

Business Analysis Report

Author: 27204650

Table of Contents

1. Data Summary	3
Sales Performance Analysis	3
Operational Correlations.....	6
Product Clustering & Recommendations.....	6
Conclusion	7
3. Data Summary	8
Findings and Summaries	8
Process Variability.....	8
Process Mean	9
Process Capability	9
SPC Rules.....	9
Insights.....	10
Recommendations	10
Conclusion	11
4.1 Type 1 error	12
4.2 Type 2 error	12
4.3 Updated Analysis	13
Data Summary	13
Sales Performance Analysis	13
Clustering	14
Clustering Results.....	14
Cluster Overview.....	15
Cluster Visualisation Insights.....	15
Business Implications	16
Conclusion	17
5. Staffing.....	18
Assumptions:.....	18
Key Findings	18

Summary	19
6 Analysis of Service Time by Number of Baristas	20
Introduction	20
Results	20
Interpretation	20
7 Reliability and Staffing Optimization in Car-Rental Operations	21
Introduction	21
Results	21
Interpretation	21
Conclusion	21
Appendix.....	22
Appendix A: Product Sale Details	22
Appendix B: Price vs Markup Clustering Results	23
Appendix C: Avg Quantity and Product Clustering Results	24
Appendix D: Monthly Quantity Sold Clustering Results	25
Appendix E: Expected Daily Revenue.....	26
Appendix F: Expected Daily Costs	27
Appendix G: Data Summary	28

1. Data Summary

Not much valuable information can be quickly summarised from the data that will necessarily help to guide any business decisions, instead I decided to perform a comprehensive analysis of the data to help uncover tangible insights.

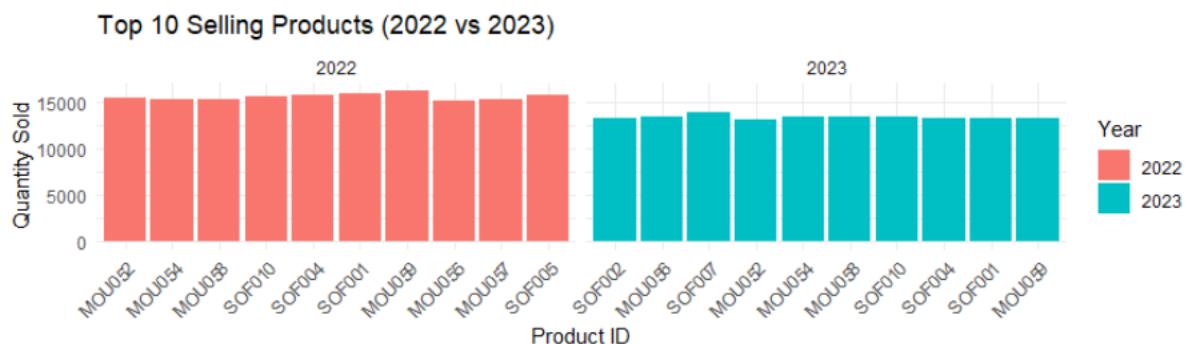
Key findings from the data:

- Revenue is concentrated in a small set of products.
- Customers are clearly segmented by income, age, and gender, with higher income groups spending significantly more.
- The company is experiencing a financial decline in sales.
- Clustering analysis highlights high-volume, low-markup products where price increases could be tested, and low-volume, low-markup products where discounts or promotions may be necessary.

Sales Performance Analysis

Order Quantities

- The order quantities for each product are very similar to one another, there is not much variance.
- Notably, the highest selling product is the “aliceblue marble” keyboard

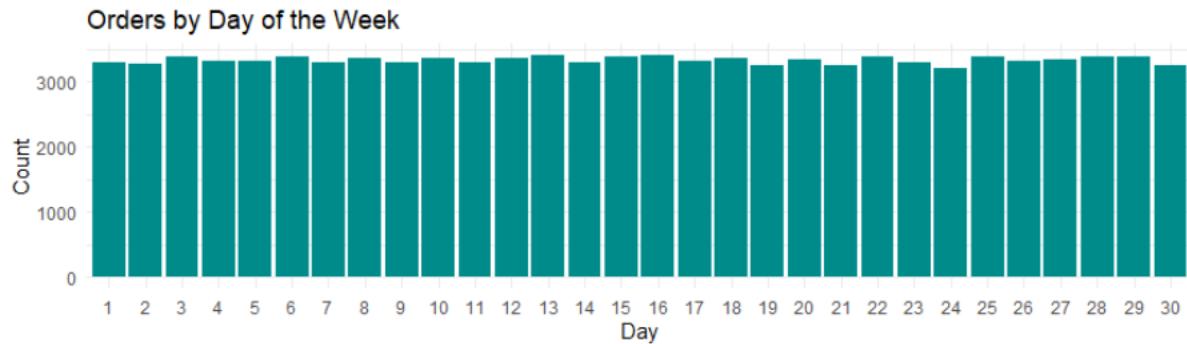


Category Performance

- Hardware categories (such as laptops and monitors) generate the highest revenue.
- Software and subscription products show consistent though smaller contributions, often with higher margins.

Daily Trends

- Daily sales appear to be relatively stable and consistent throughout the month.
- This horizontal-level trend is perfect for forecasting as demand is relatively the same each day.



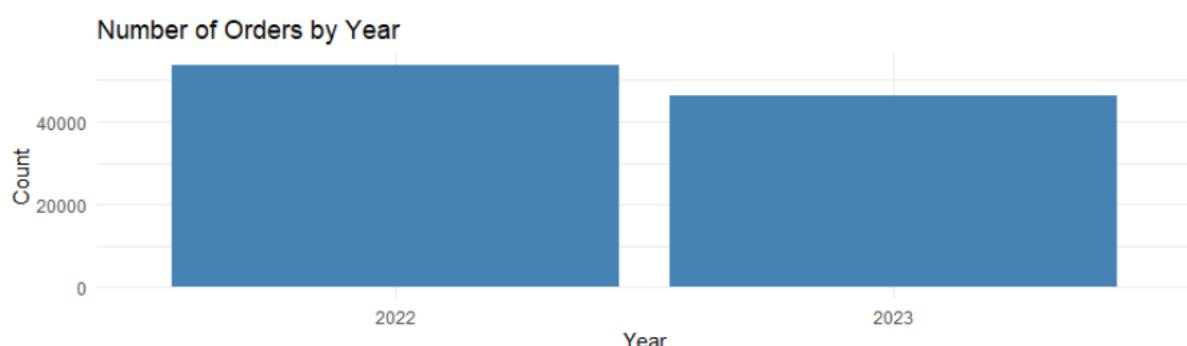
Monthly Trends

- Revenue indicates seasonality trends as well as highly consistent sales throughout most of the year.
- Staffing, marketing, and procurement should be adjusted around these cycles.



Yearly Trends

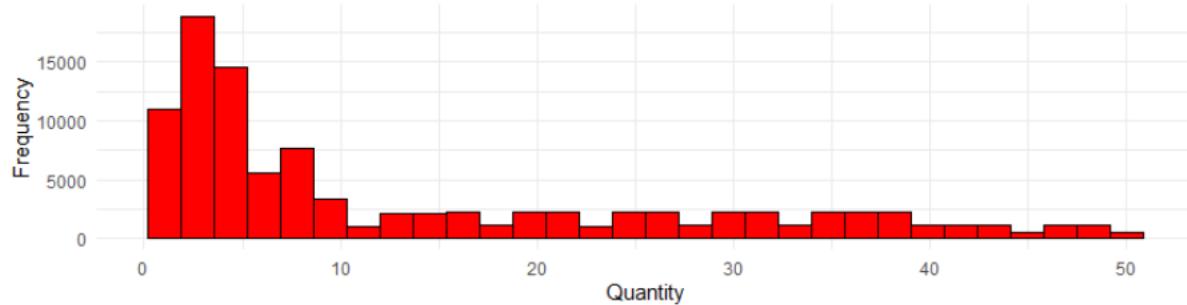
- There is a clear decline in sales from 2022 to 2023, this is a serious sign of financial decline. The company can remedy this issue through multiple efforts such as:
 - Run promotions throughout the year.
 - Adjust staffing for seasonality.
 - Adjust procurement to allow for a cost-efficient strategy if inventory costs are holding the business back from a healthy profit.



Order Quantity Analysis

- Customers appear to prefer ordering products in smaller quantities, this could be due to clients maximising their responsiveness within their own businesses and wanting flexibility. We could run promotions for higher quantity purchases to increase our lot sizing.

Distribution of Order Quantities

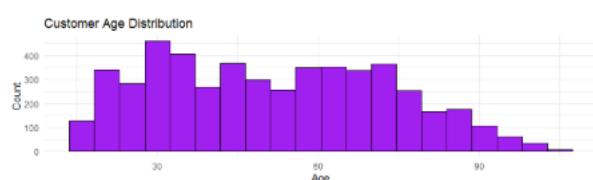


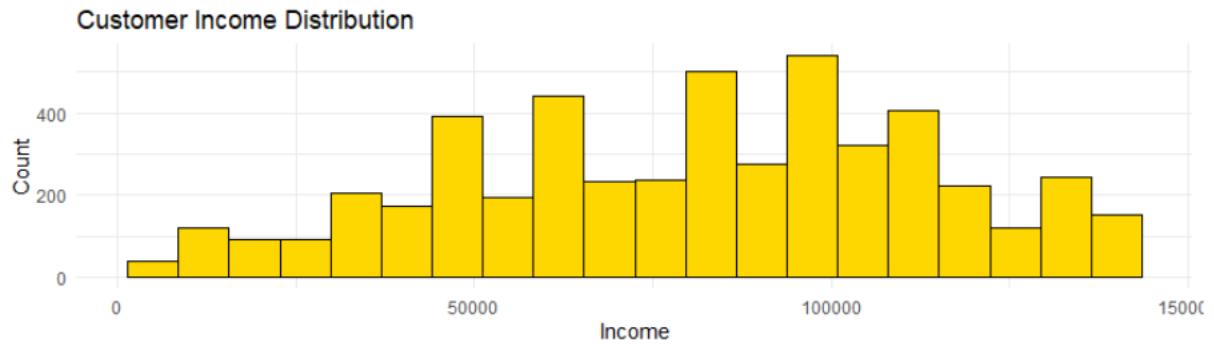
Profitability Analysis

- Margin percentages vary by product.
- High-revenue products with low margins present opportunities for small price increases.
- Low-revenue but high-margin products present opportunities to increase marketing and visibility.

Customer Demographic

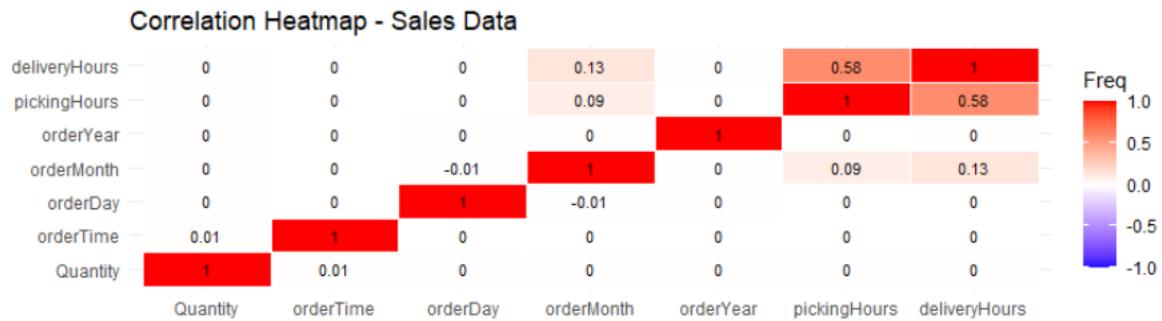
- The company has a relatively even split of male and female customers, we are able to reach a large audience due to this non-bias.
- The median age which purchases supplies from the company is approximately 30, however analysis shows that there is a large range of age groups who purchase from us as well.
- The income distribution of our customers is not stable, there are peaks over particular income ranges, and it would appear as if the median income range is approximately in the 100,000's.
- Overall, our products suit the needs of an incredibly large audience, this is great for brand awareness but could be detrimental to our competitive side. If the company were to shift and focus on one demographic to focus resources such as marketing and product placement, it would be the high-income, middle-aged segment.





Operational Correlations

- Picking hours correlate with selling price: expensive items often require longer handling.
- Delivery hours correlate with both picking time and order size.
- Streamline processes for high-value items to reduce picking/delivery time.



Product Clustering & Recommendations

Cluster 1 – High-volume, Low-markup

- Products that sell a lot of units but with thin profit margins.
- They generate stable revenue but don't maximize profitability.
 - Test a small price increase.
 - See if sales volume holds, if it holds then keep the price increase.

Cluster 2 – High-volume, High-markup

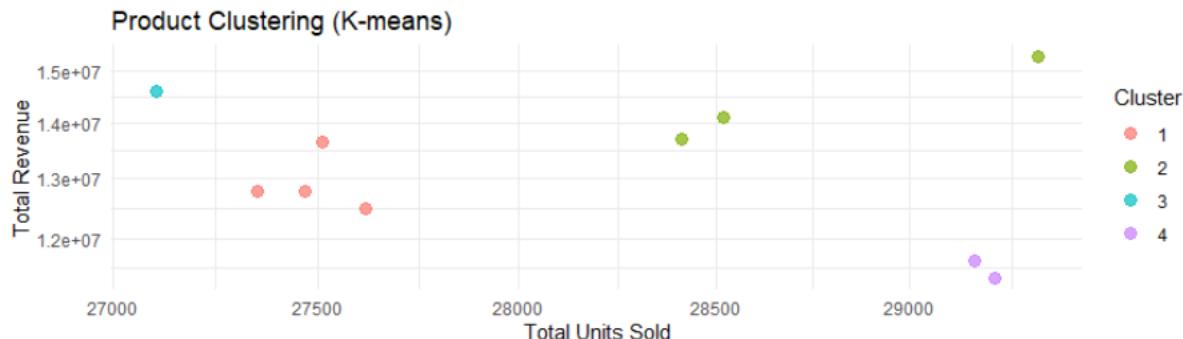
- Products that sell well and have healthy margins.
 - Make sure stock never runs out.
 - Give these products visibility in marketing.

Cluster 3 – Low-volume, Low-markup

- Products with low sales and poor margins.
- They take up inventory but don't contribute much to revenue or profit.
 - Decide if they are necessary.
 - If not, phase them out or discount them to clear up inventory.

Cluster 4 – Low-volume, High-markup

- Products with low unit sales but very high margins.
- Often premium or specialized items.
 - Market more aggressively to high-income customers.



Conclusion

The analysis indicates that the company has strong foundational assets but declining sales. At the product level, revenue is highly concentrated in a small set of products. This creates both risk and opportunity.

High-volume, low-margin products can be optimized with modest price adjustments, while low-volume, high-margin products require targeted marketing to high-income customers. Similarly, underperforming products should be carefully reviewed to free up resources and space for the better performers.

Customer demographics reinforce that while the company appeals to a very wide audience, the highest-value segment is middle-aged, higher-income customers. Concentrating marketing and loyalty efforts on this group will provide stronger returns than a broad, unfocused approach.

Operational correlations highlight efficiency gaps, particularly in the handling and delivery of high-value orders. Improving these processes could reduce costs and improve service levels. By executing these recommendations, the company can stabilize sales, improve profitability, and steady itself for sustainable growth in the coming years. In conclusion, the company should adopt a differentiated strategy.

Key Notes:

- Expand high-volume, high-margin products.
- Optimize pricing on high-volume, low-margin products.
- Reassess weak performers (low-volume, low-margin products).
- Market to customers for premium, high-margin products.
- Prioritize high-income, middle-aged customers as the primary growth segment.

3. Data Summary

Raw data was analysed to reveal any underlying information amongst the six major product families (CLO, LAP, MOU, etc.). The following insights were revealed:

- Process capability is generally below acceptable industry limits. This indicates that most product lines fail to consistently meet delivery specifications.
- The SOF product family performs exceptionally well and can achieve Cp and Cpk values exceeding 1.15. This demonstrates a stable, predictable process.
- Hardware-related families exhibit increases in mean delivery times, over time. This is a sign of inefficiency.
- S-Charts show very stable variability across all individual products, confirming that the process spread is consistent. The X-bar charts, however, indicate instability in the process mean. This implies that the system is *precise but not accurate*.
- The SPC rule evaluation found frequent violations of rules 2 and 3. This indicates gradual, non-random shifts in delivery performance.

Findings and Summaries

Process Variability

- Across all products, variability remained stable and within control limits.
- No major outlier (with respect to variation) was observed.

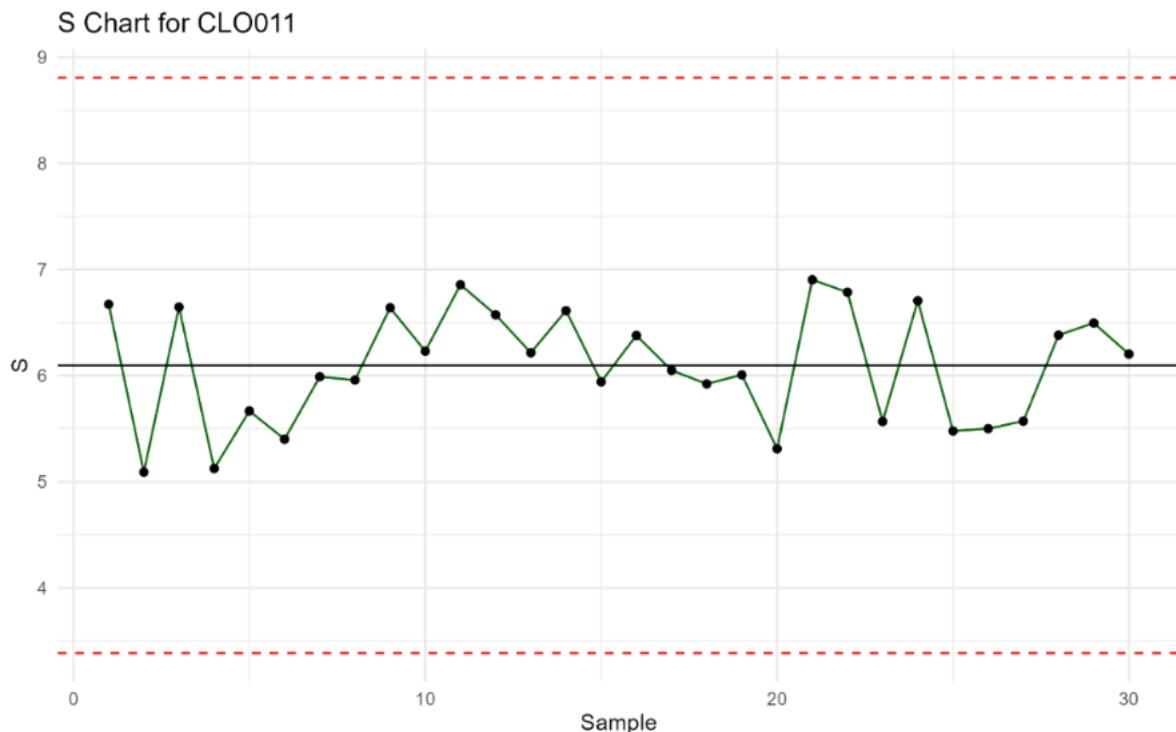


Figure 1 example of general trend

Process Mean

- Many products showed gradual upward trends (with regards to delivery times), this indicates that customers are waiting longer over time.
- SPC rule 2 and 3 violations indicate trends, not random noise.

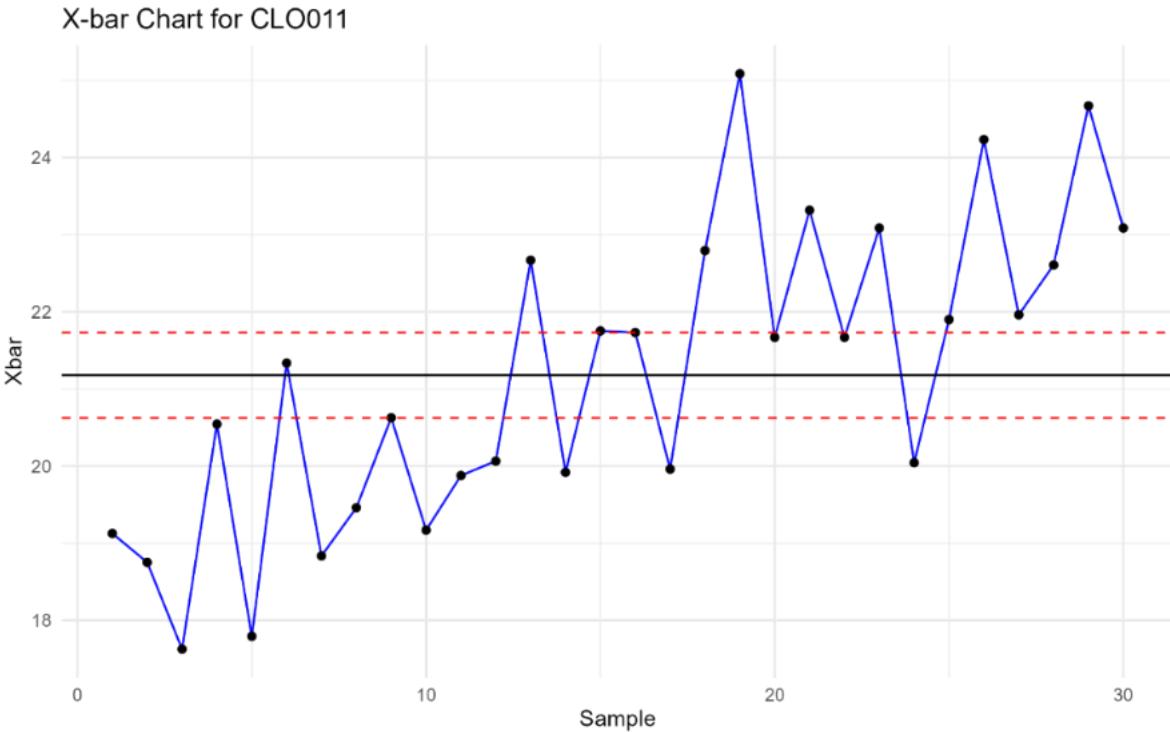


Figure 2 example of general trend

Process Capability

- Most product families are not capable ($C_p < 1$ and $C_{pk} < 1$).
- Products, SOF001, SOF003, SOF004, SOF006, SOF008 and SOF010 show very high C_p and acceptable C_{pk} . This indicates that these products are capable and reliable.
- This trend suggests that software delivery processes are better controlled compared to hardware deliveries.

SPC Rules

- Rule 1, rarely violated, the process is consistent.
- Rule 2, frequently violated, this indicates process mean shift.
- Rule 3, violated often, this confirms drift in delivery performance.
- Rule 4, no cycling patterns are viewable.

Product	Cp	Cpk	Cpu	Cpl	Capable	Rule1	Rule2	Rule3	Rule4
KEY001	0.9142	0.5716	0.5716	1.2567	No	TRUE	TRUE	FALSE	FALSE
CLO001	0.9201	0.5587	0.5587	1.2814	No	TRUE	TRUE	TRUE	FALSE
LAP001	0.9128	0.5661	0.5661	1.2595	No	TRUE	TRUE	TRUE	FALSE
MOU001	0.9067	0.5733	0.5733	1.2400	No	TRUE	TRUE	TRUE	FALSE

MON001	0.8891	0.5554	0.5554	1.2229	No	TRUE	TRUE	TRUE	FALSE
SOF001	17.7013	1.2337	1.2337	18.1690	Yes	TRUE	TRUE	TRUE	FALSE
SOF002	0.8625	0.5750	0.5750	1.1500	No	TRUE	TRUE	FALSE	FALSE
SOF003	17.9072	1.1938	1.1938	18.6207	Yes	TRUE	TRUE	TRUE	FALSE
SOF004	17.7168	1.2089	1.2089	18.2248	Yes	TRUE	TRUE	TRUE	FALSE
SOF005	0.8626	0.5692	0.5692	1.1560	No	TRUE	TRUE	FALSE	FALSE
SOF006	17.7391	1.2255	1.2255	18.2526	Yes	TRUE	TRUE	TRUE	FALSE
SOF007	0.8587	0.5667	0.5667	1.1507	No	TRUE	TRUE	FALSE	FALSE
SOF008	17.7446	1.1560	1.1560	18.3332	Yes	TRUE	TRUE	TRUE	FALSE
SOF009	0.8673	0.5616	0.5616	1.1729	No	TRUE	TRUE	TRUE	FALSE
SOF010	17.8387	1.1870	1.1870	18.4904	Yes	TRUE	TRUE	TRUE	FALSE

Figure 3 Snapshot of several products and their properties

Insights

- Capability among the products shows clear weakness. Despite stable variability, most processes are not consistently meeting specification limits.
- Hardware related deliveries show the greatest risk and are falling behind customer expectations.
- Software deliveries are a strong benchmark for success. These products could provide process improvement lessons for other the other products.
- Delivery times trends suggest inefficiencies in the company's structure.

ProductGroup	Avg_Cp	Avg_Cpk	Capable_Count	Total_Products	Proportion_Capable
SOF	17.63	1.18	10	10	100
CLO	0.87	0.58	0	10	0
MON	0.88	0.58	0	10	0
LAP	0.88	0.57	0	10	0
MOU	0.87	0.57	0	10	0
KEY	0.87	0.56	0	10	0

Recommendations

Focus on Non-Capable Processes

- Prioritize improving hardware-related products.
- Use *root cause analysis* on procurement, supplier performance, and transport processes to remedy any issues within those sectors

Utilize Best Practices from SOF products

- Use software delivery as a benchmark for efficiency and stability.
- Apply similar scheduling, planning, or monitoring practices/tools to hardware-related deliveries.

Address Mean Delivery Trend

- Make changes to the process to keep delivery times consistent. Examples include,
 - Better supplier contracts and safety stock improvements.

Capability Goals

- Aim for $C_p \geq 1$ and $C_{pk} \geq 1$ across all products. This ensures that the process capability meets industry standards.

Conclusion

My analysis shows that most product delivery processes are not capable, despite being stable in variability. Only SOF product deliveries are capable and meet industry benchmarks. The main issue is mainly with respect to the upward trend in delivery times for hardware related products, which must be addressed to improve customer satisfaction and business efficiency.

By focusing on hardware deliveries while making use of the SOF products practices and methods, the company can achieve greater consistency, reduce delays, and significantly improve overall process capability.

4.1 Type 1 error

Question A

- $P(Z > 3) = 1 - P(Z < 3)$
- $P(Z > 3) = 1 - 0.998650102 = 0.001349898$
- **0.001349898**

Question B

- $P(|Z| \leq 1) = X(1) - X(-1)$
- $0.841344746 - 0.158655254 = 0.682689492$
- $(0.682689492)^k$
 - $k=7: (0.682689492)^7 = 0.06811344283 = 6.9\%$
- **6.9%**

Question C

- $P(Z > 2) = 1 - P(Z < 2)$
- $P(Z > 2) = 1 - 0.977249868 = 0.022750132$
- $P(Z > 2)^4$
 - $= (0.022750132)^4 = 0.00000026787$
- **0.0000267%**

4.2 Type 2 error

- $P(25.011 < X < 25.089)$
- $= P([25.089-25.028]/0.017) - P([25.011 - 25.028]/0.017)$
- $= P(3.588235294) - P(-1.00)$
- $= 0.99983 - 0.158655254 = 0.84117$
 - **P(Type 2 Error) = 84.1%**
 - **Power = 1 - P(Type 2 Error) = 15.9%**

4.3 Updated Analysis

Not many changes were observed during the re-analysis of my data due to the bulk of my analysis during week 1 being mainly focused on the files, “sales2022and2023” as well as “customer_data”. This was done due to the errors present in the products_data and products_Headoffice files, rendering them essentially useless for data collection. I have introduced the new products_data and products_headoffice 2025 files into my analysis this round and the results have been compiled.

Data Summary

- Categories differ strongly

Laptops and Monitors often have much higher selling price and contribute large portions of revenue, all other products contribute very little to overall revenue by comparison.

The week 1 analysis of files “products_data.csv” and “products_Headoffice.csv” data revealed substantial inconsistencies between the two, this included mismatched product identifiers, missing prefixes “NA”, and incorrect selling price and markup values after the first ten records per product type.

After applying fixes:

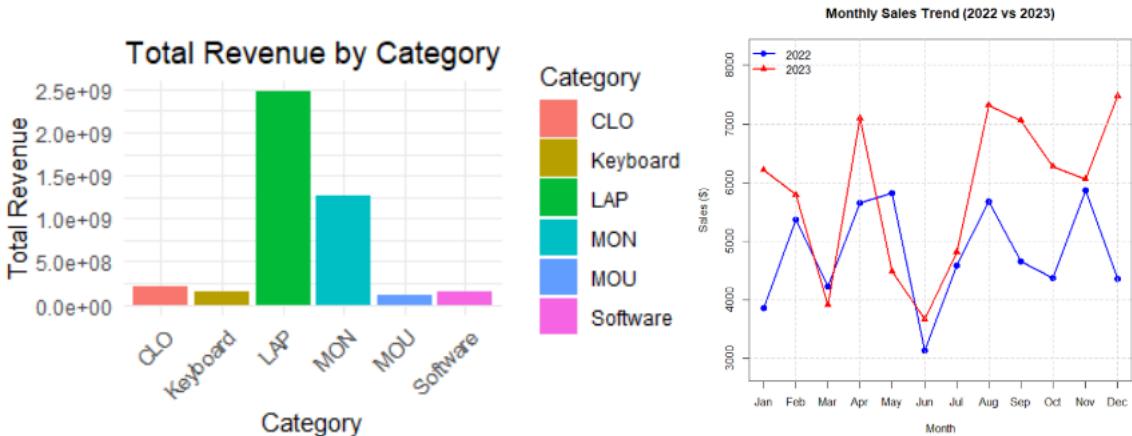
- Data quality improved drastically from almost unusable to excellent.
- All prefixes and categories have been corrected and now match.
- Price and markup differences have been eliminated.
- Forecasting and product analysis is now a viable option.

Sales Performance Analysis

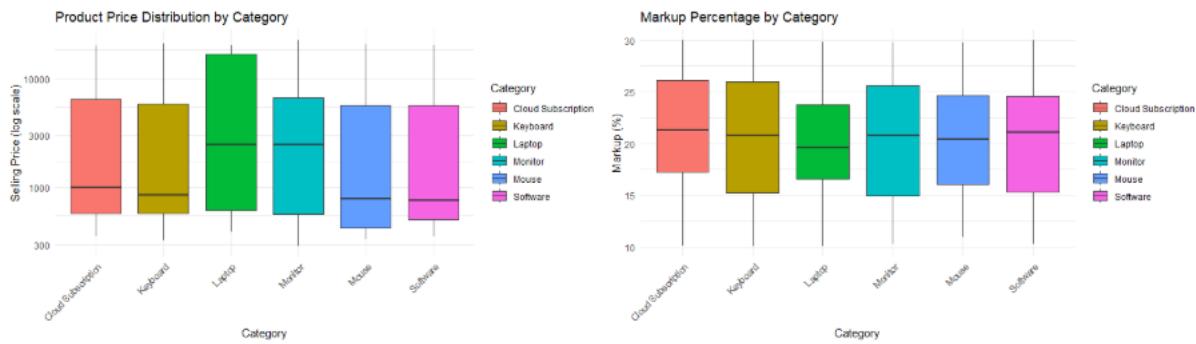
Revenue Insights

- New data analysis reveals that sales have declined not only on the yearly timeline, but also on the monthly, this drop in sales needs to be assessed to determine why it has occurred and what can be done to reverse this decline.
- Previous data analysis revealed that the highest selling product (with respect to quantity) is the “aliceblue marble keyboard”. It is clear now that although it is the most sold product, it contributes significantly less to profitability due to its low revenue generated. Analysis shows that the biggest revenue contributors are Laptops and Monitors, the reason for the drastic increase in revenue generated by these two products is due to their high-mean-pricing, which is considerably higher than all the other products price means.

- The mean markup percentage is relatively the same across each product line, this means that it would be best to focus on selling high priced products as the markup on them would be better than the low selling products.



Product Characteristics



- Data has been compiled and analysed, the top 10 products characteristics across quantity and selling price have been highlighted. It is clear that the highest selling category of products is “mouse” and “software”, however the highest selling priced product is laptops.
- With respect to previously determined insights, although the “mouse” and “software” products sell the most, they do not contribute significantly to revenue generated and should therefore either have their price increased or they should be removed from the product catalogue so that the company can focus their attention on the heavy revenue contributors. This will save time and money.

A more insightful summary for each product is provided by [Appendix A](#)

Clustering

Clustering Results

Clustering was performed on many parameters, and the following data has been collected.

- Clustering reveals that markup percentage vs price is pretty consistent between products.

- The three clusters show clear separation between low, mid, and high volume products, this confirms natural groupings in the data.
- Cluster 3 drives the majority of sales.
- Cluster 2 products exhibit very weak selling performance and should be considered for discontinuation, discounting, or “bundling strategies”.
- Cluster 1 products maintain moderate but steady sales, suggesting potential for marketing efforts or pricing adjustments to improve turnover.

Cluster Overview

- **Cluster 1: Medium Selling Products**

These products display stable demand across each month.

- **Cluster 2: Low Selling Products**

These products display consistently low monthly sales, these are often specialised items.

- **Cluster 3: High Selling Products**

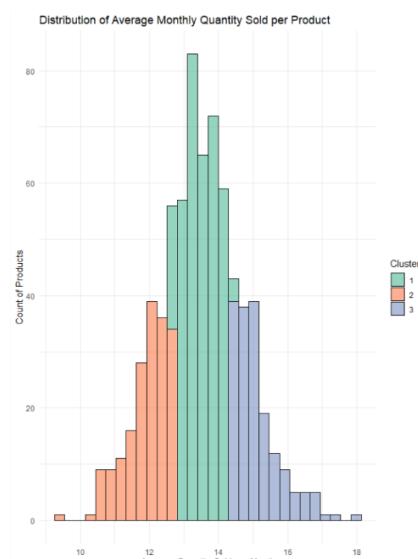
These products are in high demand and they contribute significantly to sales volumes and they should in theory contribute significantly to total revenue.

Cluster Visualisation Insights

The histogram below shows three distinct regions:

- **Cluster 1** spans a broader mid-range of monthly quantities.
- **Cluster 2** forms a concentrated group at the lower end of the distribution.
- **Cluster 3** stands apart as a small but high-performing group at the top end.

These visual distinctions confirm that the clustering captures genuine segmentation of the products.



Business Implications

- **Inventory Optimization:**

Prioritize restocking high selling products [cluster 3] while reducing low performing inventory [cluster 2].

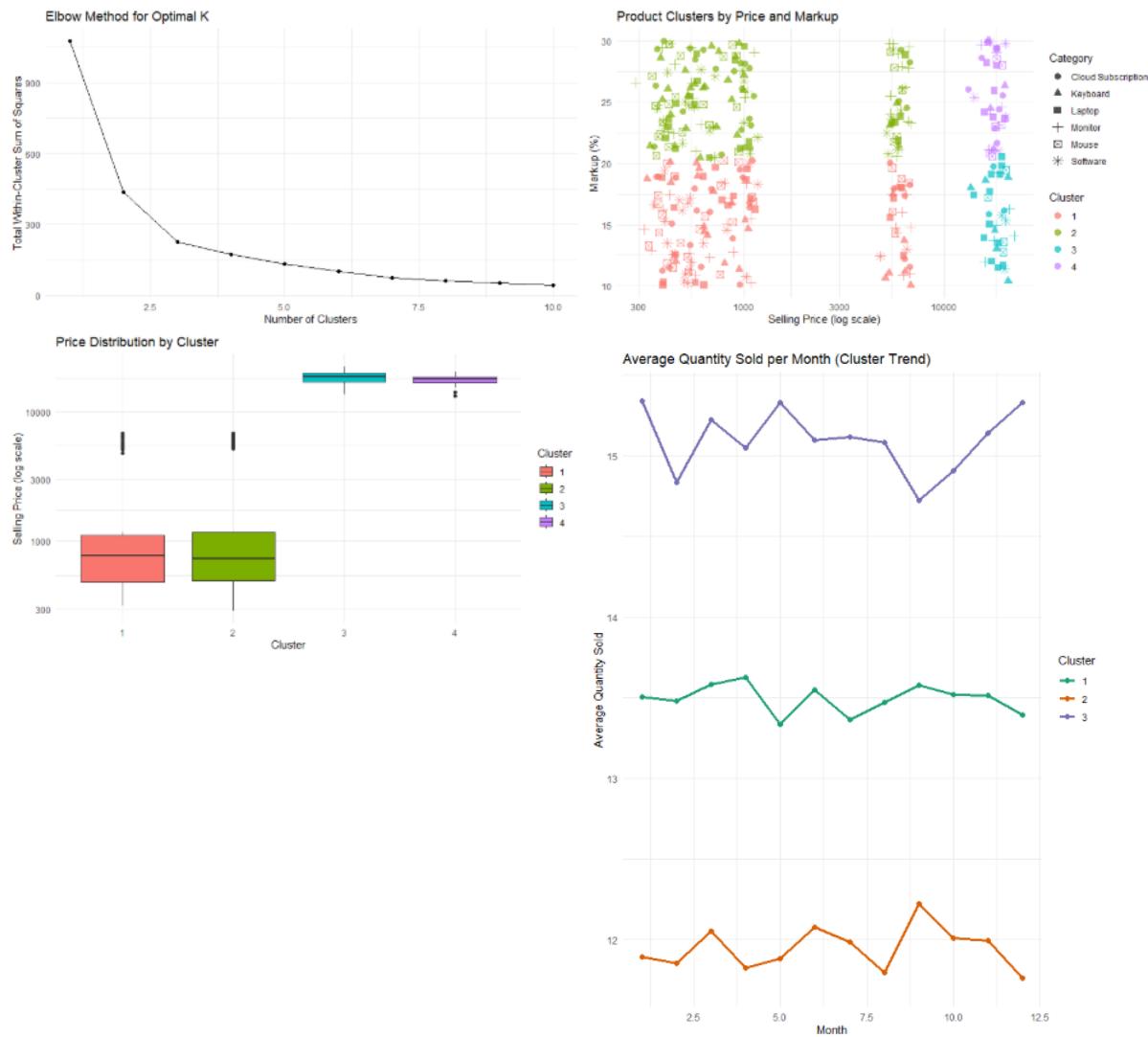
- **Marketing and Promotions:**

Target cluster 1 products with promotions or “bundle offers” to increase their demand.

- **Product Strategy:**

Review low selling [cluster 2] items for profitability. If unprofitable, consider phasing them out or integrating them into “bundles” with complementary products.

Additional information is provided in [Appendix B](#), [Appendix C](#), and [Appendix D](#)



Conclusion

The re-analysis confirmed that data integrity has significantly improved after fixing and replacing the corrupted product files with clean and consistent datasets. The findings now present a clearer picture of the company's performance in several key areas.

Revenue remains concentrated within a few high priced categories, mainly laptops and monitors, while high volume and low priced items like mice and software contribute very little to profit. This product imbalance indicates the need for a new sales strategy, one where the company streamlines or reprices low margin products and allocates greater marketing and inventory focus to "premium" lines.

K-means clustering was applied to the data to identify natural groupings amongst products based on their sales quantity, markup, and pricing characteristics. The clustering analysis yielded three distinct clusters, confirming natural segmentation within the product line:

- **Cluster 1: Medium Selling Products**

Products with moderate but steady sales that respond well to marketing or pricing adjustments.

- **Cluster 2: Low Selling Products**

Products with consistently low demand, often specialized or niche items that may require bundling or discontinuation.

- **Cluster 3: High Selling Products**

Top performing products that generate most sales and should remain the focus of inventory and promotional strategies.

The histogram visualisation of these clusters revealed three distinct sales regions, with cluster 3 clearly outperforming the others. Cluster 1 shows moderate dispersion and cluster 2 forms a concentrated low end group. This confirms that the clustering effectively represents true sales behavior and segmentation patterns across the company's product line.

From this analysis, the following recommendations are encouraged:

- **Inventory Optimization:**

Prioritize restocking of cluster 3 products while minimizing stock for cluster 2 items.

- **Marketing and Promotions:**

Target cluster 1 products with promotional or "bundling" campaigns to increase sales.

- **Product Strategy:**

Conduct profitability assessments for cluster 2 products to consider phasing out or repackaging low performers to boost overall revenue.

The incorporation of the corrected 2025 product datasets has transformed the analytical accuracy and reliability of results.

While sales volume remains stable, profitability insights now highlight where the company can refine its product mix, adjust pricing strategies, and optimize marketing focus for sustainable growth.

Future work should integrate temporal and customer-based features into clustering to further uncover patterns in seasonality, customer segments, and potential co-selling opportunities.

5. Staffing

This report analyses the effect of staffing on average service time, expected daily revenue, personnel costs, and gross profit. Analysis was performed with all parameters in mind via R-studio and the following insights and information was determined.

Assumptions:

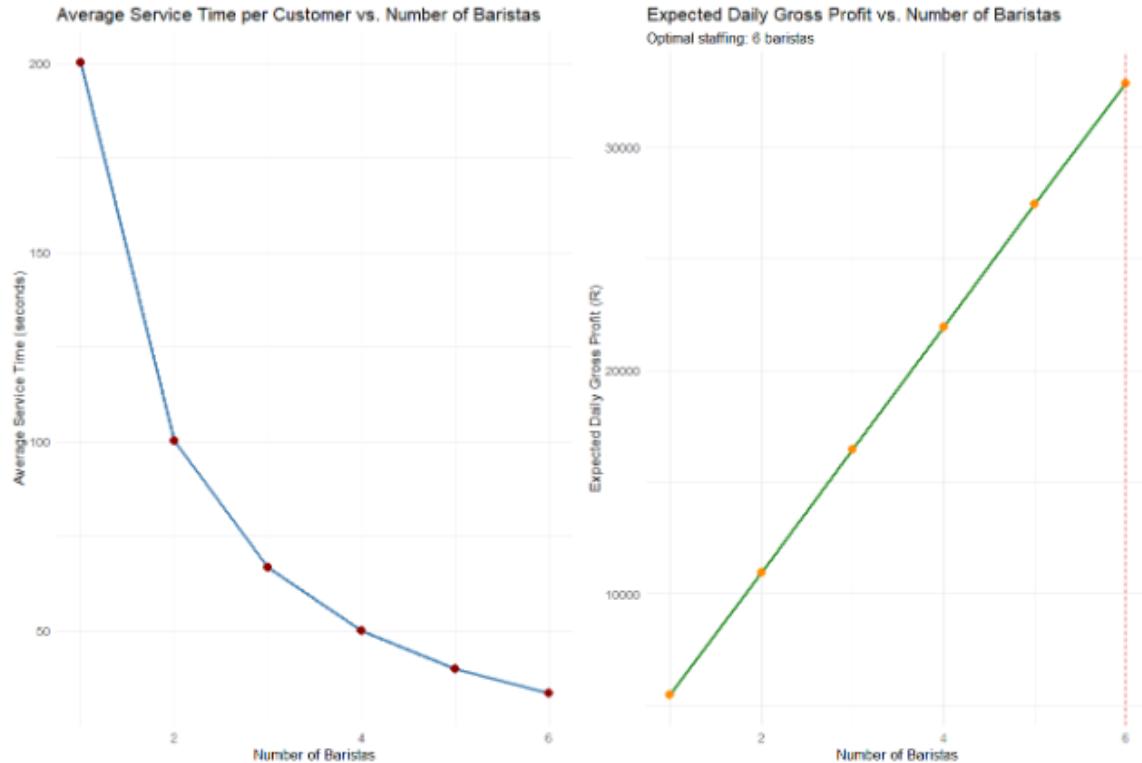
- Operating hours/day = 12
- Material profit/customer = R30.00
- Personnel cost/barista/day = R1000.00
- Reliable threshold = 120 sec.

Key Findings

- Optimal number of baristas (for max gross profit): 6
- Expected daily gross profit at optimal number of baristas: R32853.99
- Expected daily revenue at optimal number of baristas: R38853.99
- Expected daily personnel cost at optimal number of baristas: R6000.00
- Average service time at optimal number of baristas: 33.36 seconds
- The average wait time decreases exponentially as number of baristas increases.
- There is a clear, linear trend depicting how as the number of barista's increase, the coffee shop's profit decreases.
- With two barista's present, you may expect the highest average waiting time, followed by an incredibly sharp decrease as soon as you increase your number of barista's.

Summary

- Consider implementing a 6-person barista team to maximise profits.



Further summaries of insights are available in [Appendix E](#), [Appendix F](#) and [Appendix G](#)

6 Analysis of Service Time by Number of Baristas

Introduction

The timeToServe.csv data file was analysed using a one-way ANOVA, where service time (in seconds) was treated as the dependent variable and number of baristas as the factor. Additionally, Levene's test was used to verify homogeneity of variances, and residuals were checked for normality using the Shapiro–Wilk test (as well as diagnostic plots).

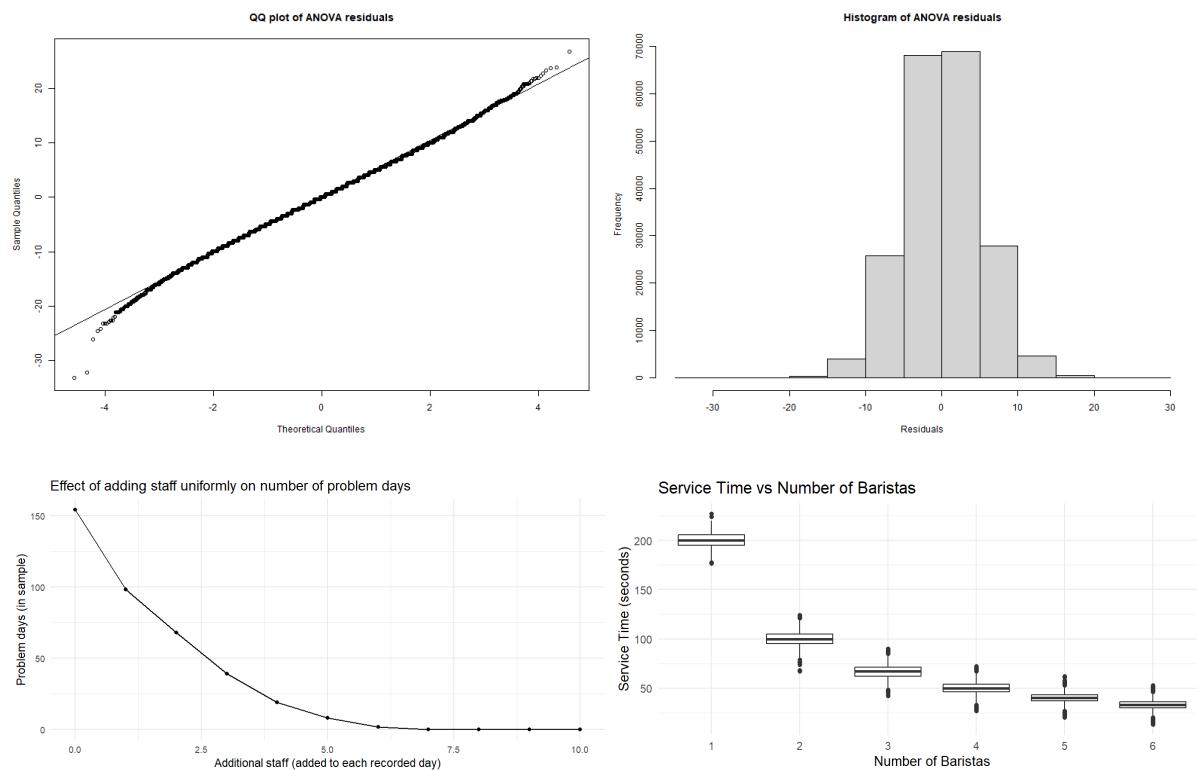
Results

The ANOVA showed a large effect of number of baristas vs service time, the following was determined: $F(5, 200000) = 310,453, p < 0.001$

This indicates that service time differs significantly between barista staffing levels. Levene's test [$F = 881.36, p < 0.001$] revealed variance among groups, this suggests that the homogeneity assumption is violated. Given the large, balanced dataset, the ANOVA remains robust to this violation. A few other tests were attempted but were unable to be completed due to the large size of the dataset.

Interpretation

The results confirm that staffing level has a huge influence on average service time. By increasing the number of baristas, the company can greatly reduce customer waiting time, this will overall improve throughput rates and daily profit. Although variance across groups differs, the findings remain true (given the sample size and test robustness).



7 Reliability and Staffing Optimization in Car-Rental Operations

Introduction

Daily staffing levels were simulated, and the results were analysed over nearly 400 days. Days that consisted of fewer than 15 staff members were classified as “unreliable”. Reliability was calculated as the proportion of reliable days. A cost model was created which combined two parameters:

- Loss per unreliable day: R20,000
- Personnel cost per staff member (per month): R25,000

The simulation tested the impact of adding between 0 and 10 extra staff members uniformly to every day. The model calculated the number of problem days, the expected annual losses, additional personnel costs and overall net effect.

Results

Out of 397 days, 243 days (61.2%) were reliable. This data corresponds to around 223 reliable days per year under the current staffing numbers. The optimization analysis showed that adding staff steadily reduced the number of unreliable days but also reduced financial return. The optimal point occurred at approximately 4 additional staff members, around this point expected annual losses decreased to R349,370 while total personnel cost increased to R1.2 million. This resulted in the least overall annual loss of approximately R1.55 million. Adding more than four staff members produced higher total costs and reduced profitability.

Interpretation

Operational reliability is currently below optimal levels, with only about 61% of days meeting reliability targets. The cost benefit model shows that by increasing staffing by around four employees per day, the company will experience the best balance between reliability improvement and additional labour expense. Increasing staff beyond this point does not provide proportional reliability gains and instead it will lead to higher overall costs.

Conclusion

The analyses in questions 6 and 7 demonstrate that staffing levels greatly effects both service quality and operational reliability. For question 6 (café), service times decrease significantly with additional baristas, improving efficiency and profitability. For question 7(car rental), reliability improves with added staff up to an optimal threshold of around 4 employees, beyond which returns diminish. These findings highlight the importance of balancing staffing costs against service performance metrics to achieve overall process stability and profitability.

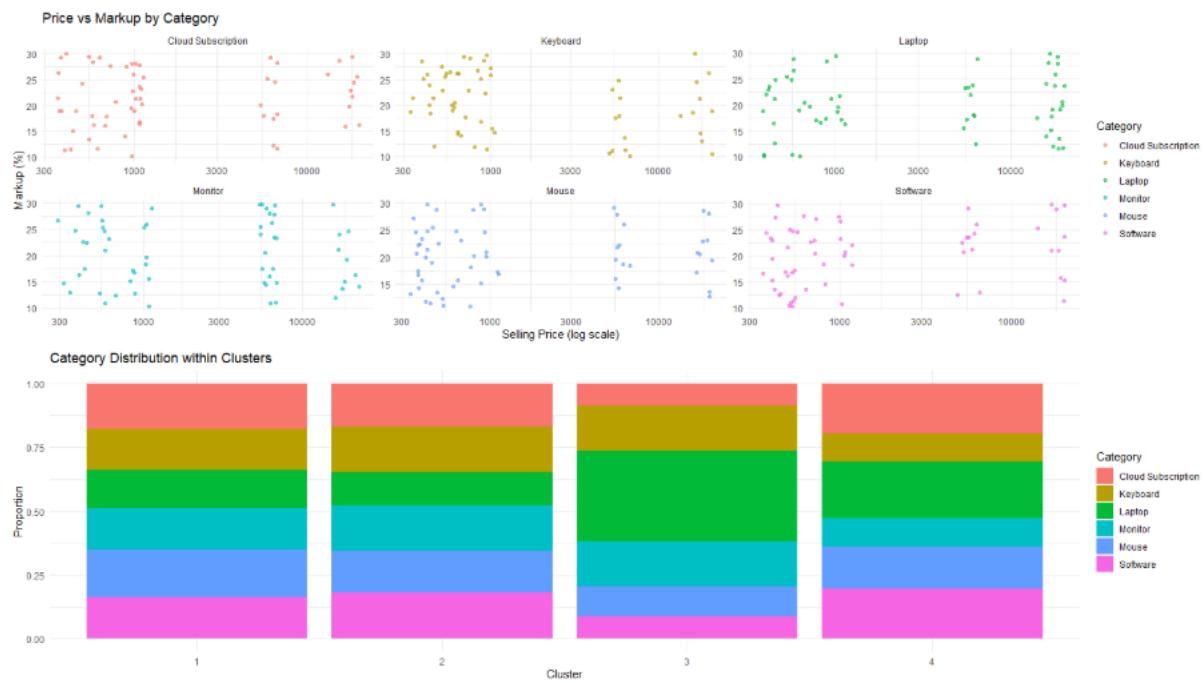
Appendix

Appendix A: Product Sale Details

ProductID	Total_Quantity_Sold	Avg_SellingPrice_HO
CLO011	21478	1070.54
CLO012	20110	963.14
CLO013	20550	1067.54
CLO014	21387	1083.11
CLO015	20834	728.26
CLO016	21676	959.51
CLO017	20934	991.81
CLO018	22094	1105.66
CLO019	20067	1092.07
CLO020	20877	1128.98
KEY041	26015	530.51
KEY042	24265	607.41
KEY043	24012	516.41
KEY044	22817	627.92
KEY045	24112	662.16
KEY046	24186	708.18
KEY047	23402	693.24
KEY048	24415	512.4
KEY049	23561	752.75
KEY050	24140	835.62
LAP021	12853	19494.91
LAP022	14058	16644.21
LAP023	13635	19452.72
LAP024	13952	18366.92
LAP025	14284	19725.18
LAP026	12892	18711.72
LAP027	14767	17202.28
LAP028	12989	18554.28
LAP029	13266	15851.74
LAP030	14025	16860.33
MON031	19078	6806.08
MON032	19463	6634.13
MON033	19090	5572.82
MON034	21962	6191.14
MON035	21059	6396.18
MON036	19796	6192.01
MON037	19141	6777.62
MON038	20231	6478.1
MON039	19344	6711.03

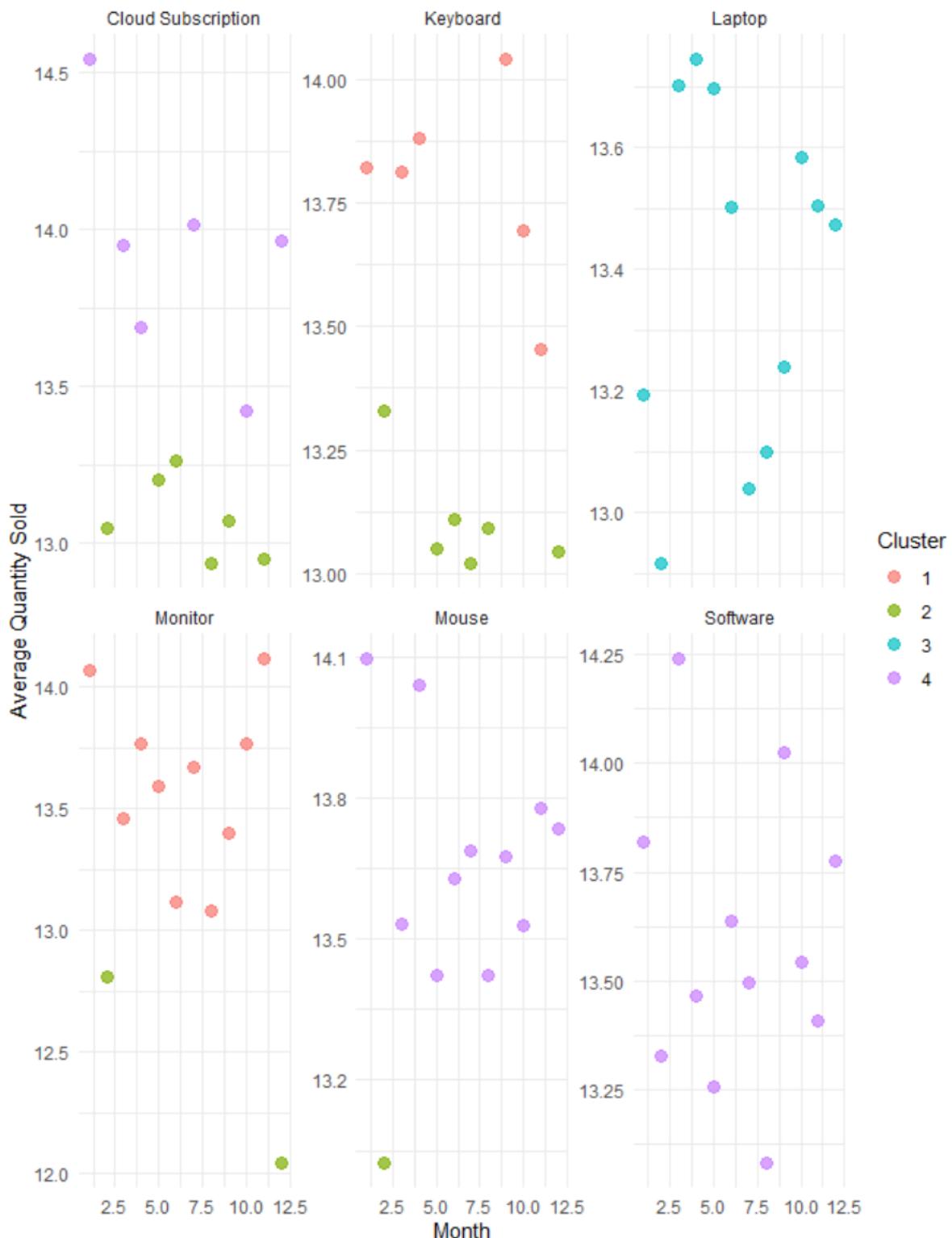
MON040	20527	5346.14
MOU051	26315	375.59
MOU052	28804	425.14
MOU053	27191	424.79
MOU054	28875	417.4
MOU055	27259	366.7
MOU056	28119	364.75
MOU057	28423	394.3
MOU058	28924	373.82
MOU059	29675	454.04
MOU060	27715	350.45
SOF001	29336	511.53
SOF002	27352	505.26
SOF003	27510	493.69
SOF004	29219	542.56
SOF005	28412	516.15
SOF006	27103	478.93
SOF007	28517	527.56
SOF008	27467	549.02
SOF009	27619	540.41
SOF010	29168	396.72

Appendix B: Price vs Markup Clustering Results

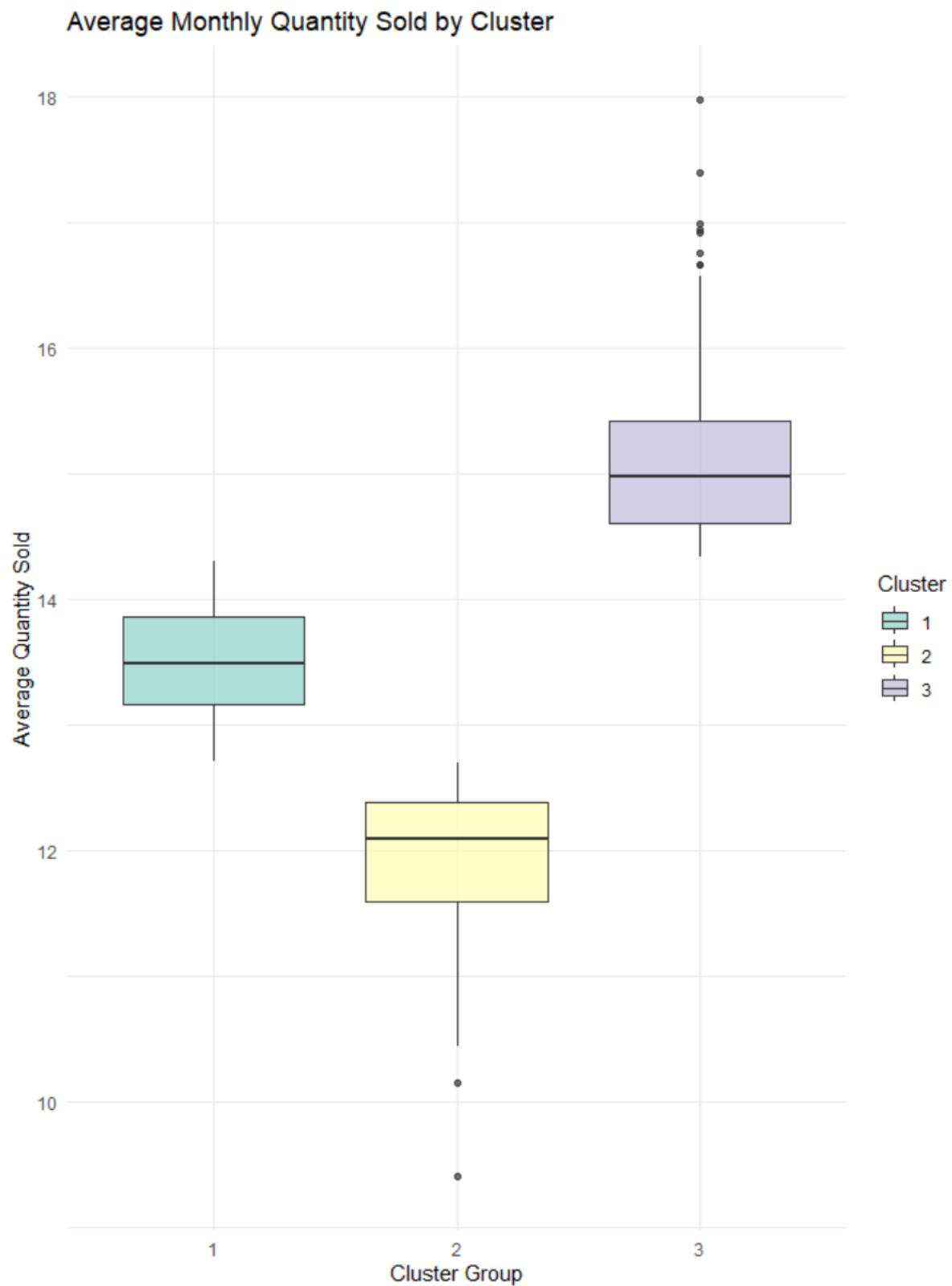


Appendix C: Avg Quantity and Product Clustering Results

Customer- and Time-Enhanced K-means Clustering



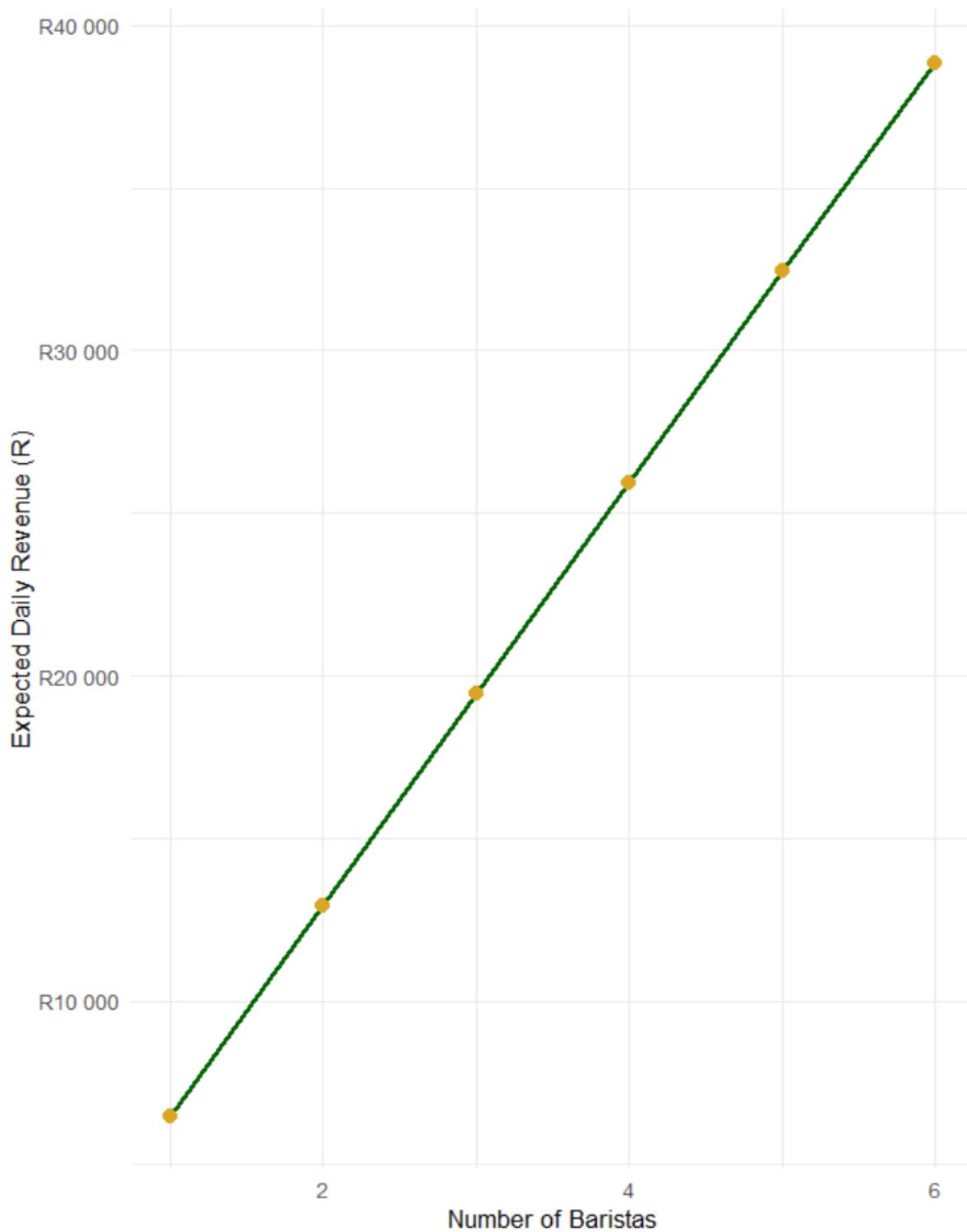
Appendix D: Monthly Quantity Sold Clustering Results



Appendix E: Expected Daily Revenue

Expected Daily Revenue vs. Number of Baristas

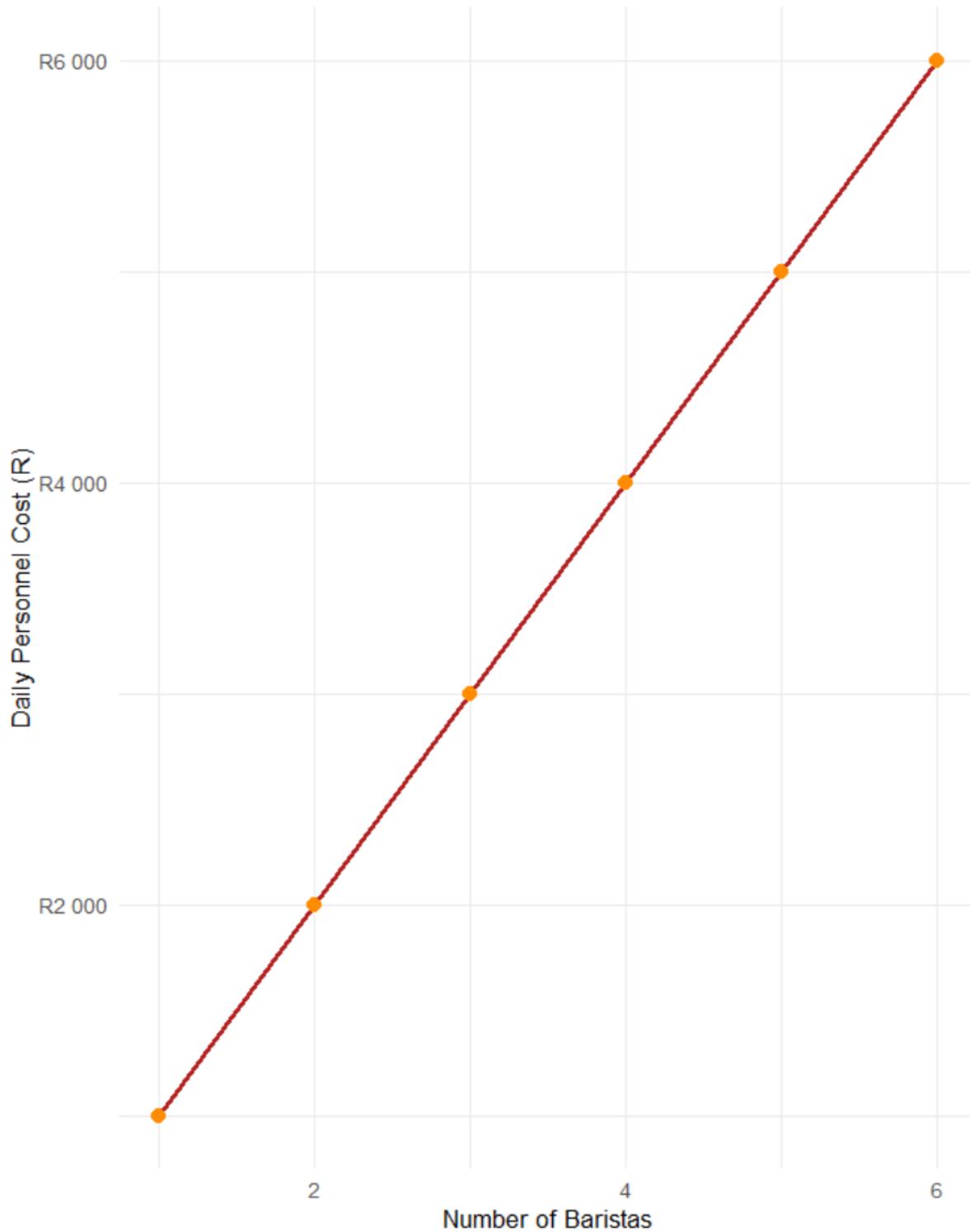
Revenue increases with faster service but plateaus at high staffing



Appendix F: Expected Daily Costs

Personnel Costs vs. Number of Baristas

Linear growth — each barista adds R1000/day



Appendix G: Data Summary

Baristas	n_obs	Mean service (s)	SD (s)	% Reliable	Exp customers/day	Exp revenue (R)	Personnel cost (R)	Gross profit (R)
1	417	200.16	8.02	0.00%	215.8	R6474.95	R1000.00	R5474.95
2	3556	100.17	7.10	99.72%	431.3	R12937.88	R2000.00	R10937.88
3	12126	66.61	6.27	100.00%	648.5	R19456.03	R3000.00	R16456.03
4	29305	49.98	5.53	100.00%	864.3	R25930.18	R4000.00	R21930.18
5	56701	39.96	4.99	100.00%	1081.0	R32430.94	R5000.00	R27430.94
6	97895	33.36	4.57	100.00%	1295.1	R38853.99	R6000.00	R32853.99