

# Student Performance Data Info

## Attributes

1 school - student's school (Categorical: 'GP' - Gabriel Pereira or 'MS' - Mousinho da Silveira)

2 sex - student's sex (Categorical: 'F' - female or 'M' - male)

3 age - student's age (numeric: from 15 to 22)

4 address - student's home address type (categorical: 'U' - urban or 'R' - rural)

5 famsize - family size (categorical: 'LE3' - less or equal to 3 or 'GT3' - greater than 3)

6 Pstatus - parent's cohabitation status (categorical: 'T' - living together or 'A' - apart)

7 Medu - mother's education (numeric: 0 - none, 1 - primary education (4th grade), 2 “ 5th to 9th grade, 3 secondary education or 4 “ higher education)

8 Fedu - father's education (numeric: 0 - none, 1 - primary education (4th grade), 2 “ 5th to 9th grade, 3 “ secondary education or 4 “ higher education)

9 Mjob - mother's job (categorical: 'teacher', 'health' care related, civil 'services' (e.g. administrative or police), 'at\_home' or 'other')

10 Fjob - father's job (categorical: 'teacher', 'health' care related, civil 'services' (e.g. administrative or police), 'at\_home' or 'other')

11 reason - reason to choose this school (categorical: close to 'home', school 'reputation', 'course' preference or 'other') 12 guardian - student's guardian (nominal: 'mother', 'father' or 'other')

13 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)

14 studytime - weekly study time (numeric: 1 - <2 hours, 2 - 2 to 5 hours, 3 - 5 to 10 hours, or 4 - >10 hours)

15 failures - number of past class failures (numeric: n if  $1 \leq n < 3$ , else 4)

16 schoolsup - extra educational support (categorical: yes or no)

17 famsup - family educational support (categorical: yes or no)

18 paid-extra paid classes within the course subject (Math or Portuguese) (categorical: yes or no)

19 activities - extra-curricular activities (categorical: yes or no)

20 nursery - attended nursery school (categorical: yes or no)

21 higher - wants to take higher education (categorical: yes or no)

22 internet - Internet access at home (categorical: yes or no)

23 romantic - with a romantic relationship (categorical: yes or no)

24 famrel - quality of family relationships (numeric: from 1 - very bad to 5 - excellent)

25 freetime - free time after school (numeric: from 1 - very low to 5 - very high)

26 goout - going out with friends (numeric: from 1 - very low to 5 - very high)

27 Dalc - workday alcohol consumption (numeric: from 1 - very low to 5 - very high)

28 Walc - weekend alcohol consumption (numeric: from 1 - very low to 5 - very high)

29 health - current health status (numeric: from 1 - very bad to 5 - very good)

30 absences - number of school absences (numeric: from 0 to 93)

## Grades

31 G1 - first period grade (numeric: from 0 to 20)

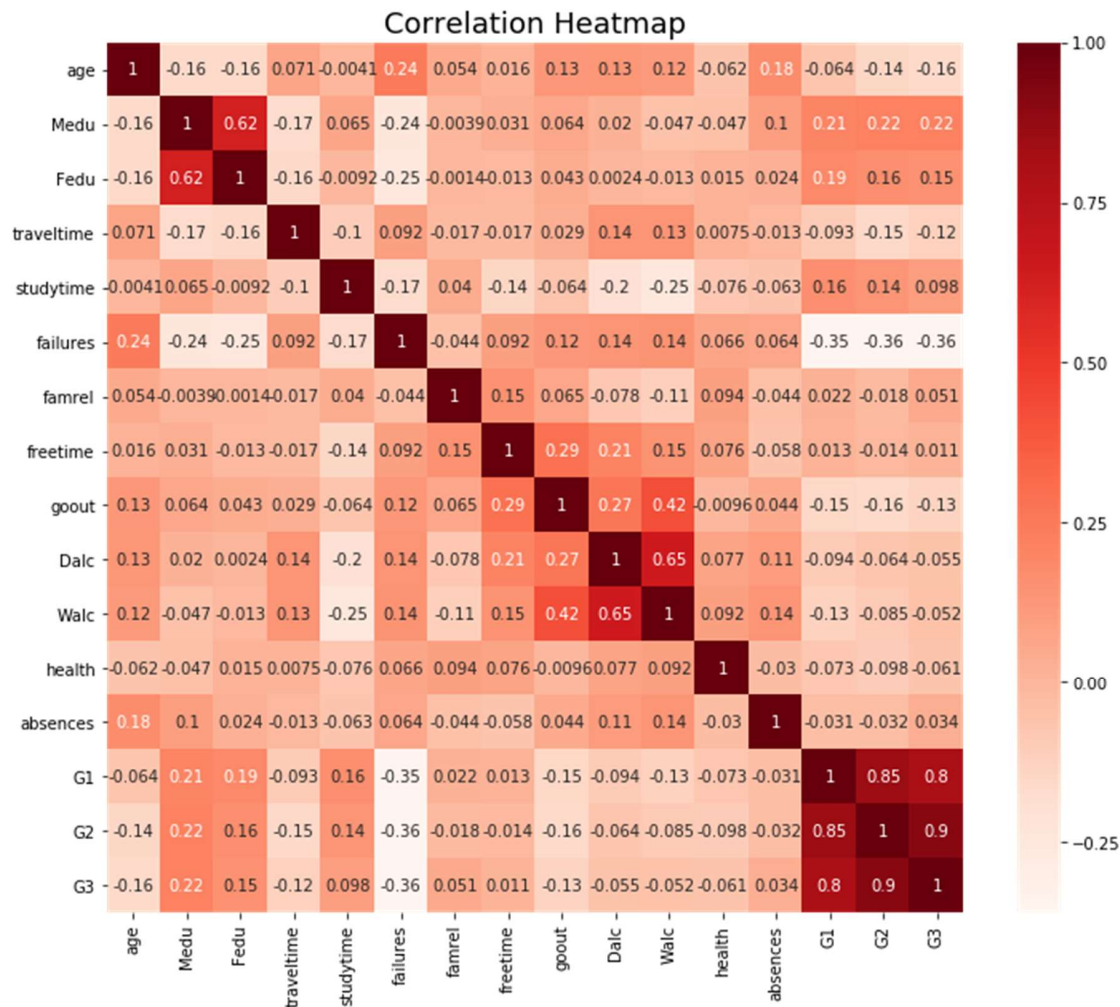
31 G2 - second period grade (numeric: from 0 to 20)

32 G3 - final grade (numeric: from 0 to 20, output target)

## Summary statistics of each attribute

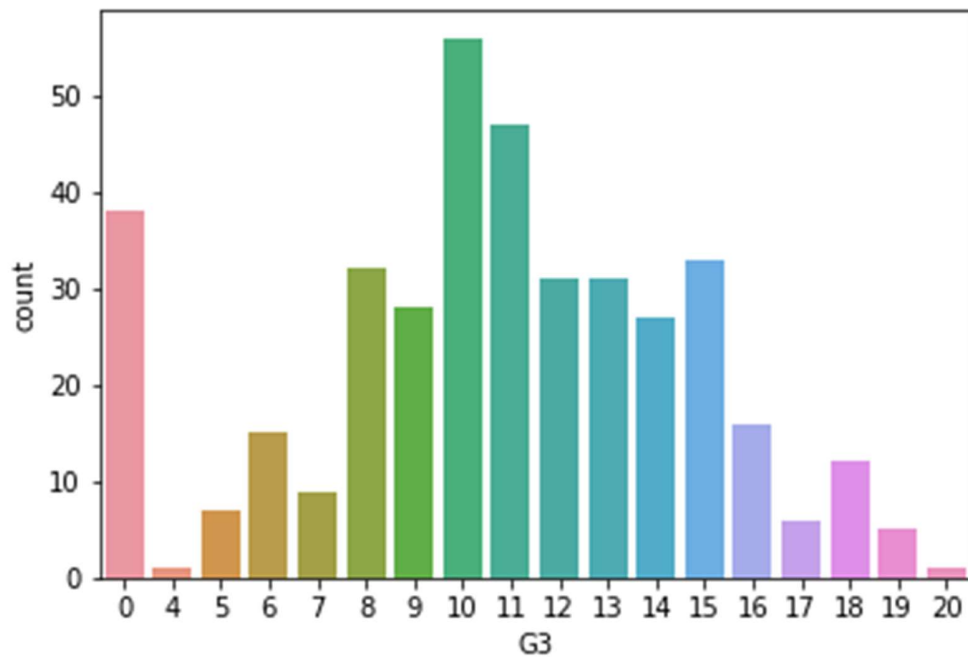
|       | age        | traveltime | studytime  | freetime   | goout      | G1         | G2         | G3         |
|-------|------------|------------|------------|------------|------------|------------|------------|------------|
| count | 395.000000 | 395.000000 | 395.000000 | 395.000000 | 395.000000 | 395.000000 | 395.000000 | 395.000000 |
| mean  | 16.696203  | 1.448101   | 2.035443   | 3.235443   | 3.108861   | 10.908861  | 10.713924  | 10.415190  |
| std   | 1.276043   | 0.697505   | 0.839240   | 0.998862   | 1.113278   | 3.319195   | 3.761505   | 4.581443   |
| min   | 15.000000  | 1.000000   | 1.000000   | 1.000000   | 1.000000   | 3.000000   | 0.000000   | 0.000000   |
| 25%   | 16.000000  | 1.000000   | 1.000000   | 3.000000   | 2.000000   | 8.000000   | 9.000000   | 8.000000   |
| 50%   | 17.000000  | 1.000000   | 2.000000   | 3.000000   | 3.000000   | 11.000000  | 11.000000  | 11.000000  |
| 75%   | 18.000000  | 2.000000   | 2.000000   | 4.000000   | 4.000000   | 13.000000  | 13.000000  | 14.000000  |
| max   | 22.000000  | 4.000000   | 4.000000   | 5.000000   | 5.000000   | 19.000000  | 19.000000  | 20.000000  |

# Correlations

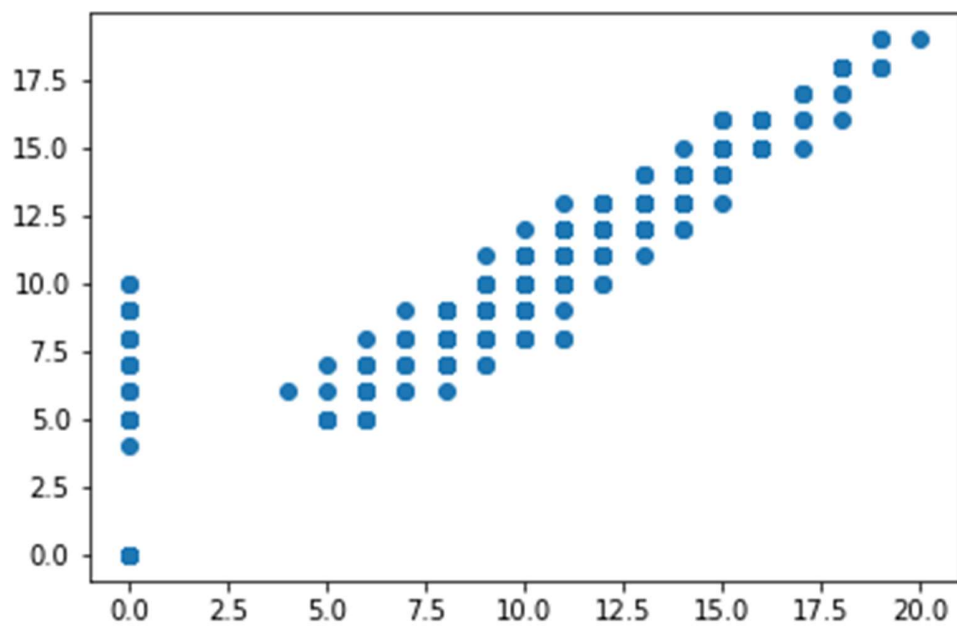


|           |           |
|-----------|-----------|
| G3        | 1.000000  |
| G2        | 0.904868  |
| G1        | 0.801468  |
| Medu      | 0.217147  |
| Fedu      | 0.152457  |
| studytime | 0.097820  |
| famrel    | 0.051363  |
| absences  | 0.034247  |
| freetime  | 0.011307  |
| Walc      | -0.051939 |

Table1: Correlations according to Final Grade

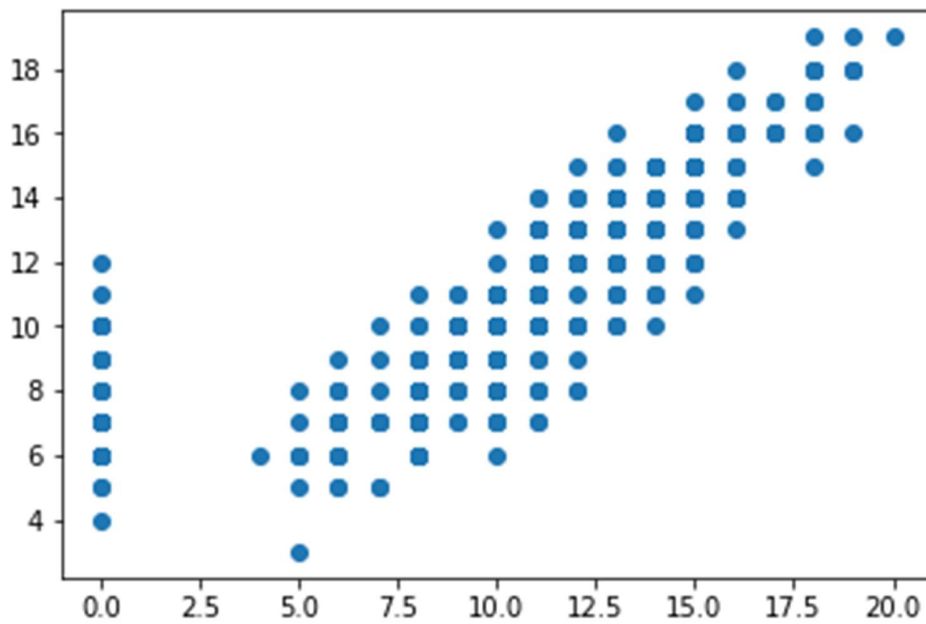


**Note: There is a huge amount of failure student in mat class**



(0.9048679892693012, 7.62571915404493e-148)

**G1-G2 → Highly correlated and meaningful (p value < 0,05)**



(0.8014679320174141, 9.001430312276602e-90)

**G1-G3 → Highly correlated and meaningful(p value <0,05)**

**-Find out the percentage of missing data for each attribute.**

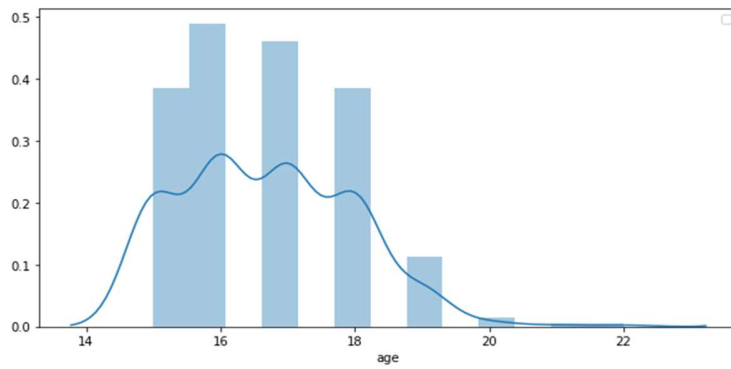
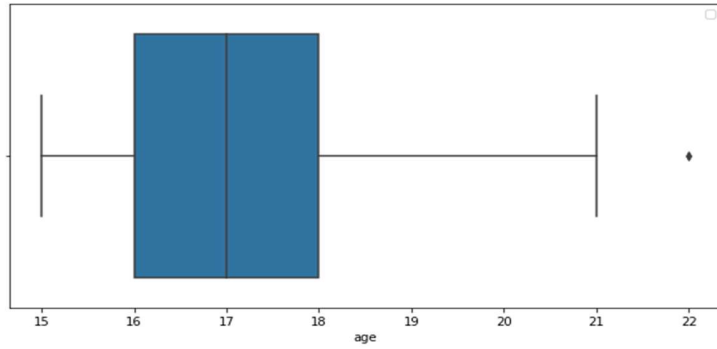
There is no missing data.

**-Find out duplicate data if there are any.**

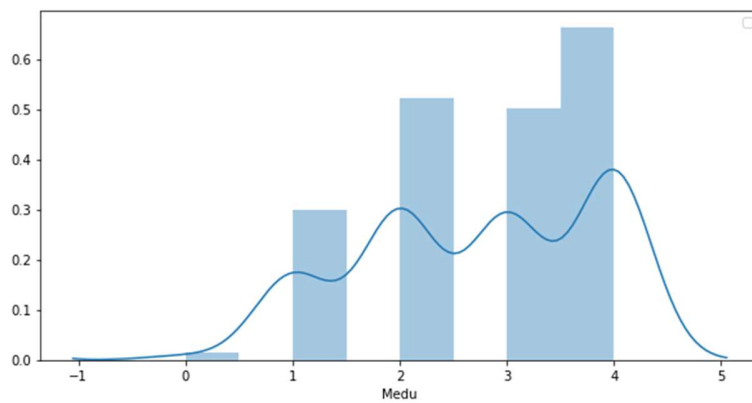
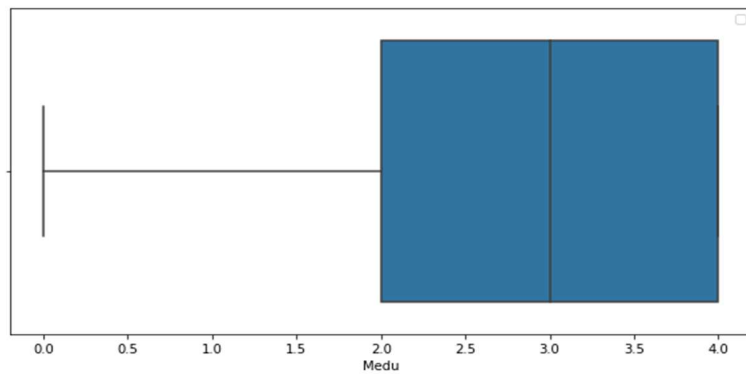
There is no duplication in df.

## ATTRIBUTES HISTOGRAM VS. BOXPLOT

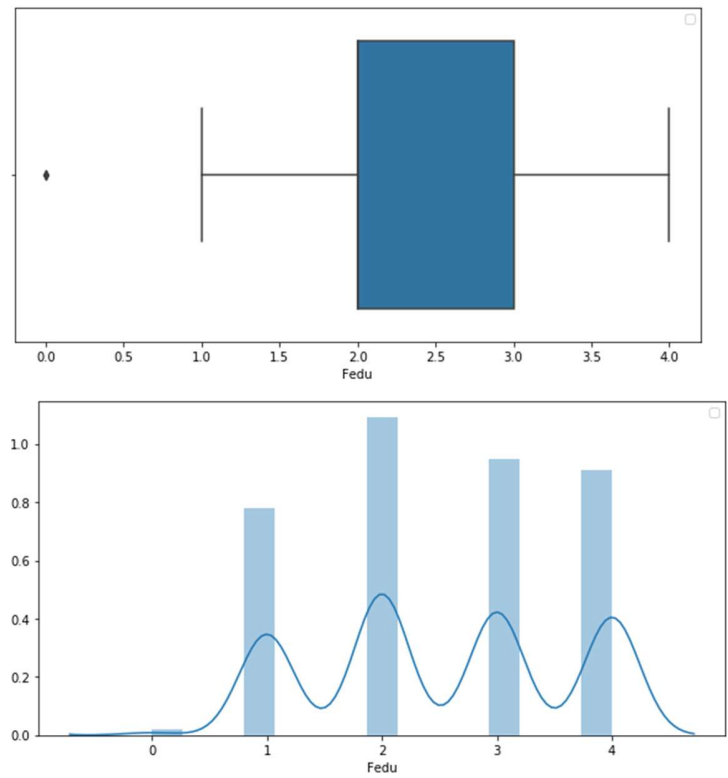
### 1. AGE



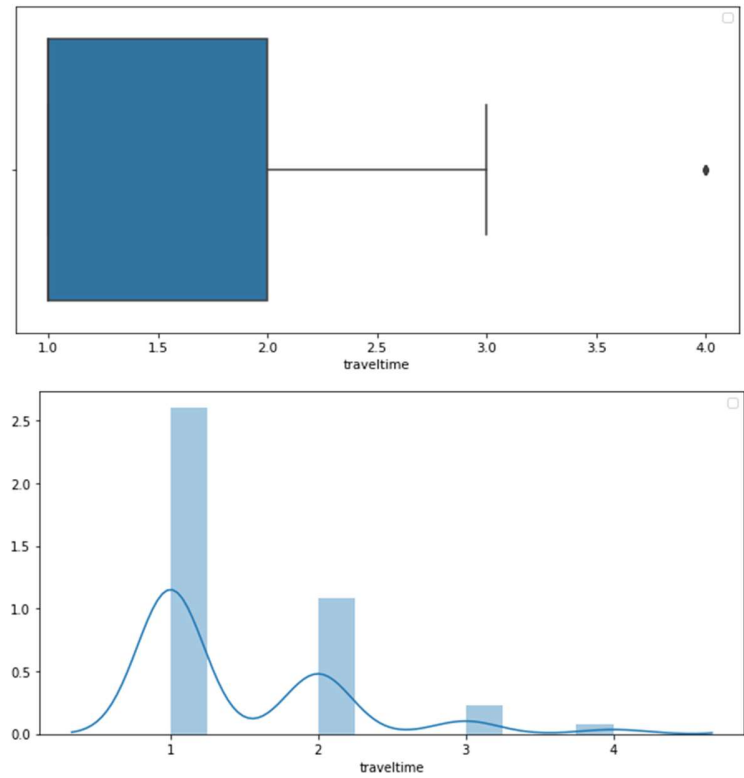
### 2. MEDU



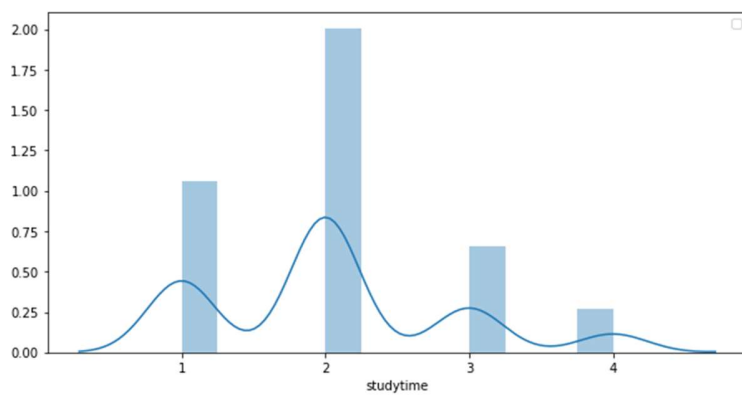
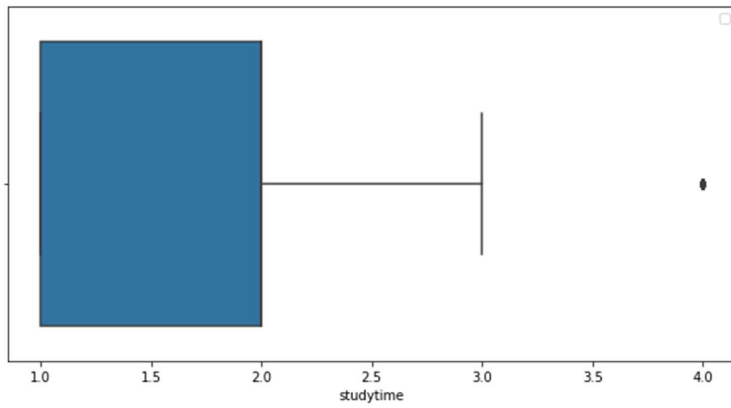
3.FEDU



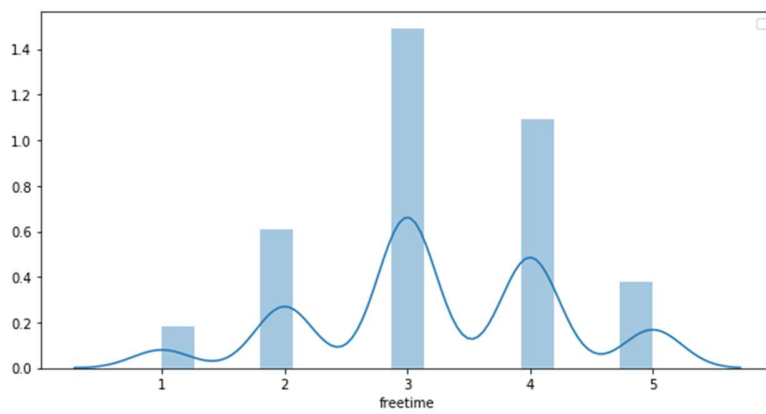
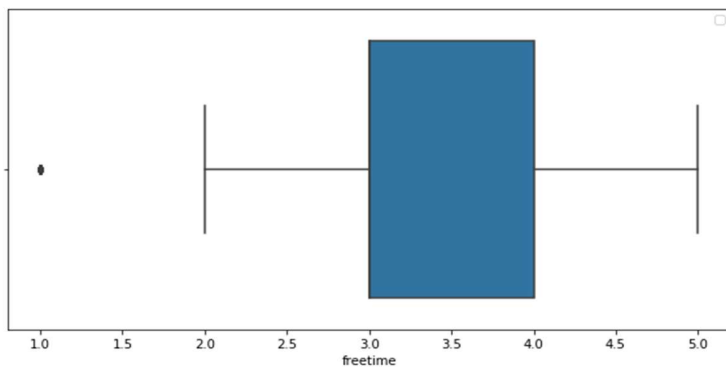
4.TRAVELTIME



## 5.STUDYTIME

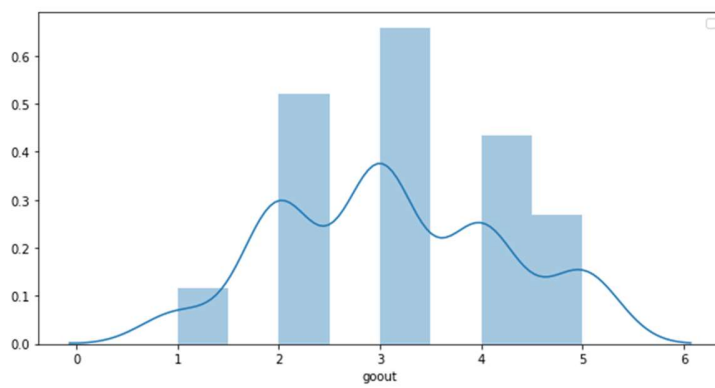
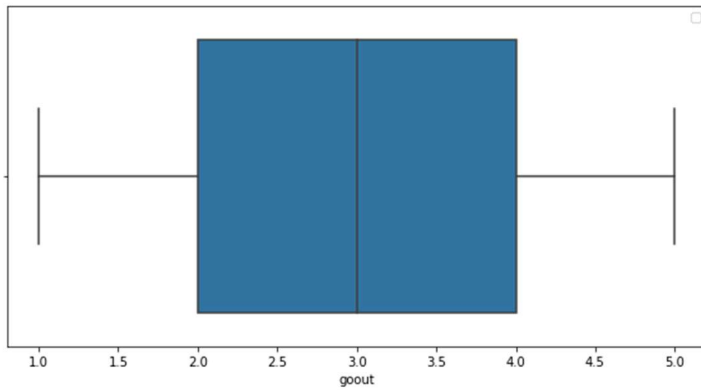


## 6.FREETIME

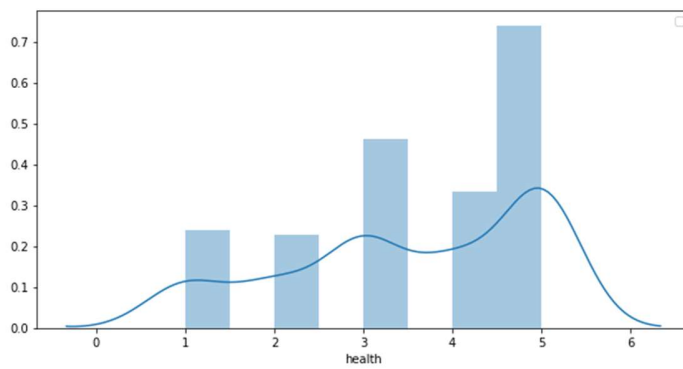
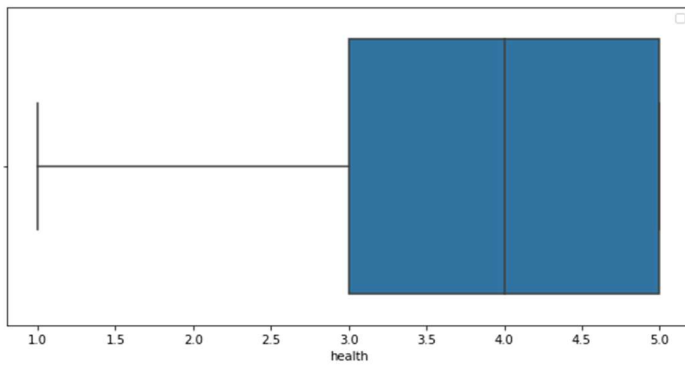




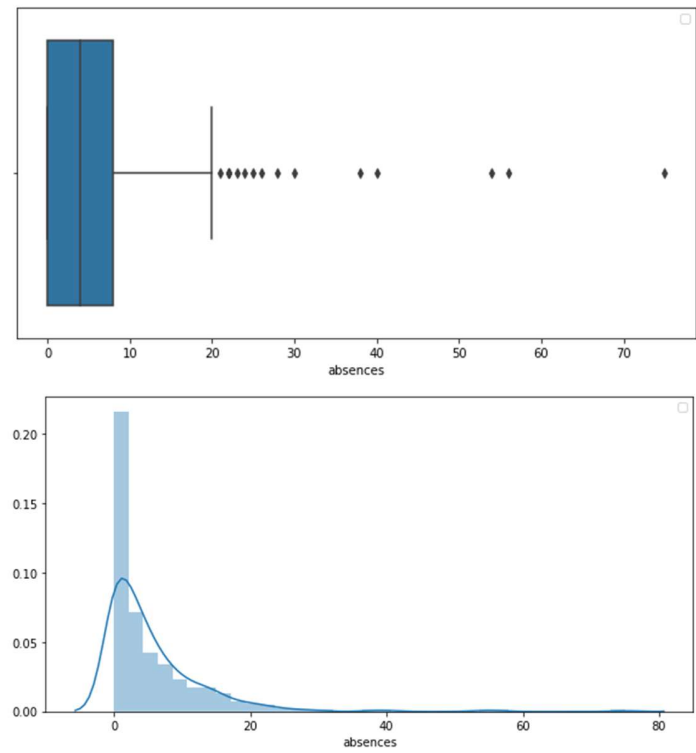
## 7.GOOUT



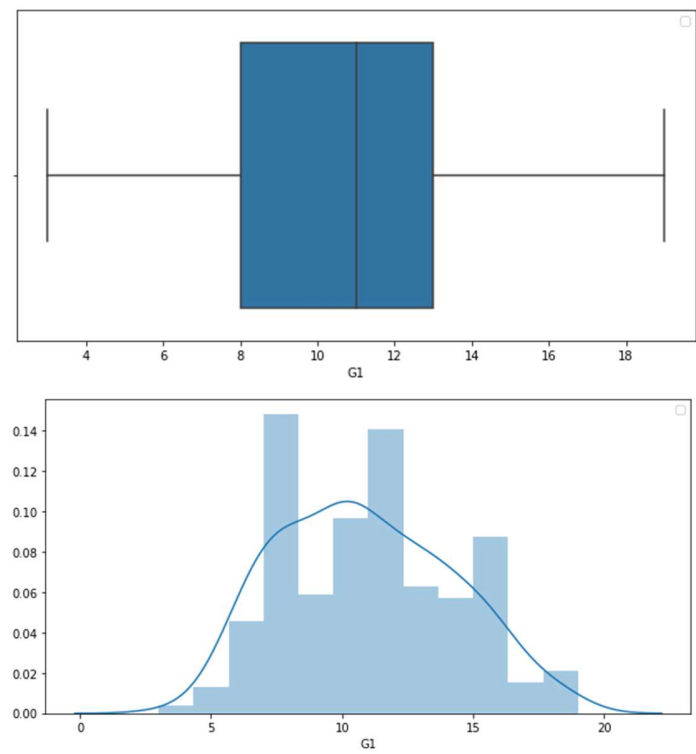
## 8.HEALTH



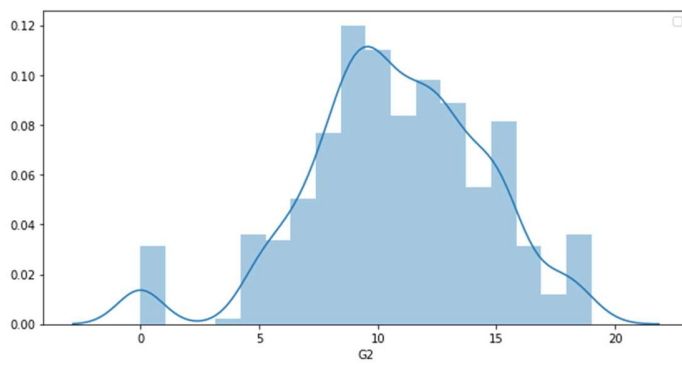
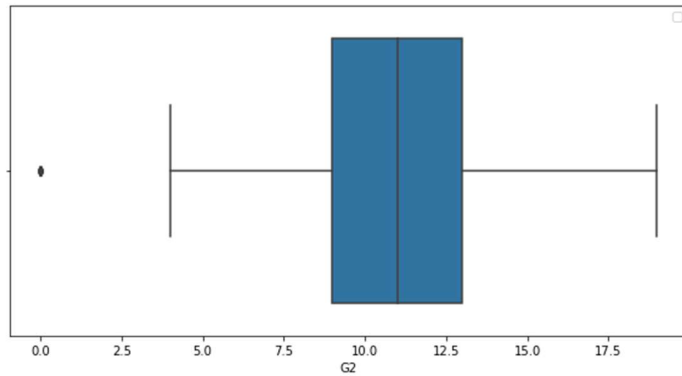
9.ABSENCES



10.G1



## 11.G2



## 12.G3

