

Walmart Sales Analysis Using SQL

By: Sinora Rai

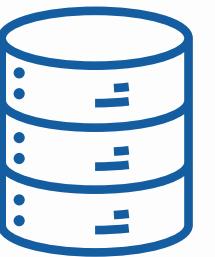




- Founded in 1962 by Sam Walton in Rogers, Arkansas. Rapid expansion across the United States throughout the 1960s and 1970s. Went public in 1970 and continued to grow nationally. First international store opened in Mexico in 1991. Currently operates in 27 countries under various banners, including Walmart and Sam's Club.
- **Mission:** "*Saving people money so they can live better.*" **Core values:** Respect for individuals, service to customers, striving for excellence, and acting with integrity.
- Largest retailer by revenue globally. Diverse offerings including groceries, electronics, clothing, and more. Extensive supply chain network for efficient distribution. Strong presence in both physical and e-commerce retail segments. Continuously adapting to market trends through innovation and technology.

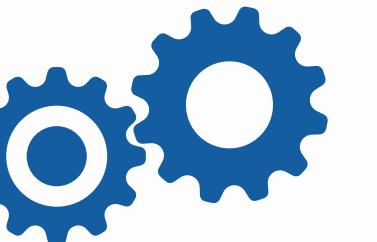
Objective

The major aim of this project is to gain insight into the sales data of Walmart to understand the different factors that affect sales of the different branches.



Step 1: **Data Wrangling**

This is the first step where inspection of data is done to make sure ****NULL**** values and missing values are detected and data replacement methods are used to replace, missing or ****NULL**** values.



Step 2: **Feature Engineering**

Added a new column named `time_of_day` to give insight of sales in the Morning, Afternoon and Evening. This will help answer the question on which part of the day most sales are made.



Step 3: **EDA**

We delve into the data to extract insights essential for addressing the business problem. Through exploration, we identify patterns and trends, enabling informed decision-making and strategic solutions.

Data Overview

This dataset contains sales transactions from a three different branches of Walmart, respectively located in Mandalay, Yangon and Naypyitaw. The data contains 17 columns and 1000 rows

Column	Description	Data Type
-----	-----	-----
invoice_id	Invoice of the sales made	VARCHAR(30)
branch	Branch at which sales were made	VARCHAR(5)
city	The location of the branch	VARCHAR(30)
customer_type	The type of the customer	VARCHAR(30)
gender	Gender of the customer making purchase	VARCHAR(10)
product_line	Product line of the product sold	VARCHAR(100)
unit_price	The price of each product	DECIMAL(10, 2)
quantity	The amount of the product sold	INT
VAT	The amount of tax on the purchase	FLOAT(6, 4)
total	The total cost of the purchase	DECIMAL(10, 2)
date	The date on which the purchase was made	DATE
time	The time at which the purchase was made	TIMESTAMP
payment_method	The total amount paid	DECIMAL(10, 2)
cogs	Cost Of Goods sold	DECIMAL(10, 2)
gross_margin_percentage	Gross margin percentage	FLOAT(11, 9)
gross_income	Gross Income	DECIMAL(10, 2)
rating	Rating	FLOAT(2, 1)

Data Wrangling

Build a data base

```
1    -- Create database  
2 • CREATE DATABASE IF NOT EXISTS walmartSales;
```

Create table and insert the data.



```
-- Create table  
CREATE TABLE IF NOT EXISTS sales(  
    invoice_id VARCHAR(30) NOT NULL PRIMARY KEY,  
    branch VARCHAR(5) NOT NULL,  
    city VARCHAR(30) NOT NULL,  
    customer_type VARCHAR(30) NOT NULL,  
    gender VARCHAR(30) NOT NULL,  
    product_line VARCHAR(100) NOT NULL,  
    unit_price DECIMAL(10,2) NOT NULL,  
    quantity INT NOT NULL,  
    tax_pct FLOAT(6,4) NOT NULL,  
    total DECIMAL(12, 4) NOT NULL,  
    date DATETIME NOT NULL,  
    time TIME NOT NULL,  
    payment VARCHAR(15) NOT NULL,  
    cogs DECIMAL(10,2) NOT NULL,  
    gross_margin_pct FLOAT(11,9),  
    gross_income DECIMAL(12, 4),  
    rating FLOAT(2, 1)  
);
```

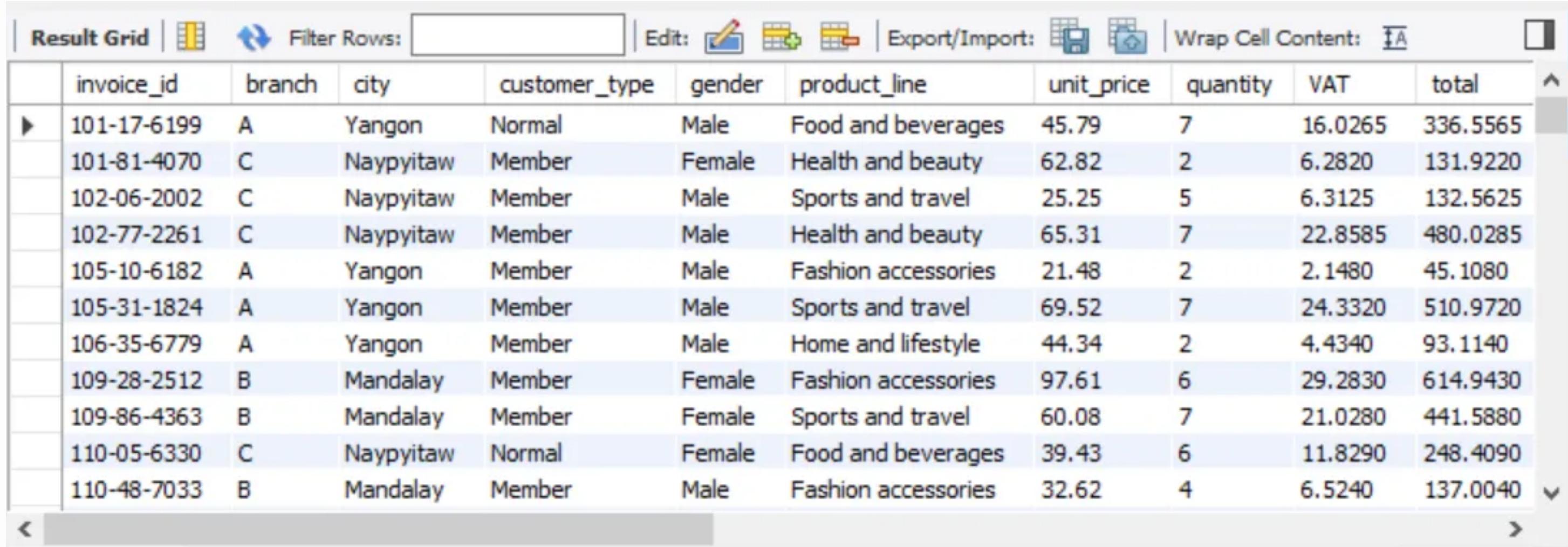
Import Data from CSV



Result Grid	Filter Rows:	Edit:	Export/Import:	Wrap Cell Content:					
invoice_id	branch	city	customer_type	gender	product_line	unit_price	quantity	VAT	total

After creating a database, it's time to import data from a CSV file. For this select import and a new window will open. Select the path from the local machine.

A snippet of the table view



	invoice_id	branch	city	customer_type	gender	product_line	unit_price	quantity	VAT	total
▶	101-17-6199	A	Yangon	Normal	Male	Food and beverages	45.79	7	16.0265	336.5565
	101-81-4070	C	Naypyitaw	Member	Female	Health and beauty	62.82	2	6.2820	131.9220
	102-06-2002	C	Naypyitaw	Member	Male	Sports and travel	25.25	5	6.3125	132.5625
	102-77-2261	C	Naypyitaw	Member	Male	Health and beauty	65.31	7	22.8585	480.0285
	105-10-6182	A	Yangon	Member	Male	Fashion accessories	21.48	2	2.1480	45.1080
	105-31-1824	A	Yangon	Member	Male	Sports and travel	69.52	7	24.3320	510.9720
	106-35-6779	A	Yangon	Member	Male	Home and lifestyle	44.34	2	4.4340	93.1140
	109-28-2512	B	Mandalay	Member	Female	Fashion accessories	97.61	6	29.2830	614.9430
	109-86-4363	B	Mandalay	Member	Female	Sports and travel	60.08	7	21.0280	441.5880
	110-05-6330	C	Naypyitaw	Normal	Female	Food and beverages	39.43	6	11.8290	248.4090
	110-48-7033	B	Mandalay	Member	Male	Fashion accessories	32.62	4	6.5240	137.0040

Feature Engineering

1. Add a new column named `time_of_day` to give insight of sales in the Morning, Afternoon and Evening. This will help answer the question on which part of the day most sales are made.

```
SELECT
    time,
    > (CASE
        WHEN `time` BETWEEN "00:00:00" AND "12:00:00" THEN "Morning"
        WHEN `time` BETWEEN "12:01:00" AND "16:00:00" THEN "Afternoon"
        ELSE "Evening"
    - END) AS time_of_day
FROM sales;
ALTER TABLE sales ADD COLUMN time_of_day VARCHAR(20);
UPDATE sales
> SET time_of_day = (
    > CASE
        WHEN `time` BETWEEN "00:00:00" AND "12:00:00" THEN "Morning"
        WHEN `time` BETWEEN "12:01:00" AND "16:00:00" THEN "Afternoon"
        ELSE "Evening"
    END
- );
```

Feature Engineering

2. Add a new column named `day_name` that contains the extracted days of the week on which the given transaction took place (Mon, Tue, Wed, Thur, Fri). This will help answer the question on which week of the day each branch is busiest.

```
SELECT
```

```
    date,
```

```
    DAYNAME(date)
```

```
FROM sales;
```

```
ALTER TABLE sales ADD COLUMN day_name VARCHAR(10);
```

```
UPDATE sales
```

```
SET day_name = DAYNAME(date);
```

Feature Engineering

3. Add a new column named `month_name` that contains the extracted months of the year on which the given transaction took place (Jan, Feb, Mar). Help determine which month of the year has the most sales and profit.

```
SELECT
```

```
    date,
```

```
    MONTHNAME(date)
```

```
FROM sales;
```

```
ALTER TABLE sales ADD COLUMN month_name VARCHAR(10);
```

```
UPDATE sales
```

```
SET month_name = MONTHNAME(date);
```

Feature Engineering

An updated data set with day_name, month_name and time_of_day

	time	payment_method	cogs	gross_margin_pct	gross_income	rating	time_of_day	day_name	month_name
1	19:44:00	Credit card	320.53	4.761904716	16.0265	7.0	Evening	Wednesday	March
2	12:36:00	Ewallet	125.64	4.761904716	6.2820	4.9	Afternoon	Thursday	January
3	17:52:00	Cash	126.25	4.761904716	6.3125	6.1	Evening	Wednesday	March
4	18:02:00	Credit card	457.17	4.761904716	22.8585	4.2	Evening	Tuesday	March
5	12:22:00	Ewallet	42.96	4.761904716	2.1480	6.6	Afternoon	Wednesday	February
6	15:10:00	Credit card	486.64	4.761904716	24.3320	8.5	Afternoon	Friday	February
7	11:26:00	Cash	88.68	4.761904716	4.4340	5.8	Morning	Wednesday	March
8	15:01:00	Ewallet	585.66	4.761904716	29.2830	9.9	Afternoon	Monday	January
9	11:36:00	Credit card	420.56	4.761904716	21.0280	4.5	Morning	Thursday	February
10	20:18:00	Credit card	236.58	4.761904716	11.8290	9.4	Evening	Monday	March
11	14:12:00	Cash	130.48	4.761904716	6.5240	9.0	Afternoon	Tuesday	January

Exploratory Data Analysis (EDA)

Business Questions To Answer

Generic Question

1. How many unique cities does the data have?
2. In which city is each branch?

-- How many unique cities does the data have?

SELECT

 DISTINCT city

FROM sales;

-- In which city is each branch?

SELECT

 DISTINCT city,

 branch

FROM sales;

Result Grid	
	city
▶	Yangon
	Naypyitaw
	Mandalay

Result Grid	
city	branch
Yangon	A
Naypyitaw	C
Mandalay	B

Exploratory Data Analysis (EDA)

Product Analysis

1. How many unique product lines does the data have?

```
SELECT  
    DISTINCT product_line  
FROM sales;
```



Result Grid		Filter Rows:
count(distinct product_line)		
▶	6	

Exploratory Data Analysis (EDA)

Product Analysis

2. What is the most common payment method?
3. What is the most selling product line?
4. What is the total revenue by month?

payment_method
Credit card
Ewallet
Cash

	customer_type	total_count
▶	Normal	496
▶	Member	499

`SELECT`

```
    SUM(quantity) as qty,  
    product_line
```

`FROM sales`

`GROUP BY product_line`

`ORDER BY qty DESC;`

`SELECT`

```
    SUM(quantity) as qty,  
    product_line
```

`FROM sales`

`GROUP BY product_line`

`ORDER BY qty DESC;`

`SELECT`

```
    month_name AS month,  
    SUM(total) AS total_revenue
```

`FROM sales`

`GROUP BY month_name`

`ORDER BY total_revenue;`

Exploratory Data Analysis (EDA)

Product Analysis

5. What month had the largest COGS?
6. What product line had the largest revenue?
7. What is the city with the largest revenue?
8. What product line had the largest VAT?

```
-- What month had the largest COGS?  
SELECT  
    month_name AS month,  
    SUM(cogs) AS cogs  
FROM sales  
GROUP BY month_name  
ORDER BY cogs;  
  
-- What product line had the largest revenue?  
SELECT  
    product_line,  
    SUM(total) as total_revenue  
FROM sales  
GROUP BY product_line  
ORDER BY total_revenue DESC;
```

```
-- What is the city with the largest revenue?  
SELECT  
    branch,  
    city,  
    SUM(total) AS total_revenue  
FROM sales  
GROUP BY city, branch  
ORDER BY total_revenue;  
  
-- What product line had the largest VAT?  
SELECT  
    product_line,  
    AVG(tax_pct) as avg_tax  
FROM sales  
GROUP BY product_line  
ORDER BY avg_tax DESC;
```

Exploratory Data Analysis (EDA)

Product Analysis

10. Which branch sold more products than average product sold?
11. What is the most common product line by gender?
12. What is the average rating of each product line?

```
SELECT  
    branch,  
    SUM(quantity) AS qnty  
FROM sales  
GROUP BY branch  
HAVING SUM(quantity) > (SELECT AVG(quantity) FROM sales);  
-- What is the most common product line by gender  
  
SELECT  
    gender,  
    product_line,  
    COUNT(gender) AS total_cnt  
FROM sales  
GROUP BY gender, product_line  
ORDER BY total_cnt DESC;
```

```
-- Which branch sold more products than average product sold?  
  
SELECT  
    branch,  
    SUM(quantity) AS qnty  
FROM sales  
GROUP BY branch  
HAVING SUM(quantity) > (SELECT AVG(quantity) FROM sales);
```

Exploratory Data Analysis (EDA)

Sales Analysis

1. Number of sales made in each time of the day per weekday
2. Which of the customer types brings the most revenue?
3. Which city has the largest tax percent/ VAT (**Value Added Tax**)?
4. Which customer type pays the most in VAT?

```
-- Number of sales made in each time of the day per weekday
SELECT
    time_of_day,
    COUNT(*) AS total_sales
FROM sales
WHERE day_name = "Sunday"
GROUP BY time_of_day
ORDER BY total_sales DESC;
```

Result Grid | Filter Rows:

	time_of_day	total_sales
▶	Evening	58
	Afternoon	52
	Morning	22

Exploratory Data Analysis (EDA)

Sales Analysis

-- Which of the customer types brings the most revenue?

```
SELECT
customer_type,
round(SUM(total),2) AS total_revenue
FROM sales
GROUP BY customer_type
ORDER BY total_revenue;
```

Result Grid		Filter Rows:
	customer_type	total_revenue
>	Normal	157261.29
	Member	163625.10

-- Which city has the largest tax/VAT percent?

```
SELECT
city,
ROUND(AVG(tax_pct), 2) AS avg_tax_pct
FROM sales
GROUP BY city
ORDER BY avg_tax_pct DESC;
```

Result Grid		Filter Rows:
	city	value_added_tax
>	Naypyitaw	16.09010850
	Mandalay	15.13020824
	Yangon	14.87020798

Exploratory Data Analysis (EDA)

Sales Analysis

```
-- Which customer type pays the most in VAT?
```

```
SELECT  
    customer_type,  
    AVG(tax_pct) AS total_tax  
FROM sales  
GROUP BY customer_type  
ORDER BY total_tax;
```

Result Grid | Filter Rows:

	customer_type	value_added_tax
▶	Member	15.61457214
	Normal	15.09805040

Exploratory Data Analysis (EDA)

Customer Analysis

```
-- How many unique customer types does the data have?
```

```
SELECT  
    DISTINCT customer_type  
FROM sales;
```

Result Grid		Filter Rows:
	customer_type	
▶	Normal	
	Member	

```
-- How many unique payment methods does the data have?
```

```
SELECT  
    DISTINCT payment  
FROM sales;
```

Result Grid		Filter
	payment	
▶	Credit card	
	Ewallet	
	Cash	

Exploratory Data Analysis (EDA)

Customer Analysis

-- What is the most common customer type?

```
SELECT
    customer_type,
    count(*) as count
FROM sales
GROUP BY customer_type
ORDER BY count DESC;
```

-- What is the gender of most of the customers?

```
SELECT
    gender,
    COUNT(*) as gender_cnt
FROM sales
GROUP BY gender
ORDER BY gender_cnt DESC;
```

Result Grid		Filter R
	customer_type	count
▶	Member	499
	Normal	496

Result Grid		Filter F
	gender	gender_cnt
▶	Male	498
	Female	497

Exploratory Data Analysis (EDA)

Customer Analysis

-- What is the gender distribution per branch?

```
SELECT
    gender,
    COUNT(*) as gender_cnt
FROM sales
WHERE branch = "C"
GROUP BY gender
ORDER BY gender_cnt DESC;
```

-- Which time of the day do customers give most ratings?

```
SELECT
    time_of_day,
    AVG(rating) AS avg_rating
FROM sales
GROUP BY time_of_day
ORDER BY avg_rating DESC;
```

gender	gender_cnt
Female	177
Male	150

Gender per branch is more or less the same hence,
I don't think has an effect of the sales per branch
and other factors.

	time_of_day	avg_rating
▶	Afternoon	7.02340
	Morning	6.94474
	Evening	6.90536

Exploratory Data Analysis (EDA)

Customer Analysis

Which time of the day do customers give most ratings per branch?

```
SELECT  
    time_of_day,  
    AVG(rating) AS avg_rating  
FROM sales  
WHERE branch = "A"  
GROUP BY time_of_day  
ORDER BY avg_rating DESC;
```

```
SELECT  
    time_of_day,  
    AVG(rating) AS avg_rating  
FROM sales  
WHERE branch = "B"  
GROUP BY time_of_day  
ORDER BY avg_rating DESC;
```

```
SELECT  
    time_of_day,  
    AVG(rating) AS avg_rating  
FROM sales  
WHERE branch = "C"  
GROUP BY time_of_day  
ORDER BY avg_rating DESC;
```

	time_of_day	avg_rating
▶	Afternoon	7.18889
	Morning	7.00548
	Evening	6.87143

	time_of_day	avg_rating
▶	Morning	6.83793
	Afternoon	6.81129
	Evening	6.75102

	time_of_day	avg_rating
▶	Evening	7.09859
	Afternoon	7.06667
	Morning	6.97458

Branch A and C are doing well in ratings, branch B needs to do a little more to get better ratings.

Exploratory Data Analysis (EDA)

Customer Analysis

Which time of the day do customers give most ratings per branch?

```
SELECT  
    time_of_day,  
    AVG(rating) AS avg_rating  
FROM sales  
WHERE branch = "A"  
GROUP BY time_of_day  
ORDER BY avg_rating DESC;
```

	time_of_day	avg_rating
▶	Afternoon	7.18889
	Morning	7.00548
	Evening	6.87143

```
SELECT  
    time_of_day,  
    AVG(rating) AS avg_rating  
FROM sales  
WHERE branch = "B"  
GROUP BY time_of_day  
ORDER BY avg_rating DESC;
```

	time_of_day	avg_rating
▶	Morning	6.83793
	Afternoon	6.81129
	Evening	6.75102

```
SELECT  
    time_of_day,  
    AVG(rating) AS avg_rating  
FROM sales  
WHERE branch = "C"  
GROUP BY time_of_day  
ORDER BY avg_rating DESC;
```

	time_of_day	avg_rating
▶	Evening	7.09859
	Afternoon	7.06667
	Morning	6.97458

Branch A and C are doing well in ratings, branch B needs to do a little more to get better ratings.

Conclusion

In our Walmart sales data analysis project, we cleaned the dataset, engineered features, and used SQL queries to extract insights. We explored product performance, sales trends, and customer behavior. Key metrics like COGS, VAT, revenue, and gross profit were computed to inform Walmart's sales strategies and future optimizations.

Keypoints:

- **Identification of top-performing product lines and branches.**
- **Analysis of sales trends, which can inform sales strategies and modifications.**
- **Profiling of customer segments and their profitability.**



THANK YOU!



Sinora Rai

Data Analytics Enthusiast



+123-456-7890



www.sweetwhisper07@gmail.com



Kolkata, West Bengal

