

O'REILLY®

Погружение в аналитику данных

От Excel к Python и R



Джордж Маунт

Оглавление

Предисловие	9
Цель обучения.....	9
Предварительные условия	9
Технические требования.....	9
Требования к предварительной подготовке.....	10
Как я пришел к аналитике.....	10
«Excel — плохо, программирование — хорошо».....	11
Преимущества Excel при обучении аналитике	12
Обзор книги.....	13
Упражнения в конце глав.....	13
Эта книга — не список готовых решений.....	14
Без паники!.....	14
Условные обозначения.....	14
Использование примеров кода	15
Контакты	16
Благодарности.....	16
ЧАСТЬ I. ОСНОВЫ АНАЛИТИКИ В EXCEL	17
Глава 1. Основы разведочного анализа данных	19
Что такое разведочный анализ данных?.....	19
Наблюдения	21
Переменные	21
Категориальные переменные.....	22
Количественные переменные	24
Закрепление материала: классификация переменных.....	25
Резюме: типы переменных.....	27
Исследование переменных в Excel	27
Исследование категориальных переменных.....	27
Исследование количественных переменных	30
Заключение.....	41
Упражнения.....	41
Глава 2. Понятие вероятности	42
Вероятность и случайность	42
Вероятность и выборочное пространство	42
Вероятность и эксперименты	43
Безусловная и условная вероятность	43

Распределение вероятностей	43
Дискретное распределение вероятностей	44
Непрерывное распределение вероятностей	47
Заключение.....	55
Упражнения.....	55
Глава 3. Основы инференциальной статистики	56
Базовые понятия статистического вывода	56
Сбор данных для репрезентативной выборки	57
Формулирование гипотез	58
Разработка плана анализа	59
Анализ данных.....	62
Принятие решения	64
Это ваш мир... данные только живут в нем.....	71
Заклучение.....	72
Упражнения.....	73
Глава 4. Корреляция и регрессия.....	74
«Корреляция не подразумевает причинно-следственную связь»	74
Понятие корреляции.....	75
От корреляции к регрессии	80
Линейная регрессия в Excel.....	81
Переосмысление результатов: ложные зависимости	87
Заклучение.....	88
Переход к программированию	89
Упражнения.....	89
Глава 5. Стек анализа данных.....	90
Статистика, аналитика и наука о данных	90
Статистика	90
Аналитика данных.....	90
Бизнес-аналитика	91
Наука о данных.....	91
Машинное обучение	91
Различия без взаимоисключения	92
Значение стека анализа данных.....	92
Электронные таблицы.....	93
VBA.....	94
Современный Excel.....	95
Базы данных.....	96
Платформы бизнес-аналитики (BI).....	97
Языки программирования для анализа данных	98
Заклучение.....	99
Что будет дальше.....	100
Упражнения.....	100
ЧАСТЬ II. ОТ EXCEL К R.....	101
Глава 6. Первые шаги в R для пользователей Excel	103
Загрузка R.....	103
Начало работы с RStudio.....	103

Пакеты в R.....	112
Обновление R, RStudio и пакетов R.....	114
Заключение.....	114
Упражнения.....	116
Глава 7. Структуры данных в R.....	117
Векторы.....	117
Индексирование и подмножества векторов.....	119
От таблиц Excel к кадрам данных R.....	120
Импорт данных в R.....	122
Исследование кадра данных.....	126
Индексирование и подмножества кадров данных.....	128
Запись кадров данных.....	129
Заключение.....	130
Упражнения.....	130
Глава 8. Обработка и визуализация данных в R.....	131
Обработка данных с помощью пакета <i>dplyr</i>	131
Постолбцовые операции.....	132
Построчные операции.....	135
Агрегирование и объединение данных.....	137
<i>dplyr</i> и оператор <i>pipe</i> (%>%).....	141
Преобразование данных с помощью <i>tidyr</i>	142
Визуализация данных с помощью <i>ggplot2</i>	145
Заключение.....	151
Упражнения.....	151
Глава 9. Кульминация: R для анализа данных.....	152
Разведочный анализ данных.....	153
Проверка гипотез.....	157
t-тест для независимых выборок.....	157
Линейная регрессия.....	159
Разделение и проверка данных для обучения и тестирования.....	161
Заключение.....	164
Упражнения.....	164
ЧАСТЬ III. ОТ EXCEL К PYTHON.....	165
Глава 10. Первые шаги в Python для пользователей Excel.....	167
Загрузка Python.....	167
Начало работы с Jupyter.....	168
Модули в Python.....	176
Обновление Python, Anaconda и пакетов Python.....	178
Заключение.....	178
Упражнения.....	178
Глава 11. Структуры данных в Python.....	180
Массивы <i>NumPy</i>	181
Индексирование и подмножества массивов <i>NumPy</i>	182
Кадры данных <i>Pandas</i>	184

Импорт данных в Python	185
Исследование кадра данных	187
Индексирование и подмножества кадров данных	188
Запись кадров данных	189
Заключение	189
Упражнения	189
Глава 12. Обработка и визуализация данных в Python	191
Постолбцовые операции	192
Построчные операции	194
Агрегирование и объединение данных	195
Преобразование данных	197
Визуализация данных	199
Заключение	203
Упражнения	203
Глава 13. Кульминация: Python для анализа данных	204
Разведочный анализ данных	205
Проверка гипотез	207
t-тест для независимых выборок	207
Линейная регрессия	208
Разделение и проверка данных для обучения и тестирования	210
Заключение	212
Упражнения	212
Глава 14. Заключение и дальнейшие шаги	213
Дополнительные элементы стека анализа данных	213
План исследований и бизнес-эксперименты	213
Дополнительные статистические методы	214
Наука о данных и машинное обучение	214
Контроль версий	214
Этика	215
Двигайтесь вперед и выбирайте то, что нравится	215
Напутствие	216
Предметный указатель	217
Об авторе	221
Об изображении на обложке	222