

Proyecto Empresa Aliada: Entregable 4

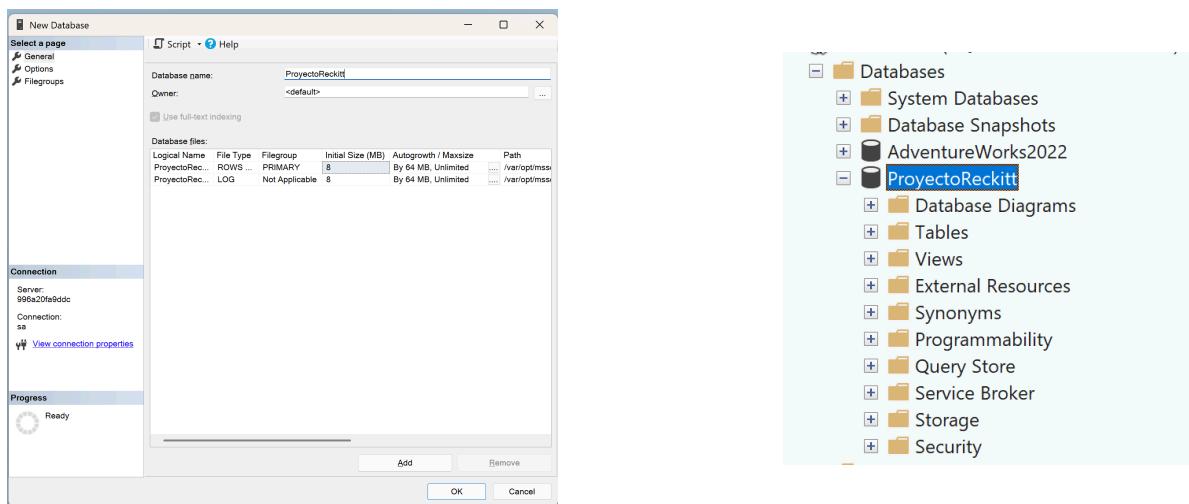
Consultas Avanzadas con SQL

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:

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



- 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.

```

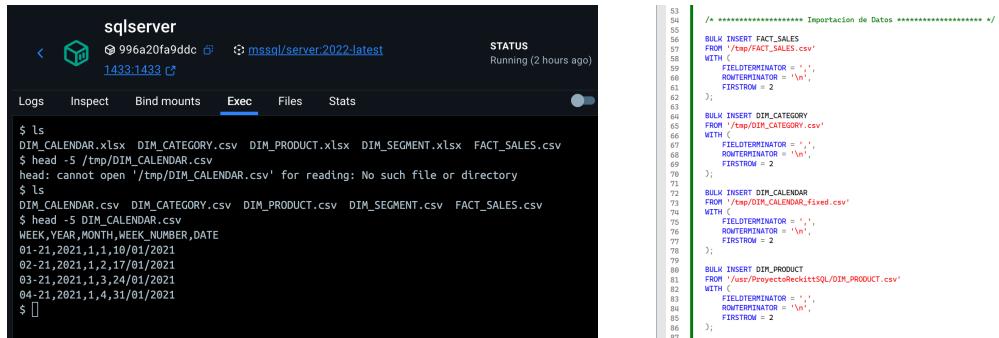
1  CREATE TABLE FACT_SALES (
2      WEEK VARCHAR(50),
3      ITEM_CODE VARCHAR(50),
4      TOTAL_UNIT_SALES DECIMAL(10, 4),
5      TOTAL_VALUE_SALES DECIMAL(10, 4),
6      TOTAL_UNIT_AVG_WEEKLY_SALES DECIMAL(10, 4),
7      REGION VARCHAR(255)
8  );
9
10
11  CREATE TABLE DIM_CALENDAR(
12      WEEK SMALLINT PRIMARY KEY,
13      YEAR TINYINT,
14      MONTH TINYINT,
15      WEEK_NUMBER TINYINT,
16      DATE_DATE
17  );
18
19  CREATE TABLE DIM_CATEGORY(
20      ID_CATEGORY TINYINT PRIMARY KEY,
21      CATEGORY VARCHAR(50)
22  );
23
24  CREATE TABLE DIM_PRODUCT(
25      MANUFACTURER VARCHAR(100),
26      BRAND VARCHAR(100),
27      ITEM VARCHAR(255) PRIMARY KEY,
28      ITEM_DESCRIPTION VARCHAR(255),
29      CATEGORY TINYINT,
30      ITEM_CODE VARCHAR(50),
31      ATTR1 VARCHAR(100),
32      ATTR2 VARCHAR(100),
33      ATTR3 VARCHAR(100),
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 conversion 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
$ []
```



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

180	-- Verificación de los primeros 5 registros
181	SELECT TOP 5 * FROM FACT_SALES
182	SELECT TOP 5 * FROM DIM_CALENDAR
183	SELECT TOP 5 * FROM DIM_CATEGORY
184	SELECT TOP 5 * FROM DIM_PRODUCT
185	SELECT TOP 5 * FROM DIM_SEGMENT
186	

75 % ▾ No issues found

Result Messages

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	7501058716422	0.0220	1.4810	1.8330	TOTAL AUTOS AREA 5

WEEK YEAR MONTH WEEK_NUMBER DATE

01-21	2021	1	1	2021-01-10
01-22	2021	1	1	2021-01-09
01-23	2021	1	1	2021-01-08
02-21	2021	1	2	2021-01-17
02-22	2021	1	2	2021-01-16

ID_CATEGORY CATEGORY

1	FABRIC CARE & SANIT
2	ARM CARE
3	LAVAVAJILLAS
4	MEGA SUPERFICIES
5	LAVATORY CARE & SHC

MANUFACTURER ITEM_CODE ITEM_DESCRIPTION CATEGORY FORMAT ATTR1 ATTR2 ATTR3

INDS ALLEN	CLORALEX 0000070000092	CLORALEX EL RENDERD BOT PLAST. 250ML NAL 00000...	CATEGORY	LIQUIDO	CLORO	CLORO	NO DEFINIDO
INDS ALLEN	CLORALEX 0000070000098	CLORALEX EL RENDERD BOT PLAST. 500ML NAL 00000...	1	LIQUIDO	CLORO	CLORO	NO DEFINIDO
INDS ALLEN	CLORALEX 0000070000015	CLORALEX EL RENDERD BOT PLAST. 950ML NAL 00000...	1	LIQUIDO	CLORO	CLORO	NO DEFINIDO
INDS ALLEN	CLORALEX 0000070000062	CLORALEX EL RENDERD BOT PLAST. 2000ML NAL 00000...	1	LIQUIDO	CLORO	CLORO	NO DEFINIDO
INDS ALLEN	CLORALEX 0000070000068	CLORALEX EL RENDERD BOT PLAST. 3750ML NAL 00000...	1	LIQUIDO	CLORO	CLORO	NO DEFINIDO

CATEGORY ATTR1 ATTR2 ATTR3 FORMAT SEGMENT

1	CLORO	CLORO	BAMBIN	LIQUIDO	BLEACH
1	CLORO	CLORO	GERMICA	LIQUIDO	BLEACH
1	CLORO	CLORO	MASCOT..	LIQUIDO	BLEACH
1	CLORO	CLORO	MULTUS..	GEL	BLEACH
1	CLORO	CLORO	MULTUS..	LIQUIDO	BLEACH

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

75 % ▾ No issues found

Result Messages

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).

-- Ventas Totales por Categoría y Región

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

5 % No issues found

Results Messages

CATEGORY	REGION	TotalUnits	TotalValue
FABRIC TREATMENT and SANIT	TOTAL AUTOS SCANNING MEXICO	195880.0540	551249.3200
FABRIC TREATMENT and SANIT	TOTAL AUTOS AREA 2	45319.9160	1188796.1500
FABRIC TREATMENT and SANIT	TOTAL AUTOS AREA 5	36266.9000	115335.5380
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

-- Verificación de Precio Aproximado de Productos para comprensión de valores de "TotalSales"

```
179
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;
```

75 % No issues found

Results Messages

ITEM_CODE	TOTAL_UNIT_SALES	TOTAL_VALUE_SALES	PrecioImplicito
1	0000075000622	504.8110	11377.2830
2	0000075000622	481.6090	10528.0890
3	0000075000622	446.4400	9867.2370
4	0000075000615	438.9640	5811.8790
5	0000075000615	437.0430	6290.0010
6	0000075000622	411.8590	9334.2370
7	0000075000615	406.8520	5261.9160
8	0000075000622	403.9030	9044.8790
9	0000075000622	402.4770	8617.1480
10	0000075000622	383.8170	8925.8700