Sales Data Analysis Report

Submitted by: Pragya Gupta
PGDM – Research & Business Analytics

Abstract

This project focuses on analyzing sales performance data from an Excel workbook (Sales_Data_Analysis.xlsm). The objective is to extract key performance indicators (KPIs), identify trends, and highlight the contribution of products, categories, regions, and customers. The methodology includes dataset profiling, data preprocessing, KPI computation, monthly revenue aggregation, and segmentation analysis. The project delivers insights into business performance, data quality considerations, and recommendations for decision-making.

1. Introduction

Sales data analysis plays a crucial role in understanding business performance, monitoring growth, and improving strategic decision-making. By tracking revenue, quantity sold, and gross margin across time and business segments, companies can identify key growth drivers and address underperforming areas.

The provided Excel workbook contained multiple sheets of raw sales records. This project systematically processes and summarizes the dataset to compute relevant metrics, uncover patterns, and generate actionable insights.

2. Dataset Description

The workbook comprised multiple sheets with varying row and column counts. The **main sheet**, selected for detailed profiling, included structured sales records. Columns identified (through heuristic detection) were:

- Date of transaction
- Quantity sold
- Revenue / Sales amount

- Price per unit
- Cost per unit (where available)
- Product / Item details
- Category / Segment
- Region / State / Market
- Customer / Client
- Sales Channel

From these fields, key metrics such as **total revenue**, **total quantity sold**, **average selling price**, and **gross margin** were derived.

3. Methodology

The following steps were followed to analyze the data:

- **Data Preprocessing:**
  Parsed sheets, standardized column names, coerced data types (date, numeric), and handled missing values.
- **Field Detection:**
  Auto-detected key business fields such as date, quantity, revenue, price, cost, product, category, region, customer, and channel.
- **KPI Computation:**
  Calculated total revenue, total quantity sold, average selling price, gross profit, and gross margin %.
- **Trend Analysis:**
  Aggregated sales data on a monthly basis to identify seasonality and growth/decline patterns.
- **Segmentation Analysis:**
  Generated top contributors by product, category, region, customer, and channel, showing where revenue is concentrated.

4. Results & Evaluation

- The dataset spanned multiple months/years, enabling **time-series trend analysis**.
- **Total revenue and total quantity** were computed successfully.
- **Average selling price** was derived, offering a benchmark for pricing analysis.
- Where cost data was available, **gross profit** and **gross margin %** were calculated.
- **Monthly revenue trends** revealed potential seasonality in sales.
- **Top contributors** by product, category, region, customer, and channel highlighted concentration of revenue among limited segments.

5. Conclusion & Insights

- Revenue trends provide valuable insights into **seasonality** and can support **demand forecasting**.
- A small set of **top-performing products and regions** accounted for a majority of revenue, suggesting areas for focused resource allocation.
- **Gross margin monitoring** helps track profitability and highlights cost-driven risks.
- **Underperforming segments** may require corrective actions or strategic review.
- Incorporating additional details such as **returns, discounts, and promotions** would strengthen KPI accuracy.

6. Limitations & Future Work

- Field detection was based on **column name heuristics**; mislabeling of columns could lead to missed metrics.
- Embedded **macros in the .xlsm workbook** were not executed; analysis was performed only on sheet data.
- **Future work** could include:

- Developing interactive dashboards (Power BI, Python-based dashboards).
- Implementing forecasting models for sales and margin predictions.
- Integrating external data (economic indicators, seasonal factors) for deeper insights.
- Automating the pipeline for **real-time monitoring**.