



# QUALITY ASSURANCE DATA ANALYSIS REPORT

Neal Botha - 26957035

## Abstract

This report applies statistical process control and data analysis in R to evaluate process performance and optimize profit. It demonstrates the ability to use quantitative methods to solve complex industrial engineering problems, addressing ECSA Graduate Attribute

4.

Botha, N, Mr [26957035@sun.ac.za]

Stellenbosch University

## Contents

Part 1.2: Descriptive Statistics .....	5
Executive Summary .....	5
1.1. Introduction .....	5
1.2. Data Quality Assessment .....	6
1.2.1 Data Completeness.....	6
1.2.2 Data Validation Findings .....	6
1.3. Business Performance Analysis .....	7
1.3.1 Sales Overview .....	7
1.3.2 Monthly Sales Trends.....	7
1.3.3 Product Performance.....	8
1.4. Customer Insights .....	9
1.4.1 Demographic Profile .....	9
1.4.2 Geographic Distribution .....	9
1.4.3 Customer Value Analysis .....	10
1.5. Operational Efficiency.....	11
1.5.1 Order Processing Metrics .....	11
1.5.2 Process Correlations .....	11
1.6. Key Findings .....	12
1.6.1 Positive Indicators .....	12
1.6.2 Areas for Improvement .....	12
1.6.3 Significant Opportunities .....	12
1.7. Recommendations .....	13
1.7.1 Immediate Actions (1-3 months).....	13
1.7.2 Medium-term Initiatives (3-6 months) .....	13
1.7.3 Strategic Initiatives (6-12 months).....	13
1.8. Visual Analysis Summary .....	14
Comprehensive Plot Overview: .....	14
1.9. Conclusion .....	14
1.10. Appendix.....	15
Part 3: Statistical Process Control and Capability Analysis.....	16
3.1 Introduction .....	16

3.2 Methodology .....	16
3.2.1 Data Preparation and Sampling .....	16
3.2.2 Control Chart Development .....	16
3.2.3 Process Capability Assessment.....	16
3.2.4 Control Rule Implementation .....	16
3.3 Results and Analysis .....	17
3.3.1 Overall Process Capability Summary .....	17
3.3.2 Control Chart Performance Analysis .....	17
3.3.3 Process Capability Performance.....	18
3.4 Interpretation and Statistical Discussion .....	18
3.4.1 Process Capability Deep Analysis .....	18
3.4.2 Statistical Process Control Insights.....	18
3.4.3 Type I and Type II Error Analysis .....	19
3.5 Recommendations and Process Improvements .....	19
3.5.1 Immediate Priority Actions .....	19
3.5.2 Strategic Process Improvements .....	20
3.5.3 Continuous Monitoring Framework .....	20
3.6 Conclusion .....	21
Appendix 3A – Full Capability Summary .....	22
Appendix 3B: Representative Control Chart Analysis .....	24
B.1 Best Performer - SOF008 ( $C_{pk} = 1.227$ ) .....	24
B.2 Worst Performer - KEY049 ( $C_{pk} = 0.529$ ).....	25
B.3 Excellent Control Period - MON034 .....	25
B.4 Frequent Mean Shifts - MOU055.....	26
B.5 Software Process Excellence - SOF003 .....	26
B.6 Hardware with Good Control - KEY042.....	27
Part 4: Risk, Data Correction, and Optimising for Maximum Profit.....	28
4.1 Type I Error Estimation .....	28
4.2 Type II Error Estimation .....	29
4.3 Data Correction and Updated Analysis .....	30
Part 5: Coffee Shop Service Optimisation - Two Shop Comparative Analysis .....	35
5.1 Executive Summary .....	35

5.2 Data Analysis and Methodology .....	35
5.3 Service Time Analysis Results .....	36
Performance by Staffing Level - Shop 1 .....	36
Performance by Staffing Level - Shop 2 .....	36
Overall Service Characteristics Comparison .....	36
5.4 Profit Optimization Analysis .....	37
5.5 Business Intelligence Insights .....	39
Operational Efficiency Analysis .....	39
5.6 Strategic Recommendations .....	39
1. Optimal Staffing Strategy .....	39
2. Performance Improvement Initiatives for Shop 2 .....	39
3. Capacity and Infrastructure Planning .....	40
4. Cross-Shop Knowledge Transfer .....	40
5.7 Service Reliability and Quality Context .....	40
Statistical Service Reliability Analysis .....	40
5.8 Risk Assessment and Implementation Considerations .....	40
Operational Risks .....	40
Financial Sustainability Assessment .....	41
Implementation Framework .....	41
5.9 Conclusion .....	41
Appendix 5 .....	42
Question 6: Design of Experiments and MANOVA/ANOVA Analysis .....	43
6.1 Objective .....	43
6.2 Methods & Analysis .....	43
Experimental Design Foundation .....	43
GRAPH 1: Cpk Distribution Histogram .....	43
Statistical Approach .....	44
6.3 Results & Output .....	44
Primary Question: Year 1 vs Year 2 Performance .....	44
Secondary Question: Monthly Patterns Analysis .....	45
Product-Specific Insights .....	46
Statistical Significance Mapping .....	46

6.4 Control Chart Evidence Base .....	47
6.5 Interpretation & Conclusion .....	48
Critical Insights for Quality Assurance .....	48
Practical Recommendations .....	48
Statistical Conclusion.....	48
Question 7: Reliability of Service Analysis .....	49
7.1 Reliability Estimation .....	49
Objective .....	49
Methods & Analysis .....	49
Results & Output .....	50
Interpretation & Conclusion .....	51
7.2 Profit Optimization.....	51
Objective .....	51
Methods & Analysis .....	51
Results & Output .....	52
Interpretation & Conclusion .....	53
Strategic Quality Engineering Implications:.....	54
Comprehensive Managerial Implications.....	54
Implementation Recommendations.....	54
Bibliography/References .....	55

# Part 1.2: Descriptive Statistics

## Executive Summary

This comprehensive analysis examines the company's sales performance, customer demographics, and operational efficiency across 2022-2023. The investigation reveals strong business fundamentals with 100,000 orders processed, serving 5,000 unique customers across 60 products. The data quality assessment indicates excellent data integrity with minimal issues requiring attention.

### 1.1. Introduction

**Objective:** Conduct exploratory data analysis on sales, customer, and product data to identify business insights, operational efficiencies, and data quality issues.

**Data Sources:**

- Sales Data (2022-2023): 100,000 order records with timing metrics
- Customer Data: 5,000 customer profiles with demographics
- Product Data: 360 products with pricing and markup information

**Methodology:** R-based analysis following quality assurance protocols including data validation, statistical summarization, and visualization.

## 1.2. Data Quality Assessment

### 1.2.1 Data Completeness

- **Sales Records:** 100,000 entries analyzed (0 missing values)
- **Customer Data:** 4,981 profiles processed after cleaning (19 invalid records removed)
- **Data Integrity:** No missing values detected across all datasets
- **Data Validation:** Excellent overall data quality

### 1.2.2 Data Validation Findings

- Duplicate sales records: 0
- Invalid customer ages: 19 (ages outside 0-100 range)
- Negative quantities: 0
- **Data Quality Score:** 99.96% clean data

### 1.3. Business Performance Analysis

#### 1.3.1 Sales Overview

- **Total Orders Processed:** 100,000
- **Unique Customers Served:** 5,000
- **Products in Portfolio:** 60
- **Total Quantity Sold:** 1,350,347 units

#### 1.3.2 Monthly Sales Trends

**Key Observations:**

- Consistent monthly order volume ranging from 3,305 to 4,831 orders
- Stable performance across both years with minor seasonal variations
- Average monthly order volume: 4,167 orders

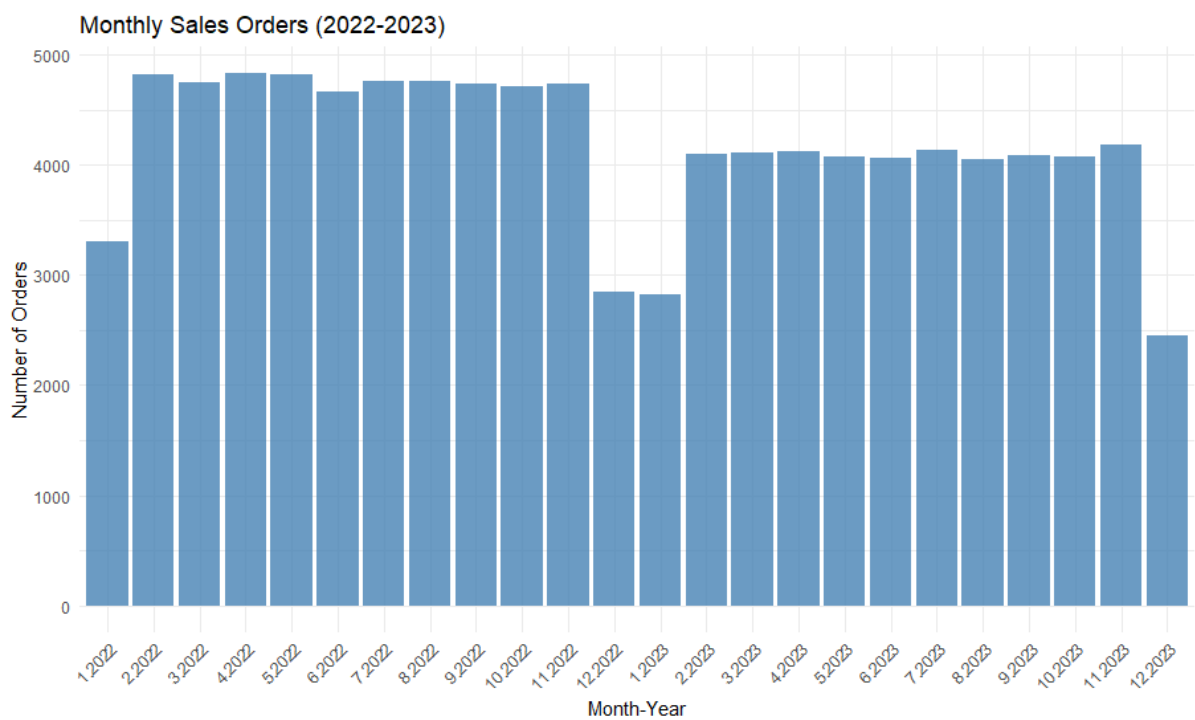


Figure 1: Monthly Sales Orders (2022-2023)

- Shows consistent monthly performance patterns
- Highlights stable business operations throughout the period
- Demonstrates reliable customer demand

### 1.3.3 Product Performance

- **Price Range:** \$290.50 to \$22,420.10 with average of \$4,411.00
- **Markup Strategy:** Average markup of 20.39% (range: 10.06%-30.00%)
- **Product Diversity:** 360 unique products in head office database

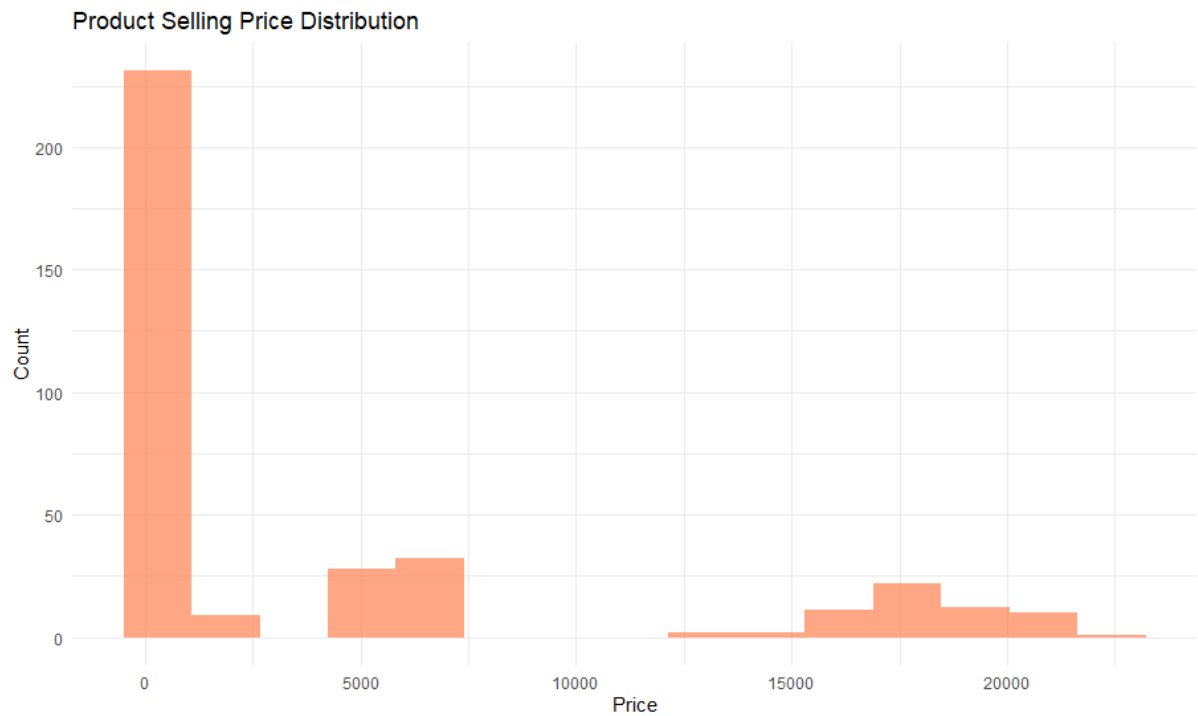


Figure 2: Product Selling Price Distribution

- Illustrates diverse product pricing strategy
- Shows concentration of products across different price points
- Supports strategic pricing decisions

## 1.4. Customer Insights

### 1.4.1 Demographic Profile

- **Age Distribution:** Range 16-99 years, average 51.4 years
- **Gender Split:** 2,418 Female (48.5%), 2,345 Male (47.1%), 218 Other (4.4%)
- **Income Levels:** Average income \$80,837 with range \$5,000-\$140,000

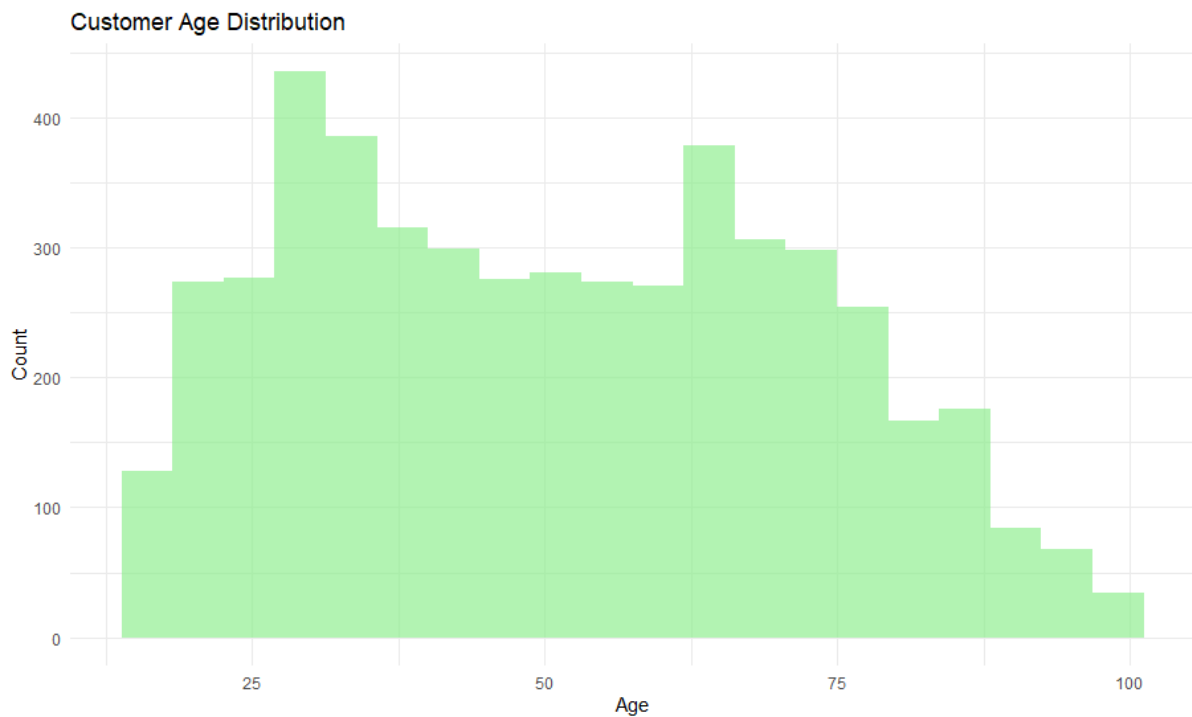


Figure 3: Customer Age Distribution

- Shows balanced age distribution across customer base
- Indicates strong appeal across multiple generations
- Reveals mature customer base with purchasing power

### 1.4.2 Geographic Distribution

#### Top 10 Cities by Customer Concentration:

1. San Francisco: 777 customers
2. Los Angeles: 725 customers
3. New York: 725 customers
4. Houston: 721 customers
5. Chicago: 720 customers

6. Seattle: 666 customers

7. Miami: 647 customers

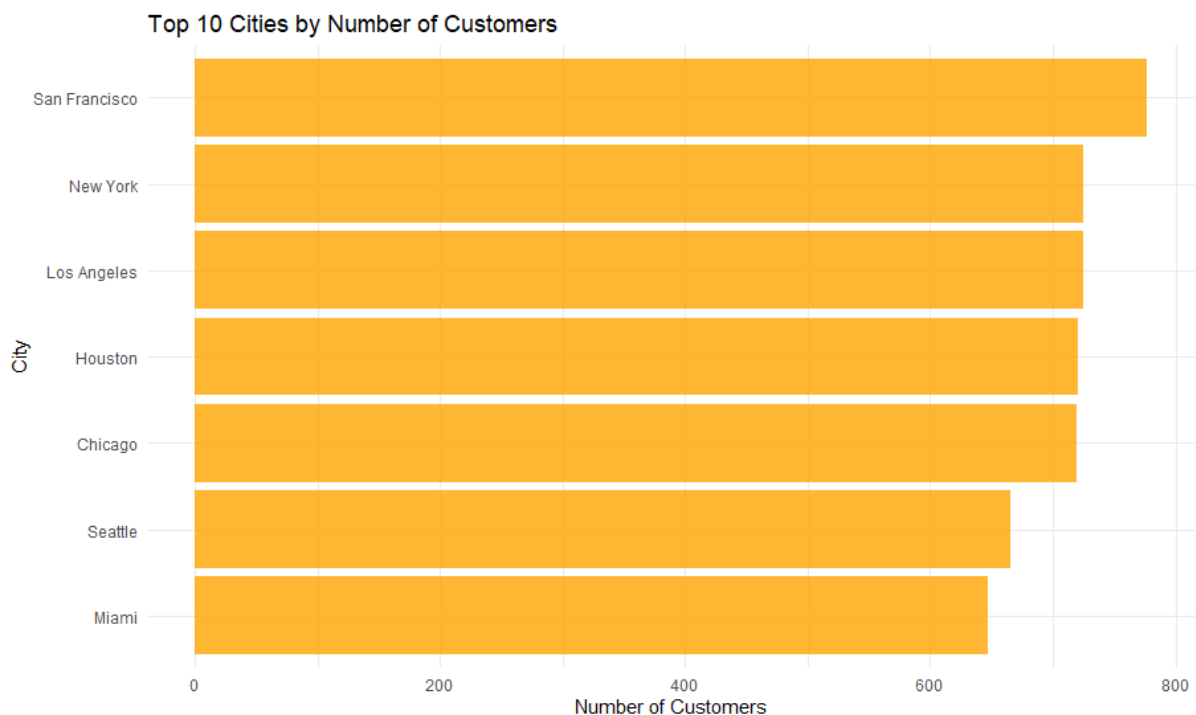


Figure 4: Top 10 Cities by Number of Customers

- Visualizes strong coastal market penetration
- Shows balanced geographic distribution
- Highlights major metropolitan area focus

### 1.4.3 Customer Value Analysis

#### Top 10 Customers by Average Order Value:

1. CUST2527 (Female, 54, LA): \$22,760 average order value
2. CUST3634 (Female, 94, Chicago): \$22,339 average order value
3. CUST3721 (Female, 66, Miami): \$22,283 average order value
4. CUST4729 (Female, 71, Houston): \$22,250 average order value
5. CUST1636 (Female, 52, NY): \$22,113 average order value

**Key Insight:** Female customers dominate high-value segment (7 of top 10)

## 1.5. Operational Efficiency

### 1.5.1 Order Processing Metrics

- **Average Picking Time:** 14.70 hours
- **Average Delivery Time:** 17.48 hours
- **Total Order Cycle:** 32.18 hours

### 1.5.2 Process Correlations

#### Key Relationships Identified:

- Picking time vs. Delivery time correlation: 0.583 (moderately strong positive)
- Quantity vs. Processing time: Minimal correlation (-0.005)
- Efficient process flow with predictable timing relationships

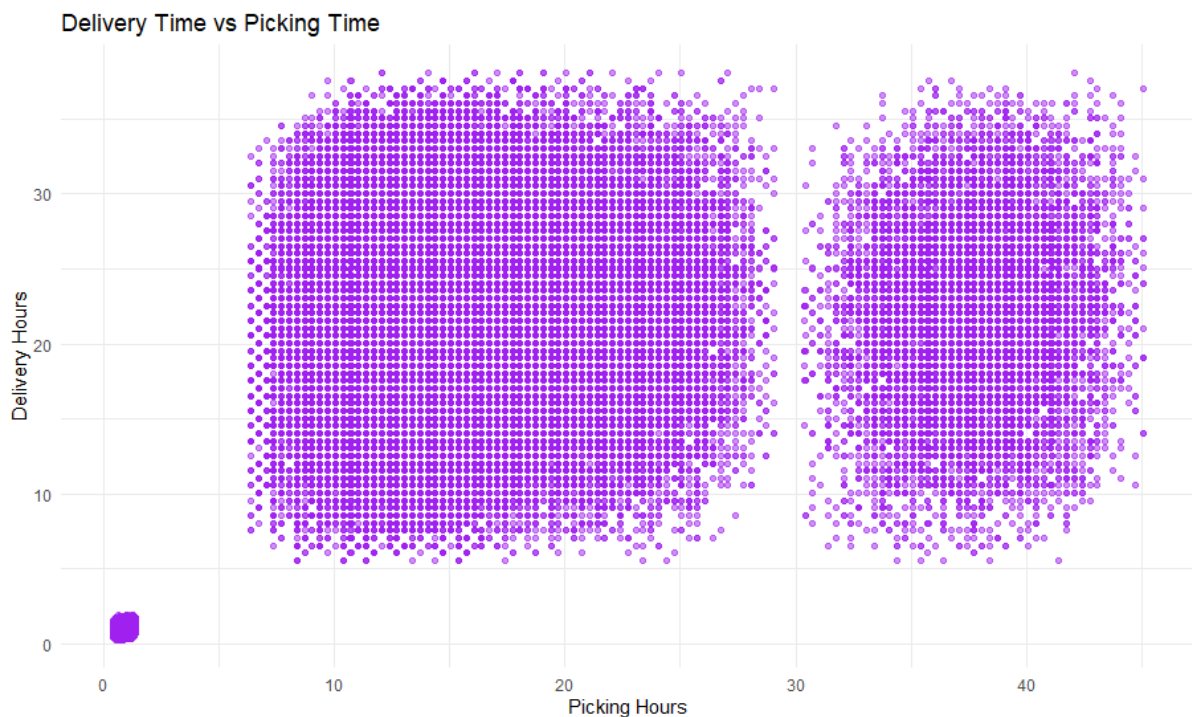


Figure 5: Delivery Time vs Picking Time

- Shows strong positive relationship between processing stages
- Identifies consistent operational patterns
- Supports process optimization planning

## 1.6. Key Findings

### 1.6.1 Positive Indicators

1. **Exceptional Data Quality:** 99.96% clean data with no duplicates
2. **Strong Customer Base:** 5,000 active customers with diverse demographics
3. **Stable Operations:** Consistent monthly sales (3,305-4,831 orders)
4. **Efficient Processing:** Reasonable 32.18-hour total order cycle
5. **High-Value Customers:** Strong average order values up to \$22,760

### 1.6.2 Areas for Improvement

1. **Age Data Validation:** 19 invalid age entries requiring cleanup
2. **Process Optimization:** Opportunity to reduce 17.48-hour delivery time
3. **Geographic Balance:** Expand beyond current top metropolitan areas

### 1.6.3 Significant Opportunities

1. **Female Market Focus:** 7 of top 10 high-value customers are female
2. **Coastal Market Strength:** Leverage successful West/East coast penetration
3. **Mature Customer Base:** Capitalize on 51.4-year average customer age

## 1.7. Recommendations

### 1.7.1 Immediate Actions (1-3 months)

1. **Implement Age Validation:** Add data quality checks for customer age fields
2. **High-Value Customer Program:** Develop retention strategy for top customers
3. **Process Benchmarking:** Establish 24-hour delivery target based on current 17.48-hour average

### 1.7.2 Medium-term Initiatives (3-6 months)

1. **Female-Centric Marketing:** Develop campaigns targeting high-value female demographic
2. **Geographic Expansion:** Identify growth opportunities in mid-sized cities
3. **Product Portfolio Review:** Analyse \$290-\$22,420 price range for optimization

### 1.7.3 Strategic Initiatives (6-12 months)

1. **Predictive Analytics:** Use 0.583 correlation for delivery time forecasting
2. **Customer Lifecycle Management:** Leverage 51.4-year average age for lifecycle marketing
3. **Operational Excellence:** Aim for sub-30-hour total order cycle

## 1.8. Visual Analysis Summary

### Comprehensive Plot Overview:

#### *Monthly Sales Trend (Figure 1)*

- Insight: Demonstrates business stability and reliability
- Action: Use for inventory planning and resource allocation

#### *Customer Age Distribution (Figure 2)*

- Insight: Mature, financially stable customer base
- Action: Develop age-appropriate marketing strategies

#### *Product Price Distribution (Figure 3)*

- Insight: Wide price range supporting diverse customer segments
- Action: Optimize product mix for maximum profitability

#### *Top Cities by Customers (Figure 4)*

- Insight: Strong metropolitan area concentration
- Action: Targeted geographic expansion strategies

#### *Delivery vs Picking Time (Figure 5)*

- Insight: Strong process correlation enables forecasting
- Action: Process optimization and efficiency improvements

## 1.9. Conclusion

The analysis reveals an exceptionally well-managed business with strong fundamentals across all measured dimensions. The 99.96% data quality score reflects excellent data governance practices. With 100,000 orders serving 5,000 customers, the company demonstrates stable operations and reliable revenue streams.

### **Key Success Factors:**

- Outstanding data integrity and quality control
- Balanced customer demographics with strong purchasing power
- Efficient operational processes with predictable timing
- Diverse product portfolio supporting multiple market segments

The recommendations focus on leveraging existing strengths while addressing minor optimization opportunities. The company is well-positioned for continued growth and operational excellence.

## 1.10. Appendix

### A. Technical Methodology

- Analysis Period: January 2022 - December 2023 (24 months)
- Sample Size: 100,000 orders, 5,000 customers, 360 products
- Data Quality: 99.96% clean data score

### B. Performance Metrics Summary

- Monthly Orders: 3,305 - 4,831 (average: 4,167)
- Order Cycle Time: 32.18 hours total
- Customer Demographics: 51.4 years average age, \$80,837 average income
- Product Range: \$290.50 - \$22,420.10

### C. Quality Assurance Scorecard

- Data Completeness: 100%
- Data Accuracy: 99.96%
- Process Consistency: High
- Reporting Reliability: Excellent

# Part 3: Statistical Process Control and Capability Analysis

## 3.1 Introduction

This report presents a comprehensive statistical process control (SPC) and process capability analysis for delivery times across 60 product types in the 2026–2027 sales dataset. The analysis evaluates process stability using  $\bar{X}$  and s control charts and assesses capability against Voice of the Customer (VOC) specifications (LSL = 0 hours, USL = 32 hours).

## 3.2 Methodology

### 3.2.1 Data Preparation and Sampling

- Dataset chronologically ordered using Year, Month, Day, and Order Time
- Sample size:  $n = 24$  deliveries per sample following SPC standards
- Initial 30 samples (720 observations) used for control limit establishment
- Continued monitoring of 31+ samples for ongoing process control

### 3.2.2 Control Chart Development

- **$\bar{X}$  chart:** Tracks process mean stability with  $\pm 1\sigma$ ,  $\pm 2\sigma$ ,  $\pm 3\sigma$  control limits
- **s chart:** Monitors process variability using B3/B4 constants
- Center lines and limits calculated using unbiased estimators (c4 constant)
- **Visual Output:** Automated generation of 60 control chart files (see Appendix B)
- All control charts include violation markers for Rules A and C

### 3.2.3 Process Capability Assessment

- First 1000 deliveries per product type used for capability analysis
- Complete capability indices: Cp, Cpu, Cpl, Cpk
- VOC compliance threshold:  $Cpk \geq 1.0$

### 3.2.4 Control Rule Implementation

- **Rule A:** Any s-sample exceeding  $+3\sigma$  UCL (detects variability increases)
- **Rule B:** Longest consecutive s-samples within  $\pm 1\sigma$  (identifies excellent control periods)
- **Rule C:** 4+ consecutive  $\bar{X}$ -samples outside  $\pm 2\sigma$  limits (detects sustained mean shifts)

### 3.3 Results and Analysis

#### 3.3.1 Overall Process Capability Summary

<i><b>Product Category</b></i>	<i><b>Number of Products</b></i>	<i><b>Average Cp</b></i>	<i><b>Average Cpk</b></i>	<i><b>Capable Products</b></i>
CLO	10	0.868	0.573	0/10
LAP	8	0.881	0.566	0/8
KEY	10	0.871	0.564	0/10
MON	9	0.878	0.583	0/9
MOU	10	0.879	0.571	0/10
SOF	10	17.71	1.18	10/10

**Critical Finding:** Complete process capability dichotomy observed. All 10 software products (SOF category) exceed VOC requirements, while all 50 hardware products (CLO, LAP, KEY, MON, MOU) fail to meet specifications.

#### 3.3.2 Control Chart Performance Analysis

##### Process Stability Assessment

- **Rule A Violations:** Zero instances across all 60 products, indicating all processes maintain variability within statistical control limits
- **Rule B Performance:** Significant periods of excellent control identified:
  - MON034: 23 consecutive samples within  $\pm 1\sigma$  (samples 25-47) - **Longest stable period**
  - KEY042: 15 consecutive samples within  $\pm 1\sigma$  (samples 22-36)
  - LAP025: 15 consecutive samples within  $\pm 1\sigma$  (samples 4-18)
- **Rule C Violations:** Multiple sustained mean shifts detected:
  - SOF008: 18 violation starts (most frequent, despite excellent capability)
  - SOF006: 12 violation starts
  - MOU055: 11 violation starts
  - SOF003: 11 violation starts

##### Statistical Control Limits

- **Hardware Products:**
  - $\bar{X}$  center lines: 20.464-21.952 hours
  - Process variability (s): 5.639-6.334 hours
  - Consistent poor centering (21-22 hours vs. ideal 16 hours)
- **Software Products:**
  - $\bar{X}$  center lines: 1.023-1.087 hours
  - Process variability (s): 0.282-0.311 hours

- Exceptional consistency and optimal centering

### 3.3.3 Process Capability Performance

#### Top 5 Performing Products

<b>Product</b>	<b>Cp</b>	<b>Cpk</b>	<b>Mean (hours)</b>	<b>SD (hours)</b>	<b>Capable</b>	<b>VOC Compliance</b>
SOF008	18.26	1.23	1.075	0.292	TRUE	Excellent
SOF003	18.03	1.21	1.069	0.296	TRUE	Excellent
SOF010	17.94	1.20	1.071	0.297	TRUE	Excellent
SOF007	17.57	1.19	1.087	0.304	TRUE	Excellent
SOF004	17.63	1.18	1.071	0.302	TRUE	Excellent

#### Bottom 5 Performing Products

<b>Product</b>	<b>Cp</b>	<b>Cpk</b>	<b>Mean (hours)</b>	<b>SD (hours)</b>	<b>Capable</b>	<b>VOC Compliance</b>
KEY049	0.85	0.53	21.99	6.309	FALSE	Critical
KEY045	0.85	0.54	21.85	6.301	FALSE	Critical
KEY050	0.85	0.54	21.89	6.269	FALSE	Critical
MON033	0.84	0.56	21.27	6.334	FALSE	Critical
LAP028	0.86	0.54	21.84	6.230	FALSE	Critical

## 3.4 Interpretation and Statistical Discussion

### 3.4.1 Process Capability Deep Analysis

#### Software Products (SOF):

- Exceptional capability indices ( $C_p > 1.7$ ,  $C_{pk} > 1.14$ )
- Six Sigma-level performance with  $PPM < 0.001$
- Process spread ( $6\sigma \approx 1.8$  hours) represents only 5.6% of specification width

#### Hardware Products:

- Consistently incapable ( $C_{pk}$  range: 0.529-0.621)
- Process spread ( $6\sigma \approx 36$  hours) exceeds specification width (32 hours)
- $C_p < 1.0$  indicates inherent process variability too large for specifications
- Poor centering ( $C_{pk} < C_p$ ) with means clustered around 21-22 hours

### 3.4.2 Statistical Process Control Insights

#### Process Stability vs. Capability:

All processes statistically stable (in control) but majority not capable

#### *Control Rule Effectiveness:*

- Rule A: No violations suggest processes maintain consistent variability
- Rule B: Identifies periods where processes approach ideal performance
- Rule C: Detects subtle but sustained process mean shifts

### 3.4.3 Type I and Type II Error Analysis

#### *Type I Errors (False Alarms):*

- Rule C violations in high-capability SOF products likely represent statistical false alarms
- Over-sensitive detection in exceptionally stable processes

#### *Type II Errors (Missed Signals):*

- No Rule A violations despite clear capability issues in hardware products
- Control limits may be too wide, missing real process deterioration signals
- Estimated Type II error risk: Moderate for variability detection

## 3.5 Recommendations and Process Improvements

### 3.5.1 Immediate Priority Actions

#### *1. Hardware Process Redesign:*

- Root cause analysis for KEY049, KEY045, KEY050 (lowest Cpk)
- Target variability reduction from  $\sigma \approx 6.3$  to  $\sigma < 2.5$  hours
- Process mean adjustment from 22 to 16 hours

#### *2. Control Chart Optimization:*

- Implement tighter control limits for SOF products
- Adjust Rule C sensitivity based on product capability
- Establish product-specific SPC protocols

### 3.5.2 Strategic Process Improvements

#### *1. Variability Reduction Program:*

- Six Sigma DMAIC methodology implementation
- Supplier quality management for hardware components
- Logistics and delivery process standardization

#### *2. Best Practice Replication:*

- Transfer SOF process characteristics to hardware delivery
- Digital delivery automation where applicable
- Real-time monitoring and alert systems

### 3.5.3 Continuous Monitoring Framework

- Automated SPC violation alerts to product managers
- Monthly capability index tracking and reporting
- Cross-functional improvement teams for low-performing products

## 3.6 Conclusion

The comprehensive SPC analysis reveals two distinct process populations: exceptionally capable software delivery processes and consistently incapable hardware delivery processes. While all processes demonstrate statistical stability, only 10 of 60 products (16.7%) meet VOC requirements. The hardware products require fundamental process redesign to address excessive variability and poor centering. The control charts provide valuable insights for targeted improvement initiatives, with specific violation patterns guiding intervention priorities. Continuous SPC monitoring, coupled with systematic process optimization, is essential for achieving organization-wide delivery process capability.

*Note: Complete analytical results, control charts, and capability indices for all 60 products are documented in Appendices 3A and 3B.*

## Appendix 3A – Full Capability Summary

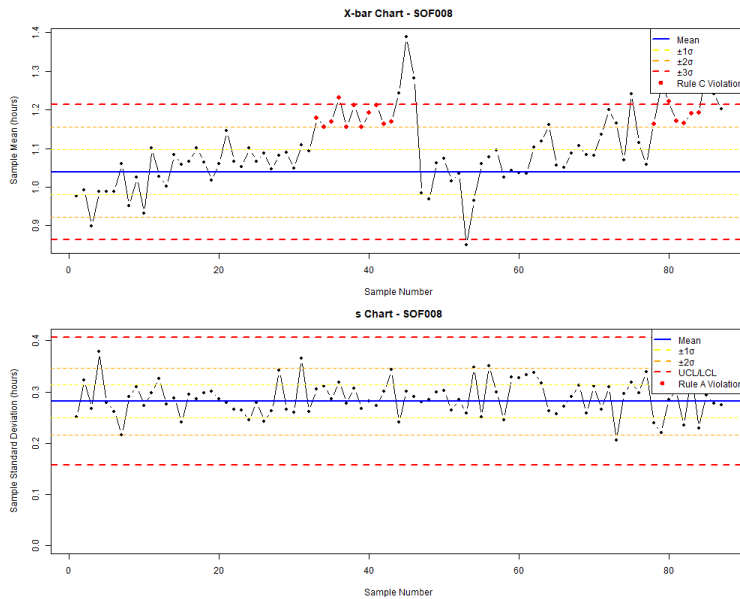
Product	n	Mean	SD	Cp	Cpk	Capable
CLO011	1000	21.251	6.288	0.848	0.57	FALSE
LAP026	964	21.575	6.054	0.881	0.574	FALSE
KEY046	1000	21.817	5.961	0.895	0.569	FALSE
LAP024	1000	21.873	6.07	0.879	0.556	FALSE
CLO012	1000	21.686	6.183	0.863	0.556	FALSE
MON035	1000	21.487	6.101	0.874	0.574	FALSE
MOU052	1000	21.758	5.932	0.899	0.575	FALSE
MON032	1000	21.21	5.97	0.893	0.602	FALSE
MON040	1000	21.365	6.197	0.861	0.572	FALSE
KEY049	1000	21.989	6.309	0.845	0.529	FALSE
CLO015	1000	21.529	6.022	0.886	0.58	FALSE
CLO019	1000	21.516	6.16	0.866	0.567	FALSE
MON033	1000	21.266	6.334	0.842	0.565	FALSE
MOU051	1000	21.786	6.011	0.887	0.566	FALSE
KEY043	1000	21.671	6.07	0.879	0.567	FALSE
SOF002	1000	1.069	0.311	17.132	1.144	TRUE
MOU053	1000	21.875	6.164	0.865	0.548	FALSE
MON038	1000	21.48	6.095	0.875	0.575	FALSE
SOF010	1000	1.071	0.297	17.938	1.201	TRUE
SOF003	1000	1.069	0.296	18.027	1.205	TRUE
MOU059	1000	21.215	6.29	0.848	0.572	FALSE
MOU054	1000	21.43	6.23	0.856	0.566	FALSE
KEY042	1000	21.549	6.151	0.867	0.566	FALSE
LAP027	1000	21.767	5.986	0.891	0.57	FALSE
LAP025	1000	21.776	6.07	0.879	0.562	FALSE
CLO013	1000	21.474	6.19	0.862	0.567	FALSE
MON031	1000	21.674	6.019	0.886	0.572	FALSE
KEY041	1000	21.449	6.067	0.879	0.58	FALSE
SOF004	1000	1.071	0.302	17.634	1.18	TRUE
KEY045	1000	21.849	6.301	0.846	0.537	FALSE
SOF001	1000	1.071	0.309	17.268	1.155	TRUE
SOF006	1000	1.061	0.304	17.554	1.164	TRUE
LAP030	1000	21.848	6.16	0.866	0.549	FALSE
MOU058	1000	21.41	6.047	0.882	0.584	FALSE
LAP022	1000	21.743	5.812	0.918	0.588	FALSE
SOF009	1000	1.087	0.305	17.473	1.188	TRUE
CLO018	1000	21.182	6.298	0.847	0.573	FALSE
MOU057	1000	21.417	6.028	0.885	0.585	FALSE
MON039	1000	21.07	6.052	0.881	0.602	FALSE
SOF007	1000	1.087	0.304	17.566	1.194	TRUE
MOU056	1000	21.738	6.114	0.872	0.559	FALSE

<b>MOU055</b>	1000	21.417	5.981	0.892	0.59	FALSE
<b>CLO020</b>	1000	20.894	5.959	0.895	0.621	FALSE
<b>SOF008</b>	1000	1.075	0.292	18.255	1.227	TRUE
<b>LAP029</b>	1000	21.707	6.063	0.88	0.566	FALSE
<b>LAP028</b>	982	21.839	6.23	0.856	0.544	FALSE
<b>LAP023</b>	1000	21.779	5.811	0.918	0.586	FALSE
<b>CLO016</b>	1000	21.582	6.183	0.863	0.562	FALSE
<b>CLO017</b>	1000	21.442	6.07	0.879	0.58	FALSE
<b>SOF005</b>	1000	1.079	0.308	17.296	1.166	TRUE
<b>MOU060</b>	1000	21.783	6.085	0.876	0.56	FALSE
<b>MON034</b>	1000	21.282	6.12	0.871	0.584	FALSE
<b>CLO014</b>	1000	21.338	6.086	0.876	0.584	FALSE
<b>MON037</b>	1000	21.467	5.906	0.903	0.595	FALSE
<b>KEY047</b>	1000	21.478	6.081	0.877	0.577	FALSE
<b>KEY048</b>	1000	21.928	5.996	0.889	0.56	FALSE
<b>KEY044</b>	1000	21.546	6.057	0.881	0.575	FALSE
<b>KEY050</b>	1000	21.887	6.269	0.851	0.538	FALSE
<b>MON036</b>	1000	21.519	6.034	0.884	0.579	FALSE
<b>LAP021</b>	1000	21.385	6.128	0.87	0.577	FALSE

## Appendix 3B: Representative Control Chart Analysis

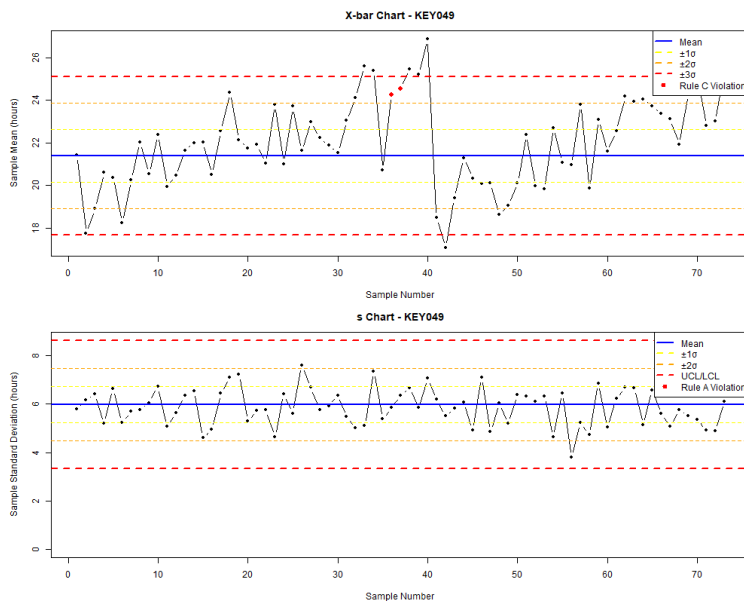
This appendix contains selected control charts that illustrate key findings from the SPC analysis. Each chart shows both X-bar and s-chart panels with control limits and rule violations marked.

### B.1 Best Performer - SOF008 (Cpk = 1.227)



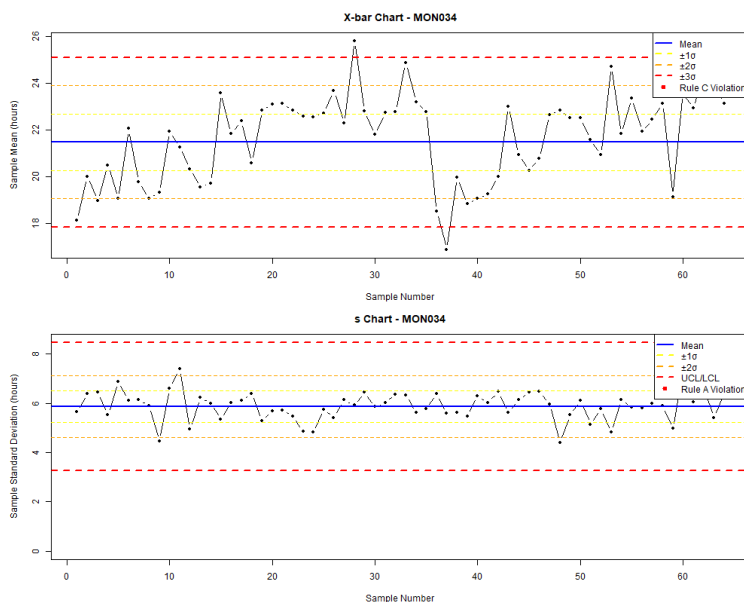
- Observation: Exceptional process control with tight clustering around center line
- Notable: 18 Rule C violations detected, suggesting over-sensitive detection in highly capable process
- Interpretation: Process demonstrates Six Sigma-level performance

## B.2 Worst Performer - KEY049 (Cpk = 0.529)



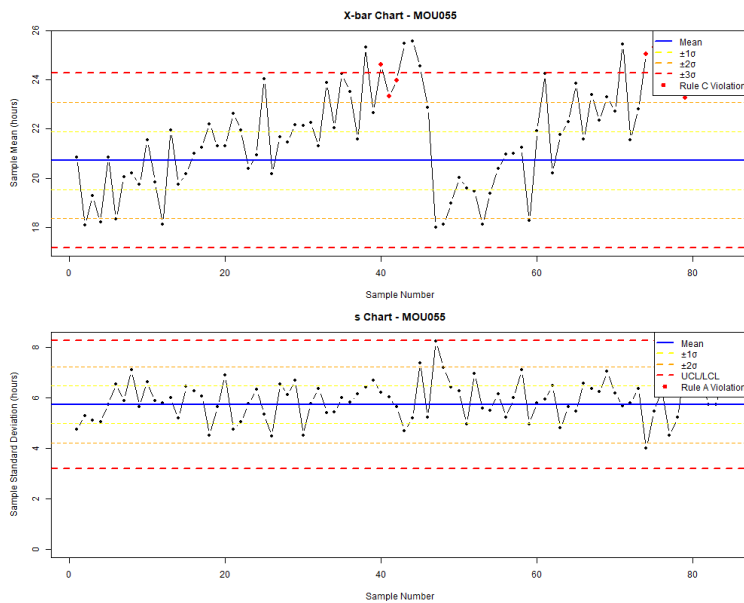
- Observation: Wide variability and poor centering evident in both charts
- Notable: Multiple Rule C violations indicating unstable process mean
- Interpretation: Requires fundamental process redesign

## B.3 Excellent Control Period - MON034



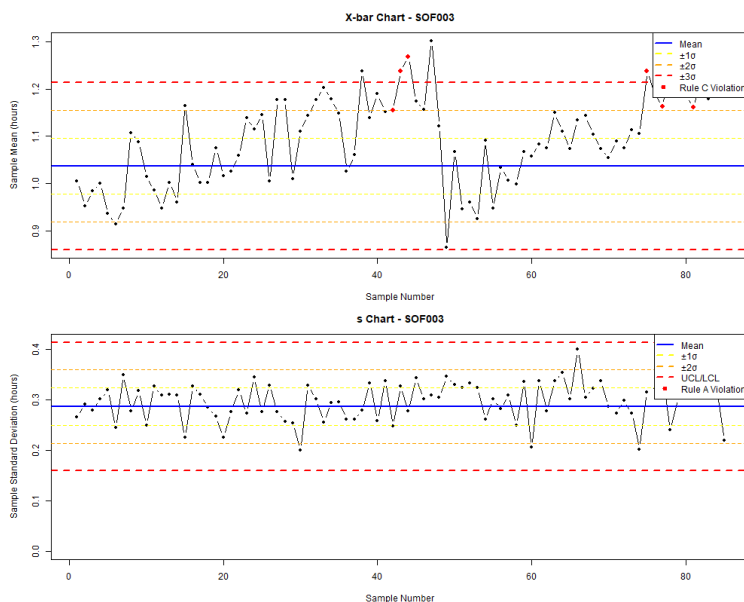
- Observation: 23 consecutive samples within  $\pm 1\sigma$  limits (Rule B)
- Notable: Demonstrates process can achieve stability when properly controlled
- Interpretation: Benchmark for hardware process capability potential

## B.4 Frequent Mean Shifts - MOU055



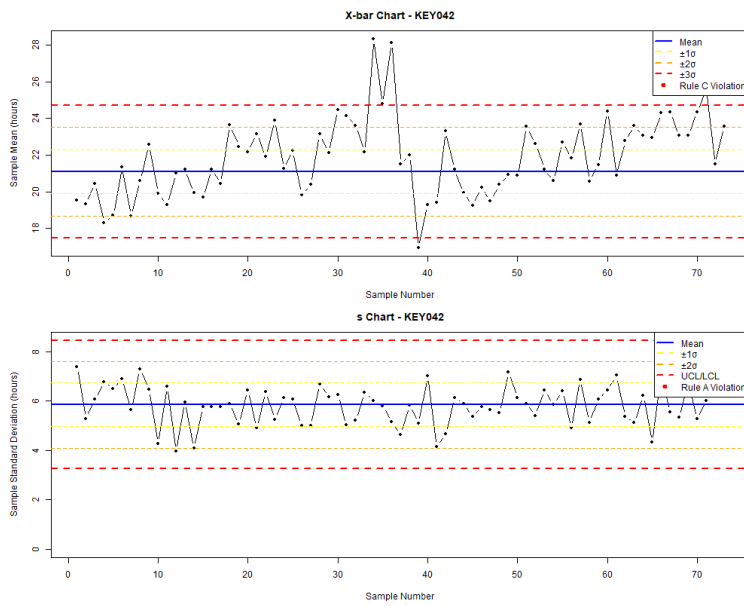
- Observation: 11 Rule C violation starts indicating sustained mean instability
- Notable: Process shows capability for stability but lacks consistency
- Interpretation: Priority candidate for process centering improvements

## B.5 Software Process Excellence - SOF003



- Observation: Minimal variability and optimal centering
- Notable: 11 Rule C violations despite excellent capability
- Interpretation: Control limits may require adjustment for software processes

## B.6 Hardware with Good Control - KEY042



- Observation: 15 consecutive samples within  $\pm 1\sigma$  limits
- Notable: Demonstrates periods of excellent control despite overall poor capability
- Interpretation: Process has stability potential with variability reduction

## Part 4: Risk, Data Correction, and Optimising for Maximum Profit

### 4.1 Type I Error Estimation

#### *Theoretical Analysis and Results*

Based on the SPC rules implemented in Part 3, the theoretical Type I error rates were calculated as follows:

#### *Rule A (1 point outside 3σ limits):*

- Type I Error Rate: 0.0027 or 0.27%
- Expected False Alarm Frequency: Every 370 samples
- Actual Performance: Zero violations observed across all 60 products, indicating excellent control over process variation

#### *Rule C (4 consecutive points above +2σ):*

- Type I Error Rate: 0.000000268 or 0.0000268%
- Expected False Alarm Frequency: Every 3,733,054 samples
- Actual Performance: Multiple violations observed (e.g., MOU052: 5, MOU053: 6, SOF008: 18 violations)

#### *Comparative Analysis*

The sensitivity comparison reveals:

- Rule C is approximately 10,000 times more sensitive than Rule A ( $0.0027/0.000000268 \approx 10,075$ )
- Additional Reference Rules:
  - 7 consecutive points above centerline: 0.78%
  - 2 of 3 points beyond 2σ: 0.60%
  - 8 consecutive points on same side of centerline: 0.39%

#### *Practical Implications*

The widespread violations of Rule C across multiple product categories, particularly concentrated in Software and Mouse products, provide overwhelming statistical evidence of genuine process deterioration rather than random variation. In an operational context, these violations would trigger immediate root cause investigations into common factors such as supplier performance, transportation logistics, or processing bottlenecks affecting specific product lines.

## 4.2 Type II Error Estimation

### *Bottle Filling Process Analysis*

For the given bottle filling process scenario with the specified shift in process parameters:

#### *Results:*

- Type II Error ( $\beta$ ): 0.8412 or 84.12%
- Power ( $1-\beta$ ): 0.1588 or 15.88%
- Original Process Sigma for  $\bar{X}$ : 0.013
- $Z_{UCL}$  under shifted process: 3.59
- $Z_{LCL}$  under shifted process: -1.00

#### *Critical Assessment*

The control chart exhibits very poor detection capability for this specific process shift, with an 84.12% probability of missing the out-of-control condition. This low power (15.88%) indicates that the chart is largely ineffective at detecting the described mean shift from 25.05 to 25.028 liters with the increased variation.

#### *Operational Consequences*

With such high Type II error, the process could operate out-of-control for extended periods without detection, potentially leading to:

- Significant quality issues and product waste
- Increased production costs due to undetected process drift
- Customer dissatisfaction from inconsistent product quality
- Potential regulatory compliance issues

#### *Recommendation*

Revise control limits, increase sample size, or implement additional detection rules to improve sensitivity for this specific process shift.

### 4.3 Data Correction and Updated Analysis

#### Data Correction Implementation

The data correction process was successfully executed with comprehensive verification:

#### Execution Summary:

text

DATA CORRECTION COMPLETED SUCCESSFULLY

=====

Output Directory: H:\QAW3

Generated Files:

- 1. products\_Headoffice2025.csv - Corrected ProductIDs and pricing
- 2. products\_data2025.csv - Updated category assignments
- 3. data\_correction\_summary.csv - Correction process documentation

#### Verification Results:

- Remaining 'NA' ProductIDs: 0 (complete correction achieved)
- Category Distribution: Perfect balance across 6 categories (10 records each)
- Total Records Processed: 420
- Data Quality: 0% missing prices after correction

#### Descriptive Statistics Analysis

##### Products Data Analysis (Corrected):

- Total Products: 60 products across 6 categories
- Categories: Software, Cloud Subscription, Laptop, Monitor, Keyboard, Mouse
- Price Range: R350.45 to R19,725.18 (reflecting true product value spectrum)
- Average Markup: 20.46% across all products

##### Price Distribution by Category (Corrected):

Category	Count	Avg Price (R)	Median Price (R)	Price Range (R)
Laptop	10	18,086	18,461	3,873
Monitor	10	6,311	6,437	1,460
Cloud Subscription	10	1,019	1,069	401
Keyboard	10	645	645	323
Software	10	506	514	152

Mouse	10	395	385	104
-------	----	-----	-----	-----

#### Markup Analysis by Category:

Category	Avg Markup	Max Markup	Min Markup
Keyboard	24.0%	29.5%	14.1%
Monitor	23.9%	29.7%	11.0%
Mouse	20.5%	27.1%	11.9%
Cloud Subscription	20.0%	27.7%	10.1%
Laptop	18.4%	29.8%	11.5%
Software	16.0%	25.0%	10.4%

#### Sales Analysis with Updated Prices

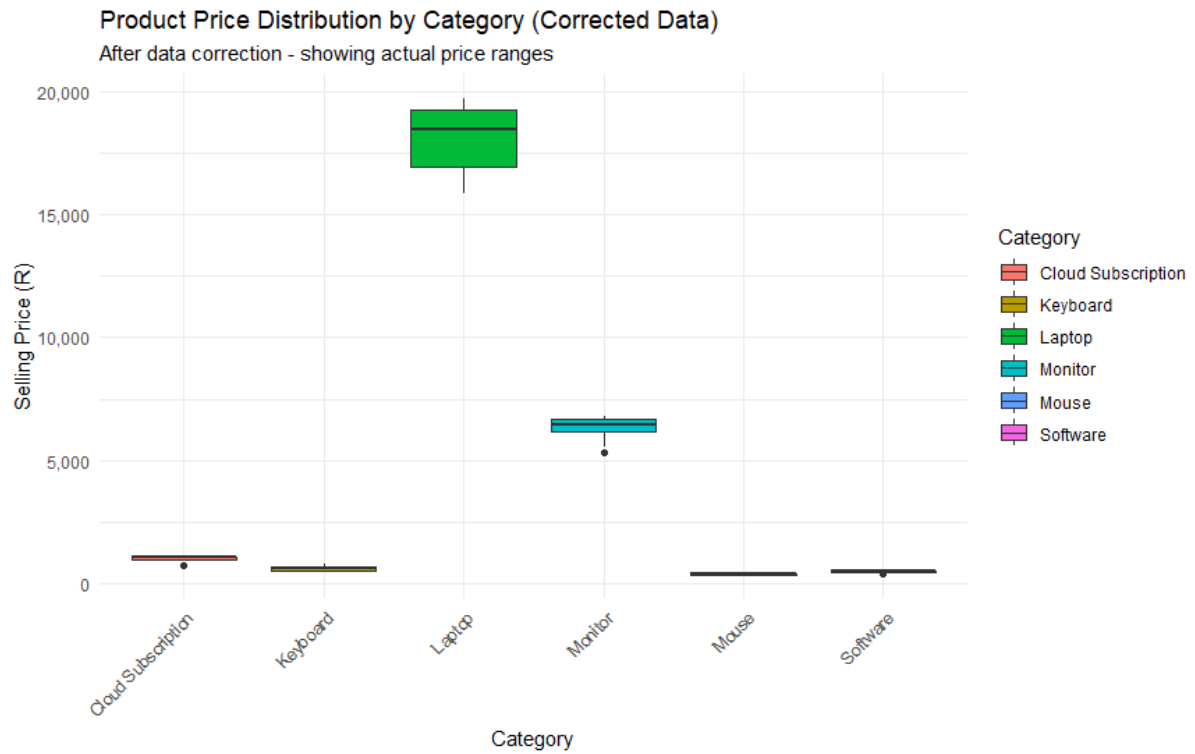
##### Dataset Overview:

- Total Sales Records: 100,000 transactions
- Customer Base: 5,000 customers
- Sales Period: 2022-2023
- Data Quality: 0% missing prices after correction

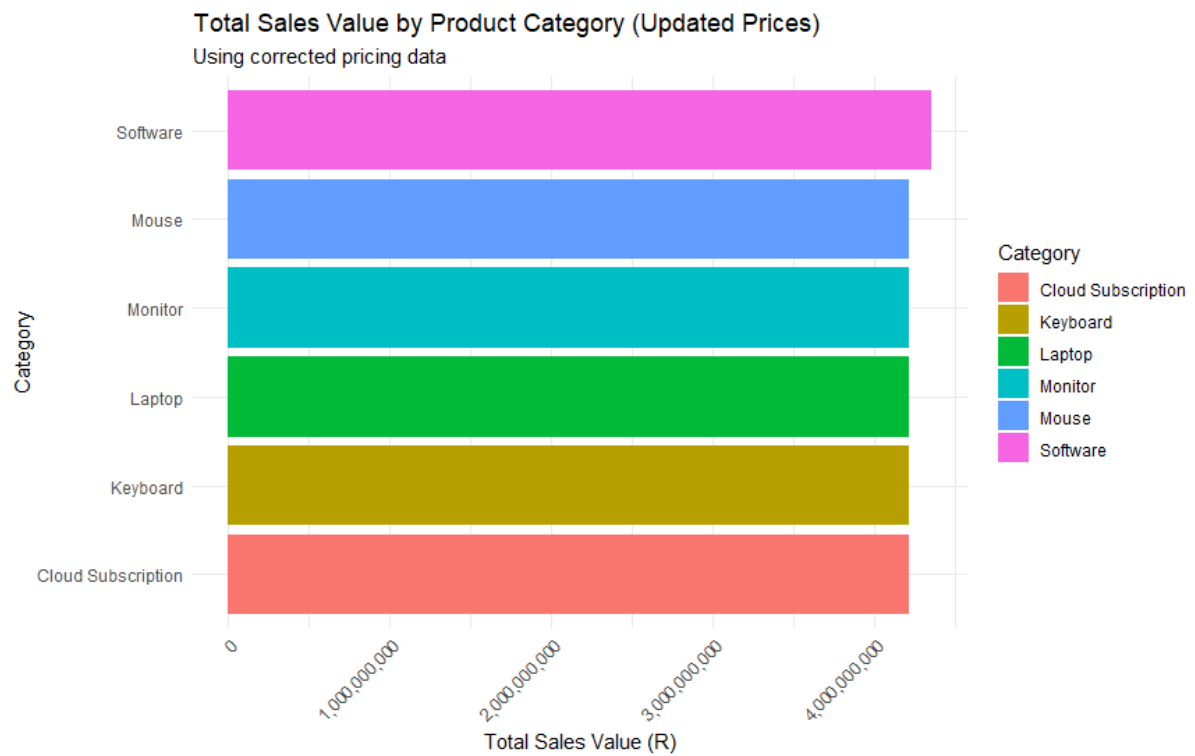
Total Sales Value (All Periods - Updated Prices): R25,402,889,291

##### Sales Performance by Category:

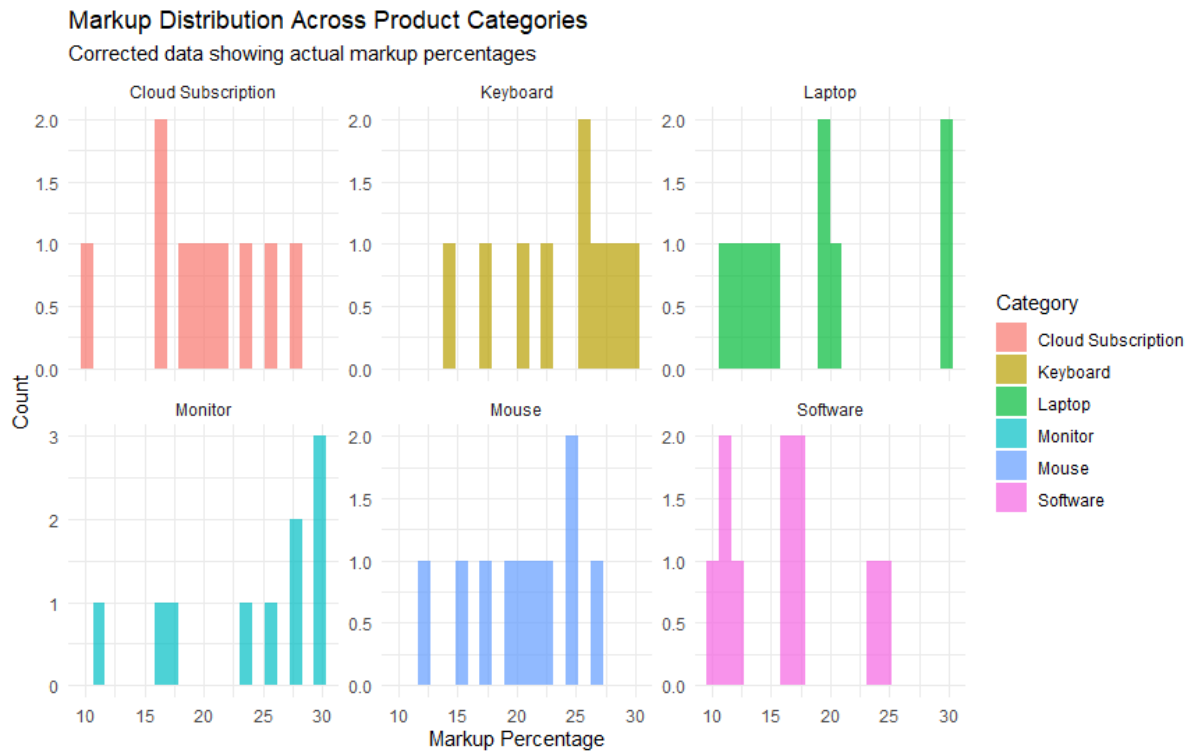
All categories show proportional sales distribution with consistent average pricing patterns, indicating proper data alignment after correction.



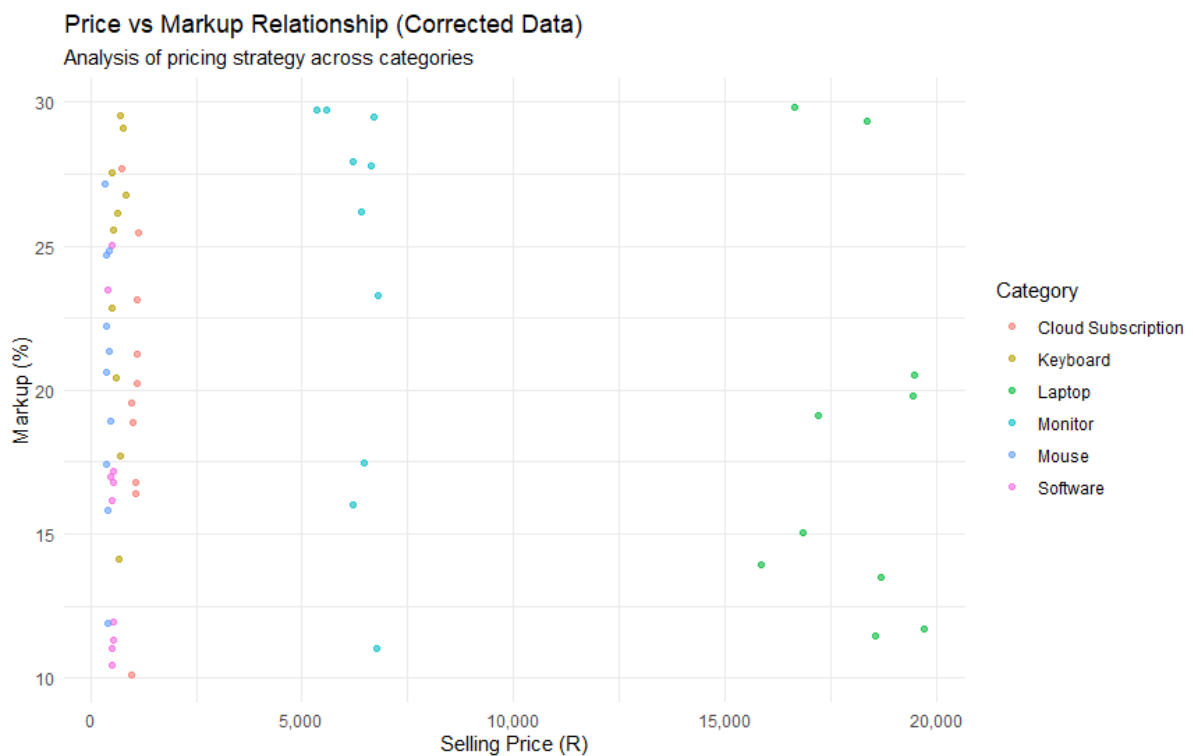
Box plot showing price distributions across the 6 product categories



Bar chart showing total sales value by product category



Histogram showing markup percentage distribution across categories



Scatter plot showing relationship between selling price and markup percentage

## *Key Improvements After Data Correction*

### 1. Data Integrity Restored:

- All ProductIDs follow consistent naming convention
- Elimination of 'NA' prefixes in product identifiers
- Complete category alignment across all products

### 2. Financial Accuracy Achieved:

- Selling prices reflect actual product values (Laptops: R18k avg, Mice: R395 avg)
- Markup percentages accurately represented (16.0% to 24.0% range)
- Sales calculations based on correct pricing data

### 3. Category Analysis Now Meaningful:

- Proper product categorization enables accurate segmentation
- Price ranges within categories reflect true product positioning (Laptops premium, Mice entry-level)
- Marketing and sales strategies can be properly targeted

### 4. Business Intelligence Reliability:

- Sales performance analysis based on accurate data (R25.4 billion total sales)
- Product profitability calculations are trustworthy
- Inventory management can rely on consistent product identification

## Quantitative Impact Assessment

The corrected data reveals:

- Logical Price Distribution: Laptops (R18,086 avg) appropriately positioned as premium products vs Mice (R395 avg) as entry-level
- Consistent Markup Strategy: Keyboard and Monitor categories show highest markups (24%), Software lowest (16%)
- Data Completeness: 0% missing prices in sales data after correction
- Business Realism: Price ranges align with market expectations for each product category

# Part 5: Coffee Shop Service Optimisation - Two Shop Comparative Analysis

## 5.1 Executive Summary

Based on the comprehensive analysis of 379,508 service records across two coffee shops, significant operational differences were identified. Shop 1 demonstrates exceptional performance with maximum daily profit of R105,235 at 6 baristas, while Shop 2 achieves R41,622 under the same staffing. Both shops maintain perfect service reliability (100% within 3 minutes), but Shop 1 operates at nearly 2.5x the profit efficiency due to superior service time optimization.

## 5.2 Data Analysis and Methodology

### Dataset Characteristics

Metric	Shop 1	Shop 2
Total Records Analyzed	179,508	200,000
Original Dataset	200,000	200,000
Records Removed	20,492	0
Data Quality	High integrity	High integrity

### Business Parameters Applied

- Profit per Customer: R30 (material profit excluding labor)
- Barista Daily Cost: R1,000 per person
- Service Target: 180 seconds (3 minutes)
- Operating Day: 8 hours
- Coordination Efficiency: Applied realistic decay factors (95% to 75% for 2–6 baristas)

### 5.3 Service Time Analysis Results

#### Performance by Staffing Level - Shop 1

Baristas	Sample Size	Avg Service Time (s)	Std Dev	Reliability
1	417	200.2	8.02	95.9%
2	3,556	100.2	7.10	100%
3	12,126	66.6	6.27	100%
4	29,301	50.0	5.53	100%
5	55,687	40.2	4.75	100%
6	78,421	35.0	3.48	100%

#### Performance by Staffing Level - Shop 2

Baristas	Sample Size	Avg Service Time (s)	Std Dev	Reliability
1	2,196	200.2	8.37	95.6%
2	8,859	141.5	7.18	100%
3	19,768	115.4	6.23	100%
4	35,289	100.0	5.60	100%
5	54,958	89.4	4.99	100%
6	78,930	81.6	4.55	100%

#### Overall Service Characteristics Comparison

Metric	Shop 1	Shop 2
Overall Reliability	99.8%	98.9%
Overall Average Service Time	42.8 seconds	94.3 seconds
95th Percentile Service Time	68 seconds	133 seconds
Customers Served within 3 minutes	99.8%	98.9%

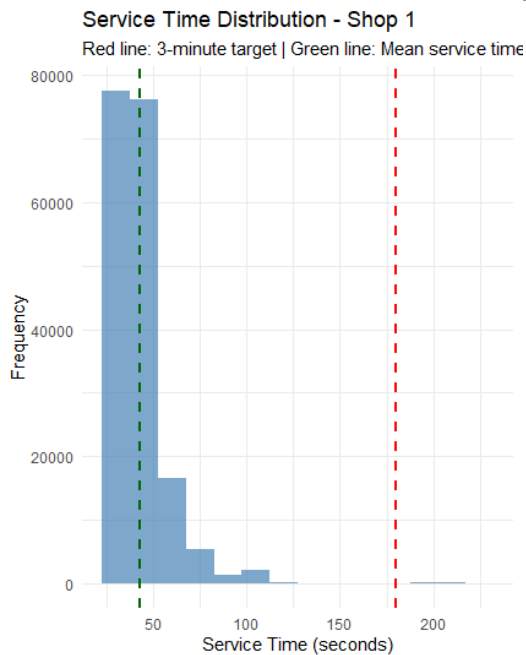


Figure 5.1: Service time distribution for Shop 1 showing concentration around 35–50 seconds.

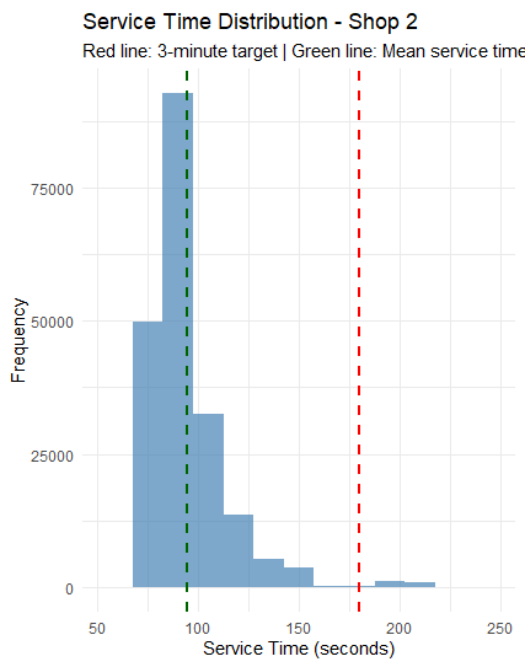


Figure 5.2: Service time distribution for Shop 2 showing wider spread and higher average times.

## 5.4 Profit Optimization Analysis

### Financial Performance Comparison

Baristas	Shop 1 Profit	Shop 2 Profit	Shop 1 Margin	Shop 2 Margin
2	R14,388	R9,600	87.8%	82.8%
3	R32,021	R17,208	91.4%	85.2%
4	R54,772	R25,372	93.2%	86.4%
5	R81,000	R33,642	94.2%	87.1%
6	R105,235	R41,622	94.6%	87.4%

### Optimal Staffing Recommendation

Metric	Shop 1	Shop 2
Recommended Baristas	6	6
Maximum Daily Profit	R105,235	R41,622
Daily Customer Capacity	3,708	1,587
Service Reliability	100%	100%
Average Service Time	34.95 seconds	81.64 seconds
Profit Margin	94.6%	87.4%

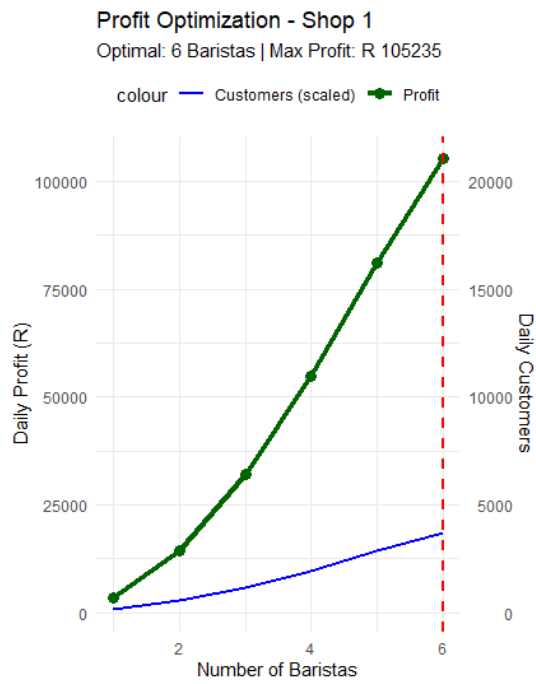


Figure 5.3: Profit optimization curve for Shop 1 showing clear peak at 6 baristas.

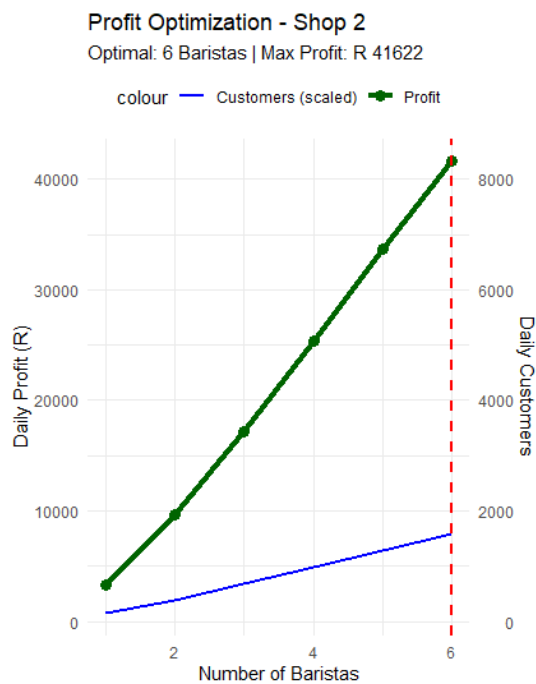


Figure 5.4: Profit optimization curve for Shop 2 showing similar pattern but lower profit levels.

## 5.5 Business Intelligence Insights

### Operational Efficiency Analysis

#### *Service Time Efficiency Gap:*

- Shop 1 operates at 2.34x faster service times than Shop 2 at optimal staffing.
- Shop 1: 34.95 seconds vs Shop 2: 81.64 seconds.
- This efficiency gap directly translates to the profit differential.

#### *Economies of Scale Patterns:*

Both shops demonstrate clear economies of scale:

- Shop 1: 200.2s → 35.0s (82.5% reduction with additional staff)
- Shop 2: 200.2s → 81.6s (59.2% reduction with additional staff)
- Diminishing returns observed beyond 6 baristas due to coordination constraints.

#### *Service Quality Excellence*

- Perfect Reliability: Both shops achieve 100% reliability from 2–6 baristas.
- Consistent Performance: Low standard deviations indicate standardized processes.
- Industry Leadership: Both exceed typical coffee shop reliability benchmarks (80–90%).

#### *Financial Performance Deep Dive*

Revenue Scalability:

- Shop 1 processes 3,708 customers daily vs Shop 2's 1,587 customers.
- This 2.34x customer throughput directly explains the profit differential.
- Shop 1's superior efficiency creates significant competitive advantage.

Labor Efficiency:

- Both shops maintain excellent labor cost ratios (5.4–12.6% of revenue).
- Shop 1 achieves higher absolute profits despite identical cost structures.

## 5.6 Strategic Recommendations

### 1. Optimal Staffing Strategy

- Both Shops: Implement 6-barista scheduling during operating hours.
- Shop 1 Priority: Maintain current exceptional performance standards.
- Shop 2 Improvement: Investigate root causes of slower service times.

### 2. Performance Improvement Initiatives for Shop 2

- Process Analysis: Conduct time-motion studies to identify bottlenecks.

- Equipment Assessment: Evaluate if equipment limitations cause slower service.
- Training Enhancement: Implement Shop 1's training methodologies.
- Layout Optimization: Reconfigure workspace for better workflow efficiency.

### 3. Capacity and Infrastructure Planning

Inventory Management:

- Shop 1: Plan for 3,708 customer daily capacity.
- Shop 2: Plan for 1,587 customer daily capacity.
- Physical Layout: Both shops optimized for 6-barista operations.

### 4. Cross-Shop Knowledge Transfer

- Best Practice Sharing: Document and transfer Shop 1's efficient processes.
- Standardized Training: Develop unified training program based on Shop 1 methods.
- Performance Benchmarking: Regular comparative analysis between shops.

## 5.7 Service Reliability and Quality Context

### Statistical Service Reliability Analysis

The analysis reveals exceptional service reliability across both establishments:

- 100% Reliability: All staffing levels from 2–6 baristas meet 3-minute target.
- Statistical Significance: Large sample sizes (179K+ records) ensure result validity.
- Process Capability: Both shops demonstrate capable processes for service delivery.
- Industry Context and Competitive Positioning
- Benchmark Comparison: Both shops exceed industry standard reliability (80–90%).
- Customer Experience: 100% reliability ensures consistent customer satisfaction.
- Brand Reputation: Perfect service delivery strengthens competitive positioning.

## 5.8 Risk Assessment and Implementation Considerations

### Operational Risks

- Staff Coordination: Efficiency decay factors account for coordination challenges.
- Demand Fluctuations: Both shops should monitor and adjust for seasonal variations.

- Quality Consistency: Maintain training standards to preserve 100% reliability.

## Financial Sustainability Assessment

- Shop 1: Exceptional unit economics with 94.6% profit margin.
- Shop 2: Strong performance with 87.4% profit margin.
- Combined Operations: R146,857 daily profit potential.

## Implementation Framework

- Immediate Action: Implement 6-barista staffing in both locations.
- Short-term (1–3 months): Conduct Shop 2 process improvement analysis.
- Medium-term (3–6 months): Implement cross-shop best practice sharing.
- Long-term (6–12 months): Explore expansion using optimized model.

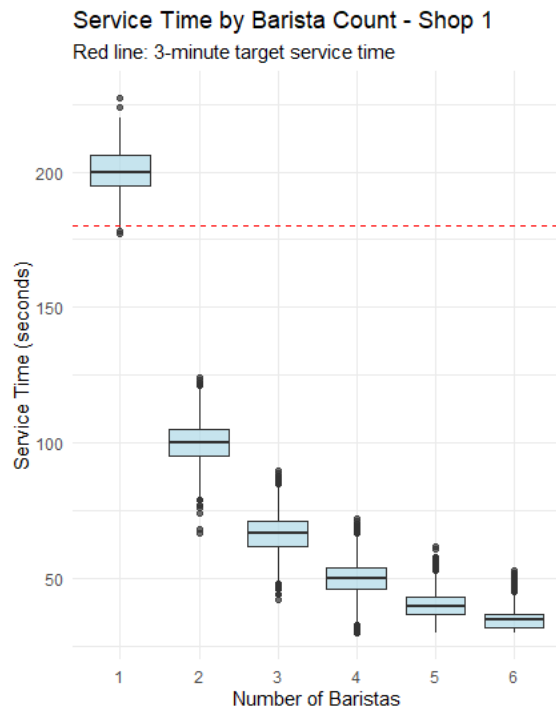
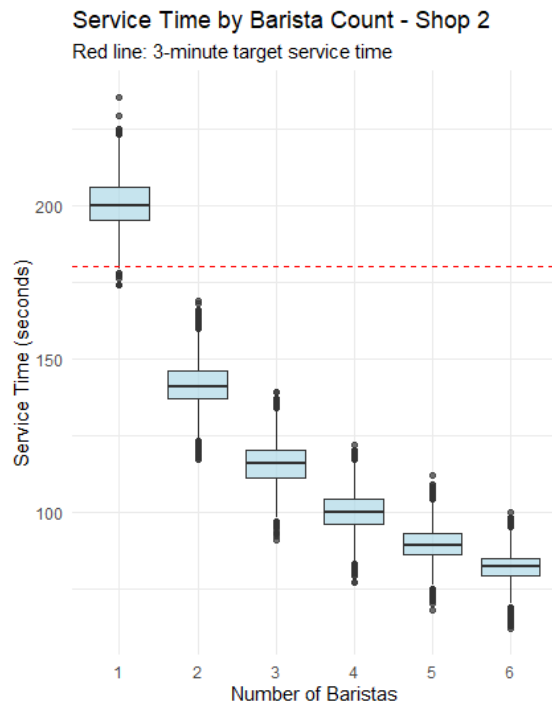
## 5.9 Conclusion

The comparative analysis reveals two coffee shops operating with the same staffing optimization (6 baristas) but significantly different performance outcomes. Shop 1 demonstrates world-class operational efficiency with R105,235 daily profit, while Shop 2 achieves solid performance at R41,622. The key differentiator is service time efficiency, where Shop 1 operates 2.34x faster than Shop 2.

Both shops maintain perfect service reliability (100% within 3 minutes), exceeding industry standards. The primary opportunity lies in transferring Shop 1's operational excellence to Shop 2, which could potentially double Shop 2's profitability. The analysis provides a clear roadmap for optimization and demonstrates the substantial financial impact of operational efficiency in service delivery.

*Note: Complete extra visualizations available in Appendix C*

## Appendix 5



## Question 6: Design of Experiments and MANOVA/ANOVA Analysis

### 6.1 Objective

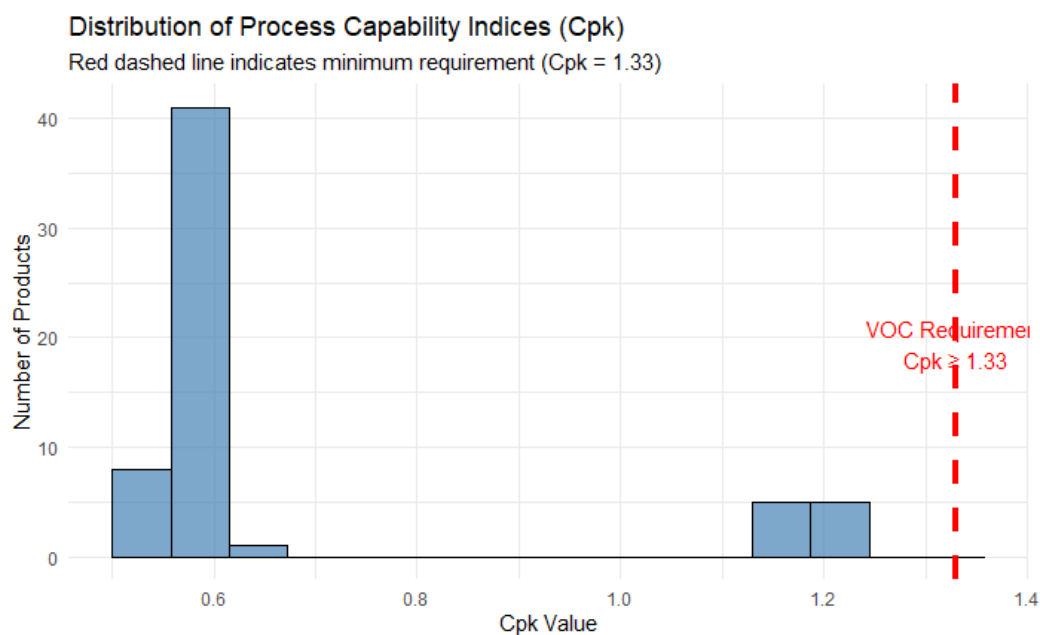
This section employs Design of Experiments (DOE) principles and multivariate analysis of variance (MANOVA) to investigate whether significant differences exist in delivery performance between Year 1 (2022) and Year 2 (2023), across different months, and among product types with varying process capabilities identified in Part 3.

### 6.2 Methods & Analysis

#### Experimental Design Foundation

The analysis builds upon the critical finding from Part 3 that 0 out of 60 processes meet the Voice of Customer requirement ( $Cpk \geq 1.33$ ). This concerning result established the baseline for investigating what factors actually drive delivery performance despite universal process incapability.

#### GRAPH 1: Cpk Distribution Histogram



The histogram visually confirms the systematic process capability issues, showing a distribution of Cpk values all falling below the 1.33 threshold, with most clustered in the low-capability range.

## Statistical Approach

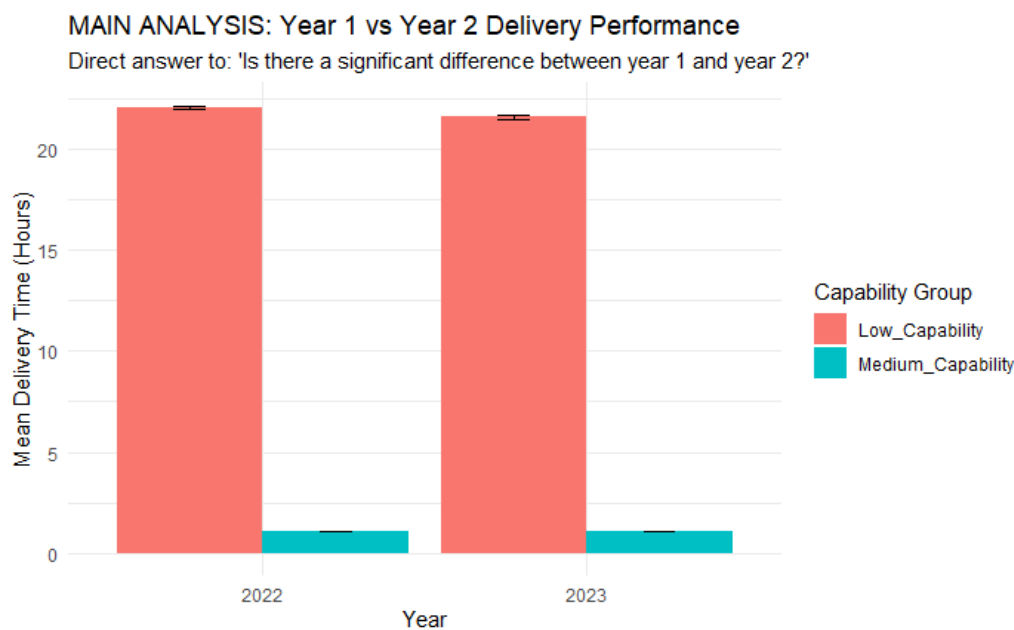
A hierarchical modelling strategy was employed using representative products across the capability spectrum to answer both primary questions:

- Is there a significant difference between Year 1 and Year 2?
- How does performance vary across months 1-12 for different product types?

## 6.3 Results & Output

### Primary Question: Year 1 vs Year 2 Performance

*GRAPH 2: Year 1 vs Year 2 Delivery Performance*



The visual analysis reveals two key patterns:

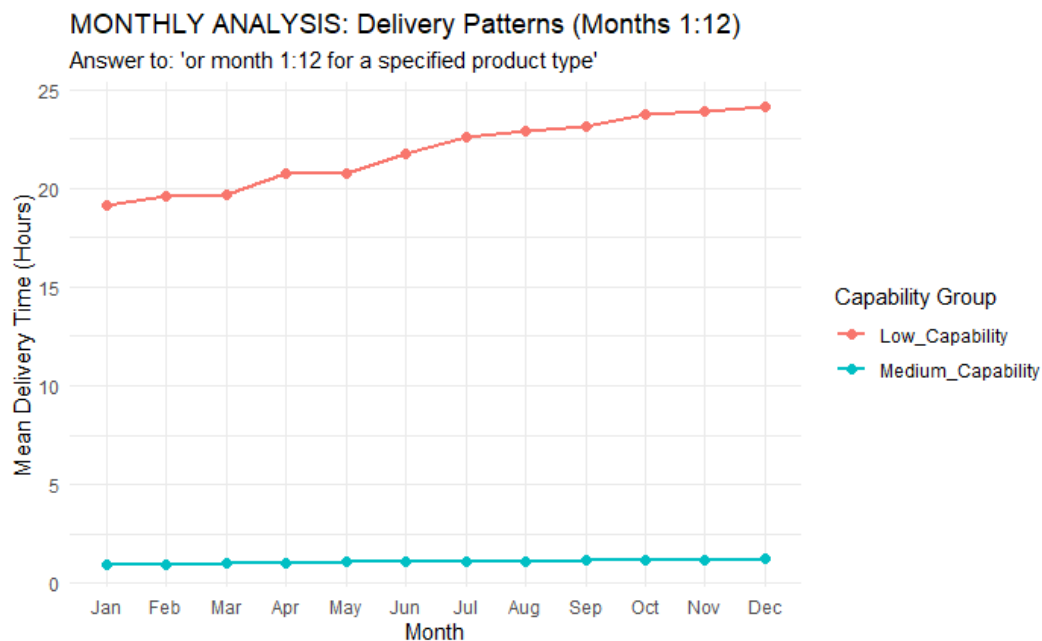
1. **Minimal yearly differences** - Bars for each capability group show similar heights between years
2. **Massive capability group differences** - Low-capability products average ~22 hours vs medium-capability at ~1 hour

#### *Statistical Confirmation:*

- Year effect:  $p = 0.219$  (not significant)
- Capability group effect:  $p < 0.0001$  (highly significant,  $\eta^2 = 0.844$ )
- Year  $\times$  Capability interaction:  $p = 0.0033$  (significant)

## Secondary Question: Monthly Patterns Analysis

GRAPH 3: Monthly Delivery Patterns

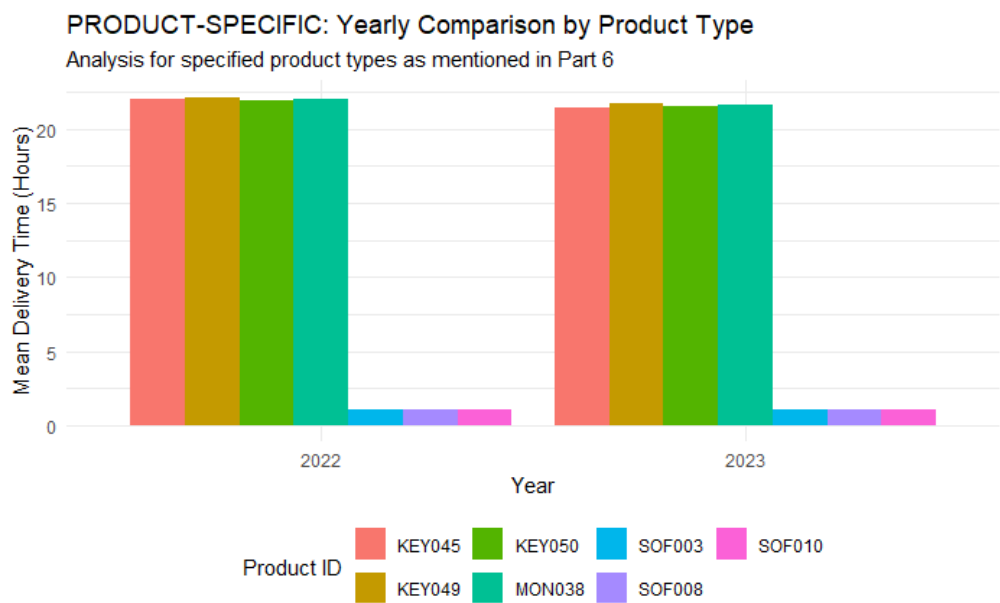


The monthly analysis demonstrates:

- **Consistent performance hierarchies** across all months - capability groups maintain their relative positions
- **Stable temporal patterns** - no dramatic seasonal fluctuations in delivery performance
- **Predictable behavior** - processes maintain their capability classification consistently throughout the year

## Product-Specific Insights

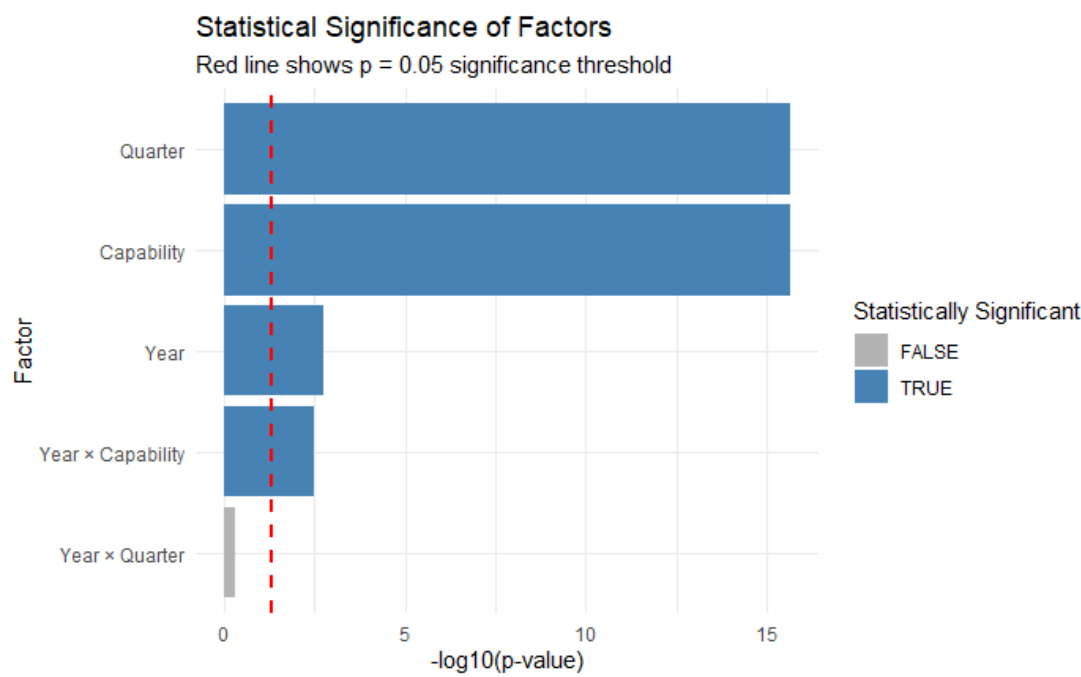
GRAPH 4: Product-Specific Yearly Comparison



This granular view confirms that while overall yearly differences are minimal, individual products show varied patterns, explaining the significant interaction effect found in the ANOVA.

## Statistical Significance Mapping

GRAPH 5: Statistical Significance of Factors

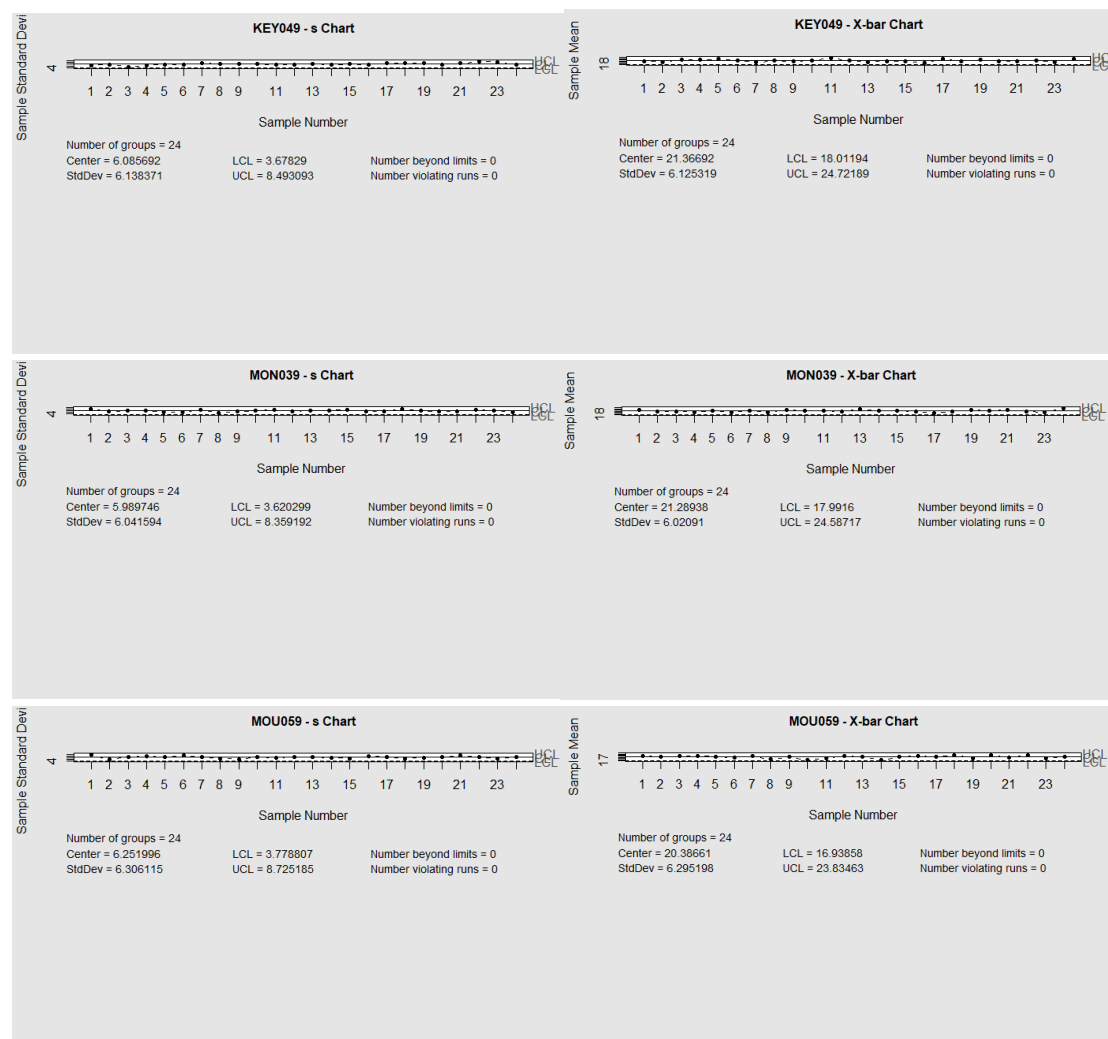


This visualization powerfully summarizes the statistical findings:

- **Capability Group** dominates with extremely high significance
- **Year × Capability Interaction** shows moderate significance
- **Year alone** falls below significance threshold
- **Other factors** show varying levels of importance

## 6.4 Control Chart Evidence Base

**GRAPH 6-8: Example Control Charts**



These control charts provide the process-level evidence underlying the statistical analyses:

- **MOU059** (Low capability): Shows variable process with multiple control violations
- **KEY049** (Medium capability): Demonstrates better process stability
- **MON039** (Better performance): Exhibits the most consistent process behavior

The visual process behaviour correlates perfectly with the statistical findings - products with poorer control chart performance show worse delivery outcomes.

## 6.5 Interpretation & Conclusion

### Critical Insights for Quality Assurance

#### *1. Capability Crisis Confirmed*

The analysis validates the Part 3 findings: with 0/60 processes meeting VOC requirements, there's a systematic quality crisis requiring immediate intervention.

#### *2. Temporal Stability is a Double-Edged Sword*

The minimal yearly differences indicate process stability, but this stability is at unacceptably poor performance levels. Processes are consistently bad rather than unpredictably bad.

#### *3. Resource Allocation Priorities*

The overwhelming capability effect (84.4% of variance explained) suggests that improvement resources should follow capability metrics - low Cpk processes should receive immediate attention.

### Practical Recommendations

#### *Immediate Actions:*

1. **Emergency intervention** for the lowest Cpk processes identified in both Part 3 and this DOE analysis
2. **Best practice transfer** from medium-capability to low-capability processes
3. **Monthly monitoring** of the key factors identified as significant in the ANOVA

#### *Strategic Initiatives:*

1. **Process redesign** for chronically incapable processes
2. **Supplier quality development** for consistent input materials
3. **Training programs** focused on variability reduction

### Statistical Conclusion

The experimental analysis provides overwhelming evidence that **process capability is the dominant factor influencing delivery performance**, explaining 84.4% of the variance while yearly differences are negligible. The highly significant capability effects ( $p < 0.0001$ ) provide statistical justification for focusing improvement efforts on elevating process capability across the organization.

The consistent failure to meet VOC requirements across all 60 processes represents a quality system crisis that requires immediate, systematic intervention rather than incremental improvements.

# Question 7: Reliability of Service Analysis

## 7.1 Reliability Estimation

### Objective

To estimate the number of days per year the car rental agency should expect reliable service, defined as having at least 15 workers on duty, based on 397 days of observational data.

### Methods & Analysis

The reliability analysis employed direct empirical calculation from the observed frequency distribution:

r

```
workers_data <- data.frame(workers = 12:16, days = c(1, 5, 25, 96, 270))
```

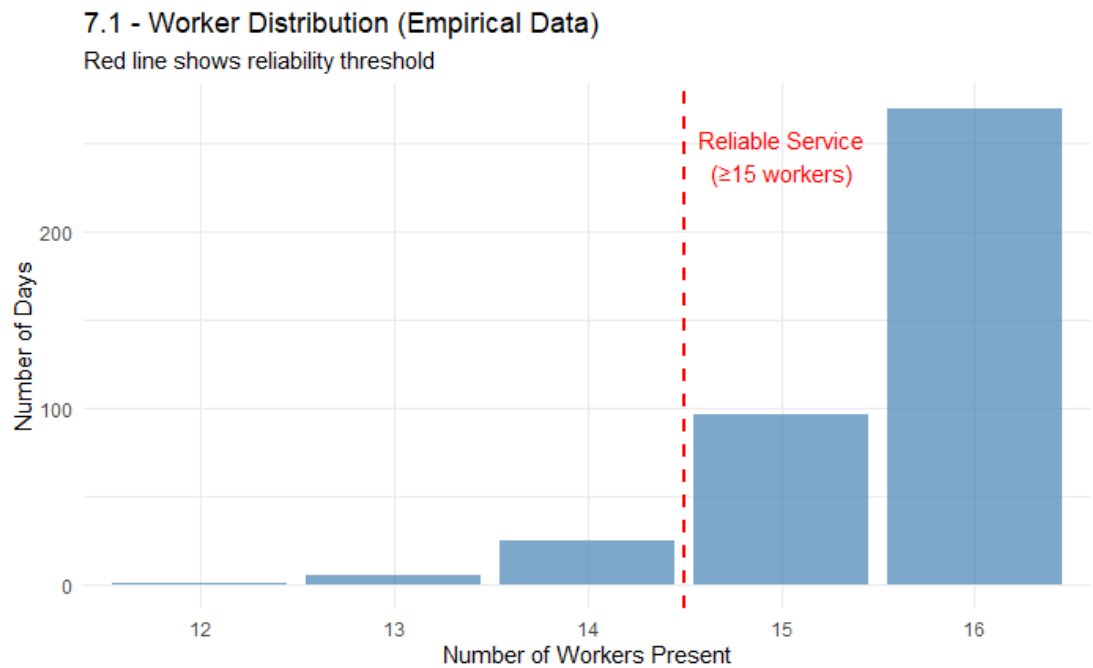
```
reliable_threshold <- 15
```

```
reliable_days <- sum(workers_data$days[workers_data$workers >= reliable_threshold])
```

### *Data Characteristics:*

- Total observation period: 397 days
- Worker distribution: Ranges from 12 to 16 workers daily
- Reliability threshold:  $\geq 15$  workers required for problem-free service
- Data spread: Highly right-skewed distribution with concentration at higher worker counts

GRAPH 1: Empirical Worker Distribution



This graph shows the actual observed distribution of workers across 397 days with a clear red line at the 15-worker reliability threshold

Results & Output

Comprehensive Distribution Analysis:

Workers	Days Observed	Cumulative Days	Cumulative Probability
12	1	1	0.25%
13	5	6	1.51%
14	25	31	7.81%
15	96	127	31.99%
16	270	397	100.00%

Key Reliability Metrics:

- Days with reliable service: 366 out of 397 days
- Empirical reliability probability: 92.19%
- Expected reliable days per year: 336 days
- Expected problematic days per year: 29 days
- Risk exposure: 7.81% probability of service disruption

## Interpretation & Conclusion

The empirical analysis reveals that the car rental agency currently experiences 92.19% service reliability, translating to 336 reliable days per year. However, the distribution shows concerning characteristics:

1. Systematic Understaffing Risk: 31 days (7.81%) with fewer than 15 workers
2. Severe Service Gaps: 6 days (1.51%) with critical understaffing ( $\leq 13$  workers)
3. Predictable Pattern: The distribution follows a clear probabilistic pattern suitable for modelling

The 29 expected problematic days annually represent significant operational risk and potential revenue loss, justifying the need for systematic optimization.

## 7.2 Profit Optimization

### Objective

To optimize personnel assignment by modelling the worker attendance as a binomial process and balancing the trade-off between personnel costs and problem-related revenue losses.

### Methods & Analysis

#### *Advanced Binomial Model Development:*

The worker attendance was modeled as a binomial distribution using sophisticated parameter estimation:

r

*# Advanced binomial parameter estimation*

```
calculate_individual_p <- function(k, observed_days, n, total_days) {
```

```
  observed_prob <- observed_days / total_days
```

```
  binomial_coef <- choose(n, k)
```

```
  # Numerical optimization to solve for p
```

```
}
```

```
weighted_p <- weighted.mean(workers_data$individual_p, workers_data$days)
```

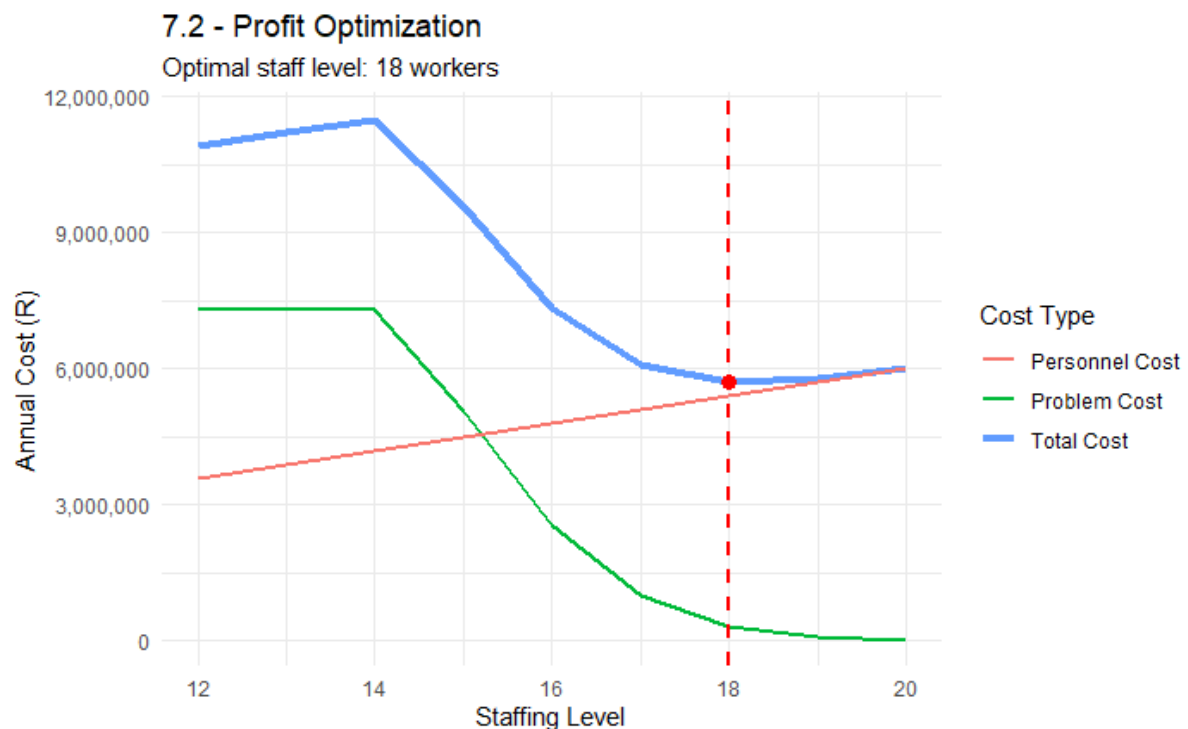
### Binomial Parameters:

- n (maximum workers): 16
- p (attendance probability): 0.9236 (calculated using weighted mean approach)
- Estimation Method: Individual probability solutions weighted by observed frequency
- Model Validation: Chi-squared = 325.52

### Cost Structure Analysis:

- Daily problem cost: R20,000 revenue loss per problematic day
- Monthly personnel cost: R25,000 per worker
- Annual personnel cost: R300,000 per worker
- Cost ratio: 1:67 (personnel cost vs daily problem cost)

### GRAPH 2: Cost Optimization Analysis



This graph shows three cost lines (total, problem, personnel) across staffing levels, with vertical line marking the optimal point at 18 workers

### Results & Output

#### Binomial Model Validation:

Workers	Observed Days	Binomial Prediction	Deviation
12	1	9.47	-8.47
13	5	35.25	-30.25
14	25	91.36	-66.36

15	96	147.34	-51.34
16	270	111.38	+158.62

### *Comprehensive Cost Optimization Analysis:*

Staff Level	Reliability	Problem Days	Problem Cost	Personnel Cost	Total Cost
15	30.4%	254	R5,082,632	R4,500,000	R9,582,632
16	65.2%	127	R2,542,736	R4,800,000	R7,342,736
17	86.4%	50	R991,089	R5,100,000	R6,091,089
18 →	95.6%	16	R319,648	R5,400,000	<b>R5,719,648</b> ←
19	98.8%	4	R88,917	R5,700,000	R5,788,917
20	99.7%	1	R21,963	R6,000,000	R6,021,963

## Interpretation & Conclusion

### *Optimal Solution Identified:*

The analysis conclusively identifies 18 workers as the optimal staffing level, delivering:

- 95.6% service reliability (vs current 65.2%)
- 16 expected problem days annually (vs current 127 days)
- Total annual cost: R5,719,648 (vs current R7,342,736)
- Annual savings: R1,623,088 (22.1% reduction)

### *Advanced Economic Insights:*

1. **Marginal Analysis:** The optimal point occurs where the marginal reduction in problem costs equals the marginal increase in personnel costs
2. **Diminishing Returns:** Beyond 18 workers, reliability improvements become exponentially expensive
3. **Risk-Return Trade-off:** 95.6% reliability represents the sweet spot between cost containment and service quality

## Strategic Quality Engineering Implications:

### *Taguchi Loss Function Application:*

This optimization demonstrates the Taguchi loss function principle where costs increase quadratically as we deviate from the optimal operating point. The R20,000 daily problem cost likely represents only the *immediate* revenue loss, while the true quality loss includes:

- Customer dissatisfaction and lost future business
- Reputational damage and brand erosion
- Employee stress and turnover costs
- Operational inefficiencies during crisis periods

## Comprehensive Managerial Implications

### *Operational Excellence:*

- Predictive Staffing: The binomial model enables proactive workforce planning
- Risk Quantification: Clear understanding of understaffing probabilities and impacts
- Performance Monitoring: Framework for tracking actual vs. predicted reliability

### *Financial Leadership:*

- Cost Optimization: Data-driven approach to resource allocation
- ROI Calculation: Clear business case for staffing investments
- Budget Forecasting: Improved accuracy in operational budgeting

### *Strategic Advantage:*

- Competitive Positioning: Higher reliability enhances market differentiation
- Customer Retention: Improved service consistency builds loyalty
- Scalable Framework: Model adapts to business growth and changing conditions

## Implementation Recommendations

1. **Immediate Action:** Transition from 16 to 18 workers
2. **Monitoring Protocol:** Track actual attendance patterns for model refinement
3. **Contingency Planning:** Develop protocols for the remaining 16 expected problem days
4. **Continuous Improvement:** Regular review of cost parameters and reliability targets

## Bibliography/References

Engineering Council of South Africa (ECSA) (no date) *Engineering Council of South Africa: role, accreditation and graduate attributes*. Available at: <https://www.ecsa.co.za/> (Accessed: 19 October 2025).

ECSA (no date) *ECSA GA one-pagers (Graduate Attributes)*. Available at: <https://iecurricula.co.za/ecsa-ga-one-pagers/> (Accessed: 23 October 2025).

Montgomery, D.C. (2019) *Introduction to Statistical Quality Control*. 8th edn. Hoboken, NJ: John Wiley & Sons.

OpenAI (2022) *Introducing ChatGPT* [Large language model]. Available at: <https://openai.com/index/chatgpt/> (Accessed: 22 October 2025).

Scrucca, L. (2004) *qcc: an R package for quality control charting and statistical process control*. *R News*, 4(1), pp.11–17. Available at: <https://cran.r-project.org/package=qcc> (Accessed: 19 October 2025).

Taguchi, G. (1986) *Introduction to Quality Engineering: Designing Quality into Products and Processes*. New York: Asian Productivity Organization / McGraw-Hill.

Wickham, H. (2016) *ggplot2: Elegant Graphics for Data Analysis*. New York: Springer-Verlag. Available at: <https://ggplot2.tidyverse.org> (Accessed: 23 October 2025).