

---

## EXAMEN FINAL

---

1. Mencione cuales son los 2 tipos de objetos principales al trabajar con Pandas
2. Al usar Pandas loc y iloc nos sirven para:
3. ¿Qué función usamos para agregar una fila de un dataframe de Pandas a otro?
4. Para calcular la transpuesta de un dataset, tanto en NumPy como en Pandas podemos usar
5. Se posee una matriz con 1000 filas y 300 columnas, al aplicar `sum(axis=0)` sobre esta matriz , devuelve un vector de cuantos elementos?
6. Cuando trabajamos con dataframes, que elementos (de la estructura del dataframe) permiten usar nombres, etiquetas o identificadores:
7. Para un vector cualquiera de tamaño "n", ordenado de menor a mayor , al aplicar `np.argmax(vector)` resulta en
8. ¿Qué es Broadcasting en Python y para que sirve?
9. Se tiene una matriz X de 3000 filas y 500 columnas, al aplicar `np.median(X, axis=1)` resulta en :
10. ¿Por qué es importante trabajar con tensores en ciencia de datos?
11. En NumPy se posee un vector definido de la siguiente forma: `v = np.array([1,2,4,5])` ¿El rango del Tensor es?
12. En data science llamamos \_\_\_\_ a la estructura de datos resultante de convertir datos o variables categóricas a una representación numérica ,comúnmente vectorial
13. ¿Un tensor de rango 0 también se puede llamar?
14. ¿Qué configuraciones de seguridad son necesarias hacer para descargar y subir archivos a y desde AWS S3?
15. ¿Por qué las operaciones vectorizadas son más eficientes que los ciclos?
16. ¿Qué características hacen que AWS Redshift sea superior a otras soluciones de DW en la nube?
17. ¿Qué es S3 y como se utiliza para trabajar con Data Lakes en AWS?
18. ¿Qué es Boto3 y como lo utilizamos en el curso?
19. ¿Qué es un Parquet file y para qué se utiliza en ingeniería de datos?
20. ¿Cuál es la diferencia entre AWS RDS y AWS Redshift?