Sales Data Analysis & Visualization Project

Executive Summary

This project analyzed four years (2014-2017) of Superstore sales data to understand performance, identify key trends, and isolate drivers of business loss.

The analysis confirmed strong revenue growth but revealed a major vulnerability: **unsustainable deep discounting** on certain product categories.

Key Findings:

- **Growth:** Sales and Profit grew by approximately 50% from 2014 to 2017.
- **Loss Driver:** Any **discount of 30% or higher** resulted in an average net loss per transaction.
- **Worst Offenders:** The **Tables** sub-category alone accounted for over $\$-17,700$ in net losses.
- **Profit Engines: Copiers**, **Phones**, and **Accessories** are the most profitable products and should be prioritized.

Methodology and Technical Stack

This project followed a standard Data Science workflow: Data Acquisition, Cleaning/Preparation, Exploratory Data Analysis (EDA), Deep Dive Analysis, and Recommendation Generation.

Tools and Environment

Phase	Tool/Library	Purpose
IDE	VS Code (Visual Studio Code)	Integrated development environment for Python and Jupyter Notebooks.
Data Cleaning	Python (Pandas, NumPy)	Data manipulation, transformation, and numerical calculations.
Visualization	Python (Matplotlib, Seaborn)	Creation of time series and comparative charts.
Code Structure	Jupyter Notebook (.ipynb file)	Documenting code, output, and analysis in a single, reproducible file.

Data Acquisition and Preparation

- **Source:** Sample Superstore Sales Dataset (`Sample_Superstore.csv`).
- **Data Quality:** The dataset had no missing values but required type conversion.
    - **Action 1 (Data Types):** `Order Date` and `Ship Date` were converted from `object` (string) to `datetime` format.
    - **Action 2 (Cleanup):** The redundant `Row ID` column was dropped, and `Postal Code` was converted to a string (`object`).

Feature Engineering (New Metrics Created)

New columns were calculated to enrich the data for analysis:

New Feature	Formula / Derivation	Purpose
Profit Margin	(Profit / Sales) * 100	Measures efficiency and profitability of each transaction.
Order Year	Extracted from Order Date	Facilitates annual trend analysis.
Shipping Duration (Days)	Ship Date - Order Date	Measures logistics efficiency.

Exploratory Data Analysis (EDA) and Trends

Initial analysis focused on establishing the baseline performance and identifying overall patterns.

Statistical Overview

Metric	Value	Observation
Total Rows	9,994	
Date Range	2014-01-03 to 2017-12-30	4 years of data available.
Max Sale	$22,638.48	Sales data is heavily skewed by outliers (large orders).
Min Profit	**$-6,599.98**	Confirms the existence of transactions resulting in massive losses.

Time-Series Trend Analysis

- **Annual Growth:** Sales and Profit showed a consistent, powerful upward trajectory over the 4-year period.
    - **2014 Sales:** $\text{\$484,247}$
    - **2017 Sales:** $\text{\$733,215}$
- **Seasonal Pattern:** The business is highly seasonal, with the **fourth quarter (Q4: Oct, Nov, Dec)** showing massive, recurring spikes in both sales and profit across all four years.

Deep Dive: Investigation into Profit Loss

The primary objective was to diagnose the cause of negative profitability.

Profit vs. Discount Analysis

A quantitative analysis was performed to determine the threshold where discounts cease to be profitable.

Discount Rate	Average Profit per Transaction	Conclusion
0% to 20%	**Positive Profit**	Sustainable for volume/revenue generation.
30% and higher	**Negative Profit**	Unprofitable and must be reviewed.

Conclusion: The discount rate of $\mathbf{30\%}$ marks the point of diminishing return, where the increase in volume no longer compensates for the reduction in unit price.

Loss Drivers by Product Category

We isolated which product sub-categories contributed to the losses, despite the high discounts.

Product Sub-Category	Total Net Profit/Loss	Overall Category
Tables	**$-17,725.48**	Furniture
Bookcases	**$-3,472.56**	Furniture
Supplies	**$-1,189.10**	Office Supplies

Conclusion: The **Furniture** category, particularly **Tables**, is the overwhelming source of net profit loss. This indicates a likely issue with the cost of goods sold (COGS), supply chain, or pricing strategy for these specific large, bulky items.

Final Recommendations

Based on the data analysis, the following actions are recommended to maximize profitability:

1. **Implement a Strict Discount Cap: Immediately restrict all discounts to a maximum of 20%**. Discounts of 30% or more must be eliminated unless justified by extreme overstock or unique promotional circumstances.
2. **Re-evaluate Loss-Making Items:** Launch a cost-of-goods-sold (COGS) audit on **Tables** and **Bookcases**. If they remain unprofitable even with reduced discounts, consider raising the base price or removing them from the product catalog.
3. **Invest in Profit Engines:** Increase marketing and inventory investment in the highest-margin categories: **Copiers, Phones, and Accessories**.
4. **Strategic Q4 Planning:** Since the business is highly seasonal, pre-plan stock and marketing efforts to capitalize on the predictable surge in sales during October through December.