



**Thenmozhi R**

**BY**

PHARMA DATA ANALYSIS USING SQL

**INTERNSHIP PROJECT**

1. Retrieve all columns for all records in the dataset.

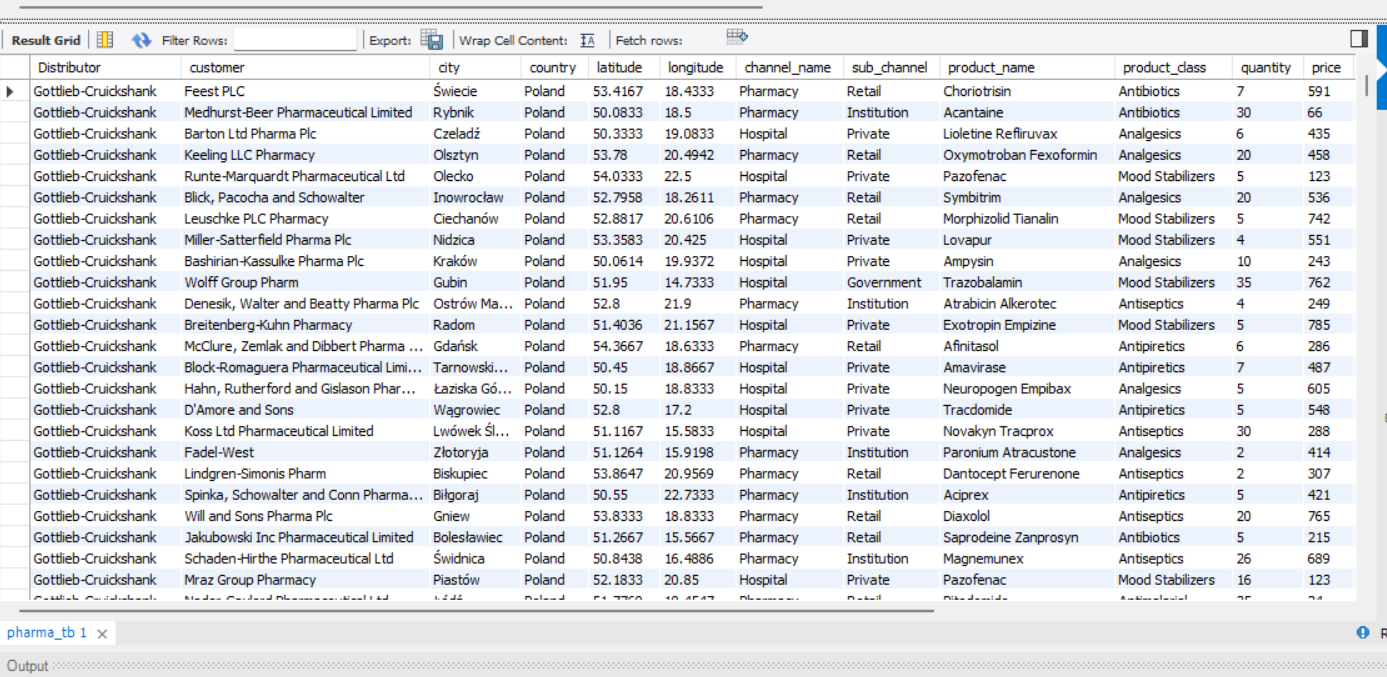
LOAD DATA

INTO TABLE pharma\_tb

FIELDS TERMINATED BY ','

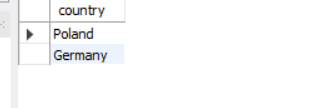
ENCLOSED BY '"'

LINES TERMINATED BY '\n'

IGNORE 1 ROWS;

2. How many unique countries are represented in the dataset?

SELECT DISTINCT country from pharma\_tb;

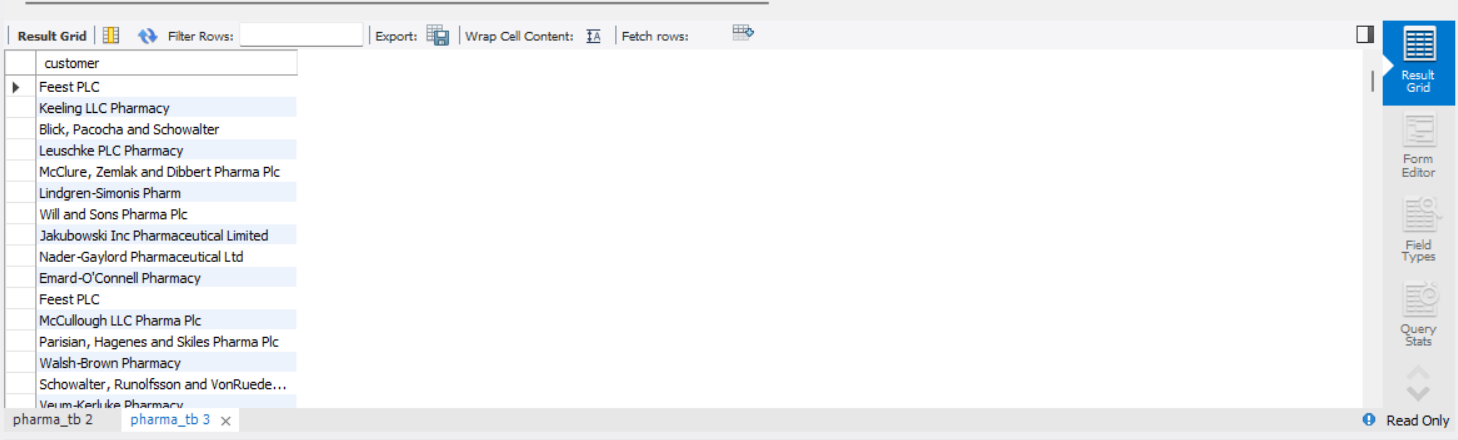


3. Select the names of all the customers on the 'Retail' channel.

SELECT customer

FROM pharma\_tb

WHERE sub\_channel = 'Retail';



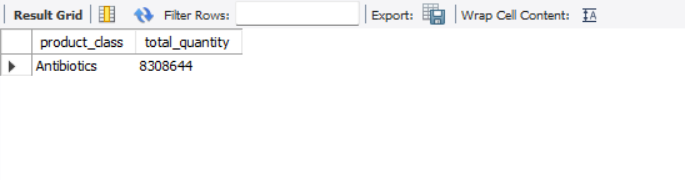
4. Find the total quantity sold for the ' Antibiotics' product class.

SELECT product\_class ,sum(quantity) total\_quantity

FROM pharma\_tb

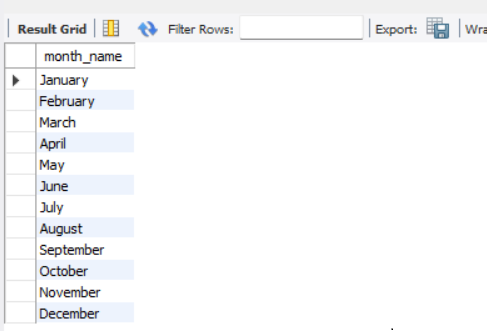
group by product\_class

having product\_class='Antibiotics';



5. List all the distinct months present in the dataset

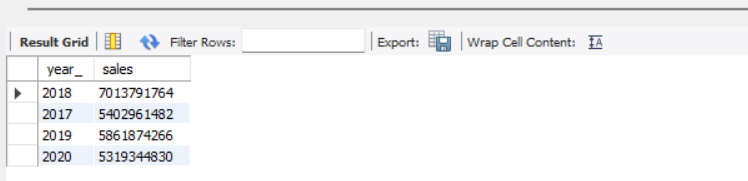
select distinct month\_name from pharma\_tb;



6. Calculate the total sales for each year

select year\_,sum(sales) as sales

from pharma\_tb group by year\_;



7. Find the customer with the highest sales value.

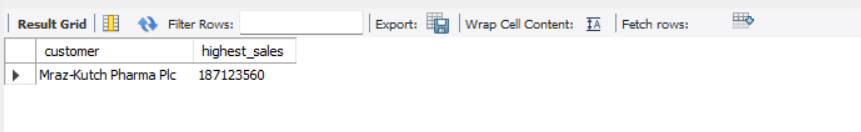
SELECT customer, SUM(sales) AS highest\_sales

FROM pharma\_tb

GROUP BY customer

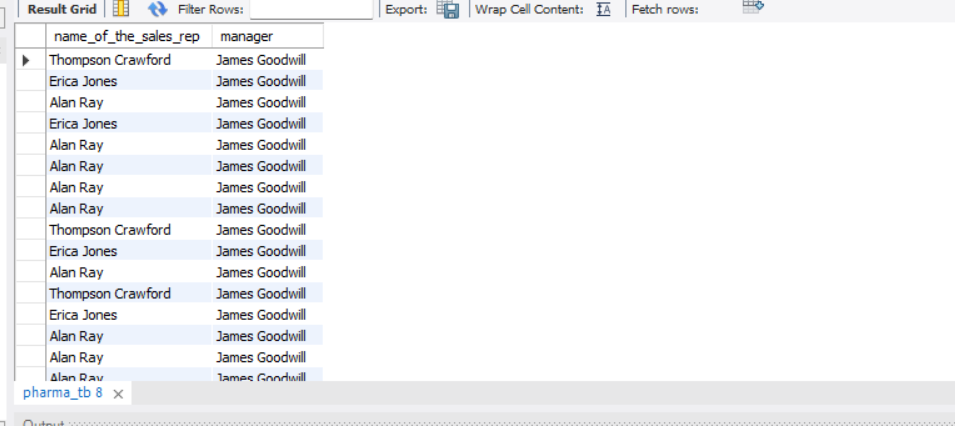
ORDER BY highest\_sales DESC

LIMIT 1;



8. Get the names of all employees who are Sales Reps and are managed by 'James Goodwill'.

select name\_of\_the\_sales\_rep,manager from pharma\_tb where manager='James Goodwill';



9. Retrieve the top 5 cities with the highest sales

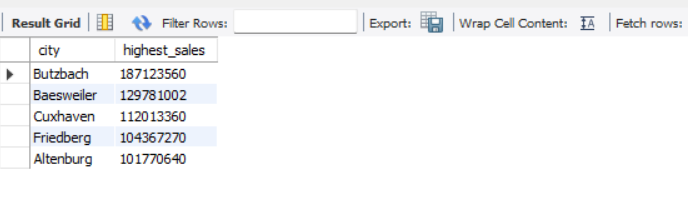
SELECT city, SUM(sales) AS highest\_sales

FROM pharma\_tb

GROUP BY city

ORDER BY highest\_sales DESC

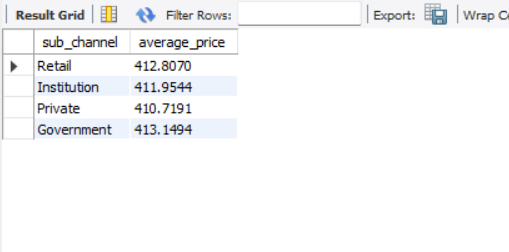
LIMIT 5;



10. Calculate the average price of products in each sub-channel.

select sub\_channel,avg(price) as average\_price from pharma\_tb

group by sub\_channel;



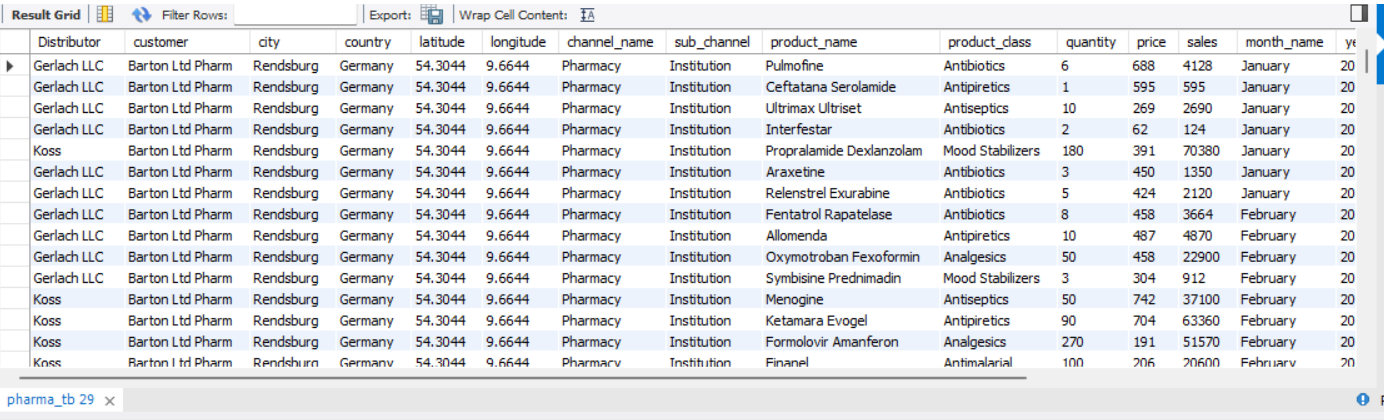
11. Join the 'Employees' table with the 'Sales' table to get the name of the Sales Rep and the corresponding sales records

select name\_of\_the\_sales\_rep,sum(sales) from pharma\_tb group by name\_of\_the\_sales\_rep;



12. Retrieve all sales made by employees from ' Rendsburg ' in the year 2018.

select\*from pharma\_tb where city='Rendsburg' and year\_=2018;



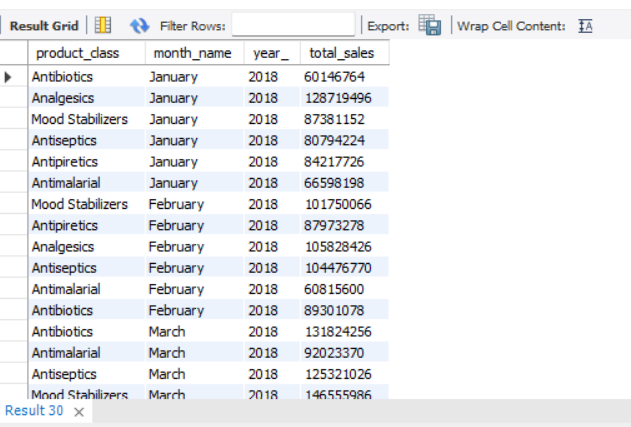
13. Calculate the total sales for each product class, for each month, and order the results by year, month, and product class

SELECT product\_class, month\_name, year\_, SUM(sales) as total\_sales

FROM pharma\_tb

GROUP BY product\_class, month\_name, year\_

LIMIT 50000;



14. Find the top 3 sales reps with the highest sales in 2019

SELECT name\_of\_the\_sales\_rep, SUM(sales) AS total\_sales

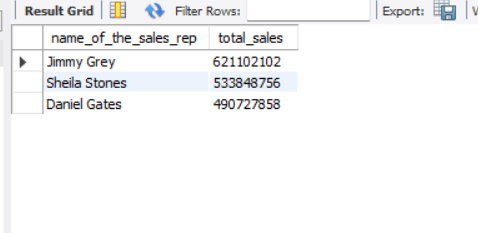
FROM pharma\_tb

WHERE YEAR\_ = 2019

GROUP BY name\_of\_the\_sales\_rep

ORDER BY total\_sales DESC

LIMIT 3;



15. Calculate the monthly total sales for each sub-channel, and then calculate the average monthly sales for each sub-channel over the years.

SELECT

YEAR\_ AS sales\_year,

month\_name AS sales\_month,

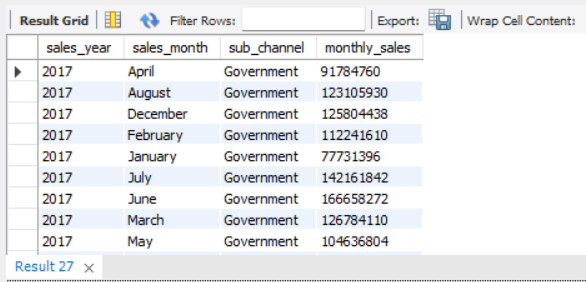
sub\_channel,

SUM(sales) AS monthly\_sales

FROM pharma\_tb

GROUP BY sales\_year, sales\_month, sub\_channel

ORDER BY sub\_channel, sales\_year, sales\_month;



16. Create a summary report that includes the total sales, average price, and total quantity sold for each product class.

SELECT

product\_class,

SUM(sales) AS total\_sales,

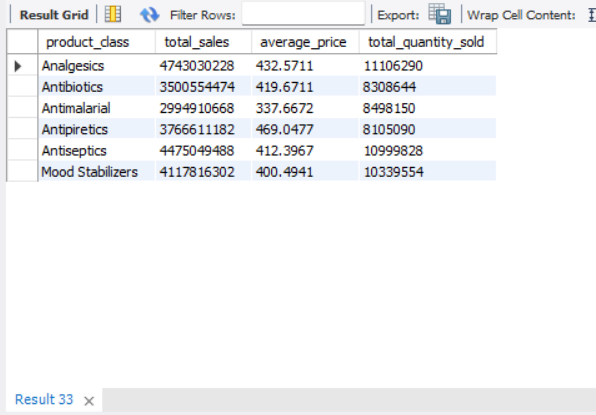
AVG(price) AS average\_price,

SUM(quantity) AS total\_quantity\_sold

FROM pharma\_tb

GROUP BY product\_class

ORDER BY product\_class;



17. Find the top 5 customers with the highest sales for each year.SELECT year\_, customer, total\_sales

FROM (

SELECT year\_, customer, SUM(sales) AS total\_sales,

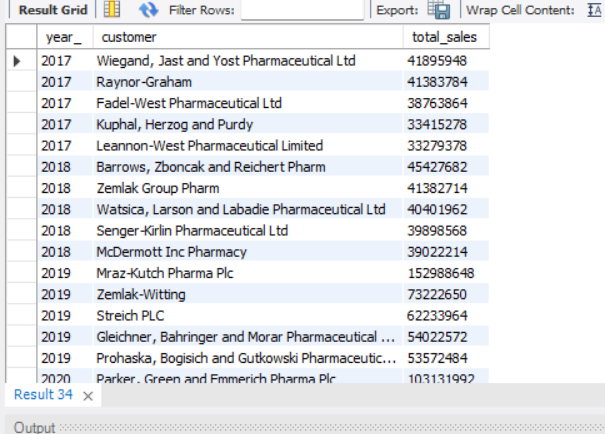
ROW\_NUMBER() OVER (PARTITION BY year\_ ORDER BY SUM(sales) DESC) AS sales\_rank

FROM pharma\_tb

GROUP BY year\_, customer

) ranked\_sales

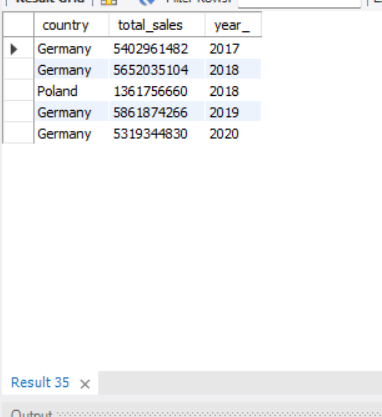
WHERE sales\_rank <= 5;



18. Calculate the year-over-year growth in sales for each country.

select country,round(sum(sales))as total\_sales,year\_ from pharma\_tb group by year\_,country

order by year\_;



19. List the months with the lowest sales for each year

SELECT year\_, month\_name, MIN(total\_sales) AS lowest\_sales

FROM (

SELECT YEAR\_ AS year\_, MONTH\_NAME AS month\_name, SUM(sales) AS total\_sales

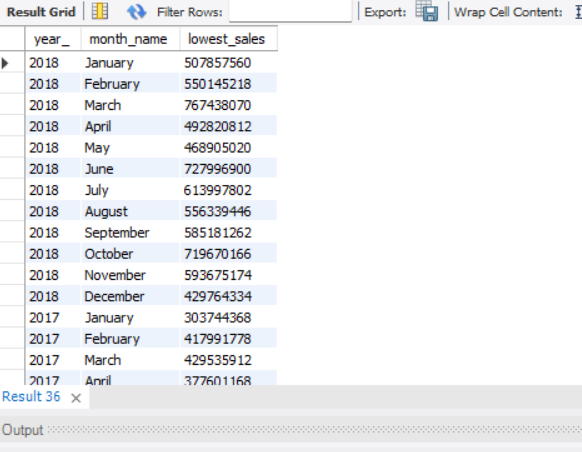
FROM pharma\_tb

GROUP BY YEAR\_, MONTH\_NAME

) AS sales

GROUP BY year\_, month\_name

LIMIT 0, 50000;



20. Calculate the total sales for each sub-channel in each country, and then find the country with the highest total sales for each sub-channel.

WITH tot\_sales AS (

SELECT

RANK() OVER(PARTITION BY sub\_channel ORDER BY SUM(sales)) AS rnk,

SUM(sales) AS total\_sales,

country,

sub\_channel

FROM pharma\_tb

GROUP BY country, sub\_channel

)

SELECT

tot\_sales.country,

tot\_sales.sub\_channel,

tot\_sales.total\_sales

FROM tot\_sales

WHERE tot\_sales.rnk = 1;

