



...

E-COMMERCE RETAIL DATA ANALYSIS

Presentation

2024

By: Sankar M



AGENDA

- ✓ **INTRODUCTION**
- ✓ **PROJECT OBJECTIVE**
- ✓ **DATABASE TABLES**
- ✓ **DATABASE SCHEMA**
- ✓ **PRIMARY KEY, FOREIGN KEY**
- ✓ **OBJECTIVE QUERY**
- ✓ **Performance Insights**
- ✓ **Key Recommendations**



Introduction

This project focuses on analyzing data from an e-commerce retail business to uncover valuable insights that can inform strategic decision-making. By examining key areas such as customer behaviour, transaction trends, and product performance, we aim to provide actionable recommendations that enhance customer engagement, optimise sales, and improve overall business efficiency.



Project Objective

The objective of this project is to extract valuable insights from e-commerce retail data to enhance business performance and customer experience. By analysing customer retention and loyalty, we aim to develop personalised marketing strategies and incentives. Additionally, this analysis will identify trends, patterns, customer behaviours, sales growth, profitability, and transactional dynamics. The insights gained will provide actionable recommendations and support informed decision-making for future growth.



DATABASE SCHEMA

Customer Table



The file contains the information about customers in the E-Commerce Retail Dataset. It includes variables such as customer_id, DOB, Gender and city_code. The schema of the dataset is 4 variables and 5647 records

The file contains the information about all transactions occur in the E-Commerce Retail Dataset. It includes variables such as transaction_id, customer_id, transaction_id, prod_subcat_code, prod_cat_code, quantity, rate, tax, total_amount and store_type. The schema of the dataset is 10 variable and 23053 records.



Transaction Table

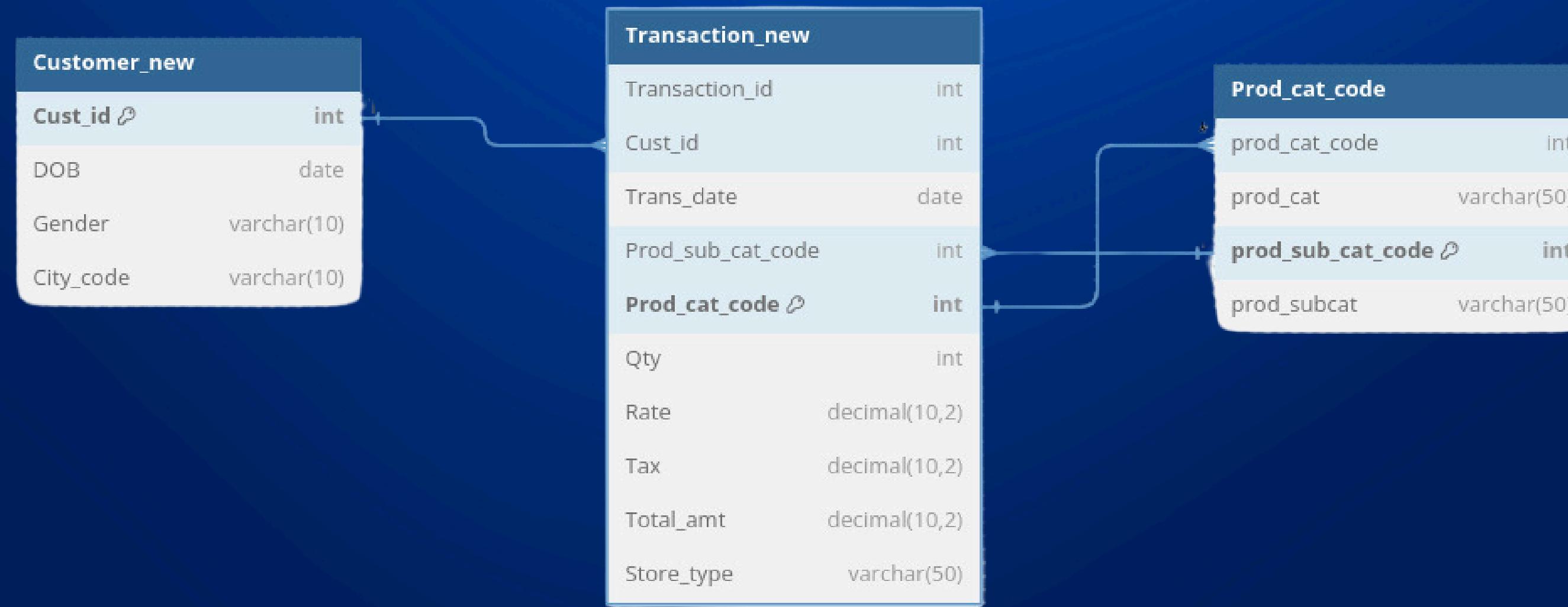
Product Table



The file contains the information about the product availability in the E-Commerce Retail Dataset. It includes variables such as prod_cat_code, prod_cat, prod_sub_cat_code and prod_subcat. The schema of the dataset is 4 variables and 23 records.

DATABASE SCHEMA

6 ...



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 KEY

A foreign key is a field (or collection of fields) in one table that refers to the primary key in another table.

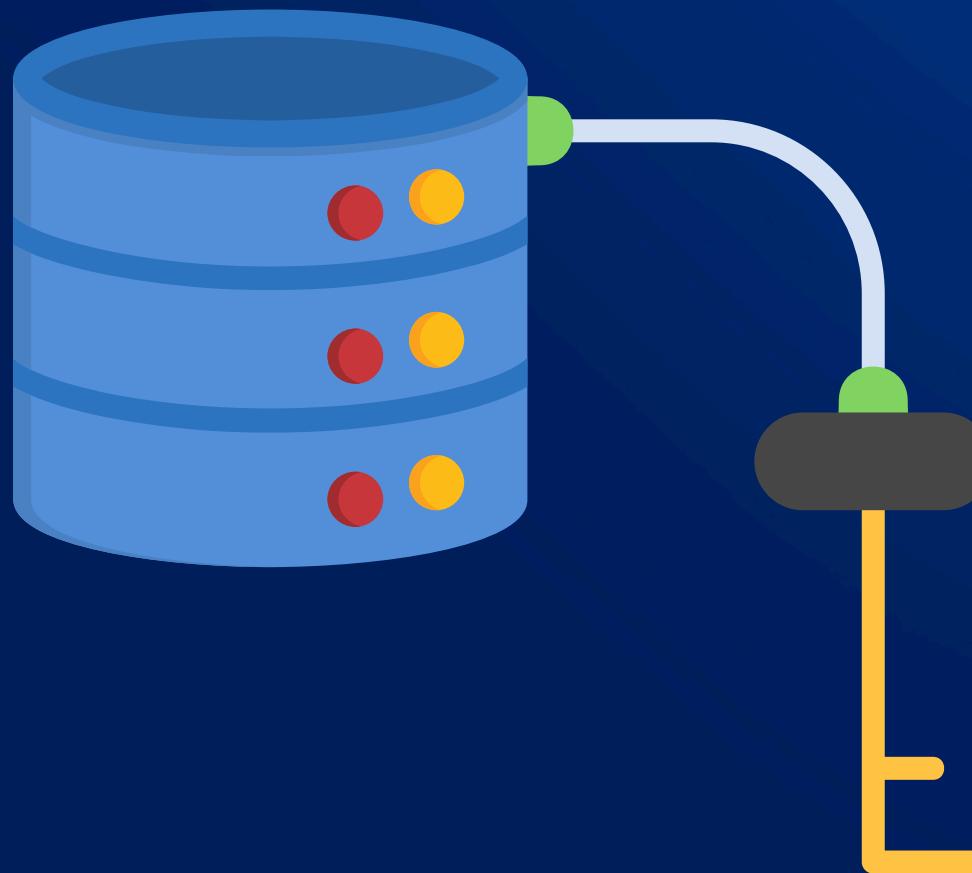
EXAMPLE:

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

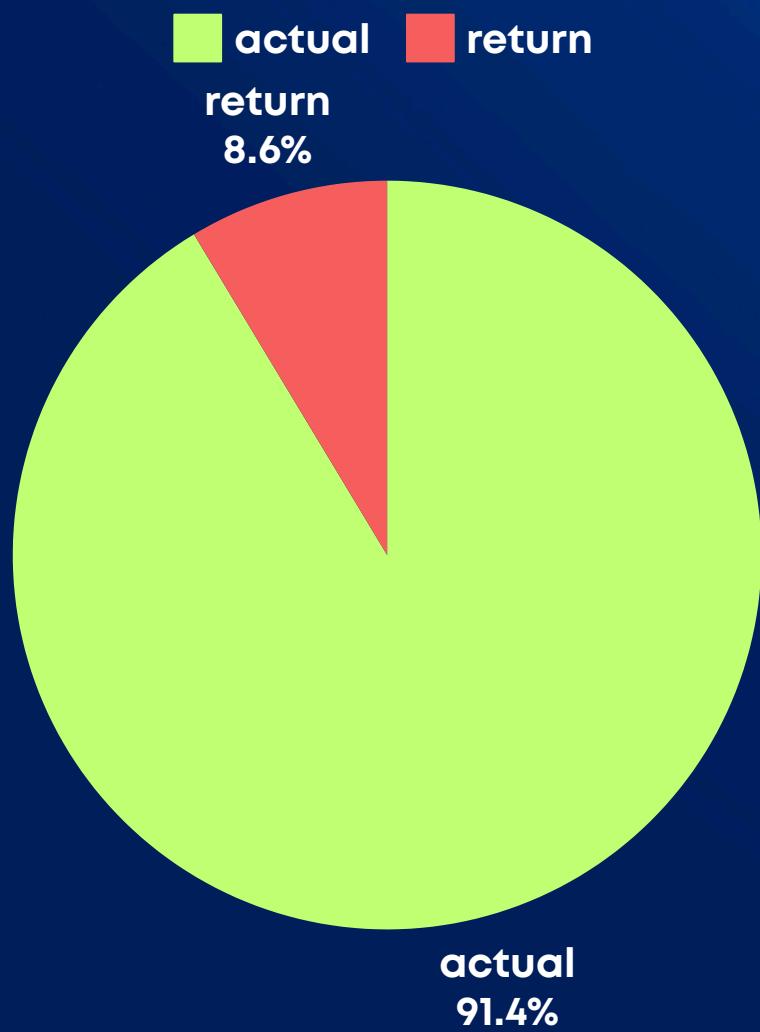
8 ...

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

```
select count(*) from Transactions_new tn ;
```

```
select count(*) from Customers_new cn ;
```

```
select count(*) from prod_cat_info pci ;
```



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

```
select count(*) from Transactions_new tn where total_amt < 0 ;
```

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 trans_start_date,  
    MAX(tran_date) AS tras_end_date,  
    DATEDIFF(MAX(tran_date), MIN(tran_date)) AS total_trans_days,  
    TESTAMPDIFF(MONTH, MIN(tran_date), MAX(tran_date)) AS  
total_trans_in_months,  
    TESTAMPDIFF(YEAR, MIN(tran_date), MAX(tran_date)) AS  
total_trans_in_years  
FROM  
    Transactions_New;
```

1	start date ↕	tras end date ↕	123 total trans days ↕	123 total trans in months ↕	123 total trans in years ↕
1	2011-01-02	2014-12-02	1,430	47	3



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

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

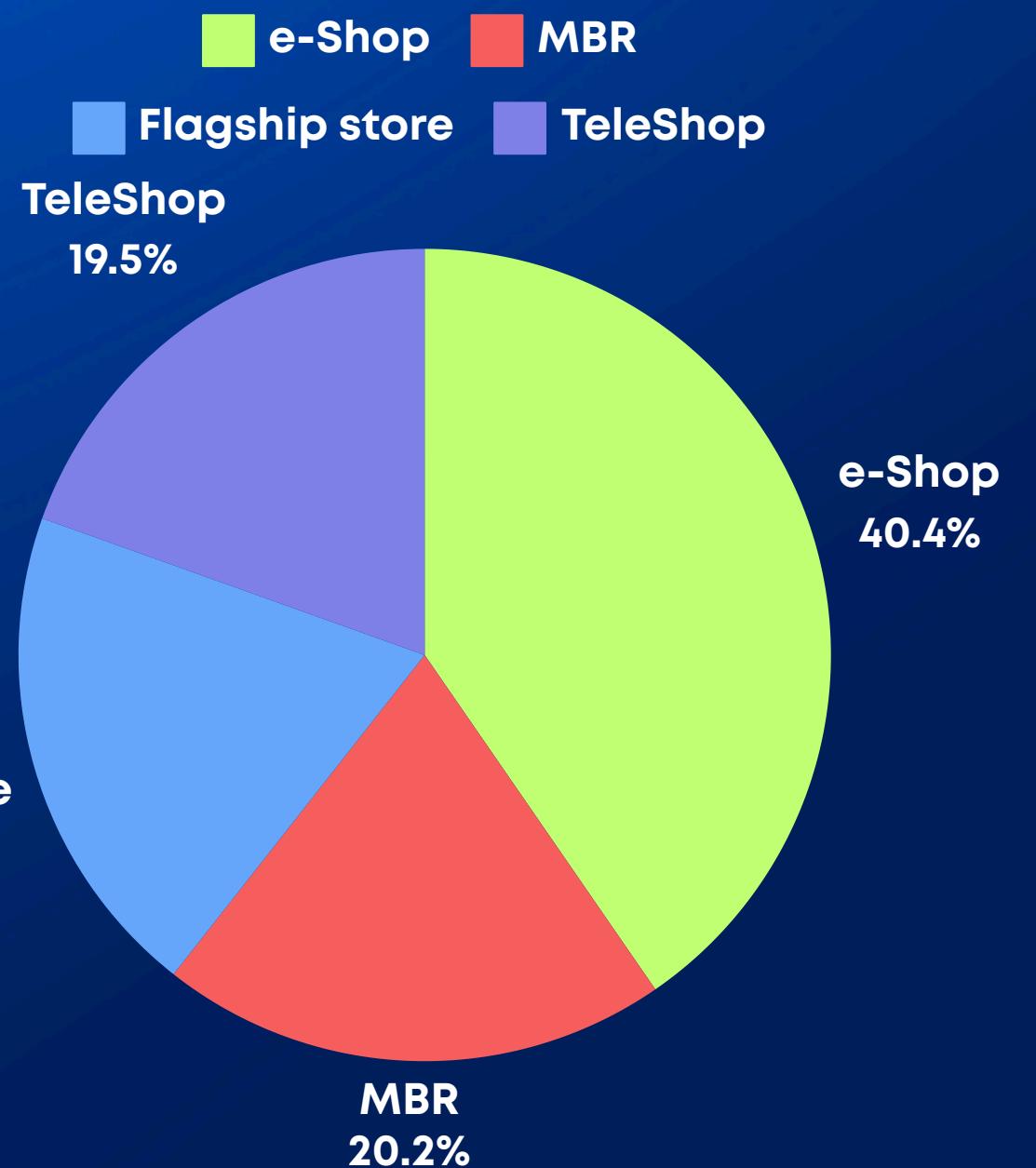
DATA ANALYSIS

10 •••

Which channel is most frequently used for transactions?

```
SELECT  
    Store_type,  
    COUNT(*) AS transaction_count  
FROM  
    Transactions_new  
GROUP BY  
    Store_type  
ORDER BY  
    transaction_count DESC
```

	Store type	transaction count
1	e-Shop	9,311
2	MBR	4,661
3	Flagship store	4,577
4	TeleShop	4,504

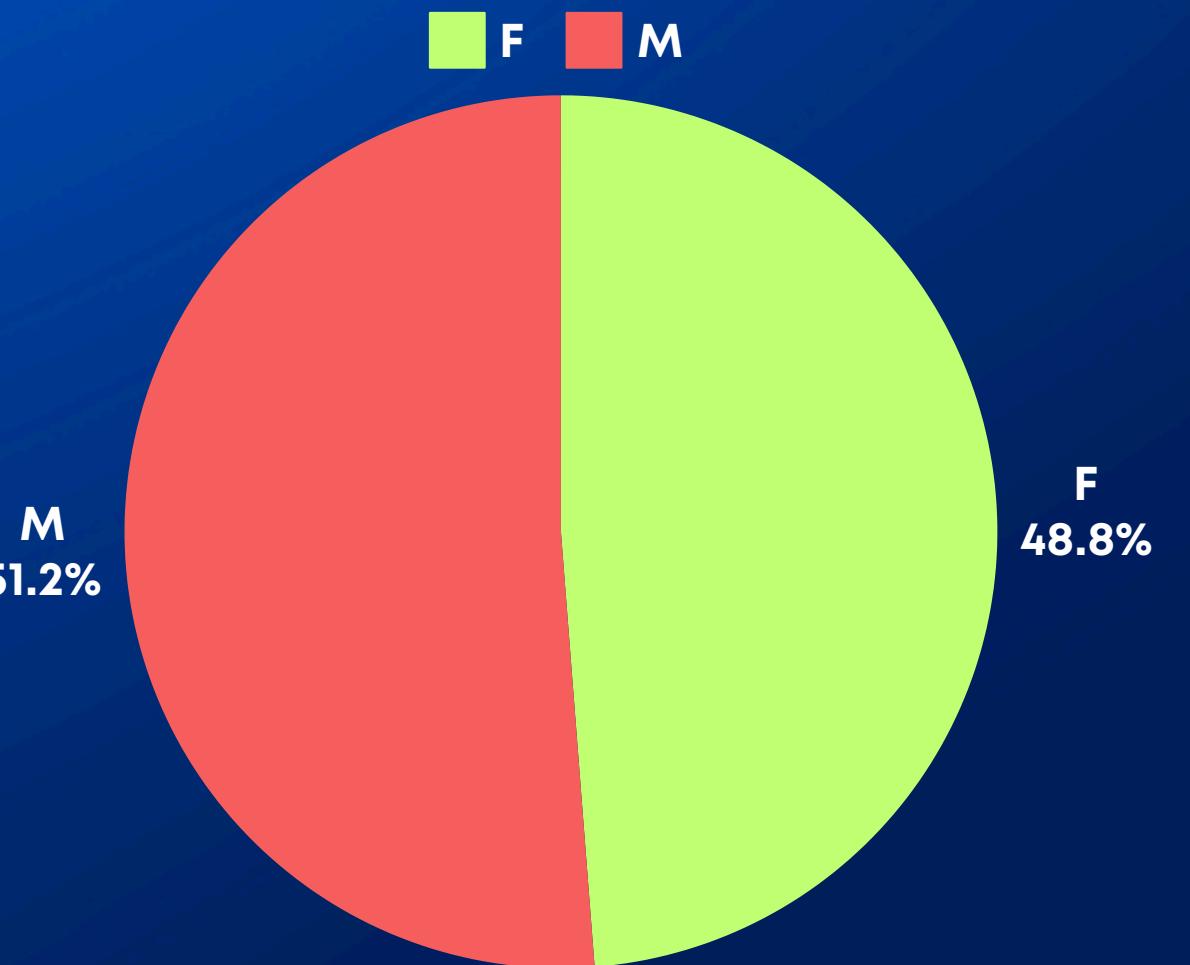


DATA ANALYSIS

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

```
SELECT  
    Gender,  
    COUNT(DISTINCT customer_id) AS customer_count  
FROM  
    Customers_new  
where Gender in ('M','F')  
GROUP BY  
    Gender;
```

	ABC Gender	123 customer count
1	F	2,753
2	M	2,892



DATA ANALYSIS

12 •••

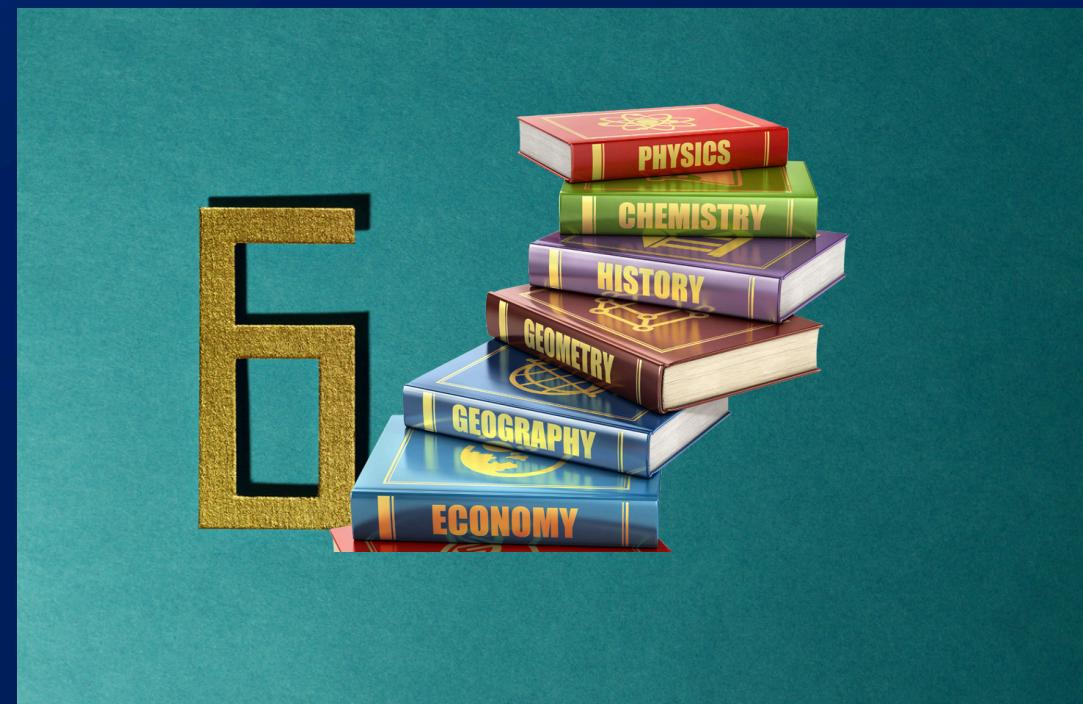
From which city do we have the maximum number of customers and how many?

```
select city_code,count(DISTINCT customer_Id) as  
total_customer  
from Customers_new cn  
group by city_code  
order by count(DISTINCT customer_Id) desc  
limit 1
```

	123 city code ↑↓	123 total customer ↑↓
1	3	595

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

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



DATA ANALYSIS

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

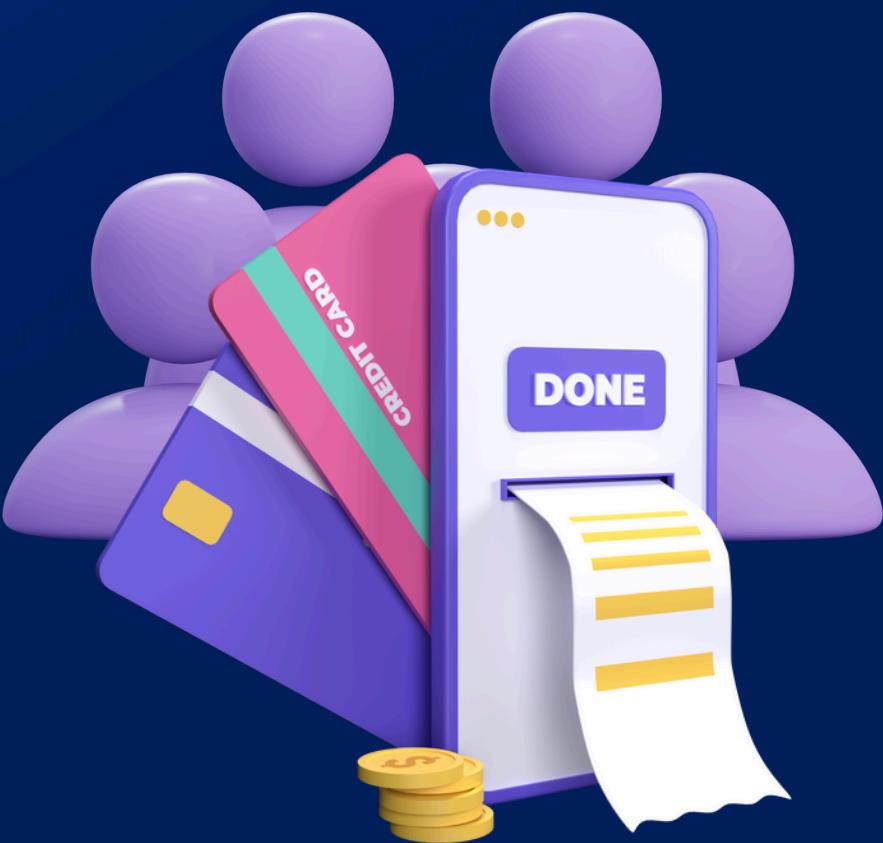
```
select pci.prod_cat,sum(total_amt) as total_amt from
Transactions_new tn
join prod_cat_info pci on pci.prod_cat_code =
tn.prod_cat_code and pci.prod_sub_cat_code =
tn.prod_subcat_code
where (pci.prod_cat = 'Electronics' or pci.prod_cat = 'Books')
GROUP by pci.prod_cat
```

123	total amt	T
1	23,545,157.67499993	

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

```
with answer as (select cust_id,count(transaction_id) from
Transactions_new tn
where total_amt > 0
group by cust_id
having count(transaction_id) > 10)
select count(cust_id)as customers_count from answer
```

6



DATA ANALYSIS

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

```
select tn.Store_type,sum(total_amt) as revenue from Transactions_New tn  
Left join prod_cat_info pci  on  
tn.prod_subcat_code = pci.prod_sub_cat_code and  
tn.prod_cat_code = pci.prod_cat_code  
where tn.Store_type = 'Flagship store' and (prod_cat = 'Electronics' or prod_cat = 'Clothing');
```



	ABC Store type	123 revenue
1	Flagship store	3,409,559.2700000056

DATA ANALYSIS

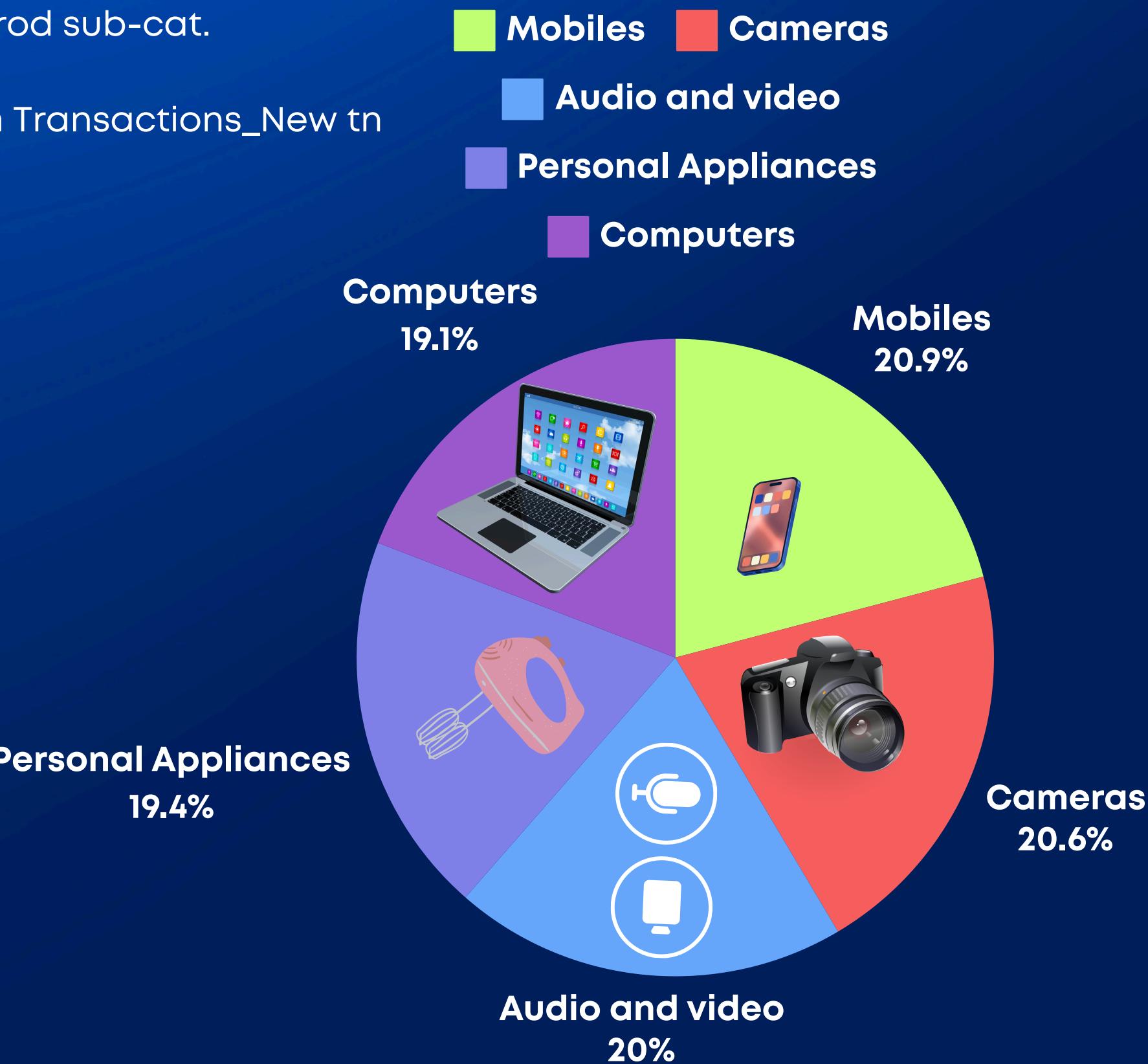
15 •••

What is the total revenue generated from "Male" customers in "Electronics" category? Output should display total revenue by prod sub-cat.

```
select prod_subcat,sum(total_amt) as revenue from Transactions_New tn  
left join Customers_new cn on  
cn.customer_id = tn.cust_id  
Left join prod_cat_info pci on  
tn.prod_subcat_code = pci.prod_sub_cat_code and  
tn.prod_cat_code = pci.prod_cat_code  
where gender = 'M' and prod_cat = 'Electronics'  
group by prod_subcat  
order by revenue desc;
```



	ABC prod subcat	123 revenue
1	Mobiles	1,192,413.2349999996
2	Cameras	1,172,702.244999998
3	Audio and video	1,138,983.1700000002
4	Personal Appliances	1,107,593.4349999987
5	Computers	1,091,417.34



DATA ANALYSIS

16 •••

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, total_sales, total_returns,
        RANK() OVER (ORDER BY total_sales DESC) AS sales_rank,
        (total_sales / SUM(total_sales) OVER ()) * 100 AS sales_percentage,
        (total_returns / SUM(total_returns) OVER ()) * 100 AS returns_percentage
    FROM SubcategorySales
)
SELECT prod_subcat, total_sales, sales_percentage, total_returns, returns_percentage
FROM TopSubcategories
WHERE
    sales_rank <= 5;

```

	ABC prod subcat ↑	123 total sales ↑	123 sales percentage ↑	123 total returns ↑	123 returns percentage ↑
1	Mens	12,392,937.44000004	10.3130298779	-1,545,406.5900000026	10.4376877969
2	Mens	8,451,811.290000023	7.0333432068	-1,111,458.725	7.5068006347
3	Kids	8,451,811.290000023	7.0333432068	-1,111,458.725	7.5068006347
4	Mobiles	8,451,811.290000023	7.0333432068	-1,111,458.725	7.5068006347
5	Women	8,451,811.290000023	7.0333432068	-1,111,458.725	7.5068006347

DATA ANALYSIS

17 •••

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?

```
WITH max_tran_date AS (
    SELECT MAX(tran_date) AS max_date
    FROM Transactions_new
),
last_30days_sales AS (
    SELECT t.cust_id, t.tran_date, t.total_amt, m.max_date
    FROM Transactions_new t
    CROSS JOIN max_tran_date m
    WHERE t.tran_date BETWEEN DATE_SUB(m.max_date, INTERVAL 30 DAY) AND m.max_date
),
age_btwn_2530 AS (
    SELECT c.customer_id, YEAR(m.max_date) - YEAR(c.DOB) AS age
    FROM Customers_new c
    CROSS JOIN max_tran_date m
    WHERE YEAR(m.max_date) - YEAR(c.DOB) BETWEEN 25 AND 35
),
net_rev AS (
    SELECT SUM(t.total_amt) AS net_total_revenue
    FROM last_30days_sales t
    JOIN age_btwn_2530 e ON t.cust_id = e.customer_id
)
SELECT net_total_revenue
FROM net_rev;
```



wwwmax_tran_date AS (SELECT	
1	123 net total revenue ↑
1	74,885.85

DATA ANALYSIS

18 ...

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

```
WITH MaxTranDate AS (
    SELECT MAX(tran_date) AS max_date
    FROM Transactions_new
),
Last90DaysReturns AS (
    SELECT
        SUM(CASE WHEN tn.total_amt < 0 THEN tn.total_amt ELSE 0 END) AS return_amount,
        pci.prod_cat
    FROM
        Transactions_new tn
    JOIN
        MaxTranDate m ON tn.tran_date BETWEEN DATE_SUB(m.max_date, INTERVAL 90 DAY) AND m.max_date
    LEFT JOIN
        prod_cat_info pci ON tn.prod_subcat_code = pci.prod_sub_cat_code
            AND tn.prod_cat_code = pci.prod_cat_code
    GROUP BY
        pci.prod_cat
)
SELECT
    prod_cat, return_amount FROM Last90DaysReturns
ORDER BY return_amount
LIMIT 1;
```



	ABC prod cat	T <small>±</small>	123 return amount	T <small>±</small>
1	Home and kitchen		-9,840.025	

DATA ANALYSIS

19 ...

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

```
SELECT  
    Store_type,  
    SUM(total_amt) AS total_amt,  
    SUM(Qty) AS total_qty  
FROM  
    Transactions_New  
GROUP BY  
    Store_type  
ORDER BY  
    SUM(total_amt) DESC,  
    SUM(Qty) DESC  
LIMIT 1;
```



	Store type	total amt	total qty
1	e-Shop	19,824,816.05000001	22,763

DATA ANALYSIS

20 ...

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

```
SELECT
    p.prod_cat,
    AVG(t.total_amt) AS avg_cat_rev
FROM
    Transactions_new t
JOIN
    prod_cat_info p ON t.prod_cat_code = p.prod_cat_code
GROUP BY
    p.prod_cat
HAVING
    AVG(t.total_amt) > (SELECT AVG(total_amt) FROM Transactions_new);
```

	ABC prod cat ↑↓	123 avg cat rev ↑↓
1	Electronics	2,189.151415884
2	Books	2,112.8182633053
3	Clothing	2,111.8707736487



DATA ANALYSIS

21 •••

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
)
SELECT
  p.prod_cat,
  AVG(t.total_amt) AS avg_revenue,
  SUM(t.total_amt) AS total_revenue
FROM
  Transactions_New t
JOIN
  prod_cat_info p ON t.prod_cat_code = p.prod_cat_code
JOIN
  TopCategories tc ON t.prod_cat_code = tc.prod_cat_code
GROUP BY
  p.prod_cat ;
```



Performance Insights

Overview

- ✓ Customer Preference: E-shop is favored.
- ✓ Data Period: January 2011 - December 2014.
- ✓ City Code 3: Highest customer count (595).
- ✓ Categories: Books, Electronics, Clothing, Kitchen, and Home.
- ✓ Revenue: \$74,885 from 30 sales.



Key Recommendations

23 ...

Enhance Online Shopping Experience:

- Improve website and mobile app.
- Ensure faster and reliable delivery.

Focus on High-Performing Categories:

- Expand product range and offer exclusive deals in Books and Electronics.

Target City Code 3:

- Localized marketing campaigns and special promotions to leverage the high customer base.

Optimize Other Categories:

- Analyze Clothing and Kitchen/Home for improvements based on customer feedback.



Thank You



8124310594



vmgomathisankar400@gmail.com