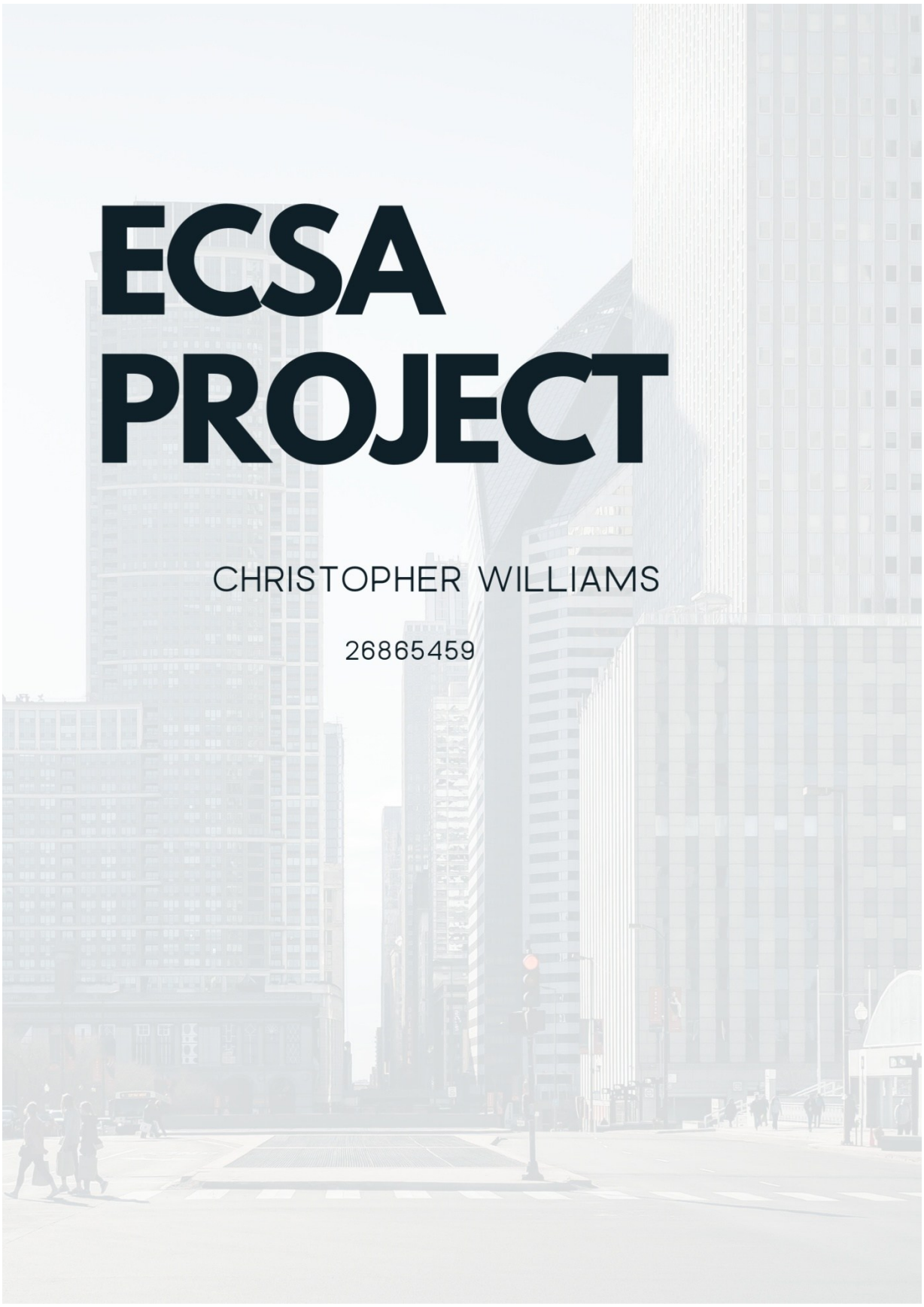


ECSA PROJECT

CHRISTOPHER WILLIAMS

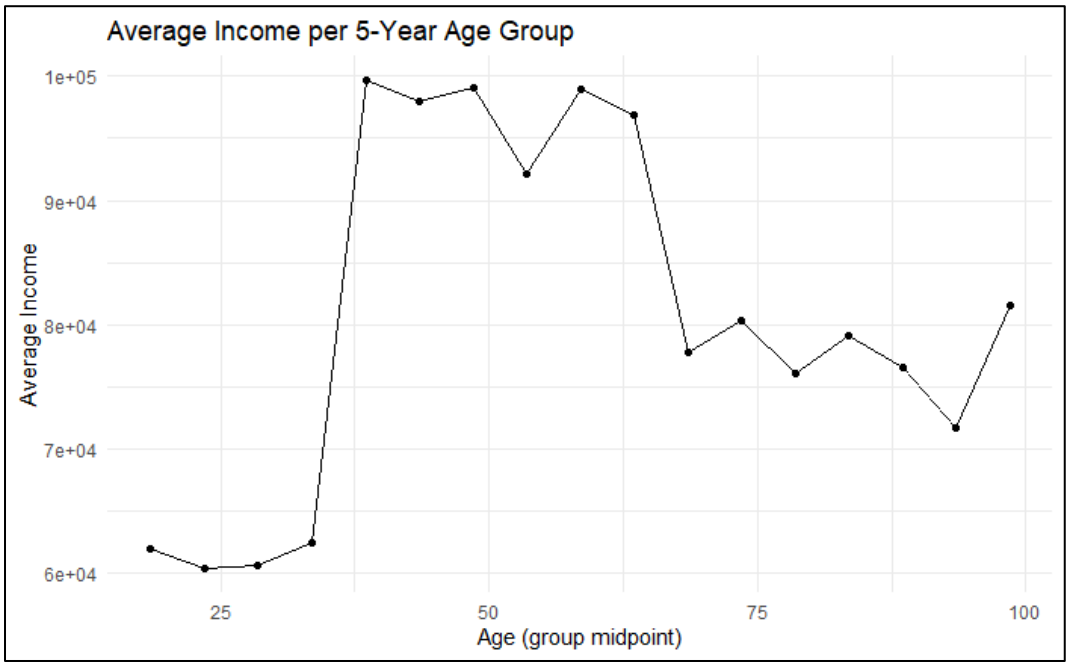
26865459

Table of Contents

1-2. Descriptive statistics on sales, products, and customers data

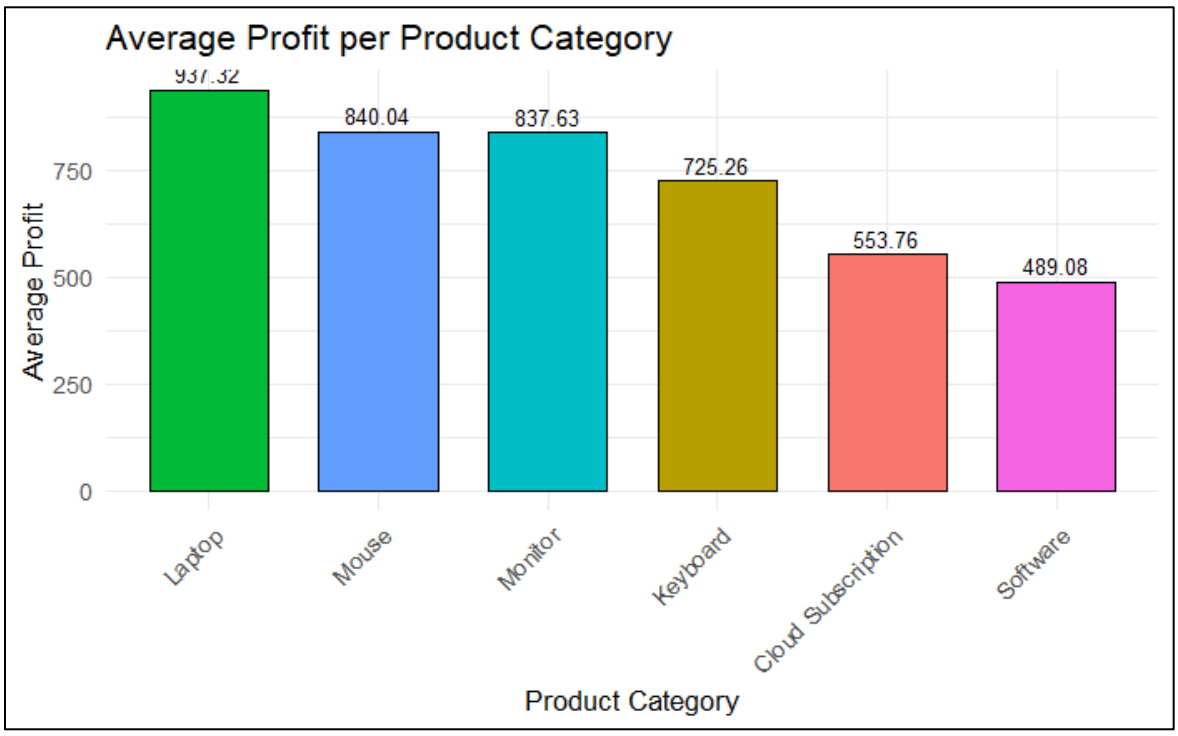
City <chr>	Average_Income <dbl>
Chicago	82244.48
Houston	80248.62
Los Angeles	80475.21
Miami	83346.21
New York	79752.07
San Francisco	79852.56
Seattle	79947.99

Above is an attachment of the average income of each city as derived from the customer income data. It is evident that Miami has the highest average income. Although the data does not vary a lot, a higher income average is a good indication of higher spending ability, and could be used by the company as a decision element as to where expansion could possibly be beneficial (where there is higher customer spending possibility).

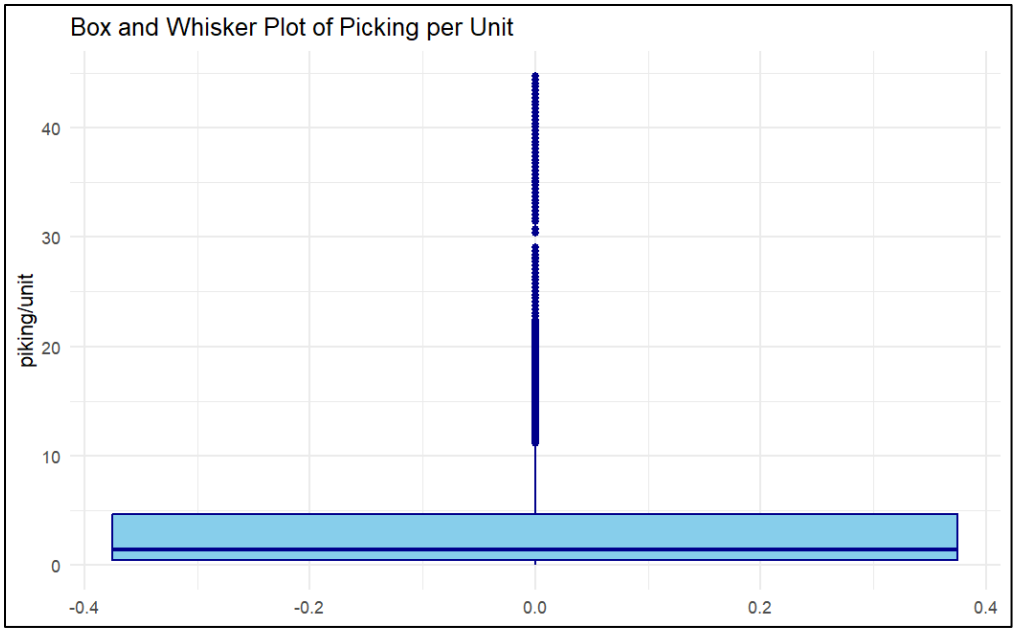


The above line graph indicates the average income per age group (groups of 5 years). As can be seen the average income of age groups between 30

and 60 have the highest average income. A higher income directly relates to higher spending ability. This can be used by the company to determine their focus market, and that can direct decisions such as which type of advertising and which products to sell. Because some models of advertising serve some age groups better than others and some products are bought more by some age groups.



The above graph shows the average profit generated by each product category. This was calculated from the products data given. Laptops generate the highest profits in comparison to the other product categories. This data can however not be used in isolation, and other data such as number of sales should be used in conjunction to determine which products generate the most profit for the company. But knowing that laptops generate the highest profit per sale, it will be good to maximise laptop sales to generate high profit.



The above box and whisker diagram shows the distribution of picking times per unit. This was calculated by dividing the order picking time by the number of units. The mean value of picking times per product is 3.86 hours. This number should be assessed by the company to determine if it is efficient or not depending on the procedures of picking a specific unit. But the plot shows that there is a large spread of the data with picking hours per unit stretching up to 45 hours per unit. These picking activities should be assessed carefully to determine the cause of the possible hold up time. Lowering the picking time of orders will have a positive effect on the efficiency of the company and can increase profits.

3. Statistical process control

The initialisation process was done using sample sizes of 24, and the first 30 samples were used to calculate all the control limits and the centre line.

	UCL	Pos2sig	Pos1sig	CL	Neg1sig	Neg2sig	LCL
1	22.762450	21.591254	20.420057	19.2488611	18.0776649	16.9064686	15.735272
2	22.819707	21.611138	20.402569	19.1940000	17.9854311	16.7768621	15.568293
3	1.139702	1.078347	1.016992	0.9556375	0.8942827	0.8329278	0.771573
4	22.782828	21.563867	20.344906	19.1259444	17.9069832	16.6880220	15.469061
5	23.170076	21.954671	20.739266	19.5238611	18.3084562	17.0930514	15.877647
6	23.092376	21.870232	20.648088	19.4259444	18.2038007	16.9816570	15.759513

The above table shows the x-bar (sample mean) centre line, control limits and +-1sigma and +2sigma limits for each product category. The categories are numbered and labelled as follows.

Product category numbering

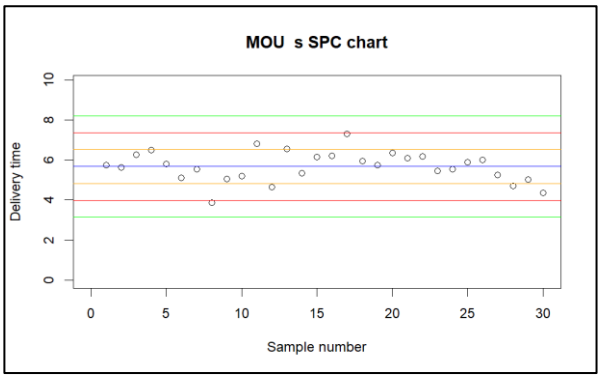
1 MOU - Mouse
2 KEY - Keyboard
3 SOF - Software
4 CLO - Cloud Subscription
5 LAP - Laptop
6 MON - Monitors

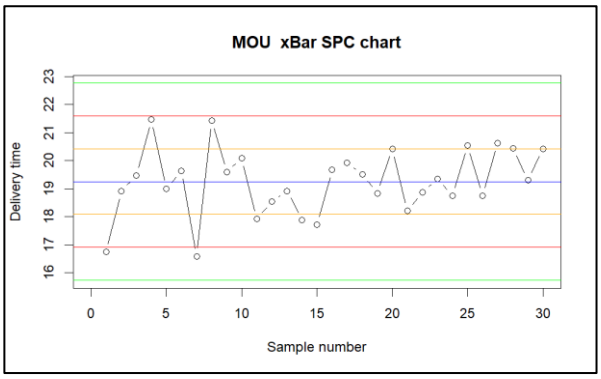
	UCL	Pos2sig	Pos1sig	CL	Neg1sig	Neg2sig	LCL
1	8.2021578	7.3601831	6.518208	5.6762338	4.8342591	3.9922844	3.1503097
2	8.4638875	7.5950456	6.726204	5.8573616	4.9885197	4.1196777	3.2508357
3	0.4296822	0.3855741	0.341466	0.2973579	0.2532498	0.2091417	0.1650336
4	8.5366672	7.6603542	6.784041	5.9077281	5.0314151	4.1551021	3.2787891
5	8.5117611	7.6380047	6.764248	5.8904921	5.0167358	4.1429794	3.2692231
6	8.5589548	7.6803539	6.801753	5.9231521	5.0445512	4.1659503	3.2873494

The above diagram shows all the s (sample standard deviation) control limits for each product category. The categories are numbered and labelled the same as the above diagram.

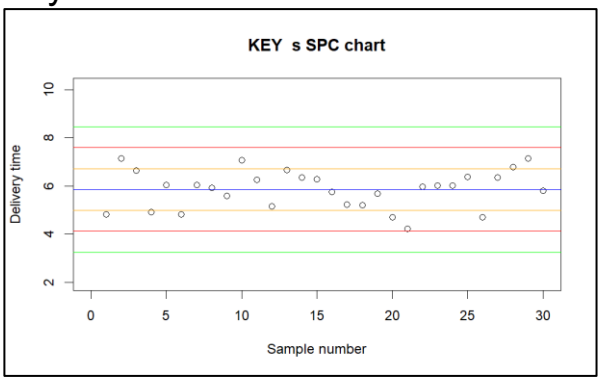
These values are used to control the output of a system, visually depicted on SPC graphs. Certain rules are set up to flag possible dysfunctional system processes.

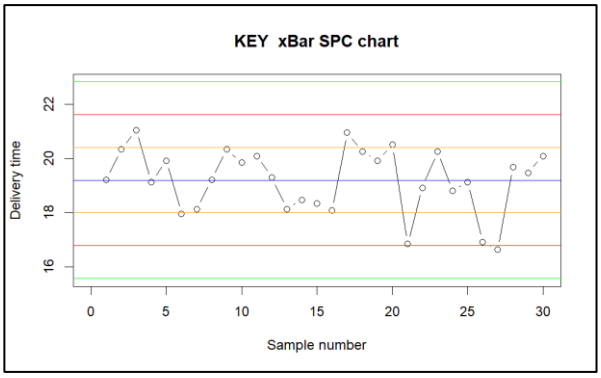
Mouse:



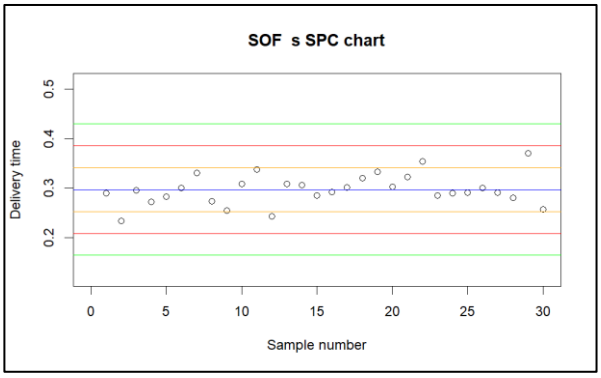


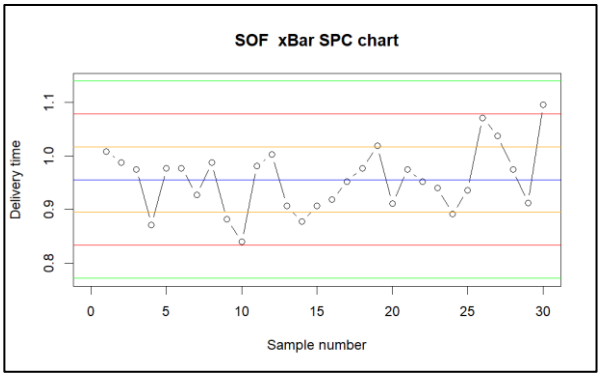
Keyboards:



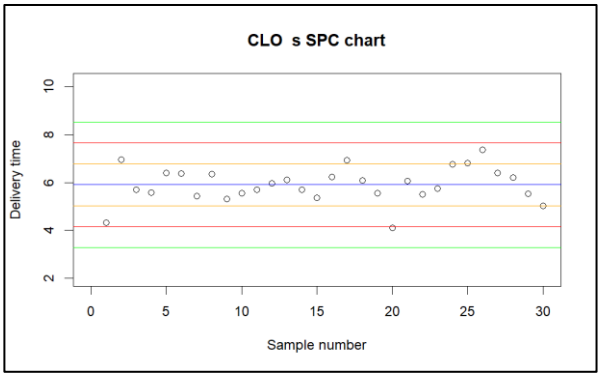


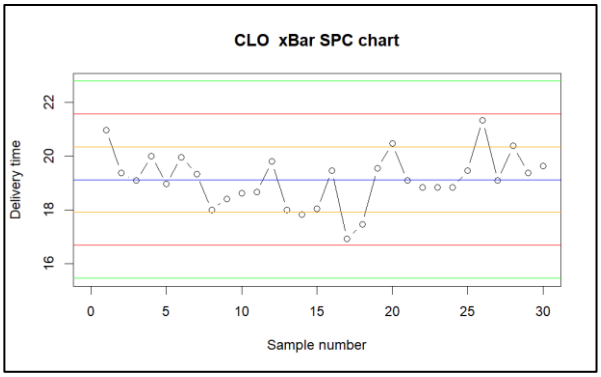
Software:



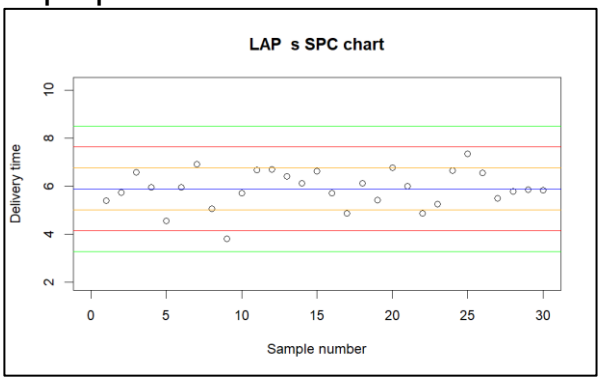


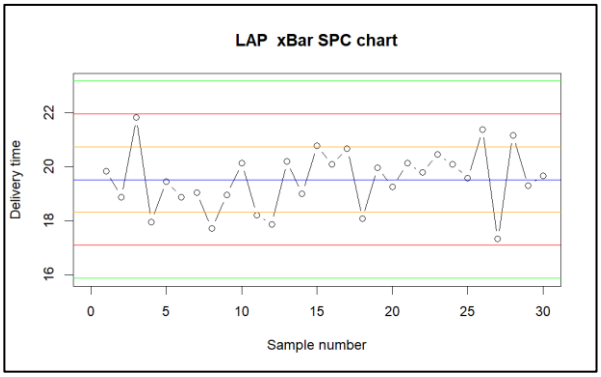
Cloud Subscription:



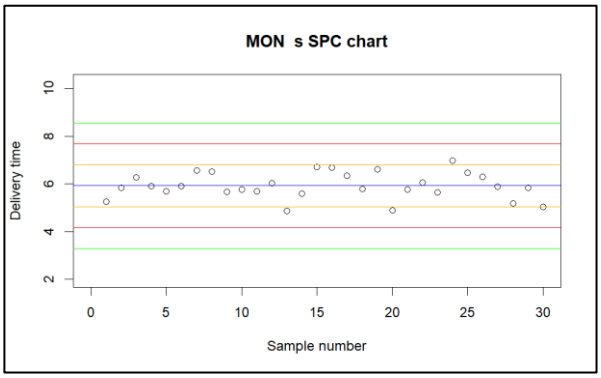


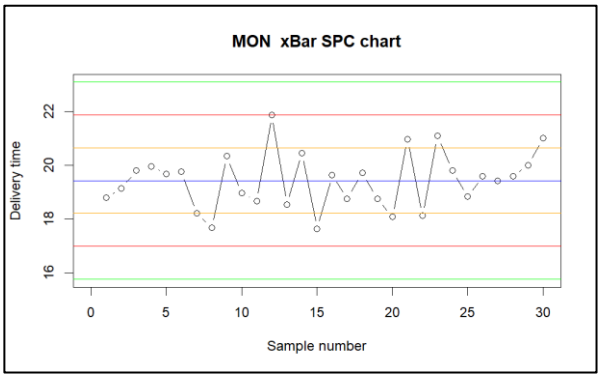
Laptop:





Monitor:





Above are the SPC graphs for the product category MOU (mouse). Figure 5 is the sample standard deviation graph. The control limit lines are colour coded:

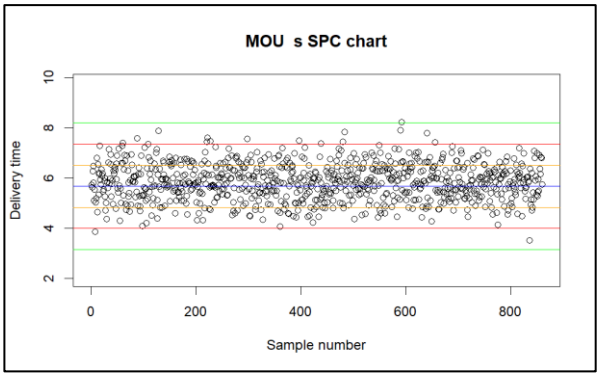
- Blue: centre line
- orange: +- 1sigma from the centre line
- red: +-2sigma from centre line
- green the upper and lower control limits

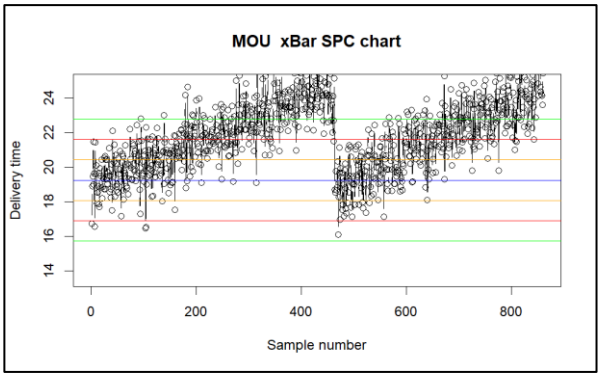
As can be seen, there are no instances outside the outer control limits for the first 30 samples, making it viable for the initialisation process.

Now, further samples are taken and plotted for process control, and is used by managers and operators to flag possible predicted errors.

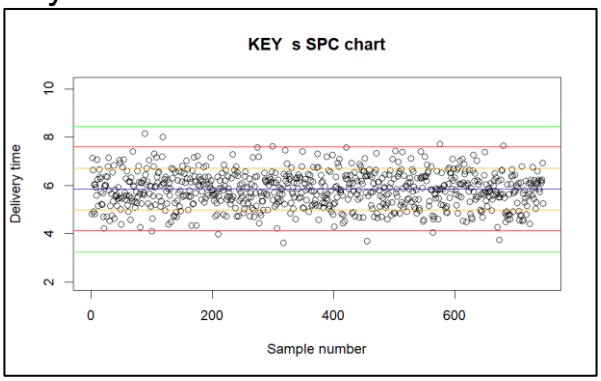
SPC charts for all samples per product category:

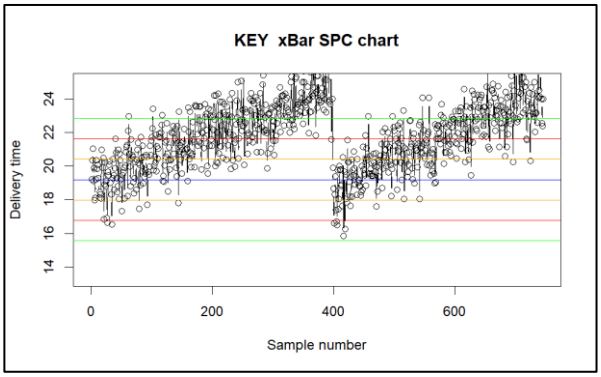
Mouse:



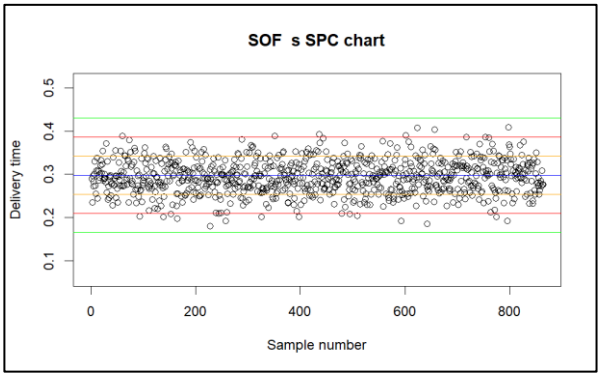


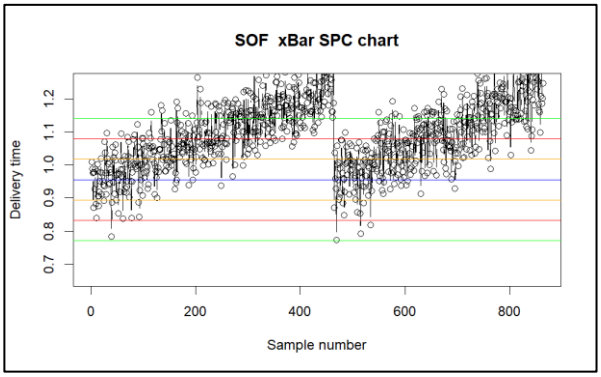
Keyboard:



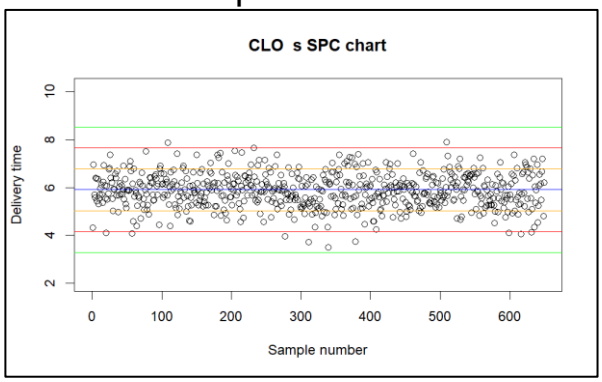


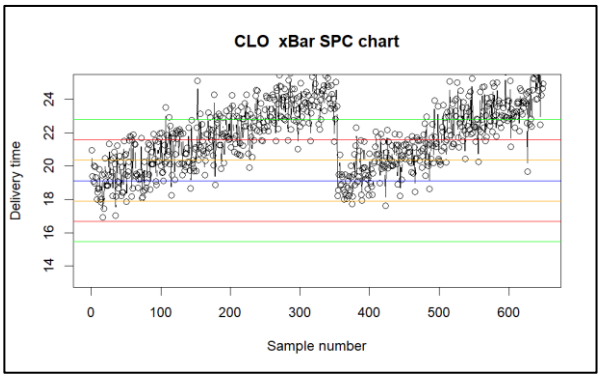
Software:



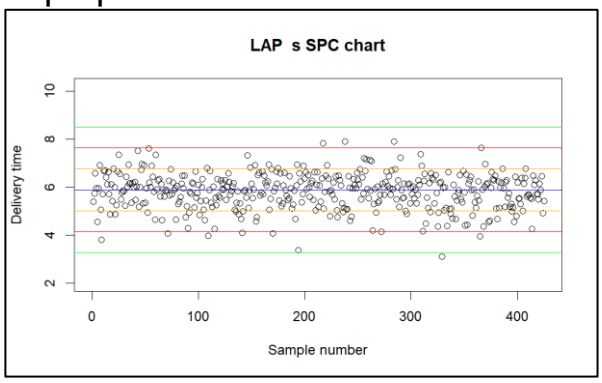


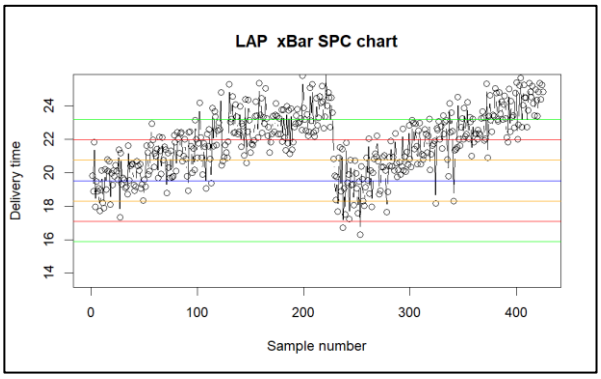
Cloud Subscription:



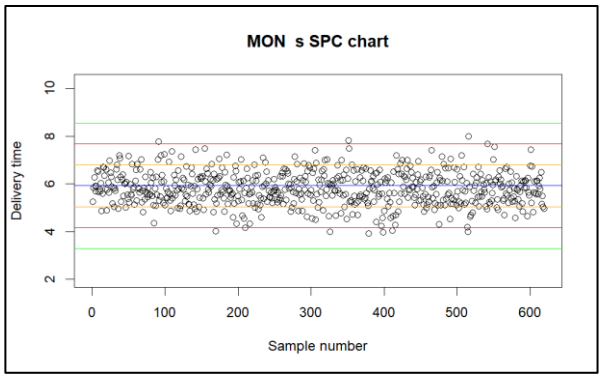


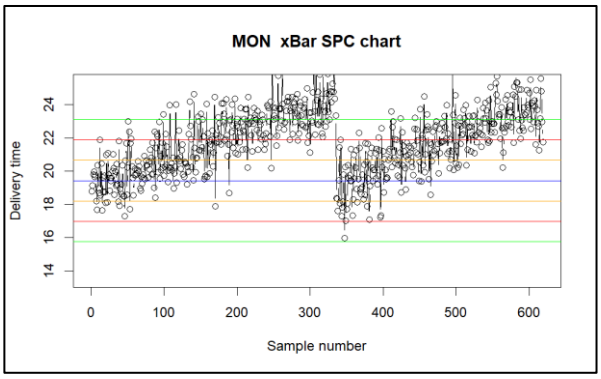
Laptop:





Monitor:





Above are the SPC charts for all the product categories. We first look at the s-charts. The s-charts show no clear deviation from the centre line. This means that the spread of the samples show no clear problem and the xbar-charts can be analysed normally. For all the products, the sample means form a similar pattern where sample means seem to deviate from the centre line up to a peak at just before the 400th sample, and then returns to

normal. Because this is past information, it can be an indication that the management team realised a possible error as the sample means started to deviate, identified the problem and corrected it, returning the process back to optimal functionality.

But the pattern then repeats, the company should investigate this pattern. It could possibly be that delivery drivers “slack” from time to time, and once disciplinary action is taken, they operate more effectively. A more strict driver control system might have to be implemented.

Process (Product) capability

	Cp	Cpu	Cpl	Cpk
1	0.8693094	0.5549161	1.183703	0.5549161
2	0.8754917	0.5613367	1.189647	0.5613367
3	17.3266425	33.4740492	1.179236	1.1792359
4	0.8721414	0.5605899	1.183693	0.5605899
5	0.8816744	0.5631485	1.200200	0.5631485
6	0.8819683	0.5657164	1.198220	0.5657164

The above diagram shows the capability measures for each product category. (Product categories will have the same numbering system throughout the report)

The Cp is the Capability potential. It measures how well the process spread fits into the customer specification range. A Cp of >1.3 is desired by most companies (Quality-One, 2020). This means that all product categories except for category 3 (Software) is below the desired capability potential.

The Cpu is the Upper capability index. It measures how close the process mean is to the upper specification limit (USL) relative to the process variation. A Cpu >1 indicates that sufficient space between the USL and centre line for the process to stay within the limit. Once again product category 3 (Software) is the only category with a desired upper capability index value.

The Cpl is the lower capability index. Similar to the Cpu, it measures how close the process mean is to the lower specification limit (USL) relative to

the process variation. Also with desired values >1. Here all the products categories have desired values of above 1, and there is little variation between the different values. This can simply mean that the lower control limit is at a minimum possible delivery value of 0 hours, therefor all category deliveries have sufficient space to always be above the the lower control limit.

The Cpk is the Capability index and measures the actual capabilities of a process by measuring weather the process mean falls within the specification limits. A Cpk > 1 indicates that a process is capable and is producing within specification limits with a safety margin.

From the above capability measures we can see that product category 3 (software) is capable of meeting customer demands in terms of delivery time.

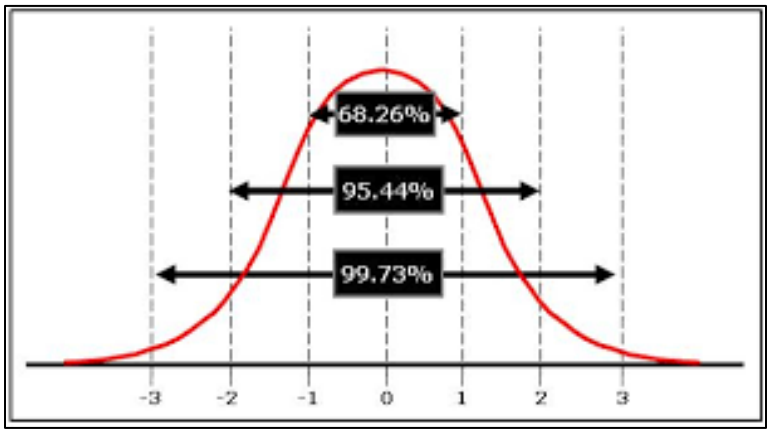
Process control with specific flag rules:

Rule A:

```
[[1]]
[1] 592

[[2]]
integer(0)

[[3]]
integer(0)

[[4]]
integer(0)

[[5]]
integer(0)

[[6]]
integer(0)
```

The attached picture is a screenshot from the r-workspace that indicates for each product category which sample s is above the UCL. And for category 1 (Mouse) the only sample above the UCL is sample 592.

4. Risk, Data correction and optimising profit



The above graph depicts an in control process normal distribution. It shows the probability percentages of an instances being within the referred to control limit lines.

- Within UCL and LCL = 99.73%
- Within 2sigma lines = 95.44%
- Within 1sigma lines = 68.26%

This will be used to determine the probability of making type I and II errors with regard to the above rules.

Type I error calculation: (In control system, error wrongly flagged)

- Rule A

For a in control system, a instance has a probability of 0.28% to fall outside the upper or lower control limit. Rule A only flags if the instance is above the upper control limit, meaning half the probability, 0.14%.

This means that the probability of a single sample standard deviation falling above the UCL in an in-control system is 0.14%, which is very small.

We can conclude that due to the small probability of this happening, if it happens, it most probably indicates some kind of error. Therefor making a type I error is small.

- Rule B

The probability of a single instance falling within the 1sigma limit lines is 68.26%. If more instances fall within the 1sigma limits, it indicates that the system spread is limited and under control.

Taking that into account, a possible out of control mean value is most probably due to a system issue, and unlikely a case of a wide spread of data.

Concluding that the probability of making a type I error is small.

- Rule C

The probability of a instance being above the 2sigma control limit line is 2.28%. Calculated by:

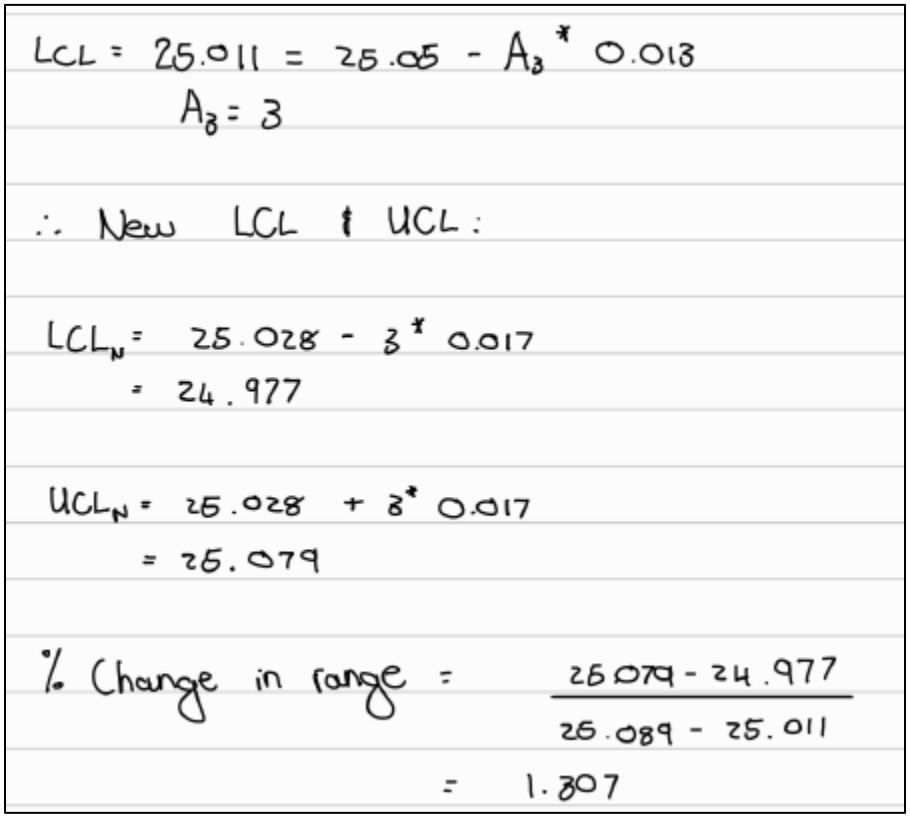
$$2.28 = (100 - 95.44)/2$$

For consecutive instances falling above the control line, we take the component to the power of the number of consecutive instances. Therefor:

$$0.02228^4 = 0.00000027$$

Which is a very small percentage. This is the probability of 4 consecutive sample means of a in control system being above the 2sigma limit. This means that if it occurs, it is very unlikely, and therefor highly possible of indicating a system error.

Type II error calculation:



Above is the calculation to determine the change in control range. As can be seen by the final answer of 1.3, there is a 30% range size increase in control range. This means that there is a 30% increase in possibly “missing” an error flag outside of the control limits.

Data correction

ProductID	Category	Description	SellingPrice	Markup
SOF001	Software	coral silk	511.53	25.05
SOF002	Software	black silk	505.26	10.43
SOF003	Software	burlywood marble	493.69	16.18
SOF004	Software	black marble	542.56	17.19
SOF005	Software	chartreuse sandpaper	516.15	11.01
SOF006	Software	cornflowerblue marble	478.93	16.99
SOF007	Software	blue marble	527.56	16.79
SOF008	Software	cornflowerblue marble	549.02	11.95
SOF009	Software	black bright	540.41	11.34
SOF010	Software	cornflowerblue matt	396.72	23.47
SOF011	Software	aliceblue silk	511.53	25.05
SOF012	Software	coral marble	505.26	10.43

ProductID	Category	Description	SellingPrice	Markup
CLO004	Cloud Subscription	chocolate marble	1083.11	21.25
CLO005	Cloud Subscription	chocolate marble	728.26	27.7
CLO006	Cloud Subscription	aliceblue bright	959.51	19.55
CLO007	Cloud Subscription	blueviolet marble	991.81	18.87
CLO008	Cloud Subscription	burlywood silk	1105.66	20.23
CLO009	Cloud Subscription	black sandpaper	1092.07	23.14
CLO010	Cloud Subscription	chocolate marble	1128.98	25.48
CLO011	Cloud Subscription	aliceblue silk	1070.54	16.41
CLO012	Cloud Subscription	burlywood silk	963.14	10.13
CLO013	Cloud Subscription	blueviolet silk	1067.54	16.8
CLO014	Cloud Subscription	azure silk	1083.11	21.25

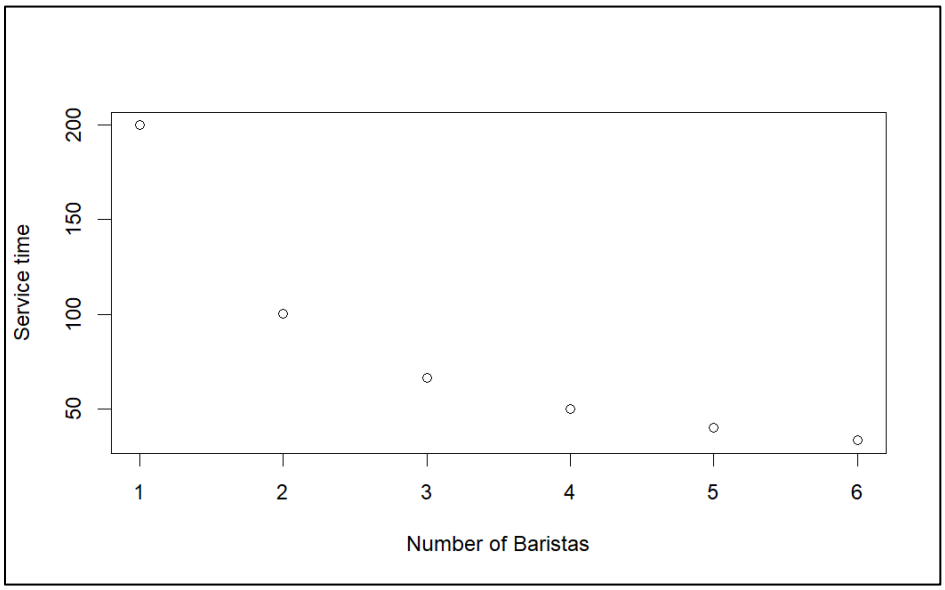
Above are two extractions from the improved and corrected products_Headoffice2025.csv file. The first 10 selling price and markup values extracted from products_data.csv is repeated for each category, and productID is also corrected to match the product category.

Using the corrected datasets the total sales value of 2023 is:

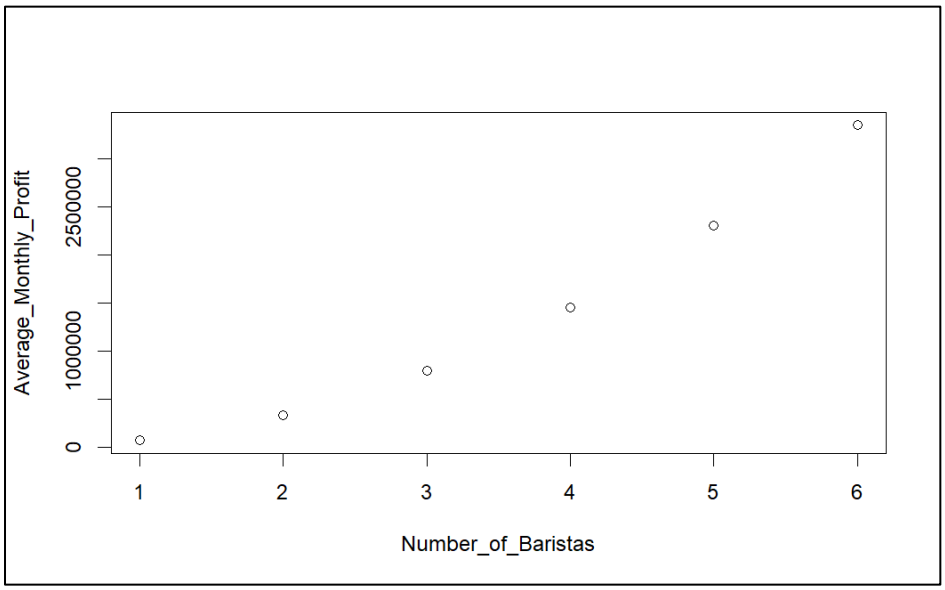
```
[1] "Total Revenue: 4352587677.92"
```

5. Barista profit optimisation

Shop 1



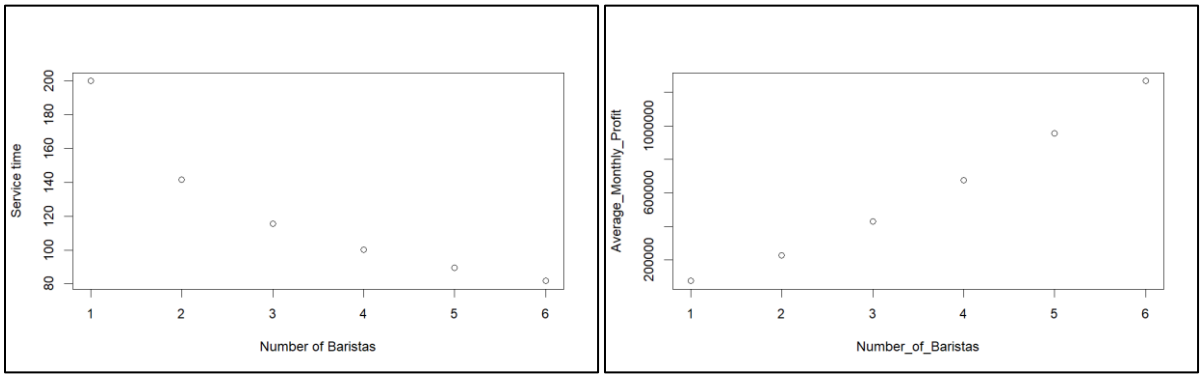
As can be seen in the above diagram, the more baristas there are, the shorter the average service time per customer order. The service time plateaus toward the minimum value.



Taken that R30 profit is made from every customer helped, and that the appointment of each barista costs R1000 per day, the above graph depicts the monthly profit made per number of baristas appointed.
As can be seen, the monthly profit made elevates the more baristas are

appointed. This is mainly due to the service times that are faster, and more customers can be served.
It can be assumed that the more baristas are appointed, that profits would grow even more, but it is said that the maximum amount of baristas that are allowed is 6. So for optimal profit, 6 baristas must be appointed.

Shop 2



As can be seen for shop 2, a similar result is achieved. Profit grows the more baristas are appointed. Although for shop 2 a more exponential growth in profit as baristas increase is achieved than for shop 1.

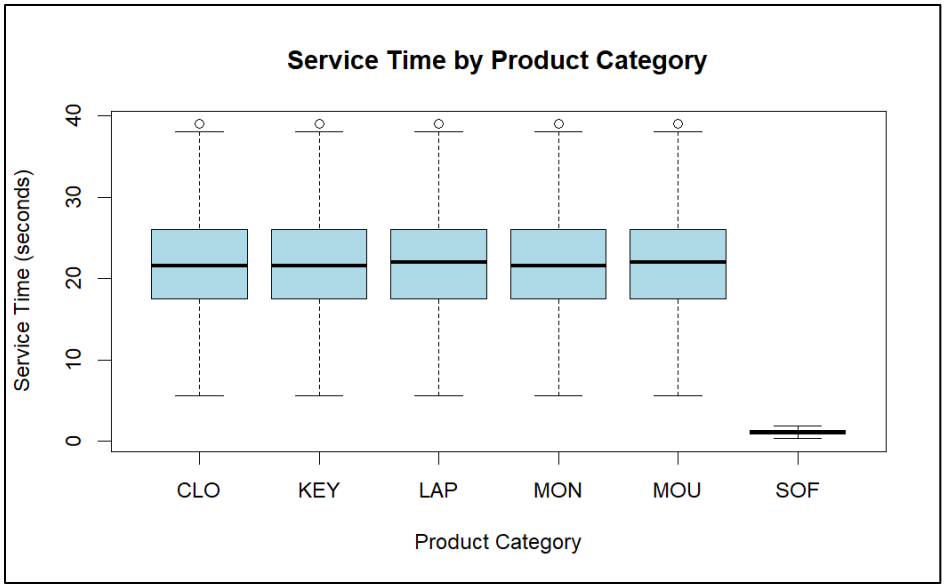
6. ANOVA analysis

The null hypothesis (H0) is:

That the product category has no effect on the delivery time of the product.

Alternative hypothesis (H1):

At least one product category has a different mean delivery time value.

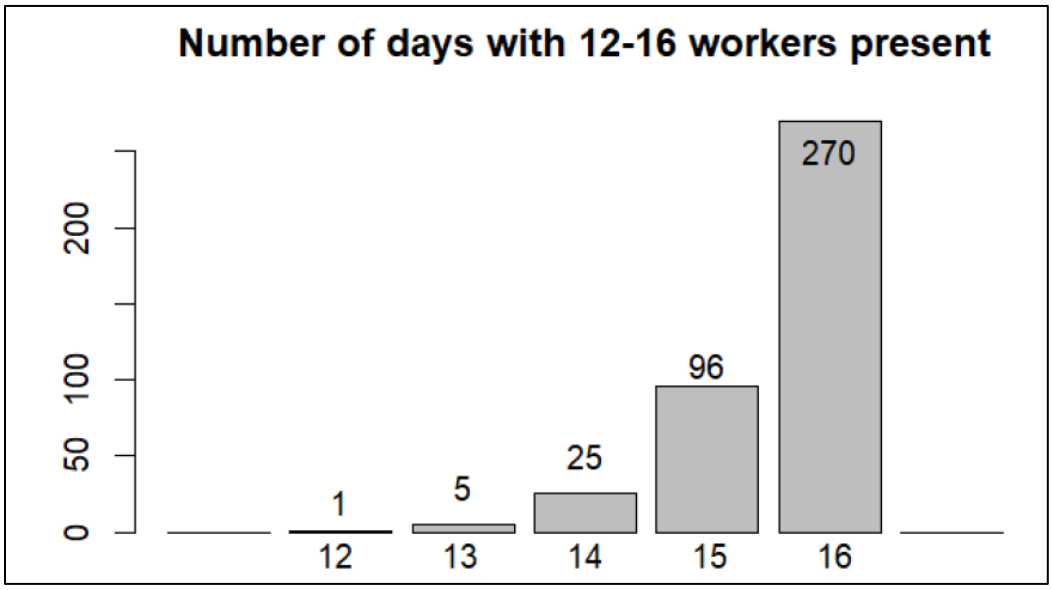


As can be seen in the above box and whisker plot diagram, all the product categories have about the same delivery time mean, except for software. It has a much lower mean delivery time.

```
              Df  Sum Sq Mean Sq F value Pr(>F)
Category       5 7022905 1404581   47694 <2e-16 ***
Residuals  99994 2944793      29
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1
```

The above results from the R code reflects a p-value of 2e-16 which is below 0.05. This further supports that the null hypothesis is rejected. Meaning that not all the product categories have the same delivery time. Software products have a much lower delivery time compared to the rest.

7. Reliability of service

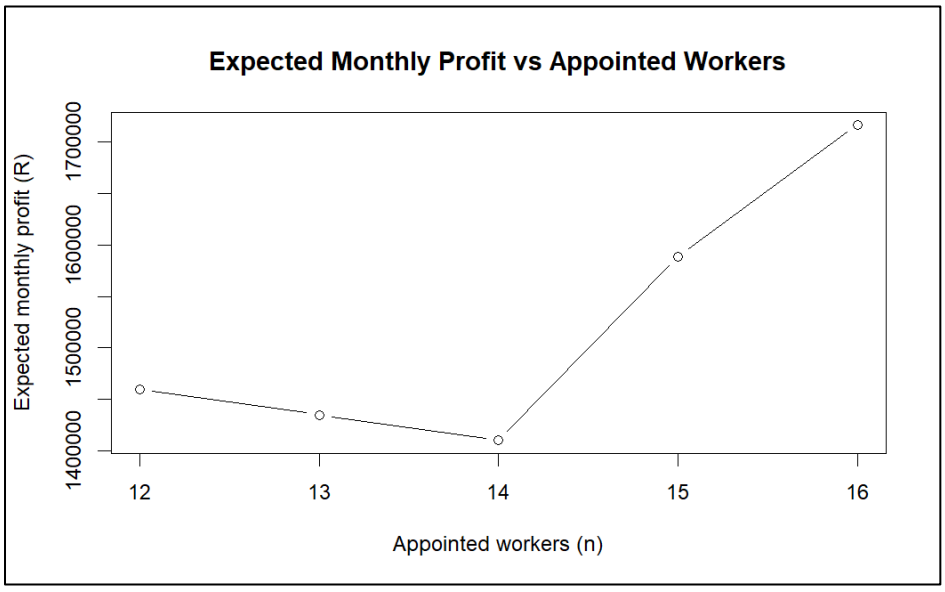


We can assume that a reliable service is delivered when there are 15 or more workers present per day. Therefor we can do our calculation as follows:

$$\frac{270 + 96}{397} = 0.9219$$

$$0.9219 \times 365 = 336.5$$

From the above calculations, it can be concluded that about 336 days out of 365 reliable service can be expected by clients.



According to the given information, the above graph depicts the monthly profit generated according to the number of workers appointed. It was assumed that normal daily profit is R100000 per worker. Up to 14 workers, there is a loss in profit due to understaffing. The graph shows that 16 workers on duty delivers the highest profit within the range of staff. Therefor the company should appoint 16 staff members to optimise profit.

This upward trend is because we assume that the cost per worker is less than the profit they generate.

Bibliography

Quality-One, 2020. *Process Capability.* [Online]
[Accessed 2025].