
Task Summary and Outcome

This project successfully analyzed the e-commerce dataset spanning 2009 to 2011 to determine key sales trends. The primary goal was to calculate the **Total Monthly Revenue** and **Total Monthly Order Volume** using advanced SQL aggregation techniques.

Key Steps Completed:

1. **Data Consolidation & Cleaning:** Two separate retail CSV files were combined, and cleaning was performed to remove returns and cancellations (rows with Quantity≤0 or Price≤0), resulting in a final dataset of **1,041,671 records**.
2. **Database Setup:** A database table named **online_sales** was created with the correct schema (VARCHAR for Invoice, TIMESTAMP for InvoiceDate, etc.).
3. **Data Import:** The consolidated CSV was imported into the MySQL/PostgreSQL table using the COPY command (or LOAD DATA INFILE), ensuring all data was available for querying.
4. **SQL Aggregation:** A single SQL query was executed to calculate the required metrics.

Final SQL Analysis Query

The final query demonstrates the use of core aggregation functions for trend analysis:

SQL

```
SELECT
    EXTRACT(YEAR FROM InvoiceDate) AS sales_year,
    EXTRACT(MONTH FROM InvoiceDate) AS sales_month,

    -- Calculation for Monthly Revenue
    SUM(Quantity * Price) AS total_revenue,

    -- Calculation for Monthly Order Volume
```

```
    COUNT(DISTINCT Invoice) AS order_volume

FROM
    online_sales

GROUP BY
    sales_year,
    sales_month

ORDER BY
    sales_year,
    sales_month;
```

Core SQL Concepts Used:

Concept	Purpose in Task	Formula Used
Grouping	Breaks the data into 26 distinct monthly periods (from Dec-2009 to Dec-2011) for aggregation.	GROUP BY sales_year, sales_month
Revenue	Calculates the sum of all money generated from positive sales transactions for the month.	SUM(Quantity∗Price)
Order Volume	Counts the unique number of distinct transactions, ignoring multiple line items within a single order.	COUNT(DISTINCT Invoice)