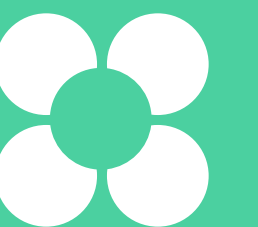


# Основы pandas

Константин Башевой  
Аналитик-разработчик в Яндексе



# Константин Башевой

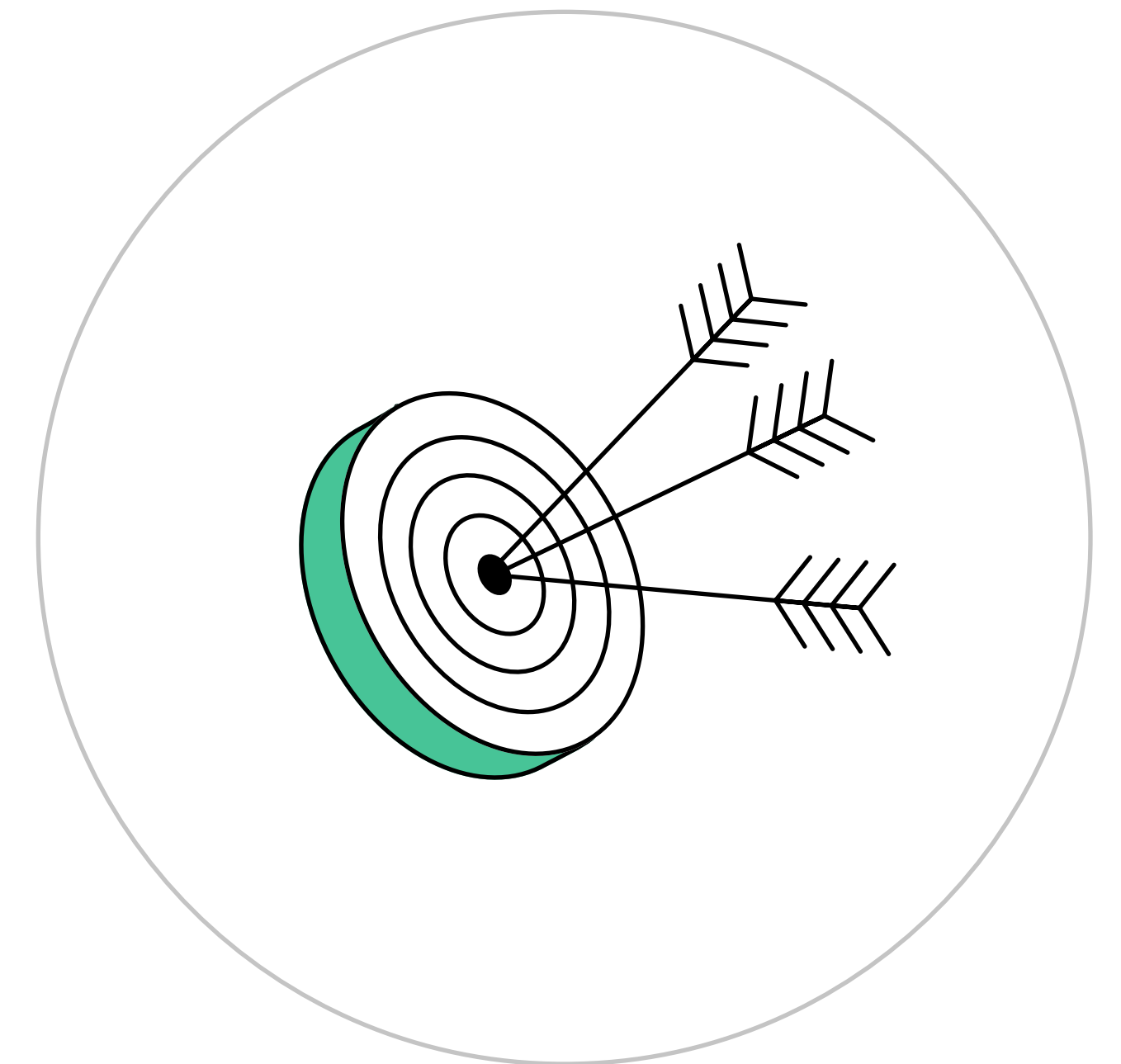
О спикере:

- аналитик-разработчик в Яндексе
- опыт работы за последние 10 лет:  
Rambler&Co, Ростелеком, Яндекс



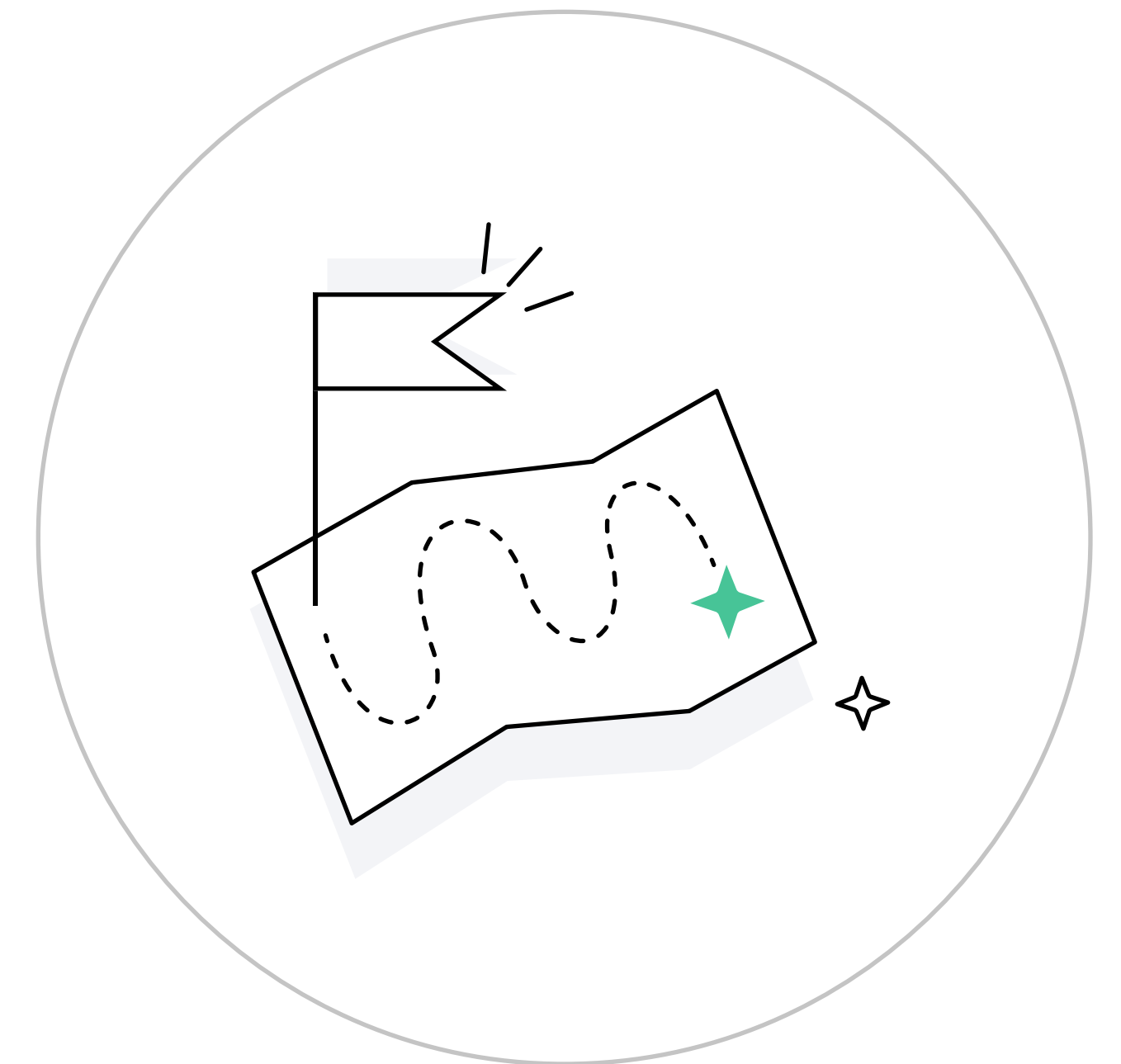
# Цели занятия

- Узнать, что такое библиотека pandas
- Создать презентацию в Jupyter Notebook
- Научиться выполнять базовые операции в pandas



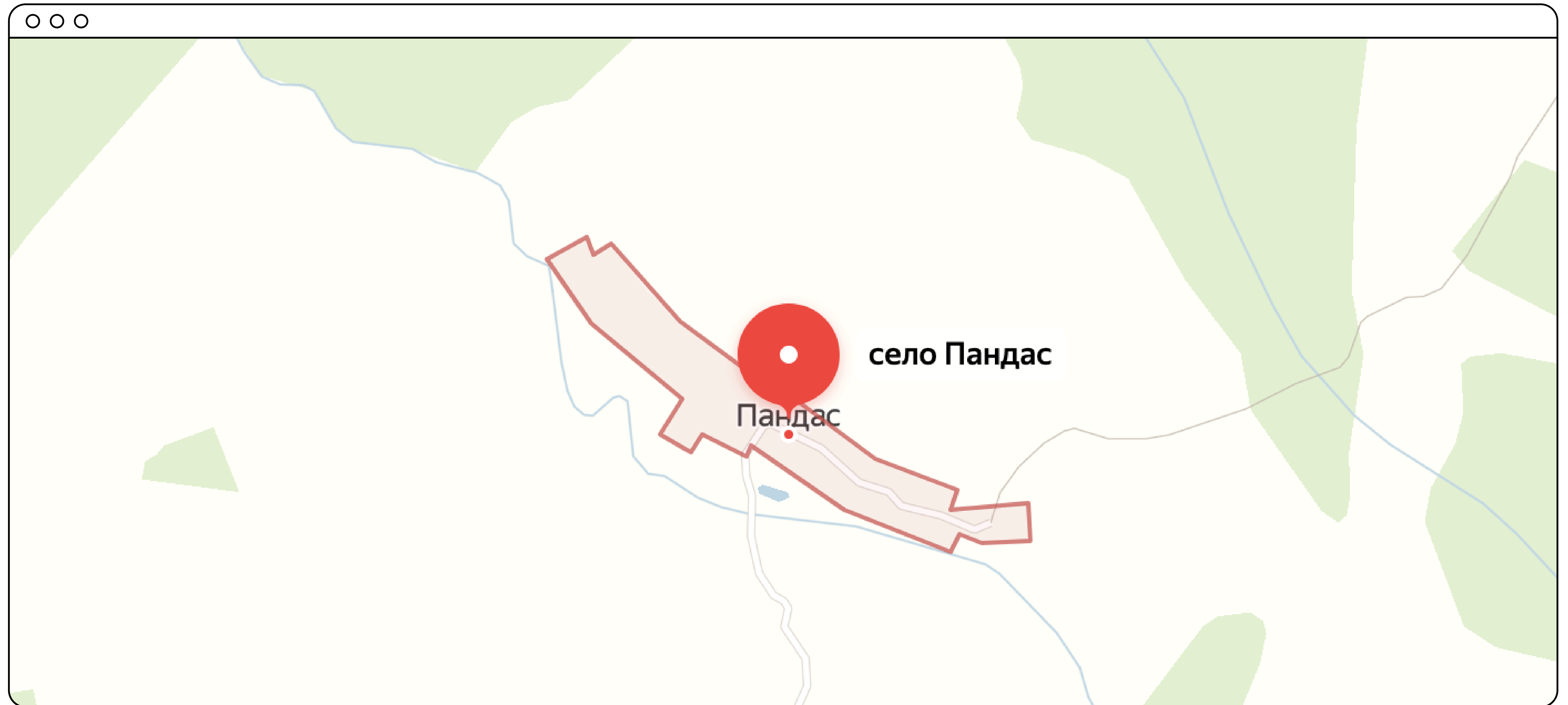
# План занятия

- 1 Обзор возможностей библиотеки pandas
- 2 Создание презентации из кода в Jupyter Notebook
- 3 Базовые операции в pandas



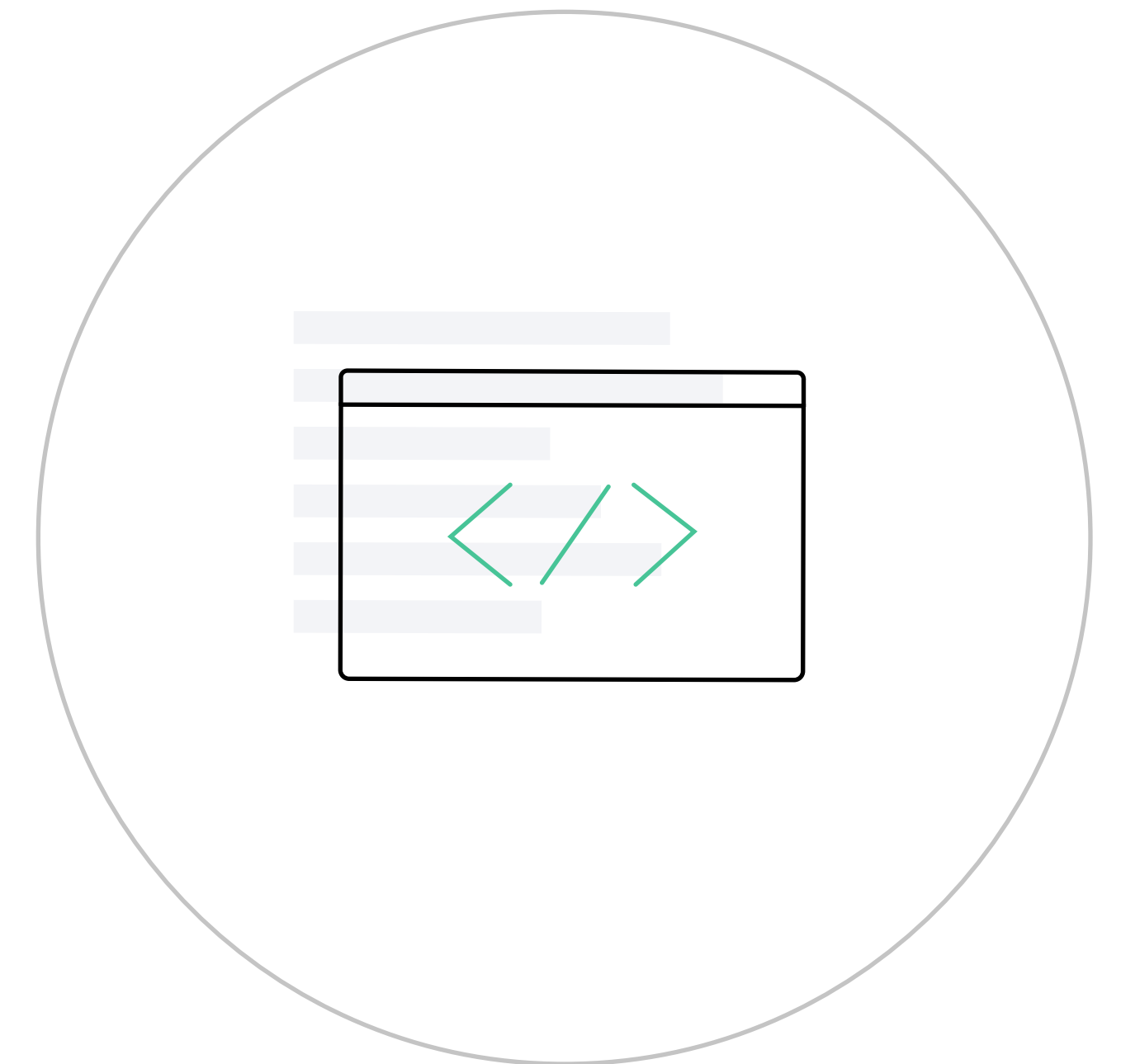
# Обзор возможностей библиотеки pandas

# Pandas = panel + data



# Основные источники, которые понимает pandas

- CSV- и Excel-файлы
- Буфер обмена (Ctrl + C)
- HTML-страницы (тег <table>)
- JSON-файлы
- Формат Parquet
- SQL-запросы



# Все вычисления в pandas происходят в оперативной памяти

DataFrame хранится в оперативной памяти.

Пример:

- достаточно иметь ноутбук с 4 ГБ RAM
- 2 ГБ свободной памяти
- с данными до 1,5 ГБ можно работать





# Pandas и Jupyter Notebook

В сочетании с Jupyter Notebook  
pandas даёт отличные  
возможности для просмотра  
промежуточных результатов

ooo

```
data = pd.read_csv('power.csv')  
data.head()
```

	country	year	quantity	category
0	Austria	1996	5.0	1
1	Austria	1995	17.0	1
2	Belgium	2014	0.0	1
3	Belgium	2013	0.0	1
4	Belgium	2012	35.0	1

# Pandas и SQL

В pandas содержится огромное количество методов, которые реализуют возможности SQL.

Если ваши данные помещаются в оперативную память, то pandas заменит базу данных.

- GROUP BY — groupby с любыми функциями
- JOIN — join и merge
- ORDER BY — sort\_values



# Цепочки вычислений

С помощью pandas можно  
выстраивать цепочки вычислений

ooo

```
(data.groupby('country').count()  
 .sort_values('quantity')  
 .query('category > 16')  
 .reset_index()[['country', 'quantity']]  
)
```

	country	quantity
0	Yemen Arab Rep. (former)	45
1	Yemen, Dem. (former)	61
2	Pacific Islands (former)	68
3	Antarctic Fisheries	90
4	German Dem. R. (former)	106

# Сложные вычисления

В pandas вы можете пользоваться не только встроенными методами, но и производить вычисления любой сложности

o o o

```
def baltic(country):  
    """Объединение стран Прибалтики"""  
  
    if country in ['Lithuania', 'Latvia', 'Estonia']:  
        return 'Прибалтика'  
  
    return 'Other'
```

```
data['baltic'] = data.country.apply(baltic)  
data.baltic.value_counts()
```

```
Other          1162176  
Прибалтика      27306  
Name: baltic, dtype: int64
```

# Создание презентации из кода в Jupyter Notebook

# Заключение



# Итоги занятия

## В этой теме мы:

- Узнали про возможности библиотеки pandas
- Создали презентацию в Jupyter Notebook
- Научились выполнять базовые операции в pandas



# Основы pandas

Константин Башевой  
Аналитик-разработчик в Яндексе

