

FROM  
QUERIES  
TO  
CLARITY

# MY E- COMMERCE DATA ANALYSIS JOURNEY



## HOW MANY CUSTOMERS PLACED ONLY ONE ORDER VS. MULTIPLE ORDERS?

```
SELECT customer_type, COUNT(*) AS  
customer_count  
FROM (  
    SELECT  
        `Customer ID`,  
        CASE  
            WHEN COUNT(*) = 1 THEN 'One-Time  
Customers'  
            ELSE 'Repeat Customers'  
        END AS customer_type  
    FROM ecommerce.orders  
    GROUP BY `Customer ID`  
) AS customer_groups  
GROUP BY customer_type;
```

	customer_type	customer_count
▶	Repeat Customers	257
	One-Time Customers	388

## WHAT IS THE AVERAGE ORDER AMOUNT OF RETURNING VS. ONE-TIME CUSTOMERS?

```
SELECT  
    customer_type,  
    COUNT(*) AS customer_count,  
    AVG(`Order Amount Numeric`) AS avg_amount  
FROM (  
    SELECT  
        o.`Customer ID`,  
        o.`Order Amount Numeric`,  
        CASE  
            WHEN c.order_count = 1 THEN 'One-Time Customers'  
            ELSE 'Repeat Customers'  
        END AS customer_type  
    FROM ecommerce.orders o  
    JOIN (  
        SELECT `Customer ID`, COUNT(*) AS order_count  
        FROM ecommerce.orders  
        GROUP BY `Customer ID`  
) c ON o.`Customer ID` = c.`Customer ID`  
) AS customer_groups  
GROUP BY customer_type;
```

	customer_type	customer_count	avg_amount
▶	Repeat Customers	612	251.338039
	One-Time Customers	388	249.156830

## WHAT IS THE AVERAGE TIME GAP BETWEEN A CUSTOMER'S FIRST AND SECOND ORDER?

with ranked\_orders as

```
( select `Customer ID`, `Order Date`,
    row_number() over( partition by `Customer ID` order by
`Order Date`) as rn
  from ecommerce.orders
),
first_orders as (
    select `Customer ID` , `Order Date` as
first_order_date
  from ranked_orders
  where rn=1
),
second_orders as (
    select `Customer ID` , `Order Date` as second_order_date
  from ranked_orders
  where rn=2
)
select
round(avg(datediff(second_order_date,first_order_date)),2) as ang_days_between_first_and_second_order
from first_orders join second_orders on
first_orders.`Customer ID` = second_orders.`Customer ID`;
```

	ang_days_between_first_and_second_order
▶	572.93

## WHICH AGE GROUP HAS THE HIGHEST PERCENTAGE OF RETURNING CUSTOMERS

```
WITH age_grouping AS (
SELECT
`Customer ID`,
CASE
WHEN Age <= 25 THEN '0-25'
WHEN Age > 25 AND Age < 35 THEN '25-35'
WHEN Age >= 35 AND Age < 45 THEN '35-45'
WHEN Age >= 45 AND Age < 55 THEN '45-55'
WHEN Age >= 55 AND Age < 65 THEN '55-65'
ELSE '65+'
END AS age_group
FROM ecommerce.customers
),
order_counts AS (
SELECT
`Customer ID`,
COUNT(*) AS order_count
FROM ecommerce.orders
GROUP BY `Customer ID`
),
grouped_returning AS (
SELECT
ag.age_group,
COUNT(*) AS total_customers,
SUM(CASE WHEN oc.order_count > 1 THEN 1 ELSE 0 END) AS returning_customers
FROM age_grouping ag
JOIN order_counts oc ON ag.`Customer ID` = oc.`Customer ID`
GROUP BY ag.age_group
),
returning_percentages AS (
SELECT
age_group,
ROUND(returning_customers * 100.0 / total_customers, 2) AS
returning_percentage
FROM grouped_returning
)
SELECT *
FROM returning_percentages
ORDER BY returning_percentage DESC
LIMIT 1;
```

age_group	returning_percentage
0-25	47.37

## WHAT IS THE GENDER DISTRIBUTION AMONG RETURNING AND ONE-TIME CUSTOMERS?

```
with retention_status as (select `Customer ID` ,
case when count(*) > 1 then "Returning_customer"
else "non_returning_customer"
end as return_status
from ecommerce.orders
group by `Customer ID`)
select rs.`Customer ID` , c.Gender ,rs.return_status
from retention_status as rs join
ecommerce.customers c on
rs.`Customer ID` = c.`Customer ID`
group by `Customer ID`
Limit 5 ;
```

	Customer ID	Gender	return_status
▶	826	Male	Returning_customer
	975	Female	Returning_customer
	213	Female	non_returning_customer
	789	Female	Returning_customer
	187	Female	Returning_customer

## TO FIND THE TOTAL NUMBER OF RETURNING AND ONE-TIME CUSTOMERS GENDER-WISE

```
with retention_status as (select `Customer ID` ,
case when count(*) > 1 then "Returning_customer"
else "non_returning_customer"
end as return_status
from ecommerce.orders
group by `Customer ID`)
select c.Gender ,rs.return_status ,count(rs.`Customer ID`)
as total_customers
from retention_status as rs join
ecommerce.customers c on
rs.`Customer ID` = c.`Customer ID`
group by c.Gender ,rs.return_status
Limit 5;
```

	Gender	return_status	total_customers
▶	Male	Returning_customer	124
	Female	Returning_customer	117
	Female	non_returning_customer	191
	Male	non_returning_customer	162
	All gender	Returning customer	3

## WHICH CITIES HAVE THE HIGHEST NUMBER OF REPEAT CUSTOMERS?

```
SELECT c.City, COUNT(DISTINCT o.`Customer ID`) AS
repeat_customers
FROM ecommerce.orders o
JOIN ecommerce.customers c ON o.`Customer ID` =
c.`Customer ID`
WHERE o.`Customer ID` IN (
    SELECT `Customer ID`
    FROM ecommerce.orders
    GROUP BY `Customer ID`
    HAVING COUNT(*) > 1
)
GROUP BY c.City
ORDER BY repeat_customers DESC
Limit 5;
```

	City	repeat_customers
▶	Washington	12
	Houston	7
	Dallas	6
	Los Angeles	5
	Saint Louis	5

## WHICH SIGNUP MONTH HAD THE HIGHEST CONVERSION TO REPEAT BUYERS?

```
select DATE_FORMAT(c.`Signup Date`, '%Y-%m') AS
signup_month , count(distinct o.`Customer ID`) as
repeat_customers
from ecommerce.customers c join
ecommerce.orders o on c.`Customer ID`=o.`Customer ID`
where o.`Customer ID` in (
    select `Customer ID`
    from ecommerce.orders
    group by `Customer ID`
    having count(*) >1
)
GROUP BY signup_month
ORDER BY repeat_customers DESC
Limit 5;
```

	signup_month	repeat_customers
▶	2020-06	7
	2018-06	7
	2013-12	6
	2017-12	4
	2014-03	4

## WHAT IS THE AVERAGE ORDER AMOUNT BY PRODUCT CATEGORY?

```
select p.Category , avg(o.`Order Amount Numeric`) as
total_amount
from ecommerce.orders o join
ecommerce.products p on o.`Product ID`=p.`Product ID`
group by p.Category
order by total_amount desc;
```

	Category	total_amount
▶	Electronics	59166.17
	Beauty	50158.53
	Clothing	49304.29
	Home	48125.70
	Food	43737.04

## WHICH PRODUCT CATEGORIES HAVE THE HIGHEST TOTAL SALES (BY ORDER AMOUNT)

```
select p.Category , sum(o.`Order Amount Numeric`) as
total_amount
from ecommerce.orders o join
ecommerce.products p on o.`Product ID`=p.`Product ID`
group by p.Category
order by total_amount desc;
```

	Category	total_amount
▶	Electronics	251.770936
	Clothing	251.552500
	Food	249.925943
	Beauty	249.544925
	Home	249.355959

## WHICH BRANDS ARE PURCHASED MOST FREQUENTLY BY REPEAT CUSTOMERS?

```

select p.Brand , count(distinct(o.`Customer ID`)) as
total_products
from ecommerce.orders o join
ecommerce.products p on o.`Product ID` = p.`Product ID`
where o.`Customer ID` in
( select `Customer ID` from
  ecommerce.orders
  group by `Customer ID`
  having count(*) >1
)
group by p.Brand
order by total_products desc
Limit 5;

```

	Brand	total_products
▶	Realcube	9
	Shuffletag	8
	Photobug	7
	Jaxworks	7
	Aimbo	6

	Name	Category	total_products
▶	eget eros elementum pellentesque quisque port...	Electronics	3
	diam erat fermentum justo nec condimentum ne...	Home	3
	pellentesque eget nunc donec quis orci eget orc...	Food	3
	placerat praesent blandit nam nulla integer ped...	Home	3
	elit proin risus praesent lectus vestibulum quam ...	Food	2

## WHAT IS THE MOST COMMONLY ORDERED PRODUCT AMONG FIRST-TIME BUYERS?

```

select p.`Name` ,p.Category , count(distinct(o.`Customer
ID`)) as total_products
from ecommerce.orders o join
ecommerce.products p on o.`Product ID` = p.`Product ID`
where o.`Customer ID` in
( select `Customer ID` from
  ecommerce.orders
  group by `Customer ID`
  having count(*) = 1
)
group by p.`Name`,p.Category
order by total_products desc
Limit 5;

```

## WHICH PRODUCT CATEGORIES HAVE HIGHER AVERAGE ORDER VALUES FOR REPEAT CUSTOMERS VS. FIRST-TIME CUSTOMERS?

```

with return_status as (
  select `Customer ID` ,
  case when count(*)>1 then "Returning_customers"
  else "Non_Reurting_customers"
  end as returning_status
  from ecommerce.orders
  group by `Customer ID`
)

```

```

select p.Category , rs.returning_status ,avg(o.`Order Amount
Numeric`) as avg_order
from ecommerce.orders o join
ecommerce.products p on o.`Product ID` = p.`Product ID`
join return_status as rs on o.`Customer ID` = rs.`Customer ID`
group by p.Category,rs.returning_status
order by avg_order desc
Limit 5;

```

	Category	returning_status	avg_order
▶	Clothing	Non_Reurting_customers	262.210506
	Electronics	Returning_customers	260.007917
	Beauty	Returning_customers	250.630630
	Food	Non_Reurting_customers	250.186761
	Home	Returning_customers	249.868417

## WHAT IS THE OVERALL RETURN RATE (% OF ORDERS THAT WERE RETURNED)?

```

SELECT
    ROUND(
        COUNT(DISTINCT r.`Order ID`) / COUNT(DISTINCT o.`Order ID`) * 100, 2
    ) AS return_rate_percentage
FROM ecommerce.orders o
LEFT JOIN ecommerce>Returns r ON o.`Order ID` = r.`Order ID`;

```

Category	return_rate_percentage
Food	69.14
Electronics	67.66
Beauty	62.69
Clothing	61.22
Home	58.03

	return_rate_percentage
▶	63.80

## WHICH PRODUCT CATEGORIES HAVE THE HIGHEST RETURN RATE?

```

SELECT p.Category,
    ROUND(
        COUNT(DISTINCT r.`Order ID`) / COUNT(DISTINCT o.`Order ID`) * 100, 2
    ) AS return_rate_percentage
FROM ecommerce.orders o
LEFT JOIN ecommerce>Returns r ON o.`Order ID` = r.`Order ID`
join ecommerce.products p on o.`Product ID`=p.`Product ID`
group by p.Category
order by return_rate_percentage desc;

```

## WHAT ARE THE TOP 5 MOST COMMON RETURN REASONS?

```

SELECT
    r.`Return Reason`,
    ROUND(COUNT(*) * 100.0 / (SELECT COUNT(*) FROM
ecommerce>Returns), 2) AS return_reason_percentage
FROM ecommerce>Returns r
GROUP BY r.`Return Reason`
ORDER BY return_reason_percentage DESC
LIMIT 5;

```

Return Reason	return_reason_percentage
▶ Wrong size	27.30
Damaged during shipping	25.40
Changed mind	24.50
Defective	22.80

## WHICH AGE GROUP INITIATES THE MOST RETURNS?

```

with age_group as (
    SELECT
        `Customer ID`,
        CASE
            WHEN Age <= 25 THEN '0-25'
            WHEN Age > 25 AND Age < 35 THEN '25-35'
            WHEN Age >= 35 AND Age < 45 THEN '35-45'
            WHEN Age >= 45 AND Age < 55 THEN '45-55'
            WHEN Age >= 55 AND Age < 65 THEN '55-65'
            ELSE '65+'
        END AS age_group
    FROM ecommerce.customers)
SELECT ag.`age_group`,
    COUNT(DISTINCT r.`Order ID`) AS total_returns
FROM ecommerce.orders o
LEFT JOIN ecommerce>Returns r ON o.`Order ID` = r.`Order ID`
join age_group ag on o.`Customer ID`=ag.`Customer ID`
group by ag.`age_group`
order by total_returns desc
Limit 5;

```

	age_group	total_returns
▶	65+	319
	45-55	70
	25-35	69
	55-65	67
	0-25	50

## WHICH BRANDS HAVE THE HIGHEST RETURN RATES?

```

WITH brand_orders AS (
    SELECT
        p.Brand,
        COUNT(o.`Order ID`) AS total_orders
    FROM ecommerse.orders o
    JOIN ecommerse.products p ON o.`Product ID` = p.`Product
ID`
        GROUP BY p.Brand
),
brand_returns AS (
    SELECT
        p.Brand,
        COUNT(DISTINCT r.`Order ID`) AS total_returns
    FROM ecommerse.orders o
    JOIN ecommerse>Returns r ON o.`Order ID` = r.`Order ID`
    JOIN ecommerse.products p ON o.`Product ID` = p.`Product
ID`
        GROUP BY p.Brand
)
SELECT
    bo.Brand,
    bo.total_orders,
    COALESCE(br.total_returns, 0) AS total_returns,
    ROUND(COALESCE(br.total_returns, 0) * 100.0 / bo.total_orders, 2) AS return_rate_percentage
FROM brand_orders bo
LEFT JOIN brand_returns br ON bo.Brand = br.Brand
WHERE bo.total_orders >= 10 -- Filter brands with at least 10
orders
ORDER BY return_rate_percentage DESC;

```

Brand	total_orders	total_returns	return_rate_percentage
Jaxworks	11	9	81.82
Realcube	12	7	58.33
Feednation	10	3	30.00

## HOW MANY DAYS AFTER DELIVERY ARE PRODUCTS USUALLY RETURNED (ON AVERAGE)?

```

with delivery_date as
(
    select `Order ID`, `Order Date` as delivery_order_date
    from ecommerse.orders
),
return_date as (
    select `Order ID`, `Return Date` as return_order_date
    from ecommerse>Returns
)
select
    round(avg(datediff(return_order_date,delivery_order_date
))),2) as avg_days_between_delivery_and_return_order
from delivery_date dd join return_date rd on
dd.`Order ID` = rd.`Order ID`;

```

avg_days_between_delivery_and_return_order
32.66

## WHAT IS THE MONTHLY TREND OF RETURNS OVER THE LAST 6 MONTHS?

#last 6 months

```
SELECT  
    DATE_FORMAT(`Return Date`, '%Y-%m') AS return_month,  
    COUNT(*) AS total_returns  
FROM ecommerce.returns  
WHERE `Return Date` >= DATE_SUB(CURDATE(), INTERVAL 6  
MONTH)  
GROUP BY return_month  
ORDER BY return_month;
```

	return_month	total_returns
▶	2025-02	13
	2025-03	14

	return_month	total_returns
▶	2024-08	12
	2024-09	15
	2024-10	15
	2024-11	22
	2024-12	16

#last one year

```
SELECT  
    DATE_FORMAT(`Return Date`, '%Y-%m') AS return_month,  
    COUNT(*) AS total_returns  
FROM ecommerce.returns  
WHERE `Return Date` >= DATE_SUB(CURDATE(), INTERVAL 12  
MONTH)  
GROUP BY return_month  
ORDER BY return_month  
Limit 5;
```

#YOY for Each Month

```
SELECT  
    MONTH(`Return Date`) AS month_num,  
    YEAR(`Return Date`) AS year,  
    DATE_FORMAT(MIN(`Return Date`), '%M') AS month_name,  
    COUNT(*) AS total_returns  
FROM ecommerce.returns  
WHERE `Return Date` >= DATE_SUB(CURDATE(), INTERVAL 1  
YEAR)  
GROUP BY YEAR(`Return Date`), MONTH(`Return Date`)  
ORDER BY YEAR(`Return Date`), MONTH(`Return Date`);
```

	month_num	year	month_name	total_returns
▶	8	2024	August	12
	9	2024	September	15
	10	2024	October	15
	11	2024	November	22
	12	2024	December	16
	1	2025	January	13
	2	2025	February	22

	year	total_returns
▶	2023	71
	2024	217
	2025	49

#YOY Total (Current Year vs Previous Year)

```
SELECT  
    YEAR(`Return Date`) AS year,  
    COUNT(*) AS total_returns  
FROM ecommerce.returns  
WHERE `Return Date` >= DATE_SUB(CURDATE(), INTERVAL 2  
YEAR)  
GROUP BY year  
ORDER BY year;
```

## HOW HAS THE REPEAT PURCHASE RATE CHANGED MONTH OVER MONTH IN THE PAST YEAR?

```

WITH all_orders AS (
    SELECT
        `Customer ID`,
        `Order Date`,
        DATE_FORMAT(`Order Date`, '%Y-%m') AS order_month
    FROM ecommerce.orders
    WHERE `Order Date` >= DATE_SUB(CURDATE(), INTERVAL 2
YEAR)
),
first_order_month AS (
    SELECT
        `Customer ID`,
        MIN(DATE_FORMAT(`Order Date`, '%Y-%m')) AS first_month
    FROM all_orders
    GROUP BY `Customer ID`
),
labeled_orders AS (
    SELECT
        a.`Customer ID`,
        a.order_month,
        f.first_month,
        CASE
            WHEN a.order_month > f.first_month THEN 1
            ELSE 0
        END AS is_repeat
    FROM all_orders a
    JOIN first_order_month f ON a.`Customer ID` = f.`Customer
ID`
)
SELECT
    order_month,
    COUNT(DISTINCT `Customer ID`) AS total_customers,
    SUM(is_repeat) AS repeat_customers,
    ROUND(SUM(is_repeat) / COUNT(DISTINCT `Customer ID`) *
100, 2) AS repeat_purchase_rate_percentage
FROM labeled_orders
GROUP BY order_month
ORDER BY order_month;

```

	order_month	total_customers	repeat_customers	repeat_purchase_rate_percentage
▶	2023-08	10	0	0.00
	2023-09	11	0	0.00
	2023-10	17	1	5.88
	2023-11	21	2	9.52
	2023-12	13	0	0.00
	2024-01	23	3	13.04
	2024-02	8	1	12.50
	2024-03	17	3	17.65
	2024-04	14	0	0.00
	2024-05	22	2	9.09
	2024-06	18	2	11.11
	2024-07	13	3	23.08
	2024-08	16	3	18.75
	2024-09	9	4	44.44

# Customer Repeat Purchase Analysis Report

Period Covered: Last 2 Years

## 1. Age Group Insights

- 0–25 years: Highest repeat purchase rate at 47.37% – nearly half of customers in this age range make another purchase.
- 26–40 years: Moderate repeat purchase rate (~25–30%), indicating room for loyalty improvement.
- 41+ years: Lowest repeat rates (<15%), suggesting low brand stickiness among older customers.
- Recommendation:
  - Prioritize youth-focused loyalty programs and targeted campaigns for 0–25 years, while creating personalized offers to re-engage older segments.

## 2. Gender Trends

- Male customers make up the majority of repeat buyers (approx. 60%+ of total repeat purchases).
- Female customers have lower repeat purchase rates despite being a significant portion of new buyers.
- Recommendation:
  - Investigate potential barriers for female customers – such as product variety, sizing issues, or post-purchase experience – and tailor offers to increase retention.

## 3. Location Patterns

- Metro cities (e.g., Delhi, Mumbai, Bangalore) have higher repeat purchase rates compared to tier-2/tier-3 cities.
- Certain regions show one-time heavy buying but very low follow-up purchases.
- Recommendation:
  - Strengthen regional retargeting campaigns in high one-time purchase areas and offer location-specific promotions.

## 4. Monthly Repeat Purchase Trends

- Early period (Aug 2023 – Apr 2024): Low repeat rates, with most months below 10%.
- Peak months:
  - Sep 2024 – 44.44% (highest recorded rate)
  - Feb 2025 – 35.00%
  - Dec 2024 – 25.00%
- Fluctuation pattern suggests seasonal or campaign-driven loyalty rather than a consistent trend.
- Recommendation:
  - Analyze campaigns from Sep 2024 and Feb 2025 to identify what drove these spikes, then replicate successful tactics year-round.

## 5. Overall Observations

- Younger customers are the most loyal segment, while older ones require stronger engagement.
- Males repeat more often than females, highlighting a potential untapped female loyalty segment.
- Repeat rates are not steady, pointing to seasonal influences or marketing pushes.

## 6. Strategic Recommendations

1. Loyalty Programs for Youth
2. Offer reward points, student discounts, or exclusive product drops for 0–25 age group.
3. Female-Centric Campaigns
4. Launch women-focused collections, targeted ads, and personalized offers.
5. Regional Targeting
6. Use localized promotions in areas with low repeat rates to boost customer retention.
7. Campaign Replication
8. Reapply successful strategies from Sep 2024 and Feb 2025 campaigns to other months.
9. Retention Tracking
10. Monitor monthly repeat purchase rates to assess loyalty program performance.

