Uncovering Product Return Patterns with SQL

E-Commerce Case Study

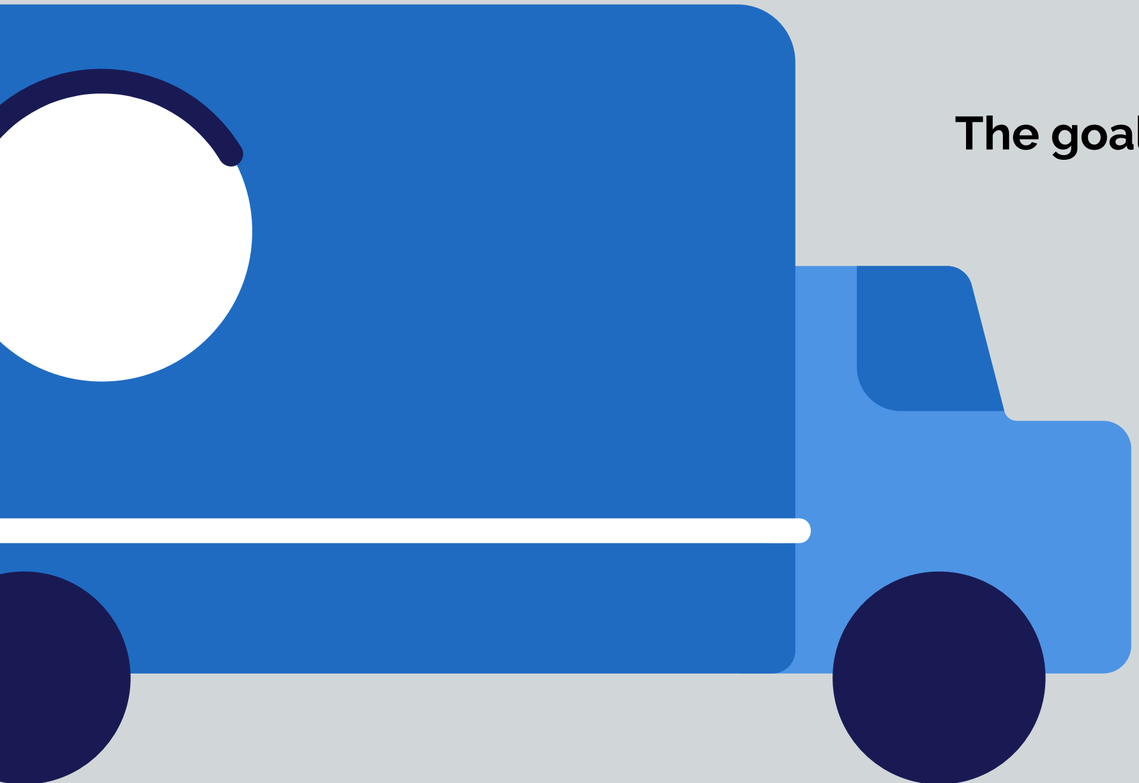
This SQL project explores return behavior and customer purchase patterns in a synthetic e-commerce dataset of 49,000+ transactions.

BY Sugandha Saini

ABOUT THE PROJECT

I performed data-driven analysis using intermediate to advanced SQL techniques including:

- Joins
- CTEs
- Aggregations
- Window Functions
- Conditional Logic (CASE WHEN)



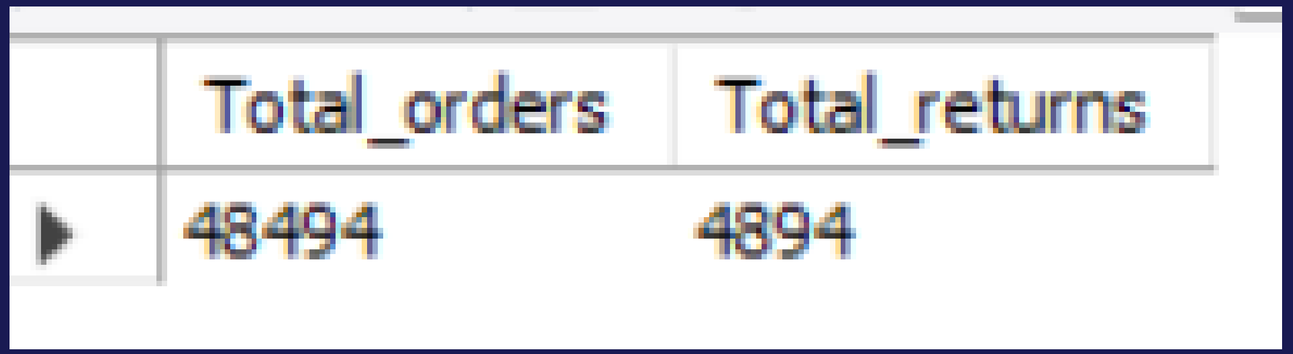
The goal was to identify high-return categories, potential fraudulent behavior, and gain insights that help in:

- Reducing return rates
- Improving inventory & logistics decisions
- Spotting patterns in high-discount misuse

Query 1

Total orders and Total returns

```sql
SELECT
    count(distinct InvoiceNo) as Total_orders,
    count( case when ReturnStatus ='Returned' then 1 end) as Total_returns
FROM online_sales_dataset;
```



Total_orders	Total_returns
48494	4894

Query 2

Most sold product categories

```sql
-- 2. Most sold product categories
SELECT Category , sum(Quantity) as Total_Quantity_Sold
FROM online_sales_dataset
GROUP BY Category
ORDER BY Total_Quantity_Sold  desc;
```

Category	Total_Quantity_Sold
Furniture	225603
Stationery	224590
Apparel	222044
Accessories	221667
Electronics	219836

Query 3

Top 5 countries with highest number of orders

```sql
SELECT
  Country,
  sum( distinct InvoiceNo) as Number_of_Orders
FROM online_sales_dataset
GROUP BY Country
ORDER BY Number_of_Orders desc
LIMIT 5;
```

Country	Number_of_Orders
France	2327921407
United Kingdom	2319868280
Belgium	2311070747
Portugal	2301712728
Germany	2299413726

Query 4

Return count by payment method

```sql
SELECT
    count(distinct InvoiceNo) as Total_orders,
    PaymentMethod
FROM online_sales_dataset
GROUP BY PaymentMethod
ORDER BY Total_orders desc;
```

Total_orders	PaymentMethod
16594	Bank Transfer
16394	Credit Card
16362	paypall

Query 5

Distribution of sales channels (Online vs In-store)

```sql
SELECT
    count(distinct InvoiceNo) as Total_orders,
    SalesChannel
FROM online_sales_dataset
GROUP BY SalesChannel
ORDER BY Total_orders desc;
```

Total_orders	SalesChannel
24717	Online
24399	In-store

Query 6

What is the monthly trend of total orders vs returns?

```sql
select
    date_format(InvoiceDate, '%Y-%m') as Months,
    count(distinct InvoiceNo) as Total_Orders,
    count(case when ReturnStatus ='Returned' then 1 end) as Total_Returns
from online_sales_dataset
group by Months
order by Months;
```

Months	Total_Orders	Total_Returns
2020-01	744	81
2020-02	696	69
2020-03	744	75
2020-04	720	76
2020-05	744	68
2020-06	720	73
2020-07	744	62
2020-08	744	74
2020-09	720	85

Query 7

Which product categories have the highest return rates?

```sql
select
     Category,
     count(case when ReturnStatus ='Returned' then 1 end) as Total_Returns,
      count(distinct InvoiceNo) as Total_Orders,
      round(count(case when ReturnStatus ='Returned' then 1 end) * 100/count(distinct InvoiceNo),2) as return_rate
from online_sales_dataset
group by Category
order by return_rate;
```

Category	Total_Returns	Total_Orders	return_rate
Accessories	965	9921	9.73
Apparel	963	9809	9.82
Electronics	977	9886	9.88
Stationery	985	9879	9.97
Furniture	1004	10028	10.01

Query 8

Compare average discount given on returned vs non-returned items.

```sql
select
      round(avg(Discount),2) as Avg_Discount,
      ReturnStatus
from online_sales_dataset
group by ReturnStatus
order by ReturnStatus;
```

Avg_Discount	ReturnStatus
0.28	Not Returned
0.28	Returned

Query 9

Which shipment providers have the highest return percentage?

```sql
Select
    ShipmentProvider,
    count(case when ReturnStatus ='Returned' then 1 end) as Total_Returns,
    count(distinct InvoiceNo) as Total_Orders,
    round(count(case when ReturnStatus ='Returned' then 1 end)*100/count(distinct InvoiceNo),2) as Return_Percentage
from online_sales_dataset
group by ShipmentProvider
order by Total_Returns;
```

ShipmentProvider	Total_Returns	Total_Orders	Return_Percentage
UPS	1178	12349	9.54
DHL	1237	12349	10.02
FedEx	1238	12426	9.96
Royal Mail	1241	12335	10.06

Query 10

Which product categories are most returned by VIP customers?

```sql
select
    OrderPriority,
    COUNT(*) AS Total_Orders,
    COUNT(case when ReturnStatus = 'Returned' then 1 end) as Returned_Orders,
    ROUND(
        COUNT(case when ReturnStatus = 'Returned' then 1 end) * 100.0 / COUNT(*),
        2
    ) AS Return_Rate_Percentage
from online_sales_dataset
group by OrderPriority
order by Return_Rate_Percentage desc;
```

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

OrderPriority	Total_Orders	Returned_Orders	Return_Rate_Percentage
High	16562	1648	9.95
Medium	16678	1638	9.82
Low	16542	1608	9.72

Query 11

Rank customers by return frequency using RANK()

```sql
with ReturnCount as(
select CustomerID,count(*) as ReturnFreq
from online_sales_dataset
where ReturnStatus ='Returned'
and CustomerID IS not null
group by CustomerID
)
select CustomerID, ReturnFreq,
rank()over(order by ReturnFreq desc ) as Return_rank
from ReturnCount
order by Return_rank
```

CustomerID	ReturnFreq	Return_rank
	507	1
29607.0	3	2
68423.0	3	2
53125.0	3	2
96879.0	3	2
37608.0	2	6

Query 12

Detect products with 3+ returns across different invoices

```sql
with return_count as(
select  InvoiceNo, Category,
count(*) as No_of_Returns
from online_sales_dataset
where ReturnStatus ='Returned'
group by Category,InvoiceNo
having count(*)>3
)
select InvoiceNo,Category,No_of_Returns
from return_count
order by No_of_Returns desc;
```

	InvoiceNo	Category	No_of_Returns

Result 10

Output

Action Output

#	Time	Action	Message
17	10:42:43	with return_count as(select InvoiceNo, Category, count(*) as No_of_Returns from online_sales_dataset where...	0 row(s) returned

Query 13

Use CASE WHEN to categorize return reasons by high discount, order priority

```
with discount as(
select OrderPriority,
       count(case when ReturnStatus ='Returned'then 1 end) as Total_Returns
from online_sales_dataset
where Discount >0.50
Group by OrderPriority
)
select OrderPriority,Total_Returns
from discount
order by Total_Returns
```

OrderPriority	Total_Returns
Low	28
Medium	33
High	37

Query 14

Compare return % between online and in-store orders using CTE

```sql
with Return_cal as(
select SalesChannel,
    count(case when ReturnStatus ='Returned' then 1 end) as Total_Returns,
    round(count(case when ReturnStatus ='Returned' then 1 end)*100/count(distinct InvoiceNo),2) as Return_Percentage
 from online_sales_dataset
 group by SalesChannel
 )
 select SalesChannel,Total_Returns,Return_Percentage
 from Return_cal
 order by Return_percentage desc;
```

SalesChannel	Total_Returns	Return_Percentage
In-store	2440	10.00
Online	2454	9.93

Query 15

Flag possible frauds: high discount + returned + same customer multiple times

```sql
with Posible_Frauds as(
 select   CustomerID ,
 count(case when ReturnStatus = 'Returned' then 1 end) as Returned,
 Count(*) as Total_orders,
 Round(avg(Discount),2) as Avg_Discount
 from online_sales_dataset
 where Discount > 0.5
 group by CustomerID
 )
 select CustomerID,Total_orders,Returned,Avg_Discount
 from Posible_Frauds
where Returned >= 2
rder by  Returned desc;
```

CustomerID	Total_orders	Returned	Avg_Discount
	996	98	1.51

Key Insights

- **Electronics and Apparel** categories stood out with the **highest return rates**, revealing potential product quality or expectation mismatches.

- Interestingly, **online orders** had significantly **more returns compared to in-store purchases**, suggesting that customers may face challenges in product selection or satisfaction when buying digitally.

- Returns were also **more frequent** among orders with **discounts over 50%**, indicating possible patterns of return abuse or misuse of promotional campaigns.

- A **few shipment providers** had unusually **high return percentages**, flagging a need for deeper investigation into logistics and handling processes.

- I also flagged customers who consistently returned high-discounted items, helping simulate **a real-world fraud detection**