

Библиотека Pandas

Козлова Юлия Ханифовна

Pandas – это пакет языка программирования **Python**, который предоставляет быстрые, гибкие и выразительные структуры данных, предназначенные для того, чтобы сделать работу с **реляционными** или **размеченными** данными простой и интуитивно понятной.

Документация: <https://pandas.pydata.org/docs/index.html>

Код на GitHub: <https://github.com/pandas-dev/pandas>

Где используется?

Библиотека **Pandas** широко используется в академических и коммерческих средах.

Сферы применения

статистика

веб-аналитика

финансы

экономика

реклама

Возможности библиотеки

- быстрый и эффективный объект `DataFrame` для манипуляции над данными со встроенной индексацией
- возможность одновременной обработки данных из разных файлов
- возможность загрузки и обработки данных различных форматов: `csv`, текстовые файлы, `Microsoft Excel`, БД `SQL`, `HDF5`
- готовые методы для различных манипуляций с таблицами: добавление, удаление, модификация, агрегирование данных
- готовые методы для операций с пропущенными значениями, выявление и устранение «проблемных» данных

Как установить?

- с использованием `conda`
`conda install pandas`
- с использованием `pip`
`pip install pandas`
- если у вас вдруг до сих пор нет Python ... то через `Anaconda`

Как импортировать?

классика, но так никто не делает 😓

```
import pandas
```

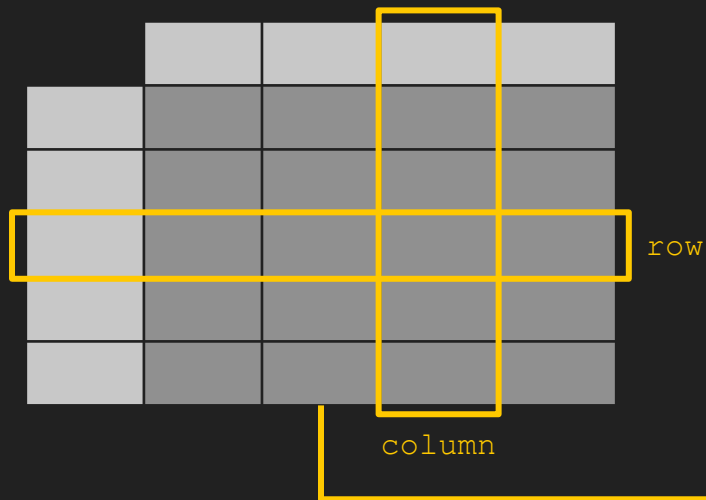
а вот так вот делают! почему?

```
import pandas as pd
```

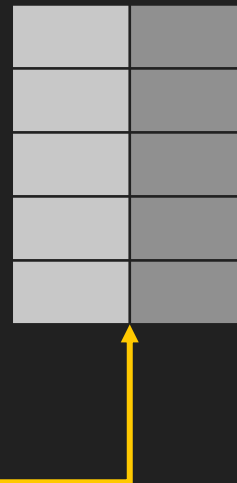
Это база...



DataFrame



Series



Сегодня в программе

1. Структуры данных `DataFrame`, `Series`
2. Доступ к данным `DataFrame`
3. Импорт и экспорт данных
4. Добавление, модификация и удаление столбцов и строк
`DataFrame`
5. Обработка отсутствующих значений
6. Фильтрация
7. Группировка

Спасибо за внимание!
Вопросы?