2022

Quality Assurance
ECSA Graduate Attributes Project
University of Stellenbosch

22966838

JPJ BEUKES

Contents

Table of figures

Introduction:

This report entails explorative data analysis of different production lines and processes.The structure of the body is divided into six different parts containing important information regarding the data given.

The first part is data wrangling where the data set is cleaned and normalised to ensure accurate predictions and correct analysis. The second part is descriptive statistic where different classes are analysed according to different business activities.

The third section is statistical process control followed by optimising the delivery time process in part four.

In part five, two MANOVA tests are done and in the last section different problems are solved by determining the reliability of service and products.

The report ends with a short conclusion, summarising what is entailed in the body and highlighting key points.

Part 1-Data wrangling:

Data wrangling, also known as data cleaning or data remediation refers to several procedures intended to convert unstructured data into forms that are easier to work with. Depending on the data being used and the objective to achieve, the precise approaches vary from project to project.

In the modern era where datasets can be substantially large and continuous, the data wrangling process should be done judiciously and with care to ensure that the data being used is correct, applicable and that the predictions derived from the data can be as accurate as possible. This allows businesses and organisations to have greater insights into their data and enables them to make informed decisions which lowers the possibility of making decisions which can negatively impact the business or organisation.

Hence, we are given a dataset, *salesTable2022*, there is no need for the process of data collection. By looking at the dataset given, one might assume that there are no data quality issues involved but proceeding with this assumption can result in incorrect analysis and inaccurate predictions.

Dataset *salesTable2022* consists of 180 000 rows (instances) and 10 descriptive features. These features can be split into qualitative and quantitative features. The quantitative features are *X, ID, AGE, Price, Year, Month, Day and Delivery time* as where the qualitative features are *Class and Why.Bought.*

FEATURE	**MININMUM**	**1ST QUARTILE**	**MEAN**	**3RD QUARTILE**	**MAXIMUM**	**MEDIAN**
AGE	18	38	54.57	70	108	53
PRICE	-588.8	482.3	12293.7	15270.7	116619	2259.6
YEAR	2021	2022	2025	2027	2029	2025
MONTH	1	4	6.521	10	12	7
DAY	1	8	15.54	23	30	16
DELIVERY.TIME	0.5	3	14.5	18.5	75	3

Table 1: Continuous feature's characteristics

The dataset is cleaned by removing missing values as well as negative price values from the dataset. The data set *salesTable2022* is split into two separate datasets, *invalidData,* which contains missing values (NA) and negative *"Price"* values as well as *validData,* which contains the valid data.

Invalid data with missing and negative values – 22 instances

Valid data that will be used – 179 978 instances

Part 2-Descriptive statistics:

Descriptive statistics are used to summarise a given data set by measuring the relations between features in a sample or population. Descriptive statistics are broken down into measures of central tendency and variability. Measures of central tendency consist of the mean, median, and mode, while measures of variability include the standard deviation, variance, minimum and maximum variables.

Class

The purchase frequency of the various product classes is not uniformly distributed. This can be due to various items that has a higher demand than others or it could also be a result of several items being more affordable than the rest. The sales frequency of gifts and technology is the highest which can be a consequence of more frequent advertising for this class of products, or the items can be subjected to discounted prices. Clothing and food products are seen as basic needs, which explains why these products are also bought regularly. Household appliances and sweets are also bought on a regular basis but isn't considered as basic needs and thus have a lower sales frequency than Food and Clothing. Luxury items has the lowest frequency which can be due to products having high purchase prices and a lower demand.

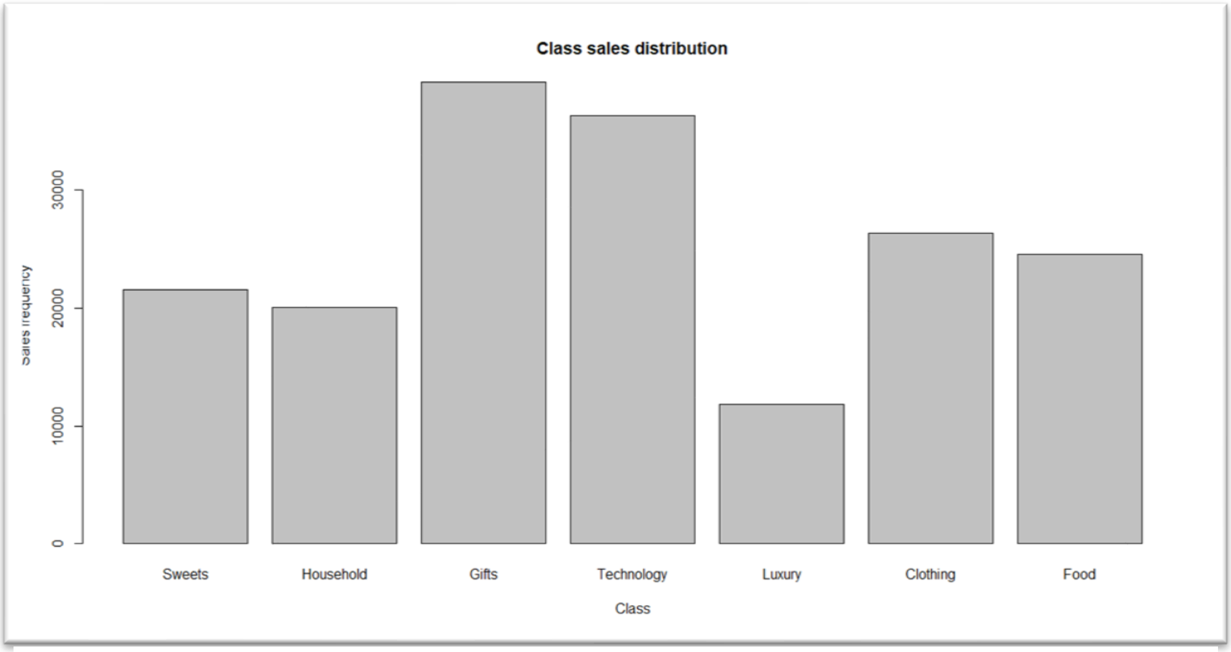


Figure 1: Class sales distribution

Price

The Price sales distribution is unevenly distributed with products that fall in the price range from 0-10 000 having the highest sales frequency. This is a result of most items from the gifts, clothing, food, sweets and household classes having a purchase price of R10 000 or lower. Another reason can be due to higher demand for these products since they fall in the lowest price bracket and is thus more affordable. Items having a value greater than R60 000, can potentially be considered outliers since the frequency of them being purchased are significantly lower than the rest of the goods.

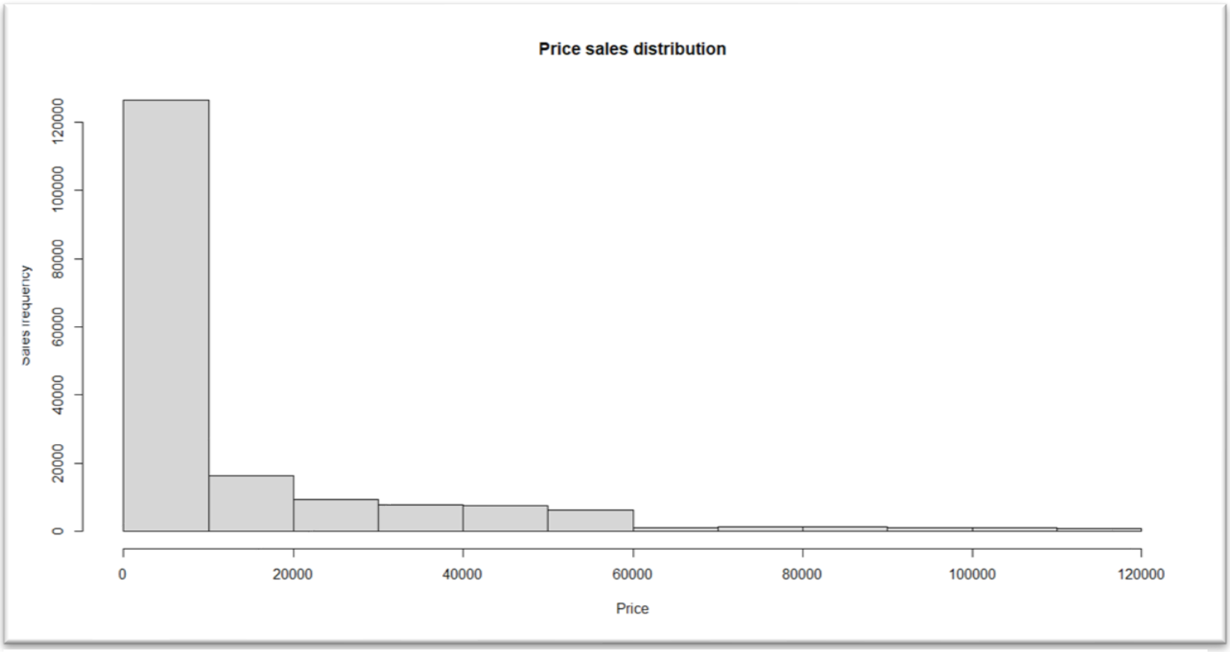


Figure 2: Price sales distribution

Age

The distribution of the age of customers is right skewed due to the mean having a higher value than the median. The graph displays a strong negative correlation starting from age 40. As the value of age increases the frequency of sales decreases as a result of people being more conservative with their money due to more financial responsibilities for example kids and retirement funding. People aged 35-40 has the highest sales frequency which can be due to stable income and not too many financial responsibilities. The lowest frequency of sales occurs at age groups 0-20 and 95-105 because these people don't earn their own income and are dependent on financial support from others who are generally aged 30 to 55. The sales frequency of people between ages 65-90 are relatively high and is declining for every age group as a result of people being conservative with their finances as they are most likely retired and living of their retirement fund.

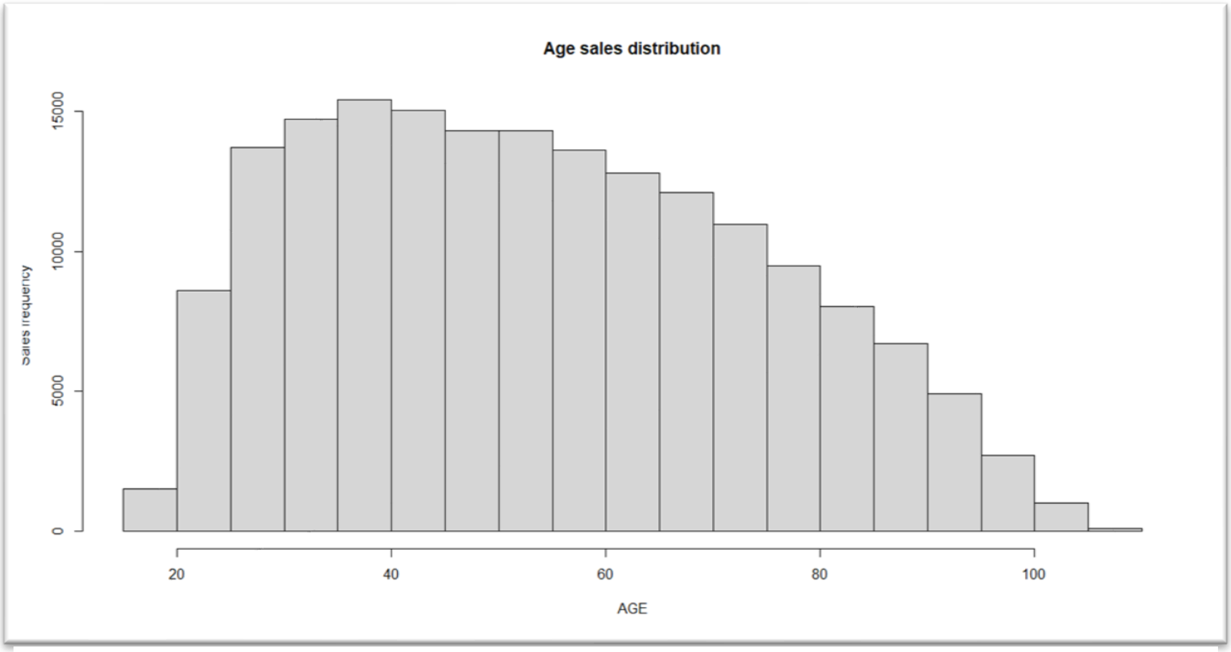


Figure 3: Age sales distribution

Why bought

The distribution below is unevenly distributed. Approximately 110 000 purchases result from products being promoted via recommendation. People who visited the website and browsed the internet resulted in 50,000 purchases. The class email makes a very small contribution to total sales; thus, businesses should not spend too much money on it for advertising.

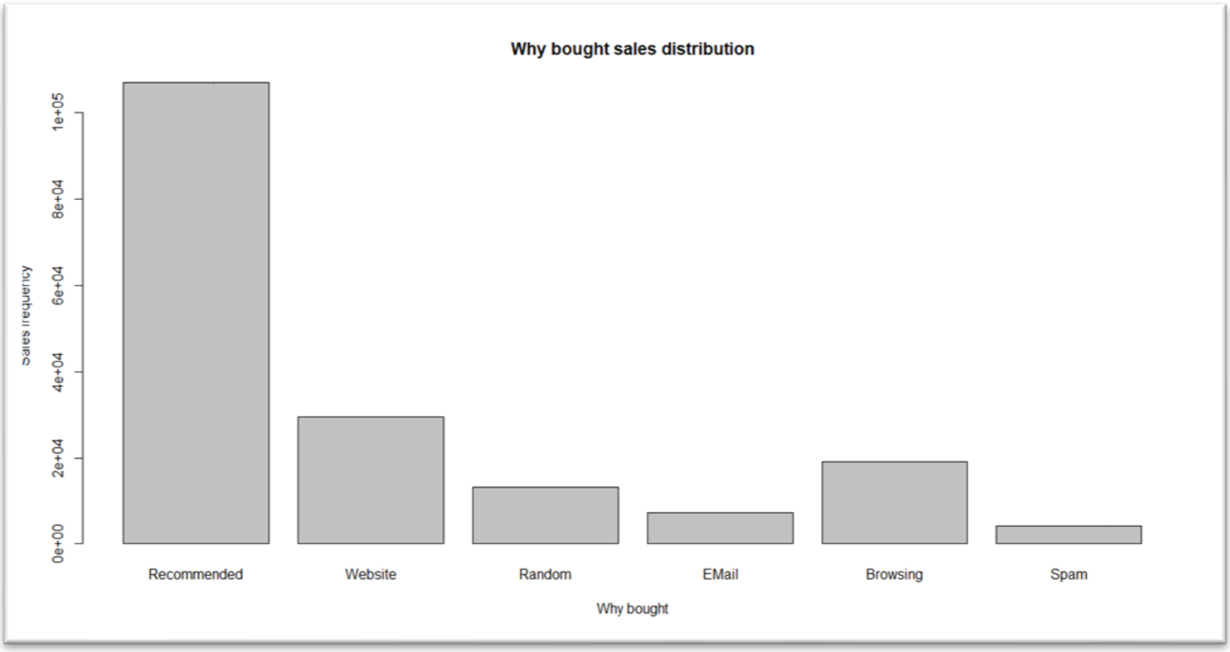


Figure 4: Why bought sales distribution

Sales vs Year

A quarter of the entire quantity of sales were made in 2021, a sharp decline in the number of sales occurred in 2022 which could be due to an economic crisis such as a global pandemic or the start of a recession. After the sharp decline, the sales started increasing and stagnated in the years 2025 and 2026. Since 2026 through to 2029, there is a definite, positive correlation between the volume of sales and the year in which the sales are made. This can be due to the economic crisis that has passed, and items being advertised on a larger scale as the company shows consistent growth.

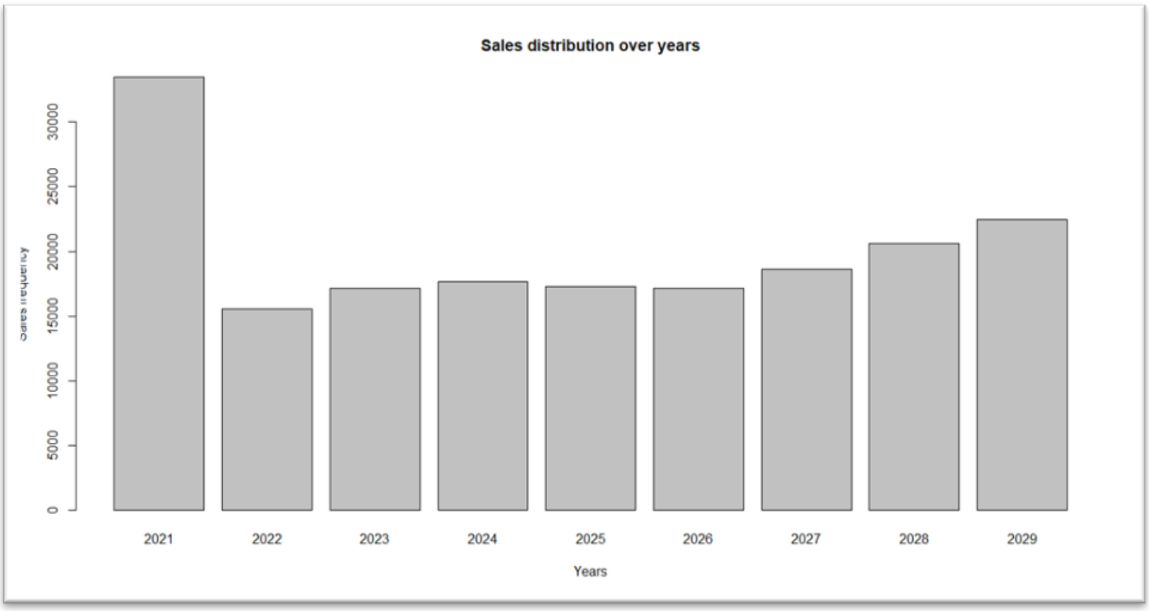


Figure 5: Sales distribution over years

Price vs Age

The boxplots depict the price range for each of the customers' various ages. For each age group, the distribution of the price range is skewed to the right. The upper quartile value of the price range for every age value is large which confirms the presence of outliers. When excluding the average values of price for younger age groups, the average value of price starts to increase from age 29 up to age 38 and decreases thereafter, confirming that people in the age group of 30 to 45 can afford more items. Businesses and organisations should regard these pupils as a priority as they largely contribute to the overall income of these institutions.

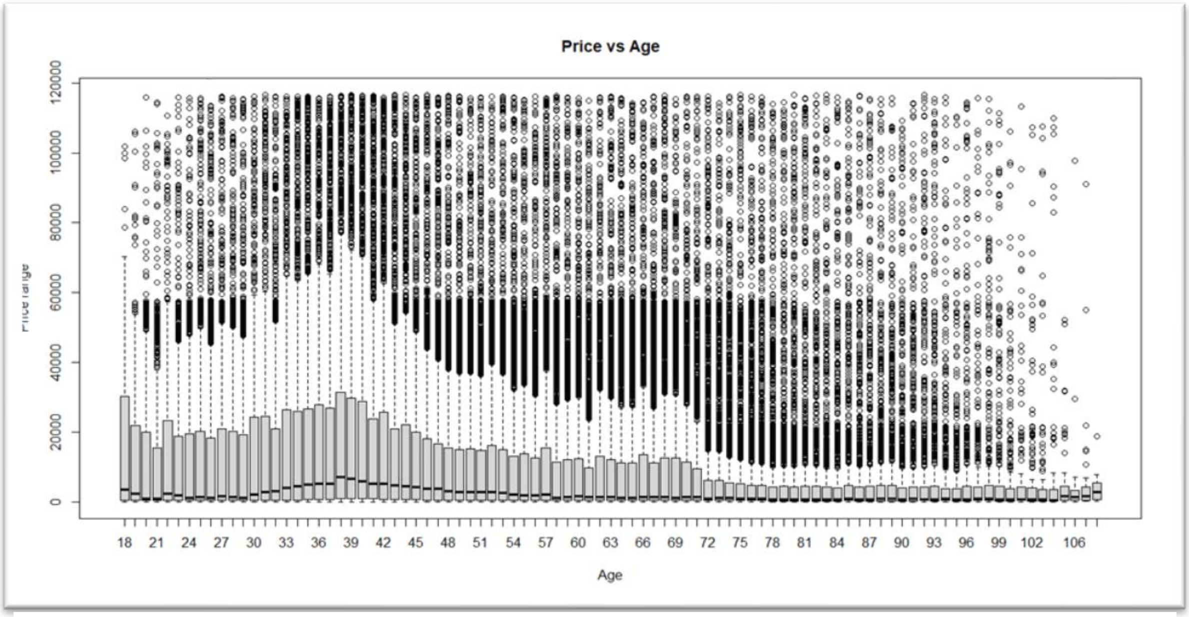


Figure 6: Price vs Age

Price vs Class

The price range of each class is normally distributed. The symmetric boxplots show that the mean price for each class is equal to that class's median price. Luxury items has the highest mean price with the mean price of Technology being second. The mean price of Clothing, Food and Sweets are exceptionally smaller than the mean price of Luxury and Technology.The mean price of Household and Gifts are also substantially low and aren't far from the lower classes' mean price. The price range for each class is independent of one another and has no influence each other.

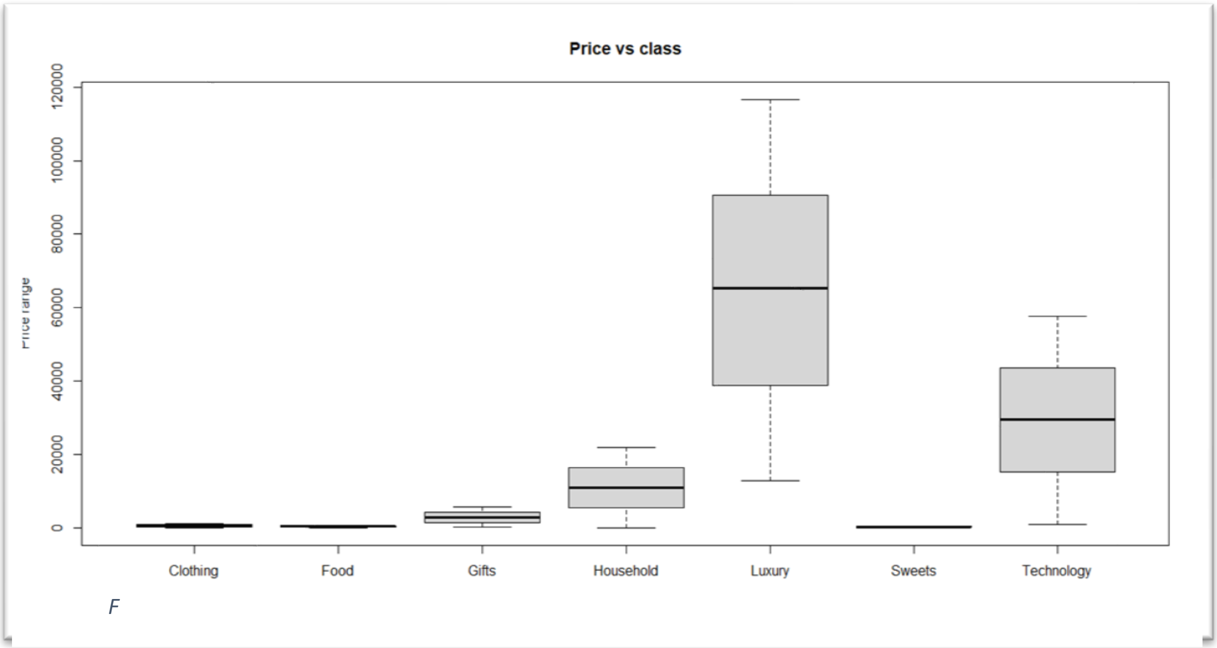


Figure 7: Price vs Class

Price vs Delivery time

As seen on the graph, the boxplots for the price range corresponding to a certain set of delivery timeframes are grouped. A particular class is represented by each collection of boxplots. When analyzing the Delivery time vs. Class graph, there is a clear representation that the time required to deliver products from different classes vary from one another. This can be due to different lot sizes being used as well as different modes of transport.

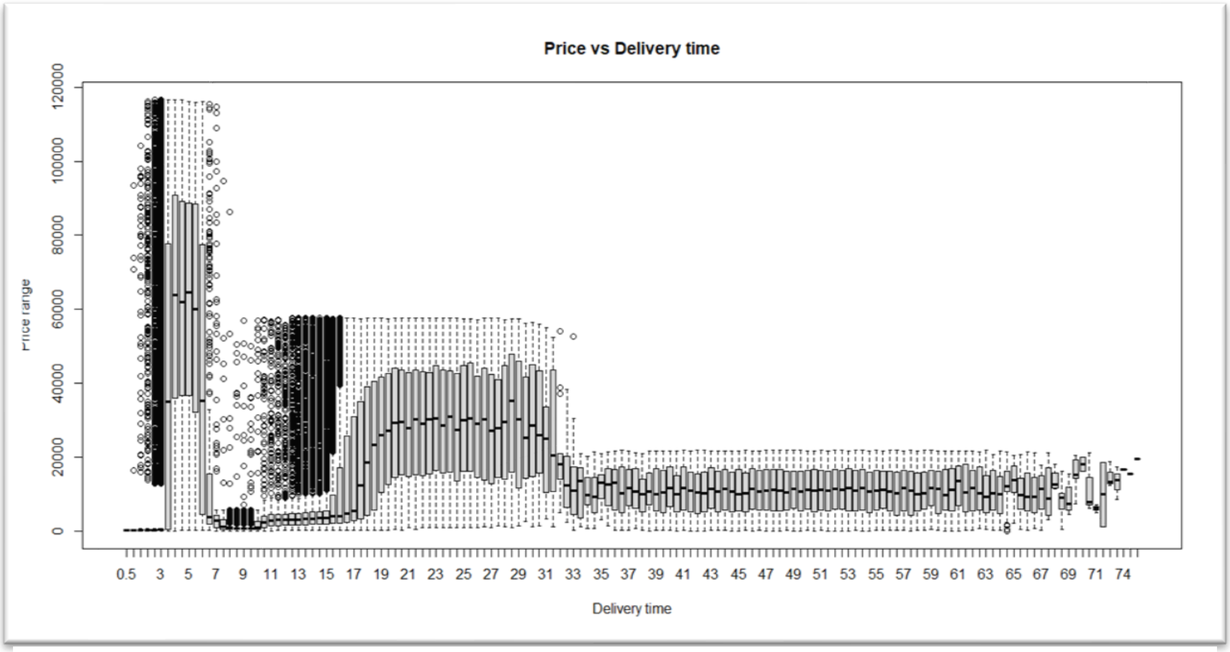


Figure 8: Price vs Delivery time

Age vs Class

The plots below provide information to the business on which age groups they should target for each particular class.

The distribution of the Age vs Class plot is unevenly distributed and there are a number of observations that can be made. The age range for sweets is skewed to the left and has a mean value of 60. This can be due to grandparents buying sweets for their grandchildren when they come and visit. The age range for technology, luxury, household appliances and clothing are all skewed right and has a mean value between 40 and 50. This shows that customers between the ages of 40 and 50 are financially stable and are able to buy several products from several classes due to their financial position. Another reason can be that customers in this age group usually have kids that they have to take care of which increases the number of purchases they make.

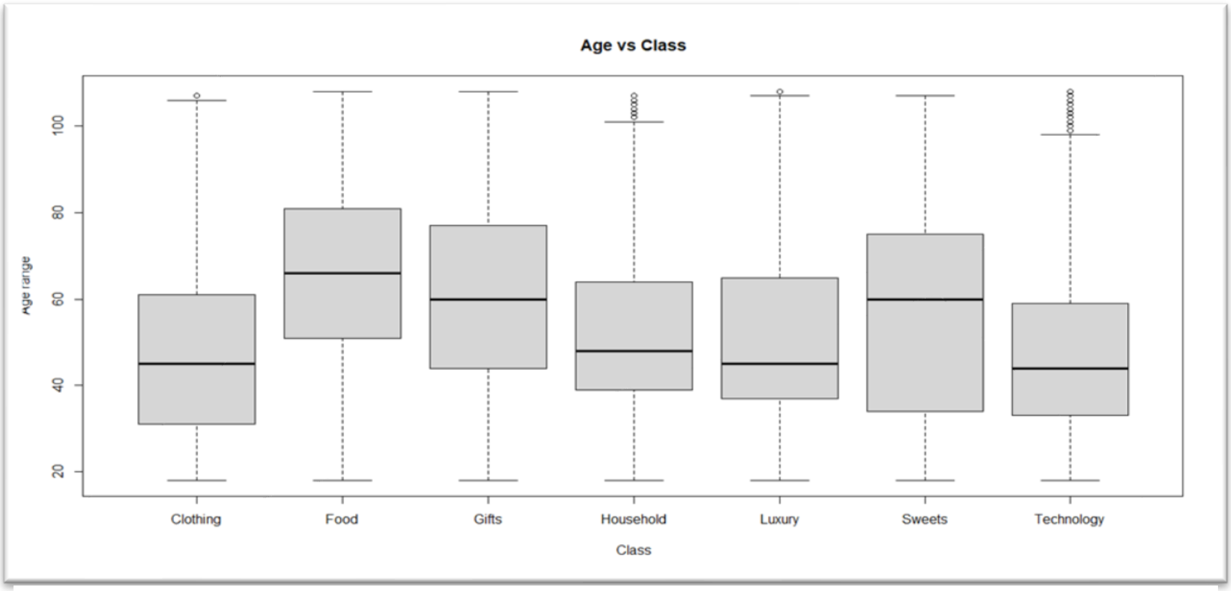


Figure 9: Age vs Class

Delivery time vs Class

When looking at the Delivery Time vs Class plot one would find that each boxplot represents the delivery time range for different items in their respective classes. Sweets, food and luxury items has the lowest delivery time out of all the classes and can be due to higher quantities being shipped at a time. Household, technology and gifts have higher delivery times which can be due to more steps being involved in the delivery. When looking at mean delivery time values, it is seen that household items has the largest value whereas sweets and food has the lowest values. The open circles at the sides of the boxplots indicates that outliers are present and should be scrutinized to obtain a more stable approximation.

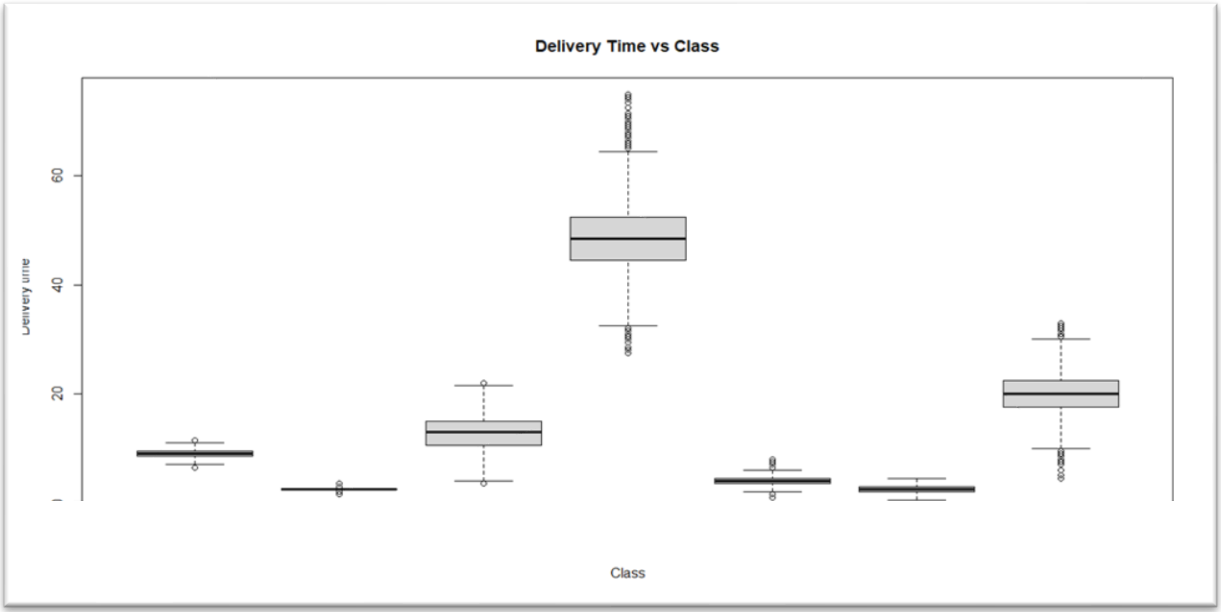


Figure 10: Delivery Time vs Class

Delivery time vs Year

When inspecting the plot below from years 2022 to 2029, a positive correlation arises between the time it takes items to be delivered and the year in which these items are delivered. As the years go by, the average delivery time increases. The increase in mean delivery time confirms that as the total sales increases over time, the mean delivery time will also increase. This can negatively affect the company as the waiting time for customers increases and customers can become unsatisfied. It is recommended that the business should investigate this problem and find a solution to it.

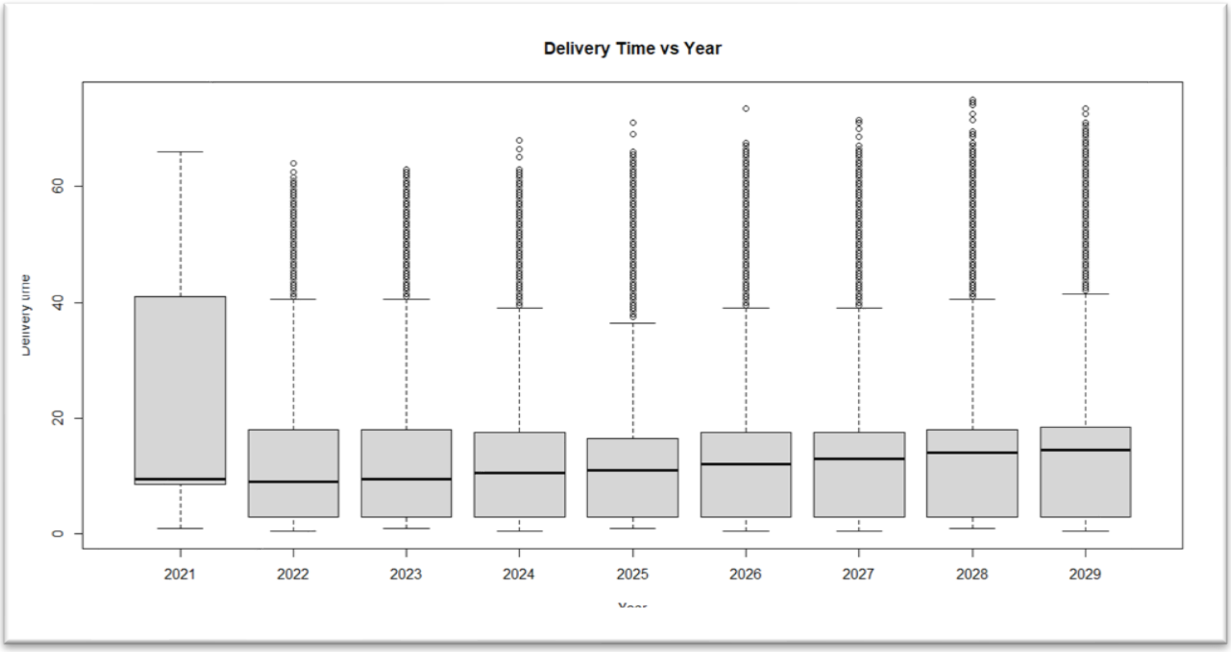


Figure 11: Delivery Time vs Year

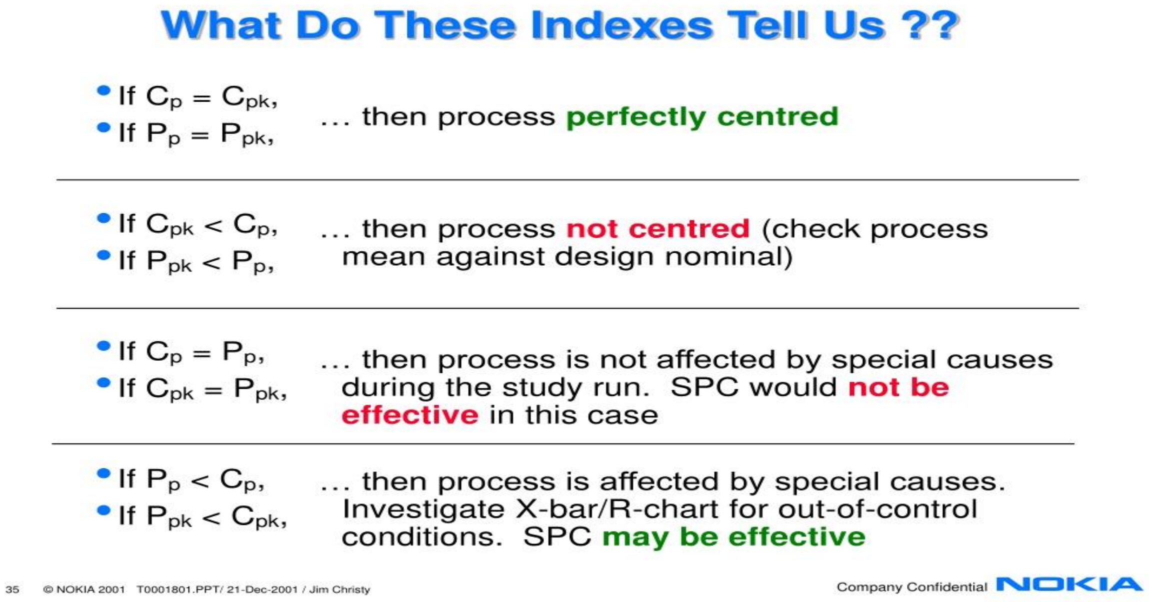
Process Capabilities

Process capability is defined as a statistical measure of the inherent process variability of a given characteristic. One can use a process-capability study to assess the ability of a process to meet specifications. When a process' capability is acknowledged and documented, it can be used to prioritise the timing of process modifications, track continuous improvement over time using trends, and assess a process' ability to meet client needs.

Index	Equation	Definition
Cp	$\frac{(USL - LSL)}{6\sigma}$	Capability potential: Illustrate a process's ability to meet specifications. It shows how well the process spread fits into the specification range.
Cpu	$\frac{(USL - Xbar)}{3\sigma}$	Process capability based on the upper specification limit

Cpl	$\frac{(Xbar - LSL)}{3\sigma}$	Process capability based on the lower specification limit
Cpk	$\min(Cpu, Cpk)$	Capability performance: Measures how close the process mean is to the target value of the specification.

Table 2: Process Capabilities



What Do These Indexes Tell Us ??

- If $C_p = C_{pk}$,
- If $P_p = P_{pk}$,

... then process **perfectly centred**

- If $C_{pk} < C_p$,
- If $P_{pk} < P_p$,

... then process **not centred** (check process mean against design nominal)

- If $C_p = P_p$,
- If $C_{pk} = P_{pk}$,

... then process is not affected by special causes during the study run. SPC would **not be effective** in this case

- If $P_p < C_p$,
- If $P_{pk} < C_{pk}$,

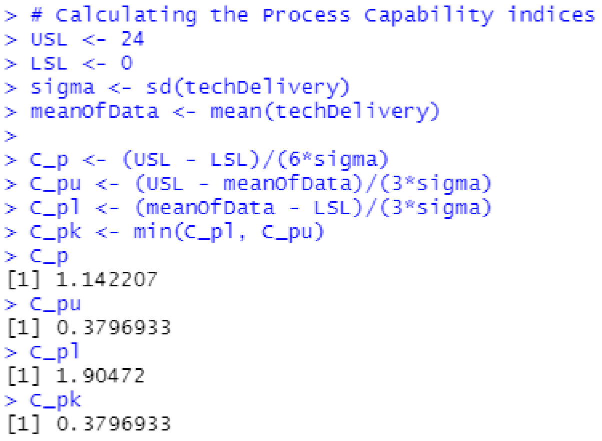
... then process is affected by special causes. Investigate X-bar/R-chart for out-of-control conditions. SPC **may be effective**

35 © NOKIA 2001 T0001801.PPT/ 21-Dec-2001 / Jim Christy

Company Confidential NOKIA

(SlideServe, 2022)

Process capability indices for the process delivery times (in hours) of the technology class items



```
> # Calculating the Process Capability indices
> USL <- 24
> LSL <- 0
> sigma <- sd(techDelivery)
> meanOfData <- mean(techDelivery)
>
> C_p <- (USL - LSL)/(6*sigma)
> C_pu <- (USL - meanOfData)/(3*sigma)
> C_pl <- (meanOfData - LSL)/(3*sigma)
> C_pk <- min(C_pl, C_pu)
> C_p
[1] 1.142207
> C_pu
[1] 0.3796933
> C_pl
[1] 1.90472
> C_pk
[1] 0.3796933
```

An LSL value of 0 for the delivery times of items is logical as time can't be a negative value. A Cp value of 1.142207 indicates that the spread of values fit into the limits. The Cpu and Cpk values are the same with a value of 0.3796933. This indicates that there are several non-conforming products and the process should be investigated.

Part 3- Statistical process control (SPC):

3.1 SPC for first 30 samples

Statistical process control during the manufacturing process uses inline data collected from the operations that produce the products in real-time. The process's state of control is established using statistical methods and techniques. By displaying a graphical representation of the variance of the process, the statistical driven process information can contribute to more knowledge and better understanding of the process. (ASQ, 2022)

The validData dataset, consisting of information with regards to sales, is sorted according to date (year, month and day) with the first instance signifying the first sale made. A total of 30 samples are used, with each sample having a size of 15 instances. This data is used to create the X- and S- charts for the delivery process times of the different processes by using the oldest data collected first.

X-chart values

CLASS	UCL	U2SIGMA	U1SIGMA	CL	L1SIGMA	L2SIGMA	LCL
SWEETS	2.897	2.7573	2.6175	2.4778	2.338	2.1983	2.0585
HOUSEHOLD	50.2483	49.0196	47.7909	46.5622	45.3335	44.1048	42.8761
GIFTS	9.4886	9.1127	8.7369	8.3611	7.9853	7.6095	7.2337
TECHNOLOGY	22.9746	22.1079	21.2412	20.3744	19.5077	18.641	17.7743
LUXURY	5.494	5.2412	4.9884	4.7356	4.4828	4.2299	3.9771
CLOTHING	9.4049	9.26	9.115	8.97	8.825	8.68	8.5351
FOOD	2.7095	2.6363	2.5632	2.49	2.4168	2.3437	2.2705

Table 3.1: X-chart values

S-chart values

CLASS	UCL	U2SIGMA	U1SIGMA	CL	L1SIGMA	L2SIGMA	LCL
SWEETS	0.8353	0.734	0.6327	0.5314	0.4301	0.3288	0.2274
HOUSEHOLD	7.3442	6.4534	5.5626	4.6719	3.7811	2.8903	1.9996
GIFTS	2.2463	1.9739	1.7014	1.429	1.1565	0.8841	0.6116
TECHNOLOGY	5.1806	4.5522	3.9239	3.2955	2.6672	2.0388	1.4105
LUXURY	1.5111	1.3278	1.1445	0.9612	0.778	0.5947	0.4114
CLOTHING	0.8666	0.7615	0.6564	0.5512	0.4461	0.341	0.2359
FOOD	0.4372	0.3842	0.3312	0.2781	0.2251	0.1721	0.119

Table 3.2: S-chart values

X and S charts – first 30 sample values

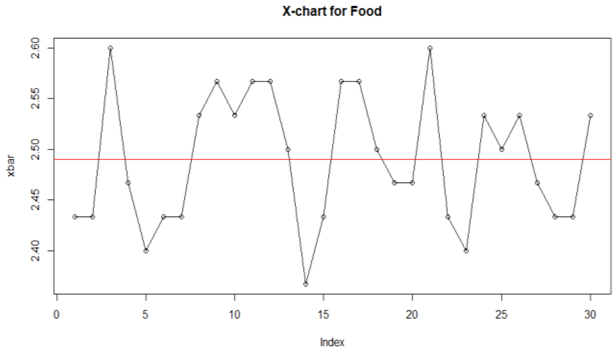


Figure 12: X-chart for Food

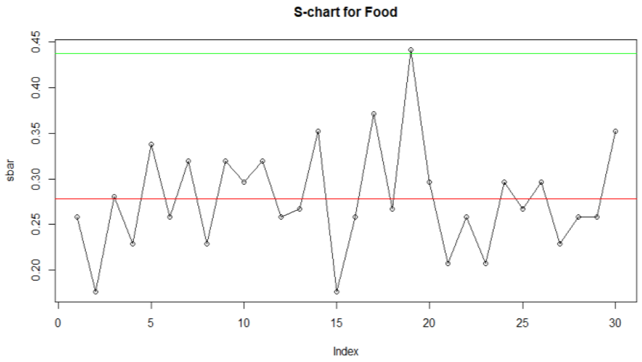


Figure 13: S-chart for Food

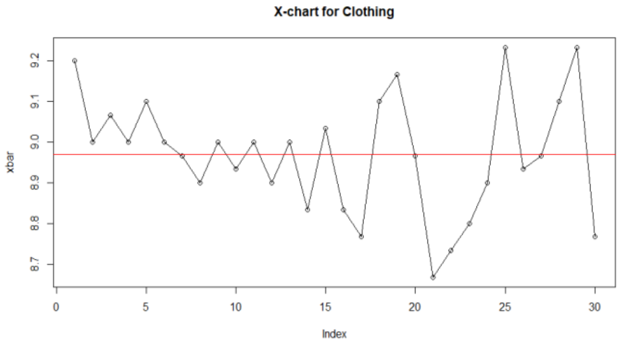


Figure 14: X-chart for Clothing

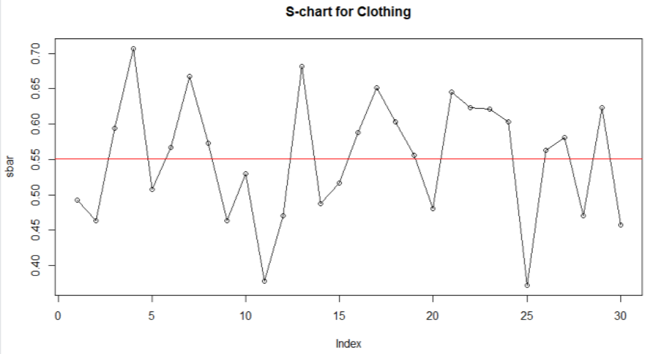


Figure 15: S-chart for Clothing

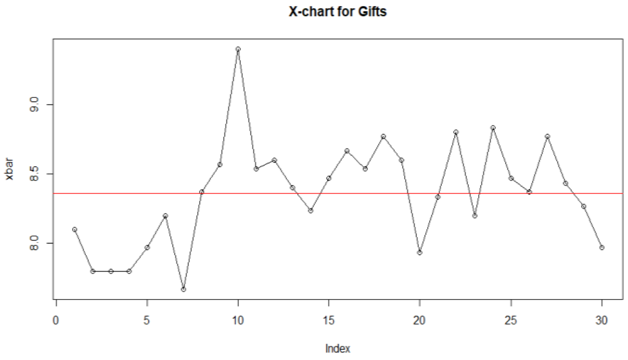


Figure 16: X-chart for gifts

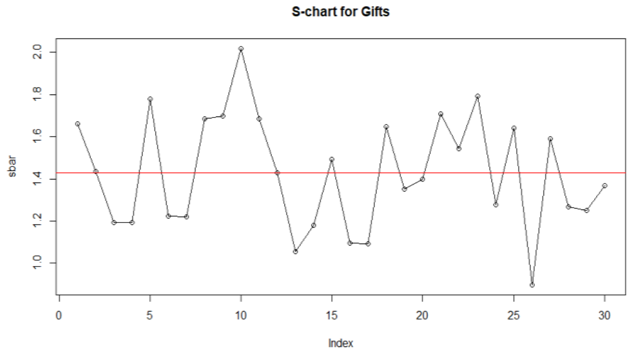


Figure 17: S-Chart for gifts

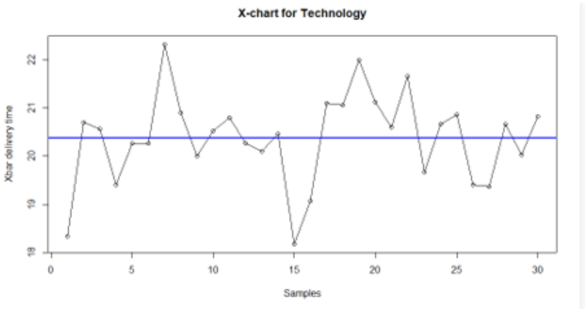


Figure 18: X-chart for technology

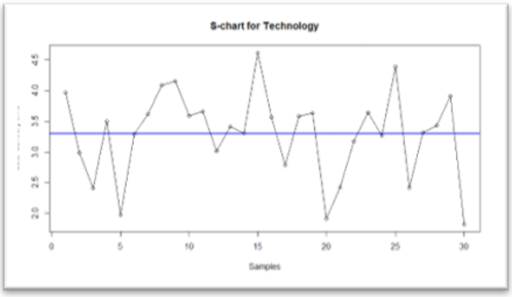


Figure 19: S-chart for technology

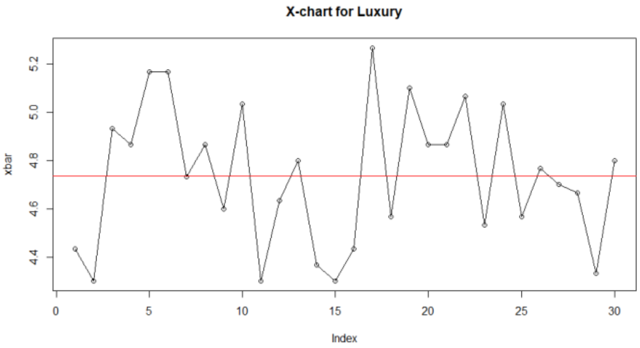


Figure 20: X-chart for Luxury

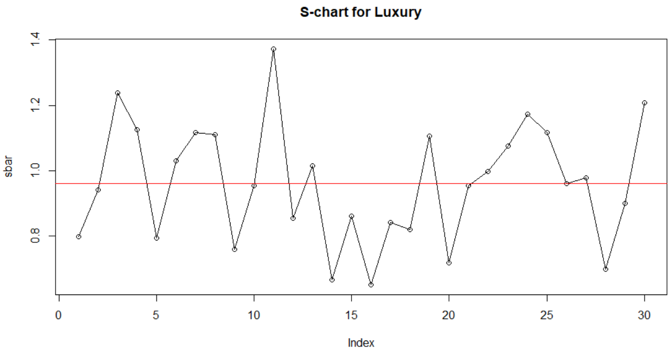


Figure 21: S-chart for Luxury

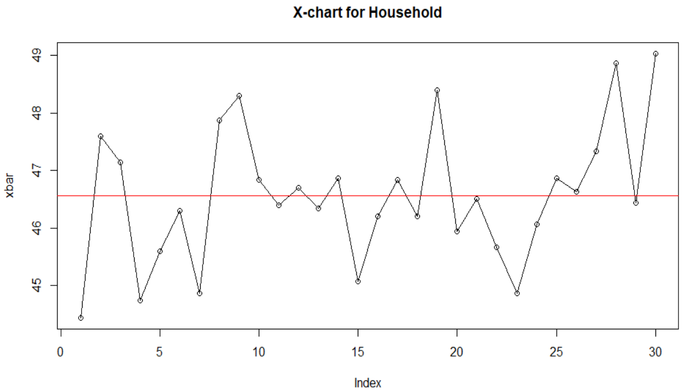


Figure 22: X-chart for Household

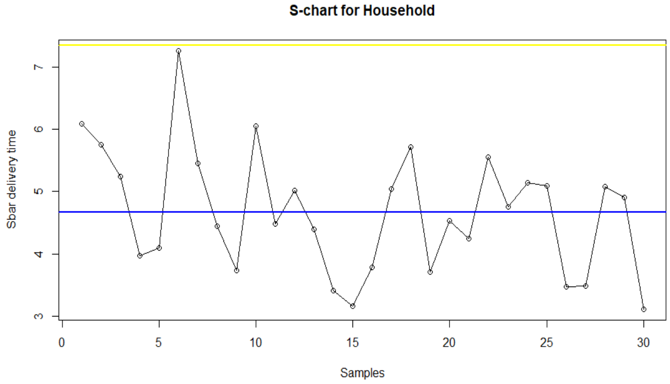


Figure 23: S-chart for Household

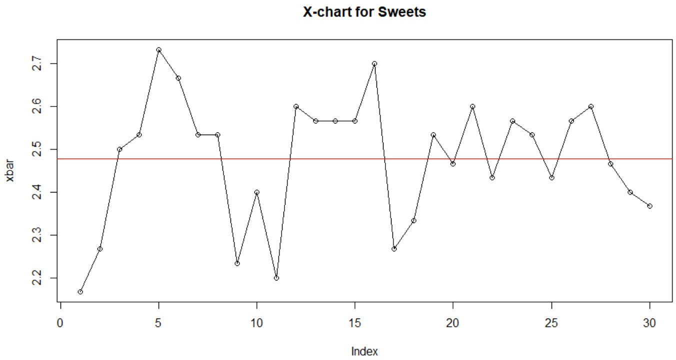


Figure 24: X- chart for Sweets

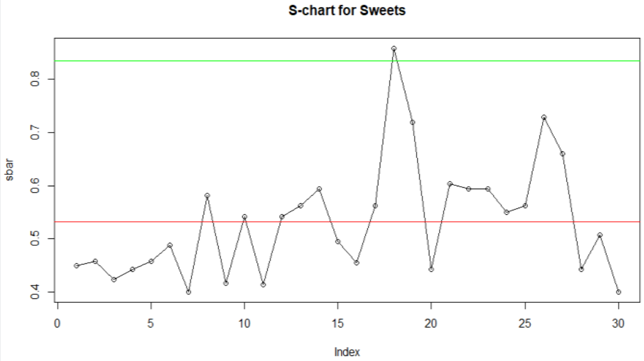


Figure 25: S-chart for Sweets

The Upper Control Limit (UCL) and the Lower Control Limit (LCL) form a corridor in whereby a quality characteristic meets the desired value or a normal deviation if it falls within. The upper and lower control limit are placed 3 sigma values above and below the center line. If a sample lies outside of these boundary limits, the process is out of control.

When inspecting the plots visually, one will find that nearly all the samples fall within the upper and lower limit, which suggests that the process for the delivery times of each class is in control, concluding that there is little to no variation present.

3.2 Statistical Process Control for all samples

X and S charts – first 30 sample values

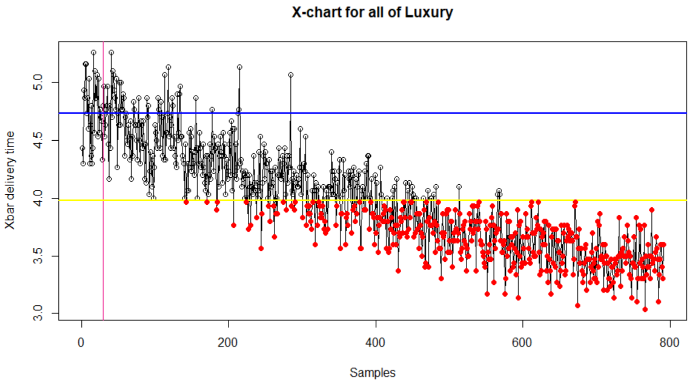


Figure 26: X-chart values for all of Luxury

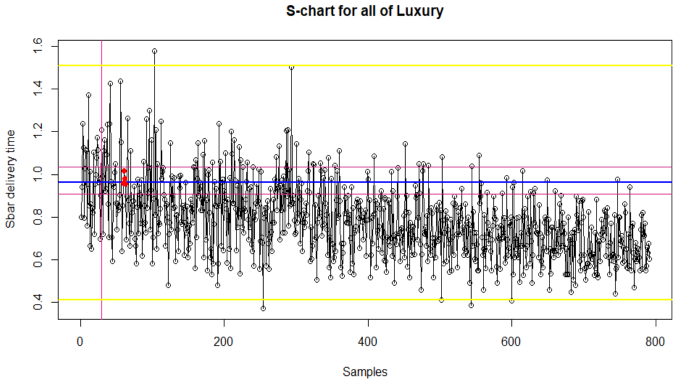


Figure 27: S-chart values for all of Luxury

The total number of samples used is 791. There are no values which lies outside of the upper and lower control limits for the first 120 samples. In total there are fewer than 4 outliers in a sample space of just below 800. This indicates that the process is stable and from graphs above it is clear that Luxury items are seen as a priority to the business as it significantly contributes to overall revenue.

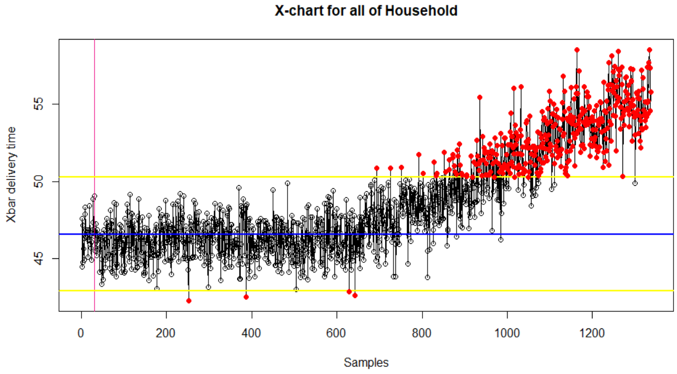


Figure 28: X-chart values for all of Household

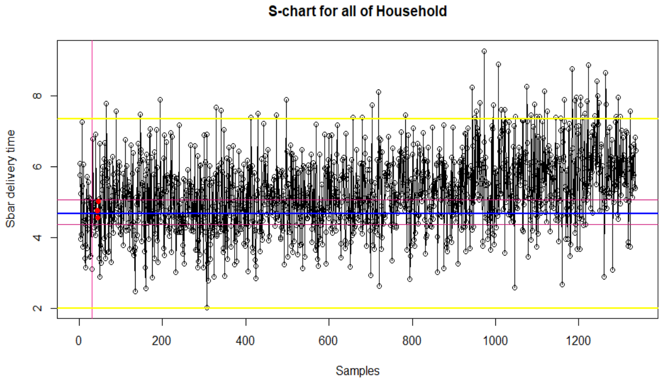


Figure 29: S-chart values for all of Household

For the Household class, the number of samples used is 1337. There are 9 outliers in the first 800 samples which indicates that the process was relatively stable but became rather unstable as time went by. The results are unwanted and should be inspected to find out where the problem lies and how it can be solved.

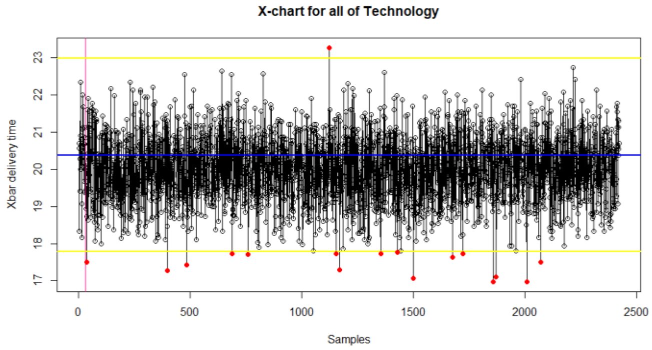


Figure 30: X-chart values for all of Technology

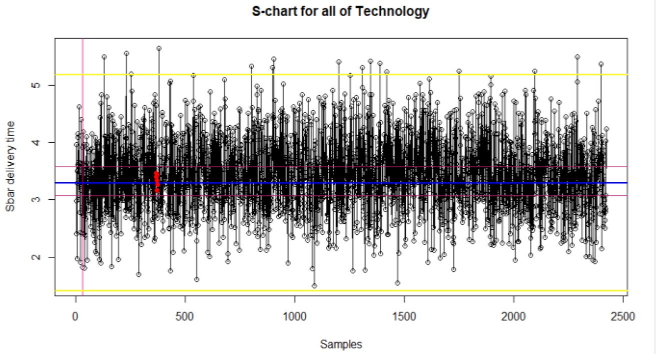


Figure 31: S-chart values for all of Technology

2423 samples were used for the plot above. Only 17 samples lie outside the UCL and LSL which is about 0.7% of the total instances. This gives an indication that the process is relatively stable and in control with little variation present.

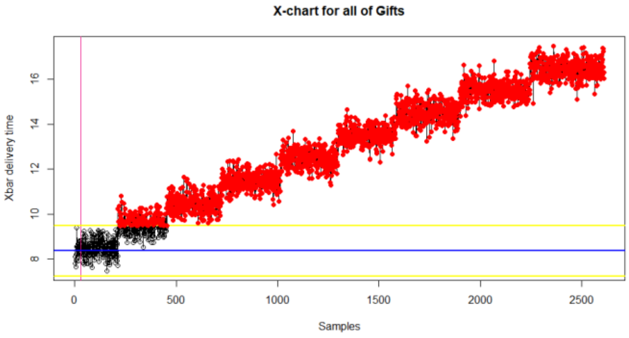


Figure 32: X-chart values for all of Gifts

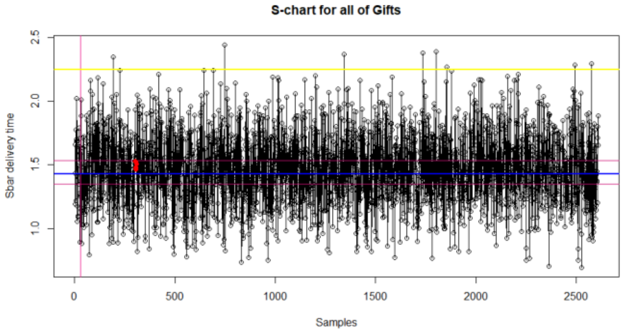


Figure 33: S-chart values for all of Gifts

The number of samples used is 2609. The mean delivery time for these samples increases every year and inspection should be done to find the root of the cause. The number of outliers present in this data indicates that the process is unstable and out of control, which can be an effect of gifts not seen as a priority or due to misdetection.

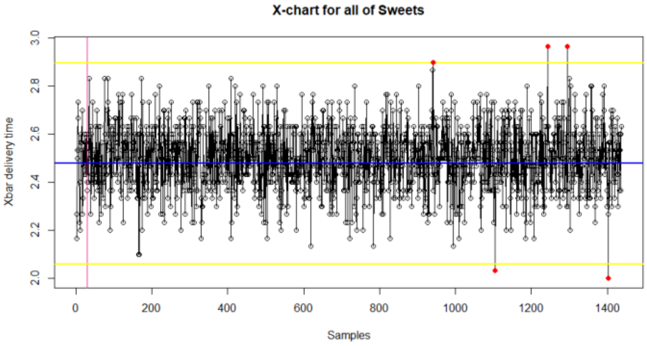


Figure 34: X-chart values for all of Sweets

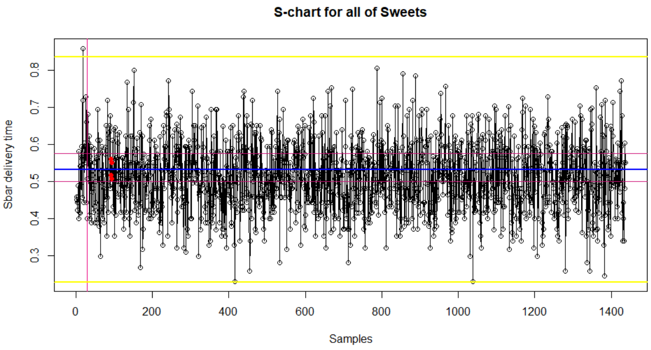


Figure 35: S-chart values for all of Sweets

A total number of 1437 samples were used. There are only 4 outliers present which indicates that the process for sweets is stable.

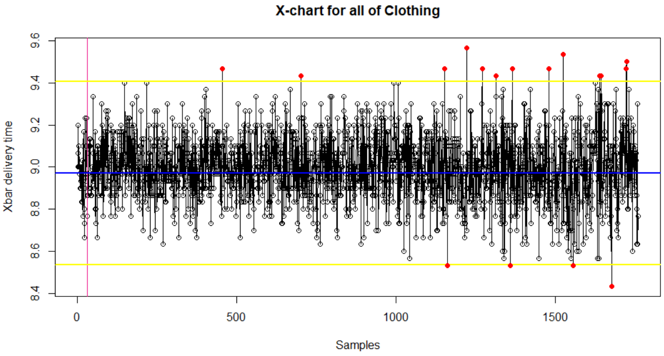


Figure 36: X-chart values for all of Clothing

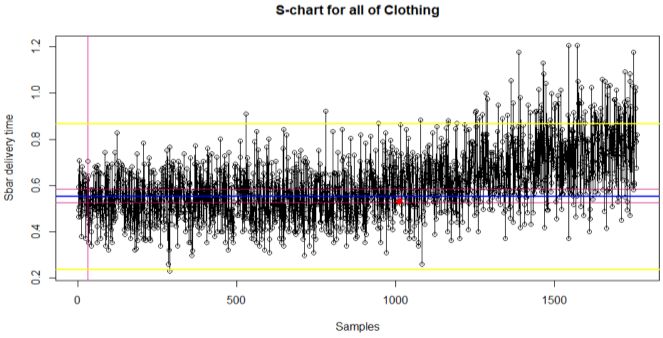


Figure 37: S-chart values for all of Clothing

In the X-chart for clothing, 1760 samples were used. There are 16 samples which do not lie within the UCL and LCL, indicating some variation within the process and a conclusion can be made that the process is relatively stable. If there is sufficient capacity and capital it would be advised to investigate the process to ensure stability and improve overall performance.

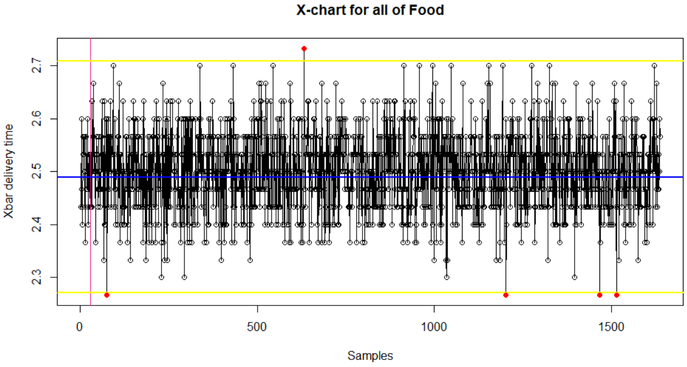


Figure 38: X-chart values for all of Food

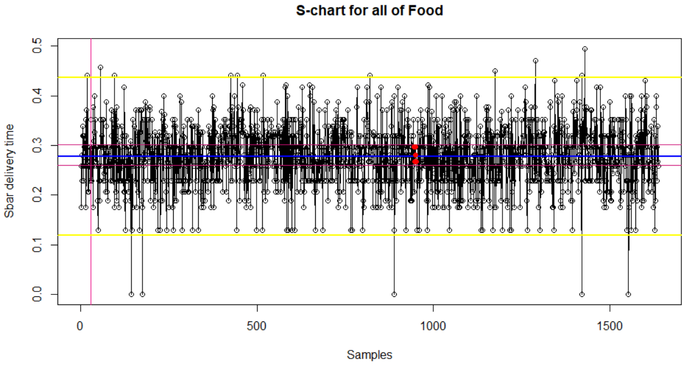


Figure 39: S-chart values for all of Food

The total number of samples used in the plot above is 1638. A total of 5 samples lies outside the upper and lower control limits. Since these values only account for 1% of the total sample values, a conclusion can be made that there is very little variation present and that the process is in control.

Part 4- Optimising the delivery process:

The instances mentioned below are referring to the delivery time means of the sample groups of the data. From the following analysis it will be determined whether the delivery times of each class has to be investigated or not.

4.1 Analysis of the X-charts

The values which fall outside the control limits of the X- test are as follows:

Class	**Total**	**1st**	**2nd**	**3rd**	**3rd last**	**2nd last**	**last**
Luxury	434	142	171	184	789	790	791
Household	400	252	387	629	1335	1336	1337
Technology	17	37	398	483	1872	2009	2071
Gifts	2290	213	216	218	2604	2605	2606
Sweets	5	942	-	-	-	-	1403
Clothing	17	455	702	1152	1677	1723	1724
Food	5	75	-	-	-	-	1515

Indicated by the red circles in the X- charts for all the individual classes an inordinate number of samples lies outside of their respective upper and lower control limits resulting in a substantial variation within all the processes.

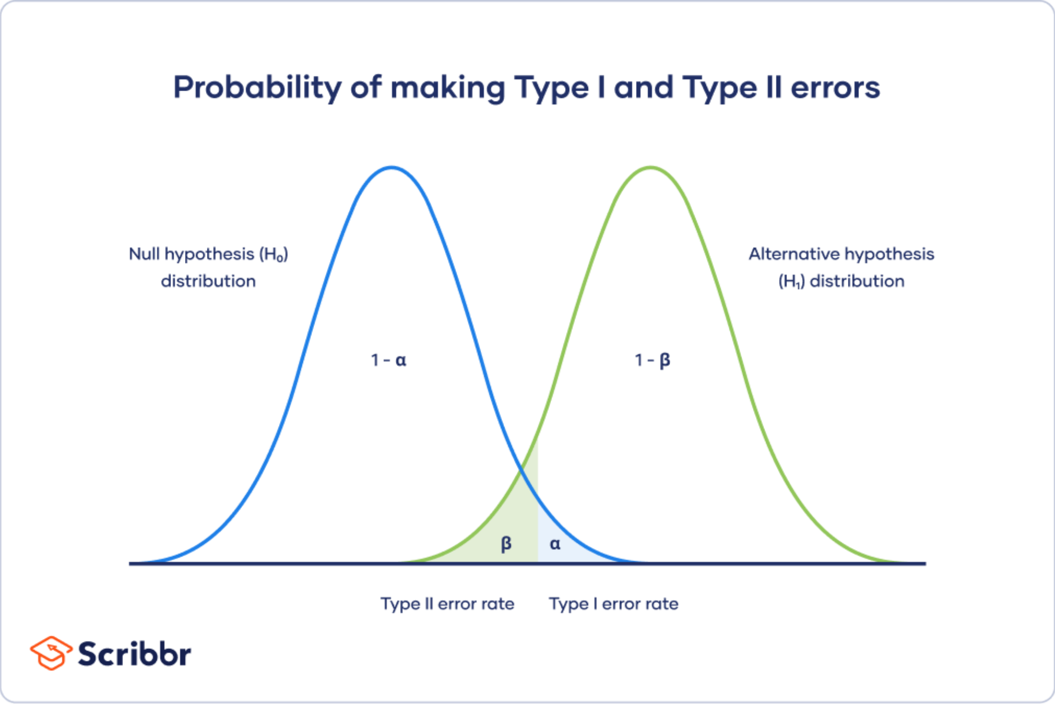
4.1.B Finding the most consecutive samples of standard deviations between -0.3 and +0.4 sigma-control limits

Class	Total samples within range	Last sample within range
Luxury	4	63
Household	4	46
Technology	6	372
Gifts	5	307
Sweets	4	94
Clothing	4	1013
Food	7	952

4.2 Estimated probability of making a Type I error

Hypothesis testing, is a type of statistical reasoning, using information from a sample to draw conclusions about a population parameter or a population probability distribution. First, an estimation is made on the parameter or distribution. The Null Hypothesis, denoted as H_0, is what is being assumed. The opposite of what is assumed in the null hypothesis is specified as the alternative hypothesis H_A. Using sample data, the hypothesis-testing technique determines whether H_0 may be rejected or not. A statistical conclusion is made that if the alternative hypothesis H_A is true, the H_0 is rejected.

In this case a sample of 30 instances is used to determine whether the process is in control or not. The null hypothesis is- H_0 : The process is in control. The alternative hypothesis is H_A : The process is not in control. The probability of making a Type I error is 0.27 %. This indicates that the odds of the business investigating a process which is stable are very low. This is a satisfying attribute as the business will not waste time on inspecting processes that are in control. (Sunlearn, 2022)


Probability of making Type I and Type II errors
Null hypothesis (H_0) distribution
Alternative hypothesis (H_1) distribution
1 - α
1 - β
β
α
Type II error rate
Type I error rate
Scribbr

(Type I & Type II Errors: Differences, Examples, Visualizations, 2022)

4.3 Best profit centre

In this subsection, the individual delivery times will be used and not the samples. When delivering items slower than 26 hours after the order or purchase is made, the loss will be R329 per item late in hours and it will cost R2.5 per item per hour to reduce the average time by one hour. An assumption will be made that the process distribution keeps the same shape when the centre is moved, and it will also be assumed that it costs less than R2.5 per item per hour if the delivery time is increased.

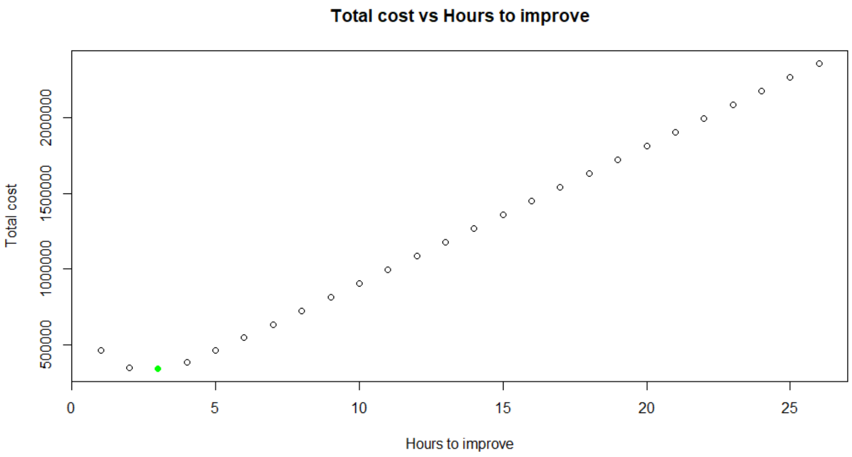


Figure 40: Best profit centre

From the graph above, the green dot indicates that the best profit center is 3 hours.

4.4 Estimated probability of making a Type II error

An estimation of the likelihood for making a type II error for A in the Technology class given that the delivery process average moves to 23 hours will be made. When making a type II error, the alternative hypothesis H_A is true, but can't be identified due to the sample being between the upper and lower control limits. The probability of making a type II error is 0.4956 which is a high probability in general. This indicates that the business should sharpen their approach to determine whether a process should be up for investigation as they fail to investigate half of the processes which are unstable as it is unknown to the business. (Sunlearn, 2022)

Part 5- MANOVA test:

A MANOVA test can be used to determine whether various levels of independent variables influence dependent variables. The independent variables can be separate or in combination with one another.

5.1.1 MANOVA 1

Dependent variables	Sales and Delivery time
Independent variables	Year
Null hypothesis (H0)	Sales and Delivery Time depend on the year in which sales are made
Alternative hypothesis (HA)	Sales and Delivery Time do not depend on the year in which sales are made
Significance level	0.05
Calculated p-value	2.2 e-16

Table 5.1 MANOVA 1

The calculated p-value is 2.2e-16 which is remarkably smaller than the significance level of 0.05. This indicates that the deviation from the null hypothesis is statistically significant, and the Null hypothesis should be rejected. A conclusion is made that Sales and Delivery Time do not depend on the year in which sales are made.

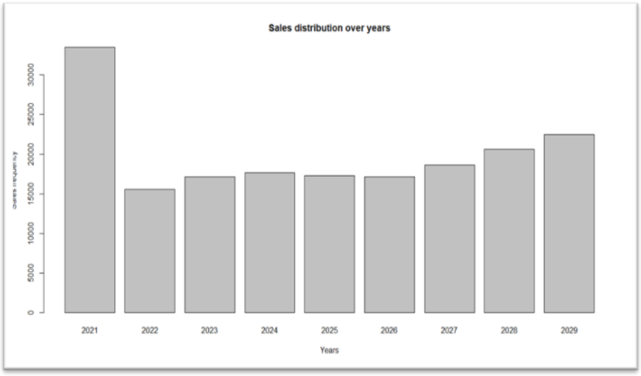


Figure 42: Sales distribution over years

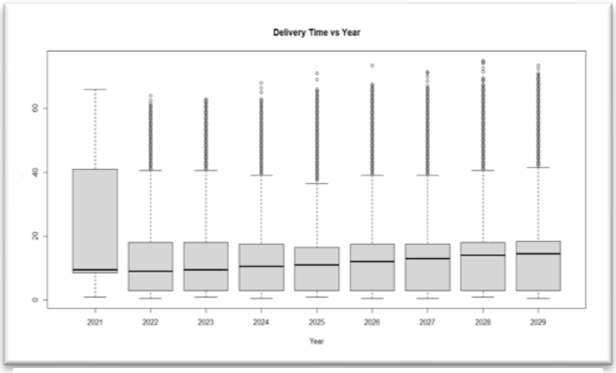


Figure 41: Delivery Time vs Year

5.1.2 MANOVA 2

Dependant variables	Price and Delivery Time
Independent variables	Class
Null hypothesis (H0)	Price and Delivery Time depend on the year in which sales are made
Alternative hypothesis (HA)	Price and Delivery Time do not depend on the year in which sales are made
Significance level	0.05
Calculated p-value	0.2787

Table 5.2 MANOVA 2

For the second MANOVA, the calculated p-value is 0.2787 which is greater than the significance level of 0.05. The Null hypothesis is not rejected since the p-value is greater than the significance level which presents that Price and Delivery Time depend on the year in which sales are made.

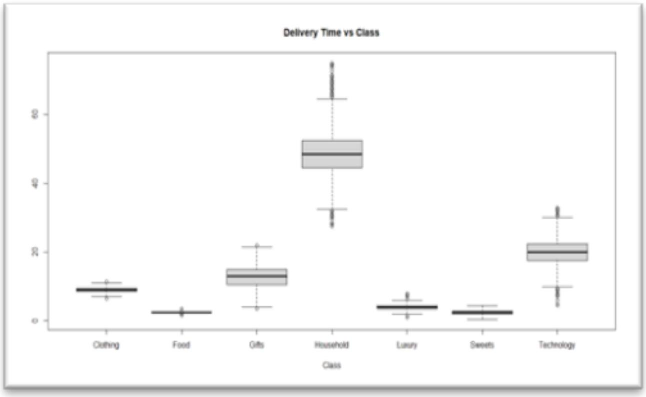


Figure 43: Delivery Time vs Class

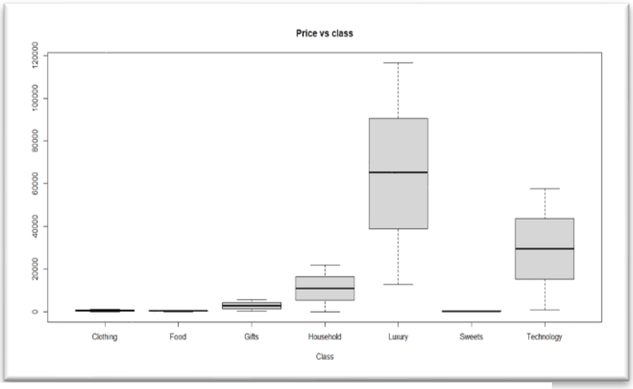


Figure 44: Price vs Class

Part 6- Reliability of service and products:

The reliability of service and products can be defined as the probability that a product, system or service will perform its intended function adequately for a specific period.

6.1.1 Problem 6

A team was assembled to investigate the refrigerator component at Cool Food, Inc. While undergoing their search for the underlying cause of scrap, they discovered a means to lower the cost of scrap to $35 per part.

Calculations:

Specification: 0.06 +- 0.04

$$k = T(x)/(\mathrm{x} - \mathrm{m})^2$$
$$= 35/(0.04)^{2}$$
$$= 21875$$

$$T(x) = 21875 \times (x - 0.06)^2$$

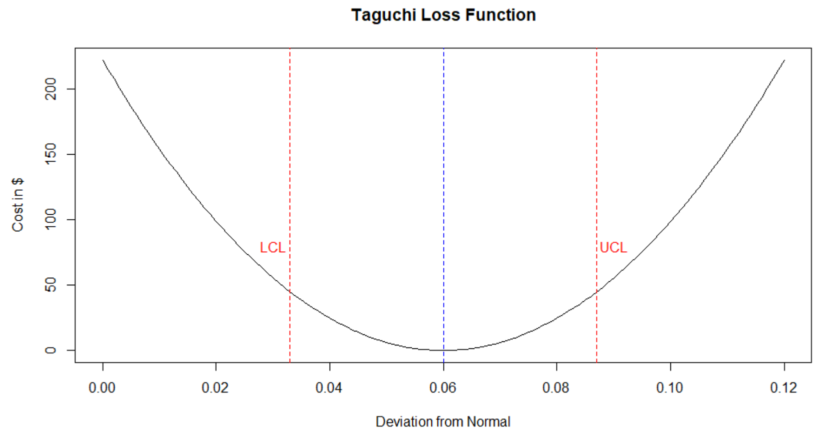


Figure 45: Taguchi Loss Function

The more a specific product's characteristic deviate from the target value 0.06, the quality of the product decreases and the cost to the business will increase. As a result, the products will be unreliable, and the service will be less effective.

6.1.2 Problem 7

At Cool Food, Inc., a blueprint specification for a refrigerator part's thickness is 0.06 +/- 0.04 centimetres. A part that doesn't fit the criteria must be scrapped, which costs $45.

Calculations

$\mathrm{k} = T(x)/(\mathrm{x} - \mathrm{m})^2$

$= 45/ (0.04)^2$

$= 28125$

$T(x) = 28125(\mathrm{x} - 0.06)^2$

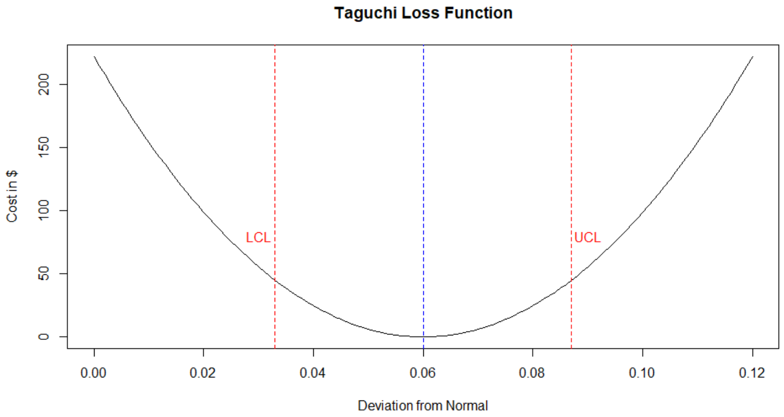


Figure 46: Taguchi Loss Function

a) Taguchi loss if process deviation from target is reduced to 0.027

T (0.027) = $28125(0.027 - 0.06)^2$

T (0.027) = \$30.63 per item

6.2 Problem 27

a) The probability of one machine at each stage

***Reliability** = Reliability(Machine A) × Reliability(Machine B) × Reliability(Machine C)*

***Reliability** =* 0.85 × 0.92 × 0.90 = 0.7038

b) Reliability if both machines are used

***Reliability** = Reliability(A1 & A2) × Reliability(B1 & B2) × Reliability(C1 & C2)*

***Reliability** =* $(1 - (1 - 0.85)^2) \times (1 - (1 - 0.92)^2) \times (1 - (1 - 0.90)^2) = 0.9779$

By putting 2 identical machines parallel to each other will improve the reliability with 27%. This is so that if one machine fails, a parallel version of the same machine can continue to function. Running the two units simultaneously would increase reliability for the business.

6.3 Problem 27

For the delivery process, there are 20 delivery vehicles available, of which 19 is required to be operating at any time to give reliable service. During the past 1560 days, the number of days that there were only 20 vehicles available was 190 days, only 19 vehicles available was 22 days, only 18 vehicles available was 3 days and 17 vehicles available only once. There are also 21 drivers, who each work an 8-hour shift per day. During the past 1560 days, the number of days that there were only 20 drivers available was 95 days, only 19 drivers available was 6 days and only 18 drivers available, once only.

The problem will be analysed using the binomial distribution. The probability of vehicles being available will be determined and combined with the probability of reliable delivery time.

Fraction of how many vehicles are not available:

$F0\_vehichle\_NA = 0.8615385$

$F1\_vehicle\_NA = 0.1217949$

$F2\_vehicles\_NA = 0.01410256$

$F3\_vehicles\_NA = 0.001923077$

$F4\_vehicles\_NA = 0.0006410256$

Probability of vehicles available:

$P0\_vehicles\_NA =$ 0.007071661

$P1\_vehicle\_NA =$ 0.006621817

$P2\_vehicles\_NA =$ 0.00892079

$P3\_vehicles\_NA =$ 0.01217169

$P4\_vehicles\_NA =$ 0.01968774

Fraction of how many drivers not available:

$F0\_drivers\_NA = 0.9346154$

$F1\_driver\_NA = 0.06089744$

$F2\_drivers\_NA = 0.003846154$

$F3\_drivers\_NA = 0.0006410256$

Probability of drivers available:

$P0\_drivers\_NA =$ 0.003224402

$P1\_driver\_NA =$ 0.003084515

$P2\_drivers\_NA =$ 0.00447595

$P3\_drivers\_NA =$ 0.008232228

Number of days having vehicles available:

# days having 21 vehicles available: 1344.312

# days having 20 vehicles available: 200.749

Vehicle and driver reliability: 0.9883481

Expected number of days for reliable delivery per year: 360.7471

Conclusion:

The report provides insight on the sales data provided. Trends in retail and purchasing have been identified and it is confirmed that the class of items bought are dependent on numerous factors included in the report. There are tendencies about several customer class groups purchasing certain products, such as the age of a customer. By using statistical control tests, problems regarding the products and processes could be identified and investigated. More insight about which factors must be investigated is also included.

The probability of making different statistical errors were also investigated and the conclusion is made that making a type II error is much more likely than making a type I error and that the business should do further investigation on the type II error.

Ultimately, the value of explorative analysis is used and will help with the overall performance of the business if implemented correctly.

References

(2022, October 9). Retrieved from SlideServe: https://www.slideserve.com/pelwood/process-capability-cp-cpk-pp-ppk-global-training-material-powerpoint-ppt-presentation

(2022, October 12). Retrieved from ASQ: https://asq.org/quality-resources/statistical-process-control#:~:text=Statistical%20process%20control%20(SPC)%20is,find%20solutions%20for%20production%20issues.

(2022, October 15). Retrieved from Sunlearn: https://learn.sun.ac.za/pluginfile.php/3514418/mod_resource/content/1/QA344%20Statistics.pdf

Type I & Type II Errors: Differences, Examples, Visualizations. (2022, October 21). Retrieved from https://www.scribbr.com/statistics/type-i-and-type-ii-errors/