Quality Assurance 344

ECSA Project Part 1 & 2

Keighan Bradfield

22837647

24 September 2022

List of figures

Table of Contents

1 Data Wrangling

The data wrangling section of this project was aimed at creating two independent data sets, one containing valid data and the other invalid. This was done by manipulating a data set of ten variables and 180 000 instances by means of sub setting. The goal of this was to use the valid data set to perform further analysis. The invalid data set is comprised of all the instances which contain missing values or negative values, which are obscurities and negatively affect the study. The invalid data set is displayed below:

	Index	X	ID	AGE	Class	Price	Year	Month	Day	Delivery.time	Why.Bought
144443	1	144443	37737	81	Food	-588.8	2022	12	10	2.5	Recommended
16320	2	16320	44142	82	Household	-588.8	2023	10	2	48.0	EMail
19998	3	19998	68743	45	Household	-588.8	2024	7	16	45.5	Recommended
155554	4	155554	36599	29	Luxury	-588.8	2026	4	14	3.5	Recommended
19540	5	19540	65689	96	Sweets	-588.8	2028	4	7	3.0	Random
98765	6	98765	64288	25	Clothing	*NA*	2021	1	24	8.5	Browsing
54321	7	54321	62209	34	Clothing	*NA*	2021	3	24	9.5	Recommended
34567	8	34567	18748	48	Clothing	*NA*	2021	4	9	8.0	Recommended
155555	9	155555	33583	56	Gifts	*NA*	2022	12	9	10.0	Recommended
177777	10	177777	68698	30	Food	*NA*	2023	8	14	2.5	Recommended
56789	11	56789	63849	51	Gifts	*NA*	2024	5	3	10.5	Website
87654	12	87654	40983	33	Food	*NA*	2024	8	27	2.0	Recommended
166666	13	166666	60188	37	Technology	*NA*	2024	10	9	21.5	Website
19541	14	19541	71169	42	Technology	*NA*	2025	1	19	20.5	Recommended
19999	15	19999	67228	89	Gifts	*NA*	2026	2	4	15.0	Recommended
12345	16	12345	18973	93	Gifts	*NA*	2026	6	11	15.5	Website
23456	17	23456	88622	71	Food	*NA*	2027	4	18	2.5	Random
65432	18	65432	51904	31	Gifts	*NA*	2027	7	24	14.5	Recommended
144444	19	144444	70761	70	Food	*NA*	2027	9	28	2.5	Recommended
76543	20	76543	79732	71	Food	*NA*	2028	9	24	2.5	Recommended
16321	21	16321	81959	43	Technology	*NA*	2029	9	6	22.0	Recommended
45678	22	45678	89095	65	Sweets	*NA*	2029	11	6	2.0	Recommended

Figure 1: Data Frame containing the invalid instances

The table above contains 17 instances with missing values and 5 instances containing negative values within the price feature. Both data frames have extra indexing columns which will prove beneficial further into the report. The coding to obtain this result was done by using functions to extract the features containing missing values or negative numbers which was then row bound to result in the above table. With this set of instances, we now can be confident to start our analysis on a data set that is impurity free.

A small preview of the valid data to be used for statistical analysis is shown below. This final data set is comprised of 179978 instances and 11 features. The focus will be on the delivery time feature and the goal is to understand what factors affect this time by using a statistical analysis.

	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
27161	13	27161	56240	45	Household	17681.94	2021	1	1	45.5	Website
27638	14	27638	24396	30	Clothing	1018.21	2021	1	1	8.5	Recommended
30778	15	30778	12235	28	Technology	21096.86	2021	1	1	15.0	Website
34277	16	34277	30290	43	Household	10573.67	2021	1	1	51.0	Recommended
34950	17	34950	40035	77	Household	16548.61	2021	1	1	51.5	Recommended
35153	18	35153	36435	53	Technology	23304.75	2021	1	1	14.0	Browsing
37187	19	37187	49974	67	Sweets	332.46	2021	1	1	2.5	Recommended
42139	20	42139	36292	75	Food	205.96	2021	1	1	3.0	Recommended
43139	21	43139	39202	34	Clothing	353.28	2021	1	1	9.0	Recommended
44379	22	44379	92277	82	Food	219.76	2021	1	1	2.5	Recommended
45422	23	45422	12068	56	Gifts	1320.89	2021	1	1	6.5	Website

Figure 2: Data Frame containing the valid data

2 Descriptive statistics

A data quality report was generated for the continuous and categorical features separately to understand and identify obscurities in the data. These are displayed below:

Feature	Count	%Missing	Cardinality	Min	1st Qrt.	Mean	Median	3rd Qrt.	Max	Std. Dev.
Index	179978	0	179978	1	44995.25	89989.5	89989.5	134983.8	179978	51955.32
X	179978	0	179978	1	45004	90005	90003	135000	180000	51960.7
ID	179978	0	15000	11126	32700	55081	55235	77637	99992	25740.27
Age	179978	0	91	18	38	53	54.57	70	108	2.038.881
Price	179978	0	78832	35.65	482.31	2259.63	12294.1	15270.97	116618.97	20889.15
Year	179978	0	9	2021	2022	2025	2025	2027	2029	2.783.364
Month	179978	0	12	1	4	7	6.521	10	12	3.453.849
Day	179978	0	30	1	8	16	15.54	23	30	8.648.721
Delivery time	179978	0	148	0.5	3	10	14.5	18.5	75	1.395.578

Feature	Count	%Missing	Cardinality	Mode	Mode Freq.	Mode %	2nd Mode	2nd Mode Freq	2nd Mode %	
Class	179978	0	7	Gifts	39149	21.75	Technology	36347	20.2	
Why Bought	179978	0	6	Recommended	106985	59.44	Website	29447	16.36	

Figure 3: Data quality report

It is now clearly visible that the valid data contains 2 categorical features as well as 9 continuous features.

2.1 Plots of features

2.1.1 Continuous features:

Price:

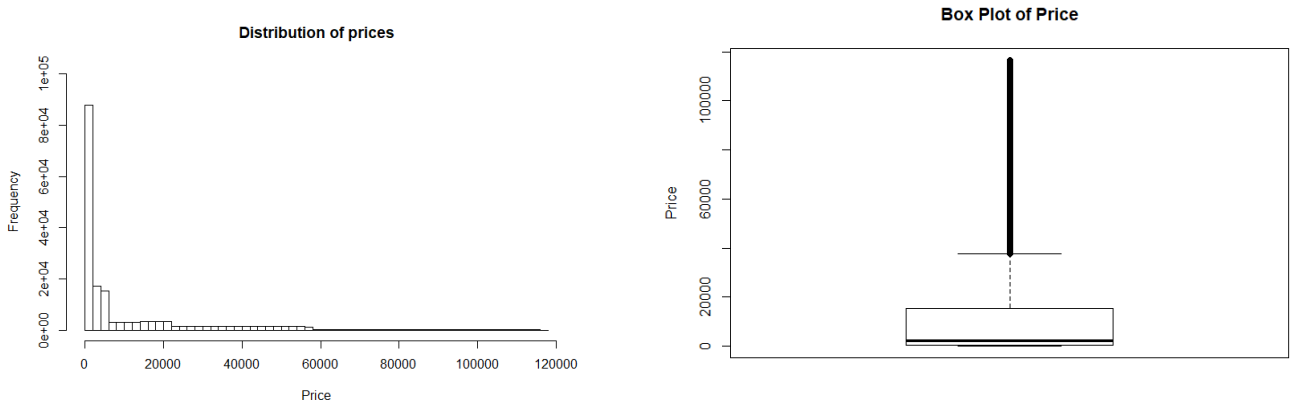


Figure 4: Plots showing the distribution of the price feature

The histogram of the price distribution shows a steep exponential distribution, where the boxplot indicates that the bulk of the prices lie below the range of 20 000, with very high extremes. These extremes could be outliers but will require further analysis to fully determine.

Delivery time:

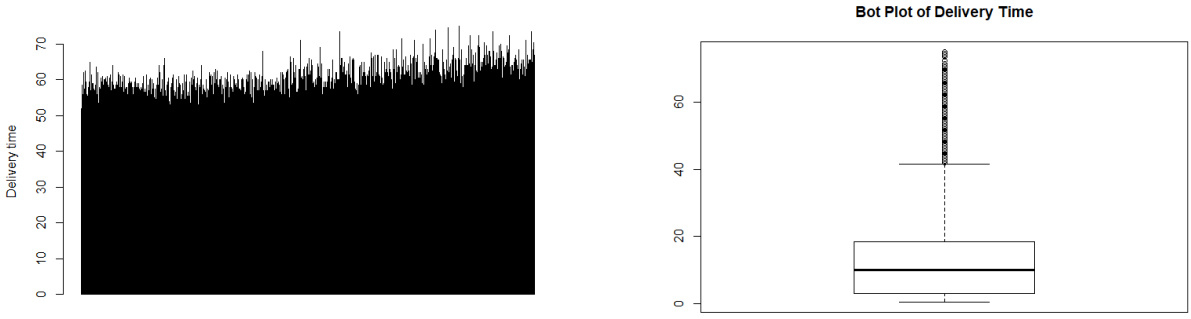


Figure 5: Plots showing the distribution of the delivery time feature

Day:

Below the bar plot of the day distribution is visible. This distribution is highly uniform, and this indicated that sale probability is equal for any day of the month. We do see marginal increases for days 4 through 11 and random small fluctuations thereafter.

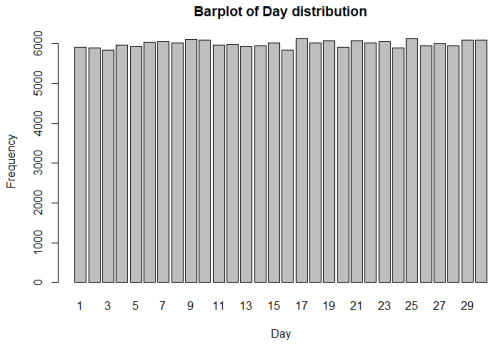
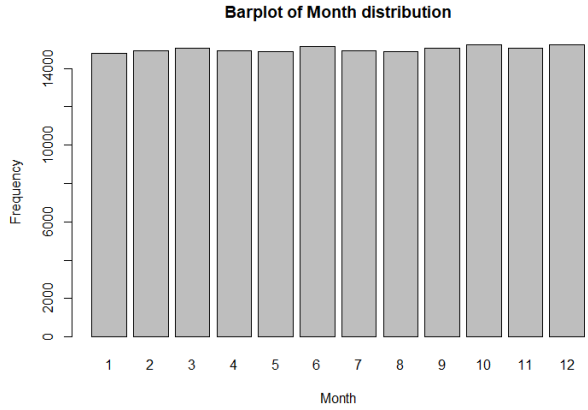


Figure 6: Plot showing the distribution of the day feature

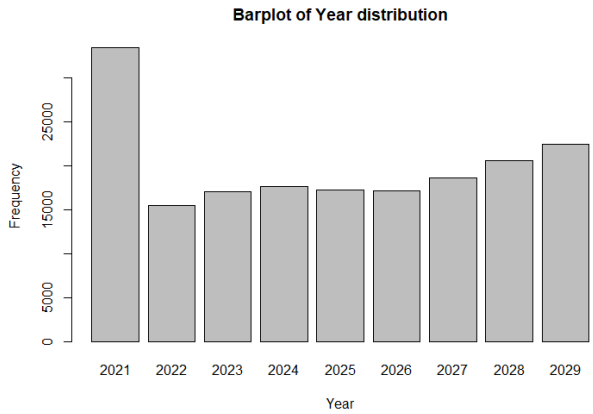
Month:



The bar plot of the month feature also follows a uniform distribution. This tells us there is no seasonality and that month will not be a beneficial feature for further analysis.

Figure 7: Plot showing the distribution of the month feature

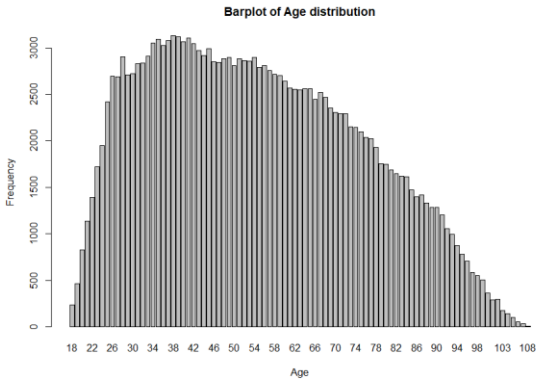
Year:



The bar plot of the feature year has no prominent shape but shows that majority of sales occurred in 2021, with a small left hand side skew from years 2022 through 2029. This means there was a large drop in sales after 2021 with a steady and slow increase for the following years with minor fluctuations.

Figure 8: Plot showing the distribution of the year feature

Age:

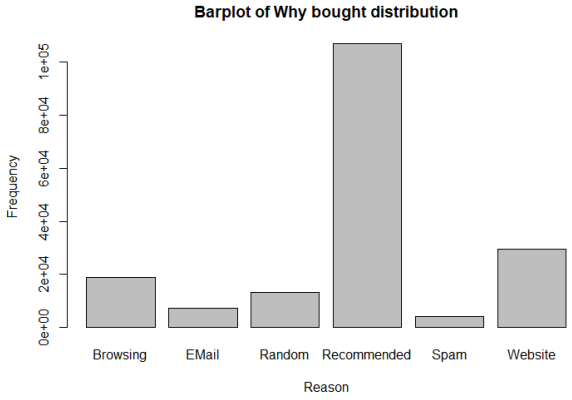


The bar plot of the age is slightly right hand skewed and shows that the most important ages range from 26 to 60, with a decreasing rate thereafter.

Figure 9: Plot showing the distribution of the age feature

2.1.2 Categorical feature:

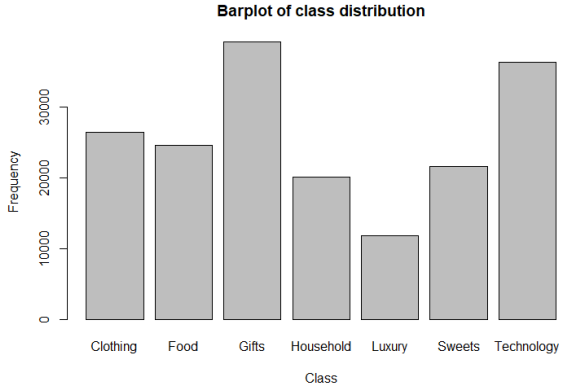
Why Bought:



The bar plot of the reason for purchase clearly indicates that the most prominent reason is through recommendations. This shows how important a good service is to consumers. This is largely impacted by a core feature of the study, namely the delivery time, as this is a large driving force for customer satisfaction.

Figure 10: Plot showing the distribution of the why bought feature

Class:



The benefit of visualising the class feature is that it outlines the key classes within the class category. It is notable that technology and gifts are the most prevalent items purchased and will furthermore mean they contribute largely to the overall revenue of the firm.

Figure 11: Plot showing the distribution of the class feature

The process capabilities for technology class items are as follows:

Cp = 1.142207

Cpl = 1.90472

Cpu = 0.3796933

Cpk = 0.3796933

When looking at these values we notice a Cpk of 0.3797, this is seen as bad as a Cpk of 1 is understood as barely capable. To improve the Cpk we would ideally reduce variation or spread of the process. It is also possible to improve this by recentring the process on the target.

The delivery time must have an LSL of at least zero, this is because delivery time is based on time which cannot be negative in the context of this situation.

3: Statistical process Control (SPC)

3.1 Control Limits and charts

The data instances provided are used to determine the control limits of the various classes. This is achieved by reordering the instances from oldest to newest and then preforming the necessary calculations on the first 450 instances to get the control limits for the classes seen below. The two important charts for this application are the X-chart as well as the S-chart.

3.1.1 S and X-Chart

X- Chart

The main purpose of an x chart is to outline the mean change of a process over time using subgroup values (subgroup of 30 samples).

	X- Chart						
Class	UCL	U2Sigma	U1Sigma	CL	L1Sigma	L2Sigma	LCL
Technology	22.97462	22.10789	21.24117	20.37444	19.50772	18.641	17.77427
Clothing	9.404934	9.259956	9.114978	8.97	8.825022	8.680044	8.535066
Household	50.24833	49.01963	47.79092	46.56222	45.33352	44.10482	42.87612
Luxury	5.493965	5.241162	4.988359	4.735556	4.482752	4.229949	3.977146
Food	2.709458	2.636305	2.563153	2.49	2.416847	2.343695	2.270542
Gifts	9.488565	9.112747	8.736929	8.361111	7.985293	7.609475	7.233658
Sweets	2.897042	2.757287	2.617532	2.477778	2.338023	2.198269	2.058514

Figure 12: Table showing the control limits for X-Chart

S-Chart

Similarly, the s chart represents the standard deviation of a process over time with this same subgroup.

S- Chart							
Class	UCL	U2Sigma	U1Sigma	CL	L1Sigma	L2Sigma	LCL
Technology	5.18057	4.552222	3.923875	3.295528	2.667181	2.038833	1.410486
Clothing	0.8665596	0.761455	0.656351	0.551247	0.446142	0.341038	0.235934
Household	7.34418	6.45341	5.56264	4.67187	3.7811	2.89033	1.99956
Luxury	1.511052	1.327777	1.144503	0.961229	0.777955	0.59468	0.411406
Food	0.4372466	0.384213	0.33118	0.278147	0.225113	0.17208	0.119047
Gifts	2.246333	1.973877	1.701421	1.428965	1.156509	0.884053	0.611597
Sweets	0.8353391	0.734022	0.632704	0.531386	0.430069	0.328751	0.227433

Figure 13: Table showing the control limits for S-Chart

The control limits for the classes; Technology, Household, Clothing, Gifts, Sweets, Luxury and Food have been specified above. These will be beneficial when plotting to get a general idea of the instances that lie outside of the control limits which can then be used to identify those instances more accurately.

3.1.2 Chart analysis of initial samples

The graphics that follow are comprised of 30 samples each containing 15 sales. These initial samples were used to obtain an estimate of the mean and standard deviation of the delivery process of each class seen below. These estimates were used to generate the control limits which are seen by the green (UCL and LCL) and red (CL) lines on each plot respectively. The purpose of the S-chart is to identify the reliability of the X-chart. An X- chart is seen as reliable if the S- chart stays within its control limits and is unreliable otherwise. When assessing the initial samples, we expect the plots to stay within the control limits and use these plots to identify if this is indeed the case.

The classes that were seen to be within the control limits are noted below.

Technology:

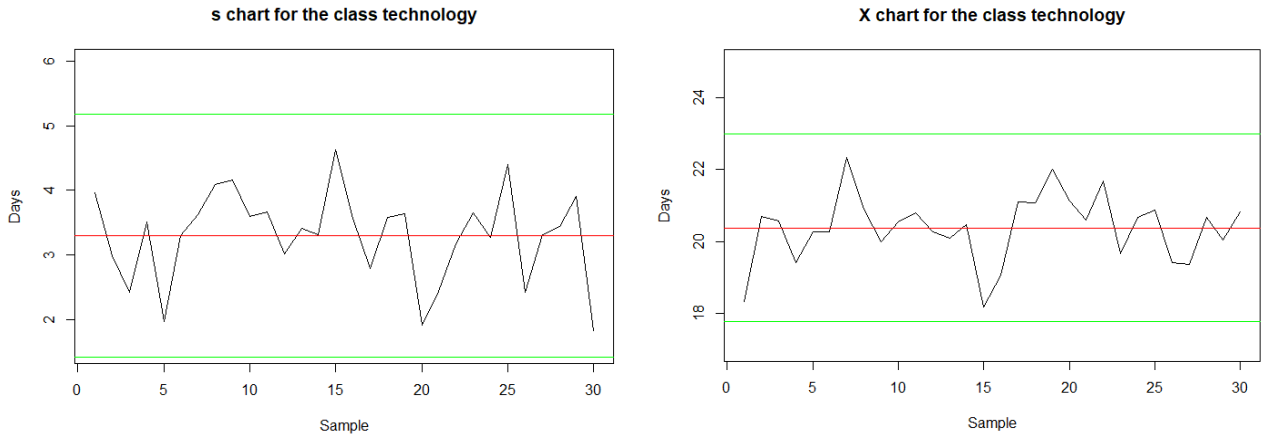


Figure 14: S and X charts for the technology class

Clothing:

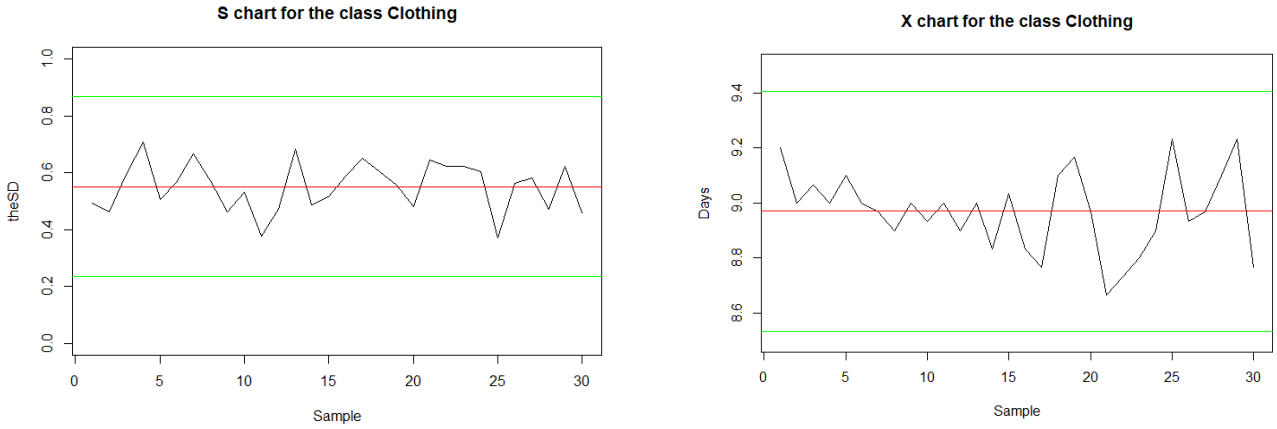


Figure 15: S and X charts for the clothing class

Household:

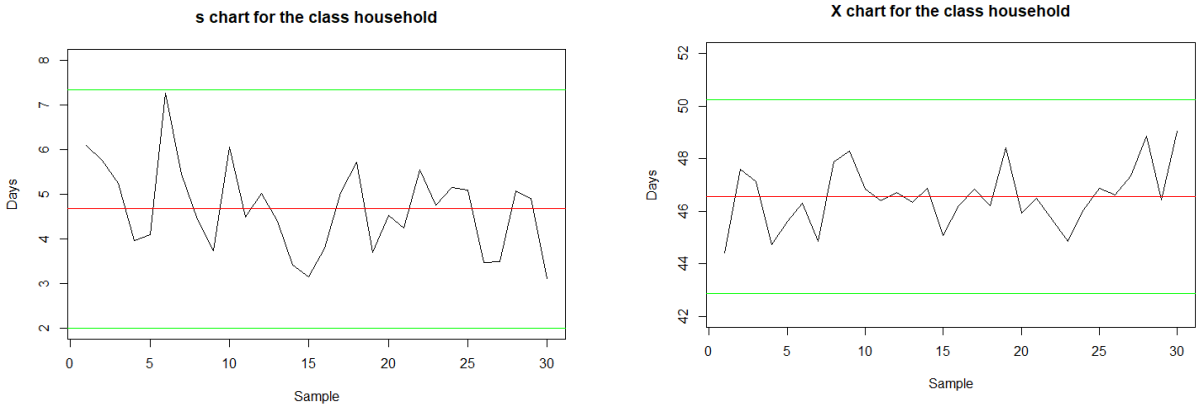


Figure 16: S and X charts for the household class

Luxury:

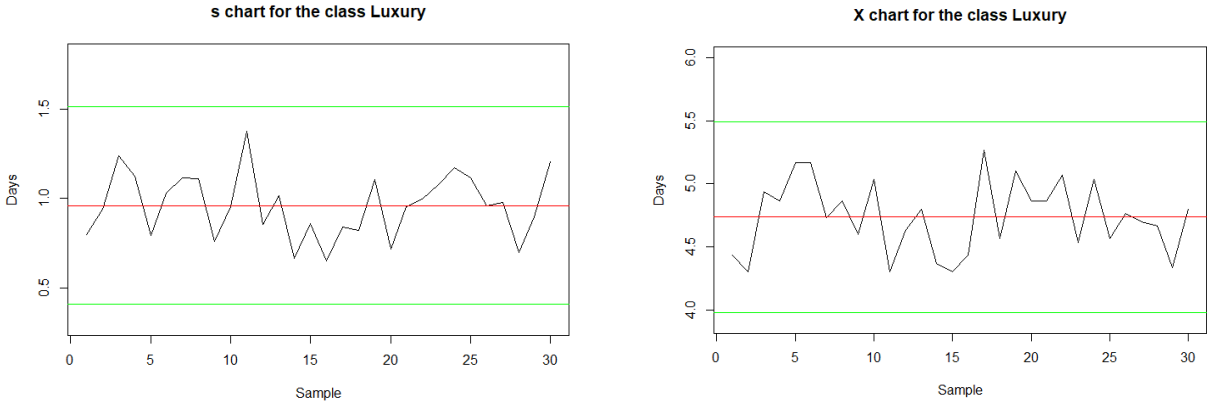


Figure 17: S and X charts for the luxury class

Food:

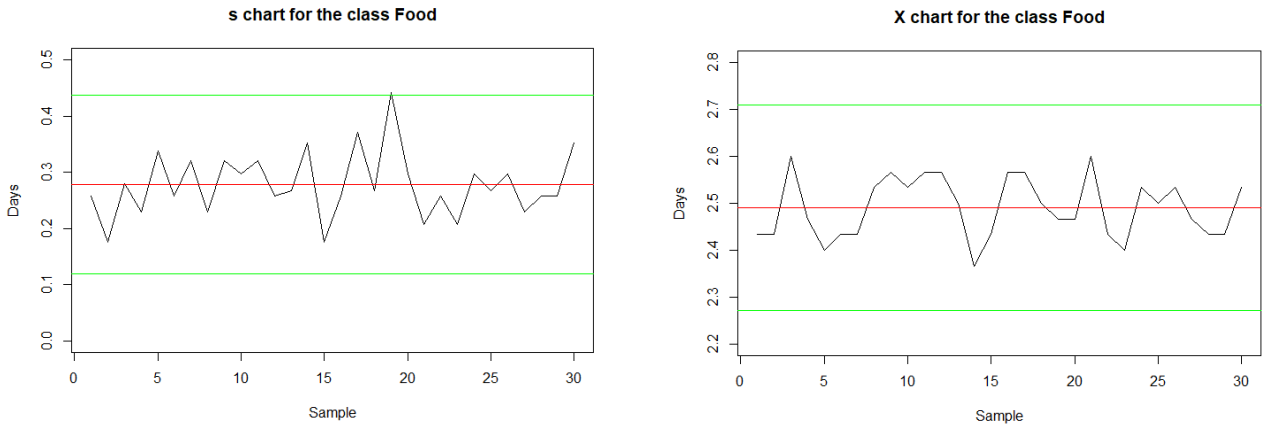


Figure 18: S and X charts for the food class

Gifts:

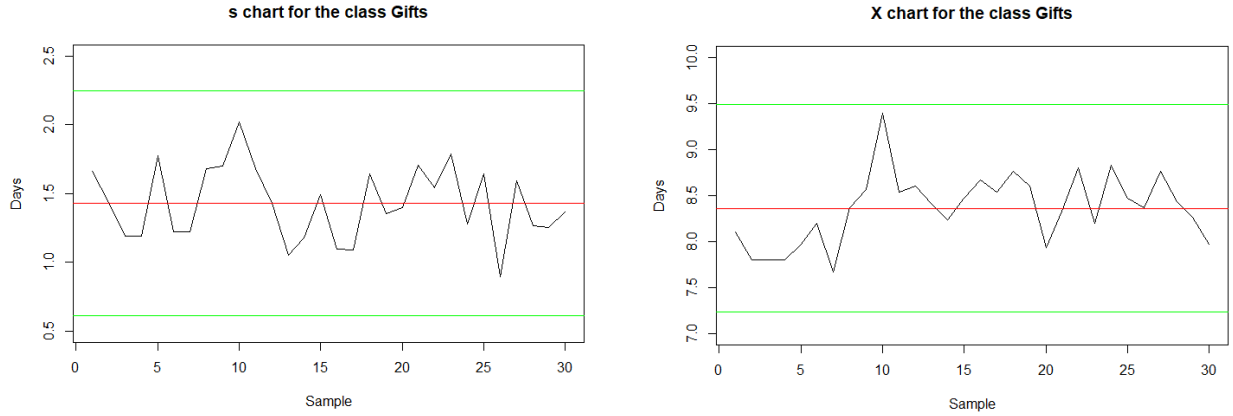


Figure 19: S and X charts for the gifts class

The only class that does not conform to the control limits is shown below, this means that the X-chart is unreliable as instances within the initial samples lie outside of the S-chart limits.

Sweets:

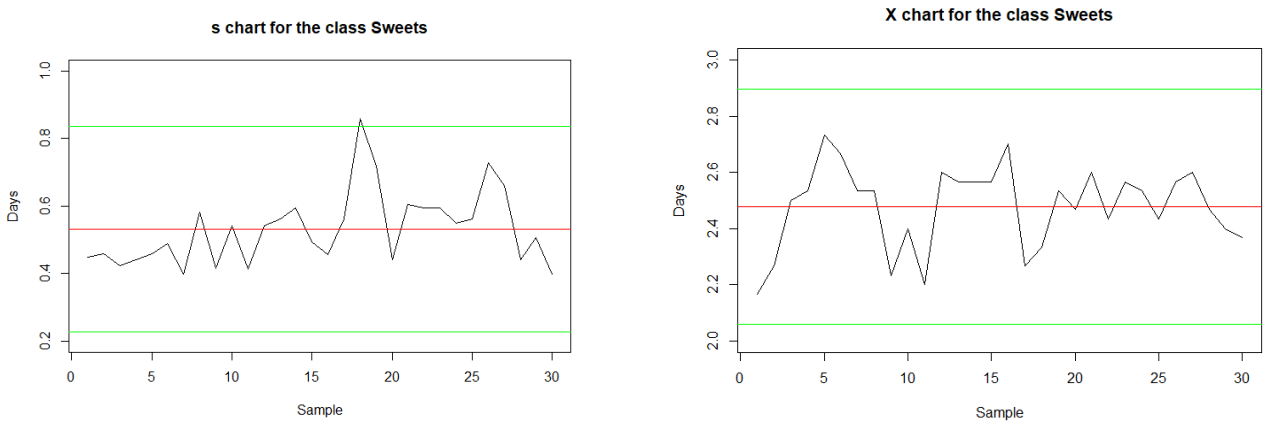


Figure 20: S and X charts for the sweets class

To rectify this issue, the specific instance that lies outside of the limits should be removed and control limits can be recalculated.

3.2 Process control analysis for all instances

The process control is now analysed for the entire class, containing all instance. The focus point of this control is based on the mean of each class. Classes that contain instances that lie within the limits are seen to be in control and are stable, otherwise the process is seen to be out of control. To identify these, it is important to first assess the S-chart to understand if the X-chart is an accurate representation of the results.

Classes that conform and are considered in control are noted below. Some instances of these charts lie outside the limits and are seen as outliers.

Technology:

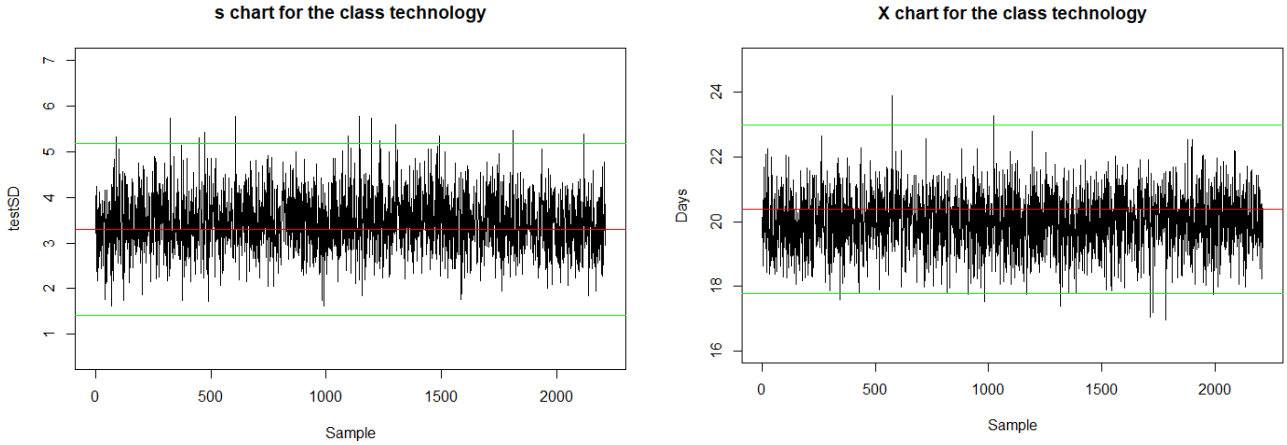


Figure 21: S and X charts for the technology class (All samples)

When looking at the S chart above, we notice the bulk of instances sit slightly higher than the specified centre line. This indicated that the process has slightly more variation than what we want. There are also quite a few instances sitting above the UCL further proving this undesirable variation above the norm. When we observe the X- chart, we note only a few instances cross both the UCL and the LCL and can then be considered outliers. We therefore conclude the technology class has a higher variation than ideal but stays within statistical control.

Clothing:

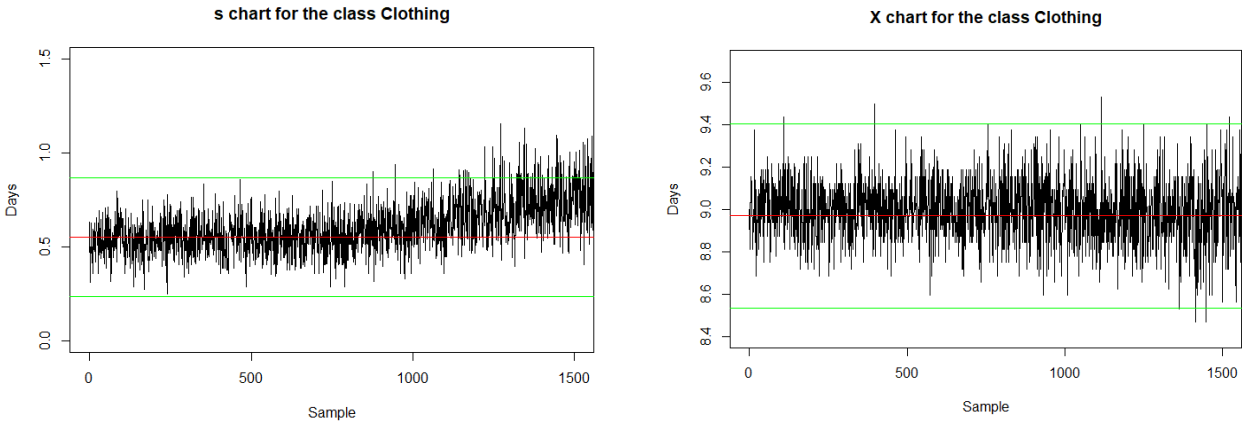


Figure 22: S and X charts for the clothing class (All samples)

Observing the S-chart of the Clothing class we see the process conforms with an acceptable variation until around 1000 samples, after which the process lags upwards and has a significantly high variation with most instances crossing the UCL. This follows closely with the X- chart as we see good statistical control up to around 1250 samples and then a spike in the variation of the mean with samples dropping below the LCL. There are not many instances, and this can therefore be accounted for as outliers.

Food:

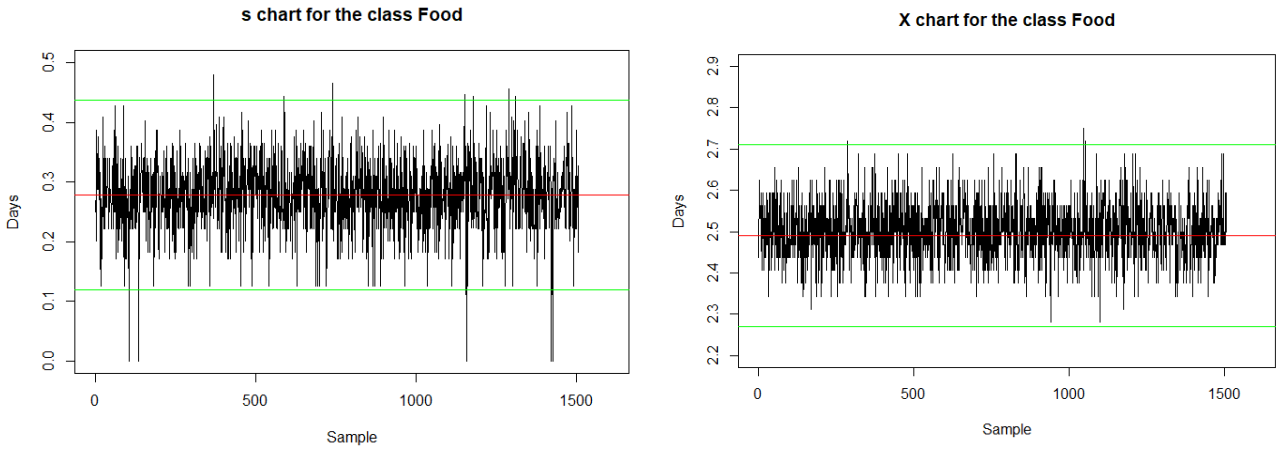


Figure 23: S and X charts for the food class (All samples)

Sweets:

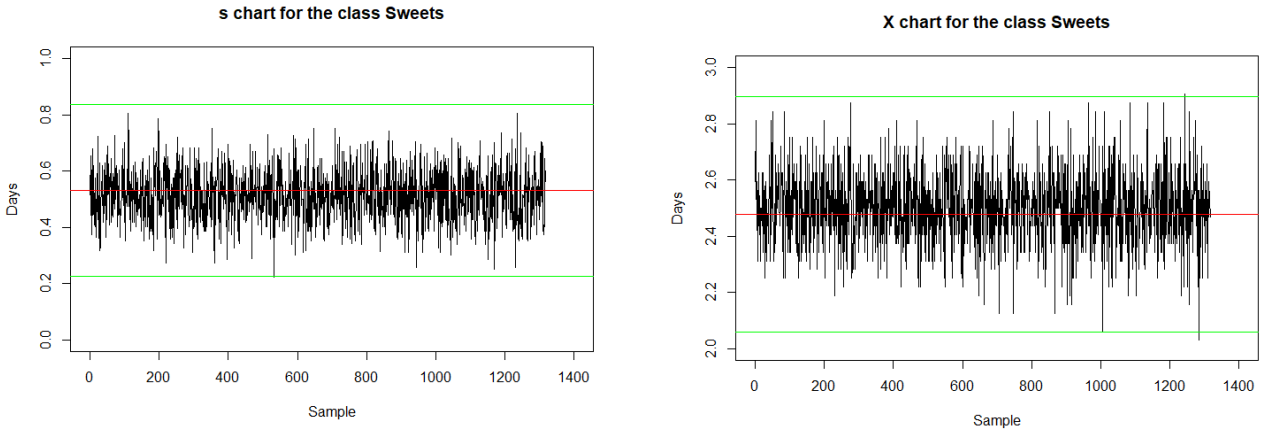


Figure 24: S and X charts for the sweets class (All samples)

Food and sweets both have S- charts both have reasonably good variations that conform to the control limits. There are however instances that far outstretch the limits and can also be seen as outliers. Their X-charts similarly show that the processes are predominantly in statistical control and are noted as processes that are in control.

Furthermore, classes that do not conform and are seen as out of control processes are depicted below. These are seen as processes that require monitoring and improvement.

Household:

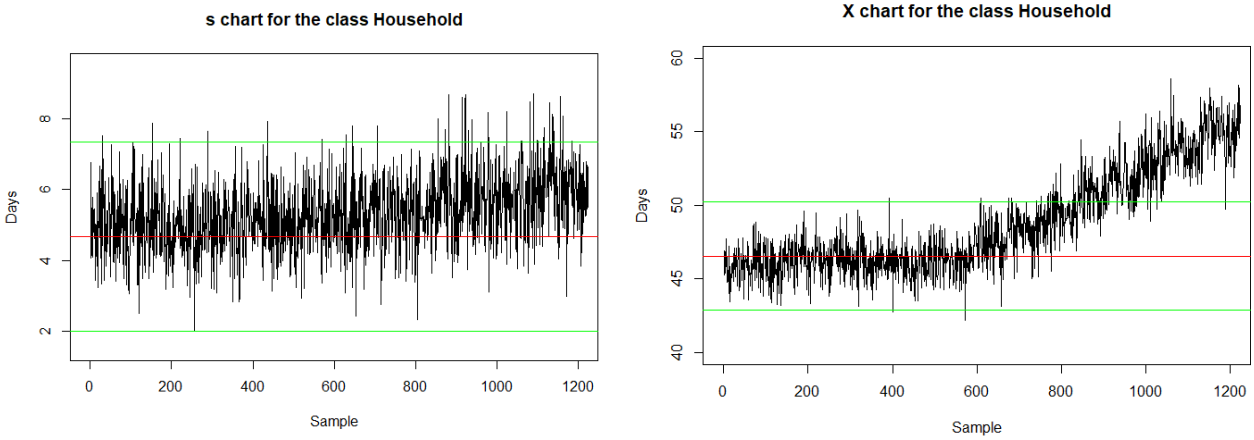


Figure 25: S and X charts for the household class (All samples)

When analysing the X-chart of the household class, it is clearly visible that there is an upward trend with just less than half the instances falling outside of the control limits. This means that delivery time is increasing and can lead to a negative impact of the company as it is seen as poor quality. Management should identify the reasons for this issue and rectify it to maintain relationships with and brand quality with relevant customers.

Luxury:

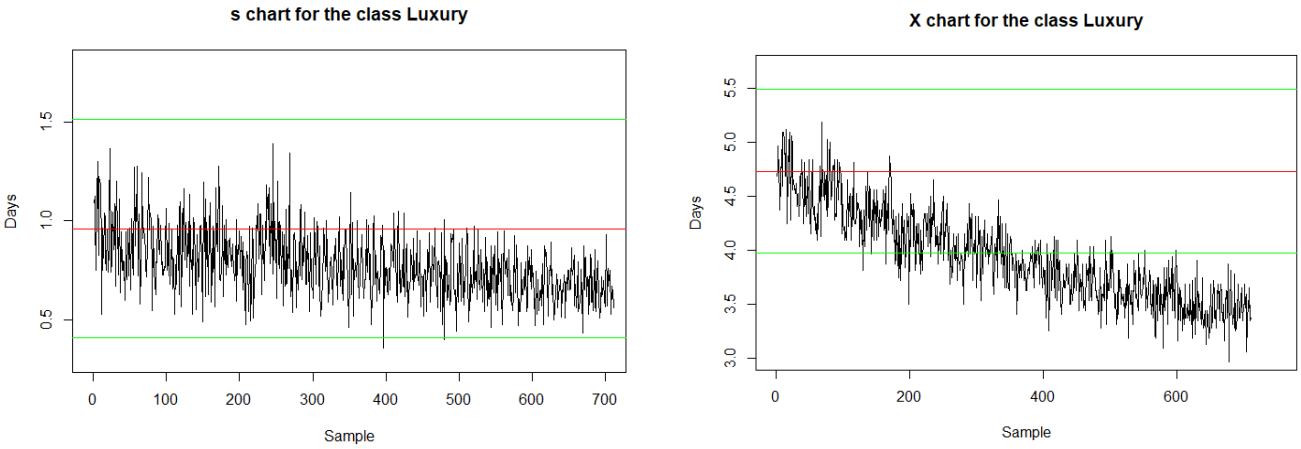


Figure 26: S and X charts for the luxury class (All samples)

As seen in the figure above, delivery time for the luxury class has a clear downward trend and breach the lower control limit. This means that delivery time has improved and is seen as a positive change. This means that customers of luxury items are receiving goods faster and will help create loyalty as well as additional revenues for the company. Management must identify if this shift is caused by a specific change and how to maintain that change. New control limits should be calculated for this class.

Gifts:

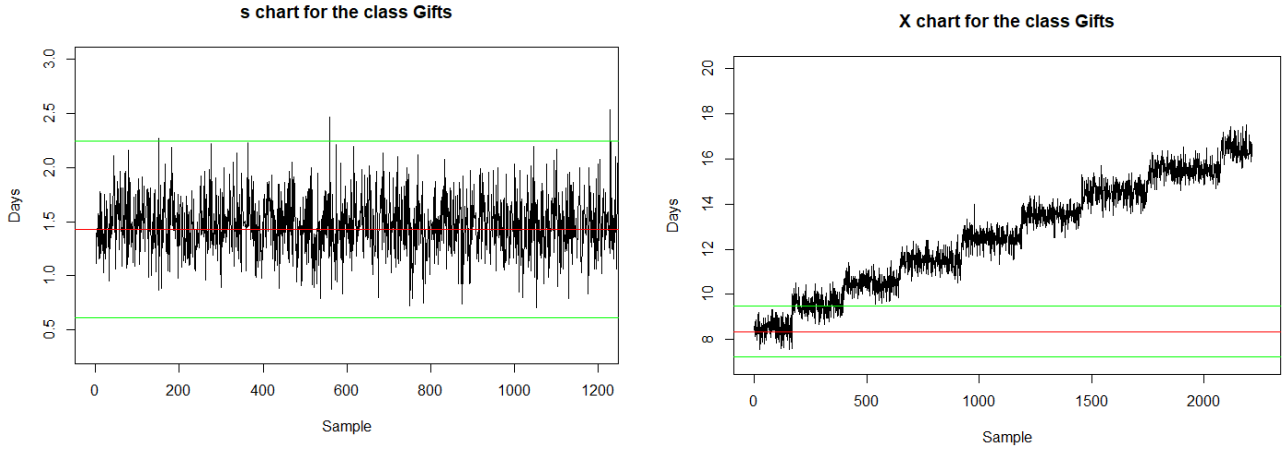


Figure 27: S and X charts for the gifts class (All samples)

Looking at the above figure shows us the class that has seen the most variation in delivery time of all the classes considered. Delivery time has a large upward trend noticed in jumps followed by a stable delivery time for around 250 instances before the next jump. This increased delivery time will have a large and negative impact of the company and issues revolving around this must be rectifies as soon as possible.

Part 4: Optimising the delivery process

4.1.1 Sample means outside of the control limits

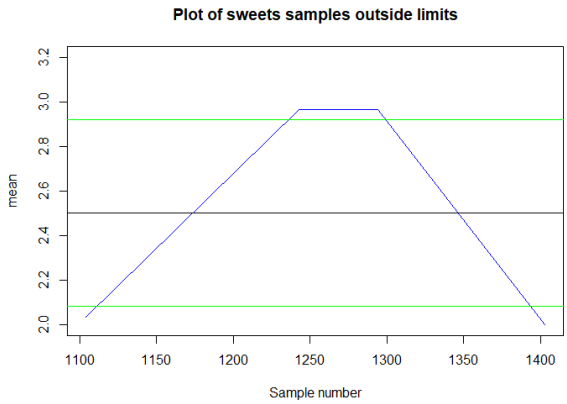
To identify and understand the problems of each class, it is beneficial to identify the samples or instances that are the root cause of the problem, these being those samples with means that lie outside of the specified control limits. Below is a table that describes each classes number of outliers and the relevant indices of these outliers. If more than six outliers are present, only the first and last three are identified for simplicity. If a class contains less than six outliers, some instances are repeated.

	Total ouliers	1st	2nd	3rd	3rd last	2nd last	last
sweets	4	1104	1243	1294	1243	1294	1403
luxury	72	3	4	5	755	766	784
technology	9	398	643	1122	1872	2009	2218
gifts	1968	1	2	3	2607	2608	2609
house	363	1	4	7	1335	1336	1337
food	5	75	633	1203	1203	1467	1515
clothing	13	455	1152	1161	1677	1723	1724

Figure 28: Display of instances with means outside control limits.

The table above shows that most classes are in statistical control and have low number of outliers. Technology is an important class that has one of the highest earnings and has a very low number of outliers which is highly favourable. Another important class, luxury, has 72 instances out of control which is high but not too concerning. This will require further investigation. Below are the plots of the instances that lie outside of the control limits for the various classes.

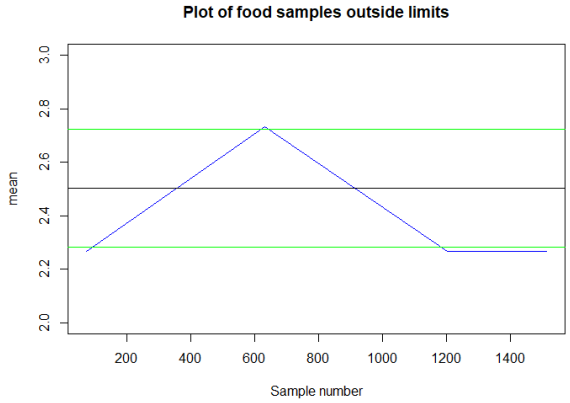
Sweets:



This class has four noticeable instances out of control in the later stages of the samples, this correlates with the S and X- charts where we noted higher variation towards the latter samples.

Figure 29: Samples outside control limits for sweets

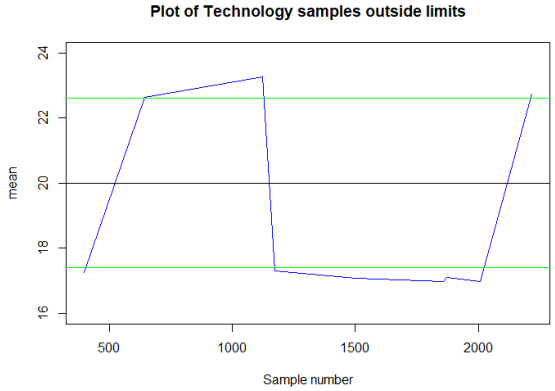
Food:



The food class also on contains a handful of instances out of control. This is presumed to be due to random variation as a process is never perfect.

Figure 30: Samples outside control limits for food

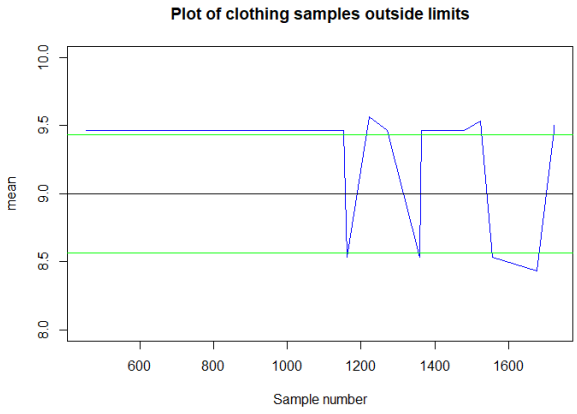
Technology:



The tech class have few instances out of control and is seen to be in control. This is very favourable as it is an important class.

Figure 31: Samples outside control limits for technology

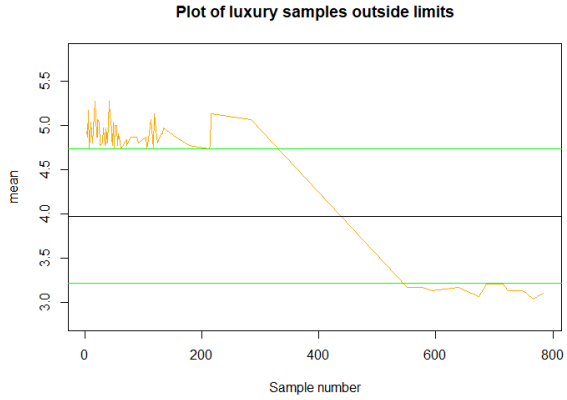
Clothing:



Clothing is like tech with only a couple concerning instances. This is seen in later samples which also correlates to S- chart variations in the latter portions of the associated plots.

Figure 32: Samples outside control limits for clothing

Luxury:



Luxury has been noted to have a noticeable number of instances out of control. The majority of these are seen to be below the LCL and although the variation is unwanted, for this application it means better delivery time which is a positive aspect for this situation.

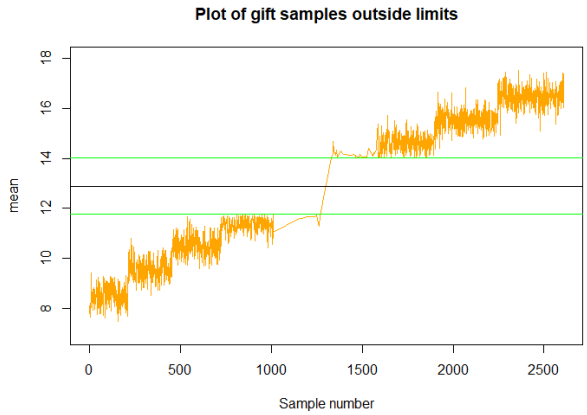
Figure 33: Samples outside control limits for luxury

Household:

The household class has many instances out of control and will require thorough investigation by the company. A few have shorter delivery times, but the vast majority are seen as a worsened time notices but the breach of the UCL which is bad for the company.

Figure 34: Samples outside control limits for household

Gifts:



Delivery time for gifts was initially seen to conform and have instances below the LCL with short delivery times. This however has rapidly increased consistently as time passes. This issue needs to be addressed as this will severely hinder the company if this trend is to continue.

Figure 35: Samples outside control limits for gifts

4.1.2 Consecutive samples of s-bar between -0.3 and +0.4 sigma limits

The table below shows the maximum number of consecutive samples that lie between the region of -0.3 and +0.4 sigma. The table also displays the ending sample number of this consecutive string of values.

	Number of consecutive samples	Ending sample number
Technology	7	737
Luxury	4	313
Clothing	6	332
Food	7	952
Sweets	5	316
Gifts	6	1282
House	5	857

Figure 36: Table showing consecutive samples between specified limits

4.2 Type I Error for A and B

A type 1 error, also known as a manufacturing error, is when a true null hypothesis is rejected. This means that the process is identified as out of control when in fact it is in control. The result is that the company spends further resources on an investigation that is not necessary as the process is performing to the correct standard.

The probability for a type 1 error for A is:

A <- pnorm(-3)*2

Which results in a probability of 0.002699796, meaning there is a 0.2699796% chance of this occurring.

The probability of a type 1 error for B is:

B <- 1-(pnorm(0.4,lower.tail = FALSE)+pnorm(-0.3))

Which results in a probability of 0.2733332, meaning there is a 27.33332% chance of being within the correct bounds. To further analyse this case for the maximum number of consecutive instances that are not within the bounds, we will raise this B to the power of the respective maximum number of consecutives to display the probabilities for each class. These are displayed below.

Food: 0.0001139846 with a max of 7

Technology: 0.0001139846 with a max of 7

Clothing: 0.0004170172 with a max of 6

Luxury: 0.005581736 with a max of 4

Household: 0.001525674 with a max of 5

Gifts: 0.0004170172 with a max of 6

Sweets: 0.001525674 with a max of 5

4.3 Optimising the delivery process

The company is said to lose R329 for each item that exceeds a delivery time of 26 hours. Assessing the number of deliveries that adhere to this criterion, the firm is incurring an unnecessary cost of R758674. To reduce this cost, the company needs to reign in their mean delivery time. To decrease the delivery time by 1 hour, they incur a cost of R2.50. A graphic of the optimal delivery time reduction is shown below.

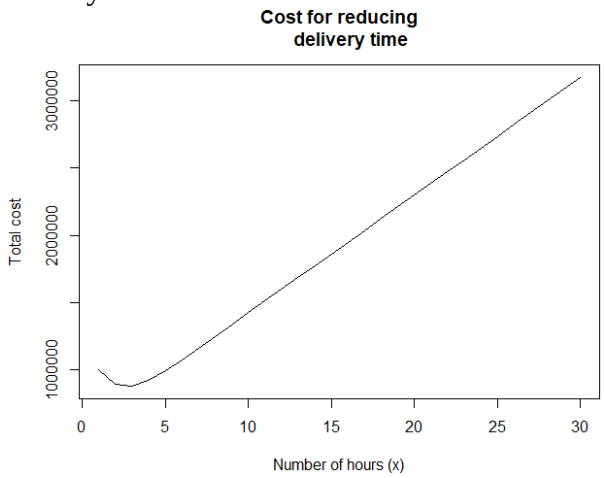


Figure 37: Graph displaying the best delivery time reduction

This graph shows that if the delivery time is reduced by 3 hours, the total cost will be minimised. This is the most optimal solution for this brute force problem.

4.4 Type ll (Consumer's) Error

The type II error, commonly known as a false negative, is an error when the null hypothesis is not rejected when it should have been. This means a process has been incorrectly identified as correct or adequate when there is in fact a problem with the process.

The probability of a type II error is calculated given a sudden jump from the current average of 20.37444 to an average of 23. This probability refers to an error that was not identified due to it falling outside the -3 and +3 sigma limits.

The probability can be calculated with the following code:

```
(pnorm(UCL_Technology,mean=45,sd=(UCL_Technology-LCL_Technology)/6)-
pnorm(LCL_Technology,mean=45,sd=(UCL_Technology-LCL_Technology)/6))*100
```

= 48.83196 %

This means that the unexpected change in mean of the technology class will result in a type II error of 48.83196%. This value means that with enough variation in the delivery process over time, the more likely the error will occur and therefore negatively influence the company as certain problems will not be correctly identified.

5 DOE and MANOVA

MANOVA is a valuable technique that can be helpful to identify the impact of a feature or multiple features on one another. For this analysis the impact of price and delivery time on the why bought feature are considered. The MANOVA has been structured with the following hypotheses as the target of the investigation.

H_O = The price and delivery time are equally distributed across the various why bought's.

H_1 = At least one differs

The result of the MANOVA was seen to be a p value = 2.2e-16, which means that the above stated null hypothesis must be rejected. This implies that at least one feature, namely the price or delivery time, have an impact on the why bought feature. This statement can be further confirmed by analysing various plots of these features against one another.

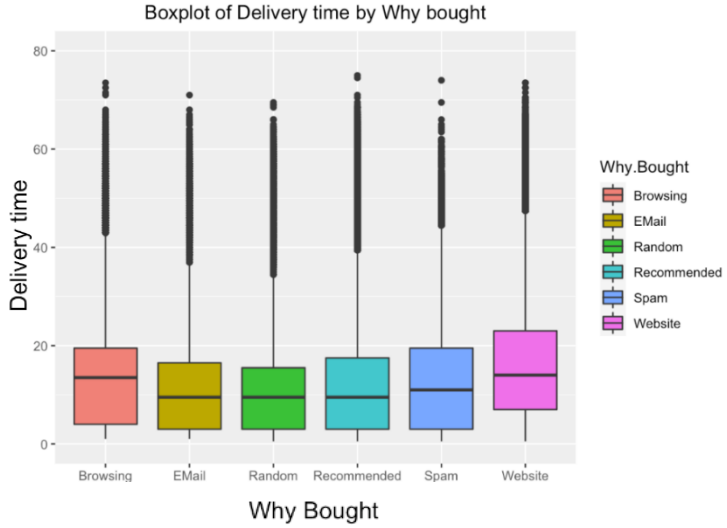


Figure 38: Graph of delivery times against Why Boughts

Figure 39: Graph of price against Why Boughts

Looking at the above distributions, it is now clear that the delivery time and price are not equally distributed across the different why boughts. This confirms the results of the MANOVA which resulted in the rejection of the null hypothesis.

Part 6: Reliability of the service and products

6.1 Lafrideradora

Problem 6:

L(X) = $k(x-T)^2$ but we know L = 45

45 = $k(0.04)^2$

k = 28125

Therefore, the Taguchi loss function becomes L(X) = $28125(x-T)^2$. Below is the plot of the loss function and it is noticeable that the larger the variation becomes, the larger the price becomes. This means the goal should be to reduce the variation as much as possible to maintain the cost. By doing this, customers will also have more stability with regards to their prices and will improve the company's reliability and maintain a good service level.

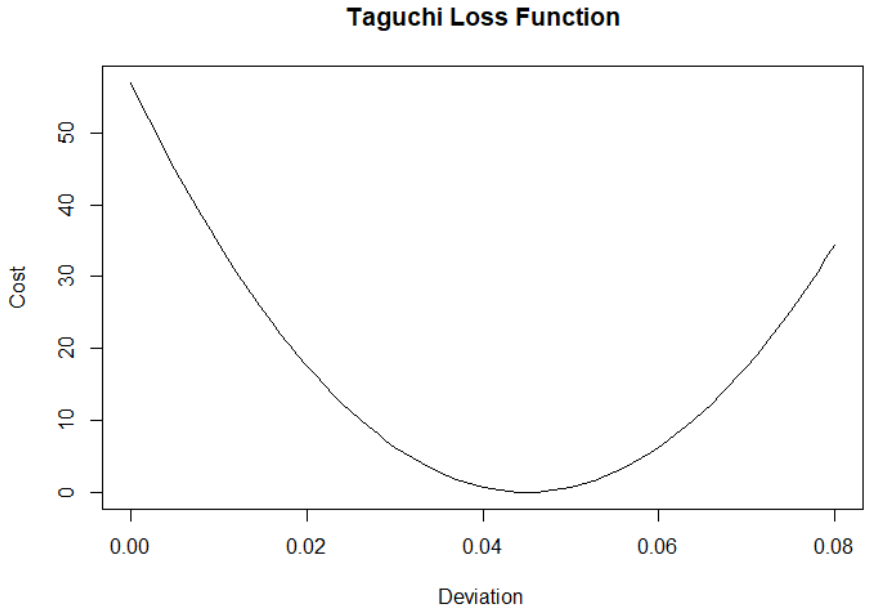


Figure 40: Graph depicting the loss function

Problem 7a & b:

The new loss is now 35

L(X) = $k(x-T)^2$ but we know L = 45

35 = $k(0.04)^2$

k = 21875

The reduced scrap price resulted in a final loss function of L(X) = $21875(x - T)^2$. If target deviation is reduced to 0.027cm we get a loss function value of L(x) = 3.696875. Below is the updated Taguchi loss function for the new scrap value.

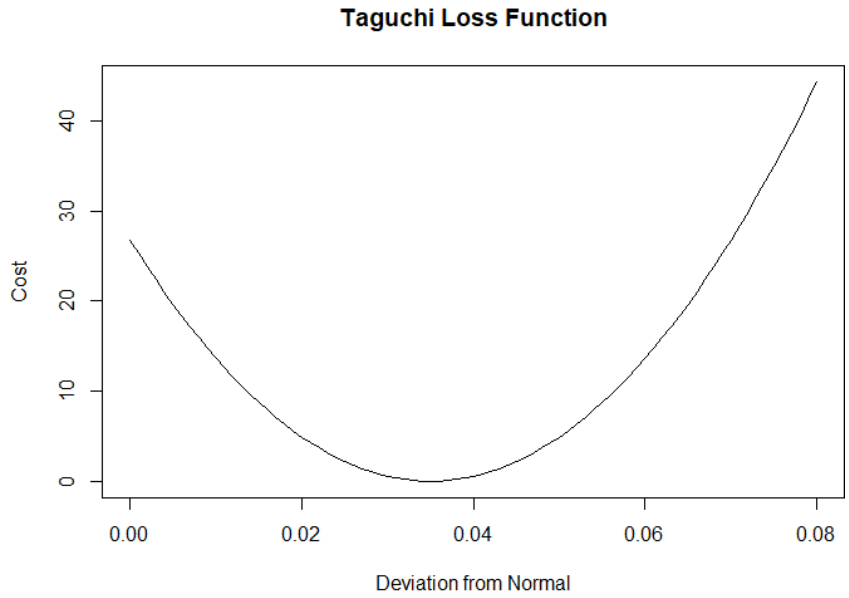


Figure 41: New loss function plot

6.2 Magnaplex problem

Problem 27a:

Rtotal = RA × RB × RC

Rtotal = 0.85 × 0.92 × 0.9

= 0.7038

= 70.38% reliability

Problem 27b:

Rparallel = [1 − (1 − RA)(1 − RA)] × [1 − (1 − RB)(1 − RB)] × [1 − (1 − RC)(1 − RC)]

= [1 − (0.15)(0.15)] × [1 − (0.08)(0.08)] × [1 − (0.1)(0.1)]

= 0.9775 × 0.9936 × 0.99

= 0.9615

= 96.15% reliability

This means that the reliability of the process increases by 25.773165% when two machines are used at each stage of the process.

6.3 Reliable delivery

Conclusion

The given data set was wrangled by using base function to remove unwanted instances and resulted in a clean and valid data set which was used for an in-depth analysis. Various methods and plots were used to understand the data, outline important influences, and affects the of the different features on one another and make key connections. These methods resulted in recommendations that the company can use to improve.

The analysis found that important features like technology and luxury items are key aspects as they result in the bulk of the sales and therefore revenue. It was noted that technology has a delivery time that is in adequate control but requires focus to ensure this remains true. Luxury class was noted to be out of control, but with a decreasing delivery time. This was a positive finding as it is now known that the luxury class is one that the firm must take time to understand the reasons for the seen delivery time changes and attempt to implement this aspect throughout the firm. Classes such as household and gifts were noted as severely out of control and have been recommended for stringent investigation.

The MANOVA analysis has found that the delivery time as well as the price affected the why bought feature.

The items from the technology and luxury classes must be monitored to maintain a high level of service whereas it has been recommended that classes such as household and gifts must be adjusted or removed.

Reference list

www.sthda.com. (n.d.). *MANOVA Test in R: Multivariate Analysis of Variance - Easy Guides - Wiki - STHDA*. [online] Available at: http://www.sthda.com/english/wiki/manova-test-in-r-multivariate-analysis-of-variance.

Evans and Lindsay (n.d.). *Managing for Quality and Performance Excellence*. 11th Edition ed.

Kelleher, J., Mac Namee, B. and D'Arcy, A., 2015. Fundamentals of Machine Learning for Predictive Data Analytics. MIT Press