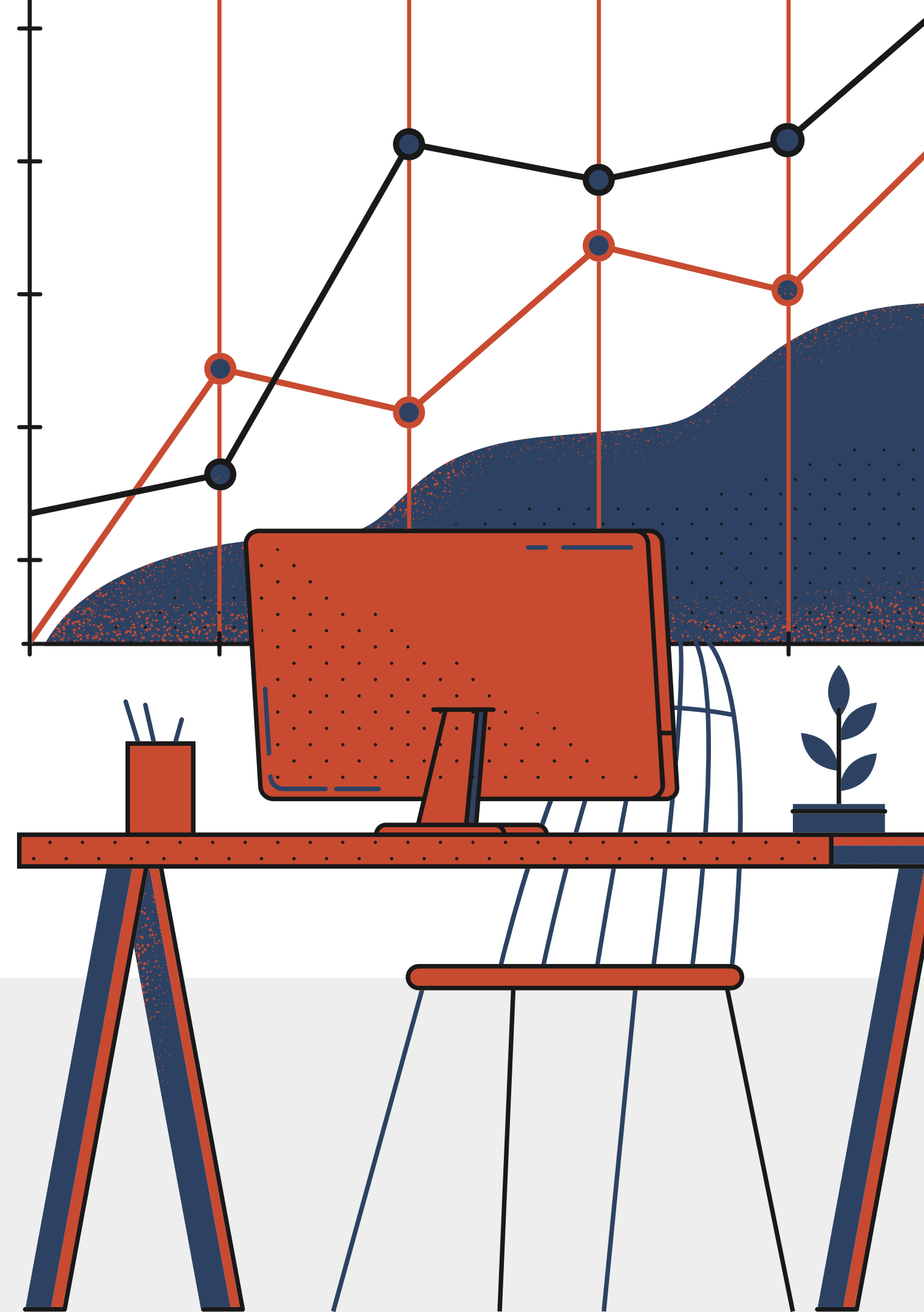


Student Performance Prediction

RATNA PERMATASARI



- I Background
- II Dataset Information
- III Exploratory Data Analysis
- IV Data Preprocessing
- V Modelling
- VI Conclusion & Recommendation

Outline

Background



- High school is an important part of the education system in many countries because it helps students acquire knowledge and skills needed for higher education and workplace and prepare them to face new challenges.
- A school with a good reputation and many high-achieving students will attract attention from parents enrolling their children in the school, donors, the government, and other institutions that want to support the school. This can increase the funds available to schools, increasing school profits.

Dataset Information

Attribute	Description (Domain)
sex	student's sex (binary: female or male)
age	student's age (numeric: from 15 to 22)
school	student's school (binary: <i>Gabriel Pereira</i> or <i>Mousinho da Silveira</i>)
address	student's home address type (binary: urban or rural)
Pstatus	parent's cohabitation status (binary: living together or apart)
Medu	mother's education (numeric: from 0 to 4 ^a)
Mjob	mother's job (nominal ^b)
Fedu	father's education (numeric: from 0 to 4 ^a)
Fjob	father's job (nominal ^b)
guardian	student's guardian (nominal: mother, father or other)
famsize	family size (binary: ≤ 3 or > 3)
famrel	quality of family relationships (numeric: from 1 – very bad to 5 – excellent)
reason	reason to choose this school (nominal: close to home, school reputation, course preference or other)
traveltime	home to school travel time (numeric: 1 – < 15 min., 2 – 15 to 30 min., 3 – 30 min. to 1 hour or 4 – > 1 hour).
studytime	weekly study time (numeric: 1 – < 2 hours, 2 – 2 to 5 hours, 3 – 5 to 10 hours or 4 – > 10 hours)
failures	number of past class failures (numeric: n if $1 \leq n < 3$, else 4)
schoolsup	extra educational school support (binary: yes or no)
famsup	family educational support (binary: yes or no)
activities	extra-curricular activities (binary: yes or no)
paidclass	extra paid classes (binary: yes or no)
internet	Internet access at home (binary: yes or no)
nursery	attended nursery school (binary: yes or no)
higher	wants to take higher education (binary: yes or no)
romantic	with a romantic relationship (binary: yes or no)
freetime	free time after school (numeric: from 1 – very low to 5 – very high)
goout	going out with friends (numeric: from 1 – very low to 5 – very high)
Walc	weekend alcohol consumption (numeric: from 1 – very low to 5 – very high)
Dalc	workday alcohol consumption (numeric: from 1 – very low to 5 – very high)
health	current health status (numeric: from 1 – very bad to 5 – very good)
absences	number of school absences (numeric: from 0 to 93)
G1	first period grade (numeric: from 0 to 20)
G2	second period grade (numeric: from 0 to 20)
G3	final grade (numeric: from 0 to 20)

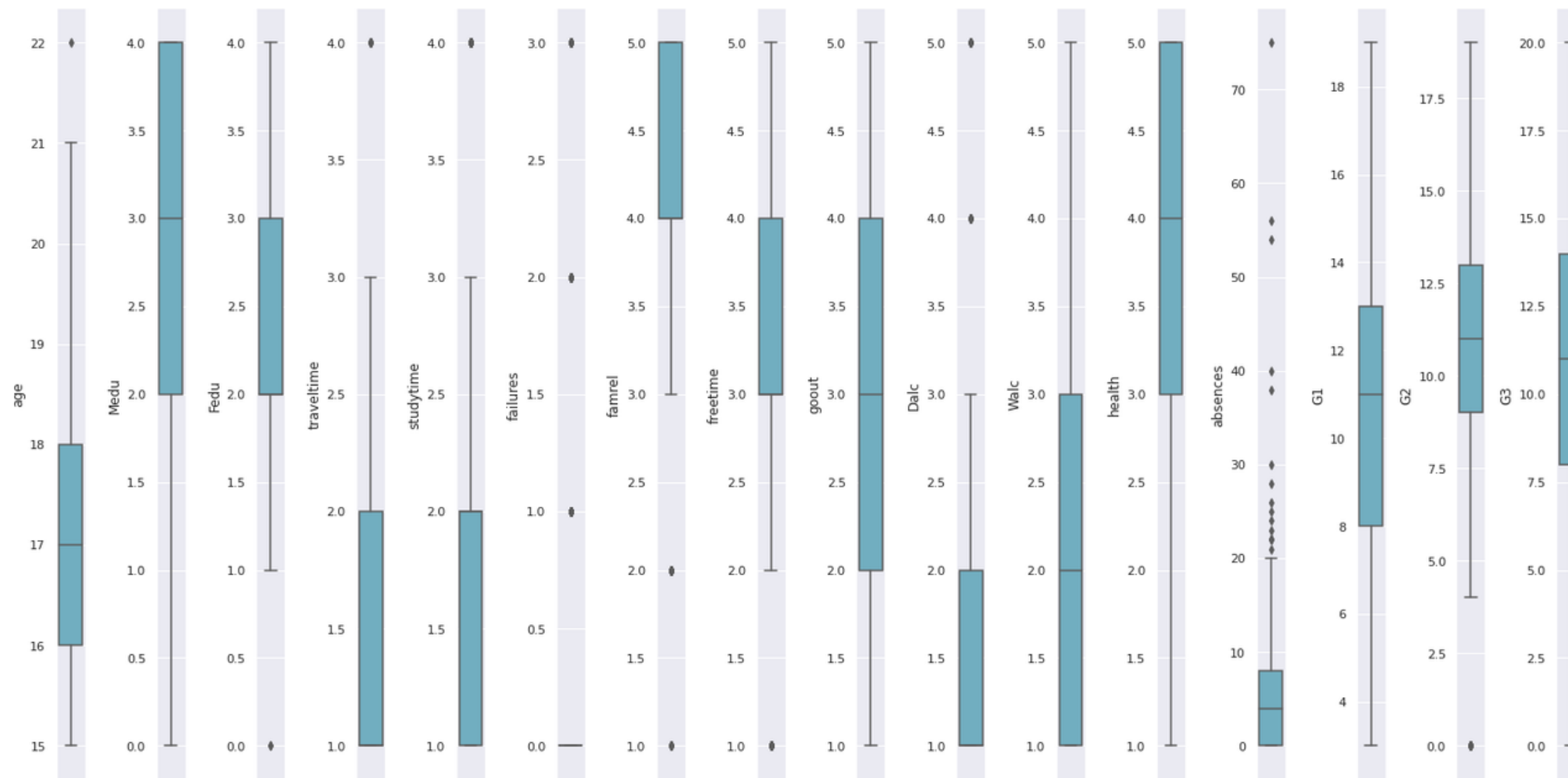
^a 0 – none, 1 – primary education (4th grade), 2 – 5th to 9th grade, 3 – secondary education or 4 – higher education.

^b teacher, health care related, civil services (e.g. administrative or police), at home or other.

- The dataset has 33 columns and 395 rows
- There are 17 categorical features and 16 numerical features
- There are no missing and no duplicate values

Exploratory Data Analysis

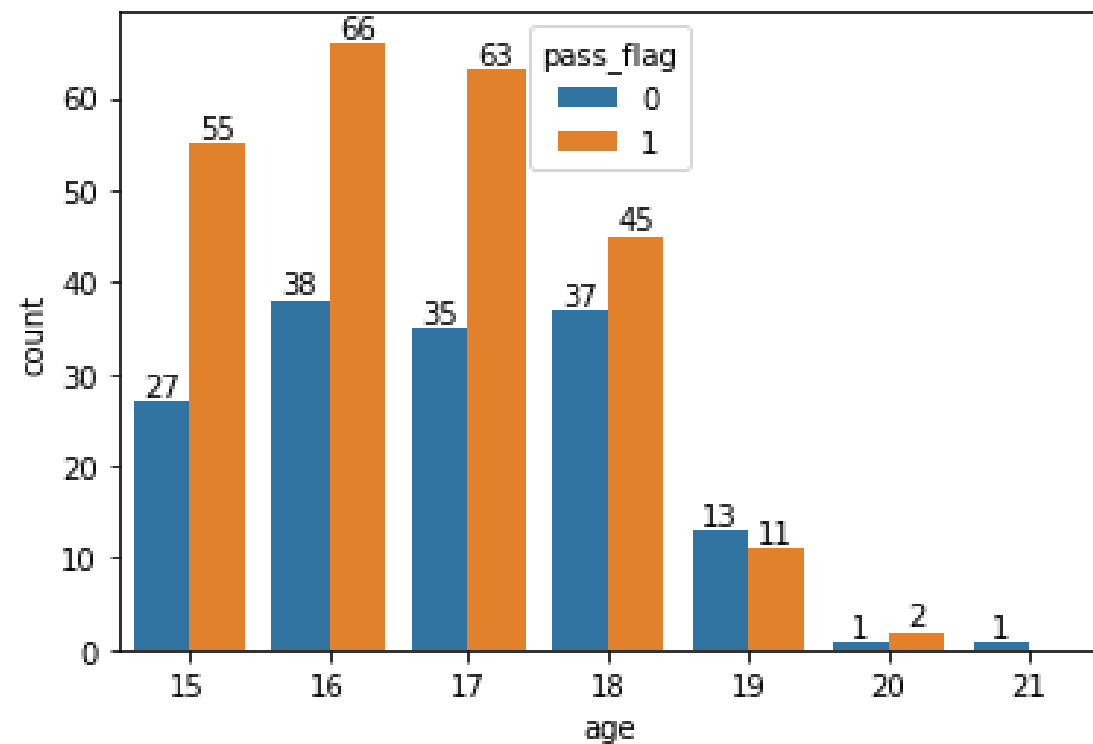
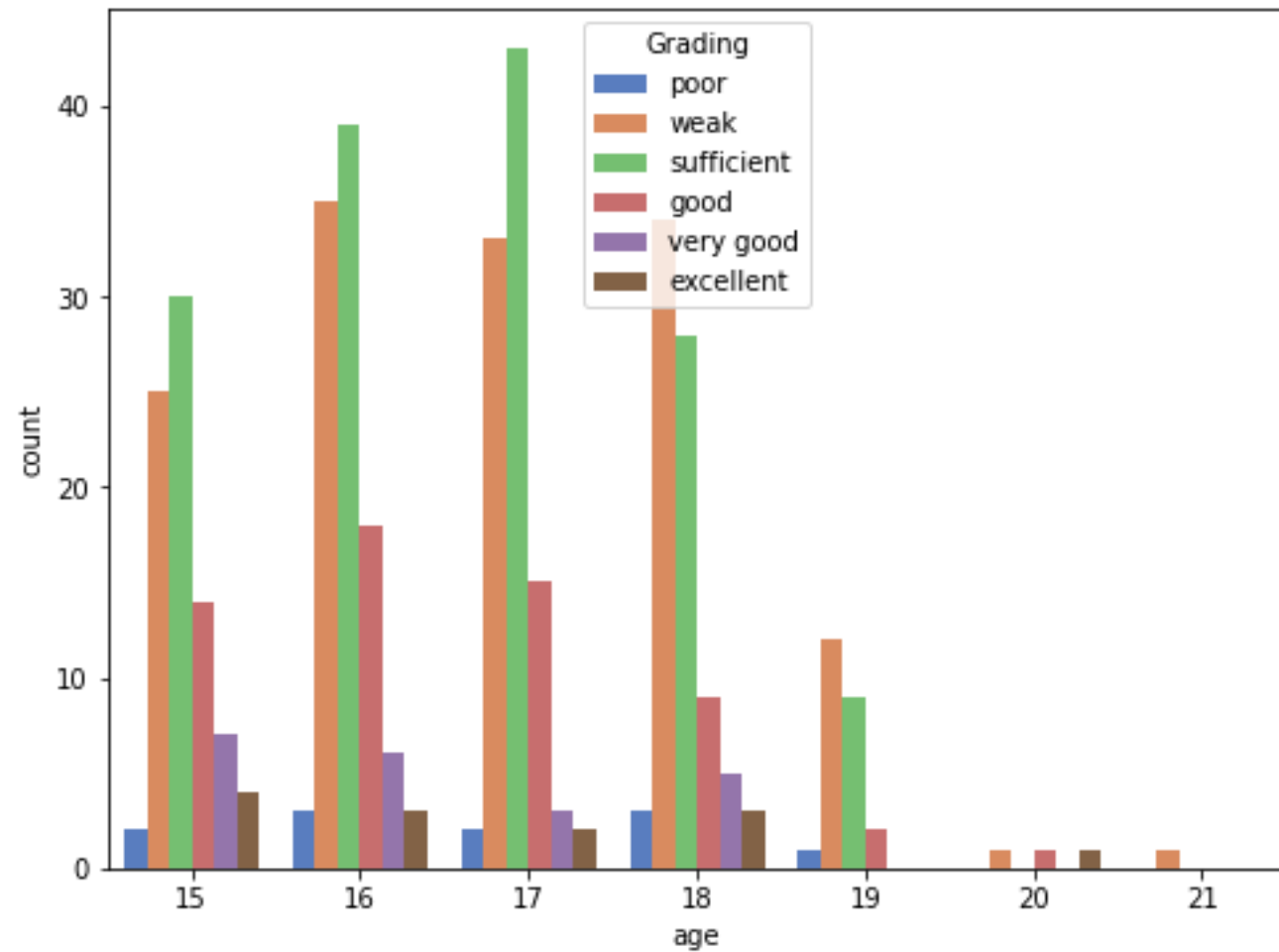
Outlier



- Find outliers with IQR method
- Remove outlier on age because 22 years old is too old for high school student
- Outliers on absences seem reasonable due to the accumulation of absences during the school year
- Other outliers are still reasonable

	school	sex	age	address	famsize	Pstatus	Medu	Fedu	Mjob	Fjob	...	goout	Dalc	Walc	health	absences	G1	G2	G3	G_avg	Grading
247	GP	M	22	U	GT3	T	3	1	services	services	...	5	5	5	1	16	6	8	8	7.333333	weak

1 rows × 35 columns



High schools and universities in Portugal use a 20-point rating scale. Qualification for each grade is as follows:

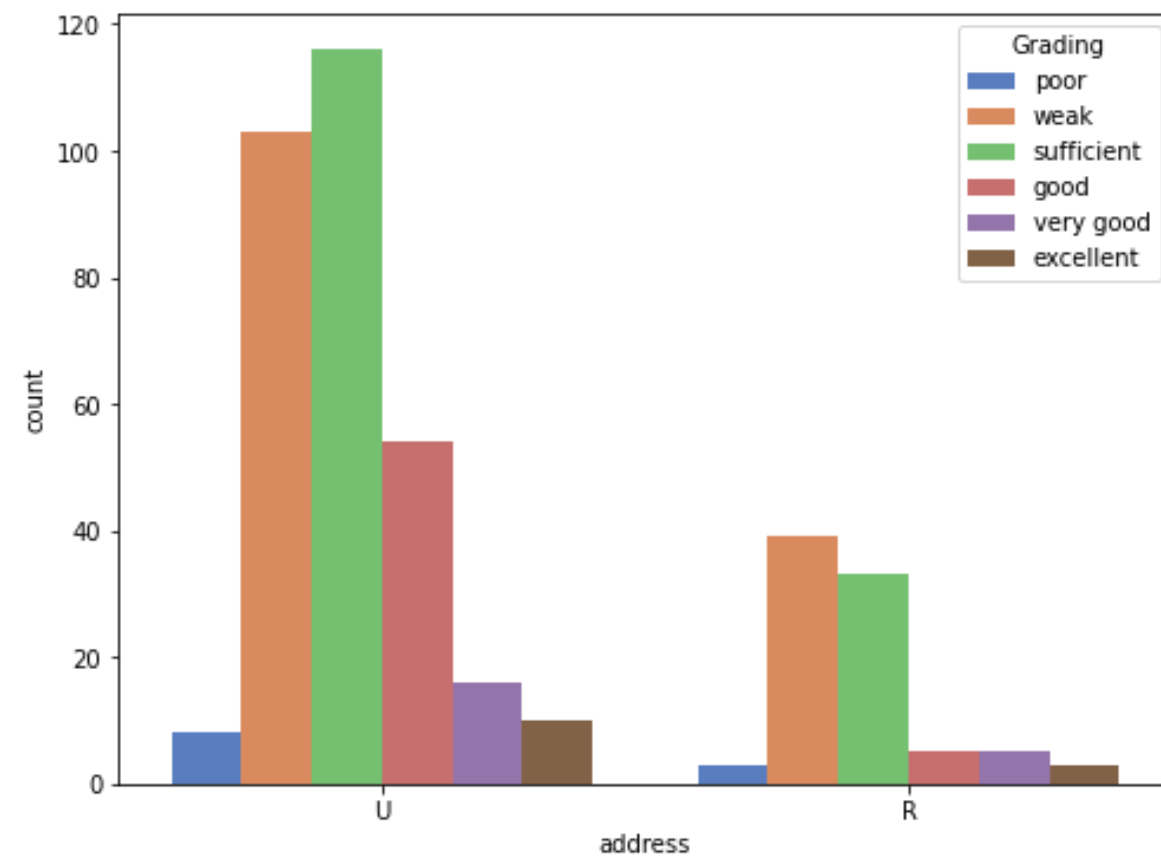
- 0 – 3.4 : Poor
- 3.5 – 9.4 : Weak
- 9.5 – 13.4 : Sufficient
- 13.5 – 15.4 : Good
- 15.5 – 17.4 : Very Good
- 17.5 – 20 : Excellent

Add New Features

If a student's score is ≥ 9.5 , they pass = 1. otherwise, they fail = 0

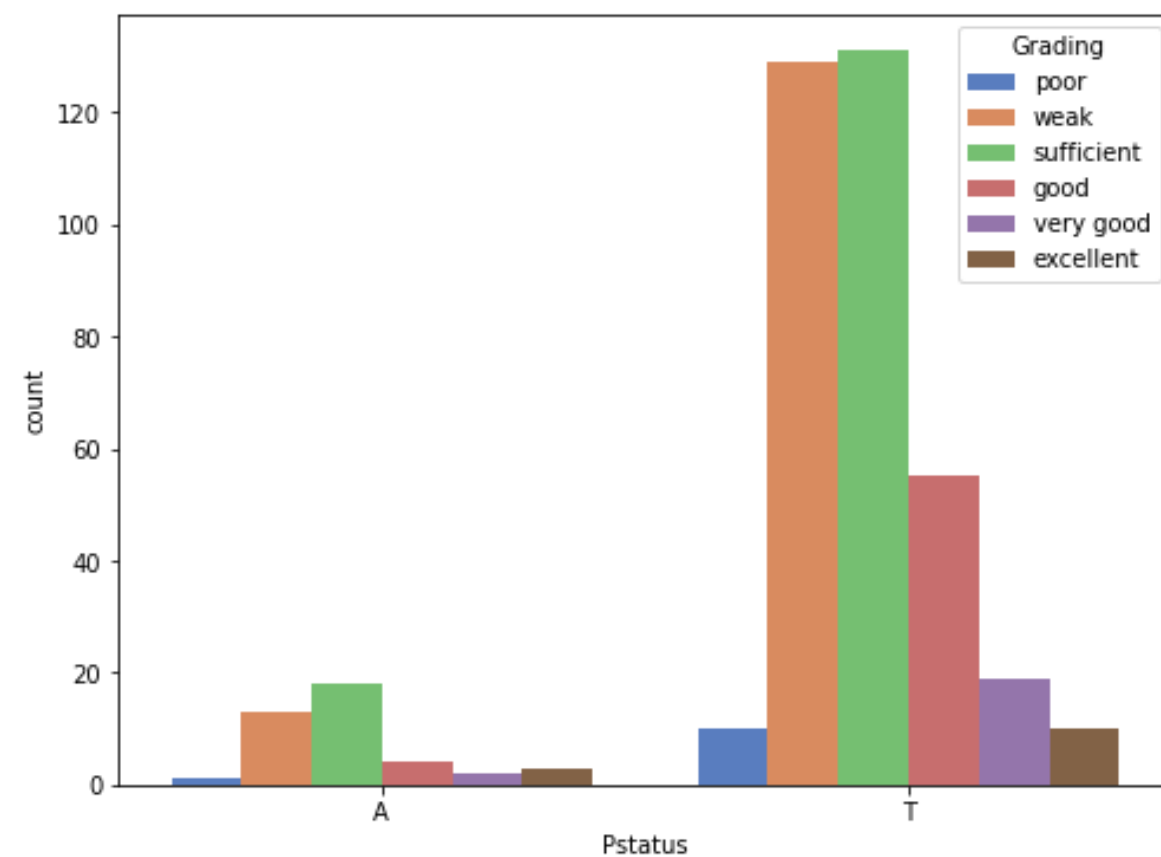
Age

From age of 18, there is a decrease in the number of students passing, which may be due to students experiencing difficulty in following the challenging lessons in Grade 3 of high school.



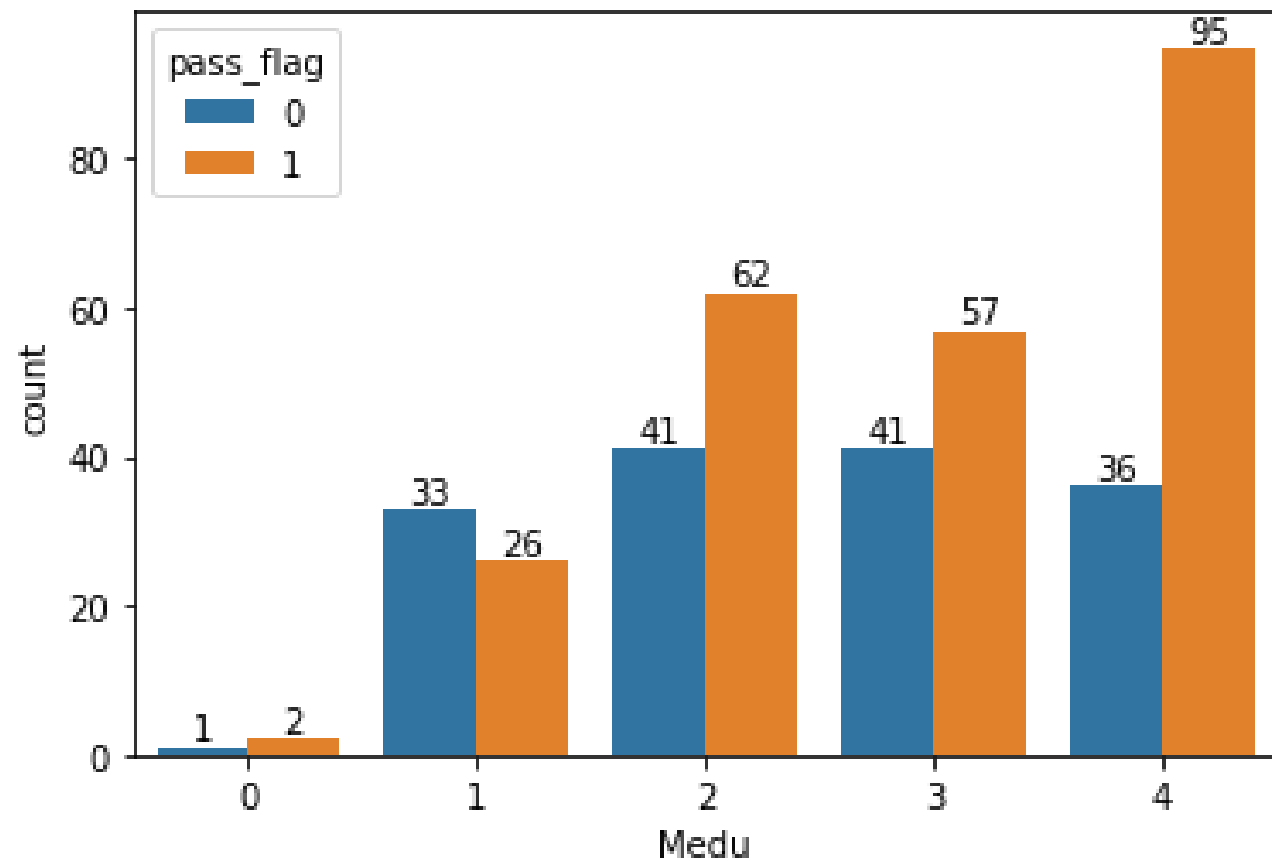
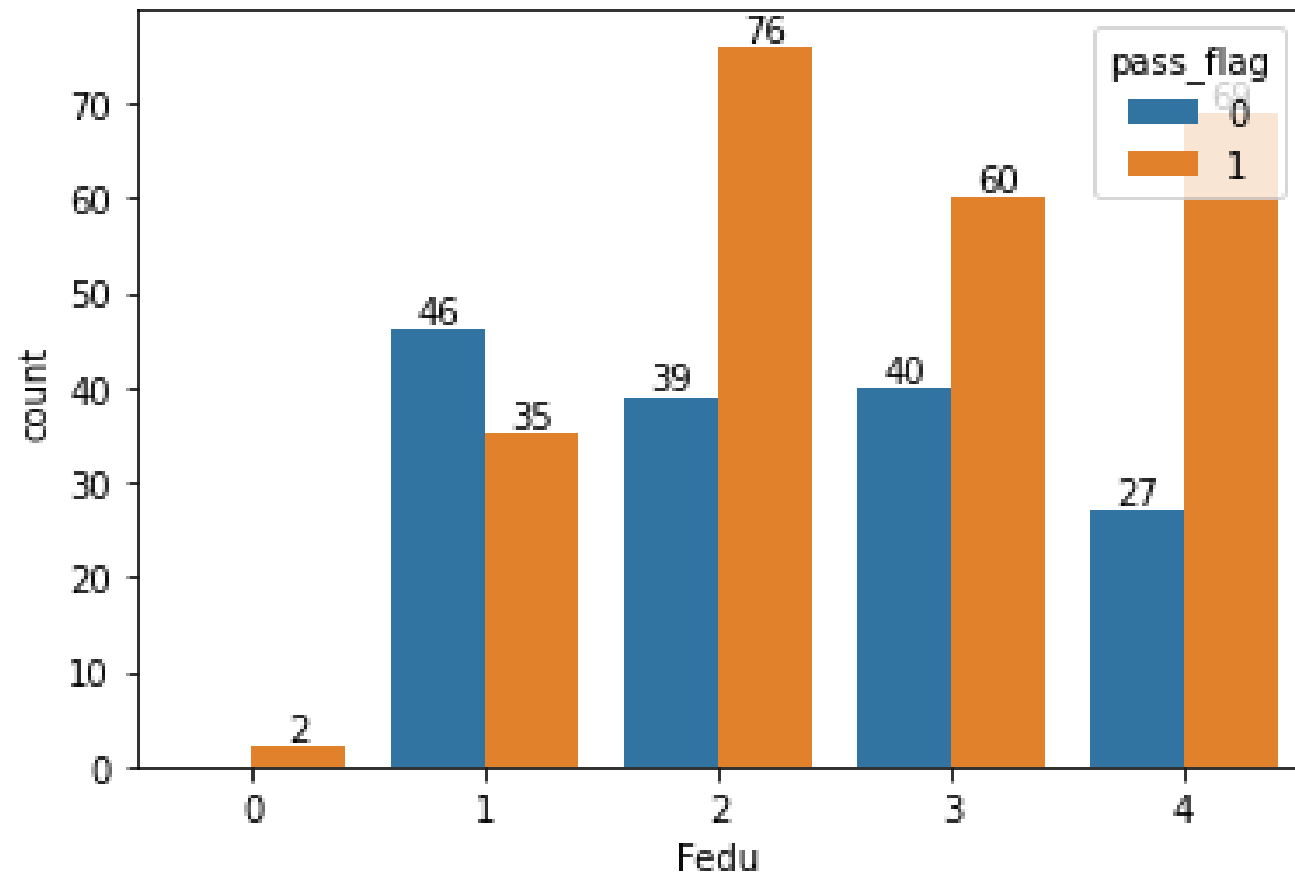
Address

Students who live in cities tend to have higher scores and pass the exam more than students who live in rural areas because they have better access to education and facilities.



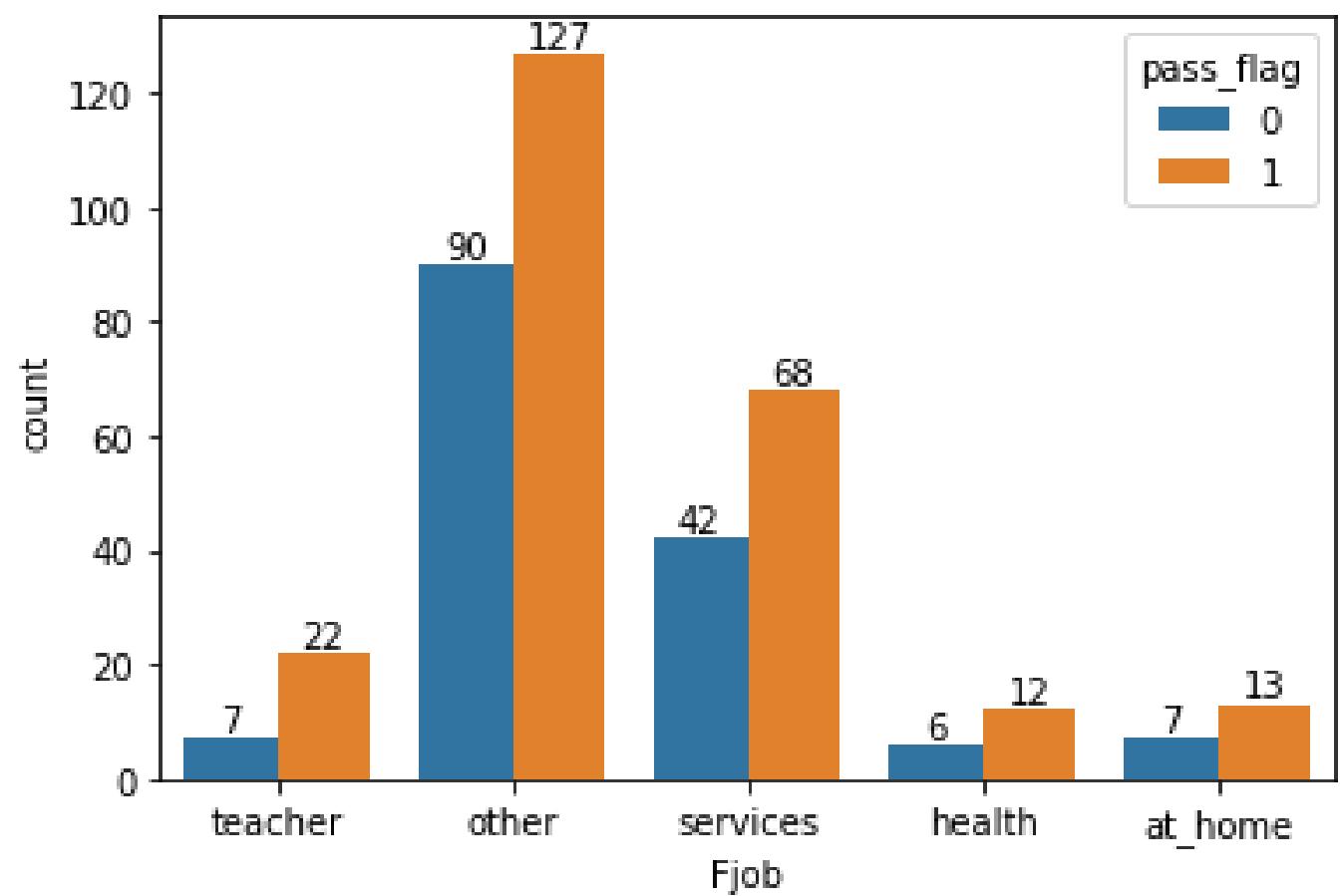
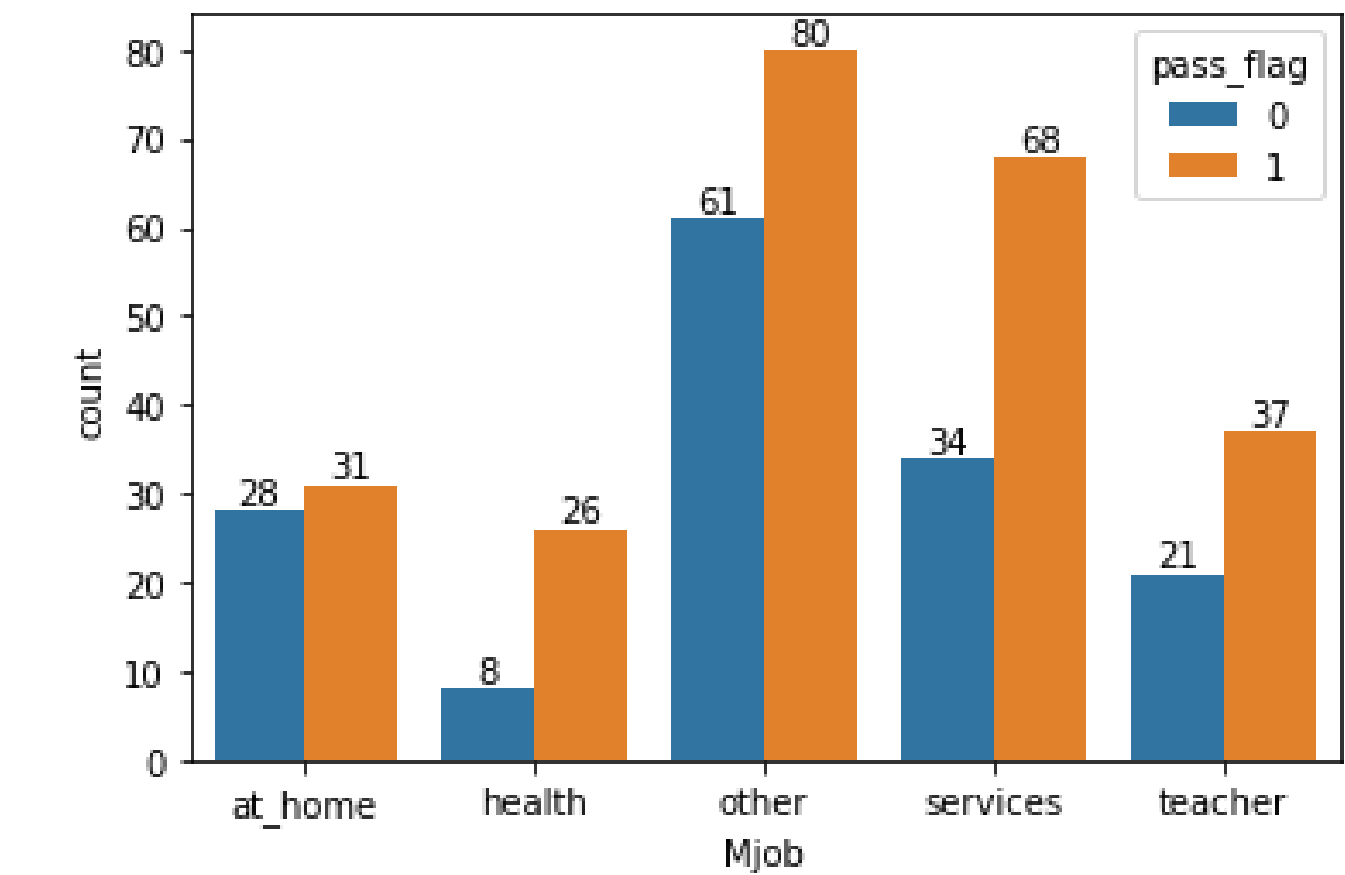
Parental Status

Students whose parents live together tend to have higher scores and pass the exam more than students whose parents live apart. This may be because living together provides a more stable and supportive home environment that can help students succeed in school.



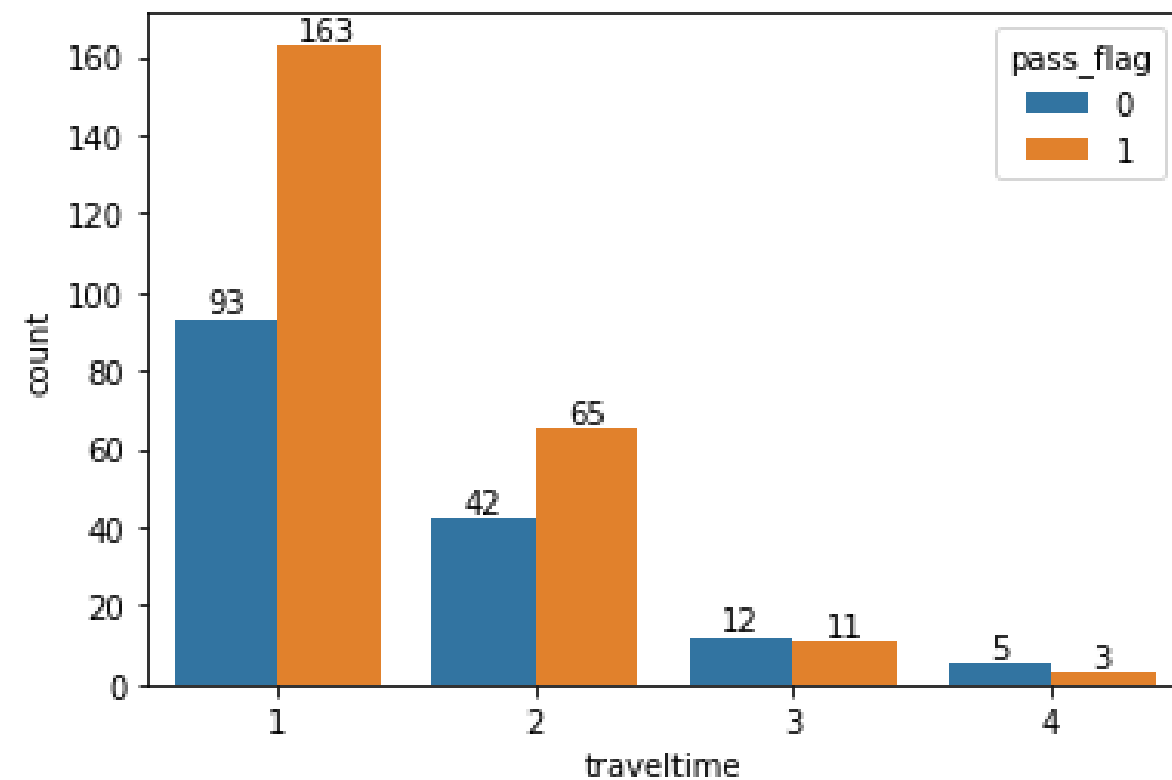
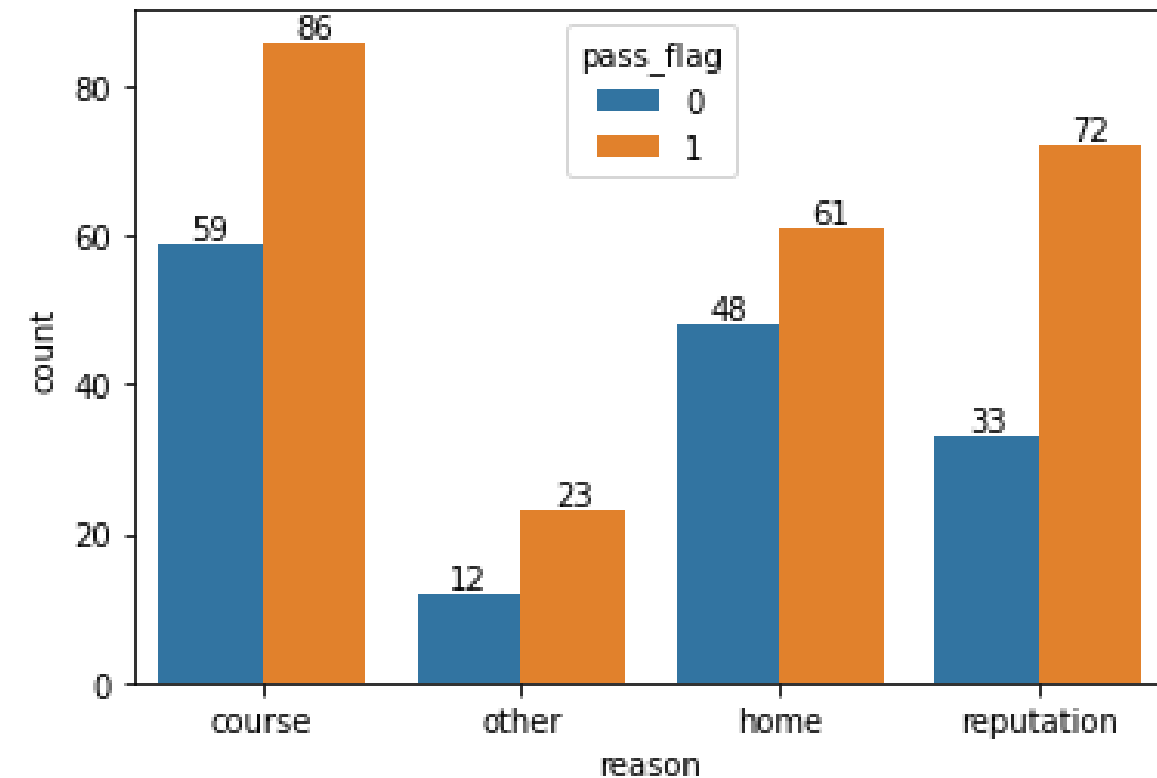
Father and Mother Education

Students whose parents graduated from elementary and higher tend to pass the exams because their parents may be more likely to value education and provide their children with guidance and support in their studies.



Father and Mother Job

Students whose parents' jobs are other tend to pass the exams than students whose parents' jobs are teachers.

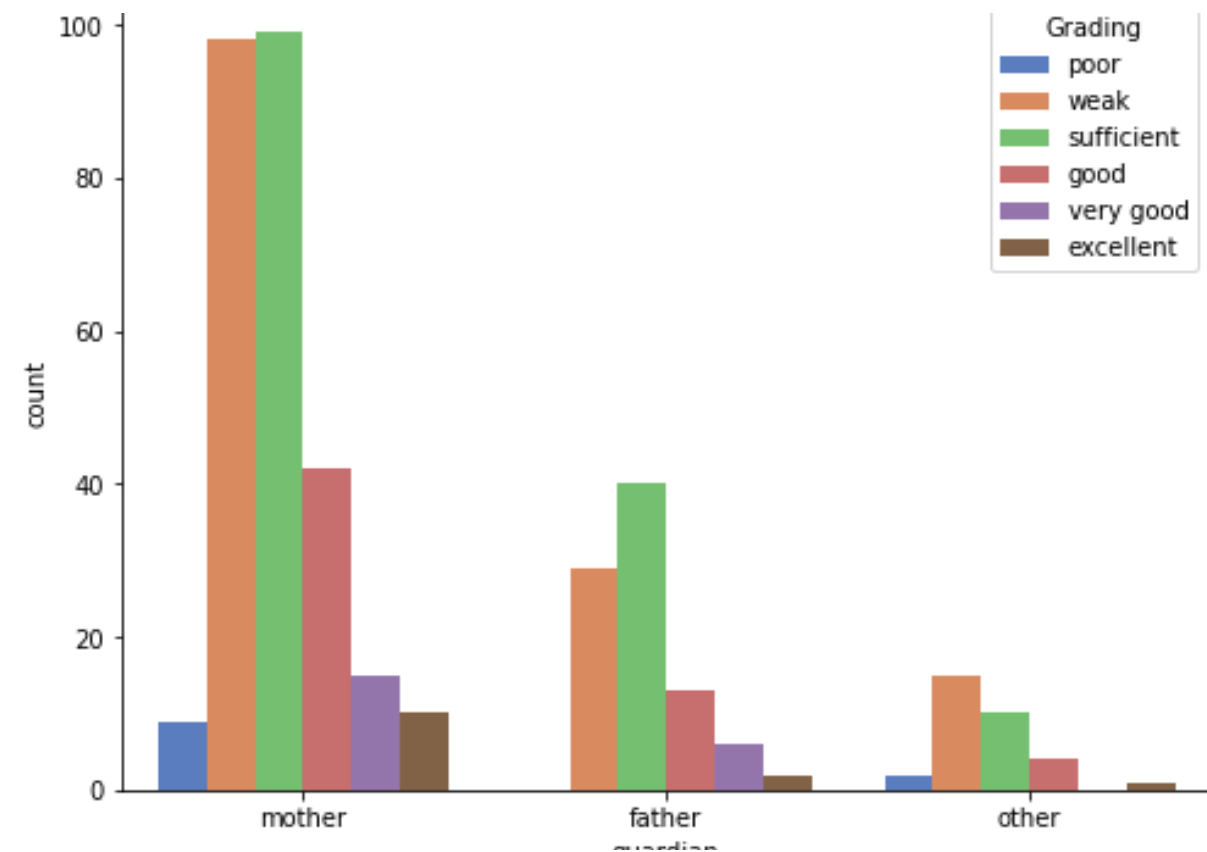
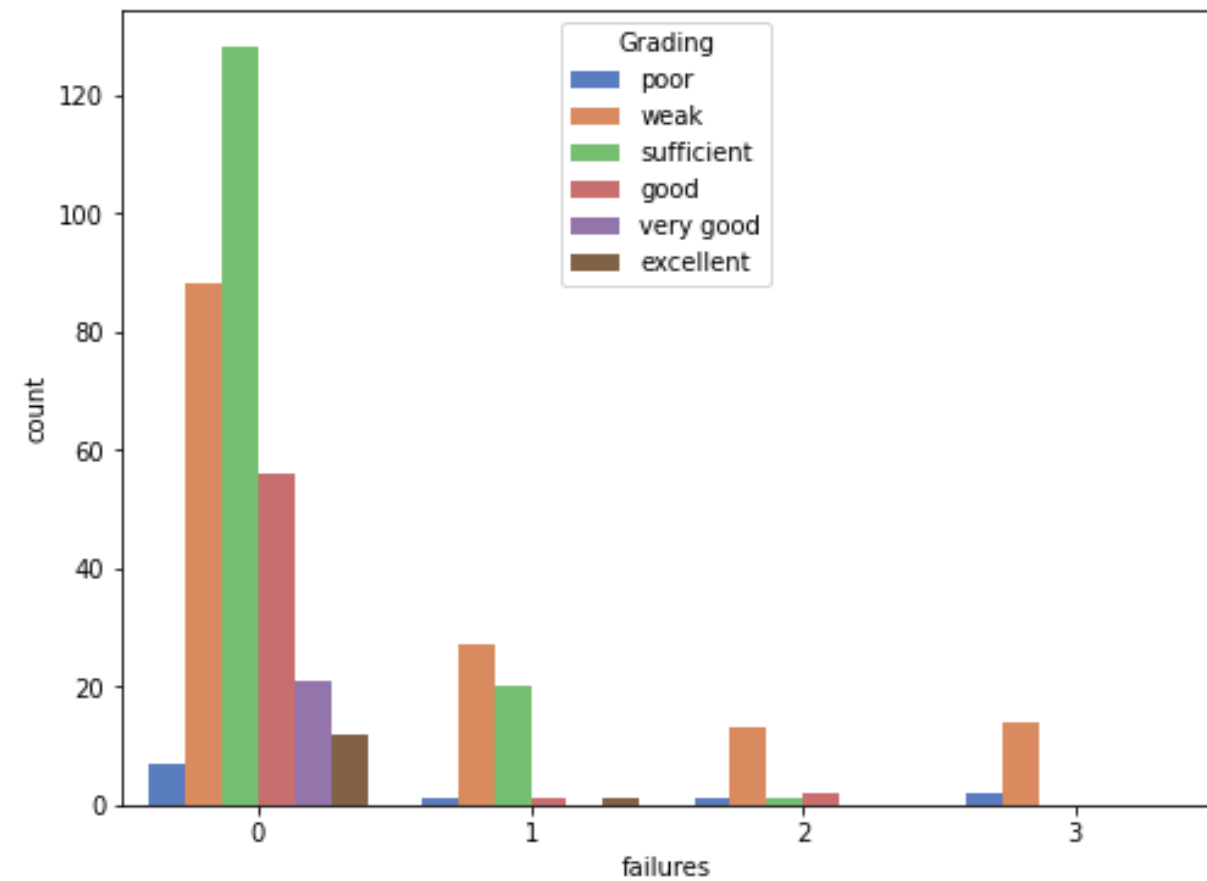


Reason to choose school

Students who choose schools based on courses usually have strong motivation and interest in that field, so they are more interested in learning and more focused on following the lessons. In addition, schools that offer certain courses usually also have better learning facilities and resources than other schools

Travel time

Students who live near their schools may be more likely to achieve good grades and pass their exams because they do not have to spend a lot of time and energy commuting. This can allow them to have more time and energy for studying and other activities that support their academic success.



Failures

Students who never fail tend to have better grades. Students who never fail may also have better study habits and more effective learning strategies

Guardian

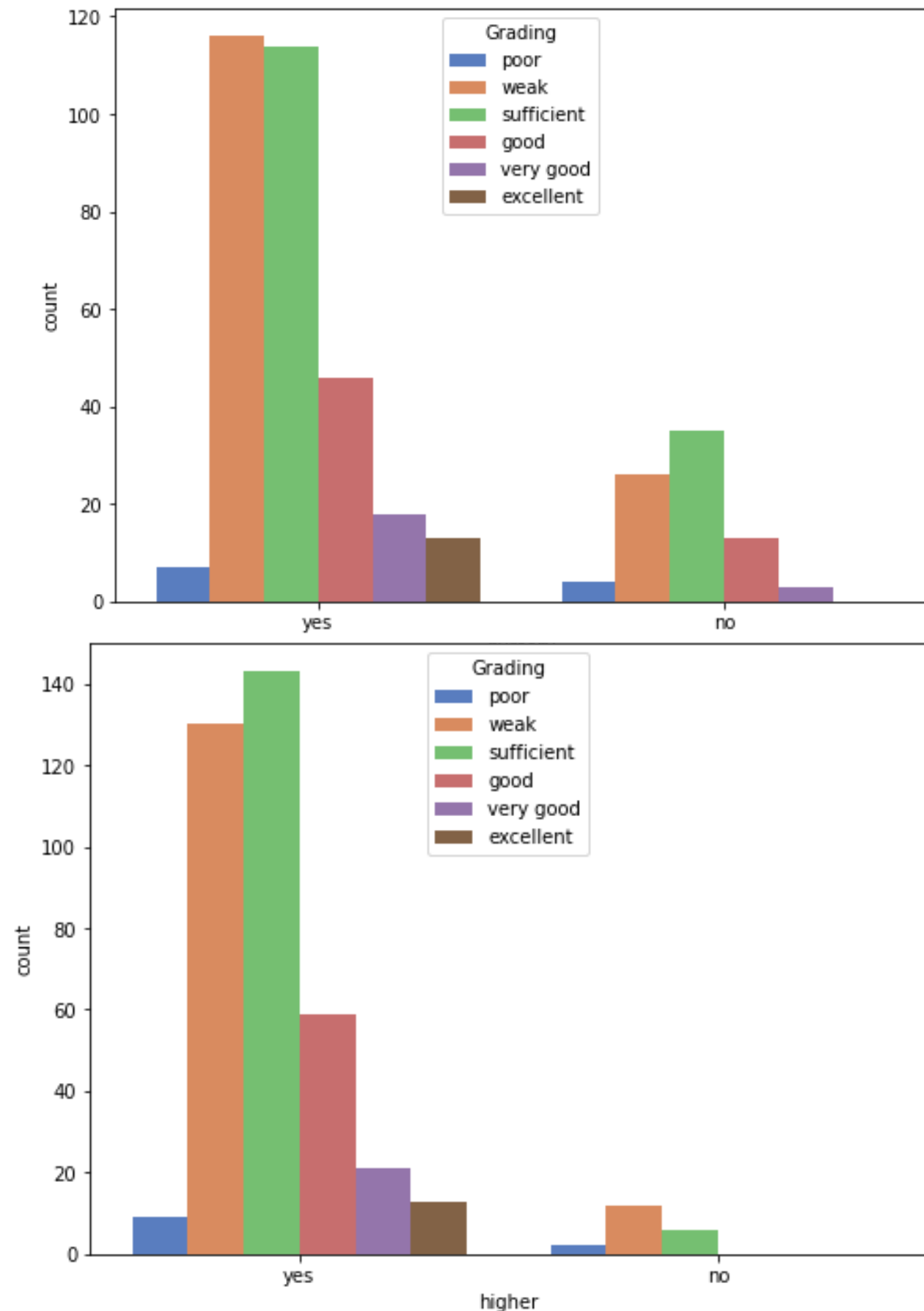
Students whose mothers are guardians tend to have better grades. When a mother is responsible for taking care of her children, she may be more involved in their lives and more invested in their academic success.

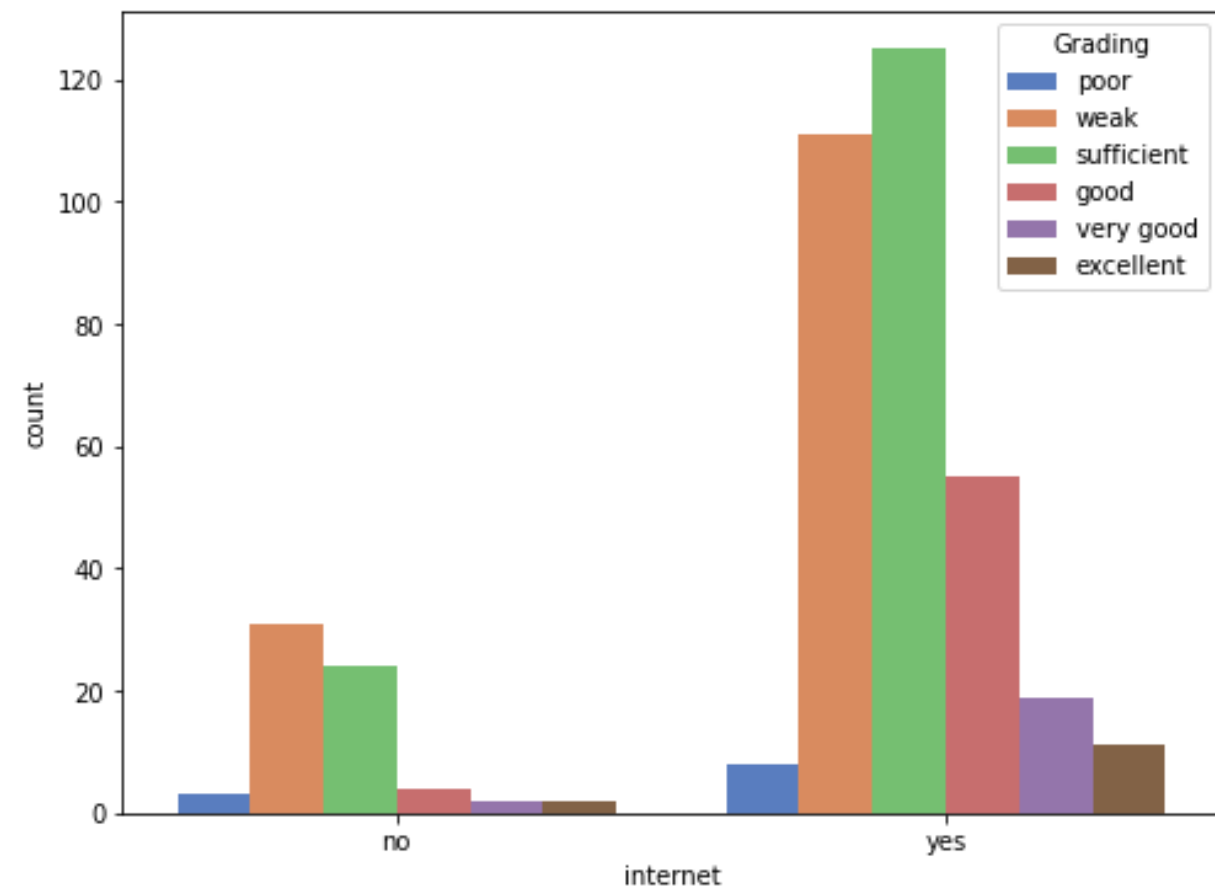
Nursery

Students who attended nursery school tend to have better grades. It means early education is important to develop cognitive, social, and emotional skills that can support their academic achievements.

Pursue higher edu

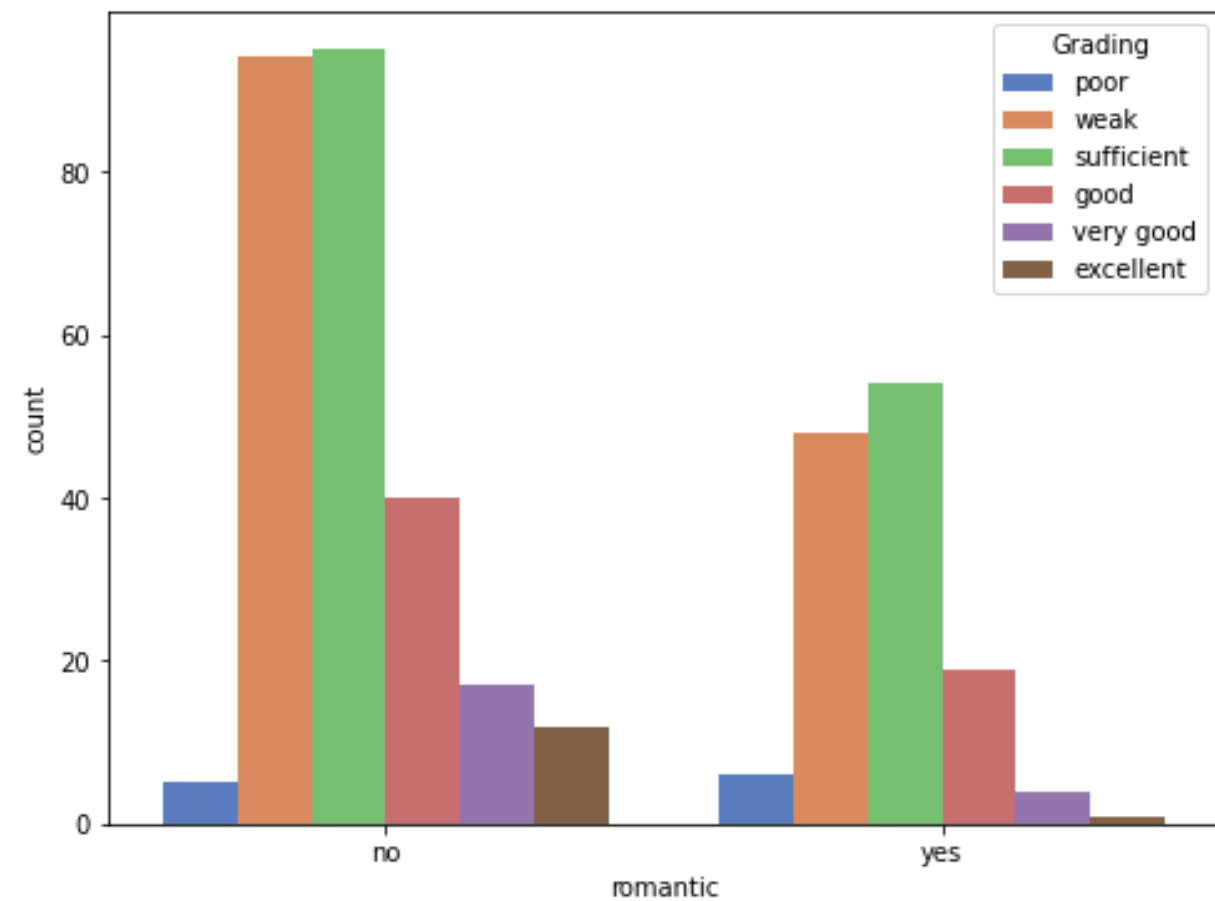
Students who are interested in pursuing higher education tend to have better grades than those who are not interested, because they are more motivated to succeed and more determined to achieve their academic goals.





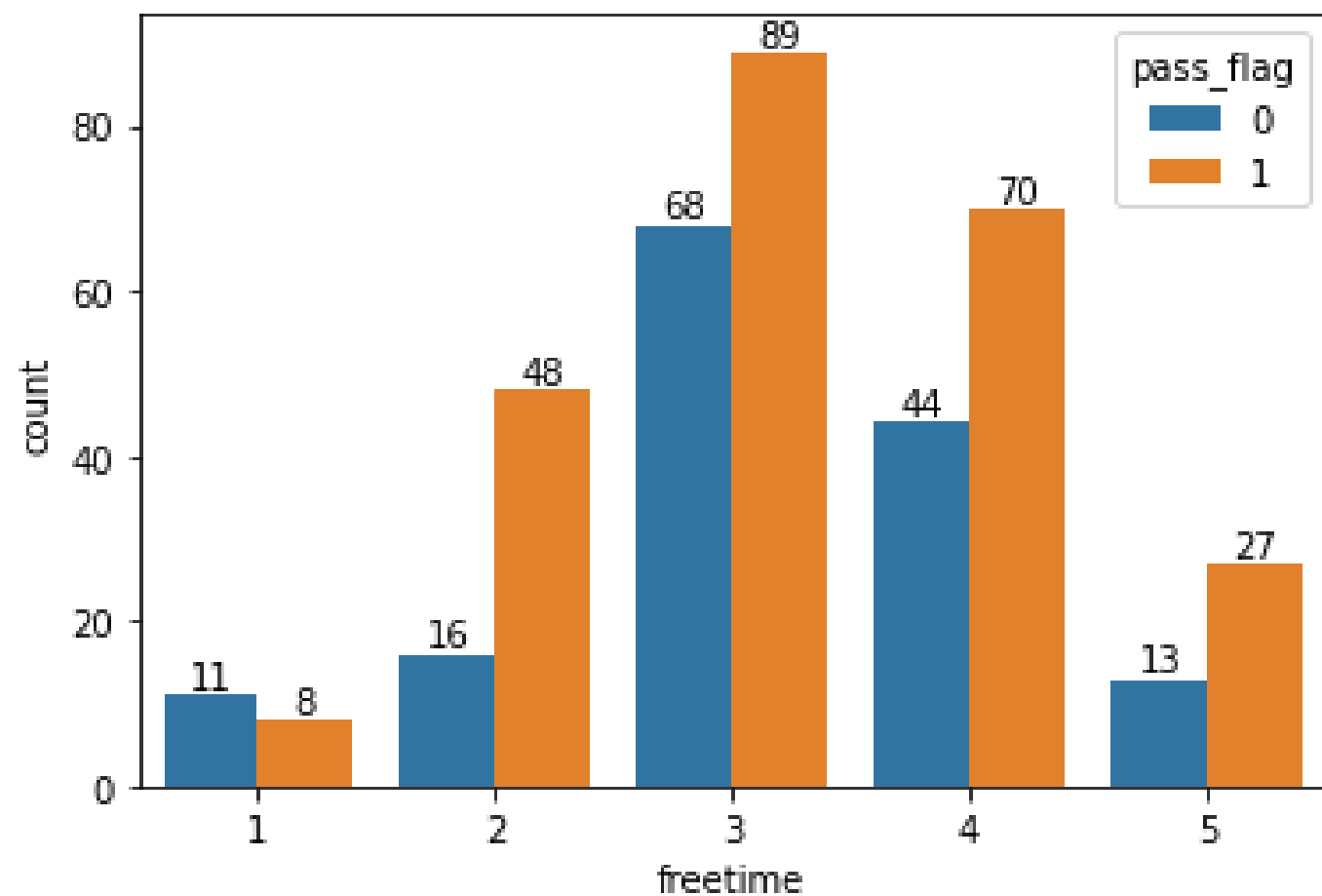
Internet at home

Students who have internet connection at their home tend to have better grades.



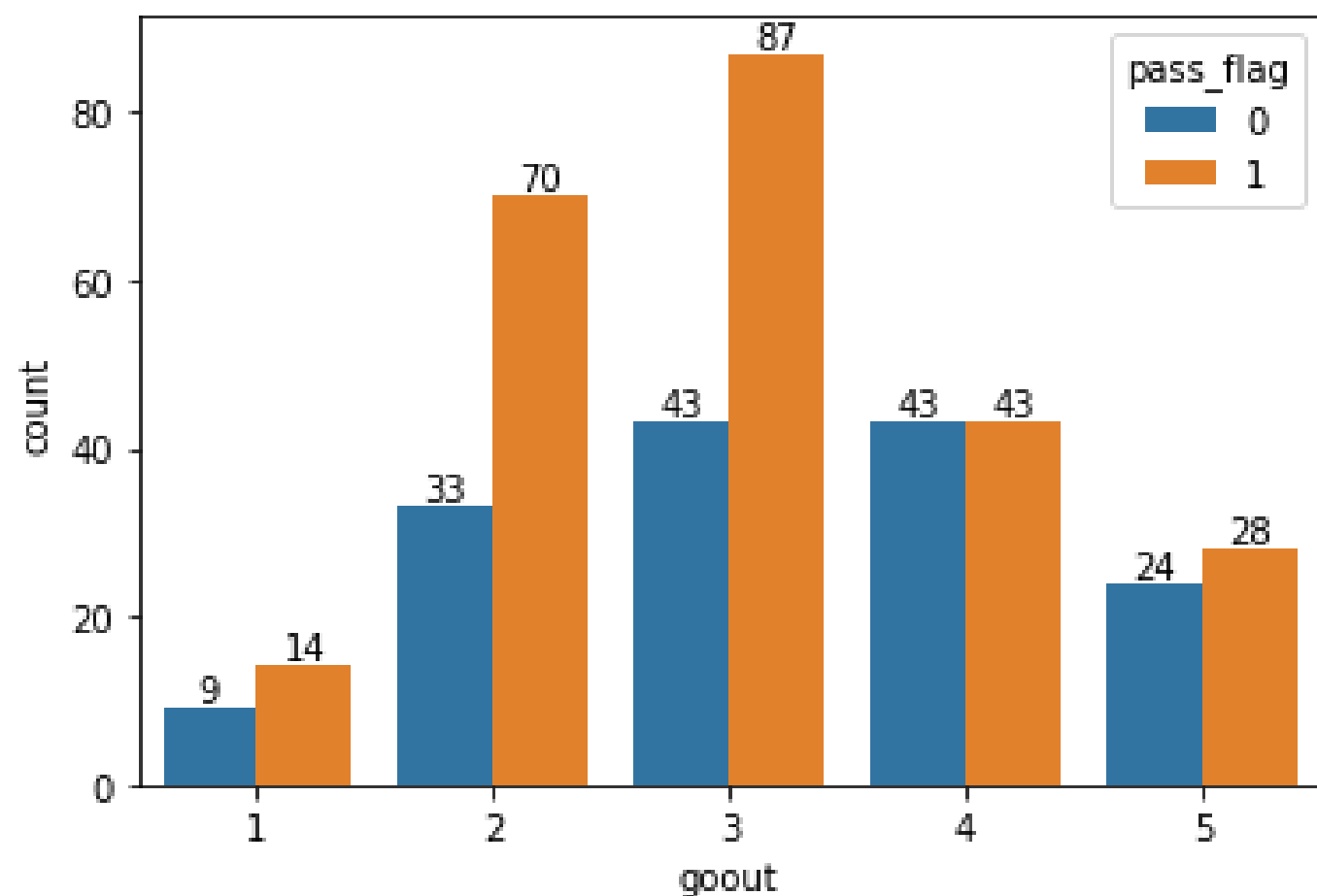
Romantic Partner

Students who do not have partners tend to have better grades and mostly pass the exams.



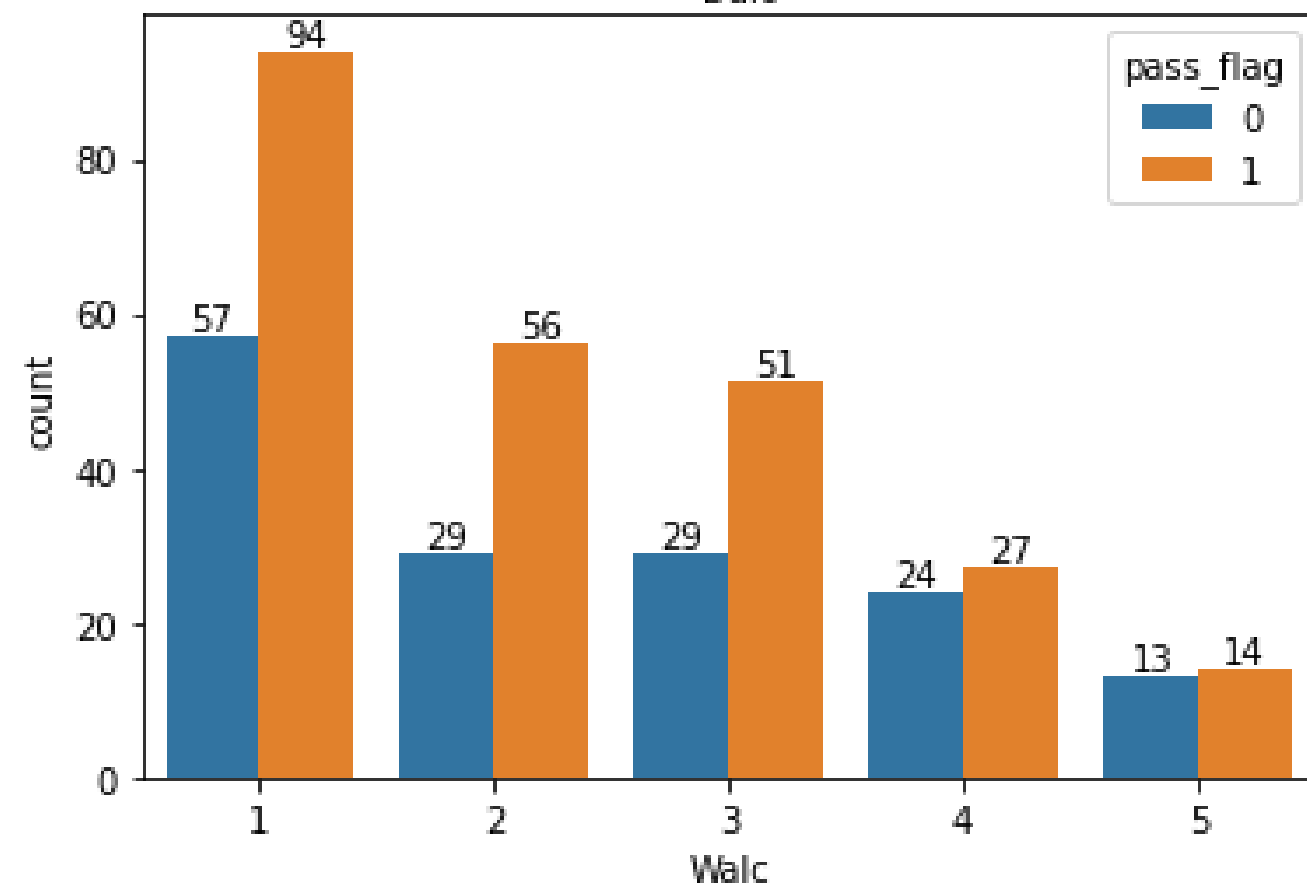
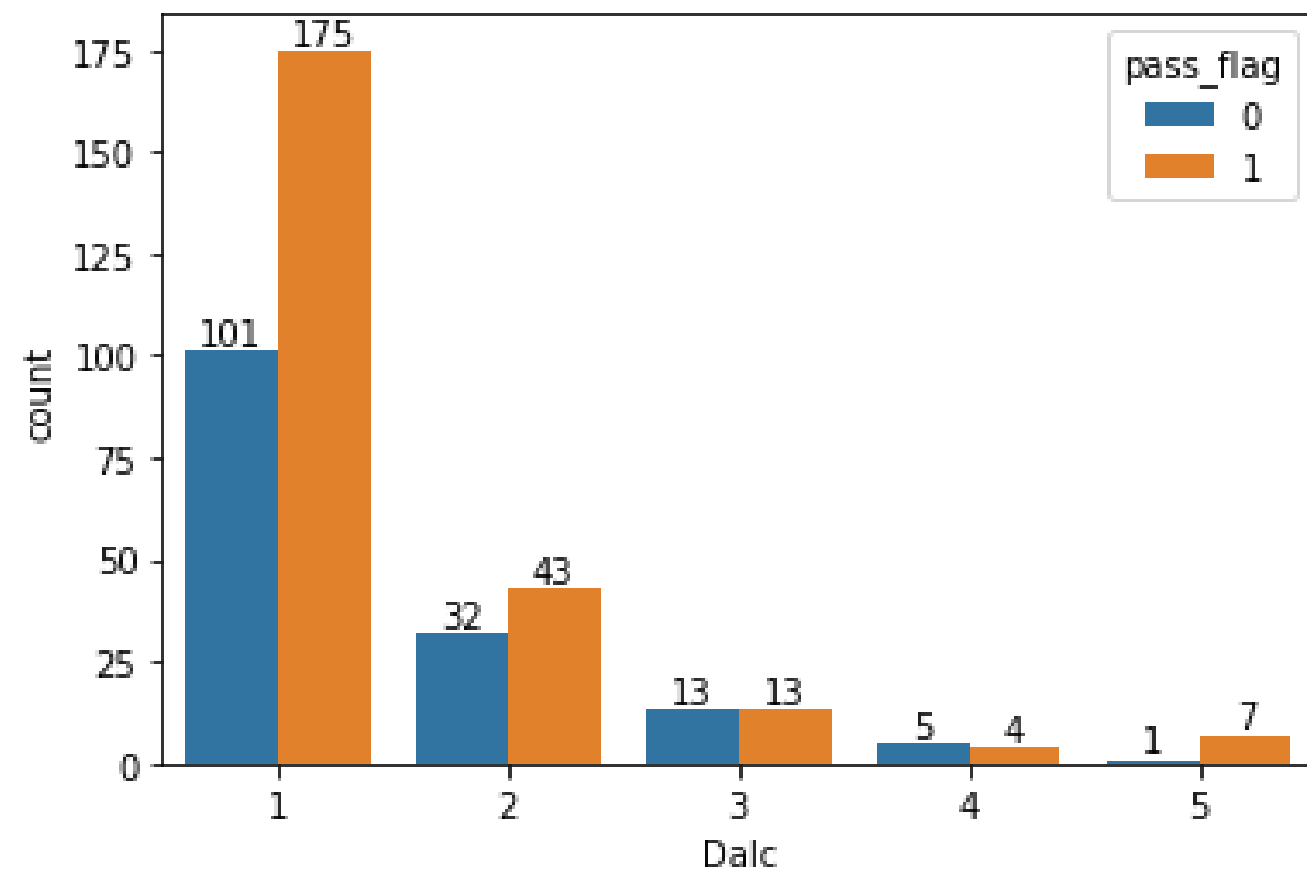
Free time

Students who have sufficient free time tend to get good grades and pass their exams because they may be better able to balance their academic and personal lives



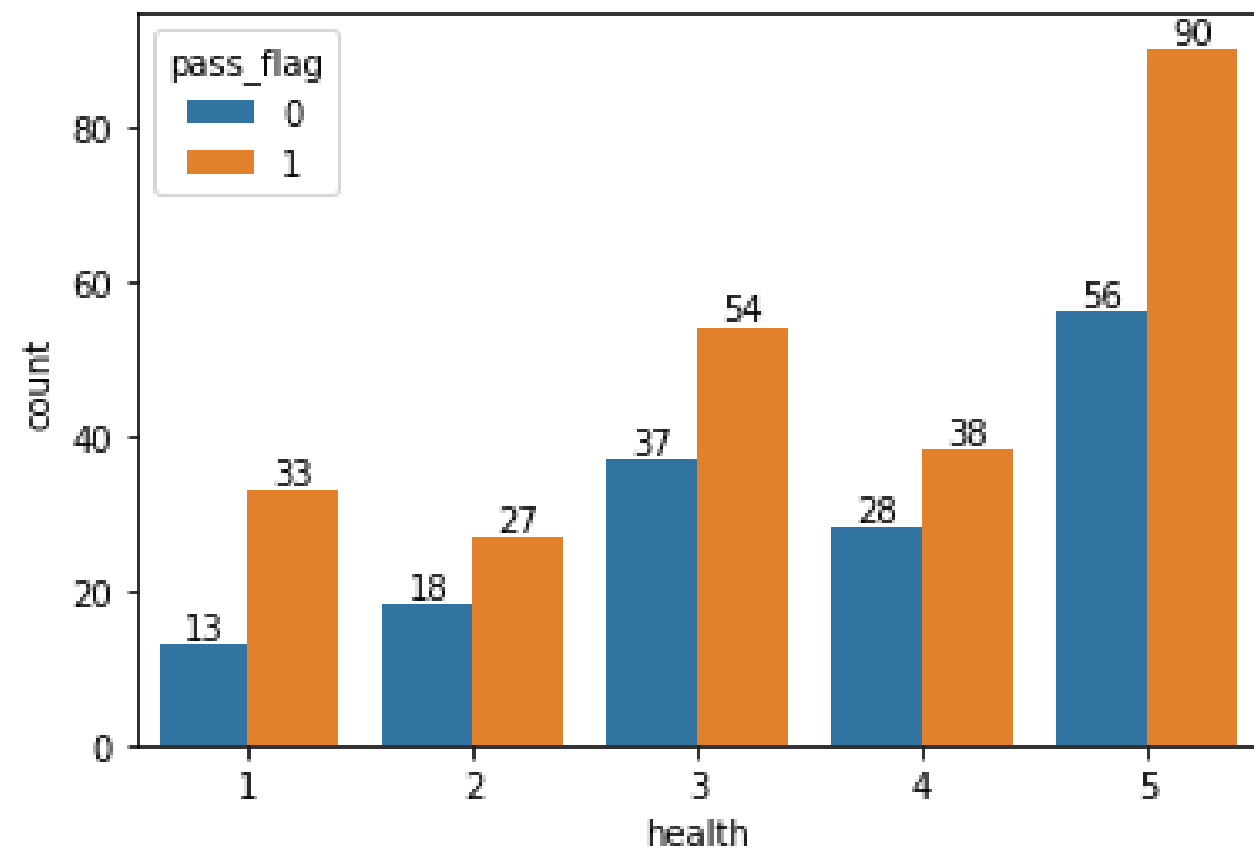
Go out with friends

Students who have sufficient time to have fun with their friends tend to get good grades and pass their exams because they may be better able to balance their academic and personal lives



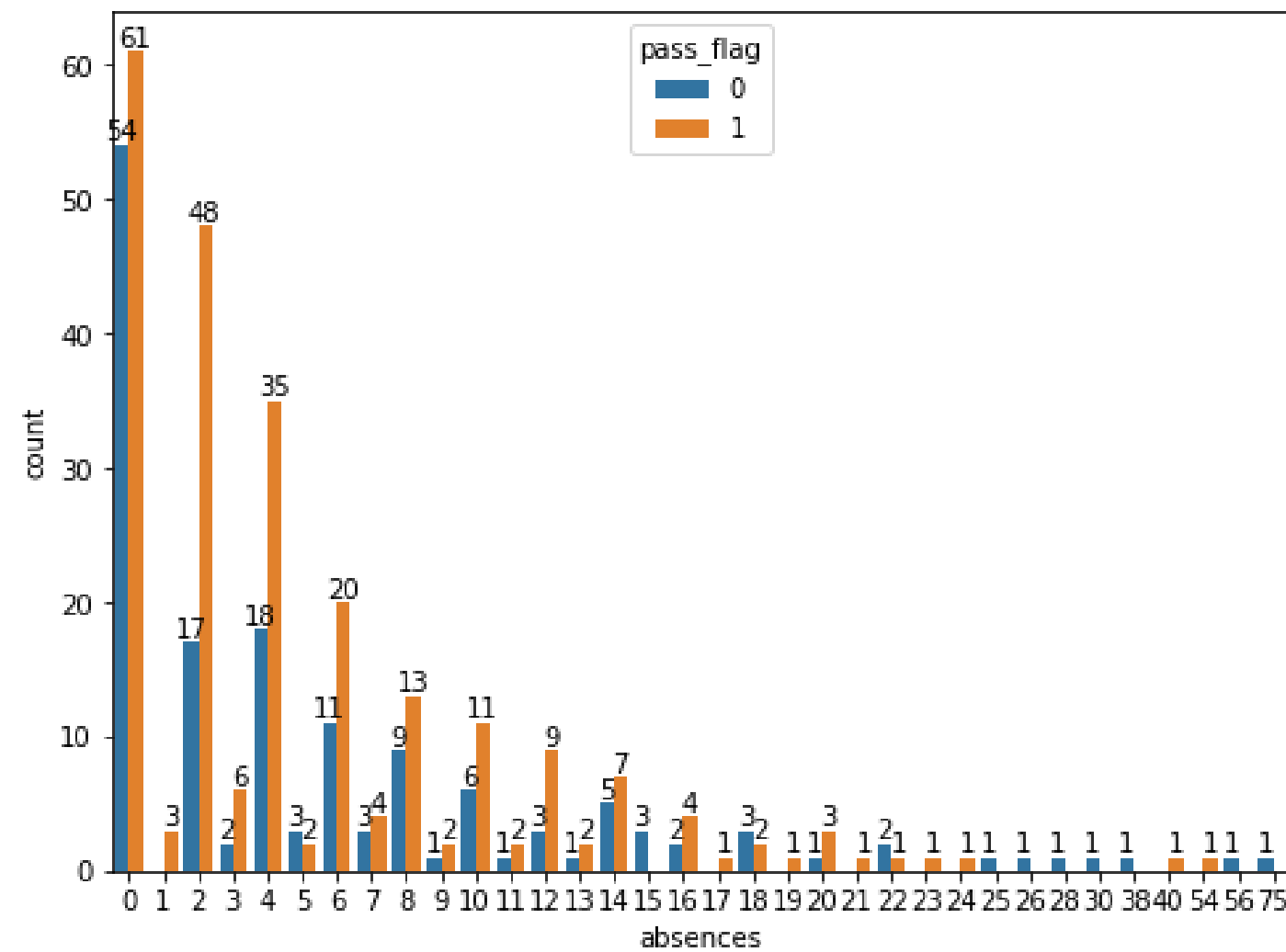
Weekday and weekend Alcohol Consumption

Students who rarely drink alcohol during the weekdays tend to get good grades, while their grades may be lower during holidays when they are more likely to drink alcohol. Therefore, avoiding or lessening alcohol during the weekend can help students to concentrate and absorb and retain information better.



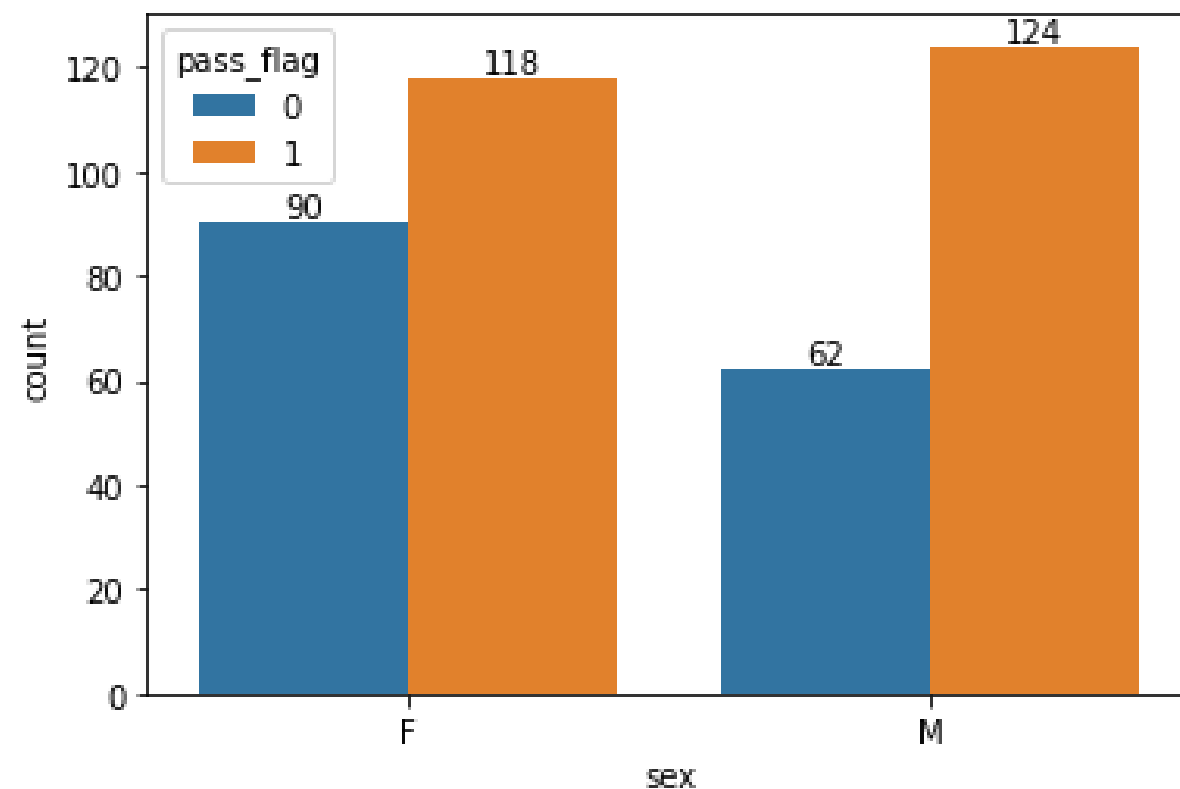
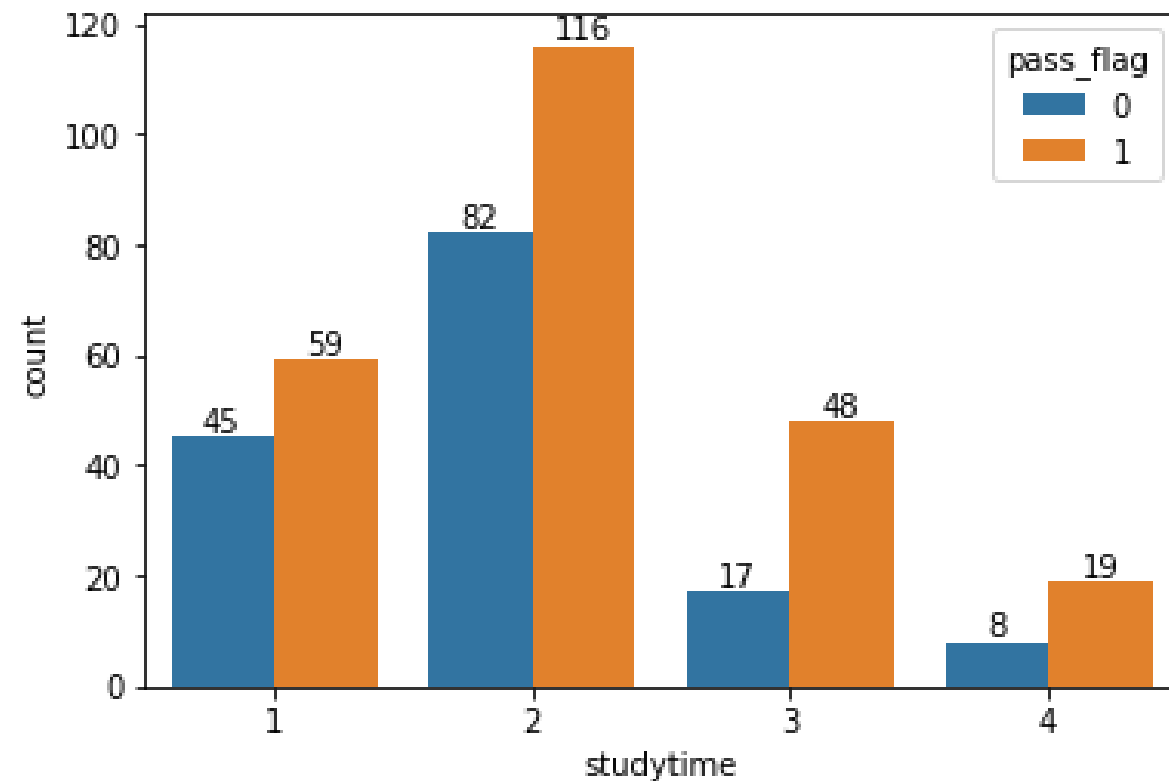
Health

Students who are healthier tend to have better grades and pass the exams.



Absences

Students who rarely absent tend to have better grades and pass the exams.

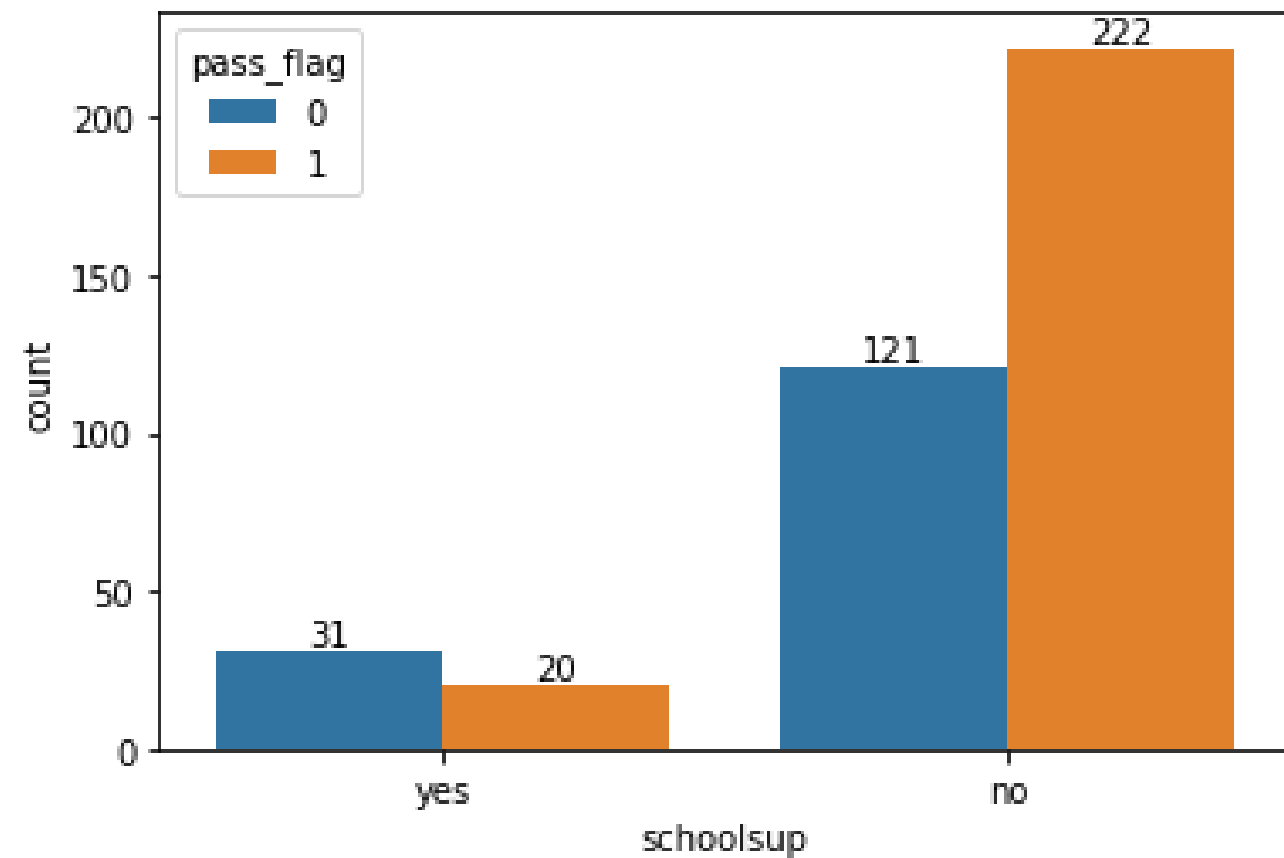


Study time

The success or failure of an exam does not depend solely on the amount of time spent studying, but also on the student's concentration while studying. Students who are able to concentrate effectively during their study time may be able to learn and retain more information, which can improve their chances of passing exams.

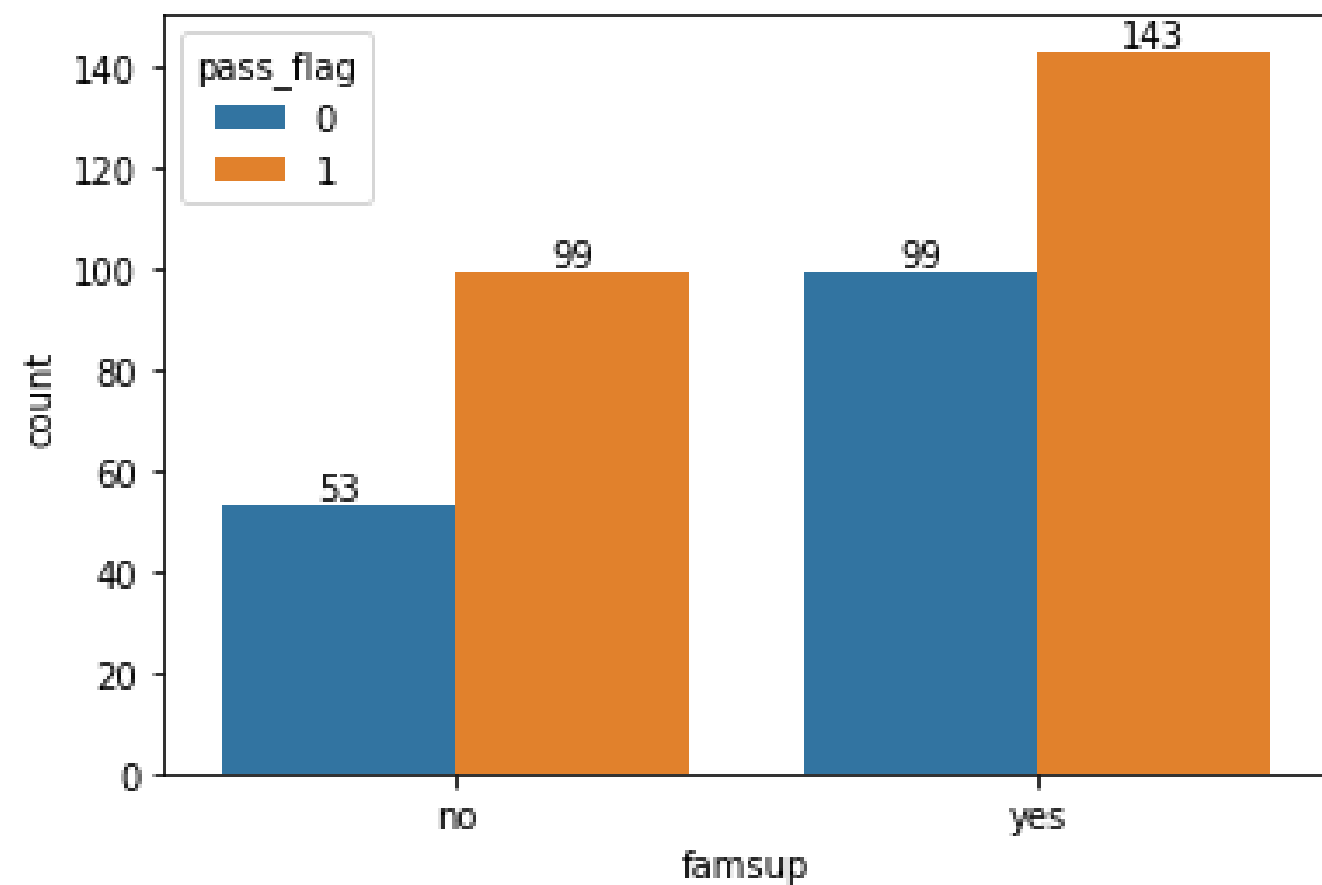
Gender

Male students tend to have a slightly higher rate of passing exams compared to female students. However, the difference in the rate of passing exams between male and female students is not particularly significant.



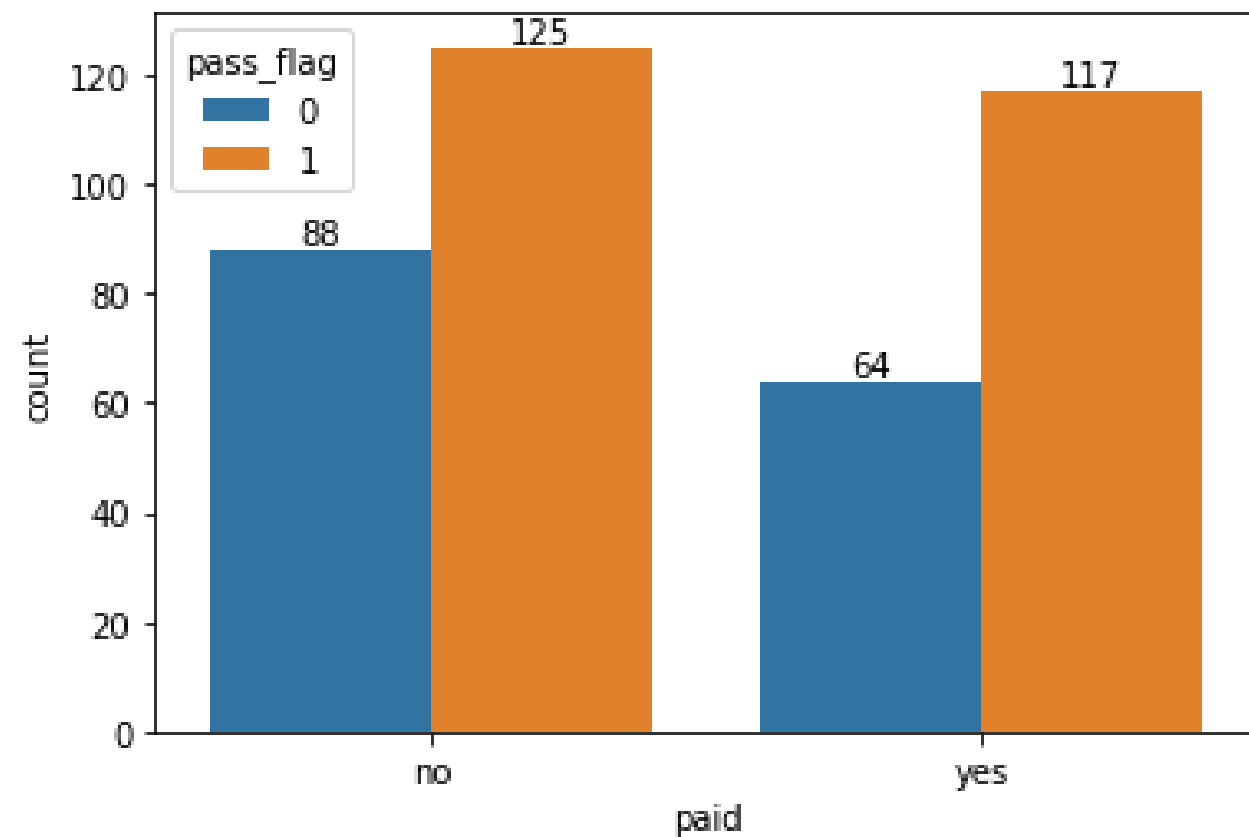
School Support

Students who get and do not get school support tend to have better grades.



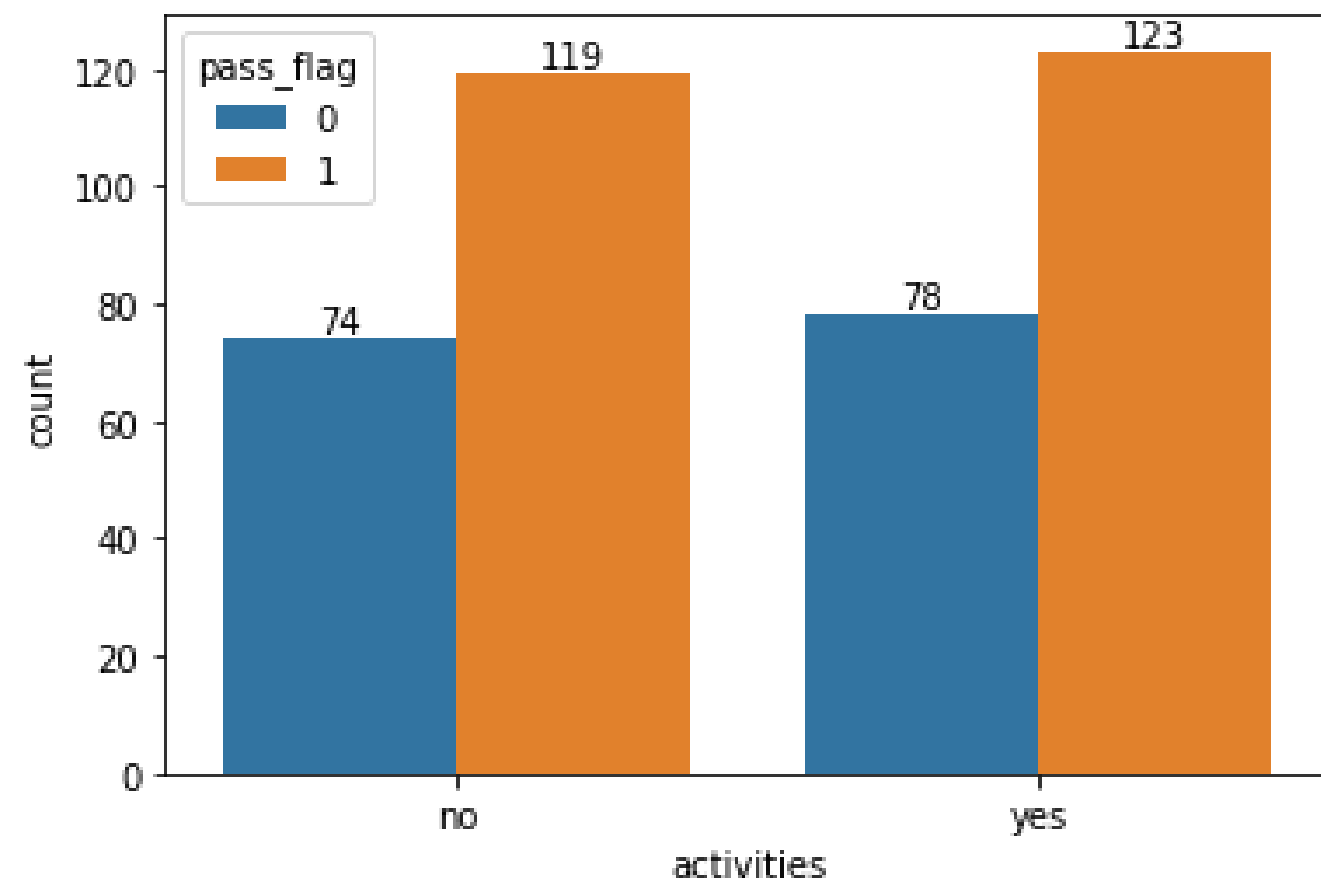
Family Support

The difference in the number of students who receive and do not receive family support is not particularly significant.



Paid extra classes

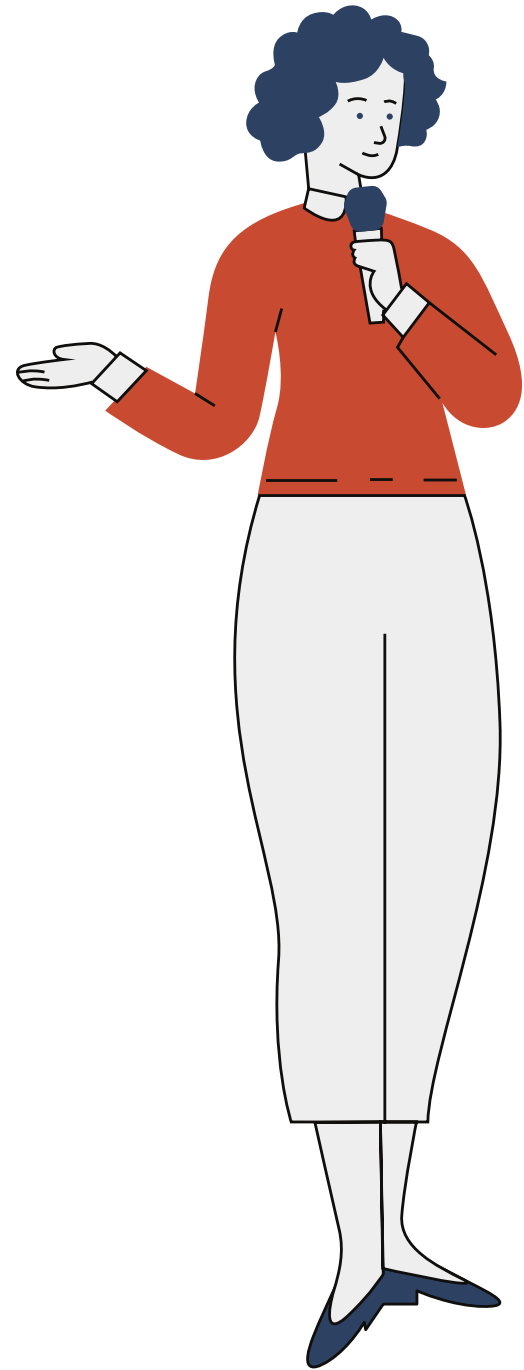
The difference in the number of students who pay for extra classes and those who do not join is not particularly significant.



Extracurricular Activities

The difference in the number of students who participate in extracurricular activities is not particularly significant.

Data Preprocessing

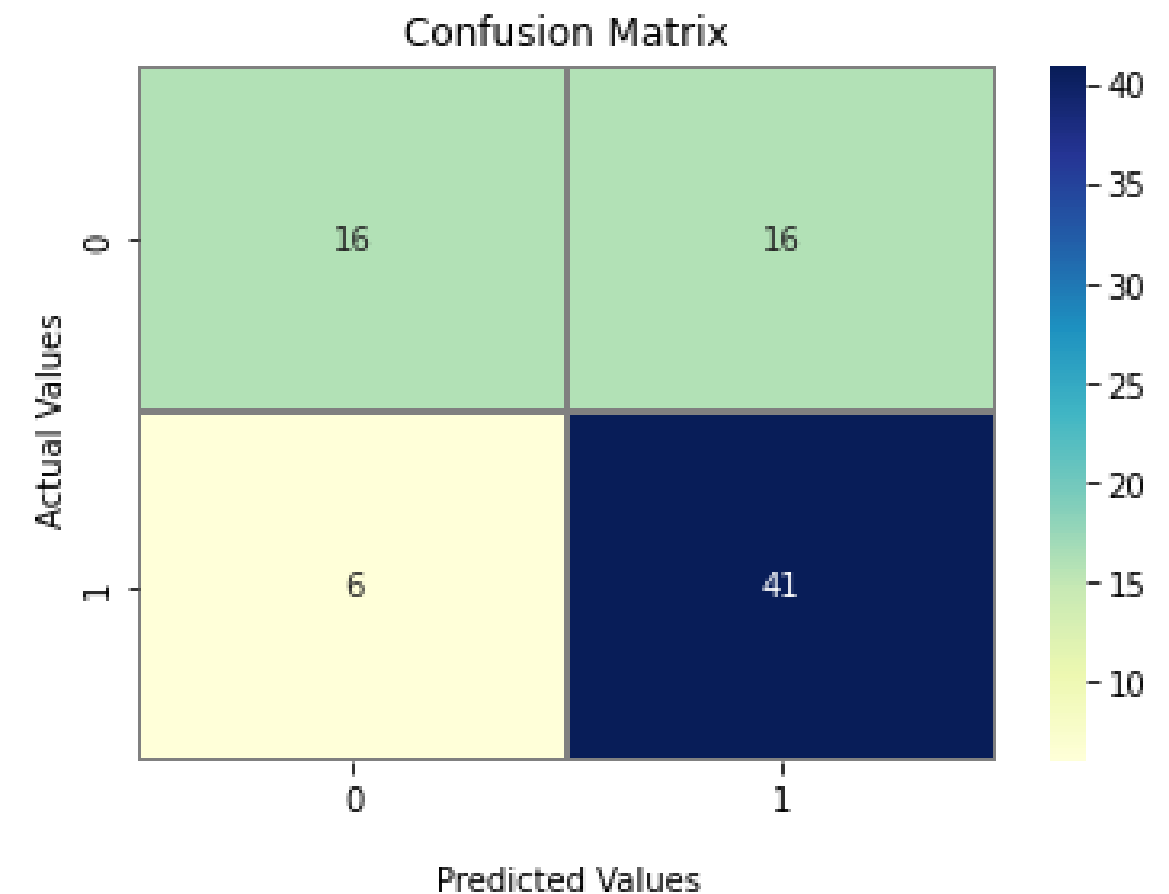


Aspect	Action
Categorical Features	Categorical feature with 2 distinct values -> Binary Encoding Categorical with more than 2 distinct values -> One Hot Encoding
Drop Features	G1, G2, G3 -> G Average

Modelling

Machine Learning Model	Precision	Recall	F1-Score
Logistic Regression	0.66	0.79	0.72
Decision Tree	0.76	0.81	0.78
Random Forest	0.66	0.87	0.75
XGBoost	0.72	0.87	0.79

- Random Forest and XG Boost have the best recall score compared to other models
- The models are able to correctly identify 87% of the students who did not pass their exams



Conclusion and Recommendation

Conclusion

- Age, parents educations, parents jobs, travel time, failures, parents relationship, address, reason to choose school, guardian, nursery, pursue higher education, internet, romantic, free time, go out, weekdays and weekend alcohol consumption, health, absences may significantly impact students' performance on exam
- Study time, gender, family support, school support, paid extra classes, and extracurricular activities may not significantly impact students' performance on exam
- Random Forest and XGBoost are able to correctly identify 87% of the students who did not pass their exams

Recommendation

- This machine learning model can be used to identify students who are at risk of failing an exam, based on their performance on a pretest. By using the model to predict which students are likely to fail, teachers can provide additional support and assistance to these students in order to help them improve their scores and increase their chances of success on the actual exam. By doing this, teachers can help to reduce the number of students who might fail the exam

The End

**Thank you
for listening**

