ECSA GA4 report

ENGINEERING COUNCIL OF SOUTH AFRICA

PS van Zyl 27070662

Stellenbosch University

Quality Assurance 344

1 INTRODUCTION

The business climate is becoming more competitive and sustainability-driven by the day. Quality Assurance is a vital part of business success. However, what does such a broad term entail? The term certainly involves soft skills like motivating employees and listening to the voice of the customer. Quality Assurance also involves hard skills like data analysis and statistics.

This report aims to apply the hard skills of Quality Assurance in business contexts. First, data analysis is done to obtain an understanding of the business. Business insights are revealed and erroneous data corrected. Next, process capabilities are calculated, and statistical process control is done. Then the risk of errors in statistical process control is calculated and discussed. To confirm suspicions during the data analysis phase, analysis of variance is done. Finally, optimization of profit is done, as well as an estimate of the reliability of service.

CONTENTS

LIST OF FIGURES

LIST OF TABLES

LIST OF ABBREVIATIONS

CL: Control limit

CLO: Cloud subscription

Corr: Correlation coefficient

CUST: Customer

ECSA: Engineering Council of South Africa

GA: Graduate Attribute

Ho: Null hypothesis

ID: Identifier

KEY: Keyboard

LAP: Laptop

LCL: lower control limit

LSL: lower service level

MON: Monitor

MOU: Mouse

Sd: standard deviation

Sigma: standard deviation

SOF: Software

SPC: Statistical Process Control

SPLOM: scatter plot matrix

UCL: upper control limit

USL: upper service level

XBar: mean

2 DESCRIPTIVE ANALYSIS

During this phase, the prescribed Descriptive analysis process was followed.

2.1 DATA LOADING AND INSPECTION.

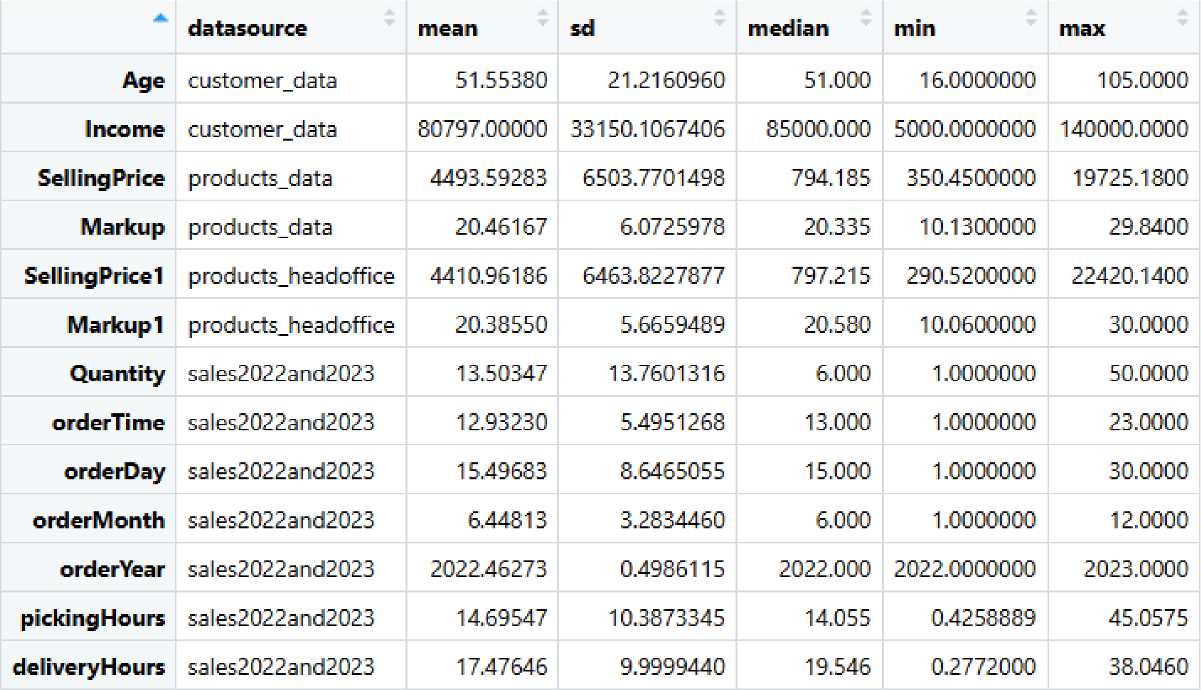
Four data sets have been supplied to do analysis, each containing:

- Customer data
- Product data
- Product data according to the head office
- Sales data for 2022 and 2023

2.2 SUMMARY STATISTICS

The functions skim and describe were used to get an initial look at the data.

Table 1 The summary statistics for the continuous features of the different data sources



	datasource	mean	sd	median	min	max
Age	customer_data	51.55380	21.2160960	51.000	16.0000000	105.0000
Income	customer_data	80797.00000	33150.1067406	85000.000	5000.0000000	140000.0000
SellingPrice	products_data	4493.59283	6503.7701498	794.185	350.4500000	19725.1800
Markup	products_data	20.46167	6.0725978	20.335	10.1300000	29.8400
SellingPrice1	products_headoffice	4410.96186	6463.8227877	797.215	290.5200000	22420.1400
Markup1	products_headoffice	20.38550	5.6659489	20.580	10.0600000	30.0000
Quantity	sales2022and2023	13.50347	13.7601316	6.000	1.0000000	50.0000
orderTime	sales2022and2023	12.93230	5.4951268	13.000	1.0000000	23.0000
orderDay	sales2022and2023	15.49683	8.6465055	15.000	1.0000000	30.0000
orderMonth	sales2022and2023	6.44813	3.2834460	6.000	1.0000000	12.0000
orderYear	sales2022and2023	2022.46273	0.4986115	2022.000	2022.0000000	2023.0000
pickingHours	sales2022and2023	14.69547	10.3873345	14.055	0.4258889	45.0575
deliveryHours	sales2022and2023	17.47646	9.9999440	19.546	0.2772000	38.0460

This table can be used to determine the mean values and variation for the various features, specifically the numerical features.

2.2.1 Insights into customer data

The customers are on average 51 years of age, and there is a customer of 105 years. Age has a mean and median close together, which can indicate a symmetrical distribution. Further analysis on the age to income distribution will follow.

2.2.2 Insights into product data

The Data for the product head office and products do not agree with one another. These discrepancies make it hard to know which datasets to merge. The discrepancies were queried. Until clarity is obtained on which datasets are correct or not, data analysis on product data is avoided, if possible, since it might lead to false discoveries.

Insights into customer data will be discussed in section 2.4.

2.3 HANDLING MISSING VALUES

There are no missing values in any of the CSV files.

2.4 EXPLORING RELATIONSHIPS

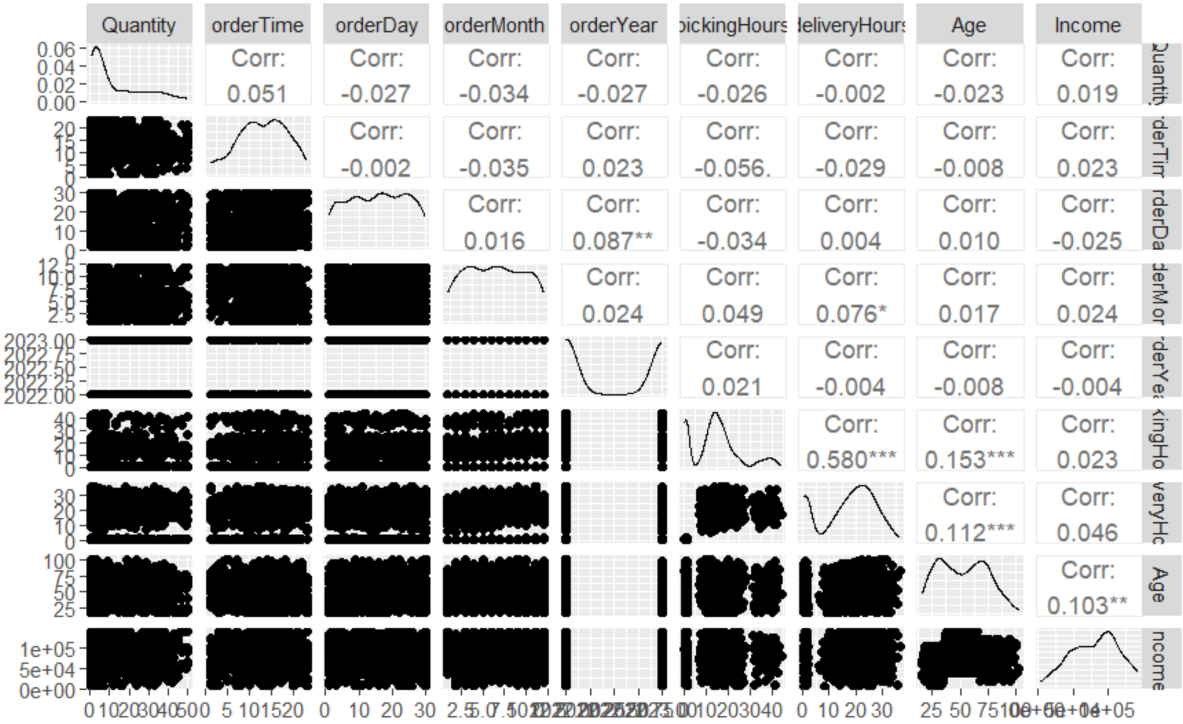
The ggpairs function was used on a merged dataset consisting of Sales and Customers.

A scatter plot matrix is a quick way to discover relationships between features of a dataset. The correlation coefficient in the upper right triangle indicates whether the compared features have a linear relationship or not. Since none of the correlation coefficients are close to 1 or -1, no linear relationships exist between the raw features.

The diagonal of the SPLOM contains a line graph of the distribution of the feature itself.

The lower left triangle contains a scatterplot of the features in relation to one another.

Table 2 Scatter plot matrix of sales and customer data



The following observations were made from the SPLOM:

- Order Year clearly only has values of 2022 and 2023.
- Picking Hours seem to have clear clusters when compared to age, delivery hours and Income.
- There are quite a few 0 values for delivery hours and picking hours.
- No strong linear correlations according to the correlation coefficients.
- Quantity is exponential, order time is bimodal, order day and order month are uniform. Order year is binary. Picking hours and delivery hours, and age are bimodal. Income is skewed to the left.

2.5 VISUAL REPRESENTATION

2.5.1 Feature distribution

In data analytics, the process of visual representation starts by displaying the distribution of each feature in each dataset visually. The continuous features are represented by blue histograms, and the categorical data is represented by red bar plots.

However, in section 2.4, the SPLOM already provided a basic understanding of the distribution of the features.

To avoid the overuse of graphs, these histograms and bar plots will be attached in an appendix for further perusal if the reader wishes to discover more about the data on their own accord.

Only noteworthy graphs will be displayed in this section.

2.5.1.1 Customer data

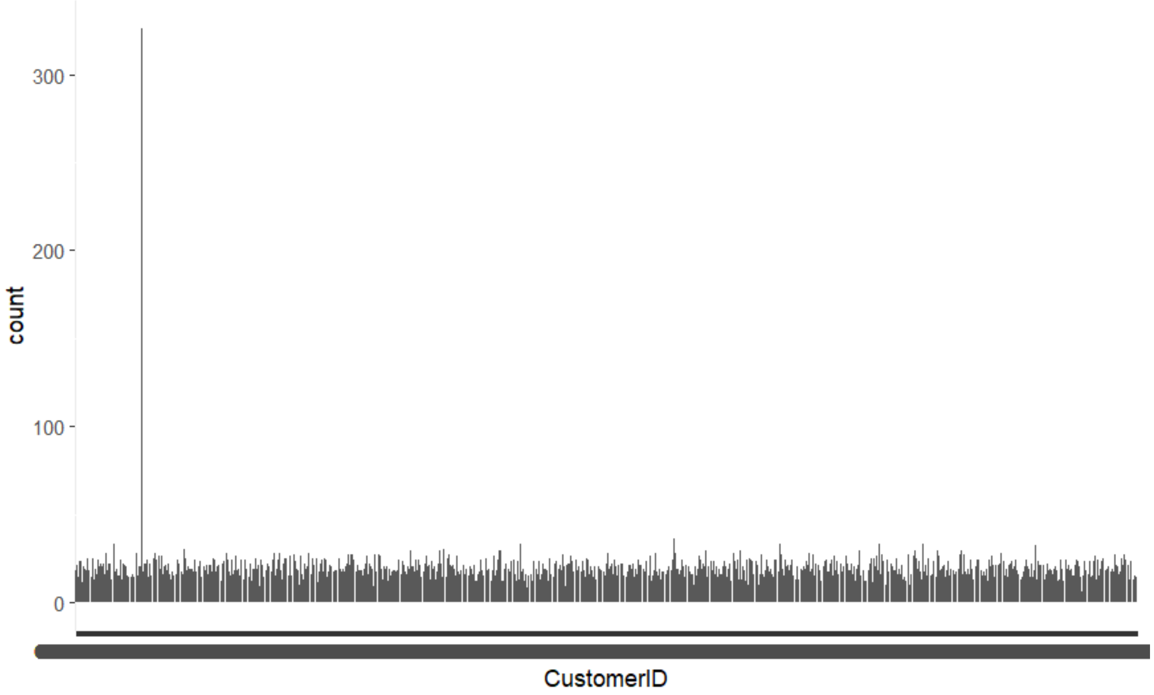


Figure 1 The distribution of sales accross the Customers

One customer buys a lot of products, CUST1193. It might be a retail store. The customer is a female of the age of 20 in San Francisco. It might be someone with an electronics business, or perhaps a purchase for employees in a business.

2.5.1.2 Product data

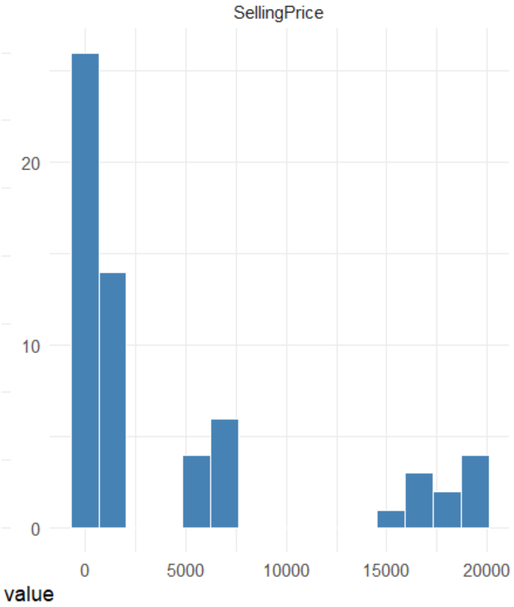


Figure 2 A histogram of the Selling Price of products according to the Product Data file

Notice the obvious clusters in selling price. These prices might coincide with the product category.

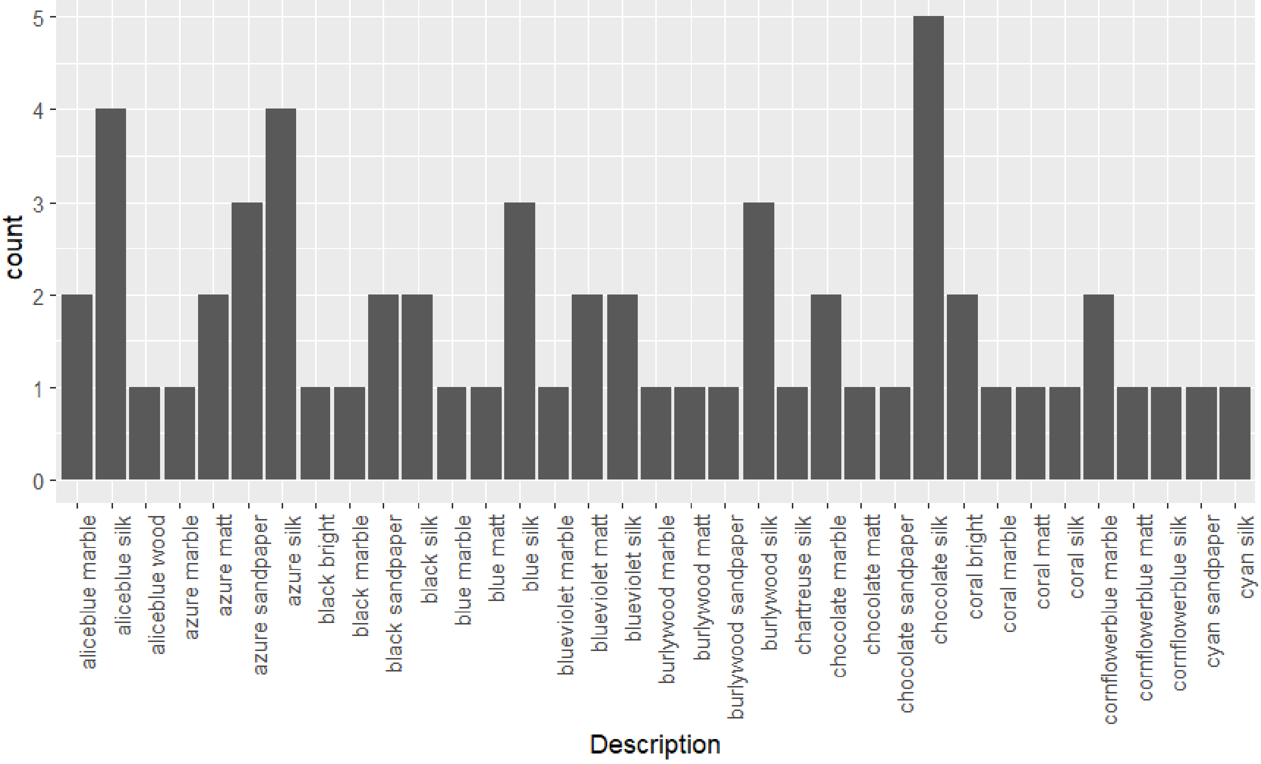


Figure 3 Bar plot of product descriptions in the product data set

There are some duplicate descriptions of products. It is due to different categories of products having the same description of colour. The product ID is a unique identifier for the products.

2.5.1.3 Product head office data

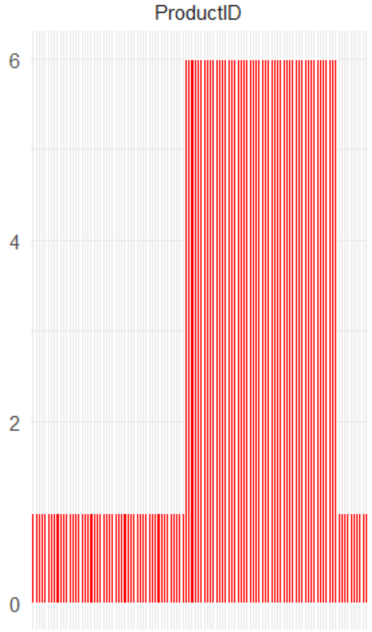


Figure 4 Bar plot of Product ID in the product dataset from head office

Some of the head office product IDs have a frequency of 6. These are the product IDs that start with “NA”. More information is required to determine what these product IDs represent. It might be products that are off the market or unidentified products. There is also a discrepancy between product IDs and descriptions between product data and product head office data. Since there is no indication of which data source is true, the product data is used minimally in analysis until more information is available.

2.5.2 Feature relationship insights

In this section, more than one feature is combined to try to find trends and interactions between features. Data anomalies are explored and potential causes are discussed.

2.5.2.1 Sales per month and year comparison

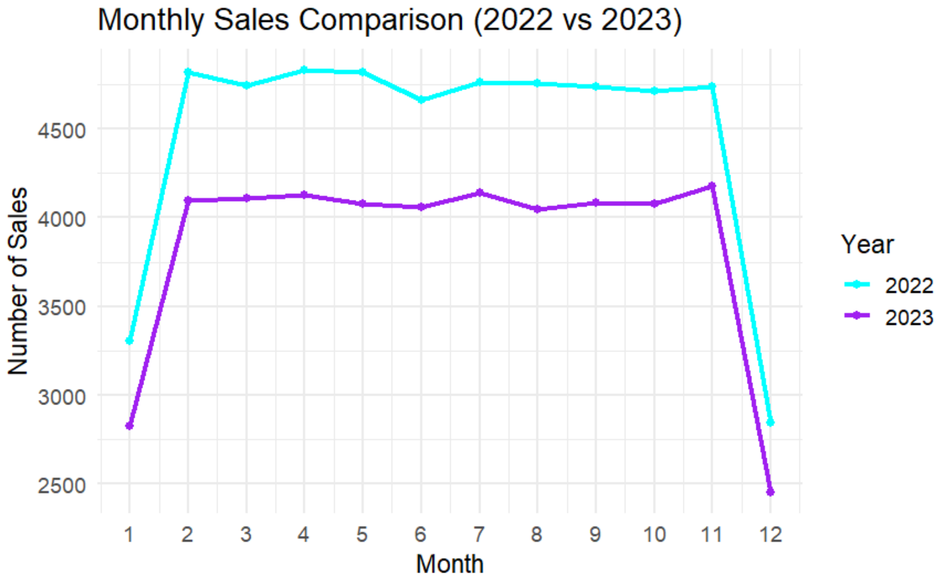


Figure 5 Monthly sales over the two years

The monthly revenue can be obtained by multiplying the quantity sold by the selling price of the product for each instance in the dataset and then adding the revenue per sale for each month.

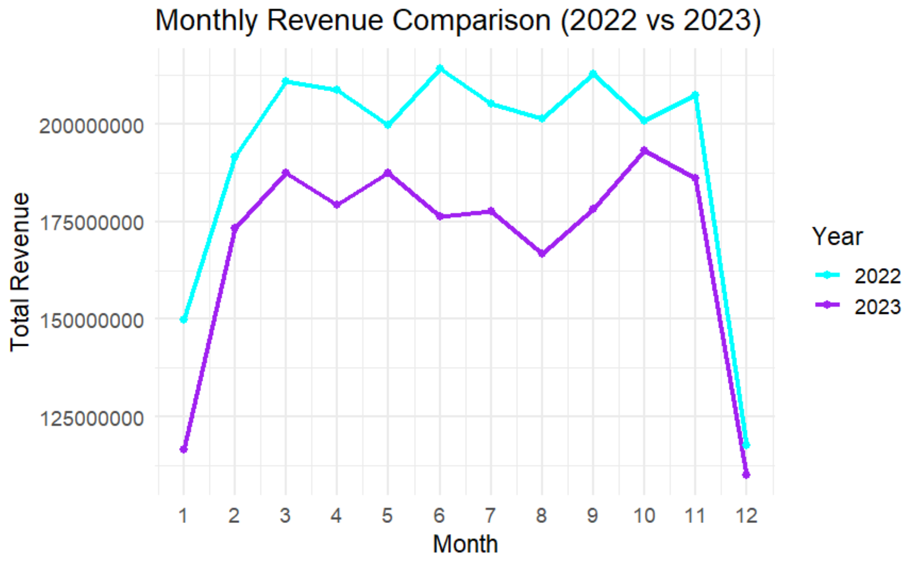


Figure 6 Monthly revenue for 2022 and 2023

From 2022 to 2023, the number of sales decreased, likely due to competition. The downward trend should be closely monitored to ensure the business stays profitable. From the two graphs, it is also obvious that sales during months 1 and 12 are low, likely due to the business closing over School holidays, for Christmas, or New Year's Eve. Staff are likely taking leave during December and January, on the days when the business is still operating.

Except for the dip in sales during months 1 and 2, the sales distribution seems uniform. However, the revenue distribution from months 2 to 11 has more variation and fluctuation.

2.5.2.2 Relations between age and income

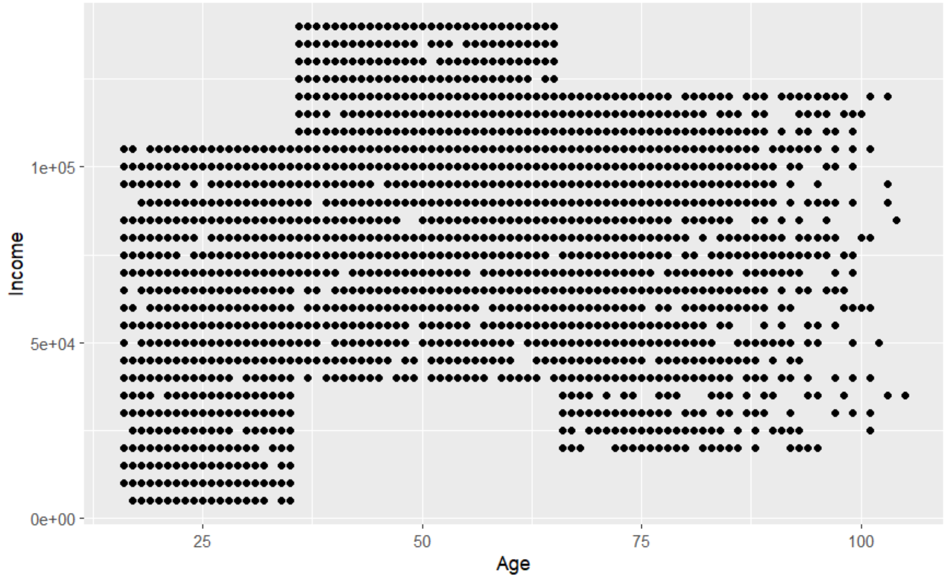


Figure 7 Distribution of age versus income of the customers

Many people younger than 30 do not earn an income, since they are most likely in school or still studying. People between 30 and 60 earn more than $50000 per year, which probably coincides with the minimum wage or a grant from the country. People older than 60 earn less than middle-aged people, but more than young people in general. The grid-like structure of the data is probably due to how the data was obtained (perhaps through a questionnaire with discrete intervals and a minimum and maximum level).

2.5.2.3 Exploring causes for clusters in picking hours and delivery hours

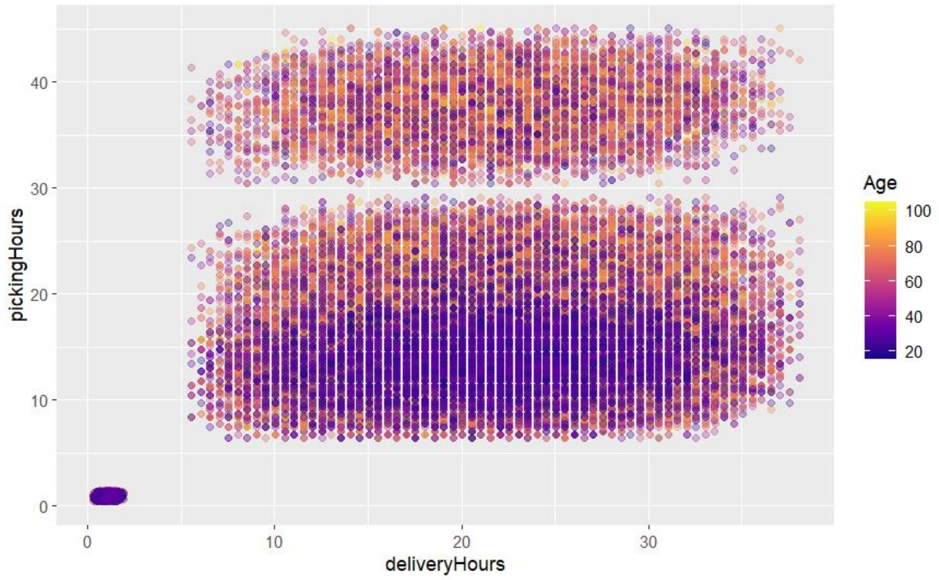


Figure 8 Scatterplot of the delivery hours and service hours distribution by age

There is some relationship between age, picking, and delivery hours. Customers younger than 30 tend to have fewer picking hours. This might just be since they order fewer products at a time.

Is it possible that no sales have picking hours between 29 and 30 hours, because of a break, maybe a weekend (if 6 hours per day for 5 days), or a shift change.

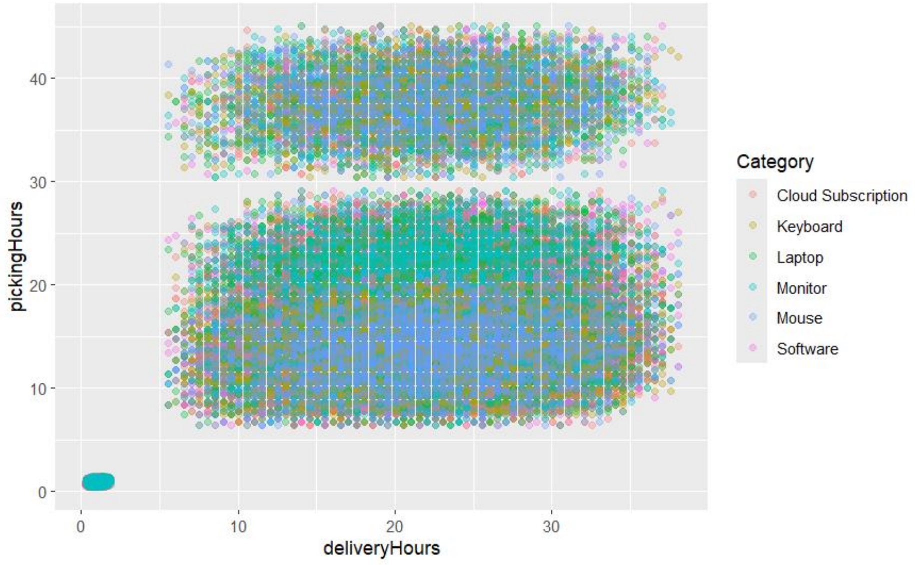


Figure 9 Scatterplot of picking and delivery hours by product category

The clusters in picking hours and delivery hours also appear to be loosely related to the product category.

All in all, more data would be required to figure out the cause of these clusters. Another possible cause is workers who work at different speeds or perhaps automated versus manual picking of products.

2.5.3 Distribution of delivery hours

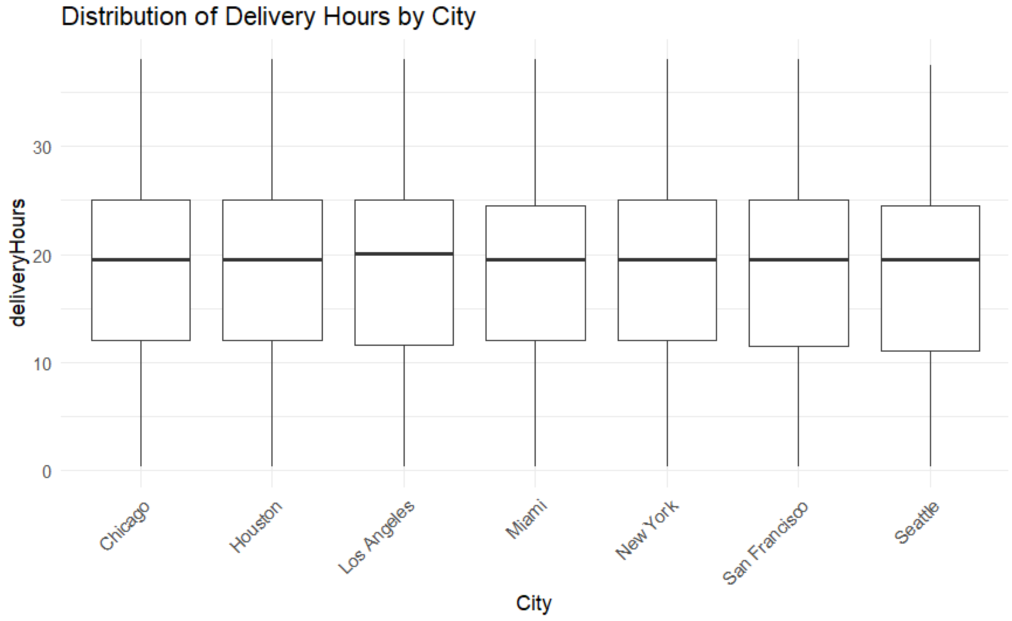


Figure 10 Box plot showcasing uniform distribution of delivery hours regardless of the city.

The distribution of Delivery Hours for the different cities is very similar. Assuming the data provided is accurate, a possible explanation for this is that the company has a branch in each of these cities and will deliver up to a certain distance away from the city. After that, the customer is responsible for shipping, or third-party couriers are used for delivery.

2.6 DATA ANALYSIS AFTER DATA CORRECTION

After an email was received from the business explaining how to correct the errors in the product data sets, the following analysis and amendments have been made.

2.6.1 Solving the mysterious clusters in delivery and picking hours

With the corrected product data set, the cause of the clusters becomes clear:

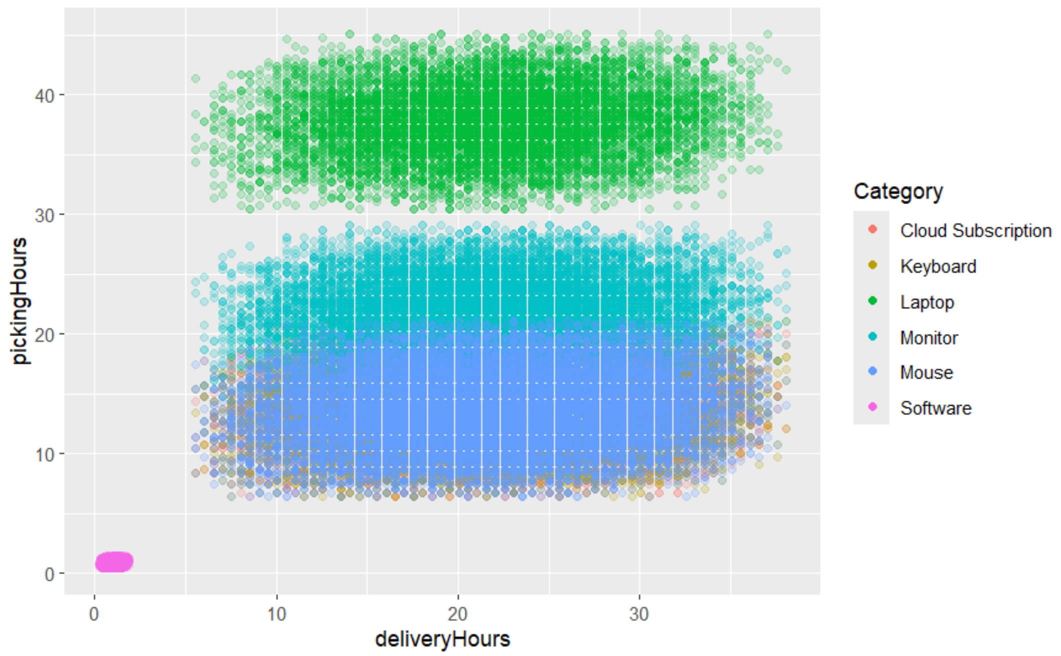


Figure 11 Scatterplot of delivery hours versus picking hours colour coded by the product category

To further investigate the distributions, boxplots are drawn to confirm the distributions of the different categories, especially of keyboards and cloud subscriptions, since those data points are obscured on the scatterplot.

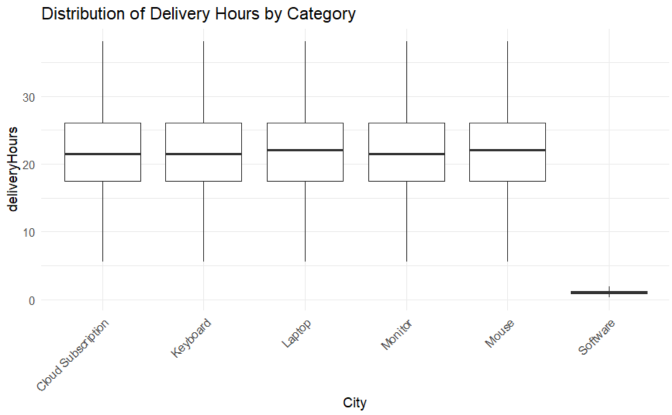


Figure 12 Boxplot of the distribution of delivery hours by product category

Software has little to no delivery time, since it can be installed via the website. It is not a physical package. This argument raises the question of why cloud subscription still takes time to deliver. It might be an error in the dataset. Further data and business insight would be required to answer this question.

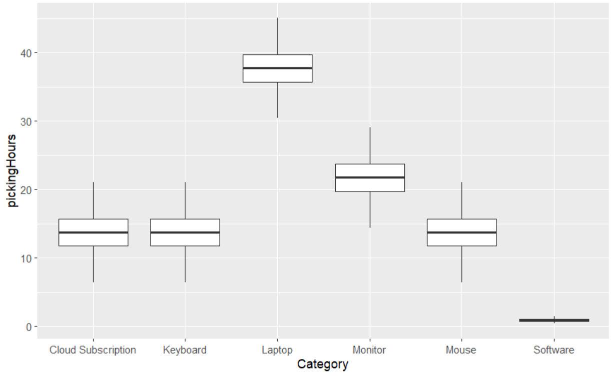


Figure 13 Boxplot of the distribution of picking hours by category

The picking hours for software are also zero, since the customer selects the software they want, and the installation happens automatically. No input is required by the picking facilities. The picking time for the laptop might be so high, since they assemble the laptops on demand. Laptops are also the most complex and expensive physical product in comparison to the rest. Additional security might be in place to prevent theft, which also leads to longer picking hours. Monitors are not as complex or expensive as laptops, but more expensive and complex than keyboards and mice.

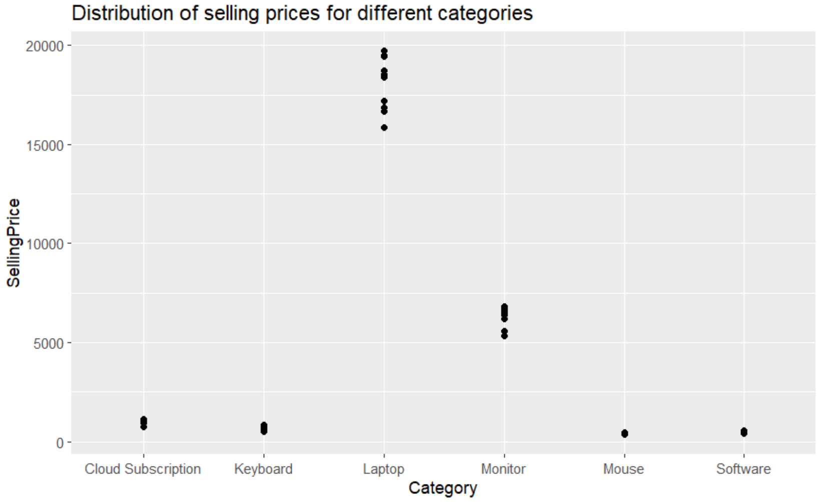


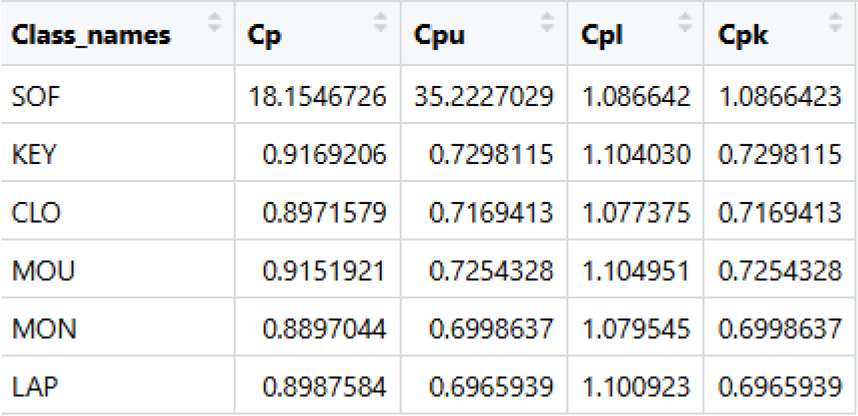
Figure 14 Distribution of selling prices of products

3 STATISTICAL PROCESS CONTROL

3.1 PROCESS CAPABILITY

The process capability factors are calculated to see whether the process in able to deliver according to the customer's specifications.

In this case the upper service level is 32 hours and lower service level is 0 hours



Class_names	Cp	Cpu	Cpl	Cpk
SOF	18.1546726	35.2227029	1.086642	1.0866423
KEY	0.9169206	0.7298115	1.104030	0.7298115
CLO	0.8971579	0.7169413	1.077375	0.7169413
MOU	0.9151921	0.7254328	1.104951	0.7254328
MON	0.8897044	0.6998637	1.079545	0.6998637
LAP	0.8987584	0.6965939	1.100923	0.6965939

Figure 15 Process capabilities for delivery

Only Software delivery can deliver within 30 hours, since its Cpk is greater than 1. From the data analysis, the software had basically 0 hours of delivery time. Thus, software is not a good category to measure delivery capability on, since it requires no delivery. it is concerning that the other product's delivery processes are less than what the customer demands.

The Cp value describes how many times the spread of the data can fit within the service limits. For software, the range of the values is 18 times less than the range that service levels allow. For the other product types, the spread of the data points is larger than the range between the service limits, because the Cp value is less than 1, although still close to 1. Since most Cp values are around 0.9, only 10% of the range of the spread will not be within the service limits, presumably the upper service limit.

3.2 PROCESS CONTROL

3.2.1 Test A: The standard deviation of a single sample is greater than the upper control limit

If the standard deviation of the sample is above the upper control limit, then the product manager must adjust or check the process controls. This may include visiting the factory floor, looking at GPS locations of delivery vehicles, tracking orders, and checking up on drivers.

Delivery only has one out-of-control signal for standard deviation: Sample number 89 for the Keyboard delivery process.

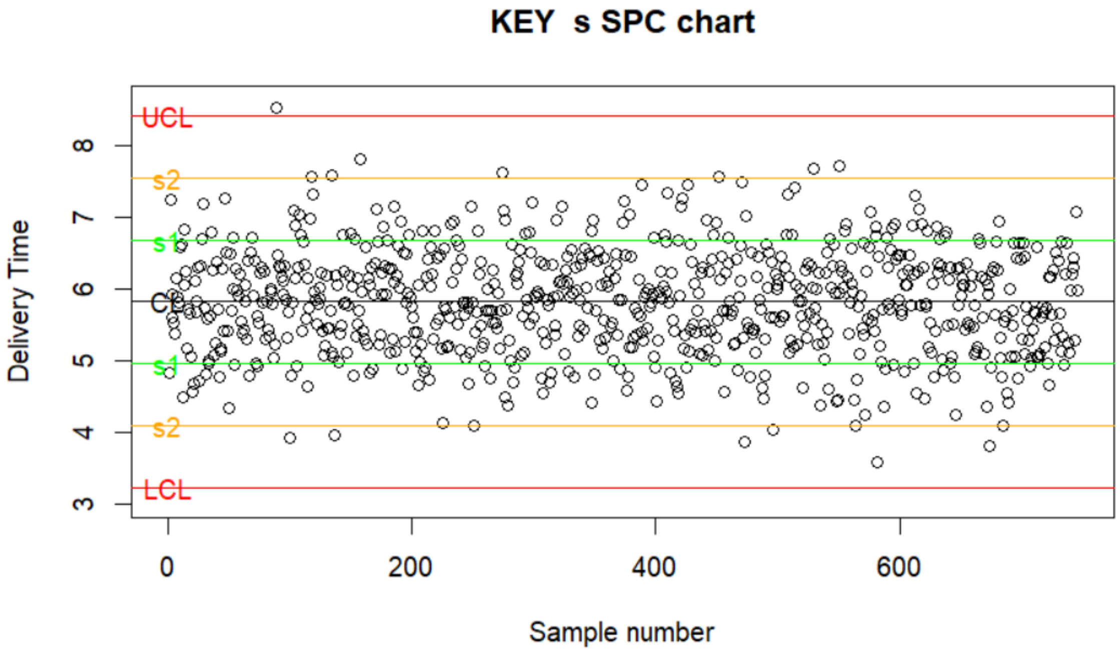


Figure 16 Statistical process control with standard deviation for the delivery time of keyboards

Out of interest, the same test was done on picking hours. A lot of the picking processes have too high standard deviations.

Notice all the samples that are above the upper control limit.

The sample numbers of these process control issues for picking are as follows:

Class Name	Sample numbers of quality issues	Number of out of control signals
Software	183 317 335 353 777 830	6
Keyboard	44 367 443 733	4
Cloud	301 353	2
Mouse	66 82 163 321 334 463 827	7
Monitor	335 437 538	3
Laptop	276	1

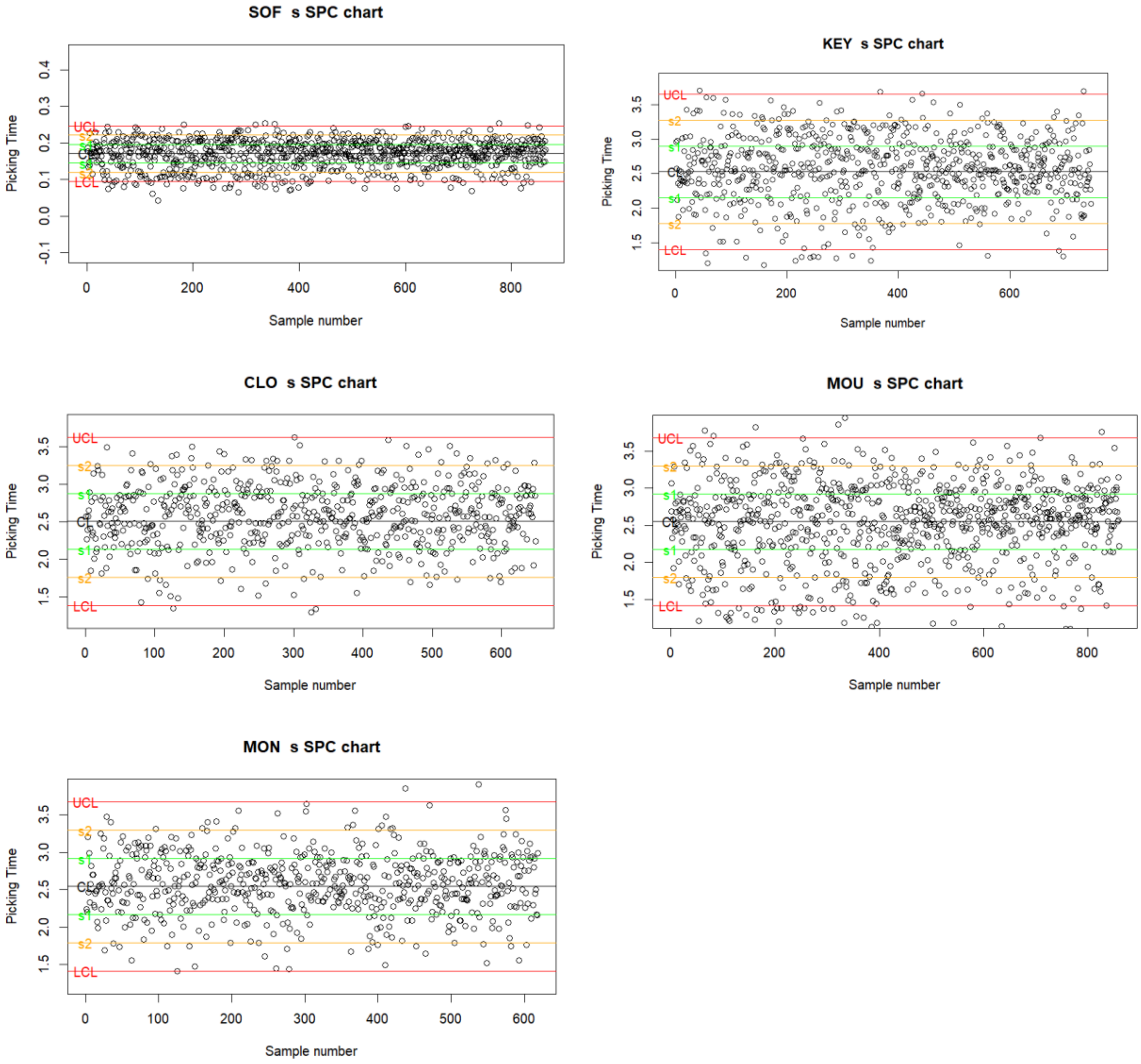
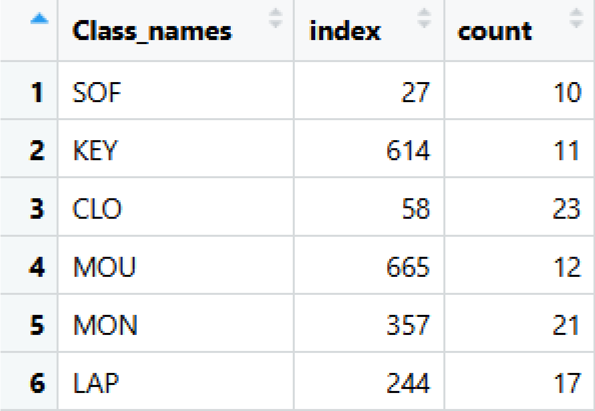


Figure 17 Statistical Process control with standard deviation for picking time

3.2.2 Test B: Most consecutive samples between 1 sigma control limits

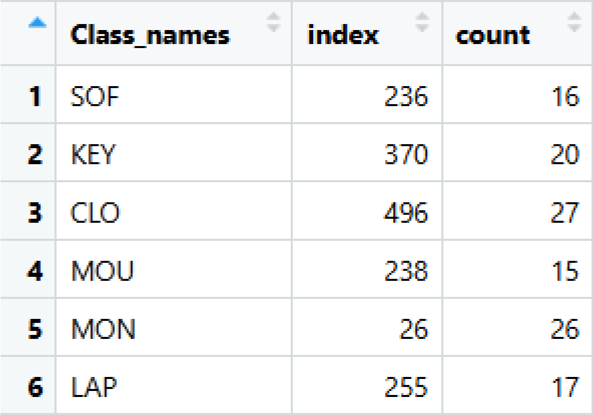
These samples give the process manager a good indication of how the process should operate. The project manager should analyse what happened in the process at these points in time to determine what is being done right. Maybe the workers are motivating each other or having a friendly competition. Maybe an ideal time of the day to do the delivery has been found. Find out why the process is performing so well and mimic those factors in the future.

Table 3 Index numbers for picking processes with the most samples within the +-1 sigma control limits



	Class_names	index	count
1	SOF	27	10
2	KEY	614	11
3	CLO	58	23
4	MOU	665	12
5	MON	357	21
6	LAP	244	17

Table 4 Index numbers for delivery processes with the most samples within the +-1 sigma control limits



	Class_names	index	count
1	SOF	236	16
2	KEY	370	20
3	CLO	496	27
4	MOU	238	15
5	MON	26	26
6	LAP	255	17

3.2.3 Test C: 4 consecutive X-bar samples above the 2-sigma upper control limit.

When this out-of-control signal is received, the process manager thinks something is wrong with the process. An example of a graph with an out-of-control signal follows.

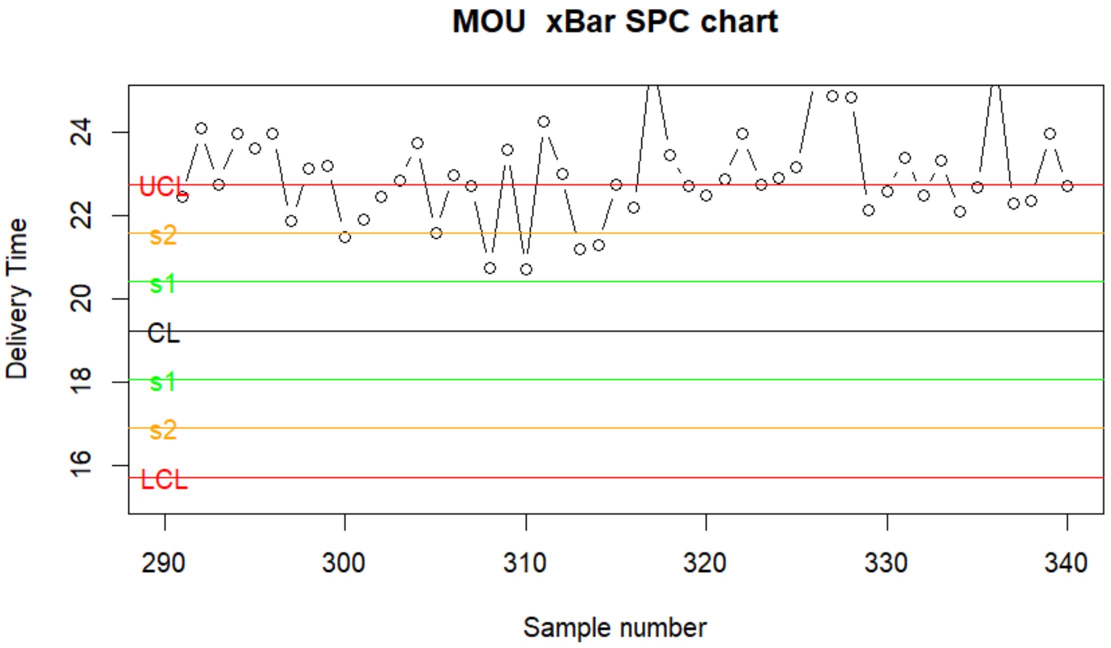
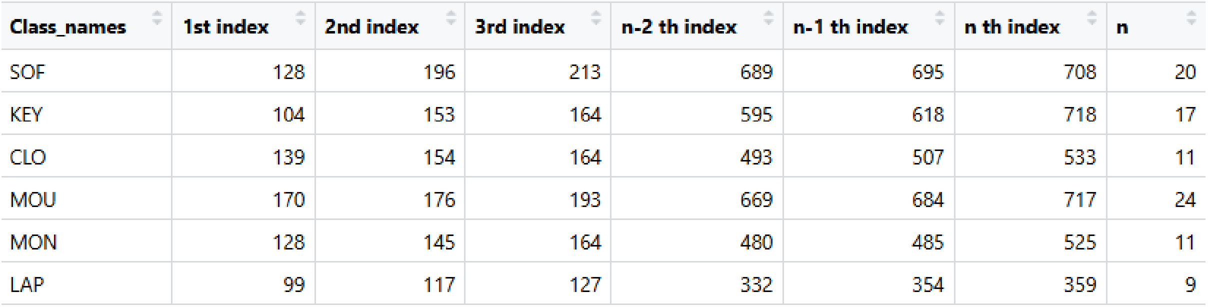


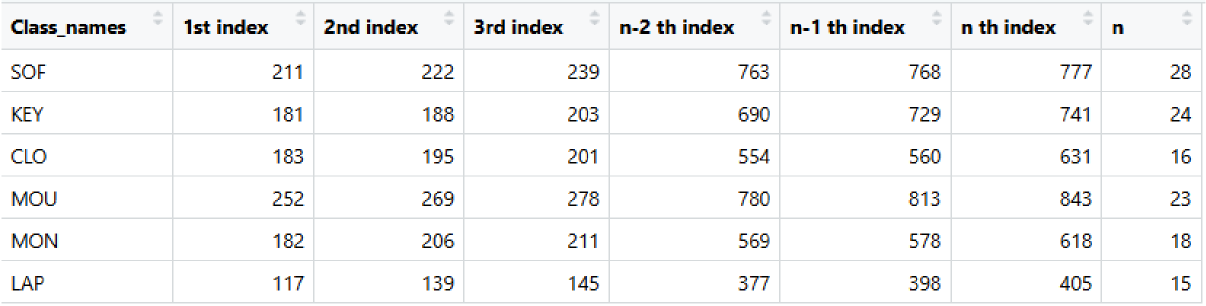
Figure 18 SPC chart for the sample means when the delivery process for mice is out of control

Table 5 Indices of samples with out-of-control signals for picking time.



Class_names	1st index	2nd index	3rd index	n-2 th index	n-1 th index	n th index	n
SOF	128	196	213	689	695	708	20
KEY	104	153	164	595	618	718	17
CLO	139	154	164	493	507	533	11
MOU	170	176	193	669	684	717	24
MON	128	145	164	480	485	525	11
LAP	99	117	127	332	354	359	9

Table 6 Indices and number of samples with out-of-control signals for delivery time



Class_names	1st index	2nd index	3rd index	n-2 th index	n-1 th index	n th index	n
SOF	211	222	239	763	768	777	28
KEY	181	188	203	690	729	741	24
CLO	183	195	201	554	560	631	16
MOU	252	269	278	780	813	843	23
MON	182	206	211	569	578	618	18
LAP	117	139	145	377	398	405	15

Tables 5 and 6 indicate every time the process manager thought the picking and delivery processes were out of control. The letter n represents the number of times the process gave an out-of-control signal due to test C.

4 RISK, DATA CORRECTION, AND OPTIMIZING FOR MAX PROFIT

4.1 THE LIKELIHOOD OF A MANUFACTURER’S ERROR

A manufacturer’s error or type 1 error occurs if there is an out-of-control signal, but the process is still in control.

Using the central limit theorem, it can be assumed that the spread of the sample means and sample standard deviations form a normal distribution.

4.1.1 Test A

The chance of the sample standard deviation being above the upper control limit of the s-chart while the process is still in control.

Let H0 = the process is stable, centred, and in control

We reject H0 if the sample standard deviation is above the upper control limit.

$$P(type\ 1\ error) = \ P(reject\ H0\ |H0\ is\ true)$$

$$= P(s > UCL|s\_bar = CL)$$

$H0\ reject$ can be calculated by calculating the area under the normal curve above the upper control limit. The probability of a manufacturer's error for this type of out-of-control signal is 0.1349898% for all the product categories.

4.1.2 Test B

The type 1 error for the test to find the most consecutive s-samples between the 1 sigma limit to represent a good process is 0. Following the definition of H0 and the formula to calculate it from test A, the type 1 error is 0, since the test does not specify when to reject H0. It only specifies whether the current process setting is a good in-control process to potentially learn from. If the s-samples are not the most consecutive samples in a row, then there is no reason to assume the process is out of control

4.1.3 Test C

The chance that 4 consecutive X-bar samples are above the upper second control limit, but the process is still in control. This can be calculated by

$P(reject\ H0\ |H0\ is\ true)$

$= P(4\ consecutive\ samples\ above\ 2nd\ control\ limit|Process\ still\ in\ control, stable\ and\ centred)$
$= P(1\ sample\ above\ 2nd\ control\ limit|Process\ in\ control) = 0.00002678772\%$

4.2 THE LIKELIHOOD OF A TYPE 2 ERROR

Process appears to be in control if the sample is between the upper and lower control limits.

Centred on process average of: 25.05 litres = CL

UCL = 25.089 L

LCL = 25.011 L

Sd = 0.013

With the type 2 error, the sample mean and standard deviation values is between the LCL and UCL, but the process is actually out of control: The average fill volume is now 25.028, and the sample standard deviation is now 0.017.

$$P(type\ II\ error) = P(accept\ H0|H0\ is\ false) = P(UCL < \mu < LCL|\mu = 25.028, \sigma = 0.017) \\ = 84.11783\%$$

This is a very big chance of a consumer's error. It is suggested that the bottle filling process should adopt more SPC rules to enable them to pick up on faulty batches. They should look at the S-chart in conjunction with the X-bar chart, for example.

5 OPTIMIZE PROFIT

In the given file timeToServe, assume the first column is the number of baristas and the second column is the time in seconds that it takes to serve a customer that day.

The profit per customer is R30 without the personnel cost.

Hiring personnel costs R1000/day/person.

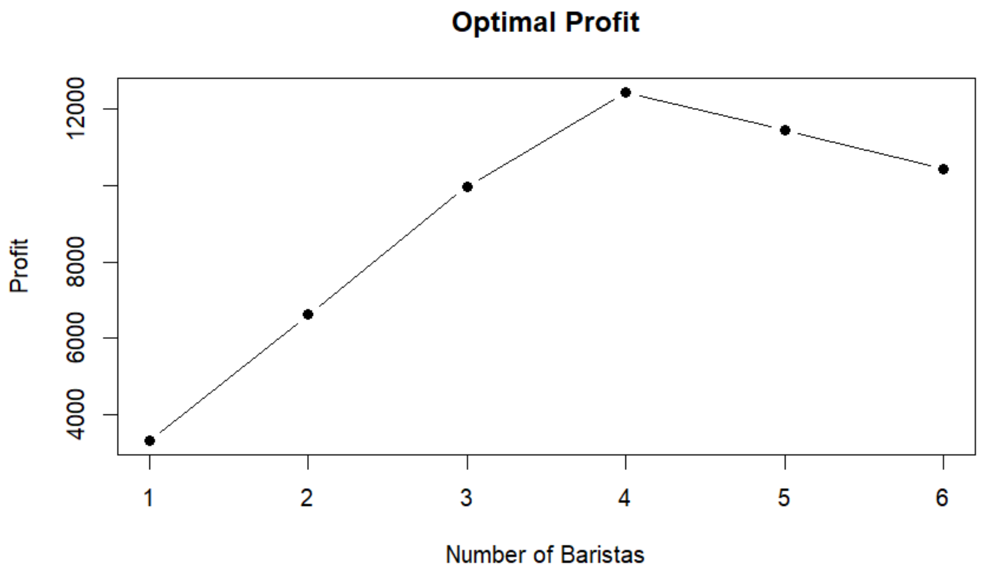
Assume 8-hour working days: 8*60*60=28800 seconds per working day.

Assume each entry in the timeToServe dataset is a customer that is being served.

The average number of customers per day can be calculated using the size of the dataset divided by the number of days in a year. 200000/365

The capacity (number of customers that can be served per day) for the number of baristas can be calculated by 28800/average timeToServe.

The profit is calculated by taking the minimum between the average number of customers per day and the capacity of the baristas, multiplied by R30 of profit per customer and subtracting the personnel cost.



The maximum profit per weekday will occur if 4 baristas are employed. The average profit per day is R12438.36 for the average number of customers per day.

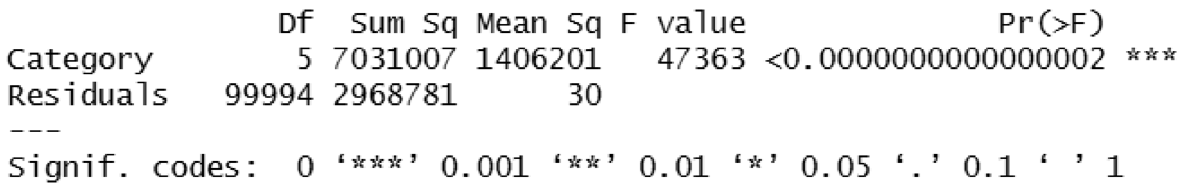
6 ANALYSIS OF VARIANCE

An analysis of variance has been done to confirm or disprove some of the suspicions from data analysis.

6.1 IS THERE A DIFFERENCE BETWEEN THE DELIVERY HOURS FOR THE DIFFERENT PRODUCT TYPES?

Ho=There is no difference in the means for the delivery hours for the different product types.

Ha=There is a significant difference in the delivery hours for the different product types.



```
               Df  Sum Sq Mean Sq F value             Pr(>F)    
Category        5 7031007 1406201   47363 <0.0000000000000002 ***
Residuals   99994 2968781      30                               
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1
```

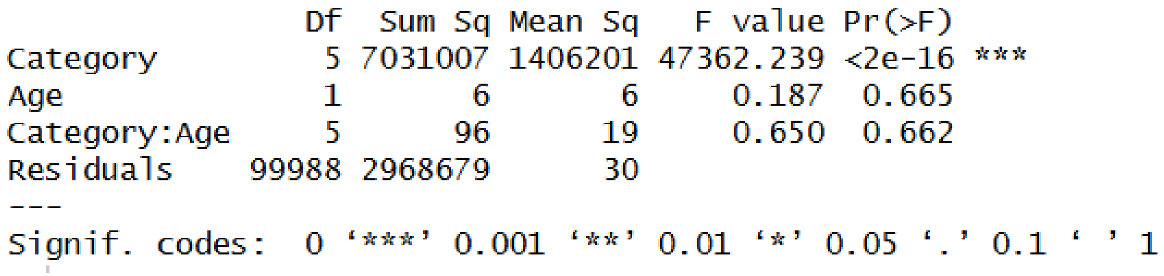
From the output of the code, it can be seen from the large F value and the small P value that there is a significant difference in delivery times for the different product types. Thus reject H0, meaning there is in fact a big difference between delivery hours for different product types.

6.2 Is there interaction between Age and Category for Delivery Hours?

A two-way ANOVA can be used to determine this.

Ho=There is no interaction between Category and Age.

Ha=There is interaction between Category and Age.



```
                 Df  Sum Sq Mean Sq   F value Pr(>F)    
Category          5 7031007 1406201 47362.239 <2e-16 ***
Age               1       6       6     0.187  0.665    
Category:Age      5      96      19     0.650  0.662    
Residuals     99988 2968679      30                     
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1
```

Since the p-value for Category: Age is 0.662, which is large, we accept Ho, which means there is no significant interaction between Category and Age.

7 Reliability of Service

7.1 Days of reliable service

The mean and variance for a binomial distribution are as follows. Since the mean and variance can be calculated in R, the two equations can be solved simultaneously to obtain p and n, where p is the probability of reliable service and n is the sample size.

$$E(x) = np = 15.58438$$

$$\sigma^2 = np(1-p) = 0.4758161$$

$$(1-p) = \frac{12811.3}{79.4} = 0.0305316$$

$$p = 1 - 0.0305316 = 0.9694684$$

$$n = \frac{E(x)}{p} = \frac{15.58438}{0.9694684} = 16.07518 \approx 16$$

The probability of reliable service is 91.5703%, which can be calculated with 1-pbinom in R.

Thus, 363.5341 out of the 397 days of the sample are expected to have reliable service.

Finally, $0.915703 \times 365 = 334.2316$ days of reliable service per year can be expected.

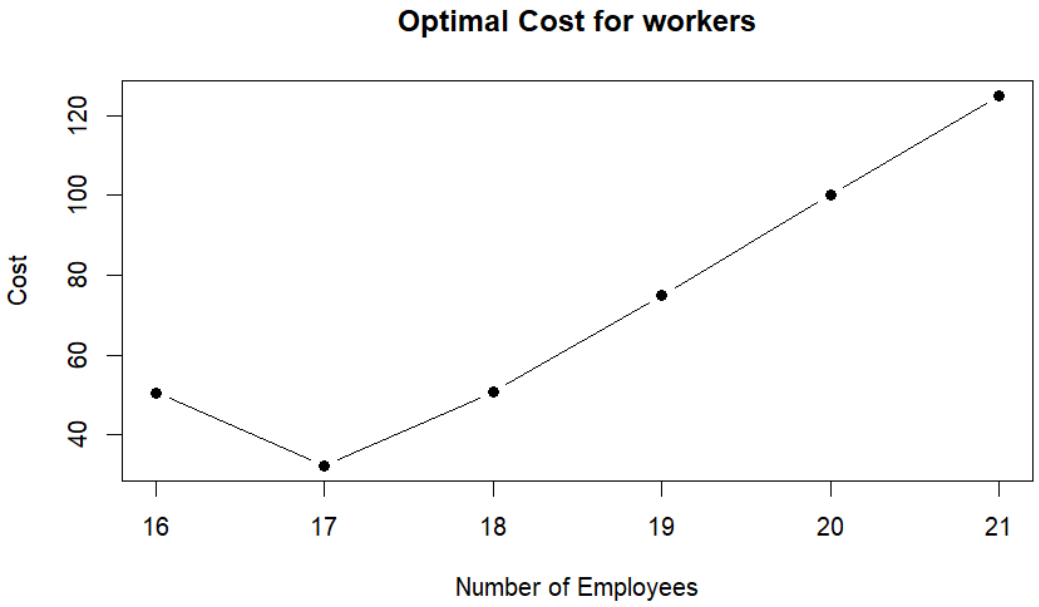
7.2 Optimizing profit for a car rental agency

It is assumed that if the number of workers is increased by one, the distribution will just be translated to the right by one. In other words, the variance stays the same, but the expected value increases by one.

It is assumed there are 30 days in a month and 365 days in a year.

It costs R20,000 in lost sales for every day the service is unreliable and R25000 for each extra worker that is employed in addition to the existing 16 workers.

With these assumptions, the optimal number of employees is found to be 17 employees (one extra employee). This will result in a minimum cost of R32127.19 per month, or R390880.8 per year. A minimum cost is assumed to result in a maximum profit.



8 Conclusion

In this project, quality assurance was applied to various scenarios: from an electronics company to a coffee shop to a car rental agency. Data analysis, statistical process control, data manipulation, correction and optimization were applied. Process capabilities, the probability of customers' and manufacturers' errors and service reliability were calculated.

Even though most of the provided data were test problems, this project gave a taste of what the real application might look like. The skills learnt would help to secure success in the ever-changing business climate and workplace

9 Bibliography

Dirkse van Schalkwyk, T. (2025) Cheat Sheet for Basic Data Analysis in R. [online] Stellenbosch: Stellenbosch University. Available at: https://stemlearn.sun.ac.za/pluginfile.php/251721/mod_resource/content/1/DataAnalysisCheatSheet.pdf (Accessed: 17 September 2025)

Dirkse van Schalkwyk, T. (2025) QA344 Statistics R. [online] Stellenbosch: Stellenbosch University. Available at: https://stemlearn.sun.ac.za/pluginfile.php/65516/mod_resource/content/5/QA344%20Statistics.pdf (Accessed: 17 September 2025)

Dirkse van Schalkwyk, T. (2025) Short summary of SPC and Limits. [online] Stellenbosch: Stellenbosch University. Available at: https://stemlearn.sun.ac.za/mod/resource/view.php?id=55515 (Accessed: 1 October 2025)

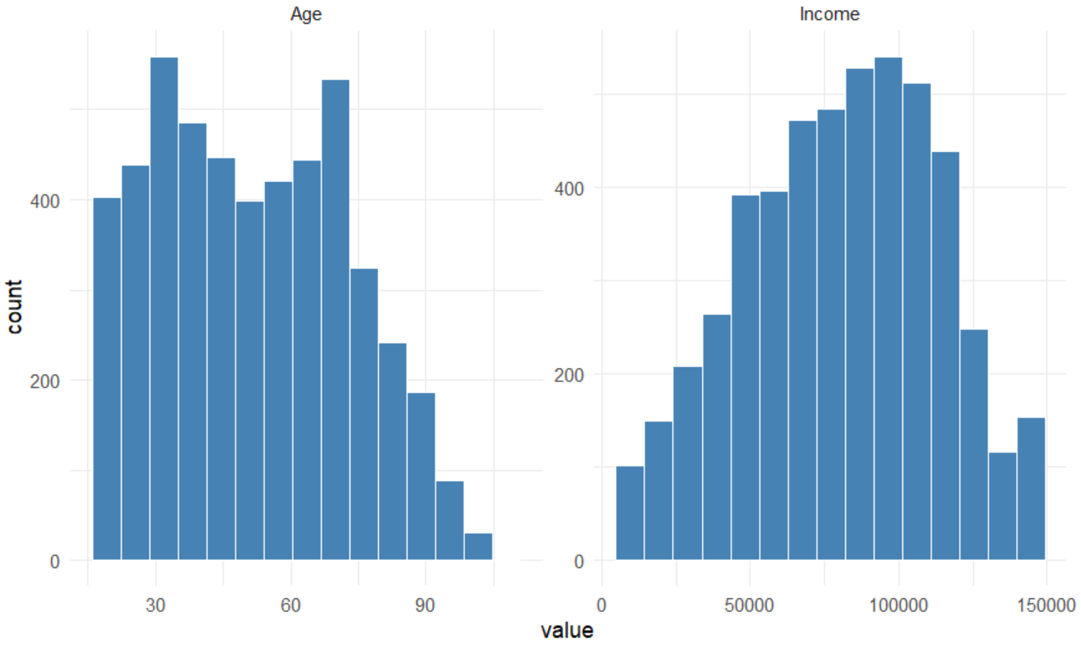
OpenAI. (2025) ChatGPT [AI language model]. Available at: https://chat.openai.com/ (Accessed: 24 October 2025).

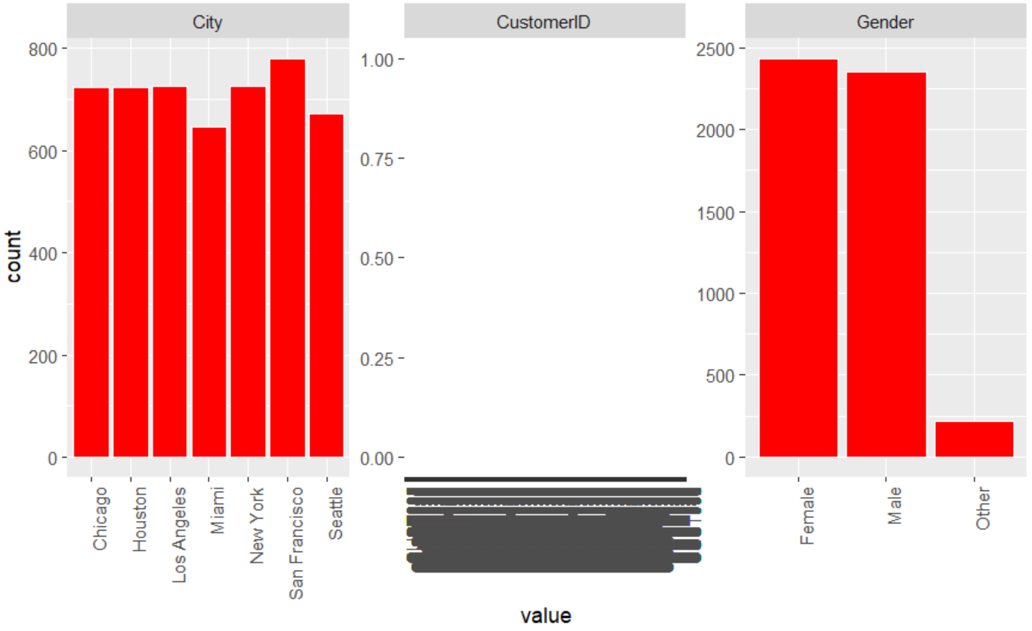
ST HDA. (no date) MANOVA Test in R: Multivariate Analysis of Variance. [online] Available at: http://www.sthda.com/english/wiki/manova-test-in-r-multivariate-analysis-of-variance (Accessed: 18 October 2025).

xAI. (2025) Grok AI [AI language model]. Available at: https://x.ai/ (Accessed: 24 October 2025).

APPPENDIX A

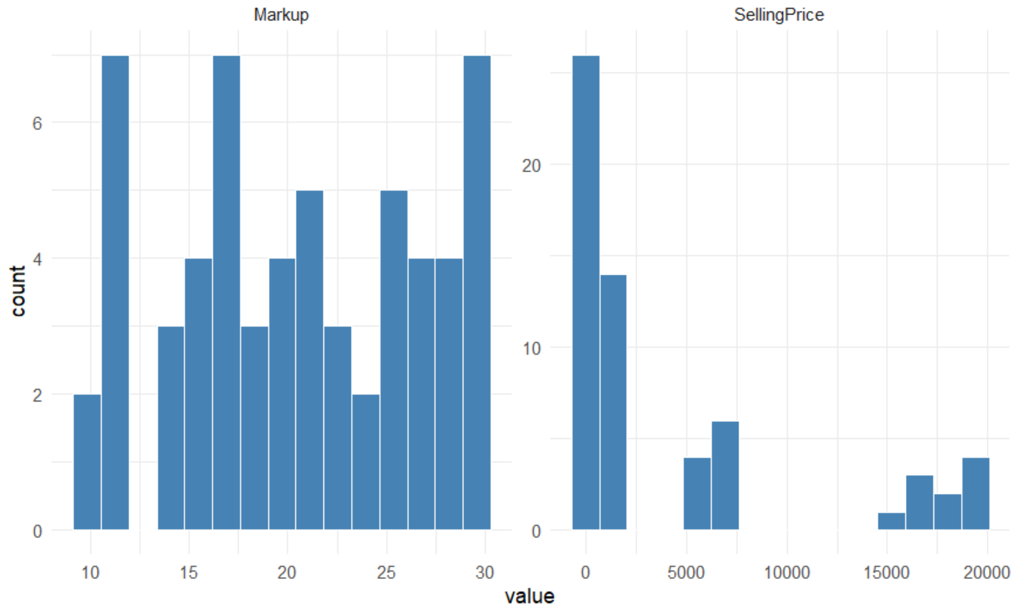
Customer Data

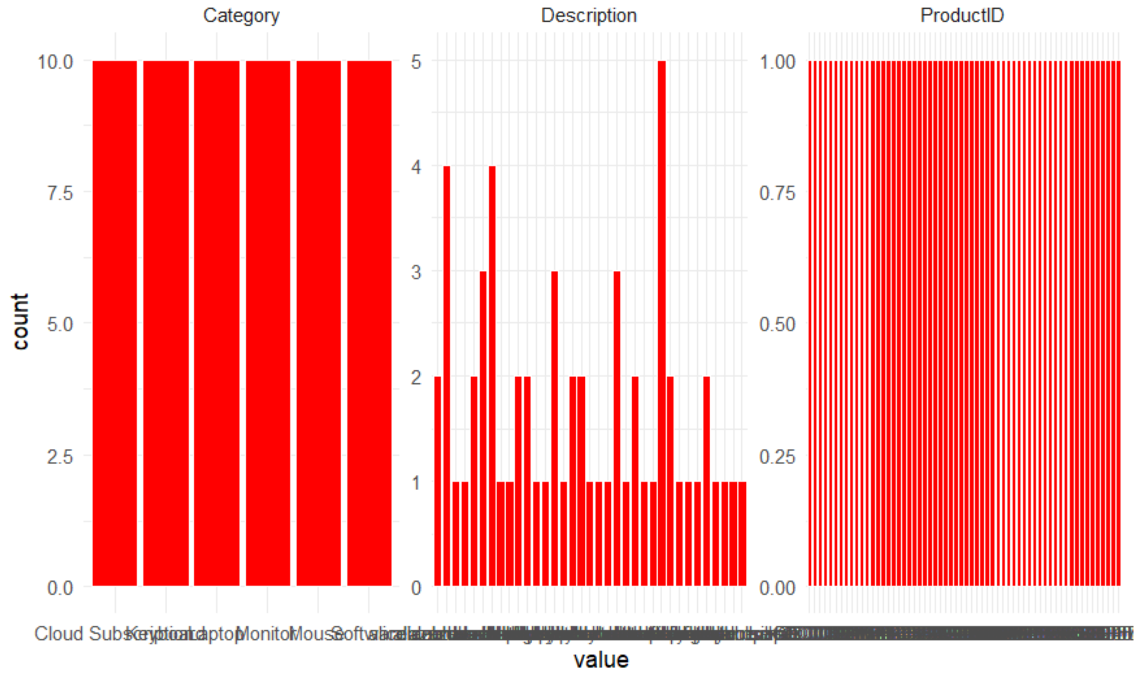




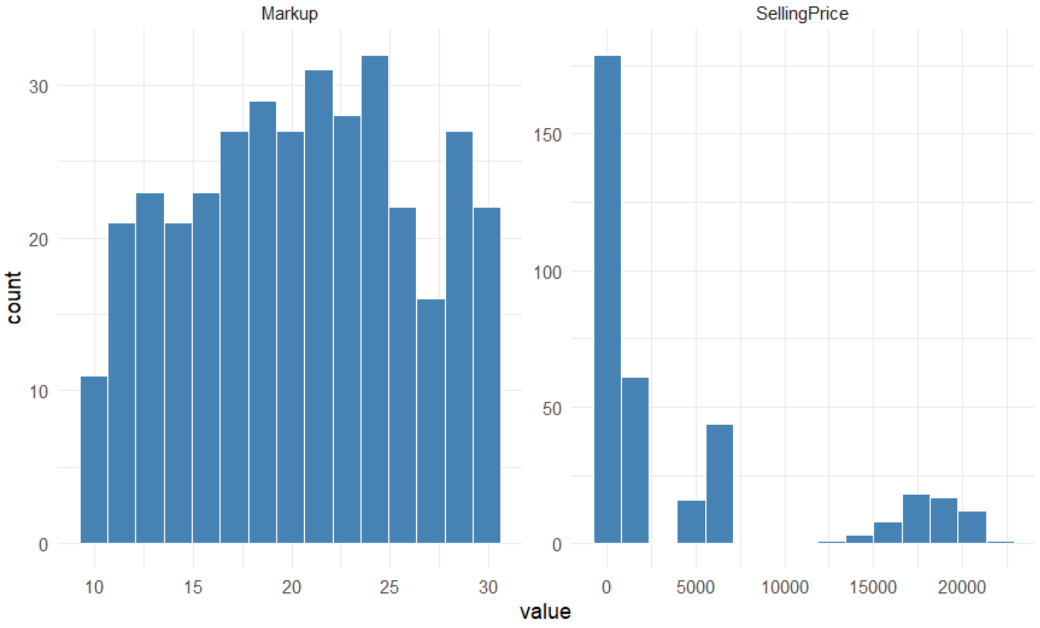
From the bar plot it is clear that Customer ID is a unique identifier for each customer.

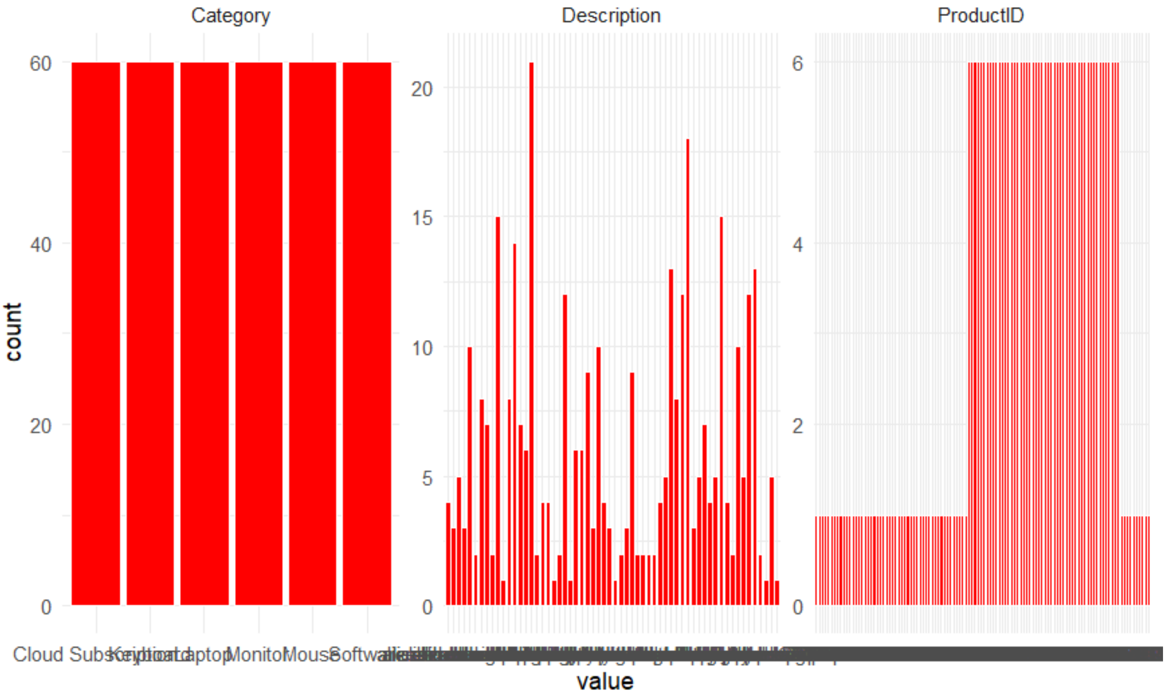
Product data



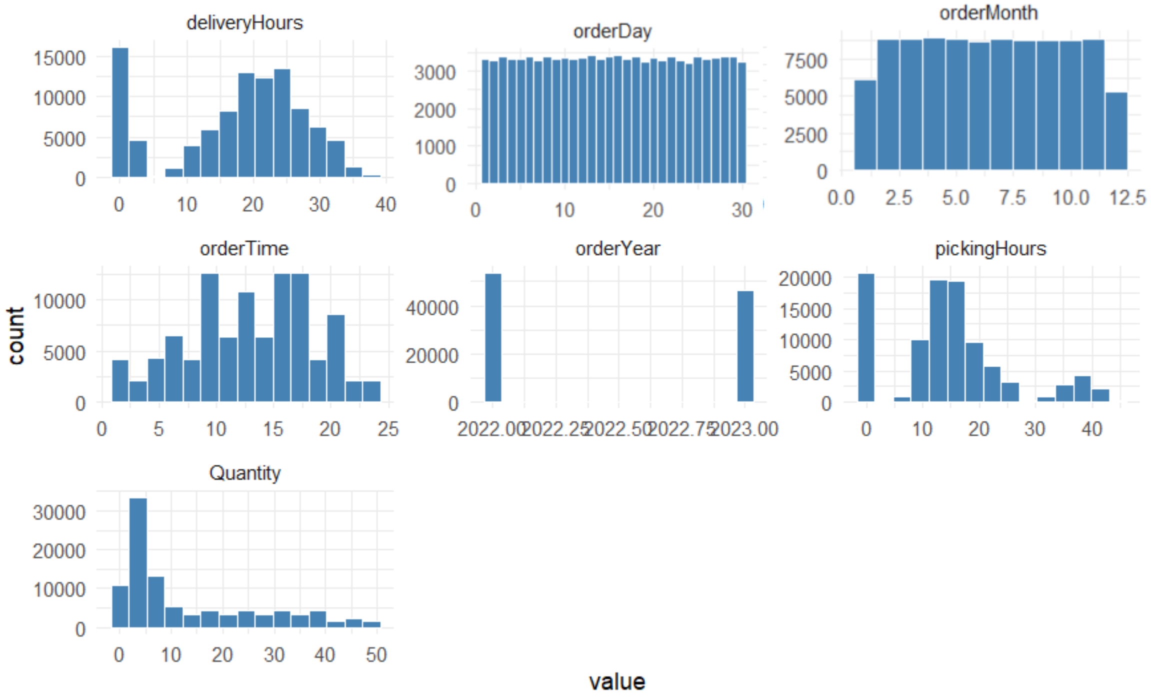


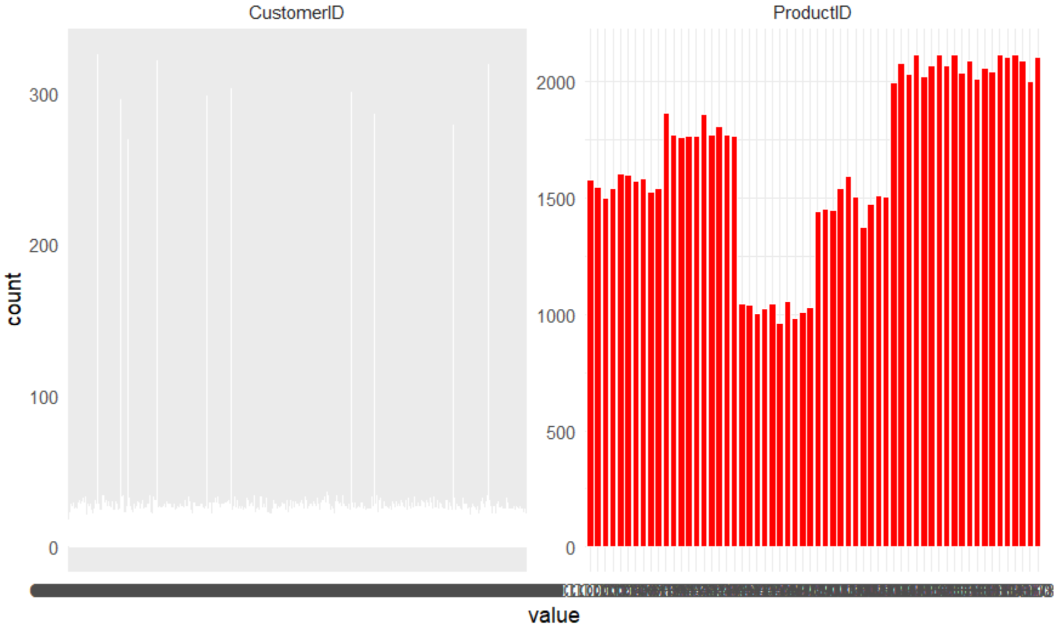
Product head office data



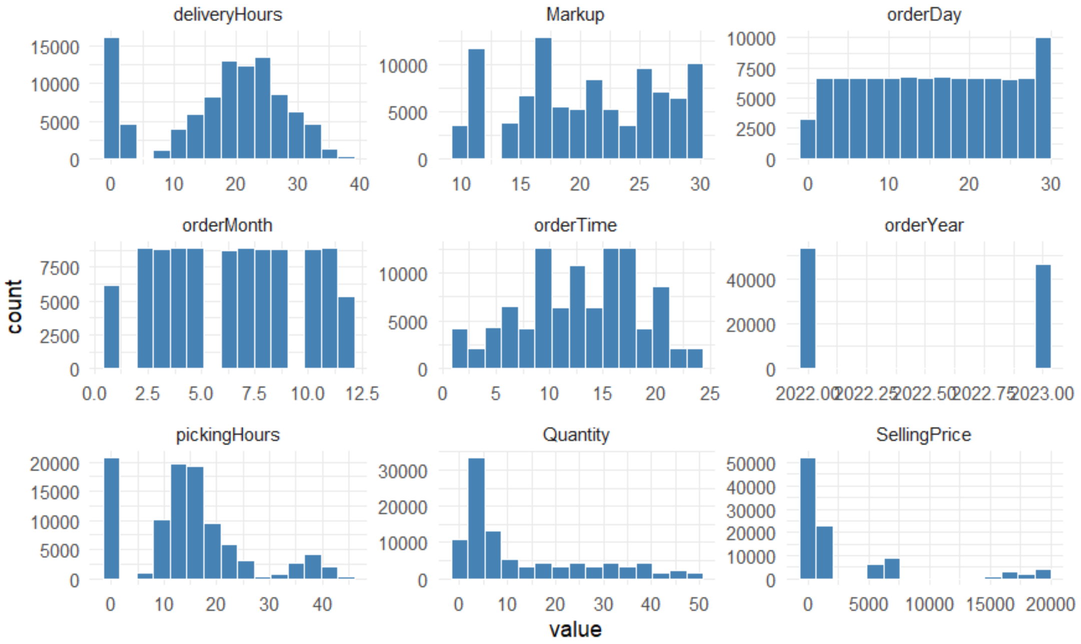


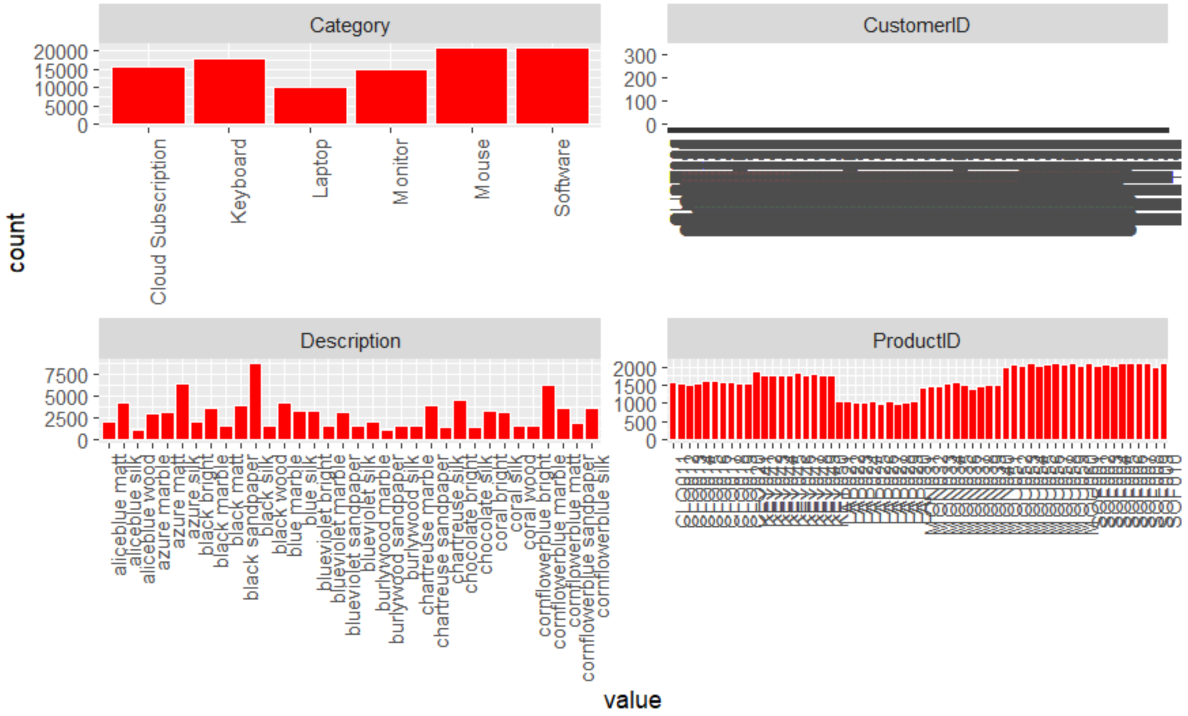
Sales 2022 and 2023 data





AFTER DATA CORRECTION


deliveryHours
Markup
orderDay
orderMonth
orderTime
orderYear
pickingHours
Quantity
SellingPrice
count
value


Category
Cloud Subscription
Keyboard
Laptop
Monitor
Mouse
Software
CustomerID
Description
aliceblue matt
aliceblue silk
aliceblue wood
azure marble
azure matt
azure silk
black bright
black marble
black matt
black sandpaper
black silk
black wood
blue marble
blue silk
blueviolet bright
blueviolet marble
blueviolet sandpaper
blueviolet silk
burlywood marble
burlywood sandpaper
burlywood silk
chartreuse marble
chartreuse sandpaper
chartreuse silk
chocolate bright
chocolate silk
coral bright
coral silk
coral wood
cornflowerblue bright
cornflowerblue marble
cornflowerblue matt
cornflowerblue sandpaper
cornflowerblue silk
ProductID
count
value