Customer Segmentation using SQL

Data Analysis Project using the Online Retail Dataset

By Maira Nawaz

Introduction

Customer segmentation is a key strategy used by businesses to understand consumer behavior.

SQL was used to extract insights from the *Online Retail Dataset*.

This analysis helps identify valuable customers, purchasing patterns, and opportunities for targeted marketing.

Objective

The objective of this project is to perform customer segmentation and behavioral analysis using SQL on the *Online Retail Dataset*. The goal is to extract meaningful insights that can support strategic decision-making for improving customer retention, product recommendations, and marketing efforts.

Key Goals:

Analyze order value distribution to identify high-revenue transactions.

Segment customers based on their purchase frequency and spending behavior.

Detect product return trends to reduce losses and improve customer satisfaction.

Identify peak purchasing times to optimize marketing campaigns.

Understand country-wise revenue contribution for global strategy planning.

Dataset Description

Fields Used:

- ❑ InvoiceNo – Unique invoice number
- ❑ StockCode – Unique product code
- ❑ Description – Product name
- ❑ Quantity – Units bought
- ❑ InvoiceDate – Transaction date & time
- ❑ UnitPrice – Price per unit
- ❑ CustomerID – Unique customer identifier
- ❑ Country – Country of purchase

Key Analysis Questions

1. What is the distribution of order values across customers?
2. How many unique products has each customer purchased?
3. Which customers only made a single purchase?
4. Which products are most commonly purchased together?
5. Who are the top 5 highest-spending customers?
6. Which countries generate the most revenue?
7. What are the peak purchasing hours or days?
8. What products have the highest return rates?

Creating Database

```
create database customer_segmentaion
```

Creating OnlineRetail Table

```
3  CREATE TABLE OnlineRetail (
4      InvoiceNo VARCHAR(20),
5      StockCode VARCHAR(20),
6      Description TEXT,
7      Quantity INT,
8      InvoiceDate DATETIME,
9      UnitPrice DECIMAL(10, 2),
10     CustomerID INT,
11     Country VARCHAR(50)
```

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

InvoiceNo	StockCode	Description	Quantity	InvoiceDate	UnitPrice	CustomerID	Country
536365	85123A	WHITE HANGING HEART T-LIGHT HOLDER	6	2010-12-01 08:26:00	2.55	17850	United Kingdom
536365	71053	WHITE METAL LANTERN	6	2010-12-01 08:26:00	3.39	17850	United Kingdom
536365	84406B	CREAM CUPID HEARTS COAT HANGER	8	2010-12-01 08:26:00	2.75	17850	United Kingdom
536365	84029G	KNITTED UNION FLAG HOT WATER BOTTLE	6	2010-12-01 08:26:00	3.39	17850	United Kingdom
536365	84029E	RED WOOLLY HOTTIE WHITE HEART.	6	2010-12-01 08:26:00	3.39	17850	United Kingdom
536365	22752	SET 7 BABUSHKA NESTING BOXES	2	2010-12-01 08:26:00	7.65	17850	United Kingdom
536365	21730	GLASS STAR FROSTED T-LIGHT HOLDER	6	2010-12-01 08:26:00	4.25	17850	United Kingdom
536366	22633	HAND WARMER UNION JACK	6	2010-12-01 08:28:00	1.85	17850	United Kingdom
536366	22632	HAND WARMER RED POLKA DOT	6	2010-12-01 08:28:00	1.85	17850	United Kingdom
536367	84879	ASSORTED COLOUR BIRD ORNAMENT	32	2010-12-01 08:34:00	1.69	13047	United Kingdom
536367	22745	POPPY'S PLAYHOUSE BEDROOM	6	2010-12-01 08:34:00	2.10	13047	United Kingdom
536367	22748	POPPY'S PLAYHOUSE KITCHEN	6	2010-12-01 08:34:00	2.10	13047	United Kingdom
536367	22749	FELTCRAFT PRINCESS CHARLOTTE DOLL	8	2010-12-01 08:34:00	3.75	13047	United Kingdom
536367	22310	IVORY KNITTED MUG COSY	6	2010-12-01 08:34:00	1.65	13047	United Kingdom
536367	84969	BOX OF 6 ASSORTED COLOUR TEASPOONS	6	2010-12-01 08:34:00	4.25	13047	United Kingdom
536367	22623	BOX OF VINTAGE JIGSAW BLOCKS	3	2010-12-01 08:34:00	4.95	13047	United Kingdom
536367	22622	BOX OF VINTAGE ALPHABET BLOCKS	2	2010-12-01 08:34:00	9.95	13047	United Kingdom
536367	21754	HOME BUILDING BLOCK WORD	3	2010-12-01 08:34:00	5.95	13047	United Kingdom
536367	21755	LOVE BUILDING BLOCK WORD	3	2010-12-01 08:34:00	5.95	13047	United Kingdom

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

```sql
SELECT
    CustomerID,
    InvoiceNo,
    Sum(Quantity * UnitPrice) as TotalOrderValue
FROM
    OnlineRetail
GROUP BY CustomerID, InvoiceNo
```

Result Grid | Filter Rows:

CustomerID	InvoiceNo	TotalOrderValue
17850	536365	139.12
17850	536366	22.20
13047	536367	278.73
13047	536368	70.05
13047	536369	17.85
12583	536370	855.86
13748	536371	204.00
17850	536372	22.20
17850	536373	259.86
15100	536374	350.40
17850	536375	259.86
15291	536376	328.80
17850	536377	22.20
14688	536378	444.98
17809	536380	34.80
15311	536381	449.98
14527	C536379	-27.50
16098	536382	430.60
15311	C536383	-4.65

2. How many unique products has each customer purchased?

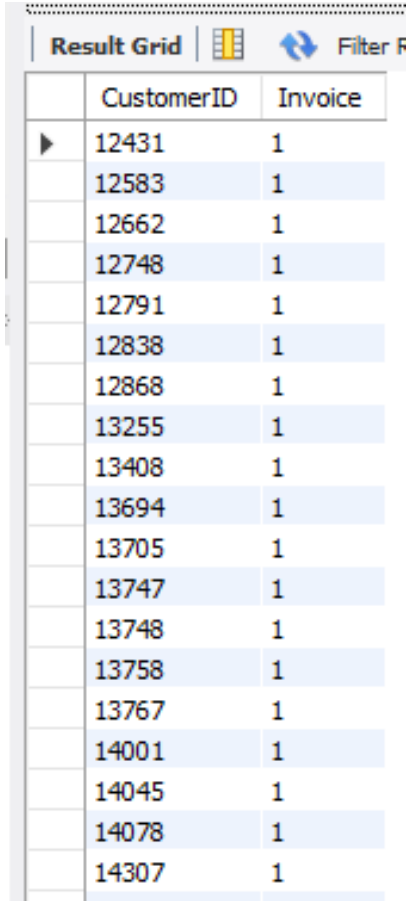
```sql
SELECT
    CustomerID,
    COUNT(DISTINCT StockCode) as Unique_Product_each_Customer_Purchased
FROM
    OnlineRetail
GROUP BY CustomerID
```

Result Grid | Filter Rows: | Export:

CustomerID	Unique_Product_each_Customer_Purchased
12431	14
12583	20
12662	15
12748	1
12791	2
12838	59
12868	12
13047	17
13255	6
13408	11
13448	18
13694	6
13705	10
13747	1
13748	1
13758	17
13767	14
14001	9
14045	1

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

```sql
SELECT
    CustomerID,
    COUNT(DISTINCT InvoiceNo )  AS Invoice
FROM
    OnlineRetail
GROUP BY CustomerID
HAVING Invoice  = 1
```

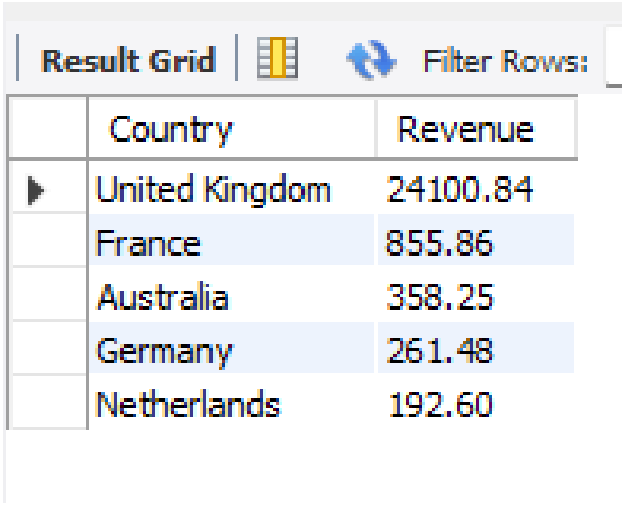


Result Grid | Filter

CustomerID	Invoice
12431	1
12583	1
12662	1
12748	1
12791	1
12838	1
12868	1
13255	1
13408	1
13694	1
13705	1
13747	1
13748	1
13758	1
13767	1
14001	1
14045	1
14078	1
14307	1

4. Which countries generate the most revenue(top 5)?

```sql
SELECT
    Country,
    Sum(Quantity * UnitPrice) as Revenue
FROM
    OnlineRetail
GROUP BY Country
order by  Revenue desc
limit 5
```



Result Grid | Filter Rows:

Country	Revenue
United Kingdom	24100.84
France	855.86
Australia	358.25
Germany	261.48
Netherlands	192.60

5. Which customers are the most valuable (top 5 spenders)?

```sql
SELECT
    CustomerID,
    Sum(Quantity * UnitPrice) as Customer_Spend
FROM
    OnlineRetail
GROUP BY CustomerID
order by  Customer_Spend desc
limit 5
```

Result Grid | Filter Rows:

CustomerID	Customer_Spend
16029	3702.12
16210	2474.74
17511	1825.74
17850	1499.34
13408	1024.68

6. Which products have the highest return rate?

```sql
SELECT
    StockCode,
    InvoiceNo,
    Description,
    SUM(Quantity) AS TotalReturned
FROM
    OnlineRetail
WHERE
    InvoiceNo LIKE 'C%'
GROUP BY
    StockCode, Description, InvoiceNo
ORDER BY
    TotalReturned ASC
LIMIT 10;
```

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

StockCode	InvoiceNo	Description	TotalReturned
21984	C536391	PACK OF 12 PINK PAISLEY TISSUES	-24
21983	C536391	PACK OF 12 BLUE PAISLEY TISSUES	-24
21980	C536391	PACK OF 12 RED RETROSPOT TISSUES	-24
22553	C536391	PLASTERS IN TIN SKULLS	-24
22556	C536391	PLASTERS IN TIN CIRCUS PARADE	-12
21484	C536391	CHICK GREY HOT WATER BOTTLE	-12
22557	C536391	PLASTERS IN TIN VINTAGE PAISLEY	-12
22960	C536506	JAM MAKING SET WITH JARS	-6
D	C536379	Discount	-1
35004C	C536383	SET OF 3 COLOURED FLYING DUCKS	-1

7. What are the peak purchasing hours?

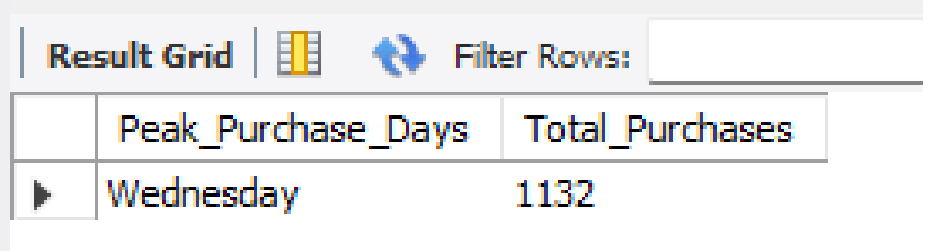
```sql
SELECT
    HOUR(InvoiceDate) AS Peak_Purchase_Hour,
    COUNT(*) AS Total_Purchases
FROM
    OnlineRetail
GROUP BY Peak_Purchase_Hour
ORDER BY Total_Purchases DESC
```

Result Grid | Filter Rows:

Peak_Purchase_Hour	Total_Purchases
12	421
11	369
9	151
10	121
8	46
13	24

8. What are the peak purchasing days?

```
SELECT
    DAYNAME(InvoiceDate) AS Peak_Purchase_Days,
    COUNT(*) AS Total_Purchases
FROM
    OnlineRetail
GROUP BY Peak_Purchase_Days
ORDER BY Total_Purchases DESC
```



Result Grid | Filter Rows:

Peak_Purchase_Days	Total_Purchases
Wednesday	1132

Conclusion

This customer segmentation project using SQL provided valuable insights into customer behaviors and sales patterns.

Key findings include:

- ❑ High-value customers and most frequently purchased products were identified.
- ❑ Specific countries contribute the majority of the revenue.
- ❑ Return trends and single-time buyers were detected, indicating areas for customer experience improvement.
- ❑ Peak purchasing hours and days were determined for targeted promotions.