# **Analysis of Superstore Dataset**

## **Project Overview**

This project involves analyzing the sales data of a Superstore to extract useful insights and perform specific analytical operations. The dataset provides information on customer orders, product sales, discounts, and shipping details.

### **Dataset**

* **Source**:[superstore\_data.csv](https://1drv.ms/x/c/a9c8f65c99bc68ba/EfYQvXMsXZNDgiBNvMocLdYB-3QZMyhlY1_HnLFH2bmNng?e=ecWeny)
* **Description**: The dataset includes sales records from a superstore, with various attributes such as customer information, sales, discounts, product categories, and shipping details.

## **Operations Performed**

The project performs the following analyses on the dataset:

1. **Top Sales per Customer**: Identify the customers with the highest sales.
2. **Average Discount by Product Category**: Calculate the average discount provided in each product category.
3. **Top 5 Cities by Total Sales**: Determine the cities generating the highest total sales.
4. **Products with Sales Greater than $5000**: Filter and display products with sales greater than $5000.
5. **Number of Orders per Shipping Mode**: Count the total number of orders for each shipping method.

## **Technologies Used**

* **Database**: PostgreSQL
* **Interface**: Valentina Studio (for database management)
* **Programming Language**: SQL
* **Tools**: Valentina Studio for interacting with PostgreSQL

## **Installation and Setup**

### **Prerequisites**

* PostgreSQL installed on your system.
* Valentina Studio or pgAdmin for managing the database.
* The Superstore dataset in .csv format.

### **Step-by-Step Instructions**

1. **Set up PostgreSQL**:

Install PostgreSQL and create a new database:  
sql  
Copy code  
CREATE DATABASE superstore\_analysis;

1. **Create Tables**:

Create the table in PostgreSQL to store the dataset:  
  
CREATE TABLE superstore\_data (

"Row ID" INTEGER,

"Order ID" TEXT,

"Order Date" DATE,

"Ship Date" DATE,

"Ship Mode" TEXT,

"Customer ID" TEXT,

"Customer Name" TEXT,

"Segment" TEXT,

"Country" TEXT,

"City" TEXT,

"State" TEXT,

"Postal Code" TEXT,

"Region" TEXT,

"Product ID" TEXT,

"Category" TEXT,

"Sub-Category" TEXT,

"Product Name" TEXT,

"Sales" NUMERIC,

"Quantity" INTEGER,

"Discount" NUMERIC,

"Profit" NUMERIC

);

1. **Import the Data**:

Use the COPY command in PostgreSQL to load the dataset:

COPY superstore\_data FROM '/path/to/superstore\_data.csv'

DELIMITER ','

CSV HEADER;

## **SQL Queries**

### **1. Top Sales per Customer**

SELECT "Customer Name", SUM("Sales") AS total\_sales

FROM superstore\_data

GROUP BY "Customer Name"

ORDER BY total\_sales DESC;

### **2. Average Discount by Product Category**

SELECT "Category", AVG("Discount") AS avg\_discount

FROM superstore\_data

GROUP BY "Category";

### **3. Top 5 Cities by Total Sales**

SELECT "City", SUM("Sales") AS total\_sales

FROM superstore\_data

GROUP BY "City"

ORDER BY total\_sales DESC

LIMIT 5;

### **4. Products with Sales Greater than $5000**

SELECT "Product Name", "Sales"

FROM superstore\_data

WHERE "Sales" > 5000;

### **5. Number of Orders per Shipping Mode**

SELECT "Ship Mode", COUNT(\*) AS total\_orders

FROM superstore\_data

GROUP BY "Ship Mode";

## **Facing Issues**

During the project, I encountered several issues while working with PostgreSQL and Valentina Studio. Below are the challenges faced and the solutions implemented:

### **1. Issue: File Not Found Error While Importing CSV in PostgreSQL**

**Problem**: While trying to import the dataset using the COPY command in PostgreSQL, I encountered the error:  
  
ERROR: could not open file "file:///C:/PostgreSQLImport/superstore\_data.csv" for reading: No such file or directory

* **Cause**: PostgreSQL couldn’t access the file because it was located on my local machine. PostgreSQL runs in a different environment, so the server couldn’t access my local file system directly.
* **Solution**:
  + Moved the CSV file to the PostgreSQL data directory or a directory accessible by PostgreSQL.

Used the absolute path to ensure correct file location:  
  
COPY superstore\_data FROM '/path/to/your/superstore\_data.csv'

DELIMITER ','

CSV HEADER;

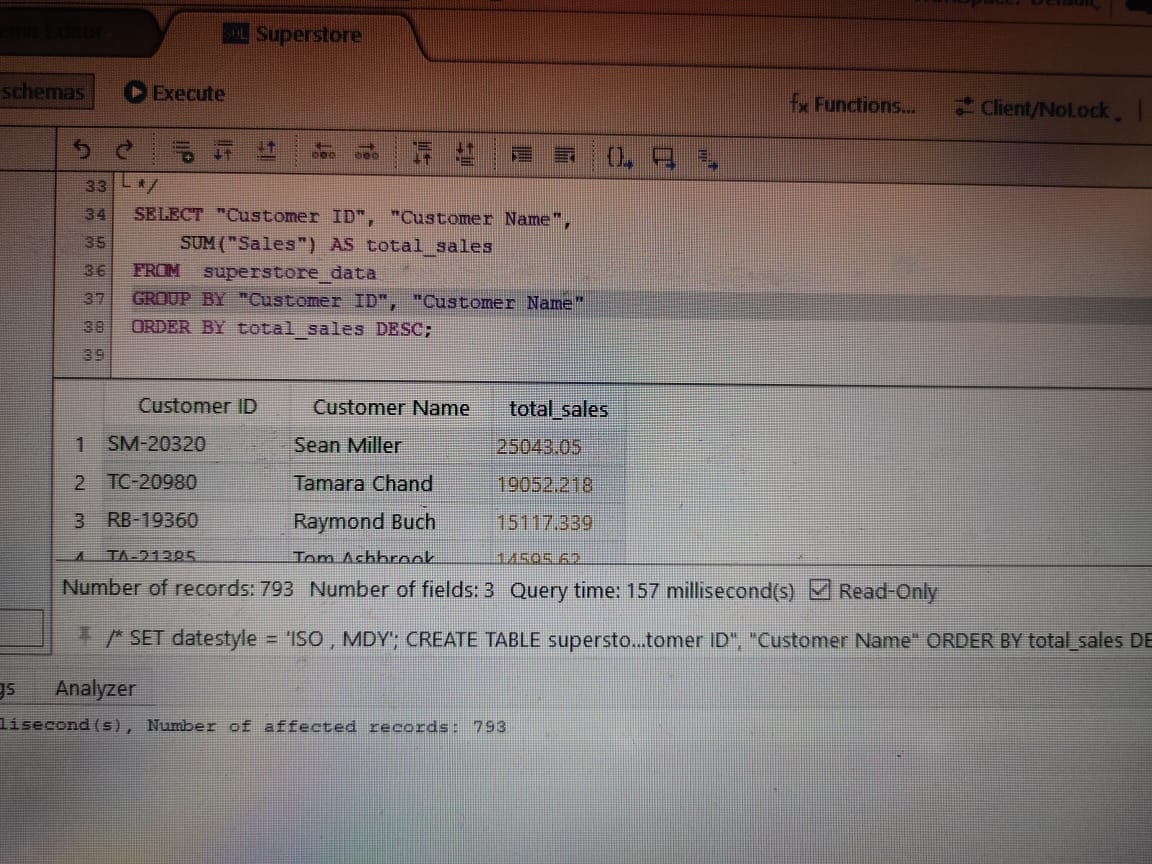
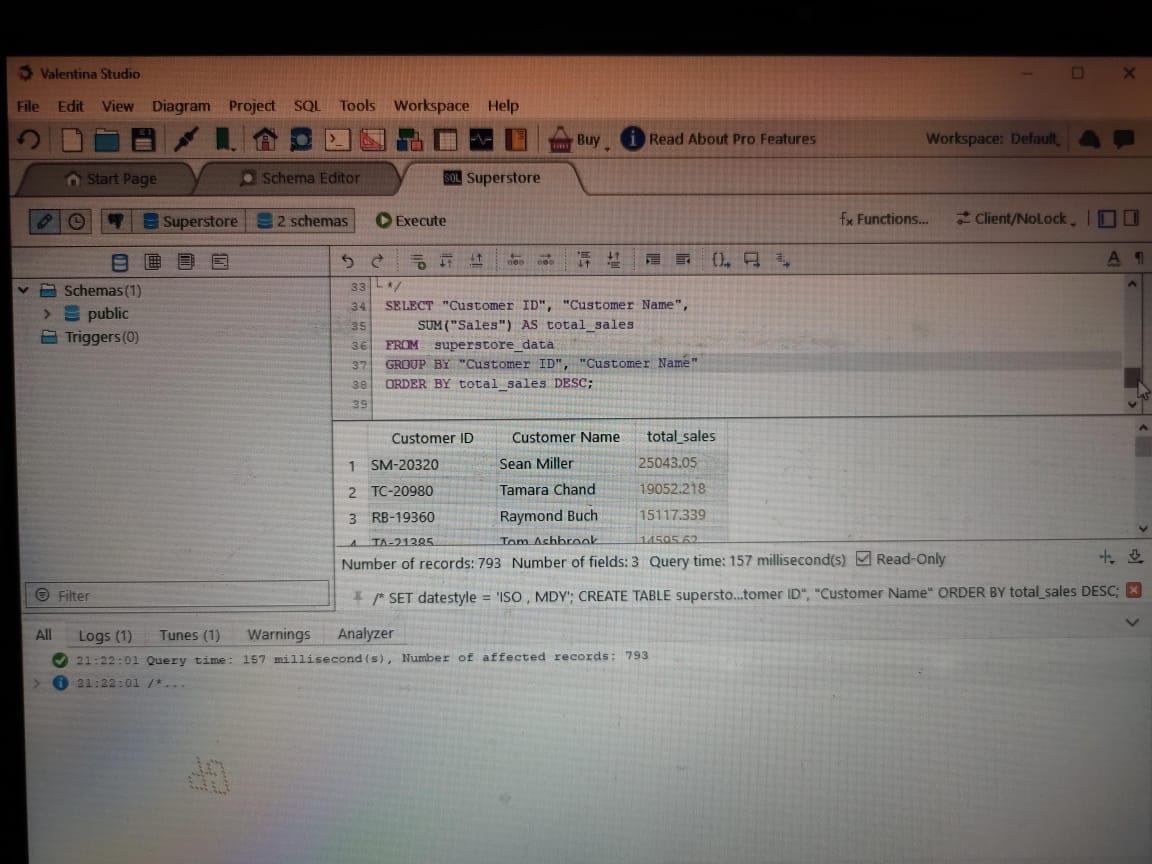
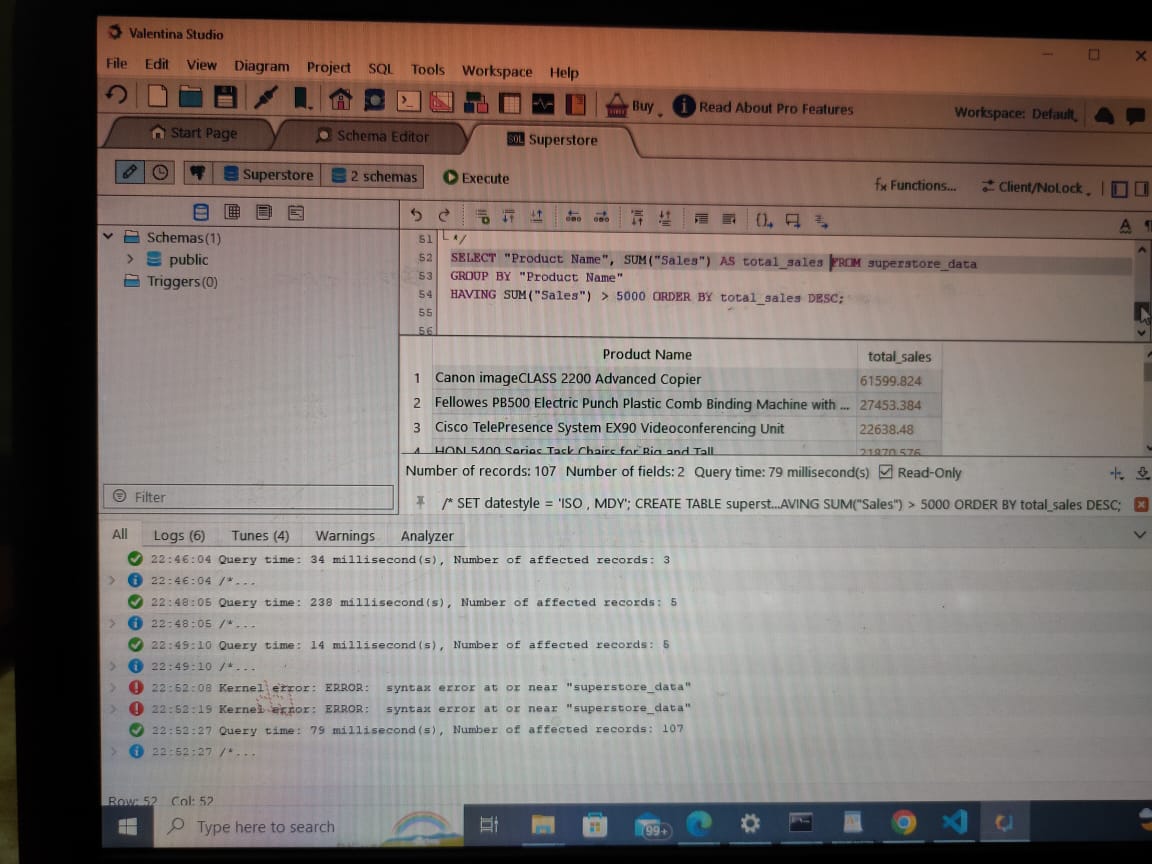
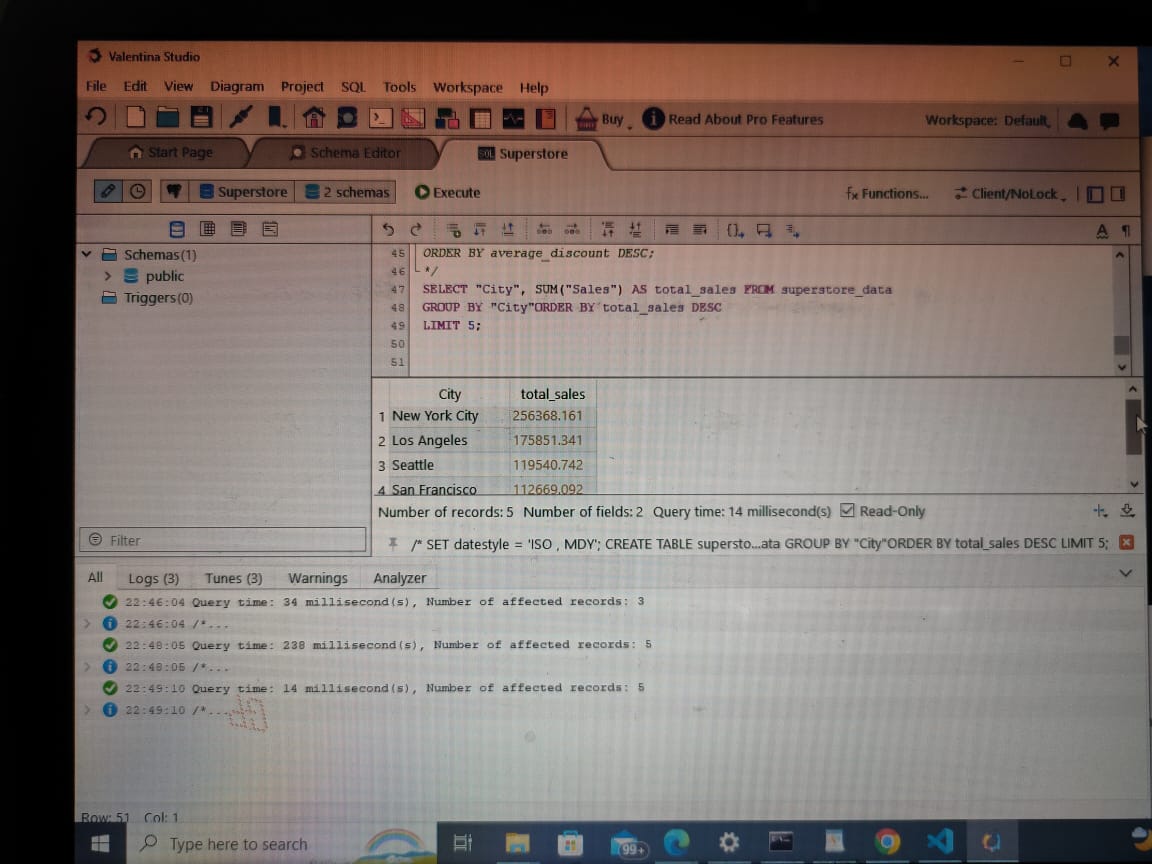
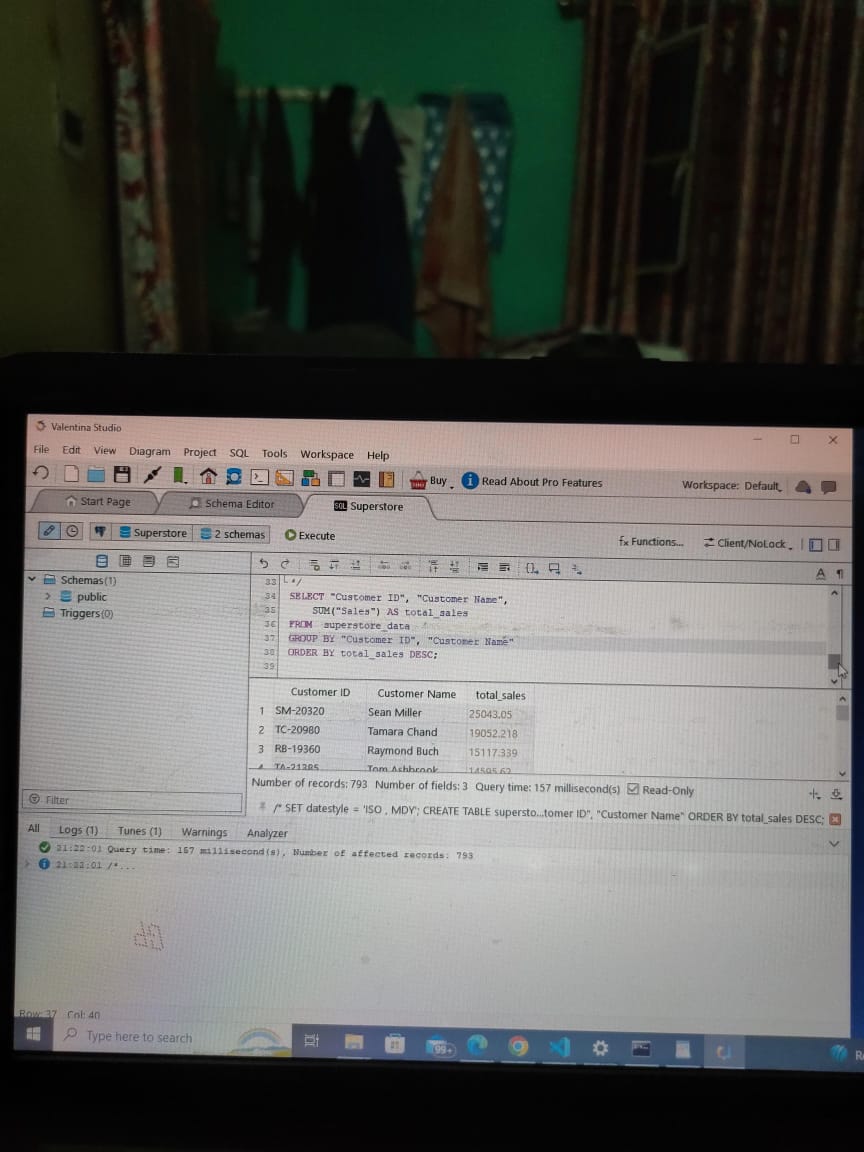
### **2. Issue: Setting Date Format in PostgreSQL**

* **Problem**: The date format in the dataset was not matching PostgreSQL’s default YYYY-MM-DD format, leading to errors when inserting the data.

**Solution**: Set the correct date style for PostgreSQL using:  
  
SET datestyle = 'ISO, DMY';

* This ensures that PostgreSQL interprets the date format as DD/MM/YYYY.

## **PROOF OF WORK**



CODE:

/\*

SET datestyle = 'ISO , MDY';

CREATE TABLE superstore\_data

(

"Row ID" INTEGER,

"Order ID" TEXT,

"Order Date" DATE,

"Ship Date" DATE,

"Ship Mode" TEXT,

"Customer ID" TEXT,

"Customer Name" TEXT,

"Segment" TEXT,

"Country" TEXT,

"City" TEXT,

"State" TEXT,

"Postal Code" TEXT,

"Region" TEXT,

"Product ID" TEXT,

"Category" TEXT,

"Sub-Category" TEXT,

"Product Name" TEXT,

"Sales" NUMERIC,

"Quantity" INTEGER,

"Discount" NUMERIC,

"Profit" NUMERIC

);

COPY superstore\_data ("Row ID", "Order ID", "Order Date", "Ship Date", "Ship Mode", "Customer ID", "Customer Name", "Segment", "Country", "City", "State", "Postal Code", "Region", "Product ID", "Category", "Sub-Category", "Product Name", "Sales", "Quantity", "Discount", "Profit")

FROM 'C:/PostgreSQLImport/superstore\_data.csv'

DELIMITER ','

CSV HEADER;

\*/

SELECT "Customer ID", "Customer Name", SUM("Sales") AS total\_sales

FROM superstore\_data

GROUP BY "Customer ID", "Customer Name"

ORDER BY total\_sales DESC;

SELECT "Category", AVG("Discount") AS average\_discount

FROM superstore\_data

GROUP BY "Category"

ORDER BY average\_discount DESC;

SELECT "City", SUM("Sales") AS total\_sales FROM superstore\_data

GROUP BY "City"ORDER BY total\_sales DESC

LIMIT 5;

SELECT "Product Name", SUM("Sales") AS total\_sales FROM superstore\_data

GROUP BY "Product Name"

HAVING SUM("Sales") > 5000 ORDER BY total\_sales DESC;

SELECT "Ship Mode", COUNT("Order ID") AS number\_of\_orders FROM superstore\_data

GROUP BY "Ship Mode" ORDER BY number\_of\_orders DESC;

## 

#### 