

**NATIONAL UNIVERSITY SCIENCE AND TECHNOLOGY (NUST)**

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

**Name: Nagina Abbas**

**Roll No: GLT-DSAI-155**

**Project Title:**

**Data Mining Project Report**

**Introduction**

* **Purpose of Customer Segmentation**: Customer segmentation helps businesses categorize clients based on demographics, purchasing patterns, and other factors, enabling them to target their marketing strategies, improve customer engagement, and enhance overall performance.
* **Dataset Overview**: The dataset used is a transactional dataset that contains information such as InvoiceNo, StockCode, Description, Quantity, InvoiceDate, UnitPrice, CustomerID, and Country. These variables allow us to analyze customer behavior and purchasing trends.

**SQL Queries and Results**

**Query 1: Distribution of Order Values Across All Customers**

sql

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

FROM online\_retail

GROUP BY CustomerID;

**Result**: A table showing each customer’s total order value.

|  |
| --- |
|  |

**Query 2: Unique Products Purchased by Each Customer**

sql

SELECT CustomerID, COUNT(DISTINCT StockCode) AS UniqueProducts

FROM online\_retail

GROUP BY CustomerID;

* + **Result**: A table indicating how many unique products each customer purchased.

|  |
| --- |
|  |

**Query 3: Single Purchase Customers**

sql

SELECT CustomerID

FROM online\_retail

GROUP BY CustomerID

HAVING COUNT(DISTINCT InvoiceNo) = 1;

* + **Result**: A list of customers who only made one purchase.

|  |
| --- |
|  |

**Query 4: Products Commonly Purchased Together**

sql

SELECT InvoiceNo, GROUP\_CONCAT(DISTINCT StockCode) AS ProductsTogether

FROM online\_retail

GROUP BY InvoiceNo;

* + **Result**: A summary of frequently purchased products.

|  |
| --- |
|  |

**Query 5: Customer Segmentation by Purchase Frequency**

sql

SELECT CustomerID,

COUNT(DISTINCT InvoiceNo) AS PurchaseFrequency,

CASE

WHEN COUNT(DISTINCT InvoiceNo) > 10 THEN 'High Frequency'

WHEN COUNT(DISTINCT InvoiceNo) BETWEEN 5 AND 10 THEN 'Medium Frequency'

ELSE 'Low Frequency'

END AS CustomerSegment

FROM online\_retail

GROUP BY CustomerID;

* + **Result**: Customers are segmented into high, medium, and low purchase frequencies.

|  |
| --- |
|  |

**Query 6: Average Order Value by Country**

sql

Copy code

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

FROM online\_retail

GROUP BY Country;

* + **Result**: The average order value for each country.

|  |
| --- |
|  |

**Challenges and Solutions**

* **Challenges**:

One challenge encountered was handling missing or incorrect customer IDs in the dataset.

* **Solutions**:

To address this, filtered out rows with null customer IDs using:

sql

SELECT \* FROM online\_retail WHERE CustomerID IS NOT NULL;

**Conclusion**

* The analysis provided meaningful insights into customer behavior. By segmenting customers by purchase frequency and average order values, I identified key customer groups (e.g., high-frequency buyers, single-purchase customers). Additionally, the product affinity analysis highlighted commonly bought-together products, which can inform cross-selling strategies.
* These insights help the business in improving customer retention, targeting promotions effectively, and enhancing sales strategies.

LINK OF PROJECT ON GITHUB:

<https://github.com/NaginaAbbas/Data-mining-project/blob/main/Data%20mining%20project%20(retail%20data%20set).sql>

THE END