Quality Control Report

Analysis of Sales Data for an online company

Van de Venter, Z, Mej
[21705283@sun.ac.za]

10/21/22

Quality Assurance

Table of Contents

Introduction

Companies use data in most area of their business. With the rise of online businesses, procurement of data is rapidly becoming easier. However, this data is only worth what can be analysed from it. Data needs to wrangled, plotted and examined for it to be of real use. The sales data from an online company went through all processes to accurately analyse it. The aim of data analysis is to find correlations and create predictions that would be of benefit to the business/industry.

Part 1: Data Wrangling

To accurately analyse and make use of the client's data, data wrangling is needed. This means all invalid or incorrect data must be filtered out. The sales data received revealed some missing entries as well as negative values and logically there can be no negative values in this type of data. Therefore, the data was split and saved into two different parts being valid and invalid data.

	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
19	19	98694	74	Sweets	546.48	2025	5	9	2.0	Recommended
20	20	98187	54	Luxury	81620.21	2027	9	14	3.0	Recommended

Figure 1: Example of original data set

The original data had approximately 180000 entries that needed to be split into a valid data set of 179978 X 10 and invalid data set of 22 X 10. Data was also ordered by year, month, day and row-index as it would have been received in real time.

	X	ID	AGE	Class	Price	Year	Month	Day	Delivery.time	Why.Bought
106820	106836	85247	64	Clothing	837.66	2021	1	1	9.0	Website
106821	106837	99057	61	Clothing	162.26	2021	9	28	9.0	Recommended
106822	106838	11647	30	Clothing	1129.88	2021	11	7	9.0	Browsing
106823	106839	24440	50	Food	375.47	2023	10	14	2.0	Recommended
106824	106840	95943	68	Household	21837.81	2021	4	23	49.5	Recommended
106825	106841	13654	77	Technology	8376.94	2029	7	29	15.5	EMail
106826	106842	46033	40	Luxury	105934.66	2022	2	30	4.0	Recommended
106827	106843	18663	89	Clothing	674.24	2028	12	17	9.0	Website
106828	106844	15125	22	Clothing	1123.55	2022	10	12	8.5	Recommended
106829	106845	34077	41	Clothing	1106.88	2021	6	12	10.0	Recommended
106830	106846	89607	75	Technology	53329.06	2027	9	16	13.0	Recommended
106831	106847	17922	41	Technology	21314.10	2027	4	30	25.5	Browsing
106832	106848	80377	25	Clothing	760.96	2022	7	7	8.5	Browsing
106833	106849	43762	79	Household	8239.98	2024	6	19	48.0	Website
106834	106850	43379	84	Clothing	1034.77	2024	6	23	9.0	Recommended
106835	106851	49084	70	Gifts	4703.61	2029	10	19	18.5	Website
106836	106852	56645	54	Technology	36370.93	2023	3	17	20.5	Recommended
106837	106853	14905	38	Technology	11821.31	2025	10	14	22.0	Recommended
106838	106854	23299	68	Household	17123.61	2023	6	9	45.5	Recommended
106839	106855	72547	65	Sweets	501.97	2027	7	8	3.0	Recommended

Figure 3:Example of filtered valid data

	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	48.0	EMail
3	16321	81959	43	Technology	NA	2029	9	6	22.0	Recommended
4	19540	65689	96	Sweets	-588.8	2028	4	7	3.0	Random
5	19541	71169	42	Technology	NA	2025	1	19	20.5	Recommended
6	19998	68743	45	Household	-588.8	2024	7	16	45.5	Recommended
7	19999	67228	89	Gifts	NA	2026	2	4	15.0	Recommended
8	23456	88622	71	Food	NA	2027	4	18	2.5	Random
9	34567	18748	48	Clothing	NA	2021	4	9	8.0	Recommended
10	45678	89095	65	Sweets	NA	2029	11	6	2.0	Recommended
11	54321	62209	34	Clothing	NA	2021	3	24	9.5	Recommended
12	56789	63849	51	Gifts	NA	2024	5	3	10.5	Website
13	65432	51904	31	Gifts	NA	2027	7	24	14.5	Recommended
14	76543	79732	71	Food	NA	2028	9	24	2.5	Recommended
15	87654	40983	33	Food	NA	2024	8	27	2.0	Recommended
16	98765	64288	25	Clothing	NA	2021	1	24	8.5	Browsing
17	144443	37737	81	Food	-588.8	2022	12	10	2.5	Recommended
18	144444	70761	70	Food	NA	2027	9	28	2.5	Recommended
19	155554	36599	29	Luxury	-588.8	2026	4	14	3.5	Recommended
20	155555	33583	56	Gifts	NA	2022	12	8	10.0	Recommended

Figure 2:Example of filtered invalid dataset

Part 2: Descriptive Statistics

The valid data set contains 179978 observations(sales) and 10 descriptive features. The descriptive features reveal details about the types of individual sales.

Standard Statistics

From the valid data each feature can be analysed using R summary functions. This produces the values in Figure 4. From the summary the features of Class and Why.Bought are of the character type, they as well as X and ID do not have numerical significance but will merely be used to classify and order data.

```
       X                ID             AGE           Class               Price        
 Min.   :     1   Min.   :11126   Min.   : 18.00   Length:179978      Min.   :    35.65  
 1st Qu.: 45004   1st Qu.:32700   1st Qu.: 38.00   Class :character   1st Qu.:   482.31  
 Median : 90005   Median :55081   Median : 53.00   Mode  :character   Median :  2259.63  
 Mean   : 90003   Mean   :55235   Mean   : 54.57                      Mean   : 12294.10  
 3rd Qu.:135000   3rd Qu.:77637   3rd Qu.: 70.00                      3rd Qu.: 15270.97  
 Max.   :180000   Max.   :99992   Max.   :108.00                      Max.   :116618.97  
      Year          Month            Day         Delivery.time    Why.Bought       
 Min.   :2021   Min.   : 1.000   Min.   : 1.00   Min.   : 0.5   Length:179978     
 1st Qu.:2022   1st Qu.: 4.000   1st Qu.: 8.00   1st Qu.: 3.0   Class :character  
 Median :2025   Median : 7.000   Median :16.00   Median :10.0   Mode  :character  
 Mean   :2025   Mean   : 6.521   Mean   :15.54   Mean   :14.5                     
 3rd Qu.:2027   3rd Qu.:10.000   3rd Qu.:23.00   3rd Qu.:18.5                     
 Max.   :2029   Max.   :12.000   Max.   :30.00   Max.   :75.0                     
```

Figure 4:Significant statistical values of each feature in the valid data set

Only 3 of the features will thus produce a useful standard deviation. Table 1 shows the standard deviation of Price, Age and Delivery Time. This indicates a large variation present in the price of different items, which is to be expected of data with such a wide range.

	Standard Deviation
Price	20889,15
Age	20,39
Delivery Time	13,96

Table 1:Standard deviations of numerical features

Data Visualization

To first understand the data the frequency of the **categorical** features can be observed.

Sale count shows that in the class feature Gifts and Technology are the most bought items across the dataset whereas luxury is the least in terms of frequency.

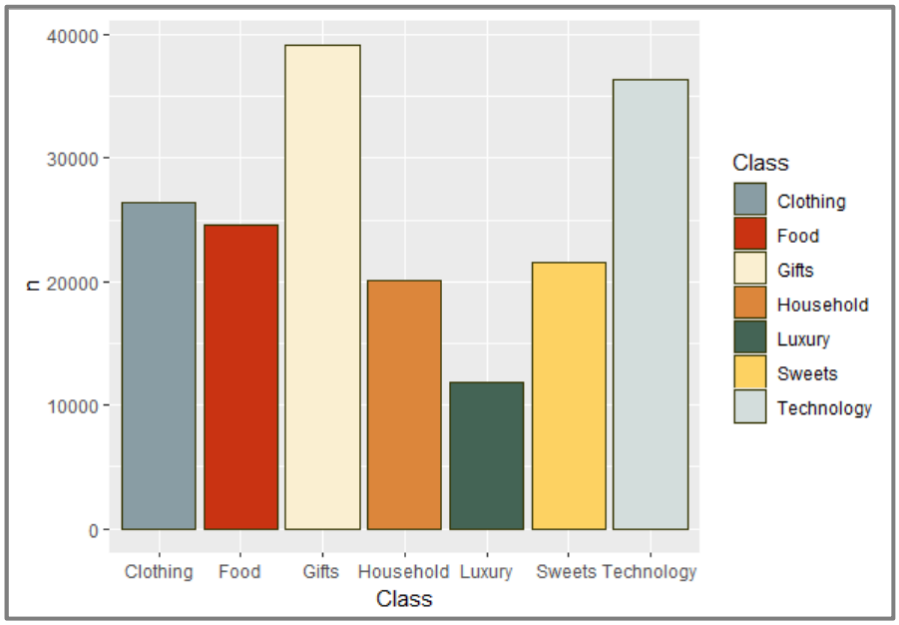


Figure 5:Frequency bar plot of each type of sales item

	Class	n
1	Clothing	26403
2	Food	24582
3	Gifts	39149
4	Household	20065
5	Luxury	11868
6	Sweets	21564
7	Technology	36347

Table 2:Frequency of each type of sales item

Contrary to figure 5 Gifts are not the greatest income generator, Technology and Luxury are the top earners. This correlates with figure 6 indicating that these classes average much higher prices than the rest. They can be seen as the priority classes since they produce majority of sales profits. Luxury can be seen as the least amount of "work" needed to generate profit since a sale in the luxury class is worth more than any other class on average.

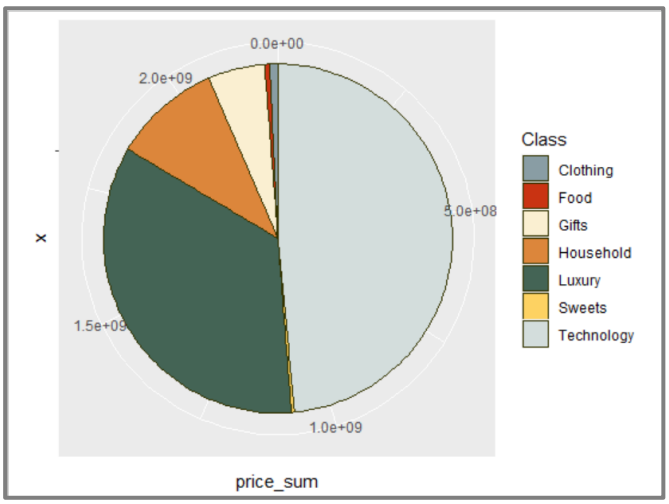


Figure 7:Division of total income generated by class

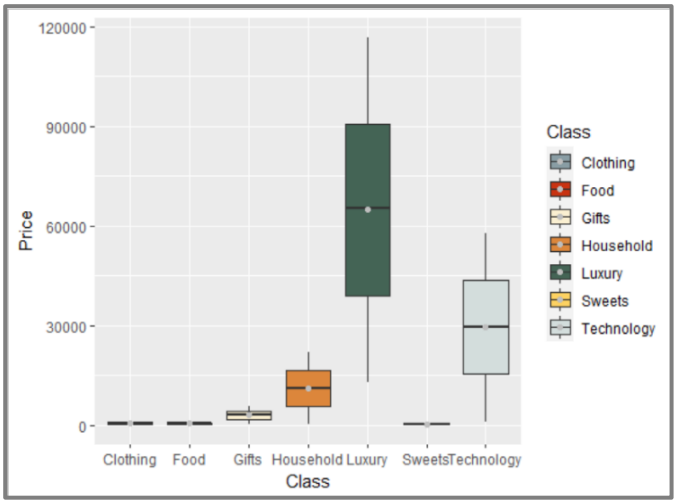


Figure 7:Boxplot of price distribution per class

The data clearly shows that in terms of frequency recommendation seems to generate the greatest number of sales while spam generates the least. Figure 9 also confirms a correlation of number of sales to income generated. This means that the method of recommendation is the most effective method used by the company to market their products.

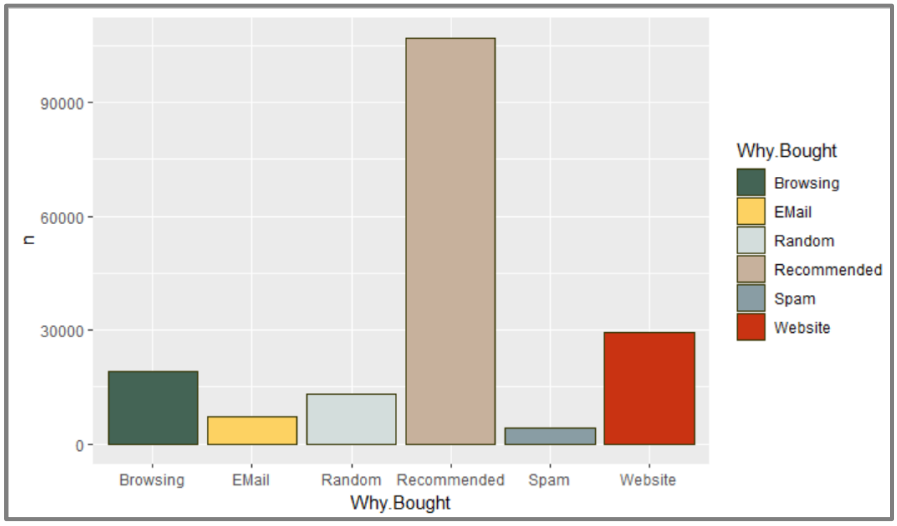


Figure 8:Frequency of reason for purchase

	Why.Bought	n
1	Browsing	18994
2	EMail	7224
3	Random	13120
4	Recommended	106985
5	Spam	4208
6	Website	29447

Table 3:Frequency of reason for purchase

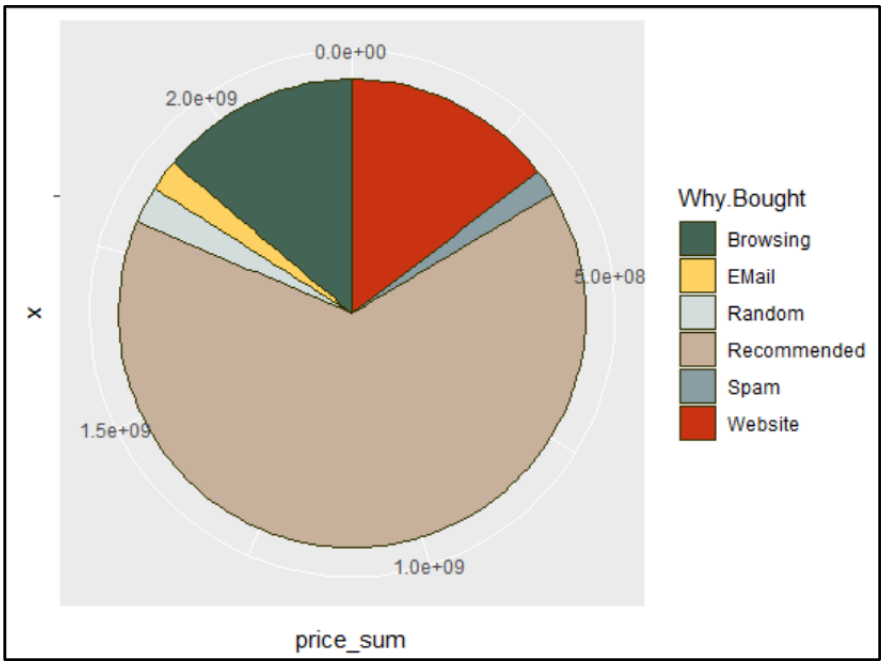


Figure 9:Division of reason for purchase of total sales income

Features like Age can be observed with Figure 9 that indicates the majority of buyers being between the ages of 30 and 60. This is however not a graph with a strong trend indicating that the company does not have a niche market in terms of age of customers. The slight decline in sale numbers from higher ages can be caused by population decreasing and internet access/skills. Figure 10 does however show a difference in Age of customers between classes. Clothing and Technology, on average, are bought by people in the range 30-60 years of age, while products like Food and Gifts are bought by a much older market ranging from 45-80 years of age. Sweets, although bought by all ages also lean towards a older market.

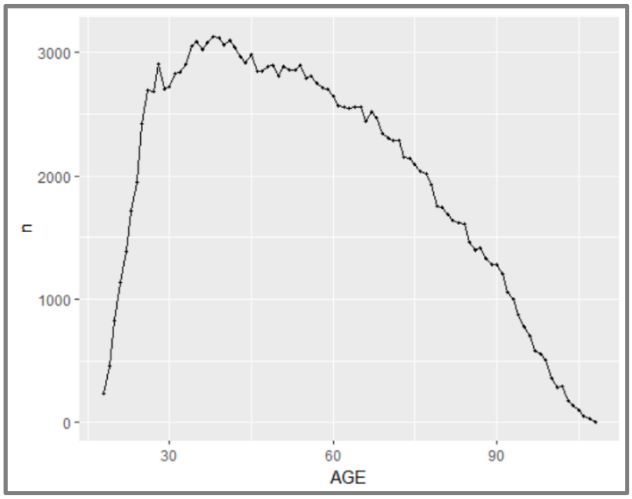


Figure 11:Age frequency distribution

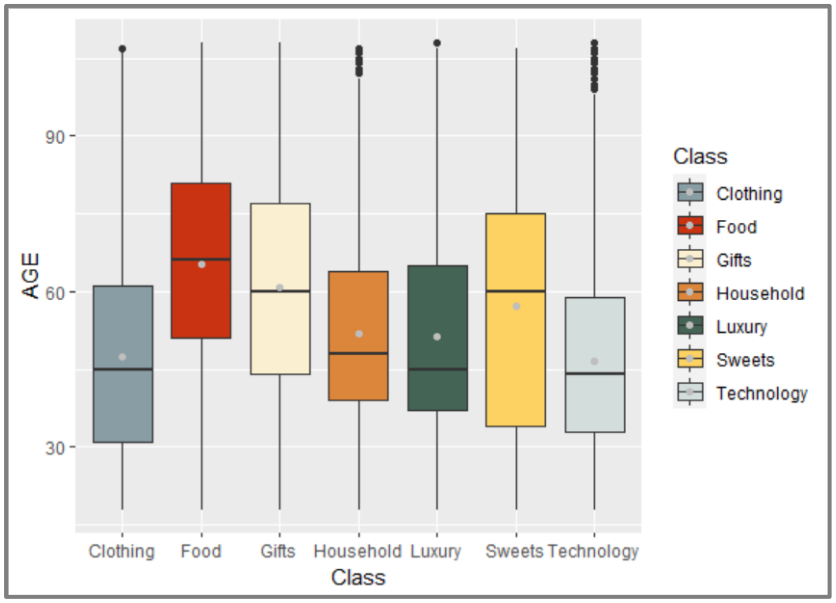


Figure 10:Box-plot of age distribution between diffrent classes

	Class	mean(AGE)
1	Clothing	47.46980
2	Food	65.37149
3	Gifts	60.82559
4	Household	51.92679
5	Luxury	51.33932
6	Sweets	57.15313
7	Technology	46.64399

Table 4:Mean age of each class

Figure 13 indicates a clear correlation between class and delivery time. Household items can generally take more than 2.3 times longer than any other class to be delivered. There can be multiple reason for this such as size of items, distribution of items or production line. Nevertheless, this seems quite high in comparison to other classes. Further investigation of the cause could be beneficial and improve customer loyalty and turnover speed.

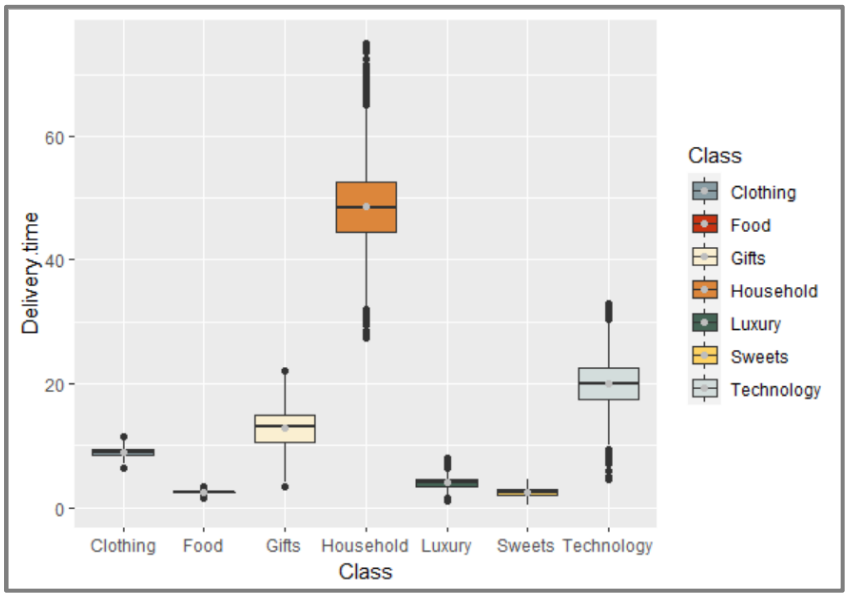


Figure 12:Boxplot of delivery time distributions per class

	Class	mean(Delivery.time)
1	Clothing	8.999527
2	Food	2.502014
3	Gifts	12.890546
4	Household	48.719561
5	Luxury	3.971520
6	Sweets	2.501206
7	Technology	20.010950

Table 5: Mean delivery time per class

A correlation matrix can be used to assess any correlation between all relevant numerical data. This will indicate whether and which classes influence each other. A correlation of 1(red) indicates a strong correlation while 0 (blue) indicate no correlation at all. Most classes have little to no correlation, however there is a small negative correlation between price and age indicating a slight decline in age as item prices rise. The small correlation of 0.13 between age and year could also indicate a small rise in age of customers as time progressed.

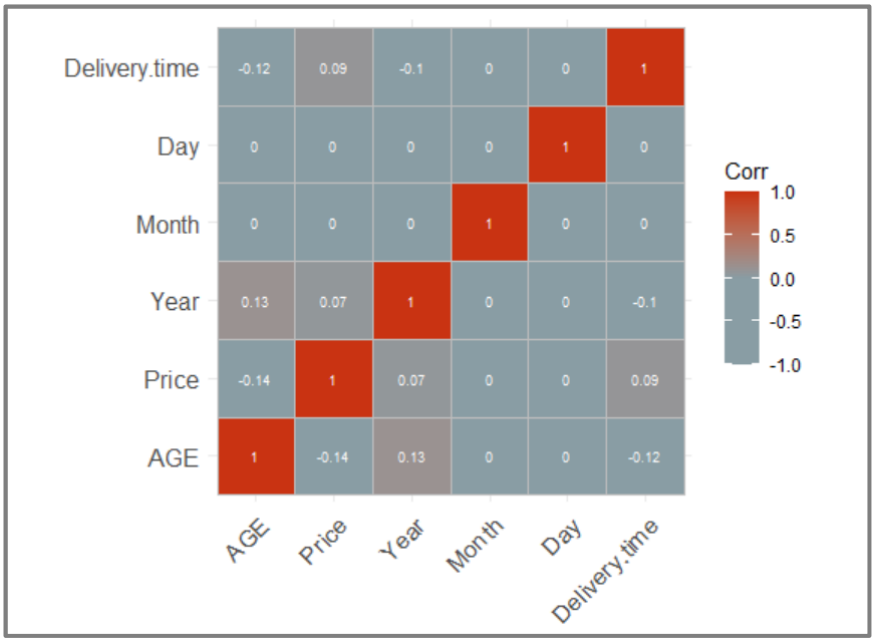


Figure 13:Correlation matrix of relevant numerical features

Process Capability Indices

Process Capability Indices can be used to evaluate how well a process is performing according to its specification limits. We use these indices with the assumption that the process is under statistical control. For these indices an upper limit of 24 hours was used and a lower limit of 0 hours. A lower limit of 0 hours is a logical starting point since time cannot be negative and no amount of time will be too little in the case of delivery time.

Indices	Values
Cpk	0,3796933
Cp	1,142207
Cpu	0,3796933
Cpl	1,90472

Table 6: Capability Indices of technology delivery time

Table 6 indicates a Cpk value of 0.3797. Although this does not mean immediate failure it does mean there is a great possibility of not meeting specifications. However, this needs to be taken into consideration with the Cp value of 1.142. This indicates that the process has potential to meet customer specifications if tight control is applied to keep it as such(Hessing, 2014).

Part 3: Statistical Process Control (SPC)

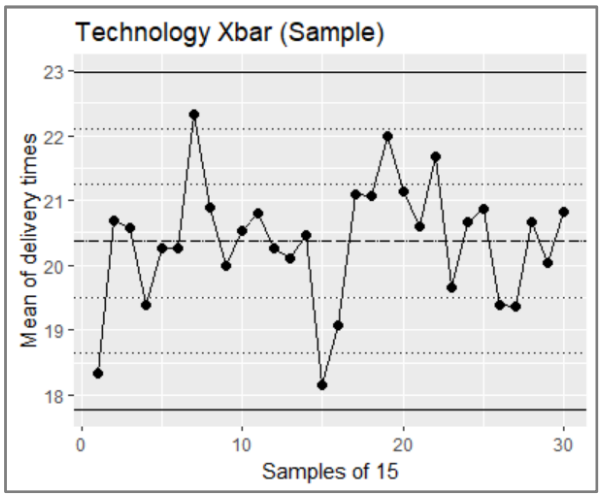
X-bar charts and S-charts are used to determine the variability and spread of data and ultimately "control" the process. For each class an X-bar chart was created with the means of 30 samples consisting of 15 sales/instances. This was also done with an S-chart by using the standard deviation of said samples. Each chart also generated limits by which the stability and control of the sample could be tested. The goal of this was to determine limits that can eventually flag out-of-control instances that need to be examined and addressed. The figures below show the charts and limits generated from the delivery times of each unique class.

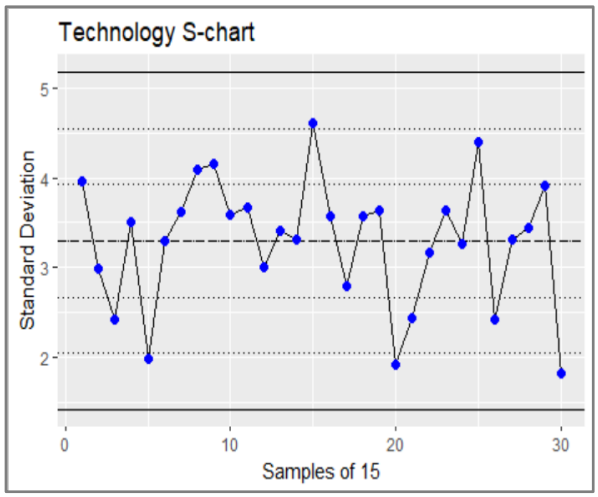
The two solid lines indicate the upper and lower limit and the dashed line indicate the mean/centre of the data. The dotted lines indicate the two inner limits on each side of the mean, used to spot patterns indicating a potentially out of control process. These patterns are identified using a set of rules/tests.

3.1 Sample Control

Technology SPC

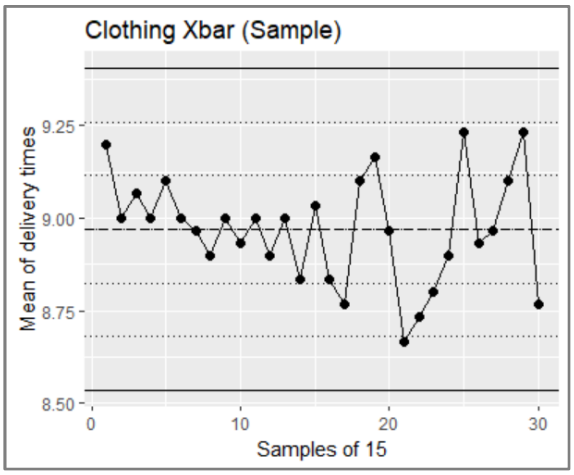
The sample plots of Technology delivery Times have no indication of being out of control and have no point outside of the limits. There are also no trends to be observed, this indicates that the set limits can be used to detect outliers in future data.

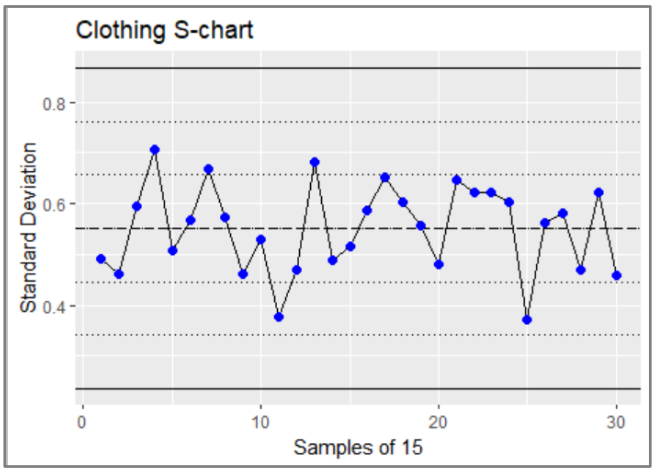




Clothing

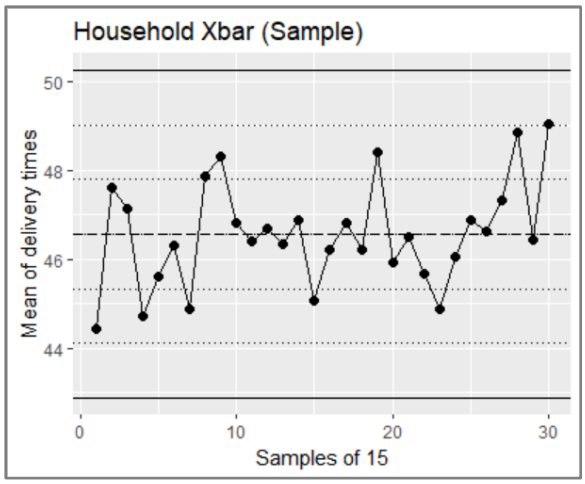
The sample charts of Clothing Delivery Times do not show any out of control points in terms of limits but there does seem to be a pattern indicating some instability. The first half of the X-bar samples have 15 consecutive points all within 1 sigma from the mean. This stratification does not necessarily indicate instability but can also be a result of the subgrouping method, in some cases it can indicate more than one process that is present e.g., different delivery methods may be in use for this class (McNeese, 2013).

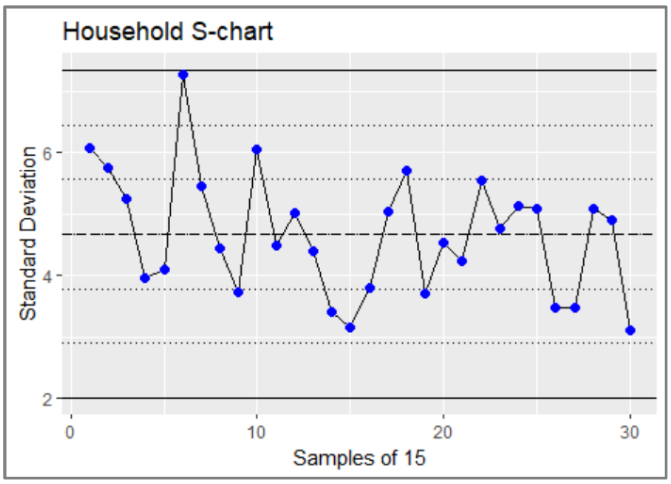




Household

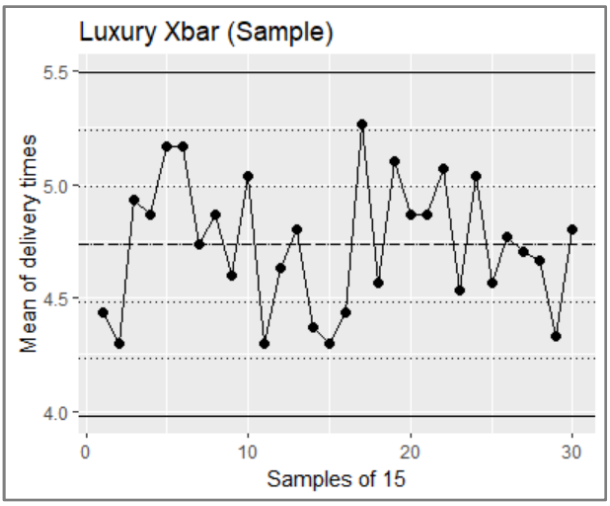
The Household Delivery Time charts indicate no specific out of control markers since all points are within limits and appear to be truly random. There is however 1 sample which has quite a high standard deviation, not high enough to recalculate limits but worth assessing to identify the cause. There is also a slight trend developing in the X Bar chart that cannot accurately be observed with this sample and subgrouping. This needs to be kept in mind if the process starts to behave strangely. It would be beneficial to update quality control if this trend persists to prevent an out-of-control process.

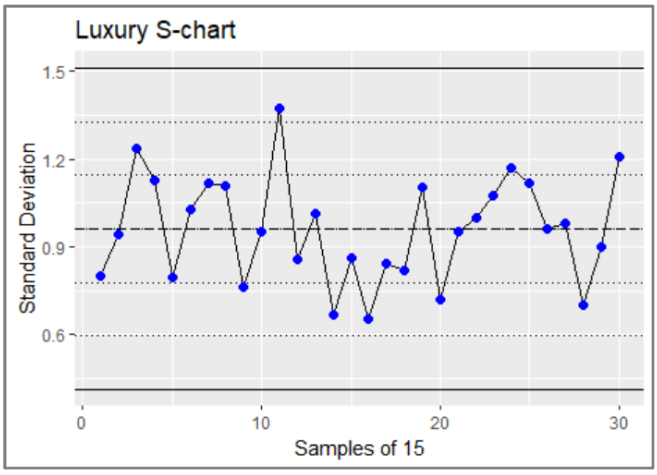




Luxury

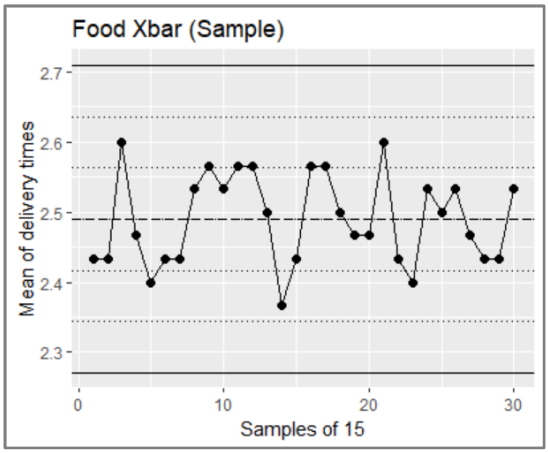
The X-Bar and S charts indicate no out-of-control patterns within the 30 Samples taken. All samples are within limits and no trends or over-control is observed. These limits can be used to monitor data in the future.

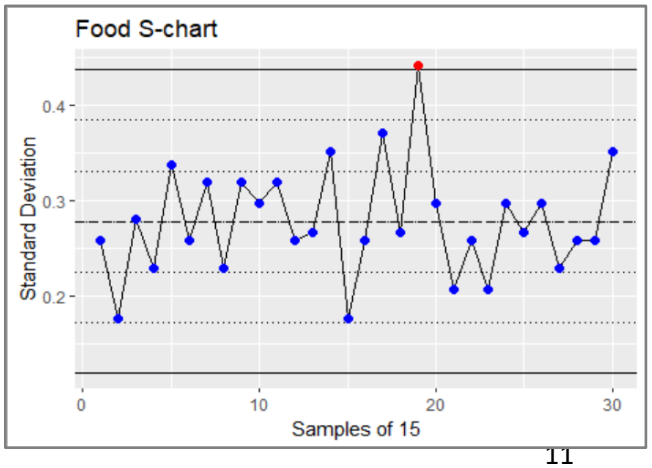




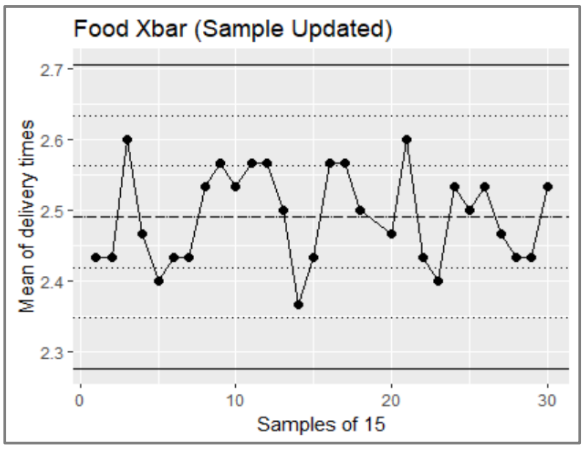
Food

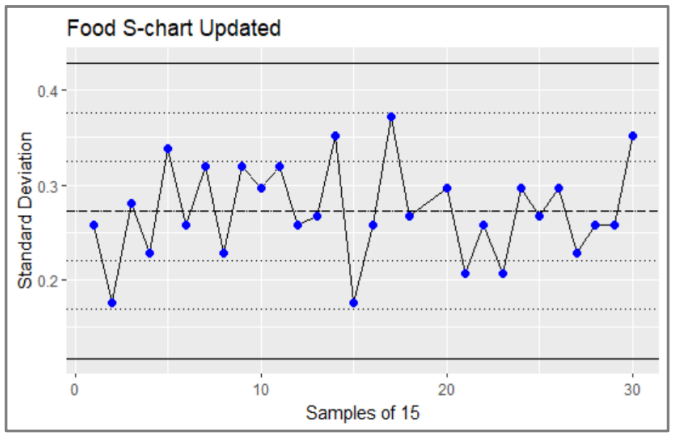
The samples of Food Delivery Times generated an S chart with an out-of-control point. This sample indicates a special cause or error that needs to be manually checked. This large shift from the mean needs to be removed to accurately create new limits and charts for Food Delivery Times.





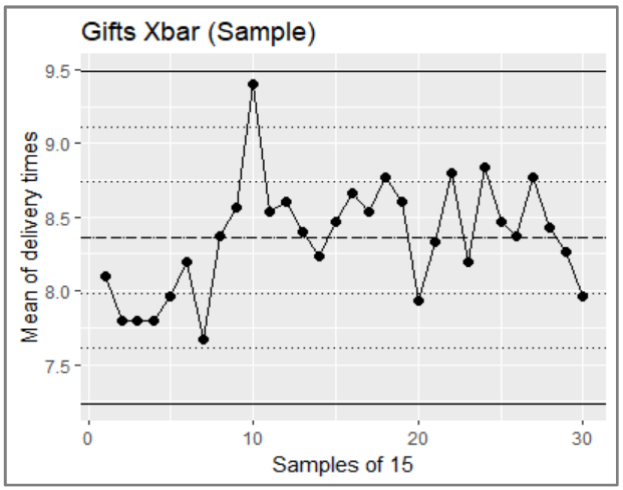
After removal the charts indicated no pattern of instability, and the new limits can be used to indicate special causes in the future. When examining the data the removed sample will be placed back as it is still an issue that needs addressing.

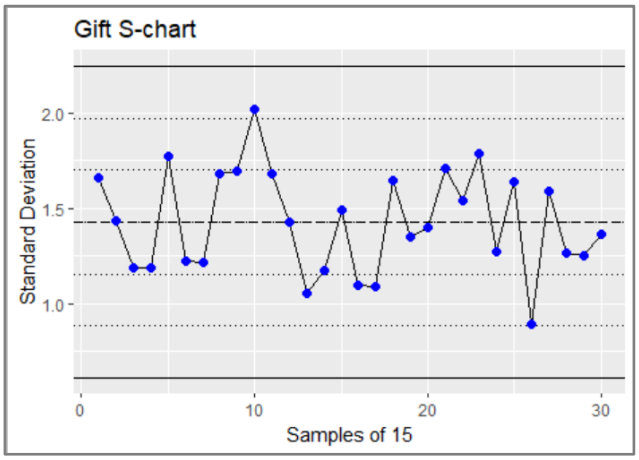




Gifts

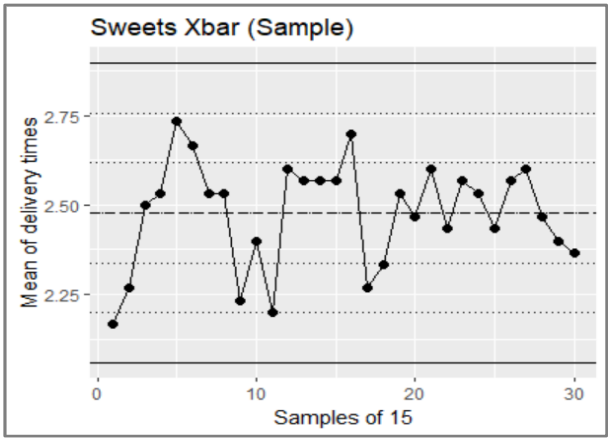
The sample charts of the Delivery times of gifts indicate no out-of-control points. There is however a small pattern of 4 consecutive points more than 1 sigma away from the mean on the X Bar chart. This small shift from the average does not raise immediate concern as it is very possible to be a change in measurement setup or a learning curve trend (Kong, 2016).

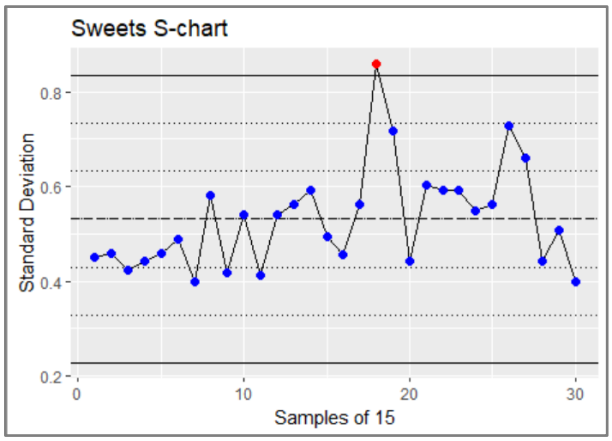




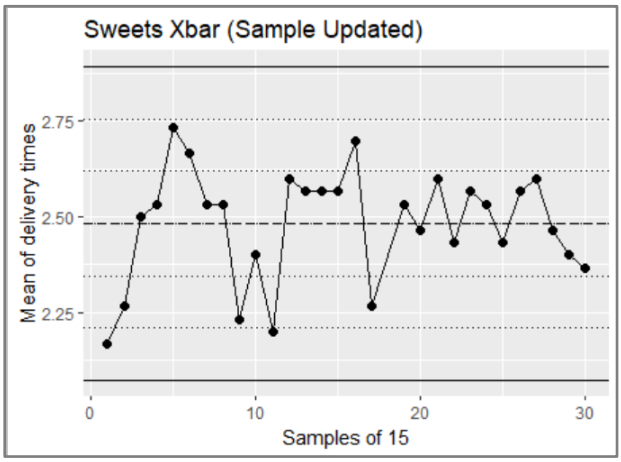
Sweets

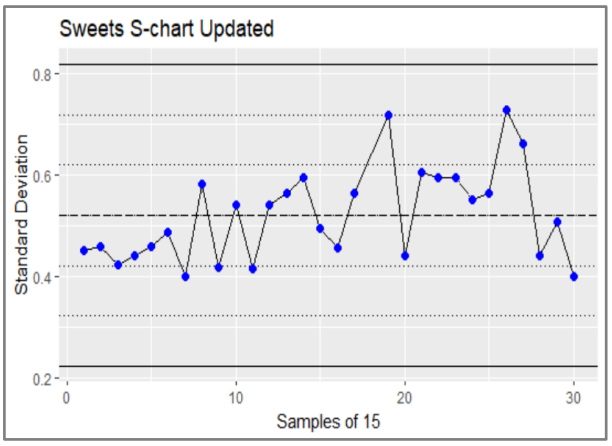
The S Chart of Sweets Delivery times generated a sample with standard deviation above the upper limit. Since no other patterns or out-of-control points were spotted the sample was removed and new charts and limits were generated. This out of bounds point indicated an error or special cause that needs further examination.





New charts indicated no pattern of instability although some instances showed a slight trend that might indicate an assessment of the quality control in that area.





Class Limits

After the relevant samples were removed and systems were controlled new and updated limits can be created to use for future systems. If any point or sample exceeds the updated limits the process should be halted, and the cause identified. Table 7 & 8 indicate the new limits of sample means and standard deviation respectively.

X-Limits

	classes	UCL	U2Sigma	U1Sigma	CL	L1Sigma	L2Sigma	LCL
1	Technology	22.974616	22.107892	21.241168	20.374444	19.507721	18.640997	17.774273
2	Clothing	9.404934	9.259956	9.114978	8.970000	8.825022	8.680044	8.535066
3	Household	50.248328	49.019626	47.790924	46.562222	45.333520	44.104818	42.876117
4	Luxury	5.493965	5.241162	4.988359	4.735556	4.482752	4.229949	3.977146
5	Food	2.705808	2.634140	2.562472	2.490805	2.419137	2.347469	2.275801
6	Gift	9.488565	9.112747	8.736929	8.361111	7.985293	7.609475	7.233658
7	Sweets	2.893106	2.756323	2.619541	2.482759	2.345976	2.209194	2.072412

Table 7:X Bar limits

S-Limits

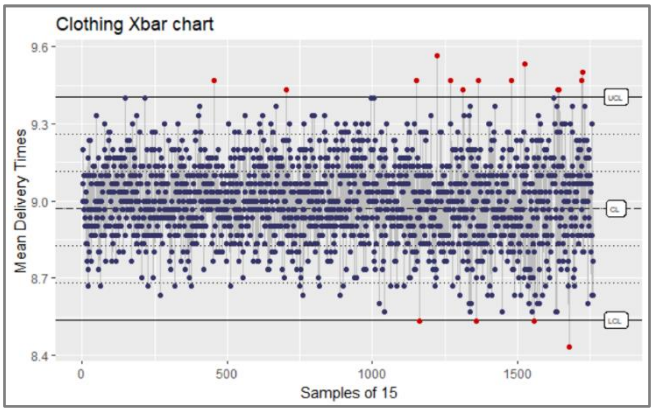
	classes_s	UCL	U2Sigma	U1Sigma	CL	L1Sigma	L2Sigma	LCL
1	Technology	5.1805697	4.5522224	3.9238751	3.2955278	2.6671805	2.0388332	1.4104859
2	Clothing	0.8665596	0.7614552	0.6563509	0.5512465	0.4461422	0.3410379	0.2359335
3	Household	7.3441801	6.4534101	5.5626402	4.6718703	3.7811003	2.8903304	1.9995605
4	Luxury	1.5110518	1.3277775	1.1445032	0.9612289	0.7779546	0.5946803	0.4114060
5	Food	0.4283723	0.3764154	0.3244584	0.2725015	0.2205445	0.1685876	0.1166306
6	Gift	2.2463333	1.9738773	1.7014213	1.4289652	1.1565092	0.8840532	0.6115971
7	Sweets	0.8175735	0.7184106	0.6192478	0.5200849	0.4209220	0.3217592	0.2225963

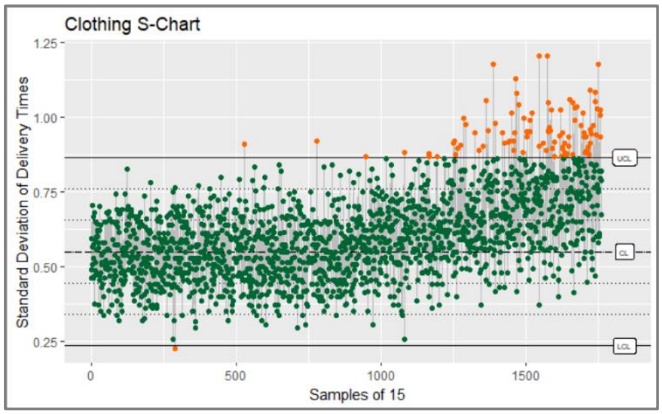
Table 8:S Chart limits

3.2 Stability of all data

With these new set limits all available data can be analysed. The figures below indicate the means and standard deviation of all data in their respective classes. Samples have been made up of 15 sales per sample. For X Bar chart, red indicates an out-of-control sample, for the S-charts orange indicates an out-of-control sample. All data should be ordered chronologically meaning in order of year, month, day, and index to accurately identify trends or instances of change.

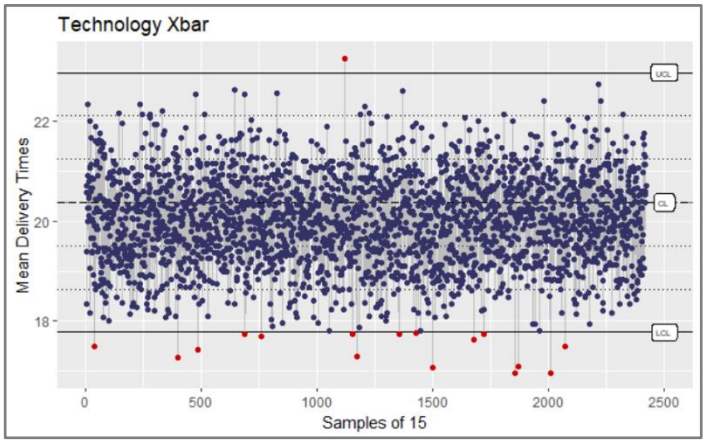
Clothing

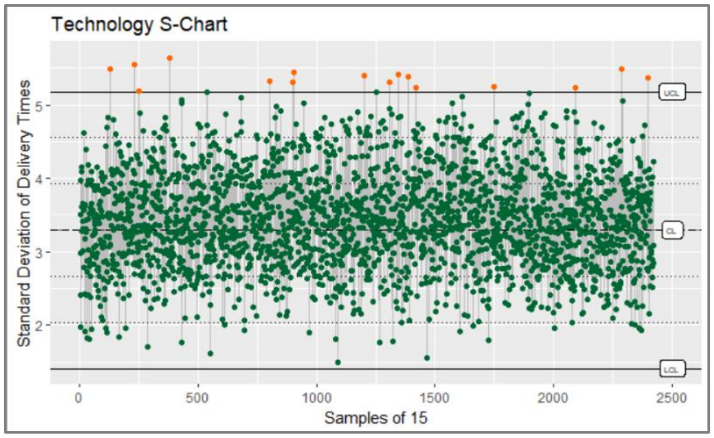




A clear trend can be seen in the overall control charts of Clothing Delivery times. With the increase of sample numbers i.e., as time is progressing the standard deviation is increasing and becoming out-of-control. The process should be stopped and reason investigated immediately as more samples will become unsatisfactory.

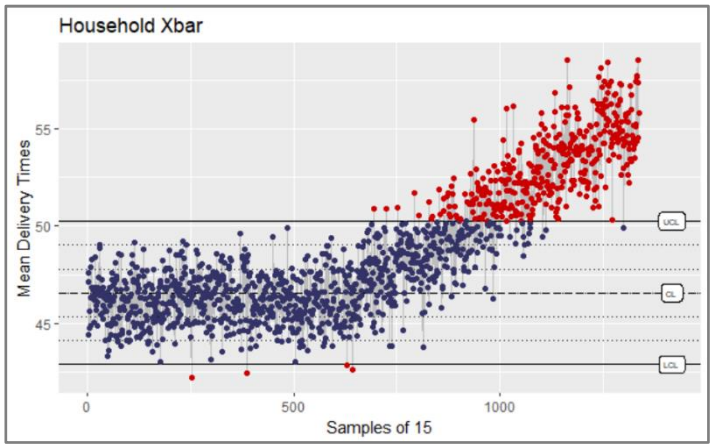
Technology

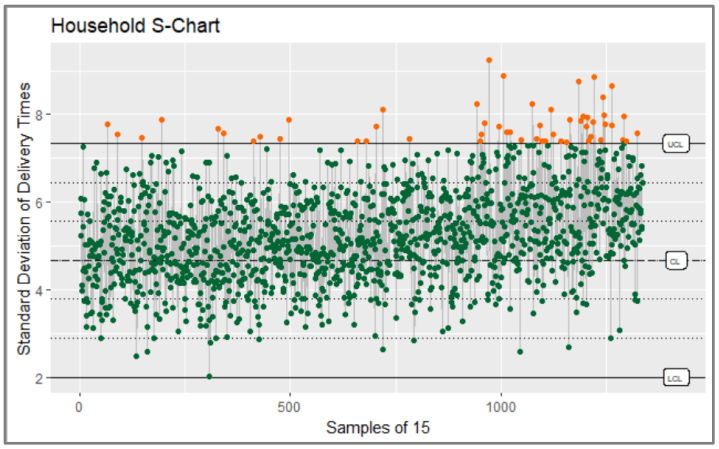




Technology Delivery Times show a reasonable number of out-of-control points with no indication of the error increasing or patterns of instability. All defective samples should still be manually examined to find and fix the cause, but there should be no concern for an increase of these errors yet.

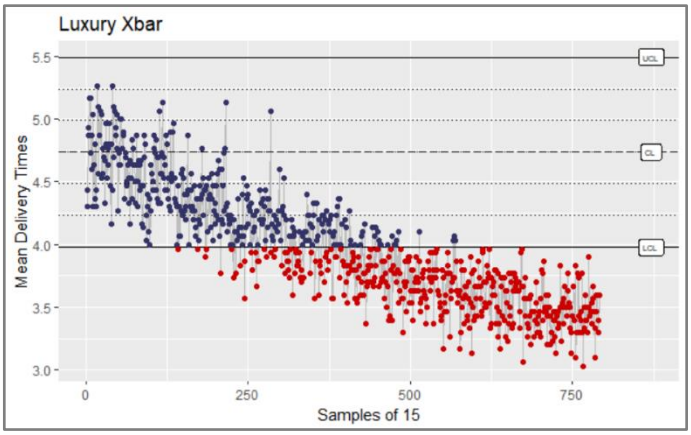
Household

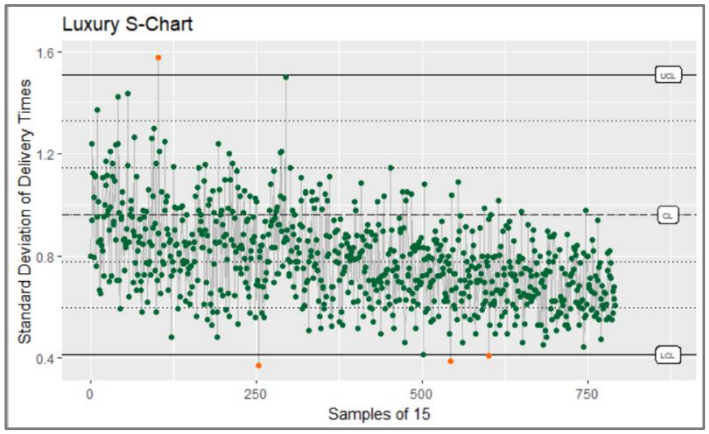




The Delivery Times of Household items indicate cause for concern over increasing delivery times. The standard deviation errors increasing also indicates a trend of errors. When observing the X Bar chart, a clear point of change can be seen where means start to rise to and above the upper limit. Investigation should start at the beginning of this trend to find cause for the change. Process should be stopped and only continued when fault is found and fixed.

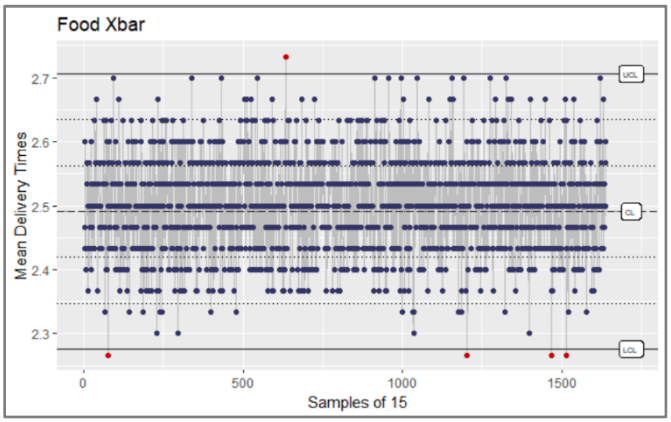
Luxury

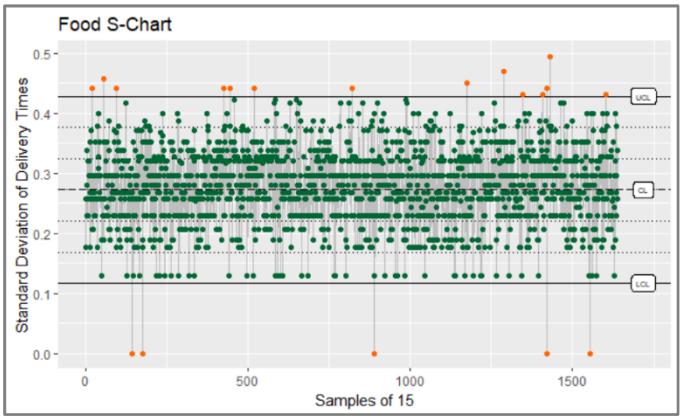




Luxury Items Delivery Time charts indicate a decreasing trend of average delivery time. Although out of bounds this indicates that the time of delivery for luxury items are decreasing as well as the variability between different deliveries. This can be because of improvements to the process or new transportation methods. It is possible that control limits need to be adjusted as a new standard can be set for these items. The expectations of the quality of service in this case can be increased.

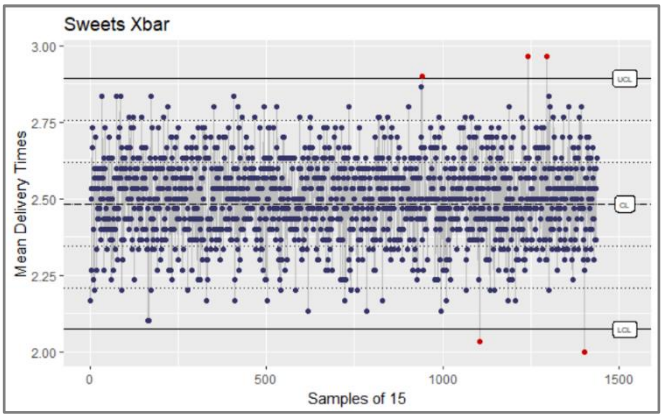
Food

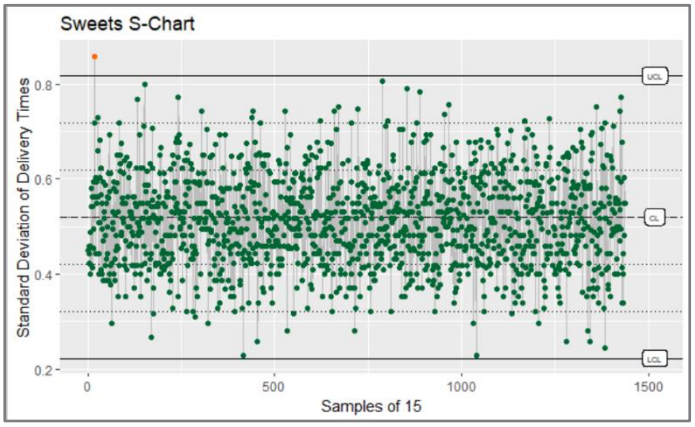




The X Bar and S Charts of Food Delivery times showcase a stable process with a reasonable amount of outliers present. In instances of this size, a few outliers are to be expected. There is no clear pattern or trend indicating future instability for the delivery times of food items. It is also clear that the delivery times are very consistently spaced, this makes prediction easier.

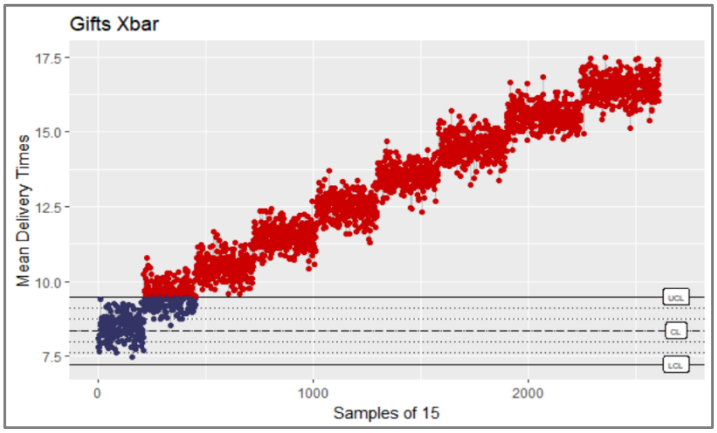
Sweets

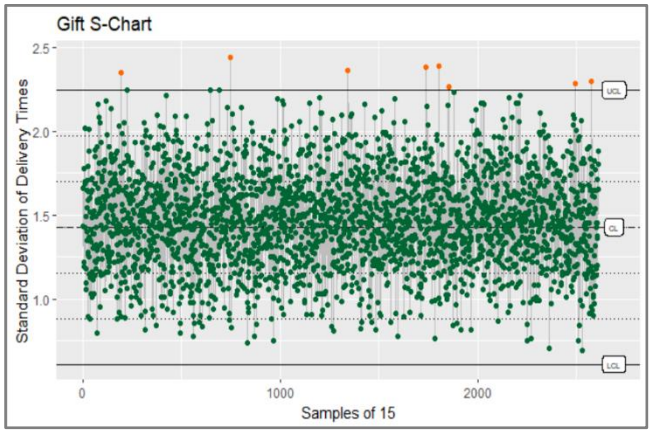




The Sweets Delivery Process is the most stable of all the classes with little outliers and no trends or patterns indicating instability. Outliers, although few should still be investigated to prevent the process from developing patterns or trends. It might be beneficial to compare delivery strategies of out-of-control processes to Sweets as it may give some insights on needed improvements or changes.

Gifts





The X Bar chart of gifts delivery times are following an increasing trend with a grouping nature. This indicates clear points of worsening delivery times periodically. This could be an indication of extreme backlog that will keep worsening as time progresses. Intervention is needed immediately to stop this from continuing. Since the samples are in clumps the standard deviation appears normal when in fact the system is greatly out-of-control. This strengthens the notion of needing multiple measures of control within the same process.

Part 4: Delivery Process Optimization

4.1 Out of Control Points

A

Using set control limits for each class the first 3 and last 3 out-of-control samples were identified as well as the total number of samples that need intervention. The most problematic would be in the category of Gifts and Household. Although Luxury items seem to have many faulty samples they are mostly beneath the lower limit. Meaning they are getting delivered faster than specified.

Class	1	2	3	3rd Last	2nd Last	Last	# Samples
Technology	37	398	483	1872	2009	2071	17
Clothing	455	702	1152	1677	1723	1724	17
Household	252	387	629	1335	1336	1337	400
Luxury	142	171	184	789	790	791	434
Food	75	633	1203	N/A	1467	1515	5
Gifts	213	216	218	2607	2608	2609	2290
Sweets	942	1104	1243	N/A	1294	1403	5

Table 10:First and last 3 out of control samples

B

To identify stratifications within the standard deviation of different classes the most consecutive samples between +0.4sigma and -0.3sigma from the mean were identified. Figure X and table X below indicates the relevant values for each class. The count indicates the maximum number of consecutive samples within above set specifications, the last sample number is the last sample identified with the range of these samples. This is still relatively low and does not indicate an out-of-control process yet.

	Count	Last Sample Number
Technology	6	372
Clothing	4	1013
Household	3	45
Luxury	4	63
Food	5	1521
Gifts	5	254
Sweets	5	316

Table11:Count of maximum consecutive samples between 0.4 and 0.3 sigma

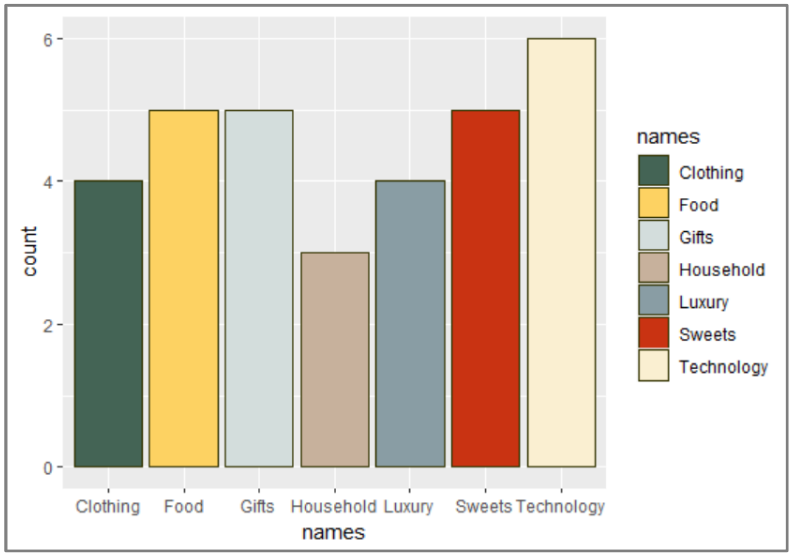


Figure 14:Bar plot of table 11

4.2 Probability of Type I error

A

As with all statistical decision making and prediction there is a risk of error in the results. To compute the likelihood of an error occurring we use hypothesis testing by means of a null hypothesis and an alternative hypothesis. In this case the null hypothesis states that the process is in control and the alternative states that it is not.

Making a Type I Error in this case would mean to predict that a process is out of control when it is in fact stable i.e., incorrectly rejecting the null hypothesis. The risk of committing such an error is valued as the significance level (alpha) (Bhandari, 2021).

Delivery time boxplots showed even distributions for each class; therefore, we can assume a normal distribution with +- 3sigma limits.

(1-pnorm(3))*2 can be used to calculate the probability of 0.27 % of a Type I error

B

The probability of a Type I error in the consecutive samples of standard distribution has to calculated separately based on the count of the specific sample.p

(pnorm(0.4) – pnorm(-0.3))^x is used where x is the maximum amount of consecutive samples in that class between -0.3 sigma and 0.4 sigma.

Column1	Count	Probability
Technology	6	0,04%
Clothing	4	0,56%
Household	3	2,04%
Luxury	4	0,56%
Food	5	0,15%
Gifts	5	0,15%
Sweets	5	0,15%

Table 12:Probability of Type I error

4.3 Cost of late delivery

The data indicates 1356 instances where technology deliveries have gone overtime. To find the best balance between losses and cost a local minimum was found that generated the least amount of loss. Logically speaking 2 decimals is the minimum amount of time that can be accurately measured and executed to a fault. Therefore, the proposed number of hours to take of would be 1,95 or 117 minutes from each Technology item's delivery time. This will bring the mean delivery time down by 1 hour and 57 minutes. This will eliminate most late delivery fees without the cost going over the losses. It is important to not that no money is saved or gained from delivering faster than 26 hours, so those instances are not relevant to the overtime calculations.

# Of hours taken off	Cost of improvement	Loss	Difference
0	0	758674	-758674
1	90867,5	370125	-279257,5
1,5	136301,2	251027	-114725,8
1,8	163561,5	201183,5	-37622
1,9	172648,2	184569	-11920,8
1,95	176261	177191,6	-930,6
1,952	177282,5	176095,6	1186,9
2	181735	167954,5	13780,5

Table 13:Balance of loss and cost of late delivery

4.4 Risk of Type II error

A shift to a mean of 23 hours from 20.3 hours will cause a larger area of the distribution to be above the upper control limit. This used to 0.05% (assuming normal distribution) but. This has however caused a calculated area of 23 %.

A type II is caused when a process is classified as in control when it is not. This will influence customers and can be seen as the type error with highest consequences.

Part 5: DOE and MANOVA

Hypothesis Test

1

Testing the hypothesis that at least 90% of delivery times will be met in all classes.

Using a MANOVA test the following was determined

Class	% of service times met
Tech	80
Clothing	73
Household	25
Luxury	97
Food	89
Gifts	15
Sweets	70

Figure 16:Percentage service reliability

Figure 15:Delivery reliability

The hypothesis was proven wrong meaning delivery times will be under 90% reliable

Part 6: Reliability of the service and products

6.1 Lafrideradora

For L(x) = $45

x-T = 0.04

$$\mathcal{L}(x) = k(x - T)^2 \qquad \text{[Taguchi Loss Function]}$$

$$45 = k\ (0.04)^2$$

$$k = 28\ 125$$

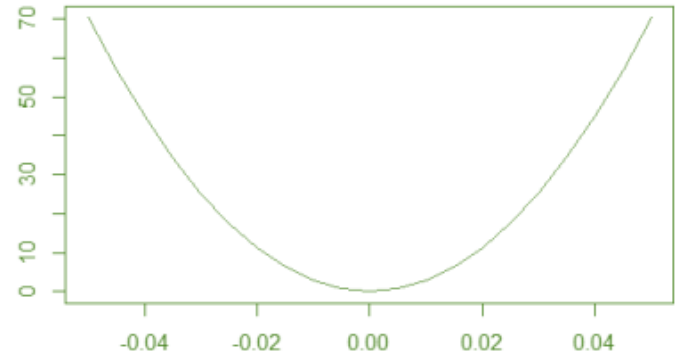


Figure 17:Taguchi Loss Function

k-> cost of scrapping a part

x-> any value from quality characteristic

T-> target value

There is direct correlation between cost and deviation, meaning if the deviation increases so will the cost.

Reducing the scrap cost to $35 per part will generate a k of 21875 and the function in the figure below.

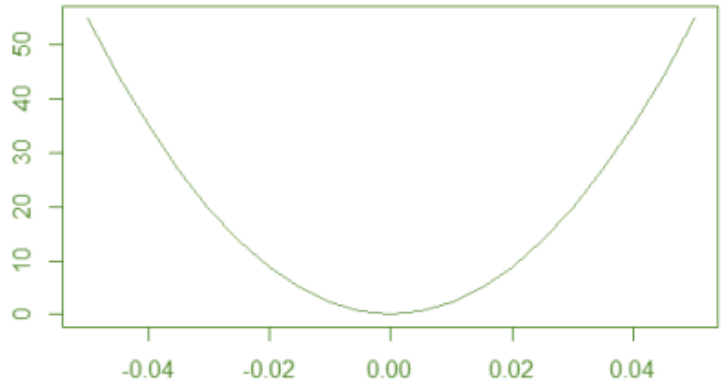


Figure 18:Taguchi Loss Function

If the deviation of the process can be reduced to 0.027cm the resulting loss is:

$$\mathcal{L}(x) = 21875(x - 0.027)^2$$

6.2 Magnaplex

To test whether any machines are wasteful, system reliability will be tested with and without backup machine.

Eliminating the backup machines causes the remaining machines to be in series.

$$total = (A\ Reliability) \text{ x } (B\ Reliability) \text{ x } (C\ Reliability) =$$

$$(0.85) * (0.92) * (0.90)$$

$$= 0.7038$$

$$=70.38\%$$

When all backup machines are included in the system, we can calculate them as a parallel system.

$$backup = (1 - (1 - 0.85) * (1 - 0.85)) * (1 - (1 - 0.92) * (1 - 0.92)) * (1 - (1 - 0.9 * (1 - 0.9))$$

$$= 0.9615$$

$$=96.15\ \%$$

Overall, it is clear that system reliability is much higher with backup machines present.

6.3 Delivery Process Reliability

Reliability based on car availability:

Taking the 1560 days as a sample, 0.26% of the time there will not be a reliable number of vehicles available for the delivery process. Meaning 99.74 % of the time there will be at least 19 cars.

Given that 19 drivers are needed to produce reliable service it can be estimated that there will be reliable service in terms of driver availability at least 99.936 % of the time.

This generates a total reliability of

$$0.9974 \text{ x } 0.99936 = 0.9968 \text{ ie } 99.68\ \% \text{ reliability.}$$

Multiplied by the total days in the year, the number of days to expect reliable service would be **363** days.

Conclusion

Data analysis revealed out of control systems in delivery times of numerous types of products. This should be addressed. The data analysis also showcased correlations between features all of which should be monitored in the future.

References

Bhandari, P. (2020). *Descriptive Statistics | Definitions, Types, Examples*. [online] Scribbr. Available at: https://www.scribbr.com/statistics/descriptive-statistics/.

Bhandari, P. (2021). *Type I & Type II Errors | Differences, Examples, Visualizations*. [online] Scribbr. Available at: https://www.scribbr.com/statistics/type-i-and-type-ii-errors/.

Hessing, T. (2014). *Process Capability (Cp & Cpk) | Six Sigma Study Guide*. [online] Six Sigma Study Guide. Available at: https://sixsigmastudyguide.com/process-capability-cp-cpk/.

Kong, J.S. (2016). *Control Chart Rules and Interpretation | BPI Consulting*. [online] Spcforexcel.com. Available at: https://www.spcforexcel.com/knowledge/control-chart-basics/control-chart-rules-interpretation.

McNeese, B. (2013). *SPC and Rational Subgrouping*. [online] BPI Consulting. Available at: https://www.spcforexcel.com/knowledge/rational-subgrouping/spc-and-rational-subgrouping.

Rian (2022). *Process Capability Index: How To Calculate CpK? | RiansClub*. [online] www.riansclub.com. Available at: https://www.riansclub.com/process-capability-index/#:~:text=The%20Process%20Capability%20Index%20%28Cpk%29%20index%20measures%20a [Accessed 23 Oct. 2022].