Semester Project

Student-mat Data Analysis

**Submitted By
Iftekhar Hossain**

Table of Contents

1. Introduction

Knowing a student's social, gender, and academic information is a crucial tool for figuring out what kind of grades they are getting. In this semester's project, I will use data on secondary school student performance to analyze them and predict the students' final grades. The dataset is based on a study conducted by Cortez and Silva (2008). From the dataset, we can find a rich pool of data about the students, including their demographic, family background, their study habits, information about their social activities, and much more.

Using all the data presented in this dataset, I will attempt to predict a final grade for a given student. I have used various methods such as exploratory data analysis (EDA), correlation, regression, and descriptive statistics for the analysis. This analysis and prediction aim to help a secondary school teacher (Ruth) devise methods to ensure the best results for each student in their class.

In the following section (Section 2), I will present my initial assumptions and hypotheses, followed by Section 3, which will present a detailed overview of the exploratory data analysis. After that, I present the identified trends, patterns, and anomalies in Section 4. In Section 5, I discuss the findings concerning the initial assumptions and hypotheses. Finally, I conclude the report by summarizing and reflecting upon the findings.

2. Initial assumptions and hypotheses

The dataset includes a wide variety of data about the students. The following table (Table 2.1) shows a list of variables and their description used in the dataset and for the analysis.

Variable name	Variable description
school	student's school (binary: 'GP' Gabriel Pereira or 'MS' Mousinho da Silveira)
sex	student's sex (binary: 'F' female or 'M' male)
age	student's age (numeric: from 15 to 22)
address	student's home address type (binary: 'U' urban or 'R' rural)
famsize	family size (binary: 'LE3' less or equal to 3 or 'GT3' greater than 3)
Pstatus	parent's cohabitation status (binary: 'T' living together or 'A' apart)
Medu	mother's education (numeric: 0 none, 1 primary education (4th grade), 2 – 5th to 9th grade, 3 – secondary education or 4 – higher education)
Fedu	father's education (numeric: 0 none, 1 primary education (4th grade), 2 – 5th to 9th grade, 3 – secondary education or 4 – higher education)
Mjob	mother's job (nominal: 'teacher', 'health' care related, civil 'services' (e.g. administrative or police), 'at_home' or 'other')
Fjob	father's job (nominal: 'teacher', 'health' care related, civil 'services' (e.g. administrative or police), 'at_home' or 'other')
reason	reason to choose this school (nominal: close to 'home', school 'reputation', 'course'

	preference or 'other')
guardian	student's guardian (nominal: 'mother', 'father' 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<=n<3, else 3)
schoolsup	extra educational support (binary: yes or no)
famsup	family educational support (binary: yes or no)
paid	extra paid classes within the course subject (Math or Portuguese) (binary: yes or no)
activities	extracurricular activities (binary: yes or no)
nursery	attended nursery school (binary: yes or no)
higher	wants to take higher education (binary: yes or no)
internet	Internet access at home (binary: yes or no)
romantic	with a romantic relationship (binary: yes or no)
famrel	quality of family relationships (numeric: from 1 very bad to 5 excellent)
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)
Dalc	workday alcohol consumption (numeric: from 1 very low to 5 very high)
Walc	weekend 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, output target)

Table 2.1: List of all the variables and their description

In the study by Cortez and Silva (2008), they tried to predict secondary students' final grades by using students' past school grades, social, demographic, and other student-related data. They have explored with different input variables combinations (both with and without past grades as input variables also). The study showed a high predictive accuracy by providing the first and/or second-period grades as input variables.

The findings from Cortez and Silva (2008) support the finding by Kotsiantis et al. (2004), which indicates that students' prior performances had a significant impact on students' future achievements. In their research, best predictive models have shown that, in some cases, there are other relevant features, such as school related (e.g. the reason to choose the school, number of absences, extra educational school support), demographic (e.g. student age, parent's job, and parent's education) and social (e.g., going out with friends, alcohol consumption) variables. According to the study, only a small part of the input variables seems to be relevant. Which is true in most of the other cases. It benefits the nonlinear function methods (e.g., SVM and NN) because these methods are more sensitive to irrelevant inputs.

I believe further research is needed (e.g., sociological studies) in order to understand why and how some variables (e.g., the reason to choose the school, parent's job, parent's education, or student's alcohol consumption) affect students' results. Thus, in my project, I will explore why and how these variables/ features (e.g. reason to choose the school, parent's education, parent's job, or student's alcohol consumption) impact the student's final grade by performing EDA, correlation, and regression analysis.

2.1 Statistical Summary of the Data

Table 2.2 presents a statistical summary of the data about age, absence and grades.

	age	absences	G1	G2	G3
Count	395	395	395	395	395
Min	15	0	3	0	0
Max	22	75	19	19	20
Range	7	75	16	19	20
Mean	16.7	5.7	10.9	10.7	10.4
Median	17	4	11	11	11
Mode	16	0	10	9	10
stdev (σ)	1.3	8.0	3.3	3.8	4.6
Q1	16	0	8	9	8
Q3	18	8	13	13	14

Table 2.2: Statistical summary of age, absences, and grades

2.2 Data quality

Missing Values: No missing values found.
Duplicate values: No duplicate value has been found.
Outliers: I think the data has already been cleaned. Because in the description of the "absence" variable the value is from 0 to 93 days. But in the given dataset the highest number of absences is 75 days. Not 93.
It seems like outliers have already been removed.
Taking care of Categorical (binary) data: Categorical data of binary variables like school, sex, address, famsize, Pstatus, Schoolsup, famsup, paid, activities, nursery, higher, internet, and romantic have been changed in to 0 and 1.

school	=IF([@school]="GP",1,0)
sex	=IF([@sex]="F",0,1)
address	=IF([@address]="U",1,0)
famsize	=IF([@famsize]="GT3",1,0)
Pstatus	=IF([@Pstatus]="A",0,1)
Schoolsup	=IF([@schoolsup]="yes",1,0)
famsup	=IF([@famsup]="yes",1,0)

paid	=IF([@paid]="yes",1,0)
activities	=IF([@activities]="yes",1,0)
nursery	=IF([@nursery]="yes",1,0)
higher	=IF([@higher]="yes",1,0)
internet	=IF([@internet]="yes",1,0)
romantic	=IF([@romantic]="yes",1,0)
Pass/Fail	=IF([@G3]>=10,1,0)

Table 2.3: Binary conversion of the variables.

2.3 Initial assumptions and hypotheses

Based on the initial insights from the study by Cortez and Silva (2008), I have the following assumptions and hypotheses.

1) Children are highly influenced by their parents. So, my first assumption is that there will be high correlation between family education background and student's performance at school.
2) Physical energy impacts a lot of our performance. So, home to school travel time can have an impact on a student's performance. Similarly, students' health condition can also influence the final grade.
3) Generally, the students who were good at past, they also will continue doing good in future. So, number of past class failures may have a significant impact on final grade. Thus, first period grade and second period grade will show positive correlation with the final grade.
4) Extra care has impact on many things. So, extra paid classes should have a positive impact on the final grade. Similarly, family educational support, extra educational school support and internet access should have a positive impact on the final grade.
5) Willingness and motivation can have positive impact. So, students who wants to take higher education will do better in the final grade.
6) It is natural to assume that the more one studies the better one can do in the exam. That's why number of study hours should have a positive impact on the final grade. Similarly, number of absences should have a negative impact on the final grade.
7) Factors such as romantic relationship, social activities like going out with friends are assumed to have no or less impact on the grade. However, alcohol consumption can have a negative impact on the final grade.

3. Exploratory Data analysis (EDA)

This section presents a detailed analysis of the exploratory data. Here I show all the different analyses of the variables, their correlations, and finally, the regression analysis to predict the grade.
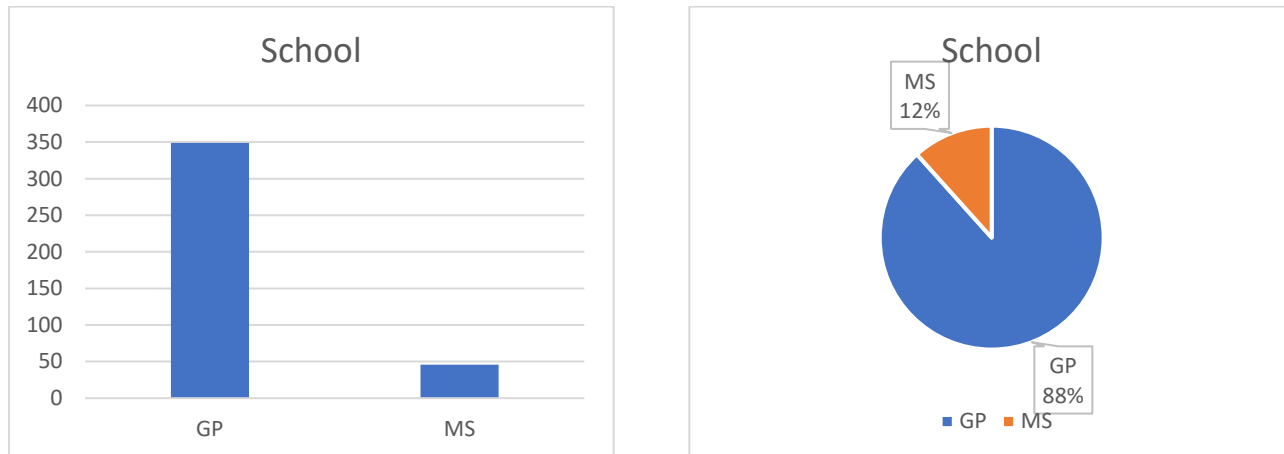
3.1 Different Features/Variables Distribution:

Figure 3.1: Student distribution per school

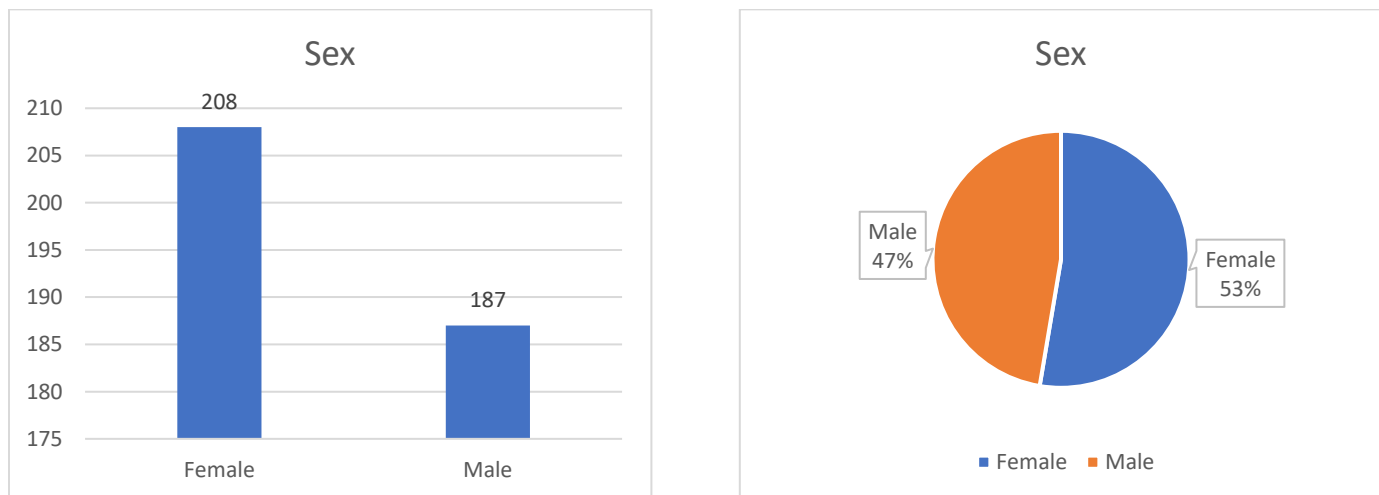- 88% information are from the students of GP Gabriel Pereira.

Sex:

Figure 3.2: Students' gender distribution

- Female students are around 6% higher than the Male student.

Age:

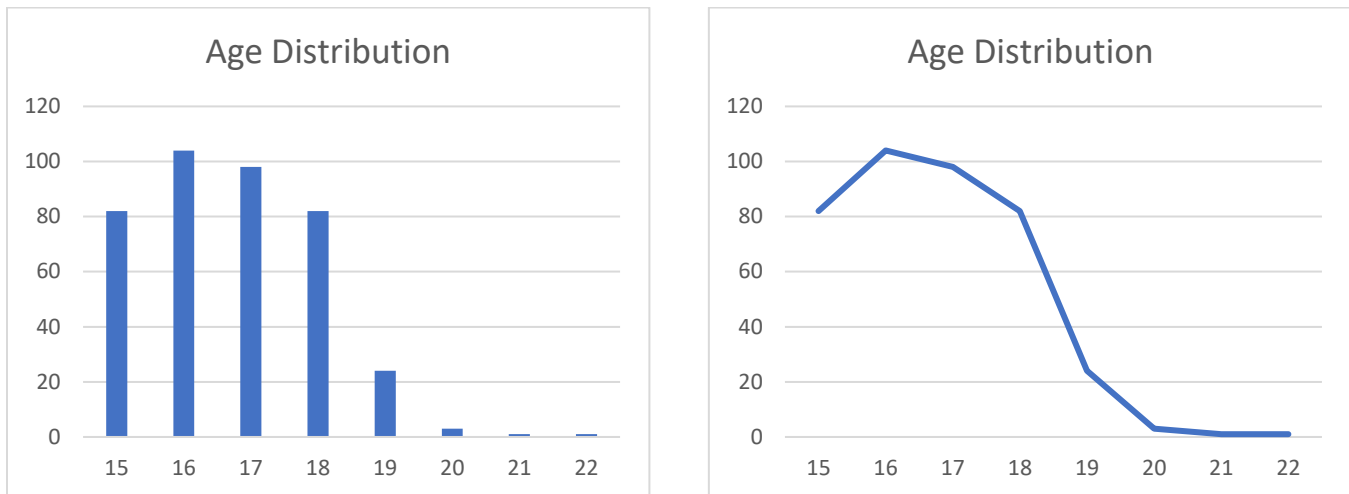
Figure 3.3: Students' age distribution

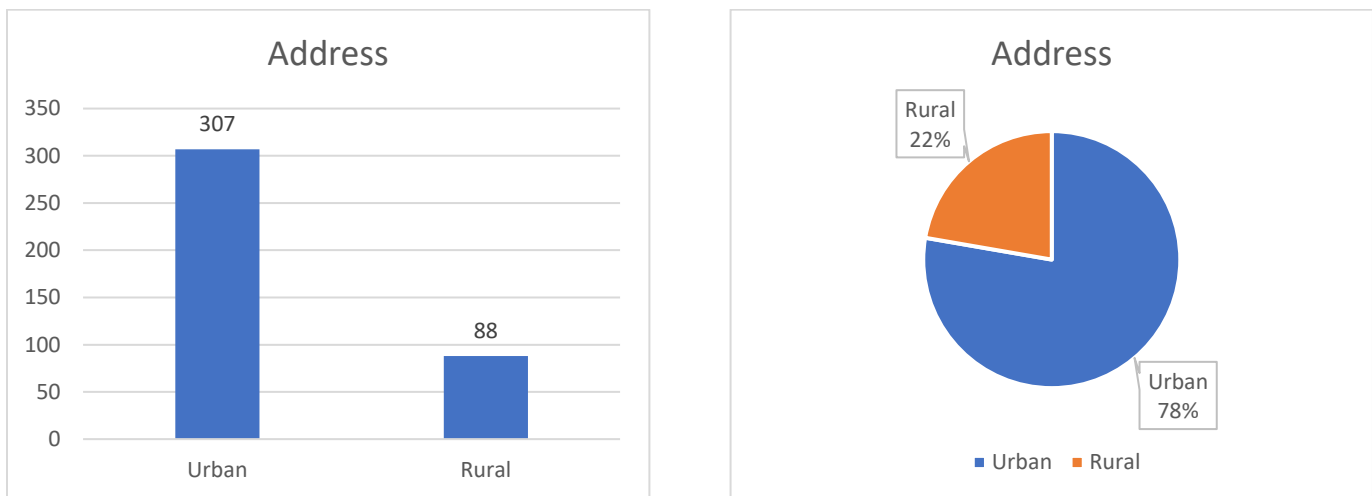- Most of the student are in between 15 to 18 years.

Address:

Figure 3.4: Students' address type distribution

- Most of the student are living in the Urban areas, almost 78%.
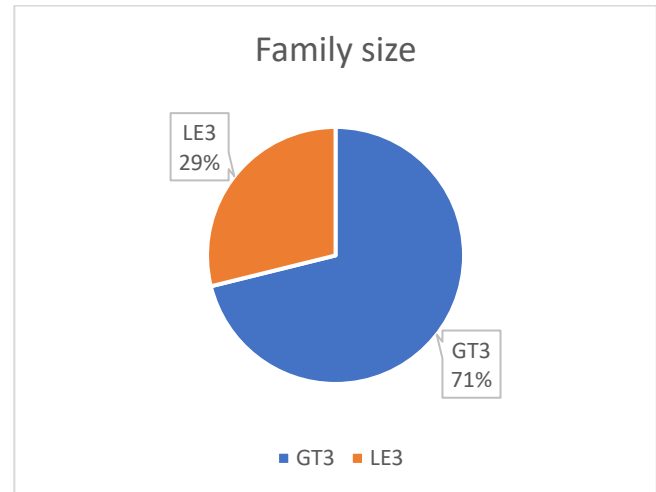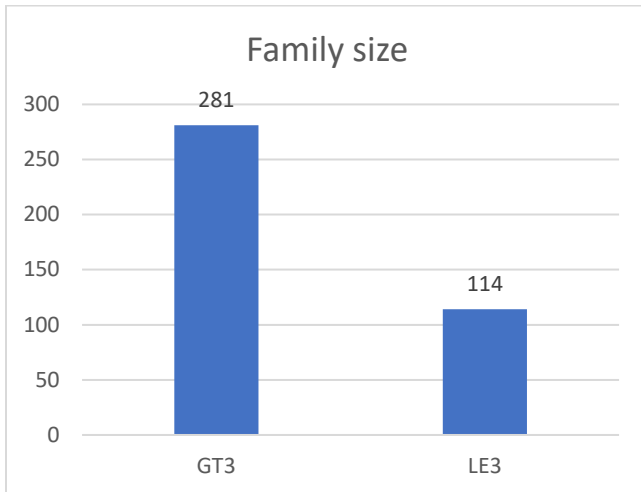
Family size:

Figure 3.5: Students' family size distribution

- About 71% students' family size is greater than 3.

Parent's cohabitation status:
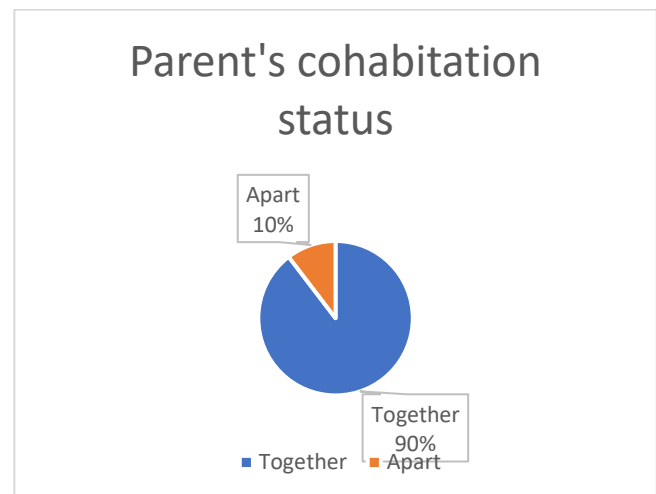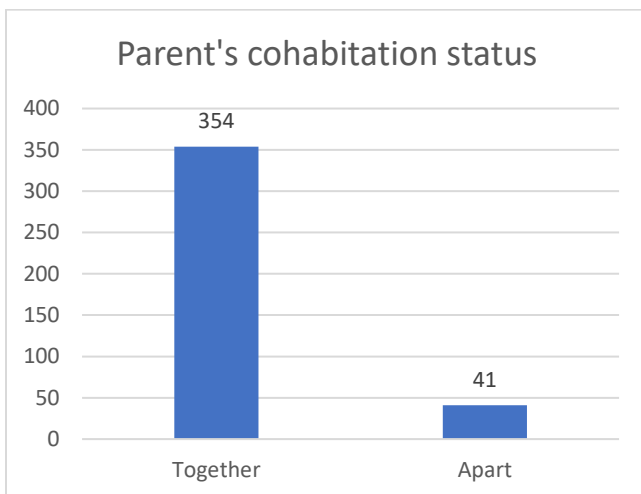
Figure 3.6: Students' parents' cohabitation distribution

- almost 90% student's parents are living together.

Mother's Education:
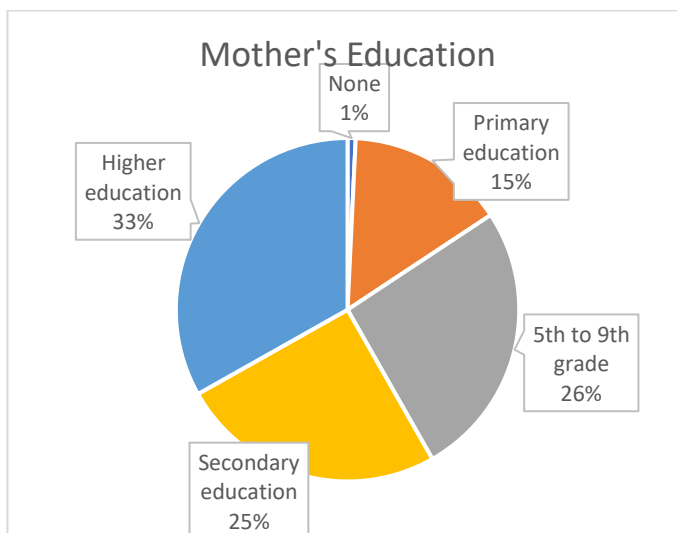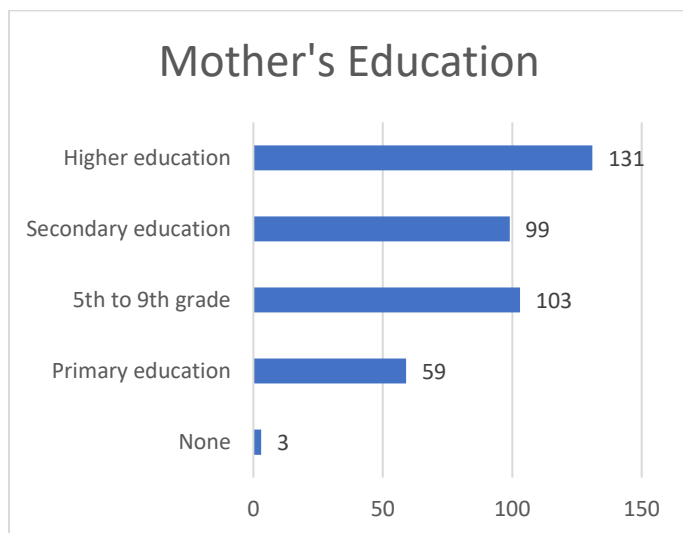
Figure 3.7: Students' Mothers' Education distribution

- Mother with higher education is higher in number. Almost 33%. Where father is only 24%.
- Mothers are more likely highly educated than the father.

Father's Education:
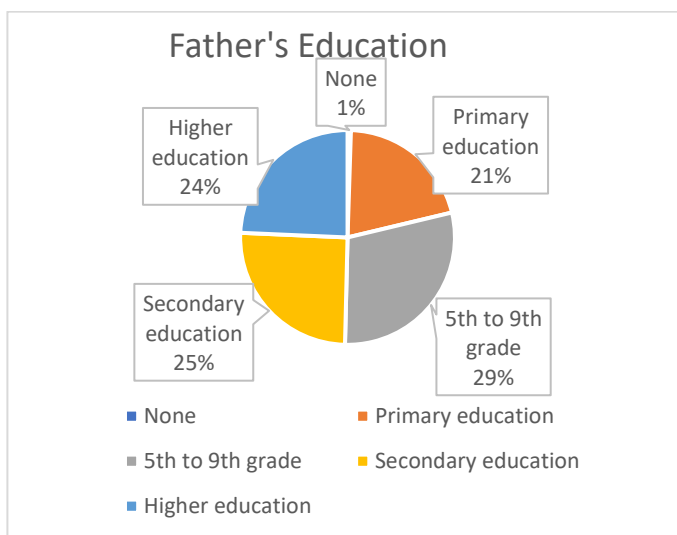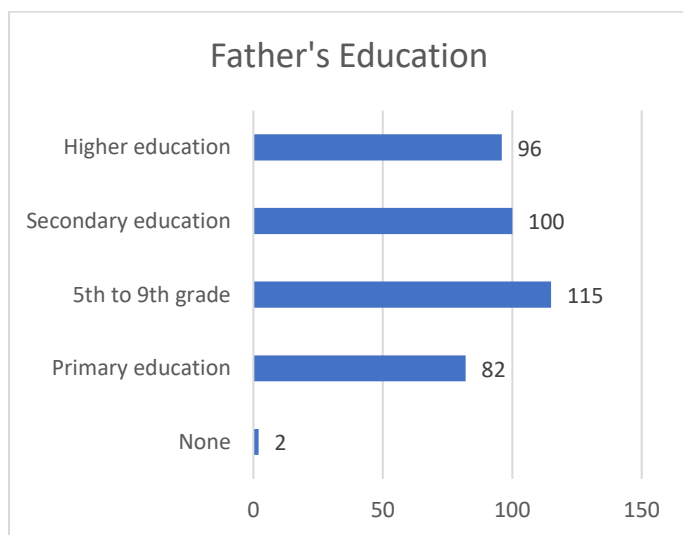
Figure 3.8: Students' fathers' Education distribution

- Most of the student's father's education is between 5th to 9th grade.

Mothers' Job:
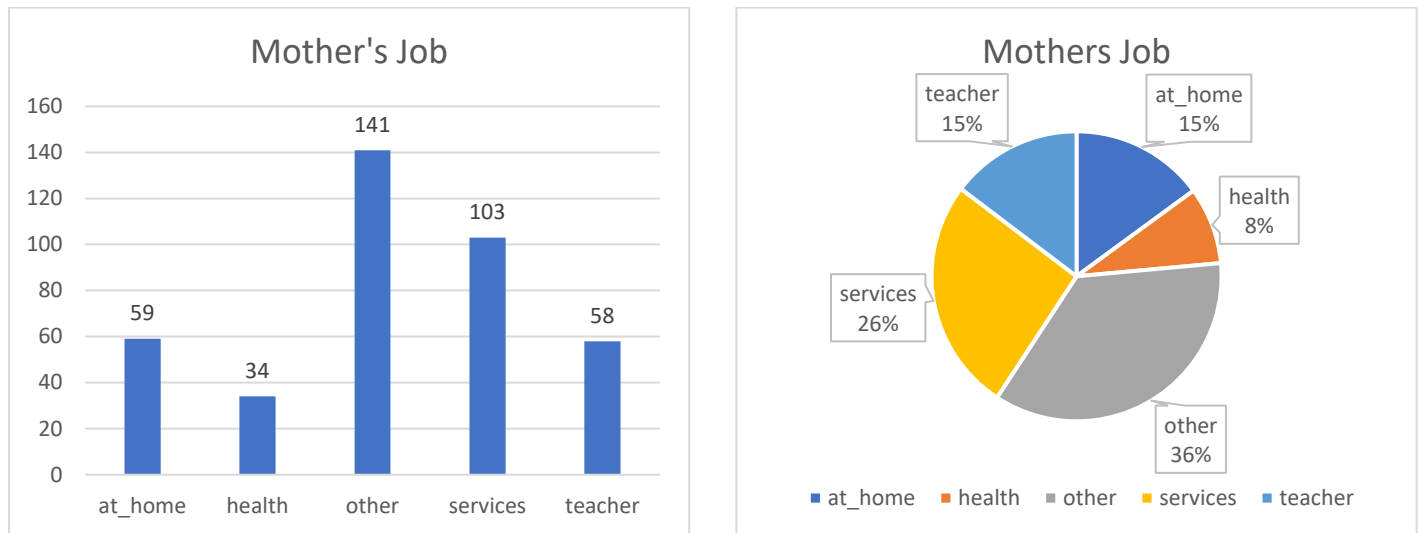
Figure 3.9: Students' Mothers' job distribution

Father's Job:
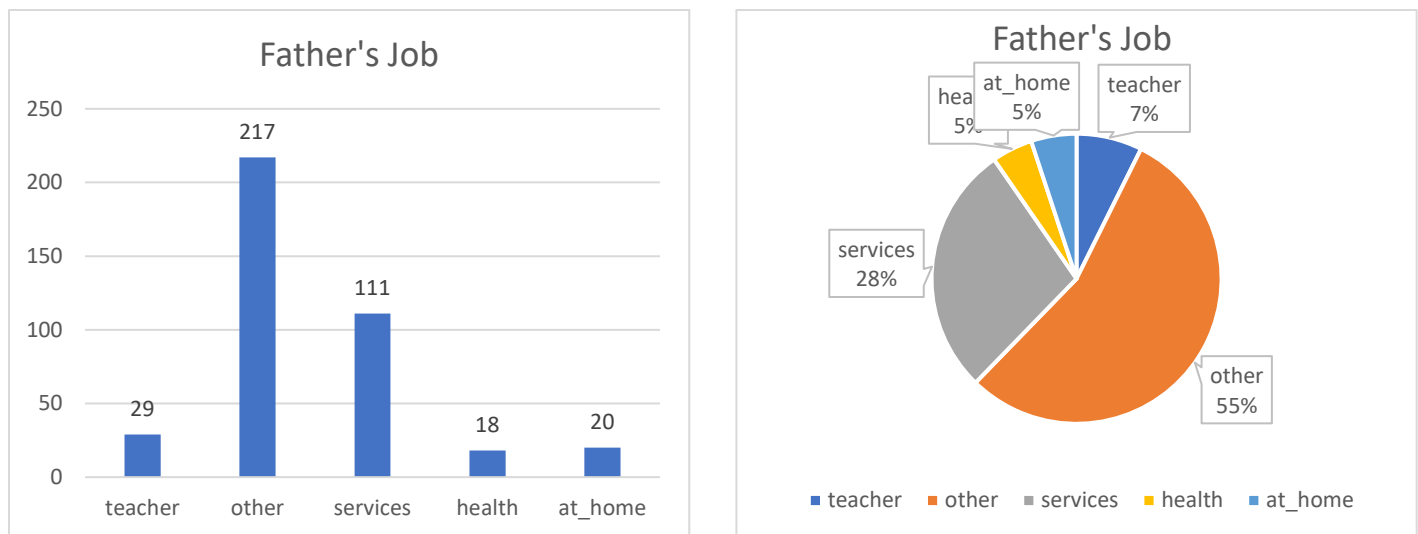
Figure 3.10: Students' fathers' job distribution

- 15% student's mother's job is at home. Where only 5% fathers' job is at home.
- Mothers are more likely at home than the father.
- Mother are more likely at teaching and health professions than the father.

Reason to choose the school:
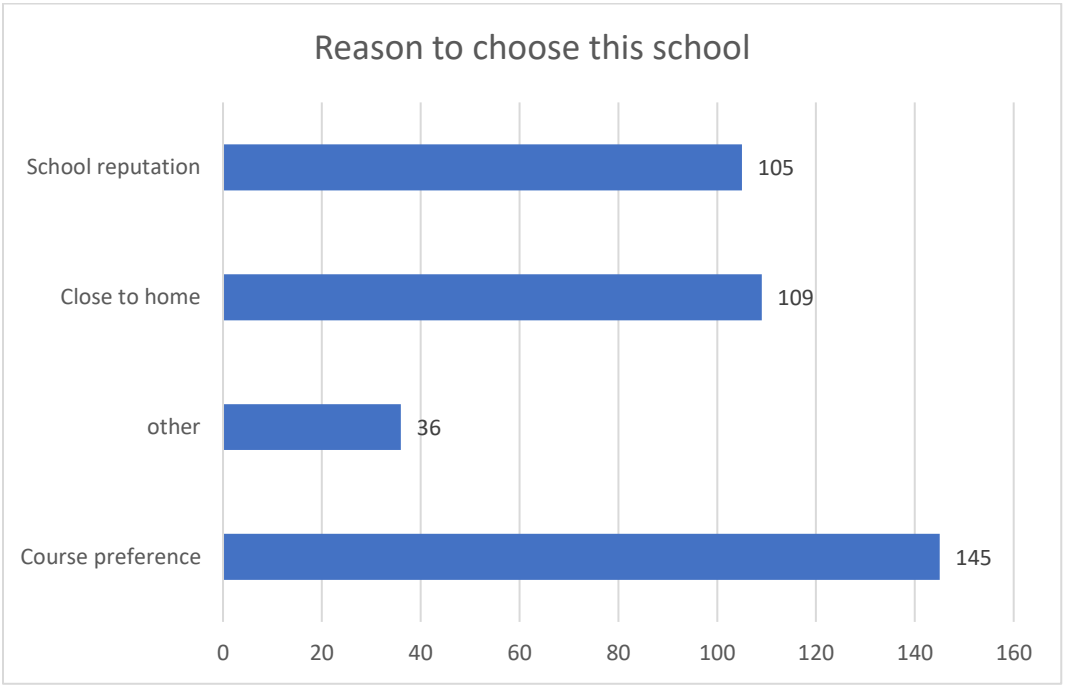
Figure 3.11: Students' Reason to choose this school

- Course preference is the most important factor to choose the school.

Guardian:
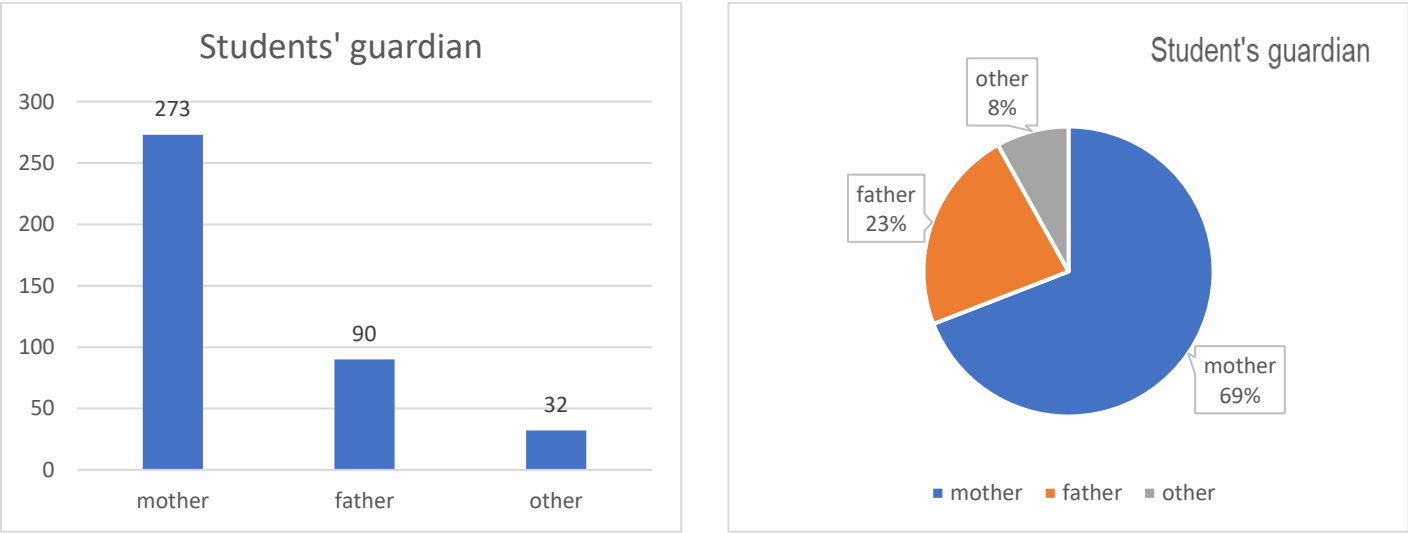
Figure 3.12: Students' guardian distribution

- Most of the cases student's guardian is his/her mother (almost 69% cases)

Home to school travel time:

Home to school travel time

	1<15 min	15 to 30 min	30 min to hour	> 1 hour
	257	107	23	8
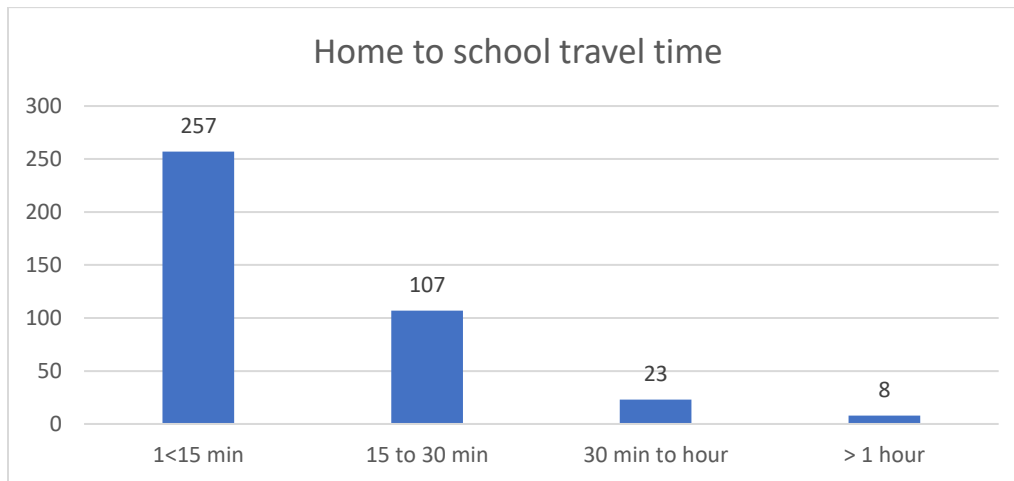
Figure 3.13: Students' home to school travel time

- Most of the student live with in a very short distance from the school (travel time <15 min).
- There is a high tendency to live closer to the school.
- Travel time distribution is right skewed.

Weekly study time:

Weekly study time

	< 2 hours	2 to 5 hours	5 to 10 hours	4 > 10 hours
	105	198	65	27
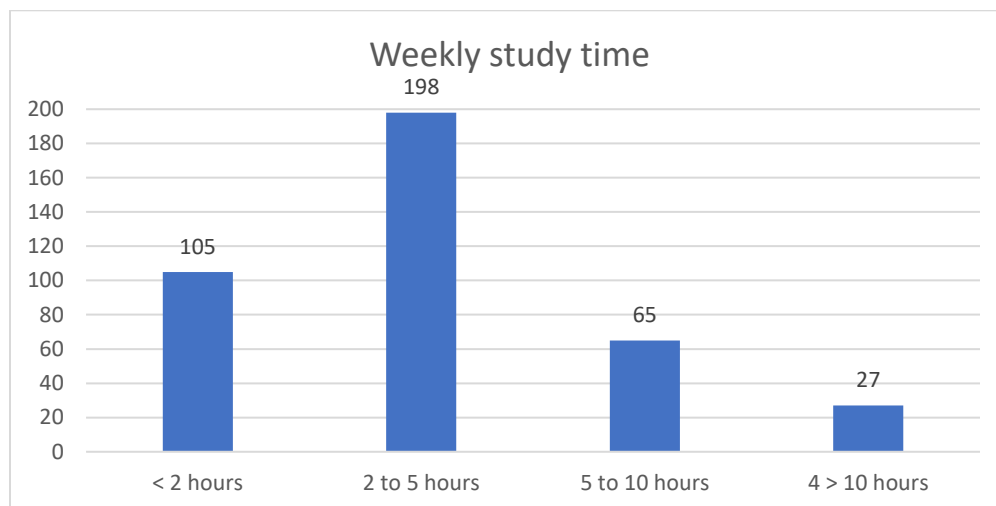
Figure 3.14: Students' Weekly study time

- Most of the students read 2 to 5 hours in a week.

Number of past class failures:
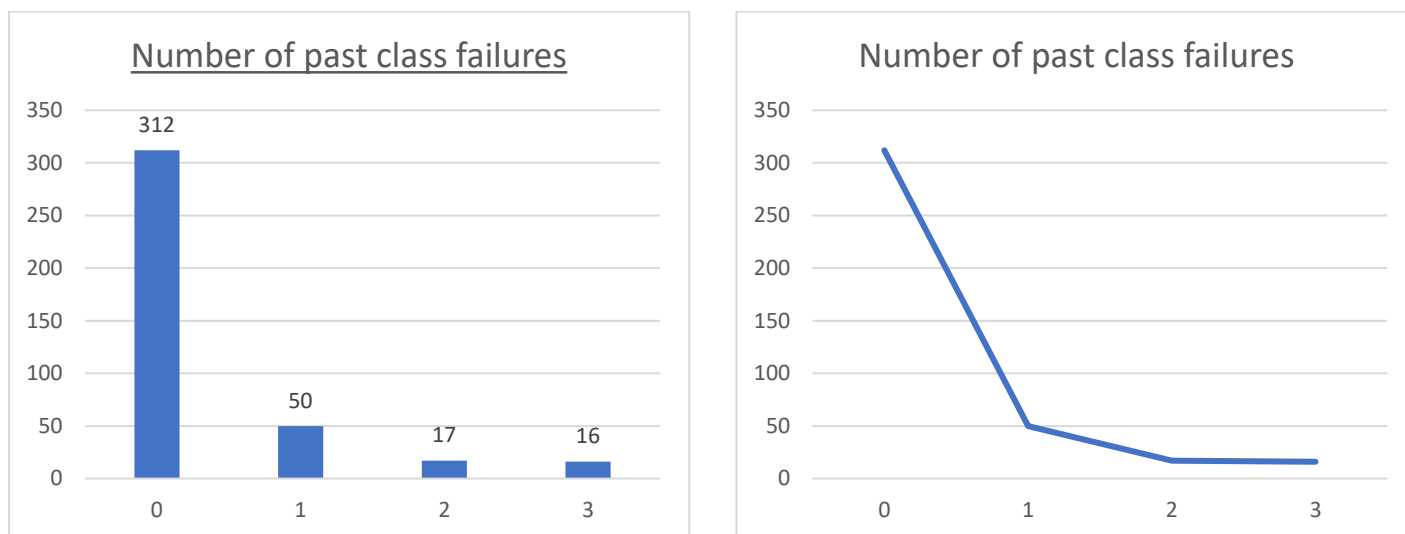
Figure 3.15: Students' number of past class failures

- Only a small number of students (around 21%) have failed before in past class.
- Failure distribution is right-skewed.

Extra educational school support:
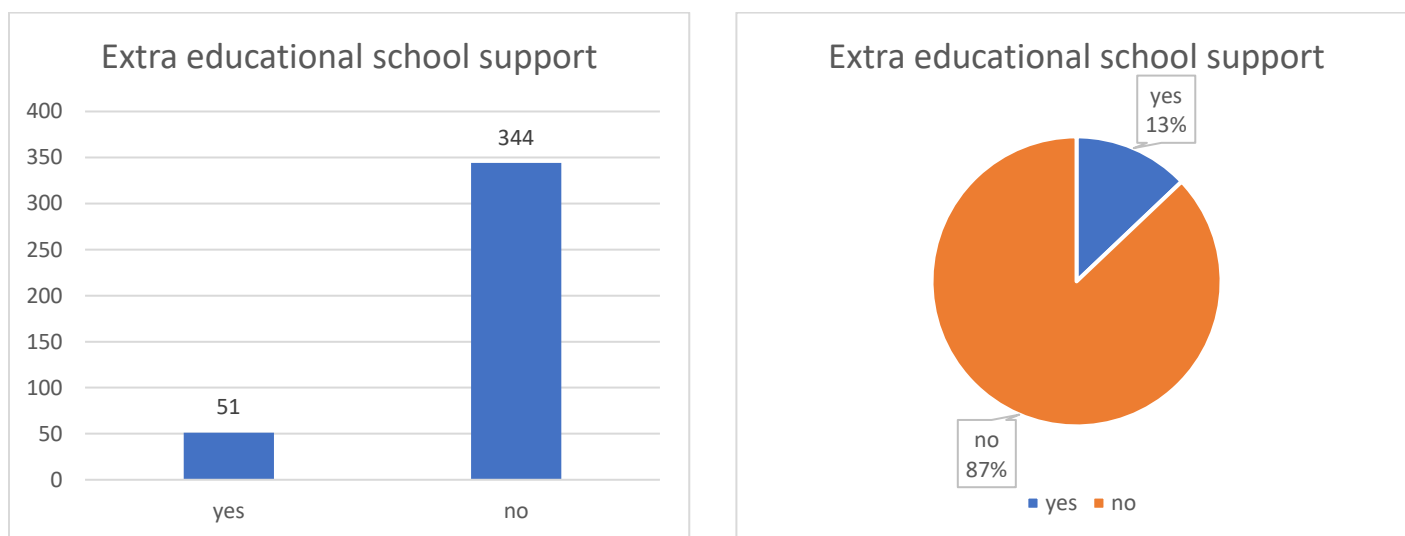
Figure 3.16: Students' extra education school support distribution

- Only 13% students have extra education school support.

Family education support:
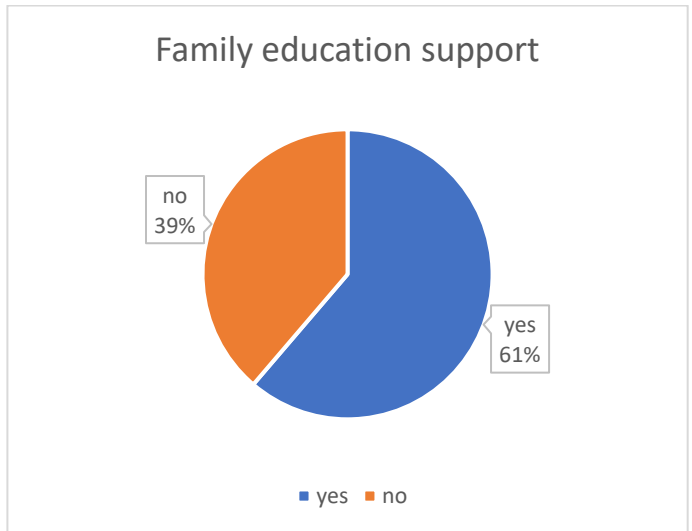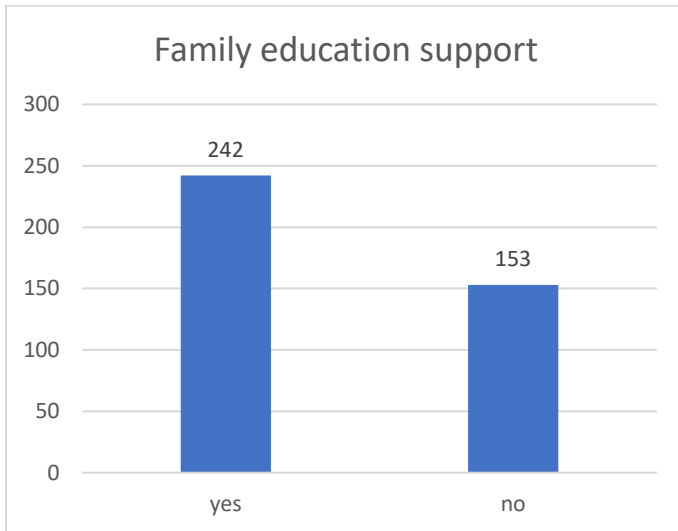
Figure 3.17: Students' family education support

- 61% student have family education support.

Extra paid classes within the course subject:
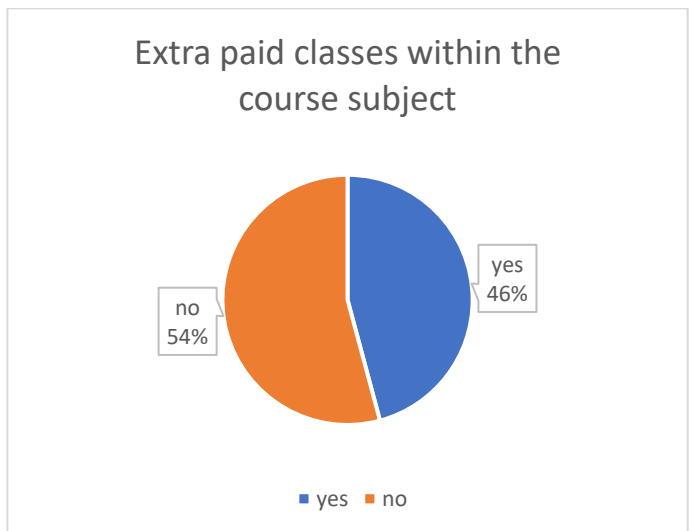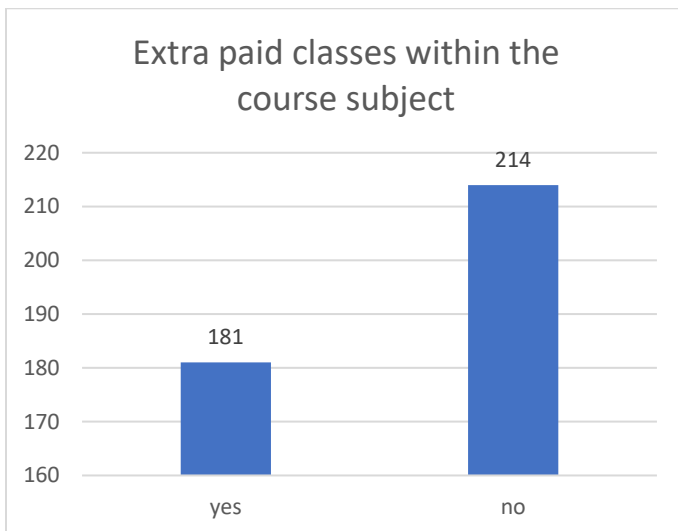
Figure 3.18: Students' extra paid classes within the course subject

- Only 46% student have extra paid classes within the course subject.

Extracurricular activities:
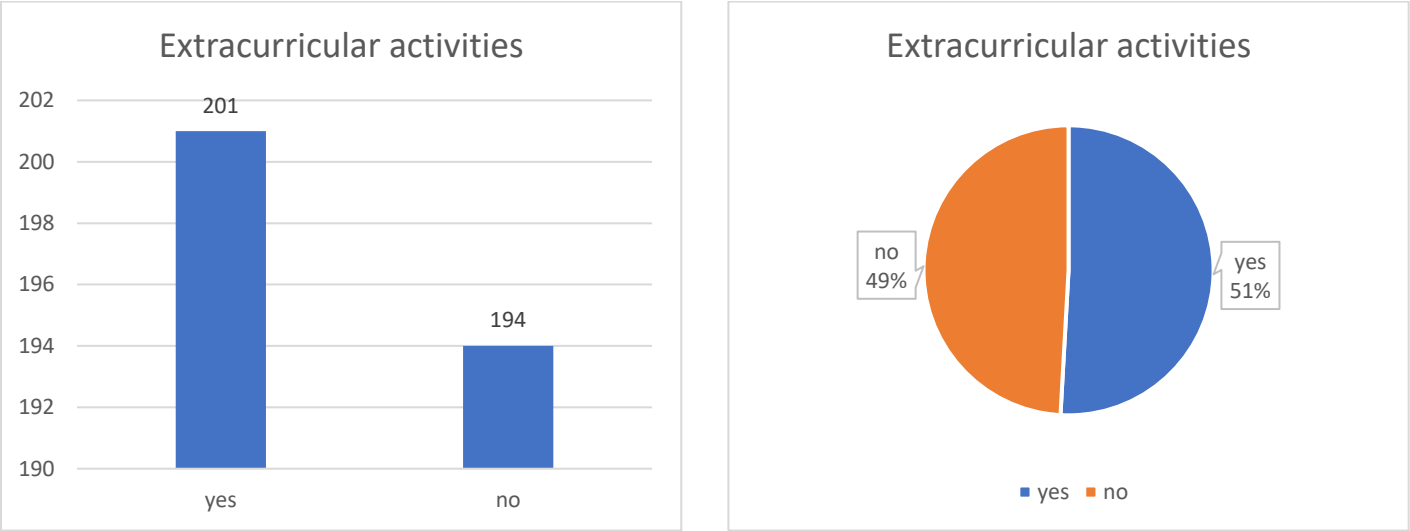
Figure 3.19: Students' extra-curricular activities

- Almost half of the students have extra-curricular activities.

Attended nursery school:
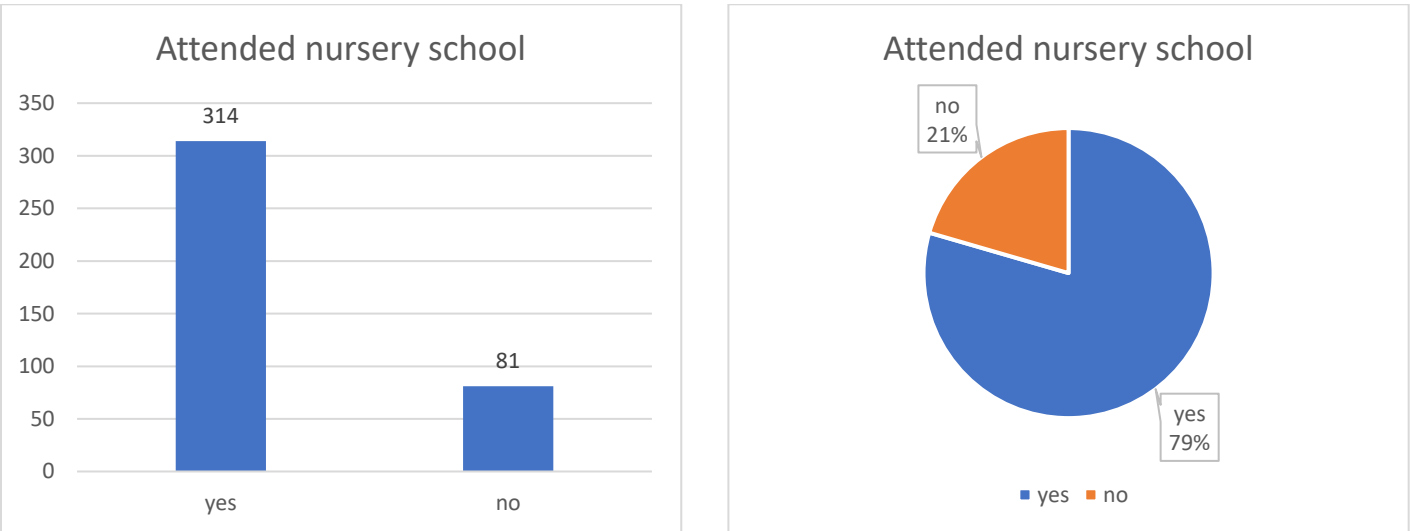
Figure 3.20: Students attended nursery school

- Majority of the student (almost 80%) had attended nursery school.

Students wants to take higher education:
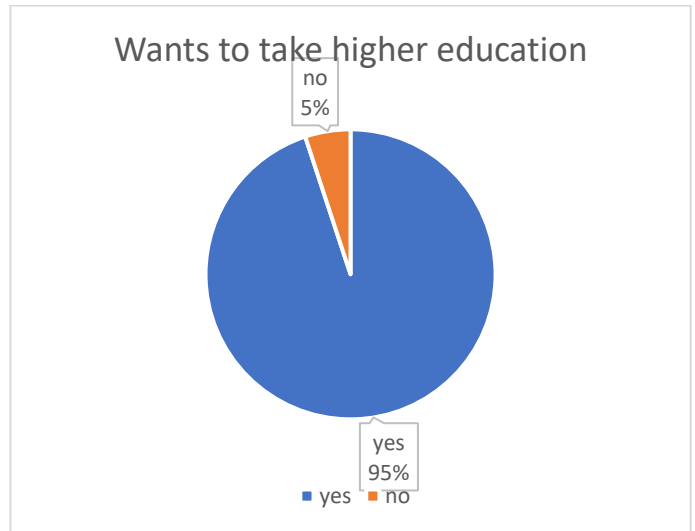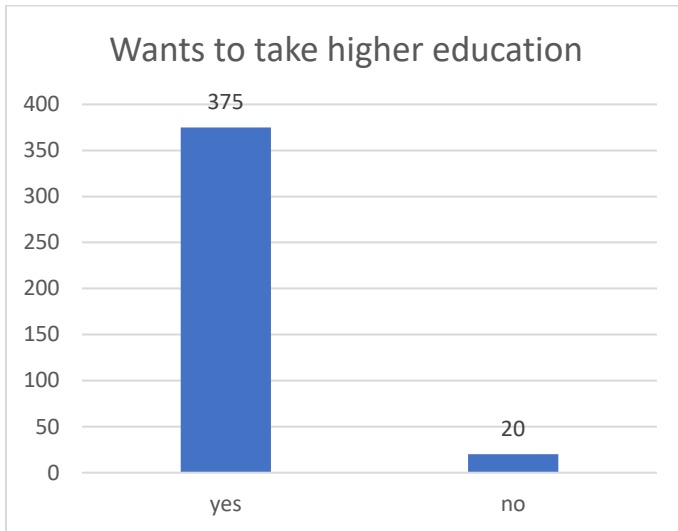
Figure 3.21: Student wants to take higher education.

- 95% students want to take higher education.

Student's Internet access at home:
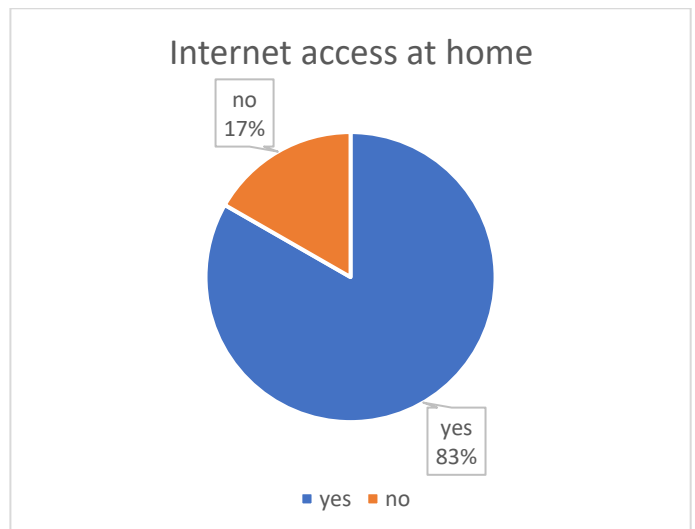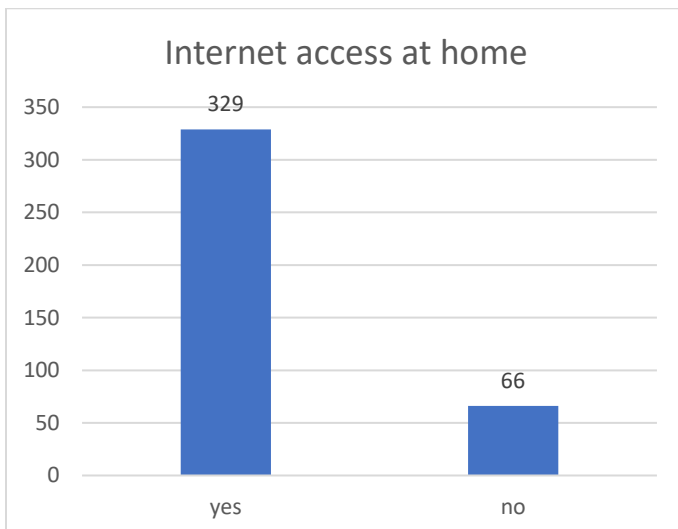
Figure 3.22: Students internet access at home

- 83% student have internet access at home.

Students with a romantic relationship:
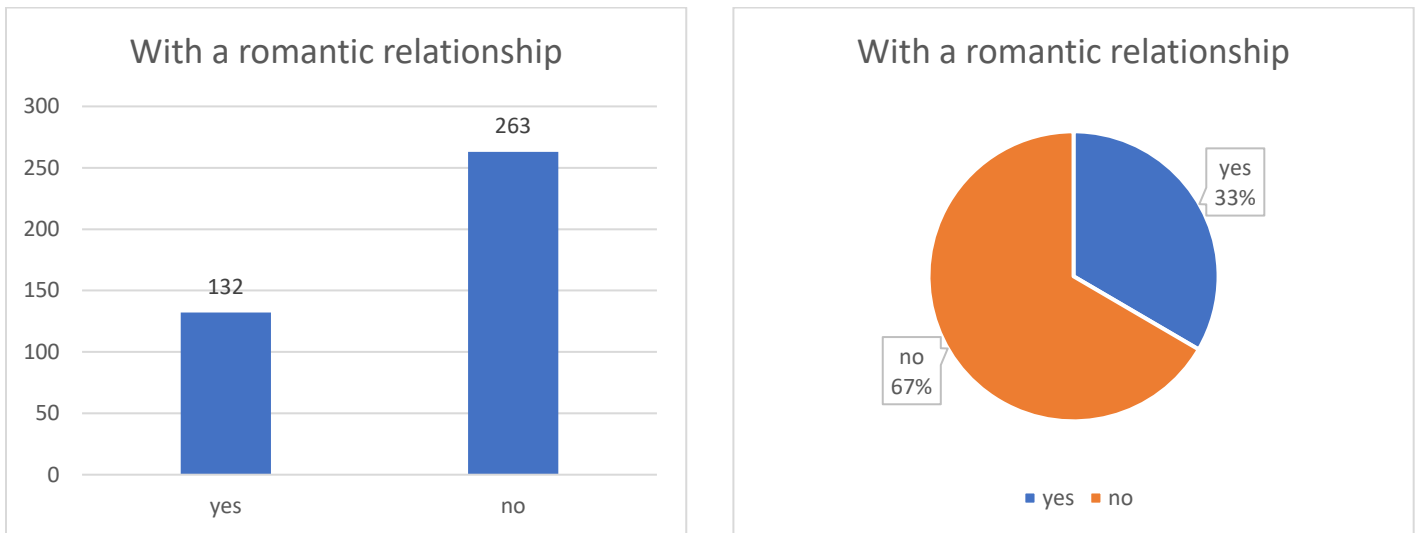
Figure 3.23: Students with a romantic relationship

- One third of the students are with a romantic relationship.

Quality of family relationship:

Figure 3.24: Students quality of family relationship
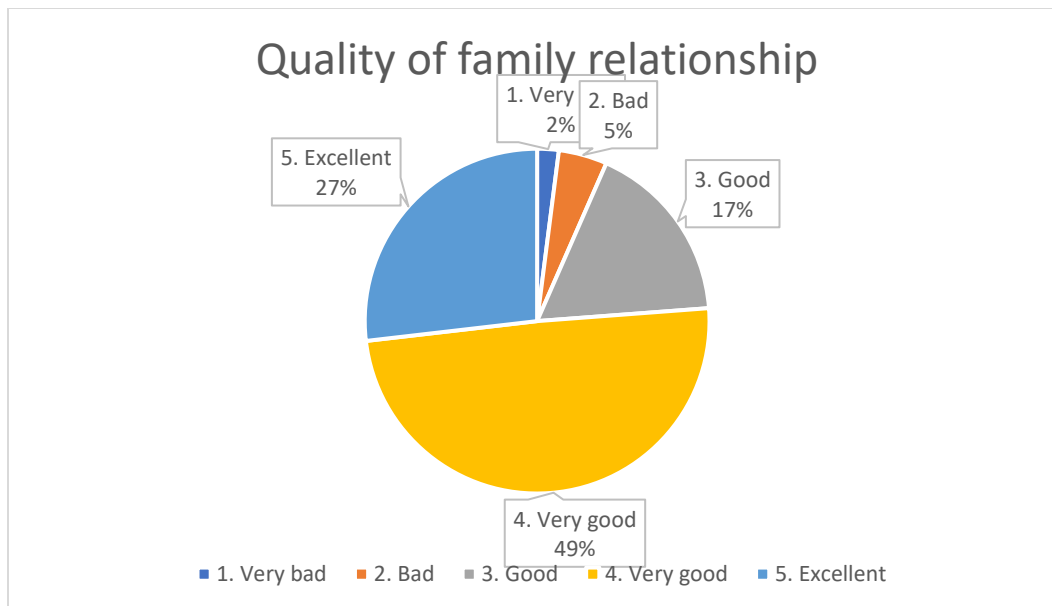
Figure 3.25: Students quality of family relationship

- 93% student have positive relationship with their family.
- Very few students have bad or very bad family relationship.
- Famrel distribution is left-skewed.

Free time after school:
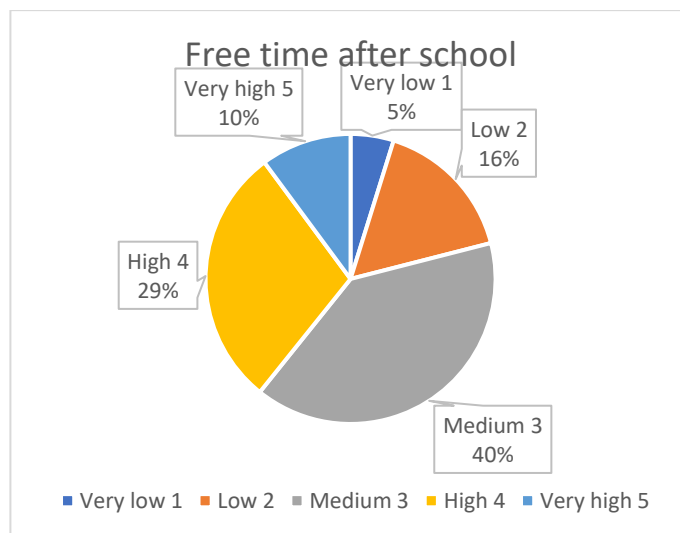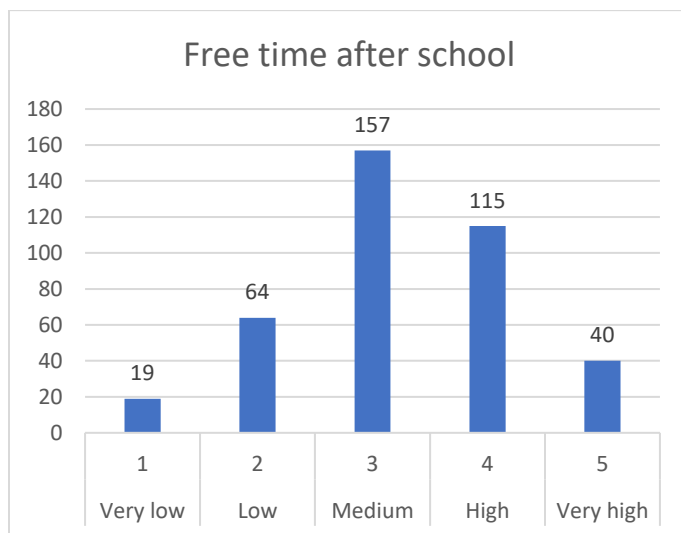
Figure 3.26: Students after school free time distribution

- Free time after school almost looks like a normal distribution.

Going out with friends:
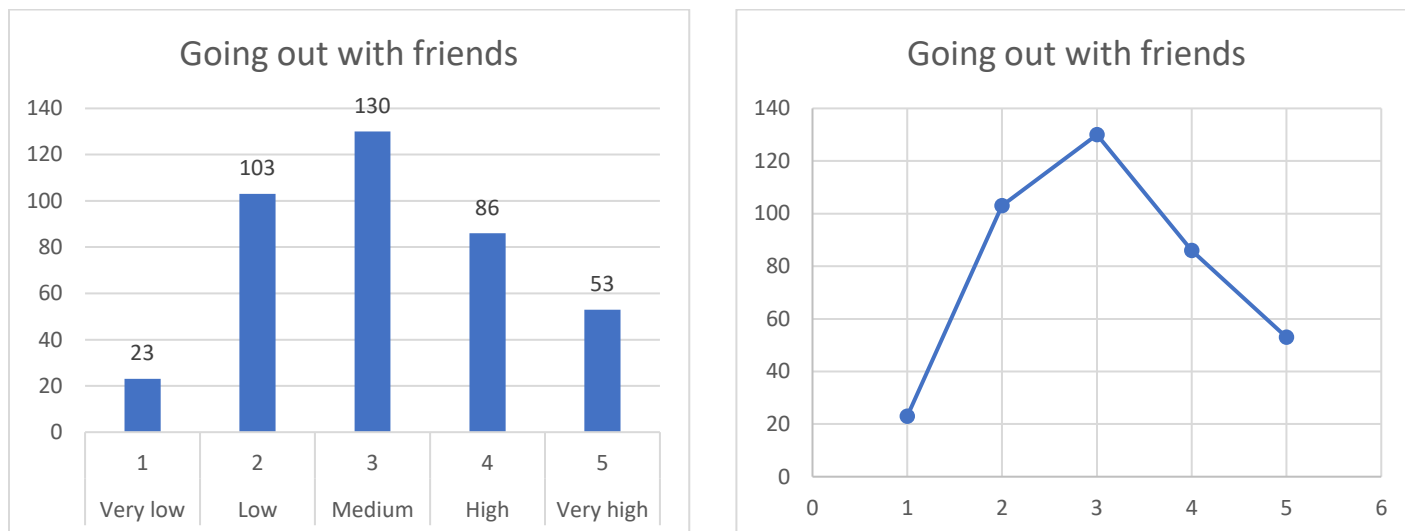
Figure 3.27: students going out with friends.

- Going out with friends' distribution almost looks like a normal distribution.

Workday alcohol consumption:
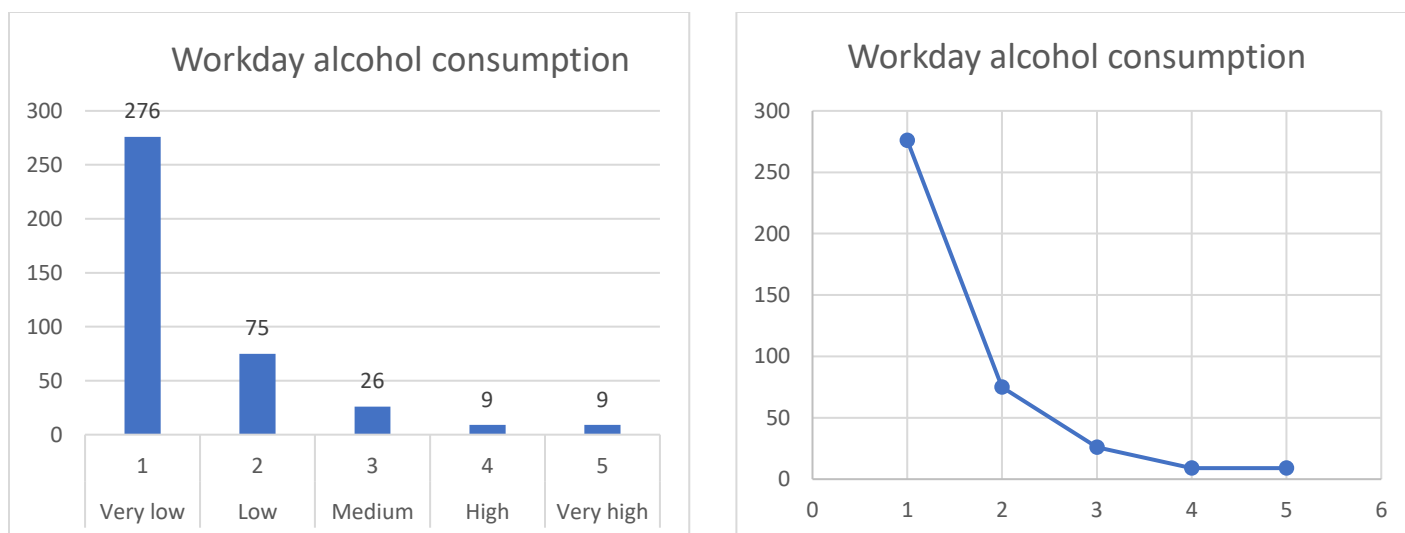
Figure 3.28: students' workday alcohol consumption

- Majority of the student drink very low alcohol in the workday.
- Compared to the weekend, they tend to drink far less during the workday.
- Dalc distribution is right-skewed.

Weekend alcohol consumption:
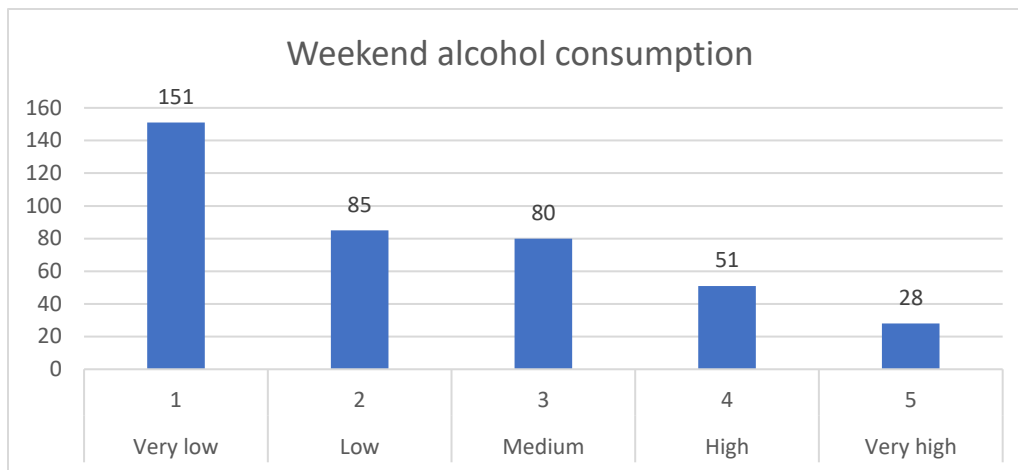
Figure 3.29: students' weekend alcohol consumption

- Weekend alcohol consumption is higher than the workdays consumption.
- Walc distribution is right-skewed.

Current health status:
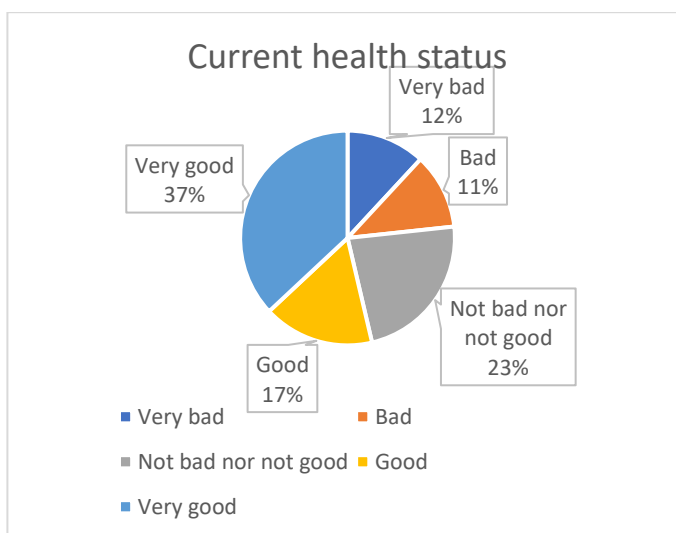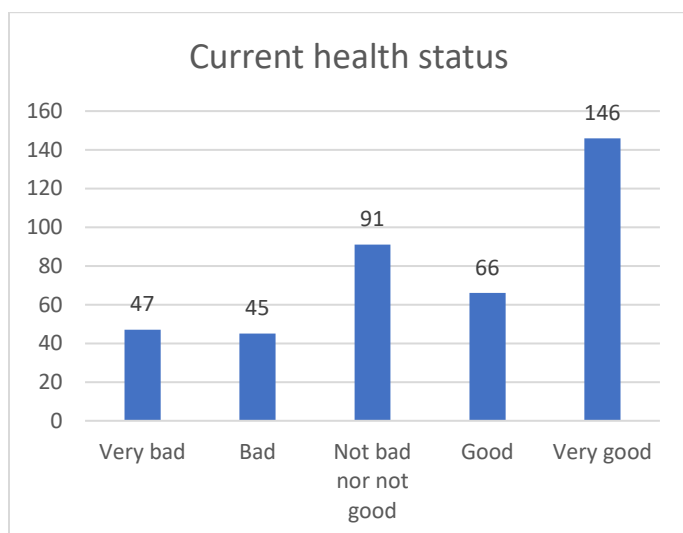
Figure 3.30: students' current health status

- Student health condition is left skewed.
- Only 23% student health is in bad or in very bad condition.

Number of school absences:
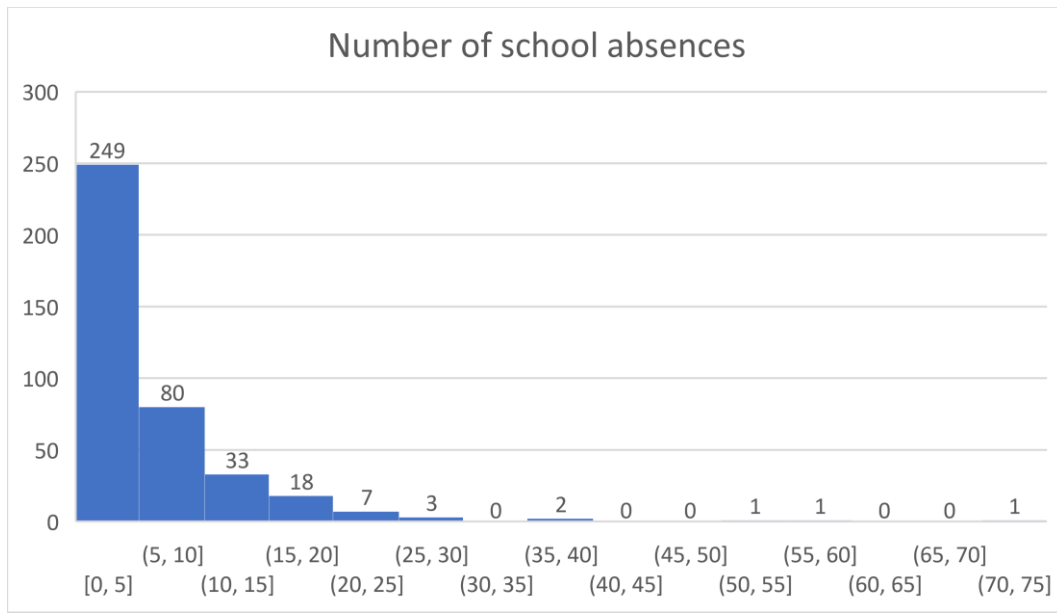
Figure 3.31: students' number of school absences

- 83% students are very attentive in the school.
- Absence distribution is right skewed.

First period grade:
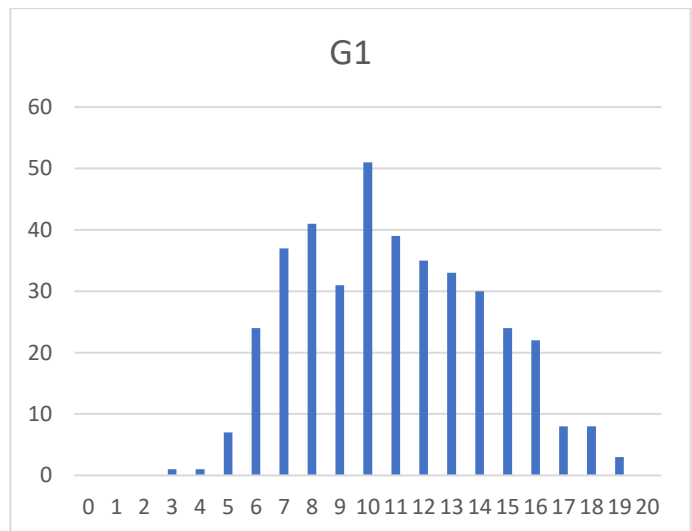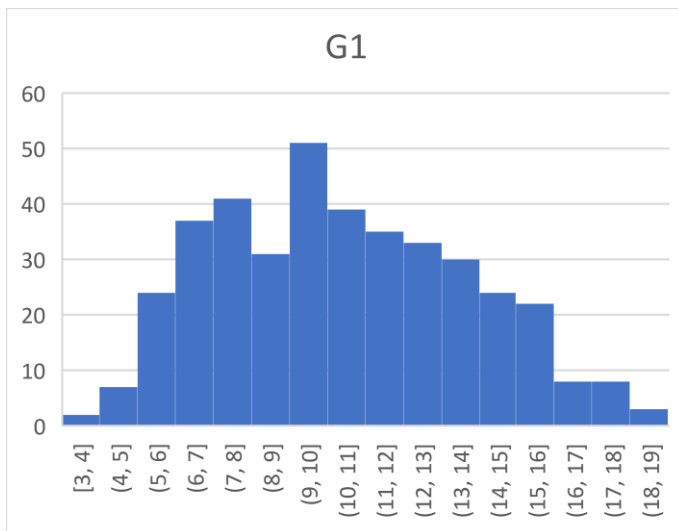
Figure 3.32: students' G1 (first period grade)

- G1 (first period grade) distribution is almost normal distribution.

Second period grade:
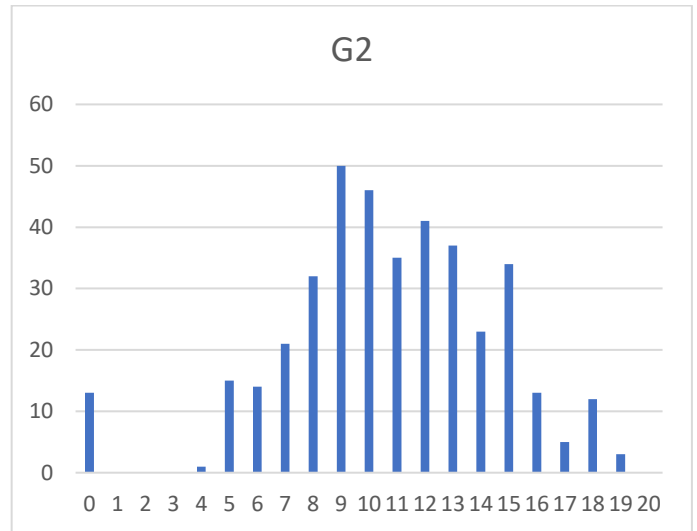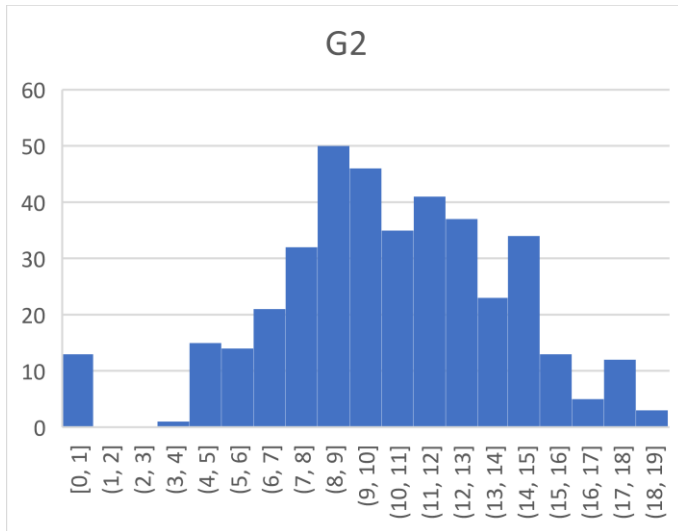
Figure 3.33: students' G2 (Second period grade) distribution

- The G2 (Second period grade) distribution is almost looks like normal distribution.

Final grade:
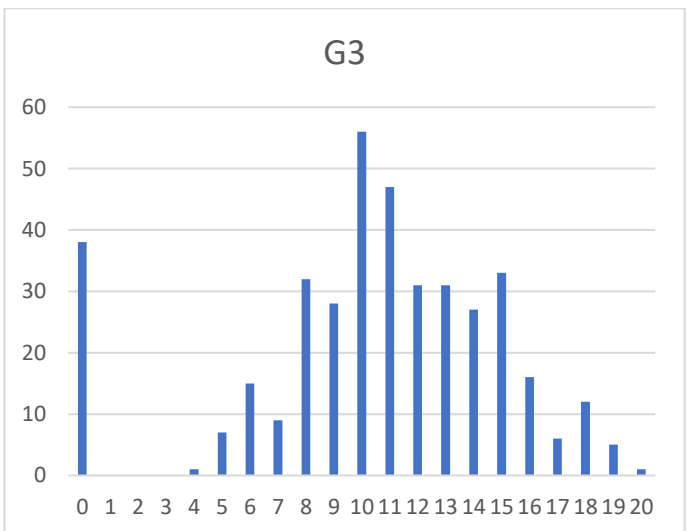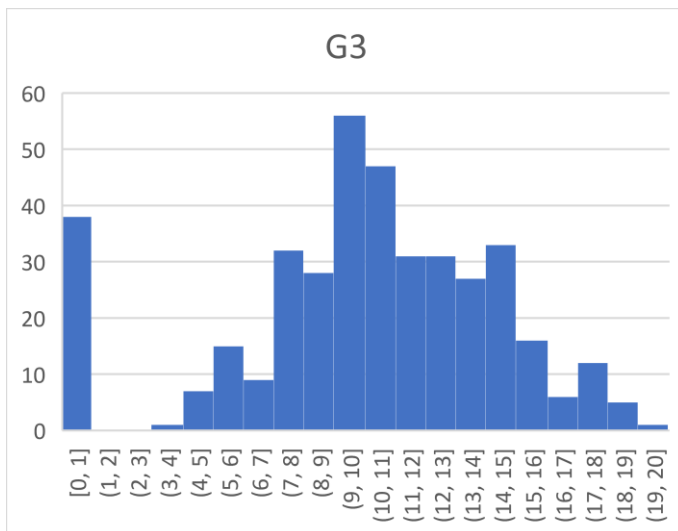
Figure 3.34: students' G3 (Final grade) distribution

- The G3 (Final grade) distribution is almost like normal distribution.

Pass and Fail in G3 (Final grade):

G3	Number of Students
Pass	265
Fail	130

Table 3.1: Students pass and fail in final grade.
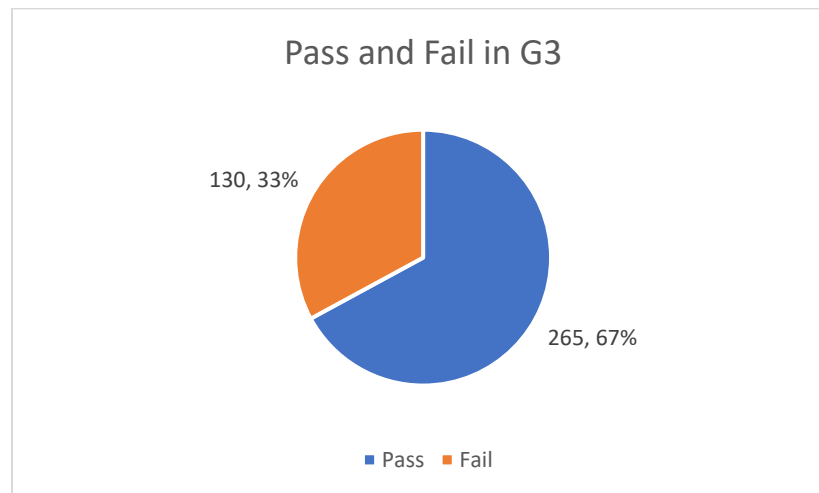
Figure 3.35: Pass and Fail in final grade.

- 67% students have passed in the final grade.
- One third of the students have failed in the final grade.

Five levels in final grade:

Levels	Number of students
Fail (0-9)	130
Sufficient (10-11)	103
Satisfactory (12-13)	62
Good (14-15)	60
Excellent (16-20)	40

Table 3.2: Students five levels in final grade.
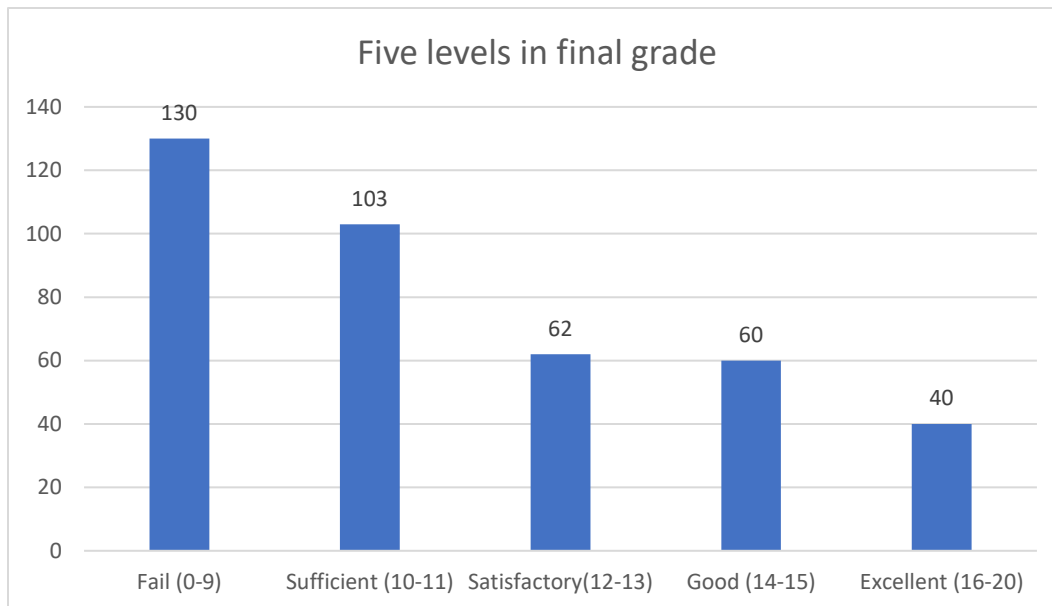
Figure 3.36: Five levels in final grade.

- Students who have passed in final grade, majority of then have passed by the grade between 10 to 13.

3.2 Relationship among the variables

Age and School:

Age	Gabriel Pereira	Mousinho da Silveire
15	82	0
16	104	0
17	86	12
18	57	25
19	18	6
20	1	2
21	0	1
22	1	0

Table 3.3: Students school by age
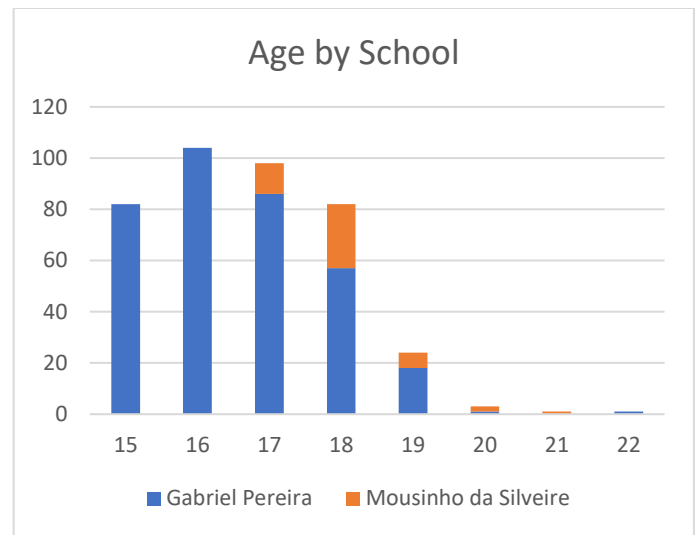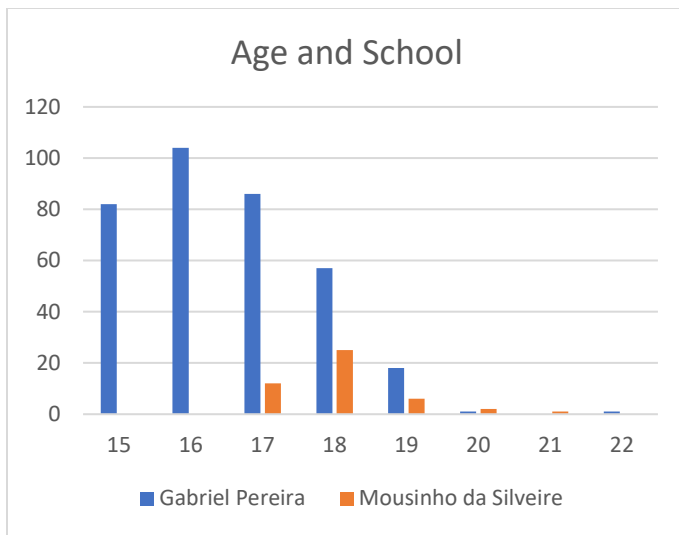
Figure 3.37: Age distribution by School

- The students at Mousinho da Silveire tend to be older.

School and address type:

Address type	Gabriel Pereira	Mousinho da Silveire
Urban	286	21
Rural	63	25

Table 3.4: Students School and address distribution
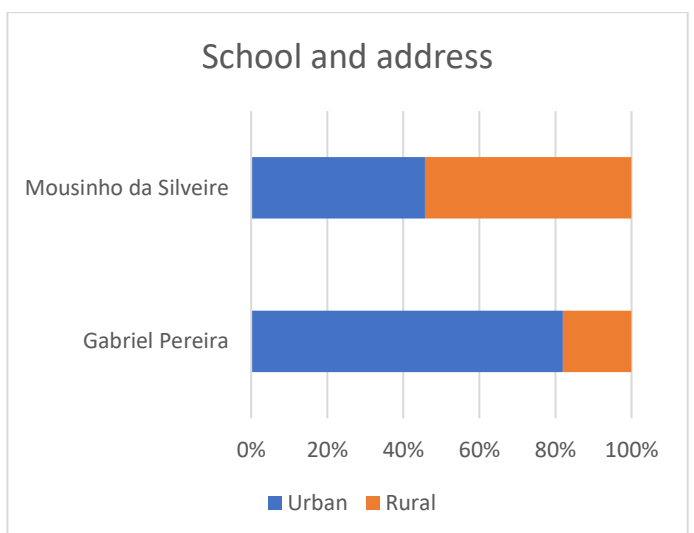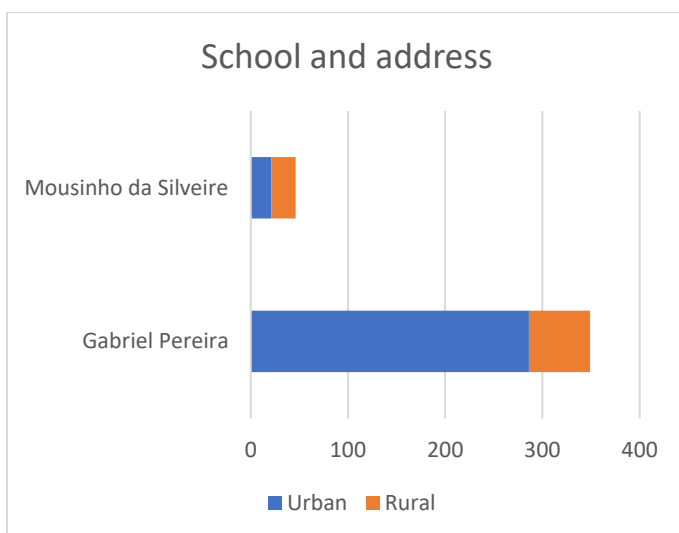
Figure 3.38: Students School and address distribution

- More than 80% of the Gabriel Pereira school's students are from the urban area.

School and travel time:

Traveltime	Gabriel Pereira	Mousinho da Silveire
1<15 min	243	14
15 to 30 min	84	23
30 min to hour	15	8
> 1 hour	7	1

Table 3.5: Students school and travel time distribution
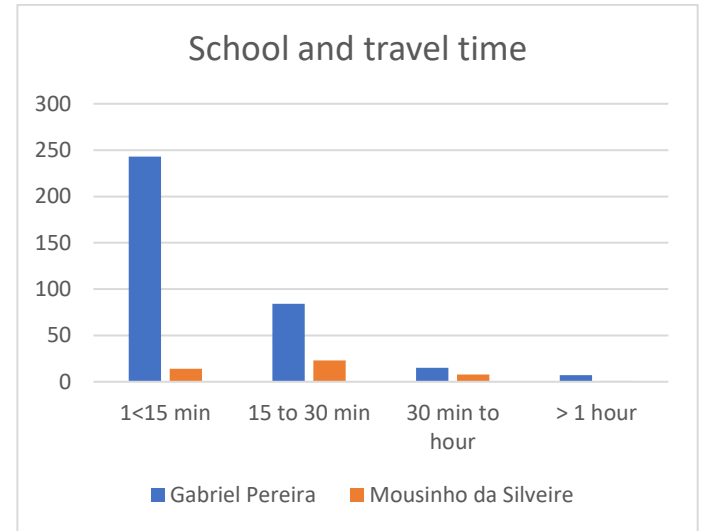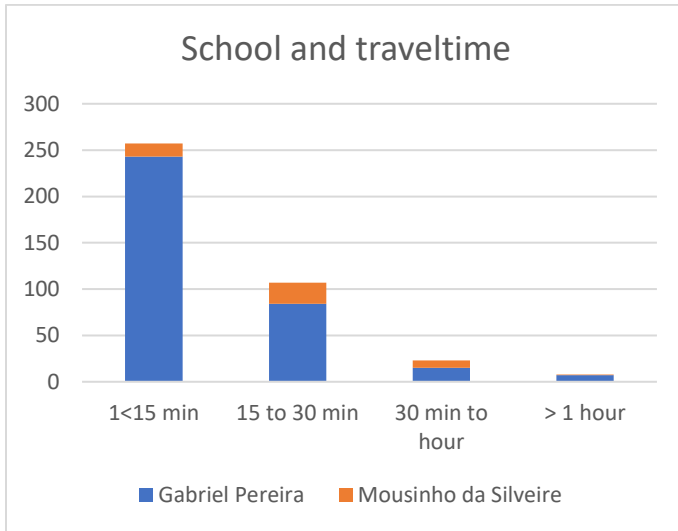
Figure 3.39: Students school and travel time distribution

- Students of Gabriel Pereira are more likely to choose school near the home than students of Mousinho da Silveire.

School and extra educational school support:

schoolsup	Gabriel Pereira	Mousinho da Silveire
Yes	51	0
No	298	46

Table 3.6: School and extra educational school support distribution
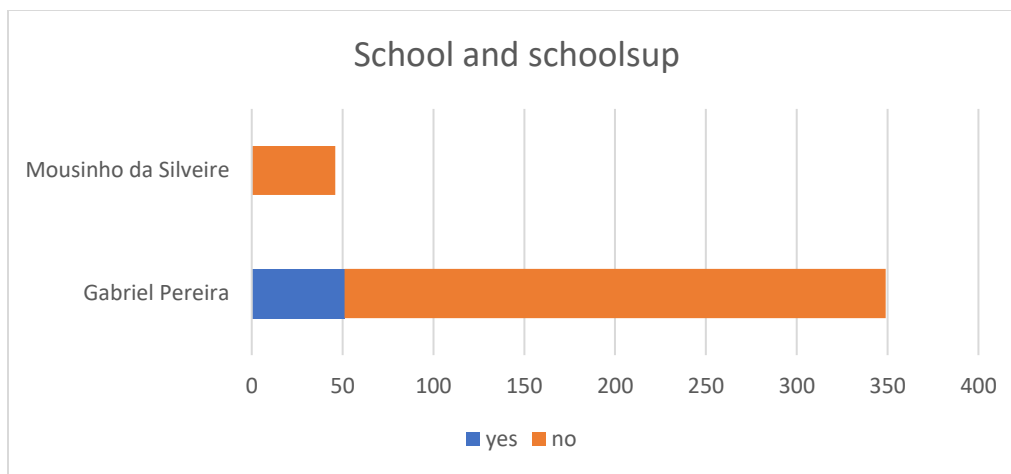
Figure 3.40: School and extra educational school support distribution

- None of the students of Mousinho da Silveire have extra educational support.

famsup	Gabriel Pereira	Mousinho da Silveire
Yes	224	18
No	125	28

Table 3.7: School and family educational support distribution
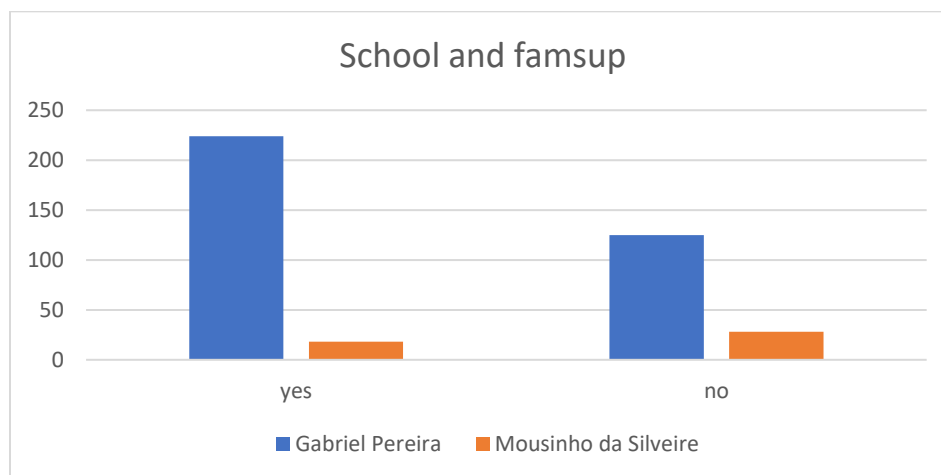
Figure 3.41: School and family educational support distribution

27
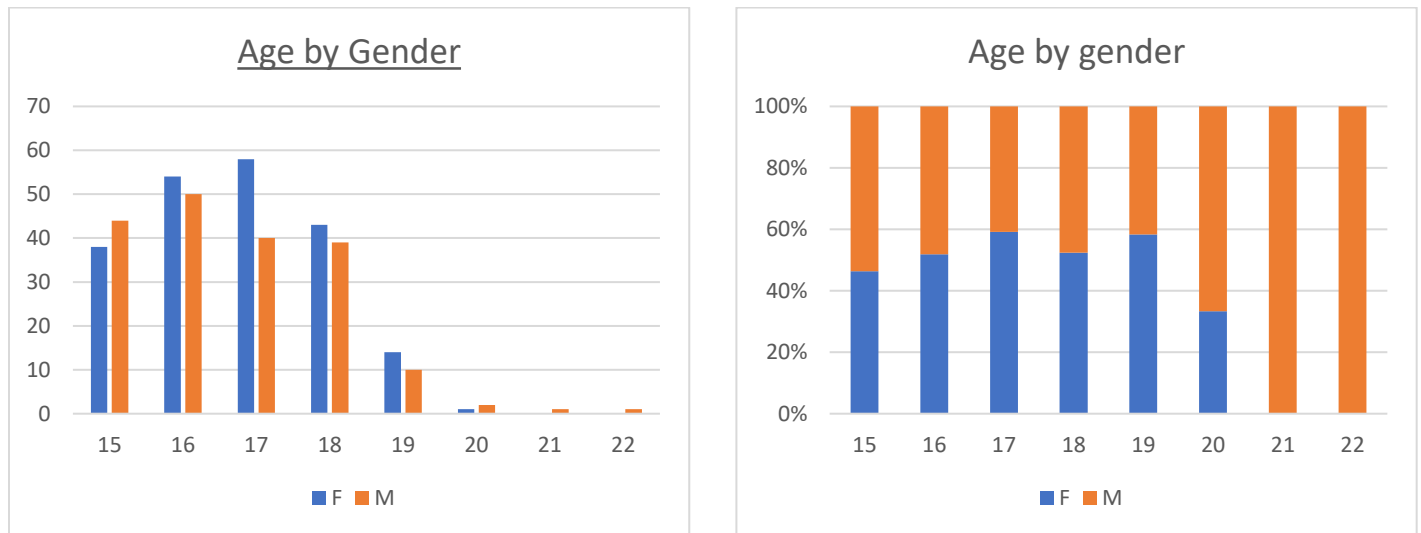
Age by Gender:

Figure 3.42: Age by gender distribution

- Female students are trends to be younger than the male students.

Gender by Final grade:
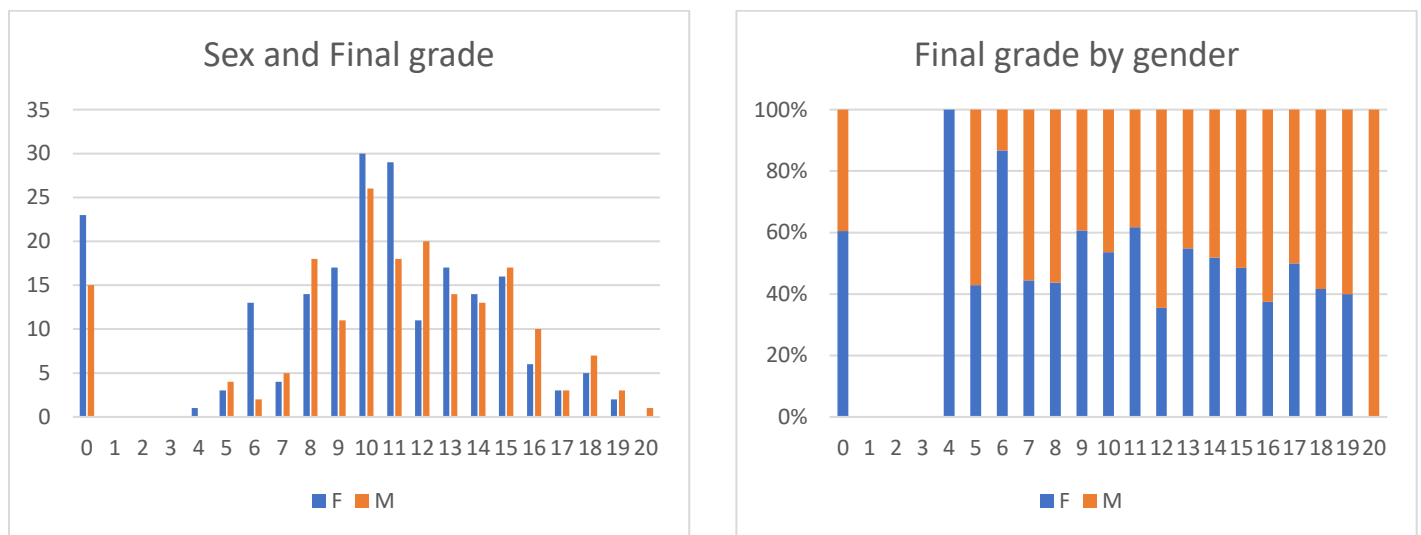
Figure 3.43: Sex and Final grade distribution
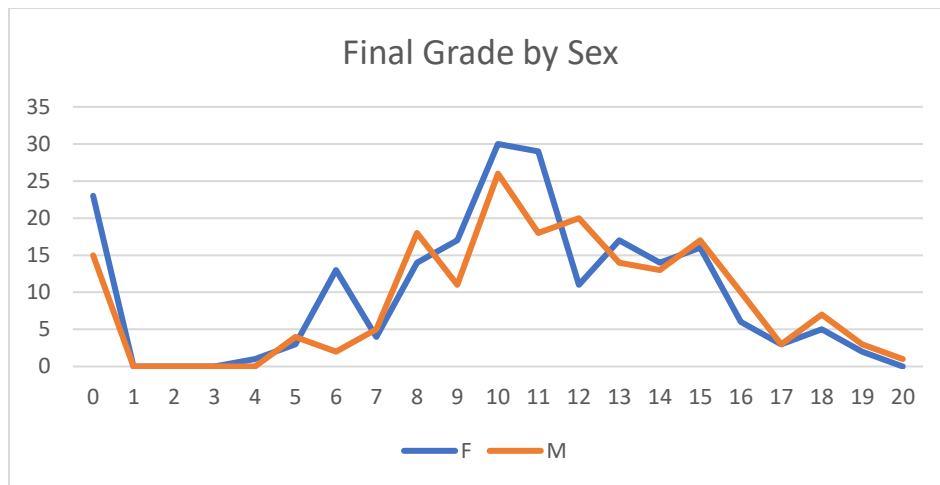
Figure 3.44: Sex and Final grade distribution

- In good and excellent grade male students are doing better than the female.

Gander and Pass/Fail:

	Pass	Fail
Female	133	75
Male	132	55

Table 3.8: Pass/fail by gender.
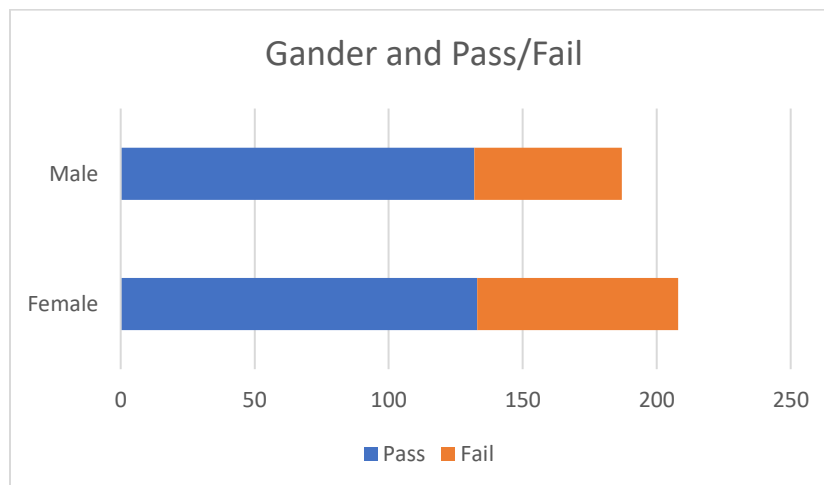
Figure 3.45: Pass/fail by gender.

- Passing ratio is slightly higher in the Male students than the Female student.

Age and Final grade:

G3	15 years	16 years	17 years	18 years	19 years	20 years	21 years	22 years
0	6	7	8	12	5	0	0	0
1	0	0	0	0	0	0	0	0
2	0	0	0	0	0	0	0	0
3	0	0	0	0	0	0	0	0
4	0	0	1	0	0	0	0	0
5	1	2	2	2	0	0	0	0
6	4	4	5	2	0	0	0	0
7	2	2	2	2	0	0	1	0
8	5	6	6	10	4	0	0	1
9	3	6	8	6	4	1	0	0
10	10	16	16	13	1	0	0	0
11	11	13	11	7	5	0	0	0
12	7	8	10	5	1	0	0	0
13	5	9	8	5	4	0	0	0
14	6	9	5	7	0	0	0	0
15	11	9	9	3	0	1	0	0
16	3	7	3	3	0	0	0	0
17	1	2	1	2	0	0	0	0
18	4	3	3	1	0	1	0	0
19	3	0	0	2	0	0	0	0
20	0	1	0	0	0	0	0	0

Table 3.9: Age and final grade distribution.

▪ Younger students' trends to do better in the exam then the older students.

Gender and study time:

studytime	Female	Male
< 2 hours	27	78
2 to 5 hours	113	85
5 to 10 hours	51	14
4 > 10 hours	17	10

Table 3.10: weekly study time by gender.
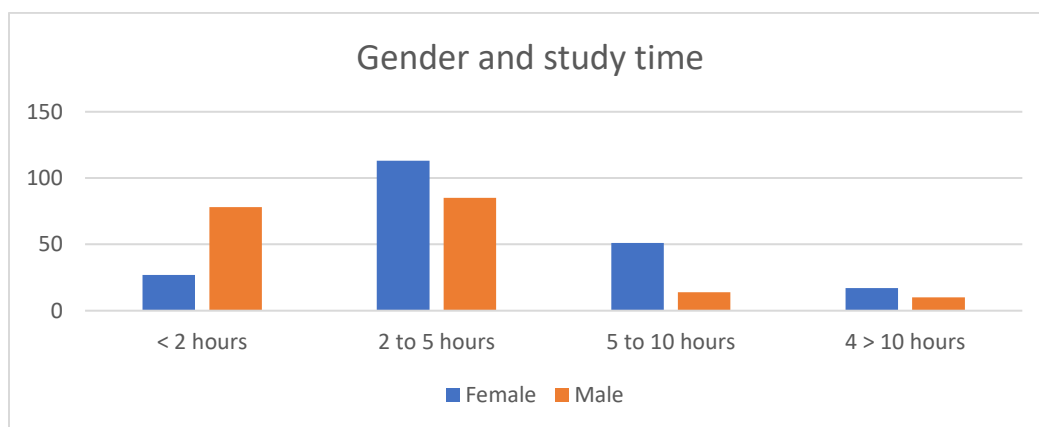
Figure 3.47: weekly study time by gender.

▪ Female students' trends to study more hours than the male students.

Gender and schoolsup:

Schoolsup	Female	Male
Yes	36	15
No	172	172

Table 3.11: Extra educational school support by gender.
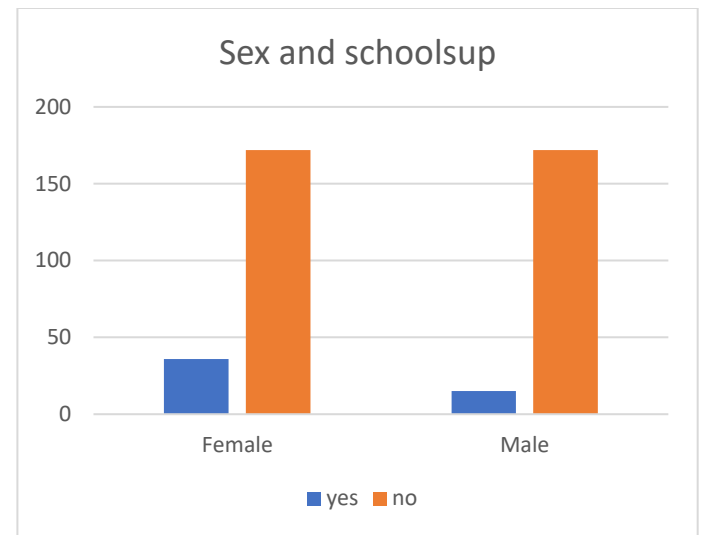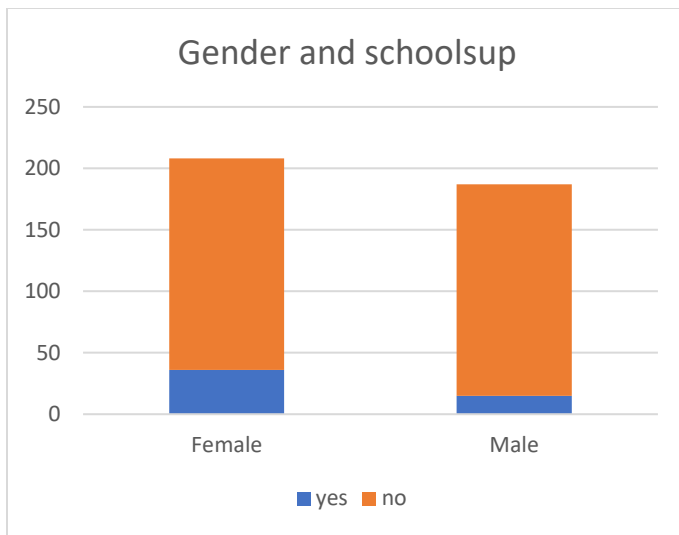
Figure 3.48: Extra educational school support by gender.

- Female students received more extra educational school support than the male student.

Gender and famsup:

famsup	Female	Male
yes	142	100
no	66	87

Table 3.12: family education support by gender.
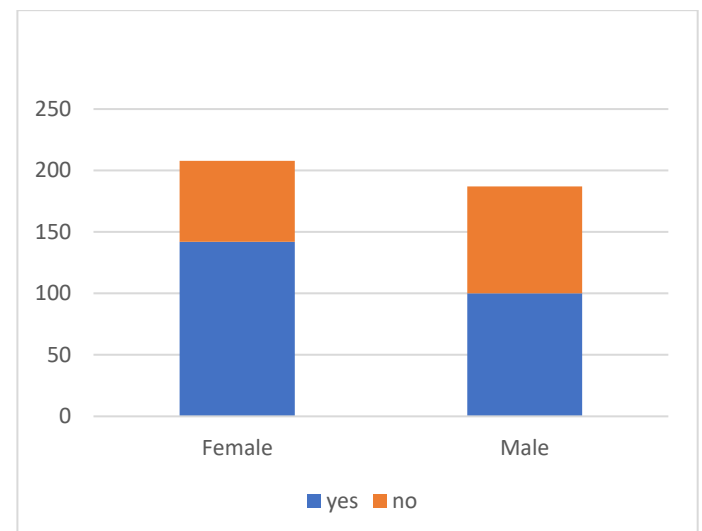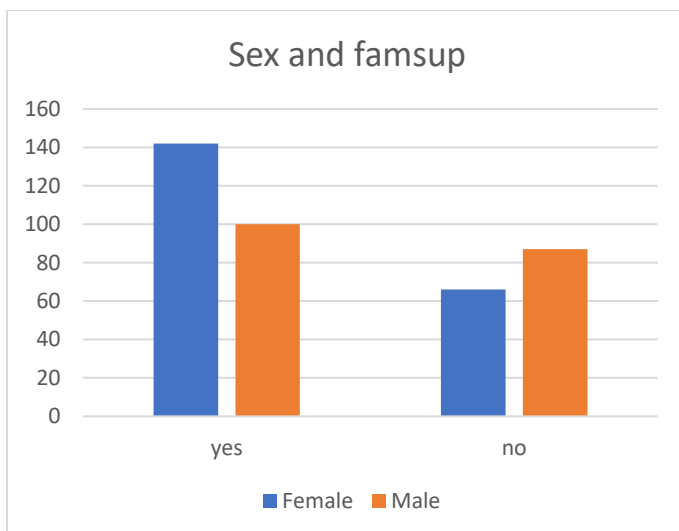
Figure 3.49: Family education support by gender.

- More female students have family education support than the male student.

- Male students are less likely than female students to receive family education support.

Gender and Paid Class:

paid		Female	Male
yes		108	73
no		100	114

Table 3.13: Extra paid classes by gender
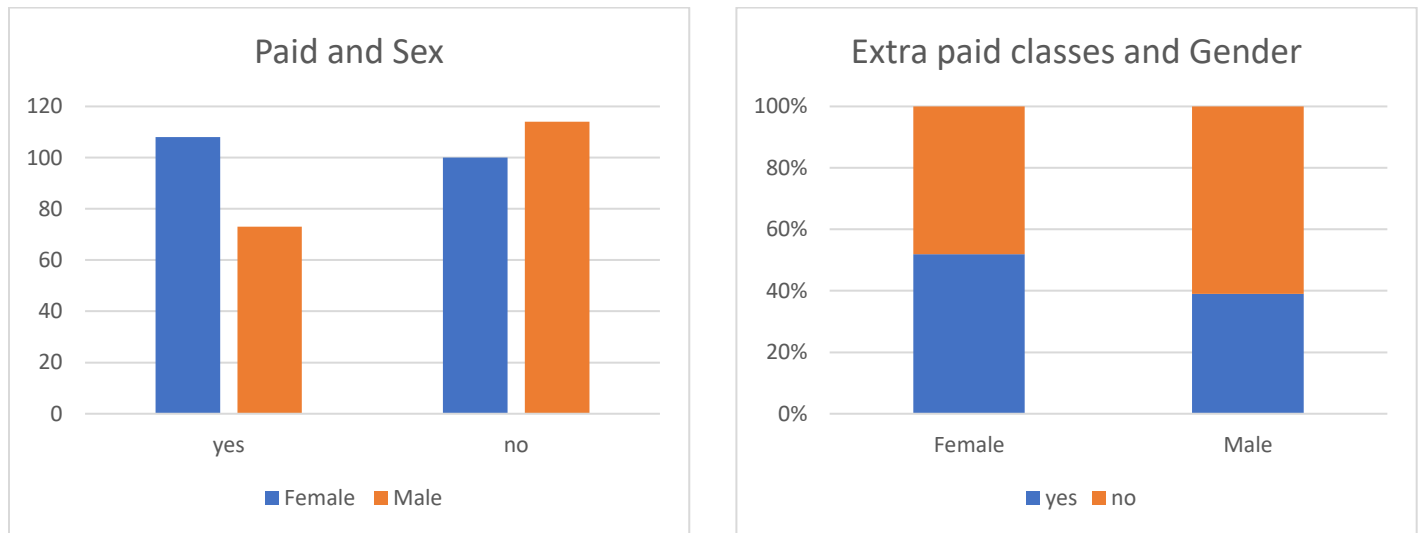
Figure 3.50: Extra paid classes by gender

- Female students are slightly (around 15%) receiving more extra paid classes within the course subject (Math and Portuguese) the Male students.

Gender and Higher education:

higher	Female	Male
yes	204	171
no	4	16

Table 3.14: Gender distribution by students wants to take higher education.
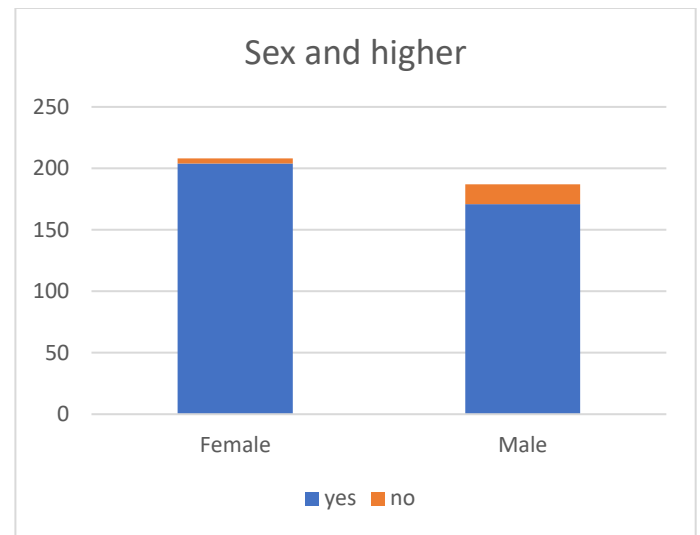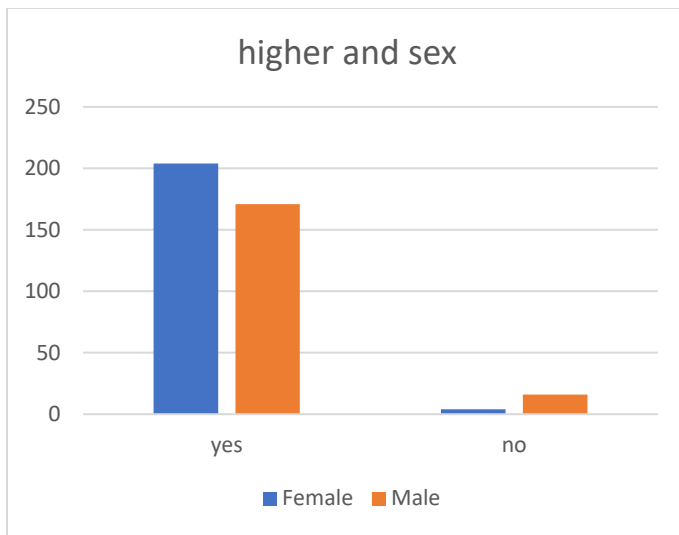
Figure 3.51: Students wants to take higher education by gender.

- Female students are more interested to take higher education than the male students.

Gender and Romantic relationship:

romantic	Female	Male
yes	79	53
no	129	134

Table 3.15: students in romantic relationship and gender.
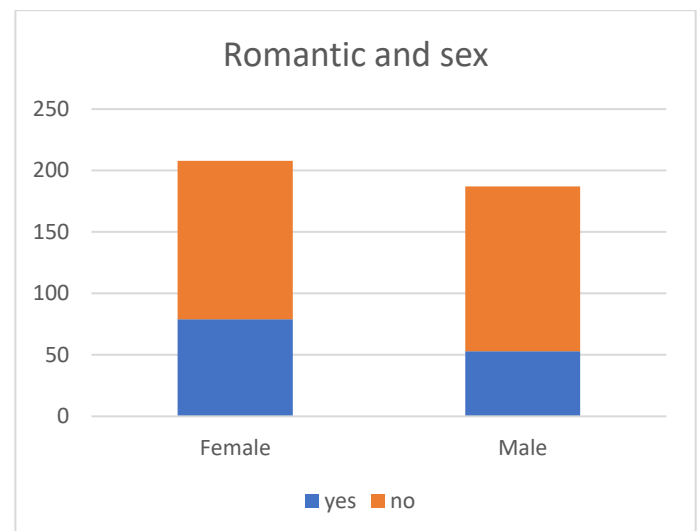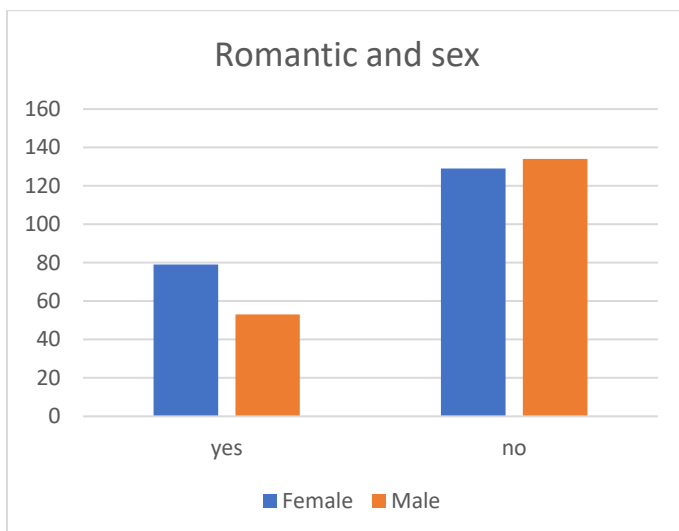
Figure 3.52: students in romantic relationship and gender.

- Most of the students are not in a romantic relationship.

- Female students are more in a romantic relationship than the male among the students who are in a relationship.

Extracurricular activities and Sex:

activities	Female	Male
yes	96	105
no	112	82

Table 3.16: Students' extra-curricular activities and gender.
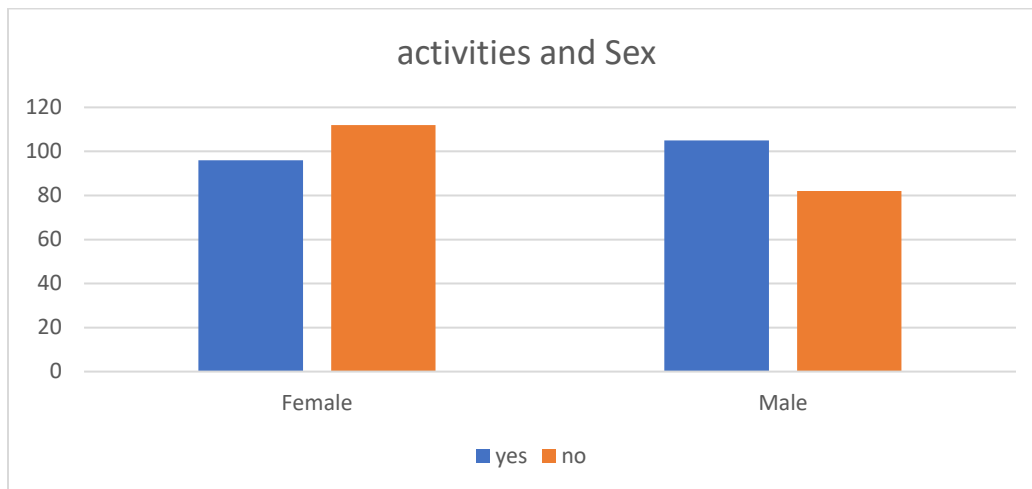
Figure 3.53: Students' extra-curricular activities and gender.

- Male students are more active in extracurricular activities than the female.

After school free time and Gender:

freetime	Female	Male
Very low	15	4
Low	38	26
Medium	93	64
High	54	61
Very high	8	32

Table 3.16: After school free time and gender distribution.
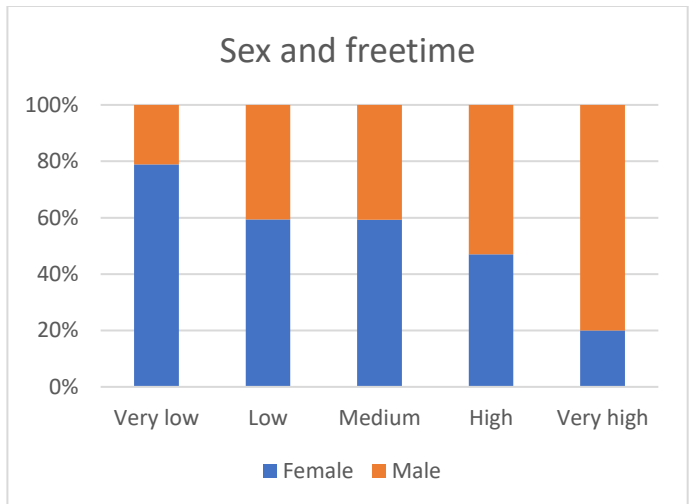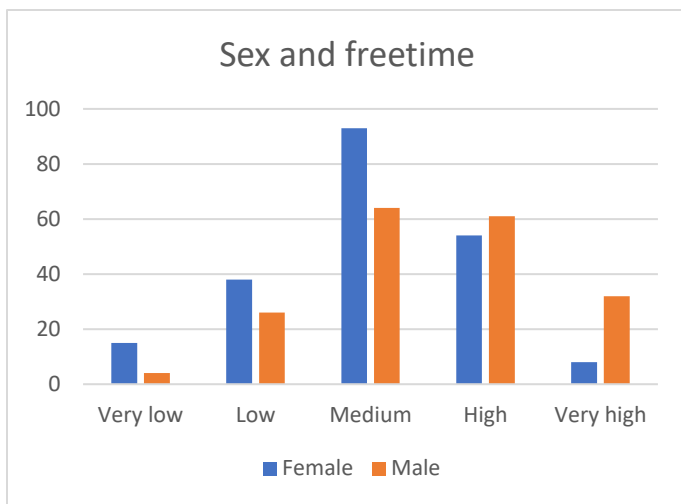
Figure 3.54: After school free time and gender distribution.

- Male students have a tendency to pass more after school free time than the females.

Working day alcohol and Sex:

Dalc	Female	Male
Very low	167	109
Low	33	42
Medium	5	21
High	2	7
Very high	1	8

Table 3.18: students' working day alcohol consumption by gender.

- Majority of the student consume very low alcohol.
- Male students have a tendency to consume more alcohol comparing to the Female.

Weekend alcohol and Sex:

Walc	Female	Male
Very low	94	57
Low	51	34
Medium	45	35
High	14	37
Very high	4	24

Table 3.19: students' weekend alcohol consumption by gender.
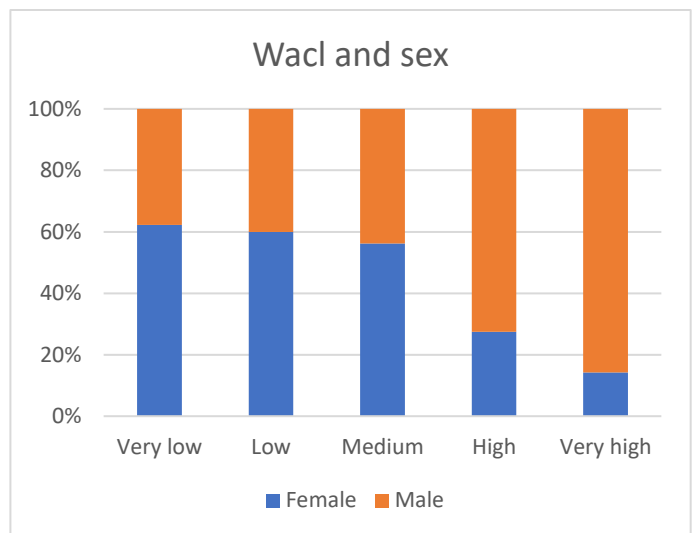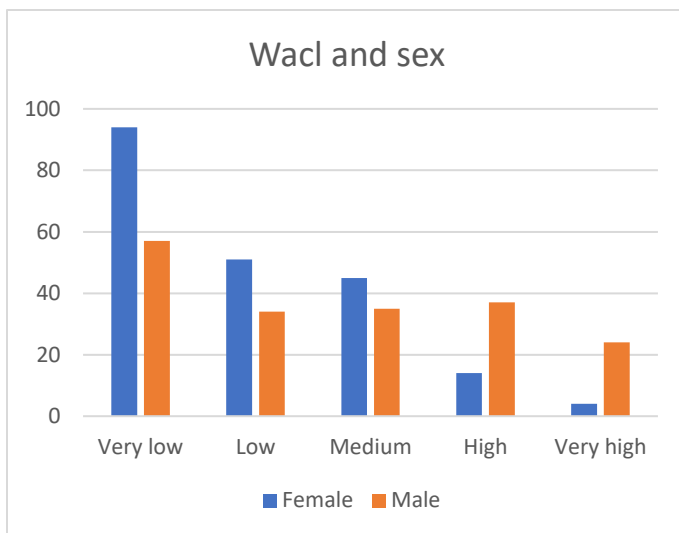
Figure 3.56: students' weekend alcohol consumption by gender.

- In the weekend they drink more alcohol than the working days.

Gender and Health status:

Health	Female	Male
Very bad	31	16
Bad	25	20

Not bad nor not good	52	39
Good	37	29
Very good	63	83

Table 3.20: Students' current health status and gender.
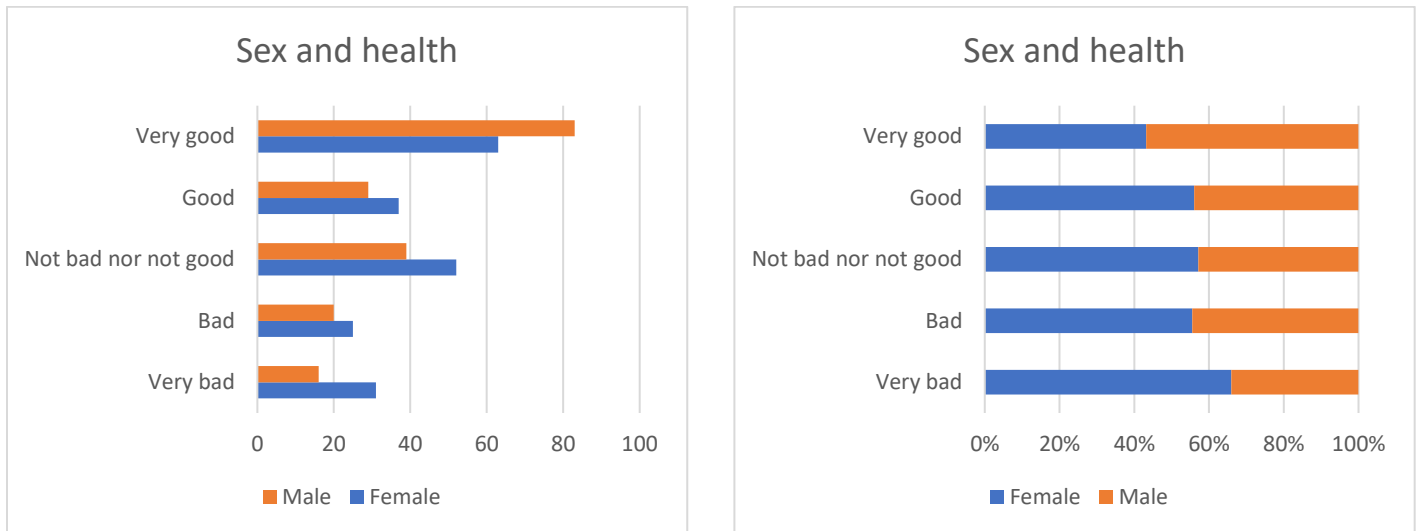
Figure 3.57: Students' current health status and gender.

- Male students' health condition is better than the female.

Mother education and age:

Age	None	Primary education	5th to 9th grade	Secondary education	Higher education
15	0	9	17	18	38
16	1	10	30	28	35
17	1	17	28	25	27
18	0	14	23	18	27
19	1	8	4	8	3
20	0	0	1	1	1
21	0	1	0	0	0
22	0	0	0	1	0

Table 3.21: Mother's education and student's age distribution.
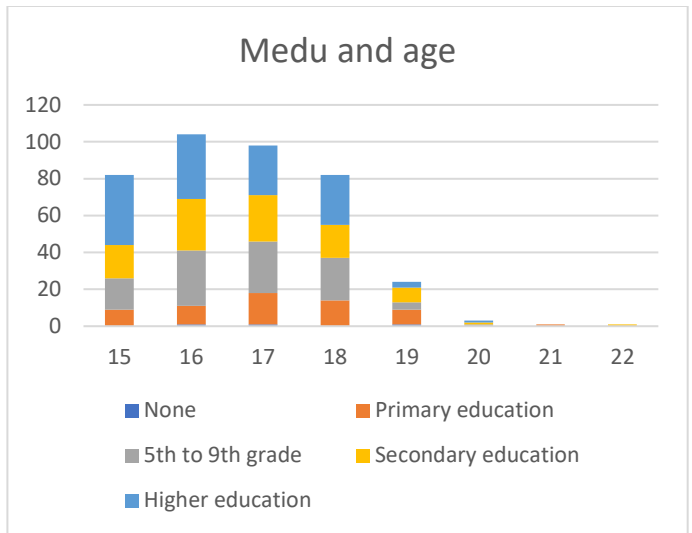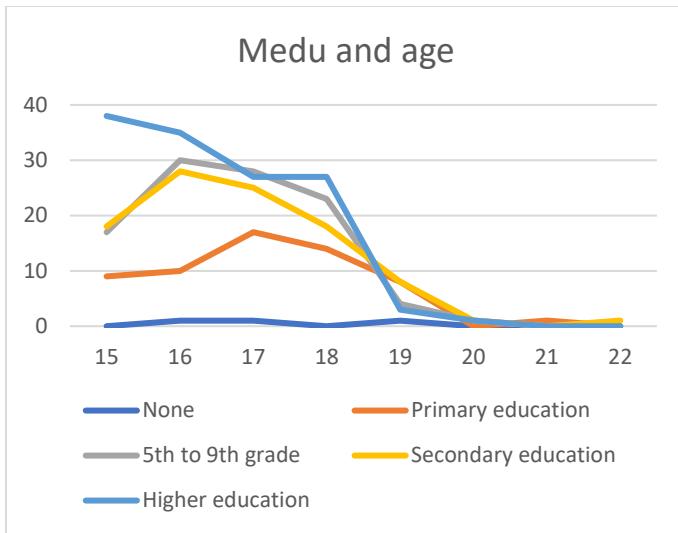
Figure 3.58: Student's age and their mother's education.
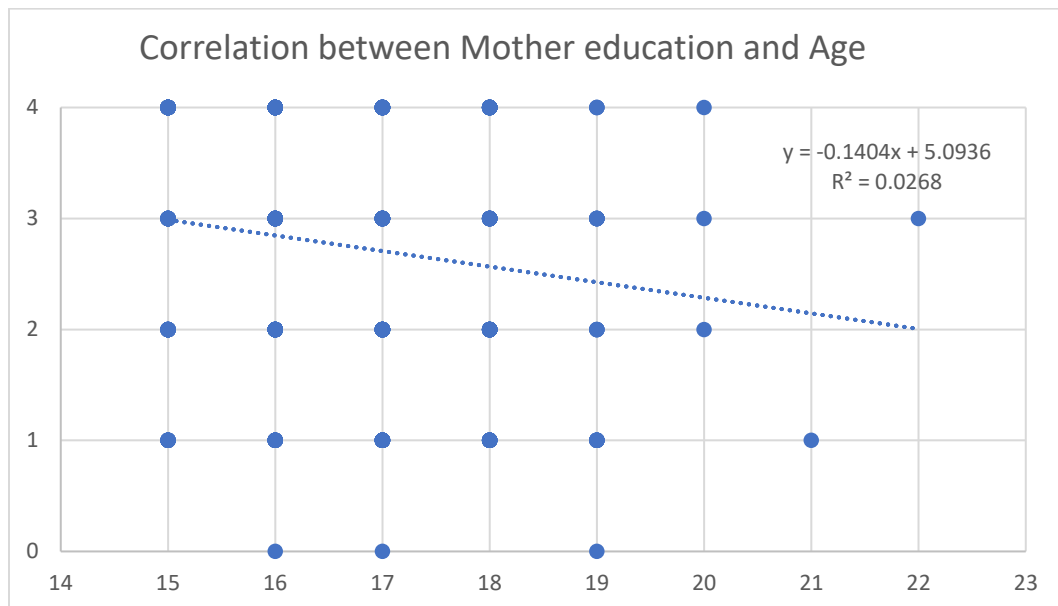
Figure 3.59: Correlation between student's age and their mother's education.

- A negative Correlation between Mother education and Age with value -0.164
- Students whose mother has higher education trends to go to school in an earlier age.

Father education and Age:

Age /Fedu	None	Primary education	5th to 9th grade	Secondary education	Higher education
15	1	10	25	18	28
16	1	17	31	31	24

17	0	22	30	24	22
18	0	21	21	20	20
19	0	10	5	7	2
20	0	0	3	0	0
21	0	1	0	0	0
22	0	1	0	0	0

Table 3.22: Student's age and their father's education.
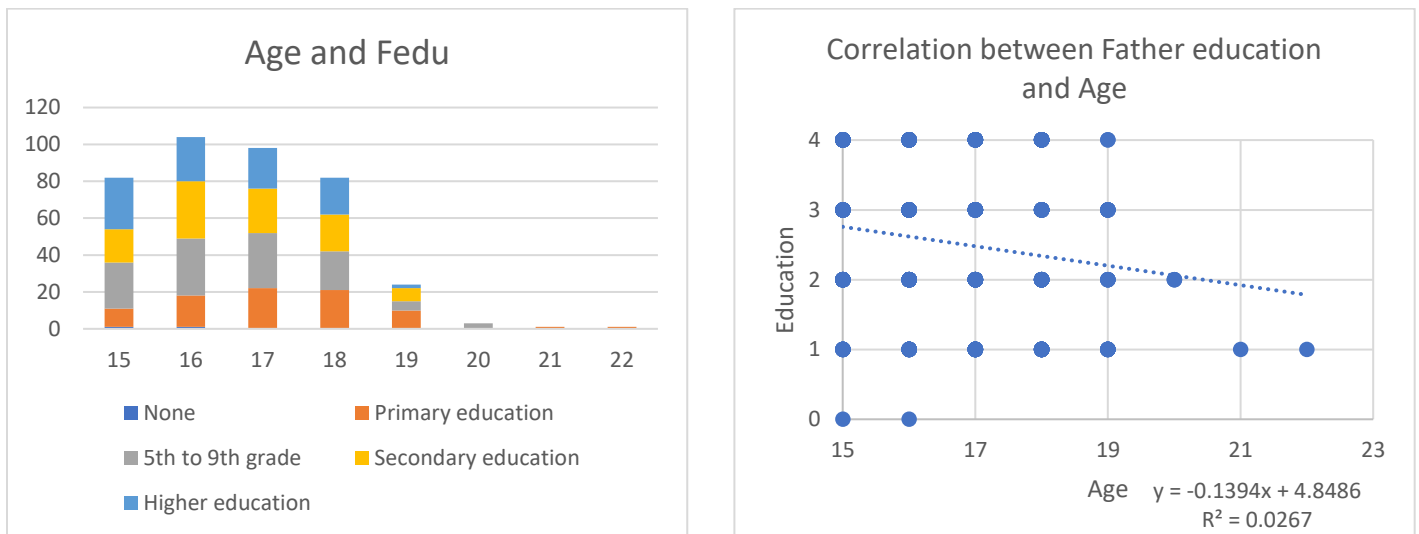
Figure 3.60: Student's age and their father's education.

- A negative Correlation between Mother education and Age with value -0.163
- Students whose father has higher education trends to go to school in an earlier age.

Age and Past failures:

Age/Failures	Zero	One	Two	Three
15	71	3	5	3
16	88	10	4	2
17	84	8	2	4
18	63	16	1	2
19	5	13	3	3
20	1	0	2	0
21	0	0	0	1
22	0	0	0	1

Table 3.23: Age and number of past class failures
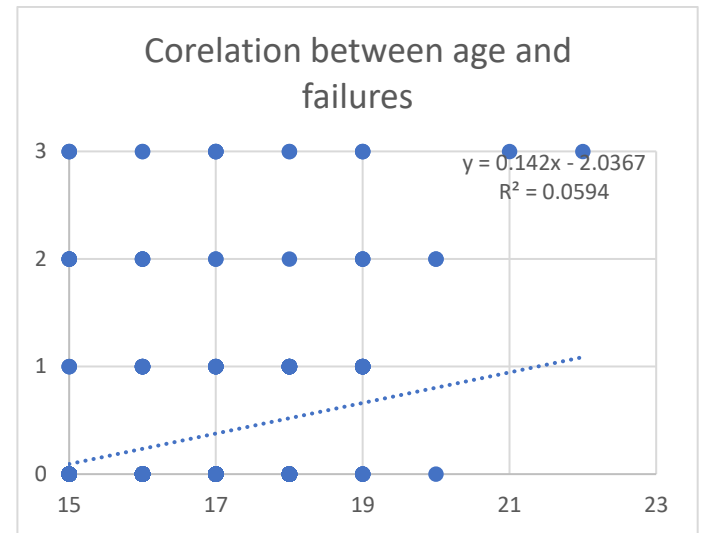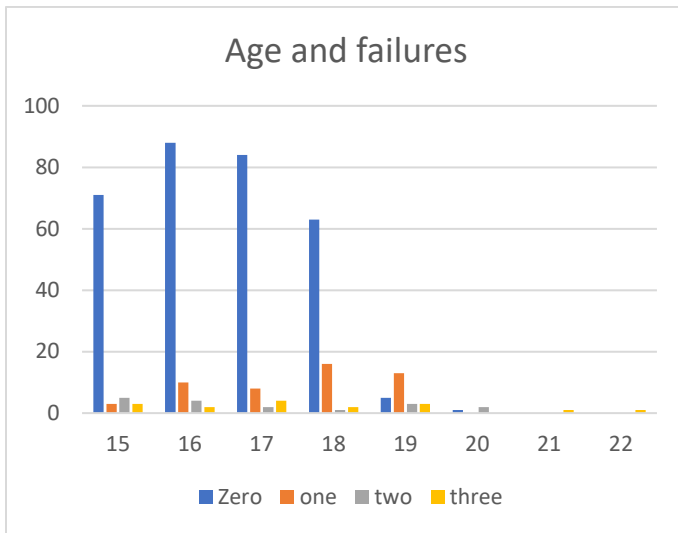
Figure 3.61: Age and number of past class failures

- A positive Correlation between Failure and Age with value 0.24
- Older students are more likely to fail than the younger one.

Age and schoolsup:

Age /schoolsup	yes	no
15	23	59
16	18	86
17	5	93
18	4	78
19	1	23
20	0	3
21	0	1
22	0	1

Table 3.24: Age and extra educational school support distribution.
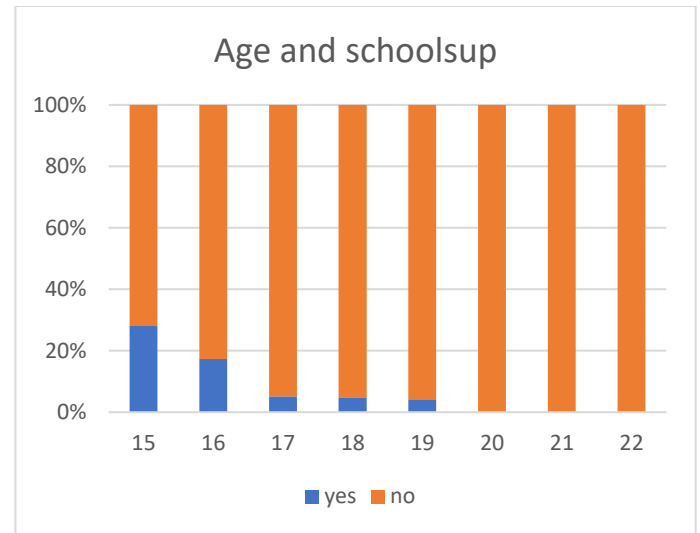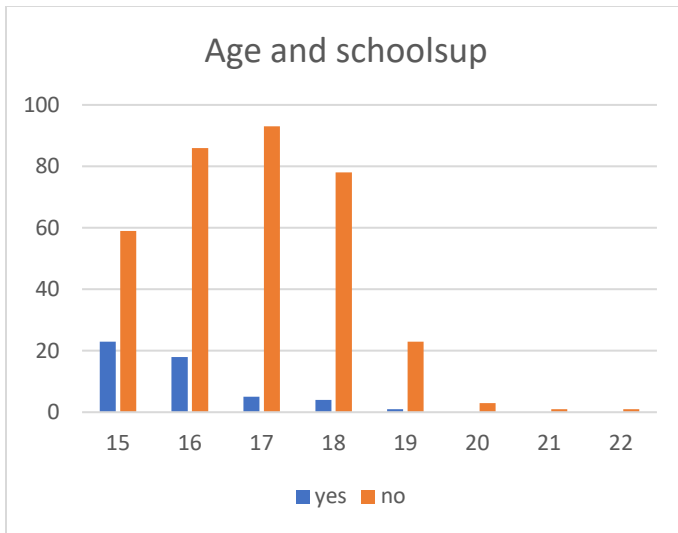
Figure 3.62: Age and extra educational school support distribution.
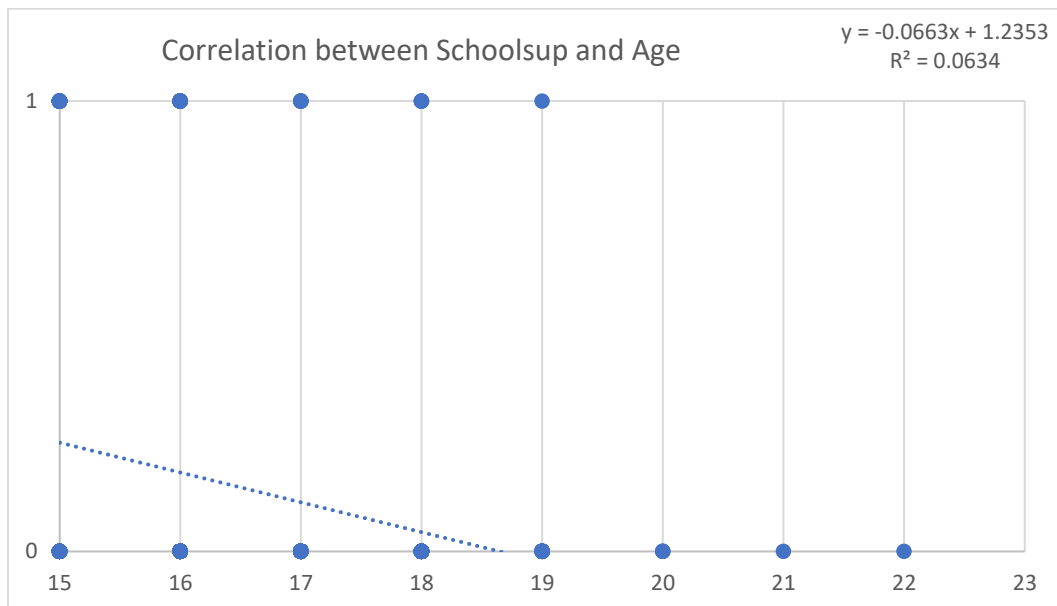
Figure 3.63: Age and extra educational school support distribution.

- A negative Correlation between schoolsup and Age with value -0.25
- Younger students are taking more extra educational school support than the older, among those who are taking it.

Age and higher education:

Age /higher	yes	no
15	82	0
16	101	3

17	94	4
18	73	9
19	21	3
20	3	0
21	1	0
22	0	1

Table 3.25: Students' age and their interest to do higher education.
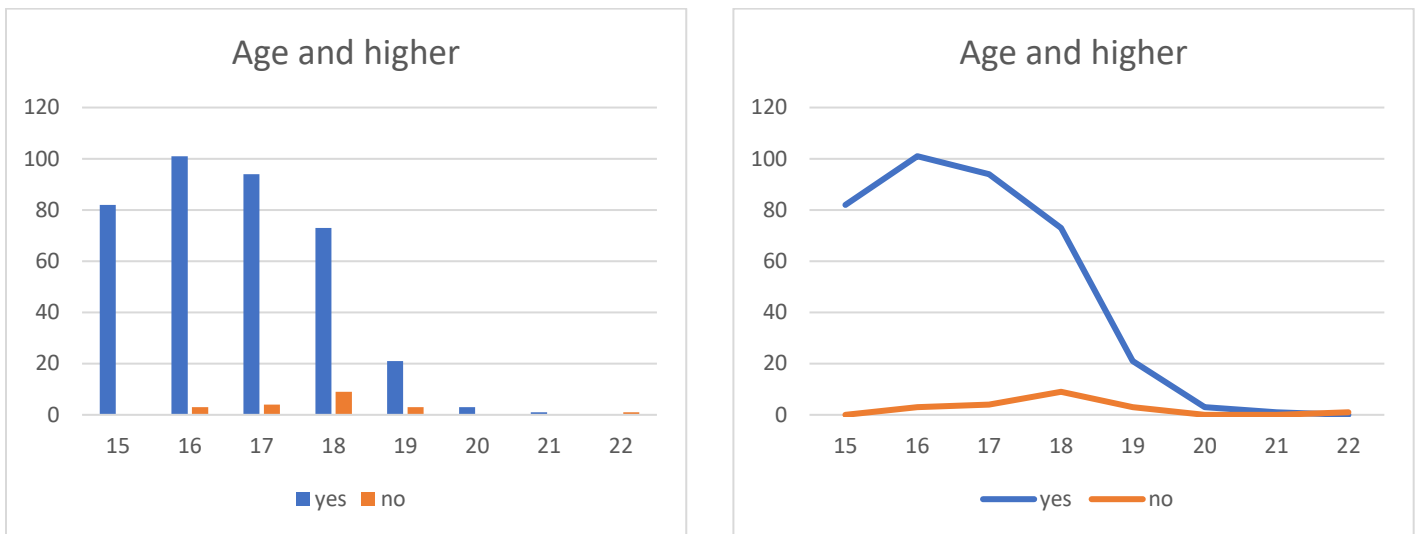
Figure 3.64: Students' age and their interest to do higher education.
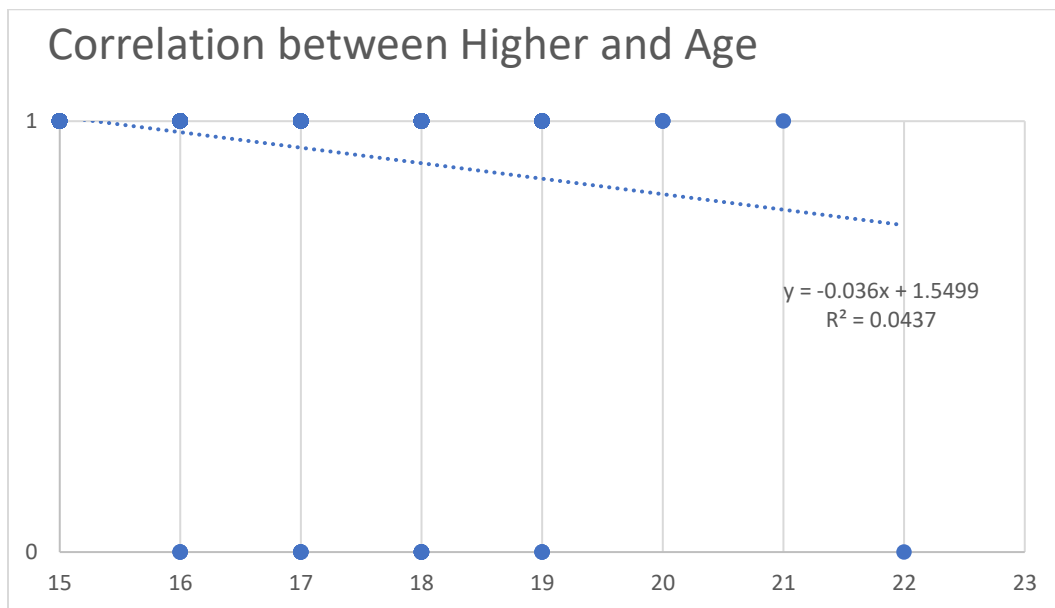
Figure 3.65: Correlation between Students' age and their interest to do higher education.

- A negative Correlation between higher and Age with value -0.209
- More Younger students tends to want higher education than the older.

Age and Romantic relationship:

Age /romantic	yes	no
15	19	63
16	28	76
17	37	61
18	35	47
19	11	13
20	1	2
21	0	1
22	1	0

Table 3.26: Students' age and romantic relationship.
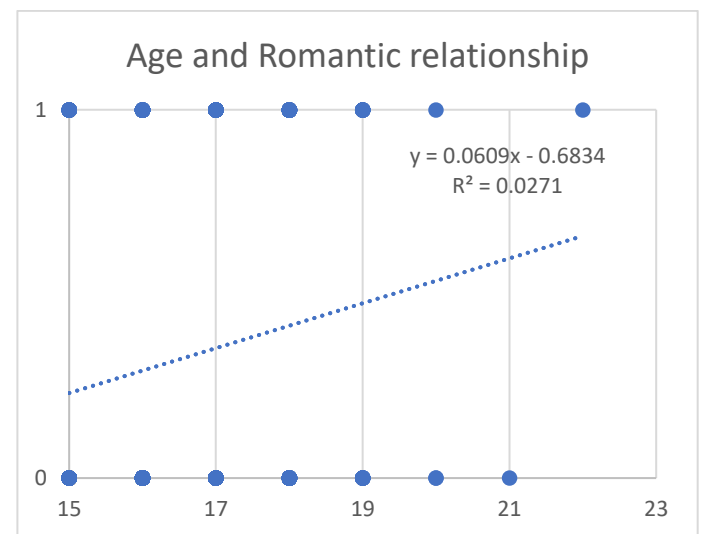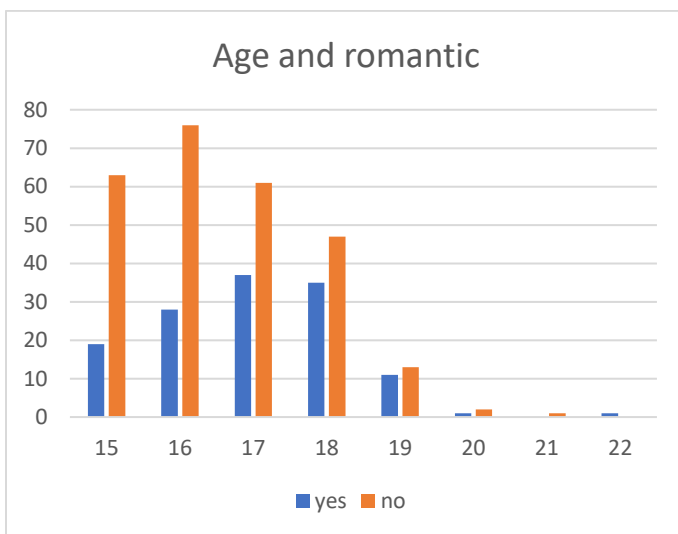
Figure 3.66: students' age and romantic relationship.

- A positive Correlation between Romantic relationship and Age with value 0.165
- Older students are more likely in a romantic relationship. One third of the students are with a romantic relationship.

44

Address and Internet access at home:

Address/Internet	yes	no
Urban	269	38
Rural	60	28

Table 3.27: Address type and internet access at home.
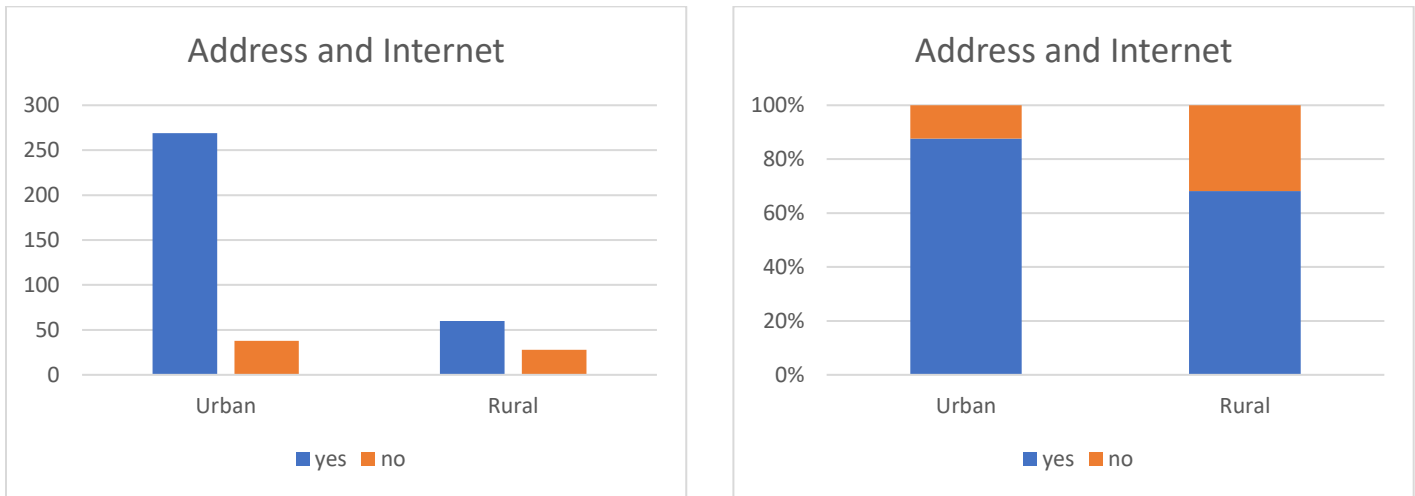
Figure 3.67: Address type and internet access at home.

- Urban students have more internet access at home than the rural student.

Address and travel time:

Traveltime	Urban	Rural
1<15 min	222	35
15 to 30 min	73	34
30 min to hour	9	14
> 1 hour	3	5

Table 3.28: students' address type and their home to school travel time.
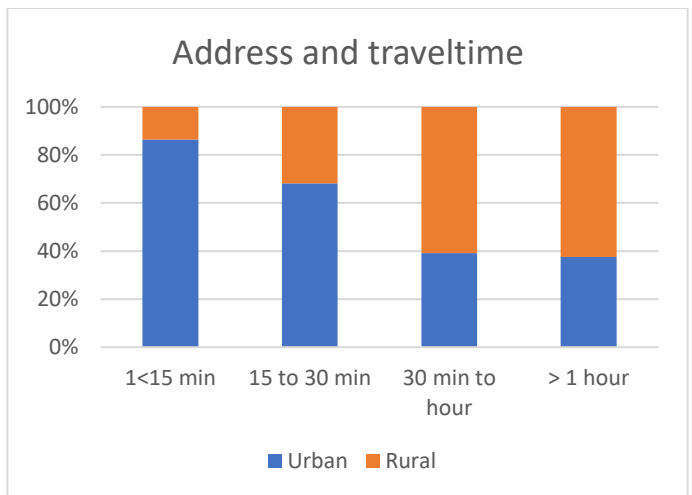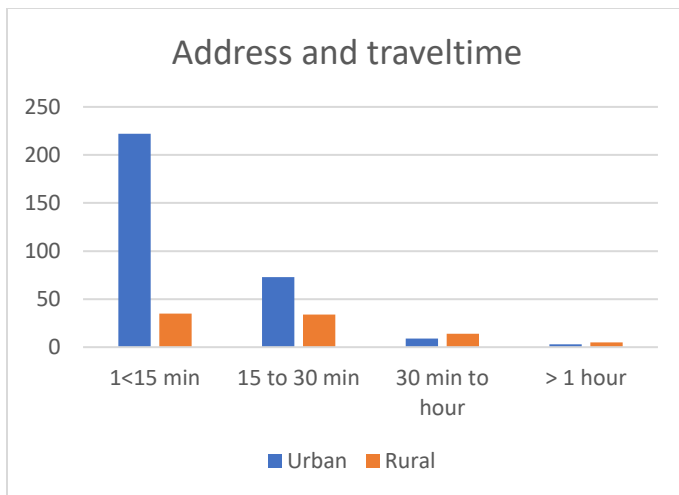
Figure 3.68: students' address type and their home to school travel time.

y = -0.5493x + 1.875
R² = 0.1076
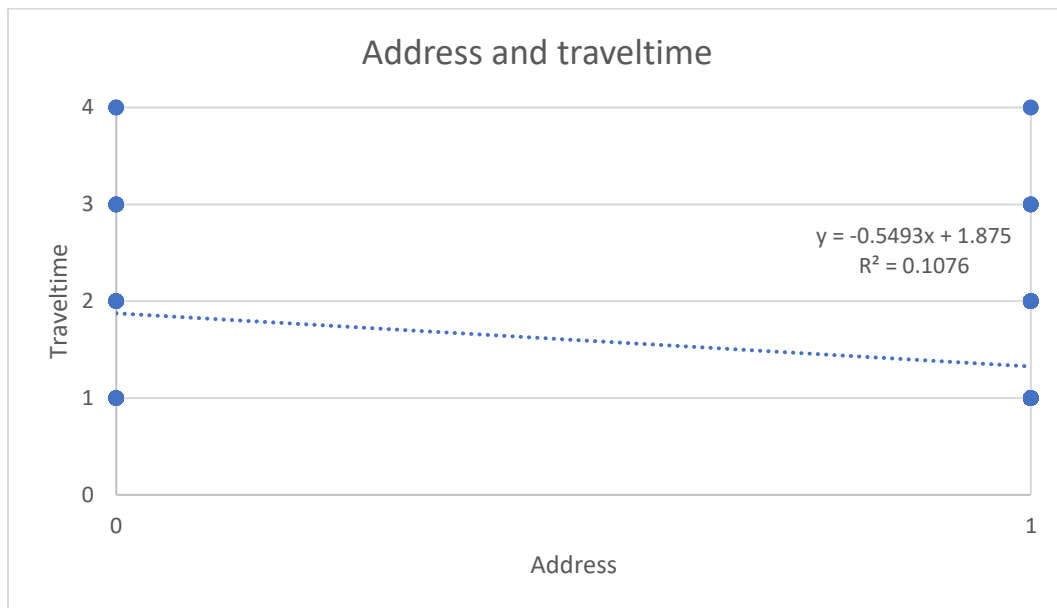
Figure 3.69: Correlation between students' address type and their home to school travel time.

- A negative Correlation between Address and travel time with value -0.33
- Urban student more likely chooses nearby school than the rural student.

Mother education and father education:

	None.Fedu	Primary education.Fedu	5th to 9th grade.Fedu	Secondary education.Fedu	Higher education.Fedu
None.Medu	0	1	2	0	0

Primary education.Medu	1	37	15	5	1
5th to 9th grade.Medu	0	28	51	17	7
Secondary education.Medu	0	15	28	38	18
Higher education.Medu	1	1	19	40	70

Table 3.29: Mother's education and father's education
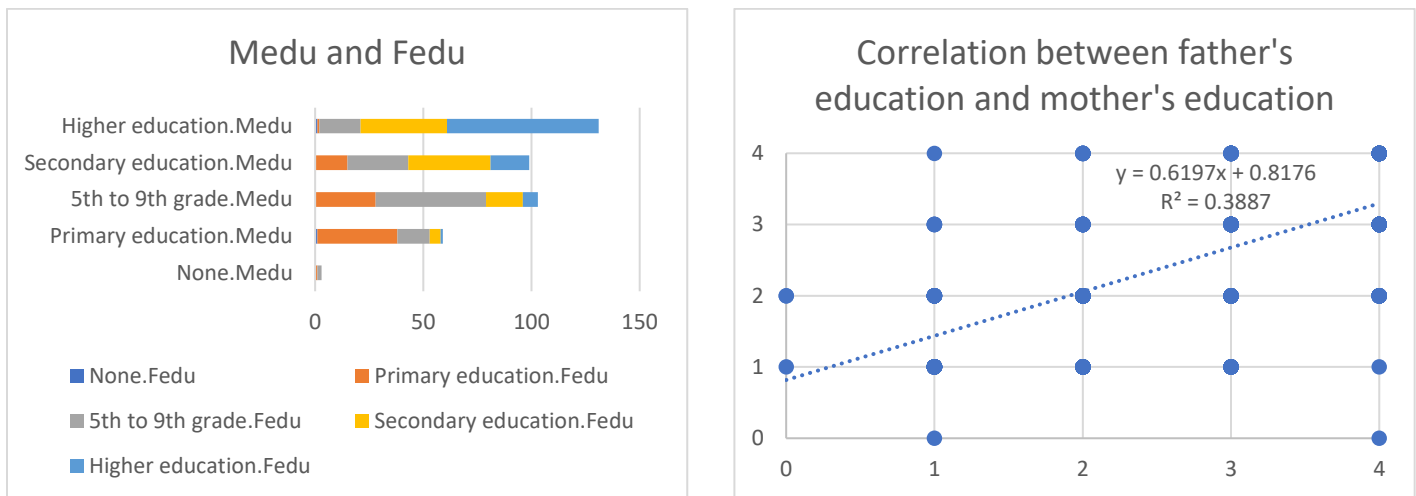
Figure 3.70: Mother's education and father's education distribution.

- A high positive Correlation between Mother education and father education with value 0.623
- Education level plays an important role in people relationships.

Mother education and travel time:

Medu/Traveltime	1<15 min	15 to 30 min	30 min to hour	> 1 hour
None.Medu	2	1	0	0
Primary education.Medu	31	17	8	3
5th to 9th grade.Medu	63	31	6	3
Secondary education.Medu	68	24	6	1
Higher education.Medu	93	34	3	1

Table 3.30: students' mothers' education and students' home to school travel time.
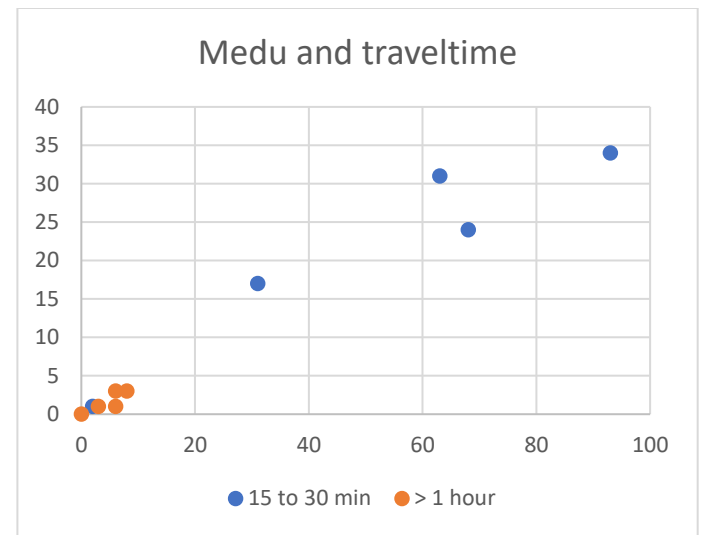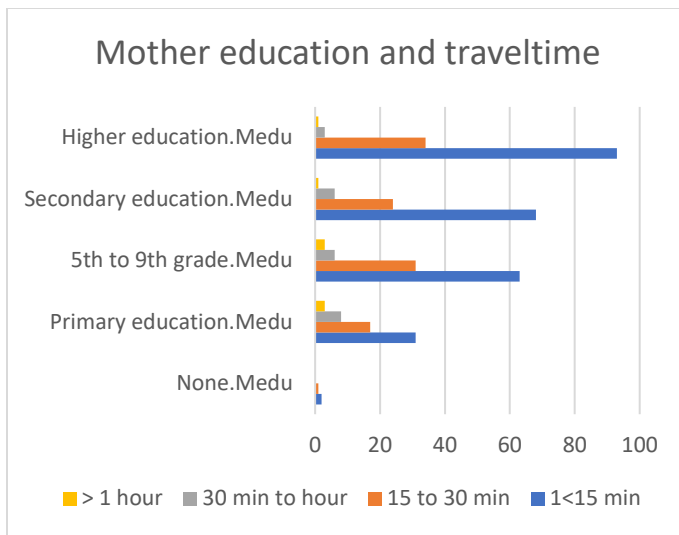
Figure 3.71: students' mothers' education and students' home to school travel time.

- A negative Correlation between Mother education and travel time from home to school with value -0.172
- High educated mother trends to live near the school. Or we can say they choose the school for their children which is near the school.

Mother education and failure:

Medu/failures	0	1	2	3
None.Medu	2	0	0	1
Primary education.Medu	35	15	2	7
5th to 9th grade.Medu	81	14	5	3
Secondary education.Medu	73	14	7	5
Higher education.Medu	121	7	3	0

Table 3.31: students' mothers' education and students' number of past class failures.
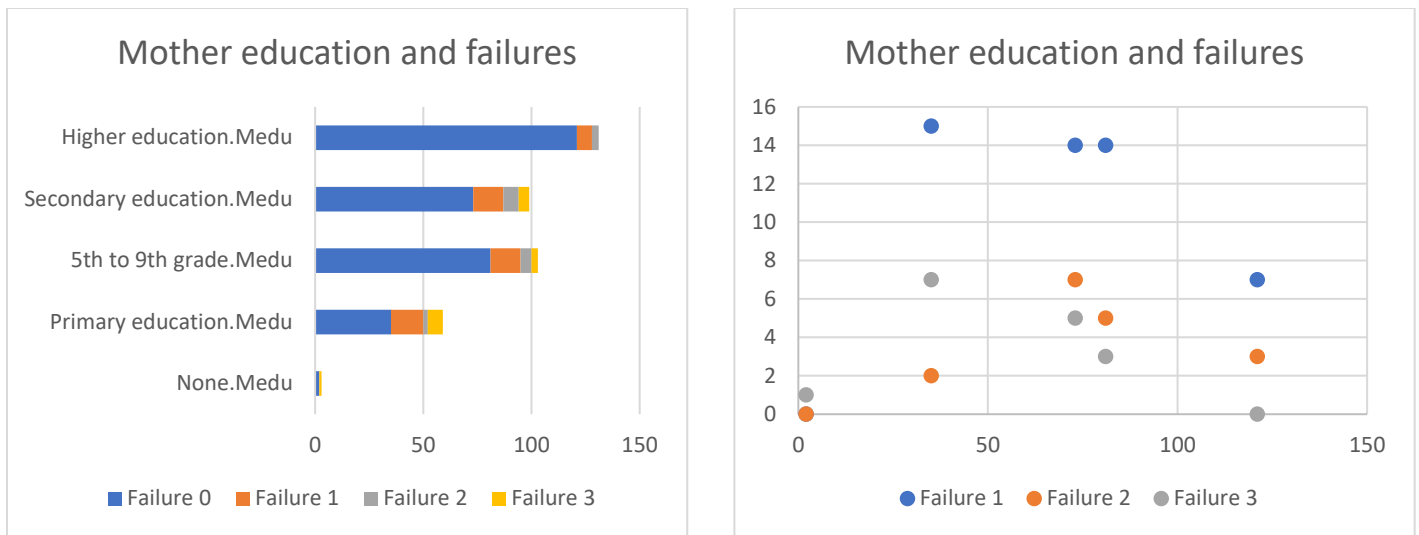
Figure 3.72: students' mothers' education and students' number of past class failures.

- A negative Correlation between Mother education and past class failures with value -0.237
- Mother education plays an important role in children education.

Mother education and family educational support:

Medu/famsup	yes	no
None.Medu	1	2
Primary education.Medu	27	32
5th to 9th grade.Medu	58	45
Secondary education.Medu	63	36
Higher education.Medu	93	38

Table 3.32: Mother education and family educational support
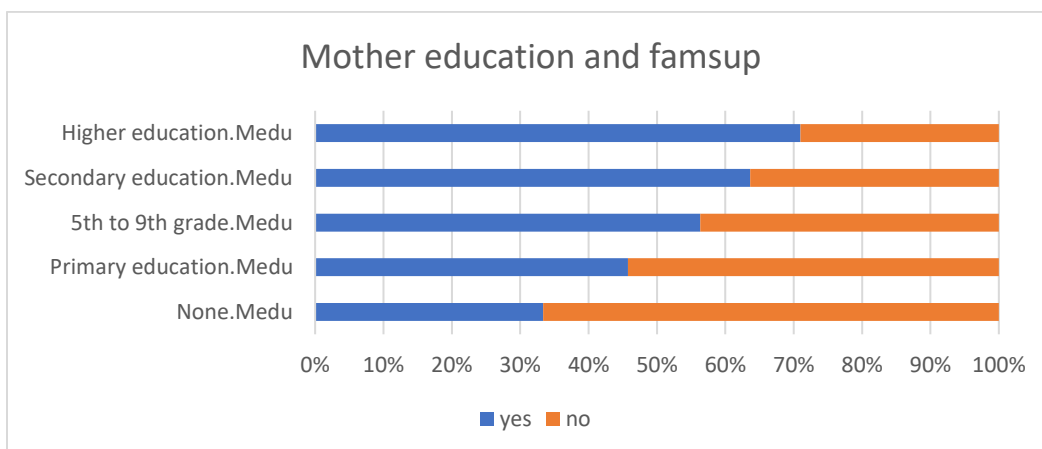
Figure 3.73: Mother education and family educational support

- Mother with more education have more family educational support for their children.

Mother education and nursery:

Medu/nursery	yes	no
None.Medu	1	2
Primary education.Medu	37	22
5th to 9th grade.Medu	84	19
Secondary education.Medu	76	23
Higher education.Medu	116	15

Table 3.33: Mother's education and student attended nursery school
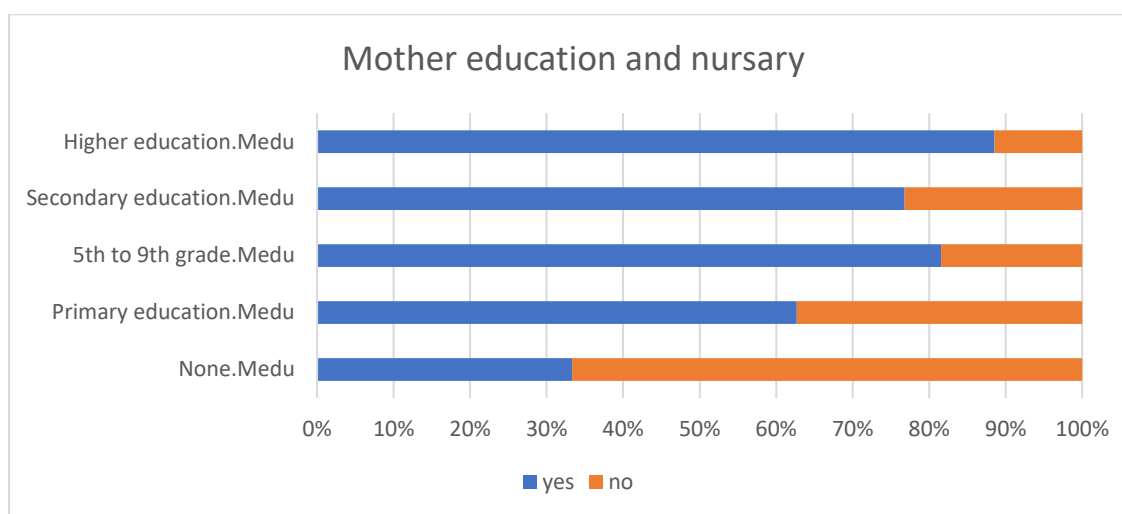
Figure 3.74: Mother's education and student attended nursery school

- Children with more educated mother attends more nursery school.
- A positive correlation between mother education level and children's attended nursery school with value 0.193

Mother education and student wants Higher education:

Medu/Higher	yes	no
None.Medu	2	1
Primary education.Medu	53	6
5th to 9th grade.Medu	96	7
Secondary education.Medu	94	5

50

Higher education.Medu	130	1

Table 3.34: Mother's education and student wants to take higher education.
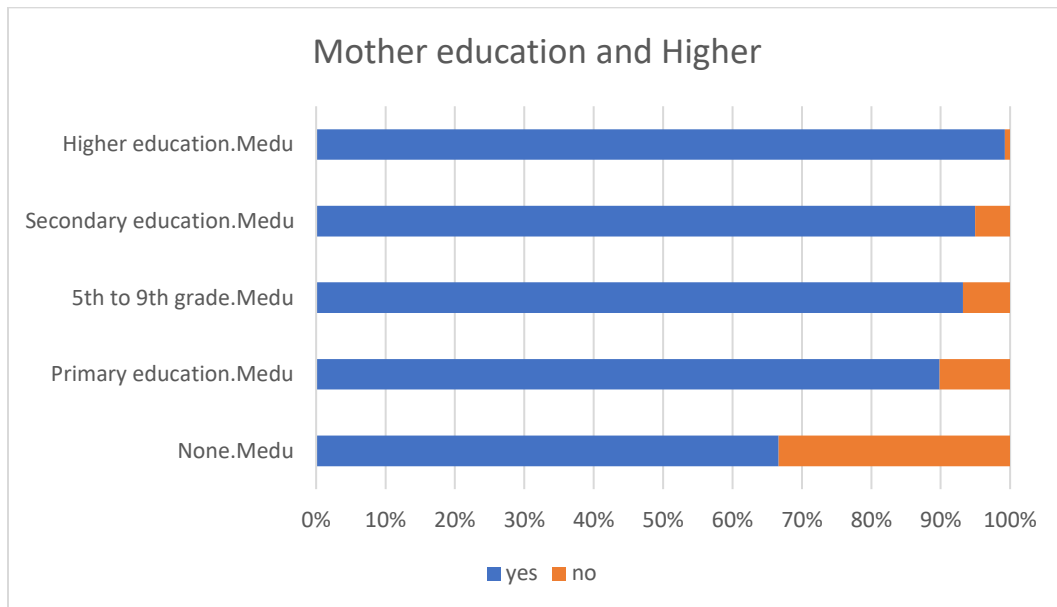
Figure 3.75: Mother's education and student wants to take higher education.

- Children's attitude towards higher education correlates with the level of the mother's educational background.
- The percentage of children's willingness for higher education increases when the mother's education level increases.

Mother education and Internet access at home:

Medu/Internet	yes	no
None.Medu	2	1
Primary education.Medu	39	20
5th to 9th grade.Medu	86	17
Secondary education.Medu	82	17
Higher education.Medu	120	11

Table 3.35: Mother's education and internet access at home
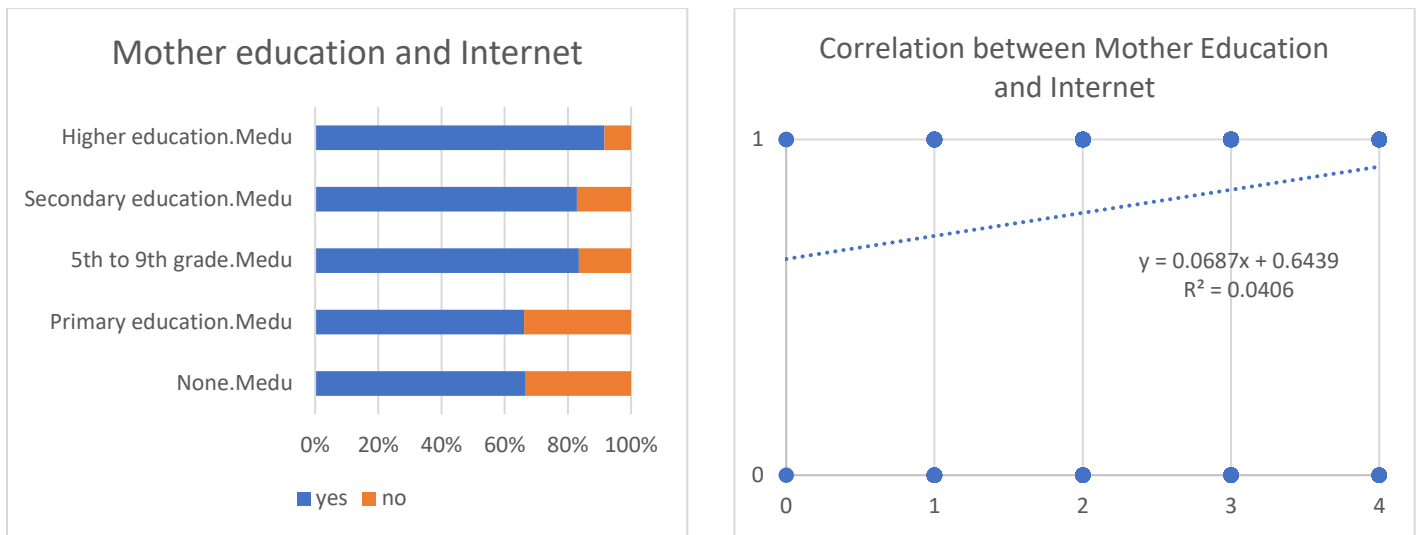
Figure 3.76: Mother's education and internet access at home

- A positive Correlation between Mother Education and Internet access at home with value 0.20
- Children with more educated mother has more internet access at home.

Father education and past failure:

Fedu/failure	Failure 0	Failure 1	Failure 2	Failure 3
None.Fedu	2	0	0	0
Primary education.Fedu	51	16	3	12
5th to 9th grade.Fedu	88	17	7	3
Secondary education.Fedu	85	10	4	1
Higher education.Fedu	86	7	3	0

Table 3.36: Father's education and number of past class failures
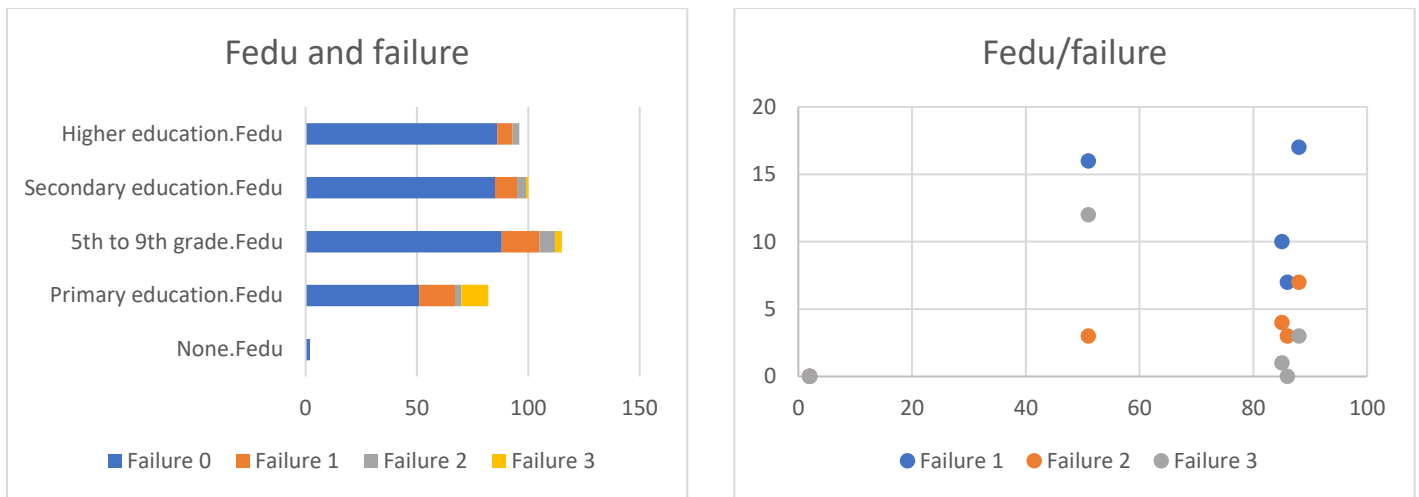
Figure 3.77: Father's education and number of past class failures

- A negative Correlation between Father education and past class failure with value -0.25
- Father education plays an important role in children education.

Father education and family educational support:

Fedu/famsup	yes	no
None.Fedu	1	1
Primary education.Fedu	39	43
5th to 9th grade.Fedu	67	48
Secondary education.Fedu	64	36
Higher education.Fedu	71	25

Table 3.37: Father's education and family educational support
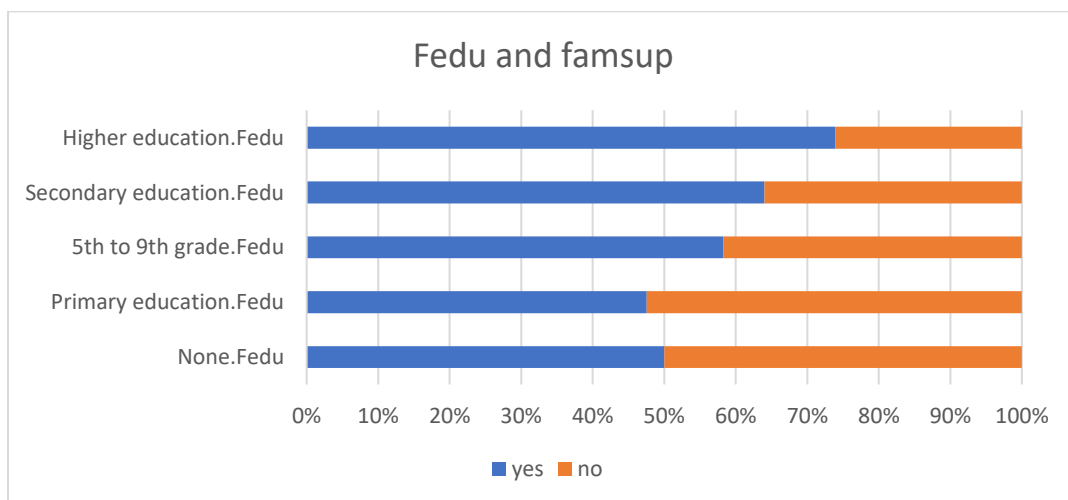
Figure 3.78: Father's education and family educational support

53

- Father with more education have more family educational support for their children.

Father education and student wants higher education:

Fedu/higher	yes	no
None.Fedu	2	0
Primary education.Fedu	72	10
5th to 9th grade.Fedu	108	7
Secondary education.Fedu	98	2
Higher education.Fedu	95	1

Table 3.37: Father's education and student wants to take higher education.
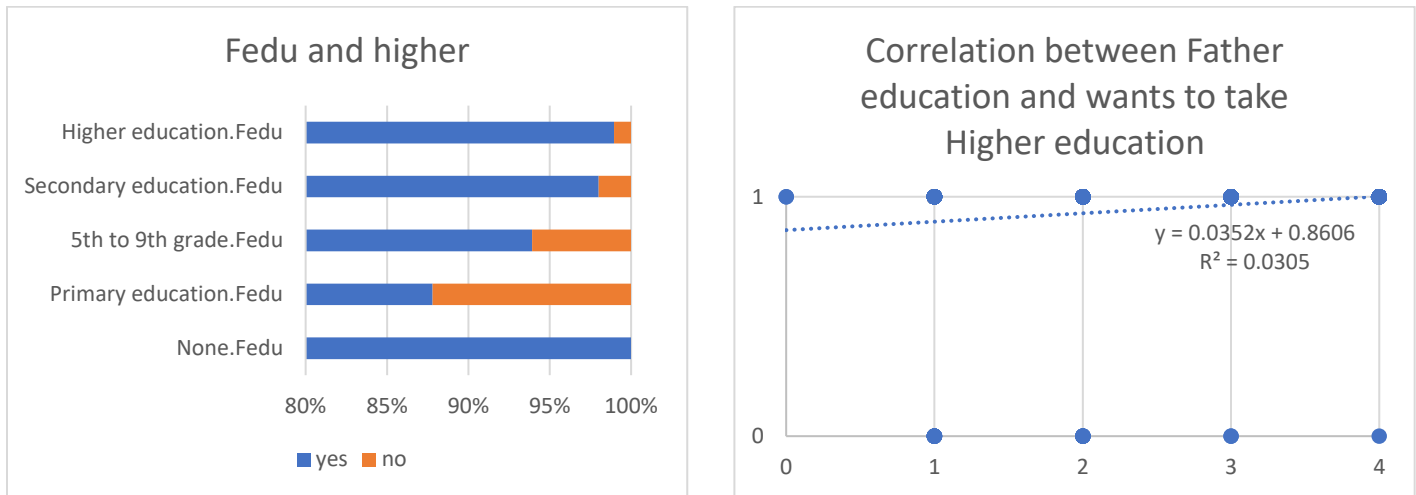
Figure 3.79: Father's education and student wants to take higher education.

- A positive Correlation between Father education and wants to take Higher education with value 0.175

Study time and past failures:

studytime/failure	Failure 0	Failure 1	Failure 2	Failure 3
< 2 hours	74	16	6	9
2 to 5 hours	158	26	7	7
5 to 10 hours	54	7	4	0
4 > 10 hours	26	1	0	0

Table 3.38: student's weekly study time and their number of past class failures
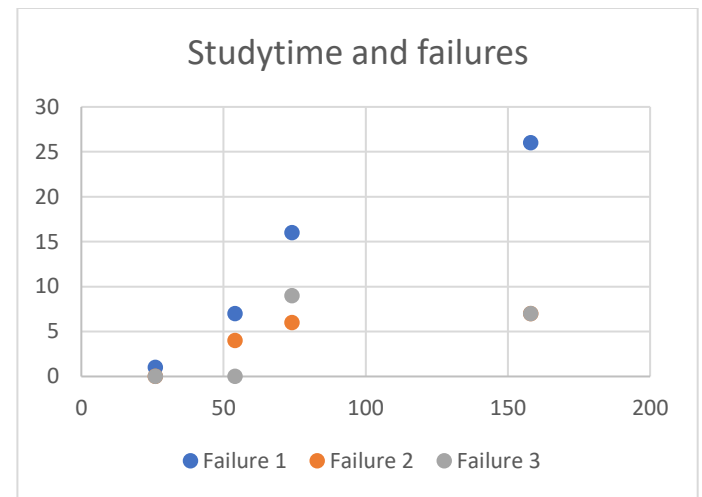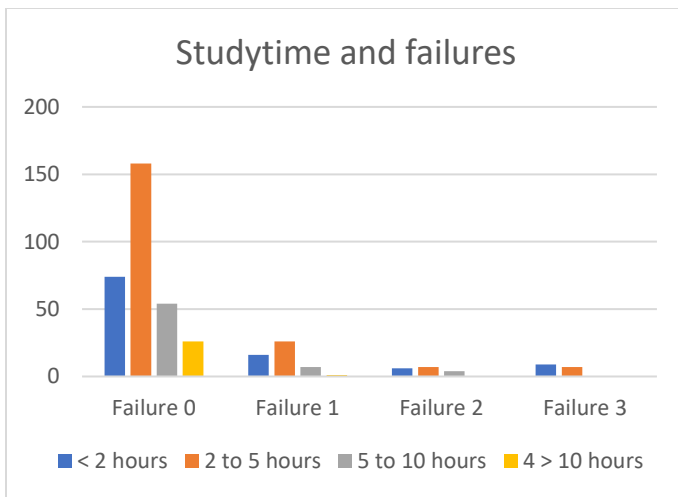
Figure 3.80: student's weekly study time and their number of past class failures
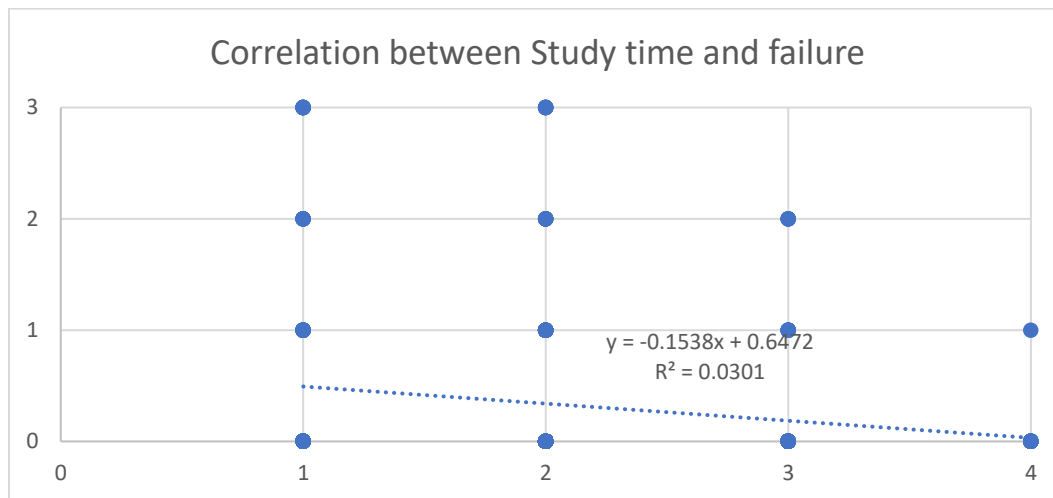
Figure 3.81: correlation between student's weekly study time and their number of past class failures

- A negative correlation between the study time and the failure with -0.174
- Study more and less likely to fail.

Study time and extra paid classes:

studytime/Paid	yes	no
< 2 hours	32	73
2 to 5 hours	99	99
5 to 10 hours	34	31
4 > 10 hours	16	11

Table 3.39: student's weekly study time and extra paid classes
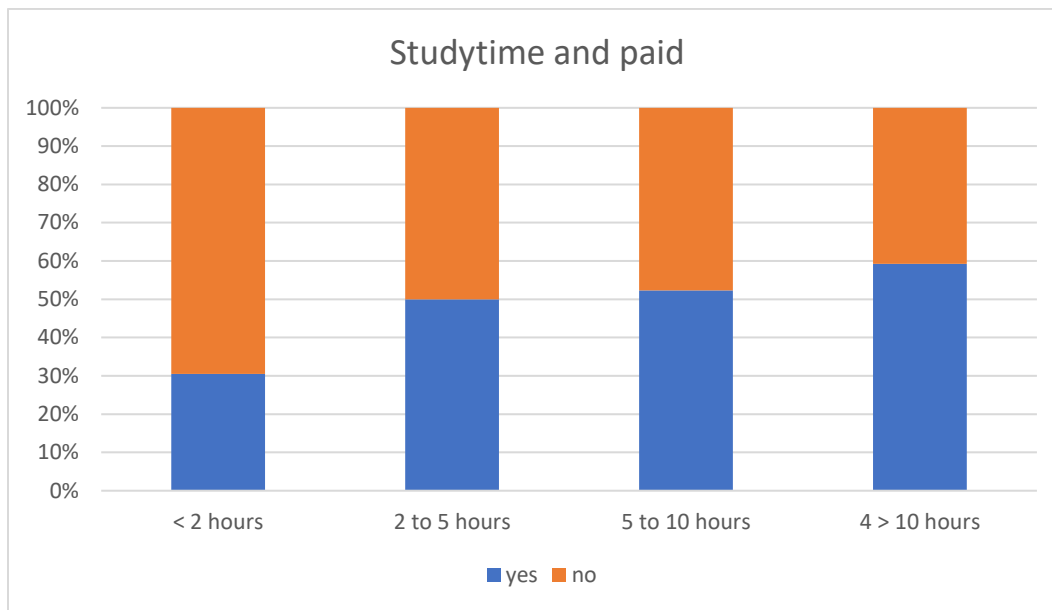
Figure 3.82: student's weekly study time and extra paid classes

- A positive correlation between study time and extra paid classes with value 0.167

Study time and students wants higher education:

studytime/higher	yes	no
< 2 hours	93	12
2 to 5 hours	190	8
5 to 10 hours	65	0
4 > 10 hours	27	0

Table 3.40: student's weekly study time and wants higher education.
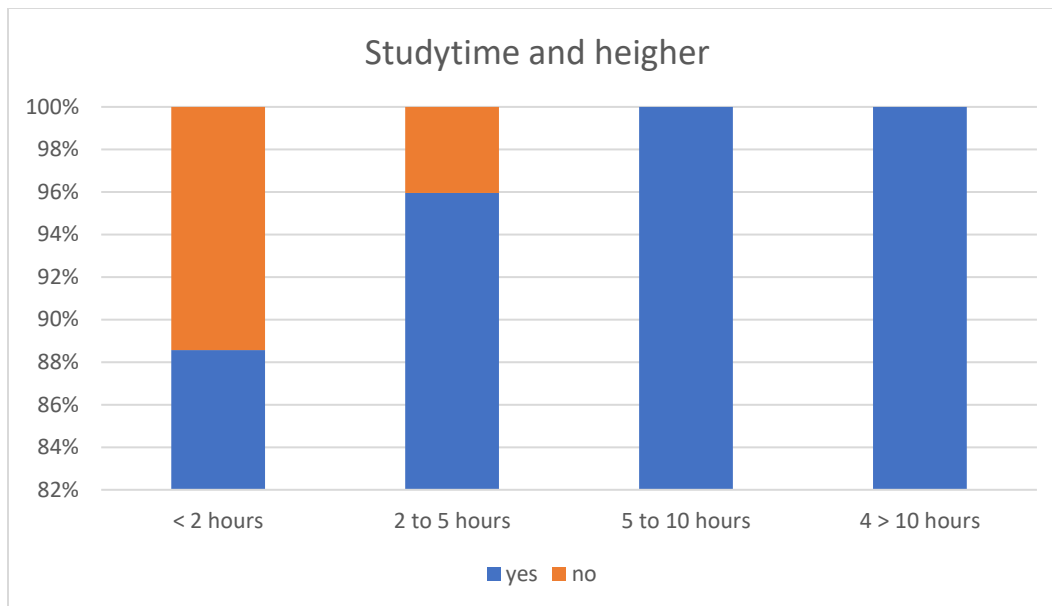
Figure 3.83: student's weekly study time and wants higher education.

- Students attitude towards higher education correlates with the amount of their study time.
- Those students who study more hours want to do more higher education.

Study time and working day alcohol consumption:

studytime/Dalc	Very low	Low	Medium	High	Very high
< 2 hours	59	24	15	2	5
2 to 5 hours	141	38	9	7	3
5 to 10 hours	53	11	1	0	0
4 > 10 hours	23	2	1	0	1

Table 3.41: student's weekly study time and working day alcohol consumption.
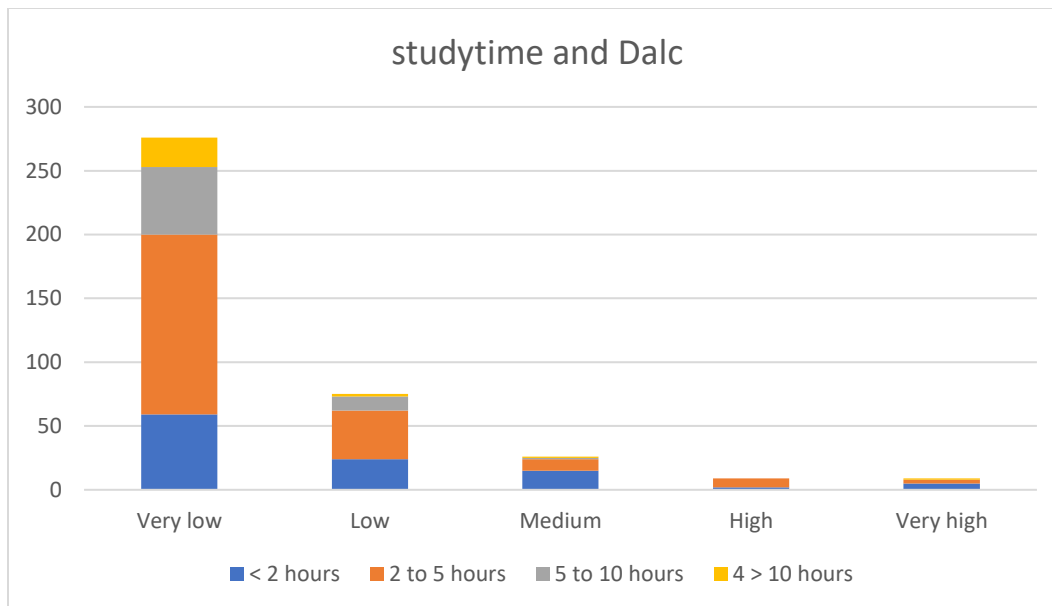
Figure 3.84: student's weekly study time and working day alcohol consumption.

- The student who studies more hours tends to consume lesser alcohol in the working days.
- A negative correlation between the study time and working day students alcohol consumption with value -0.196

.

Study time and weekend alcohol consumption:

studytime/Walc	Very low	Low	Medium	High	Very high
< 2 hours	24	25	21	22	13
2 to 5 hours	77	41	42	26	12
5 to 10 hours	33	15	14	2	1
4 > 10 hours	17	4	3	1	2

Table 3.42: student's weekly study time and weekend alcohol consumption.
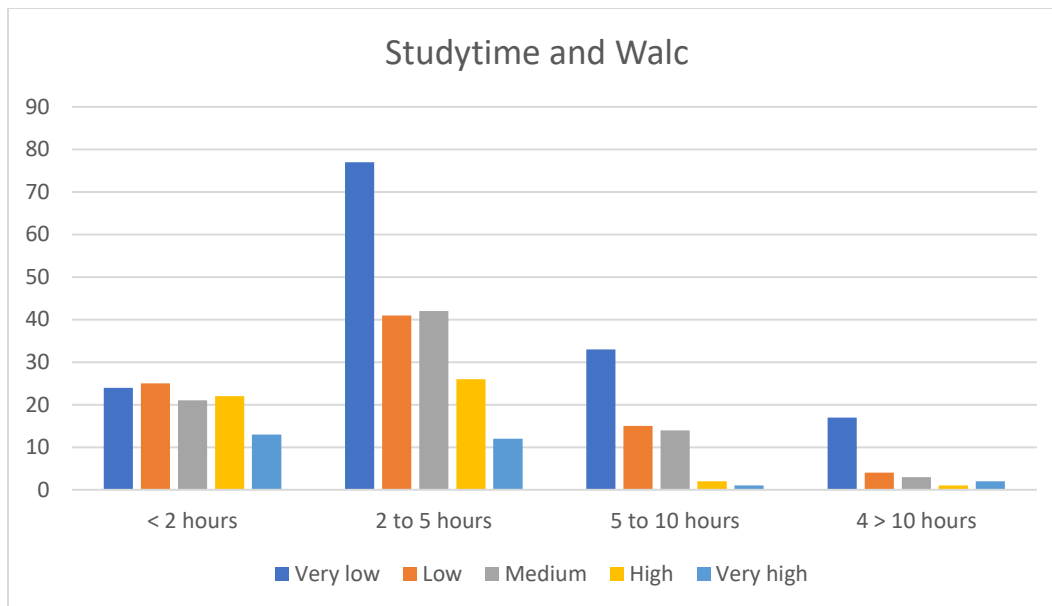
Figure 3.85: student's weekly study time and weekend alcohol consumption.

- The student who studies more hours tends to consume lesser alcohol in the weekend.
- A negative correlation between the students' study time and weekend alcohol consumption with value -0.254

Failures and paid class:
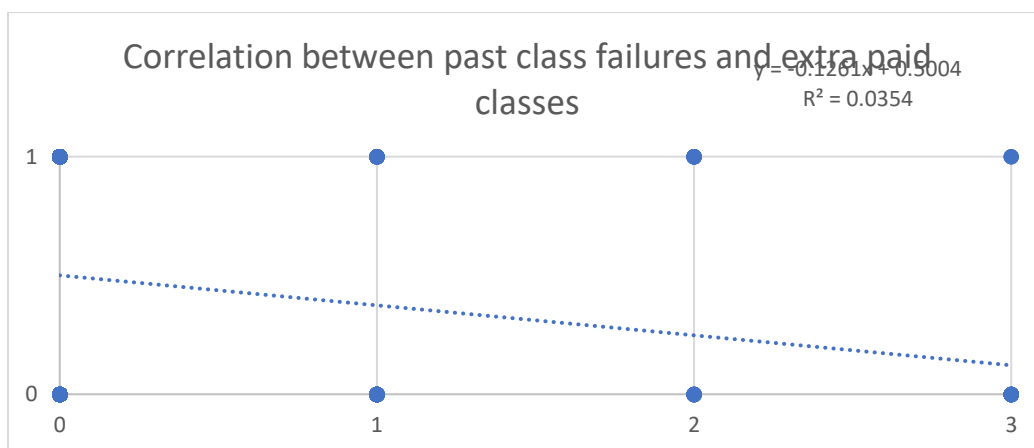
Figure 3.86: Correlation between number of past class failures and extra paid classes

- A negative Correlation between past class failure and extra paid classes with value -0.188

Past failures and students want to take higher education:

Failure/higher	Failure 0	Failure 1	Failure 2	Failure 3
yes	305	44	16	10
no	7	6	1	6

Table 3.43: Number of past failures and student wants to take higher education.
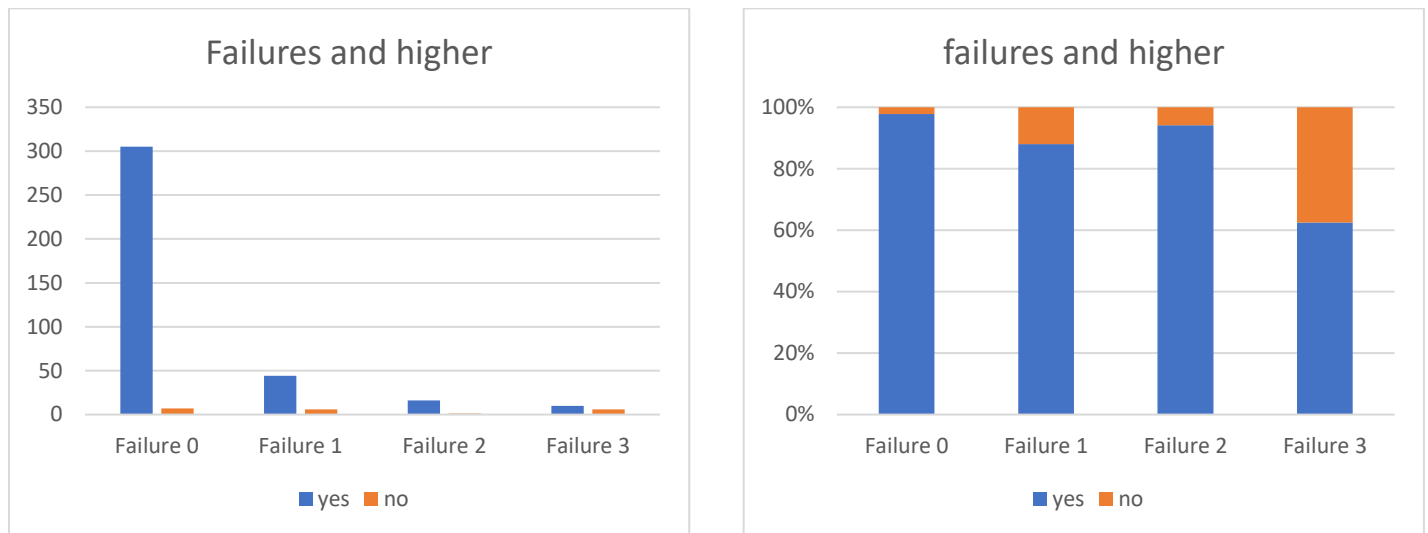
Figure 3.87: Number of past failures and student wants to take higher education.
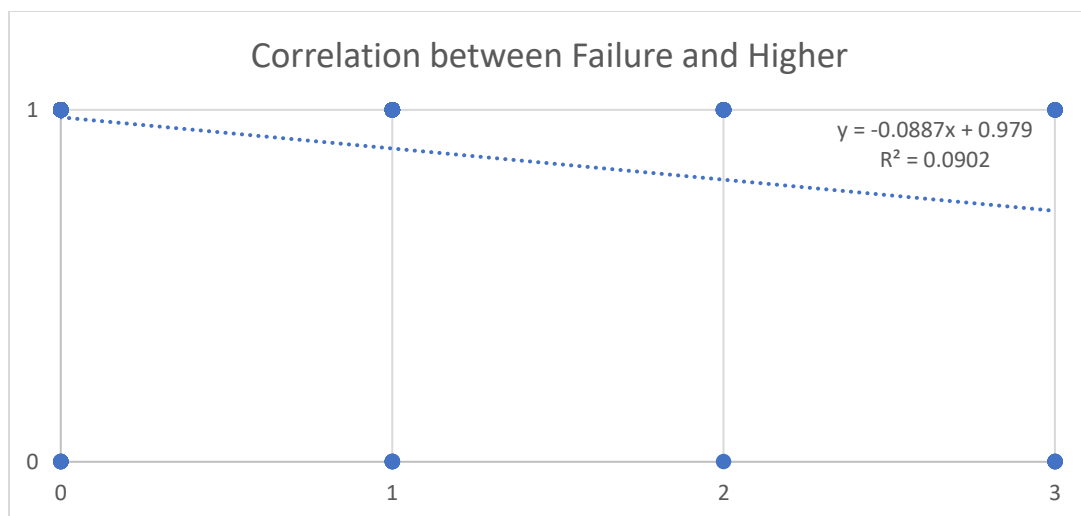
Figure 3.88: Correlation between number of past failures and student wants to take higher education.

- A negative Correlation between past class failure and wants to take higher education with value -0.30
- Students will fail less if they have a tendency to take higher education.

Paid class and student wants Higher education:

paid/higher	yes.higher	no.higher
yes.paid	180	1
no.paid	195	19

Table 3.44: Extra paid classes and student wants to take higher education.
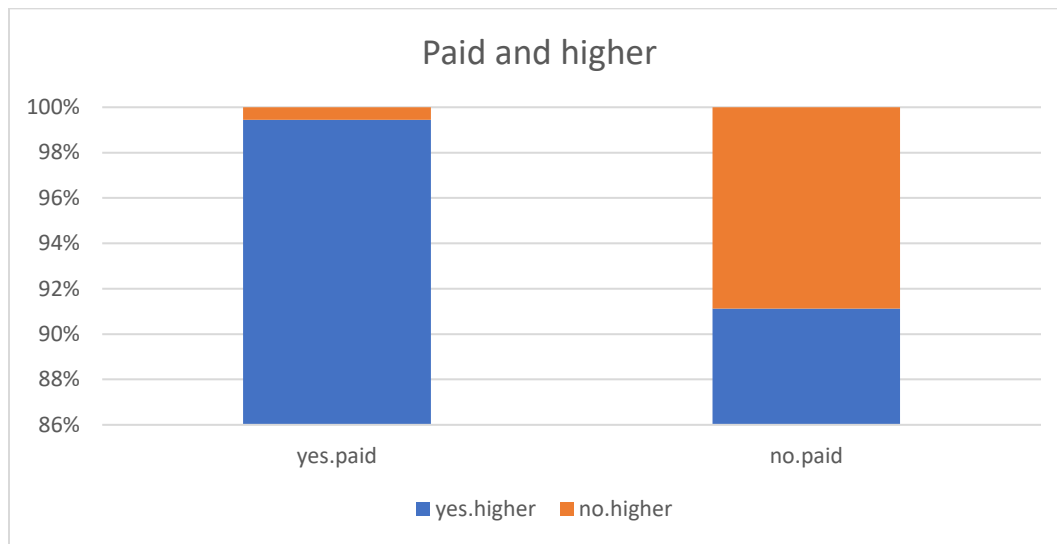
Figure 3.89: Extra paid classes and student wants to take higher education.

- A positive Correlation between extra paid classes and wants to take higher education with value 0.189

Free time after school and going out with friends:

freetime/goout	Very low	Low	Medium	High	Very high
Very low	4	8	5	1	1
Low	5	27	17	11	4
Medium	5	41	70	26	15
High	5	18	34	41	17
Very high	4	9	4	7	16

Table 3.45: Free time after school and going out with friends.
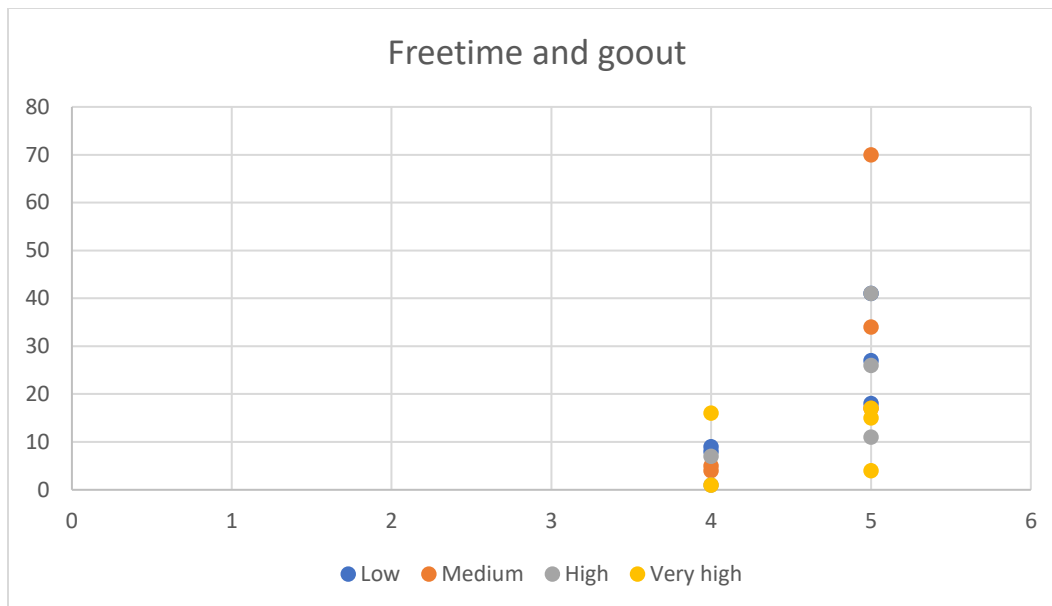
Figure 3.90: Free time after school and going out with friends

- A positive correlation between Free time after school and going out with friends with value 0.285019

Free time after school and workday alcohol consumption:

freetime/Dacl	Very low.Acl	Low.Acl	Medium.Acl	High.Acl	Very high.Acl
Very low.time	17	1	1	0	0
Low.time	51	9	3	1	0
Medium.time	113	33	6	3	2
High.time	73	23	12	3	4
Very high.time	22	9	4	2	3

Table 3.46: Free time after school and workday alcohol consumption.
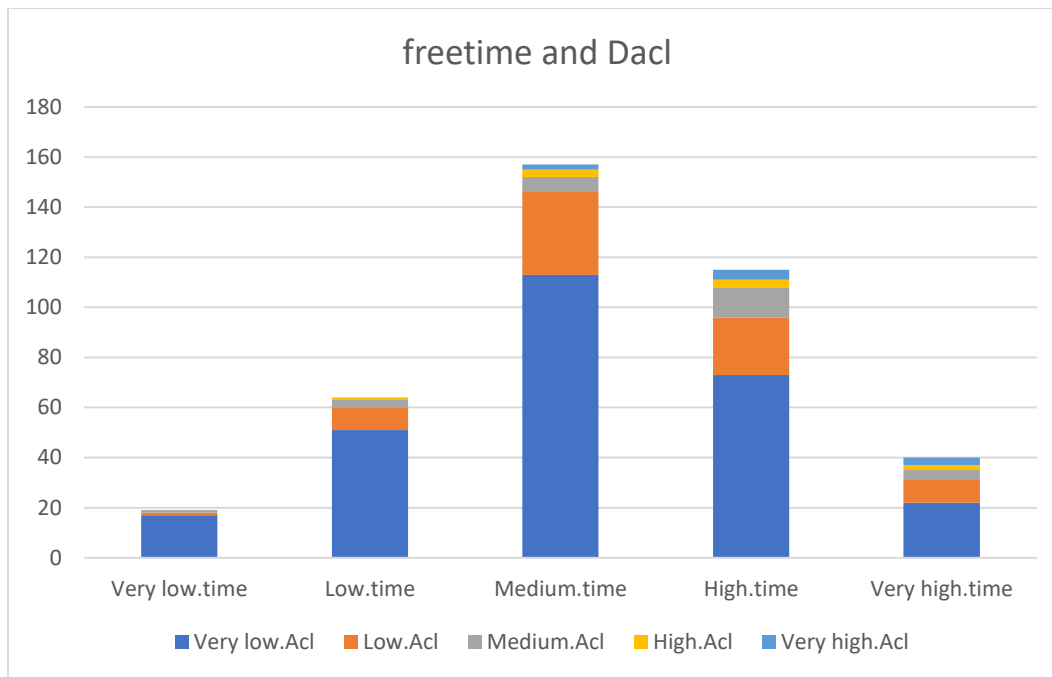
Figure 3.91: Free time after school and workday alcohol consumption.

▪ A positive correlation between Free time after school and workday alcohol consumption with value 0.209

Going out and workday alcohol consumption:

GoOut/Dacl	Very low.Acl	Low.Acl	Medium.Acl	High.Acl	Very high.Acl
Very low.goout	20	2	1	0	0
Low.goout	81	16	3	2	1
Medium.goout	99	22	6	1	2
High.goout	50	23	9	4	0
Very high.goout	26	12	7	2	6

Table 3.47: Going out with friends and workday alcohol consumption.
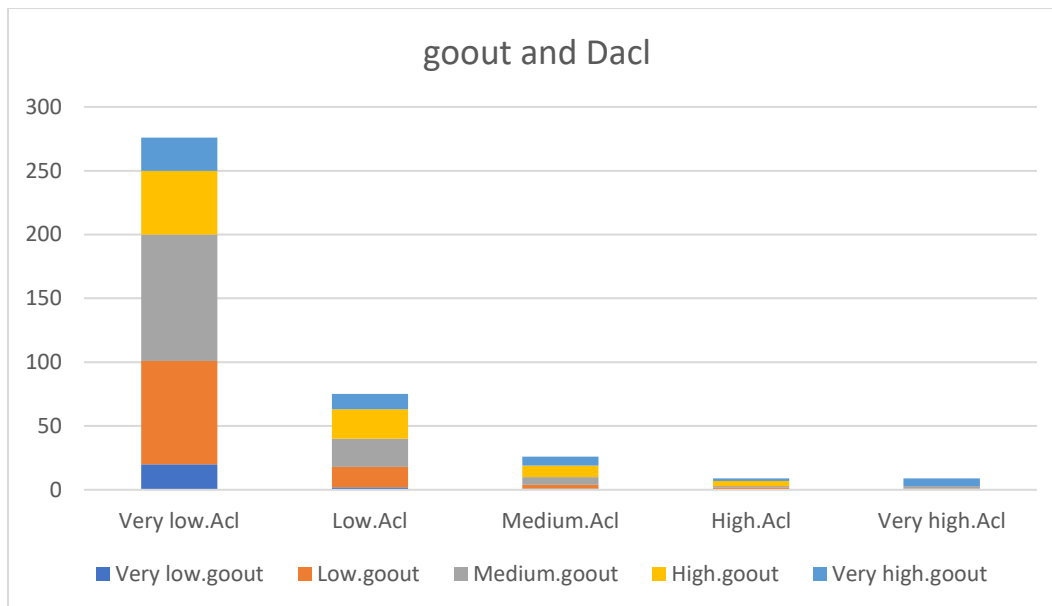
Figure 3.92: Going out with friends and workday alcohol consumption.

- A positive correlation between going out with friends and workday alcohol consumption. with value 0.267

Going out and weekend alcohol consumption:

GoOut/Wacl	Very low.Acl	Low.Acl	Medium.Acl	High.Acl	Very high.Acl
Very low.goout	15	4	2	2	0
Low.goout	54	29	12	6	2
Medium.goout	47	37	35	7	4
High.goout	26	9	22	23	6
Very high.goout	9	6	9	13	16

Table 3.48: Going out with friends and weekend alcohol consumption.
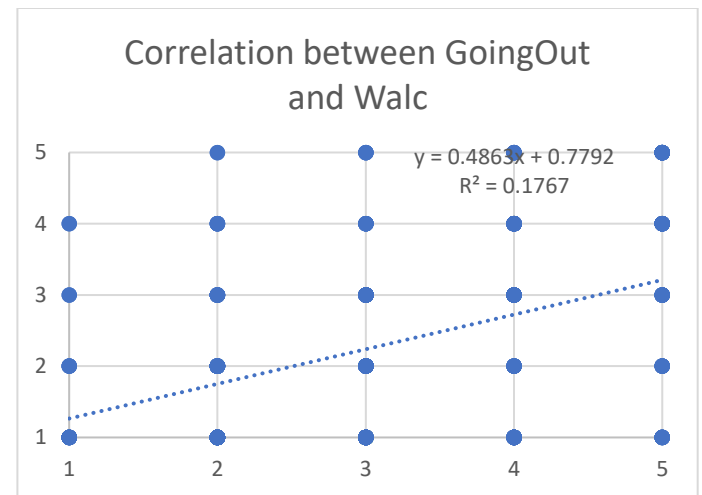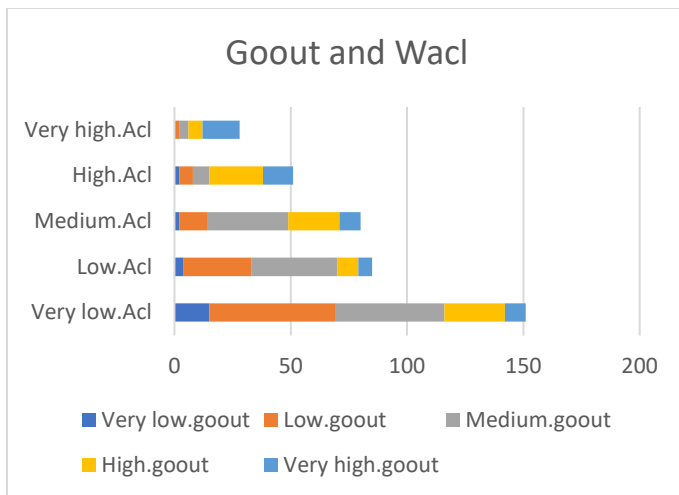
Figure 3.93: Going out with friends and weekend alcohol consumption.

- A high positive correlation between going out with friends and weekend alcohol consumption. with value 0.42
- The students who going out more with friends tends to consume more alcohol.

Workday alcohol consumption and Weekend alcohol consumption:

Dacl/Wacl	Very low.Wacl	Low.Wacl	Medium.Wacl	High.Wacl	Very high.Wacl
Very low.Dacl	150	65	42	15	4
Low.Dacl	1	18	29	22	5
Medium.Dacl	0	1	8	11	6
High.Dacl	0	1	1	3	4
Very high.Dacl	0	0	0	0	9

Table 3.49: Workday alcohol consumption and Weekend alcohol consumption.
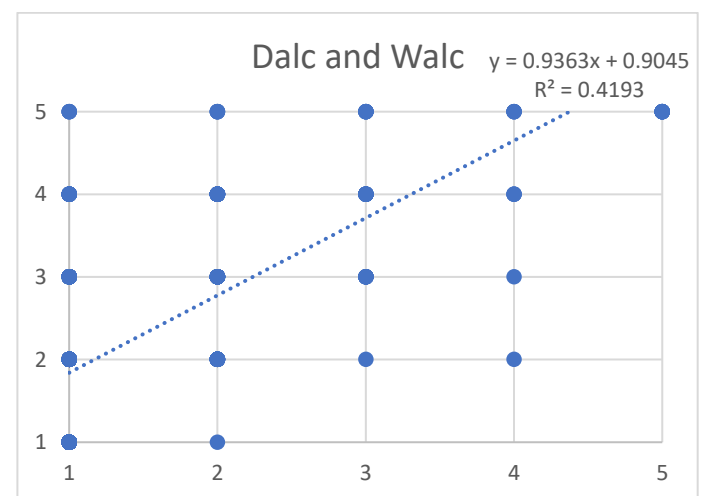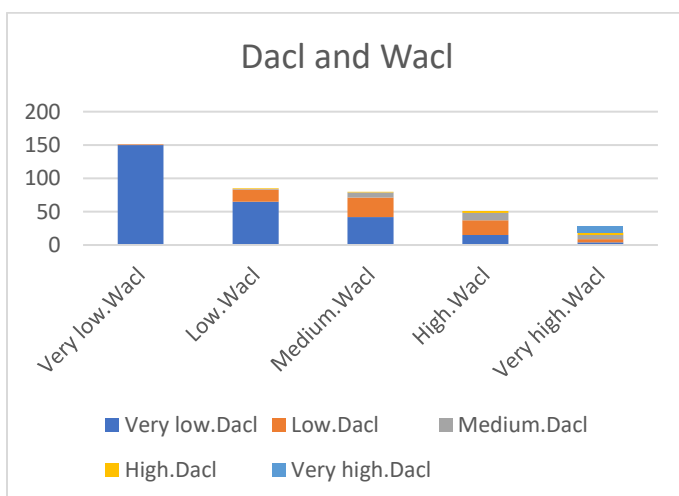
- A very high positive correlation between workdays alcohol and weekend alcohol consumption with value 0.647
- Those who consume more alcohol in the working days tends to consume more alcohol in the weekends.

Mother's education and final grade:

	None	Primary education	5th to 9th grade	Secondary education	Higher education
0	0	9	14	9	6
1	0	0	0	0	0
2	0	0	0	0	0
3	0	0	0	0	0
4	0	0	0	0	1
5	0	2	0	3	2
6	0	2	4	5	4
7	0	2	1	3	3
8	0	6	11	7	8
9	1	4	6	8	9
10	0	13	14	15	14
11	0	7	15	11	14
12	0	6	10	8	7
13	0	4	10	7	10
14	0	1	8	3	15
15	2	1	5	9	16
16	0	2	1	3	10
17	0	0	2	2	2
18	0	0	1	5	6
19	0	0	1	1	3
20	0	0	0	0	1

Table 3.50: Mother's education and final grade.
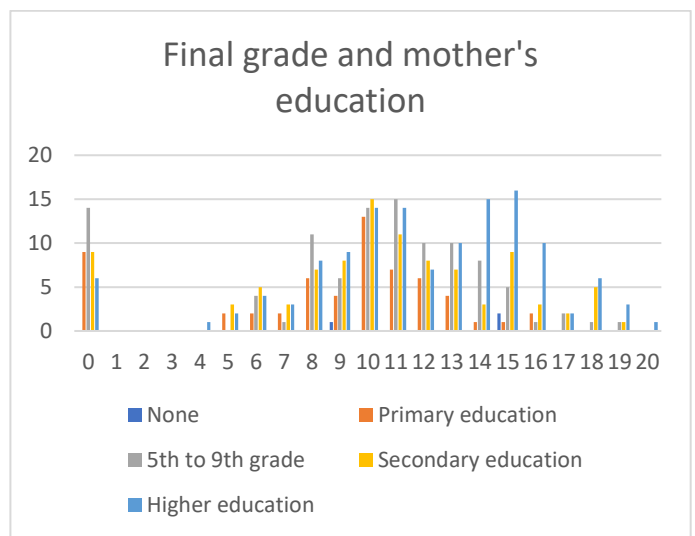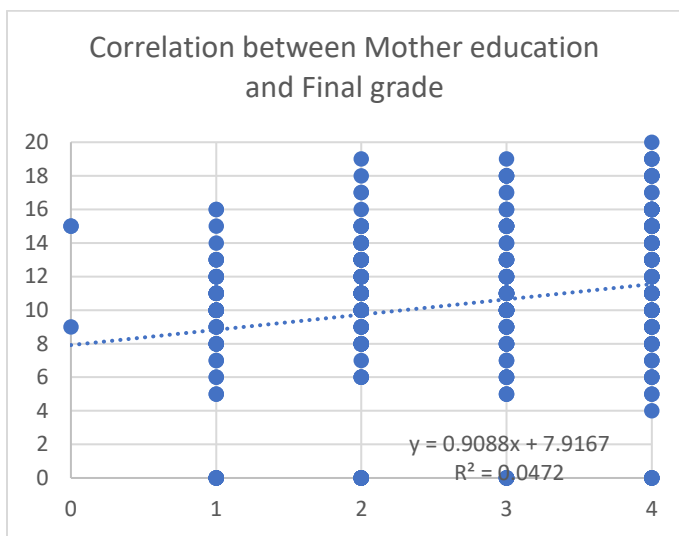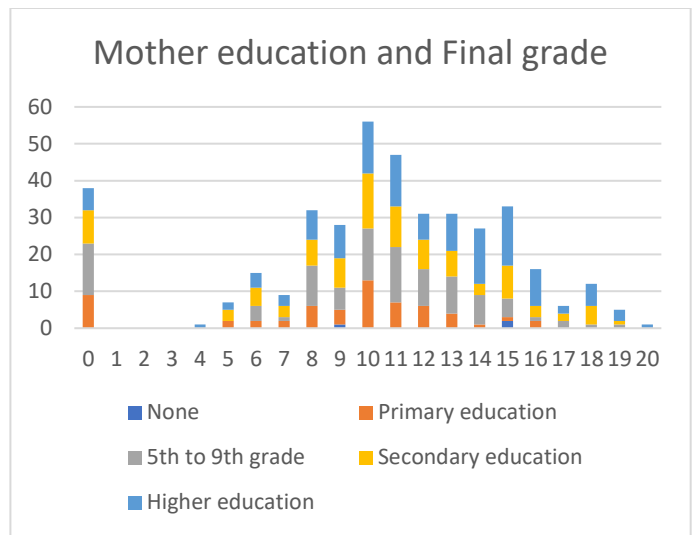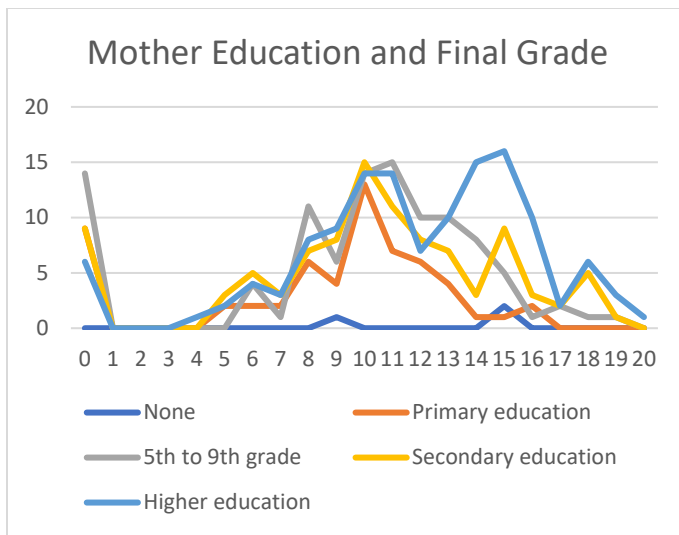
Figure 3.95: Mother's education and final grade.

- A positive correlation between mother education level and the students' final grade with 0.217

Father's education and final grade:

	None	Primary education	5th to 9th grade	Secondary education	Higher education
0	0	11	13	6	8
1	0	0	0	0	0
2	0	0	0	0	0
3	0	0	0	0	0
4	0	0	0	1	0
5	0	1	3	1	2

6	0	5	4	3	3
7	0	5	1	1	2
8	0	8	8	12	4
9	0	5	9	9	5
10	1	16	14	18	7
11	0	6	17	11	13
12	0	8	8	7	8
13	0	5	9	8	9
14	0	4	8	4	11
15	0	3	12	8	10
16	1	3	4	4	4
17	0	0	0	2	4
18	0	2	2	4	4
19	0	0	3	0	2
20	0	0	0	1	0

Table 3.51: Father's education and final grade.
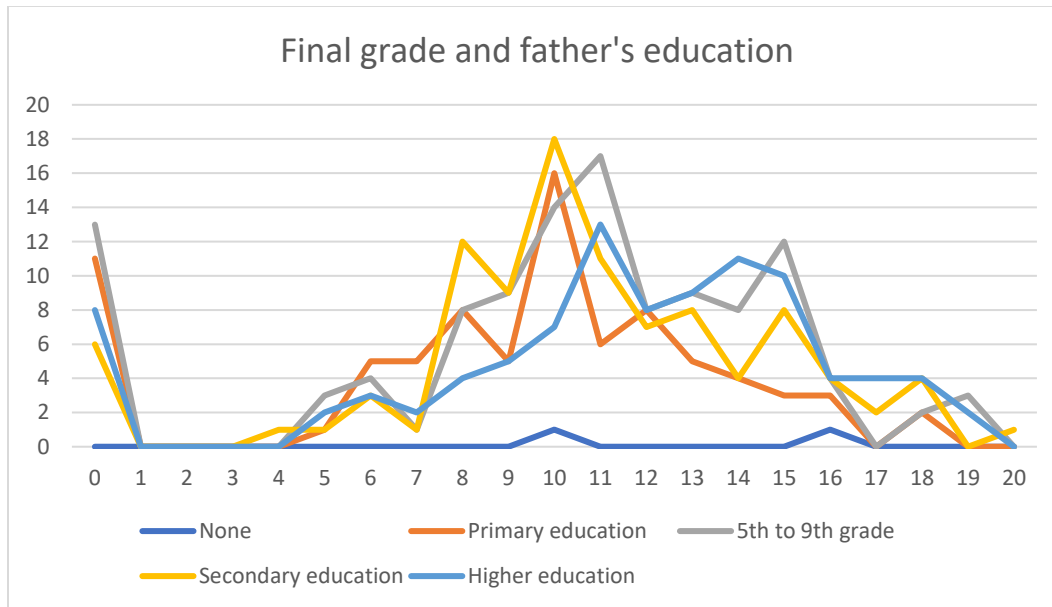
Figure 3.96: Father's education and final grade.

- A positive correlation between Father education level and the students' final grade with 0.152457 Where mother education and final grade is 0.217

- Mother education influences more on children's education than father education.

Travel time and final grade:

G3	Traveltime			
	1<15 min	15 to 30 min	30 min to hour	> 1 hour
0	21	12	4	1
1	0	0	0	0
2	0	0	0	0
3	0	0	0	0
4	1	0	0	0
5	6	1	0	0
6	8	6	1	0
7	5	3	0	1
8	19	10	3	0
9	20	7	0	1
10	29	19	5	3
11	34	10	2	1
12	20	9	2	0
13	19	8	3	1
14	22	5	0	0
15	23	8	2	0
16	12	4	0	0
17	4	2	0	0
18	9	2	1	0
19	4	1	0	0
20	1	0	0	0

Table 3.52: Home to school travel time and final grade
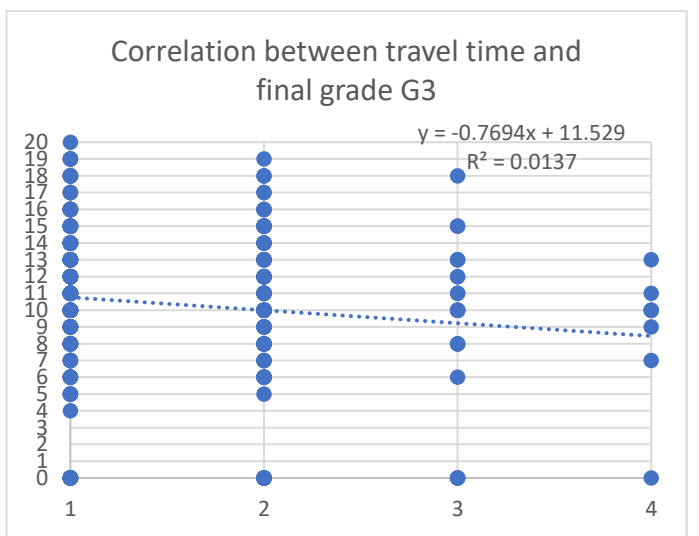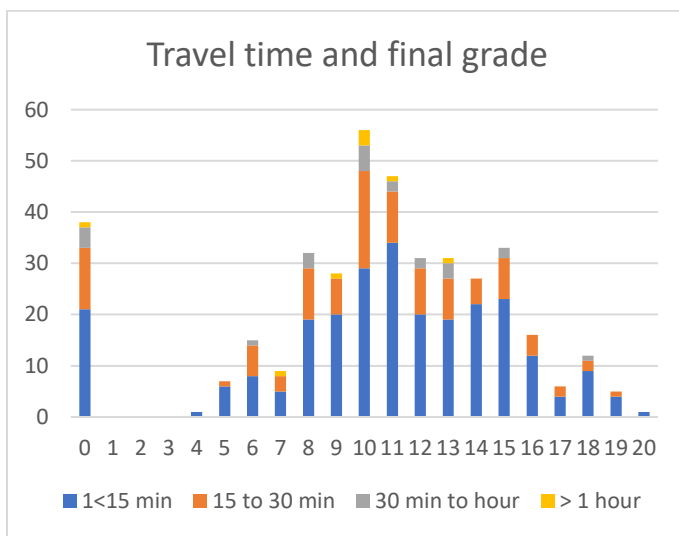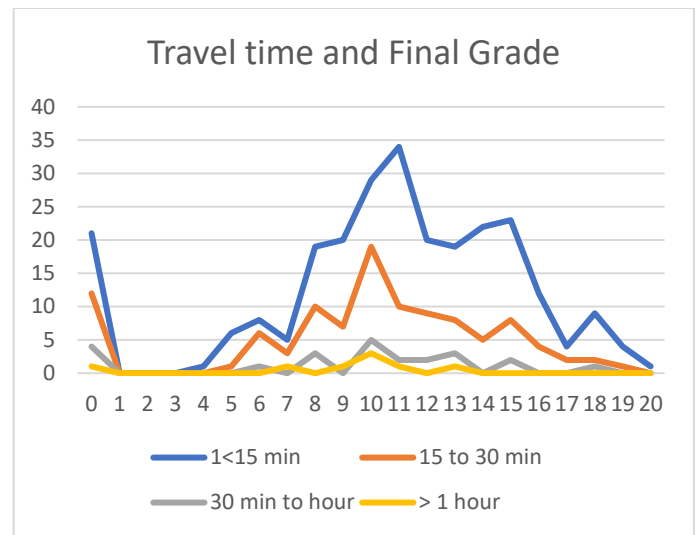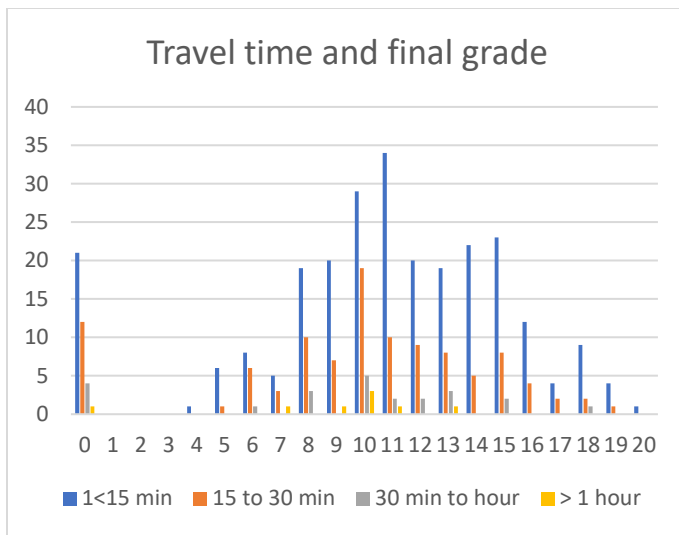
Figure 3.97: Home to school travel time and final grade.

- A negative correlation between home to school travel time and final grave with value -0.11714

Study time and final grad:

G3	studytime			
	< 2 hours	2 to 5 hours	5 to 10 hours	4 > 10 hours
0	13	16	6	3
1	0	0	0	0
2	0	0	0	0
3	0	0	0	0
4	0	1	0	0
5	4	3	0	0

6	3	11	0	1
7	2	6	1	0
8	11	15	3	3
9	4	18	6	0
10	17	28	8	3
11	7	29	8	3
12	10	15	4	2
13	9	17	3	2
14	7	8	9	3
15	7	16	8	2
16	4	7	4	1
17	0	4	1	1
18	5	3	2	2
19	2	1	2	0
20	0	0	0	1

Table 3.53: Weekly study time and final grade.
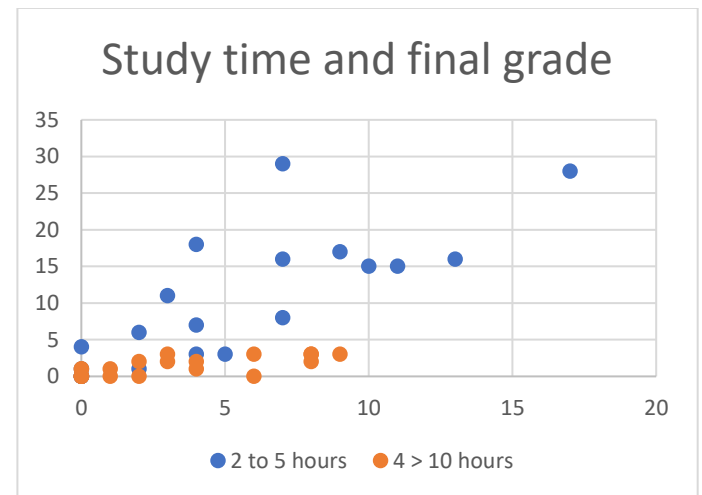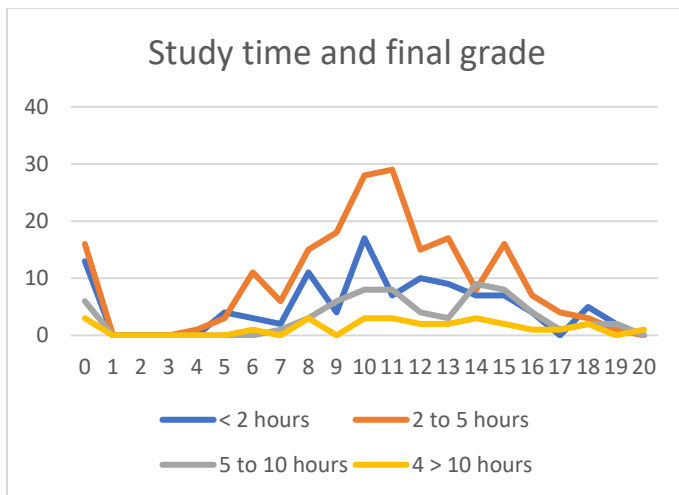
Figure 3.98: Weekly study time and final grade.

- A positive correlation between weekly study time and final grade with a value 0.098

Number of past class failures and final grade:
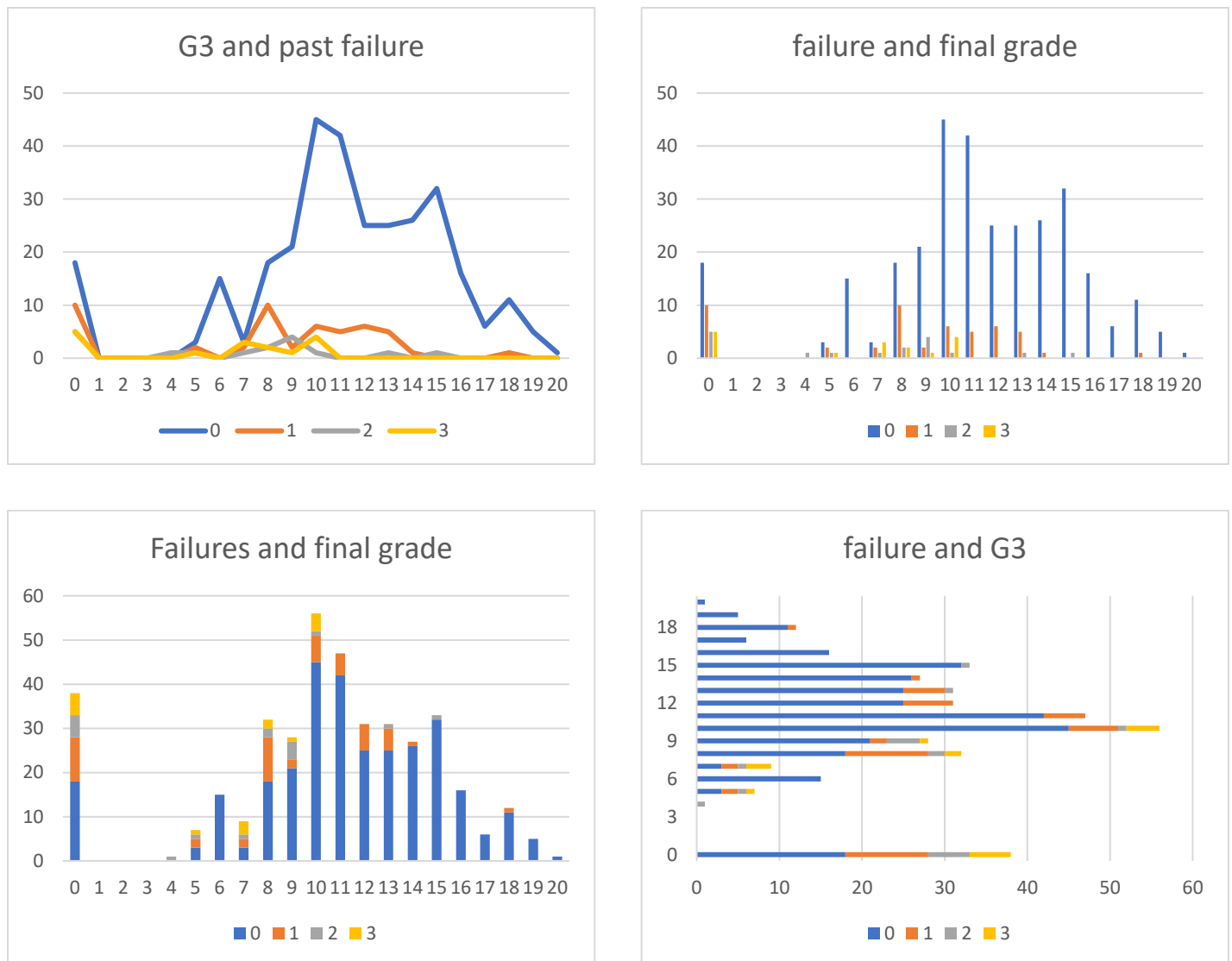
Figure 3.99: Number of past class failures and final grade.

- A negative correlation between number of past class failures and final grade with a value -0.36041
- There is a high trend in the data that student who did well in the past also going better in the following examinations.
- The correlation between G1 and G2 is 0.85 and correlation between G1 and G3 is 0.8
- The correlation between G2 and G3 is 0.904

Extra education school support and final grade:
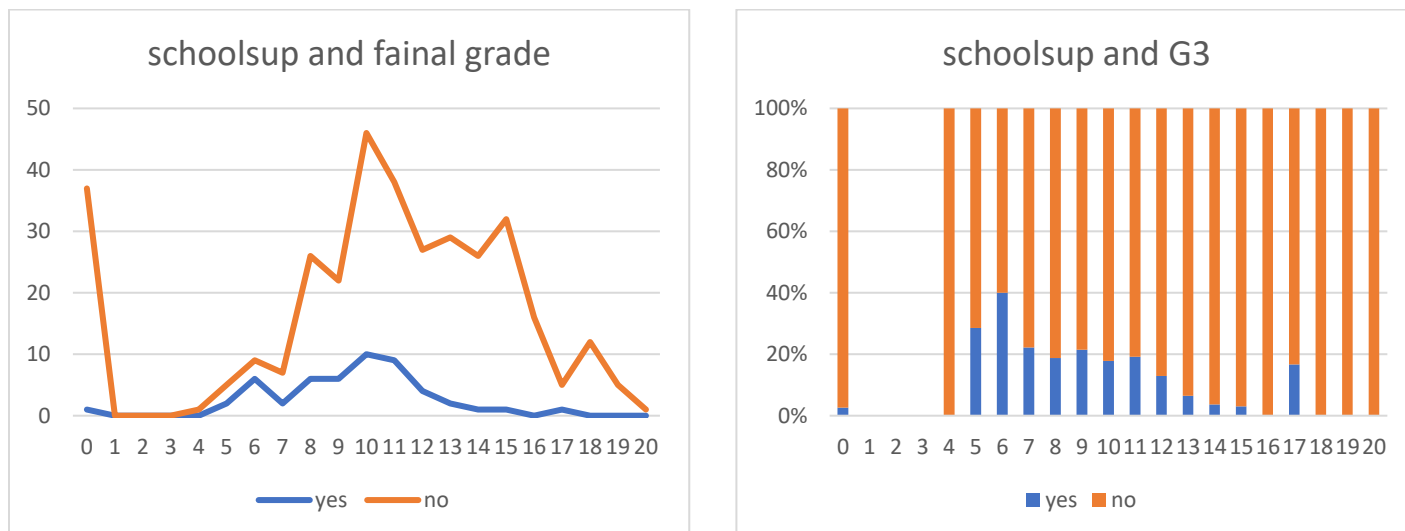
Figure 3.100: Extra education school support and final grade.

- A negative correlation between schoolsup (extra educational school support) and final grade with a value -0.083
- It seems like school is giving extra educational school support to the weak students and many of them have not passed in the final grade. There is a little improvement in the final grade.

Family education support and Final grade:
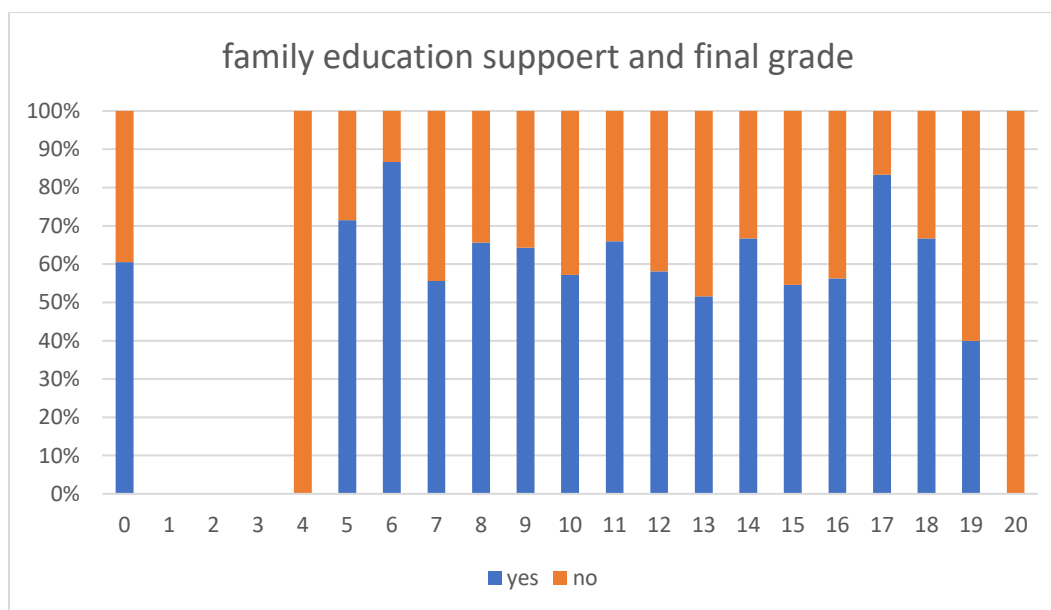
Figure 3.101: Family education support and final grade.

Students wants to take higher education and final grade:
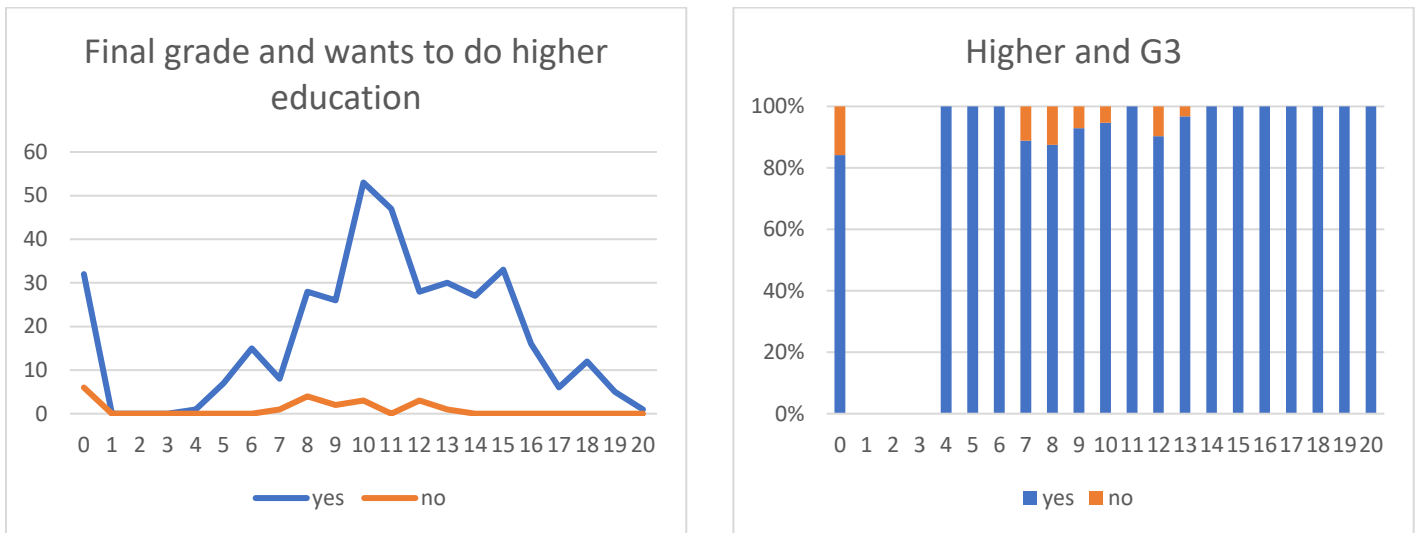
Figure 3.102: Students wants to take higher education and final grade.
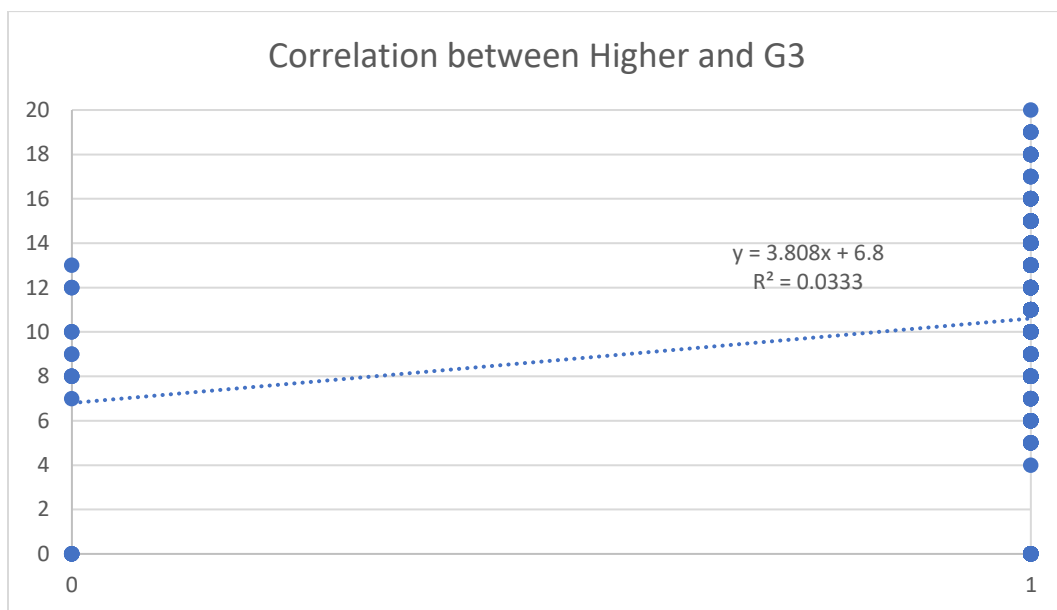
Figure 3.103: Correlation between students wants to take higher education and final grade.

- A positive correlation between higher (wants to take higher education) and final grade with value 0.182.
- Students trend to do better who wants to take higher education.

Quality of the family relationships and final grade:

G3/famrel	1. Very bad	2. Bad	3. Good	4. Very good	5.Excellent
0	1	3	7	17	10
1	0	0	0	0	0
2	0	0	0	0	0
3	0	0	0	0	0
4	0	0	1	0	0
5	0	1	2	2	2
6	0	0	3	9	3
7	0	1	0	5	3
8	1	1	5	18	7
9	0	1	8	13	6
10	0	0	8	33	15
11	1	3	8	25	10
12	2	1	5	13	10
13	2	2	6	14	7
14	0	0	5	13	9
15	0	3	6	14	10
16	1	1	0	9	5
17	0	1	0	3	2
18	0	0	3	4	5
19	0	0	1	2	2
20	0	0	0	1	0

Table 3.54: Quality of the family relationships and final grade.
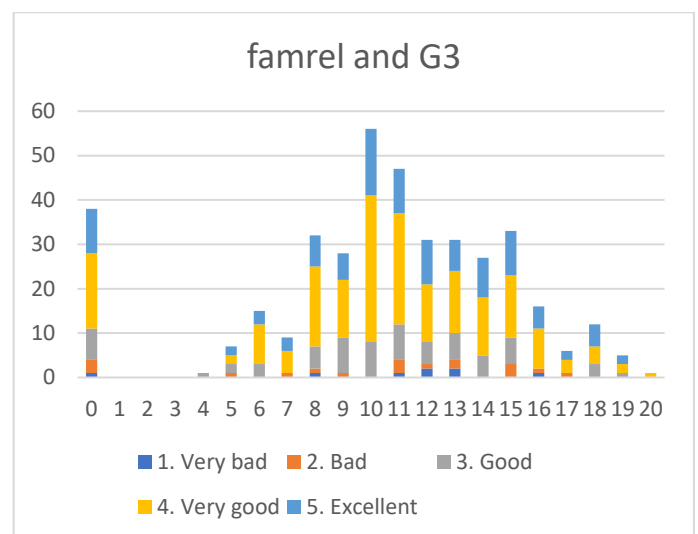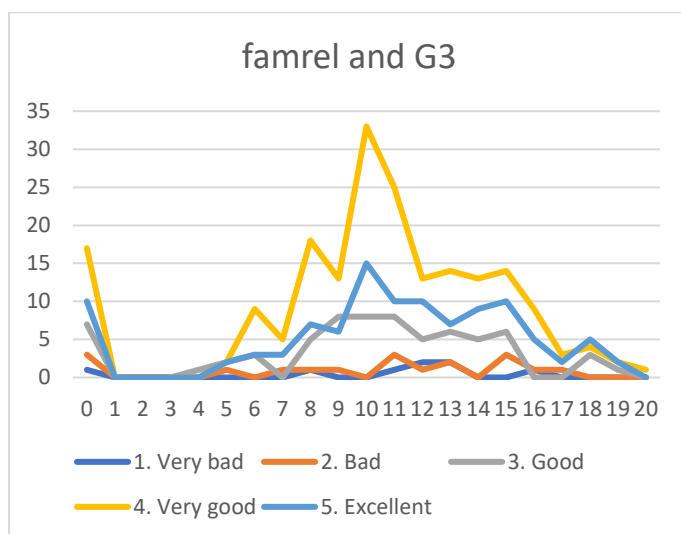
Figure 3.104: Quality of the family relationships and final grade.

- Quality of family relationships have very little influence on final grade with a correlation value of 0.051

Going out with friends and Final grade:

G3	goout				
	Very low	Low	Medium	High	Very high
0	4	9	8	9	8
1	0	0	0	0	0
2	0	0	0	0	0
3	0	0	0	0	0
4	0	0	0	0	1
5	0	0	3	1	3
6	1	3	4	5	2
7	0	1	5	2	1
8	0	5	7	14	6
9	1	6	10	7	4
10	5	13	20	11	7
11	4	14	14	9	6
12	1	12	11	4	3
13	2	10	9	7	3
14	0	5	13	8	1
15	2	11	14	5	1
16	1	7	3	1	4
17	2	1	2	0	1
18	0	3	5	2	2
19	0	2	2	1	0
20	0	1	0	0	0

Table 3.55: Going out with friends and final grade.
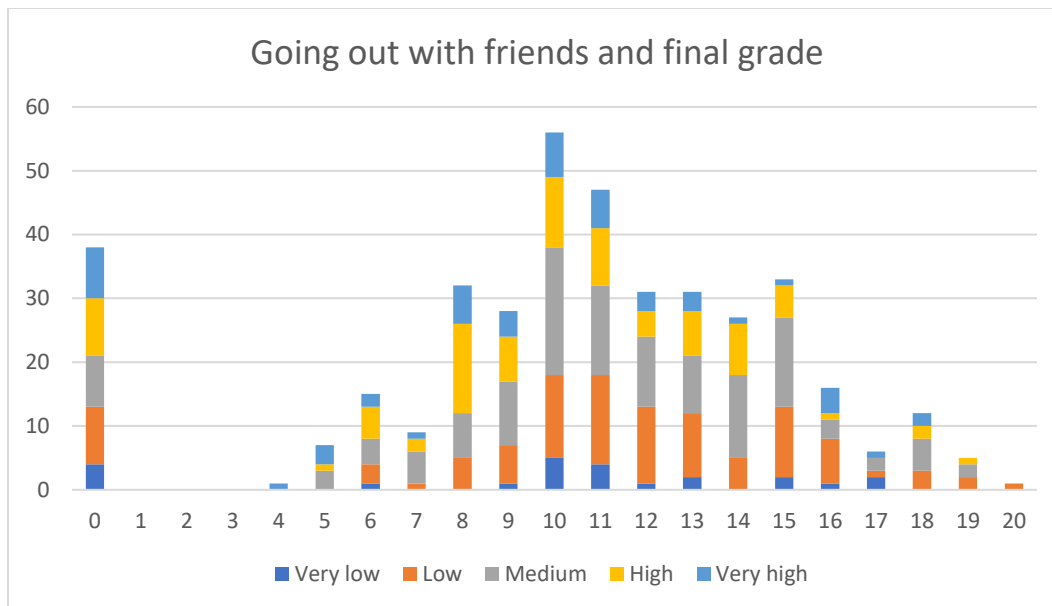
Figure 3.105: Going out with friends and final grade.

- A negative correlation between going out and final grade with value -0.133

Workday alcohol consumption and Final grade:

	Very low	Low	Medium	High	Very high
0	26	11	1	0	0
1	0	0	0	0	0
2	0	0	0	0	0
3	0	0	0	0	0
4	0	1	0	0	0
5	3	2	0	1	1
6	13	2	0	0	0
7	6	1	2	0	0
8	18	10	2	1	1
9	17	4	4	3	0
10	36	12	7	0	1
11	35	8	1	1	2
12	20	7	2	1	1
13	19	4	3	2	3
14	22	4	1	0	0
15	29	4	0	0	0
16	12	2	2	0	0
17	4	1	1	0	0

	Very low	Low	Medium	High	Very high
18	10	2	0	0	0
19	5	0	0	0	0
20	1	0	0	0	0

Table 3.56: Workday alcohol consumption and final grade.
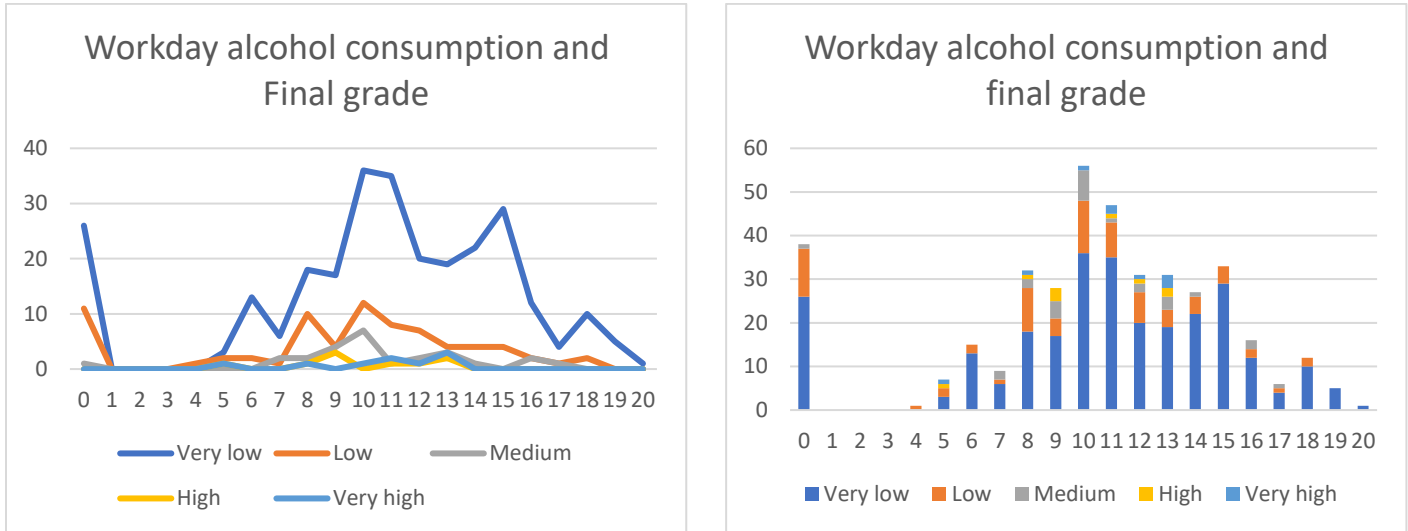
Figure 3.106: Workday alcohol consumption and final grade.

- A negative correlation between Workday alcohol consumption and final grade with value -0.05466
- Workday alcohol consumption has a very low negative influence on final grade.

Weekend alcohol consumption and Final grade:

	Very low	Low	Medium	High	Very high
0	18	12	3	3	2
1	0	0	0	0	0
2	0	0	0	0	0
3	0	0	0	0	0
4	0	0	0	1	0
5	0	1	3	2	1
6	8	5	1	1	0
7	3	1	2	1	2
8	9	3	10	7	3
9	12	3	6	5	2
10	14	9	16	11	6
11	17	15	8	5	2
12	9	7	8	6	1

13	12	6	4	4	5
14	11	9	5	1	1
15	15	8	8	2	0
16	7	4	2	1	2
17	4	0	1	1	0
18	7	1	3	0	1
19	4	1	0	0	0
20	1	0	0	0	0

Table 3.57: Weekend alcohol consumption and final grade.
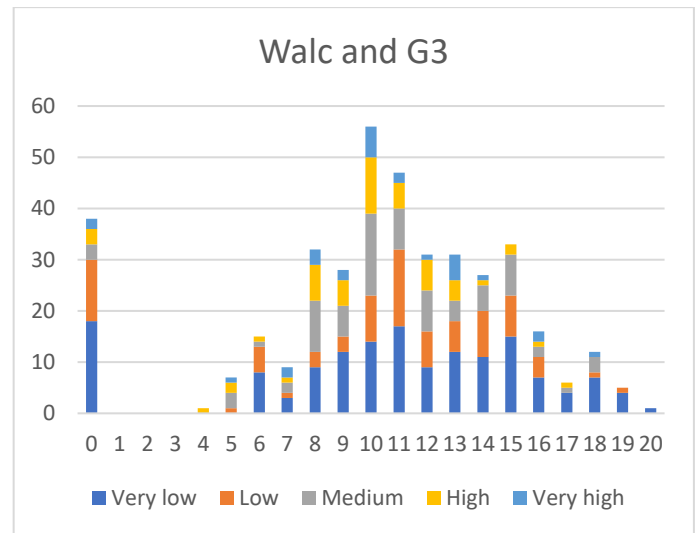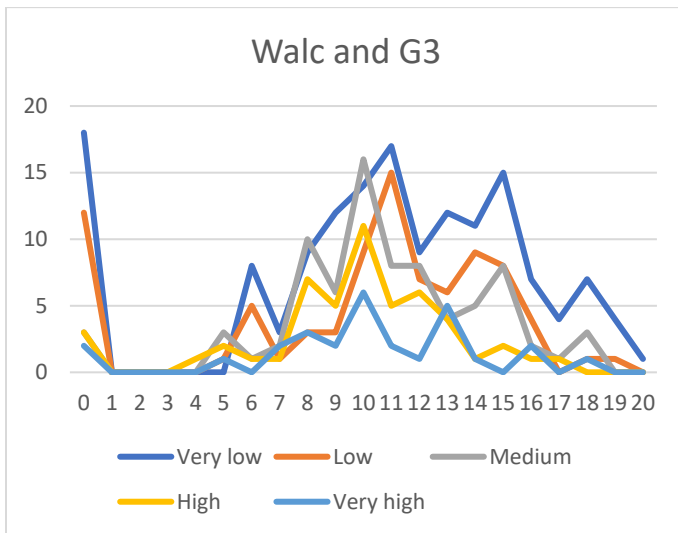
Figure 3.107: Weekend alcohol consumption and final grade.

- A negative correlation between Weekend alcohol consumption and final grade with value - -0.05194
- Weekend alcohol consumption has a very low negative influence on final grade.

Student current health status and Final grade:

G3	Health				
	Very bad	Bad	Not bad nor not good	Good	Very good
0	2	7	8	8	13
1	0	0	0	0	0
2	0	0	0	0	0
3	0	0	0	0	0
4	1	0	0	0	0
5	1	0	2	1	3

6	1	3	3	1	7
7	0	0	3	3	3
8	4	2	8	5	13
9	1	3	7	6	11
10	7	8	16	11	14
11	5	2	13	7	20
12	1	2	11	3	14
13	5	4	4	5	13
14	4	6	5	5	7
15	8	2	4	3	16
16	1	2	4	3	6
17	2	0	1	1	2
18	2	2	1	3	4
19	2	1	1	1	0
20	0	1	0	0	0

Table 3.58: Student current health status and final grade.
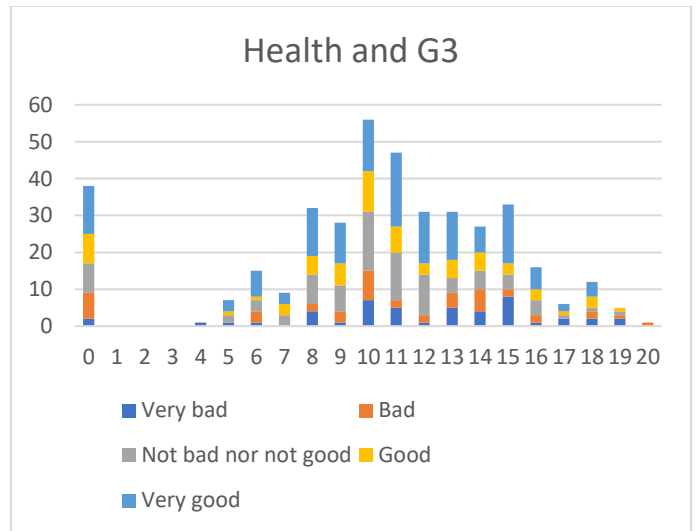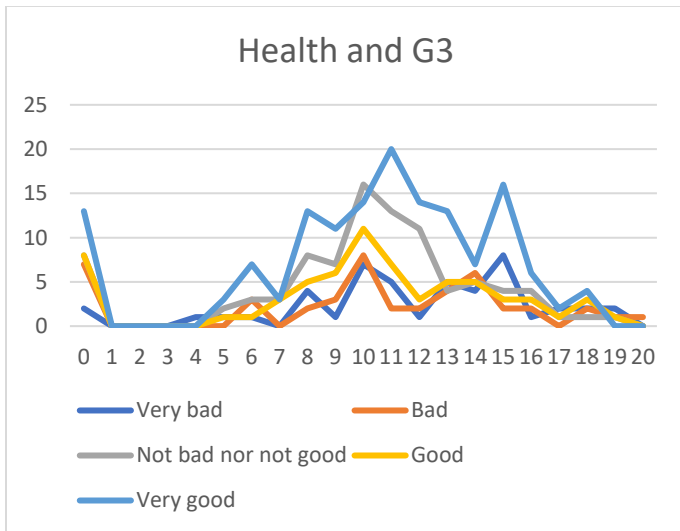
Figure 3.108: Student current health status and final grade.

- A negative correlation between health and final grade with value -0.06133

3.3 Correlation matrix:

	school01	Sex01	age	Address01	famsize01	Pstatus01	Medu	Fedu	traveltime	studytime	failures	Schoolsup01	famsup01	Paid01	Activities01	Nursery01	Higher01	Internet01	romantic01	famrel	freetime	goout	Dalc	Walc	health	absences	G1	G2	G3
school01	1																												
Sex01	0.012286	1																											
age	-0.37761	-0.02861	1																										
Address01	0.279797	-0.0285	-0.14672	1																									
famsize01	-0.04592	-0.08986	-0.03785	-0.07247	1																								
Pstatus01	0.064866	-0.08986	0.029598	-0.04257	0.149612	1																							
Medu	0.133333	0.078228	-0.16366	0.138804	0.043068	-0.12356	1																						
Fedu	0.0079007	0.034878	-0.16344	0.072178	0.058879	-0.08873	0.623455	1																					
traveltime	-0.24231	0.059722	0.070641	-0.3281	-0.06349	0.028265	-0.17164	-0.15819	1																				
studytime	0.090681	-0.30627	-0.0414	-0.02091	0.073595	0.024294	0.064944	-0.00917	-0.10091	1																			
failures	-0.0598	0.044436	0.245665	-0.07858	0.015769	-0.00334	-0.23668	-0.25041	0.092239	-0.17356	1																		
Schoolsup01	0.139769	-0.13827	-0.14061	0.023903	0.112893	0.019067	-0.04224	-0.03603	0.03753	0.037763	-0.10044	1																	
famsup01	0.164967	-0.0082	-0.08663	0.059589	0.00113	-0.09061	0.187727	0.086981	-0.00029	0.16722	-0.18804	-0.0205	1																
Paid01	0.017083	-0.12913	0.0528	0.013882	0.046435	0.1597	0.185496	0.157177	0.145238	0.146722	0.089877	-0.05307	0.194681	1															
Activities01	0.116946	0.098933	-0.10306	0.005257	0.000113	0.097352	0.108277	0.112643	-0.00777	0.089877	-0.110073	0.0045967	0.103583	0.048663	1														
Nursery01	0.089277	0.164469	-0.15162	0.029908	0.0070074	-0.04073	0.1597	0.174566	-0.09351	0.175081	-0.30002	0.046032	0.189214	0.096484	0.054803	1													
Higher01	0.02415	-0.15106	-0.20908	0.142854	0.201463	0.168845	0.185496	0.157177	-0.10015	0.081325	-0.110073	-0.0015	0.102143	-0.02075	0.194681	0.020374	1												
Internet01	0.133578	0.044113	-0.11209	0.22184	-0.00072	0.0070074	0.201463	0.127507	-0.1113	0.059422	-0.06245	-0.00968	0.103583	0.00783	0.022374	0.153332	0.0048602	1											
romantic01	-0.0607	-0.10202	0.164469	0.005257	-0.03439	-0.04045	0.039681	0.015602	0.021962	0.053285	0.093137	-0.08077	0.01244	0.005536	0.019651	0.027495	-0.10566	0.087122	1										
famrel	0.047926	0.058971	0.014258	0.022776	0.038717	0.025179	-0.00391	-0.00137	0.01681	0.039731	-0.04434	-0.00134	0.00046	0.040687	-0.00358	0.02419	0.032768	-0.06382	0.150701	1									
freetime	-0.00299	0.238744	0.034878	-0.0177	0.030891	-0.01286	-0.01702	0.039731	-0.1432	0.091987	-0.04547	0.010538	-0.06425	0.089728	-0.0247	-0.06124	0.051286	0.024319	0.032768	-0.06382	1								
goout	0.007152	0.073897	0.068835	-0.02306	0.0083439	0.0083105	0.02854	-0.01702	-0.0639	-0.0377	-0.01563	0.010093	0.046088	0.004612	0.010108	0.027837	0.00782	0.064568	0.285019	0.266904	1								
Dalc	-0.11421	0.268171	0.133125	-0.09340	-0.10152	-0.03059	0.019834	0.002386	0.138325	-0.19602	0.136047	-0.02149	-0.03158	0.062465	-0.16651	0.063465	0.179129	-0.03719	0.022168	0.013613	-0.1491	-0.09416	-0.13618	0.136291	1				
Walc	-0.06609	0.274194	0.117276	-0.10113	-0.10343	0.006045	-0.04712	-0.01263	0.134116	-0.25378	0.141962	-0.08669	-0.08454	-0.07748	-0.0953	-0.10034	0.011667	-0.01014	-0.1134	0.147822	0.420386	0.647544	-0.08493	-0.09772	-0.02994	1			
health	0.042651	0.145388	-0.06219	-0.04035	0.028992	0.022307	-0.04688	0.014742	-0.07562	0.065827	-0.03412	0.023923	-0.01347	-0.08019	0.028542	0.094056	0.075733	-0.00958	0.017718	0.092476	-0.06133	0.034247	-0.05466	-0.05194	-0.06133	1			
absences	0.08848	-0.066696	0.17523	-0.02787	-0.03578	-0.13494	0.100285	0.024473	-0.01294	0.063726	0.022326	0.024833	0.007435	-0.01362	0.019155	-0.05609	0.101701	0.028742	0.094056	0.044302	0.111008	0.136291	0.0801468	0.852118	-0.02994	1			
G1	0.025731	0.091839	-0.14447	0.126037	-0.00122	-0.01687	0.205341	0.190027	-0.09304	0.19027	-0.21261	-0.083457	0.039079	0.05701	0.069763	0.179129	0.071619	0.153384	-0.04435	0.044302	0.111008	-0.1491	-0.09416	-0.13618	0.136291	0.0801468	1		
G2	0.050086	0.091099	-0.14447	0.126037	-0.00122	-0.04138	0.215227	0.164899	-0.1532	0.13588	-0.17261	-0.083457	0.039079	0.05701	0.069763	0.179129	0.071619	0.153384	-0.04435	0.016225	-0.1491	-0.09416	-0.12618	-0.07317	-0.06133	0.03178	0.852118	1	
G3	0.043017	0.109456	-0.16158	0.105756	-0.08141	-0.05801	0.217147	0.152457	-0.11714	0.09782	-0.36041	-0.108279	-0.03916	0.101996	0.01161	0.053568	0.182465	0.098483	-0.12997	0.053363	0.011307	-0.13279	-0.05466	-0.05194	-0.06133	0.034247	0.801468	0.904868	1

Table 3.59: Correlation matrix of the variables.

81

3.4 Regression

School and Final Grade:

Regression between the independent variable School and the final grade (G3) dependent variable:

SUMMARY OUTPUT								
Regression Statistics								
Multiple R	0.045017							
R Square	0.002027							
Adjusted R Square	-0.00051							
Standard Error	4.582617							
Observations	395							
ANOVA								
	df	*SS*	*MS*	*F*	*Significance F*			
Regression	1	16.75918	16.75918	0.798042	0.372226			
Residual	393	8253.15	21.00038					
Total	394	8269.909						
	Coefficients	*Standard Error*	*t Stat*	*P-value*	*Lower 95%*	*Upper 95%*	*Lower 95.0%*	*Upper 95.0%*
Intercept	9.847826	0.67567	14.5749	8.97E-39	8.519446	11.17621	8.519446	11.17621
school01	0.642145	0.718821	0.893332	0.372226	-0.77107	2.05536	-0.77107	2.05536

Table 3.60: a simple regression model with school and final grade.

p-value of the school variable is 0.372226 which is greater than the threshold value. So, it is not statistically significant.
The lower bound of the confidence interval (Lower 95%) for the school is -0.77107 and the upper bound (Upper 95%) is 2.05536. Zero falls within this confidence interval. So, the school variable is not statistically significant.

Sex and the Final Grade:

Regression between independent variable Sex and the Final Grade (G3) dependent variable:

SUMMARY OUTPUT								
Regression Statistics								
Multiple R	0.103456							
R Square	0.010703							
Adjusted R Square	0.008186							
Standard Error	4.562653							
Observations	395							
ANOVA								
	df	*SS*	*MS*	*F*	*Significance F*			
Regression	1	88.51342	88.51342	4.251814	0.039865			
Residual	393	8181.395	20.8178					
Total	394	8269.909						
	Coefficients	*Standard Error*	*t Stat*	*P-value*	*Lower 95%*	*Upper 95%*	*Lower 95.0%*	*Upper 95.0%*
Intercept	9.966346	0.316363	31.50288	1.4E-109	9.344371	10.58832	9.344371	10.58832
Sex01	0.948092	0.459794	2.061993	0.039865	0.044128	1.852056	0.044128	1.852056

Table 3.61: a simple regression model with sex and final grade

p-value of the sex variable is 0.039865 which is smaller than the threshold value (0.05). So, it is statistically significant. The lower bound of the confidence interval (Lower 95%) for the sex variable is 0.044128 and the upper bound (Upper 95%) is 1.852056. Zero does not fall within this confidence interval.
So, the sex variable is statistically significant. By the R-squared value, this simple regression model with sex variable explains about 1% of the data.

Age and the Final Grade:

Regression between independent variable Age and the Final Grade (G3) dependent variable:

SUMMARY OUTPUT								
Regression Statistics								
Multiple R	0.161579							
R Square	0.026108							
Adjusted R Square	0.02363							
Standard Error	4.52699							
Observations	395							
ANOVA								
	df	*SS*	*MS*	*F*	*Significance F*			
Regression	1	215.9101	215.9101	10.53547	0.001271			
Residual	393	8053.999	20.49364					
Total	394	8269.909						
	Coefficients	*Standard Error*	*t Stat*	*P-value*	*Lower 95%*	*Upper 95%*	*Lower 95.0%*	*Upper 95.0%*
Intercept	20.10111	2.992784	6.716525	6.54E-11	14.21724	25.98498	14.21724	25.98498
age	-0.58013	0.178729	-3.24584	0.001271	-0.93151	-0.22874	-0.93151	-0.22874

Table 3.62: a simple regression model with age and final grade

p-value of the age variable is 0.001271 which is smaller than the threshold value (0.05). So, it is statistically significant.
The lower bound of the confidence interval (Lower 95%) for the age variable is -0.93151 and the upper bound (Upper 95%) is -0.22874. Zero does not falls within this confidence interval. So, the age variable is statistically significant.
By the R-squared value, this simple regression model with age variable explains about 2% of the data.

Address and the Final Grade:
Regression between independent variable Address and the Final Grade (G3) dependent variable:

SUMMARY OUTPUT								
Regression Statistics								
Multiple R	0.105756							
R Square	0.011184							
Adjusted R Square	0.008668							
Standard Error	4.561543							
Observations	395							
ANOVA								
	df	*SS*	*MS*	*F*	*Significance F*			
Regression	1	92.49351	92.49351	4.445164	0.035633			
Residual	393	8177.415	20.80767					
Total	394	8269.909						
	Coefficients	*Standard Error*	*t Stat*	*P-value*	*Lower 95%*	*Upper 95%*	*Lower 95.0%*	*Upper 95.0%*
Intercept	9.511364	0.486262	19.56016	5.51E-60	8.555363	10.46736	8.555363	10.46736
Address01	1.162903	0.551569	2.108356	0.035633	0.078509	2.247298	0.078509	2.247298

Table 3.63: a simple regression model with address and final grade

p-value of the address variable is 0.035633 which is smaller than the threshold value (0.05). So, it is statistically significant.

The lower bound of the confidence interval (Lower 95%) for the address variable is 0.078509 and the upper bound (Upper 95%) is 2.247298. Zero does not fall within this confidence interval. So, the address variable is statistically significant.

Family size and the Final Grade:

Regression between independent variable Family size and the Final Grade (G3) dependent variable:

SUMMARY OUTPUT								
Regression Statistics								
Multiple R	0.081407							
R Square	0.006627							
Adjusted R Square	0.004099							
Standard Error	4.572042							
Observations	395							
ANOVA								
	df	*SS*	*MS*	*F*	*Significance F*			
Regression	1	54.80566	54.80566	2.621832	0.106205			
Residual	393	8215.103	20.90357					
Total	394	8269.909						
	Coefficients	*Standard Error*	*t Stat*	*P-value*	*Lower 95%*	*Upper 95%*	*Lower 95.0%*	*Upper 95.0%*
Intercept	11	0.428211	25.68827	4.03E-86	10.15813	11.84187	10.15813	11.84187
famsize01	-0.82206	0.507695	-1.61921	0.106205	-1.8202	0.176074	-1.8202	0.176074

Table 3.64: a simple regression model with family size and final grade.

p-value of the famsize (family size) variable is 0.106205 which is greater than the threshold value (0.05). So, it is not statistically significant. The lower bound of the confidence interval (Lower 95%) for the famsize variable is -1.8202 and the upper bound (Upper 95%) is 0.176074. Zero falls within this confidence interval. So, the famsize variable is not statistically significant.

Parents' cohabitation status and the Final Grade:

Regression between independent variable parents' cohabitation status and the Final Grade (G3) dependent variable:

SUMMARY OUTPUT								
Regression Statistics								
Multiple R	0.058009							
R Square	0.003365							
Adjusted R Square	0.000829							
Standard Error	4.579543							
Observations	395							
ANOVA								
	df	*SS*	*MS*	*F*	*Significance F*			
Regression	1	27.82859	27.82859	1.326927	0.250053			
Residual	393	8242.08	20.97221					
Total	394	8269.909						
	Coefficients	*Standard Error*	*t Stat*	*P-value*	*Lower 95%*	*Upper 95%*	*Lower 95.0%*	*Upper 95.0%*
Intercept	11.19512	0.715204	15.65304	2.88E-43	9.789017	12.60123	9.789017	12.60123
Pstatus01	-0.87026	0.755487	-1.15192	0.250053	-2.35557	0.615039	-2.35557	0.615039

Table 3.65: a simple regression model with Parent's cohabitation status and final grade.

p-value of the Pstatus (Parent's cohabitation status) variable is 0.250053 which is greater than the threshold value (0.05). So, it is not statistically significant.

The lower bound of the confidence interval (Lower 95%) for the Pstatus variable is -2.35557 and the upper bound (Upper 95%) is 0.615039. Zero falls within this confidence interval. So, the Pstatus variable is not statistically significant.

Mothers' education and the Final Grade:

Regression between independent variable mother's education and the Final Grade (G3) dependent variable:

SUMMARY OUTPUT								
Regression Statistics								
Multiple R	0.217147							
R Square	0.047153							
Adjusted R Square	0.044728							
Standard Error	4.47781							
Observations	395							
ANOVA								
	df	*SS*	*MS*	*F*	*Significance F*			
Regression	1	389.9513	389.9513	19.44818	1.34E-05			
Residual	393	7879.958	20.05078					
Total	394	8269.909						
	Coefficients	*Standard Error*	*t Stat*	*P-value*	*Lower 95%*	*Upper 95%*	*Lower 95.0%*	*Upper 95.0%*
Intercept	7.916682	0.609709	12.98437	2.51E-32	6.717983	9.115381	6.717983	9.115381
Mother's education	0.908758	0.206067	4.410009	1.34E-05	0.503626	1.313889	0.503626	1.313889

Table 3.66: a simple regression model with Mother's education and final grade.

p-value of the Medu (Mother's education) variable is 0.000013 which is very smaller than the threshold value (0.05). So, it is statistically significant.

The lower bound of the confidence interval (Lower 95%) for the Medu variable is 0.503626 and the upper bound (Upper 95%) is 1.313889. Zero does not fall within this confidence interval. So, the medu variable is statistically significant.

Father's education and the Final Grade:

Regression between independent variable father's education and the Final Grade (G3) dependent variable:

SUMMARY OUTPUT								
Regression Statistics								
Multiple R	0.152457							
R Square	0.023243							
Adjusted R Square	0.020758							
Standard Error	4.533643							
Observations	395							
ANOVA								
	df	*SS*	*MS*	*F*	*Significance F*			
Regression	1	192.2185	192.2185	9.351913	0.00238			
Residual	393	8077.69	20.55392					
Total	394	8269.909						
	Coefficients	*Standard Error*	*t Stat*	*P-value*	*Lower 95%*	*Upper 95%*	*Lower 95.0%*	*Upper 95.0%*
Intercept	8.796727	0.576307	15.26395	1.23E-41	7.663696	9.929758	7.663696	9.929758
Father's Education	0.64186	0.209889	3.05809	0.00238	0.229214	1.054507	0.229214	1.054507

Table 3.67: a simple regression model between father's education and final grade.

p-value of the Fedu (father's education) variable is 0.00238 which is very smaller than the threshold value (0.05). So, it is statistically significant.

The lower bound of the confidence interval (Lower 95%) for the Fedu variable is 0.229214 and the upper bound (Upper 95%) is 1.054507. Zero does not fall within this confidence interval. So, the Fedu (Father's education) variable is statistically significant.

School travel time and the Final Grade:

Regression between independent variable home to school travel time and the Final Grade (G3) dependent variable:

SUMMARY OUTPUT								
Regression Statistics								
Multiple R	0.117142							
R Square	0.013722							
Adjusted R Square	0.011213							
Standard Error	4.555685							
Observations	395							
ANOVA								
	df	*SS*	*MS*	*F*	*Significance F*			
Regression	1	113.4818	113.4818	5.46788	0.01987			
Residual	393	8156.427	20.75427					
Total	394	8269.909						
	Coefficients	*Standard Error*	*t Stat*	*P-value*	*Lower 95%*	*Upper 95%*	*Lower 95.0%*	*Upper 95.0%*
Intercept	11.5294	0.528762	21.80453	1.18E-69	10.48984	12.56895	10.48984	12.56895
traveltime	-0.76943	0.329047	-2.33835	0.01987	-1.41634	-0.12251	-1.41634	-0.12251

Table 3.68: a simple regression model with home to school travel time and final grade.

p-value of the traveltime (house to school travel time) variable is 0.01987 which is smaller than the threshold value (0.05). So, it is statistically significant.

The lower bound of the confidence interval (Lower 95%) for the traveltime variable is -1.41634 and the upper bound (Upper 95%) is -0.12251. Zero does not fall within this confidence interval. So, the traveltime variable is statistically significant.

Study time and the Final Grade:

Regression between independent variable study time and the Final Grade (G3) dependent variable:

SUMMARY OUTPUT								
Regression Statistics								
Multiple R	0.09782							
R Square	0.009569							
Adjusted R Square	0.007049							
Standard Error	4.565268							
Observations	395							
ANOVA								
	df	*SS*	*MS*	*F*	*Significance F*			
Regression	1	79.13221	79.13221	3.796826	0.052061			
Residual	393	8190.777	20.84167					
Total	394	8269.909						
	Coefficients	*Standard Error*	*t Stat*	*P-value*	*Lower 95%*	*Upper 95%*	*Lower 95.0%*	*Upper 95.0%*
Intercept	9.328261	0.60326	15.4631	1.81E-42	8.142241	10.51428	8.142241	10.51428
study time	0.534001	0.274051	1.948545	0.052061	-0.00479	1.072791	-0.00479	1.072791

Table 3.69: a simple regression model with weekly study time and final grade.

p-value of the studytime (weekly study time) variable is 0.052061 which is slightly greater than the threshold value (0.05). So, it is not statistically significant.

The lower bound of the confidence interval (Lower 95%) for the studytime variable is -0.00479 and the upper bound (Upper 95%) is 1.072791. Zero falls within this confidence interval. So, the studytime variable is not statistically significant.

Past failures and the Final Grade:

Regression between independent variable failures and the Final Grade (G3) dependent variable:

SUMMARY OUTPUT								
Regression Statistics								
Multiple R	0.360415							
R Square	0.129899							
Adjusted R Square	0.127685							
Standard Error	4.278967							
Observations	395							
ANOVA								
	df	*SS*	*MS*	*F*	*Significance F*			
Regression	1	1074.252	1074.252	58.67167	1.47E-13			
Residual	393	7195.657	18.30956					
Total	394	8269.909						
	Coefficients	*Standard Error*	*t Stat*	*P-value*	*Lower 95%*	*Upper 95%*	*Lower 95.0%*	*Upper 95.0%*
Intercept	11.1572	0.236088	47.25869	3.4E-164	10.69305	11.62136	10.69305	11.62136
failures	-2.22042	0.289882	-7.65974	1.47E-13	-2.79034	-1.65051	-2.79034	-1.65051

Table 3.70: a simple regression model with number of past class failures and final grade.

p-value of the failures (number of past class failure) variable is 1.47E-13 which is very smaller than the threshold value (0.05). So, it is statistically significant.

The lower bound of the confidence interval (Lower 95%) for the failure variable is -2.79034 and the upper bound (Upper 95%) is -1.65051. Zero does not fall within this confidence interval. So, the failures variable is statistically significant.

Schoolsup and the Final Grade:

Regression between independent variable schoolsup and the Final Grade (G3) dependent variable:

SUMMARY OUTPUT								
Regression Statistics								
Multiple R	0.082788							
R Square	0.006854							
Adjusted R Square	0.004327							
Standard Error	4.57152							
Observations	395							
ANOVA								
	df	*SS*	*MS*	*F*	*Significance F*			
Regression	1	56.68103	56.68103	2.712167	0.100385			
Residual	393	8213.228	20.8988					
Total	394	8269.909						
	Coefficients	*Standard Error*	*t Stat*	*P-value*	*Lower 95%*	*Upper 95%*	*Lower 95.0%*	*Upper 95.0%*
Intercept	10.56105	0.24648	42.8475	3.5E-150	10.07646	11.04563	10.07646	11.04563
Schoolsup	-1.12967	0.685954	-1.64687	0.100385	-2.47827	0.218924	-2.47827	0.218924

Table 3.71: a simple regression model with extra educational school support and final grade.

p-value of the schoolsup (extra educational support) variable is 0.100385 which is greater than the threshold value (0.05). So, it is not statistically significant.

The lower bound of the confidence interval (Lower 95%) for the schoolsup variable is -2.47827 and the upper bound (Upper 95%) is 0.218924. Zero falls within this confidence interval. So, the schoolsup variable is not statistically significant.

famsup and the Final Grade:

Regression between independent variable famsup and the Final Grade (G3) dependent variable:

SUMMARY OUTPUT								
Regression Statistics								
Multiple R	0.039157							
R Square	0.001533							
Adjusted R Square	-0.00101							
Standard Error	4.58375							
Observations	395							
ANOVA								
	df	*SS*	*MS*	*F*	*Significance F*			
Regression	1	12.6801	12.6801	0.603505	0.437711			
Residual	393	8257.229	21.01076					
Total	394	8269.909						
	Coefficients	*Standard Error*	*t Stat*	*P-value*	*Lower 95%*	*Upper 95%*	*Lower 95.0%*	*Upper 95.0%*
Intercept	10.64052	0.370574	28.71361	1.56E-98	9.911967	11.36908	9.911967	11.36908
famsup01	-0.3678	0.473441	-0.77686	0.437711	-1.29859	0.562999	-1.29859	0.562999

Table 3.72: a simple regression model with family educational support and final grade.

p-value of the famsup (c) variable is 0.437711which is greater than the threshold value (0.05). So, it is not statistically significant.
The lower bound of the confidence interval (Lower 95%) for the famsup variable is -1.29859 and the upper bound (Upper 95%) is 0.562999 . Zero falls within this confidence interval. So, the famsup variable is not statistically significant.

Paid class and Final Grade:

Regression between independent variable paid and the Final Grade (G3) dependent variable:

SUMMARY OUTPUT								
Regression Statistics								
Multiple R	0.101996							
R Square	0.010403							
Adjusted R Square	0.007885							
Standard Error	4.563344							
Observations	395							
ANOVA								
	df	*SS*	*MS*	*F*	*Significance F*			
Regression	1	86.03379	86.03379	4.131451	0.042765			
Residual	393	8183.875	20.82411					
Total	394	8269.909						
	Coefficients	*Standard Error*	*t Stat*	*P-value*	*Lower 95%*	*Upper 95%*	*Lower 95.0%*	*Upper 95.0%*
Intercept	9.985981	0.311944	32.01212	1.5E-111	9.372694	10.59927	9.372694	10.59927
Paid	0.936671	0.460825	2.032597	0.042765	0.030681	1.84266	0.030681	1.84266

Table 3.73: a simple regression model with extra paid classes and final grade.

p-value of the paid (extra paid classes) variable is 0.042765 which is smaller than the threshold value (0.05). So, it is statistically significant.

The lower bound of the confidence interval (Lower 95%) for the paid variable is 0.030681 and the upper bound (Upper 95%) is 1.84266. Zero does not fall within this confidence interval. So, the paid variable is statistically significant.

Extra-curricular activities and the Final Grade:

Regression between independent variable Activities and the Final Grade (G3) dependent variable:

SUMMARY OUTPUT								
Regression Statistics								
Multiple R	0.0161							
R Square	0.000259							
Adjusted R Square	-0.00228							
Standard Error	4.586673							
Observations	395							
ANOVA								
	df	*SS*	*MS*	*F*	*Significance F*			
Regression	1	2.143564	2.143564	0.101892	0.74974			
Residual	393	8267.765	21.03757					
Total	394	8269.909						
	Coefficients	*Standard Error*	*t Stat*	*P-value*	*Lower 95%*	*Upper 95%*	*Lower 95.0%*	*Upper 95.0%*
Intercept	10.34021	0.329304	31.40019	3.6E-109	9.692789	10.98762	9.692789	10.98762
Activities	0.147356	0.461634	0.319206	0.74974	-0.76022	1.054936	-0.76022	1.054936

Table 3.74: a simple regression model with extracurricular activities and final grade.

p-value of the activities (extracurricular activities) variable is 0.74974 which is greater than the threshold value (0.05). So, it is not statistically significant.

The lower bound of the confidence interval (Lower 95%) for the activity's variable is -0.76022 and the upper bound (Upper 95%) is 1.054936. Zero falls within this confidence interval. So, the activities variable is not statistically significant.

Nursery and the Final Grade:

Regression between independent variable Nursery and the Final Grade (G3) dependent variable:

SUMMARY OUTPUT								
Regression Statistics								
Multiple R	0.051568							
R Square	0.002659							
Adjusted R Square	0.000121							
Standard Error	4.581164							
Observations	395							
ANOVA								
	df	*SS*	*MS*	*F*	*Significance F*			
Regression	1	21.99174	21.99174	1.047871	0.306628			
Residual	393	8247.917	20.98707					
Total	394	8269.909						
	Coefficients	*Standard Error*	*t Stat*	*P-value*	*Lower 95%*	*Upper 95%*	*Lower 95.0%*	*Upper 95.0%*
Intercept	9.950617	0.509018	19.54865	6.18E-60	8.949878	10.95136	8.949878	10.95136
Nursery	0.584415	0.570909	1.023656	0.306628	-0.538	1.706833	-0.538	1.706833

Table 3.75: a simple regression model with attended nursery school and final grade.

p-value of the Nursery variable is 0.306628 which is greater than the threshold value (0.05). So, it is not statistically significant.

The lower bound of the confidence interval (Lower 95%) for the nursery variable is -0.538 and the upper bound (Upper 95%) is 1.706833. Zero falls within this confidence interval. So, the Nursery variable is not statistically significant.

Want to take higher education and the Final Grade:

Regression between independent variable Higher and the Final Grade (G3) dependent variable:

SUMMARY OUTPUT								
Regression Statistics								
Multiple R	0.182465							
R Square	0.033293							
Adjusted R Square	0.030834							
Standard Error	4.510259							
Observations	395							
ANOVA								
	df	*SS*	*MS*	*F*	*Significance F*			
Regression	1	275.3329	275.3329	13.5349	0.000267			
Residual	393	7994.576	20.34243					
Total	394	8269.909						
	Coefficients	*Standard Error*	*t Stat*	*P-value*	*Lower 95%*	*Upper 95%*	*Lower 95.0%*	*Upper 95.0%*
Intercept	6.8	1.008524	6.742523	5.57E-11	4.817222	8.782778	4.817222	8.782778
Higher	3.808	1.035069	3.678981	0.000267	1.773035	5.842965	1.773035	5.842965

Table 3.76: a simple regression model with wants to take higher education and final grade.

p-value of the higher variable (wants to take higher education) is 0.000267 which is very smaller than the threshold value (0.05). So, it is statistically significant.

The lower bound of the confidence interval (Lower 95%) for the higher variable is 1.773035 and the upper bound (Upper 95%) is 5.842965. Zero does not fall within this confidence interval. So, the higher variable is statistically significant.

Internet access at home and Final Grade:

Regression between independent variable Internet and the Final Grade (G3) dependent variable:

SUMMARY OUTPUT								
Regression Statistics								
Multiple R	0.098483							
R Square	0.009699							
Adjusted R Square	0.007179							
Standard Error	4.564968							
Observations	395							
ANOVA								
	df	*SS*	*MS*	*F*	*Significance F*			
Regression	1	80.20963	80.20963	3.849029	0.05048			
Residual	393	8189.699	20.83893					
Total	394	8269.909						
	Coefficients	*Standard Error*	*t Stat*	*P-value*	*Lower 95%*	*Upper 95%*	*Lower 95.0%*	*Upper 95.0%*
Intercept	9.409091	0.561909	16.74488	6.87E-48	8.304368	10.51381	8.304368	10.51381
Internet	1.20793	0.615696	1.961894	0.05048	-0.00254	2.4184	-0.00254	2.4184

Table 3.77: A simple regression model with internet access at home and final grade.

p-value of the Internet variable is 0.05048 which is slightly greater th-0.00254an the threshold value (0.05). So, it is not statistically significant.

The lower bound of the confidence Interval (Lower 95%) for the internet variable is -0.00254 and the upper bound (Upper 95%) is 2.4184. Zero falls within this confidence interval. So, the internet variable is not statistically significant.

In a romantic relationship and the final grade:

Regression between independent variable romantic and the final grade (G3) dependent variable:

SUMMARY OUTPUT								
Regression Statistics								
Multiple R	0.12997							
R Square	0.016892							
Adjusted R Square	0.014391							
Standard Error	4.548358							
Observations	395							
ANOVA								
	df	*SS*	*MS*	*F*	*Significance F*			
Regression	1	139.6969	139.6969	6.752698	0.009713			
Residual	393	8130.212	20.68756					
Total	394	8269.909						
	Coefficients	*Standard Error*	*t Stat*	*P-value*	*Lower 95%*	*Upper 95%*	*Lower 95.0%*	*Upper 95.0%*
Intercept	10.8365	0.280464	38.63779	6.5E-136	10.2851	11.3879	10.2851	11.3879
Romantic	-1.26074	0.485164	-2.5986	0.009713	-2.21459	-0.3069	- 2.21459	-0.3069

Table 3.78: a simple regression model with in a romantic relationship and final grade.

p-value of the romantic variable is 0.009713 which is smaller than the threshold value (0.05). So, it is statistically significant.
The lower bound of the confidence interval (Lower 95%) for the romantic variable is -2.21459 and the upper bound (Upper 95%) is -0.3069. Zero does not fall within this confidence interval. So, the romantic variable is statistically significant.

Quality of family relationship and the final grade:

Regression between independent variable famrel and the Final Grade (G3) dependent variable:

SUMMARY OUTPUT								
Regression Statistics								
Multiple R	0.051363							
R Square	0.002638							
Adjusted R Square	0.0001							
Standard Error	4.581213							
Observations	395							
ANOVA								
	df	*SS*	*MS*	*F*	*Significance F*			
Regression	1	21.81769	21.81769	1.039556	0.308552			
Residual	393	8248.091	20.98751					
Total	394	8269.909						
	Coefficients	*Standard Error*	*t Stat*	*P-value*	*Lower 95%*	*Upper 95%*	*Lower 95.0%*	*Upper 95.0%*
Intercept	9.380049	1.041094	9.009796	9.09E-18	7.333238	11.42686	7.333238	11.42686
famrel	0.262439	0.257398	1.019586	0.308552	-0.24361	0.768489	-0.24361	0.768489

Table 3.79: a simple regression model with quality of family relationship and final grade.

p-value of the famrel (quality of family relationship) variable is 0.308552 which is greater than the threshold value (0.05). So, it is not statistically significant.

The lower bound of the confidence interval (Lower 95%) for the variable famrel is -0.24361 and the upper bound (Upper 95%) is 0.768489 . Zero falls within this confidence interval. So, the famrel variable is not statistically significant.

Free time and final grade:

Regression between independent variable freetime and the Final Grade (G3) dependent variable:

SUMMARY OUTPUT								
Regression Statistics								
Multiple R	0.011307							
R Square	0.000128							
Adjusted R Square	-0.00242							
Standard Error	4.586974							
Observations	395							
ANOVA								
	df	*SS*	*MS*	*F*	*Significance F*			
Regression	1	1.057338	1.057338	0.050253	0.82274			
Residual	393	8268.852	21.04033					
Total	394	8269.909						
	Coefficients	*Standard Error*	*t Stat*	*P-value*	*Lower 95%*	*Upper 95%*	*Lower 95.0%*	*Upper 95.0%*
Intercept	10.24739	0.783299	13.08236	1.02E-32	8.707412	11.78737	8.707412	11.78737
freetime	0.051862	0.231352	0.224172	0.82274	-0.40298	0.506704	-0.40298	0.506704

Table 3.80: a simple regression model with free time after school and final grade.

p-value of the freetime variable (free time after school) is 0.82274 which is greater than the threshold value (0.05). So, it is not statistically significant.

The lower bound of the confidence interval (Lower 95%) for the freetime variable is -0.40298 and the upper bound (Upper 95%) is 0.506704. Zero falls within this confidence interval. So, the freetime variable is not statistically significant.

Going out with friends and final grade:

Regression between independent variable goout and the Final Grade (G3) dependent variable:

SUMMARY OUTPUT								
Regression Statistics								
Multiple R	0.132791							
R Square	0.017634							
Adjusted R Square	0.015134							
Standard Error	4.546643							
Observations	395							
ANOVA								
	df	*SS*	*MS*	*F*	*Significance F*			
Regression	1	145.8281	145.8281	7.054389	0.008229			
Residual	393	8124.081	20.67196					
Total	394	8269.909						
	Coefficients	*Standard Error*	*t Stat*	*P-value*	*Lower 95%*	*Upper 95%*	*Lower 95.0%*	*Upper 95.0%*
Intercept	12.1141	0.679325	17.83256	1.53E-52	10.77853	13.44966	10.77853	13.44966
goout	-0.54647	0.20575	-2.65601	0.008229	-0.95098	-0.14197	-0.95098	-0.14197

Table 3.81: a simple regression model with going out with friends and final grade.

p-value of the goout variable (going out with friends) is 0.008229 which is smaller than the threshold value (0.05). So, it is statistically significant.

The lower bound of the confidence interval (Lower 95%) for the goout variable is -0.95098 and the upper bound (Upper 95%) is -0.14197. Zero does not fall within this confidence interval. So, the goout variable is statistically significant.

Working day alcohol consumption and final grade:

Regression between independent variable Dalc and the Final Grade (G3) dependent variable:

SUMMARY OUTPUT								
Regression Statistics								
Multiple R	0.05466							
R Square	0.002988							
Adjusted R Square	0.000451							
Standard Error	4.58041							
Observations	395							
ANOVA								
	df	*SS*	*MS*	*F*	*Significance F*			
Regression	1	24.70817	24.70817	1.177693	0.278491			
Residual	393	8245.201	20.98015					
Total	394	8269.909						
	Coefficients	*Standard Error*	*t Stat*	*P-value*	*Lower 95%*	*Upper 95%*	*Lower 95.0%*	*Upper 95.0%*
Intercept	10.83156	0.447572	24.20071	7.13E-80	9.951625	11.71149	9.951625	11.71149
Dalc	-0.28114	0.259063	-1.08522	0.278491	-0.79046	0.228183	-0.79046	0.228183

Table 3.82: a simple regression model with weekday alcohol consumption and final grade.

p-value of the Dalc variable (weekday alcohol consumption) is 0.278491 which is greater than the threshold value (0.05). So, it is not statistically significant.

The lower bound of the confidence interval (Lower 95%) for the Dalc variable is -0.79046 and the upper bound (Upper 95%) is 0.228183. Zero falls within this confidence interval. So, the Dalc variable is not statistically significant.

Weekend alcohol consumption and final grade:

Regression between independent variable Walc and the Final Grade (G3) dependent variable:

SUMMARY OUTPUT								
Regression Statistics								
Multiple R	0.051939							
R Square	0.002698							
Adjusted R Square	0.00016							
Standard Error	4.581076							
Observations	395							
ANOVA								
	df	*SS*	*MS*	*F*	*Significance F*			
Regression	1	22.30968	22.30968	1.063061	0.303152			
Residual	393	8247.599	20.98626					
Total	394	8269.909						
	Coefficients	*Standard Error*	*t Stat*	*P-value*	*Lower 95%*	*Upper 95%*	*Lower 95.0%*	*Upper 95.0%*
Intercept	10.83851	0.47085	23.01904	7.44E-75	9.912811	11.76421	9.912811	11.76421
Walc	-0.18476	0.1792	-1.03105	0.303152	-0.53707	0.167547	-0.53707	0.167547

Table 3.83: a simple regression model with weekend alcohol consumption and final grade.

p-value of the Walc variable (weekend alcohol consumption) is 0.303152 which is greater than the threshold value (0.05). So, it is not statistically significant.
The lower bound of the confidence interval (Lower 95%) for the Walc variable is -0.53707 and the upper bound (Upper 95%) is 0.167547. Zero falls within this confidence interval.
So, the Walc variable is not statistically significant.

Student health status and final grade:

Regression between independent variable health and the Final Grade (G3) dependent variable:

SUMMARY OUTPUT								
Regression Statistics								
Multiple R	0.061335							
R Square	0.003762							
Adjusted R Square	0.001227							
Standard Error	4.578631							
Observations	395							
ANOVA								
	df	*SS*	*MS*	*F*	*Significance F*			
Regression	1	31.11085	31.11085	1.484023	0.223877			
Residual	393	8238.798	20.96386					
Total	394	8269.909						
	Coefficients	*Standard Error*	*t Stat*	*P-value*	*Lower 95%*	*Upper 95%*	*Lower 95.0%*	*Upper 95.0%*
Intercept	11.13359	0.633124	17.58517	1.76E-51	9.888859	12.37833	9.888859	12.37833
health	-0.20211	0.165912	-1.2182	0.223877	-0.5283	0.124071	-0.5283	0.124071

Table 3.84: a simple regression model with student's current health status and final grade.

p-value of the health variable is 0.223877 which is greater than the threshold value (0.05).
So, it is not statistically significant.
The lower bound of the confidence interval (Lower 95%) for the health variable is -0.5283 and the upper bound (Upper 95%) is 0.124071. Zero falls within this confidence interval. So, the health variable is not statistically significant.

Number of absence and final grade:

Regression between independent variable absence and the Final Grade (G3) dependent variable:

SUMMARY OUTPUT								
Regression Statistics								
Multiple R	0.034247							
R Square	0.001173							
Adjusted R Square	-0.00137							
Standard Error	4.584577							
Observations	395							
ANOVA								
	df	*SS*	*MS*	*F*	*Significance F*			
Regression	1	9.6996	9.6996	0.461483	0.497332			
Residual	393	8260.209	21.01834					
Total	394	8269.909						
	Coefficients	*Standard Error*	*t Stat*	*P-value*	*Lower 95%*	*Upper 95%*	*Lower 95.0%*	*Upper 95.0%*
Intercept	10.30327	0.283471	36.34683	9.4E-128	9.745958	10.86058	9.745958	10.86058
absences	0.019605	0.02886	0.679325	0.497332	-0.03713	0.076344	-0.03713	0.076344

Table 3.85: a simple regression model with number of absence and final grade.

p-value of the absence variable is 0.497332 which is greater than the threshold value (0.05). So, it is not statistically significant.
The lower bound of the confidence interval (Lower 95%) for the absence variable is -0.03713 and the upper bound (Upper 95%) is 0.076344. Zero falls within this confidence interval. So, the absence variable is not statistically significant.

First period grade and final grade:

Regression between First period grade and final grade.

SUMMARY OUTPUT								
Regression Statistics								
Multiple R	0.801468							
R Square	0.642351							
Adjusted R Square	0.641441							
Standard Error	2.743359							
Observations	395							
ANOVA								
	df	*SS*	*MS*	*F*	*Significance F*			
Regression	1	5312.183	5312.183	705.8422	9E-90			
Residual	393	2957.726	7.52602					
Total	394	8269.909						
	Coefficients	*Standard Error*	*t Stat*	*P-value*	*Lower 95%*	*Upper 95%*	*Lower 95.0%*	*Upper 95.0%*
Intercept	-1.6528	0.474745	-3.48145	0.000555	-2.58616	-0.71945	-2.58616	-0.71945
G1	1.106256	0.041639	26.56769	9E-90	1.024393	1.188119	1.024393	1.188119

Table 3.86: a simple regression model with First period grade and final grade.

p-value of the G1 variable is 9E-90 which is smaller than the threshold value (0.05). So, it is statistically significant.

The lower bound of the confidence interval (Lower 95%) for the G1 variable is 1.024393 and the upper bound (Upper 95%) is 1.188119. Zero does not fall within this confidence interval. So, the G1 variable is statistically significant.

By the R-squared value, this simple regression model with first period grade variable explains about almost 64% of the data.

Second period grade and final grade:

Regression between second period grade and final grade.

SUMMARY OUTPUT								
Regression Statistics								
Multiple R	0.904868							
R Square	0.818786							
Adjusted R Square	0.818325							
Standard Error	1.952764							
Observations	395							
ANOVA								
	df	*SS*	*MS*	*F*	*Significance F*			
Regression	1	6771.286	6771.286	1775.708	7.6E-148			
Residual	393	1498.623	3.813289					
Total	394	8269.909						
	Coefficients	*Standard Error*	*t Stat*	*P-value*	*Lower 95%*	*Upper 95%*	*Lower 95.0%*	*Upper 95.0%*
Intercept	-1.39276	0.29694	-4.69037	3.77E-06	-1.97655	-0.80897	-1.97655	-0.80897
G2	1.102112	0.026154	42.13915	7.6E-148	1.050693	1.153532	1.050693	1.153532

Table 3.87: a simple regression model with second period grade and final grade.

p-value of the G2 variable is 7.6E-148 which is smaller than the threshold value (0.05). So, it is statistically significant.

The lower bound of the confidence interval (Lower 95%) for the G2 variable is 1.050693 and the upper bound (Upper 95%) is 1.050693. Zero does not fall within this confidence interval. So, the G2 variable is statistically significant.

By the R-squared value, this simple regression model with second period grade variable explains about almost 81 % of the data.

Multiple Linear Regression 1:

Multiple Linear Regression to final grade based on the explanatory variables of the students: Sex, age, Address, Mother's_education, Father's Education, traveltime, failures, Paid, Higher, Romantic and goout.

SUMMARY OUTPUT								
Regression Statistics								
Multiple R	0.441024854							
R Square	0.194502922							
Adjusted R Square	0.171368541							
Standard Error	4.170450391							
Observations	395							
ANOVA								
	df	*SS*	*MS*	*F*	*Significance F*			
Regression	11	1608.521436	146.2292215	8.407526578	2.71203E-13			
Residual	383	6661.387425	17.39265646					
Total	394	8269.908861						
	Coefficients	*Standard Error*	*t Stat*	*P-value*	*Lower 95%*	*Upper 95%*	*Lower 95.0%*	*Upper 95.0%*
Intercept	10.38337314	3.458882095	3.001944807	0.002858446	3.582598061	17.18414822	3.582598061	17.18414822
Sex01	1.141672244	0.43643876	2.615881881	0.009251642	0.283556309	1.999788179	0.283556309	1.999788179
age	-0.06085791	0.177601884	-0.34266478	0.732038643	-0.410054683	0.288338862	-0.410054683	0.288338862
Address01	0.617873928	0.543848729	1.136113581	0.256619419	-0.45142904	1.687176895	-0.45142904	1.687176895

Mother's_education	0.562586612	0.253338679	2.220689768	0.026956442	0.06447788	1.060695343	0.06447788	1.060695343
Father's Education	-0.12163885	0.251093574	-0.484436332	0.628353149	-0.615333308	0.372055608	-0.615333308	0.372055608
traveltime	-0.313891339	0.324464129	-0.967414611	0.333947289	-0.951845309	0.324062631	-0.951845309	0.324062631
failures	-1.725434381	0.314055696	5.49403944	7.17684E-08	-2.342923527	-1.107945234	-2.342923527	-1.107945234
Paid	0.256497034	0.43986885	0.583121614	0.560154619	-0.608363067	1.121357134	-0.608363067	1.121357134
Higher	1.480567788	1.045359778	1.416323661	0.157493538	-0.574794764	3.53593034	-0.574794764	3.53593034
Romantic	-0.817647517	0.458316942	-1.784022023	0.075211668	-1.718779834	0.083484799	-1.718779834	0.083484799
goout	-0.46048654	0.194054349	-2.372977164	0.018138291	-0.842031775	-0.078941304	-0.842031775	-0.078941304

Table 3.88: Multiple Linear Regression to final grade based on the explanatory variables of the students: Sex, age, Address, Mother's_education, Father's Education, traveltime, failures, Paid, Higher, Romantic and goout.

By the R-squared value, this multiple linear regression model explains about 19 % of the data.

Multiple Linear Regression 2:

Multiple Linear Regression to final grade based on the explanatory variables of the students: Sex, age, Address, Mother's_education, Father's Education, traveltime, failures, Paid, Higher, Romantic, goout, G1 and G2

SUMMARY OUTPUT								
Regression Statistics								

Multiple R	0.909495943							
R Square	0.827182871							
Adjusted R Square	0.821286224							
Standard Error	1.936784347							
Observations	395							
ANOVA								
	df	SS	MS	F	Significance F			
Regression	13	6840.726956	526.2097659	140.2801982	2.693E-136			
Residual	381	1429.181905	3.751133608					
Total	394	8269.908861						
	Coefficients	Standard Error	t Stat	P-value	Lower 95%	Upper 95%	Lower 95.0%	Upper 95.0%
Intercept	-0.172233488	1.633271123	-0.105453091	0.916071786	-3.383587354	3.039120377	-3.383587354	3.039120377
Sex01	0.148837931	0.204510241	0.727777398	0.467196867	-0.253272132	0.550947994	-0.253272132	0.550947994
age	-0.113947858	0.083276317	-1.368310492	0.172021254	-0.277686578	0.049790861	-0.277686578	0.049790861
Address01	-0.068242183	0.253447222	-0.269255992	0.787878451	-0.566572622	0.430088256	-0.566572622	0.430088256
Mother's_education	0.127596912	0.118237344	1.079159156	0.281199485	-0.104882525	0.360076349	-0.104882525	0.360076349
Father's Education	-0.130530848	0.117001829	-1.115630832	0.265283342	-0.360581003	0.099519307	-0.360581003	0.099519307

traveltime	0.117549849	0.151501125	0.775900833	0.438288899	-0.180333164	0.415432861	-0.180333164	0.415432861
failures	-0.187780073	0.152357377	-1.232497412	0.218523334	-0.487346658	0.111786511	-0.487346658	0.111786511
Paid	0.074635274	0.205964936	0.362368835	0.717277341	-0.330335025	0.479605573	-0.330335025	0.479605573
Higher	0.108584134	0.487840147	0.222581381	0.8239806	-0.850611998	1.067780265	-0.850611998	1.067780265
Romantic	-0.277570763	0.214703672	-1.292808645	0.196860661	-0.69972325	0.144581724	-0.69972325	0.144581724
goout	0.077513052	0.091277493	0.849202245	0.396302022	-0.10195766	0.256983764	-0.10195766	0.256983764
G1	0.164188814	0.058423597	2.810316766	0.005204709	0.049315759	0.27906187	0.049315759	0.27906187
G2	0.958260544	0.052080231	18.39969853	1.47864E-54	0.855859878	1.060661209	0.855859878	1.060661209

Table 3.89: A multiple linear regression with G1 and G2

By the R-squared value, this multiple linear regression model explains about almost 82 % of the data.
Out of all the variables only **s**ex, age, Address, Mother's_education, Father's Education, traveltime, failures, Paid, Higher, Romantic, goout, G1 and G2 are statistically significant.
And variables school, famsize, Pstatus, schoolsup, famsup, activities, nursery, internet, famrel, Dalc, Walc, health, absence are not statistically significant.

Interpret R-squared/adjusted R-squared:

Let's see how well our model fits the data. For this reason, we need to interpret R-squared and adjusted R-squared.
Generally, in Multiple Linear Regression when we have more than one explanatory/Independent variables, it is more preferable to use the adjusted R-square than the R-squared. Adjusted R-square considers the number of independent variables in the regression model. It can deliver a more accurate view of the correlation.

By the adjusted R-square, that means controlling the number of independent variables in the regression model, the variability of the features of the students explains almost 82% of the variability of the final grade. By the R-squared value, our regression model explains about 82% of the data. G1 and G2 has very high influence on the model.

In the simple regression models, the highest R-squared value was 81% with the G2 second period grade variable. But in the Multiple Linear Regression Model R-squared value is around almost 83%. Which means the multiple regression model adds about 2 % points of explanatory power compared to the simple regression model. Multiple Linear regression has more/better explanatory power.

4. Trends, Patterns and Anomalies

This section presents the patterns and trends that I have identified from the EDA presented in Section 3. Below I summarize the identified patterns and trends.

Summary of students' demographic
88% students from the dataset are from the GP Gabriel Pereira school (Figure 3.1). Female students are around 6% higher than the Male student (Figure 3.2) and most of the student are in between 15 to 18 years (Figure 3.3). Most of the student are living in the urban areas (almost 78%, Figure 3.4). 83% student have internet access at home (Figure 3.22). Urban students have more internet access at home than the rural student (Figure 3.67). Also, students from urban areas are more likely to choose the nearby school than the rural students (Figure 3.69).

Patterns in students' family background
Almost 90% of the students live with their both parents (i.e., parents are living together) (see Figure 3.6). About 71% of the students' family size is greater than 3 (Figure 3.5). Mother of the students are more like to stay at homes than father (15% of the students' mother stay at home, whereas only 5% of the fathers stay at home) (Figure 3.9 and Figure 3.10). However, it is interesting to see that, mothers with higher education are higher in number (almost 33%) than fathers (24%) (see Figure 3.7). Most of the students' fathers' education is between 5th to 9th grade (Figure 3.8). More mothers have higher educated than the fathers (Figure 3.7 and Figure 3.8). The number of mothers in teaching and health professions are greater than the number of fathers (Figure 3.10).

Patters in terms of school
The data shows that course preference is the most important factor to choose the school (see Figure 3.11). Secondly, there is a high tendency to choose a school that is the nearest from their home. Most of the student live with in a very short distance from the school (travel time <15 min) (see Figure 3.13). So, the travel time distribution is right skewed. Lastly, another pattern is that for majority of the students (almost 69%) their mother is registered as their guardian (see Figure 3.12).

Patterns in students' study habits

The data shows that most of the students involved in this study spend 2 to 5 hours in a week for study (Figure 3.14). Only 46% student have extra paid classes within the course subject (Figure 3.18). The number of students who have failed before in previous class is relatively small, around 21% (Figure 3.15). In terms of support, only 13% students have extra educational support from the school (Figure 3.16). However, majority of the students (61% student) have family education support (Figure 3.17). Almost half of the students are involved in extracurricular activities.

Other than that, majority of the student (almost 80%) had attended nursery school and 95% of them want to take higher education (Figure 3.20 and Figure 3.21).

Patterns in students' personal and family relationships

Majority of the students (93%) students have positive relationship with their family. There are very few students in the dataset that have bad or very bad family relationship (Figure 3.25). Moreover, only one third of the students are engaged in any romantic relationship (Figure 3.23). I have found a positive correlation between romantic relationship and age (with value 0.165); which means that older students are more likely to have a romantic relationship (Figure 3.66).

Patterns in students' social activities

Overall, alcohol consumption either on weekdays or on weekends, has a very low negative influence on final grade. I have found negative correlation between alcohol consumption and final grade with value -0.05466 for weekdays and with value -0.05194 on weekends (Figure 3.106 and Figure 3.107).

Compared to the weekends, students tend to drink far less during the workdays. Moreover, majority of the students drink very low alcohol in any workday. I have found the weekend alcohol consumption to be higher than the workdays consumption. In addition, the free time after school and the habit of going out with friends almost looks like a normal distribution. In both cases, the distribution is right skewed (Figure 3.26 and Figure 3.27). However, the students who going out more with friends tends to consume more alcohol.

Patterns in student's health status

Overall, the students seem to have good health conditions. The data shows that 83% students are very attentive in the school, which means less absence due to any possible health reasons (Figure 3.31). Only 23% of the students have health conditions that can be considered as bad or very bad.

Student's grade patterns

The grade distribution for all G1 (first period grade), G2(second period grade) and G3 (final grade) are normally distributed. That means majority of the students get an average grade. The data shows that 67% students have passed in the final grade and less than one third of the students have failed in the final grade (Figure 3.35). And the students who have passed in final grade, majority of them have got a grade between 10 to 13 (Figure 3.36).

Patterns based on specific schools

I have found some patterns based on specific schools. Firstly, the students at Mousinho da Silveire school tend to be older (Figure 3.37). Students of Gabriel Pereira are more likely to choose school near the home than students of Mousinho da Silveire (Figure 3.39). More than 80% of the Gabriel Pereira school's students are from the urban area (Figure 3.38). And lastly none of the students of Mousinho da Silveire have extra educational support (Figure 3.40).

Patterns based on students' age

I have found a positive correlation between failure and age with value 0.24 (Figure 3.61). Younger students tend to do better in the exam then the older students. On other words, older students are more likely to fail than the younger ones. Secondly, younger students tend to want higher education more than the older students (Figure 3.63). In addition, more younger students take extra educational school support than the older (among those who are taking it).

Patterns based on gender

There are multiple patterns in the data related to the gender of the students. There are few aspects where the male students were better than the female students. Firstly, in getting 'good' and 'excellent' grade, male students are doing better than the female (Figure 3.44). Passing ratio is also slightly higher in the male students than the female student (Figure 3.45). Male students were more active in extracurricular activities than the female (Figure 3.53). Male students' health condition was better than the female, even though male students showed a tendency to consume more alcohol comparing to the female (Figure 3.55).

On contrary, there are few aspects where the female students were better compared to the male. For example, female students' trend to study more hours than the male students (Figure 3.47). Female students received more extra educational school support than the male students (Figure 3.48). Female students are more interested to take higher education than the male students (Figure 3.51). Female students are slightly (around 15%) receiving more extra paid classes within the course subject (Math and Portuguese) than the male students (Figure 3.50). More female students have family education support than the male student (Figure 3.50). Lastly, female students are trends to be younger than the male students (Figure 3.42).

Patterns based on parents education

The data shows that parents (both father and mother) education play an important role in children's education. There are positive correlations between parent's education level (applies to both father and mother) and the students' final grade. Students whose mother or father has higher education trends to go to school in an earlier age i.e., children with more educated mother or father attends nursery schools more (Figure 3.59 and Figure 3.60).

I have seen multiple patterns related to the education background of the student's mother. For example, mothers with more education have more family educational support for their children. The percentage of children's willingness for higher education increases when the mother's education level increases (Figure 3.75). Children with more educated mother has more internet access at home (Figure 3.76). Highly educated

mother trends to live near their child's school. Or we can say they choose the school for their children which is near the home (Figure 3.71).

I have seen some similar patterns based on father education as well. Fathers with more education have more family educational support for their children. I have also found a positive correlation between father's education and children's willingness to take higher education (Figure 3.79). However, mother's educational background seems to have more influence on children's education.

Patterns based on study habits

The data shows a negative correlation between the study time and the failure (with value -0.174) and a positive correlation between weekly study time and final grade (with a value 0.098). This indicates that students who study more are less likely to fail and that students who tend to spend more time on study are more likely to get a better grade in final exam (Figure 3.98 and Figure 3.81). Nonetheless, it is important to note that in both cases the correlation value is less significant. There is also a high trend in the data that student who did well in the past exams also going to do well in the following examinations (Figure 3.99). There are multiple factors that have a negative correlation with the final grade of the students. These factors include students' health, past record of failing, extra educational school support, and school traveling time.

Moreover, the data shows that even though the school are giving extra educational school support to the weak students, many of them have not passed in the final grade. Thus, there is a little improvement in the final grade (Figure 3.100). However, I have found a positive correlation between willingness of higher education and final grade (with value 0.182) and between extra paid classes and willingness to take higher education (with value 0.189). This indicated that students who wants to take higher education trend to get do better in final grade (Figure 3.103). Similarly, students' attitude towards higher education correlates with the amount of their study time (Figure 3.83).

Lastly, quality of family relationships of the students has very little influence on final grade, with a correlation value of 0.051 (Figure 3.104). Also, the students who study for more hours tend to consume lesser alcohol in the weekdays and weekend. There is a negative correlation between the students' study time and weekend alcohol consumption (with value -0.254).

5. Discussion

As presented in Section 2, I initially assumed that children's educational performance will be highly influenced by their family education background. After analyzing the data, I have found multiple positive correlations between family education background and student's performance at school. For example, if a student's father or mother has a higher education, then that student tends to start school earlier and those students also have more interest for higher education. As shown in Figure 3.95, mother's education level also positively impacts the student's final grade. Moreover, as presented in Table 3.66, a simple regression model with Mother's education (independent variable) and final grade (dependent variable) showed that the mother's education

(medu) variable is statistically significant. Thus, it can be said that my assumption proved to be true as per the analysis.

My second assumption was that physical energy may impact students' educational performance. So, home to school travel time can have an impact on a student's performance. Similarly, students' health condition can also influence the final grade. After analyzing the data, I have found that there is a negative correlation between home to school travel time and final grade with value -0.11714 (Figure 3.97). Additionally, as presented in Table 3.68, a simple regression model with home to school travel time (independent variable) and final grade (dependent variable) showed that the home to school travel time (traveltime) variable is statistically significant. Thus, it proves my assumption to be true.

My third assumption was also proved as true. As per the analysis there is a negative correlation between number of past class failures and final grade with a value -0.36041 (Figure 3.99). There is a high trend in the data that student who did well in the past also doing better in the following examinations. The correlation between G1 and G2 is 0.85 and correlation between G1 and G3 is 0.8 .The correlation between G2 and G3 is 0.904 (Figure 3.99 , Table 3.59). Moreover, as presented in Table 3.70 (failures and final grade), Table 3.86 (first period grade and final grade) and in Table 3.87 (second period grade and final grade), all independent variables (failures, first period grade and second period grade) are statistically significant in their regression models with final grade as the dependent variable.

My fourth assumption was partially true. I found a positive correlation between extra paid classes and final grade with value 0.101996 (Table 3.59). Additionally, as presented in Table 3.73, a simple regression model with extra paid classes (independent variable) and final grade (dependent variable) showed that extra paid classes (paid) variable is statistically significant.

My assumption about the family educational support proved to be false. I assumed that it would have a positive impact on final grade. But the analysis resulted in a negative correlation between family educational support and final grade with value -0.039 (Table 3.59). Moreover, as presented in Table 3.72, a simple regression model with family educational support (independent variable) and final grade (dependent variable) showed that family educational support (famsup) variable is not statistically significant.

There is a positive correlation between internet access at home and the final grade with vale 0.098 (Table 3.59). Besides, as presented in Table 3.77, a simple regression model with internet access at home (independent variable) and final grade (dependent variable) showed that internet access at home (internet) variable is not statistically significant.

My fifth assumption was true. There is a positive correlation between willingness to take higher education and the final grade with value 0.18 (Table 3.59). Additionally, as presented in Table 3.76, a simple regression model with willingness to take higher education (independent variable) and final grade (dependent variable) showed that willingness to take higher education (Higher) variable is statistically significant.

My sixth assumption was again partially true. There was a positive correlation between weekly study time and final grade with a value 0.098 (Table 3.59). In a simple regression model with weekly study time and final grade, p-value of the studytime (weekly study time) variable is 0.052061 which is very slightly greater than the threshold value (0.05) (Table 3.69). So, it is not statistically significant. The lower bound of the confidence interval (Lower 95%) for the studytime variable is -0.00479 and the upper bound (Upper 95%) is 1.072791. Zero falls within this confidence interval. So, the study time variable is not statistically significant.

My seventh assumption was, factors such as romantic relationship, social activities like going out with friends will have no or less impact on the grade. However, I assumed that alcohol consumption can have a negative impact on the final grade. The analysis resulted in a negative correlation between in a romantic relationship and final grade with value -0.1299 (Table 3.59). Moreover, as presented in Table 3.78, a simple regression model with in romantic relationship (independent variable) and final grade (dependent variable) showed that a romantic relationship (romantic) variable is statistically significant. So, my assumption was true.

But my assumption about the alcohol consumption was not completely true. There is a negative correlation between working day alcohol consumption and final grade with value -0.0546 (Table 3.59), which is very small. Besides, as presented in Table 3.82, a simple regression model with working day alcohol consumption (independent variable) and final grade (dependent variable) showed that working day alcohol consumption (Dalc) variable is not statistically significant.

There is a negative correlation between weekend alcohol consumption and final grade with value -0.0519 (Table 3.59), which is also very small. Besides, as presented in Table 3.83, a simple regression model with weekend alcohol consumption (independent variable) and final grade (dependent variable) showed that weekend alcohol consumption (Walc) variable is not statistically significant.

6. Conclusion

In this project, I have done Exploratory data analysis (EDA), correlation, and regression analysis (both single and multiple regression) on Student-mat data in order to understand why and how some variables (e.g., the reason to choose the school, parent's job, parent's education, or student's alcohol consumption etc.) affect the students' results. Thus, in my project, I have explored different variables/ features, the relation between different variables and why and how they impact the final grade.

I have predicted the students' final grades by using students' past school grades, social, demographic, and other student-related data by different regression model from Table 3.60 to Table 3.89. Out of all the variables only sex, age, address, mother's_education, father's Education, traveltime, failures, paid, higher, romantic, goout, G1 and G2 are statistically significant in simple regression model with the final grade. And variables school, famsize, Pstatus, schoolsup, famsup, activities, nursery, internet, famrel, Dalc, Walc, health, absence are not statistically significant in simple regression model with the final grade. In multiple Linear Regression model to final grade based on the explanatory variables of the students' Sex, age, Address, Mother's_education, Father's Education, traveltime, failures, paid, higher, romantic, goout, G1 and G2 this regression model explains about almost 82 % of the data.

I found that student's family educational background is correlated to their performance at school. In terms of failure and success, student previous performance is also highly correlated to their future performance,. Extra efforts like extra support from school and extra paid class and also internet access at home have little impact on the student's final grade. If the student study more, they will tend to do little better in the exam. It seems like other variables have more influence over grade. Romantic relationship, social activities like going out with

friends will have no or less negative impact on the final grade. Alcohol consumption have very little negative impact on the final grade.

The dataset we received for the project was of very good quality. There were no missing values or duplicate values in the data. It seems like the data was already being cleaned and a little processed. I just needed to handle the categorical (binary) data variables. I converted them into binary (0 and 1) values. I have performed different statistical operations, correlations, and regressions to find the trends, patterns, and anomalies in the data to verify whether my initial assumptions and hypotheses are held or not. It was a good learning opportunity for me. I was thrilled to discover new patterns and trends, which I did not expect in the beginning, and when I found that in some cases, my assumption was wrong, especially when I saw very little negative influence of alcohol on students' grades. I assumed it to be significant. Overall, I enjoyed this project work, and doing this project helped me to reflect and revise my learning from this course throughout the semester.

7. References

- Cortez, Paulo, and Alice Maria Gonçalves Silva. "Using data mining to predict secondary school student performance." (2008).
- Kotsiantis, Sotiris, Christos Pierrakeas, and Panagiotis Pintelas. "Predicting students' performance in distance learning using machine learning techniques." *Applied Artificial Intelligence* 18.5 (2004): 411-426.