Student Performance in Exams

Project

Analyze a set of data reporting public school student's scores in different exams, and other features

Importance

- Allows to get an overall picture on what factors can influence the performances the most
- Gives hints on where to intervene to guarantee a fair academic path

Aim

Discover possible trends and built a model that predicts score performances in exams

Jeffrey Gordon, Chris Lawson, Utkarsh Mujumdar, Alberto Salvarese

Dataset Description

The dataset contains data representing the test scores along with some demographic variables for 1000 students at a public school. Feature descriptions are as follows:

- Gender: Gender of the student
- Race/ethnicity: The racial/ethnic group the student belongs to
- Parental level of education: Highest achieved level of education by the students' parents
- Lunch: Indicates if the student part of a free/reduced price lunch program
- Test preparation course: Indicates if the student completed the test preparation course
- Math score: Math test score
- Reading score: Reading test score
- Writing score: Writing test score

Source: Understanding digital participation divides by Royce Kimmons (link)

Project Discovery

- Plan to do both Exploration and Hypothesis testing

- Exploration:
  - Explore if there are any underlying trends in the data (specifically clusters)

  - For example if different financial factors have any effect on on test scores

Project Discovery

- Hypothesis Testing
- There is an inverse relationship between Math and Reading and Writing scores
- The higher the parents education level, the higher scores will be

Problem Setup

- Regression problem for predicting each test score
- 3 Models - regressing each test score on the other variables (gender, race, parent education, etc)
- Then we will add in the other test scores as features, to see if a score on one test correlates to a score on another test
- Use these models to find the most important features for predicting the score & answering our hypotheses.
- Possible regression models - Linear, Lasso, Ridge, Gradient boosting, etc. Plan to fit multiple models, choose the best fitting model & explore why that model fit the best. This will give us further insight into the data.