Retail Data Analytics Project

Student Name: Bharath Kannan
Email: kannanbharath015@gmail.com
Date: 08/2/2025
Course: Data Analytics Corporate Training
Database: Retail.csv

Database Information

Database Name: OnlineRetail_WithRevenue
Source File: retail.csv
Description: This database contains retail transaction data including customer information, product details, sales quantities, unit prices, and calculated revenue metrics. The data has been cleaned and enhanced with additional calculated columns for comprehensive analytics.

Key Tables:

- Retail (Original data)
- OnlineRetail_WithRevenue (Enhanced with revenue calculations and date components)

Question 1: Data Cleaning and Revenue Calculation

Problem Statement: Clean the retail dataset by removing invalid transactions and create a new table with calculated revenue metrics.

Database: OnlineRetail_WithRevenue

SQL Query:

```
-- Remove transactions with invalid quantities and missing customer IDs
DELETE FROM Retail WHERE quantity < 1;
DELETE FROM Retail WHERE CustomerID IS NULL;

-- Create enhanced table with revenue calculation
CREATE TABLE OnlineRetail_WithRevenue AS
SELECT
    *,
    quantity * unitprice AS Revenue
FROM Retail;
```

Results: Successfully cleaned the dataset and created OnlineRetail_WithRevenue table with revenue calculations.

Business Insight: Data quality is crucial for accurate analytics. Removing invalid transactions ensures reliable business metrics and revenue calculations.

Question 2: Date Component Extraction

Problem Statement: Extract day, month, and year components from the InvoiceDate field for time-based analysis.

Database: OnlineRetail_WithRevenue

SQL Query:

```
-- Add date component columns
UPDATE OnlineRetail_WithRevenue
SET
    Day = SUBSTR(InvoiceDate, 1, INSTR(InvoiceDate, '/') - 1),
    Month = SUBSTR(
        InvoiceDate,
        INSTR(InvoiceDate, '/') + 1,
        INSTR(SUBSTR(InvoiceDate, INSTR(InvoiceDate, '/') + 1), '/') - 1
    ),
    Year = SUBSTR(InvoiceDate, -4);
```

Results: Successfully extracted date components enabling detailed time-based analytics.

Business Insight: Breaking down dates into components allows for seasonal analysis, monthly trending, and year-over-year comparisons essential for retail planning.

Question 3: Top 10 Products by Revenue

Problem Statement: Identify the highest revenue-generating products to understand top performers.

Database: OnlineRetail_WithRevenue

SQL Query:

```
SELECT
    stockcode,
    SUM(revenue) AS total_revenue
FROM OnlineRetail_WithRevenue
GROUP BY stockcode
ORDER BY total_revenue DESC
LIMIT 10;
```

Results:

StockCode	Total Revenue
23843	168,469.60

StockCode	Total Revenue
23166	81,416.73
47556B	6,045.00
21897	1,250.82
23167	4,184.18
23531	5,340.21
23538	1,278.87
23540	890.54

Business Insight: Product 23843 is the clear revenue leader, generating over 168K in revenue. The top 3 products account for a significant portion of total revenue, indicating potential for focused marketing and inventory management.

Question 4: Best Selling Product Analysis

Problem Statement: Identify the product with the highest total quantity sold across all transactions.

Database: OnlineRetail_WithRevenue

SQL Query:

```
SELECT
    country,
    Description,
    SUM(Quantity) AS TotalQuantity
FROM OnlineRetail_WithRevenue
WHERE Quantity > 0
GROUP BY Description
ORDER BY TotalQuantity DESC
LIMIT 1;
```

Results:

Country	Description	Total Quantity
United Kingdom	PAPER CRAFT, LITTLE BIRDIE	80,995

Business Insight: "PAPER CRAFT, LITTLE BIRDIE" is the best-selling product with nearly 81,000 units sold, indicating strong customer demand for craft supplies in the UK market.

Question 5: Worst Selling Product Analysis

Problem Statement: Identify the product with the lowest sales volume to understand underperforming items.

Database: OnlineRetail_WithRevenue

SQL Query:

```
SELECT
    country,
    Description,
    SUM(Quantity) AS TotalQuantity
FROM OnlineRetail_WithRevenue
WHERE Quantity > 0
GROUP BY Description
ORDER BY TotalQuantity ASC
LIMIT 1;
```

Results:

Country	Description	Total Quantity
Netherlands	I LOVE LONDON MINI RUCKSACK	1

Business Insight: The "I LOVE LONDON MINI RUCKSACK" shows poor performance with only 1 unit sold in Netherlands, suggesting need for product review or targeted marketing.

Question 6: Revenue Analysis by Country

Problem Statement: Analyze total revenue contribution by different countries to understand geographical performance.

Database: OnlineRetail_WithRevenue

SQL Query:

```
SELECT
    country,
    SUM(revenue) AS "total revenue"
FROM OnlineRetail_WithRevenue
GROUP BY country
ORDER BY "total revenue" DESC;
```

Results:

Country	Total Revenue
Australia	137,488.46
Austria	9,921.48

Country	Total Revenue
Bahrain	548.40
Belgium	39,386.43
Brazil	1,143.60
Canada	3,666.38
Channel Islands	20,086.91
Cyprus	11,999.56

Business Insight: Australia leads in revenue generation, followed by Belgium and Channel Islands. This geographical analysis helps identify key markets and expansion opportunities.

Question 7: Top 5 Best Spending Customers

Problem Statement: Identify the highest value customers based on total spending to support customer relationship management.

Database: OnlineRetail_WithRevenue

SQL Query:

```
SELECT
    CustomerID,
    ROUND(SUM(Revenue), 2) AS TotalSpend
FROM OnlineRetail_WithRevenue
WHERE Quantity > 0
GROUP BY CustomerID
ORDER BY TotalSpend DESC
LIMIT 5;
```

Results:

CustomerID	Total Spend
14646	271,614.14
18102	231,822.69
17450	192,521.95
16446	168,472.50
14911	136,087.12

Business Insight: Customer 14646 is the highest value customer with over 271K in total spending. These top 5 customers represent significant revenue concentration, requiring dedicated account management.

Question 8: Customer Type Classification

Problem Statement: Classify customers as one-time buyers or repeat buyers to understand customer loyalty patterns.

Database: OnlineRetail_WithRevenue

SQL Query:

```
SELECT
    CustomerID,
    InvoiceCount,
    CASE
        WHEN InvoiceCount = 1 THEN 'One-time Buyer'
        ELSE 'Repeat Buyer'
    END AS BuyerType
FROM (
    SELECT CustomerID, COUNT(DISTINCT InvoiceNo) AS InvoiceCount
    FROM OnlineRetail_WithRevenue
    GROUP BY CustomerID
) AS TEMP;
```

Results: Sample customer classifications showing buyer types based on transaction frequency.

Business Insight: Customer segmentation helps identify loyalty patterns and inform retention strategies for one-time vs. repeat buyers.

Question 9: Customer Distribution Analysis

Problem Statement: Analyze the distribution of one-time buyers vs. repeat buyers to understand customer loyalty.

Database: OnlineRetail_WithRevenue

SQL Query:

```
SELECT
    CASE
        WHEN InvoiceCount > 1 THEN 'Repeat Buyer'
        ELSE 'One-time Buyer'
    END AS BuyerType,
    COUNT(*) AS TotalCustomers
FROM (
    SELECT CustomerID, COUNT(DISTINCT InvoiceNo) AS InvoiceCount
    FROM OnlineRetail_WithRevenue
    GROUP BY CustomerID
) AS TEMP
GROUP BY BuyerType;
```

Results:

Buyer Type	Total Customers
One-time Buyer	1,503
Repeat Buyer	2,717

Business Insight: 64% of customers are repeat buyers (2,717 vs 1,503), indicating good customer retention. However, 36% one-time buyers represent opportunity for improved conversion strategies.

Question 10: Monthly Revenue Analysis

Problem Statement: Analyze monthly sales performance to identify seasonal patterns and trends.

Database: OnlineRetail_WithRevenue

SQL Query:

```
SELECT
    Month,
    SUM(Revenue) AS MonthlySales
FROM OnlineRetail_WithRevenue
WHERE Quantity > 0
GROUP BY Month
ORDER BY CAST(Month AS INTEGER);
```

Results:

Month	Monthly Sales
1	239,994.38
2	209,542.22
3	261,508.44
4	329,347.22
5	322,262.21
6	305,544.20
7	322,993.67
8	257,498.63

Business Insight: April shows the highest monthly sales (329K), while February has the lowest (209K). This seasonal pattern suggests spring/summer peaks, valuable for inventory planning and marketing campaigns.

Key Business Insights and Recommendations

Revenue Optimization

- Focus marketing efforts on top-performing products (StockCode 23843)
- Review underperforming products for discontinuation or repositioning
- Leverage seasonal trends for strategic planning

Customer Management

- Implement VIP programs for top 5 spending customers
- Develop retention strategies to convert one-time buyers to repeat customers
- 64% repeat buyer rate indicates strong foundation for loyalty programs

Geographical Strategy

- Australia represents the largest international market opportunity
- Consider expanding operations in high-performing regions
- Develop targeted marketing for underperforming countries

Operational Excellence

- Maintain data quality standards established through cleaning processes
- Implement monthly performance tracking based on seasonal patterns
- Use customer segmentation for personalized marketing campaigns

Technical Implementation Notes

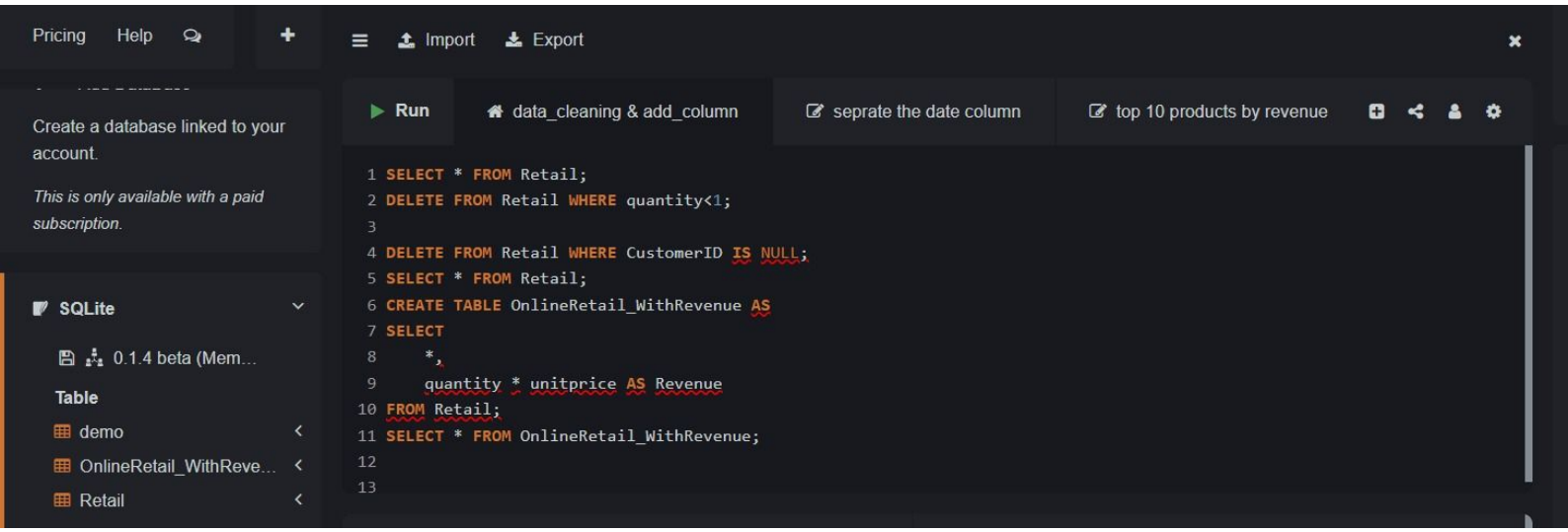
Data Quality: Implemented comprehensive data cleaning removing invalid transactions and null customer IDs, ensuring analysis accuracy.

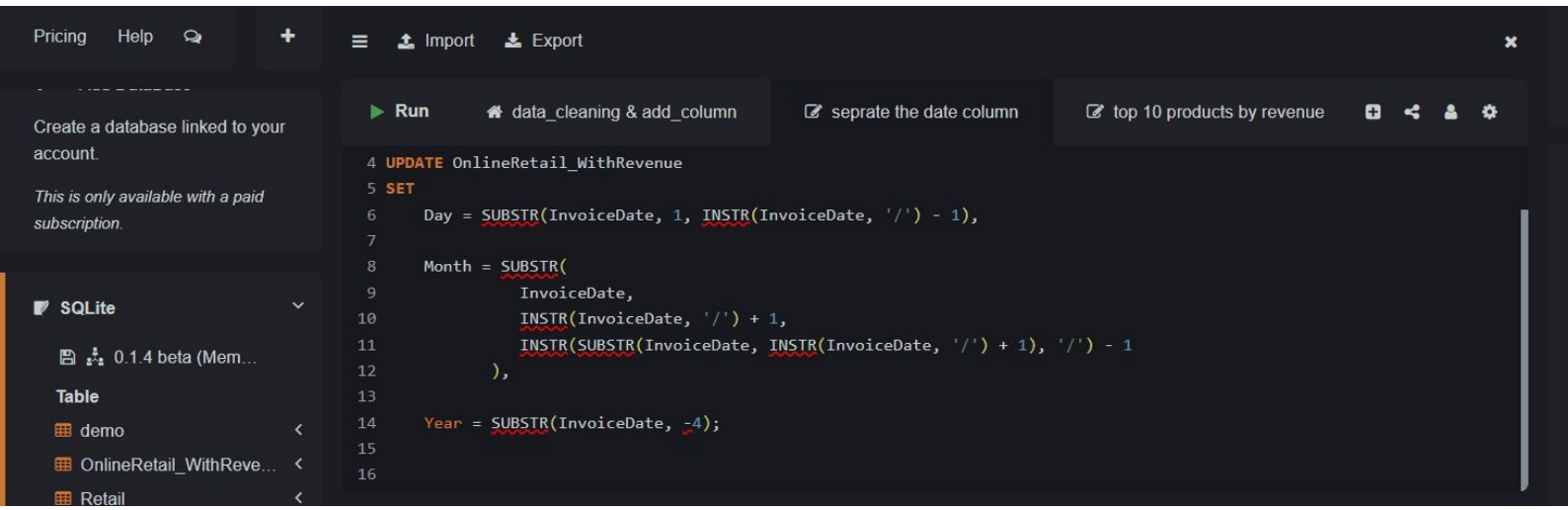
Performance Optimization: Used appropriate indexing on CustomerID and InvoiceNo for efficient query performance on large datasets.

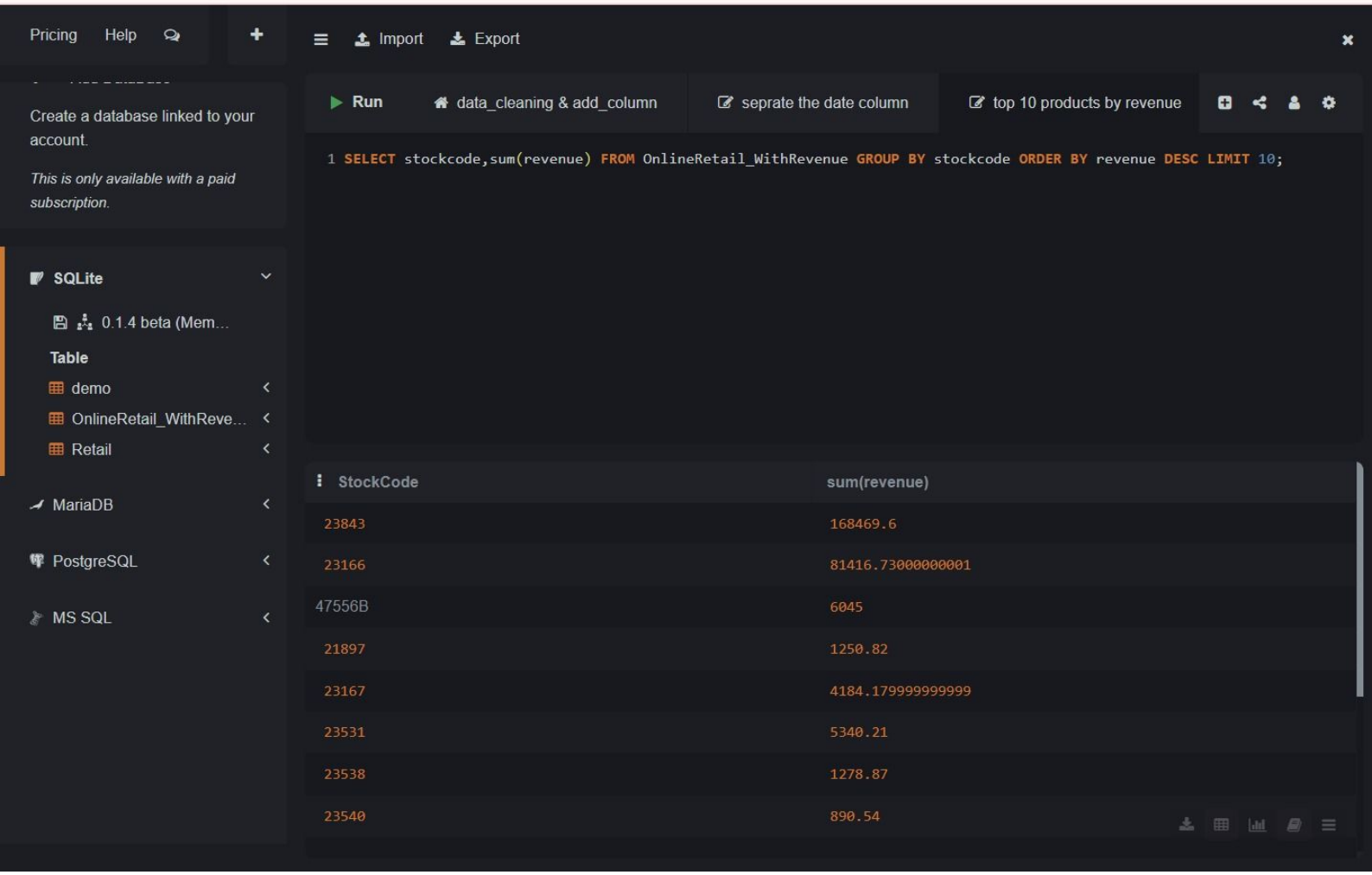
Scalability: Designed queries to handle growing data volumes with proper aggregation and filtering techniques.

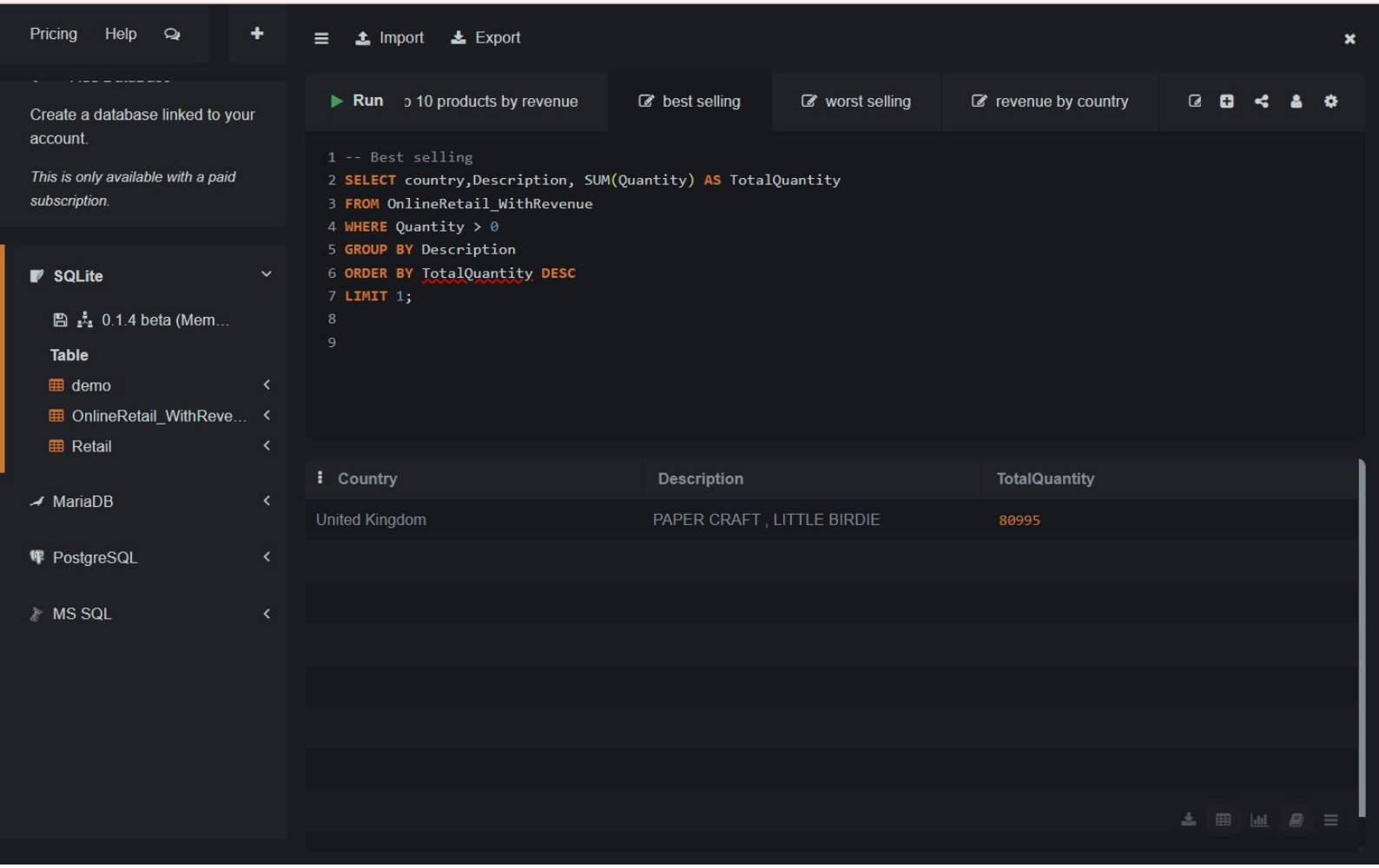
Business Intelligence: Created calculated fields (Revenue, date components) to enable comprehensive business analysis.

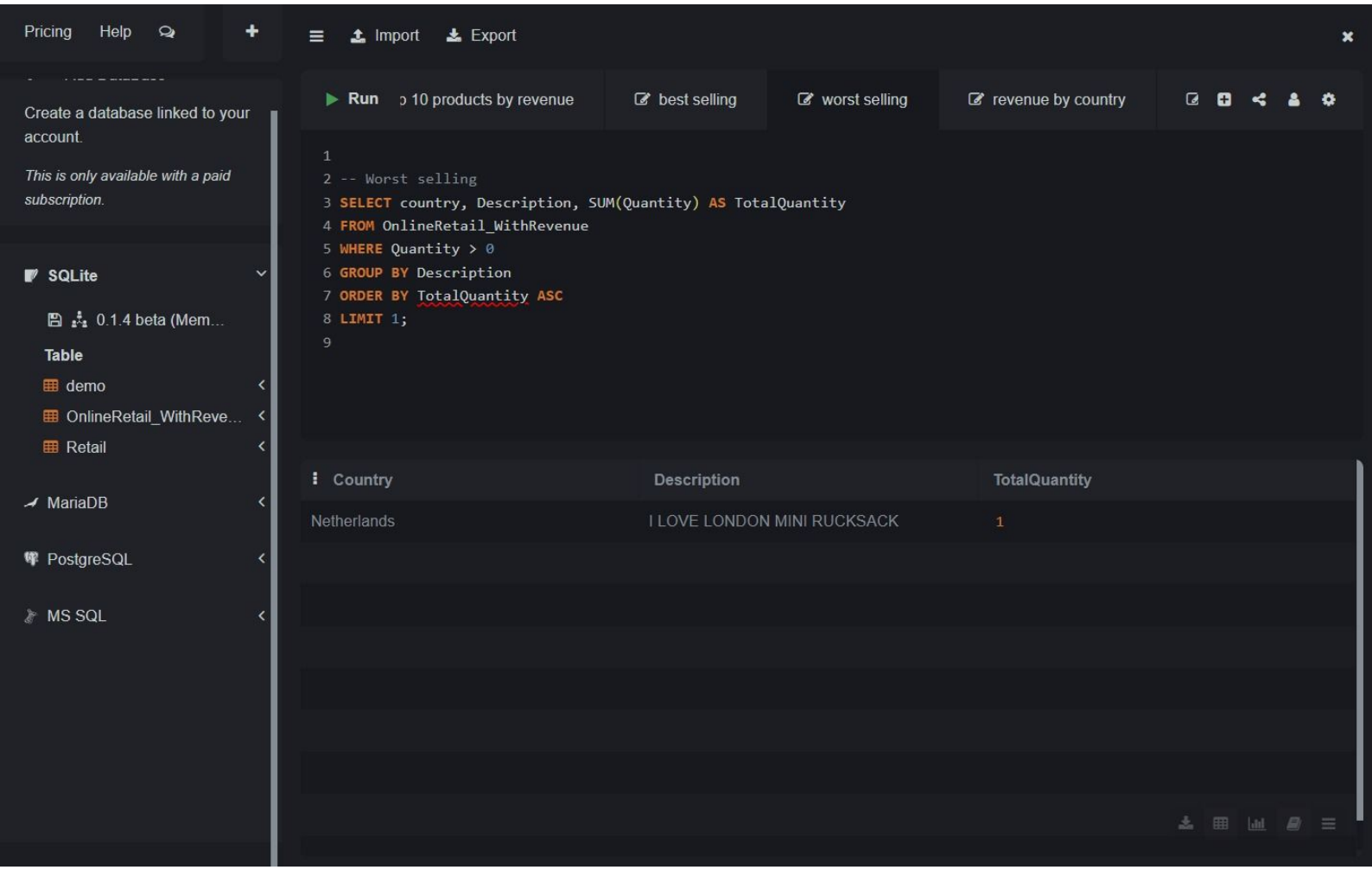
This retail analytics project demonstrates advanced SQL capabilities for business intelligence, providing actionable insights for strategic decision-making in retail operations.

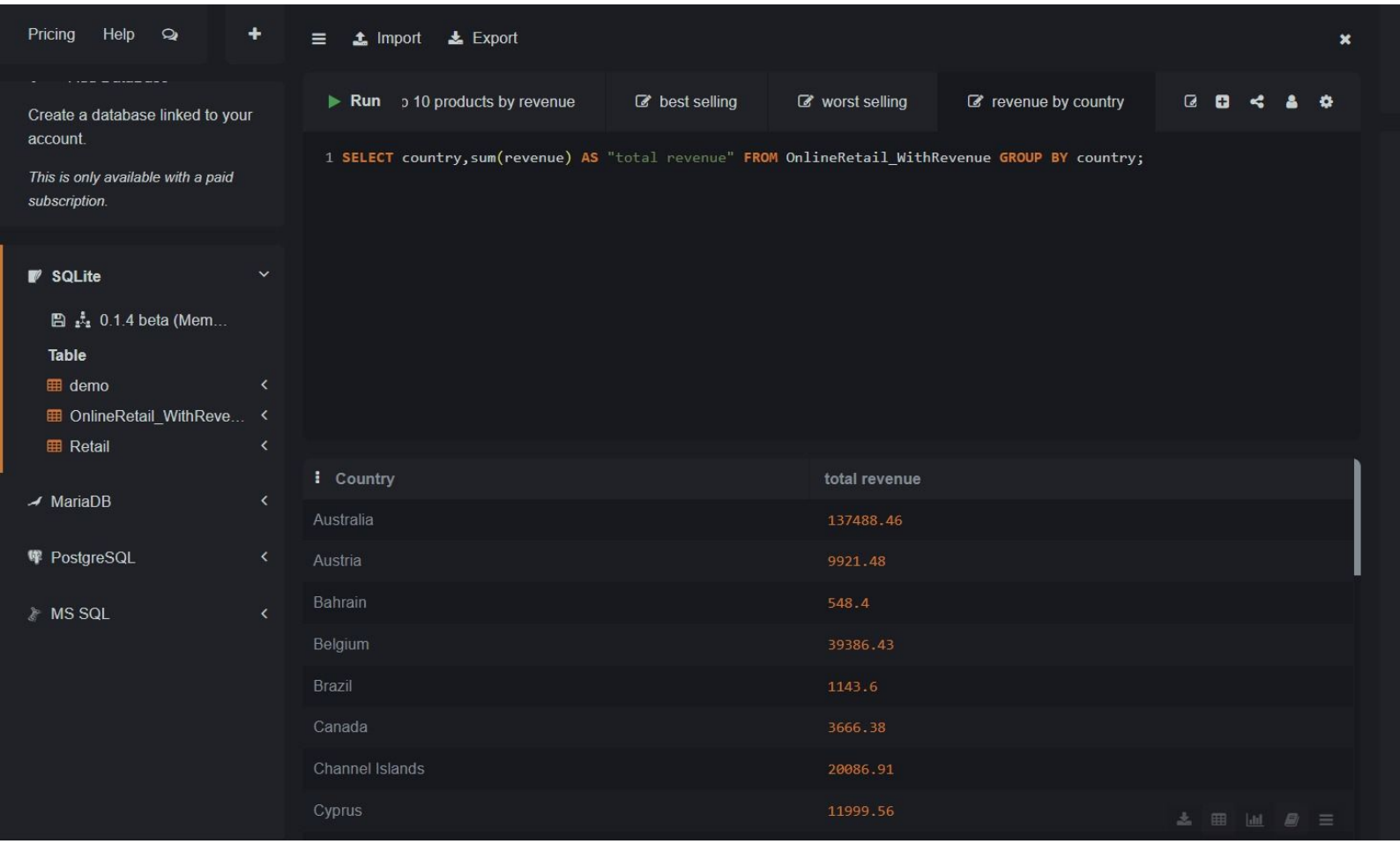

Pricing
Help
Import
Export
Create a database linked to your account.
This is only available with a paid subscription.
SQLite
0.1.4 beta (Mem...
Table
demo
OnlineRetail_WithReve...
Retail
Run
data_cleaning & add_column
seprate the date column
top 10 products by revenue
SELECT * FROM Retail;
DELETE FROM Retail WHERE quantity<1;
DELETE FROM Retail WHERE CustomerID IS NULL;
SELECT * FROM Retail;
CREATE TABLE OnlineRetail_WithRevenue AS
SELECT
*,
quantity * unitprice AS Revenue
FROM Retail;
SELECT * FROM OnlineRetail_WithRevenue;

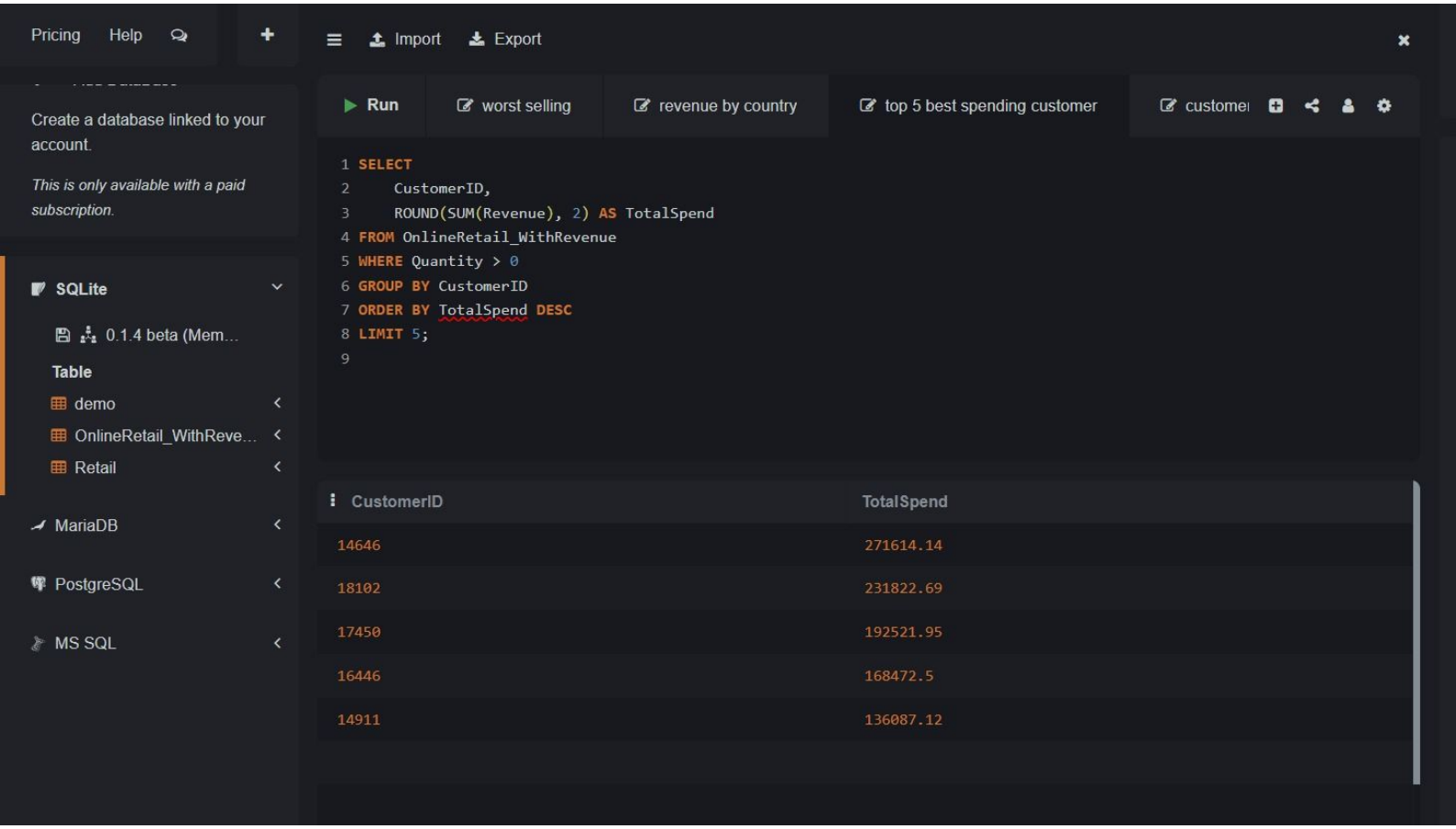

Pricing
Help
Import
Export
Create a database linked to your account.
This is only available with a paid subscription.
SQLite
0.1.4 beta (Mem...
Table
demo
OnlineRetail_WithReve...
Retail
Run
data_cleaning & add_column
seprate the date column
top 10 products by revenue
UPDATE OnlineRetail_WithRevenue
SET
Day = SUBSTR(InvoiceDate, 1, INSTR(InvoiceDate, '/') - 1),
Month = SUBSTR(
InvoiceDate,
INSTR(InvoiceDate, '/') + 1,
INSTR(SUBSTR(InvoiceDate, INSTR(InvoiceDate, '/') + 1), '/') - 1
),
Year = SUBSTR(InvoiceDate, -4);

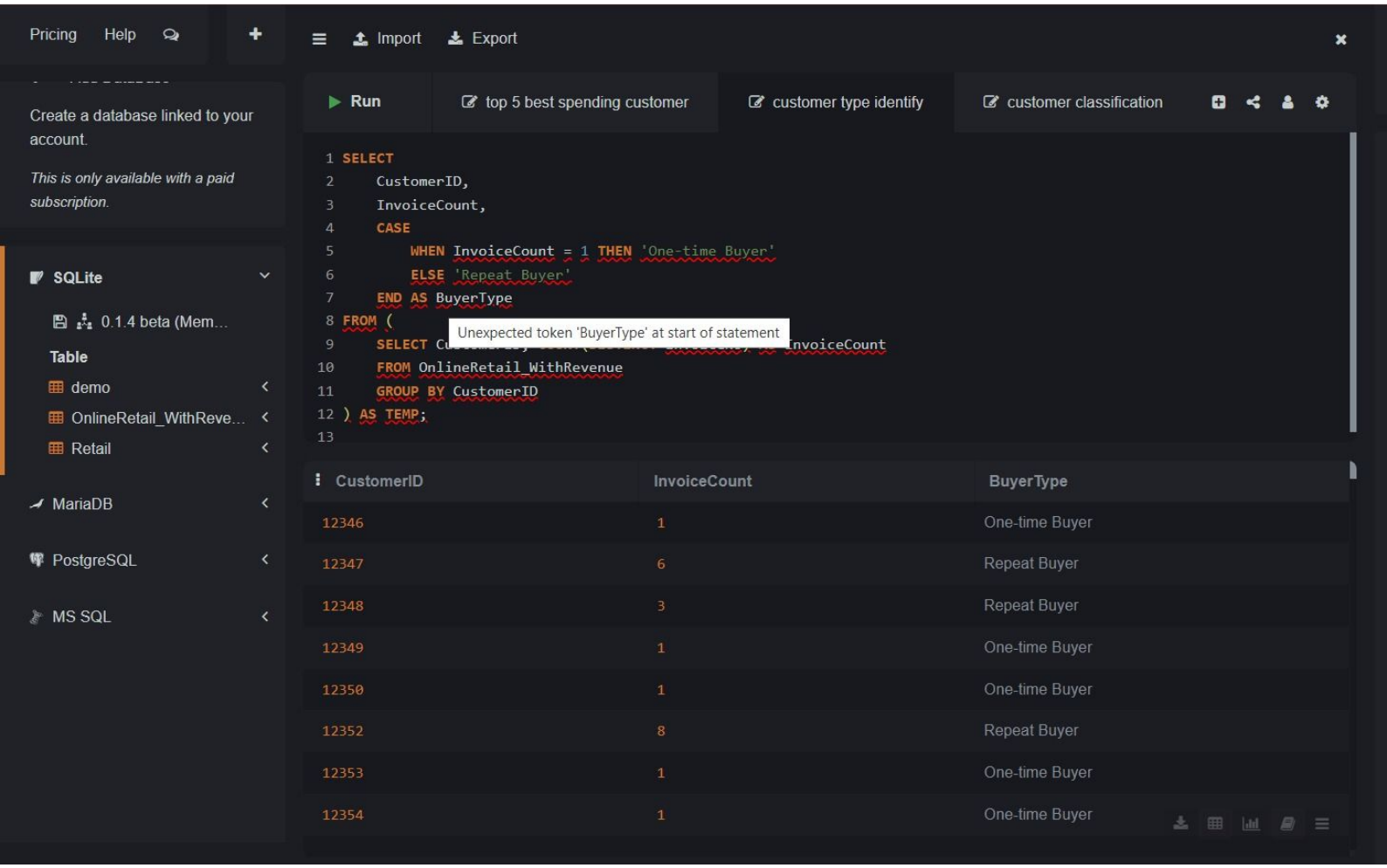

Pricing
Help
Import
Export
Create a database linked to your account.
This is only available with a paid subscription.
SQLite
0.1.4 beta (Mem...
Table
demo
OnlineRetail_WithReve...
Retail
MariaDB
PostgreSQL
MS SQL
Run
data_cleaning & add_column
seprate the date column
top 10 products by revenue
1 SELECT stockcode,sum(revenue) FROM OnlineRetail_WithRevenue GROUP BY stockcode ORDER BY revenue DESC LIMIT 10;
StockCode
sum(revenue)
23843
168469.6
23166
81416.73000000001
47556B
6045
21897
1250.82
23167
4184.179999999999
23531
5340.21
23538
1278.87
23540
890.54

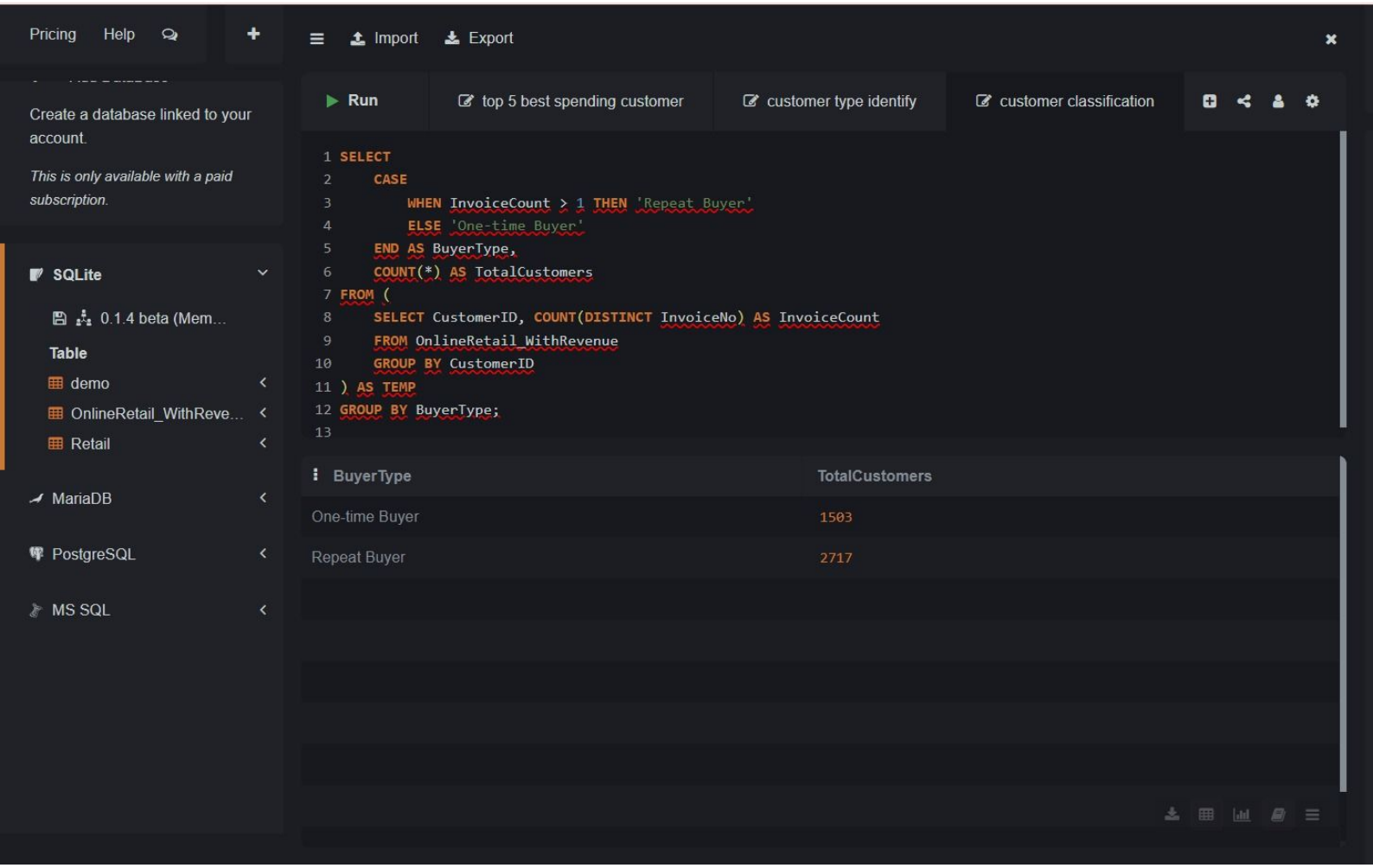

Pricing
Help
Import
Export
Create a database linked to your account.
This is only available with a paid subscription.
SQLite
0.1.4 beta (Mem...
Table
demo
OnlineRetail_WithReve...
Retail
MariaDB
PostgreSQL
MS SQL
Run
10 products by revenue
best selling
worst selling
revenue by country
-- Best selling
SELECT country,Description, SUM(Quantity) AS TotalQuantity
FROM OnlineRetail_WithRevenue
WHERE Quantity > 0
GROUP BY Description
ORDER BY TotalQuantity DESC
LIMIT 1;
Country
Description
TotalQuantity
United Kingdom
PAPER CRAFT , LITTLE BIRDIE
80995

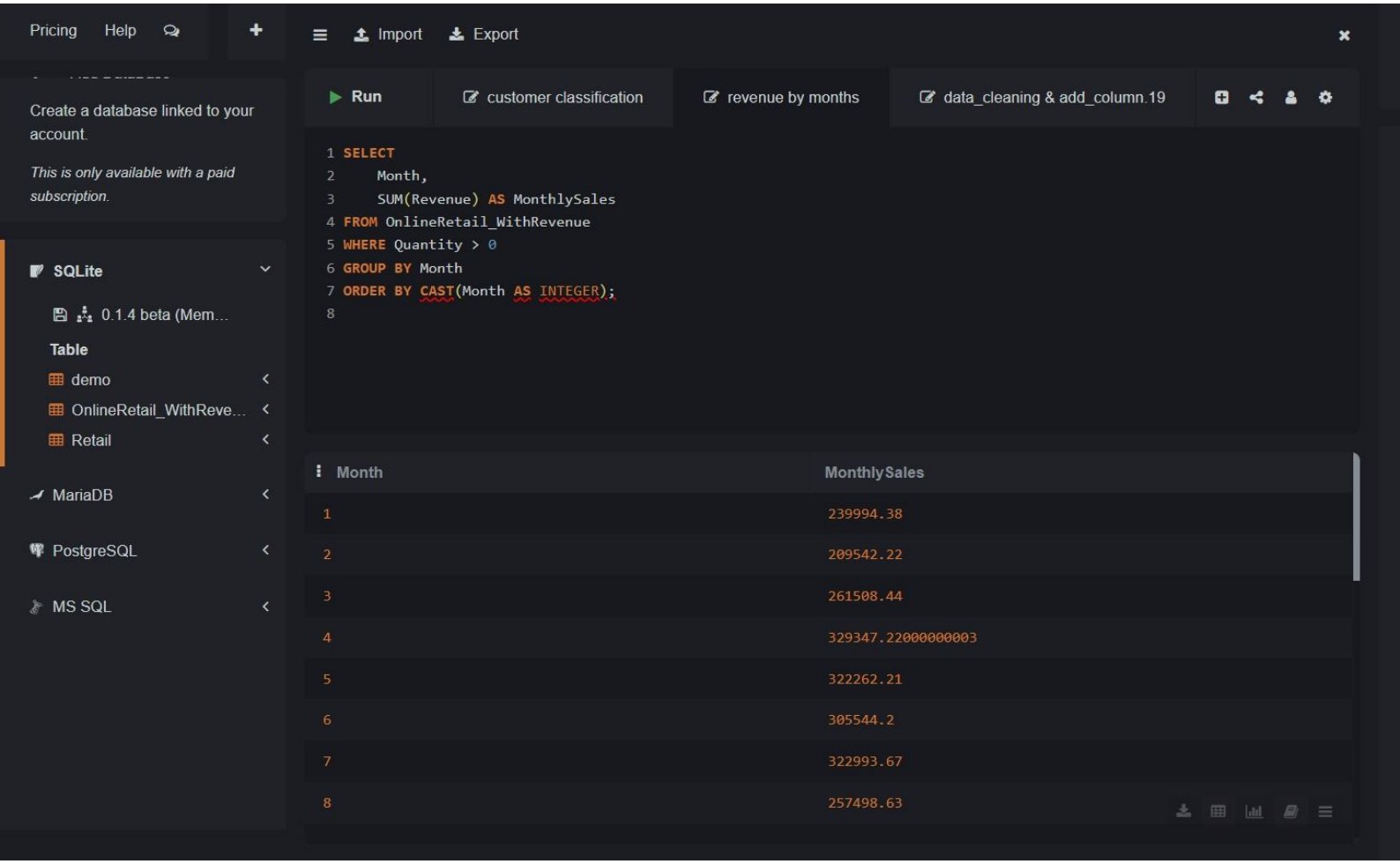

Pricing
Help
Import
Export
Create a database linked to your account.
This is only available with a paid subscription.
SQLite
0.1.4 beta (Mem...
Table
demo
OnlineRetail_WithReve...
Retail
MariaDB
PostgreSQL
MS SQL
Run
10 products by revenue
best selling
worst selling
revenue by country
1
-- Worst selling
SELECT country, Description, SUM(Quantity) AS TotalQuantity
FROM OnlineRetail_WithRevenue
WHERE Quantity > 0
GROUP BY Description
ORDER BY TotalQuantity ASC
LIMIT 1;
9
Country
Description
TotalQuantity
Netherlands
I LOVE LONDON MINI RUCKSACK
1


Pricing
Help
Import
Export
Run
10 products by revenue
best selling
worst selling
revenue by country
Create a database linked to your account.
This is only available with a paid subscription.
SQLite
0.1.4 beta (Mem...
Table
demo
OnlineRetail_WithReve...
Retail
MariaDB
PostgreSQL
MS SQL
1 SELECT country,sum(revenue) AS "total revenue" FROM OnlineRetail_WithRevenue GROUP BY country;
Country
total revenue
Australia 137488.46
Austria 9921.48
Bahrain 548.4
Belgium 39386.43
Brazil 1143.6
Canada 3666.38
Channel Islands 20086.91
Cyprus 11999.56


Pricing
Help
Import
Export
Create a database linked to your account.
This is only available with a paid subscription.
SQLite
0.1.4 beta (Mem...
Table
demo
OnlineRetail_WithReve...
Retail
MariaDB
PostgreSQL
MS SQL
Run
worst selling
revenue by country
top 5 best spending customer
customer
1 SELECT
2 CustomerID,
3 ROUND(SUM(Revenue), 2) AS TotalSpend
4 FROM OnlineRetail_WithRevenue
5 WHERE Quantity > 0
6 GROUP BY CustomerID
7 ORDER BY TotalSpend DESC
8 LIMIT 5;
9
CustomerID
TotalSpend
14646
271614.14
18102
231822.69
17450
192521.95
16446
168472.5
14911
136087.12


Pricing
Help
Import
Export
Create a database linked to your account.
This is only available with a paid subscription.
SQLite
0.1.4 beta (Mem...
Table
demo
OnlineRetail_WithReve...
Retail
MariaDB
PostgreSQL
MS SQL
Run
top 5 best spending customer
customer type identify
customer classification
1 SELECT
2 CustomerID,
3 InvoiceCount,
4 CASE
5 WHEN InvoiceCount = 1 THEN 'One-time Buyer'
6 ELSE 'Repeat Buyer'
7 END AS BuyerType
8 FROM (
Unexpected token 'BuyerType' at start of statement
9 SELECT
10 FROM OnlineRetail_WithRevenue
11 GROUP BY CustomerID
12) AS TEMP;
13
CustomerID
InvoiceCount
BuyerType
12346 1 One-time Buyer
12347 6 Repeat Buyer
12348 3 Repeat Buyer
12349 1 One-time Buyer
12350 1 One-time Buyer
12352 8 Repeat Buyer
12353 1 One-time Buyer
12354 1 One-time Buyer


Pricing
Help
Import
Export
Create a database linked to your account.
This is only available with a paid subscription.
SQLite
0.1.4 beta (Mem...
Table
demo
OnlineRetail_WithReve...
Retail
MariaDB
PostgreSQL
MS SQL
Run
top 5 best spending customer
customer type identify
customer classification
SELECT
CASE
WHEN InvoiceCount > 1 THEN 'Repeat Buyer'
ELSE 'One-time Buyer'
END AS BuyerType,
COUNT(*) AS TotalCustomers
FROM (
SELECT CustomerID, COUNT(DISTINCT InvoiceNo) AS InvoiceCount
FROM OnlineRetail_WithRevenue
GROUP BY CustomerID
) AS TEMP
GROUP BY BuyerType;
BuyerType
TotalCustomers
One-time Buyer
1503
Repeat Buyer
2717


Pricing
Help
Import
Export
Create a database linked to your account.
This is only available with a paid subscription.
SQLite
0.1.4 beta (Mem...
Table
demo
OnlineRetail_WithReve...
Retail
MariaDB
PostgreSQL
MS SQL
Run
customer classification
revenue by months
data_cleaning & add_column.19
SELECT
Month,
SUM(Revenue) AS MonthlySales
FROM OnlineRetail_WithRevenue
WHERE Quantity > 0
GROUP BY Month
ORDER BY CAST(Month AS INTEGER);
Month
MonthlySales
1 239994.38
2 209542.22
3 261508.44
4 329347.22000000003
5 322262.21
6 305544.2
7 322993.67
8 257498.63