

Lead University
Bachillerato Ingeniería en Ciencia de Datos
Administración de datos

Caso 3
Porcentaje: 10%
Valor de la Tarea: 100 puntos
Fecha de Entrega: miércoles 11 de agosto de 2021 antes de las 6:00 pm

Caso 3

Descripción:

1. El caso 3 se será desarrollado en parejas.
2. Se debe subir al aula virtual el link de Github.

Instrucciones:

1. Importar el CSV data.csv en Python. El delimitador de los valores es (;) **(5 puntos)**
2. Mostrar los valores los primeros 5 valores del dataframe **(5 puntos)**
3. Mostrar los últimos 5 valores del dataframe **(5 puntos)**
4. Mostrar el tipo de dato (Dtype) de cada columna(Ejemplo si es Integer, object, float, etc) **(5 puntos)**
5. Mostrar la cantidad de valores nulos de cada columna **(5 puntos)**
6. Mostrar el porcentaje total de valores perdidos en todo el dataframe **(10 puntos)**
7. Mostrar el porcentaje de valores perdidos por columna **(5 puntos)**
8. Eliminar aquella columna (s) que tengan más del 70% de sus valores en nulo. **(10 puntos)**
9. Eliminar aquellas **filas** que tenga menos de dos elementos. **(5 puntos)**
10. Reemplace los String "--" por valores nulos **(5 puntos)**
11. Reemplazar los valores nulos de la columna de "Gender" por U (Unknown) para aquellos empleados que no especificaron este campo. **(5 puntos)**
12. Reemplazar los valores nulos de la columna "Departamento" por "Bench" **(5 puntos)**
13. Reemplazar los valores nulos de la columna "HourlyRate" por la mediana **(5 puntos)**
14. Reemplazar los valores perdidos inesperados de la columna "Overtime" por No **(5 puntos)**
15. Salvar el dataframe una vez realizado el procesamiento de los puntos anteriores **(5 puntos)**
16. Documentar el código **(5 puntos)**
17. Escoger la estructura de carpetas y la nomenclatura de las carpetas y los archivos para los datos sin procesar y los datos procesados, además el código de Python. **(10 puntos)**