:
    - If we were to increase the number of trucks to 21 (with the same assumptions as above), then we should expect an average of **364.7984 (364) reliable delivery days per year**.

Conclusion

In conclusion, the analysis of the client data led to a number of key insights and understandings. These insights will be briefly revised and recommendations will be made where appropriate.

Technology, Luxury and Household item classes were found to be responsible for the majority of the company's revenue (93.5%) – the company's primary focus should therefore be on the service delivery and process control of these majority classes. Alternatively, the minority item classes (Clothing, Food, Gifts and Sweets) could be up-scaled, if logistically and financially possible, in order to diversify the company's revenue stream in times of potential market uncertainty of these majority classes. A customer demographic with respect to age and item class was also found to be helpful in determining the target market of each of the classes and can be used by the marketing department with regards to potential marketing campaigns.

Statistical Process Control (SPC) found that the delivery process times of the Clothing, Household and Gifts item classes were out of control – these processes should be re-centered around new target mean and standard deviation times in order to gain a greater control of these processes. The MANOVA Test found that there was no clear statistical dependency or correlation between the price and delivery time of items across the various item classes. Priority with regards to delivery times could be placed on more expensive items as to ensure these high-income generating customers are satisfied; although this should not be to the detriment of the average customer.

References

QA344 Statistics.pdf. [online] Available at: <https://learn.sun.ac.za/pluginfile.php/3514418/mod_resource/content/1/QA344%20Statistics.pdf. [Accessed 3 October 2022].