Report summarizing the findings from the Exploratory Data Analysis (EDA) and data cleaning process:

Exploratory Data Analysis and Data Cleaning Report

1. Overview

The dataset comprises 86,400 rows of customer transaction data with various attributes such as customer demographics, spending details, and transaction metadata.

2. Data Quality and Cleaning

2.1 Missing Values

- No missing values were identified across any of the columns, indicating a complete dataset.

2.2 Data Types

- The data types appear appropriate for analysis. Numeric columns include 'avg_income' and 'spend,' while categorical variables are stored as objects.

2.3 Duplicates

- No duplicate entries were found, ensuring each row represents a unique customer transaction.

3. Descriptive Statistics

Descriptive statistics provide insights into the central tendency and spread of key numeric variables:

avg_income:
 - Mean: 51,657.03
 - Standard Deviation: 14,688.31
 - Minimum: 24,816.00
 - Maximum: 86,600.00

spend:
 - Mean: 614.46
 - Standard Deviation: 661.57
 - Minimum: 6.00
 - Maximum: 10,313.00

4. Categorical Variables

4.1 Unique Values

- Age Group: ['45+' '25-34' '21-24' '35-45']
- City: ['Bengaluru' 'Hyderabad' 'Chennai' 'Delhi NCR' 'Mumbai']
- Occupation: ['Salaried IT Employees' 'Salaried Other Employees' 'Government Employees' 'Freelancers' 'Business Owners']
- Gender: ['Male' 'Female']
- Marital Status: ['Married' 'Single']
- Month:['May' 'June' 'September' 'October' 'August' 'July']
- Category: ['Electronics' 'Groceries' 'Bills' 'Apparel' 'Food' 'Health & Wellness' 'Travel' 'Entertainment' 'Others']
- Payment Type: ['Net Banking' 'Debit Card' 'Credit Card' 'UPI']

Conclusion

The dataset is well-prepared for further analysis, and the initial exploration provides valuable insights into customer behavior and spending patterns.