

**DJS-Compute**  
Machine Learning  
Assignment-2

**Topic: NumPy and Pandas**

**General Instructions:**

- The deadline for completing the tasks is 2nd October, 2023. A discussion session, addressing any doubts and task related topics, will be conducted in the following week.
- You can refer to online resources if needed .
- Add necessary/relevant comments in our code.

**Q1) Numpy :**

Create a copy of the Colab file, and complete the assigned tasks.

<https://colab.research.google.com/drive/1ISkCOPOAyhkZTYjAg2SIAGAO0t-Iy8dV?authuser=2>

**Q2) Pandas:**

Perform the mentioned tasks, on the following dataset:

<https://www.kaggle.com/datasets/geomack/spotifyclassification>

Resource:

<https://towardsdatascience.com/introduction-to-pandas-hands-on-tutorial-part-one-2e74f35ab166>

1. Initially, follow the steps as mentioned in the resource article. You do not need to set up Jupyter on your laptop, using a Kaggle notebook or Google Colab is fine.
2. Complete the 3 additional exercises, mentioned under the title, “ **Activities to sharpen skills**”, in the resource article.
3. Once you're done with that, perform the following tasks:
  - Create a dataframe of the songs that rank among top 10 for the instrumentality metric. This dataframe should only contain the instrumentality, song title and artist columns.
  - Create a dataframe of the average tempo for each of the top 10 most occurring artists.

**Note:** Make sure to explain every bit of code that you have written for the pandas task, succinctly, and **as per your understanding**.