

 **Linked In** **GitHub**

Special Offer



Walmart Data Analysis



By Alipriya Ghosh



MY SQL PROJECT

alipriyaghosh17@gmail.com





FUNCTIONS USED

- **Select**
- **From**
- **Where Clause**
- **Group By**
- **Order by – (Desc/Asc)**
- **LIMIT**
- **Window Functions –Row number**
- **CTE (WITH clause)**
- **Subquery as a temporary table**
- **CTE +Conditional function : CASE**
- **Date & Time Functions like**
Month, Date, Format , Dayname, weekday, hour
- **Arithmetic Functions:**
Count , Sum , Max ,Avg , Min

About

This project aims to explore the Walmart Sales data to understand top performing branches and products, sales trend of different products, customer behaviour. The aim is to study how sales strategies can be improved and optimized. The dataset was obtained from the Kaggle Walmart Sales Forecasting Competition.

"In this recruiting competition, job-seekers are provided with historical sales data for 45 Walmart stores located in different regions. Each store contains many departments, and participants must project the sales for each department in each store. To add to the challenge, selected holiday markdown events are included in the dataset. These markdowns are known to affect sales, but it is challenging to predict which departments are affected and the extent of the impact."

Purposes Of The Project

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.

Analysis List

Product Analysis

Conduct analysis on the data to understand the different product lines, the product lines performing best and the product lines that need to be improved.

Sales Analysis

This analysis aims to answer the question of the sales trends of product. The result of this can help use measure the effectiveness of each sales strategy the business applies and what modifications are needed to gain more sales.

Customer Analysis

This analysis aims to uncover the different customer segments, purchase trends and the profitability of each customer segment.



1. How many unique cities does the data have?

Select Count(distinct city) as number_city
from walmart;

	number_city
▶	3

2. In which city is each branch?

Select City , Branch
from Walmart
group by city , branch ;

City	Branch
Yangon	A
Naypyitaw	C
Mandalay	B

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

Select Count(distinct city) as number_city
from walmart;

Count(distinct Product_line)
6

4. What is the most common payment method

Select Payment , count(Payment) as number_times
from walmart
group by Payment
order by number_times desc
limit 1;

Payment	number_times
Ewallet	345





5. What is the most selling product line?

```
Select product_line , sum(Quantity) as most_sold
from walmart
GROUP BY product_line
ORDER BY most_sold DESC
limit 1;
```

product_line	most_sold
Electronic accessories	971

6. What is the total revenue by month?

```
Select Month(date) as Month_number, date_format(date,' % M') as
month_name, round(sum(total),1) as total_revenue
from walmart
group by Month_number , month_name;
```

Month_number	month_name	total_revenue
1	January	116291.9
3	March	109455.5
2	February	97219.4

7. What month had the largest COGS?

```
Select Month(date) as Month_number, date_format(date,'%M') as
month_name, round(sum(cogs),1) as total_cogs
from walmart
group by Month_number , month_name
order by total_cogs desc
limit 1;
```

Month_number	month_name	total_cogs
1	January	110754.2



8. What product line had the largest revenue?

```
Select product_line ,round(sum(total),2) as Revenue
from walmart
group by product_line
order by Revenue desc;
```

product_line	Revenue
Food and beverages	56144.84
Sports and travel	55122.83
Electronic accessories	54337.53
Fashion accessories	54305.9
Home and lifestyle	53861.91
Health and beauty	49193.74

9. What is the city with the largest revenue?

```
Select city ,round(sum(total),2) as Revenue
from walmart
group by city
order by Revenue desc;
```

city	Revenue
Naypyitaw	110568.71
Yangon	106200.37
Mandalay	106197.67

10.What product line had the largest VAT?

```
SELECT product_line, avg(Vat) AS Max_VAT
FROM walmart
GROUP BY product_line
ORDER BY Max_VAT DESC
LIMIT 1;
```

city	Revenue
Naypyitaw	110568.71
Yangon	106200.37
Mandalay	106197.67





11. Fetch each product line and add a column to those product, line showing "Good", "Bad". Good if its greater than average sales

```
With total_sales as
(Select product_line, sum(total) as total_s
from walmart
group by Product_line) ,
Average as ( select Avg(total_s ) as averge_sales from total_sales)
Select product_line, total_s ,
Case When total_s> averge_sales Then "Good Product"
Else "Bad product"
end as performance
from total_sales, Average;
```

product_line	total_s	performance
Health and beauty	49193.739000000016	Bad product
Electronic accessories	54337.531500000005	Good Product
Home and lifestyle	53861.91300000001	Good Product
Sports and travel	55122.826499999996	Good Product
Food and beverages	56144.844000000005	Good Product
Fashion accessories	54305.895	Good Product

12 What is the most common product line by gender

```
Select product_line , count(gender) as no_of_productline
from walmart
group by product_line
order by no_of_productline desc;
```

product_line	no_of_productline
Fashion accessories	178
Food and beverages	174
Electronic accessories	170
Sports and travel	166
Home and lifestyle	160
Health and beauty	152



13.Which branch sold more products than average product sold?

```
WITH branch_totals AS (  
    SELECT Branch, SUM(Quantity) AS total_products  
    FROM walmart  
    GROUP BY Branch),  
average_total AS (  
    SELECT AVG(total_products) AS avg_products  
    FROM branch_totals)  
SELECT b.Branch, b.total_products  
FROM branch_totals b, average_total a  
WHERE b.total_products > a.avg_products;
```

Branch	total_products
A	1859

14. What is the average rating of each product line

```
select product_line, avg(rating) as rate  
from walmart  
group by Product_line  
order by rate desc;
```

product_line	rate
Food and beverages	7.113218390804598
Fashion accessories	7.029213483146067
Health and beauty	7.003289473684212
Electronic accessories	6.92470588235294
Sports and travel	6.916265060240964
Home and lifestyle	6.8375



15.How many unique customer types does the data have?

Select count(distinct Customer_type) as unique_type
from walmart ;

unique_type
2

16. How many unique payment methods does the data have?

Select count(distinct Payment) as num_payment_types
from walmart ;

num_payment_types
3

17. What is the most common customer type?

Select Customer_type , count(*) as count
from walmart
group by Customer_type
ORDER BY count DESC;

Customer_type	count
Member	501
Normal	499

18. What is the gender of most of the customers

Select Gender , count(*) as count
from walmart
group by Gender
ORDER BY count DESC

Gender	count
Female	501
Male	499



19. What is the gender distribution per branch?

```
Select Gender , Branch ,count(*) as count
from walmart
group by Gender, branch
ORDER BY Gender;
```

Gender	Branch	count
Female	A	161
Female	B	162
Female	C	178
Male	A	179
Male	B	170
Male	C	150

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

```
Select time , Max_rating
from
(Select Hour(time) as time , count(rating) as max_rating , row_number() over
(ORDER BY COUNT(rating) DESC)as row_num
from walmart
group by time) as t
where row_num=1;
```

time	Max_rating
19	7

21. Which day fo the week has the best avg ratings?

```
SELECT
    DAYNAME(date) AS day_of_week,
    AVG(rating) AS avg_rating
FROM walmart
GROUP BY DAYNAME(date)
ORDER BY avg_rating DESC;
```

day_of_week	avg_rating
Monday	7.153599999999999
Friday	7.076258992805756
Sunday	7.011278195488723
Tuesday	7.003164556962025
Saturday	6.901829268292688
Thursday	6.88985507246377
Wednesday	6.805594405594405



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

```
WITH rating AS (  
    SELECT HOUR(time) AS time_of_the_day, branch,  
    COUNT(rating) AS most_rating FROM walmart  
    GROUP BY time_of_the_day, branch),  
ranking AS (SELECT rating.*,  
    ROW_NUMBER() OVER (PARTITION BY branch ORDER BY most_rating  
    DESC) AS row_num FROM rating)  
SELECT time_of_the_day, branch, most_rating  
FROM ranking  
WHERE row_num = 1;
```

time_of_the_day	branch	most_rating
10	A	38
19	B	50
10	C	37

23 Which day of the week has the best average ratings per branch?

```
WITH branch_ratings AS ( SELECT DAYNAME(date) AS day_of_week,  
    branch, AVG(rating) AS avg_rating FROM walmart  
    GROUP BY DAYOFWEEK(date), DAYNAME(date), branch)  
SELECT day_of_week, branch, avg_rating  
FROM (SELECT br.*, ROW_NUMBER() OVER (PARTITION BY branch ORDER  
BY avg_rating DESC) AS rn FROM branch_ratings b) ranked  
WHERE rn = 1;
```

day_of_week	branch	avg_rating
Friday	A	7.3119999999999985
Monday	B	7.335897435897434
Friday	C	7.278947368421051



24 Number of sales made in each time of the day per weekday

```
Select  weekday(date) as week_number,  
        dayname(date) as week_name ,  
        Case when time between 4 and 11 then 'Morning'  
        when time between 12 and 16 then "Afternoon"  
        when time between 17 and 20 then "Evening"  
        Else "Night"  
        End as time_of_the_day,  
count(*) as no_of_sales  
from walmart  
group by week_number,week_name ,time_of_the_day  
order by week_number;
```

week_number	week_name	time_of_the_day	no_of_sales
0	Monday	Afternoon	64
0	Monday	Evening	40
0	Monday	Morning	21
1	Tuesday	Afternoon	62
1	Tuesday	Evening	60
1	Tuesday	Morning	36
2	Wednesday	Afternoon	71
2	Wednesday	Evening	50
2	Wednesday	Morning	22
3	Thursday	Afternoon	61
3	Thursday	Evening	44
3	Thursday	Morning	33
4	Friday	Afternoon	68
4	Friday	Evening	42
4	Friday	Morning	29
5	Saturday	Afternoon	69





25.Which of the customer types brings the most revenue?

```
Select customer_type , sum(total) as total_rev
from Walmart
group by customer_type
order by total_rev desc;
```

customer_type	total_rev
Member	164223.44400000002
Normal	158743.30500000005

26. Which city has the largest tax percent/ VAT (Value Added Tax)?

```
Select City , Max(Vat)
from walmart
group by city;
```

City	Max(Vat)
Yangon	49.49
Naypyitaw	49.65
Mandalay	48.69

27. Which customer type pays the most in VAT?

```
Select customer_type , round(sum(Vat),2) as most_of_vat
from Walmart
group by customer_type
order by most_of_vat desc;
```

customer_type	most_of_vat
Member	7820.16
Normal	7559.21