4.1 Pandas 介绍

2023 年 9 月 18 日

数据分析常用库——Pandas 库

它是一个擅长用来对数据进行收集、清洗和分析的 python 第三方库。

早期多用于时间序列的分析，在金融领域使用较多，目前已经扩展到所有需要对数据进行处理的场景中。

1. Pandas 的作者

作为 pandas 的主要发明者，韦斯·麦金尼 (Wes McKinney) 讲述了他的个人经历：

"I'm an American computer programmer and the Director of Ursa Labs. I studied theoretical mathematics at MIT (graduating in late 2006) before becoming very interested in programming and tools for data analysis, especially for industry use cases, in 2007.

From August 2007 to July 2010, I worked on the front office quant research team at AQR Capital Management, a large quantitative investment manager in Greenwich, CT. During this time, I led a very successful effort to migrate research and production model building and research processes to the Python programming language. I started building pandas on April 6, 2008, as part of a skunkworks effort to reproduce some econometric research in Python. As part of my work, we formed a new Research Development team for the global macro group to drive software innovation in the front office.

I joined the PhD program in the Statistical Science Department at Duke University before taking leave in Summer 2011 to explore ways to develop open source software (such as pandas) in a sustainable way. I discovered that entrepreneurship often makes more sense than consulting to fund open source with more leverage."

— Wes McKinney, The creater of Pandas.

读完之后，你能回答一下问题么：1. pandas 是什么时间发明出来的？2. 最早 pandas 是为了做什么用的?

2. 你的 pandas 版本

如何加载 pandas 呢？想必你已经知道了。

```
import pandas as pd
```

如果需要了解你的 pandas 的版本号，请输入以下命令:

```
pd.__version__
```

3. cheatsheet

在网上搜索关键词"cheatsheet + pandas"，可以搜索到一些制作精良的 pandas 快速手册，方便随时查找所需的方法和用法。

请点击，附录：Pandas-Cheat-Sheet.pdf 文件