

Библиотека Pandas

ВВЕДЕНИЕ

Библиотека pandas
PANel DAta –
панельные данные

Востребованные функции pandas

- группировка данных по определенным признакам
- создание сводных таблиц из нескольких,
- очистка данных от дубликатов и невалидных строк или столбцов,
- вывод значений по фильтрам,
- использование агрегирующих функций (сумма, среднее арифметическое и т.д.),
- визуализация данных.

Быстродействие

Библиотека **Pandas** создана на основе библиотеки **NumPy**, что является большим плюсом в ее производительности.

Основные структуры данных

Series – одна колонка таблицы

DataFrame – таблица целиком

Установка Pandas

```
pip install pandas
```

Подключение Pandas

```
import pandas as pd
```

Объект Series

Series	
index	value
0	12
1	-4
2	7
3	9

Series состоит из двух связанных между собой одномерных массивов: значений и индексов.

Индекс может быть не только числовой.

По умолчанию: индекс как одномерном массиве - целые числа с 0.

Объект DataFrame

	Column 1	Column 2	Column 3	Column 4	Column 5
0	Value 1.1	Value 2.1	Value 3.1	Value 4.1	Value 5.1
1	Value 1.2	Value 2.2	Value 3.2	Value 4.2	Value 5.2
2	Value 1.3	Value 2.3	Value 3.3	Value 4.3	Value 5.3
3	Value 1.4	Value 2.4	Value 3.4	Value 4.4	Value 5.4
4	Value 1.5	Value 2.5	Value 3.5	Value 4.5	Value 5.5

Diagram illustrating the structure of a DataFrame:

- index**: Points to the first column (index values).
- values**: Points to the second column (data values).
- Series**: Points to the fourth column (a single column is a Series).
- columns**: Points to the header row (all columns are labeled).

DataFrame – двумерная таблица, в которой строки и столбцы имеют имена.

Каждый столбец таблицы фактически представляет из себя объект Series.

Как считывать данные

CSV-файлы

Считывание: `pd.read_csv("имя_файла.csv")`

Сохранение: `df.to_csv("имя_файла.csv")`

Excel-файлы (формат 2007+)

Считывание: `pd.read_excel("имя_файла.xlsx")`

Сохранение: `df.to_excel("имя_файла.xlsx")`

`df` объект типа DataFrame

Некоторые возможности

Группировка данных

`groupby`

Сортировка данных

`sort_values`

Конкатенация данных

`concat`

Объединение связанных таблиц

`merge`

Заполнение пропущенных значений

`fillna`

Преобразование типов данных

`astype`