The Research Monograph Series in Computing, Electrical & Communication Networks
Vol. 1 (01), April 2023, pp. 19–43

# 4

# RESULTS AND DISCUSSION

## 4.1 INTRODUCTION

This study is based on primary data recover directly from the company´s management system. In this chapter, the researcher addresses the results for business questions and analyse the results in comparison with not only the past years of sales in the company but also the results of similar research.

## 4.2 RESULTS

The first step towards the results is the data preparation process. This is the process of cleaning and transforming raw data before processing and analysing it. It is an essential step prior to processing and often requires reformatting data, making changes, and combining data sets to enrich the data. Figure 1 is a flowchart of the process required for data analysis.

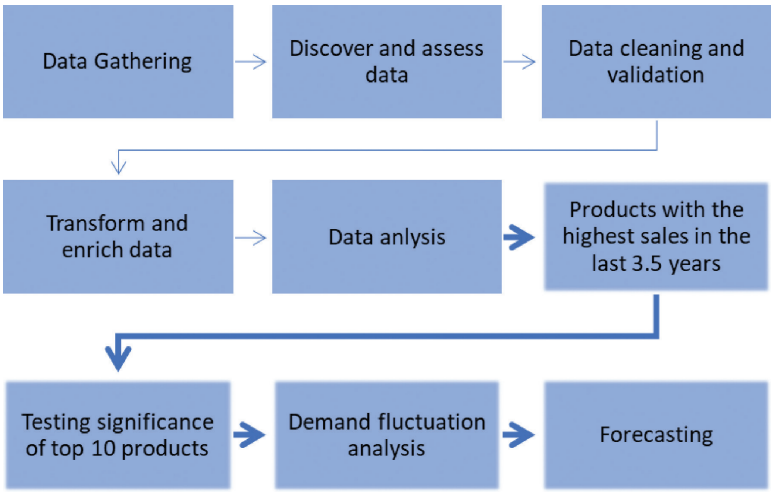



**Figure 1.** Flowchart of data analysis process.

The figure shows the different steps required for the data analysis process that will lead one to the results. The first step is data gathering, which refers to the process of data collection. Once one has obtained the data, they proceed to the discover and assess process, which refers to getting to know and understand the data collected. Then next step is data cleaning and validation; here the data set is corrected or restructured by eliminating incorrect, corrupt, incorrectly formatted, duplicate, or incomplete data. Next is the transformation and enrichment of the data, which refers to the process of updating format or value entries to make the data more easily understood by a wider audience; it also includes the action of adding or connecting data to provide deeper insights. Finally, there is the data analysis process, which can be different according to the goals and objectives. For this research, the data analysis process is focused on the four objectives: finding the products with the highest sales, testing the significance of the top 10 products, understanding the demand fluctuations, and forecasting the sales.

### 4.2.1 Gather Data

The data preparation process begins with finding the right data. Here one can see a snapshot of the data the researcher has gather from the company´s information system (Figure 2).

### 4.2.2 Discover and Assess Data

After collecting the data, it is important to discover each data set. This step is about getting to know the data and understanding what must be done before the data become useful in a particular context. The data set contains 18 variables by 19,168 entries, and each variable is explained in Table 2.

| Región | Clie | NombreZon | IdCliente | NombreClie | IdProducto | DescProduct | Total Unida | UM | Familia | Ingr. Activo | FechaFactur | UNeaUave | TransTipoDi | TransDoc | PrecioUnita | Venta | NUMFAC | TIPODOCFAC |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| - | - | | 12372 | COSMOCEL, S | ASA110 | ASA BLANCA | 1280 | PZ | ASA | Sin I.A. | 1/2/2019 | 00002-SO-21 | SO | 21941 | 2.13 | 2726.4 | 20630 | EF |
| - | - | | 3674 | AGRICULTUF | VARIOS | OTROS INGR | 0 | PZ | INTERESES N | Sin I.A. | 1/3/2019 | 00002-CD-31 | CD | 315 | 1 | -2309.97 | 2039 | CE |
| NACIONAL | LOCAL | | 67265 | AGROQUIMI | BTWS040 | BOT TWISTA | 13728 | PZ | BOTELLA | Sin I.A. | 1/3/2019 | 00002-SO-21 | SO | 21943 | 5.59 | 76739.52 | 20632 | EF |
| NACIONAL | LOCAL | | 67265 | AGROQUIMI | T63020 | TAPA 63 BLA | 13728 | PZ | TAPA | Sin I.A. | 1/3/2019 | 00002-SO-21 | SO | 21943 | 2.16 | 29652.48 | 20632 | EF |
| NACIONAL | LOCAL | | 65570 | SINTESIS Y F | EMS042 | BOT MISIL 5( | 0 | PZ | BOTELLA | Sin I.A. | 1/7/2019 | 00002-CD-31 | CD | 316 | 2.83 | -30046.11 | 2040 | CE |
| NACIONAL | LOCAL | | 65570 | SINTESIS Y F | EMS043 | BOT MISIL 5( | 0 | PZ | BOTELLA | Sin I.A. | 1/7/2019 | 00002-CD-31 | CD | 316 | 3.02 | -36747.36 | 2040 | CE |
| NACIONAL | LOCAL | | 65570 | SINTESIS Y F | K-PAX-M-001 | EMPECO 10 L | 0 | PZ | GARRAFA PC | Sin I.A. | 1/7/2019 | 00002-CD-31 | CD | 316 | 22.94 | -84098.04 | 2040 | CE |
| NACIONAL | LOCAL | | 65570 | SINTESIS Y F | T45020 | TAPA 45 BLA | 0 | PZ | TAPA | Sin I.A. | 1/7/2019 | 00002-CD-31 | CD | 316 | 1.17 | -27631.89 | 2040 | CE |
| NACIONAL | LOCAL | | 65570 | SINTESIS Y F | T63030 | TAPA 63 BL/ | 0 | PZ | TAPA | Sin I.A. | 1/7/2019 | 00002-CD-31 | CD | 316 | 2.29 | -6077.66 | 2040 | CE |
| NACIONAL | LOCAL | | 59571 | MARCO ANTI | MDES060 | CONTAMINA | 3160 | KG | DESPERDICI( | Sin I.A. | 1/7/2019 | 00002-SO-21 | SO | 21950 | 3.5 | 11060 | 20641 | EF |
| MATRIZ | VENTAS COR | | 66677 | AGRO DESAF | EMS042 | BOT MISIL 5( | 7436 | PZ | BOTELLA | Sin I.A. | 1/7/2019 | 00002-SO-21 | SO | 21951 | 5.15 | 38295.4 | 20642 | EF |
| MATRIZ | VENTAS COR | | 66677 | AGRO DESAF | T45020 | TAPA 45 BLA | 7436 | PZ | TAPA | Sin I.A. | 1/7/2019 | 00002-SO-21 | SO | 21951 | 1.05 | 7807.8 | 20642 | EF |
| NACIONAL | LOCAL | | 65556 | ESPECIALIDA | ASA010 | ASA BLANCA | 1715 | PZ | ASA | Sin I.A. | 1/7/2019 | 00002-SO-21 | SO | 21952 | 2.1 | 3601.5 | 20643 | EF |
| NACIONAL | LOCAL | | 65556 | ESPECIALIDA | GASL010 | BOT TWISTA | 1715 | PZ | GARRAFA TV | Sin I.A. | 1/7/2019 | 00002-SO-21 | SO | 21952 | 14.85 | 25467.75 | 20643 | EF |
| NACIONAL | LOCAL | | 65556 | ESPECIALIDA | T63020 | TAPA 63 BLA | 1715 | PZ | TAPA | Sin I.A. | 1/7/2019 | 00002-SO-21 | SO | 21952 | 1.71 | 2932.65 | 20643 | EF |
| NACIONAL | LOCAL | | 65555 | DOW AGROS | EMS041 | BOT EMPECO | 40320 | PZ | BOTELLA | Sin I.A. | 1/7/2019 | 00002-SO-21 | SO | 21953 | 4.68 | 188697.6 | 20644 | EF |
| NACIONAL | LOCAL | | 65555 | DOW AGROS | T45310 | TAPA 45 ROJ | 40320 | PZ | TAPA | Sin I.A. | 1/7/2019 | 00002-SO-21 | SO | 21954 | 1.51 | 60883.2 | 20645 | EF |
| NACIONAL | LOCAL | | 107456 | 4E GLOBAL S | 2091371 | PEZ KIDS RO | -75075 | PZ | BOTELLA | Sin I.A. | 1/8/2019 | 00002-CA-12 | CA | 1257 | 1.2 | -90090 | 2041 | CE |
| NACIONAL | LOCAL | | 107456 | 4E GLOBAL S | 2091328 | VERT PLANA | -6183 | PZ | BOTELLA | Sin I.A. | 1/8/2019 | 00002-CA-12 | CA | 1258 | 1.92 | -11871.36 | 2042 | CE |
| NACIONAL | LOCAL | | 107456 | 4E GLOBAL S | 2091328 | VERT PLANA | -61830 | PZ | BOTELLA | Sin I.A. | 1/8/2019 | 00002-CA-12 | CA | 1259 | 1.92 | -118713.6 | 2043 | CE |
| NACIONAL | LOCAL | | 21745 | AGROQUIMI | ASA010 | ASA BLANCA | 8232 | PZ | ASA | Sin I.A. | 1/8/2019 | 00002-SO-21 | SO | 21961 | 2.1 | 17287.2 | 20646 | EF |
| NACIONAL | LOCAL | | 21745 | AGROQUIMI | EMS050 | BOT EMPECO | 11154 | PZ | BOTELLA | Sin I.A. | 1/8/2019 | 00002-SO-21 | SO | 21961 | 4.69 | 52312.26 | 20646 | EF |
| NACIONAL | LOCAL | | 21745 | AGROQUIMI | GASL040 | BOT TWISTA | 8232 | PZ | GARRAFA TV | Sin I.A. | 1/8/2019 | 00002-SO-21 | SO | 21961 | 14.1 | 116071.2 | 20646 | EF |
| NACIONAL | LOCAL | | 21745 | AGROQUIMI | T45180 | TAPA 45 AZU | 11700 | PZ | TAPA | Sin I.A. | 1/8/2019 | 00002-SO-21 | SO | 21961 | 1.12 | 13104 | 20646 | EF |

**Figure 2.** Data sample.

**Table 2.** Variable description.

| ***Variable*** | ***Description*** | ***Number of categories*** |
|---|---|---|
| *Región Cliente* | Clients' region | 6 |
| *NombreZonaVenta Cliente* | Sales Region | 4 |
| *IdCliente* | Client´s unique number | 113 |
| *NombreCliente* | Client´s name | 112 |
| *IdProducto* | Product´s unique number (SKU) | 196 |
| *DescProducto* | Product´s description | 191 |
| *Total Unidades* | Total of pieces sold | n/a* |
| *UM* | Unit of Measure | 4 |
| *Familia* | Product´s family | 21 |
| *Ingr. Activo* | Active Ingredient | 1 |
| *FechaFactura* | Invoice Date | 4 years |
| *LineaLlave* | Key Line of invoice | n/a* |
| *TransTipoDoc* | Document type | 10 |
| *TransDoc* | Document number | n/a* |
| *PrecioUnitario* | Unite price | n/a* |
| *Venta* | Total sale | n/a* |
| *NUMFAC* | Invoice Number | n/a* |
| *TIPODOCFAC* | Invoice type | 2 |

*n/a: not applies. * Unique information per entry, not specific number of categories available.*

A descriptive analysis and a normality test can be helpful in this part of the data preparation. Both analyses have been done using SPSS, and the results for the descriptive analysis are shown in Table 3 and for the normality test in Table 4.

Table 3 shows the results of the descriptive analysis done using SPSS. The descriptive statistics can be classified into three main types: frequency measures, central tendency measures, and dispersion measures. Categorical data are usually measured by frequency, while quantitative data are measured by other measures. This analysis shows the valid frequency and the lost frequency of different variables as well as the mean (average) and the standard deviation of the numerical variables. This means that the average amount of pieces sold per entry was 16,201, the average unites price was 239.87 MXN, and the average sale was 59,920 MXN. In statistics, the standard deviation indicates how far apart values

**Table 3.** Descriptive analysis statistics.

| Statistics | | | | | | | |
|---|---|---|---|---|---|---|---|
| | | **Product´s unique number (SKU)** | **Total of pieces sold** | **Product´s family** | **Invoice Date** | **Unite price** | **Total sale** |
| N | Valid | 19,168 | 19,168 | 19,168 | 19,168 | 19,168 | 19,168 |
| | Lost | 0 | 0 | 0 | 0 | 0 | 0 |
| Mean | | | 16,201 | | | $239.87 | $59,920.11 |
| Standard deviation | | | 33,243 | | | $7,289.39 | $123,436.194 |

**Table 4.** Normality test.

| | **Kolmogorov–Smirnov[a]** | | |
|---|---|---|---|
| | **Statistic** | **df.** | **Sig.** |
| Total of pieces sold | 0.251 | 19,168 | 0.000 |

[a] Lilliefors significance correction.

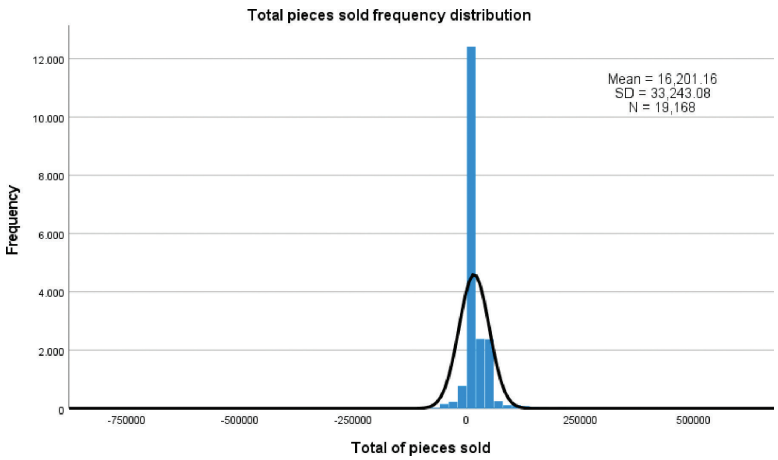



**Figure 3.** Total pieces sold distribution.

are from the mean. Figure 3 shows the distribution of the total pieces sold by frequency.

Table 4 shows the results from the normality test done using SPSS. SPSS works with two well-known tests of normality, namely, the Kolmogorov–Smirnov test and the Shapiro–Wilk test. The Shapiro–Wilk

test is more appropriate for sample sizes under 50; therefore, the Kolmogorov–Smirnov test was selected for this research. Normally distributed populations (within certain tolerances) can be determined by a normality test. There are a number of statistical tests that require a normally distributed sample population, such as the Student's t-test and the one-way and two-way ANOVA. It is not possible to trust the results of the tests if the assumption of normality is not valid. As one can see, the P-value (Sig.) in the normality test is less than 0.05, indicating that the data are not normally distributed. One can also see this in the boxplot where all the outlines are shown (Figure 4). It is important to mentioned that the researcher will not be normalizing the data in this specific case because the outside data are also helpful for the analysis and because the statistical test that will be performed (one-sample median tests) does not required normalize data.

### 4.2.3 Clean and Validate Data

Cleaning up the data can be the most time-consuming part of the data preparation process, but it is definitely crucial to removing faulty data and filling gaps.

The first step for data clean-up is filtering by family. Of the 21 families found on the data set, only 8 are products (*handle*, *bottle, lid, zeppelin bottle, polyethylene bottle, twister bottle, label,* and *vial*) and the researcher will only focus in 7 (*bottle, lid, zeppelin bottle, polyethylene bottle, twister bottle, label,* and *vial*). It is important to mention that the researcher is not working with the handle family because this will cease to exist from

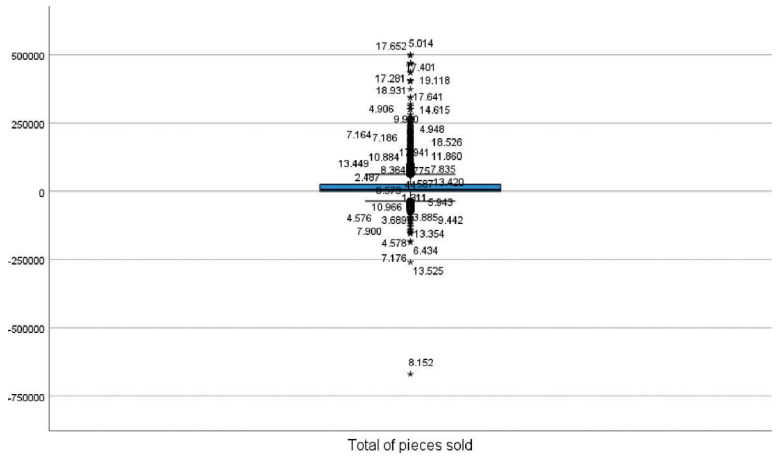



**Figure 4.** Normality boxplot.

the year 2022. With this, the researcher ends up with a data set of 18,268 entries (Figure 5).

Finally, when the researcher found that the data set has more product IDs than product description, they need to make sure that they are working with the correct information. However, since the researcher knew that some product IDs (SKU) have change in the past years, they will change the old SKU with the new one so that the analysis considers this part numbers (Table 5).

Once the researcher has made these changes in data set, they then clean the entries with "0" on the Unite Price variable (Figure 6).

| Región | Clien | NombreZor | IdCliente | NombreCliente | IdProducto | DescProducto | Total Unidad | UM | Familia | FechaFactu | LineaLlave | TransT | TransDc | PrecioUnita | Venta | NUMFAC | TIPOD |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| - | - | | 12372 | COSMOCEL, S.A. | BTWS150 | BOT TWISTAR 2 | 25,200 | PZ | BOTELLA | 1-Feb-19 | 00002-SO-21 | SO | 21941 | $ 2.80 | $ 70,560 | 20630 | EF |
| - | - | | 12372 | COSMOCEL, S.A. | T45020 | TAPA 45 BLANC | 19,500 | PZ | TAPA | 1-Feb-19 | 00002-SO-21 | SO | 21941 | $ 1.71 | $ 33,345 | 20630 | EF |
| - | - | | 12372 | COSMOCEL, S.A. | 2120G020 | BOT ZEPELIN 50 | 1,280 | PZ | GARRAFA | 1-Feb-19 | 00002-SO-21 | SO | 21941 | $ 10.29 | $ 13,171 | 20630 | EF |
| NACIONAL | LOCAL | | 65562 | GOWAN MEXICA | BTWS250 | BOT TWISTAR S | 13,338 | PZ | BOTELLA | 1-Feb-19 | 00002-SO-21 | SO | 21942 | $ 3.38 | $ 45,082 | 20631 | EF |
| NACIONAL | LOCAL | | 65562 | GOWAN MEXICA | T45180 | TAPA 45 AZUL C | 13,338 | PZ | TAPA | 1-Feb-19 | 00002-SO-21 | SO | 21942 | $ 1.05 | $ 14,005 | 20631 | EF |
| NACIONAL | LOCAL | | 67265 | AGROQUIMICOS | BTWS040 | BOT TWISTAR 1 | 13,728 | PZ | BOTELLA | 1-Mar-19 | 00002-SO-21 | SO | 21943 | $ 5.59 | $ 76,740 | 20632 | EF |
| NACIONAL | LOCAL | | 67265 | AGROQUIMICOS | T63020 | TAPA 63 BLANC | 13,728 | PZ | TAPA | 1-Mar-19 | 00002-SO-21 | SO | 21943 | $ 2.16 | $ 29,652 | 20632 | EF |
| NACIONAL | LOCAL | | 65570 | SINTESIS Y FORM | EMS042 | BOT MISIL 56 GI | - | PZ | BOTELLA | 1-Jul-19 | 00002-CD-31 | CD | 316 | $ 2.83 | -$ 30,046 | 2040 | CE |
| NACIONAL | LOCAL | | 65570 | SINTESIS Y FORM | EMS043 | BOT MISIL 56 G | - | PZ | BOTELLA | 1-Jul-19 | 00002-CD-31 | CD | 316 | $ 3.02 | -$ 36,747 | 2040 | CE |
| NACIONAL | LOCAL | | 65570 | SINTESIS Y FORM | K-PAX-M-001 | EMPECO 10 LTS | - | PZ | GARRAFA | 1-Jul-19 | 00002-CD-31 | CD | 316 | $ 22.94 | -$ 84,098 | 2040 | CE |
| NACIONAL | LOCAL | | 65570 | SINTESIS Y FORM | T45020 | TAPA 45 BLANC | - | PZ | TAPA | 1-Jul-19 | 00002-CD-31 | CD | 316 | $ 1.17 | -$ 27,632 | 2040 | CE |
| NACIONAL | LOCAL | | 65570 | SINTESIS Y FORM | T63030 | TAPA 63 BLAN | - | PZ | TAPA | 1-Jul-19 | 00002-CD-31 | CD | 316 | $ 2.29 | -$ 6,078 | 2040 | CE |
| - | - | | 3653 | GRUPO BIOQUIMI | BLGS030 | BOT LIGHT STAI | 15,896 | PZ | BOTELLA | 1-Jul-19 | 00002-SO-21 | SO | 21945 | $ 3.13 | $ 49,754 | 20636 | EF |
| - | - | | 3653 | GRUPO BIOQUIMI | BLGS030 | BOT LIGHT STAI | 5,400 | PZ | BOTELLA | 1-Jul-19 | 00002-SO-21 | SO | 21946 | $ 3.13 | $ 16,902 | 20637 | EF |
| - | - | | 3653 | GRUPO BIOQUIMI | T45111 | TAPA 45 VERDE | 15,700 | PZ | TAPA | 1-Jul-19 | 00002-SO-21 | SO | 21947 | $ 0.97 | $ 15,229 | 20638 | EF |
| - | - | | 3653 | GRUPO BIOQUIMI | T45111 | TAPA 45 VERDE | 6,400 | PZ | TAPA | 1-Jul-19 | 00002-SO-21 | SO | 21948 | $ 0.97 | $ 6,208 | 20639 | EF |
| NACIONAL | LOCAL | | 65570 | SINTESIS Y FORM | EMS042 | BOT MISIL 56 GI | 10,617 | PZ | BOTELLA | 1-Jul-19 | 00002-SO-21 | SO | 21949 | $ 2.83 | $ 30,046 | 20640 | EF |

**Figure 5.** Data filter by family.

**Table 5.** Duplicate products.

| *IdProducto* | *DescProducto* | *Family* | *Comment* |
|---|---|---|---|
| *BLGS060* | BOT LIGHT STAR 1LT ALUMINIO | Bottle | Old |
| *BLGS061* | BOT LIGHT STAR 1LT ALUMINIO | Bottle | New |
| *2091087* | CONCHA 221 ML | Bottle | Old |
| *2091064* | CONCHA 221 ML | Bottle | New |
| *T45180* | TAPA 45 AZUL C/LINNER IND | Lid | New |
| *T45230* | TAPA 45 AZUL C/LINNER IND | Lid | Old |
| *T45180AC* | TAPA 45 AZUL C/LINNER IND | Lid | Old |
| *T45100* | TAPA 45 VERDE C/LINNER IND | Lid | Correct |
| *T45110* | TAPA 45 VERDE C/LINNER IND | Lid | Correct |

| IdProd | DescPr | Total U | U | Familia | Ingr. Ac | FechaFactura | LineaLl | TransTi | TransDc | PrecioU | Venta | NUMFA | TIPODC |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| T63100 | TAPA 63 V | 96,750 | PZ | TAPA | Sin I.A. | 5/27/2019 | 00002-SO- | SO | 22782 | 0.0000 | 2 | 21455 | EF |
| T63100 | TAPA 63 V | -96,750 | PZ | TAPA | Sin I.A. | 5/30/2019 | 00002-CA- | CA | 1324 | 0.0000 | -2 | 2135 | CE |
| ETGE0025 | ETIQUETA | 8,948 | PZ | ETIQUETA | Sin I.A. | 1/23/2020 | 00002-SO- | SO | 24318 | 0.0000 | 0 | 22925 | EF |
| ETGE0100 | ETIQUETA | 182,513 | PZ | ETIQUETA | Sin I.A. | 1/23/2020 | 00002-SO- | SO | 24318 | 0.0000 | 1 | 22925 | EF |
| ETGE0100 | ETIQUETA | 19,670 | PZ | ETIQUETA | Sin I.A. | 1/23/2020 | 00002-SO- | SO | 24318 | 0.0000 | 1 | 22925 | EF |
| ETGE0025 | ETIQUETA | -8,948 | PZ | ETIQUETA | Sin I.A. | 1/29/2020 | 00002-CA- | CA | 1503 | 0.0000 | 0 | 2382 | CE |
| ETGE0100 | ETIQUETA | -19,670 | PZ | ETIQUETA | Sin I.A. | 1/29/2020 | 00002-CA- | CA | 1503 | 0.0000 | -1 | 2382 | CE |
| ETGE0100 | ETIQUETA | -182,513 | PZ | ETIQUETA | Sin I.A. | 1/29/2020 | 00002-CA- | CA | 1503 | 0.0000 | -1 | 2382 | CE |

**Figure 6.** Unite price "0".

Finally, the researcher clears out variables that do not add value to analysis. In this case, the researcher only has one, "Ingr.Activo." This variable is the same for all the entries and does not add value to the analysis.

### 4.2.4 Transform and Enrich Data

This is the process of updating format or value entries to make the data more easily understood by a wider audience. It also refers to the action of adding or connecting data to provide deeper insights.

The researcher makes sure that everything has the same format, especially the variables that they are working on, such as date, number of pieces, unit price, and total sale. Furthermore, another variable will be added, which the researcher named "Top 10," with "1" indicating that product is part of the top 10 sold products and "0" that it is not (Figure 7).

### 4.2.5 Data Analysis

The cleaning process of the data has left the researcher with 18,268 entries and 17 variables. This is because as explained during the data preparation

| Región | Clien | NombreZor | IdCliente | NombreCliente | IdProducto | DescProducto | Total Unidad | UM | Familia | FechaFactu | LineaLlave | Trans1 | TransDc | PrecioUnita | Venta | NUMFAC | TIPODOC | Top 10 |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| - | - | | 12372 | COSMOCEL, S.A. | BTWS150 | BOT TWISTAR 2 | 25,200 | PZ | BOTELLA | 1-Feb-19 | 00002-SO-215 | SO | 21941 | $ 2.80 | $ 70,560 | 20630 | EF | 1 |
| - | - | | 12372 | COSMOCEL, S.A. | T45020 | TAPA 45 BLANC | 19,500 | PZ | TAPA | 1-Feb-19 | 00002-SO-215 | SO | 21941 | $ 1.71 | $ 33,345 | 20630 | EF | 1 |
| - | - | | 12372 | COSMOCEL, S.A. | Z120G020 | BOT ZEPELIN 50 | 1,280 | PZ | GARRAFA | 1-Feb-19 | 00002-SO-215 | SO | 21941 | $ 10.29 | $ 13,171 | 20630 | EF | 0 |
| NACIONAL | LOCAL | | 65562 | GOWAN MEXICAI | BTWS250 | BOT TWISTAR 5 | 13,338 | PZ | BOTELLA | 1-Feb-19 | 00002-SO-215 | SO | 21942 | $ 3.38 | $ 45,082 | 20631 | EF | 0 |
| NACIONAL | LOCAL | | 65562 | GOWAN MEXICAI | T45180 | TAPA 45 AZUL C | 13,338 | PZ | TAPA | 1-Feb-19 | 00002-SO-215 | SO | 21942 | $ 1.05 | $ 14,005 | 20631 | EF | 1 |
| NACIONAL | LOCAL | | 67265 | AGROQUIMICOS ( | BTWS040 | BOT TWISTAR 1 | 13,728 | PZ | BOTELLA | 1-Mar-19 | 00002-SO-215 | SO | 21943 | $ 5.59 | $ 76,740 | 20632 | EF | 1 |
| NACIONAL | LOCAL | | 67265 | AGROQUIMICOS ( | T63020 | TAPA 63 BLANC | 13,728 | PZ | TAPA | 1-Mar-19 | 00002-SO-215 | SO | 21943 | $ 2.16 | $ 29,652 | 20632 | EF | 0 |
| NACIONAL | LOCAL | | 65570 | SINTESIS Y FORMI | EMS042 | BOT MISIL 56 GI | - | PZ | BOTELLA | 1-Jul-19 | 00002-CD-314 | CD | 316 | $ 2.83 | -$ 30,046 | 2040 | CE | 1 |
| NACIONAL | LOCAL | | 65570 | SINTESIS Y FORMI | EMS043 | BOT MISIL 56 G | - | PZ | BOTELLA | 1-Jul-19 | 00002-CD-314 | CD | 316 | $ 3.02 | -$ 36,747 | 2040 | CE | 1 |
| NACIONAL | LOCAL | | 65570 | SINTESIS Y FORMI | K-PAX-M-001 | EMPECO 10 LTS | - | PZ | GARRAFA | 1-Jul-19 | 00002-CD-314 | CD | 316 | $ 22.94 | -$ 84,098 | 2040 | CE | 0 |
| NACIONAL | LOCAL | | 65570 | SINTESIS Y FORMI | T45020 | TAPA 45 BLANC | - | PZ | TAPA | 1-Jul-19 | 00002-CD-314 | CD | 316 | $ 1.17 | -$ 27,632 | 2040 | CE | 1 |
| NACIONAL | LOCAL | | 65570 | SINTESIS Y FORMI | T63030 | TAPA 63 BLAN | - | PZ | TAPA | 1-Jul-19 | 00002-CD-314 | CD | 316 | $ 2.29 | -$ 6,078 | 2040 | CE | 0 |
| - | - | | 3653 | GRUPO BIOQUIMI | BLGS030 | BOT LIGHT STAI | 15,896 | PZ | BOTELLA | 1-Jul-19 | 00002-SO-215 | SO | 21945 | $ 3.13 | $ 49,754 | 20636 | EF | 0 |
| - | - | | 3653 | GRUPO BIOQUIMI | BLGS030 | BOT LIGHT STAI | 5,400 | PZ | BOTELLA | 1-Jul-19 | 00002-SO-215 | SO | 21946 | $ 3.13 | $ 16,902 | 20637 | EF | 0 |
| - | - | | 3653 | GRUPO BIOQUIMI | T45111 | TAPA 45 VERDE | 15,700 | PZ | TAPA | 1-Jul-19 | 00002-SO-215 | SO | 21947 | $ 0.97 | $ 15,229 | 20638 | EF | 0 |

**Figure 7.** Data with format and top 10 variable.

**Table 6.** Products per family.

| ***Family*** | ***Products*** |
|---|---|
| *Bottle* | 108 |
| *Label* | 11 |
| *Lid* | 32 |
| *Polyethylene bottle* | 9 |
| *Twister bottle* | 5 |
| *Vial* | 2 |
| *Zeppelin bottle* | 1 |

process, the researcher is only going to work with 7 out of 21 families. Form the total of 196 product IDs, only 168 are products under the 7 families that the researcher will be working with; the distribution of the products by family is shown in Table 6.

### 4.2.6 Products with the Highest Sales in the Past 3.5 Years

Out of the 7 selected families, the "Bottle" family is the one with the highest sales in the past 3.5 years, representing 50% of the sales (Figure 8).

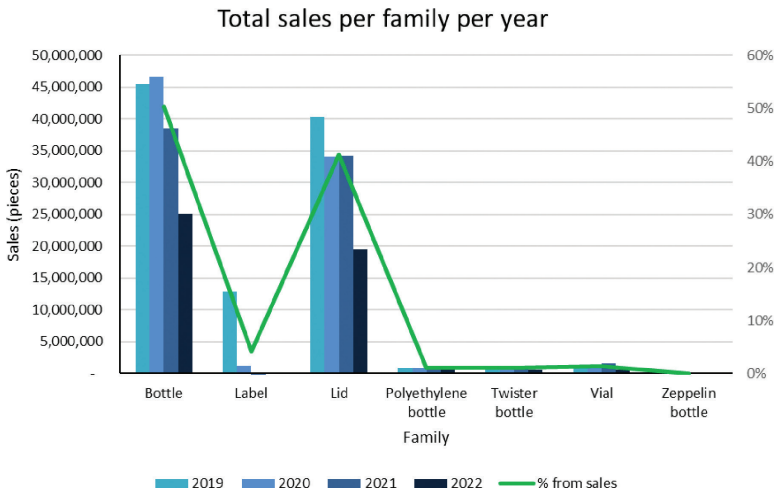



**Figure 8.** Total sales per family per year.

**Table 7.** Top 10 products.

| *Product ID* | *Product description* | *Family* |
|---|---|---|
| *T45020* | Lid 45 White Linner Ind Ing | Lid |
| *EMS042* | Bot Misil 56 Grs Blanco | Bottle |
| *BZE060* | Bot Zeneca 7 Cost 1L Blanco | Bottle |
| *BTWS150* | Bot Twistar 250 Ml Blanca | Bottle |
| *BZE040* | Bot Zeneca 7 Cost 1L Clear | Bottle |
| *ETGE01000* | Generic Label Litter | Label |
| *T45180* | Tapa 45 Azul C/Linner Ind | Lid |
| *EMS041* | Bot Empecomisil 1Lt Blanco | Bottle |
| *EMS043* | Bot Misil 56 G Cristal | Bottle |
| *BTWS040* | Bot Twistar 1Lt Blanco | Bottle |

Let us remember that the data for 2022 only come from January to July, which means this is not the total sales of the year.

Regarding the list of products that the company has under the 7 selected families, there are a total of 20 products that have been in the top 10 sales. However, the researcher focuses on the summary of the past 3.5 years. These top 10 products are shown in Table 7. The total sales for these 10 products are shown in Figure 8 and the detail per year in Figure 9. This information can also be found in Figure 7, where the researcher added an additional variable to indicate if the part number is part of the top 10 products.

Figure 9 shows the products with the top 10 sales during the past 3.5 years. There are products from 3 families, and one can see that there is a product whose sales have been considerably higher than those of the rest of the products. This product is T45020 from the *lid* family. Knowing how

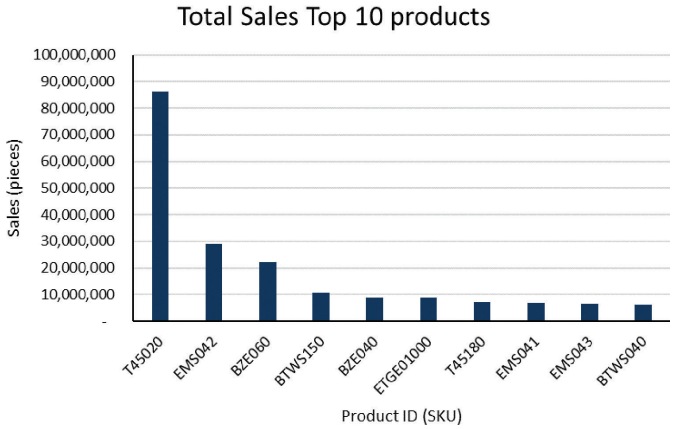



**Figure 9.** Top sales from the top 10 products.

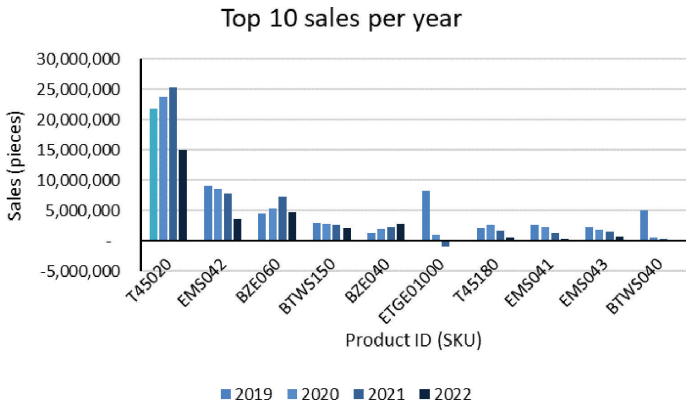



**Figure 10.** Top 10 sales per year.

this product is sold makes sense for it to be the number one in sales. One *lid* fit more than one *bottle*, which means that this specific *lid* is sold for more than one *bottle*.

Figure 10 shows the sales for the top 10 products. But in this case, one can see the detail per year for each of this part numbers. As mentioned before, the *lid* T45020 has the highest sales every year and also a continued growth year after year. In general, there are three part numbers with growth during the past 3 years (2019–2022): the *lid* T45020, the *bottle* BZE060, and the *bottle* BZE040. There are two part numbers with significant drops in sales in the same period: the *label* ETGE01000 and *bottle* BTWS040. The rest of the product had had some fluctuations.

### 4.2.7 Testing Significance of Top 10 Products

It is important for the researcher, and especially for the company, to understand if the sales of these top 10 products have an impact on the total sales of the company. As a result, the researcher performed a statistical analysis to test the significance of the top 10 products for the total sales.

For the statistical test, the researcher performed a one-sample median test, in order to compare the median of samples with hypothesized median values. For this test, compared with the one-sample t-test, the researcher does not assume that it is interval and normally distributed, instead an ordinal variable.

The researched has chosen this test because as can be seen in the data preparation process, the data are not normally distributed and that the outliners have not been removed to normalize them. In addition, one of the variables is ordinal.

The researcher has chosen SPSS as an analysis tool. The steps followed to perform the analysis are as follows: open the final data set in SPSS (Figure 11) and ensure that the variables are correctly defined (Figure 12).

| RegiónCliente | NombreZonaVentaClien | IdCliente | NombreCliente | IdProducto | DescProducto | TotalUnidades | UM | Familia | FechaFactura | LineaLlave | TransTipoDo | TransDoc | PrecioUnitario | Venta | NUMFAC | TIPOOC | Top10 |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| - | - | 12372 | COSMOCEL, S.A. | BTWS150 | BOT TWISTAR ... | 25200 | PZ | BOTELLA | 02-Jan-19 | 00002-S... | SO | 21941 | $2.80 | $70,560.00 | 20630 | EF | 1 |
| - | - | 12372 | COSMOCEL, S.A. | T45020 | TAPA 45 BLAN... | 19500 | PZ | TAPA | 02-Jan-19 | 00002-S... | SO | 21941 | $1.71 | $33,345.00 | 20630 | EF | 1 |
| - | - | 12372 | COSMOCEL, S.A. | Z120G020 | BOT ZEPELIN ... | 1280 | PZ | GARRAFA... | 02-Jan-19 | 00002-S... | SO | 21941 | $10.29 | $13,171.20 | 20630 | EF | 0 |
| NACIONAL | LOCAL | 65562 | GOWAN MEXICAN... | BTWS250 | BOT TWISTAR ... | 13338 | PZ | BOTELLA | 02-Jan-19 | 00002-S... | SO | 21942 | $3.38 | $45,082.44 | 20631 | EF | 0 |
| NACIONAL | LOCAL | 65562 | GOWAN MEXICAN... | T45180 | TAPA 45 AZUL ... | 13338 | PZ | TAPA | 02-Jan-19 | 00002-S... | SO | 21942 | $1.05 | $14,004.90 | 20631 | EF | 1 |
| NACIONAL | LOCAL | 67265 | AGROQUIMICOS C... | BTWS040 | BOT TWISTAR ... | 13728 | PZ | BOTELLA | 03-Jan-19 | 00002-S... | SO | 21943 | $5.59 | $76,739.52 | 20632 | EF | 1 |
| NACIONAL | LOCAL | 67265 | AGROQUIMICOS C... | T63020 | TAPA 63 BLAN... | 13728 | PZ | TAPA | 03-Jan-19 | 00002-S... | SO | 21943 | $2.16 | $29,652.48 | 20632 | EF | 0 |
| NACIONAL | LOCAL | 65570 | SINTESIS Y FORM... | EMS042 | BOT MISIL 56 ... | 0 | PZ | BOTELLA | 07-Jan-19 | 00002-C... | CD | 316 | $2.83 | -$30,046.11 | 2040 | CE | 1 |
| NACIONAL | LOCAL | 65570 | SINTESIS Y FORM... | EMS043 | BOT MISIL 56 ... | 0 | PZ | BOTELLA | 07-Jan-19 | 00002-C... | CD | 316 | $3.02 | -$36,747.36 | 2040 | CE | 1 |
| NACIONAL | LOCAL | 65570 | SINTESIS Y FORM... | K-PAX-M-001 | EMPECO 10 L ... | 0 | PZ | GARRAFA... | 07-Jan-19 | 00002-C... | CD | 316 | $22.94 | -$84,098.04 | 2040 | CE | 0 |
| NACIONAL | LOCAL | 65570 | SINTESIS Y FORM... | T45020 | TAPA 45 BLAN... | 0 | PZ | TAPA | 07-Jan-19 | 00002-C... | CD | 316 | $1.17 | -$27,631.89 | 2040 | CE | 1 |
| NACIONAL | LOCAL | 65570 | SINTESIS Y FORM... | T63030 | TAPA 63 BLA... | 0 | PZ | TAPA | 07-Jan-19 | 00002-C... | CD | 316 | $2.29 | -$6,077.66 | 2040 | CE | 0 |
| - | - | 3663 | GRUPO BIOQUIMIC... | BLGS030 | BOT LIGHT ST... | 15896 | PZ | BOTELLA | 07-Jan-19 | 00002-S... | SO | 21945 | $3.13 | $49,754.48 | 20636 | EF | 0 |
| - | - | 3663 | GRUPO BIOQUIMIC... | BLGS030 | BOT LIGHT ST... | 5400 | PZ | BOTELLA | 07-Jan-19 | 00002-S... | SO | 21946 | $3.13 | $16,902.00 | 20637 | EF | 0 |
| - | - | 3663 | GRUPO BIOQUIMIC... | T45111 | TAPA 45 VERD... | 15700 | PZ | TAPA | 07-Jan-19 | 00002-S... | SO | 21947 | $0.97 | $15,229.00 | 20638 | EF | 0 |
| - | - | 3663 | GRUPO BIOQUIMIC... | T45111 | TAPA 45 VERD... | 6400 | PZ | TAPA | 07-Jan-19 | 00002-S... | SO | 21948 | $0.97 | $6,208.00 | 20639 | EF | 0 |

**Figure 11.** Final data in SPSS.

Now that the data are entered in SPSS and the variables are defined, the following steps are followed to do the test (also see Figures 13 and 14):

1. Click the analyse tab.
2. Click nonparametric tests.

| | Name | Type | Width | Decimals | Label | Values | Missing | Columns | Align | Measure | Role |
|---|---|---|---|---|---|---|---|---|---|---|---|
| 1 | RegiónCliente | String | 24 | 0 | Clients region | None | None | 24 | Left | Nominal | Input |
| 2 | NombreZonaVentaCliente | String | 19 | 0 | Sales Region | None | None | 19 | Left | Nominal | Input |
| 3 | IdCliente | Numeric | 7 | 0 | Clients unique number | None | None | 12 | Right | Ordinal | Input |
| 4 | NombreCliente | String | 40 | 0 | Clients Name | None | None | 40 | Left | Nominal | Input |
| 5 | IdProducto | String | 12 | 0 | Products unique number (SKU) | None | None | 12 | Left | Nominal | Input |
| 6 | DescProducto | String | 30 | 0 | Products Description | None | None | 30 | Left | Nominal | Input |
| 7 | TotalUnidades | Numeric | 7 | 0 | Total Pieces Sold | None | None | 12 | Right | Scale | Input |
| 8 | UM | String | 2 | 0 | Unite of Measure | None | None | 2 | Left | Nominal | Input |
| 9 | Familia | String | 19 | 0 | Products Family | None | None | 19 | Left | Nominal | Input |
| 10 | FechaFactura | Date | 11 | 0 | Invoice Date | None | None | 11 | Right | Ordinal | Input |
| 11 | LineaLlave | String | 21 | 0 | Key Line of Invoice | None | None | 21 | Left | Nominal | Input |
| 12 | TransTipoDoc | String | 2 | 0 | Document Type | None | None | 2 | Left | Nominal | Input |
| 13 | TransDoc | Numeric | 6 | 0 | Document Number | None | None | 12 | Right | Nominal | Input |
| 14 | PrecioUnitario | Dollar | 21 | 2 | Unite Price | None | None | 16 | Right | Scale | Input |
| 15 | Venta | Dollar | 19 | 2 | Total Sale | None | None | 16 | Right | Scale | Input |
| 16 | NUMFAC | Numeric | 6 | 0 | Invoice Number | None | None | 12 | Right | Nominal | Input |
| 17 | TIPODOCFAC | String | 2 | 0 | Invoice Type | None | None | 2 | Left | Nominal | Input |
| 18 | Top10 | Numeric | 1 | 0 | Top 10 category | None | None | 12 | Right | Nominal | Input |

**Figure 12.** SPSS variables view.

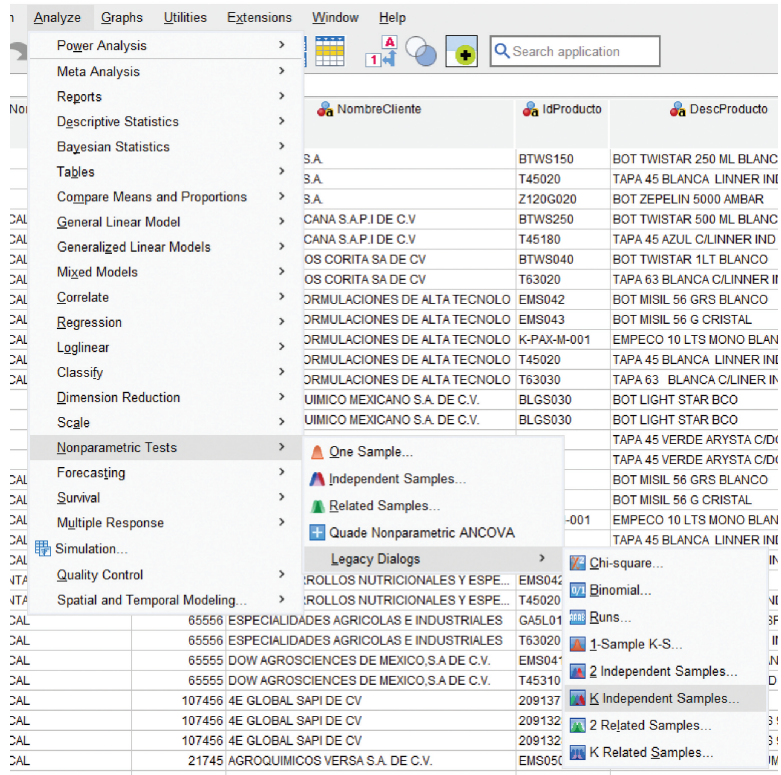



**Figure 13.** Performing test, part 1.

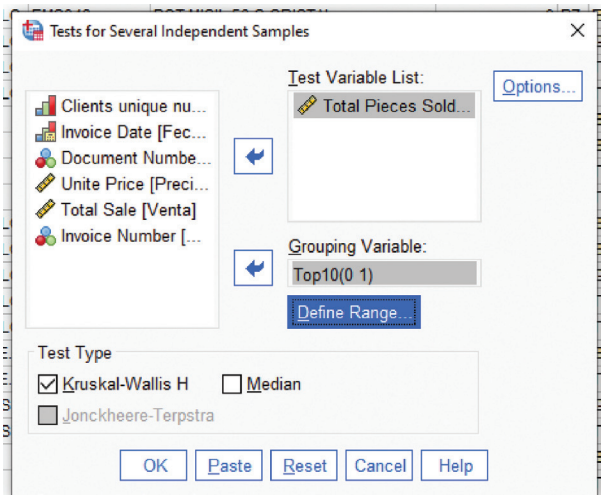



**Figure 14.** Performing test, part 2.

3. Click Legacy Dialogs.
4. Click K independent samples.
5. Drag the "Total of pieces sold" variable into the box labelled Test Variable List.
6. Drag "Top 10" into the box labelled Grouping Variable.
7. Click Define Range.
8. Set the minimum value to "0" and the maximum value to "1."
9. Click OK.
10. Ensure the box is checked next to Kruskal–Wallis H.
11. Finally, click OK.

Now is time to discuss the results of the test. SPSS shows two tables as part of this analysis: *Ranks* and *Test Statistics* (Figure 15).

The second table (Test Statistics) displays the results of the test:

- Kruskal–Wallis H: This is the chi-square ($X^2$) test statistic that measures the difference between observed and expected frequency of outcomes of a set of events or variables and is useful for analysing such differences in categorical variables.
- df: This is the degrees of freedom. A data sample is defined as having a maximum number of logically independent values, meaning they are values that can vary freely and is calculated as #groups–1 = 2–1 = 1.
- Asymp. Sig: This is the P-value associated with a $X^2$ test statistic of 645.659 with 1 degrees of freedom.

Finally, the researcher set a hypothesis and the test performed before answers if the hypothesis is true or not. The set hypothesis are as follows:

Null hypothesis: The top 10 products are not significant for the total sales per year.

Alternative hypothesis: The top 10 products are significant for the total sales per year.

Since the P-value (0.000) is less than 0.05, the researcher rejects the null hypothesis and therefore the researcher goes with the alternative hypothesis.

**Ranks**

| Top 10 category | | N | Mean Rank |
|---|---|---|---|
| Total of pieces sold | 0 | 10,400 | 8,277.65 |
| | 1 | 7,868 | 10,267.09 |
| | Total | 18,268 | |

**Test statistics** [a,b]

| | Total of pieces sold |
|---|---|
| Kruskal-Wallis H | 645.659 |
| df | 1 |
| Asymp. Sig. | 0.000 |

a. Kruskal Wallis Test

b. Group variable: Top 10

**Figure 15.** Test results.


**Figure 16.** Total pieces sold according to the top 10 category per year.

This means that the company should be aware of this part number because they are significant for the total pieces sold. Figure 16 shows the representation of the total pieces sold for the top 10 products (YES category) vs the products that are not in the top 10 (NO category). As shown in figure, every year from 2019 to 2021, the top 10 products have been more that the rest of the products, which means that the top 10 products represent more than 50% of the total pieces sold per year. Even though 2022 does not show the same information, it is likely that the changes can be seen in the next 5 months.

### 4.2.8 Demand Fluctuation Analysis

It is important that one understands the behaviour and fluctuations of the top 10 products during the past 3.5 years.

As mentioned before, there are three part numbers with an increase from year to year: T45020 (*Lid*), BZE060 (*Bottle*), and BZE040 (*Bottle*), and two part numbers that decrease: ETGE01000 (*Label*) and BTWS040 (*Bottle*).

Figure 17 shows the comparison of the sales of each year by trimester. It seems that *Bottles* sales have been steady, *Label* suffered a fall in 2021 which is recovering in 2022, and *Lid*s is the family with the highest fluctuations; therefore, the researcher is going to do a deeper analysis of sales by family to have a better understanding of this.

In Figure 18, one can see the sales for B*ottles* per quarter and year and that, on average, there is a drop in sales quarter after quarter. The first and

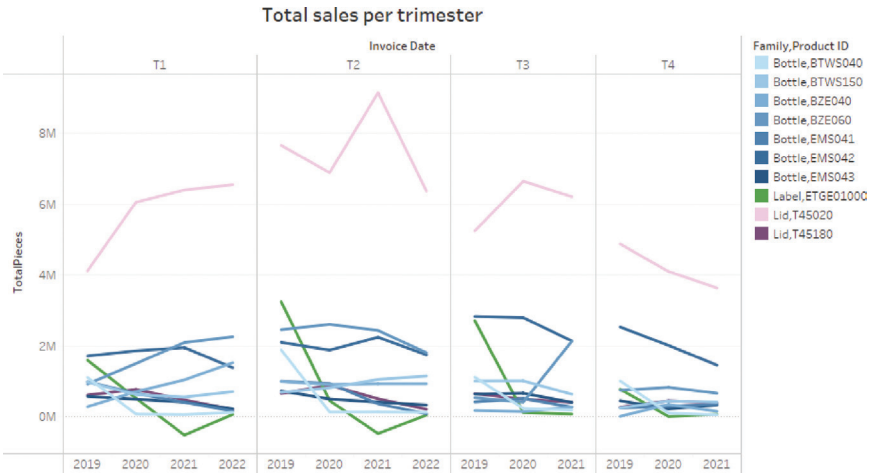



**Figure 17.** Top 10 total sales per trimester.

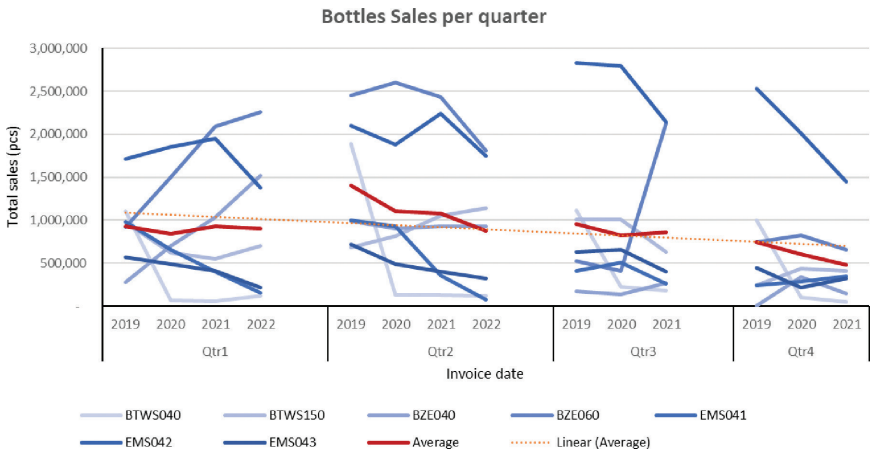



**Figure 18.** Bottles sales per quarter.

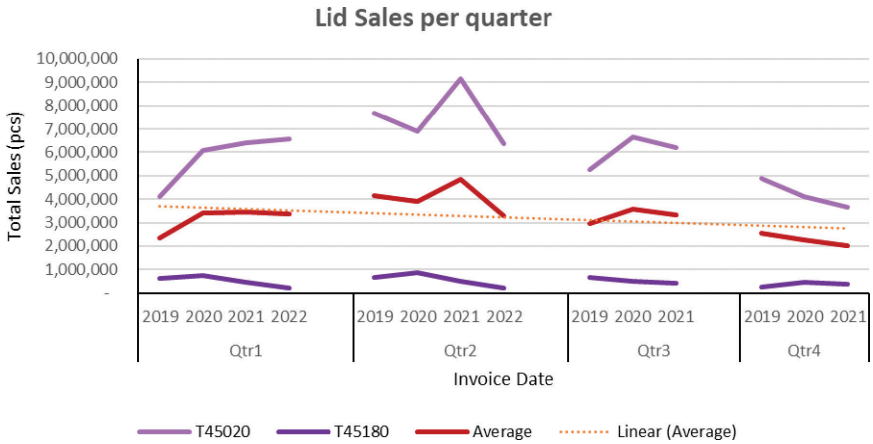



**Figure 19.** Lid sales per quarter.

third quarters have been steady for the past years, while the second and fourth quarters have suffered a year-over-year drop.

Figure 19 shows the summary of the sales for the *lids* and *bottles*, with a drop for lids quarter after quarter and having the second quarter of each year the best in sales. Now, in the case of the fluctuations of each year by quarter, one can see that there was an increase for the first quarter from 2019 to 2020 and stays steady for the next years. The second quarter is the one with the highest fluctuations, being mostly steady from 2019 to 2020, and then having an important increase in 2021 and an even bigger decrease for 2022. Quarter 3 had a little increase in 2020 and then a little for in 2021; these changes are not as significant and those in quarter 2. Finally, quarter 4 has had a steady drop from 2019 to 2021.

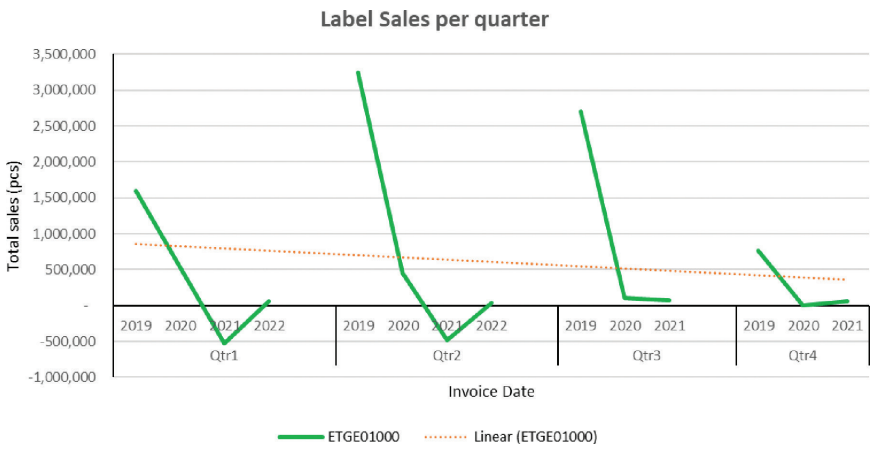



**Figure 20.** Label sales per quarter.

Figure 20 shows the sales for the *Label*. This item is the one with the highest drop having negative numbers for the first two quarters of 2021 and having almost no sales for the rest of 2021 and the first two quarters of 2022. The negative sales are due to credit notes and will need to be review separately. Label also has a continue decrease from the first quarter to the last quarter of the year.

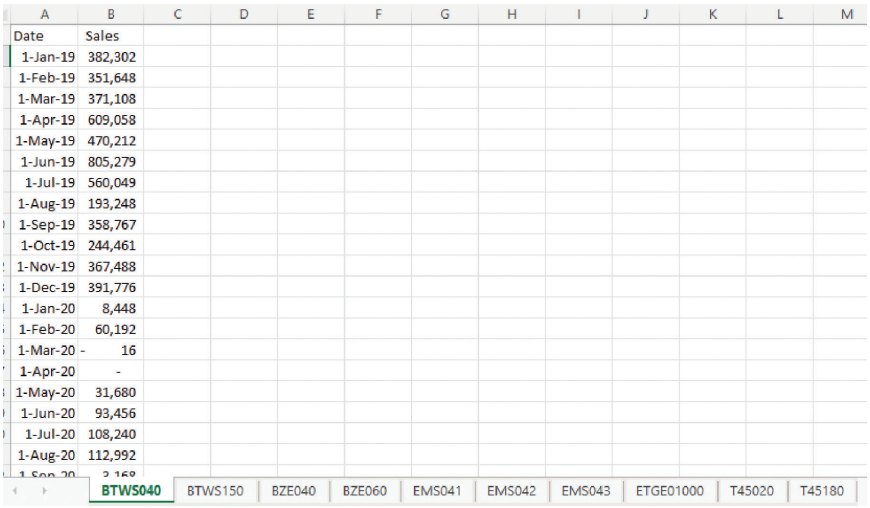
With this information, one can understand that the company has the best sales during the second quarter of the year and the lowest sales during the last quarter.

### 4.2.9 Forecasting

Once the researcher has identified the top 10 products and found the significance of these products for the total sales and analysed the sales fluctuations, they will forecast the sales of these products for 2023.

Because the researcher only has part of the sales for 2022, they will create a forecast for both years (2022 and 2023). Furthermore, they will also graph the actual sales for the first 7 months of 2022 to confirm if the forecast is accurate.

As mentioned in a previous chapter, the researcher will work with excels forecasting tool, called "Forecast sheet." For the researcher to be able to use this tool properly, it is important that the researcher creates a new sheet per SKU and with the resume per month and year (see Figure 21); as one can see, the researcher has 10 sheets (one per SKU) and each sheet has 2 columns, date and sales.



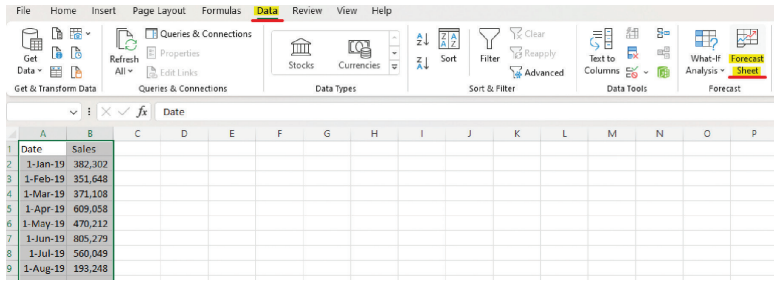
| Date | Sales |
|---|---|
| 1-Jan-19 | 382,302 |
| 1-Feb-19 | 351,648 |
| 1-Mar-19 | 371,108 |
| 1-Apr-19 | 609,058 |
| 1-May-19 | 470,212 |
| 1-Jun-19 | 805,279 |
| 1-Jul-19 | 560,049 |
| 1-Aug-19 | 193,248 |
| 1-Sep-19 | 358,767 |
| 1-Oct-19 | 244,461 |
| 1-Nov-19 | 367,488 |
| 1-Dec-19 | 391,776 |
| 1-Jan-20 | 8,448 |
| 1-Feb-20 | 60,192 |
| 1-Mar-20 | 16 |
| 1-Apr-20 | - |
| 1-May-20 | 31,680 |
| 1-Jun-20 | 93,456 |
| 1-Jul-20 | 108,240 |
| 1-Aug-20 | 112,992 |



**Figure 21.** Sales data per SKU worksheets.

Once the researcher has sales data divided per SKU, they can proceed with the forecast, with the following the steps (Figures 22 and 23):

1. Select the data.
2. Go to "Data" tab.
3. Click "Forecast Sheet".
4. In the Forecast End box, pick an end date; in this case, it will be December 31, 2023.
5. Click Options.
6. In the Forecast Start box, pick a start date; in this case, this will be January 1, 2022.
7. In the Seasonality section, one clicks on Set Manually and select "12" (this represents the 12 months of the year).
8. Click on Create.





| Date | Sales |
|---|---|
| 1-Jan-19 | 382,302 |
| 1-Feb-19 | 351,648 |
| 1-Mar-19 | 371,108 |
| 1-Apr-19 | 609,058 |
| 1-May-19 | 470,212 |
| 1-Jun-19 | 805,279 |
| 1-Jul-19 | 560,049 |
| 1-Aug-19 | 193,248 |

**Figure 22.** Forecast sheet, steps 1–3.

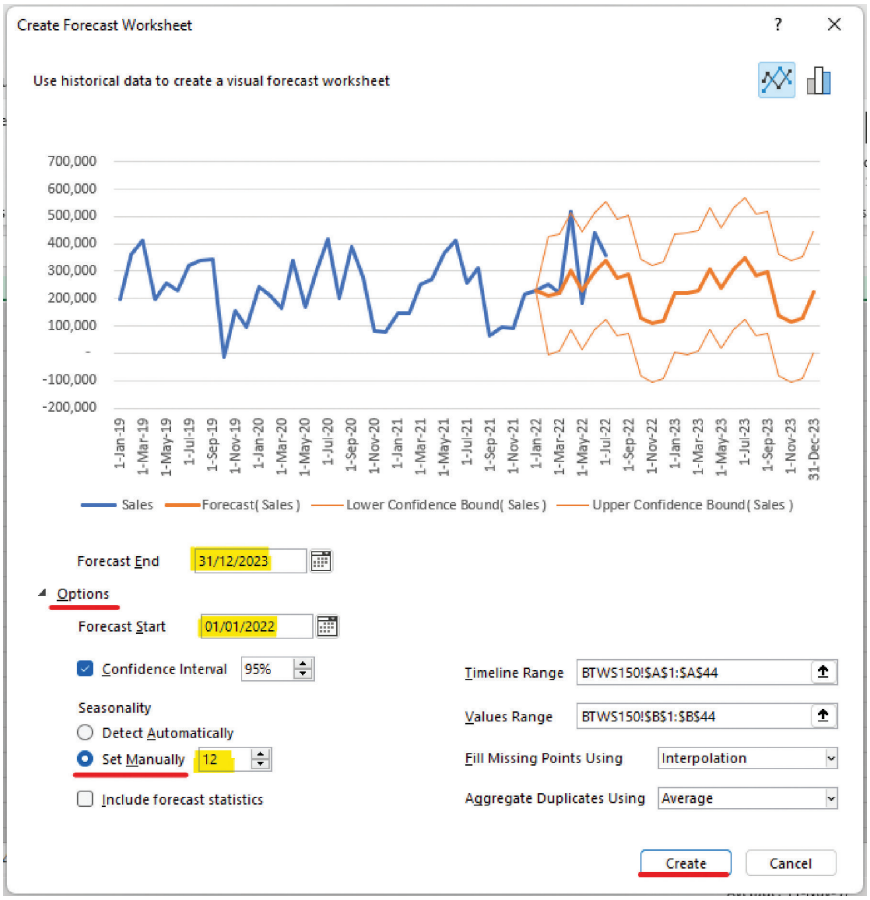



**Figure 23.** Forecast sheet, steps 4–8.

With this, excel will create a new worksheet that contains a table and a chart with the information of the historical values and predicted values (Figure 24). The table contains five columns: the first two correspond to the historical data that one provided and the last three correspond to the predicted values where "Forecast" shows the estimated prediction and "Lower Confidence Bound" and "Upper Confidence Bound" show the interval range where future are expected to fall. The researcher will repeat this process for the 10 products the researcher is analysing.

Figure 25 shows the forecast for the part number T45020 from the *Lid* family. The blue line represents the actual pieces sold, the thick orange line represents the predicted forecast, and the confidence bounds where 95% of the sales are expected to fall. Here one can see that almost all the sales for the first semester of 2022 fall inside the confidence bounds, apart

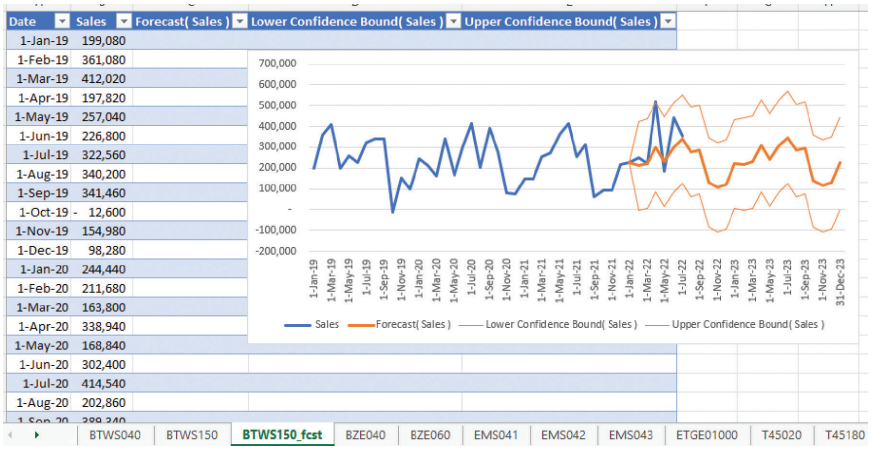



**Figure 24.** Forecast results (table and chart).

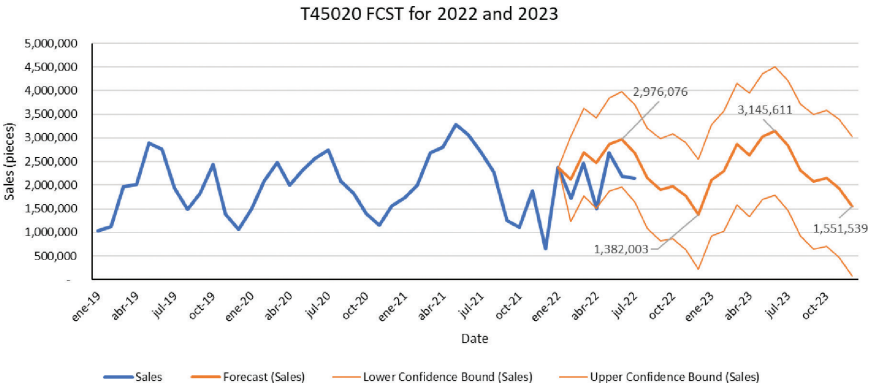



**Figure 25.** T45020 forecast for 2022 and 2023.

from April, which falls 9,774 pieces under the confidence bound, and 3 months are very close to the forecast (February, March, and May). One can also see that the highest pick for 2023 is 3.1 million pieces at the end of the second quarter (June 2023), while the lowest sale is predicted to be during the last quarter, specifically in December 2023. This is persistent with the sales fluctuation analysis the researcher performed previously (see Figure 19).

Figure 26 shows the forecast for the *bottle* EMS042, which has had a fall since 2019. Here all the sales for the first semester of 2022 fall into the confidence bounds and there are 3 months that are very close to the forecast (February, March, and June). For the 2023 forecast, the highest sale will be in September 2023 (third quarter) with 783 thousand pieces and the lowest sale would during December 2023.

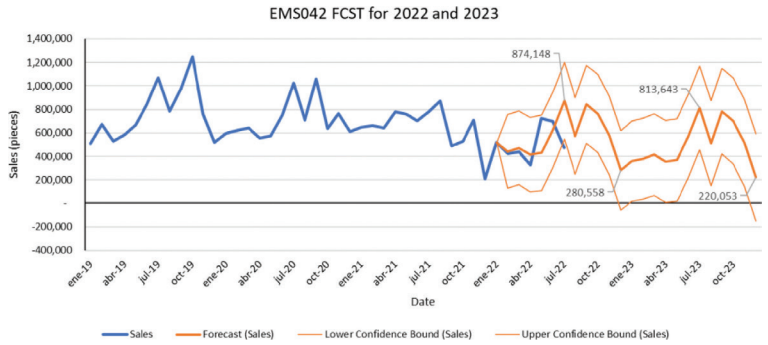



**Figure 26.** EMS042 forecast for 2022 and 2023.

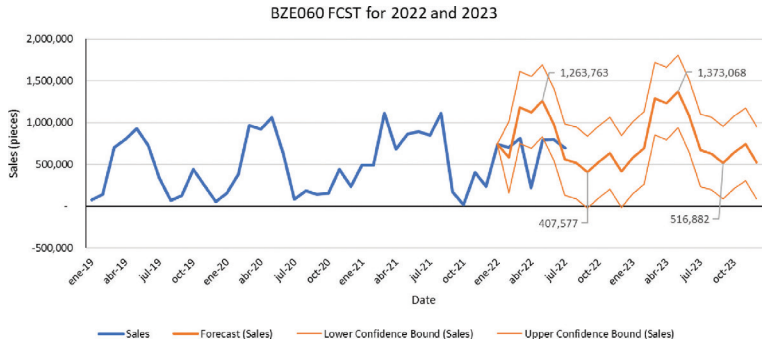



**Figure 27.** BZE060 forecast for 2022 and 2023.

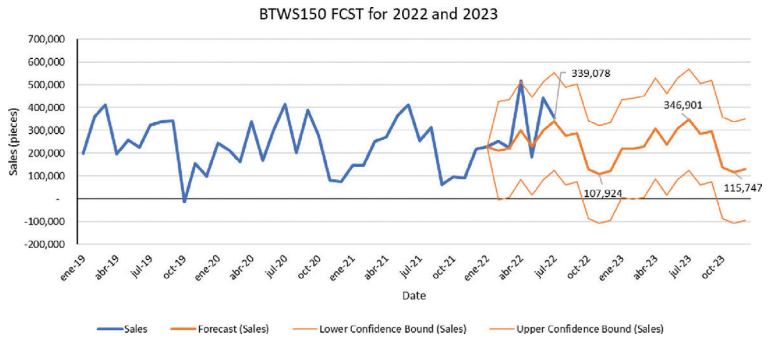



**Figure 28.** BTWS150 forecast for 2022 and 2023.

In Figure 27, one can observe the forecast for the *bottle* BZE060, which has been increasing each year since 2019. For this part number, one can see that there is 1 month sale for 2022 way under the forecast and that the sales of the first semester of 2022 have not been as expected. For 2023, the highest sale is expected to be during May 2023 around 1.3 million pieces (second quarter) and the lowest in September.

Figure 28 shows the forecast for the part number BTWS150 from the *bottle* family. In this case, there is one point falling out of the confidence bounds – April 2022 – where the actual sales were 517 thousand and the upper confidence bound is 515 thousand and there is 1 month pretty close to the forecast – March 2022 – where the actual sales were 222 thousand and the researched had predicted sales of 221.2 thousand. For 2023, the highest expected sale will be at the beginning of the third quarter (July 2023) with 346.9 thousand pieces, and for the lowest sale, this is expected to be on the last quarter of the year, during November 2023.

Figure 29 shows the sales forecast for the *bottle* BZE040. In this case, there is once again 1 month actual sales falling out of the confidence bounds – May 2022 – where the lower confidence bound was 167 thousand and the sales were of only 5,214 pieces. For 2023, the company expects to have the highest sale during the second quarter, in May 2023 with 489 thousand pieces and the lowest sale on August 2023.

Figure 30 shows the forecast for the *label* ETGE01000; Due to a big level of rejection and customers return, the forecast predicted is basically

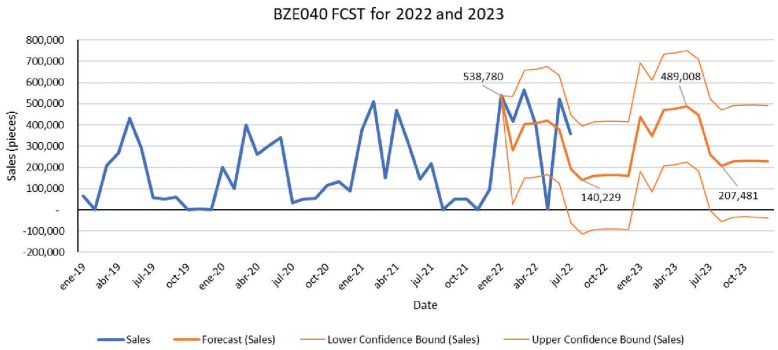



**Figure 29.** BZE040 forecast for 2022 and 2023.

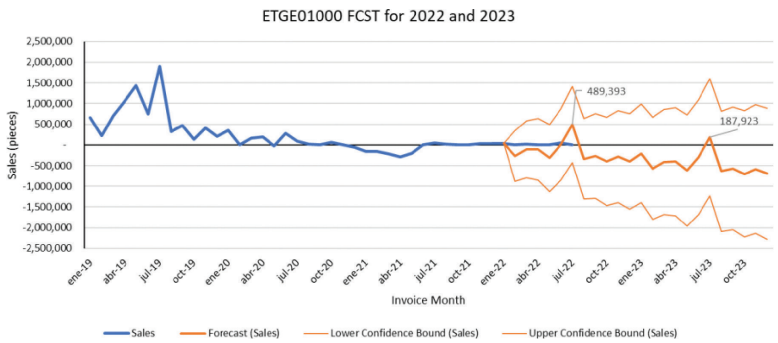



**Figure 30.** ETGE01000 forecast for 2022 and 2023.

negative, having only one positive month for 2023, being this July. Now, the forecast for 2022 showed something similar; however, 4 out of 7 months have had positive sales on 2022 and the other 3 months are not in negative sales. This means that it would be necessary to update this forecast continuously until the rejection rate is back to normal.

Figure 31 shows the forecast for the part number T45180 from the *Lid* family. For the information of 2022, there are 2 months which forecast was not far to what the researcher had predicted: March 2022 has a negative difference of 29 thousand pieces and July 2022 has a negative difference of 46 thousand pieces. For 2023, the company is expecting the highest sale pick on the second quarter, specifically in June 2023 with 222 thousand pieces and the lowest sale at the begging of the last quarter with a negative sale of 11 thousand pieces. Here it is important to remember that negative sales are possible due to credit notes; These documents represent returns our quality claims.

Figure 32 shows the forecast for the *bottle* EMS041. For this part number, there is 1 month of 2022 out of the confidence bound – April

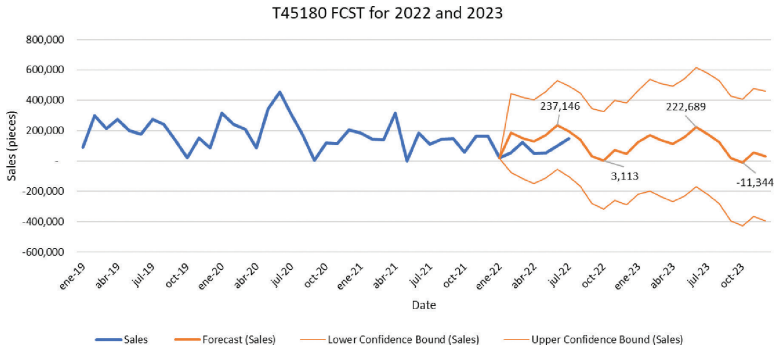



**Figure 31.** T45180 forecast for 2022 and 2023.

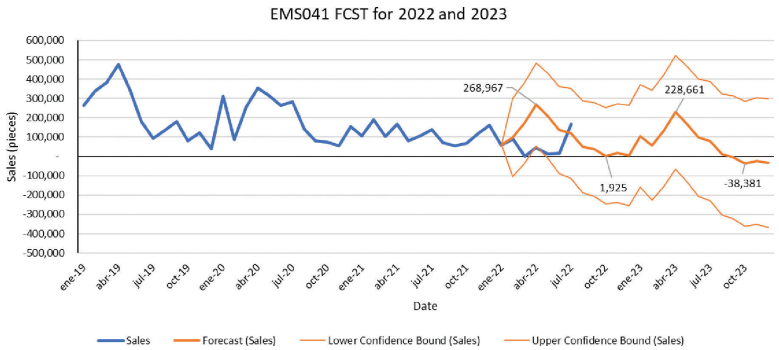



**Figure 32.** EMS041 forecast for 2022 and 2023.

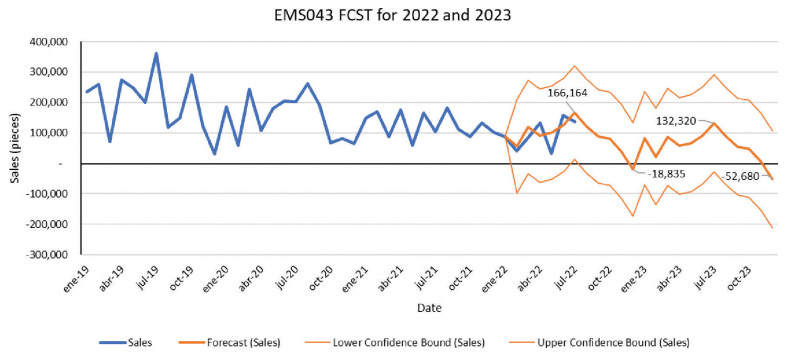



**Figure 33.** EMS043 forecast for 2022 and 2023.

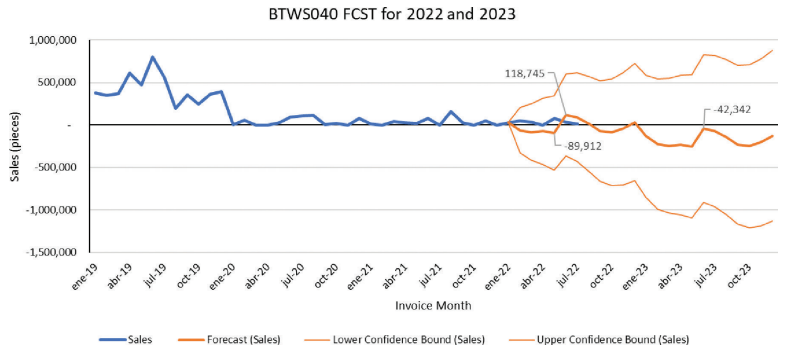



**Figure 34.** BTWS040 forecast for 2022 and 2023.

2022 which fell 7 thousand pieces under the lower confidence bound. February 2022 has the closest pieces sold in comparison with the forecast, with a difference of only 6 thousand pieces. For 2023, the company expects the highest sale to occur during the first month of the second quarter (April 2023) with 228.6 thousand pieces, while the lowest quarter is expected to be the last quarter of the year with negative sales every month.

Figure 33 represents the sales and forecast for the *bottle* with part number EMS043. For 2022, there are 6 out of 7 months with very close accuracy level between actual sales and the forecast, with February being the closest one with a difference of only 15 thousand pieces. This means that there are no points falling out of the confidence bounds. For the 2023 forecast, the company expects July 2023 to be the month with the highest sales, being these 132 thousand pieces, and December the lowest with a negative sale.

Figure 34 is another case of high rejection rates having a negative impact in the company´s forecast. In this case, it is the *bottle* BTWS040,

which has a similar issue as the label ETGE01000. This part number is not showing any positive months for 2023, but again the sales during the first semester of 2022 have been better than expected, having 6 out of 7 months in positive sales and 1 month in ceros (not negative sales). This part number will also need to be updated as the rejection rate is back to normal.

## 4.3 SUMMARY

The overall aim of this study was to identify the top 10 sold products during the past 3.5 years and forecast the sales for the year 2023. The research was based on primary data from the company and a quantitative analysis was overdone. This research started with a data set of 18 variables by 19,168 entries. After cleaning and validating the data, the researcher ends up with a data set of 17 variables and 18,268 entries. From that data set, the researcher selected and filter by some variables so that the research ended up working with 168 part numbers out of the 196 that are contained in the data.

These 168 part numbers correspond to 7 family products, of which only the top 10 sold products are found. As the researcher mentioned during the analysis, 20 part numbers were found in the top 10 sales positions over the past 3.5 years but the top 10 products with the higher sales were selected during all this time. These products are divided as follows:

- 7 products from the *bottle* family
- 2 products from the *lid* family
- 1 product from the *label* family

From the top 10 products, 3 products have shown constant grow during the study period, 2 part numbers with important drops, and 5 products with minor fluctuations during the same period.

During the significance test, the researcher rejected the null hypothesis, which stated that there was no statistical significance between the products categorize as top 10 and those that were not part of this group, which translates into the top 10 products having an impact and a relationship with those not listed as top 10.

Then the researcher worked on the sales fluctuation analysis and found that the *bottles* sales have been steady during the past years, *labels* suffer an important fall, and *lids* are the family with highest fluctuations. In addition, the company has the best sales during the second quarter of the year and the lowest sales during the last quarter.

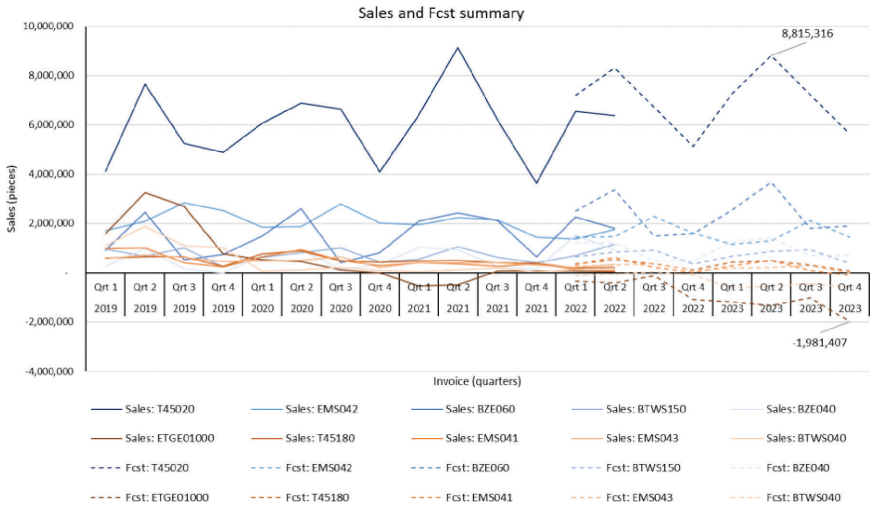



**Figure 35.** Top 10 products sales and forecast summary.

Finally for the forecast analysis, Figure 35 shows the summary of the sales and the forecast by quarter. The second quarter of 2023 shows the best sales of the year, corresponding to the part number T45020. The lowest sale is shown for the fourth quarter of 2023 for part number ETGE01000. As mentioned before, "negative sales" are possible due to returns and rejections from the customers and two part numbers (the same with the important drops) will need to be updated because the forecast is showing mostly "negative sales" do to past rejection rates.