WALMART

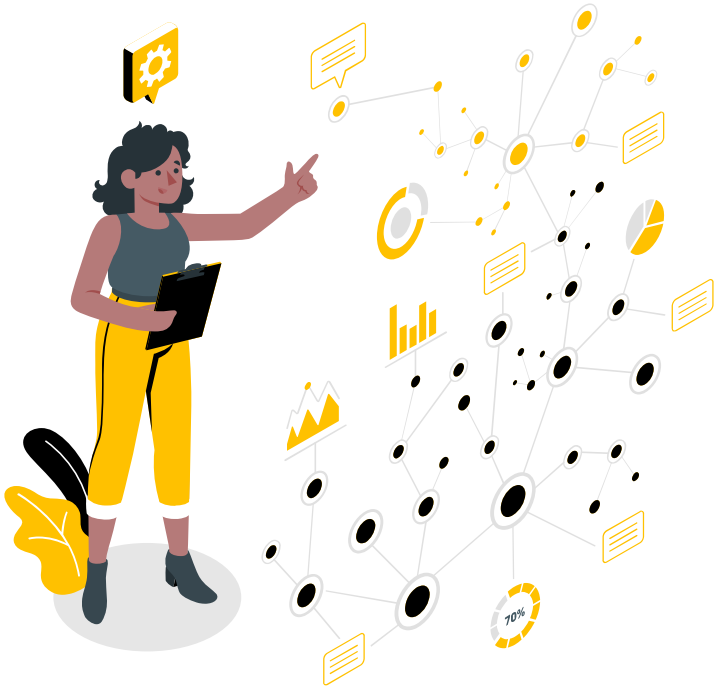
SALES DATA ANALYSIS.

By ***Debonil Saha***
(Dec 01 Batch)

Table of Contents

TASK 01

Identifying the Top Branch by Sales Growth Rate



TASK 01: Walmart wants to identify which branch has exhibited the highest sales growth over time. Analyze the total sales for each branch and compare the growth rate across months to find the top performer.

QUERY:

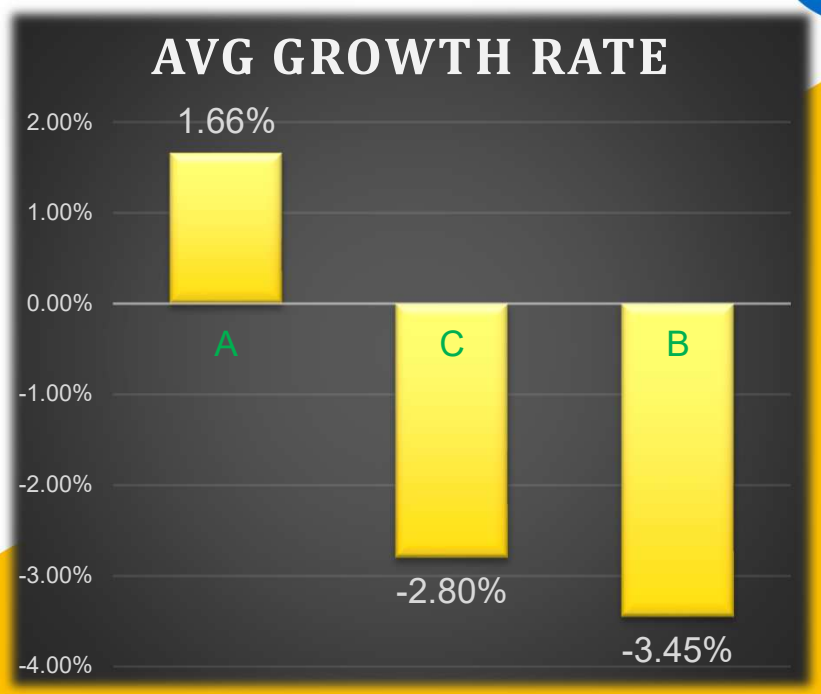
```
WITH MonthlySales AS (
                    SELECT Branch, DATE_FORMAT(Date, '%Y-%m') AS Month,
                    ROUND(SUM(Total),2) AS Sale_by_month
                    FROM walmartsales
                    GROUP BY Branch, Month
                    ORDER BY Month, Branch), prev_sale AS(
                    SELECT Branch, Month, Sale_by_month,
                    LAG(Sale_by_month) OVER (PARTITION BY Branch ORDER BY Month) AS Previous_Sales
FROM monthlysales),

growth_rate AS(
                    SELECT Branch, Month, Sale_by_month, Previous_sales,
                    ROUND(((Sale_by_month - Previous_sales)/ Previous_sales) * 100,2) AS mon_Growth_Rate
FROM prev_sale)

SELECT              Branch,
                    ROUND(AVG(mon_growth_rate),2) as Average_growth_rate
FROM growth_rate
WHERE mon_growth_rate is NOT NULL
GROUP BY Branch
ORDER BY Average_growth_rate DESC
LIMIT 1;
```

RESULT: 01

BRANCH	AVG GROWTH RATE
A	**1.66%**
C	-2.8%
B	-3.45%

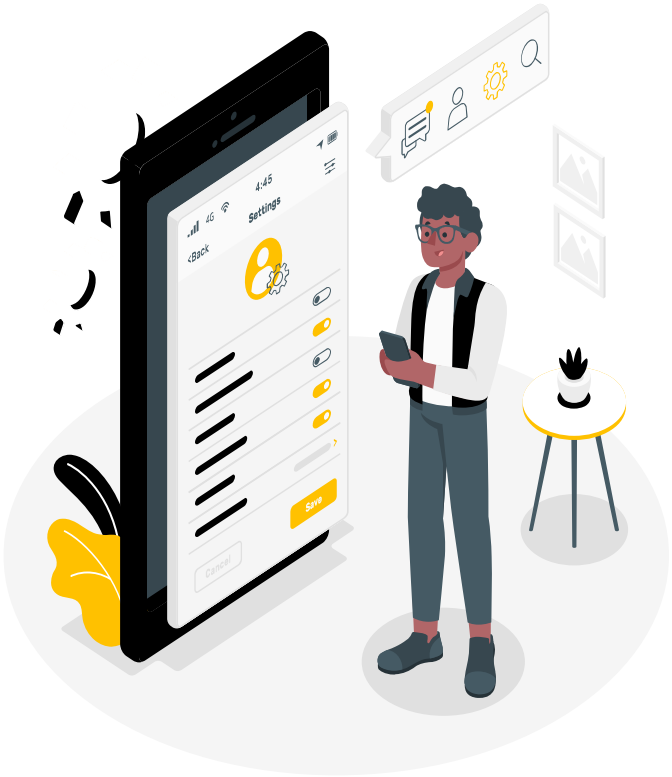


Insights:

*Branch A showed the highest average sales growth at **1.66%**, indicating positive performance. In contrast, Branch B (**-3.45%**) and Branch C **(-2.8%)** experienced declines, with B performing the worst. This suggests potential operational or market challenges for these branches, requiring further investigation and strategic intervention.*

TASK 02

Finding the Most Profitable Product Line for Each Branch



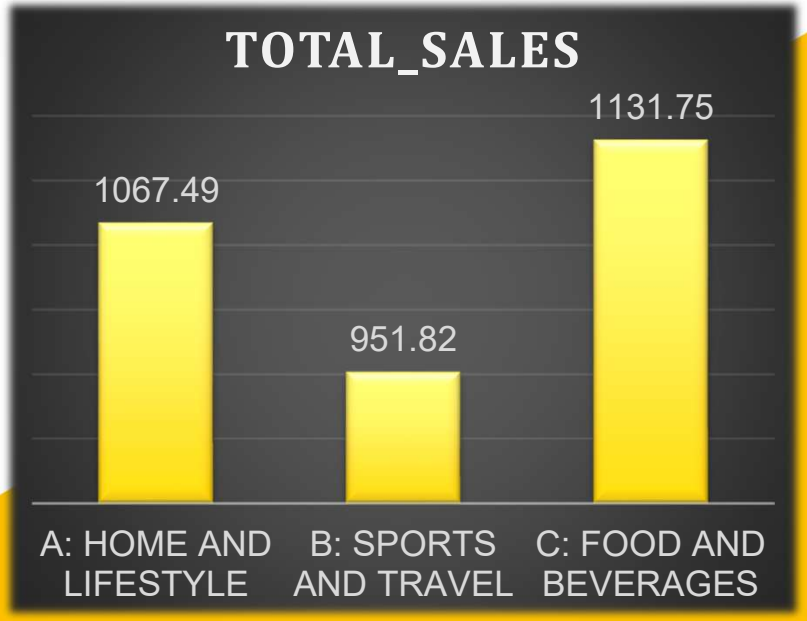
TASK 02: Walmart needs to determine which product line contributes the highest profit to each branch. The profit margin should be calculated based on the difference between the gross income and cost of goods sold.

QUERY:

```
WITH ProductProfit AS (
        SELECT   Branch, product_line,
        round(SUM(gross_income),2) AS Total_Profit
        FROM walmartsales
        GROUP BY Branch, product_line),
RankedProducts AS (
        SELECT   Branch,   product_line,  Total_Profit,
        RANK() OVER (PARTITION BY Branch ORDER BY Total_Profit DESC) AS prod_rank
        FROM ProductProfit)
SELECT Branch, product_line, Total_Profit
FROM RankedProducts
WHERE prod_rank = 1;
```

RESULT: 02

BRANCH	PRODUCT LINE	TOTAL SALES
A	HOME AND LIFE STYLE	$ 1067.49
B	SPORTS AND TRAVEL	$ 951.82
C	FOOD AND BEVERAGES	$ 1131.75

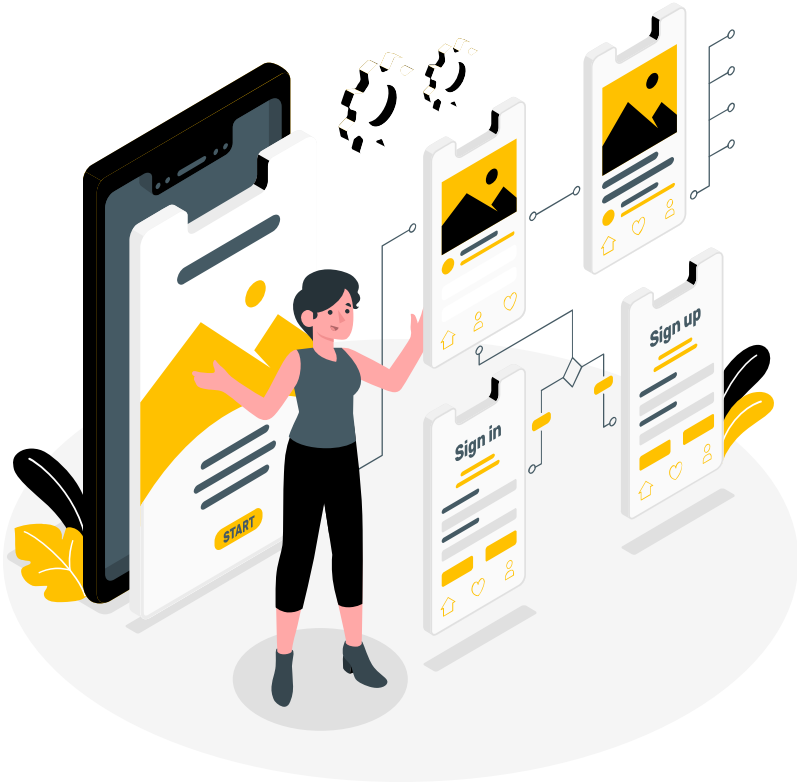


Insights:

Branch C (Food and Beverages) recorded the highest total profit at $1131.75, indicating strong demand or efficient operations in this category. Branch A (Home and Lifestyle) followed with $1067.49, showing solid performance. Branch B (Sports and Travel) had the lowest profit at $951.82, suggesting potential challenges such as lower sales volume or higher costs. Optimizing pricing, promotions, or product mix in Branch B could improve profitability.

TASK 03

Analyzing Customer Segmentation Based on Spending



TASK 03: Walmart wants to segment customers based on their average spending behavior. Classify customers into three tiers: High, Medium, and Low spenders based on their total purchase amounts.

QUERY:

```
WITH spending_cte AS(
        SELECT customer_id, Round(SUM(Total),4) as total_spending
        FROM walmartsales
        GROUP BY customer_id
        ORDER BY total_spending DESC),
Percentile_cte AS (
        SELECT customer_id, total_spending,
        NTILE(3) OVER (ORDER BY Total_Spending DESC) AS Spending_Tier
         FROM spending_cte)SELECT customer_id, total_spending,
        CASE
                WHEN spending_tier =1 THEN "High"
                WHEN spending_tier =2 THEN "Medium"
                ELSE "Low"
        END AS Customer_Class
FROM percentile_cte;
```

RESULT: 03

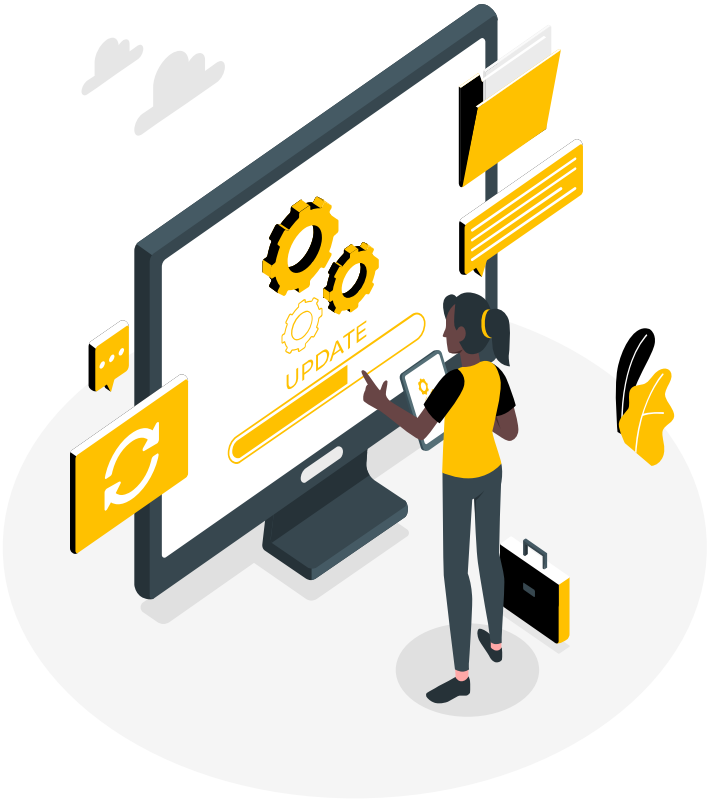
CUSTOMER ID	TOTAL_SPENDING	CUSTOMER_CLASS
1	22634.55	High
2	23392.28	High
3	23402.26	High
4	17656.72	Low
5	19632.04	Low
6	20693.96	Low
7	20628.09	Low
8	26634.34	High
9	19661.6	Low
10	20723.93	Medium
11	21398.82	Medium
12	21720.65	Medium
13	21063.66	Medium
14	21049.4	Medium
15	22674.46	High

Insights:

Customers have been classified into High, Medium, and Low categories based on their total spending. The High-spending customers generally exceed $22,600, indicating strong purchasing power and potential for premium product targeting. Medium-spending customers range around $20,700–$21,700, representing a stable, mid-tier segment. Low-spending customers, spending between $17,600–$20,700, may require targeted promotions or loyalty programs to increase their spending.

TASK 04

Detecting Anomalies in Sales Transactions



TASK 04: Walmart suspects that some transactions have unusually high or low sales compared to the average for the product line. Identify these anomalies.

QUERY:

```
WITH ProductStats AS (
                SELECT product_line,
                AVG(Total) AS Avg_Sales,   STDDEV(Total) AS Std_Dev
                FROM walmartsales
                GROUP BY product_line    ),
Anomalies AS (
                SELECT w.Invoice_ID, w.Branch, w.product_line, w.Total,
                p.Avg_Sales, p.Std_Dev,
                (w.Total - p.Avg_Sales) / p.Std_Dev AS Z_Score
                FROM walmartsales AS w
                JOIN ProductStats AS p
                ON w.product_line = p.product_line    )
SELECT *
FROM Anomalies
WHERE ABS(Z_Score) > 3
ORDER BY Z_Score DESC;
```

RESULT: 04

Invoice_ID	Branch	Product_line	Total	Avg_Sales	Std_Dev	Z_Score
860-79-0874	C	Fashion accessories	1042.65	305.09	242.88	3.04
687-47-8271	A	Fashion accessories	1039.29	305.09	242.88	3.02

Insights:

*The Z-Scores for both transactions (3.04 for Branch C and 3.02 for Branch A**) indicate that these sales are more than 3 standard deviations above the average for the Fashion Accessories product line.*

TASK 05

Most Popular Payment Method by City

TASK 05: Walmart needs to determine the most popular payment method in each city to tailor marketing strategies.

QUERY:

```
WITH payment_frequency_cte AS (
                SELECT City, Payment,
                count(invoice_id) as frequency_method
                FROM walmartsales
                GROUP BY city, Payment),
frequency_rank_cte AS (
                SELECT City, Payment, frequency_method,
                RANK() OVER(PARTITION BY City ORDER BY frequency_method DESC) as pay_rank
                FROM payment_frequency_cte)
SELECT City, Payment, pay_rank
FROM frequency_rank_cte
WHERE pay_rank =1;
```

RESULT:05

City	Payment	Pay_Rank
Mandalay	Ewallet	1
Naypyitaw	Cash	1
Yangon	Ewallet	1

Insights:

The pay_rank shows the top payment in the cities from data. People in Mandalay city pay the most using e-wallet, people in Naypyitaw pay the most using cash mode of payment and people of Yangon city pay the most using e-wallet.

TASK 06

Monthly Sales Distribution by Gender

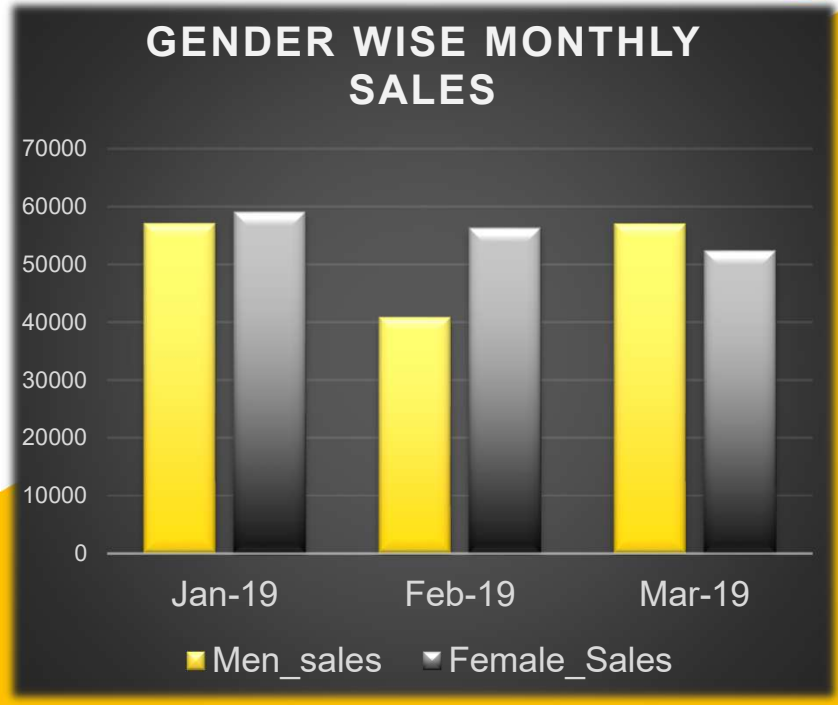
TASK 06: Walmart wants to understand the sales distribution between male and female customers on a monthly basis.

QUERY:

```
With men_sales_cte AS(
                SELECT DATE_FORMAT(Date, '%Y-%m') AS Month,
                ROUND(sum(total),2) AS Men_sales
                FROM Walmartsales
                WHERE gender="Male"
                GROUP BY Gender, Month),
Female_sales_cte AS(
                SELECT DATE_FORMAT(Date, '%Y-%m') AS Month,
                ROUND(sum(total),2) AS female_sales
                FROM Walmartsales
                WHERE gender="Female"
                GROUP BY Gender, Month)
SELECT m.Month, m.Men_sales, f.Female_Sales
FROM men_sales_cte as m
JOIN female_Sales_cte as f
ON m.Month=f.Month
ORDER BY m.Month;
```

RESULT:06

Month	Men_sales	Female_Sales
Jan-19	57152.89	59138.98
Feb-19	40883.82	56335.56
Mar-19	57047.12	52408.39

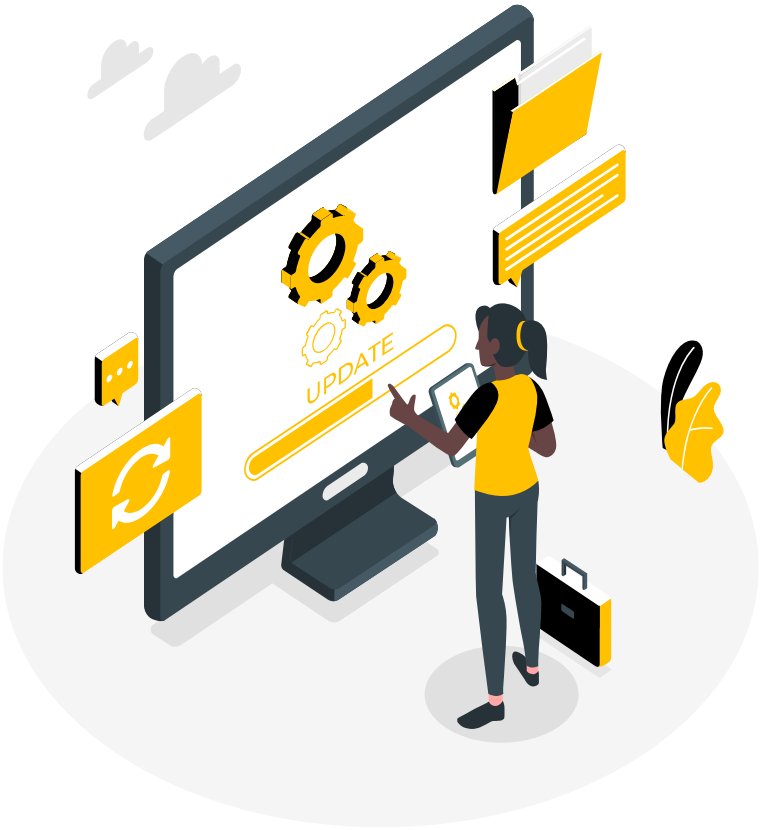


Insights:

The monthly sales distribution indicates that female sales consistently outperformed male sales across all three months. While female sales remained relatively stable, peaking in January at $59,138.98 and slightly declining afterward, male sales showed significant fluctuations. Male sales dropped sharply from $57,152.89 in January to $40,883.82 in February before rebounding to $57,047.12 in March. This suggests a potential seasonal or external factor affecting male purchasing behavior more than females.

TASK 07

Best Product Line by Customer Type



TASK 07: Walmart wants to know which product lines are preferred by different customer types(Membe vs. Normal).

QUERY:

```
WITH sales_cte AS(
                SELECT Customer_type, Product_line,
                ROUND(Sum(total) ,2) as Sales
                FROM walmartsales
                GROUP BY Customer_type, Product_line),
rank_cte AS (
                SELECT Customer_type, Product_line, Sales,
                RANK() OVER (PARTITION BY Customer_type ORDER BY Sales DESC) as Sales_rank
                FROM Sales_cte)
SELECT *
FROM rank_cte
WHERE Sales_rank =1;
```

RESULT: 07

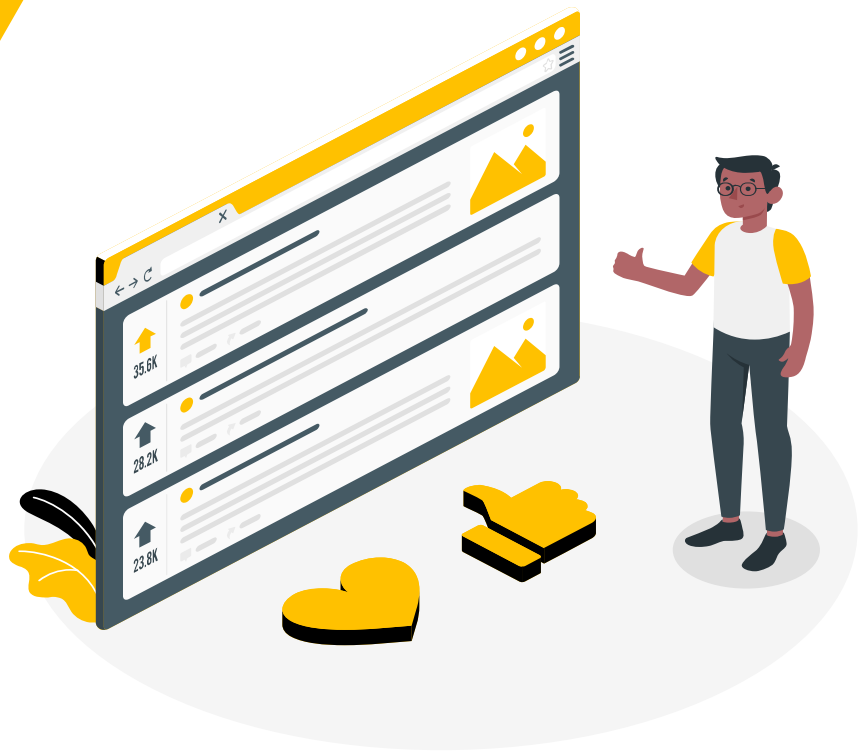
Customer_type	Product_line	Sales	Sales_rank
Member	Food and beverages	31357.62	1
Normal	Electronic accessories	29839.04	1

Insights:

The analysis shows that in the customer type Member the top product line by sales is "food and beverages" and in the normal customer type the top product line is Electronics accessories.

TASK 08

Identifying Repeat Customers



TASK 08: Walmart needs to identify customers who made repeat purchases within a specific time frame (e.g., within 30 days).

QUERY:

```
SELECT    Customer_ID,
                COUNT(Invoice_ID) AS Purchase_Count,
                MIN(date) AS First_Purchase_Date,
                MAX(date) AS Last_Purchase_Date
FROM    walmartsales
WHERE    date BETWEEN DATE_SUB('2019-03-31', INTERVAL 30 DAY) AND '2019-03-31'
GROUP BY    Customer_ID
HAVING    COUNT(Invoice_ID) > 1;
```

RESULT: 08

Customer_ID	Purchase_Count	First_Purchase_Date	Last_Purchase_Date
5	**30**	**02-03-2019**	**30-03-2019**
9	**29**	**01-03-2019**	**29-03-2019**
12	**27**	**01-03-2019**	**30-03-2019**
7	26	02-03-2019	28-03-2019
8	25	01-03-2019	28-03-2019
4	24	02-03-2019	29-03-2019
14	24	01-03-2019	30-03-2019
2	23	02-03-2019	27-03-2019
3	22	02-03-2019	30-03-2019
11	22	01-03-2019	29-03-2019
15	22	03-03-2019	30-03-2019
6	20	02-03-2019	30-03-2019
10	18	01-03-2019	28-03-2019
1	18	02-03-2019	30-03-2019
13	15	02-03-2019	25-03-2019

Insights:

All listed customers made multiple purchases, with purchase counts ranging from 15 to 30. The "First_Purchase_Date" and "Last_Purchase_Date" columns indicate that these customers engaged in repeat purchases within the 30-day window, demonstrating strong retention. High-frequency shoppers, such as Customer_ID 5 (30 purchases) and Customer_ID 9 (29 purchases), suggest potential loyalty or bulk purchasing behavior.

TASK 09

Finding Top 5 Customers by Sales Volume

TASK 09: Walmart wants to reward its top 5 customers who have generated the most sales Revenue.

QUERY:

```
WITH Sales_cte AS (
                SELECT Customer_ID,
                ROUND(Sum(Total),2) AS Sales
                FROM walmartsales
                GROUP BY Customer_ID),
rank_cte AS(
                SELECT Customer_ID, Sales,
                RANK() OVER (ORDER BY Sales DESC) as Sales_rank
                FROM Sales_cte)
SELECT * FROM rank_cte
WHERE Sales_rank<=5;
```

RESULT: 09

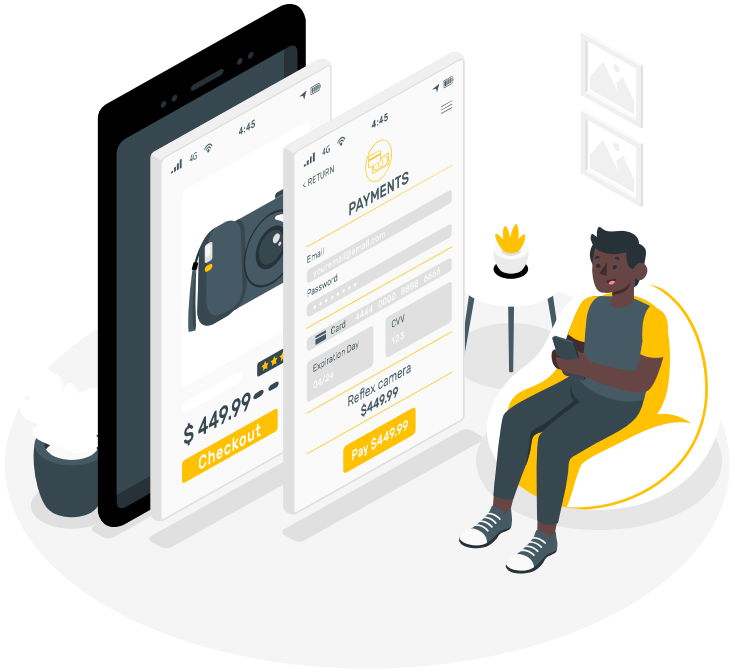
Customer_ID	Sales	Sales_rank
8	26634.34	1
3	23402.26	2
2	23392.28	3
15	22674.46	4
1	22634.55	5

Insights:

The sales ranking data highlights Walmart's top-performing customers based on total sales. Customer_ID 8 leads with the highest sales of $26,634.34, followed by Customer_IDs 3 and 2, with sales just over $23,400. The top five customers all generated sales exceeding $22,600, indicating a strong concentration of revenue among a few high-value customers. Understanding these top spenders enables Walmart to implement personalized marketing strategies, such as exclusive discounts, early access to promotions, or loyalty rewards, to further strengthen customer retention and increase lifetime value.

TASK 10

Analyzing Sales Trends by Day of the Week



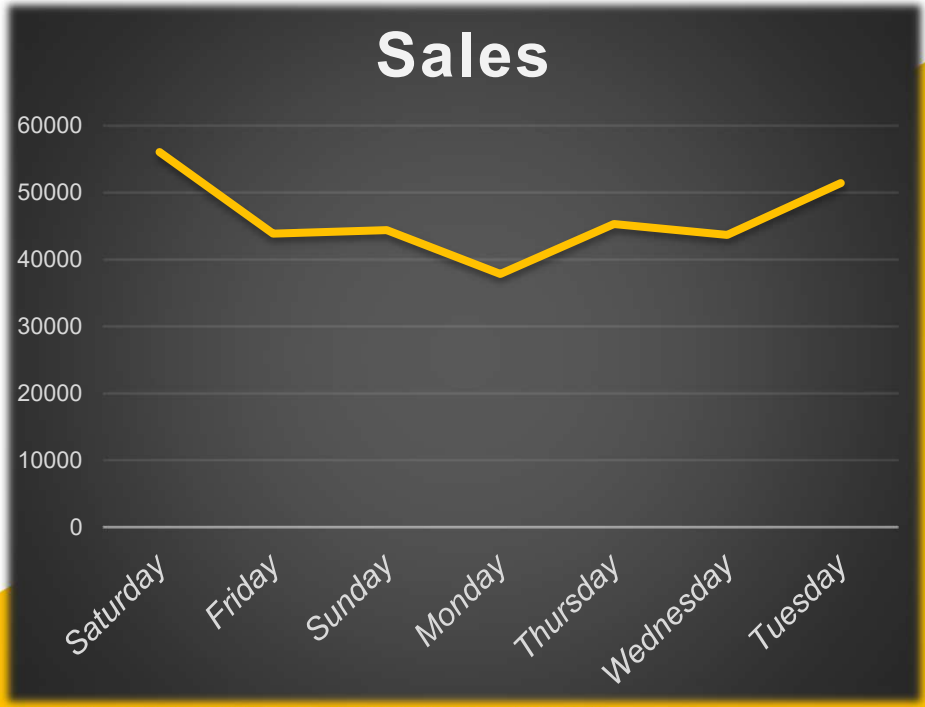
TASK 10: Walmart wants to analyze the sales patterns to determine which day of the week brings the highest sales.

QUERY:

```
SELECT
        ROUND(SUM(total),2) as Sales,
        DAYNAME(Date) AS Weekday
FROM walmartsales
GROUP BY Weekday
ORDER BY Sales DESC
LIMIT 1;        #last line is optional
```

RESULT: 10

Weekday	Sales
Saturday	**56120.81**
Friday	43926.34
Sunday	44457.89
Monday	37899.08
Thursday	45349.25
Wednesday	43731.14
Tuesday	51482.25



Insights:

The sales data by weekday reveals that Saturday generates the highest sales ($56,120.81), followed by Tuesday ($51,482.25), indicating peak shopping days. Friday, Sunday, and Thursday maintain relatively strong sales, while Monday records the lowest sales ($37,899.08). This trend suggests that weekends and early-week shopping are preferred by customers, possibly due to payday cycles, promotions, or weekend stock-ups.

Video Explanation

https://drive.google.com/file/d/14XyhUGgMk4ODvEi_X__rb2rbcq7VQWj_/view?usp=sharing

THANKS FOR WATCHING

BY Debonil Saha
Dec 01 Batch