create database walmart;

show tables;

select \* from sales;

-----------------------------------------------------------------------------------------------------------------------------

-----------------------------------------feature engineering------------------------------------------------------------------

--time\_of\_day-----------------------------------------------

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 'Noon'

else 'Evening'

end) as time\_of\_day

from sales;

Alter table sales add column time\_of\_day varchar (20);

select \* from sales;

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 'Noon'

else 'Evening'

end ;

---------------day\_name col to be added-------------

select \* from sales;

SELECT DAYNAME(STR\_TO\_DATE(date, '%d-%m-%Y')) AS day\_name

FROM sales;

alter table sales add column day\_name varchar (30);

update sales set day\_name = dayname(str\_to\_date(date,'%d-%m-%Y'));

-------------------month\_name col to be added------------

select \* from sales;

select date, monthname(str\_to\_date(date,'%d-%m-%Y')) as month\_name from sales;

alter table sales add column month\_name varchar (20);

update sales set month\_name = monthname(str\_to\_date(date,'%d-%m-%Y'));

------------------------------------------------------------------------------------------------------------------------------------------

------------------------------------------------------EDA---------------------------------------------------------------------------------

---------------city and branch analysis---------------------------

select distinct city, count(city) as city\_count from sales group by city;

select distinct branch, city, count(city) as city\_count from sales group by branch, city;

--------------------------------------------------------product\_line----------------------------------------------------------------------

select distinct product\_line, count(product\_line) as product\_line\_count from sales group by product\_line order by count(product\_line) desc;

select \* from sales;

select payment, count(payment) as count from sales group by payment order by count desc;

--------------------------------------------------------revenue---------------------------------------------------------------------------

select \* from sales;

select distinct month\_name from sales;

select distinct month\_name, round(sum(total),2) as sum from sales group by month\_name order by sum desc;

-------largest cogs month-----------------------------------------------------------------------------------------------------------------

select \* from sales;

select distinct month\_name, round(sum(cogs),2) as cogs\_sum from sales group by month\_name order by cogs\_sum desc;

---------------------------largest revenue product line------------------------------------------------------------------------------------

select \* from sales;

select distinct product\_line, round(sum(total),2) as total\_revenue from sales group by product\_line order by total\_revenue desc;

----------------city having largest revenue------------------------------------------------------------------------------------------

select \* from sales;

select distinct city, round(sum(total),2) as sum from sales group by city order by sum desc;

-------------product line with largest VAT-------------------------------------------------------------------------------------------

select \* from sales;

SELECT product\_line, round(AVG(`Tax\_5%`),2) AS avg\_tax

FROM sales

GROUP BY product\_line

ORDER BY avg\_tax desc;

-------------fetching product line and categorising it into good and bad based on avg sales-------------------------------------------

select \* from sales;

select round(avg(total),2) as avg\_sales from sales;

----which branch sold more products than avg products sold-------

select \* from sales;

select round(avg(total),2) as avg\_sales from sales ;

SELECT branch, total,

(SELECT ROUND(AVG(total), 2) FROM sales) AS avg\_total

FROM sales

WHERE total > (SELECT AVG(total) FROM sales);

select avg(quantity) as avg\_qty from sales;

-------which branch sold more products than avg products sold-------

select branch, sum(quantity) as total\_qty,(select round(avg(quantity),2) from sales) as avg\_qty from sales

where quantity > (select avg(quantity) from sales)

group by branch order by total\_qty desc;

------------------------most common product line by gender--------------------------------------

select \* from sales;

select distinct product\_line, gender, count(gender) as cnt\_gender from sales

group by product\_line, gender order by cnt\_gender desc limit 10;

--------------------avg rating of each product line----------------------------------------------

select \* from sales;

select round(avg(rating),2) as avg\_rating from sales;

select distinct product\_line, round(avg(rating),2) as avg\_rating from sales

group by product\_line order by avg\_rating desc;

----------------------------------------------------------------sales------------------------------------------------------------------------

---------------------------------------------------------------------------------------------------------------------------------------------

----number of sales made in each time of the day per weekday---------------------------------------------------------------------------------

select \* from sales;

select round(sum(total),2) as total\_sales, time\_of\_day,day\_name from sales

where day\_name in ('monday','tuesday','wednesday','thursday','friday')

group by time\_of\_day,day\_name order by total\_sales desc limit 10;

---------------------------------just for monday---------------------------------------------------------------------------------------------

select time\_of\_day, count(\*) as total\_sales from sales where day\_name = 'monday'

group by time\_of\_day

order by total\_sales desc;

-----------------------------------which customer type brings the most revenue---------------------------------------------------------------

select \* from sales;

alter table sales rename column `Customer type` to Customer\_type;

select customer\_type, round(sum(total),2) as total\_revenue from sales group by customer\_type order by total\_revenue;

----------------------------------which city has the largest tax percent or vat--------------------------------------------------------------

select \* from sales;

alter table sales rename column `Tax\_5%` to Tax\_percent;

select city, round(avg(tax\_percent),2) as vat from sales group by city order by vat desc;

-----------------------------------which customer type most in terms of vat------------------------------------------------------------------

select \* from sales;

select customer\_type, round(avg(tax\_percent),2) as vat from sales group by customer\_type order by vat desc;

---------------------------------------------------------------------------------------------------------------------------------------------

----------------------------------------------customer---------------------------------------------------------------------------------------

-------------------------------------how many unique customer type are there---------------------------------------------------------------------

select \* from sales;

select distinct customer\_type from sales;

-----------------------------------------------how many unique payment methods does the data have---------------------------------------------

select \* from sales;

select distinct payment from sales;

-----------------------------------------------most common customer type----------------------------------------------------------------------

select \* from sales;

select customer\_type, count(\*) as cnt from sales group by customer\_type order by cnt desc;

------------------------------------------which customer type buys the most-------------------------------------------------------------------

select \* from sales;

select customer\_type, count(\*) as cnt from sales group by customer\_type order by cnt desc;

-----interms of revenue----------------

select customer\_type, round(sum(total),2) as total\_revenue\_generated from sales group by customer\_type order by total\_revenue\_generated desc;

---------------------------------------------------gender of most of the csutomers------------------------------------------------------------

select \* from sales;

select gender, count(\*) cnt from sales group by gender order by cnt desc;

select customer\_type, gender, count(gender) as cnt from sales group by customer\_type, gender

order by cnt desc;

--------------------------------------------------gender distribution per branch--------------------------------------------------------------

select \* from sales;

select branch, gender, count(gender) as gender\_cnt from sales group by branch, gender order by branch, gender\_cnt desc;

-------------------------------------------------what time of the day do customers give most ratings------------------------------------------

select \* from sales;

select round(avg(rating),2) from sales;

select time\_of\_day, Rating from sales where rating > (select round(avg(rating),2) from sales)

order by Rating desc;

select time\_of\_day, avg(rating) as avg\_rating from sales

group by time\_of\_day order by avg\_rating desc;

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

select \* from sales;

select time\_of\_day, branch, round(avg(rating),2) as avg\_ratings from sales group by time\_of\_day, branch

order by avg\_ratings desc;

----------------------------------which day of the week has the best avg ratings--------------------------------------------------------------

select \* from sales;

select day\_name, round(avg(rating),2) as avg\_ratings from sales

group by day\_name order by avg\_ratings desc limit 1;

----------------------------------which day of the week has the best avg ratings per branch---------------------------------------------------

select \* from sales;

select day\_name, branch, round(avg(rating),2) as avg\_ratings from sales

group by day\_name, branch

order by branch, avg\_ratings desc;