ECSA Project: Quality Assurance 344

Hough Joubert

26970678

24 October 2025

Contents

Introduction

In this project we will be focusing on data analysis and data manipulation. The goal of the project is to evaluate our ability to produce a structured report, perform efficient data wrangling and applying statistical methods to solve real-world problems. The key tasks that will be performed are data analysis, descriptive statistics, statistical process control (SPC), process capability analysis, risk assessment, data correction, profit optimization and advanced statistical modelling using ANOVA.

Phase 1

Introduction

In this phase of the project, four Excel datasheets are imported into R studio to perform data analysis. Since the previous analyst left no documentation or notes, the project starts with a fresh, independent exploration of the data. The primary objectives are data loading and inspection, data cleaning, performing data analysis and reporting the findings.

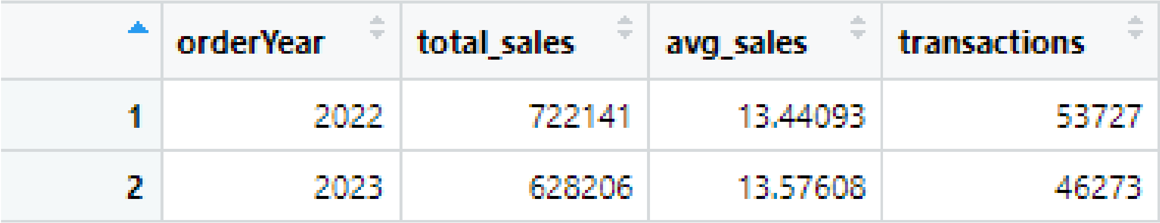
Data loading and inspection

When loading data from an Excel sheet into R, there is a possibility of Data Mismatches happening. Examples of these mismatches are when numeric values are converted to characters, or when dates that are stored in Excel's internal date system is changed to numeric values. Hidden rows or columns might also still be imported from Excel. There is a risk that formulas used to calculate the values of features can be read as a text and does not display the values. The best practice to prevent these errors is to convert the Excel files to .csv files. The read.csv function can handle larger datasets and is faster than packages such as readxl or openxlsx.

An important part of the data inspection is testing for missing values. If a missing value is found in the data it is advised to remove that instance from the data, as it may lead to biased results, the misinterpretation of trends and a reduced statistical power. The following steps were taken to test for any missing values. Firstly, there was checked if there are any columns that contains any missing values. After performing this operation on each of the four data sheets, it was found that no columns contained any missing values. Therefore, it is not necessary to perform operations such as removing the rows with missing values. After knowing that there are no missing values, we can perform the data analysis.

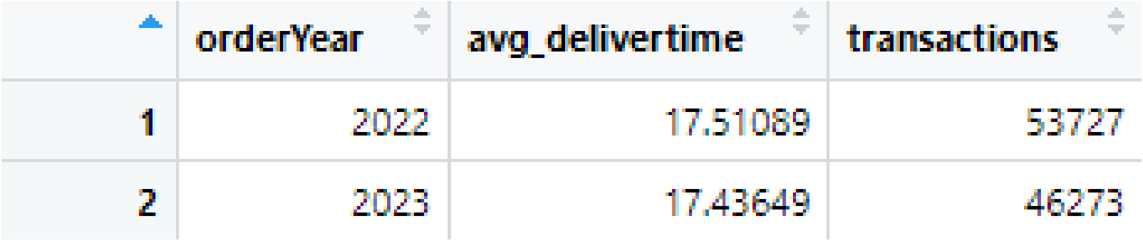
Data Analysis and findings

Sales 2022 and 2023



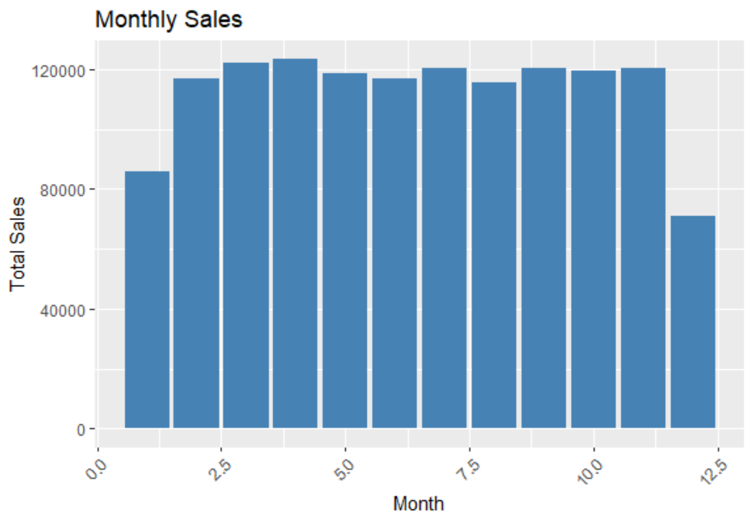
	orderYear	total_sales	avg_sales	transactions
1	2022	722141	13.44093	53727
2	2023	628206	13.57608	46273

In 2022 the total sales were 722 141 products with 53 727 different transactions. This accumulated to an average of 13.44 units sold per transaction. In 2023 the company experienced a major decline in the number of sales. The number of sales declined to 628 206 with a total of 46 273 transactions. This shows a slight increase in the average number of units sold per transaction. By interpreting these statistics, it can be found that either a new competitor entered the market which led to a decrease in customers, or due to the change in the economic circumstances people bought less products.



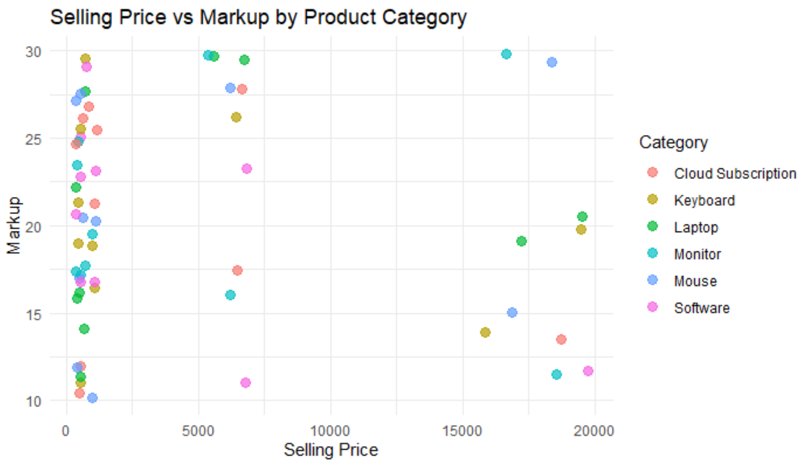
	orderYear	avg_delivertime	transactions
1	2022	17.51089	53727
2	2023	17.43649	46273

The following statistic should be worrying for the company. Although there was a significant decrease in the number of orders received, the average delivery time only decreased slightly. This shows that the productivity decreased along with the number of orders received. This can possibly be due to the company firing employees due to the loss in income as result of the lower number of orders.

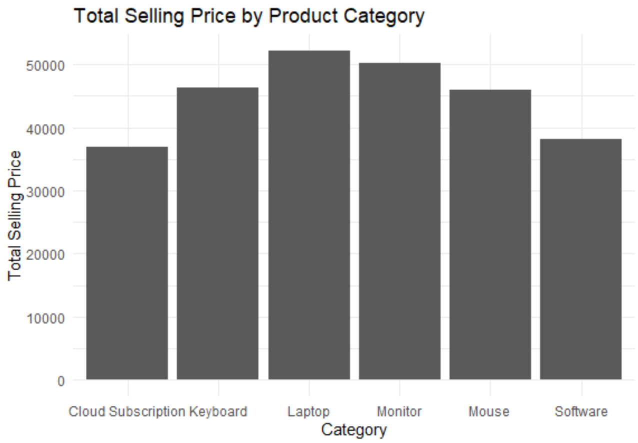


When analysing this histogram that illustrates the total monthly sales for both 2022 and 2023 it was discovered that during the 1st and 12th months of the year there is a notable decrease in the number of monthly sales. For the rest of the year sales are relatively stable between 100 000 and 120 000. The drop-off in sales during December and January is most likely due to most companies closing during Christmas and New Year. Therefore, the drop-off in sales is not something that the company should be concerned about.

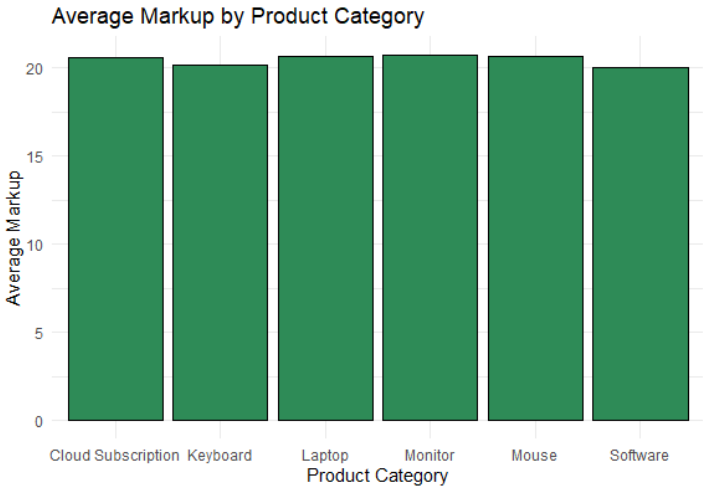
Product data



When analysing the scatterplot that shows the relationship between the selling price and the markup of products in different categories, no clear trend could be observed. It was found that the vast majority of product that the company sells is lower priced. For the lower priced products, the percentage markup was evenly distributed between 10-30 percent. The few products that are more expensive had a majority markup ranging between 10 and 20 percent with two products having a markup of close to 30 percent.

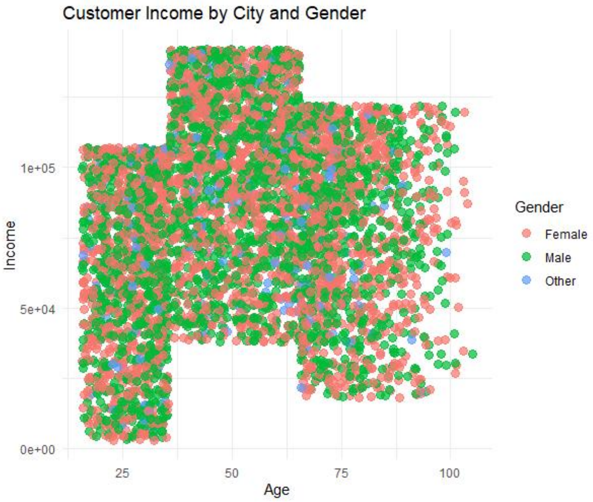


This histogram shows the total selling price of products from all the different product categories. By analysing this graph, it can clearly be seen that the seen that the categories that contributes the most to the total sales are laptops and monitors. Cloud subscriptions and software had the lowest contribution. From these statistics, it would be preferred that the company should focus on laptops and monitors, but the average markup should first be evaluated for each category to confirm this suggestion. Although cloud subscriptions and software contribute the least to the selling price, would they not require any storing space in a warehouse. It would not be advised to stop selling these products as they are products that has the least capital costs for the company.

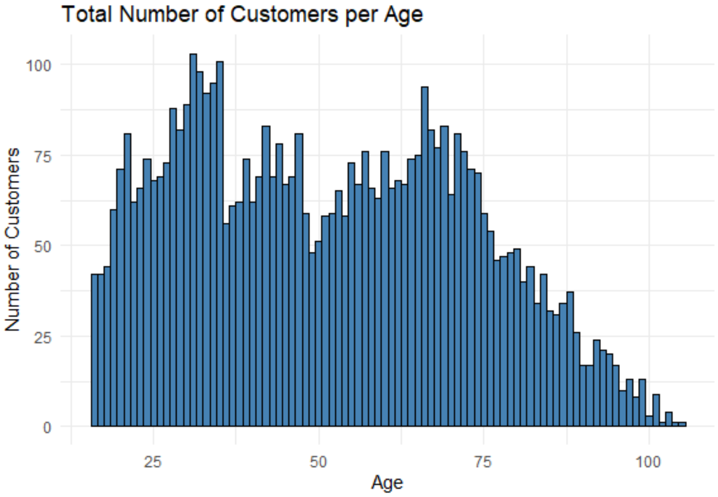


By analysing the average markup of each category, the previous statement can be supported. Laptops and monitors have the largest average markup. There is very little variation in the markups of the different categories, therefore it would not be suggested that the company should stop selling any of the products.

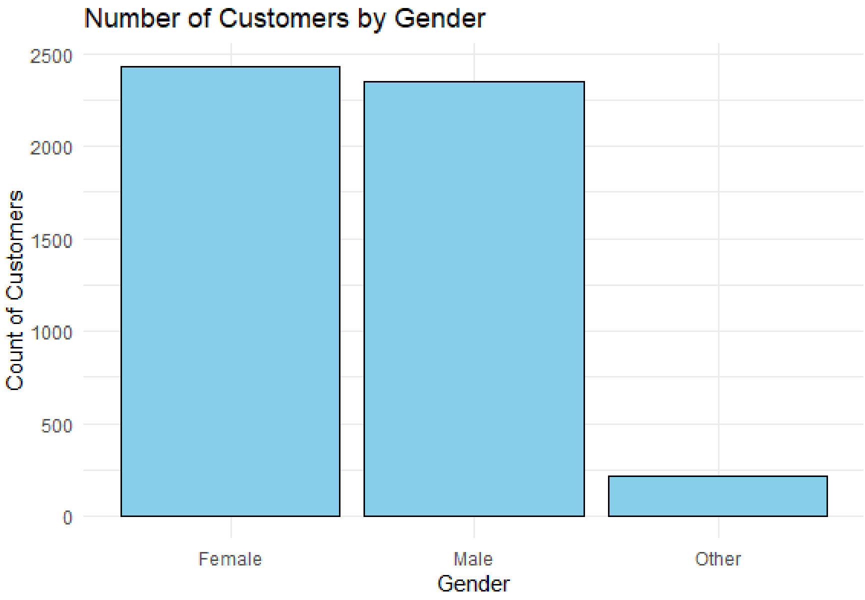
Customers data

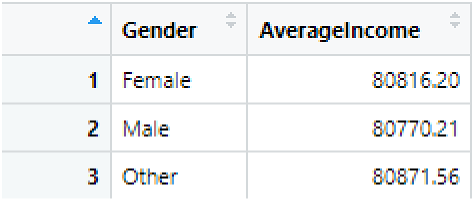


This scatterplot that shows how the income levels of different age groups that is categorized according to gender. These plots suggest that there are no patterns that could be recognised regarding a difference in income per gender, but rather how the income levels differ between age categories. It shows that middle-aged customers between 35 and 65 years of age tend to have a higher level of income. Customers aged younger than 35 tend to have an income ranging between low to medium-high income with customers aged 65 and older having an income ranging between lower-medium to higher-medium. It would be suggested that the company focusses on providing products that middle-aged people need, as they have the highest buying power.



When analysing this graph, it is seen that the younger population is more likely to buy products from the company. It would be advised to do market research to find out how to make the products more attractive for middle-aged people. This would lead to an increase in products sold.



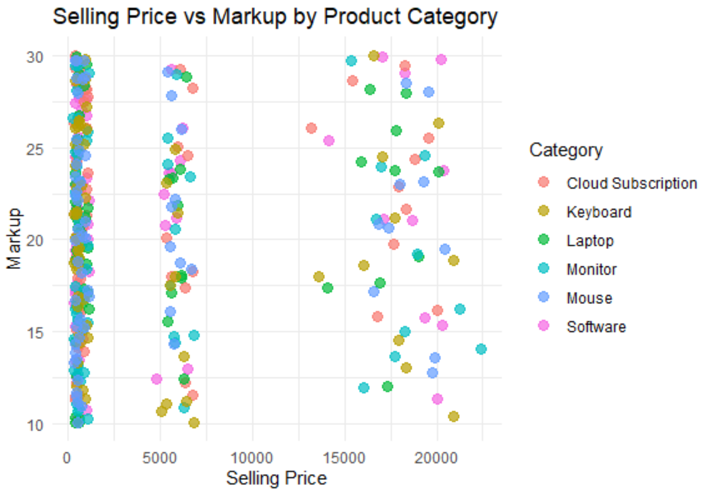


	Gender	AverageIncome
1	Female	80816.20
2	Male	80770.21
3	Other	80871.56

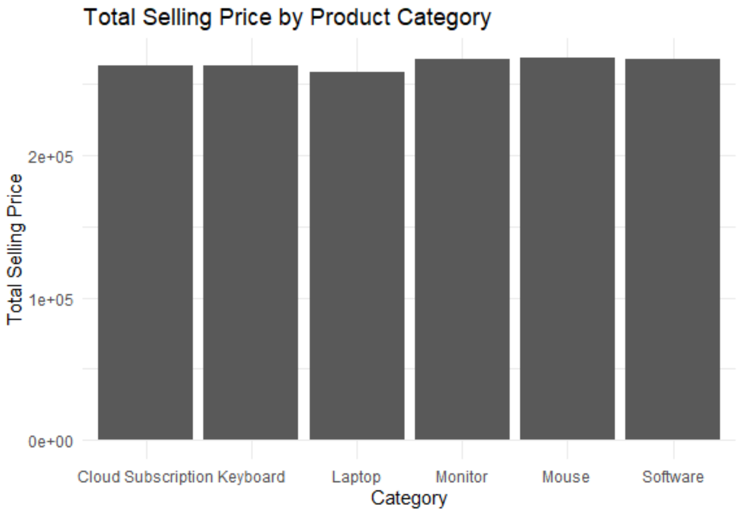
The company sells products that are of technological nature. After analysing the gender of the customers, it was found that the distribution between male and female is very even, along with their average income. It would therefore not be advised to target a specific gender, as technological devices are a necessity for all people.

Products Headoffice

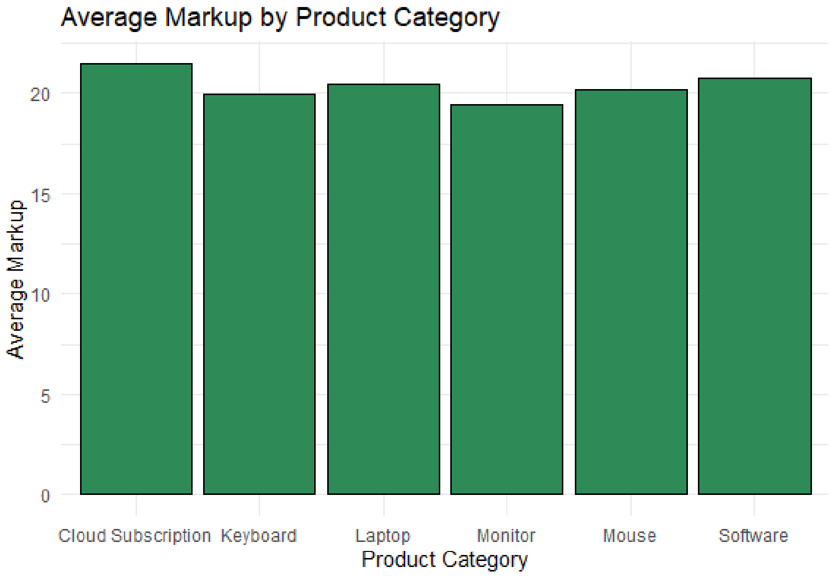
The head office products will now be compared to our company's product data, to find if the performance of our branch aligns with the head office.



In this scatterplot the head office sells products of the same categories, but on a much larger scale. Although there are more products that are being sold, the same pattern is observed. Most products are around $5000 or less, while they are equally spread out on the markup axis. This shows that our branch is on par with their pricing when comparing it to the head office.



When analysing the total contribution to the sales of the different categories at the head office, it was noted that the contribution of the different categories was very equal, unlike those of our branch. At the head office laptops had the lowest contribution, while at our branch it had the highest contribution.



At the head office, cloud subscriptions had the highest markup, with all the other categories also being in the range of a 20% markup. This would mean that all the products are equally important for the company. The outcome of this graph aligns with the graph that was generated when analysing the product data file. This is most likely due to the company having a set selling price for the different products and other branches of the company is not permitted to sell products at their own price.

Conclusion

In conclusion, the analysis of these four Excel datasheets provided valuable insight into the company's sales, products and customers. The data revealed that the company had a significant decline in sales and transactions from 2022 to 2023. This major decline was accompanied by a slight reduction in delivery time. The monthly sales indicated a stable pattern from February to November, with expected drop offs in December and January due to possible closures. The product analysis showed that the lower-priced products dominated the sales, but the markups on products were equally distributed through the products of all price categories. Customer data highlighted that middle-aged individuals have the highest buying power due to their higher income. Currently the age group that buys most frequently is younger people. It is suggested that the company makes use of marketing that would attract middle-aged people. By analysing the head offices' data, it is found that our branch aligns with their pricing and markup although the number of sales differ. Overall, the company should focus on optimizing productivity, targeting middle-aged customers and maintaining a diverse product portfolio to enhance their performance.

Phase 2

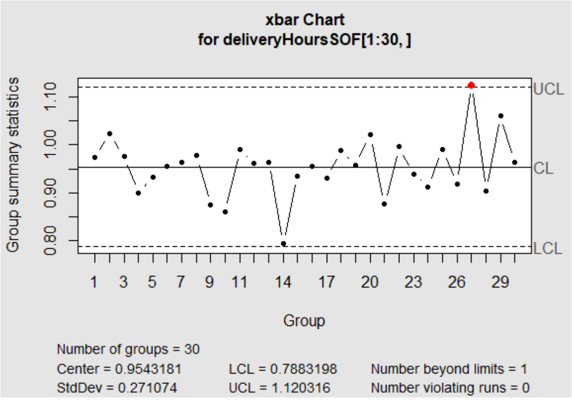
Introduction

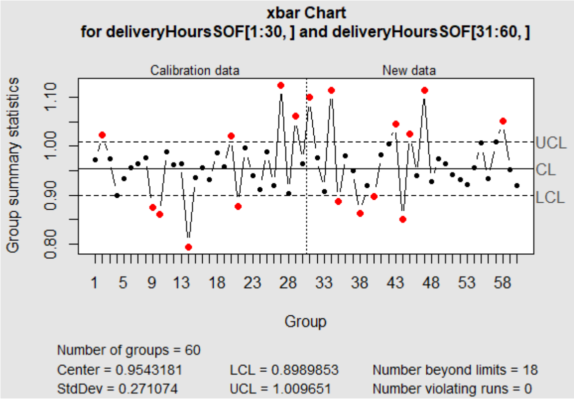
Phase 2 of the project aims to monitor and enhance the delivery process for six product types using statistical process control techniques (SPC). In this phase Xbar- and S-Charts will be used to track the mean and standard deviation of the delivery times. A sample size of 24 deliveries will be used. The first 30 samples will be plotted to calculate the mean and standard deviation. Thereafter a further 30 samples will be added to calibrate the data. Then the rest of the samples will be plotted on the same graph with the calibrated limits. This would allow us to identify the number of samples that fall outside the upper and lower limits.

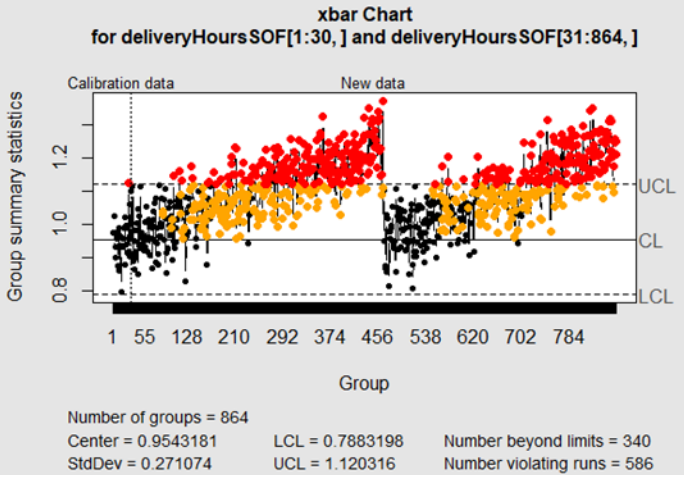
Questions 3.1 and 3.2

SOF

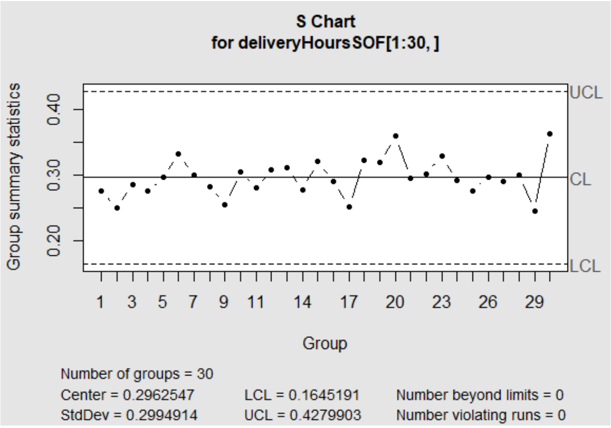
Xbar Chart



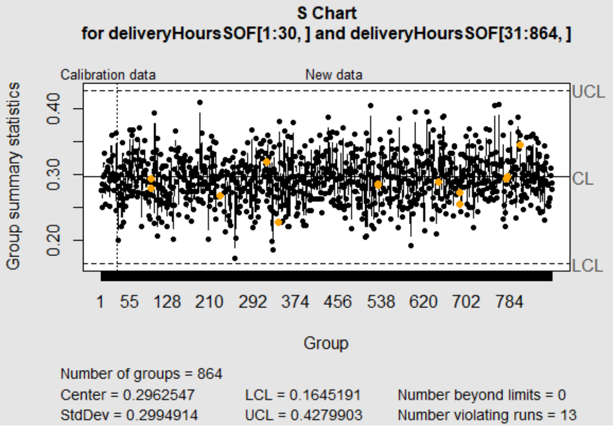




S Chart



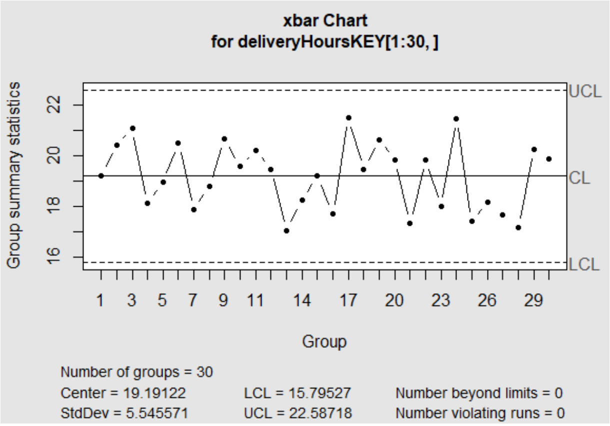
Number of groups = 30
Center = 0.2962547
StdDev = 0.2994914
LCL = 0.1645191
UCL = 0.4279903
Number beyond limits = 0
Number violating runs = 0



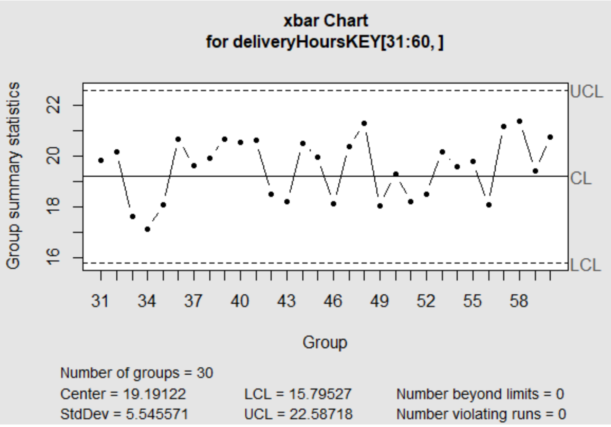
Number of groups = 864
Center = 0.2962547
StdDev = 0.2994914
LCL = 0.1645191
UCL = 0.4279903
Number beyond limits = 0
Number violating runs = 13

KEY

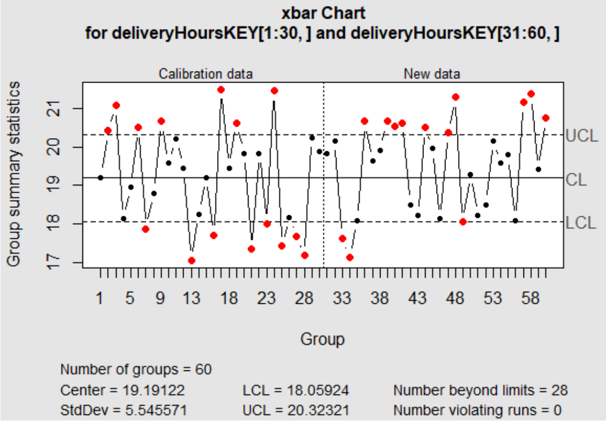
Xbar Chart



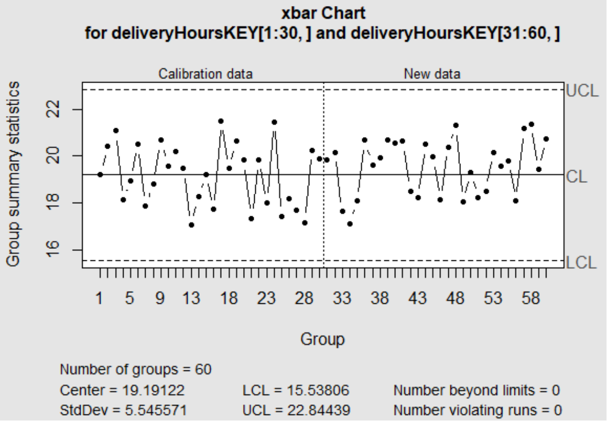
```
Number of groups = 30
Center = 19.19122         LCL = 15.79527       Number beyond limits = 0
StdDev = 5.545571         UCL = 22.58718       Number violating runs = 0
```



```
Number of groups = 30
Center = 19.19122         LCL = 15.79527       Number beyond limits = 0
StdDev = 5.545571         UCL = 22.58718       Number violating runs = 0
```

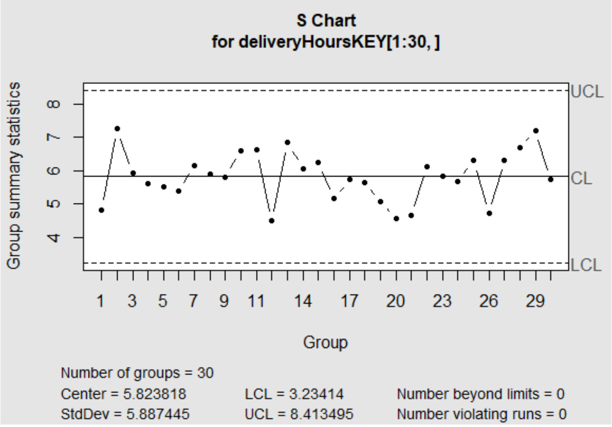


```
Number of groups = 60
Center = 19.19122         LCL = 18.05924       Number beyond limits = 28
StdDev = 5.545571         UCL = 20.32321       Number violating runs = 0
```

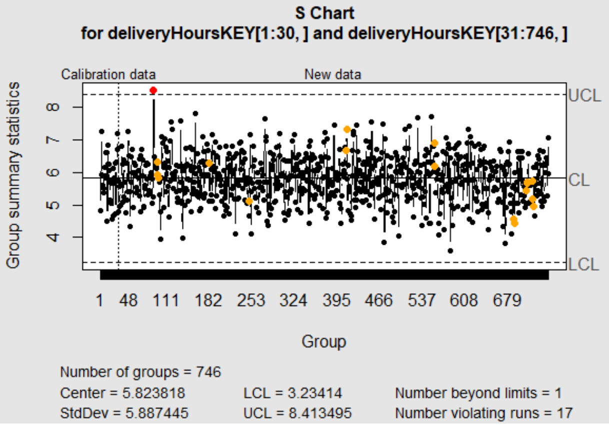


```
Number of groups = 60
Center = 19.19122         LCL = 15.53806       Number beyond limits = 0
StdDev = 5.545571         UCL = 22.84439       Number violating runs = 0
```

S Chart



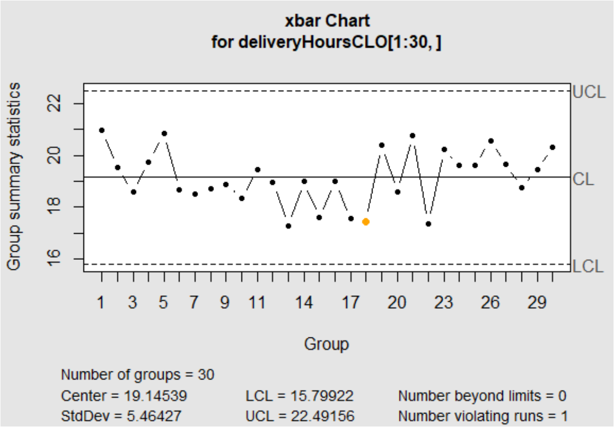
```
Number of groups = 30
Center = 5.823818        LCL = 3.23414        Number beyond limits = 0
StdDev = 5.887445        UCL = 8.413495       Number violating runs = 0
```

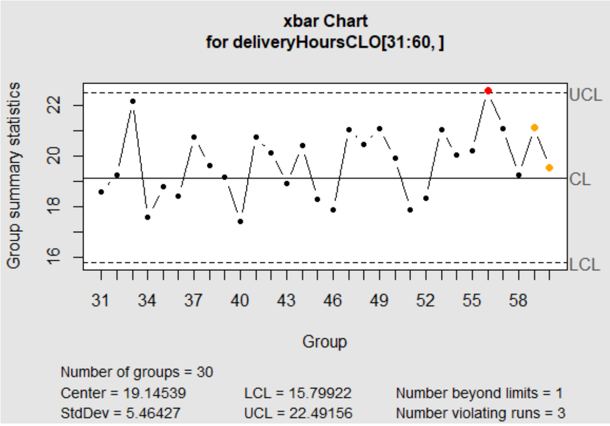


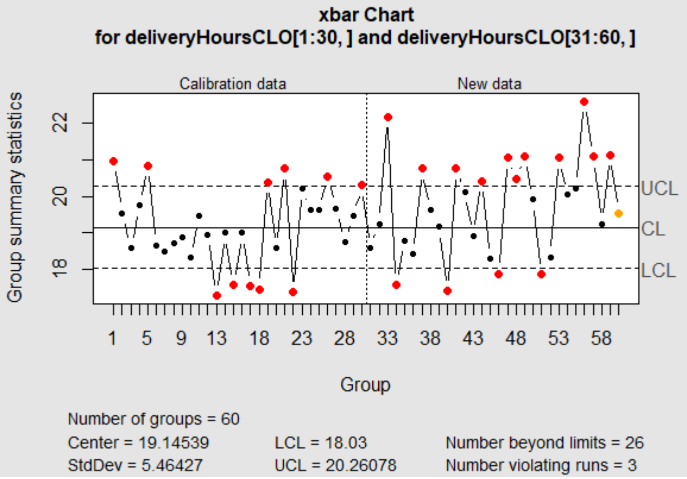
```
Number of groups = 746
Center = 5.823818        LCL = 3.23414        Number beyond limits = 1
StdDev = 5.887445        UCL = 8.413495       Number violating runs = 17
```

CLO

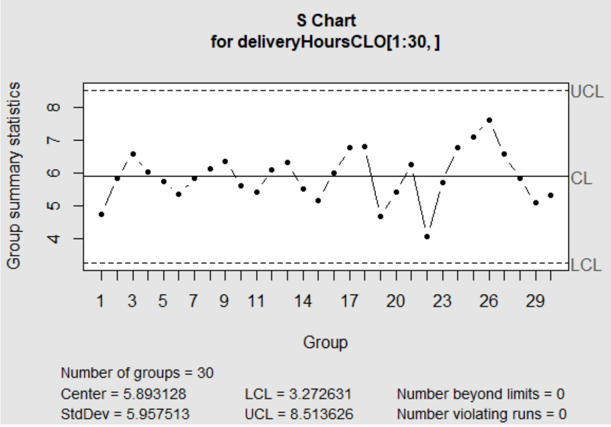
Xbar chart



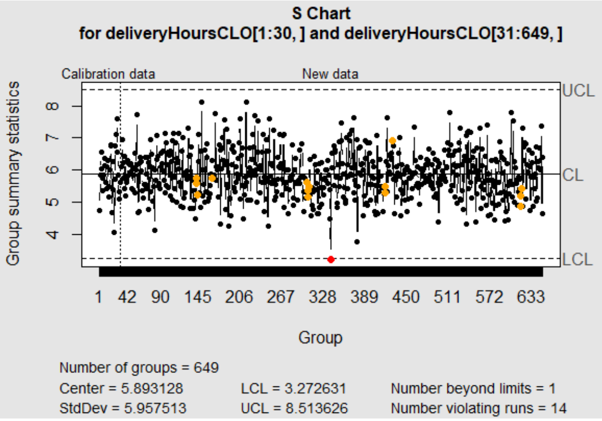




S Chart



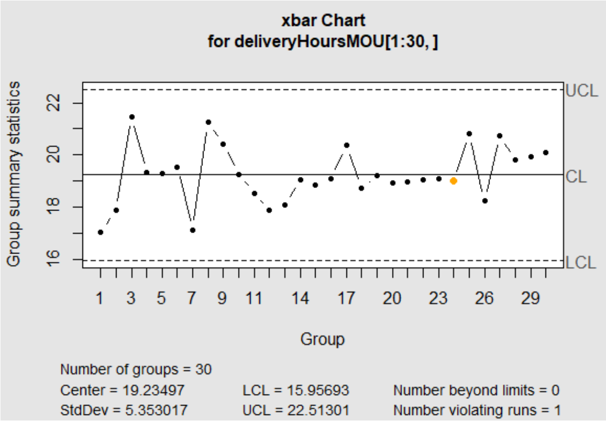
```
Number of groups = 30
Center = 5.893128        LCL = 3.272631        Number beyond limits = 0
StdDev = 5.957513        UCL = 8.513626        Number violating runs = 0
```

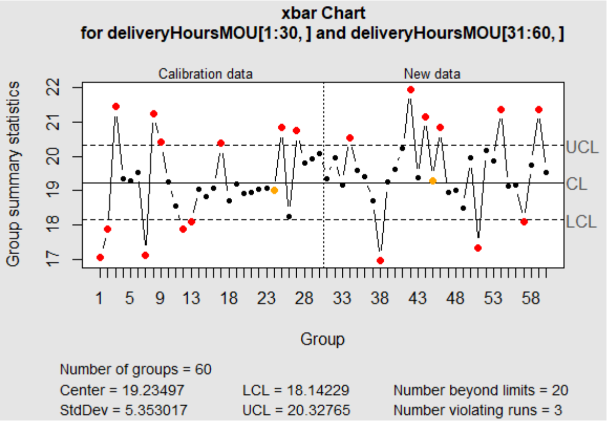


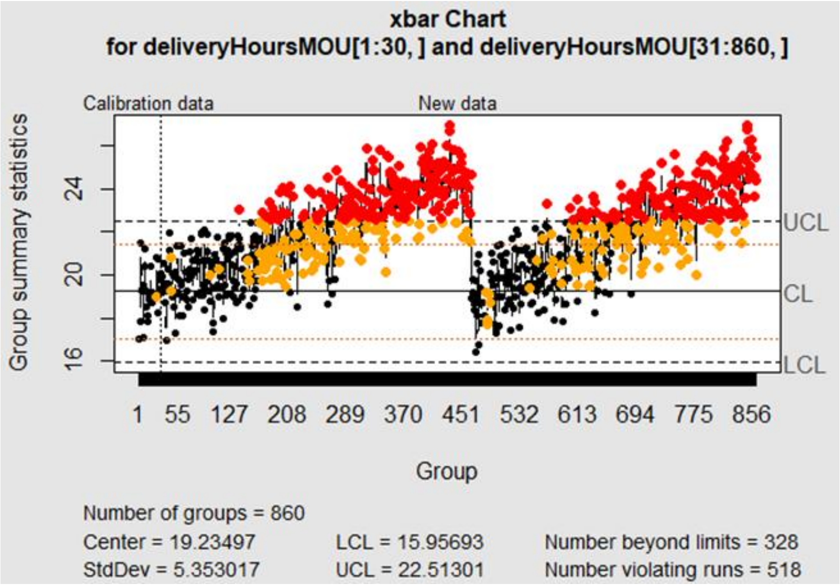
```
Number of groups = 649
Center = 5.893128        LCL = 3.272631        Number beyond limits = 1
StdDev = 5.957513        UCL = 8.513626        Number violating runs = 14
```

MOU

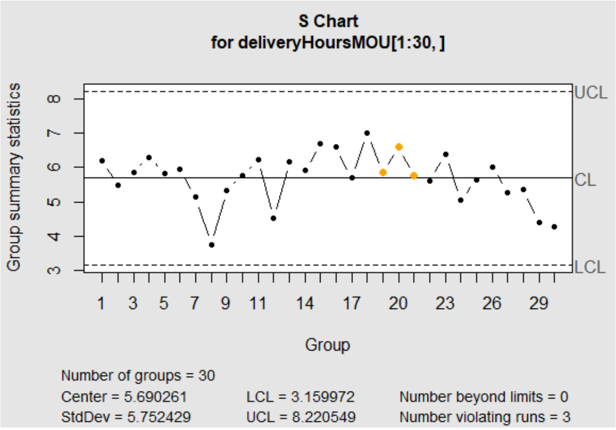
Xbar Chart

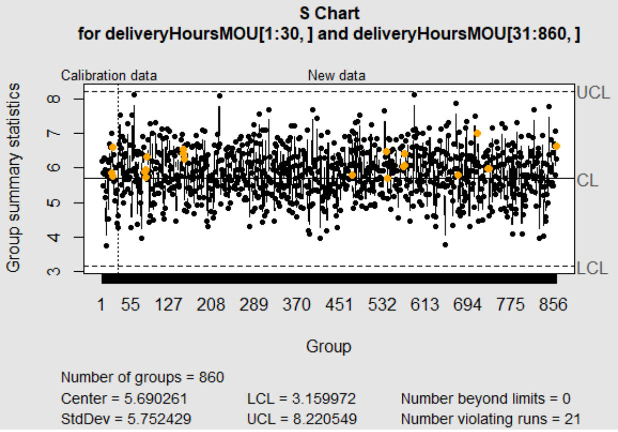






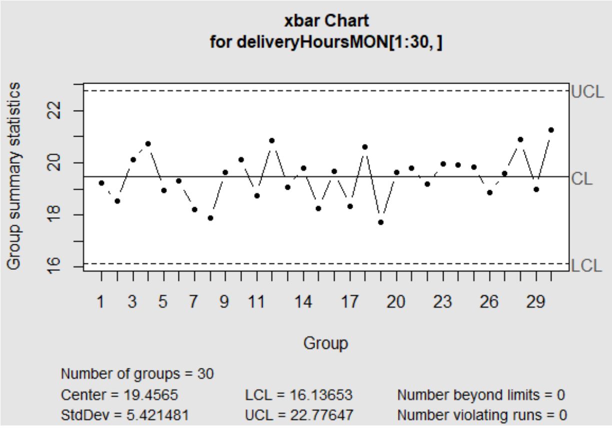
S Chart


S Chart
for deliveryHoursMOU[1:30,]
Group summary statistics
3
4
5
6
7
8
UCL
CL
LCL
1
3
5
7
9
11
14
17
20
23
26
29
Group
Number of groups = 30
Center = 5.690261
StdDev = 5.752429
LCL = 3.159972
UCL = 8.220549
Number beyond limits = 0
Number violating runs = 3

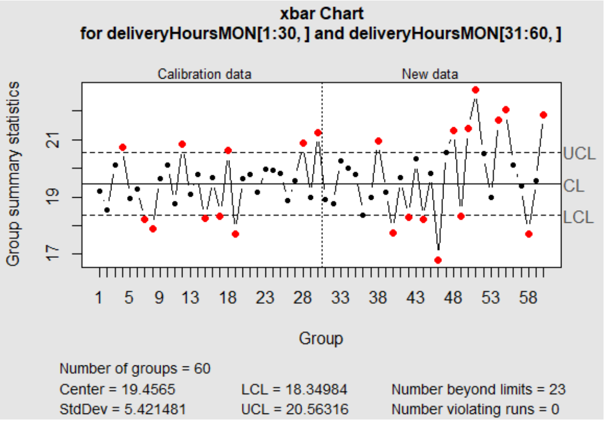

S Chart
for deliveryHoursMOU[1:30,] and deliveryHoursMOU[31:860,]
Calibration data
New data
Group summary statistics
3
4
5
6
7
8
UCL
CL
LCL
1
55
127
208
289
370
451
532
613
694
775
856
Group
Number of groups = 860
Center = 5.690261
StdDev = 5.752429
LCL = 3.159972
UCL = 8.220549
Number beyond limits = 0
Number violating runs = 21

MON

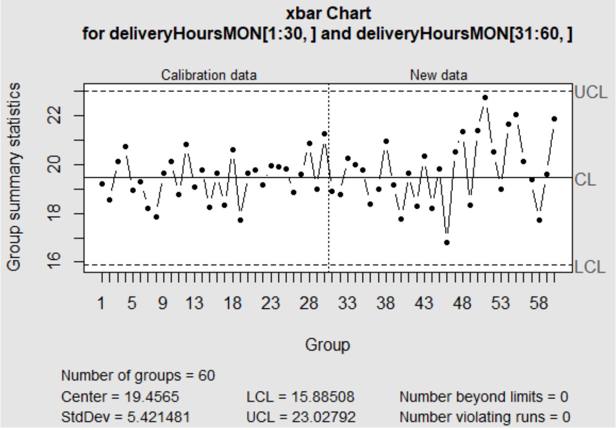
Xbar Chart


xbar Chart
for deliveryHoursMON[1:30,]

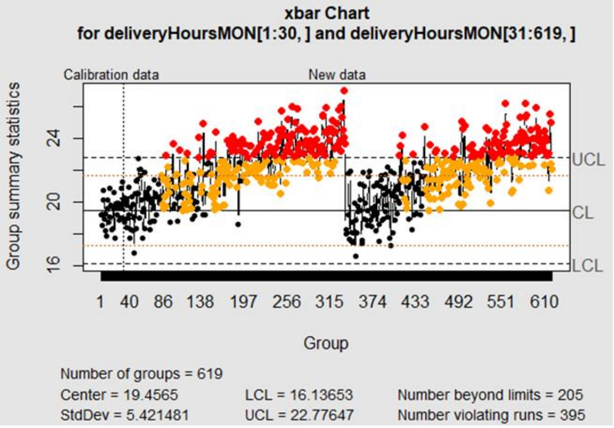
Number of groups = 30
Center = 19.4565 LCL = 16.13653 Number beyond limits = 0
StdDev = 5.421481 UCL = 22.77647 Number violating runs = 0


xbar Chart
for deliveryHoursMON[1:30,] and deliveryHoursMON[31:60,]

Number of groups = 60
Center = 19.4565 LCL = 18.34984 Number beyond limits = 23
StdDev = 5.421481 UCL = 20.56316 Number violating runs = 0

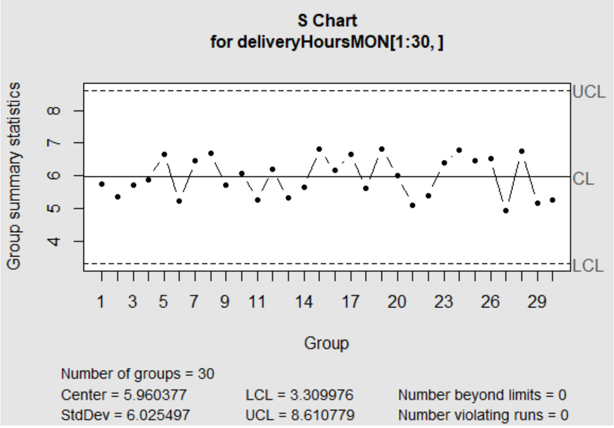

xbar Chart
for deliveryHoursMON[1:30,] and deliveryHoursMON[31:60,]

Number of groups = 60
Center = 19.4565 LCL = 15.88508 Number beyond limits = 0
StdDev = 5.421481 UCL = 23.02792 Number violating runs = 0

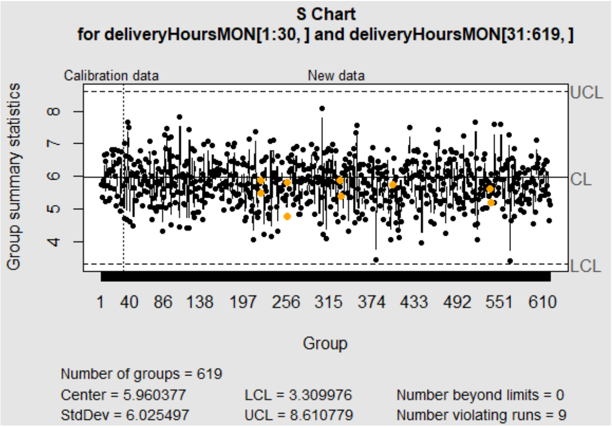

xbar Chart
for deliveryHoursMON[1:30,] and deliveryHoursMON[31:619,]

Number of groups = 619
Center = 19.4565 LCL = 16.13653 Number beyond limits = 205
StdDev = 5.421481 UCL = 22.77647 Number violating runs = 395

S Chart



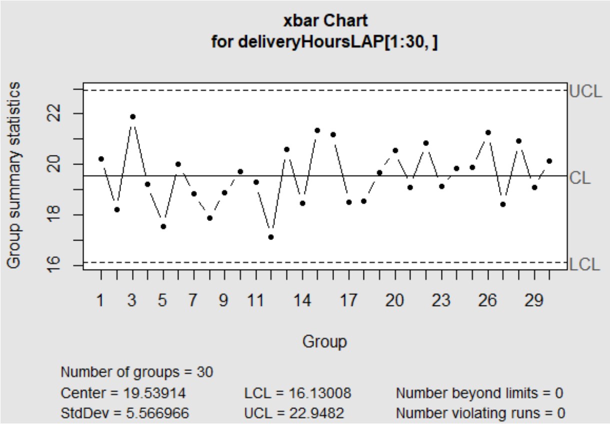
```
Number of groups = 30
Center = 5.960377          LCL = 3.309976          Number beyond limits = 0
StdDev = 6.025497          UCL = 8.610779          Number violating runs = 0
```

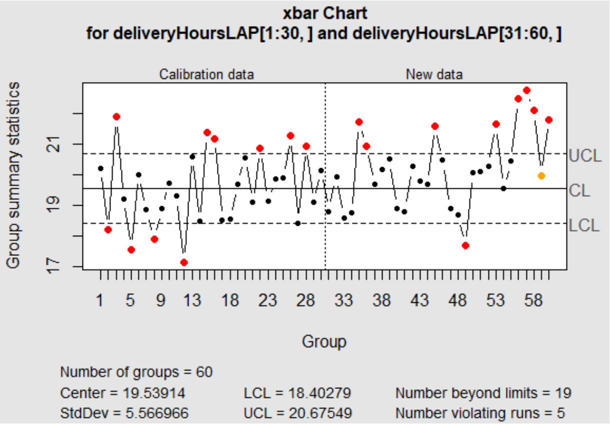


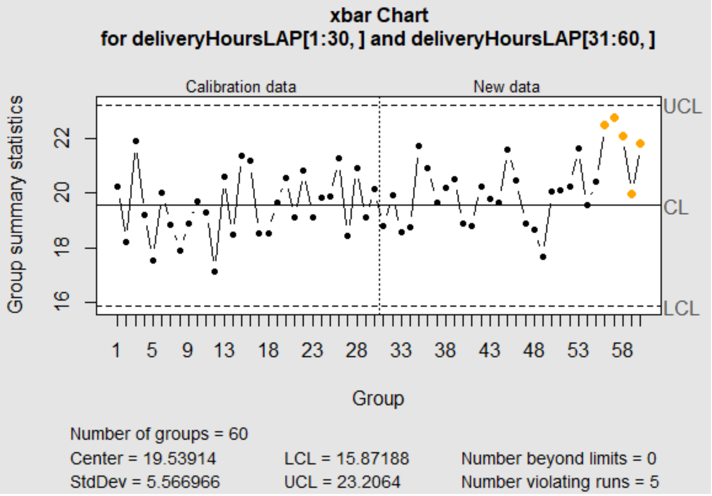
```
Number of groups = 619
Center = 5.960377          LCL = 3.309976          Number beyond limits = 0
StdDev = 6.025497          UCL = 8.610779          Number violating runs = 9
```

LAP

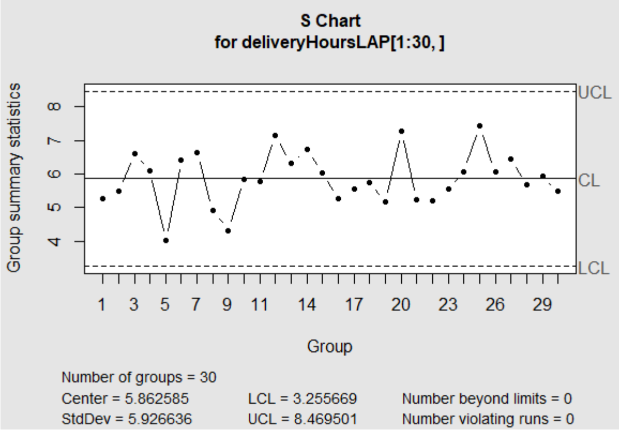
Xbar Chart



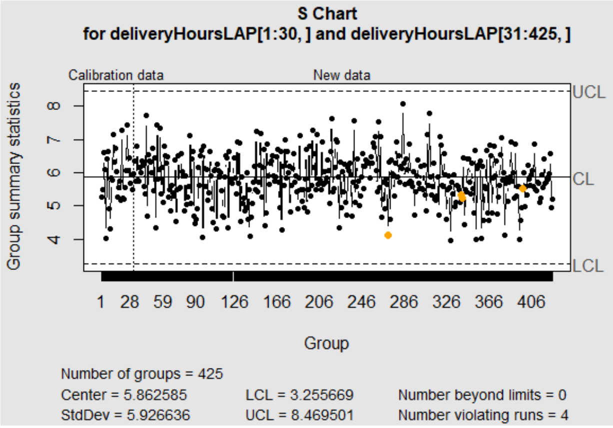




S Chart



```
Number of groups = 30
Center = 5.862585        LCL = 3.255669        Number beyond limits = 0
StdDev = 5.926636        UCL = 8.469501        Number violating runs = 0
```



```
Number of groups = 425
Center = 5.862585        LCL = 3.255669        Number beyond limits = 0
StdDev = 5.926636        UCL = 8.469501        Number violating runs = 4
```

Discussion of Charts

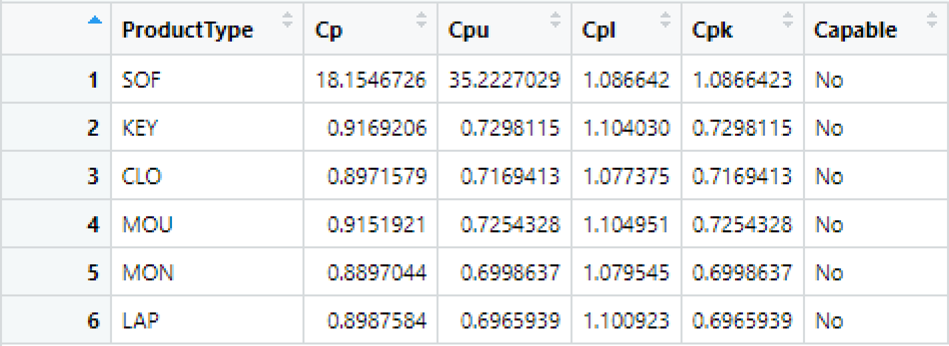
Xbar Chart

The Xbar charts display the mean delivery times for samples of 24 deliveries. For each product type the first 30 samples are used to create the mean as well as the 1 sigma upper and lower limits. After the limits is established, the next 30 samples are plotted on the same limits. The mean and control limits are then adjusted. After the mean and limits are adjusted all the remaining samples are plot on a graph with these limits. We can identify the number of samples that fall without these limits. The goal is to minimize the number of samples outside the limits to improve the productivity. It is noticed that most samples that are beyond the limits exceeds the upper limit rather than the lower limit. This indicates that the processes are most likely to be slower. It was also noticed that all the Xbar charts that plotted all the samples, illustrated that there is a pattern that emerges in the data, where the first few samples start within the calibrated boundaries and then gradually more samples fall outside the boundaries until halfway. At the halfway mark the samples the samples fall within the boundaries, but as time goes on it gradually starts falling outside the boundaries. This may be due to negligent workers or due to poor calibration techniques. It is also noted that initially the calibrated boundaries are set to narrow which causes that the boundaries should be readjusted due to too many samples falling outside the boundaries.

S Chart

The S chart monitors the process variation, by controlling the variability and spread within each subgroup. Like the Xbar chart, the first 30 samples established the centre line and control limits. After these limits are set the rest of the samples are plotted onto the graph. It is observed that that very few samples exceed the S Chart limits, while many samples exceed the Xbar Chart limits. This is due to a spread of data within the subgroups being detected by the S Chart. When the process means shift, the S Chart would remain in control, while the Xbar Chart goes out of control. Due to the low variability in the standard deviation, the original limits that were determined upon analysing the first 30 samples did not change when all the samples were plotted against the same limits. Across all the samples of all 6 types of products, only 2 samples exceeded the original limits.

Question 3.3

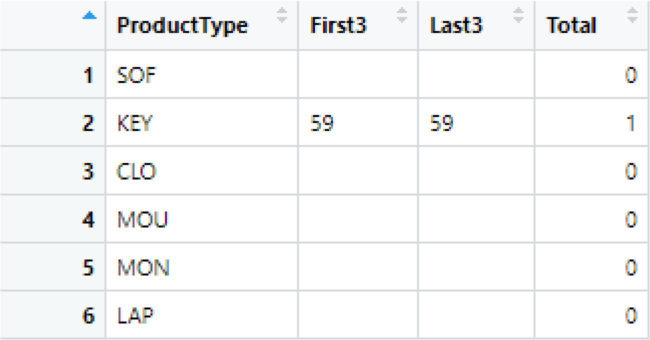


	ProductType	Cp	Cpu	Cpl	Cpk	Capable
1	SOF	18.1546726	35.2227029	1.086642	1.0866423	No
2	KEY	0.9169206	0.7298115	1.104030	0.7298115	No
3	CLO	0.8971579	0.7169413	1.077375	0.7169413	No
4	MOU	0.9151921	0.7254328	1.104951	0.7254328	No
5	MON	0.8897044	0.6998637	1.079545	0.6998637	No
6	LAP	0.8987584	0.6965939	1.100923	0.6965939	No

This analysis reveals significant insights into the process performance, by making use of process capability indices (Cp, Cpu, Cpl, Cpk). These indices were calculated by using the first 1000 deliveries per product type. The results indicates that none of the product types can meet the voice of the customer. It is observed that SOF has very high Cp and Cpu values. This means that the process is stable and normally distributed, with a highly consistent delivery process. The high Cpu indicates that the process mean is well below the upper specification limit, which means that it has minimal risk of exceeding the limit of 32 hours.

Question 3.4

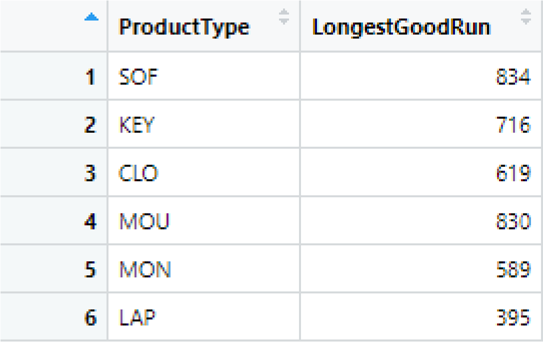
a)



	ProductType	First3	Last3	Total
1	SOF			0
2	KEY	59	59	1
3	CLO			0
4	MOU			0
5	MON			0
6	LAP			0

It is observed that only one sample falls outside the 3-sigma limit. This product is of type KEY and is the 59th sample.

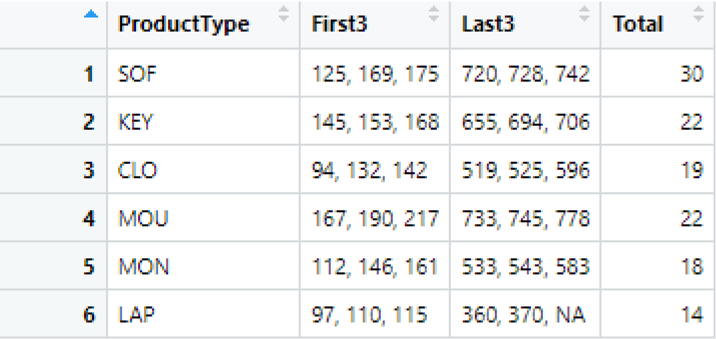
b)



	ProductType	LongestGoodRun
1	SOF	834
2	KEY	716
3	CLO	619
4	MOU	830
5	MON	589
6	LAP	395

It is observed that both products of type SOF and MOU have very long runs with consecutive samples within + or – 1 standard deviation. This indicates that the data has very little variation.

c)

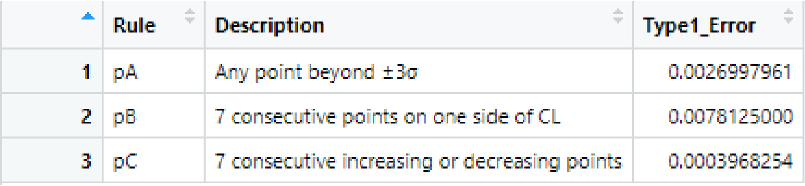


	ProductType	First3	Last3	Total
1	SOF	125, 169, 175	720, 728, 742	30
2	KEY	145, 153, 168	655, 694, 706	22
3	CLO	94, 132, 142	519, 525, 596	19
4	MOU	167, 190, 217	733, 745, 778	22
5	MON	112, 146, 161	533, 543, 583	18
6	LAP	97, 110, 115	360, 370, NA	14

This chart highlights the process control issues across all six product types. It shows that SOF has the most violations of 4 consecutive samples exceeding the second upper control limit, while LAP has the least.

Phase 3

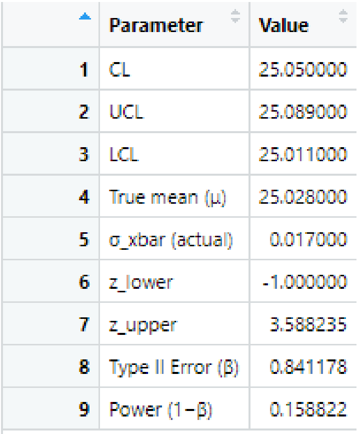
Question 4.1



	Rule	Description	Type1_Error
1	pA	Any point beyond ±3σ	0.0026997961
2	pB	7 consecutive points on one side of CL	0.0078125000
3	pC	7 consecutive increasing or decreasing points	0.0003968254

A Type I error is when you conclude that a process is out of control, when it is in control. Making a Type I error would penalize the manufacturer, as it leads to unnecessary investigation, down time or the rejection of good products. For problem A we can see that there is a 0.27% chance that a sample would appear out of the ±3σ boundaries. It is stated that each sample of problem B has a 50% chance to be above or below the centre line. It is calculated that there is a 0.78% chance that 7 consecutive points from a random sample would be above or below the centre line. To calculate the probability of 7 consecutive samples in problem C decreasing or increasing it was found that the probability is 0.039% this is regarded as very low, as it would mean that less than 4 samples out of 10 000 would conform to this rule. If we were to compute for only increasing the probability would even be smaller.

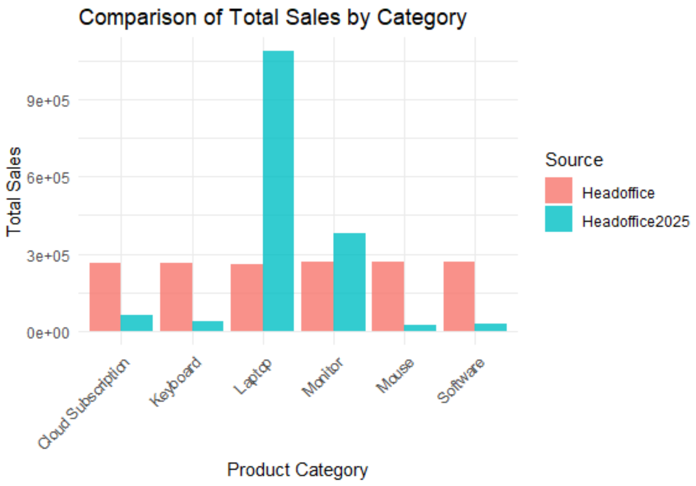
Question 4.2



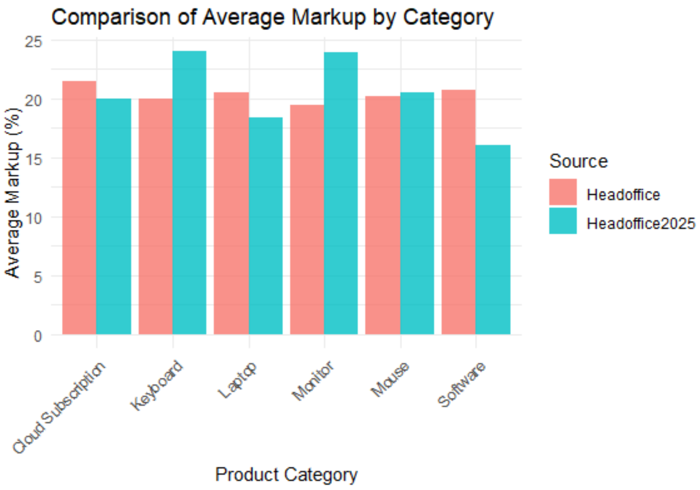
	Parameter	Value
1	CL	25.050000
2	UCL	25.089000
3	LCL	25.011000
4	True mean (μ)	25.028000
5	σ_xbar (actual)	0.017000
6	z_lower	-1.000000
7	z_upper	3.588235
8	Type II Error (β)	0.841178
9	Power (1−β)	0.158822

A Type II error occurs when it is failed to detect whether a process is out of control. This would mean that the null hypothesis would be selected, even though the alternative hypothesis is true. In this problem the probability of making a Type II error is 84.1%. This is a very worrying sign for the consumers, as there is a very large probability that the manufacturer would not detect any problems with the product, which would lead to them buying a faulty product.

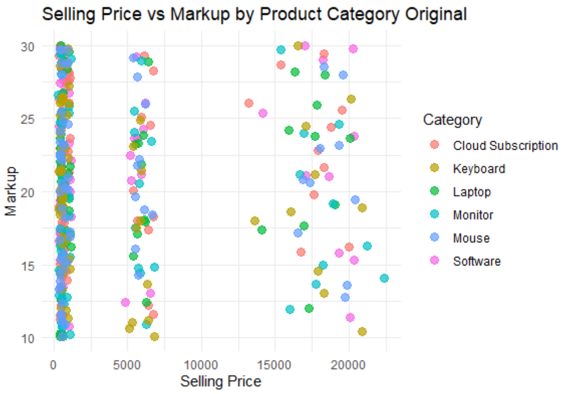
Question 4.3

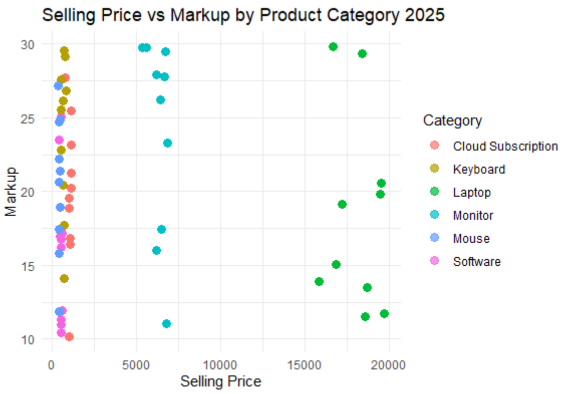


When comparing the total sales per category of the previous head office data with the current head office data, it is noticed that there was a major error in the original data. According to the 2025 data, it shows that Laptops are by far the highest contributor to the companies' income. The total income generated from Monitors had a slight increase, while Cloud Subscriptions, Keyboards, Mouses and Software were much lower.



The average markups of the different categories stayed relatively consistent, compared to the total sales, however there were a few noticeable changes. Keyboards and Monitors are now the two categories with the highest markup, while Laptops and Software are now the lowest. By analysing this data, it can be suggested to the company that they stop selling software, as it is the lowest contributor to total sales, as well as markup.

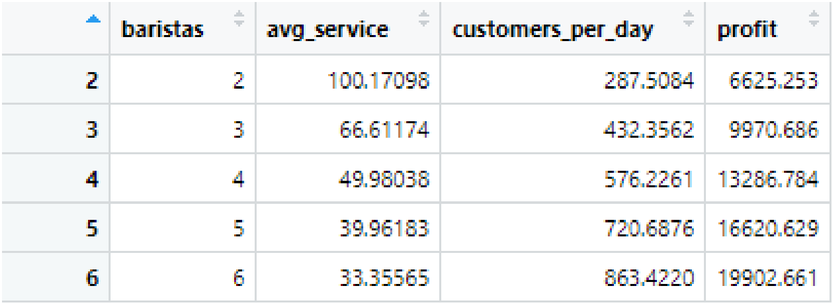




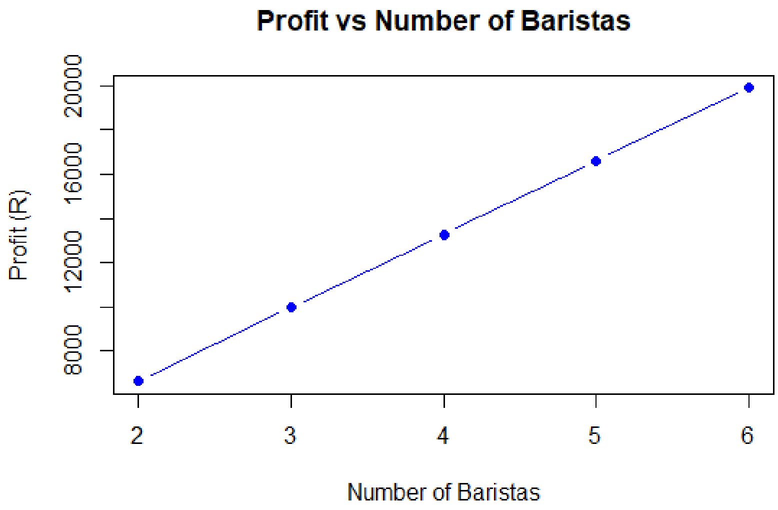
By analysing this scatterplot, we can identify that the original data was faulty, as there were no correlations between the Selling Price and Markup of the different categories. In the corrected data, we can clearly see that Laptops and Monitors have the highest selling prices, while Monitors has the highest markup. These observations align with the observations in the previous histograms.

Question 5.

Shop 1:

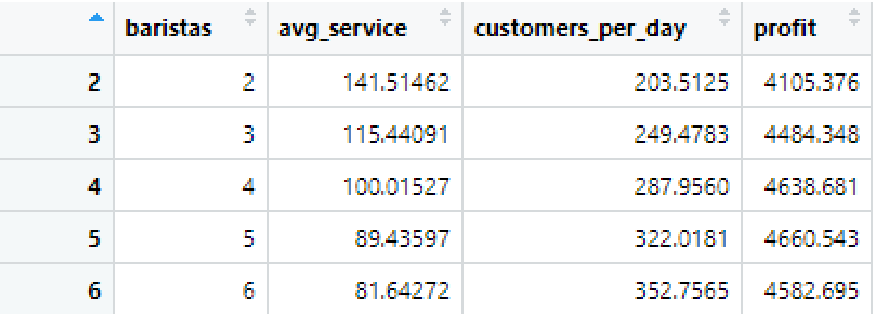


	baristas	avg_service	customers_per_day	profit
2	2	100.17098	287.5084	6625.253
3	3	66.61174	432.3562	9970.686
4	4	49.98038	576.2261	13286.784
5	5	39.96183	720.6876	16620.629
6	6	33.35565	863.4220	19902.661

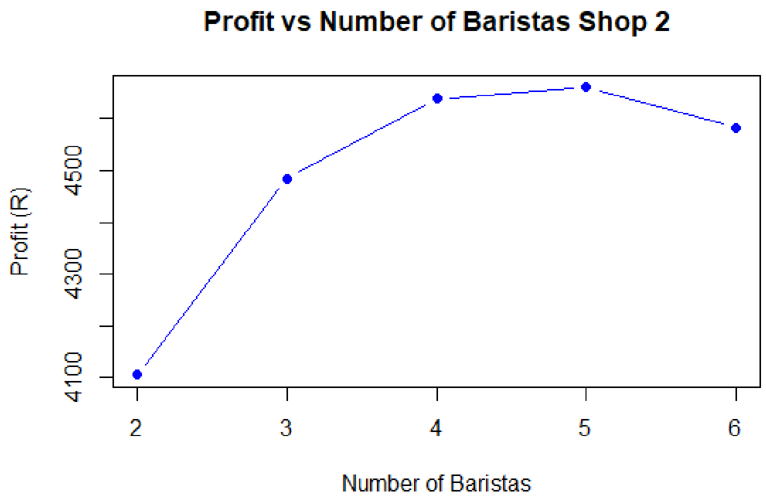


When analysing the results of shop 1 it is found that that the profit linearly increases for each barista that is added. According to our calculations, the coffee shop would make the maximum profit when they appoint 6 baristas. When looking at the results logically it can be concluded that an error may occur in the data set that was constructed to perform the analysis, as there would be a point where the number of customers that the shop can service reaches a maximum.

Shop 2:



	baristas	avg_service	customers_per_day	profit
2	2	141.51462	203.5125	4105.376
3	3	115.44091	249.4783	4484.348
4	4	100.01527	287.9560	4638.681
5	5	89.43597	322.0181	4660.543
6	6	81.64272	352.7565	4582.695



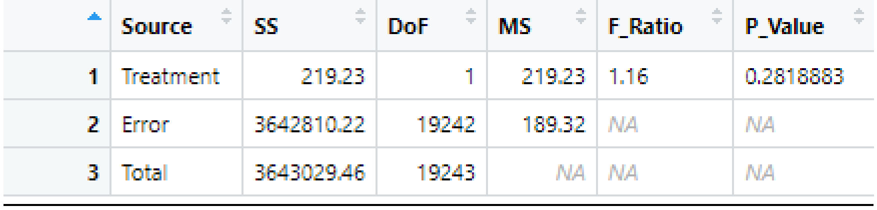
For shop 2 the model calculated that 5 baristas would be the optimal amount. For 6 baristas it is noticed that the profit starts decreasing. This is due to the number of customers that are serviced not increasing by the same rate as with the other increments in baristas. It would therefore not be feasible to appoint another barista, as the income would not increase significantly enough to cover their daily wage and increasing the profit. It would be advised that the coffee shop monitors their peak hours to analyse if it would be profitable to appoint a casual barista that only works during rush hours, while only keeping 5 permanent baristas.

Phase 4

Question 6

Both ANOVA tables would measure products of type SOF. Both models would be measured against a significance level of 0.05. If the model finds that $p < 0.05$, reject H_0 due to significant difference. If $p \geq 0.05$ we fail to reject H_0 due to no significant difference.

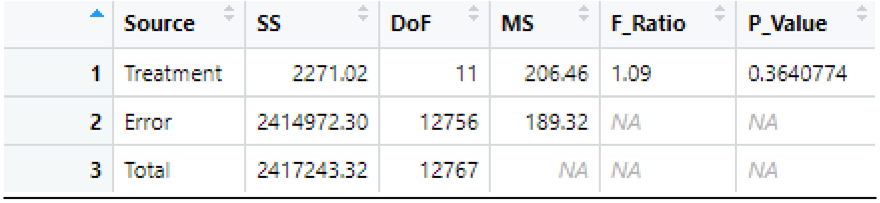
ANOVA Year



	Source	SS	DoF	MS	F_Ratio	P_Value
1	Treatment	219.23	1	219.23	1.16	0.2818883
2	Error	3642810.22	19242	189.32	NA	NA
3	Total	3643029.46	19243	NA	NA	NA

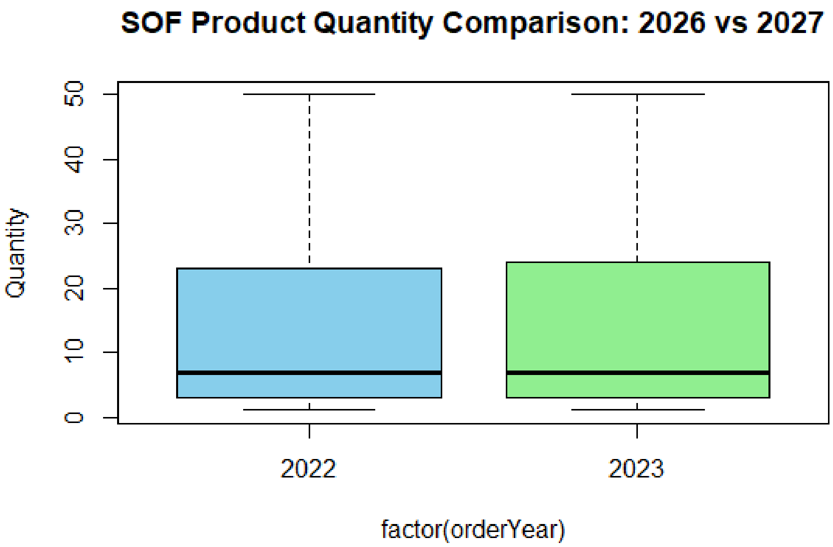
This one-way ANOVA table compares the means between years 2026 and 2027. This analysis test is to test the null hypothesis that the population means for the outcomes of products of type SOF are equal across the two years. After analysing the outcomes, the p-value is compared to the chosen significance level of 0.05. Due to the p value being greater than the significance level, it indicates that there is no meaningful difference between the years and shows a stable outcome across the two years.

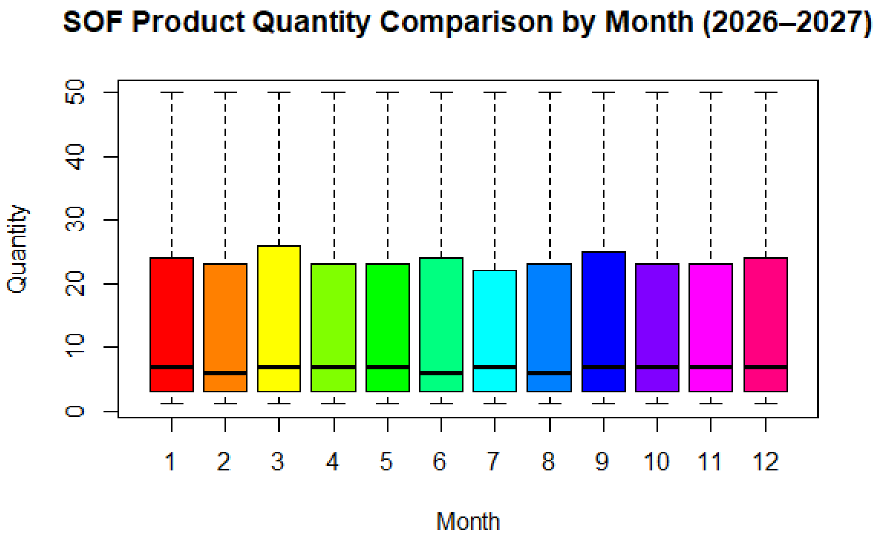
ANOVA Month



	Source	SS	DoF	MS	F_Ratio	P_Value
1	Treatment	2271.02	11	206.46	1.09	0.3640774
2	Error	2414972.30	12756	189.32	NA	NA
3	Total	2417243.32	12767	NA	NA	NA

This ANOVA table compares the sales of the category SOF over two years to determine whether there is a significant difference between the months in the year. Due to comparing the data of 12 months, the degree of freedom is 11. Upon evaluation, it is determined that the p-value is 0.3640774, which exceeds the significance level of 0.05. Due to this observation, it is determined that there is insufficient evidence to reject the null hypothesis. This indicates that the sales of the specific category do not vary significantly over the 12 months.





By analysing the following boxplots, it is observed that the quantities stay very consistent over the periods that it is measured on. This indicates that the data does not vary, which supports the conclusion that was made on the previous page when the ANOVA tables were analysed.

Question 7

7.1

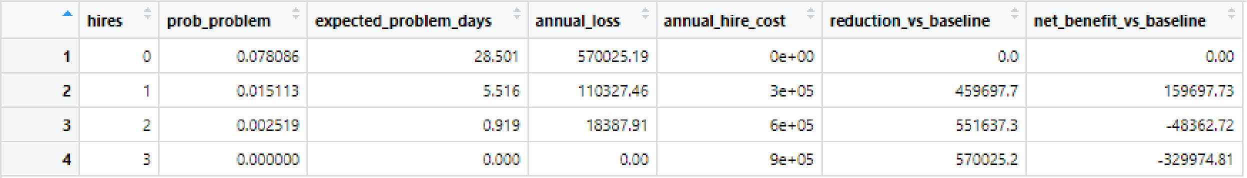
Assume reliable day is when 15 or more workers arrive to work.

Reliable days observed: 366 out of 397 days.

$$reliable\ service\ days\ per\ year = \frac{366}{397} \times 365$$

Estimated reliable service days per year: 336.

7.2



	hires	prob_problem	expected_problem_days	annual_loss	annual_hire_cost	reduction_vs_baseline	net_benefit_vs_baseline
1	0	0.078086	28.501	570025.19	0e+00	0.0	0.00
2	1	0.015113	5.516	110327.46	3e+05	459697.7	159697.73
3	2	0.002519	0.919	18387.91	6e+05	551637.3	-48362.72
4	3	0.000000	0.000	0.00	9e+05	570025.2	-329974.81

When optimizing this model that predicts the number of additional employees that should be appointed to avoid that the company suffers a loss of income when less than 15 employees are on duty, it was determined that one additional employee should be appointed. When one additional employee is appointed, the projected probability of experiencing a problem is 0.015113, which is a big improvement from the current probability of 0.078086. It is projected that that the company would make a benefit of R159 697.73 if an extra employee is appointed. In contrast if 2 or more employees are appointed it would improve the projected problem days to close to none, however the increased hiring cost would lead to a loss of profit, as it would exceed the additional income.

Conclusion

The project successfully demonstrated how methods and theory that is presented throughout the Industrial Engineering course can be applied to real world data analysis and optimization problems. Descriptive statistics, SPC and process capability indices provided significant insights into delivery processes and product performance. Data correlations were used to improve the accuracy of models, while optimization models were used to maximize company profit by predicting future tendencies based on historical data. The project provided us with valuable experience of what can be required of us in a work environment.

References

OpenAI, 2023. ChatGPT. San Francisco: OpenAI. Available at: https://chat.openai.com.

Stellenbosch University, QA344 Statistics.pdf