# PHARMA DATA ANALYSIS

In our project, we learned how to leverage SQL to manage data effectively in databases. Mastering about 20 different types of SQL queries, we were able to extract specific information, perform calculations, and organize data neatly.

From simple tasks like finding a name in a list to more complex tasks like analyzing large amounts of data, SQL proved to be an invaluable tool.

This project underscored the importance of SQL in today's data-driven world, where efficiently handling large datasets is crucial for businesses.

Now, we're applying our SQL skills to analyze pharmaceutical sales data, enabling us to uncover valuable insights and drive informed decision-making for the pharmaceutical industry.

AAKASH SHARMA

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

```
21 •  SELECT * FROM pharma;
22
```

Result Grid | Filter Rows: | Export: | Wrap Cell Content: | Fetch rows:

| Distributor | Customer Name | City | Country | Latitude | Longitude | Channel | Sub-channel | Product Name | Product Class | Quantity | Price | Sales | Month | Year | Nam Rep |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| Gottlieb-Cruickshank | Zieme, Doyle and Kunze | Lublin | Poland | 51.2333 | 22.5667 | Hospital | Private | Topipizole | Mood Stabilizers | 4 | 368 | 1472 | January | 2018 | Mary |
| Gottlieb-Cruickshank | Feest PLC | ?wiecie | Poland | 53.4167 | 18.4333 | Pharmacy | Retail | Choriotrisin | Antibiotics | 7 | 591 | 4137 | January | 2018 | Jessi |
| Gottlieb-Cruickshank | Medhurst-Beer Pharmaceutical Limited | Rybnik | Poland | 50.0833 | 18.5 | Pharmacy | Institution | Acantaine | Antibiotics | 30 | 66 | 1980 | January | 2018 | Steve |
| Gottlieb-Cruickshank | Barton Ltd Pharma Plc | Czelad? | Poland | 50.3333 | 19.0833 | Hospital | Private | Lioletine Refliruvax | Analgesics | 6 | 435 | 2610 | January | 2018 | Mary |
| Gottlieb-Cruickshank | Keeling LLC Pharmacy | Olsztyn | Poland | 53.78 | 20.4942 | Pharmacy | Retail | Oxymotroban Fexoformin | Analgesics | 20 | 458 | 9160 | January | 2018 | Anne |

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

```
25 •    select  count(distinct  country) as  Unique_countries
26      from pharma;
27
```

Result Grid | Filter Rows: | Export: | Wrap Cell Content:

| Unique_countries |
| --- |
| 2 |

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

```
28 •  select `customer name`
29    from pharma
30    where `sub-channel` = 'Retail';
31
```

Result Grid | Filter Rows: | Export: | Wrap Cell Content: | Fetc

| customer name |
| --- |
| Feest PLC |
| Keeling LLC Pharmacy |
| Blick, Pacocha and Schowalter |
| Leuschke PLC Pharmacy |
| McClure, Zemlak and Dibbert Pharma Plc |
| Lindgren-Simonis Pharm |
| Will and Sons Pharma Plc |

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

```
33 •  select sum(quantity) as total_quantity
34    from pharma
35    where `product class` = 'Antibiotics';
36
```

Result Grid | Filter Rows: | Export: | Wrap Cell Cor

| total_quantity |
| --- |
| 3957348 |

# 5. List all the distinct months present in the dataset.

```
37 •    select distinct month
38      from pharma;
39
```

Result Grid | Filter Rows: | Export:

| month |
| --- |
| January |
| February |
| March |
| April |
| May |
| June |
| July |

# 6. Calculate the total sales for each year.

```
40 •    select year, sum(sales) as total_sales
41      from pharma
42      group by year;
43
```

Result Grid | Filter Rows: | Export:

| year | total_sales |
|---|---|
| 2018 | 3506897354 |
| 2017 | 2701480741 |
| 2019 | 2930937133 |
| 2020 | 2108573686 |

# 7. Find the customer with the highest sales value.

```
44 •  select `customer name` , sum(sales) as total_sales
45    from pharma
46    group by 1
47    order by 2 desc
48    limit 1;
49
```

Result Grid | Filter Rows: | Export: | Wrap Cell Content:

| customer name | total_sales |
|---|---|
| Mraz-Kutch Pharma Plc | 92441208 |

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

```
31 •  select `Name of Sales Rep` ,manager
32    from pharma
33    where manager = 'James Goodwill';
```

Result Grid | Filter Rows: | Export: | Wrap Cel

| Name of Sales Rep | manager |
|---|---|
| Thompson Crawford | James Goodwill |
| Erica Jones | James Goodwill |
| Alan Ray | James Goodwill |
| Erica Jones | James Goodwill |
| Alan Ray | James Goodwill |
| Alan Ray | James Goodwill |

# 9. Retrieve the top 5 cities with the highest sales.

```
54 •  select city, sum(sales) as total_sales
55    from pharma
56    group by 1
57    order by 2 desc
58    limit 5;
59
60 •  select `sub-channel` , round(avg(price),2) as avg
```

Result Grid | Filter Rows: | Export: | Wrap Cel

| city | total_sales |
|---|---|
| Butzbach | 92441208 |
| Baesweiler | 62935628 |
| Cuxhaven | 54910769 |
| Friedberg | 48843679 |
| Emsdetten | 45368998 |

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

```
60 •  select `sub-channel` , round(avg(price),2) as avg_price
61    from pharma
62    group by 1
63    order by 2 desc;
64
```

Result Grid | Filter Rows: | Export: | Wrap Cell Content:

| sub-channel | avg_price |
|---|---|
| Government | 413.27 |
| Retail | 412.63 |
| Institution | 411.86 |
| Private | 410.63 |

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

SQL Output Statistics

```
select p.name_of_sales_rep, sum(p.sales)
    from pharma_data p
   group by p.name_of_sales_rep;
```

| | NAME_OF_SALES_REP | SUM(P.SALES) |
|---|---|---|
| 1 | Mary Gerrard | 875270762.91093 |
| 2 | Morris Garcia | 901195482.5 |
| 3 | Erica Jones | 871372192 |
| 4 | Abigail Thompson | 981056993.864903 |
| 5 | Jimmy Grey | 985969993.944742 |
| 6 | Jessica Smith | 881698369.002429 |
| 7 | Stella Given | 888340902.41899 |
| 8 | Steve Pepple | 875449982.57143 |
| 9 | Alan Ray | 842637242.2 |
| 10 | Thompson Crawford | 866964886.178331 |
| 11 | Sheila Stones | 958203898.244147 |
| 12 | Anne Wu | 920168301.173581 |
| 13 | Daniel Gates | 950658635.185934 |

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

```
65 •  select `Name of Sales Rep` ,city , year
66    from pharma
67    where city = 'Rendsburg' and year = 2018 ;
68
```

Result Grid | Filter Rows: | Export: | Wrap Cell Cont

| Name of Sales Rep | city | year |
|---|---|---|
| Jessica Smith | Rendsburg | 2018 |
| Daniel Gates | Rendsburg | 2018 |
| Alan Ray | Rendsburg | 2018 |
| Sheila Stones | Rendsburg | 2018 |
| Thompson Crawford | Rendsburg | 2018 |
| Jimmy Grey | Rendsburg | 2018 |

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

```
69 •  select `product class` , sum(sales) as total_sales, month,year
70    from pharma
71    group by 1,3,4
72    order by 4,3,1;
73
```

Result Grid | Filter Rows: | Export: | Wrap Cell Content:

| product class | total_sales | month | year |
|---|---|---|---|
| Analgesics | 32223716 | April | 2017 |
| Antibiotics | 40029226 | April | 2017 |
| Antimalarial | 17789675 | April | 2017 |
| Antipiretics | 22868812 | April | 2017 |
| Antiseptics | 42712211 | April | 2017 |
| Mood Stabilizers | 33176944 | April | 2017 |
| Analgesics | 49744520 | August | 2017 |

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

```
74 •  select `Name of Sales Rep`, sum(sales) as total_sales
75    from pharma
76    where year = 2019
77    group by 1
78    order by 2 desc
79    limit 3;
80
```

Result Grid | Filter Rows: | Export: | Wrap Cell Content:

| Name of Sales Rep | total_sales |
|---|---|
| Jimmy Grey | 310551051 |
| Sheila Stones | 266924378 |
| Daniel Gates | 245363929 |

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

```
60 •  select `Sub-channel` , sum(sales) as total_sales , month,year, avg(sales) as avg_monthsales
61    from pharma
62    group by 1,3,4;
```

Result Grid | Filter Rows: | Export: | Wrap Cell Content:

| Sub-channel | total_sales | month | year | avg_monthsales |
|---|---|---|---|---|
| Private | 12071005 | January | 2018 | 10635.2467 |
| Retail | 12160561 | January | 2018 | 13678.9213 |
| Institution | 8343003 | January | 2018 | 10682.4622 |
| Government | 9230114 | January | 2018 | 12160.8880 |
| Private | 23646583 | February | 2018 | 21516.4540 |

```
64 •  select `sub-channel`, avg(total_sales) as avg_monthsales
65    from (select `Sub-channel` , sum(sales) as total_sales , month,yea
66    from pharma
67    group by 1,3,4) as month_sales
68    group by 1;
```

Result Grid | Filter Rows: | Export: | Wrap Cell Content:

| sub-channel | avg_monthsales |
|---|---|
| Private | 19351184.8571 |
| Retail | 18760376.0000 |
| Institution | 15471572.0714 |
| Government | 15280331.2857 |

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

```
91 •  select `product class`, sum(sales) as total_sales, round(avg(price),2) as avg_price ,sum(quantity) as total_quantity
92    from pharma
93    group by 1;
94
```

Result Grid | Filter Rows: | Export: | Wrap Cell Content:

| product class | total_sales | avg_price | total_quantity |
|---|---|---|---|
| Mood Stabilizers | 1949739660 | 400.43 | 4884402 |
| Antibiotics | 1662994558 | 419.99 | 3957348 |
| Analgesics | 2249708536 | 432.37 | 5299414 |
| Antiseptics | 2145954075 | 412.32 | 5295097 |
| Antipiretics | 1798780695 | 468.85 | 3875536 |
| Antimalarial | 1440711390 | 337.52 | 4053219 |

# 17. Find the top 5 customers with the highest sales for each year.

```
95 •  select `name of sales rep` ,sum(sales) as total_sales ,year
96    from pharma
97    group by 1,3
98    order by 2
99    limit 5;
100
```

Result Grid | Filter Rows: | Export: | Wrap Cell Content: | Fet

| name of sales rep | total_sales | year |
|---|---|---|
| Jessica Smith | 124776812 | 2020 |
| Mary Gerrard | 136087388 | 2020 |
| Steve Pepple | 148347538 | 2020 |
| Jimmy Grey | 149288327 | 2020 |
| Thompson Crawford | 154275883 | 2020 |

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

```sql
SELECT country,year,total_sales,
    LAG(total_sales) OVER (PARTITION BY country ORDER BY year) AS previous_year_sales,
    ((total_sales - LAG(total_sales) OVER (PARTITION BY country ORDER BY year)) /
    LAG(total_sales) OVER (PARTITION BY country ORDER BY year)) * 100 AS yoy_growth
FROM (SELECT country,YEAR,SUM(sales) AS total_sales
    FROM pharma
    GROUP BY country,year
) AS yearly_sales;
```

Result Grid | Filter Rows: | Export: | Wrap Cell Content:

| country | year | total_sales | previous_year_sales | yoy_growth |
|---|---|---|---|---|
| Germany | 2017 | 2701480741 | NULL | NULL |
| Germany | 2018 | 2826017552 | 2701480741 | 4.6099 |
| Germany | 2019 | 2930937133 | 2826017552 | 3.7126 |
| Germany | 2020 | 2108573686 | 2930937133 | -28.0580 |
| Poland | 2018 | 680879802 | NULL | NULL |

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

```
115 •   select month,year, sum(sales) as total_sales
116     from pharma
117     group by 1,2
118     order by 3 ;
119
```

Result Grid | Filter Rows: | Export: | Wrap Cell C

| month | year | total_sales |
|---|---|---|
| November | 2020 | 59968305 |
| January | 2019 | 97664076 |
| April | 2020 | 135409908 |
| January | 2017 | 151872184 |
| January | 2020 | 170724633 |
| October | 2019 | 187630559 |
| April | 2017 | 188800584 |

# 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

```
135 • with subchannelsales as (select country,`sub-channel`,sum(sales) as total_sales
136   from pharma
137   group by 1,2),
138   maxsubchannel as (select `sub-channel`,max(total_sales) as max_sales
139   from subchannelsales
140   group by 1)
141   select sc.country,sc.`sub-channel`,sc.total_sales from subchannelsales  sc
142   join maxsubchannel msc on sc.`sub-channel` = msc.`sub-channel` and sc.total_sales = msc.max_sales;
143
```

Result Grid | Filter Rows: | Export: | Wrap Cell Content:

| country | sub-channel | total_sales |
|---|---|---|
| Germany | Government | 2763800068 |
| Germany | Private | 2217743137 |
| Germany | Institution | 2607947184 |
| Germany | Retail | 2977518723 |