****

**Work Integrated Learning Programmes Division**

**ML System Optimization**

## S2-23\_AIMLCZG516

**2023 -24**

**Assignment 2 – PS1 [Weightage 10%]**

**Instructions for Assignment Evaluation**

1. Please follow the naming convention as <Group no>\_<PS\_Number>.ipynb.

Eg – for group 1 with PS1 should be named as - Group1\_ PS1.ipynb.

1. Inside each jupyter notebook, you are required to mention your name, Group details and the Assignment dataset you will be working on.
2. Organize your code in separate sections for each task. Add comments to make the code readable.
3. Notebooks without output shall not be considered for evaluation.
4. Delete unnecessary error messages and long outputs.
5. Display the analysis of attributes in one frame rather than one after one. However, special treatment to attributes can be displayed separately.
6. Prepare a jupyter notebook (recommended - Google Colab) to build, train and evaluate the model.
7. Each group consists of up to 4 members. All members of the group will work on the same problem statement.
8. Each group should upload in CANVAS in respective locations under ASSIGNMENT Tab. Assignment submitted via means other than through CANVAS will not be graded.
9. Submission: Files should be uploaded on canvas without zipping them. One is ipynb file and other one html or pdf with output of the ipynb file.

**Problem Statement**

Implement a DataFrame Manipulation with cuDF.

Create, manipulate, and perform operations on cuDF DataFrames.

**Criteria:**

* **cuDF DataFrames** (6 Points):
  + Create a cuDF DataFrame with 100k rows, containing three columns: age (random integers between 18 and 70), income (random floats between 30,000 and 150,000), and state (randomly selected from a list of India states).
  + Filter the DataFrame to include only individuals who are over 30 years old and have an income greater than $50,000.
* **Functionality** (2 Points):
  + Group the filtered DataFrame by state and compute the mean income for each state.
* **Code Quality and Efficiency** (2 Points):
  + Submit the Python script used to generate the DataFrame, perform the filtering, and compute the group-by operation.

**\*\*\*\*\*\*\*\*\***