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:

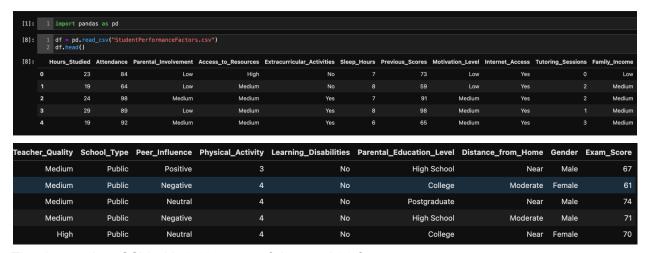
https://www.kaggle.com/datasets/lainguyn123/student-performance-factors

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?



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