Quality Assurance 344

ECSA Graduate Attributes Report

C.J. Rossouw

23796693

21 October 2022

Abstract

The report covers a deep analysis of sales data from an online business. The raw data underwent data wrangling to prepare it for analysis usage. The data was visualized and better understood with the use of descriptive statistics. X-bar and S-bar control charts were plotted for each product class, to determine the stability od the class's delivery time. The probability of predicting type I and II errors were calculated. MANOVA was performed to identify the impact that features have on each o

Table of Contents

List of Tables

Introduction

This report will analyze and discuss sales and client data that has been gathered from an online business. Statistical analysis will be performed to extract information and gain insights regarding the current performance of the business systems. The aim is to use these insights to improve the performance of the business systems and minimise current shortcomings.

The data will undergo data wrangling to remove unwanted data that will make predictions inaccurate. There after all data that is in agreement with the validation process of the data wrangling will be analysed. With the use of descriptive statistics valuable insights and trends will be identified, as well as also help to better visualize the data. Once a better understanding of the data is reached statistical process control will be performed on the delivery time data of different product classes sold by the business. The statistical process control will be used with the help of X-bar & S control charts. Further analysis of the type I and type II errors of the delivery time data will be used, in conjunction with the process control information, to optimize the delivery time of sales and reduce overall costs. Information from another business will be looked at to determine the reliability of the service and products.

1. Data Wrangling

The process of data wrangling involves cleaning and unifying a raw data source, with the aim of increasing the usefulness of the data for usage in data analytics. (Patel, 2018). The data that was obtained from the business contains data entries that should be deemed as invalid, as these entries might skew the data or create an inaccurate representation of the data. These invalid data entries are entries that contain negative values or entries with missing values, represented by "NA". The invalid data needs to be removed from the valid data and these two datasets needs to be stored separately.

Of the 180 000 entries within the dataset, there are 17 entries where the price is not entered (NA). There are also 5 entries where the price of the entry has a negative value, R -588.80 for all 5. This is illogical, as a sale item can not have a negative price. This might be a result of a clerical error, where a purchased item that was returned was recorded as part of the sales data.

Upon further inspection of the different numerical features of the dataset, there were no other invalid cases. The 22 cases of invalid data was removed from the dataset. The valid data set was created, which contains 179978 entries of complete data. The invalid dataset can be seen below in Table 1. A segment of the last 6 entries of the valid dataset can be seen in Table 2.

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

Table 1: Invalid Dataset

Primary Key	X	ID	AGE	Class	Price	Year	Month	Day	Delivery.time	Why.Bought
179973	179995	49178	82	Food	505.88	2024	2	20	2.5	Website
179974	179996	65414	31	Gifts	3147.66	2026	2	1	13	Recommended
179975	179997	57864	34	Gifts	1111.36	2023	6	4	10	Recommended
179976	179998	48301	77	Gifts	3943.92	2028	4	29	17	Website
179977	179999	96502	56	Sweets	243	2023	5	26	2	Website
179978	180000	71587	53	Household	15362.39	2021	8	22	43.5	Website

Table 2: Valid Dataset

2. Descriptive Statistics

After the invalid data has been removed, the valid dataset contains 179978 entries, each with 11 features. The dataset as a whole will be visualised and analysed in R studio, before being separated into smaller selections, such as separating it by class or reason for purchase. Table XXX shows the data type and cardinality of each feature in the dataset. The dataset has 9 numerical features, of which 7 are integers, and 2 categorical features. Based on the low cardinalities of Year, Class and Why Bought these features are categorical features. Price and delivery time are continuous features.

Feature	Primary Key	X	ID	AGE	Class	Price	Year	Month	Day	Delivery Time	Why Bought
Data Type	Integer	Integer	Integer	Integer	Categorical	Numerical	Integer	Integer	Integer	Numerical	Categorical
Cardinality	179978	179978	150000	91	7	78832	9	12	30	148	6

Table 3: Data Overview

2.1. Analysis of Numerical Features

Table XXX displays a 5 point summary of all the numerical features, along with the mean and standard deviation of each feature.

	Primary Key	X	ID	AGE	Price	Year	Month	Day	Delivery.time
Minimum	1	1	11126	18	35.65	2021	1	1	0.5
1st Quartile	44995	45004	32700	38	482.31	2022	4	8	3
Median	89990	90005	55081	53	2259.63	2025	7	16	10
3rd Quartile	134984	135000	77637	70	15270.97	2027	10	23	18.5
Maximum	179978	180000	99992	108	116618.97	2029	12	30	75
Mean	89990	90003	55235	54.57	12294.1	2025	6.521	15.54	14.5
Standard Deviation	51955.32	51960.7	25740.27	20.38881	20889.15	2.783364	3.453849	8.648721	13.95578

Table 4: Summary of Numerical Features

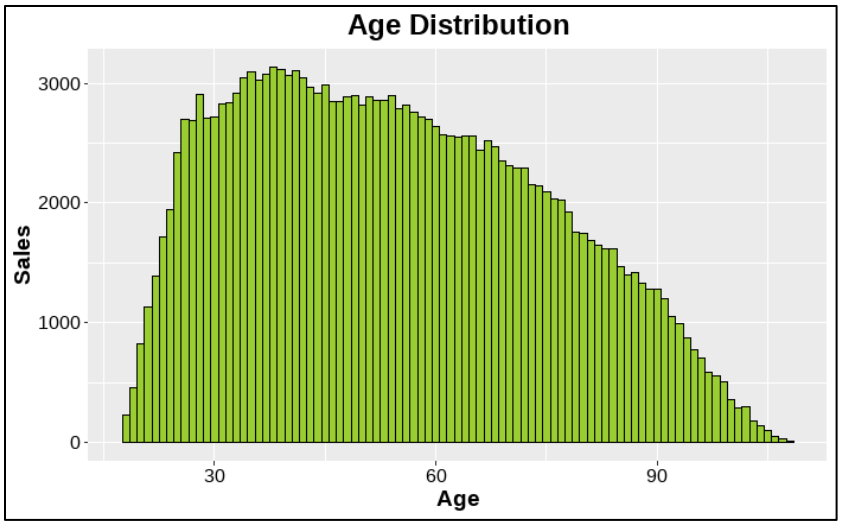
Primary Key, X and ID: All three these features have a very high cardinality, with Primary Key and ID having a cardinality that is the same as the number of instances. Due to this all three these features will not be able to provide insightful and accurate information with regards to the data. These features will not be used in any further analysis or visualisations.

AGE: The age of the customers ranges from 18 to 108. The mean is 54.57 years and the standard deviation is 20.39 years. Since the age data has a large range and standard deviation further analysis will need to find possible trends in the age data. An example of this can be to find the target audience age of different product classes.

Price: The cheapest product has a price of R 35.65, whilst the most expensive product costs R 116 618.97. With a standard deviation of R20 889.15 it is evident that the price feature will have to be analysed further. Similarly to AGE, it might be insightful to break up the feature by class or an other categorical feature to find meaningful information.

Delivery Time: The mean delivery time is 14.5 days, with a standard deviation of 13.96 days. The shortest delivery time is half a day, whilst the longest is 75 days. The delivery time feature will also need to be analysed further.

2.1.1. Age



The distribution of customer age is unimodal, and it is skewed to the right. The mean age is 54.57. 50% of all sales are made to people between the age of 38 and 70.

Figure 1: Age Distribution

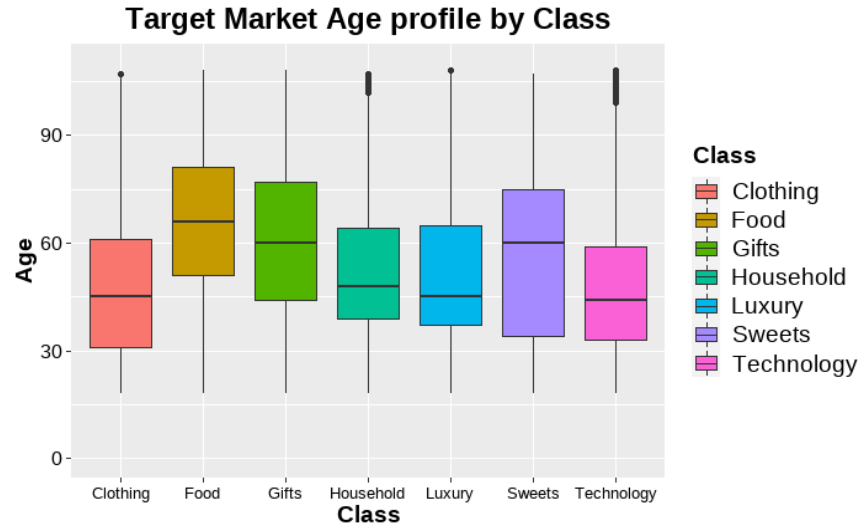


Figure 2: Age Distribution per Class

From Figure 2 it can be seen that all classes are distributed quite similarly over a large age range. Food has the highest mean age, at around 75. This is likely due to old people, who are not physically able to bay food in stores make use of online services. Technology has the lowest mean age This makes sense, as younger people are more likely to buy technology online, rather than in stores, as they are likely to know more about the item they want to purchase, and will not need in store assistance. The age distribution of sweets is bimodal. The rest of the features all have unimodal distributions. The distribution of clothing, household, Luxury and Technology is skewed to the right. Food is skewed to the left and gifts appears to not be skewed.

2.1.2. Price

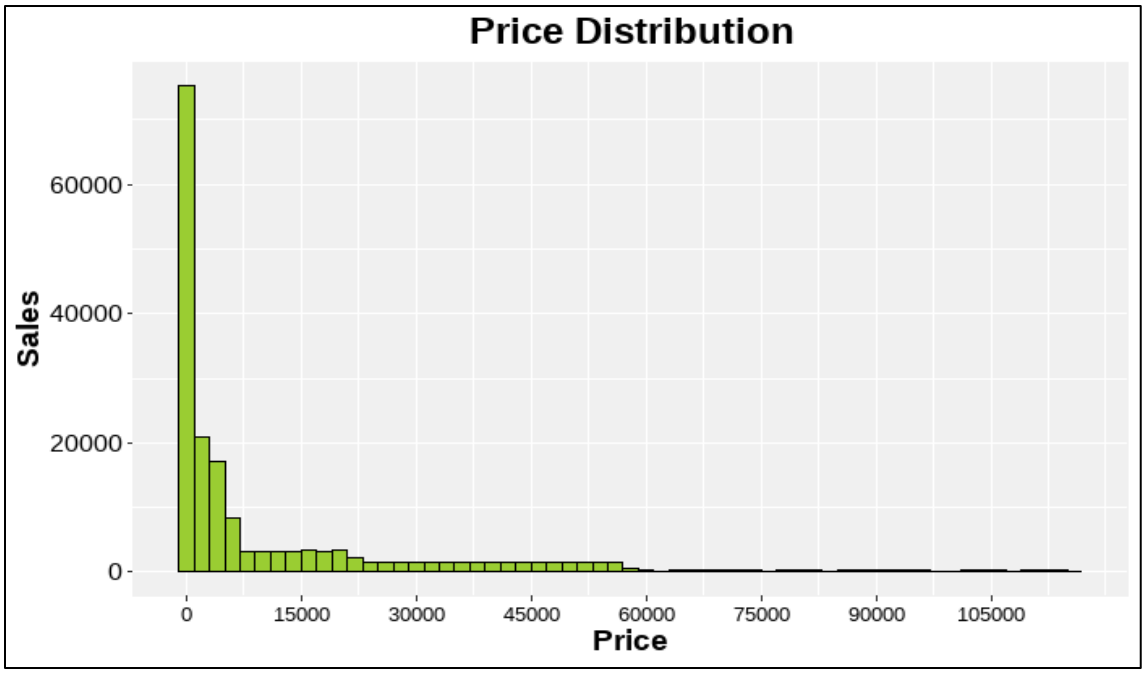


Figure 3: Price Distribution

Price is distributed unimodally and skewed to the right. With a mean price of R 12 294.10, Figure 3 visualizes just how drastically the distribution is skewed to the right. Most items have a low price, but there are items with an extremely large price.

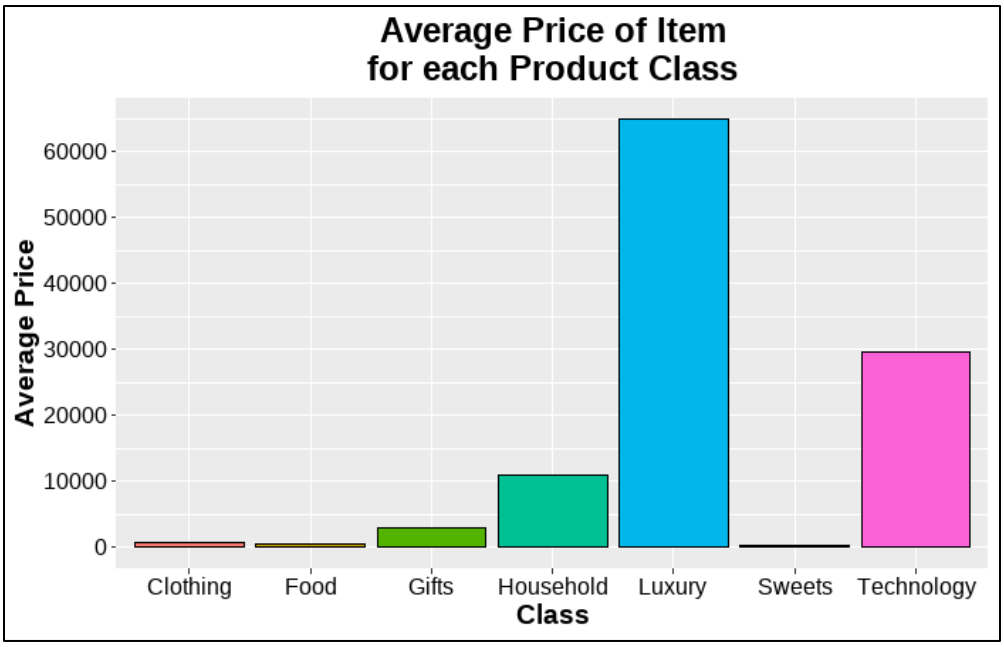


Figure 4:Average Price of Each Class

Figure 4 shows the average price per item for each class. 4 Classes has an average price well below R 5000, with household and technology having an average price between R 1000 and R 30 000. Luxury items, however, has an average price of over R 60 000. This is likely the cause of the dramatic skew to the right of the overall distribution.

2.1.3. Delivery Time

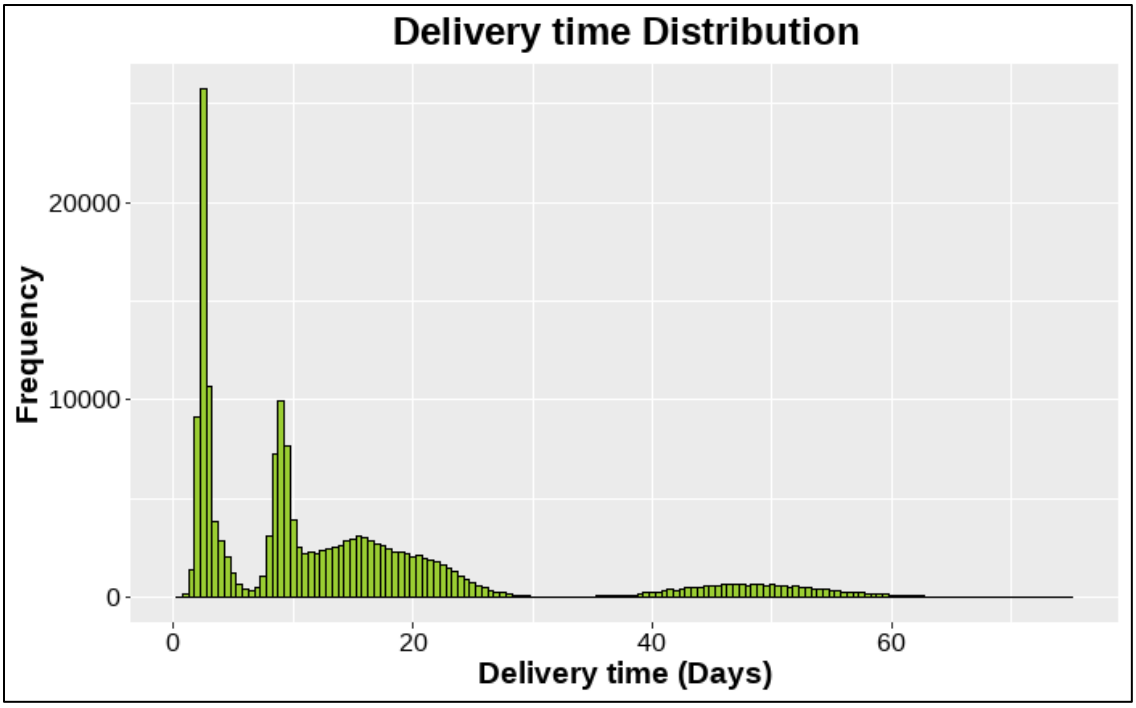


Figure 5: Delivery Time Distribution

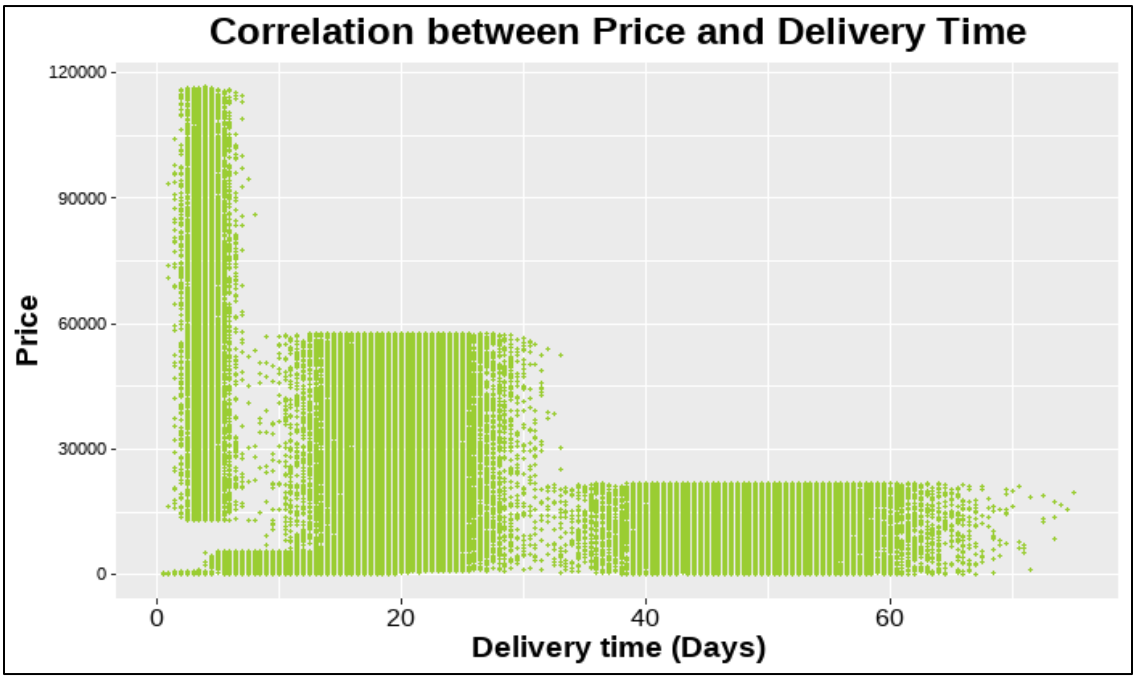


Figure 6: Correlation between Price and Delivery Time

Figure 5 shows that the delivery time distribution is has a bimodal shape and is skewed to the right. 3 – 5 day delivery time has the highest frequency. The distribution also follows a normal approximation between 30 and 70 days. Figure 6 illustrates that there is a significant correlation between the price of an item and the delivery time. There are 3 clear groups: items with a price between R 30 000 and R 120 000, that are delivered within about 8 days. A second grouping of items is items with a price below R 60 000, which have a delivery time anywhere between 0 and 33 days. The final grouping that can be seen are the items that has a delivery time that is larger 35 days. These items have a price below R 25 000. These groupings will be further analysed with the help of Figure 8.

Figure 7: Delivery Time Distribution per Class

Figure 7, in conjunction with Figure 5, provides a lot of information regarding the distribution of delivery time. It shows that the bimodal distribution that is skewed to the right is influenced differently by each class. The normal distribution of delivery times between 30 and 70 days is almost entirely just from Household items. Luxury, sweets, clothing and food items form the most frequently occurring delivery time between 3 – 5 days. The second peak of the bimodal distribution represents the delivery time of gifts and technology.

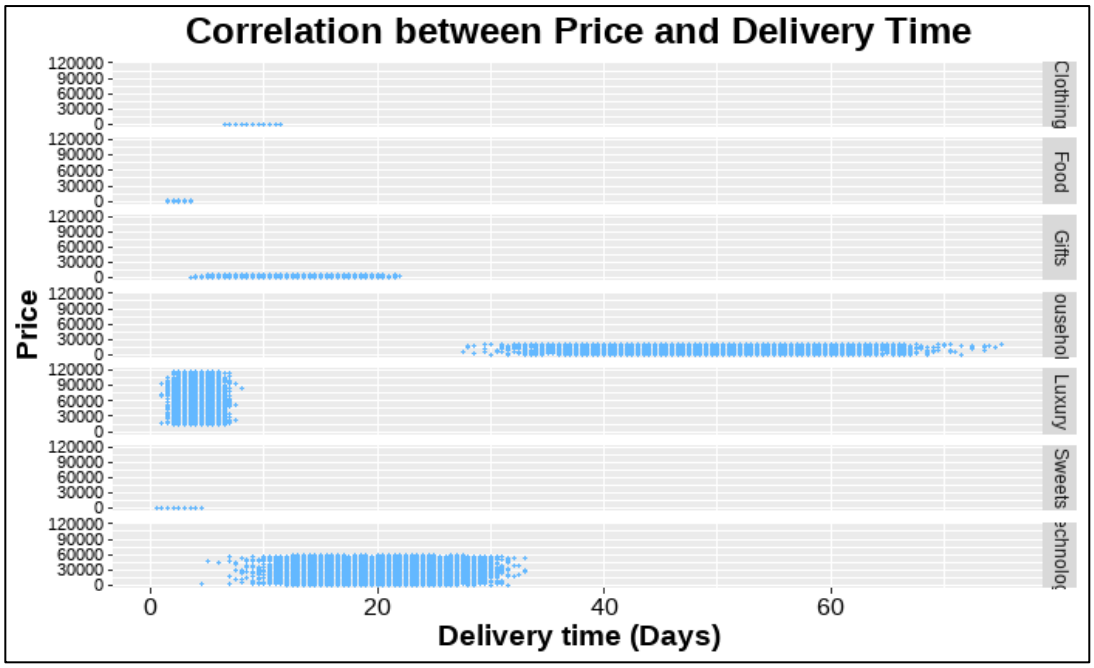


Figure 8: Correlation between Price and delivery Time of each Class

Figure 6 can be divided into its parts by Figure 8 in the same way that Figure 7 helped to explain Figure 5. It is clear that the first grouping, of high value items with short delivery times, are the luxury items.

Clothing, Food, Gifts, Sweets and Technology make up the middle group. The final group of items with a very long delivery time is the Household items.

2.2. Analysis of Categorical Features

2.2.1. Yearly Trends

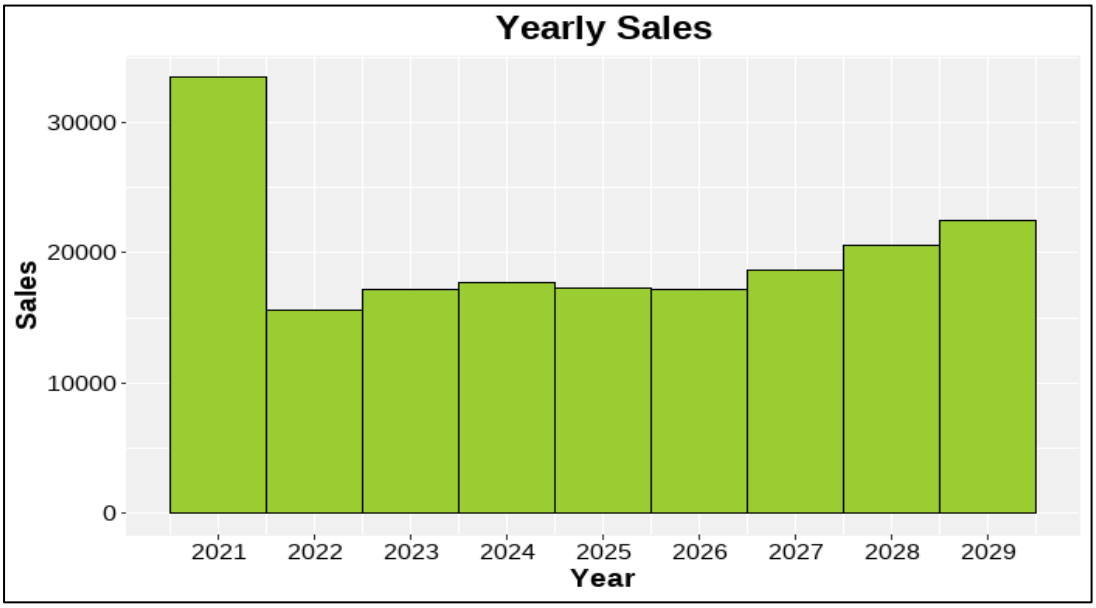


Figure 9: Yearly Sales

Figure 9 shows that the business had the most sales in 2021. 2022 was a steep drop off in the number of sales. The number of sales dropped by more than 50% from 2021 to 2022. From 2022 to 2029 there is a steady increase in the amount of yearly sales. Figure 10 compliments Figure 9 in illustrating that 2021 was the year with the most sales. Clothing and Household were the two largest contributors to 2021s success. These were also the 2 classes that had the biggest drop in the number of sales in 2022, where the number of yearly sales remained almost the same until 2029.Food, sweets and technology shows a trend of steady increase in yearly sales across the whole period. There are slight fluctuations in the yearly sales of Luxury items, but it remains largely the same.

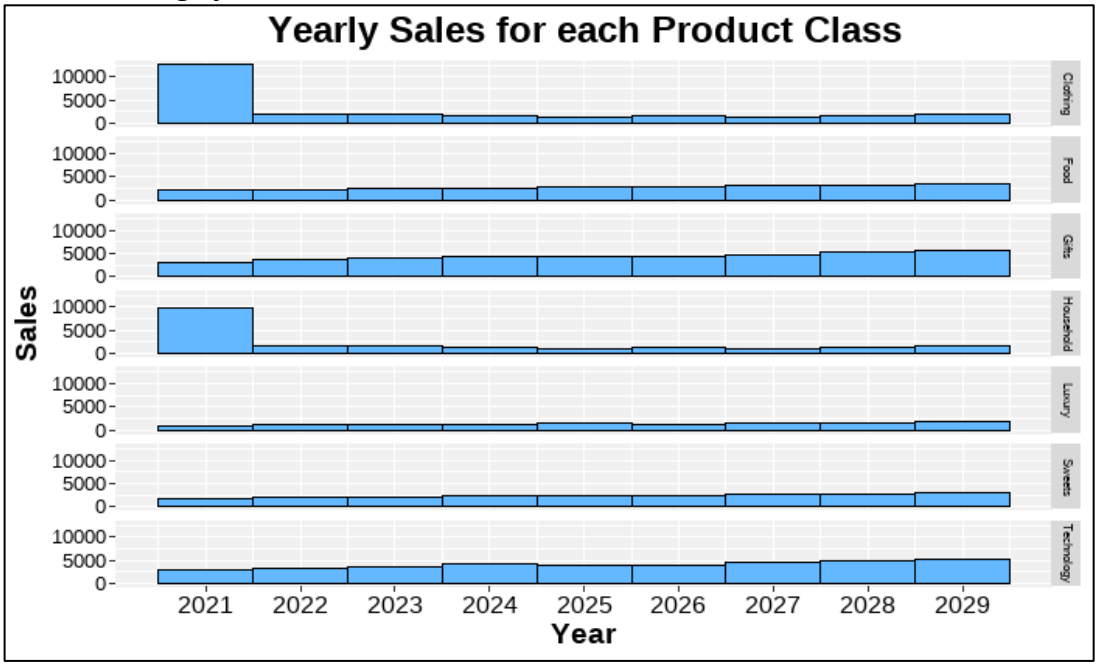


Figure 10:Yearly sales per Class

2.2.2. Why Bought

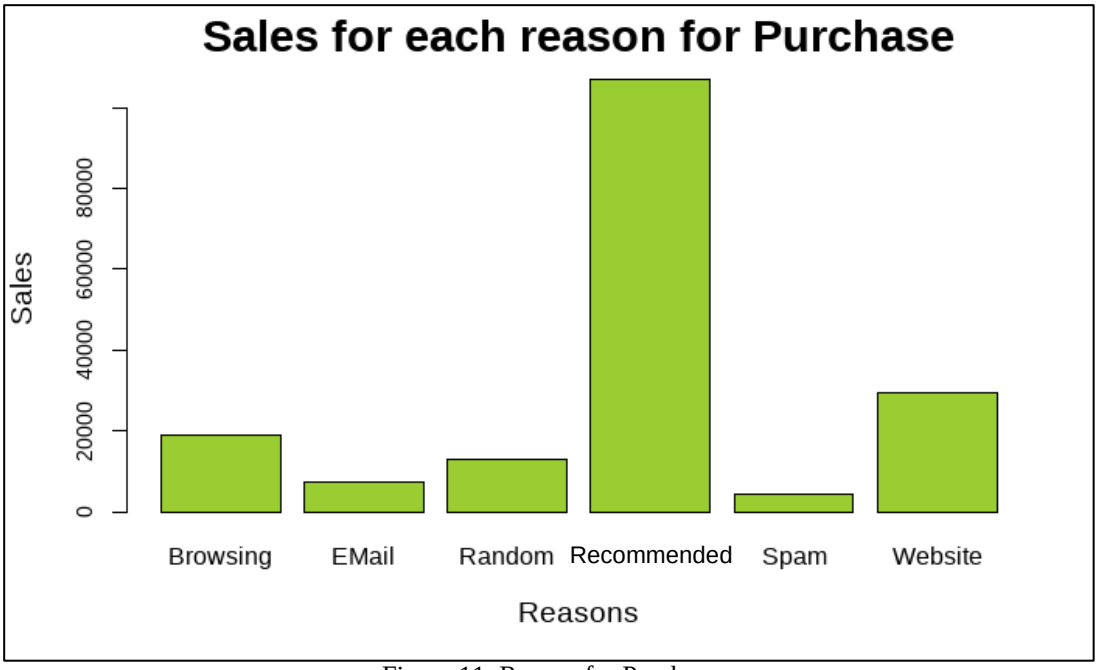


Figure 11: Reason for Purchase

Almost half of all the sales that were made were due to recommendations. Spam and email marketing were the smallest contributors to sales. Figure 11 is very useful for the marketing department of the business. They need to focus on their internet presence, as browsing and their website were 2 of the largest reasons that items were sold. The most effective way, however, is word of mouth.

2.2.3. Class

Figure 12: Sales per Class

Gifts and Technology were the items with the highest quantity of sales. Whilst a smaller number of luxury and household items were sold, due to the relatively higher unit price, it would have made up a larger % in revenue than it did in the number of sales.

2.3. Process Capabilities Indices

Process capability indices are used to measure the ability of a process to deliver an output that conforms to the specifications that was set out for the process (Evans & Lindsay, 2018). In order to evaluate the performance of the online business the sales data needs to be analyzed by computing the following indices:

Upper Specification Limit (USL) = 24 hours

Lower Specification Limit (LSL) = 0 hours

Standard Deviation of the Process (σ) = 3.501993

Process Mean (μ) = 20.01095

Process Capability (C_P)

$$C_P = \frac{USL - LSL}{6\sigma}$$
$$C_P = 1.142207$$

Upper One-Sided Index (C_{PU})

$$C_{PU} = \frac{USL - \mu}{3\sigma}$$
$$C_{PU} = 0.3796933$$

Lower One-Sided Index (C_{PL})

$$C_{PL} = \frac{\mu - LSL}{3\sigma}$$
$$C_{PL} = 1.90472$$

Process Capability Index (C_{PK})

$$C_P = \min(C_{PL}, C_{Pu})$$
$$C_P = 0.3796933$$

Process Capability Indices	Value
C_P	1.14
C_{PU}	0.380
C_{LU}	1.90
C_{PK}	0.380

Table 5: Process Capability Indices

Why is a LSL of 0 logical?

A Lower Specification Limit of 0 is logical, as delivery time cannot be negative. An item can not be delivered in less than 0 days.

3. Statistical Process Control (SPC)

Statistical Process Control (SPC) is a method used to monitor outliers and problems that can effect the system. A control chart is plotted with the UCL (Upper Control Limit), CL (Centreline) and LCL (Lower Control Limit). These lines are used as a reference for the data to see whether it is under control or not.

For the control charts the delivery time of each class is plotted and compared to the reference lines. The data needs to be in a ascending date order. The data therefor needs to be ordered so the oldest data is selected first.

3.1. X- and S- Chart Limits

The control limits for each class is represented in Table 6, which contains the values used for the X-bar charts, and in Table 7, which represents the values for the S-Chart. The headings of each column represents the following:

- UCL: (Upper Control Limit) Three standard deviations above the mean.
- U2Sigma: (Upper 2-sigma Control Limit) Two standard deviations above the mean.
- U1Sigma: (Upper 1-sigma Control Limit) One standard deviation above the mean.
- CL: (Center Line) Mean delivery time of each class.
- L1Sigma: (Lower 1-sigma Control Limit) One standard deviation below the mean.
- L2Sigma: (Lower 2-sigma Control Limit) Two standard deviation below the mean.
- LCL: (Lower Control Limit) Three standard deviation below the mean.

X-Chart

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

Table 6: X Chart Limits

S-Chart

Class	UCL	U2Sigma	U1Sigma	CL	L1Sigma	L2Sigma	LCL
Technology	5.180570	4.552222	3.923875	3.295528	2.667181	2.038833	1.410486
Clothing	0.866560	0.761455	0.656351	0.551247	0.446142	0.341038	0.235934
Household	7.344180	6.453410	5.562640	4.671870	3.781100	2.890330	1.999560
Luxury	1.511052	1.327777	1.144503	0.961229	0.777955	0.594680	0.411406
Food	0.437247	0.384213	0.331180	0.278147	0.225113	0.172080	0.119047
Gifts	2.246333	1.973877	1.701421	1.428965	1.156509	0.884053	0.611597
Sweets	0.835339	0.734022	0.632704	0.531386	0.430069	0.328751	0.227433

Table 7: S Chart Limits

3.2. Examples of X- and S- Chart initializations

The first 30 samples are used to estimate the mean and standard deviation of each class's delivery time. These means and standard deviations are used to construct the control limits for each class. The first 30 samples of each product class is plotted on both the X- and S-charts. Figure XXX and XXX are examples of each X- and S- chart initialization for the first 300 samples. The other initializations can be viewed in Appendix A -B.

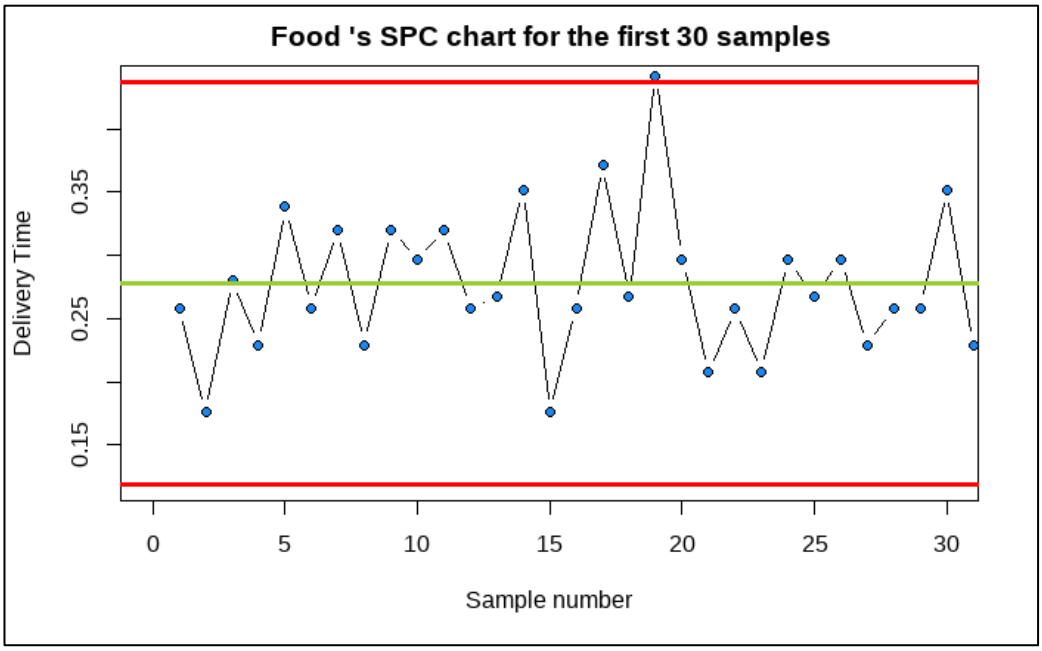


Figure 13: Food's S-Chart for the first 30 Samples

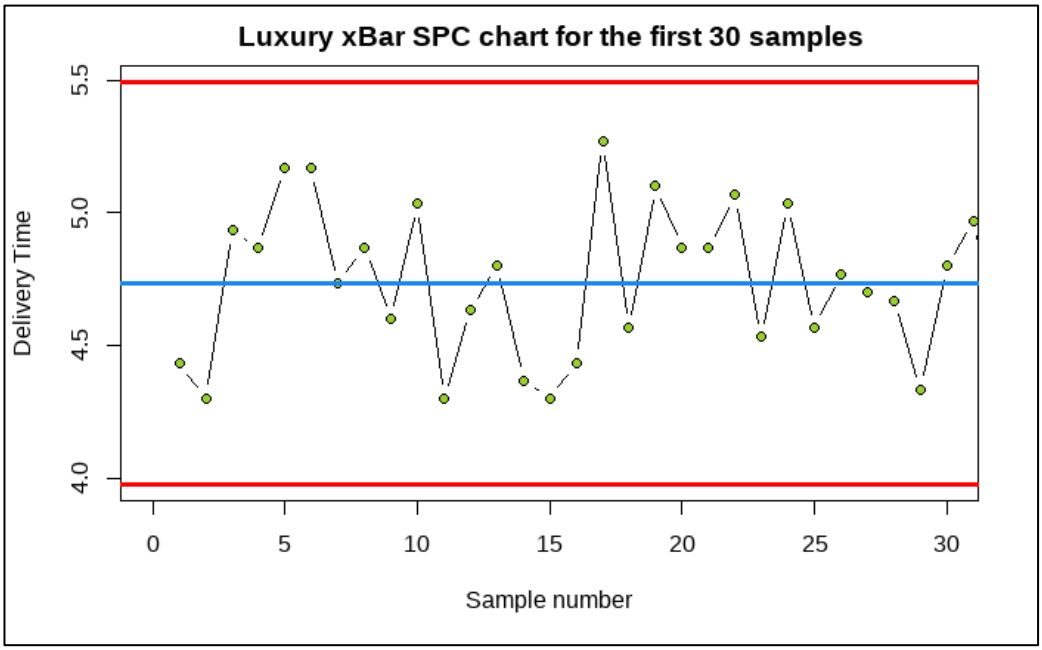


Figure 14: Luxury's X-Chart of the first 30 Samples

The X-Charts of all 7 classes are under control (between the control limits). The S-Charts of Food and Sweets shows these two classes are not under control (Figure 13). The S-charts of the other classes are within the control limits.

3.3. Process Control

The process of plotting the samples from each class on the X- and S-charts is now done with all samples. Points that are outside the control limit, indicates that the mean of the class is out of control. These graphs are a useful tool to determine whether a process is stable or unstable. Figure 15, 16 and 17 are examples of the control chart initializations for all samples. The initializations for the remaining classes on the X- and S-Chart can be seen in Appendix C – D.

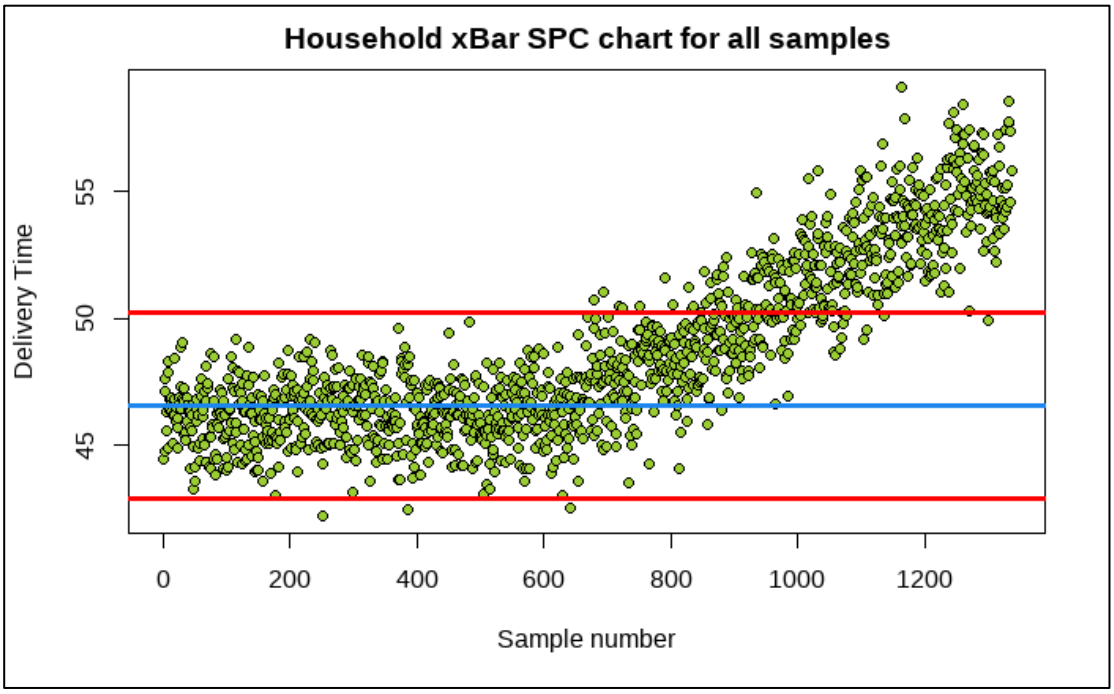


Figure 15: Household X-Chart for all Samples

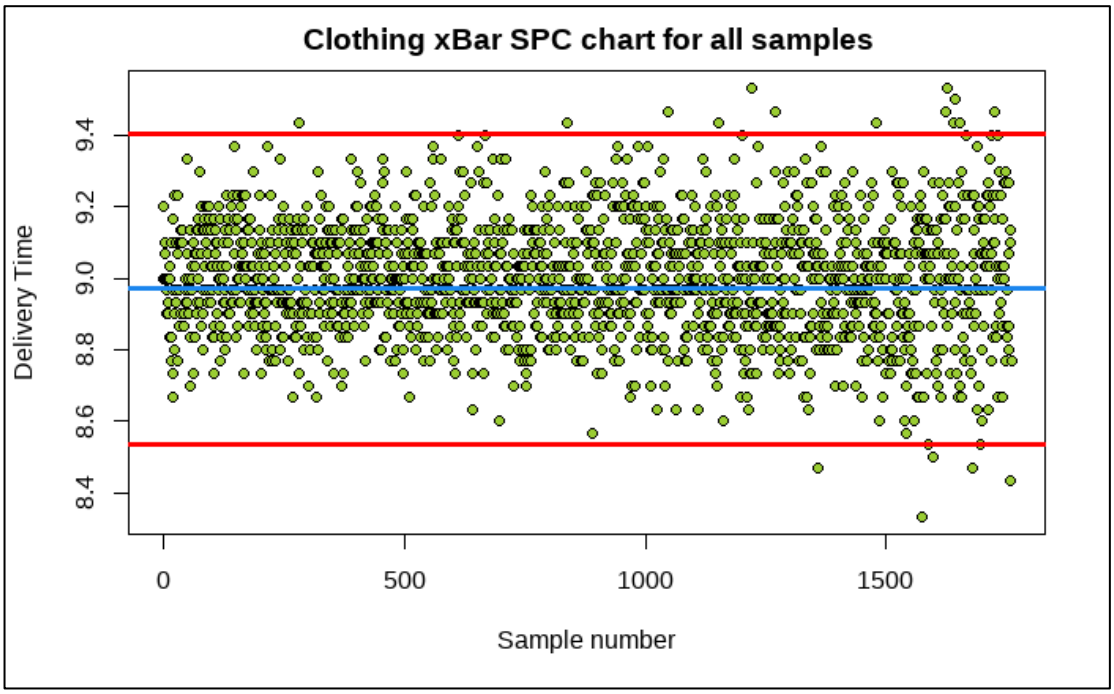


Figure 16: Clothing X-Chart for all Samples

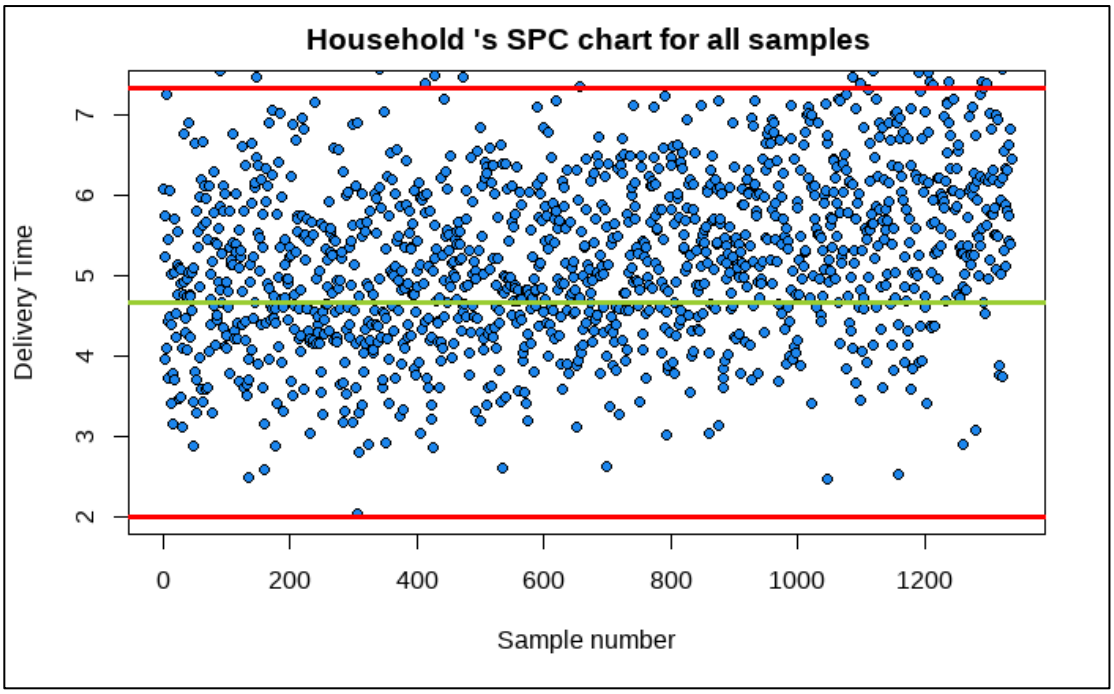


Figure 17: Household S-Chart for all Samples

The X-Charts of gifts, household and luxury show that the classes are unstable and out of control. The performance of these classes needs to be improved. The X-charts of the other classes are for the most part within the control limits. There are some samples that are slightly outside the limits, but the majority of samples are within the control limits. The S-charts of all the samples for all classes are very similarly distributed. The majority of samples are within the control limits, and are therefor mostly stable. Each class has a handful of samples that are outside the control limits, but not so far outside as seen in some of the X-charts.

4. Optimizing the delivery processes

4.1. Indications of out of control samples

4.1.1. (A) X-bar/ Sample means outside the outer control limits

Class	1st	2nd	3rd	3rd Last	2nd Last	Last	Total
Clothing	282	837	1048	2695	2723	1756	20
Household	252	387	643	2335	2336	2337	395
Food	75	432	-	-	1149	1408	4
Technology	37	345	353	1933	2009	2071	19
Sweets	942	1243	-	-	1294	1358	4
Gifts	213	216	218	1320	1321	1322	2287
Luxury	142	171	184	789	790	791	440

Table 8: Samples Outside the X-Chart Outer Control Limits

4.1.2. (B) Most consecutive samples of S-bar between sigma control limits

Table 9 shows the most consecutive samples of outside the -0.3 and +0.4 sigma control limits.

Class	Maximum number of consecutive samples Count	Last Sample Number
Clothing	4	223
Household	4	94
Food	5	756
Technology	6	1776
Sweets	3	45
Gifts	7	2477
Luxury	4	63

Table 9: Most Consecutive Samples outside S-bar Control Limits

The values of Table 8 and 9 was calculated in R. The graphs under Appendix E help to visualize all the samples that are outside the X-bar control limits. Figure 18 is an example of these graphs. The figure clearly illustrates that the first 3 samples that are outside the means are far from each other, with the majority of out of control samples being at the end.

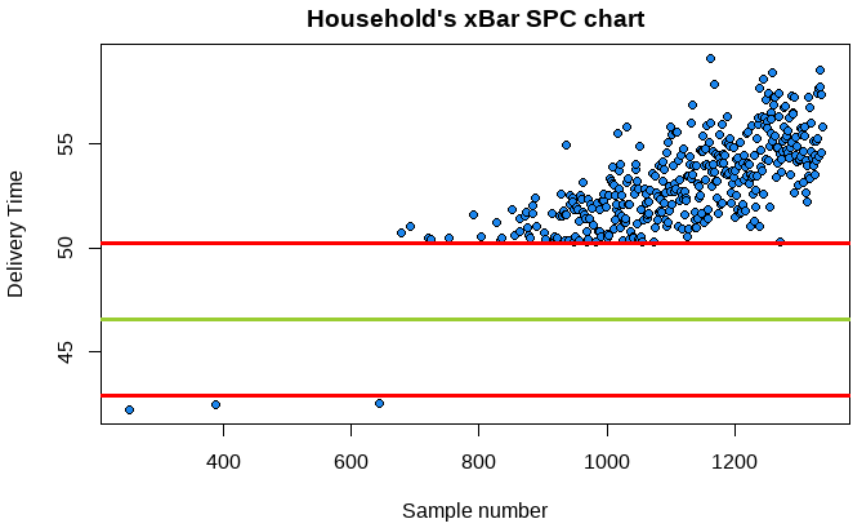


Figure 18: Household Samples outside of the X-Chart Limits

4.2. Type I Error

A Type I error, also described s a false positive error, is when the null hypothesis is rejected, when it should be accepted. The null hypotheses for the Manufacturer's Error states that a process is in control. A type I error is therefor the probability that The probability of making a type I error is therefor the probability of a manufacturer predicting a failure at a process, when in actual fact the process is actually working.

For A: The probability of making a type I error can be calculated in R with the following formula:

$$\big(1 - \text{pnorm}(3)\big) + \big(\text{pnorm}(-3)\big)$$

$$= 0.002699796$$

For B: The probability of making a type I error can be calculated in R with the following formula:

$$\big(\text{pnorm}(-0.3)\big) + \big(1 - \text{pnorm}(0.4)\big)$$

$$= 0.7266668$$

Therefor, the probability of making a Type I error for A, identifying X-bar samples that are not within the control limits, is 0.26998%. The probability of making a Type I error for B, finding the most consecutive S-bar samples between -0.3 and 0.4 sigma, is 72.667%.

4.3. Optimizing the delivery process

There is a cost involved with each item that is delivered late. The penalty for late delivery is R 329 per hour. It costs R2.5 per item to reduce the average delivery time by 1 hour. The optimal delivery time was calculated in R. All delivery times were considered. In order to ensure maximal profit, the delivery needs to be centered. The optimal delivery time (in hours), to center the delivery process, is 12.72 h. Figure 19 illustrates the calculation process.

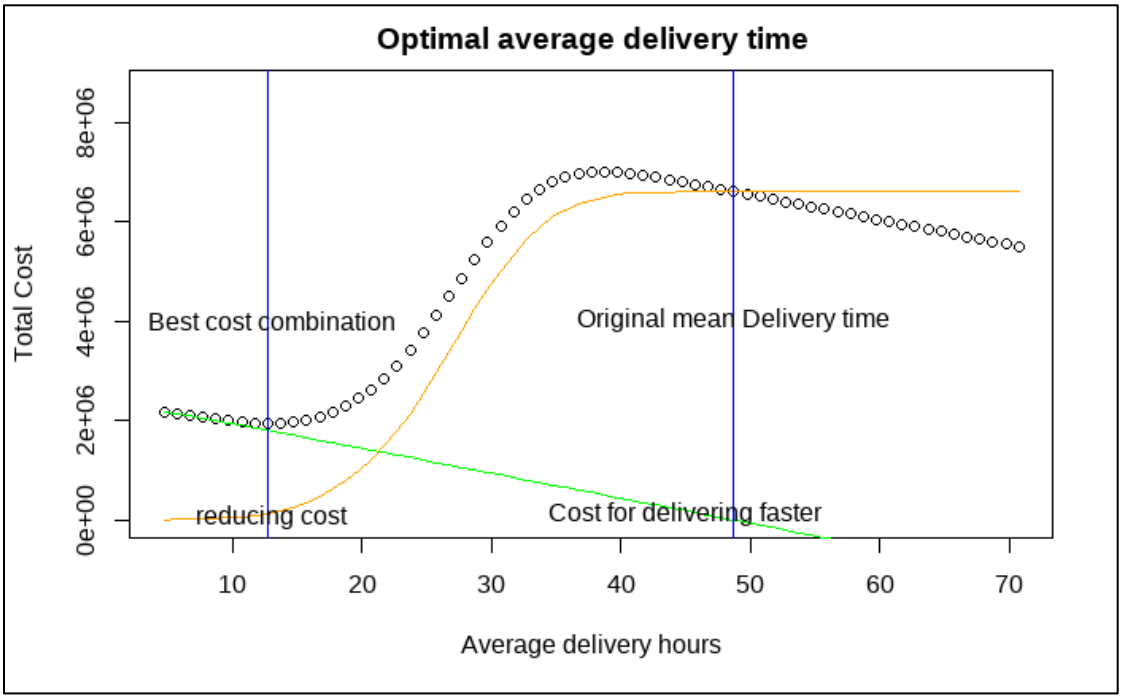


Figure 19: Optimal Average Delivery Time

4.4. Type II Error

A Type II error is a false negative. The null hypothesis is not rejected, but it should be. The null hypothesis tests if the process is under control. A Type II error happens when a test predicts that a process is under control, when in reality it is not. The probability of a Type II error for a is calculated as follows:

```
sdev2<- (UCLx[1]-LCLx[1])/6
type2A<- pnorm(UCLx[1],mean = 23, sd = sdev2)-pnorm(LCLx[1], mean = 23, sd = sdev2)
type2A #0.4883177
```

Figure 20: Type II Error

The probability of making a type II error for A is 48.83%.

5. DOE and MANOVA

Multivariate analysis of variance, MANOVA, is an expansion of analysis of variance, ANOVA. This type of analysis uses more than one dependent variable. The MANOVA will be used to determine if class has an impact on the price and delivery time of a sales item. The MANOVA will be performed with $\alpha = 0.05$ and the following Null- and Alternative Hypotheses will be tested:

Price:

H_0 = The class of an item does not have an significant impact on the price of it.

H_1 = The class of an item has an significant impact on the price of it.

Delivery Time:

H_0 = The class of an item does not have an significant impact on the delivery time of it.

H_1 = The class of an item has an significant impact on the delivery time of it.

Using the functions in R the MANOVA can be calculated. The results can be seen in Figure XXXX:

```
 Response Price :
               Df         Sum Sq       Mean Sq F value    Pr(>F)
Class           6 57168427663229 9528071277205   80258 < 2.2e-16 ***
Residuals  179971 21365723828547     118717592
---
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 33458565 5576427  629429 < 2.2e-16 ***
Residuals  179971  1594452       9
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1
```

Figure 21: MANOVA Results

The P value for both price and delivery time is 2.2×10^{-16}. This is significantly smaller than the chosen level, $\alpha = 0.05$. Therefor both null hypotheses are rejected with a high level of certainty. The class of an item does have a significant affect on the price and delivery time of the item. The result of the MANOVA is backed up by Figure XXX and Figure XXXXX.

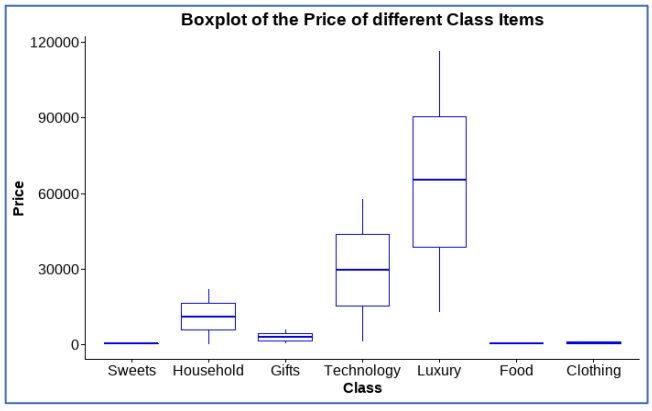


Figure 22: Boxplot of Price per Class

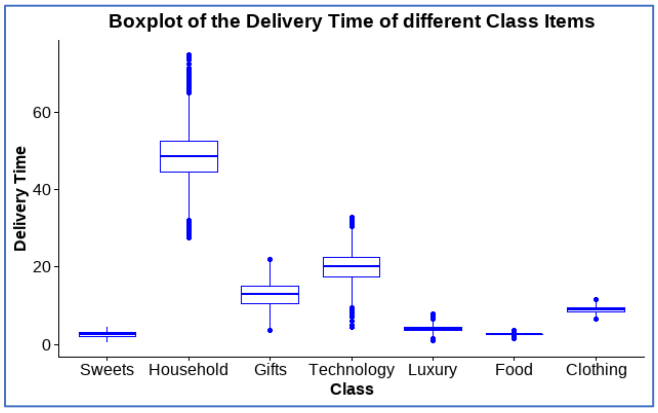


Figure 23: Boxplot of Delivery Time per Class

These boxplots clearly illustrate that the class of an item has a significant impact on both the price and the delivery time of an item. For either feature, price or delivery time, it is illustrated that the distribution of the feature is vastly different from class to class. Household items, for example have a minimum delivery time above 20 days, where as the maximum delivery time for sweets are well below 10 days.

6. Reliability of the service and products

6.1. Chapter 7 (p.359)

Problem 6

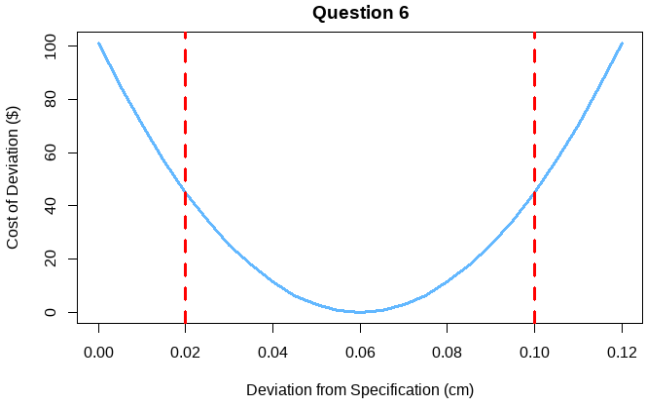


Figure 24: Problem 6 Taguchi Loss Function

Taguchi Loss Function:

$$L(x) = k(x-T)^2$$
$$\therefore k = \frac{L(x)}{(x-T)^2}$$
$$k = \frac{45}{(0.04)^2}$$
$$k = 28125$$
$$\therefore L(x) = 28125(x-0.06)^2$$

Problem 7

(a)

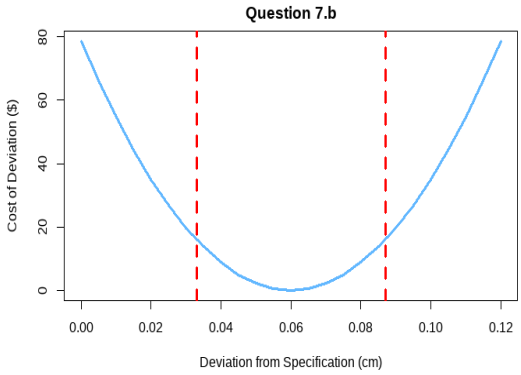


Figure 25: Problem 7.a Taguchi Loss Function

Taguchi Loss Function:

$$L(x) = k(x-T)^2$$
$$\therefore k = \frac{L(x)}{(x-T)^2}$$
$$k = \frac{35}{(0.04)^2}$$
$$k = 21875$$
$$\therefore L(x) = 21875(x-0.06)^2$$

(b)

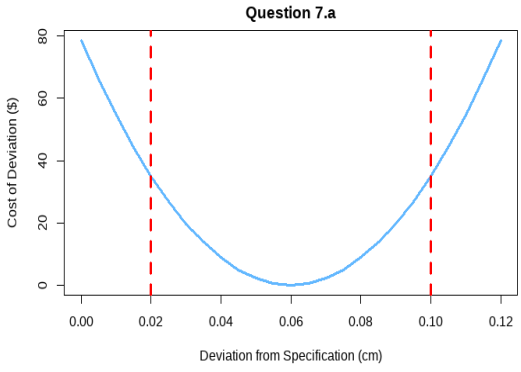


Figure 26: Problem 7.a Taguchi Loss Function

Taguchi Loss:

$$L(x) = 21875(x-0.06)^2$$
$$L(x) = 21875(0.027)^2$$
$$L(x) = 15.946875$$
$$\therefore \$\,\mathbf{15.95}$$

6.2. Problem 7: Magnaplex Production System

(a) System Reliability with one machine per station

$$Reliability = Machine\ A\ Reliability \times Machine\ B\ Reliability \times Machine\ C\ Reliability$$
$$Reliability = 0.85 \times 0.92 \times 0.90$$
$$Reliability = 0.7038$$

(b) System Reliability with two machines per station

$$Reliability = Reliability_{Station\ A} \times Reliability_{Station\ B} \times Reliability_{Station\ C}$$
$$Reliability = [1 - (1 - 0.85)^2] \times [1 - (1 - 0.92)^2] \times [1 - (1 - 0.90)^2]$$
$$Reliability = 0.96153156$$

Running two machines per station increases the overall production system's reliability by 25.77% from 70.38% to 96.15%. Running two machine's in parallel, is therefor a good decision, as it drastically improves the reliability of the production system, in comparison to a single machine per station system.

6.3. Binomial Probabilities

During last 1560 days:

20 Delivery vehicles were available, of which 19 are required for reliable operations. There are 21 drivers who each work a daily shift of 8 hours.

Availability of vehicles and drivers:

Vehicles:
20 Vehicles were available on 190 days
19 Vehicles were available on 22 days
18 Vehicles were available 3 on days
17 Vehicles were available 1 on day

Drivers:
20 Drivers were available on 95 days
19 Drivers were available on 6 days
18 Drivers were available on 1 day

Estimated number of days per year to expect reliable delivery times (with 20 delivery vehicles):

```
bi1 <- (1560-22-3-1)/1560
binom1 <- dbinom(0,10,prob = 1-bi1, log = FALSE)

bi2 <- (1560-6-1)/1560
binom2 <- dbinom(0,10, prob = 1-bi2, log = FALSE)

t1 <- binom1*binom2
days20 <- t1*365
days20
```

Figure 27: Binomial Calculations (20 vehicles)

days20 = 294.964

Therefor 295 days of reliable delivery can be expected per year.

Estimated number of days per year to expect reliable delivery times (with 21 delivery vehicles):

```
bi3 <- (1560-1)/1560
binom3 <- dbinom(0,10,prob = 1-bi3, log = FALSE)

t2 <- binom3*binom2
days21 <- t2*365
days21
```

Figure 28: Binomial Calculations (21 vehicles)

days21 = 346.7182

Therefor 347 days of reliable delivery can be expected per year.

The probability of vehicle and driver unavailability was calculated first. The binomial distribution was used to calculate the probability of neither being unavailable. The following estimates can be made based on the calculations:

The estimated number of reliable delivery days are 295 days per year, with 20 delivery vehicles. The estimated number of reliable delivery days are 347 days per year, with 21 delivery vehicles. Therefor, a single additional delivery vehicle will lead to 52 additional days of reliable delivery.

Conclusion

The report analyzed and discussed sales data of an inline business. After the data was wrangled, descriptive statistics were used to better understand the data. Valuable insights were found and trends were identified. Statistical process was performed on the valid data. This process helped to determine the mean of the delivery time of each product class. X-bar and S-bar control charts were used to analyze the data and identify samples that were outside of these control limits. The probability of making wrong predictions, type I and II errors, were calculated. The delivery time was optimized. MANOVA calculations were used to determine that the class of an item impacts the price it has as well as the delivery time. The reliability of services and products for similar businesses was calculated.

References

Fernando Hernandez. (2015). *Data Analysis with R – Exercises* [Online]. Available from: http://fch808.github.io/Data-Analysis-with-R-Exercises.html [Accessed 18/10/22]

Kenton, W., 2019. *Descriptive Statistics.* [Online] Available at: https://www.investopedia.com/terms/d/descriptive_statistics.asp [Accessed 18/10/22]

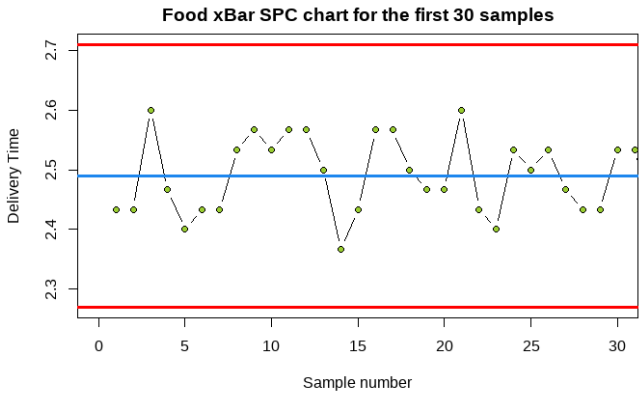
Evans, J. R. & Lindsay, W. M., 2018. *Managing for Quality and Performance Excellence.* Boston: Cengage Learning.

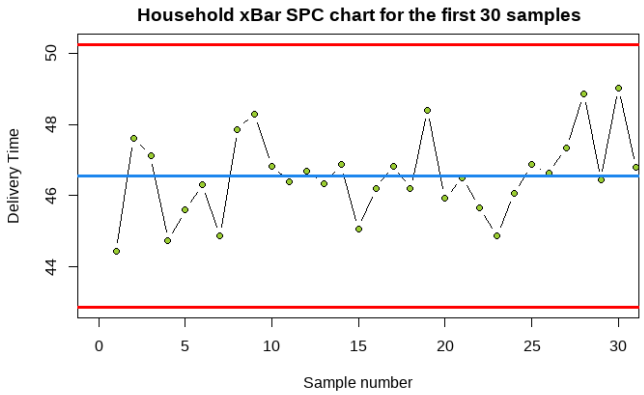
Patel, D., 2018. *What is data Wrangling.* [Online] Available at: https://www.digitalvidya.com/blog/what-is-data-wrangling/ [Accessed 18/10/22]

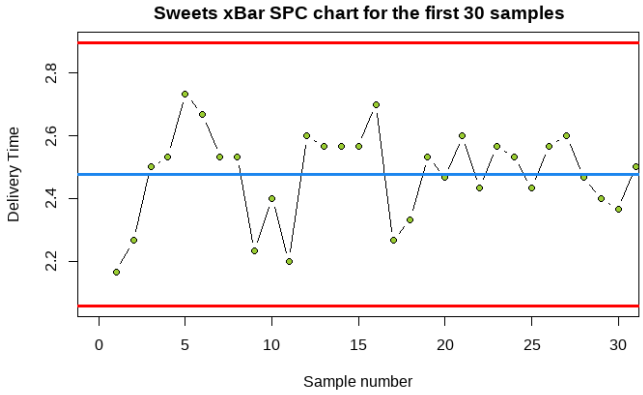
Quality System, Inc.(2022) . *X-bar and sigma chart formulas* [Online]. Available from: https://www.pqsystems.com/qualityadvisor/formulas/x_bar_sigma_f.php [Accessed 18/10/22]

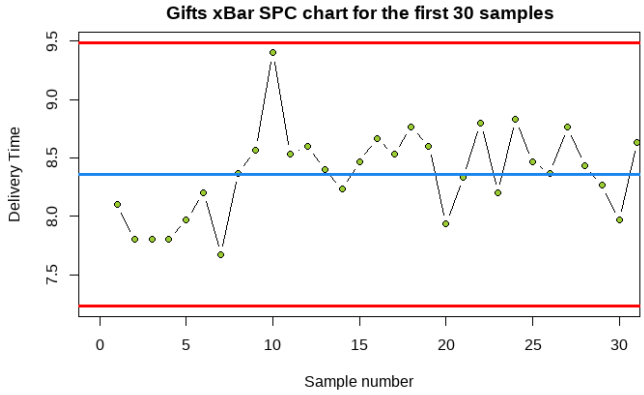
Appendices

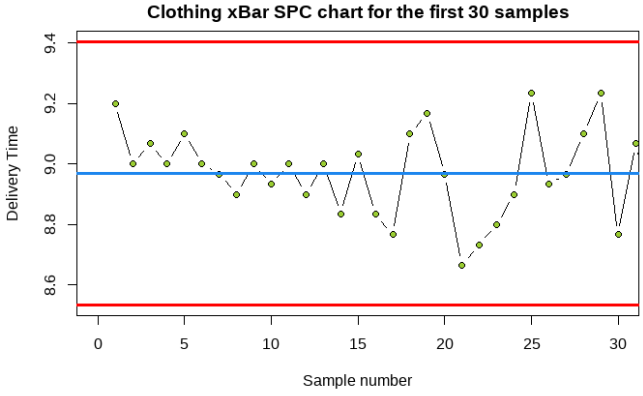
Appendix A: X-Charts of First 30 Samples

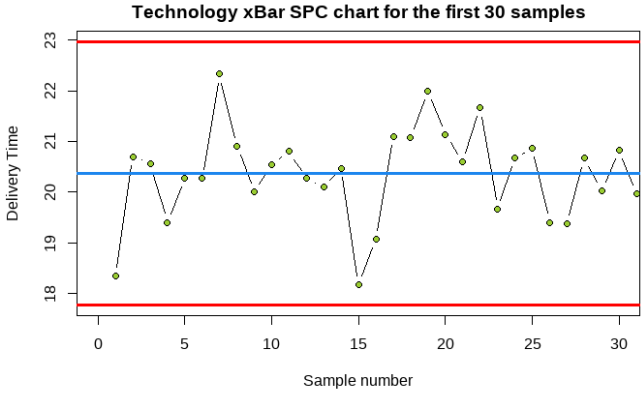




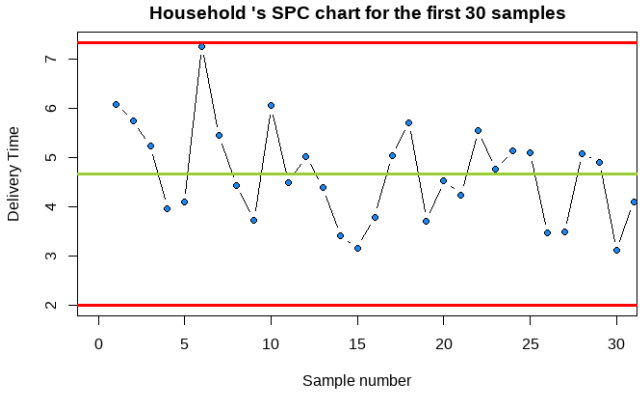


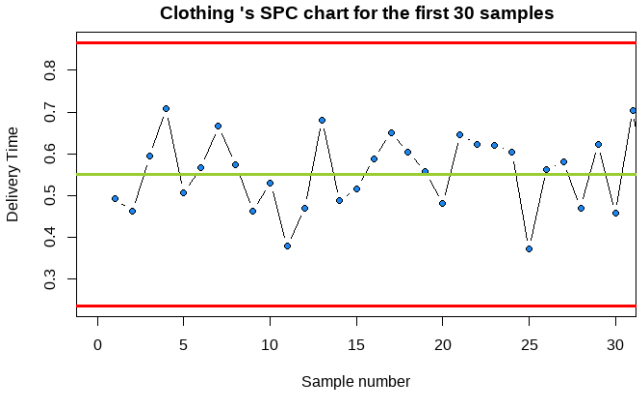


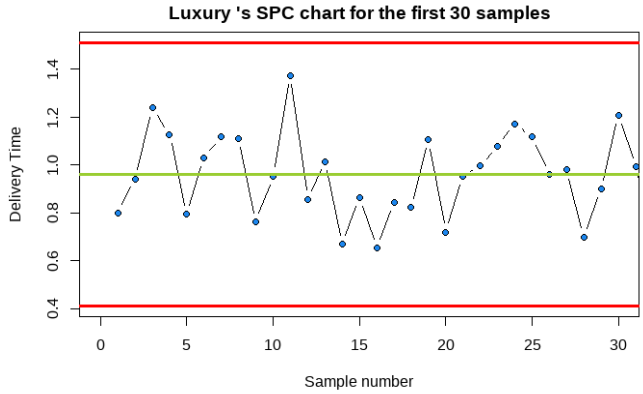


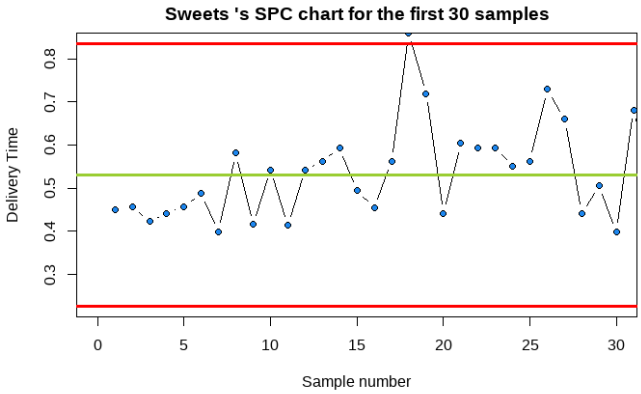


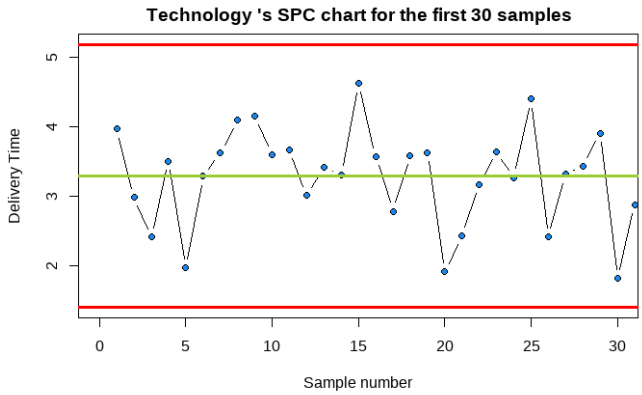
Appendix B: S-Charts of First 30 Samples

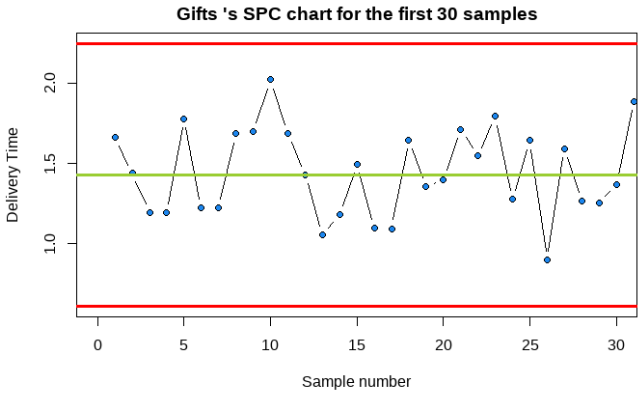




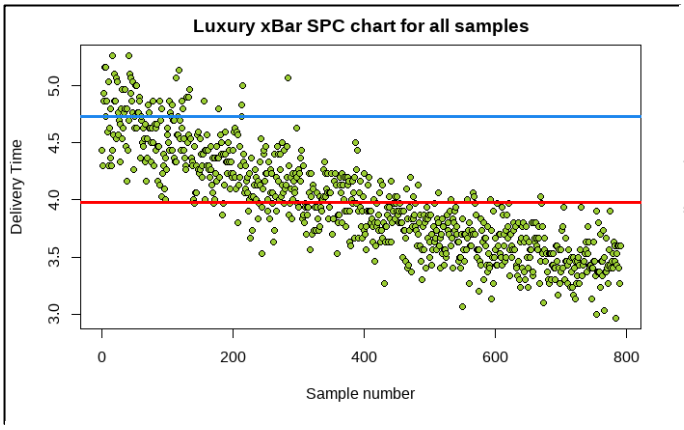


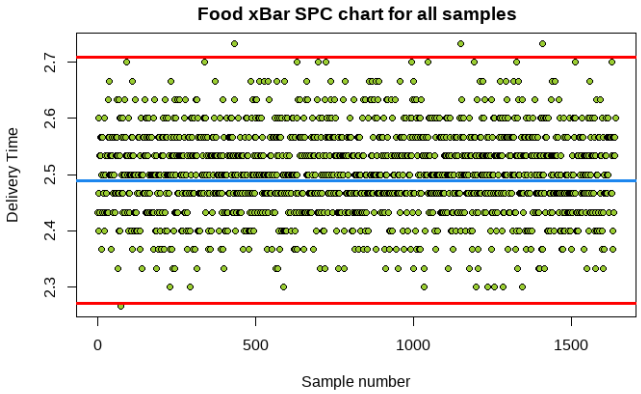


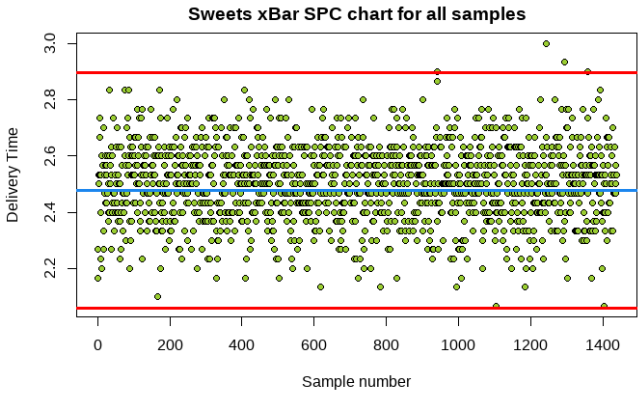


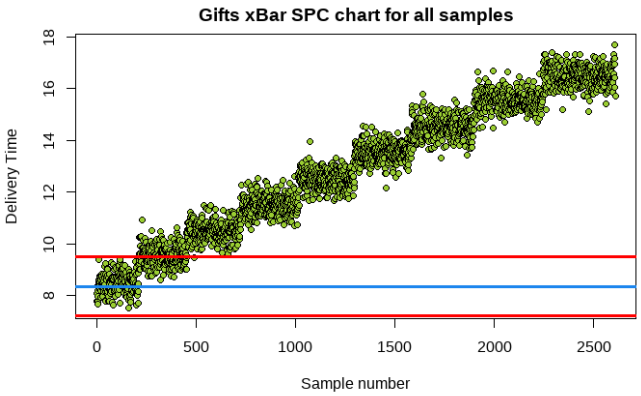


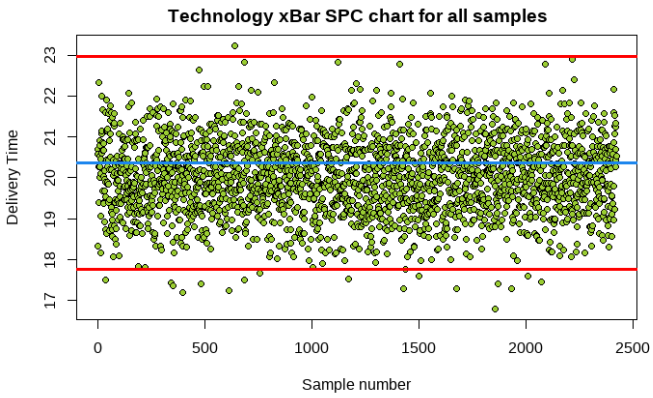
Appendix C: X-Charts All Samples



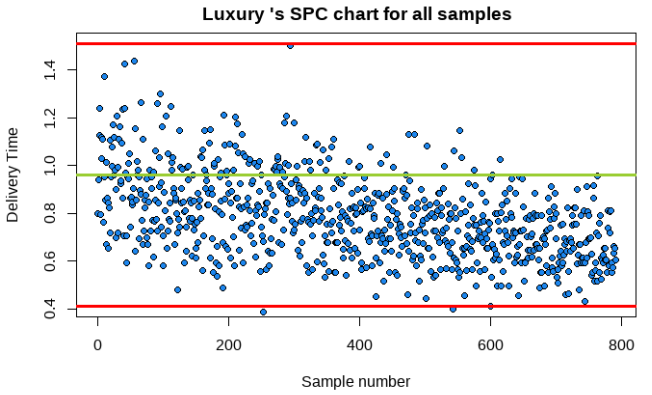


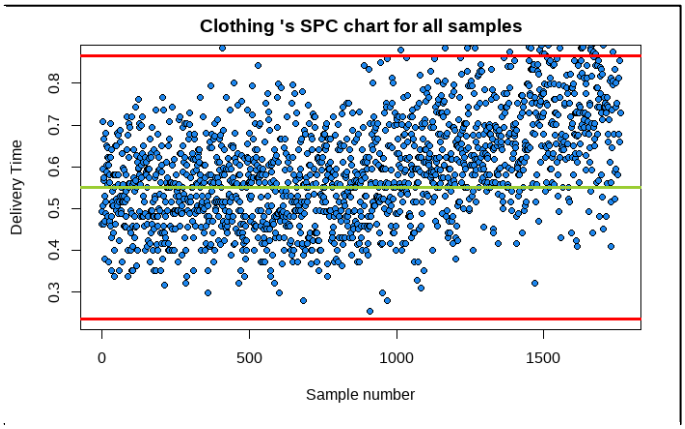


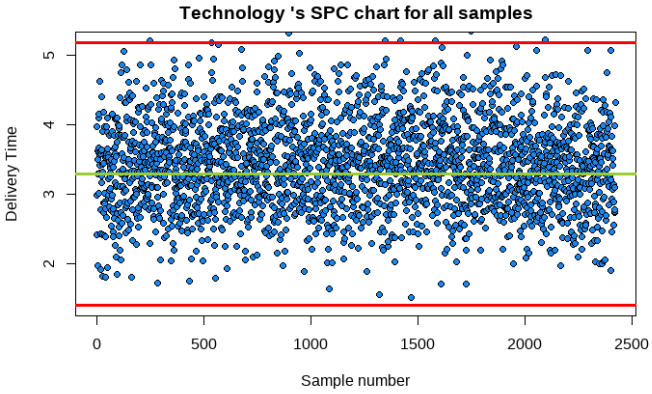


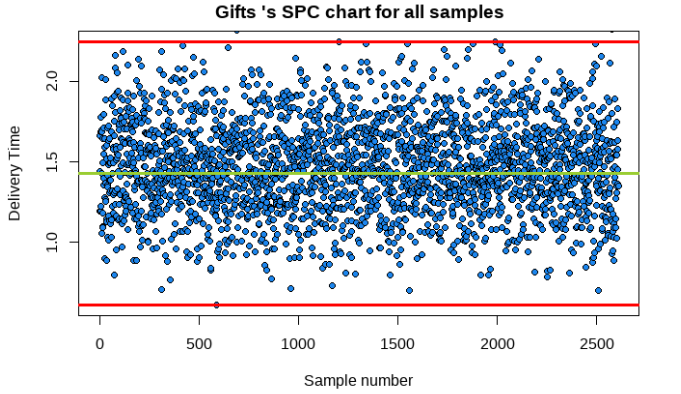


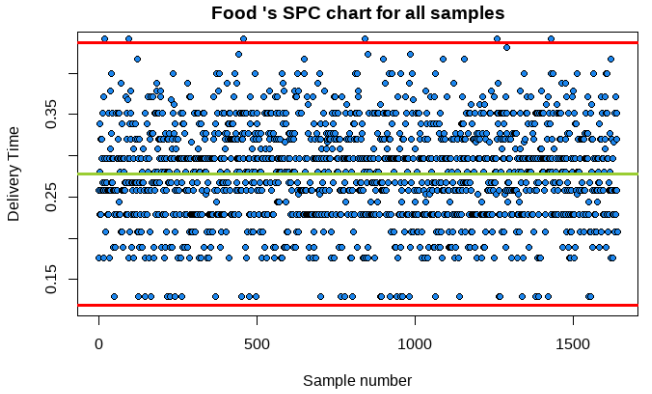
Appendix D: S-Charts All Samples

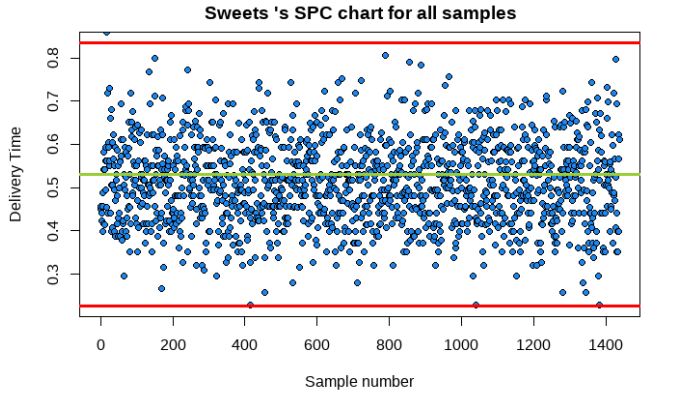












Appendix E: Samples outside the X-Chart control limits