|  |  |
| --- | --- |
|  |  |
|  |
|  |

**High Impact Skills Development Program for Gilgit Baltistan**

**Data Mining Module Project**

**Student Name: SUMAIR IRSHAD**

**GitHub Link: https://github.com/Sumair218/**

**Project Title: Online Retail Segmentation Project**

**Introduction:**

In this report, we present the findings and insights obtained from the "Online Retail Segmentation" project. The project aimed to explore customer segmentation using data mining techniques and SQL queries. The primary objectives were to understand data mining fundamentals, apply SQL for data analysis, and implement mining concepts in a real-world dataset.

**Dataset Overview:**

The dataset used for this project contains information about customers, their purchasing behavior, and interactions with an online retail company. It consists of variables such as InvoiceNo, StockCode, Description, Quantity, InvoiceDate, UnitPrice, CustomerID, and Country.

**Procedural Aspects of Queries:**

**Beginner Queries:**

1. **What is the distribution of order values across all customers in the dataset?**

**Query:**

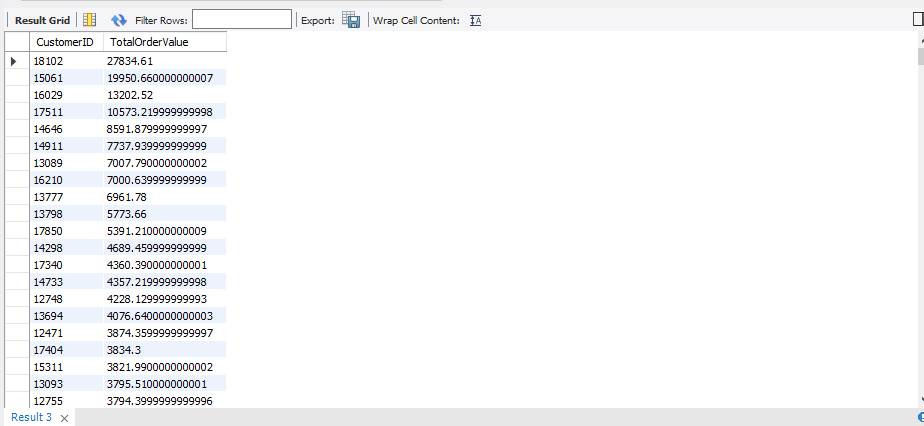
SELECT CustomerID, SUM(Quantity \* UnitPrice) AS TotalOrderValue

FROM online\_retail

GROUP BY CustomerID

ORDER BY TotalOrderValue DESC;

**Result:**

****

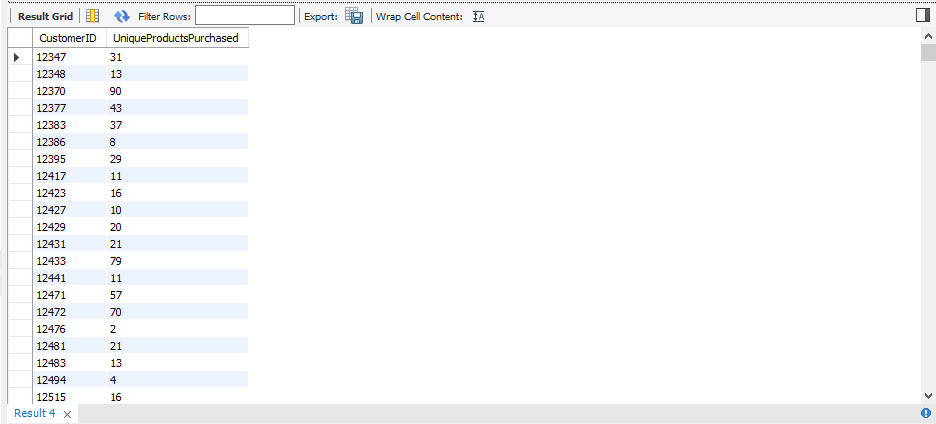
1. **How many unique products has each customer purchased?**

**Query:**

SELECT CustomerID, COUNT(DISTINCT StockCode) AS UniqueProductsPurchased

FROM online\_retail

GROUP BY CustomerID;

**Result:**

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

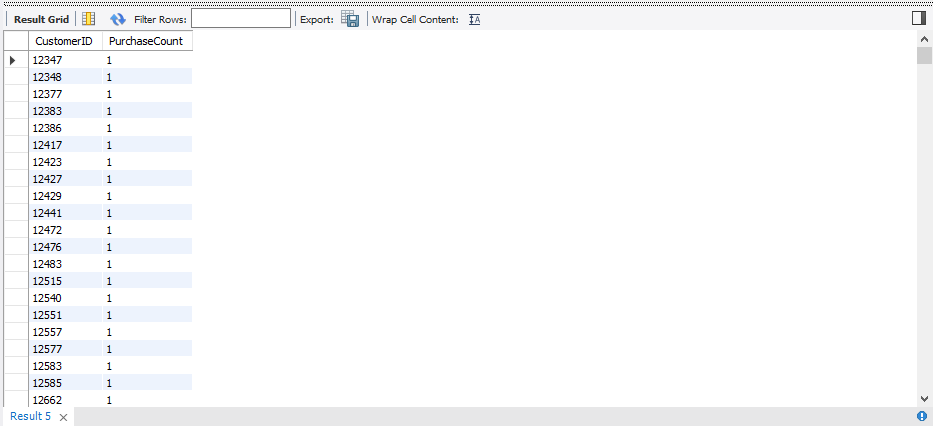
**Query:**

SELECT CustomerID, COUNT(DISTINCT InvoiceNo) AS PurchaseCount

FROM online\_retail

GROUP BY CustomerID

HAVING PurchaseCount = 1;

**Result:**

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

**Query:**

SELECT a.StockCode AS ProductA, b.StockCode AS ProductB, COUNT(\*) AS PurchaseCount

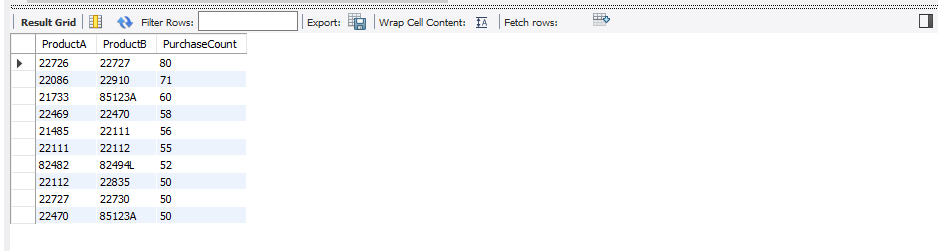
FROM online\_retail a

JOIN online\_retail b ON a.InvoiceNo = b.InvoiceNo AND a.StockCode < b.StockCode

GROUP BY ProductA, ProductB

ORDER BY PurchaseCount DESC

LIMIT 10;

**Result:**

**Advance Queries**

1. **Customer Segmentation by Purchase Frequency?**

**Qurey:**

SELECT CustomerID, COUNT(DISTINCT InvoiceNo) AS PurchaseFrequency

FROM online\_retail

GROUP BY CustomerID;

SELECT CustomerID,

CASE

WHEN PurchaseFrequency >= 10 THEN 'High'

WHEN PurchaseFrequency >= 5 THEN 'Medium'

ELSE 'Low'

END AS FrequencySegment

FROM (

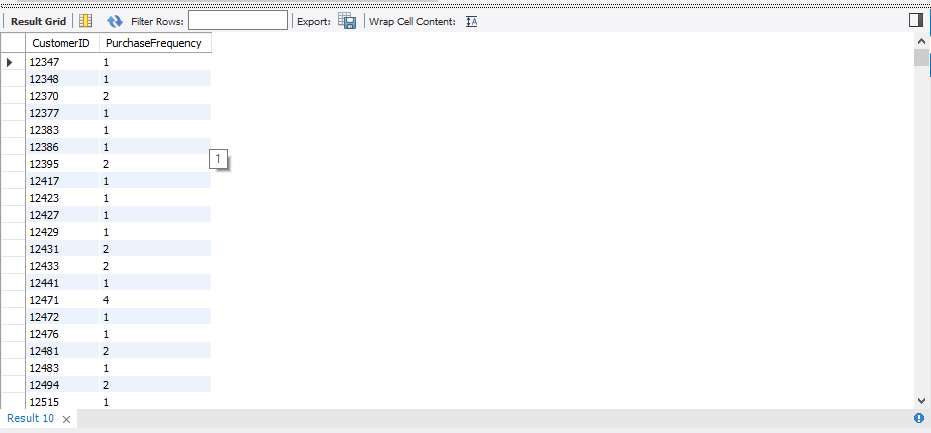
SELECT CustomerID, COUNT(DISTINCT InvoiceNo) AS PurchaseFrequency

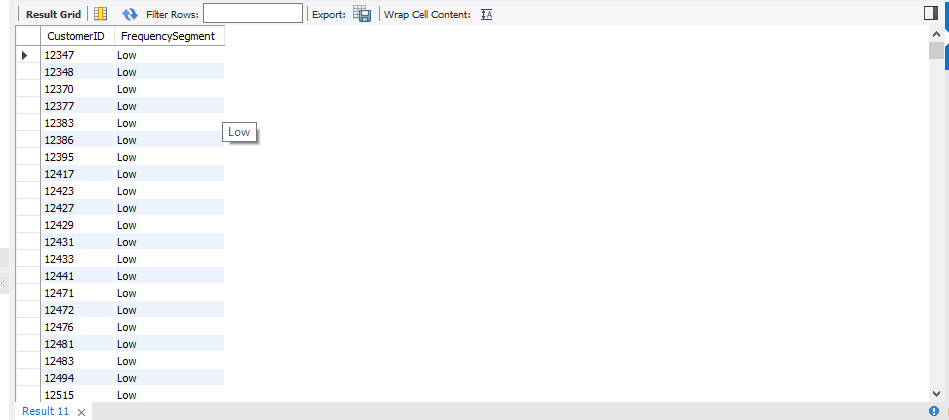
FROM online\_retail

GROUP BY CustomerID

) AS PurchaseFrequencies;

**Result:**

****

1. **Average Order Value by Country?**

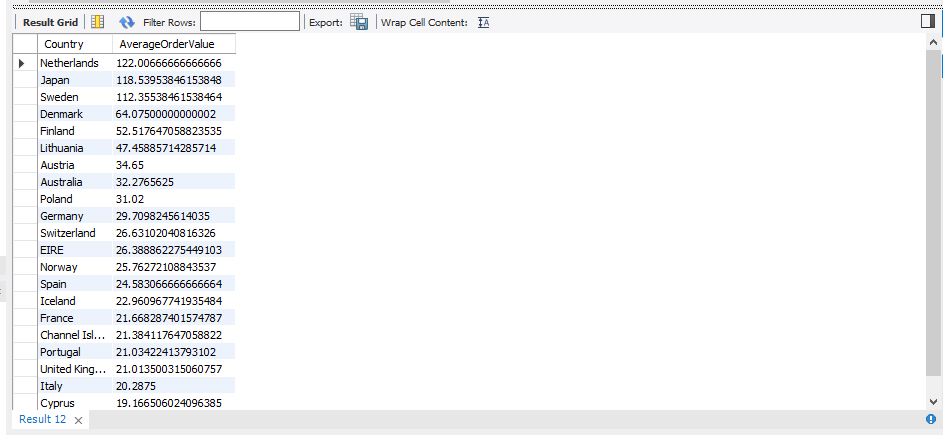
**Query:**

SELECT Country, AVG(Quantity \* UnitPrice) AS AverageOrderValue

FROM online\_retail

GROUP BY Country

ORDER BY AverageOrderValue DESC;

**Result:**

1. **Customer Churn Analysis?**

**Query:**

SELECT CustomerID

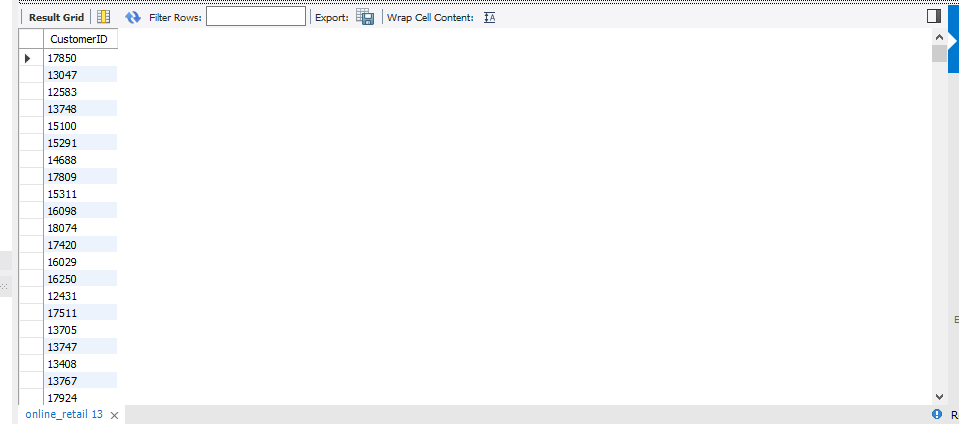
FROM online\_retail

WHERE InvoiceDate <= DATE\_SUB(NOW(), INTERVAL 6 MONTH)

GROUP BY CustomerID

HAVING MAX(InvoiceDate) < DATE\_SUB(NOW(), INTERVAL 6 MONTH);

**Result:**

****

1. **Product Affinity Analysis:**

**Query:**

CREATE TEMPORARY TABLE ProductPairs AS

SELECT A.StockCode AS ProductA, B.StockCode AS ProductB

FROM online\_retail A

JOIN online\_retail B ON A.InvoiceNo = B.InvoiceNo AND A.StockCode < B.StockCode;

SELECT ProductA, ProductB, COUNT(\*) AS Occurrences

FROM ProductPairs

GROUP BY ProductA, ProductB

ORDER BY Occurrences DESC;

SELECT

pp.ProductA,

pp.ProductB,

COUNT(\*) AS CommonOccurrences,

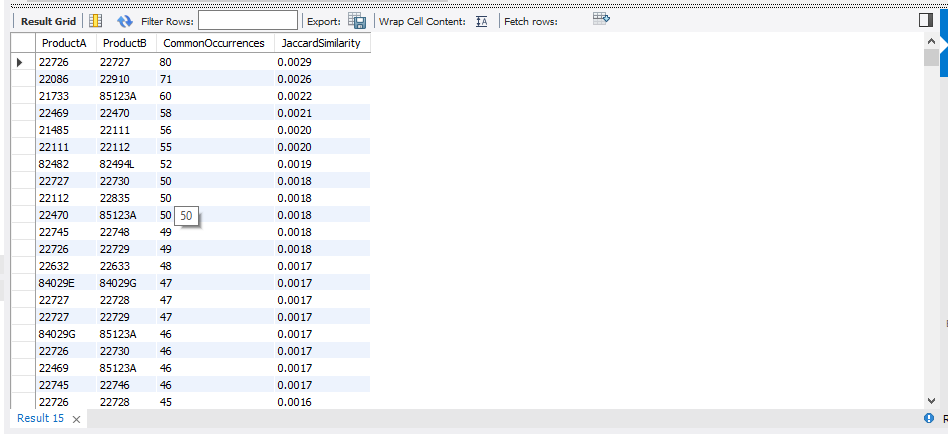
COUNT(\*) / (SELECT COUNT(\*) FROM online\_retail) AS JaccardSimilarity

FROM ProductPairs pp

GROUP BY pp.ProductA, pp.ProductB

ORDER BY CommonOccurrences DESC, JaccardSimilarity DESC;

**Result:**

1. **Time-based Analysis?**

**Query:**

SELECT

YEAR(InvoiceDate) AS Year,

MONTH(InvoiceDate) AS Month,

SUM(Quantity) AS TotalQuantitySold

FROM online\_retail

GROUP BY Year, Month

ORDER BY Year, Month;

SELECT

YEAR(InvoiceDate) AS Year,

QUARTER(InvoiceDate) AS Quarter,

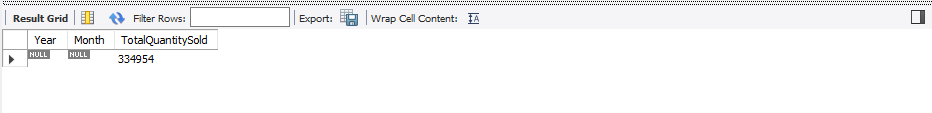
SUM(Quantity) AS TotalQuantitySold

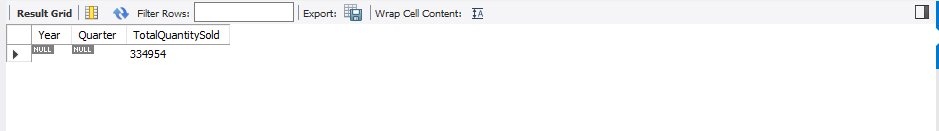
FROM online\_retail

GROUP BY Year, Quarter

ORDER BY Year, Quarter;

**Result:**

****



**Conclusion:**

The "Online Retail Segmentation" project provided valuable insights into customer segmentation and data mining concepts. Through SQL queries, we gained a better understanding of customer purchasing behavior, identified high-value customers, uncovered patterns in product purchases, and pinpointed customers with single purchases. These insights can be utilized for targeted marketing, personalized customer engagement, and improved business performance.

**Future Directions:**

This project opens the door to further exploration and enhancement. Future directions could involve implementing more advanced data mining techniques such as clustering and association rule mining. Additionally, incorporating data visualization tools could provide a more intuitive representation of customer segments and their behavior.

**Acknowledgments:**

We would like to express our gratitude to the creators of the dataset used in this project, as well as our mentors for their guidance and support throughout the project.

**References:**

1. SQL Tutorial - W3Schools.
2. MySQL Workbench Documentation.
3. Personal communication with mentors.