Case Study

Question 1: Data Transformation and Analysis

You are working as a data analyst for a company that deals with large datasets. Your team receives a 2D NumPy array containing customer transaction data, where each row represents a different customer, and columns represent transaction amounts in different months.

Using NumPy functions:

1. Find the maximum, minimum, and total transaction amount across all customers.
2. Compute the cumulative sum of transactions column-wise.
3. Flatten the 2D array into a 1D array and sort the transactions in ascending order.
4. Transform the dataset by reshaping it into a different structure.

Write Python code to achieve these tasks and explain the significance of each function in data transformation.

Sample transaction data (4 customers, 5 months)

Question 2: Matrix Operations and Computation

A company is working on an AI-based recommendation system and needs to perform several matrix operations. The dataset consists of two matrices:

- **Matrix A** (User-Product Interaction): Contains ratings given by users to different products.
- **Matrix B** (Product-Feature Mapping): Contains product attributes that help in recommendations.

Using NumPy:

1. Perform element-wise multiplication of the user-product interaction matrix with another similar matrix.
2. Compute the matrix multiplication of A and B using two different NumPy functions.
3. Find the transpose of matrix A .
4. Identify the shape, number of dimensions, and data type of matrix B.

Write Python code to perform these operations and explain how these transformations are used in recommendation systems.

Defining matrices

A = np.array([[5, 3, 2], [4, 1, 5], [3, 2, 4]]) # User-Product Interaction

B = np.array([[1, 2], [3, 4], [5, 6]]) # Product-Feature Mapping

Question 3: Analyzing Student Performance Using NumPy

A school wants to analyze the performance of students in different subjects using Python and NumPy. The scores of students are stored in a NumPy array, where each row represents a student, and each column represents the marks obtained in a specific subject. The school's data analytics team needs to perform various operations on this dataset to derive insights and generate reports.

Problem Statement:

You have been given a dataset containing students' scores in multiple subjects. Using NumPy functions, perform the following tasks:

- Create a 2D NumPy array to represent the student scores.
- Reshape the array if needed to match the required format.
- Find the highest and lowest scores across all students and subjects.
- Calculate the total and average scores of students.
- Sort student scores in ascending order for each subject.
- Compute the square root of each score.
- Find the cumulative sum of scores row-wise and column-wise.
- Transpose the dataset to swap rows and columns for analysis.
- Perform matrix multiplication to analyze weighted scores.
- Use element-wise operations to scale or modify scores.
- Extract scores of a specific student.
- Extract marks of a particular subject using column slicing.
- Allow users to input scores and reshape them into a structured NumPy array.