Walmart sales detailed project analysis report

Sql Queries:

```
create database siva_walmart_salesdata;

use siva_walmart_salesdata;

show tables;

select * from walmart_sales;
```

Invoice_id	Branch	City	customer_type	Gender	Product_line	Unit_price	Quantity	VAT	Total	date	time	Payment_method	Cogs
750-67-8428	A	Yangon	Member	Female	Health and beauty	74.69	7	26.1415	548.9715	2005-01-19	13:08:00	Ewallet	522.83
226-31-3081	C	Naypyitaw	Normal	Female	Electronic accessories	15.28	5	3.82	80.22	2008-03-19	10:29:00	Cash	76.4
631-41-3108	A	Yangon	Normal	Male	Home and lifestyle	46.33	7	16.2155	340.5255	2003-03-19	13:23:00	Credit card	324.31
123-19-1176	A	Yangon	Member	Male	Health and beauty	58.22	8	23.288	489.048	2027-01-19	20:33:00	Ewallet	465.76
373-73-7910	A	Yangon	Normal	Male	Sports and travel	86.31	7	30.2085	634.3785	2008-02-19	10:37:00	Ewallet	604.17
699-14-3026	C	Naypyitaw	Normal	Male	Electronic accessories	85.39	7	29.8865	627.6165	2025-03-19	18:30:00	Ewallet	597.73
355-53-5943	A	Yangon	Member	Female	Electronic accessories	68.84	6	20.652	433.692	2025-02-19	14:36:00	Ewallet	413.04
315-22-5665	C	Naypyitaw	Normal	Female	Home and lifestyle	73.56	10	36.78	772.38	2024-02-19	11:38:00	Ewallet	735.6
665-32-9167	A	Yangon	Member	Female	Health and beauty	36.26	2	3.626	76.146	2010-01-19	17:15:00	Credit card	72.52

general analysis:

-- How many unique cities does the city have?

```
select  COUNT( distinct City)
from walmart_sales;
```

COUNT(distinct City)
3

3 unique cities does the city column have.

-- in which city is each branch?

```
select branch,city
from walmart_sales
order by branch;
```

branch	city
A	Yangon
B	Mandalay
C	Naypyitaw

 # product related questions:

-- how many unique product lines does the data have?

```
select Product_line,count(distinct Product_line)
from walmart_sales
group by Product_line;
```

	count(distinct Product_line)
▶	6

-- what is the most common payment method?

```
with data as(select Payment_method, count( Payment_method) count
from walmart_sales
group by payment_method)
select *
from data
where count=(select max(count) from data);
```

	payment_method
▶	Ewallet

Ewallet is the most common payment method.

-- what is the most selling product line?

```
with data as(select Product_line,count( Product_line) count
from walmart_sales
group by Product_line
order by count desc)
select *
from data
where count=(select max(count) from data);
```

	product_line
▶	Fashion accessories

Fashion accessories is the most common product line.

-- what is the total revenue by month?

```
select Month_name,SUm(Total) as Total_revenue
from walmart_sales
GROUP BY Month_name;
```

	Month_name	Total_revenue
▶	January	116291.86800000005
	March	109455.50700000004
	February	97219.37399999997

-- what month had the largest cogs?

```
with data as ( select Month_name,sum(cogs) cogs
from walmart_sales
group by Month_name
order by sum(cogs) desc)
select *
from data
where cogs=(select max(cogs) from data);
```

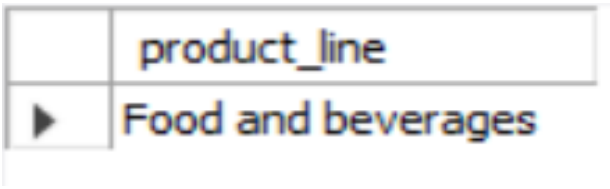
	Month_name	cogs
▶	January	110754.16000000002

january month had the largest cogs

-- what product line had the largest revenue?

```
with data as(select Product_line,sum(Total) revenue
```

```
from walmart_sales
group by product_line
order by sum(Total) desc)
select * from data
where revenue=(select max(revenue) from data);
```

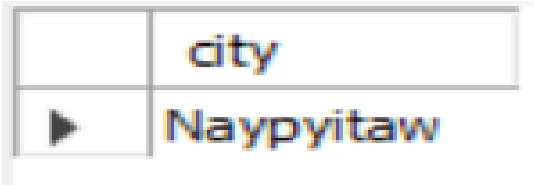


	product_line
▶	Food and beverages

food and beverages had the largest revenue.

-- what is the city with largest revenue?

```
with data as(select City,sum(Total) revenue
from walmart_sales
group by City
order by sum(Total) desc)
select * from data
where revenue=(select max(revenue) from data);
```



	city
▶	Naypyitaw

naypyitaw city had the largest revenue.

-- what product line had the largest VAT?

```
with data as(select Product_line,sum(VAT) vat
from walmart_sales
group by product_line
order by sum(VAT) desc)
```

```
select * from data
where vat=(select max(vat) from data);
```

product_line
Food and beverages

Food and beverages had the largest vat.

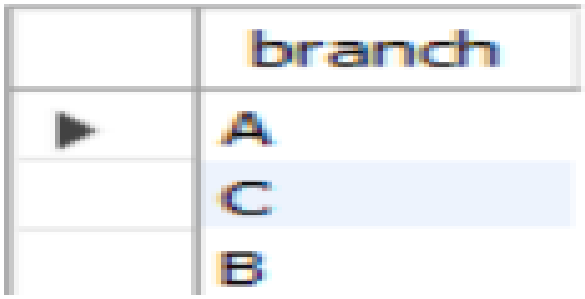
-- Fetch each product line and add a column to those product line showing "Good""Bad" Good if its greaterthan average sales?

```
SELECT
product_line,
quantity,
CASE
   WHEN quantity >= 5.5100 THEN 'Good'
   ELSE 'Bad'
END AS sales_evaluation
from walmart_sales;
```

product_line	quantity	sales_evaluation
Health and beauty	7	Good
Electronic accessories	5	Bad
Home and lifestyle	7	Good
Health and beauty	8	Good
Sports and travel	7	Good
Electronic accessories	7	Good
Electronic accessories	6	Good
Home and lifestyle	10	Good
Health and beauty	2	Bad
Food and beverages	3	Bad

-- Which branch sold more products than average product sold?

```
select branch
from walmart_sales
where quantity>(select avg(quantity) from walmart_sales);
```



	branch
▶	A
	C
	B

-- What is the most common product line by gender?

```
with femaledata as(select product_line, gender,count(gender) count
from walmart_sales
where gender='female'
group by product_line,gender
order by count desc),
maledata as(select product_line, gender,count(gender) count
from walmart_sales
where gender='male'
group by product_line,gender
order by count desc),
femalerank as (select *,dense_rank() over(order by count desc) as dr from femaledata),
malerank as(select *,dense_rank() over(order by count desc) as dr from maledata),
```

```
data as(select * from femalerank
union all
select * from malerank)
select * from data
where dr=1;
```

	product_line	gender	count	dr
▶	Fashion accessories	Female	96	1
	Health and beauty	Male	88	1

Fashion accessories is the common_product line by female

Health and beauty is the common_product line by male

##----------(or)----------##

```
with data as(select gender,product_line,count(gender) count
from walmart_sales
group by gender,product_line
order by count desc),
rankdata as(select *,dense_rank() over(partition by gender order by count
desc) as rnk from data)
select * from rankdata
where rnk =1;
```

-- What is the average rating of each product line?

```
select product_line,avg(rating)
```

```
from walmart_sales
group by product_line;
```

product_line	avg(rating)
Health and beauty	7.003289473684212
Electronic accessories	6.92470588235294
Home and lifestyle	6.8375
Sports and travel	6.916265060240964
Food and beverages	7.113218390804598
Fashion accessories	7.029213483146067

```
# (or)
select product_line,avg(rating) over( partition by product_line) avg_rating
from walmart_sales;
```

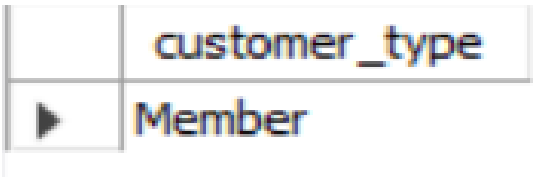
sales analysis:

```
-- Number of sales made in each time of the day per weekday?
select day_name,time_of_day,count(*) as sales_count
from walmart_sales
group by day_name,time_of_day
order by
field(day_name, "Monday", "Tuesday", "wednesday", "Thursday", "Friday", "Saturday", "Sunday");
```

day_name	time_of_day	sales_count
Monday	morning	23
Monday	afternoon	64
Monday	evening	45
Tuesday	evening	55
Tuesday	morning	23
Tuesday	afternoon	72
Wednesday	afternoon	71
Wednesday	morning	27
Wednesday	evening	43
Thursday	afternoon	62

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

```
with data as(select customer_type,sum(total) total
from walmart_sales
group by customer_type)
select * from data
where total= (select max(total) from data);
```

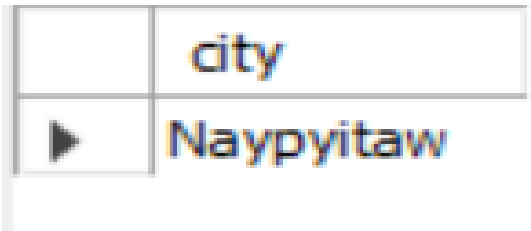


	customer_type
▶	Member

member customer type brings the most revenue.

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

```
with data as(select City,sum(vat) vat
from walmart_sales
group by city)
select * from data
where vat=(select max(vat) from data) ;
```



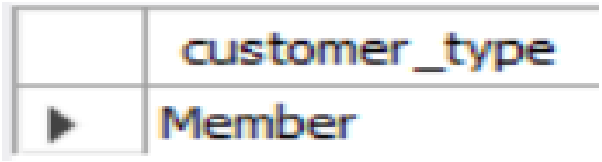
	city
▶	Naypyitaw

naypyitaw city has the largest tax percent.

-- Which customer type pays the most in VAT?

```
with data as(select customer_type,sum(vat) vat
from walmart_sales
group by customer_type)
select * from data
```

where vat=(select max(vat) from data);



	customer_type
▶	Member

member customer type pays the most vat.

customer related analysis:

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

select count(distinct customer_type)

from walmart_sales;

	count
▶	2

2 unique customers does the data have.

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

select count(distinct payment_method)

from walmart_sales;

3 unique payment methods does the data have.

-- What is the most common customer type?

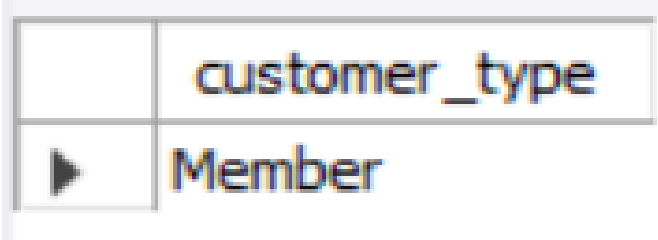
with data as(select customer_type, count(customer_type) count

from walmart_sales

group by(customer_type))

select * from data

where count=(select max(count) from data);



	customer_type
▶	Member

member is the most customer type.

-- **Which customer type buys the most?**

```
with data as(select customer_type,sum(quantity) quantity
from walmart_sales
group by (customer_type))
select * from data
where quantity=(select max(quantity) from data);
```

	customer_type
▶	Member

member customer type buys the most.

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

```
with data as(select gender,count(gender) count
from walmart_sales
group by gender)
select *
from data
where count=(select max(count) from data);
```

	gender
▶	Female

most of the customers gender is female.

-- **What is the gender distribution per branch?**

```
select branch,gender, count(gender) as frequency
from walmart_sales
group by branch,gender
```

```
order by branch,frequency desc;
```

	branch	gender	frequency
▶	A	Male	179
	A	Female	161
	B	Male	170
	B	Female	162
	C	Female	178
	C	Male	150

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

```
with data as(select Time_of_day,count(rating) rating
from walmart_sales
group by Time_of_day)
select * from data
where rating=(select max(rating) from data);
```

	time_of_day
▶	afternoon

afternoon time customers give most ratings.

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

```
with data as(select time_of_day,branch,count(rating) rating
from walmart_sales
group by time_of_day,branch
order by branch,count(rating) desc),
        rankdata as(select *,dense_rank() over(partition by branch order by
rating desc) as dr from data)
select *
from rankdata
```

```
where dr=1;
```

	time_of_day	branch	rating	dr
▶	afternoon	A	152	1
	afternoon	B	137	1
	afternoon	C	150	1

```
-- Which day of the week has the best avg ratings?
with data as(select day_name, avg(Rating) rating
from walmart_sales
group by day_name)
select *
from data
where rating=(select max(rating) from data);
```

	day_name
▶	Wednesday

wednesday has the best avg rating.

```
-- Which day of the week has the best average ratings per branch?
with data as( select branch,day_name,avg(rating) rating
from walmart_sales
group by branch,day_name
order by avg(rating) desc),
rankdata as( select *,dense_rank() over(partition by branch order by rating desc) as rnk from data)
select * from rankdata where rnk=1;
```

branch	day_name	rating	rnk
A	Tuesday	7.234042553191487	1
B	Friday	7.226190476190476	1
C	Sunday	7.480434782608695	1

Possible insights from this data:

- There are 3 unique cities in this data those are “yangon”, “mandalay”, “Naypyitaw”.
- This data consists of 6 unique product_lines.
- E-wallet is the most common payment method used by everyone.
- Fashion accessories is the most selling product in this data.
- Food and beverages created the largest revenue in this data.
- Naypyitaw city collected largest tax.
- Member customer type payed higher amount of VAT.
- Most of the customers are from female category.
- Most of the customers given rating in afternoon.

END