

Módulo 2  
Clase 1

# Obtención y Preparación de Datos

Especialización en Ciencia de Datos

# Presentación del módulo



Duración:  
24 horas



Competencia del módulo:  
“Codificar piezas de software de  
baja/mediana complejidad en  
lenguaje Python para resolver  
una problemática de acuerdo al  
uso de Librerías”

# Aprendizaje esperado

1. Manipula datos utilizando estructuras de vectores y matrices utilizando biblioteca Numpy para resolver un problema.
2. Manipular datos utilizando estructuras de Series y DataFrames utilizando biblioteca Pandas para resolver un problema.
3. Recuperar datos desde distintas fuentes utilizando librerías utilitarias de Python para su posterior utilización.
4. Aplica técnicas de limpieza y preparación de datos utilizando librerías Python para su depuración.
5. Aplica técnicas de unión, combinación y redimensionamiento de estructuras de dato utilizando librerías de Python para el reacomodo de datos.



# Evaluación del Módulo

- **20% Pruebas de conocimiento**
- **30% Ejecuciones prácticas parciales**
- **10% Participación en clases**
  - **40% Trabajo final del módulo**

# Contenido



1. Librería Numpy
2. Librería Pandas
3. Obtención de datos desde archivos

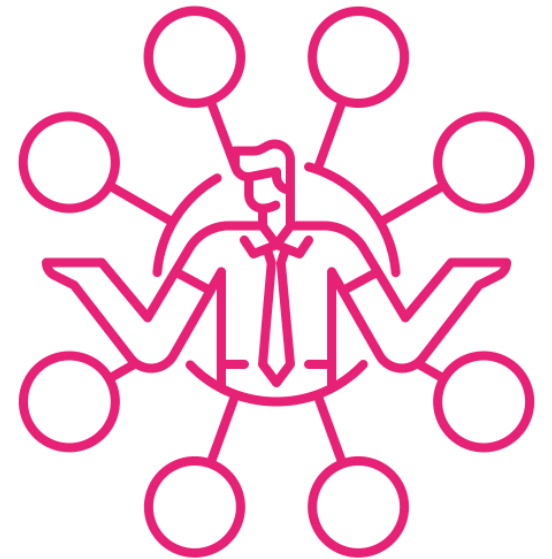


4. Limpieza de datos
5. Data Wrangling I
6. Data Wrangling II

# Consejos para el éxito!

- Practicar lo más que puedan (recomendación: media hora por día)
- No sentirse frustrado o con miedo, (Es parte de aprender tener errores al inicio)
- Ser parte de una comunidad, compartir errores y soluciones

- **“Tu FUTURO se decide por lo que haces HOY, no MAÑANA”**



# Pre-procesamiento de Datos



