

# Student (Math) Performance



# Introduction

- Case:
- This dataset contains the final scores of students at the end of a math programs with several features that might or might not impact the future outcome of these students.
- Dataset:
- This is a dataset from the UCI datasets repository.
- 395 entries and 33 features.
- Target G3: (final grade numeric: from 0 to 20)



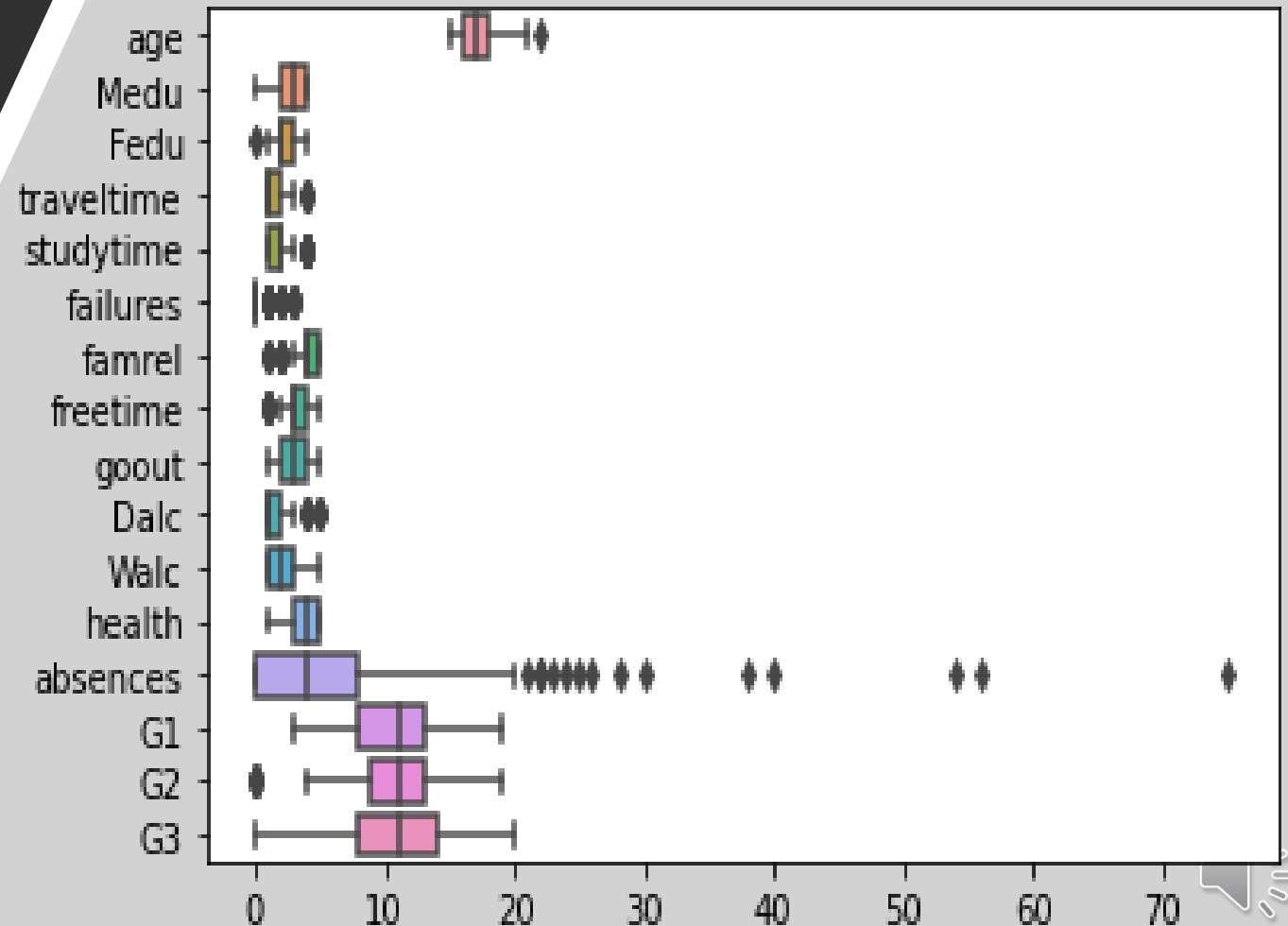
# Data Cleaning

- Data has no missing values
- Data has no duplicates
- Did no feature engineering.



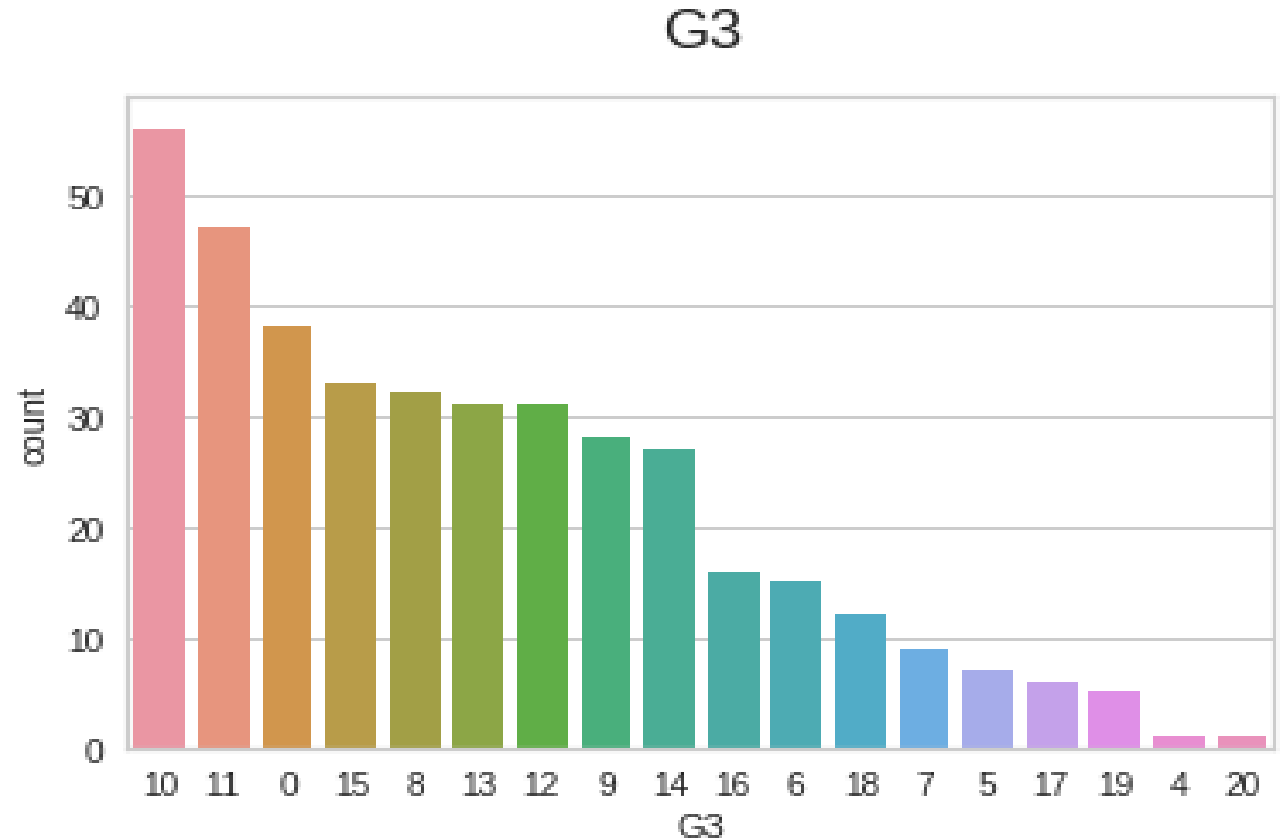
# Exploratory Viz

Absences has the most outliers;  
we can see that this makes a  
difference later.



# Exploratory Viz

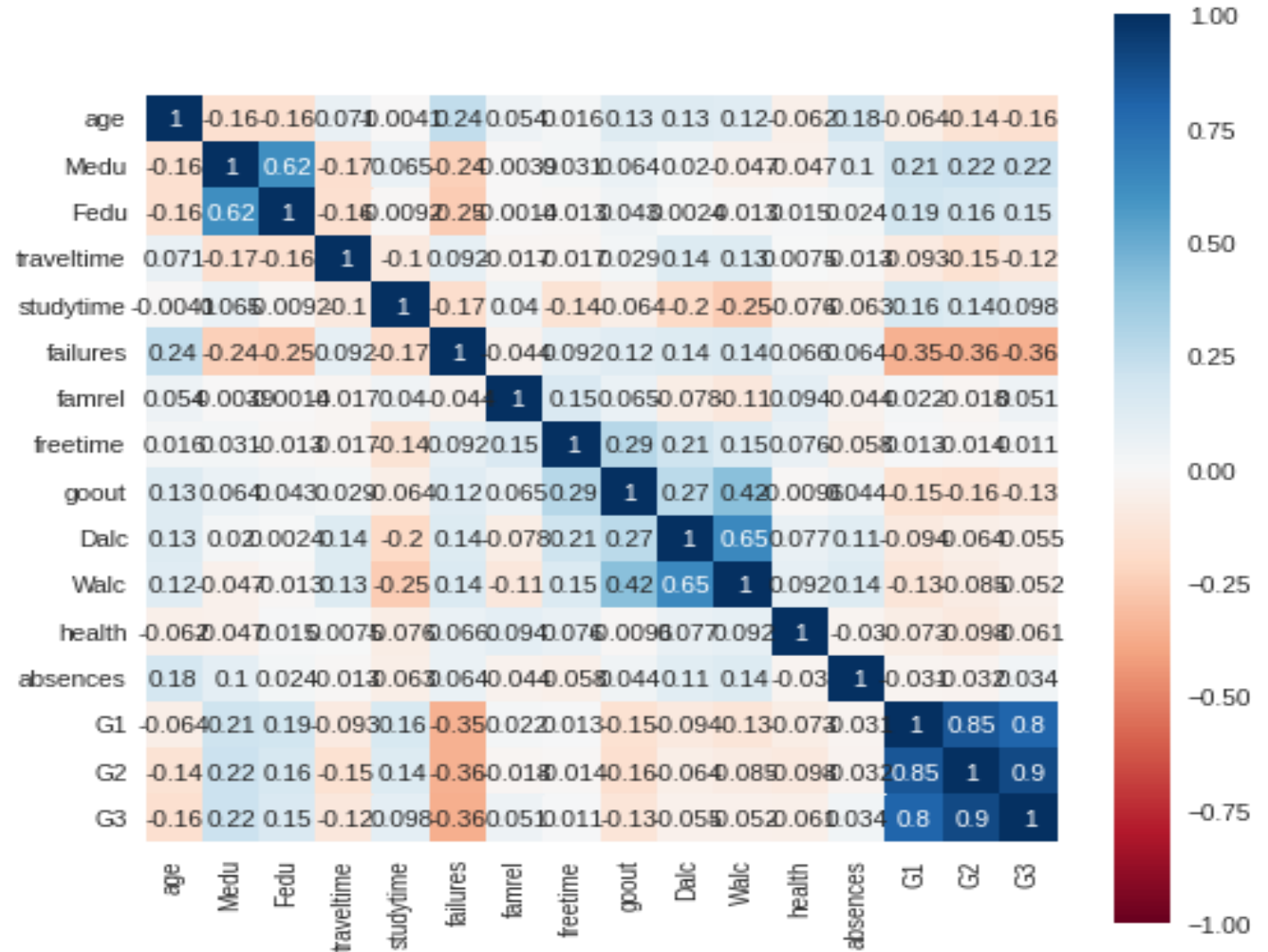
We see that there is right tail distribution;  
we do not have normal distribution—most  
students are not doing well.



# Exploratory Viz

We can see that: G2, which is second period grade has high corr with G3, as well as G1.

We also see that failures has negative corr with G3.



# Exploratory Viz

We see that age given; we see an increase with G3; we also see that outliers are mostly males.

