

Proyecto Empresa Aliada: Entregable 4

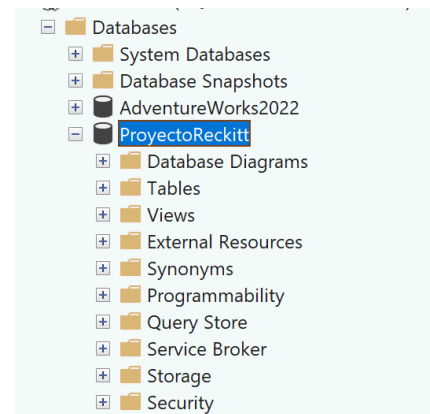
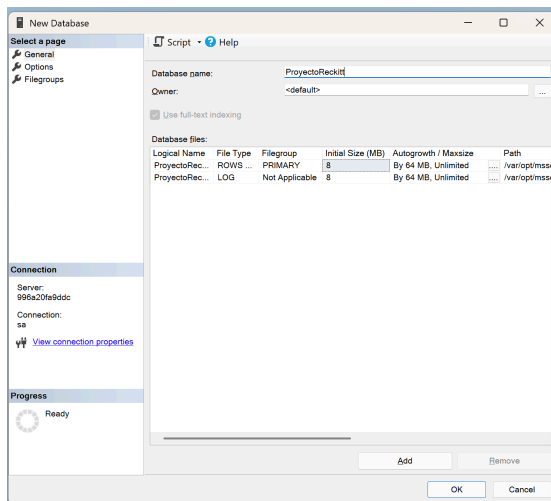
Consultas Avanzadas con SQL

En esta fase del proyecto se implementó una base de datos en **SQL Server** para **almacenar y analizar** los datos de ventas proporcionados por la empresa aliada. El modelo de datos está compuesto por **cinco tablas**: (DIM_CATEGORY, DIM_PRODUCT, DIM_SEGMENT, DIM_CALENDAR y FACT_SALES).

El objetivo principal fue estructurar la información para realizar consultas que permitan obtener insights sobre el comportamiento de las ventas por categoría, región y periodo de tiempo.

A continuación se detallan los pasos realizados:

1. **Creación de la base de datos:** La base de datos fue creada por medio del asistente de SSMS, dándole como nombre “ProyectoReckitt”.



2. **Creación de la estructura de las tablas:** Por medio de código (query) sql, se definió la estructura de cada una de las tablas, tratando de respetar la estructura de los datos compartidos por la empresa.

```

1 CREATE TABLE FACT_SALES (
2     WEEK VARCHAR(50),
3     ITEM_CODE VARCHAR(50),
4     TOTAL_UNIT_SALES DECIMAL(10,4),
5     TOTAL_VALUE_SALES DECIMAL(10,4),
6     TOTAL_UNIT_AVG_WEEKLY_SALES DECIMAL(10,4),
7     REGION VARCHAR(255)
8 );
9
10 CREATE TABLE DIM_CALENDAR(
11     WEEK VARCHAR(50) PRIMARY KEY,
12     YEAR SMALLINT,
13     MONTH TINYINT,
14     WEEK_NUMBER TINYINT,
15     DATE DATE
16 );
17
18 CREATE TABLE DIM_CATEGORY(
19     ID_CATEGORY TINYINT PRIMARY KEY,
20     CATEGORY VARCHAR(50)
21 );
22
23 CREATE TABLE DIM_PRODUCT(
24     MANUFACTURER VARCHAR(100),
25     BRAND VARCHAR(100),
26     ITEM VARCHAR(255) PRIMARY KEY,
27     ITEM_DESCRIPTION VARCHAR(255),
28     CATEGORY TINYINT,
29     FORMAT VARCHAR(50),
30     ATTR1 VARCHAR(100),
31     ATTR2 VARCHAR(100),
32     ATTR3 VARCHAR(100),
33 );
34
35 CREATE TABLE DIM_SEGMENT (
36     CATEGORY TINYINT,
37     ATTR1 VARCHAR(100),
38     ATTR2 VARCHAR(100),
39     ATTR3 VARCHAR(100),
40     FORMAT VARCHAR(50),
41     SEGMENT VARCHAR(50)
42 );
  
```

3. **Importación de los datos:** Para este paso, cabe destacar que en mi entorno, SQL server se encuentra corriendo en un contenedor de docker, y se esta ejecutando SMSS en una maquina virtual de “Parallels” en MacOS, por lo que:

- Los datos se cargaron al contenedor de docker.
- Se realizo conversión a CSV para los archivos XLSX.
- Se les dio un formato estándar a los datos, eliminando comillas dobles, y separadores que pudieran causar problemas en la carga.
- Y finalmente se cargaron los datos por medio de diversas queries.

```
$ head -5 /usr/ProyectoReckittSQL/FACT_SALES.csv
"WEEK","ITEM_CODE","TOTAL_UNIT_SALES","TOTAL_VALUE_SALES","TOTAL_UNIT_AVG_WEEKLY_SALES","REGION"
"34-22","7501058792808BP2","0.0060","0.1390","1.0000",TOTAL AUTOS AREA 5
"34-22","7501058715883","0.4870","116.5190","2.9160",TOTAL AUTOS AREA 5
"34-22","7702626213774","1.3910","68.4530","5.1710",TOTAL AUTOS AREA 5
"34-22","7501058716422","0.0220","1.4810","1.8330",TOTAL AUTOS AREA 5
$
```

```
sqlserver
996a20fa9ddc mssql/server:2022-latest
1433.1433

Logs Inspect Bind mounts Exec Files Stats

$ ls
DIM_CALENDAR.xlsx DIM_CATEGORY.csv DIM_PRODUCT.xlsx DIM_SEGMENT.xlsx FACT_SALES.csv
$ head -5 /tmp/DIM_CALENDAR.csv
head: cannot open '/tmp/DIM_CALENDAR.csv' for reading: No such file or directory
$ ls
DIM_CALENDAR.csv DIM_CATEGORY.csv DIM_PRODUCT.csv DIM_SEGMENT.csv FACT_SALES.csv
$ head -5 DIM_CALENDAR.csv
WEEK,YEAR,MONTH,WEEK_NUMBER,DATE
01-21,2021,1,1,10/01/2021
02-21,2021,1,2,17/01/2021
03-21,2021,1,3,24/01/2021
04-21,2021,1,4,31/01/2021
$
```

```
53  /* ***** Importacion de Datos ***** */
54
55  BULK INSERT FACT_SALES
56  FROM '/tmp/FACT_SALES.csv'
57  WITH (
58    FIELDTERMINATOR = ',',
59    ROWTERMINATOR = '\n',
60    FIRSTROW = 2
61  );
62
63  BULK INSERT DIM_CATEGORY
64  FROM '/tmp/DIM_CATEGORY.csv'
65  WITH (
66    FIELDTERMINATOR = ',',
67    ROWTERMINATOR = '\n',
68    FIRSTROW = 2
69  );
70
71  BULK INSERT DIM_CALENDAR
72  FROM '/tmp/DIM_CALENDAR.csv'
73  WITH (
74    FIELDTERMINATOR = ',',
75    ROWTERMINATOR = '\n',
76    FIRSTROW = 2
77  );
78
79  BULK INSERT DIM_PRODUCT
80  FROM '/usr/ProyectoReckittSQL/DIM_PRODUCT.csv'
81  WITH (
82    FIELDTERMINATOR = ',',
83    ROWTERMINATOR = '\n',
84    FIRSTROW = 2
85  );
86
87
```

4. **Comprobación de datos:** Se realizaron consultas básicas para comprobar que los datos se cargaron correctamente.

```
-- Verificación de los primeros 5 registros
101 SELECT TOP 5 * FROM FACT_SALES
102
103 SELECT TOP 5 * FROM DIM_CALENDAR
104
105 SELECT TOP 5 * FROM DIM_CATEGORY
106
107 SELECT TOP 5 * FROM DIM_PRODUCT
108
109 SELECT TOP 5 * FROM DIM_SEGMENT
110
```

| WEEK | ITEM_CODE | TOTAL_UNIT_SALES | TOTAL_VALUE_SALES | TOTAL_UNIT_AVG_WEEKLY_SALES | REGION |
|-------|------------------|------------------|-------------------|-----------------------------|--------------------|
| 34-22 | 7501058792808BP2 | 0.0060 | 0.1390 | 1.0000 | TOTAL AUTOS AREA 5 |
| 34-22 | 7501058715883 | 0.4870 | 116.5190 | 2.9160 | TOTAL AUTOS AREA 5 |
| 34-22 | 7702626213774 | 1.3910 | 68.4530 | 5.1710 | TOTAL AUTOS AREA 5 |
| 34-22 | 7501058716422 | 0.0220 | 1.4810 | 1.8330 | TOTAL AUTOS AREA 5 |
| 34-22 | 7501058794353 | 2.0370 | 182.8390 | 5.3750 | TOTAL AUTOS AREA 5 |

| WEEK | YEAR | MONTH | WEEK_NUMBER | DATE |
|-------|------|-------|-------------|------------|
| 01-21 | 2021 | 1 | 1 | 2021-01-10 |
| 01-22 | 2022 | 1 | 1 | 2022-01-09 |
| 01-23 | 2023 | 1 | 1 | 2023-01-09 |
| 02-21 | 2021 | 1 | 2 | 2021-01-17 |
| 02-22 | 2022 | 1 | 2 | 2022-01-16 |

| 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 |
|--------------|----------|---------------|--|----------|---------|--------|--------|-------------|
| INDS ALLEN | CLORALEX | 0000750000002 | CLORALEX EL RENDIDOR BOT PLAST 250ML NAL 00000... | 1 | LIQUIDO | COLORO | COLORO | NO DEFINIDO |
| INDS ALLEN | CLORALEX | 0000750000006 | CLORALEX EL RENDIDOR BOT PLAST 500ML NAL 00000... | 1 | LIQUIDO | COLORO | COLORO | NO DEFINIDO |
| INDS ALLEN | CLORALEX | 0000750000015 | CLORALEX EL RENDIDOR BOT PLAST 950ML NAL 00000... | 1 | LIQUIDO | COLORO | COLORO | NO DEFINIDO |
| INDS ALLEN | CLORALEX | 0000750000021 | CLORALEX EL RENDIDOR BOT PLAST 2000ML NAL 00000... | 1 | LIQUIDO | COLORO | COLORO | NO DEFINIDO |
| INDS ALLEN | CLORALEX | 0000750000039 | CLORALEX EL RENDIDOR BOT PLAST 3750ML NAL 00000... | 1 | LIQUIDO | COLORO | COLORO | NO DEFINIDO |

| CATEGORY | ATTR1 | ATTR2 | ATTR3 | FORMAT | SEGMENT |
|----------|--------|--------|--------|---------|---------|
| 1 | COLORO | COLORO | COLORO | LIQUIDO | BLEACH |
| 1 | COLORO | COLORO | COLORO | LIQUIDO | BLEACH |
| 1 | COLORO | COLORO | COLORO | LIQUIDO | BLEACH |
| 1 | COLORO | COLORO | COLORO | LIQUIDO | BLEACH |
| 1 | COLORO | COLORO | COLORO | LIQUIDO | BLEACH |

```
-- Total de Registros por Tabla
107
108 SELECT 'FACT_SALES' AS Tabla, COUNT(*) AS Registros FROM FACT_SALES
109
110 SELECT 'DIM_CALENDAR', COUNT(*) FROM DIM_CALENDAR
111
112 SELECT 'DIM_CATEGORY', COUNT(*) FROM DIM_CATEGORY
113
114 SELECT 'DIM_PRODUCT', COUNT(*) FROM DIM_PRODUCT
115
116 SELECT 'DIM_SEGMENT', COUNT(*) FROM DIM_SEGMENT
117
118
```

| Tabla | Registros |
|----------------|-----------|
| 1 FACT_SALES | 122002 |
| 2 DIM_CALENDAR | 156 |
| 3 DIM_CATEGORY | 5 |
| 4 DIM_PRODUCT | 505 |
| 5 DIM_SEGMENT | 53 |

5. **Insights clave de los datos:** Como ultimo paso, se realizaron consultas para obtener una perspectiva clave de los datos. Como análisis, se obtuvieron las ventas totales desglosadas por region y las semanas correspondientes de venta. También, las semanas con mas venta por region, las ventas totales por categoría y region. Como análisis extra, se calculo el precio aproximado por producto para una mejor comprensión de los valores de TotalSales, el cual nos parece indicar que los datos vienen de un sistema que registra ventas fraccionadas (por ejemplo, ventas por peso, o promedios semanales).

```

119 -- Ventas Totales Desglosadas por Region y Semanas de Venta
120 WITH Sales AS (
121     SELECT
122         ITEM_CODE,
123         SUM(TOTAL_UNIT_SALES) AS TotalSales,
124         REGION,
125         WEEK
126     FROM FACT_SALES
127     GROUP BY ITEM_CODE, REGION, WEEK
128 )
129 SELECT dp.ITEM_DESCRIPTION,
130        dc.CATEGORY,
131        fs.REGION,
132        fs.WEEK,
133        s.TotalSales
134 FROM DIM_PRODUCT dp
135 INNER JOIN Sales s ON dp.ITEM = s.ITEM_CODE
136 INNER JOIN DIM_CATEGORY dc ON dp.CATEGORY = dc.ID_CATEGORY
137 ORDER BY s.REGION, TotalSales DESC
138
139

```

75 % No issues found

| ITEM_DESCRIPTION | CATEGORY | REGION | WEEK | TotalSales |
|--|----------------------------|--------------------|-------|------------|
| CLORELEX EL RENDIDOR BOT PLAST. 950ML NAL. 000007... | FABRIC TREATMENT and SANIT | TOTAL AUTOS AREA 1 | 22-22 | 71.3870 |
| CLORELEX EL RENDIDOR BOT PLAST. 950ML NAL. 000007... | FABRIC TREATMENT and SANIT | TOTAL AUTOS AREA 1 | 39-22 | 68.6620 |
| CLORELEX EL RENDIDOR BOT PLAST. 950ML NAL. 000007... | FABRIC TREATMENT and SANIT | TOTAL AUTOS AREA 1 | 23-22 | 66.8830 |
| CLORELEX EL RENDIDOR BOT PLAST. 950ML NAL. 000007... | FABRIC TREATMENT and SANIT | TOTAL AUTOS AREA 1 | 38-22 | 65.8840 |
| CLORELEX EL RENDIDOR BOT PLAST. 950ML NAL. 000007... | FABRIC TREATMENT and SANIT | TOTAL AUTOS AREA 1 | 38-22 | 64.6770 |
| CLORELEX EL RENDIDOR BOT PLAST. 950ML NAL. 000007... | FABRIC TREATMENT and SANIT | TOTAL AUTOS AREA 1 | 22-23 | 64.4470 |
| CLORELEX EL RENDIDOR BOT PLAST. 950ML NAL. 000007... | FABRIC TREATMENT and SANIT | TOTAL AUTOS AREA 1 | 35-22 | 63.9840 |
| CLORELEX EL RENDIDOR BOT PLAST. 950ML NAL. 000007... | FABRIC TREATMENT and SANIT | TOTAL AUTOS AREA 1 | 26-22 | 63.5500 |
| CLORELEX EL RENDIDOR BOT PLAST. 950ML NAL. 000007... | FABRIC TREATMENT and SANIT | TOTAL AUTOS AREA 1 | 37-22 | 60.6620 |
| CLORELEX EL RENDIDOR BOT PLAST. 2000ML NAL. 00000... | FABRIC TREATMENT and SANIT | TOTAL AUTOS AREA 1 | 22-23 | 60.5860 |
| CLORELEX EL RENDIDOR BOT PLAST. 2000ML NAL. 00000... | FABRIC TREATMENT and SANIT | TOTAL AUTOS AREA 1 | 44-22 | 60.5340 |
| CLORELEX EL RENDIDOR BOT PLAST. 950ML NAL. 000007... | FABRIC TREATMENT and SANIT | TOTAL AUTOS AREA 1 | 23-23 | 60.2730 |
| CLORELEX EL RENDIDOR BOT PLAST. 950ML NAL. 000007... | FABRIC TREATMENT and SANIT | TOTAL AUTOS AREA 1 | 33-22 | 59.9060 |
| CLORELEX EL RENDIDOR BOT PLAST. 2000ML NAL. 00000... | FABRIC TREATMENT and SANIT | TOTAL AUTOS AREA 1 | 02-23 | 59.3010 |
| CLORELEX EL RENDIDOR BOT PLAST. 950ML NAL. 000007... | FABRIC TREATMENT and SANIT | TOTAL AUTOS AREA 1 | 40-22 | 58.8520 |
| CLORELEX EL RENDIDOR BOT PLAST. 2000ML NAL. 00000... | FABRIC TREATMENT and SANIT | TOTAL AUTOS AREA 1 | 01-23 | 58.7480 |
| CLORELEX EL RENDIDOR BOT PLAST. 950ML NAL. 000007... | FABRIC TREATMENT and SANIT | TOTAL AUTOS AREA 1 | 31-22 | 58.4580 |
| CLORELEX EL RENDIDOR BOT PLAST. 2000ML NAL. 00000... | FABRIC TREATMENT and SANIT | TOTAL AUTOS AREA 1 | 01-23 | 58.1780 |

```

148 -- Semana con mas Ventas por Region
149 WITH SalesByWeekRegion AS (
150     SELECT
151         REGION,
152         SUM(TOTAL_UNIT_SALES) AS TotalSales,
153         WEEK,
154         FROM FACT_SALES
155     GROUP BY REGION, WEEK
156 )
157 TopWeekByRegion AS (
158     SELECT
159         REGION,
160         TotalSales,
161         ROW_NUMBER() OVER (PARTITION BY REGION ORDER BY TotalSales DESC) AS rn
162     FROM SalesByWeekRegion
163 )
164 SELECT dp.ITEM_DESCRIPTION,
165        dc.CATEGORY,
166        fs.REGION,
167        fs.WEEK,
168        s.TotalSales
169 FROM FACT_SALES fs
170 INNER JOIN TopWeekByRegion tw ON fs.WEEK = tw.WEEK AND fs.REGION = tw.REGION
171 INNER JOIN DIM_PRODUCT dp ON dp.ITEM = s.ITEM_CODE
172 INNER JOIN DIM_CATEGORY dc ON dp.CATEGORY = dc.ID_CATEGORY
173 WHERE tw.rn = 1
174 GROUP BY dp.ITEM_DESCRIPTION, dc.CATEGORY, fs.REGION, fs.WEEK
175 ORDER BY fs.REGION, TotalSales DESC
176
177

```

75 % No issues found

| ITEM_DESCRIPTION | CATEGORY | REGION | WEEK | TotalSales |
|---|----------------------------|--------------------|-------|------------|
| CLORELEX EL RENDIDOR BOT PLAST. 950ML NAL. 0000075000... | FABRIC TREATMENT and SANIT | TOTAL AUTOS AREA 1 | 22-22 | 71.3870 |
| CLORELEX EL RENDIDOR BOT PLAST. 2000ML NAL. 0000075000... | FABRIC TREATMENT and SANIT | TOTAL AUTOS AREA 1 | 22-22 | 53.8210 |
| ELANCATEL CONCENTRADO BOT PLAST 3750ML NAL. 750103... | FABRIC TREATMENT and SANIT | TOTAL AUTOS AREA 1 | 22-22 | 43.9770 |
| CLORELEX EL RENDIDOR BOT PLAST. 3750ML NAL. 0000075000... | FABRIC TREATMENT and SANIT | TOTAL AUTOS AREA 1 | 22-22 | 22.0830 |
| CLORELEX ULTRAGEL CLORO GEL CONCENTRADO BOT 600 ... | FABRIC TREATMENT and SANIT | TOTAL AUTOS AREA 1 | 22-22 | 20.9680 |
| CLOREX REGULAR CONCENTRADO BOTPLAST 930 ML 750107... | FABRIC TREATMENT and SANIT | TOTAL AUTOS AREA 1 | 22-22 | 17.6010 |
| CLOREX CLORO EN GEL BOT 950ML 7501025405151 | FABRIC TREATMENT and SANIT | TOTAL AUTOS AREA 1 | 22-22 | 12.6920 |
| CLOREX ULTRAGEL CLORO EN GEL CONCENTRADO 950ML... | FABRIC TREATMENT and SANIT | TOTAL AUTOS AREA 1 | 22-22 | 12.5930 |
| CLOREX EL RENDIDOR BOT PLAST. 500ML NAL. 0000075000... | FABRIC TREATMENT and SANIT | TOTAL AUTOS AREA 1 | 22-22 | 8.4630 |
| CLOREX REGULAR CONCENTRADO BOTPLAST 3800ML 750107... | FABRIC TREATMENT and SANIT | TOTAL AUTOS AREA 1 | 22-22 | 7.5140 |
| CLOREX MASCOTAS BOT PLAST 950ML NAL. 7501025411992 | FABRIC TREATMENT and SANIT | TOTAL AUTOS AREA 1 | 22-22 | 7.4970 |
| CLOREX ULTRAGEL CITRICO BOTELLA 600ML 7501025401801 | FABRIC TREATMENT and SANIT | TOTAL AUTOS AREA 1 | 22-22 | 5.3820 |
| VANISH SUPER BARRA QUITA MANCHAS 80GR 7501058714996 | FABRIC TREATMENT and SANIT | TOTAL AUTOS AREA 1 | 22-22 | 5.2280 |
| VANISH SUPER BARRA QUITA MANCHAS 80GR 7501058... | FABRIC TREATMENT and SANIT | TOTAL AUTOS AREA 1 | 22-22 | 5.1290 |
| CLOREX ULTRAGEL CITRICO BOTELLA 950ML 7501025401795 | FABRIC TREATMENT and SANIT | TOTAL AUTOS AREA 1 | 22-22 | 4.8880 |
| CLOREX AROMA CLORO LIQUIDO FLORAL BP 950ML 750102... | FABRIC TREATMENT and SANIT | TOTAL AUTOS AREA 1 | 22-22 | 4.8520 |
| CLOREX REGULAR CONCENTRADO BOTPLAST 1800ML 750107... | FABRIC TREATMENT and SANIT | TOTAL AUTOS AREA 1 | 22-22 | 4.6530 |
| CLOREX AROMA CLORO LIQUIDO LAVANDIA BP 950ML 75010... | FABRIC TREATMENT and SANIT | TOTAL AUTOS AREA 1 | 22-22 | 4.5970 |

```

168 -- Ventas Totales por Categoría y Región
169 SELECT dc.CATEGORY,
170        fs.REGION,
171        SUM(fs.TOTAL_UNIT_SALES) AS TotalUnits,
172        SUM(fs.TOTAL_VALUE_SALES) AS TotalValue
173 FROM FACT_SALES fs
174 INNER JOIN DIM_PRODUCT dp ON dp.ITEM = fs.ITEM_CODE
175 INNER JOIN DIM_CATEGORY dc ON dp.CATEGORY = dc.ID_CATEGORY
176 GROUP BY dc.CATEGORY, fs.REGION
177 ORDER BY dc.CATEGORY, TotalValue DESC
178
179

```

5 % No issues found

| CATEGORY | REGION | TotalUnits | TotalValue |
|----------------------------|-----------------------------|-------------|--------------|
| FABRIC TREATMENT and SANIT | TOTAL AUTOS SCANNING MEXICO | 195880.0540 | 5521429.3200 |
| FABRIC TREATMENT and SANIT | TOTAL AUTOS AREA 2 | 45319.9160 | 1188796.1500 |
| FABRIC TREATMENT and SANIT | TOTAL AUTOS AREA 5 | 36266.9000 | 1153335.5380 |
| FABRIC TREATMENT and SANIT | TOTAL AUTOS AREA 6 | 38585.2220 | 983957.5710 |
| FABRIC TREATMENT and SANIT | TOTAL AUTOS AREA 3 | 25896.0540 | 803655.3370 |
| FABRIC TREATMENT and SANIT | TOTAL AUTOS AREA 1 | 25791.9330 | 714249.9790 |
| FABRIC TREATMENT and SANIT | TOTAL AUTOS AREA 4 | 24020.1490 | 677435.9980 |

```

179 -- Verificacion de Precio Aproximado de Productos para comprension de valores de "TotalSales"
180 SELECT TOP 10
181     ITEM_CODE,
182     TOTAL_UNIT_SALES,
183     TOTAL_VALUE_SALES,
184     CASE
185         WHEN TOTAL_UNIT_SALES > 0
186         THEN TOTAL_VALUE_SALES / TOTAL_UNIT_SALES
187         ELSE 0
188     END AS PrecioImplicito
189 FROM FACT_SALES
190 WHERE TOTAL_UNIT_SALES > 0
191 ORDER BY TOTAL_UNIT_SALES DESC
192
193

```

75 % No issues found

| ITEM_CODE | TOTAL_UNIT_SALES | TOTAL_VALUE_SALES | PrecioImplicito |
|---------------|------------------|-------------------|--------------------|
| 0000075000622 | 504.6810 | 11377.2830 | 22.543513625438643 |
| 0000075000622 | 481.6090 | 10528.0890 | 21.860241399143288 |
| 0000075000622 | 446.4400 | 9867.2370 | 22.102048067846268 |
| 0000075000615 | 438.9640 | 5811.8790 | 13.239990067622621 |
| 0000075000615 | 437.0430 | 6290.0010 | 14.392178801628215 |
| 0000075000622 | 411.8590 | 9334.2370 | 22.66367130486444 |
| 0000075000615 | 406.8520 | 5261.9160 | 12.933243538190791 |
| 0000075000622 | 403.9030 | 9044.8790 | 22.393691059487055 |
| 0000075000622 | 403.4770 | 8617.1480 | 21.357222344768103 |
| 0000075000622 | 383.8170 | 8925.8700 | 23.255535841299369 |