

Оглавление

Предисловие	14
Благодарности	16
О книге	18
Об авторах	21
Об обложке	24

Часть 1. DATA SCIENCE. С чего начать..... 25

1 Что такое Data Science? 26

1.1. Что такое Data Science?	28
1.1.1. Математика/статистика	29
1.1.2. Базы данных и программирование	31
1.1.3. Понимание бизнеса	33
1.2. Различные типы вакансий в Data Science	35
1.2.1. Аналитики	35
1.2.2. Машинное обучение	36
1.2.3. Теория принятия решений	36
1.2.4. Смежные специальности	38
1.3. Выбор пути	40
1.4. Интервью с Робертом Чангом, дата-сайентистом из Airbnb	40

2 Типы компаний в Data Science 43

2.1. КИТк: крупная информационно-технологическая компания	44
2.1.1. Команда: одна из многих в КИТк	44
2.1.2. Технология: продвинутая, но неупорядоченная	45
2.1.3. Плюсы и минусы КИТк	46
2.2. HandbagLOVE: устоявшийся ритейлер	47
2.2.1. Команда: небольшая группа, стремящаяся к росту	48
2.2.2. Технология: устаревшие методы, которые начинают меняться	49
2.2.3. Плюсы и минусы HandbagLOVE	49

2.3. Seg-Metra: стартап на ранней стадии	50
2.3.1. Команда (какая еще команда?)	51
2.3.2. Технология: передовые методы, собранные воедино	51
2.3.3. Плюсы и минусы Seg-Metra	52
2.4. Videory: успешный технологический стартап на поздней стадии	54
2.4.1. Команда: специализированная, но с разнообразием	55
2.4.2. Технология: стараемся не увязнуть в устаревшем коде	55
2.4.3. Плюсы и минусы Videory	56
2.5. Global Aerospace Dynamics: гигантский государственный подрядчик	57
2.5.1. Команда: дата-сайентист в море инженеров	58
2.5.2. Технологии: старые, ржавые и с сильными ограничениями системы безопасности	59
2.5.3. Плюсы и минусы GAD	59
2.6. Делаем выводы	61
2.7. Интервью с Рэнди Ау, специалистом в области количественного UX Research в Google	61
3 Приобретение навыков	64
3.1. Получение образования в Data Science	65
3.1.1. Выбор учебного заведения	66
3.1.2. Поступление	69
3.1.3. Заключение по академическому образованию	70
3.2. Буткемпы	71
3.2.1. Чему можно научиться	72
3.2.2. Цена	74
3.2.3. Выбор программы	74
3.2.4. Заключение по DS-буткемпам	74
3.3. Работа с Data Science в вашей компании	75
3.3.1. Выводы об обучении на работе	77
3.4. Самообучение	78
3.4.1. Выводы о самообучении	79
3.5. Как сделать выбор	79
3.6. Интервью с Джулией Силдж, дата-сайентистом и инженером-программистом RStudio	80

4 *Создание портфолио* 83

- 4.1. Создание проекта 84
 - 4.1.1. *Найдите данные и задайте вопрос* 84
 - 4.1.2. *Выбор направления* 87
 - 4.1.3. *Заполнение GitHub README* 88
- 4.2. Создание блога 89
 - 4.2.1. *Возможные темы* 89
 - 4.2.2. *Выбор платформы* 91
- 4.3. Работа с примерами проектов 92
 - 4.3.1. *Фрилансеры в Data Science* 92
 - 4.3.2. *Обучение нейронной сети на «неприличных» автомобильных номерах* 94
- 4.4. Интервью с Дэвидом Робинсоном, дата-сайентистом 96

Часть 2. Как попасть в DATA SCIENCE 101

5 *Поиск: как определиться с подходящей работой* 102

- 5.1. Поиск работы 103
 - 5.1.1. *Расшифровка описания вакансий* 104
 - 5.1.2. *Поиск тревожных сигналов* 106
 - 5.1.3. *Большие надежды* 107
 - 5.1.4. *Посещение митапов* 108
 - 5.1.5. *Использование социальных сетей* 110
- 5.2. На какие вакансии откликаться 112
- 5.3. Интервью с Джесси Мостипак, developer advocate в Kaggle 113

6 *Отклик на вакансию: резюме и сопроводительное письмо* 116

- 6.1. Резюме: основы 117
 - 6.1.1. *Структура* 119
 - 6.1.2. *Подробнее о разделе опыта: наполнение* 124
- 6.2. Сопроводительное письмо: основные положения 126
 - 6.2.1. *Структура* 127

6.3. Адаптация	129
6.4. Реферальная программа	130
6.5. Интервью с Кристен Керер, инструктором по Data Science и создателем курсов	132
7 <i>Интервью: чего ожидать и что делать</i>	135
7.1. Чего хотят компании?	136
7.1.1. Процесс интервью	137
7.2. Этап 1: первое телефонное интервью	138
7.3. Этап 2: интервью в офисе	140
7.3.1. Техническое интервью	142
7.3.2. Поведенческое интервью	146
7.4. Этап 3: решение кейса	148
7.5. Этап 4: итоговое интервью	151
7.6. Оффер	151
7.7. Интервью с Райаном Уильямсом, старшим специалистом по принятию решений в Starbucks	152
8 <i>Оффер: знайте, на что соглашаться</i>	155
8.1. Процесс	156
8.2. Получение оффера	156
8.3. Переговоры	158
8.3.1. Что можно обсуждать?	159
8.3.2. О какой сумме договариваться	162
8.4. Тактика переговоров	164
8.5. Как выбрать между двумя «хорошими» офферами	165
8.6. Интервью с Брук Уотсон Мадубуонву, старшим дата-сайентистом в ACLU	167

Часть 3. Осваиваемся в DATA SCIENCE 171**9 Первые месяцы на работе 172**

- 9.1. Первый месяц 173
 - 9.1.1. Онбординг в крупной организации: хорошо отлаженный процесс 173
 - 9.1.2. Онбординг новых сотрудников в небольшой компании: «Онбординг? Нет, не слышали» 174
 - 9.1.3. Понимание и установка ожиданий 174
 - 9.1.4. Знайте данные, с которыми работаете 176
- 9.2. Становимся продуктивными 179
 - 9.2.1. Задавайте вопросы 180
 - 9.2.2. Выстраивайте взаимоотношения 182
- 9.3. Если вы первый дата-сайентист 184
- 9.4. Если работа не соответствует обещанию 186
 - 9.4.1. Ужасная работа 186
 - 9.4.2. Токсичная рабочая среда 187
 - 9.4.3. Решение уволиться 188
- 9.5. Интервью с Джарвисом Миллером, дата-сайентистом в Spotify 190

10 Создание эффективного анализа 193

- 10.1. Запрос 196
- 10.2. План анализа 199
- 10.3. Выполнение анализа 201
 - 10.3.1. Импорт и очистка данных 201
 - 10.3.2. Просмотр и моделирование данных 203
 - 10.3.3. Важные моменты для анализа и моделирования 205
- 10.4. Завершение 209
 - 10.4.1. Итоговая презентация 210
 - 10.4.2. Длительное хранение работы 211
- 10.5. Интервью с Хилари Паркер, дата-сайентистом в Stitch Fix 212

11 *Развертывание модели в производство* 215

- 11.1. А что вообще развертывается в производство? 216
- 11.2. Создание производственной системы 218
 - 11.2.1. Сбор данных 219
 - 11.2.2. Построение модели 220
 - 11.2.3. Обслуживание моделей с API 221
 - 11.2.4. Построение API 222
 - 11.2.5. Документация 224
 - 11.2.6. Тестирование 225
 - 11.2.7. Развертывание API 226
 - 11.2.8. Нагрузочное тестирование 229
- 11.3. Поддержание работоспособности системы 230
 - 11.3.1. Мониторинг системы 230
 - 11.3.2. Переобучение модели 231
 - 11.3.3. Внесение изменений 232
- 11.4. В завершение 232
- 11.5. Интервью с Хизер Нолис, инженером МО в T-Mobile 232

12 *Работа со стейкхолдерами* 235

- 12.1. Типы стейкхолдеров 236
 - 12.1.1. Бизнес-стейкхолдеры 236
 - 12.1.2. Инженеры-стейкхолдеры 238
 - 12.1.3. Высшее руководство компании 239
 - 12.1.4. Ваш непосредственный руководитель 240
- 12.2. Работа со стейкхолдерами 241
 - 12.2.1. Понимание целей стейкхолдеров 241
 - 12.2.2. Постоянное общение 243
 - 12.2.3. Будьте системным 245
- 12.3. Расстановка приоритетов 247
 - 12.3.1. Инновационная и полезная работа 249
 - 12.3.2. Не инновационная, но все же полезная работа 249
 - 12.3.3. Инновационная, но не полезная работа 250
 - 12.3.4. Не инновационная, не полезная работа 251
- 12.4. В завершение 252
- 12.5. Интервью с Сейд Сноуден-Акинтунде, дата-сайентистом в Etsy 252

Часть 4. Как подняться по карьерной лестнице в DATA SCIENCE 257

13 Если DS-проект провалился 258

- 13.1. Почему проваливаются DS-проекты 260
 - 13.1.1. У вас не те данные, что вы хотели 260
 - 13.1.2. У данных нет сигнала 261
 - 13.1.3. Прделанная работа оказалась не нужна 263
- 13.2. Управление риском 265
- 13.3. Что делать, если проекты терпят неудачу 266
 - 13.3.1. Что делать с проектом 267
 - 13.3.2. Как справиться с негативными эмоциями 269
- 13.4. Интервью с Мишель Кейм, руководителем отдела
Data Science и машинного обучения Pluralsight 270

14 Вступление в сообщество Data Science 273

- 14.1. Расширение портфолио 275
 - 14.1.1. Больше публикаций 275
 - 14.1.2. Больше проектов 276
- 14.2. Посещение конференций 277
 - 14.2.1. Как справиться с социофобией 281
- 14.3. Выступление с докладом 282
 - 14.3.1. Получение возможности 283
 - 14.3.2. Подготовка 286
- 14.4. Вклад в открытый исходный код 287
 - 14.4.1. Участие в работе других людей 287
 - 14.4.2. Создание собственного пакета или библиотеки 289
- 14.5. Распознавание и предотвращение выгорания 290
- 14.6. Интервью с Рене Теате, директором отдела Data Science
в HelioCampus 291

15 Уходим красиво 294

- 15.1. Решение уволиться 295
 - 15.1.1. Оценка прогресса в знаниях 296
 - 15.1.2. Заручитесь поддержкой руководителя 296

- 15.2. В чем разница между первым и последующими поисками работы 298
 - 15.2.1. *Определитесь, чего хотите* 299
 - 15.2.2. *Интервью* 300
- 15.3. Поиск новой работы для трудоустроенных 301
- 15.4. Сообщение об увольнении 303
 - 15.4.1. *Рассмотрение контроффера* 304
 - 15.4.2. *Как сказать команде* 304
 - 15.4.3. *Упрощение передачи дел* 306
- 15.5. Интервью с Аmandой Касари, техническим менеджером Google 307

16 *Вверх по карьерной лестнице* 310

- 16.1. Путь руководителя 312
 - 16.1.1. *Преимущества работы руководителем* 313
 - 16.1.2. *Недостатки должности руководителя* 314
 - 16.1.3. *Как стать руководителем* 315
 - 16.2. Путь ведущего дата-сайентиста 317
 - 16.2.1. *Преимущества работы ведущим дата-сайентистом* 318
 - 16.2.2. *Недостатки должности ведущего дата-сайентиста* 319
 - 16.2.3. *Как стать ведущим дата-сайентистом* 320
 - 16.3. Путь независимого консультанта 321
 - 16.3.1. *Преимущества работы в качестве независимого консультанта* 322
 - 16.3.2. *Недостатки работы в качестве независимого консультанта* 323
 - 16.3.3. *Как стать независимым консультантом* 324
 - 16.4. Выбор своего пути 325
 - 16.5. Интервью с Анджелой Басса, руководителем отдела Data Science, инженерии данных и машинного обучения в iRobot 325
- Эпилог 331
- Приложение. Вопросы интервью 333