Francois Schoonbee

23535121

Quality Assurance 344

ECSA Project

Table of Contents

Table of Figures

Part 1: Data wrangling

Before any data analysis can be done, the data must first be prepared. To do this, the negative and missing values must be removed. Initially the sales data contained 180 000 instances with 10 descriptive features: X, ID, Age, Class, Price, Year, Month, Day, Delivery time and Why Bought.

X <int>	ID <int>	A... <int>	Class <chr>	Price <dbl>	Year <int>	Month <int>	Day <int>	Delivery.time <dbl>	Why.Bought <chr>
1	19966	54	Sweets	246.21	2021	7	3	1.5	Recommended
2	34006	36	Household	1708.21	2026	4	1	58.5	Website
3	62566	41	Gifts	4050.53	2027	8	10	15.5	Recommended
4	70731	48	Technology	41843.21	2029	10	22	27.0	Recommended
5	92178	76	Household	19215.01	2027	11	26	61.5	Recommended
6	50586	78	Gifts	4929.82	2027	4	24	14.5	Random
7	73419	35	Luxury	108953.53	2029	11	13	4.0	Recommended
8	32624	58	Sweets	389.62	2025	7	2	2.0	Recommended
9	51401	82	Gifts	3312.11	2025	12	18	12.0	Recommended
10	96430	24	Sweets	176.52	2027	11	4	3.0	Recommended
11	87530	33	Technology	8515.63	2026	7	15	21.0	Browsing
12	14607	64	Gifts	3538.66	2026	5	13	13.5	Recommended
13	24299	52	Technology	27641.97	2024	5	29	17.0	Browsing
14	77795	92	Food	556.83	2025	6	3	3.0	Random
15	62567	73	Clothing	347.99	2024	3	29	8.5	Website
16	14839	47	Technology	54650.41	2027	12	30	18.5	Recommended
17	96208	44	Technology	14739.09	2028	3	17	13.0	Recommended
18	39674	69	Technology	22315.17	2026	8	20	20.5	Recommended
19	98694	74	Sweets	546.48	2025	5	9	2.0	Recommended
20	99187	54	Luxury	81620.21	2027	9	14	3.0	Recommended
21	59365	72	Gifts	3314.76	2028	4	30	13.0	Recommended

1-21 of 180,000 rows Previous 1 2 3 4 5 6 ... 48 Next

Figure 1: Sales Data

After removing the missing and negative values the contained 179 978 entries (meaning 22 entries were removed).

Primary Key <int>	X <int>	ID <int>	AGE <int>	Class <chr>	Price <dbl>	Year <int>	Month <int>	Day <int>	Delivery.time <dbl>	Why.Bought <chr>
1	1	19966	54	Sweets	246.21	2021	7	3	1.5	Recommended
2	2	34006	36	Household	1708.21	2026	4	1	58.5	Website
3	3	62566	41	Gifts	4050.53	2027	8	10	15.5	Recommended
4	4	70731	48	Technology	41843.21	2029	10	22	27.0	Recommended
5	5	92178	76	Household	19215.01	2027	11	26	61.5	Recommended
6	6	50586	78	Gifts	4929.82	2027	4	24	14.5	Random
7	7	73419	35	Luxury	108953.53	2029	11	13	4.0	Recommended
8	8	32624	58	Sweets	389.62	2025	7	2	2.0	Recommended
9	9	51401	82	Gifts	3312.11	2025	12	18	12.0	Recommended
10	10	96430	24	Sweets	176.52	2027	11	4	3.0	Recommended
11	11	87530	33	Technology	8515.63	2026	7	15	21.0	Browsing
12	12	14607	64	Gifts	3538.66	2026	5	13	13.5	Recommended
13	13	24299	52	Technology	27641.97	2024	5	29	17.0	Browsing
14	14	77795	92	Food	556.83	2025	6	3	3.0	Random
15	15	62567	73	Clothing	347.99	2024	3	29	8.5	Website
16	16	14839	47	Technology	54650.41	2027	12	30	18.5	Recommended
17	17	96208	44	Technology	14739.09	2028	3	17	13.0	Recommended
18	18	39674	69	Technology	22315.17	2026	8	20	20.5	Recommended
19	19	98694	74	Sweets	546.48	2025	5	9	2.0	Recommended
20	20	99187	54	Luxury	81620.21	2027	9	14	3.0	Recommended
21	21	59365	72	Gifts	3314.76	2028	4	30	13.0	Recommended

1-21 of 179,978 rows Previous 1 2 3 4 5 6 ... 48 Next

Figure 2: Valid Data

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

1-21 of 22 rows Previous 1 2 Next

Figure 3: Invalid Data

For both data sets a primary key was added to keep track of the data. It can be observed that of the data that was removed that all the problems were found in the price column. This shows that the data collection process for the price of the object has some issues, therefore that data collection method of the price of the object must be improved.

Part 2: Descriptive statistics

Data analysis is systematically application of logical or statistical techniques to illustrate and describe, compress and recap, and evaluate data (ori.hhs.gov, 2022)

2.1 Data Analysis

First, the type of data was analysed:

```
tibble [179,978 x 11] (S3: tbl_df/tbl/data.frame)
 $ Primary Key  : int [1:179978] 1 2 3 4 5 6 7 8 9 10 ...
 $ X            : int [1:179978] 1 2 3 4 5 6 7 8 9 10 ...
 $ ID           : int [1:179978] 19966 34006 62566 70731 92178 50586 73419 32624 51401 96430 ...
 $ AGE          : int [1:179978] 54 36 41 48 76 78 35 58 82 24 ...
 $ Class        : chr [1:179978] "Sweets" "Household" "Gifts" "Technology" ...
 $ Price        : num [1:179978] 246 1708 4051 41843 19215 ...
 $ Year         : int [1:179978] 2021 2026 2027 2029 2027 2027 2029 2025 2025 2027 ...
 $ Month        : int [1:179978] 7 4 8 10 11 4 11 7 12 11 ...
 $ Day          : int [1:179978] 3 1 10 22 26 24 13 2 18 4 ...
 $ Delivery.time: num [1:179978] 1.5 58.5 15.5 27 61.5 14.5 4 2 12 3 ...
 $ Why.Bought   : chr [1:179978] "Recommended" "Website" "Recommended" "Recommended" ...
```

Figure 4: Data Feature Types

All the features are numerical except Class and Why Bought, which are characters. This means that calculations can be done with all the features except Class and Why Bought.

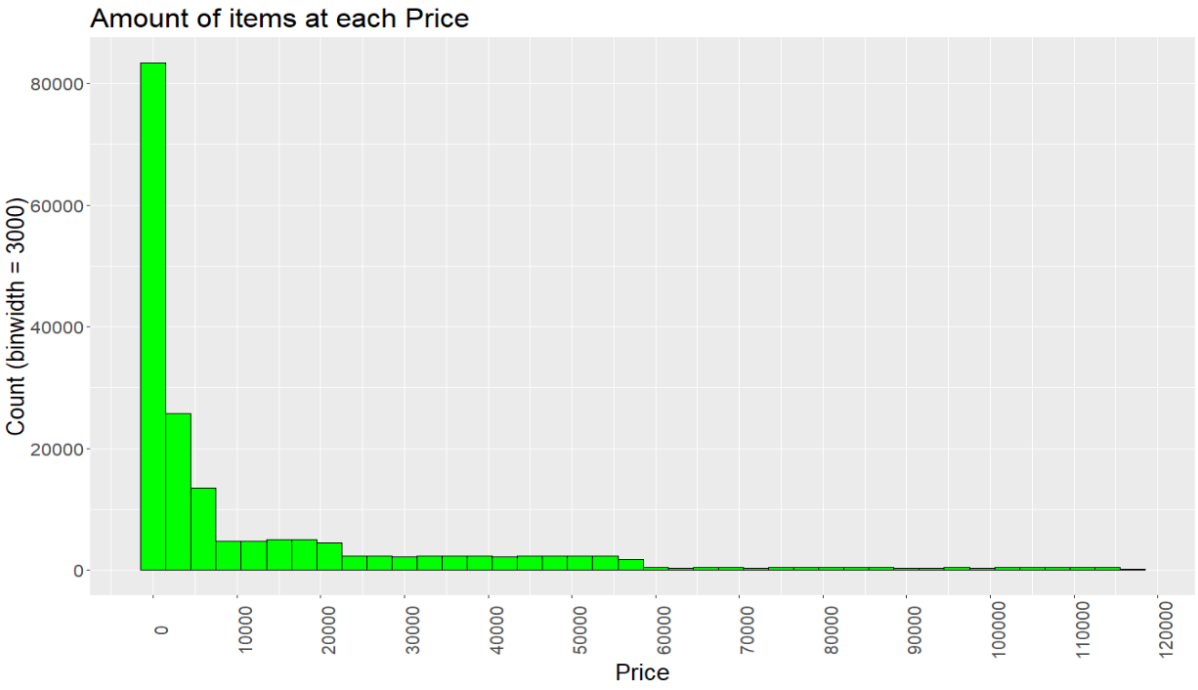

Figure 5: Histogram of item Prices

This graph is a histogram of the number of items sold at different price points. It can be observed that most of the items sold for between 0 and 5 000, followed by a relatively uniform distribution between 9 000 and 54 000 and relatively few were sold at a price higher than 60 000. To find out why, this will be further analysed. First, seeing as it is clear that the first column holds most of the items, the binwidth must be decreased to better understand the price distribution.

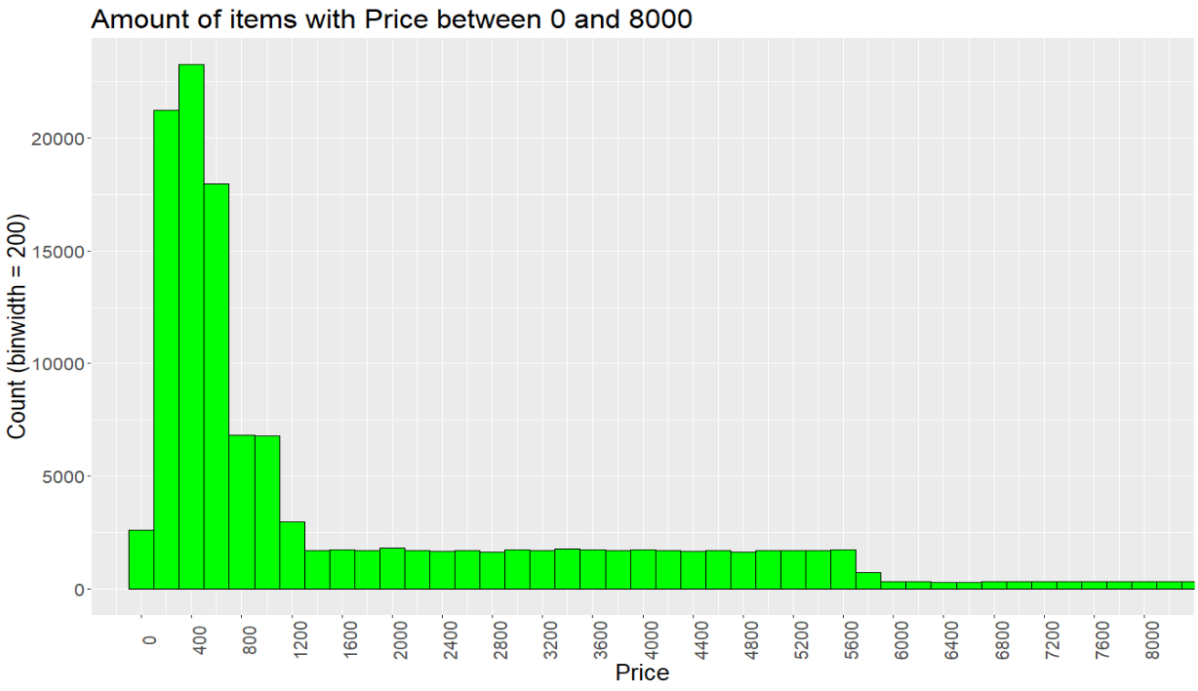

Figure 6: Histogram of item Prices between 0 and 8 000

This histogram shows the number of items sold between 0 and 8 000. Once again most of the items are between 200 and 600.

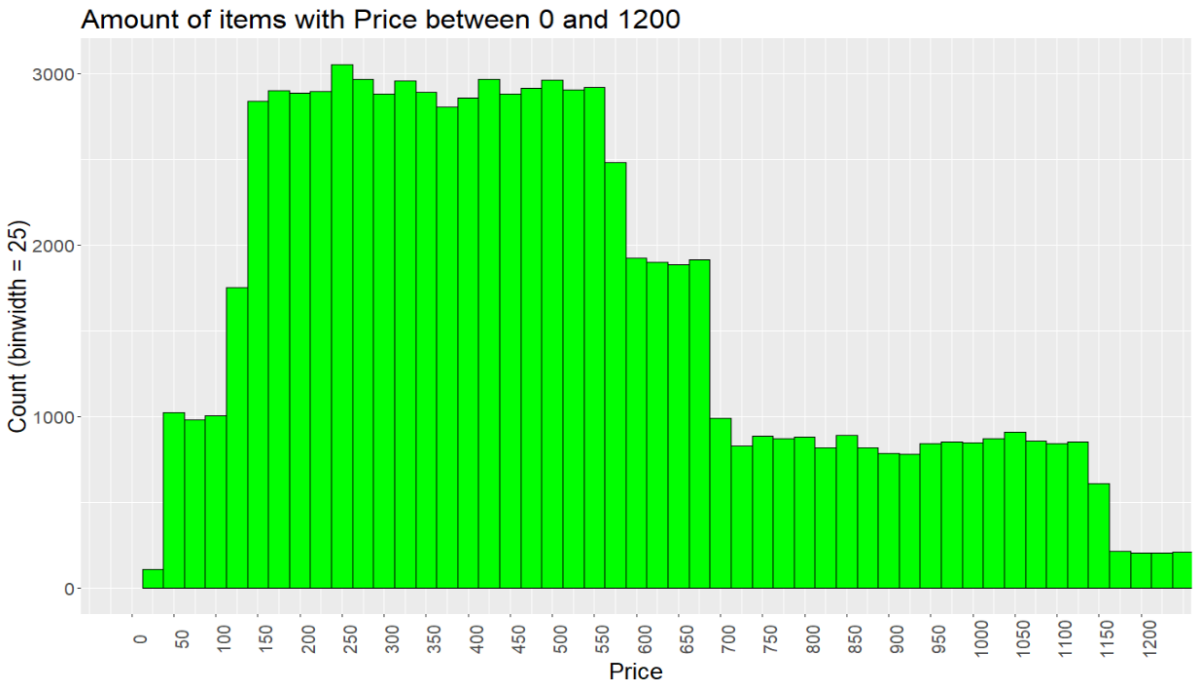

Figure 7: Histogram of item Prices between 0 and 1200

This histogram further shows that most of the items were sold for between 150 and 650.

This distribution will be further analysed to understand what causes this unevenness in the data.

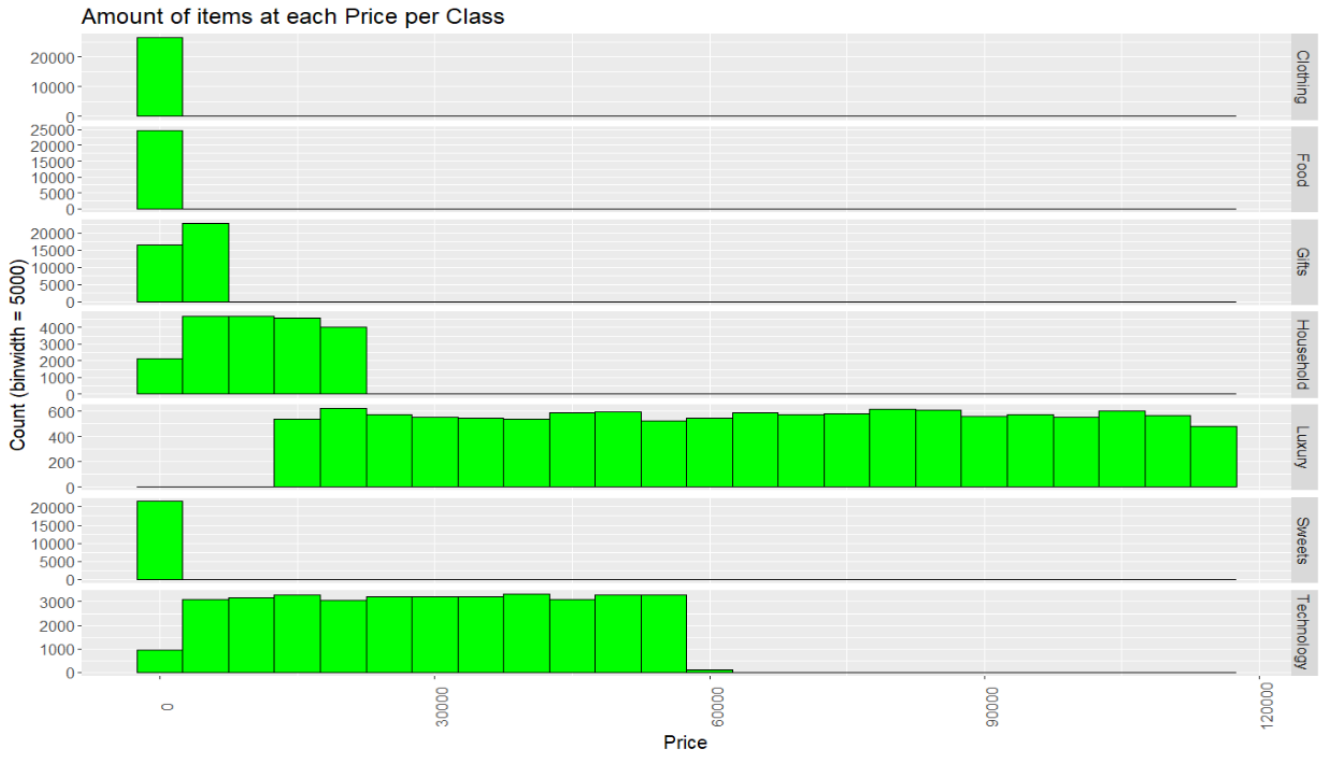

Figure 8: Histograms of item Prices for each Class

These histograms plot the number of items sold at each price point for each Class. There is a correlation between the class and the price it sold at. For Clothing, Food, Gifts and Sweets, most of the items were sold at lower price points. Household items were sold for a bit higher prices, while both Technology and Luxury have a large uniform distribution of prices.

For classes Clothing, Food, Gifts and Sweets a histogram will be plotted with a smaller binwidth to better understand the price distribution.

Figure 9: Histogram of item Prices for Clothing

This histogram shows that the Price of Clothing items have a uniform distribution between 0 and 1200.

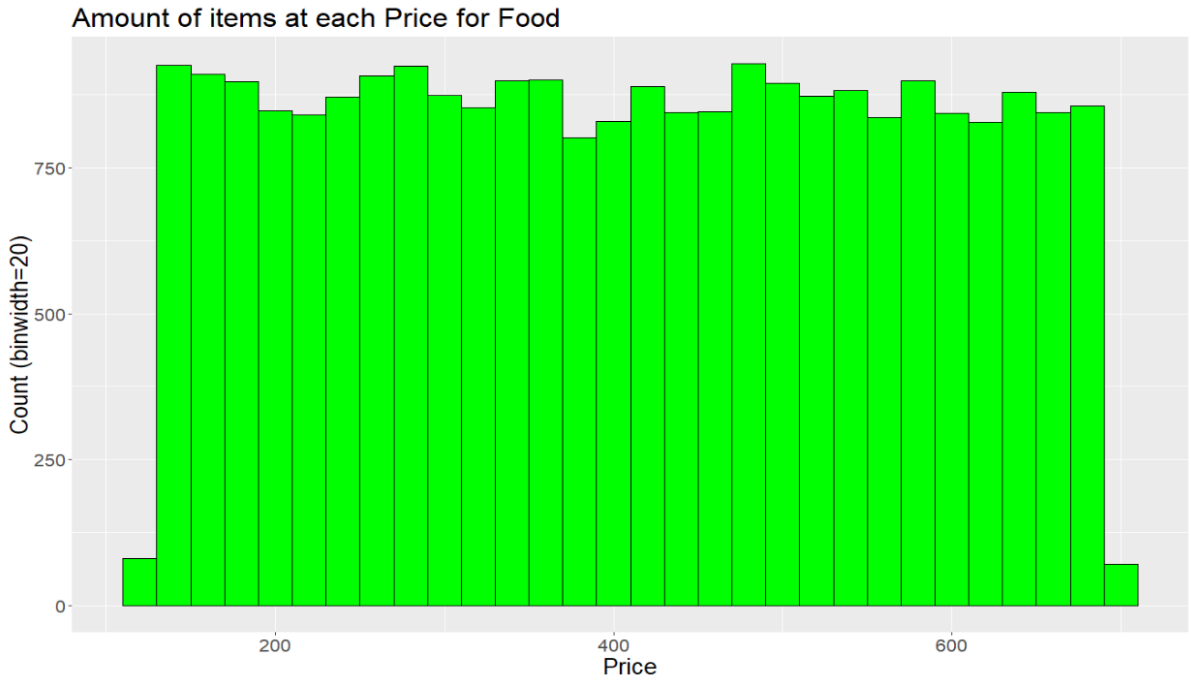

Figure 10: Histogram of item Prices for Food

This histogram shows that the Price of Food items have a uniform distribution between 20 and 680. With a few items being sold for both more and less than that.

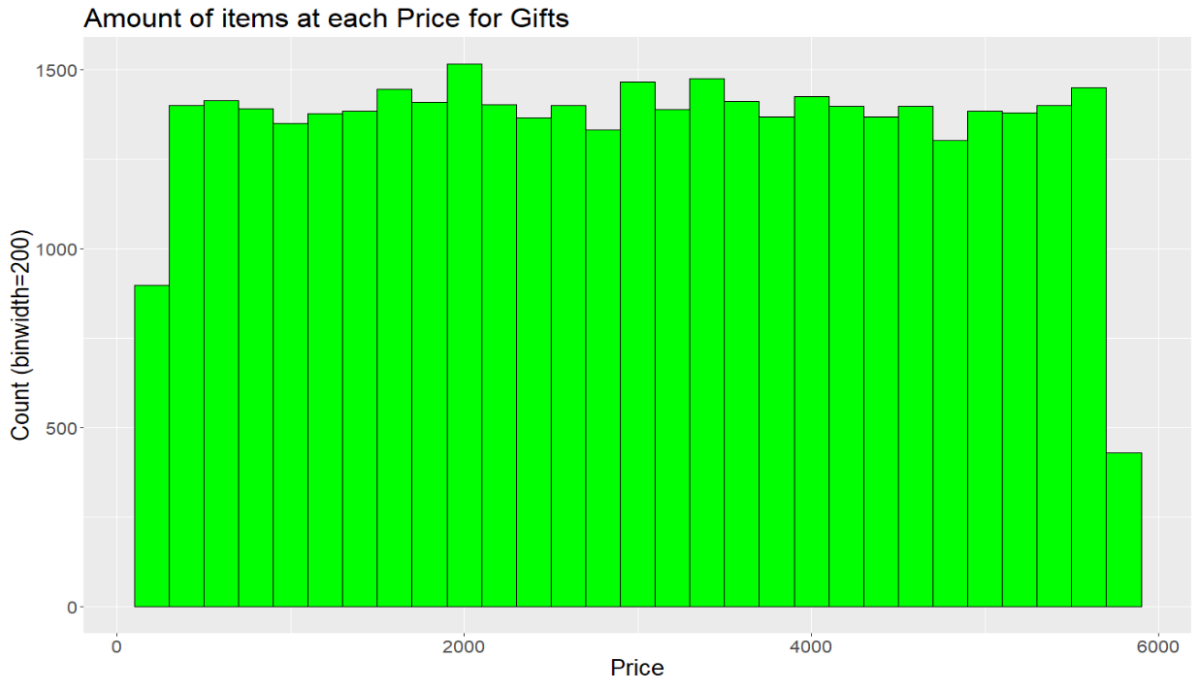

Figure 11: Histogram of item Prices for Gifts

This histogram shows that the Price of Gifts items have a uniform distribution between 0 and 6000.

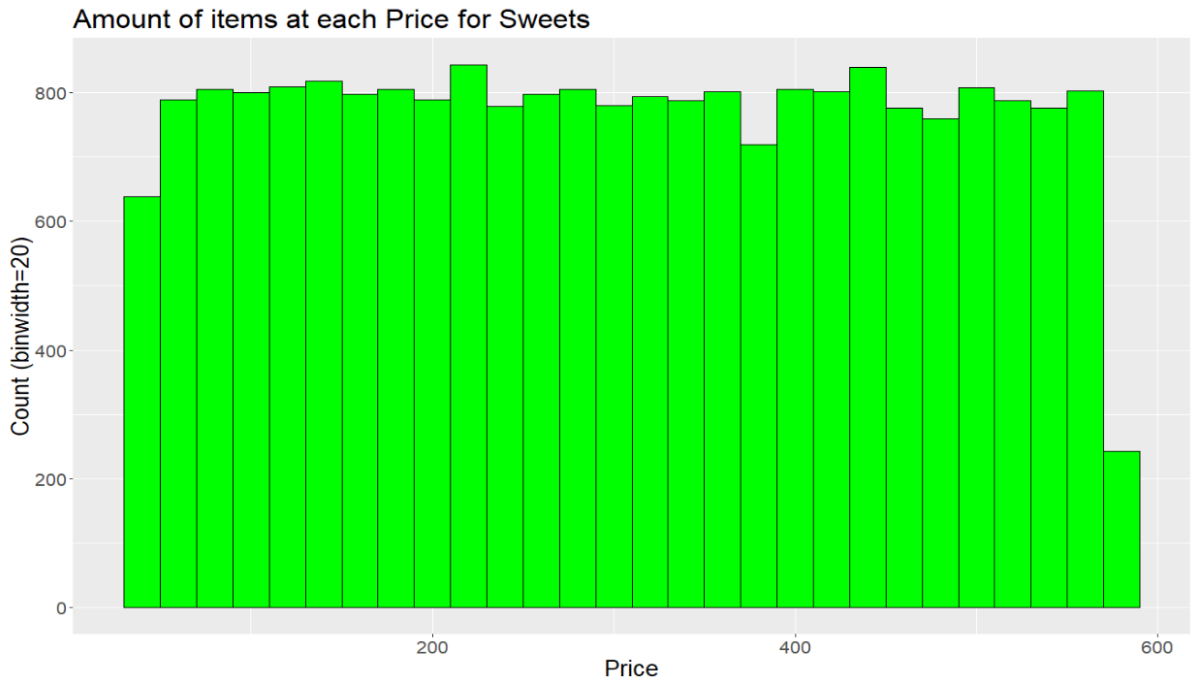

Figure 12: Histogram of item Prices for Gifts

This histogram shows that the Price of Clothing items have a uniform distribution between 0 and 600.

Class <chr>	Max_Price <dbl>	Min_Price <dbl>	Range <dbl>	Median_Price <dbl>	Mean_Price <dbl>	Count <int>	Sum_Price <dbl>
Clothing	1154.02	127.76	1026.26	642.04	640.5253	26403	16911791
Food	691.96	127.76	564.20	408.37	407.8153	24582	10024915
Gifts	5774.49	172.61	5601.88	2961.59	2961.8414	39149	115953130
Household	21935.33	127.76	21807.57	10960.88	11009.2738	20065	220901079
Luxury	116618.97	12825.37	103793.60	65342.14	64862.6386	11868	769789795
Sweets	576.38	35.65	540.73	303.25	304.0704	21564	6556974
Technology	57735.40	935.18	56800.22	29653.90	29508.0626	36347	1072529553

Figure 13: Analysis Table of Price for each Class

This table show the maximum, minimum, range, median, mean, amount and sum of the Price for each Class of items. Luxury items have the largest rage while Sweets and Food items have a small range. Furthermore, Luxury items sold on average for at the highest Price, followed by Technology, Household, Gifts, Clothing, Food then Sweets. Gifts sold the greatest number of items while Luxury sold the fewest. Lastly, Technology brought in the most money followed by Luxury, Household, Gifts, Clothing, Food then Sweets. Although, the cost of goods sold is not available here, it might be beneficial to focus more on technology items seeing as they bring in more revenue. Furthermore, not many Luxury items were sold therefore, it might be beneficial to spend more money on advertising for Sweets items.

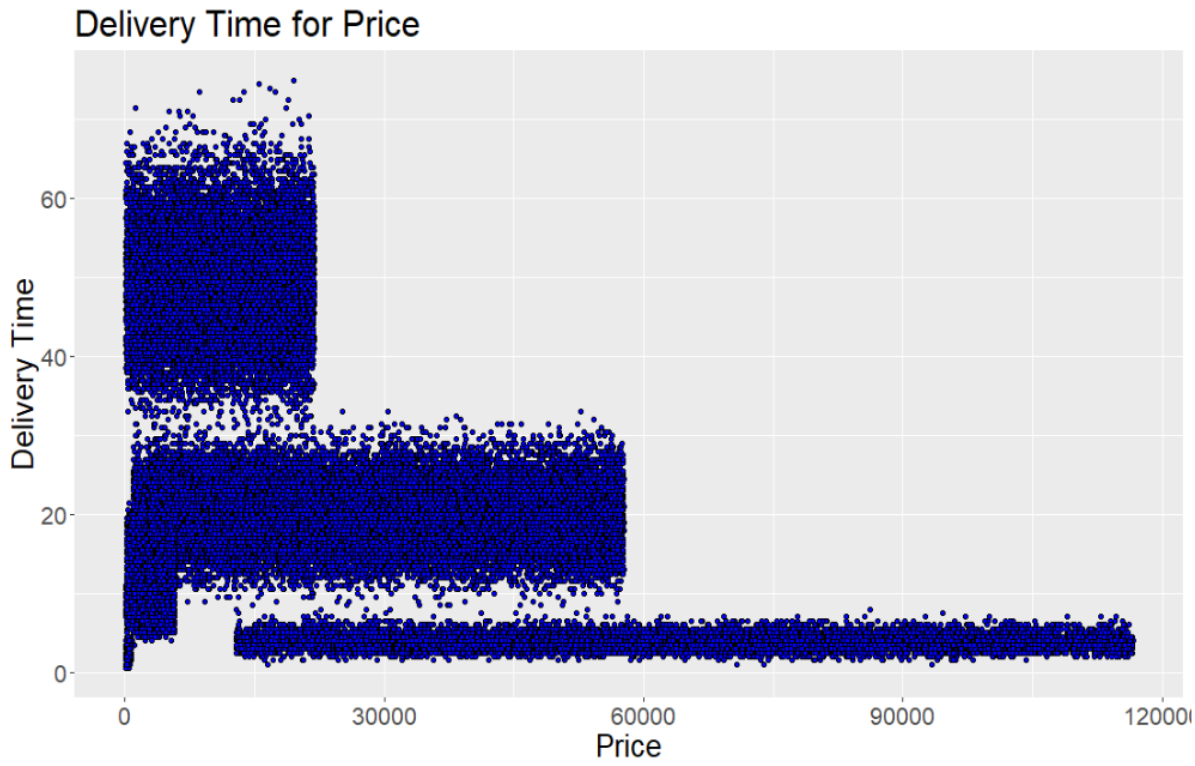

Figure 14: Graph of Delivery Time and Price

This point plot plots items sold at each Price compared to their Delivery Time. This graph shows that there might be a correlation between Delivery Time and Price, however it is not obvious to see.

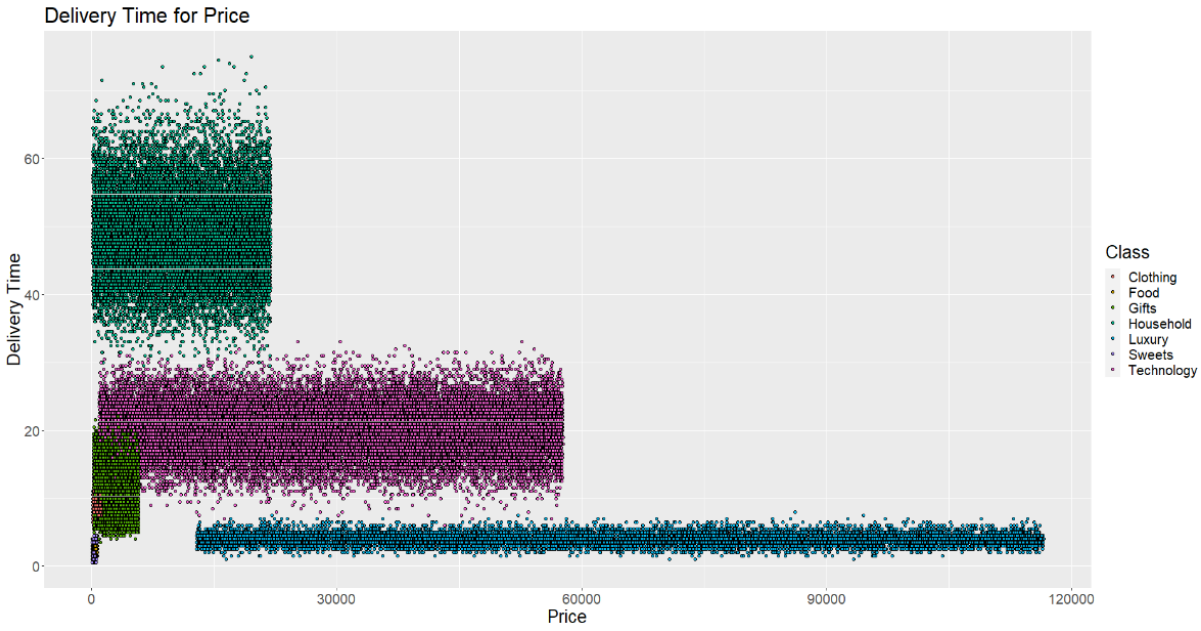

Figure 15: Graph of Delivery Time and Price with colours to identify Class

This graph is the same one as before except each Class of items have a different colour. It can be observed that each class of items is grouped together. Therefore, a plot will be made that plots Delivery Time and Price for each individual Class.

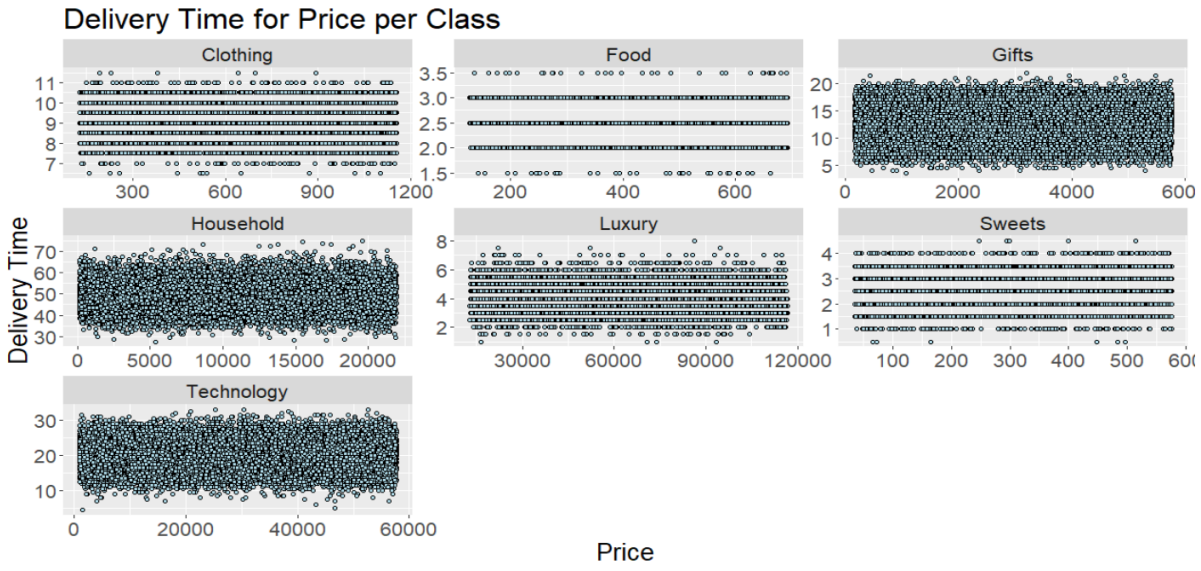

Figure 16: Graphs of Delivery Time and Price for each Class

These graphs plot the Delivery Time to Price for each Class. By studying the y-axes, it can be observed that Food, Luxury and Sweets have a low Delivery Time, Gifts, Clothing and Technology

slightly higher and Household items have higher Delivery Times. It can also be observed that there is no correlation between Price and Delivery Time.

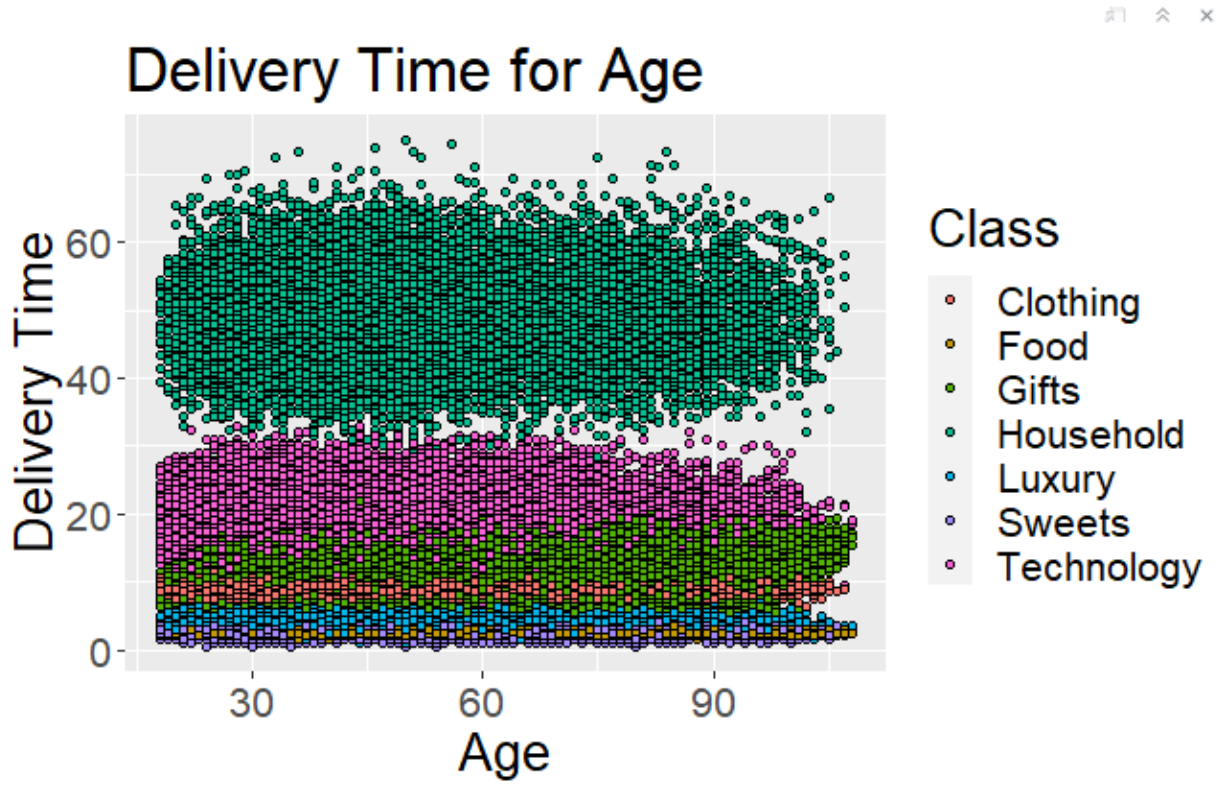

Figure 17: Graph of Delivery Time and Age

This graph shows the Delivery Time compared to Age. Even though each Class is a different colour, it can be observed that there is no correlation between Delivery Time and Age.

Class <chr>	Max_Delivery.time <dbl>	Min_Delivery.time <dbl>	Range <dbl>	Median_Delivery.time <dbl>	Mean_Delivery.time <dbl>	Count <int>	Sum_Delivery.time <dbl>
Clothing	11.5	6.5	5.0	9.0	8.999527	26403	237614.5
Food	3.5	1.5	2.0	2.5	2.502014	24582	61504.5
Gifts	22.0	3.5	18.5	13.0	12.890546	39149	504652.0
Household	75.0	27.5	47.5	48.5	48.719561	20065	977558.0
Luxury	8.0	1.0	7.0	4.0	3.971520	11868	47134.0
Sweets	4.5	0.5	4.0	2.5	2.501206	21564	53936.0
Technology	33.0	4.5	28.5	20.0	20.010950	36347	727338.0

7 rows

Figure 18: Analysis Table of Delivery Time for each Class

This table show the maximum, minimum, range, median, mean, amount and sum of the Delivery Time for each Class of items. It can be observed that Sweets and Food have the lowest Delivery Time, followed by Luxury, Clothing, Gifts, Technology then Household. This shows that an effort should be made to lower the Delivery Time of Household items seeing as it is much higher than other items.

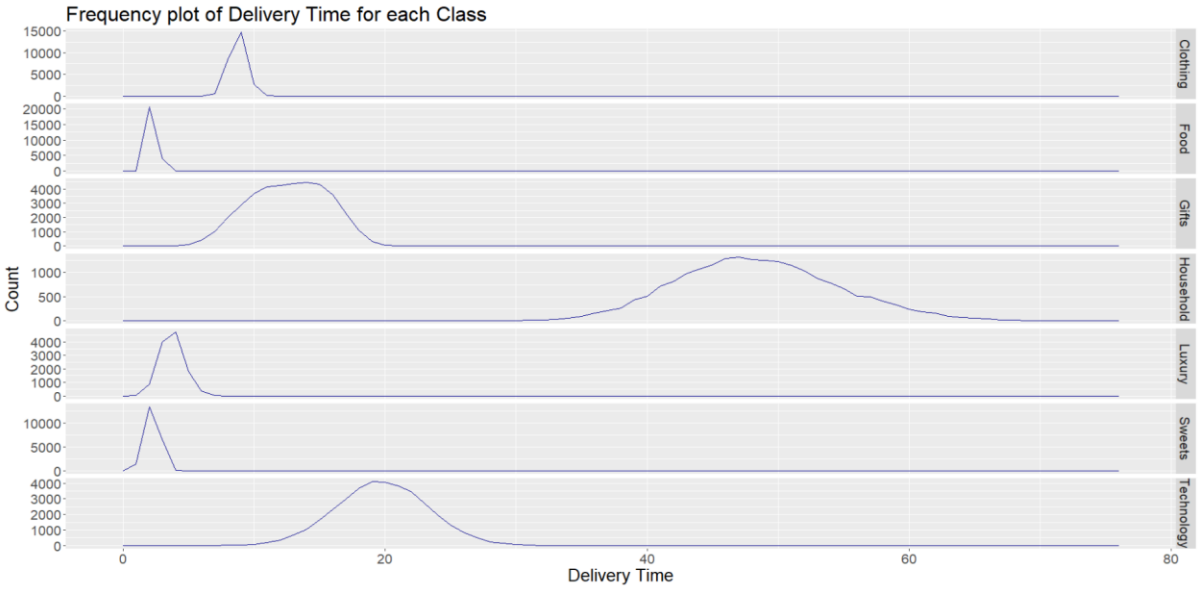

Figure 19: Frequency plot for Delivery Time for each Class

This frequency plot clearly shows the distribution of the Delivery Time for each Class.

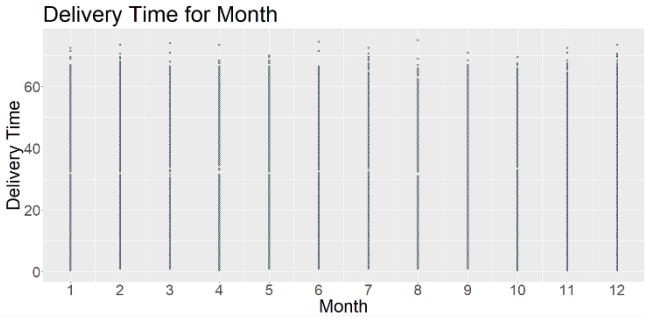

Figure 21: Graph of Delivery Time for each Month

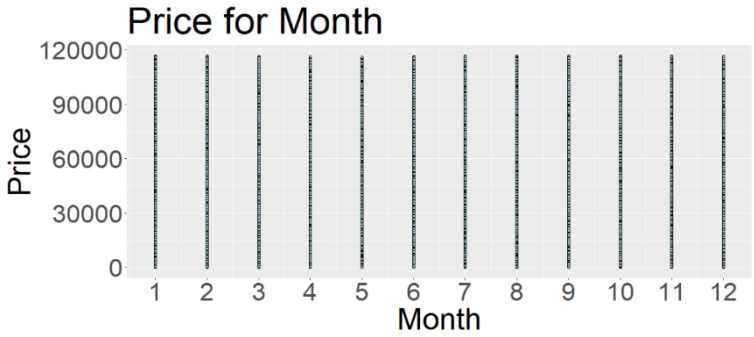

Figure 20:Graph of Price for each Month

The above graphs plot the Delivery Time for each Month and the Price for each Month. It can be observed that there is no correlation for both graphs.

Year <int>	mean_price <dbl>	Stan_dev <dbl>	count <int>
2021	8232.412	15023.02	33443
2022	12599.245	21132.28	15546
2023	12715.317	21385.75	17128
2024	13406.740	21938.89	17698
2025	13518.231	22308.36	17267
2026	13329.081	22009.43	17152
2027	13311.933	22006.25	18656
2028	13247.453	21863.87	20613
2029	13480.112	22337.36	22475

9 rows

Figure 22: Analysis table of Price for each Year

The above table shows the mean and standard deviation of the Price for each Year as well as the number of items sold in each year. It can be observed that the most items were sold in 2021, followed by a drop in sales which then slowly increased. This can be due to the increase in mean Price after 2021.

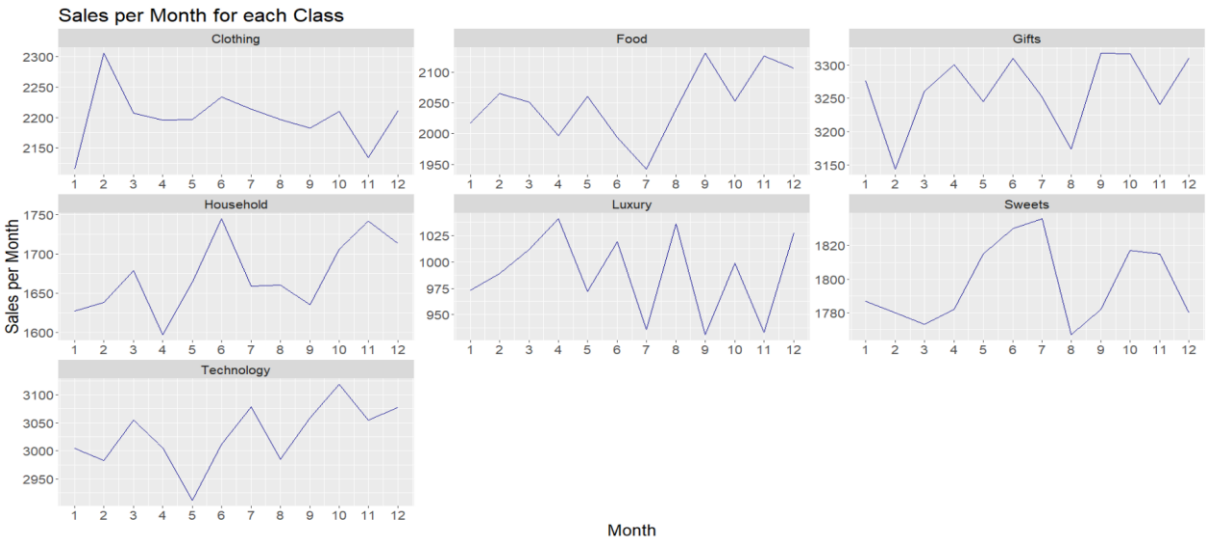

Figure 23: Graphs of sales per Month for each Class

These graphs show the Sales for each Month for each Class. This is to check if there is any seasonality in the different Class of items. It can be observed that no Class of items follow a clear seasonality, except for Sweets. The Sweets sales peak during June and July then a drop in sales in August, it peaks again in October and November, followed by a drop from January to March.

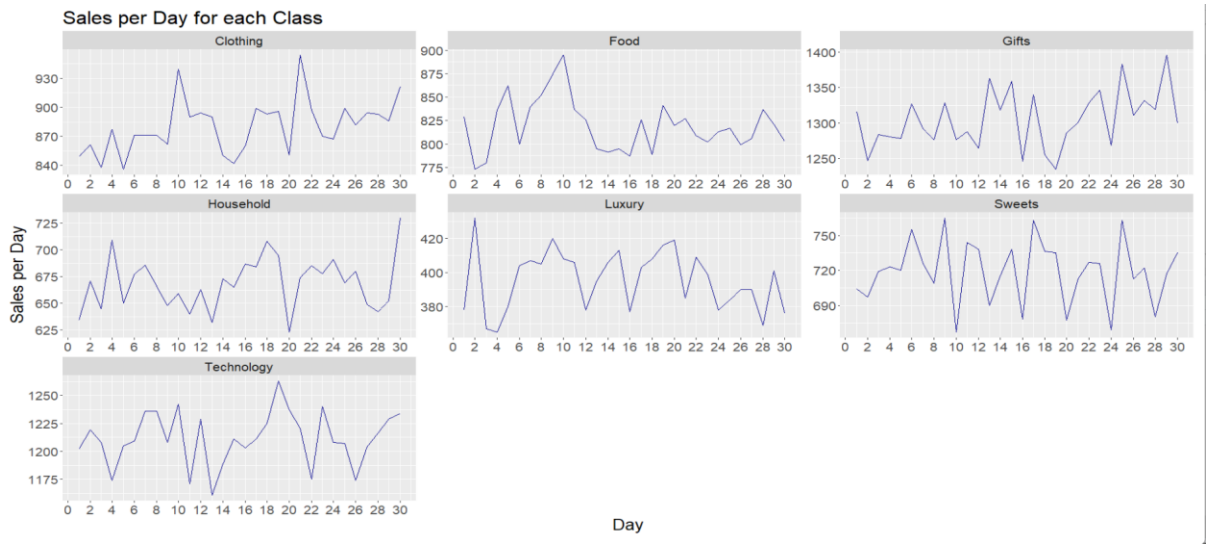

Figure 24: Graphs of number of sales per Day for each Class

These graphs plot the sales for each Day per Class. It can be observed that there is no correlation between number of sales and Day and follows a random distribution.

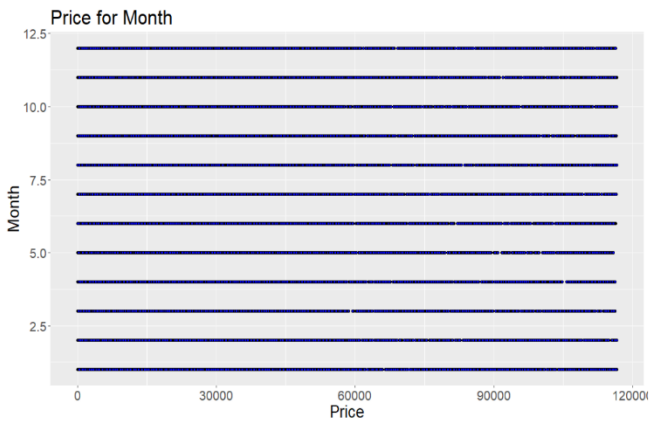

Figure 26: Graph of Price for Month

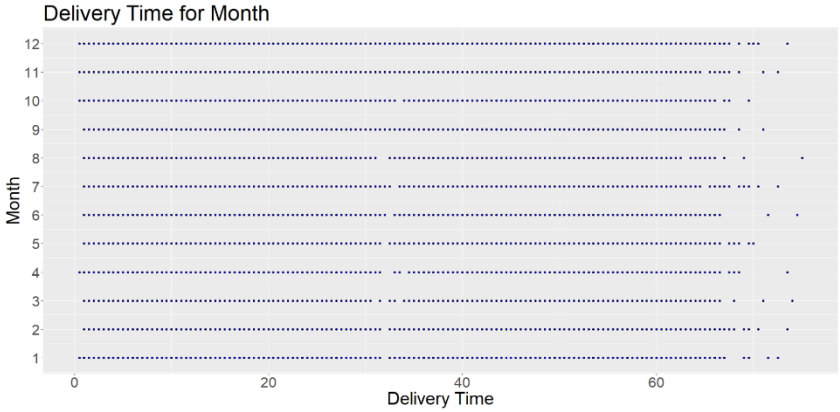

Figure 25: Graph of Delivery Time for Month

These graphs plot Price and Delivery Time with Month. In both graphs it can se seen that there is no correlation in both cases.

2.2 Process Capability

Process Capability is a measure that calculates how consistently a process can manufacture parts to be within specification. Cp measures whether the spread of the process is smaller the the width of the specifications. Cpl measures the capability of a process based on the LSL (lower specificaion limit). Cpu measures the capability of a process based on the USL (Upper specificaion limit) Cpk measures both the process spread in relation to the specification width and the centering of the manufacturing process. And equals the minimum of the Cpu and Cpl values. (1 Factory, 2022)

To calculate the Process capabilities for items of Class Technology, the specification limits (USL and LSL) were given to be 24 and zero respectively.

Cp <dbl>	Cpu <dbl>	Cpl <dbl>	Cpk <dbl>
1.14	0.38	1.9	0.38

Figure 27: Table of Process Capabilities

The above table shows what the process capabilities Cp, Cpk, and Cpl. The Cp of 1.14 indicates that the process is relatively capable. Generally a Cp of 1.33 indicates a capable process, therfore the consistency of Delivery Time must be improved. The Cpu of 0.380 is low and this indicates that the process capability is very poor is realtion to the upper specification limit (USL). The Cpl of 1.90 is high and indicates that in relation to the lower specification limit (LSL) the process is capable. The large diffrence between the Cpu and the Cpl indicates that the process is poorly centered (as seen in

the Cpk value of 0.380) and because Cpu < Cpl the process is more likely to produce products that exeed the upper specification limit (USL) of 24.

Part 3: Sampling

3.1 Control Charts

To analyse the Delivery Times of the data, the data was split into samples of 15 items each (containing only the value of Dilevery Time. First the data was oragnised into chronological order. Then the first 30 samples of 15 items (the first 450 items) for each individual Class were used to claculate their mean and standard deviation to calculate the different limits the Class has.

S-charts for first 30 samples

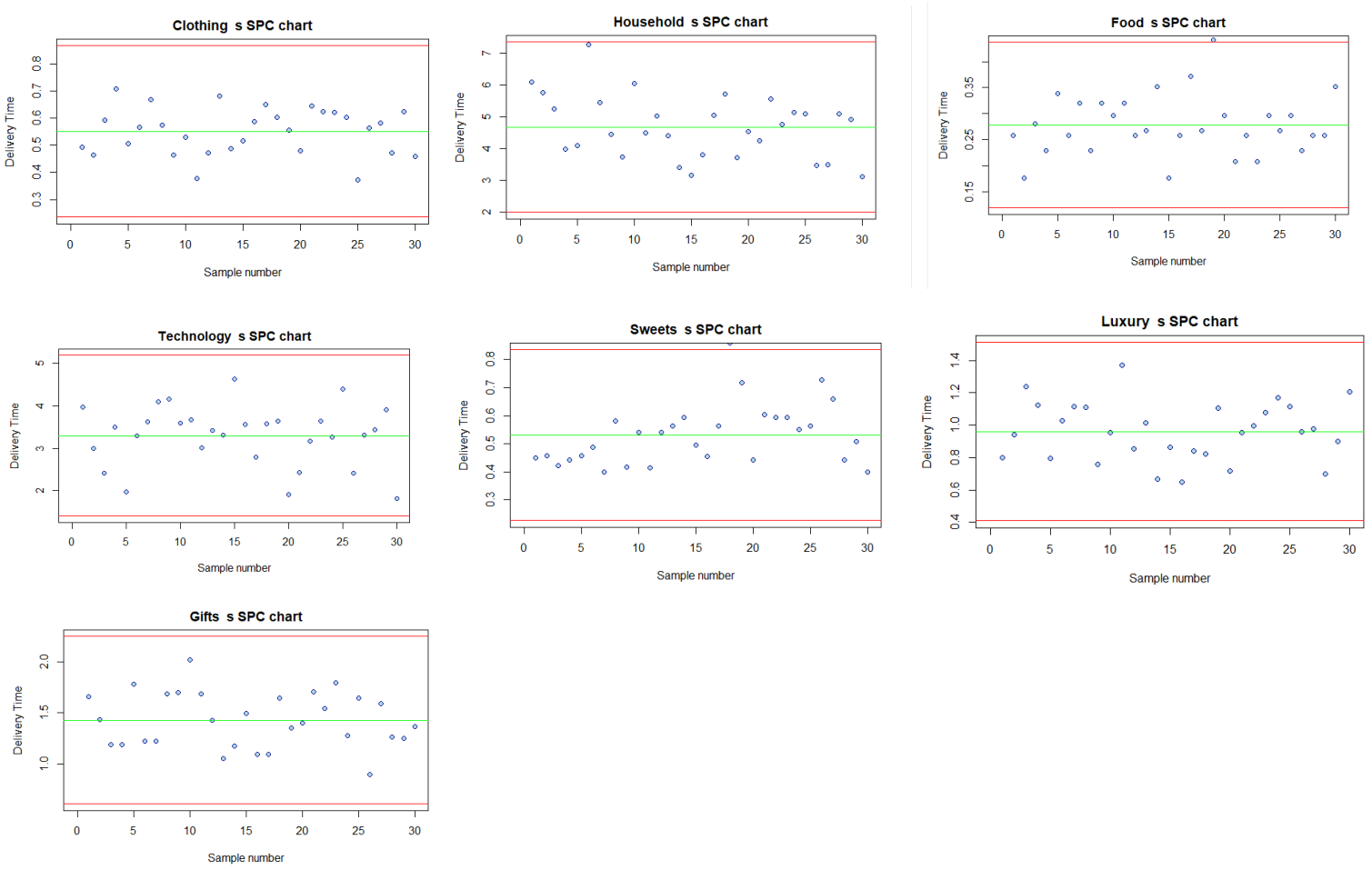

Figure 28: S-charts of first 30 samples for each Class

These charts plot the first 30 samples' standard deviation. The green line is the Control Limit (CL), the red line at the top is the Upper Control Limit (UCL) and the bottom red line is the Lower Control Limit (LCL) of the standard deviation of Delivery Time

X-charts for first 30 samples

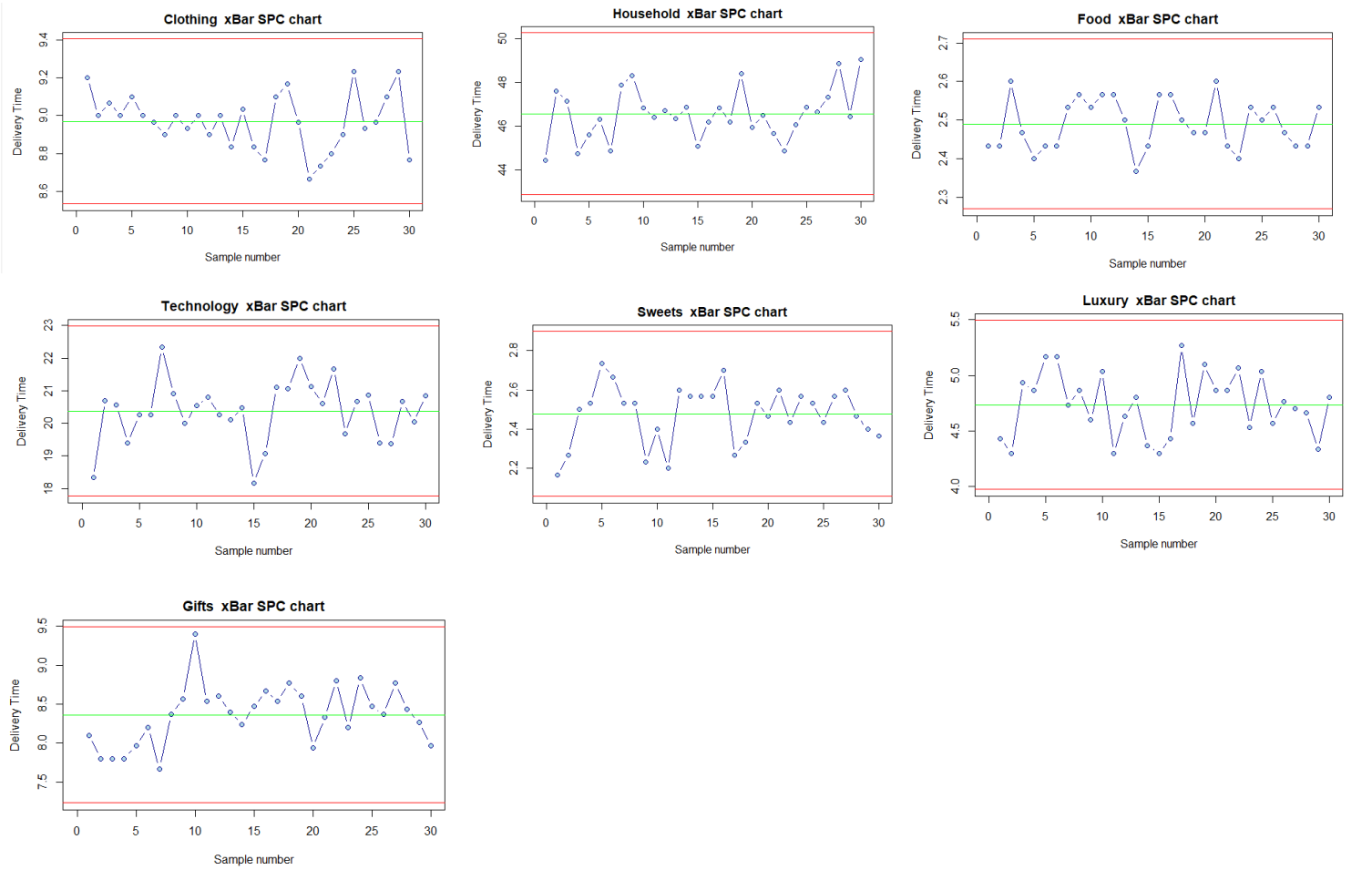

Figure 29: X-charts of first 30 samples for each Class

These charts plot the first 30 samples' mean. The green line is the Control Limit (CL), the red line at the top is the Upper Control Limit (UCL) and the bottom red line is the Lower Control Limit (LCL) of the mean Delivery Time.

S-charts for all samples

These charts plot the standard deviation of all the samples. The green line is the Control Limit (CL), the red line at the top is the Upper Control Limit (UCL) and the bottom red line is the Lower Control

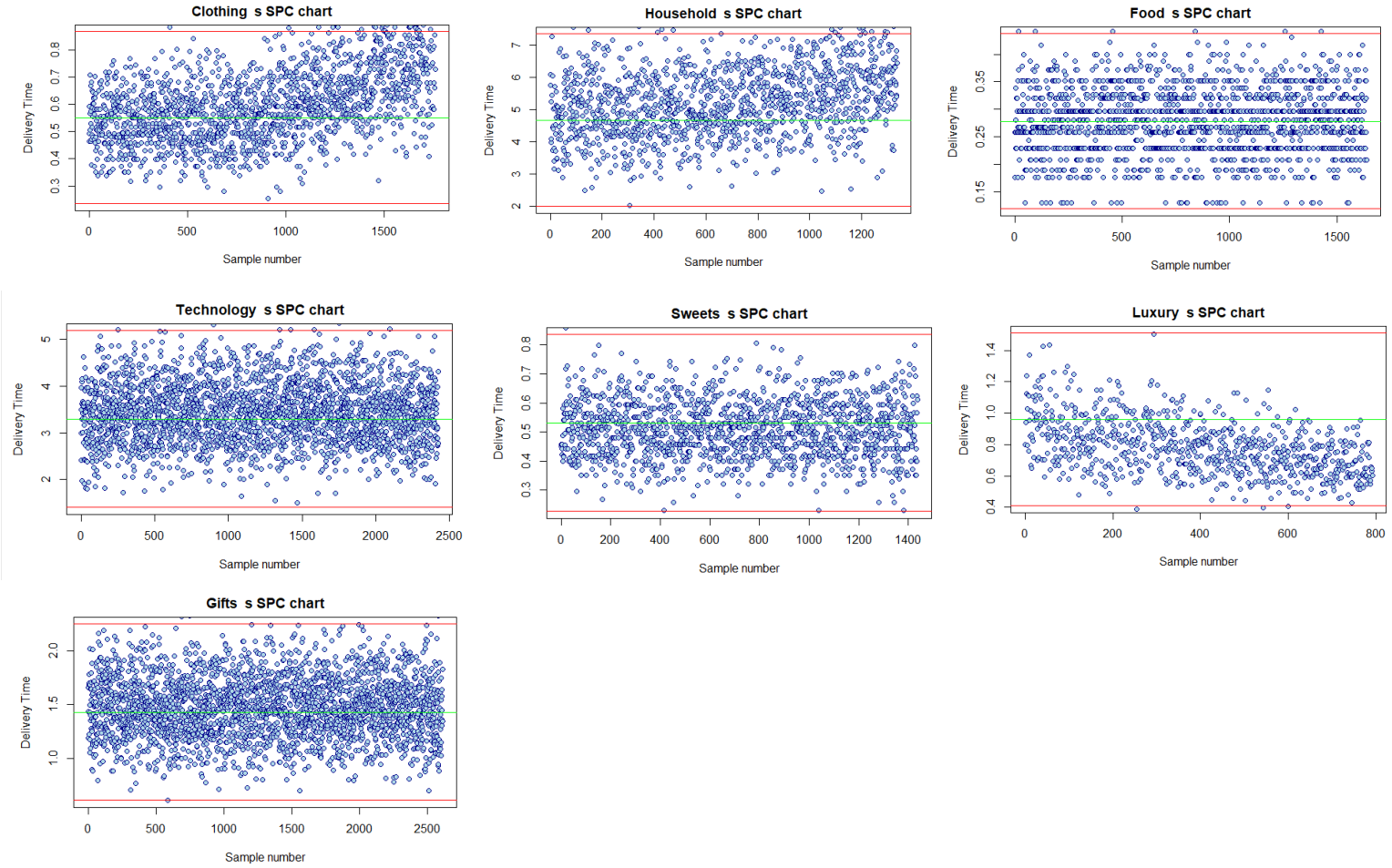

Figure 30: S-charts of all samples for each Class

Limit (LCL) of the standard deviation of Delivery Time.

X-charts for all samples

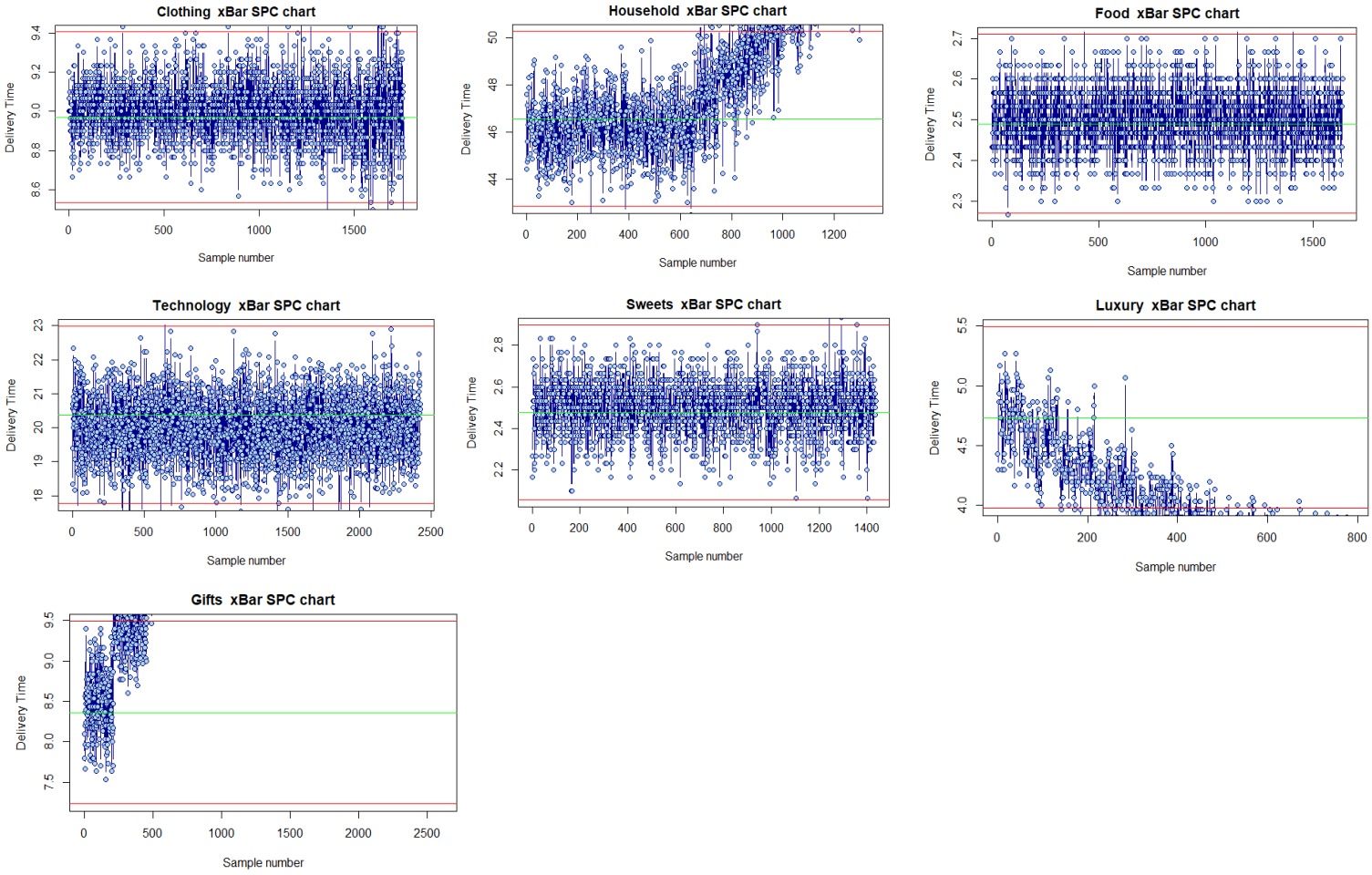

Figure 31: X-charts of all samples for each Class

These charts plot the mean of all the samples. The green line is the Control Limit (CL), the red line at the top is the Upper Control Limit (UCL) and the bottom red line is the Lower Control Limit (LCL) of the mean Delivery Time.

Class <chr>	UCLs <dbl>	U2Sigmas <dbl>	U1Sigmas <dbl>	CLs <dbl>	L1Sigmas <dbl>	L2Sigmas <dbl>	LCLs <dbl>
Clothing	0.8665596	0.7233611	0.6373038	0.5512465	0.4651893	0.3791320	0.2359335
Household	7.3441801	6.6537848	5.6628275	4.6718703	3.6809130	2.6899557	1.9995605
Food	0.4372466	0.3944580	0.3363023	0.2781467	0.2199910	0.1618354	0.1190468
Technology	5.1805697	4.7120218	4.0037748	3.2955278	2.5872808	1.8790338	1.4104859
Sweets	0.8353391	0.7496954	0.6405408	0.5313862	0.4222316	0.3130770	0.2274333
Gifts	2.2463333	1.9636522	1.6963087	1.4289652	1.1616217	0.8942782	0.6115971
Luxury	1.5110518	1.3241636	1.1426962	0.9612289	0.7797615	0.5982941	0.4114060

Figure 32: Table of the standard deviation of Delivery Time limits

Class <chr>	UCLx <dbl>	U2Sigmax <dbl>	U1Sigmax <dbl>	CLx <dbl>	L1Sigmax <dbl>	L2Sigmax <dbl>	LCLx <dbl>
Clothing	9.404934	9.261041	9.115521	8.970000	8.824479	8.678959	8.535066
Household	50.248328	48.926693	47.744458	46.562222	45.379987	44.197751	42.876117
Food	2.709458	2.620016	2.555008	2.490000	2.424992	2.359984	2.270542
Technology	22.974616	22.252544	21.313494	20.374444	19.435394	18.496345	17.774273
Sweets	2.897042	2.770909	2.624343	2.477778	2.331212	2.184647	2.058514
Gifts	9.488565	9.124865	8.742988	8.361111	7.979234	7.597357	7.233658
Luxury	5.493965	5.307688	5.021622	4.735556	4.449489	4.163423	3.977146

Figure 33: Table of the mean of Delivery Time limits

Above the tables containing the Upper Control Limit (UCL), Upper 2 sigma limit (U2 Sigma), Upper 1 sigma limit (U1 Sigma), Control Limit (CL), Lower 1 sigma limit (L1 Sigma), Lower 2 sigma limit (L2 Sigma) and Lower Control Limit (LCL) of the standard deviation (labelled -s) and the mean (labelled -x) of the Delivery Time for each Class.

Part 4: Optimising the delivery processes

4.1 Out of control samples

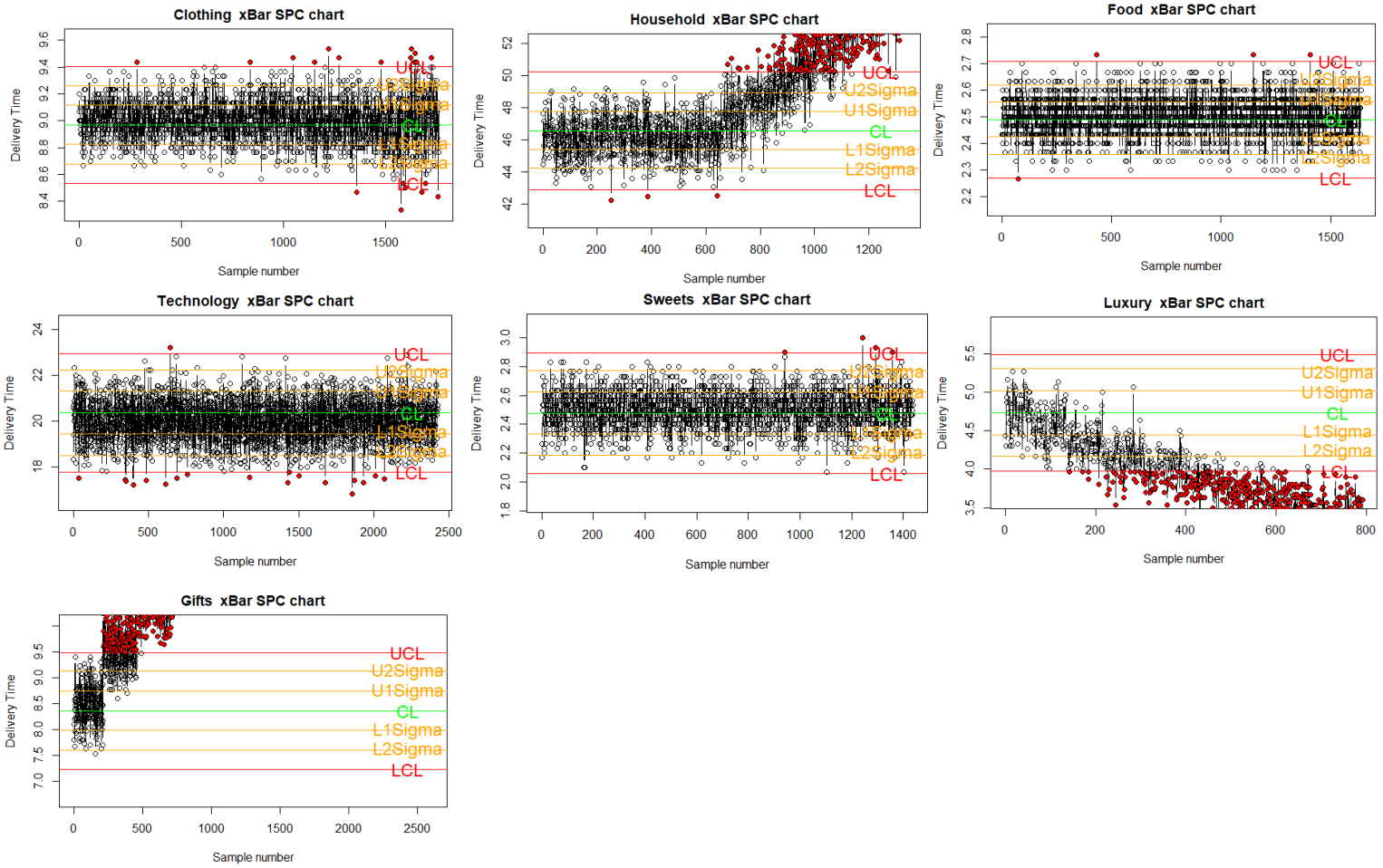

Figure 34: X-charts of each class with outliers highlighted

These charts plot the mean Delivery Times of all the samples. All the values that are outside of the outer control limits (UCL and LCL) have been highlighted in red.

4.1 A. one X-bar outside limits

Class	Total found	First	Second	Third	Third Last	Second Last	Last
Clothing	20	282	837	1048	1695	1723	1756
Household	395	252	387	643	1335	1336	1337
Food	4	75	432	1149	1408	*NA*	*NA*
Technology	19	37	345	353	1933	2009	2071
Sweets	4	942	1243	1294	1358	*NA*	*NA*
Gifts	2287	213	216	218	2607	2608	2609
Luxury	440	142	171	184	789	790	791

Figure 35: Table of values outside Control Limits

The table above shows the first 3 and last three samples that are outside of the outer control limits as well the number of samples found for each Class.

4.1 B. -0.3 and +0.4 sigma-control limits

Class Names <chr>	Consecutive <int>	Ending Sample <dbl>
Clothing	5	1579
Household	4	1017
Food	4	986
Technology	2	114
Sweets	2	175
Gifts	5	2484
Luxury	10	778

Figure 36: Table of consecutive values outside limits

The table above shows the most consecutive values that lie outside of -0.3 and +0.4 sigma-control limits as well as the ending sample number.

4.2 Type 1 error

A type 1 error (false positive) occurs when a null hypothesis is rejected but is actually true while a type 2 error (false negative) accepts a null hypothesis that is actually false. (Amitav Banerjee, 2009) In this case the H_0 is "the process is in control and centred on the centreline calculated using the first 30 samples".

A) For question A the probability of making a type 1 error is 0.002699796. This means that there is a 0.27% chance that the H_0 is rejected but it should have been accepted. There is a 0.27% chance that if the process is determined to be out of control and not centred on the centreline that is actually in control and centred. This is low and is very unlikely, therefore if the H_0 is rejected it should be rejected.

B) For question B the probability of making a type 1 error is 0.03751032. This means that there is a 3.75% chance that the H_0 is rejected but it should have been accepted. There is a 03.75% chance that if the process is determined to be out of control and not centred on the centreline that is actually in control and centred. This is low and is unlikely, therefore if the H_0 is rejected it should be rejected.

4.3 Delivery time reduction optimisation

To determine the minimum cost possible a solution determined using brute force. Assuming that the process can only be reduced by whole hours and not parts (only integers), to incur the least amount of cost the Delivery Time must be reduced with 2 hours. This is mainly due to the penalty cost incurred

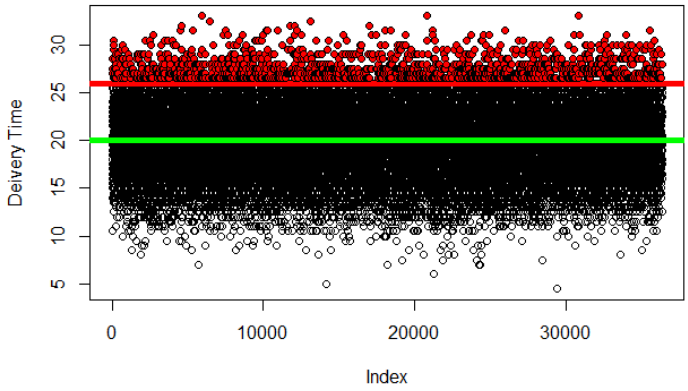

Figure 38: Graph of Delivery Times where penalties are incurred

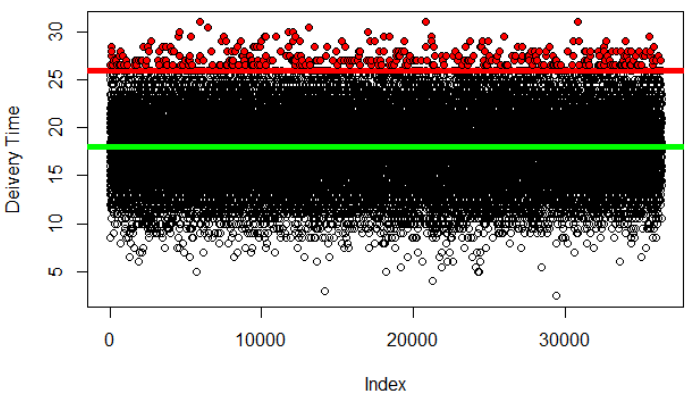

Figure 37: Graph of Delivery Times where penalties are incurred with reduction

if the Delivery Time is more than 26.

In the graphs above the Delivery Time for Technology items are plotted. All the items that incur penalty costs are marked in red. The red line is the threshold for where penalties are incurred (26) and the green line is the mean. It can be observed that by reducing the Delivery Time with 2 hours most of the items that incur penalties will no longer receive penalties. This leads the t0 least amount of cost.

4.4 Type 2 error

The type 2 error was calculated to be 0.4105766. Meaning that there is a 42.06% chance that if the null hypothesis is accepted it is actually false. This is large and it is due to the Xbar being between the LCL and UCL.

Part 5. DOE and MANOVA

For the manova test the relationships between all numerical features (Age, Year, Day, Month, Delivery Time) and Class will be examined as well as Delivery Time with all the other numerical features. The manova results will then be compared to the relationships identified in part 2.

For the manova test all the H_0 will read as follows: “Variables do not influence each other”. Then H_a will read: “Variables do influence each other”. Any p-value smaller than 0.05 will be seen as significant and will result in a rejection of the null hypothesis.

Class MANOVA

Variable <chr>	p-value <dbl>
Age	0.0000000
Year	0.0000000
Day	0.1776729
Month	0.2912988
Price	0.0000000
Delivery.time	0.0000000

Figure 39: Table of p-values for Class

Above the p-values for all the variables when compared to Class can be seen. It can be observed that only the p-values for Day and Month are larger than 0.05 and therefore their null hypothesis' will be accepted and all other will be rejected. This means that Class has a relationship with Age, Year, Price and Delivery Time and does not have a relationship with Day and Month. This correlates with the results in part 2 except for Age. This relationship was not identified.

Delivery Time MANOVA

Variable <chr>	p-value <dbl>
Age	0.0000000
Year	0.0000000
Day	0.2415335
Month	0.1839459
Price	0.0000000

Figure 40: Table of p-values for Delivery Time

Above the p-values for all the variables when compared to Delivery Time can be seen. It can be observed that only the p-values for Day and Month are larger than 0.05 and therefore their null hypothesis' will be accepted and all other will be rejected. This means that Delivery Time has a relationship with Age, Year and Price and does not have a relationship with Day and Month. This correlates with the results in part 2 except for Age. This relationship was not identified.

Part 6: Reliability of the service and products

6.1 Problem 6 and 7 of chapter 7

Problem 6

The Taguchi loss function for this problem was calculated to be:

Taguchi loss = $28125(y-0.06)^2$.

This graph can be seen plotted below.

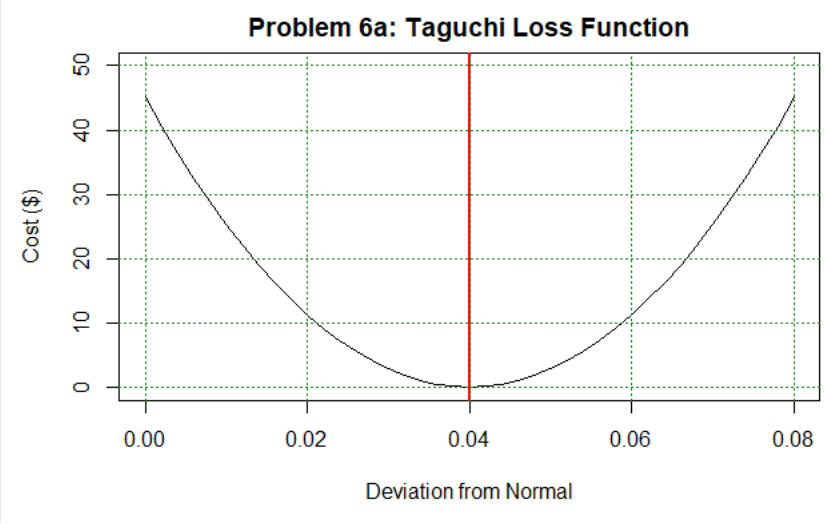

Figure 41: Problem 6 Taguchi loss function

Problem 7

a) The Taguchi loss function for this problem was calculated to be: Taguchi loss = $28175(y-0.06)^2$.

This graph can be seen plotted below.

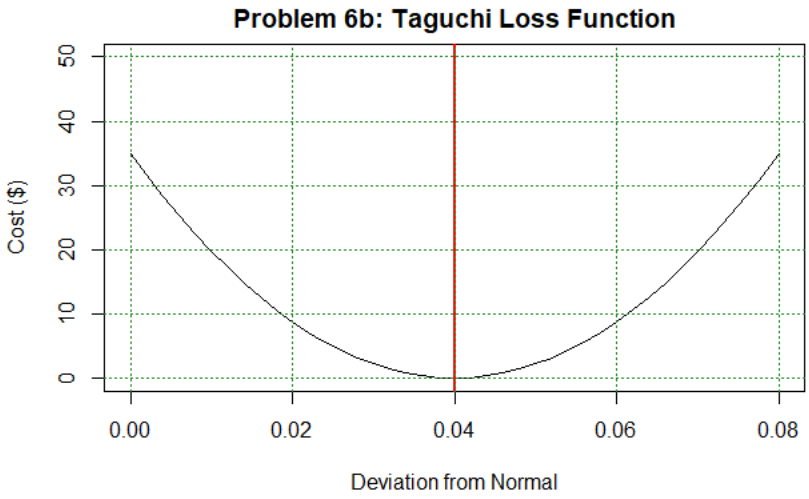

Figure 42: Problem 7 Taguchi loss function

b) If the process deviation is reduced to 0.027cm the Taguchi loss is 15.94687.

6.2 Problem 27 of chapter 7

a) The reliability of the system without backup machines is 0.7038.

b) The reliability of the system with backup machines is 0.9615316. This is an improvement of 26%.

6.3 Binomial probabilities

By using the available information, it was calculated that with 21 vehicles there will be reliable deliveries 364.8885 days of the year.

References

1 Factory. (2022, 10 15). *Process Capability Analysis Cp, Cpk, Pp, Ppk - A Guide*. Retrieved from 1factory: https://www.1factory.com/quality-academy/guide-to-process-capability-analysis-cp-cpk-pp-ppk.html

Amitav Banerjee, U. B. (2009). Hypothesis testing, type I and type II errors. *Industrial Psychiatry Journal* , 1.

ori.hhs.gov. (2022, 10 15). *Data Analysis*. Retrieved from Responsible Conduct in Data Management: https://ori.hhs.gov/education/products/n_illinois_u/datamanagement/datopic.html#