

Библиотека Pandas

Светлана Обухова

Data Analyst в Ernst&Young

Skillbox

Библиотека Pandas

Skillbox

Pandas

Цели модуля



Знакомство с библиотекой
Pandas



Чтение файлов с помощью
Pandas



Подсчёт простых статистик
в Pandas

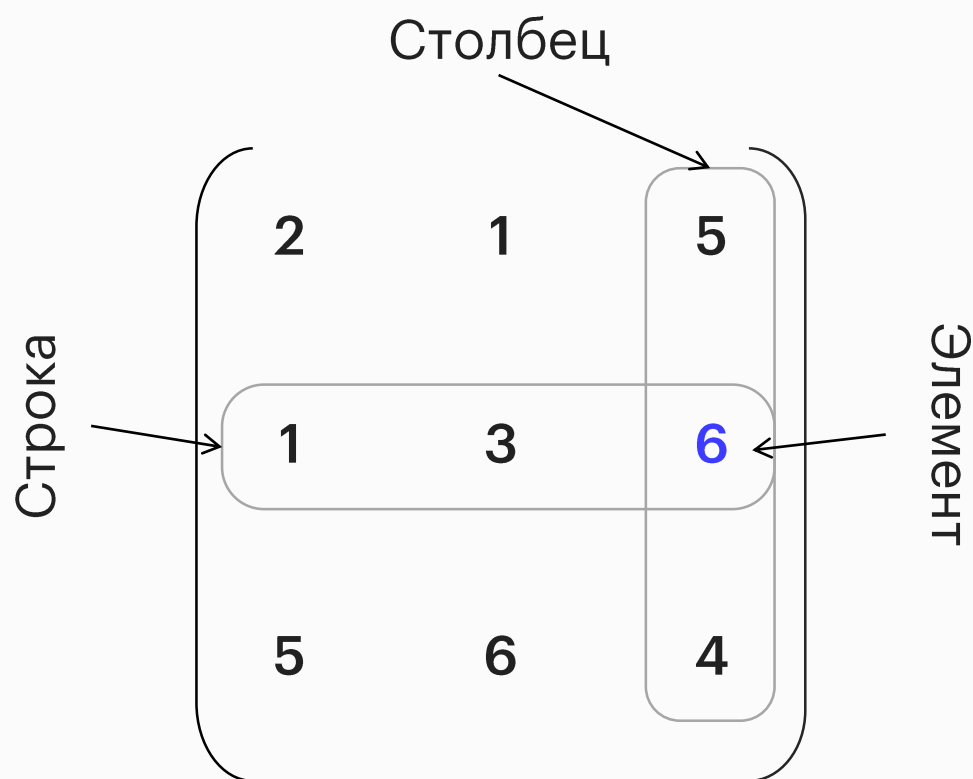


Объединение датафреймов
в Pandas

Pandas — программная библиотека на языке Python для обработки и анализа данных.

Датафрейм (DataFrame) —
двумерная структура
данных, матрица,
содержащая метки
для строк и столбцов.

Матрица



Автомобильный код региона

City	Code
Tomsk	70
Omsk	55

Pandas

DataFrame

```
class pandas.DataFrame(  
    data=None,  
    index=None,  
    columns=None,  
    dtype=None,  
    copy=None)
```


pd.DataFrame(...)

Параметр	Описание
data	Массив ndarray, словарь (dict) или другой DataFrame
index	Список меток для записей (имена строк таблицы)
columns	Список меток для полей (имена столбцов таблицы)
dtype	Объект numpy.dtype , определяющий тип данных
copy	Создаёт копию массива данных, если параметр равен True , в ином случае ничего не делает

Series — маркированная
одномерная структура
данных, представленная
в виде таблицы с одной
строкой.

Списки

Length=5					
	<div>1</div>	<div>2</div>	<div>5</div>	<div>1</div>	<div>3</div>
Index	0	1	2	3	4
Negative index	- 5	- 4	- 3	- 2	- 1

Pandas

Series

Length=5

Название
метки

