Global Superstore Analytics Project

Prepared by: Aleeza Zia

1. Executive Summary

The Global Superstore dataset (**51,290 rows × 21 columns**) was analyzed to address seven key research questions on discounts, customer segmentation, regional performance, forecasting, shipping efficiency, product profitability, and customer lifetime value.

The notebook outputs reveal:

- **Total Sales:** 12.64M
- **Total Profit:** 1.47M (**Profit Ratio = 11.62%**)
- **Leading countries:** United States, India, and China drive sales and profit.
- **Category performance:** Technology is most profitable; Furniture suffers due to losses in the Tables subcategory.
- **Customer segment:** Consumer segment dominates both sales and profit.
- **Forecasting:** Predicts consistent sales growth with Q4 peaks.
- **Shipping efficiency:** Standard Class shipping balances volume and profitability; faster modes erode margins.
- **Product profitability:** Top 10 profitable products concentrated in Technology and Office Supplies; worst 10 mainly loss-making Tables.
- **Customer Lifetime Value (CLV):** Small group of high-value customers contributes disproportionately to long-term profitability.

This project demonstrates skills in **EDA, visualization, segmentation (RFM), profitability analysis, and time-series forecasting (NeuralProphet).**

2. Research Questions

RQ1 – Discounts vs. Profitability

- How do discount levels affect profit across product categories and sub-categories?
- Are deep discounts correlated with negative profits for particular categories or segments?

RQ2 – Customer Segmentation (RFM)

- How can customers be segmented using Recency, Frequency, and Monetary (RFM) metrics?
- Which segments represent high-value, loyal, and at-risk customers?
- How should marketing and retention strategies be prioritized by segment?

RQ3 – Regional & Category Performance

- Which regions and product categories deliver the highest sales and the highest/lowest profitability?
- Which sub-categories are underperforming in specific regions and warrant further investigation?

RQ4 – Forecasting Monthly Sales

- What are the expected monthly sales for the next 12 months using a time-series model with yearly seasonality?
- What seasonal patterns and trend components should inventory and planning teams consider?

RQ5 – Shipping Mode Impact

- Which shipping modes contribute most to total profit and order efficiency?
- Are there trade-offs between speed (or cost) of shipping mode and profitability?

RQ6 – Product-Level Performance

- Which products are the top 10 most profitable and which are the 10 worst-performing products by profit?
- Do these products share characteristics (category, price point, or shipping method) that explain performance?

RQ7 – Customer Lifetime Value (CLV)

- What is the distribution of CLV across customers using AOV × Purchase Frequency × Estimated Customer Lifespan?
- Which customers should be targeted for retention investments based on CLV?

3. Findings — Question by Question

RQ1 – Discounts vs. Profitability

- Strong negative relationship found: higher discounts = lower profits.
- In Furniture → Tables, profits turn negative when discounts are steep.
- Scatterplots and trend lines confirm that deep discounts destroy margins instead of boosting volume.
- **Recruiter takeaway:** shows ability to link pricing strategy to profitability outcomes.

RQ2 – Customer Segmentation (RFM)

- Customers were segmented based on Recency, Frequency, and Monetary (RFM) metrics.
- **Loyal/High-Value group:** frequent, recent buyers with high spend.
- **At-Risk group:** older purchases, low recency, dropping frequency.
- **Low-Value/New group:** one-off or low-spend customers.

- Outputs showed clear segmentation, with loyal customers forming a core retention group for long-term profitability.
- **Recruiter takeaway:** demonstrates advanced segmentation beyond raw sales.

RQ3 – Regional & Category Performance

- United States dominates sales and profit, followed by India and China.
- **Technology:** most profitable category (Phones, Copiers, Accessories).
- **Office Supplies:** strong sales volume, moderate margins.
- **Furniture:** risk category; Tables drag down profit despite revenue.
- Subcategory outputs highlight:
  - Phones, Chairs, Binders = consistently high performers
  - Tables = consistently loss-making
- **Recruiter takeaway:** ability to use data to spot strong and weak areas geographically and by product line.

RQ4 – Forecasting Monthly Sales

- Forecast using NeuralProphet shows:
  - Steady upward sales trend over next 12 months
  - Strong Q4 peaks every year, likely linked to holiday cycles
  - Seasonality and trend captured clearly in model outputs
- **Recruiter takeaway:** applied predictive analytics (time-series modeling) to guide business planning.

RQ5 – Shipping Mode Impact

- **Standard Class:** most orders and highest profitability
- **Second Class & First Class:** moderate contributions, balanced cost vs. profit
- **Same Day:** least profitable due to higher delivery costs despite faster speed
- Analysis highlights the trade-off between customer convenience and profitability
- **Recruiter takeaway:** shows understanding of operational efficiency and margin trade-offs

RQ6 – Product-Level Performance

- Top 10 most profitable products: mostly from Technology and Office Supplies
- Bottom 10 worst-performing products: concentrated in Furniture (especially Tables)
- Losses linked to high discounting + low margin structure
- **Recruiter takeaway:** ability to drill down from category to individual SKU-level performance

RQ7 – Customer Lifetime Value (CLV)

- CLV calculated as **AOV × Purchase Frequency × Lifespan**

- Distribution is heavily skewed: a small percentage of customers generate most of the lifetime value
- High CLV customers = candidates for loyalty programs, special offers, retention campaigns
- Low CLV = high churn risk, not cost-effective for heavy retention spending
- **Recruiter takeaway:** shows capacity to connect analytics with long-term customer relationship strategies

4. Practical Implications

- **Discounting strategy:** Avoid excessive discounts, particularly in Furniture
- **Customer focus:** Prioritize loyal/high-CLV customers, recover at-risk ones, limit investment in low-CLV groups
- **Regional priorities:** Double down on U.S., India, China; monitor weak subcategories in other regions
- **Seasonal planning:** Stock and market aggressively before Q4
- **Shipping optimization:** Promote Standard Class; reprice faster options to protect margins
- **Product portfolio:** Scale winners (Phones, Chairs, Copiers); fix or drop chronic loss-makers (Tables)

5. Closing Statement

This Global Superstore analysis answers seven real-world commercial questions, blending descriptive analytics, segmentation, forecasting, and profitability analysis.

The results deliver clear recommendations for pricing, product management, customer strategy, and supply chain efficiency. By grounding findings in hard numbers (sales, profit ratios, forecasts, CLV), this project demonstrates not just technical skills in Python and analytics but also **business acumen** recruiters look for.