

Build a Career in Data Science

EMILY ROBINSON
AND JACQUELINE NOLIS



MANNING
SHELTER ISLAND

brief contents

PART 1	GETTING STARTED WITH DATA SCIENCE	1
	1 ■ What is data science?	3
	2 ■ Data science companies	18
	3 ■ Getting the skills	37
	4 ■ Building a portfolio	55
PART 2	FINDING YOUR DATA SCIENCE JOB.....	71
	5 ■ The search: Identifying the right job for you	73
	6 ■ The application: Résumés and cover letters	85
	7 ■ The interview: What to expect and how to handle it	101
	8 ■ The offer: Knowing what to accept	119
PART 3	SETTLING INTO DATA SCIENCE.....	135
	9 ■ The first months on the job	137
	10 ■ Making an effective analysis	155
	11 ■ Deploying a model into production	174
	12 ■ Working with stakeholders	192

PART 4	GROWING IN YOUR DATA SCIENCE ROLE	213
13	■ When your data science project fails	215
14	■ Joining the data science community	228
15	■ Leaving your job gracefully	246
16	■ Moving up the ladder	261

contents

preface xvii
acknowledgments xix
about this book xxi
about the authors xxiv
about the cover illustration xxvii

PART 1 GETTING STARTED WITH DATA SCIENCE.....1

I What is data science? 3

1.1 What is data science? 5
Mathematics/statistics 6 ■ *Databases/programming 7*
Business understanding 9

1.2 Different types of data science jobs 10
Analytics 11 ■ *Machine learning 11* ■ *Decision science 12*
Related jobs 13

1.3 Choosing your path 14

1.4 Interview with Robert Chang, data scientist at Airbnb 15
What was your first data science journey? 15 ■ *What should people look for in a data science job? 16* ■ *What skills do you need to be a data scientist? 16*

- ## 2 *Data science companies* 18
- 2.1 MTC: Massive Tech Company 19
 - Your team: One of many in MTC* 19
 - The tech: Advanced, but siloed across the company* 20
 - The pros and cons of MTC* 21
 - 2.2 HandbagLOVE: The established retailer 22
 - Your team: A small group struggling to grow* 22
 - Your tech: A legacy stack that's starting to change* 23
 - The pros and cons of HandbagLOVE* 23
 - 2.3 Seg-Metra: The early-stage startup 24
 - Your team (what team?)* 25
 - The tech: Cutting-edge technology that's taped together* 26
 - Pros and cons of Seg-Metra* 26
 - 2.4 Videory: The late-stage, successful tech startup 28
 - The team: Specialized but with room to move around* 28
 - The tech: Trying to avoid getting bogged down by legacy code* 29
 - The pros and cons of Videory* 29
 - 2.5 Global Aerospace Dynamics: The giant government contractor 31
 - The team: A data scientist in a sea of engineers* 31
 - The tech: Old, hardened, and on security lockdown* 32
 - The pros and cons of GAD* 32
 - 2.6 Putting it all together 33
 - 2.7 Interview with Randy Au, quantitative user experience researcher at Google 34
 - Are there big differences between large and small companies?* 34
 - Are there differences based on the industry of the company?* 35
 - What's your final piece of advice for beginning data scientists?* 35
- ## 3 *Getting the skills* 37
- 3.1 Earning a data science degree 38
 - Choosing the school* 39
 - Getting into an academic program* 42
 - Summarizing academic degrees* 43
 - 3.2 Going through a bootcamp 44
 - What you learn* 44
 - Cost* 46
 - Choosing a program* 46
 - Summarizing data science bootcamps* 47
 - 3.3 Getting data science work within your company 47
 - Summarizing learning on the job* 49
 - 3.4 Teaching yourself 49
 - Summarizing self-teaching* 50

- 3.5 Making the choice 51
- 3.6 Interview with Julia Silge, data scientist and software engineer at RStudio 52

Before becoming a data scientist, you worked in academia; how have the skills learned there helped you as a data scientist? 52 ■ *When deciding to become a data scientist, what did you use to pick up new skills? 53* ■ *Did you know going into data science what kind of work you wanted to be doing? 53* ■ *What would you recommend to people looking to get the skills to be a data scientist? 53*

4 Building a portfolio 55

- 4.1 Creating a project 56

Finding the data and asking a question 56 ■ *Choosing a direction 59* ■ *Filling out a GitHub README 60*

- 4.2 Starting a blog 60

Potential topics 60 ■ *Logistics 61*

- 4.3 Working on example projects 63

Data science freelancers 63 ■ *Training a neural network on offensive license plates 64*

- 4.4 Interview with David Robinson, data scientist 65

How did you start blogging? 66 ■ *Are there any specific opportunities you have gotten from public work? 66* ■ *Are there people you think would especially benefit from doing public work? 66* ■ *How has your view on the value of public work changed over time? 66* ■ *How do you come up with ideas for your data analysis posts? 67* ■ *What's your final piece of advice for aspiring and junior data scientists? 67*

PART 2 FINDING YOUR DATA SCIENCE JOB.....71

5 The search: Identifying the right job for you 73

- 5.1 Finding jobs 74

Decoding descriptions 75 ■ *Watching for red flags 77*
Setting your expectations 77 ■ *Attending meetups 78*
Using social media 80

- 5.2 Deciding which jobs to apply for 81

- 5.3 Interview with Jesse Mostipak, developer advocate at Kaggle 83

What recommendations do you have for starting a job search? 83
How can you build your network? 83 ■ *What do you do if you*

don't feel confident applying to data science jobs? 83 ■ *What would you say to someone who thinks "I don't meet the full list of any job's required qualifications?"* 84 ■ *What's your final piece of advice to aspiring data scientists?* 84

6 The application: Résumés and cover letters 85

6.1 Résumé: The basics 86

Structure 88 ■ *Deeper into the experience section: generating content* 93

6.2 Cover letters: The basics 94

Structure 95

6.3 Tailoring 96

6.4 Referrals 97

6.5 Interview with Kristen Kehrer, data science instructor and course creator 99

How many times would you estimate you've edited your résumé? 99
What are common mistakes you see people make? 99 ■ *Do you tailor your résumé to the position you're applying to?* 100 ■ *What strategies do you recommend for describing jobs on a résumé?* 100
What's your final piece of advice for aspiring data scientists? 100

7 The interview: What to expect and how to handle it 101

7.1 What do companies want? 102

The interview process 103

7.2 Step 1: The initial phone screen interview 104

7.3 Step 2: The on-site interview 106

The technical interview 108 ■ *The behavioral interview* 111

7.4 Step 3: The case study 113

7.5 Step 4: The final interview 115

7.6 The offer 116

7.7 Interview with Ryan Williams, senior decision scientist at Starbucks 117

What are the things you need to do knock an interview out of the park? 117 ■ *How do you handle the times where you don't know the answer?* 117 ■ *What should you do if you get a negative response to your answer?* 118 ■ *What has running interviews taught you about being an interviewee?* 118

8 *The offer: Knowing what to accept* 119

- 8.1 The process 120
- 8.2 Receiving the offer 120
- 8.3 Negotiation 122
 - What is negotiable?* 122 ▪ *How much you can negotiate* 125
- 8.4 Negotiation tactics 127
- 8.5 How to choose between two “good” job offers 128
- 8.6 Interview with Brooke Watson Madubuonwu, senior data scientist at the ACLU 129
 - What should you consider besides salary when you’re considering an offer?* 130 ▪ *What are some ways you prepare to negotiate?* 130
 - What do you do if you have one offer but are still waiting on another one?* 130 ▪ *What’s your final piece of advice for aspiring and junior data scientists?* 131

PART 3 SETTLING INTO DATA SCIENCE 135

9 *The first months on the job* 137

- 9.1 The first month 138
 - Onboarding at a large organization: A well-oiled machine* 138
 - Onboarding at a small company: What onboarding?* 139
 - Understanding and setting expectations* 139 ▪ *Knowing your data* 141
- 9.2 Becoming productive 144
 - Asking questions* 145 ▪ *Building relationships* 146
- 9.3 If you’re the first data scientist 148
- 9.4 When the job isn’t what was promised 149
 - The work is terrible* 149 ▪ *The work environment is toxic* 150
 - Deciding to leave* 151
- 9.5 Interview with Jarvis Miller, data scientist at Spotify 152
 - What were some things that surprised you in your first data science job?* 153 ▪ *What are some issues you faced?* 153 ▪ *Can you tell us about one of your first projects?* 153 ▪ *What would be your biggest piece of advice for the first few months?* 154

10 *Making an effective analysis* 155

- 10.1 The request 158
- 10.2 The analysis plan 160

- 10.3 Doing the analysis 162
 - Importing and cleaning data* 162 ▪ *Data exploration and modeling* 164 ▪ *Important points for exploring and modeling* 166
- 10.4 Wrapping it up 169
 - Final presentation* 170 ▪ *Mothballing your work* 171
- 10.5 Interview with Hilary Parker, data scientist at Stitch Fix 172
 - How does thinking about other people help your analysis?* 172
 - How do you structure your analyses?* 172 ▪ *What kind of polish do you do in the final version?* 172 ▪ *How do you handle people asking for adjustments to an analysis?* 173

11 Deploying a model into production 174

- 11.1 What is deploying to production, anyway? 175
- 11.2 Making the production system 177
 - Collecting data* 178 ▪ *Building the model* 178
 - Serving models with APIs* 179 ▪ *Building an API* 180
 - Documentation* 182 ▪ *Testing* 183 ▪ *Deploying an API* 184
 - Load testing* 187
- 11.3 Keeping the system running 187
 - Monitoring the system* 187 ▪ *Retraining the model* 188
 - Making changes* 189
- 11.4 Wrapping up 189
- 11.5 Interview with Heather Nolis, machine learning engineer at T-Mobile 189
 - What does “machine learning engineer” mean on your team?* 189
 - What was it like to deploy your first piece of code?* 190 ▪ *If you have things go wrong in production, what happens?* 190 ▪ *What’s your final piece of advice for data scientists working with engineers?* 191

12 Working with stakeholders 192

- 12.1 Types of stakeholders 193
 - Business stakeholders* 193 ▪ *Engineering stakeholders* 194
 - Corporate leadership* 195 ▪ *Your manager* 196
- 12.2 Working with stakeholders 197
 - Understanding the stakeholder’s goals* 197 ▪ *Communicating constantly* 199 ▪ *Being consistent* 201

- 12.3 Prioritizing work 203
 - Both innovative and impactful work* 204 ▪ *Not innovative but still impactful work* 205 ▪ *Innovative but not impactful work* 205 ▪ *Neither innovative nor impactful work* 206
- 12.4 Concluding remarks 206
- 12.5 Interview with Sade Snowden-Akintunde, data scientist at Etsy 207
 - Why is managing stakeholders important?* 207 ▪ *How did you learn to manage stakeholders?* 207 ▪ *Was there a time where you had difficulty with a stakeholder?* 207 ▪ *What do junior data scientists frequently get wrong?* 208 ▪ *Do you always try to explain the technical part of the data science?* 208 ▪ *What's your final piece of advice for junior or aspiring data scientists?* 208

PART 4 GROWING IN YOUR DATA SCIENCE ROLE.....213

13 *When your data science project fails* 215

- 13.1 Why data science projects fail 216
 - The data isn't what you wanted* 217 ▪ *The data doesn't have a signal* 218 ▪ *The customer didn't end up wanting it* 220
- 13.2 Managing risk 221
- 13.3 What you can do when your projects fail 222
 - What to do with the project* 223 ▪ *Handling negative emotions* 224
- 13.4 Interview with Michelle Keim, head of data science and machine learning at Pluralsight 226
 - When was a time you experienced a failure in your career?* 226
 - Are there red flags you can see before a project starts?* 226
 - How does the way a failure is handled differ between companies?* 226 ▪ *How can you tell if a project you're on is failing?* 227 ▪ *How can you get over a fear of failing?* 227

14 *Joining the data science community* 228

- 14.1 Growing your portfolio 230
 - More blog posts* 230 ▪ *More projects* 231
- 14.2 Attending conferences 231
 - Dealing with social anxiety* 234
- 14.3 Giving talks 235
 - Getting an opportunity* 236 ▪ *Preparing* 239

- 14.4 Contributing to open source 239
 - Contributing to other people's work* 240 ▪ *Making your own package or library* 241
- 14.5 Recognizing and avoiding burnout 242
- 14.6 Interview with Renee Teate, director of data science at HelioCampus 243
 - What are the main benefits of being on social media?* 243 ▪ *What would you say to people who say they don't have the time to engage with the community?* 244 ▪ *Is there value in producing only a small amount of content?* 244 ▪ *Were you worried the first time you published a blog post or gave a talk?* 244

15 *Leaving your job gracefully* 246

- 15.1 Deciding to leave 247
 - Take stock of your learning progress* 247 ▪ *Check your alignment with your manager* 248
- 15.2 How the job search differs after your first job 250
 - Deciding what you want* 250 ▪ *Interviewing* 251
- 15.3 Finding a new job while employed 252
- 15.4 Giving notice 254
 - Considering a counteroffer* 255 ▪ *Telling your team* 255
 - Making the transition easier* 257
- 15.5 Interview with Amanda Casari, engineering manager at Google 258
 - How do you know it's time to start looking for a new job?* 258
 - Have you ever started a job search and decided to stay instead?* 258
 - Do you see people staying in the same job for too long?* 258
 - Can you change jobs too quickly?* 259 ▪ *What's your final piece of advice for aspiring and new data scientists?* 259

16 *Moving up the ladder* 261

- 16.1 The management track 263
 - Benefits of being a manager* 264 ▪ *Drawbacks of being a manager* 264 ▪ *How to become a manager* 265
- 16.2 Principal data scientist track 267
 - Benefits of being a principal data scientist* 268 ▪ *Drawbacks of being a principal data scientist* 269 ▪ *How to become a principal data scientist* 270

- 16.3 Switching to independent consulting 271
Benefits of independent consulting 272 ▪ *Drawbacks of independent consulting* 272 ▪ *How to become an independent consultant* 273
- 16.4 Choosing your path 274
- 16.5 Interview with Angela Bassa, head of data science, data engineering, and machine learning at iRobot 275
What's the day-to-day life as a manager like? 275 ▪ *What are the signs you should move on from being an independent contributor?* 275 ▪ *Do you have to eventually transition out of being an independent contributor?* 275 ▪ *What advice do you have for someone who wants to be a technical lead but isn't quite ready for it?* 276 ▪ *What's your final piece of advice to aspiring and junior data scientist?* 276
- Epilogue* 280
- appendix Interview questions* 282
- index* 311