# Final Project Report
## PSTAT100: Data Science Concepts and Analysis

**STUDENT NAME**

- Xinyao Song (xinyaosong)
- Matthew Chan (hchan837)
- Darren Colianni (dcolianni)
- Nic Chan (nicholaschan)
- STUDENT 5 (NetID 5)

# 1 Abstract

This project investigates how various social, familial, and behavioral factors are correlated with academic performance and average free times among Portuguese high school students. Using the Student Performance dataset compiled by Cortez and Silva in 2008, we explore three specific research questions: the first two concerning the influence of parental education and alcohol consumption on students' final Portuguese grades (G3), and the last one focusing on how one's relationship status influence the amount of free time students report. From the result, we can see that [FINISH AT THE END]

# 2 Introduction

This project focuses on how different factors influence school performance in Portuguese secondary school students. The dataset our team used is the Student Performance dataset compiled by Cortez and Silva in 2008. The dataset contains 649 observations and 30 features collected through questionnaires from two Portuguese high schools, with features covering aspects of family background, social environment, academic records, and types of lifestyles. The primary objective of our analysis is to examine how different social and behavioral factors relate to the final Portuguese grade (G3), which is measured on a scale from 0 to 20. By answering the three focused research questions below, we aim to understand how various background and behavioral variables correlate with or predict academic performance. The results from this analysis could help inform educational strategies and potentially help students with their stress and mental health.

## 2.1 Research Question 1

After statistically controlling for weekly study time (on a 1–4 numeric scale where 1 indicates less than 2 hours and 4 indicates more than 10 hours) and internet access (coded as "yes" or "no"), how is the father's highest education level (on a 0–4 numeric scale where 0 indicates no formal education and 4 indicates higher education) associated with Portuguese high school students' final grades in Portuguese class (G3), which are measured on a 0–20 numeric scale?

**Hypothesis:**

$H_0$: After controlling for weekly study time and internet access, the father's highest education level (0 = no formal education to 4 = higher education) has no significant association with students' final Portuguese grade (G3, measured on a 0–20 scale).

$H_1$: After controlling for weekly study time and internet access, the father's highest education level ($0 =$ no formal education to $4 =$ higher education) is significantly positively associated with students' final Portuguese grade (G3, 0–20 scale).

## 2.2 Research Question 2

Based on the number of alcoholic beverages consumed by Portuguese high school students per weekday and weekend (on a 1-5 numeric scale where 1 indicates very low and 5 indicating very high), how does it affect Portuguese high school students' final grades in Portuguese class (G3), which are measured on a 0–20 numeric scale?

**Hypothesis:**

$H_0$: There is no significant association between students' alcohol consumption levels during weekdays and weekends (1 = very low to 5 = very high) and their final Portuguese grade (G3, 0–20 scale).

$H_1$: Students' alcohol consumption levels during weekdays and weekends (1 = very low to 5 = very high) are significantly associated with their final Portuguese grade (G3, 0–20 scale).

## 2.3 Research Question 3

Is there a statistically significant difference in reported average free time (on a 1-5 categorical scale where 1 indicates very low and 5 indicates very high) between Portuguese high school students who are in romantic relationships (coded as 1), and those who are not in romantic relationships (coded as 0)?

**Hypothesis:**

$H_0$: There is no significant difference in average free time between students in a romantic relationship and those who are not.

$H_1$: There is a significant difference in average free time (1 = very low to 5 = very high) between students in a romantic relationship and students who are not in romantic relationships.

# 3 Data Processing

The Student Performance dataset does not contain any missing values or fields with unclear meanings. Therefore, no data cleaning or preprocessing was necessary for this project.

# 4 Research Question 1

## 4.1 Modeling Process

## 4.2 Results

## 4.3 Interpretation

# 5 Research Question 2

## 5.1 Modeling Process

## 5.2 Results

## 5.3 Interpretation

# 6 Research Question 3

## 6.1 Modeling Process

## 6.2 Results

## 6.3 Interpretation

# 7 Conclusion