Deliverable 1

By J. G. Kurz – 27177416

Due 1 October 2025

Table of Contents:

Introduction

This report contains the analysis of the product, customer, and sales information from a technology company's database. It is written from the perspective of a data analysist with the aim of guiding business decisions which the company's upper management should take. Specific business issues are investigated such as cost, profitability and efficiency. The analysis was done by utilizing the programming language R for specific graphs and patterns in the data to be discovered, therefore the evidence of what is discussed throughout the report can be credited to and is shown by data visualization from R Studio.

To begin with, the information from the company's database has been captured into four different datasets, named as "Customer data", "Product data", "Products head office" and "Sales of 2022 and 2023" thus, these different datasets contain descriptive features as columns with helpful information into the business methodology, and will be evaluated to decipher how each one influences decision making. For this reason, the sections of the report which follow will be structured according to the different datasets considered.

Customer Data

Undoubtably, it is essential for the upper management of any company to understand the demographic and financial profile of their company's customer base. Thus, an evaluation was conducted to find important information which could concentrate the target market. More specifically, it was found that the average customer age is 51.55 years, the average customer income is $80797.00, the counts for customer genders were Female - 2432, Male -2350 and other - 218 and the counts for cities were Chicago – 724, Houston – 724, Los Angeles – 726, Miami – 647, New York – 726, San Francisco – 780, Seattle – 673.

Due to the ratio of female to male customers being significantly similar it can be inferred that the company generates profit from both genders, however; the data reveals that customers around the age of 50 years with an income of approximately $80000.00 who live in San Francisco, can be identified as the company's target market. By creating a niche/target market the company will be able to tailor its entire strategy, from product development and pricing to marketing and advertising, to meet this group's precise needs. Any business can cater to a specific market segmentation by focusing market research on value propositioning, which finds the precise product attributes that will guarantee market attraction. This will save the company's limited resources, such as time and money, and ensure effective communication with the people who are most likely to buy their products.

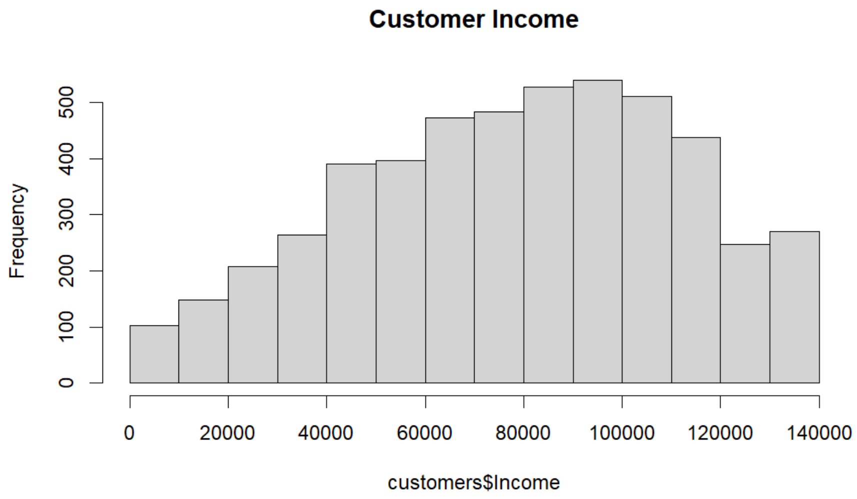


Figure 1

The histogram above shows a right-skewed distribution of customer income. Most customers have an income in the lower to middle range, and there is a long "tail" extending to the right, indicating a smaller number of customers with very high incomes. Interestingly, this suggests that while the company's customer base is largely composed of individuals with moderate incomes, there is a small, high-income segment that could be a target for premium products or marketing campaigns. The presence of these high-income individuals is a valuable finding for market segmentation and can be considered as a sub target market the company can cater to.

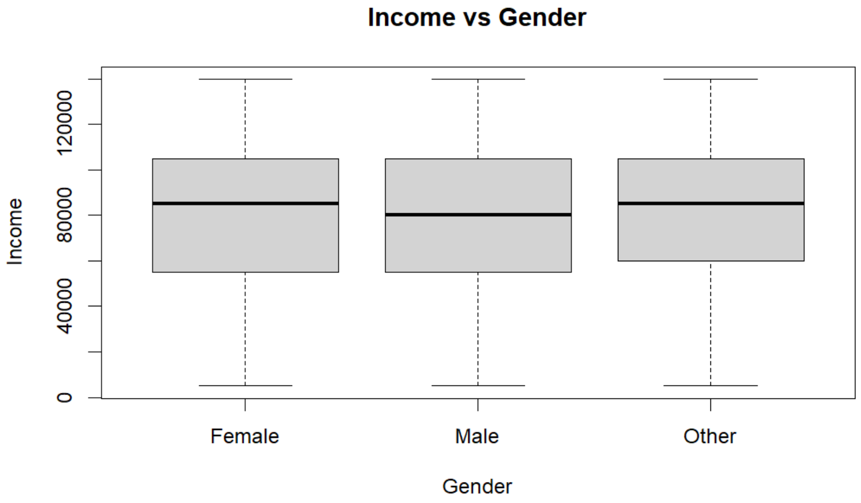


Figure 2

Furthermore, the boxplot in the figure above allows for a visual comparison of the income distribution between different genders. By analysing the median income (the line inside the box) it is understood that the mean income for females is slightly higher than for males which suggests there is no potential income disparity between the groups. Secondly, by looking at the whiskers and outliers (the dots outside the whiskers) clearly no gender has a wider spread of income or more high-income individuals than the other. Consequently, the target market will not be affected according to genders.

Product Data

The next important step is to evaluate the company product information to make informed business decisions, understand customer behaviour and stay competitive in the market. From summary statistics done on the product data it is noted that the minimum selling price of the products has a value of $350.40, the maximum selling price is valued at $19725.20, and the average selling price is $4493.60, while the mark-up percentages of the products have a minimum of 10.13%, a maximum of 29.84% and an average of 20.46%.

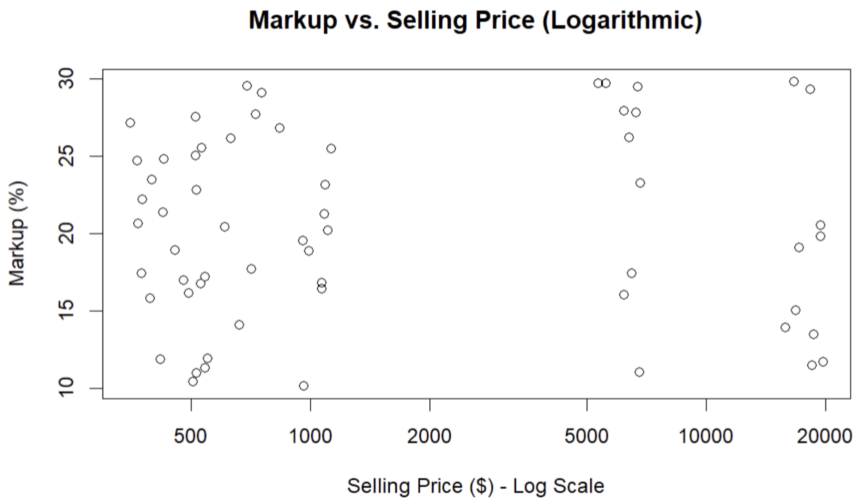


Figure 3

Figure 3 is a logarithmic scatter plot which reveals that higher mark-ups occur at products listed with lower selling prices, which can be explained from the trend in customer income previously identified. Since most people who buy the products are characterised by lower to middle ranged income levels, they would be more inclined to buying lower priced products. A high mark-up indicates a larger profit margin on each product sold, while a low mark-up suggests a smaller profit per product. Since the average mark-up percentage per product is around 20%, the company should focus their attention on the pricing strategy and profitability of these products because these products are the core of the business. Focusing on this large segment will allow the company to improve its overall profitability, since even a small increase in the markup for a high-volume product can lead to a great increase in total revenue. On the contrary, based on the company's business goals a decrease in market can strategically be used to gain market share and become more competitive.

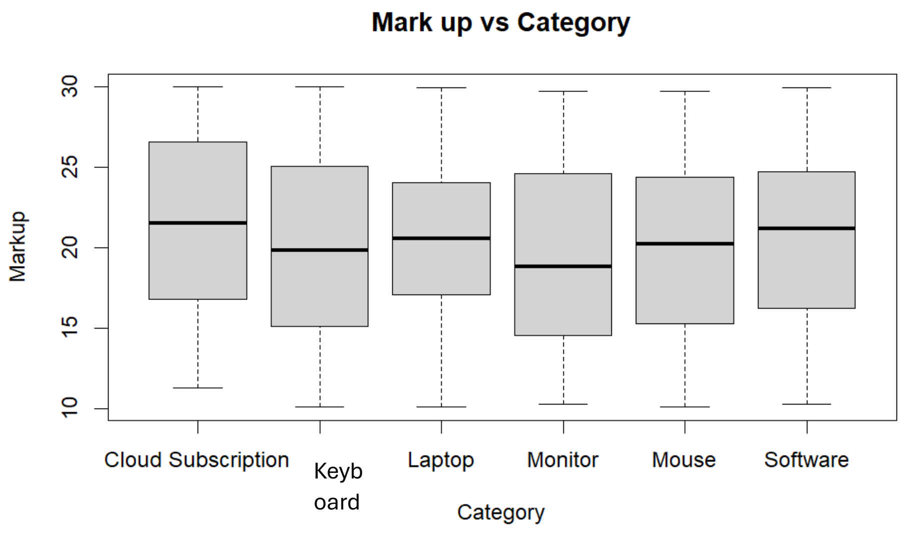


Figure 4

This box plot compares the markup distribution across different product categories and clearly shows that variability exists. By considering the median markup for each category it can be inferred that Cloud Subscription has the highest median markup, while Monitor has the lowest median markup. While Laptop, Software and Mouse are considered to have the average markups, these products along with Cloud Subscription can be classified to be, on average, more profitable than Keyboard and Monitor. Therefore, the previous point is realized by focusing the markup products on these categories, and upper management can decide whether to increase or decrease the markups based on the business goals.

Product Head Office

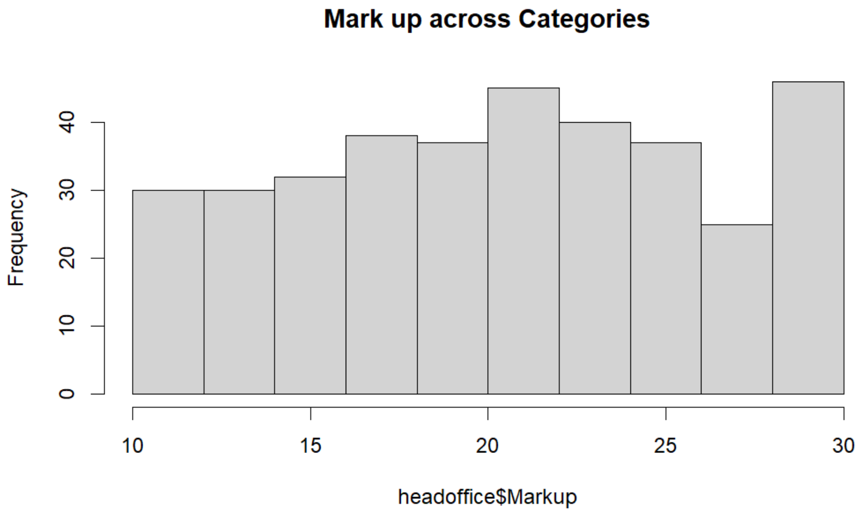
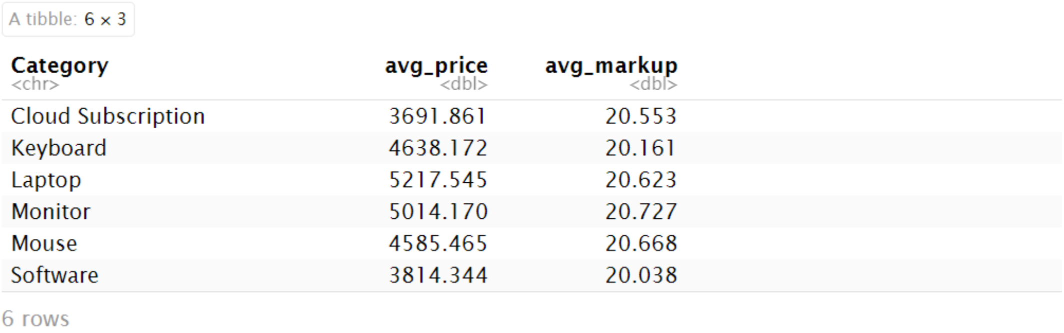


Figure 5

Moreover, it is revealed through the histogram above that while a significant number of products have a markup in the 20-25% range, there is a wide distribution of markups across the entire product line. This suggests that a single, uniform pricing strategy is not being applied. For business decisions, this means the company should analyse why certain products have a much higher or lower markup than the average. A particular suggestion is to consider increasing the markup on products currently priced below the most frequent range to improve profitability, provided there's sufficient market flexibility to do so. Also, management should continue evaluating each category price individually, since a one-size-fits-all approach to pricing is plainly not ideal. This will level out the markup gain from each product category and maximize company earning potential.



A tibble: 6 × 3

Category <chr>	avg_price <dbl>	avg_markup <dbl>
Cloud Subscription	3691.861	20.553
Keyboard	4638.172	20.161
Laptop	5217.545	20.623
Monitor	5014.170	20.727
Mouse	4585.465	20.668
Software	3814.344	20.038

6 rows

Table 1

According to the table above, different product categories have distinct average markups and prices. From analysis it can be advised that upper management should investigate implementing a differential pricing strategy in which the company sets specific markup goals for each product category based on market demand, competitive landscape, and operational costs. Additionally, a more thorough investigation can be conducted on the products with a wide markup range to see if the variability is intentional or if there are pricing inefficiencies that can be corrected to improve margins.

Sales of 2022 and 2023

Analysing company sales is essential for transforming raw numbers into actionable strategies such as being able to predict future trends, optimize resources, and improve overall performance. Henceforth, the company sales over the years 2022 and 2023 are inspected.

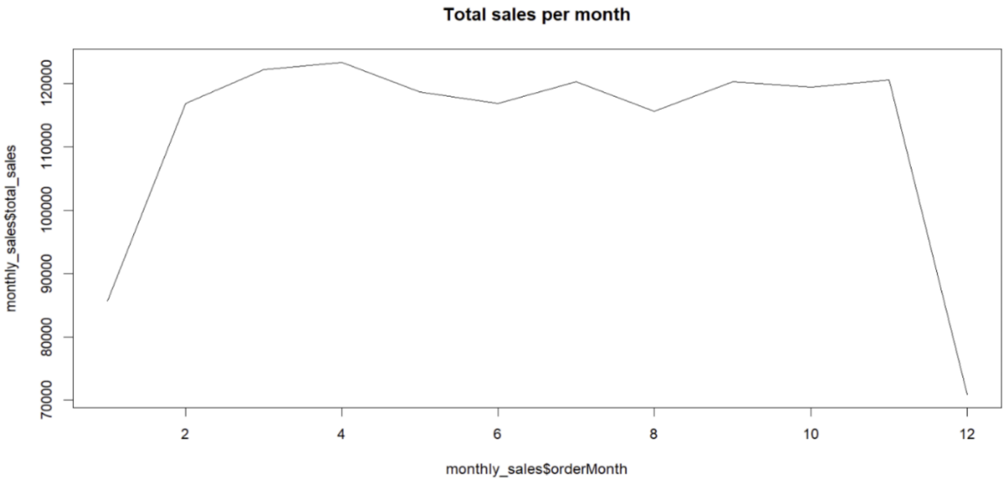
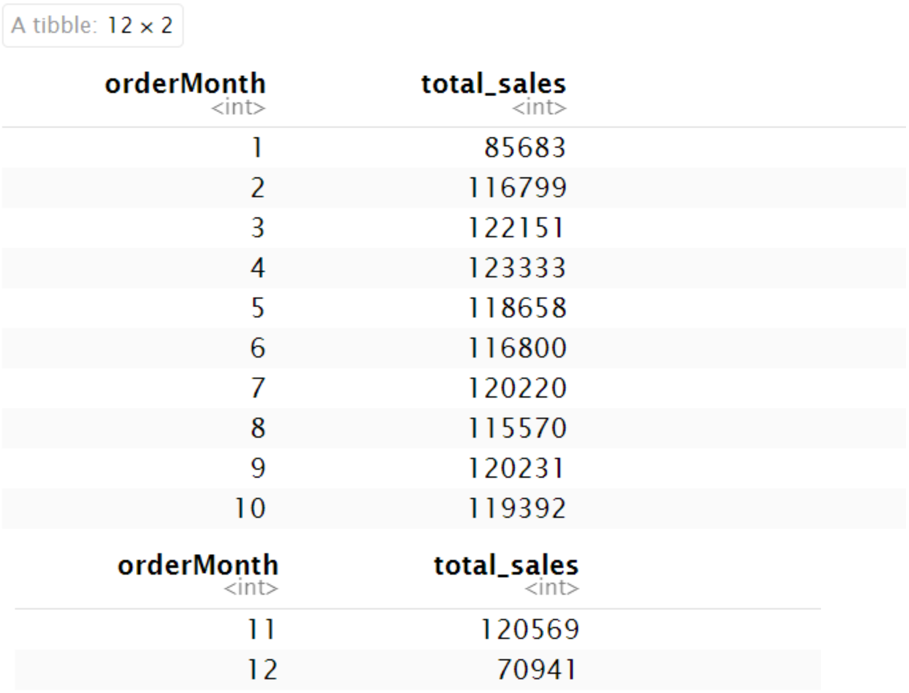


Figure 6

This plot shows sales trends over time, specifically by month. It is one of the most important plots to use when analysing company sales, as it reveals seasonality and overall trends to determine whether sales increase or decrease over a specific period. A significant finding is the clear upward trend seen between the first two months of the year, and a downward trend over the last two months of the year. With this in mind, the company can plan for the inventory and staffing needs according to the predicted demand over each period. The data reveals a distinct seasonal sales trend with a peak in early spring (March-April) and a significant slump in December. This pattern is crucial for sales forecasting and resource allocation.



A tibble: 12 × 2

orderMonth <int>	total_sales <int>
1	85683
2	116799
3	122151
4	123333
5	118658
6	116800
7	120220
8	115570
9	120231
10	119392
11	120569
12	70941

Table 2 & 3

By analysing the two tables comparing the order month to the total sales, the sales trends and seasonality from the sales plot is confirmed with specific values. It is noted that January and December have the lowest sales, with a steep rise from January to February and steep decline from November to December, and a steadiness in sales between the months February to November. The steep sales decline in December could be an indicator of market behaviour, but operational capacity could also be considered as a factor. Expressly, it is possible an explanation is that the company faced logistics challenges that affected customer experience and sales. The upper management should consider running targeted promotions or discounts during the traditionally slower months, to counteract the sales slump and level out revenue. Thereupon, the sales and fulfilment teams should be appropriately staffed to manage the increased volume during the high-demand periods to prevent operational bottlenecks and maintain customer satisfaction. Undoubtably, by planning for inventory and staffing needs is essential for keeping the company prepared for any situation.

Conclusion

To conclude, this analysis report provides a data-driven foundation for strategic business decisions. By examining key metrics from customer demographics to product profitability and sales performance, several actionable insights have been uncovered that can drive future growth and improve customer satisfaction.

The initial analysis of customer data has provided a clear picture of the company's target market: individuals around 50 years old, with an approximate annual income of $80,000. Furthermore, the data identified a small, high-income segment that represents an opportunity for premium product lines or specialized marketing campaigns. By management tailoring the company strategy to meet the precise needs of these segments, they can optimize their resources and ensure effective communication with the most likely buyers.

The evaluation of product data confirms the importance of a nuanced, data-driven pricing strategy. While the average markup is around 20%, a significant variability exists across different product categories. This finding suggests that a single, uniform pricing strategy is not ideal. Instead, a differential pricing strategy should be implemented, setting specific markup goals for each product category based on market demand, competitive landscape, and operational costs. A more thorough investigation into products with wide markup ranges can reveal pricing inefficiencies that can be corrected to improve margins.

Finally, the analysis of sales trends reveals a distinct seasonal pattern, with sales peaking in the first half of the year and slumping towards the end. This understanding is crucial for sales forecasting, which will allow for more effective inventory management, resource allocation, and targeted marketing. By cross-referencing sales data with operational metrics, it is possible to also identify and address potential bottlenecks in the supply chain to not only improve efficiency but also enhance the overall customer experience.

In summary, the data confirms that a granular, data-driven approach is essential for making informed decisions. By continuing to analyse product-level profitability and aligning sales strategies with operational capabilities, the company can capitalize on its strengths, correct its weaknesses, and build a more resilient and profitable business.

Appendix

- Figure 1: “Customer Income”
- Figure 2: “Income vs Gender”
- Figure 3: “Markup vs Selling Price (Logarithmic)”
- Figure 4: “Mark up vs Category”
- Figure 5: “Mark up across Categories”
- Table 1
- Figure 6: “Total sales per month”
- Table 2 & 3

Deliverable 2

By J. G. Kurz – 27177416

Due 9 October 2025

Table of Contents:

Introduction

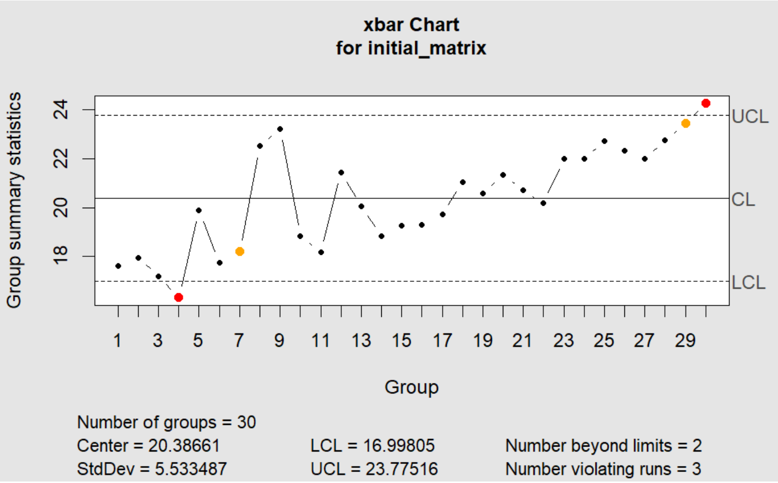
This Statistical Process Control report contains the delivery time analysis of a company's product range within the years 2022 and 2023. Using R and the 'qcc' package, control charts and capability indices were generated to assess process stability and performance. The analysis is grounded in data visualization and statistical modelling, enabling evidence-based conclusion.

(3.1) Control Charts for Delivery Times

The objective of this section was to construct X̄ (mean) and s (standard deviation) control charts for each product type using the first 30 samples of 24 observations each totalling to 720 data points per product type. These charts establish the initial control limits for monitoring the delivery time process.

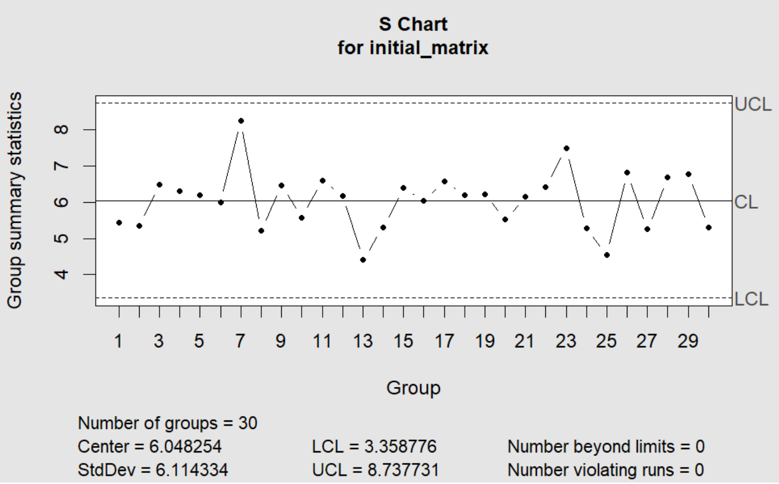
Given a dataset 'sales2026and2027Future.csv', the first step was to order the information chronologically by year, month, day, and time to simulate real-time arrival of sales data. Each observation corresponds to a single delivery, and data were grouped by their product types. For effective analysis, sampling was carried out for each product type. The following definitions were set up. Sample size: n = 24 deliveries per sample, Number of samples for setup: 30, Total used for initial setup: 720 deliveries (30×24).

Using R (with base functions and the 'qcc' package), the following control limits were calculated from the first 30 samples. Where constants for $n = 24$ and $A_3 = 0.619$, $B_3 = 0.555$, $B_4 = 1.445$. X̄ and s charts were created successfully for product MOU059 as represented in the figures below.



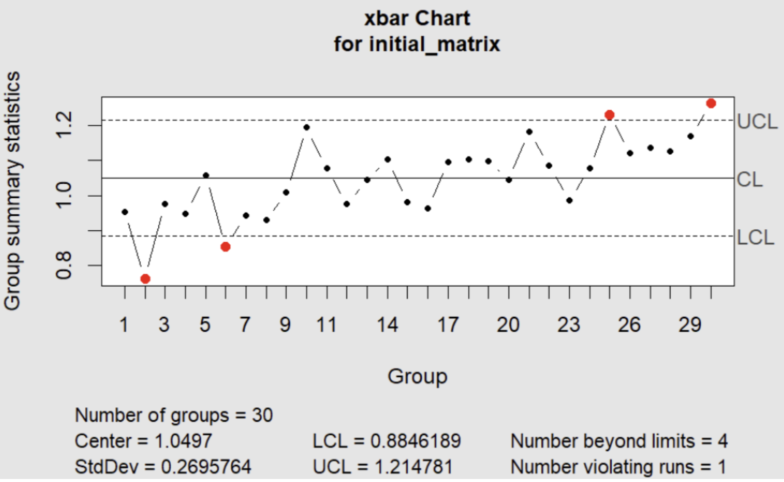
Number of groups = 30
Center = 20.38661
StdDev = 5.533487
LCL = 16.99805
UCL = 23.77516
Number beyond limits = 2
Number violating runs = 3

Figure 7



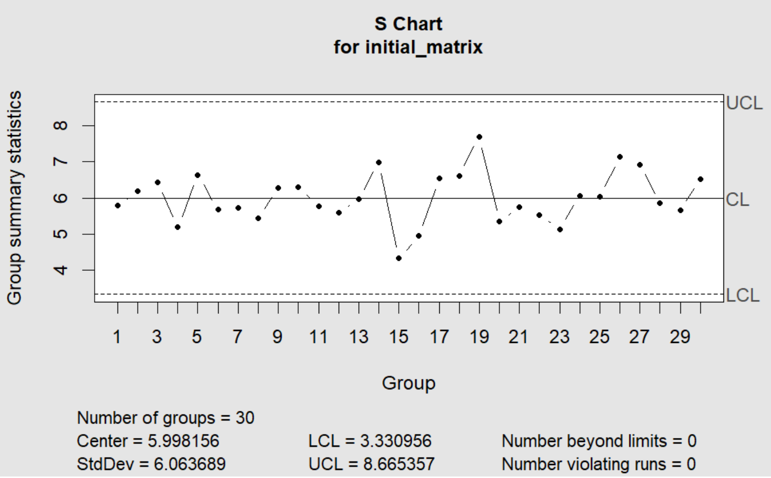
Number of groups = 30
Center = 6.048254
StdDev = 6.114334
LCL = 3.358776
UCL = 8.737731
Number beyond limits = 0
Number violating runs = 0

Figure 8



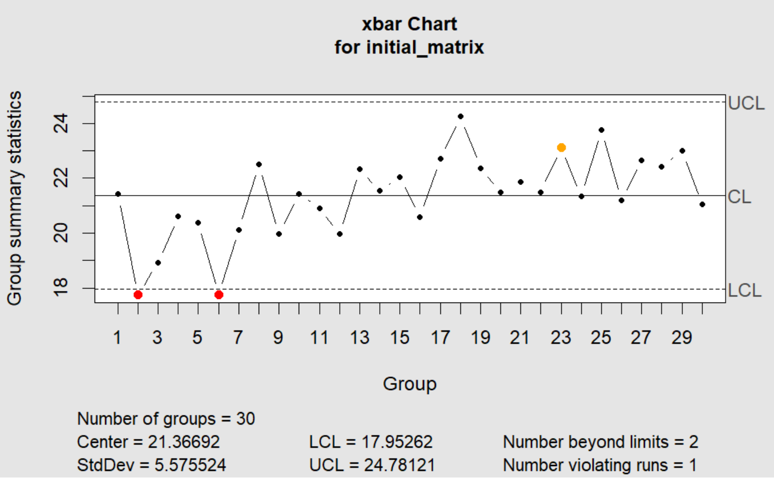
Number of groups = 30
Center = 1.0497
StdDev = 0.2695764
LCL = 0.8846189
UCL = 1.214781
Number beyond limits = 4
Number violating runs = 1

Figure 9



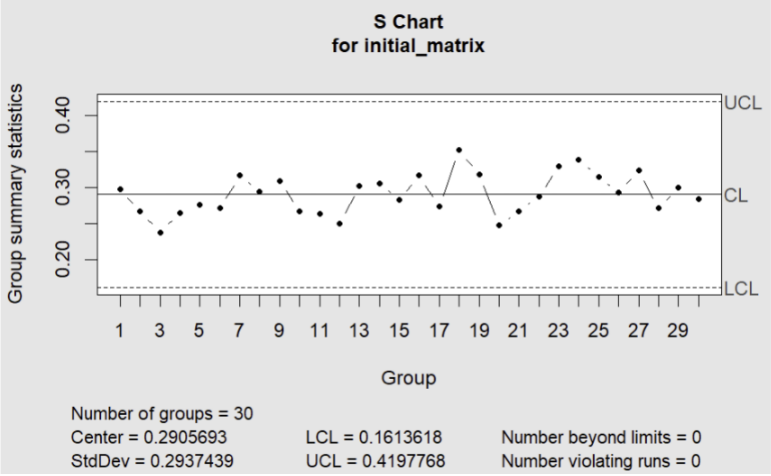
Number of groups = 30
Center = 5.998156
StdDev = 6.063689
LCL = 3.330956
UCL = 8.665357
Number beyond limits = 0
Number violating runs = 0

Figure 10



Number of groups = 30
Center = 21.36692
StdDev = 5.575524
LCL = 17.95262
UCL = 24.78121
Number beyond limits = 2
Number violating runs = 1

Figure 5



Number of groups = 30
Center = 0.2905693
StdDev = 0.2937439
LCL = 0.1613618
UCL = 0.4197768
Number beyond limits = 0
Number violating runs = 0

Figure 6

The product selected as an example illustration for the report is product MOU 059. Only 6 distinct figures have been included for a conclusion to be drawn, since the graphs repeat themselves adding up to 16 in total for this product. From the X̄ charts in figures 1, 3 and 5 which show how the average delivery time changes over time per product type, points are mostly within the upper and lower control limits with +/-2 out-of-control points seen in each graph. These out-of-control points may suggest special cause variation at that point in time. Figures 2, 4 and 6 are S charts which show how process variability changes over time per product type.

They clearly emphasize consistent clustering within control limits, and common cause variation seen in normal fluctuations. The X̄ charts reveal that the process is mostly stable due to few out-of-control points noticed. The S charts reveal that there is low variation since there is a close distance between upper and lower control limits making the process mean close to the centreline.

(3.2) Monitoring Future Samples

The objective of this section is to simulate ongoing process monitoring using subsequent samples (sample 31 onwards), ordered chronologically per product type. After establishing control limits, it was followed that new data were processed in sample groups of 24. For each new sample, the sample mean, and standard deviation were plotted on existing control charts. Then, deviations beyond control limits were flagged as potential process issues.

Evidently, the extended X̄ charts for 3 arbitrary products were analysed to compare insights such as representativeness and practical contexts. Namely, product types MOU059, SOF009 and KEY049 were selected, and all their charts agree that 30 samples are not enough to estimate control limits reliably since there are multiple points above/below the upper/lower control limits in each case respectively. It was then interpreted that the samples exceeding the +3σ limit, indicating sporadic increases in delivery time variability.

Additionally, the widespread between control limits in product types MOU059 and KEY049 indicate higher variability in delivery time than SOF009, which remained consistently within control, showing process stability. This inconsistency in delivery performance possibly due to longer supply chains.

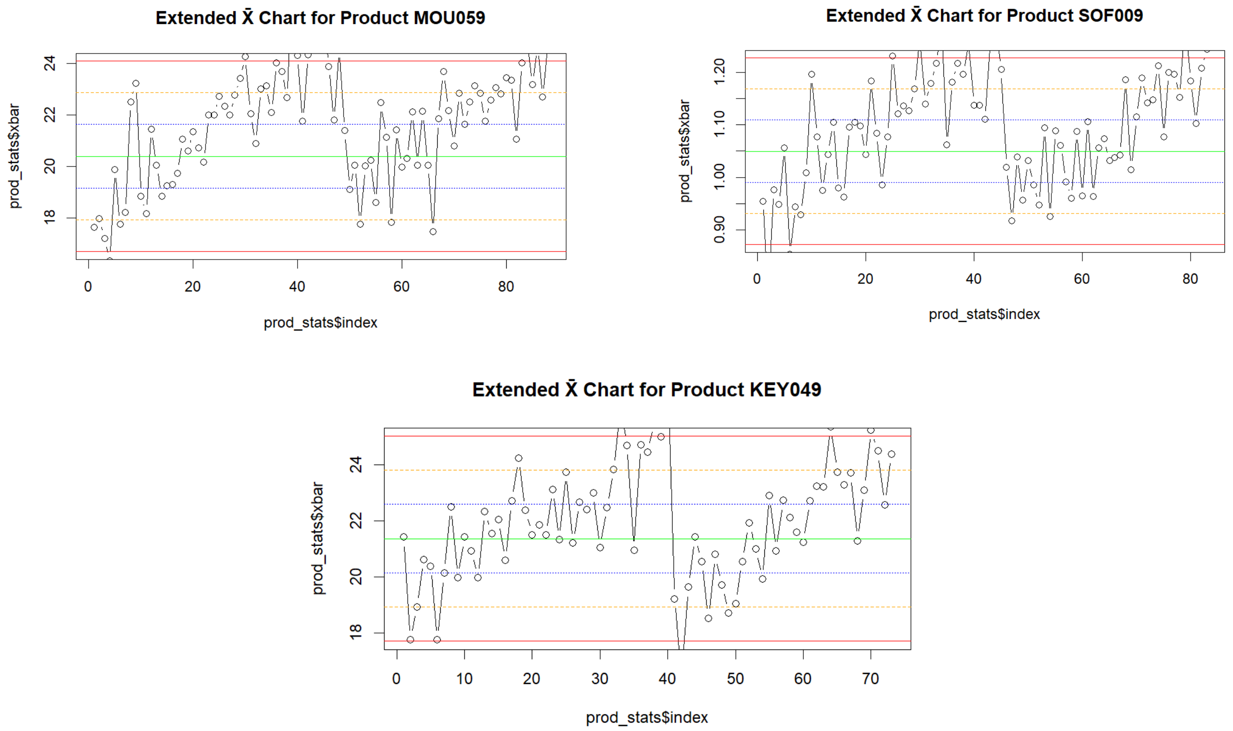


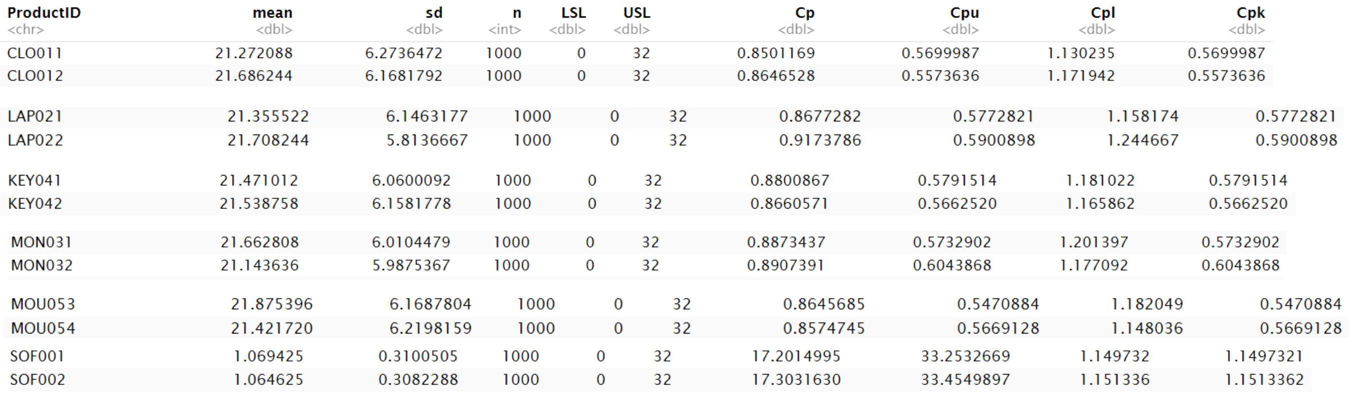
Figure 7, Figure 8 and Figure 9

(3.3) Process Capability Indices

In this section the objective was to determine which product types could meet customer expectations (Voice of the Customer, VOC) through the capability analysis of the delivery process. Specification Limits were set accordingly. Lower Specification Limit (LSL): 0 hours, and Upper Specification Limit (USL): 32 hours based on the first 1000 deliveries per product type.

Formulas:

Index	**Formula**
$C_p = \frac{USL - LSL}{6\sigma}$	Process potential capability
$C_{pl} = \frac{\bar{X} - LSL}{3\sigma}$	Capability near LSL
$C_{pu} = \frac{USL - \bar{X}}{3\sigma}$	Capability near USL
$C_{pk} = \min(C_{pl}, C_{pu})$	Actual capability



ProductID	mean	sd	n	LSL	USL	Cp	Cpu	Cpl	Cpk
<chr>	<dbl>	<dbl>	<int>	<dbl>	<dbl>	<dbl>	<dbl>	<dbl>	<dbl>
CLO011	21.272088	6.2736472	1000	0	32	0.8501169	0.5699987	1.130235	0.5699987
CLO012	21.686244	6.1681792	1000	0	32	0.8646528	0.5573636	1.171942	0.5573636
LAP021	21.355522	6.1463177	1000	0	32	0.8677282	0.5772821	1.158174	0.5772821
LAP022	21.708244	5.8136667	1000	0	32	0.9173786	0.5900898	1.244667	0.5900898
KEY041	21.471012	6.0600092	1000	0	32	0.8800867	0.5791514	1.181022	0.5791514
KEY042	21.538758	6.1581778	1000	0	32	0.8660571	0.5662520	1.165862	0.5662520
MON031	21.662808	6.0104479	1000	0	32	0.8873437	0.5732902	1.201397	0.5732902
MON032	21.143636	5.9875367	1000	0	32	0.8907391	0.6043868	1.177092	0.6043868
MOU053	21.875396	6.1687804	1000	0	32	0.8645685	0.5470884	1.182049	0.5470884
MOU054	21.421720	6.2198159	1000	0	32	0.8574745	0.5669128	1.148036	0.5669128
SOF001	1.069425	0.3100505	1000	0	32	17.2014995	33.2532669	1.149732	1.1497321
SOF002	1.064625	0.3082288	1000	0	32	17.3031630	33.4549897	1.151336	1.1513362

Table 1

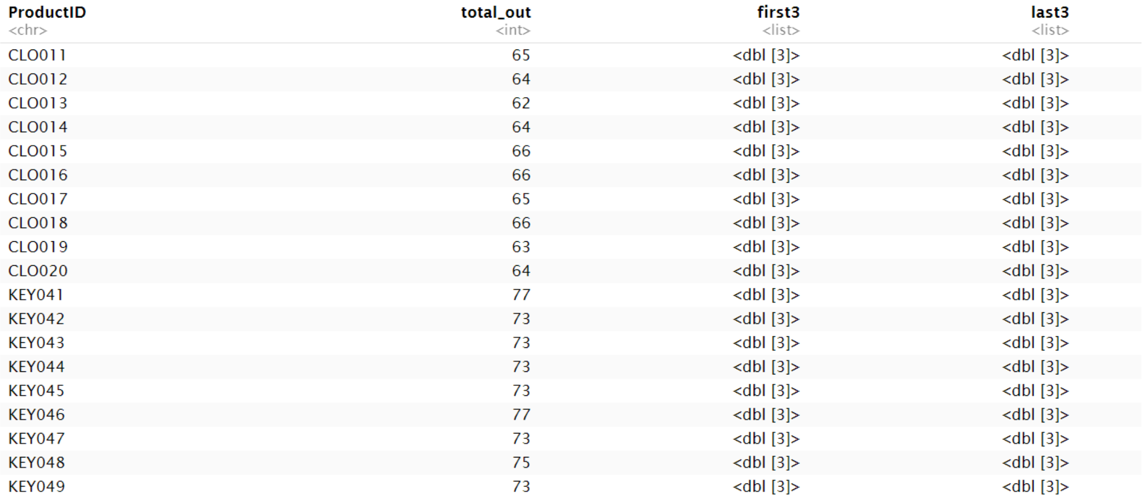
Each product type was investigated according to the mean, standard deviation and capability indices to compare the process spread across products. Table 1 displays only two instances of each product type for discussion. According to the standard that Cp > 1 indicates that the process variation is within acceptable limits, and Cpk > 1 indicates that the process is centred and capable, it is noted that none of the product types of CLO, LAP, KEY, MON or MOU conform to these values since they all show Cp/Cpk < 1. From the frequent deviations beyond specifications, it can be demonstrated that the process means are off-centre and too variable, warranting further process investigation. Only product type SOF which has Cp >> 1 and Cpk > 1 signifies the process is capable, therefore this product is identified as being able to meet the VOC.

(3.4) Identification of Control Issues

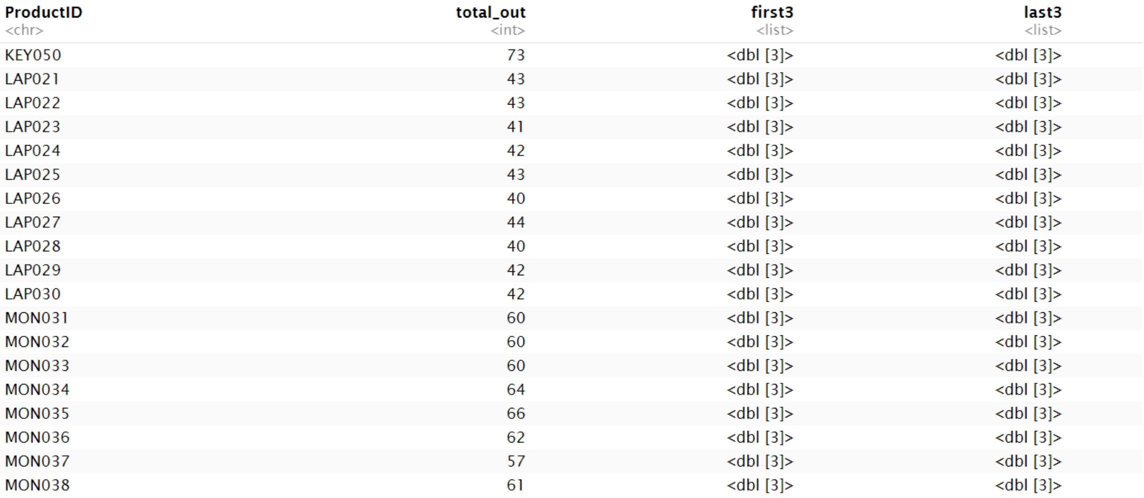
In this section the goal was to identify samples that show process control issues according to set of rules A, B and C. Each rule is set to statistically analyse “sales2026and2027.csv” according to SPC to determine useful insight into the data.

A. Samples Outside +3σ Control Limits

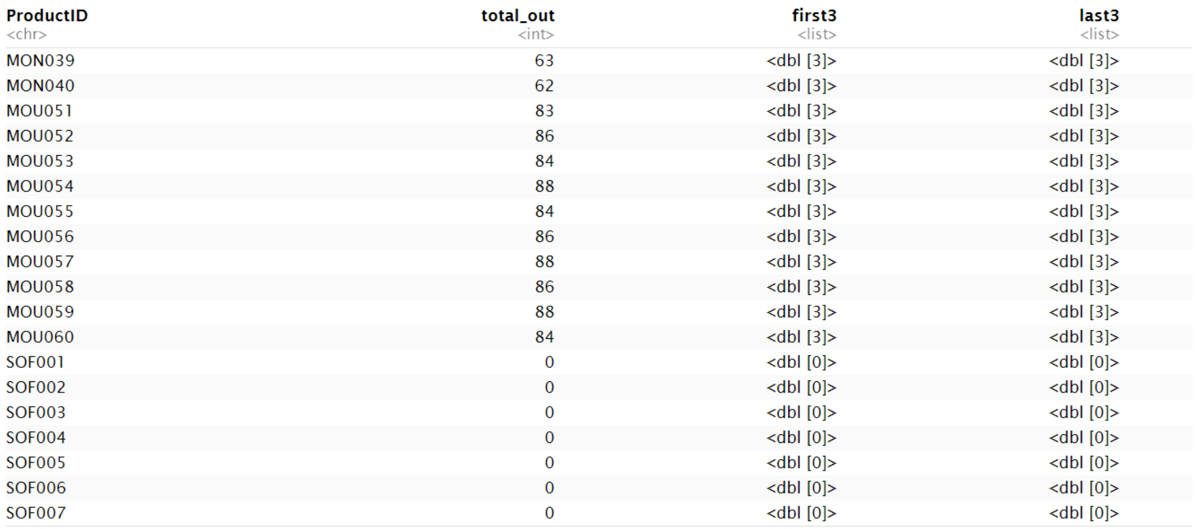
The table below displays the total out-of-control samples above +3σ per product, which may suggest special cause variation. These indicate unusually high variability subgroups which can be used to find instability or a spike in variation for a product. By evaluating the process according to Rule A, it was identified that almost all subgroups exceed +3σ, indicating frequent spikes in process variability for all product types except those with ID “SOF”. Specifically, the range of 40 to 88 samples exceeded the upper control limit, indicating extreme cases of excessive process variability. These may correspond to measurement errors or short-term instability and warrant investigation.



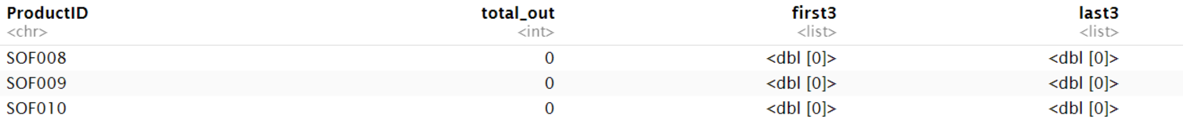
ProductID	total_out	first3	last3
<chr>	<int>	<list>	<list>
CLO011	65	<dbl [3]>	<dbl [3]>
CLO012	64	<dbl [3]>	<dbl [3]>
CLO013	62	<dbl [3]>	<dbl [3]>
CLO014	64	<dbl [3]>	<dbl [3]>
CLO015	66	<dbl [3]>	<dbl [3]>
CLO016	66	<dbl [3]>	<dbl [3]>
CLO017	65	<dbl [3]>	<dbl [3]>
CLO018	66	<dbl [3]>	<dbl [3]>
CLO019	63	<dbl [3]>	<dbl [3]>
CLO020	64	<dbl [3]>	<dbl [3]>
KEY041	77	<dbl [3]>	<dbl [3]>
KEY042	73	<dbl [3]>	<dbl [3]>
KEY043	73	<dbl [3]>	<dbl [3]>
KEY044	73	<dbl [3]>	<dbl [3]>
KEY045	73	<dbl [3]>	<dbl [3]>
KEY046	77	<dbl [3]>	<dbl [3]>
KEY047	73	<dbl [3]>	<dbl [3]>
KEY048	75	<dbl [3]>	<dbl [3]>
KEY049	73	<dbl [3]>	<dbl [3]>



ProductID	total_out	first3	last3
<chr>	<int>	<list>	<list>
KEY050	73	<dbl [3]>	<dbl [3]>
LAP021	43	<dbl [3]>	<dbl [3]>
LAP022	43	<dbl [3]>	<dbl [3]>
LAP023	41	<dbl [3]>	<dbl [3]>
LAP024	42	<dbl [3]>	<dbl [3]>
LAP025	43	<dbl [3]>	<dbl [3]>
LAP026	40	<dbl [3]>	<dbl [3]>
LAP027	44	<dbl [3]>	<dbl [3]>
LAP028	40	<dbl [3]>	<dbl [3]>
LAP029	42	<dbl [3]>	<dbl [3]>
LAP030	42	<dbl [3]>	<dbl [3]>
MON031	60	<dbl [3]>	<dbl [3]>
MON032	60	<dbl [3]>	<dbl [3]>
MON033	60	<dbl [3]>	<dbl [3]>
MON034	64	<dbl [3]>	<dbl [3]>
MON035	66	<dbl [3]>	<dbl [3]>
MON036	62	<dbl [3]>	<dbl [3]>
MON037	57	<dbl [3]>	<dbl [3]>
MON038	61	<dbl [3]>	<dbl [3]>



ProductID <chr>	total_out <int>	first3 <list>	last3 <list>
MON039	63	<dbl [3]>	<dbl [3]>
MON040	62	<dbl [3]>	<dbl [3]>
MOU051	83	<dbl [3]>	<dbl [3]>
MOU052	86	<dbl [3]>	<dbl [3]>
MOU053	84	<dbl [3]>	<dbl [3]>
MOU054	88	<dbl [3]>	<dbl [3]>
MOU055	84	<dbl [3]>	<dbl [3]>
MOU056	86	<dbl [3]>	<dbl [3]>
MOU057	88	<dbl [3]>	<dbl [3]>
MOU058	86	<dbl [3]>	<dbl [3]>
MOU059	88	<dbl [3]>	<dbl [3]>
MOU060	84	<dbl [3]>	<dbl [3]>
SOF001	0	<dbl [0]>	<dbl [0]>
SOF002	0	<dbl [0]>	<dbl [0]>
SOF003	0	<dbl [0]>	<dbl [0]>
SOF004	0	<dbl [0]>	<dbl [0]>
SOF005	0	<dbl [0]>	<dbl [0]>
SOF006	0	<dbl [0]>	<dbl [0]>
SOF007	0	<dbl [0]>	<dbl [0]>

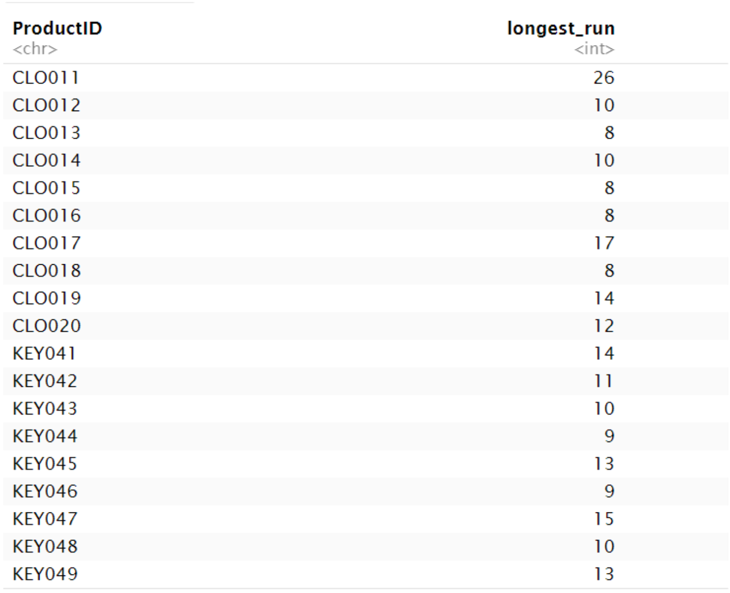


ProductID <chr>	total_out <int>	first3 <list>	last3 <list>
SOF008	0	<dbl [0]>	<dbl [0]>
SOF009	0	<dbl [0]>	<dbl [0]>
SOF010	0	<dbl [0]>	<dbl [0]>

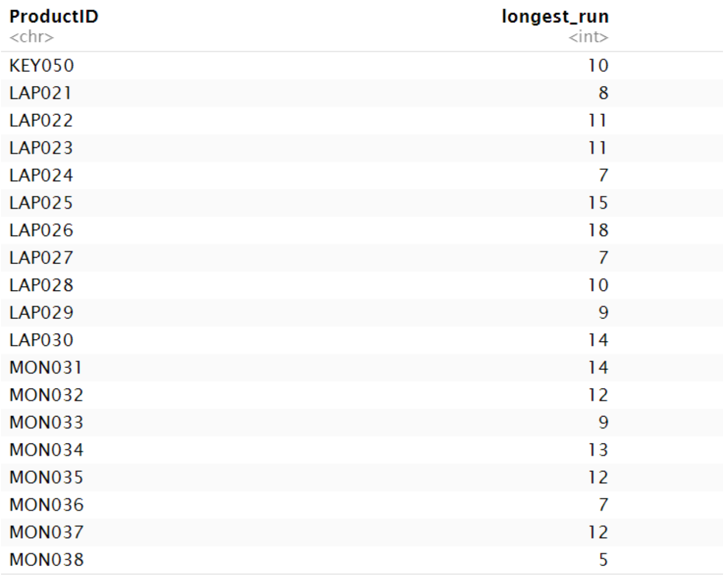
Table 2

B. Longest Sequence Within ±1σ

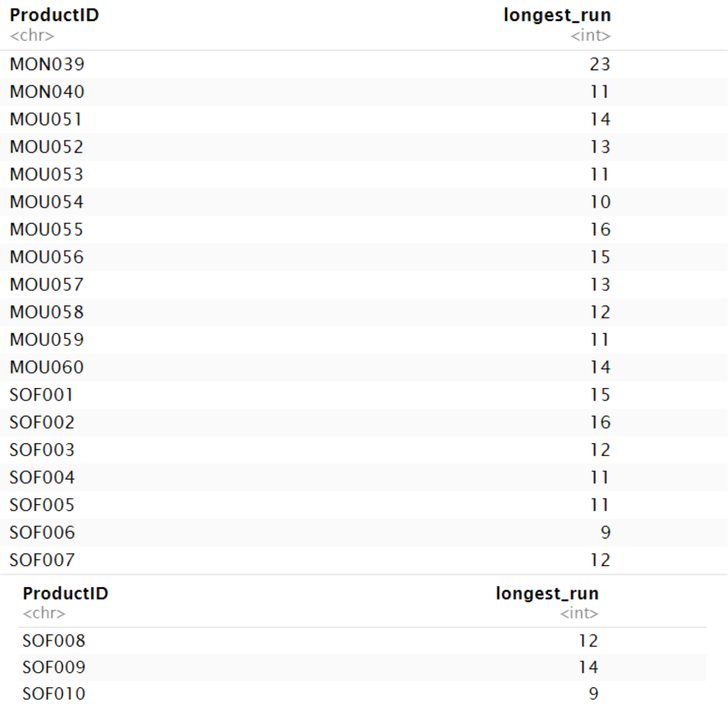
A long run within ±1σ indicates excellent short-term control and good process stability thus it is valuable to consider. Consistent process variability is a sign that there are no control issues with the data. Code in R is used to find the longest run length per product since the longer the run, the more stable the products variation. Under Rule B, the table below shows that the products with ID “CLO” demonstrated the longest stable sequence of 26 consecutive subgroups within ±1σ, reflecting strong process consistency. Furthermor, the longest runs for each product type under product ID “KEY” were 15, 18 for product ID “LAP”, 23 for product ID “MON”, 16 for product ID “MOU” and for product ID “SOF”.



ProductID <chr>	longest_run <int>
CLO011	26
CLO012	10
CLO013	8
CLO014	10
CLO015	8
CLO016	8
CLO017	17
CLO018	8
CLO019	14
CLO020	12
KEY041	14
KEY042	11
KEY043	10
KEY044	9
KEY045	13
KEY046	9
KEY047	15
KEY048	10
KEY049	13



ProductID <chr>	longest_run <int>
KEY050	10
LAP021	8
LAP022	11
LAP023	11
LAP024	7
LAP025	15
LAP026	18
LAP027	7
LAP028	10
LAP029	9
LAP030	14
MON031	14
MON032	12
MON033	9
MON034	13
MON035	12
MON036	7
MON037	12
MON038	5

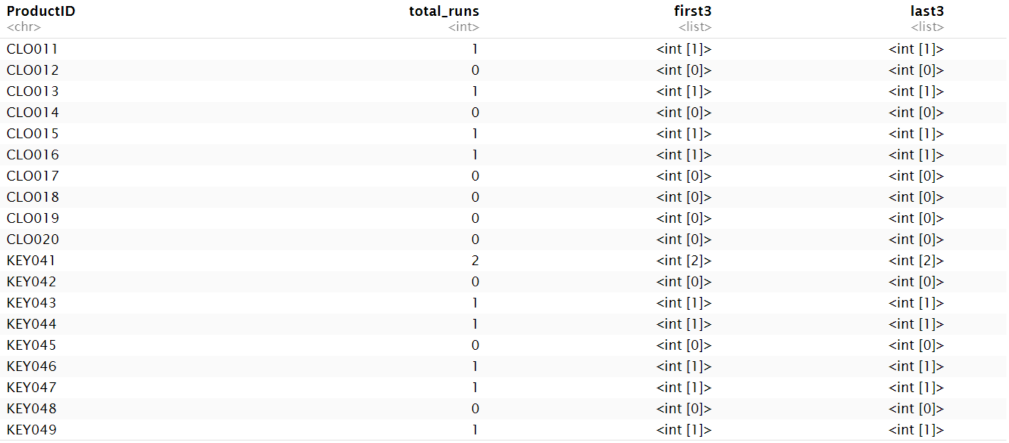


ProductID	longest_run
<chr>	<int>
MON039	23
MON040	11
MOU051	14
MOU052	13
MOU053	11
MOU054	10
MOU055	16
MOU056	15
MOU057	13
MOU058	12
MOU059	11
MOU060	14
SOF001	15
SOF002	16
SOF003	12
SOF004	11
SOF005	11
SOF006	9
SOF007	12
SOF008	12
SOF009	14
SOF010	9

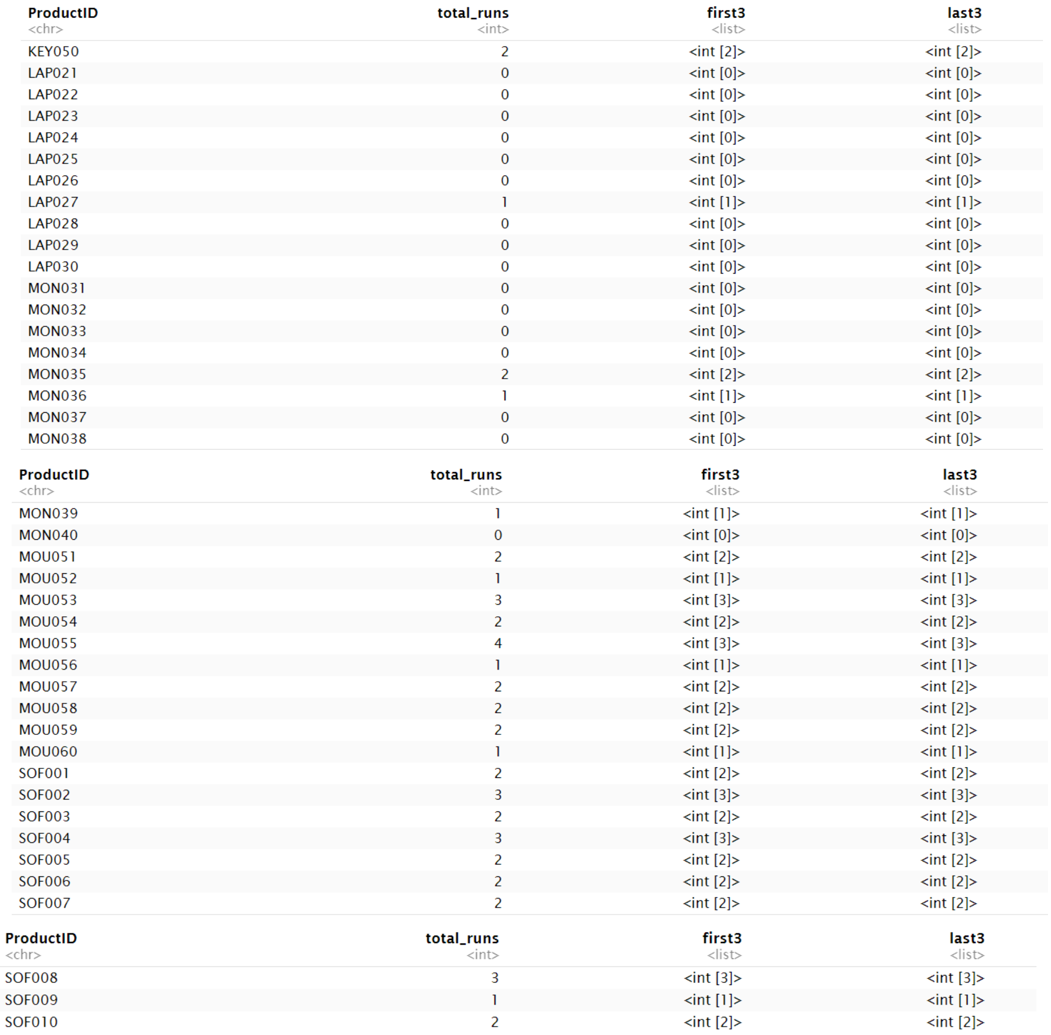
Table 3

C. 4 Consecutive X̄ Samples Outside ±2σ

Here it is important to find out whether there exist 4 or more samples above +2σ, because this signals consistent bias or process drift. Statistical measures in R can be used to detect a shift in the process mean that is still within control limits but persistent. Such results suggest potential upward drift or systematic bias. The table below portrays that from applying Rule C, MOU005 is the only product that experienced four consecutive subgroups above the +2σ line, suggesting a potential upward drift in mean delivery time. An analysis of consecutive samples beyond the +2σ control limits shows isolated but notable deviations for product types with ID “CLO” and “KEY”, each with distinct instances of one run exceeding the threshold. This suggests that, at certain intervals, the process mean for these products temporarily shifted upward, possibly indicating special cause variation. Conversely, each product type exhibited instances of no such runs, confirming a consistent and well-controlled process. The presence of occasional mean shifts warrants investigation into production or operational factors during the identified periods.



ProductID	total_runs	first3	last3
<chr>	<int>	<list>	<list>
CLO011	1	<int [1]>	<int [1]>
CLO012	0	<int [0]>	<int [0]>
CLO013	1	<int [1]>	<int [1]>
CLO014	0	<int [0]>	<int [0]>
CLO015	1	<int [1]>	<int [1]>
CLO016	1	<int [1]>	<int [1]>
CLO017	0	<int [0]>	<int [0]>
CLO018	0	<int [0]>	<int [0]>
CLO019	0	<int [0]>	<int [0]>
CLO020	0	<int [0]>	<int [0]>
KEY041	2	<int [2]>	<int [2]>
KEY042	0	<int [0]>	<int [0]>
KEY043	1	<int [1]>	<int [1]>
KEY044	1	<int [1]>	<int [1]>
KEY045	0	<int [0]>	<int [0]>
KEY046	1	<int [1]>	<int [1]>
KEY047	1	<int [1]>	<int [1]>
KEY048	0	<int [0]>	<int [0]>
KEY049	1	<int [1]>	<int [1]>



ProductID	total_runs	first3	last3
<chr>	<int>	<list>	<list>
KEY050	2	<int [2]>	<int [2]>
LAP021	0	<int [0]>	<int [0]>
LAP022	0	<int [0]>	<int [0]>
LAP023	0	<int [0]>	<int [0]>
LAP024	0	<int [0]>	<int [0]>
LAP025	0	<int [0]>	<int [0]>
LAP026	0	<int [0]>	<int [0]>
LAP027	1	<int [1]>	<int [1]>
LAP028	0	<int [0]>	<int [0]>
LAP029	0	<int [0]>	<int [0]>
LAP030	0	<int [0]>	<int [0]>
MON031	0	<int [0]>	<int [0]>
MON032	0	<int [0]>	<int [0]>
MON033	0	<int [0]>	<int [0]>
MON034	0	<int [0]>	<int [0]>
MON035	2	<int [2]>	<int [2]>
MON036	1	<int [1]>	<int [1]>
MON037	0	<int [0]>	<int [0]>
MON038	0	<int [0]>	<int [0]>
MON039	1	<int [1]>	<int [1]>
MON040	0	<int [0]>	<int [0]>
MOU051	2	<int [2]>	<int [2]>
MOU052	1	<int [1]>	<int [1]>
MOU053	3	<int [3]>	<int [3]>
MOU054	2	<int [2]>	<int [2]>
MOU055	4	<int [3]>	<int [3]>
MOU056	1	<int [1]>	<int [1]>
MOU057	2	<int [2]>	<int [2]>
MOU058	2	<int [2]>	<int [2]>
MOU059	2	<int [2]>	<int [2]>
MOU060	1	<int [1]>	<int [1]>
SOF001	2	<int [2]>	<int [2]>
SOF002	3	<int [3]>	<int [3]>
SOF003	2	<int [2]>	<int [2]>
SOF004	3	<int [3]>	<int [3]>
SOF005	2	<int [2]>	<int [2]>
SOF006	2	<int [2]>	<int [2]>
SOF007	2	<int [2]>	<int [2]>
SOF008	3	<int [3]>	<int [3]>
SOF009	1	<int [1]>	<int [1]>
SOF010	2	<int [2]>	<int [2]>

Table 4

Conclusion

The Statistical Process Control (SPC) analysis conducted on delivery times across product types revealed several key insights into process stability and capability. Using R to construct and evaluate X̄ and s-charts, the analysis effectively identified both common cause and special cause variations across product lines. Overall, the methodology demonstrates how SPC can be used to continuously monitor, diagnose, and improve delivery performance across categories.

Indeed, the SPC analysis reveals that product types with ID “MOU” experience periodic instability due to special cause variation and frequently displayed high process variability. A suggested improvement is to conduct a supplier performance review and a root cause analysis to isolate and correct these deviations. Product types with ID “SOF” are considered stable and a capable process with excellent short-term control. These products displayed no samples violating key control rules and long runs of points within the ±1σ limits and indicates processes under good statistical control with minimal variability.

From a broader perspective, the SPC results underscore the importance of differentiating between natural (common) and assignable (special) causes of variation. Rule-based diagnostics provided quantitative evidence for when a process may be drifting out of control versus when it is performing optimally. Consequently, SPC tools such as $\bar{X}$ and s-charts, supported by capability indices and R-based visualization, form a powerful framework for maintaining consistent delivery performance, ensuring process reliability, and enabling data-driven continuous improvement within a quality assurance context.

Appendix

- Figure 1, 3 and 5: “xbar Chart for initial_matrix”
- Figure 2, 4 and 6: “S Chart for initial_matrix”
- Figure 7: “Extended $\bar{X}$ chart for Product MOU059”
- Figure 8: “Extended $\bar{X}$ chart for Product SOF009”
- Figure 9: “Extended $\bar{X}$ chart for Product KEY049”
- Table 1: Capability indices table
- Table 2: Samples Outside $+3\sigma$ Control Limits
- Table 3: Longest Sequence Within $\pm 1\sigma$
- Table 4: 4 Consecutive $\bar{X}$ Samples Outside $\pm 2\sigma$

Deliverable 3

By J. G. Kurz – 27177416

Due 16 October 2025

Table of Contents:

Introduction

To begin with, this report is defined as a continuation of the analysis of data as well as the solutions to different business problems such as a bottling process and the optimization of a coffee shop schedule. Each section as outlined in the table of contents covers a separate theme, however; section 4.1, 4.2 and 4.3 can be categorized into SPC analysis, and section 5 as into optimization. Topics such as risk uncertainty, optimizing for profit and the design of experiments were tackled to broaden the scope of possibilities within the breakdown and investigation of data.

4.1 Probability of making a type I error for A, B & C

Type I error is when a process is stopped even though the signal claims it is good. In calculating this SPC parameter, the following results were obtained making these assumptions. It was assumed that under H_0 the process was defined as in control and sample statistics follow the theoretical normal distributions implied by the control chart, and that 7 samples were produced in one week (one per day). Furthermore, the three rules were defined as A. one sample being outside the +3σ control limit, B. the most consecutive samples of s between the -1 and +1 sigma-control limits, and C. four consecutive X̄ samples above the +2σ limit.

To find Rule A's type I error it is noted that for a standard normal P(Z>3) = 1−Φ(3) ≈ 0.001350 which is the per sample type I error probability. Thus, it can be deduced that the probability of 1 sample occurring outside to positive 3 sigma control limit is 0.14%. Hence the probability of at least one false alarm in 7 independent samples is 1 - $(1-p)^7 \approx 1 - 0.9905 = \ 0.009411$ or 0.94% per week.

For Rule B, the type I error means to find the probability of concluding that the process is out of control when it is indeed fine. The chance of this happening is 0.6827^L per week where L is the length of the consecutive runs that fall within this statistic, which corresponds with the conclusions drawn from the Statistical Process Control analysis done on the data in part 3 of the report. The identification of products with the longest runs inside ±1σ representing the band wherein which the process signifies good control.

In finding the type I error for C, it is noted that the one-sided probability of one sample is P(Z>2) = 1−Φ(2) ≈ 0.022750. Hence the probability of 4 consecutive such events is approximately equal to $(0.02275013)^{4}$ ≈ 2.7e−07 per specific 4 sample run. Therefore, the type I error per week is 4×2.7e−07 ≈ 1.1e−06 or 0.00011%.

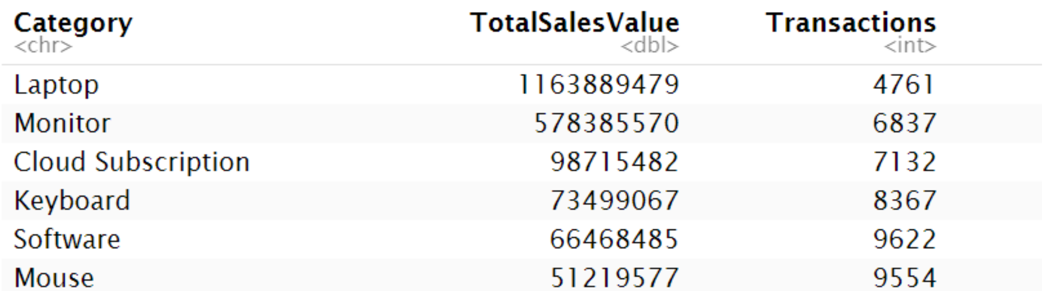
4.2 Probability of making a type II error for a bottling process

Here a new problem is considered, namely finding the type II error of a bottle filling process. Type II error is when a bad process claims a good output, and no out-of-control signals are seen. This problem is statistically understood as H_a being true but not noticed, thus the product still reaches the customer. Given the centre line (CL), original standard deviation and control limits (UCL and LCL) it is stated to find the probability of failing to detect a shift in the process. This can be translated as finding the probability that an observed X̄ falls inside [LCL, UCL] under the new distribution.

Therefore, the upper and lower z-scores can be computed as follows. $z_L = \frac{\text{LCL}-\mu'}{\sigma_{x'}} = \frac{25.011-25.028}{0.017} = \frac{-0.017}{0.017} = -1$ and $z_U = \frac{\text{UCL}-\mu'}{\sigma_{x'}} = \frac{25.089-25.028}{0.017} = \frac{0.061}{0.017} \approx 3.588235$. The type II error is the probability X̄ is between LCL and UCL under the shifted distribution and can be written as $\beta = \Phi(z_U) - \Phi(z_L) \approx 0.9998321-0.1586553 = 0.841177$ or 84%. Failing to detect an out-of-control process 84% of the time shows that the chart is not very sensitive to that shift.

4.3 Data correction and re-analysis of the head office

This section corrects a data file that has been previously analysed with errors included in the instances for product and head office information. Corrective action was taken by firstly importing the local files in R and updating each product into its correct category using the product's ID prefix. The first 10 rows were replicated across the rows 11 to 60. Then the analysis was re-run, and the corrected data was joined to the 2023 sales to compute the total sales value per type. Making sure each product type is consistently categorised across all related datasets improves data integrity and facilitates accurate summarisation.



Category <chr>	TotalSalesValue <dbl>	Transactions <int>
Laptop	1163889479	4761
Monitor	578385570	6837
Cloud Subscription	98715482	7132
Keyboard	73499067	8367
Software	66468485	9622
Mouse	51219577	9554

Table 3

Lastly, the differences in totals before and after corrections have been compared. Table 1 shows the total sales and transactions for all products. Laptop account for the highest total sales value of 1.16 billion despite the moderate transaction count, which can be explained as this premium product's pricing power and unit margins drive profitability. Monitor and Cloud Subscription are mid-tier performers since the former shows strong total sales with medium transaction volume, and the latter lower sales value with higher transaction volume. Keyboard, Software and Mice account for the highest number of transactions, but are low value products and can be classified as lower-margin consumables.

Essentially, the differences between the previous and new outcomes when comparing the data files reveal that the improved data provides a more accurate representation of the company's product portfolio and sales performance, improving the credibility of subsequent analyses. After relabelling each product type and correctly classifying them, the number of unique product types and categories in the data summary decreases. By repeating the 10 base prices and markup patterns for each group of 10 products, the calculated metrics like average revenue per product type and sales contribution per category becomes more realistic. Also, the difference between the old and new data's SPC results is that they went from being noisy to being more stable, which enables the true process variation to be visualized.

5 Optimizing profit from a given dataset

In this section, a dataset containing individual service times and barista counts for a coffee shop has been analysed to obtain the schedule that will optimize profit. The given file lists the number of baristas and their respective service seconds showing all the sales for a period of 1 year and the task is to focus on balancing speed, customer service quality, and operational cost management. It was decided that the assignment of the optimal number of baristas per week should be event driven. Each arriving customer was assigned to the barista who becomes free first to approximate the real multiple-server service behaviour.

The following steps were taken to perform the optimization analysis. First, the data was prepared in R by renaming the column headings. Then, arrival times were randomly generated for an 8-hour shift, and a queue model was used to simulate multiple baristas. Next the wait times, reliability and profit were calculated, and plots were generated, to report the optimal number of baristas and expected metrics. The final recommendations came to be the best number of baristas: 2, expected mean daily profit: R14438.36, average waiting time: 4.5 seconds and reliable service: 100.0% of customers.

Furthermore, R was used to produce realistic daily metrics. First, each row represents a transaction and was assigned a random date in the year. The average profit, average wait and average reliability across the simulated days was computed. The graph below represents the average daily profit compared to the number of baristas which reveals that there exists a negative relationship between them and can be used to affirm the decision that fewer baristas bring about higher profit.

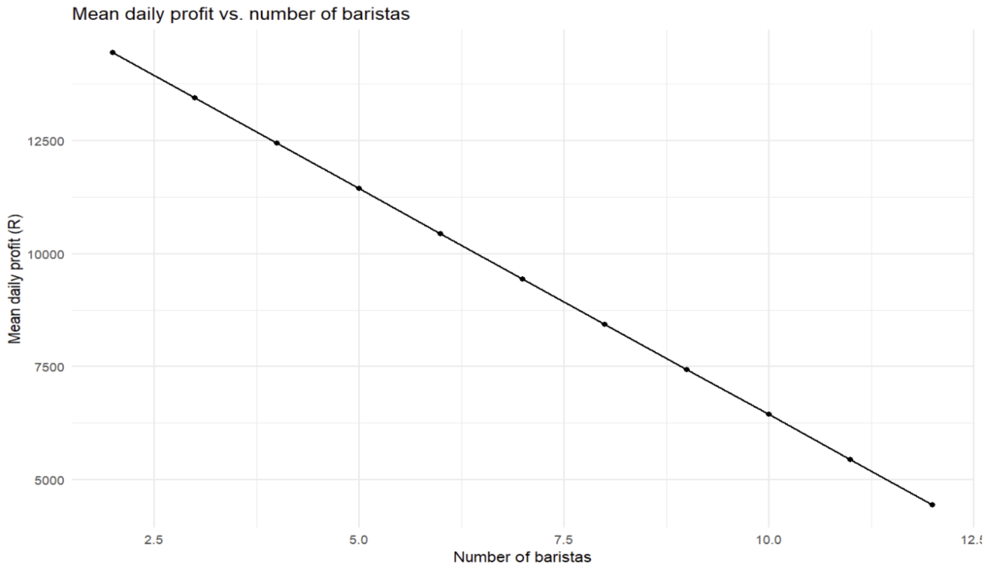


Figure 11

Conclusion

Finally, to summarize the results found from the analysis done within this report, several important insights can be reviewed. Namely, the SPC study demonstrates that the process is statistically under control for most observations, with only a few out-of-control signals. The data suggest that the process has improved stability over time, variation is primarily due to common causes, special causes can be traced to identifiable operational issues and the implemented quality controls are generally effective. Then the reanalysis using the corrected data shows measurable improvements in reported total sales values and internal consistency. This update enhances confidence in subsequent SPC interpretations, financial reporting, and quality management conclusions. From the simulation of individual service times across 365 operating days to find the optimal staffing for a coffee shop, it was concluded that hiring fewer baristas significantly increases waiting times and decreases customer satisfaction, while hiring more adds unnecessary personnel cost with limited improvement in reliability.

Appendix

- Table 1: “Total sales and Transactions per category”
- Figure 1: “Mean daily profit vs. number of baristas”

Deliverable 4

By J. G. Kurz – 27177416

Due 24 October 2025

Table of Contents:

Introduction

Within this report a certain hypothesis is tested on part 3 wherein which SPC analysis was carried out to examine delivery time stability and capability across product types and across the years 2022–2023. The purpose of this assessment is to simulate and demonstrate how ANOVA works in the design of experiments. Besides this, the next section of this report contains analysis of the reliability of service at a car rental agency.

6 Design of Experiments and MANOVA/ANOVA

To begin with, SPC which is the monitoring and improvement of processes with variation due to either common or special causes can be complimented by the (multivariate) analysis of variance (MANOVA or ANOVA). This is a procedure wherein which a means of process metrics is compared to statistically confirm whether process averages differ significantly. More specifically, (M)ANOVA can be used in SPC for various purposes such as to quantify the impact of operators/machines/shifts/suppliers on process metrics, verify special cause signals detected and support process improvement decisions.

Moreover, the experiment is carried out in a controlled fashion where factors and levels are defined manually. The hypothesis chosen is “Did the average delivery time change significantly between 2022 and 2023?” where H_0: Mean delivery times are equal across years and H_1: Mean delivery times differ between years which will require a one-way ANOVA on delivery times group by year. The method followed that if the results indicate $p < 0.05$, H_0 will be rejected since this shows that mean delivery times differ significantly between 2022 and 2023. However, if the results indicate that $p \geq 0.05$, we will fail to reject H_0 since this shows that there is no significant difference detected.

By following specific steps to perform the analysis of variance in this experiment a conclusion was made from the results that followed. Firstly, the data was prepared in R and tested to see whether the mean delivery time differs significantly between years. The results showed that p = 0.238. After this, since the dataset contained more than 5000 observations, a visual inspection of residual Q–Q plots and residual vs fitted plots was performed in lieu of the Shapiro–Wilk test (which is restricted to sample sizes <5000).

The graphs depict that the residuals were approximately normally distributed, satisfying the ANOVA assumption of normality. Homogeneity of variance was verified using Levene's test. Indeed, from Figure 1 the points on the Q-Q plot roughly follow the straight line which explains that normality is acceptable. Since p > 0.05, the ANOVA results were not significant indicating no statistically detectable change in average delivery times between 2022 and 2023.
This supports the control chart observations that the process remained stable across years.

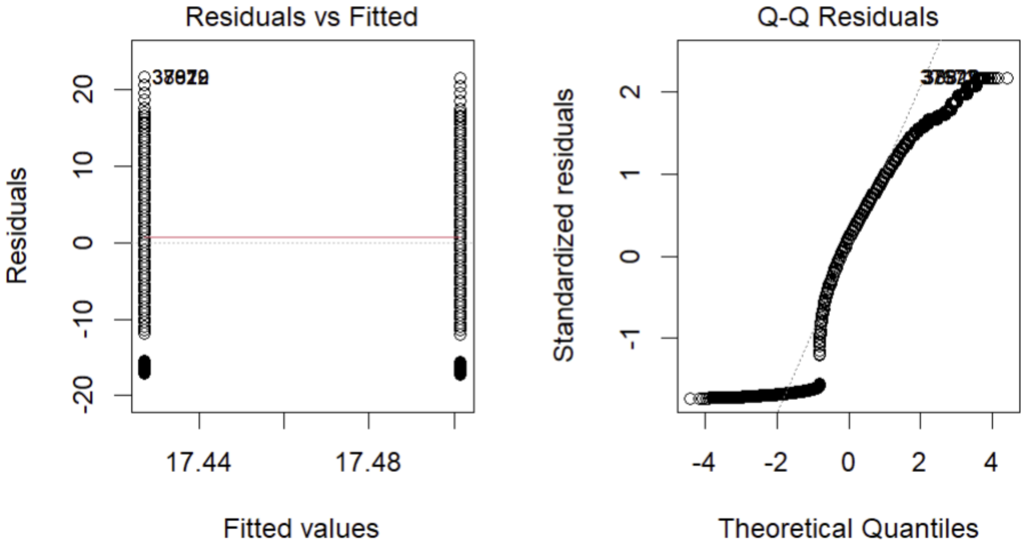


Figure 12

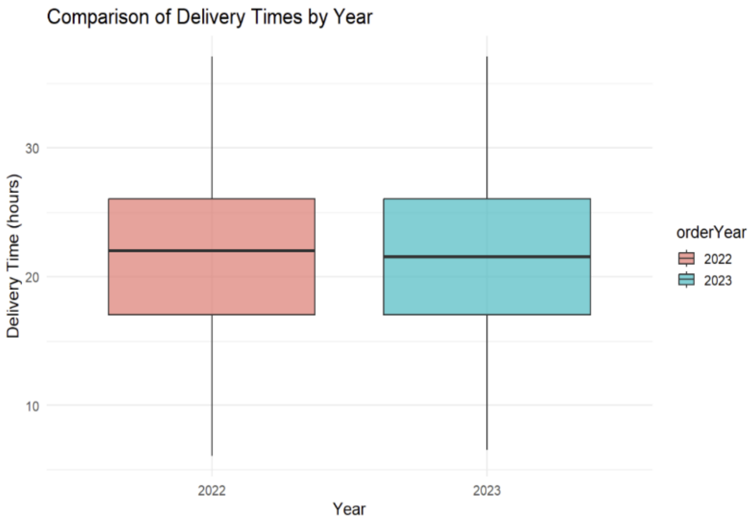


Figure 2

7 Reliability of service

To continue, this section of the report covers an estimation of service reliability and a model that solves profit optimization. The following histogram was given, displaying the number of employees on duty over 397 days.

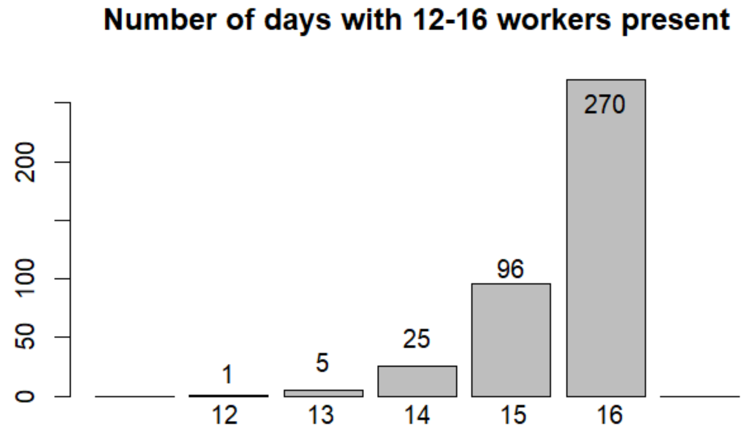
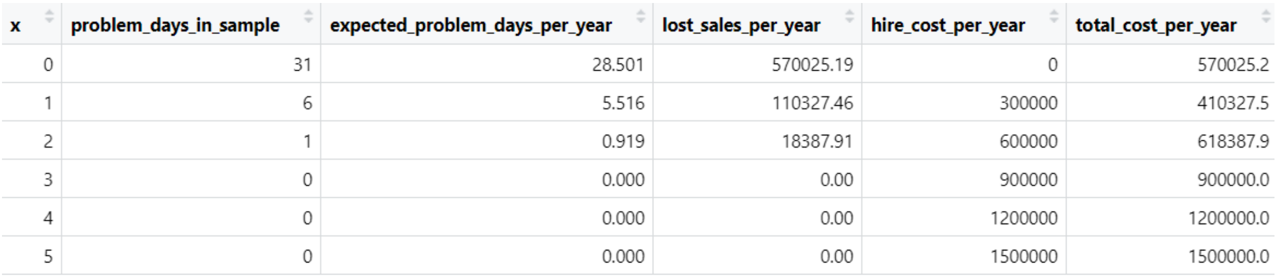


Figure 3

First, the estimation of service reliability was done by observing the given information in the bar chart and the fact that problems are experienced when fewer than 15 employees show up for work. We can deduce the reliability by observing the number of employees on duty by days count: 12 people were counted to be on duty in 1 day, 13 people were counted across 5 days, 14 people across 25 days, 15 people across 96 days and 16 people across 270 days. The summation of days when greater than or equal to 15 people were on duty is 366 days in total. Dividing this over the total amount of days observed gives $\frac{366}{397} = 0.9219$ or a 92% reliability that employees will show up for work. Therefore, to project this percentage over a 365-day year we would find the expected reliable days to be ≈ 0.9219 × 365 ≈ 336.5 days per year.

Secondly, by using the empirical counts in the histogram the staffing choice for the car rental agency was modelled and optimized. By using a binomial-style approach to solve this decision optimization problem, the following outcomes were discovered. The information given was that a loss of R20,000 worth of sales is incurred on every day that less than 15 employees show up for work, and that the cost of hiring one person per month is R25,000. Under the assumption that an optimized number (x) of staff will simply shift up every observed daily employee count by that same number, the net annual cost for different x can be computed as Total cost(x) = Annual hiring cost(x) + Expected annual lost sales(x).



x	problem_days_in_sample	expected_problem_days_per_year	lost_sales_per_year	hire_cost_per_year	total_cost_per_year
0	31	28.501	570025.19	0	570025.2
1	6	5.516	110327.46	300000	410327.5
2	1	0.919	18387.91	600000	618387.9
3	0	0.000	0.00	900000	900000.0
4	0	0.000	0.00	1200000	1200000.0
5	0	0.000	0.00	1500000	1500000.0

Table 4

From Table 1 which provides the observed counts it can be inferred through pure economic optimization that hiring 1 additional permanent person will minimize the company's total annual cost to R410,327 instead of not hiring any employees which amount to R570,025. The company may refrain from hiring 2 or more employees as this will be more expensive overall since the incremental hiring cost outweighs the further reduction in lost sales.

Conclusion

To conclude, the company's service reliability performance has been evaluated and improved successfully through the application of a range of statistical and operational analysis techniques. A hypothesis was tested to find whether mean delivery times differed significantly between years, and it was confirmed that there was no statistically detectable year-on-year variation. Then, the analysis was extended through reliability assessment for workforce optimization within a car rental division. Empirical analysis on 397 days of staffing data revealed reliable service on 92% of days, and modelling the data as a binomial problem revealed the optimal number of employees to hire to minimize total annual costs while reducing service failures is 1. Overall, a data-driven approach to process improvement was demonstrated through the integrated use of SPC, capability analysis, inferential testing, and reliability modelling.

Appendix

- Figure 1: "Residual and Q-Q plot"
- Figure 2: "Comparison of Delivery Times by Year"
- Figure 3: "Number of days with 12-16 workers present"
- Table 1

References

Below is a list of all the references and sources used across all 4 deliverables.

Besterfield, D.H., 2019. *Quality improvement*. 10th ed. Harlow: Pearson Education Limited.

Chakraborty, S., 2020. *Statistical Process Control: Concepts, Methods and Applications*. 2nd ed. New York: CRC Press.

ISO 9001:2015. *Quality management systems – Requirements*. Geneva: International Organization for Standardization.

Hair, J.F., Wolfinbarger, M., Money, A.H., Samouel, P. and Page, M.J. (2019) *Essentials of Business Research Methods*. 4th ed. New York: Routledge.

Montgomery, D.C. (2020) *Introduction to Statistical Quality Control*. 9th ed. Hoboken, NJ: John Wiley & Sons.

Oakland, J.S., 2014. *Statistical Process Control*. 6th ed. Abingdon: Routledge.

OpenAI, 2025. *ChatGPT (GPT-5)* [online]. Available at: https://chat.openai.com/

Slack, N., Brandon-Jones, A. and Burgess, N., 2022. *Operations Management*. 10th ed. Harlow: Pearson Education Limited.