Quality Assurance

ECSA Project

Hunter, GA, Mr [26896648@sun.ac.za]
STELLENBOSCH UNIVERSITY

Contents

Table of Figures

Lists of Tables

Introduction

This report contains data-driven investigations that simulate real industrial engineering practises. The simulations focus on attributes such as service delivery, performance, operational efficiency and process control.

This project involves analysing and datasets related to operations like sales, delivery times and optimal service operations for profit. Through the use of R programming, statistical and analytical methods were used to gain valuable insights from the data. The tasks of the project included statistical process control (SPC), descriptive statics, process capacity analysis and risk estimation through type 1 and 2 errors. The project also incorporated data correction, profit optimisation and reliability modelling.

This report will show how analytical tools and techniques can be used be to identify trends and learn business insights. The findings in the report reflect both practical interpretation and technical accuracy.

Part 1.2 Descriptive Statistics

This component of the project was to explore and make sure understanding of the data was sound. The datasets investigated were customers, products_data, products_hq, and sales2022and2023. An overview of the dataset was initially done followed by an analysis to identify key patters or trends.

Overview of Datasets

The following datasets had no missing values, this shows that the datasets are somewhat consistent and are likely well-maintained records. The following table summarises the main numerical values.

dataset <chr>	variable <chr>	n <int>	n_missing <int>	complete_rate <dbl>	mean <dbl>	sd <dbl>	p0 <dbl>
customers	age	5000	0	1	51.55380	2.121610e+01	16.0000000
customers	income	5000	0	1	80797.00000	3.315011e+04	5000.0000000
products	markup	60	0	1	20.46167	6.072598e+00	10.1300000
products	selling_price	60	0	1	4493.59283	6.503770e+03	350.4500000
products_hq	markup	360	0	1	20.38550	5.665949e+00	10.0600000
products_hq	selling_price	360	0	1	4410.96186	6.463823e+03	290.5200000
sales	delivery_hours	100000	0	1	17.47646	9.999944e+00	0.2772000
sales	order_day	100000	0	1	15.49683	8.646506e+00	1.0000000
sales	order_month	100000	0	1	6.44813	3.283446e+00	1.0000000
sales	order_time	100000	0	1	12.93230	5.495127e+00	1.0000000

Table 1: Overview of Dataset

The dataset sizes changed quite a bit based on their function. The product dataset contained 60 unique products while the customer dataset had 5000 observations. The descriptive statistics in table 1 show that the mean customer age was 51.5 and mean income was 80797. The data being complete and not having an out of order entries emphasises on the fact this this data is accurate complete and a good starting point for further investigation.

Customer Demographics

The age distribution in the below figure is relatively uniform between the ages of 20 and 90. There is a slightly higher number of customers between ages 25 to 60 which indicated that this company likely services a broad adult category. Ages around 30 have the highest count of customers.

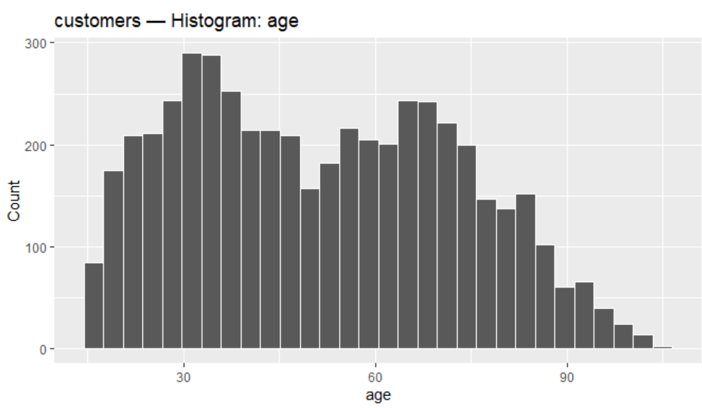

Figure 1: Customer Demographics - Age

Sales Patterns

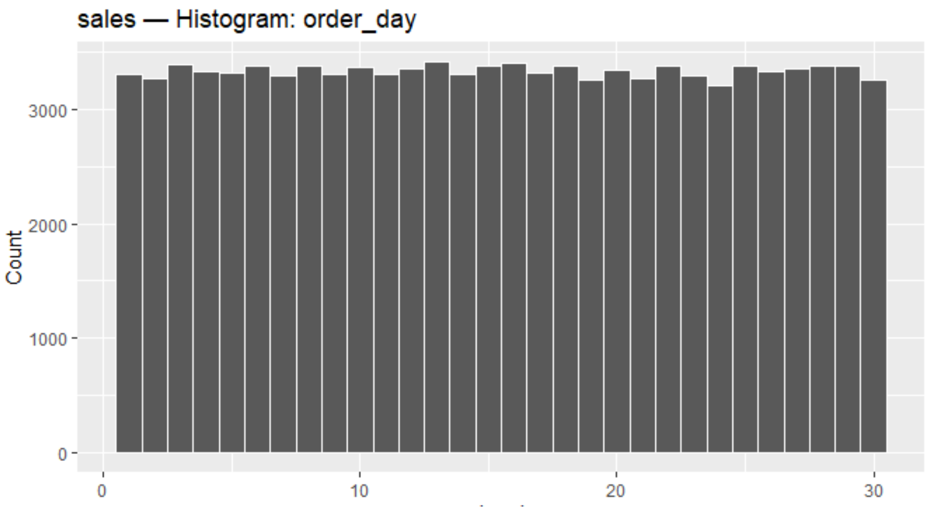

Figure 2: Sales

Sales throughout the month from Day 1 to Day 30 show a uniform distribution as there is no dramatic change in sales from one day compared to another. This continuous regularity suggests the idea that there is continuous operation throughout the month with little to no downtime or any major stock out issues. The pattern also suggested that there were no sales presents because there were no increased spikes in order quantity or if there was a sale it was unsuccessful in improving customer interaction.

Product Markup

The markup up distribution of products has distinct peaks compared to a smooth, bell-shaped curve showing that markups are not normal distributed. Instead, clusters are found near certain markup values such as 10, 15, 20-25 and 30. This suggests that the company makes use of tier pricing strategies

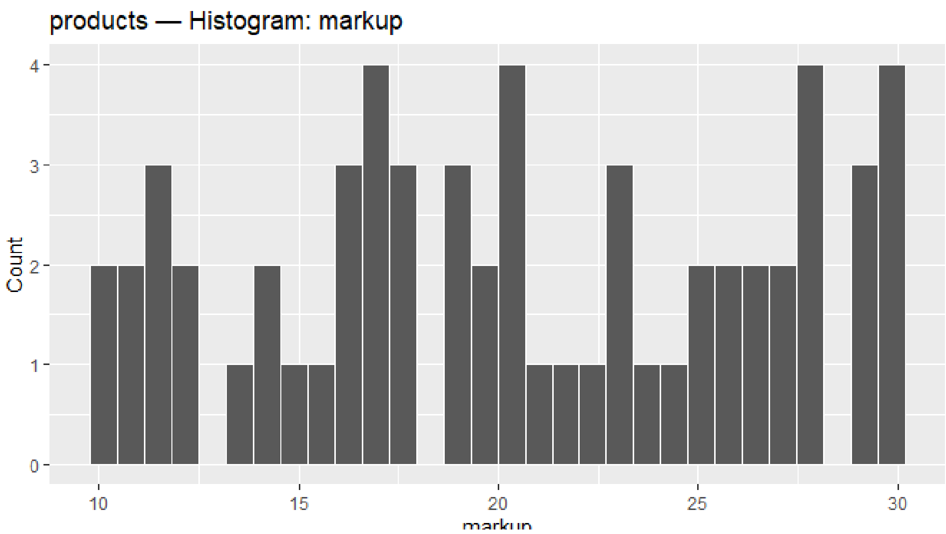

Figure 3: Product Markup

Part 3: Statistical Process Control

Statistical Process Control methods were used to monitor delivery performance and identify process stability over time.

X-bar and S-Chart

X-bar and S-Charts were used to monitor the average delivery time and the spread of the data. Sample size of 24 was used per a subgroup and the first 30 samples were used to determine the initial centre lines and control limits. The limits show the natural range of variation if the process is stable.

The below figures 4 to 7 show the S-charts for a different product. Throughout the graphs most points are located within the 3σ control limit. This indicates stable performance overall, however, some samples briefly exceed the control limit. Suggesting that there are short term process disturbances. These out-of-control points need to be investigated in a real industrial setting to potentially identify root causes like delayed shipments or system interruptions.

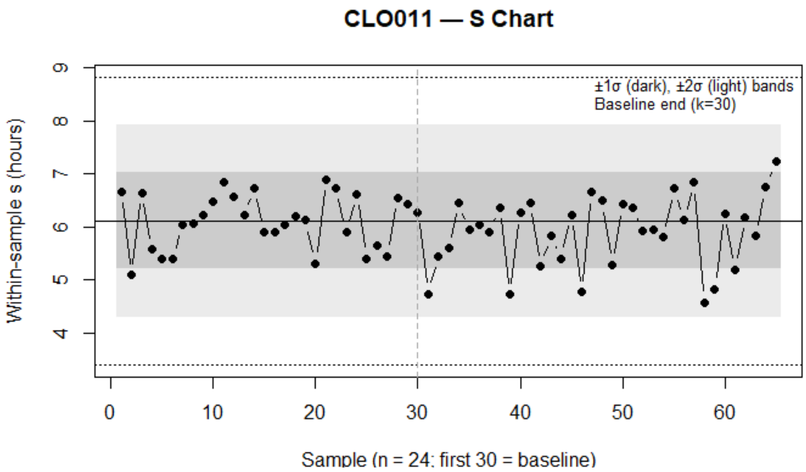

Figure 4: S-Chart

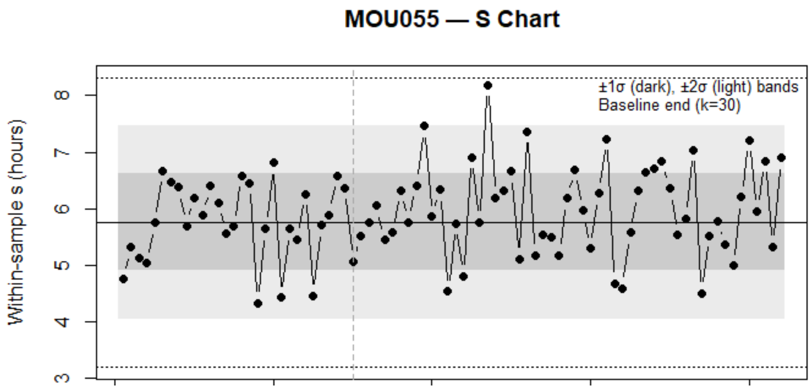

Figure 5: S-Chart

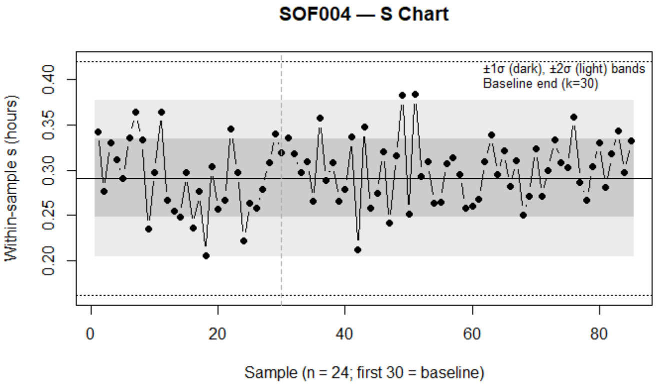

Figure 7: S-Chart

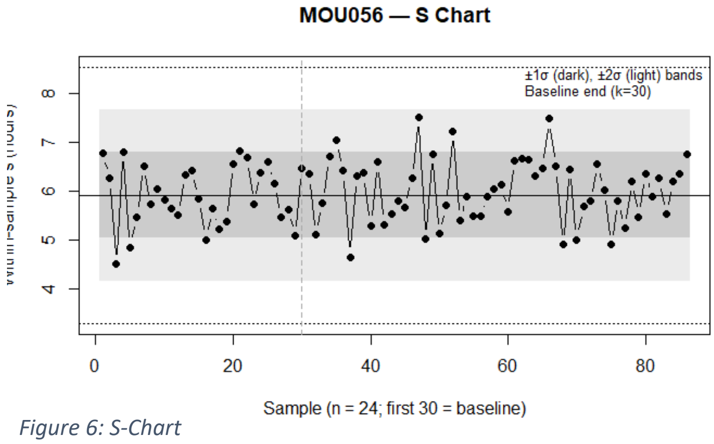

Figure 6: S-Chart

Monitoring and Simulation

After the first 30 samples additional subgroups were analyses to simulate real time monitoring. The control limits were retained to evaluate whether new data behaved within the expected behaviours. The charts showed that while most products remained stable, a few products experienced temporary spikes in variation. This shows the importance of continuous update monitoring and timely corrective action to maintain quality performances.

Process Capability Indices

Process Capability Indices analysis was done on the delivery time data to determine how ell each process meets the specific delivery target limits. The Upper Specification limit (USL) was set to 32 hours and the Lower Specification limit was set to 0 hours. These represented the maximum and minimum appropriate delivery times.

Each products mean delivery time as well as standard deviation were calculated. The Cp value measures the overall process spread relative to the tolerance band. The Cpk shows how centred the process mean is within the band.

ProductID <chr>	mean_delivery <dbl>	sd_delivery <dbl>	Cp <dbl>	Cpu <dbl>	Cpl <dbl>	Cpk <dbl>
CLO011	21.474786	6.2372126	0.8550828	0.5624956	1.147670	0.5624956
CLO012	21.849878	6.1958087	0.8607970	0.5460746	1.175519	0.5460746
CLO013	21.812999	6.1184810	0.8716760	0.5549853	1.188367	0.5549853
CLO014	21.768569	6.0907628	0.8756429	0.5599425	1.191343	0.5599425
CLO015	21.853156	5.9981690	0.8891602	0.5638857	1.214435	0.5638857
CLO016	21.854791	6.1472805	0.8675923	0.5501191	1.185066	0.5501191
CLO017	21.752861	6.0729882	0.8782058	0.5624436	1.193968	0.5624436
CLO018	21.593204	6.1356682	0.8692343	0.5653715	1.173097	0.5653715
CLO019	21.906919	6.1510839	0.8670559	0.5469541	1.187158	0.5469541
CLO020	21.326189	5.9877608	0.8907058	0.5942016	1.187210	0.5942016

Table 2: Capability Indices

All products processes had a Cp value of below 1 which shows that the overall variation in delivery hours is larger or wider than the specifications. This shows that the deliveries are not yet able to be consistently delivery between the 0 to 32 hours delivery time.

Variation across product types does not differ much (the standard deviation is only around 6 hours). This combined with the low Cp and Cpk values show that there is a system wide problem s oppose to a single product type having issues. This shows a need for system tightening and/or better control systems to achieve full capability.

4: Risk and Error in Process Control

This section covers understanding the risk involved in interpreting control chart signals. There are two main error types, Type 1 (false alarm) and type 2 (missed detection). Which together describe the reliability of the SPC system.

4.1 Type 1 Error

This occurs when a stable process is incorrectly identified as out of control. The table below summarises the calculated results.

Rule <chr>	TypeI_Probability <dbl>
A: s > +3σ (upper only)	1.349898e-03
B: diagnostic (no alarm)	0.000000e+00
C: 4× $\bar{X}$ > +2σ	2.678772e-07

Table 3: Error 1

Rule A has a probability of 0.135 percent which shows that the chance of a false alarm very low. Only 1 in every 740 samples are incorrect false alarms. Rule C and B also show little to no error probabilities. Proving again that the control limits are sufficiently wide enough to prevent false alarms.

4.2 Type 2 Error

Type 2 error pick up when the process mean or variability change, however the control chart is not able to identify it. The calculation shows that the Type II error (β) = 0.8412 and Power (1−β) = 0.1588. This is a bad results as it shows that the control chart misses true shifts 84.1 percent of the time and only correctly identifies it 15.9 percent. This shows that the control chart is not sensitive to small process changes.

To improve this poor detection sensitive the company could implement supplementary rules, reduce subgroup size or sampling interval, or use cumulative sum.

Metric <chr>	Value <dbl>
Type II error (β)	0.8411783
Power (1-β)	0.1588217

Table 4: Error 2

4.3 Data correction and Analysis

All data graphs reviewed in question 1.2 (orders a day, markup and customer age), remained the exact same. One graph that changed was the Units sold per product type where monitors and laptops significantly decreased from around 250 000 to around 140 000 each. The figure below shows the new updated graph.

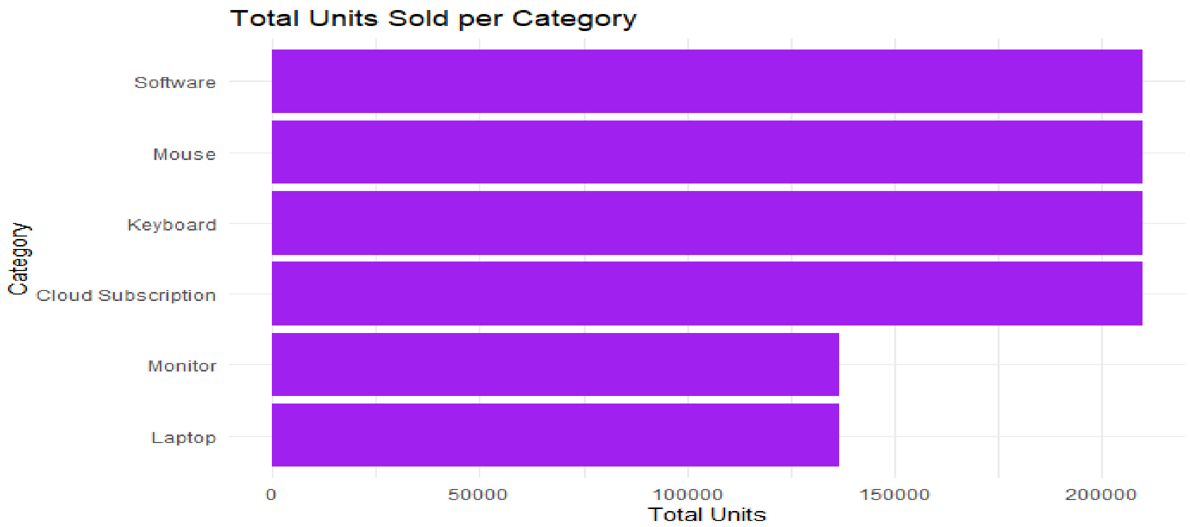

Figure 8: Data Correction

Part 5: Optimising profit for the coffee shops

This component of the project looks at the relationships between number of baristas, service speed and profitability to determine the optimal number of staff for efficiency and overall shop profitability.

5.1 Seconds per Service vs number of Baristas

Both figures 9 and 10 show how as the number of baristas increase for both shops 1 and 2 the average seconds per service decrease. These graphs show a inverse relationship between the number of baristas and service time. This is excepted and makes logical sense, the more employees a company has the more efferently they are able to serve customers.

However, the rate of improvement diminishes as the graphs begin to flatten out. This specifically happens occurs around 4 baristas. This suggests that increasing staff beyond the point contributes to less significantly to improvement. This trend shows the law of diminishing returns.

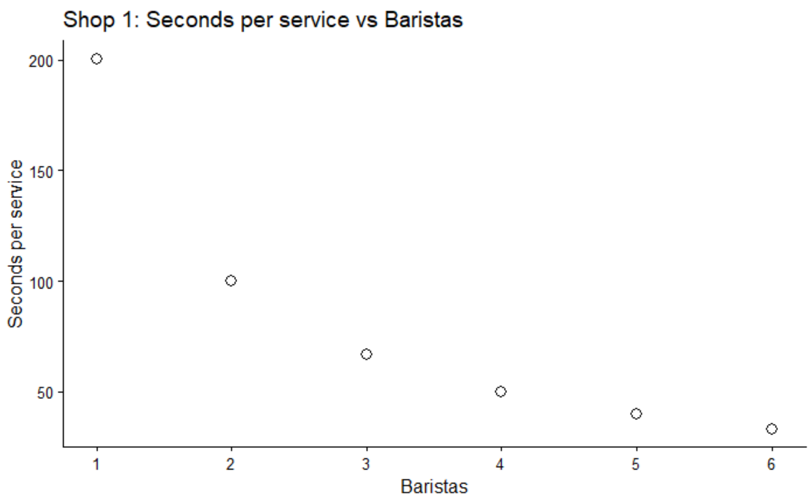

Figure 9: Seconds per Service Shop 1

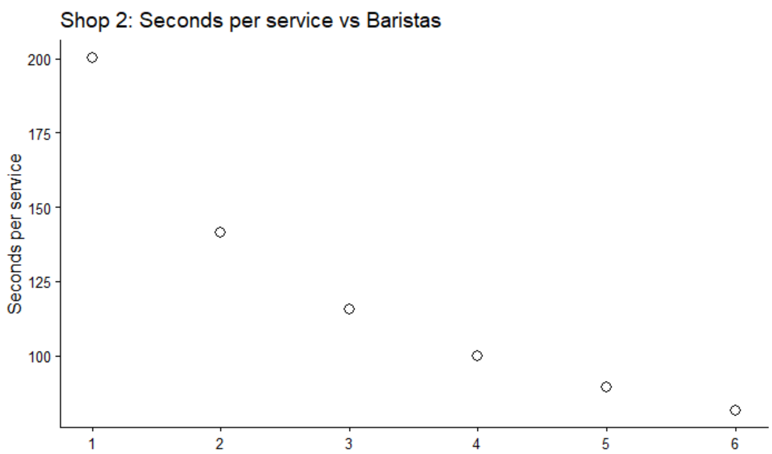

Figure 10:Seconds per Service Shop 2

5.2 Profit per a Barista

Figures 11 and 512 show the profit per a barista across staffing levels at both shops.

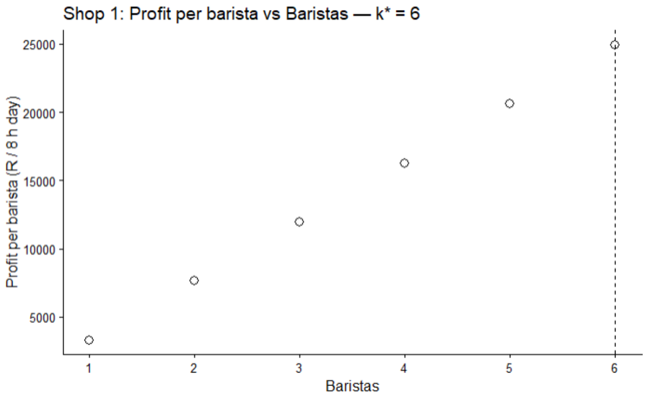

Figure 11: Profit Shop 1

For shop 1 (shown in figure 11) the graph displays a strong, linear positive trend. It is seen that as the number of baristas increase, there is a consistent profit increase. This positive upward trend explains the efficiency gained through higher customer throughput and reduced service time as baristas are increased. Profit peaks at approximately R25 000 per an 8-hour day, where 6 staff members are working. The dotted vertical line at k* = 6 indicated the optimal staffing level. Shop 1 demonstrates great scalability as profitability and service speed both increase consistently until optimal profitability is reached. It is seen that having 6 baristas is the most optimal point for shop 1.

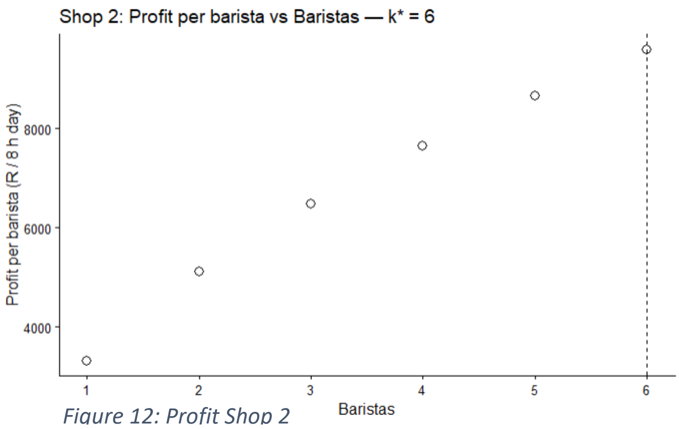

Figure 12: Profit Shop 2

Figure 12 shows a similar relationship between staffing and profitability as shop 1 because as the number of baristas increases so does profitability per a barista. Showing again that better service times and a higher customer throughput directly relates to profitability. The slope of the curve is lower and slightly decrease but still in an upwards direction. This shows that the law of diminishing returns is applicable when looking at the optimal number of baristas, each barista adds less incremental profit than the previous one. 6 baristas are the optimal number of baristas for shop 6 to maximise profitability.

Part 6: DOE and MANOVA / ANOVA Analysis

The purpose of this component of the project was to investigate if there was a delivery time difference between years 2022 and 2023 and across months within these years. A two factor ANOVA was conducted to evaluate the effects of month and year on delivery time.

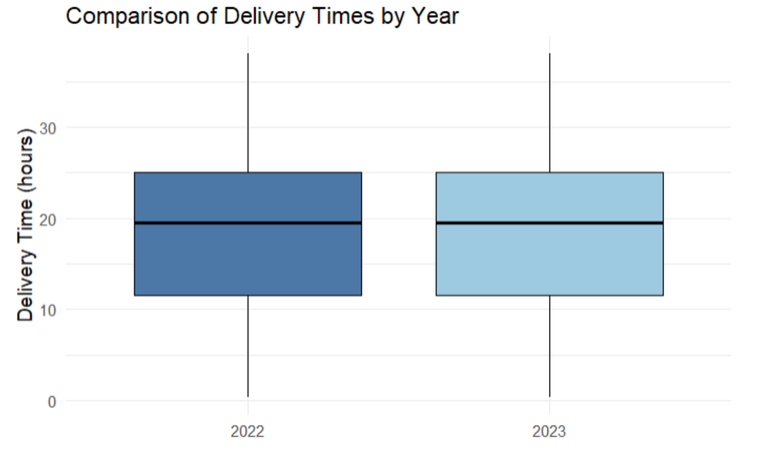

Figure 13: Delivery Time Years

The box plot that summarises the delivery time analysis of years 2022 and 2023 shows that median delivery times for both 2022 and 2023 were almost identical at around 19.5 hours. There is only a marginal difference between in mean the delivery time between years. As year 2022 has a mean time of 17.51 hours while 2023 had a mean time of 17.43. However, this difference is so marginal it should not be investigated in depth. There were no effective improvements in delivery efficiency between years. The year effect was statistically insignificant as p = 0.2366, showing that performance remained stable across years.

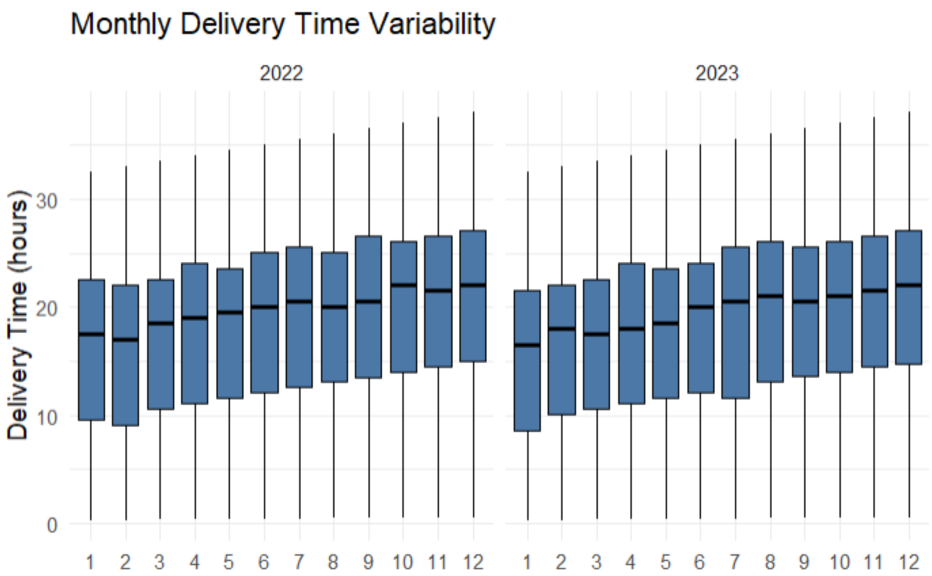

Figure 14: Delivery Time Month

However monthly delivery variability shows in figure 6.2, shows a clear pattern. Delivery times increased gradually from the beginning months towards the final months of the year. The ANOVA results support this trend with month showing a large effect ($p<0.001$), explaining a small but meaningful portion of variation ($\eta^2 = 0.0173$). This factor emphasises on the fact that seasonal or demand related factors are likely to influence delivery performance throughput the year.

Part 7

7.1 Reliable Service Days

The dataset had recorded the number of employees on duty over 397 days. Days with less that 15 employees were classified as problem days, and there were a resulting 31 problem days out of the total 397 days. This figured was normalised to a standard calendar year and resulted in their being 28.5 problem days. The calculated service reliability was 92.2 percent, this is a good baseline measure of operational reliability and a good starting point to making further staffing adjustments.

7.2 Profit Optimisation

A binomial reliability model was developed to help determine the most profitable staffing setup in order to balance cost of hiring additional employees against expected loss from unreliable services. Each problem day was estimated to reduce sales by R20 000 whilst hiring additional employees would cost the company R25 000 per a month per an employee. Increasing staff members from 0 to 6 were investigated. For each scenario profitability was calculated and the best outcome was to hire one additional employee. This can be seen in figure 7.1 below as profitability increases and spikes when k = 1, which indicates employee 1 additional employee. Hiring an additional employee reduces problem days from 28.5 to 5.5 per a year and increases the company's reliability from 92.2% and 98.5% leading to an annual net profit gain of approximately R159 700.

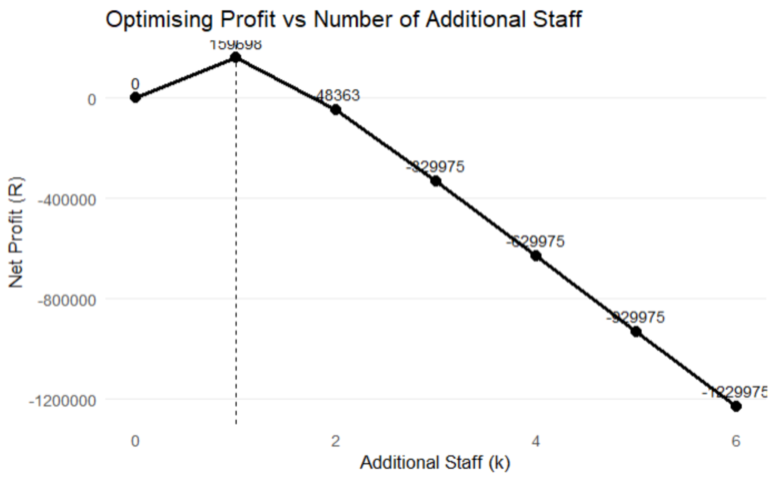

Figure 15: Profit Additional Employees

Conclusion

This report shows how the application of process control as well as data analytics methods were used to improve overall efficiency of multiple simulations. Through the use of R, descriptive statistics revealed key sales and product trends. SPC Indices identified process stability and performance gaps. Type 1 and 2 errors helped managed risk and data correction ensured reliable results. Finally profit optimisation was completed.

References

1factory. "A Guide to Process Capability (Cp, Cpk) and Process Performance (Pp, Ppk) | 1factory." *1factory.com*, 2024, www.1factory.com/quality-academy/guide-process-capability.html.

Basic Tools for Process Improvement Module 10 CONTROL CHART.

Bhandari, Pritha. "Type I & Type II Errors | Differences, Examples, Visualizations." *Scribbr*, 18 Jan. 2021, www.scribbr.com/statistics/type-i-and-type-ii-errors/.

Hayes, Adam. "Descriptive Statistics: Definition, Overview, Types, Example." *Investopedia*, 2024, www.investopedia.com/terms/d/descriptive_statistics.asp.

Hessing, Ted. "X Bar S Control Chart | Six Sigma Study Guide." *Six Sigma Study Guide*, 21 June 2019, sixsigmastudyguide.com/x-bar-s-chart/.

OpenAI. "ChatGPT." *ChatGPT*, OpenAI, 2025, chatgpt.com/.