EDA Insights Report

Overview

This report summarizes the findings from the Exploratory Data Analysis (EDA) conducted on the sales dataset. The EDA process uses various visualizations to uncover patterns, trends, and relationships in the data. Each visualization is explained in order, highlighting the main insight it provides.

Dataset Summary

The dataset contains sales transaction records, including order details, customer information, product categories, sales amounts, profit, discounts, and indicators for holidays.

Data Quality and Preparation

- All columns were checked for missing values, and none were found, ensuring the dataset is complete and reliable for analysis.

- The data was reviewed for duplicates and inconsistencies; no duplicate records were detected.

- Data types were standardized, and date fields were converted to datetime format for accurate time-based analysis.

- Outliers in sales and profit were identified and retained for further business review, as they may represent significant or unusual transactions.

- New features such as "is_holiday" and "holiday_name" were created from date columns to enable analysis of holiday effects on sales and profit.

Summary Statistics

- Key numerical variables such as sales, profit, and discount were summarized using mean, median, minimum, maximum, and standard deviation.

- The profit distribution is slightly skewed, with some negative values indicating loss-making transactions.
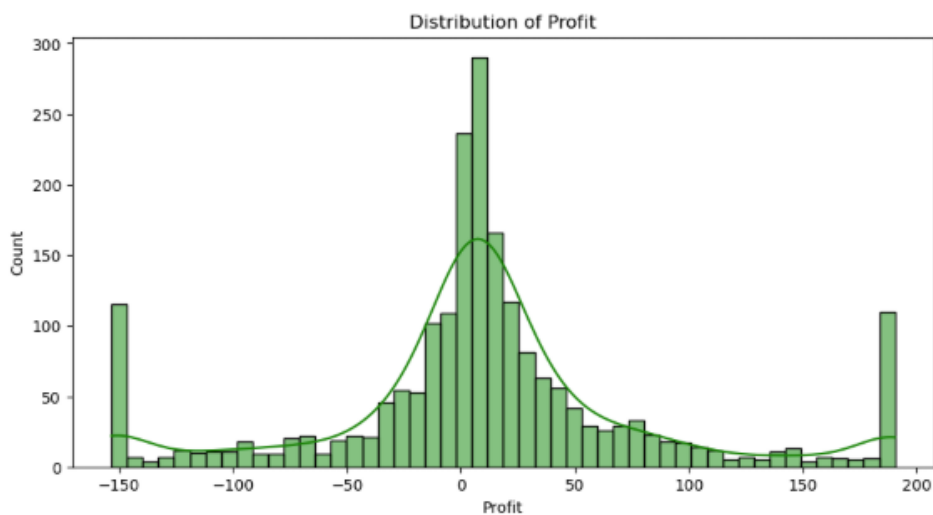
Visualization Insights

1. Distribution of Sales

- **Visualization:** Histogram of Sales

- **Insight:** Shows how sales amounts are distributed. Most transactions are of lower sales values, with a few high-value sales (potential outliers). This helps understand what a typical sale looks like and highlights any unusually large transactions.
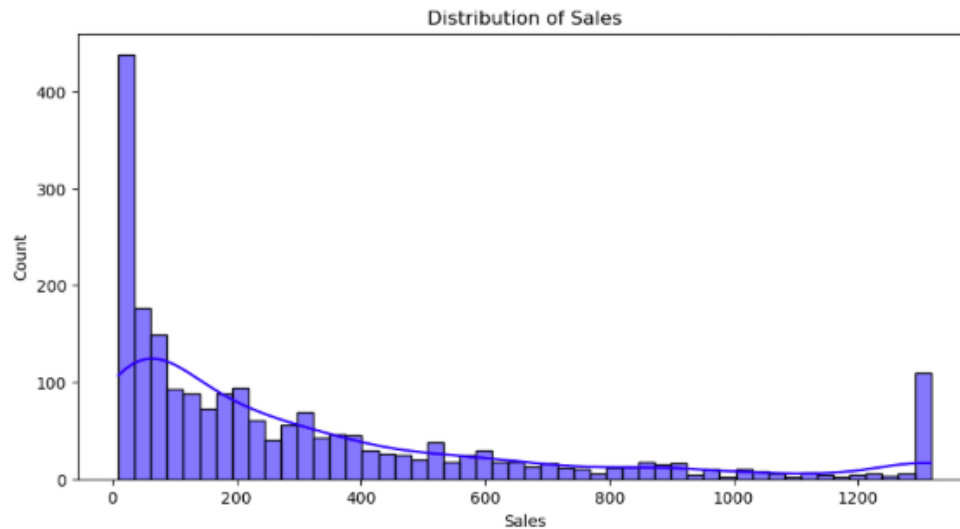
2. Profit Distribution

- **Visualization:** Histogram of Profit

- **Insight:** Displays the spread of profit values. Both positive and negative profits are present, indicating that some sales result in losses, possibly due to high discounts or costs. This helps identify areas where profitability can be improved.

3. Sales vs. Profit

- **Visualization:** Scatter Plot of Sales vs. Profit

- **Insight:** Shows the relationship between sales and profit for each transaction. Generally, higher sales lead to higher profits, but some high-sales transactions still result in losses. This helps identify where high sales do not guarantee profitability.


4. Sales by Category

- **Visualization:** Bar Chart of Total Sales per Category

- **Insight:** Highlights which product categories contribute most to overall sales. This can guide inventory and marketing focus.
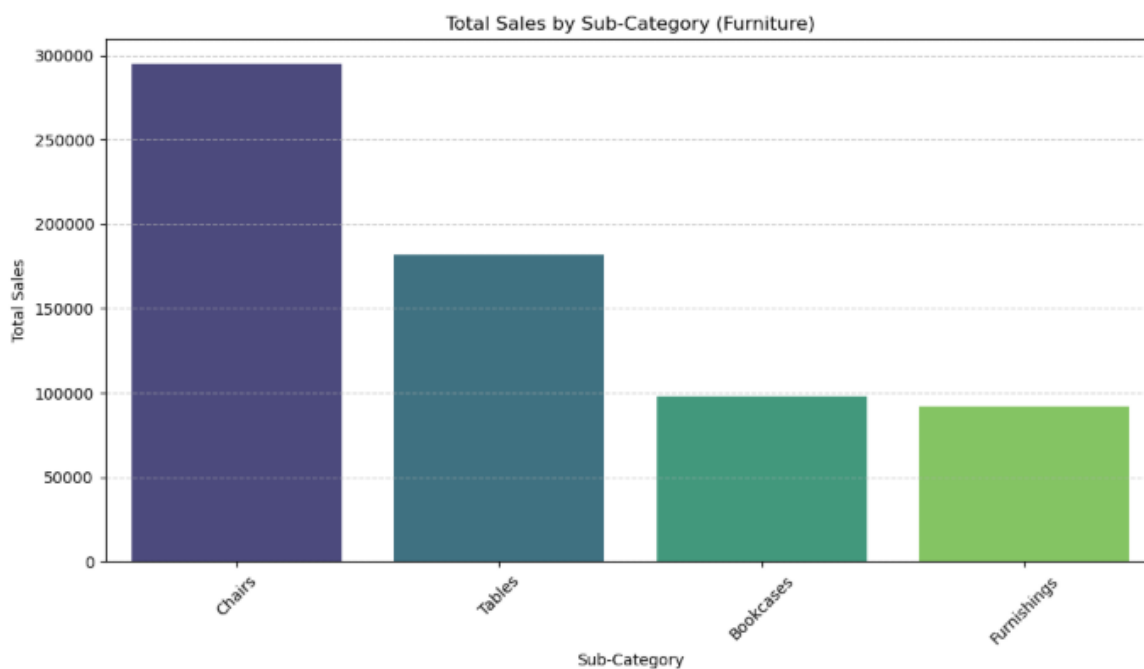
5. Profit by Sub-Category

- **Visualization:** Bar Chart of Total Profit per Sub-Category

- **Insight:** Reveals which sub-categories are most and least profitable. Some sub-categories may have high sales but low or negative profit, indicating a need for further analysis.
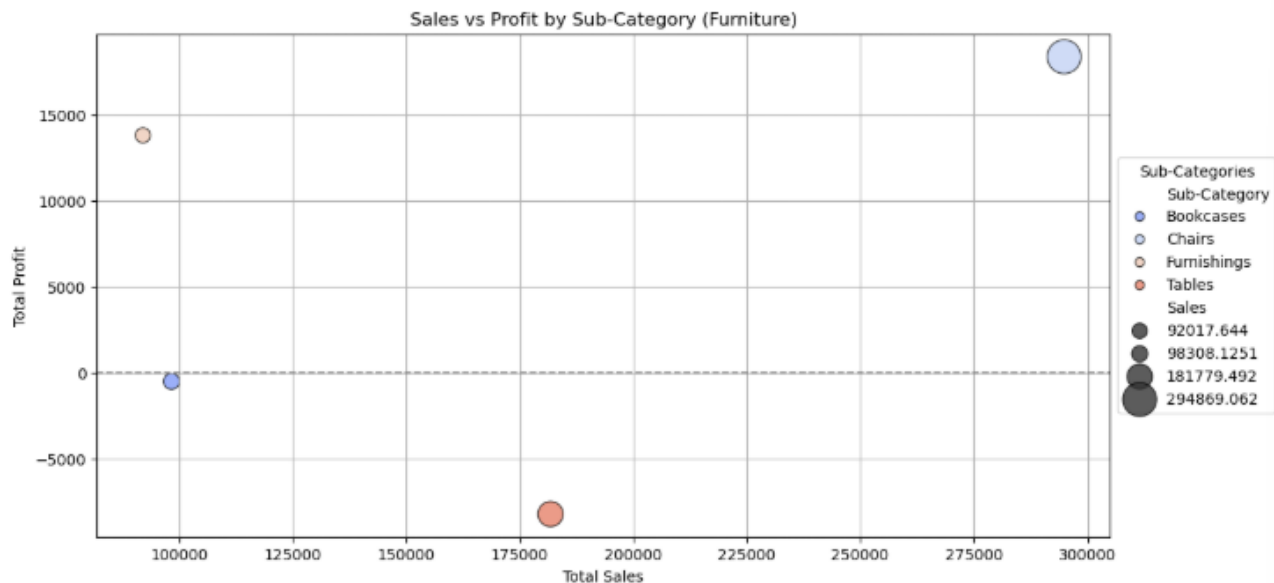

6. Orders Over Time

- **Visualization:** Line Chart of Number of Orders by Month/Year

- **Insight:** Shows trends and seasonality in order volume. Peaks may align with holidays or promotions, while dips indicate off-seasons. This helps in planning for demand fluctuations.
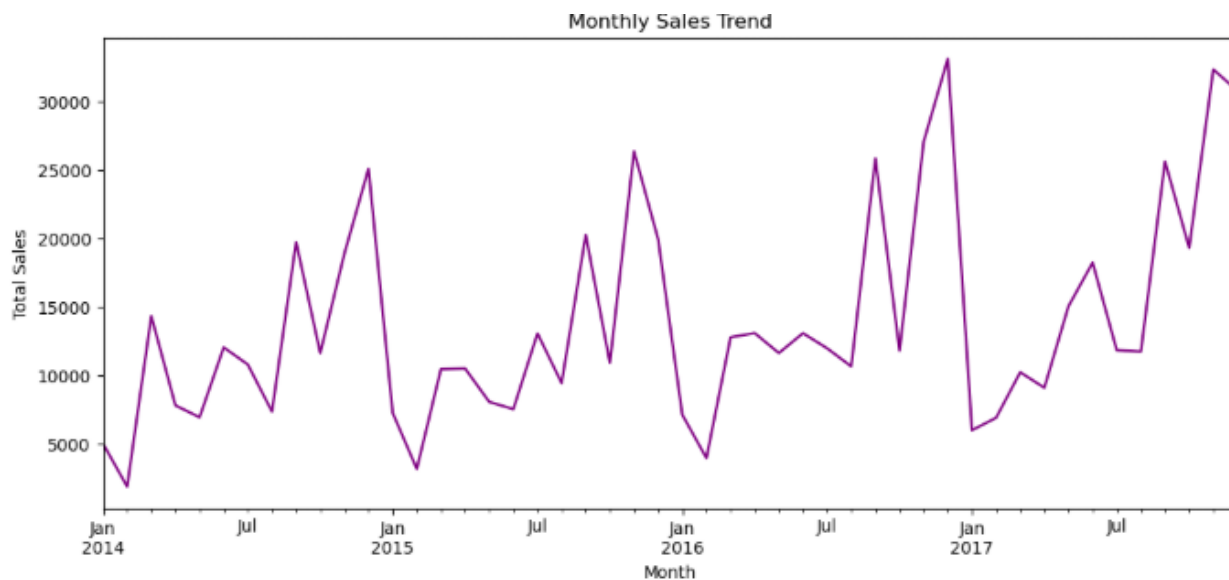
7. Holiday Impact

- **Visualization:** Comparison of Sales/Profit on Holidays vs. Non-Holidays

- **Insight:** Assesses whether holidays drive higher sales or profit. If holidays show higher averages, targeted campaigns during these periods could be beneficial.
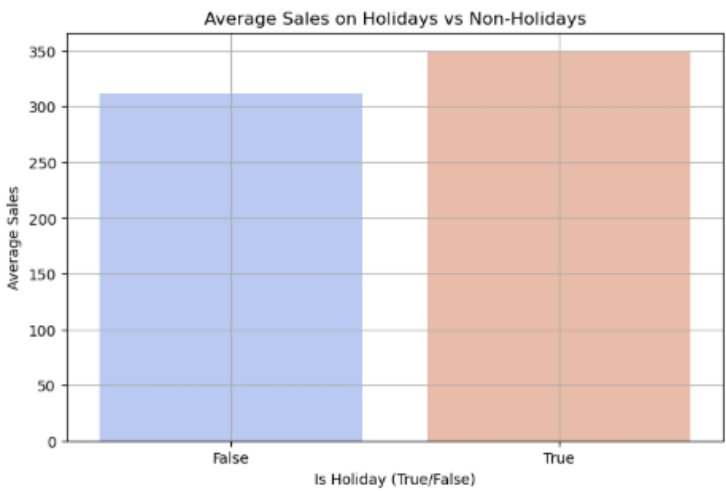

8. Correlation Matrix

- **Visualization:** Heatmap of Correlations between Numeric Variables

- **Insight:** Shows how variables like sales, profit, discount, and quantity are related. For example, a strong negative correlation between discount and profit suggests that higher discounts reduce profit.
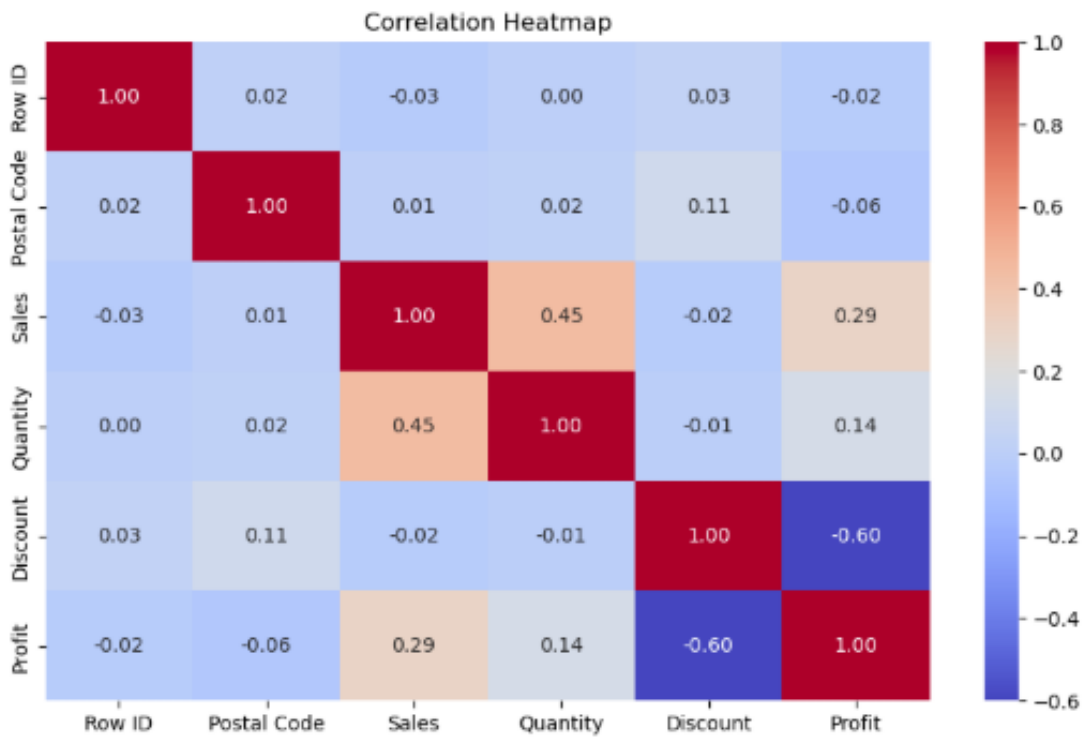
9. Discount Analysis

- **Visualization:** Scatter Plot (Discount vs. Profit, Furniture)

- **Insight:** This scatter plot shows that in the Furniture category, higher discounts often lead to losses (red points), especially when the discount is above 0.2. High discounts have a strong negative impact on profit.
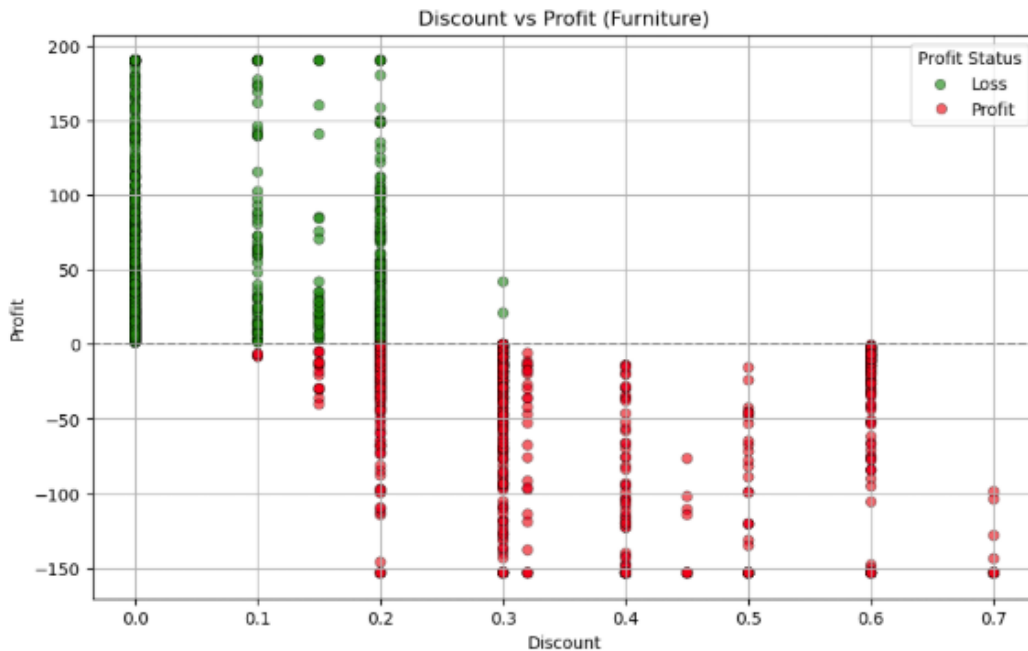
Key Takeaways

- Most sales transactions are of low to moderate value, with a few exceptionally large orders.
- Profitability is inconsistent; some transactions result in losses, often due to high discounts.
- A small number of categories and sub-categories contribute the majority of sales and profit.
- Order volumes show clear seasonality, with peaks during holidays and promotional periods.
- Discount strategies have a direct and significant impact on profit, highlighting the need for careful optimization.
- Holidays have a noticeable positive effect on both sales and profit, presenting opportunities for targeted marketing.
- Correlations between key variables (such as discount and profit) provide actionable insights for business decision-making.

Conclusion

The EDA process, through Nine diverse visualizations, has uncovered important patterns and actionable insights in the sales data. These findings can guide further analysis, inform business strategies, and support decision-making for improved performance and profitability.