

T5 - SDAIA Academy

Students final grades prediction

Sara AlAbdulsalam
Fatimah AlShammari

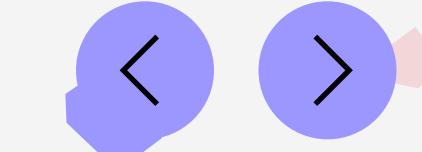


Table of content

01

Introduction

02

Dataset -
EDA

03

Pre-
processing

04

Modeling

05

Conclusion

Introduction



seeks to have an early alert of the performance of students to take actions based on it

Introduction

We will use classification models to predict whether the student will pass or fail



Pass

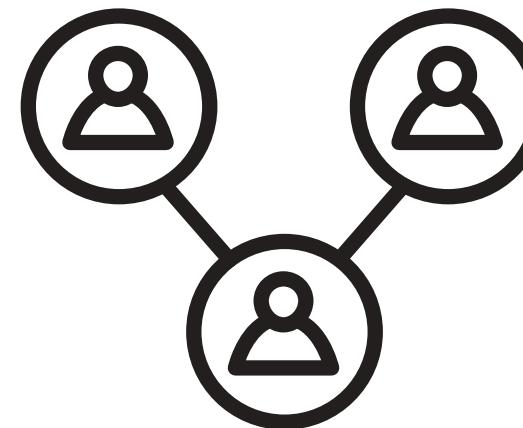
Fail

Dataset

33 Features related to students
395 Observations



Demographic



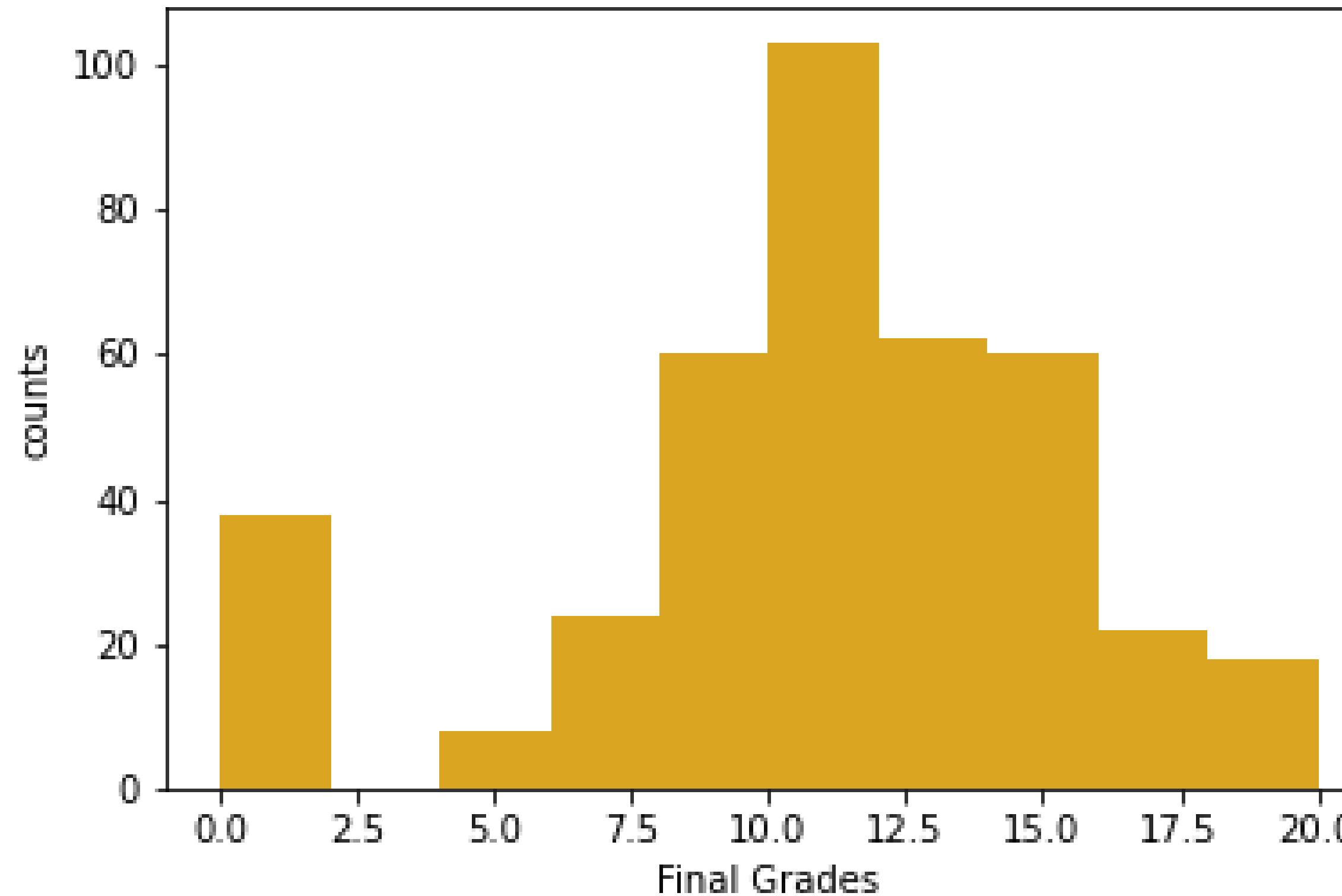
Social



School

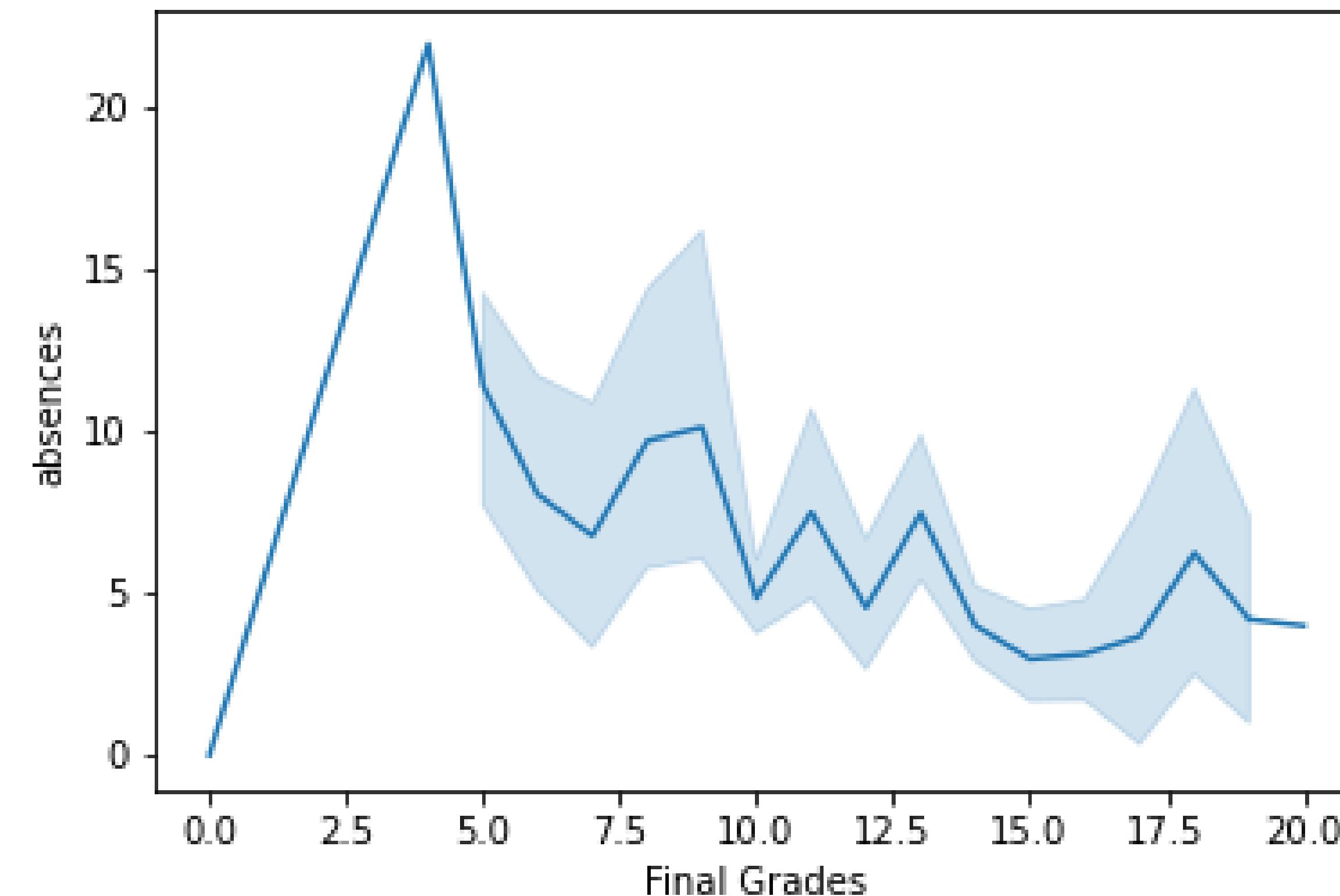
Dataset EDA

Student Final Grade Distribution

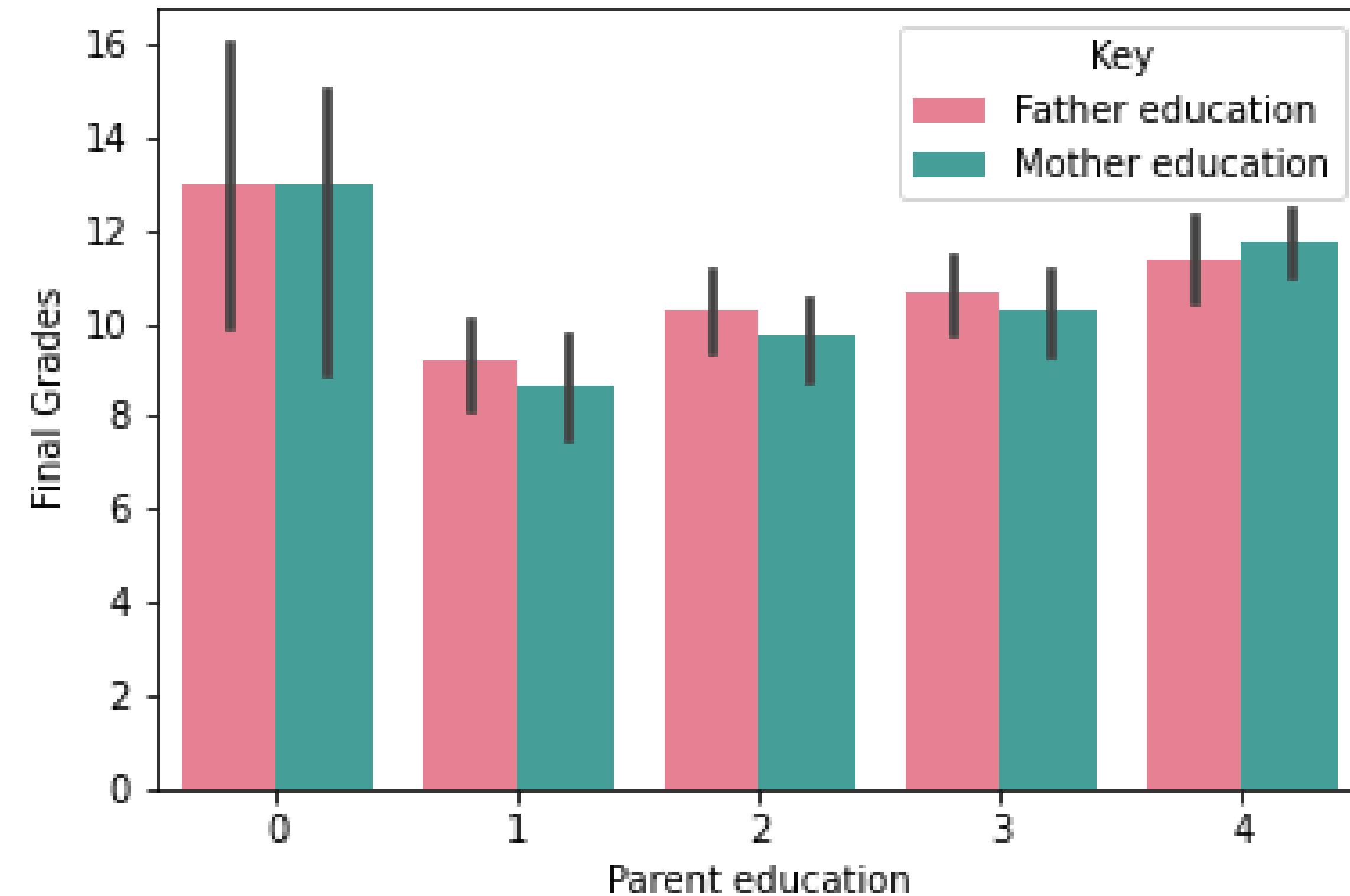


Dataset EDA

Students Grade vs Absences

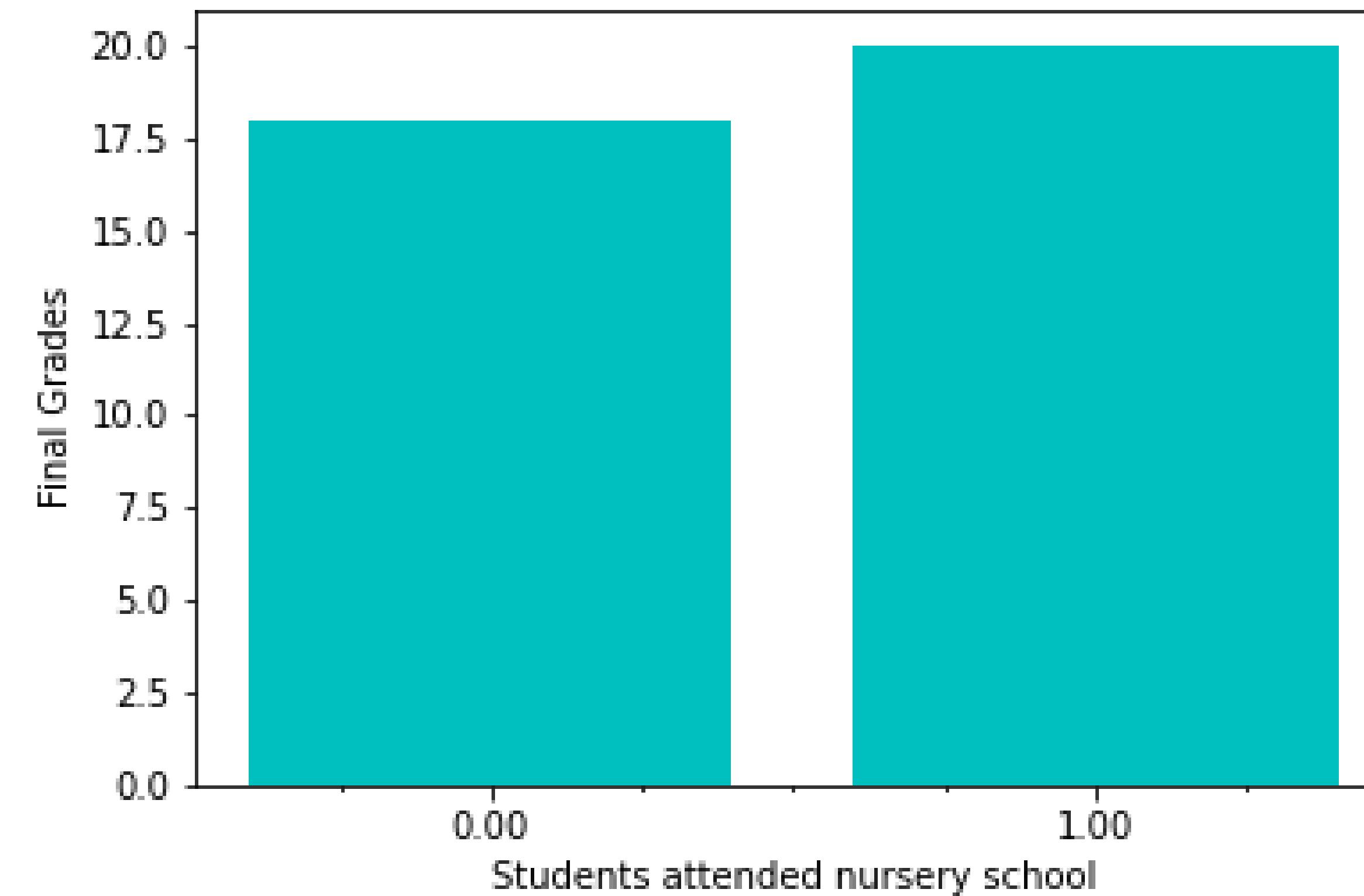


Dataset EDA



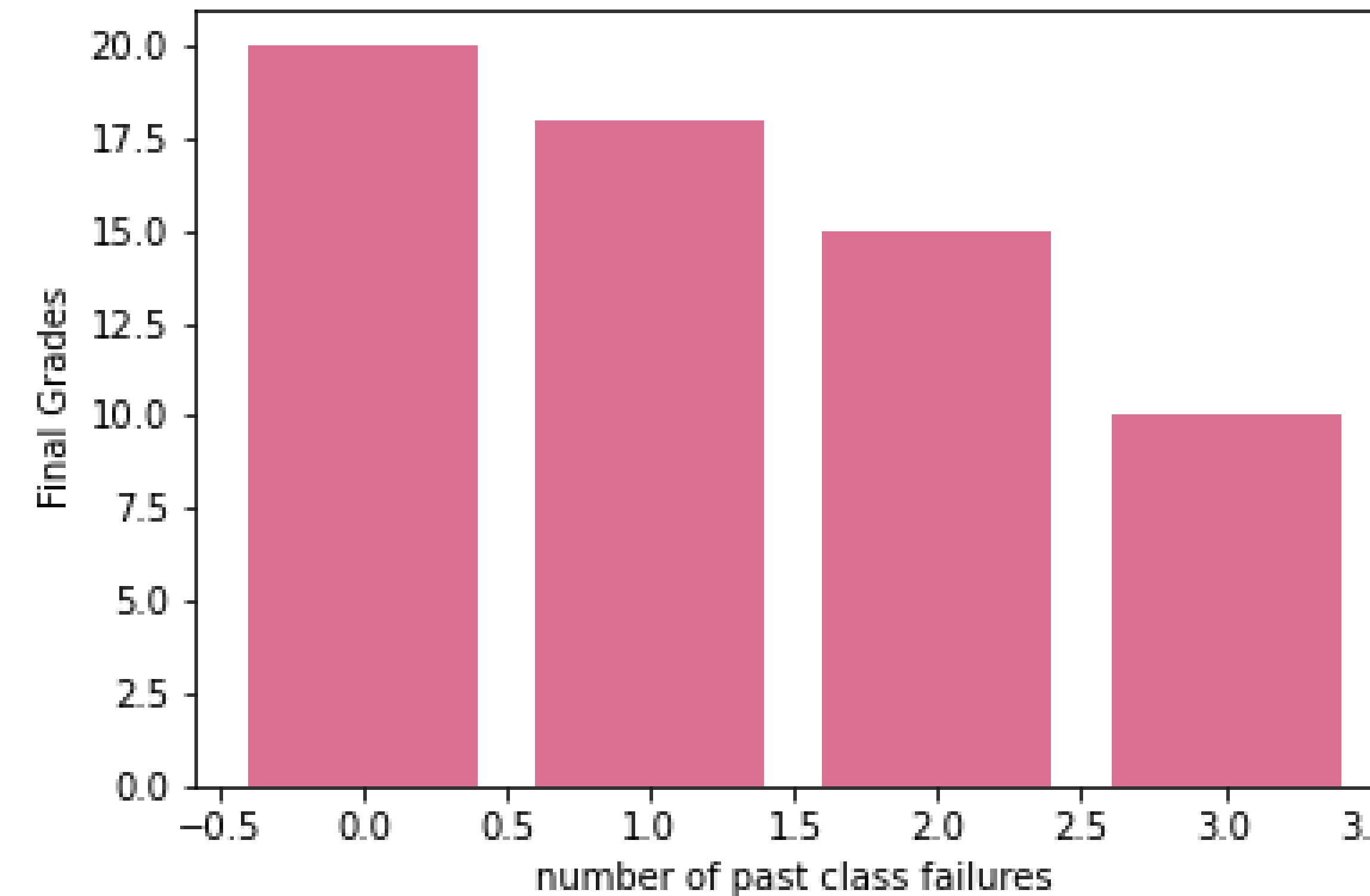
Dataset EDA

Students attended nursery school vs Final Grades



Dataset EDA

Students number of past class failures vs Final Grades



Pre-processing

Feature Engineering

convert the Final Grade to
'Pass' / 0 or 'Fail'/1 feature

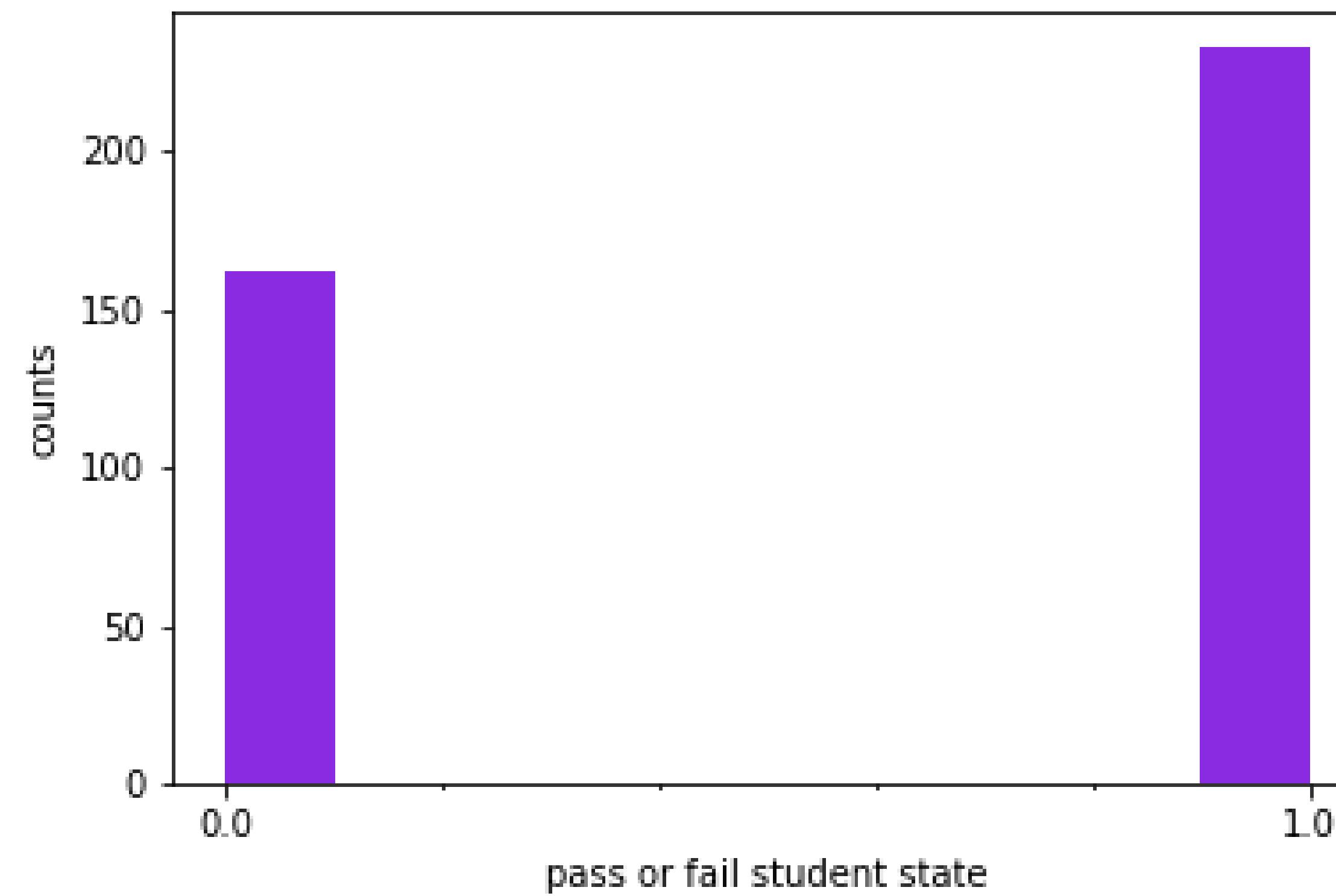
Create a feature that is the
average grade of first and
second period grades

Add dummy variables for the
categorical features

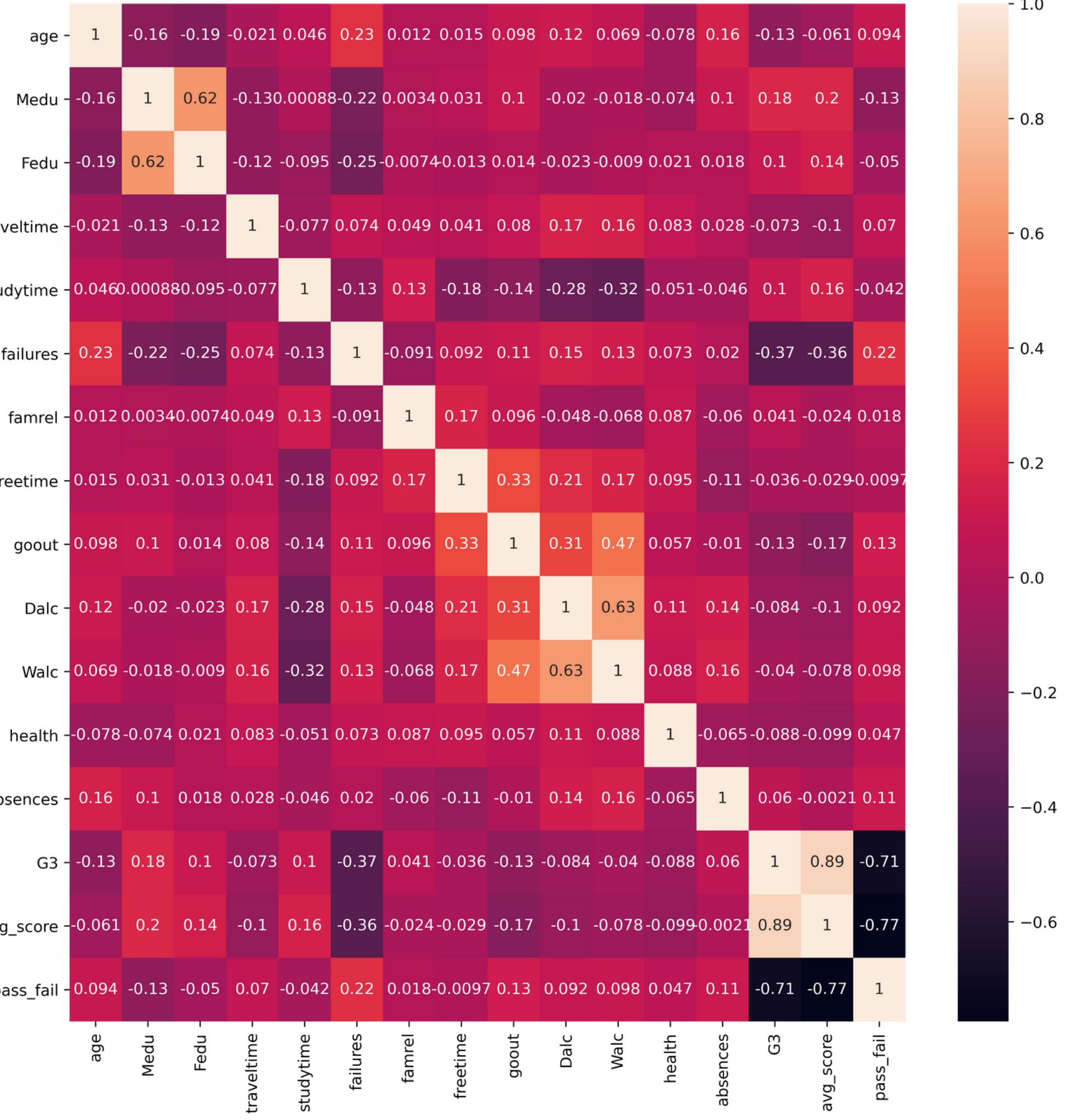
Oversampling the training set

Dataset EDA

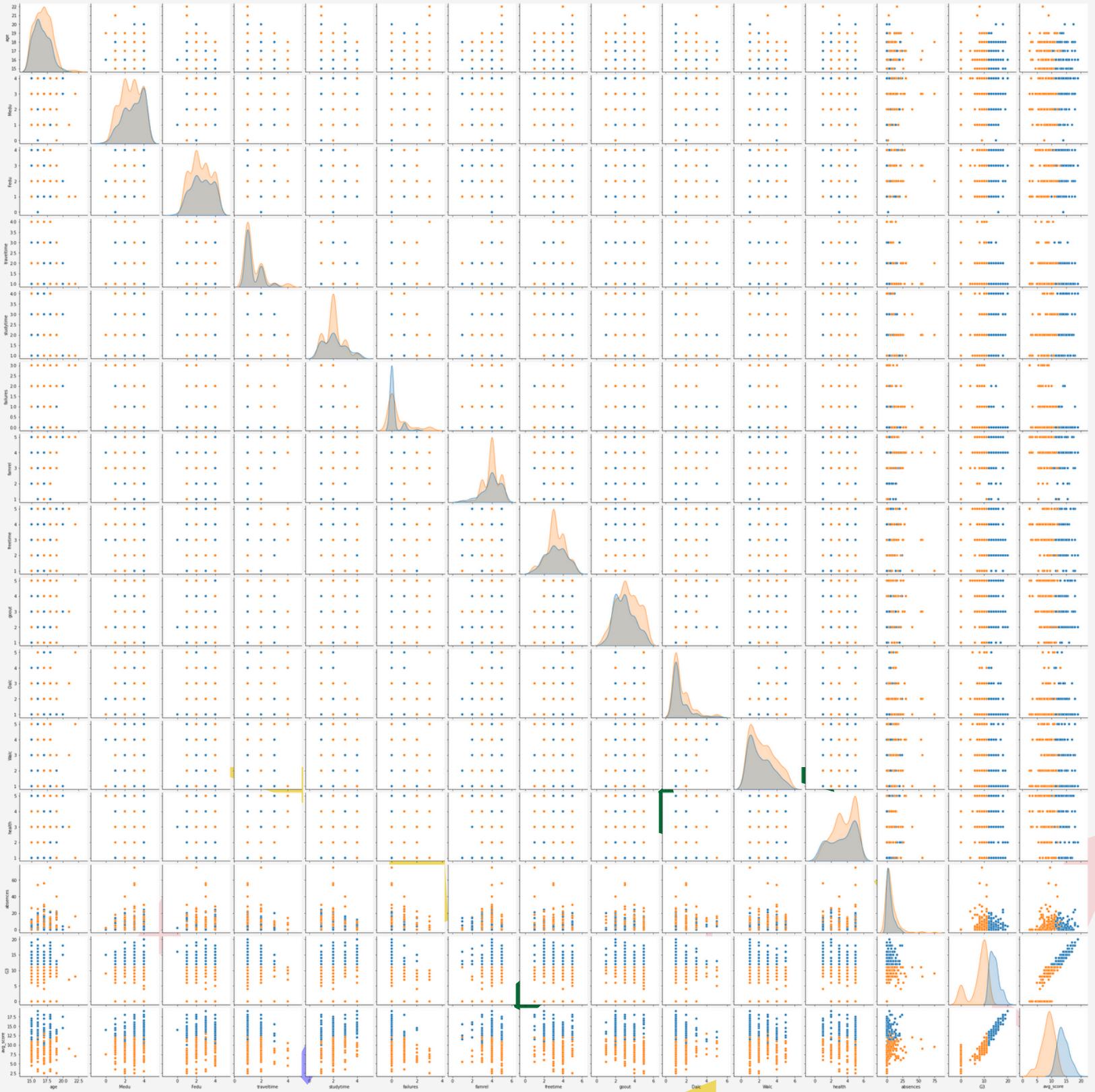
Student State Distribution



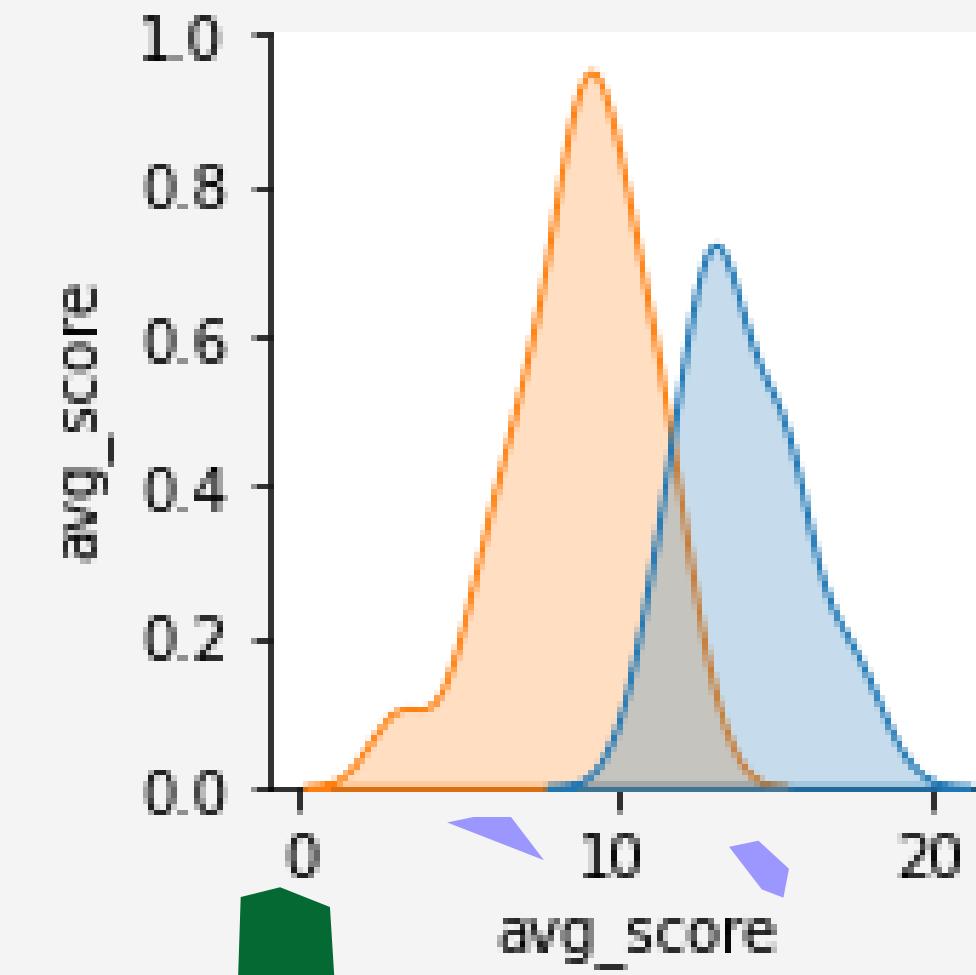
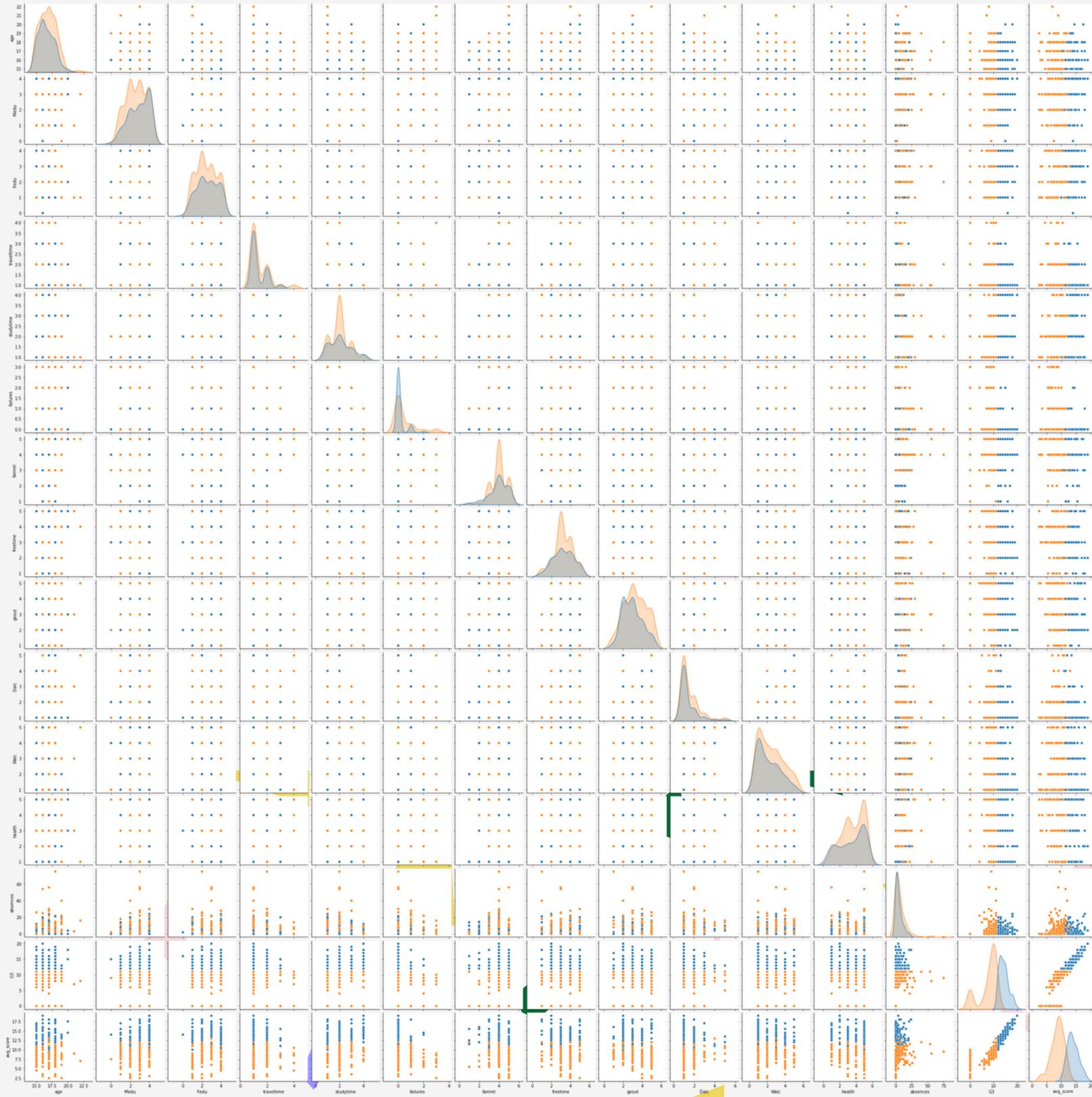
Dataset EDA



Dataset EDA

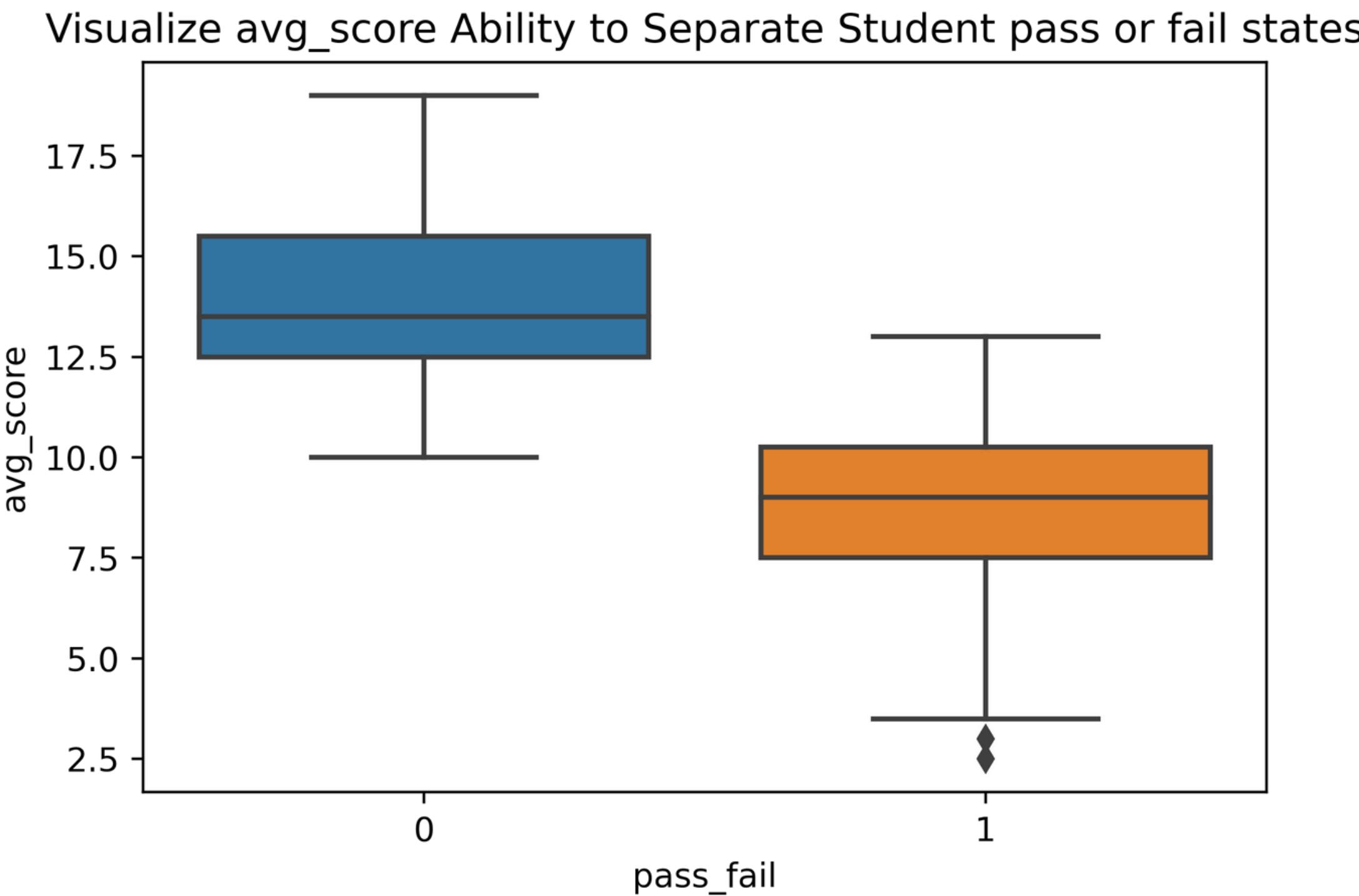


Dataset EDA



`pass_fail`

Dataset EDA



Modeling

Classification - Logistic Regression

Model 1

feature: avg_score

Cross Validation: 10 Folds

Cross Validation score:
91%

Model 2

feature: avg_score,
Fedu, Medu

Cross Validation: 10 Folds

Cross Validation score:
90%

Model 3

feature: All features

Cross Validation: 10 Folds

Cross Validation score:
93%

Model 1

Modeling

Classification - KNN

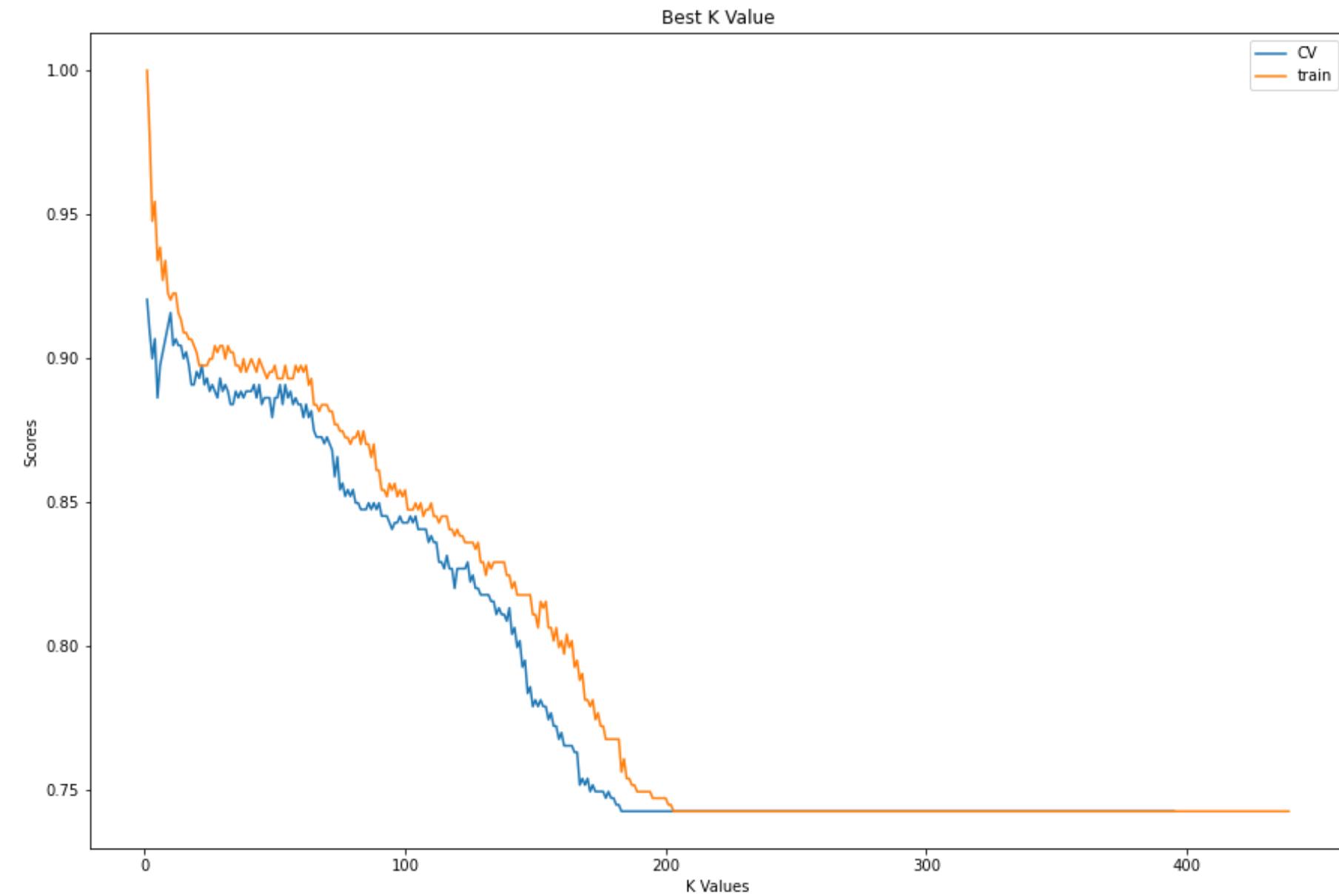
feature: All features

Best K-Value = 1

Cross Validation: 10 Folds

Cross Validation score: 93%

Test set Accuracy: 84%



Model 2

Modeling

Classification - KNN

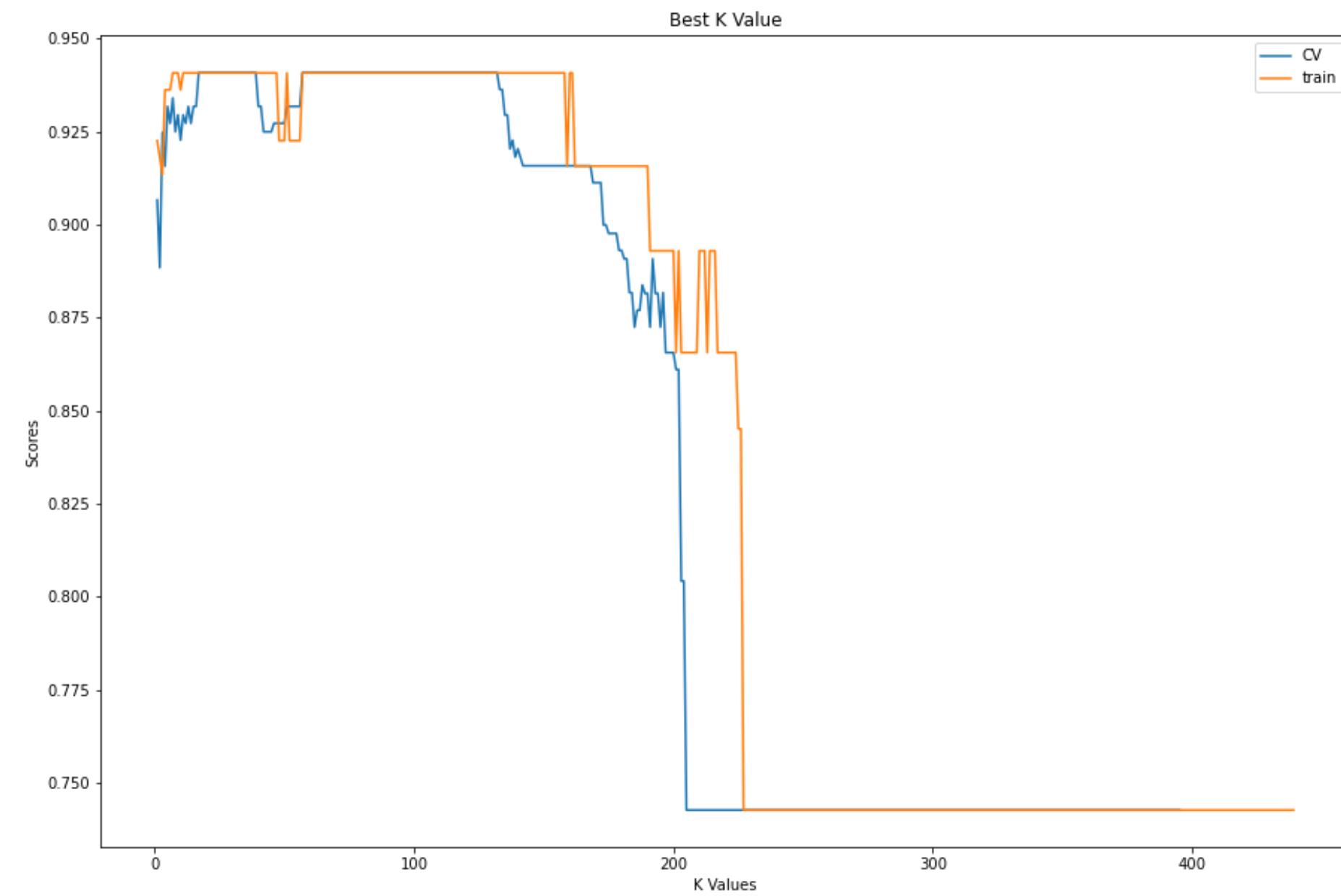
feature: avg_score

Best K-Value = 15

Cross Validation: 10 Folds

Cross Validation score: 92%

Test set Accuracy: 95%



Modeling

Classification - Decision Tree

Model 1

feature: avg_score

Cross Validation: 10 Folds

Cross Validation score:

92%

Model 2

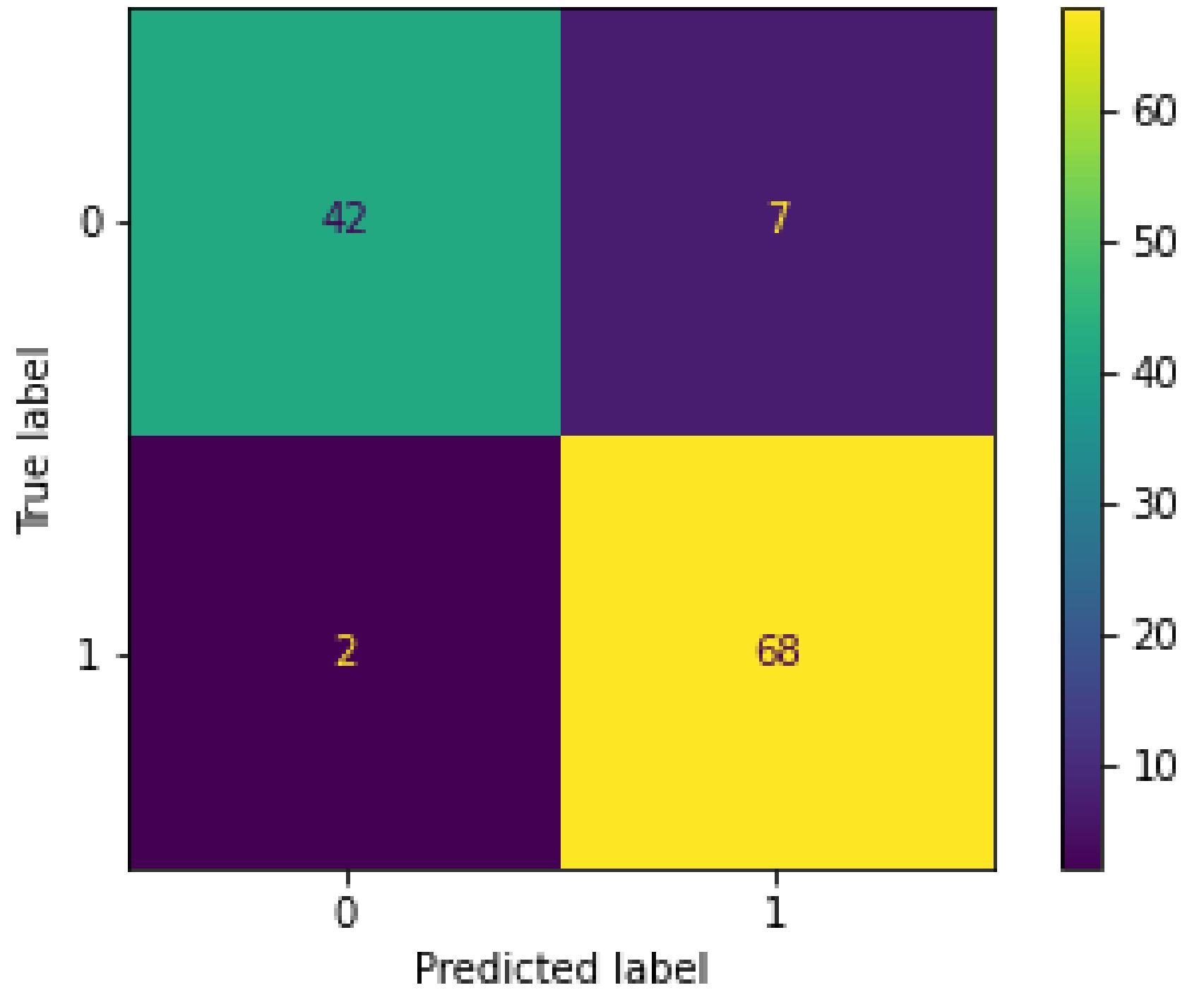
feature: All features

Cross Validation: 10 Folds

Cross Validation score:

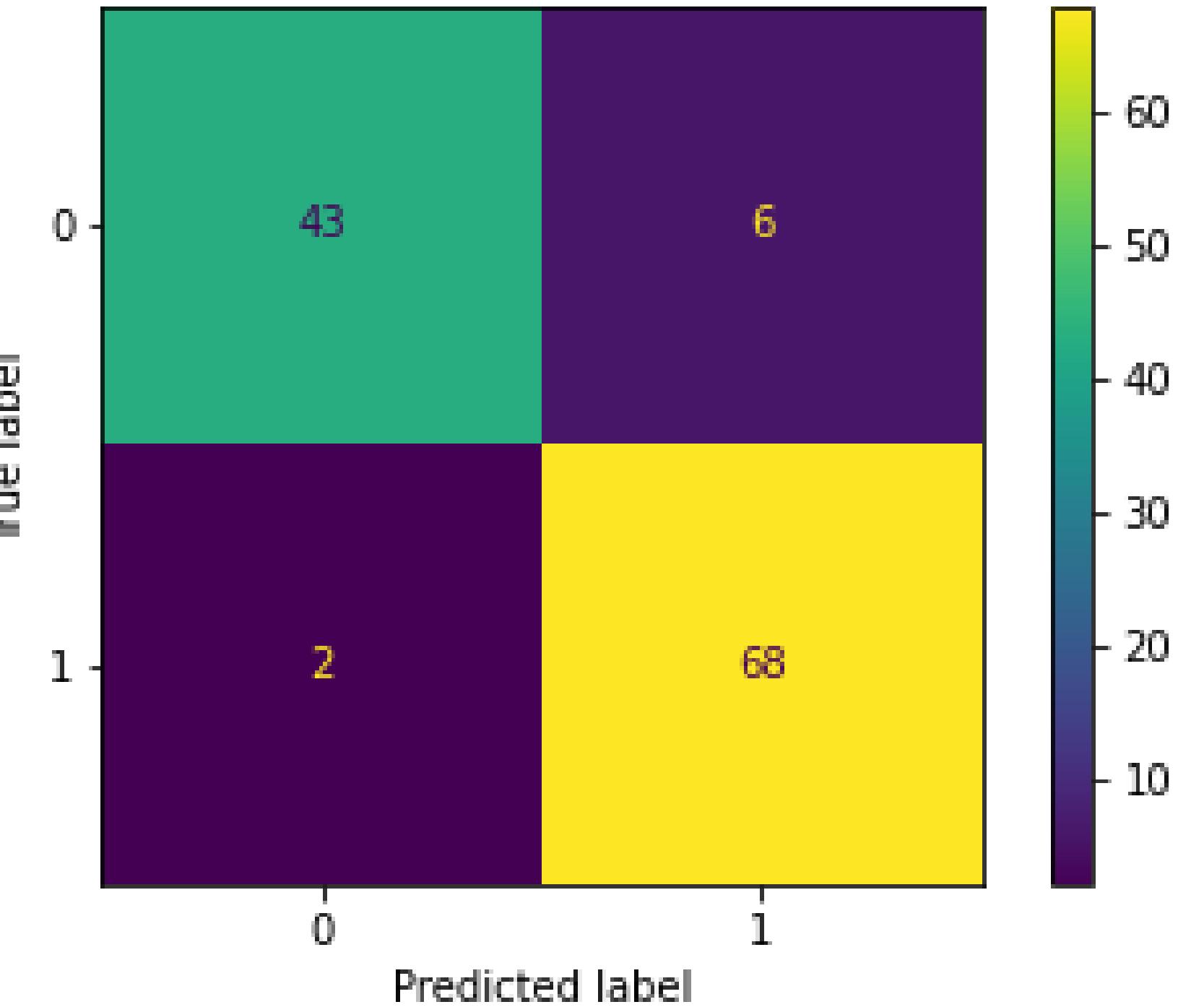
94%

confusion matrix for Logistic Regression



Confusion Matrix
for Logistic
Regression

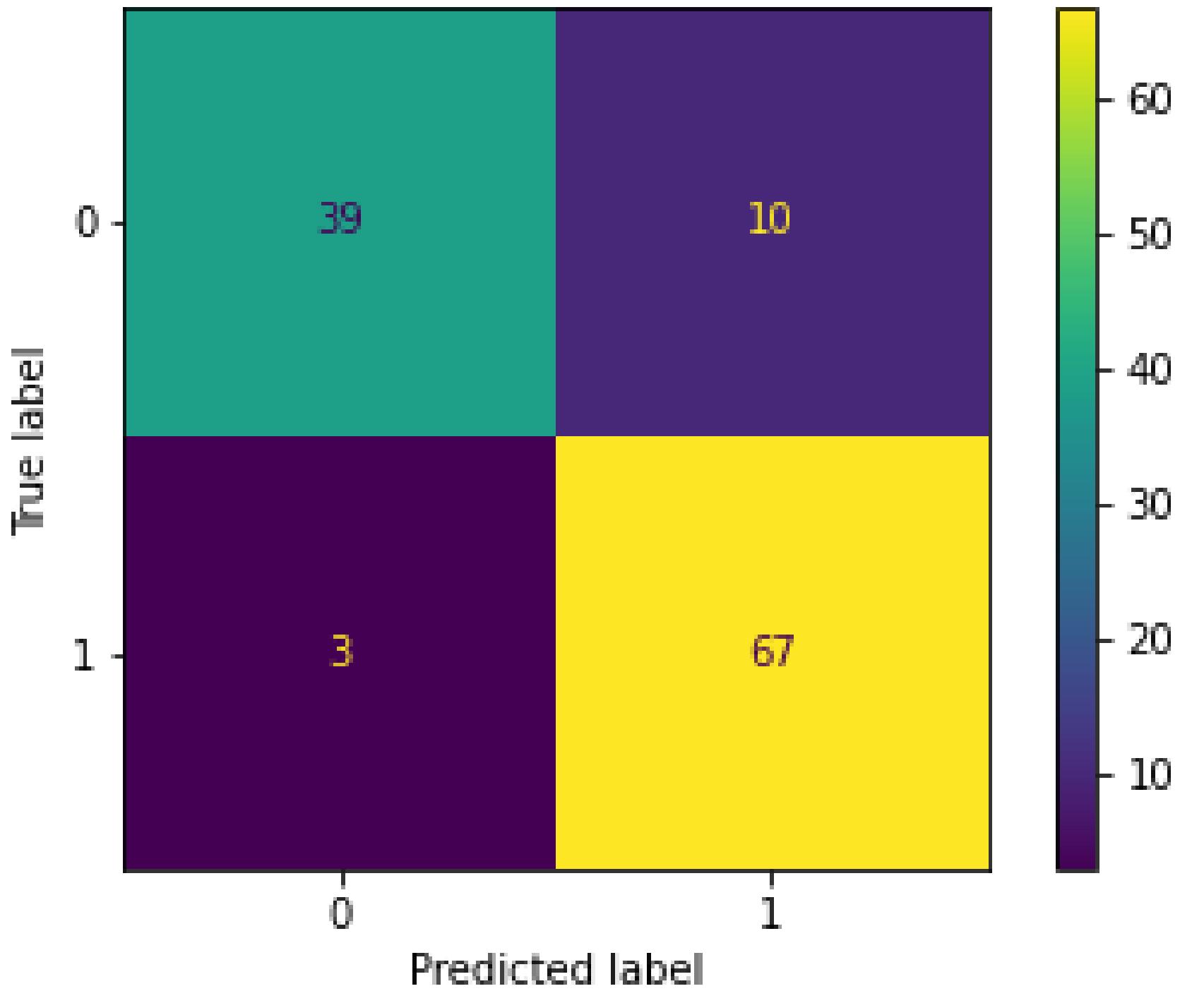
Confusion Matrix for KNN



<

Confusion Matrix
for KNN

confusion matrix for Decision Tree



<

Confusion Matrix
for Decision Tree

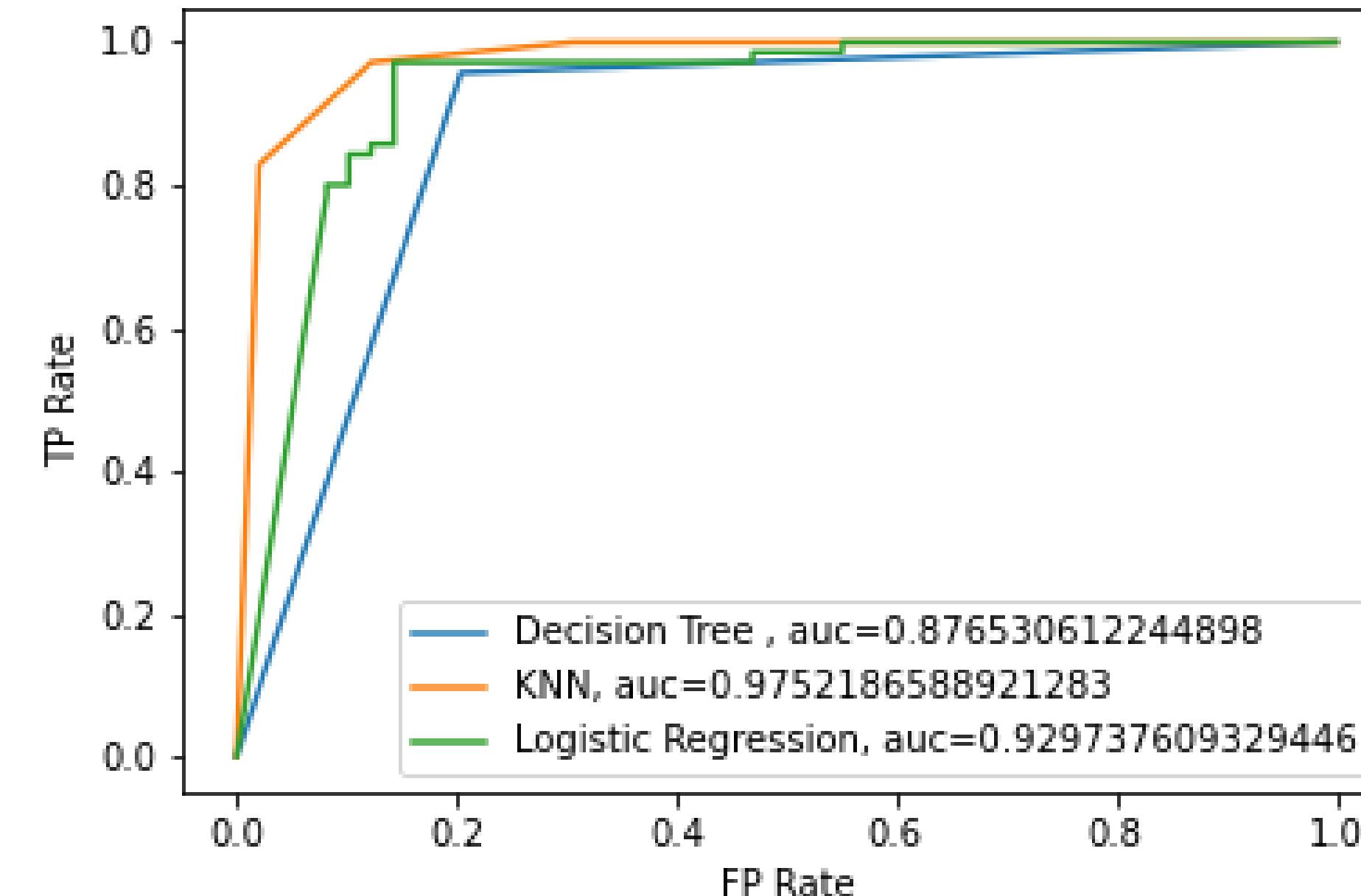
Results Comparison

Machine Learning Classifiers	Accuracy	Recall	Precision
Logistic Regression	92%	97%	95%
K-Nearest Neighbors	93%	97%	96%
Decision Tree	89%	96%	93%

Results Comparison

ROC-AUC

ROC-AUC



Conclusion

Logistic Regression



KNN

1

2

Decision Tree

3

Fatimah AlShammari
Sara AlAbdulsalam

Thank You

