

Pandas

Pandas – пакет для языка программирования Python, который предназначен для работы с табличными данными. Пакет дает возможность строить сводные таблицы, выполнять группировки, предоставляет удобный доступ к данным, взаимодействует с пакетом matplotlib и дает возможность с помощью него отображать данные на графиках.

Основными типами данных Pandas являются Series – проиндексированный одномерный массив значений, DataFrame – проиндексированный многомерный массив значений, каждый столбец которого является элементом типа Series. Пакет позволяет работать с множеством типов источников данных, таких как SQL, текстовые данные, Excel файлы, HTML.

Примеры использования:

```
>>> import pandas as pd
>>>
>>> df = pd.DataFrame({
...     'country': ['Kazakhstan', 'Russia', 'Belarus', 'Ukraine'],
...     'population': [17.04, 143.5, 9.5, 45.5],
...     'square': [2724902, 17125191, 207600, 603628]
... }, index=['KZ', 'RU', 'BY', 'UA'])
>>> df
```

	country	population	square
KZ	Kazakhstan	17.04	2724902
RU	Russia	143.50	17125191
BY	Belarus	9.50	207600
UA	Ukraine	45.50	603628

```
>>> df.index = ['KZ', 'RU', 'BY', 'UA']
>>> df.index.name = 'Country Code'
>>> df
```

	country	population	square
Country Code			
KZ	Kazakhstan	17.04	2724902
RU	Russia	143.50	17125191
BY	Belarus	9.50	207600
UA	Ukraine	45.50	603628

```
>>>
>>> titanic_df = pd.read_csv('titanic.csv')
>>> print(titanic_df.head())
```

	PassengerID	Name	PClass	Age	\
0	1	Allen, Miss Elisabeth Walton	1st	29.00	
1	2	Allison, Miss Helen Loraine	1st	2.00	
2	3	Allison, Mr Hudson Joshua Creighton	1st	30.00	
3	4	Allison, Mrs Hudson JC (Bessie Waldo Daniels)	1st	25.00	

Sex	Survived	SexCode
0	female	1
1	female	0
2	male	0
3	female	0

Книги:

- Chen D. Y. *Pandas for Everyone: Python Data Analysis*. // Addison-Wesley, 2018.
- McKinney W. *Python for Data Analysis: Data Wrangling with Pandas, NumPy, and IPython*. // O'Reilly, 2017.
- Molin S. *Hands-On Data Analysis with Pandas: Efficiently perform data collections, wrangling, analysis, and visualization using Python*. // Packt Publishing, 2019.
- Petrou T. *Pandas Cookbook: Recipes for Scientific Computing, Time Series Analysis and Data Visualization using Python*. // Packt Publishing, 2017.
- Klosterman S. *Data Science Projects with Pandas: A case study approach to successful data science projects using Python, pandas, and scikit-learn*. // Packt Publishing, 2019.

Сетевые ресурсы:

- *Python Data Analysis Library – pandas*. <https://pandas.pydata.org>
- *pandas Wiki*. [https://en.wikipedia.org/wiki/Pandas_\(software\)](https://en.wikipedia.org/wiki/Pandas_(software))
- *Top 8 pandas resources*. <https://www.dataschool.io/best-python-pandas-resources>

Консультация по вопросам использования:

vtasks@jscs.ru