QUALITY ASSURANCE 344

ECSA GA4 PROJECT

Juan van Wyk
23913762@sun.ac.za

Contents

LIST OF FIGURES & TABLES

INTRODUCTION

Client data and information have been collected from a business to perform statistical analysis, to expand or monitor the performance and systems of the business. The results will be analysed and a conclusion will be made about the situation.

PART 1: DATA WRANGLING

Data wrangling is the process of removing errors and combining complex data sets to make them more accessible and easier to analyse. (What is Data Wrangling, 2022)

The data given was very well put together, except for some “NA’s” and “negative values”. To get the most accurate and true calculations and analyses from the data, only valid data should be used. By removing the invalid data from the rest, the calculations and analysis will be accurate.

All the invalid entries were located in the “Price” column. There is a total of 22 data points that had an entry of “NA” or a negative value. Seventeen entries had a “Price” value of “NA” and five entries had a negative “price value”. Note that all the negative entries had a value of “-588.8”. A screenshot of the removed data is shown below in figure 1.

	X	ID	AGE	Class	Price	Year	Month	Day	Delivery.time	Why.Bought
12345	12345	18973	93	Gifts	NA	2026	6	11	15.5	Website
16320	16320	44142	82	Household	-588.8	2023	10	2	48.0	EMail
16321	16321	81959	43	Technology	NA	2029	9	6	22.0	Recommended
19540	19540	65689	96	Sweets	-588.8	2028	4	7	3.0	Random
19541	19541	71169	42	Technology	NA	2025	1	19	20.5	Recommended
19998	19998	68743	45	Household	-588.8	2024	7	16	45.5	Recommended
19999	19999	67228	89	Gifts	NA	2026	2	4	15.0	Recommended
23456	23456	88622	71	Food	NA	2027	4	18	2.5	Random
34567	34567	18748	48	Clothing	NA	2021	4	9	8.0	Recommended
45678	45678	89095	65	Sweets	NA	2029	11	6	2.0	Recommended
54321	54321	62209	34	Clothing	NA	2021	3	24	9.5	Recommended
56789	56789	63849	51	Gifts	NA	2024	5	3	10.5	Website
65432	65432	51904	31	Gifts	NA	2027	7	24	14.5	Recommended
76543	76543	79732	71	Food	NA	2028	9	24	2.5	Recommended
87654	87654	40983	33	Food	NA	2024	8	27	2.0	Recommended
98765	98765	64288	25	Clothing	NA	2021	1	24	8.5	Browsing
144443	144443	37737	81	Food	-588.8	2022	12	10	2.5	Recommended
144444	144444	70761	70	Food	NA	2027	9	28	2.5	Recommended
155554	155554	36599	29	Luxury	-588.8	2026	4	14	3.5	Recommended
155555	155555	33583	56	Gifts	NA	2022	12	9	10.0	Recommended
166666	166666	60188	37	Technology	NA	2024	10	9	21.5	Website
177777	177777	68698	30	Food	NA	2023	8	14	2.5	Recommended

Figure 1: Removed Data

PART 2: DESCRIPTIVE STATISTICS

Descriptive statistics are brief informational coefficients that summarize a given data set, which can be either a representation of the entire population or a sample of a population. (Descriptive Statistics: Definition, Overview, Types, Example, 2022)

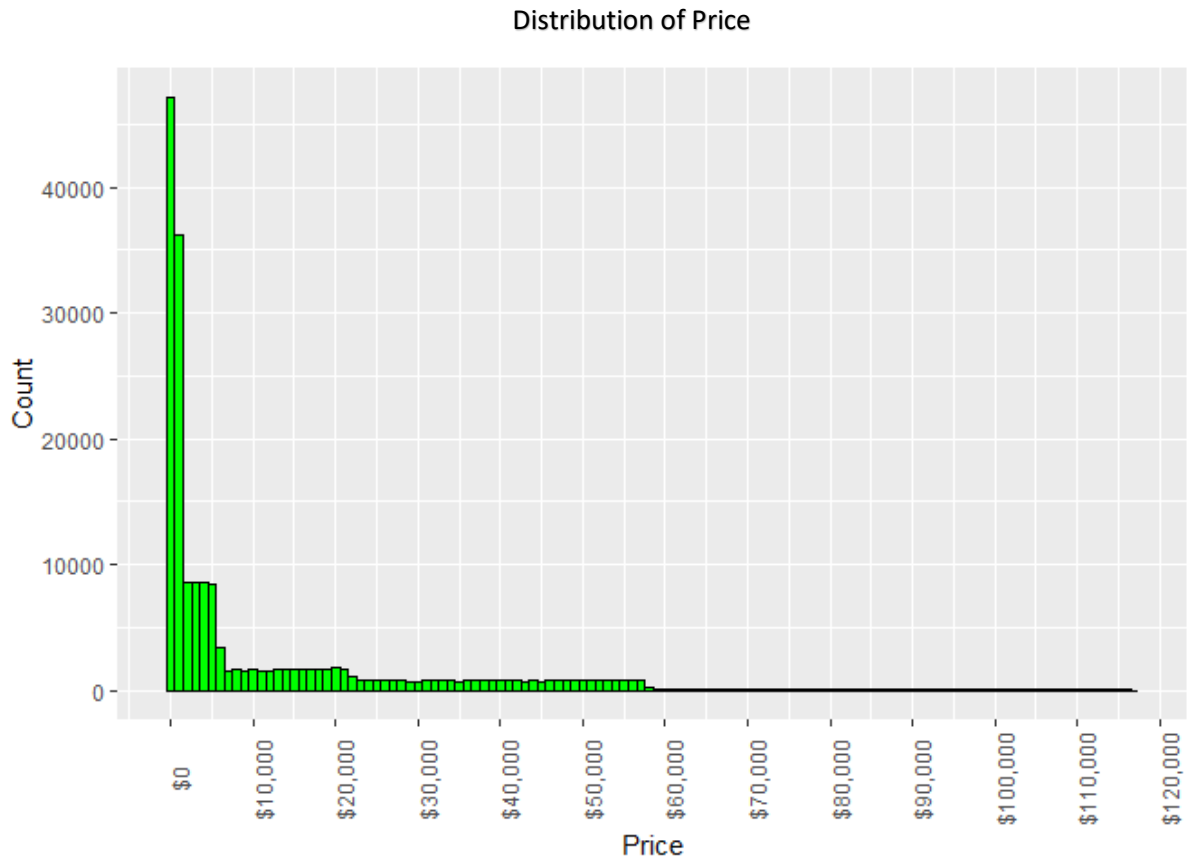


Figure 2: Distribution of Price

The prices of the various classes vary greatly, it was discovered after analyzing the reliable data. The first thing noticed when analizing the data, is the outlier in the first two columns. The majority of them are comparable, but the luxury ones have a relatively small number and has very expensive pricing. Most of the items sold was cheap items and very little expensive items were sold. That is made clear when looking at the histogram of the amount that fall into which price range.

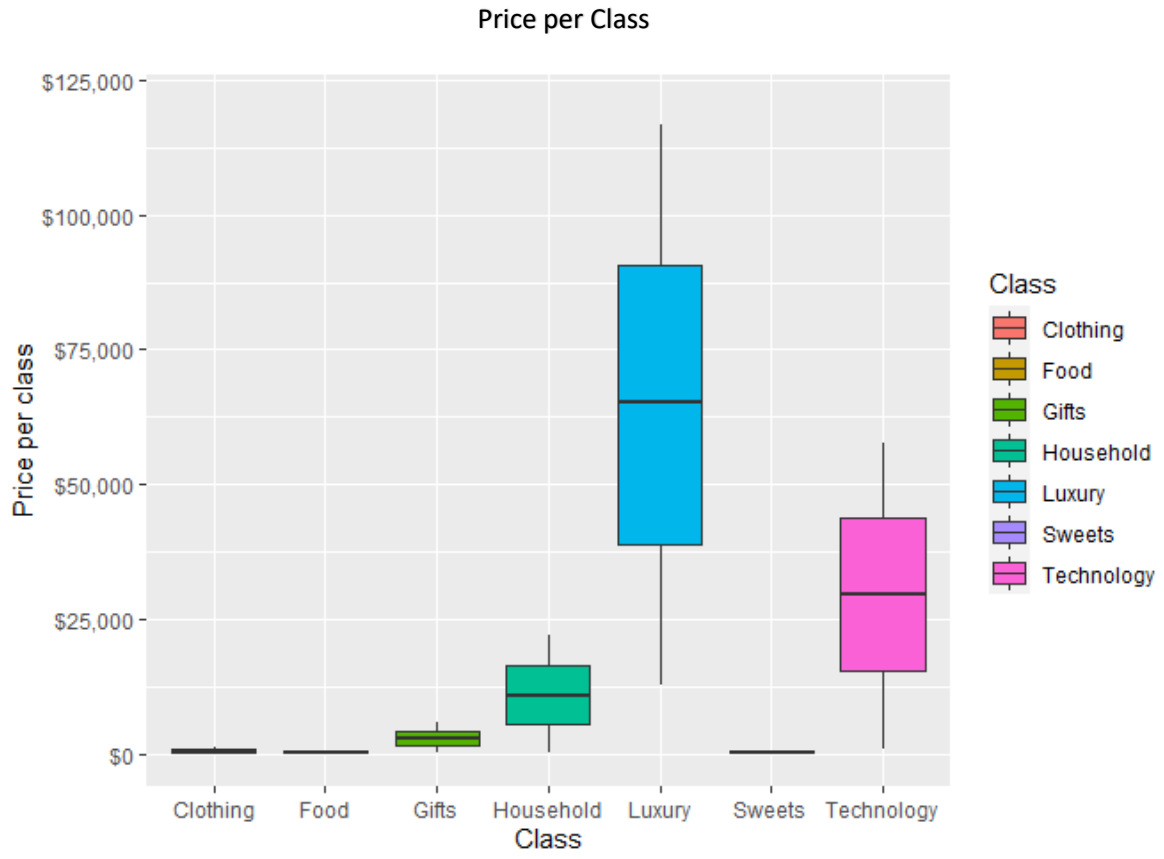


Figure 3: Price per Class

Figure 3 above is a box chart where the minimum price, first quartile, median, third quartile and maximum price of each class can be observed. The luxury class items have the highest cost, followed by technology, household, gifts and so on. Luxury items also have the biggest spread of all the classes, going from to about $115 000 to about $ 12 000, while the spread of the classes Sweets, Food and Clothing are very little in comparison to the Luxury items. The interquartile range of the Luxury items is the largest, this show that the luxury items have the most difference in price throughout the items.

The distribution of age, price and delivery time

Distribution of Age

MINIMUM	1ST QUARTILE	MEDIAN	3RD QUARTILE	MAXIMUM	MEAN
18	38	53	70	108	54.57

Table 1: Distribution of Age

The youngest person represented in the data given is 18 years old, the oldest person is 108 years and the average person is 55 years old. Given the data, it can be said that most of the people represented in the data set are between the ages 38 and 70.

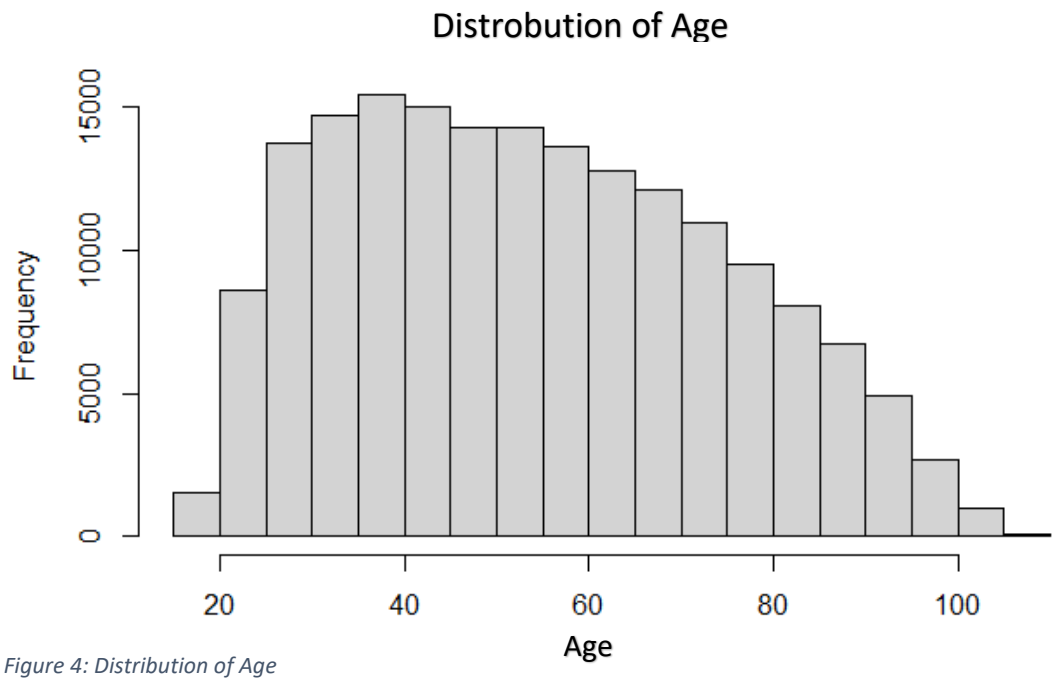


Figure 4: Distribution of Age

In the distribution curve for age, the data is skewed to the right (positively skewed), this is because the peak of the histogram is at the left hand side with the tail of the graph moving to the right. This show that more younger people is represented in the data, where the age group just below their forties is represented the most out of all the people.

Distribution of Price

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

Table 2: Distribution of Price

The minimum price represented in the data is $35.65, maximum of $116 618.97 and an average of $12 294.10. It can be observed that most of the prices fall in the interval between $482.31 and $15 270.97.

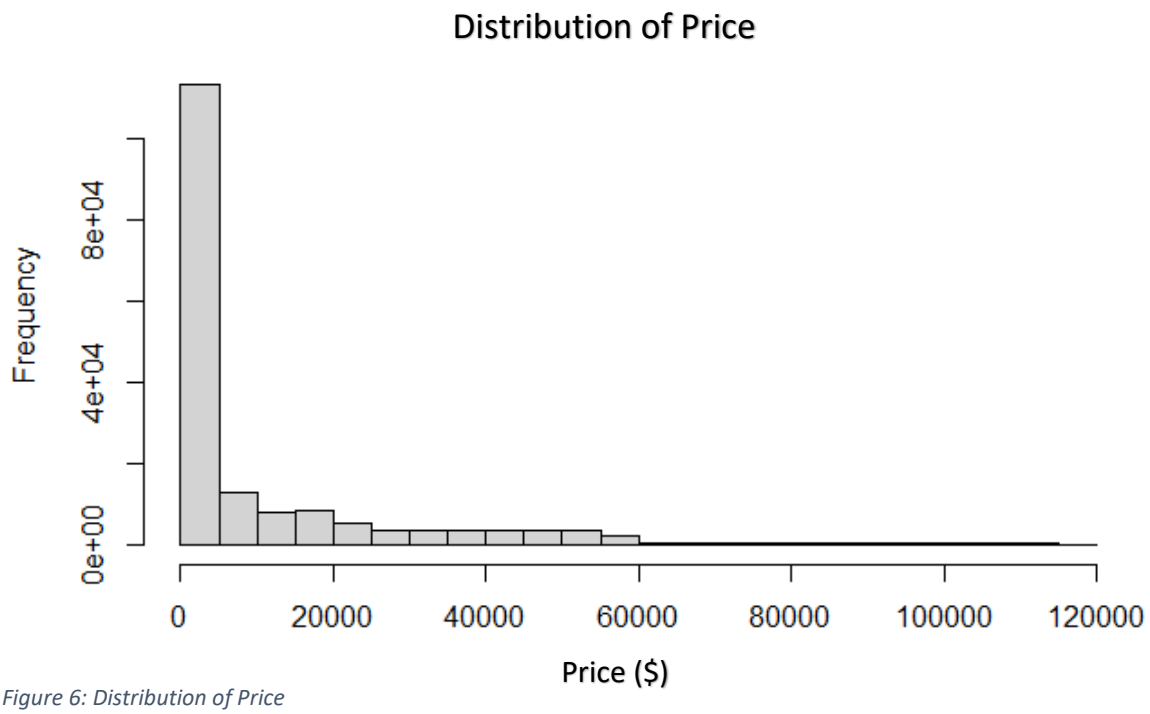


Figure 6: Distribution of Price

The first thing observed from the Distribution of Price graph, is the outlier. The luxury class has a very wide price range compared to the other classes. It can be observed that most items represented in the data, is cheap items like food, sweets and clothing relative to Luxury items, that generally cost more.

Distribution of Delivery Time (hours)

MINIMUM	1ST QUARTILE	MEDIAN	3RD QUARTILE	MAXIMUM	MEAN
0.5	3	10	18.5	75	14.5

Table 3: Distribution of Delivery Time

The minimum delivery time is 0.5 hours (30min), the maximum is 75 hours and the data represents an average delivery time of 14.5hours. The data represent most of there delivery times, in the interval between 3 hours and 18.5 hours.

Figure 7:Distribution of Delivery Time

A bimodal normal curve characterizes the delivery time distribution, because of the two peaks. This leads to the conclusion that there may be different delivery times for local and international orders, resulting in shorter delivery times for local orders and longer for international orders.

PROCESS CAPABILITY INDICES

Process capability index measures the extent of variation a process experiences relative to its specification limits. It also helps us to compare different processes with respect to the optimal situation or if they come up to our expectations. (Kiran, 2017)

Cp: Cp is an index used to assess the width of the process spread in comparison to the width of the specification. (Calculate Cp, 2022)

Cp = 1.142

Cpu: CPU is a measure of the potential capability of the process based on its upper specification limit. (Potential (within) capability for Normal Capability Analysis - Minitab, 2022)

Cpu = 0.380

Cpl: CPL is a measure of the potential capability of the process based on its lower specification limit. (Potential (within) capability for Normal Capability Analysis - Minitab, 2022)

Cpl = 0.1.905

Cpk: Cpk represents the lowest value of the capability against the upper or lower specification. (*Process capability (CP) and performance (CPK),* 2022)

Cpk = 0.380

The Cpk and Cpu values are the same. (0.380). This means the average of the specification is equal to the target value. A LSL value of zero is used, because it is impossible to deliver an item in less than zero days, this means the lower specification limit can't be less than zero.

PART 3: STATISTICAL PROCESS CONTROL

Statistical process control (SPC) is defined as the use of statistical techniques to control a process or production method. (What is Statistical Process Control? SPC Quality Tools | ASQ, 2022)

The information was rearranged according to the date of sales. The control limits were then initialized using 30 samples with 15 instances each. The X-Bar chart and the s chart produced by this procedure are shown below.

X-Chart

class	UCLX	U1SIGMAX	U2SIGMAX	CLX	L1SIGMAs1X	L2SIGMAX	LCLX
Clothing	9.404934	9.114978	9.042489	8.970000	8.825022	8.897511	8.535066
Food	2.709458	2.563153	2.526576	2.490000	2.416847	2.453424	2.270542
Luxury	5.493965	4.988359	4.861957	4.735556	4.482752	4.609154	3.977146
Technology	22.974616	21.241168	20.807806	20.374444	19.507721	19.941083	17.774273
Gifts	9.488565	8.736929	8.549020	8.361111	7.985293	8.173202	7.233658
Household	50.248328	47.790924	47.176573	46.562222	45.333520	45.947871	42.876117
Sweets	2.897042	2.617532	2.547655	2.477778	2.338023	2.407900	2.058514

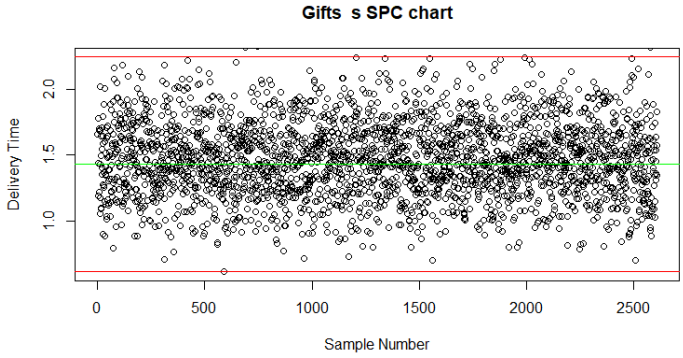
Figure 8: X-Chart

S-Chart

class	UCLS	U1SIGMAS	U2SIGMAS	CLS	L1SIGMAS1S	L2SIGMAS	LCLS
Clothing	0.8665596	0.6563509	0.6037987	0.5512465	0.4461422	0.4986944	0.2359335
Food	0.4372466	0.3311800	0.3046633	0.2781467	0.2251134	0.2516300	0.1190468
Luxury	1.5110518	1.1445032	1.0528660	0.9612289	0.7779546	0.8695917	0.4114060
Technolgy	5.1805697	3.9238751	3.6097015	3.2955278	2.6671805	2.9813542	1.4104859
Gifts	2.2463333	1.7014213	1.5651932	1.4289652	1.1565092	1.2927372	0.6115971
Household	7.3441801	5.5626402	5.1172552	4.6718703	3.7811003	4.2264853	1.9995605
Sweets	0.8353391	0.6327039	0.5820450	0.5313862	0.4300686	0.4807274	0.2274333

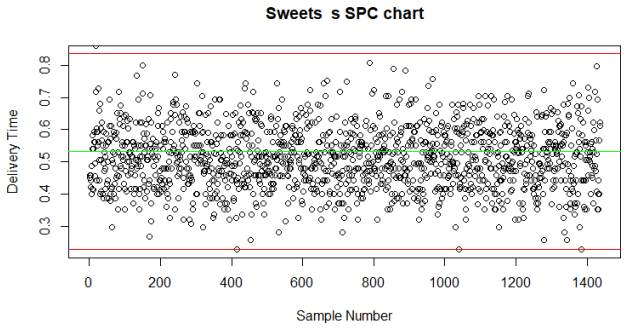
Figure 9: S-Chart

CONTROL CHART SAMPLES



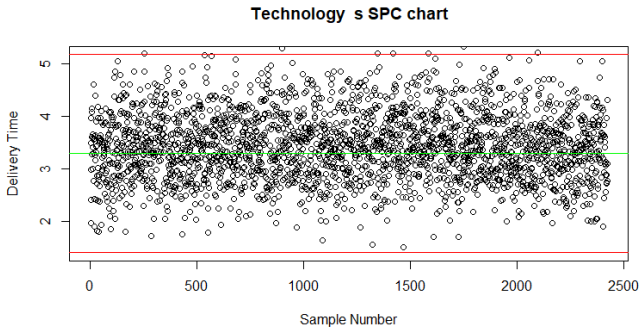
Gifts delivery time, is very wide spead. This makes sence considering you can order a lot of different types of gifts.

Figure 10: Control chart of Gifts



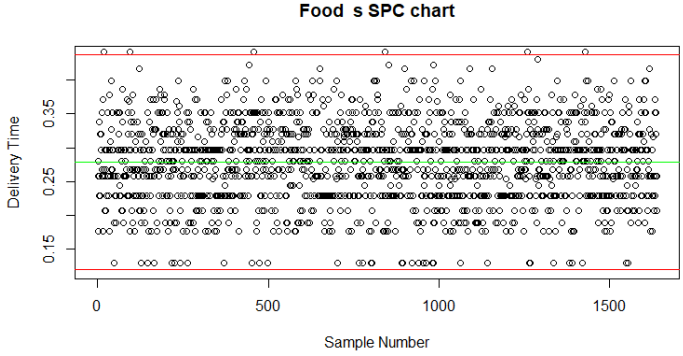
Sweets products have a delivery time bewteen 0.3 hours and 0.8 hours.

Figure 11: Control Chart of Sweets



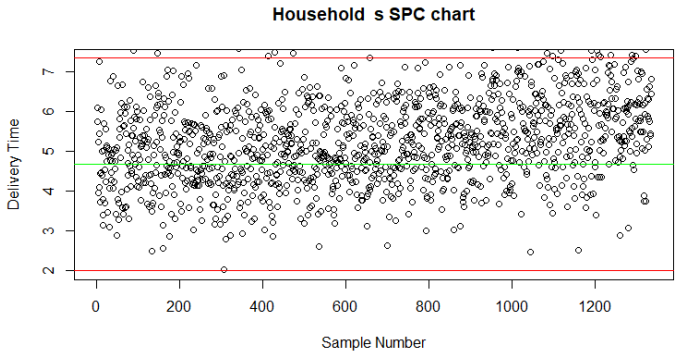
Technology delivery time, is also very wide spead. This makes sence considering you can order a lot of different types of technology products.

Figure 12: Control Chart of Technology



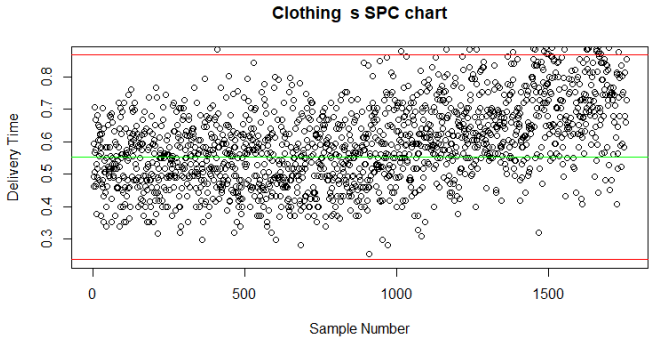
Food has a very short delivery time, ranging from about 0.15 hours to about 0.4 hours. This makes it very comfortable for the client to expect their product.

Figure 13: Control Chart of Food



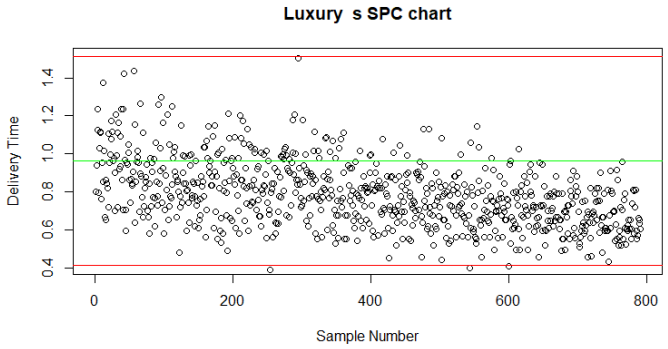
The houshold product takes a very long time to deliver. The difference between the upper and lower limit is also large, around 5 hours. This makes it difficult to predict when to expect the product.

Figure 14: Control Chart of Household



Clothing products has a delivery time between 0.3 hours and 0.8 hours.

Figure 15: Control Chart of Clothing



Luxury items are delivered between 0.4 hours, and 1.4 hours which is understandable, considering your product.

Figure 16: Control Chart of Luxury

SPC OF ALL THE DATA

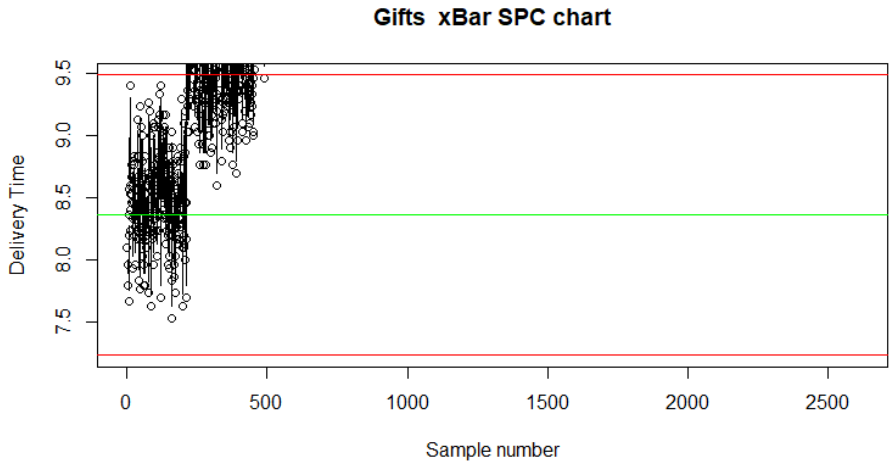


Figure 17: Control Chart of Gifts for all Entries

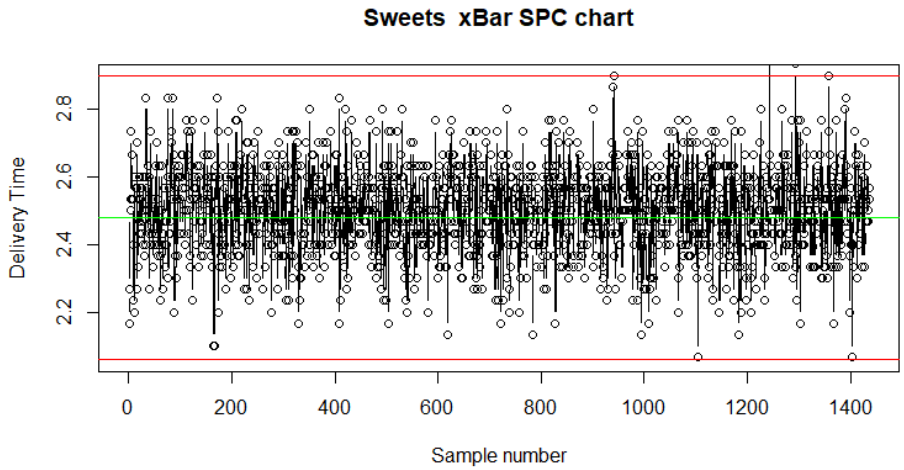


Figure 18: Control Chart of Sweets for all Entries

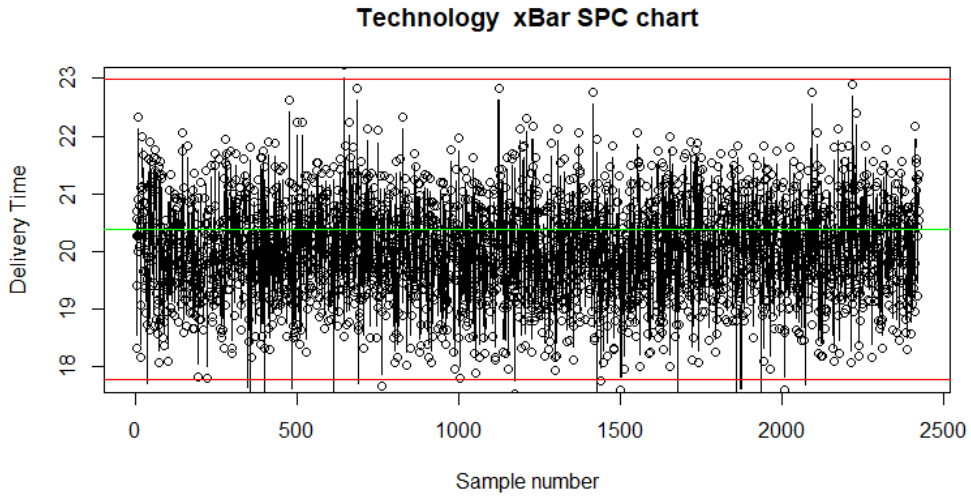


Figure 19:Control Chart of Technology for all Entries

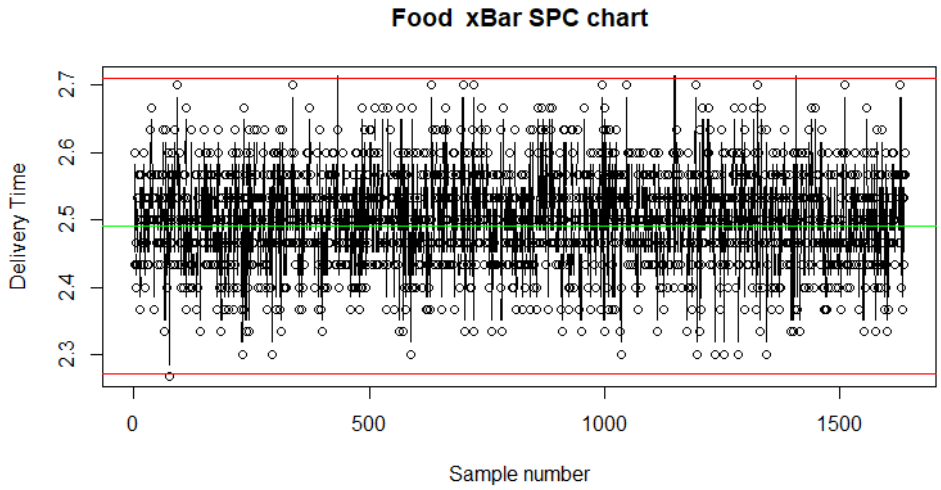


Figure 20: Control Chart of Food for all Entries

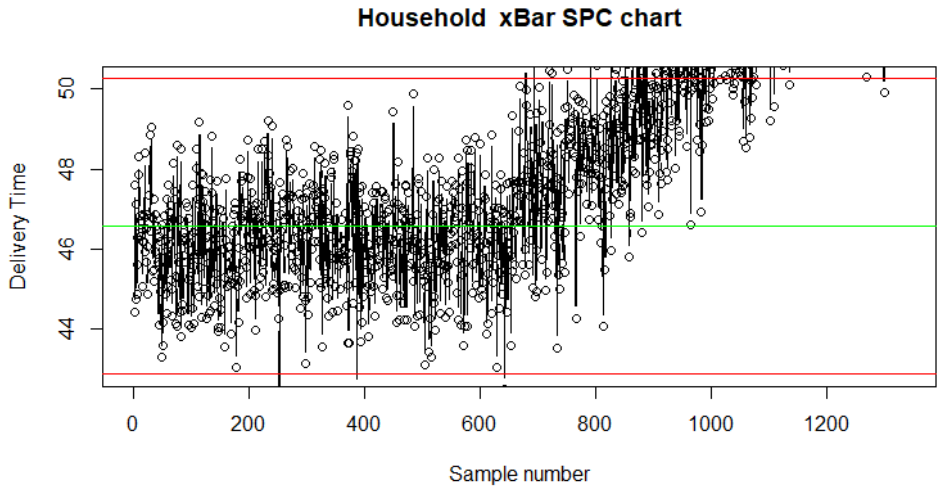


Figure 21: Control Chart of Household for all Entries

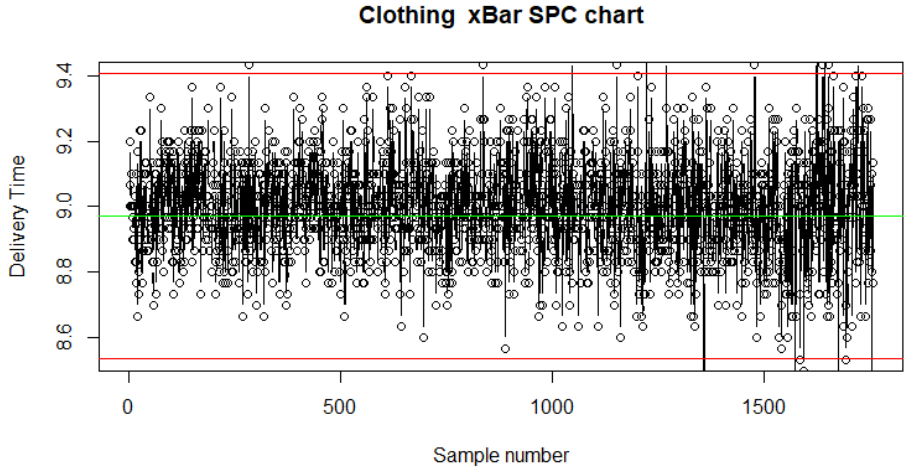


Figure 22: Control Chart of Clothing for all Entries

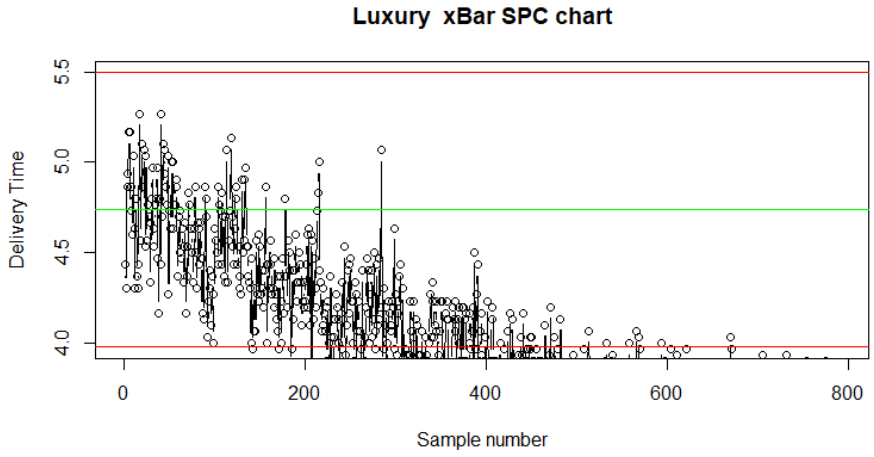


Figure 23: Control Chart of Luxury for all Entries

ALL ENTRIES CHARTS SUMMARY

For Gifts and Household items, the delivery time increased a lot. The company should come up with a solution considering the delivery times of these two classes will probably follow the same path. Increase in delivery time may result in loss of customers.

The Clothing, Sweets, Food and Technology classes, have a very consistent delivery times. The total delivery time differs depending on the class, but all four classes' delivery times stays consistent. This makes it acceptable, because the client/customer will have a very good idea, of when to expect the arrival of their product.

Luxury products are the only class whose delivery time is improving. The improvement of delivery times will result in an increase of customers, due to people that are generally impatient and wants their products as soon as possible. This is something that can be improved in the other products.

PART 4: OPTIMISING THE DELIVERY PROCESS

4.1

A)

CLASS	1ST	2ND	3RD	3RD FROM LAST	2ND FROM LAST	LAST	TOTAL
Technology	2.9	2.033	2.967	2.967	2.967	2	5
Clothing	17.5	17.267	17.433	17.1	16.967	17.5	17
Household	3.967	3.967	3.9	3.4	3.3	3.6	434
Luxury	2.267	2.733	2.667	2.667	2.667	2.667	5
Food	9.467	9.433	9.467	8.433	9.467	9.5	17
Gifts	42.233	42.467	42.833	57.367	54.567	55.8	400
Sweets	10.233	9.667	9.7	16.567	16.3	16.033	2290

Table 4: Sample Means

B)

CLASS	LONGEST CONSECUTIVE	LAST SAMPLE NUMBER
Technology	636	787
Clothing	224	1171
Household	497	791
Luxury	125	783
Food	406	529
Gifts	71	568
Sweets	225	644

Table 5: Consecutive Samples

4.2)

The R-code (1 - pnorm(3))*2 is used to determine the likelihood of making a type 1 error in a control chart utilizing 3 sigma control limits. This causes a type 1 error with a change of 0.2699%.

4.3)

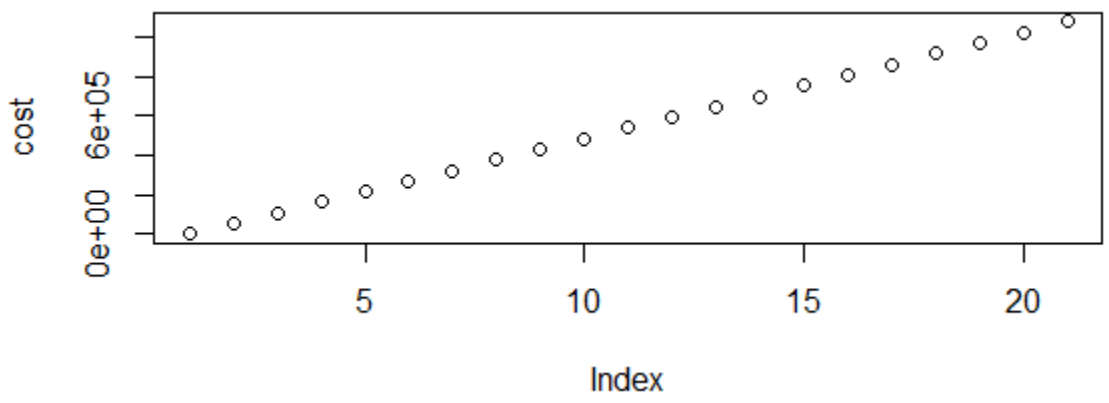


Figure 24: Price of late deliveries in Technology Class

4.4) The minimum delivery time is 0.5 hour, With a total cost of $ 387517.

PART 5: DOE & MANOVA

The following figure is the MANOVA results.

```
 Response 6 :
                Df  Sum Sq Mean Sq F value               Pr(>F)
Delivery.time    1    12.6 12.5779  145.98 < 0.00000000000000022 ***
Residuals   179976 15507.3  0.0862
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

 Response 7 :
                Df  Sum Sq Mean Sq F value               Pr(>F)
Delivery.time    1    18.3 18.2666  196.78 < 0.00000000000000022 ***
Residuals   179976 16706.3  0.0928
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

 Response 8 :
                Df Sum Sq Mean Sq F value               Pr(>F)
Delivery.time    1      7  6.9908  68.966 < 0.00000000000000022 ***
Residuals   179976  18244  0.1014
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

 Response 9 :
                Df  Sum Sq Mean Sq F value     Pr(>F)
Delivery.time    1     1.7 1.68115  15.385 0.00008773 ***
Residuals   179976 19666.7 0.10927
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

 Response 1 :
                Df  Sum Sq Mean Sq F value               Pr(>F)
Delivery.time    1   934.2  934.17  6394.8 < 0.00000000000000022 ***
Residuals   179976 26291.4    0.15
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

 Response 2 :
                Df  Sum Sq Mean Sq F value               Pr(>F)
Delivery.time    1    14.8 14.8268  188.07 < 0.00000000000000022 ***
Residuals   179976 14188.4  0.0788
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

 Response 3 :
                Df  Sum Sq Mean Sq F value               Pr(>F)
Delivery.time    1    14.9  14.867  172.83 < 0.00000000000000022 ***
Residuals   179976 15482.3   0.086
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

 Response 4 :
                Df  Sum Sq Mean Sq F value               Pr(>F)
Delivery.time    1    28.3 28.2769  319.43 < 0.00000000000000022 ***
Residuals   179976 15931.8  0.0885
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

 Response 5 :
                Df  Sum Sq Mean Sq F value               Pr(>F)
Delivery.time    1    33.4  33.399  385.91 < 0.00000000000000022 ***
Residuals   179976 15576.2   0.087
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1
```

Figure 25: MANOVA Results

Ho: The cost or timing of a product's delivery are unaffected by a person's class. Ha: The Price and Delivery Time for the Products are Affected by Class. Given that both the price and the delivery time

of the product are significantly influenced by the class of the product, the null hypothesis was strongly rejected.

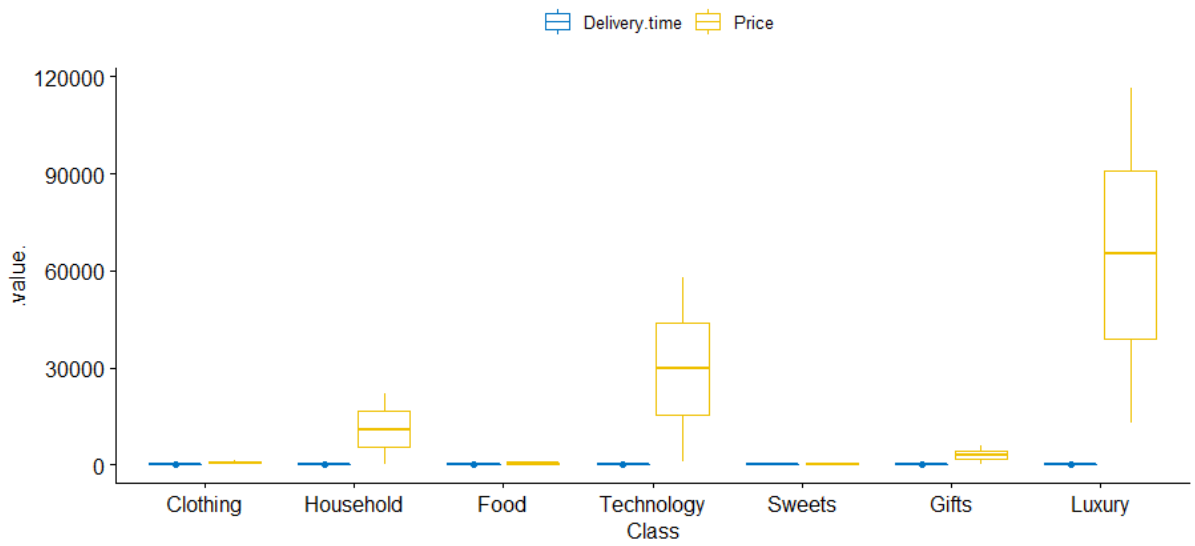


Figure 26: Box Plot of MANOVA results

This box plot clearly shows the relationship between price and class; but, because the y axis is not comparable, it is challenging to understand the relationship between delivery times.

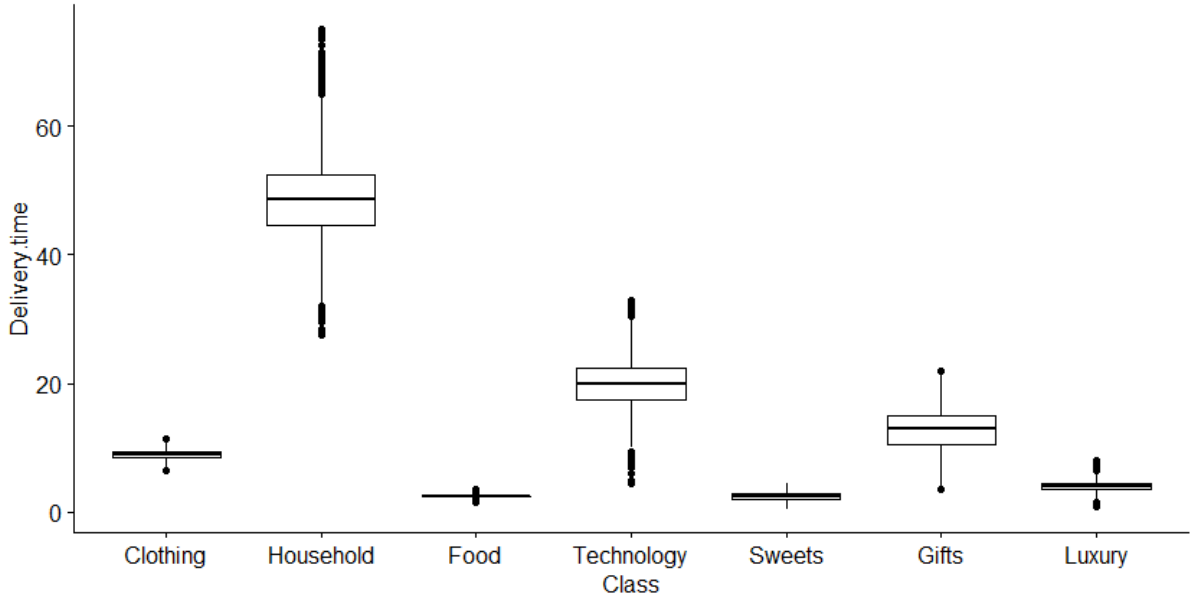


Figure 27: Box Plot of Delivery Time vs Class

This plot's y axis is more appropriate, making the relationship between class and delivery time much more obvious. In contrast to the wide variety of household goods, it is also easy to notice how accurately food is delivered.

PART 6: RELIABILITY OF SERVICE AND PRODUCTS

6.1)

6) L = k(y–m)2 = 45 when (y-m)=0.04

k = 45/(0.042) = 28125

L = 28125(y-0.06)2

7 a) k = 35/(0.042)

= 21875

L = 21875(y-0.06)2

b) L= 21875*0.0272

= 15.95

6.2) The reliability is simply: If only one machine in each of A, B, and C is operational. RA * RB * RC = 0.85 * 0.92 * 0.9 = 0.7038. The first process will be more reliable if both of each machine, A, B, and C, are operational. The combined dependability of components A that are connected in parallel is 1 - probability (both fail). When both A's fail, the combined reliability is RAA = 1- 0.0225 = 0.9775, with a failure probability of (1-0.85)2 = 0.0225.

The total reliability of both Bs is RBB = 1- 0.0064 = 0.9936, and both Bs fail with a probability of (1-0.92)2 = 0.0064. The combined dependability of both Cs is RCC = 1- 0.01 = 0.99 when they fail with a probability of (1-0.9)2 = 0.01. RAA*RBB*RCC = 0.9775*0.9936*0.99 = 0.9615, or 26% improvement, is the reliability.

CONCLUSION

The business has a variety of options for improving the dependability of its distribution system. The measurement of how the business is doing would be significantly impacted by the implementation of items like control charts. Keeping consumers pleased by reducing delivery times is also a smart idea because it is simpler to keep an existing client than to acquire a new one. A smart concept is to encourage the selling of expensive goods because they sell for more money and you do not have to sell as many of them.

Overall, I'd say the organization has a distribution system that is somewhat dependable, could use some minor adjustments, but is generally enough for their needs.

REFERENCES

Simplilearn. 2022. *What is Data Wrangling*. [online] Available at: <https://www.simplilearn.com/data-wrangling-article#:~:text=Data%20wrangling%20is%20the%20process%20of%20removing%20errors%20and%20combining,analysis%20is%20becoming%20increasingly%20necessary.> [Accessed 17 October 2022].

Investopedia. 2022. *Descriptive Statistics: Definition, Overview, Types, Example*. [online] Available at: <https://www.investopedia.com/terms/d/descriptive_statistics.asp> [Accessed 17 October 2022].

Kiran, D., 2017. *Process Capability*. [online] Science Direct. Available at: <https://www.sciencedirect.com/topics/engineering/process-capability-index#:~:text=Process%20capability%20index%20measures%20the,come%20up%20to%20our%20expectations.> [Accessed 18 October 2022].

Asq.org. 2022. *What is Statistical Process Control? SPC Quality Tools | ASQ*. [online] Available at: <https://asq.org/quality-resources/statistical-process-control#:~:text=Statistical%20process%20control%20(SPC)%20is,find%20solutions%20for%20production%20issues.> [Accessed 19 October 2022].

Pqsystems.com. 2022. *Calculate Cp*. [online] Available at: <https://www.pqsystems.com/qualityadvisor/DataAnalysisTools/capability_4.6.1.php#:~:text=Cp%20is%20an%20index%20used,upper%20and%20lower%20specification%20limits.> [Accessed 19 October 2022].

Support.minitab.com. 2022. *Potential (within) capability for Normal Capability Analysis - Minitab*. [online] Available at: <https://support.minitab.com/en-us/minitab/21/help-and-how-to/quality-and-process-improvement/capability-analysis/how-to/capability-analysis/normal-capability-analysis/interpret-the-results/all-statistics-and-graphs/potential-within-capability/#:~:text=CPU%20is%20a%20measure%20of,the%20variation%20within%20the%20subgroups> [Accessed 19 October 2022].

Support.minitab.com. 2022. *Potential (within) capability for Normal Capability Analysis - Minitab*. [online] Available at: <https://support.minitab.com/en-us/minitab/21/help-and-how-to/quality-and-process-improvement/capability-analysis/how-to/capability-analysis/normal-capability-analysis/interpret-the-results/all-statistics-and-graphs/potential-within-capability/#:~:text=CPL%20is%20a%20measure%20of,the%20within%2Dsubgroup%20standard%20deviation> [Accessed 19 October 2022].

Process capability (CP) and performance (CPK) (no date) *InfinityQS*. InfinityQS. Available at :https://www.infinityqs.com/statistical-process-control/spc-chart-guide/process-capability-index-cp#:~:text=Cpk%20represents%20the%20lowest%20value,upper%20and%20lower%20specification%20limit. (Accessed: October 20, 2022).