Practice Project - 2

Learning Objective:

Part - I: Numpy

Part - II: Pandas

Part - I: Numpy

Q1) State differences between Numpy Array and lists.

Q2) Write a Code to create 50 evenly spaced numbers between 0 to 10.

Q3) Write a Code to create an array of shape (4,5) with random samples with random seed 7.

Q4) Write a Code to reshape the array created in Q3 into shape (2,10).

Q5) Create a 1D array starting from 0 ending with 100 with step size 10 and another array starting from 1000 ending with 10000 with step size 1000 and then concatenate two arrays created.

Q6) Use the array created in Q5 and replace the values less than 50 as <50.

Q7) Write a Code to create 1D, 2D, and 3D arrays randomly.

Q8) Write a Code to create an array with 10 random integers within 100.

Q9) Create an array with sample data with a shape of (5,4) with standard normal distribution with the random seed of 201.

Q10) Use the array from Q9 and create a Dataframe with index=['L','M','N','O','P'], columns=['A','B','C','D']

Q11) Find the most frequent value in an numpy 1D array. Use array x = np.array([10,20,66,20,10,30,50,60,70,60,66,12,13,66]).

Q12) Write a Code to Flatten a 2d array into 1d array. Use array: np.array([[1,2,3],[8,9,10],[11,22,33]])

Part - II: Pandas

Dataset Description: youtube_dislike_dataset.csv

This dataset contains information about trending YouTube videos from August 2020 to December 2021 for the USA, Canada, and Great Britain.

Context:

YouTube announced the decision to hide the number of dislikes from users around November 2021. However, the official YouTube Data API allowed you to get information about dislikes until December 13, 2021.
This dataset contains the latest possible information about dislikes, likes, views and more which was collected just before December 13. The information was collected by videos that had been trending in the USA, Canada, and Great Britain for a year prior.

Steps:

Q1) Import required libraries and read the provided dataset (youtube_dislike_dataset.csv) and retrieve top 5 and bottom 5 records.
Q2) Check the info of the data frame and write your inferences on datatypes and the shape of the dataset.
Q3) Check for the Percentage of the Null values and drop it.
Q4) Check the statistical summary of the dataset.

Q5) Convert data type of column published_at from object to pandas DateTime.

Q6) Create a column as 'published_month' using the column published_at.

Q7) Replace the numbers in the column published_month as names of the months i,e., 1 as 'Jan', 2 as 'Feb', and so on.....

Q8) Find the minimum and maximum value of the view_count, likes, dislikes, and comment_count and write your inferences.

Q9) Find the count of unique video_id, channel_id, and channel_title.

Q10) Find the top10 channel names having the highest number of videos in the dataset and the bottom10 having the lowest number of videos.

Q11) Find the title of the video which has the maximum number of likes and the title of the video has minimum likes and write your inferences.

Q12) Find the title of the video which has the maximum number of dislikes and the title of the video having minimum dislikes and write your inferences.