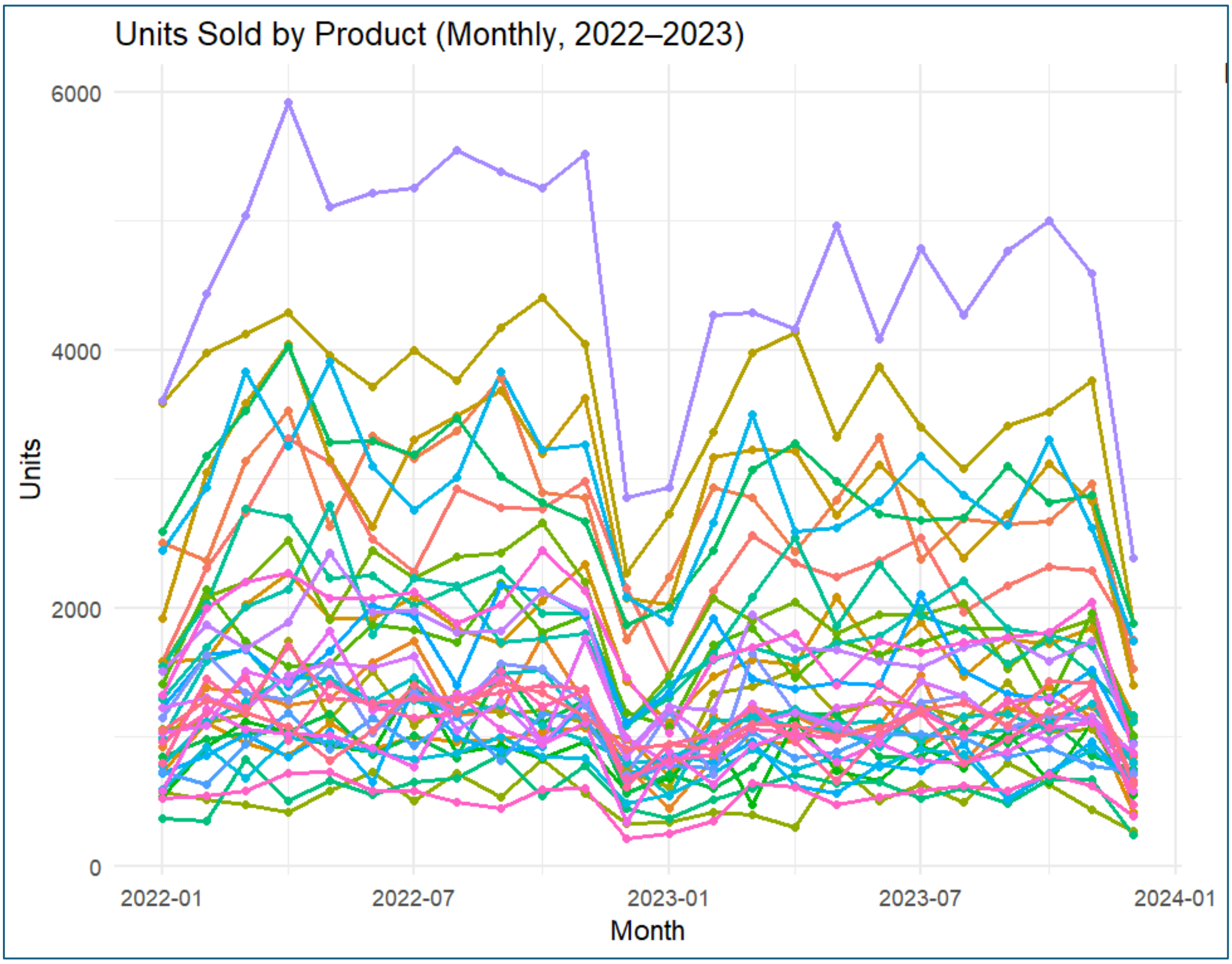
Quality Assurance 344

Final Report



Name: M. le Roux

Student number: 26999757

October 2025

Contents

Part 1

1.1 Introduction

This part analyses a tech retailer/wholesaler that sells laptops, monitors, mice, keyboards, software, and cloud subscriptions. We will analyse the data and highlight the most important findings. Graphs are included to help illustrate these findings clearly.

1.2 Initial data observation

The part looks at the company's 2022–2023 sales using the following provided files: customer_data, products_data, products_Headoffice and sales2022and2023.

During a quick review we noticed:

- Mistakes in the datasheet products_Headoffice where there were mistakes in the productid column. There were number of IDs that shows NA at the beginning.
- Mistakes in the datasheet of products_data as the category column does not correspond to the productid column.

We will ignore these mistakes for now after we concluded this analysis and heard back from the Head office. We will thus only focus on the analysis of the files of customer_data and sales2022and2023 until we hear back from the headoffice.

1.3 Observations

1.3.1 Categories vs Sales

The different categories that this company sells is listed in the visual below which includes Monitor, software, Keyboard and so on. When we look at the sales between categories across 2022–2023 (Visual A below) we can see that there isn't a clear runaway winner. Monitors sell the most overall, with Software, Keyboard, and Cloud Subscription close behind, and Mouse and Laptop only slightly lower.

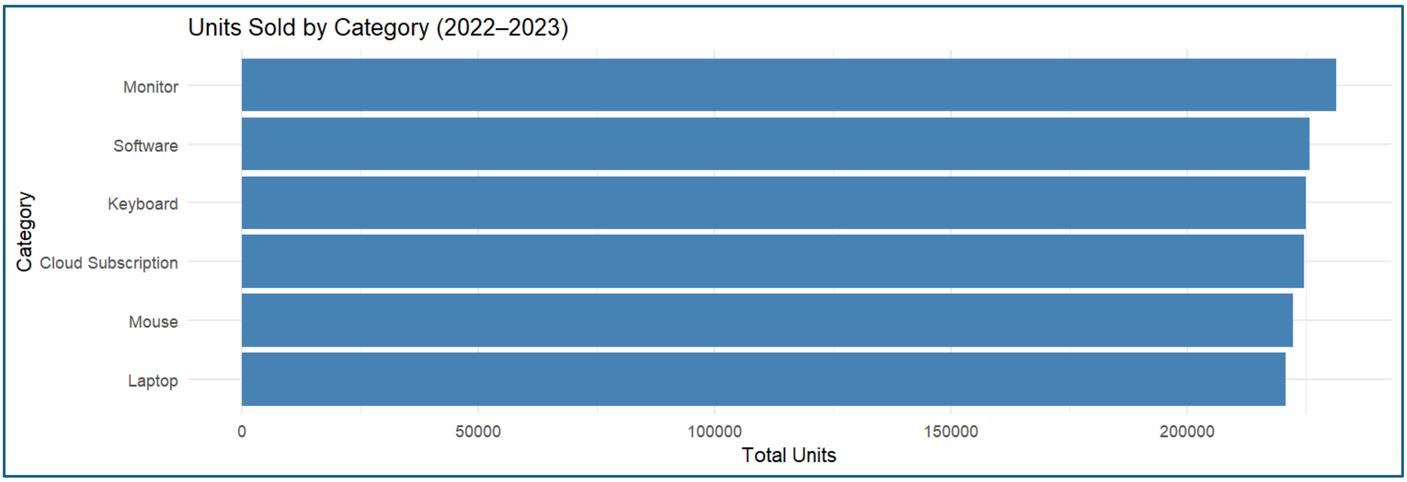


Figure A: Units sold by category for 2022-2023

Looking at 2022 and 2023 separately, the picture hardly changes: the ranking is the same and the gaps between categories stay small. The main difference however is that the sales in 2023 dropped between each category (seen in Visual B below).

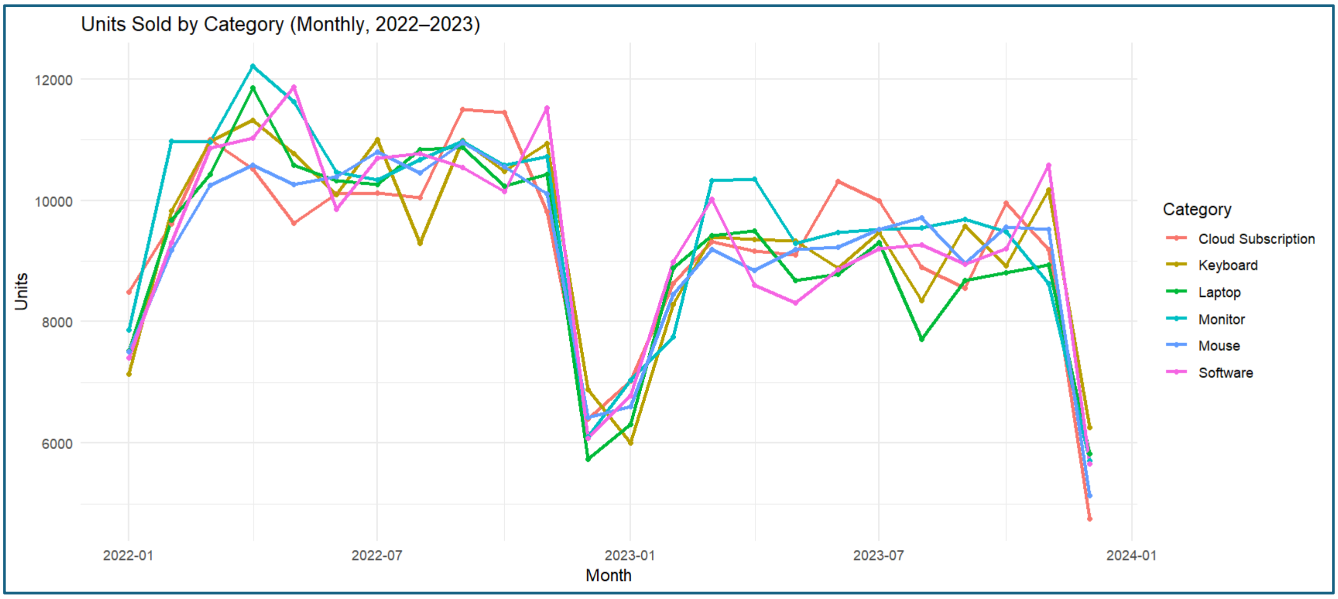


Figure B: Units sold by category monthly for 2022-2023

We can see from the graph that there was a fast climb in early 2022, broad peaks between the year of 2022, then a sharp year-end dip (Dec). 2023 repeats the same seasonality but at slightly lower levels overall (most months are ~5–10% below 2022), with smaller peaks mid-year and another drop at the end of the year. Across the two years, Monitors and Cloud Subscriptions tend to sit at the top of the pack, Mouse and Laptops a touch lower, but the gaps between categories are modest compared with the shared pattern.

1.3.2 Description vs Sales

There are different description types that customers can choose from in each category, such as *chocolate silk*, *azure silk*, *azure sandpaper*, and so on. When we look at how the sales performed across these description groups (see Visual C below), we can see a clear leading group. *Chocolate silk* is the runaway leader, followed by *azure silk* and a cluster including *azure sandpaper*, *burlywood silk*, *blue silk*, and *aliceblue silk*. It's clear that almost all the best-sellers have "silk" in their names, which suggests that customers are particularly drawn to products with "silk" in the description.

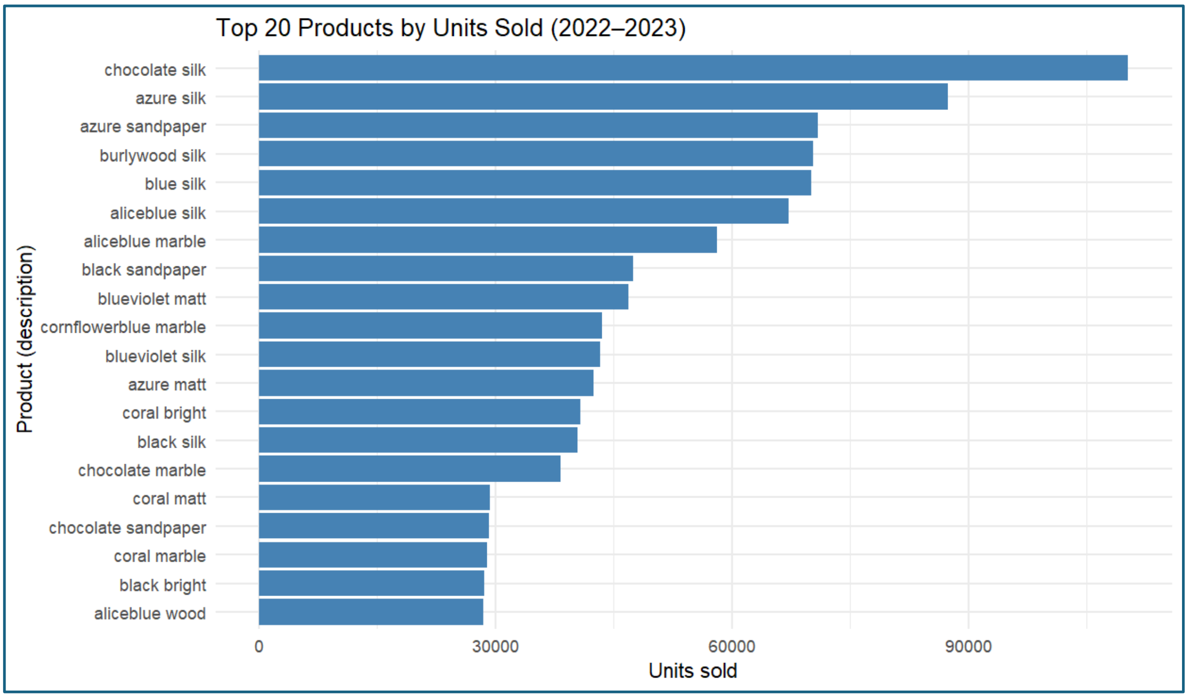


Figure C: Top 20 Products according to description sold from 2022-2023

When we look at 2022 and 2023 separately, we don't see a significant difference except that we know that the sales dropped in 2023 (Visual D below). Chocolate silk still was a favourite in 2023 and 2022

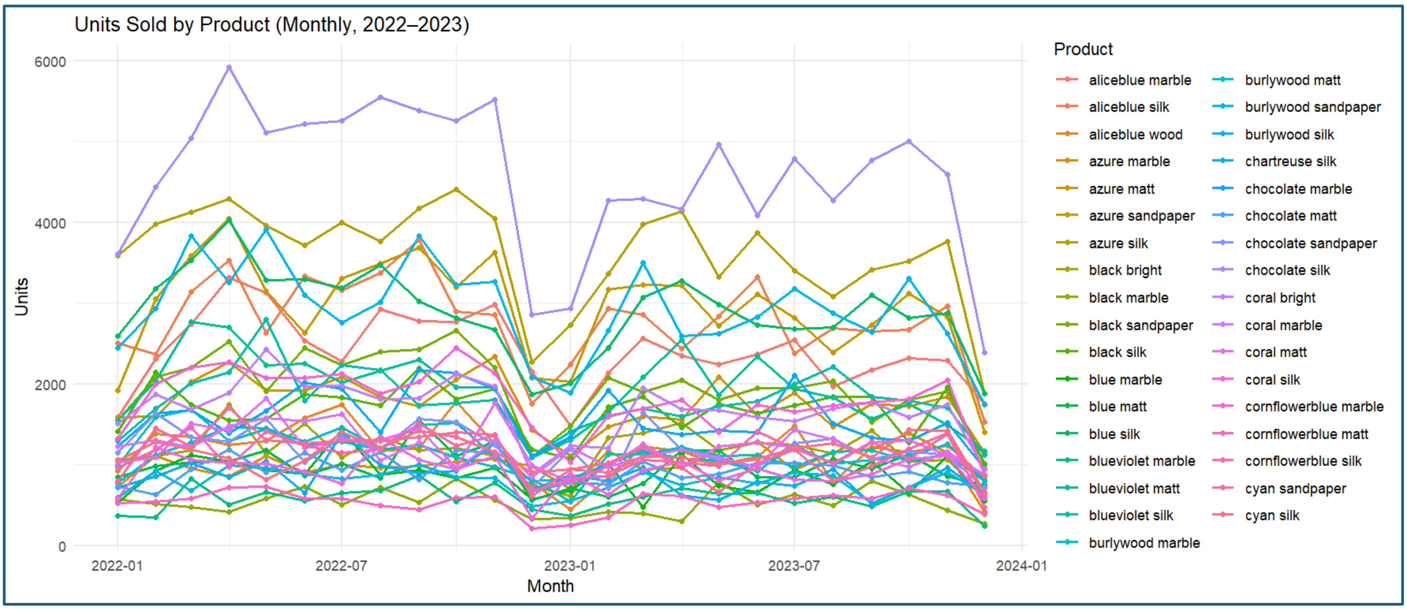


Figure D: Units Sold Monthly by Product description from 2022-2023

1.3.3 Description vs Category

We can look at which description did the best within each category (Visual E below). Cloud Subscription has the largest single type of description sold across all categories, which is "burlywood silk". These items are the main volume drivers in their categories and are good candidates for priority stocking and promotion.

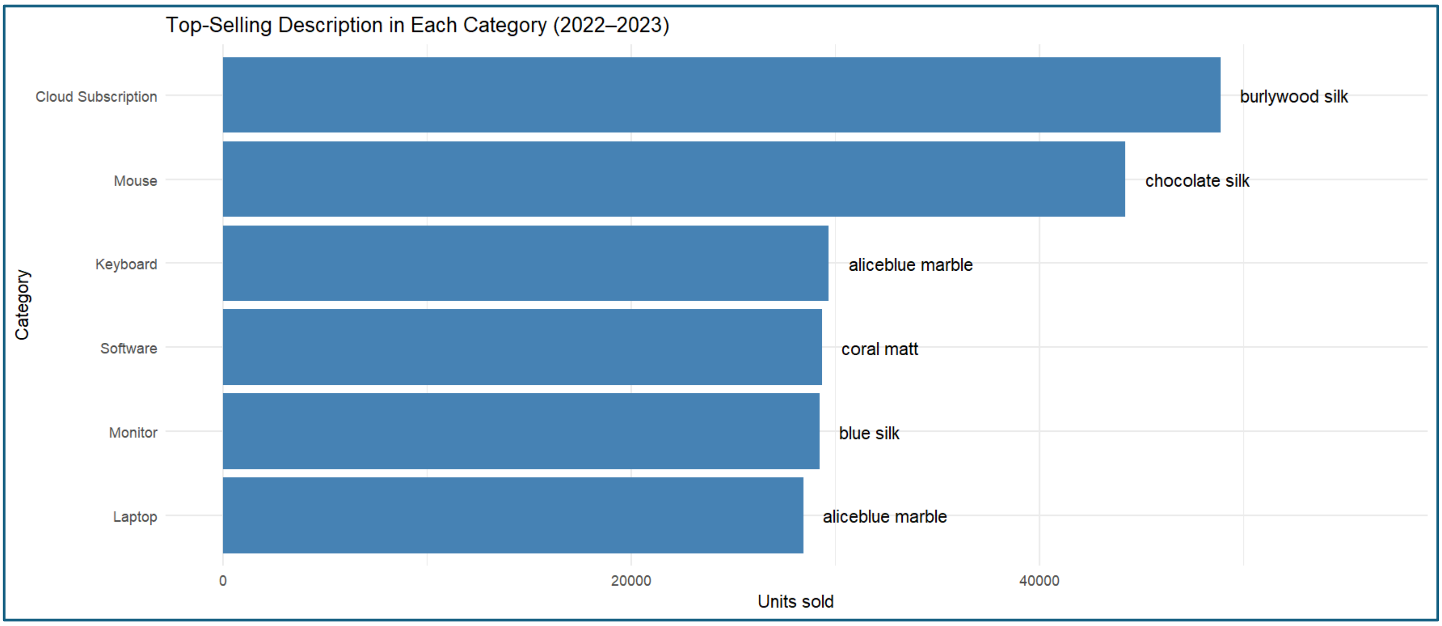


Figure E: Top selling description in each category from 2022-2023

1.3.4 Weekday vs Sales

When we look at the sales during the weekdays, we can see a clear pattern (Visual F below). In both years, Friday is at the top. A noticeable difference between 2022 and 2023 appears when we look at Tuesday—sales on that day dropped significantly from 2022 to 2023, which suggests that head office should investigate the reason behind this change. Weekend volumes (Saturday–Sunday) fall in the middle range—lower than Friday but not the lowest overall. Overall, 2022 sales are higher than 2023 across all days, although the general weekday pattern remains more or less the same. The biggest differences between the two years are seen on Monday and Tuesday.

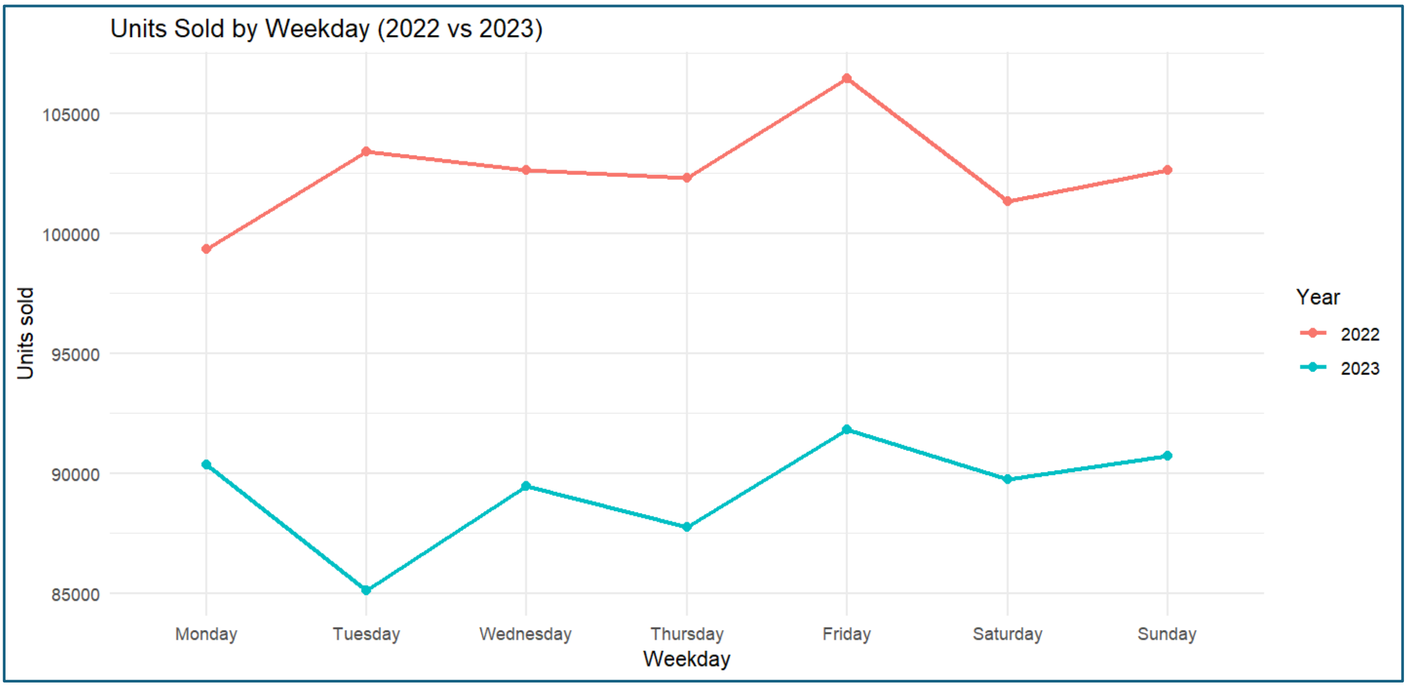


Figure F: Units sold by weekday from 2022-2023

1.3.5 Day of the month vs Sales

When we look at the unit sold during each day of the month, we can see that there looks like a bit of a dip from day 3 to day 9. But as we can see there are some changes from 2022 to 2023. On day 10 the sales starts to spike in 2023 while the sales drop in 2022. It is established that 2023 did worse than 2022. The highest spike for 2022 was at day 16 while the highest spike for 2023 was at day 10. The lowest spike for 2022 was at day 24, while the lowest spike for 2023 was at day 26.

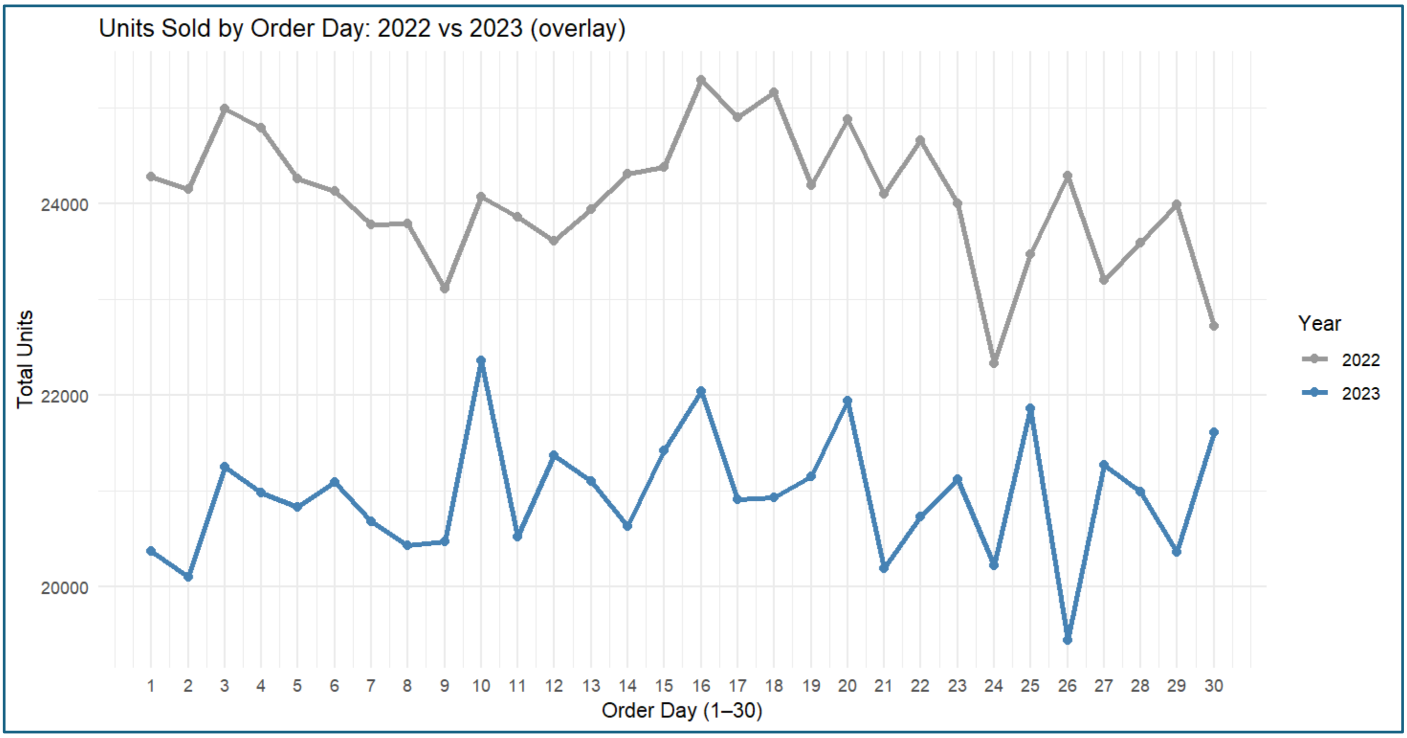


Figure G: Units sold by order day from 2022-2023

1.3.6 Picking hours

When we look at the picking hours, we can see that the distribution is cantered in the 10–20 hour range (most orders are picked within roughly half a day), with a long right tail out to 30–45 hours that marks slower, exceptional cases. There's also a visible block at 0–1 hours, which likely reflects missing or unlogged times rather than true same-hour picking.

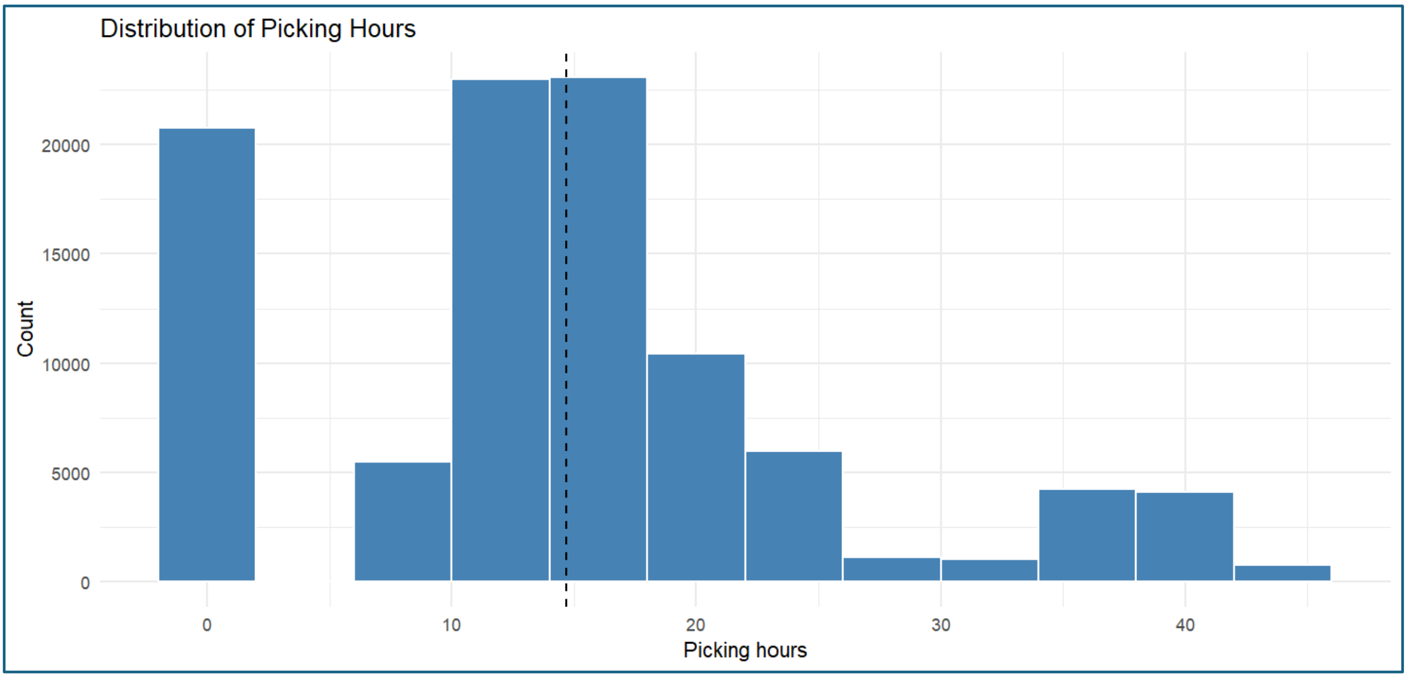


Figure H: : Distribution of picking hours

1.3.7 Delivery hours

The delivery times are mostly in the 18–28 hour range, with a right tail that stretches to about 35–40 hours for slower, exceptional cases. There's again a small spike at 0–1 hours, which likely reflects missing or unlogged timestamps rather than true same-hour delivery.

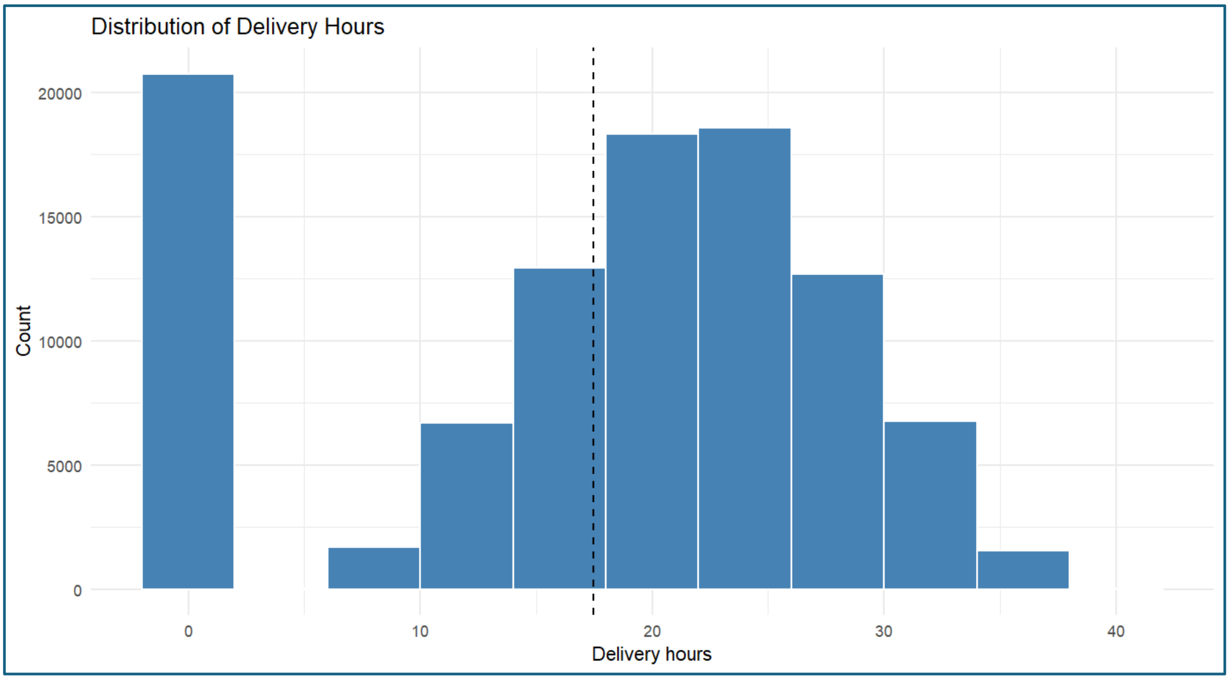


Figure I: Distribution of delivery hours

1.3.8 Customer observations

Table 1 shows how many units each customer bought in 2022 and 2023. Table 2 shows how many orders were placed by each customer in 2022 and 2023. We can see from the tables that CUST1193, CUST1791and CUST596 are the top 3 customers as they ordered the most units and ordered the most frequently.

Table 1: How many orders a customer placed

	customerid	units
1	CUST1193	14704
2	CUST1791	14626
3	CUST596	14212
4	CUST3721	13852
5	CUST2527	13773
6	CUST2277	13538
7	CUST1427	13335
8	CUST4729	12938
9	CUST3944	12855
10	CUST1501	11958
11	CUST1203	542
12	CUST4498	531
13	CUST1048	527
14	CUST2941	520
15	CUST3487	502

Table 2: How many units a customer bought

	customerid	orders
67	CUST1193	326
1	CUST1791	322
119	CUST596	319
514	CUST2527	303
4	CUST3721	301
399	CUST2277	298
32	CUST1427	296
368	CUST3944	286
325	CUST4729	279
12	CUST1501	269
3586	CUST3471	36
1253	CUST584	34
1432	CUST1719	34
1434	CUST1249	34
1881	CUST1446	34
2073	CUST3974	34
2228	CUST3576	34

In 2022 there are 53,727 orders and in 2023 there are 46,273 orders. The table 3 below shows the categories and how much revenue and gross profit the business made. Laptops made the most gross profit, then monitors and so on.

Table 3: revenue and gross margin for each category

	category	revenue	margin
1	Laptop	821533851	18423872712
2	Monitor	809104952	16939751513
3	Mouse	721090260	16463653608
4	Keyboard	723693159	13972439149
5	Cloud Subscription	621799523	11722600585
6	Software	655365933	10261426453

The ratio between male and female customers is basically identical with a ratio of 0.4864 being female, 0.4700 male and 0.0436 other/unspecified. The most customers are between the ages of 30-35 years. The most customers live in San Fransisco. The most customers have an income of 105000. No customer under 50000 income purchased high quantities, so we can conclude that the more money the customers have the more they are willing to buy. Visual J below shows the top 20 customers with their income.

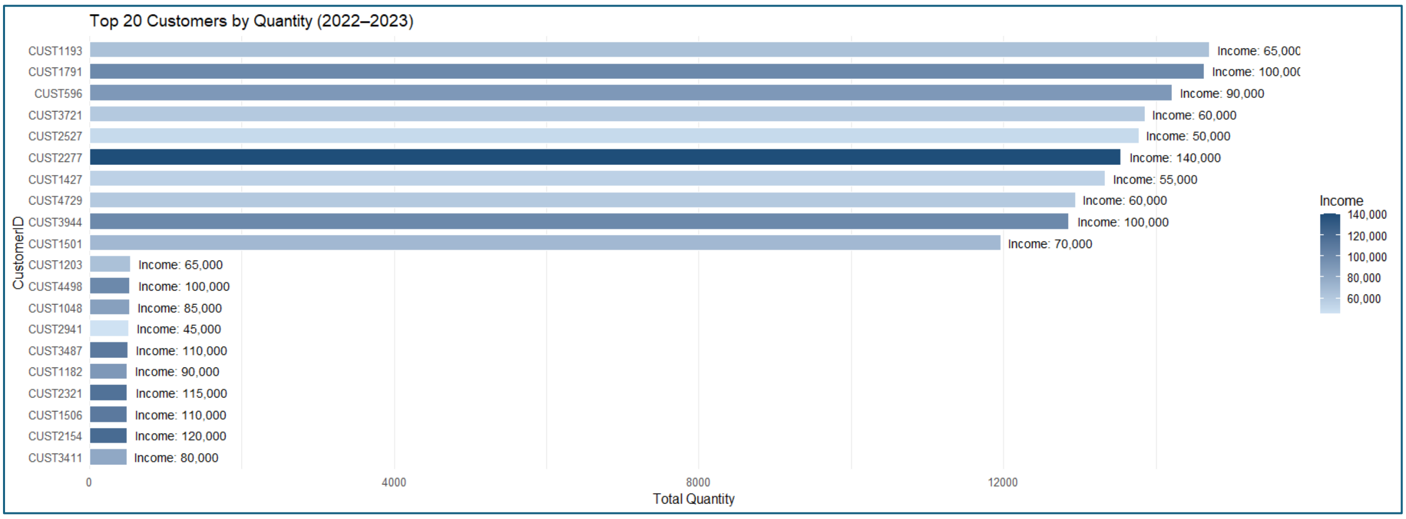


Figure J: How many units a customer bought

We can not really conclude that age has a factor in buying larger quantities, but we can say that no customer over 77 and under 18 bought large quantities. Visual K below shows the top 20 customers with their ages.

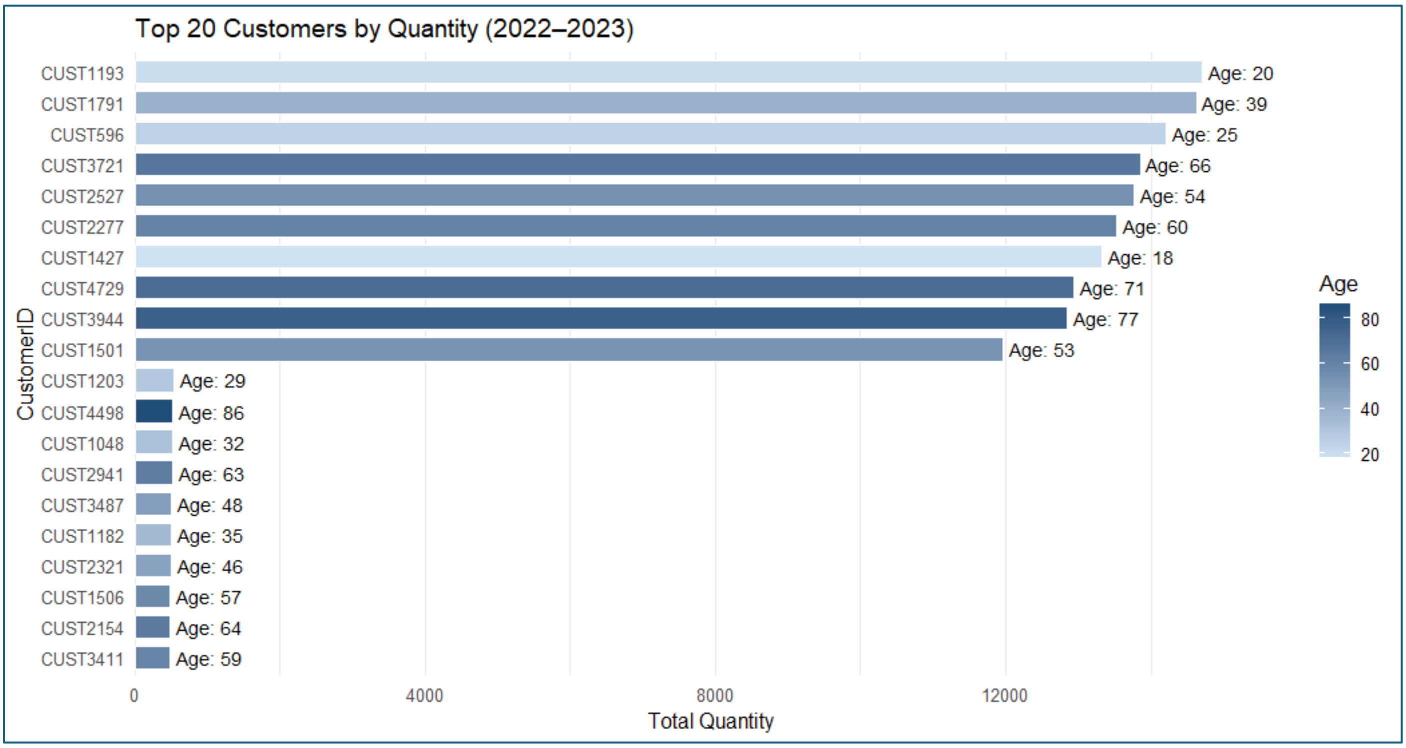


Figure K: Top 20 Customers by Quantity vs age

The top 10 customers are mostly female. 6/10 of the top 10 customers are female and 12/20 of the top 20 customers are female. Visual K below shows the top 20 customers, how much they bought in total and whether they are male or female.

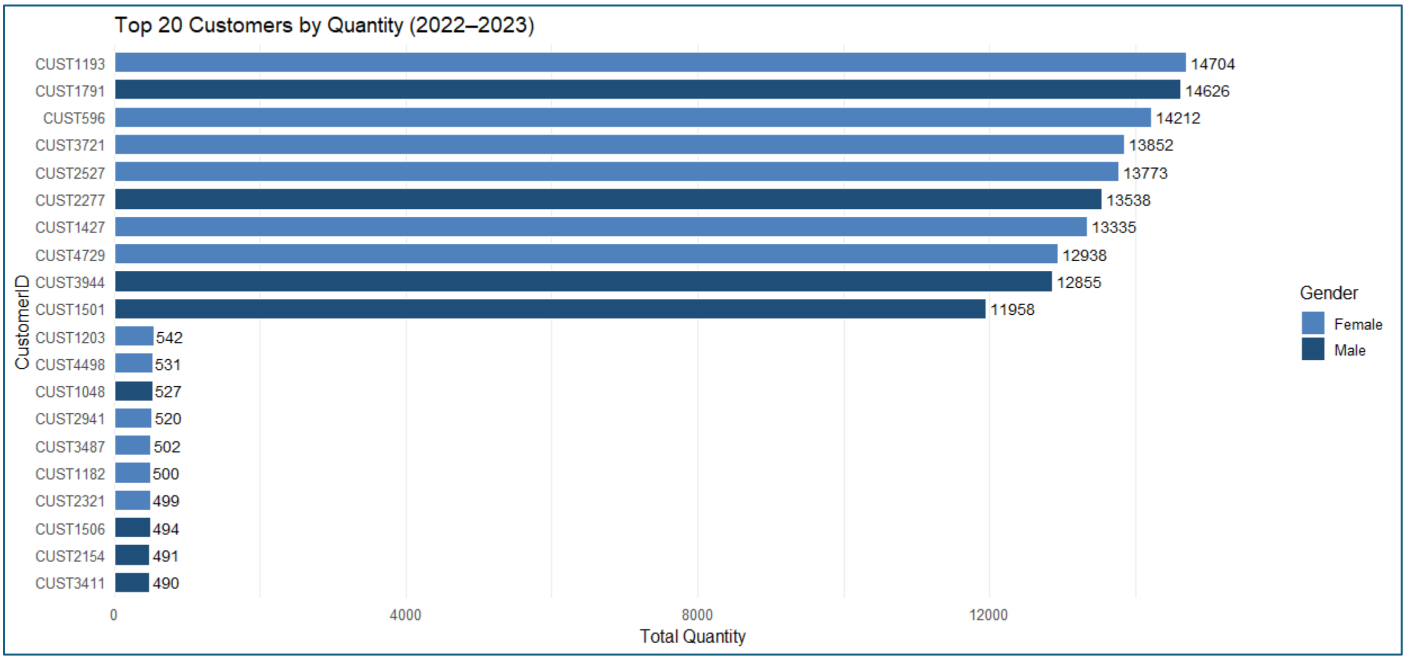


Figure L: Top 20 Customers by Quantity vs Gender

1.4 Insights & Recommendation

From the analysis above, we can see that the company's sales performance between 2022 and 2023 followed very similar seasonal and category trends, although total sales volumes were slightly lower in 2023. Fridays continue to be the strongest sales day, while Mondays and Tuesdays show a noticeable drop that should be investigated further. Monitors and Cloud Subscriptions remain consistent best-sellers, suggesting steady customer demand for both physical and digital products.

We also found that descriptions containing "silk" performed far better than others, which may indicate that customers are drawn to the way these products are named or marketed. This could be used in future branding strategies by applying similar naming conventions across other product lines.

The customer data shows that the gender distribution is almost equal, meaning the business attracts a balanced audience. However, higher-income customers tend to purchase larger quantities, while lower-income customers generally buy smaller volumes. This suggests that targeted promotions for mid- to high-income customers could increase total revenue. Age appears to have less of an effect on sales behaviour, but no customers below 18 or above 77 made large purchases.

Several data quality issues were also identified, including missing timestamps, incomplete customer IDs, and unmatched product IDs between files. Before making final business decisions, the company should clean and validate these datasets to avoid incorrect insights. Future data collection should ensure that product and customer identifiers are consistent across all systems.

Overall, the company's performance remains stable, but improvements can be made by:

- Investigating the sales decline on Tuesdays,
- Refining product naming and marketing to mirror "silk"-style successes,
- Focusing promotions on higher-income customers,
- and improving data accuracy for future analysis.

Part 2 and 3

2.1 Introduction

This part continues the analysis of the tech retailer and wholesaler that sells laptops, monitors, mice, keyboards, software, and cloud subscriptions. However, we now focus specifically on the future sales data. We will analyse this dataset and highlight the most important findings, with graphs included to clearly illustrate the results. In this section, we apply Statistical Process Control (SPC) to monitor and evaluate process stability, focusing particularly on the delivery times to assess how consistent and reliable the process is over time.

2.2 Initial data setup

This part looks at the company's 2026–2027 forecasted sales, using the provided file *sales2026and2027*. We created s-charts and X-bar charts for all product categories to monitor process stability. However, since the results across all categories showed similar patterns, only the CLO charts are included in this report for illustration. The focus of this section is therefore on drawing conclusions and insights rather than showing every individual chart.

We applied X-bar and s control charts to the *sales2026and2027Future.csv* delivery times, ordered as they would arrive in real time (by Year → Month → Day → orderTime). Following the brief, we drew samples of 24 deliveries per product type (CLO, SOF, LAP, KEY, MON, MOU). The first 30 samples (30 × 24 observations) per product type were used to initialise the charts: we estimated the centre lines and the ±1σ, ±2σ, and ±3σ control limits from these baseline samples. After initialisation, we continued with sample 31, 32, ... in an accelerated simulation, monitoring each process as if data were arriving in sets of 24. As in real life, we check the s-chart first (spread must be in control before interpreting the mean), and then evaluate the X-bar chart for shifts or trends. When a chart shows a rule violation, the relevant product manager should be alerted to check/adjust the process.

2.3 Observations

2.3.1 CLO charts

Figure M below shows the s and x-bar chart for CLO. In the s-chart, most points fall within the control limits, showing a consistent level of variation with only a few points approaching the upper control zones. This suggests that the process variation is stable and predictable, meaning the process spread is in control. In the X-bar chart, the overall mean remains steady around the centre line, but there are visible sections where the average delivery time shifts slightly upward or downward over time. These small but noticeable shifts may indicate assignable causes, such as changes in courier performance, scheduling, or workload distribution. However, because the points remain within the ±3σ limits, the process can be considered under control, though it still requires ongoing monitoring to prevent drift.

In conclusion, the CLO delivery process appears stable in terms of variation but shows periodic mean shifts that should be reviewed by management. Since the process spread is in control, we can confidently use this baseline data to calculate the process capability indices (Cp, Cpu, Cpl, and Cpk) for the first 1,000 deliveries per product family in the following section.

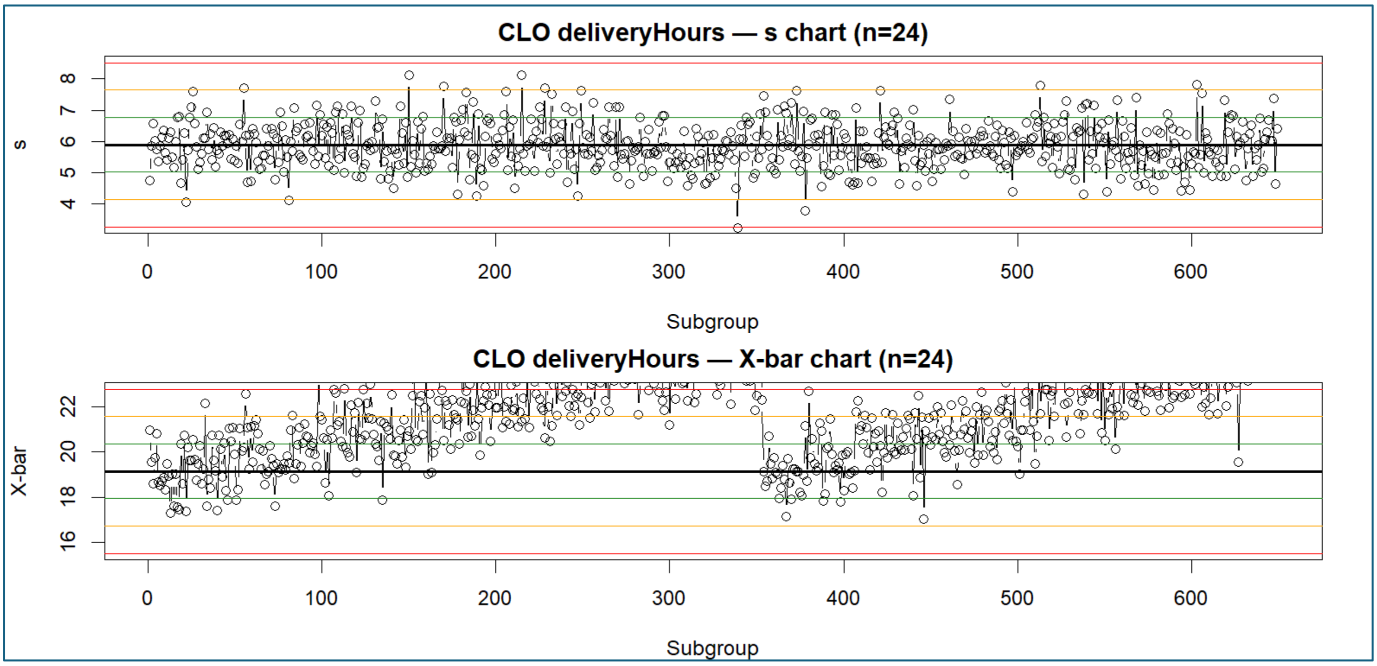


Figure M: CLO dileveryHours s and X-bar charts

Table below shows all the Cp, Cpu, Cpl and Cpk answers for each category. Because customers mainly care about how quickly they receive their orders, the most important measure in this table is Cpu, which represents how close the average delivery time is to the upper limit of 32 hours. In other words, Cpu shows how safely the company stays below the delivery-time promise made to customers. A higher Cpu value means the company is performing better and delivering well within the 32-hour limit, while a lower Cpu means deliveries are getting closer to being late.

When we look at the table, the Software (SOF) category performs extremely well, with deliveries far below the 32-hour limit, showing almost no risk of delay. The other product families (KEY, MOU, CLO, MON, and LAP) have average delivery times around 19–20 hours and much higher variation, meaning that while most deliveries are on time, a few might exceed 32 hours. This suggests that these product lines are not as consistent as Software and could benefit from improvements in scheduling, courier coordination, or dispatch efficiency.

Table 4: ProductFamily and Cp, Cpu, Cpl and Cpk calculations

ProductFamily	n	mean_dh	sd_dh	Cp	Cpu	Cpl	Cpk
SOF	1000	0.955	0.294	18.135	35.188	1.083	1.083
KEY	1000	19.276	5.815	0.917	0.729	1.105	0.729
MOU	1000	19.298	5.828	0.915	0.727	1.104	0.727
CLO	1000	19.226	5.941	0.898	0.717	1.079	0.717
MON	1000	19.410	5.999	0.889	0.700	1.079	0.700
LAP	1000	19.606	5.934	0.899	0.696	1.101	0.696

2.4 Insights & Recommendation

From the process capability results, it is clear that the Software (SOF) category performs exceptionally well. Its average delivery time is under one hour with very little variation, resulting in a very high Cpu value. This means that SOF deliveries are both fast and consistent, and the process easily meets the customer requirement of being completed within 32 hours.

In contrast, the other product families—Keyboard (KEY), Mouse (MOU), Clothing (CLO), Monitor (MON), and Laptop (LAP)—show Cpu values of around 0.7, which indicates that while their average delivery times are still well below the 32-hour limit, the variation in their process is much higher. This larger spread means that some orders could occasionally take longer than expected, creating a small but noticeable risk of late deliveries. These products are therefore not yet capable of fully meeting the Voice of the Customer (VOC) under all conditions.

To improve consistency and capability, the company should focus on reducing delivery variation by standardising processes such as courier scheduling, pick-and-pack procedures, and dispatch timing. Improving coordination between departments and ensuring that logistics partners meet strict service-level agreements will also help bring down variability. Once these improvements are in place, the company should re-evaluate the process capability to confirm that the Cpk values move closer to 1.33 or higher, which would indicate a capable and stable process.

In summary, while overall delivery performance remains satisfactory, the goal should be to make every product category as reliable and consistent as the Software line. By reducing variability and maintaining regular process monitoring through X-bar and s-charts, the company can ensure that customer expectations for fast and dependable delivery are consistently met.

Part 4

4.1 Introduction

This part continues the analysis of the tech retailer and wholesaler. However, the focus now shifts towards evaluating process risks, correcting data inconsistencies, and optimising performance for profit. Graphs and summaries are included to clearly illustrate the findings and improvements.

4.2 Initial data setup

Before continuing with the analysis, it was necessary to correct errors found in the original product data files. The head office dataset contained several incorrect ProductIDs, where some entries were labelled with "NA" instead of the correct category prefixes (for example, "SOF" for Software or "CLO" for Cloud Subscription). These prefixes were corrected using the Category column as a reference, ensuring that each ProductID accurately matched its product family.

Similarly, the products_data file contained mismatched category names that did not correspond with their ProductID prefixes. This file was updated by referencing the first three letters of the ProductID to assign the correct category name (for example, all "SOF" codes were updated to "Software," and all "LAP" codes to "Laptop").

After all corrections were made, two new, cleaned datasets were created : products_Headoffice2025 and products_data2025. These updated files now provide accurate, consistent information across all product families and will be used for all further analysis and comparisons in this section.

4.3 Likelihoods od Errors

4.3.1 Type I (Manufacture's) error

This section looks at the likelihood of making a Type I error, also known as a *false alarm*. In our delivery process, this would mean that the control chart signals a problem. For example, several delivery time samples appear above the upper control limit — even though the delivery process is actually stable and functioning normally.

In this case, H_0 (null hypothesis) is that the process is in control and centred on the established average delivery time, while H_1 (alternative hypothesis) is that the process has shifted or become unstable. The probability of incorrectly rejecting H_0 (when the process is actually fine) is called α, or the Type I error rate.

Since the probability of any single point falling above the centreline by random variation is 0.5, the chance of seeing seven consecutive points above the centreline is $0.5^7 = 0.0078$, or about 8 in 1000. That means there is a 0.78% chance that our SPC chart could falsely indicate that deliveries are taking too long when, in reality, they are still consistent. In our context, a Type I error would cause unnecessary adjustments — for example, management might change delivery staff schedules or vehicle routes even though no real issue exists.

In our case, a Type I error would mean that the company mistakenly believes its delivery process is out of control, even though it is actually stable. Based on the control charts from Part 3, most of the delivery data stayed well within the limits, with only small variations. This suggests that there are no signs of frequent false alarms, and the process remains generally stable.

Therefore, the likelihood of a Type I error in this process is very low, meaning the system is unlikely to trigger unnecessary interventions when everything is operating normally.

4.3.2 Type II (Customer's) error

A Type II error, also called consumer's risk, happens when the process has actually changed but the control chart fails to detect it. In our delivery-time context, this means the average delivery duration or variation has shifted — perhaps deliveries are taking longer than usual — but all points still fall within the control limits, so the company incorrectly believes everything is fine.

Here, H_0 is again that the process is in control, and H_1 is that the process has shifted (for example, due to traffic, staff delays, or system issues). The probability of failing to detect this shift is called β, the Type II error rate.

Using the example from the question, if the target process mean is 25.05 hours with UCL = 25.089 and LCL = 25.011, and the true mean quietly drifts to 25.028 hours, the process has moved but still lies within control limits. This means β is high — the SPC chart is not sensitive enough to catch small drifts.

In our case, a Type II error would occur if customers started receiving their products later than expected, but the company's control chart still appeared to show that everything was under control. From the SPC analysis, we can see that the delivery times were mostly consistent and close to the target, without significant shifts in the mean.

This means that the chance of a Type II error is also relatively low — the company is unlikely to miss major delivery issues. However, small undetected drifts could still occur over time, especially if the control limits are too wide or the sampling frequency is low. To further reduce this risk, the company could use tighter control limits or monitor delivery performance more often.

4.4 Observations

4.4.1 Products observations from products_Headoffice2025 file

According to the products_Headoffice data sheet the product that has the highest markup is KEY026, which is keyboard and description of blueviolet silk. It has a with a markup of 30 and a selling price of 16569.13. Figure N below shows the average markup per product description. From the figure we can see that the product description with the highest average markup is chocolate wood.

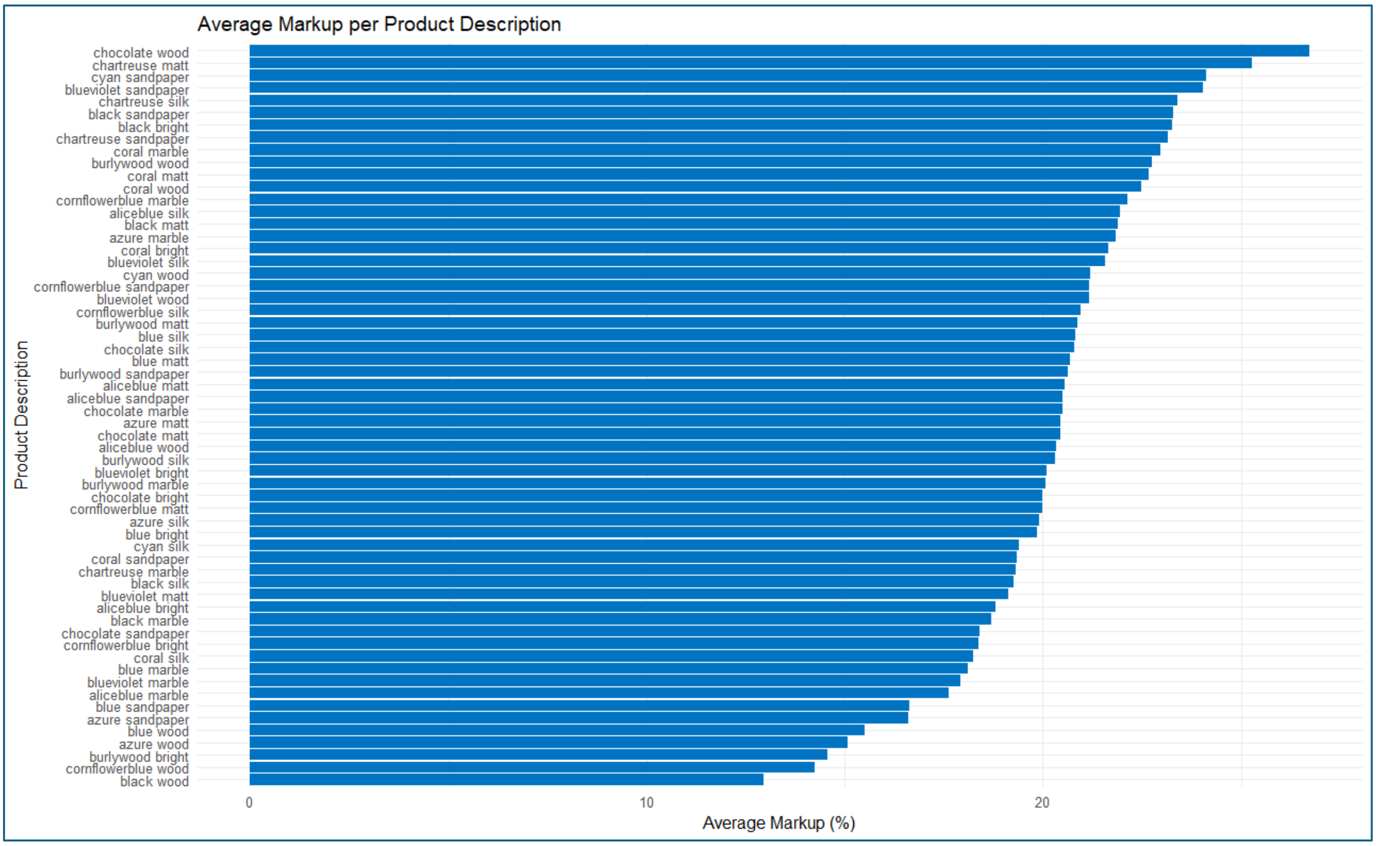


Figure N: Average markup product description in products_Headoffice file

This graph shows the average markup percentage for each product category. We can see that *Cloud Subscription* and *Software* have the highest average markups, indicating higher profit margins on these products. The other categories such as *Laptop*, *Mouse*, *Keyboard*, and *Monitor* have slightly lower but still consistent markups. Overall, the company maintains a fairly uniform pricing strategy across all categories, with only small differences in markup percentages.

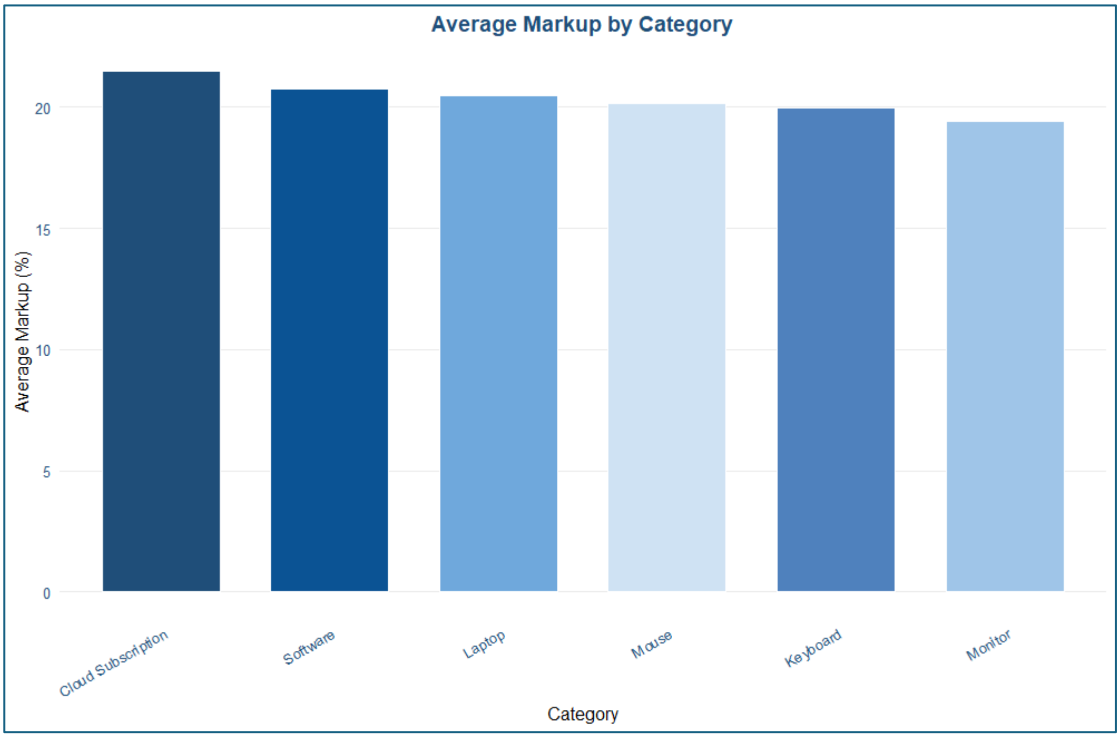


Figure O: Average markup category from products_Headoffice2025 data

This graph shows the top 20 product descriptions ranked by their average selling price. The highest-priced items include *Cyan wood*, *Aliceblue marble*, and *Cornflowerblue wood*, which stand out well above the rest. Most other descriptions fall within a moderate price range between R5,000 and R10,000. The variation in prices suggests that material or finish types (such as "wood" and "marble") tend to be associated with higher-value products compared to "sandpaper" or "silk" descriptions.

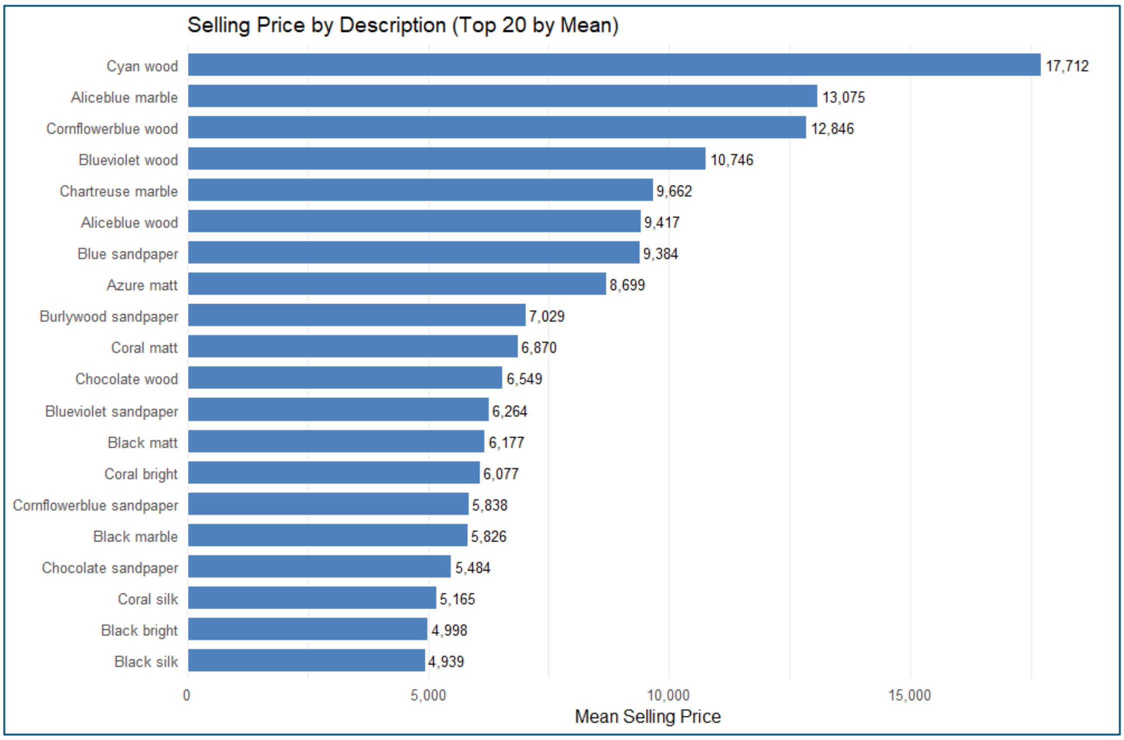


Figure P: Selling price by Category (Top 20 Mean) from products_Headoffice2025 data

This graph shows the top 20 product descriptions ranked by how many items share each description. *Black silk*, *Chocolate silk*, and *Coral silk* appear most frequently, suggesting that products described with "silk" are both popular and widely used across multiple categories. Other common descriptions such as *Black marble* and *Cornflowerblue silk* also occur often, showing that smooth, elegant finishes are dominant in the company's product range. Overall, this indicates that "silk" and "marble" themed descriptions are the most common design styles in the company's inventory.

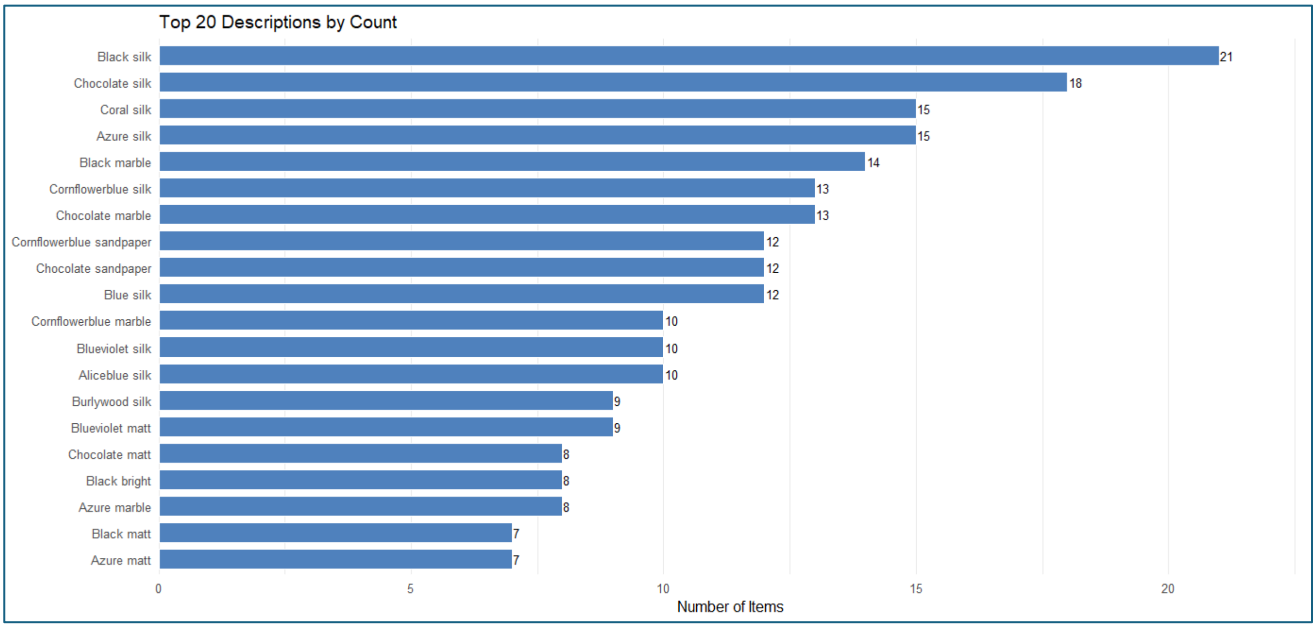


Figure Q: Top 20 Description by Count from products_Headoffice2025 data

4.3.1 Products observations from products_data2025 file

From figure R below, it is clear that burlywood sandpaper, blue matt, and blueviolet marble have the highest average markups, suggesting these products bring in the greatest profit relative to their selling price. On the other end, descriptions such as cyan silk and cornflowerblue matt have noticeably lower markups, indicating smaller profit margins.

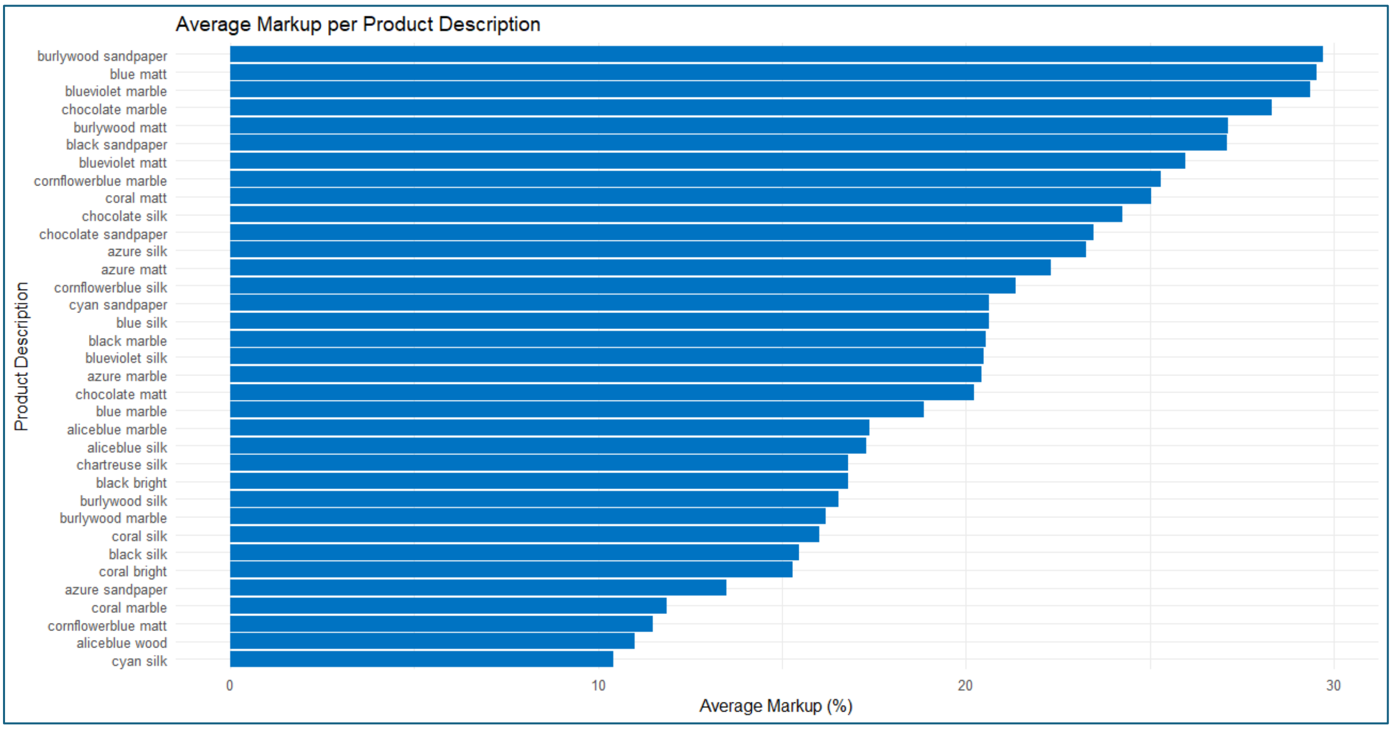


Figure R: Average Markup per Product Description from products_data2025

From the chart, it is clear that Keyboards and Monitors have the highest markups, both reaching close to 24%. This indicates that these two categories are among the most profitable product lines for the business. Mice and Cloud Subscriptions follow closely behind, showing stable profitability as well.

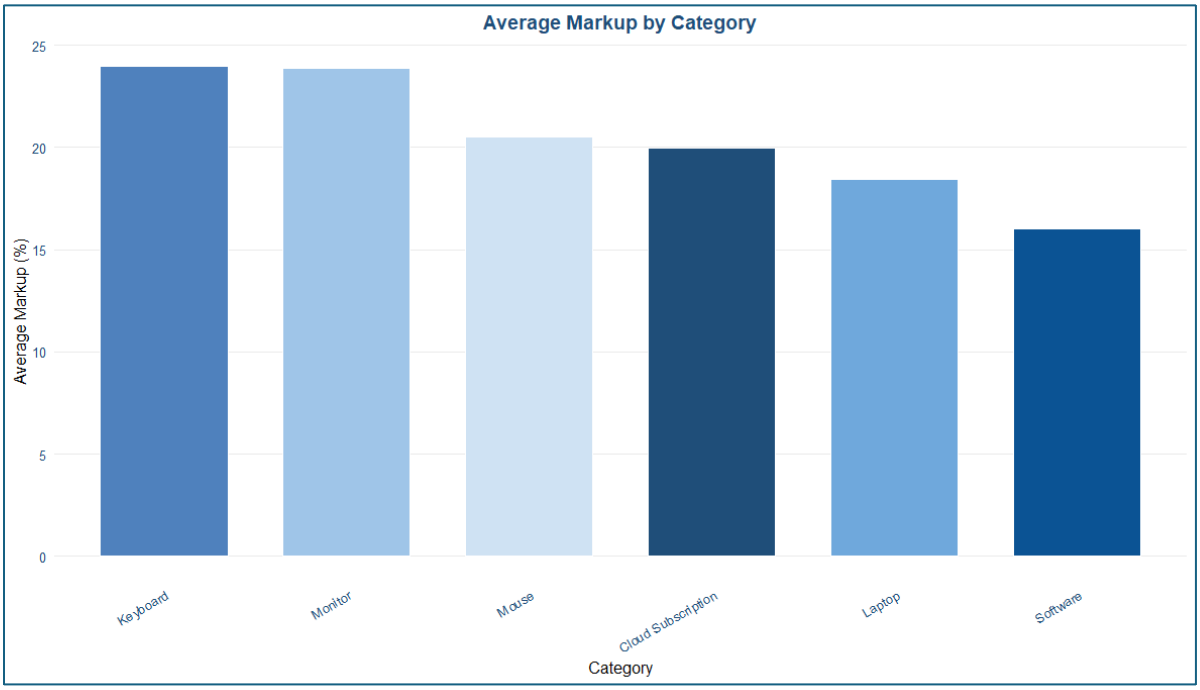


Figure S: Average Markup by Category from products_data2025

From the chart, Black marble, Cornflowerblue matt, and Blueviolet marble stand out as the highest-priced descriptions, all averaging between roughly R18,000 and R19,500. These premium-priced items likely reflect specialized or high-quality materials that justify their higher cost.

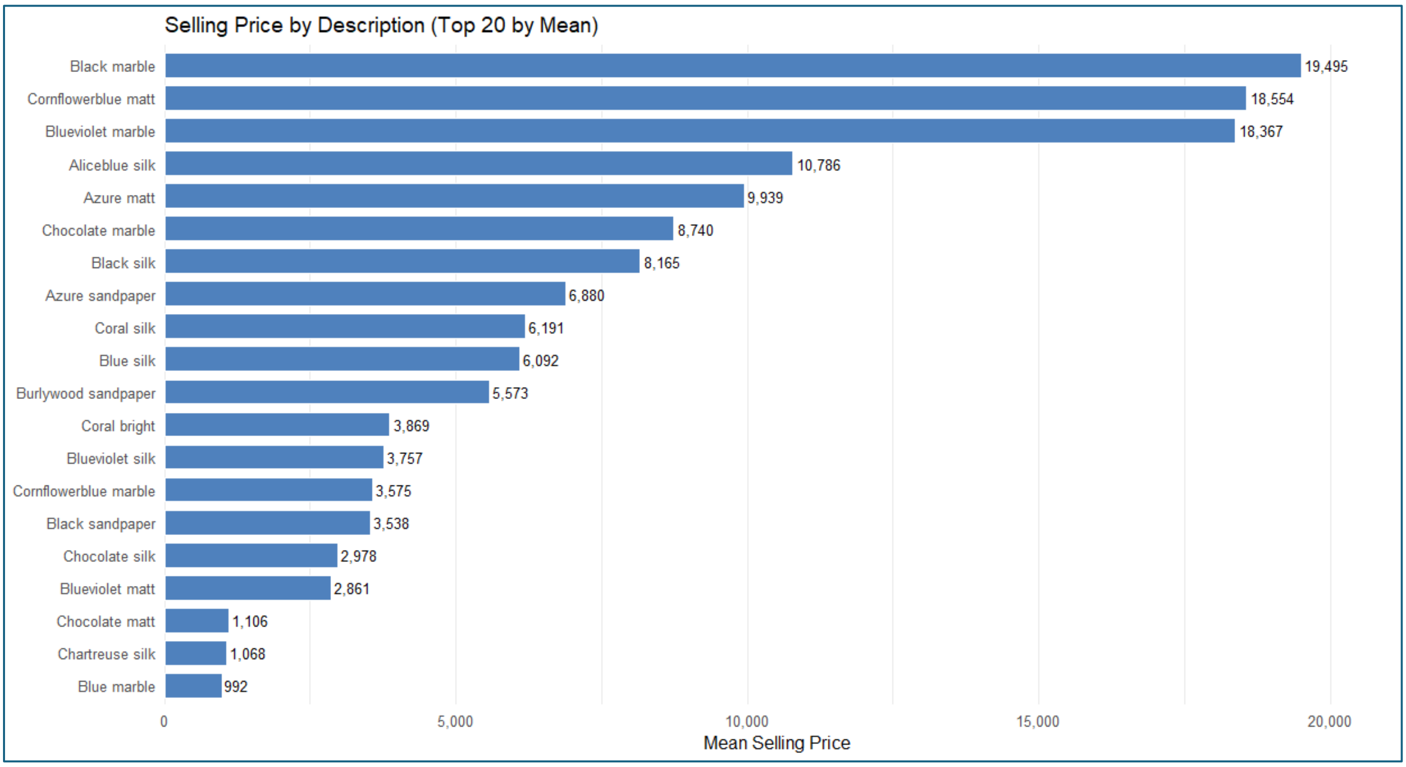


Figure T: Selling Price by Description (Top 20 Mean) form products_data2025

4.4 Insights and Recommendations

From the updated catalogues (products_Headoffice2025 and products_data2025), markups are broadly consistent across categories, with Cloud Subscription and Software showing the highest average margins. Descriptions linked to premium finishes (e.g., “wood” and “marble”) tend to have the highest selling prices, while many “standard” finishes sit in a mid-price band. The product descriptions that appear most often in the range are the “silk” variants.

A key insight is that “silk” is the top-selling description, but it also shows up among the lower selling-price items. This can be read in two ways:

1. There may be headroom to increase price on selected high-volume “silk” SKUs without hurting demand; or
2. Customers may be actively choosing more affordable options, suggesting the business should prioritise value—keeping “silk” products competitively priced while ensuring quality.

It would help to run small, controlled price tests on a subset of high-volume “silk” items (e.g., +3–5%) to measure elasticity. If conversion holds, roll out gradually. Keep a clear “value” tier for price-sensitive buyers. Try and maintain premium “wood/marble” lines as high-margin anchors, while using “silk” as the volume driver. Consider good–better–best bundles to guide customers up the price ladder.

In short, keep premium finishes as margin leaders, protect affordability where demand is strongest (“silk”), and use targeted price tests to lift profit without sacrificing the volume that drives the business.

Part 5

5.1 Introduction

This part focuses on optimizing the performance and profitability of two coffee shops using the provided datasets, *timeToServe.csv* and *timeToServe2.csv*. Both files contain one year's worth of service-time data, showing the number of baristas on duty and the corresponding time taken to serve each customer. Graphs and summaries are included to clearly illustrate the findings and improvements.

5.2 Initial data setup

For all calculations, the following assumptions were used: service level agreement (SLA) = 60 seconds, opening hours = 10 hours per day, revenue = R30 per customer served, and labour cost = R1 000 per barista per day. Since the datasets do not include information about customer demand limits, we performed two separate analyses. In the first analysis, we assumed an unlimited customer demand (constant flow of customers throughout the day for up to six baristas). In the second analysis, we introduced a daily space cap of 800 customers, representing the realistic physical limitation of each coffee shop. These two scenarios allow the head office to compare outcomes and decide which assumption best fits their operational environment and business goals.

5.3 Observations

5.3.1 Daily profit vs baristas with no demand and space caps

Figure U below illustrates the relationship between the number of baristas on duty and the daily profit for two coffee shops, Shop I and Shop II. As shown, both shops experience a clear upward trend — profit consistently increases as more baristas are added. The optimal staffing level for both locations is currently six baristas, where each shop achieves its highest recorded profit (R188,250 for Shop I and R73,350 for Shop II).

These results suggest that neither shop has yet reached the point of diminishing returns, meaning that adding additional staff could potentially continue to improve profitability — provided that customer demand remains steady and space constraints (in this case, capped at 800 customers) allow for efficient operation.

Therefore, both Shop I and Shop II could experiment with employing more baristas, testing whether the profit growth continues or begins to plateau. If additional hiring is not feasible, they should instead focus on optimizing workflow efficiency and service speed among the existing team to approach the same profitability gains without increasing labor costs.

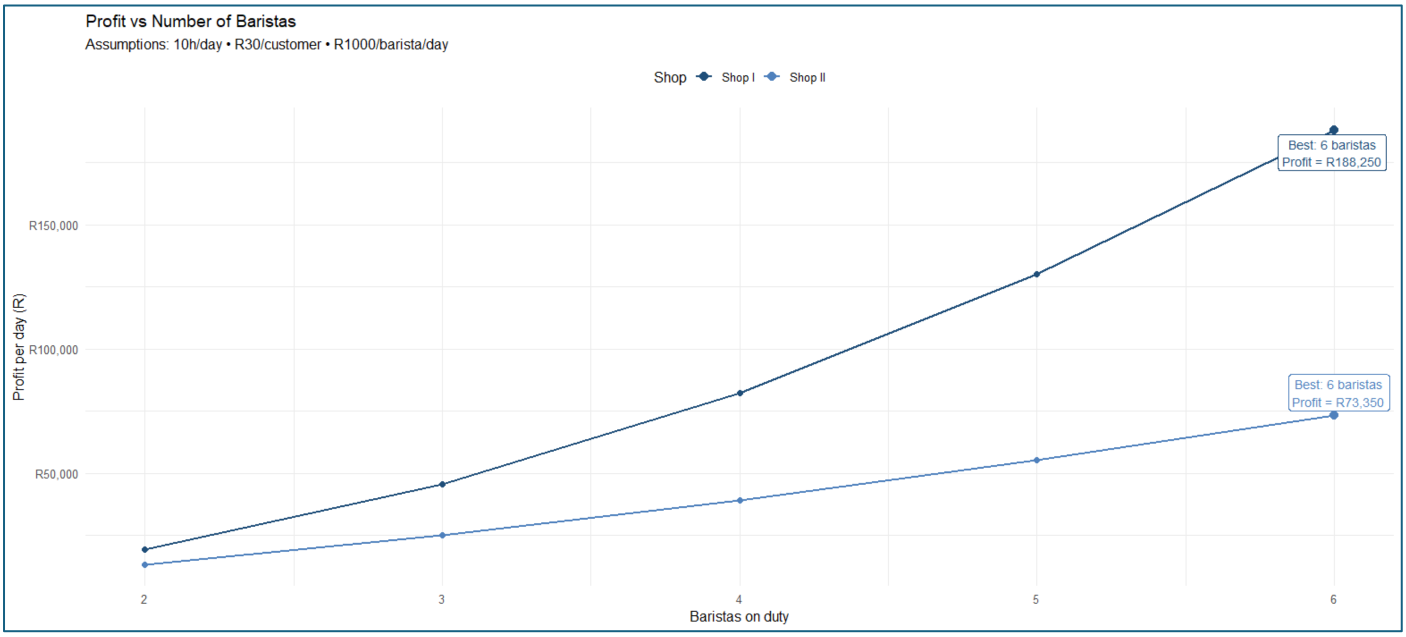


Figure U: Daily Profit vs Baristas with no demand and space caps for Shop I and II

5.3.2 Daily profit vs baristas with demand and space caps

Figure W below shows the relationship between the number of baristas and the daily profit when applying an 800-customer-per-day space cap for both shops. Under this constraint, profits initially rise as more baristas are added, but after a certain point, they begin to decline.

The optimal point for both Shop I and Shop I occurs at three baristas, where each shop reaches its highest achievable profit of approximately R21,000 per day, serving the full capacity of 800 customers. Adding more baristas beyond this point reduces overall profit, as the additional labor cost outweighs any improvement in service time—since customer capacity is already maxed out.

This finding suggests that when space or demand is limited, efficiency—not staff quantity—is the key driver of profitability. Both shops should maintain around three baristas per shift, focusing instead on streamlining service flow and maintaining quality rather than expanding staff numbers.

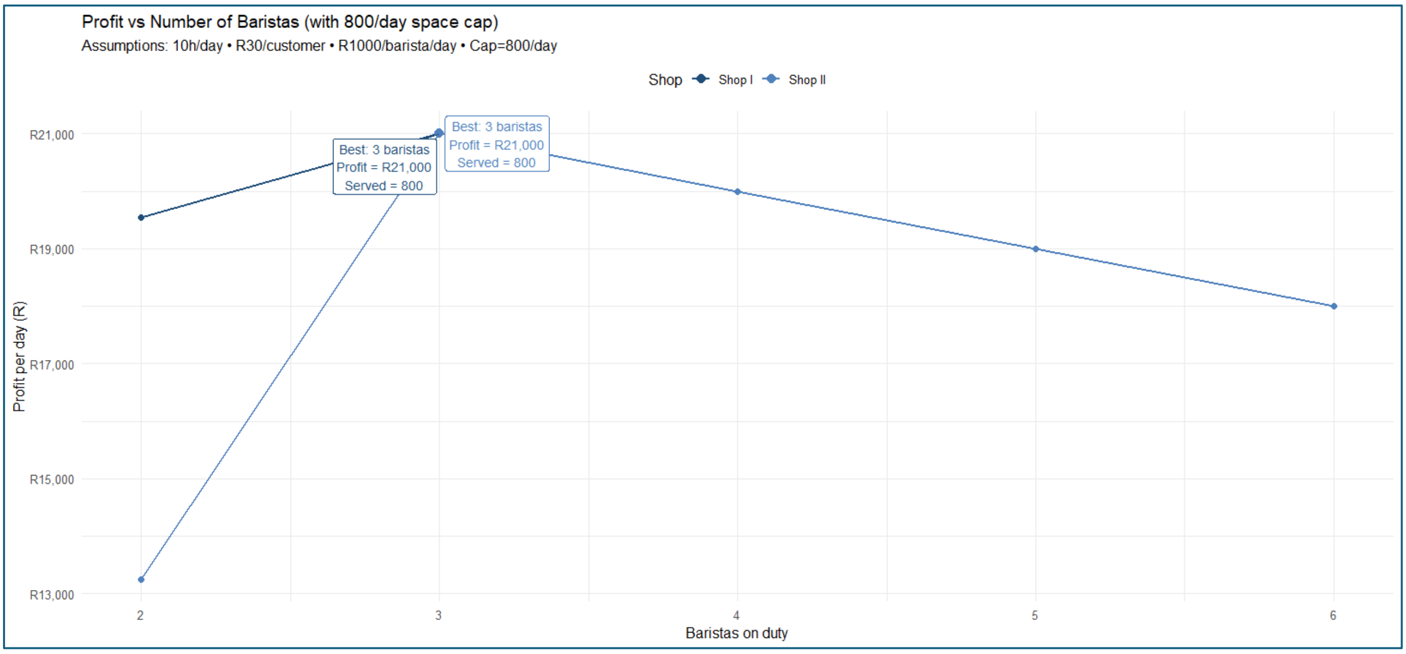


Figure V: Profit vs Number of Baristas with demand and space caps for Shop I and II

5.4 Insights & Recommendation

From the profit optimization analysis, it is evident that both coffee shops benefit from a balance between staffing levels and operational efficiency. When no customer space limitation was considered, profits steadily increased with the number of baristas, peaking at six. This indicated that under unrestricted demand, adding more staff directly improved customer throughput and overall profitability. However, when a realistic space cap of 800 customers per day was applied, the results changed significantly—showing that profit no longer increases indefinitely with more baristas. Instead, the highest profit for both Shop I and Shop II was achieved with three baristas, after which profits began to decline due to higher staffing costs without any increase in customers served.

This highlights a critical insight for management: beyond a certain point, adding more staff creates diminishing returns. In a constrained environment, such as limited shop space or a fixed number of daily customers, operational efficiency and smart scheduling become far more important than simply increasing manpower. The current results suggest that both shops can maximize profitability by maintaining three baristas per shift, ensuring that service is fast and consistent without incurring unnecessary labor costs.

Therefore, I recommend that Head Office investigate the actual space capacity of each shop and assess the weekly customer demand to perform a more accurate and thorough analysis.

Part 6

6.1 Introduction

Part 6 tests, with statistics, whether our process times really change across the calendar and are big enough to matter operationally. Using the 2026–2027 data, we analysed picking and delivery hours for each product family with one-way/two-way ANOVA and a combined MANOVA, and then used Tukey post-hoc comparisons to see exactly which months differ. The focus is not on code but on decisions: do averages differ by year and by month, and when should Operations expect slower service? For the SOF family (shown), the year effect is negligible, while the month effect is strong—times climb from early in the year and peak in the last quarter, with December clearly highest. This evidence supports seasonal staffing and capacity planning: keep baseline staffing earlier in the year and add capacity from September onward, especially in December, to protect reliability and profit.

6.2 Observations

6.2.1 SOF – Delivery hours by YEAR

This plot compares the average delivery time for SOF across the two years, with the thin bars showing 95% confidence intervals. The two points sit at almost the same level (≈1.09 h in 2022 and ≈1.09 h in 2023), and the error bars overlap a lot. That means any year-to-year difference is tiny and not statistically meaningful. This matches the ANOVA result we ran earlier ($p \approx 0.67$), which says there's no evidence that the average delivery time changed between 2022 and 2023.

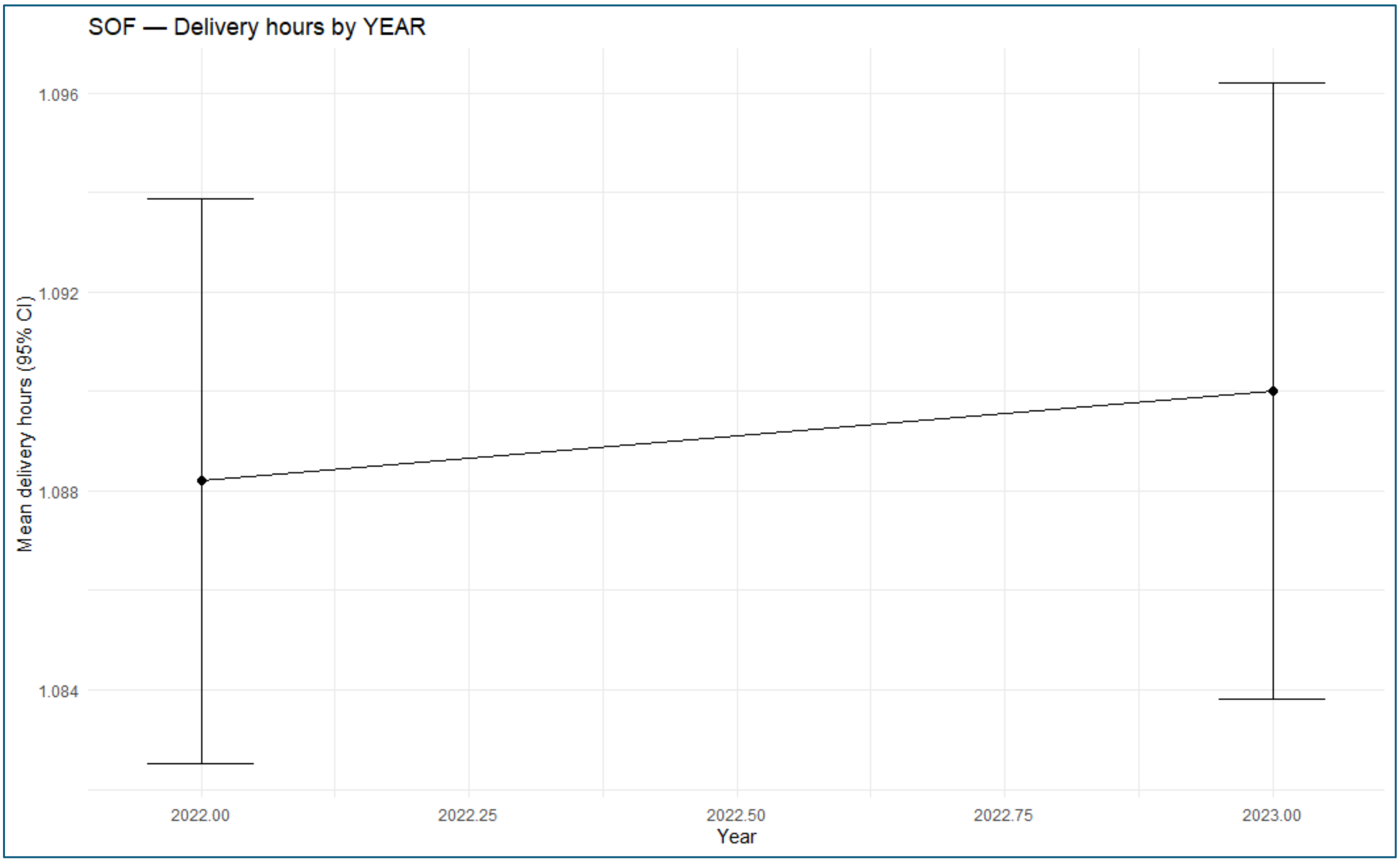


Figure W: SOF delivery hours by year

6.2.2 SOF – Delivery hours by MONTH

This figure plots the average delivery time for SOF by month with 95% confidence bars, split by 2022 and 2023. In both years there's a clear upward trend as the year progresses: times are lowest at the start of the year (≈0.95–1.00 h) and steadily climb towards year-end, peaking around November–December (≈1.20–1.25 h). The error bars mostly don't overlap between early months and late months, which backs up the stats we ran (the month effect was highly significant). 2023 follows the same pattern as 2022 with very similar levels, with a tiny dip around July before continuing up again.

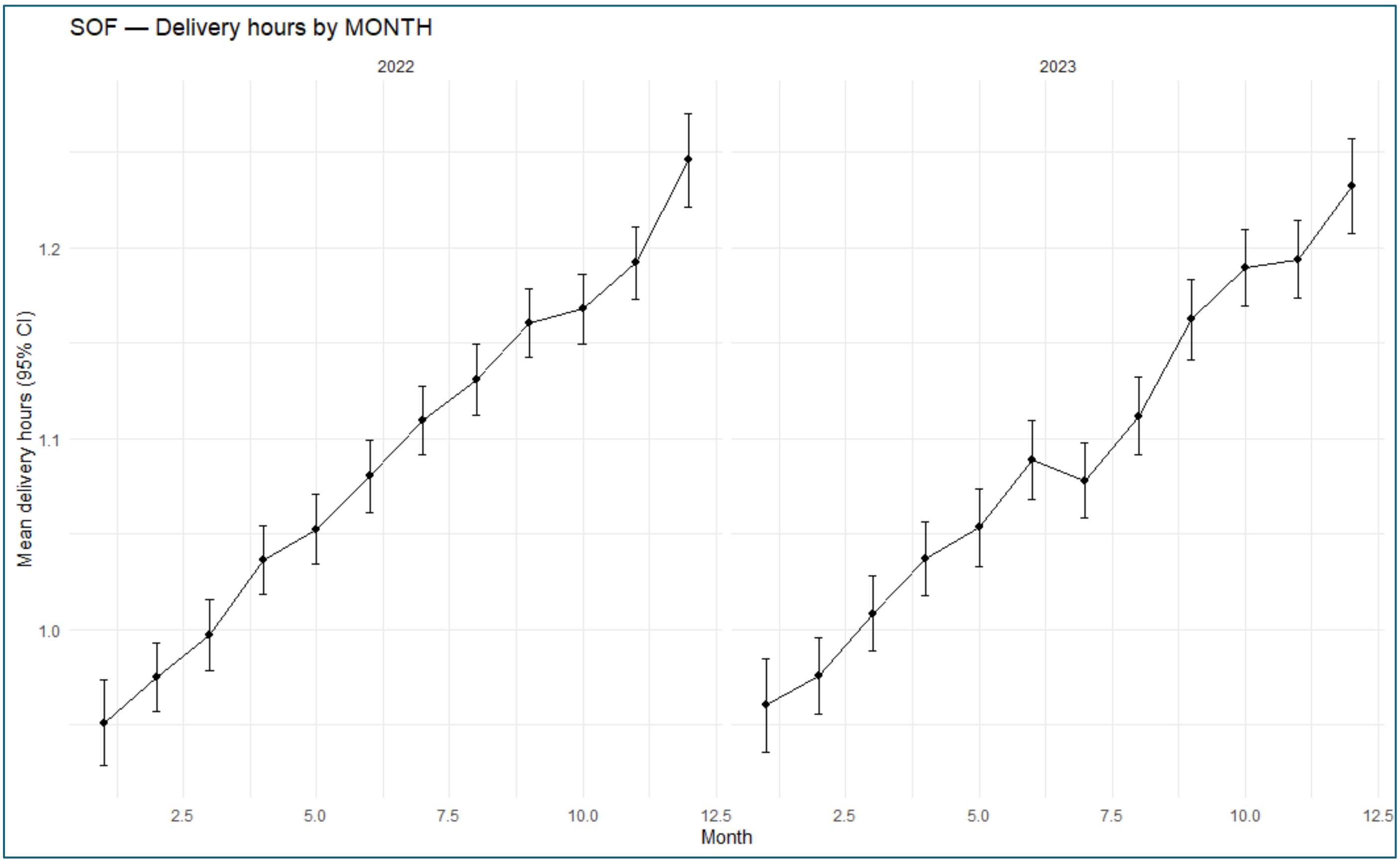


Figure X: SOF delivery hours by month

6.3 Insights & Recommendation

From the ANOVA/MANOVA we ran on SOF, the month of the year has a strong, statistically significant effect on delivery time, while the year itself does not. In other words, performance is broadly the same between 2022 and 2023, but it changes a lot within a year. The month-by-month plots back this up: delivery hours are lowest at the start of the year and rise steadily towards Nov–Dec. The Tukey post-hoc test also shows that late-year months are significantly slower than early-year months. Because the year effect is not significant, these seasonal patterns are the main driver rather than a permanent year-to-year shift.

Operationally this means we should plan around seasonality. For Q4 we should assume longer service times and add capacity before the curve bends upward. That can be as simple as scheduling extra staff, extending peak shifts, and pulling forward prep work so baristas/pickers spend less time per order. We should set different monthly targets rather than one flat SLA for the whole year, using early-year performance as a realistic baseline and tightening processes to keep late-year months closer to that level.

It's also worth linking this back to SPC. The s-charts showed stable variation, so the spread is under control. The issue is the mean drifting upward in peak months. We should focus improvements on steps that reduce average time—queueing, batching, station layout, and hand-offs—then re-check the X-bar charts after the changes. If we keep the monthly mean from creeping up while variation stays stable, we'll protect reliability without needing to over-staff.

Finally, we should repeat the same analysis for the other product families and for picking hours, and use the same seasonal staffing and process adjustments wherever we see the same pattern. This keeps the plan consistent: hold the process stable (SPC), plan for seasonal load (ANOVA/MANOVA), and review after changes to confirm the gains.

Part 7

7.1 Introduction

This part focuses on evaluating and optimising the reliability of service at a car rental agency based on staffing levels. Using the provided data, we estimate how often the company can expect reliable service and identify the optimal number of employees needed to minimise financial losses caused by understaffing. The analysis also considers the cost of hiring additional personnel and helps determine a balance between operational reliability and profitability.

7.2 Initial data setup

For this section, the available data showed how many staff members were on duty each day over a total of 397 days. The number of workers ranged from 12 to 16 per day, with each level associated with a specific number of days worked. This data was summarised into a table showing the number of workers, number of days, and probability of occurrence, and then visualised in a graph. The graph helps to clearly illustrate how often the company operated with sufficient staff to provide reliable service.

7.3 Observations

7.3.1 Reliability of Service

Figure below shows the probability distribution of staffing levels at the car rental agency over 397 days, ranging from 12 to 16 workers on duty. Each bar represents how often a specific staffing level occurred, expressed as both number of days and percentage probability. The light-blue bars represent unreliable service (fewer than 15 workers), while the dark-blue bars indicate reliable service (15 or more workers on duty).

From the graph, we can see that the agency had 16 workers on 270 days, which accounts for 68% of the total days. Therefore, the company can expect to have reliable service on approximately 68% of the days in a year, assuming current staffing patterns continue. This means that for most of the year, operations are likely to run smoothly and meet service expectations, but there is still room to improve staffing consistency to reduce the number of unreliable days

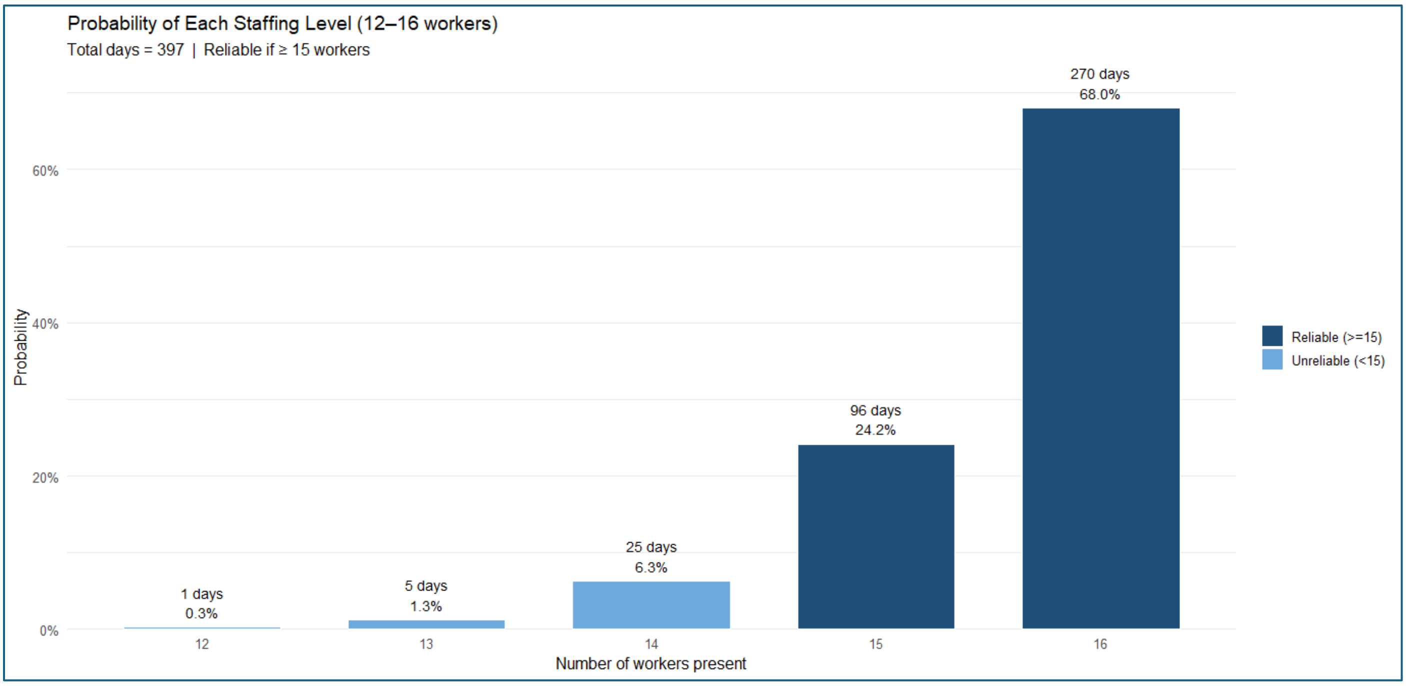


Figure Y: Probability of each staffing level (12-16 workers)

7.3.2 Optimise Profit

This graph shows how the expected total monthly cost changes as the company increases or decreases the number of employees on duty. It combines both the monthly personnel cost (R25 000 per person) and the expected daily revenue losses when there are too few workers (fewer than 15).

From the analysis, the model identifies 17 employees as the optimal staffing level, giving the lowest total monthly cost of approximately R430 443. Staffing fewer than 17 people leads to a sharp increase in costs due to the higher likelihood of service disruptions and the resulting R20 000 loss per problem day. Conversely, hiring more than 17 people raises expenses unnecessarily, as additional salaries outweigh the benefits of improved reliability.

In summary, employing 17 staff members minimizes total monthly costs while maintaining reliable service levels. This provides a balanced point between efficiency and reliability — ensuring smooth operations without overspending on personnel.

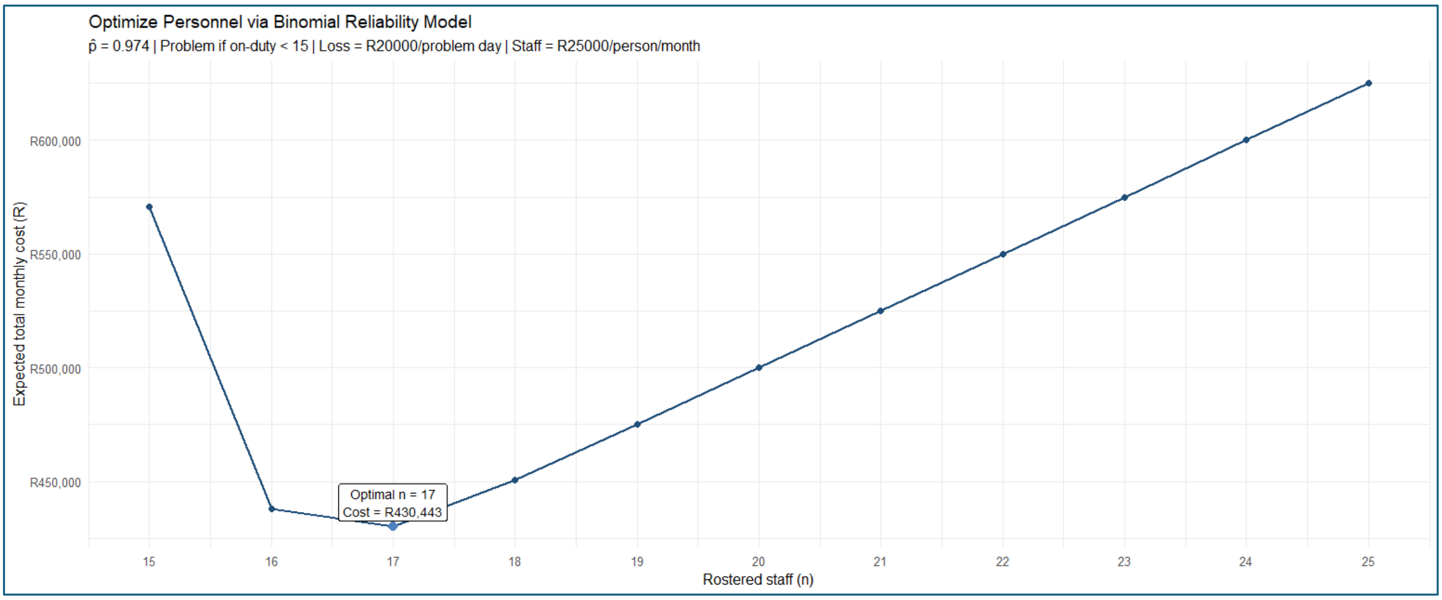


Figure Z: Binomial reliability model

7.4 Insights and recommendations

From the analysis, it is clear that the company's reliability and profit are highly sensitive to staffing levels. The reliability model shows that the business experiences service problems when there are fewer than 15 people on duty — leading to significant daily revenue losses of about R20 000 per day. By modelling the situation using a binomial reliability approach, it was determined that 17 employees represent the optimal staffing level, minimizing total expected monthly costs to approximately R430 443.

Having fewer than 17 staff members significantly increases the probability of service issues, while hiring more than 17 results in diminishing returns, as the additional wages exceed the potential revenue gains. Therefore, maintaining the staff complement at around 17 employees provides the most cost-effective balance between reliability and labour expense.

The company should set a staffing policy that ensures at least 17 employees are scheduled per day to maintain reliable service and minimize financial losses. Keeping the daily workforce at or above this level will help prevent disruptions that lead to reduced sales and customer dissatisfaction. It is also essential to monitor attendance and daily staffing levels closely, as even minor absences can push the operation below the reliability threshold, increasing the likelihood of service problems.

In addition, management should track seasonal demand fluctuations — for instance, during holidays or peak travel times — when customer volumes are typically higher. During these periods, hiring temporary or part-time staff can help sustain reliability without overcommitting to long-term employment costs. Lastly, the company should continuously track and analyse data on sales, service delays, and staff attendance. This will enable the organisation to refine the staffing model over time and make data-driven adjustments to personnel levels as business patterns and customer demands evolve.

Reference table

R Core Team. (2024). *R: A language and environment for statistical computing*. Vienna: R Foundation for Statistical Computing. Available at: https://www.r-project.org/ [Accessed 24 Oct. 2025].

Wickham, H. and Grolemund, G. (2017). R for Data Science: Import, Tidy, Transform, Visualize, and Model Data. Sebastopol: O'Reilly Media. Available at: https://r4ds.had.co.nz/ [Accessed 24 Oct. 2025].

OpenAI. (2025). *ChatGPT (GPT-5)* [Large language model]. Available at: https://chat.openai.com/ [Accessed 24 Oct. 2025].

DataCamp. (2023). *What is data analysis? Expert guide*. Available at: https://www.datacamp.com/blog/what-is-data-analysis-expert-guide [Accessed 24 Oct. 2025].

Figures

Tables