**Comprehensive Data Analysis with Pandas, NumPy, and Python**

**Objective**

Perform in-depth data analysis and generate actionable insights from a dataset, using Pandas, NumPy, Python functions, and custom logic.

**Dataset**

Use the following CSV file structure. Save it as **ecommerce\_data\_advanced.csv.**

| **OrderID** | **Product** | **Category** | **Price** | **Quantity** | **CustomerID** | **OrderDate** | **Rating** | **Discount** | **City** | **Country** | **ReturnStatus** | **PaymentMethod** |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| 1 | Laptop | Electronics | 800 | 1 | 1001 | 2024-10-01 | 4.5 | 0.10 | New York | USA | Returned | Credit Card |
| 2 | Phone | Electronics | 500 | 2 | 1002 | 2024-10-03 | 4.2 | 0.15 | Los Angeles | USA | Completed | PayPal |
| 3 | Chair | Furniture | 150 | 3 | 1003 | 2024-10-10 | 4.0 | 0.05 | Toronto | Canada | Completed | Debit Card |
| 4 | Headphones | Electronics | 100 | 4 | 1004 | 2024-10-12 | 3.8 | 0.20 | New York | USA | Returned | Credit Card |
| 5 | Desk | Furniture | 200 | 1 | 1005 | 2024-10-15 | 4.7 | 0.10 | London | UK | Completed | Credit Card |
| 6 | Gaming Console | Electronics | 300 | 1 | 1006 | 2024-10-20 | 4.6 | 0.05 | Chicago | USA | Completed | PayPal |
| 7 | Sofa | Furniture | 500 | 2 | 1007 | 2024-10-25 | 4.4 | 0.15 | Toronto | Canada | Completed | Debit Card |

**Tasks**

**1. Data Preparation**

* Load the dataset and ensure all date columns (OrderDate) are in datetime format.
* Handle missing data:
  + Drop rows with more than 2 missing values.
  + For numerical columns, use **interpolation** to fill missing values.
  + For categorical columns, replace missing values with "Unknown".

**2. Advanced Data Transformations**

* Create a new column EffectivePrice:
  + Formula: EffectivePrice=Price×Quantity×(1−Discount)
* Add a column CustomerType:
  + If EffectivePrice > 1000, label as Premium.
  + If EffectivePrice between 500-1000, label as Standard.
  + Otherwise, label as Economy.

**3. Data Insights and Queries**

* Using **groupby** operations:
  + Find the total revenue, average rating, and number of unique customers per Category.
  + Determine the city with the highest total revenue.
* Identify:
  + Products with a rating lower than the average for their category.
  + The total refund amount for Returned orders.

**4. Advanced Analysis with NumPy**

* Use NumPy to perform the following:
  + Calculate the **moving average** of the Price column with a window size of 3 and add it as a new column PriceMovingAvg.
  + Use np.random to simulate 5% inflation across all product prices and store the adjusted prices in a column InflatedPrice.

**5. Custom Functions and Logical Challenges**

* Write a function simulate\_discount that:
  + Accepts a Discount percentage.
  + Simulates a new EffectivePrice for all products with the given discount.
  + Returns the top 3 products with the highest simulated effective prices.
* Create a Python function analyze\_customer:
  + Accepts CustomerID as input.
  + Returns:
    - Number of orders placed.
    - Total spending.
    - Percentage of orders marked as Returned.

**6. Visualization (Mandatory)**

* Generate the following visualizations:
  + A **line plot** showing revenue trends over time.
  + A **heatmap** of revenue contributions by City and Category (use a pivot table).
  + A **histogram** of product ratings.

**7. Export**

* Save the final transformed dataset to ecommerce\_final.csv.
* Generate a summary text file (summary.txt) containing:
  + Total revenue.
  + Number of orders.
  + Most frequently used payment method.