EXPLORATORY DATA ANALYSIS OF CUSTOMER SPENDING HABITS

1. Introduction

This project focuses on performing Exploratory Data Analysis (EDA) on a synthetically generated customer dataset. The goal is to understand spending patterns, income distribution, and the relationship between income and transaction amounts.

2. Dataset Description

A total of 500 customer records were generated using Python. The dataset contains the following variables:

- Age (18 to 70 years)
- Annual Income ($30,000 to $150,000)
- Transaction Amount ($10 to $500)
- Customer Segment (A, B, C)

The data was generated using NumPy random functions and stored in a Pandas DataFrame.

3. Data Cleaning and Inspection

The dataset was inspected using: - data.info() - data.isnull().sum()

No missing values were found in the dataset, indicating it is clean and ready for analysis.

4. Descriptive Statistics

Descriptive statistics were calculated for: - Age - Annual Income - Transaction Amount

The analysis included: - Mean - Median - Standard Deviation - Minimum - Maximum

The income distribution centers around the mid-range with moderate variation. Transaction amounts also show a normal-like distribution with limited extreme values.

5. Central Tendency

Mean and median values were calculated for Age and Transaction Amount.
The mean and median values are relatively close, suggesting that the data is approximately symmetrically distributed without heavy skewness.

6. Correlation Analysis

The correlation between Annual Income and Transaction Amount was calculated.

A positive correlation was observed, indicating that customers with higher annual income tend to spend slightly more per transaction.

7. Visual Analysis

Histogram: The histogram of Transaction Amount shows that most purchases fall within a moderate spending range. Extreme low and high transactions are less frequent.

Scatter Plot: The scatter plot between Annual Income and Transaction Amount demonstrates a mild upward trend, reinforcing the positive correlation between income and spending.

8. Conclusion

The analysis suggests that:

- Customer spending behavior is moderately distributed.
- Higher-income customers tend to spend slightly more.
- The dataset is clean and suitable for analysis.
- Simulated customer segments are distributed across income levels.

This project demonstrates the complete workflow of data generation, descriptive statistics, visualization, and interpretation using Python.