create database amazon\_data;

use amazon\_data;

describe sales\_analysis;

select \* from sales\_analysis;

-- SQL Capstone Project on Amazon Slaes Data

-- Creating a New column time\_of\_day

alter table sales\_analysis

add column time\_of\_day varchar(20) after Time;

-- updating values in time\_of\_day column

set sql\_safe\_updates = 0;

update sales\_analysis

set time\_of\_day =

case

when time(time) >= '00:00:00' and time(time) < '12:00:00' then 'Morning'

when time(time) >= '12:00:00' and time(time) < '18:00:00' then 'Afternoon'

else 'Evening'

end;

select \* from sales\_analysis;

-- Creating a New column day\_name

alter table sales\_analysis

add column day\_name varchar(20) after Date;

update sales\_analysis

set day\_name=

case dayofweek(date)

when 1 then 'Sun'

when 2 then 'Mon'

when 3 then 'Tue'

when 4 then 'Wed'

when 5 then 'Thu'

when 6 then 'Fri'

when 7 then 'Sat'

end;

select \* from sales\_analysis;

-- Adding a New column month\_name

alter table sales\_analysis

add column month\_name varchar(20) after day\_name;

-- updating values in month\_name column

update sales\_analysis

set month\_name =

case month(date)

when 1 then 'Jan'

when 2 then 'Feb'

when 3 then 'Mar'

when 4 then 'Apr'

when 5 then 'May'

when 6 then 'Jun'

when 7 then 'Jul'

when 8 then 'Aug'

when 9 then 'Sep'

when 10 then 'Oct'

when 11 then 'Nov'

when 12 then 'Dec'

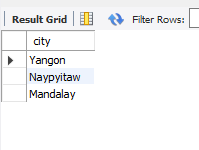
end;

select \* from sales\_analysis;

-- Business Que

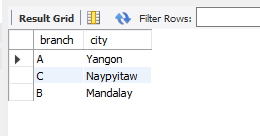
-- 1.What is the count of distinct cities in the dataset?

select distinct city from sales\_analysis;



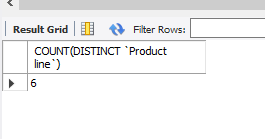
-- 2.For each branch, what is the corresponding city?

select distinct branch,city from sales\_analysis;

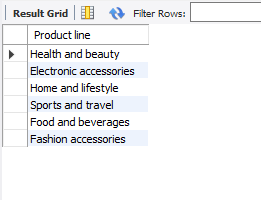


-- 3.What is the count of distinct product lines in the dataset?

SELECT COUNT(DISTINCT `Product line`) FROM sales\_analysis;



SELECT DISTINCT `Product line` FROM sales\_analysis;



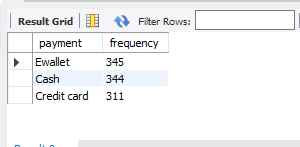
-- 4.Which payment method occurs most frequently?

select payment, count(\*) as frequency

from sales\_analysis

group by payment

order by frequency desc;



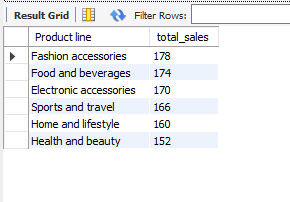
-- 5.Which product line has the highest sales?

select `Product line`, count(\*) as total\_sales

from sales\_analysis

group by `Product line`

order by total\_sales desc;



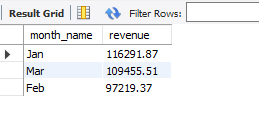
-- 6.How much revenue is generated each month?

select month\_name, round(sum(total),2) as revenue

from sales\_analysis

group by month\_name

order by revenue desc;



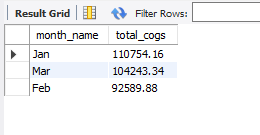
-- 7.In which month did the cost of goods sold reach its peak?

select month\_name, round(sum(cogs),2) as total\_cogs

from sales\_analysis

group by month\_name

order by total\_cogs desc;



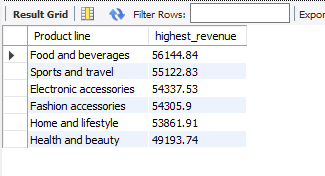
-- 8.Which product line generated the highest revenue?

select `Product line`, round(sum(total),2) as highest\_revenue

from sales\_analysis

group by `Product line`

order by highest\_revenue desc;



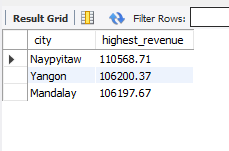
-- 9.In which city was the highest revenue recorded?

select city, round(sum(total),2) as highest\_revenue

from sales\_analysis

group by city

order by highest\_revenue desc;



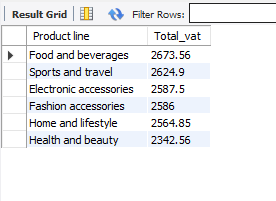
-- 10.Which product line incurred the highest Value Added Tax?

select `Product line`, round(sum(`Tax 5%`), 2) as Total\_vat

from sales\_analysis

group by `Product line`

order by `Total\_vat` desc;



-- 11.For each product line, add a column indicating "Good" if its sales are above average, otherwise "Bad."

select `product line`, round(sum(total),2) as Total\_sales,

case

when sum(total) > (

select avg(total\_sales)

from(

select `product line`, sum(total) as Total\_sales

from sales\_analysis

group by `product line`

) as subquery

) then 'Good'

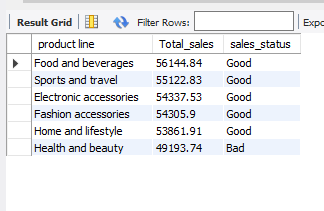
else 'Bad'

end as sales\_status

from sales\_analysis

group by `product line`

order by total\_sales desc;



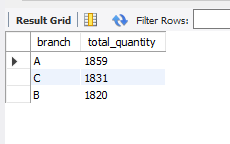
-- 12.Identify the branch that exceeded the average number of products sold.

select branch, sum(quantity) as total\_quantity

from sales\_analysis

group by branch

having sum(quantity) > (select avg(quantity) from sales\_analysis);



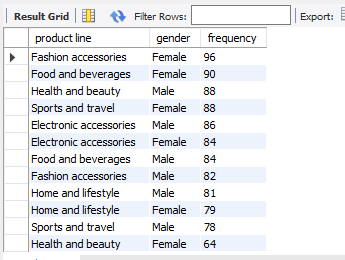
-- 13.Which product line is most frequently associated with each gender?

select `product line`, gender, count(\*) as frequency

from sales\_analysis

group by gender, `product line`

order by frequency desc;



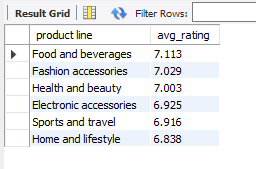
-- 14.Calculate the average rating for each product line.

select `product line`, round(avg(rating),3) as avg\_rating

from sales\_analysis

group by `product line`

order by avg\_rating desc;



-- 15.Count the sales occurrences for each time of day on every weekday.

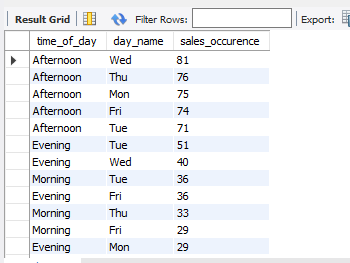
select time\_of\_day, day\_name, count(\*) as sales\_occurence

from sales\_analysis

where day\_name in ('Mon','Tue','Wed','Thu','Fri')

group by time\_of\_day, day\_name

order by sales\_occurence desc;



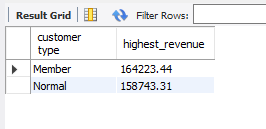
-- 16.Identify the customer type contributing the highest revenue.

select `customer type`, round(sum(total), 2) as highest\_revenue

from sales\_analysis

group by `customer type`

order by highest\_revenue desc;



-- 17.Determine the city with the highest VAT percentage.

select city,

round(sum(`Tax 5%`), 2) as higest\_vat,

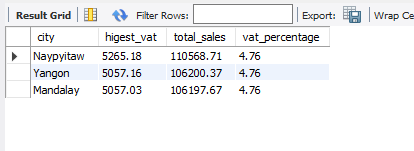
round(sum(Total), 2) as total\_sales,

round((sum(`Tax 5%`) / sum(Total)) \* 100,2) as vat\_percentage

from sales\_analysis

group by city

order by higest\_vat desc;



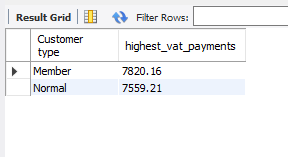
-- 18.Identify the customer type with the highest VAT payments.

select `Customer type`, round(sum(`Tax 5%`),2) as highest\_vat\_payments

from sales\_analysis

group by `Customer type`

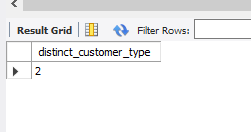
order by highest\_vat\_payments desc;



-- 19.What is the count of distinct customer types in the dataset?

select count(distinct `Customer type`) as distinct\_customer\_type

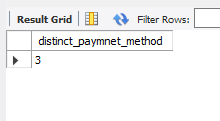
from sales\_analysis;



-- 20.What is the count of distinct payment methods in the dataset?

select count(distinct payment) as distinct\_paymnet\_method

from sales\_analysis;



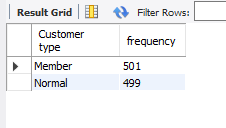
-- 21.Which customer type occurs most frequently?

select `Customer type`, count(`Customer type`) as frequency

from sales\_analysis

group by `Customer type`

order by frequency desc;



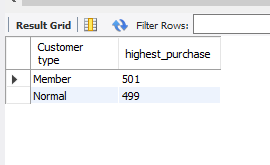
-- 22.Identify the customer type with the highest purchase frequency.

select `Customer type`, count(\*) as highest\_purchase

from sales\_analysis

group by `Customer type`

order by highest\_purchase desc;



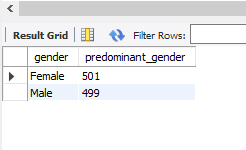
-- 23.Determine the predominant gender among customers.

select gender, count(\*) as predominant\_gender

from sales\_analysis

group by gender

order by predominant\_gender desc;



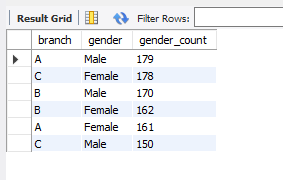
-- 24.Examine the distribution of genders within each branch.

select branch,gender, count(\*) as gender\_count

from sales\_analysis

group by gender,branch

order by gender\_count desc;



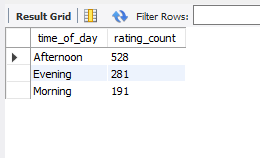
-- 25.Identify the time of day when customers provide the most ratings.

select time\_of\_day, count(rating) as rating\_count

from sales\_analysis

group by time\_of\_day

order by rating\_count desc;



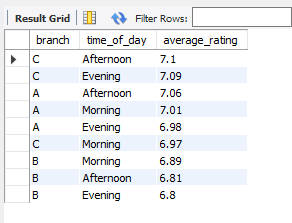
-- 26.Determine the time of day with the highest customer ratings for each branch.

select branch,time\_of\_day, round(avg(rating),2) as average\_rating

from sales\_analysis

group by branch,time\_of\_day

order by average\_rating desc;



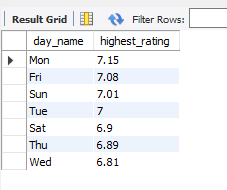
-- 27.Identify the day of the week with the highest average ratings.

select day\_name, round(avg(rating),2) as highest\_rating

from sales\_analysis

group by day\_name

order by highest\_rating desc;



-- 28.Determine the day of the week with the highest average ratings for each branch.

select branch,day\_name, round(avg(rating),2) as highest\_average\_rating

from sales\_analysis

group by branch,day\_name

order by highest\_average\_rating desc;

