



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.

```
SELECT * FROM pharma_data;
```

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

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

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

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

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

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

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

```
SELECT DISTINCT Month
FROM pharma_data
```

6. Calculate the total sales for each year.

```
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

Object Explorer

Connect

CRYPTOGRAPHIC-A\SQLEXPRESS (SQL Se

Databases

System Databases

master

Tables

System Tables

External Tables

Graph Tables

dbo.data_dictionary\$

dbo.employee_survey_data

dbo.GENERAL_DATA

dbo.manager_survey_data

dbo.Pharma_data

Columns

Distributor (nvarchar(

Customer Name (nvar

City (nvarchar(255), n

Country (nvarchar(255

Latitude (float, null)

Longitude (float, null)

Channel (nvarchar(25

Sub-channel (nvarchar

Product Name (nvarchar

Product Class (nvarchar

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

~vs9268.sql - CRYPT...Cryptographic (64))

SQLQuery1 hr analy...sql - not connected*

100 %

Results

Messages

	Distributor	Customer Name	City	Country	Latitude	Longitude	Channel	Sub-channel	Product Name	Product Class	Quantity	Price	Sales	Month	Year	Name of Sales Rep
1	Stehr-Champlin	West Group Pharmacy	Kluczbork	Poland	50.9833	18.2167	Hospital	Government	Adriacaine	Mood Stabilizers	2	361	722	August	2018	Sheila Stones
2	Stehr-Champlin	Leback, Yost and Heller Pharmaceutical Limited	Wieliczka	Poland	49.9894	20.0661	Pharmacy	Institution	Kinovatol	Analgesics	4	370	1480	August	2018	Erica Jones
3	Stehr-Champlin	Bartolotti-Altenwerth	Police	Poland	53.55	14.5708	Hospital	Government	Lovepion	Antiseptics	6	544	3264	August	2018	Mary Gerrard
4	Stehr-Champlin	McDullough LLC Pharma Plc	Czajka Stochowa	Poland	50.8	19.1167	Pharmacy	Retail	Primapion	Antibiotics	4	492	1968	August	2018	Steve Pepple
5	Stehr-Champlin	McDullough LLC Pharma Plc	Czajka Stochowa	Poland	50.8	19.1167	Pharmacy	Retail	Ketastadil	Analgesics	20	763	15260	August	2018	Steve Pepple
6	Stehr-Champlin	Rohan Inc	Chojnów	Poland	51.2667	15.9333	Pharmacy	Institution	Formolovir Amantferon	Analgesics	5	191	955	August	2018	Abigail Thompson
7	Stehr-Champlin	Crona Group Pharma Plc	Kwidzyn	Poland	53.7358	18.9308	Pharmacy	Institution	Relenstrel Exurabine	Antibiotics	3	424	1272	August	2018	Daniel Gates
8	Stehr-Champlin	Mraz-Kutch Pharm	Bieruń, Stary	Poland	50.0897	19.0928	Pharmacy	Retail	Diprotaine	Antiseptics	40	318	12720	August	2018	Sheila Stones

UniqueCountriesCount

12

Customer Name

TotalQuantitySold

1NULL

Month

1February

2June

3August

4April

5May

6Decem...

7January

8Septe...

Year TotalSales

12018680879801.8

22017715206489

Customer Name HighestSales

1Leannon-West Pharmaceutical Limited9216000

Name of Sales Rep

City TotalSales

1Grevembroich12365561

Query executed successfully.

CRYPTOGRAPHIC-A\SQLEXPRESS ... CRYPTOGRAPHIC-A\Crypto... master 00:00:02

Object Explorer: CRYPTOGRAPHIC-A\SQLEXPRESS (SQL Se)

Connect: master

Tables: System Tables, External Tables, Graph Tables, dbo.data_dictionary\$, dbo.employee_survey_data, dbo.GENERAL_DATA, dbo.manager_survey_data, dbo.Pharma_data

Columns: Distributor (nvarchar), Customer Name (nvarchar), City (nvarchar(255)), Country (nvarchar(255)), Latitude (float, null), Longitude (float, null), Channel (nvarchar(255)), Sub-channel (nvarchar), Product Name (nvarchar), Product Class (nvarchar), Quantity (float, null), Price (float, null), Sales (float, null), Month (nvarchar(255)), Year (float, null), Name of Sales Rep (nvarchar), Manager (nvarchar(255)), Sales Team (nvarchar)

Keys: Constraints, Triggers

Query Results:

Sub-channel	AveragePrice
1 Institution	412.323580686728
2 Government	415.166565671868
3 Private	407.404106220802
4 Retail	409.593620282223

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
5	2017	April	Antiseptics	26023969													
6	2017	April	Mood Stabl...	23802971													
7	2017	Febr...	Analgesics	32478774													
8	2017	Febr...	Antibiotics	38999789													
9	2017	Febr...	Antimalarial	22944087													
10	2017	Febr...	Antipiretics	37327196													
11	2017	Febr...	Antiseptics	39137404													
12	2017	Febr...	Mood Stabl...	38108639													

Name of Sales Rep TotalSales

Sub-channel	AverageMonthlySales
1 Government	20287540.8125
2 Institution	21138475.05
3 Private	22377592.625
4 Retail	23451784.6875

Product Class	TotalSales	AveragePrice	TotalQuantity
1 Mood Stabilizers	251680301	401.767704839483	640685
2 Antimalarial	173946104	336.956598864711	508100
3 Analgesics	249676552	430.496897468728	586699
4 Antipiretics	217104762	468.683821749607	469580
5 Antiseptics	2725208...	409.135476756023	668003.2
6 Antibiotics	231157726	418.528268129771	558295

Customer Name SalesYear TotalSales

1 Top-Fisher 2017 12365561

2 Leannon-West Pharmaceutical Limited 2017 10157604

3 Rutherford and Sons Pharm 2017 7011679

Query executed successfully.

CRYPTOGRAPHIC-A\SQLEXPRESS ... CRYPTOGRAPHIC-A\Crypto... master 00:00:02 57,624 rows

Customer Name (nvarchar(255), null)	Customer Name	SalesYear	TotalSales		
City (nvarchar(255), null)	3	Rutherford and Sons Pharm	2017	7011679	
Country (nvarchar(255), null)	4	Veum and Sons Pharmaceutical Limit...	2017	6886028	
Latitude (float, null)	5	Raynor-Graham	2017	6139348	
Longitude (float, null)	6	Doyle-Tillman Pharmaceutical Ltd	2018	7158142	
Channel (nvarchar(255), null)	7	Kreiger Inc Pharma Plc	2018	7063000	
Sub-channel (nvarchar(255), null)	8	Carter-Corn Pharma Plc	2018	6825490	
Product Name (nvarchar(255), null)	9	Boyer and Sons Pharmaceutical Limit...	2018	6574331	
Product Class (nvarchar(255), null)	10	Koss Ltd Pharmaceutical Limited	2018	6361234	
Quantity (float, null)					
Price (float, null)					
Sales (float, null)					
Month (nvarchar(255), null)					
Year (float, null)					
Name of Sales Rep (nvarchar(255), null)					
Manager (nvarchar(255), null)					
Sales Team (nvarchar(255), null)					
Keys					
Constraints					
Triggers					

Country	SalesYear	TotalSales	PreviousYearSales	YearOverYearGrowth
Germany	2017	715206489	NULL	NULL
Poland	2018	680679801.8	NULL	NULL

Year	Month	LowestSales
2017	Apr	139570460
2018	Apr	36361345.8

Country	Sub-channel	TotalSales
Germany	Government	187273786
Germany	Institution	182605510
Poland	Private	207133274
Germany	Retail	194418985

Query executed successfully.

CRYPTOGRAPHIC-A\SQLXPRESS ... CRYPTOGRAPHIC-A\Crypto... master 00:00:02 57,624 rows