Proyecto - 3 (Estudio de caso - 3) Parte 1 | Análisis de datos con Python Pandas

Enlace al conjunto de datos : https://www.kaggle.com/wenruliu/adult...

----------------------------------------------------------------

En este proyecto, vamos a discutir:

- ¿Cómo obtener muestras aleatorias del conjunto de datos?

- es decir

- entre

- único

- dropna

- reemplazar

- duplicado

- drop\_duplicados

- astype

- aplicar

- ¿Qué es el análisis univariante?

- ¿Qué es el análisis bivariante?

- Optimización de la memoria

--------------------------------------------------

Preguntas :

1.Mostrar las 10 primeras filas del conjunto de datos

2. Comprobar las 10 últimas filas del conjunto de datos

3. Encontrar la forma de nuestro conjunto de datos (número de filas y número de columnas)

4. Obtener información sobre nuestro conjunto de datos como el número total de filas, el número total de columnas, los tipos de datos de cada columna y la memoria necesaria

5. Obtener una muestra aleatoria del conjunto de datos (50%)

6.Comprobar los valores nulos en el conjunto de datos

7.Realice la limpieza de datos [ Reemplace '?' con NaN ]

8. 8. Eliminar todos los valores perdidos

9. 9. Comprobar los datos duplicados y eliminarlos

Proyecto - 3 (Estudio de caso - 3) Parte 2 | Análisis de datos con Python Pandas

Enlace al conjunto de datos: https://www.kaggle.com/wenruliu/adult...

----------------------------------------------------------------

En este proyecto, vamos a discutir:

- ¿Cómo obtener muestras aleatorias del conjunto de datos?

- es decir

- entre

- único

- dropna

- reemplazar

- duplicado

- drop\_duplicados

- astype

- aplicar

- ¿Qué es el análisis univariante?

- ¿Qué es el análisis bivariante?

- Optimización de la memoria

Preguntas:

10. Obtener estadísticas generales sobre el marco de datos

11. Eliminar las columnas education-num, capital-gain y capital-loss

12. ¿Cuál es la distribución de la columna de edad?

13. Encuentre el número total de personas con edades comprendidas entre los 17 y los 48 años (inclusive) utilizando el método entre

14. ¿Qué es la columna de distribución de la clase de trabajo?

15. 15. ¿Cuántas personas tienen un título de grado y de máster?

16. Análisis bivariante

17. Sustitución de los valores salariales por 0 y 1

18. ¿Qué clase de trabajo obtiene el salario más alto?

19.¿Cómo tiene más posibilidades de obtener un salario superior a 50K el hombre o la mujer?

20. 20. Cambiar el tipo de datos de las columnas de la clase de trabajo por el tipo de datos de la categoría