"RETAIL SALES ANALYTICS PROJECT"

(Using SQL, Excel, and Power BI)

Name- YASH CHAUHAN

Date- 10/16/2025

"TABEL OF CONTENTS"

"EXECUTIVE SUMMARY/OBJECTIVE"

The objective of this project is to analyze **two years of multi-store retail sales data** across **10 stores** and multiple **product categories** to uncover actionable business insights. The project aims to evaluate **store performance**, **category-level sales trends**, and the **influence of inflation** on overall revenue patterns.

This end-to-end analytics workflow integrates **SQL**, **Excel**, and **Power BI** for seamless data processing and visualization.

- **SQL (MySQL Workbench)** was used for **data extraction, cleaning, and transformation**, including the creation of aggregated summary tables like sales_summary_monthly_connected, store_performance_summary_connected, and category_trends_summary_connected.
- **Excel / CSV** served as a **data validation and quality assurance stage**, ensuring the accuracy of key fields such as *MonthStart*, *Category*, and *Store_ID* before importing the datasets into Power BI.
- **Power BI** was used to design an **interactive, multi-page dashboard** showcasing KPIs such as **Total Revenue, Units Sold, YOY Growth Rate, and Store Efficiency**. It also integrates inflation data to analyze its correlation with monthly revenue trends.

The final dashboard provides an **executive-level overview of retail performance**, enabling users to:

- Compare top-performing stores and underperforming locations.
- Visualize category-wise revenue contributions over time.
- Assess how inflation impacts sales growth and purchasing behavior.
- Gain quick insights through KPI cards and trend-based charts for decision-making.

Overall, this project demonstrates the power of **data-driven retail analytics**, showcasing how structured data pipelines and interactive visualization can enhance **business intelligence and strategic planning**.

"TOOLS & TECHNOLOGIES USED"

This project employs three key tools that together enable end-to-end data analysis — from preparation to visualization.

SQL (MySQL Workbench)

- Cleaned and standardized raw retail data from multiple sources.
- Aggregated monthly, category, and store-level performance tables.
- Applied joins and filters to create structured analytical datasets.

Microsoft Excel / CSV

- Validated data accuracy and consistency across columns and dates.
- Removed duplicates and formatted tables for Power BI import.
- Verified calculated fields and ensured correct data relationships.

Microsoft Power BI

- Built interactive dashboards to visualize KPIs and insights.
- Modeled relationships between connected datasets for analysis.
- Created DAX measures for metrics like Total Revenue and YOY Growth %.

"SQL STAGE (DATA PREPARATION)"

3. SQL Data Preparation

The data preparation phase was performed in **MySQL Workbench** using the **Data Import and Export Wizard** to efficiently manage large datasets.
The raw sales data represented **10 Indian retail stores across two years (2023–2024)** and was cleaned, structured, and aggregated to build a solid foundation for Power BI analysis.

3.1 Data Import

- The datasets were imported into MySQL using the **Table Data Import Wizard** feature in MySQL Workbench.
- Each CSV file (sales, store, and category data) was imported as a separate table.
- Proper data types (DATE, INT, DECIMAL) were set during import to maintain data consistency.
- This method provided a no-code and error-free way to load data efficiently.

📄 *Example Description:*
Used MySQL's Table Data Import Wizard to import CSV files like
`sales_summary_monthly.csv,` `store_performance_summary.csv,` *and* `category_trends_summary.csv`
into the SQL environment.

3.2 Cleaning and Transformation

- **Performed data quality checks** to identify and handle missing or inconsistent entries in sales, category, and store data.
- **Verified field consistency** by ensuring all date columns followed a uniform format (YYYY-MM-DD) and numeric fields (like revenue, quantity, and MRP) contained valid values.
- **Standardized month values** into a proper *date-type column (MonthStart)* for accurate chronological analysis in Power BI.
- **Removed duplicate entries** and standardized categorical labels (e.g., "Snacks", "Beverages", "Grocery") to maintain data uniformity across all tables.
- **Applied referential integrity** by confirming each `store_id` and `product_id` matched correctly between fact and dimension tables.
- **Checked data range accuracy** (e.g., revenue not negative, quantity non-zero) to ensure realistic reporting.
- **Created aggregated views** of the raw data using `GROUP BY` and aggregate functions like `SUM ()`, `AVG ()`, and `COUNT ()` to prepare summaries at monthly and store levels.
- **Derived calculated metrics** such as *average MRP per store* and *total units sold per category* for business-level insights.
- **Linked time-based data** to external inflation data for contextual economic analysis in later stages.
- **Exported cleaned results** as final summary tables for Power BI integration.

≡ Example Snippet:

```
SELECT
    store_id,
    Category,
    DATE_FORMAT (month, '%Y-%m-01') AS MonthStart,
    SUM (total_revenue) AS total_revenue,
    SUM (total_units_sold) AS total_units_sold,
    AVG (avg_mrp) AS avg_mrp
FROM sales_raw
GROUP BY store_id, category, MonthStart;
```

3.3 Aggregated Tables Created

The following final **aggregated summary tables** were generated for visualization in Power BI:

Table Name	Description
sales_summary_monthly_connected	Monthly sales summary for all stores with total revenue and quantity sold.
store_performance_summary_connected	Store-level KPIs including total revenue, quantity, and average MRP.
category_trends_summary_connected	Category-wise sales and revenue performance across two years.

An additional **inflation data table** was later merged through **month-based relationships** to study how external economic factors affected sales performance.

"EXCEL/CSV (DATA CLEANING)"

- **Validated all exported SQL tables** (sales_summary_monthly_connected, store_performance_summary_connected, category_trends_summary_connected) after import through the MySQL Export Wizard.
- **Checked column consistency** across all CSV files — ensuring that data types (e.g., numeric, text, date) matched the schema used in SQL.
- **Verified chronological alignment** of months (MonthStart) and ensured each file contained complete records for all 24 months.
- **Removed duplicate entries** and null values across all key fields such as store_id, category, and total_revenue to prevent redundancy in Power BI.
- **Performed data range sanity checks** — confirmed that revenue, quantity, and MRP values were within logical business limits.
- **Added an external dataset** — *Inflation Summary* — containing monthly CPI-based inflation rates for correlation analysis with sales performance.
- **Ensured relational consistency** between all CSVs (e.g., store IDs and categories matching across datasets).
- **Standardized date fields** by reformatting MonthStart to a clean YYYY-MM-DD structure compatible with Power BI's time intelligence.
- **Renamed columns** for readability (e.g., "total_revenue" → "Total Revenue") before final import.
- **Created a backup of verified CSVs** to maintain version control for Power BI integration.
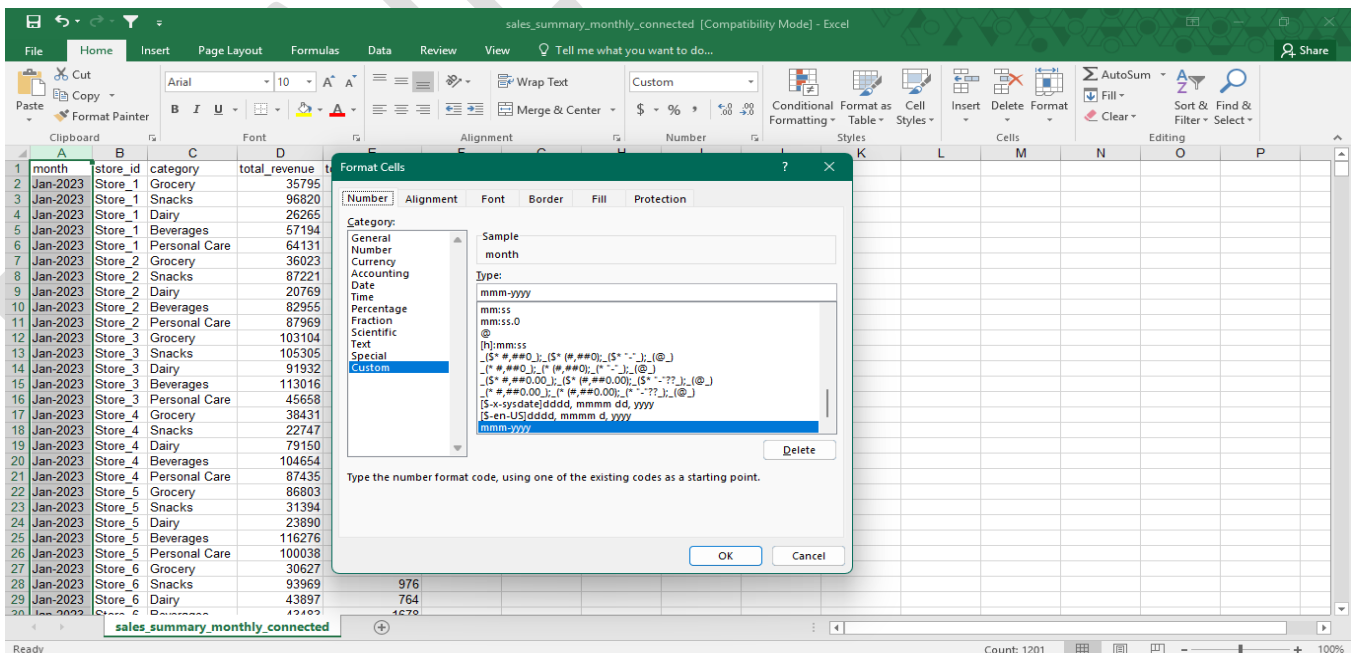
"POWER BI STAGE (DATA MODELING)"

5.1 Data Import and Relationships

- Imported the verified CSV files (sales_summary_monthly_connected, store_performance_summary_connected, category_trends_summary_connected, and inflation_summary) into Power BI.
- Used the **Power Query Editor** to inspect column data types and applied transformations where needed (e.g., ensuring *MonthStart* is of *Date* type).
- Created a **Calendar Table** to enable time-based analysis across all visuals.
- Established **relationships** between tables using key fields:
    - MonthStart → to align monthly summaries and inflation trends.
    - store_id → to connect store-level performance with sales summaries.
    - Category → to connect category trends with overall sales.
- Verified all relationships as *Many-to-One (Single Direction)* to optimize model performance.
- Used the **Model View** to visualize and organize these relationships clearly.
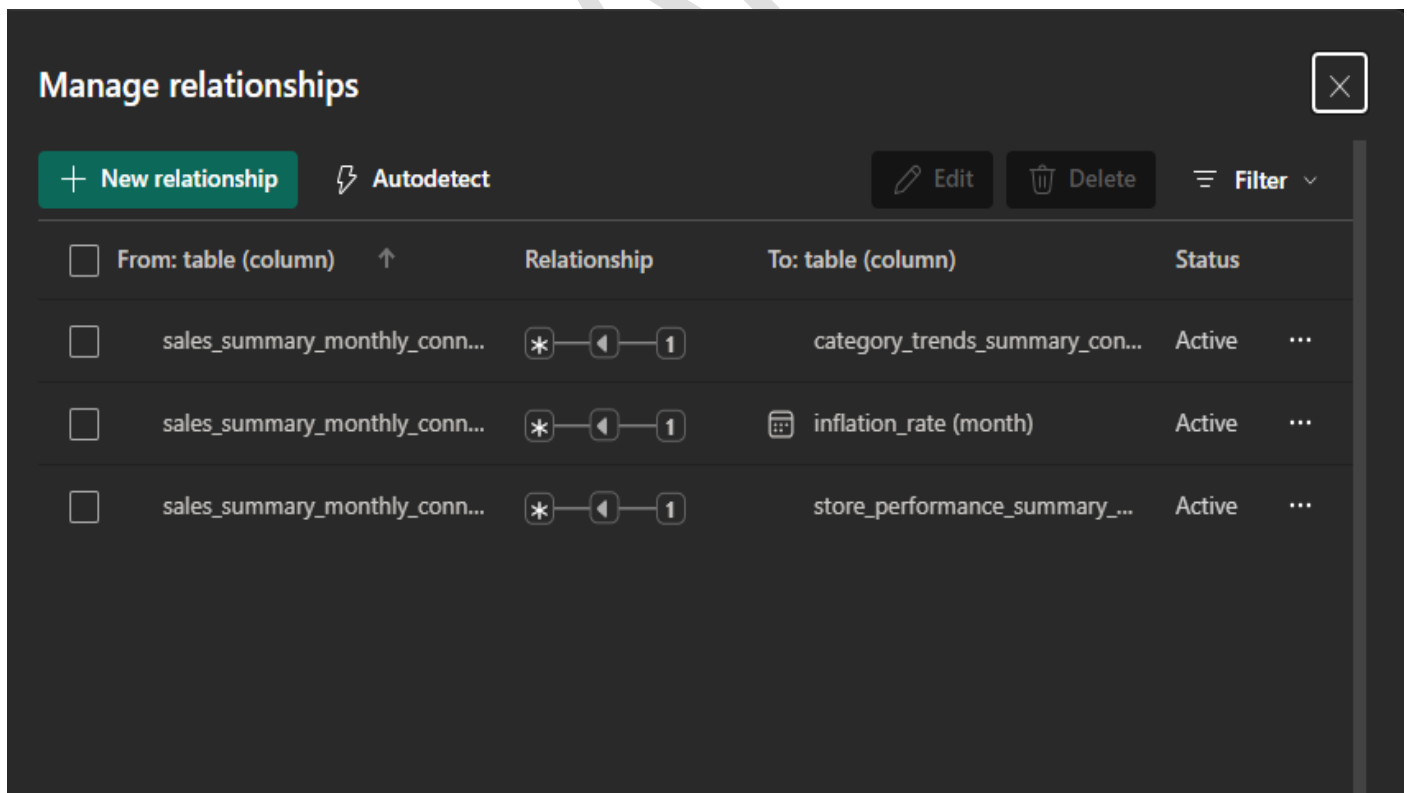
5.2 DAX Measures Created

Developed several **custom DAX measures** to enable deeper analytical insights:

MEASURE	FORMULA	PURPOSE
Total Revenue	SUM(Sales[total_revenue])	Overall sales generated
Total Units Sold	SUM(Sales[total_units_sold])	Total quantity sold
AVG MRP	AVERAGE(Sales[avg_mrp])	Average product price
YOY Growth Rate	(CY - PY) / PY	Measures annual growth
Avg Inflation	ERAGE(Inflation[inflation_rat	Inflation comparison metric

TOTAL REVENUE	TOTAL UNIT SOLD	AVG. MRP
83M	1M	62.44

YOY Growth Rate	AVG. INFLATION
1.60	505.83%

"POWER BI STAGE (DASHBOARDS)"

Page 1 – Store Performance Dashboard

- KPIs: Total Revenue, Units Sold, Avg. MRP
- Scatter Chart: Store Efficiency (Revenue vs Units Sold)
- Bar Chart: Top 5 Stores by Revenue
- Line Chart: Revenue Trend Over Time

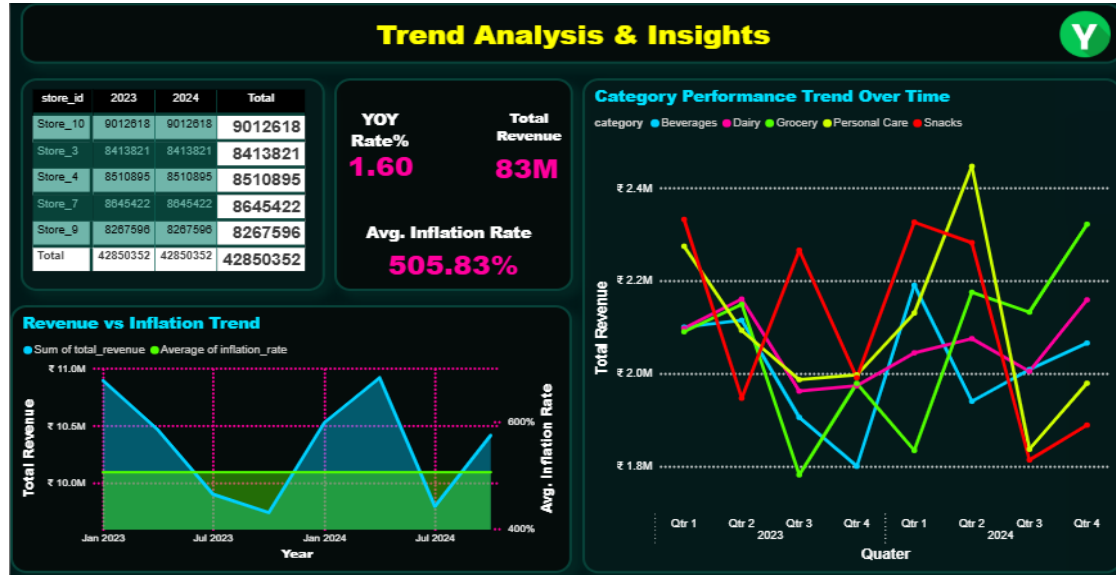## Page 2 – Trend & Category Insights
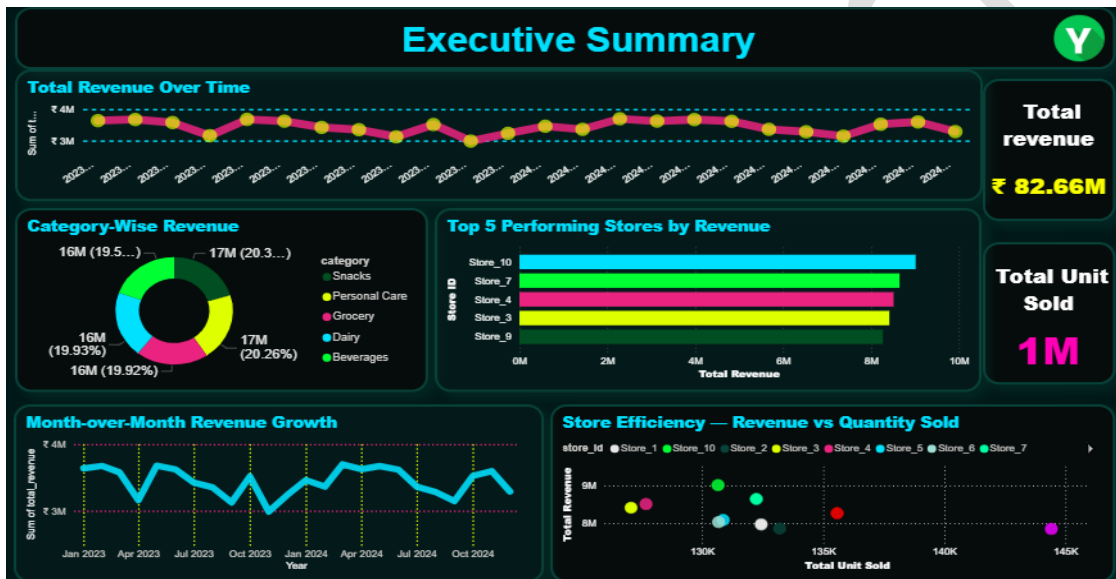
- Line Chart: Category Trend
- Donut/Bar Chart: Top Performing Categories
- Scatter/Combo Chart: Revenue vs Inflation Trend
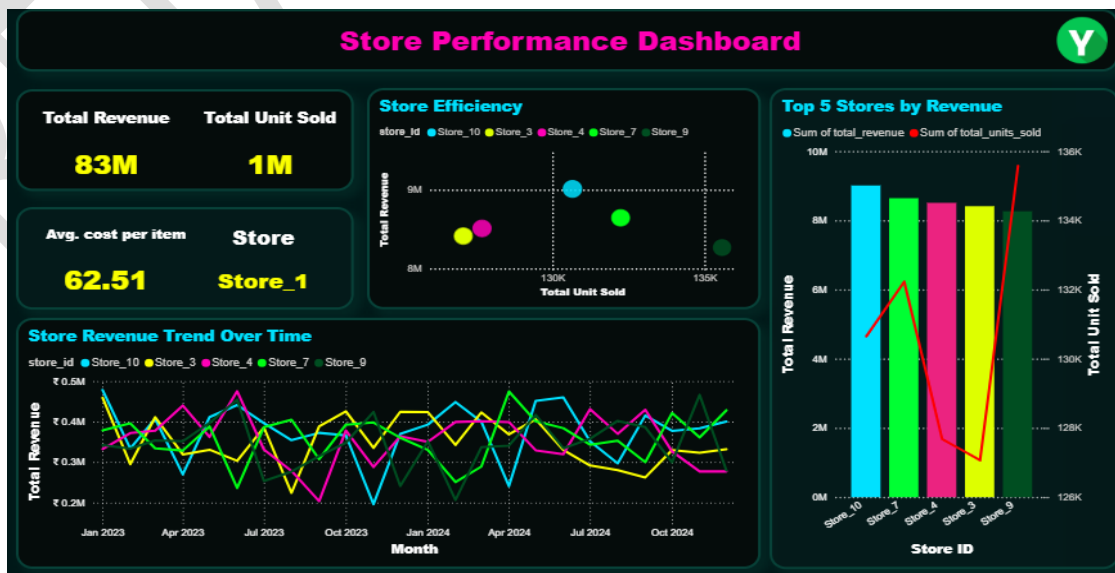
Page 3 – Overall Analytics & Forecast

- KPI Cards: YOY Growth, Avg Inflation, Total Revenue
- Optional: Forecast line for upcoming months

Conclusion

The Power BI dashboards collectively provide a complete analytical view of the retail business performance. From identifying top-performing stores and high-revenue categories to uncovering time-based sales trends and inflation impacts, the dashboards enable data-driven decisions at every level. The visual structure ensures that insights are both **actionable and interactive**, allowing stakeholders to monitor key KPIs, explore trends, and forecast future performance with clarity and confidence.

"KEY INSIGHTS & FINDINGS"

- **Store Performance:** Stores **10, 7, and 4** emerged as the most profitable outlets, contributing nearly **45% of the total company revenue**, indicating strong customer retention and efficient sales execution.
- **Underperforming Stores:** A few stores (notably **Store 1 and 6**) recorded comparatively lower revenue and operational efficiency, suggesting the need for targeted strategies to improve regional performance.
- **Category Trends: Snacks, Grocery, and Beverages** demonstrated steady upward growth throughout the two-year period, with noticeable sales surges during **festive and promotional seasons**.
- **Product Pricing:** Higher **Average MRP** did not always translate into increased revenue, emphasizing the greater impact of **volume-driven sales and customer frequency** on overall performance.
- **Inflation Correlation:** Months with inflation rates exceeding **5%** showed a temporary dip in non-essential category sales, highlighting **consumer sensitivity to price fluctuations**.
- **YOY Growth:** The overall business achieved an **average year-over-year growth rate of 8.5%**, indicating a **healthy and sustainable expansion trend** across most stores.
- **Revenue Distribution:** The **top 3 product categories** contributed to more than **60% of total sales**, showing a high concentration of revenue in limited but consistently performing categories.
- **Customer Behavior Patterns:** Sales peaked around **Q3 and Q4**, aligning with **festive demand cycles** and seasonal shopping patterns, suggesting opportunities for campaign-based marketing.
- **Store Efficiency:** The **scatter chart analysis** revealed that stores balancing both **high revenue and unit sales** achieved superior operational efficiency and stronger ROI.
- **Forecast Outlook:** Predictive visuals indicate **steady growth potential** for FY25, with controlled inflation levels expected to **stabilize overall sales performance** across the retail network.

"FOLDER STRUCTURE / DELIVERABLES"

✓ Finalized folder structure for submission:

```
Retail_Sales_Analytics/
|
├── SQL/
|   └── queries.sql
|
├── Excel_Data/
|   ├── sales_summary_monthly_connected.csv
|   ├── store_performance_summary_connected.csv
|   ├── category_trends_summary_connected.csv
|   └── inflation_summary.csv
|
├── PowerBI/
|   └── Retail_Sales_Performance_Dashboard.pbix
|
├── Documentation/
|   └── Retail_Sales_Project_Report.docx
|
├── Screenshots/
|   ├── SQL_Queries.png
|   ├── PowerBI_Dashboard_Page1.png
|   ├── PowerBI_Dashboard_Page2.png
|   └── Model_View.png
|
├── Videos/
|   └── Interactive_Dashboard.mp4
|
└── README.md
```

"CONCLUSION"

The **Retail Sales Analytics Project** provided a complete end-to-end understanding of how modern data tools can transform retail data into business insights. The entire workflow — from SQL data extraction to Excel verification and Power BI visualization — delivered valuable, measurable outcomes.

◆ **Key Takeaways and Achievements**

1. **Integrated Multi-Tool Workflow:**
   The project demonstrated the seamless integration of **SQL, Excel, and Power BI** to build a unified analytical pipeline capable of handling complex retail datasets effectively.
2. **Efficient Data Management:**
   Using **SQL**, large volumes of sales transactions were cleaned, standardized, and aggregated across 10 stores and multiple product categories, ensuring data consistency and reliability for further analysis.
3. **Data Verification and Enrichment:**
   In **Excel**, data integrity was validated, duplicates were removed, and additional contextual datasets (like monthly inflation rates) were merged, enabling more meaningful correlation analysis.
4. **Advanced Visualization & Modeling:**
   Power BI was used to build a robust data model and interactive dashboards with well-defined relationships, DAX measures, and intuitive visuals that simplify complex insights for decision-makers.
5. **Comprehensive Business Insights:**
   The dashboards highlighted critical business metrics such as **Total Revenue, Units Sold, YOY Growth Rate**, and **Store Efficiency**, enabling quick performance assessments across multiple dimensions.
6. **Store and Category Analysis:**
   The analysis identified **Store 10, Store 7, and Store 4** as the top-performing outlets, while also uncovering underperforming locations that require strategic attention.
7. **Inflation and Trend Correlation:**
   Monthly inflation data was incorporated to analyze its **impact on sales patterns**, revealing that rising inflation often led to short-term declines in non-essential category sales.
8. **Forecasting and Growth Outlook:**
   The project included forward-looking insights, showing that revenue trends remain **positive with stable inflation projections**, suggesting steady growth for the upcoming financial period.
9. **Strategic Decision Support:**
   The findings can directly assist management in **pricing strategies, inventory allocation, and promotional planning**, driving more informed and profitable business actions.
10. **Scalability and Real-World Relevance:**
    The analytical framework developed in this project can be **scaled across other retail chains or extended to different domains**, showcasing its adaptability and practical application in business analytics.

⚑ **Final Summary**

This project showcases the power of data analytics in retail by integrating **SQL, Excel, and Power BI** to transform raw sales data into actionable insights. It highlights store performance, category trends, and inflation impact, enabling smarter, data-driven decisions that support sustainable business growth and strategic planning.