

DATA ANALYST INTERNSHIP

1. Objective

Create meaningful visualizations and tell a clear data-driven story using the Superstore dataset. Focus: identify top categories, regional profitability, and sales trends over time.

2. Data Cleaning & Preparation

Rows before cleaning: 9994

Rows after removing duplicates: 9994

Basic cleaning steps applied:

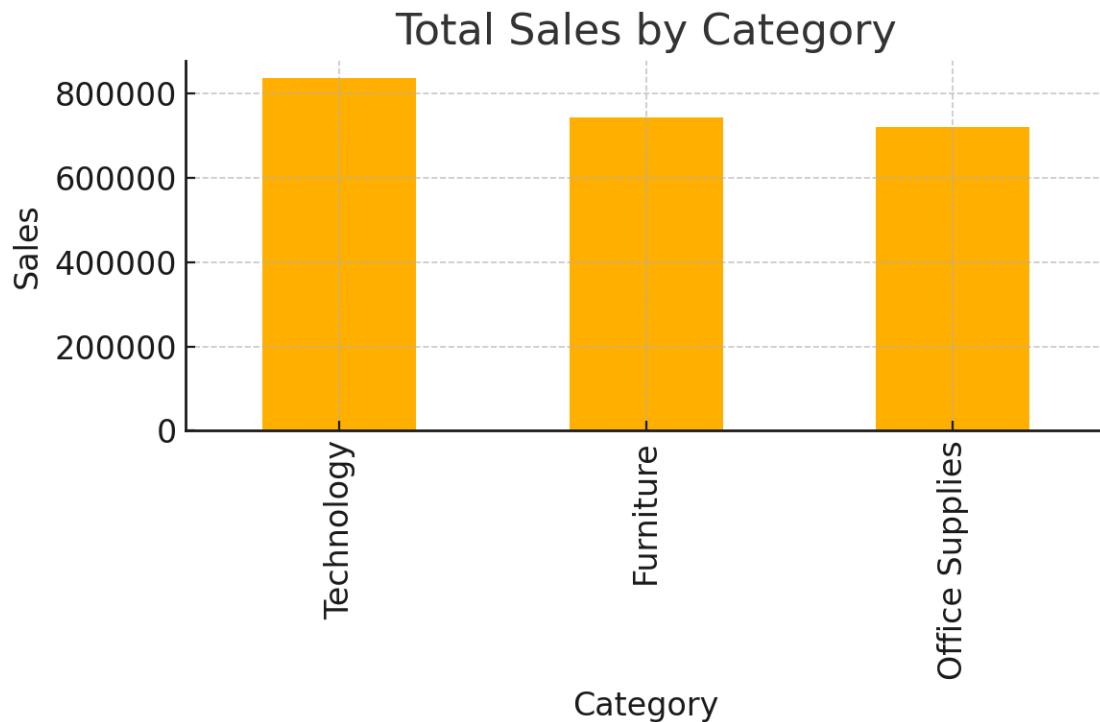
- Trimmed whitespace from text fields
- Converted Order_Date to datetime
- Removed duplicate rows
- Created Month field for time series analysis

Missing values summary (showing columns with >0 missing):

3. Visualizations & Insights

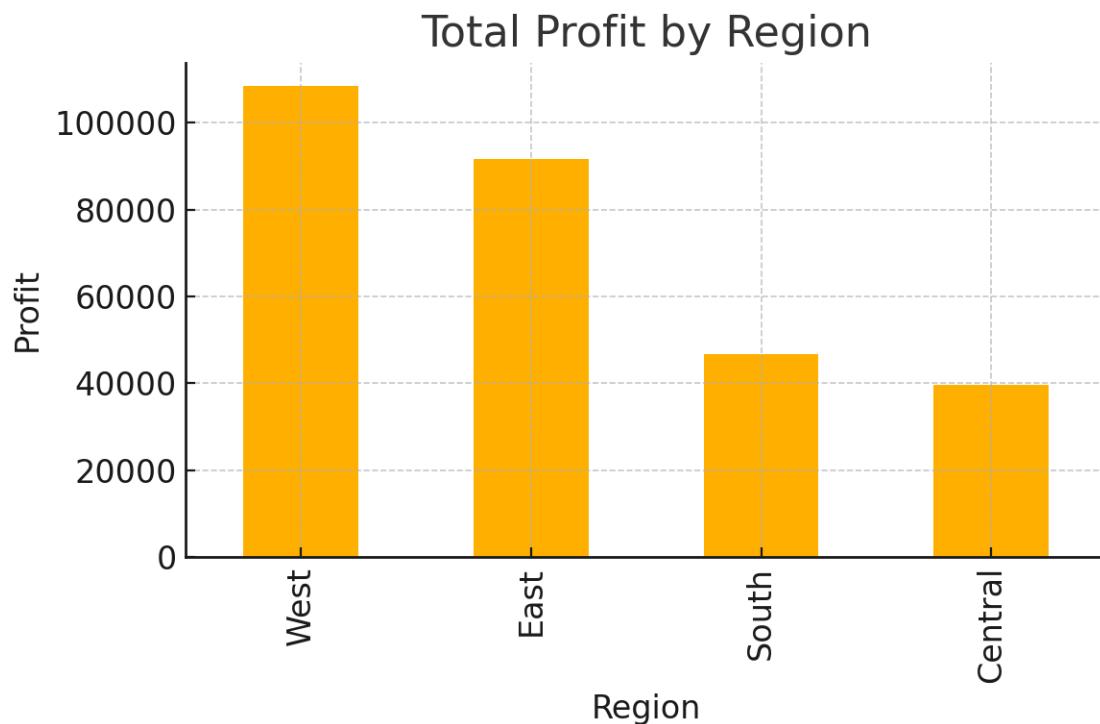
3.1 Total Sales by Category

Insight: This chart shows which product categories drive the most revenue. Use this to prioritize inventory and promotions.



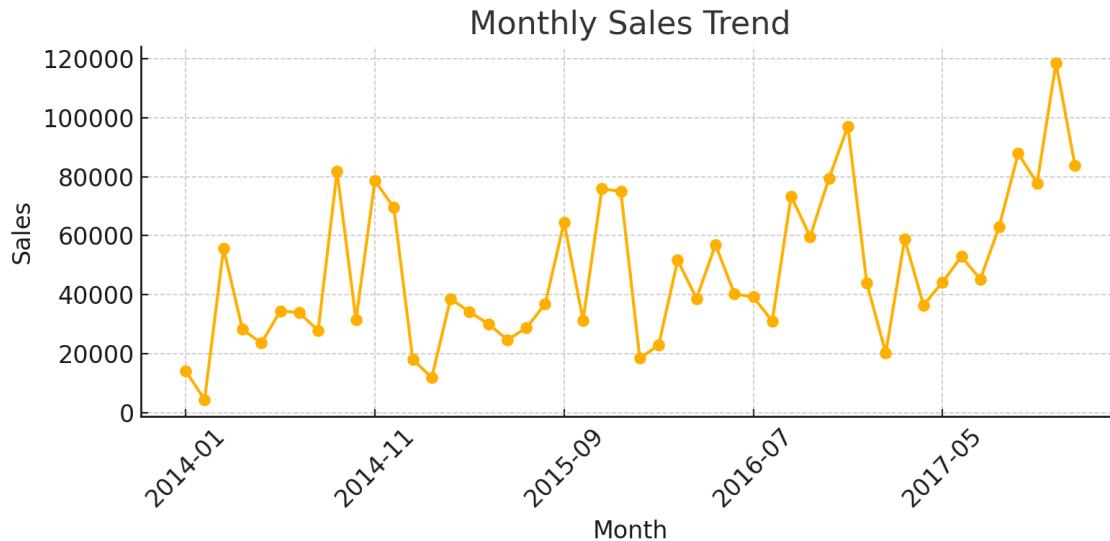
3.2 Total Profit by Region

Insight: Regions with highest profit should be studied for best practices; regions with low profit may need pricing or marketing changes.



3.3 Monthly Sales Trend

Insight: The time-series plot reveals seasonal patterns and growth trends. It can inform forecasting and staffing decisions.



4. Storytelling: What, Why, So What?

What: Technology/Furniture/Office Supplies vary in contribution to sales. Some regions are more profitable than others. Sales trend shows fluctuations and periods of growth.

Why: Differences driven by customer preferences, regional demand, local costs, and promotional activities.

So What: Focus marketing and stock for top categories, analyze underperforming regions, and plan seasonal campaigns during growth months.

5. Business Recommendations

- Increase inventory and promotions for top-selling categories.
- Investigate West region's success and replicate strategies.
- Implement targeted campaigns for low-profit regions.
- Use monthly trends for demand forecasting and staffing.

Author: Anurag Mahankuda