



EXAMEN FINAL

1. Mencione cuales son los 2 tipos de objetos principales al trabajar con Pandas

R// Existen para Pandas tipos de objetos como: Series (listas, 1D), DataFrame (tablas, 2D)

2. Al usar Pandas loc y iloc nos sirven para:

R// Seleccionar los registros de las filas y columnas, iloc permite seleccionar los elementos en base a la posición, y loc permite seleccionar mediante etiquetas o declaraciones condicionales.

3. ¿Qué función usamos para agregar una fila de un dataframe de Pandas a otro?

R// Se puede utilizar el método.append para añadir una fila.append para añadir filas de otros dataframes al final del dataframe original, y devolver un nuevo dataframe.

4. Para calcular la transpuesta de un dataset, tanto en NumPy como en Pandas podemos usar

R// se puede utilizar df.transpose() ó df.T

5. Se posee una matriz con 1000 filas y 300 columnas, el aplicar sum(axis=0) sobre esta matriz , devuelve un vector de cuantos elementos?

R// Los elementos que devuelve son 300

6. Cuando trabajamos con dataframes, que elementos (de la estructura del dataframe) permiten usar nombres, etiquetas o identificadores:

R// Las filas y las columnas

7. Para un vector cualquiera de tamaño "n", ordenado de menor a mayor , el aplicar np.argmax(vector) resulta en

R// La cantidad máxima de los datos del vector

8. ¿Qué es Broadcasting en Python y para que sirve?

R// Broadcasting es la difusión masiva de información y una estrategia de numpy para operar con arreglos multidimensionales de diferentes dimensiones.

9. Se tiene una matriz X de 3000 filas y 500 columnas, el aplicar `np.median(X, axis=1)` resulta en :

R// Calcular la mediana de las filas

10. ¿Por qué es importante trabajar con tensores en ciencia de datos?

R// Porque los tensores son objetos matemáticos que almacenan valores numéricos y que pueden tener distintas dimensiones, un tensor de 1D es un vector, de 2D una matriz, de 3D un cubo etc.

11. En NumPy se posee un vector definido de la siguiente forma: `v = np.array([1,2,4,5])` ¿El rango del Tensor es?

R// el rango del tensor es 1

12. En data science llamamos ____ a la estructura de datos resultante de convertir datos o variables categóricas a una representación numérica ,comúnmente vectorial

R// Algoritmo

13. ¿Un tensor de rango 0 también se puede llamar?

R// Escalar

14. ¿Qué configuraciones de seguridad son necesarias hacer para descargar y subir archivos a y desde AWS S3?

R// Administrar una configuración de IAM con asignación de usuario con accesos que correspondan

15. ¿Por qué las operaciones vectorizadas son más eficientes que los ciclos?

R// Actúan con mayor rapidez y menos capacidad las operaciones vectorizadas vs los ciclos

16. ¿Qué características hacen que AWS Redshift sea superior a otra soluciones de DW en la nube?

R// AWS es masivo, dinámico, seguro, permite aprovechar al máximo las últimas tecnologías para impulsar la innovación.

17. ¿Qué es S3 y como se utiliza para trabajar con Data Lakes en AWS?

R// Es un espacio de almacenamiento utilizado en la nube, se le da uso por la forma ilimitada del escalado.

18. ¿Qué es Boto3 y como lo utilizamos en el curso?

R// Función que integra Python con AWS, utilizado para el S3

19. ¿Qué es un Parquet file y para qué se utiliza en ingeniería de datos?

R// Es un formato de almacenamiento de código abierto para almacenar datos, se almacenan con la extensión.parquet, es un formato muy estructurado, proporciona optimizaciones para acelerar las consultas.

20. ¿Cuál es la diferencia entre AWS RDS y AWS Redshift?

R// RDS se dirige a bases de datos pequeñas y AWS Redshift administra datos en enormes cantidades combinado con queries.