La siguiente table es un head del data set correspondiente para la elaboracion y ejecucion del parcial.

|  |  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| **school** | **sex** | **age** | **address** | **famsize** | **Pstatus** | **Medu** | **Fedu** | **Mjob** | **Fjob** | **reason** |
| **0** GP | F | 18 | U | GT3 | A | 4 | 4 | at\_home | teacher | course |
| **1** GP | F | 17 | U | GT3 | T | 1 | 1 | at\_home | other | course |
| **2** GP | F | 15 | U | LE3 | T | 1 | 1 | at\_home | other | other |
| **3** GP | F | 15 | U | GT3 | T | 4 | 2 | health | services | home |
| **4** GP | F | 16 | U | GT3 | T | 3 | 3 | other | other | home |
| **5** GP | M | 16 | U | LE3 | T | 4 | 3 | services | other | reputation |
| **6** GP | M | 16 | U | LE3 | T | 2 | 2 | other | other | home |
| **7** GP | F | 17 | U | GT3 | A | 4 | 4 | other | teacher | home |
| **8** GP | M | 15 | U | LE3 | A | 3 | 2 | services | other | home |
| **9** GP | M | 15 | U | GT3 | T | 3 | 4 | other | other | home |

**¿DATOS FALNTANTES?**

Se observan datos falntantes por variable

school 0

sex 0

age 0

address 0

famsize 0

Pstatus 0

Medu 0

Fedu 0

Mjob 0

Fjob 0

reason 0

guardian 0

traveltime 0

studytime 0

failures 0

schoolsup 0

famsup 0

paid 0

activities 0

nursery 0

higher 0

internet 0

romantic 0

famrel 0

freetime 0

goout 0

Dalc 0

Walc 0

health 0

absences 0

G1 0

G2 0

G3 0

dtype: int64

**Como se puede evidenciar no existen datos faltantes en ninguna de las variables del data set**

## ¿Cuántos estudiantes viven en zona urbana o rural? ¿Cuántos de los estudiantes rurales son mujeres y cuántas son hombres?

### ¿Cuántos estudiantes viven en zona urbana o rural?

Urbana 307

Rrural 88

### ¿Cuántos de los estudiantes rurales son mujeres y cuántas son hombres?

Female 44

Male 44

## ¿Cuántos estudiantes que viven en la zona rural se demoran en transportarse para llegar de la casa al colegio más de una hora?

5 estudiantes que viven en la zona rural se demoran mas de una hora en llegar de la casa al colegio

### ¿Cuántos de ellos son mujeres y cuántos son hombres?

Male 4

Female 1

## ¿Cuántos estudiantes que viven en la zona rural tienen más de 5 fallas?

38

### ¿Cuántos de ellos son mujeres y cuántos son hombres?

Female 20

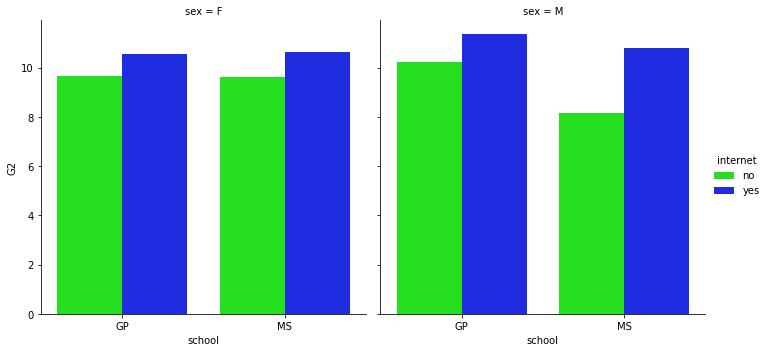
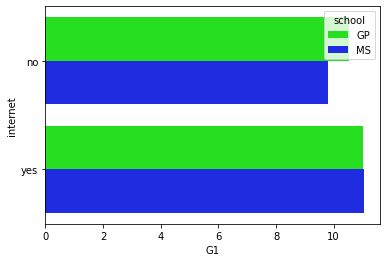
Male 18

## ¿Cuántos estudiantes tienen madres y padres con un grado de escolaridad por encima de la educación secundaria?

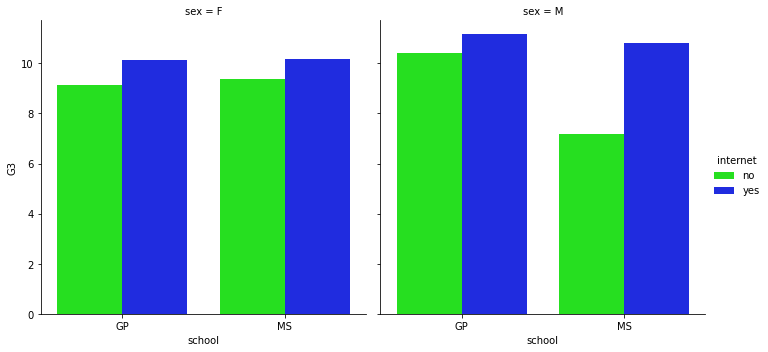
Se tomo la condicion de si los padres y madres tenian un grado de escolaridad por encima de educacion segundaria donde el resultado fue 69 estudiantes

## Se quiere ver la media de las notas de matemáticas de los estudiantes del colegio Gabriel Pereira vs Mousinho da Silveira, teniendo en cuenta si los estudiantes tienen o no internet en sus

**casas. Realice una visualización de datos teniendo en cuenta estas 3 variables.**



< >



## Cree una nueva variable en la tabla que corresponda a la

**calificación total para cada uno de los estudiantes y responda ¿Cuál es la media, mediana, mínima y máxima calificación total?**

|  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- |
| **Medu** | **Fedu** | **traveltime** | **studytime** | **failures** | **famrel** | **freetime** | **go** |
| 395.000000 | 395.000000 | 395.000000 | 395.000000 | 395.000000 | 395.000000 | 395.000000 | 395.000 |
| 2.749367 | 2.521519 | 1.448101 | 2.035443 | 0.334177 | 3.944304 | 3.235443 | 3.108 |
| 1.094735 | 1.088201 | 0.697505 | 0.839240 | 0.743651 | 0.896659 | 0.998862 | 1.113 |
| 0.000000 | 0.000000 | 1.000000 | 1.000000 | 0.000000 | 1.000000 | 1.000000 | 1.000 |
| 2.000000 | 2.000000 | 1.000000 | 1.000000 | 0.000000 | 4.000000 | 3.000000 | 2.000 |
| 3.000000 | 2.000000 | 1.000000 | 2.000000 | 0.000000 | 4.000000 | 3.000000 | 3.000 |
| 4.000000 | 3.000000 | 2.000000 | 2.000000 | 0.000000 | 5.000000 | 4.000000 | 4.000 |
| 4.000000 | 4.000000 | 4.000000 | 4.000000 | 3.000000 | 5.000000 | 5.000000 | 5.000 |

# media

32.037974683544306

# mediana

32.0

# minimo y maximo

(4, 58)

Teniendo en cuenta que el dataset contiene tanto variables numéricas como variablescategóricas responda ¿Qué tipo de análisis realizaría: Regresión o clasificación? Si fueraregresión o clasificación: ¿Qué variable usaría para predecir o clasificar? ¿Cuáles variables tiene la hipótesis que podrían servir para pronosticar o clasificar?

Con el data set studen\_ mat realizaria una regresion con la variable tiempo de estudio (studytime) y de acuerdo a las variables como lo son traveltime, age, sex, activities,higher,internet,romantic,freetime,absences,failures pronosticar el tiempo de estudio empleado de acuerdo a las actividades o frecuencias observadas.

#### 

