

# **Unlocking Insights: E-Commerce Retail Data Analysis**

2024

By : Vishnu G Nath



vishnugs32@gmail.com



vishnugnath



vishnugnath

---

# AGENDA

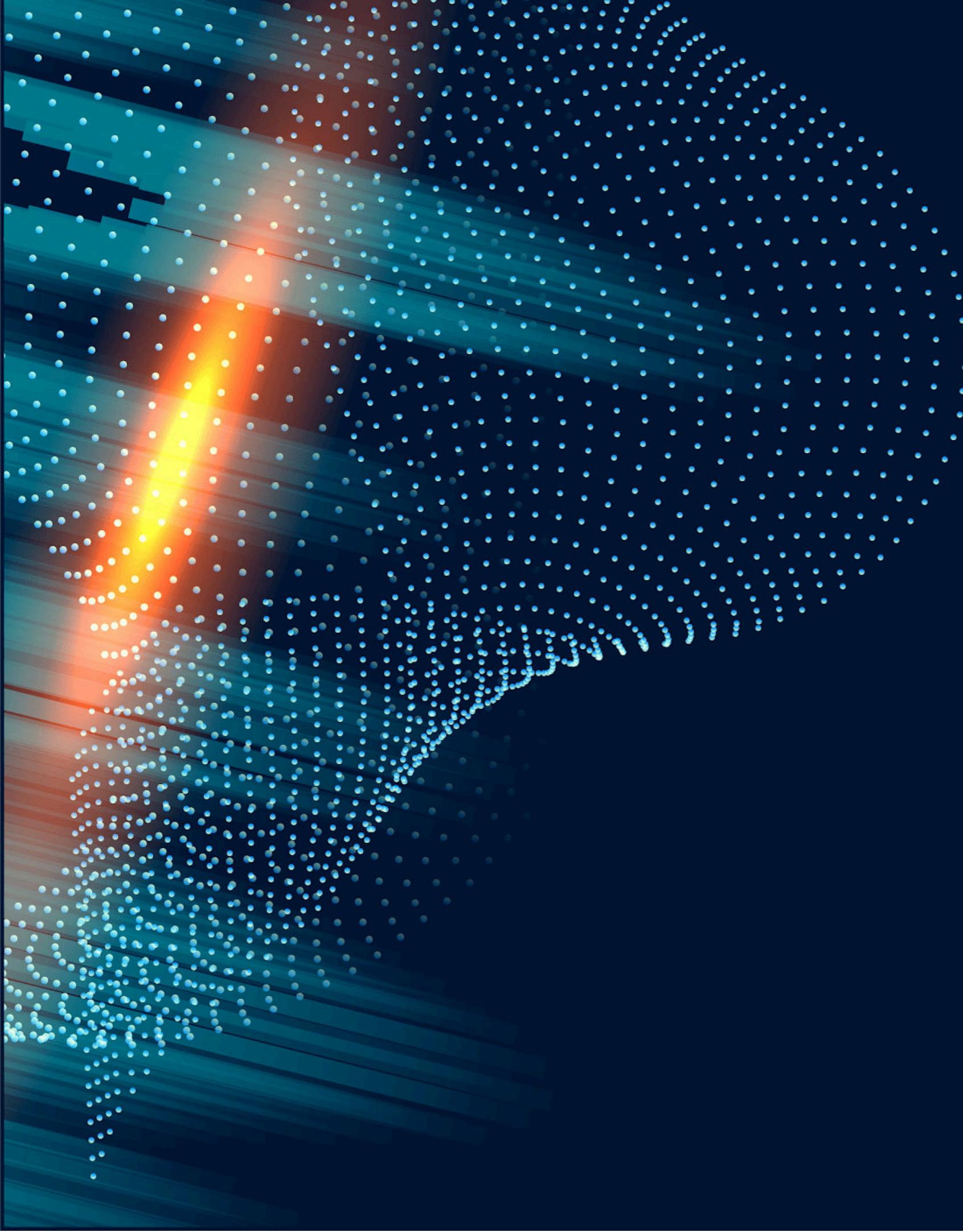
- INTRODUCTION
- PROJECT OBJECTIVE
- DATABASE TABLES
- DATABASE SCHEMA
- PRIMARY KEY, FOREIGN KEY
- OBJECTIVE QUERY
- PERFORMANCE INSIGHTS
- KERY RECOMMENDATION



# INTRODUCTION

This analysis aims to derive meaningful insights from e-commerce data to inform strategic decision-making. By examining customer behaviour, transaction trends, and product performance, we will provide actionable recommendations to enhance engagement, optimise sales, and improve overall business efficiency.





# Project Objective

---

Leveraging e-commerce data, we can derive actionable insights to optimize business performance and elevate customer experience. By analyzing customer retention, loyalty patterns, and emerging trends, we can devise personalized marketing strategies and support data-driven decision-making to propel future growth.

# DATABASE SCHEMA OVERVIEW

## Customer TableDescription:

This table contains detailed information about customers in the E-Commerce Retail Dataset. Variables: customer\_idDOB (Date of Birth)Gendercity\_codeSchema Details: Total Variables: 4Total Records: 5,647

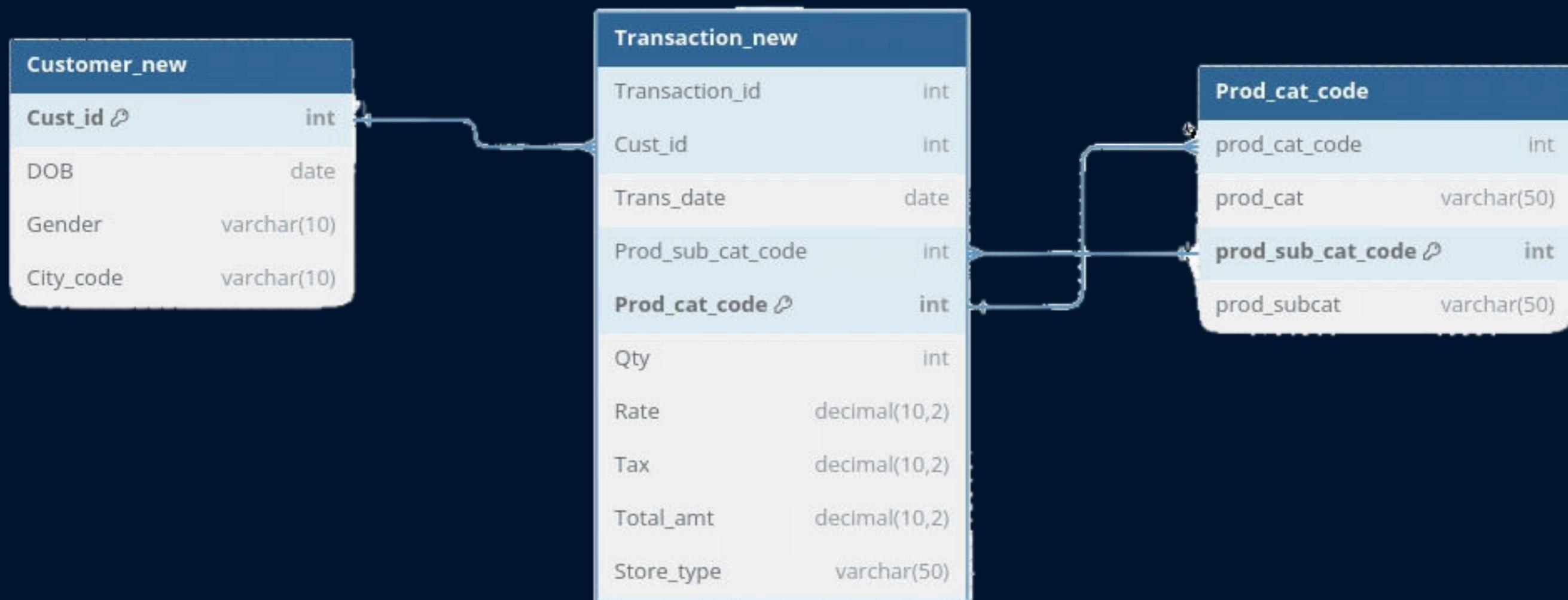
## Transaction TableDescription:

This table records all transactions that occur in the E-Commerce Retail Dataset. Variables: customer\_idtransaction\_idprod\_subcat\_codeprod\_cat\_codequantityrateaxtotal\_amountstore\_typeSchema Details: Total Variables: 10Total Records: 23,053

## Product TableDescription:

This table holds information about the products available in the E-Commerce Retail Dataset. Variables: prod\_cat\_codeprod\_catprod\_sub\_cat\_codeprod\_subcatSchema Details: Total Variables: 4Total Records: 23

# DATABASE SCHEMA OVERVIEW



# **PRIMARY KEY**

---

A PRIMARY KEY IS A UNIQUE IDENTIFIER FOR EACH RECORD IN A DATABASE TABLE.

EXAMPLE :

CUSTOMER\_NEW - CUST\_ID PROD\_CAT\_CODE -  
PROD\_SUB\_CAT\_CODE  
TRANSACTION\_NEW-  
PROD\_CAT\_CODE  
FOREIGN

# **FORIEGN KEY**

A FOREIGN KEY IS A FIELD (OR COLLECTION OF FIELDS) IN ONE TABLE THAT REFERS TO THE PRIMARY KEY IN ANOTHER TABLE.

EXAMPLE :

TRANSACTION\_NEW :  
CUST\_ID (REFERENCE TO CUSTOMER\_NEW)  
PROD\_SUB\_CAT\_CODE(REFERENCE TO PROD\_CAT\_CODE)  
PROD\_CAT\_CODE : PROD\_CAT\_CODE(REFERENCE  
TO TRANSACTION\_NEW)

# DATA PREPARATION AND UNDERSTANDING

1. What is the total number of rows in each of the 3 tables in the database?

```
select count(*) from customers_new;  
select count(*) from prod_cat_info;  
select count(*) from transactions_new;
```

customers\_new

Result Grid	
	Filter Rows:
count(*)	
5645	

prod\_cat\_info

Result Grid	
	Filter Rows:
count(*)	
23	

transcation\_new

Result Grid	
	Filter Rows:
count(*)	
23053	

2. What is the total number of transactions that have a return?

```
select count(transaction_id) from transactions_new  
where Qty>0;
```

Result Grid	
	count(transaction_id)
▶	20876

3. What is the time range of the transaction data available for analysis? Show the output in number of days, months and years simultaneously in different columns.

```
SELECT
    MIN(`tran_date`) AS min_order_date,
    MAX(`tran_date`) AS max_delivery_date,
    TIMESTAMPDIFF(DAY, MIN(`tran_date`), MAX(`tran_date`)) AS total_days,
    TIMESTAMPDIFF(MONTH, MIN(`tran_date`), MAX(`tran_date`)) AS total_months,
    TIMESTAMPDIFF(YEAR, MIN(`tran_date`), MAX(`tran_date`)) AS total_years
FROM
    `mini_project`.`transactions_new`;
```

	min_order_date	max_delivery_date	total_days	total_months	total_years
▶	2011-01-02	2014-12-02	1430	47	3

4. Which product category does the sub-category "DIY" belong to?

```
select prod_cat from prod_cat_info  
where prod_subcat="DIY";
```

Result Grid		Filter Rows:	Export:	Wrap Cell Content:
	prod_cat			
▶	Books			

# DATA ANALYSIS

1. Which channel is most frequently used for transactions?

```
select A.Store_type from
(select count(distinct(transaction_id)) as TOTAL_COUNT,Store_type from transactions_new
group by Store_type
order by TOTAL_COUNT desc
limit 1 ) A
```

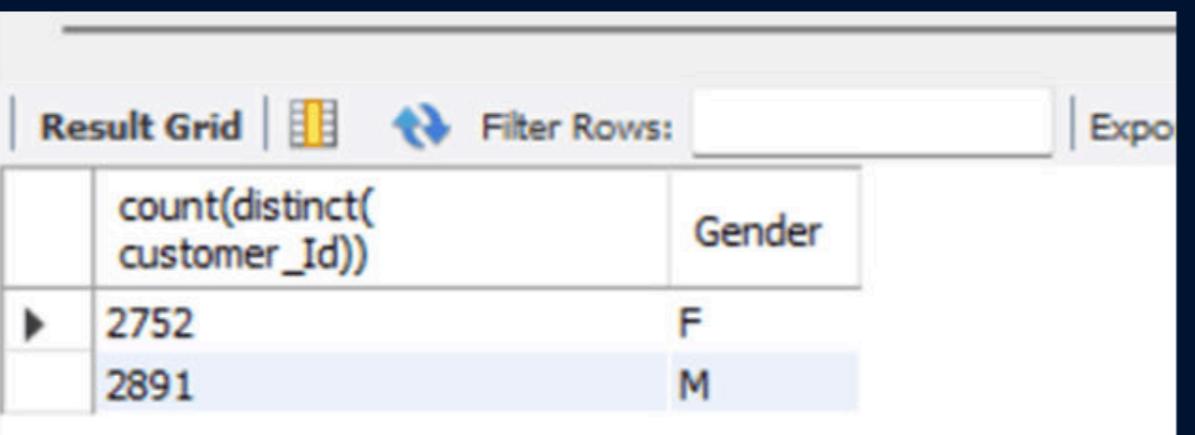


A screenshot of a database query result grid. The grid has a header row with a single column labeled "Store\_type". Below the header, there is one data row containing the value "e-Shop". The grid includes standard navigation icons for rows and columns, and a "Filter Rows" button.

Store_type
e-Shop

2. What is the count of Male and Female customers in the database?

```
select count(distinct( customer_Id)),Gender from customers_new  
where Gender in ('M','F')  
group by Gender;
```



The screenshot shows the MySQL Workbench interface with a query editor and a result grid. The query editor contains the SQL code provided above. The result grid displays two rows of data: one for females (F) with a count of 2752, and one for males (M) with a count of 2891.

	count(distinct( customer_Id))	Gender
▶	2752	F
	2891	M

3. From which city do we have the maximum number of customers and how many?

```
select count(distinct(customer_Id))as ID,city_code from customers_new  
group by city_code  
order by ID desc  
limit 1;
```

Result Grid		
	ID	city_code
▶	595	3

4.How many sub-categories are there under the Books category?

```
select count(prod_sub_cat_code) from prod_cat_info  
where prod_cat='Books'
```

Result Grid	
	count(prod_sub_cat_code)
▶	6

5. What is the maximum quantity of products ever ordered?

```
select A.qty from
  (SELECT prod_cat_code,COUNT(DISTINCT transaction_id) AS I,SUM(Qty) as qty
   FROM transactions_new
   GROUP BY prod_cat_code
   order by qty desc
   limit 1) A
```

Result Grid	
	qty
▶	14669

6. What is the net total revenue generated in categories Electronics and Books?

```
select sum(T.total_amt)AS Total_Revenue,P.prod_cat
from transactions_new T
Left join prod_cat_info P on
T.prod_subcat_code = P.prod_sub_cat_code and
T.prod_cat_code = P.prod_cat_code
where P.prod_cat in ('Electronics','Books')
GROUP BY P.prod_cat;
```

Result Grid | Filter Rows: Export:

	Total_Revenue	prod_cat
▶	10722463.634999989	Electronics
	12822694.040000048	Books

7. How many customers have >10 transactions with us, excluding returns?

```
select count(transaction_id),cust_id from transactions_new  
GROUP BY cust_id  
HAVING count(transaction_id)>10;
```

	count(transaction_id)	cust_id
▶	11	268934
	11	273896
	12	266794
	12	270831
	12	272286
	13	268819
	11	275131
	11	269216
	11	270908
	11	270535
	11	268408
	13	269449
	11	274948
	12	275252
	11	267014
	12	273014

8. What is the combined revenue earned from the "Electronics" & "Clothing" categories, from "Flagship stores"?

```
|  
| select *  from transactions_new T  
| Left join prod_cat_info P on  
| T.prod_subcat_code = P.prod_sub_cat_code and  
| T.prod_cat_code = P.prod_cat_code  
| HAVING P.prod_cat in ('Electronics','Clothing') and T.Store_type = 'Flagship stores';
```

The screenshot shows a database query results grid. The top row contains the column names: transaction\_id, cust\_id, tran\_date, prod\_subcat\_code, prod\_cat\_code, Qty, Rate, Tax, total\_amt, Store\_type, prod\_cat\_code, prod\_cat, prod\_sub\_cat\_code, prod\_subcat, and prod\_cat. The grid below is currently empty, indicating no data has been returned by the query.

transaction_id	cust_id	tran_date	prod_subcat_code	prod_cat_code	Qty	Rate	Tax	total_amt	Store_type	prod_cat_code	prod_cat	prod_sub_cat_code	prod_subcat	prod_cat

9. What is the total revenue generated from "Male" customers in "Electronics" category?

Output

should display total revenue by prod sub-cat.

```
select sum(T.total_amt)AS Total_Revenue,T.prod_subcat_code  from customers_new C
Left join transactions_new T on
C.customer_Id = T.cust_id
Left join prod_cat_info P on
T.prod_subcat_code = P.prod_sub_cat_code and
T.prod_cat_code = P.prod_cat_code
where C.Gender ='M' and P.prod_cat ='Electronics'
group by T.prod_subcat_code;
```

	Total_Revenue	prod_subcat_code
▶	1138983.1700000006	10
	1167845.7699999977	9
	1107593.4349999987	8
	1192413.235	4
	1090794.12	5

10. What is percentage of sales and returns by product sub category; display only top 5 sub categories in terms of sales?

```
WITH SubcategorySales AS (
    SELECT
        t.prod_subcat_code,
        p.prod_subcat,
        SUM(t.total_amt) AS total_sales,
        SUM(CASE WHEN t.Qty < 0 THEN t.total_amt ELSE 0 END) AS total_returns
    FROM
        transactions_new t
    JOIN
        prod_cat_info p ON t.prod_subcat_code = p.prod_sub_cat_code
    GROUP BY
        t.prod_subcat_code, p.prod_subcat
),
TopSubcategories AS (
    SELECT
        prod_subcat_code,
        prod_subcat,
        total_sales,
        total_returns,
        RANK() OVER (ORDER BY total_sales DESC) AS sales_rank
    FROM
        SubcategorySales
)
SELECT
    prod_subcat,
    total_sales,
    (total_sales / (SELECT SUM(total_sales) FROM SubcategorySales) * 100) AS sales_percentage,
    total_returns,
    (total_returns / (SELECT SUM(total_returns) FROM SubcategorySales) * 100) AS returns_percentage
FROM
    TopSubcategories
WHERE
    sales_rank <= 5;
```

prod_subcat	total_sales	sales_percentage	total_returns	returns_percentage
Mens	12392937.44000004	10.313029877909456	-1545406.5900000026	10.437687796889916
Women	8451811.290000023	7.03334320682429	-1111458.725	7.506800634698539
Mobiles	8451811.290000023	7.03334320682429	-1111458.725	7.506800634698539
Kids	8451811.290000023	7.03334320682429	-1111458.725	7.506800634698539
Mens	8451811.290000023	7.03334320682429	-1111458.725	7.506800634698539

11. For all customers aged between 25 to 35 years find what is the net total revenue generated by these consumers in last 30 days of transactions from max transaction date available in the data?

```
-- Step 1: Determine the maximum transaction date
WITH MaxTranDate AS (
SELECT MAX(tran_date) AS max_date
FROM transactions_new
),

-- Step 2: Filter transactions within the last 30 days from the maximum transaction date
RecentTransactions AS (
SELECT t.*, m.max_date
FROM transactions_new t
CROSS JOIN MaxTranDate m
WHERE DATE(t.tran_date) BETWEEN DATE_SUB(m.max_date, INTERVAL 30 DAY) AND m.max_date
),

-- Step 3: Calculate the age of customers and filter those aged between 25 and 35 years
EligibleCustomers AS (
SELECT c.customer_Id,
c.DOB,
YEAR(m.max_date) - YEAR(c.DOB) - (DATE_FORMAT(m.max_date, '%m%d') < DATE_FORMAT(c.DOB, '%m%d')) AS age
FROM customers_new c
CROSS JOIN MaxTranDate m
WHERE YEAR(m.max_date) - YEAR(c.DOB) - (DATE_FORMAT(m.max_date, '%m%d') < DATE_FORMAT(c.DOB, '%m%d')) BETWEEN 25 AND 35
),

-- Step 4: Join recent transactions with eligible customers and calculate net total revenue
NetTotalRevenue AS (
SELECT SUM(t.total_amt) AS net_total_revenue
FROM RecentTransactions t
JOIN EligibleCustomers e ON t.cust_id = e.customer_Id
)

-- Step 5: Select the result
SELECT net_total_revenue
FROM NetTotalRevenue;
```

Result Grid	
	net_total_revenue
▶	74885.84999999999

## 12. Which product category has seen the max value of returns in the last 3 months of transactions?

```
-- Step 1: Determine the maximum transaction date
WITH MaxTranDate AS (
    SELECT MAX(tran_date) AS max_date
    FROM transactions_new
),

-- Step 2: Filter transactions within the last 3 months from the maximum transaction date
RecentTransactions AS (
    SELECT t.*, m.max_date
    FROM transactions_new t
    CROSS JOIN MaxTranDate m
    WHERE DATE(t.tran_date) BETWEEN DATE_SUB(m.max_date, INTERVAL 3 MONTH) AND m.max_date
),

-- Step 3: Join recent transactions with product category information and calculate returns
ReturnsByCategory AS (
    SELECT
        p.prod_cat_code,
        p.prod_cat,
        SUM(t.total_amt) AS total_returns
    FROM
        RecentTransactions t
    JOIN
        prod_cat_info p ON t.prod_cat_code = p.prod_cat_code
    WHERE
        t.total_amt < 0
    GROUP BY
        p.prod_cat_code, p.prod_cat
),
```

```
-- Step 4: Find the product category with the maximum value of returns
MaxReturnsCategory AS (
    SELECT
        prod_cat,
        total_returns
    FROM
        ReturnsByCategory
    ORDER BY
        total_returns DESC
    LIMIT 1
)

-- Step 5: Select the result
SELECT
    prod_cat,
    total_returns
FROM
    MaxReturnsCategory;
```

Result Grid		Filter Rows:	...
	prod_cat	total_returns	
▶	Electronics	-3491.8	

13. Which store-type sells the maximum products; by value of sales amount and by quantity sold?

```
WITH SalesByStoreType AS
(
    SELECT Store_type,
           SUM(Qty) AS total_quantity_sold,
           SUM(total_amt) AS total_sales_amount
      FROM transactions_new
     GROUP BY Store_type
)

SELECT Store_type, total_quantity_sold, total_sales_amount
  FROM |
    SalesByStoreType
 WHERE
    total_quantity_sold = (SELECT MAX(total_quantity_sold) FROM SalesByStoreType)
  OR total_sales_amount = (SELECT MAX(total_sales_amount) FROM SalesByStoreType);
```

	Store_type	total_quantity_sold	total_sales_amount
▶	e-Shop	22763	19824816.05000001

14. What are the categories for which average revenue is above the overall average.

```
WITH CategoryAvgRevenue AS (
    SELECT prod_cat_code, AVG(total_amt) AS avg_revenue
    FROM transactions_new
    GROUP BY prod_cat_code),

OverallAvgRevenue AS (
    SELECT AVG(total_amt) AS overall_avg_revenue
    FROM transactions_new)

SELECT p.prod_cat, c.avg_revenue
FROM CategoryAvgRevenue c
JOIN OverallAvgRevenue o ON c.avg_revenue > o.overall_avg_revenue
JOIN prod_cat_info p ON c.prod_cat_code = p.prod_cat_code;
```

Result Grid		
	prod_cat	avg_revenue
▶	Clothing	2111.8707736486494
	Clothing	2111.8707736486494
	Clothing	2111.8707736486494
	Electronics	2189.151415884032
	Books	2112.8182633053298

15. Find the average and total revenue by each subcategory for the categories which are among top 5 categories in terms of quantity sold.

```
WITH TopCategories AS (
  SELECT
    prod_cat_code,
    SUM(Qty) AS total_quantity_sold
  FROM
    transactions_new
  GROUP BY
    prod_cat_code
  ORDER BY
    total_quantity_sold DESC
  LIMIT 5
),
SubcategoryRevenue AS (
  SELECT
    t.prod_subcat_code,
    p.prod_subcat,
    SUM(t.total_amt) AS total_revenue,
    AVG(t.total_amt) AS avg_revenue
  FROM
    transactions_new t
  JOIN
```

```
    prod_cat_info p ON t.prod_subcat_code = p.prod_sub_cat_code
  JOIN
    TopCategories tc ON t.prod_cat_code = tc.prod_cat_code
  GROUP BY
    t.prod_subcat_code, p.prod_subcat
)

SELECT
  prod_subcat,
  total_revenue,
  avg_revenue
FROM
  SubcategoryRevenue
ORDER BY
  total_revenue DESC;
```

prod_subcat	total_revenue	avg_revenue
Mens	8143423.469999979	2091.2746456086234
Mens	6451859.635000015	2145.6134469571048
Kids	6451859.635000015	2145.6134469571048
Mobiles	6451859.635000015	2145.6134469571048
Women	6451859.635000015	2145.6134469571048
Audio and video	6360967.860000008	2125.281610424326
Non-Fiction	6360967.860000008	2125.281610424326
Kitchen	6360967.860000008	2125.281610424326
Kids	6296891.119999999	2053.110896641669
Women	6296891.119999999	2053.110896641669
Comics	6296891.119999999	2053.110896641669
Children	4318677.024999997	2098.4825194363443
Bath	4318677.024999997	2098.4825194363443
Academic	4205226.674999994	2072.561200098568
Tools	4205226.674999994	2072.561200098568
Women	4071711.734999994	2091.2746456086284

# Performance Insights

## Summary

E-shop  
is preferred by customers.

January 2011–December 2014 is the data period.

City Code 3: Maximum number of clients (595).  
Books, Electronics, Clothes, Kitchen, and Household are the categories.

\$74,885 in revenue from 30 sales.

---

## Key Recommendations:-

Enhance Online Shopping Experience:- Upgrade the website and mobile app for improved user experience.- Ensure quicker and more dependable delivery services.

Focus on High-Performing Categories:- Broaden the product selection in Books and Electronics.- Introduce exclusive deals in these categories.

Target City Code 3:- Implement localized marketing campaigns and special promotions to capitalize on the high customer base.

Optimize Other Categories:- Evaluate the Clothing and Kitchen/Home categories for enhancements based on customer feedback.

---

# THANKS !



vishnugs32@gmail.com



vishnugnath



vishnugnath