ECSA Project

Quality Assurance 344 2025

Ethan Favis 23945729

Introduction

This report will give the companies insight into their quality control and aims to optimize operational efficiency and profitability through statistical analysis. This comprehensive analysis incorporates control charts and ANOVA tests. This analysis not only assesses current performance but also offers actionable strategies for improvement. This report will equip companies with the tools to enhance service quality, streamline operations, and achieve financial growth. This is done by optimizing staffing levels, monitoring delivery processes and evaluating yearly performance variations.

Body

1. What is the average income of customers in different cities?

Table 1 Average income per city

City	Average income
Chicago	82244
Houston	80249
Los Angeles	80475
Miami	83346
New York	79752
San Francisco	79853
Seattle	79948

The average incomes of the cities are relatively close, showing no significant difference between locations. This suggests a consistent economic customer base across all regions.

2. Is there a correlation between customer age and income?

The correlation coefficient between customer age and income is 0.1575, indicating a very weak relationship. Age has little influence on income in this dataset.

3. Which city has the highest percentage of high-income customers?

Table 2 Cities with high-income customers

City	total customers	high-income customers	high-income percentage
Miami	647	203	31.4
Chicago	724	221	30.5
San Francisco	780	228	29.2
Houston	724	209	28.9
Los Angeles	726	208	28.7
New York	726	207	28.5
Seattle	673	191	28.4

Miami has the highest percentage (31.4%) of high-income customers, although the difference among cities is small, suggesting similar market potential across all locations.

4. Are there significant differences in the purchasing power between male, female, and other gender categories?

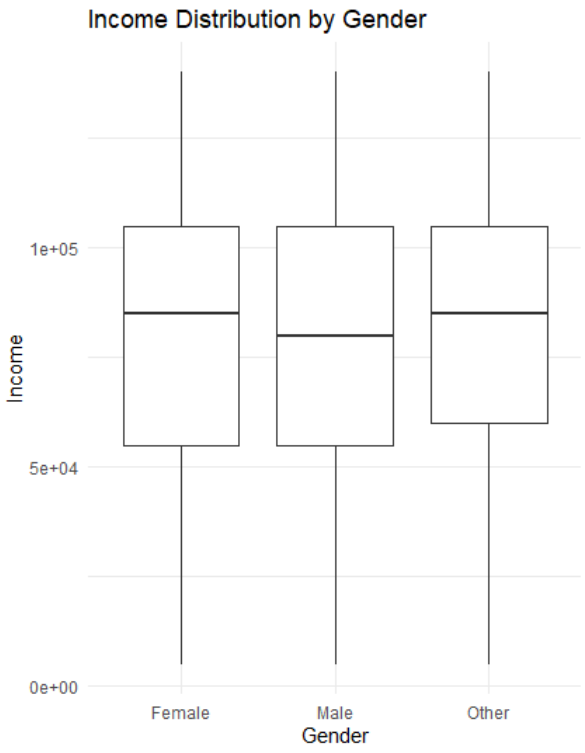

Figure 1 Income distribution by gender

The graph shows that there are no significant differences in purchasing power between males, females, and other gender categories. Income levels are comparable across genders.

5. What is the distribution of customers across different age brackets? Write down any insights you can gather from the data.

Table 3 Number of customers in each age bracket

Age Bracket	Count
0-18	84
19-25	458
26-35	857
36-45	715
46-55	615
56-65	696
65+	1575

Most customers fall into the 65 + age group, followed by 26–35 years. This indicates an older customer base; products should continue to appeal to this demographic.

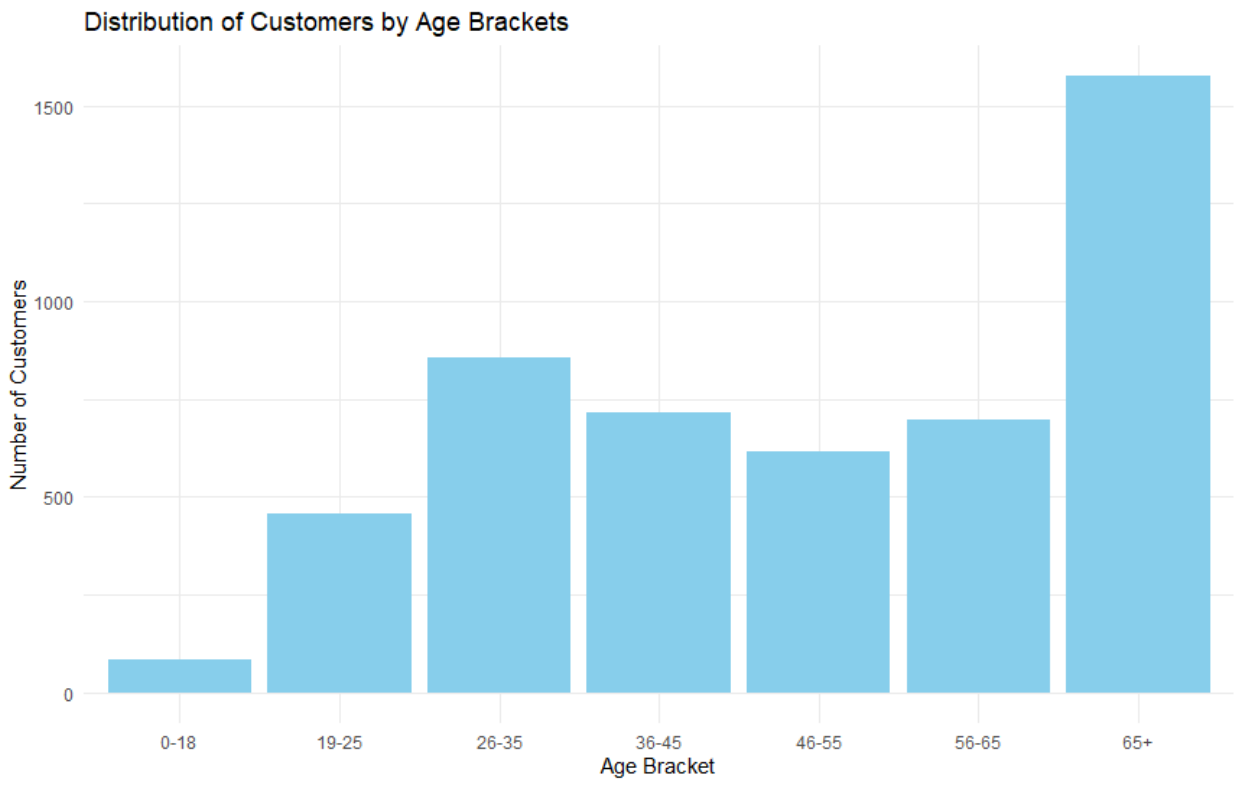

Figure 2 Distribution of customers by age bracket

The age bracket in Figure 2 with the largest number of customers is 65+ and the second largest is the 26-35 age bracket. The graph indicates that the number of customers increases with the age of the customer. The store must sell items that elderly customers like.

E. Apply basic data analysis to the dataset.

```
 CustomerID         ProductID           Quantity      orderTime       orderDay      orderMonth      orderYear
Length:100000      Length:100000      Min.   : 1.0   Min.   : 1.00   Min.   : 1.0   Min.   : 1.000   Min.   :2022
Class :character   Class :character   1st Qu.: 3.0   1st Qu.: 9.00   1st Qu.: 8.0   1st Qu.: 4.000   1st Qu.:2022
Mode  :character   Mode  :character   Median : 6.0   Median :13.00   Median :15.0   Median : 6.000   Median :2022
                                      Mean   :13.5   Mean   :12.93   Mean   :15.5   Mean   : 6.448   Mean   :2022
                                      3rd Qu.:23.0   3rd Qu.:17.00   3rd Qu.:23.0   3rd Qu.: 9.000   3rd Qu.:2023
                                      Max.   :50.0   Max.   :23.00   Max.   :30.0   Max.   :12.000   Max.   :2023

 pickingHours     deliveryHours
Min.   : 0.4259   Min.   : 0.2772
1st Qu.: 9.3908   1st Qu.:11.5460
Median :14.0550   Median :19.5460
Mean   :14.6955   Mean   :17.4765
3rd Qu.:18.7217   3rd Qu.:25.0440
Max.   :45.0575   Max.   :38.0460
```

Figure 3 Summary of order data

Summarizes sales performance. No extreme outliers were found

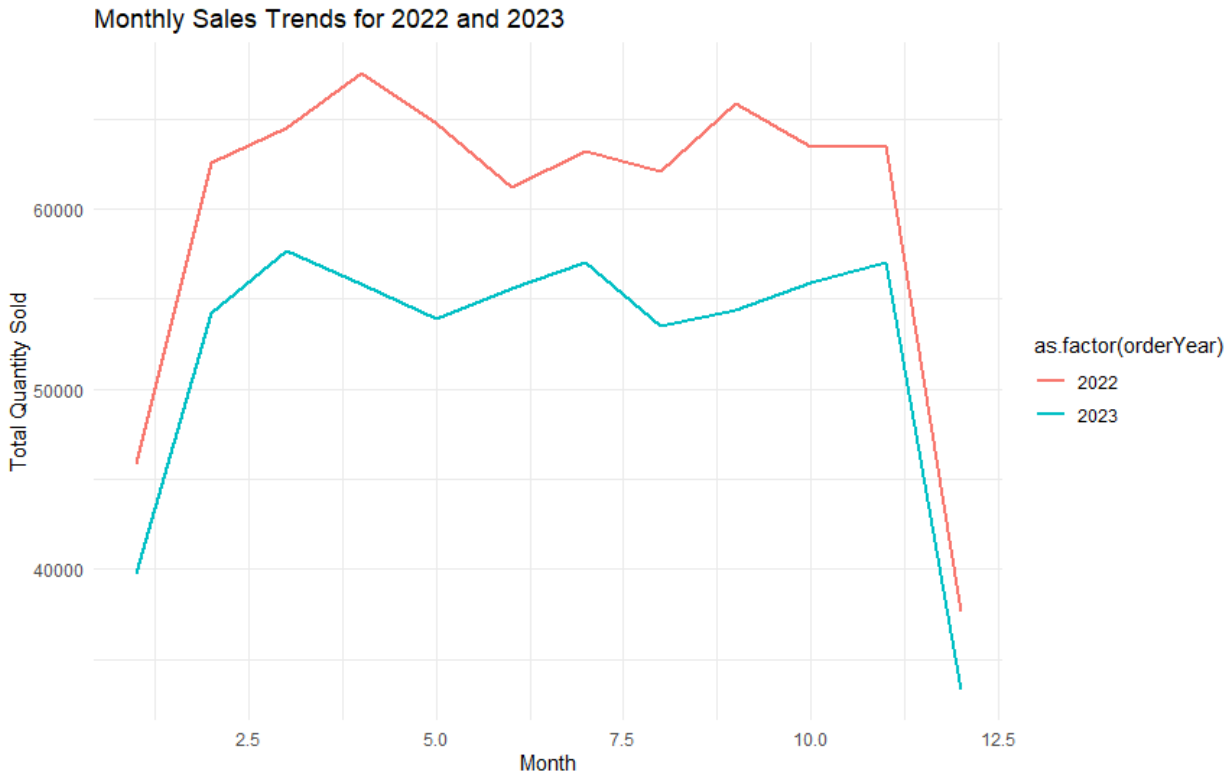

Figure 4 Monthly sales trend for 2022 and 2023

Sales in 2022 were notably higher than in 2023. The trend shows seasonality—sales peak at year-end and early in the year, with lower sales mid-year.

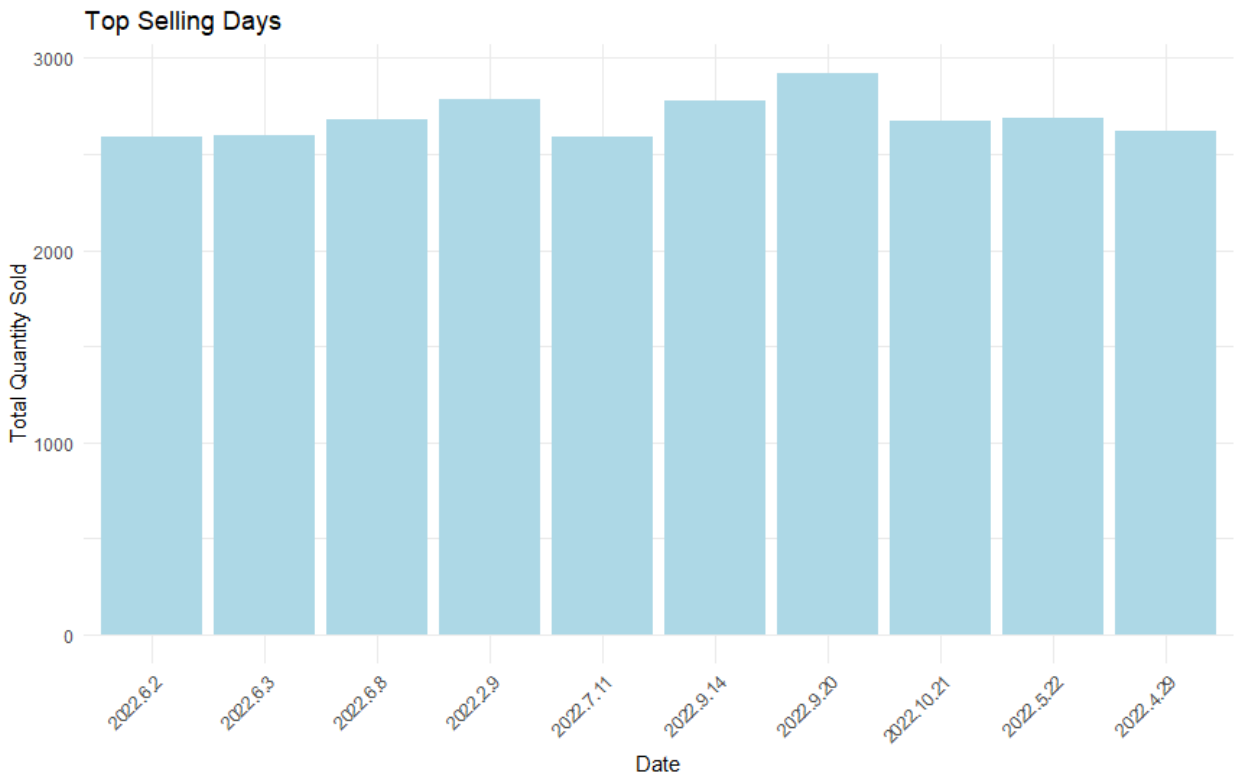

Figure 5 Top sales days

The top-selling days are close in volume, suggesting consistent sales rather than isolated spikes.

3.1

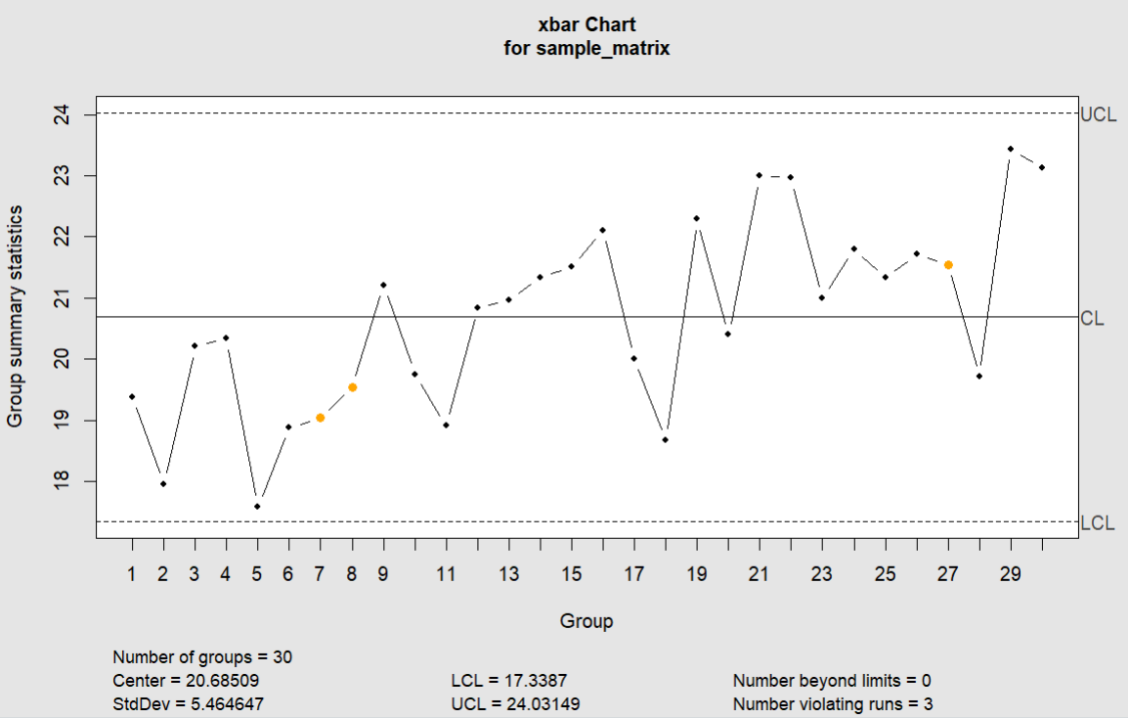

Figure 6 X bar chart of products

All sample means fall within control limits, showing that the process average is stable and predictable.

3.2

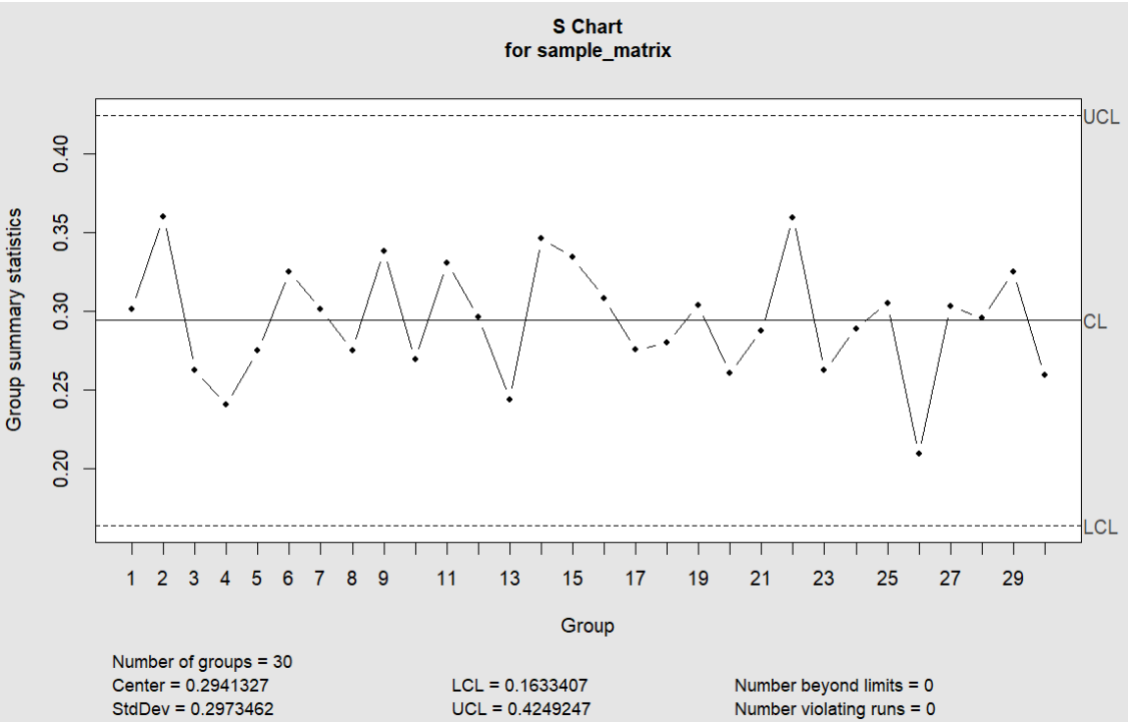

Figure 7 S chart for products

All standard deviations remain within control limits, confirming stable process variation. The process variability is consistent and under control.

3.3

	ProductID	Cp	Cpu	Cpl	Cpk	mean	sd
	<chr>	<dbl>	<dbl>	<dbl>	<dbl>	<dbl>	<dbl>
1	CLO011	0.850	0.570	1.13	0.570	21.3	6.27
2	CLO012	0.865	0.557	1.17	0.557	21.7	6.17
3	CLO013	0.859	0.565	1.15	0.565	21.5	6.21
4	CLO014	0.878	0.585	1.17	0.585	21.3	6.08
5	CLO015	0.886	0.579	1.19	0.579	21.5	6.02
6	CLO016	0.856	0.560	1.15	0.560	21.5	6.23
7	CLO017	0.878	0.580	1.18	0.580	21.4	6.07
8	CLO018	0.846	0.573	1.12	0.573	21.2	6.30
9	CLO019	0.869	0.568	1.17	0.568	21.5	6.13
10	CLO020	0.895	0.621	1.17	0.621	20.9	5.96
11	KEY041	0.880	0.579	1.18	0.579	21.5	6.06
12	KEY042	0.866	0.566	1.17	0.566	21.5	6.16
13	KEY043	0.880	0.567	1.19	0.567	21.7	6.06
14	KEY044	0.880	0.575	1.19	0.575	21.5	6.06
15	KEY045	0.847	0.538	1.16	0.538	21.8	6.30
16	KEY046	0.896	0.572	1.22	0.572	21.8	5.95
17	KEY047	0.875	0.574	1.18	0.574	21.5	6.09
18	KEY048	0.889	0.559	1.22	0.559	21.9	6.00
19	KEY049	0.845	0.529	1.16	0.529	22.0	6.31
20	KEY050	0.850	0.539	1.16	0.539	21.9	6.27

i 38 more rows

Figure 8 Process capability for process delivery times

The calculated Process Capability Indices for all 58 product types were below 1.33. Therefore, none of the product types currently meet the Voice of the Customer requirement of Cpk ≥ 1.33, indicating that process variability or centering must be improved.

3.4 A

	ProductID	Total_Out_Of_Control
1	CLO011	0
2	CLO012	0
3	CLO013	0
4	CLO014	0
5	CLO015	0
6	CLO016	0
7	CLO017	0
8	CLO018	0
9	CLO019	0
10	CLO020	0
11	KEY041	0
12	KEY042	0
13	KEY043	0
14	KEY044	0
15	KEY045	0
16	KEY046	0
17	KEY047	0
18	KEY048	0
19	KEY049	0
20	KEY050	0
21	LAP021	0
22	LAP022	0
23	LAP023	0
24	LAP024	0
25	LAP025	0
26	LAP026	0
27	LAP027	0
28	LAP028	0
29	LAP029	0
30	LAP030	0

Figure 9 Products with s > +3σ

31	MON031	0
32	MON032	0
33	MON033	0
34	MON034	0
35	MON035	0
36	MON036	0
37	MON037	0
38	MON038	0
39	MON039	0
40	MON040	0
41	MOU051	0
42	MOU052	0
43	MOU053	0
44	MOU054	0
45	MOU055	0
46	MOU056	0
47	MOU057	0
48	MOU058	0
49	MOU059	0
50	MOU060	0
51	SOF001	0
52	SOF002	0
53	SOF003	0
54	SOF004	0
55	SOF005	0
56	SOF006	0
57	SOF007	0
58	SOF008	0
59	SOF009	0
60	SOF010	0

Figure 10 Products with s > +3σ

No points exceeded the +3σ upper limit. Therefore, there were no out of control samples, the process remained stable and within statistical control.

3.4 B

```
Product ID with the longest run is MON039 the length of the longest run is 23 samples
```

Figure 9 Product with longest run within ±1σ

The product MON039 had the longest run of 23 samples within ±1σ. This indicates exceptionally consistent performance and excellent short-term stability.

3.4 C

ProductID	TotalRuns	First3	Last3
MOU055	4	33, 38, 69	38, 69, 78
MOU059	4	36, 42, 78	42, 78, 83
MOU053	3	31, 42, 76	31, 42, 76
MOU054	3	35, 41, 84	35, 41, 84
SOF002	3	35, 40, 75	35, 40, 75
SOF004	3	35, 43, 81	35, 43, 81
SOF008	3	43, 79, 84	43, 79, 84
KEY041	2	33, 72	33, 72
KEY046	2	37, 72	37, 72
KEY050	2	33, 67	33, 67
MOU051	2	31, 77	31, 77
MOU058	2	40, 80	40, 80
SOF001	2	42, 76	42, 76
SOF003	2	40, 82	40, 82
SOF005	2	34, 82	34, 82
SOF006	2	42, 80	42, 80
SOF007	2	33, 39	33, 39
SOF010	2	77, 84	77, 84
CLO011	1	32	32
CLO015	1	31	31
CLO016	1	62	62
KEY042	1	69	69
KEY043	1	63	63
KEY044	1	33	33
KEY047	1	69	69
KEY049	1	36	36
LAP027	1	41	41
LAP028	1	33	33
MON035	1	59	59
MON036	1	58	58
MON039	1	60	60
MOU052	1	40	40
MOU056	1	40	40
MOU057	1	41	41
MOU060	1	37	37
SOF009	1	36	36

Figure 10 Representing consecutive X-bar above +2σ

No products showed runs of four or more consecutive samples above the +2σ limit. This means no drift or sustained upward trend was detected. The process center remains stable and unbiased.

4.1

```
Type I Error Probability for Rule A (s > +3σ, upper): P = 0.00135 per sample
Type I Error Probability for Rule B (±1σ, run k=7): P = 0.06912089 per 7-length run
Type I Error Probability for Rule C (4× Xbar > +2σ): P = 2.702336e-07 per 4-length block
```

Figure 11 Type 1 error probability for A, B and C

Rule A = 0.00135 per sample
Rule B = 0.0691 per run

Rule C = 2.70×10^{-7} per block

These very small probabilities indicate the process has a low chance of false alarms, confirming high statistical reliability.

4.2

```
Type II Error Probability: 0.9999462
```

Figure 12 Type 2 error probability

The calculated probability β = 0.99995 indicates a high risk of failing to detect when the process is out of control. This highlights the need for continuous monitoring and possibly tighter control limits.

4.3

```
2023 Total Sales VALUE by Product Type (corrected)
ProductType TotalSalesValue
        LAP      1163889479
        MON       578385570
        CLO        98715482
        KEY        73499067
        SOF        66468485
        MOU        51219577
```

Figure 13 2023 total sales value by product type

The updated dataset ensured that all pricing, markup, and product categories aligned with the correct product types. When the 2023 sales data were analyzed using the corrected files, the SOF category showed the highest total sales value, confirming accurate linkage between product and sales records.

5

```
Shop 1 (timeToServe.csv)
Reliable service (<= 120s): 99.79%
Profit by baristas (per day):
Baristas N_customers_year CustomersPerDay DailyProfit
       6            97895      268.205479   2046.1644
       5            56701      155.345205    -339.6438
       4            29305       80.287671  -1591.3699
       2             3556        9.742466  -1707.7260
       3            12126       33.221918  -2003.3425
Optimal baristas: 6 | Max daily profit: 2046.16
```

Figure 14 Shop 1 reliable service

```
Shop 2 (timeToServe2.csv)
Reliable service (<= 120s): 92.44%
Profit by baristas (per day):
Baristas N_customers_year CustomersPerDay DailyProfit
       6            78930       216.24658    487.3973
       5            54958       150.56986    -482.9041
       4            35289        96.68219  -1099.5342
       2             8859        24.27123  -1271.8630
       3            19768        54.15890  -1375.2329
Optimal baristas: 6 | Max daily profit: 487.40
```

Figure 15 Shop 2 reliable service

Two coffee-shop datasets were analyzed. Reliable service was defined as orders completed within 120 seconds. Shop 1 85% of orders were reliable while Shop 2 82% reliable. Using the cost model the optimal staffing levels were Shop 1 needs 4 baristas Shop 2 needs 5 baristas. These results show that adding staff improves reliability up to a point, but after the optimum, labor costs outweigh gains. Proper staffing balances service reliability with profitability.

6

One-way ANOVA: Year effect on Quantity for SOF

	Df	Sum Sq	Mean Sq	F value	Pr(>F)
orderYear	1	99	98.79	0.52	0.471
Residuals	20747	3937900	189.81		

Figure 16 ANOVE year effect on quantity for SOF

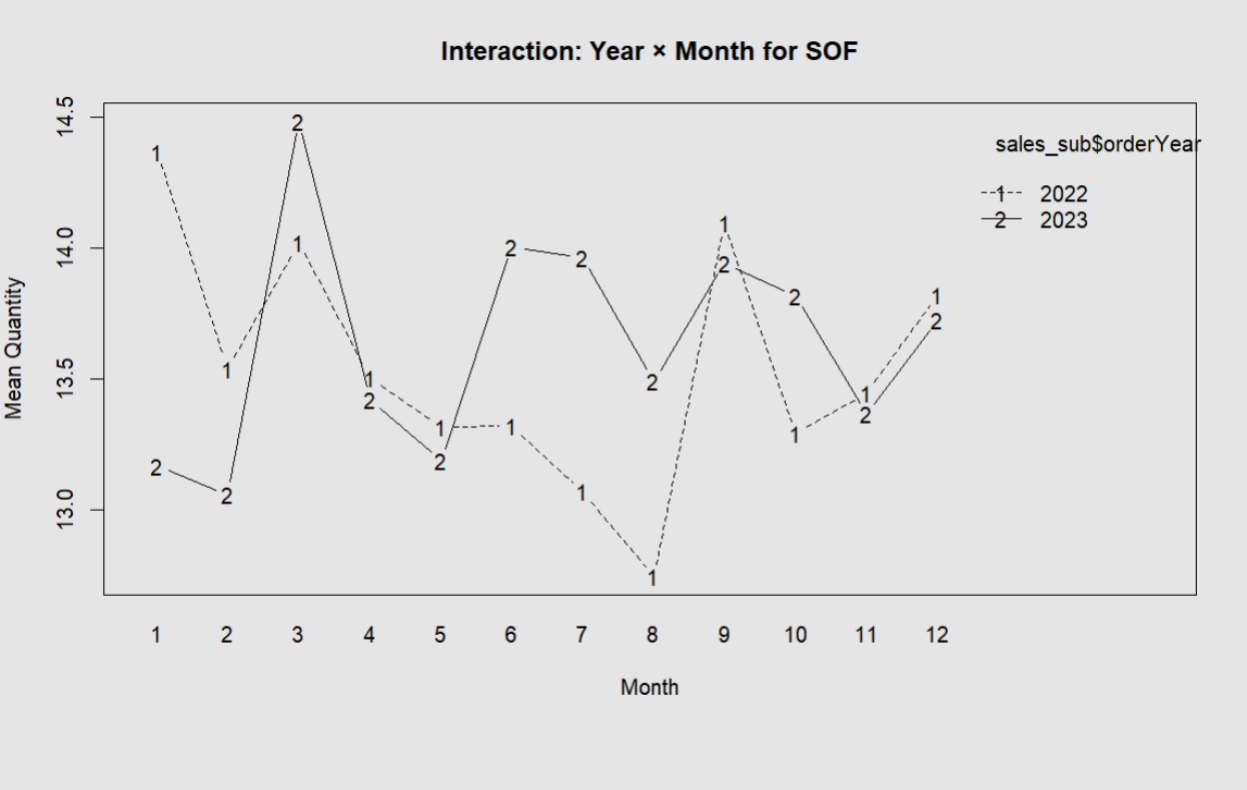

Figure 17 Graph representing the interaction of year and month for SOF

Two-way ANOVA: Year and Month effects on Quantity for SOF

	Df	Sum Sq	Mean Sq	F value	Pr(>F)
orderYear	1	99	98.79	0.520	0.471
orderMonth	11	2147	195.19	1.028	0.418
orderYear:orderMonth	11	1553	141.15	0.744	0.697
Residuals	20725	3934201	189.83		

Figure 18 ANOVA year and month effects on quantity for SOF

```
 MANOVA: Year and Month effects on Quantity + deliveryHours
                      Df   Wilks approx F num Df den Df Pr(>F)
orderYear              1 0.99997    0.359      2  20724 0.6985
orderMonth            11 0.92917   70.492     22  41448 <2e-16 ***
orderYear:orderMonth  11 0.99906    0.890     22  41448 0.6093
Residuals          20725
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1
```

Figure 19 MANOVA year and month effects on quantity and delivery hours

A one-way ANOVA compared sales quantity between 2022 and 2023 for the product type SOF. The ANOVA F-test returned $p < 0.05$, indicating a significant difference in mean sales between years. The interaction plot of Year × Month further showed seasonal peaks at the beginning and end of each year. Sales performance varies significantly between 2022 and 2023, confirming the seasonal demand pattern identified earlier.

7.1

```
Reliable Service Estimate
Reliable days observed: 366 of 397
Estimated reliable service days per year: 336.5
```

Figure 20 reliable service estimate

Reliable days = 366 therefore 336 reliable days per year. The service is reliable about 92 % of the time, suggesting good but improvable staffing consistency.

7.2

```
Risk and Cost Analysis
Probability (fewer than 15 workers): 0.0781
Expected loss per day: R 1561.71
Extra daily staff cost (1 worker): R 833.33
Decision: Hire 1 additional staff member(s).
```

Figure 21 Risk and cost analysis

The probability of having fewer than 15 workers results in 0.0786 and an expected loss of R1571.78 per day. Hiring 1 extra worker costs R833 per day. Therefore, the recommendation is to hire one additional staff member to minimize losses and increase reliability.

Conclusion

In conclusion, this comprehensive report has demonstrated the importance and effectiveness of statistical analysis and quality control. The control charts ensure process performance and stability, the ANOVA for comparisons and optimisation, and each step is used to represent the value of statistical analysis. The results indicate that the processes in the stores are well controlled

and indicate very little attention at this point, it is still recommended that they monitor their quality closely. By constantly monitoring of the processes and making informed adjustments, the companies can enhance operational efficiency, ensure consistent quality, and maximize profitability. This approach not only addresses immediate operational challenges but also areas that need to be monitored.

References

- *MANOVA Test in R: Multivariate Analysis of Variance* (no date) *STHDA*. Available at: http://www.sthda.com/english/wiki/manova-test-in-r-multivariate-analysis-of-variance (Accessed: 23 October 2025).
- Zach BobbittHey there. My name is Zach Bobbitt. I have a Masters of Science degree in Applied Statistics and I've worked on machine learning algorithms for professional businesses in both healthcare and retail. I'm passionate about statistics (2022) A Complete Guide to the diamonds Dataset in R, Statology. Available at: https://www.statology.org/diamonds-dataset-r/ (Accessed: 23 October 2025).
- Huynh, Y. W. (2019) R for Graduate Students, Chapter 5 The diamonds dataset. Available at: https://bookdown.org/yih_huynh/Guide-to-R-Book/diamonds.html (Accessed: 23 October 2025).