

Pandas

Introducción a la biblioteca Pandas

CertiDevs

Índice de contenidos

1. Introducción a Pandas	1
2. Propósito de Pandas en el análisis de datos.....	1
3. Relación entre Pandas y NumPy.....	1
4. Instalación de Pandas	2

1. Introducción a Pandas

Pandas es una biblioteca de Python de código abierto que proporciona estructuras de datos y herramientas de análisis de datos de alto rendimiento y fácil de usar.

Fue desarrollada por Wes McKinney en **2008** y desde entonces se ha convertido en una de las bibliotecas más populares y ampliamente utilizadas en el análisis de datos y la ciencia de datos en Python.

[Sitio oficial de Pandas](#)

2. Propósito de Pandas en el análisis de datos

Pandas está diseñado para trabajar con datos estructurados y semiestructurados, como **tablas** y **series temporales**.

Sus **principales características** incluyen:

- Estructuras de datos flexibles y potentes: **Series** y **DataFrames**.
- Herramientas para cargar datos desde diferentes formatos de archivo, como CSV, Excel, SQL y JSON.
- Funciones de manipulación y limpieza de datos, como filtrar, ordenar, agregar y unir.
- Funciones estadísticas y de cálculo integradas.
- Integración con otras bibliotecas de Python, como NumPy, Matplotlib y scikit-learn.

Pandas facilita el análisis y la manipulación de datos en Python, permitiendo a los científicos de datos y analistas centrarse en extraer información valiosa de los datos en lugar de preocuparse por problemas técnicos.

3. Relación entre Pandas y NumPy

NumPy es otra biblioteca popular de Python para realizar operaciones numéricas y matriciales.

Pandas se basa en **NumPy** y utiliza sus arrays y funciones para almacenar y operar con datos. Sin embargo, **Pandas extiende** las capacidades de **NumPy** al proporcionar estructuras de datos etiquetadas y herramientas específicas para el análisis de datos.

Aunque Pandas y NumPy pueden usarse de forma independiente para ciertas tareas, a menudo se utilizan juntas en proyectos de análisis de datos y ciencia de datos.

NumPy es especialmente útil para operaciones matemáticas y de álgebra lineal, mientras que **Pandas** es la elección preferida para **manipular** y **analizar datos estructurados** y semiestructurados.

4. Instalación de Pandas

Antes de comenzar a trabajar con Pandas, es necesario instalarlo en su entorno de Python.

Puede **instalar Pandas** utilizando **pip**, el administrador de paquetes de Python, con el siguiente comando:

```
pip install pandas
```

Una vez que Pandas esté instalado, puede importarlo en su código de Python junto con NumPy (ya que Pandas depende de NumPy) utilizando las siguientes declaraciones de importación:

```
import numpy as np
import pandas as pd
```