# Capstone 2 Project Proposal

## Problem statement

What measures can be taken in the secondary school to increase students' final year grades (G3) by 10% in the next period by introducing new policies?

## Context

A survey about student achievement was implemented in secondary education of two Portuguese schools. The data attributes included student grades, demographic, and social- and school-related features) The target attribute G3 has a strong correlation with attributes G2 and G1. This occurs because G3 is the final year grade (issued at the 3rd period), while G1 and G2 correspond to the 1st and 2nd period grades. The board of school board would like to propose new policies based on the survey data to enhance students' performance in the future.

## Criteria for success

- The students' final year grade (G3) in the secondary school is enhanced by 10% in the next period.

## Scope of solution space

- This project will focus on the categorical and numerical features of the survey to analyze the key factors in the G3
- A regression model will be developed to represent the dependence of G3 on the key features and test possible policies.

## Constraints

- Some features in the dataset can be treated as either categorical or numerical. The use of various datatype brings uncertainties to the modeling.
- Some features, e.g., education of parents, may impact student's performance but not be able to consider in the policy

## Stakeholders

- Board of School
- Teachers
- Parents

## Data sources

- The dataset was collected from Kaggle website.

https://www.kaggle.com/datasets/devansodariya/student-performance-data

- A more detailed explanation of column names was collected from UCI Machine Learning Repository.

  https://archive.ics.uci.edu/ml/datasets/Student+Performance

**Procedure to complete the project**

1. Problem Identification (completed)
2. Data Wrangling
   - Check data integrity (completed)
   - Clean data: missing values, duplicates, etc. (in progress)
3. Exploratory Data Analysis
   - Summarize important features
   - Statistical property analysis
   - Visualize data
   - Identify trends and patterns
4. Pre-Processing and Training Data Development
   - Finalize the dataset by removing dispensable features
   - Splitting the dataset into testing and training subsets
   - Train the model
5. Modeling
   - Retrain the model with a whole set of data
   - Predictions and result analysis
6. Documentation
   A project report, a slide deck, and Jupyter notebooks will be generated and shared in a GitHub repo.
   https://github.com/z407f022/Capstone-2.git