**Project : DATA-MINING**

**Name : Hasnain Asad**

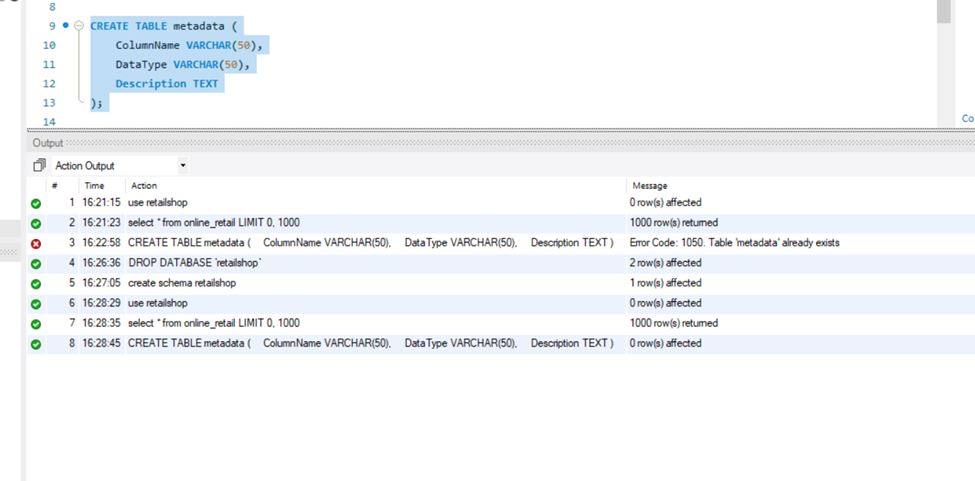
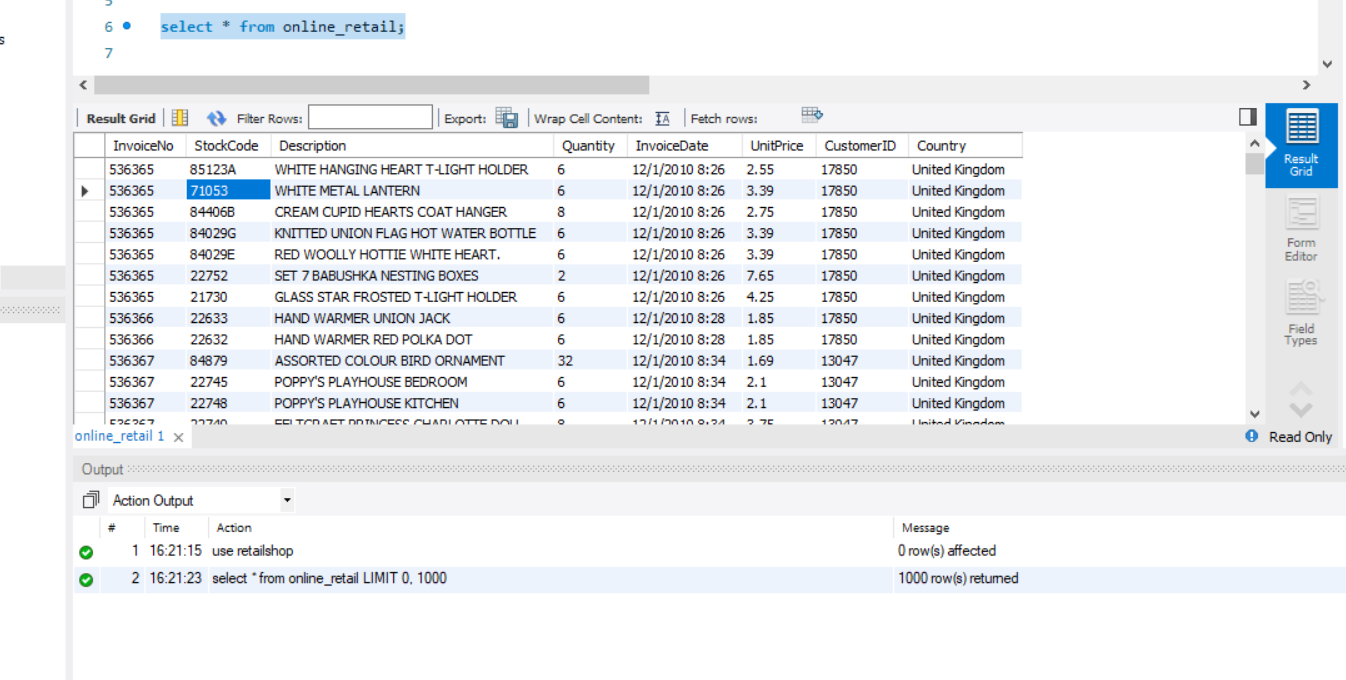
**Roll No:GIL-DASI-137**

**Section : 3**

* **Retail Shop Database Project Report**
* **Introduction**

The objective of this project is to produce insightful business analysis using the dataset of an online retail store. The dataset includes a variety of parameters, including transaction amounts, customer and product details, invoice numbers, and other data. Our goal is to extract valuable data from SQL queries, including purchase patterns, consumer behavior, product affinity, and sales trends.

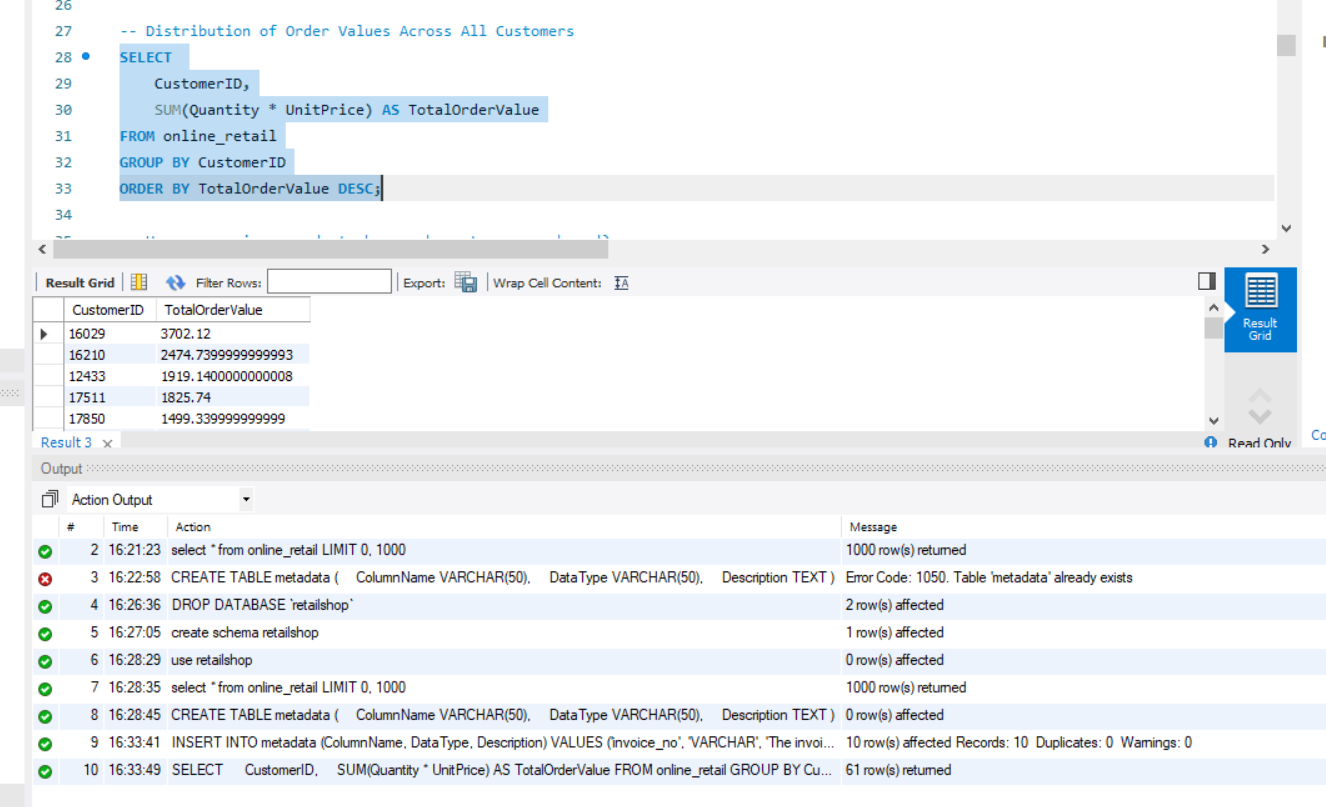
#### 2. ****Schema Setup****

To start, the retailshop schema was created and they are followed by selecting the working database using the USE retailshop; command. A metadata table was also created to maintain information about the columns and their respective data types. This metadata helps to better understand the structure of the online\_retail table. 

#### 3. ****Query Execution and Results****

* **Distribution of Order Values Across All Customers**

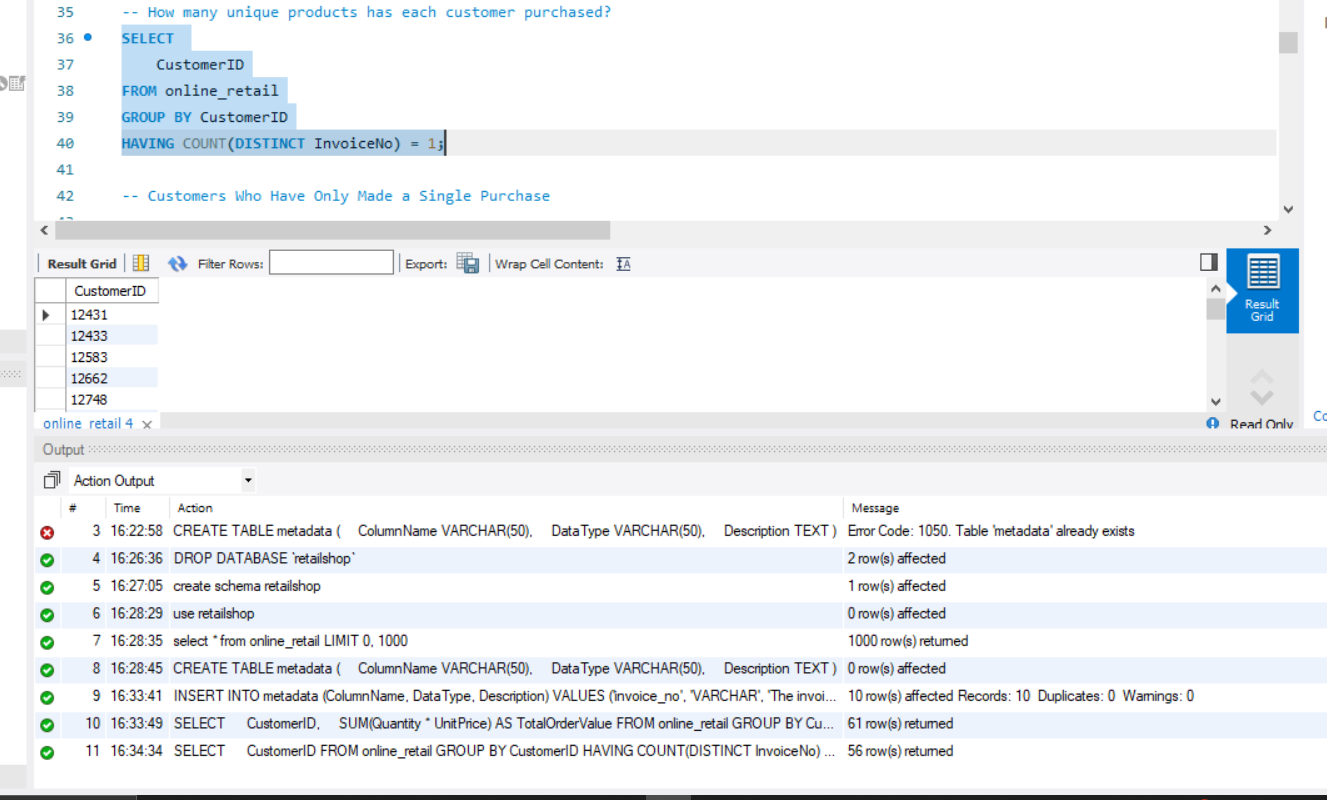
In order to determine which customers spend the most money, this query totals the purchases made by each and sorts the results in decreasing order.



The query calculates the total value of all purchases by multiplying the Quantity and UnitPrice for each transaction and grouping them by CustomerID.

* **Unique Products Purchased by Each Customer**

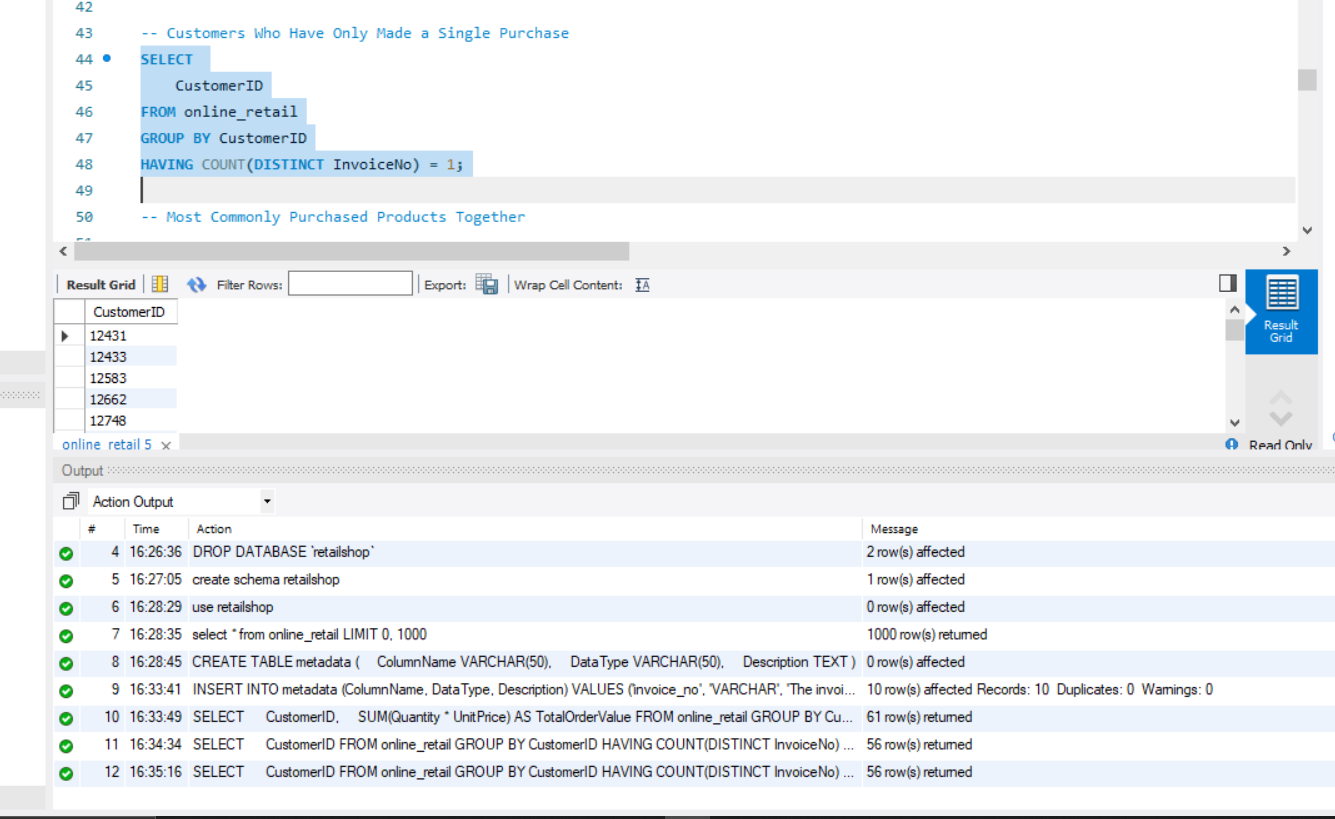
Customers that have only made a single unique purchase are identified by this query.



The HAVING clause offers insight into one-time buyers by removing clients who have made many separate purchases.

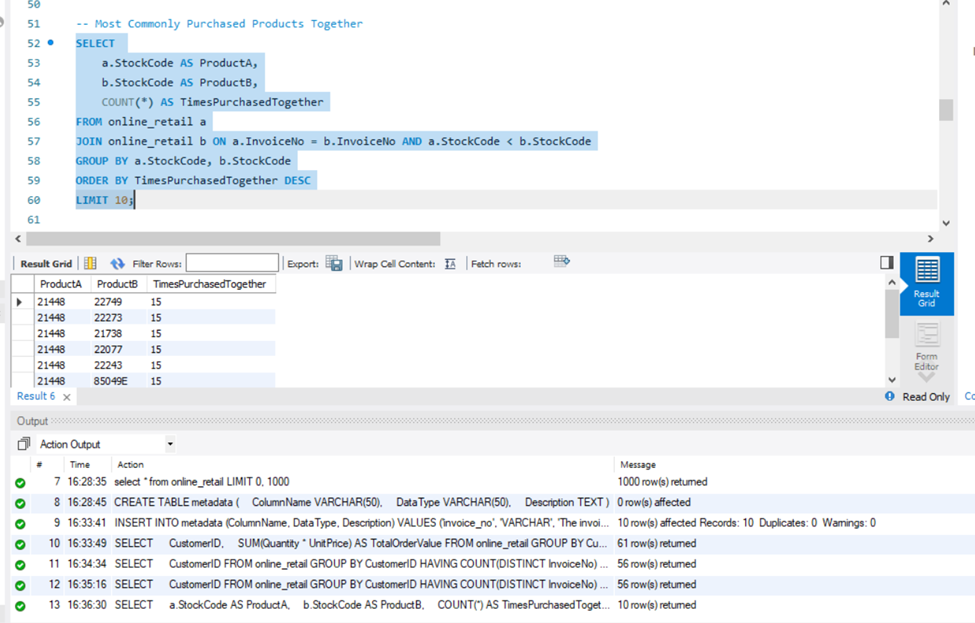
* **Customers Who Have Made a Single Purchase**

The query retrieves customers who have made only a single purchase based on the InvoiceNo field.



* **Most Commonly Purchased Products Together**

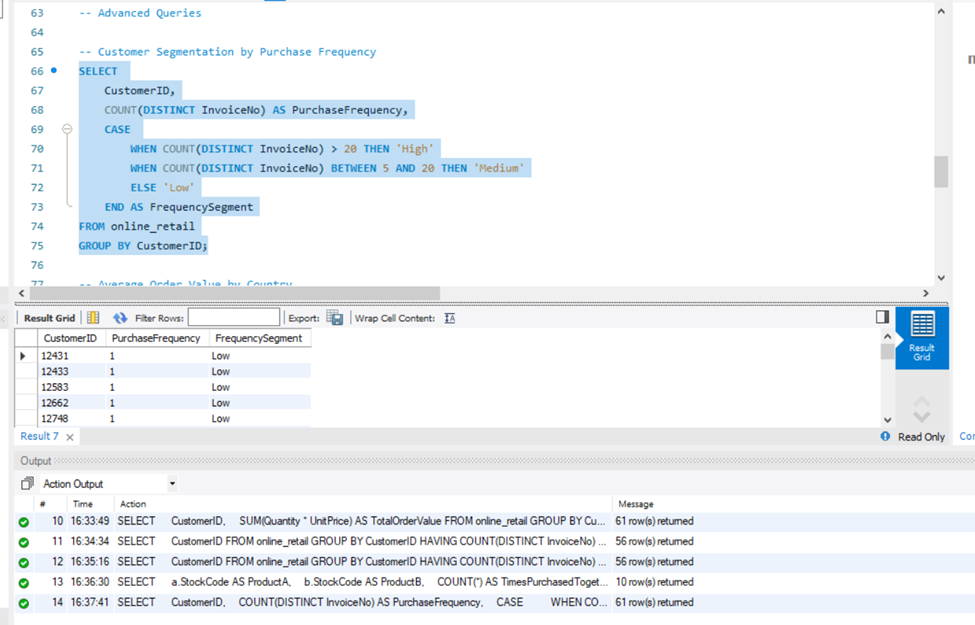
This query joins the same InvoiceNo and counts product pairings to investigate the products that are frequently purchased together.



Giving insights on product affinity and cross-selling opportunities, the data display the top 10 product pairs most frequently purchased together.

* **Customer Segmentation by Purchase Frequency**

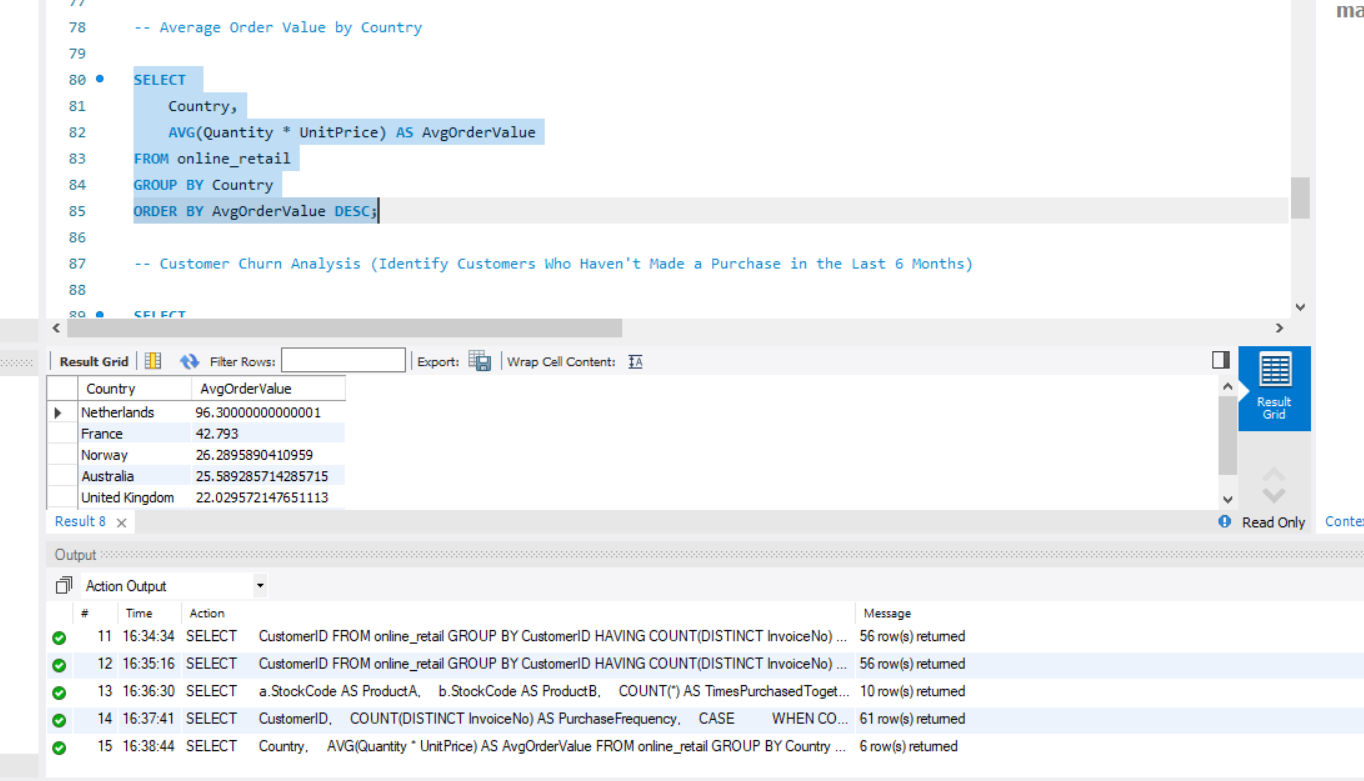
The number of purchases a customer has made divides them into segments, which is helpful for focused marketing.



The CASE statement categorizes customers into high, medium, and low-frequency segments.

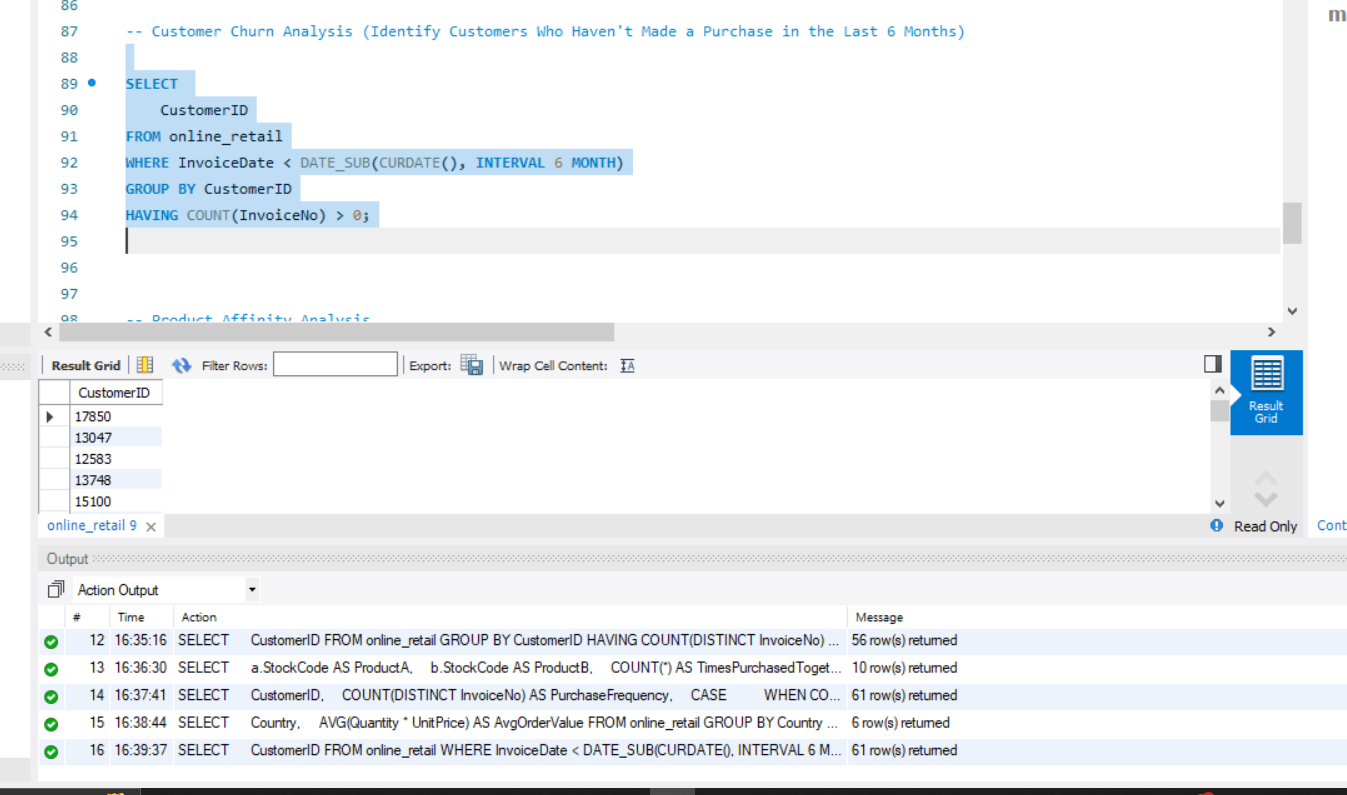
* **Average Order Value by Country**

In order to determine the locations of higher-value purchases, this query computes the average order value across several nations.



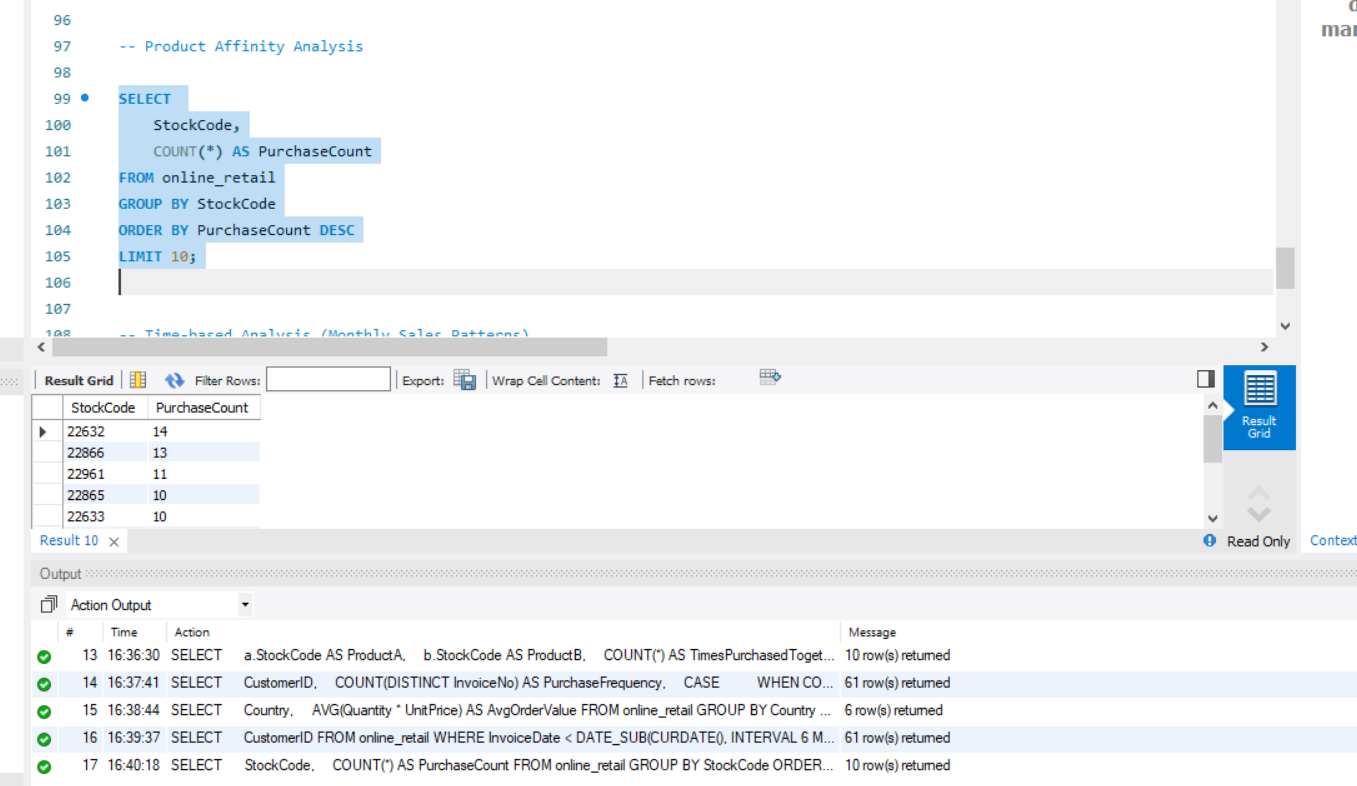
* **Customer Churn Analysis**

For the purpose of creating retention strategies, it is imperative to identify consumers who have not made a purchase within the last six months as part of a churn analysis.



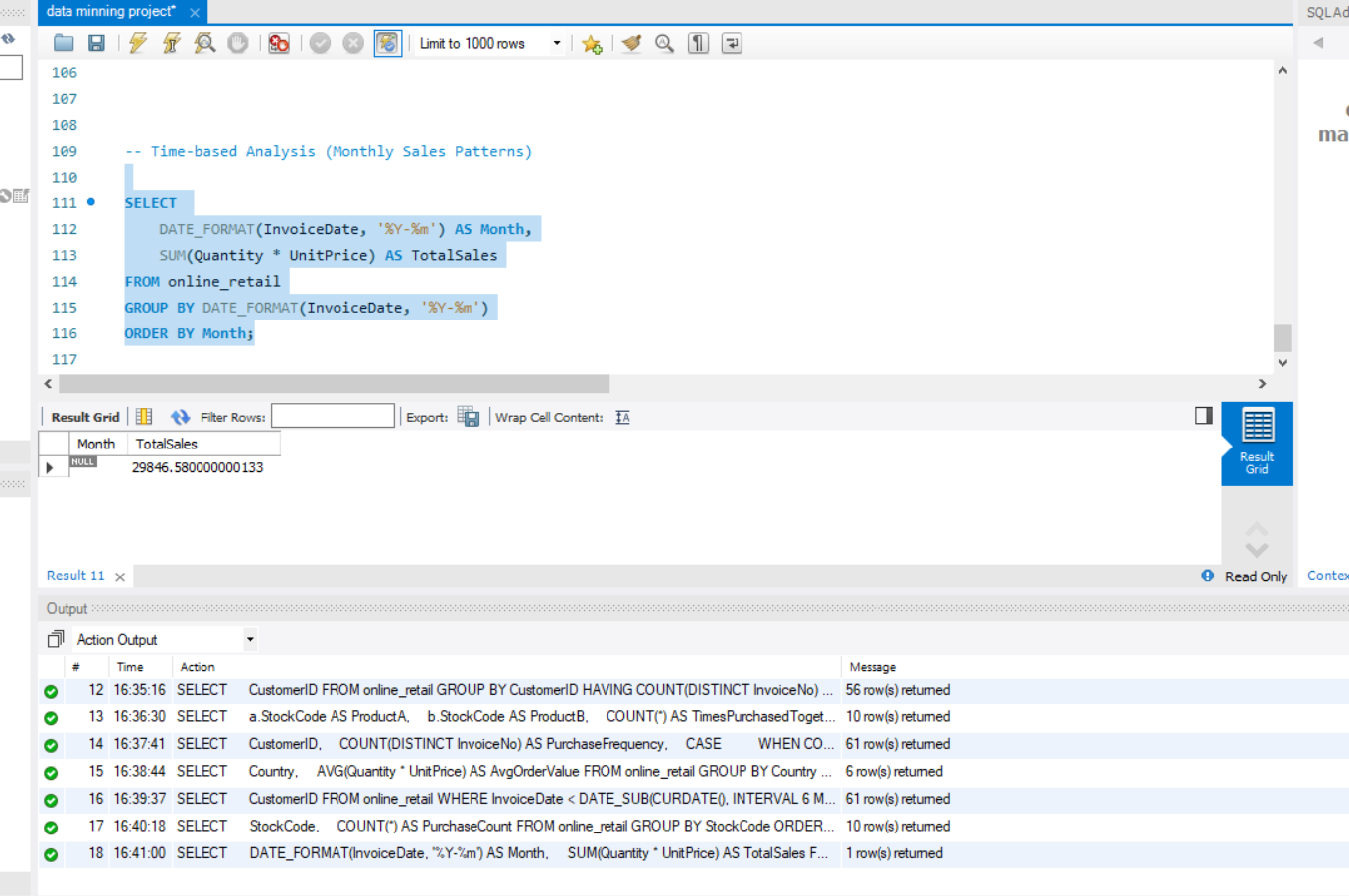
* **Product Affinity Analysis**

This query helps identify the most popular products based on the number of purchases.



* **Time-Based Analysis (Monthly Sales Patterns)**

This query aggregates sales on a monthly basis in order to detect patterns in sales over time.



Forecasting and spotting seasonality might be aided by the monthly sales performance information this query offers.

### Conclusion

This research used SQL queries on the online\_retail dataset to provide important insights into product trends, customer behavior, and sales patterns. We classified customers by frequency of purchases, found high-spending customers, and discovered common product combinations. Monthly sales patterns provided a clear picture of peak times, and customer retention prospects were identified via churn analysis. These insights have the potential to enhance product offers, marketing tactics, and overall business performance. Deeper analytics has the potential to improve customer engagement and decision-making processes in the future.

Project Link : https://github.com/hasnainasad1/Data-mining-Project

Github Link: <https://github.com/hasnainasad1>