

EDA Report Summary: Sales Dataset

Dataset Overview:

- **Total Rows:** 2,823
 - **Total Columns:** 25
 - **Data Type:** Mix of categorical and numerical features.
 - **Date Column:** ORDERDATE .
-

Key Findings:

1. Data Quality:

- No missing values in major numerical fields like SALES, QUANTITYORDERED, PRICEEACH.

Some missing data found in:

- STATE, TERRITORY, and ADDRESSLINE2 — Not critical for sales-related analysis.
 - Date formats corrected and parsed.
-

2. Sales Insights:

- **Average Sales per Order:** \$3,554
 - **Sales Range:** From \$482 to \$14,083.
 - Significant **outliers** in SALES and QUANTITYORDERED — few orders bring disproportionately high revenue.
 - **Large Deal Sizes** are associated with the highest sales values.
-

3. Price and Quantity Patterns:

- **Average Price Each:** ~\$83.65
- **Quantity Ordered:**
 - Mean: 35 items/order
 - Orders vary widely — from 6 to 97 units per order.

- Products with **higher unit prices (PRICEEACH)** generally lead to **higher total sales (SALES)**.
-

4. Correlations (Bivariate Analysis):

- **Strong Positive Correlation** between:
 - PRICEEACH and SALES (0.65)
 - QUANTITYORDERED and SALES (0.58)
 - **Moderate Relationship** between:
 - MSRP (Manufacturer's Suggested Retail Price) and PRICEEACH.
 - Other fields (e.g., postal codes, months) show **very weak or no correlations**.
-

5. Deal Size Analysis:

- **Small Deals:** Lower SALES and PRICEEACH values.
 - **Medium & Large Deals:** Significantly higher SALES figures, contributing majorly to overall revenue.
-

6. Outlier Detection:

- **Boxplots** revealed:
 - Some very high-value sales.
 - Extremely high quantities ordered in rare cases.
 - Recommendation: Investigate extreme sales cases to ensure data integrity or detect special promotions/bulk orders.
-

7. Visual Analysis (Graphs and Charts):

- **Histograms:** Sales, Price, Quantity show right-skewed distributions.
 - **Heatmap:** Visual correlation among numeric variables.
 - **Pair plots:** Clear linear trend between price and sales.
 - **Scatterplot:** Clear segmentation among different deal sizes.
-

Summary of Insights:

"Sales performance is highly influenced by both unit pricing and quantity ordered. Larger deal sizes drive significantly more revenue. Although the overall dataset quality is good, specific high-sales outliers should be reviewed for special campaigns or anomalies. Future analyses can focus on temporal patterns (by quarter, month) and regional performance."

Thank you..