राष्ट्रीय प्रौद्योगिकी संस्थान सिक्किम
NATIONAL INSTITUTE OF TECHNOLOGY SIKKIM
(An Institute of National Importance, Ministry of Education, Govt. of India)

Weekly Assignment - II

1. Write a python program to perform the following tasks:
   a. Write a Python function that takes a DataFrame containing transaction data and aggregates the total transaction amount per user. The DataFrame has columns 'user_id', 'item_name' and 'transaction_amount'.
   b. Write a Python function that takes a DataFrame and returns a random sample of 10% of the rows. Ensure reproducibility by setting a random seed.

2. Write a python program to perform the following tasks:
   a. Write a Python script to perform PCA (Principal Component Analysis) on a dataset. Assume the dataset has numerical features only.
   b. Write a Python function that selects a subset of features from a DataFrame based on a list of column names provided as input.
   c. Write a Python function that creates a new feature 'transaction_ratio' in a DataFrame. This feature is the ratio of 'transaction_amount' to 'account_balance'.

3. Write a Python function that discretizes a continuous feature 'age' into age groups [0-18, 19-35, 36-60, 61+] and binarizes a categorical feature 'gender' into binary columns.

4. Write a Python function that applies a logarithmic transformation to a numerical feature 'income' to reduce skewness.

5. Write python program to perform the following task:
    a. To compute the cosine similarity between two vectors.
    b. To compute the Euclidean distance between two points.
    c. To compute the Jaccard similarity between two sets.
    d. To compute the Pearson correlation coefficient between two arrays.
    e. To compute the Manhattan distance between two points.
    f. To compute the Hamming distance between two binary vectors.