ECSA Project

Quality Assurance

Nicola van der Merwe

23931779

Table of Content

Contents

List of Figures

Introduction

The client data for an online business was given to me to analyze. The data was first cleaned and sorted into two groups: valid data and invalid data. Also, any data that is Not Applicable (NA) to the business, should be removed.

The report consists of the visualization of the data that was given, and a descriptive understanding of the visuals is made thereafter. The rest of the report entails data wrangling, descriptive statistics, and process control, MANOVA tests were done, and reliability calculations conclude the report.

Within the report, I familiarized myself with the data with the necessary help of RStudio to plot graphs of the data. Trends and relationships are identified through the visualization of the data. The company can gain a lot of knowledge of their business through this report by identifying which classes of products are problematic and where the investigation is needed by the sales department of the company.

The focus shifts to the delivery time of the products when the X-bar charts and s-charts are formulated and drawn. By looking at this, the company will have an idea of where to solve problems regarding delivery times for certain classes of products. Hereby, the quality of service and products will also increase.

PART 1: Valid & Invalid Data

Valid data (n)

The valid data instances are instances where the data does not contain any missing data. This part of the data is separated from the invalid dataset. With RStudio, the amount of valid data instances equals 179 983 out of 180 000 instances. This is a large observation and contains 99.99% of the data. Therefore, only the first couple of instances are shown in the table below to illustrate that the valid data, not containing any missing values.

In row 12 345 the first missing value is found, where the price is saved as 'NA'. This row will be scratched and row 12 346 will be the having the new values for row 12 345. This is shown in figure 1 below.

X	ID	AGE	Class	Price	Year	Month	Day	Delivery time	Why Bought
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	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	108953.5	2029	11	13	4	Recommended
8	32624	58	Sweets	389.62	2025	7	2	2	Recommended
9	51401	82	Gifts	3312.11	2025	12	18	12	Recommended
10	96430	24	Sweets	176.52	2027	11	4	3	Recommended
11	87530	33	Technology	8515.63	2026	7	15	21	Browsing
12	14607	64	Gifts	3538.66	2026	5	13	13.5	Recommended
13	24299	52	Technology	27641.97	2024	5	29	17	Browsing
14	77795	92	Food	556.83	2025	6	3	3	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	Recommended
18	39674	69	Technology	22315.17	2026	8	20	20.5	Recommended
19	98694	74	Sweets	546.48	2025	5	9	2	Recommended
20	99187	54	Luxury	81620.21	2027	9	14	3	Recommended

Figure 1: Valid data

Old row:

12344	90260	34	Luxury	42891.66	2025	8	4	4	Recommended
12345	18973	93	Gifts	NA	2026	6	11	15.5	Website
12346	92286	32	Technology	38167.24	2028	7	6	19.5	Website

Figure 2: Old row

New row:

12343	27986	37	Clothing	712.19	2021	10	10	9	Recommended
12344	90260	34	Luxury	42891.66	2025	8	4	4	Recommended
Removed									
12345	92286	32	Technology	38167.24	2028	7	6	19.5	Website
12346	89263	44	Clothing	891.71	2021	7	2	8.5	Recommended

Figure 3: New row

Row 12 345 now took the place of the (old) row 12 346.

Invalid data (m)

The invalid data instances are instances where the instance contains missing values. In the dataset, seventeen instances out of the 180 000 are seen to have missing values. Figure 4 below shows all the instances where the price was not given in the dataset.

ID	AGE	Class	Price	Year	Month	Day	Delivery.time	Why.Bought
18973	93	Gifts	NA	2026	6	11	15.5	Website
81959	43	Technology	NA	2029	9	6	22	Recommended
71169	42	Technology	NA	2025	1	19	20.5	Recommended
67228	89	Gifts	NA	2026	2	4	15	Recommended
88622	71	Food	NA	2027	4	18	2.5	Random
18748	48	Clothing	NA	2021	4	9	8	Recommended
89095	65	Sweets	NA	2029	11	6	2	Recommended
62209	34	Clothing	NA	2021	3	24	9.5	Recommended
63849	51	Gifts	NA	2024	5	3	10.5	Website
51904	31	Gifts	NA	2027	7	24	14.5	Recommended
79732	71	Food	NA	2028	9	24	2.5	Recommended
40983	33	Food	NA	2024	8	27	2	Recommended
64288	25	Clothing	NA	2021	1	24	8.5	Browsing
70761	70	Food	NA	2027	9	28	2.5	Recommended
33583	56	Gifts	NA	2022	12	9	10	Recommended
60188	37	Technology	NA	2024	10	9	21.5	Website
68698	30	Food	NA	2023	8	14	2.5	Recommended

Figure 4: Invalid data

PART 2: Descriptive Statistics

Analysis of the valid data set: Continuous features

The following table shows the values of certain numerical features in the data set:

Variable	Min	Q1	Mean	Med	Q3	Max	Outlier	n_miss	card_con
X	1.00	45004.25	90002.55	90004.5	134999.75	18000	0	0	179978
Age	18.00	38.00	54.57	53.00	70.00	108	0	0	91
Price	36.65	482.31	12294.10	2259.63	15270.97	116619	22171	0	78832
DelTime	0.50	3.00	14.50	10.00	18.50	75	17516	0	148

The feature, age, ranges from 18 years old up to 108 years old, with an average of fifty-five. This means that most sales are made by customers around the age of 55 years old. This age is then clearly the target audience. The substantial difference between the minimum and maximum Prices, can be because different classes of products are sold for various prices. Luxurious items and Sweets are not in the same price range and that is why there is such a big gap between R36.65 and R116 619. The major difference in the minimum and maximum of the delivery times is shocking and should be investigated. The minimum of half an hour and maximum of 75 hours are days apart from each other and it is a problem in the company that should be solved. The X-feature has the same number for the cardinality as it has for the number of instances. This feature is therefore disregarded. The outliers for the Price and Delivery Time class are necessary to investigate and would be kept in the data and further analysis should take place.

Categorical features

Class name	CountC	Miss_val	Card	Modes	Mode_freq	Mode_perc
ID	179978	0	15000	41842	27	0.0150018
Class	179978	0	7	Gifts	39149	21.752103
Year	179978	0	9	2021	33443	18.581716
Month	179978	0	12	12	15225	8.4593673
Day	179978	0	30	17	6126	3.4037493
WhyBought	179978	0	6	Recommended	106985	59.443376

Gifts are the most popular class that is sold and the most popular time that this is bought is on the 17th of December in 2021. This is in the middle of the holiday season and is a reason for the sales to be so popular. And the most common reason customers buy products from the company is because of a recommendation given to them.

Graphs

Delivery time vs Class of Product

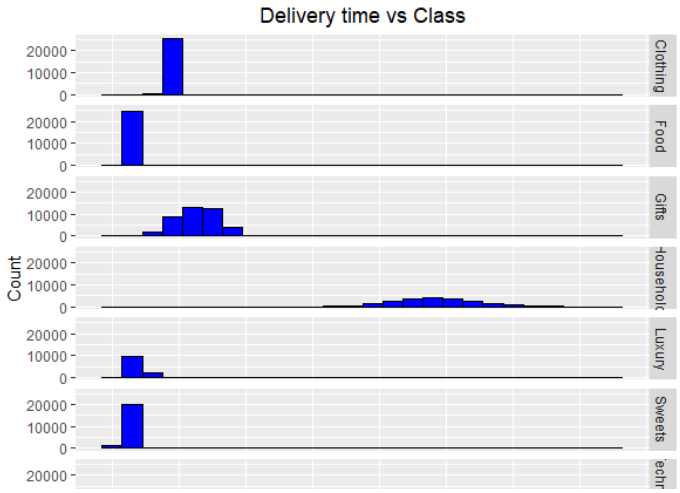


Figure 5: Delivery time per Class

The graph shows that food, luxury items, and sweets are delivered the fastest while households take the longest time for delivery to take place. The company needs to figure out why deliveries of household products are taking so long, to not lack service quality. Foods are often delivered quickly since they are portable, easily made, and transported, and should be delivered fresh. Household goods typically require a longer time and more labor to load onto the van and transport to the customer's location, which may account for the longer delivery times. Household products are dispersed throughout delivery windows, maybe because they include both large and heavy (time-consuming) and little parts.

Monthly Count for sales

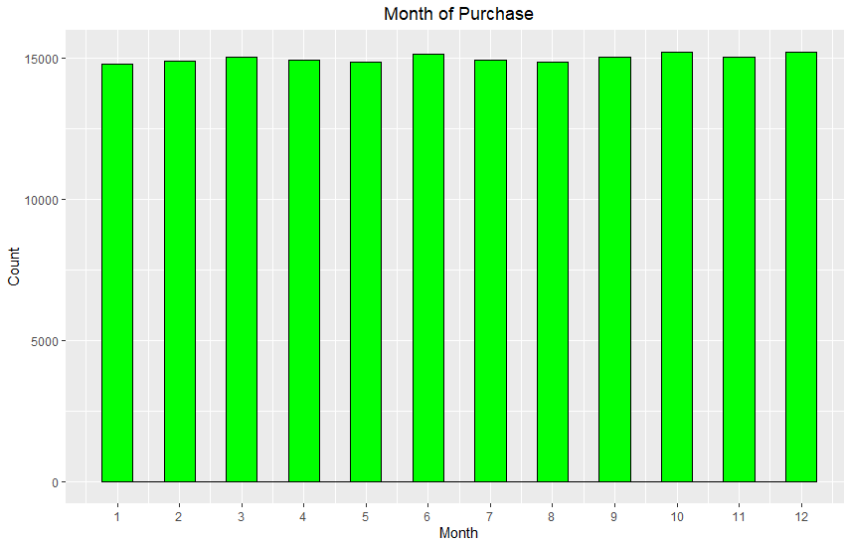


Figure 6: Monthly count for sales

There is a uniform distribution between the months during the year. No trends are visibly identified for months of sales.

Day Count for sales

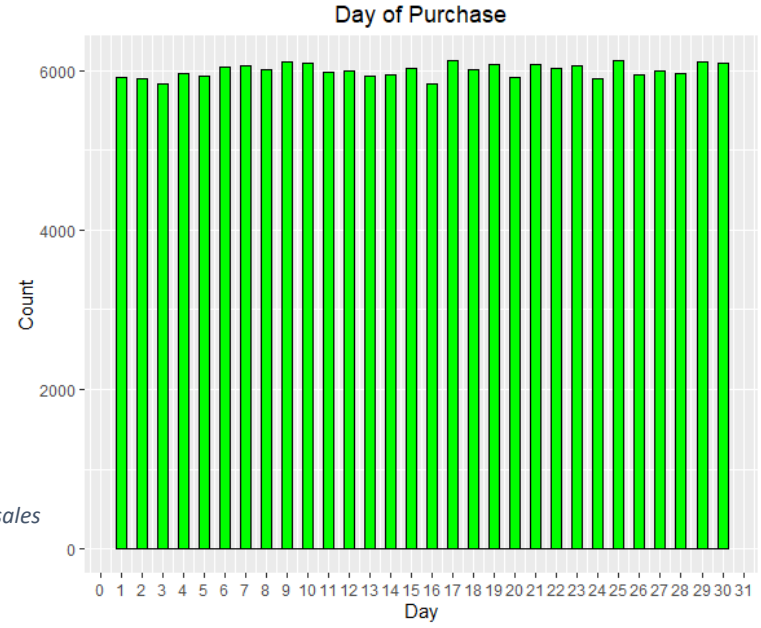


Figure 7: Day count for sales

Day Count vs Class:

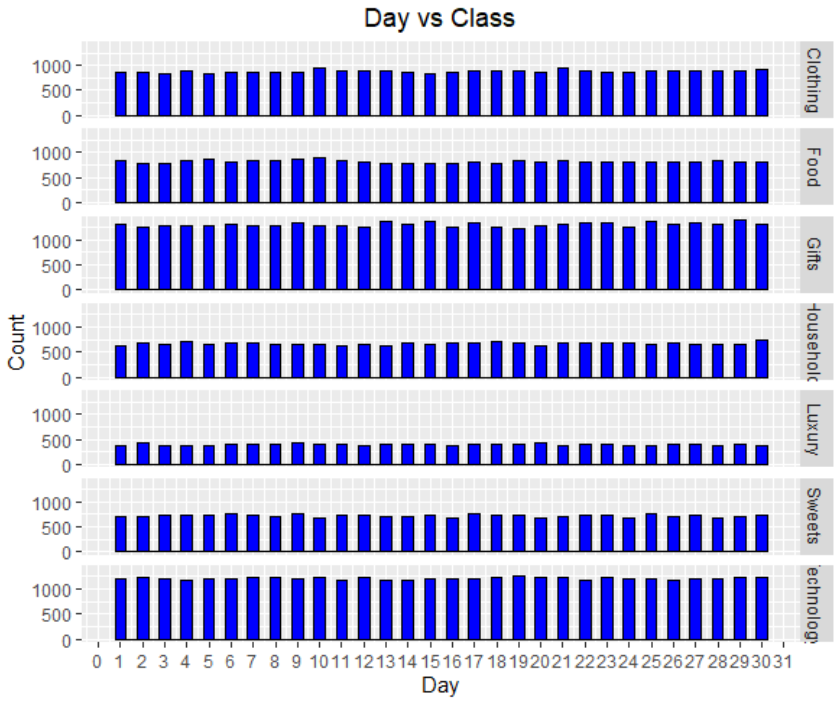


Figure 8: Day vs Class

The sales count is also uniformly distributed throughout each day. The sales are on average equal each day with the minimum sales at the beginning of the month. Each class has uniform distributed counts throughout all the days of the month. There is no trend to be identified for each day of sales.

Age vs Class of Product

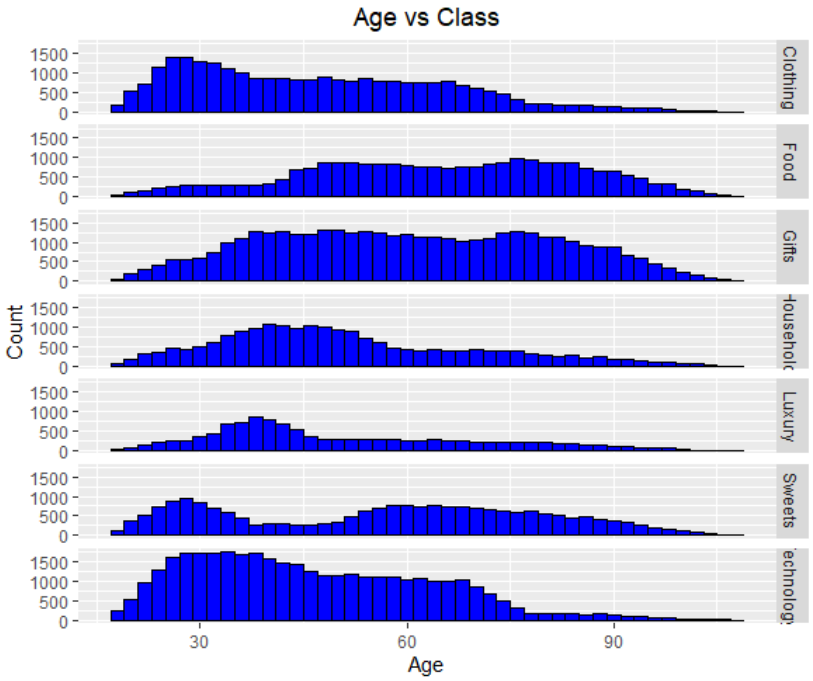


Figure 9: Age vs Class

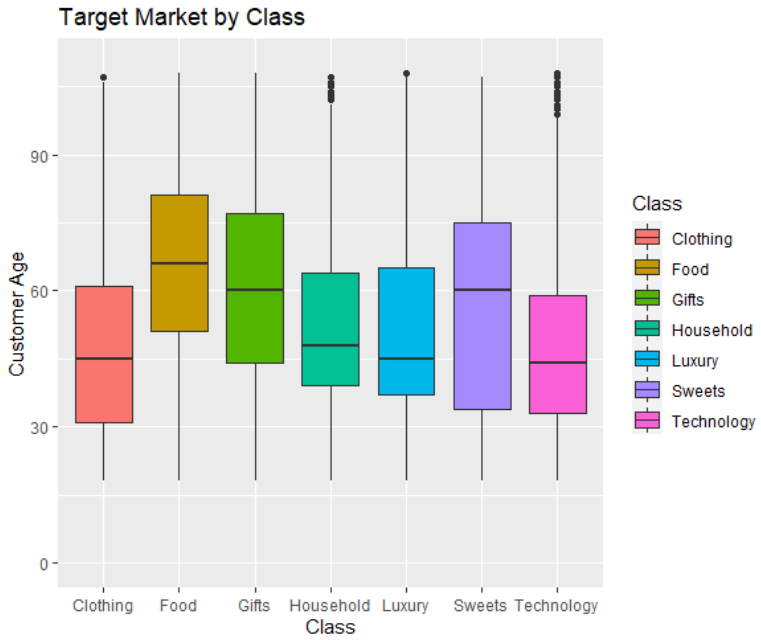


Figure 10: Target Market by Class

All classes include customers who are older than ninety-eight. It is assumed that retirement communities use online marketplaces to buy goods more often. The age range between 28-33 is most common for the sales of clothing, which is exponentially distributed. This makes sense logically, given that younger people are more likely to purchase fashionable clothing. The most common age range for purchasing food is between 45-80 years old. People over sixty-five are more likely to buy food online than to go to the store to buy it themselves. Gifts are distributed across all age groups, given that any person at any age buys gifts often. Households are exponentially dispersed, with 35-55 years old being the most popular. The age of around thirty-five is when luxury is dispersed most frequently.

Technology is unimodal (right tailed in figure 9), with the most common age range falling between 28-40 years old. Sales managers should look for techniques to advertise technology to the younger age group since it has the youngest mean age group out of all the classes. The study's findings indicate that younger and middle-aged consumers favor luxury, technology, and home goods.

Figure 11: Sales per Class per Month

Looking at the graphs above, all classes have uniformly distributed sales during all the months. There are no trends that can be picked up from the graph but show sales remain constant throughout the year.

Year vs Class:

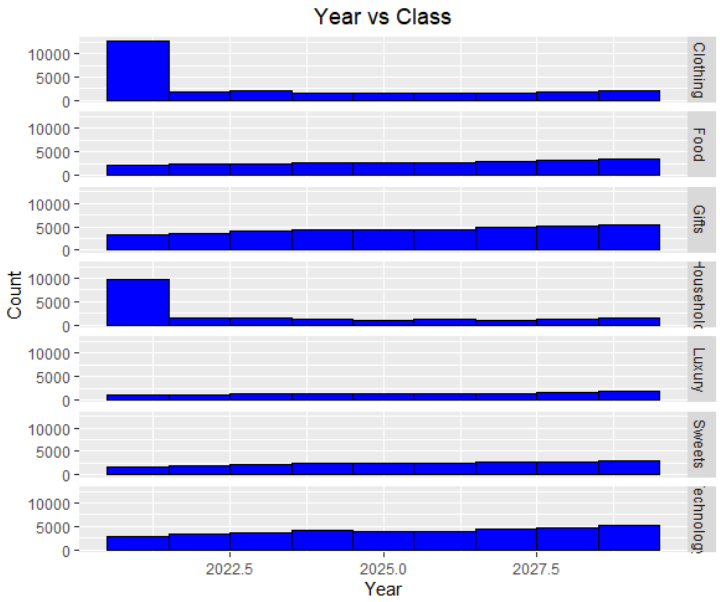


Figure 12: Year vs Class

Sales count for household and clothing peaked in the first years and thereafter stayed constant. Whereas for the other class, the count is uniformly distributed, and no trends could be picked up from the distribution. The count for technology seems to be increasing throughout the years, seeing that technology is a class that improves as the years pass by.

Why Bought Count

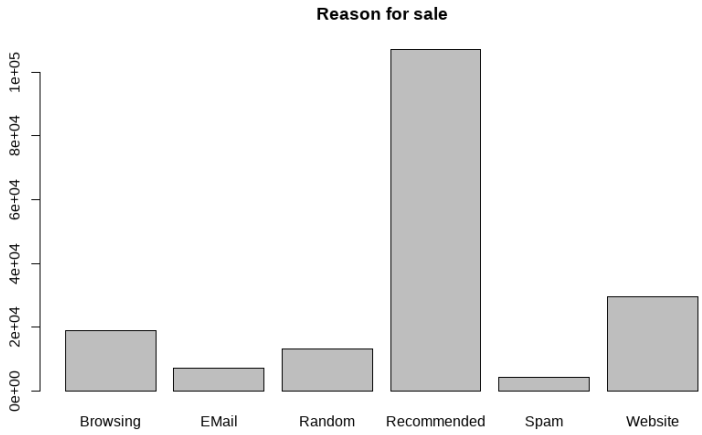


Figure 13: Reason for sale

The most frequent reason for buying a product is a recommendation from another person. Spam is the least common reason. Customers should receive high-quality services to increase word-of-mouth marketing of the products.

Class of Product Sold

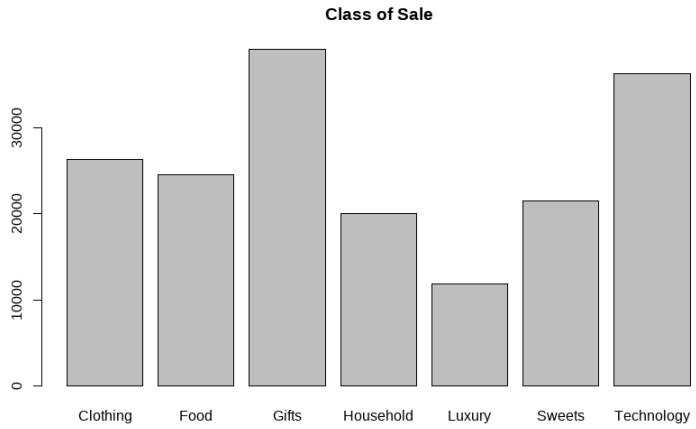


Figure 14: Class of Sales

Gifts are the most often purchased item, followed by technology. Luxury goods are the least often purchased products, because of the prices that are so high for these items. Since gifts and technology are the most popular items purchased and should have the highest quality. Clothing, Food, Household, and Sweets products have an average amount of sales, and the quality of these classes should be kept to ensure that the sales do not decrease, hopefully increasing with time.

Delivery time vs Price

In figure 16 shown above, we can see that by lowering the price, the delivery time would increase. The option for faster delivery will therefore result in a higher price asked per item. The luxury of having a fast delivery would therefore come a at cost. The cost can go to the company to improve their delivery times to have an all-over more successful business.

Figure 15: Delivery Time vs Price

Delivery time of technology

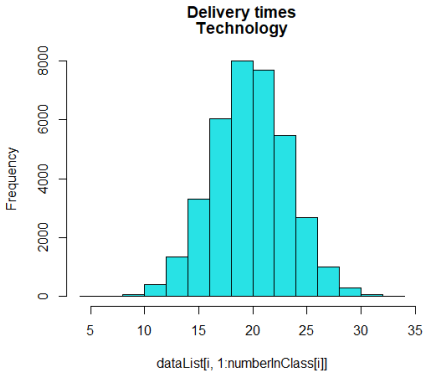


Figure 16: Delivery time for Technology

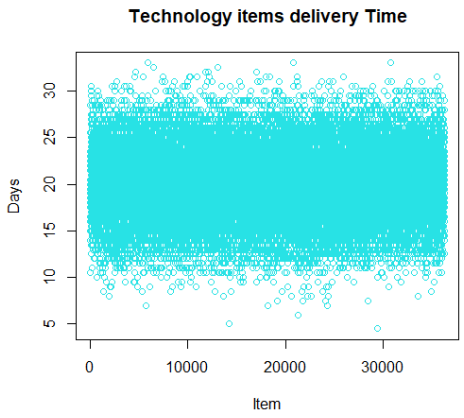


Figure 18

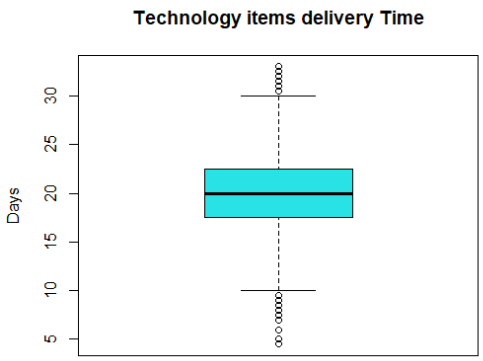


Figure 17

As seen above in the figures for the delivery time of the class of technology, the data is normally distributed with an almost centered mean between the minimum and maximum of the data (seen in figure 19).

Process capability indices

USL = 24

LSL = 0

The LSL is zero since the lower specification limit would ensure that all the data points that are used are positive, and that no instances contain any negative values. This is also the reason all negative instances were removed from the dataset before any of the graphs were obtained. Any data points that are less than the LSL, should be out of bounds. Therefore, the choice of setting the LSL equal to zero is logical and gives us an accurate display of all the samples in the dataset.

The following was obtained:

CP = 1.142207 = 1.14 **CPL = 1.880884 = 1.88**

CPU = 0.4035293 = 0.40 **CPK = 0.4035293 = 0.40**

Looking at the obtained values, the process is capable due to the CP being greater than one (CP>1). CPK is less than CP (CPK < CP) and that indicates that the process is not centered between the specified limits. The process, therefore, needs to be improved by moving the mean to the left.

PART 3: Statistical process control (SPC) for the X&s-charts

An X&s chart showing delivery times is created with thirty samples, each with 15 Sales. The data must first be arranged in chronological order before being used to compute and build the charts. The year, month, and day are used to sort the data from oldest to newest.

X-charts values

Class	UCL	U2Sigma	U1Sigma	CL	L1Sigma	L2Sigma	LCL
Technology	22.974616	22.107892	21.241168	20.374444	19.507721	18.640997	17.774273
Clothing	9.404934	9.259956	9.114978	8.970000	8.825022	8.680044	8.535066
Household	50.248328	49.019626	47.790924	46.562222	45.333520	44.104818	42.876117
Luxury	5.493965	5.241162	4.988359	4.735556	4.482752	4.229949	3.977146
Food	2.709458	2.636305	2.563153	2.490000	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 19: X-bar values

S-charts values

Class	UCL	U2Sigma	U1Sigma	CL	L1Sigma	L2Sigma	LCL
Technology	5.1805697	4.5522224	3.9238751	3.2955278	2.6671805	2.0388332	1.4104859
Clothing	0.8665596	0.7614552	0.6563509	0.5512465	0.4461422	0.3410379	0.2359335
Household	7.3441801	6.4534101	5.5626402	4.6718703	3.7811003	2.8903304	1.9995605
Luxury	1.5110518	1.3277775	1.1445032	0.9612289	0.7779546	0.5946803	0.4114060
Food	0.4372466	0.3842133	0.3311800	0.2781467	0.2251134	0.1720801	0.1190468
Gifts	2.2463333	1.9738773	1.7014213	1.4289652	1.1565092	0.8840532	0.6115971
Sweets	0.8353391	0.7340215	0.6327039	0.5313862	0.4300686	0.3287509	0.2274333

Figure 20: S-chart values

Graphs for the first thirty samples

Sweets X-bar

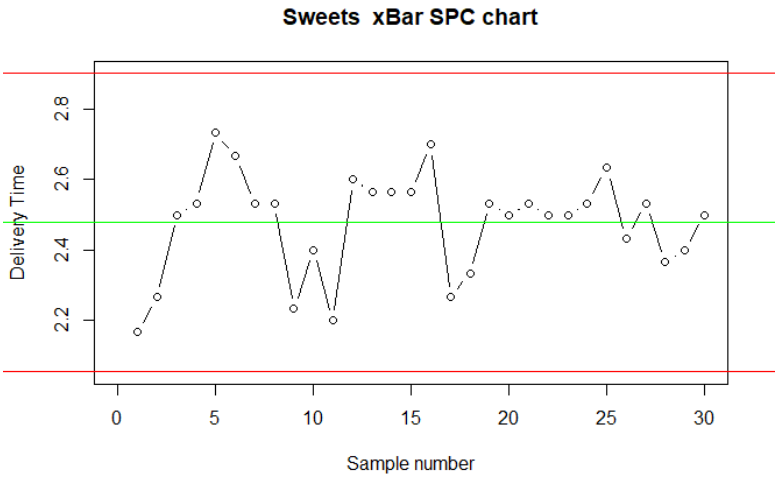


Figure 21: Sweets

Technology X-bar

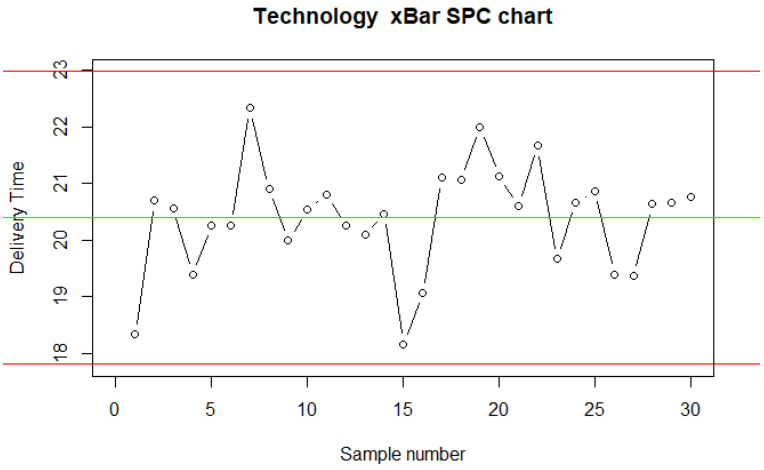


Figure 22: Technology

Clothing X-bar

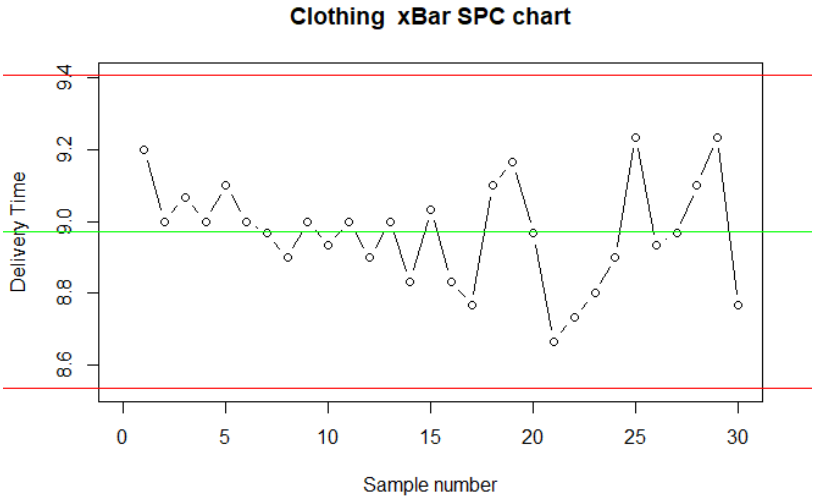


Figure 23: Clothing

Household X-bar

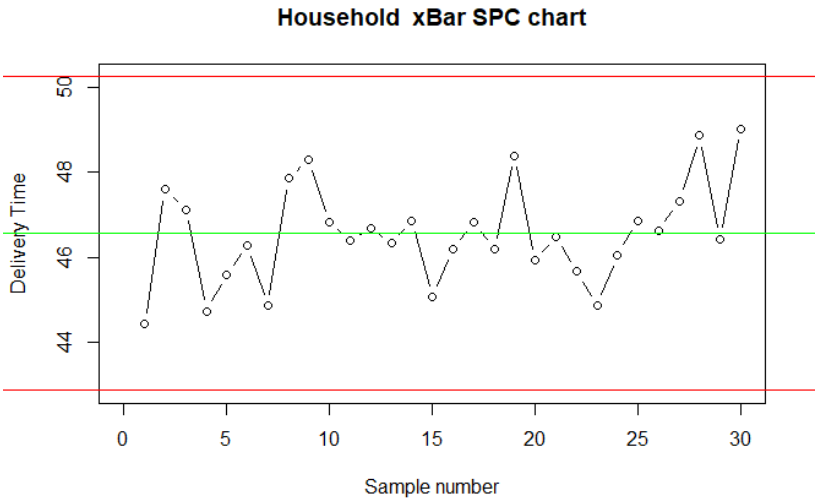


Figure 24: Household

Luxury X-bar

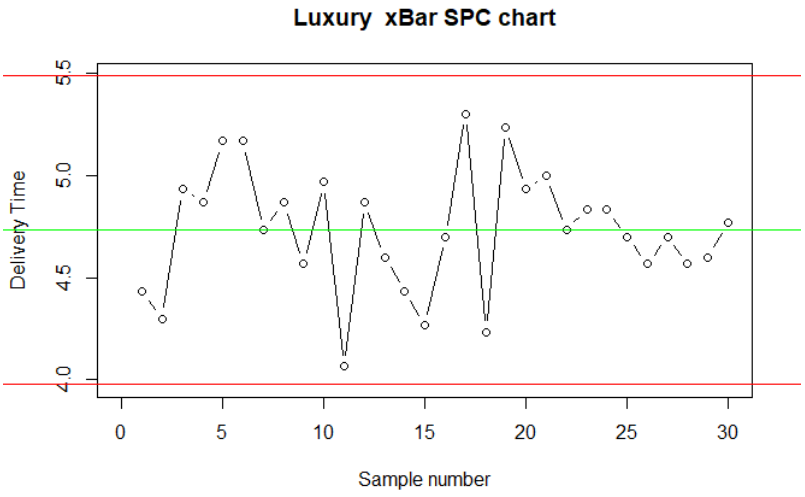


Figure 25: Luxury

Food X-bar

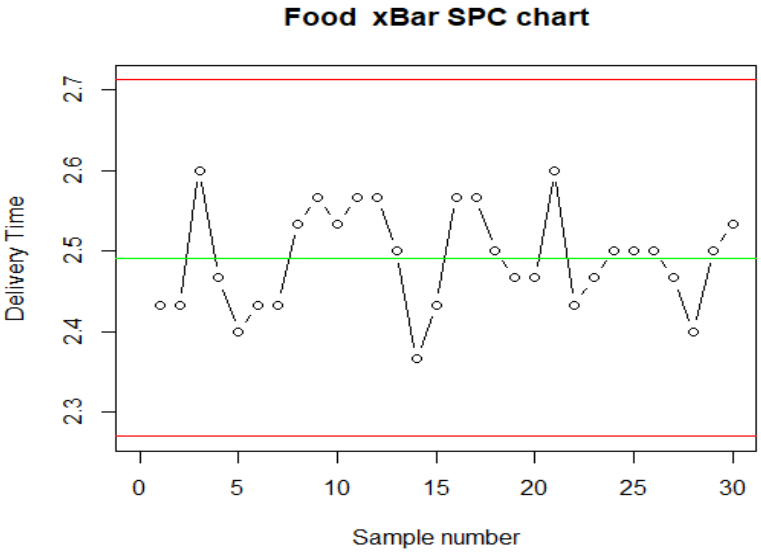


Figure 26: Food

Gifts X-bar

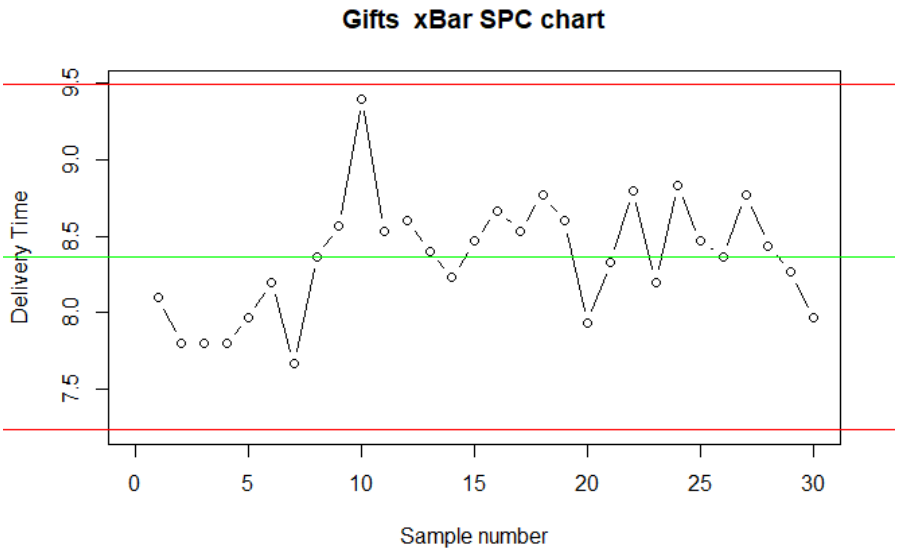


Figure 27: Gifts

As seen above in figures 22 to 28, the first thirty samples of all the different classes are distributed on the x-bars and all the samples are between the control limits. Therefore, the conclusion can be made that all the different classes are under control if you look at the graphs printed out above. Not one of the classes has data that goes above or below the red lines (that indicate the UCL and LCL).

After seeing that all the charts are under control, the graphs on all samples are drawn now. In the x-bar and s-charts that will follow, the limits are shown through a shaded grey area on the chart. The graph will indicate an instance that is out of control/limits, with a red dot. When there are multiple red dots present, the company will know that they should investigate what exactly causes the problem and should find a reason of why the product class

Graphs for all samples

Gifts

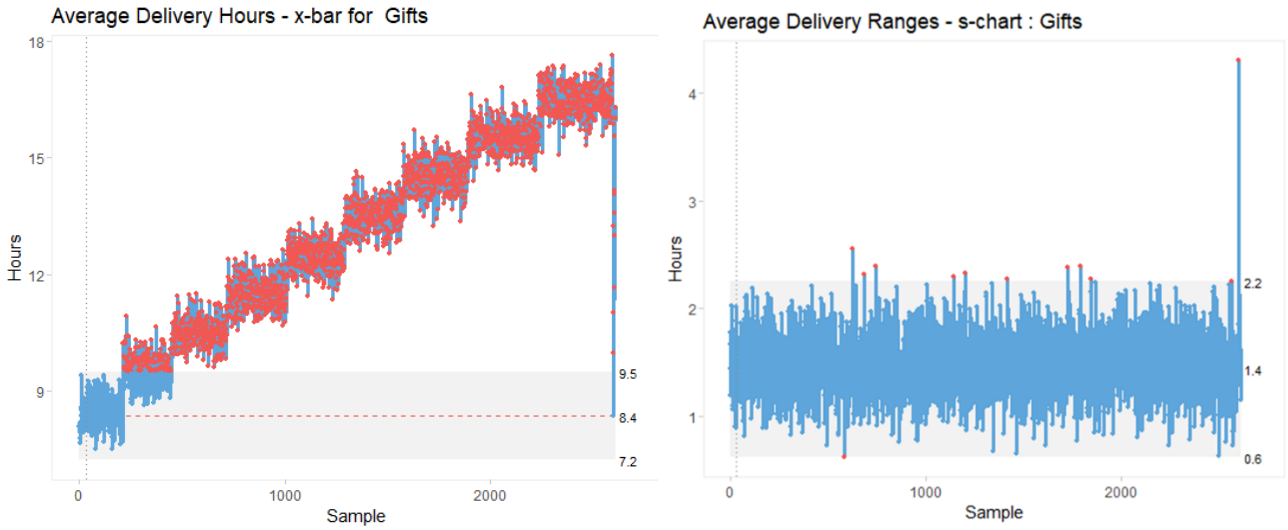


Figure 28: Gifts x- and s-chart

The delivery time for gifts increased and an investigation should be done to find the reason behind the increase. Because of this increase, the delivery time for gifts is not in control and not stable. One reason might be that the company experienced an increase in demand and cannot deliver as fast as possible. A logistical problem could have occurred where more products should be delivered, but capacity does not allow everything to be shipped. Looking at the s-chart in figure 29, there are only a few samples that are out of bounds. Therefore, the conclusion made on the x-bar is appropriate.

Technology

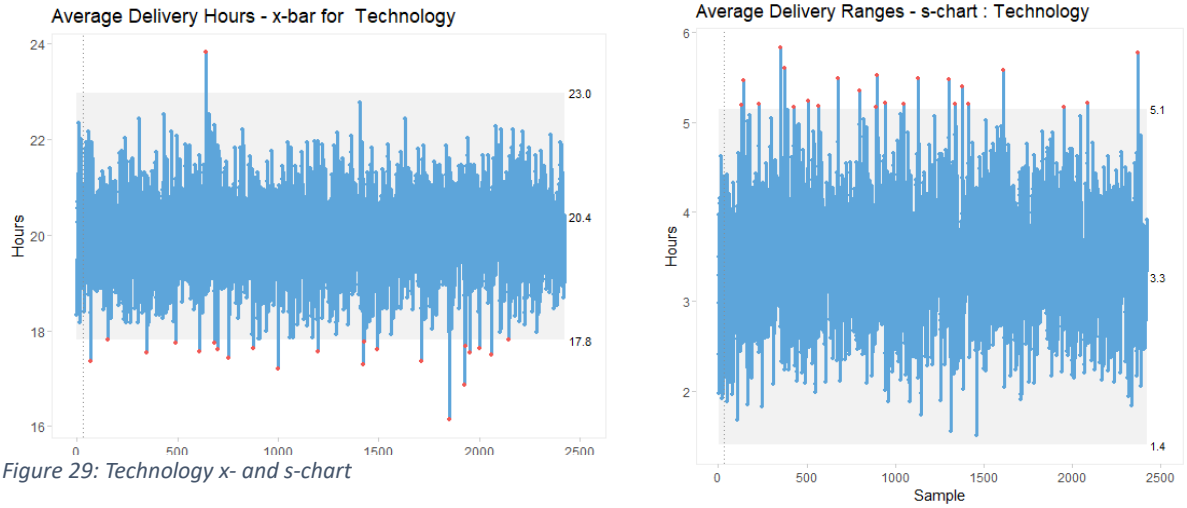
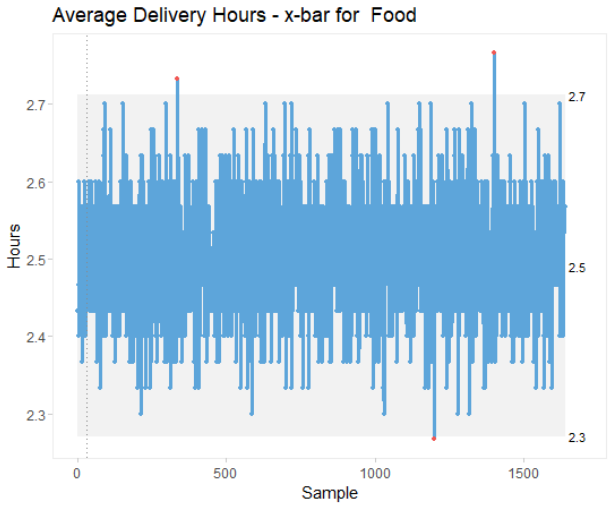


Figure 29: Technology x- and s-chart

The majority of all the samples are under control and do not exceed the limits too much. The s-chart shows that most of the samples are under control, which makes the conclusion made on the x-bar appropriate.

Food



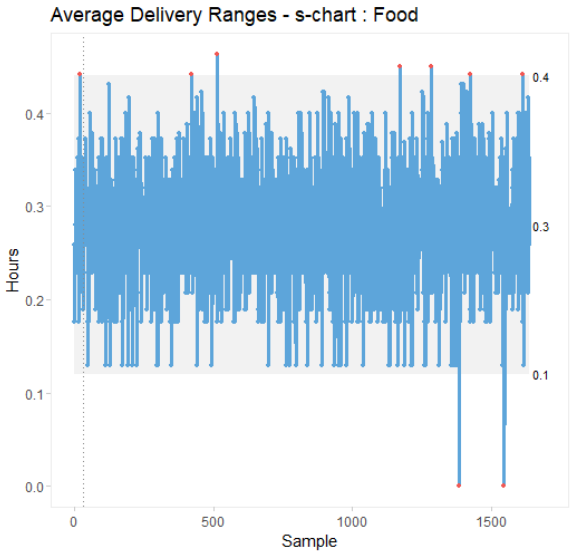
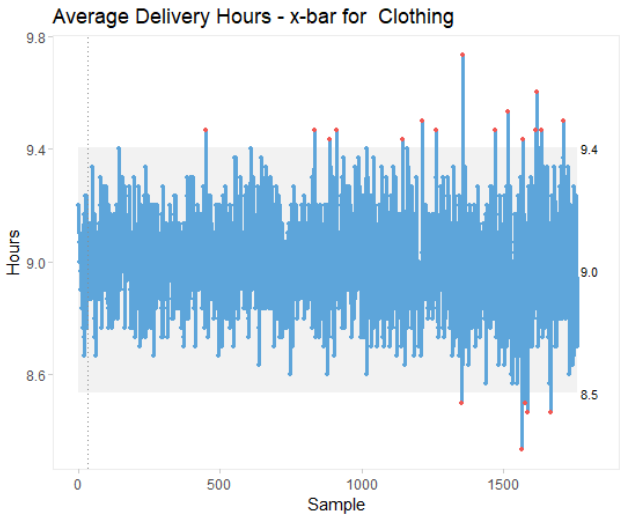
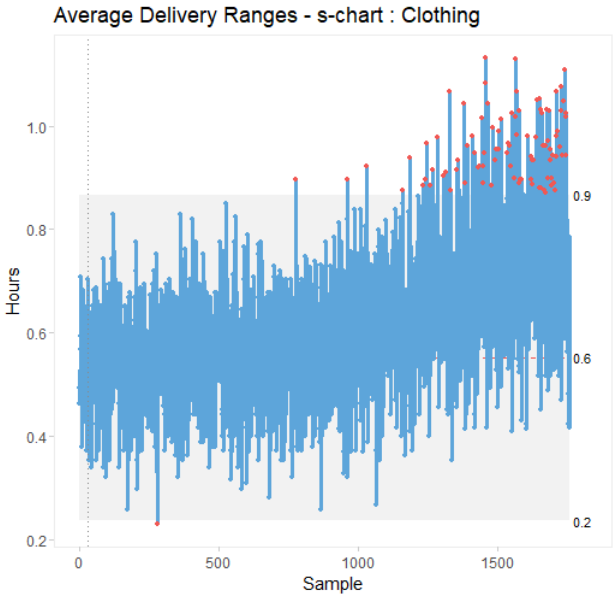


Figure 30: Food x- and s-chart

The class, Food, seems to be under control with only a few instances where the sample is not under control. The s-chart looks stable and under control, with only a few instances that are out of control and therefore concluding that the conclusion made on the x-bar is appropriate.

Clothing





Most of the samples are between the control limits. Clothing is under control, but there are some odd occasions where samples are beyond the limits. A likely reason for this cause could be due to seasonal changes. There are quite a few samples out of the control limits for the S-bar chart, but if those samples that are out of bounds would be removed, the X-bar chart will still indicate the same results. Therefore, the conclusion of the X-bar chart is appropriate.

Sweets

Figure 32: Sweets x- and s-chart

Sweets are under control, with only a few samples that cross the control limits. In the s-chart, there are few samples that cross the limits and therefore the conclusion made on the x-bar seems to be appropriate.

Household

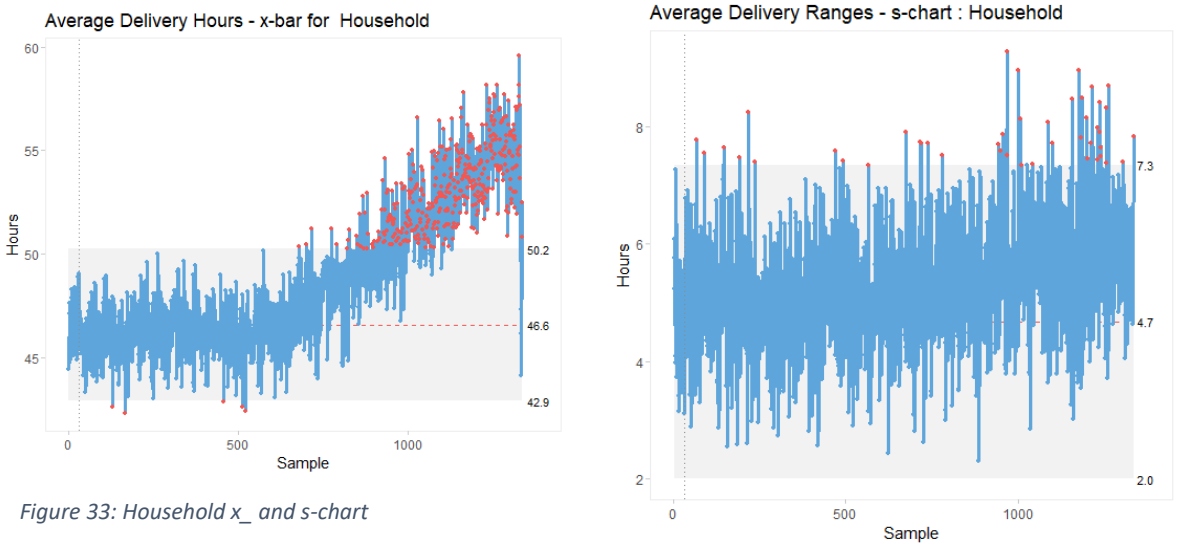


Figure 33: Household x_ and s-chart

The delivery time for households increased rapidly and the reason behind this should be investigated. A reason behind this can be due to other classes that had a fast delivery time and caused the household goods to be delivered slower. The s-chart only has a few samples that are not between the limits, and this indicates that the conclusion made on the x-bar is appropriate.

Luxury

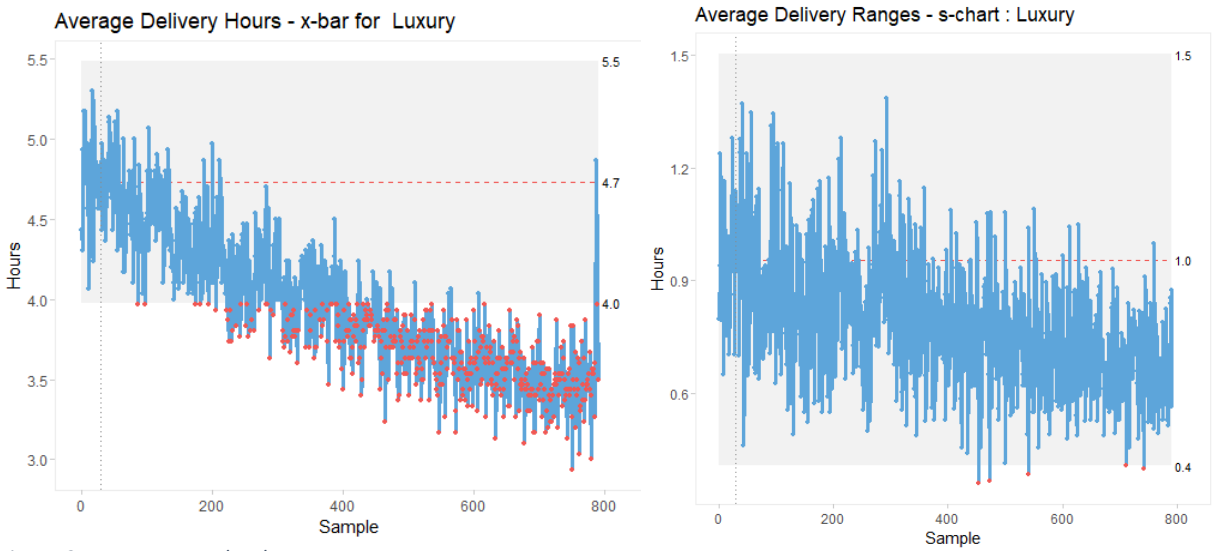


Figure 34: Luxury x- and s-chart

The delivery time for luxury products decreased significantly. The reason behind this should be investigated by the company. A likely reason could be since luxurious products are expensive and of high value. The time between when the product is bought and when it is delivered should be much faster, to prevent any damage done to the product. This rapid decrease in delivery time for luxury products can also be a reason household products are being delivered much slower. The s-chart shows that the samples are under control and the conclusion made on the x-bar is appropriate.

PART 4: Optimizing the delivery processes

4.1 Inspection of x-bar: x-bar sample means outside of the outer control limit

Samples out of control:

Class	Total found	1st	2nd	3rd	3rd Last	2nd Last	Last
Clothing	20	450	832	885	1635	1667	1713
Household	393	128	165	457	1331	1336	1337
Food	3	336	1197	1401	NA	NA	NA
Technology	23	67	152	344	2000	2062	2147
Sweets	3	1099	1238	1351	NA	NA	NA
Gifts	2288	212	215	217	2607	2608	2609
Luxury	442	87	97	175	787	790	791

Figure 35: Samples that are out of control

By looking at the data displayed in figure 25 above, one can see that clothing, food, technology, and sweets are in control. Household, luxury, and gifts are out of control, because of the multiple samples that were outside of the control limits. The problems that cause this should be further investigated to determine what the exact problem is that causes the problem.

Samples of S-chart that are between-0.3 and 0.4 sigma

Class	maximum between -0.3 & 0.4 sigma	Position of first	Last Sample position
Clothing	5	665	665
Household	4	253	761
Food	5	752	905
Technology	6	1191	1191
Sweets	5	692	692
Gifts	6	1334	1334
Luxury	3	230	230

Figure 36: S-chart samples between -0.3 and 0.4 sigma

The maximum number of samples that are between -0.3 and 0.4 sigma, is equal to 6. This is low, indicating a lot of samples for all the different classes are beyond the -0.3 and 0.4 sigma control limits.

The classes Gifts and Technology have the most consecutive samples between -0.3 and 0.4 sigma control limits. This means they will therefore be most stable within the specific limits, compared to the other classes. In the figure below the green line represents the s-chart mean value.

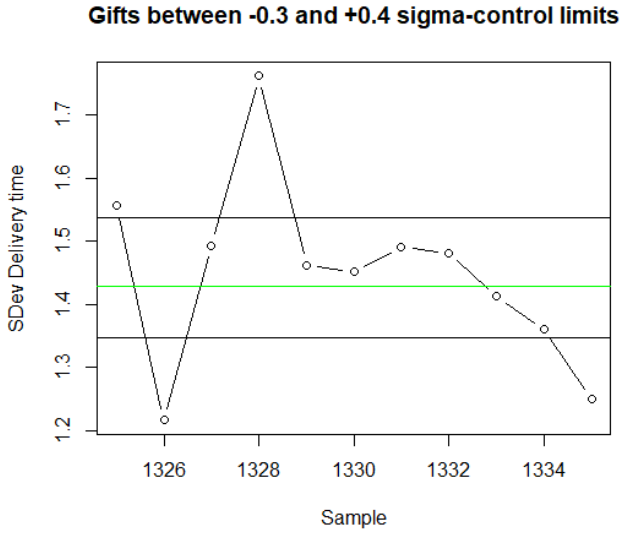


Figure 38: Gifts

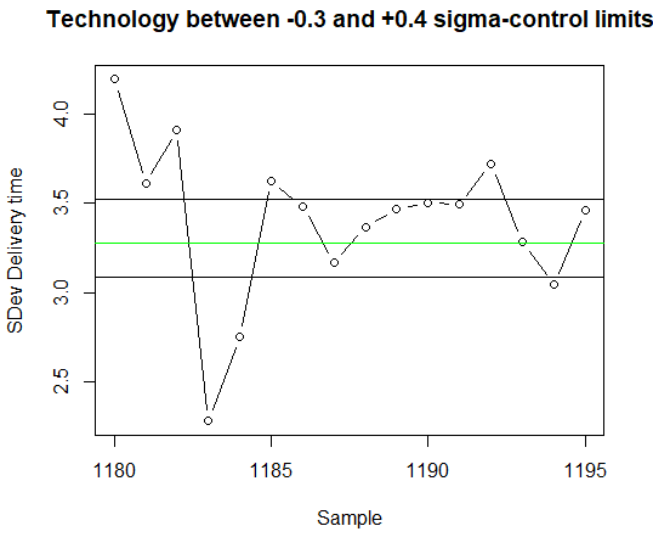


Figure 37: Technology

4.2 Type l error

Rejecting the null hypothesis when it is true is a Type I mistake. It entails concluding outcomes that are statistically significant when, in fact, they were just the consequence of chance or unrelated causes.

The H0 hypothesis is used to calculate a type l error:

H0 = The mean of all the classes is processing between the limits

H1 = The mean of all the classes is not processing within the limits.

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

Figure 39

The conclusion was made that a Type I Error has been made. The probabilities of making a type I error for A and B are shown in figure 41 below.

Rule	Probabilities	Probability %
A	0.00269979606326019	0.269979606326019
B	0.131659416692719	13.1659416692719

Figure 40: Probabilities

4.3 Minimizing technology delivery costs

To successfully determine the minimum delivery cost for the technology, it is needed to compare the costs of all the different hours and then find the exact hour with the lowest cost associated with it. The results can be obtained by going through all the delivery times and their associated cost.

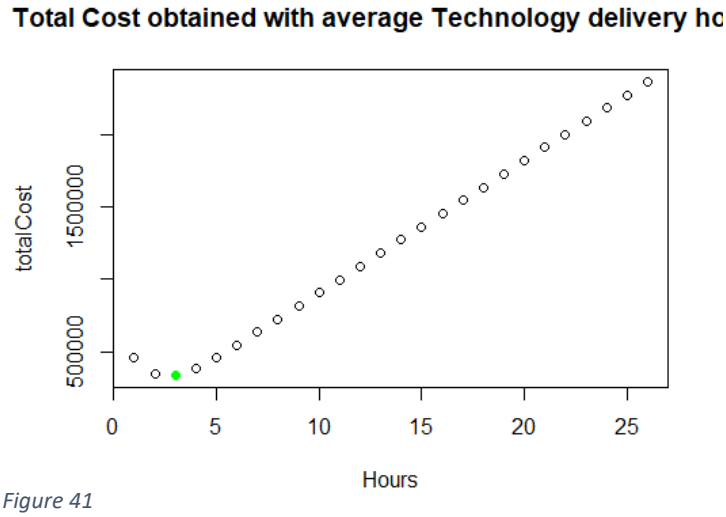


Figure 41

As seen in the figure above, the best delivery time center is three hours, indicated by the green dot in the graph above. The cost of R340 870 will be associated with the delivery of a product if the delivery time of each delivery is reduced by 3 hours (therefore also reducing the weighted average from 20 to 17 hours). Delivery times that would exceed 26 hours would then be more expensive compared to reducing the delivery times to 3 hours. Figure 42 has a similar form compared to a Taguchi loss function graph, showing that there is a similarity. The loss is present although the product is within specifications and by finding the exact point on the graph, the amount of time lost during delivery can be minimized. The customer will therefore be completely satisfied with the quality of service, because of the delivery that is perfectly on target.

4.4 Type ll Error

The likelihood of mistakenly failing to reject the null hypothesis when it does not apply to the entire population is known as a type II error. A type II error is a false negative.

The probability of making this type of error for the technology class if the new average equals twenty-three, is 0.48819. This means that there is a 48.82% chance of making this error. This can cause dissatisfaction to the customers because the delivery time will be longer than what was stated to them.

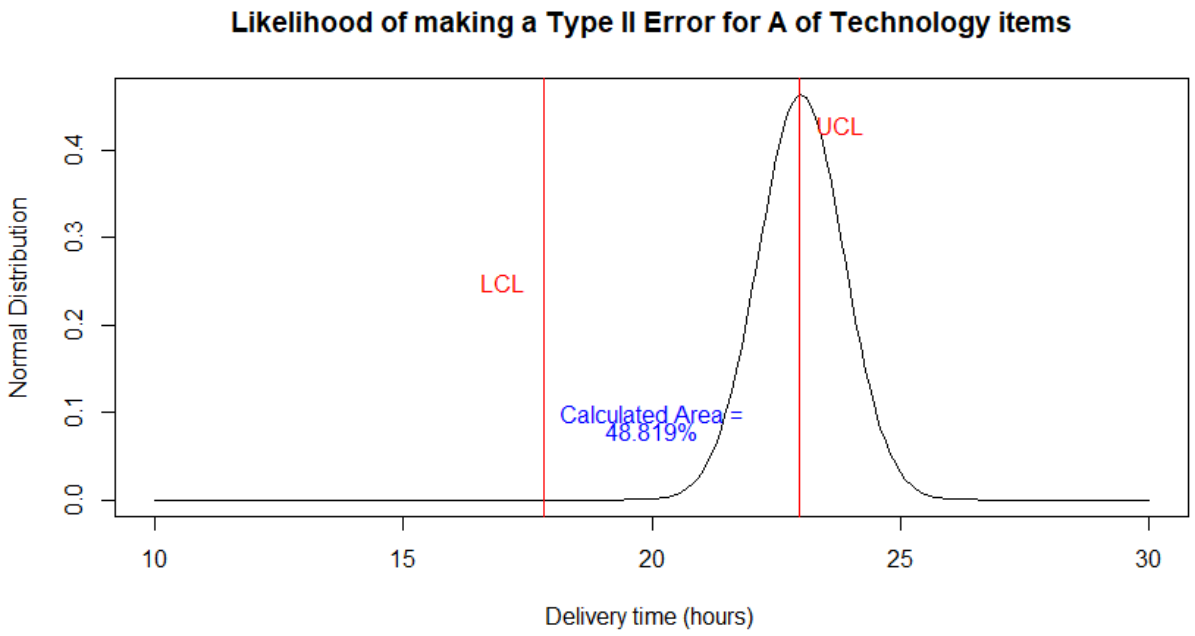


Figure 42: Type ll Error probability

PART 5: DOE and MANOVA

Determining if there is any connection between the dependent variables and the independent factors, involves using the MANOVA table. In the table, the p-value is provided for each dependent variable. I chose p = 0.05 because it is the most popular p-value universally.

First hypothesis

Dependent variable: Price, Delivery time & Age

Independent variable: Class of each product

H0: Price, Delivery time & Age did not make a meaningful change to the buying pattern for all classes

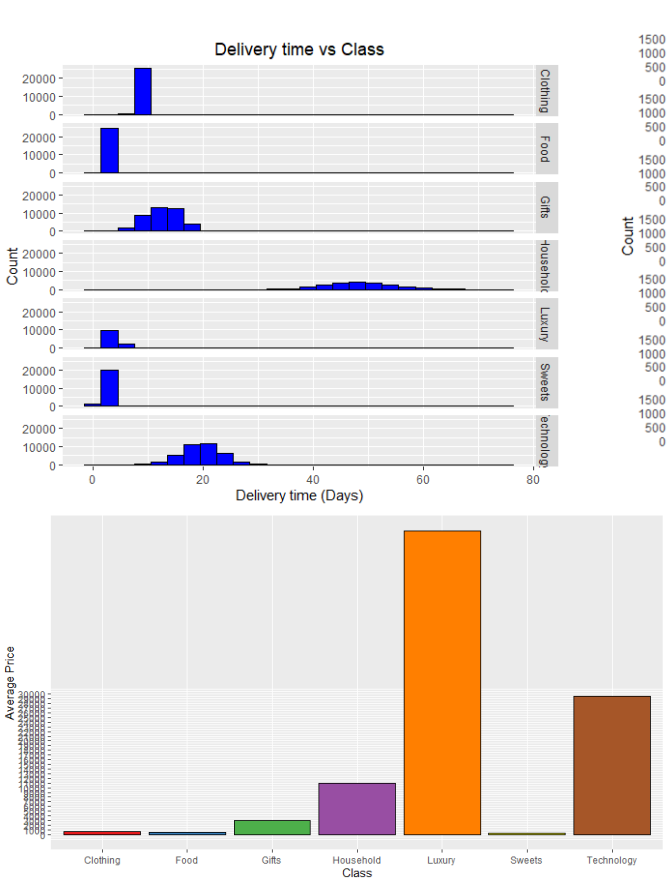
H1: At least one feature affects the buying pattern

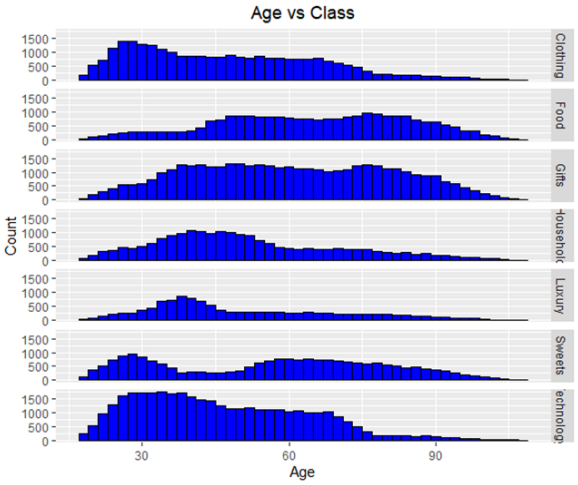
P-value < 2.2e-16

Because p = 0.05 is larger than the p-value equaling 2.2e-16, one should reject the null hypothesis. The average of at least one variable varies. Price, Delivery Time, or Age is significantly different among classes.

Dependent variable	**P-Value**	**Analyses**
Price	2.2e-16	P-value < 0.05, therefore Price differs depending on what class the product is.
Delivery Times	2.2e-16	P-value < 0.05, therefore Delivery Times are different for different classes of products.
Age	2.2e-16	P-value < 0.05, therefore Age differs depending on what class the product is.

Visualizing the analyses:





Conclusion for MANOVA Test One:

The household delivery time is much speedier compared to other classes' delivery times, but the conclusion is unfortunately not dependable due to the x-bar chart that shows the process is not in control. The reliability of the service delivery reliability will decrease. The buying patterns of luxury items show that the items are expensive, because of their high value of the items. Therefore, to increase the revenue of luxury products, one should have high reliability of service. Age is distributed among each class, but the advertisement of technology should be focused on younger people because they are the target market for technological items.

Second hypothesis

Dependent variable: Price, Delivery time & Age

Independent variable: Why Bought

H0: Price, Delivery time & Age did not make a meaningful change to the buying pattern for the reason a product is bought.

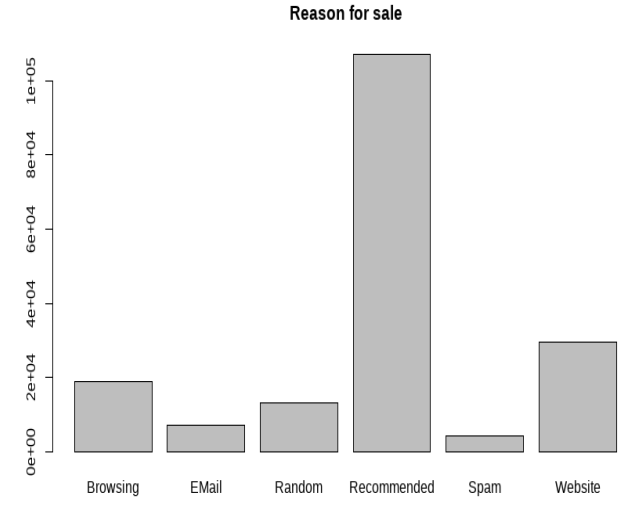
H1: At least one feature affects the buying pattern for the reason a product is bought.

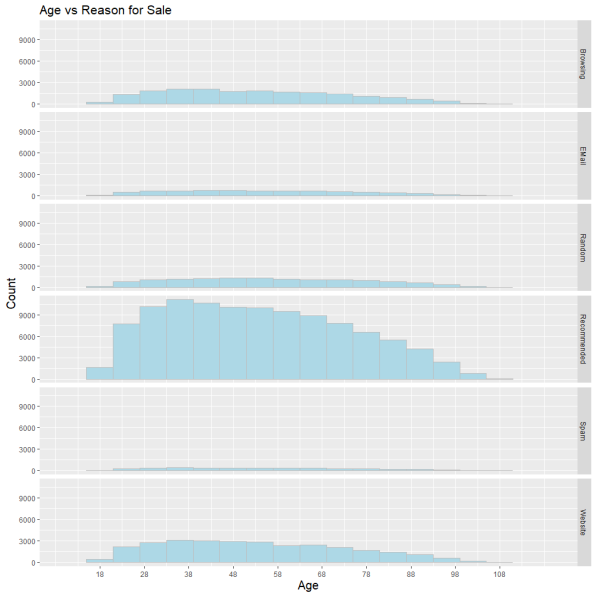
P-value < 2.2e-16

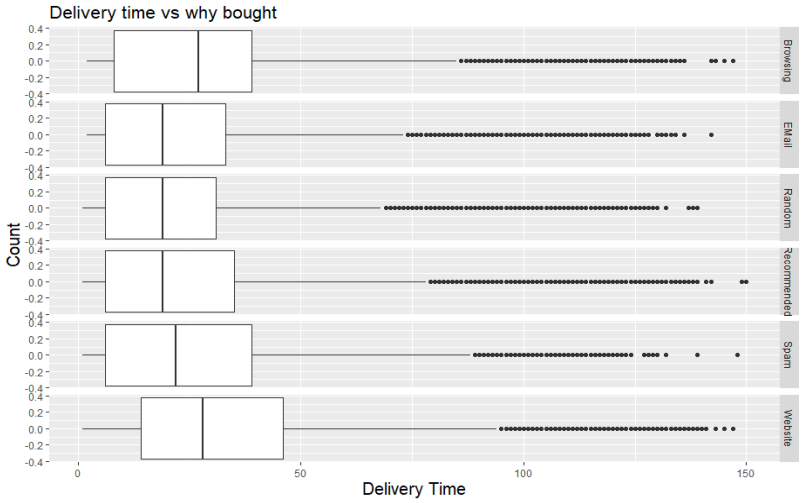
Because p = 0.05 is larger than the p-value equaling 2.2e-16, one should reject the null hypothesis. The average of at least one variable varies. Price, Delivery Time, or Age is significantly different among Why Bought.

Dependent variable	**P-Value**	**Analyses**
Price	2.2e-16	P-value < 0.05, therefore Price differs depending on the reason a product is bought.
Delivery Times	2.2e-16	P-value < 0.05, therefore Delivery Time is different for different classes of products on the reason a product is bought.
Age	2.2e-16	P-value < 0.05, therefore Age differs depending on why the product is bought by the customer.

Analyses visualization:







Conclusion for MANOVA Test Two:

The age of the reason a product is bought is distributed normally over the different age groups. The delivery time when a product is bought because of a website is longer than the other groups. All the classes have almost the same delivery time when compared to the reason they bought the product.

Third hypothesis

Dependent variables: Day, Year

Independent variable: Class

H0: Day, Year made no meaningful change to the buying pattern for a type of class of a product.

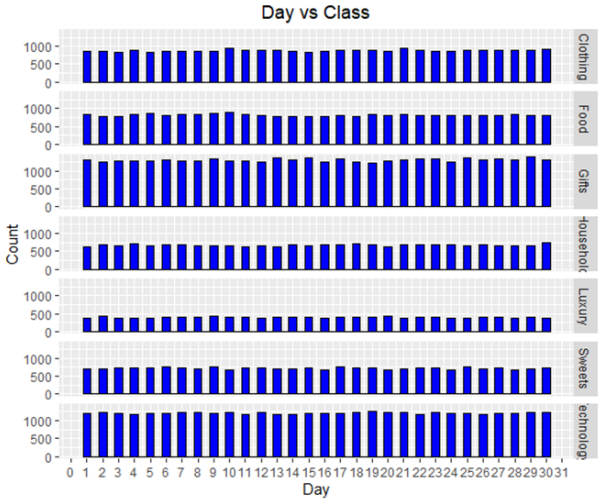
H1: At least one feature influences the buying pattern for the specific class of the product that is bought.

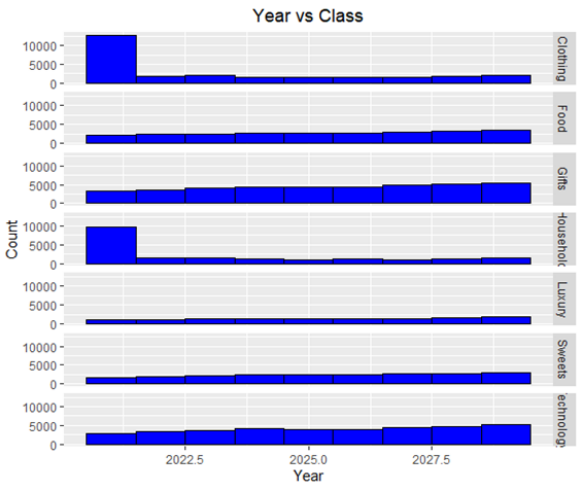
P-value < 2.2e-16

Reject the null hypothesis. At least one dependent variable's average differs. Due to this output, day and year are different among the different classes of products.

Dependent variable	**P-value**	**Analyses**
Day	2.2e-16	The p-value < 0.05. This indicates that days are different depending on which class of product you are looking at.
Year	2.2e-16	The p-value < 0.05. This indicates that years are different depending on which class of product you are looking at.

Visualization of the analyses:





Conclusion for MANOVA Test Three:

Looking at the graphs above, the demand for household products and clothing decreased after the year 2020/2021 to 2029. To find the reason behind the decreasing demand should be investigated by the sales department to ensure that the reason does not imply that quality/service is not up to standards. For the other classes, through the years the demand is distributed equally. Looking at the day vs class graph, all the classes are distributed equally with luxury having a low count. This can be due to luxury items being expensive.

PART 6: Reliability of service and products

6.1 Problem 6

6. A blueprint specification for the thickness of a refrigerator part at Cool Food, Inc. is 0.06 ± 0.04 centimeters (cm). It costs \$45 to scrap a part that is outside the specifications. Determine the Taguchi loss function for this situation.

Calculate the constant

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

$$45 = \mathrm{k}\,(0.04)^2$$

$$\mathrm{k} = 45/\,(0.04)^2 = 28125$$

Calculate the loss function

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

$$L(x) = 28125\,(x - 0.06)^2$$

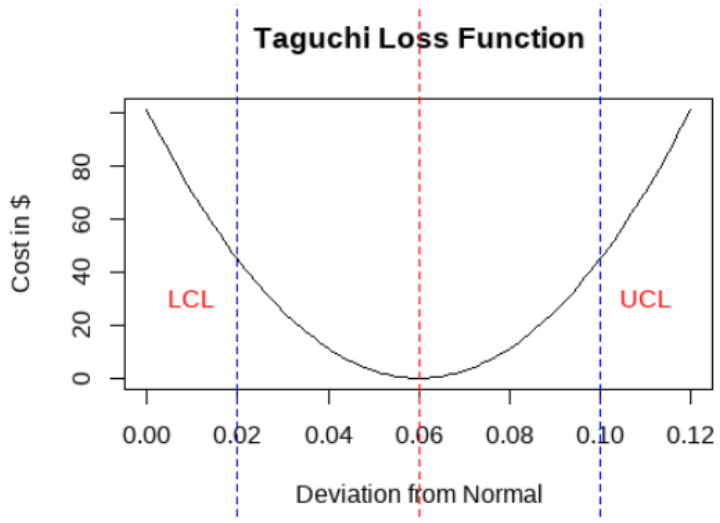


Figure 43

The product's characteristics deviate from the target value of 0.06cm. When the quality of the product is reduced, the company will make a bigger loss as seen in figure 24. If this occurs, the products and services would be unreliable, and efficiency would decrease.

6.1 Problem 7

7. A team was formed to study the refrigerator part at Cool Food, Inc. described in Problem 6. While continuing to work to find the root cause of scrap, they found a way to reduce the scrap cost to $35 per part.
   a. Determine the Taguchi loss function for this situation.
   b. If the process deviation from target can be reduced to 0.027 cm, what is the Taguchi loss?

a) Taguchi Loss function

<u>Calculate the constant of the equation</u>

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

$$35 = \mathrm{k}\ (0.04)^2$$

$$\mathrm{k} = 35/\ (0.04)^2 = 21875$$

Calculate the Loss function

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

$L(x) = 21875\ (x - 0.06)^2$

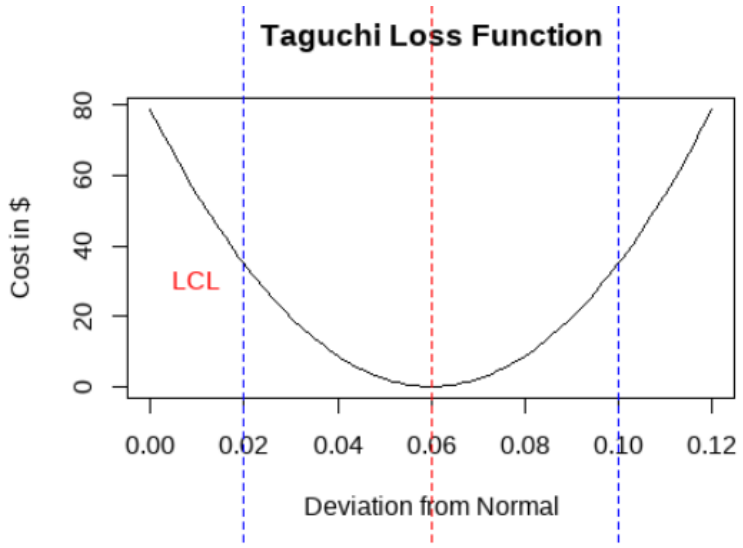


Figure 44

The product's characteristics deviate from the target value of 0.06cm. When the quality of the product is reduced, the company will make a bigger loss as seen in figure 24. If this occurs, the products and services would be unreliable, and efficiency would decrease.

b) Taguchi loss if process deviations = 0.027

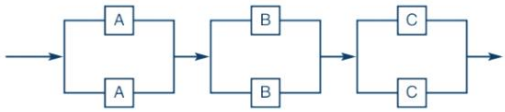
$L(x) = k(x - T)^2$

$L\ (0.027) = 21875\ (0.027)^2 = \$15.95$

This means that the company makes a loss of $15.95 per item if the deviation from the target in the process equals 0.027 cm. This will cause bad service quality.

6.2 Problem 27

27. Magnaplex, Inc. has a complex manufacturing process, with three operations that are performed in series. Because of the nature of the process, machines frequently fall out of adjustment and must be repaired. To keep the system going, two identical machines are used at each stage; thus, if one fails, the other can be used while the first is repaired (see accompanying figure).



Production System

The reliabilities of the machines are as follows:

Machine	Reliability
A	0.85
B	0.92
C	0.90

a. Analyze the system reliability, assuming only one machine at each stage (all the backup machines are out of operation).
b. How much is the reliability improved by having two machines at each stage?

a) $\boldsymbol{Reliability} = Reliability\ (Machine\ A) \times Reliability\ (Machine\ B) \times Reliability\ (Machine\ C)$

$\boldsymbol{Reliability} = 0.85 \times 0.92 \times 0.90 = 0.7038$

b) $\boldsymbol{Reliability} = Reliability\ (A1\ \&\ A2) \times Reliability\ (B1\ \&\ B2) \times Reliability\ (C1\ \&\ C2)$

$\boldsymbol{Reliability} = \left(1 - (1 - 0.85)^2\right) \times \left(1 - (1 - 0.92)^2\right) \times \left(1 - (1 - 0.90)^2\right) = 0.9615$

Thus equaling 96.15%

By placing the two machines parallel to one another, productivity will improve by 25.77%. the reason for this increase in productivity is that if one machine breaks/fails, the other one can operate independently. This will improve the reliability of the company as well.

6.3 Problem: Expecting reliable delivery

Calculations that were used for the vehicle and driver reliability were done with the help of dbinom() functions and the constants given in the questions.

Case 1: 20 vehicles available

RESULTS:

R(v) = ProbabilityOfReliableNrOfVehicles = 0.9904

R(D) = ProbabilityOfReliableNrOfDrivers = 0.9979

TotalReliability = R(v) * R(D) * 365 = 0.98834

TotalReliability = 360.7449

Case 2: 21 vehicles available

RESULTS:

R(v) = ProbabilityOfReliableNrOfVehicles = 0.99998

R(D) = ProbabilityOfReliableNrOfDrivers = 0.9979

TotalReliability = R(v) * R(D) * 365 = 0.99788

TotalReliability = 364.229

CONCLUDING:

By adding one extra delivery vehicle, the number of days available for deliveries increased by 3.48 days.

Conclusion

After cleaning and ordering the data that was provided for all the different classes, valid data was obtained and was used further in the project. I gained a thorough understanding of all the data and thereafter constructed the necessary tables and graphs that visualize the data. After that, control charts were constructed, and enough information is now present to have a clear understanding of all the different classes of products.

After setting up the control charts and seeing which products lie outside of the control limits, I was able to assume that gifts, luxury, and household products are out of control. The reason behind the unstable control of these three products should be thoroughly investigated by the sales department to pinpoint the problem and resolve the problem. Regarding the logistic partner, changes need to be made to reduce the delivery time.

The MANOVA test results show this as well, where the demand for household products decreased. This can be due to the service quality that is not up to date and that the customers being dissatisfied with the company.

In part 5, one can see that the probability of making a type 1 error is much smaller than making a type 2 error. The company should thus focus on ensuring that the company delivers products on time and not simply assuming that it is done on time.

By using explorative analyses and visualizing all the different classes to compare to one another, I understand the importance that it holds for a company to detect where problems are and to focus on problem-solving for that specific area of the company.

References

Hernandez, F. (2015, 01 10). *Data-Analysis with R exercises*. Retrieved from http://fch808.github.io/Data-Analysis-with-R-Exercises.html

Honorlock. (2022). *Six Sigma in Higher Education*. Retrieved from https://honorlock.com/blog/what-is-six-sigma-and-how-does-it-apply-to-higher-education/#:~:text=Six%20Sigma%20is%20a%20process,processes%2C%20and%20improving%20student%20satisfaction.

STHDA. (2022). *Statisical Tools For High-Throughput Data Analysis*. Retrieved from http://www.sthda.com/english/wiki/manova-test-in-r-multivariate-analysis-of-variance