

Pandas 教程



Pandas 是 **Python** 语言的一个扩展程序库，用于数据分析。

Pandas 名字衍生自术语 "**panel data**"（面板数据）和 "**Python data analysis**"（Python 数据分析）。

Pandas 是一个开源、BSD 许可的库，提供高性能、易于使用的数据结构和数据分析工具。

Pandas 一个强大的分析结构化数据的工具集，基础是 **Numpy**（提供高性能的矩阵运算）。

Pandas 可以从各种文件格式比如 CSV、JSON、SQL、Microsoft Excel 导入数据。

Pandas 可以对各种数据进行运算操作，比如归并、再成形、选择，还有数据清洗和数据加工特征。

学习本教程前你需要了解

在开始学习 **Pandas** 教程之前，我们需要具备基本的 **Python** 基础，如果你对 **Python** 还不了解，可以阅读我们的教程：

- [Python 2.x 版本](#)
- [Python 3.x 版本](#)

Pandas 应用

Pandas 的主要数据结构是 **Series**（一维数据）与 **DataFrame**（二维数据）。

Pandas 广泛应用在学术、金融、统计学等各个数据分析领域。

数据结构

Series 是一种类似于一维数组的对象，它由一组数据（各种 **Numpy** 数据类型）以及一组与之相关的数据标签（即索引）组成。

DataFrame 是一个表格型的数据结构，它含有一组有序的列，每列可以是不同的值类型（数值、字符串、布尔型值）。**DataFrame** 既有行索引也有列索引，它可以被看做由 **Series** 组成的字典（共用一个索引）。

相关链接

- **Pandas** 官网 <https://pandas.pydata.org/>
- **Pandas** 源代码: <https://github.com/pandas-dev/pandas>

📝 点我分享笔记

