Diwali Sales Data Analysis

Objective

Analyze Diwali sales data to understand customer purchasing behavior and identify factors that influence sales.

Dataset

- File: Diwali Sales Data.csv
- Original rows: 11,251
- Cleaned rows (after removing rows with null Amount): 11,239
- Columns after cleaning: 13 (dropped 'Status' and 'unnamed1')
- Key columns: User_ID, Cust_name, Product_ID, Gender, Age Group, Age, is_Married (formerly Marital_Status), State, Zone, Occupation, Product_Category, Orders, Amount

Data cleaning (performed)

1. Dropped 'Status' and 'unnamed1' columns.
2. Removed rows with null 'Amount' values (12 rows removed).
3. Converted 'Amount' to integer.
4. Renamed 'Marital_Status' to 'is_Married' (ensure rename uses inplace=True or assignment).
5. Verified data with info() and describe().

EDA findings

- Female shoppers are the majority and contribute higher total spending than male shoppers.
- The largest buyer segment is age 26–35, especially females in this group.
- Top states by orders and amount: Uttar Pradesh, Maharashtra, Karnataka.
- Married customers contribute a larger share of orders and amount (stored as 0/1 in the dataset; rename to is_Married for clarity).
- Top occupations by count/amount include IT, Healthcare, and Aviation.
- Product category counts show which categories are most frequently purchased.

Tools

Python, Pandas, NumPy, Matplotlib, Seaborn