**Proyecto de finalización del módulo Coding & Programming SIC-2022**

**Objetivo:**

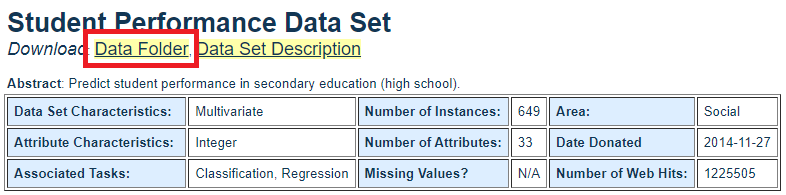
Desarrollar una automatización que permita extraer e interpretar la data del archivo Student-mat.csv y Student-por.csv de manera que una la información solicitada de ambos archivos y realice los requerimientos solicitados

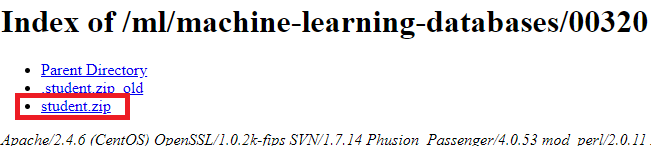
**Instrucciones:**

Acceda a la base de datos a través del siguiente link:

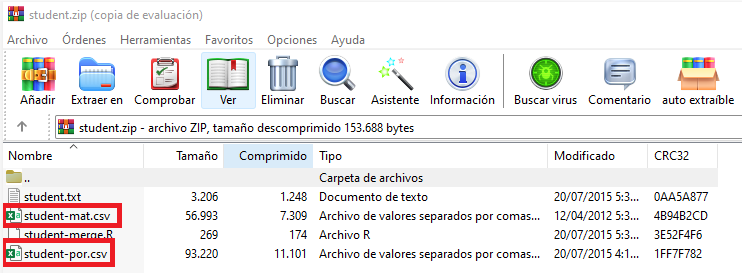
<https://archive.ics.uci.edu/ml/datasets/student+performance>

Allí acceder a Data folder y descargar el archivo Student.zip





Descomprima el archivo.zip y extraiga los dos archivos .csv con los que desarrollará el proyecto



En estos archivos.csv encontrara información relacionada a dos cursos, uno de matemáticas y otro de portugués, donde la información de cada columna esta especificada a continuación:

**Información de atributos:**

# Atributos para los conjuntos de datos student-mat.csv (curso de matemáticas) y student-por.csv (curso de portugués):

1. school - escuela del estudiante (binario: 'GP' - Gabriel Pereira o 'MS' - Mousinho da Silveira)
2. sex - sexo del estudiante (binario: "F" - mujer o "M" - hombre)
3. age - edad del estudiante (numérico: de 15 a 22 años)
4. adress - tipo de domicilio del estudiante (binario: "U" - urbano o "R" - rural)
5. famsize - tamaño de la familia (binario: "LE3" - menos o igual a 3 o "GT3" - más de 3)
6. Pstatus - estado de convivencia de los padres (binario: "T" - convivencia o "A" - separación)
7. Medu - educación de la madre (numérico: 0 - ninguna, 1 - educación primaria (4º grado), 2 - 5º a 9º grado, 3 - educación secundaria o 4 - educación superior)
8. Fedu - educación del padre (numérico: 0 - ninguna, 1 - educación primaria (4º grado), 2 - 5º a 9º grado, 3 - educación secundaria o 4 - educación superior)
9. Mjob - trabajo de la madre (nominal: 'teacher', 'health' care related, civil 'services' (e.g. administrative or police), 'at\_home' or 'other')
10. Fjob - trabajo del padre (nominal: 'teacher', 'health' care related, civil 'services' (e.g. administrative or police), 'at\_home' or 'other')
11. reason - razón para elegir este centro educativo (nominal: close to 'home', school 'reputation', 'course' preference or 'other')
12. guardian - tutor del alumno (nominal: 'mother', 'father' or 'other')
13. traveltime - tiempo de viaje de casa a la escuela (numérico: 1 - <15 min., 2 - 15 a 30 min., 3 - 30 min. a 1 hora, o 4 - >1 hora)
14. studytime - tiempo de estudio semanal (numérico: 1 - <2 horas, 2 - 2 a 5 horas, 3 - 5 a 10 horas, o 4 - >10 horas)
15. failures - número de fracasos de clase anteriores (numérico: n si 1<=n<3, sino 4)
16. schoolsup - apoyo educativo adicional (binario: sí o no)
17. famsup - apoyo educativo familiar (binario: sí o no)
18. paid - clases extra pagadas dentro de la asignatura del curso (matemáticas o portugués) (binario: sí o no)
19. activities - actividades extraescolares (binario: sí o no)
20. nursery - asistió a la guardería (binario: sí o no)
21. higher - quiere cursar estudios superiores (binario: sí o no)
22. internet - acceso a internet en casa (binario: sí o no)
23. romantic - con una relación romántica (binario: sí o no)
24. famrel - calidad de las relaciones familiares (numérico: de 1 - muy malo a 5 - excelente)
25. freetime - tiempo libre después de la escuela (numérico: de 1 - muy bajo a 5 - muy alto)
26. goout - salir con amigos (numérico: de 1 - muy bajo a 5 - muy alto)
27. Dalc - consumo de alcohol en el día de trabajo (numérico: de 1 - muy bajo a 5 - muy alto)
28. Walc - consumo de alcohol en fin de semana (numérico: de 1 - muy bajo a 5 - muy alto)
29. health - estado de salud actual (numérico: de 1 - muy malo a 5 - muy bueno)
30. absences - número de ausencias escolares (numérico: de 0 a 93)

# Estas calificaciones están relacionadas con la asignatura del curso, Matemáticas o Portugués:

G1 - calificación del primer período (numérico: de 0 a 20)

G2 - calificación del segundo periodo (numérico: de 0 a 20)

G3 - nota final (numérico: de 0 a 20, objetivo de salida)

|  |
| --- |
|  |

De acuerdo a las indicaciones presentadas a continuación, esta automatización debe ser desarrollado en Python.

**Actividades:**

Obtenga la fecha en la que se ejecuta el script

Para ambos archivos:

Cree un DataFrame para cada escuela, con el formato de nombre “nombrecolegio\_curso” que reúna los datos de las columnas: school, sex, age, address, Pstatus, guardian, traveltime, studytime, failures, paid, internet, health, absences, G1,G2,G3

Verifique que no haya data de valor nulo (NaN), en caso de encontrar algún valor NaN se deberá eliminar toda la fila

Para cada escuela muestre un gráfico circular(pastel) donde se evidencie el porcentaje de estudiantes hombres y mujeres de cada curso

Para cada escuela muestre un gráfico de barras donde se muestre la cantidad de estudiantes que tienen la misma edad

Muestre el promedio de las edades de cada curso de cada escuela

Muestre el promedio de las notas G1, G2, G3 de cada curso de cada escuela

Grafique el promedio de las notas en un gráfico de barras horizontal

Halle el valor máximo de las ausencias y considere dicho valor como el total de clases del curso, de manera que pueda sacar un porcentaje de asistencia para cada estudiante

Cree una nueva columna llamada “extra”

Cree una nueva columna llamada “approved” en la cual determine si el estudiante aprueba o reprueba el curso (1 para aprueba, 0 para reprueba), considerando:

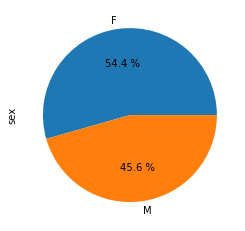
* Si el estudiante tiene un porcentaje de asistencia menor al 80% del curso reprueba
* Si el estudiante tiene un porcentaje de asistencia mayor al 80% del curso, pero la nota G3 es menor a 10 reprueba
* Si el estudiante tiene un porcentaje de asistencia mayor al 80% del curso, pero la nota G3 está entre 10 y 15 aprueba y se asigna 1 a la columna extra
* Si el estudiante tiene un porcentaje de asistencia mayor al 80% del curso y la nota G3 es mayor a 15 aprueba y se asigna 0 a la columna extra

Grafique el porcentaje de alumnos aprobados de cada curso

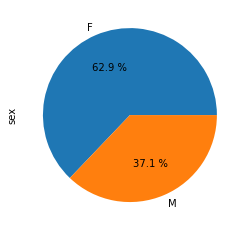
**Resultado:**

1. Generar un archivo en formato csv con nombre: “Resultado + Fecha en la que se ejecuta el script”, este archivo debe quedar en la carpeta donde se encuentren los datos originales y el archivo de Python
2. Entregar en un archivo PDF todas las gráficas solicitadas

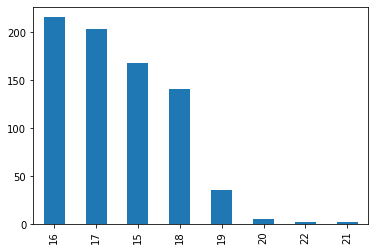
escuela1["sex"].value\_counts().plot(*kind*="pie", *autopct*="%0.1f %%")



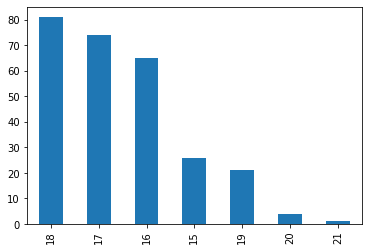
escuela2["sex"].value\_counts().plot(*kind*="pie", *autopct*="%0.1f %%")



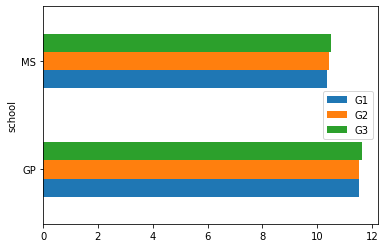
escuela1["age"].value\_counts().plot(*kind*="bar")



escuela2["age"].value\_counts().plot(*kind*="bar")



estudiantes[["school", "G1", "G2", "G3"]].groupby("school").mean().plot(*kind*="barh")



estudiantes["approved"].value\_counts().plot(*kind*="pie", *autopct*="%0.1f %%")

