Stat451 Project Proposal
Group 27: Eve Lundin, Friday Mishra, Naydeen Mahmoud, Joel Myers

Write a one-page proposal (4 points) including a few lines of code to read data, descriptions of the question(s), variable(s), and methods you will use. Turn in a proposal.html (or .pdf), once per group

DataSet:
https://www.kaggle.com/datasets/lainguyn123/student-performance-factors

Description:
With this project, we aim to identify and analyze key factors that influence student performance, with a focus on personal, familial, and socioeconomic characteristics. By exploring which factors such as parental involvement, study habits, and health- most significantly impact academic outcomes, we want to gain insights into the challenges shaping students' success. We are thinking about using a KNN regression to evaluate the relative importance of each factor in question 1, and either a Decision tree or multiple linear regression for question 2. We could use a multiclass SVM for our third question.

Questions:

1. Which factors, like parental involvement, students' health, or study time, affect the probability of a student achieving a higher grade?
2. How do socioeconomic factors influence a students' achievement?
    a. Family Income, Access to Resources, Tutoring sessions, Parental Education Level
3. Do students with a high physical activity level and participation in extracurriculars have a high level of motivation in school?

```
[1]: 1 import pandas as pd

[8]: 1 df = pd.read_csv("StudentPerformanceFactors.csv")
     2 df.head()
```

[8]:

	Hours_Studied	Attendance	Parental_Involvement	Access_to_Resources	Extracurricular_Activities	Sleep_Hours	Previous_Scores	Motivation_Level	Internet_Access	Tutoring_Sessions	Family_Income
0	23	84	Low	High	No	7	73	Low	Yes	0	Low
1	19	64	Low	Medium	No	8	59	Low	Yes	2	Medium
2	24	98	Medium	Medium	Yes	7	91	Medium	Yes	2	Medium
3	29	89	Low	Medium	Yes	8	98	Medium	Yes	1	Medium
4	19	92	Medium	Medium	Yes	6	65	Medium	Yes	3	Medium

Teacher_Quality	School_Type	Peer_Influence	Physical_Activity	Learning_Disabilities	Parental_Education_Level	Distance_from_Home	Gender	Exam_Score
Medium	Public	Positive	3	No	High School	Near	Male	67
Medium	Public	Negative	4	No	College	Moderate	Female	61
Medium	Public	Neutral	4	No	Postgraduate	Near	Male	74
Medium	Public	Negative	4	No	High School	Moderate	Male	71
High	Public	Neutral	4	No	College	Near	Female	70

The dataset is a CSV with 6,607 rows of data and 20 features.