**Rahat karim (GIL-DSAI-258)**

**Data Mining**

**Customer Segmentation and Product Analysis Using SQL**

**1. Introduction**

Customer segmentation is a vital strategy that allows organizations to categorize their clients based on factors such as demographics, shopping patterns, and purchasing behavior. The aim of this project is to use SQL to perform various analyses on a dataset that includes customer purchase data, such as invoices, products, and customer details. By segmenting customers and analyzing product affinity, businesses can tailor their marketing and sales strategies to meet the specific needs of each segment.

This report outlines the SQL queries used to analyze the dataset and the insights derived from it. We focus on beginner and advanced queries to understand customer behavior and trends, such as purchase frequency, average order value, and product affinity.

**2. Dataset Overview**

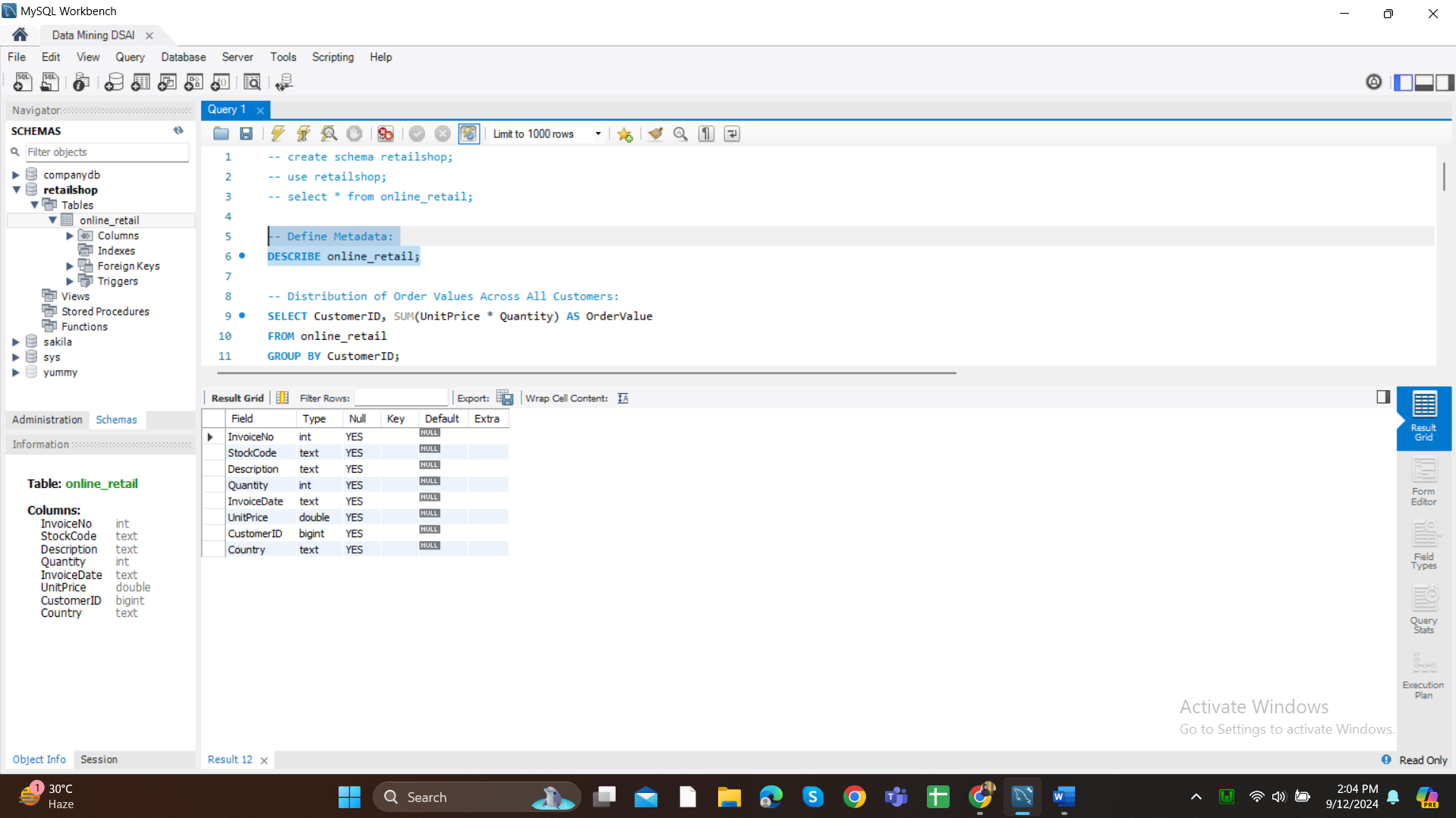
The dataset used in this project contains the following variables:

* **InvoiceNo**: Unique invoice number for each transaction
* **StockCode**: Product code for each item sold
* **Description**: Description of the products
* **Quantity**: Quantity of the product sold in each transaction
* **InvoiceDate**: Date and time of the transaction
* **UnitPrice**: Price of each product sold
* **CustomerID**: Unique identifier for each customer
* **Country**: Country where the transaction occurred

**3. SQL Queries and Insights**

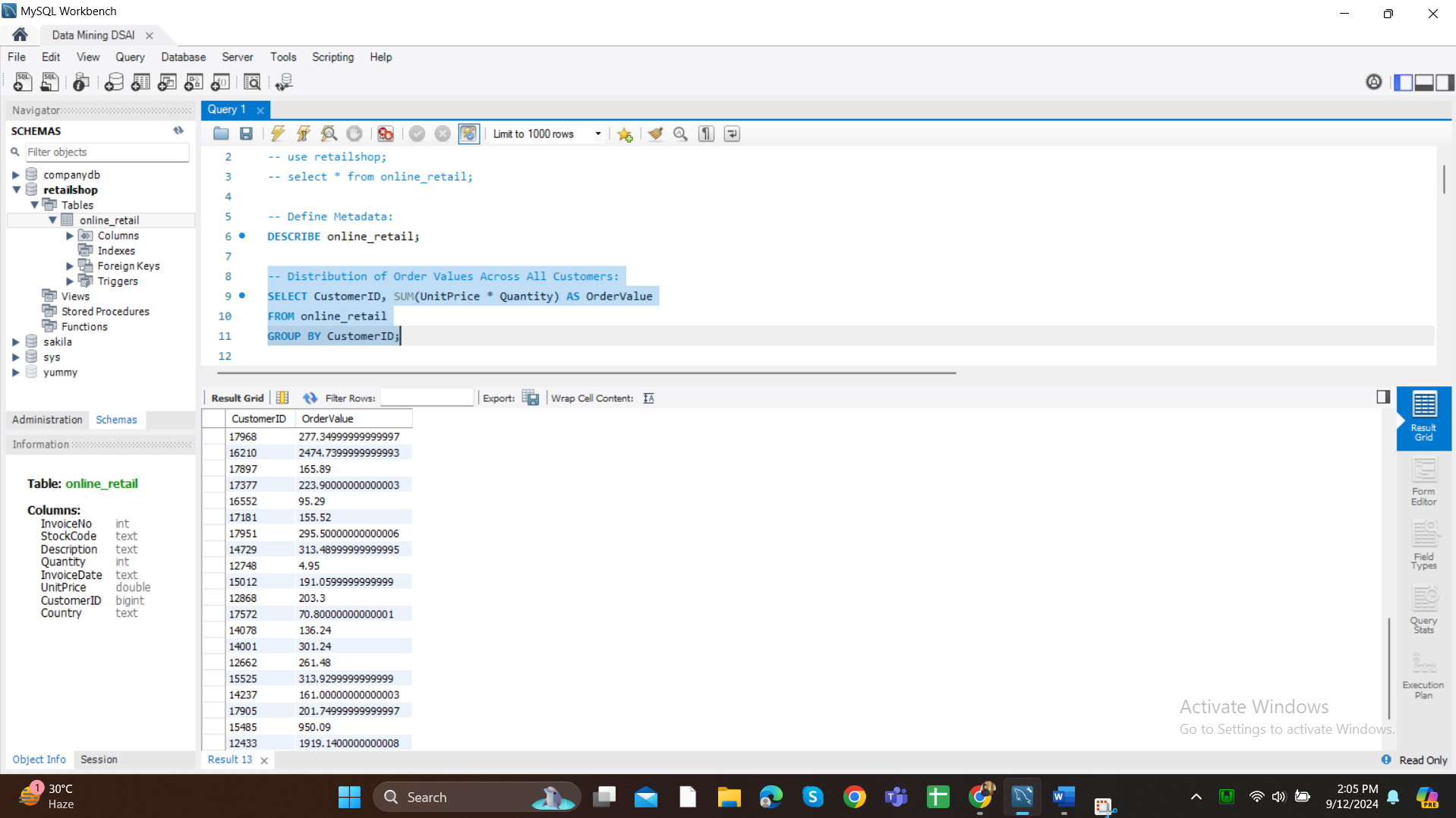
**Beginner Queries**

1. **Defining Metadata**



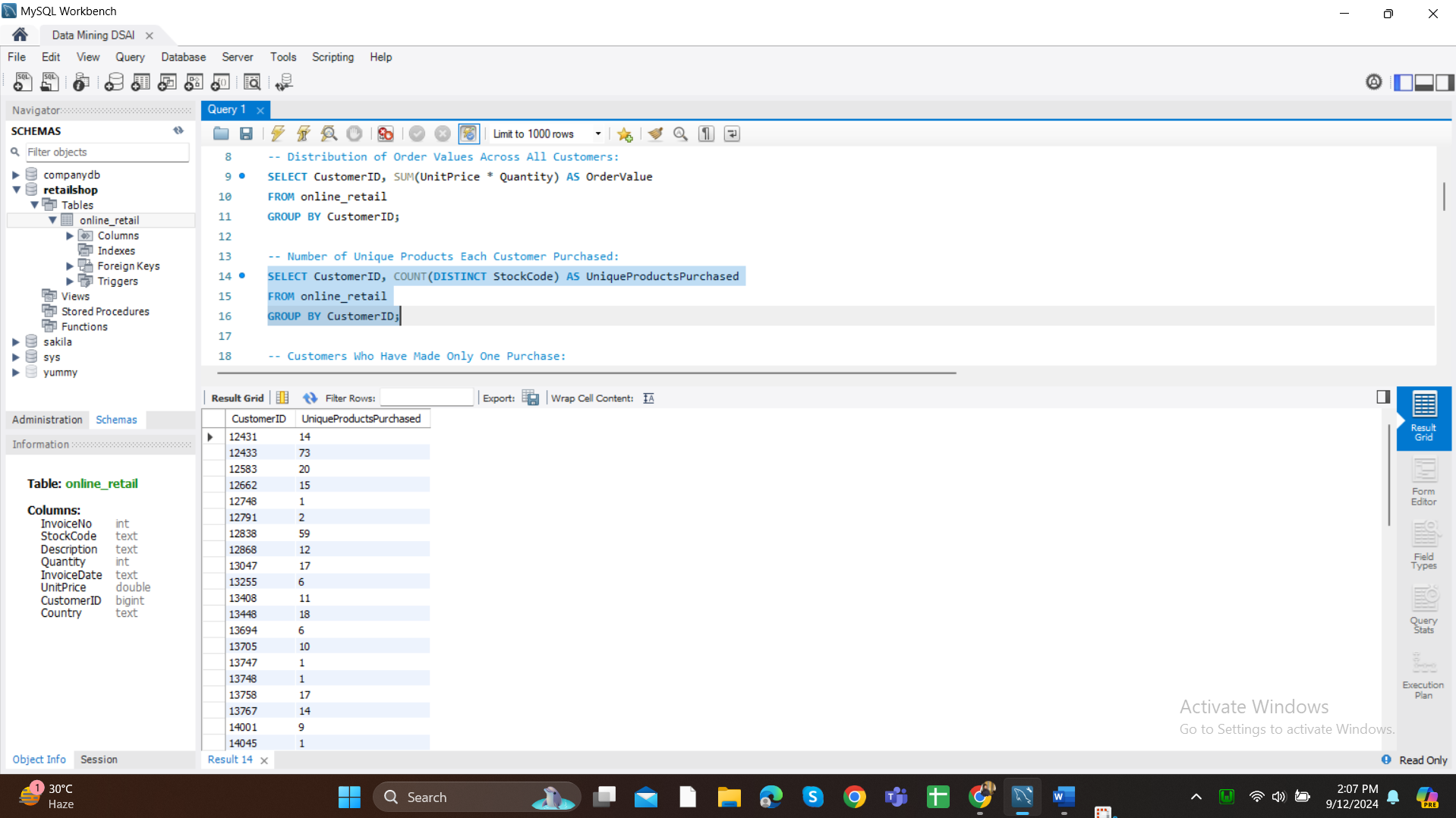
**Insight**: This query provides a description of the table structure, including column names and data types, helping us understand the dataset schema.

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



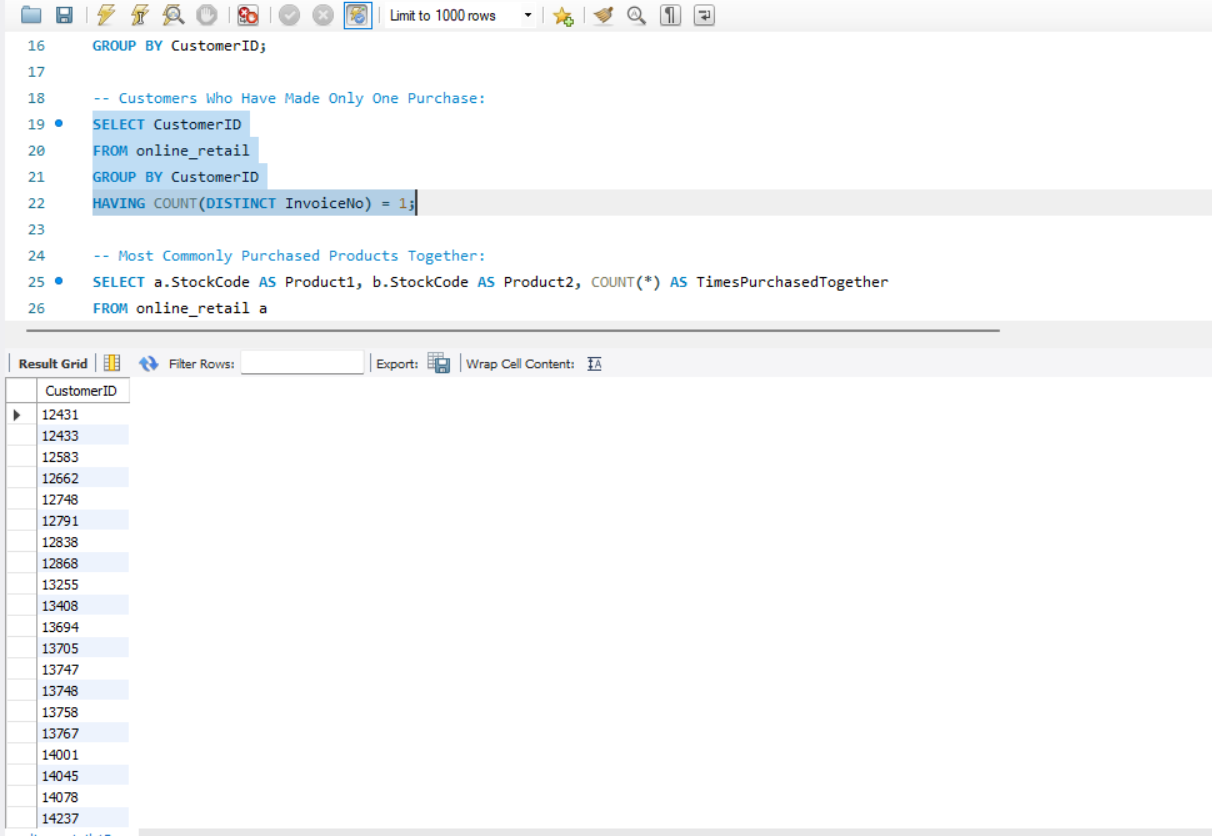
**Insight**: By calculating the total order value per customer, we can see which customers contribute the most to the company’s revenue.

1. **Unique Products Purchased by Each Customer**



**Insight**: This query helps identify how diverse each customer’s purchasing behavior is, which can be useful for product recommendation systems.

1. **Customers with Only One Purchase**



**Insight**: Identifying customers who have only made one purchase helps target them with marketing campaigns to increase retention.

1. **Most Commonly Purchased Products Together**

A screenshot of a computer

Description automatically generated

**Insight**: Understanding which products are commonly purchased together allows businesses to bundle products or make strategic recommendations.

**Advanced Queries**

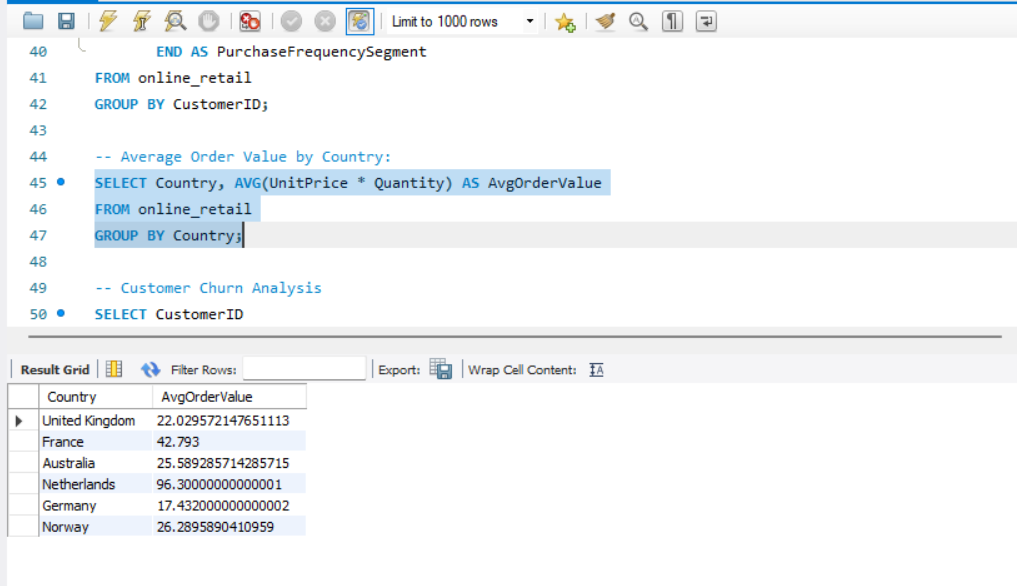
1. **Customer Segmentation by Purchase Frequency**

A screenshot of a computer

Description automatically generated

**Insight**: Segmenting customers based on purchase frequency helps identify loyal customers (high-frequency) and those who need re-engagement (low-frequency).

1. **Average Order Value by Country**



**Insight**: This query helps identify countries with the highest average order values, allowing the company to focus more marketing resources on high-value regions.

1. **Customer Churn Analysis**

A screenshot of a computer

Description automatically generated

**Insight**: Identifying customers who haven’t purchased in the last 6 months can help develop retention strategies or special offers to win them back.

1. **Product Affinity Analysis**

A screenshot of a computer

Description automatically generated

**Insight**: By determining product affinity, businesses can offer personalized bundles or cross-sell products to customers.

1. **Time-Based Analysis**

A screenshot of a computer

Description automatically generated

**Insight**: Analyzing sales trends over time helps businesses prepare for peak periods and adjust inventory or staffing levels accordingly.

**4. Customer Segmentation Strategy**

One of the key aspects of this project is customer segmentation, which categorizes customers based on purchase frequency. By segmenting customers into high, medium, and low-frequency groups, businesses can:

* **High-Frequency Customers**: Focus on loyalty programs and rewards to retain top customers.
* **Medium-Frequency Customers**: Provide additional value, such as discounts, to encourage more frequent purchases.
* **Low-Frequency Customers**: Develop targeted marketing campaigns or personalized offers to increase engagement.

Using advanced SQL queries, we also examined customer churn, helping to identify customers at risk of leaving. By analyzing the time since their last purchase, we can create targeted campaigns to reduce churn rates.

**5. Conclusion**

This project showcases the power of SQL in analyzing customer data to segment customers, understand purchasing behavior, and identify trends over time. Customer segmentation, product affinity analysis, and churn prediction are crucial strategies for enhancing customer satisfaction and driving business growth. Through the queries executed in this project, we can gain valuable insights that can shape marketing, sales, and retention strategies.

The SQL queries and insights derived from the dataset enable businesses to create more personalized experiences for their customers, improving both engagement and profitability.

**GitHub Profile Link**

[Rahat Karim GitHub](https://github.com/Rahat-Karim)

This report covers the procedural steps and insights gathered from the SQL queries in customer segmentation and product analysis.