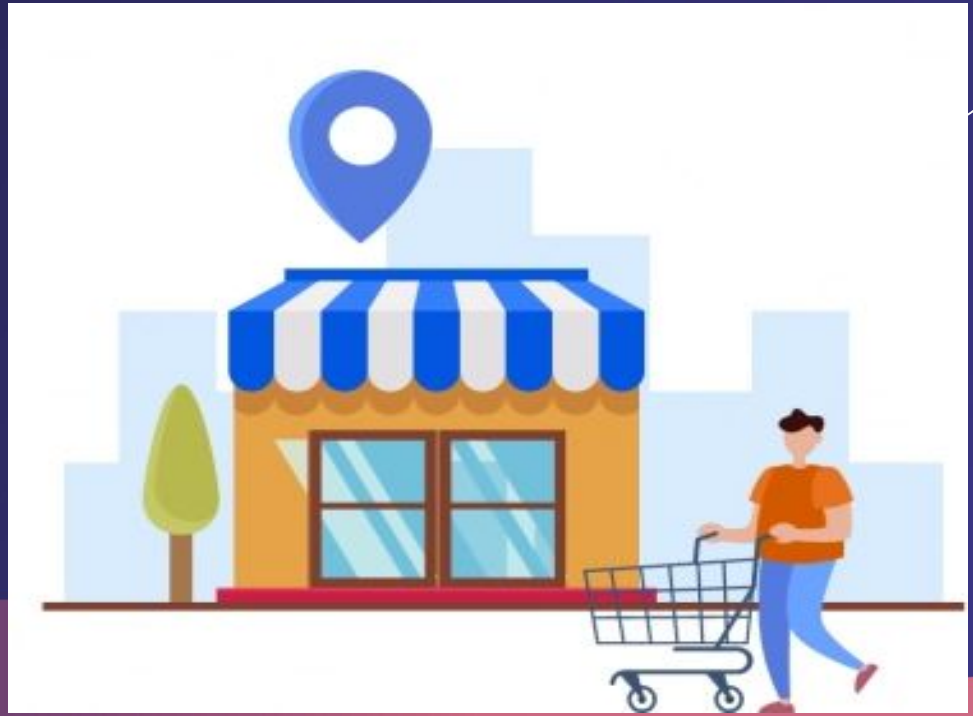
E-COMMERCE RETAIL DATA ANALYSIS

Presentation by

Ritu Kumari



AGENDA

- ✓ INTRODUCTION
- ✓ DATASET OVERVIEW
- ✓ DATABASE SCHEMA
- ✓ BUSINESS GOAL
- ✓ DATA PREPARATION & UNDERSTANDING
- ✓ DATA ANALYSIS
- ✓ KEY RECOMMENDATION

INTRODUCTION

01 By leveraging three key datasets—Transactions, Customers, and Product Categories—we will uncover valuable insights to inform strategic business decisions and enhance overall performance

02 This project aims to provide a comprehensive analysis of customer behavior, sales trends, and product performance within an e-commerce retail business context.

DATASET OVERVIEW

Transactionsnew.csv

- Description: Contains detailed records of customer transactions, capturing essential information such as transaction ID, customer ID, transaction date, product category, product subcategory, and transaction amount.

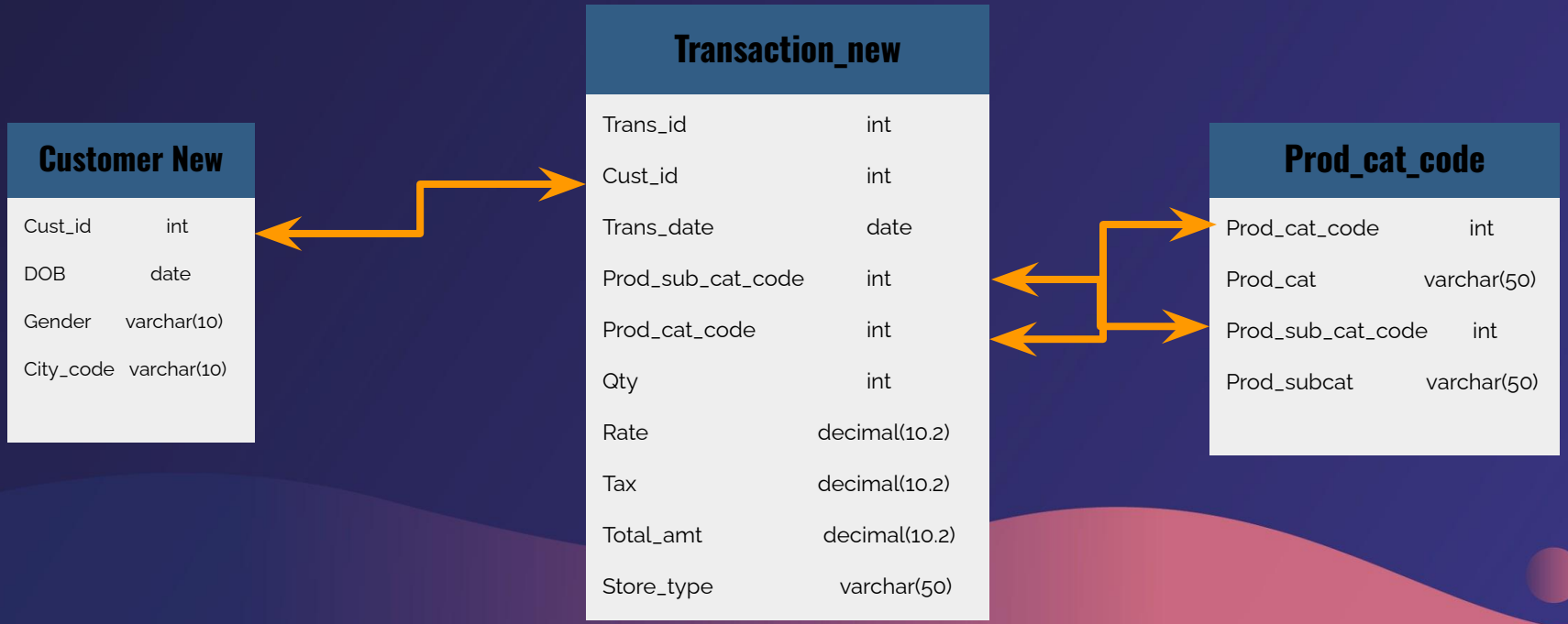
Customers_new.csv

- Description: Provides information about customers, including customer ID, name, date of birth, gender, and city code, enabling demographic analysis and customer segmentation.

prod_cat_info.csv

- Description: Includes product category information, detailing product category codes, names, subcategory codes, and names, which is crucial for analyzing product performance and market basket analysis.

DATABASE SCHEMA



BUSINESS GOAL

- **Enhance Customer Understanding:**

  Objective: Develop a deep understanding of customer segments, preferences, and behaviors using demographic and transaction data.

  Outcome: Create targeted marketing campaigns and personalized customer experiences to increase customer satisfaction and loyalty

- **Increase Sales and Revenue:**

  Objective: Identify trends, peak periods, and product performance to optimize sales strategies.

  Outcome: Implement effective sales tactics and promotions to boost revenue and improve overall sales performance.

- **Strategic Decision Making:**

  Objective: Provide actionable insights from data analysis to inform strategic business decisions.

  Outcome: Enable data driven decision-making processes to support business growth and operational efficiency.

- **Enhance Cross Selling and Upselling**:

  Objective: Discover product associations through market basket analysis to enhance cross selling and upselling strategies.

  Outcome: Increase average transaction value and customer satisfaction by recommending complementary products.

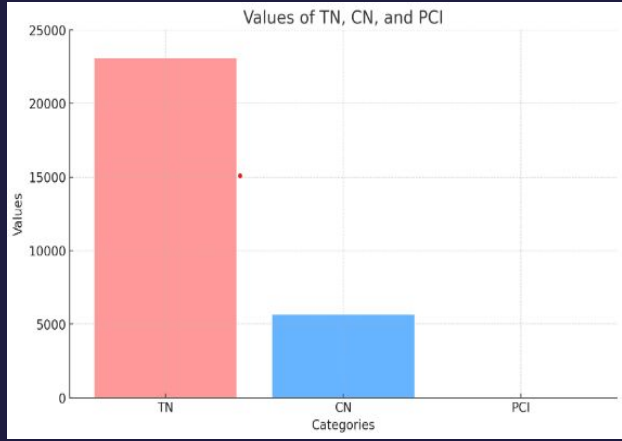
DATA PREPARATION & UNDERSTANDING

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

```
select count(*) from e_commerce.transactions_new;

select count(*) from e_commerce.customers_new;

select count(*) from e_commerce.prod_cat_info;
```

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

```
select count(*) from  e_commerce.transactions_new where total_amt < 0 ;
```

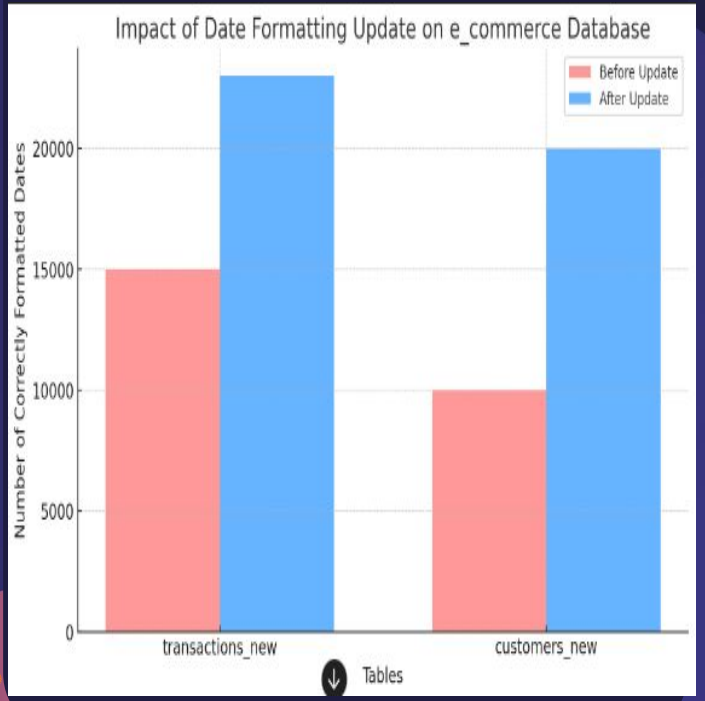


Result Grid

	count(*)
▶	2177

You would have noticed, the dates provided across the datasets are not in a correct format. As first steps, please convert the date variables into valid date formats before proceeding ahead?

```
UPDATE e_commerce.transactions_new
    SET tran_date =
    STR_TO_DATE(tran
    _date, '%d-%m-%Y')
    WHERE tran_date
    IS NOT NULL;

    ALTER TABLE e_commerce.transactions_newMODIFY COLUMN tran_date DATE NOT

    NULL;


    UPDATE e_commerce.customers_new


    SET DOB = STR_TO_DATE(DOB, '%d-%m-%Y')


    WHERE DOB IS NOT NULL;ALTER TABLE
    e_commerce.customers_new MODIFY
    COLUMN DOB DATE NOT NULL;
```

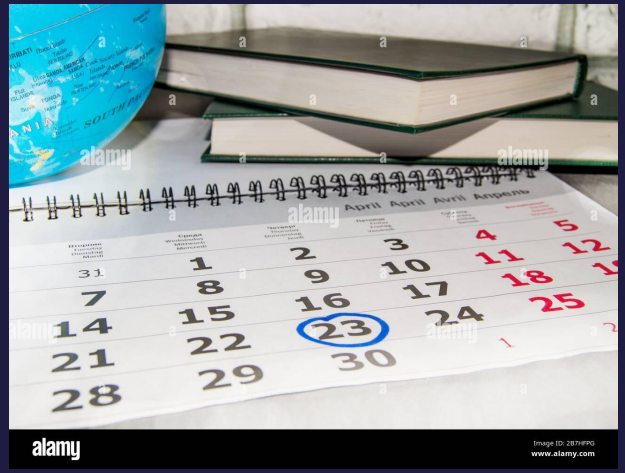


What is the time range of the transaction data available for analysis? Show the output in number of days, months and years simultaneously in different columns?

```
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,
TIMESTAMPDIFF(MONTH, MIN(tran_date), MAX(tran_date)) AS
total_trans_in_months,
TIMESTAMPDIFF(YEAR, MIN(tran_date), MAX(tran_date)) AS
total_trans_in_years
FROM   e_commerce.transactions_new;
```

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

trans_start_date	tras_end_date	total_trans_days	total_trans_in_months	total_trans_in_years
2011-01-02	2014-12-02	1430	47	3



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

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

Result Grid

prod_cat
Books

DATA ANALYSIS

Which channel is most frequently used for transactions?

```
SELECT  Store_type,  COUNT(*) AS transaction_countFROM e_commerce.transactions_newGROUP BY   Store_type ORDER BY transaction_count DESC;
```

Result Grid | Filter Rows:

Store_type	transaction_count
e-Shop	9311
MBR	4661
Flagship store	4577
TeleShop	4504

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

```
SELECT Gender, COUNT(DISTINCT ï»¿customer_Id) AS customer_countFROM e_commerce.customers_newwhere Gender in ('M','F')GROUP BY Gender;
```

Result Grid | Filter Rov

Gender	customer_count
F	2752
M	2891

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

```
select city_code,count(DISTINCT ï»¿customer_Id) as total_customer  from e_commerce.customers_new group by city_code order by count(DISTINCT ï»¿customer_Id) desc limit 1;
```

Result Grid | Filter Rows:

	city_code	total_customer
▶	3	595

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

```
select count(*) from e_commerce.prod_cat_info where prod_cat = 'Books‘
```

What is the maximum quantity of products ever ordered?

```
select max(Qty) as maximum_quantity    from e_commerce.transactions_new;
```

Result Grid

	maximum_quantity
▶	5

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

```
select round(SUM(total_amt),2) as net_revenue,prod_cat from e_commerce.transactions_new join e_commerce.prod_cat_info on
e_commerce.transactions_new.prod_cat_code= e_commerce.prod_cat_info.prod_cat_code where prod_cat in
("ELECTRONICS","BOOKS") group by prod_cat;
```

Result Grid | Filter Row

	net_revenue	prod_cat
▶	53612318.18	Electronics
	76936164.24	Books

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

Select cust_id, count(cust_id) as transactions from e_commerce.transactions_new where Qty>=0 group by cust_id having count(cust_id) >10;

Result Grid | Filter

cust_id	transactions
266794	11
270535	11
273014	11
274227	11
272741	11
270803	11

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

Select Store_type,round(sum(total_amt),0) as net_revenue,prod_cat from e_commerce.transactions_new join e_commerce.prod_cat_info on e_commerce.transactions_new. prod_subcat_code = e_commerce.prod_cat_info.prod_cat_codewhere prod_cat in ("ELECTRONICS","Clothing") group by prod_cat,store_typehaving store_type="Flagship store";

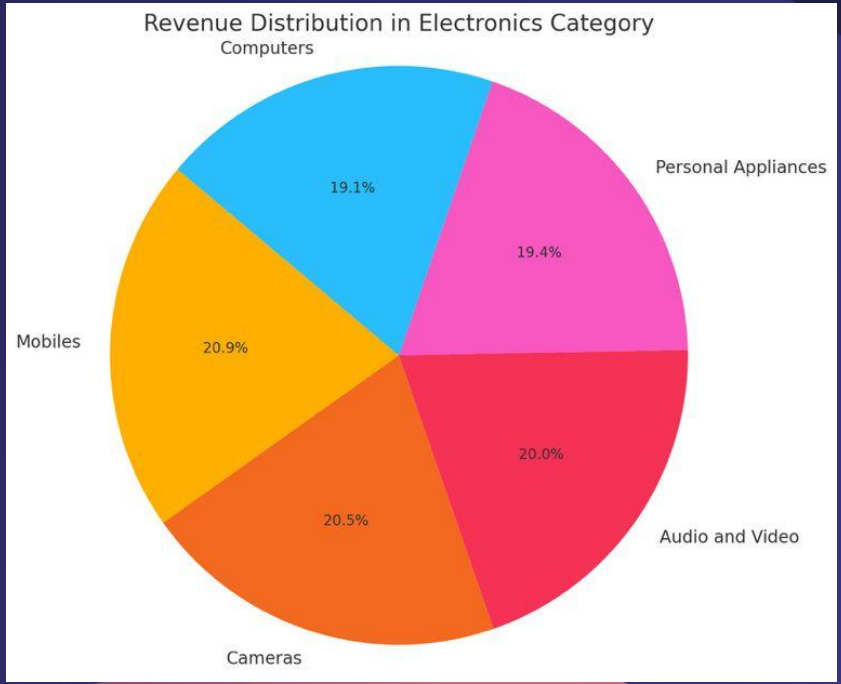
Store_type	net_revenue	prod_cat
Flagship store	6214421	Electronics
Flagship store	3470076	Clothing

What is the total revenue generated from "Male" customers in "Electronics" category? Output should display total revenue by prod subcat?

```
select prod_subcat ,sum(total_amt) as revenue from e_commerce.transactions_new t
left join e_commerce.customers_new c
on c.ï»¿customer_Id = t.cust_Id
Left join e_commerce.prod_cat_info pci on
t.prod_subcat_code = pci.prod_sub_cat_code and
t.prod_cat_code = pci.prod_cat_code
where gender = 'M' and prod_cat = 'Electronics'
group by prod_subcat
order by revenue desc;
```

Result Grid | Filter Rows:

prod_subcat	revenue
Mobiles	1192413.2349999996
Cameras	1167845.7699999986
Audio and video	1138983.1700000002
Personal Appliances	1107593.4349999987
Computers	1090794.1199999999



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 e_commerce.transactions_new),last_30days_sales AS (
          SELECT t.cust_id, t.tran_date, t.total_amt, m.max_date FROM e_commerce.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.ï»¿
¿customer_Id, YEAR(m.max_date) - YEAR(c.DOB) AS age                    FROM e_commerce.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.ï»¿¿customer_Id)SELECT net_total_revenueFROM net_rev;
```

Result Grid | Filter R

net_total_revenue
74885.85

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_qtyFROM e_commerce.transactions_new GROUP BY
Store_type ORDER BY     SUM(total_amt) DESC,  SUM(Qty) DESCLIMIT 1;
```

Result Grid | Filter Rows:

Store_type	total_amt	total_qty
e-Shop	19824816.05000001	22763

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 e_commerce.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 e_commerce.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 e_commerce.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;
```

Result Grid | Filter Rows:

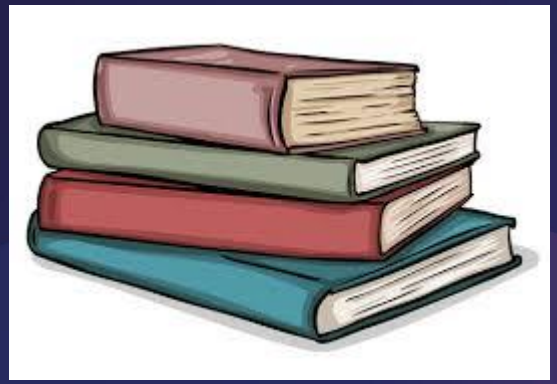
prod_cat	return_amount
Home and kitchen	-9840.025

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

```
SELECT p.prod_cat,  AVG(t.total_amt) AS avg_cat_revFROM  e_commerce.transactions_new tJOIN e_commerce.prod_cat_info p ON t.prod_cat_code = p.prod_cat_codeGROUP BY  p.prod_catHAVING AVG(t.total_amt) > (SELECT AVG(total_amt) FROM e_commerce.transactions_new);
```

Result Grid | Filter Rows:

prod_cat	avg_cat_rev
Clothing	2111.870773648651
Electronics	2189.1514158840396
Books	2112.818263305291



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

```
WITH TopCategories AS (  SELECT  prod_cat_code,   SUM(Qty) AS total_quantity_sold FROM e_commerce.transactions_new
GROUP BY prod_cat_code   ORDER BY total_quantity_sold DESC   LIMIT5)SELECT     p.prod_cat,   AVG(t.total_amt) AS
avg_revenue, SUM(t.total_amt) AS total_revenueFROM e_commerce.transactions_new tJOIN   e_commerce.prod_cat_info p ON
t.prod_cat_code = p.prod_cat_codeJOIN TopCategories tc ON t.prod_cat_code = tc.prod_cat_codeGROUP BY       p.prod_cat;
```

Result Grid | Filter Rows:

Store_type	total_amt	total_qty
e-Shop	19824816.05000001	22763

Key Recommendations

1. Customer Transaction Analysis:

- **Increase Average Transaction Value (ATV):**
  - Offer bundle deals or discounts for multiple purchases.
  - Implement upselling and cross-selling strategies during checkout.

2. New Product Category Analysis:

- **Market Research and Validation:**
  - Conduct thorough market research to understand demand and competition.
  - Use focus groups or surveys to gather feedback on potential new products.

3. Implementation Strategy

- **Prioritize Actionable Insights:**
  - Focus on insights that directly impact revenue, customer satisfaction, and operational efficiency.
  - Allocate resources to initiatives with the highest potential ROI based on data analysis.

- **Continuous Improvement:**
  - Foster a culture of data-driven decision-making within the organization.
  - Regularly review and update strategies based on new data and market dynamics

Thank you

kritu8578@gmai.com