

| Plan Formativo: Ciencia de Datos | Nivel de Dificultad |
|--|---------------------|
| Módulo 2: Obtención y Preparación de Datos | Bajo / medio |
| Tema: Obtención de Datos desde Archivos CSV y Excel | |

Intención del aprendizaje o aprendizaje esperado:

 Aplica técnicas de limpieza y preparación de datos utilizando librerías Python para su depuración.

Ejercicios planteados

1. El fichero titanic.csv contiene información sobre los pasajeros del Titanic. Escribir un programa con los siguientes requisitos:



- a) Generar un DataFrame con los datos del fichero.
- b) Mostrar por pantalla 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 por pantalla los datos del pasajero con identificador 148.
- d) Mostrar por pantalla las filas pares del DataFrame.
- e) Mostrar por pantalla los nombres de las personas que iban en primera clase ordenadas alfabéticamente.
- f) Mostrar por pantalla el porcentaje de personas que sobrevivieron y murieron.



- g) Mostrar por pantalla el porcentaje de personas que sobrevivieron en cada clase.
- h) Eliminar del DataFrame los pasajeros con edad desconocida.
- i) Mostrar por pantalla la edad 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 por pantalla el porcentaje de menores y mayores de edad que sobrevivieron en cada clase.

Datos de apoyo al planteamiento

Preguntas guía

- ¿Qué son los datos perdidos en Python?
- ¿Cómo manipular datos perdidos?

Recursos Bibliográficos:

[1] Lectura de ficheros CSV

https://www.youtube.com/watch?v=wmNsecoZ Go

[2] Escritura de ficheros CSV

https://www.youtube.com/watch?v=uCv5zGPGgb4