

# Pandas en Python



---

La herramienta esencial para el análisis de datos

Ing. Ms Rudy Luis Manzaneda Veizaga

# ¿Qué es Pandas?

---

01

- Biblioteca de código abierto escrita en Python.

02

- Creada en 2008 por Wes McKinney.

03

- Popular entre científicos de datos, analistas y estadísticos.

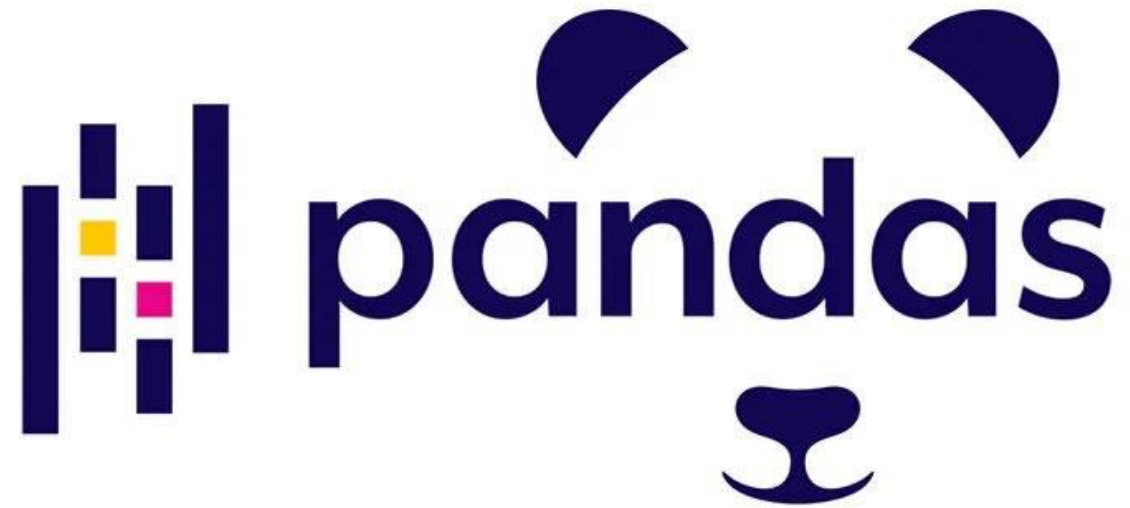
04

- Proporciona herramientas para explorar, transformar y analizar datos de forma estructurada.

# Usos comunes de Pandas

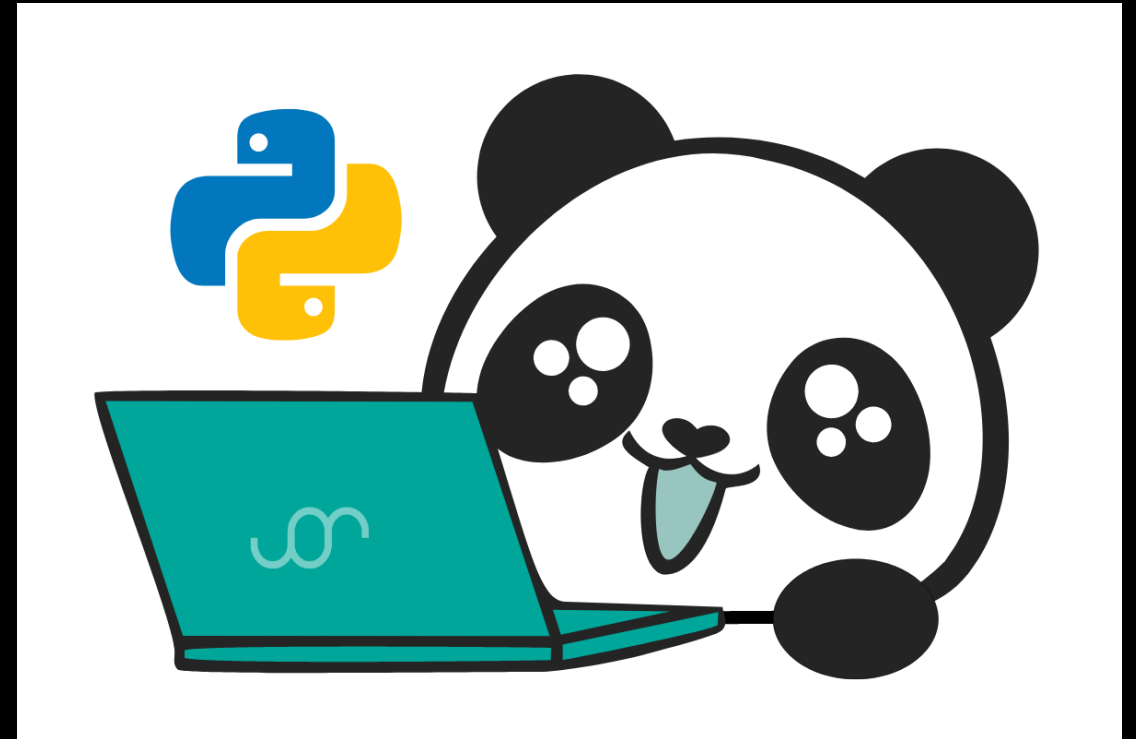
---

- 1. Administrar y limpiar conjuntos de datos.
- 2. Realizar análisis de datos complejos.
- 3. Generar informes comprensibles.
- 4. Construir modelos de aprendizaje automático.



# Características clave de Pandas

- - Estructuras de datos flexibles (Series y DataFrames).
- - Indexación rápida y sencilla.
- - Métodos avanzados de limpieza de datos.
- - Herramientas para manipulación de datos (clasificación, filtrado, agregación).
- - Visualización de datos integrada.



# Estructuras principales de Pandas

- 1. Series
  - - Objeto unidimensional similar a un array.
  - - Contiene un índice asociado para identificar elementos.
- 2. DataFrames
  - - Estructura bidimensional (filas y columnas).
  - - Similar a una hoja de cálculo, ideal para trabajar con grandes conjuntos de datos.



# Ventajas de Pandas

- Fácil de aprender: Basado en Python, con una sintaxis intuitiva.
- Rápido y eficiente: Escrito en Cython para un rendimiento óptimo.
- Versátil: Soporta múltiples formatos de datos y tareas.
- Confiable y mantenido: Comunidad activa y actualizaciones regulares.
- De código abierto: Gratis y accesible para todos.

# Limitaciones de Pandas

- - Depuración compleja: Identificar errores puede ser difícil.
- - No ideal para datos muy grandes: Recomendado usar Dask o PySpark.
- - Curva de aprendizaje: Requiere conocimientos previos de Python.
- - Limitado en aprendizaje profundo: Mejor complementarlo con TensorFlow o PyTorch.



# Conclusión

- - Pandas es una herramienta fundamental en el análisis de datos.
- - Su flexibilidad y eficiencia la convierten en una biblioteca indispensable.
- - Ideal para tareas de manipulación, limpieza y análisis en ciencia de datos.

