



PROGRAMACION ORIENTADA A OBJETOS (CC201)

Ciclo 2024 - 02

Taller Semana 12

numpy-pandas

Caso 1:

El archivo titanic.csv contiene información sobre los pasajeros del Titanic. Escribir un programa en Python que realice lo siguiente:

- a) Generar un DataFrame con los datos del archivo.
- b) Mostrar las dimensiones del DataFrame, el número de datos que contiene, los nombres de sus columnas y filas, los tipos de datos de las columnas, las 10 primeras filas y las 10 últimas filas
- c) Mostrar los datos del pasajero con identificador 148.
- d) Mostrar las filas pares del DataFrame.
- e) Mostrar los nombres de las personas que iban en primera clase ordenadas alfabéticamente.
- f) Mostrar el porcentaje de personas que sobrevivieron y murieron.
- g) Mostrar el porcentaje de personas que sobrevivieron en cada clase.
- h) Eliminar del DataFrame los pasajeros con edad desconocida.
- i) Mostrar la edad promedio o media de las mujeres que viajaban en cada clase.
- j) Añadir una nueva columna booleana para ver si el pasajero era menor de edad o no.
- k) Mostrar el porcentaje de menores y mayores de edad que sobrevivieron en cada clase.

Caso 2

Se tiene el archivo Excel colesterolos.xlsx, escribir un programa en Python que:

- a) Genere un DataFrame con esos datos. Observar que el archivo tiene la coma como separador de decimales, se debe asegurar que sea vea como puntos en el DataFrame
- b) Adicionar un registro con los siguientes datos:
 - a. Nombre: Carlos Rivas
 - b. Edad: 28
 - c. Sexo: Hombre
 - d. Peso: 80 kg
 - e. Altura: 1.78
 - f. Colesterol: 245
- c) Verificar que se ingresó correctamente imprimiendo los últimos 7 registros
- d) Agregar una columna al DataFrame, llamada diabetes, que tendrá los siguientes valores para los registros: ["No", "No", "Si", "No", "Si", "No", "Si", "No", "Si", "No", "Si", "No", "Si", "No"]
- e) Grabar un nuevo archivo Excel, volverlo a leer generando otro DataFrame y comprobar que los datos sean correctos

Caso 3:

Se tiene el siguiente archivo bupa.csv, los datos vienen sin encabezado, pero se sabe que los que debe corresponder son los siguientes: mcv, alkphos, sgpt, sgot, gammagt, drinks, selector

Se pide:

- a) Leer los datos y crear un DataFrame
- b) Poner los encabezados de las columnas
- c) Calcular la media, la mediana y la moda de la columna alkphos
- d) Crear una copia de DataFrame y poner 50 nan de manera aleatoria
- e) Contar cuantos nan hay por columna
- f) Eliminar las filas con NaN e imprimir el DataFrame