# High Impact Skills Development Program AI & Data Science

# NAME: MUHAMMAD SAQLAIN

# ROLL NO: GIL\_DSAI\_255

# SECTION: 05

**Project Title: Online Retail Segmentation**

**Dataset : Retail\_shop.**

**MetaData:**

 Look for a dataset that contains information about customers such as demographic information, purchasing history, and customer interactions. The data set contains the following variables:

**InvoiceNo:** The invoice number for each transaction

**StockCode**: The unique code for each product sold

**Description**: The description of each product sold

**Quantity**: The quantity of each product sold in each transaction

**InvoiceDate**: The date and time of each transaction

**UnitPrice:** The price of each product sold

**CustomerID**: The unique identifier for each customer

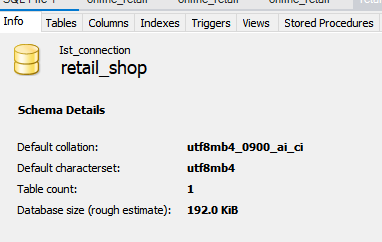
**Country**: The country where each transaction occurred

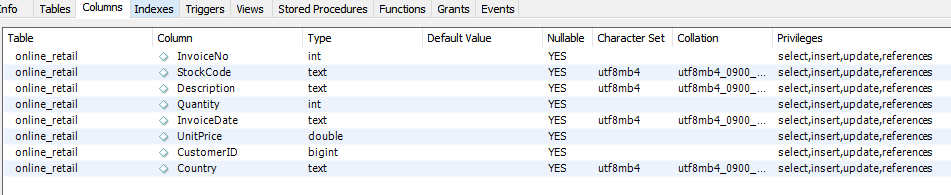
**Beginner Queries.**

Q1:        **Define meta data in mysql workbench or any other SQL tool**

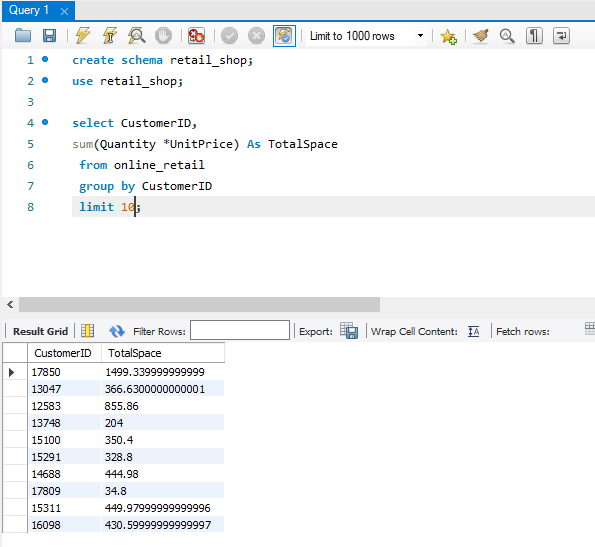
Metadata refers to data about data

**INFORMATION ;**





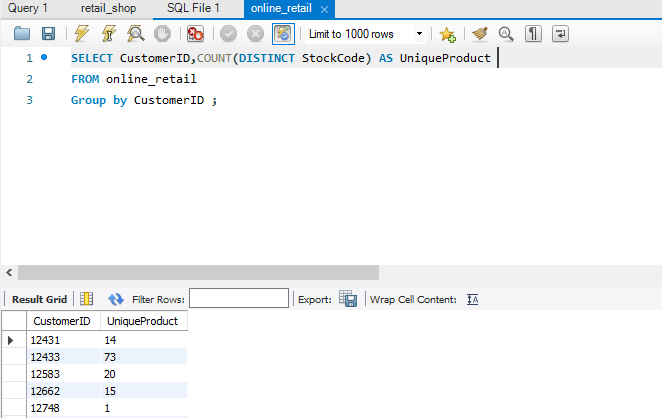
Q2:     **What is the distribution of order values across all customers in the dataset?**



The SQL query calculates the total amount of money each customer has spent on their orders

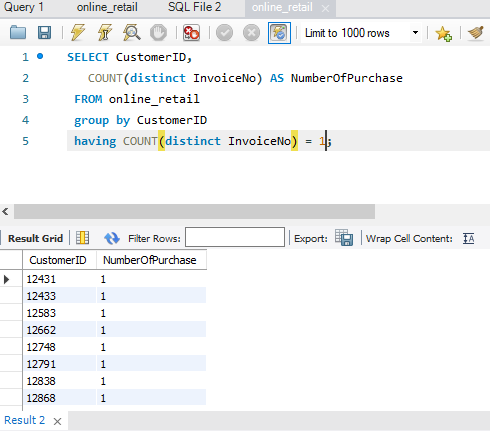
By summing up the total price of products they have purchased .

**Q3    How many unique products has each customer purchased?**



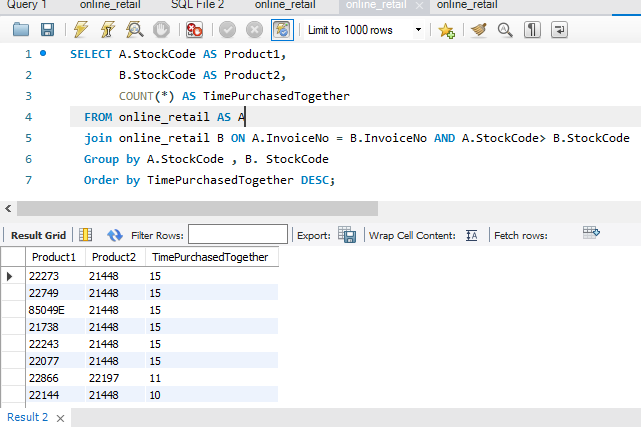
This query will output a list of customers along with the number of unique products they have purchased.

**Q4   Which customers have only made a single purchase from the company?**



This query returns the list of customer ID ‘s for those who have made only single purchase from the company. By using HAVING COUNT(DISTINCT INVOIC NO ) = 1, It only filter out customers with more than purchase ,focusing only on those who have made only single transition .

**Q5     Which products are most commonly purchased together by customers in the dataset?**

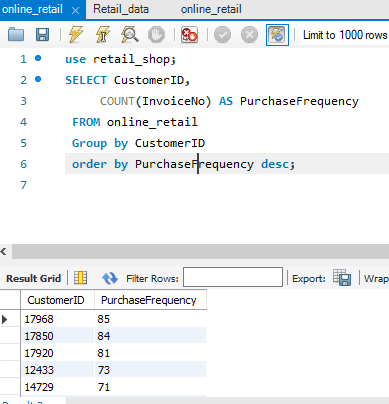


This query returns a list of product pairs (Product1 and Product2) that are frequently purchased together, sorted by how often they appear in the same invoice (TimesPurchasedTogether). This is useful for understanding customer buying habits and can be leveraged for cross-selling strategies, product placement, or inventory management.

**Advance Queries**

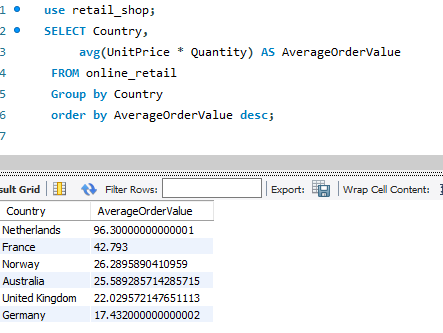
**1.      Customer Segmentation by Purchase Frequency**

Group customers into segments based on their purchase frequency, such as high, medium, and low frequency customers. This can help you identify your most loyal customers and those who need more attention.



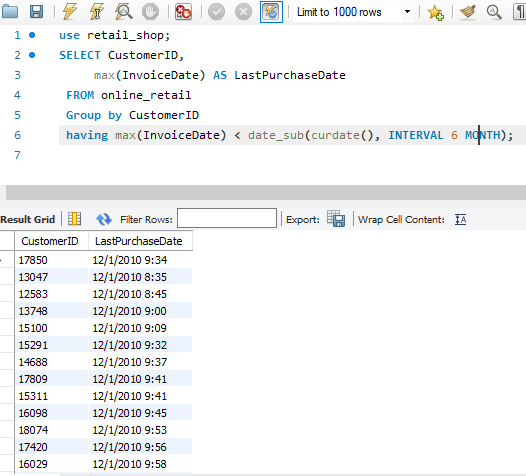
**2. Average Order Value by Country**

Calculate the average order value for each country to identify where your most valuable customers are located.



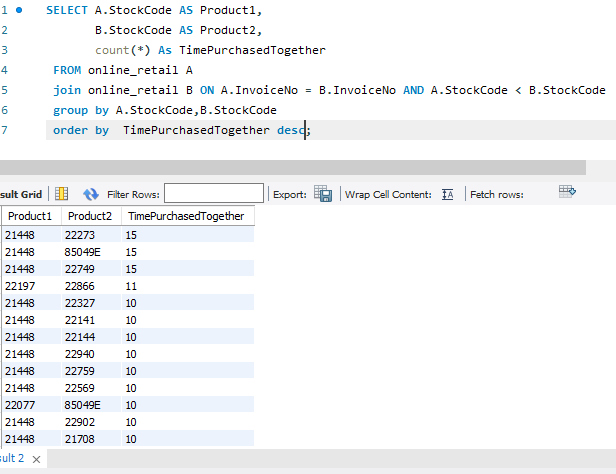
**3. Customer Churn Analysis**

 Identify customers who haven't made a purchase in a specific period (e.g., last 6 months) to assess churn.



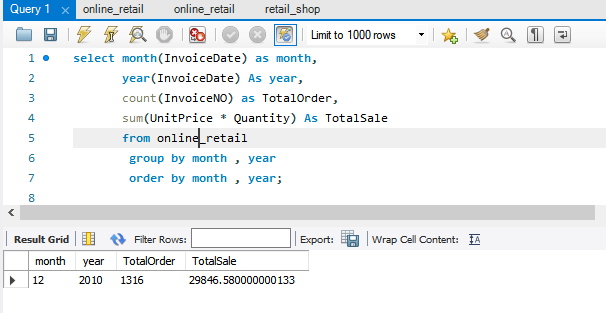
**4. Product Affinity Analysis**

Determine which products are often purchased together by calculating the correlation between product purchases.



**5. Time-based Analysis**

  Explore trends in customer behavior over time, such as monthly or quarterly sales patterns.



Time-based analysis provides valuable insights into customer behavior by exploring monthly and quarterly sales trends. By understanding these patterns, businesses can make better strategic decisions, increase profitability, and improve customer satisfaction.

**LINKS**

GITHUB

https://github.com/Saqlain772/Saqlain772.git