# Optimized Batch & Stream Data Processing for Enhanced Inventory Management

*Submitted in partial fulfilment of the requirements for the degree of*

**Post Graduate Diploma in Data Engineering**

by

**Sudarshan P (G23AI1046)**

**Utkarsh Gupta (G23AI1048)**

**Under the guidance of**

**Dr. Pradip Samal**

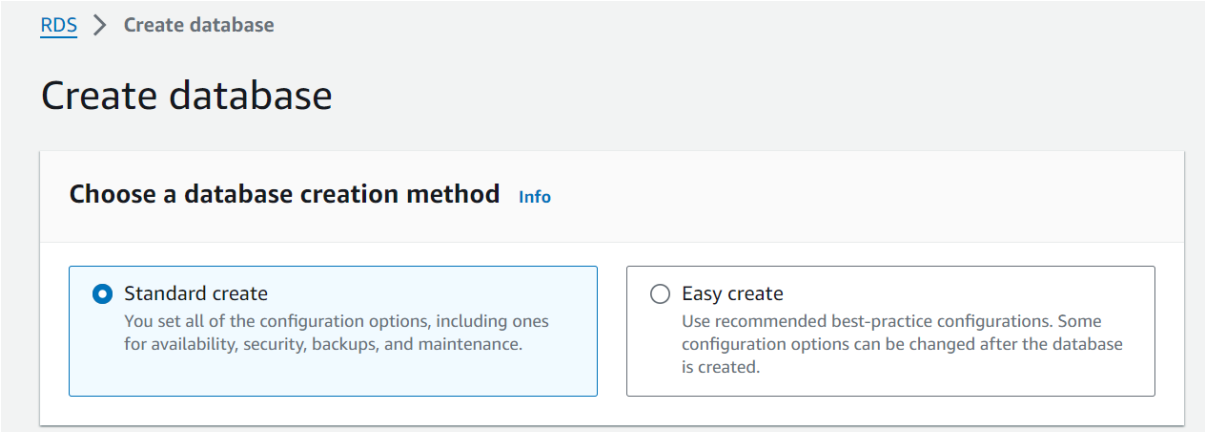
**IIT Jodhpur**

**Indian Institute of Technology Jodhpur**
**Advance Data Engineering in Cloud**
**Trimester-3 (July 2024)**

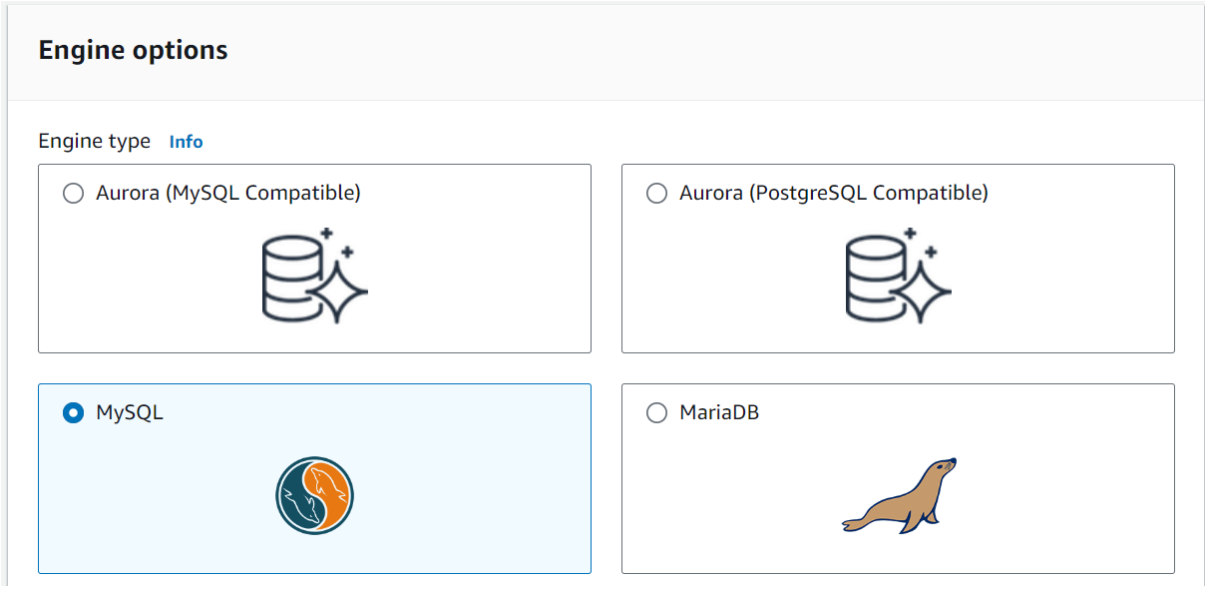
# Assignment – 3

## Part 5: Create RDS instance

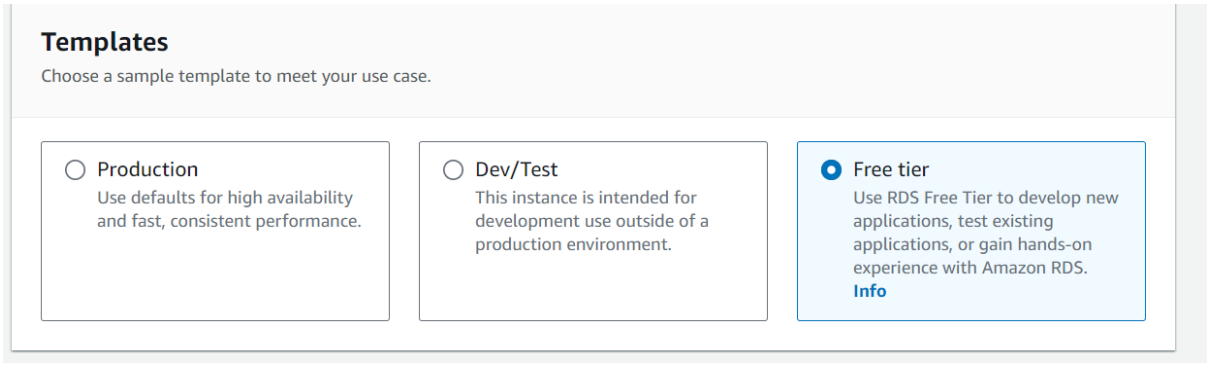
1. Go to RDS console and Create DB instance



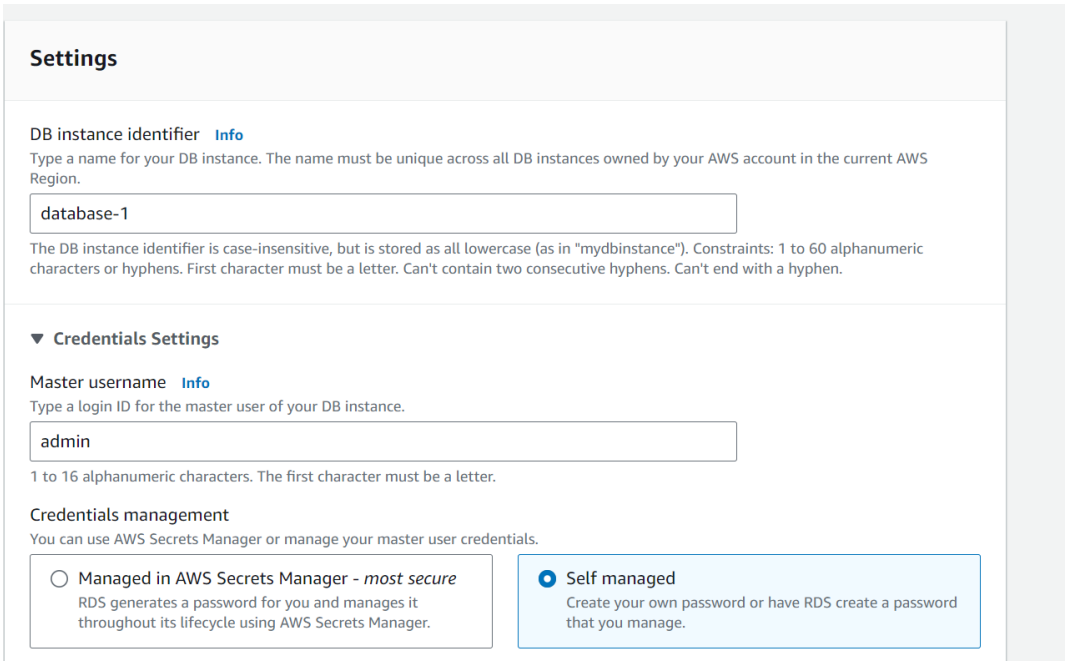

2. Choose MySQL

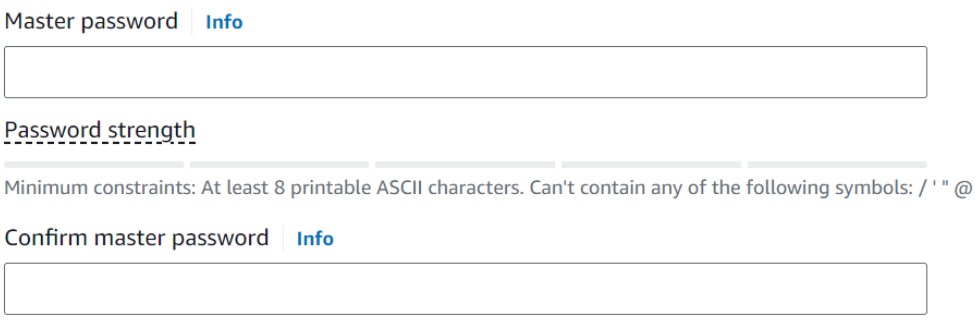


3. Choose Free Tier

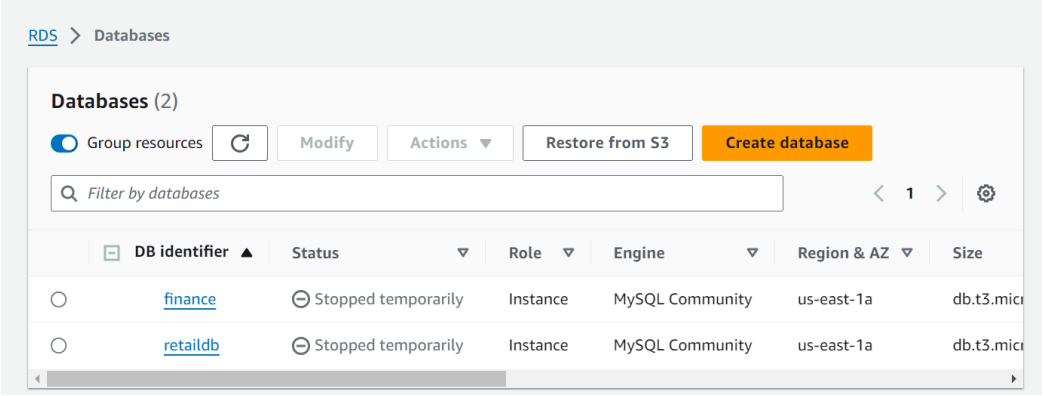



4. Give DB instance and DB username and password



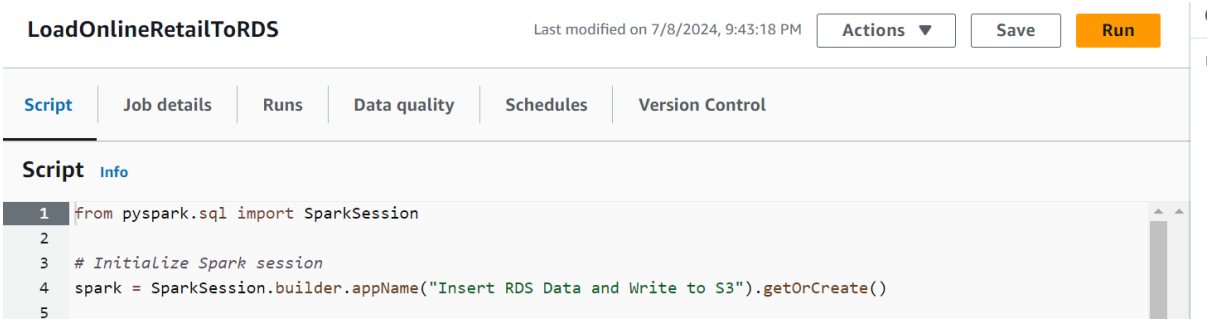



5. Keep everything else as it is.
6. Once the you create it should have status as available (Here I have stopped it for cost purpose)



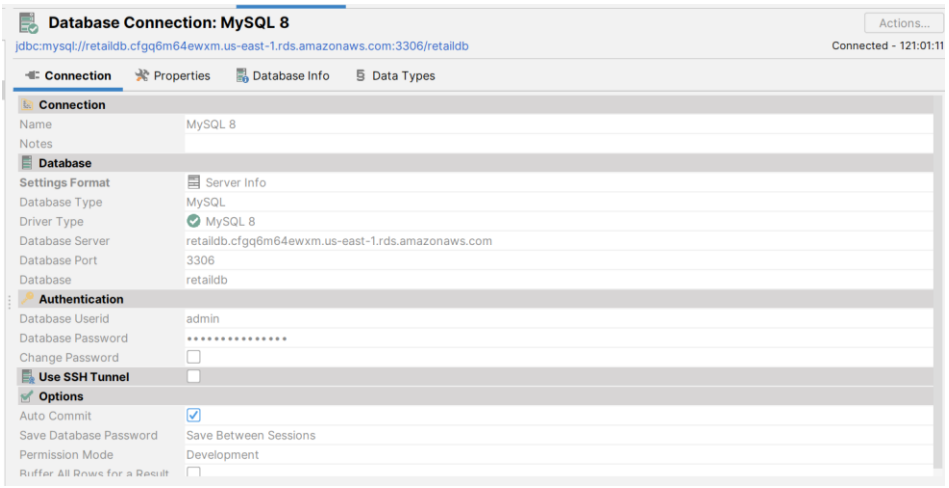

## Part 6: Load Transformed Data To RDS

1. Create Spark Glue Job to load s3 data to RDS

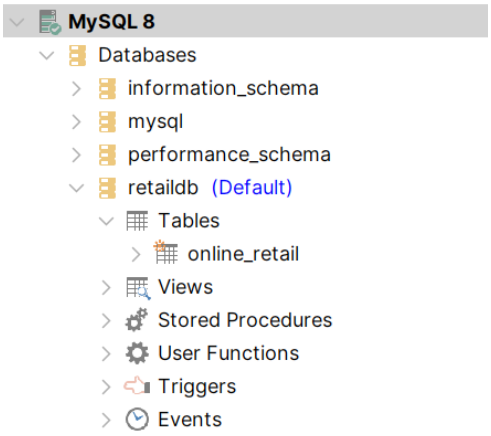



2. Provide connection details

3. Once the data is loaded check the data from local as well
    a. Create connection using DBVisualizer

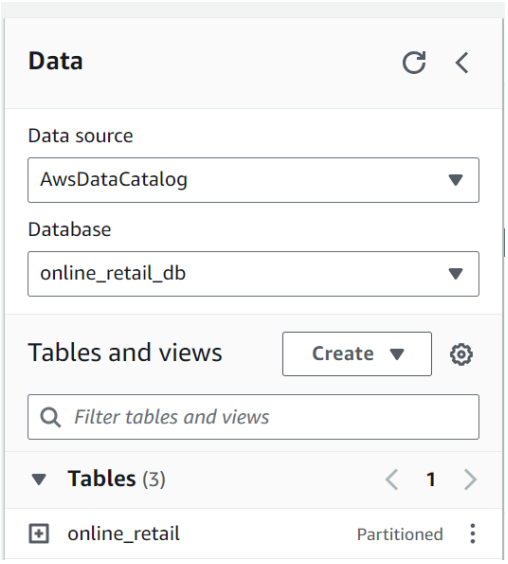



    b. Check if table is created

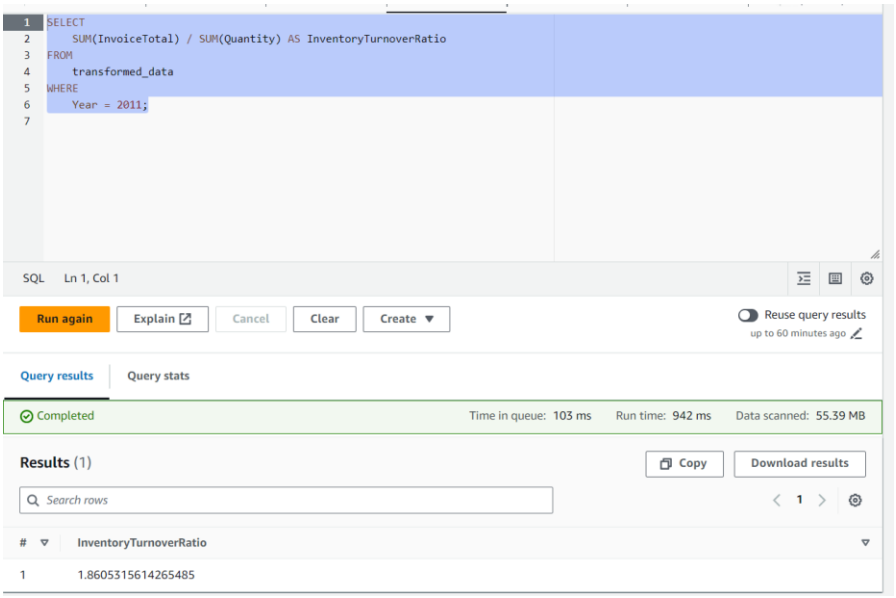



## Part 7: Setup Athena and KPIs

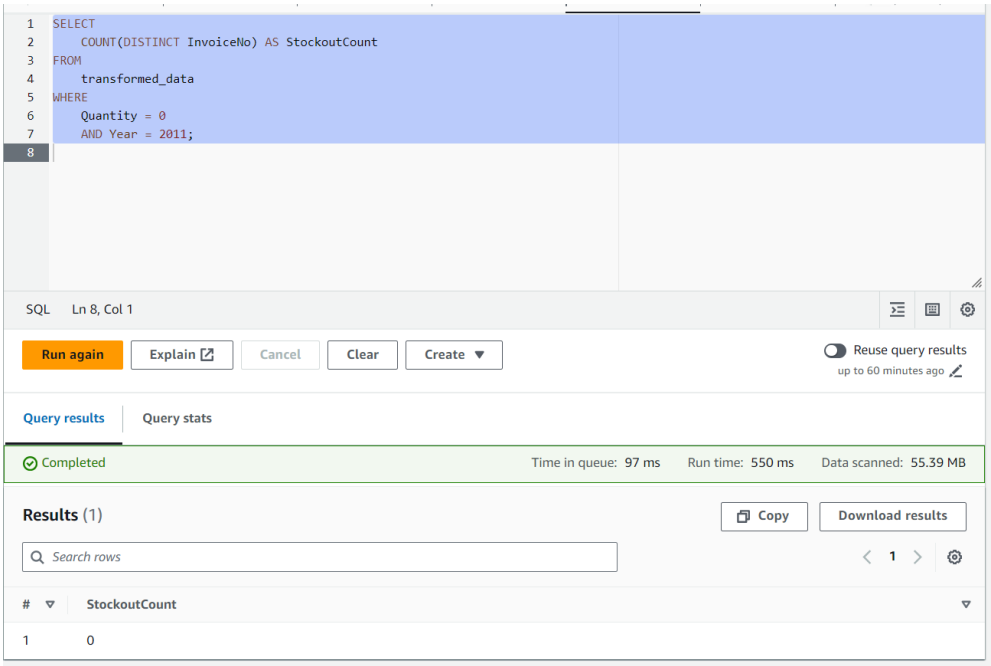
1. Go to Athena Console
2. Use the correct database and Table



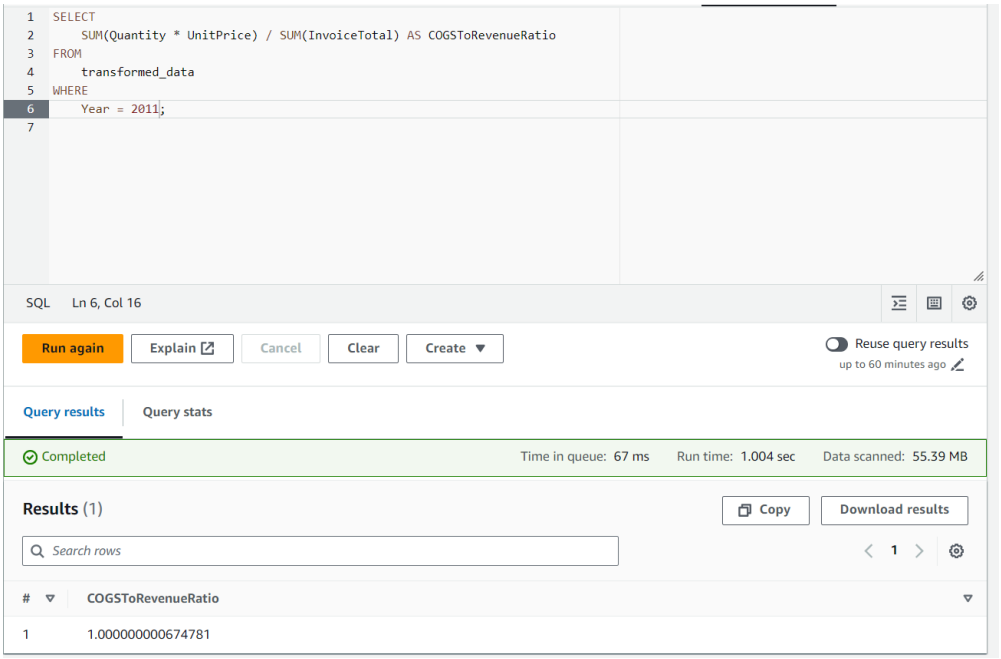

3. Run aggregation KPI queries:
    a. **Inventory Turnover Ratio** measures how efficiently inventory is managed by indicating how many times inventory is sold and replaced over a period.



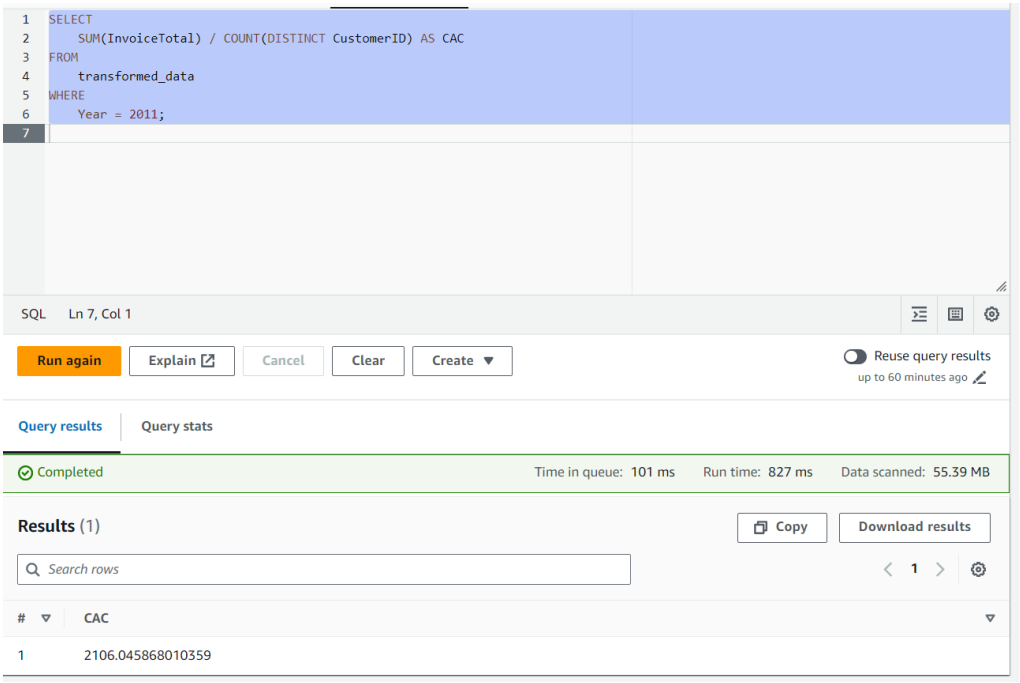
b. **Stockout Rate** measures the percentage of time products are out of stock.



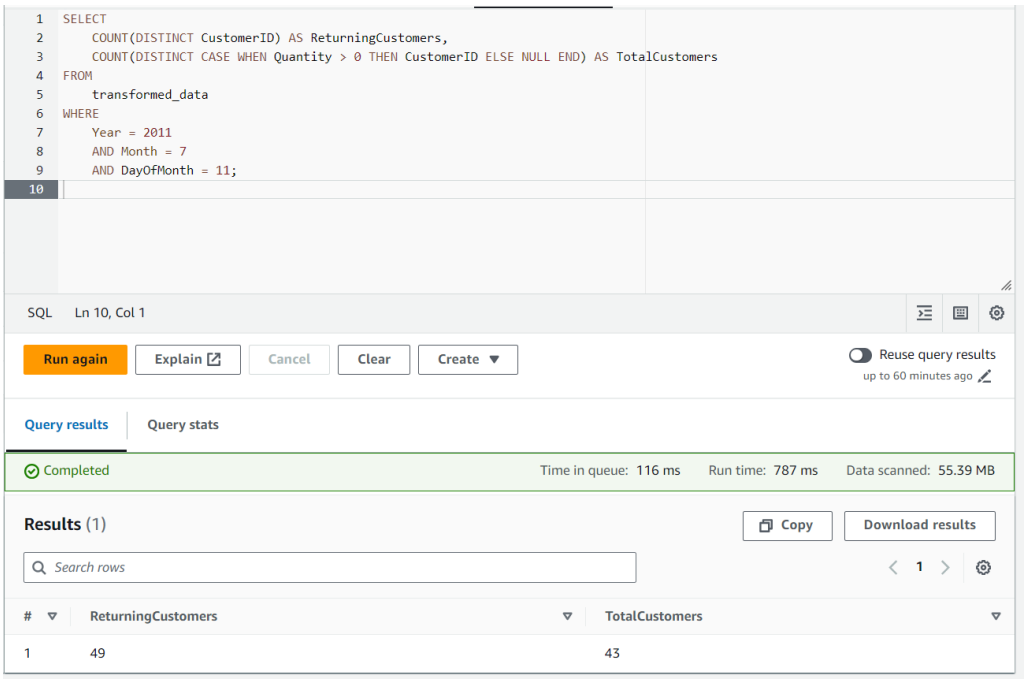

c. **COGS to Revenue Ratio** measures the efficiency of managing inventory costs relative to revenue.



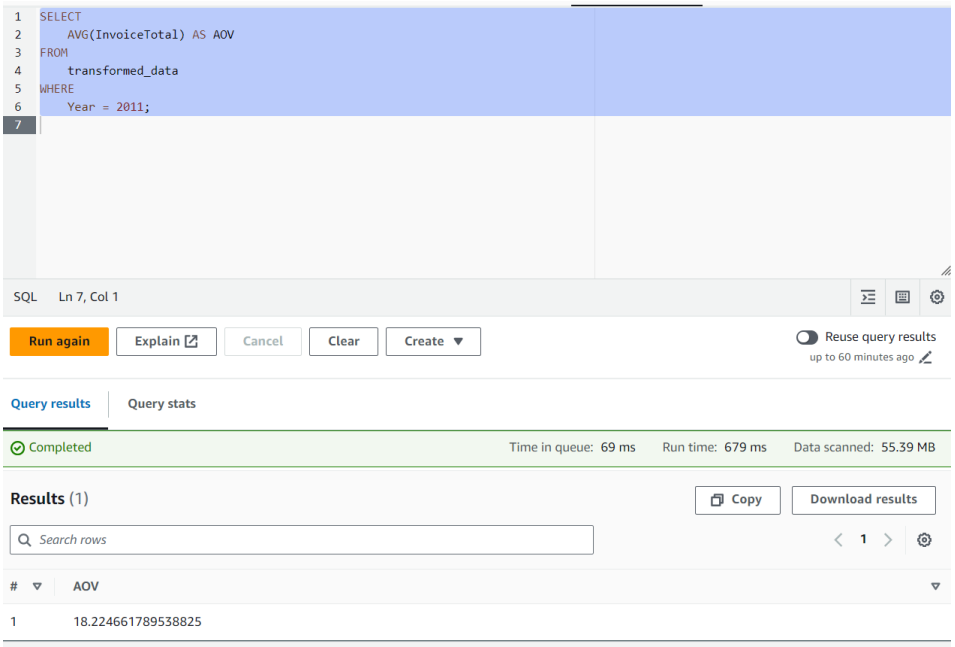
d. **Customer Acquisition Cost (CAC)** measures the average cost of acquiring a new customer.



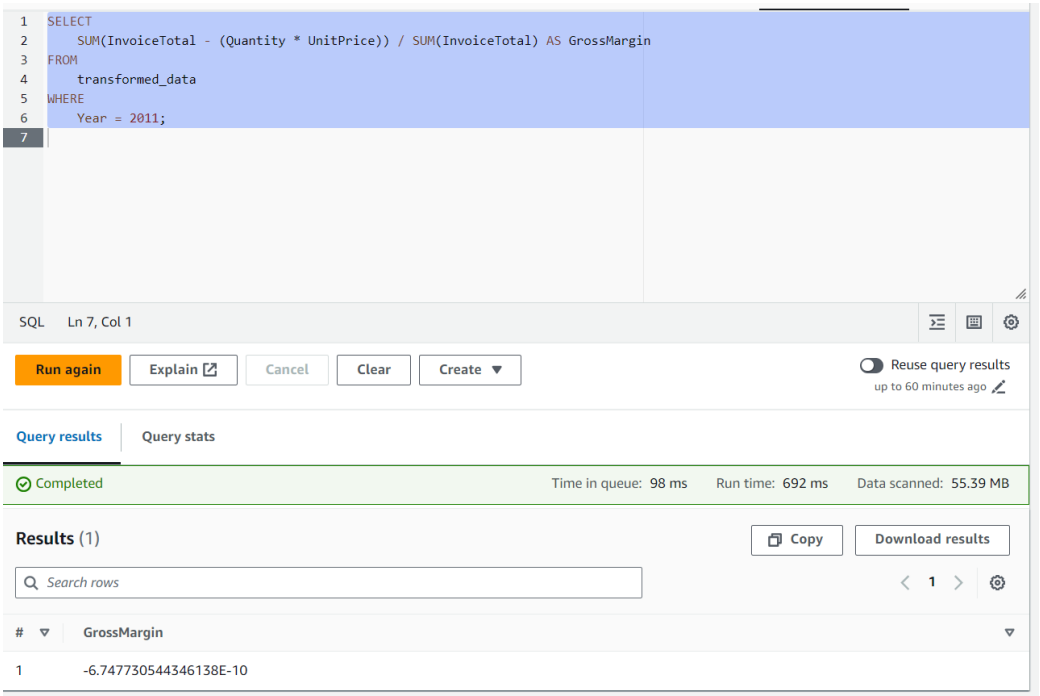

e. **Returning Customers**: The total number of unique customers who made purchases on July 11, 2011.



f. **Average Order Value (AOV)**: Measures the average amount spent per order over the course of the year 2011.

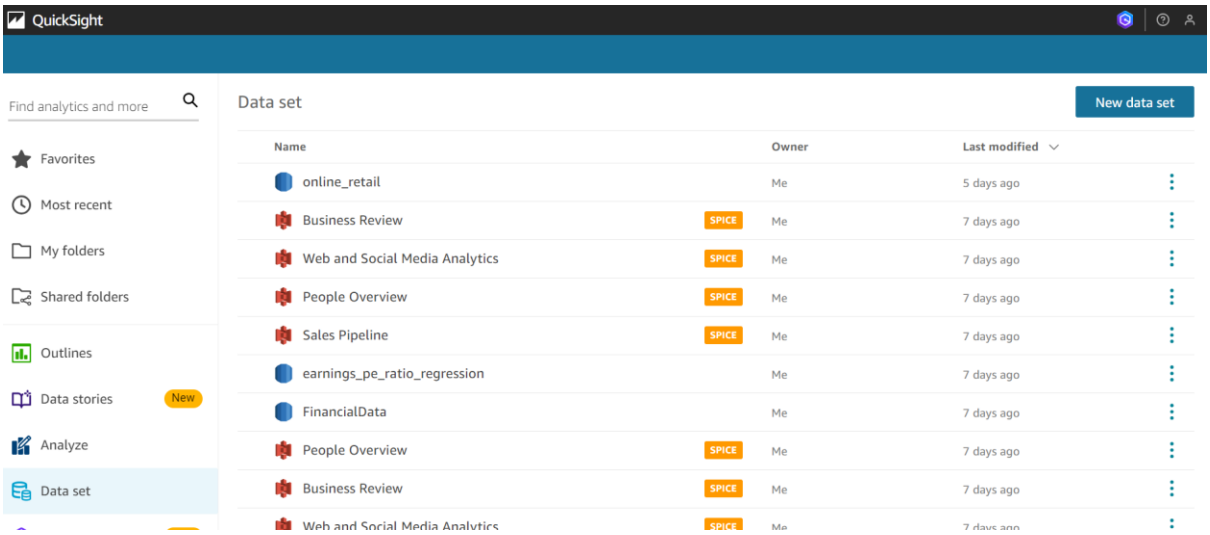



**g. Gross Margin**: Measures the profitability of products by comparing the revenue (InvoiceTotal) to the cost of goods sold (Quantity * UnitPrice).

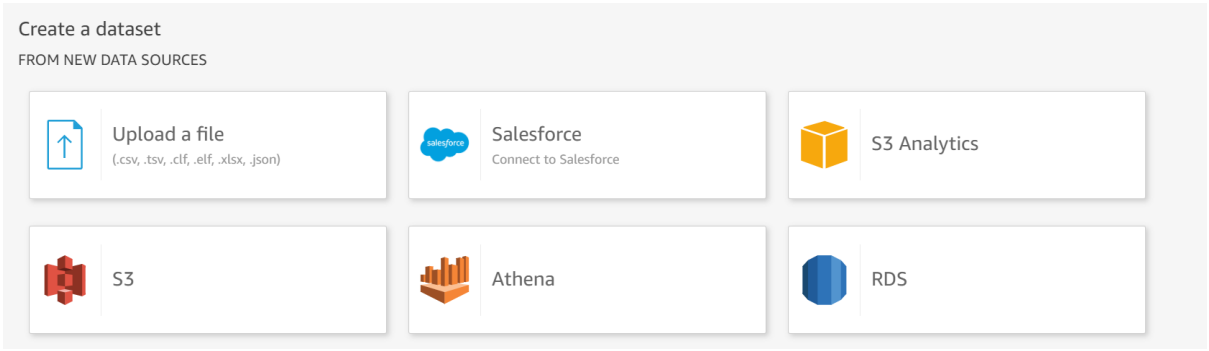


## Part 8: Quicksight and Visualizations

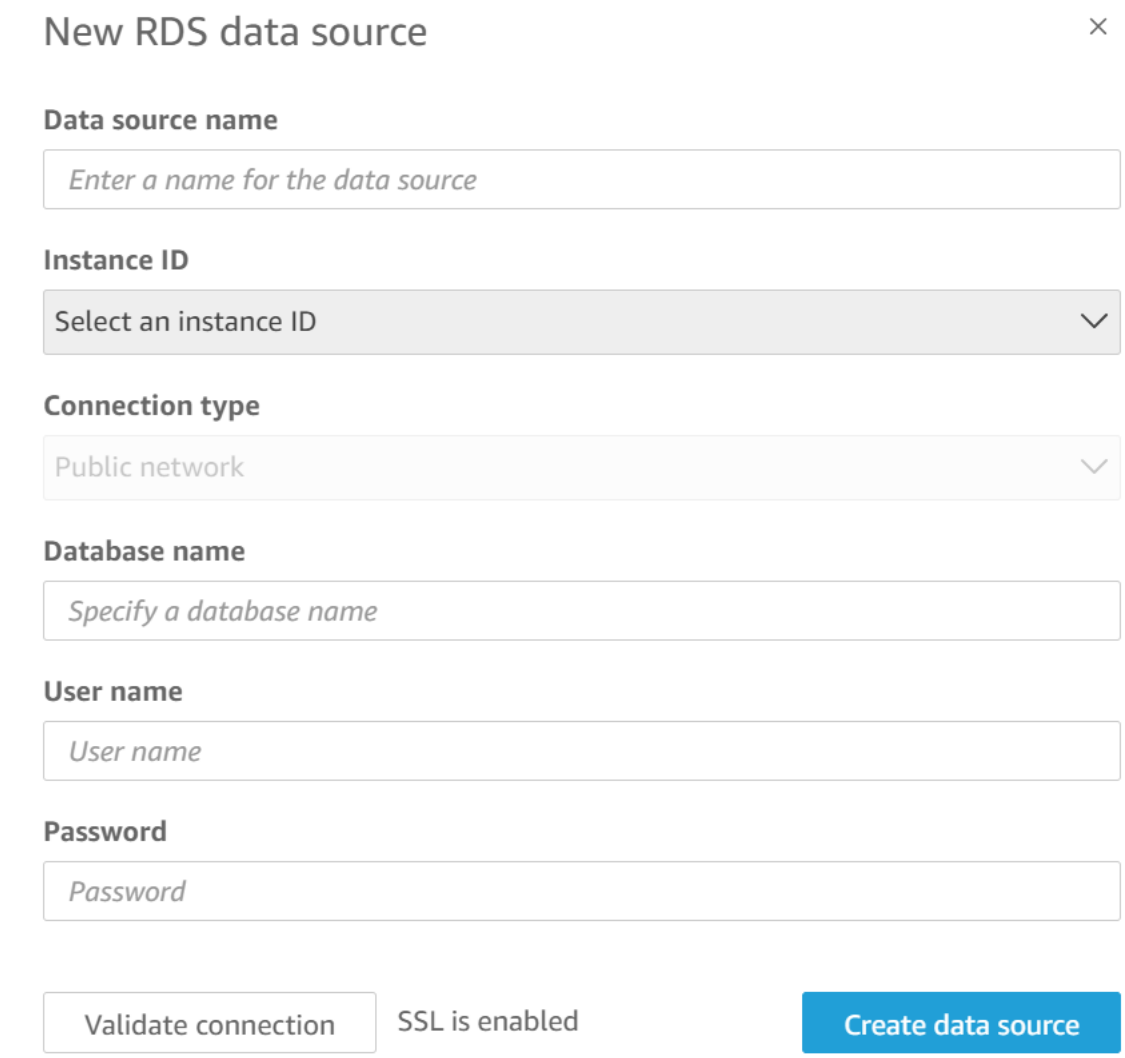
1. Go to Quicksight and login
2. Choose the dataset to have visualization on



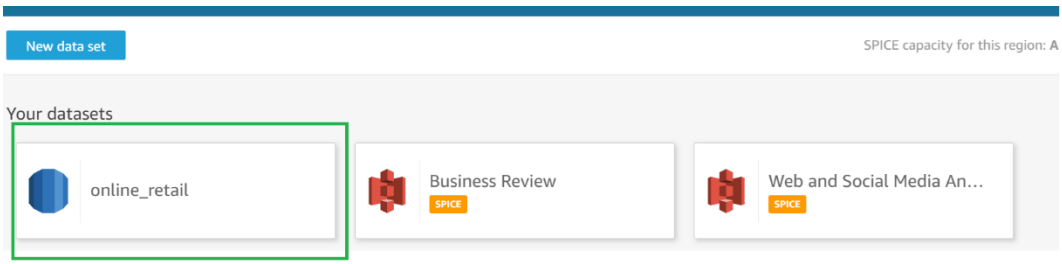

3. Click on new dataset and choose RDS



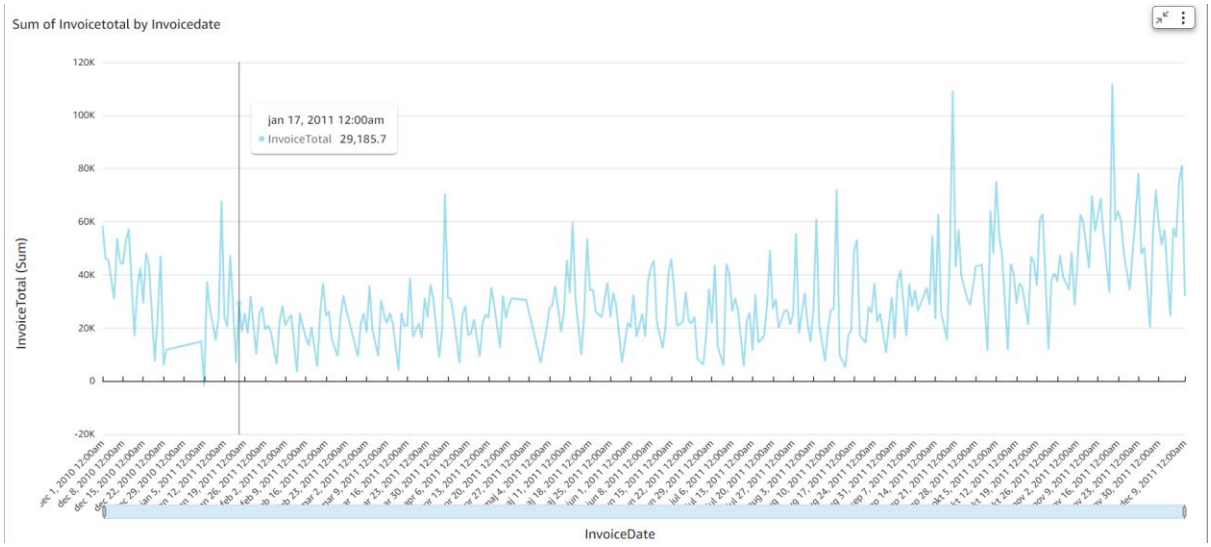
4. Enter RDS details and test the connection



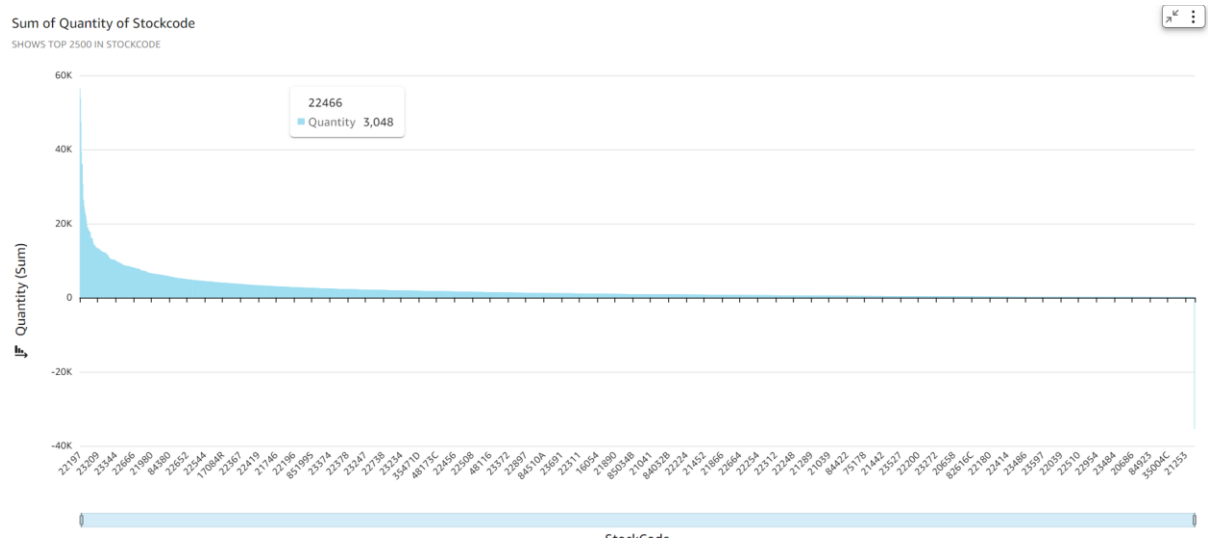

5. Once created, it will be listed in the datasets



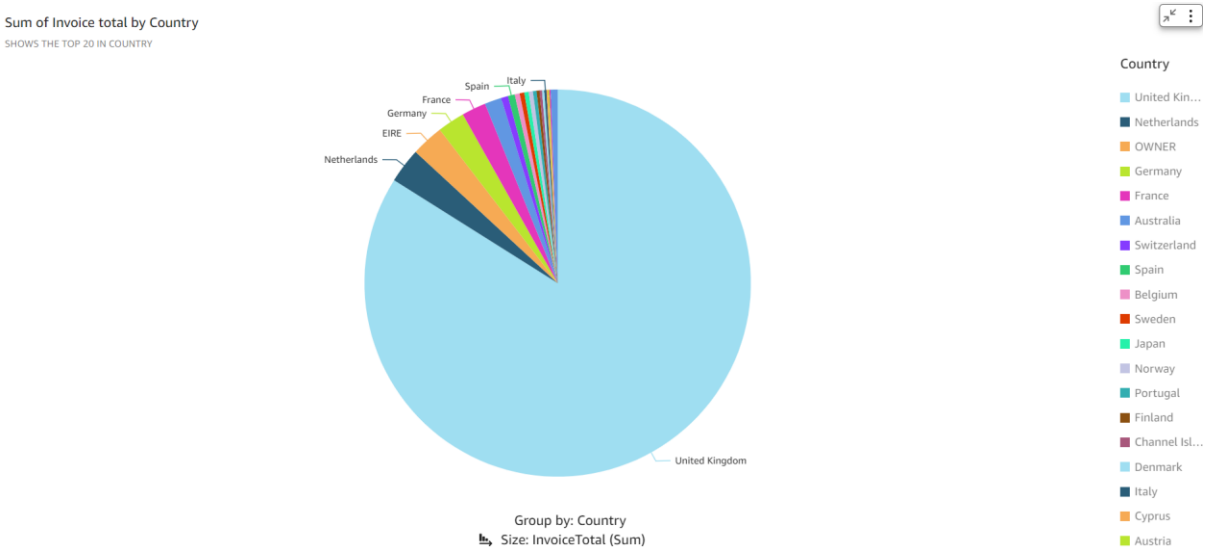
6. Create Dashboard
    a. Sum of Invoice total by invoice date



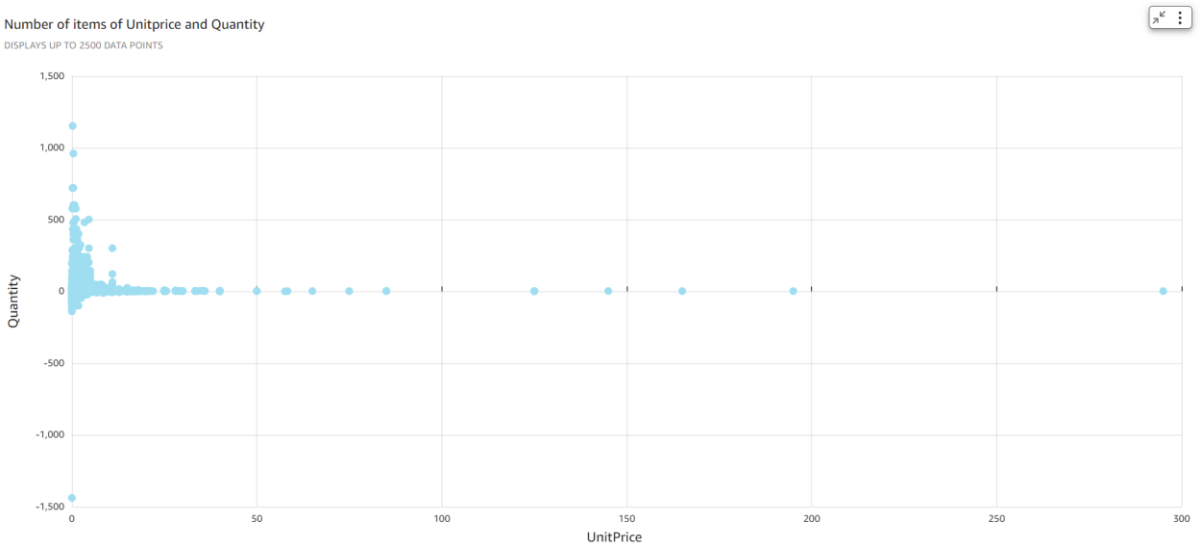

    b. Sum of Quantity of Stockcode



c. Sum of invoice total by country




d. Number of items of Unitprice and Quantity



e. Sum of Invoicetotal, Sum of Quantity, and Sum of Unitprice by Month

Sum of Invoicetotal, Sum of Quantity, and Sum of Unitprice by Month

| Rows | InvoiceTotal | Quantity | UnitPrice |
|---|---|---|---|
| 1 | 558,448.56 | 308,281 | 172,003.69 |
| 2 | 497,026.41 | 277,374 | 126,841.95 |
| 3 | 682,013.98 | 351,165 | 170,778.3 |
| 4 | 492,367.84 | 288,237 | 128,689.46 |
| 5 | 722,094.1 | 379,652 | 190,058.09 |
| 6 | 689,977.23 | 340,945 | 200,032.62 |
| 7 | 680,156.99 | 389,051 | 171,424.58 |
| 8 | 681,386.46 | 405,450 | 149,831.85 |
| 9 | 1,017,596.68 | 548,669 | 198,308.68 |
| 10 | 1,069,368.23 | 569,749 | 261,626.83 |
| 11 | 1,456,145.8 | 737,182 | 323,943.25 |
| 12 | 1,179,424.67 | 566,747 | 392,533.67 |

f. Sum of Quantity by Month and Country