Retail Sales EDA Summary

Dataset: Retail Store Sales (Dirty) for Data Cleaning | Source: Kaggle

Objective: Perform data cleaning and EDA on a retail sales dataset to reveal trends, top products, time-based behavior, and insights into customer and category spend.
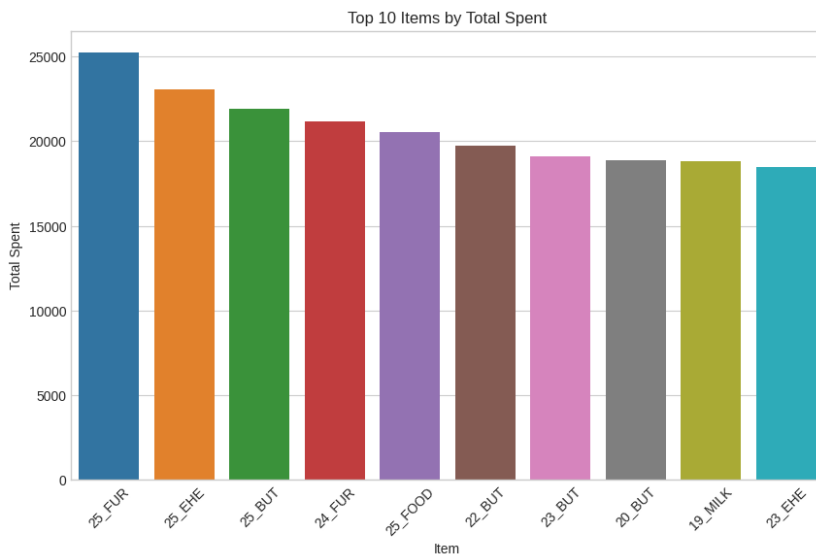
Highlighted Code Snippets

```python
# Cleaning
df['Transaction ID'] = df['Transaction ID'].str.replace('TXN_', '').astype(int)
df['Category'] = df['Category'].str.lower()

# Time Features
df['Transaction Date'] = pd.to_datetime(df['Transaction Date'])
df['Month_year'] = df['Transaction Date'].dt.to_period('M')

# Analysis
df.groupby('Category')['Total Spent'].sum().sort_values(ascending=False)
```
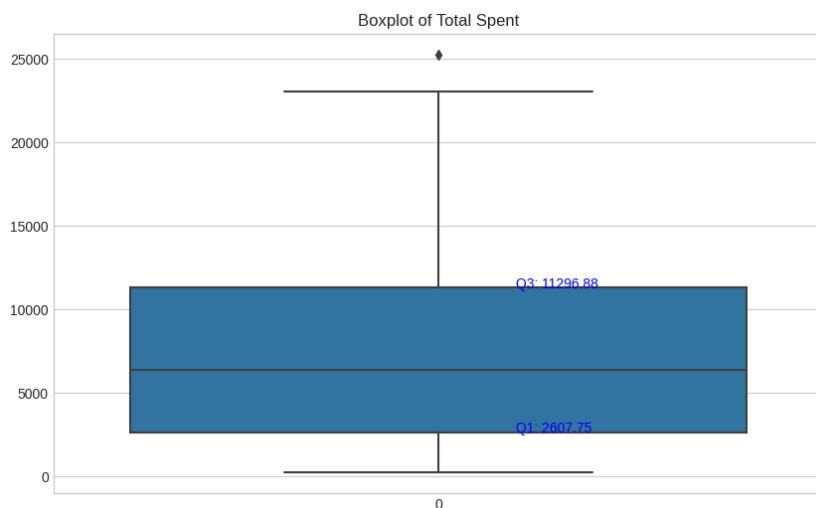
Top 10 Items by Spend

Spending Distribution (Boxplot)

Key Insights

- Butchers and electronics generate highest revenue
- Furniture leads in quantity sold
- Spending fluctuates across months
- Online purchases slightly higher than in-store
- Discount usage is evenly split; 33% unknown