1. Data Connection and Preparation

The solution leverages a **Star Schema** model, with `fact_sales` as the central table and three dimension tables: `dim_product`, `dim_geography`, and `dim_client_sale`.

1.1. Database Structure (DDL)

The following SQL DDL was used to establish the PostgreSQL database structure:

SQL

```
DROP TABLE IF EXISTS fact_sales CASCADE;
DROP TABLE IF EXISTS dim_product CASCADE;
-- ... (Remaining DDL for dim_geography, dim_client_sale, and fact_sales with FKs) ...
```

1.2. Data Cleaning and Normalization

A custom Python script (`AutomatedDataCleaner` and `normalize_sales_data`) was executed to process the raw `RWventas.csv` file.

- **Cleaning:** The script performed automated column mapping, data type conversions, text cleaning (e.g., standardizing product names and types, auto-correcting country based on city), and calculation of the `total` revenue field.
- **Strict Null Removal:** Crucially, the script implemented a **strict cleanup** phase, removing any row in the source data and resulting dimension/fact tables that contained a `NaN`, `Unknown`, or other null value in any column. * **Normalization:** The cleaned data was normalized into the four target CSV files (`fact_sales.csv`, `dim_product.csv`, `dim_geography.csv`, `dim_client_sale.csv`).

1.3. Data Load Method

The data was loaded into the PostgreSQL tables using a **manual import process** via a database management tool (e.g., pgAdmin, DBeaver.):

1. The DDL was executed to create the four empty tables.
2. For each table, the **"Import Data"** feature was selected by right-clicking the table name.
3. The corresponding, cleaned CSV file from the normalized_data folder was imported, following the guided steps (Browse, then Siguiente and loading the data).

1.4. Connecting Power BI to PostgreSQL

The connection in Power BI was established as follows[4]:

1. **Get Data** -> **PostgreSQL database**.
2. Enter the Server and Database name (e.g., RIWI_VENTAS).

3. Select **Data Connectivity Mode** as **Import** (or DirectQuery, based on performance requirements).
4. In the Navigator, select the four required tables: public.dim_product, public.dim_geography, public.dim_client_sale, and public.fact_sales.
5. Click **Transform Data** (Power Query Editor) for final review, ensuring data types are correct before loading.

2. DAX Measures Defined

The following DAX measures were created for analysis and visualization:

DAX Measure	**Purpose**	**Formula**
Ventas Totales	Total sales revenue.	SUM('public fact_sales'[total])
Unidades_Vendidas	Total quantity of units sold.	SUM('public fact_sales'[quantity])
Descuento_Total	Sum of total discounts applied.	SUM('public fact_sales'[discount])
Ventas_años_Anterior	Sales from the same period last year (YoY comparison).	CALCULATE([Ventas Totales], SAMEPERIODLASTYEAR('public fact_sales'[date]))
Ranking Productos	Ranks products based on Ventas Totales.	RANKX(ALL('public dim_product'[product]), [Ventas Totales], , DESC, DENSE)
Top 5 Productos	Filters total sales to only show the Top 5 ranked products.	IF([Ranking Productos] <= 5, [Ventas Totales], BLANK())

Ventas_País	Total sales calculated at the Country level, ignoring other filters in the dim_geography table.	CALCULATE([Ventas Totales], ALLEXCEPT('public dim_geography', 'public dim_geography'[country]))

3. Required Visualizations and Interactivity

The dashboards were built to meet the mandatory visualization and interactivity requirements.

3.1. Mandatory Visualizations

Requirement	Visualization Type	Data Used
Top 5 Products	**Bar Chart**	Products (dim_product) vs. Top 5 Productos (Measure) 666
Top 5 Clients	**Table / Bar Chart**	Client Types (dim_client_sale) vs. Ventas Totales (Measure) 7
Sales by Region	**Choropleth Map (Filled Map)**	Country (dim_geography) vs. Ventas Totales (Measure) 8888
Top Categories	**Donut Chart (Gráfico de anillos)**	Product Type (dim_product) vs. Ventas Totales (Measure) 10101010

3.2. Interactivity and Filters

The dashboard includes **Slicers (Segmentadores)** to enable detailed, dynamic analysis:

- **Region:** Slicer based on the **Country** field from `dim_geography`.
- **Category:** Slicer based on the **Product_Type** field from `dim_product`.
- **Date Range:** Slicer based on the **Date** field from `fact_sales`.

4. Insights and Strategic Storytelling

Based on the visualized data, the following two actionable insights and recommendations are provided:

Insight 1: Regional Sales Dominance and Focus

Observation: The analysis using the Choropleth Map and the `Ventas Totales` measure clearly shows that **spain** and **colombian** generate the highest total sales volume. Specifically, españa leads by a significant margin.

Actionable Insight: While españa is the leading market, the **Year-over-Year (YoY) Sales Comparison** reveals that colombianhas the highest growth rate, indicating a rapidly scaling market.

Strategic Recommendation: Maintain high focus and resource allocation on the spain market due to its sheer volume, but **aggressively increase investment (marketing and distribution)** in the colombian market to capitalize on its high growth trajectory and secure future market share.