Stellenbosch
UNIVERSITY
IYUNIVESITHI
UNIVERSITEIT

**Department of Industrial Engineering**

# Quality Assurance 344 – 2025

# ECSA Project

**Emmanuel Fako**
**26090376**

# Contents

## Table of Figures

# 1. Introduction

As the new data analyst, I have conducted a data analysis on our customer, product and sales data. The goal of this report is to provide an initial overview of the customers, product landscape and sales performance, to provide insights into the current position, as well as areas of concern and further improvement.

Thereafter, SPC analysis is done on the delivery times of future sales data, where use of control charts is demonstrated to monitor delivery time of the products as well as the process capability indices. Risk was analyzed, where the probability of Type I and II errors were assessed. Profit models were then used to optimize staffing levels to maximize profit, and the results were discussed.

ANOVA was conducted on the product type, Keyboard, to determine whether there was a significant difference in the mean delivery times. The report concludes with an analysis of the reliability of a car rental company, where the data was fitted to a binomial distribution to optimize the cost of services and maximize profit.

# 2. Analysis of Company Data (Part 1.2 & 4.3)

## 2.1 Data Quality Issues

Initial data preparation and cleaning revealed several insights into the quality of the data presented. The datasets revealed no missing values, irregular cardinality and abnormal outliers since product prices vary according to the quality. However, during the initial data loading phase, a data integrity issue was identified in the *products_data.csv* file. Duplicate *ProductID*'s were identified in *products_data.csv* and *products_headOffice.csv*, that had differed in *Category*, *Description*, *Selling Price* and *Markup*. To ensure the accuracy of this analysis, the first 10 rows of *product_data.csv* where the problem was present were removed and it was assumed that the correct information was in the file from the head office.

Products_data

| | ProductID | Category | Description | SellingPrice | Markup |
|---|---|---|---|---|---|
| 1 | SOF001 | Software | coral matt | 511.53 | 25.05 |

Products_headOffice

| | ProductID | Category | Description | SellingPrice | Markup |
|---|---|---|---|---|---|
| 1 | SOF001 | Software | coral silk | 521.72 | 15.65 |

*Figure 1:Duplicate Case in Products Dataset*

Within *products_data.csv*, other instances of *ProductID*'s not corresponding to the *Category* of the product were identified. This is indicative of an error with documenting the products and is a cause for serious concern as these products had been purchased by customers in the 2022 and 2023 financial years. This problem was not resolved before the analysis because removing the dataset would drastically affect the analysis of sales for the financial years. This is an issue that requires action from the warehousing department, and it is recommended that the problem be addressed for future purposes.

Products_data

| | ProductID | Category |
|---|---|---|
| 11 | CLO011 | Keyboard |
| 12 | CLO012 | Mouse |
| 13 | CLO013 | Software |
| 14 | CLO014 | Cloud Subscription |
| 15 | CLO015 | Laptop |
| 16 | CLO016 | Monitor |
| 17 | CLO017 | Keyboard |
| 18 | CLO018 | Mouse |
| 19 | CLO019 | Software |
| 20 | CLO020 | Cloud Subscription |

Products_headOffice

| | ProductID | Category |
|---|---|---|
| 61 | CLO001 | Cloud Subscription |
| 62 | CLO002 | Cloud Subscription |
| 63 | CLO003 | Cloud Subscription |
| 64 | CLO004 | Cloud Subscription |
| 65 | CLO005 | Cloud Subscription |
| 66 | CLO006 | Cloud Subscription |
| 67 | CLO007 | Cloud Subscription |
| 68 | CLO008 | Cloud Subscription |
| 69 | CLO009 | Cloud Subscription |
| 70 | CLO010 | Cloud Subscription |

*Figure 2:ProductCategory Mismatch in Products_data*

### Comments on cleaned data

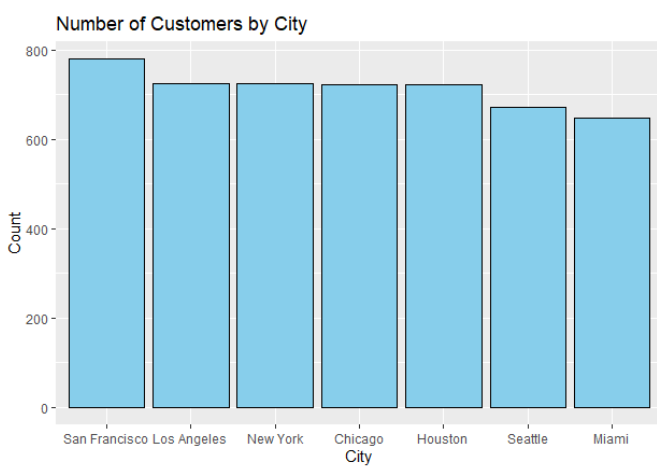
The data quality issues that were pointed out earlier in the report were corrected and updated to *products_HeadOddice2025.csv* and *products_data2025.csv*. Customer analysis was largely unaffected by the data error, but sales and product performance were re-analyzed. The previous analysis remains, and the corrected analysis can be found below the old analysis.

## 2.2 Customer Demographic Analysis

This analysis dives into the customer base and the insights that can be found from the information and characteristics of the customers. It answers the key questions of who the customers are, where they live, and what their age and income distribution is.

### Customer Locations

A review of the customer data reveals that San Francisco is the premier market, containing 780 customers. It is followed by a consistent tier of four major cities, showing a uniform market presence across locations. Los Angeles and New York each have 726 customers, while Chicago and Houston are very close with 726 customers each. The next tier consists of our smaller markets in Seattle and Maimi, with 673 and 647 customers respectively. This distribution shows a clear market leader in San Francisco, a stable core of four markets with a comparable market presence, and a pair of cities representing opportunities for targeted growth such as regional marketing or local offers to increase the customer base and bring gains to the company.




Number of Customers by City

| City | Number of Customers |
|---|---|
| San Francisco | 780 |
| Los Angeles | 726 |
| New York | 726 |
| Chicago | 724 |
| Houston | 724 |
| Seattle | 673 |
| Miami | 647 |

*Figure 3: Number of Customers by City*

## Customer Age Distribution

The customer ages reveal two primary customer segments which create a bimodal distribution. The largest and most concentrated group contains younger adults, with a sharp peak between 25 and 35. The second group consists of older adults with a wider peak between 60-75. There is also a clear dip in customer frequency for the middle-aged customers and shows that the group is currently underrepresented in the data.

Marketing efforts used by the company should not be one-size-fits-all. Promotions that attract 25-30 year olds are different to those that would appeal to 60-75 year olds. The company should consider developing separate marketing strategies for the two core demographics.

Linking the age distribution to the relationship between age and income, we see that the group with the largest purchasing power is the underrepresented demographic of middle-aged customers. The company should investigate this gap, as a targeted campaign to attract this missing segment could open large revenue streams.

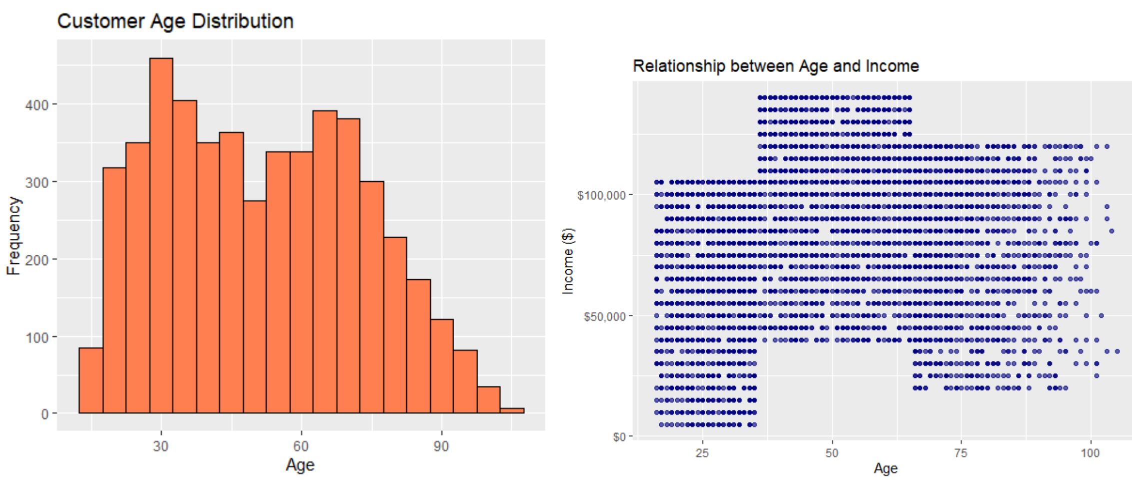



*Figure 4: Customer Age Analysis*

## Customer Income by Gender

An analysis of customer income by gender reveals that income is not a distinguishing characteristic between male and female customers in this dataset. All segments show a relatively similar central income, the same level of income inequality, and the same overall distribution. This implies that marketing or product strategy based on customer income levels should not be segmented by gender as it would not make that big of a difference.

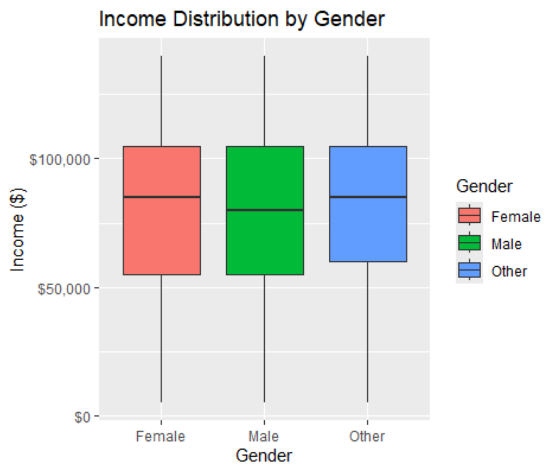



*Figure 5: Income Distribution by Gender*

## 2.3 Product Portfolio Analysis

The cleaned product data was merged and analyzed to reveal insights into what products the company sells and the distribution of the prices.

### Product Categories

The product portfolio of the company shows uniformity across all categories. With four categories (Software, Monitor, Laptop, Cloud Subscription) containing 18 products and the remaining two (Mouse, Keyboard) containing 19, it shows a product management strategy that is rigid and deliberate as opposed to one that has grown over time to meet the needs of customers. Customers may expect different levels of variety in a physical category like mice or laptops compared to a service category like cloud subscriptions. The company should investigate if this balanced product count directly leads to balanced sales performance or if some items are going unsold because customers do not see the need for it. It could be the case that a small number of products in each category drive most of the revenue, and if so, the poorly performing products should be considered for discontinuation to reduce inventory and production costs.

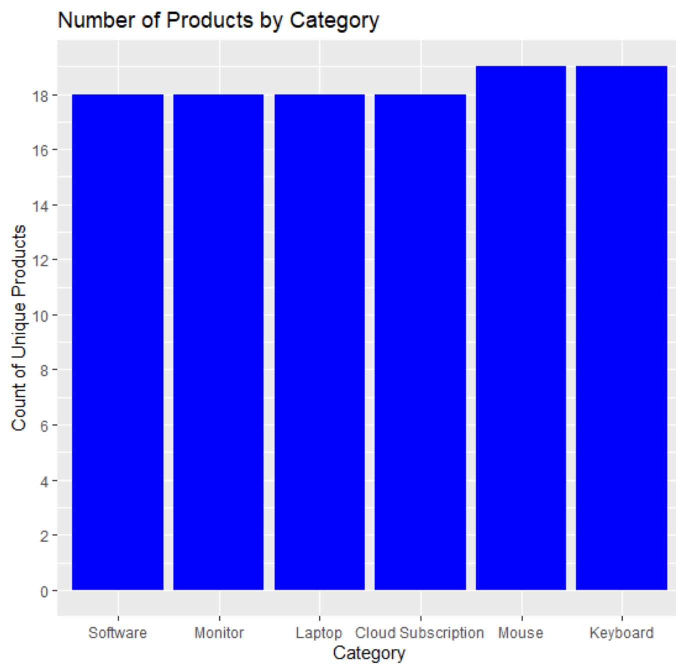



*Figure 6: Number of Products by Category*

## Rectified Product Categories

Product variety shows a largely expanded version of the previous analysis, with an even amount of 60 products per Category as opposed to 18. The question remains whether this deliberate product management strategy is what the customers wish for and leads to balanced sales performance, or if some products are going unsold.

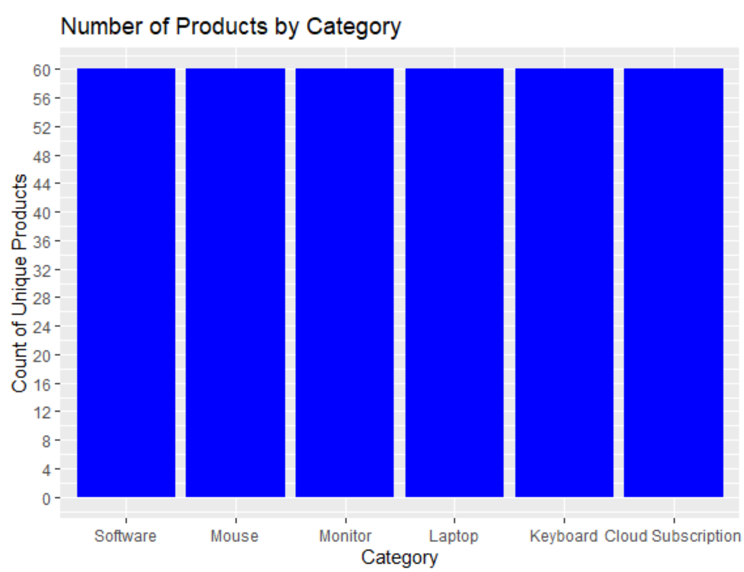



*Figure 7: Updated Product Count*

## Pricing Distribution

A pricing analysis reveals that selling prices are right skewed, with a high volume of products ranging from $300 to $1000. The graph also contains small selection of premium products with a higher price range. This is favorable for the company as they can be able to target a broad mainstream market with its core products, and are able to cater to the higher income customers that desire quality with no price cap.

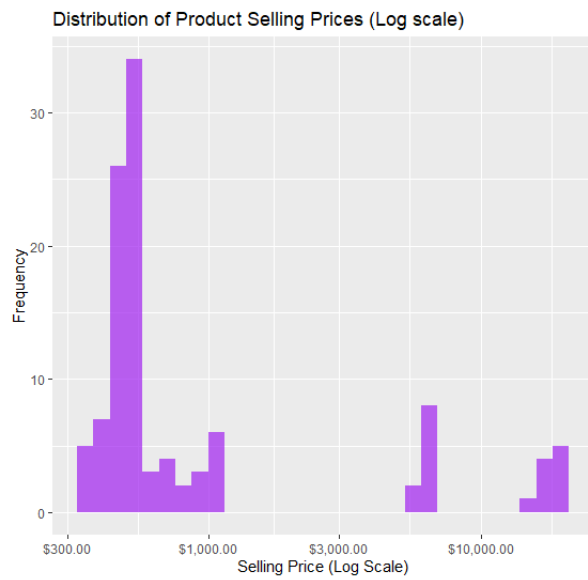



*Figure 8: Distribution of Selling Prices*

## Rectified Pricing Distribution

The Pricing analysis shows a similar right skewed pattern, with a high volume of products ranging from $300 to $1200. With the correction of the product data, the frequencies of selling prices increased, however the proportions remained relatively the same.

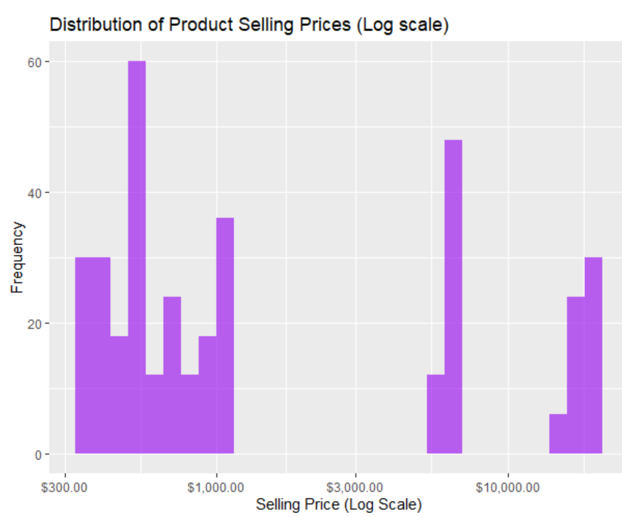



*Figure 9: Updated Distribution of Selling Prices*

## 2.4 Sales Performance Analysis

This section dives into the sales data, answering the key questions of how the company's revenue has trended over time, as well as which products generate the most revenue.

### Total Sales Over Time

The total sales revenue over time reveals a pattern of strong performance in 2022, followed by a more moderate but stable trend through 2023. In 2022, revenue begins around $150 million in January and quickly climbs. There are several months in 2022 that peak above $200 million between March and November, this shows a period of growth and high demand. In December 2022, there is a large drop in revenue to just above $117 million. Moving into 2023, sales start low in January but quickly increase and stabilizes mostly between $175 million and $190 million across the year. The figures for 2023 are consistently below those of 2022, experience the same year-end dip as the previous year and reach their lowest point across both years, of $110 million.

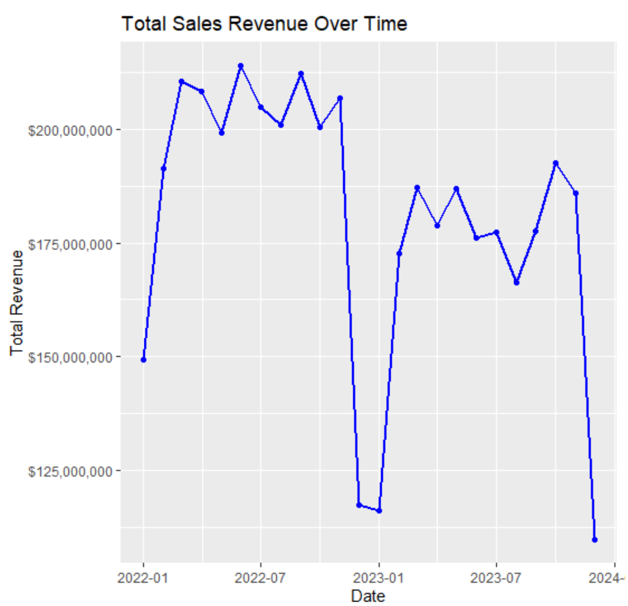
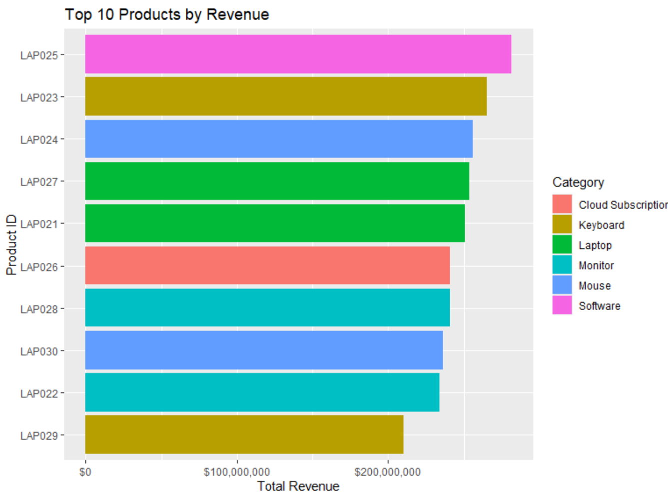



*Figure 10: Total Sales Revenue Over Time*

These patterns could be influenced by seasonality (demand surging mid-year and declining at year-end), consumer spending cycles (tighter budgets towards the end of the year and at the beginning). A consistent end of year drop may suggest that December is just a slow month for the business.

### Best and Worst Selling Products by Revenue

An analysis of the top products by revenue reveals that the company has had huge success with the products LAP021-LAP030, with LAP025 bringing in the most revenue. Keeping track of the best performing products is crucial as it will allow the company to keep tighter control of the stock that brings the most value and ensure that there are no stock outs or problems regarding these products.

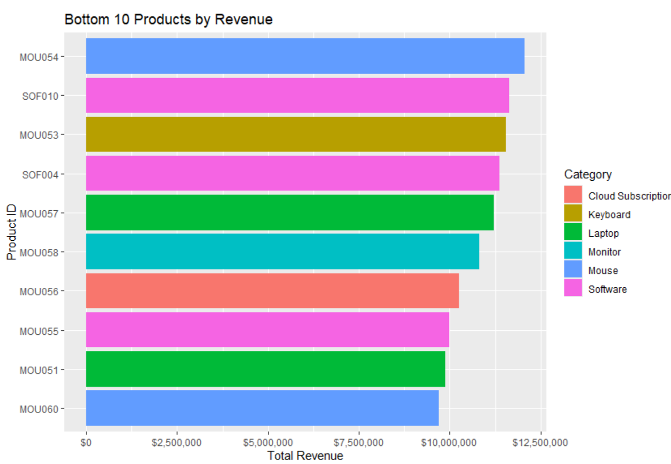



Top 10 Products by Total Revenue

| Product ID | Category | Total Revenue |
|---|---|---|
| LAP025 | Software | $281,754,471 |
| LAP023 | Keyboard | $265,237,837 |
| LAP024 | Mouse | $256,255,268 |
| LAP027 | Laptop | $254,026,069 |
| LAP021 | Laptop | $250,568,078 |
| LAP026 | Cloud Subscription | $241,231,494 |
| LAP028 | Monitor | $241,001,543 |
| LAP030 | Mouse | $236,466,128 |
| LAP022 | Monitor | $233,984,304 |
| LAP029 | Keyboard | $210,289,183 |

*Figure 11: Top 10 Products by Total Revenue*

In contrast, the figure below shows the bottom 10 products in terms of total revenue. These products likely do not cost as much as premium products but require further investigation as to whether these products are being sold in high volume or if they are premium products that are not generating enough revenue.



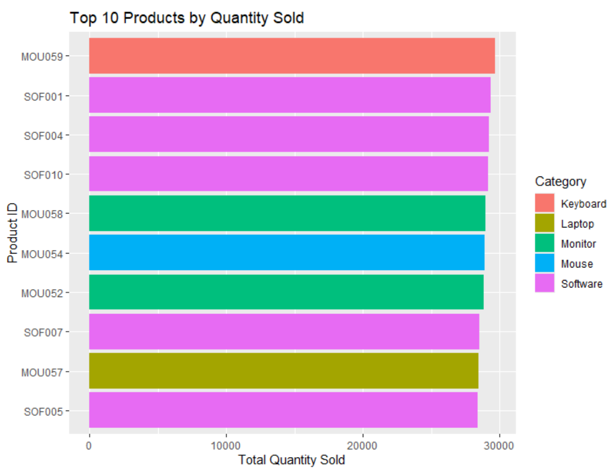

Bottom 10 Products by Total Revenue

| Product ID | Category | Total Revenue |
|---|---|---|
| MOU054 | Mouse | $12,052,425 |
| SOF010 | Software | $11,650,574 |
| MOU053 | Keyboard | $11,550,465 |
| SOF004 | Software | $11,375,833 |
| MOU057 | Laptop | $11,207,189 |
| MOU058 | Monitor | $10,812,370 |
| MOU056 | Cloud Subscription | $10,256,405 |
| MOU055 | Software | $9,995,875 |
| MOU051 | Laptop | $9,883,651 |
| MOU060 | Mouse | $9,712,722 |

*Figure 12: Bottom 10 Products by Total Revenue*

## Best and Worst Selling Products by Quantity

Regarding the top product quantities sold, we see some familiar products appearing that were present in the low ranks of products sold by revenue. This supports the claim that even though these are the low-cost, high-volume products that the company sells and even though they do not bring the most revenue, products like these are essential as they are a staple for the customers and generate the most movement. An example is the product MOU058 which sits 5th highest in total quantity with a total of 28924 units sold but is the 5th lowest product by total revenue with $10 812 370.

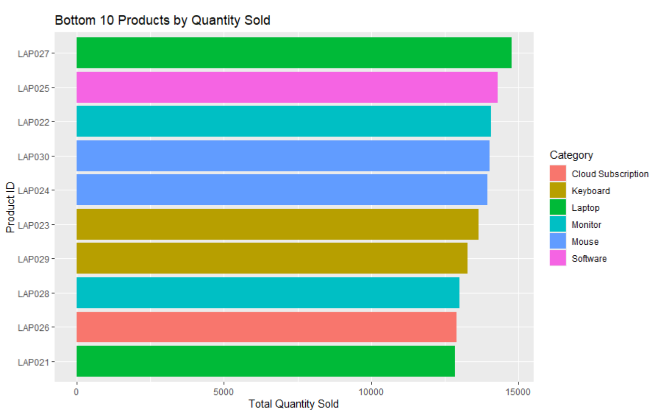



Top 10 Products by Total Quantity Sold

| Product ID | Category | Total Quantity Sold |
|---|---|---|
| MOU059 | Keyboard | 29,675 |
| SOF001 | Software | 29,336 |
| SOF004 | Software | 29,219 |
| SOF010 | Software | 29,168 |
| MOU058 | Monitor | 28,924 |
| MOU054 | Mouse | 28,875 |
| MOU052 | Monitor | 28,804 |
| SOF007 | Software | 28,517 |
| MOU057 | Laptop | 28,423 |
| SOF005 | Software | 28,412 |

*Figure 13: Top 10 Products by Quantity Sold*

The bottom 10 products by total quantity sold show a very interesting insight, in that the lowest products sold by quantity are the same products that yield the most revenue for the company. This indicates that these are the premium products of the company, and even though they do not generate large sales volume, they are the catalyst to the company's success in that they bring enormous gains.



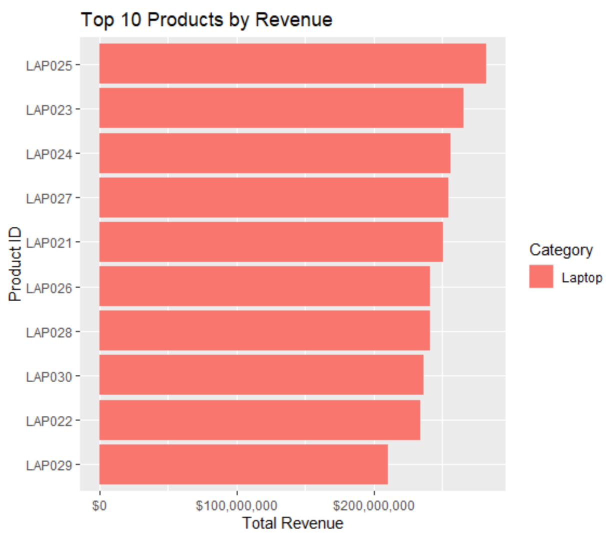

Bottom 10 Products by Total Quantity Sold

| Product ID | Category | Total Quantity Sold |
|---|---|---|
| LAP027 | Laptop | 14,767 |
| LAP025 | Software | 14,284 |
| LAP022 | Monitor | 14,058 |
| LAP030 | Mouse | 14,025 |
| LAP024 | Mouse | 13,952 |
| LAP023 | Keyboard | 13,635 |
| LAP029 | Keyboard | 13,266 |
| LAP028 | Monitor | 12,989 |
| LAP026 | Cloud Subscription | 12,892 |
| LAP021 | Laptop | 12,853 |

*Figure 14: Bottom 10 Products by Quantity Sold*

## Rectified Best and Worst Selling Products by Revenue

Previously, the Top 10 Products were split between several categories but contained similar *ProductIDs*. The rankings and total revenue remained the same, but the new graph shows the dominance of the *laptop* category and the significant amount of income that this category brings. Focus remains the same, about keeping tighter control of the best-selling products to prevent stockouts.

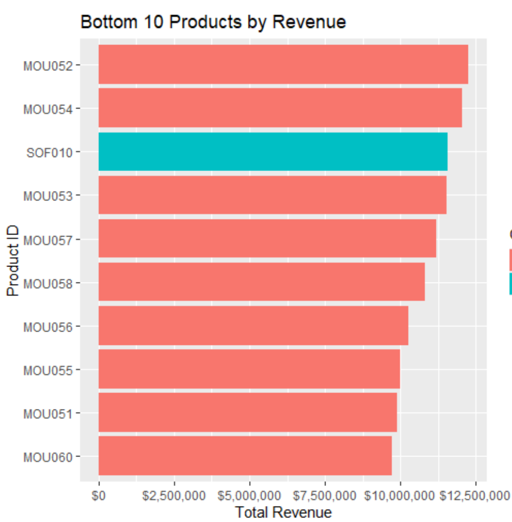



Top 10 Products by Total Revenue

| Product ID | Category | Total Revenue |
|---|---|---|
| LAP025 | Laptop | $281,754,471 |
| LAP023 | Laptop | $265,237,837 |
| LAP024 | Laptop | $256,255,268 |
| LAP027 | Laptop | $254,026,069 |
| LAP021 | Laptop | $250,568,078 |
| LAP026 | Laptop | $241,231,494 |
| LAP028 | Laptop | $241,001,543 |
| LAP030 | Laptop | $236,466,128 |
| LAP022 | Laptop | $233,984,304 |
| LAP029 | Laptop | $210,289,183 |

*Figure 15: Updated Top 10 Products by Total Revenue*

The analysis of the bottom 10 products by revenue shows that the category that produces the least revenue is *Mouse*, and it should be investigated whether this is since they are low value, high volume products, or if it is an indication of underperforming products.



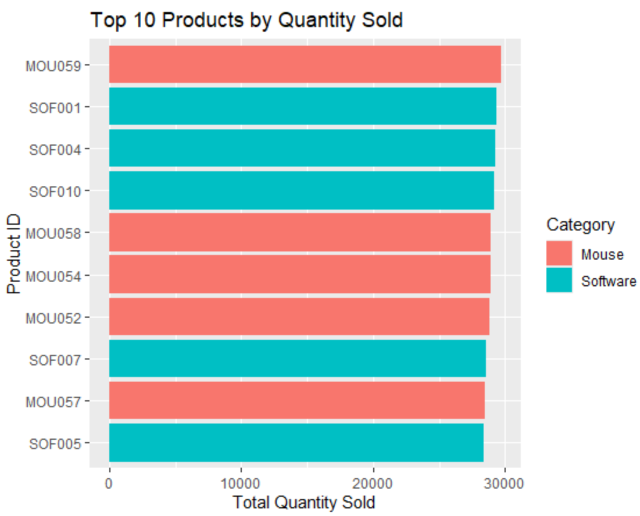

Bottom 10 Products by Total Revenue

| Product ID | Category | Total Revenue |
|---|---|---|
| MOU052 | Mouse | $12,245,733 |
| MOU054 | Mouse | $12,052,425 |
| SOF010 | Software | $11,571,529 |
| MOU053 | Mouse | $11,550,465 |
| MOU057 | Mouse | $11,207,189 |
| MOU058 | Mouse | $10,812,370 |
| MOU056 | Mouse | $10,256,405 |
| MOU055 | Mouse | $9,995,875 |
| MOU051 | Mouse | $9,883,651 |
| MOU060 | Mouse | $9,712,722 |

*Figure 16: Updated Bottom 10 Products by Total Revenue*

## Rectified Best and Worst Selling Products by Quantity

The Rank of ProductID stayed the same, however the category paints correct picture of what category is performing, namely *Mouse* and *Software*. A few of the products that have the highest quantity sold are products that appear on the bottom 10 products by revenue, suggesting that they may be low-value, high-volume products.

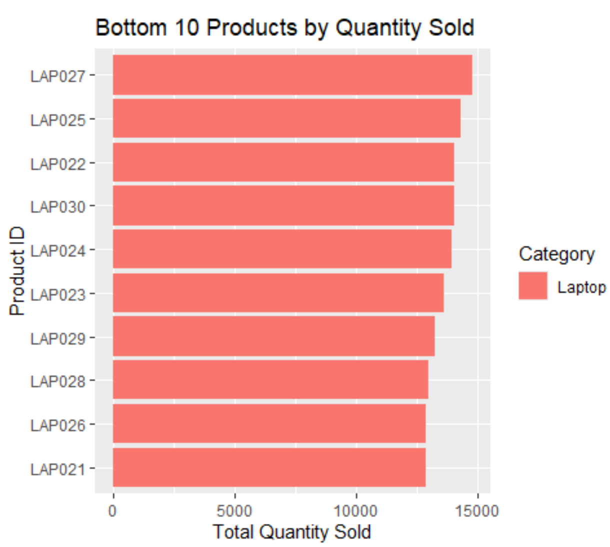



Top 10 Products by Total Quantity Sold

| Product ID | Category | Total Quantity Sold |
|---|---|---|
| MOU059 | Mouse | 29,675 |
| SOF001 | Software | 29,336 |
| SOF004 | Software | 29,219 |
| SOF010 | Software | 29,168 |
| MOU058 | Mouse | 28,924 |
| MOU054 | Mouse | 28,875 |
| MOU052 | Mouse | 28,804 |
| SOF007 | Software | 28,517 |
| MOU057 | Mouse | 28,423 |
| SOF005 | Software | 28,412 |

*Figure 17:Updated Top 10 Products by Quantity*

The bottom 10 remains unchanged except for the correct category type, showing that the category *Laptop* has the least quantity sold. All 10 of the products appear in the top 10 products by revenue, which means that they are likely to be premium laptops that do not have much sales volume, but are high in value.



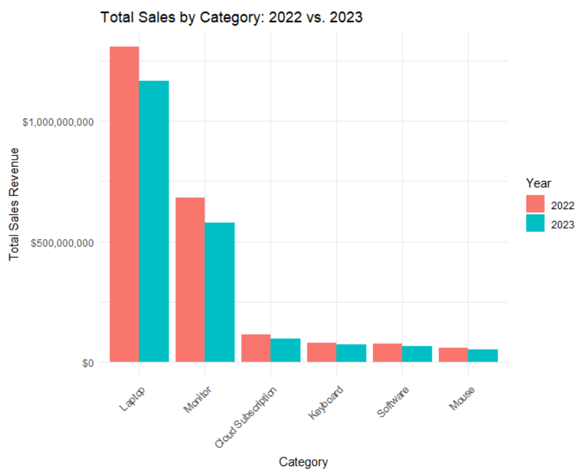

Bottom 10 Products by Total Quantity Sold

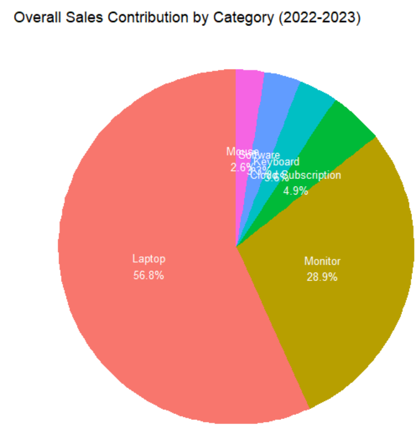
| Product ID | Category | Total Quantity Sold |
|---|---|---|
| LAP027 | Laptop | 14,767 |
| LAP025 | Laptop | 14,284 |
| LAP022 | Laptop | 14,058 |
| LAP030 | Laptop | 14,025 |
| LAP024 | Laptop | 13,952 |
| LAP023 | Laptop | 13,635 |
| LAP029 | Laptop | 13,266 |
| LAP028 | Laptop | 12,989 |
| LAP026 | Laptop | 12,892 |
| LAP021 | Laptop | 12,853 |

*Figure 18:Updated Bottom 10 Products by Quantity*

## 2022 and 2023 Sales

The Sales volume per product for 2022 and 2023 based on the updated data shows two key dominating product types, Laptop and Monitor. The other product types contribute only a fraction to the total sales value. An area of concern is the category-wide decrease in sales of 2023 compared to 2022. Laptops sales make up 56.8% of total sales in 2022 and 2023. Together with Monitors, these two product types make up almost 86% of total sales. This is a significant portion, and without them the company fails. Efforts should be made to keep tighter control of these product types and ensure efficiency and quality of the products.




Total Sales by Category: 2022 vs 2023

| **Category** | **Sales 2022** | **Sales 2023** | **Total Sales (All Years)** |
|---|---|---|---|
| Laptop | 1,306,924,896 | 1,163,889,479 | 2,470,814,376 |
| Monitor | 680,557,278 | 578,385,570 | 1,258,942,847 |
| Cloud Subscription | 115,394,937 | 98,715,482 | 214,110,418 |
| Keyboard | 81,503,144 | 73,499,067 | 155,002,210 |
| Software | 76,058,870 | 66,468,485 | 142,527,355 |
| Mouse | 59,970,894 | 51,219,577 | 111,190,471 |

*Figure 19: Updated Sales Data 2022 and 2023*

# 2.5 Operational Efficiency and Product Strategy Analysis

## Fulfilment Time by City

The Fulfilment time for all cities show a slightly skewed chart where more orders are fulfilled at a longer time. Although it is even across cities, slower fulfilment times lead to more unsatisfied customers and lower customer retention. The company should work to decrease the fulfilment time by having more distribution centers or reconfigure their distribution network to be geared towards efficiency and reliability. The analysis also revealed many outliers in every city, where the fulfilment time took extremely long and likely resulted in unsatisfied customers.

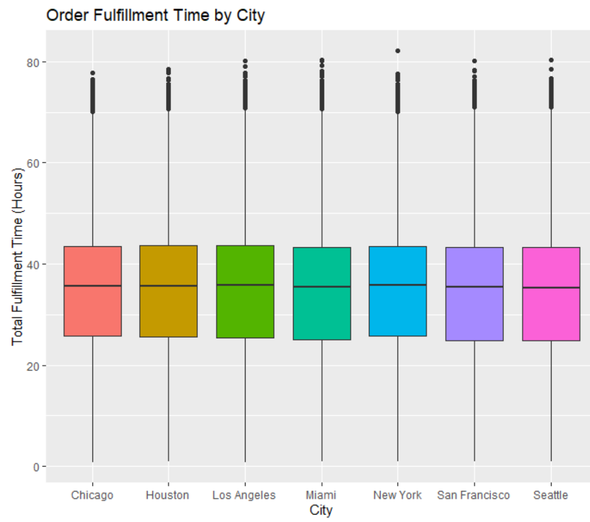



*Figure 20: Order Fulfilment Time by City*

# 3. Conclusion of Part 1.2

This analysis shows a company with a strong product portfolio that serves both a high-volume and a high-revenue function. The customer base is well established but mostly split among younger and older adults, leaving an opportunity with the underrepresented middle-aged demographic. The company faces a problem with poor data integrity related to product documentation and inefficient order fulfilment processes that show long and inconsistent delivery times across all markets.

Moving forward, the immediate priorities should be to rectify the data quality issues at their source to allow for accurate analysis and to re-evaluate the distribution network to improve fulfilment speed and reliability. Addressing these challenges while developing marketing strategies for the middle-aged demographic and smaller markets like Seattle and Miami, will be key to future growth and maintaining customer satisfaction.

# 4. Statistical Process Control (Part 3)

## 4.1 Chart initialization (3.10

The process under review is the Delivery times of the company's products in 2026 and 2027 per product type. To prepare the data, the data was arranged so that the orders arrived in an ordered timeline. The process was managed to an error-free period until 30 samples of 24 sales instances were collected. The UCL, LCL, CL, 2Sigma(green) and 1Sigma(blue) were calculated from these samples. During this period, no samples appeared outside the UCL and LCL. The control chart for Laptop can be found below, and the rest can be found in Appendix 1A.

### Laptop

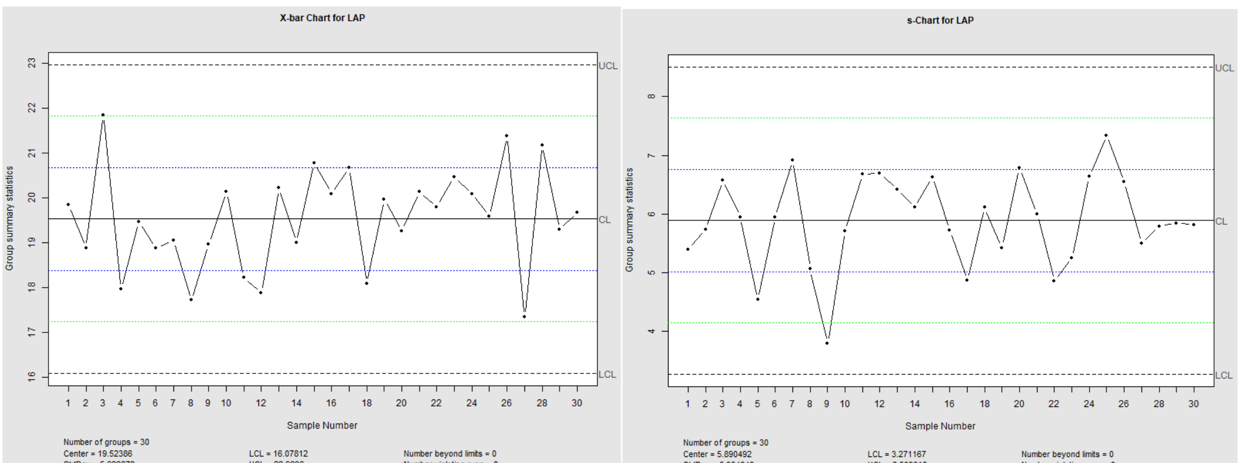



*Figure 21: X-bar and S Chart for Laptop*

## 4.2 Accelerated simulation (3.2)

The processes continued to run, and samples were recorded. The initial control limits were used, producing detailed control charts that can be further evaluated for control purposes. The full control chart for Laptop can be found below, and the rest of the control charts can be found in Appendix 1B

### Laptop

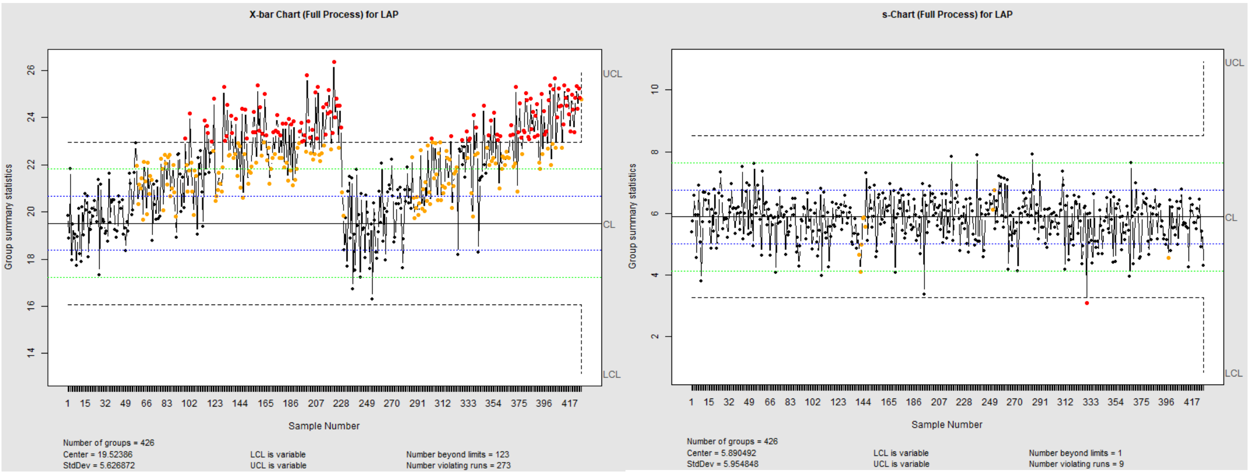



*Figure 22: Full X-bar and S Chart for Laptop*

The s-chart represents the stability and consistency of the process. It is important to analyze the s-chart first, if the process variation is out of control, the control limits on the corresponding $\bar{X}$ chart are meaningless because the control limits are calculated directly from the average variation shown in the s-chart. The s-charts for most categories lie inside the control limits, except for Laptop and Mouse with one point either above the UCL or below the LCL. An unstable s-chart is a red flag that requires immediate attention before any other analysis can be trusted.

The $\bar{X}$ charts show a similar pattern in all product categories, where the process veers severely out of statistical control. There is an upward trend as time progresses, and the samples cross the UCL around sample 117 and move far beyond this limit before resetting and repeating this cycle. If delivery times are consistently becoming longer, action should be taken. It is at this point that the processes should not be allowed to continue, and the special cause variations should be investigated and rectified to prevent the system from continuing in this trend.

## 4.3 Process Capability (3.3)

The Process Capability indices were calculated for the process delivery times of all product types using the first 1000 deliveries per product type. It was assumed that all product types have LSL of 0, and USL of 32h. The indices can be found under the respective product graphs. To find which products can meet the VOC, the $Cp_k$ measure is the most important indicator, as it measures both spread and centering.

### Cloud Subscription

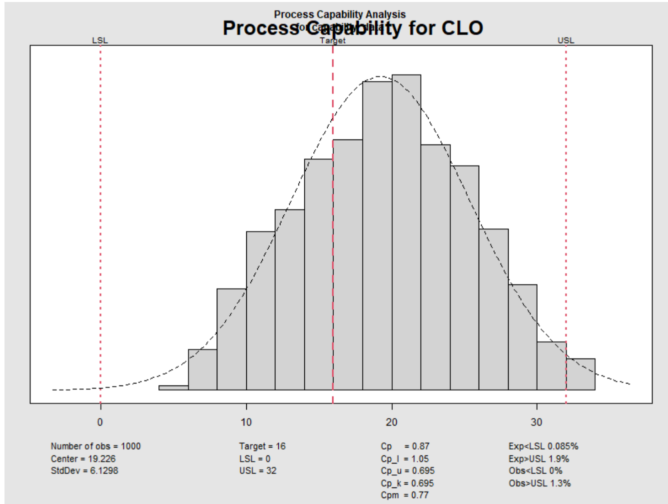



*Figure 23: Process Capability Analysis for Cloud Subscription*

Much like most of the categories, the process capability analysis shows that the process is not capable of consistently meeting specifications. The process mean is above the target of 16, showing that it is not centered. Overall, the process requires improvement by reducing variation and centering the process mean on the target to achieve acceptable capability levels.

The highest $Cp_k$ products are Software- 1.1, Keyboard- 0.714, Mouse- 0.707. The Software is the only process with a $Cp_k$ value greater than 1, so it would be the most capable of meeting specifications. This product would be classified as good as it is below the USL. This makes sense as software does not have to be delivered most of the time. The upper specification limit for software, however, can be much smaller as the process capability graph is too far from the USL. The process capability analysis for the remaining categories can be found in Appendix 1C.

## Software

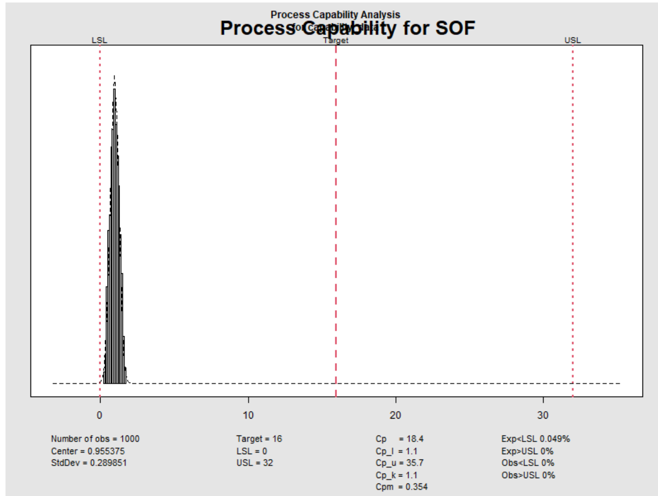



*Figure 24: Process Capability Analysis for Software*

# 4.4 Process Control Issues. (3.4)

Tests were conducted on all product types to find process control issues according to the following rules:
Test A: 1 ‘s’ sample outside of the upper +3 sigma-control limits for all product types (if many, list only the first 3 and last 3 and total number identified).
Test B: Find the most consecutive samples of s between the -1 and +1 sigma-control limits for all product types. This signifies good control.
Test C: 4 consecutive X-bar samples outside of the upper, second control limits for all product types (if many, list only the first 3 and last 3 and total number identified)
The results can be found below.

## Cloud Subscription

```
Rule A (s-chart > UCL): Found 0 violations.
Rule B (Good Control): Longest run of s-values within +/- 1 sigma is 35 samples.
  This run occurred from sample 474 to 508 .
Rule C (4 consecutive X-bar samples > +2 sigma): Found 14 instances of this pattern.
  Starting sample numbers of these runs: 122, 179, 192, 557, 604, 628
```

## Keyboard

```
Rule A (s-chart > UCL): Found 0 violations.
Rule B (Good Control): Longest run of s-values within +/- 1 sigma is 15 samples.
  This run occurred from sample 730 to 744 .
Rule C (4 consecutive X-bar samples > +2 sigma): Found 27 instances of this pattern.
  Starting sample numbers of these runs: 99, 112, 172, 687, 698, 726
```

## Laptop

```
Rule A (s-chart > UCL): Found 0 violations.
Rule B (Good Control): Longest run of s-values within +/- 1 sigma is 19 samples.
  This run occurred from sample 116 to 134 .
Rule C (4 consecutive X-bar samples > +2 sigma): Found 11 instances of this pattern.
  Starting sample numbers of these runs: 119, 129, 153, 348, 359, 374
```

## Monitor

```
Rule A (s-chart > UCL): Found 0 violations.
Rule B (Good Control): Longest run of s-values within +/- 1 sigma is 34 samples.
  This run occurred from sample 238 to 271 .
Rule C (4 consecutive X-bar samples > +2 sigma): Found 22 instances of this pattern.
  Starting sample numbers of these runs: 134, 171, 179, 566, 610, 615
```

## Mouse

```
Rule A (s-chart > UCL): Found 1 violations.
  Sample numbers: 592
Rule B (Good Control): Longest run of s-values within +/- 1 sigma is 16 samples.
  This run occurred from sample 672 to 687 .
Rule C (4 consecutive X-bar samples > +2 sigma): Found 25 instances of this pattern.
  Starting sample numbers of these runs: 194, 233, 249, 768, 777, 807
```

## Software

```
Rule A (s-chart > UCL): Found 0 violations.
Rule B (Good Control): Longest run of s-values within +/- 1 sigma is 21 samples.
  This run occurred from sample 659 to 679 .
Rule C (4 consecutive X-bar samples > +2 sigma): Found 27 instances of this pattern.
  Starting sample numbers of these runs: 133, 202, 237, 774, 803, 842
```

One product, Mouse, had a s-value above the UCL. An out-of-control sample in the s-chart needs to be addressed in real life. If the process variation is out of control, the control limits on the corresponding $\bar{X}$ chart are meaningless because the control limits are calculated directly from the average variation shown in the s-chart, as mentioned earlier. This could lead to the process being out of control without the operator's knowledge. The operator can stop the process and investigate the issue.

# 5. Risk, Data correction and optimizing for maximum profit (Part 4)

## 5.1 Likelihood of Type I Errors (4.1)

Type I Errors occur when processes are perfectly in control, but the process is stopped because of an unstable signal that was wrong. It is a false alarm that tells you to investigate a problem that doesn't exist.

### Rule A: 1 s sample outside of the +3 sigma control limits

Since the 3-sigma control limits contain 99.73% of all sample statistics if they are in control, the probability of a sample falling outside the 3-sigma limits is *1 – 0.9973 = 0.0027*

The rule only mentions the UCL, so it is one-sided. So, the probability should be halved, and the Likelihood of a Type I error for Rule A is

$$\mathrm{P(Type\ I\ Error\ for\ Rule\ A)} = \frac{0.0027}{2} = 0.00135$$

### Rule B: Most consecutive samples of s between -1 and 1 sigma control limits

This rule is more of a descriptive measure of periods of stability and good control, not an alarm rule. A Type I error is a false signal of a problem, so Rule B does not have an associated Type I error.

### Rule C: 4 consecutive X-bar samples outside +2-sigma

2-sigma control limits contain 95,45% of all samples, so the probability of a sample falling outside 2-sigma is *1 – 0.9545 = 0.0455*

Again, the rule only applies to the upper limit, we take half of the probability. Since each sample is an independent event, probabilities are multiplied and the likelihood of a Type I error for Rule B becomes

$$P(Type\ I\ Error\ for\ Rule\ C) = (P_{>2\sigma})^4 = (0.02275)^4 = 2.76 \times 10^{-7}$$

## 5.2 Likelihood of Type II Errors (4.2)

A Type II errors occur when the process has gone out of control, but the sample happens to fall inside the original control limits, so we fail to detect the problem and conclude that the process is in control. Ha is true but we do not reject $H_0$.

The given parameters are:

- Old Control Limits ($H_0$):
  - UCL = 25.089 litres
  - LCL = 25.011 litres
  - CL = 25.05 litres
  - $\sigma_{\bar{x},old}$ = 0.013
- Shifted Process ($H_a$):
  - $\mu_{new}$ = 25.028
  - X-bar Standard Deviation ($\sigma_{\bar{x},new}$) = 0.017

Probability of a Type II error $\beta$ is

$$\beta = P(LCL < X < UCL)$$

$$\beta = P(X < 25.089) - P(X < 25.011)$$

With the UCL

$$P(X < 25.089)$$

$$Z = \frac{X - \mu}{\sigma} = \frac{25.089 - 25.028}{0.017} = 3.58824$$

$$P(Z < 3.58824) = 0.999834$$

$$= 99.98\%$$

With the LCL

$$P(X < 25.011)$$

$$Z = \frac{X - \mu}{\sigma} = \frac{25.011 - 25.028}{0.017} = -1$$

$$P(Z < -1) = 0.158655$$

$$= 15.87\%$$

Final Probability

$$\beta = 0.999834 - 0.158655 = 0.841179$$

$$= 84.12\ \%$$

An 84% probability indicates that the current control chart is very ineffective at detecting a shift in the process mean. Bottles can constantly be under-filled on average and there is a high chance that the sample will still fall within the control limits, even though the process is out of control and corrective action needs to be taken. This high probability signals a high risk of unknowingly producing a large volume of out-of-spec products, as the system designed to catch such a problem is likely to miss it.

## 6. Coffee Shop Profit Optimization (Part 5)

A profit optimization model was run on two coffee shops to find the optimal number of baristas per weekday to maximize profit.

To ensure consistency, the following assumptions were made:

- Demand is constant, there is always a customer ready to be served.
- Barista's work 8-hour shifts, 7 days a week.
- The same number of barista's work for the entirety of the shift.
- The model assumes that the average service time calculated for each barista group is a reliable measure of that group's efficiency to prevent groups with more recorded transactions from seeming more profitable simply because they have more data.
- The definition of reliable service was assumed to be 90 seconds or less.

The model works by using a brute force methodology to determine the most profitable number of baristas. It begins by establishing financial and operational constants such as the profit per customer and daily cost per barista. The core of the model is to evaluate each possible staffing level found in the dataset. It normalizes performance by calculating the average service time for each distinct group of baristas to ensure a fair comparison. This average time acts as a benchmark for that group's efficiency.

Using the efficiency benchmark, the model simulates a standard 8-hour workday for each level, calculating how many customers they could theoretically serve if there was no idle time. It calculates the daily revenue based on the customer throughput and subtracts the total daily personnel cost for that group to find the daily profit. The daily profit is then scaled into a weekly profit. The optimization is achieved by comparing the final weekly profit values for all levels and selecting the number of baristas that yield the highest profit. A secondary analysis is also performed to assess the percentage of customers who receive reliable service, assumed to be 90 seconds, under the optimal barista level. The profit model tables for both shops can be found in Appendix 1D.

## Coffee Shop 1

The model concluded that to maximize profit, the optimal barista level is 6, producing a profit of R139319. The graph below depicts a story of how the number of baristas working is directly proportional to the expected weekly profit. This is because with more workers, the average service time is much faster, and the coffee shop can increase its weekly throughput as more customers are being served and contribute R30 profit. With the optimal number, 863 customers are served per day. This vastly outweighs the cost of operating 6 baristas per day. Once scaled to weeks, it is evident that a 6-barista schedule would produce the highest profit.

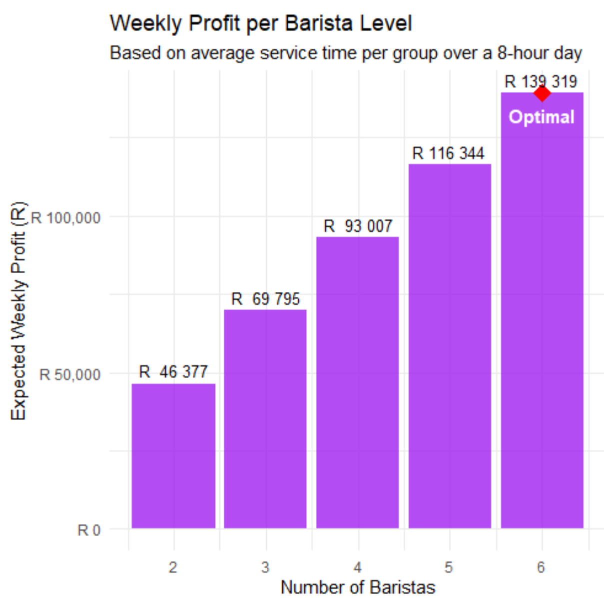



*Figure 25: Shop 1 Weekly Profit Graph*

Under this schedule, service reliability achieved 100%, with all 97895 customers served within 90 seconds. This is a promising figure as it decreases the cost of poor quality due to lost opportunities for sales revenue, ensuring all customers receive attention reliably. Achieving 100% reliability would also keep customer satisfaction high.

```
Service Reliability:
 Baristas: 6
 Reliable service: 100.00%
 Service Threshold: 90
 Customers served: 97895.00
```

*Figure 26: Shop 1 Service Reliability*

## Coffee Shop 2

A model analysis on shop 2 revealed that the optimal number of baristas needed to maximize profit is 5, producing a profit of R32623.80 per week and serving 352 customers per day. Unlike shop 1, the relationship of profit vs number of baristas shows somewhat of a parabolic curve, where profits maximize at 5 baristas, and the addition of more personnel will result in less profits as seen when 6 baristas are on duty. The chart was visualized as a line graph to see the extent of this.  It is a common misconception that more workers would bring more money, and although it is true for the first scenario, operations differ for many businesses, and incorporating the same strategy across the board would lead to inefficiency.

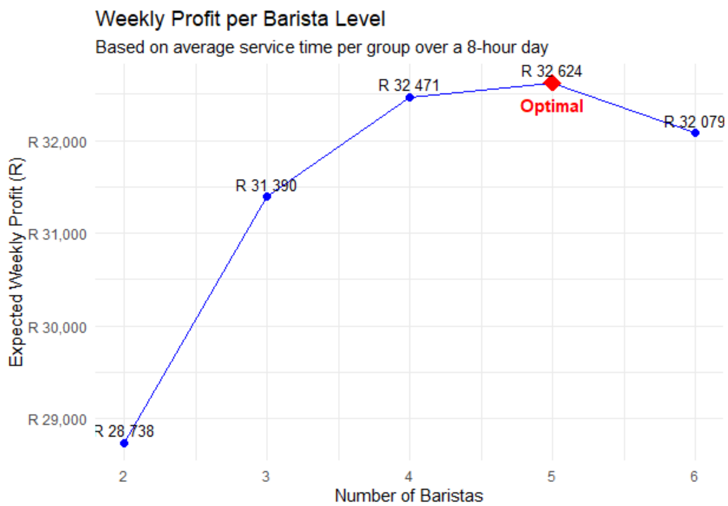



*Figure 27:Shop 2 Weekly Profit Graph*

The average service times for this shop is significantly more than that of shop 1, so it was assumed that reliable service would sit at a threshold of 95 seconds as opposed to the previous assumption. Under this threshold, the service reliability is 88.89%. This means that about 88 out of 100 customers would receive their orders in under 95 seconds, this amounts to 48853 customers.

```
Service Reliability:
 Baristas: 5
 Reliable service: 88.89%
 Service Threshold: 95s
 Customers served: 48853.00
```

*Figure 28: Shop 2 Service Reliability*

This model proves its robustness in how it factors in the profit from customers served, as well as the operational cost of personnel into the equation to find the optimal configuration under the assumption of constant demand for different datasets. In a situation where demand fluctuates, personnel should be assigned to meet the demand of the customers during that season but not incur unnecessary costs from scheduling more workers than required. This model shows similarity to the Taguchi Loss Function, that shows that any deviation from the optimal setting results in increased loss. It shows that operating with too few or too many workers leads to inefficiency and a rise in expected costs, so staffing levels should be kept close to the optimal point to maximize profitability.

# 7. DOE and ANOVA (Part 6)

The results of the SPC were analyzed and the chosen product type for further analysis is Keyboard. This product type is chosen because its s-chart shows stable variability, as opposed to a product like Mouse, that has a point above the UCL and is a poor candidate for ANOVA. We will use a confidence level of 95% for a two-tailed test. ($\alpha = 0.05 \rightarrow \frac{\alpha}{2} = 0.025$)

$H_0$: There is no significant difference in the mean delivery hours among the 12 months of the year for Keyboard. All months have the same average delivery time.

$H_1$: There is a significant difference in the mean delivery hours of at least one month for Keyboard.

## ANOVA and LSD Results:

As seen above, the ANOVA table produces a p-value of 0, indicating a low probability of making an error when rejecting $H_0$. Although the p value is not truly zero, it is much smaller than the chosen alpha value of 0.05, therefore we reject the null hypothesis. The month of the year has a significant effect on the average delivery hours. The average delivery times are not the same across all months.

```
"Totals of every row:"
 17890.63 18104.62 19133.63 19817.60 19805.61 20389.88 21066.18 21189.90 21716.62 22222.62 22882.23 22938.73
"Means of every row:"
 18.87197 19.09770 20.18315 20.90465 20.89200 21.50831 22.22171 22.35221 22.90783 23.44158 24.13737 24.19697
```

```
"Results in a standard ANOVA table. p-values => Type I errors"
 [,1]        [,2]        [,3]    [,4]      [,5]    [,6]
 "Source"    "SS"        "DoF"   "MS"      "fo"    "P-value"
 "Treatment" "33890.65"  "11"    "3080.97" "90.47" "0"
 "Error"     "386989.39" "11364" "34.05"   "---"   "---"
 "Total"     "420880.04" "11375" "---"     "---"   "---"
```

*Figure 29: ANOVA Table for Keyboard DeliveryHours*

The ANOVA tells us that there is a difference, and the LSD tells us which specific months are different from each other. Here, the LSD value is 0.525, meaning any difference between the average delivery hours that is greater than that is considered a statistically significant difference.

The second table takes the absolute differences of every pair and divides it by the LSD, a value greater than 0 indicates a significant difference.

Comparing months 1 and 3: Since the value is 2.4956 and is greater than 0, January delivery times are significantly different from March.
Comparing months 1 and 12: The value is 10.1352, which indicates a large significant difference between the start and the end of the year.
Comparing months 11 and 12: The value is 0, indicating that although the averages are slightly different, the difference is not statistically significant.

```
LSD value is:  0.525398822298403


Differences between averages are (larger are more significant):

      [,1] [,2]     [,3]     [,4]     [,5]     [,6]     [,7]     [,8]     [,9]     [,10]    [,11]    [,12]   
 [1,] "0"  "0.2257" "1.3112" "2.0327" "2.02"   "2.6363" "3.3497" "3.4802" "4.0359" "4.5696" "5.2654" "5.325" 
 [2,] "0"  "0"      "1.0855" "1.8069" "1.7943" "2.4106" "3.124"  "3.2545" "3.8101" "4.3439" "5.0397" "5.0993"
 [3,] "0"  "0"      "0"      "0.7215" "0.7088" "1.3252" "2.0386" "2.1691" "2.7247" "3.2584" "3.9542" "4.0138"
 [4,] "0"  "0"      "0"      "0"      "0.0126" "0.6037" "1.3171" "1.4476" "2.0032" "2.5369" "3.2327" "3.2923"
 [5,] "0"  "0"      "0"      "0"      "0"      "0.6163" "1.3297" "1.4602" "2.0158" "2.5496" "3.2454" "3.305" 
 [6,] "0"  "0"      "0"      "0"      "0"      "0"      "0.7134" "0.8439" "1.3995" "1.9333" "2.6291" "2.6887"
 [7,] "0"  "0"      "0"      "0"      "0"      "0"      "0"      "0.1305" "0.6861" "1.2199" "1.9157" "1.9753"
 [8,] "0"  "0"      "0"      "0"      "0"      "0"      "0"      "0"      "0.5556" "1.0894" "1.7852" "1.8448"
 [9,] "0"  "0"      "0"      "0"      "0"      "0"      "0"      "0"      "0"      "0.5338" "1.2295" "1.2891"
[10,] "0"  "0"      "0"      "0"      "0"      "0"      "0"      "0"      "0"      "0"      "0.6958" "0.7554"
[11,] "0"  "0"      "0"      "0"      "0"      "0"      "0"      "0"      "0"      "0"      "0"      "0.0596"
[12,] "0"  "0"      "0"      "0"      "0"      "0"      "0"      "0"      "0"      "0"      "0"      "0"     


Significant differences are:

      [,1] [,2] [,3]     [,4]     [,5]     [,6]     [,7]     [,8]     [,9]     [,10]    [,11]     [,12]    
 [1,] "0"  "0"  "2.4956" "3.8688" "3.8448" "5.0178" "6.3756" "6.624"  "7.6815" "8.6974" "10.0217" "10.1352"
 [2,] "0"  "0"  "2.066"  "3.4392" "3.4151" "4.5882" "5.946"  "6.1944" "7.2519" "8.2678" "9.5921"  "9.7055" 
 [3,] "0"  "0"  "0"      "1.3732" "1.3492" "2.5222" "3.88"   "4.1284" "5.1859" "6.2018" "7.5261"  "7.6396" 
 [4,] "0"  "0"  "0"      "0"      "0"      "1.149"  "2.5068" "2.7552" "3.8127" "4.8286" "6.1529"  "6.2663" 
 [5,] "0"  "0"  "0"      "0"      "0"      "1.173"  "2.5309" "2.7792" "3.8368" "4.8527" "6.177"   "6.2904" 
 [6,] "0"  "0"  "0"      "0"      "0"      "0"      "1.3578" "1.6062" "2.6637" "3.6796" "5.0039"  "5.1174" 
 [7,] "0"  "0"  "0"      "0"      "0"      "0"      "0"      "0"      "1.3059" "2.3218" "3.6461"  "3.7595" 
 [8,] "0"  "0"  "0"      "0"      "0"      "0"      "0"      "0"      "1.0575" "2.0734" "3.3977"  "3.5112" 
 [9,] "0"  "0"  "0"      "0"      "0"      "0"      "0"      "0"      "0"      "1.0159" "2.3402"  "2.4537" 
[10,] "0"  "0"  "0"      "0"      "0"      "0"      "0"      "0"      "0"      "0"      "1.3243"  "1.4378" 
[11,] "0"  "0"  "0"      "0"      "0"      "0"      "0"      "0"      "0"      "0"      "0"       "0"      
[12,] "0"  "0"  "0"      "0"      "0"      "0"      "0"      "0"      "0"      "0"      "0"       "0"      
```

*Figure 30: LSD Test for Keyboard DeliveryHours*

The means show a consistent upward trend throughout the year, starting at 18.87 hours and increasing to 24.20 hours in December. The LSD confirms this trend by showing that the earlier, faster months are significantly different to the delivery times of the later months.

This leads to the conclusion that there is a significant seasonal effect on delivery times for Keyboard. A trend reveals that the delivery times get progressively worse as the year goes by. It is not just a few random months that are bad, there is an issue causing performance and reliability to degrade as the year progresses. The cause of this issue needs to be investigated to solve the problem in the future and keep delivery times as consistent as possible.

# 8. Reliability of Service (Part 7)

## Reliable Service Estimate (7.1)

To estimate how many days per year the company can expect reliable service, assumed to be 15. The attendance data was modelled by fitting a binomial distribution with a maximum of 16 workers and an estimated probability *p* of attendance on that day.

$$P(X = k) = \binom{16}{k} p^k (1 - p)^{16-k}$$

Individual p-values were attained at each level of workers by finding the value that makes the binomial equal to the empirical, then the weighted mean p value was calculated, giving p = 0.9236363. Using the fitted probability, p, we find

$$P(X \geq 15) = P(X = 15) + P(X = 16)$$

$$= 0.3711264 + 0.2805536$$

$$= 0.65168$$

Multiplying the probability by a year gives 237 days of reliable service per year.

## Car Rental Profit Optimization (7.2)

A profit optimization model was run on the car rental company to find the optimal staffing level to maximize annual profit. The data was fitted to a binomial distribution and for each potential staffing level, the probability of having at least 15 workers on duty was used to calculate the expected cost together with the fixed cost of the staffing level. The lowest expected cost is chosen as the optimal solution. The optimal number of workers to contribute the lowest expected cost of R5 719 648, is 18 workers. This configuration would reduce income the least and therefore contribute to higher profit.

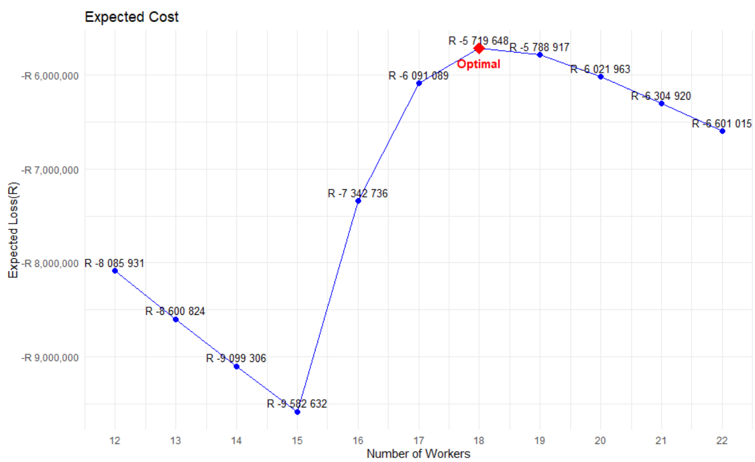



*Figure 31: Expected Cost Graph*

# 9. References


Burger, E. (2024) *R for Data Analytics.* [pdf]

RStudio, Inc. (2015) *Data Wrangling with dplyr and tidyr Cheat Sheet*. [pdf] RStudio.

RStudio, Inc. (2015) *Data Visualization with ggplit2 Cheat Sheet.* [pdf] RStudio.

# Appendices

## Appendix 1A: Initialized Charts

### Cloud Subscription

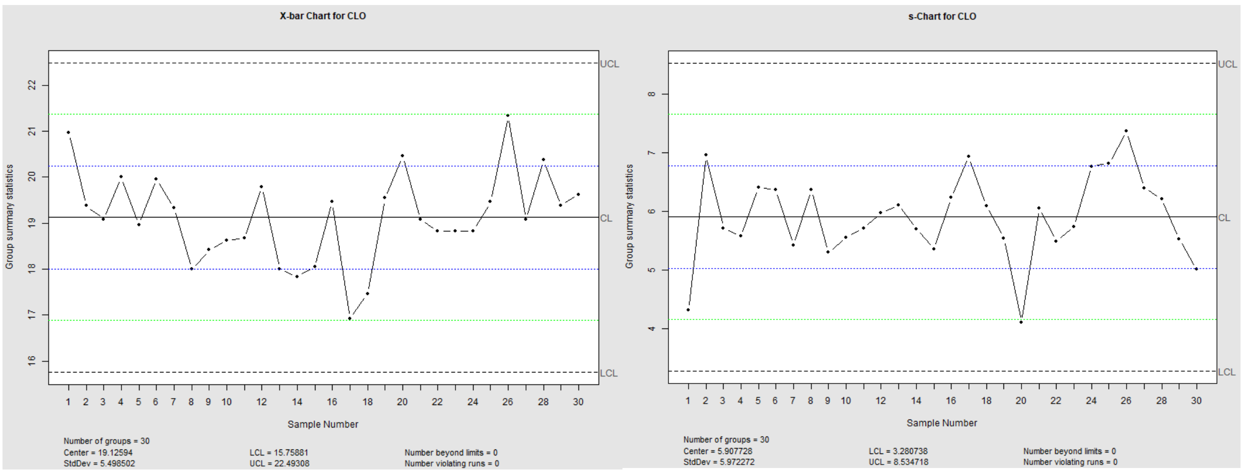



*Figure 32: X-bar and S Chart for Cloud Subscription*

### Keyboard

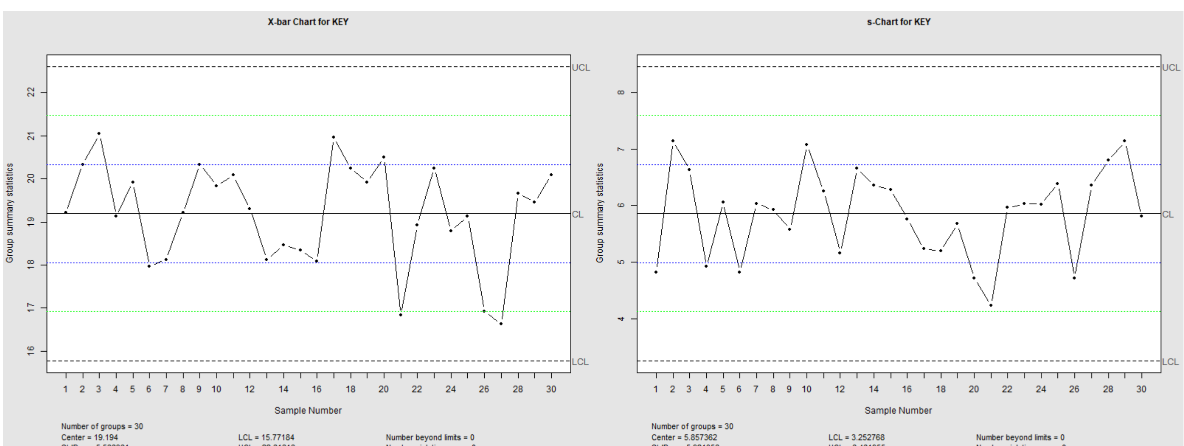



*Figure 33: X-bar and S Chart for Keyboard*

## Monitor

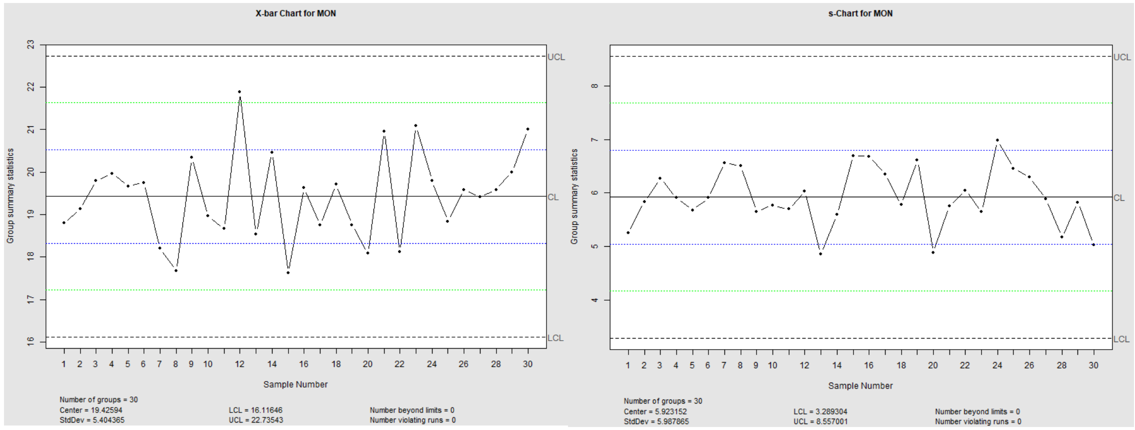



*Figure 34: X-bar and S Chart for Monitor*

## Mouse

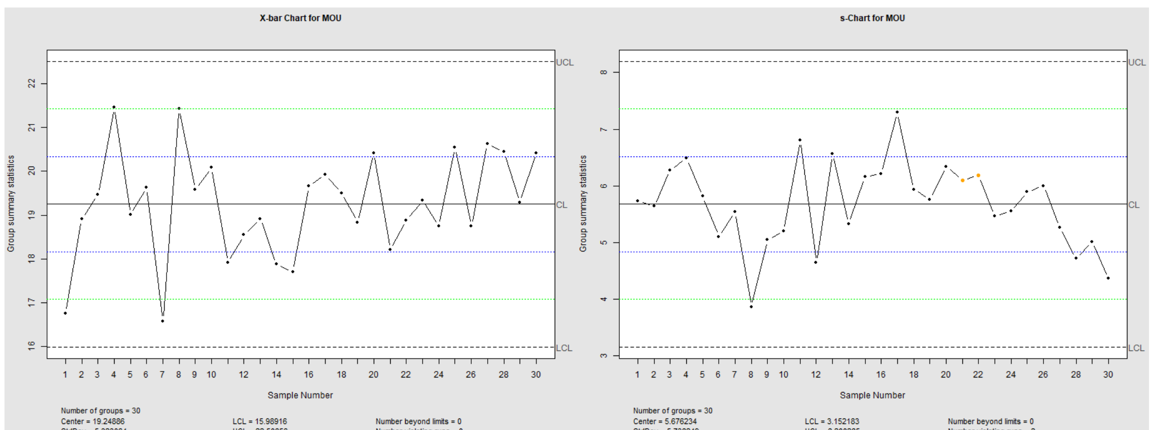



*Figure 35: X-bar and S Chart for Mouse*

## Software

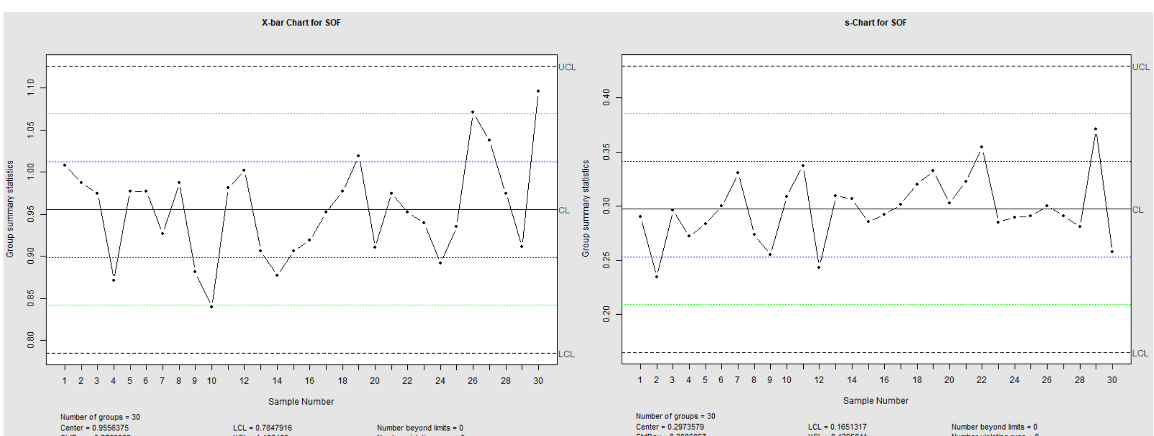



*Figure 36: X-bar and S Chart for Software*

# Appendix 1B: Accelerated Simulation

## Cloud Subscription

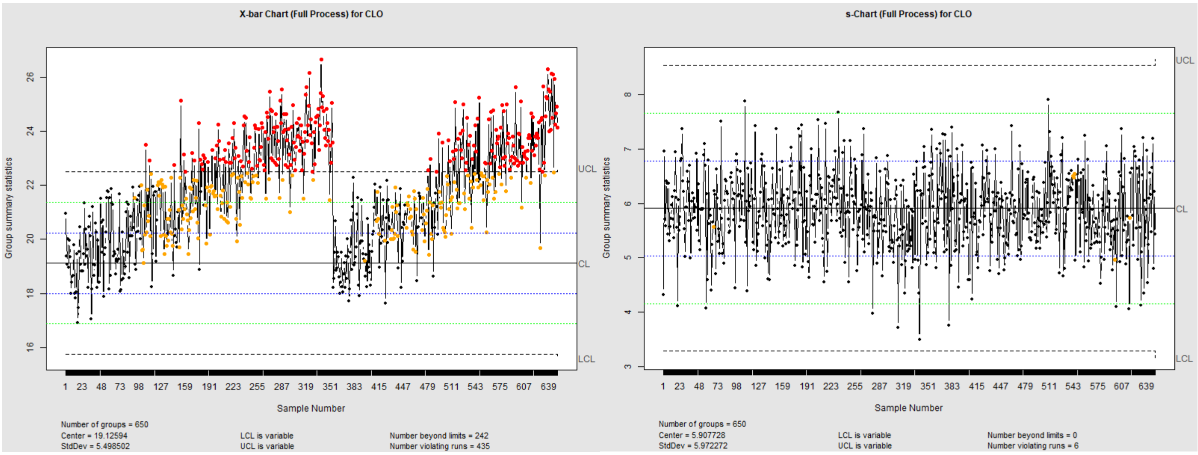



*Figure 37: Full X-bar and S Chart for Cloud Subscription*

## Keyboard

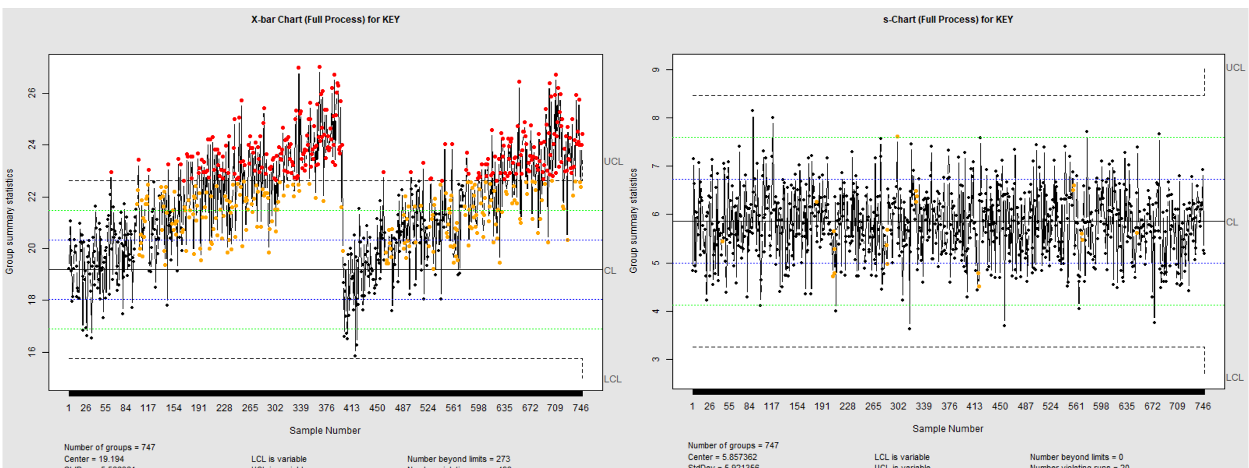



*Figure 38: Full X-bar and S Chart for Keyboard*

## Monitor

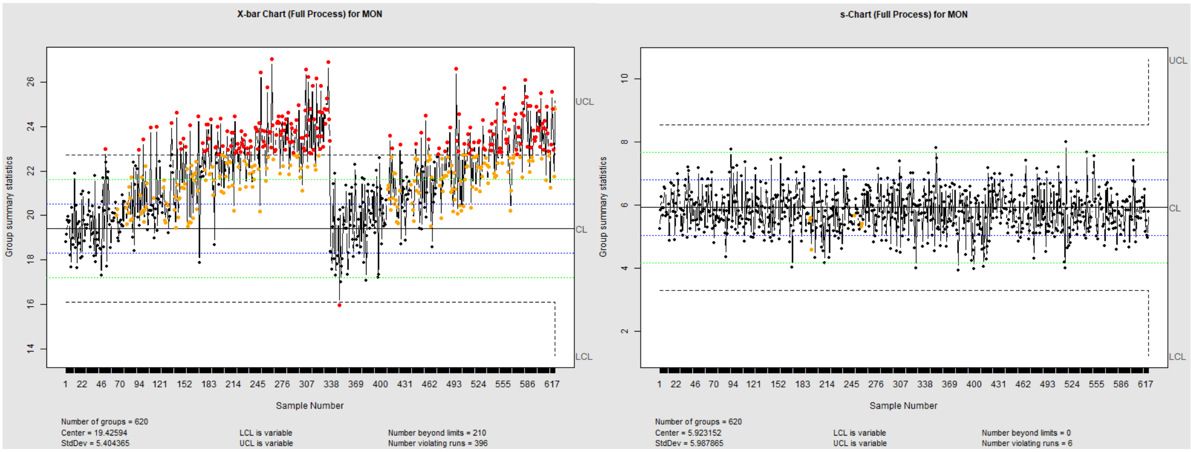



*Figure 39: Full X-bar and S Chart for Monitor*

## Mouse

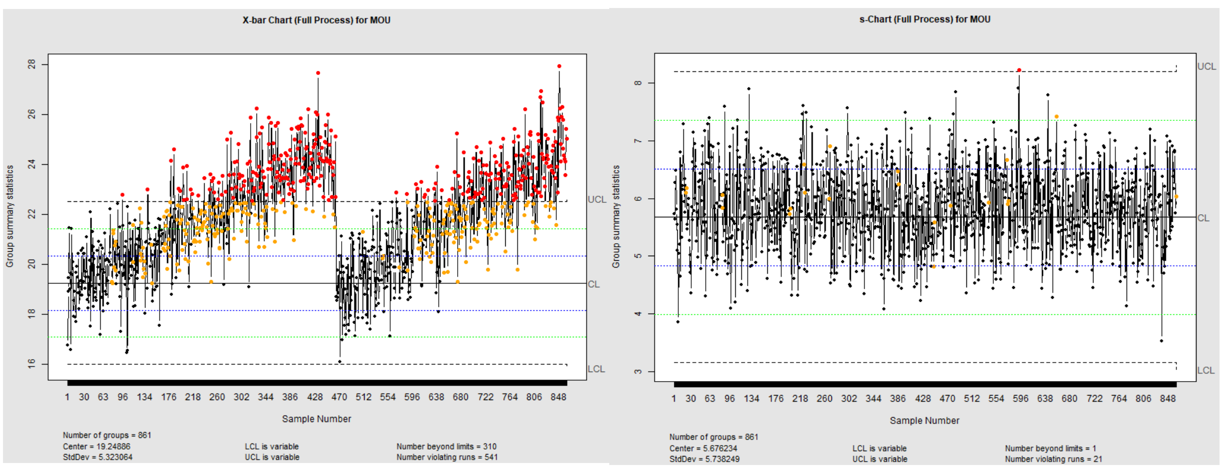



*Figure 40: Full X-bar and S Chart for Mouse*

## Software

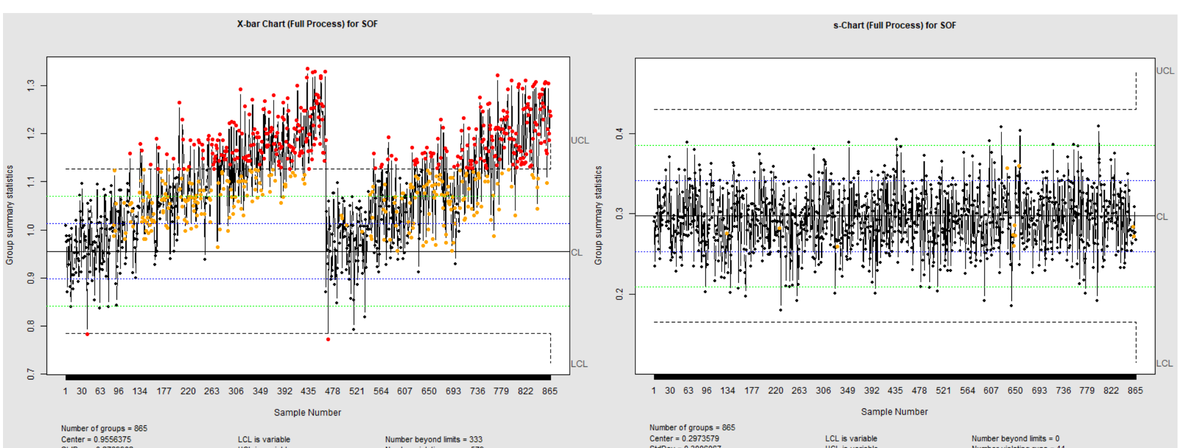



*Figure 41: Full X-bar and S Chart for Software*

# Appendix 1C: Process Capability

## Keyboard

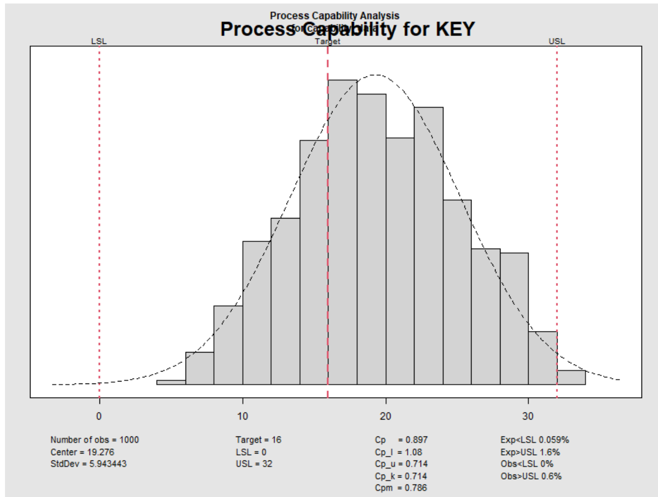



*Figure 42: Process Capability Analysis for Keyboard*

## Laptop

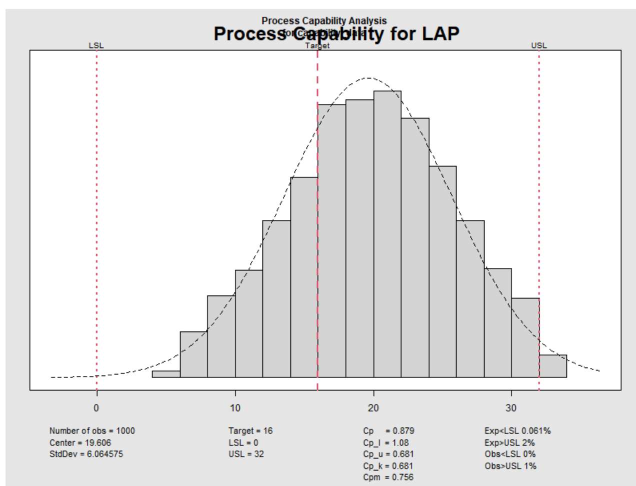



*Figure 43: Process Capability Analysis for Laptop*

## Monitor

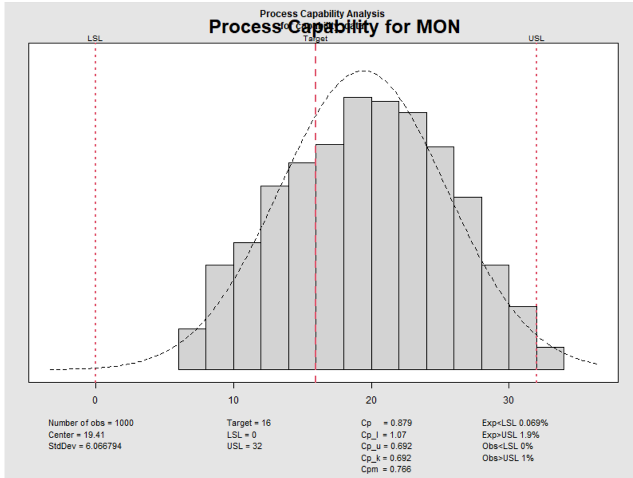



*Figure 44: Process Capability Analysis for Monitor*

## Mouse

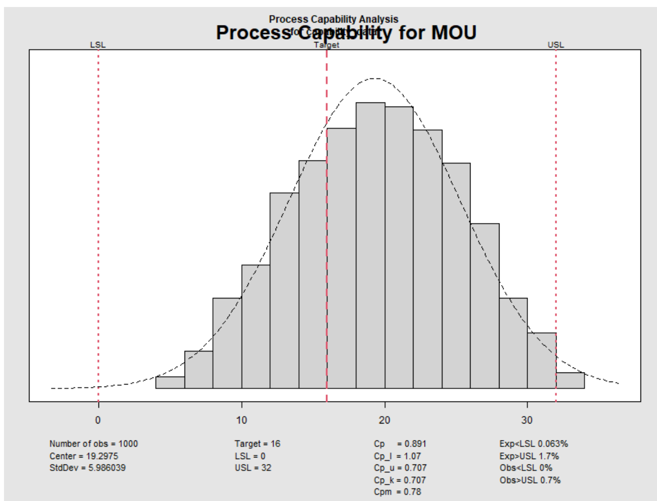



*Figure 45: Process Capability Analysis for Mouse*

## Appendix 1D: Profit Optimization (Part 5)

| NumBarista | AvgServiceTime | CustomersPerDay | DailyRevenue | DailyCost | DailyProfit | WeeklyProfit |
|---|---|---|---|---|---|---|
| 2 | 100.17098 | 287.5084 | 8625.253 | 2000 | 6625.253 | 46376.77 |
| 3 | 66.61174 | 432.3562 | 12970.686 | 3000 | 9970.686 | 69794.80 |
| 4 | 49.98038 | 576.2261 | 17286.784 | 4000 | 13286.784 | 93007.49 |
| 5 | 39.96183 | 720.6876 | 21620.629 | 5000 | 16620.629 | 116344.40 |
| 6 | 33.35565 | 863.4220 | 25902.661 | 6000 | 19902.661 | 139318.63 |

*Figure 46: Shop 1 Profit Model*

| | NumBarista | AvgServiceTime | CustomersPerDay | DailyRevenue | DailyCost | DailyProfit | WeeklyProfit |
|---|---|---|---|---|---|---|---|
| **1** | 2 | 141.51462 | 203.5125 | 6105.376 | 2000 | 4105.376 | 28737.63 |
| **2** | 3 | 115.44091 | 249.4783 | 7484.348 | 3000 | 4484.348 | 31390.44 |
| **3** | 4 | 100.01527 | 287.9560 | 8638.681 | 4000 | 4638.681 | 32470.76 |
| **4** | 5 | 89.43597 | 322.0181 | 9660.543 | 5000 | 4660.543 | 32623.80 |
| **5** | 6 | 81.64272 | 352.7565 | 10582.695 | 6000 | 4582.695 | 32078.86 |

*Figure 47: Shop 2 Profit Model*