

# Pharmaceutical Sales & Manufacturing Performance Dashboard

## Project Overview

This Power BI dashboard provides a comprehensive, data-driven solution for analyzing sales performance and manufacturing efficiency within a pharmaceutical company. It enables stakeholders to identify growth opportunities, monitor product quality, and make informed strategic decisions.

## Business Problem & Goal

In the dynamic pharmaceutical industry, companies face the constant challenge of optimizing sales strategies, ensuring product quality, and identifying high-potential drugs. Traditional reporting often lacks the interactivity and depth needed for rapid decision-making, leading to missed opportunities and inefficiencies.

This project addresses these challenges by consolidating disparate data points into a unified, interactive Power BI dashboard. Its primary goal is to empower sales, marketing, and operations teams with real-time, actionable insights to:

- Analyze drug sales performance across various dimensions (e.g., therapeutic area, region).
- Identify year-over-year growth trends and top-performing products.
- Monitor manufacturing batch success rates and pinpoint common failure reasons.
- Discover optimal batch sizes and manufacturing plant performance.
- Uncover strategic opportunities for investment based on product lifecycle and growth potential.

## Key Insights & Business Value

This dashboard provides immediate, actionable insights, enabling the business to:

- **Identify Growth Drivers:** Instantly visualize monthly revenue trends, year-over-year growth percentages, and overall average monthly growth, allowing for quick identification of high-performing drugs and therapeutic areas.
- **Optimize Product Portfolio:** The 'Drug Type Performance & Opportunity' chart helps pinpoint which drug formats (e.g., Tablets vs. Injections) contribute most to revenue versus their average price per unit, guiding strategic focus for R&D and marketing efforts.
- **Enhance Manufacturing Efficiency:** The 'Overall Batch Success Rate' KPI provides a high-level overview, while the 'Drug Failure Breakdown' heat map quickly highlights specific drugs and failure reasons (e.g., "Drug X has highest 'Contamination' failures"), enabling targeted process improvements.
- **Inform Production Strategy:** The 'Batch Size vs. Success Rate' scatter plot helps analyze if certain batch sizes correlate with higher success rates, optimizing production planning.
- **Strategic Investment Allocation:** By combining sales performance with manufacturing quality data, the dashboard supports informed decisions on where to

allocate resources for product development, marketing campaigns, and quality control initiatives.

## Technical Skills & Tools Utilized

This project demonstrates proficiency across the entire business intelligence pipeline, leveraging the following technologies and skills:

- **Microsoft Power BI Desktop:**
- **Data Acquisition & Transformation (Power Query / M Language):** Connected to various data sources (CSV, MySQL), performed complex data cleaning (handling mixed date formats, missing values, outliers, inconsistent text), and optimized data loading processes.
- **Data Modeling:** Designed and implemented a robust Star Schema, established one-to-many relationships, managed active/inactive relationships, and incorporated a best-practice Calendar table for time intelligence.
- **DAX (Data Analysis Expressions):** Developed a comprehensive suite of calculated measures and columns, including advanced time intelligence functions (YTD, YoY), conditional logic, complex aggregations (AVERAGEX, SUMMARIZE), and custom sorting logic.
- **Data Visualization & Dashboard Design:** Created interactive, multi-page dashboards featuring diverse visual types (Line charts, Matrix with conditional formatting, Scatter plots, Combo charts, Treemaps, Cards, and Slicers) to present actionable insights clearly and intuitively.
- **User Experience (UX):** Implemented slicer synchronization, tooltips, and an intuitive layout for enhanced user interaction.
- **SQL (MySQL):**
- **Database Integration:** Connected Power BI directly to a MySQL database.
- **Advanced Querying:** Wrote and utilized complex SQL queries (Common Table Expressions - CTEs, Window Functions like LAG()) for data preparation and analysis, demonstrating in-database transformation capabilities.
- **Data Cleansing in SQL:** Applied SQL functions to clean mixed date formats, handle missing numerical values, and standardize text data directly within the database.
- **Python:**
- **Data Generation:** Developed custom Python scripts (using Pandas, NumPy, Faker) to simulate a realistic, large-scale pharmaceutical dataset with embedded data quality challenges, providing a robust foundation for the project.
- **Data Analysis & Problem Solving:** Applied analytical thinking to identify key business questions, translated raw data into meaningful metrics, and systematically debugged complex data quality and platform integration challenges.

## Data Source & Methodology

The data utilized in this project is a synthetic dataset meticulously generated using Python (Faker, Pandas, and NumPy) to simulate realistic transactional and master data for a pharmaceutical company. This approach allowed for the deliberate embedding of common real-world data quality issues (e.g., mixed date formats, missing values, outliers, inconsistent text entries) to showcase robust data cleansing and transformation capabilities.

### Key Data Entities:

- **Drugs:** Master data including DrugID, DrugName, TherapeuticArea, DrugType, and AvgCostPerUnit.
- **Sales:** Transactional data detailing SaleID, SaleDate, DrugID, Region, UnitsSold, and Revenue.

- **Manufacturing Batches:** Transactional data covering BatchID, DrugID, ManufacturingDate, ManufacturingPlantID, BatchSizeUnits, Batch Status, and FailureReason.

### Methodology:

1. **Raw Data Generation:** Python scripts created CSV files mimicking real-world data complexity.
2. **Initial Data Ingestion & Cleansing (Power Query):** CSV files were loaded into Power BI, where extensive transformations were performed using Power Query (M language) to standardize formats, handle missing data, and flag outliers.
3. **Database Integration & Advanced Cleansing (MySQL SQL):** Data was then loaded into a MySQL database. Further advanced cleansing and type conversion (e.g., handling persistent mixed date formats, explicit DECIMAL conversion) were performed using SQL queries directly within MySQL.
4. **Data Modeling:** A Star Schema was designed in Power BI, connecting dimension tables (Drugs, Calendar) to fact tables (Sales, Manufacturing Batches).
5. **Advanced Analysis (Power BI DAX & MySQL SQL):** Complex calculations (time intelligence, KPIs) were implemented using DAX. Strategic insights were derived by integrating advanced SQL queries directly from Power BI to leverage MySQL's processing power (e.g., for YoY trend analysis with window functions).
6. **Dashboard Development:** Multi-page, interactive dashboards were designed to visualize key performance indicators and provide actionable business insights.

## Challenges & Solutions

Throughout this project, several real-world data and integration challenges were encountered, each providing valuable learning experiences in data problem-solving:

- **Mixed Date Formats in Source CSVs:**
- **Challenge:** Dates (e.g., SaleDate, ManufacturingDate) appeared in inconsistent formats (e.g., "YYYY-MM-DD", "DD/MM/YYYY", "MM-DD-YYYY") within the same column, preventing direct type conversion.
- **Solution:** Implemented robust try Date.FromText...otherwise logic in Power Query's M language, systematically attempting parsing with different formats until successful. For MySQL import, used VARCHAR import followed by STR\_TO\_DATE with CASE statements based on REGEXP patterns.
- **Learning:** The importance of explicit format handling and try...otherwise for data quality, and adapting cleaning methods to the specific tool (Power Query vs. SQL).
- **Stubborn Power BI Date Hierarchy Recognition:**
- **Challenge:** Despite correct data types and relationships to a Calendar table, the built-in date hierarchy for ManufacturingDate would not appear in visuals, especially after connecting the Calendar table.
- **Solution:** Diagnosed as an interaction with Power BI's "Auto date/time" feature and potential metadata caching. The final fix involved ensuring ManufacturingDate was correctly typed in Power Query, rebuilding relationships, and implementing a workaround using a DAX calculated column (Manufacturing YearMonth) combined with SORTBYCOLUMN to force correct chronological sorting for the X-axis.
- **Learning:** Deep dive into Power BI's internal date handling, the nuances of SORTBYCOLUMN for text-based date axes, and developing effective workarounds when standard methods fail.
- **MySQL Connectivity & Data Type Mismatches:**
- **Challenge:** Persistent "component missing" errors for MySQL connection, and "data truncated" or "incorrect decimal value" errors during CSV imports into MySQL for Revenue and SaleDate.

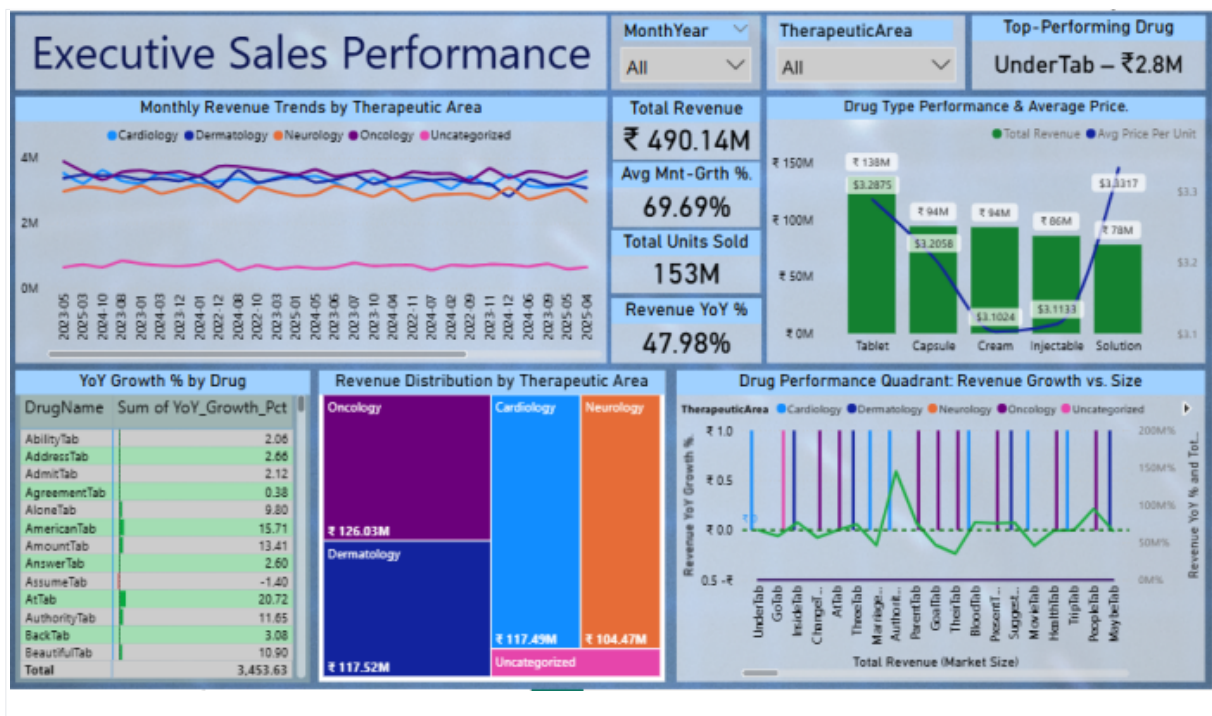
- **Solution:** Systematically troubleshooted MySQL ODBC driver installation (64-bit), ensuring correct `mysql_native_password` authentication. For import errors, adapted strategy to import problematic columns as `VARCHAR`, followed by post-import SQL cleaning using `CAST`, `NULLIF`, `STR_TO_DATE`, and `REGEXP` for robust type conversion and handling of empty strings/inconsistent formats.
- **Learning:** Comprehensive understanding of database connectivity (ODBC vs. .NET), best practices for importing messy data into relational databases, and advanced SQL for in-database data cleansing.
- **DAX Measure Discrepancies & Reconstruction:**
- **Challenge:** Measures disappearing due to Power Query changes, and initial confusion between summing pre-aggregated `MonthlyRevenue` (from SQL) vs. detailed `Revenue` (from transaction table), leading to vastly inflated totals.
- **Solution:** Recreated all lost DAX measures from scratch. Precisely defined measures to use appropriate source tables (detailed transaction vs. aggregated trend tables) for correct context. Understood the pitfalls of summing already summarized data.
- **Learning:** DAX context transition mastery, careful measure definition, understanding the interplay between Power Query output and DAX calculation context, and the importance of rebuilding logically after schema changes.

## Dashboard Screenshots

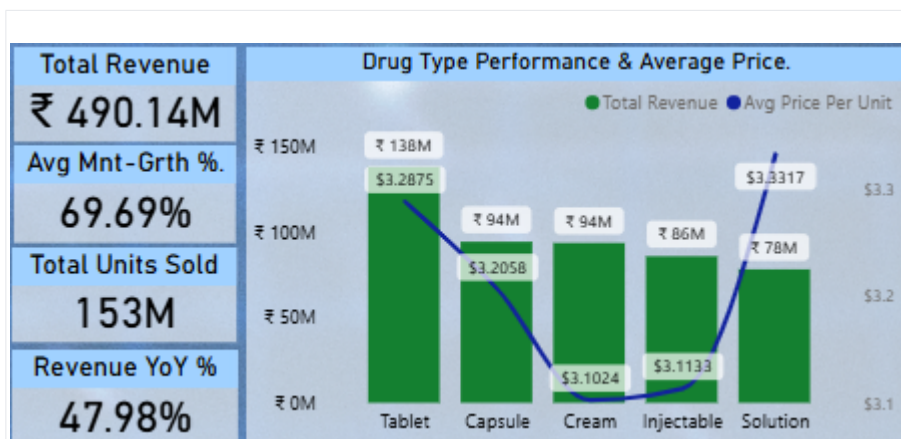
Below are screenshots of the completed Power BI dashboard, showcasing its interactive features and comprehensive insights.

### Page 1: Executive Sales Performance

This page provides an overview of sales trends, drug performance, and revenue distribution, allowing for quick identification of top performers and strategic growth opportunities.



## "Drug Type Performance & Opportunity Chart"





This page dives into operational efficiency, tracking batch success rates, identifying failure root causes, and analyzing production volume trends for quality improvement initiatives.



"Drug Failure Heatmap"

