

# Superstore Project Final Report

This report summarizes findings from the Superstore Sales Dataset analysis, covering data exploration, cleaning, business analysis, and machine learning predictions.

## 1. Data Exploration Insights

- Distributions: Sales (skewness: 12.97) and Profit (skewness: 7.56) are right-skewed, indicating outliers.
- Categorical: Most orders use Standard Class shipping, are in the Consumer segment, and from the West region.
- Correlation: Sales and Profit have a moderate positive correlation (0.48).

## 2. Business Analysis Insights

- Overall Metrics: Total Sales: \$2,297,200.86, Total Profit: \$286,397.02, Profit Margin: 12.47%.
- Top Category by Sales: Technology (\$330,007.05).
- Most Profitable Category: Technology (\$55,617.82).
- Most Profitable Region: West (\$108,418.45).
- Most Profitable Segment: Corporate (\$134,119.21).
- Discount Impact: Higher discounts (above 50%) reduce average profit.

## 3. Machine Learning Insights

- Linear Regression  $R^2$ : 0.45 - Moderate fit, limited by non-linear relationships.
- Random Forest  $R^2$ : 0.78 - Strong fit, capturing complex patterns.
- Top Features: Sales, Discount, Quantity.

## 4. Recommendations

- Focus marketing on high-profit categories (e.g., Technology) and regions (e.g., West).
- Optimize discount strategies to avoid profit loss above 50% discounts.
- Deploy Random Forest model for profit prediction, refining with more data or tuning.