QUALITY ASSURANCE ECSA PROJECT

Jordan Waugh

22743758

Table of contents

Table of figures

Table of tables

Abstract

This report is set out to analyze and investigate an online sales business services and quality functions. This analysis of the data with statistical tables, equations, and figures to determine the state of control and quality. Recommendations are set out to help improve the online business.

Introduction

The data supplied in the SalesTable2022 refers to the sales that took place which includes variables such as ID, class, age, price, tear, month, day, delivery time and why bought. An investigation will take place to analyze the data provided for customers. The process includes wrangling of the data, statistical analyzes and descriptive statistics and observing the relationship of the data.

Data Wrangling

The data provided in the SalesTable2022 was analyzed and separated into invalid and valid data. Valid data was all positive and non-NA values whereas the invalid data was all the data with NA values, all negative values. This was placed into another file to separate from the valid data.

- **Invalid data**

The invalid data was separated from the valid data (the remaining data leftover from the valid data). Furthermore, the 17 instances of the 180000 are shown below:

	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	86622	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

Table 1

- **Valid data**

The valid data was separated from the invalid data. Furthermore, a total of 179983 instances were obtained from the 180000. The first section of the valid data is shown below:

	X	ID	AGE	Class	Price	Year	Month	Day	Delivery.time	Why.Bought
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

Table 2

Part 2: Descriptive statistics

- 5-point summary of age

AGE	
MINIMUM	18.00
1^{ST} QUARTILE	38.00
MEDIAN	53.00
MEAN	54.57
3^{RD} QUARTILE	70.00
MAXIMUM	108.00

Table 3

The minimum age is 18 and the maximum age is 108. The most common age in the dataset is 54.57 i.e., 55. The maximum age of 108 could be a mistake or even an outlier as the is fairly unrealistic with the average ages.

- 5-point summary of price

PRICE	
MINIMUM	35.65
1ST QUARTILE	482.31
MEDIAN	2259.63
MEAN	12294.10
3RD QUARTILE	15270.97
MAXIMUM	116618.97

Table 4

The minimum price from the valid data set is 35.65 and the maximum is 116618.97. the most common is 12294.10. This is realistic as there are no negative values.

- 5-point summary of delivery time

DELIVERY TIME	
MINIMUM	0.5
1ST QUARTILE	3.0
MEDIAN	10.0
MEAN	14.5
3RD QUARTILE	18.5
MAXIMUM	75.0

Table 5

The minimum delivery time is 0.5 days (12 hours), and the maximum is 75 days (1800 hours). The average is 14.5 days (348 hours).

Graphs

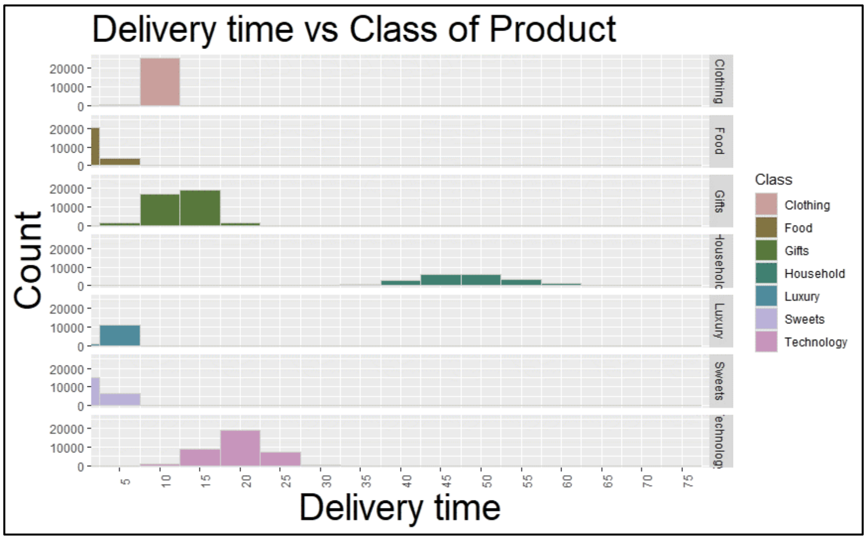


Figure 1

figure 1: The classes sweets, food and luxury have the shortest delivery times with gifts just thereafter. The class with the longest delivery time is the household class. The reason for these delivery periods might be due that household goods are large and bulky making it exceptionally difficult and time consuming to deliver whereas food, luxury and sweets are smaller everyday items making them less time consuming and easier to deliver.

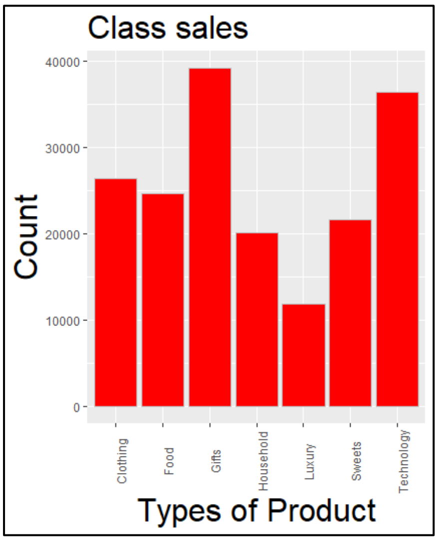
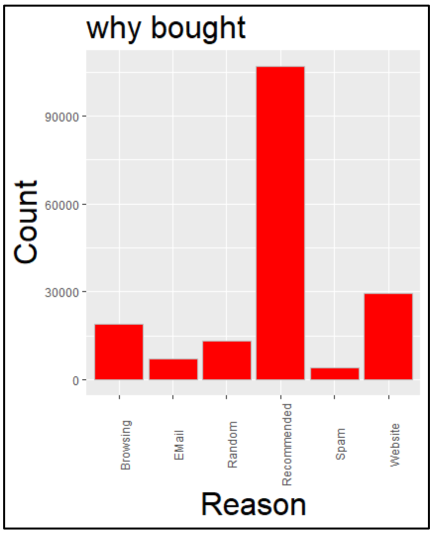


Figure 2

Figure 2: gifts was the most frequent purchased class followed by technology. The least frequent was luxury, this could be due to gifts and technology are affordable and good quality products whereas luxury was not meeting criteria of the target audience (too expensive).



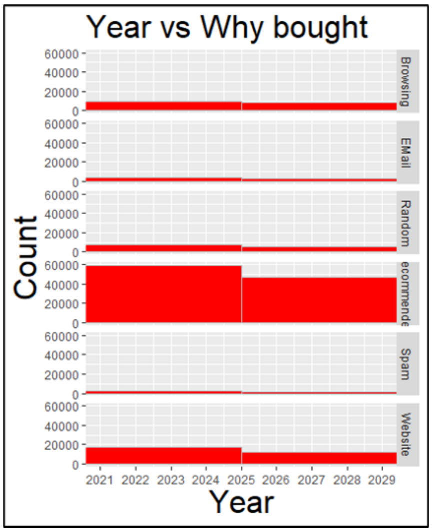


Figure 3, 4

Figures 3: the reason for a purchase is due to a recommendation by a fellow person (family member or friend). The reason for least number of purchases was spam.

Figure 4: the reason for a purchase in a specific year, from year 2021 to 2025 for the recommended reason was the highest and the lowest from 2025 to 2029 for spam.

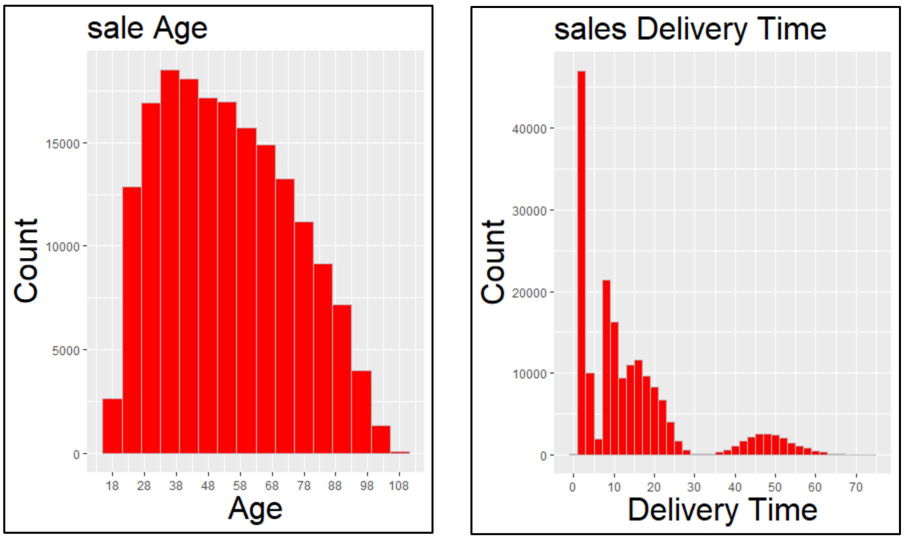
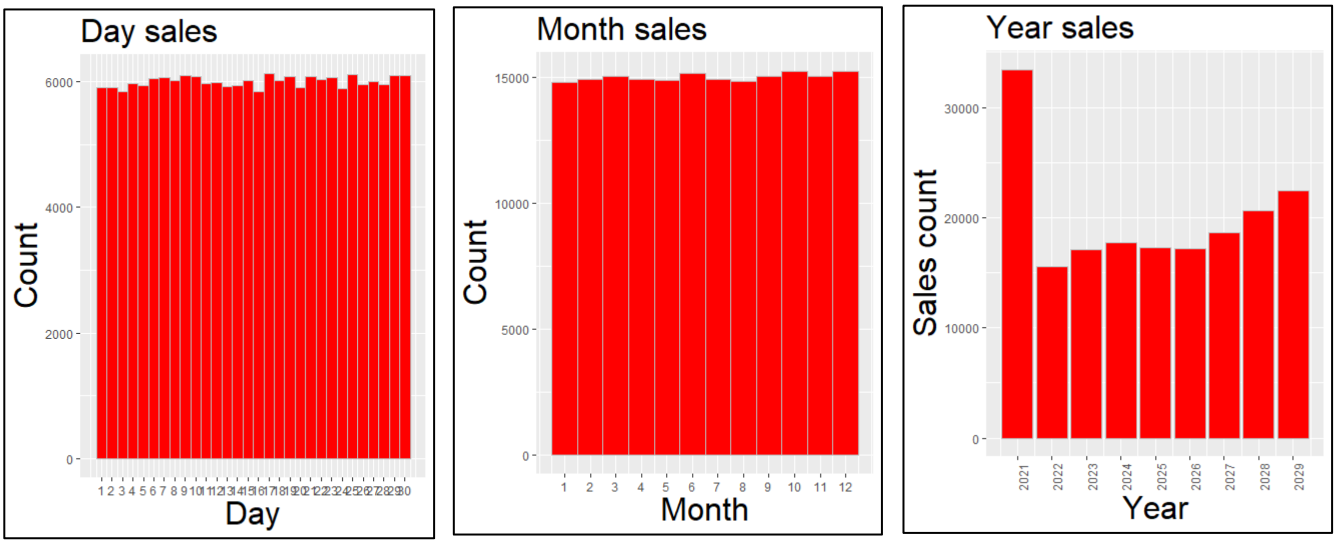


Figure 5,6

Figure 5: the sales increase then decrease showing a pattern of skewness to the right. The sales are very popular between ages 33 to 43 and low number of sales for ages 10 to 20 and 60 plus years.

Figure 6: the most sales with shortest delivery time between 1 to 4 days. The least sales were between 28 and 35 delivery time.



Figures 7, 8, 9

Figure 7: The daily sales are consistent with minimal variation.

Figure 8: Monthly sales are fairly consistent with minimal variation throughout.

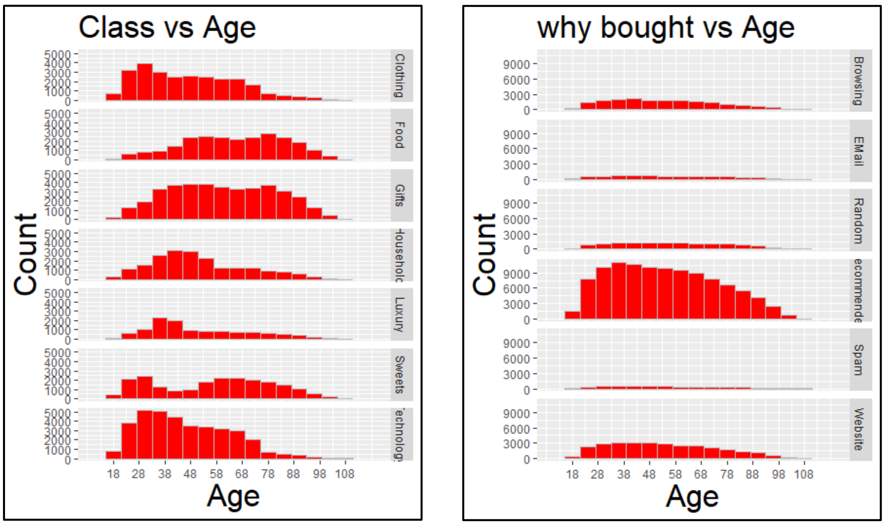
Figure 9: 2022 has the least number of sales in a particular year whereas 2021 had the most

Figures 10,11,12

Figure 10: most frequent sales regarding class and day was technology and gifts. The least was luxury. This could be due that gifts and technology items are cheaper and have a higher demand.

Figure 11: the most frequent sales was between months 3 and throughout the class of products. Technology class had the most from month 0 to 3 and 9 to 12 whereas luxury was the least.

Figure 12: the highest was clothing from year 2021 to 2025 followed by gifts and technology from year 2025 to 2029. The least throughout was luxury.



Figures 13, 14

Figure 12: as seen above there was a lot of variation between age and class of the product. For clothing and technology, the data is skewed to the right, and perhaps luxury. The rest are fairly centred and have some variation.

Figure 13: the recommended and website class skewed to the right. The rest of the classes are fairly flat with some variation.

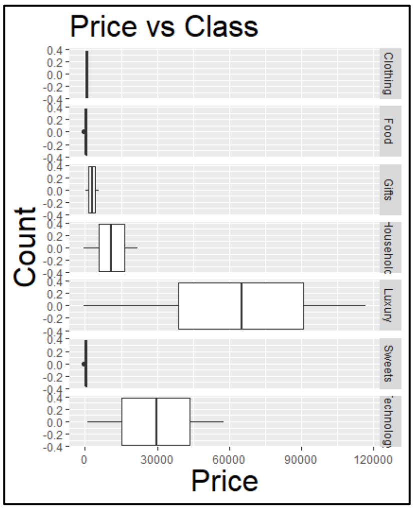


Figure 15

Figure 14: sweets, clothing and food are the cheapest items to purchase. Whereas luxury is the most expensive followed by technology and then household. This is expected as luxury, technology and household items are more expensive than sweets, food, and clothing.

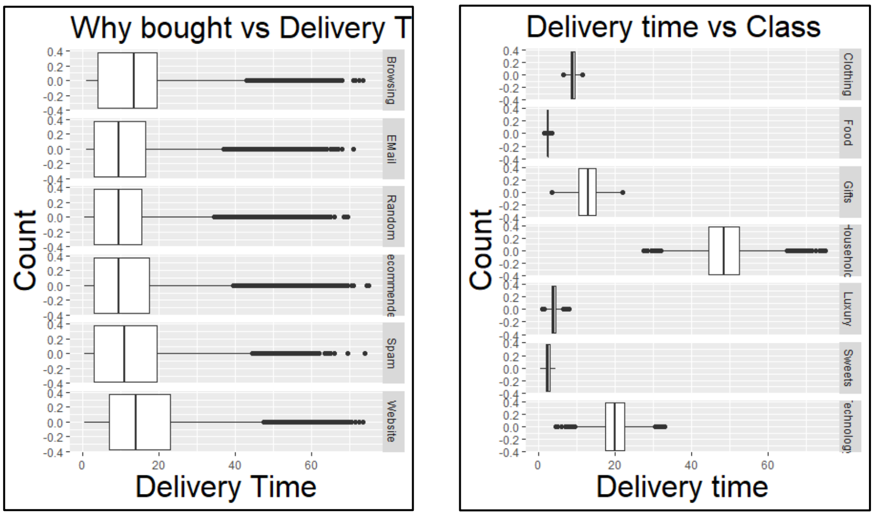


Figure 16, 17

Figure 15: all boxplots show a similar distribution but the website reason as well as browsing have a bit longer delivery time than the rest.

figure 16: it shows that food, sweets, luxury, and clothing have good delivery times and short. Technology and gifts have an average delivery time and household being the longest.

Process capabilities

The process capabilities were calculated for delivery times for technology class items

Used USL=24 and LSL=0.

Cp	0.286634399833
CPU	0.573268799667
Cpk	0.346342334911
Cpl	0.346342334911

Table 6

CP: The CP is less than 1. This means the data is not capable.

CPK: CPK is less than 1 this means the data is not centred and not capable.

LSL: LSL can’t be less than zero.

PART 3: SPC

To calculate the control limits, the first 30 samples in each class were used.

X-Chart							
Class	UCL	U2Sigma	U1Sigma	CL	L1Sigma	L2Sigma	LCL
Technology	22.9731035296432	22.1068838345769	21.2406641395107	20.37444	19.5082247493782	18.642005054312	17.7757853592457
Clothing	9.40468055141378	9.25978703427586	9.11489351713793	8.97	8.82510648286207	8.68021296572415	8.53531944858622
Household	50.246183898822	49.0181966732887	47.7902094477555	46.56222	45.334234996689	44.1062477711557	42.8782605456224
Luxury	5.49352400900969	5.24086785785831	4.98821170670693	4.735556	4.48289940440418	4.2302432532528	3.97758710210142
Food	2.70933008763795	2.6362200584253	2.56311002921265	2.49	2.41688997078735	2.3437799415747	2.27066991236205
Gifts	9.48790890716703	9.11230964181506	8.73671037646309	8.361111	7.98551184575914	7.60991258040716	7.23431331505519
Sweets	2.89679765130182	2.75712436012714	2.61745106895246	2.477778	2.3381044866031	2.19843119542842	2.05875790425374

Table 7

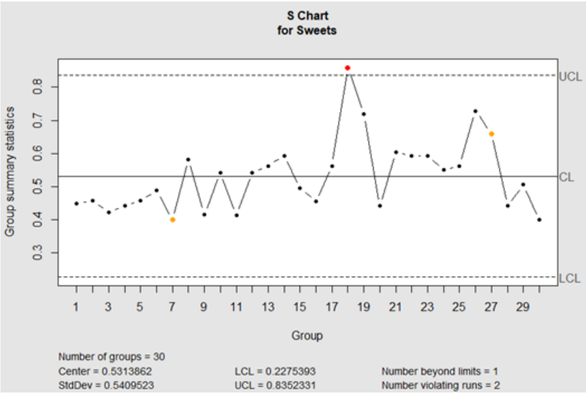
S-Chart							
Class	UCL	U2Sigma	U1Sigma	CL	L1Sigma	L2Sigma	LCL
Technology	5.179912106856	4.55178400502195	3.9236559031879	3.295528	2.66739969951979	2.03927159768574	1.41114349585169
Clothing	0.866449571514653	0.761381895951185	0.656314220387717	0.5512465	0.44617886926078	0.341111193697312	0.236043518133844
Household	7.34324783058395	6.45278864402426	5.56232945746456	4.67187	3.78141108434516	2.89095189778547	2.00049271122577
Luxury	1.51085996274949	1.3276495928345	1.1444392278174	0.9612289	0.778018492885311	0.594808125419265	0.411597757953219
Food	0.437191081671539	0.384176281689576	0.331161481707612	0.2781467	0.225131881743686	0.172117081761723	0.11910228177976
Gifts	2.24604819424492	1.97368720437184	1.70132621449875	1.428965	1.15660423475259	0.884243244879508	0.611882255006426
Sweets	0.835233112432091	0.733950816771798	0.632668521111506	0.5313862	0.430103929790922	0.328821634130629	0.227539338470337

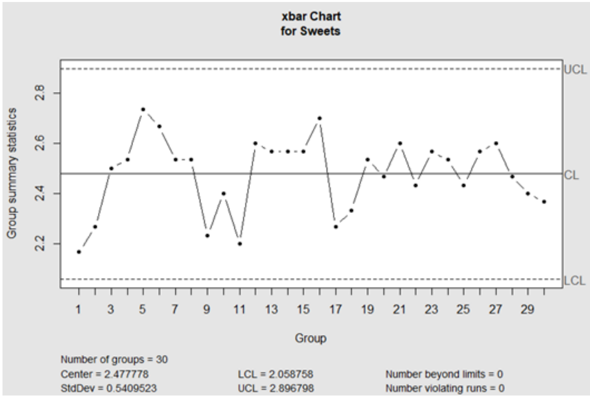
Table 8

XBAR AND S GRAPHS FOR FIRST 30 SAMPLES:

The control graphs that follow show whether the process is in control or not and if any sample points fall outside the limits(UCL and LCL).

Sweets:



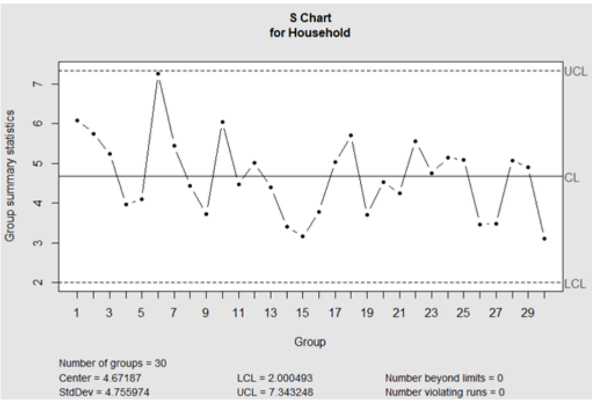


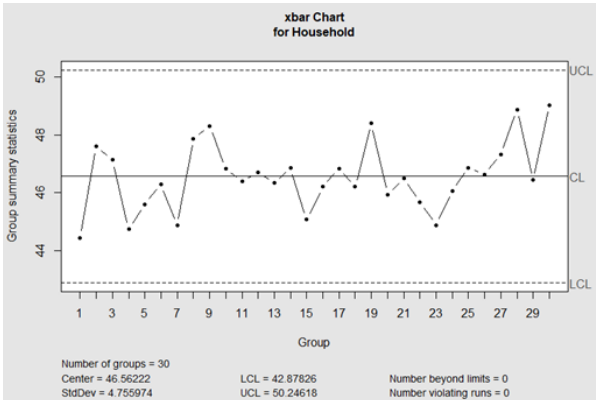
Figures 18,19

The S chart shows there is 1 sample point outside the LCL and UCL. This means the sample point must be removed.

The xbar chart show there aren't any sample points outside the LCL and UCL. This means the process is in control.

Household:

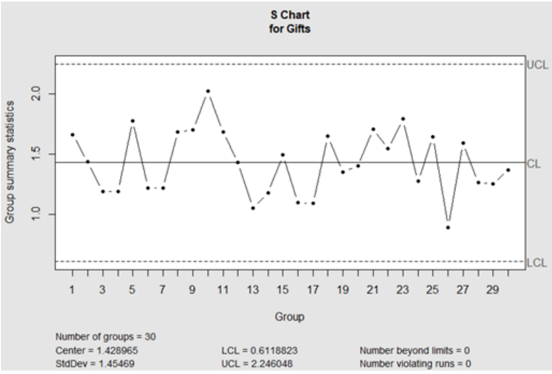


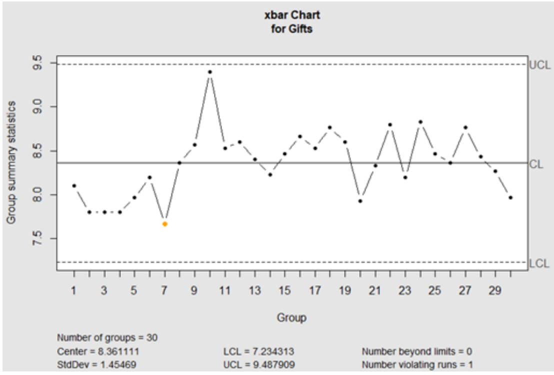


Figures 20,21

The S chart and xbar chart show there aren't any sample points outside the LCL and UCL. This means the process is in control.

Gifts:

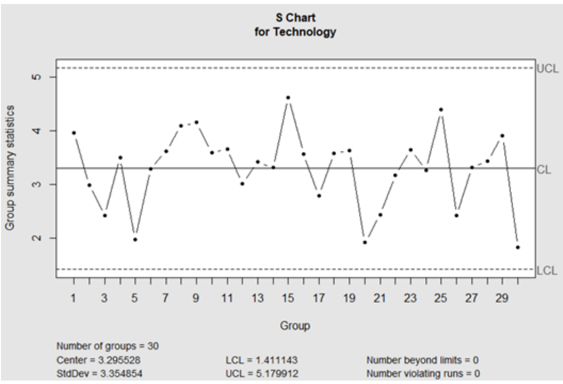


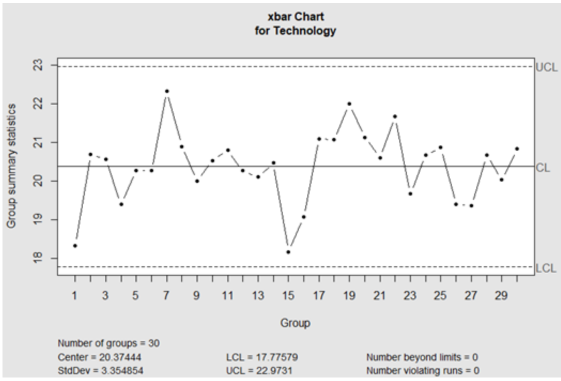


Figures 22,23

The S chart and xbar chart show there aren't any sample points outside the LCL and UCL. This means the process is in control.

Technology:

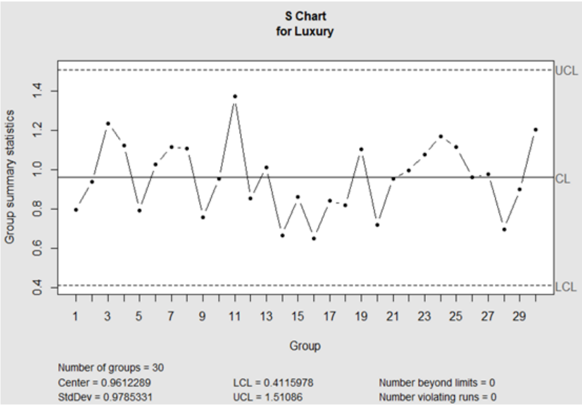


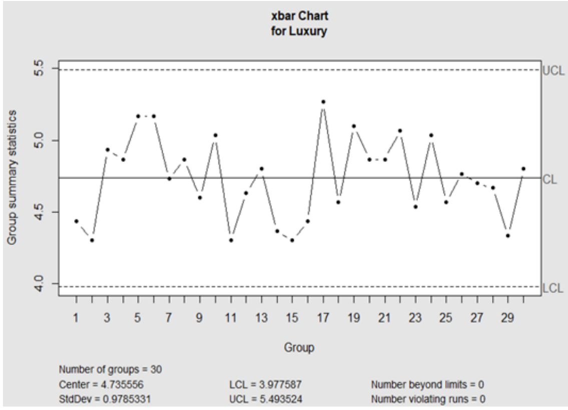


Figures 24,25

The S chart and xbar chart show there aren't any sample points outside the LCL and UCL. This means the process is in control.

Luxury:

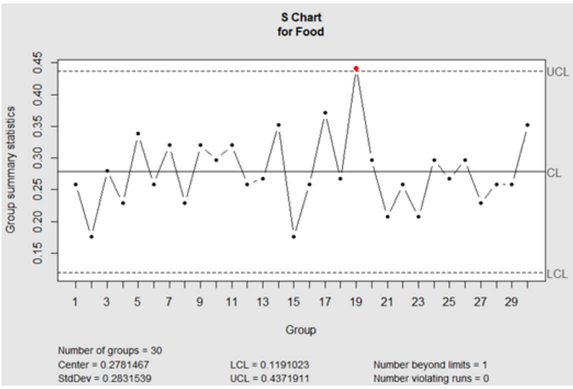


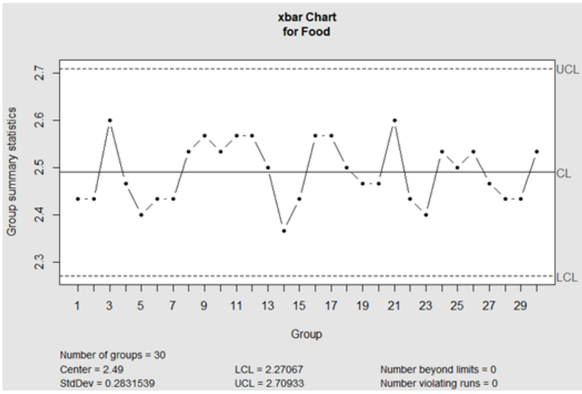


Figures 26,27

The S chart and xbar chart show there aren't any sample points outside the LCL and UCL. This means the process is in control.

Food:



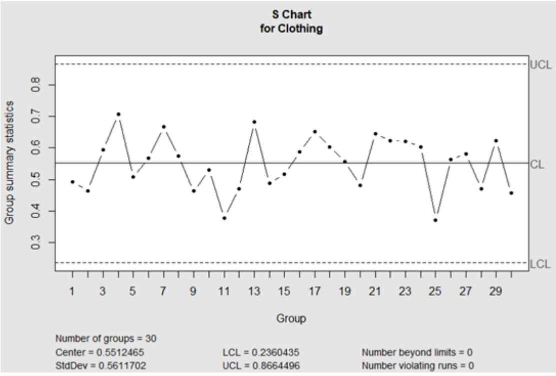


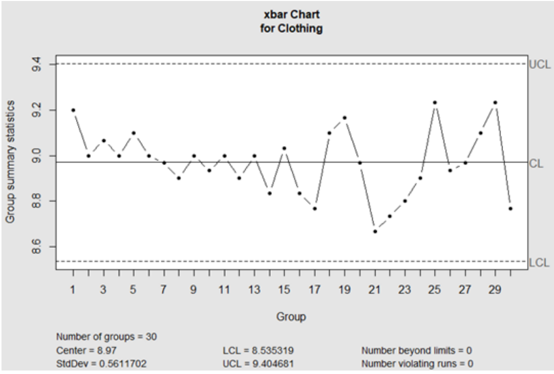
Figures 28,29

The S chart shows there is 1 sample point outside the LCL and UCL. This means the sample point must be removed.

The xbar chart show there aren't any sample points outside the LCL and UCL. This means the process is in control.

Clothing:



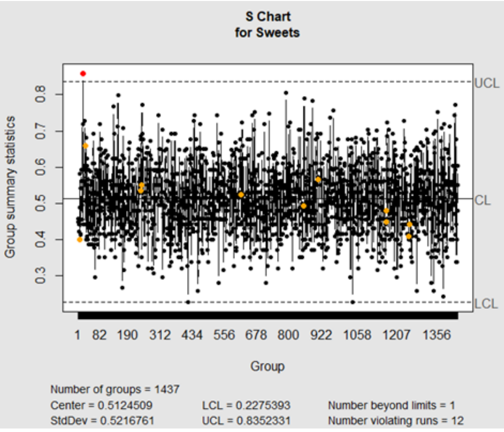


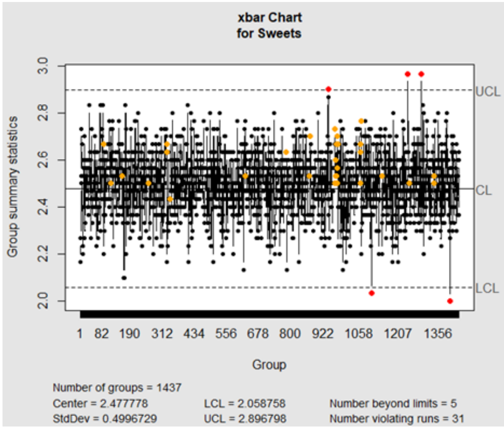
Figures 30,31

The S chart and xbar chart show there aren't any sample points outside the LCL and UCL. This means the process is in control.

XBAR AND S GRAPHS ON ALL THE SAMPLES:

Sweets:

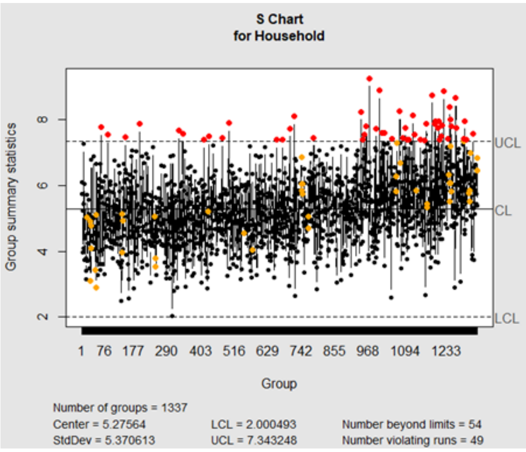


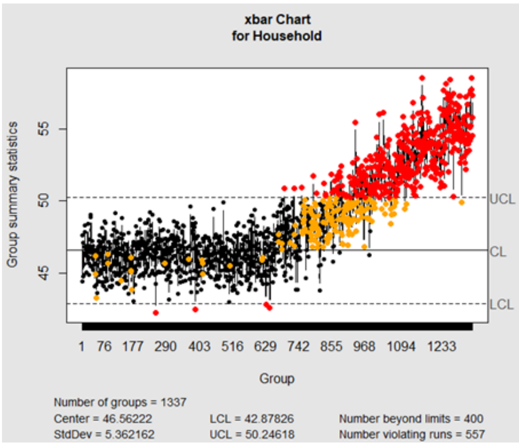


Figures 32,33

In the sweets class it can be seen the process is in control. The reason for it being in control is there are only a few sample points exceeding the UCL and LCL. The sample points are distributed evenly over the CL.

Household:

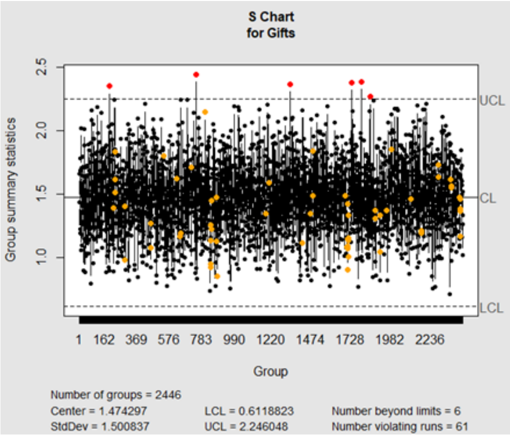


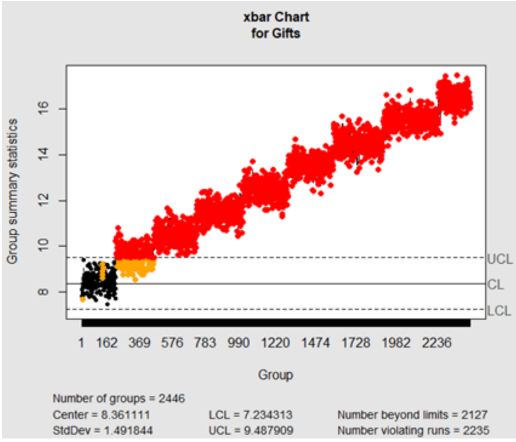


Figures 34,35

The Household class is not in control and unstable as seen in figure 35 above. The data keeps increasing and thus causing most of the sample points to be exceeding the UCL. A reason for the data to keep increasing is due that the household class is using an alternative method of delivery.

Gifts:

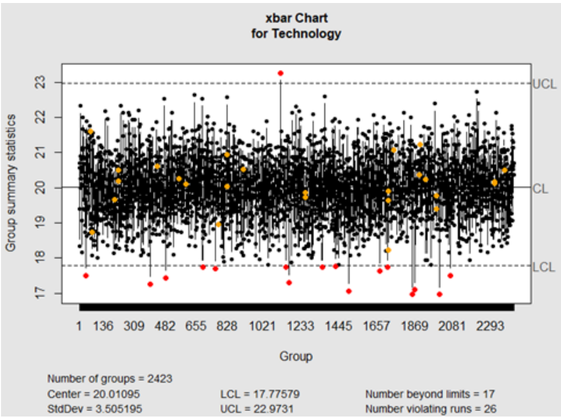


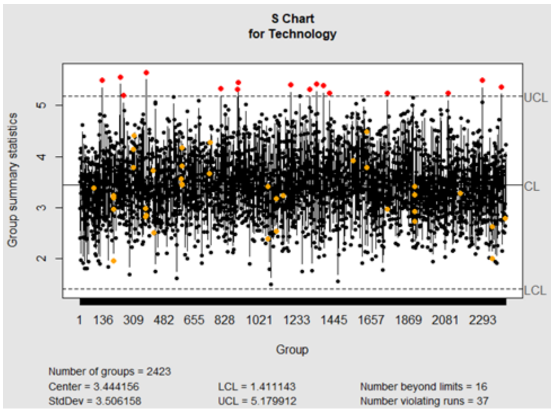


Figures 36,37

The gifts class is not in control and unstable as seen in figure 37 above. The data keeps increasing and thus causing most of the sample points to be exceeding the UCL. A reason for the data to keep increasing is due that the gifts class is used an alternative method of delivery.

Technology:

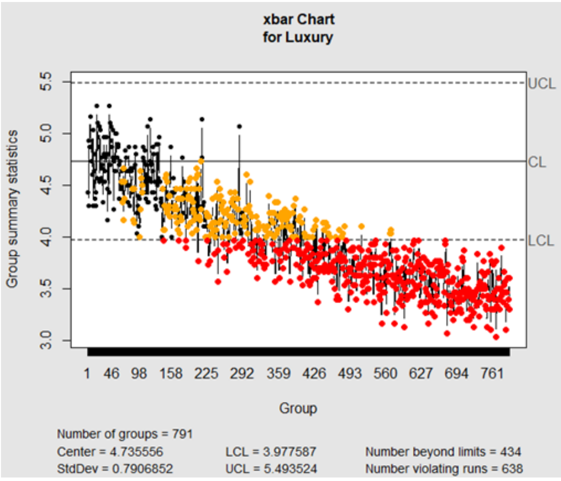


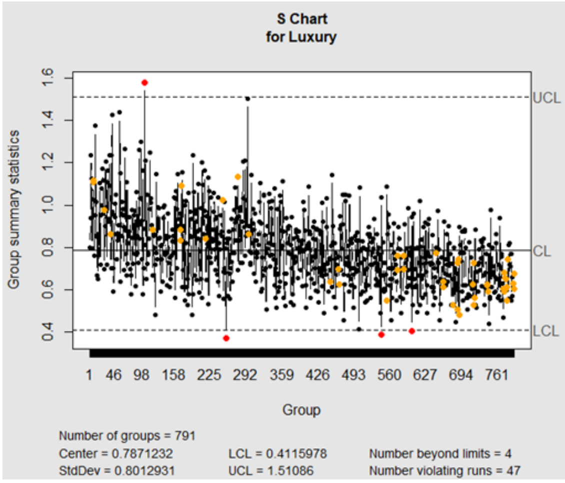


Figures 38,39

In figures 38 and 39 for the technology class it can be seen the process is in control. The reason for it being in control is there are only a few sample points exceeding the UCL and LCL. The sample points are distributed evenly over the CL.

Luxury:

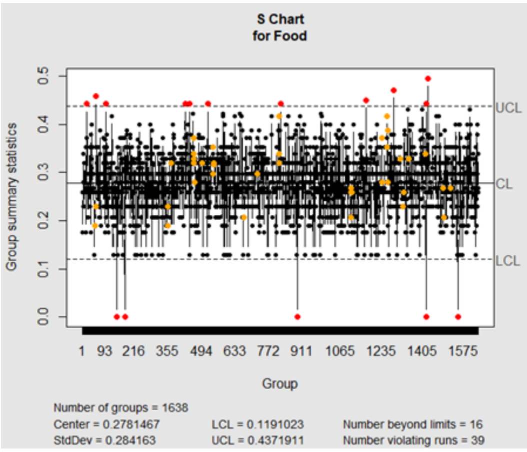


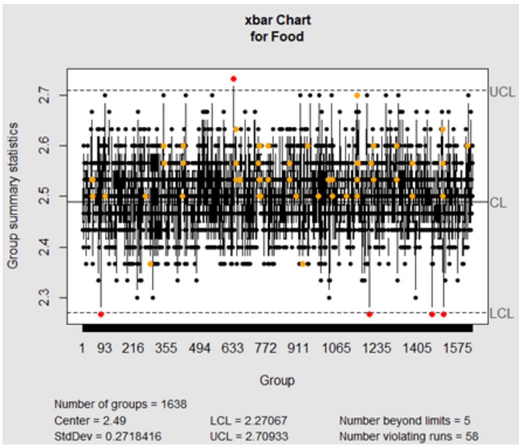


Figures 40,41

The luxury class is not in control and unstable as seen in figure above. The data keeps decreasing and thus causing most of the sample points to be exceeding the LCL. A reason for the data to keep decreasing is due that the luxury class having more expensive products leading to a faster delivery.

Food:

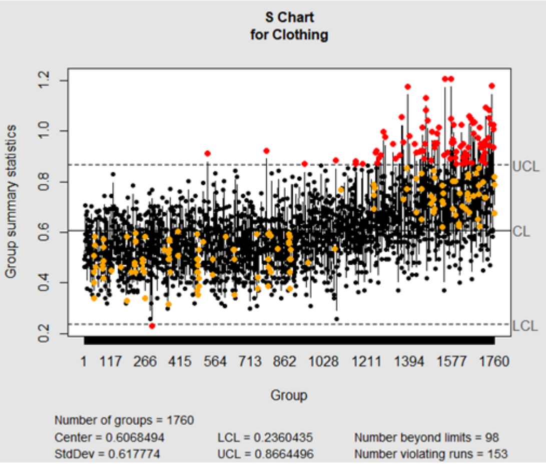


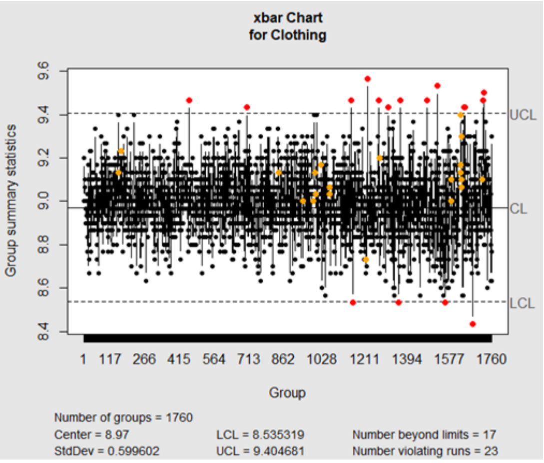


Figures 42,43

In figures 42 and 43 for the food class it can be seen the process is in control. The reason for it being in control is there are only a few sample points exceeding the UCL and LCL. The sample points are distributed evenly over the CL.

Clothing:





Figures 44,45

In figures 44 and 45 for the food class it can be seen the process is moderately in control. The reason for it being moderately in control is there are several sample points exceeding the UCL and one sample point exceeding the LCL. The sample points are distributed evenly but tend to increase more.

PART 4: Optimizing delivery time

Graph of first and last 3 outliers regarding the control limits:

6 plots occur in this section from Household, gifts, and luxury as these have the most sample points exceeding the LCL and UCL limits. This means that household, gifts, and luxury classes are not in control and the delivery will not be effective and achieved as desired. Although the rest of the classes' technology, sweets, clothing, and food will have effective and desired delivery.

First 3 outliers' household: (In region 245:641)

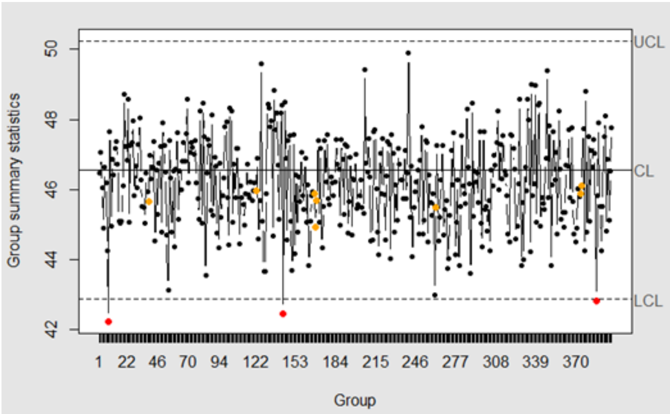


Figure 46

Last 3 outliers for household: (In region 1335:1337)

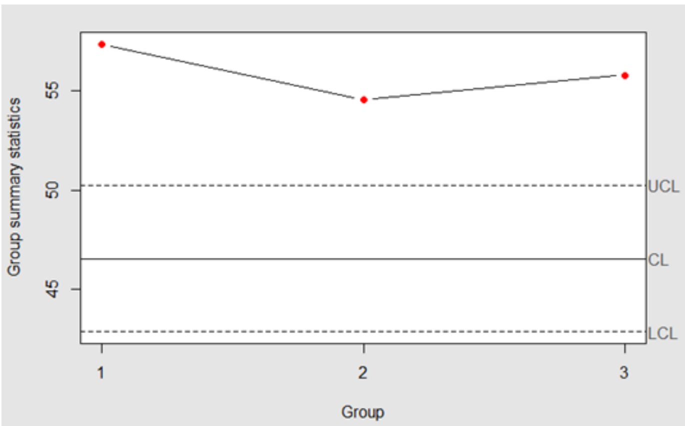


Figure 47

first 3 outliers for Gifts: (In region 210:218)

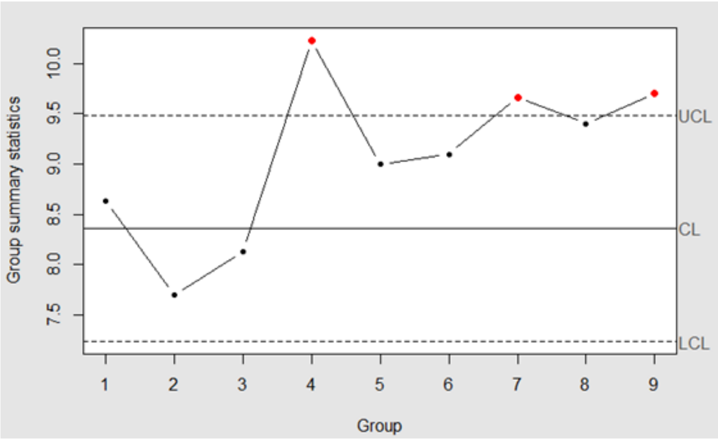


Figure 48

Last 3 outliers for Gifts: (In region 2444:2446)

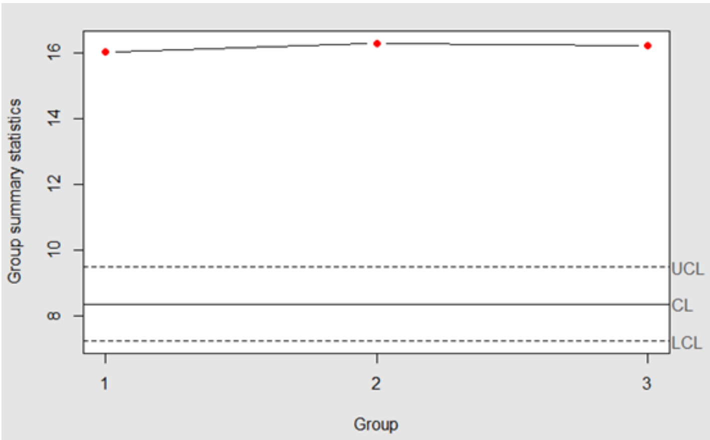


Figure 49

First 3 outliers for Luxury: (in region 137:184)

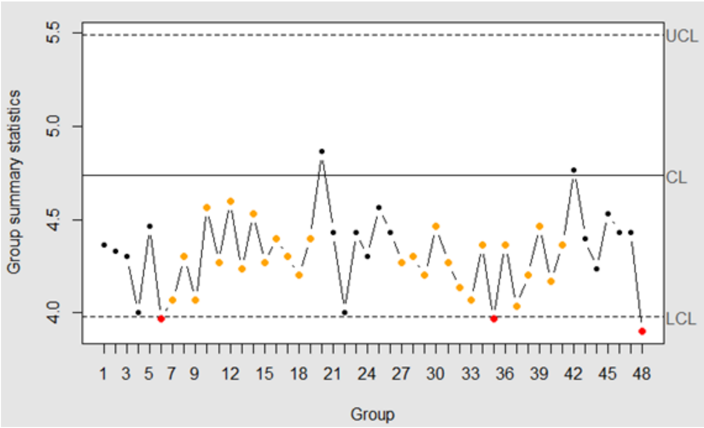


Figure 50

Last 3 outliers for Luxury: (In region 789:791)

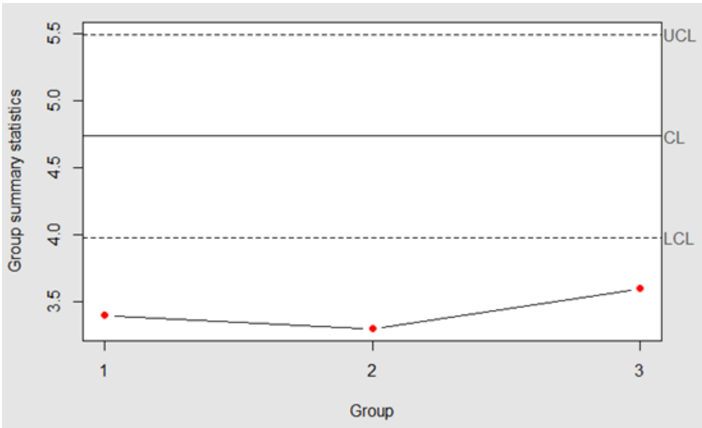


Figure 51

Table showing the most consecutive samples in each class:

Class	Most consecutive
Technology	6
Clothing	4
Household	3
Luxury	-
Food	7
Gifts	5
Sweets	4

Table 9

The most/maximum consecutive samples is7, this value is low.

Likelihood of making Type I Error:

Ho: The process is in control and centred on the centreline calculated using the first 30 samples.

H1: The process in not in control and has moves from the centreline or has increased or decreased in variation.

Type I error occurs when a process is said to be out of control when the process is in control. Probability of making a type 1 error: (calculated using the pnorm () function)

- A: 0.002699796
  There is a 0.27% chance of making an error.
- B: 0.7266668
  There is a 72.66% chance of making an error.

4.3

Currently the lost sales are 1356 which equate to R446124.

4.4 Type II error

Figure 52

A type II error will occur when the online business assumes the delivery services for technology is on time when it is late. The probability of making a type II error is 48.76%.

PART 5: DOE and MANOVA

Multiple manova tests can be done on the given data set as a manova test is a statistical test done to observe results from multiple dependent variables and 1 independent variable.

A p value of 0.05 will be chosen for the manova tests as provided in project description.

In this section there will be 3 tests done on the given data.

1st test:

Hypothesis test : Day, month and year have an effect on sales for a class of product.

- Dependent variables: day, month, year.
- Independent variables: class of product

Dependent variables P value:

- Day 0.1808
- Month 0.2866
- Year 2.2x10^-16

Independent variable p value:

- P value is 2.2x10^-16 which is smaller than 0.05.

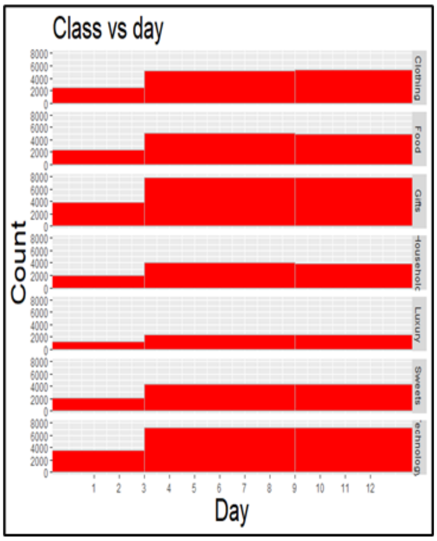
```
              Df  Pillai approx F num Df den Df    Pr(>F)
Class          6 0.10987   1140.2     18 539913 < 2.2e-16
Residuals 179971

Class     ***
Residuals
---
Signif. codes:
0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

 Response Day :
                Df   Sum Sq Mean Sq F value Pr(>F)
Class            6      664 110.617  1.4789 0.1808
Residuals   179971 13461317  74.797

 Response Month :
                Df  Sum Sq Mean Sq F value Pr(>F)
Class            6      88  14.684   1.231 0.2866
Residuals   179971 2146861  11.929

 Response Year :
                Df  Sum Sq Mean Sq F value    Pr(>F)
Class            6  153063 25510.5  3698.9 < 2.2e-16 ***
Residuals   179971 1241223     6.9
---
Signif. codes:
0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

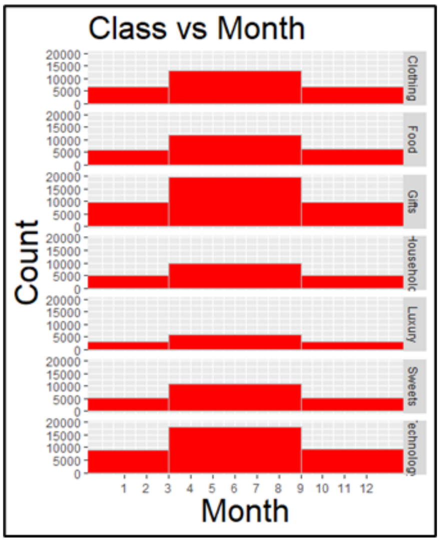
Figure 53

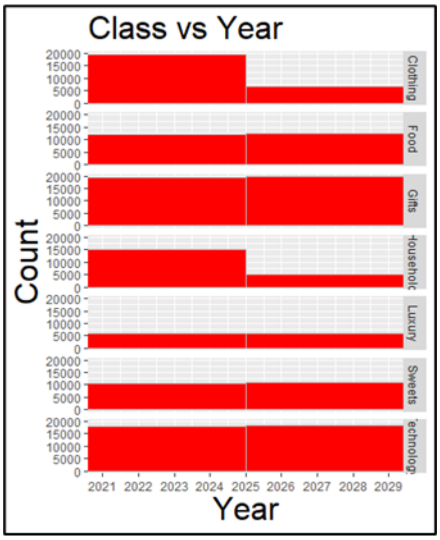
Discussion manova 1

The 2 dependent variables (day and month) have p values bigger than 0.05. this means these dependent variables will not change or effect the class of product.

The dependent variable (year) has a p value smaller than 0.05. this means that the dependent variables change regarding the independent variable, in this case “class of product”







Visual analysis to prove the discussion above. Class vs day and class vs month have similar form with a bit of variation but class vs year differs. Action should be taken to look why the sales dropped from year 2025 to 2029, mostly for clothing and household.

2nd test:

Hypothesis test: delivery time, price and age have an effect on sales for the reason why bought.

- Dependent variables: delivery time, price, and age.
- Independent variables: why bought

Dependent variables P value:

- Age 2.2x10^-16
- Price 2.2x10^-16
- Delivery time 2.2x10^-16

Independent variable p value:

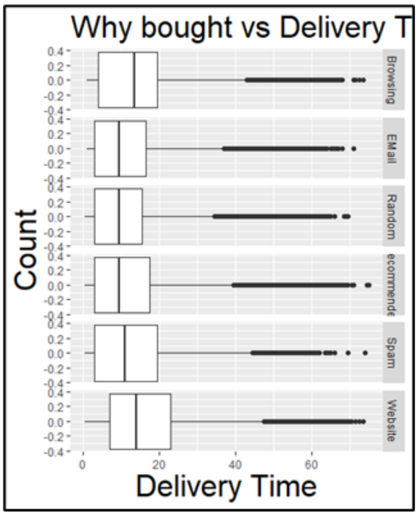
- P value is 2.2x10^-16 which is smaller than 0.05.

```
              Df   Pillai approx F num Df den Df
Why.Bought      5 0.044138   537.43     15 539865
Residuals  179955
              Pr(>F)
Why.Bought < 2.2e-16 ***
Residuals
---
Signif. codes:
0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

 Response Delivery.Time :
                Df   Sum Sq Mean Sq F value    Pr(>F)
Why.Bought       5   783098  156620  822.54 < 2.2e-16 ***
Residuals   179955 34265114     190
---
Signif. codes:
0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

 Response Price :
                Df     Sum Sq    Mean Sq F value    Pr(>F)
Why.Bought       5 1.5738e+12 3.1477e+11   736.1 < 2.2e-16
Residuals   179955 7.6952e+13 4.2762e+08

Why.Bought  ***
Residuals
---
Signif. codes:
0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

 Response AGE :
                Df   Sum Sq Mean Sq F value    Pr(>F)
Why.Bought       5   106825 21365.0  51.464 < 2.2e-16 ***
Residuals   179955 74707634   415.1
---
Signif. codes:
0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1
```

Figure 54

Discussion manova 2

Age, price, and delivery time all have p values less than 0.05. this means that all 3 dependent variables change regarding the independent variable, in this case “why bought”.



Visual analysis for why bought and delivery time. It can be seen that website has the longest delivery time followed by browsing.

3rd test:

Hypothesis test: delivery time, price and age have an effect on sales for class of product.

- Dependent variables: Delivery time, price, and age.
- Independent variable: class of products

Dependent variables P value:

- Age 2.2x10^-16
- Price 2.2x10^-16
- Delivery time 2.2x10^-16

Independent variable p value:

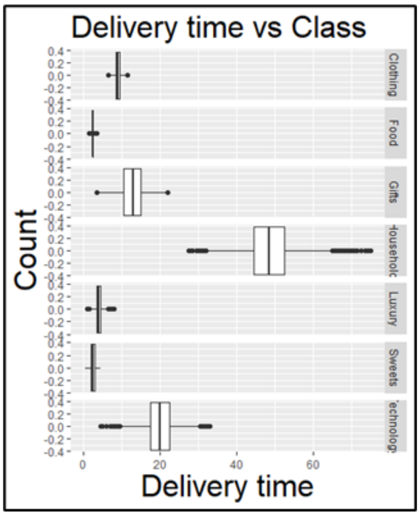
- P value is 2.2x10^-16 which is smaller than 0.05.

```
                Df Pillai approx F num Df den Df    Pr(>F)
Class            6 1.7576    42432     18 539862 < 2.2e-16
Residuals 179954

Class        ***
Residuals
---
Signif. codes:
0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
 Response Delivery.time :
                Df   Sum Sq Mean Sq F value    Pr(>F)
Class            6 33454043 5575674  629397 < 2.2e-16 ***
Residuals   179954  1594169       9
---
Signif. codes:
0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

 Response Price :
                Df     Sum Sq    Mean Sq F value    Pr(>F)
Class            6 5.7156e+13 9.5260e+12   80218 < 2.2e-16
Residuals   179954 2.1370e+13 1.1875e+08

Class         ***
Residuals
---
Signif. codes:
0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

 Response AGE :
                Df   Sum Sq Mean Sq F value    Pr(>F)
Class            6  8423938 1403990  3805.6 < 2.2e-16 ***
Residuals   179954 66390521     369
---
Signif. codes:
0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

Figure 55

Discussion manova 3

Age, price, and delivery time all have p values less than 0.05. this means that all 3 dependent variables change regarding the independent variable, in this case class of products.



Visual analysis of delivery time and class. It can be seen the household class has the longest delivery time but this class is not in control so this information will be misleading and inaccurate.

PART 6: Reliability of service and products

Problem 6:

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

$45 = k(0.04)^2$

$k = 45/(0.04)^2 = 28125$

loss function and shown in figure 56:

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

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

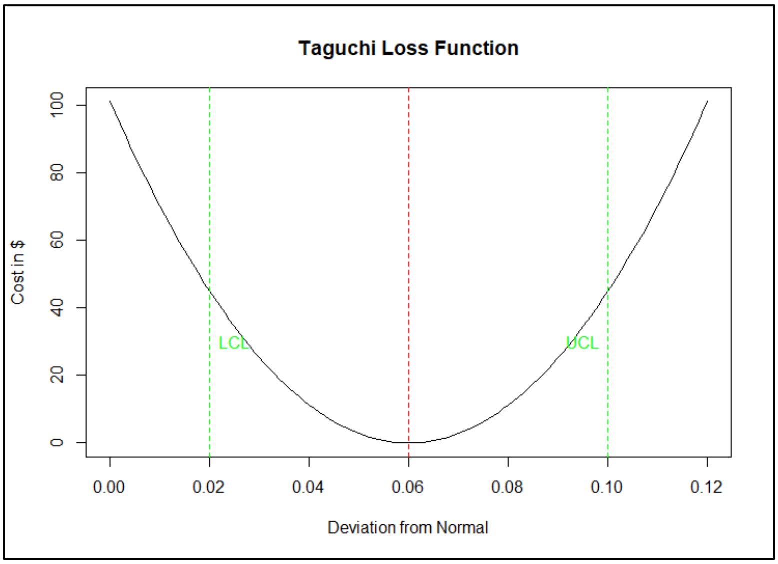


Figure 56

Problem 7:

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

$35 = k\ (0.04)^2$

k = 35/(0.04) 2 = 21875

loss function and shown in figure 57:

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

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

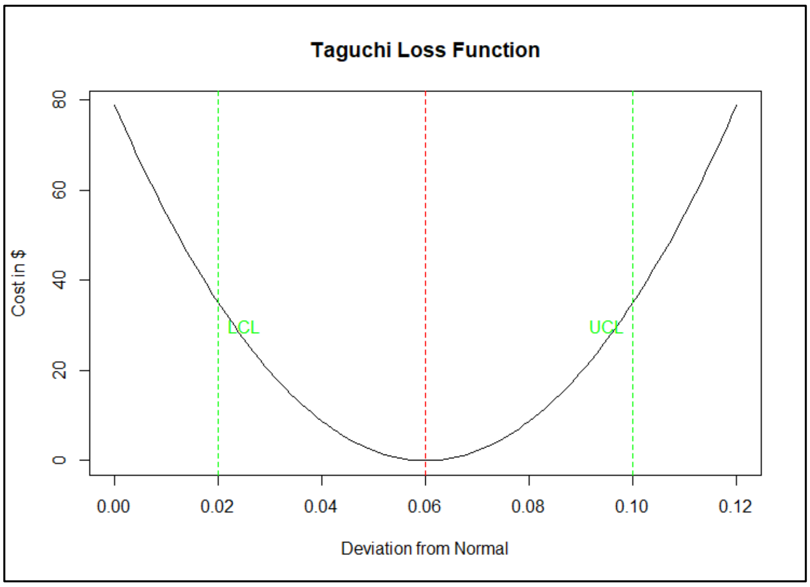


Figure 57

b) Process deviation (T) = 0.027

Taguchi loss: L (0.027) = 21875(0.027)^2 = \$15.95

Problem 27:

a) If there are no backup machines available, the system is in series and therefor the system reliability is as follows:

$R_T = R_a \times R_b \times R_c = 0.85 \times 0.92 \times 0.90 = 0.7038 = 0.704$

Therefore, the system is 70.38% (70.4%) reliable.

b) reliability of the system in parallel:

$R_T = 1 - [\,(1 - (R_a)^2) \times (1 - (R_b)^2) \times (1 - (R_c)^2]$

$= 1 - [(1 - (0.85)^2) \times (1 - (0.92)^2) \times (1 - (0.90)^2)]$

$= 0.9615$

The reliability of the system improved from 70.38% to 96.15%. The process went from a series system to parallel system.

6.3

Using the following formula:

$$P(x) = \binom{n}{x} p^x q^{n-x} = \frac{n!}{(n-x)!\,x!} p^x q^{n-x}$$

where
n = the number of trials (or the number being sampled)
x = the number of successes desired
p = probability of getting a success in one trial
$q = 1 - p$ = the probability of getting a failure in one trial

Figure 58

Days (1560)	Vehicles (21)	Days (1560)	Drivers (21)
190	20	95	20
22	19	6	19
3	18	1	18
1	17	-	-

Table 10

6.3.1: Expected delivery times in days per year: first we calculate the reliable vehicles in given year:

- We have 5 calculations to do using the binomial equation due to the fact we have 4 different sets of days provided and included the total days from 1560 days. This is seen in table 3 above.
- Calculations below used the binomial equation in figure 56:

- p (0) for 1344 days
  p= 0.0070716611
- p (1) for 190 days
  p= 0.0066218917
- p (2) for 22 days
  p= 0.0089080848
- p (3) for 3 days
  p= 0.1217169284
- p (4) for 1 days
  p= 0.0197086603

- Weighted p= (0.0070716611x1344) +(0.0066218917x190) +(0.0089080848x22) +(0.1217169284x3) +(0.0197086603x1)/1560 = 0.0070606882

Expected reliability:

- P = 0.8617411094
- P = 314.53550496 (In days)

Reliability drivers:

$$P(x) = \binom{n}{x} p^x q^{n-x} = \frac{n!}{(n-x)!\,x!} p^x q^{n-x}$$

where
n = the number of trials (or the number being sampled)
x = the number of successes desired
p = probability of getting a success in one trial
$q = 1 - p$ = the probability of getting a failure in one trial

- There are 4 calculations required, firstly the total number of days out of 1560 days, plus the 3 sets of days provided. This is seen in table 3 above.

- P (0) for 1458 days
  p= 0.0032397424
- P (1) for 95 days
  p= 0.0030613121
- P (2) for 6 days
  p= 0.0044668560

- P (3) for 1 days
  p= 0.0082453462

- Weighted p = (0.0032397424x1458) +(0.0030613121x95) +(0.0044668560x6) +(0.0082453462x1)/ 1560 = 0.0068181750

Expected reliability:

- P = 0.8661717832
- P = 316.152700879 (In days)

Total reliability:

- $P_{rel} = P_{veh} \times P_{driv}$= 0.8617411094 x 0.8661717832= 0.7464158339

Total reliability = 272.4417794 (In days)

Therefore, 273 out of 365 days of the year will have a reliable delivery service.

6.3.2: Number of vehicles increased by 1 to 22.

- P (21) = 0.8556566241
- P (20) = 0.1338586773
- P (19) = 0.0099944790
- P (18) = 0.0004737980
- P (17) = 0.00001600336

Reliability =0.8556566241+0.1338586773+0.0099944790+0.0004737980+0.00001600336= 0.9999995819

Reliability in days= 364.9998474 (In days)

To conclude, increasing the number of vehicles will increase the reliability. This should be implemented as it was only increased by 1 vehicle from 21 to 22.

Conclusion

With the client data provided for the online business multiple steps were taken to sort into valid and invalid data, multiple charts were constructed to analyse and get a better understanding of the data.

The graphs in part 2 gave an understanding of how the different classes link and the effects each one has on another. In part 3 SPC charts and tables were constructed, as seen household, gifts and luxury classes were proven to be out of control regarding delivery times with technology, food, gifts, sweets were all in control. The classes proven out of control need attention for improvement or change.

Reliability probabilities were calculated for vehicles and drivers and was seen that increasing the vehicles from 21 to 22 would significantly benefit the online business delivery times.

References

Hernandez F, 2015. Data Analysis with R. [Online]. Available at: http://fch808.github.io/Data-Analysis-with-R-Exercises.html . [Accessed on 20 September 2022].

Hessing T. X Bar R Control Charts. [Online]. Available at: https://sixsigmastudyguide.com/x-bar-r-control-charts/ . [Accessed on 2 30 September 2022].

STHDA. MANOVA Test in R: Multivariate Analysis of Variance. [Online]. Available at: http://www.sthda.com/english/wiki/manova-test-in-r-multivariate-analysis-of-variance . [Accessed on 28 September 2022].

Salazar, R, 2020. Loss Function Analysis with R. [Online]. Available at: https://towardsdatascience.com/loss-function-analysis-with-r-6d14ee79f6c4#:~:text=The%20Taguchi%20loss%20function%20has,risk%20of%20having%20more%20variation .[Accessed on 2 October 2022].

Wickham, H. Ggplot2. [Online]. Available at: https://ggplot2.tidyverse.org/ . [Accessed 5 October 2022].