# Data Driven Insights Report

Name : Mohammed Taqi Uddin Omer

Role : Business Analyst Intern

Date : 11/07/2025

## 1. Executive Summary

This project utilized a synthetic retail sales dataset consisting of 1,000 records to demonstrate a complete data-driven insights workflow — from data ingestion and cleaning to exploratory analysis, visualization, and actionable business recommendations.

Key Outcomes

- Conducted thorough data cleaning and aggregated sales across dimensions such as month, product, region, customer, and sales channel.
- Developed four key visualizations:
    1. Monthly Revenue Trend
    2. Monthly Profit Trend
    3. Top 10 Products by Revenue
    4. Regional Revenue Distribution
- Compiled a comprehensive analytical report with data-backed business insights addressing areas such as inventory optimization, marketing strategy, regional expansion, and customer retention.

---

## 2. Dataset Overview

The dataset simulates transactional retail sales data between 2023 and 2024, representing a diverse mix of customers, products, and regions.

Columns Description

- order_id: Unique identifier for each transaction
- date: Order date (YYYY-MM-DD)
- customer_id: Anonymized customer identifier
- product: Product SKU (30 unique products)
- category: Product category (Electronics, Home, Clothing, Sports, Beauty)
- region: Geographic region (North, South, East, West)
- quantity: Number of units sold
- unit_price: Price per unit (in currency)
- discount_pct: Discount percentage applied (0–20%)
- discount: Discount amount
- revenue: Net revenue after discount (quantity * unit_price - discount)
- cost: Estimated product cost to company

- profit: Profit margin (revenue - cost)
- payment_method: Mode of payment (Card, UPI, NetBanking, Cash, Wallet)
- channel: Sales channel (Online or Offline)

---

## 3. Data Cleaning and Preparation

Process Summary

1. Data Type Standardization:
   Converted date fields to datetime and ensured all numeric columns (e.g., revenue, cost, profit) were properly formatted.
2. Derived and Verified Fields:
   Verified consistency between revenue, quantity, and unit_price. Created derived columns as needed to maintain data integrity.
3. Handling Missing Values:
   No significant missing data was detected. In practical settings, missing numeric values would be imputed conservatively (zeros or mean values), with affected fields flagged for review.
4. Aggregation Keys:
   Added a month field (date truncated to month) to support time-series aggregations.

Why this step matters:
Ensuring data integrity before analysis prevents aggregation errors and ensures accurate and meaningful visualization results.

---

## 4. Exploratory Data Analysis

The following key metrics were computed to understand the business performance:

- Total Revenue — cumulative sales over the period
- Total Profit — overall profit generated
- Total Orders — count of unique transactions
- Average Order Value (AOV) — mean revenue per order
- Monthly Revenue and Profit — time-series trends
- Top Products by Revenue — product-level contribution analysis
- Top Customers by Revenue — most valuable customers
- Revenue by Region and Channel — geographic and sales channel distribution

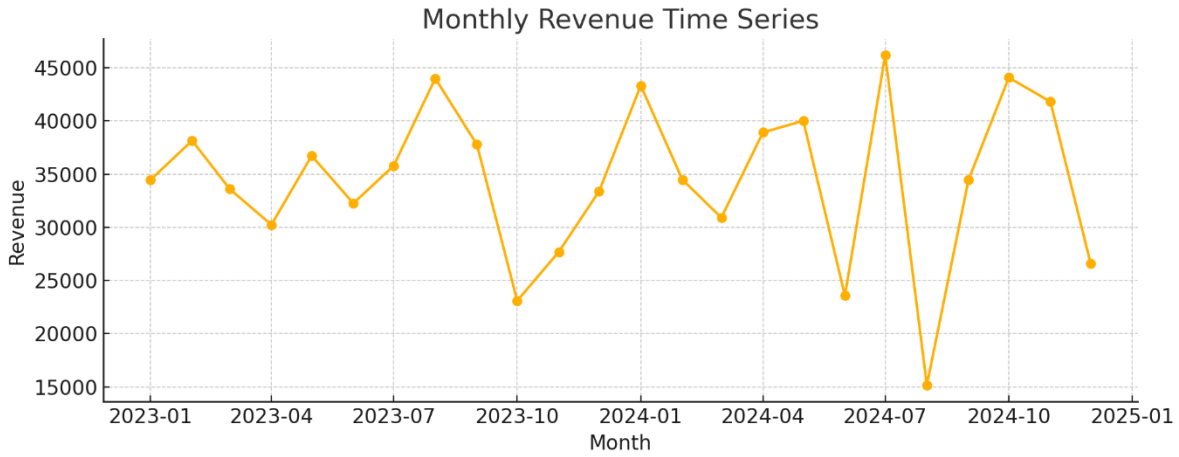
---

## 5. Visualizations

The following visualizations were produced to support key insights.

5.1 Monthly Revenue Trend

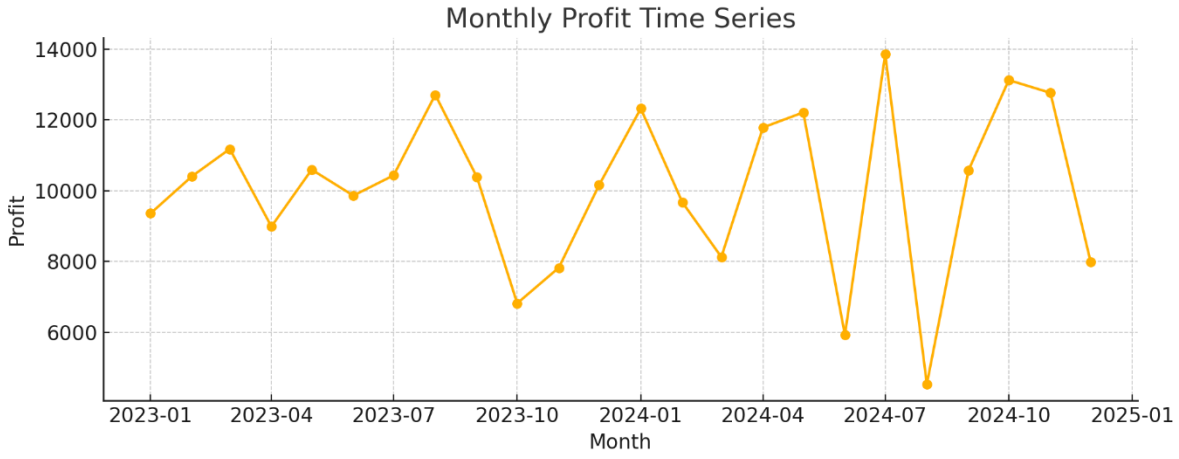
Purpose: Illustrates monthly fluctuations and seasonality in revenue.




---

## 5.2 Monthly Profit Trend

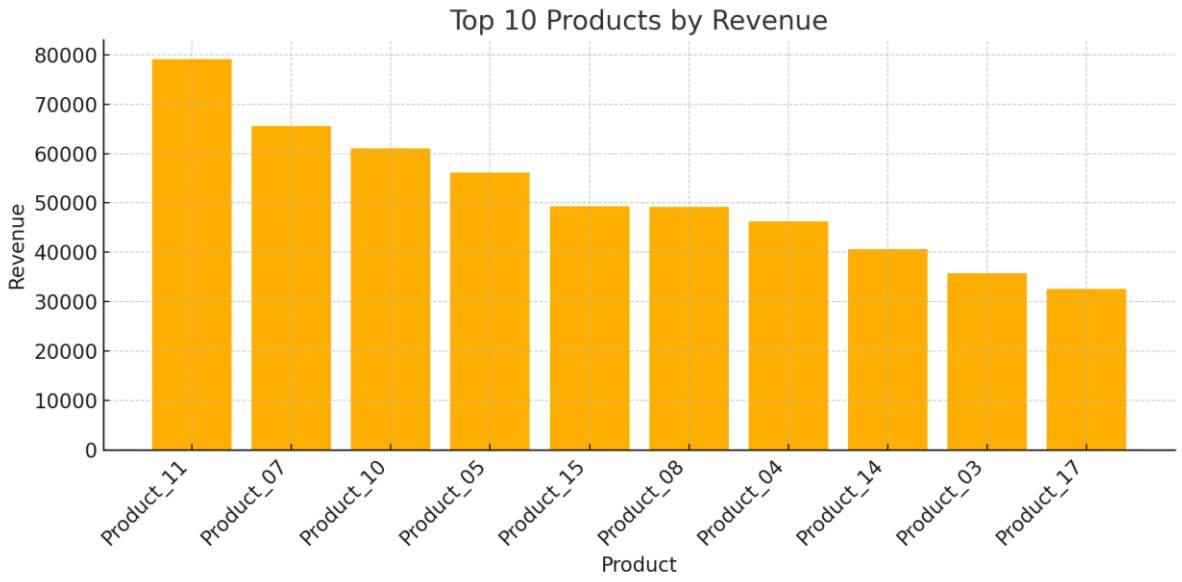
Purpose: Tracks profitability independent of revenue to highlight changes in margin performance.




---

5.3 Top 10 Products by Revenue

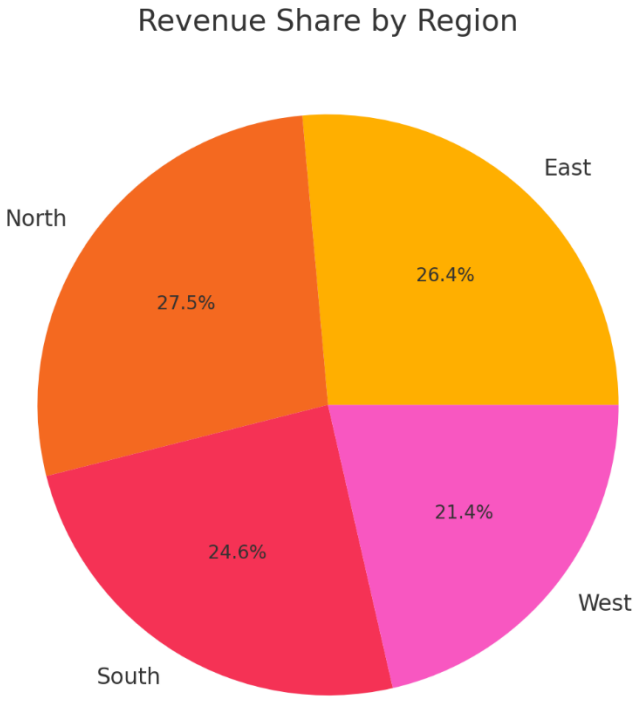
Purpose: Identifies the highest-grossing products contributing most to sales.




---

## 5.4 Revenue Share by Region

Purpose: Displays the distribution of total revenue across regions to identify strong and weak market areas.




---

## 6. Key Findings

1. Seasonality in Revenue:
   Revenue varies across months, indicating potential seasonal trends. Aligning marketing and inventory efforts with these peaks could maximize sales.
2. Profit Margin Fluctuations:
   Profit trends do not perfectly follow revenue trends, implying that discounting or cost variations influence margins.
3. Product Concentration:
   A small subset of products drives a major portion of revenue, reinforcing the 80/20 rule (Pareto Principle).
4. Regional Disparity:
   Revenue is unevenly distributed across regions, signaling potential growth opportunities in weaker markets.
5. High-Value Customers:
   A few customers contribute disproportionately to overall revenue, suggesting the need for targeted loyalty and retention programs.

---

## 7. Business Insights and Recommendations

Inventory & Supply Chain

- Prioritize restocking top-selling products before expected peak seasons.
- For long-lead-time products, initiate orders early to prevent stockouts.

Marketing & Promotions

- Schedule promotions strategically before high-demand months.
- Deploy localized marketing campaigns to boost underperforming regions.

Pricing & Profitability

- Review discount policies if profit margins decline despite high sales.
- Focus marketing on higher-margin products and consider bundled offers.

Customer Retention & Engagement

- Launch loyalty programs for top-spending customers.
- Implement RFM (Recency, Frequency, Monetary) analysis for personalized outreach.

Data Quality & Measurement

- Enforce mandatory data capture (e.g., product, price, region).
- Introduce campaign tracking (UTMs) to measure marketing ROI.

Advanced Analytical Opportunities

- Conduct cohort analysis to study customer retention patterns.
- Apply RFM segmentation for targeted communication.
- Use A/B testing to measure promotional effectiveness.
- Implement forecasting models (ARIMA, Prophet) for demand prediction.

---

## 8. Limitations

- Synthetic Data: The dataset used is simulated; while realistic, results should be validated with live transactional data.
- Estimated Costs: The cost field was modeled; actual COGS data would refine profit accuracy.
- Single Data Source: Integration of inventory, returns, and marketing data would allow for deeper causal insights.

---

## 9. Appendix — Outputs and Reference Tables

Sample Python Code (Pandas Snippets)

```
# Load dataset
df = pd.read_csv("sample_retail_1000.csv", parse_dates=['date'])


# Compute monthly revenue
df['month'] = df['date'].dt.to_period('M').dt.to_timestamp()
monthly_rev = df.groupby('month')['revenue'].sum().sort_index()


# Identify top 10 products
top_products = df.groupby('product')['revenue'].sum().sort_values(ascending=False).head(10)


# Compute regional revenue
region_rev = df.groupby('region')['revenue'].sum()
```