Final Project - Step 1 (8 Points)

PSTAT100: Data Science Concepts and Analysis

> 🔥 Due Date
>
> The deadline for this step is Friday, April 25.

Formation of groups:

Monday 5pm Section Group 4

- Xinyao Song (xinyaosong)
- Nic Chan (nicholaschan)
- Jay Chang (jchang206)
- Darren Colianni (dcolianni)
- Matthew Chan (hchan837)

1 Tentative Role Distribution Table

Name	Responsibilities
Xinyao Song	GG-plot Interweaving & Analysis
Nic Chan	Research / Background Information
Jay Chang	Coding / Meeting Coordinator
Darren Colianni	Hypothesis Tests / Prediction Modeling
Matthew Chan	Writing the report / Research

2 Data Information

Data Name/Title: Student Performance

Author/Owner: Paulo Cortez

Date of Publication: 11/26/2014

Publication Venue: UC Irvine Machine Learning Repository

Retrieval Date: 4/22/25

Link: https://archive.ics.uci.edu/dataset/320/student+performance

Citation (APA): Cortez, P. (2008). Student Performance [Dataset]. UCI Machine Learning Repository. https://doi.org/10.24432/C5TG7T.

3 Initial Insights:

The data consists of two Portuguese high schools and follows a trimester system for student performance: G1, G2, and G3, each scored from 0 to 20. Among the features, study time, travel time, and absences seem to be the most important. Past performance is also likely a major influence. Our team finds it particularly interesting how good study habits and learning techniques developed during high school can significantly improve for certain students from 9th to 12th grade. This dataset presents us a unique opportunity to explore how behavioral factors shape student's educational performances.