



HIGH IMPACT SKILLS DEVELOPMENT PROGRAM

AI & DATA SCIENCE

PROJECT OF DATA MINING

NAME: Nadir Abbas
ROLL NO: GIL_DSAI_152
SECTION: 3rd
Email: nadir03128743177@gmail.com
Project Title: Online Retail Segmentation
Dataset: Retail shop

Overview:

Customer segmentation constitutes a widely adopted strategic approach within organizations, aimed at systematically categorizing clients by analyzing a spectrum of factors such as demographics, shopping patterns, and other discernible traits. By effectively segmenting the customer base, businesses gain the ability to tailor their offerings, communications, and marketing efforts with a higher degree of precision, ultimately leading to enhanced customer satisfaction, targeted engagement, and improved overall business performance.

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:

Invoice No: The invoice number for each transaction **Stock**

Code: The unique code for each product sold

Description: The description of each product sold

Quantity: The quantity of each product sold in each transaction

Invoice Date: The date and time of each transaction

Unit Price: The price of each product sold

Customer ID: The unique identifier for each customer

Country: The country where each transaction occurred

SQL Project Idea: Use SQL queries to answer the following questions:

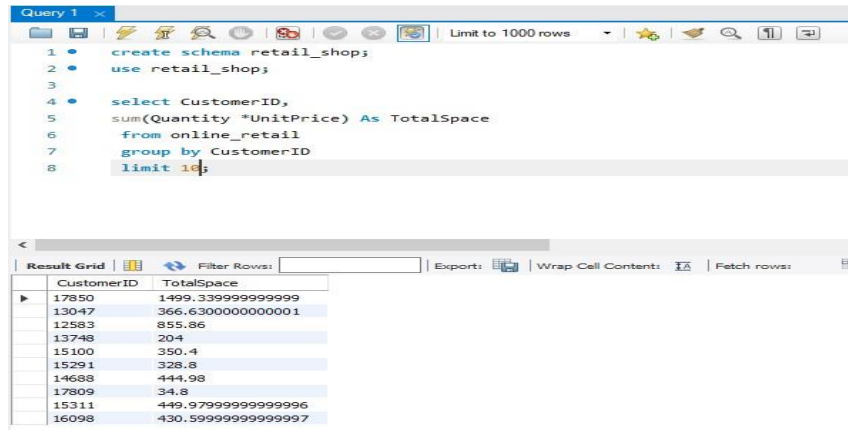
Beginner Queries.

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

Metadata in SQL tools like MySQL Workbench refers to data that describes other data within the database. This includes:

- **Schema Information:** Details about database structures, such as tables, columns, data types, and constraints.
- **Table Definitions:** Information about table names, column names, data types, and relationships between tables.
- **Index Information:** Data about indexes on tables, including their names and columns they cover. □
- **User Privileges:** Data about permissions granted to users or roles.

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



The screenshot shows a MySQL Workbench interface with a query window titled 'Query 1'. The query is as follows:

```
1 • create schema retail_shop;
2 • use retail_shop;
3
4 • select CustomerID,
5     sum(Quantity *UnitPrice) As TotalSpace
6     from online_retail
7     group by CustomerID
8     limit 10;
```

Below the query window, the 'Result Grid' is displayed, showing the results of the query. The columns are 'CustomerID' and 'TotalSpace'.

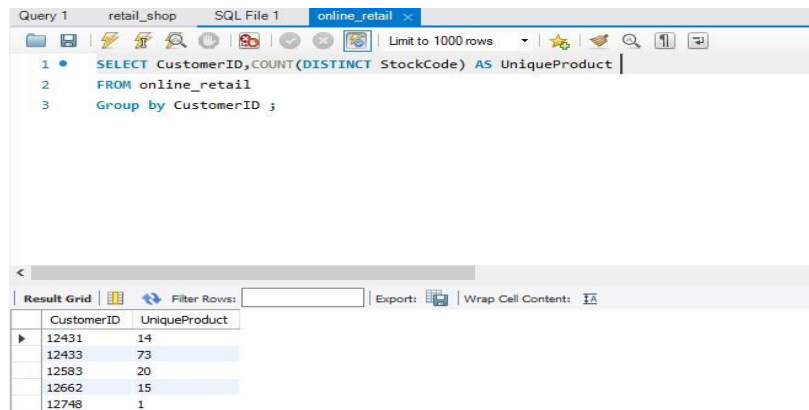
CustomerID	TotalSpace
17850	1499.3399999999999
13047	366.63000000000001
12583	855.86
13748	204
15100	350.4
15291	328.8
14688	444.98
17809	34.8
15311	449.97999999999996
16098	430.59999999999997

SUMMARY:

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?



The screenshot shows a MySQL Workbench interface with a query window titled 'Query 1'. The query is as follows:

```
1 • SELECT CustomerID, COUNT(DISTINCT StockCode) AS UniqueProduct
2     FROM online_retail
3     Group by CustomerID ;
```

Below the query window, the 'Result Grid' is displayed, showing the results of the query. The columns are 'CustomerID' and 'UniqueProduct'.

CustomerID	UniqueProduct
12431	14
12433	73
12583	20
12662	15
12748	1

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

Query 1 online_retail SQL File 2 online_retail online_retail online_retail

```

1 • SELECT CustomerID,
2       COUNT(distinct InvoiceNo) AS NumberOfPurchase
3       FROM online_retail
4       group by CustomerID
5       having COUNT(distinct InvoiceNo) = 1;

```

Result Grid | Filter Rows: | Export: | Wrap Cell Content: |

CustomerID	NumberOfPurchase
12431	1
12433	1
12583	1
12662	1
12748	1
12791	1
12838	1
12868	1

Result 2 x

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?

Query 1 online_retail SQL File 2 online_retail online_retail online_retail online_retail

```

1 • SELECT A.StockCode AS Product1,
2       B.StockCode AS Product2,
3       COUNT(*) AS TimePurchasedTogether
4       FROM online_retail AS A
5       join online_retail B ON A.InvoiceNo = B.InvoiceNo AND A.StockCode > B.StockCode
6       Group by A.StockCode , B. StockCode
7       Order by TimePurchasedTogether DESC;

```

Result Grid | Filter Rows: | Export: | Wrap Cell Content: | Fetch rows: |

Product1	Product2	TimePurchasedTogether
22273	21448	15
22749	21448	15
85049E	21448	15
21738	21448	15
22243	21448	15
22077	21448	15
22866	22197	11
22144	21448	10

Result 2 x

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.

online_retail x Retail_data online_retail

```

1 • use retail_shop;
2 • SELECT CustomerID,
3       COUNT(InvoiceNo) AS PurchaseFrequency
4       FROM online_retail
5       Group by CustomerID
6       order by PurchaseFrequency desc;

```

Result Grid | Filter Rows: | Export: | Wrap

CustomerID	PurchaseFrequency
17968	85
17850	84
17920	81
12433	73
14729	71

2. Average Order Value by Country

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

```

1 • use retail_shop;
2 • SELECT Country,
3       avg(UnitPrice * Quantity) AS AverageOrderValue
4   FROM online_retail
5   Group by Country
6   order by AverageOrderValue desc;
7

```

Country	AverageOrderValue
Netherlands	96.30000000000001
France	42.793
Norway	26.2895890410959
Australia	25.589285714285715
United Kingdom	22.029572147651113
Germany	17.432000000000002

3. Customer Churn Analysis

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

```

1 • use retail_shop;
2 • SELECT CustomerID,
3       max(InvoiceDate) AS LastPurchaseDate
4   FROM online_retail
5   Group by CustomerID
6   having max(InvoiceDate) < date_sub(curdate(), INTERVAL 6 MONTH);
7

```

CustomerID	LastPurchaseDate
17850	12/1/2010 9:34
13047	12/1/2010 8:35
12583	12/1/2010 8:45
13748	12/1/2010 9:00
15100	12/1/2010 9:09
15291	12/1/2010 9:32
14688	12/1/2010 9:37
17809	12/1/2010 9:41
15311	12/1/2010 9:41
16098	12/1/2010 9:45
18074	12/1/2010 9:53
17420	12/1/2010 9:56
16029	12/1/2010 9:58

4. Product Affinity Analysis

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

```

1 • SELECT A.StockCode AS Product1,
2       B.StockCode AS Product2,
3       count(*) As TimePurchasedTogether
4   FROM online_retail A
5   join online_retail B ON A.InvoiceNo = B.InvoiceNo AND A.StockCode < B.StockCode
6   group by A.StockCode,B.StockCode
7   order by TimePurchasedTogether desc;

```

Product1	Product2	TimePurchasedTogether
21448	22273	15
21448	85049E	15
21448	22749	15
22197	22866	11
21448	22327	10
21448	22141	10
21448	22144	10
21448	22940	10
21448	22759	10
21448	22569	10
22077	85049E	10
21448	22902	10
21448	21708	10

5. Time-based Analysis

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

