

# Homework 2

DS440 Data Mining: Fall 2022

Grading Rubric

**100 points**

## Question 1: (40 points)

1. Problem 1.1 (5 points)
  - Correctly extracting the alcohol column and 12 other features (2.5 points)
  - Correctly displaying the pandas dataframe with the alcohol as the 13th column (2.5 points)
2. Problem 1.2 (10 points)
  - Correct scatterplot for all 12 features (6 points)
  - Correct subplot structure (4 points)
3. Problem 1.3 (10 points)
  - Most helpful feature with explanation (5 points)
  - Least helpful feature with explanation (5 points)
4. Problem 1.4 (15 points)
  - Valid answer + explanation for each part (5 + 5 + 5 points)

## Question 2: (60 points)

1. Problem 2.1 (20 points)
  - Fitting the straight line (10 points)
  - Correctly plotting the line (5 points)
  - Computing  $R^2$  correctly (5 points)
2. Problem 2.2 (30 points)
  - Plots +  $R^2$  for each degree 2, 3, 4, 5 (7 + 7 + 7 + 7 points)
  - Correctly plotting everything in one plot (2 points)
3. Problem 2.3 (10 points)
  - Valid model choice + valid explanation (5+5 points)