**Homework 4**

Question 1

a) Create a 4x4 Numpy array with random integers between 1 and 20.

c) Replace all elements in the array that are greater than 10 with the value 10.

d) Calculate the sum of each row in the modified array.

Question 2

Given the array data = np.array([5, 10, 15, 20, 25, 30, 35, 40]):

a) Compute the standard deviation and variance of the array.

b) Find the median and the 25th percentile of the array.

c) Normalize the array so that it has a mean of 0 and a standard deviation of 1.

Question 3

a) Create a Numpy array of integers from 1 to 15.

b) Use boolean indexing to extract all even numbers from the array.

c) Create a new array where all numbers divisible by 3 are replaced with -3.

Question 4

a) Create two arrays: x = np.array([1, 2, 3]) and y = np.array([4, 5, 6]).

b) Stack the arrays vertically and horizontally.

c) Concatenate x and y to form a single array.

Question 5

a) Create a Pandas Series from a dictionary: {'a':10, 'b':20, 'c':30, 'd':40}.

b) Multiply all values in the Series by 2.

c) Check if the index 'e' exists in the Series.

d) Convert the Series into a DataFrame with a single column named 'Numbers'.

Question 6

a) Create a DataFrame df with columns 'A', 'B', and 'C', where each column contains 5 random integers between 0 and 50.

b) Add a new column 'D' which is the sum of columns 'A' and 'B'.

c) Drop column 'C' from the DataFrame.

d) Rename column 'D' to 'Total'.

Question 7

Given the DataFrame df from Question 6:

a) Set the index of the DataFrame to a range of dates starting from today.

b) Select the rows where the 'Total' is greater than 50.

c) Select only the 'A' and 'Total' columns for the first three rows.

Question 8

1. Create a DataFrame with some missing values given the following:

data = {'Name': ['Tom', 'James', 'Ricky', 'Vin', 'Steve'],

'Age': [28, 34, np.nan, 29, np.nan],

'Score': [85.0, np.nan, 79.0, np.nan, 92.0]}

b) Identify the columns with missing values.

c) Fill the missing 'Age' values with the median age.

d) Drop the rows where 'Score' is missing.

Question 9

Using the following DataFrame:

data = {'Team': ['A', 'A', 'B', 'B', 'C', 'C'],

'Player': ['John', 'Alice', 'Bob', 'Emily', 'Frank', 'Grace'],

'Points': [10, 15, 20, 25, 30, 35]}

df = pd.DataFrame(data)

a) Group the DataFrame by 'Team' and calculate the total 'Points' for each team.

b) Find the player with each team's maximum 'Points'.

Question 10

Given the following DataFrame:

data = {'Name': ['Anna', 'Ben', 'Charlie', 'Diana', 'Ethan', 'Fiona'],

'Age': [28, 22, 35, 30, 40, 25],

'City': ['New York', 'Los Angeles', 'New York', 'Chicago', 'Chicago', 'Los Angeles'],

'Income': [50000, 60000, 75000, 82000, 90000, 45000]}

df = pd.DataFrame(data)

a) Filter the DataFrame to include only individuals who are over 30 years old.

b) Further filter the DataFrame to include only those from 'New York' or 'Chicago'.

c) Calculate the average income of the filtered DataFrame.

d) Reset the index of the filtered DataFrame.

**Include all code and outputs in a Jupyter Notebook!**