

Pharmaceutical Sales Data Analysis Report

1. Project Overview

This project analyzes a large pharmaceutical sales dataset containing over **254,083 records** from two countries: **Germany** and **Poland**. The goal was to explore trends, measure performance, and extract business insights using **SQL Server**.

The key objectives of the project include:

- Cleaning and preparing the raw dataset
- Calculating key performance indicators (KPIs)
- Identifying top-performing products, teams, and sales representatives
- Analyzing sales across channels and product categories
- Evaluating year-over-year growth and cumulative trends

All queries and metrics were documented clearly, and the dataset was filtered for outliers to ensure accurate results.

Data Cleaning Process

Before analysis, the dataset was thoroughly cleaned to ensure accuracy and reliability. The following steps were taken:

1. Null Value Handling

- Null values were found in **Sales** and **Quantity** columns.
- Records with nulls in these critical fields (around **2,500 rows**) were **deleted** as imputing zeros could skew the sales analysis.

2. Duplicate Records

- Duplicate rows were identified using all columns (distributor, customer, product, sales, etc.).
- **No duplicates were found**

3. Outlier Detection

- Outliers in `Sales`, `Price`, and `Quantity` were detected using the **Interquartile Range (IQR)** method.
- Over **34,000 records** were identified as outliers.
- Instead of deleting them, a **filtered view** was created (only including records where quantity < **60,000**), allowing analysis on a more realistic subset without losing valuable insights.

These cleaning steps improved the quality of the data and enhanced the accuracy of all KPIs and visual findings.

KPI & Sales Insights

Total Sales

- The ****total pharmaceutical sales**** amounted to ****\$11,945,457,289.20****, reflecting the large scale of transactions in the dataset.

Top Performing Products

- ****Highest-selling product****: ` Sumanazole `
- ****Lowest-selling product****: ` Amphisirox `

product_name	Total_sales	Rank_type
Sumanazole	113907983.00	Highest
Amphisirox	2628775.00	lowest

- The ****top 5 products by total sales**** were identified using aggregate `SUM(sales)` grouped by product name.

product_name	Total_Sales
Sumanazole	113907983.00
Betanem	107541499.00
Docstryl Rivacin	105260490.00
Travoloride	102568100.00
Propratecan	101231306.00

These insights help identify which drugs generate the most revenue and which might need reconsideration in marketing or distribution.

Total Sales by year

- Total sales over the four year and best year sales

year	year_total_sales	Total_quantity	num_customers	mum_products_perYear	num_sales_rep
2017	2717929120.82	6594434.1	551	240	13
2018	3545594979.60	8587404.19	751	240	13
2019	2872327291.78	6978875.50	551	240	13
2020	2671542097.00	6577645	551	240	13

- From the above table we found the highest sales year is '2018' but no change in number of products, sales representative or price of products but there is a significant and sudden increase in the number of customers – 200 new client appeared only in that year.
- All those 200 customers are in Poland.
- Possible reasons behind the 2018 customer surge in Poland:
 - o Short-term promotional campaign or discounts
 - o Temporary distribution partnership
 - o Change in customer registration policy

Highest and lowest Months in sales

- The three months which have the highest sales

Month	month_total_sales
November	1133482678.00
August	1115816472.00
March	1114970562.00

- The three months which have the Lowest sales

Month	month_total_sales
January	679603097.00
April	810870504.80
May	877610718.26

Product Sales by Year

- Sales for each product were analyzed across years to find ****best-performing years****
- ****Best year**** varies per product, showing how demand shifts over time.

Year-over-Year Sales Growth

- Using ``LAG()`` and ``relative growth rate``, we calculated:
- The difference in sales compared to the previous year.
- Percentage growth per year per product.
- This identifies growth trends and declining products.
- ex. Sample:

product_name	year	total_sales	previous_year_sales	sales_difference	relative_growth	growth_rank
Abatatriptan	2017	19266030.00	NULL	NULL	NULL	4
Abatatriptan	2018	15668072.00	19266030.00	-3597958.00	-0.19	3
Abatatriptan	2019	20850942.00	15668072.00	5182870.00	0.33	1
Abatatriptan	2020	21708694.00	20850942.00	857752.00	0.04	2
Abilovir Aprotasol	2017	7169119.00	NULL	NULL	NULL	4
Abilovir Aprotasol	2018	13311841.60	7169119.00	6142722.60	0.86	1
Abilovir Aprotasol	2019	6139656.00	13311841.60	-7172185.60	-0.54	3
Abilovir Aprotasol	2020	8351643.00	6139656.00	2211987.00	0.36	2

From here we can display the highest relative growth for each product over the four years where the growth rank equal 1.

Cumulative Product Sales

- Calculated ****cumulative yearly sales**** per product using ``SUM() OVER(...)``.
- Useful for understanding long-term product performance.
- ex. Sample:

product_name	year	yearly_sales	cumulative_sales
Abatatriptan	2017	19266030.00	19266030.00
Abatatriptan	2018	15668072.00	34934102.00
Abatatriptan	2019	20850942.00	55785044.00
Abatatriptan	2020	21708694.00	77493738.00
Abilovir Aprotasol	2017	7169119.00	7169119.00
Abilovir Aprotasol	2018	13311841.60	20480960.60
Abilovir Aprotasol	2019	6139656.00	26620616.60
Abilovir Aprotasol	2020	8351643.00	34972259.60

Product Distribution by Country & Channel

Number of Products per Country

- Countries like **Poland** and **Germany** had the same number of products and the same products

Sales within each country and country's sales percent

It turns out that Germany's sales significantly exceed Poland's, as shown in the following table:

country	country_sales	sales_percent
Poland	681056670.80	0.06%
Germany	11126336818.40	0.94%

Product Diversity by Sub-channel

- Evaluated how many unique products exist in sub-channels such as **retail pharmacy** or **hospital direct**.
- Helpful to understand **channel specialization** within each country.

channel	num_sub_Channel	num_products
Hospital	2	240
Pharmacy	2	240

Each channel has the same number of Sub_channels and the same number of products.

Cumulative Sales Over Years

- Tracks **long-term growth** and helps in forecasting and performance analysis.

Total sales per subchannel and per country

It turns out that in **Germany 'retail'** has the highest sales and in **Poland 'private'** has the highest sales.

sub_channel	country	Total_sales
Retail	Germany	3079134909.40
Private	Poland	207195997.00

Product Variety per Channel Over Time

Checking the number and type of products in each channel over the years to see their impact on total sales, but it turns out that all channels have the same number of products and the same products.

Team Performance Analysis

Number of Teams and Total Sales per Country

- Analyzed team presence and performance across countries.
- To determine whether the number of teams in each country is different or equal and thus determine the impact of the number on total sales.
- The number of teams is balanced across both countries, and making comparisons valid.

country	num_teams	total_sales
Poland	4	681056670.80
Germany	4	11126336818.40

Total Sales and Quantity per Team

- Identified **Delta Team** as the top-performing team in both revenue and quantity.

sales_team	total_quantity	total_sales
Delta	8980978.14005303	3678610089.34
Bravo	6779539.28995056	2769508902.94
Charlie	6581881.46999145	2737334360.55
Alfa	6395959.90997601	2621940136.37

Best Sales Team over all and per Year

-showed **team winner** for the 4 years

Top_sales_team	total_quantity	total_sales
Delta	8980978.14005303	3678610089.34

- Showed **yearly winners**, useful for internal benchmarking.

sales_team	year	total_quantity	total_sales
Delta	2017	2144135.55	897592896.51
Delta	2018	2691369	1110005009.00
Delta	2019	2243892.59	899561438.83
Delta	2020	1901581	771450745.00

All Teams: Sales in Germany vs. Poland

- Broader comparison across all teams to understand **country-wise performance**.

sales_team	country	total_sales
Delta	Poland	206526991.00
Delta	Germany	3472083098.34
Charlie	Germany	2577393509.55
Charlie	Poland	159940851.00
Bravo	Germany	2612792725.14
Bravo	Poland	156716177.80
Alfa	Poland	157872651.00
Alfa	Germany	2464067485.37

Top-Selling Product per Team per Country

- Identified each team's **highest revenue-generating product per country**.

sales_team	product_name	country	T_sales
Alfa	Cephozumab Synmethate	Germany	32524800.00
Alfa	Magnemunex	Poland	3849443.00
Bravo	Sumanazole	Germany	39089397.00
Bravo	Angioparin Brimosumab	Poland	3507526.00
Charlie	Propratecan	Germany	44535964.00
Charlie	Propratecan	Poland	3764640.00
Delta	Antaparin Varizyme	Germany	45555939.00
Delta	Flutacerol	Poland	4255440.00

Average Sales per Order and per Rep

- Calculated average sales:

- **Per order**: total sales ÷ number of orders.

- **Per rep**: total sales ÷ number of sales representatives.

- Balanced comparison across teams regardless of size.

- It is evident from the table that the **Delta team has the highest total sales** because its team has a **larger number of sales_rep than the other teams**. Therefore, by calculating the **average total sales** for each sales representative in each team, we find that the **Bravo team outperforms the Delta team**.

sales_team	num_sales_rep	num_orders	T_sales	AVG_order	AVG_per_rep
Bravo	3	58028	2769508902.94	47727.11	923169634.31
Delta	4	77450	3678610089.34	47496.58	919652522.34
Alfa	3	58076	2621940136.37	45146.71	873980045.46
Charlie	3	57888	2737334360.55	47286.73	912444786.85

Sales Representative Performance Analysis

Top Sales Representative Overall

- Identified the ****highest performing sales rep**** across the entire dataset based on ``SUM(sales)``.

name_of_sales_rep	sales_team	total_rep_sales	ranking
Abigail Thompson	Bravo	990329559.87	1

Top Sales Rep per Team

- Finding the ****leading rep**** per team.

name_of_sales_rep	sales_team	total_rep_sales
Abigail Thompson	Bravo	990329559.87
Sheila Stones	Delta	969481143.24
Daniel Gates	Charlie	965562042.19
Thompson Crawford	Alfa	884001321.17

Top Sales Rep per Channel

- Ranked sales reps per ****channel**** (e.g., Pharmacy, Hospital) by total sales.

- Supports assigning reps to the ****channels where they perform best****.

name_of_sales_rep	channel	num_orders	total_rep_sales
Sheila Stones	Hospital	9633	502361643.24
Abigail Thompson	Pharmacy	9846	538189928.65

Yearly Sales Growth per Rep

- Tracked annual sales for each sales rep using ``LAG()`` and calculated:

- ****Sales difference**** year-over-year.

- ****Relative growth rate****.

- Highlights reps with ****consistent growth**** or ****declining performance****, guiding training or incentives.

- 2018 recorded the highest relative growth for most sales reps.

- Eg. Sample

name_of_sales_rep	year	sales_year_rep	prev_year_sales	sales_diff	relatvie_growth
Abigail Thompson	2017	229172884.31	NULL	NULL	NULL
Abigail Thompson	2018	307070055.80	229172884.31	77897171.49	0.34
Abigail Thompson	2019	213468726.76	307070055.80	-93601329.04	-0.3
Abigail Thompson	2020	240617893.00	213468726.76	27149166.24	0.13
Alan Ray	2017	217625234.00	NULL	NULL	NULL

Alan Ray	2018	243916404.20	217625234.00	26291170.20	0.12
Alan Ray	2019	203442483.00	243916404.20	-40473921.20	-0.17
Alan Ray	2020	189913825.00	203442483.00	-13528658.00	-0.07

Manager Performance Analysis

Total Sales per Manager

- Calculated total sales managed by each ****manager**** across their sales teams.
- Helps evaluate ****managerial impact**** on revenue performance.

manager	sales_team	mang_sales
Britanny Bold	Delta	3678610089.34
Tracy Banks	Bravo	2769508902.94
Alisha Cordwell	Charlie	2737334360.55
James Goodwill	Alfa	2621940136.37

Product Category Analysis

Total Sales per Product Class

- Aggregated total sales for each `product class` using `SUM(sales)`.
- **Analgesics class** drive the most revenue

product_class	T_class_sales
Analgesics	2312211230.29
Antiseptics	2263031818.65
Mood Stabilizers	2079304774.64
Antipiretics	1911154411.18
Antibiotics	1776080974.54
Antimalarial	1465610279.90

Yearly Sales per Product Class

- Analyzed sales by `product_class` per `year` to track ****annual performance trends****.
- Helps identify:
 - Growth patterns of specific categories.
 - Product classes that need ****strategic adjustments****.
- we can find the best sales over years for each product class
- sample..

product_class	year	T_class_sales
Analgesics	2017	545584061.00
Analgesics	2018	703889698.00
Analgesics	2019	562021570.29
Analgesics	2020	500715901.00
Antibiotics	2017	391431326.00
Antibiotics	2018	531661940.00
Antibiotics	2019	432297531.54
Antibiotics	2020	420690177.00

Pharmaceutical Sales Data Analysis - Summary Report

Key Findings and Insights

Overall Performance

- Total pharmaceutical sales analyzed: ****\$11.95 billion**** (Germany: 94% Poland: 6%)
- Dataset contained ****254,083 records**** from Germany and Poland
- Data cleaning removed **2,500 records with null values** and identified **34,000** outliers

Top Products

- Best-selling product^{**}: **Sumanazole** (\$113.9 million)^{**}
- **Top 5 products**^{**} accounted for over \$530 million in sales^{**}
- **Worst-performing product**^{**}: Amphesirox (\$2.63 million)^{**}

Sales Trends

- **Best Year**: 2018 (3.55B \$) due to a sudden surge of 200 new clients **in Poland** (likely from a short-term campaign)
- **** Analgesics **** was the highest-performing product class (\$2.31 billion)
- **Germany's retail** sub-channel and **Poland's private** sub-channel were the top performers
- ****November**** was the best month in sales (1.13 B \$)
- ****January**** was the worst month in sales (679.6 M\$)

Team Performance

- **** Delta Team **** was the overall top performer (\$3.68 billion)
- **** Bravo Team **** had the highest average sales per representative (\$923 (million/rep
- **** Abigail Thompson **** (Bravo Team) was the top sales representative (\$990 (million

Channel Analysis

- Both hospital and pharmacy channels carried the same 240 products
- No product differentiation between channels was observed
- Sales distribution between channels was balanced

Manager Performance

- **** Brittany Bold **** (Delta Team manager) oversaw the highest sales (\$3.68 (billion
- Manager performance directly correlated with team performance rankings

Business Recommendations

1. ****Product Strategy****:

- Increase focus on top-performing products (Sumanazole, Betanem, Docstryl (Rivacin
- Review marketing and distribution strategy for low-performing products like Amphesirox

2. **Geographical Focus**:

- Germany dominates sales (94%) - consider expanding resources there
- Poland Growth:
 - Investigate the 200-customer surge (likely a promo campaign) and replicate successful strategies.

3. **Team Optimization**:

- Replicate **Bravo Team's** high per-representative performance across other teams
- **Balance Delta Team's size** (4 reps vs. 3 in other teams) to improve efficiency.
- Provide additional **training/resources to Alfa Team** which showed lowest average performance

4. **Channel Management**:

- Consider product differentiation between channels to better target customer needs
- Focus on retail sub-channel in Germany and private sub-channel in Poland

5. **Growth Opportunities**:

- Analyze 2018 success factors to replicate growth patterns
- Focus on Analgesics category which shows consistent strong performance

6. **Seasonal Adjustments**:

- **Capitalize on peak months** (November, August, March).
- **Address January slump** (lowest sales: 680M).

Technical Approach

- Comprehensive data cleaning process (null handling, outlier detection)
- Advanced SQL techniques used:
 - Window functions (DENSE_RANK(), LAG())
 - CTEs and subqueries
 - Aggregation and growth rate calculations
- Balanced analysis considering both absolute performance and normalized metrics
- **YOY Growth & cumulative Sales Trends.**

This analysis provides actionable insights for sales strategy optimization, resource allocation, and performance improvement across markets, products, and teams.