Name: Subhamoy Das
Email: subhamoy1873@gmail.com
LinkedIn: https://linkedin.com/in/itssubhamoydas/

Introduction

Hello there! Welcome to our online internship on SQL for Pharma Data Analysis . In this program, we delve into the fascinating realm of data science and healthcare, combining the power of SQL with the critical task of data analysing. As a student pursuing B.Sc. Hons. in Computer Science with a keen interest in software development, you'll find this internship to be a valuable opportunity to apply your skills and expand your knowledge.

Throughout the internship, we'll be working with a dataset encompassing various factors such as: pharma(Distributor, Customer_Name, City, Country, Latitude, Longitude, Channel, Sub-channel, Product Name, Product_Class, Quantity, Price, Sales, Month, Year, Name_of_Sales_Rep, Manager, Sales_Team). This real-world dataset mirrors the complexity of healthcare data and provides a rich environment for honing your SQL skills.

Aim: To analyze the given dataset 'Pharma_data_analysis.xlsx' and perform the following queries in MySQL.

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

Ans: SELECT * FROM pharma;

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

Ans: SELECT COUNT(DISTINCT Country) AS UniqueCountriesCount FROM pharma;

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

Ans:

```
SELECT Customer_Name
FROM pharma
WHERE Channel = 'Retail';
```

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

Ans:

```
SELECT SUM(Quantity) AS TotalQuantitySold
```

Name: Subhamoy Das
Email: subhamoy1873@gmail.com
LinkedIn: https://linkedin.com/in/itssubhamoydas/

FROM pharma

WHERE Product_Class = 'Electronics';

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

Ans: SELECT DISTINCT Month

FROM pharma;

6. Calculate the total sales for each year.

Ans:

SELECT Year, SUM(Sales) AS TotalSales

FROM pharma

GROUP BY Year;

7. Find the customer with the highest sales value.

Ans:

SELECT Customer_Name, MAX(Sales) AS HighestSales

FROM pharma

GROUP BY Customer_Name

ORDER BY HighestSales DESC

LIMIT 1;

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

Ans:

SELECT DISTINCT srep.Name_of_Sales_Rep

FROM pharma AS srep

JOIN pharma AS manager ON srep.Manager = manager.Name_of_Sales_Rep

WHERE manager.Manager = 'John Smith'

AND srep.Sales_Team = 'Sales Rep';

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

Ans:

SELECT City, SUM(Sales) AS TotalSales

Name: Subhamoy Das
Email: subhamoy1873@gmail.com
LinkedIn: https://linkedin.com/in/itssubhamoydas/

```
FROM pharma

GROUP BY City

ORDER BY TotalSales DESC

LIMIT 5;
```

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

Ans:

```
SELECT Sub_channel, AVG(Price) AS AveragePrice

FROM pharma

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.

Ans:

```
SELECT e.Employee_Name, p.*

FROM Employees AS e

JOIN pharma AS p ON e.Name_of_Sales_Rep = p.Name_of_Sales_Rep;
```

12. Retrieve all sales made by employees from 'New York' in the year 2022.

Ans:

```
SELECT *

FROM pharma

WHERE City = 'New York' AND Year = 2022;
```

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

Ans:

```
SELECT Year, Month, Product_Class, SUM(Sales) AS TotalSales

FROM pharma

GROUP BY Year, Month, Product_Class

ORDER BY Year, Month, Product_Class;
```

14. Find the top 3 sales reps with the highest sales in 2023.

Name: Subhamoy Das
Email: subhamoy1873@gmail.com
LinkedIn: https://linkedin.com/in/itssubhamoydas/

Ans:

```
SELECT Name_of_Sales_Rep, SUM(Sales) AS TotalSales

FROM pharma

WHERE Year = 2023

GROUP BY Name_of_Sales_Rep

ORDER BY TotalSales 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.

Ans:

```
SELECT Sub_channel, Month, SUM(Sales) AS MonthlyTotalSales

FROM pharma

GROUP BY Sub_channel, Month

ORDER BY Sub_channel, Month;


WITH MonthlyAvgSales AS (

   SELECT Sub_channel, Month, AVG(Sales) AS AvgMonthlySales

   FROM pharma

   GROUP BY Sub_channel, Month

)

SELECT Sub_channel, AVG(AvgMonthlySales) AS OverallAvgMonthlySales

FROM MonthlyAvgSales

GROUP BY Sub_channel

ORDER BY Sub_channel;
```

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

Ans:

```
SELECT
```

Name: Subhamoy Das
Email: subhamoy1873@gmail.com
LinkedIn: https://linkedin.com/in/itssubhamoydas/

```
    Product_Class,
    SUM(Sales) AS TotalSales,
    AVG(Price) AS AveragePrice,
    SUM(Quantity) AS TotalQuantitySold
FROM
    pharma
GROUP BY
    Product_Class
ORDER BY
    Product_Class;
```

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

Ans:

```
SELECT
    Year,
    Customer_Name,
    SUM(Sales) AS TotalSales
FROM
    pharma AS p1
WHERE
    (
        SELECT COUNT(DISTINCT Customer_Name)
        FROM pharma AS p2
        WHERE p1.Year = p2.Year AND p1.Sales <= p2.Sales
    ) <= 5
GROUP BY
    Year, Customer_Name
ORDER BY
```

Name: Subhamoy Das
Email: subhamoy1873@gmail.com
LinkedIn: https://linkedin.com/in/itssubhamoydas/

```
    Year, TotalSales DESC;
```

18. Calculate the year-over-year growth in sales for each country. 2 of 2

Ans:

```
SELECT
    Country,
    Year,
    AVG(Sales) AS AverageSales,
    LAG(AVG(Sales)) OVER (PARTITION BY Country ORDER BY Year) AS PreviousYearAverageSales,
    ((AVG(Sales) - LAG(AVG(Sales)) OVER (PARTITION BY Country ORDER BY Year)) / LAG(AVG(Sales)) OVER (PARTITION BY Country ORDER BY Year)) * 100 AS YoYGrowth
FROM
    pharma
GROUP BY
    Country, Year
ORDER BY
    Country, Year;
```

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

Ans:

```
SELECT
    Year,
    Month,
    MIN(TotalSales) AS LowestSales
FROM (
    SELECT
        Year,
        Month,
        SUM(Sales) AS TotalSales
    FROM
```

Name: Subhamoy Das
Email: subhamoy1873@gmail.com
LinkedIn: https://linkedin.com/in/itssubhamoydas/

```
        pharma
    GROUP BY
        Year, Month
) AS MonthlySales
GROUP BY
    Year
ORDER BY
    Year, LowestSales;
```

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.

Ans:

```
SELECT
    t1.Country,
    t1.Sub_channel,
    t1.TotalSales
FROM (
    SELECT
        Country,
        Sub_channel,
        SUM(Sales) AS TotalSales
    FROM
        pharma
    GROUP BY
        Country, Sub_channel
) AS t1
JOIN (
    SELECT
        Sub_channel,
```

Name: Subhamoy Das
Email: subhamoy1873@gmail.com
LinkedIn: https://linkedin.com/in/itssubhamoydas/

```
    MAX(TotalSales) AS MaxSales
  FROM (
    SELECT
      Country,
      Sub_channel,
      SUM(Sales) AS TotalSales
    FROM
      pharma
    GROUP BY
      Country, Sub_channel
  ) AS t2
  GROUP BY
    Sub_channel
) AS t3
ON t1.Sub_channel = t3.Sub_channel AND t1.TotalSales = t3.MaxSales;
```