Sales Performance Analysis of Walmart Stores Using Advanced MySQL Techniques

Presented by:
Vandana Bhavsar

Internshala Trainings Project
1-sep-2025

INTERNSHALA
internships that matter

Project Overview

This project analyzes detailed transaction data from multiple Walmart branches to uncover actionable insights into sales patterns, customer behavior, and operational efficiency.

- **Dataset:** Contains rich information on customer demographics, product lines, sales figures, and payment methods.
- **Objective:** To employ advanced MySQL techniques to answer critical business questions and inform strategic decisions.

Business Problem

Walmart aims to **optimize its sales strategies**. By analyzing historical data, we can identify key trends and performance indicators related to:

- **Sales Performance:** Which branches and products are performing best?
- **Customer Segmentation:** How do different customer types behave?
- **Product & Payment Trends:** What are the most popular products and payment methods?

Task 1: Identifying the Top Branch by Sales Growth Rate

Business Question: Which branch has exhibited the highest sales growth over time?

```
-- This CTE calculates the average monthly growth rate for each branch.
-- Step 1: Calculate total sales for each branch for each month.
-- We use DATE_FORMAT to extract month to group sales data.
WITH MonthlySales AS (
      SELECT
        Branch,
        DATE_FORMAT(Date, '%m') AS sales_month,
        round(SUM(Total),2) AS monthly_sales
    FROM
        walmartsales
    GROUP BY
        Branch,
        sales_month
        order by branch, sales_month
),
```

OUTPUT

	Branch	sales_month	monthly_sales
▶	A	01	38681.13
	A	02	29860.12
	A	03	37659.12
	B	01	37176.06
	B	02	34424.27
	B	03	34597.34
	C	01	40434.68
	C	02	32934.98
	C	03	37199.04

```sql
-- Step 2: For each branch, get the sales of the previous month.
-- We use the LAG() window function to look back one row (one month) within each branch's data.
SalesWithPreviousMonth AS (
    SELECT
        Branch,
        sales_month,
        monthly_sales,
        LAG(monthly_sales, 1) OVER (PARTITION BY Branch ORDER BY sales_month) AS previous_month_sales
    FROM
        MonthlySales
) ,
-- Step 3: Calculate the month-over-month growth rate for each branch.
-- The formula is ((Current Month Sales - Previous Month Sales) / Previous Month Sales) * 100
-- We handle the case where previous_month_sales is 0 to avoid division by zero errors.
MonthlyGrowthRate AS (
      SELECT
        Branch,
        sales_month,
        ((monthly_sales - previous_month_sales) / previous_month_sales) * 100 AS growth_rate
    FROM
        SalesWithPreviousMonth
    where previous_month_sales IS NOT NULL AND previous_month_sales > 0
)
```

```sql
-- Step 4: Calculate the average growth rate for each branch and find the top performer.
-- We group by branch, calculate the average of the growth rates, and order the results
-- to see the branch with the highest average growth at the top.
SELECT
    Branch,
    round(AVG(growth_rate),2) AS average_growth_rate
FROM
    MonthlyGrowthRate
GROUP BY
    Branch
ORDER BY
    average_growth_rate DESC limit 1;
```

All Branched Growth Rate Over Time(without ' **limit 1** ')

	Branch	average_growth_rate
▶	A	1.66
	C	-2.8
	B	-3.45

OUTPUT (Only Top Performer)

	Branch	average_growth_rate
▶	A	1.66

Key Insight & Recommendation

Insight

- Branch A demonstrated the highest average monthly sales growth, making it the top-performing branch in terms of growth momentum.

Recommendation

- Walmart should investigate the strategies at Branch A.

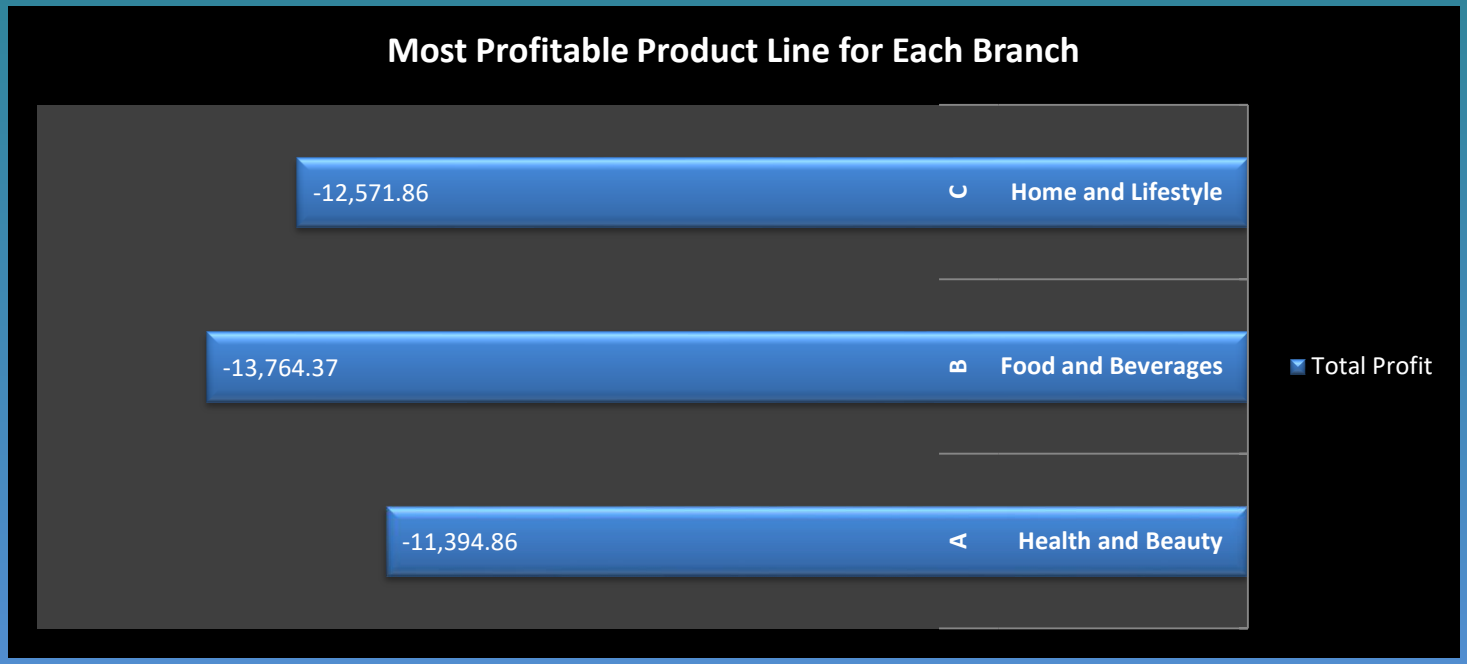
Task 2: Finding the Most Profitable Product Line for Each Branch

Business Question: Which product line generates the highest profit for each Walmart branch based on gross income and cost of goods sold?

```sql
/*
This query identifies the most profitable product line within each branch.
It first calculates the total profit for each product line per branch,
then ranks them to find the top performer in each branch.
*/
SELECT branch, product_line, round(profit,2)
FROM (
    SELECT branch, product_line, SUM(gross_income - cogs) AS profit,
           RANK() OVER (PARTITION BY branch ORDER BY SUM(gross_income - cogs) DESC) AS rnk
    FROM walmartsales
    GROUP BY branch, product_line
) AS ranked
WHERE rnk = 1;
```

OUTPUT

	Branch	product_line	total_profit
▶	A	Health and beauty	-11394.86
	B	Food and beverages	-13764.37
	C	Home and lifestyle	-12571.86



Key Insight & Recommendation

Insight:

- All three branches (A, B, and C) show losses in their top product lines.
- Branch B has the highest loss in Food and Beverages (−13,764.37).
- Branch A and C also show significant losses in Health and Beauty and Home and Lifestyle, respectively.

Recommendation:

- Review pricing and cost structure for these product lines—COGS may be too high or pricing too low.
- Analyze sales volume vs. cost to identify if low sales or high operational costs are driving losses.
- Consider promotional strategies or bundling to boost sales in underperforming categories.
- Evaluate product viability—discontinue or reposition products that consistently underperform.

Task 3: Analyzing Customer Segmentation Based on Spending

Business Question: How can Walmart segment its customers into High, Medium, and Low spenders based on their total purchase behavior to optimize targeting and retention strategies?

```sql
/*
This query segments customers based on their total spending behavior.
It calculates the total spend for each customer, then classifies them into
High, Medium, or Low tiers using conditional logic.
This helps identify top-value customers and tailor marketing or loyalty strategies.
*/
SELECT
    Customer_ID,
    CONCAT(ROUND(SUM(total), 0), ' $') AS Total_Spend,
    CASE
        WHEN SUM(total) > 25000 THEN 'High'
        WHEN SUM(total) > 20000 THEN 'Medium'
        ELSE 'Low'
    END AS Spending_Tier
FROM
    walmartsales
GROUP BY customer_ID
ORDER BY Total_spend DESC;
```

OUTPUT

Customer_ID	Total_Spend	Spending_Tier
8	26634 $	High
3	23402 $	Medium
2	23392 $	Medium
15	22674 $	Medium
1	22635 $	Medium
12	21721 $	Medium
11	21399 $	Medium
13	21064 $	Medium
14	21049 $	Medium
10	20724 $	Medium
6	20694 $	Medium
7	20628 $	Medium
9	19662 $	Low
5	19632 $	Low
4	17657 $	Low

Key Insight & Recommendation

Insight:

- Majority of top spenders fall under the **Medium Spending Tier**, despite spending over $20,000.
- Only **one customer** is classified as **High Tier**, even though others have similar spend levels.
- Customers with spend between **$17,000–$20,000** are placed in the **Low Tier**, which may not reflect their actual value.

Recommendation:

- **Reevaluate tier thresholds** to better align with actual spending behavior.
- Consider creating a **"Premium Medium" or "Upper Medium"** tier to capture high-value Medium customers.

Task 4: Detecting Anomalies in Sales Transactions

Business Question: Which sales transactions significantly deviate from the average for their product line, indicating potential pricing or data anomalies?

```sql
-- This query detects Anomalies in walmart Sales Transactions
-- Step 1: Create a CTE to calculate the average of the 'Total' for each 'Product line'.
WITH avg_sales AS (
    SELECT product_line,
           AVG(`Total`) AS avg_total
    FROM walmartsales
    GROUP BY product_line
)
-- Step 2: Join the main sales table with our new statistics table.
SELECT s.invoic_id, s.product_line, s.total,
       ROUND(a.avg_total, 2) AS AvgTotal,
       CASE
           WHEN s.total > a.avg_total * 1.5 THEN 'High Anomaly'
           WHEN s.total < a.avg_total * 0.5 THEN 'Low Anomaly'
           ELSE 'Normal'
       END AS AnomalyType
FROM walmartsales s
JOIN avg_sales a
  ON s.product_line = a.product_line
WHERE s.total > a.avg_total * 1.5
   OR s.total < a.avg_total * 0.5;
```

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

invoic_id	product_line	total	AvgTotal	AnomalyType
233-67-5758	Health and beauty	42.3675	323.64	Low Anomaly
430-53-4718	Health and beauty	633.108	323.64	High Anomaly
552-44-5977	Health and beauty	520.8	323.64	High Anomaly
764-44-8999	Health and beauty	30.996	323.64	Low Anomaly
148-41-7930	Health and beauty	734.706	323.64	High Anomaly
605-03-2706	Health and beauty	49.77	323.64	Low Anomaly
503-21-4385	Health and beauty	125.7165	323.64	Low Anomaly
333-23-2632	Health and beauty	74.088	323.64	Low Anomaly
433-08-7822	Health and beauty	579.8415	323.64	High Anomaly
450-42-3339	Health and beauty	888.405	323.64	High Anomaly
545-07-8534	Health and beauty	122.472	323.64	Low Anomaly
595-94-9924	Health and beauty	145.5825	323.64	Low Anomaly
725-54-0677	Health and beauty	629.16	323.64	High Anomaly
699-88-1972	Health and beauty	832.944	323.64	High Anomaly
101-81-4070	Health and beauty	131.922	323.64	Low Anomaly
176-78-1170	Health and beauty	106.5015	323.64	Low Anomaly
160-22-2687	Health and beauty	503.7375	323.64	High Anomaly
528-14-9470	Health and beauty	95.865	323.64	Low Anomaly
222-42-0244	Health and beauty	681.4395	323.64	High Anomaly
565-91-4567	Health and beauty	90.3	323.64	Low Anomaly
559-61-5987	Health and beauty	18.6375	323.64	Low Anomaly
359-94-5395	Health and beauty	97.419	323.64	Low Anomaly

Result 26

Output

Action Output

#	Time	Action	Message
54	17:14:49	WITH avg_sales AS (SELECT product_line, AVG(`Total`) AS avg_total FROM walmartsales G...	569 row(s) returned

OUTPUT

Key Insight & Recommendation

Key Insights

- Most of the unusual transactions are in the **Health and beauty** product line.
- The **average sale amount** is around ₹323, but some sales are much higher or much lower.
- **High anomalies** are above ₹485, and **low anomalies** are below ₹160.
- This shows that some products might be **overpriced or underpriced**, or there could be **mistakes in data entry**.
- These differences can affect how we understand customer spending and product performance.

Recommendations

- **Check the pricing** of Health and beauty items to make sure it's consistent.
- **Look into very low sales** to see if there were discounts or errors.
- **Review high sales** to confirm if they were correct or part of a bundle.
- Try to **group customers** based on their buying patterns to learn more.

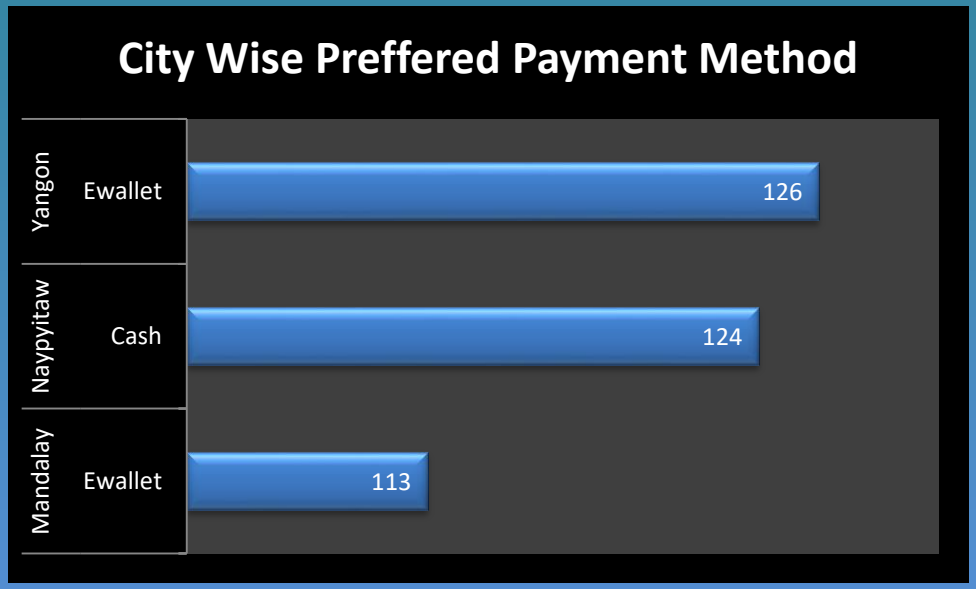
Task 5: Most Popular Payment Method by City

Business Question: What is the most preferred payment method in each city to help Walmart customize its local marketing and payment strategies?

```sql
/* This query finds the most frequently used payment method
in each city to help Walmart adjust its marketing and payment options. */
-- step 1 : find total count of payment per city
WITH payment_counts AS (
    SELECT city, payment, COUNT(*) AS total_pay_count
    FROM walmartsales
    GROUP BY city, payment
),
-- step 2: find most frequently used payment method in each city by giving rank to all
ranked_methods AS (
    SELECT *,
           RANK() OVER (PARTITION BY city ORDER BY total_pay_count DESC) AS rnk
    FROM payment_counts
)
SELECT city, payment, total_pay_count
FROM ranked_methods
WHERE rnk=1
ORDER BY total_pay_count desc;
```

OUTPUT

city	payment	total_pay_count
Yangon	Ewallet	126
Naypyitaw	Cash	124
Mandalay	Ewallet	113



Key Insight & Recommendation

Key Insights

- **Yangon and Mandalay** customers mostly use **Ewallet** for payments.
- **Naypyitaw** prefers **Cash**, showing less use of digital methods.
- Ewallet is the **most popular overall**, but preferences vary by city.

Recommendations

- **Promote Ewallet offers** in Yangon and Mandalay to match customer habits.
- In Naypyitaw, **encourage digital payments** by offering discounts or rewards for Ewallet use.
- Use this data to **customize payment promotions** for each city to improve customer experience and sales.

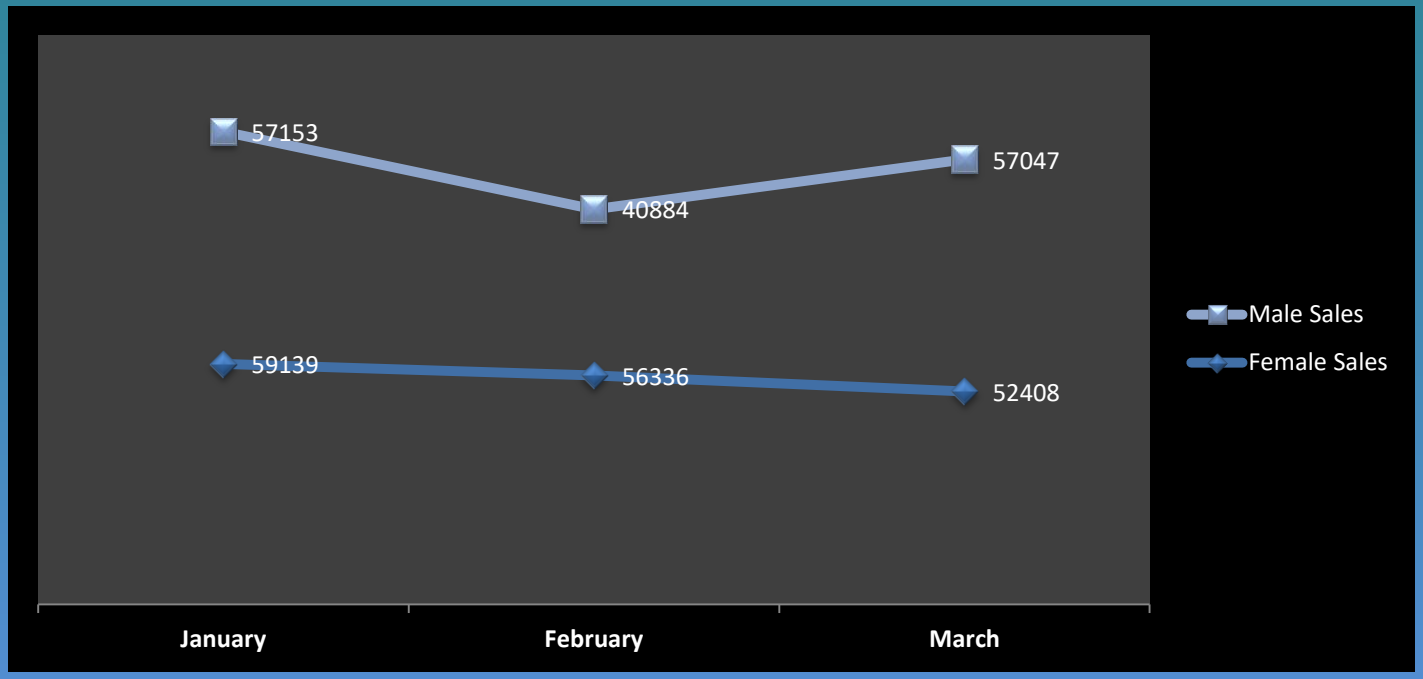
Task 6: Monthly Sales Distribution by Gender

Business Question: How do monthly sales differ between male and female customers, and which gender contributes more to sales each month?

```sql
-- This query help to identify male monthly sales and  females monthly sales
SELECT
    YEAR(date) AS Year,
    MONTHNAME(date) AS Month,
    ROUND(SUM(CASE
                WHEN Gender = 'Female' THEN Total
                ELSE 0
            END),
            0) AS 'Female Sales',
    ROUND(SUM(CASE
                WHEN Gender = 'Male' THEN Total
                ELSE 0
            END),
            0) AS 'Male Sales'
FROM
    walmartsales
GROUP BY year , month
ORDER BY MIN(Date);
```

OUTPUT

	Year	Month	Female Sales	Male Sales
▶	2019	January	59139	57153
	2019	February	56336	40884
	2019	March	52408	57047



Key Insight & Recommendation

Key Insights

- In all three months (January to March 2019), **female customers spent slightly more** than male customers.
- **February** shows the biggest gap, with female sales at ₹56,336 vs male sales at ₹40,884.
- Overall, both genders contribute strongly to monthly sales, but **female customers lead in total spending**.

Recommendations

- Create **special offers for female customers**, especially in months where their spending is higher.
- Use this data to **plan gender-focused promotions** during peak months like February.
- Keep tracking monthly trends to see if these patterns continue and adjust marketing strategies accordingly.

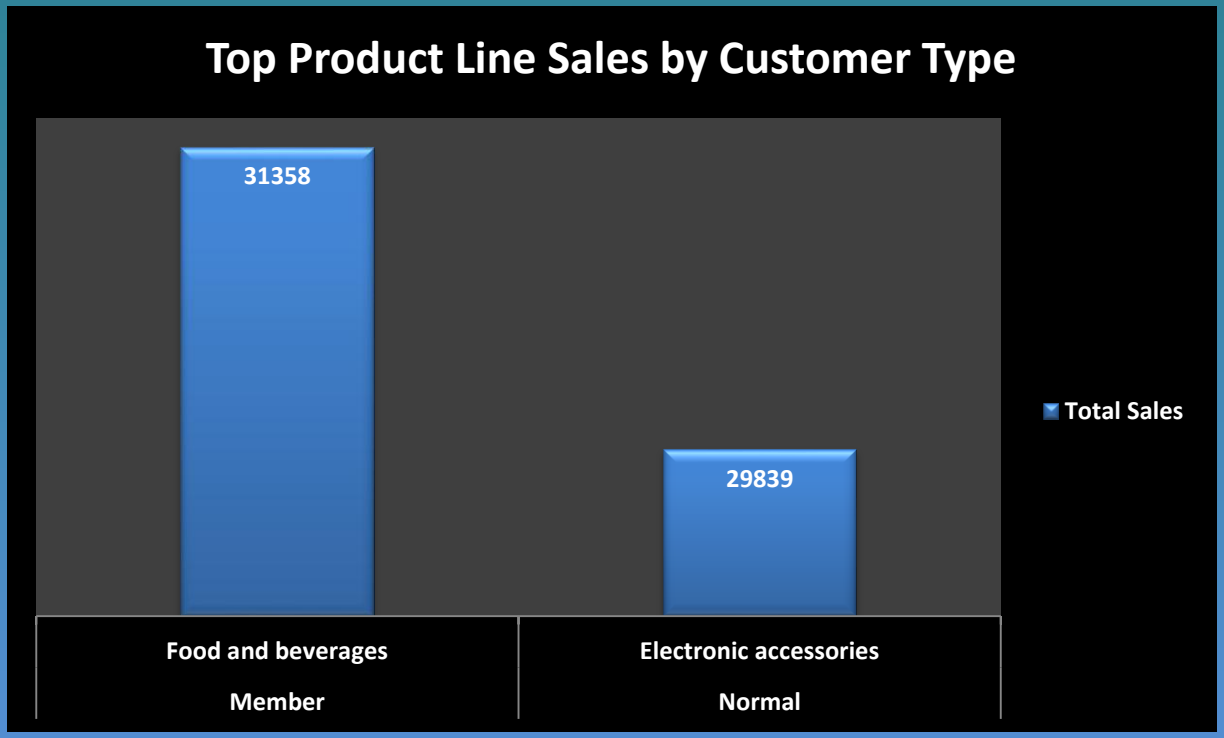
Task 7: Best Product Line by Customer Type

Business Question: Which product lines are most preferred by Walmart's Member and Normal customers?

```sql
/* This query finds the most popular product line for each customer type,
showing both the total number of purchases and the total sales revenue.
Popularity is ranked based on the purchase count.*/
WITH cutomer_type_count AS
(
SELECT customer_type, product_line, count(*) AS purchage_count, sum(total) AS total_sales
FROM walmartsales
GROUP BY customer_type,product_line
),
customer_type_rank As
(
 SELECT *,
 RANK() OVER (partition by customer_type order by total_sales desc) as rnk
 FROM cutomer_type_count
) SELECT customer_type AS 'Customer Type',
        product_line AS 'Product Line',
        purchage_count AS 'Purchage Count',
        round(total_sales,0) AS 'Total Sales'
        FROM customer_type_rank WHERE rnk=1;
```

OUTPUT

Customer Type	Product Line	Purchage Count	Total Sales
Member	Food and beverages	94	31358
Normal	Electronic accessories	92	29839



Key Insight & Recommendation

Key Insights

- **Members** prefer **Food and beverages**, both in number of purchases and total sales.
- **Normal customers** spend the most on **Electronic accessories**, showing interest in tech products.
- Each customer type has a **distinct preference**, which can guide targeted promotions.

Recommendations

- Offer **combo deals or loyalty points** on Food and beverages for Members.
- Promote **gadgets and tech accessories** to Normal customers through seasonal campaigns.
- Use this data to **personalize marketing** and improve customer satisfaction.

Task 8: Identifying Repeat Customers

Business Question: Which customers made repeat purchases within 30 days, helping Walmart understand loyalty and buying behavior?

```sql
/* This query identifies repeat customers within a 30-day timeframe. */
SELECT DISTINCT(customer_id) AS 'Repeat Customers'
FROM
    (SELECT
        a.customer_id,
        a.date AS FirstPurchase,
        b.date AS RepeatPurchase,
        DATEDIFF(b.date, a.date) AS DaysBetween
    FROM
        walmartsales a
    JOIN walmartsales b ON a.customer_id = b.customer_id
        AND b.date > a.date
        AND DATEDIFF(b.date, a.date) <= 30
    ORDER BY a.customer_id , a.date) AS WS
ORDER BY customer_id;
```

OUTPUT

	Repeat Customers
▶	1
	2
	3
	4
	5
	6
	7
	8
	9
	10
	11
	12
	13
	14
	15

Note: Consider customer_id as customer

Key Insight & Recommendation

Key Insights

- A total of **15 customers** returned within 30 days — showing early signs of loyalty.
- These repeat buyers may be more **engaged**, and could become long-term customers if nurtured.
- Even with a small sample, this behavior shows **potential for growth** in customer retention.

Recommendations

- **Identify and reward** these 15 repeat customers with special offers or loyalty perks.
- Consider adding **customer feedback** or satisfaction surveys to understand why they returned.
- Track repeat behavior monthly to see if the number grows — and adjust marketing accordingly.

Task 9: Finding Top 5 Customers by Sales Volume

Business Question: Which customers have generated the highest sales revenue, so Walmart can recognize and reward its top contributors?

```sql
/* Walmart wants to reward its top 5 customers who have
generated the most sales Revenue.*/
SELECT
    customer_id,
    ROUND(SUM(total), 0) AS Total_Sales_Volume
FROM
    walmartsales
GROUP BY customer_id
ORDER BY total_sales_volume DESC
LIMIT 5;
```

OUTPUT

customer_id	Total_Sales_Volume
8	26634
3	23402
2	23392
15	22674
1	22635

Note: Consider **customer_id** as **customer**

Key Insight & Recommendation

Key Insights

- **Customer ID 8** is Walmart's top spender, generating ₹26,634 in revenue.
- The top 5 customers together contribute a **significant portion of total sales**.
- These customers are likely **high-value and loyal**, worth special attention.

Recommendations

- **Reward these top 5 customers** with exclusive offers, early access, or loyalty bonuses.
- Analyze their buying patterns to **replicate success** with other customers.
- Create a **VIP program** to retain and grow relationships with high-value buyers.

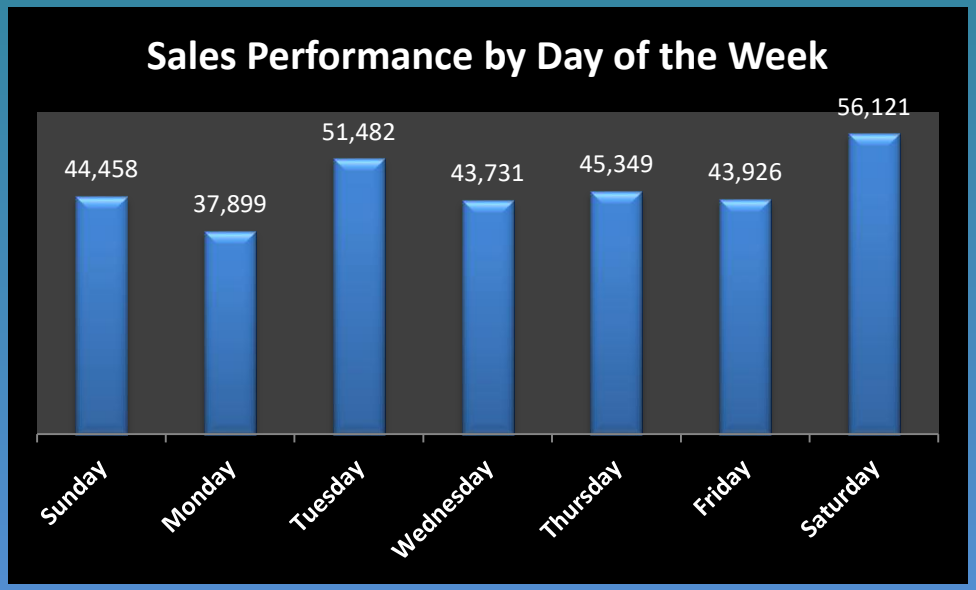
Task 10: Analyzing Sales Trends by Day of the Week

Business Question: Which day of the week generates the highest total sales, helping Walmart optimize staffing and promotions?

```sql
/*
Task 10: Analyzing Sales Trends by Day of the Week
Walmart wants to analyze the sales patterns to determine
which day of the week
brings the highest sales.
*/
SELECT
    DAYNAME(date) AS Days,
    ROUND(SUM(total), 0) AS Sales
FROM
    walmartsales
GROUP BY days , DAYOFWEEK(date)
ORDER BY DAYOFWEEK(date);
```

OUTPUT

Days	Sales
Sunday	44458
Monday	37899
Tuesday	51482
Wednesday	43731
Thursday	45349
Friday	43926
Saturday	56121



Key Insight & Recommendation

Key Insights

- **Saturday** has the highest sales — ₹56,121.
- **Tuesday** also does well — ₹51,482.
- **Monday** has the lowest sales — ₹37,899.
- Other days are in the middle, with steady sales.

Recommendations

- Add more staff and stock on **Saturday** — it's the busiest day.
- Run special offers on **Monday** to boost slow sales.
- Use **Tuesday** for midweek deals — people are already buying.
- Plan your store and marketing based on these day-wise trends.

Video Link

https://www.loom.com/share/a9fc2cb7b0234409ad22ded5324f870b?sid=7d8d6750-f22d-40d9-b53f-c43af99cd8e9

Thank you