Quality Assurance

ECSA FINAL

Stellenbosch University

Wihan Meyer - 27024466

24 Oct 2025

Table of Contents

Part 2

Introduction

1. CUSTOMER

```
  CustomerID Gender   Age Income City
  <chr>      <chr>  <dbl>  <dbl> <chr>
1 CUST001    Male      16  65000 New York
2 CUST002    Female    31  20000 Houston
3 CUST003    Male      29  10000 Chicago
4 CUST004    Male      33  30000 San Francisco
5 CUST005    Female    21  50000 San Francisco
```

Example of Data

Variable	Age	Income
Min	16	5000
Q1	33	55000
Median	51	85000
Q3	68	105000
Max	105	140000
Mean	51.5538	80797.0000
SD	21.2161	33150.1067

Table 1: Customer Numerical Data

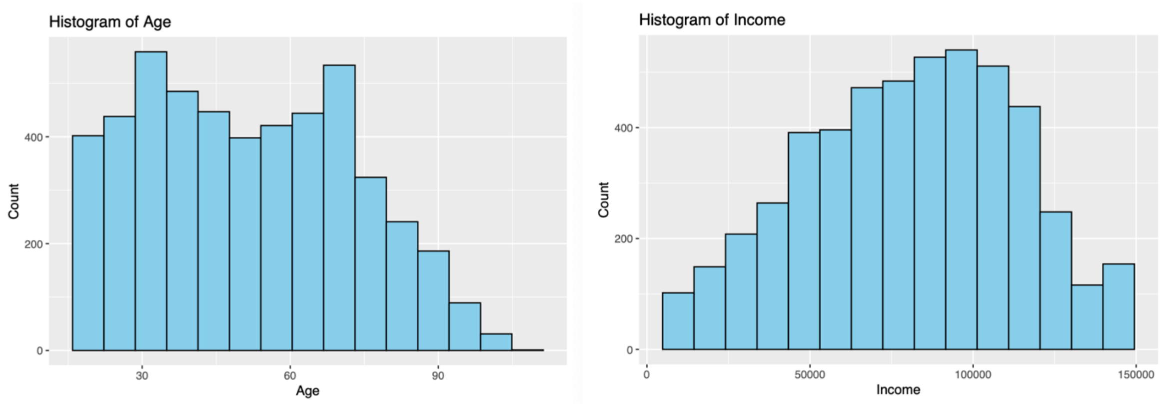
The table shows that age follows a roughly normal and balanced distribution (considering <16 cannot be a customer), indicating moderate variation among respondents. In contrast, income is skewed toward mid- to high-earning individuals with substantial spread (SD = 33 150). This suggests that income may play a more significant role than age in defining the customer base.

```
--- Gender ---
       col Freq
1 Female 2432
2   Male 2350
3  Other  218
```

```
--- City ---
             col Freq
1        Chicago  724
2        Houston  724
3    Los Angeles  726
4          Miami  647
5       New York  726
6  San Francisco  780
7        Seattle  673
```

Table 2: Customer Non-Numerical Data

The customer base appears well balanced in terms of both gender and city representation.

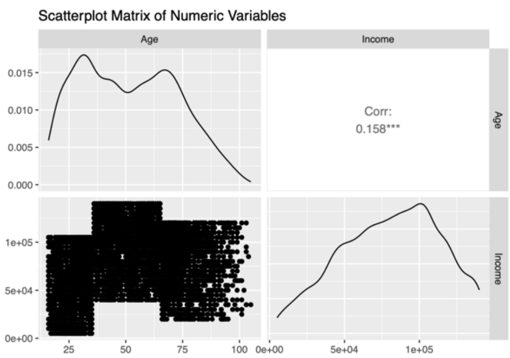


Univariate Data: Customers

Age is evenly distributed between 16 and 75, after which a decline occurs, which is typical of a normal population pattern (Department of Economic and Social Affairs, 2017). Overall, this reflects a standard population distribution excluding those under 16, who are unlikely to be customers. Income shows a wide range, with values increasing toward the mid- to high-income group, indicating that most customers fall within this bracket.

Relational Data: Customers

In terms of income, the data shows no indication that city or gender have an effect. What we do see is a high percentage of middle-aged (35–65) customers with higher incomes, which could suggest that this group forms the core target market with the strongest purchasing power and brand loyalty.



Scatterplot Matrix: Customers

In terms of the scatterplot matrix there is no linear correlation between the age and income of the customers.

2. PRODUCTS

```
  ProductID Category           Description      SellingPrice Markup
  <chr>     <chr>              <chr>                   <dbl>  <dbl>
1 SOF001    Software           coral matt               512.   25.0
2 SOF002    Cloud Subscription cyan silk                505.   10.4
3 SOF003    Laptop             burlywood marble         494.   16.2
4 SOF004    Monitor            blue silk                543.   17.2
5 SOF005    Keyboard           aliceblue wood           516.   11.0
```

Example of Data

Variable	Min	Q1	Median	Q3	Max	Mean	SD
SellingPrice	350.45	512.18	794.18	6416.66	19725.18	4493.59	6503.77
Markup	10.13	16.14	20.34	25.71	29.84	20.46	6.07

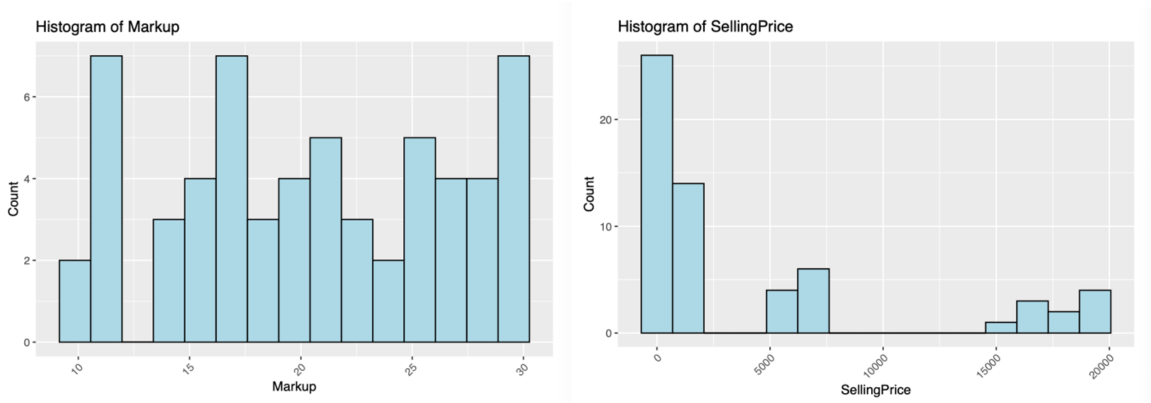
Table 3: Product Numerical Data

The selling price shows high variability, ranging from 350 to 19 725, indicating a wide variety of products. The median (794) is much lower than the mean (4 493), suggesting a right-skewed distribution. This means most products are lower priced, while a smaller number of high-end items contribute disproportionately to overall sales. The markup is evenly distributed around 20.46%.

```
--- Category ---
                     col Freq
1 Cloud Subscription   10
2           Keyboard   10
3             Laptop   10
4            Monitor   10
5              Mouse   10
6           Software   10
```

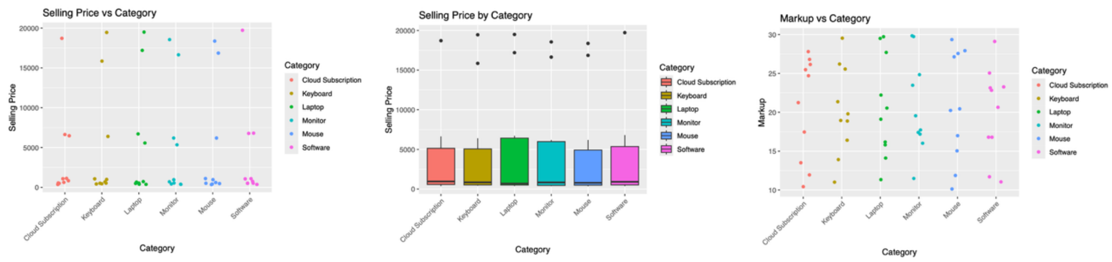
Table 4: Product Non-Numerical Data

The store carries equal amounts of each product type.



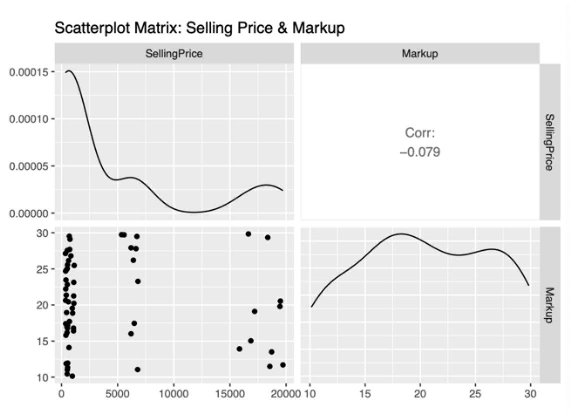
Univariate Data: Products

Markup values are fairly evenly distributed, showing consistent pricing margins across products. In contrast, selling prices are highly uneven, with a large concentration of sales in the sub-1 000 range and a smaller, distinct surge in high-end products priced above 15 000.



Relational Data: Products

The relational data shows no clear relationship between selling price, markup, and product type. While this may initially appear uninformative, it actually suggests a well-balanced product mix, indicating that maintaining variety across product types can help maximize both sales and overall profitability.



Scatterplot Matrix: Products

The scatterplot matrix shows no linear correlation between selling price and markup. This is insightful, as it suggests that the store does not rely solely on one pricing strategy. While some businesses focus on high-price, high-markup niche products and others on high-volume, low-markup commercial items, this store appears to maintain a balanced position between the two approaches.

3. PRODUCTS HEADOFFICE

```
  ProductID Category Description            SellingPrice Markup
  &lt;chr&gt;     &lt;chr&gt;    &lt;chr&gt;                         &lt;dbl&gt;  &lt;dbl&gt;
1 SOF001    Software coral silk                     522.   15.6
2 SOF002    Software black silk                     467.   28.4
3 SOF003    Software burlywood marble               496.   20.1
4 SOF004    Software black marble                   389.   17.2
5 SOF005    Software chartreuse sandpaper           483.   17.6
```

Example of Data

Variable	Min	Q1	Median	Q3	Max	Mean	SD
SellingPrice	290.52	495.94	797.22	5843.33	22420.14	4410.96	6463.82
Markup	10.06	15.84	20.58	24.84	30.00	20.39	5.67

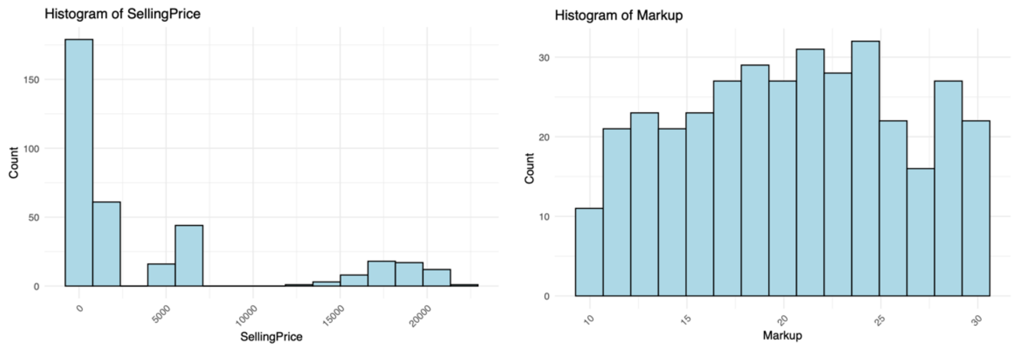
Table 5: Product Headoffice Numerical Data

Selling prices display significant variation, spanning from 290 to 22 420, which reflects a broad product range. The median (797) falls well below the mean (4 411), pointing to a right-skewed pattern where most items are lower priced, and a few high-value products elevate the average. Markup values remain consistent at around 20%, indicating stable profit margins across the product range.

```
                 col Freq
1 Cloud Subscription   60
2           Keyboard   60
3             Laptop   60
4            Monitor   60
5              Mouse   60
6           Software   60
```

Table 6: Product Headoffice Non-Numerical Data

Again, a balanced offering including the same number of products within each type.

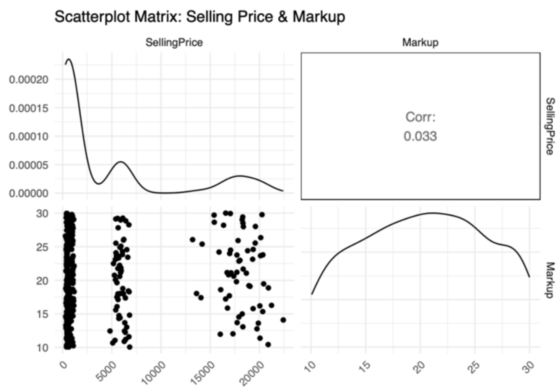


Univariate Data: Products Headoffice

The markup distribution appears fairly uniform, indicating consistent profit margins across products. In contrast, selling prices are heavily concentrated at the lower end, with most products priced below 1 000 and only a few high-value items extending beyond 15 000. This suggests that while the product range includes premium options, the majority of sales come from lower-priced items, however most of the turnover comes from the higher end products.

Relational Data: Products Headoffice

The relational data shows no clear relationship between selling price, markup, and product type. We can confidently confirm this by looking at the difference in averages (none). As said with the products in store "While this may initially appear uninformative, it actually suggests a well-balanced product mix, indicating that maintaining variety across product types can help maximize both sales and overall profitability."



Scatterplot Matrix: Products Headoffice

The scatterplot matrix shows no linear correlation between selling price and markup. This confirms the company's pricing strategy: " While some businesses focus on high-price, high-markup niche products and others on high-volume, low-markup commercial items, this store appears to maintain a balanced position between the two approaches.”

4. SALES

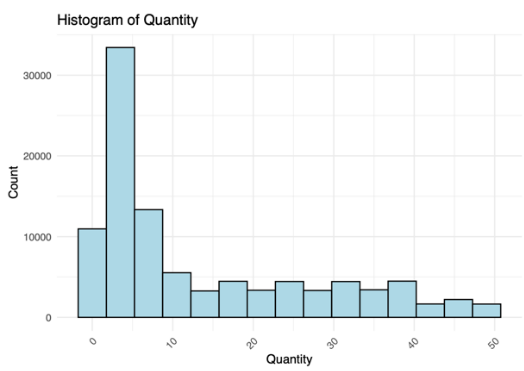
```
  CustomerID ProductID Quantity orderTime orderDay orderMonth orderYear
  <chr>      <chr>        <dbl>     <dbl>    <dbl>      <dbl>     <dbl>
1 CUST1791   CLO011          16        13       11         11      2022
2 CUST3172   LAP026          17        17       14          7      2023
3 CUST1022   KEY046          11        16       23          5      2022
4 CUST3721   LAP024          31        12       18          7      2023
5 CUST4605   CLO012          20        14        7          2      2022
```

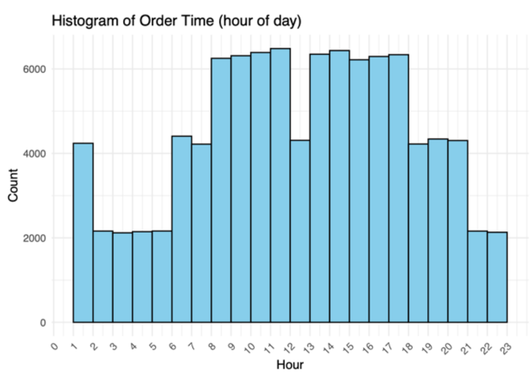
Example of Data

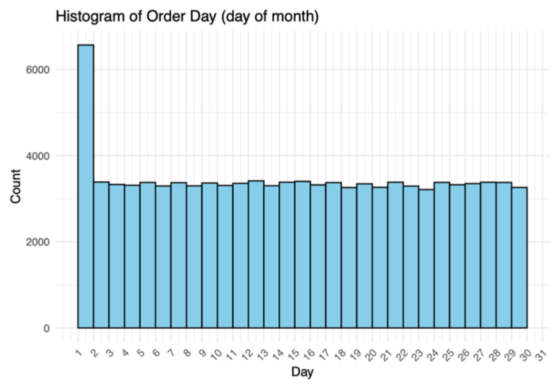
Variable	Min	Q1	Median	Q3	Max	Mean	SD
Quantity	1.00	3.00	6.00	23.00	50.00	13.50	13.76
orderTime	1.00	9.00	13.00	17.00	23.00	12.93	5.50
orderDay	1.00	8.00	15.00	23.00	30.00	15.50	8.65
orderMonth	1.00	4.00	6.00	9.00	12.00	6.45	3.28
orderYear	2022.00	2022.00	2022.00	2023.00	2023.00	2022.46	0.50
pickingHours	0.43	9.39	14.05	18.72	45.06	14.70	10.39
deliveryHours	0.28	11.55	19.55	25.04	38.05	17.48	10.00

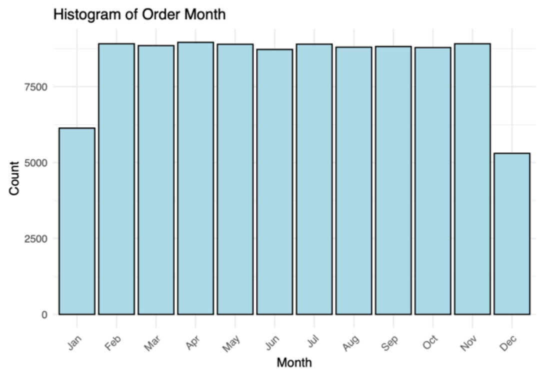
Table 7: Sales Numerical Data

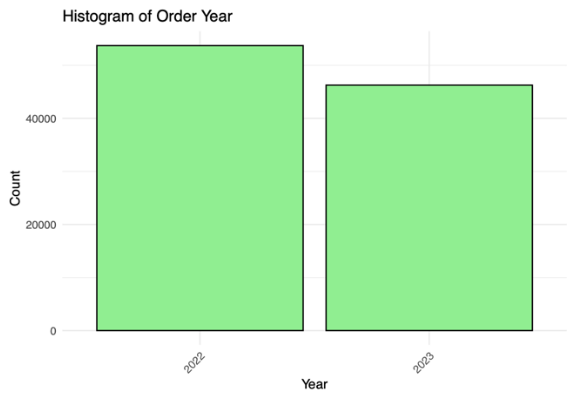
The high range and SD of quantity show they have high variability in order size. This could lead to issues with bottlenecks. Order Time, Day, month and year are evenly spread over the given time period (no daily, monthly, yearly seasonality).

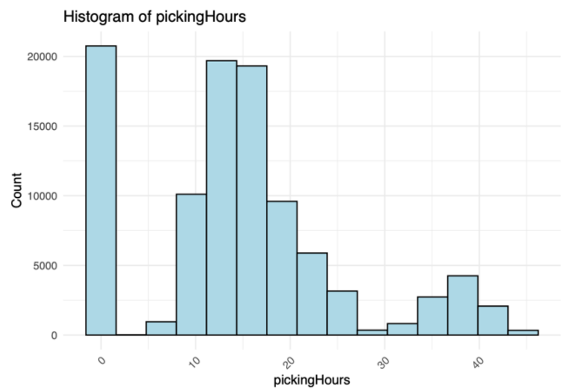


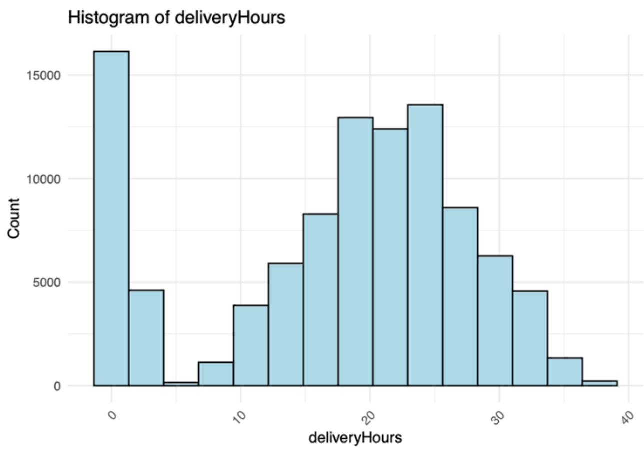






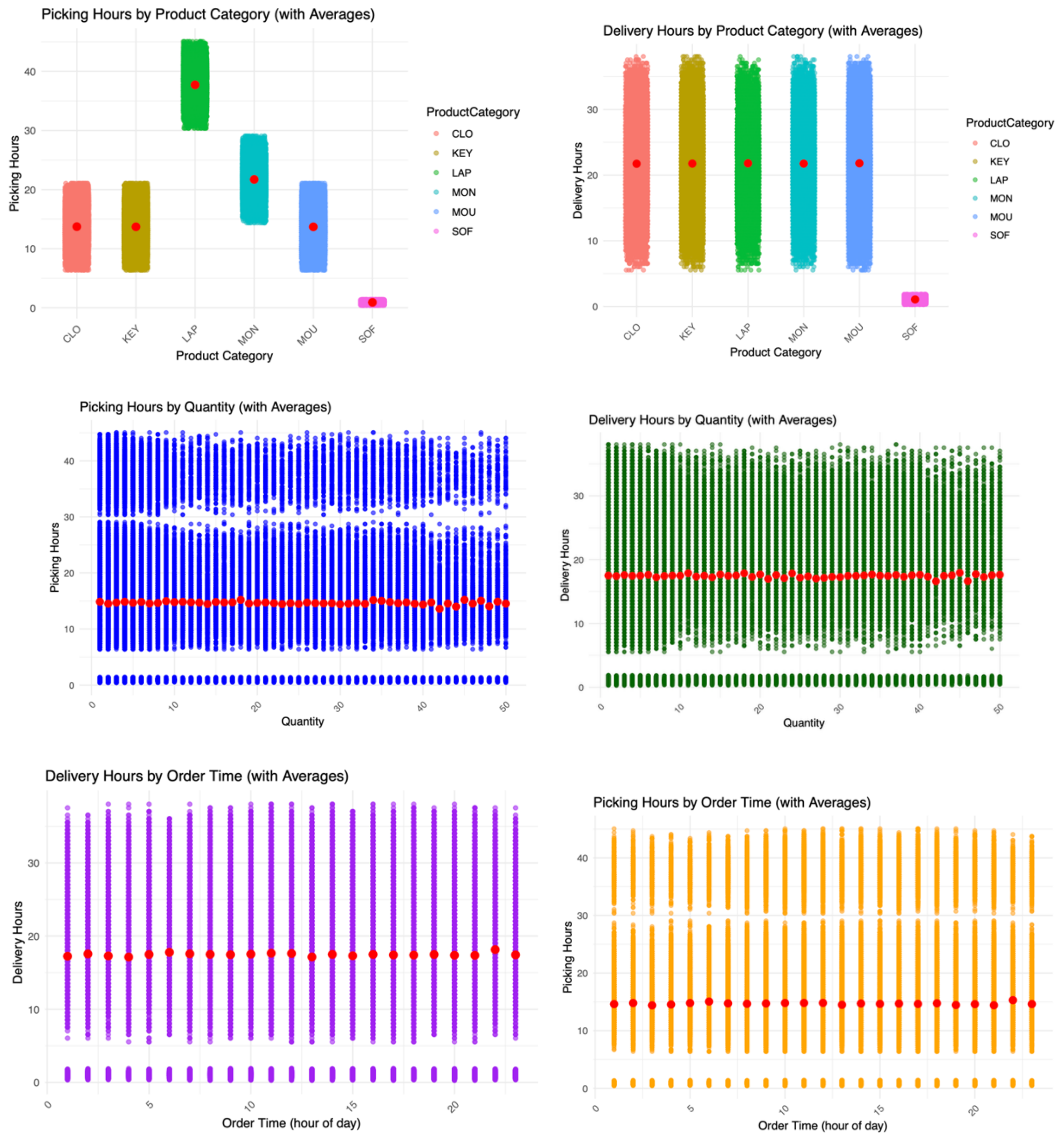






Univariate Data: Sales

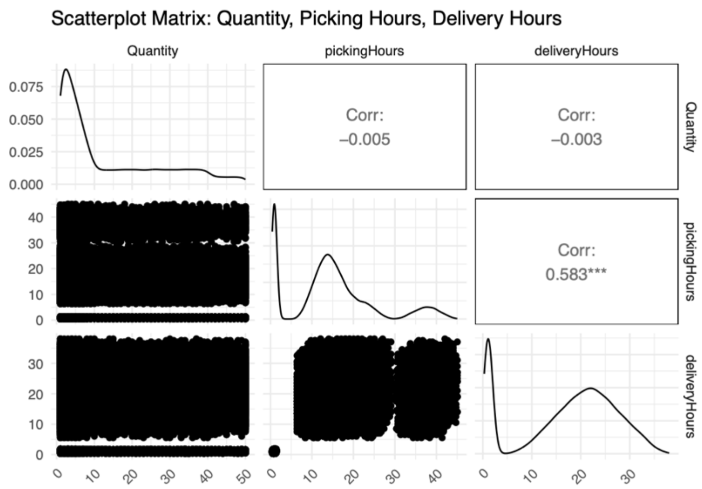
Order sizes are spread across the full 0–50 range, with a clear peak around five items per order. Orders are placed consistently throughout the day, with higher volumes during business hours and fewer outside them. Across the month, sales remain steady, though there is a pronounced spike on the first day, likely linked to commercial or bulk purchasing patterns rather than typical consumer behaviour. No strong seasonality trends appear in the monthly data, aside from a slight dip in December and January. This, together with the day-one surge, suggests the company's sales are primarily driven by business clients. A notable decrease of roughly 10 000 orders from 2022 to 2023 is visible and warrants further investigation. Delivery hours show two clusters — one group of nearinstant deliveries (likely digital or software products) and another concentrated around 20 hours, representing standard physical product deliveries.



Relational Data: Sales

We see several important insights from the relational data. Firstly, picking hours for laptops are significantly higher, with monitors following as the second most time-consuming category. Despite this, delivery hours remain consistent across all product types, while software items—being digital—require virtually no delivery time. Interestingly, order quantity and order time show no noticeable impact on either picking or delivery duration. This indicates that the company's fulfilment process is well-optimized and capable of handling varying order sizes and times without major slowdowns.

The stability in delivery times across categories suggests that logistics operations are efficient and standardized, regardless of product complexity. The variation in picking hours is likely due to differences in handling or packaging requirements rather than process inefficiency. Overall, the data points to a reliable and scalable system with minimal risk of bottlenecks, even as order volume or product diversity increases.



Scatterplot Matrix: Sales

Although the scatterplot matrix shows a clear correlation between delivery hours and picking hours, this relationship is not particularly meaningful. The two variables share similar averages and naturally increase over time, which explains their apparent connection. Rather than indicating a causal relationship, the trend simply reflects two parallel processes within the same operational system. Both metrics represent the system's overall capability, which appears to decline gradually over time — a point that will be explored further in the control charts.

CONCLUSION:

The overall analysis shows that the company maintains a balanced and diverse customer and product base. Age and gender distributions are even, while income is concentrated among middle-aged, higher-earning customers**,** identifying this group as the core target market**.** Product data indicates a wide range of selling prices but consistent markups around 20%**,** suggesting stable pricing policies and a well-structured product mix.

Sales analysis reveals high variability in order size but no strong seasonality, with demand peaks on the first day of each month—consistent with commercial purchasing behaviour. Operational data highlights an efficient fulfilment system: picking and delivery hours remain stable across products and unaffected by order size or timing. Although delivery and picking times appear correlated, this mainly reflects parallel process trends rather than inefficiency.

Overall, the company demonstrates operational stability, balanced product offerings, and a dependable logistics process, positioning it well to maintain profitability and scalability.

Part 3

	ProductType	n_samples
	<chr>	<int>
1	CLO	649
2	KEY	746
3	LAP	425
4	MON	619
5	MOU	860
6	SOF	864

Table 8: Samples per Product Type

	ProductType	sample_id	n	xbar	s	phase
	<chr>	<dbl>	<int>	<dbl>	<dbl>	<chr>
1	CLO	1	24	21.0	4.32	Baseline
2	CLO	2	24	19.4	6.96	Baseline
3	CLO	3	24	19.1	5.71	Baseline
4	CLO	4	24	20.0	5.58	Baseline
5	CLO	5	24	19.0	6.41	Baseline
6	CLO	6	24	20.0	6.37	Baseline

Example of Sample Data

	ProductType	xbarbar	sbar	CLx	UCLx	LCLx	CLs	UCLs	LCLs	X_1sig_up	X_2sig_up	X_1sig_lo
	<chr>	<dbl>	<dbl>	<dbl>	<dbl>	<dbl>	<dbl>	<dbl>	<dbl>	<dbl>	<dbl>	<dbl>
1	CLO	19.1	5.91	19.1	19.7	18.6	5.91	8.53	3.28	19.3	19.5	18.9
2	KEY	19.2	5.86	19.2	19.7	18.7	5.86	8.46	3.25	19.4	19.5	19
3	LAP	19.5	5.89	19.5	20.1	19	5.89	8.51	3.27	19.7	19.9	19.3
4	MON	19.4	5.92	19.4	20	18.9	5.92	8.56	3.29	19.6	19.8	19.2
5	MOU	19.2	5.68	19.2	19.8	18.7	5.68	8.2	3.15	19.4	19.6	19.1
6	SOF	0.956	0.297	0.956	0.983	0.929	0.297	0.43	0.165	0.965	0.974	0.947

Example of Control Data

As required by the assignment, the data was grouped by product type and arranged in chronological order before sampling. The first 30 samples from each product category were used to establish the baseline performance. Based on these baseline samples, the control limits, averages, and standard deviations were calculated and later applied in the development of the control charts.

```
Step 5 complete - SPC rules evaluated by PRODUCT TYPE.
Rule A (3 spikes) → 1 type-violations recorded.
Rule B (±1 stable runs) → 6 types evaluated.
Rule C ( 4 above +2 ) → 122 mean-shift events.
CSV tables saved for inclusion in report.
```

```
ruleA
```

```
## # A tibble: 1 x 6
##   ProductType Rule  Description                              total_violations first3 last3
##   <chr>       <chr> <chr>                                               <int> <chr>  <chr>
## 1 MOU         A     s > UCLs (+3 ) → process variation spike                1 592    592
```

```
ruleB
```

```
## # A tibble: 6 x 7
##   ProductType run_id Rule  Description                              run_len start   end
##   <chr>        <int> <chr> <chr>                                      <int> <dbl> <dbl>
## 1 CLO             97 B     Longest run of s within ±1  (good control)     35   474   508
## 2 KEY            175 B     Longest run of s within ±1  (good control)     15   730   744
## 3 LAP             20 B     Longest run of s within ±1  (good control)     19   116   134
## 4 MON             47 B     Longest run of s within ±1  (good control)     34   238   271
## 5 MOU            198 B     Longest run of s within ±1  (good control)     16   672   687
## 6 SOF            171 B     Longest run of s within ±1  (good control)     21   659   679
```

```
ruleC
```

```
## # A tibble: 122 x 6
##    ProductType Rule  Description                              total_runs first3
##    <chr>       <chr> <chr>                                         <int> <chr>
##  1 CLO         C     4 X-bar above +2  → mean drift suspected         19 47-50; 53-57; 81-84
##  2 CLO         C     4 X-bar above +2  → mean drift suspected         19 47-50; 53-57; 81-84
##  3 CLO         C     4 X-bar above +2  → mean drift suspected         19 47-50; 53-57; 81-84
##  4 CLO         C     4 X-bar above +2  → mean drift suspected         19 47-50; 53-57; 81-84
##  5 CLO         C     4 X-bar above +2  → mean drift suspected         19 47-50; 53-57; 81-84
##  6 CLO         C     4 X-bar above +2  → mean drift suspected         19 47-50; 53-57; 81-84
##  7 CLO         C     4 X-bar above +2  → mean drift suspected         19 47-50; 53-57; 81-84
##  8 CLO         C     4 X-bar above +2  → mean drift suspected         19 47-50; 53-57; 81-84
##  9 CLO         C     4 X-bar above +2  → mean drift suspected         19 47-50; 53-57; 81-84
## 10 CLO         C     4 X-bar above +2  → mean drift suspected         19 47-50; 53-57; 81-84
## 11 CLO         C     4 X-bar above +2  → mean drift suspected         19 47-50; 53-57; 81-84
## 12 CLO         C     4 X-bar above +2  → mean drift suspected         19 47-50; 53-57; 81-84
```

A, B and C Rule Data

Rule A – Extreme Variation (+3 SD Control Limit):

The +3 standard deviation control limit is used to detect extreme variation or special-cause events within the process. In this dataset, only one instance exceeded this threshold, associated with the Mouse product type. While this point represents an extreme deviation, it is an isolated case and does not indicate a sustained or systemic loss of control. Therefore, the delivery process remains generally well-controlled with respect to extreme variation under Rule A.

Rule B – Process Consistency (Stable Runs within ±1 SD):

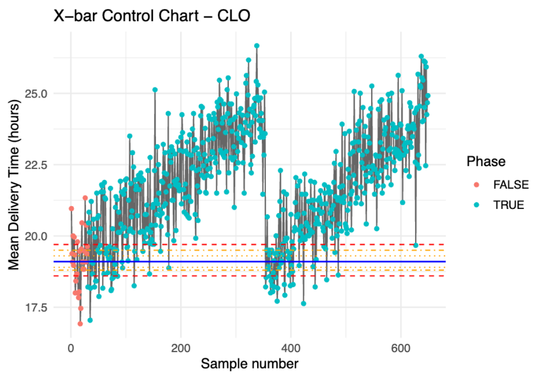
Rule B evaluates the length of consecutive samples that remain within ±1 standard deviation from the mean. This measure reflects process stability and consistency. Observed runs ranged from 15 to 35 consecutive samples, demonstrating that variability is minimal and the process operates within a stable range for extended periods. These long runs confirm good short-term control and consistent performance across most product types.

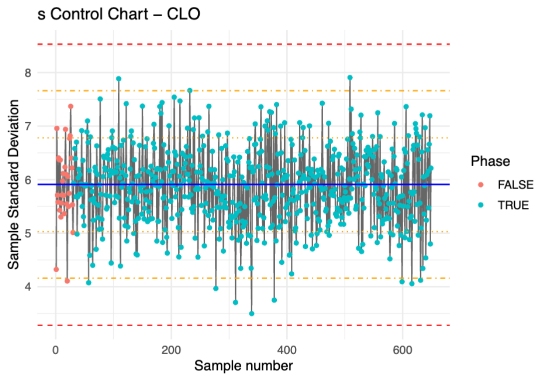
Rule C – Mean Shifts (Sustained Movement Beyond +2 SD):

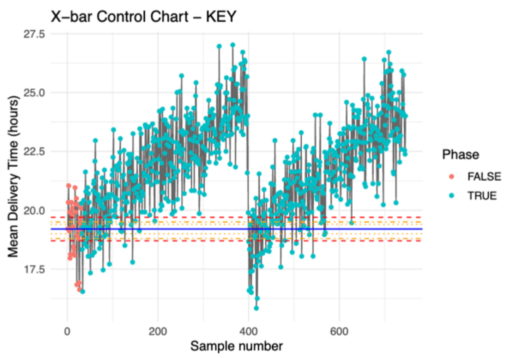
Rule C identifies sustained mean shifts, which signal gradual changes in the process average or capability. A total of 122 mean-shift events were detected across all product types, indicating that while the process is not subject to frequent extreme spikes, systematic shifts in the mean are occurring regularly. This suggests that the process is gradually drifting out of control, likely due to underlying factors such as process wear, operator adjustments, or external variation sources. Continuous monitoring and corrective actions are recommended to restore stability and maintain long-term process capability.

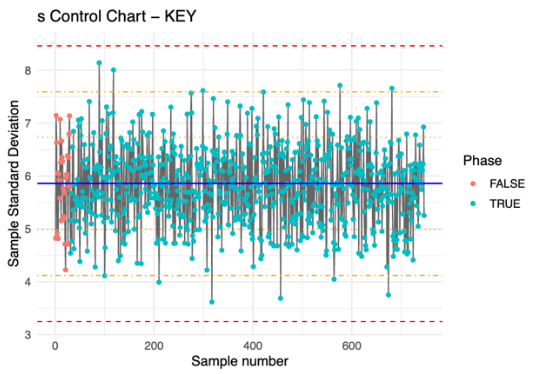
ProductType	n_records	mean_delivery	sd_delivery	sbar	sigma_hat	Cp	Cpu	Cpl	Cpk	Capability_Status
<chr>	<int>	<dbl>	<dbl>	<dbl>	<dbl>	<dbl>	<dbl>	<dbl>	<dbl>	<chr>
SOF	1000	0.955375	0.2940868	0.297	0.3002448	17.800	34.500	1.06	1.060	⚠ Marginal capability
MOU	1000	19.297500	5.8276023	5.680	5.7420563	0.929	0.737	1.12	0.737	❌ Not Capable
KEY	1000	19.276000	5.8151950	5.860	5.9240228	0.900	0.716	1.08	0.716	❌ Not Capable
CLO	1000	19.226000	5.9408054	5.910	5.9745691	0.893	0.713	1.07	0.713	❌ Not Capable
MON	1000	19.410000	5.9989192	5.920	5.9846784	0.891	0.701	1.08	0.701	❌ Not Capable
LAP	1000	19.606000	5.9339589	5.890	5.9543506	0.896	0.694	1.10	0.694	❌ Not Capable

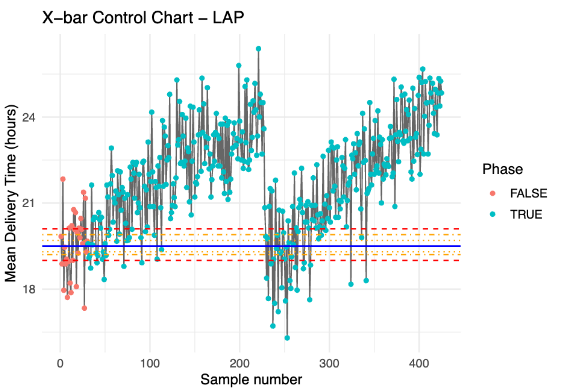
Capability Data

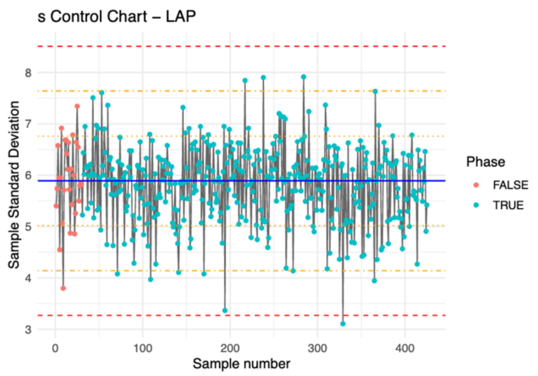


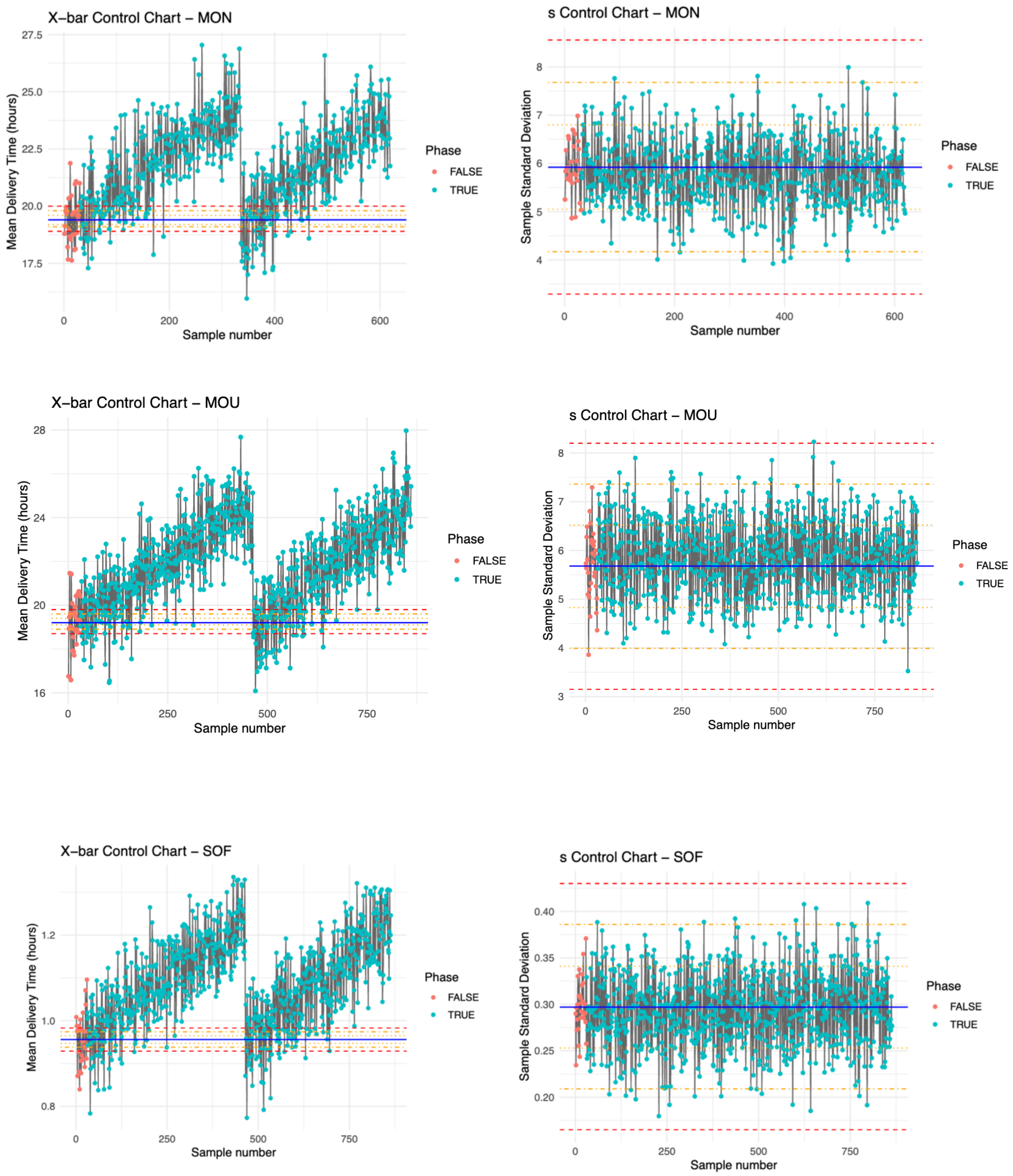












X- and S-plot Charts

early_mean	late_mean
<dbl>	<dbl>
0.956	1.09

Test Data ~ confirmation of mean shift

CONCLUSION:

As shown in the capability chart, all processes except the delivery of software are out of control. The reasoning behind this is clear when examining the X-bar charts. Software delivery shows extremely limited delivery times, likely because it is handled digitally, which makes it less susceptible to operational delays or process variation.

In contrast, the X-bar and S-bar charts for physical products display consistent variability (S-bar) but reveal a gradual loss of control in the X-bar values over time. Each year, the process drifts out of control but later stabilizes again, suggesting cyclical internal factors rather than random variation.

Extensive analysis of sales data showed no signs of seasonality—sales volumes remain evenly distributed throughout the year. This means that the observed instability is not caused by sales demand fluctuations, but rather by a systematic operational issue that develops progressively over time. Possible causes may include gradual process inefficiencies, workforce fatigue, equipment wear, or management and scheduling inconsistencies.

These findings align with the SPC rule evaluation: Rules A and B confirmed strong short-term control and process consistency, while Rule C detected numerous mean-shift events, reinforcing the presence of ongoing shifts in process capability.

In summary, the process itself is consistent but experiences progressive drift, leading to periods of instability. Since demand and sales volume are not the root causes, further investigation into internal operational practices is recommended to maintain long-term control and process reliability.

Part 4

4.1 TYPE 1 ERROR:

Rule A:

The probability that a value falls **outside** of 3 standard deviations given a normal distribution:

P(mean +- 3sd) = 0.00135

Considering the probability of consecutive errors:

P=0.00135^x

We know there was only one sample above, thus P=0.00135. Thus, for the in-control process 1.35 in every 1000 will be flagged (Lane, n.d.).

Rule B:

Type 1 error is the following: Out of control signal in and in control process. Since rule B is an indication of process stability (it does not give out of control signals) – it has 0 probability of a Type 1 error.

Rule C:

The probability that a value falls **outside** of +2 standard deviations given a normal distribution:

P(mean + 2sd) = 0.02275

Considering the probability of consecutive errors:

P=0.02275^x

For Rule C there has to be 4 consecutive ~ P= 2.7 x 10^-7

4.2 TYPE 2 ERROR:

Original Mean: 25.050

UCL: 25.089

LCL: 25.011

Mean: 25.028

Original SD: 0.013

SD: 0.017

Zu = UCL – Mean / SD ~ 3.588

Zl = LCL – Mean / SD ~ -1

B = Φ(Zu)− Φ(Zl) ~ 0.841

That means there is an 84% chance the sample mean will stay in the control limits (Type 2 error – missed detection). This is logical as the shift is minimal and still comfortably between control limits.

4.3 UPDATED DATA:

1. PRODUCTS

	ProductID	SellingPrice	Markup
	<chr>	<dbl>	<dbl>
1	SOF001	512.	25.0
2	SOF002	505.	10.4
3	SOF003	494.	16.2
4	SOF004	543.	17.2
5	SOF005	516.	11.0
6	SOF006	479.	17.0
7	SOF007	528.	16.8
8	SOF008	549.	12.0
9	SOF009	540.	11.3
10	SOF010	397.	23.5
11	SOF011	512.	25.0
12	SOF012	505.	10.4
13	SOF013	494.	16.2
14	SOF014	543.	17.2
15	SOF015	516.	11.0

Example of Fixed Data

	Category	Example_Products
	<chr>	<chr>
1	Cloud Subscription	CLO011, CLO012, CLO013
2	Keyboard	KEY041, KEY042, KEY043
3	Laptop	LAP021, LAP022, LAP023
4	Monitor	MON031, MON032, MON033
5	Mouse	MOU051, MOU052, MOU053
6	Software	SOF001, SOF002, SOF003

Example of Fixed Data

	ProductID	Category	Description	SellingPrice	Markup
	<chr>	<chr>	<chr>	<dbl>	<dbl>
1	SOF001	Software	coral matt	512.	25.0
2	SOF002	Software	cyan silk	505.	10.4
3	SOF003	Software	burlywood marble	494.	16.2
4	SOF004	Software	blue silk	543.	17.2
5	SOF005	Software	aliceblue wood	516.	11.0

Example of Fixed Data

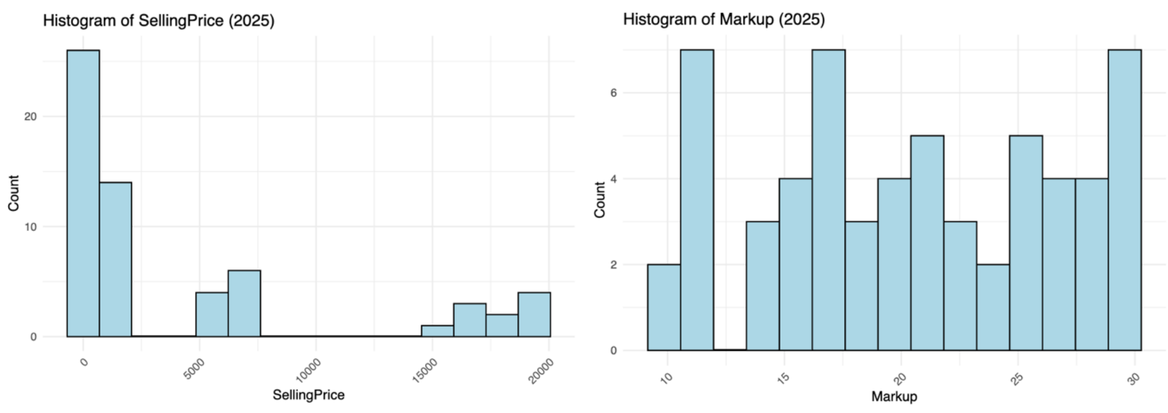
Variable	Min	Q1	Median	Q3	Max	Mean	SD
SellingPrice	350.45	512.18	794.18	6416.66	19725.18	4493.59	6503.77
Markup	10.13	16.14	20.34	25.71	29.84	20.46	6.07

Products Numerical Data: Fixed

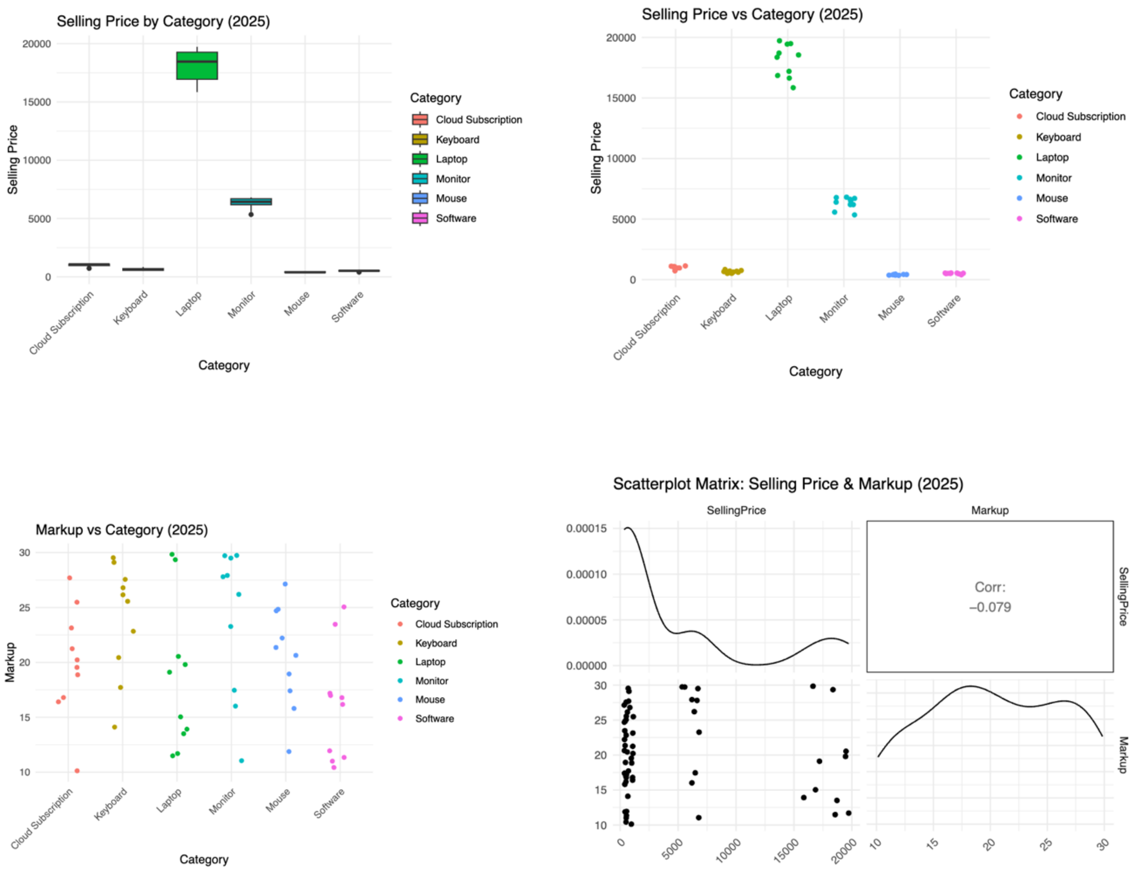
```
                 col Freq
1 Cloud Subscription   10
2           Keyboard   10
3             Laptop   10
4            Monitor   10
5              Mouse   10
6           Software   10

--- Description ---
                     col Freq
1       aliceblue marble    2
2         aliceblue silk    4
3         aliceblue wood    1
4           azure marble    1
5             azure matt    2
6         azure sandpaper   3
7             azure silk    4
8           black bright    1
9           black marble    1
10       black sandpaper    2
11            black silk    2
12           blue marble    1
13             blue matt    1
14             blue silk    3
```

Products Non-Numerical Data: Fixed



Univariate Data



Relational Data

CONCLUSION:

No data analysis was done on the updates products file. As I identified the data issues early on I used the prefix in product ID for categorical data. Thus, the results are 100% identical.

2. PRODUCTS HEADOFFICE

```
  ProductID Category Description          SellingPrice Markup
  <chr>     <chr>    <chr>                       <dbl>  <dbl>
1 SOF001    Software coral silk                   512.   25.0
2 SOF002    Software black silk                   505.   10.4
3 SOF003    Software burlywood marble             494.   16.2
4 SOF004    Software black marble                 543.   17.2
5 SOF005    Software chartreuse sandpaper         516.   11.0
```

Example of Fixed Data

Variable	Min	Q1	Median	Q3	Max	Mean	SD
SellingPrice	350.45	512.18	794.18	6416.66	19725.18	4493.59	6458.32
Markup	10.13	16.14	20.34	25.71	29.84	20.46	6.03

Products Headoffice Numerical Data: Fixed

The updated dataset shows only minor changes compared to the original. For selling price, the mean increased slightly from 4410.96 to 4493.59, while the median remained nearly the same, indicating that the overall center of the distribution has not shifted but a few higher-priced items may have influenced the mean. The standard deviation decreased marginally, suggesting slightly reduced overall variability, while the interquartile range widened, showing a greater spread among mid-range values. Both the minimum and maximum prices changed, narrowing the range and indicating fewer extreme outliers. The distribution remains right-skewed, with most products at lower prices and a few expensive ones raising the average. For markup, the mean and median remained stable, showing no change in pricing structure, while the standard deviation and interquartile range increased slightly, suggesting a small rise in variability. Overall, the differences indicate data refinement or outlier adjustment rather than any major change in product pricing or profitability patterns.

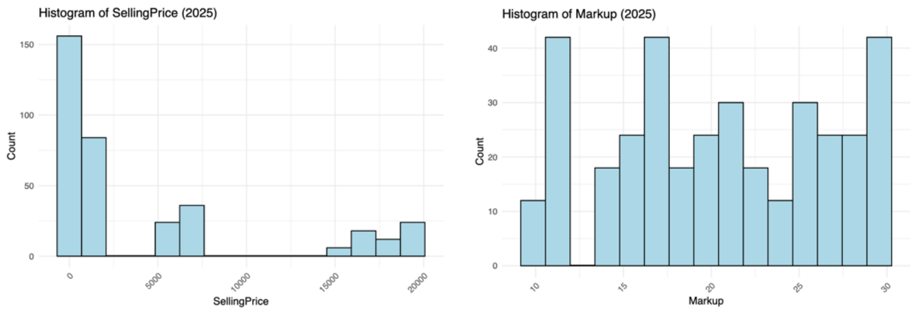
--- Category ---

	col	Freq
1	Cloud Subscription	60
2	Keyboard	60
3	Laptop	60
4	Monitor	60
5	Mouse	60
6	Software	60

--- Description ---

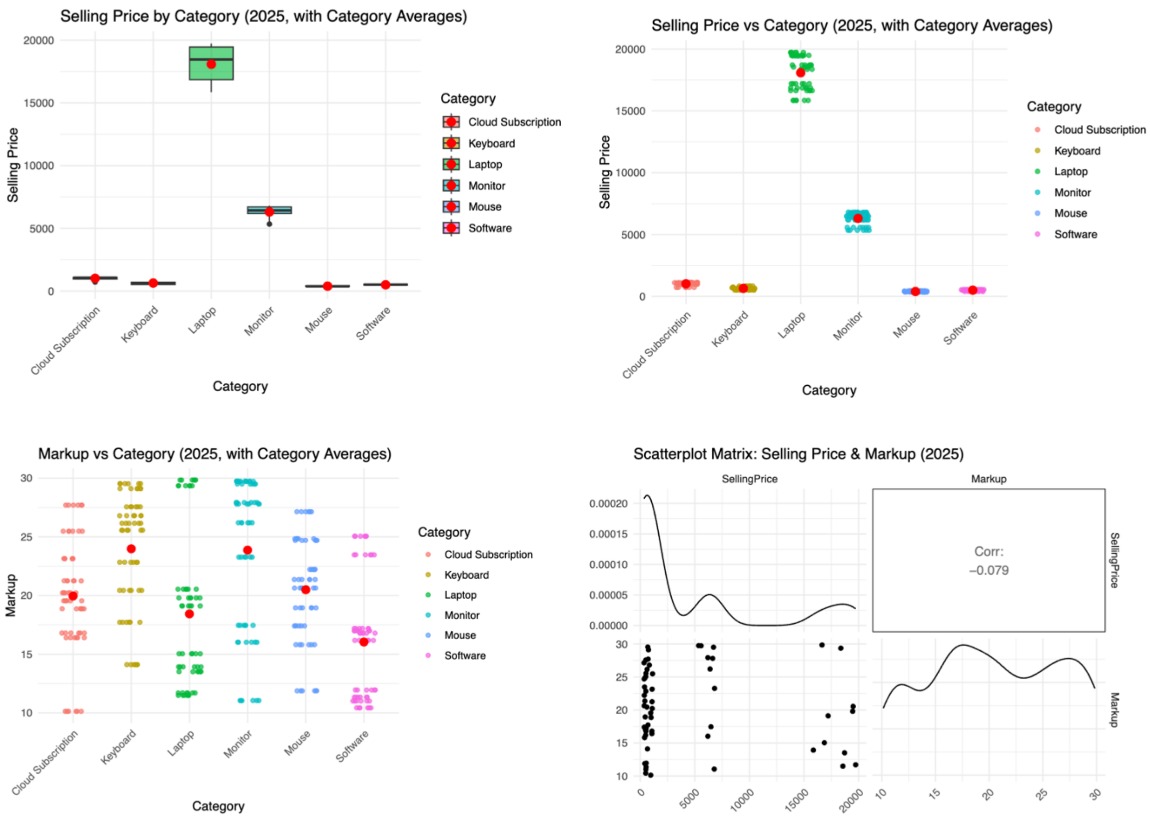
	col	Freq
1	aliceblue bright	4
2	aliceblue marble	3
3	aliceblue matt	5
4	aliceblue sandpaper	3
5	aliceblue silk	10
6	aliceblue wood	2
7	azure marble	8
8	azure matt	7
9	azure sandpaper	2
10	azure silk	15
11	azure wood	1
12	black bright	8
13	black marble	14
14	black matt	7
15	black sandpaper	6
16	black silk	21
17	black wood	2
18	blue bright	4
19	blue marble	4
20	blue matt	1
21	blue sandpaper	2
22	blue silk	12
23	blue wood	1
24	blueviolet bright	6
25	blueviolet marble	6
26	blueviolet matt	9
27	blueviolet sandpaper	3
28	blueviolet silk	10
29	blueviolet wood	4
30	burlywood bright	3

Products Headoffice Non-Numerical Data: Fixed



Univariate Data

Data is near identical, with no changes relevant to company.

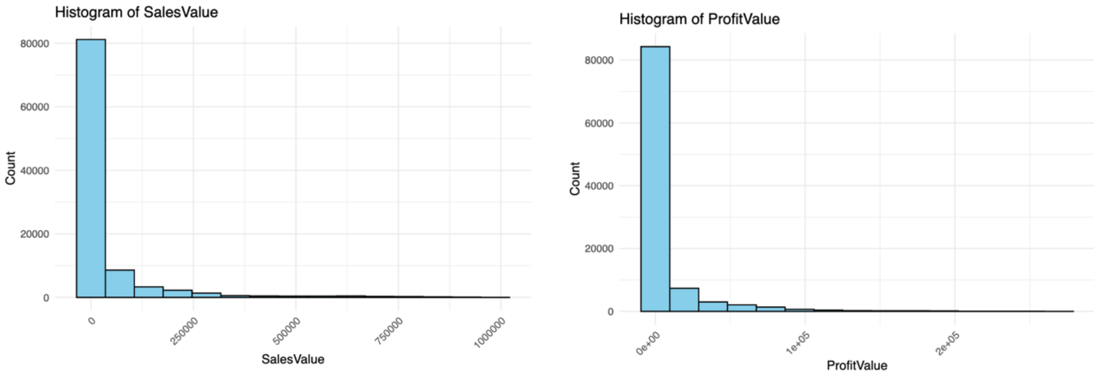


Relational Data

CONCLUSION:

Minimal changes were observed in the overall markup and selling price summaries; however, the relational data reveals significant differences across product types. Previously, selling prices appeared nearly identical across categories, but the updated data provides a more realistic differentiation. Cloud Subscriptions, Keyboards, Mouses, and Software now average below 1 000, while Laptops range between 17 000 and 19 000, and Monitors average around 6 000. This adjustment aligns better with expected product value, as laptops are inherently high-value items that drive total revenue despite their smaller sales volume. Markup values changed moderately—Laptops and Software generally show lower markups, whereas Mouses, Keyboards, Monitors, and Cloud Subscriptions carry slightly higher ones. The combination of high selling price and strong markup for Monitors suggests they contribute substantially to profit margins, while Laptops, though valuable, operate on tighter margins due to their competitive pricing and higher cost base.

SALES: Missing Financial Analysis



Univariate Data

Relational Data

	ProductType	Total_Sales	Total_Profit
	<chr>	<dbl>	<dbl>
1	LAP	2470814376.	455643479.
2	MON	1258942847.	297175938.
3	CLO	214110418.	42548481.
4	KEY	155002210.	37213011.
5	SOF	142527355.	22679478.
6	MOU	111190471.	22577052.

Sales per Type

	Grand_Sales	Grand_Profit
	<dbl>	<dbl>
1	4352587678.	877837440.

Sales and Profit Totals

	CustomerID	Total_Sales
	<chr>	<dbl>
1	CUST1791	56873415.
2	CUST2527	48935615.
3	CUST2277	46827905.
4	CUST3944	46032081.
5	CUST596	45713549.
6	CUST3721	44034697.
7	CUST1193	41344056.
8	CUST4729	39675224.
9	CUST1427	34785958.
10	CUST1501	31923505.

Top 10 Customers

CONCLUSION:

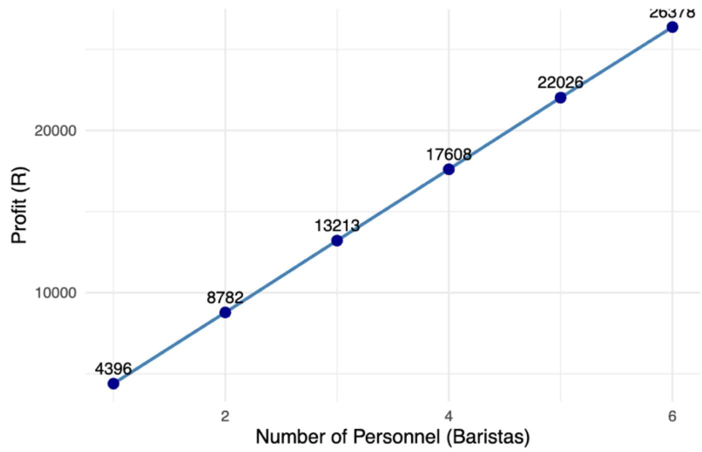
The financial data shows that the company generated total sales of approximately 4.3 billion with a total profit of around 878 million, reflecting a strong overall performance. Among product types, laptops lead both in sales (247 million) and profit (45.6 million), confirming their role as the company's most valuable segment despite higher costs and lower markups. Monitors follow with sales of 125.9 million and profit of 29.7 million, benefiting from both high unit value and strong margins. Cloud Bases Subscriptions and keyboards also contribute notably, with profits exceeding 42 million and 37 million respectively, indicating steady demand and efficient pricing. In contrast, software and mouses have lower total sales and profits, though they add consistent revenue due to high transaction frequency.

Customer analysis reveals a concentrated sales base, with the top 10 customers accounting for a significant share of total sales—the largest customer alone contributing over 56 million. This highlights the importance of maintaining strong relationships with key clients while diversifying the customer base to reduce dependency risk. The company should place massive emphasis on their top 10 (or more) customers. Promotions, personalised service and other measures should be taken to retain clients.

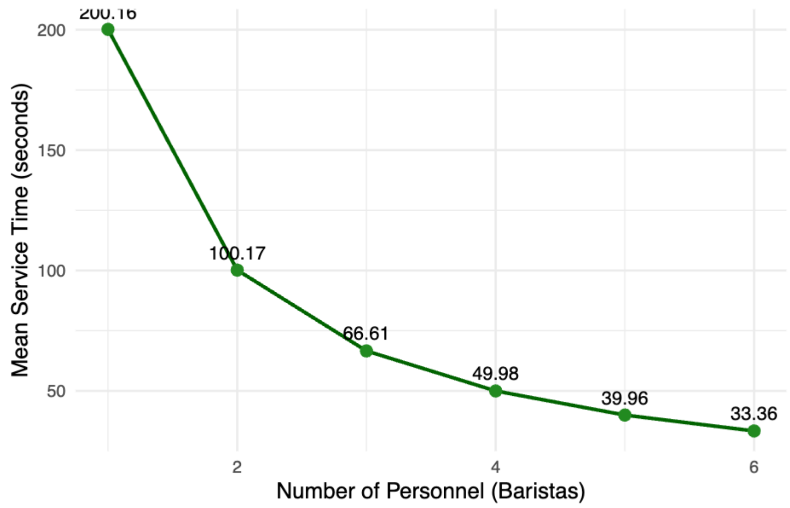
Part 5

	Personnel	Mean_Service_sec	Samples	Customers_Served	Sales_R	Cost_R	Profit_R
1	1	200.16	417	179.9	5396	1000	4396
2	2	100.17	3556	359.4	10782	2000	8782
3	3	66.61	12126	540.4	16213	3000	13213
4	4	49.98	29305	720.3	21608	4000	17608
5	5	39.96	56701	900.9	27026	5000	22026
6	6	33.36	97895	1079.3	32378	6000	26378

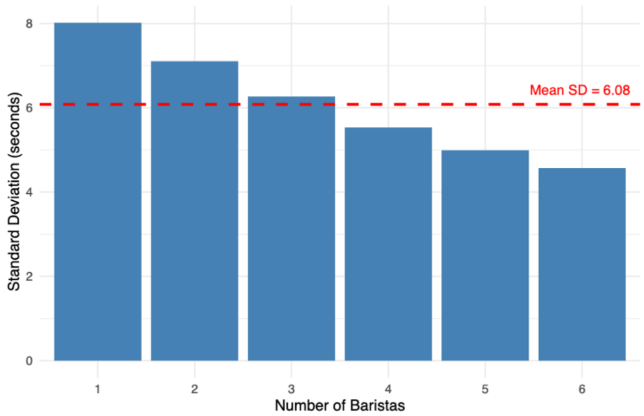
Example of Sample Data



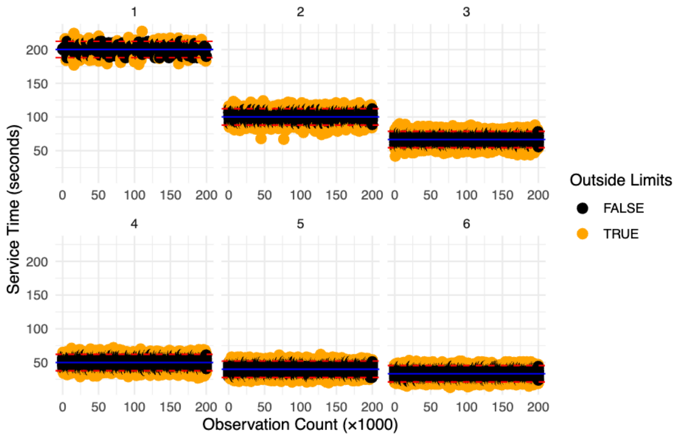
Shop 1: Profit vs Baristas Graph



Shop 1: Mean Service Time vs Baristas Graph



Shop 1: Standard Deviation vs Baristas Graph



Shop 1: "Control Charts"

	baristas	outside_points	total_points
	<dbl>	<int>	<int>
1	1	51	417
2	2	312	3556
3	3	678	12126
4	4	653	29305
5	5	683	56701
6	6	834	97895

Shop 1: Control Data

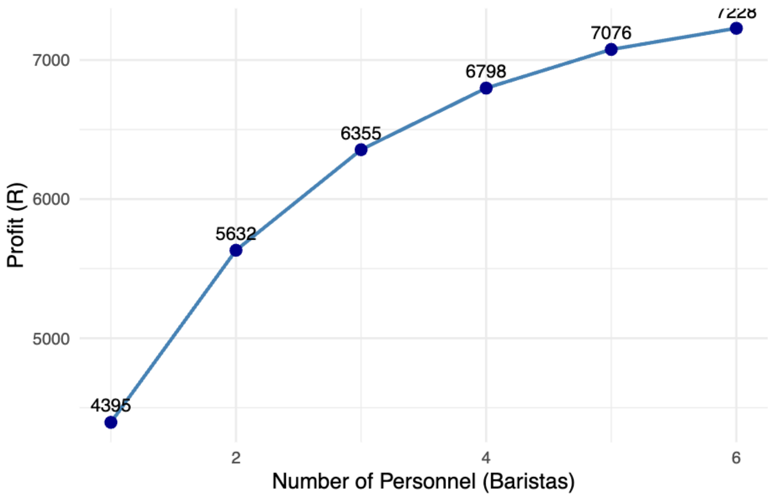
	baristas	reliability_percent
	<dbl>	<dbl>
1	1	87.8
2	2	91.2
3	3	94.4
4	4	97.8
5	5	98.8
6	6	99.1

Shop 1: Reliability Data

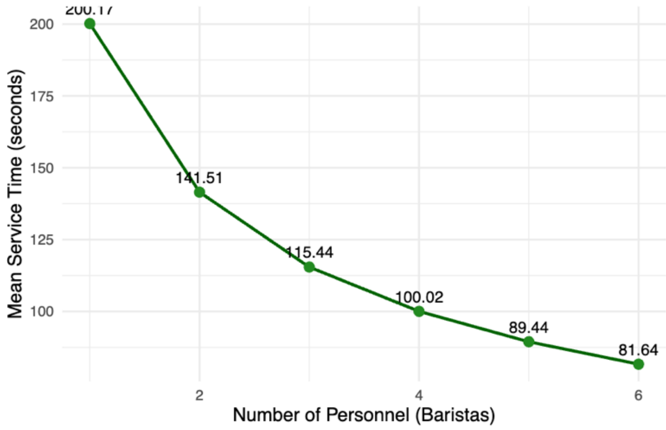
SHOP 2

	Personnel	Mean_Service_sec	Samples	Customers_Served	Sales_R	Cost_R	Profit_R
1	1	200.17	2196	179.8	5395	1000	4395
2	2	141.51	8859	254.4	7632	2000	5632
3	3	115.44	19768	311.8	9355	3000	6355
4	4	100.02	35289	359.9	10798	4000	6798
5	5	89.44	54958	402.5	12076	5000	7076
6	6	81.64	78930	440.9	13228	6000	7228

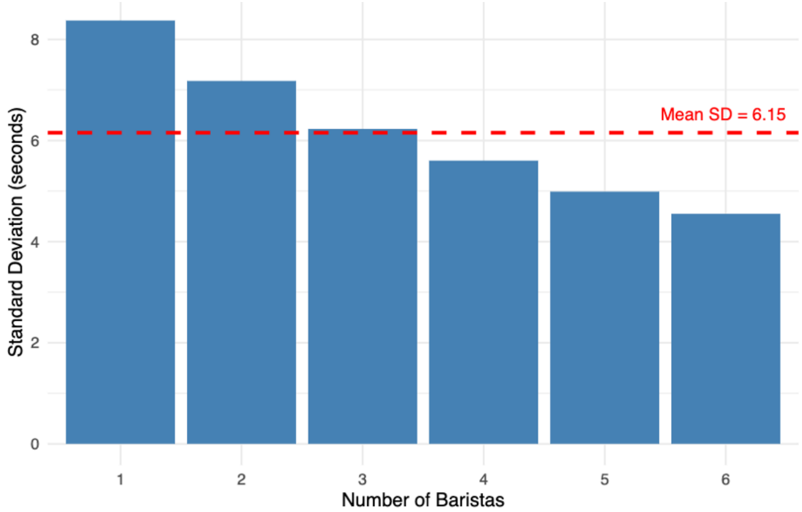
Example of Sample Data



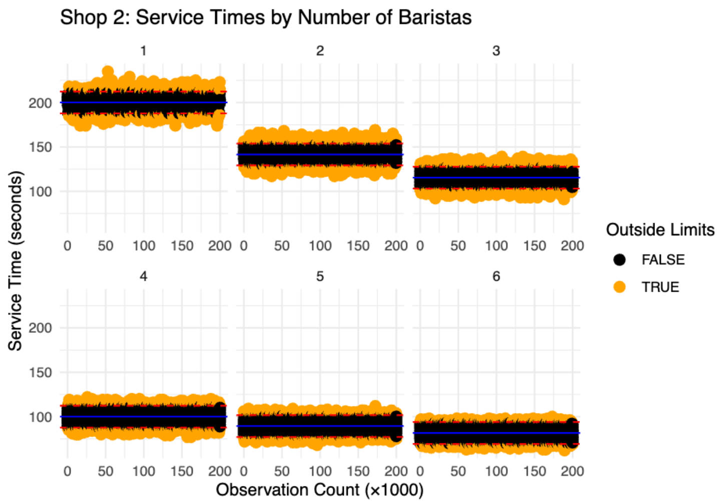
Shop 2: Profit vs Baristas Graph



Shop 2: Mean Service Time vs Baristas Graph



Shop 2: Standard Deviation vs Baristas Graph



Shop 2: "Control Charts"

	baristas	outside_points	total_points
	<dbl>	<int>	<int>
1	1	304	2196
2	2	872	8859
3	3	1071	19768
4	4	885	35289
5	5	870	54958
6	6	652	78930

Shop 2: Control Data

	baristas	reliability_percent
	<dbl>	<dbl>
1	1	86.2
2	2	90.2
3	3	94.6
4	4	97.5
5	5	98.4
6	6	99.2

Shop 2: Reliability Data

CONCLUSION: SHOP 1 & 2

How the Data Was Compiled:

Profit was calculated in two parts. First, the mean service time for a given number of baristas was used to estimate the total number of customers served. This figure was multiplied by the average turnover per customer to obtain total sales. The total cost was then calculated as the sum of labor costs for all baristas, and profit was derived as sales minus total cost.

Service reliability was assessed using variation in service times. The standard deviation (SD) of service time was calculated for each barista configuration, and control limits were established at ±2 SD. The percentage of service times falling within these limits represented the reliability score, reflecting how consistently customers were served within expected timeframes.

Shop 1:

For Shop 1, increasing the number of baristas had a clear linear effect across key performance metrics. Profit rose steadily from 4 396 with one barista to 26 378 with six, while service time variability (SD) declined from 8.0 to 4.8. This indicates that adding baristas not only improved throughput and sales but also enhanced process stability and customer service consistency.

However, reliability gains began to plateau beyond five baristas, improving only marginally from 98.9% to 99.1% between five and six baristas. This suggests diminishing returns in operational reliability beyond a certain staffing level. From a business standpoint, the optimal number of baristas may not necessarily be the maximum (six), as the marginal improvement in reliability does not justify the added labor cost if profit growth continues without stabilization.

A key takeaway for management is the opportunity to optimize staffing levels for efficiency rather than simply maximizing headcount. With reliability already near 99%, the focus should shift to balancing labor costs with service speed and customer volume. Implementing dynamic staffing models—where baristas are scheduled based on expected demand peaks—could further enhance profitability without compromising reliability. Additionally, since variability in service times declines predictably with more staff, cross-training employees or improving workflow efficiency could replicate similar stability gains at lower staffing costs. However only implementing 6 baristas will not maximise profit, as linear increase will continue.

Shop 2:

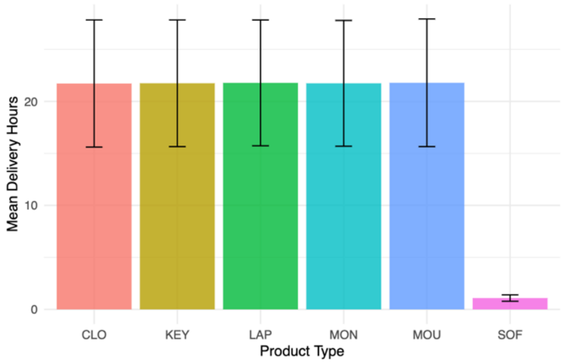
For Shop 2, all key performance indicators show strong improvement with each additional barista, reaching a point of stabilization at six baristas. Profit increases consistently from 4 395 with one barista to 7 228 with six, while the standard deviation of service time decreases steadily from 8.2 to 4.6, reflecting greater consistency and reduced variability in operations. Average service time also improves significantly, dropping from 200 to 81 seconds, which indicates a much faster and more efficient service process. Reliability follows the same trend, rising from 86.2% to 99.2%, meaning nearly all customer interactions fall within expected and acceptable time limits.

These results suggest that six baristas represent the optimal staffing level for this shop. At this point, all performance measures—profit, reliability, and service consistency—have reached their practical maximums, and further increases in staffing would likely result in diminishing returns. The process appears to have achieved operational equilibrium, where both productivity and quality are maximized without introducing excess cost.

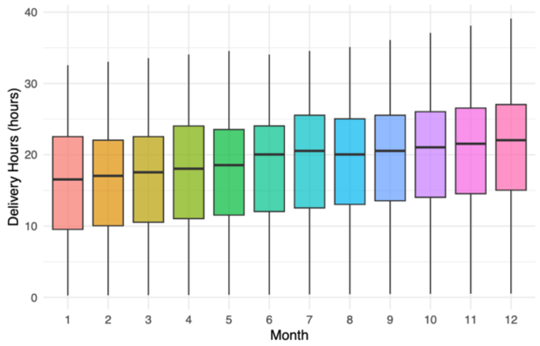
From a managerial standpoint, this indicates that Shop 2 operates most efficiently at six baristas, balancing workload, customer satisfaction, and profitability. The shop should maintain this staffing level during normal operating conditions but consider flexible scheduling or rotational support during peak hours to preserve reliability without overspending on labor. Additionally, with reliability already exceeding 99%, any further improvement would likely depend on process optimization, such as workflow layout or order-handling efficiency, rather than additional staffing.

In conclusion, the process shows a clear trend of improvement as variability decreases and performance stabilizes at the optimal staffing level. The small losses seen when deviating from this point are similar to the **Taguchi Loss** (Gurus, n.d.), where any deviation from the target (optimal number of baristas) results in incremental losses in quality, efficiency, or profitability. In this case, operating below six baristas increases service time variability and reduces customer satisfaction, while exceeding it yields minimal additional benefit at a higher cost—illustrating that consistent operation near the target condition minimizes total loss to the business.

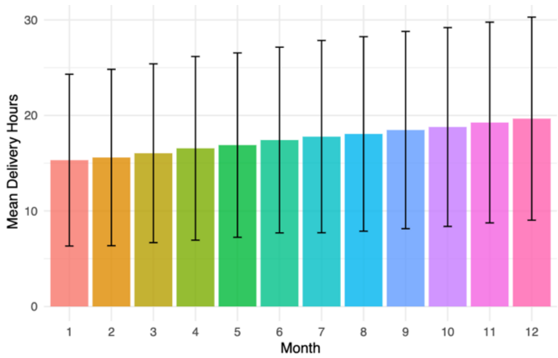
Part 6



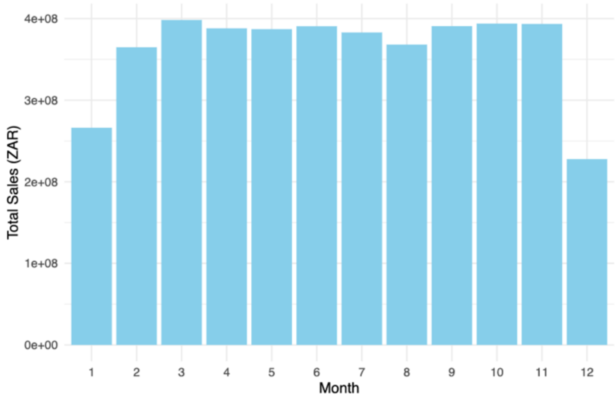
Anova 1: Delivery Hour vs Product Type (incl SD)



Anova 2: Delivery Hour vs Month (Boxplot)



Anova 2: Delivery Hour vs Month (Graph)



Anova 3: Sales vs Month

CONCLUSION:

ANOVA 1:

In earlier analysis, a group of delivery times showed almost zero duration, creating skewed results that initially lacked a clear explanation. The hypothesis tested whether product type had a significant influence on delivery time. The ANOVA confirmed this relationship, showing that software products had negligible delivery times compared to other product types. This finding makes logical sense, as software is delivered digitally, unlike physical products that require transport and handling.

ANOVA 2:

Control chart analysis revealed that the delivery process began the year in control but gradually moved out of control over time. The hypothesis proposed that delivery times increase as the year progresses. The ANOVA supported this, showing that mean delivery times rose from approximately 15.2 hours early in the year to over 19 hours later. This confirms a statistically significant upward trend, suggesting that process performance deteriorates gradually, possibly due to operational strain, resource limitations, or seasonal inefficiencies.

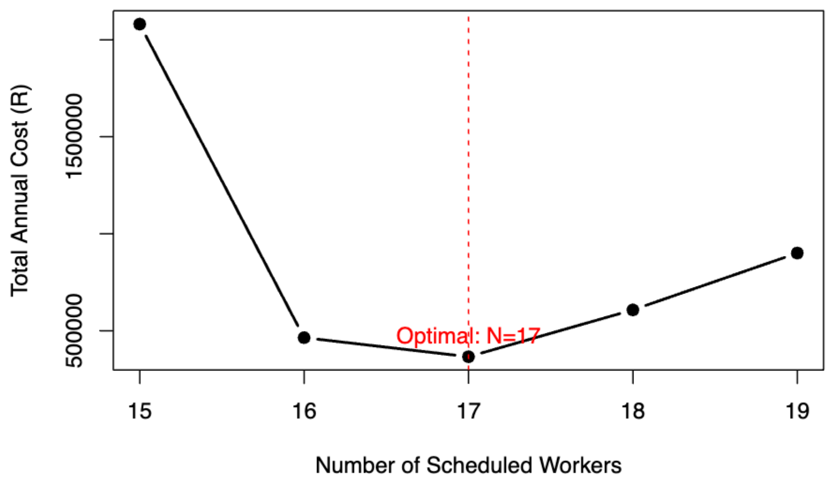
ANOVA 3:

Following the findings of ANOVA 2, a third hypothesis was developed to determine whether the increasing delivery times were caused by higher sales volumes rather than internal inefficiencies. The hypothesis that sales increased per month was tested and rejected, as no significant upward trend in sales volume was observed. This indicates that the rising delivery times are not demand-driven but instead point to a structural or systemic issue within the delivery process—such as capacity constraints, workflow inefficiency, or insufficient resource management—that should be addressed to restore process control and efficiency.

Part 7

N	Prob_Problem	Problem_Days	Annual_Loss	Hire_Cost	Total_Cost
15	0.326179	119.06	2381108	-3e+05	2081108
16	0.063631	23.23	464506	0e+00	464506
17	0.009071	3.31	66220	3e+05	366220
18	0.001040	0.38	7593	6e+05	607593
19	0.000101	0.04	740	9e+05	900740

Table 9: Annual Cost vs Number of Workers



Annual Cost vs Number of Workers (Graph)

CONCLUSION:

To evaluate and optimise the reliability of service at the car rental agency, the number of workers on duty was analysed over 397 recorded days. Using this distribution, the mean number of workers per day was calculated and converted into an estimated attendance probability of approximately 0.974 (97.4 percent). A binomial probability model was then applied to estimate the likelihood of having fewer than 15 workers on duty—defined as a "problem day" where service reliability decreases and average daily sales drop by R20 000.

A cost–benefit optimisation model was developed to determine the most profitable staffing level. The model considered both the expected annual loss from unreliable service and the annual hiring cost (R25 000 per month, or R300 000 per year, per additional employee). Calculations were performed for staffing levels between 15 and 19 workers, with the total expected annual cost computed as the sum of reliability-related losses and hiring expenses for each staffing level.

The results show that with 16 scheduled workers, the agency would experience around 23 unreliable days per year, resulting in an estimated annual loss of about R464 506. Increasing the workforce to 17 workers reduces the probability of unreliable service to only 0.9 percent, or roughly 3 unreliable days per year, lowering the total annual cost to approximately R366 220—the minimum across all scenarios. Beyond 17 workers, the additional salary expenses outweigh the savings gained from improved reliability, leading to higher overall costs.

In conclusion, scheduling 17 workers per day provides the optimal balance between cost and reliability, achieving an estimated 99.1 percent reliability (about 362 reliable service days per year). This staffing level minimises total annual costs while maintaining consistently high service quality, ensuring efficient operations and maximised profitability for the company.

Bibliography

Department of Economic and Social Affairs, 2017. *Distribution of the world's population by age and sex, 2017,* New York: United Nations.

Lane, D. M., n.d. *Normal Distribution.* [Online]
Available at: https://onlinestatbook.com/2/calculators/normal_dist.html
[Accessed 20 October 2025].

Gurus, Q., n.d. *Quality Gurus.* [Online]
Available at: https://www.qualitygurus.com/taguchi-loss-function/
[Accessed 20 October 2025].