



# ОСНОВЫ PYTHON ДЛЯ DATA SCIENCE

ОСВОЙТЕ PYTHON И СРАЗУ ПРИСТУПАЙТЕ К РЕШЕНИЮ  
ПРИКЛАДНЫХ ЗАДАЧ, СВЯЗАННЫХ С АНАЛИЗОМ ДАННЫХ



# ОСНОВЫ PYTHON ДЛЯ DATA SCIENCE

Кеннеди Берман



Санкт-Петербург · Москва · Минск

2023

ББК 32.973.2-018.1  
УДК 004.43  
Б50

## **Берман Кеннеди**

Б50 Основы Python для Data Science. — СПб.: Питер, 2023. — 272 с.: ил. — (Серия «Библиотека программиста»).

ISBN 978-5-4461-2251-6

Python — язык программирования № 1 для машинного обучения и Data Science. Но как же сложно решить, с чего начать изучение Python, ведь у него огромный инструментарий! Кеннеди Берман фокусируется на тех навыках программирования, которые понадобятся вам для решения задач в области Data Science и машинного обучения.

Вы познакомитесь с блокнотами Jupyter — лучшей средой для профессиональной работы с данными. После этого перейдете к ключевым библиотекам, которые упрощают процесс математических вычислений, визуализации, решение задач машинного обучения и обработки естественного языка. После этого, овладев основами, вы перейдете к продвинутым техникам, позволяющим решать более сложные задачи.

**16+** (В соответствии с Федеральным законом от 29 декабря 2010 г. № 436-ФЗ.)

ББК 32.973.2-018.1  
УДК 004.43

Права на издание получены по соглашению с Pearson Education Inc.

Все права защищены. Никакая часть данной книги не может быть воспроизведена в какой бы то ни было форме без письменного разрешения владельцев авторских прав.

Информация, содержащаяся в данной книге, получена из источников, рассматриваемых издательством как надежные. Тем не менее, имея в виду возможные человеческие или технические ошибки, издательство не может гарантировать абсолютную точность и полноту приводимых сведений и не несет ответственности за возможные ошибки, связанные с использованием книги.

Издательство не несет ответственности за доступность материалов, ссылки на которые вы можете найти в этой книге. На момент подготовки книги к изданию все ссылки на интернет-ресурсы были действующими.

ISBN 978-0136624356 англ.

Authorized translation from the English language edition, entitled Foundational Python for Data Science, 1st Edition, by Kennedy Behrman, published by Pearson Education, Inc, publishing as Addison Wesley Professional  
© 2022 Pearson Education, Inc.

ISBN 978-5-4461-2251-6

© Перевод на русский язык ООО «Прогресс книга», 2022  
© Издание на русском языке, оформление ООО «Прогресс книга», 2022  
© Серия «Библиотека программиста», 2022

# Краткое содержание

Предисловие .....	16
Благодарности .....	18
Об авторе .....	19
От издательства .....	20

## ЧАСТЬ I PYTHON В JUPYTER NOTEBOOK

<b>Глава 1.</b> Введение в Jupyter Notebook .....	22
<b>Глава 2.</b> Основы Python .....	35
<b>Глава 3.</b> Последовательности .....	48
<b>Глава 4.</b> Прочие структуры данных .....	61
<b>Глава 5.</b> Управление выполнением .....	82
<b>Глава 6.</b> Функции .....	95

## ЧАСТЬ II БИБЛИОТЕКИ DATA SCIENCE

<b>Глава 7.</b> Библиотека NumPy .....	114
<b>Глава 8.</b> Библиотека SciPy .....	134

<b>Глава 9.</b> Библиотека Pandas .....	145
<b>Глава 10.</b> Библиотеки визуализации данных .....	173
<b>Глава 11.</b> Библиотеки машинного обучения .....	193
<b>Глава 12.</b> Инструментарий естественного языка (NLTK) .....	199

**ЧАСТЬ III  
СРЕДНИЙ УРОВЕНЬ PYTHON**

<b>Глава 13.</b> Функциональное программирование .....	214
<b>Глава 14.</b> Объектно-ориентированное программирование .....	229
<b>Глава 15.</b> Прочие вопросы .....	245
<b>Приложение А.</b> Ответы к вопросам в конце глав .....	261

# Оглавление

<b>Предисловие</b> .....	16
Примеры кода .....	17
<b>Благодарности</b> .....	18
<b>Об авторе</b> .....	19
<b>От издательства</b> .....	20

## ЧАСТЬ I PYTHON В JUPYTER NOTEBOOK

<b>Глава 1.</b> Введение в Jupyter Notebook .....	22
Выполнение операторов в Python .....	23
Блокноты Jupyter .....	24
Блокноты Google Colab .....	24
Текстовые ячейки Colab .....	26
LaTeX .....	29
Ячейки кода Colab .....	29
Файлы Colab .....	29
Управление документами Colab .....	31
Фрагменты кода Colab .....	31
Существующие коллекции .....	32
Системные псевдонимы .....	33
Магические функции .....	33
Резюме .....	34
Вопросы для закрепления .....	34

<b>Глава 2. Основы Python</b> .....	35
Основные типы в Python .....	36
Высокоуровневые и низкоуровневые языки .....	37
Операторы .....	38
Выполнение базовых математических операций .....	46
Классы, объекты и точечная нотация .....	47
Резюме .....	47
Вопросы для закрепления .....	47
<b>Глава 3. Последовательности</b> .....	48
Общие операции .....	49
Проверка принадлежности .....	49
Индексирование .....	50
Слайсинг .....	50
Сбор информации .....	51
Математические операции .....	52
Списки и кортежи .....	53
Создание списков и кортежей .....	53
Добавление и удаление элементов списка .....	54
Распаковка .....	56
Сортировка списков .....	57
Строки .....	57
Диапазоны .....	59
Резюме .....	60
Вопросы для закрепления .....	60
<b>Глава 4. Прочие структуры данных</b> .....	61
Словари .....	62
Создание словарей .....	62
Доступ, добавление и обновление с помощью ключей .....	63
Удаление элементов из словарей .....	64
Представления словаря .....	65

---

Проверка наличия ключа в словаре . . . . .	68
Метод get . . . . .	69
Допустимые типы ключей . . . . .	70
Метод hash . . . . .	71
Множества . . . . .	72
Операции над множествами . . . . .	75
Замороженные множества . . . . .	80
Резюме . . . . .	81
Вопросы для закрепления . . . . .	81
<b>Глава 5. Управление выполнением . . . . .</b>	<b>82</b>
Составные операторы . . . . .	83
Структура составных операторов . . . . .	83
Оценка True или False . . . . .	84
Операторы if . . . . .	87
Циклы while . . . . .	91
Циклы for . . . . .	92
Операторы break и continue . . . . .	93
Резюме . . . . .	93
Вопросы для закрепления . . . . .	94
<b>Глава 6. Функции . . . . .</b>	<b>95</b>
Объявление функций . . . . .	96
Управляющий оператор . . . . .	96
Строки документации . . . . .	96
Параметры . . . . .	98
Операторы возврата . . . . .	104
Область видимости в функциях . . . . .	105
Декораторы . . . . .	105
Анонимные функции . . . . .	110
Резюме . . . . .	110
Вопросы для закрепления . . . . .	111



**ЧАСТЬ II**  
**БИБЛИОТЕКИ DATA SCIENCE**

<b>Глава 7. Библиотека NumPy</b> .....	114
Установка и импорт NumPy .....	115
Создание массивов .....	116
Индексация и слайсинг .....	119
Поэлементные операции .....	121
Фильтрация значений .....	122
Представления и копии .....	124
Методы массива .....	125
Бродкастинг .....	129
Математические функции NumPy .....	130
Резюме .....	132
Вопросы для закрепления .....	132
<b>Глава 8. Библиотека SciPy</b> .....	134
Обзор SciPy .....	135
Подмодуль <code>scipy.misc</code> .....	135
Подмодуль <code>scipy.special</code> .....	136
Подмодуль <code>scipy.stats</code> .....	137
Дискретные распределения .....	137
Непрерывные распределения .....	140
Резюме .....	144
Вопросы для закрепления .....	144
<b>Глава 9. Библиотека Pandas</b> .....	145
Структура датафреймов .....	146
Создание датафреймов .....	146
Создание датафреймов из словаря .....	146
Создание датафреймов из списка списков .....	148
Создание датафрейма из файла .....	149
Взаимодействие с данными датафреймов .....	150
Головы и хвосты .....	150

---

Описательная статистика .....	152
Доступ к данным .....	155
Синтаксис со скобками .....	155
Оптимизированный доступ по метке .....	158
Оптимизированный доступ по индексу .....	160
Маски и фильтрация .....	161
Булевы операторы библиотеки Pandas .....	163
Управление датафреймами .....	164
Управление данными .....	167
Метод <code>replace</code> .....	169
Интерактивный дисплей .....	171
Резюме .....	172
Вопросы для закрепления .....	172
<b>Глава 10. Библиотеки визуализации данных .....</b>	<b>173</b>
Библиотека <code>matplotlib</code> .....	174
Оформление графиков .....	175
Маркировка данных .....	179
Построение графиков для множества наборов данных .....	180
Объектно-ориентированный стиль .....	182
Библиотека <code>Seaborn</code> .....	184
Темы <code>Seaborn</code> .....	185
Библиотека <code>Plotly</code> .....	188
Библиотека <code>Vokeh</code> .....	189
Другие библиотеки визуализации .....	191
Резюме .....	191
Вопросы для закрепления .....	192
<b>Глава 11. Библиотеки машинного обучения .....</b>	<b>193</b>
Популярные библиотеки машинного обучения .....	194
Принцип работы машинного обучения .....	194
Преобразования .....	195
Разделение тестовых и тренировочных данных .....	196

Обучение и тестирование .....	197
Подробнее о Scikit-learn .....	198
Резюме .....	198
Вопросы для закрепления .....	198
<b>Глава 12.</b> Инструментарий естественного языка (NLTK) .....	199
Образцы текстов NLTK .....	200
Частотное распределение .....	202
<i>Текстовые объекты</i> .....	206
Классификация текста .....	207
Резюме .....	210
Упражнения .....	211

### **ЧАСТЬ III СРЕДНИЙ УРОВЕНЬ PYTHON**

<b>Глава 13.</b> Функциональное программирование .....	214
Знакомство с функциональным программированием .....	215
Область видимости и состояние .....	215
Зависимость от глобального состояния .....	216
Изменение состояния .....	217
Изменение изменяемых данных .....	218
Функции функционального программирования .....	219
Списковые включения .....	222
Базовый синтаксис списковых включений .....	222
Замена <code>map</code> и <code>filter</code> .....	222
Множественные переменные .....	224
Словарные включения .....	224
Генераторы .....	224
Выражения-генераторы .....	225
Функции-генераторы .....	226
Резюме .....	228
Вопросы для закрепления .....	228

---

<b>Глава 14. Объектно-ориентированное программирование</b> . . . . .	229
Связывание состояния и функции . . . . .	230
Классы и экземпляры . . . . .	230
Закрытые методы и переменные . . . . .	233
Переменные класса . . . . .	233
Специальные методы . . . . .	234
Методы представления . . . . .	235
Расширенные методы сравнения . . . . .	236
Методы математических операторов . . . . .	238
Наследование . . . . .	239
Резюме . . . . .	243
Вопросы для закрепления . . . . .	243
<b>Глава 15. Прочие вопросы</b> . . . . .	245
Сортировка . . . . .	246
Списки . . . . .	246
Чтение и запись файлов . . . . .	250
Контекстные менеджеры . . . . .	250
Объекты <code>datetime</code> . . . . .	251
Регулярные выражения . . . . .	254
Наборы символов . . . . .	255
Классы символов . . . . .	255
Группы . . . . .	256
Именованные группы . . . . .	256
Найти все . . . . .	257
Найти итератор . . . . .	257
Замена . . . . .	258
Замена с использованием именованных групп . . . . .	258
Компиляция регулярных выражений . . . . .	258
Резюме . . . . .	259
Вопросы для закрепления . . . . .	260

<b>Приложение А. Ответы к вопросам в конце глав</b> .....	261
Глава 1 .....	261
Глава 2 .....	261
Глава 3 .....	261
Глава 4 .....	262
Глава 5 .....	262
Глава 6 .....	262
Глава 7 .....	263
Глава 8 .....	263
Глава 9 .....	263
Глава 10 .....	264
Глава 11 .....	264
Глава 12 .....	264
Глава 13 .....	265
Глава 14 .....	265
Глава 15 .....	266