## Topic of Individual Project
## Due at 11:59 pm on Sunday, Nov 2

Based on the introductions of individual projects, please decide the topic of your individual project. Fill in the following questions and submit the filled PDF version before the deadline.

1. What's the topic? Add a short description here

Predicting Student Academic Performance Using Study Habits and Attendance Data
This project will explore how various academic behaviors, such as class attendance, study time, and participation, affect student performance in college courses. The analysis aims to identify which behavioral factors most strongly predict final grades and visualize performance patterns across different groups of students.

2. If the dataset exists, what's the link of the dataset/datasets? If not, how will you create this dataset?

I plan to use the Student Performance Dataset from the UCI Machine Learning Repository, available at: https://archive.ics.uci.edu/dataset/320/student+performance If additional features are needed, I will simulate or augment data by creating a small synthetic dataset representing additional student metrics.

3. What's the three questions to solve/answer to this topic?

Which study-related factors (attendance, study hours, parental involvement, etc.) most strongly correlate with final grade outcomes?

Can we accurately predict whether a student will pass or fail using their behavioral and demographic features?

How do study habits differ between high-performing and low-performing students?

4. Among your questions, which one need quantitative analysis (Training a model) to get your answer. What model/method do you plan to use?

The second question, predicting whether a student will pass or fail, requires quantitative analysis. I plan to use classification models, such as Logistic Regression and Random Forest Classifier, to train and evaluate model performance using accuracy, precision, recall, and F1 score as metrics. Visualization of feature importance will also be included.

5. Are there any suggestions you want to provide for the instructor?

It would be helpful if the instructor could provide a short list of recommended public datasets suitable for beginner-level model training, along with examples of successful data-cleaning workflows in Python.