# Assigment4

Miray Hozer

2025-10-25

## Topic

An Analysis of Socio-Demographic and Lifestyle Factors Influencing the Academic Performance of High School Students

## Summary

Student success is a really important topic for all of us. It's not just about grades; it also plays a big role in students' futures and for society as a whole. It wouldn't be right to say a student's success only depends on their intelligence or the school they go to. A lot of different things like their family, environment, and social life affect their grades too. That's why understanding what helps or hurts success is really necessary to take the right steps to help both schools and students.

In this project, my main goal is to visualize the factors that affect student success using a dataset from two high schools in Portugal. I'll be using the final grade, G3, to measure success. To understand the topic better, I will focus on three main areas: the family situation, the student's own study habits, and their social life. For family factors, I'll look at the mother's and father's education levels (Medu, Fedu). For the student's own habits, I'll focus on how many hours they study per week (studytime) and whether they've failed a class before (failures). Under the social life category, I'll examine their weekday and weekend alcohol consumption (Dalc, Walc) and how often they go out with friends (goout).

There are some basic questions I'll try to answer in this study. For example, does the parents' education really affect a student's grades? Do grades always improve as study time increases? Or do habits like drinking alcohol and going out a lot have a negative impact on grades? My own guess is that students whose parents are more educated, who study regularly, and don't have harmful habits will probably have higher grades.

In the next stages of the project, I will create various visuals like box plots and scatter plots using the R programming language and the ggplot2 package to answer these questions. My goal is to turn these numbers and data into visuals that tell a story that everyone can understand.

## Dataset Description

- **Dataset Name:** Student Performance Data Set

- **Description:** This dataset contains student achievement data collected from two Portuguese high schools: Gabriel Pereira and Mousinho da Silveira. The data covers the students' demographic features, family information, social habits, and academic performance in three different exams (G1, G2, and the final grade, G3).

- **Creators/Institution:** P. Cortez and A. Silva. The data was collected as part of a project at the University of Minho, Portugal.

- **Source and Access Link:** The dataset is publicly available through the Source of the Dataset, which is frequently used in the field of data science.

```r
student_data <- read.csv("student_data.csv")
summary(student_data)
```

```
##     school              sex                 age          address
##  Length:395         Length:395         Min.   :15.0   Length:395
##  Class :character   Class :character   1st Qu.:16.0   Class :character
##  Mode  :character   Mode  :character   Median :17.0   Mode  :character
##                                        Mean   :16.7
##                                        3rd Qu.:18.0
##                                        Max.   :22.0
##    famsize            Pstatus               Medu            Fedu
##  Length:395         Length:395         Min.   :0.000   Min.   :0.000
##  Class :character   Class :character   1st Qu.:2.000   1st Qu.:2.000
##  Mode  :character   Mode  :character   Median :3.000   Median :2.000
##                                        Mean   :2.749   Mean   :2.522
##                                        3rd Qu.:4.000   3rd Qu.:3.000
##                                        Max.   :4.000   Max.   :4.000
##     Mjob               Fjob               reason             guardian
##  Length:395         Length:395         Length:395         Length:395
##  Class :character   Class :character   Class :character   Class :character
##  Mode  :character   Mode  :character   Mode  :character   Mode  :character
##
##
##
##    traveltime      studytime       failures       schoolsup
##  Min.   :1.000   Min.   :1.000   Min.   :0.0000   Length:395
##  1st Qu.:1.000   1st Qu.:1.000   1st Qu.:0.0000   Class :character
##  Median :1.000   Median :2.000   Median :0.0000   Mode  :character
##  Mean   :1.448   Mean   :2.035   Mean   :0.3342
##  3rd Qu.:2.000   3rd Qu.:2.000   3rd Qu.:0.0000
##  Max.   :4.000   Max.   :4.000   Max.   :3.0000
##    famsup              paid             activities          nursery
##  Length:395         Length:395         Length:395         Length:395
##  Class :character   Class :character   Class :character   Class :character
##  Mode  :character   Mode  :character   Mode  :character   Mode  :character
##
##
##
##    higher             internet           romantic            famrel
##  Length:395         Length:395         Length:395         Min.   :1.000
##  Class :character   Class :character   Class :character   1st Qu.:4.000
##  Mode  :character   Mode  :character   Mode  :character   Median :4.000
##                                                           Mean   :3.944
##                                                           3rd Qu.:5.000
##                                                           Max.   :5.000
##    freetime          goout            Dalc            Walc
##  Min.   :1.000   Min.   :1.000   Min.   :1.000   Min.   :1.000
##  1st Qu.:3.000   1st Qu.:2.000   1st Qu.:1.000   1st Qu.:1.000
##  Median :3.000   Median :3.000   Median :1.000   Median :2.000
##  Mean   :3.235   Mean   :3.109   Mean   :1.481   Mean   :2.291
##  3rd Qu.:4.000   3rd Qu.:4.000   3rd Qu.:2.000   3rd Qu.:3.000
##  Max.   :5.000   Max.   :5.000   Max.   :5.000   Max.   :5.000
##     health          absences            G1              G2
##  Min.   :1.000   Min.   : 0.000   Min.   : 3.00   Min.   : 0.00
```

```
##   1st Qu.:3.000    1st Qu.: 0.000    1st Qu.: 8.00    1st Qu.: 9.00
##   Median :4.000    Median : 4.000    Median :11.00    Median :11.00
##   Mean   :3.554    Mean   : 5.709    Mean   :10.91    Mean   :10.71
##   3rd Qu.:5.000    3rd Qu.: 8.000    3rd Qu.:13.00    3rd Qu.:13.00
##   Max.   :5.000    Max.   :75.000    Max.   :19.00    Max.   :19.00
##         G3
##   Min.   : 0.00
##   1st Qu.: 8.00
##   Median :11.00
##   Mean   :10.42
##   3rd Qu.:14.00
##   Max.   :20.00
```