UNIVERSITEIT•STELLENBOSCH•UNIVERSITY
jou kennisvennoot • your knowledge partner

FACULTY OF ENGINEERING

Department Industrial Engineering

Quality Assurance 344
ECSA report

Date: 21 October 2022

INDUSTRIAL ENGINEERING
Stellenbosch University

This report done by Jakobus Pierre de Wet (23751916, 3rd year)
For the Engineering Counsel of South Africa in October 2022

Relevant Department of study:
The Department of Industrial Engineering, Stellenbosch University

Abstract

This report describes work done for the ECSA report. It contains the results of process analysis done on an online retail company. Insights on the company's sales are drawn from data analysis. Results and other outcomes are considered, and recommendations are also stated. These recommendations aim to improve the overall performance of the company's sales.

List of figures

Disclaimer: All illustrations are used with the consent of the company.

List of Tables

Table of Contents

Introduction

An online retail company keeps track of their sales by creating a data set which contains information on their sales. To gain competitive advantage in the online shopping industry a sales analysis is needed to gain valuable insights. The data is in raw form and needs some processing before any analysis can be done. Hereafter, the valid data is summarized and … by means of descriptive statistics. Statistical process control (SPC) is implemented on the delivery times of the various product classes. A SPC table is set up with valuable SPC variables for each class. After this the optimization phase for the delivery process starts by discovering inconsistencies and making informed recommendations. The probability of making Type 1 and Type 2 errors is also discussed. A DOE and MANOVA analysis are made to gain insight on. The report ends off with calculations regarding the reliability of the service and products.

Part 1: Data wrangling.

The sales data that we've been given has data quality issues that need to be fixed before we can go further. All the issues can be found in the feature called "Price".

Invalid Data set:

A data set that is a combination of 2 separate data sets, one containing all the missing values and another containing all the negative values.

Negative values subset:
5 Observations of 10 variables.

RStudio Source Editor

negvalues

Filter

	X	ID	AGE	Class	Price	Year	Month	Day	Delivery.time	Why.Bought
16320	16320	44142	82	Household	-588.8	2023	10	2	48.0	EMail
19540	19540	65689	96	Sweets	-588.8	2028	4	7	3.0	Random
19998	19998	68743	45	Household	-588.8	2024	7	16	45.5	Recommended
144443	144443	37737	81	Food	-588.8	2022	12	10	2.5	Recommended
155554	155554	36599	29	Luxury	-588.8	2026	4	14	3.5	Recommended

Figure 1: Table showing negative values subset

Incomplete cases subset:
17 Observations of 10 variables.

RStudio Source Editor

InComplete

Filter

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

Figure 2: Table showing incomplete values subset

These two sets are combined to create an overall invalid data set.

Invalid Data Set:
22 Observations of 11 variables with mCount as integer values ranging from 1 to 22.

RStudio Source Editor

InvalidSet

Filter

	mCount	X	ID	AGE	Class	Price	Year	Month	Day	Delivery.time	Why.Bought
12345	1	12345	18973	93	Gifts	*NA*	2026	6	11	15.5	Website
16321	2	16321	81959	43	Technology	*NA*	2029	9	6	22.0	Recommended
16322	3	16322	84530	31	Food	359.03	2021	6	16	2.5	Recommended
19541	4	19541	71169	42	Technology	*NA*	2025	1	19	20.5	Recommended
19543	5	19543	93483	89	Gifts	1828.96	2023	6	19	9.5	Recommended
19999	6	19999	67228	89	Gifts	*NA*	2026	2	4	15.0	Recommended
20002	7	20002	43366	33	Food	413.08	2027	9	17	2.5	EMail
23456	8	23456	88622	71	Food	*NA*	2027	4	18	2.5	Random
34567	9	34567	18748	48	Clothing	*NA*	2021	4	9	8.0	Recommended
45678	10	45678	89095	65	Sweets	*NA*	2029	11	6	2.0	Recommended
54321	11	54321	62209	34	Clothing	*NA*	2021	3	24	9.5	Recommended
56789	12	56789	63849	51	Gifts	*NA*	2024	5	3	10.5	Website
65432	13	65432	51904	31	Gifts	*NA*	2027	7	24	14.5	Recommended
76543	14	76543	79732	71	Food	*NA*	2028	9	24	2.5	Recommended
87654	15	87654	40983	33	Food	*NA*	2024	8	27	2.0	Recommended
98765	16	98765	64288	25	Clothing	*NA*	2021	1	24	8.5	Browsing
144444	17	144444	70761	70	Food	*NA*	2027	9	28	2.5	Recommended
144457	18	144457	47570	67	Household	1235.33	2028	12	16	41.0	Browsing
155555	19	155555	33583	56	Gifts	*NA*	2022	12	9	10.0	Recommended
155569	20	155569	25588	55	Food	251.39	2026	7	14	2.5	Random
166666	21	166666	60188	37	Technology	*NA*	2024	10	9	21.5	Website
177777	22	177777	68698	30	Food	*NA*	2023	8	14	2.5	Recommended

Figure 3: Complete invalid data set

Valid Data Set:

Data set that contains instances that have no data quality issues. The missing and negative values have been removed by means of code.
179978 observations of 11 variables with nCount as integer values ranging from 1 to 179978.

	nCount	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
47329	24	47329	63397	53	Technology	50791.93	2021	1	1	20.0	Recommended
49966	25	49966	85562	39	Household	14913.66	2021	1	1	28.0	Random
50553	26	50553	31911	54	Food	426.88	2021	1	1	2.0	Random
56879	27	56879	57698	50	Clothing	234.71	2021	1	1	9.5	Recommended

Figure 4: First 27 entries of the valid data set

Part 2: Descriptive statistics

Now that we have a valid data set to work with, we can start analysing the sales data. The analysis will result in a better understanding of the data. Valuable insights can be drawn out from the data to help the company perform better.

5 Number statistical summary of each feature:

```
     nCount             X                ID
Min.   :     1   Min.   :     1   Min.   :11126
1st Qu.: 44995   1st Qu.: 45004   1st Qu.:32700
Median : 89990   Median : 90005   Median :55081
Mean   : 89990   Mean   : 90003   Mean   :55234
3rd Qu.:134984   3rd Qu.:135000   3rd Qu.:77637
Max.   :179978   Max.   :180000   Max.   :99992
      AGE            Class               Price
Min.   : 18.00   Length:179978      Min.   :  -588.8
1st Qu.: 38.00   Class :character   1st Qu.:   482.3
Median : 53.00   Mode  :character   Median :  2259.6
Mean   : 54.57                      Mean   : 12294.1
3rd Qu.: 70.00                      3rd Qu.: 15271.0
Max.   :108.00                      Max.   :116619.0
      Year          Month             Day
Min.   :2021   Min.   : 1.000   Min.   : 1.00
1st Qu.:2022   1st Qu.: 4.000   1st Qu.: 8.00
Median :2025   Median : 7.000   Median :16.00
Mean   :2025   Mean   : 6.521   Mean   :15.54
3rd Qu.:2027   3rd Qu.:10.000   3rd Qu.:23.00
Max.   :2029   Max.   :12.000   Max.   :30.00
Delivery.time   Why.Bought
Min.   : 0.5   Length:179978
1st Qu.: 3.0   Class :character
Median :10.0   Mode  :character
Mean   :14.5
3rd Qu.:18.5
Max.   :75.0
```

Figure 5:Feature summary (R-format)

Plots:

Histograms:

Count of sales per product class:

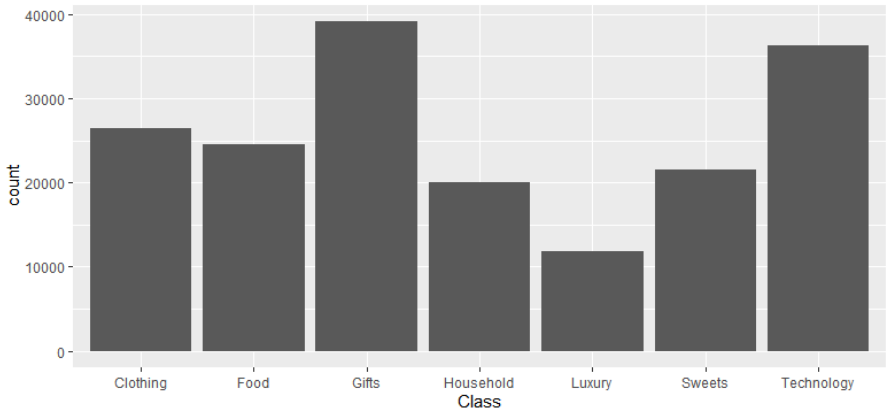

Figure 6: Count of sale per product class

This sales count graph shows that the two product classes that had the most sales, were Gifts and technology. The class with the least sales is the Luxury class.

Count of sales at different prices:

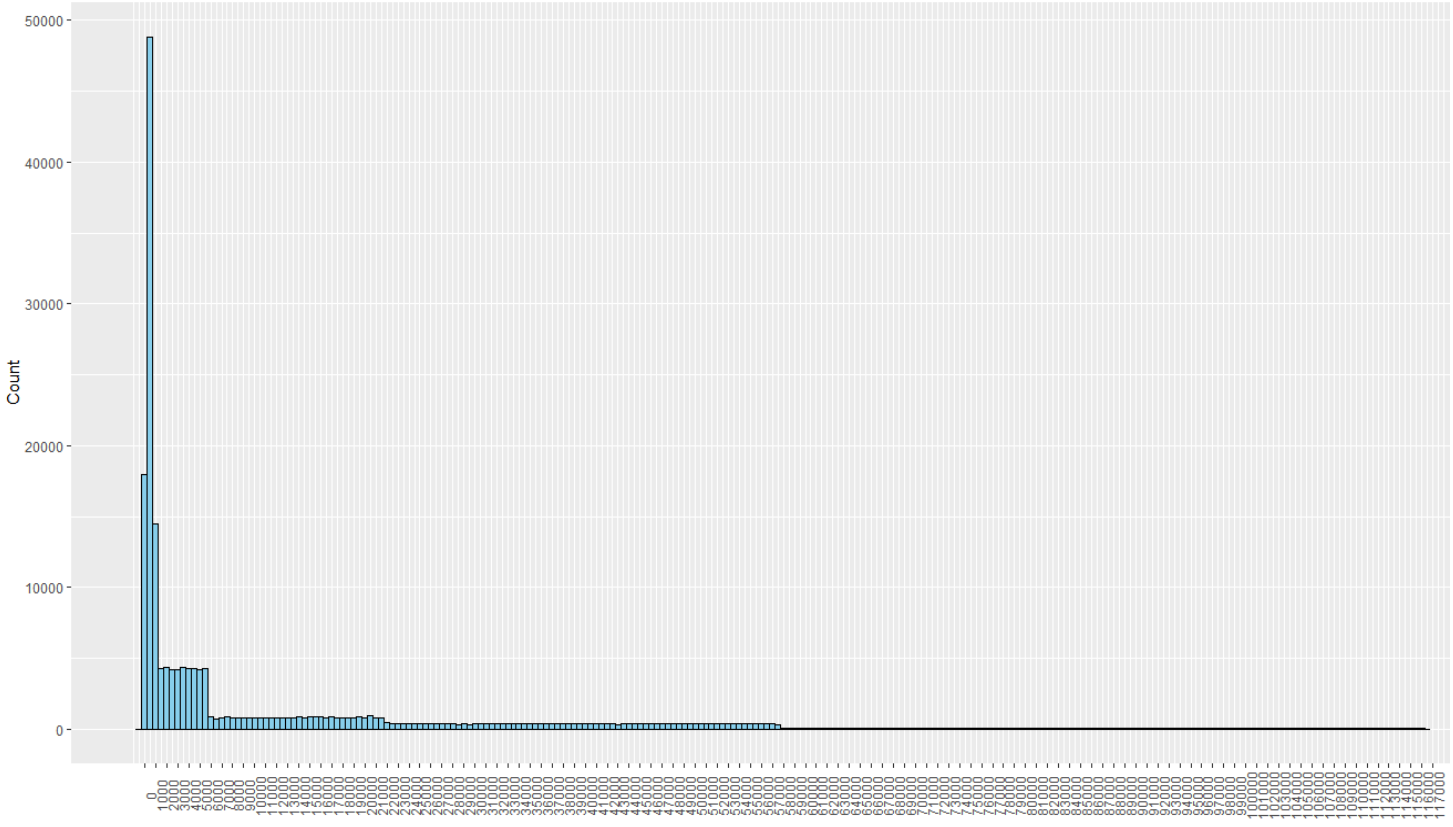

Figure 7: Sales count at different prices

The above graph shows the distribution of the sales count per price in thousands.

Price distribution summary:

- 0<Price<500: 47119 instances
- 500<Price<1000: 28074 instances
- 1000<Price<2000:12504

- 2000<Price<4000: 17161
- 4000<Price<6000: 15379
- 6000<Price<10000: 6297
- 10000<Price<15000: 7977
- 15000<Price<20000: 8402
- 20000<Price<30000: 9232

See the graph for the rest of the distribution…

Closer look at the price between 300 and 5000:

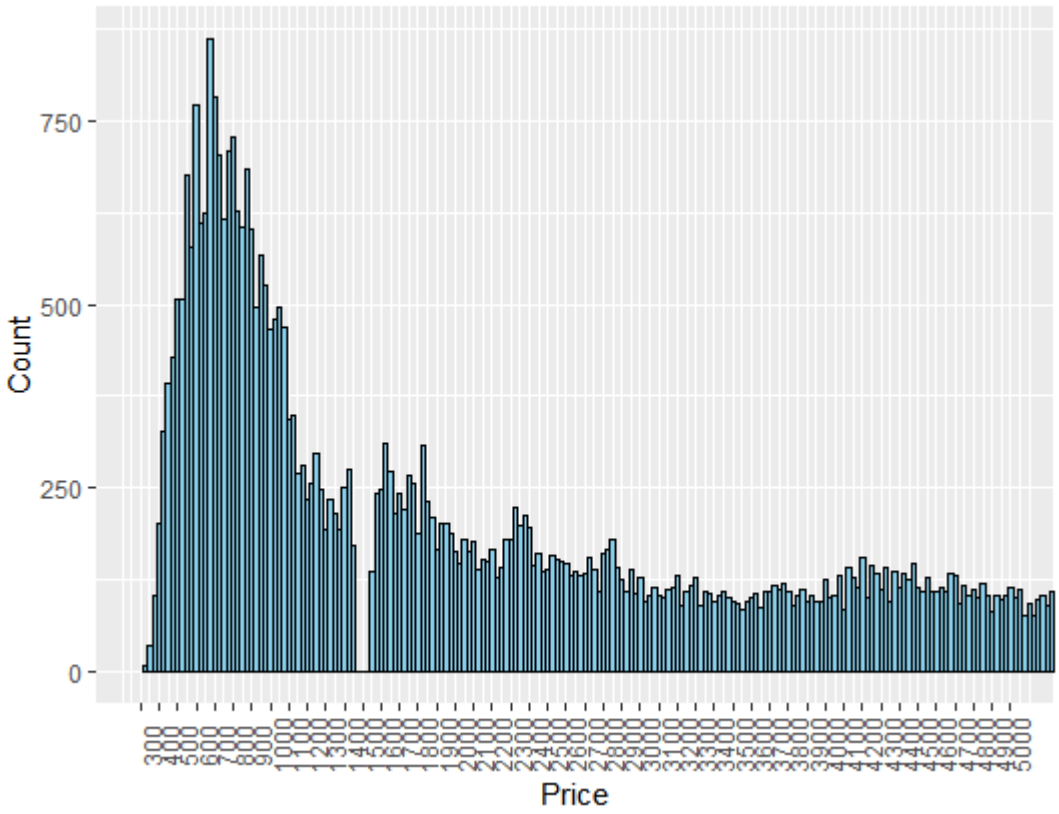

Figure 8: Sales count between R300 and R5 000

This data set follows skewed-right distribution. It shows that a lower prices accounts for more sales.
Summary of prices per class:

Table 1: Summary of each product class before removing any entries

Class	Maximum Price	Minimum Price	Median Price
Clothing	1154.02	127.76	642.04
Food	691.96	-558.80	408.37
Gifts	5774.49	172.61	2961.65
Household	21935.33	-588.80	10959.78
Luxury	116618.97	-588.80	65340.93
Sweets	576.38	-588.80	303.25
Technology	5775.40	935.18	29653.90

This table summarizes the price feature per class. Household, Luxury, and Sweets all have a minimum value that's negative.

This can be a result of an error while imputing the price of those certain sales. These values are removed before any further data processing can be done.

Count of sales per reason for buying a product:

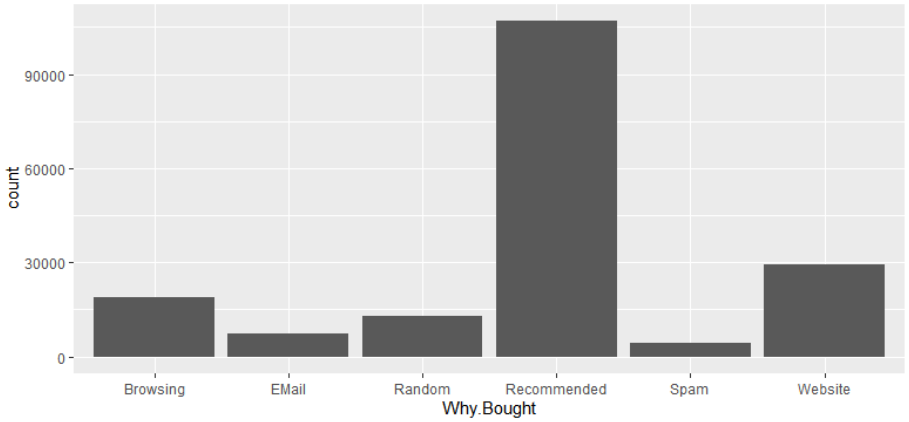

Figure9: Count of sales per reason for buying a product

This graph shows how the customers got informed of the product. Recommendations resulted by far in the most sales. The company can use this information as a reason for spending more money in further advancing their recommendations.

Count of sales per year:

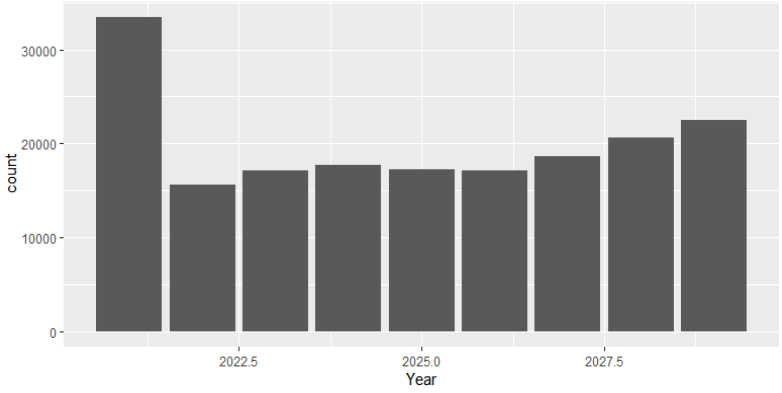

Figure 10:Count of sales per year

This graph shows the sales per year. The first year resulted in the most sales. In the following year the number of sales halved, but here after it steadily started to increase again up until the previous where there were just above twenty-two thousand sales.

Reasons for buying top 3 sales product classes namely, Gifts, Technology, and Clothing:
The 3 following graphs shows the Why.Bought feature count for the top performing product classes.
In all 3 recommendations again beat all the reasons. It also becomes clear after analyzing the graphs that categories like Browsing and Websites show potential and that the business can retain more sales if there is invested into these channels.

Gifts:

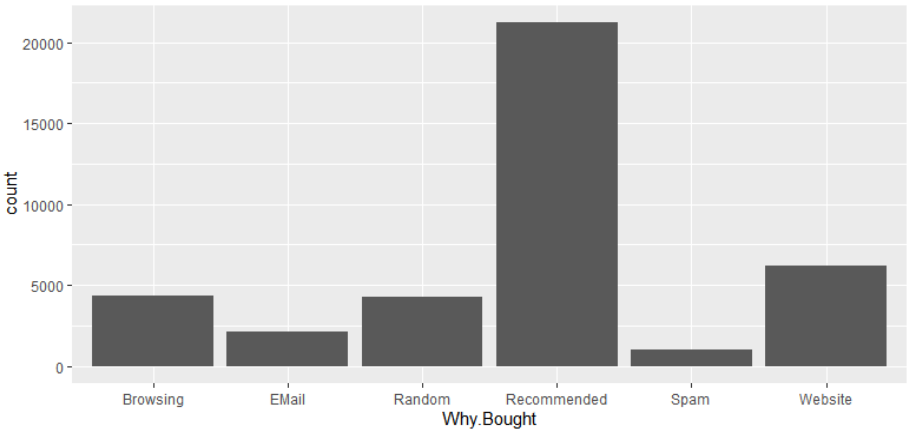

Figure 11: Reasons for buying Gift items

Technology:

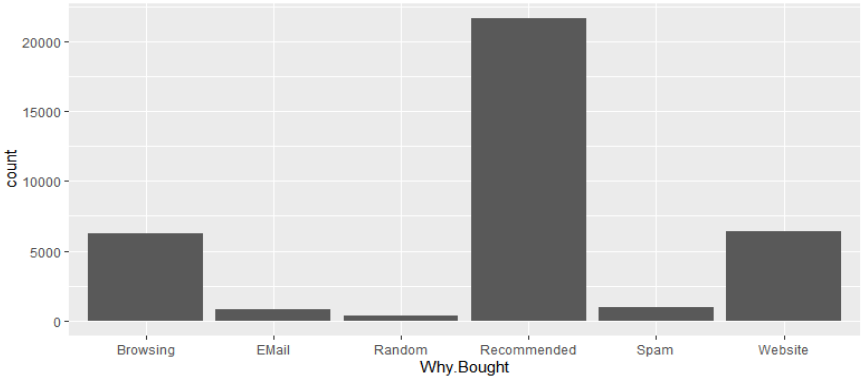

Figure 12: Reasons for buying Technology items

Clothing:

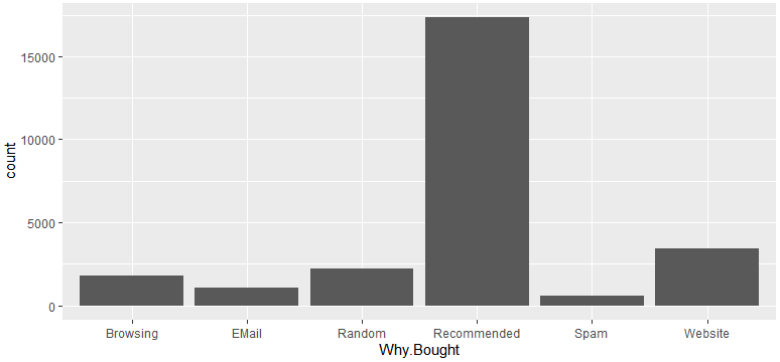

Figure 13: Reasons for buying Clothing items

Sales count per age:

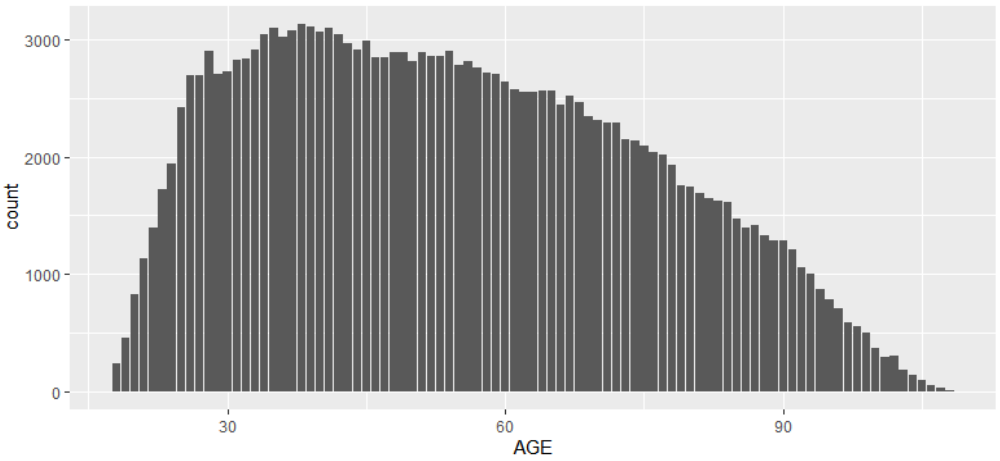

Figure 14: Sales count per age group

The graph of age versus sales count follows a skewed-right distribution. It shows that younger customers accounts for more frequent purchases. It should also be noted that maximum age is 108 years. This might again, like the negative values in the Price feature, be a result of errors when imputing values to the data set.

Bar plots:

Average times of delivery per product class in days:

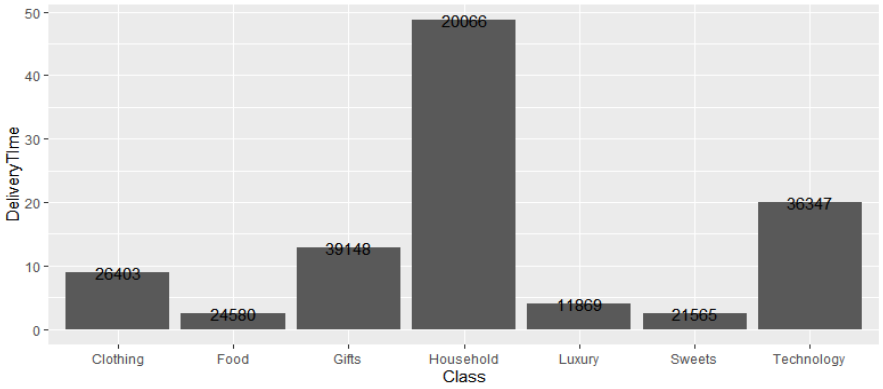

Figure 15: Average delivery time per product class in days

The household product class has the highest mean delivery time of plus minus 49 days. This is understandable seeing that this class has physically the biggest products and this in part will result more time being spent in delivery because of the transport mode being trucks that slower than small package carriers that is the mode of transport for most of the other classes. The Technology class comes in second and this might be a result of the fact that these are often required to managed with seeing that they are more fragile. This will also take away from the delivery time

Average sales price per product class:

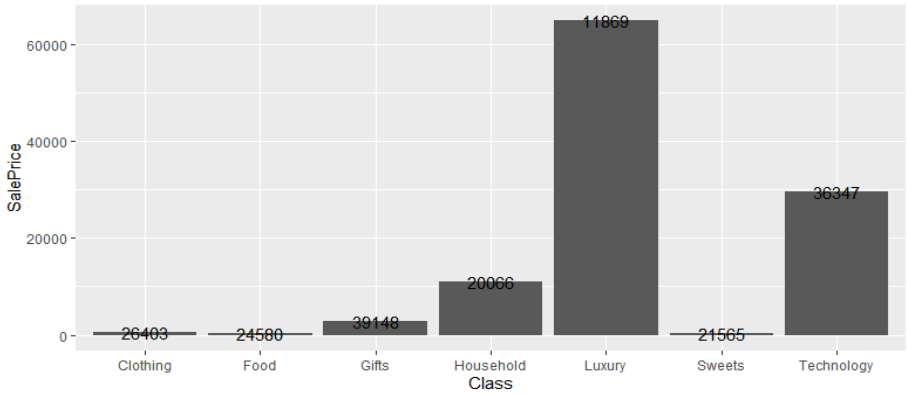

Figure 16: Mean per product class

This graph shows the mean price for the various product classes. Luxury products are on the steep end compared to other classes. Technology also has higher sales prices than the other classes.

Boxplots:

Price distribution per class:

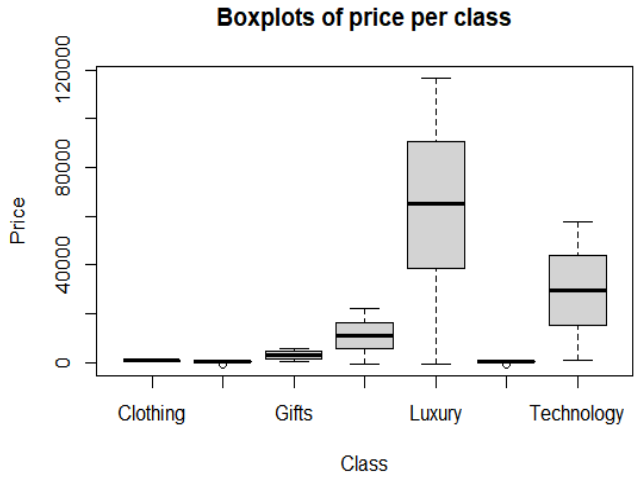

Figure 17: Boxplot of price distribution per product class

These boxplots show the price ranges per Class. The Luxury class and Technology class has the biggest price ranges. This might be a result of the variety of product that can be purchased in theses classes. The Food, Sweets, Gifts, and Clothing classes have relatively no margins compared to the other 3 classes. Practically this makes sense seeing that the prices in these classes don't often go over R500, except if you'd be buying a whole sheep or some very fancy imported food. These online shops normally don't sell these types of products.

Delivery time distribution per class:

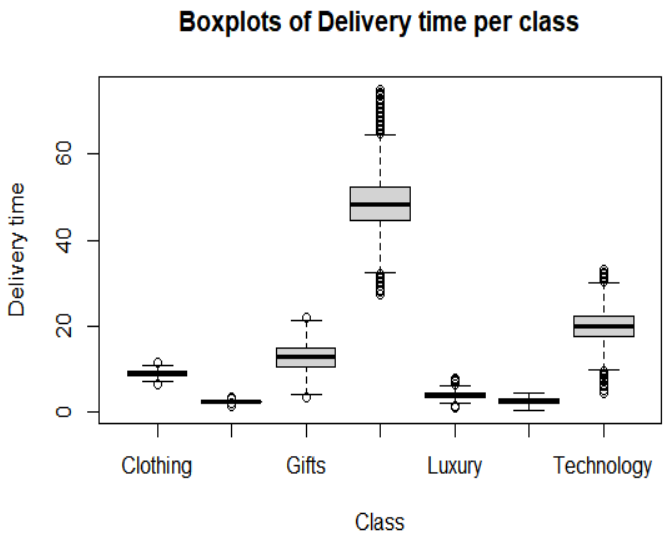

Figure 18: Boxplot of delivery time distribution per product class

These boxplots confirm the insights from the histograms of the delivery times per class. Another insight that can be drawn is that the Household class has quite a big range for their delivery times. This accounts for the varying size of household products that can be bought, but it still shows that Household and Technology are the two classes that have the biggest delivery winows.

Line plots:

Change in sales throughout the month:

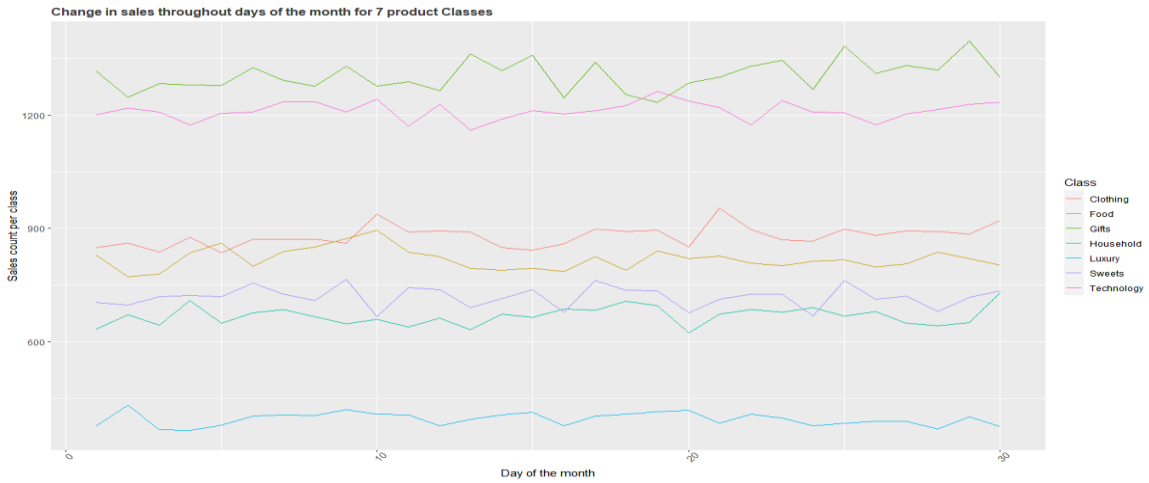

Figure 19: Line plot of change in product class throughout the month

By looking at the above given graph, it's clear that there aren't really any tendencies when it comes to what day of the month there are more sales.

Change in sales throughout the year:

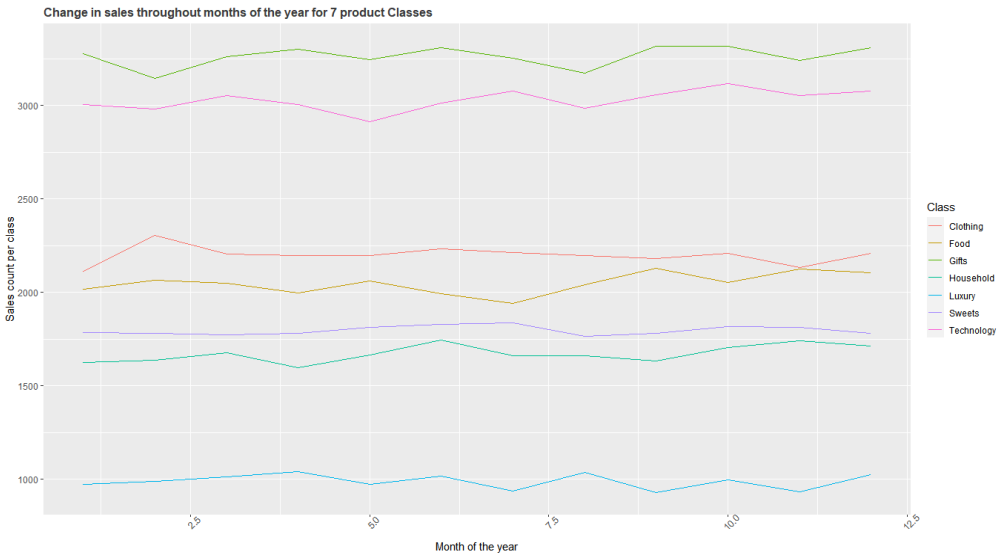

Figure 20: Line plot of change in product class throughout the year

The same can be seen for the sales per month. Here there also is not any noticeable tendencies.

Change in sales according to customer age:

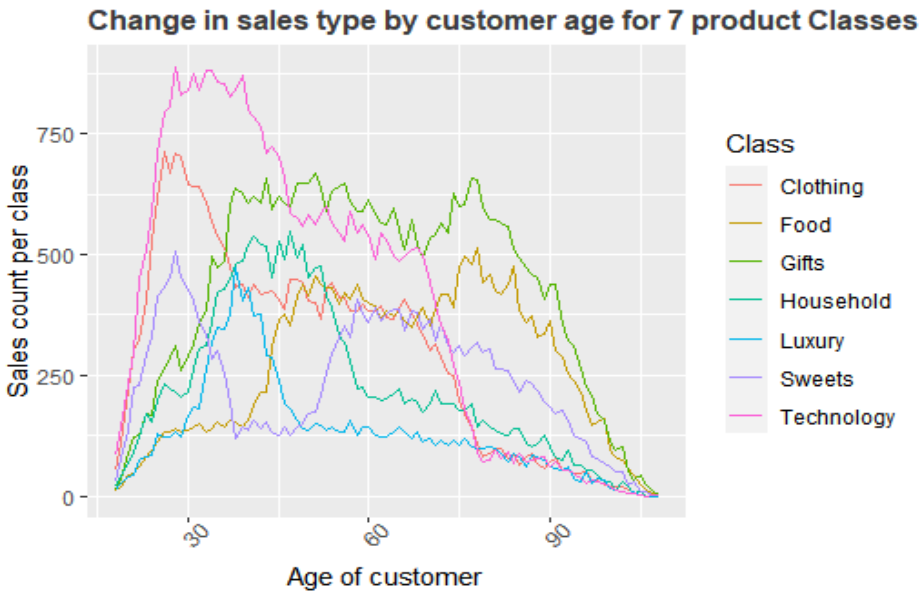

Figure 21: Line plot showing change in product sales according to customer age

Quite a few insights can be made from this graph. Younger customers spend more money on clothing and technology. Food purchases increases from the age of plus minus 40 years. This a result of parents buying food for their families. The middle age group don't really buy that much sweet compared to the younger and older group. Household purchases has a peak at plus minus 45 years. Until that point there is a steady increase in purchases, and thereafter, there is a decline as age of the customers increase. Luxury items has a peak at plus

minus 40 years. Maybe a sign of customers dealing with a mid-life crisis…

The company can use all this information to run promotions on certain products with the according customer age group. They can also use this information for recommendations, emails, and their website information channels.

Change in the reason for buying according to customer age:

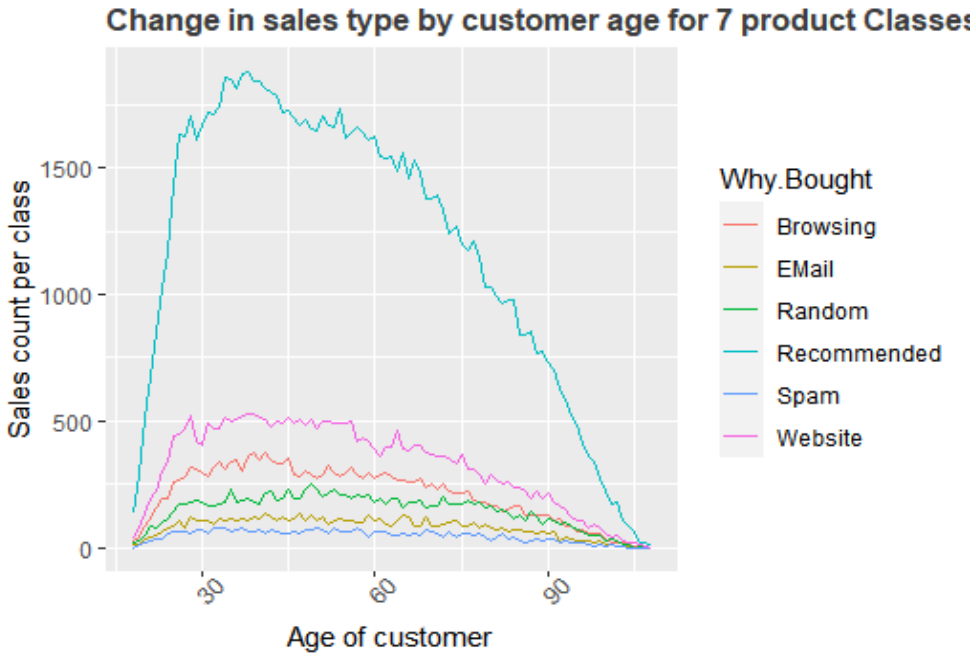

Figure 22: Line plot showing change in purchasing reason according to customer age

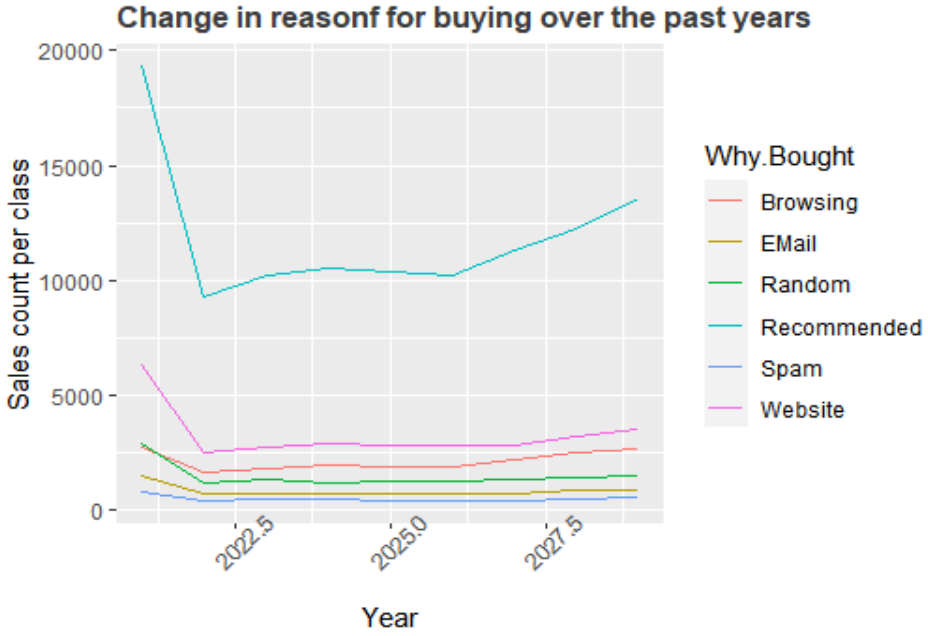

Figure 23: Line plot showing change in purchasing reason over the past few years

Process Capability indices:

The process capability index gauges how much variance a process encounters in comparison to its specification parameters. We might compare various procedures in terms of the ideal circumstance or whether they live up to our expectations.

The Technology class's delivery time information was used to find the indices. A Upper and Lower Specification Limit was given. The Lower Specification Limit of 0 makes sense because a negative value for delivery time is impossible. The product can't be received before it's even ordered…
The indices were found by using the following formulas:

$$C_p = \frac{USL - LSL}{6\sigma}$$

$$C_{pu} = \frac{USL - \mu}{3\sigma}$$

$$C_{pl} = \frac{\mu - LSL}{3\sigma}$$

$$C_{pk} = \min(C_{pl}, C_{pu})$$

All the following values are calculated from the delivery time of the Technology product class:

Table 2: Table of process capability indices

Variable	Value
Standard Deviation	3.501
Mean	20.106
Cp	1.142
Cpu	1.905
Cpl	0.380
Cpk	0.380

Part 3: Statistical process control (SPC):

X-bar and S charts serve as a statistical approach to determine whether a process is in control or not. An indication to a product class that is in control are samples that are within the control limits. More samples outside the control limits corresponds to a more out of control process. These control limits are determined beforehand for each class by looking at the first 30 samples of that class.

3.1. Analyzing the first 30 Samples

The complete data is arranged the in samples of 15 entries each. The critical SPC values were calculates with

these samples. The critical values were then used to plot X-charts for each of the seven classes. The critical values can be found in the following tables:

X-Bar chart table:

Table 3: Table of control limits for each product class according to X-bar charts approach

Class	UCL	U2Sigma	U1Sigma	CL	L1Sigma	L2Sigma	LCL
Clothing	9.40493	9.25996	9.11498	8.97	8.82502	8.68004	8.53507
Household	50.24833	49.01963	47.79092	46.56222	45.33352	44.10482	42.87612
Food	2.70946	2.63631	2.56315	2.49	2.41685	2.34369	2.27054
Technology	22.97462	22.10789	21.24117	20.37444	19.50772	18.641	17.77427
Sweets	2.89704	2.75729	2.61753	2.47778	2.33802	2.19827	2.05851
Gifts	9.48856	9.11275	8.73693	8.36111	7.98529	7.60948	7.23366
Luxury	5.49397	5.24116	4.98836	4.73556	4.48275	4.22995	3.97715

S-chart table:

Table 4: Table of control limits for each product class according to S charts approach

Class	UCL	U2Sigma	U1Sigma	CL	L1Sigma	L2Sigma	LCL
Clothing	0.86656	0.76146	0.65635	0.55125	0.44614	0.34104	0.23593
Household	7.34418	6.45341	5.56264	4.67187	3.7811	2.89033	1.99956
Food	0.43725	0.38421	0.33118	0.27815	0.22511	0.17208	0.11905
Technology	5.18057	4.55222	3.92388	3.29553	2.66718	2.03883	1.41049
Sweets	0.83534	0.73402	0.6327	0.53139	0.43007	0.32875	0.22743
Gifts	2.24633	1.97388	1.70142	1.42897	1.15651	0.88405	0.6116
Luxury	1.51105	1.32778	1.1445	0.96123	0.77795	0.59468	0.41141

Charts:

These process control charts can be visualized with line graphs. Only the Luxury and Gift classes' charts will be displayed to be concise and effective. Observations will be made on these chart's classes

Luxury:
For the Luxury class the first 30 samples can be quite deceiving. At first this process seems in control,but

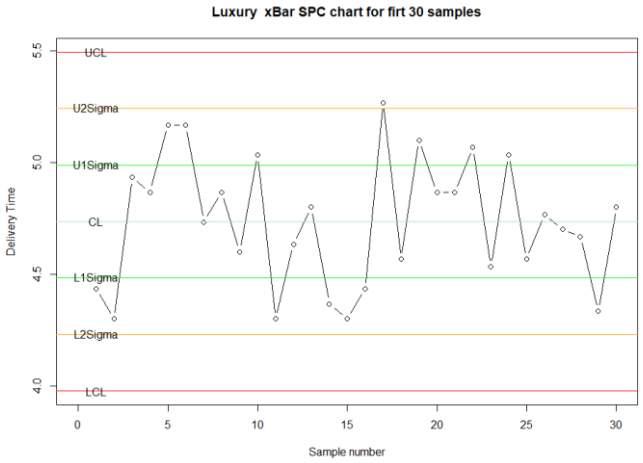

Figure 24: First 30 X-Bar Chart samples for Luxury

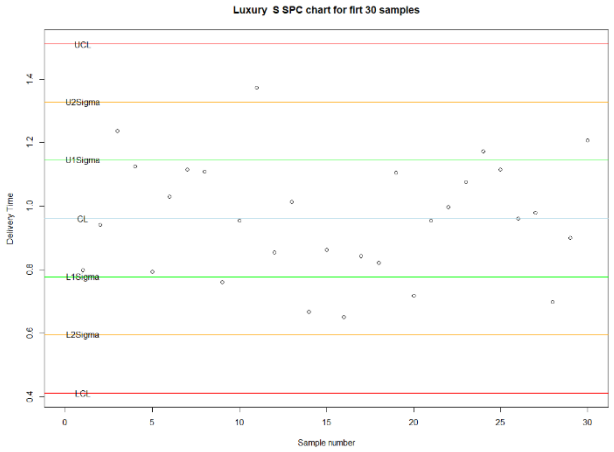

Figure 25: First 30 S-Chart samples for Luxury

Gifts:
The same can be said for the Gift product class. The first 30 samples shows some deviation from the centre line.

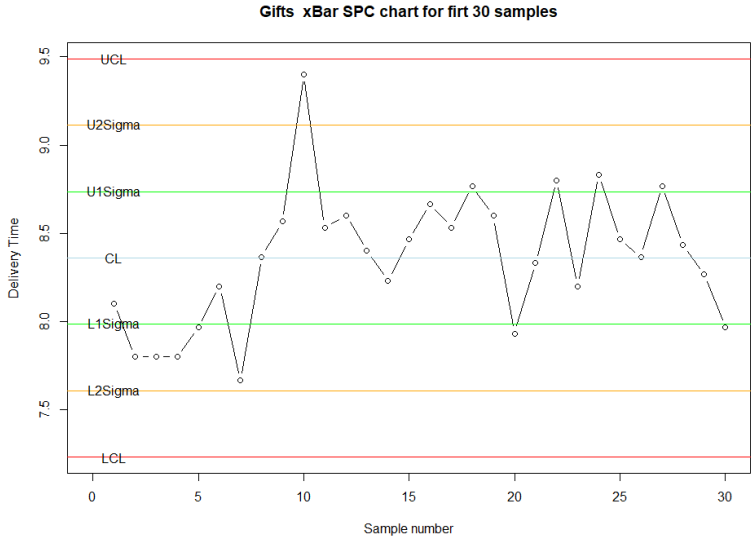

Figure 26:First 30 X-Bar Chart samples for Gifts

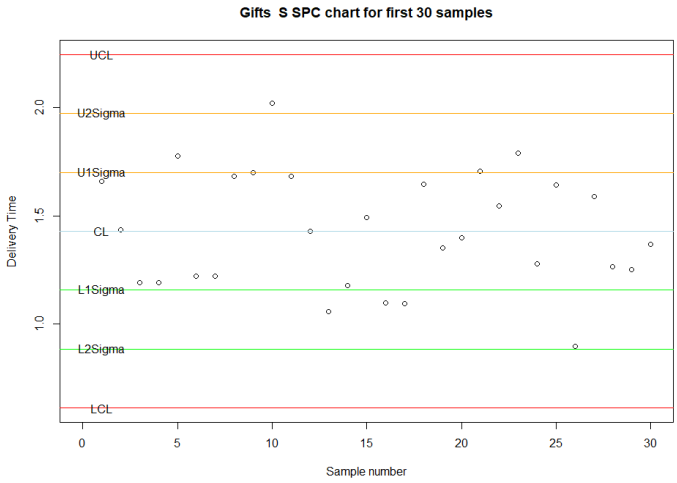

Figure 27:First 30 S-Chart samples for Gifts

3.2 Analyzing the full set Samples

Luxury:
When all the other samples are also plotted, then it becomes clear that this class has a lot of sample means outside of the critical value limits. These outliers could only be accounted for after a few sets of samples are analyzed.

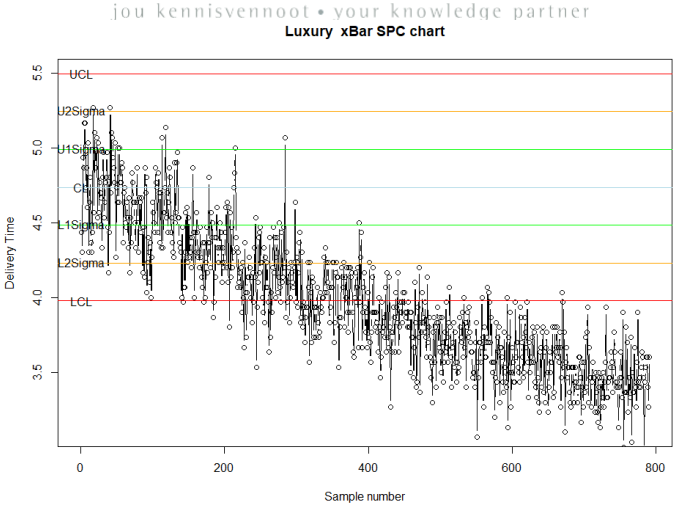

Figure 28: X-Bar Chart for all samples of Luxury

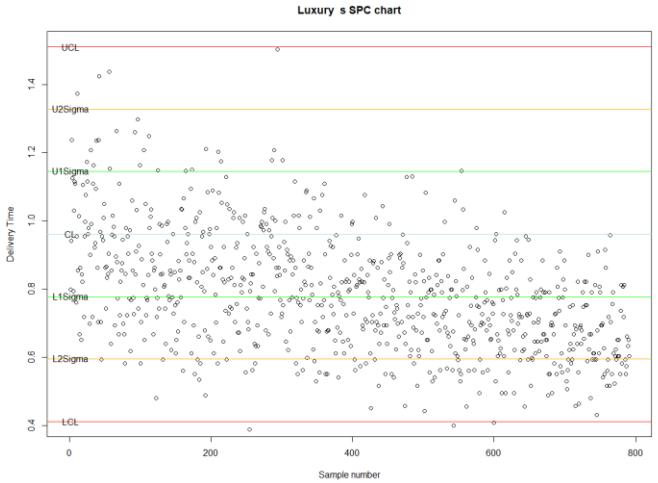

Figure 29: S-Chart for all samples of Luxury

Gifts:
Just like the Luxury class the Gifts class only starts deviating from the control limits only after a few sample sets. From here the charts follows an aggressive upwards trend. These actions would have been hard to predict by only looking at the first 30 samples of each of these two classes

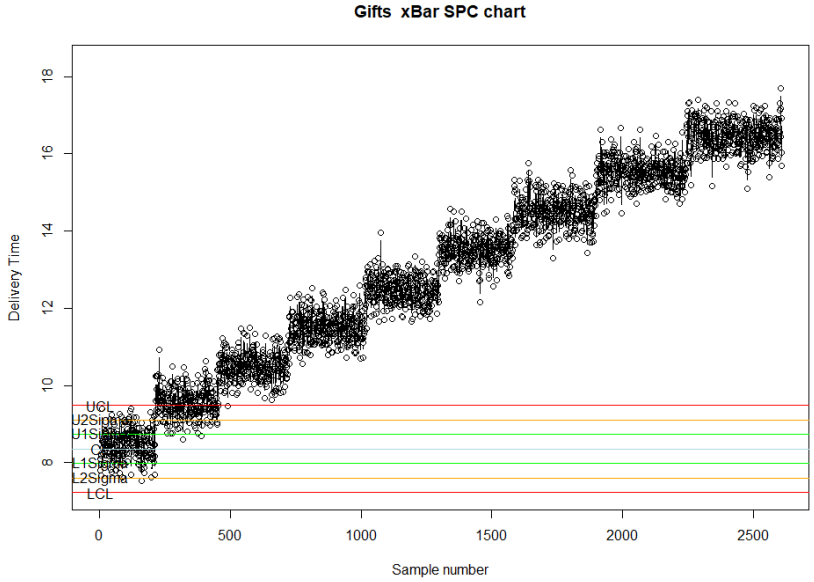

Figure 30: X-Bar Chart for all samples of Gifts

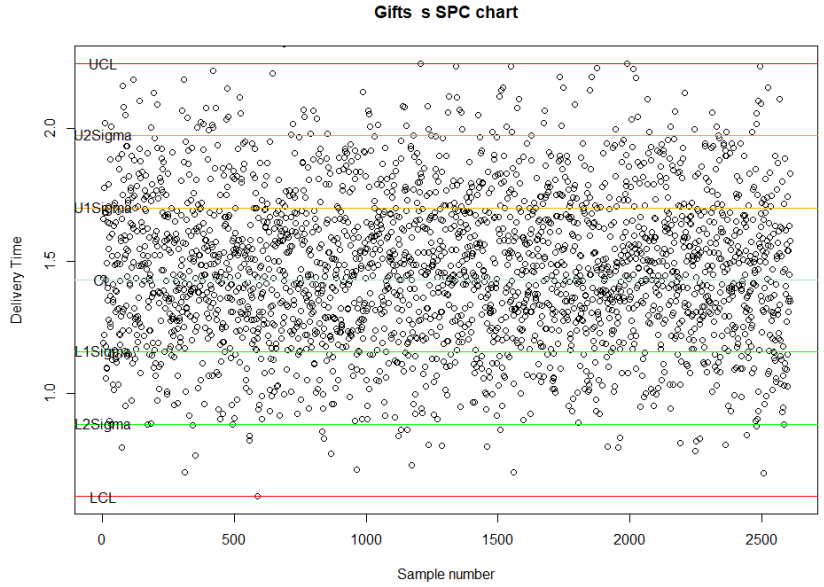

Figure 31: S-Chart for all samples of Gifts

Part 4: Optimizing the delivery processes

4.1 Out of control samples

As previously stated, SPC helps to determine whether a process is in control or not. After creating the critical values in part 3, we can use these limits are used to analyze the various classes. An easy way to understand this concept is by visualizing these classes on X-bar and S charts. The following table shows a summary of the out-of-control samples for each class's delivery time.

Part A:

Table 5: Table of outliers per class

Class	Total found	1st	2nd	3rd	3rd Last	2nd 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	135	NA	NA
Gifts	2287	213	216	218	2607	2608	2609
Luxury	440	142	171	184	789	790	791

By looking at this table it's clear that some of the cases are more severe than other. The Gifts class is the worst off with 2287 sample means outside of the control limits. This class's delivery system needs the most refining. Hereafter attention can be given to the Luxury and Household classes. The rest of the product classes seems fairly in control.

These outliers can now be marked on our charts to show how bad the control in the various classes are. In the following charts the outliers are indicated with red crosses on the instances.

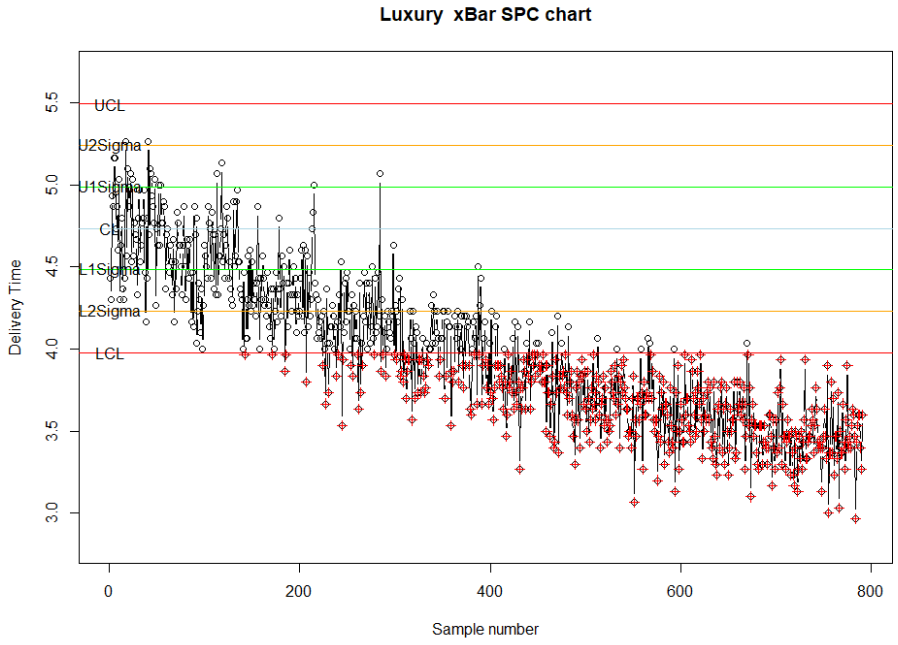

Figure 32: Outliers in Luxury class

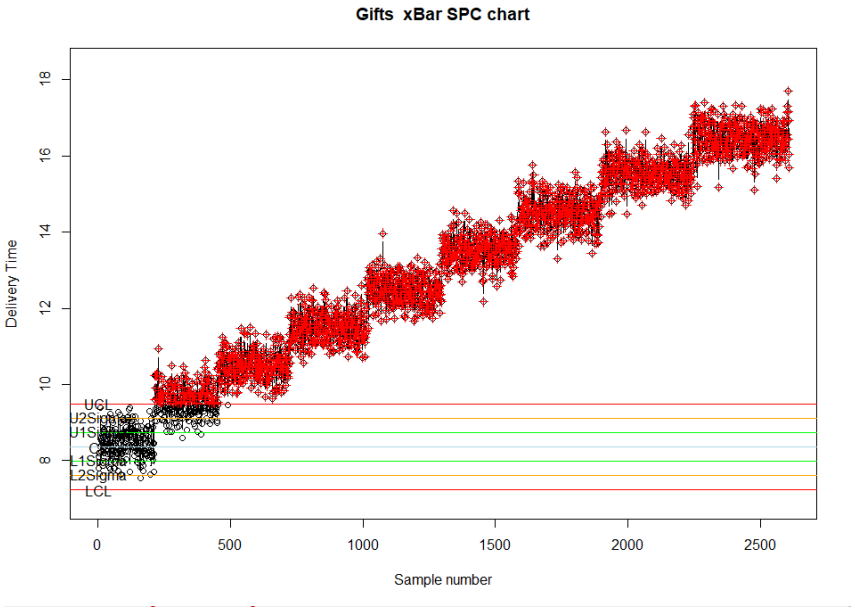

Figure 33: Outliers in Gifts class

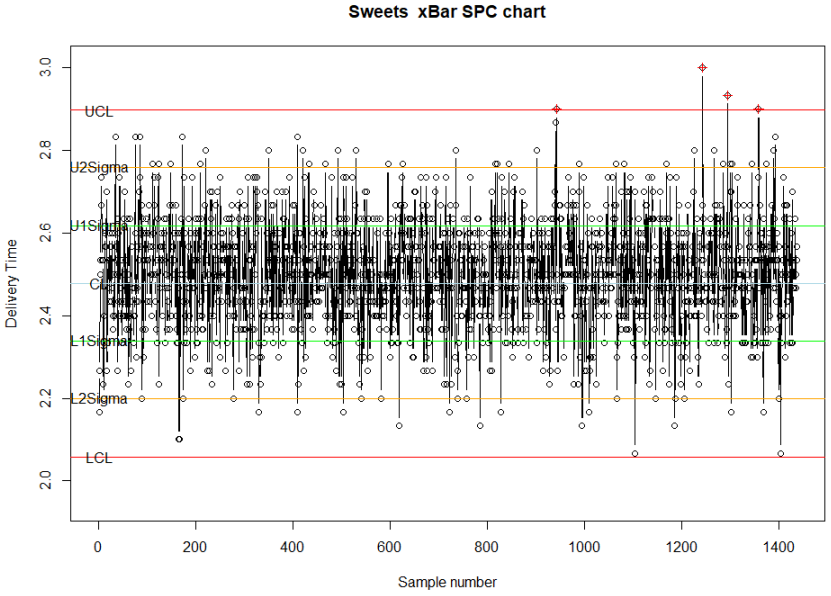

Figure 34: Outliers in Sweets class

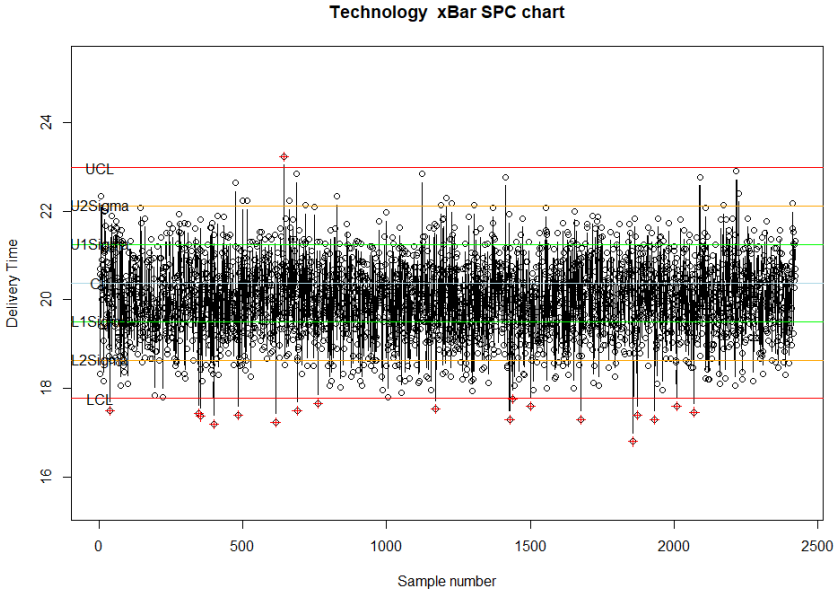

Figure 35: Outliers in Technology class

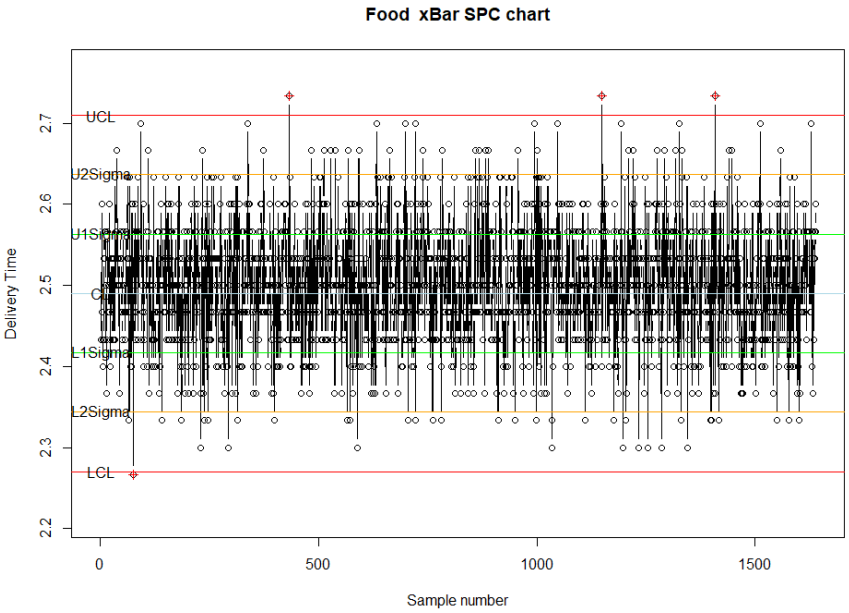

Figure 36: Outliers in Food class

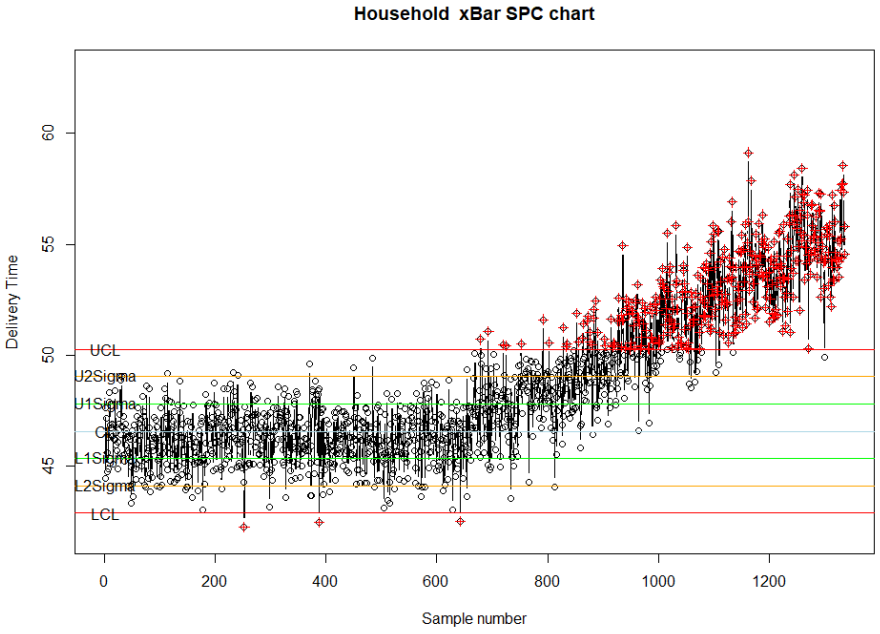

Figure 37: Outliers in Household class

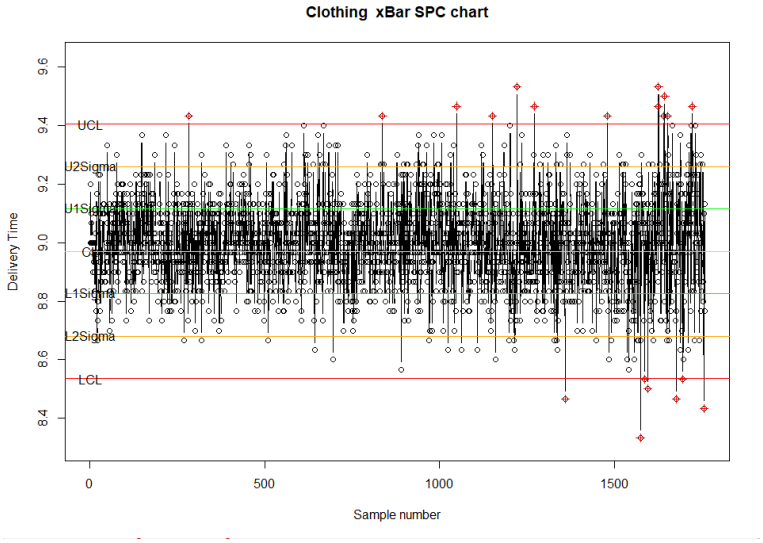

Figure 38: Outliers in Clothing class

Part B:

This part requires the analysis of consecutive samples of "s-bar or sample standard deviations" between -0.3 and +0.4 sigma-control limits. The last or ending sample number will also be stated.

Table 6: Table of consecutive samples inside -0.3 and +0.4 sigma-control limits

Classes	Index	Total
Clothing	1186	14
Household	1202	7
Food	1598	6
Technology	2384	21
Sweets	1371	10
Gifts	2476	20
Luxury	271	5

The Luxury product class has the biggest amount of sample standard deviations within these sigma-control limits with a value of 20. Hereafter is the Gifts class with 20 samples. Upon analyzing the graph it's clear that the last sample index number and total amount of samples within the limits has a fairly linear relationship. The higher the total number of samples found, the larger the index number

4.2 Type I (Manufacturer's) Error for A and B

The following table describes the types of mistakes that can be made:

Table 7: Type 1 and Type 2 error summary

	Process is fine	Process is not fine
SPC indicated the Process is not fine	Type I Error or Manufacturer's Error	Correct to fix process
SPC indicated the Process is fine	Correct to do nothing	Type II Error or Consumer's Error

Type 1 error for A:
When SPC indicates the and X-Bar sample is outside the control limits of plus and minus 3 sigmas while it's actually not then a Type I Error or Manufacturer's Error is made. The probability of this happening can be calculated by the following:

$$type1_A = \big(1 - pnorm(3)\big) * 2 = 0.270\%$$

Type 1 error for B:
When SPC indicates the and X-Bar sample is inside the control limits of plus and minus 3 sigmas while it's actually not then a Type 2 Error or Consumer's Error is made. The probability of this happening can be calculated by the following:

$$type1_B = pnorm(0.4) - pnorm(-0.3) = 27.333\%$$

4.3 Delivery time optimization calculations

In the online shopping market, the delivery process is getting more important by the day. Customers are not willing anymore to wait for long periods of time to receive their items. This has placed a lot of pressure on retail companies and has led to a whole new competitive side to the industry that can more than often make or break a deal.

For the Technology product class, the current average delivery time is 48.71859 hours.

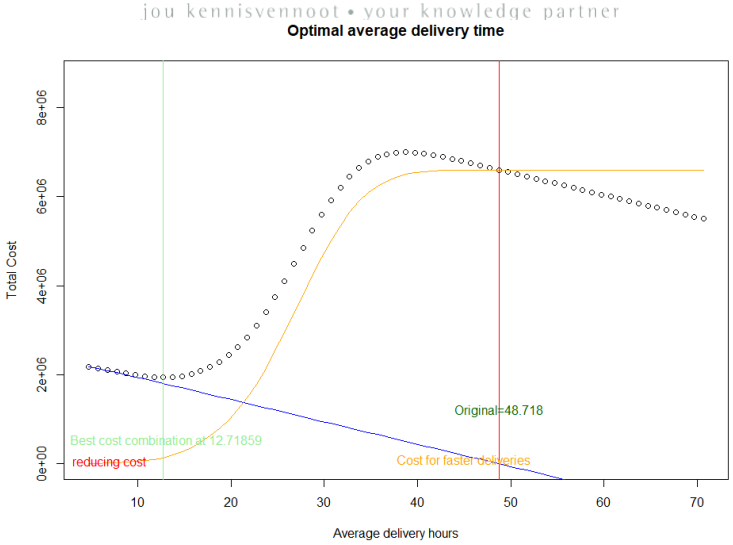

Figure 39: Total cost per average delivery hour

As seen on the given above graph, this current average delivery corresponds to a combined cost of R 6 450 898 The dotted line is the combination of the 2 costs. The minimum of this line will result in the minimum combined cost. The minimum value can be attained at a delivery time of 12.718 hours. This corresponds to a combined cost of R2 009 137. By optimizing the delivery time to 12.718 hours an amount of R 4 441 761 can be saved. This will ensure a much higher ROI for this company by reducing their delivery related costs.

4.4 Type 2 (Consumer's) Error for A

When a SPC indicates the X-Bar sample is inside the control limits while it's actually not then a Type 2 Error or Manufacturer's Error is made. In this scenario the delivery process average moves to 23 hours without being noted. The probability of making a Consumer's error in such a case can be calculated by the following:

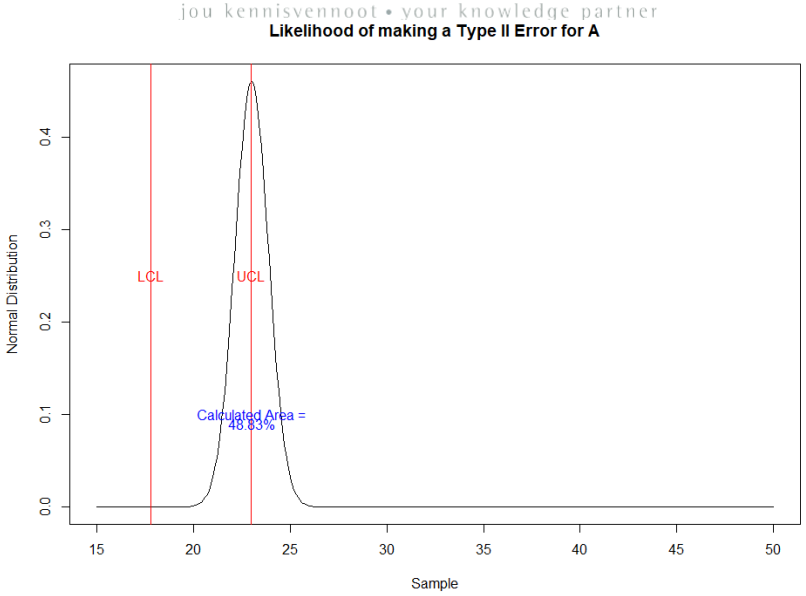

Figure 40: Normal distribution of samples

Part 5: DOE and MANOVA

By analyzing many dependent variables at once, multivariate ANOVA (MANOVA) expands the possibilities of analysis of variance (ANOVA). ANOVA uses statistics to compare the means of three or more groups.

A MANOVA can be set up to determine what the influence of the class of a product has on the product's price and delivery time. This results in the class being the dependent variable and the price and delivery time as the two independent variables. (Frost, 2022)

The null hypothesis is that all the different product classes have the same mean prices and delivery times. The $H_0 \& H_1$ hypotheses are set up as the following:

$\boldsymbol{H_0\ Price}$: The product class has no significant influence on the price.
$\boldsymbol{H_1 Price}$: The product class plays significant role on the price.

$\boldsymbol{H_0\ Delivery\ time}$: The product class has no significant influence on the delivery time.
$\boldsymbol{H_1 Delivery\ time}$: The product class plays significant role on the delivery time.

By means of
Response delivery time:

Table 8: Table for MANOVA for Delivery Time

Value	DF	Sum Sq	Mean Sq	F value	Pr(>F)
Class	6	33456906	5576151	629515<	2.2e-16 ***
Residuals	179954	1594005	9		

Response price:

Table 9:Table for MANOVA for Price

Value	DF	Sum Sq	Mean Sq	F value	Pr(>F)
Class	6	5.7156e+13	9.5259e+12	80224 <	2.2e-16 ***
Residuals	179954	2.1368e+13	1.1874e+08		

17 Observations were deleted due to missingness.

The p-Value for both the hypothesis has a very small number. This proves that the product classes influence the delivery time and the product price. These two findings can also be backed up by means of box plots of the delivery time and price per product class as seen below. These graphs support the fact that the delivery time and price of the product differs from class to class. Therefore, both $\boldsymbol{H_0\,Delivery\ time}$ and $\boldsymbol{H_0\,Price}$ is rejected and $\boldsymbol{H_1 Price}$ and $\boldsymbol{H_1 Delivery\ time}$ accepted**.**

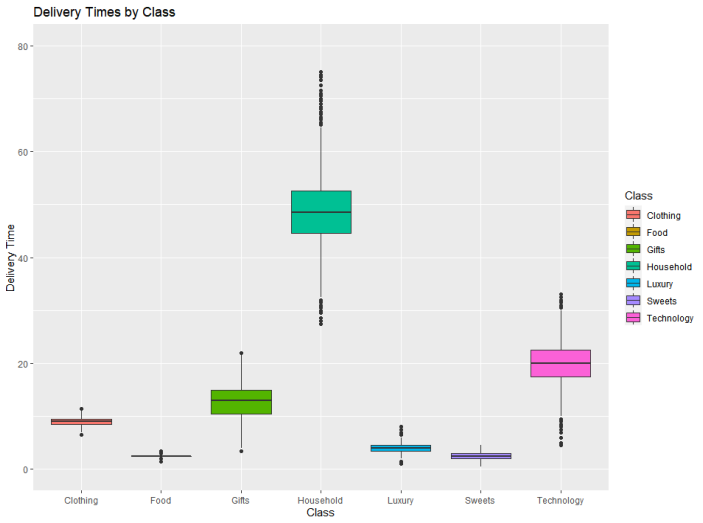

Figure 41: Boxplot for Delivery time per class

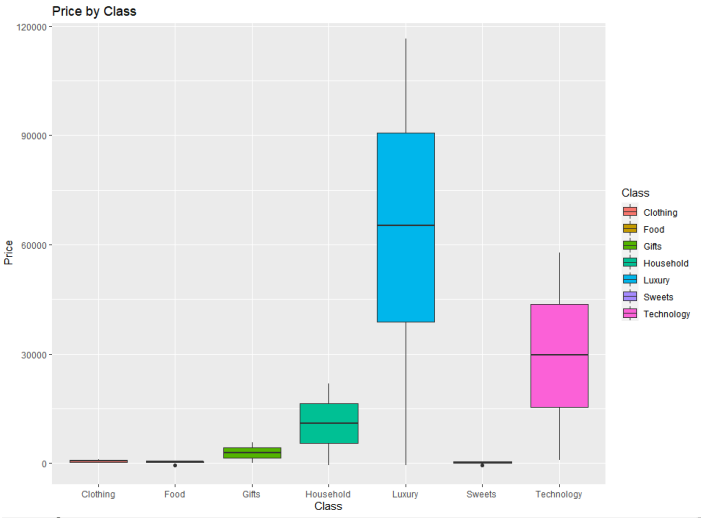

Figure 42: Boxplot of Price per class

Part 6: Reliability of the service and products.

6.1 Problem 6 &7

Problem 6:

Specifications= 0.060 ±0.04cm
Scrapping cost = $45 per part

$$L(x) = k(x - T)^2$$
$$45 = k(0.04)^2$$
$$k = 28125$$

There, the Taguchi loss function for this problem is:
$$L(x) = 28125(x - T)^2$$

Problem 7:

a) New scrapping cost = $35 per part
Therefore, the Taguchi loss function is represented by the following equation:
$$L(x) = k(x - T)^2$$
$$35 = k(0.04)^2$$
$$k = 21875$$
There, the Taguchi loss function for this problem is:
$$L(x) = 21875(x - T)^2$$

b) Process deviation (T) = 0.027
New Taguchi Loss: $L(0.027) = 21875(0.027 - 0.04)^2$= $3.67
Therefore, for a part produced with a deviation of 0.027cm from target, the additional cost will be $3.67.

6.2 Problem 27

a) $R_s = R_1 \times R_2 \times R_3 = 0.85 \times 0.92 \times 0.9 = 0.7038$
With only one machine at each stage there is a reliability of 70.38%

b) $R_s = 1 - (1 - R_1) \times (1 - R_2) \times (1 - R_3)$
$R_s = 1 - (1 - 0.85) \times (1 - 0.92) \times (1 - 0.9) = 0.9988$
By having 2 machines at each stage the reliability is improved to 99.88%
This results in a much higher reliability.

6.3 Vehicles:

```
#6.3
#Binomial distribution
{
  #21 Vehicles
  x <-18
  n<-21
  p<-(1560-22-190)/1560

  p21Trucks<- 1- pbinom(x,n,p, lower.tail = TRUE, log.p = FALSE)

  #22 Vehicles
  xx <-18
  nn<-22
  pp<-(1560-22-190-53)/1560

  p22Trucks<- 1- pbinom(xx,nn,pp, lower.tail = TRUE, log.p = FALSE)

  #21 Drivers
  x3  <-18
  n3<-21
  p3<-(1560-95-6)/1560

  p21Drivers<- 1- pbinom(x3,n3,p3, lower.tail = TRUE, log.p = FALSE)

  data.frame(p21Trucks,p22Trucks,p21Drivers)


}
```

Figure 43: Binomial distribution in R- Format

Therefore, the answer to this question is as follows:

Probability of 21 Trucks being reliable = 0.4420258 = **44.203%**
Probability of 22 Trucks being reliable = 0.6399049 = **63.990%**
Probability of 21 Drivers being reliable = 0.8484614 = **84.846 %**

$$\boldsymbol{Reliable\ days\ per\ year = Truck\ reliability \times Driver\ reliability\ \times Days\ per\ year}$$

Therefore, the reliable days per year for **21 trucks and 21 Drivers:**

$$Reliable\ days\ per\ year = 0.4420258\ \times 0.8484614\ \times 365 = 136.89\ Days\ per\ year$$
$$\approx \boldsymbol{136\ Days\ per\ year}$$

Therefore, the reliable days per year for **22 trucks and 21 Drivers:**

$$Reliable\ days\ per\ year = 0.6399049\ \times 0.8484614\ \times 365 = 198.17\ Days\ per\ year$$
$$\approx \boldsymbol{198\ Days\ per\ year}$$

The days are rounded down seeing that the number of days can only be integer values.

Conclusion

The company is making the most profit from the Technology and Luxury class. With this being said the management team can run more promotions on these classes. They can also make their margins bigger for these classes seeing that there is such a bid demand for Technology and Luxury items.

The insights drawn from the descriptive statistics can also be used by management to improve their overall sales by running the right type of promotions by looking at the most effective reasons for buying per class. This information can be used to target the right group customers for the according products.

In their delivery department there is a few recommendations that can be made. Firstly, the delivery time for Household items can be improved. This will have a positive effect on their customer satisfaction and their reputation as an online retail company. Secondly, the delivery times of Gifts and Household items are out of control as seen in part 4 in the SPC analysis. Improvement in their delivery reliability will be beneficial for building more loyal customers in a market where delivery time is of crucial importance according to an article on customer satisfaction in the online shopping industry. (Thangavel, 2015)
And lastly, the optimal delivery time for Technology items were found in part 4.3. A trial period can be run on this class to determine whether this strategy to decrease delivery costs is feasible or not. If found feasible, the same strategy can then be implemented on the other product classes.

The knowledge gained from the MANOVA analysis is also valuable for the company to better understand how the product class effects the price and delivery time of the product.

References

- Frost, J., 2022. *Multivariate ANOVA (MANOVA) Benefits and When to Use It*. [online] Statistics By Jim. Available at: https://statisticsbyjim.com/anova/multivariate-anova-manova-benefits-use/ [Accessed 19 October 2022].
- Hernandez, F., 2015. *Data Analysis with R - Exercises.* [Online] Available at: http://fch808.github.io/Data-Analysis-with-R-Exercises.html [Accessed 28 September 2022].
- Thangavel, S. (2015) *A study on customer satisfaction towards online shopping - researchgate*, *A Study on customer Satisfaction towards Online Shopping*. ResearchGate. Available at: https://www.researchgate.net/publication/329026968_A_Study_on_customer_Satisfaction_towards_Online_Shopping (Accessed: October 20, 2022).