Student Performance in Mathematics Classroom

Maheen Syed - MS Data Science - Brown University

Github: https://github.com/msyed96/DATA1030-Project

25 - October - 2019

AGENDA:

- Introduction
- Dataset overview
- Preprocessing
- EDA
- Conclusion/Questions

INTRO:

- **Problem:** Can we predict academic performance based on the different features we know about the students in the Math class?
- **Importance:** Make education more accessible and inclusive by looking at factors that affect student performance
- **Type of problem:** Classification
- **Source:** UCI Machine Learning Repository

DATASET OVERVIEW:

- Student performance in Secondary School Math Class, in two schools based in Portugal. (Data collected in 2006)
- Grades - 0 to 20 points scale
- Students Evaluated three times (G1, G2 and G3), with G3 being the final grade.

PREPROCESSING: STEPS

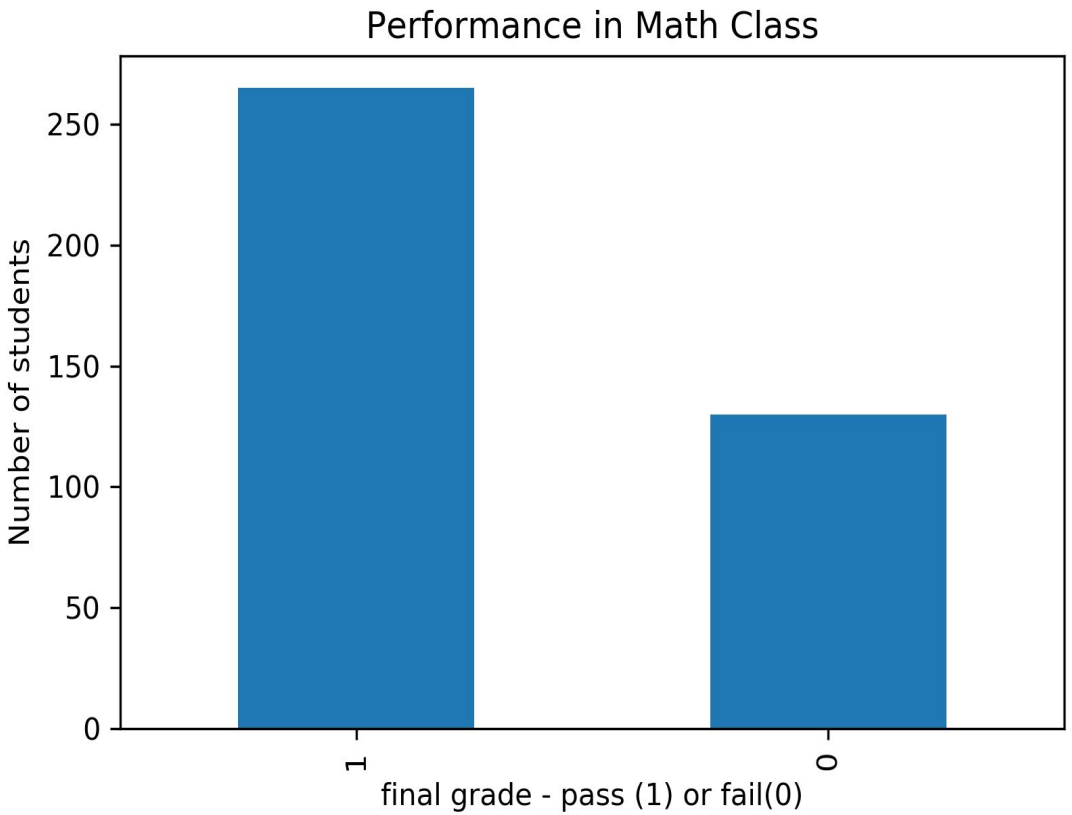
- Number of Data points before preprocessing: (395,33)
- Number of Missing Values: 0
- Number of Data points after preprocessing: (395, 50)
- Target Variable: G3 - Final Grade in Class <10 is Fail | >= 10 is Pass

PREPROCESSING: STEPS

Features	Preprocess
Categorical Features - 17 ftrs	One Hot Encoding
Categorical - Ordinal Features - 11 ftrs	Already Preprocessed
Numerical Features - 4 ftrs	Min-Max Scaling

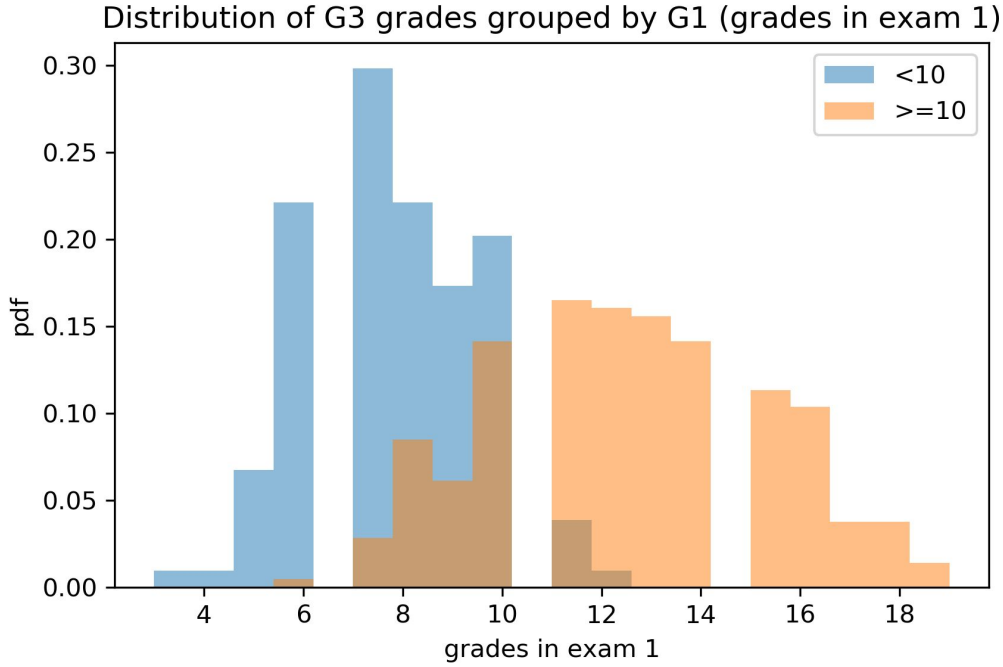
- Target Variable - G3 - Label Encoding
  0 - Fail | 1 - Pass

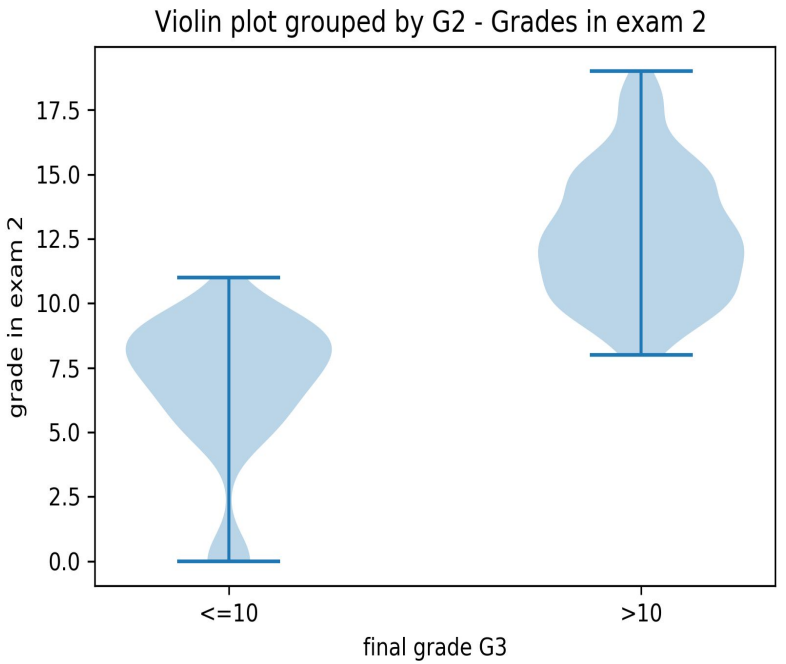
EDA:



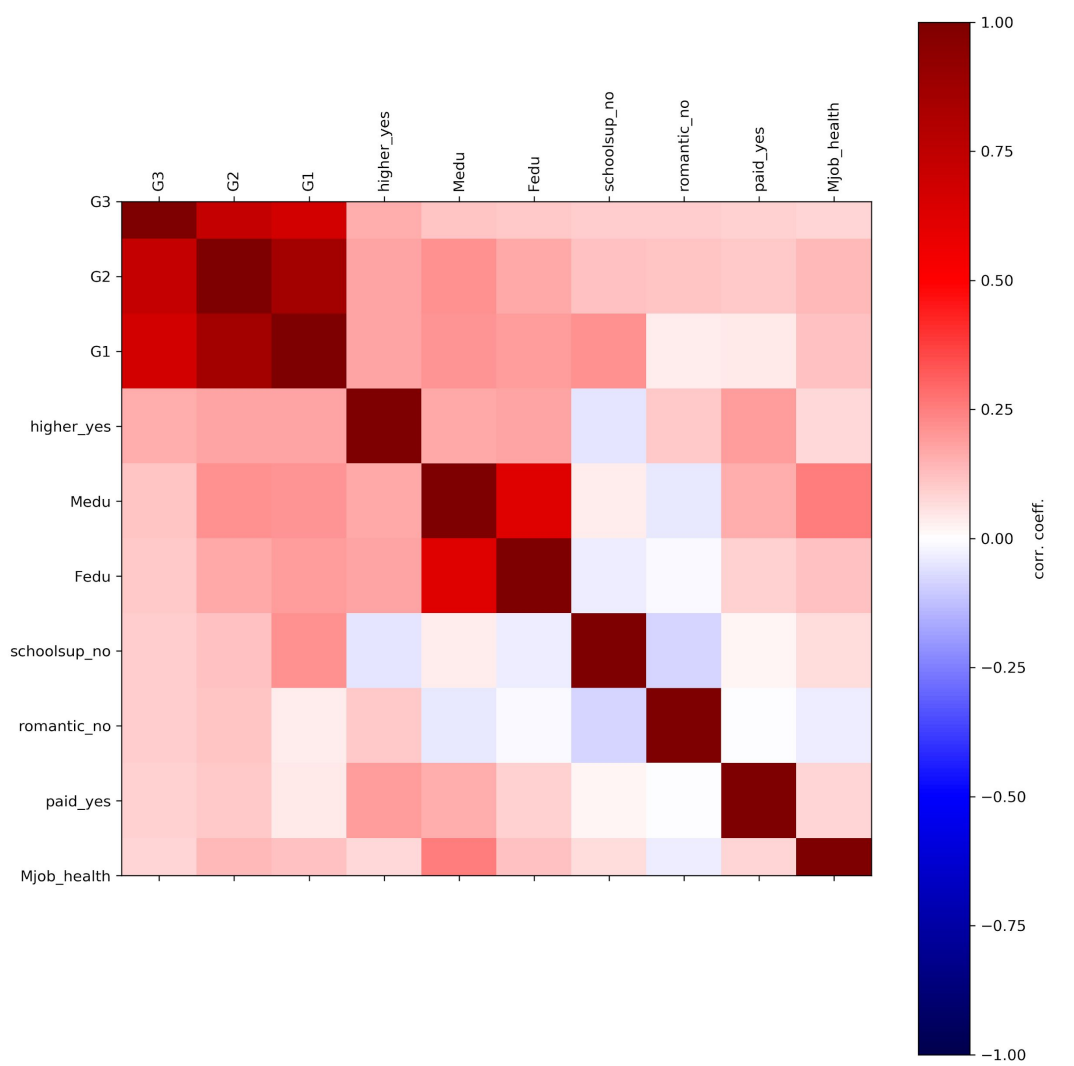
- Classification Problem
- Calculated Balance of Data Set
- Balanced!

EDA:

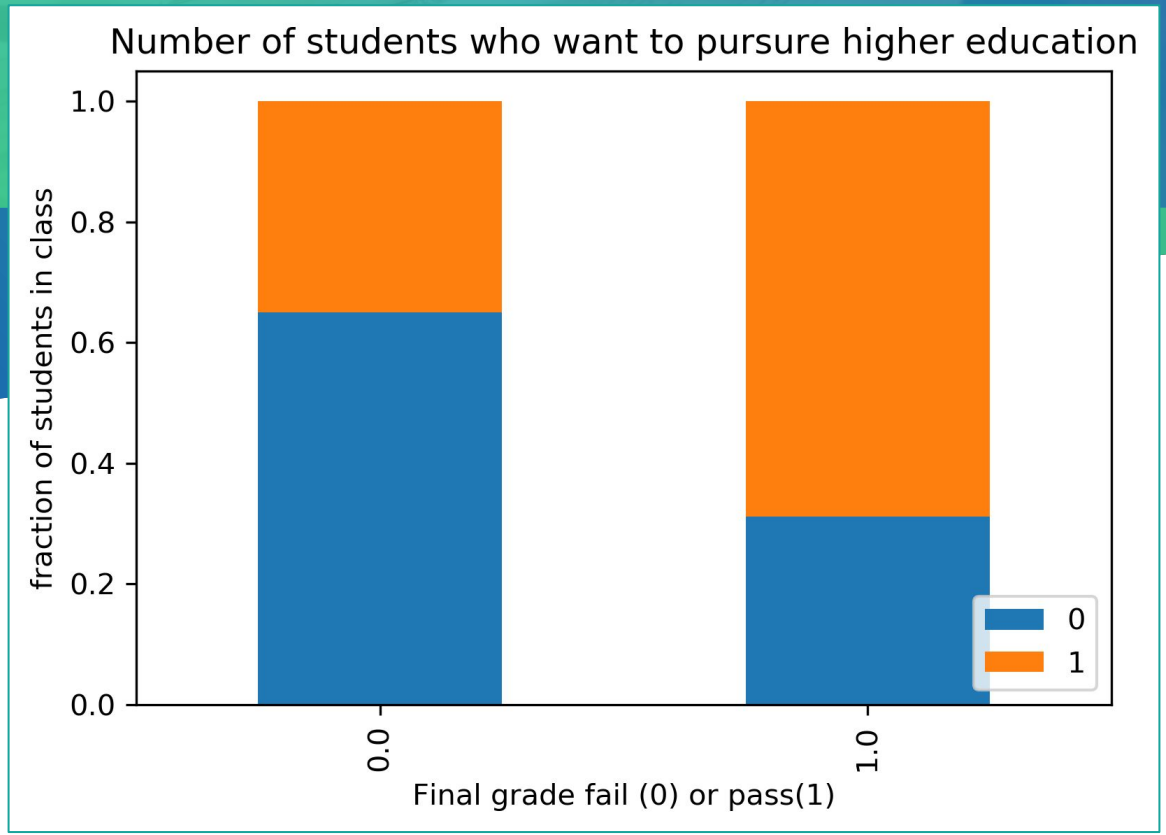

Distribution of G3 grades grouped by G1 (grades in exam 1)
<10
>=10
pdf
0.30
0.25
0.20
0.15
0.10
0.05
0.00
4
6
8
10
12
14
16
18
grades in exam 1


Violin plot grouped by G2 - Grades in exam 2
grade in exam 2
17.5
15.0
12.5
10.0
7.5
5.0
2.5
0.0
<=10
>10
final grade G3

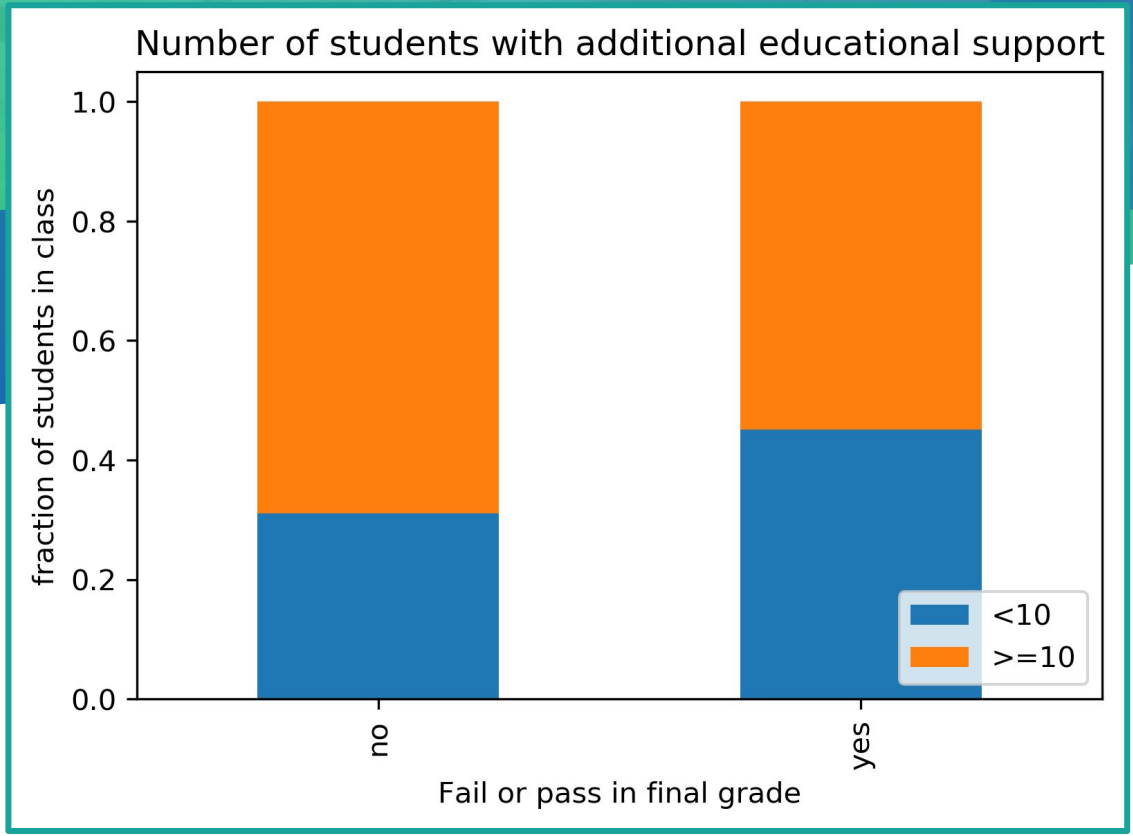
EDA:
Correlation Matrix:



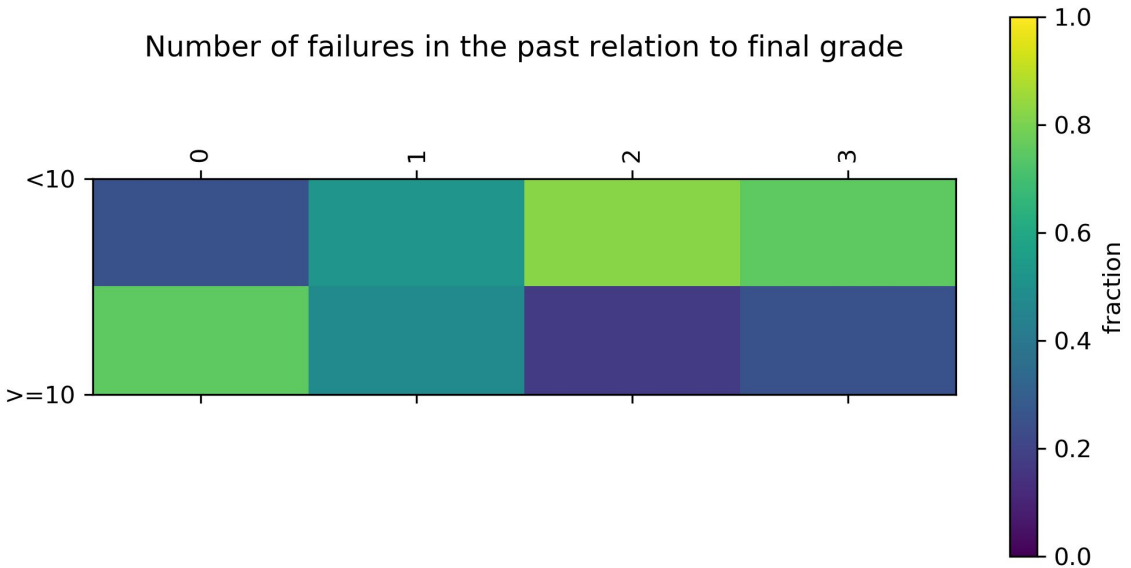
EDA:


Number of students who want to pursure higher education
fraction of students in class
1.0
0.8
0.6
0.4
0.2
0.0
0
1
0.0
1.0
Final grade fail (0) or pass(1)

EDA:


Number of students with additional educational support
fraction of students in class
1.0
0.8
0.6
0.4
0.2
0.0
<10
>=10
no
yes
Fail or pass in final grade

EDA:


Number of failures in the past relation to final grade
0
1
2
3
<10
>=10
1.0
0.8
0.6
0.4
0.2
0.0
fraction

Questions?

Thank you!