Sales Analysis Dashboard - Overview

1. Objective

The goal of this project was to build a **3-page interactive Sales Analysis Dashboard in Power BI** that helps business stakeholders understand:

- Key revenue drivers
- Product profitability
- Customer value
- Regional performance
- Operational inefficiencies (delivery delays)
- Customer retention trends

This dashboard supports **faster decisions, better insights, and data-driven strategy**.

2. Dataset

Global Superstore Dataset (Kaggle)

Contains:

- Orders
- Customers
- Products
- Categories & Sub-categories
- Regions / States
- Sales, Profit, Quantity
- Discounts
- Shipping dates & modes

3. Key Business Questions Solved

The dashboard answers the following real-world questions:

Sales & Product

1. Which products and categories drive the most sales and profit?
2. How do sales and profit trend over time?

Customers

1. Which customer segments are most valuable?

2. What is customer repeat behavior and retention?

Regions

1. Which regions/states underperform or outperform?

Operations

1. What operational issues (shipping delays) affect customer satisfaction?

Leadership Insights

1. What are the top actionable insights for business leaders?

4. Assumptions

- Shipping Delay = **Ship Date – Order Date**
- Profit Margin % = **Profit / Sales**
- Avg Order Value = **Total Sales / Total Orders**
- Returning customer = customers with **>1 order**
- Missing numeric values were filled with 0 when appropriate, or removed if they could not be used reliably
- Date dimensions follow standard calendar year-month format

5. Data Cleaning (Power Query)

- Removed rows with no Order ID or Customer ID
- Blank numeric cells replaced with 0
- understood data using column quality, column profile and column distribution

6. SQL Transformations

Performed before loading into Power BI:

- **Created database and staging table (stg_superstore)** using SELECT INTO
- **Removed duplicate records** using ROW_NUMBER() OVER (PARTITION BY Order_ID, Product_ID ORDER BY Row_ID)
- **Cleaned and standardized text fields** using LTRIM(), RTRIM(), and UPPER() for customer, region, state, city, and product names
- **Created computed columns** for business logic:
    - ShippingDelay (Ship_Date – Order_Date)
    - ProfitMargin (Profit / Sales, with divide-by-zero handling)

    - OrderYear, OrderMonth, OrderQuarter
- **Validated date quality** by checking for invalid rows where Ship_Date < Order_Date
- **Ensured consistent datatypes** for all columns using INFORMATION_SCHEMA.COLUMNS
- **Maintained clean backup and staging layers** to separate raw data from cleaned data

7. Data Modeling (Star Schema)

Fact Table

stg_superstore

Dimension Tables

- **Dim Customers**
- **Dim Date**
- Dim Products

Relationships

- 1–Many: DimDate → stg_superstore
- 1–Many: DimCustomer → stg_superstore
- 1–Many: DimProduct → stg_superstore

8. Key DAX Measures

Core Metrics:

Total Sales = SUM(Orders[Sales])

Total Profit = SUM(Orders[Profit])

Profit Margin % = DIVIDE([Total Profit], [Total Sales])

Customer Count = DISTINCTCOUNT(Orders[Customer ID])

Avg Shipping Days = AVERAGE(Orders[Shipping Delay])

Avg Order Value = DIVIDE([Total Sales], COUNTROWS(Orders))

Returning Customers = COUNTROWS(FILTER(ADDCOLUMNS('Dim Customer',"Order Count", CALCULATE(COUNTROWS(stg_superstore))),[Order Count]>1))

```
Retention Count =
CALCULATE(
  DISTINCTCOUNT('Dim Customer'[Customer_ID]),
  FILTER(
    ALL(stg_superstore),
    stg_superstore[Order Year Month] = SELECTEDVALUE('Dim Date'[YearMonth])),
  FILTER(
    ALL('Dim Customer'),
    'Dim Customer'[Cohort Month] = SELECTEDVALUE('Dim Customer'[Cohort Month])))
```

Time Intelligence:

```
Rolling 3M Sales = CALCULATE([Total Sales], DATESINPERIOD(DimDate[Date],
MAX(DimDate[Date]), -3, MONTH))
```

Product Page (Top N Filter):

```
Rank = RANKX(ALLSELECTED(stg_superstore[Product_Name]),[Product Sales],,DESC)
Is In TopN =
VAR SelectedTopN = SELECTEDVALUE('TopN'[TopNValue], 10)
RETURN
IF(SelectedTopN = 0, 1, IF([Rank] <= SelectedTopN, 1, 0))
```

9. Dashboard Pages

Page 1 — Sales Overview

Contains:

- Six KPI Cards
- Monthly Sales Trend (3M rolling average)
- Sales by Category
- Top 5 Products
- Sales by State Map
- Segment × Category Profit Matrix
- Quick Insights Box

Insights:

- Technology is top profit driver
- Furniture → lower margins
- Avg Shipping Delay around 4 days

Page 2 — Product Performance

Contains:

- Product slicer + Top N selector
- Pareto Chart (Sales + Cumulative %)
- Sales vs Profit Scatter Plot (quantity = bubble size)
- Detailed product-level table with sparklines
- Recommendations card

Insights:

- 20% of products drive ~80% of sales
- Some high-volume products generate negative profit

Page 3 — Region & Customers

Contains:

- Returning Customer %
- Avg Order Frequency
- Avg Discount
- Avg Delivery Delay
- Region Profit Bars
- Delivery Delay Heatmap
- CLV Bucket Chart
- Cohort Retention Matrix
- Quick Insights Panel

Insights:

- High delivery delays in Standard Class
- Medium CLV group offers biggest uplift potential
- Early retention drops fast → need onboarding interventions

10. Limitations

- Dataset is simulated
- No COGS → cannot compute true gross margin
- No marketing data
- Static data: no real-time refresh
- No return data

11. Future Enhancements

- Add **Churn Prediction** model
- Add **Sales Forecasting (ML)**
- Add **RFM Segmentation**
- Build **What-If Pricing Simulator**
- Add **Mobile layout**
- Include **Inventory analytics**

12. Final Summary

This project demonstrates:

- End-to-end BI development (SQL → Power Query → DAX → Visualization)
- Clean modeling using Star Schema
- Insight-driven storytelling
- Ability to answer real business questions

This Sales Dashboard equips stakeholders with a **overview of performance** and enables **data-driven decisions across sales, operations, customer behavior, and product strategy**.