Oils_Sergii Data Analysis Project

Project Overview

This project was created as part of a real client request to analyze sales data and provide key insights into product and customer behavior.
The goal was to **rank products** and calculate:

- The **total purchase amount** for each product
- The **total purchase amount per product per client**
- The **most recent sale price** for each product by client

Data Source and Preparation

The client provided a raw `.xlsx` file containing detailed sales data.
I performed the following steps to prepare the dataset:

1. **Data cleaning and formatting** — corrected structure and formats in **Google Sheets** (dates, decimals, column names, etc.)
2. **Database import** — uploaded the cleaned file into **PostgreSQL** using **DBeaver**.
3. **Data type adjustments** — fixed incorrect column types (e.g., numeric vs text, integer overflow, date parsing).

After data preparation, the table `Oils_Sergii` was ready for SQL analysis.

Analysis Process

All analysis was done using **SQL (PostgreSQL)** inside **DBeaver**.
Key operations included:

- Aggregating data by product and client
- Applying window functions (`SUM() OVER`) to calculate totals
- Using subqueries to extract the **latest sale price** per product and client

---

Final SQL Query

The final query calculates:

- `total_sum` — how much each client spent on a given product
- `total_company_sum` — total purchase amount per client
- `total_product_sum` — total purchase amount per product across all clients
- `last_price` — the last known sale price for that product and client

All results are rounded to two decimal places.

Tools & Technologies

- **Google Sheets** — data cleaning & formatting
- **PostgreSQL** — data storage & analysis
- **DBeaver** — SQL development environment
- **SQL** — main analysis language

Example Use Cases

This analysis can be extended to:

- Identify top-performing products
- Compare client purchasing behavior
- Track price changes over time
- Build dashboards (e.g., in Power BI or Tableau) for business monitoring

---

How to Reproduce

1. Import the cleaned dataset into a PostgreSQL database
2. Execute the final SQL script in DBeaver
3. Export or visualize the results

Key Takeaways

- Proper data cleaning before database import saves significant time during analysis
- Handling numeric and date formats correctly is critical when importing from Excel
- SQL window functions are powerful for comparative and ranking analysis

---

Author: Andrii Drozda
Project: Oils_Sergii Data Analysis (PostgreSQL, SQL, DBeaver)