Data Analysis Program Summary

1. Overview

This Python program demonstrates a complete Exploratory Data Analysis (EDA) workflow. It uses NumPy, Pandas, Matplotlib, and Seaborn to generate, process, and visualize a synthetic retail dataset of 500 records.

2. Data Generation & Processing

The dataset simulates a retail environment using various statistical distributions:
- Purchase Amount: Modeled using an exponential distribution (scale=120) to reflect realistic spending.
- Demographics: Randomly assigned ages (18-70) and product categories.
- Temporal Data: A daily date range starting from January 1, 2025.

3. Statistical Analysis

The program calculates key metrics for all numerical columns, including:
- Central Tendency: Mean and Median.
- Dispersion: Standard Deviation, Minimum, and Maximum.
- Distribution: First (25%) and Third (75%) Quartiles.
- Categorical Frequency: Counts for Product Categories and Return Status.

4. Visualizations Included

The program generates three distinct plots to provide visual insights:
1. Histogram: Shows the frequency distribution of Purchase Amounts.
2. Bar Chart: Compares the volume of transactions across different Product Categories.
3. Scatter Plot: Analyzes the relationship between Customer Age and Purchase Amount.