Prashant Singh Chauhan
Email - chauhanp72@gmail.com
LinkedIn - https://www.linkedin.com/in/prashant-singh-chauhan-1072298a/

Project Abstract

Welcome to my immersive SQL-driven Pharma Data Analysis project done as part of my Data Analyst Internship with PSYLIQ, where the fusion of data science and healthcare opens doors to a realm of possibilities! This program is a gateway to unlocking the potential of SQL in dissecting and comprehending healthcare data. As an MSc grad in Physics passionate about software development and Data Analytics, this internship promises an invaluable opportunity to wield my skills and broaden my horizons. Throughout this enriching journey, we embark on an exploration of a multifaceted dataset encompassing pivotal elements such as Pharma_data (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 comprehensive real-world dataset mirrors the intricacies of healthcare data, serving as a robust platform to sharpen your SQL prowess.

Pharma Data Assessment Details

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

```sql
SELECT * FROM pharma_data;
```

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

```sql
SELECT COUNT(DISTINCT Country) AS UniqueCountriesCount FROM pharma_data;
```

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

```sql
SELECT [Customer Name]
FROM pharma_data
WHERE Channel = 'Retail';
```

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

```sql
SELECT SUM(Quantity) AS TotalQuantitySold
FROM pharma_data
WHERE [Product Class] = 'Electronics';
```

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

```sql
SELECT DISTINCT Month
FROM pharma_data
```

6. Calculate the total sales for each year.

```sql
SELECT Year, SUM(Sales) AS TotalSales
FROM pharma_data
GROUP BY Year;
```

7. Find the customer with the highest sales value.

```
SELECT top 1 [Customer Name], MAX(Sales) AS HighestSales
FROM pharma_data
GROUP BY [Customer Name]
ORDER BY HighestSales DESC
```

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

```
SELECT DISTINCT(a.[Name of Sales Rep])
FROM pharma_data AS a
JOIN pharma_data AS m ON a.Manager = m.[Name of Sales Rep]
WHERE m.Manager = 'John Smith'
 AND a.[Sales Team] = 'Sales Rep';
```

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

```
SELECT top 5 City, SUM(Sales) AS TotalSales
FROM pharma_data
GROUP BY City
ORDER BY TotalSales DESC
```

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

```
SELECT [Sub-channel], AVG(Price) AS AveragePrice
FROM pharma_data
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 *
FROM pharma_data
WHERE City = 'Rendsburg'
AND YEAR([year]) = 2018;
```

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

```
--SELECT e.Employee_Name, p.*
--FROM Employees AS e
--JOIN pharma_data AS p ON e.Name_of_Sales_Rep = p.Name_of_Sales_Rep;
```

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

```
SELECT *
FROM pharma_data
WHERE City = 'Rendsburg'
AND YEAR([year]) = 2018;

SELECT
    [Year],
    [Month],
    [Product Class],
    SUM(Sales) AS TotalSales
FROM pharma_data
GROUP BY [Year], [Month], [Product Class]
ORDER BY [Year], [Month], [Product Class];
```

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

```
SELECT top 3
    [Name of Sales Rep],
    SUM(Sales) AS TotalSales
FROM pharma_data
WHERE YEAR([year]) = 2019
GROUP BY [Name of Sales Rep]
ORDER BY TotalSales DESC
```

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 top 3
    [Name of Sales Rep],
    SUM(Sales) AS TotalSales
FROM pharma_data
WHERE YEAR([year]) = 2019
GROUP BY [Name of Sales Rep]
ORDER BY TotalSales DESC
```

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

```
;WITH MonthlyTotalSales AS (
    SELECT
        [Year] AS SalesYear,
        [Month] AS SalesMonth,
        [Sub-channel],
        SUM(Sales) AS MonthlySales
    FROM pharma_data
    GROUP BY [Year], [Month], [Sub-channel]
),
AverageMonthlySales AS (
    SELECT
        [Sub-channel],
        AVG(MonthlySales) AS AvgMonthlySales
    FROM MonthlyTotalSales
    GROUP BY [Sub-channel]
)
SELECT
    [Sub-channel],
    AVG(AvgMonthlySales) AS AverageMonthlySales
FROM AverageMonthlySales
GROUP BY [Sub-channel];
```

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

```
SELECT
    [Product Class],
    SUM(Sales) AS TotalSales,
    AVG(Price) AS AveragePrice,
    SUM(Quantity) AS TotalQuantity
FROM pharma_data
GROUP BY [Product Class];
```

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

```
;WITH RankedSales AS (
    SELECT
        [Customer Name],
        [Year] AS SalesYear,
        SUM(Sales) AS TotalSales,
        ROW_NUMBER() OVER (PARTITION BY [Year] ORDER BY SUM(Sales) DESC) AS
SalesRank
    FROM pharma_data
    GROUP BY [Customer Name], [Year]
)
SELECT
    [Customer Name],SalesYear,TotalSales FROM RankedSales WHERE SalesRank <= 5;
```

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

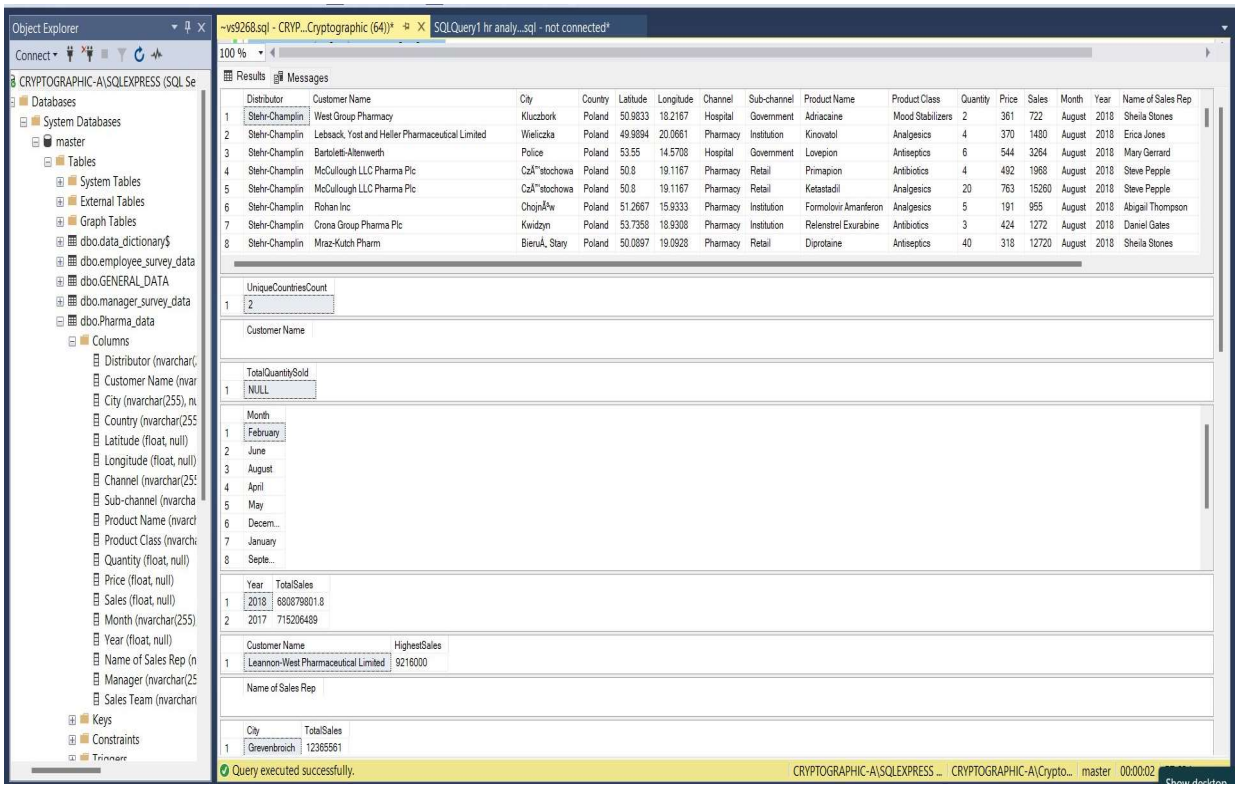
```
;WITH SalesByYear AS (
    SELECT
        [Year] AS SalesYear,
        Country,
        SUM(Sales) AS TotalSales
    FROM pharma_data
    GROUP BY [Year], Country
)
SELECT
    Country,
    SalesYear,
    TotalSales,
    LAG(TotalSales) OVER (PARTITION BY Country ORDER BY SalesYear) AS
PreviousYearSales,
    CASE
        WHEN LAG(TotalSales) OVER (PARTITION BY Country ORDER BY SalesYear) IS
NULL THEN NULL
        ELSE (TotalSales - LAG(TotalSales) OVER (PARTITION BY Country ORDER BY
SalesYear)) / LAG(TotalSales) OVER (PARTITION BY Country ORDER BY SalesYear) * 100
    END AS YearOverYearGrowth
FROM SalesByYear
ORDER BY Country, SalesYear;
```

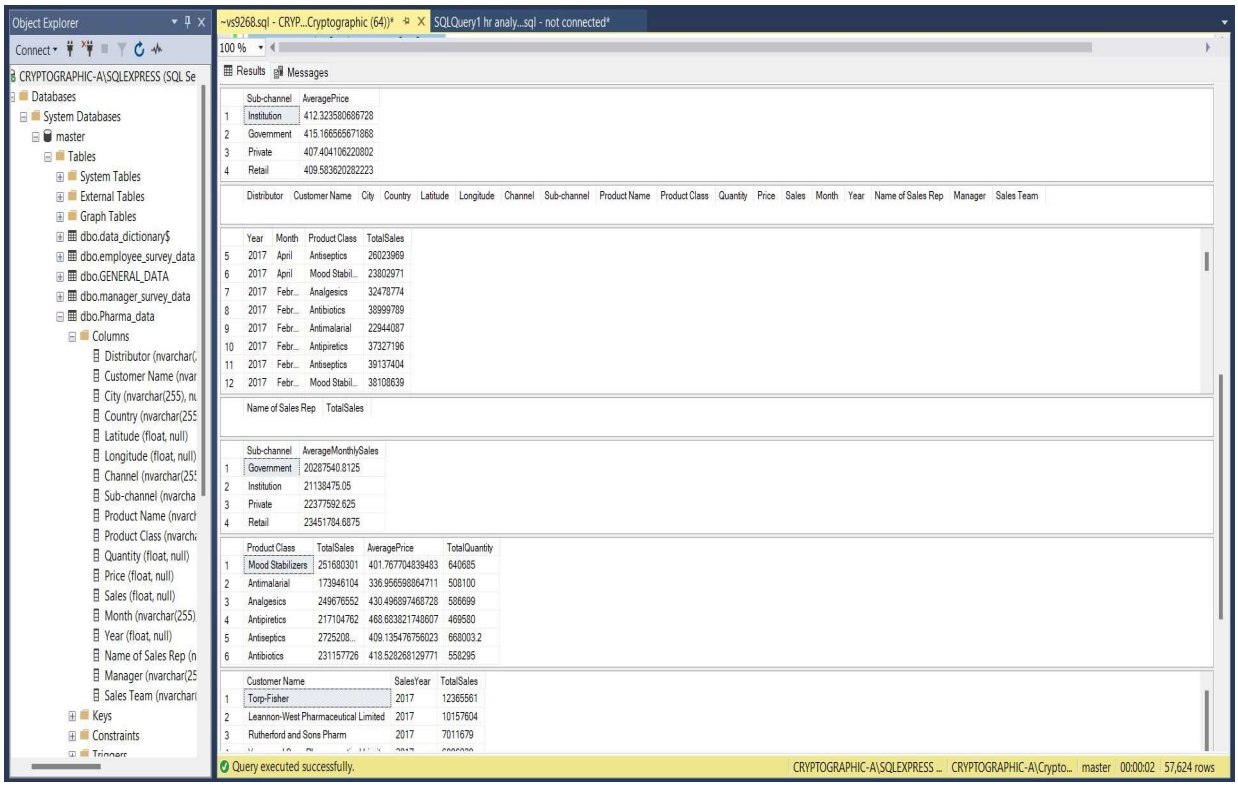
20. Calculate the total sales for each sub-channel in each country, and then find the country with the highest total sales for each sub-channel.

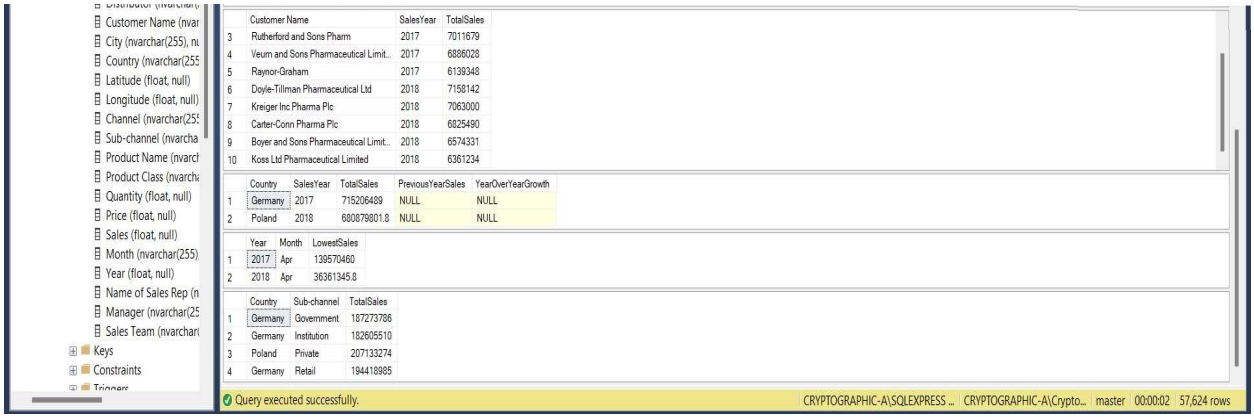
```
;WITH SubChannelSales AS (
    SELECT
        Country,
        [Sub-channel],
        SUM(Sales) AS TotalSales,
        ROW_NUMBER() OVER(PARTITION BY [Sub-Channel] ORDER BY SUM(Sales) DESC) AS
CountryRank
    FROM
        Pharma_data -- Replace 'YourTableName' with your actual table name
    GROUP BY
        Country,
        [Sub-channel]
)

SELECT
    Country,
    [Sub-channel],
    TotalSales
FROM
    SubChannelSales
WHERE
    CountryRank = 1 ORDER BY [Sub-Channel];
```

RESULTS from SQL Server






Customer Name (nvar
City (nvarchar(255), nu
Country (nvarchar(255
Latitude (float, null)
Longitude (float, null)
Channel (nvarchar(25
Sub-channel (nvarcha
Product Name (nvarch
Product Class (nvarch
Quantity (float, null)
Price (float, null)
Sales (float, null)
Month (nvarchar(255)
Year (float, null)
Name of Sales Rep (n
Manager (nvarchar(25
Sales Team (nvarchar(
Keys
Constraints
Triggers
Customer Name SalesYear TotalSales
3 Rutherford and Sons Pharm 2017 7011679
4 Veum and Sons Pharmaceutical Limit... 2017 6886028
5 Raynor-Graham 2017 6139348
6 Doyle-Tillman Pharmaceutical Ltd 2018 7158142
7 Kreiger Inc Pharma Plc 2018 7063000
8 Carter-Conn Pharma Plc 2018 6825490
9 Boyer and Sons Pharmaceutical Limit... 2018 6574331
10 Koss Ltd Pharmaceutical Limited 2018 6361234
Country SalesYear TotalSales PreviousYearSales YearOverYearGrowth
1 Germany 2017 715206489 NULL NULL
2 Poland 2018 680879801.8 NULL NULL
Year Month LowestSales
1 2017 Apr 139570460
2 2018 Apr 36361345.8
Country Sub-channel TotalSales
1 Germany Government 187273786
2 Germany Institution 182605510
3 Poland Private 207133274
4 Germany Retail 194418985
Query executed successfully.
CRYPTOGRAPHIC-A\SQLEXPRESS ...
CRYPTOGRAPHIC-A\Crypto...
master
00:00:02
57,624 rows