**Data Science Intermedio**

Evaluaci**ón diagnóstica**

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| **Nombre**: |  |  |  | **Fecha**: | 07/10/2022 |
|  | Apellido Paterno | Apellido Materno | Nombre(s) |  |  |
| **Tiempo: 40 min.** | **Calificación:** | | | |  |

|  |
| --- |
| **Objetivo:** El participante identificará su dominio sobre los temas básicos de ciencia de datos y Python necesarios para abordar este curso nivel intermedio. Asimismo, se reforzarán los conceptos que sean necesarios. |

Instrucciones: Responde las siguientes preguntas y, cuando lo amerite, puedes escribir código o hacer captura de pantalla.

1. En Python, ¿cuál es la diferencia entre un diccionario, una lista, una tupla y un data-frame? Puedes redactar las diferencias, o bien, escribir cómo crearías cada uno de estas estructuras de datos en Python.

**Valor: 3 puntos**

2. ¿Qué es la ciencia de datos?, ¿A qué se refiere el término Big Data?

**Valor: 2 puntos**

3. Realiza lo siguiente en Python.

3.1. Genera un numpy array aleatorio de 100 filas y 2 columnas con números enteros entre 1 y 50.  
**(1 punto)**.

3.2. Convierte el numpy array en un data-frame de pandas e imprime los primeros 10 elementos.  
**(1 punto)**.

3.3. Calcula, para los valores de la segunda columna, su promedio, desviación estándar y moda.  
**(1 punto)**.

3.4. Despliega dos gráficas: un histograma de frecuencias de los valores de la segunda columna con matplotlib y una gráfica de dispersión con seaborn donde el eje x corresponda a la primer columna y el eje y a la segunda columna.  
**(2 puntos)**.

**Valor: 5 puntos**