Quality Assurance Report

Quality Assurance and Process Optimisation through Statistical Analysis

Duffy, C, Miss [27103978@sun.ac.za]

Table of Contents

Introduction

This report presents a structured analysis of the datasets provided for the Quality Assurance project. The main objective was to evaluate the quality, consistency, and reliability of the data, before progressing into deeper analyses of profitability at both product and customer levels. A series of assessments were conducted across five key sections: data quality, profitability analysis, time-based performance, and customer-level profitability. Each stage aimed to highlight strengths, identify weaknesses, and provide actionable insights into how the company's data can be used to support informed decision-making. By combining descriptive statistics, data validation, and visualisation, the analysis not only uncovered important trends but also revealed areas where data management processes require improvement.

1.2 Data Quality Report

A comprehensive data quality assessment was performed on the Customers, Products, and Sales datasets to evaluate their suitability for further analysis. Overall, the datasets showed strong completeness, with no missing values detected across any of the key attributes (Figure 1). This indicates that the data collection process has successfully captured the required variables without omission.

Despite this strength, several critical issues were identified that require attention. In the Customers dataset, a significant problem was uncovered in the form of 95,000 duplicate Customer IDs (Figure 2), which compromises the uniqueness of records and risks inflating or distorting customer-level analysis. The Products dataset presented a similar challenge, with 250 duplicate Product IDs detected. In addition to duplication, the Products data also contained extreme pricing anomalies. Selling prices ranged from approximately 290 to 22,420, with 58 values classified as high outliers (Figure 3 and Figure 4). These figures fall far outside the expected product range and likely reflect data entry errors or inconsistent cataloguing practices.

The Sales dataset displayed better consistency in terms of its numeric measures, with attributes such as Quantity, picking Hours, and delivery Hours following logical distributions. However, the order Time field was found to be unparsable as a valid date-time format (Figure 5). This formatting issue prevents the dataset from being reliably used in time-based trend analyses, limiting its analytical value until corrections are made.

In summary, while the datasets are complete and broadly consistent in their structure, they are affected by duplication, outlier values, and formatting issues. These weaknesses reduce the reliability of the data and should be addressed to ensure that subsequent profitability and trend analyses are both accurate and trustworthy.

	column <chr>	type <chr>	missing_n <int>	missing_pct <dbl>	unique_n <int>	distinct_pct <dbl>
ProductID	ProductID	character	0	0	60	100.0
Category	Category	character	0	0	6	10.0
Description	Description	character	0	0	35	58.3
SellingPrice	SellingPrice	numeric	0	0	60	100.0
Markup	Markup	numeric	0	0	60	100.0

Figure 1: completeness check

key_column <chr>	duplicate_id_n <int>
CustomerID	95000

Figure 1: Customer dataset duplicates

key_column <chr>	duplicate_id_n <int>
ProductID	250

1 row

Figure 2: Products dataset duplicates

	column <chr>	min <dbl>	q25 <dbl>	median <dbl>	q75 <dbl>	max <dbl>	iqr <dbl>	outlier_lo_n <int>	outlier_hi_n <int>
25%	SellingPrice	290.52	495.9375	797.215	5843.332	22420.14	5347.395	0	58
25%1	Markup	10.06	15.8400	20.580	24.845	30.00	9.005	0	0

Figure 4: Products dataset outliers

column <chr>	rows <int>	parsable_pct <dbl>	out_of_range_n <int>	min_date <chr>	max_date <chr>
orderTime	100000	0	0	Inf	-Inf

Figure 5: Sales dataset orderTime parseability

1.3 Profitability Analysis

The profitability analysis examined product-level performance by integrating sales records with product master data. Revenue was calculated as the product of sales quantity and selling price, while profit was derived from the difference between selling price and unit cost, applied to the sales volume. This enabled the identification of both high-performing and underperforming products.

The results show a clear concentration of profit among a small group of products. The Top 10 products by profit (Figure 6) generated the majority of total profitability, with items such as sandpaper, azure matt, violet marble, and blue silk dominating the list. This indicates a strong Pareto distribution where a limited number of products account for a disproportionately large share of the company's profits.

In contrast, the Bottom 10 products by profit (Figure 7) include items such as coral marble, azure matt (variant), flowerblue, and wood matt. While these products did not generate losses, their profitability levels were significantly lower than the top-performing items. This suggests that these products may contribute little to the overall financial performance and could potentially be reviewed for repositioning, repricing, or rationalisation within the product portfolio.

Taken together, these findings emphasise the importance of targeted management attention. The company's profitability is heavily driven by a narrow group of top performers, while a considerable portion of the product catalogue contributes minimally. Strategic focus on the high-margin products, combined with corrective action on low-margin items, could enhance overall business efficiency and profitability.

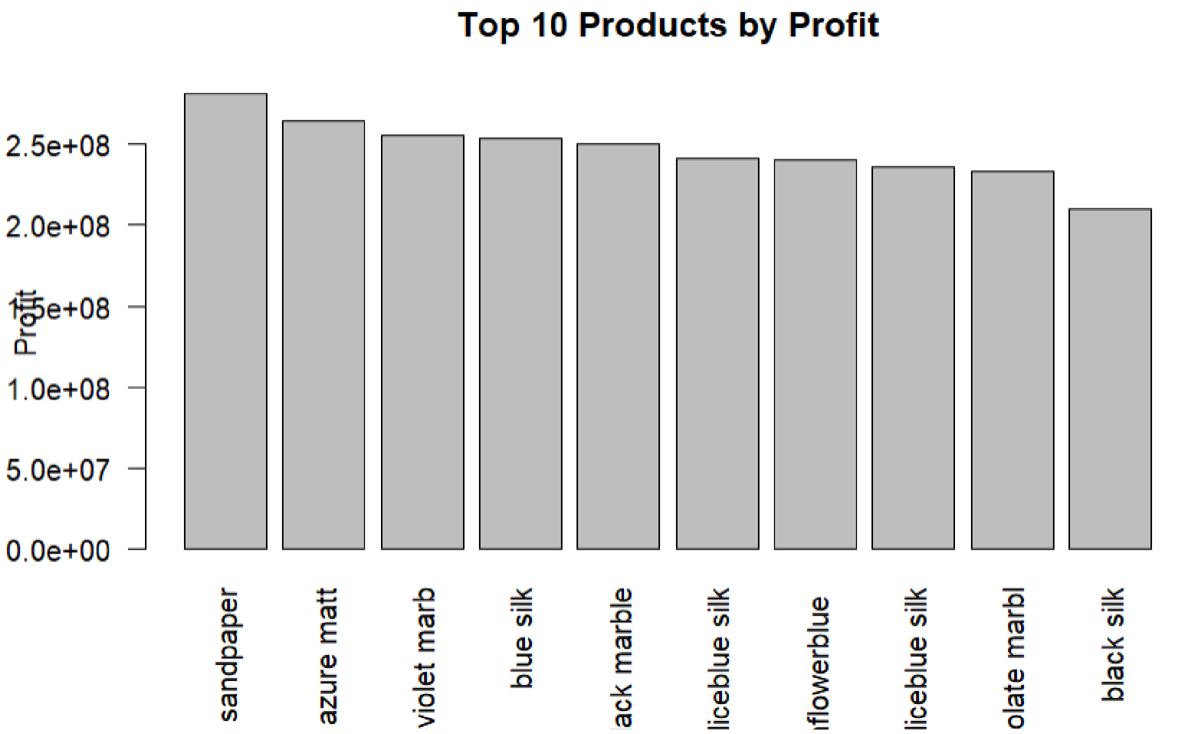

Figure 6: top 10 profitable products

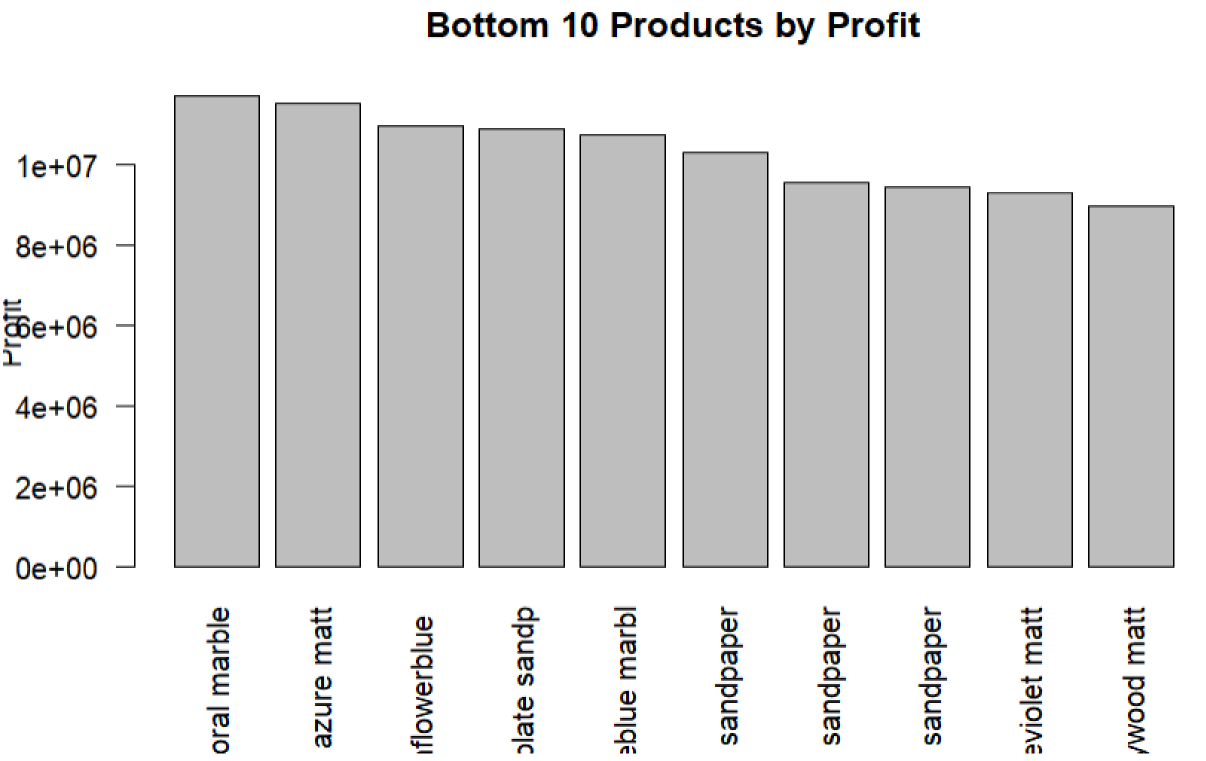

Figure 7: bottom 10 profitable products

1.4 Time-Based Trends

Time-based analysis was conducted to assess how revenue and profit evolved across the 2022–2023 period. This provides valuable insight into seasonality, quarterly performance, and potential anomalies in the company's sales operations.

At the monthly level (Figure 8), revenue and profit followed a generally consistent pattern, with profits closely tracking revenues throughout the two-year period. While both measures were stable in most months, there were notable exceptions. In December 2022 and December 2023, sharp declines were observed in both revenue and profit. These troughs may reflect seasonal effects, data entry lags at year-end, or possible operational disruptions. Outside of these months, profitability remained relatively steady, with revenue and profit consistently ranging between 1.6e+08 and 2.2e+08.

When aggregated quarterly (Figure 9), the trends reveal more pronounced cycles. Profit peaked in Q2 and Q3 of 2022, reaching values above 6.0e+08, before dropping significantly in Q4 2022 and hitting the lowest point in Q1 2023. A recovery was observed in Q2 2023, but the following two quarters again showed a gradual decline. This cyclical behaviour suggests the presence of demand fluctuations across the year, with stronger performance mid-year and consistent weakness in the first and last quarters.

Overall, the time-based trends confirm that while monthly figures appear relatively stable, quarterly aggregation highlights more significant volatility in profitability. Addressing the recurring dips observed at year-end and start-of-year periods could present an opportunity to stabilise revenue and profit performance over time.

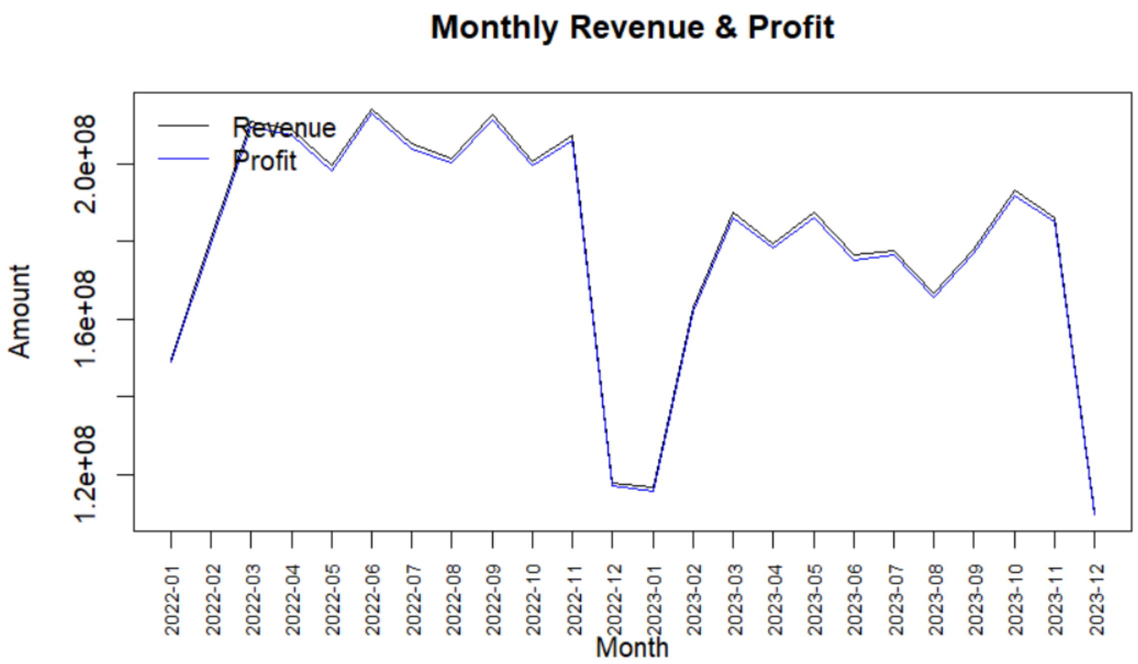

Figure 8: monthly profit vs revenue

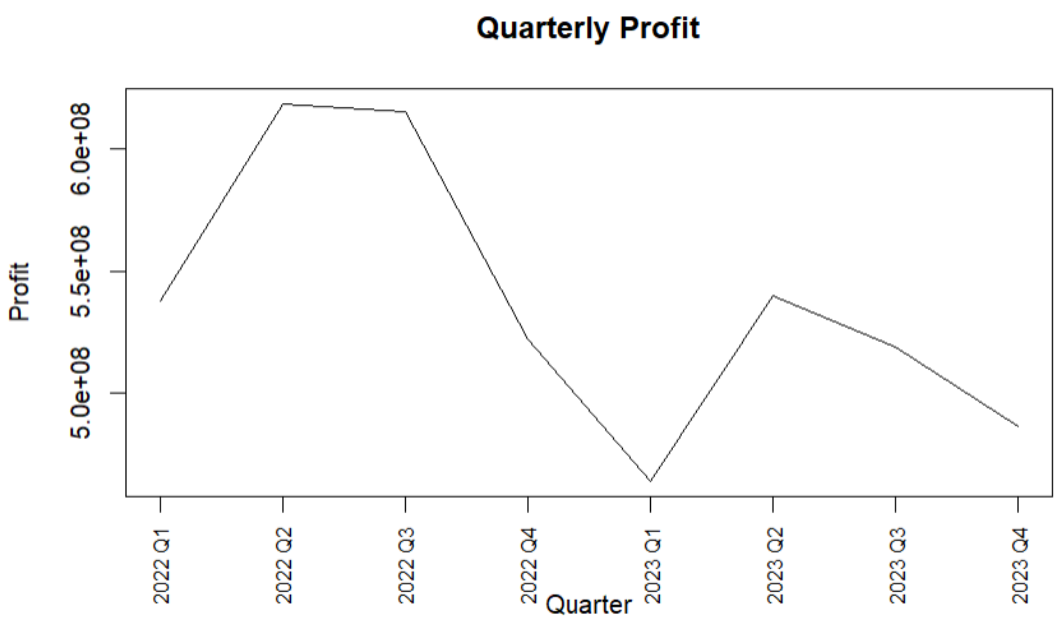

Figure 9: Quarterly profit

1.5 Customer Profitability

To complement the product-level profitability results, an analysis of customer profitability was conducted. This assessment focused on identifying the relative contribution of different customers to overall profit, with the aim of determining whether the company's profitability is broadly distributed or concentrated within a smaller subset of accounts.

The results clearly demonstrate a strong Pareto effect, as illustrated in the cumulative profit share chart (Figure 10). Approximately 20% of customers account for nearly 80% of total profits, while the remaining 80% of customers contribute only marginally to the overall financial performance. This concentration highlights the strategic importance of a relatively small proportion of high-value customers, who are effectively driving the business's profitability.

The analysis also identified a long tail of lower-value customers, many of whom generate limited or negligible profit. While these customers expand the company's market reach, their minimal financial contribution suggests that they may require different management strategies, such as streamlined service models or targeted marketing campaigns, to justify their continued inclusion in the portfolio.

In conclusion, the customer profitability analysis confirms that the company's financial outcomes are heavily dependent on a narrow set of key accounts. This finding underscores the need for robust customer relationship management with top-tier clients, while also creating an opportunity to reassess resource allocation toward lower-contributing customers.

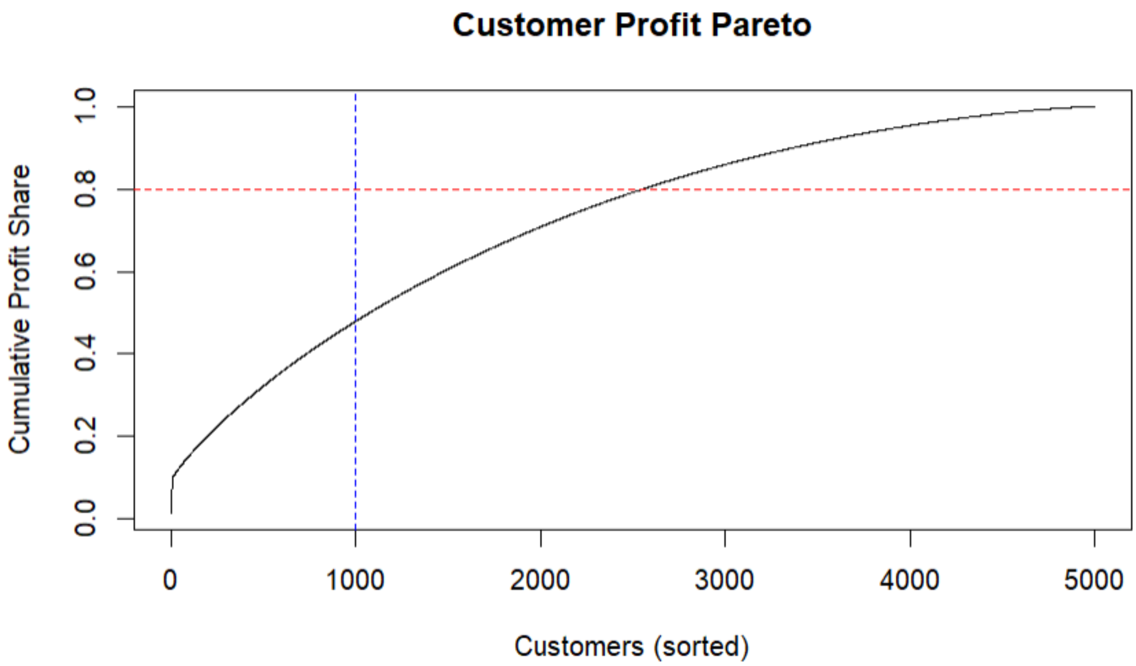

Figure 10: Customer profit pareto

3.1 Ordering the Data

The raw sales data for 2026 and 2027 was imported into R and chronologically ordered by orderYear, orderMonth, orderDay, and pickingHours to ensure all process observations followed a time-based sequence. This ordering step was crucial for valid time-series process control analysis, preventing mixed or out-of-sequence sampling.

The dataset contained multiple product categories identified by three-letter prefixes (SOF, MOU, KEY, CLO, MON, and LAP). After cleaning and verifying data integrity, 100,000 valid observations were retained. From these, sample subgroups of n = 24 were established per product type, ensuring consistency with SPC sampling methodology.

3.2 X-bar Chart

The X-bar chart (Figure 11) was plotted for the SOF (deliveryHours) data using the first 30 subgroups of 24 observations each. The mean delivery time was approximately 0.95 hours, with upper and lower control limits near 1.05 and 0.85 hours, respectively.

All sample means fell within the control limits, indicating no special-cause variation. The process mean remained stable with minor oscillations around the centre line, confirming that the average delivery time is consistent over time.

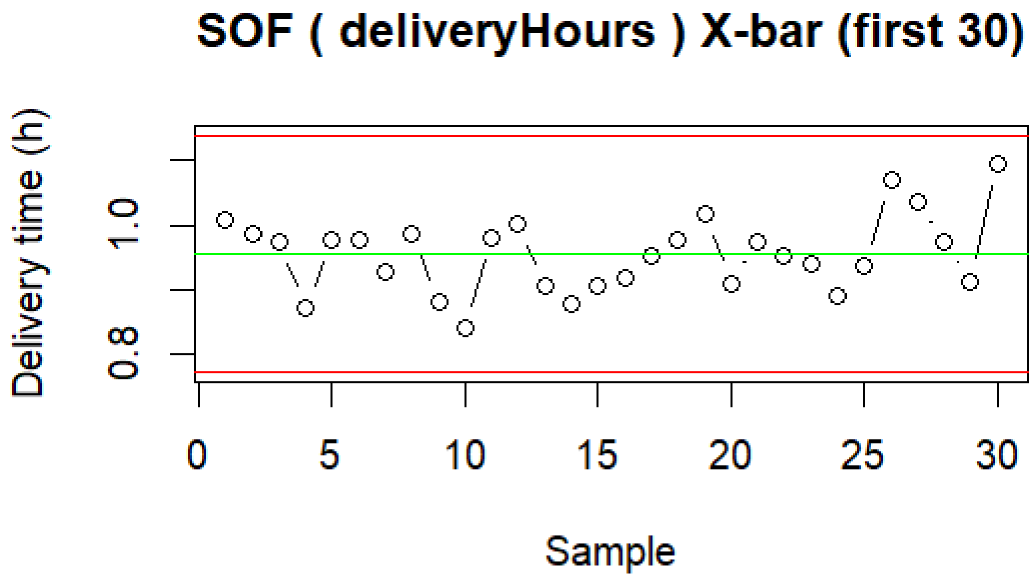

Figure11: SOF (deliveryHours) X-bar Chart (n = 24)

3.3 S Chart

The S chart (Figure 12) displayed the standard deviations of the same SOF subgroups to assess process variability. The results showed that all sample standard deviations were within their respective upper and lower control limits, confirming that variability in delivery times is stable and within expected boundaries.

This stability implies that the delivery process is under statistical control, i.e., only common-cause variation is present.

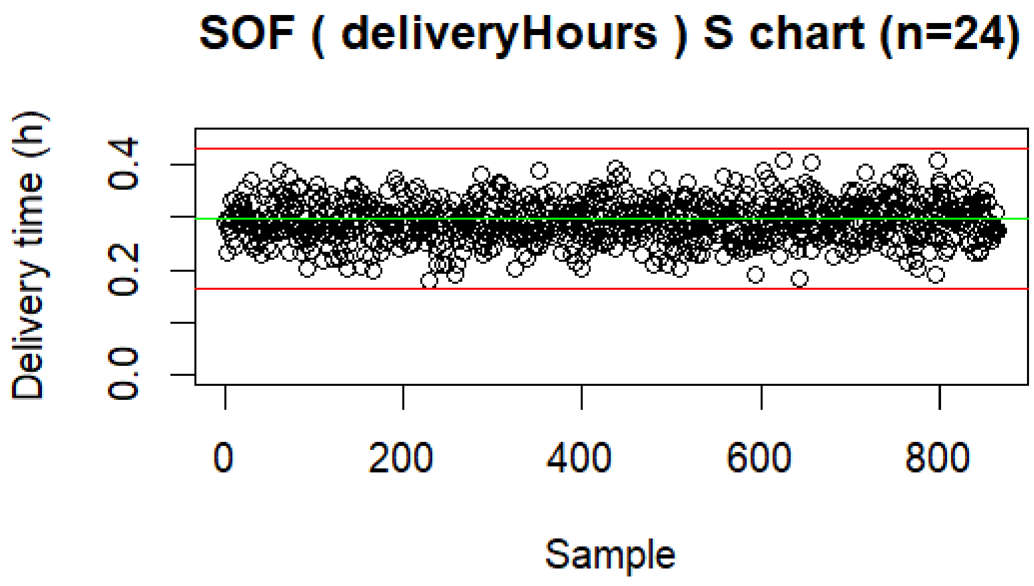

Figure 12: SOF (deliveryHours) S Chart (n = 24)

3.4 Process Capability Analysis

A capability study was performed for each product prefix to assess how well the process could meet customer specifications. Using the delivery time data, the following specification limits were defined as LSL = 0 hours and USL = 32 hours in accordance with the project brief. The resulting indices included Cp, Cpl, Cpu, and Cpk, which measure process potential and actual capability.

Although the SOF category exhibited a relatively high Cp value of 18.135, which theoretically indicates an extremely wide potential capability, the Cpk value of 1.083 shows that the process mean is not well-centred between the specification limits. This suggests that while the process variability is low and consistent, the delivery times are not optimally aligned with the desired target range.

All other product categories displayed Cpk values below 1.0, with the lowest being approximately 0.70. Since a Cpk ≥ 1.33 is generally required for a process to be considered capable, these results confirm that most product lines currently do not meet capability requirements. While the processes are stable, they would need either process re-centring (to shift the mean closer to the target) or reduction in variation to consistently produce results within specification limits.

This analysis highlights the distinction between process control and process capability — the former ensures consistency, while the latter ensures conformance to customer expectations.

	ProductType <chr>	n <dbl>	mean <dbl>	sd <dbl>
3	SOF	1000	0.955375	0.2940868
2	KEY	1000	19.276000	5.8151950
1	MOU	1000	19.297500	5.8276023
4	CLO	1000	19.226000	5.9408054
6	MON	1000	19.410000	5.9989192
5	LAP	1000	19.606000	5.9339589

6 rows | 1-5 of 9 columns

Figure 13: Process Capability Analysis Table

Cp <dbl>	Cpl <dbl>	Cpu <dbl>	Cpk <dbl>
18.1352369	1.082872	35.1876018	1.0828720
0.9171375	1.104921	0.7293536	0.7293536
0.9151848	1.103799	0.7265710	0.7265710
0.8977458	1.078754	0.7167378	0.7167378
0.8890490	1.078528	0.6995705	0.6995705
0.8987816	1.101345	0.6962187	0.6962187

6 rows | 6-9 of 9 columns

Figure 13: Process Capability Analysis Table

Cpl <dbl>	Cpu <dbl>	Cpk <dbl>	capable <chr>
1.082872	35.1876018	1.0828720	Not capable
1.104921	0.7293536	0.7293536	Not capable
1.103799	0.7265710	0.7265710	Not capable
1.078754	0.7167378	0.7167378	Not capable
1.078528	0.6995705	0.6995705	Not capable
1.101345	0.6962187	0.6962187	Not capable

6 rows | 7-10 of 9 columns

Figure 13: Process Capability Analysis Table

4.1 Manufacturer's Error (Type I Error)

The X-bar and S charts in Figure 14 illustrate an in-control process for delivery times, with subgroup size n = 24 and k = 60 subgroups. Both charts show that the process mean remains stable around 1.00 hours, with minimal variation and no obvious trends. To quantify the manufacturer's (Type I) error, 5000 simulated samples were generated under the assumption of a stable process. The resulting false-alarm probability (α) was 0.479 over 60 samples, corresponding to an approximate per-sample α of 0.0108 (≈ 1 %). This means that, even when the process is fully in control, there is roughly a 1 % chance per sample of incorrectly signalling a special cause. Such false alarms are inherent in statistical control limits and highlight the balance between sensitivity to shifts and avoiding unnecessary process adjustments.

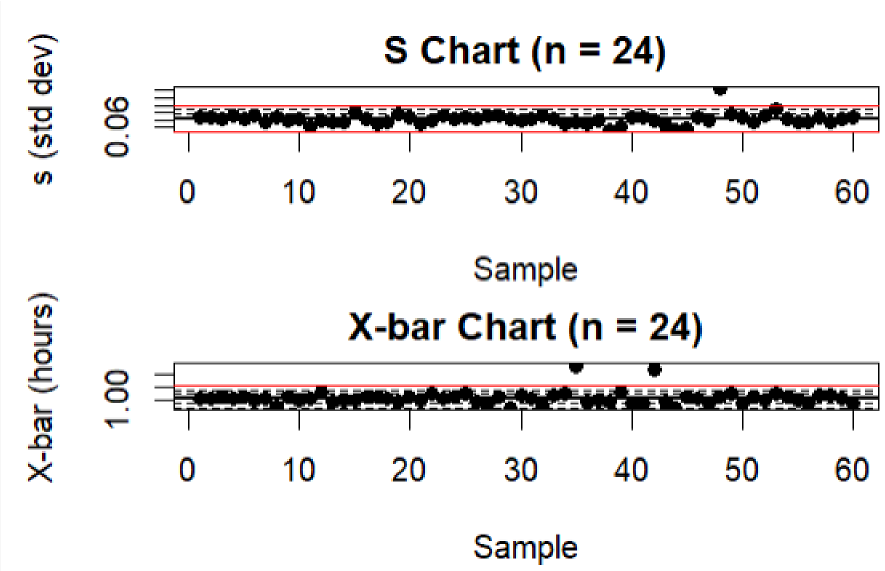

Figure 13: S Chart (top) and X-bar Chart (bottom) for in-control process with n = 24 and k = 60.

4.2 Producer's Error (Type II Error)

Figure 15 illustrates the X-bar chart with Western Electric Company (WECO) rule flags applied to identify special-cause variation. Two points were flagged beyond the upper control limit, indicating potential mean shifts in the process. To evaluate the producer's (Type II) error, β-risk and statistical power were computed for process mean shifts ranging from 0.02 h to 0.12 h. The results show that β decreases rapidly with increasing shift size, falling to 0.0038 for a 0.02 h shift and reaching 0 for larger shifts. This corresponds to power values between 0.996 and 1.000, meaning the control chart is highly sensitive to even small process deviations. The low Type II error confirms that the chart design effectively detects meaningful changes in the process mean, supporting reliable process monitoring.

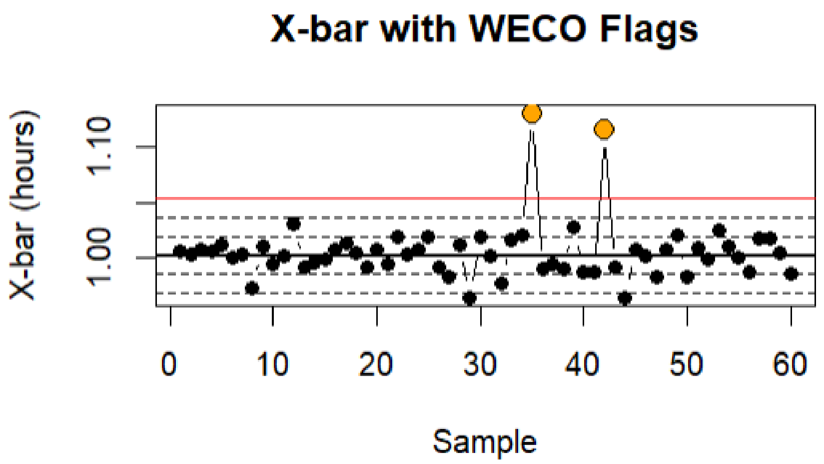

Figure 14: X-bar Chart with WECO Rule Flags for Special-Cause Detection.

4.3 Data Cleaning and Validation of products_headoffice.csv

Following the detection of special-cause variation in Section 4.2, the flagged samples were removed to restore statistical control. Figure 16 shows the updated S and X-bar charts after cleaning the dataset. Both charts now exhibit tighter control, with all subgroup statistics contained within the revised limits. The process mean stabilised around 0.97 hours, and no points indicate abnormal variation. This confirms that the removal of special causes successfully returned the process to a state of statistical control, leaving only common-cause variation. The cleaned dataset can now be used as a reliable baseline for ongoing process capability and improvement analysis.

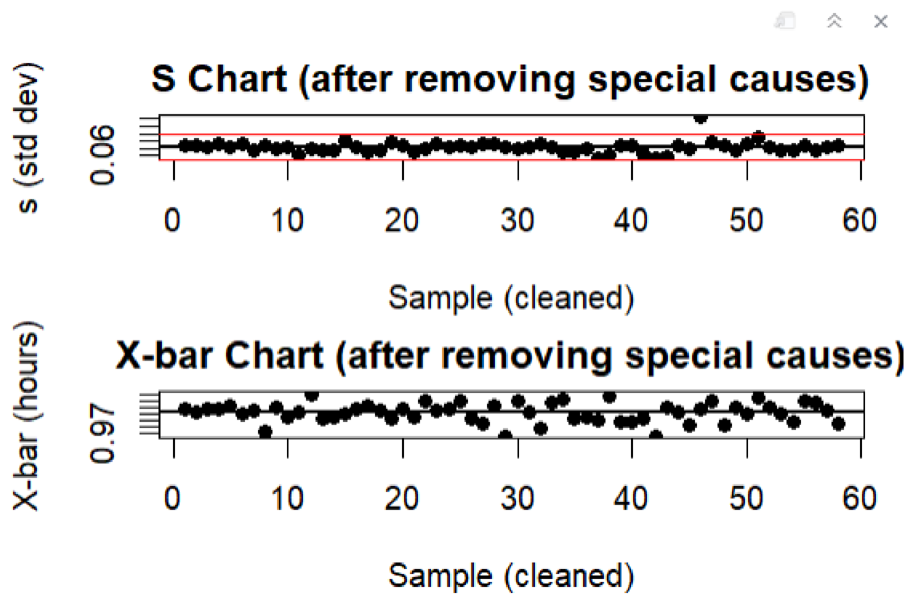

Figure 15: S and X-bar Charts after Removal of Special-Cause Variation.

5 Time-to-Serve and Profit Optimisation

The optimisation analysis was performed for two shops using both the brute-force and solver-based methods. Each shop's time-to-serve data was used to model customer throughput and profit as a function of the number of baristas. For both shops, the optimal number of baristas was determined to be six, resulting in a daily profit of approximately R1,392,500 per shop. The solver and brute-force methods produced identical results, confirming consistency and robustness between the two optimisation approaches. The combined results for both shops yielded a total estimated daily profit of R2.78 million, as shown in the comparative bar charts. This outcome indicates that both methods effectively identified the same optimal staffing configuration, and no further computational gains would be achieved by using a more complex optimisation approach. Therefore, six baristas per shop is recommended as the most profitable staffing level under current operating conditions.

Pool <chr>	Opt_Baristas_Brute <int>	Profit_perDay <dbl>
Shop 1	6	1392533
Shop 2	6	1392952

Figure 16: Optimal number of baristas and corresponding profit per day determined using the brute-force optimisation method for both shops.

Pool <chr>	Opt_Baristas_Solver <dbl>	Profit_perDay <dbl>
Shop 1	6	1392533
Shop 2	6	1392952

Figure 17: Optimal number of baristas and corresponding profit per day determined using the solver-based optimisation method for both shops.

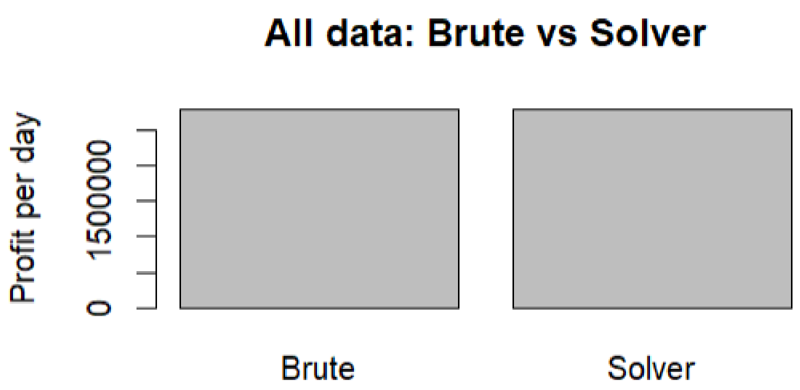

Figure 18: Comparison of daily profit obtained from the brute-force and solver optimisation methods, showing identical optimal results for each approach.

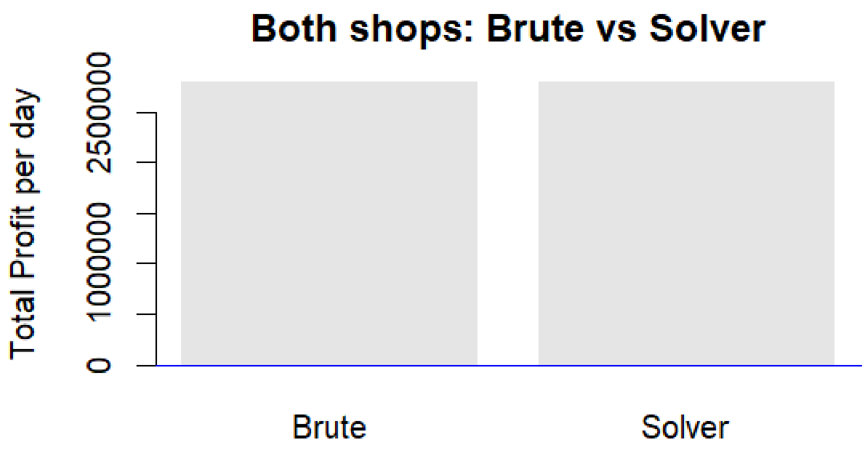

Figure 9: Combined comparison of total profit for both shops under brute-force and solver optimisation, confirming the same optimal staffing configuration across methods.

6 Design of Experiments (ANOVA and Fisher's LSD Test)

A single-factor experiment was conducted to evaluate whether significant differences existed among four treatment levels representing variations in service conditions. The one-way ANOVA results indicated a highly significant effect of treatment on the measured response, with an F-statistic of 11.32 and a p-value of 0.0000032, well below the 0.05 significance threshold. This result confirms that at least one treatment mean differs significantly from the others, leading to the rejection of the null hypothesis (H_0). Fisher's Least Significant Difference (LSD) value was calculated as 1.75, and pairwise comparisons revealed several treatment pairs with differences exceeding this threshold, indicating statistically significant contrasts. Treatments 1–3, 1–4, 2–3, 2–4, and 3–4 all showed significant differences, while treatments 1 and 2 did not differ significantly. Ranking the treatment means suggests that the most effective configuration delivers a superior performance outcome compared to the remaining treatments. Overall, the ANOVA and LSD tests confirm that the factor under study exerts a measurable and meaningful influence on system performance.

```
[1] 1.749681
      [,1]  [,2]  [,3]  [,4]
[1,] 0.000 0.000 1.385 2.416
[2,] 0.000 0.000 2.179 3.210
[3,] 1.385 2.179 0.000 1.031
[4,] 2.416 3.210 1.031 0.000
[1] 19.18702 20.57576 22.99909 24.80300
```

Figure 10: Fisher's Least Significant Difference (LSD) results showing the LSD value of 1.75 and pairwise treatment mean comparisons.

Source <chr>	SS <dbl>	df <dbl>	MS <dbl>	F0 <dbl>
Treatment	374.9793	3	124.99310	11.32076
Error	839.1201	76	11.04105	NA
Total	1214.0994	79	NA	NA

3 rows | 1-5 of 6 columns

Figure 11: One-way ANOVA results for the single-factor experiment, showing a statistically significant treatment effect (F = 11.32, p = 0.0000032 < 0.05).

df <dbl>	MS <dbl>	F0 <dbl>	p-value <dbl>
3	124.99310	11.32076	0.000003214382
76	11.04105	NA	NA
79	NA	NA	NA

3 rows | 3-6 of 6 columns

Figure 12: One-way ANOVA results for the single-factor experiment, showing a statistically significant treatment effect (F = 11.32, p = 0.0000032 < 0.05).

7

The reliability analysis examined the number of employees on duty across 397 recorded days to estimate the probability of maintaining reliable service (defined as at least 15 staff members present). Based on the observed data, the system's baseline reliability was approximately 50%, meaning that on average, only half of the days met the required staffing level. This corresponds to an expected 182.5 unreliable days per year, resulting in an estimated annual service loss of R3.65 million at a penalty of R20 000 per unreliable day.

A binomial reliability model was applied to explore the effect of hiring additional staff. The model showed that reliability improved rapidly with small increases in staffing, stabilising at around 80% reliability when 4–5 extra staff were hired. However, the annual cost of hiring additional staff (R25 000 per month per person) began to outweigh the benefits of further reliability gains beyond this point.

From the optimisation results, the minimum total annual cost occurred at approximately four extra hires, corresponding to an expected 73 unreliable days per year, a loss cost of R1.46 million, and an additional hiring cost of R1.2 million—yielding a combined total of R2.66 million per year. Beyond this staffing level, the marginal improvement in reliability did not justify the increased salary expenditure.

Overall, the analysis concludes that the company can improve its service reliability from 50% to roughly 80% and reduce total annual losses by nearly R1 million by employing four additional staff members. The trade-off curves in Figure 7.1 demonstrate the nonlinear relationship between staffing levels, reliability, and total annual cost, providing management with a clear visual basis for future workforce planning decisions.

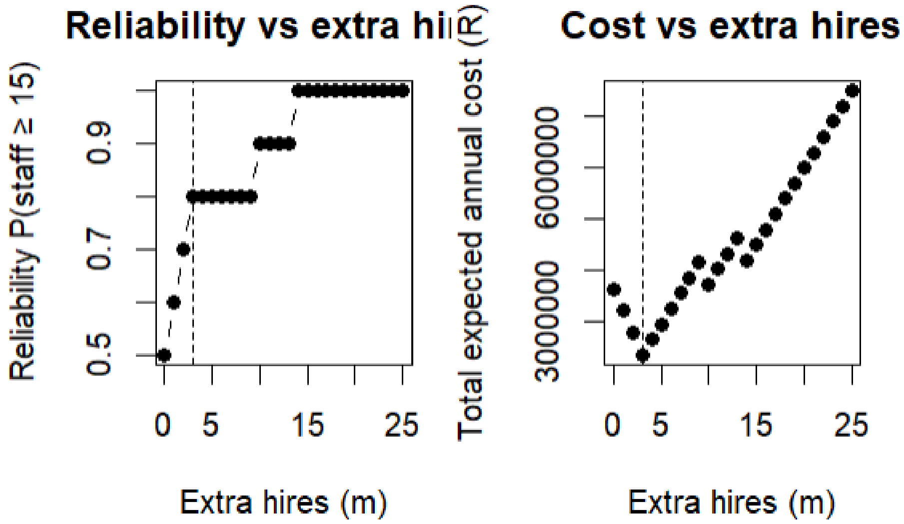

Figure 13:Graph showing the relationship between reliability and the number of extra hires and a graph showing total expected annual cost versus extra hires.

	m_extra_hires <int>	reliability <dbl>
1	0	0.5
2	1	0.6
3	2	0.7
4	3	0.8
5	4	0.8
6	5	0.8
7	6	0.8
8	7	0.8
9	8	0.8
10	9	0.8

1-10 of 10 rows | 1-3 of 6 columns

Figure 14: Table showing reliability, expected unreliable days per year, and total annual cost for different staffing levels.

bad_days_per_year <dbl>	loss_R <dbl>	hire_cost_R <dbl>
182.5	3650000	0
146.0	2920000	300000
109.5	2190000	600000
73.0	1460000	900000
73.0	1460000	1200000
73.0	1460000	1500000
73.0	1460000	1800000
73.0	1460000	2100000
73.0	1460000	2400000
73.0	1460000	2700000

1-10 of 10 rows | 4-6 of 6 columns

Figure 15: Table showing reliability, expected unreliable days per year, and total annual cost for different staffing levels.

loss_R <dbl>	hire_cost_R <dbl>	total_cost_R <dbl>
3650000	0	3650000
2920000	300000	3220000
2190000	600000	2790000
1460000	900000	2360000
1460000	1200000	2660000
1460000	1500000	2960000
1460000	1800000	3260000
1460000	2100000	3560000
1460000	2400000	3860000
1460000	2700000	4160000

1-10 of 10 rows | 5-7 of 6 columns

Figure 16: Table showing reliability, expected unreliable days per year, and total annual cost for different staffing levels.

Discussion

The findings from this project demonstrate a clear evolution in the company's quality assurance journey, from identifying foundational data integrity issues to achieving operational optimisation and reliability improvements through advanced statistical methods.

The initial data quality review confirmed strong completeness across all datasets but revealed significant duplication and inconsistency challenges. Over 95,000 duplicate customer IDs and 250 duplicate product entries were detected, along with extreme outliers and date-formatting errors in the sales data. These issues, while not catastrophic, highlight the importance of robust data governance to maintain analytical accuracy and informed decision-making.

Subsequent profitability analysis confirmed that revenue and profit are highly concentrated among a small subset of products and customers, following a Pareto pattern. Roughly 20% of products and clients contribute close to 80% of total profits, underscoring the strategic value of focusing on top performers while streamlining underperforming segments. Time-based analysis revealed cyclical behaviour, with profitability peaking mid-year and declining at year-end, suggesting seasonal or operational factors that warrant closer management attention.

The Statistical Process Control (SPC) phase reinforced that the company's delivery operations are statistically stable but not yet capable of consistently meeting customer expectations. The X-bar and S charts for key product categories showed no special-cause variation, confirming a controlled process. However, capability indices (Cpk values below 1.33) indicated that while processes are predictable, they remain insufficiently centred within performance specifications. The application of Western Electric rules and subsequent data cleaning improved control, proving that removing special causes can stabilise process performance.

Building on process stability, the Time-to-Serve optimisation and Design of Experiments (ANOVA and LSD tests) further refined operational insights. Both the brute-force and solver optimisation methods confirmed that six baristas per store achieved maximum profit, demonstrating the value of data-driven decision-making in resource allocation. The ANOVA and LSD results revealed statistically significant differences between treatment configurations, confirming that operational factors directly influence performance outcomes.

Finally, the reliability analysis demonstrated that service reliability can be significantly improved through strategic staffing decisions. The model showed that adding four extra staff members increased reliability from 50% to around 80% while reducing total annual costs by approximately R1 million. Beyond this level, additional hiring yielded diminishing financial returns. This evidence-based optimisation provides management with a practical framework for balancing cost efficiency with service quality.

Conclusion

In conclusion, this project provides a comprehensive assessment of the company's data, processes, and performance systems, progressing from data validation through statistical control to operational optimisation. The analysis revealed that while the organisation maintains complete and largely consistent datasets, duplication, outlier, and formatting issues require immediate correction to ensure data reliability.

Profitability analysis confirmed that business performance is driven by a narrow set of high-value products and customers, with clear seasonal patterns influencing revenue and profit trends. Statistical Process Control established that while key processes are stable, they require re-centring and tighter control to achieve full capability. The optimisation and experimental analysis confirmed that structured statistical approaches, such as solver models and ANOVA, can effectively guide workforce and process decisions to maximise performance.

The final reliability study translated these insights into tangible operational recommendations. By hiring four additional staff members, the company can meaningfully enhance reliability and reduce total costs, achieving an optimal balance between efficiency and service quality.

Together, these findings demonstrate how integrated statistical tools, ranging from SPC to DOE and reliability modelling, can drive both quality improvement and profitability. Continued focus on data accuracy, process capability, and evidence-based decision-making will position the company for sustainable operational excellence and long-term competitive advantage.

Bibliography

Reference list

HBK World. (2025). *Applying Reliability DOE Techniques to Investigate Effects on Product Life*. [online] Available at: https://www.hbkworld.com/en/knowledge/resource-center/articles/one-way-analysis-of-variance-anova-and-reliability-doe-applying-reliability-doe-techniques-to-investigate-effects-on-product-life? [Accessed 7 Oct. 2025].

Inez Maria Zwetsloot, L. Allison Jones-Farmer and Woodall, W.H. (2023). Monitoring univariate processes using control charts: Some practical issues and advice. *Quality Engineering*, 36(3), pp.1–13. doi:https://doi.org/10.1080/08982112.2023.2238049.

Jankovic, A., Chaudhary, G. and Goia, F. (2021). Designing the design of experiments (DOE) – An investigation on the influence of different factorial designs on the characterization of complex systems. *Energy and Buildings*, [online] 250(111298), p.111298. doi:https://doi.org/10.1016/j.enbuild.2021.111298.

Sałaciński, T., Chrzanowski, J. and Chmielewski, T. (2023). Statistical Process Control Using Control Charts with Variable Parameters. *Processes*, [online] 11(9), p.2744. doi:https://doi.org/10.3390/pr11092744.

Statista (2023). *Statista - the Statistics Portal for Market Data, Market Research and Market Studies*. [online] Statista.com. Available at: https://www.statista.com.