

# Using Data Science to improve sales performance of sanitizing products

Víctor De la Torre  
Jr. Data Scientist  
May 1st 2025

# Content

1. Project Introduction
2. Data Cleaning and Transformation with Python
3. Exploratory Data Analysis (EDA) and Visualizations
4. Product or Region Segmentation with Clustering
5. Data Analysis with SQL
6. Creating a Dashboard in PowerBI
7. Sales Forecasting with Machine Learning
8. Conclusions and Recommendations

# 1. Project Introduction

**Purpose of the project:** to use Data Science to evaluate the current sales strategy of Reckitt's Bleach and Laundry Bleach products.



## Brands to be analyzed: Vanish and Lysol.



# 1. Project Introduction

## Relevance of the analysis

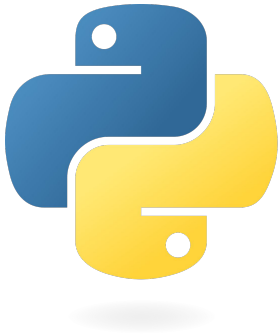
Analyzing and processing large volumes of data will make it possible to:

- Identify consumer behavior patterns
- Anticipate trends
- Optimize marketing, positioning and pricing strategies
- Improve sales efficiency



# 1. Project Introduction

Tools:



**Python:**

- Data cleaning, manipulation and analysis



**SQL Server:**

- Database creation and management



**Power BI:**

- Dashboard creation and publication

## 2. Data Cleaning and Data Transformation with Python

### Data loading

The Pandas library was used to load the data:

Dataframe 1: fact\_sales\_df

	WEEK	ITEM_CODE	TOTAL_UNIT_SALES	TOTAL_VALUE_SALES	TOTAL_UNIT_AVG_WEEKLY_SALES	REGION
0	34-22	7501058792808BP2	0.006	0.139	1.000	TOTAL AUTOS AREA 5
1	34-22	7501058715883	0.487	116.519	2.916	TOTAL AUTOS AREA 5
2	34-22	7702626213774	1.391	68.453	5.171	TOTAL AUTOS AREA 5
3	34-22	7501058716422	0.022	1.481	1.833	TOTAL AUTOS AREA 5
4	34-22	7501058784353	2.037	182.839	5.375	TOTAL AUTOS AREA 5

Dataframe 2: dim\_product\_df

	MANUFACTURER	BRAND	ITEM	ITEM_DESCRIPTION	CATEGORY	FORMAT	ATTR1	ATTR2	ATTR3
0	INDS. ALEN	CLORALEX	0000075000592	CLORALEX EL RENDIDOR BOT.PLAST. 250ML NAL. 000...	1	LIQUIDO	CLORO	CLORO	NO DEFINIDO
1	INDS. ALEN	CLORALEX	0000075000608	CLORALEX EL RENDIDOR BOT.PLAST. 500ML NAL. 000...	1	LIQUIDO	CLORO	CLORO	NO DEFINIDO
2	INDS. ALEN	CLORALEX	0000075000615	CLORALEX EL RENDIDOR BOT.PLAST. 950ML NAL. 000...	1	LIQUIDO	CLORO	CLORO	NO DEFINIDO
3	INDS. ALEN	CLORALEX	0000075000622	CLORALEX EL RENDIDOR BOT.PLAST. 2000ML NAL. 000...	1	LIQUIDO	CLORO	CLORO	NO DEFINIDO
4	INDS. ALEN	CLORALEX	0000075000639	CLORALEX EL RENDIDOR BOT.PLAST. 3750ML NAL. 000...	1	LIQUIDO	CLORO	CLORO	NO DEFINIDO

Dataframe 3: dim\_segment\_df

	CATEGORY	ATTR1	ATTR2	ATTR3	FORMAT	SEGMENT
0	1	CLORO	CLORO	BAMBINO	LIQUIDO	BLEACH
1	1	CLORO	CLORO	GERMICIDA	LIQUIDO	BLEACH
2	1	CLORO	CLORO	MASCOTAS	LIQUIDO	BLEACH
3	1	CLORO	CLORO	MULTIUSOS	GEL	BLEACH
4	1	CLORO	CLORO	MULTIUSOS	LIQUIDO	BLEACH

Dataframe 3: dim\_calendar\_df

	WEEK	YEAR	MONTH	WEEK_NUMBER	DATE
0	01-21	2021	1	1	2021-01-10
1	02-21	2021	1	2	2021-01-17
2	03-21	2021	1	3	2021-01-24
3	04-21	2021	1	4	2021-01-31
4	05-21	2021	2	5	2021-02-07

Dataframe 5: dim\_category\_df

	ID_CATEGORY	CATEGORY
0	1	FABRIC TREATMENT and SANIT\y\n
1	2	AIR CARE
2	3	LAVAJILLAS
3	4	MEGA SUPERFICIES
4	5	LAVATORY CARE & BRC

## 2. Data Cleaning and Data Transformation with Python

### Data cleaning:

Each dataframe was cleaned using the following methods:

Method	Use
head()	Show dataframe's first rows
info()	Dataframe's general information
value_counts()	Information about all values in each column
isnull()	Null values identification
fillna()	Null values substitution
uplicated()	Identification of duplicated values
drop_duplicates()	Elimination of duplicated values
drop()	Elimination of unnecessary rows or columns
rename()	Change column or index name
iloc()	Modification of typographical errors
merge()	Dataframe merging using common column
to_csv()	Exportation of .csv files

## 2. Data Cleaning and Data Transformation with Python

### Data Loading:

The regions corresponding to 'TOTAL AUTOS' were replaced by numbers as follows:

Original Name	Updated Name
TOTAL AUTOS SCANNING MEXICO	0
TOTAL AUTOS AREA 1	1
TOTAL AUTOS AREA 2	2
TOTAL AUTOS AREA 3	3
TOTAL AUTOS AREA 4	4
TOTAL AUTOS AREA 5	5
TOTAL AUTOS AREA 6	6



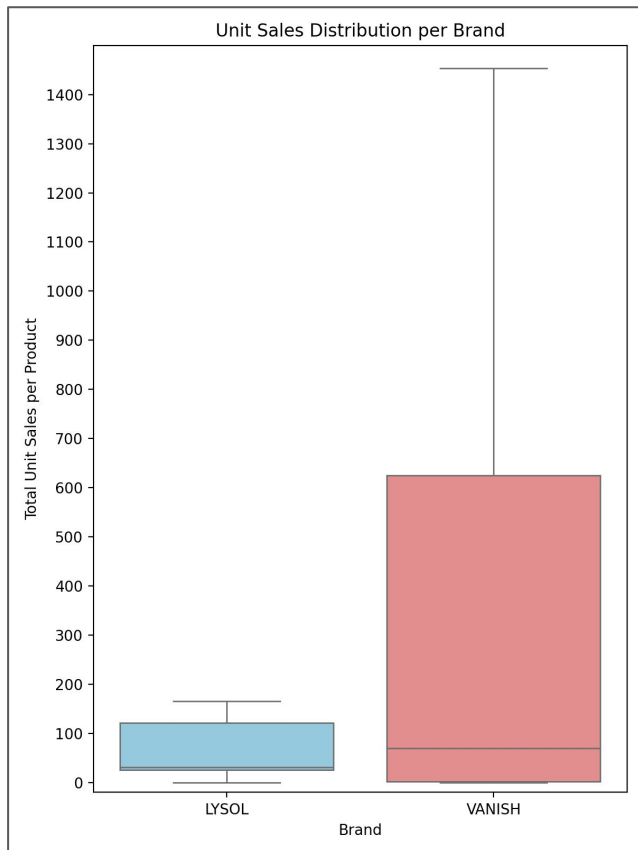
## 2. Data Cleaning and Data Transformation with Python

As a final result, a consolidated dataframe was obtained with information on sales and earnings realized in 2022 and 2023.

ITEM_CODE	TOTAL_UNIT_SALES	TOTAL_VALUE_SALES	TOTAL_UNIT_AVG_WEEKLY_SALES	REGION	MANUFACTURER	BRAND	ITEM_DESCRIPTION	FORMAT	ATTR1	ATTR2	ATTR3	ID_CATEGORY	CATEGORY	WEEK	YEAR	MONTH	WEEK_NUMBER	DATE
501058792808BP2	0.006	0.139	1.000	5	RECKITT	VANISH	1/4+MMCRYSTALWHITE...	POLVO	SAFE BLEACH	FABRIC TREATMENT	ROSA	1	FABRIC TREATMENT AND SANITIZER	34-22	2022	8	34	2022-08-28
7501058715883	0.487	116.519	2.916	5	RECKITT	VANISH	ANCHAS BOLSA 1.8K...	POLVO	SAFE BLEACH	FABRIC TREATMENT	ROSA	1	FABRIC TREATMENT AND SANITIZER	34-22	2022	8	34	2022-08-28
7702626213774	1.391	68.453	5.171	5	RECKITT	VANISH	INCHAS DOYPACK 24...	POLVO	SAFE BLEACH	FABRIC TREATMENT	ROSA	1	FABRIC TREATMENT AND SANITIZER	34-22	2022	8	34	2022-08-28
7501058716422	0.022	1.481	1.833	5	RECKITT	VANISH	NCHA AHORRO DEL ...	POLVO	SAFE BLEACH	FABRIC TREATMENT	ROSA	1	FABRIC TREATMENT AND SANITIZER	34-22	2022	8	34	2022-08-28
7501058784353	2.037	182.839	5.375	5	RECKITT	VANISH	JTE 450 GR NAL 7501...	POLVO	SAFE BLEACH	FABRIC TREATMENT	ROSA	1	FABRIC TREATMENT AND SANITIZER	34-22	2022	8	34	2022-08-28

For exploratory analysis, k-means segmentation and time series prediction using ARIMA, the data were grouped with the `group_by()` function by `ITEM_CODE`.

# 3. Exploratory Data Analysis (EDA) and Visualization



## Unit Sales Distribution

### Vanish products (pre-washers and bleaches):

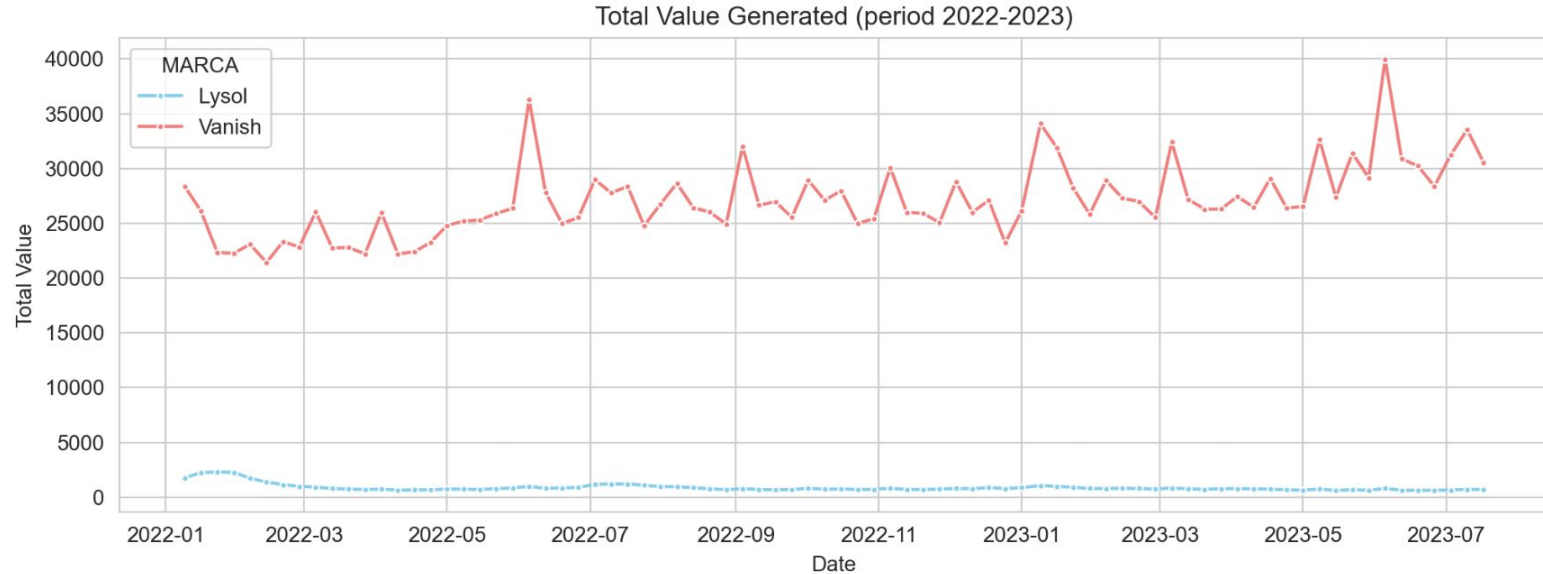
They have high unit sales (products have been sold a greater number of times) and variable unit sales.

### Lysol products (sanitizers):

Unit sales are low (products have been sold a smaller number of times) but at the same time are consistent.

# 3. Exploratory Data Analysis (EDA) and Visualization

Total value generated by Vanish and Lysol in the period 2022 - 2023



- **Vanish** : earnings have increased over time.
- **Lysol**: earnings have remained stable over the years.

# 3. Exploratory Data Analysis (EDA) and Visualization

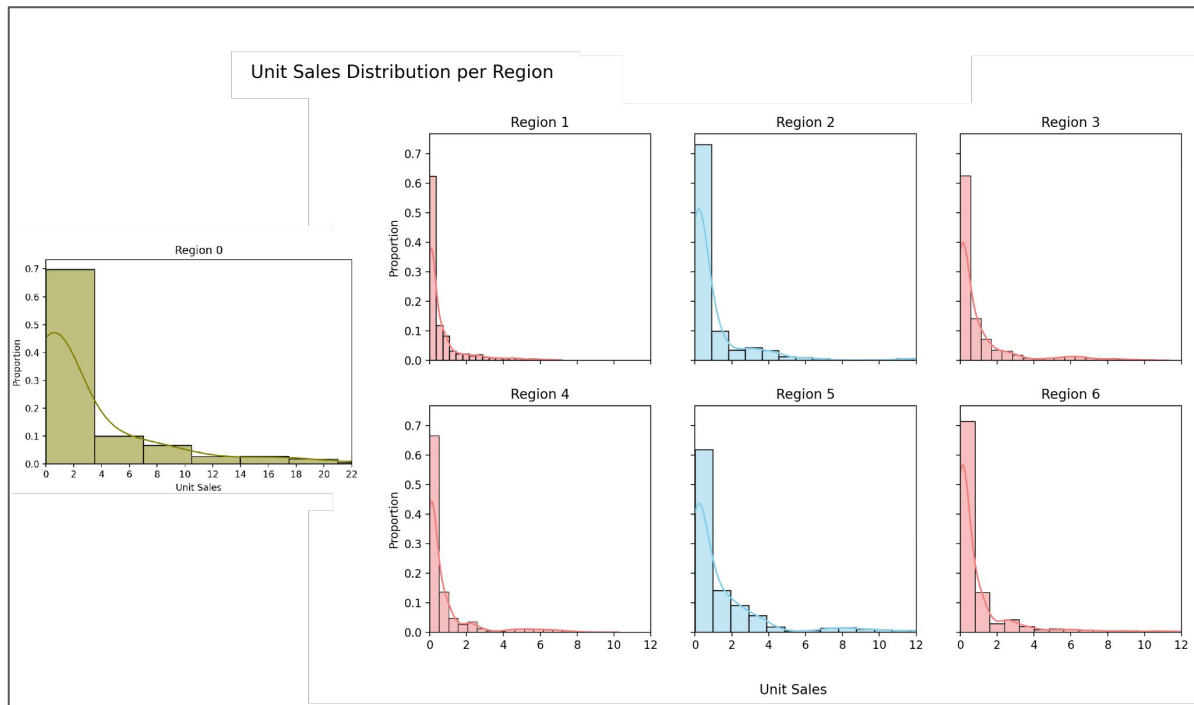
## General:

Products in all regions are sold in small quantities (right skewed distribution).

## Particularities:

Region 0 presents the largest variety of unit sales (longest tail)

Region 2, 3 and 6 present the lowest concentration of sales (80% of sales are between 0 and 1 products).



# 3. Exploratory Data Analysis (EDA) and Visualization



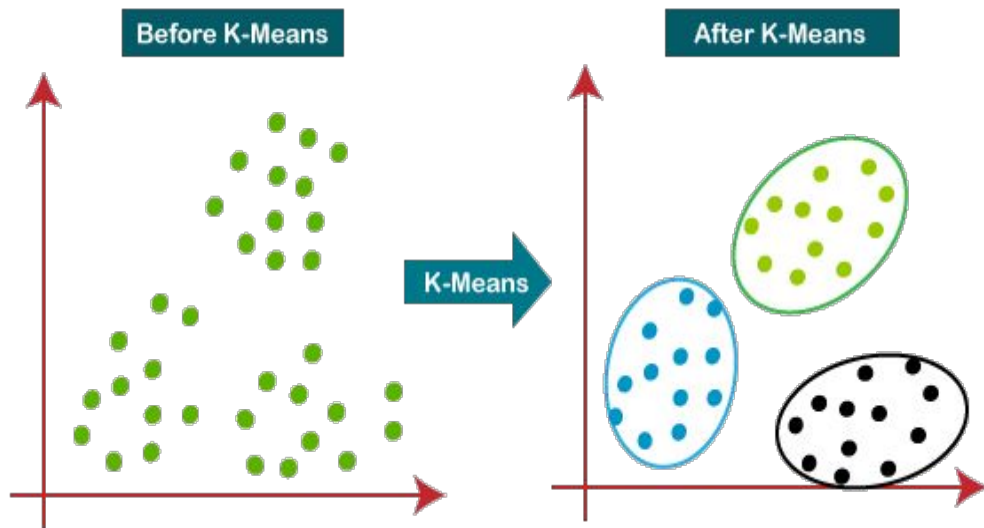
Reckitt's main competitors are Clorox and Inds. Alen, the latter being the ones that stand out for their maximum earnings in this comparative graph.

## 4. Product Segmentation

The K-means clustering algorithm was used to segment the products based on key variables such as total sales, total profits, product type, etc.

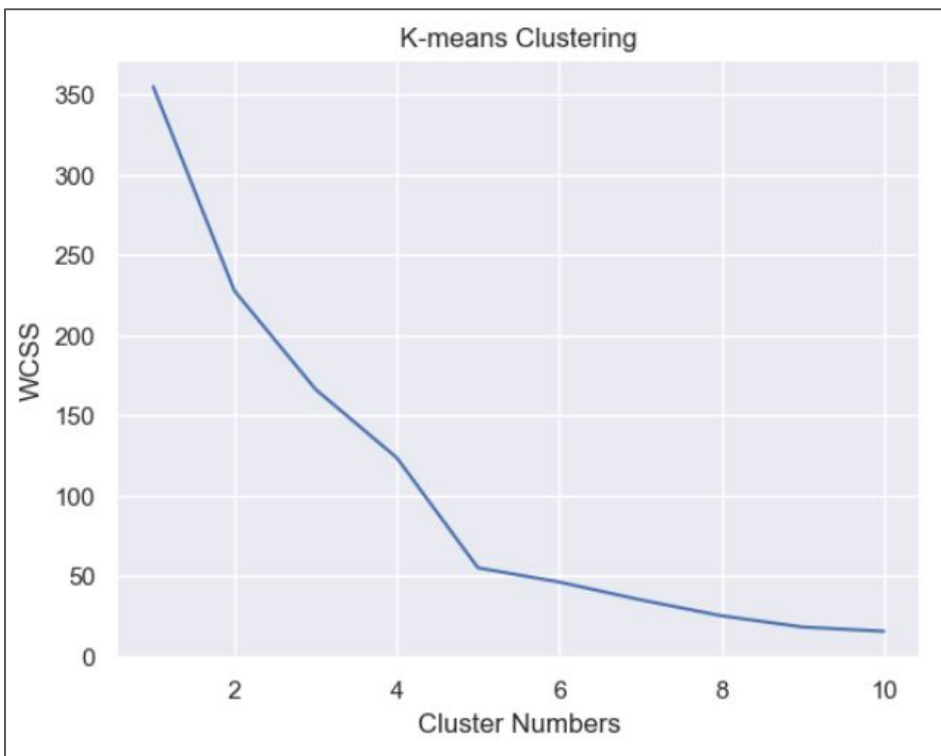
The data were transformed using the following tools:

Tool	Use
Target Encoder	Transformation of categorical variables into numerical variables
Standard Scaler	Data standardization
PCA	Dimension reduction



## 4. Product Segmentation

### Cluster Optimal Number Determination

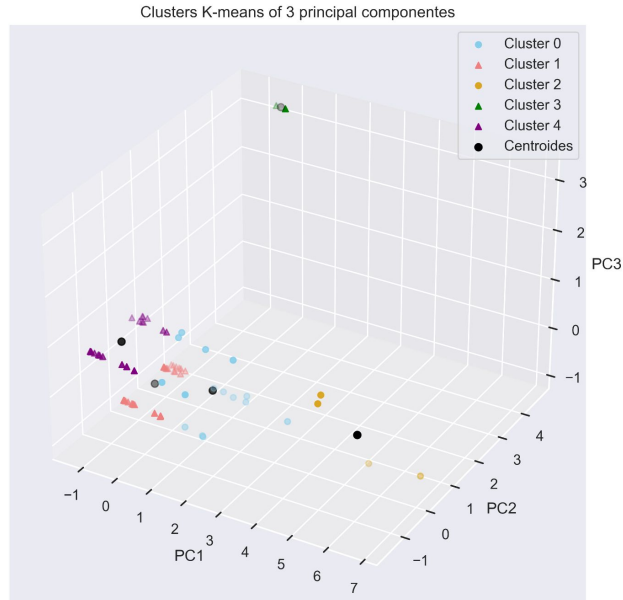


The elbow method was applied to determine the optimal number of clusters.

According to the graph, the optimal number of clusters is  $k=5$ .

# 4. Product Segmentation

## K-means application



Five clusters were identified using data reduced to 3 dimensions by PCA and a value of  $k = 5$ .

The clusters appear to be related to the number of units sold and total profits (next slide).



## 4. Product Segmentation

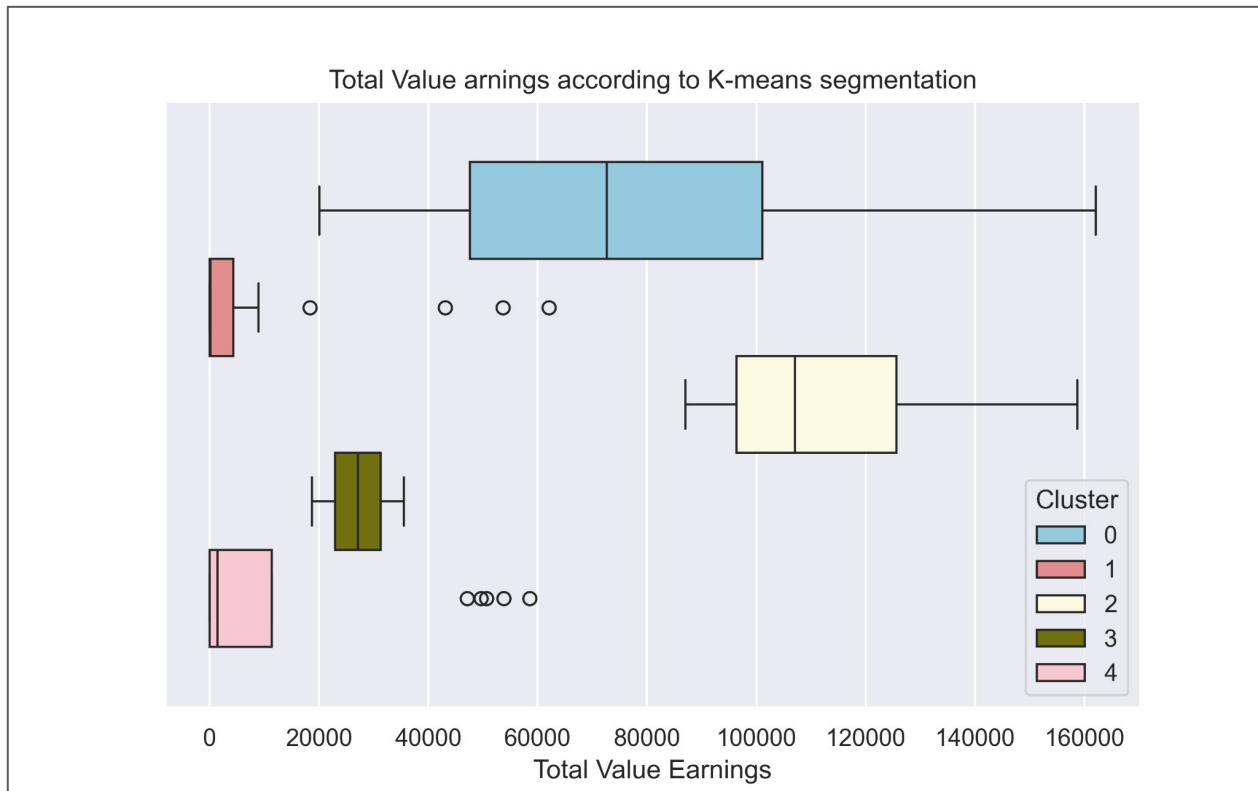
### Adding cluster label to the original data

#### Value of the medians:

- Cluster 1 = 175.16
- Cluster 4 = 1 466.41
- Cluster 3 = 27 122.88
- Cluster 0 = 72 658.36
- Cluster 2 = 107 064.16

#### Number of observations per cluster:

- Cluster 1 : 35
- Cluster 4 : 21
- Cluster 3 : 2
- Cluster 0 : 15
- Cluster 2 : 4

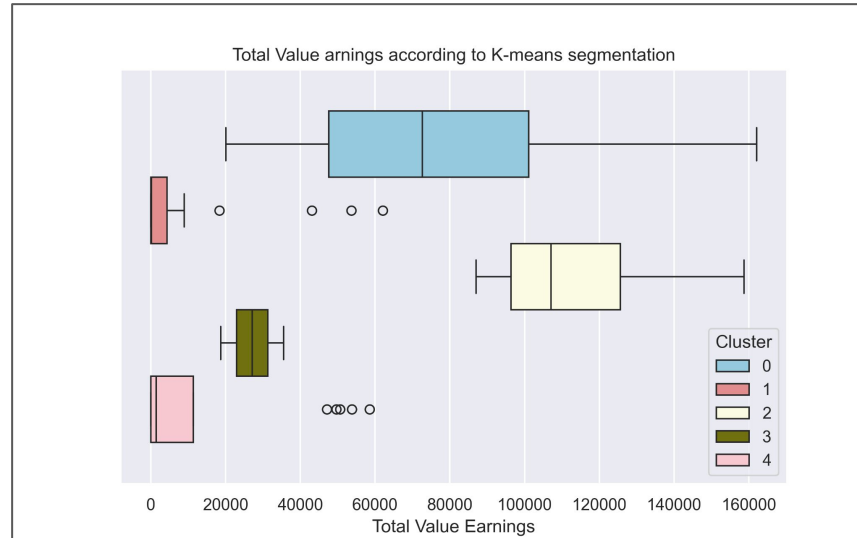


## 4. Product Segmentation

According to the cluster assignment, clusters 1 and 4 belong to products whose profits are 'low' (less than 1500 units).

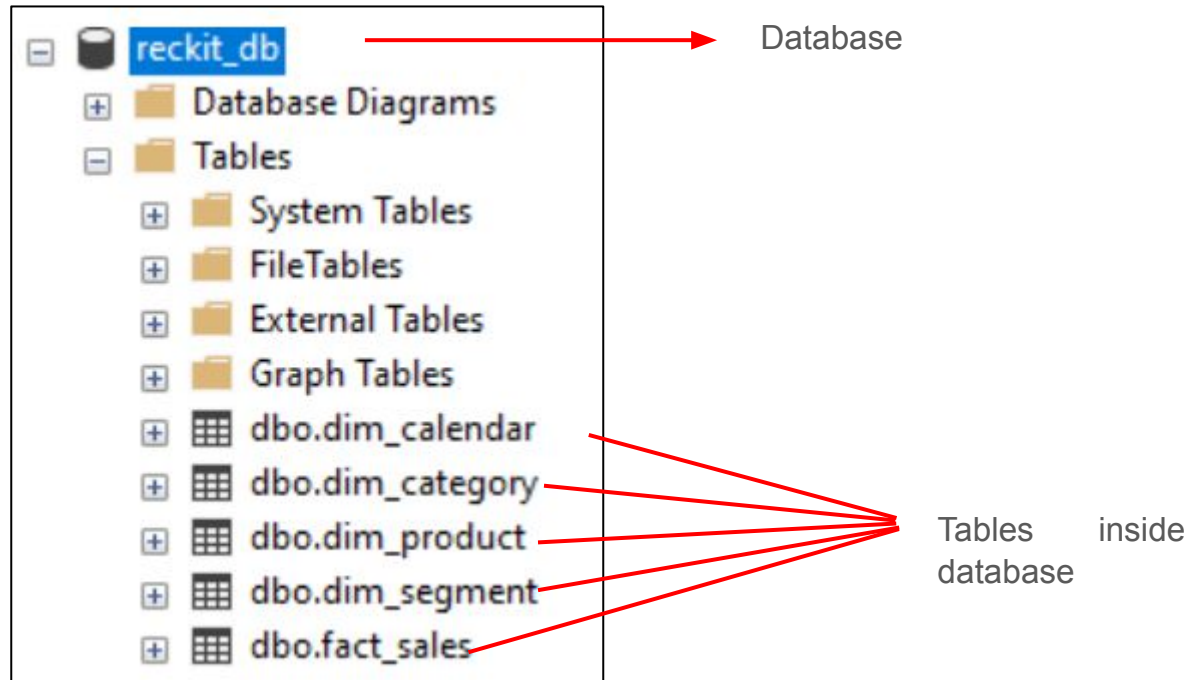
Cluster 3 belongs to products with intermediate profits (approximately 27 000 units).

Clusters 0 and 2 belong to products that generated high profits (greater than 70,000 units). It should be noted that sales in these last groups were variable compared to the other clusters.



## 5. Data Analysis with SQL

The rekit\_db database was created in SQL Server and the tables dim\_calendar, dim\_category, dim\_product, dim\_segment and dbo.fact\_sales were imported:



## 5. Data Analysis with SQL

Basic queries were run to verify correct data loading:

```
SELECT*  
FROM dbo.dim_calendar  
  
SELECT*  
FROM dbo.dim_category  
  
SELECT*  
FROM dbo.dim_product  
  
SELECT*  
FROM dbo.dim_segment  
  
SELECT*  
FROM dbo.fact_sales
```

102 %

Results Messages

	WEEK	YEAR	MONTH	WEEK_NUMBER	DATE
1	01-21	2021	1	1	2021-01-10
2	02-21	2021	1	2	2021-01-17
3	03-21	2021	1	3	2021-01-24
4	04-21	2021	1	4	2021-01-31
5	05-21	2021	2	5	2021-02-07
6	06-21	2021	2	6	2021-02-14
7	07-21	2021	2	7	2021-02-21
8	08-21	2021	2	8	2021-02-28

ID_CATEGORY	CATEGORY
1	FABRIC TREATMENT and SANIT
2	AIR CARE
3	LAVAVAJILLAS
4	MEGA SUPERFICIES
5	LAVATORY CARE & BRC

MANUFACTURER	BRAND	ITEM	ITEM_DESCRIPTION	CATEGORY	FORMAT	ATTR1	ATTR2	ATTR3	
1	INDS. ALEN	CLORALEX	0000075000592	CLORALEX EL RENDIDOR BOT. PLAST. 250ML NAL. 00000...	1	LIQUIDO	COLORO	COLORO	NO DEFINIDO
2	INDS. ALEN	CLORALEX	0000075000608	CLORALEX EL RENDIDOR BOT. PLAST. 500ML NAL. 00000...	1	LIQUIDO	COLORO	COLORO	NO DEFINIDO
3	INDS. ALEN	CLORALEX	0000075000615	CLORALEX EL RENDIDOR BOT. PLAST. 950ML NAL. 00000...	1	LIQUIDO	COLORO	COLORO	NO DEFINIDO
4	INDS. ALEN	CLORALEX	0000075000622	CLORALEX EL RENDIDOR BOT. PLAST. 2000ML NAL. 00000...	1	LIQUIDO	COLORO	COLORO	NO DEFINIDO
5	INDS. ALEN	CLORALEX	0000075000639	CLORALEX EL RENDIDOR BOT. PLAST. 3750ML NAL. 00000...	1	LIQUIDO	COLORO	COLORO	NO DEFINIDO
6	INDS. ALEN	CLORALEX	0000075000646	CLORALEX FANTASIA FLORAL BOT PLAST 950ML NAL 000...	1	LIQUIDO	COLORO	COLORO	OTR. TIPOS
7	INDS. ALEN	CLORALEX	0000075000653	CLORALEX FANTASIA FLORAL BOT PLAST 2000 ML NAL 0...	1	LIQUIDO	COLORO	COLORO	OTR. TIPOS
8	INDS. ALEN	CLORALEX	0000075000677	CLORALEX FRESCURA CITRICA BOT PLAST 950ML NAL 0...	1	LIQUIDO	COLORO	COLORO	NO DEFINIDO

CATEGORY	ATTR1	ATTR2	ATTR3	FORMAT	SEGMENT
37	1	SAFE...	FABR...	PRE LAVA...	BARRA BAR
38	1	SAFE...	FABR...	PRE LAVA...	GEL LIQUID &...
39	1	SAFE...	FABR...	PRE LAVA...	LIQUIDO PRETRE...
40	1	SAFE...	FABR...	PRE LAVA...	POLVO POWDER
41	1	SAFE...	FABR...	QUITAMA...	GEL LIQUID &...
42	1	SAFE...	FABR...	QUITAMA...	LIQUIDO LIQUID &...
43	1	SAFE...	FABR...	QUITAMA...	POLVO POWDER
44	1	SAFE...	FABR...	ROPA BL...	LIQUIDO LIQUID &...

WEEK	ITEM_CODE	TOTAL_UNIT_SALES	TOTAL_VALUE_SALES	TOTAL_UNIT_AVG_WEEKLY_SALES	REGION	
1	34-22	7501058792808BP2	0.00600000005215406	0.138999998569489	1	TOTAL AUTOS AREA 5
2	34-22	7501058715883	0.486999988555908	116.518997192383	2.91599988937378	TOTAL AUTOS AREA 5
3	34-22	7702626213774	1.39100003242493	68.4530029296875	5.1710000038147	TOTAL AUTOS AREA 5
4	34-22	7501058716422	0.0219999998807907	1.48099994659424	1.83299994468689	TOTAL AUTOS AREA 5
5	34-22	7501058784353	2.03699994087219	182.839004516602	5.375	TOTAL AUTOS AREA 5
6	34-22	7501058716064	0.0049999988824129	0.679000020027161	1.25	TOTAL AUTOS AREA 5

## 5. Data Analysis with SQL

Table joins were also performed to run more advanced queries and obtain insights on sales and profits generated.

### Consulta # 1: query per category

```
3 SELECT ROUND(SUM(fs.TOTAL_UNIT_SALES),2) AS TOTAL_UNIT_SALES,  
4         ROUND(SUM(fs.TOTAL_VALUE_SALES),2) TOTAL_VALUE_SALES,  
5         ROUND(SUM(fs.TOTAL_UNIT_AVG_WEEKLY_SALES),2) TOTAL_UNIT_AVG_WEEKLY_SALES,  
6         dp.CATEGORY,  
7         ds.ATTR1  
8 FROM dbo.fact_sales fs  
9     INNER JOIN dbo.dim_product dp ON fs.ITEM_CODE = dp.ITEM  
10    INNER JOIN dbo.dim_segment ds ON dp.CATEGORY = ds.CATEGORY  
11 GROUP BY dp.CATEGORY, ds.ATTR1  
12 ORDER BY ROUND(SUM(fs.TOTAL_VALUE_SALES),2) DESC
```



Results Messages				
TOTAL_UNIT_SALES	TOTAL_VALUE_SALES	TOTAL_UNIT_AVG_WEEKLY_SALES	CATEGORY	ATTR1
10577526.16	298157217.1	33269629.1	1	SAFE BLEACH
5484643.19	154600038.5	17250918.79	1	CLORO
4309362.51	121471458.82	13554293.34	1	PRELAVADOR
391760.23	11042859.89	1232208.49	1	SANITIZANTE

**Insight obtained:** there is only sales information related to category 1, where products categorized as 'safe bleach' generated the most value.


## 5. Data Analysis with SQL

### Consulta # 2: query per region

```

20 SELECT ROUND(SUM(TOTAL_UNIT_SALES),2) AS TOTAL_UNIT_SALES,
21        ROUND(SUM(TOTAL_VALUE_SALES),2) AS TOTAL_VALUE_SALES,
22        ROUND(SUM(TOTAL_UNIT_AVG_WEEKLY_SALES),2) AS TOTAL_UNIT_AVG_WEEKLY_SALES
23        REGION
24 FROM dbo.fact_sales
25 GROUP BY REGION
26 ORDER BY TOTAL VALUE SALES DESC

```



	TOTAL_UNIT_SALES	TOTAL_VALUE_SALES	TOTAL_UNIT_AVG_WEEKLY_SALES	REGION
1	195880.05	5521429.32	228291.49	TOTAL AUTOS SCANNING MEXICO
2	45319.92	1188796.15	180553.22	TOTAL AUTOS AREA 2
3	36266.9	1153335.54	175129.12	TOTAL AUTOS AREA 5
4	38585.22	983957.57	183537.89	TOTAL AUTOS AREA 6
5	25896.05	803655.34	181328.12	TOTAL AUTOS AREA 3
6	25791.93	714249.98	128718.31	TOTAL AUTOS AREA 1
7	24020.15	677436	154650.33	TOTAL AUTOS AREA 4

**Insight gained:** the region 'TOTAL AUTOS SCANNING MEXICO' (named as region 0 in the python dataframes) generated the most sales value, while region 4 generated the least.

### Consulta # 3: query per period time

```

35 SELECT ROUND(SUM(fs.TOTAL_UNIT_SALES),2) AS TOTAL_UNIT_SALES,
36        ROUND(SUM(fs.TOTAL_VALUE_SALES),2) AS TOTAL_VALUE_SALES,
37        ROUND(SUM(fs.TOTAL_UNIT_AVG_WEEKLY_SALES),2) AS TOTAL_UNIT_AVG_WEEKLY_SALES,
38        dc.YEAR
39 FROM dbo.fact_sales fs
40     INNER JOIN dbo.dim_calendar dc ON fs.WEEK = dc.WEEK
41 GROUP BY dc.YEAR
42 ORDER BY TOTAL_VALUE_SALES DESC

```

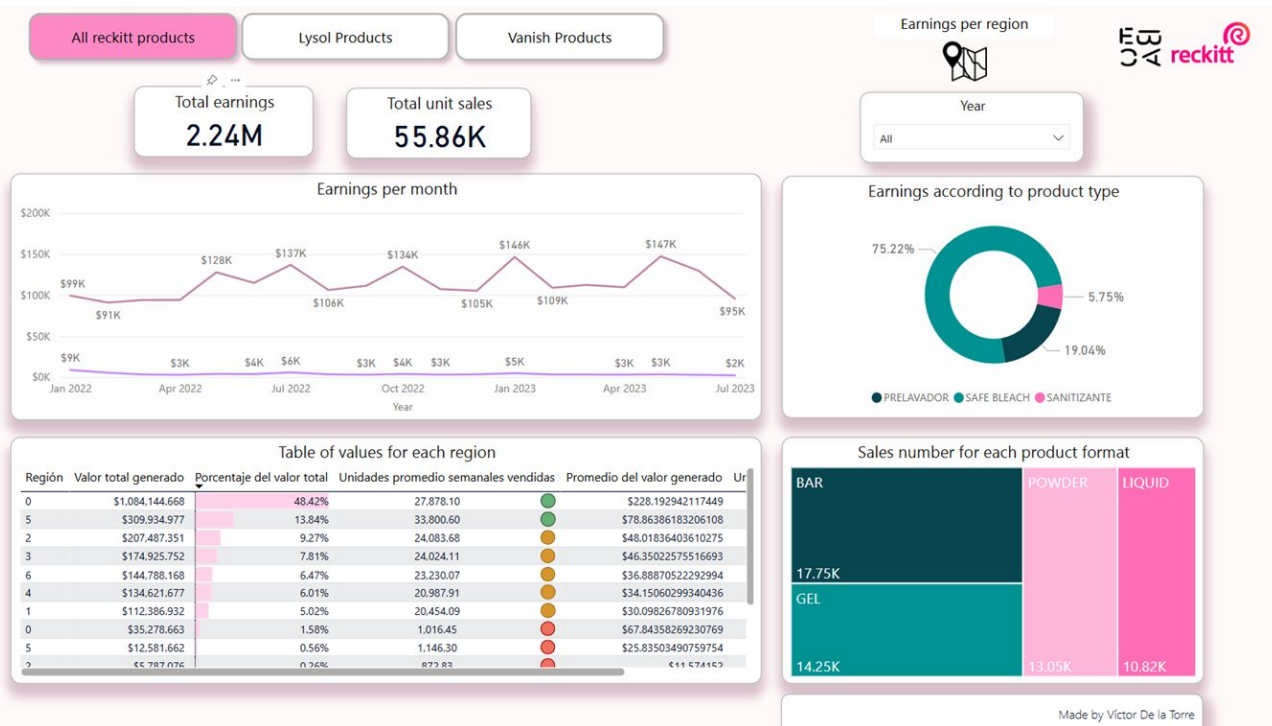


	TOTAL_UNIT_SALES	TOTAL_VALUE_SALES	TOTAL_UNIT_AVG_WEEKLY_SALES	YEAR
1	255405.64	7008619.05	823699.31	2022
2	136354.59	4034240.84	408509.17	2023

**Insight gained:** 2022 generated more sales value than 2023

# 6. Creating a Dashboard in PowerBI

Page 1 of 2



## Dashboard Structure:

The dashboard includes a summary of total profits and units sold along with graphs and tables showing performance by product type.

## Interactive displays:

The user can filter the information by Lysol, Vanish or both products and view it in the graphs and table. It can also be filtered by product format (stick, powder, liquid and gel).

# 6. Creating a Dashboard in PowerBI

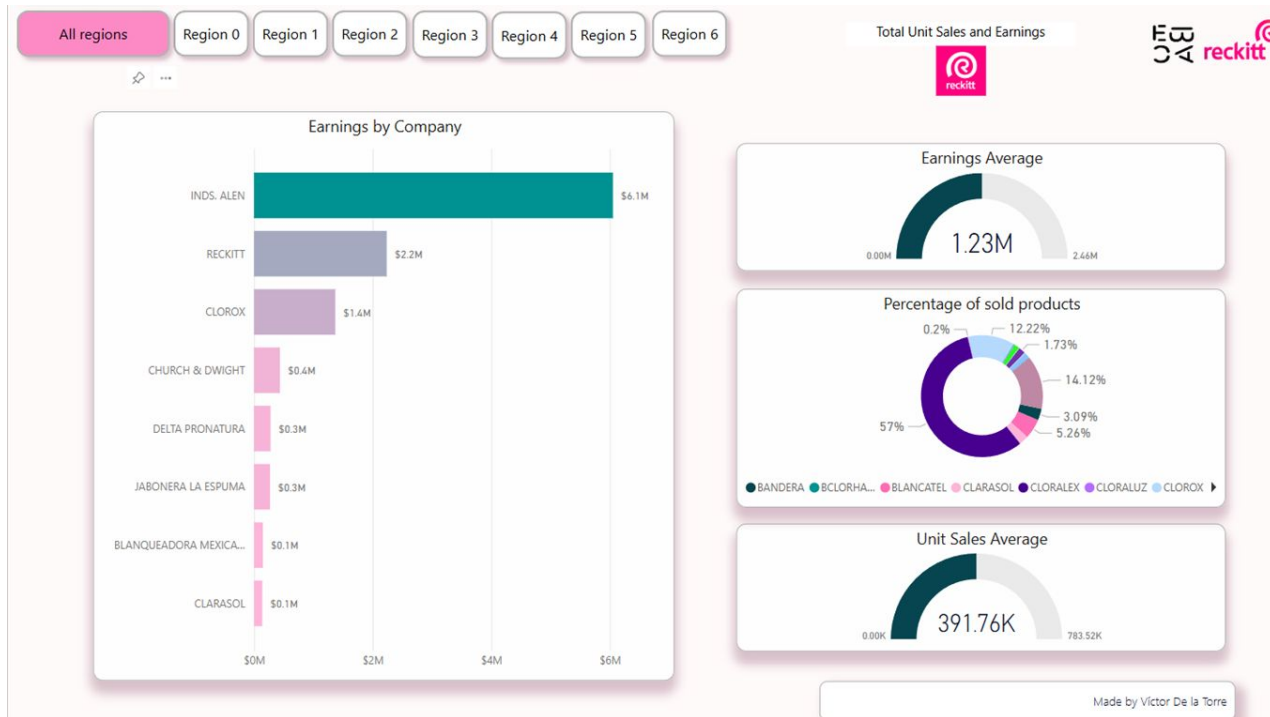
Page 2 de 2

## Dashboard Structure:

The dashboard includes profit and unit sales information by region and a comparison of Reckitt's main competitors has been added.

## Interactive Visualizations:

The user can filter the profit and unit sales information by one or more regions.

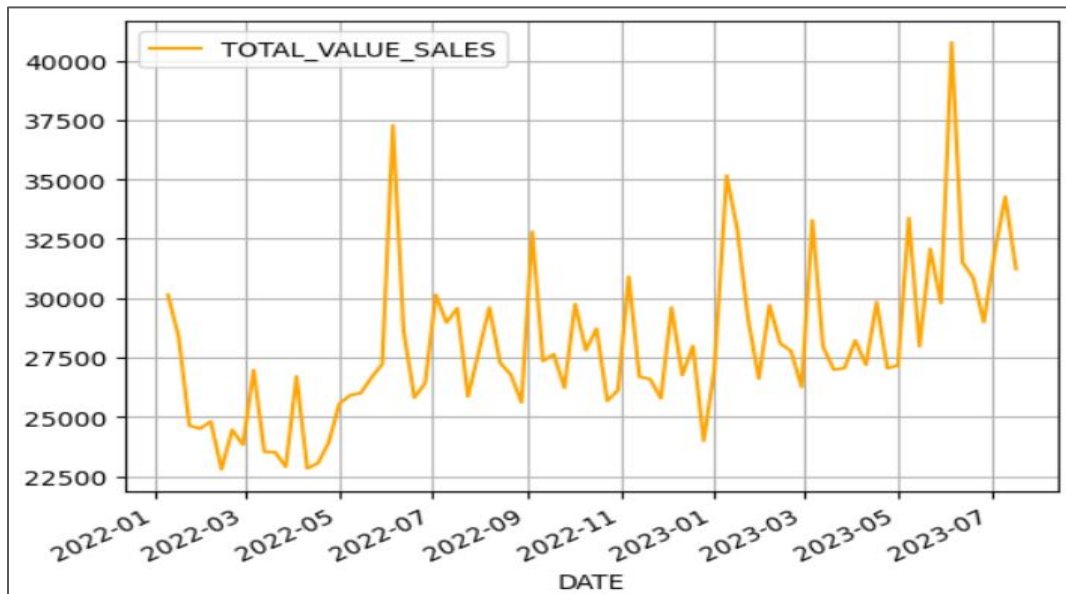




## 7. Sales Prediction with Machine Learning

An ARIMA time series model was used to predict earnings, based on sales patterns observed in historical data. The variable `TOTAL_VALUE_SALES` was used, which represents the total value generated.

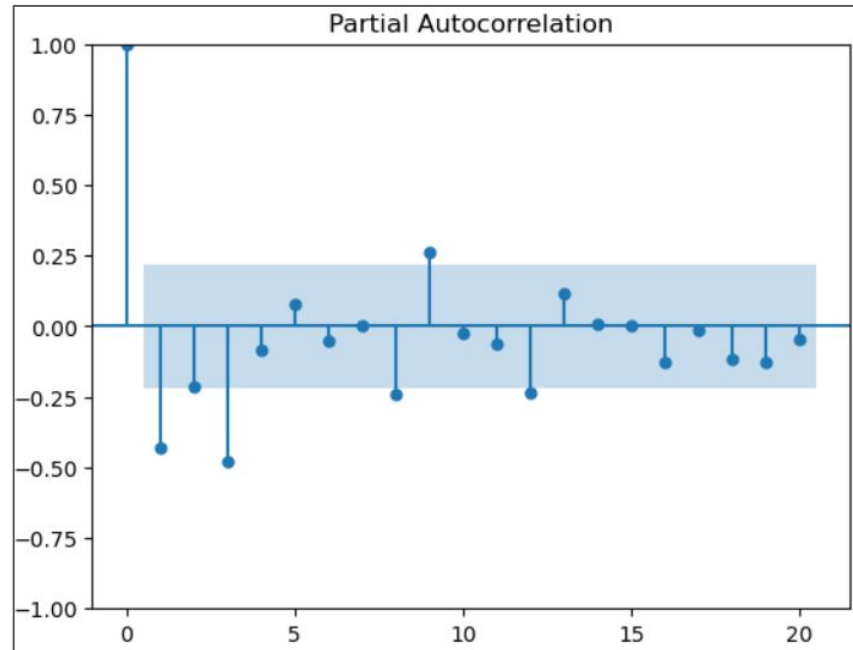
A graph of historical Reckitt product earnings is shown below:



## 7. Sales Prediction with Machine Learning

Before performing the forecasts with ARIMA, the calculation of differencing between observations ( $\Delta y_t = y_t - y_{t-1}$ ) was applied and the Dickey-Fuller test was applied to check that the data were stationary.

The partial autocorrelation function was also applied to obtain the optimal p value:



Optimal value  $p = 3$

## 7. Sales Prediction with Machine Learning

To determine the hyper parameters of the ARIMA model, the Akaike Information Criterion (AIC) was used to compare different statistical models and select the most appropriate model.

### Code

```
# Importation of ARIMA module to estimate AIC
from statsmodels.tsa.arima.model import ARIMA
lista= []
for p in range (1,4): # iteration to find p optimal value
    for q in range (1,4): # q optimal value
        for d in range (1,2):
            mod_ar = ARIMA(df['diferencias_1'], order=(p,d,q)) # ARIMA object
            res_ar = mod_ar.fit() # Model application
            lista.append(res_ar.aic)
        try:
            print(f'AIC for ARIMA({p},{d},{q}) es: {res_ar.aic}')
        except Exception as e:
            print(f'Error ARIMA({p},1,{q}): {e}')

# List sorting
lista.sort()

# Let's drop the column diferencias_1
df.drop(columns='diferencias_1', inplace=True)
lista
```

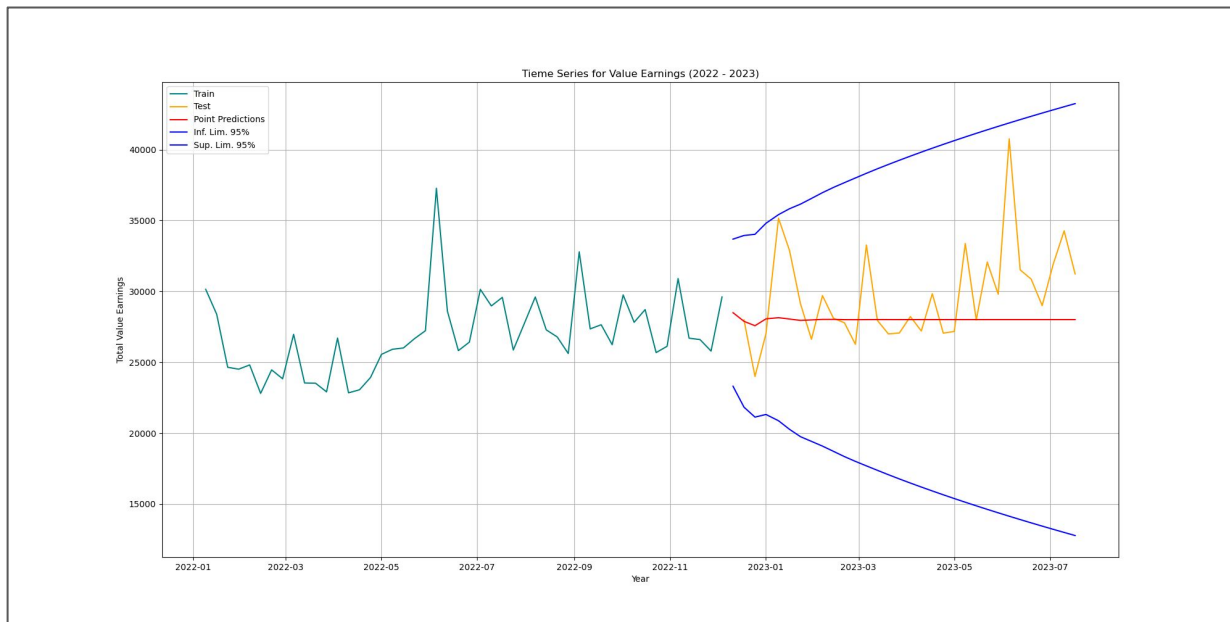
**Result** the ARIMA (3,1,1) model had the best score according to the AIC.

```
AIC for ARIMA(1,1,1) es: 1494.1131496833377
AIC for ARIMA(1,1,2) es: 1481.2716959983861
AIC for ARIMA(1,1,3) es: 1481.351866186811
AIC for ARIMA(2,1,1) es: 1492.9202777505338
AIC for ARIMA(2,1,2) es: 1494.4187461560534
AIC for ARIMA(2,1,3) es: 1483.9881518427444
AIC for ARIMA(3,1,1) es: 1474.4103954361237
AIC for ARIMA(3,1,2) es: 1475.815520782338
AIC for ARIMA(3,1,3) es: 1478.6962534869217
```

## 7. Sales Prediction with Machine Learning

The following data were calculated using the ARIMA (3,1,1) model:

- Point predictions using the test basis (orange line in the graph)
- Confidence intervals of the predictions using the test basis (blue line in the graph).



## 7. Sales Prediction with Machine Learning

A calculation of the point forecasts and confidence intervals over the next 8 weeks was also made using the above model and this was the result:

	<b>predicciones_puntuales</b>	<b>limite_inferior</b>	<b>limite_superior</b>
<b>DATE</b>			
<b>2023-05-15</b>	28010.741042	14872.568916	41148.913168
<b>2023-05-22</b>	28010.717283	14623.381723	41398.052843
<b>2023-05-29</b>	28010.689002	14378.752387	41642.625616
<b>2023-06-05</b>	28010.691888	14138.473530	41882.910247
<b>2023-06-12</b>	28010.703954	13902.293674	42119.114234
<b>2023-06-19</b>	28010.705728	13669.989900	42351.421556
<b>2023-06-26</b>	28010.701267	13441.383454	42580.019079
<b>2023-07-03</b>	28010.699481	13216.312849	42805.086113
<b>2023-07-10</b>	28010.700849	12994.618776	43026.782922
<b>2023-07-17</b>	28010.701902	12776.149778	43245.254026

## 7. Sales Prediction with Machine Learning

Finally, to evaluate the accuracy of the model, the mean square error (MSE) and the mean absolute percentage error (MAPE) were calculated:

RMSE = 3724.72

MAPE = 7.98%.

According to the ARIMA (3,1,1) model, a model with an RMSE of 3748 units and a MAPE of 8% was obtained, suggesting that the model is adequate to predict the future earnings of VANISH and LYSOL products.



## 8. Conclusions and recommendations

### Summary of key findings

In all regions, the majority of Reckitt's unit sales consist of products that have been sold a minimum number of times. This means that the same item or product was sold only a few times. However, it should be noted that region '0' was the region that had the most sales variability, having up to more than 20 unit sales per product.

Most of Reckitt's profits were shown to come from the Vanish products (pre-washers and bleaches) and these were shown to increase between 2022 and 2023. In the case of Lysol products (sanitizers), profits have been minimal and have remained stable over this time period.

## 8. Conclusions and recommendations

### Summary of key findings

Segmentation by K-means showed that there are items that generate minimal gains (clusters 1 and 4) while there are items that generate high gains (clusters 0 and 2). These items come mainly from pre-wash and bleach products (Vanish).

The ARIMA model provides a reasonably accurate prediction (MAPE of 8%) of the total value generated. However, this model also suggests a slightly decreasing trend in the future, so it is important to identify external factors that may influence the company's total earnings in these next periods.



## 8. Conclusions and recommendations

### Impact of Analytics on Business Strategy

This project demonstrated that there are certain products for which demand is high, although, unfortunately, most products have low demand and therefore low sales. The magnitude of sales and profits are similar in most of the regions, with the exception of region 0 and 2, which had a higher variability of product sales.

The results of this project will serve to prioritize products that have minimal sales (including Lysol products), as well as improve the sales strategy for these products in the regions that generated minimal profits. Some questions that could help to improve the strategy are:

- Why are some products sold infrequently?
- What is the frequency with which the different products offered by Reckitt are consumed (mainly those with low sales)?
- Are the products with low sales meeting customer needs?
- Is the result of low sales because the customer is not interested in the product or because our competitors offer something better?
- Which products sell more each season?

## 8. Conclusions and recommendations

### Recommendations

It is recommended that priority be given to regions with low sales and minimal profits, which in this case were regions 2, 3 and 6. Sales could be improved by determining which products the population of each region considers suitable for consumption and which are not of interest to them.

Products or items with minimal sales should be upgraded or replaced by products that are more attractive to the population. While Vanish brand prewashers and bleaches showed acceptable sales, Lysol brand sanitizers did not. This means that consumers are not interested in the product, so it would be appropriate to replace or upgrade it.

Study the external factors in each region that could influence product sales in particular periods of time. This is in order to prevent declines in total profits (as predicted by the ARIMA point forecasts).