ECSA Quality assurance project

Tiaan van Antwerpen

23537027

2022-10-19

Table of Contents

Abstract

A company engaged in internet sales has supplied sales data. It was established through data investigation and analysis that delivery times, price of sales products, and client age can be used as input into a mathematical modeling providing statistical conclusions and conduct hypothesis testing of the data. The data can be categorized as follows: technology, clothing, gifts, luxury, sweets, food, and household goods. On the delivery times for the sold items, business development systems, such as statistical process control and optimization techniques, can be implemented to enhance the delivery times. Xbar-charts and s-charts will be drawn up for samples of delivery times per class and this can be used to form or draw conclusions on how delivery times can be improved. Additionally, the deployment of conclusion on the service and reliability statistical data can be utilized to solve or appoint a variety of additional problems.

List of figures

List of tables

Introduction

The following project schematic will use the statistical methodology for measuring quality consistency as a structure methodology for quality control.

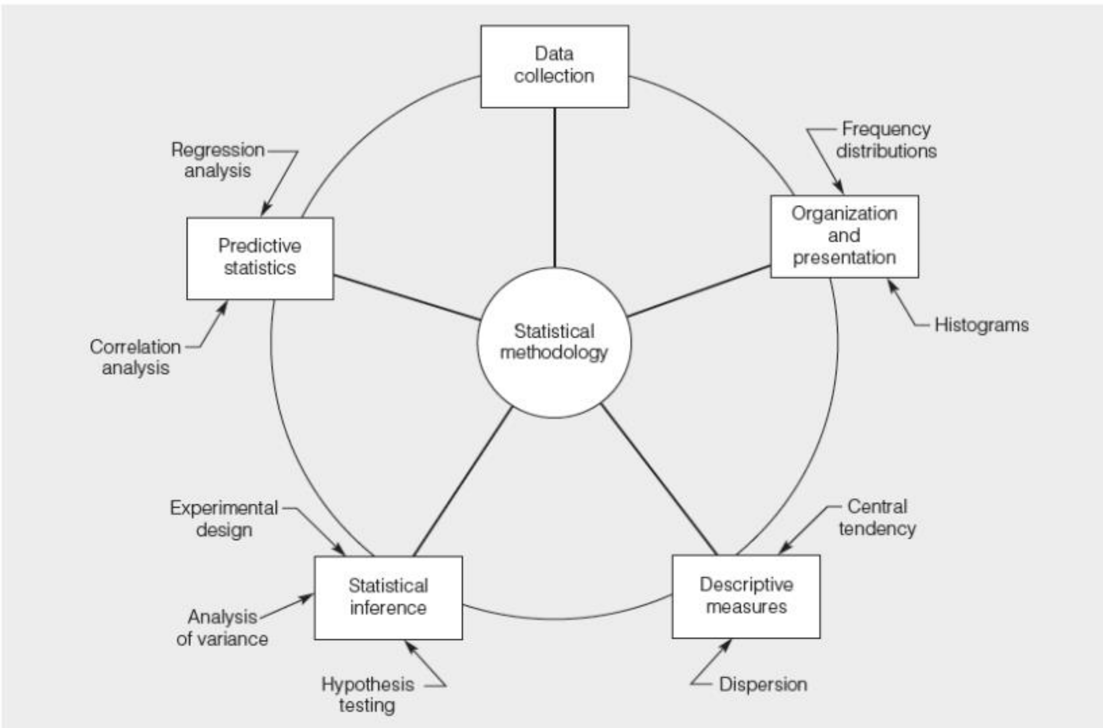


Figure 1: Statistical methodology - R.Evans and M.Lindsay,2014

The data will be sourced, compiled, validated synthesized or analysis producing a data quality report from the data to examine data quality nonconformance, problems, validation issues within the data set. From this data quality report, descriptive statistics will be utilized to graphically and numerically illustrate the sales data and distribution, spread analysis. Following the resolution of data quality issues, data investigation will occur. To better comprehend the data, graphs and plots will be used to portray the analysis results. Understanding the facts will generate certain hypotheses. It is possible to test the hypothesis and make predictions based on a comprehension of the evidence as deduced from the data modeling

1. Part 1

Part 1 of the procedure will involve data interpretation. The data will be collated, validated, reviewed and will be analyzed and the data quality will be reported on to resolve data quality issues observed in the provided data set.

Feature selecting

It is importing to select the features into continuous features and categorical features. The target feature is not specified at the moment but may be introduced later on.

The following features are identified as continuous features: ID, AGE, Price, Delivery time.

The following features are identified as categorical features: The “Class” and “WhyBought”.

A data quality plan will be generated to evaluate important data quality issues in the data set.

1.1.1. The data quality report Continuous features

- For continuous features, the data quality report describes the central tendency and variation of each feature using the minimum, 1st quartile, mean, median, 3rd quartile, maximum and standard deviation,
- For each feature, the number of instances is shown, the percentage of missing values and the cardinality.
- A histogram is created for each continuous feature, except when the cardinality of the feature is less than ten, in this case, a barplot is created instead,

Feature	Count	Miss.	Card.	Min	Q1	Mean	Median	Q3	Max	SD
X	180,000	0	180,000	1.0	45,000.8	90,000.5	90,000.5	135,000.2	180,000	51,961.7
ID	180,000	0	15,000	11,126.0	32,700.0	55,235.1	55,081.0	77,637.0	99,992	25,739.7
AGE	180,000	0	91	18.0	38.0	54.6	53.0	70.0	108	20.4
Price	180,000	17	78,834	-588.8	482.3	12,293.7	2,259.6	15,270.7	116,619	20,889.0
Year	180,000	0	9	2,021.0	2,022.0	2,024.9	2,025.0	2,027.0	2,029	2.8
Month	180,000	0	12	1.0	4.0	6.5	7.0	10.0	12	3.5
Day	180,000	0	30	1.0	8.0	15.5	16.0	23.0	30	8.6
Delivery.time	180,000	0	148	0.5	3.0	14.5	10.0	18.5	75	14.0

Table 1: Data quality report of continuous features

1.1.2. Comment on continuous feature data quality report

The x data and ID has a too high carnality. This could mean that this feature can’t be used for descriptive statistics and merely gives information about the data and its place holder in the data set. Year month and day can’t be used to do calculations on and this can be labeled as a categorical feature in this case.

Features: age, price and delivery time will be plotted to visually examine data quality errors.

Figure 2: Continuous/ numeric features histogram plot

1.1.3. Comments on data quality issues:

There are 17 missing data entries for sale price. There are also negative sales entries, which could not be possible. There are no other data quality issues observed during the examination.

1.1.4. Comments on plots of continuous features

Delivery time seems to have 2 normal distributions in its data that could have an effect on the statistical process control on the data. Most of the deliveries tends to be at a lower delivery time.

The distribution of age is skewed to the right normal distribution. Which means most of the clients tends to be younger. This is plausible since younger people tend to buy more products online.

Most of the priced items is at lower price. This is due to low value items that needs to be sold at higher volumes to generate profits.

1.2. Data quality report: categorical features

Feature	Count	Missing	Cardinality	Mode	Mode Freq.	Mode %	2nd Mode	2nd Mode Freq.	2nd Mode %
Class	180,000	0	7	Gifts	39,154	21.8	Technology	36,350	20.2
Why.Bought	180,000	0	6	Recommended	107,000	59.4	Website	29,450	16.4

Table 2: Data quality report of categorical features

1.2.1. Comments on data quality issues

There are no data quality issues visible. The data will be plotted to draw more conclusions on the distributions of each class group or reason for buying each product.

Figure 3: Categorical features histogram plot

1.2.2. Comments on plots of categorical features

Gifts is the highest sold item with a frequency of 39154 items being sold over a time period of 7 years. Technology items has the second highest frequency with 36350 items being sold over a time period of 7 years. Both these items constituents each more than 20% of all the total sales.

Customers tends to buy more products that is recommended with 107000 products being sold soley from recommendation contributing 59% of the total sales. The company should focus on these types of customers when working on their websites and online sales departments.

Handling/Wrangling data quality issues

A cleaned data set without data quality issues will be created to do data exploration on

The cleaned data set will be modified to create a date column to do better calculations on the data. The first 6 instances of the data set will be displayed to show the new column created and the data in ordered format.

X	ID	AGE	Class	Price	Year	Month	Day	Delivery.time	Why.Bought	Dates
463	47,101	50	Clothing	1,030.9	2,021	1	1	9.0	Recommended	2021-01-01
2,627	88,087	21	Clothing	428.0	2,021	1	1	10.0	Recommended	2021-01-01
3,374	25,418	68	Household	13,184.4	2,021	1	1	48.5	Website	2021-01-01
5,288	13,566	94	Household	7,021.9	2,021	1	1	42.0	Recommended	2021-01-01
8,182	84,692	35	Clothing	475.2	2,021	1	1	9.0	Recommended	2021-01-01
9,272	46,305	72	Clothing	581.0	2,021	1	1	8.5	Random	2021-01-01

Table 3: First 6 data instances

2. Part 2

For part 2 data exploration will be performed to further characterize the data and what correlations we can make with the data.

From the data quality report, a couple of graphs has been generated. Understanding the data steps will require deeper exploration. A correlation matrix will be constructed to identify which continuous features is correlated with each other.

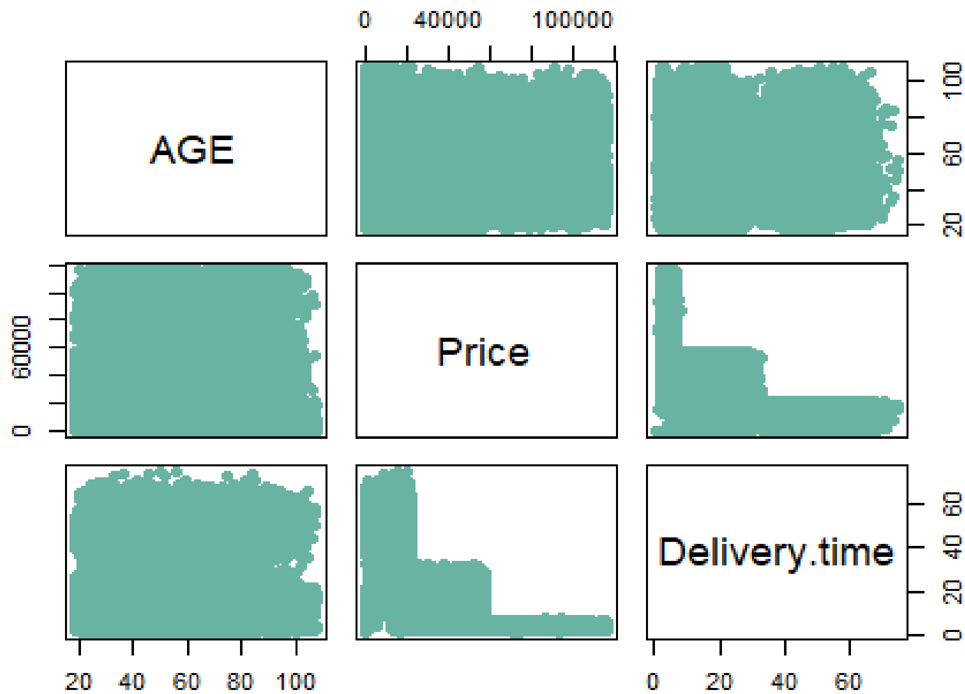


Figure 4correlation matrix of continuous/numeric features

Comments on correlation matrix:

From the plot we can see that price and delivery time is correlated. However it is required to break up each continuous feature in a class and reason for buying the product.

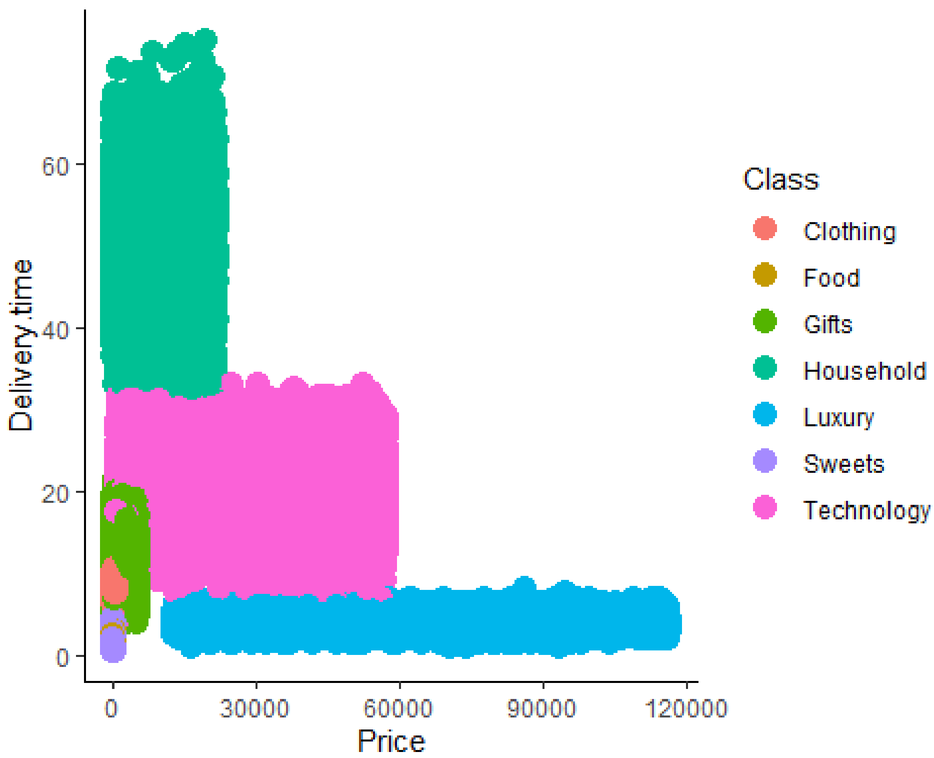


Figure 5: Price per delivery time scatter plot

From the plot it almost appears as if the data from price and delivery time could be used to cluster the data into its respective class. This could be utilized as a prediction model to build to classify data and improve on the companies price to delivery time implementation.

Splitting the data even more to demonstrate the different groups of clusters.

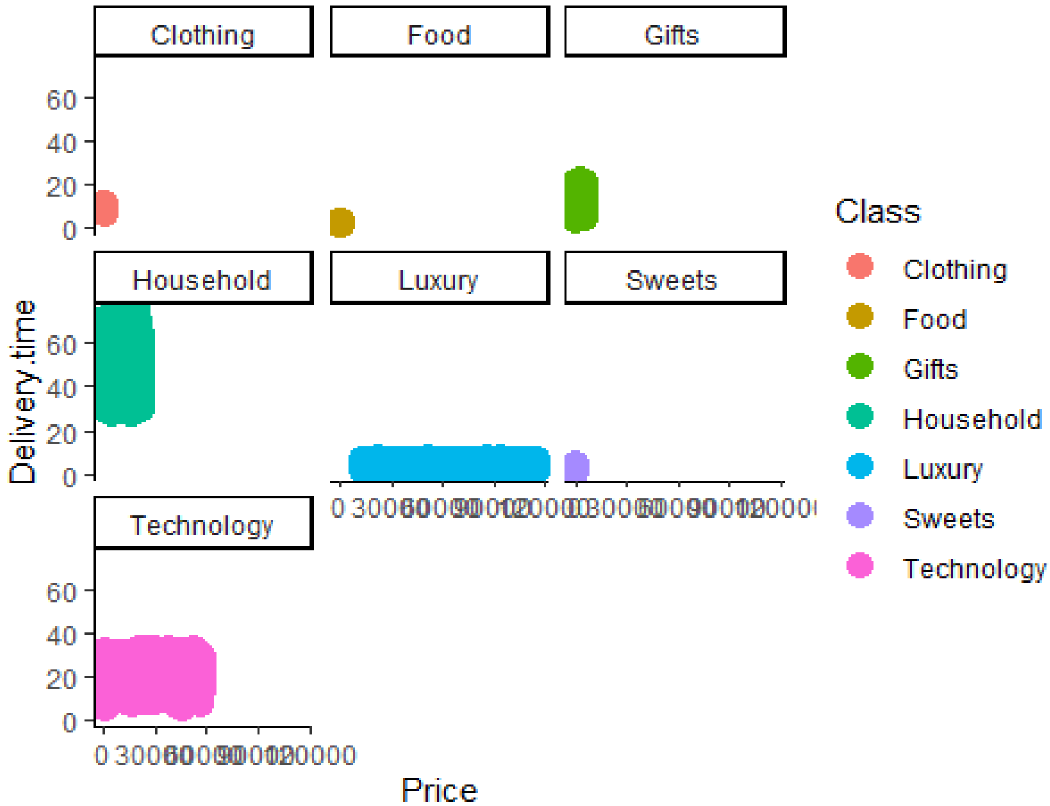


Figure 6: Price per delivery time scatter plot per class

It is evident that each class has its own delivery time and price correlated to it. Luxury items tend to cost the most and it is the most widely distributed from low to high values. Household items and Technology items tend to have the highest delivery times. These items could easily be impacted by stock out items and may influence the companies service levels resulting in customers paying less or being dissatisfied by the long delivery lead times. Food and sweets are classed as low valued items with short delivery time as these items could parish over time.

2.1. Price data exploration

Most information can already be obtained from the data quality report.

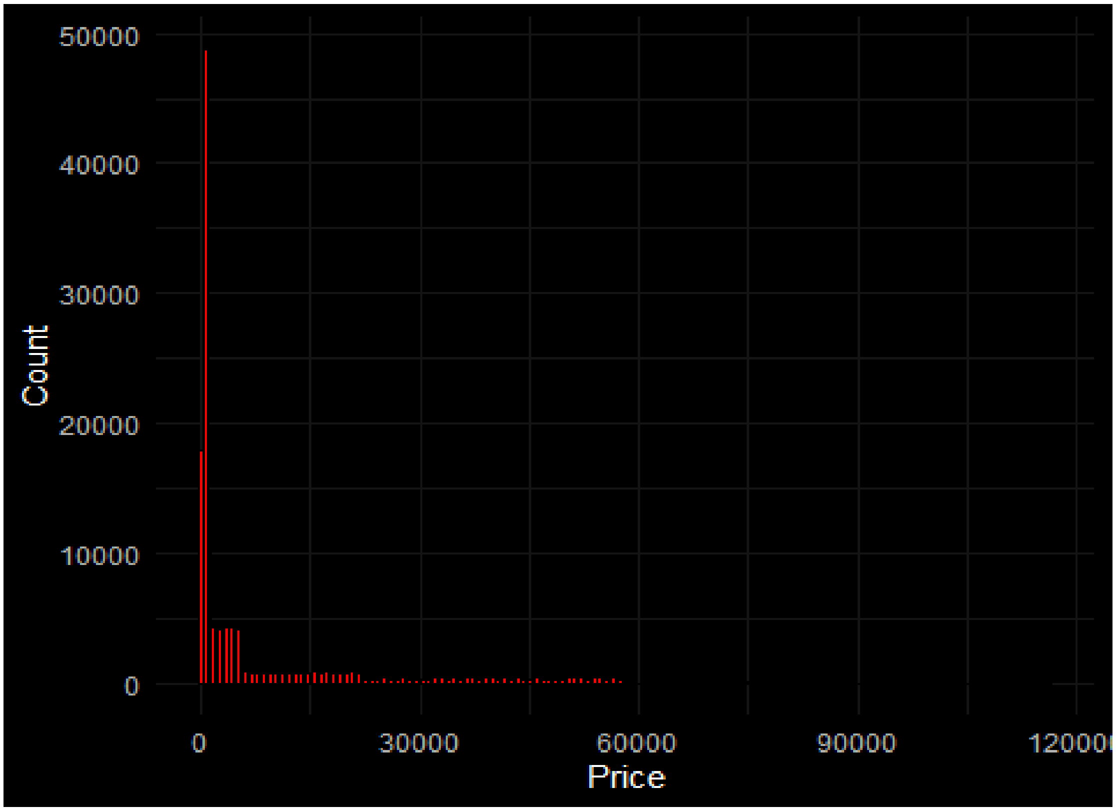


Figure 7: Distribution of priced items

Comment: Low priced products is sold in higher volumes than higher priced products. To generate a profit from low value items, more needs to be sold.

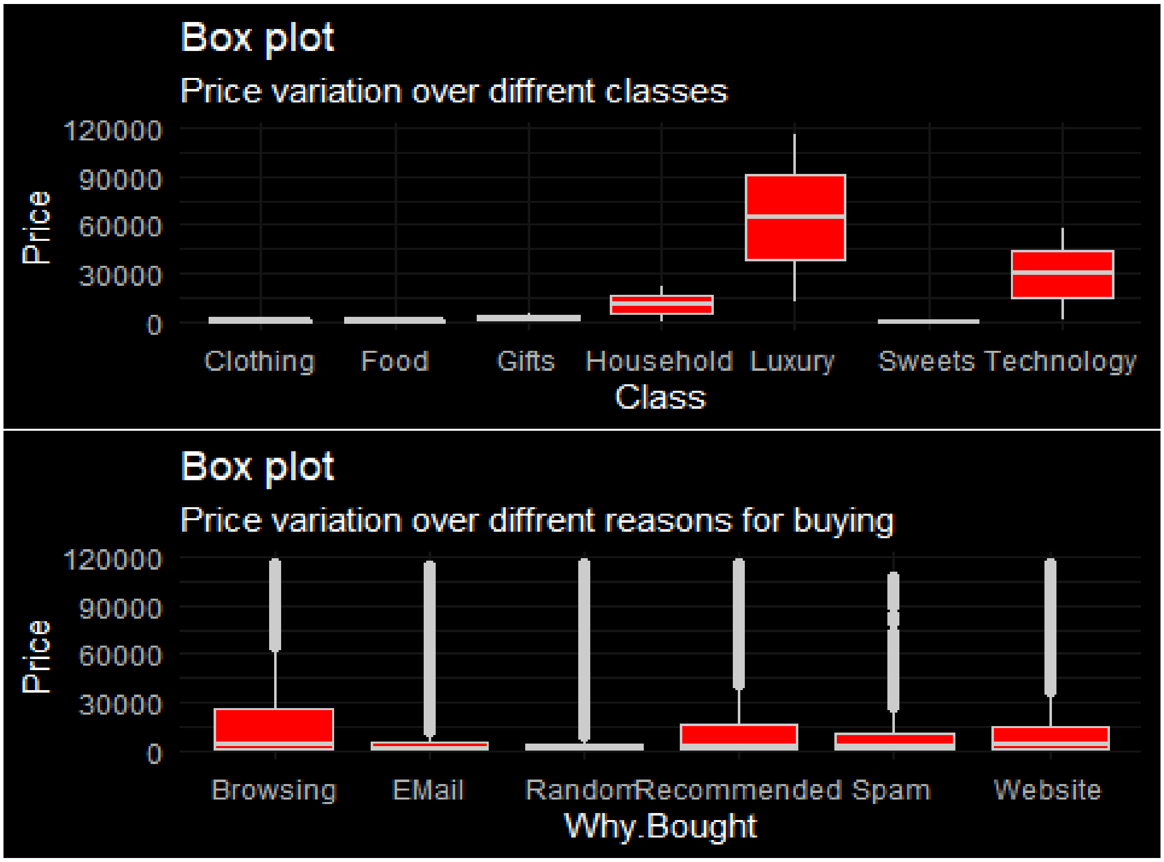


Figure 8: Boxplots for price variation per class and reasons for buying

Observations: Luxury items tend to cost the highest as well as Food. Clothing, food, gifts and sweets can be categorized as low value classes.

Observations: Email, Random and Recommended reasons for buying tends to be customers that will pay lower prices. Even though recommend as a reason for buying is most of the sales, it is usually for low value products. Customers that browse for products tend to pay higher prices.

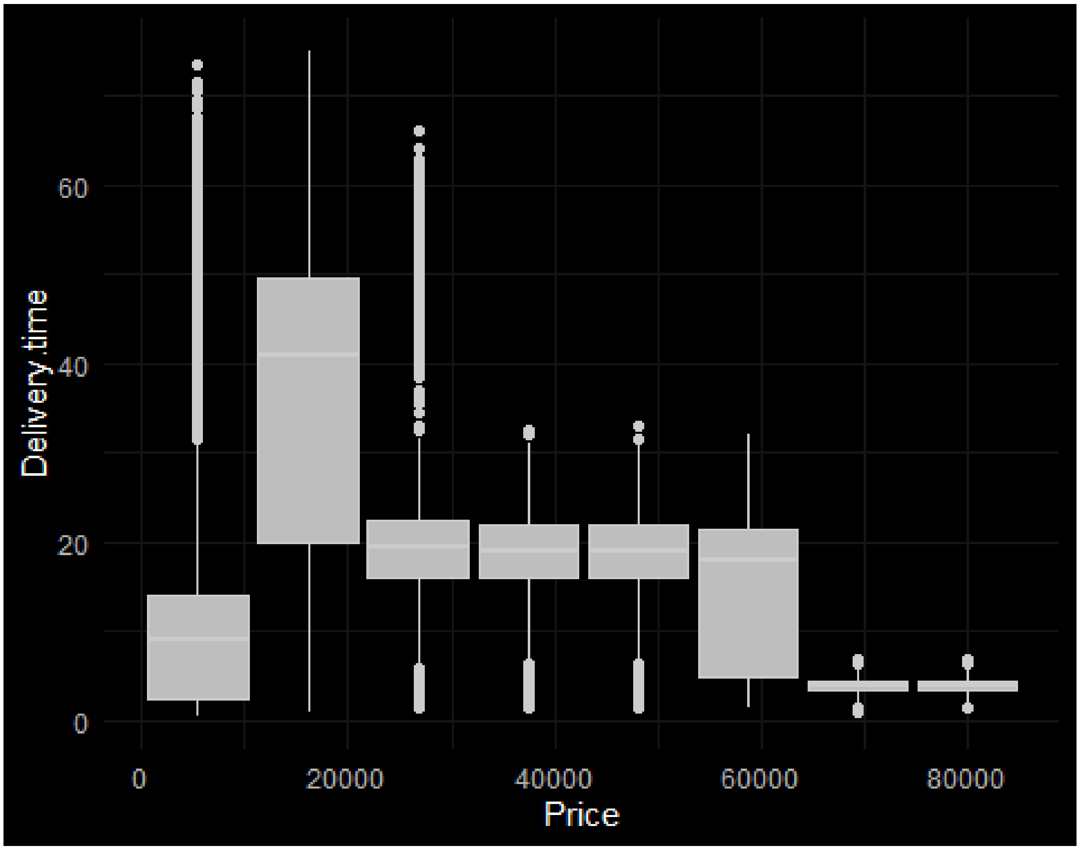


Figure 9: Boxplot of variation of delivey times per price

Comment: It's not easy to predict whether price of the product does in fact have an impact on delivery time. The highest value items tend to have the shortest delivery time. Products with a price range between 2000-6000 per item tends to have the same average delivery time. Thus, more accurate predictions can be made between certain price ranges. For prediction models price range will be grouped and delivery time can be predicted.

How did price change over time:

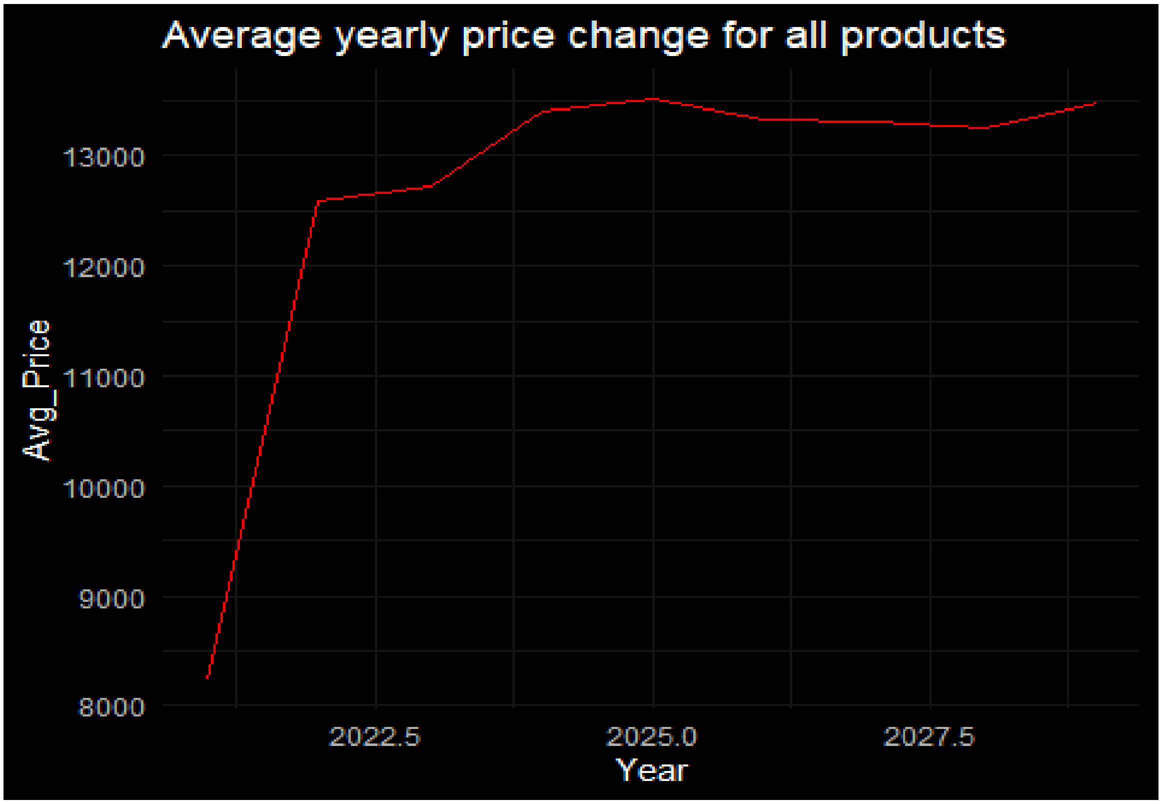


Figure 10: Average yearly price change for all products

Comment: Price significantly increased from 2021 to 2022

How did price per class vary over time:

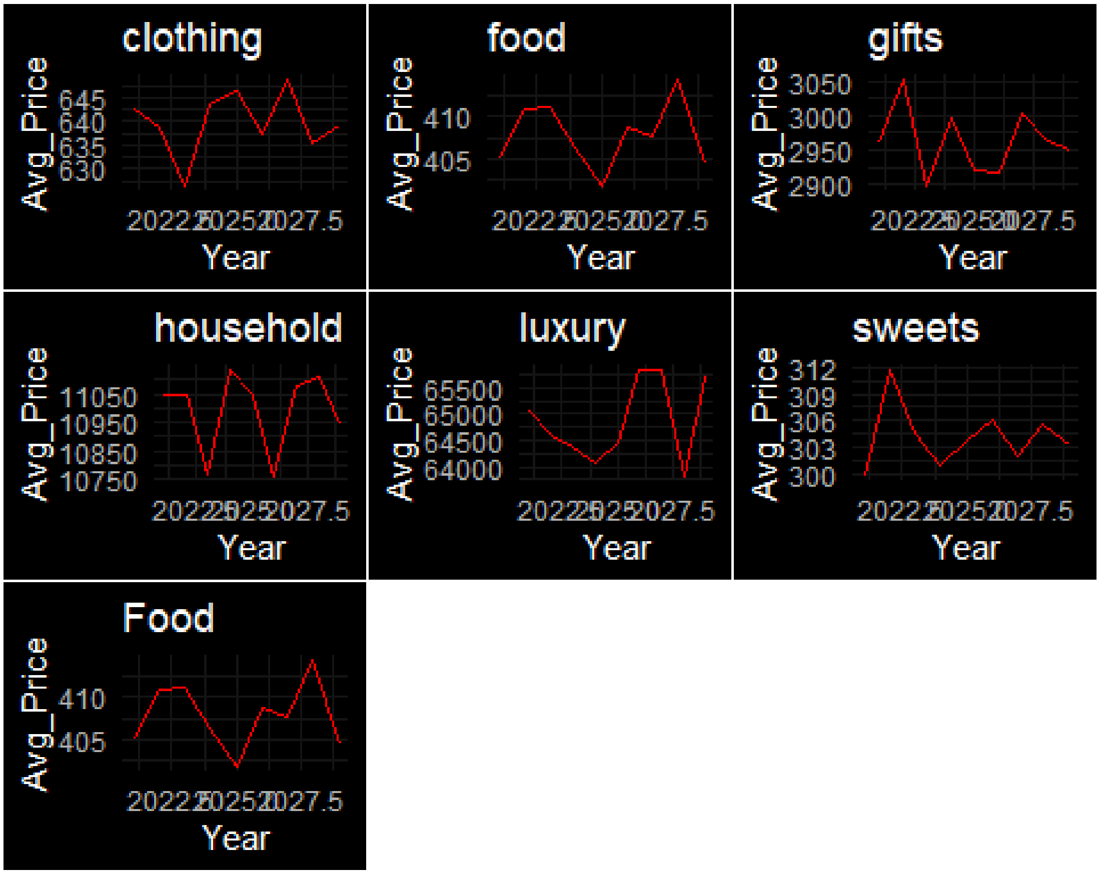


Figure 11: Average yearly price change per class

Comment: Time did not have a huge impact on the change in price for most items.

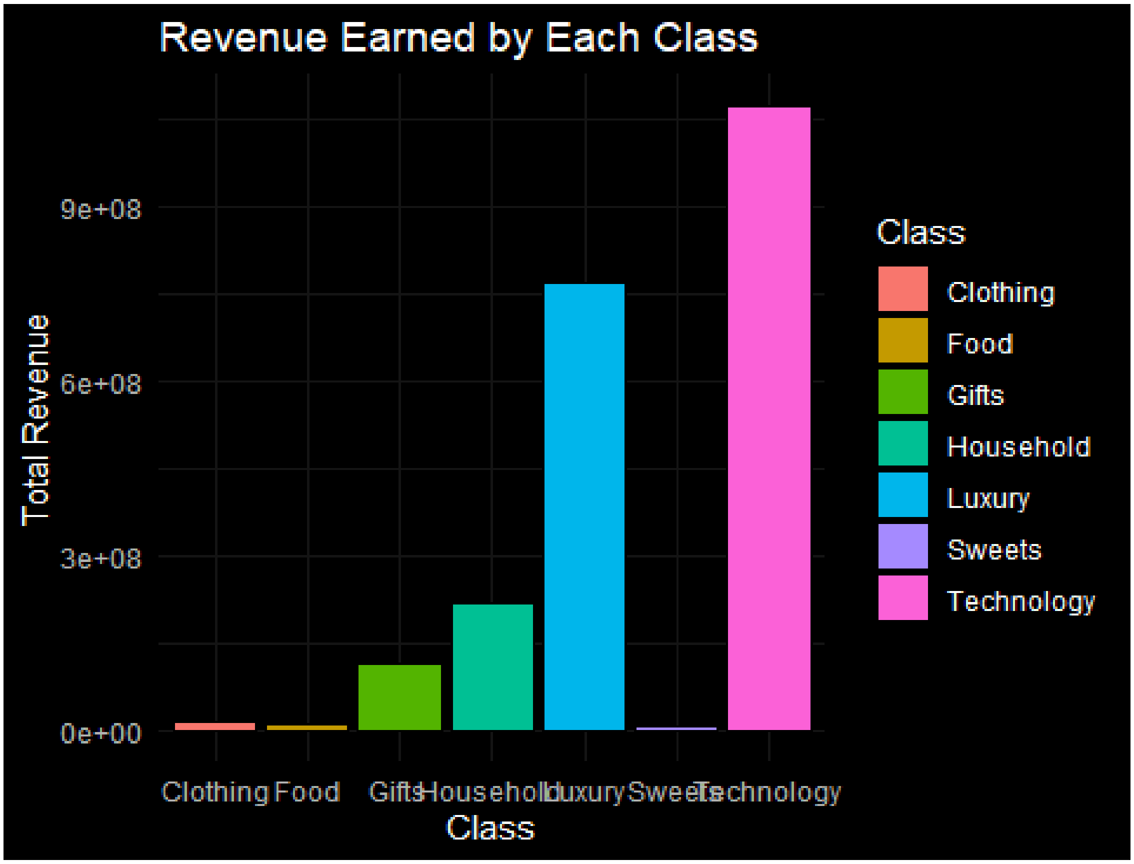


Figure 12: Revenue earned by each class

Comment: Food items made the most revenue. Clothing, food and sweets did not make nearly as much revenue. These items can be considered to be replaced in the future. 20% of all items sold was gifts, but gifts did not generate a large enough profit. Technology items contributes 20% of the total sales and generated the most profit. According to promotions principal 20% of product generated over 80% of the revenue, thus the company should focus on the technology items and make sure customers are satisfied the most for the class item.

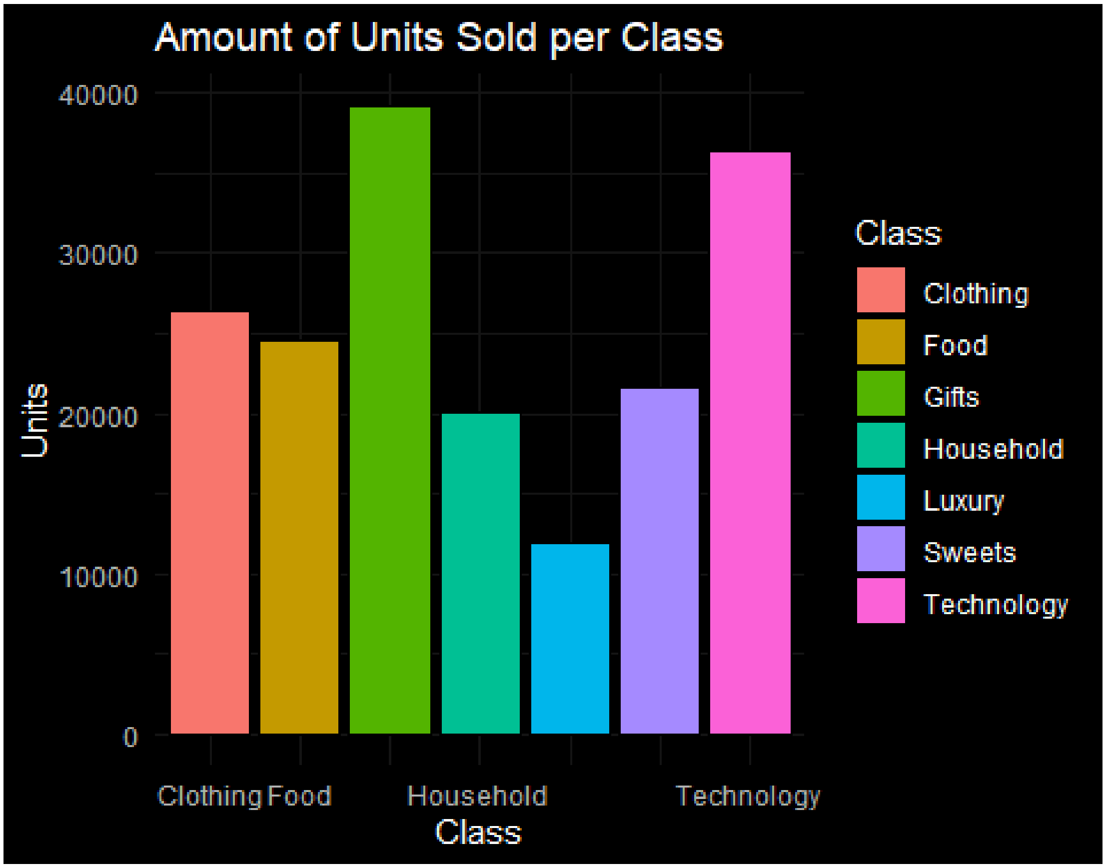


Figure 13: Amounts of units sold per class

Comment: Gifts was sold the most but did not make a large revenue. Where luxury items were sold the least but generated a large revenue.

2.2. Age impact on categorical features

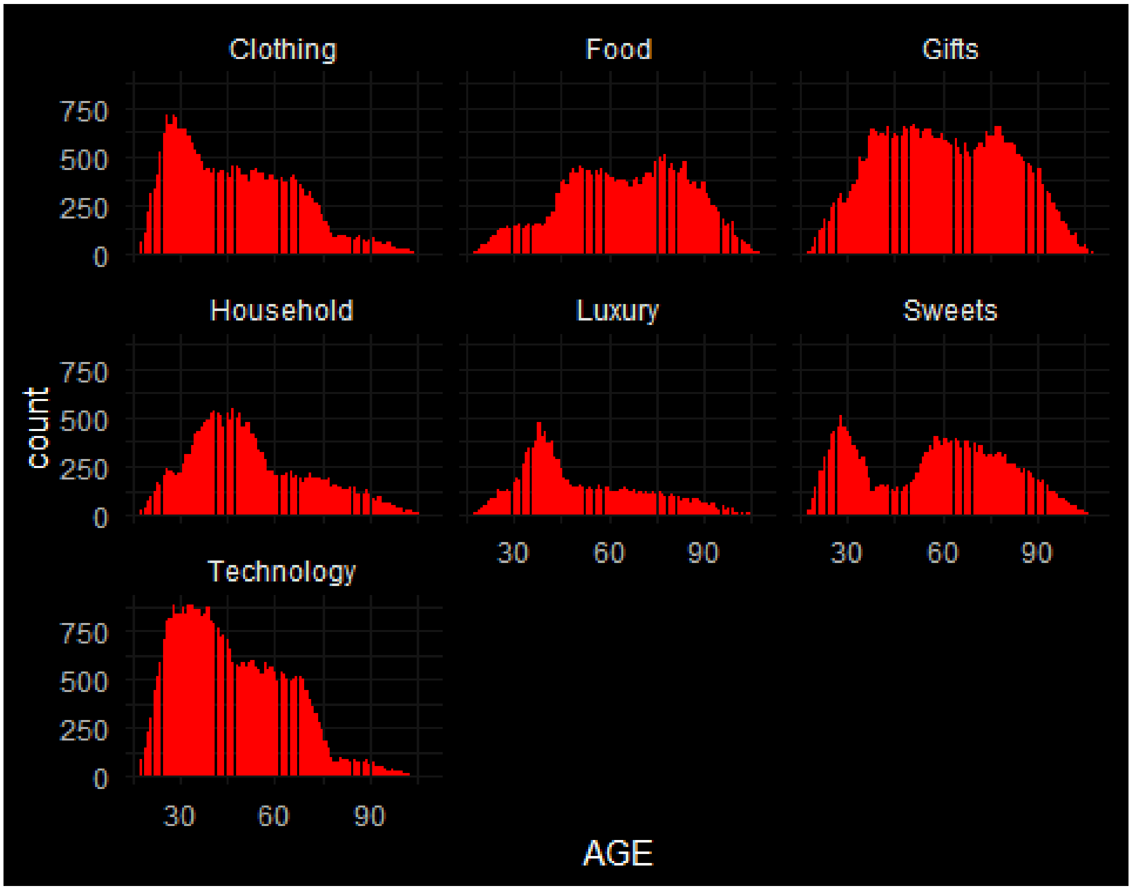


Figure 14: Age distribution per class

Comments: It can easily be spotted what type of class each age group buys the most. This is helpful to identify the target audience

Different distributions consist of the different age groups, Technology items are mostly bought from younger customers. We also see dips in distributions for example younger people and older people tends to buy sweets, but not adults of age between 40-50.

Age 40-70 shows that these group of customers tends to buy more gifts and food, suggesting them being parents and supporting children.

2.3. Delivery time impact on categorical features

Figure 15: Delivery time distribution per class

Comments on plot: Technology has a high delivery time and since technology is mostly sold the company should focus on this area to prevent customers from being dissatisfied. Technology also is also normally distributed. Household items also has the largest distribution tending to a high delivery time. Gifts has a normally distribution skewed to the left tending to longer lead teams on the mean. Food has many items being delivered on a low lead time. This could be because food can be ordered in large lots at the same time, also food has a low shelf life thus delivery time needs to be the lowest on average for food items.

Class comparisons

Finally viewing AGE, delivery time and price per item all at ounce. Visual conclusion can be made much easier. The strengths and weakness of each class can easily be spotted. A weakness could be too large delivery times. And strengths could be more balanced for price and age.

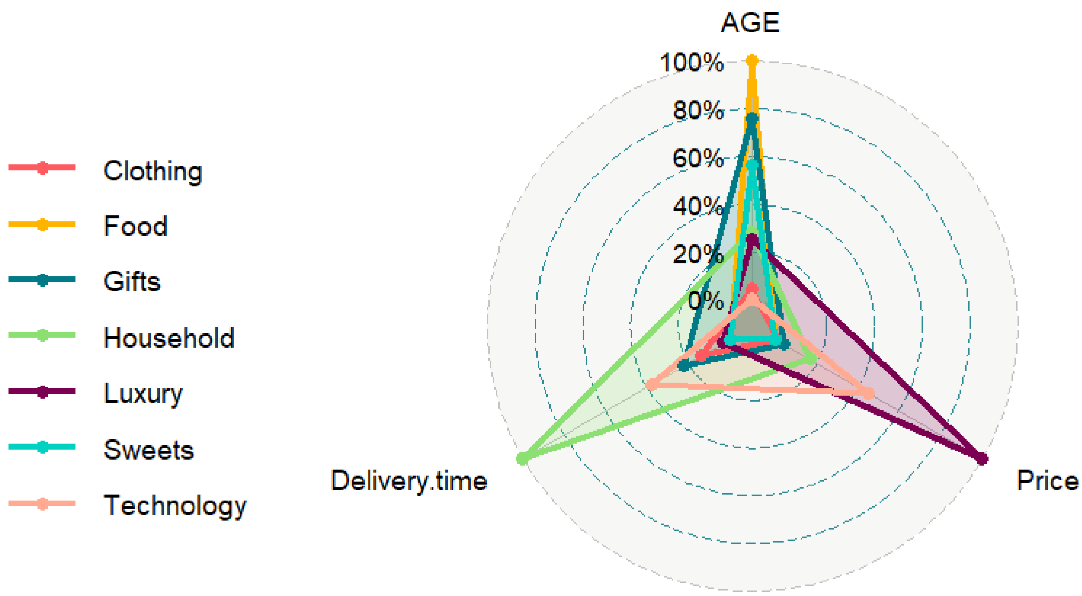


Figure 16 Radar graph of delivery time, Age, price per class

From the radar graph its visible that household items have the longest delivery time. Luxury items is the most expensive and that food is targeted by all age groups.

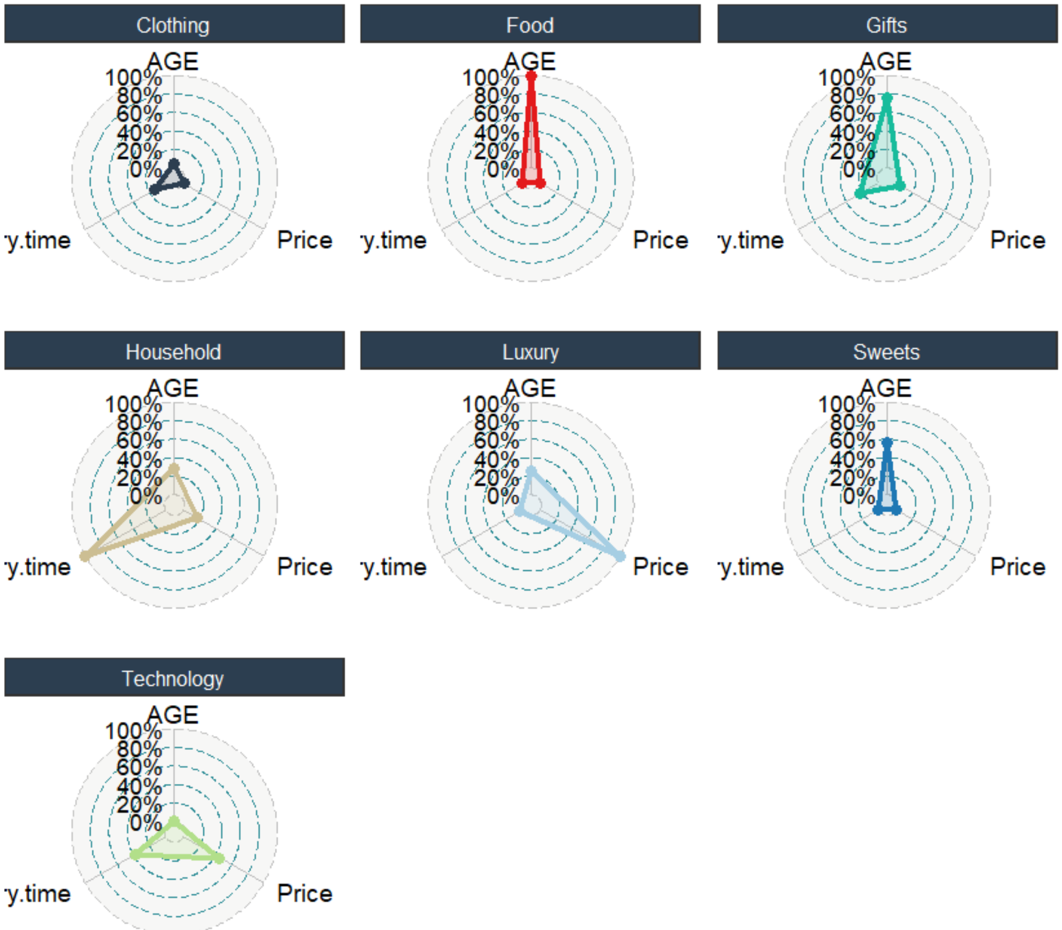


Figure 17 Facet radar graph

The following conclusions can be made:

Clothing: Clothing is relatively balanced as AGE, delivery time and price fall in the 20% range.

Food: Food has the highest age distribution of 100% (most) of all customers buys food.

Gifts: Gifts also has more strengths in being sold to 80% (second most) of customers.

Household: Household items has the longest delivery times.

Luxury: Luxury items has the highest price and the lowest delivery times. Meaning customers will be satisfied when buying luxury items.

Sweets: Sweets comes 3rd when selling to customers of all ages, with a very low delivery time.

Technology: Technology is all around the best distributed when it comes to balance AGE, price and delivery time. Technology however has the second largest delivery time.

Process Capability Analysis

Process capability analysis represents a significant component of the measure phase from the DMAIC (Define, Measure, Analysis, Improve, Control) cycle during a Six Sigma project. This analysis measures how a process performance fits the customer's requirements, which are translated into specification limits for the interesting characteristics of the product to be manufactured or produced. The results from this analysis may help identify variation within a process and develop further action plans that lead to better yield, lower variation and less number of defects.

The process capability analysis will be determined by:

The delivery time for all class technology items

Given: USL= 24 and LSL = 0

Delivery time can't be less than zero thus its logical that LSL could be zero

$$sigma = sd(Deliveytime)$$

$$Cp = (UL - LSL)/(6 * sigma)$$

$$Cpu = (USL - mean(deliverytime)/(3 * sigma)$$

$$Cpl = (mean(deliverytime) - LSL)/(3 * sigma)$$

$$Cpk = min(Cpl, Cpu)$$

```
Cp = 1.14220682164984

Cpu = 0.38

Cpl = 1.905

Cpk = 0.38
```

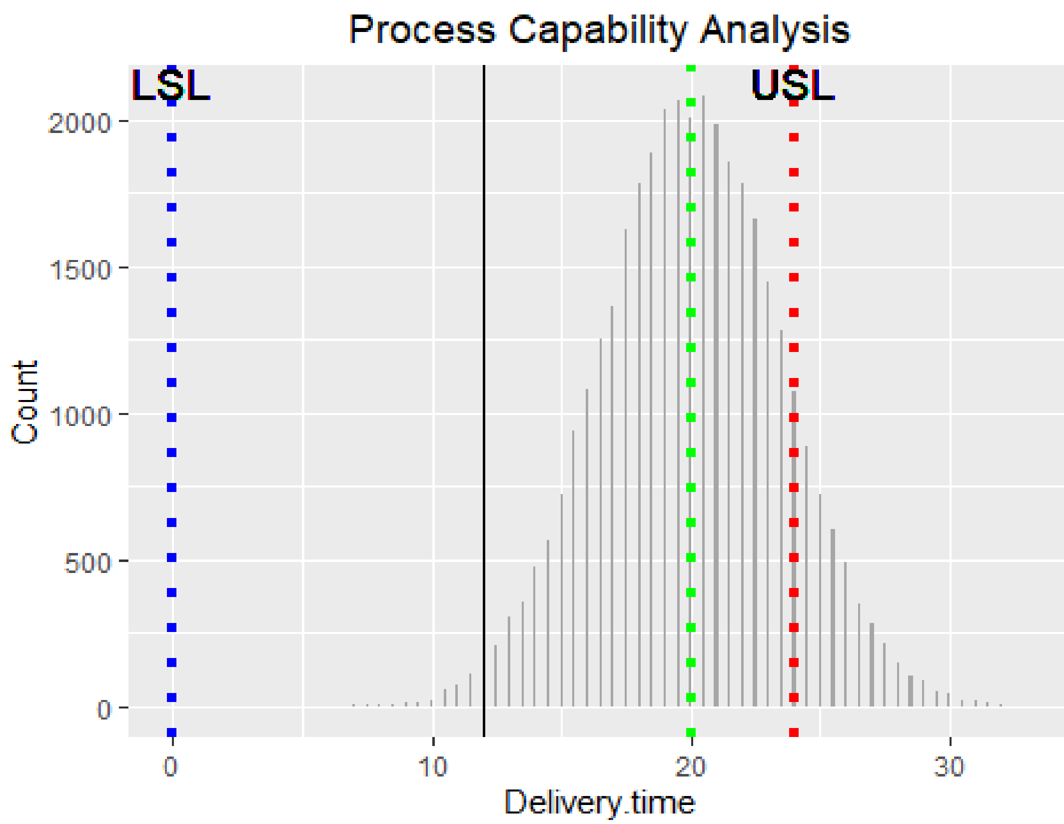


Figure 18: Process capability for technology delivery times

Comment: Cp>1 thus the process of delivery time is capable of meeting specifications because the process variation is smaller than the specification range. Cpk is smaller than Cp and it means that a part of the distribution exceeds the upper boundary limit.

Technology has one of the highest delivery times. A certain percentage of delivery time does not meet the requirements specified. To improve Cpk, we may either center the process on the target (by adjusting a dial or setting) or reduce the variation and spread of the process.

```
## 11.1426
```

11.1426% of technology items sold does not conform to the specifications and exceeds the upper boundary limit. This a high number and steps should be taken to improve on this.

3. Part 3

For this part statistical process control will be performed with the companies' data. The companies' processes will be visuals and improvements can be made by analyzing the statistical control charts. The statistical process will be done on the companies' delivery times. As seen from previous data visualization and statistical description that the delivery times should be broken up in different class groups as there means and limits may vary for delivery times.

From SPC we could determine whether a process is stable and if the process meets customer specifications. We can identify samples out of customer specification and implement root cause analysis on the problems.

Methodology for drawing up the tables for x-bar chart and s-chart:

The first 30 samples will be used to calculate the values in the tables. Afterwards these calculated values will be used in the process control charts to draw up the samples between the ranges calculated from the first 30 samples.

Table containing information about x-chart

The following table will contain the needed information to calculate the xbar-charts

Class	UCL	U2Sigma	U1Sigma	CL	L1Sigma	L2Sigma	LCL
Technology	22.975	22.108	21.241	20.374	19.508	18.641	17.774
Clothing	9.405	9.260	9.115	8.970	8.825	8.680	8.535
Household	50.248	49.020	47.791	46.562	45.334	44.105	42.876
Luxury	5.494	5.241	4.988	4.736	4.483	4.230	3.977
Food	2.709	2.636	2.563	2.490	2.417	2.344	2.271
Gifts	9.489	9.113	8.737	8.361	7.985	7.609	7.234
Sweets	2.897	2.757	2.618	2.478	2.338	2.198	2.059

Table 4: x-chart control limits

Table containing information about s-chart

The following table will require the needed information to calculate the s-charts

Class	UCL	U2Sigma	U1Sigma	CL	L1Sigma	L2Sigma	LCL
Technology	5.181	7.378	6.279	3.296	0.312	-0.787	1.410
Clothing	0.867	1.234	1.050	0.551	0.052	-0.132	0.236
Household	7.344	10.459	8.901	4.672	0.442	-1.115	2.000
Luxury	1.511	2.152	1.831	0.961	0.091	-0.229	0.411
Food	0.437	0.623	0.530	0.278	0.026	-0.066	0.119
Gifts	2.246	3.199	2.723	1.429	0.135	-0.341	0.612
Sweets	0.835	1.190	1.012	0.531	0.050	-0.127	0.227

Table 5: s-chart control limits

Statistical process control charts

Statistical process control charts will be plotted and visually inspected whether the delivery times is stable or unstable.

3.1.1. Clothing SPC chart

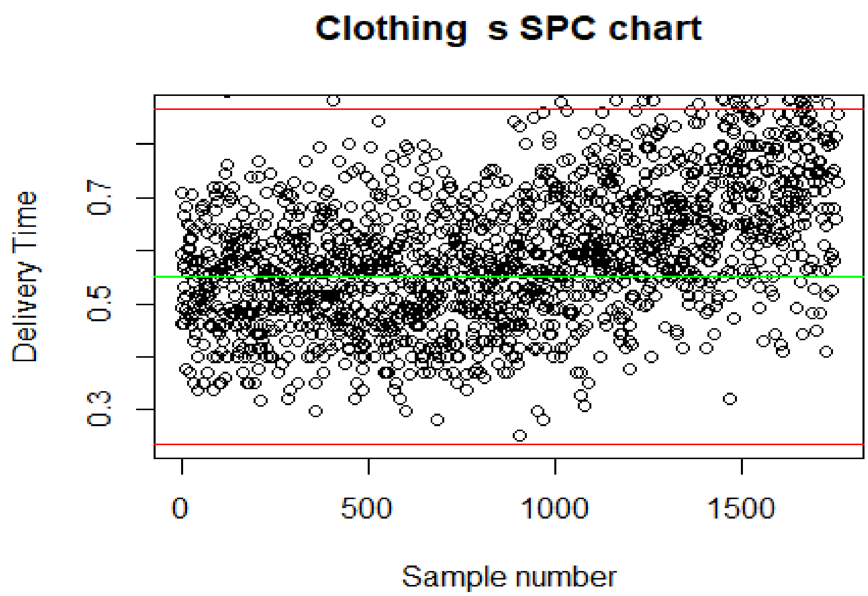


Figure 19: clothing s-chart

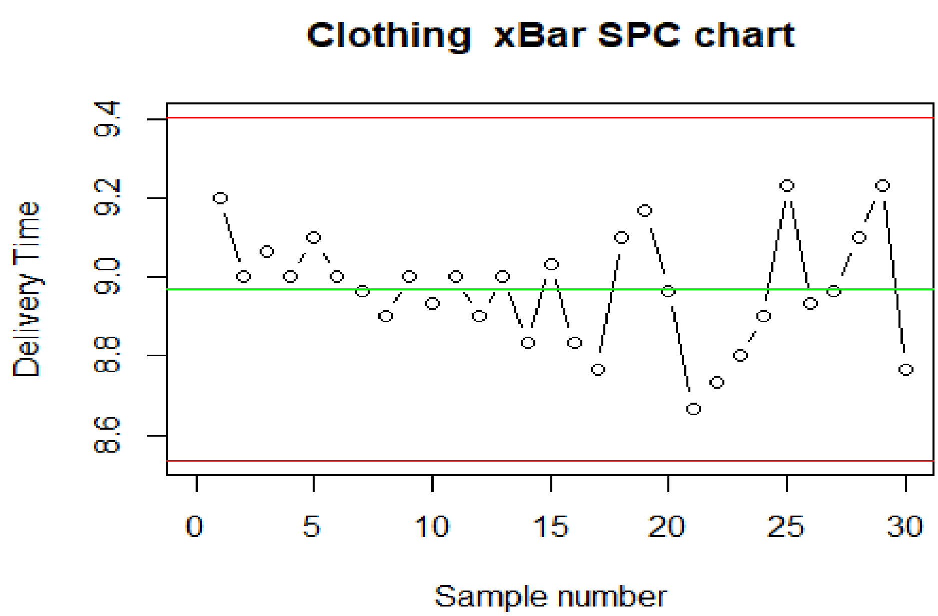


Figure 20: clothing xbar-chart 30 samples

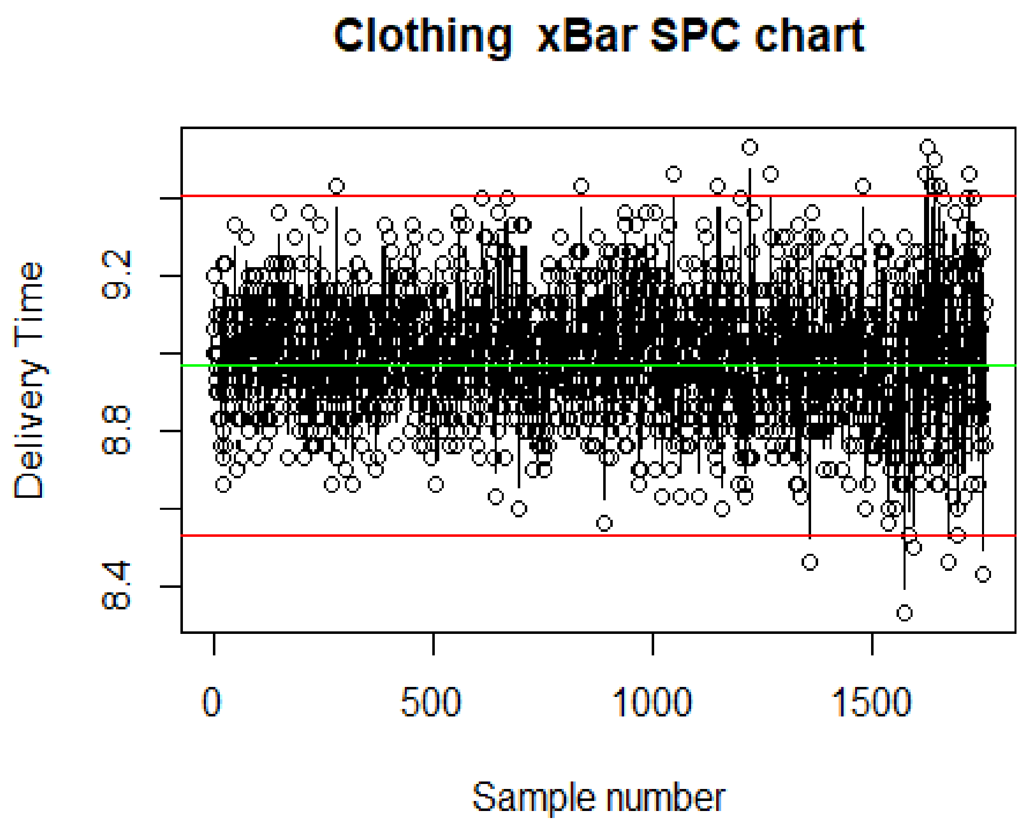


Figure 21: clothing xbar-chart all samples

Examining the s-chart, x-bar chart for the clothing items reveals the following characteristics: the points appear to fall arbitrarily above and below the middle line. The majority of points are located along the middle line. There are no samples exceeding the control limits consecutively. There are the same number of points above and below the center line. Therefore, it is apparent that clothing delivery times are under control. However, as time progresses, a growing number of samples are outside the control limits. Consequently, a modest change in the process could cause the impacts to grow over time. Figure 18 demonstrates that the standard deviation is growing with time.

3.1.2. Household SPC chart

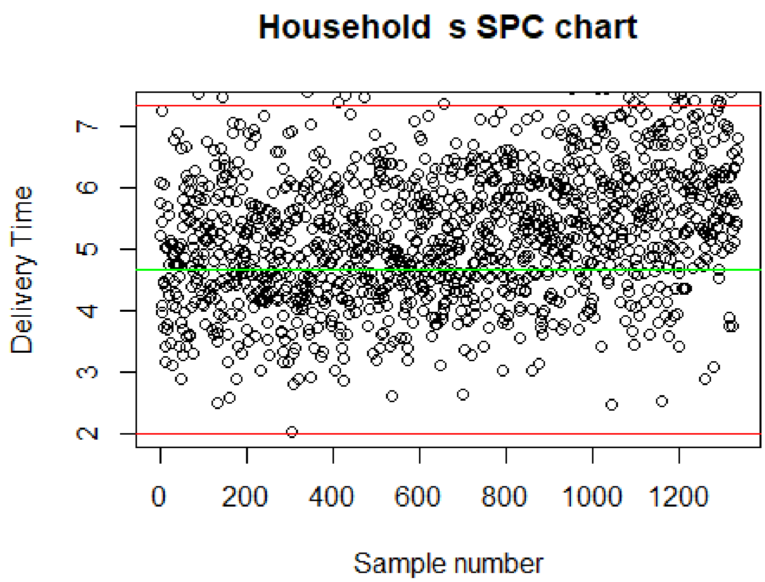


Figure 22: Household s-chart

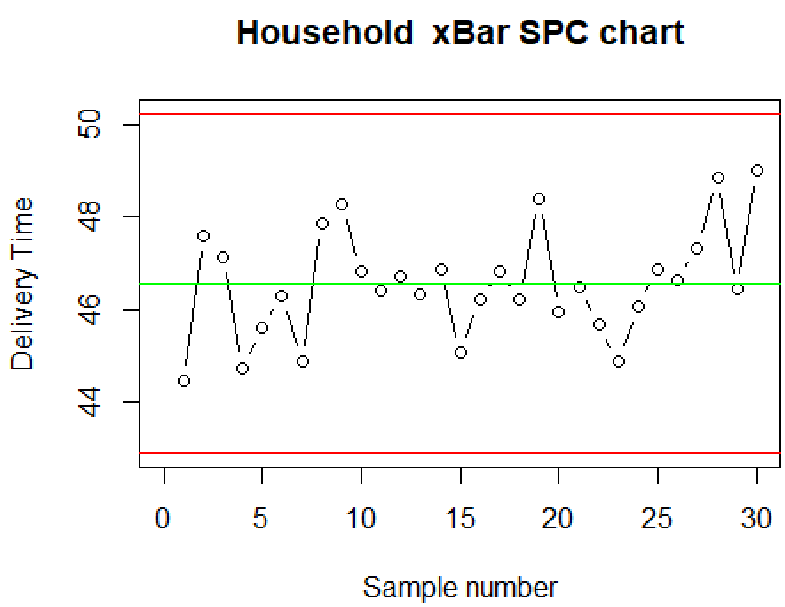


Figure 23: Household xbar-chart 30 samples

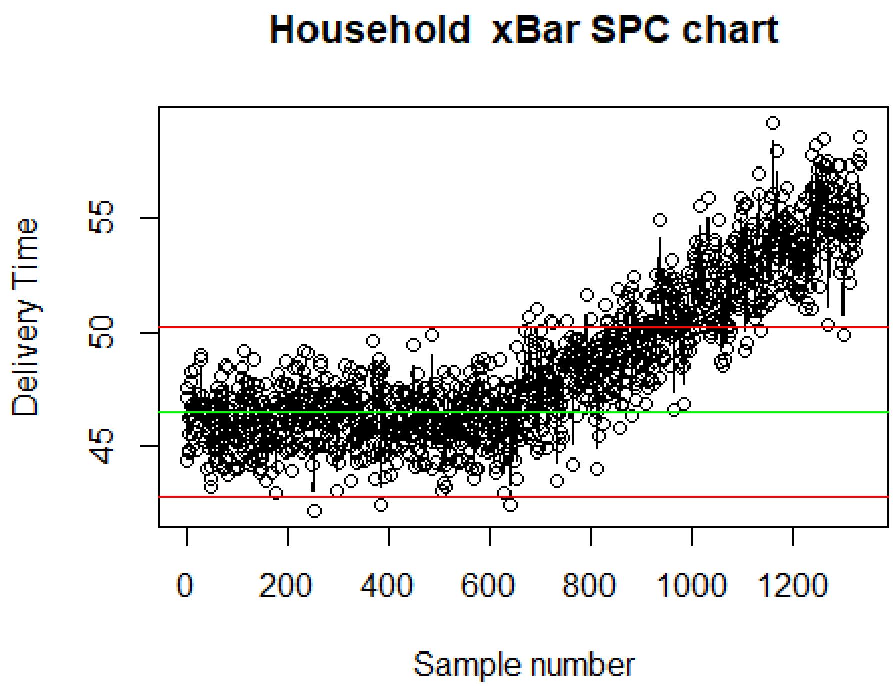


Figure 24: Household xbar-chart all samples

Figure 21 indicates that the process appears stable. Nonetheless, as time passes, it is evident that the standard deviation from figure 20 is also growing, but not by a significant amount. Up to the first 600 samples, delivery times for home goods stay stable and under control, as there are few samples that fall outside the control boundaries. Beginning with the 600th sample, delivery timeframes are becoming increasingly out of control. There is an abrupt increase trend in the lengthening of delivery times. This could be an indication that supply chain marketplaces are experiencing a disruption, resulting in products becoming scarcer, or that providers are gradually becoming less dependable. When examining the first 30 samples, it is clear that the process average has shifted, since an excessive number of consecutive samples fall between the upper control limit and the center line.

3.1.3. Food SPC chart

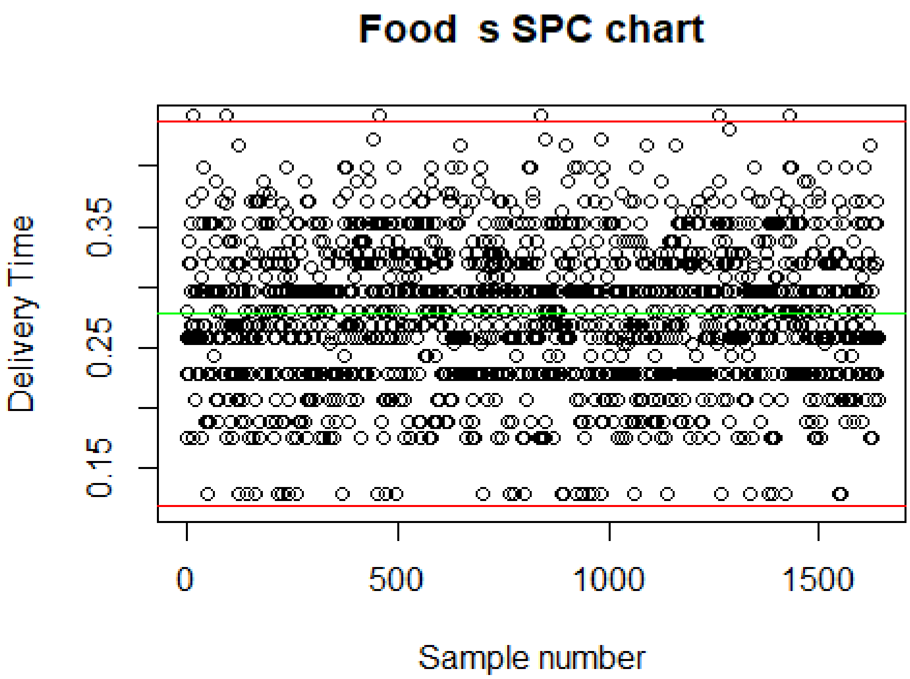


Figure 25 Food s-chart

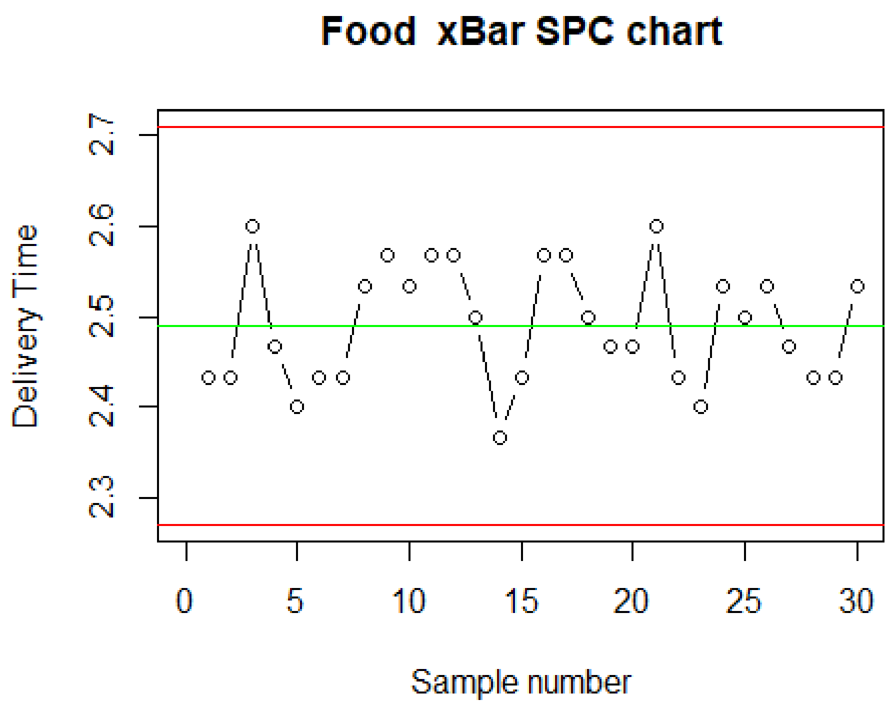


Figure 26 Food xBar-chart 30 samples

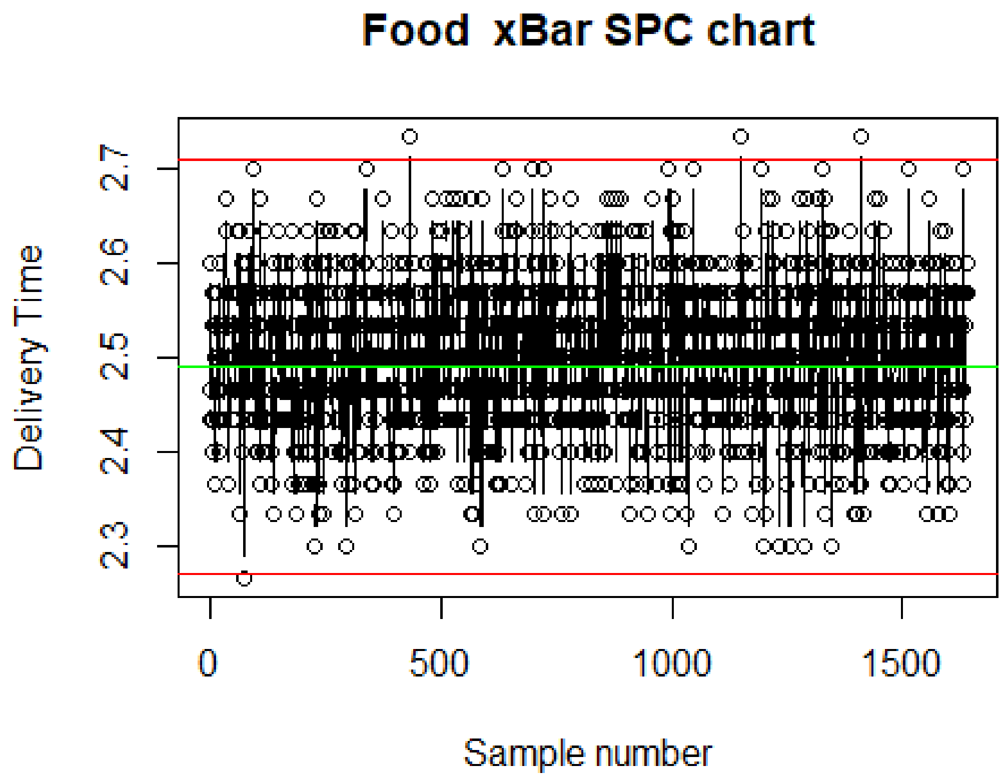


Figure 27 Food xbar-chart all samples

Examining the s-chart and x-bar chart for Food products will reveal the following characteristics: The points appear to be distributed arbitrarily above and below the central line. The majority of points are located along the middle line. There are no samples exceeding the control limits consecutively. There are the same number of points above and below the center line. Thus, it is apparent that food delivery times are under control. Examining figure 24 it is evident that the graph is hugging the center line. This indicates that the control limits are to wide. The distribution of the standard deviations from figure 23 has lines in the distributions with values not occurring at a certain delivery time. This could explains the large number of distributions at a certain delivery time for all samples in the xbar- chart.

The reason for this close hugging to the center line is that food items can spoil and its critical for delivery times to be between the control limits.

3.1.4. Technology SPC chart

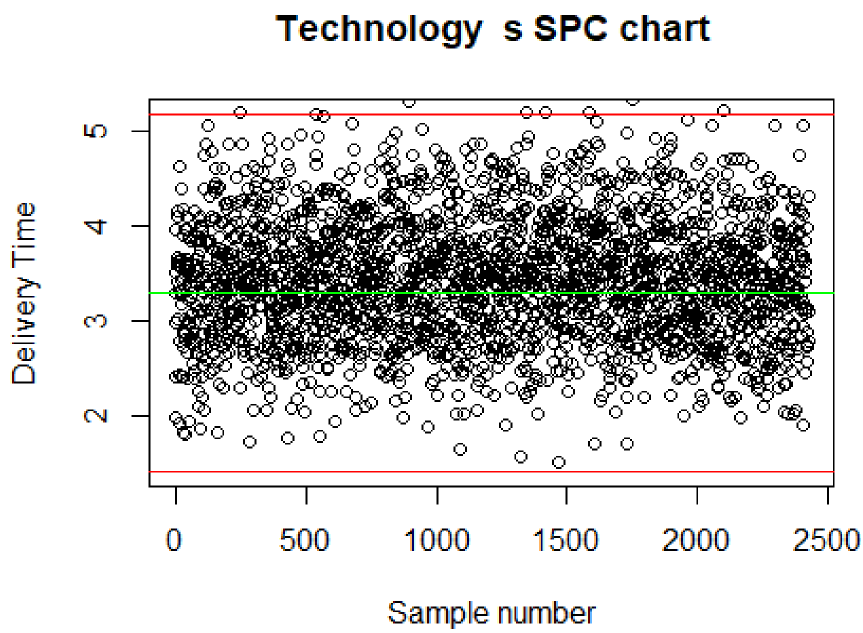


Figure 28: Technology s-chart

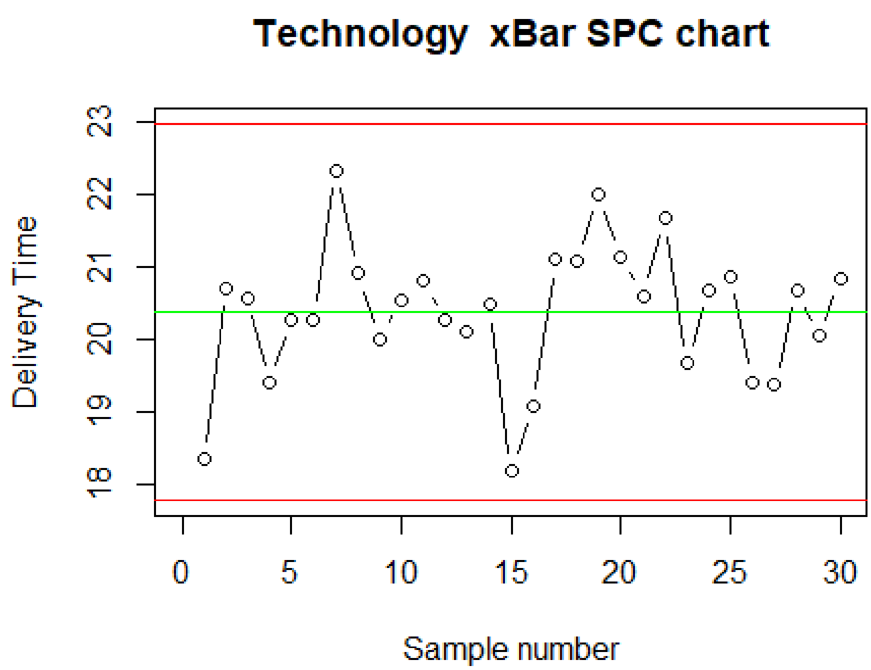


Figure 29: Technology xBar-chart 30 samples

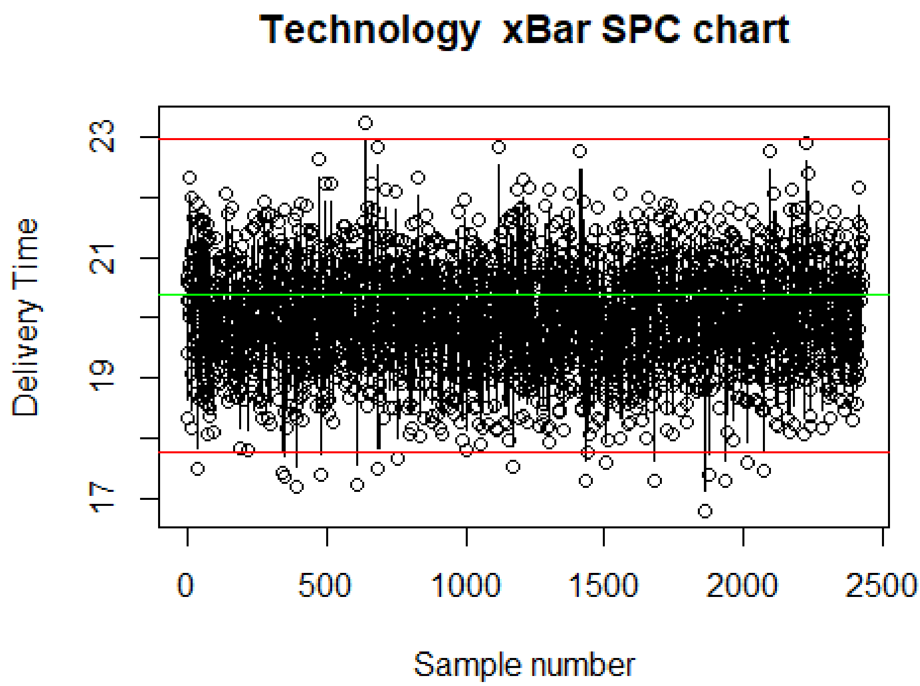


Figure 30: Technology xbar-chart all samples

Examining closely figure 28 it is possible to see a cycle in the samples over time. Cycles are short repeating patterns in the chart, alternating between high and low peaks. These results are due to certain causes that come and go on regulare basis. Reasons could be: the demand peaks differs and when demand risines , more products needs to be delivered and thus it can take on average slower to reach customers. There is however few samples outsiode the control limits and it occurs randomly. In the perspective of the samples meeting the control limits, the delivery times for technology can be seen as stable.

3.1.5. Sweets SPC charts

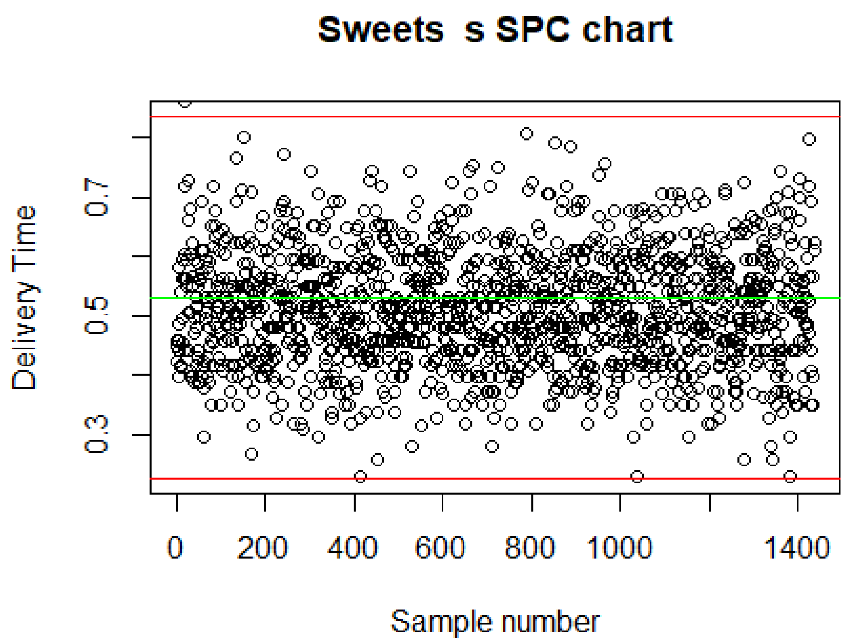


Figure 31: Sweets s-chart

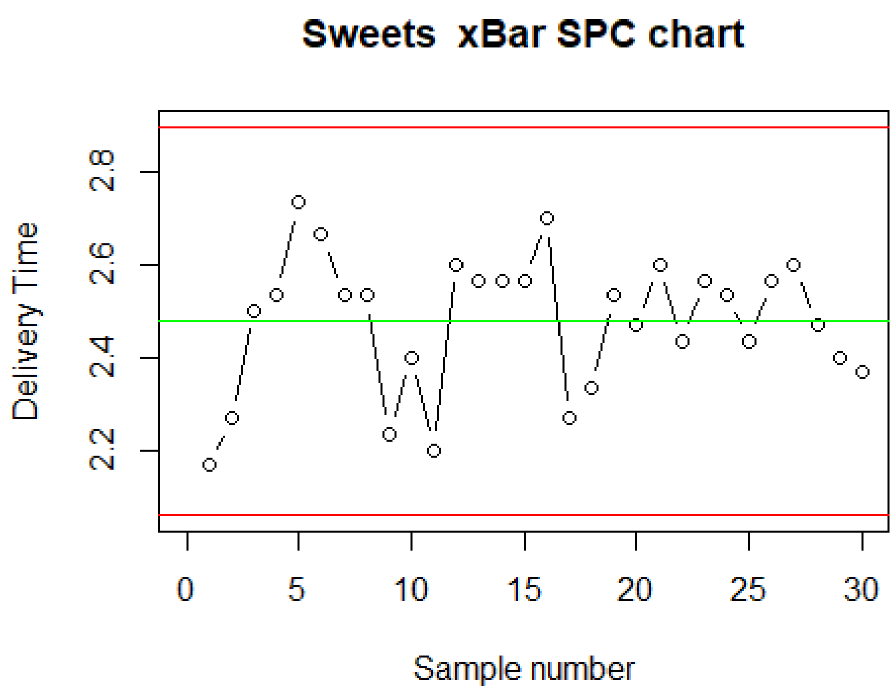


Figure 32: Sweets xBar-chart 30 samples

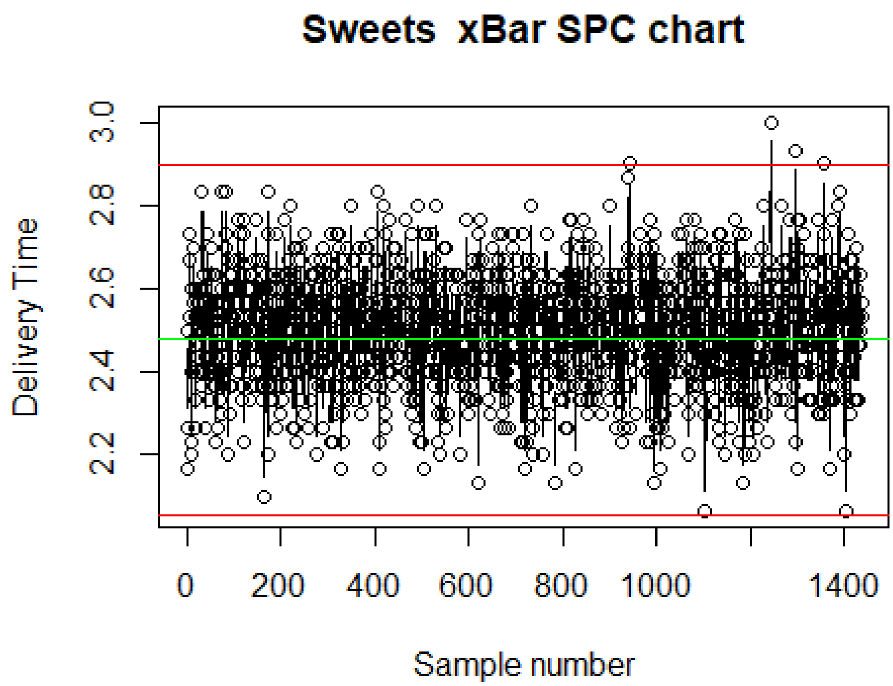


Figure 33: Sweets xBar-chart all samples

Sweets delivery times has almost no points outside the control limits, the number of points above and below the center line is the same, only a few points fall randomly outside the control limits, most points are close to the center line. Thus, sweets delivery time can be seen as a stable process. Sweets has the least number of random samples outside the control limits from figure 31. The UCL and LCL could be moved a bit closer to the center (mean?). This will greatly improve on the delivery times the company promises the customers and by promising more stable delivery times, the company could charge more for their sweets products.

3.1.6. Gifts SPC chart

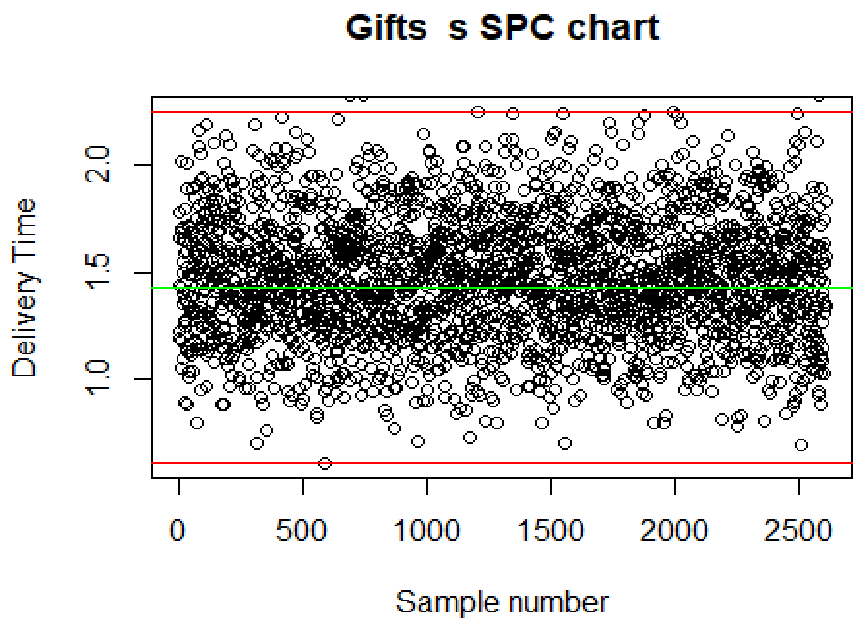


Figure 34: Gifts s-chart

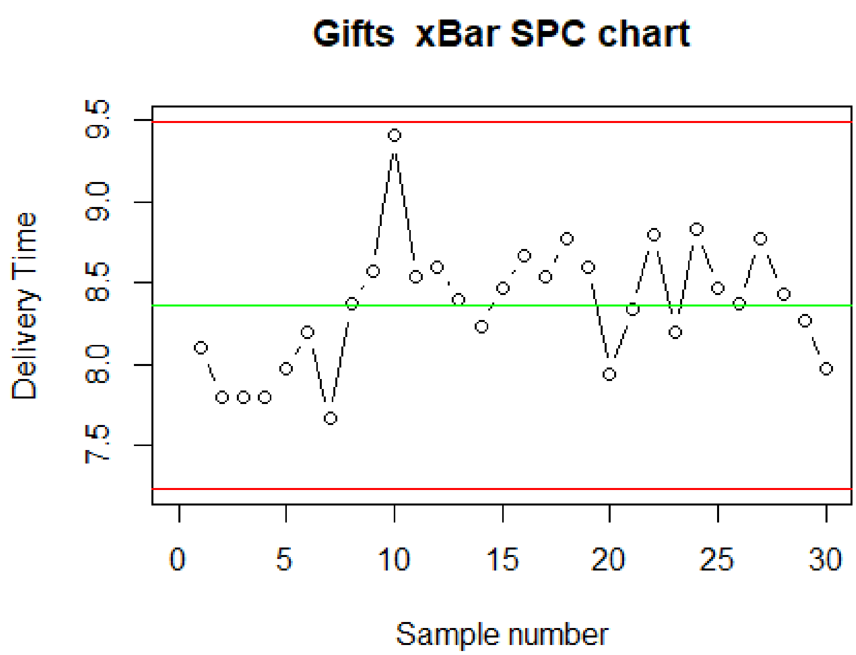


Figure 35: Gifts xBar-chart 30 samples

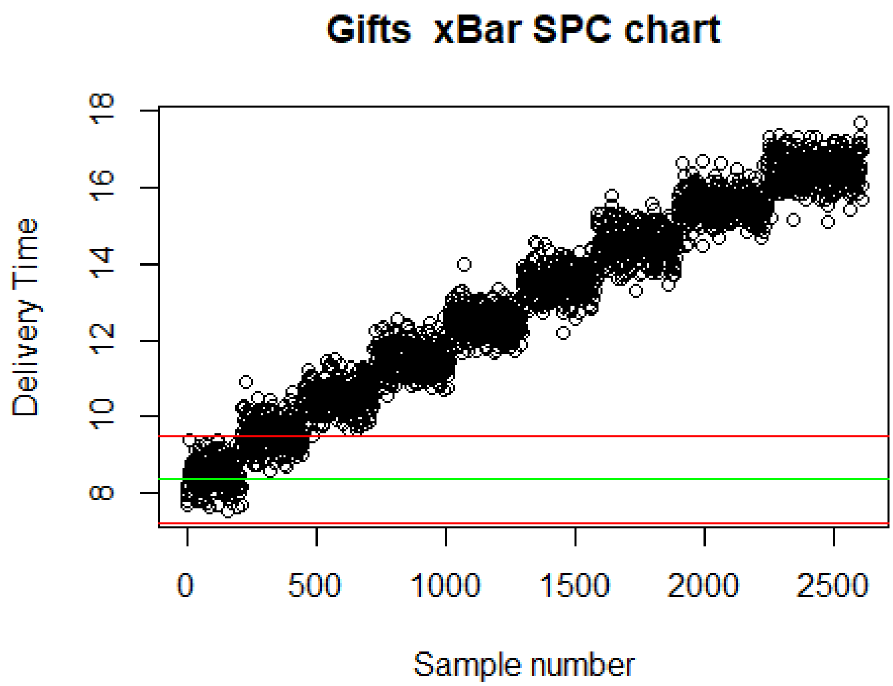


Figure 36: Gifts xBar-chart all samples

Examining Figure 32 and figure 33. The sample deviations remain in the control limits over time and the first 30 samples is relatively stable between the control limits. However, from Figure 34 there is an upward trend in gifts delivery times. Samples seems to be grouped over a certain time interval. These samples seem to have the same height of lower to higher delivery times, thus this explains why the standard deviations remain the same and the s-chart did not display or portray any problems nor deviations. The samples are increasingly jumping per chunk over time. There is an increase from 8 days delivery time anticipated to almost 18 days delivery times received. The delivery times is out of control and increasing in an upward trend. This could be a result of the company making large changes to improve delivery times, but instead worsen the effect every time the change is made. The sudden upward jumps give evidence that there was a sudden change made.

3.1.7. Luxury SPC chart

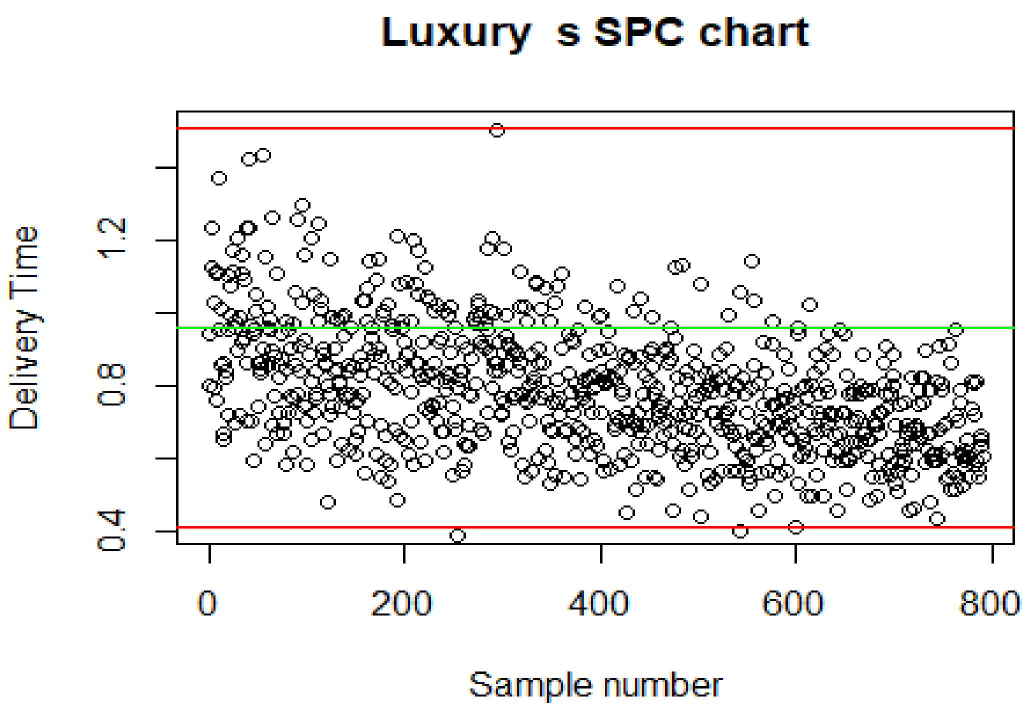


Figure 37: Luxury s-chart

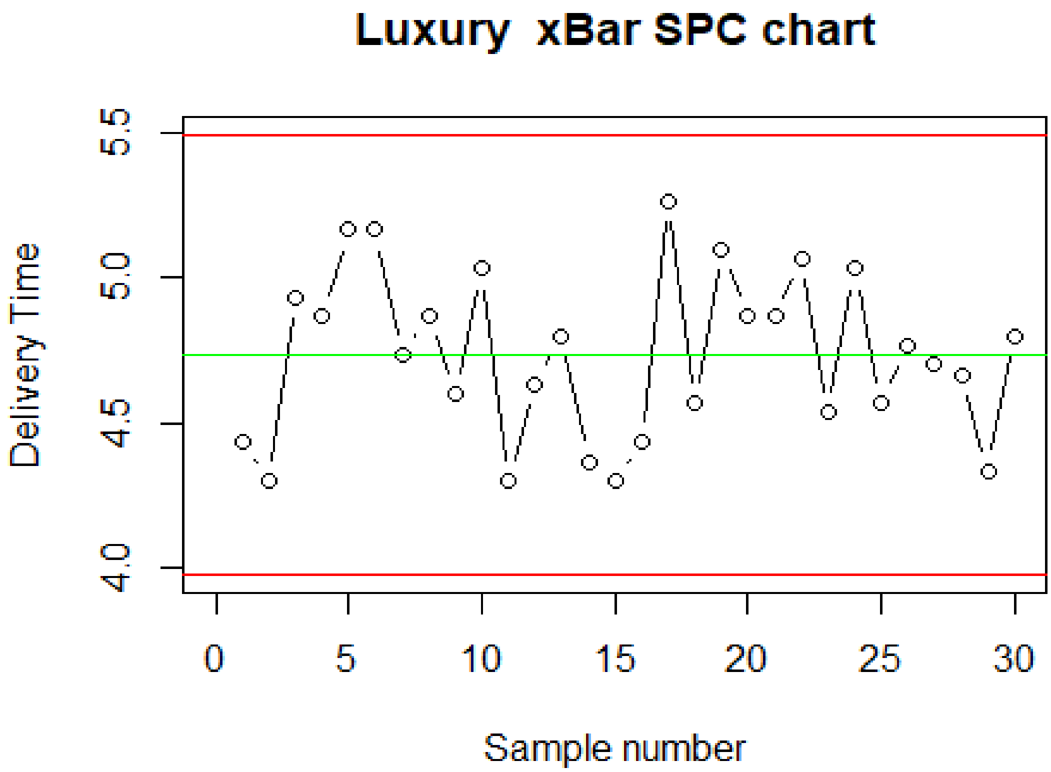


Figure 38: Luxury xBar-chart 30 samples

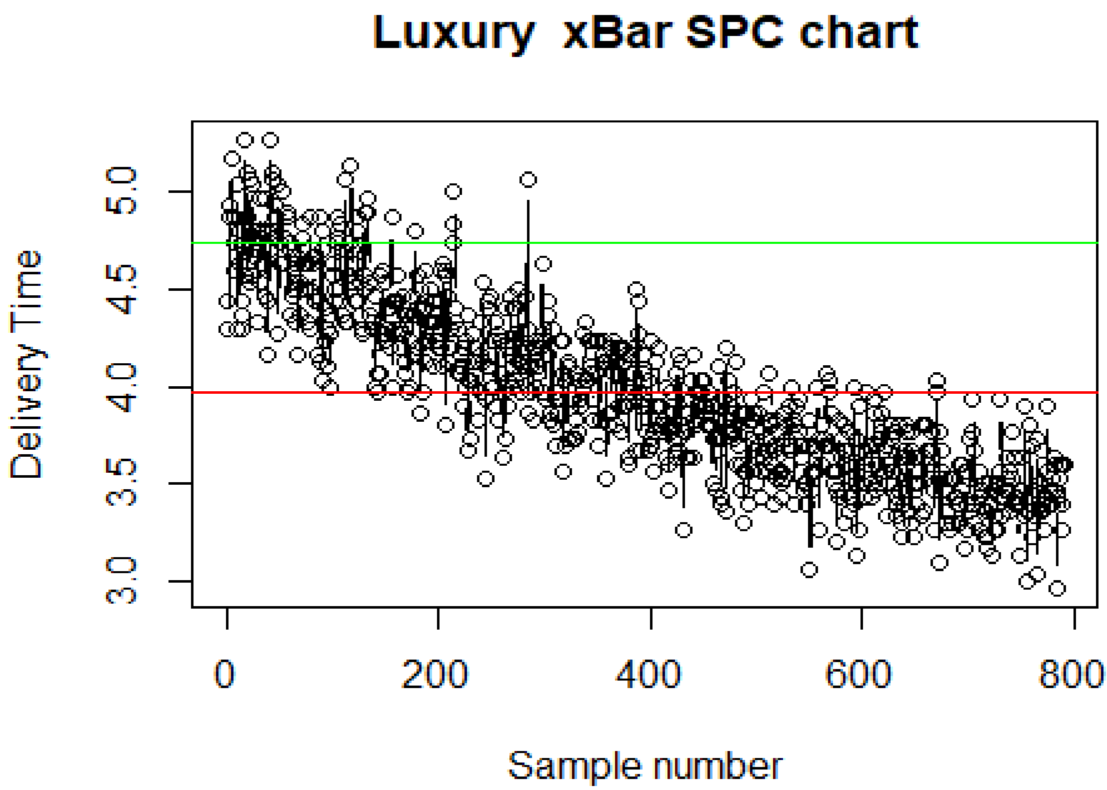


Figure 39: Luxury xBar-chart all samples

From figure 36 it is evident that the graph is depicting the the fisrt 30 samples. With a short delivery time ptomisedAs time progresses the delivery time samples decreases over time as seen in figure 37. The reason for this could be that a dial setting or a change was made, since the standerdeviations for delivery times started decreasing overtime. This is not neccerly a negative impact on the bussiness but in the context of delivery time to decrease the delivey times over a time period while the company does not promote or promise its customers lower delivey times. This means there could potetially be the commencemen and decrease in the UCL over time aswell as the LCL. There could however be costimplication when decreasing delivery times to much, thus at some point the company should start reaching an equilibrium and commence production of a greater delivery time.

4. Part 4: Optimizing the delivery processes

The aim will be to optimize the delivery time processes. The first step is to identify how many samples is out of the specifications. This will give an indication of whether the process specifications should be adjusted. Too few samples out of the boundary limits could mean there is no problem with the process currently and that these random occurrences don't happen very often.

On the other hand, if too many samples occur out of the boundary limits, this could mean the process should be inspected and corrective action should be taken in place

A - Out of control samples from X-bar chart

Examining where samples occur outside of the boundary limits could give an indication of out-of-control processes. From these analysis in the attempt to identify these samples occur consecutively. If samples occur consecutively more than 7 times above or below the control limits the process is unstable according to R.Evans and M.Lindsay,2014.

4.1.1. Non-conforming lower limit table

Lowerb_class	1st	2nd	3rd	3rd-Last	2nd-Last	Last	Found
Clothing	1,359	1,574	1,587	1,677	1,695	1,756	7
Food	75						1
Gifts							0
Household	252	387	643				3
Luxury	142	171	184	789	790	791	440
Sweets							0
Technology	37	345	353	1,933	2,009	2,071	18

Table 6: Lower limit non conforming samples from Xbar-chart

For Sweets and Gifts there is no samples below the lower limits. This could either mean they are stable or that the process is unstable and increasing over time at the upper bound.

Most of clothing out of boundaries for lower limits occurs with high sample numbers. Analysis indicates that over time there is an increasing in unitability.

4.1.2. Non-conforming upper limit table

Uperb_Class	1st	2nd	3rd	3rd-Last	2nd-Last	Last	Found
Clothing	282	837	1,048		282	1,723	13
Food	432	1,149	1,408				3
Gifts	213	216	218		213	2,609	2,287
Household	679	693	720		679	1,337	392
Luxury							0
Sweets	942	1,243	1,294			1,358	4
Technology	643						1

Table 7: Uper limit non conforming samples from Xbar-chart

Table 7 depicts:

Here we can see gifts delivery times is unstable and is increasing over time. Luxury items delivery times seems to increase over time.

4.1. B

For this part knowing where the most consecutive samples of the s-bar chart or sample deviations between -0.3 and 0.4 sigma .

Method: A function was written to loop trough the delivery times per class. Counting the consecutive numbers between the range specified.

Class	Number consecutive	Last index
Technology	6	1776
Clothing	4	223
Household	3	45
Luxury	4	63
Food	5	756
Gifts	7	2477
sweets	4	94

Table 8: Consecutive samples between range -0.3 and 0.4 sigma

4.2. Type I error

For a type I error we will determine what is the probability of the manufacturer or in this case delivery planner for the company's sales to only make an error. An error could be made in an attempt to adjust the delivery times to improve on the delivery times, but instead unknowing increasing the chances of errors.

typ1 <- (1 - pnorm(3))*2

prob1 <- typ1*100

prob1

```
## [1] 0.2699796
```

4.3. Process optimization of technology delivery times

As previously observed in the process capability analysis for technology, it is obvious that the process exceeds the defined upper limit. This out-of-control process could be improved by lowering the data's mean and cutting all delivery times by one hour per iteration. However, there are costs associated with altering the mean delivery time.

Nevertheless, according to the Taguchi loss function, maintaining systems that exceed the upper limits will increase expenses. Moving the samples between the control limits will thereby reduce the Taguchi loss function's related cost.

The delivery time process will be optimized to locate the sweet spot between the cost of reducing delivery times and the cost of being outside the regulatory limits.

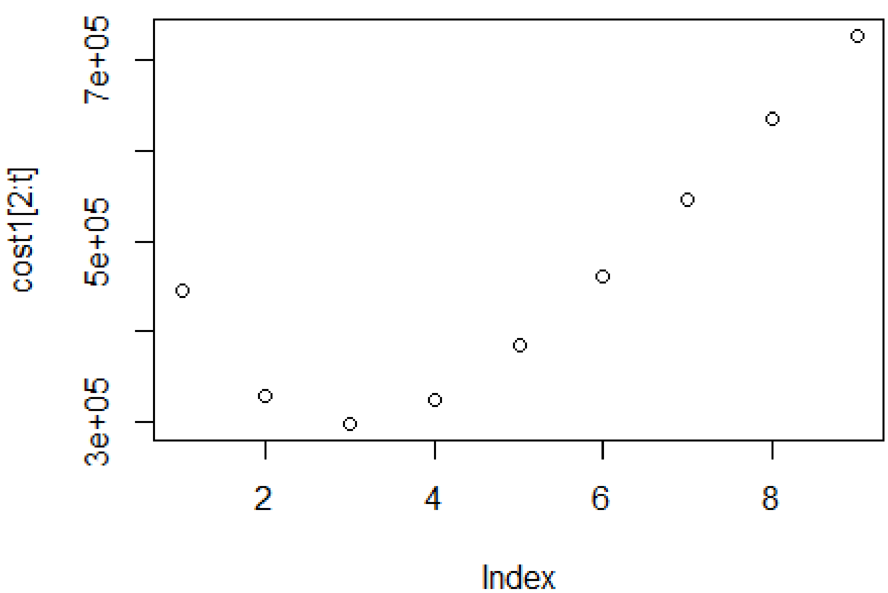


Figure 40: Cost per mean delivery time decrease

```
## [1] 298201
```

The lowest cost found is 298201. For technology items average delivery time per items will be reduced by 3 hours. Shifting the delivery times distribution to the left by 3 hours.

Visual inspection resulted in the plotting of data forming a normal graph to see the move of the process capability plot changed.

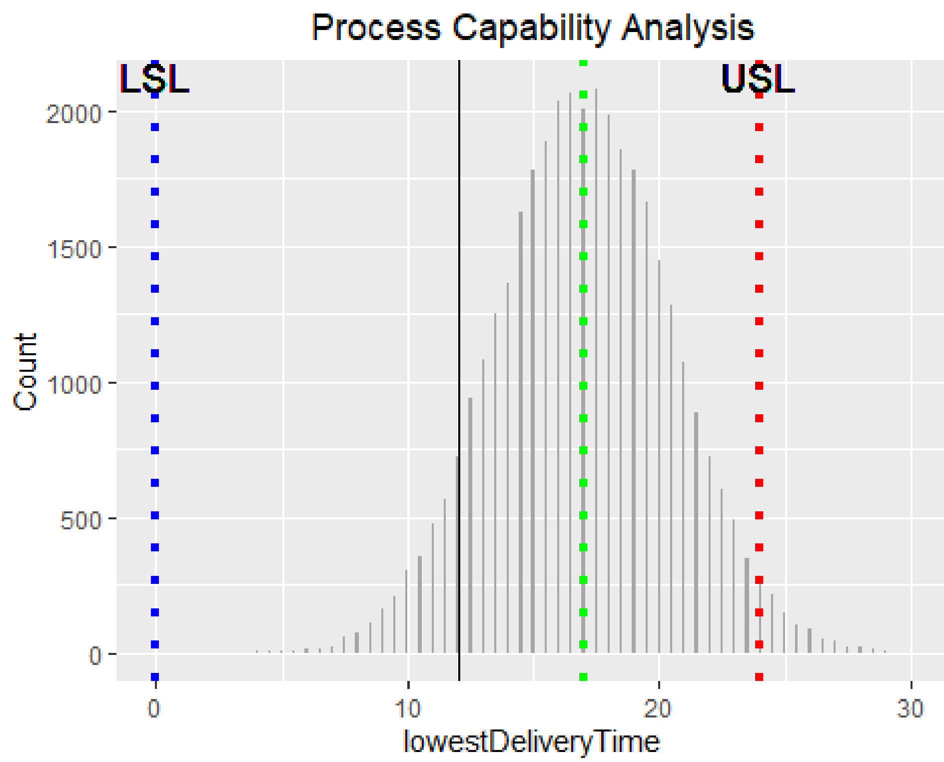


Figure 41: New distrubution of technolgy items delivery times

4.4 - Type II error

When a process is out of control the following could be a reason: The process variation became too large, or the process is outside of the limits.

In the type II error, the Ha is true, but we fail to identify this, due to the sample Xbar value being between LCL and UCL. The type II error will be calculated:

$$UCL = 22.974616$$

$$LCL = 17.774$$

$$sd = (UCL - LCL)/6$$

$$Type\ II\ error = pnorm(UCL, 23, sd) - pnorm(LCL, 23, sd)$$

```
## [1] 0.4883183
```

5. Part 5 - Hypothesis Testing

For the Hypothesis testing, tests will be made to see if the mean of certain data is the same or if it differs.

Hypothesis 1: Testing if delivery time and sales price is different for all sales classes

The following hypothesis will be formed that the mean for delivery time and price is the same for every class. This is critical for the company in two ways:

The company's reliability will be impacted if they would assume each class has the same mean delivery time. Since the company would promise customers a certain delivery time. If this hypothesis is incorrect the company would lose customers due to low reliability and the companies service levels will drop.

Ho: Class has no impact on the price or the delivery time of the products.

Ha: Class has an impact on Price as well as the Delivery time for the products.

```
## Call:
## cbind(Delivery.time, Price) ~ Class
##
## Descriptive:
##        Class     n Delivery.time     Price
## 1   Clothing 26396         9.000   640.524
## 2       Food 24582         2.502   407.785
## 3      Gifts 39141        12.891  2961.922
## 4  Household 20065        48.719 11008.295
## 5     Luxury 11867         3.971 64855.630
## 6     Sweets 21565         2.501   304.029
## 7 Technology 36345        20.011 29508.585
##
## Wald-Type Statistic (WTS):
##       Test statistic df   p-value
## Class "1157398.628"  "12" "<0.001"
##
## modified ANOVA-Type Statistic (MATS):
##       Test statistic
## Class        1158250
##
## p-values resampling:
##       paramBS (WTS) paramBS (MATS)
## Class "<0.001"      "<0.001"
```

Inspecting from the Anova table the P value is extremely small. This means that the mean for all classes isn't the same for delivery time and price

The null hypothesis was rejected with high certainty seeing as the class of the product has a significant impact on the price as well as the delivery time of the product.

Thus, the company cannot undertake the same delivery time per product and in order to be reliable they should inform customers on the average delivery time per class. The company cannot promise the same service level for every product.

A graph will be plotted to validate whether the answer to the hypothesis is true

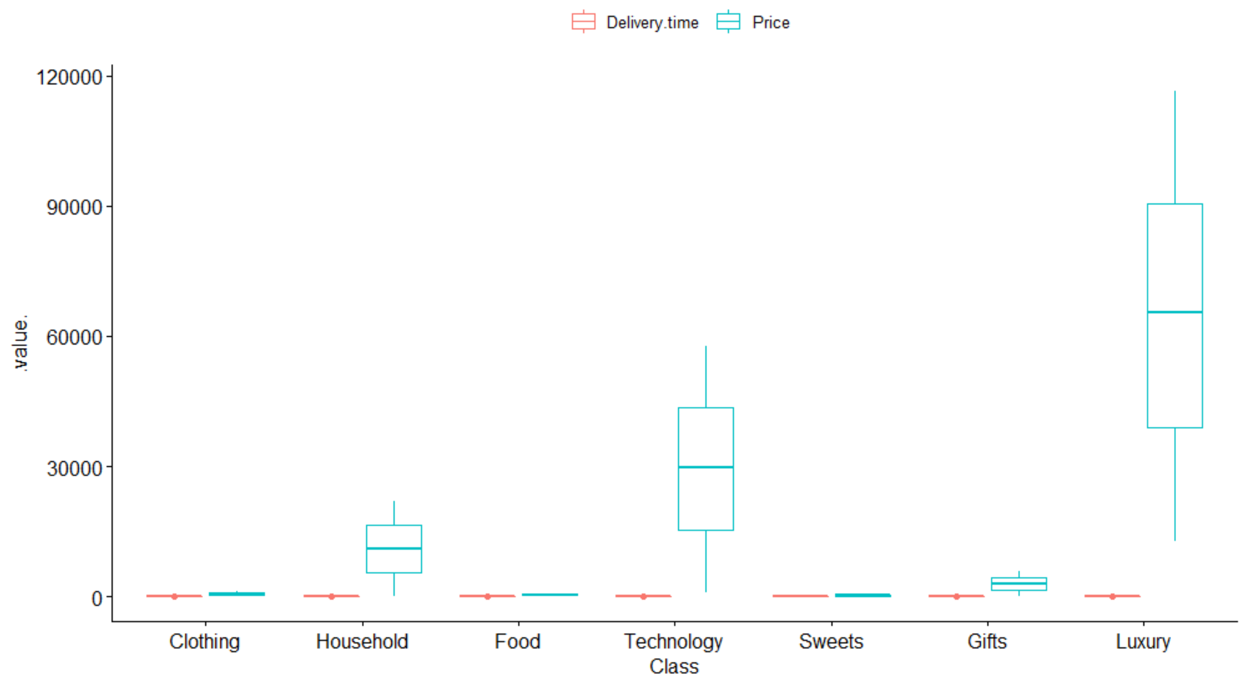


Figure 42: Boxplot of delivey time and price per class

From the plot it is evident that thee mean values for both delivery time and price is different for each class.

Hypothesis 2: Whether or not certain ages prefers certain class items

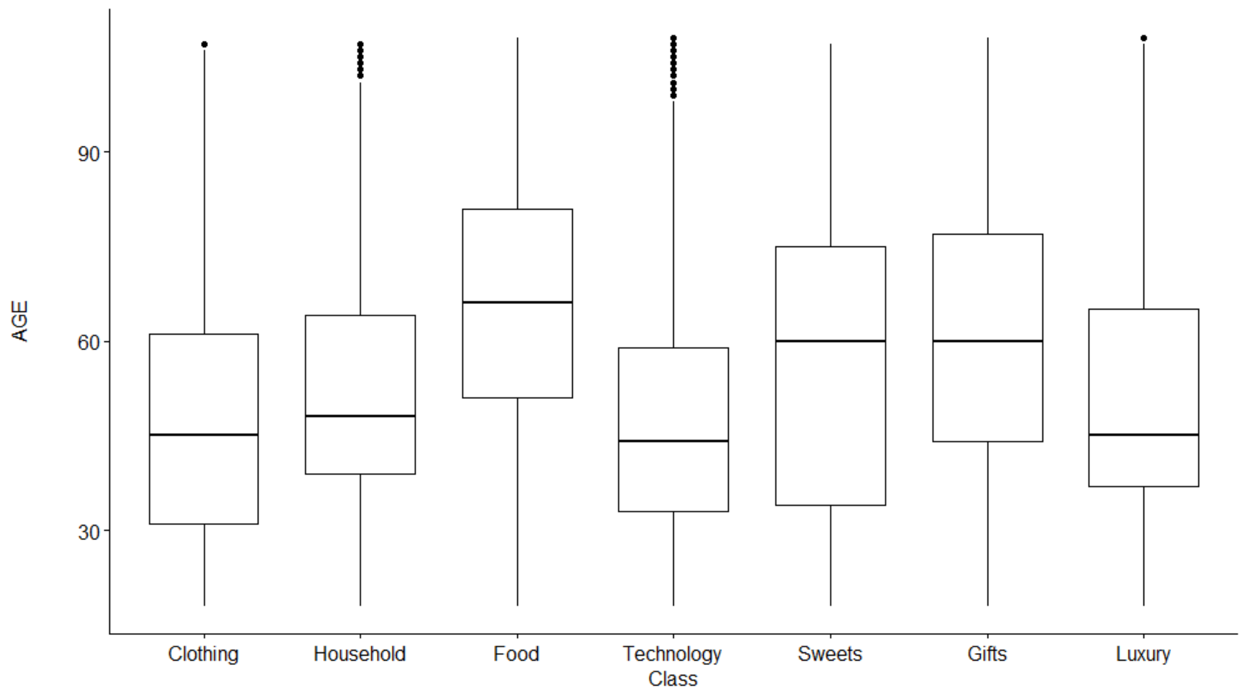
This hypothesis will be inspected since the company should determine whether they should solely focus on a certain age group per class.

Ho: Customer age has no impact on the class they are buying

Ha: Customer age has an impact on the class they are buying

```
## Call:
## AGE ~ Class
##
## Descriptive:
##        Class     n    AGE
## 1   Clothing 26399 47.468
## 2       Food 24587 65.371
## 3      Gifts 39146 60.827
## 4  Household 20065 51.929
## 5     Luxury 11867 51.338
## 6     Sweets 21566 57.155
## 7 Technology 36348 46.643
##
## Wald-Type Statistic (WTS):
##       Test statistic df  p-value
## Class "24066.781"    "6" "<0.001"
##
## modified ANOVA-Type Statistic (MATS):
##       Test statistic
## Class       24066.78
##
## p-values resampling:
##       paramBS (WTS) paramBS (MATS)
## Class "<0.001"      "<0.001"
```

The p value is small thus we reject the hypothesis that age does not have an impact on the number of products sold



	Clothing	Household	Food	Technology	Sweets	Gifts	Luxury
n	26403	24582	39149	20065	11868	21564	36347
mean	47	65	61	52	51	57	47
median	45	66	60	48	45	60	44
min	18	18	18	18	18	18	18
max	107	108	108	107	108	107	108

Figure 43: Boxplot of age per class group

Visual study of the graph reveals that the average age of each sold goods is not identical. Thus, the corporation should target specific age groups more precisely when selling products.

6. Part 6 – Reliability of the service and products

Following this section, diverse consulting tasks will be performed for various firms and problems.

6.1. Reliability of service and products

The Taguchi loss function for manufacturing refrigerator parts will be determined for Cool Food Inc.

<u>Problem 6 p359 11th edition</u>

Specification: 0.06 ± 0.04 cm

Cost to scrap: $45

$L(x) = k\,(x - T)^2$

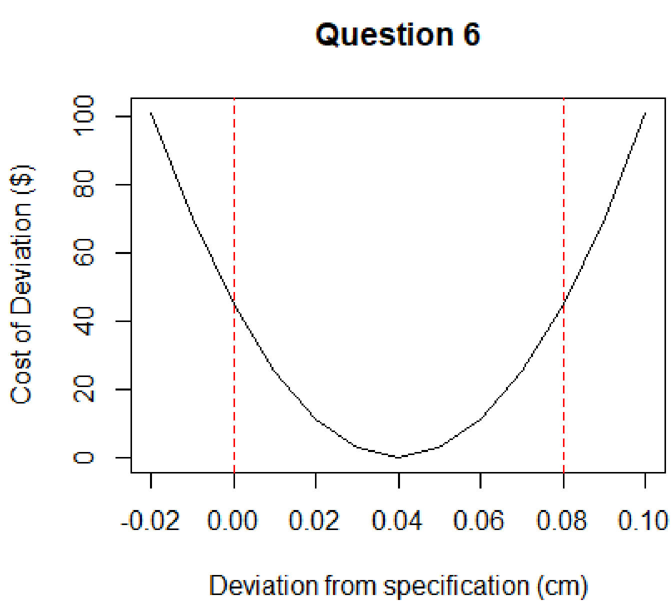
$45 \;\; = \; k\,(0.04)^2$

```
## [1] 28125
```

Thus, the Taguchi loss function can be derived as:

$L(x) = 28125(x - 0.04)^2$

Drawing the function in a plot:



Comments: Thus, Cool Food Inc. can determine the cost associated when tolerance of part does not meet specifications.

Problem 7 p359 11th edition

Scrap cost reduced to $35 per part

a) $L(x) = k\,(x - T)^2$

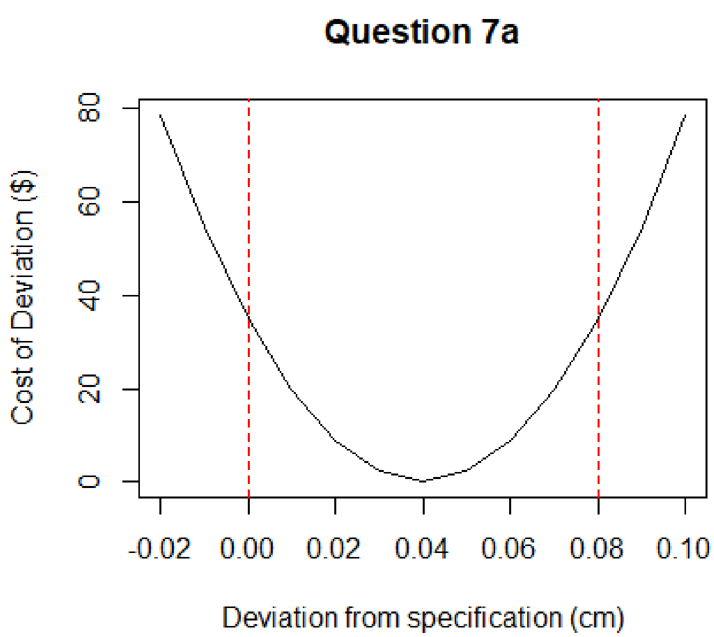
   $35 \;=\; k\,(0.04)^2$

```
## [1] 21875
```

Thus, the Taguchi loss function can be derived as:

$L(x) = 21875(x - 0.04)^2$

Drawing the function in a plot:



Comment: In the graph it can be seen that reducing scrap cost will save Cool Food Inc.

b) The Taguchi loss will be determined when the deviation from the target is reduced to 0.027 cm.

   $L(x) = 21875(x - 0.04)^2$

   $L(0.027) = 21875(0.027)^2$

   = $ 15.94

Comment: The new divaiation will result in a lower scrap cost value

6.2. A and B

The likelihood that Magnaplex's work station will not fail will be determined:

Machine	Reliability
A	0.85
B	0.92
C	0.90

a.) Only one works at each stage thus the machines are in series

$R_s = R_A \times R_B \times R_C$

$= 0.85 \times 0.92 \times 0.90$

```
## [1] 0.7038
```

Comment: The manufacturing process is 70.38% reliable if only one machine at each stage is operational

b)

A_reliable <- 1- (1-a)^2

B_reliable <- 1- (1-b)^2

C_relaible <- 1- (1-c)^2

total_reliable <- A_reliable*B_reliable*C_relaible

```
## [1] 0.9615316
```

Comment: The manufacturing process is 96.15% reliable when 2 identical machines are run per stage. This will greatly improve on the process reliability of only one machine being operational.

6.3. Binomial question

Vehicles (/20)	Days available (/1560)
20	190
19	22
18	3
17	1

Available days of drivers

Drivers (/21)	Days available (/1560)
20	95
19	6
18	1

The binomial distribution will be used to calculate the probability of a reliable process to calculate the number of days:

$$f(x) = \binom{n}{x} \times p^x \times (1-p)^{n-x} = \frac{n!}{x!\,(n-x)!} \times p^x \times (1-p)^{n-x}$$

P(21 vehicles) $= \binom{21}{0} \times \frac{1560-190-22-3-1}{1560}^0 \times (1 - \frac{1560-190-22-3-1}{1560})^{21-0}$
= 0.007071612

P(20 vehicles)= $\binom{21}{1} \times \frac{190}{1560}^1 \times (1 - \frac{190}{1560})^{21-1}$
= 0.006643289

P(19 vehicles)= $\binom{21}{2} \times \frac{22}{1560}^2 \times (1 - \frac{22}{1560})^{21-2}$
= 0.008927524

P(18 vehicles)= $\binom{21}{3} \times \frac{3}{1560}^3 \times (1 - \frac{3}{1560})^{21-3}$
= 0.01217067

P(17 vehicles)= $\binom{21}{3} \times \frac{1}{1560}^3 \times (1 - \frac{1}{1560})^{21-3}$
= 0.01967464

Then the weighted average will be calculated. The weighted average will be multiplied with 365 days. This yields a result of 364.8445 days of reliable delivery.

7. Conclusion

There are several key drivers and actions that the company can deploy and execute to improve the reliability of its delivery system. Implementing elements such as control charts would have a significant impact on the company's performance measurement. As retaining a client is easier than acquiring a new one, reducing delivery times to keep consumers satisfied is also a smart strategy. Likewise, promoting the sale of luxury things items is a good idea because they command a higher price and generate more money. As indicated by te statistical process control charts overall, the organization has a dependable delivery system that could use a few minor adjustments but generally meets their needs.

8. References

R.Evans, J. and M. Lindsay, W., 2022. *Managing for Quality and performance Excellence*. 9th ed.

Lean Six Sigma. (2018). Taguchi Loss Function [online]. Available from: https://www.leansixsigmadefinition.com/glossary/taguchi-loss-function/

[Accessed 18/10/22]

Quality System, Inc.(2021) . X-bar and sigma chart formulas [online].

Available from: https://www.pqsystems.com/qualityadvisor/formulas/x_bar_sigma_f.php [Accessed 18/10/22]