

Jacqueline Nolis, Emily Robinson

SKAZANY NA SUKCES

Kariera w **Data Science**



Tytuł oryginału: Build A Career in Data Science

Tłumaczenie: Anna Zawila, Tadeusz Zawila

ISBN: 978-83-283-7293-1

Original edition copyright © 2020 by Emily Robinson and Jacqueline Nolis.

All rights reserved.

Polish edition copyright © 2021 by Helion SA.

All rights reserved.

No part of this book may be reproduced or transmitted in any form or by any means, electronic or mechanical, including photocopying, recording or by any information storage retrieval system, without permission from the Publisher.

Wszelkie prawa zastrzeżone. Nieautoryzowane rozpowszechnianie całości lub fragmentu niniejszej publikacji w jakiegokolwiek postaci jest zabronione. Wykonywanie kopii metodą kserograficzną, fotograficzną, a także kopiowanie książki na nośniku filmowym, magnetycznym lub innym powoduje naruszenie praw autorskich niniejszej publikacji.

Wszystkie znaki występujące w tekście są zastrzeżonymi znakami firmowymi bądź towarowymi ich właścicieli.

Autor oraz Helion SA dołożyli wszelkich starań, by zawarte w tej książce informacje były kompletne i rzetelne. Nie biorą jednak żadnej odpowiedzialności ani za ich wykorzystanie, ani za związane z tym ewentualne naruszenie praw patentowych lub autorskich. Autor oraz Helion SA nie ponoszą również żadnej odpowiedzialności za ewentualne szkody wynikłe z wykorzystania informacji zawartych w książce.

Helion SA

ul. Kościuszki 1c, 44-100 Gliwice

tel. 32 231 22 19, 32 230 98 63

e-mail: helion@helion.pl

WWW: <http://helion.pl> (księgarnia internetowa, katalog książek)

Drogi Czytelniku!

Jeżeli chcesz ocenić tę książkę, zajrzyj pod adres

<http://helion.pl/user/opinie/sksukd>

Możesz tam wpisać swoje uwagi, spostrzeżenia, recenzję.

Printed in Poland.

- [Kup książkę](#)
- [Poleć książkę](#)
- [Oceń książkę](#)

- [Księgarnia internetowa](#)
- [Lubię to! » Nasza społeczność](#)

Spis treści

<i>Wstęp</i>	13
<i>Podziękowania</i>	15
<i>O niniejszej książce</i>	17
<i>O autorkach</i>	21

CZĘŚĆ I ROZPOCZĘCIE PRZYGODY Z BADANIEM DANYCH 23

1. Czym jest danologia?	25
1.1. Czym jest danologia?	27
1.1.1. <i>Matematyka/statystyka</i>	29
1.1.2. <i>Bazy danych/programowanie</i>	30
1.1.3. <i>Zrozumienie biznesu</i>	32
1.2. Różne rodzaje prac związanych z badaniem danych	33
1.2.1. <i>Analiza</i>	34
1.2.2. <i>Uczenie maszynowe</i>	34
1.2.3. <i>Nauka o podejmowaniu decyzji</i>	35
1.2.4. <i>Pokrewne prace</i>	36
1.3. Wybór swojej drogi	37
1.4. Wywiad z Robertem Changiem, badaczem danych w Airbnb	38
<i>Jak zaczęła się twoja podróż badacza danych?</i>	39
<i>Czego ludzie powinni szukać w pracy związanej z danologią?</i>	39
<i>Jakich umiejętności potrzeba, aby zostać badaczem danych?</i>	40
2. Firmy działające w obszarze danologii	41
2.1. MTC — potężna firma technologiczna	42
2.1.1. <i>Twój zespół — jeden z wielu w MTC</i>	42
2.1.2. <i>Technologia — zaawansowana, ale obecna w firmie w systemie silosowym</i>	43
2.1.3. <i>Zalety i wady pracy w MTC</i>	44
2.2. HandbagLOVE — uznany sprzedawca detaliczny	45
2.2.1. <i>Twój zespół — grupka ludzi walczących o rozwój</i>	46
2.2.2. <i>Twoja technologia — stos technologiczny, który zaczyna się zmieniać</i>	46
2.2.3. <i>Wady i zalety pracy w HandbagLOVE</i>	47
2.3. Seg-Metra — nowo powstały start-up	48
2.3.1. <i>Twój zespół (jaki zespół?)</i>	48
2.3.2. <i>Technologia — najnowsza, wewnętrznie spójna technologia</i>	49
2.3.3. <i>Wady i zalety pracy w Seg-Metra</i>	50
2.4. Videory — udany start-up technologiczny na późnym etapie rozwoju	52
2.4.1. <i>Zespół — wyspecjalizowany, ale z przestrzenią do działania</i>	52
2.4.2. <i>Technologia — próby pokonania kodu odziedziczonych</i>	53
2.4.3. <i>Zalety i wady pracy w Videory</i>	53

- 2.5. Global Aerospace Dynamics — wielki dostawca rządowy 54
 - 2.5.1. *Zespół — danolog w morzu inżynierów* 55
 - 2.5.2. *Technologia — stara, skostniała i wyposażona w blokadę bezpieczeństwa* 56
 - 2.5.3. *Zalety i wady pracy w CAD* 56
- 2.6. Podsumowanie 57
- 2.7. Wywiad z Randym Au, specjalistą Google’a ds. badań ilościowych w zakresie doświadczenia użytkownika (ang. user experience) 58
 - Czy istnieją duże różnice pomiędzy dużymi a małymi firmami?* 58
 - Czy istnieją różnice w zależności od branży, w jakiej działa dana firma?* 59
 - Jaka jest twoja ostatnia rada dla początkujących badaczy danych?* 59
- 3. Zdobywanie umiejętności 61
 - 3.1. Uzyskiwanie tytułu zawodowego lub stopnia naukowego w zakresie danologii 62
 - 3.1.1. *Wybór uczelni* 64
 - 3.1.2. *Dostanie się do programu akademickiego* 66
 - 3.1.3. *Podsumowanie kwestii tytułów zawodowych oraz stopni naukowych* 68
 - 3.2. Przejście przez intensywny kurs 69
 - 3.2.1. *Czego się nauczysz* 69
 - 3.2.2. *Koszt* 71
 - 3.2.3. *Wybór programu* 71
 - 3.2.4. *Podsumowanie intensywnych kursów danologicznych* 72
 - 3.3. Zdobywanie pracy z zakresu badania danych w swojej firmie 72
 - 3.3.1. *Podsumowanie nauki w pracy* 75
 - 3.4. Uczenie się na własną rękę 75
 - 3.4.1. *Podsumowanie nauki na własną rękę* 76
 - 3.5. Dokonanie wyboru 77
 - 3.6. Wywiad z Julią Silge, badaczką danych i programistką w RStudio 78
 - Zanim zostałaś badaczką danych, pracowałaś na uczelni. Jak zdobyte w środowisku akademickim umiejętności pomogły ci w danologii?* 78
 - Po podjęciu decyzji o zostaniu badaczką danych w jaki sposób zaczęłaś nabywać nowych umiejętności?* 79
 - Czy wiedziałaś, wkraczając do domeny badania danych, jaką pracę chciałabyś wykonywać?* 79
 - Co poleciałabyś osobom, które chciałyby zdobyć umiejętności potrzebne do zostania badaczem danych?* 80
- 4. Tworzenie portfolio 81
 - 4.1. Tworzenie projektu 82
 - 4.1.1. *Identyfikowanie danych i zadanie pytania* 82
 - 4.1.2. *Wybór kierunku* 85
 - 4.1.3. *Wypełnianie pliku README na GitHubie* 86
 - 4.2. Założenie bloga 87
 - 4.2.1. *Potencjalne tematy* 87
 - 4.2.2. *Logistyka* 88
 - 4.3. Praca nad przykładowymi projektami 90
 - 4.3.1. *Freelancerzy zajmujący się badaniem danych* 90
 - 4.3.2. *Uczenie sieci neuronowej na podstawie kontrowersyjnych tablic rejestracyjnych* 91

- 4.4. Wywiad z Davidem Robinsonem, danologiem 93
Jak zaczęła się twoja przygoda z blogowaniem? 93
Czy działalność publiczna przyniosła ci konkretne korzyści? 93
Czy są ludzie, którzy twoim zdaniem szczególnie skorzystałoby na działalności publicznej? 93
Jak zmieniał się z czasem twój pogląd na wartość działalności publicznej? 94
Jak powstają pomysły na posty dotyczące analizy danych? 94
Jaka jest twoja ostatnia rada dla początkujących i młodszych badaczy danych? 94
- Zasoby do rozdziałów 1. – 4. 95
Książki 95
Posty na blogu 96

CZĘŚĆ II ZNALEZIENIE PRACY ZWIĄZANEJ Z ANALIZĄ DANYCH 99

5. Poszukiwania — znalezienie odpowiedniej dla siebie pracy 101
- 5.1. Znalezienie pracy 102
- 5.1.1. *Dekodowanie opisów* 103
- 5.1.2. *Szukanie znaków ostrzegawczych* 105
- 5.1.3. *Określanie swoich oczekiwań* 106
- 5.1.4. *Uczestnictwo w spotkaniach* 107
- 5.1.5. *Korzystanie z mediów społecznościowych* 109
- 5.2. *Podjęmowanie decyzji o tym, o jaką pracę się ubiegać* 110
- 5.3. Wywiad z Jesse Mostipak, rzeczniką ds. rozwoju oprogramowania w Kaggle'u 112
Jakie masz porady co do rozpoczęcia poszukiwania pracy? 112
Jak można zbudować swoją sieć kontaktów? 112
Co robić, gdy brak nam pewności, żeby odpowiadać na oferty pracy z zakresu badania danych? 113
Co byś powiedziała komuś, kto uważa, że „nie spełnia wszystkich wymogów na to stanowisko”? 113
Jaka będzie twoja ostatnia rada dla początkujących badaczy danych? 113
6. Aplikowanie — życiorysy i listy motywacyjne 115
- 6.1. Życiorys — podstawy 116
- 6.1.1. *Struktura* 118
- 6.1.2. *Zanurzenie się w szczegółach dotyczących doświadczenia zawodowego — generowanie treści* 123
- 6.2. Listy motywacyjne — podstawy 125
- 6.2.1. *Struktura* 126
- 6.3. *Szycie na miarę* 127
- 6.4. *Polecenie* 128
- 6.5. Wywiad z Kristen Kehrer, wykładowczynią i autorką kursów danologicznych 130
Ile razy, według twoich szacunków, przeredagowywałaś swój życiorys? 130
Jakie są częste błędy, które twoim zdaniem ludzie popełniają? 131
Czy dostosowujesz swoje CV do stanowiska, na które aplikujesz? 131
Jakie strategie polecasz do opisanie stanowisk pracy w CV? 131
Jaką masz ostatnią radę dla początkujących danologów? 132

7. Rozmowa kwalifikacyjna — czego należy się spodziewać i jak sobie z tym poradzić 133
- 7.1. Czego pragną firmy? 134
 - 7.1.1. *Proces rozmowy rekrutacyjnej* 135
 - 7.2. Krok 1. Wstępna rozmowa telefoniczna 136
 - 7.3. Krok 2. Rozmowa w siedzibie firmy 138
 - 7.3.1. *Rozmowa na tematy techniczne* 140
 - 7.3.2. *Pytania behawioralne* 144
 - 7.4. Krok 3. Analiza przypadku 146
 - 7.5. Krok 4. Rozmowa końcowa 148
 - 7.6. Oferta 149
 - 7.7. Wywiad z Ryanem Williamsem, doświadczonym analitykiem decyzyjnym w Starbucksie 150
 - Co trzeba zrobić, żeby świetnie wypaść na rozmowie o pracę?* 150
 - Jak sobie radzisz w sytuacjach, gdy nie znasz odpowiedzi?* 150
 - Co należy zrobić w przypadku, gdy odpowiedź spotka się z negatywną reakcją?* 151
 - Czego na temat kandydatów nauczyło cię prowadzenie rozmów o pracę?* 151
8. Oferta — co można zaakceptować 153
- 8.1. Proces 154
 - 8.2. Otrzymanie oferty 154
 - 8.3. Negocjacje 156
 - 8.3.1. *Co podlega negocjacji?* 157
 - 8.3.2. *Jak bardzo możesz negocjować* 159
 - 8.4. Taktyki negocjacyjne 162
 - 8.5. Jak wybrać jedną spośród dwóch „dobrych” ofert pracy 163
 - 8.6. Wywiad z Brooke Watson Madubuonwu, starszą badaczką danych w ACLU 165
 - Co należy wziąć pod uwagę, oprócz wynagrodzenia, gdy rozważa się przyjęcie oferty?* 165
 - Jakie są sposoby na przygotowanie się do negocjacji?* 165
 - Co zrobić, kiedy mamy już jedną ofertę, ale nadal czekamy na inną?* 166
 - Jaka jest twoja ostatnia rada dla początkujących i młodszych badaczy danych?* 166
- Zasoby do rozdziałów 5. – 8. 167
- Książki* 167
 - Posty na blogach i kursy* 168

CZĘŚĆ III ZADOMOWIENIE SIĘ W ŚWIECIE BADANIA DANYCH 171

9. Pierwsze miesiące w nowym miejscu pracy 173
- 9.1. Pierwszy miesiąc 174
 - 9.1.1. *Wdrożenie w dużej organizacji — dobrze naoliwiona maszyna* 174
 - 9.1.2. *Wdrożenie w małej firmie. Jakie wdrożenie?* 175
 - 9.1.3. *Zrozumienie i ustalenie oczekiwań* 175
 - 9.1.4. *Znajomość własnych danych* 177
 - 9.2. Stawianie się produktywnym pracownikiem 180
 - 9.2.1. *Zadawanie pytań* 181
 - 9.2.2. *Budowanie relacji* 182

- 9.3. Jeśli jesteś pierwszym zatrudnionym badaczem danych 184
- 9.4. Kiedy praca nie jest tym, co Ci obiecywano 185
 - 9.4.1. *Charakter pracy jest okropny* 185
 - 9.4.2. *Środowisko pracy jest toksyczne* 186
 - 9.4.3. *Decyzja o odejściu z pracy* 187
- 9.5. Wywiad z Jarvisem Millerem, badaczem danych w firmie Spotify 189
 - Co cię zaskoczyło w twojej pierwszej pracy naukowej w dziedzinie danologii?* 189
 - Z jakimi problemami się stykałeś?* 190
 - Możesz nam opowiedzieć o jednym ze swoich pierwszych projektów?* 190
 - Jaka byłaby twoja najważniejsza rada na kilka pierwszych miesięcy pracy?* 191
- 10. Przeprowadzanie skutecznej analizy 193
 - 10.1. Wniosek 196
 - 10.2. Plan analizy 198
 - 10.3. Przeprowadzenie analizy 201
 - 10.3.1. *Importowanie i czyszczenie danych* 201
 - 10.3.2. *Eksplorowanie i modelowanie danych* 203
 - 10.3.3. *Ważne punkty dotyczące eksplorowania i modelowania* 205
 - 10.4. Odpowiednia otoczka 209
 - 10.4.1. *Finalna prezentacja* 210
 - 10.4.2. *Kończenie pracy* 211
 - 10.5. Wywiad z Hilary Parker, badaczką danych w Stitch Fix 211
 - W jaki sposób myślenie o innych ludziach pomaga w twoich analizach?* 212
 - Jaką strukturę nadajesz swoim analizom?* 212
 - Jakiego rodzaju korektę przeprowadzasz w ramach wersji ostatecznej?* 212
 - Jak sobie radzisz z osobami proszącymi o wprowadzenie zmian do analizy?* 213
- 11. Wdrażanie modelu do środowiska produkcyjnego 215
 - 11.1. Czym w ogóle jest wdrożenie do środowiska produkcyjnego? 216
 - 11.2. Tworzenie systemu produkcyjnego 218
 - 11.2.1. *Gromadzenie danych* 219
 - 11.2.2. *Budowa modelu* 220
 - 11.2.3. *Serwowanie modeli przy użyciu interfejsu programowania aplikacji* 221
 - 11.2.4. *Budowa interfejsu programowania aplikacji* 222
 - 11.2.5. *Dokumentacja* 224
 - 11.2.6. *Testowanie* 225
 - 11.2.7. *Wdrażanie API* 225
 - 11.2.8. *Testy obciążeniowe* 229
 - 11.3. Utrzymanie działającego systemu 229
 - 11.3.1. *Monitorowanie systemu* 230
 - 11.3.2. *Ponowne trenowanie modelu* 230
 - 11.3.3. *Wprowadzanie zmian* 231
 - 11.4. Na zakończenie 232
 - 11.5. Rozmowa z Heather Nolis, inżynierem zajmującym się uczeniem maszynowym w T-Mobile 232
 - Co oznacza bycie „inżynierem zajmującym się uczeniem maszynowym” w twoim zespole?* 232
 - Jakie to było uczucie, gdy po raz pierwszy wdrażałaś fragment kodu do środowiska produkcyjnego?* 233

Jeśli coś pójdzie nie tak w środowisku produkcyjnym, to co wtedy? 233
Jaka będzie twoja ostatnia rada dla danologów pracujących z inżynierami? 234

12. Współpraca z interesariuszami 235

- 12.1. Typy interesariuszy 236
 - 12.1.1. *Interesariusze biznesowi* 236
 - 12.1.2. *Interesariusze inżynierijni* 237
 - 12.1.3. *Korporacyjne kierownictwo* 239
 - 12.1.4. *Twój menedżer* 240
 - 12.2. Współpraca z interesariuszami 240
 - 12.2.1. *Zrozumienie celów interesariusza* 241
 - 12.2.2. *Ciągła komunikacja* 243
 - 12.2.3. *Bycie konsekwentnym* 245
 - 12.3. Ustalanie priorytetów w pracy 247
 - 12.3.1. *Praca zarówno innowacyjna, jak i wywierająca wpływ* 248
 - 12.3.2. *Praca, która nie jest innowacyjna, ale wciąż wywiera wpływ* 249
 - 12.3.3. *Praca innowacyjna, która jednak nie wywiera żadnego wpływu* 249
 - 12.3.4. *Praca, która ani nie jest innowacyjna, ani też nie wywiera żadnego wpływu* 250
 - 12.4. Uwagi końcowe 251
 - 12.5. Wywiad z Sade Snowden-Akintunde, badaczką danych w Etsy 251
 - Dlaczego zarządzanie interesariuszami jest ważne?* 252
 - Jak nauczyłaś się zarządzać interesariuszami?* 252
 - Czy miałaś kiedyś problemy z interesariuszem?* 252
 - Co młodzi badacze danych często robią źle?* 252
 - Czy zawsze starasz się wyjaśniać techniczne aspekty danologii?* 253
 - Jaka jest twoja ostatnia porada dla młodszych lub początkujących danologów?* 253
- Zasoby do rozdziałów 9. – 12. 253
 Książki 253
 Blogi 255

CZĘŚĆ IV TWÓJ ROZWÓJ W ROLI BADACZA DANYCH 257

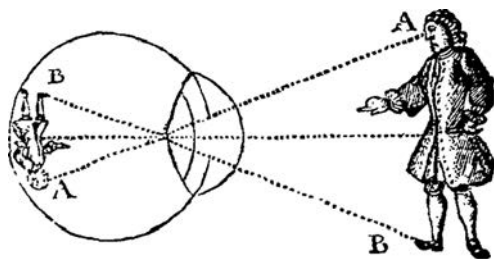
13. Kiedy Twój projekt danologiczny kończy się niepowodzeniem 259

- 13.1. Dlaczego projekty danologiczne kończą się niepowodzeniem 261
 - 13.1.1. *Dane nie spełniają Twoich oczekiwań* 261
 - 13.1.2. *Brak sygnału w zakresie danych* 262
 - 13.1.3. *Utrata zainteresowania klientów* 264
- 13.2. Zarządzanie ryzykiem 266
- 13.3. Co możesz zrobić, gdy Twój projekt kończy się niepowodzeniem? 267
 - 13.3.1. *Co należy zrobić z projektem?* 267
 - 13.3.2. *Radzenie sobie z negatywnymi emocjami* 269
- 13.4. Wywiad z Michelle Keim, dyrektorką ds. danologii i uczenia maszynowego w Pluralsight 270
 - Czy zdarzyło ci się doświadczyć porażki w swojej karierze?* 271
 - Czy potrafisz rozpoznać sygnały ostrzegawcze przed rozpoczęciem projektu?* 271
 - Czy firmy różnią się w sposobie podejścia do niepowodzeń?* 271
 - Skąd wiesz, że projekt, który realizujesz, skończy się niepowodzeniem?* 272
 - Jak pokonać strach przed porażką?* 272

14. Dołączenie do środowiska danologicznego 273
- 14.1. Rozwijanie własnego portfolio 275
 - 14.1.1. Więcej wpisów na blogu 275
 - 14.1.2. Więcej projektów 276
 - 14.2. Uczestnictwo w konferencjach 277
 - 14.2.1. Radzenie sobie z lękiem społecznym 280
 - 14.3. Wystąpienia 281
 - 14.3.1. Poszukiwanie możliwości wystąpienia 281
 - 14.3.2. Przygotowanie 284
 - 14.4. Przyczynianie się do rozwoju otwartego oprogramowania 285
 - 14.4.1. Wkład w pracę innych ludzi 285
 - 14.4.2. Tworzenie własnego pakietu lub biblioteki 287
 - 14.5. Rozpoznawanie i unikanie wypalenia 288
 - 14.6. Wywiad z Renee Teate, dyrektor ds. danologii w HelioCampus 289
 - Jakie są główne korzyści płynące z działalności w mediach społecznościowych?* 289
 - Co byś powiedziała ludziom, którzy mówią, że nie mają czasu na angażowanie się w życie społeczności?* 290
 - Czy warto wytwarzać niewielką ilość treści?* 290
 - Stresowałaś się publikacją pierwszego wpisu na blogu lub pierwszym przemówieniem?* 290
15. Jak odejść z pracy z wdziękiem 293
- 15.1. Decyzja o odejściu 294
 - 15.1.1. Podsumuj swoje postępy w nauce 294
 - 15.1.2. Omów swoją sytuację z menedżerem 295
 - 15.2. Jak wygląda poszukiwanie pracy, kiedy masz już za sobą pierwszą pracę 297
 - 15.2.1. Decydowanie o tym, czego chcesz 298
 - 15.2.2. Rozmowa kwalifikacyjna 298
 - 15.3. Znalezienie nowej pracy w trakcie zatrudnienia 300
 - 15.4. Złożenie wypowiedzenia 301
 - 15.4.1. Rozważanie przyjęcia kontroferty 302
 - 15.4.2. Poinformowanie zespołu 303
 - 15.4.3. Ułatwienie przeprowadzenia zmiany 304
 - 15.5. Wywiad z Amandą Casari, menedżerem ds. technicznych w Google'u 305
 - Po czym poznać, że nadszedł czas na szukanie nowej pracy?* 305
 - Czy kiedykolwiek rozpoczęłaś poszukiwania pracy i zdecydowałaś się jednak pozostać?* 306
 - Widzisz ludzi, którzy zbyt długo wykonują tę samą pracę?* 306
 - Czy można zmienić pracę zbyt szybko?* 306
 - Jaka jest twoja ostatnia rada dla ambitnych i początkujących danologów?* 307
16. Wspinanie się po szczeblach kariery 309
- 16.1. Ścieżka menedżerska 311
 - 16.1.1. Zalety bycia menedżerem 312
 - 16.1.2. Wady bycia menedżerem 312
 - 16.1.3. Jak zostać menedżerem 313

16.2.	Ścieżka głównego danologa	315
16.2.1.	Zalety bycia głównym danologiem	317
16.2.2.	Wady bycia głównym danologiem	318
16.2.3.	Jak zostać głównym danologiem	318
16.3.	Zostanie niezależnym konsultantem	319
16.3.1.	Zalety bycia niezależnym konsultantem	321
16.3.2.	Wady bycia niezależnym konsultantem	321
16.3.3.	Jak zostać niezależnym konsultantem	322
16.4.	Wybór ścieżki	323
16.5.	Wywiad z Angelą Bassą, dyrektorką ds. danologii, inżynierii danych i uczenia maszynowego w iRobot	323
	<i>Jaka wygląda codzienna praca na stanowisku menedżerskim?</i>	324
	<i>Jakie są oznaki tego, że należy przestać być szeregowym pracownikiem?</i>	324
	<i>Czy musisz w końcu całkiem przestać wykonywać robotę szeregowego pracownika?</i>	324
	<i>Jakiej rady udzieliłabyś komuś, kto chce zostać technicznym liderem, ale jeszcze nie jest na to gotowy?</i>	325
	<i>Jaka jest twoja ostatnia rada dla początkujących i młodszych badaczy danych?</i>	325
	Zasoby dla rozdziałów 13. – 16.	326
	Książki	326
	Blogi	327
	Epilog	329
	Dodatek. Pytania z rozmów kwalifikacyjnych	330

1



Czym jest danologia?

Niniejszy rozdział obejmuje:

- trzy główne obszary danologii,
- różne rodzaje stanowisk pracy związanych z badaniem danych.

„Najseksowniejsza praca XXI wieku”. „Najlepsza praca w Ameryce”. Danolog, inaczej badacz danych (ang. *data scientist*), to tytuł, który przed 2008 rokiem nawet nie istniał, a teraz określa stanowisko pracy, na które pracodawcy rekrutują na potęgę i o którym marzą rzesze osób poszukujących pracy. Istnieje ku temu dobry powód, ponieważ danologia to prężnie rozwijająca się dziedzina, z medianą podstawowego wynagrodzenia przekraczającą w Stanach Zjednoczonych (w 2019 roku) 100 tysięcy dolarów (<http://mng.bz/XpMp>). W dobrej firmie badacze danych cieszą się dużą autonomią i wciąż uczą się czegoś nowego. Wykorzystują swoje umiejętności do rozwiązywania istotnych problemów, np. współpracują z lekarzami w celu analizy eksperymentów klinicznych, pomagają drużynom sportowym w wyborze nowych graczy lub przeprojektowują modele cenowe dla branży widżetów. W końcu, tak jak to omawiamy w rozdziale 3., nie ma jednego sposobu na to, żeby stać się badaczem danych. Danolodzy mają zróżnicowane wykształcenie, więc nie jesteś w żaden sposób ograniczony kierunkiem swoich studiów.

Jednak nie wszystkie prace związane z badaniem danych są doskonałe. Zarówno firmy, jak i osoby poszukujące pracy mogą mieć nierealistyczne oczekiwania. Przedsiębiorstwa stawiające pierwsze kroki w dziedzinie nauki o danych mogą mieć mylne wyobrażenie, że jedna osoba zdoła rozwiązać wszystkie problemy przedsiębiorstwa, które są powiązane z danymi. Kiedy taki danolog zostanie w końcu zatrudniony, może stanąć

w obliczu niekończącej się listy życzeń i zażaleń. Może się też zdarzyć, że badaczowi danych powierzone zostanie natychmiastowe wdrożenie systemu uczenia maszynowego, chociaż nie wykonano wcześniej żadnej pracy w celu przygotowania lub oczyszczenia danych. Istnieje też szansa, że taki danolog nie będzie miał nikogo, kto mógłby mu doradzić lub udzielić wskazówek, a nawet wykazać się zwykłą empatią w kontekście problemów, z którymi borykają się badacze danych. Tematy te omówimy jednak bardziej szczegółowo w rozdziałach 5. i 7. — pomożemy Ci w nich uniknąć firm, które najprawdopodobniej okażą się złym wyborem dla świeżo upieczonego badacza danych. W rozdziale 9. poradzimy Ci zaś, co zrobić, jeśli znajdziesz się w niekorzystnej sytuacji.

Z drugiej strony osoby poszukujące pracy mogą sądzić, że na swojej nowej ścieżce kariery nigdy nie zazną chwili nudy. Mogą one bowiem oczekiwać, że interesariusze będą stosować się do ich zaleceń wręcz odruchowo, że inżynierowie danych będą mogli natychmiast rozwiązywać wszelkie problemy związane z jakością danych oraz że otrzymają najszybsze dostępne zasoby mocy obliczeniowej do wdrożenia swoich modeli. W rzeczywistości naukowcy zajmujący się danymi spędzają dużo czasu na czyszczeniu i przygotowywaniu danych, a także na zarządzaniu oczekiwaniami oraz priorytetami innych zespołów. Projekty nie zawsze się udają. Kadra zarządzająca wyższego szczebla potrafi składać klientom nierealistyczne obietnice dotyczące tego, co mogą dostarczyć firmowe modele badania danych. Może się też okazać, że głównym zadaniem zatrudnionego danologa jest praca z archaicznym systemem danych, którego nie da się zautomatyzować i który wymaga każdego tygodnia poświęcenia wielu godzin tytanicznej wręcz pracy umysłowej, i to tylko po to, by oczyścić dane. Bywa również tak, że badacze danych dostrzegają wiele statystycznych lub technicznych błędów w przeprowadzanych wcześniej analizach danych, które to błędy mają realne konsekwencje, ale nikt nie jest tym tematem zainteresowany, oni sami zaś są tak przeciążeni pracą, że nie mają czasu, by chociaż spróbować je naprawić. Danolog może też zostać poproszony o przygotowanie raportu, który będzie potwierdzał to, o czym zdecydowało już wcześniej kierownictwo wyższego szczebla, więc taki badacz danych może się obawiać zwolnienia z pracy, jeśli udzieli niezależnej odpowiedzi.

Niniejsza książka ma służyć temu, żeby poprowadzić Cię przez proces stawania się danologiem i rozwijania swojej kariery. Chcemy mieć pewność, że Ty, Czytelniku, poznasz wszystkie plusy tego zawodu i unikniesz większości czyhających na Ciebie pułapek. Być może pracujesz w podobnej branży, np. zajmujesz się analizą marketingową, i zastanawiasz się, jak dokonać zmiany w swoim życiu zawodowym. A może już jesteś danologiem, ale rozglądasz się za nową robotą i masz wrażenie, że za pierwszym razem źle zabrałeś się za szukanie pracy. Albo też, by rozwijać swoją karierę, zamierzasz przemawiać na konferencjach branżowych, przyczyniać się do rozwoju otwartego oprogramowania lub stać się niezależnym konsultantem. Niezależnie od Twojego poziomu zaawansowania jesteśmy pewne, że ta książka będzie dla Ciebie pomocna.

W pierwszych czterech rozdziałach omawiamy główne możliwości zdobycia umiejętności z zakresu danologii oraz tworzenie portfolio w celu obejścia paradoksu polegającego na tym, że trzeba mieć doświadczenie, żeby zdobyć doświadczenie. Część II pokazuje, jak należy napisać list motywacyjny i życiorys, które zapewnią Ci zaproszenie

na rozmowę kwalifikacyjną, oraz jak zbudować swoją sieć kontaktów, żeby ktoś Cię polecił. Omawiamy też strategie negocjacyjne, które — jak wykazały badania — zapewnią Ci najlepszą możliwą ofertę.

W pracy związanej z danologią będziesz tworzyć analizy, pracować z interesariuszami, a może nawet wprowadzać model danych do środowiska produkcyjnego. Część III pomoże Ci zrozumieć, jak wyglądają wszystkie te procesy i jak przygotować się na odniesienie sukcesu. W części IV znajdziesz strategie, które pomogą Ci stanąć na nogi po nieuchronnej porażce projektu. A kiedy już będziesz gotowy, my przeprowadzimy Cię przez proces podejmowania decyzji o tym, w jakim kierunku rozwinąć swoją karierę: czy przejść do kadry zarządzającej, pozostać indywidualnym pracownikiem, czy też wystartować jako niezależny konsultant.

Zanim jednak rozpoczniesz tę podróż, musisz mieć jasność co do tego, kim są badacze danych i jaką pracę wykonują. Danologia to szeroka dziedzina, która obejmuje wiele rodzajów pracy, a im lepiej rozumiesz różnice między nimi, tym lepiej będziesz mógł się w nich rozwijać.

1.1. Czym jest danologia?

Badanie danych to praktyka wykorzystywania danych w próbach zrozumienia i rozwiązania rzeczywistych problemów. Ta koncepcja nie jest całkiem nowa; ludzie analizują dane dotyczące sprzedaży i trendy od czasu wynalezienia zera. W ostatniej dekadzie jednak ilość danych, do których mamy dostęp, wzrosła wykładniczo. Pojawienie się komputerów pomogło w ich wygenerowaniu, ale komputery są również naszym jedynym sposobem, aby przetworzyć te tony informacji. Za pomocą kodu komputerowego zajmujący się danymi badacz może przetwarzać lub agregować dane, przeprowadzać analizy statystyczne czy też szkolić modele uczenia maszynowego. Rezultatem może być raport lub tzw. dashboard, przeznaczony do oglądania przez ludzi, albo też model uczenia maszynowego, który będzie wdrażany w sposób ciągły.

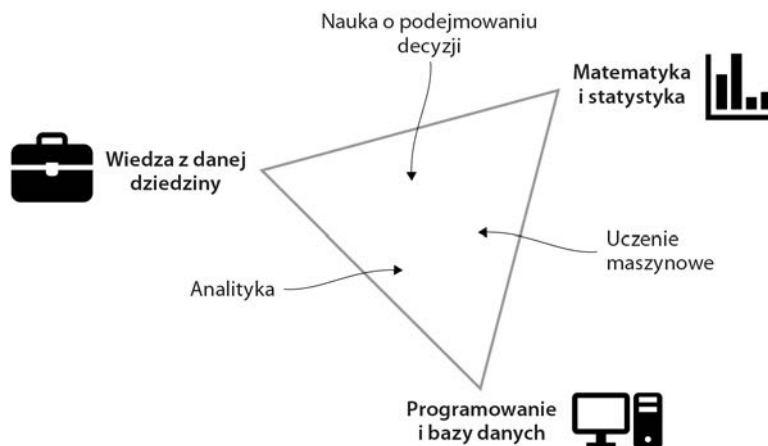
Jeżeli np. firma trudniąca się sprzedażą detaliczną ma problem z podjęciem decyzji o lokalizacji nowego sklepu, może wezwać na pomoc badacza danych, żeby przeprowadził analizę. Danolog mógłby przyjrzeć się danym historycznym dotyczącym miejsc, do których wysyłane są zamówienia online, żeby zrozumieć, jak rozkłada się zapotrzebowanie klientów na dane towary. Badacze danych mogą również połączyć dane dotyczące lokalizacji klientów z danymi demograficznymi oraz informacjami o dochodach dla tych miejscowości, pochodzącymi z rejestrów. Mając do dyspozycji takie zestawy danych, mogliby oni znaleźć optymalne miejsce do otwarcia nowego sklepu i przygotować prezentację w programie Microsoft PowerPoint, w której przedstawiliby swoje rekomendacje wiceprezesowi ds. sprzedaży detalicznej.

W innej sytuacji ta sama firma zajmująca się sprzedażą detaliczną mogłaby chcieć zwiększyć wielkość zamówień dokonywanych online przez polecenie produktów klientom, podczas gdy ci robią zakupy. Danolog może w takiej sytuacji wczytać historyczne dane dotyczące zamówień internetowych i stworzyć model uczenia maszynowego, który — biorąc pod uwagę zestaw przedmiotów znajdujących się w koszyku — będzie przewidywał,

jaki przedmiot najlepiej polecić kupującemu. Po stworzeniu takiego modelu badacz danych współpracowałby z zespołem inżynierów tej firmy w taki sposób, żeby za każdym razem, gdy klient dokonuje zakupów, nowy model uczenia maszynowego był wykorzystywany do rekomendowania produktów.

Wiele osób, gdy zaczynają zajmować się danologią, jest przytłoczonych ze względu na ilość rzeczy, których muszą się nauczyć, takich jak programowanie („A w jakim języku?”), statystyka („A które metody są najczęściej wykorzystywane w praktyce, a jakie z nich mają w dużej mierze akademicki charakter?”), uczenie maszynowe („A czym różni się uczenie maszynowe od statystyki czy SI?”) czy też wiedza z zakresu branży, w której chcą pracować („A jeśli nie wiesz, gdzie chcesz pracować?”). Ponadto osoby te muszą się nauczyć umiejętności biznesowych, takich jak skuteczne przekazywanie wyników różnym odbiorcom, od innych badaczy danych po prezesa firmy. Niepokój ten może się nasilić w związku ze stanowiskami pracy, na których wymaga się od kandydatów stopnia naukowego doktora, wieloletniego doświadczenia w badaniu danych oraz wiedzy specjalistycznej w zakresie metod statystycznych i programistycznych. Jak można się nauczyć tych wszystkich umiejętności? Od których powinieneś zacząć? Jakie są podstawy?

Jeśli rzuciłeś okiem na różne obszary danologii, to być może jesteś zaznajomiony z popularnym diagramem Venna w zakresie danologii autorstwa Drew Conwaya. Zdaniem Conwaya (w momencie tworzenia diagramu) danologia znajdowała się na skrzyżowaniu wiedzy matematycznej i statystycznej, wiedzy specjalistycznej w danej dziedzinie oraz umiejętności hakerskich (czyli programowania). Rysunek ten jest często wykorzystywany w charakterze kamienia węgielnego służącego do określenia, czym jest danologia. Z naszej perspektywy komponenty danologii są nieco odmienne od tego, co zaproponował Conway (rysunek 1.1).



Rysunek 1.1. Umiejętności, które łączą się w danologię, i to, jak łączą w sobie różne role

Zmieniliśmy oryginalny diagram Venna opracowany przez Conwaya na trójkąt, ponieważ nie chodzi o to, czy posiadasz daną umiejętność, czy też nie, lecz o to, że możesz mieć ją wyrobioną w innym stopniu niż inne osoby w tej dziedzinie. I chociaż prawdą

jest to, że wszystkie trzy umiejętności są fundamentalne i że trzeba mieć każdą z nich opanowaną w pewnym stopniu, to jednak nie musisz być ekspertem we wszystkich dziedzinach. W trójkącie umieściliśmy różne rodzaje specjalizacji z zakresu danologii. Te specjalizacje nie zawsze przekładają się jeden do jednego w kwestii nazw stanowisk pracy, a nawet jeśli tak jest, to różne firmy czasami nazywają je po swojemu.

Co więc oznacza każdy z tych komponentów?

1.1.1. Matematyka/statystyka

Na poziomie podstawowym wiedza z zakresu matematyki i statystyki to biegłość w posługiwaniu się danymi. Dzielimy tę umiejętność na trzy poziomy wiedzy:

- *Ta technika istnieje* — Jeżeli nie wiesz, że coś jest możliwe, nie możesz tego wykorzystać. Gdyby badacz danych próbował grupować podobnych klientów — wiedząc, że można to zrobić przy użyciu metody statystycznej (zwanej *grupowaniem*) — byłby to pierwszy krok w tym kierunku.
- *Jak stosować techniki* — Chociaż danolog może znać wiele technik, to jednak musi być w stanie zrozumieć złożoność ich stosowania, czyli nie tylko wiedzieć, jak pisać kod potrzebny do stosowania metod, ale także jak je konfigurować. Jeśli badacz danych chce użyć metody takiej jak algorytm centroidów w celu pogrupowania klientów, musi zrozumieć, jak wykonać algorytm centroidów w języku programowania, takim jak R lub Python. Danolog musiałby również zrozumieć, jak dostosować parametry metody, i np. wybrać, ile grup należy utworzyć.
- *Jak wybrać techniki, które należy wypróbować* — Ponieważ istnieje tak wiele dających się zastosować w danologii technik, ważne jest to, aby badacz był w stanie szybko ocenić, czy dana technika będzie dobrze działać. W naszym przykładzie dotyczącym grupowania klientów nawet wtedy, gdy skupia się na grupowaniu danych, musi rozważyć kilkadziesiąt różnych metod i algorytmów. Zamiast próbować każdej z metod, musi być w stanie szybko wykluczyć konkretne metody i skupić się tylko na kilku.

Tego rodzaju umiejętności są stale wykorzystywane w pracy związanej z badaniem danych. Aby rozważyć inny przykład, załóżmy, że pracujesz w przedsiębiorstwie e-commerce. Twój partner biznesowy może być zainteresowany tym, które kraje mają najwyższą średnią wartość zamówień. Jeśli masz te dane pod ręką, to łatwo możesz odpowiedzieć na to pytanie. Ale zamiast po prostu przedstawiać te informacje klientowi i pozwolić mu wyciągnąć własne wnioski, możesz pójść o krok dalej. Jeśli masz jedno zamówienie z kraju A za 100 dolarów, a z kraju B tysiąc zamówień o średniej wartości zamówienia 75 dolarów, będzie prawdą, że kraj A ma wyższą średnią wartość zamówienia. Jednak czy możesz mieć pewność, że Twój partner biznesowy powinien inwestować w reklamę w kraju A, żeby zwiększyć liczbę zamówień? Raczej nie. Masz tylko jeden punkt danych dla kraju A, a może to być wartość odstająca. Jeśli w kraju A miałbyś zamiast tego 500 zamówień, mógłbyś użyć testu statystycznego, żeby sprawdzić, czy wartość zamówienia była znacząco różna, co oznacza, że jeśli rzeczywiście nie byłoby różnicy między pomiarami A i B, to byłoby mało prawdopodobne, żebyś zobaczył tę

różnicę, którą dostrzegłeś na pierwszy rzut oka. W tym długim na cały akapit przykładzie dokonaliśmy wielu różnych ocen dotyczących tego, jakie podejścia są rozsądne, co należy rozważyć i jakie wyniki uznano za nieistotne.

1.1.2. Bazy danych/programowanie

Programowanie i bazy danych odnoszą się do możliwości pobierania danych z firmowych baz danych i pisania czystego, wydajnego, łatwego do utrzymania kodu. Umiejętności te są pod wieloma względami podobne do tego, co musi wiedzieć programista, z tą różnicą, że naukowcy zajmujący się danymi muszą napisać kod, który przeprowadzi otwartą analizę, a nie wygeneruje z góry zdefiniowanego wyniku. Każdy stos danych w firmie jest unikatowy, więc nie wymaga się od badacza danych żadnego zestawu umiejętności technicznych. Ale ogólnie rzecz ujmując, musisz wiedzieć, jak pozyskiwać dane z bazy danych, jak je czyścić, manipulować nimi, podsumowywać je, wizualizować i udostępnić.

W większości zawodów związanych z danologią R lub Python stanowią główne języki programowania. R to język, który ma swoje korzenie w statystyce, więc jest najsilniejszy w analizie i modelowaniu statystycznym, wizualizacji oraz generowaniu raportów z wynikami. Python zaczął swoją karierę jako ogólny język tworzenia oprogramowania i z czasem zyskał niezwykłą popularność w danologii. W powszechnej opinii Pythona uważa się za język lepszy od R w pracy z dużymi zbiorami danych, w uczeniu maszynowym i zasilaniu algorytmów w czasie rzeczywistym (takich jak silniki rekomendacyjne firmy Amazon). Jednak dzięki pracy wielu członków społeczności możliwości obu języków są obecnie prawie równe. Badacze danych z powodzeniem używają R do tworzenia modeli uczenia maszynowego, które są uruchamiane miliony razy w tygodniu, a także wykonują czyste, nadające się do przedstawienia analizy statystyczne w Pythonie.

R i Python są najpopularniejszymi językami programowania w danologii z kilku powodów:

- Są darmowe i dostępne w ramach otwartego oprogramowania, co oznacza, że wiele osób, a nie tylko jedna firma czy jedna grupa, udostępnia kod programistyczny, którego można użyć. Mają mnóstwo pakietów lub *bibliotek* (zbiorów kodu) do gromadzenia danych, manipulowania nimi, wizualizacji, analizy statystycznej i uczenia maszynowego.
- Co ważne, ponieważ każdy z tych języków jest tak powszechnie stosowany, badaczom danych łatwo jest znaleźć pomoc, gdy napotykają na problemy. Chociaż niektóre firmy nadal używają SAS, SPSS, STATA, MATLAB i innych komercyjnych programów, wiele przedsiębiorstw zaczyna przechodzić na R lub Pythona.

Chociaż większość analiz danologicznych jest wykonywana w R lub Pythonie, często trzeba pracować z bazą danych, aby pozyskać te dane. W tym miejscu pojawia się SQL. SQL jest językiem programowania, którego większość baz danych używa do wewnętrznego manipulowania danymi lub do ich wyodrębniania. Wczuj się w sytuację badacza danych, który chce przeanalizować setki milionów rekordów zamówień złożonych w danej firmie, aby przewidzieć, jak będą się zmieniać w ciągu dnia. Po pierwsze, taki danolog prawdopodobnie napisałby zapytanie SQL, aby uzyskać liczbę zamówień z każdego dnia.

Następnie wykorzystalby tak uzyskane (dzienne) wyniki, żeby wykonać prognozę statystyczną w R lub Pythonie. Z tego powodu SQL jest niezwykle popularny w środowisku danologów i dlatego też nie zajdziesz zbyt daleko bez jego znajomości.

Inną podstawową umiejętnością jest wykorzystanie *systemów kontroli wersji* — metody śledzenia zmian kodu w czasie. Kontrola wersji umożliwia zapisywanie plików, przywracanie ich stanu z poprzedniego czasu i sprawdzanie, kto, jak i kiedy zmienił dany plik. Opanowanie tej umiejętności jest niezwykle ważne dla danologii oraz inżynierii oprogramowania, ponieważ jeśli ktoś przypadkowo zmienił plik, przez co zepsuł Twój kod, chcesz mieć możliwość odwrócenia takiej zmiany lub zobaczenia, co się zmieniło.

Git to zdecydowanie najczęściej stosowany system kontroli wersji, często używany w połączeniu z GitHubem, internetowym serwisem hostingowym dla Gita. Git pozwala zapisać (zatwierdzić) Twoje zmiany, jak również zobaczyć całą historię projektu i to, jak zmieniał się on wraz z każdym zatwierdzeniem (ang. *commit*). Jeśli dwie osoby pracują oddzielnie nad tym samym plikiem, Git dba o to, żeby żadne zmiany nie zostały przypadkowo usunięte lub nadpisane. W wielu firmach, szczególnie tych z silnymi zespołami programistycznymi, będziesz musiał użyć Gita, jeśli zechcesz podzielić się swoim kodem lub wprowadzić coś do środowiska produkcyjnego.

Czy można być badaczem danych bez umiejętności programowania?

Możliwe jest wykonywanie ogromnej ilości pracy z danymi przy użyciu wyłącznie Excela, Tableau lub innego narzędzia do analizy biznesowej, które ma graficzny interfejs użytkownika. I chociaż nie piszesz kodu, to twórcy tych narzędzi twierdzą, że mają one wiele takich samych funkcji, co języki programowania — takie jak R czy Python — i wielu badaczy danych czasem z nich korzysta. Ale czy mogą one być kompletnym zestawem narzędzi danologicznych? Naszym zdaniem — nie. Z pragmatycznego punktu widzenia — naprawdę niewiele przedsiębiorstw ma zespół zajmujący się badaniem danych, w którym nie trzeba programować. Ale nawet gdyby tak nie było, programowanie miałoby przewagę nad stosowaniem tych narzędzi.

Pierwszą zaletą programowania jest jego powtarzalność. Kiedy piszesz kod programistyczny, zamiast korzystać z oprogramowania typu „wskaż myszą i kliknij”, możesz uruchomić go ponownie za każdym razem, gdy zmieniają się Twoje dane, niezależnie od tego, czy dzieje się to codziennie czy raz na 6 miesięcy. Tę zaletę ma także system kontroli wersji: zamiast zmieniać nazwę swojego pliku za każdym razem, gdy zmienia się kod, możesz zachować jeden plik, ale zarazem widzieć całą historię zmian.

Drugą zaletą jest elastyczność. Jeśli np. w Tableau konkretny typ wykresu nie jest dostępny, to nie będziesz mógł go utworzyć. Jednak dzięki programowaniu możesz napisać własny kod, aby zrobić coś, o czym twórcy i serwisanci narzędzia nigdy nie pomyśleli.

Trzecią zaletą języków programistycznych dostępnych w ramach ruchu otwartego oprogramowania, takich jak Python i R, jest wkład społeczności. Tysiące ludzi tworzą pakiety i publikują je otwarcie na GitHubie i/lub CRAN-ie (R) i czy pipie (Python). Możesz pobrać ten kod i wykorzystać go do rozwiązania własnych problemów programistycznych. Nie jesteś zależny od jednej firmy lub grupy ludzi, którzy dodają funkcje.

1.1.3. Zrozumienie biznesu

Każda wystarczająco zaawansowana technologia jest nie do odróżnienia od magii.

Arthur C. Clarke

Środowiska biznesowe w różny sposób, delikatnie rzecz ujmując, rozumieją to, jak funkcjonuje danologia. Często się zdarza, że kierownictwo po prostu chce coś zrobić i dlatego zwraca się do swoich niesamowitych badaczy danych, aby to zrobili. Kluczową umiejętnością w danologii jest wiedza, jak przełożyć sytuację biznesową na pytanie dotyczące danych, znaleźć na nie odpowiedź i wreszcie przetłumaczyć tę odpowiedź na język biznesu. Kierownik może na przykład zapytać: „Dlaczego nasi klienci odchodzą?”. Jednak nie istnieje pakiet Pythona o nazwie „dlaczego-klienci-odchodzą”, który można by zaimportować — to od Ciebie zależy, jak odpowiesz na to pytanie przy użyciu danych.

O zrozumieniu biznesu można mówić wtedy, gdy Twoje danologiczne ideały spotykają się z praktyczną stroną świata rzeczywistego. Nie wystarczy chcieć konkretnej informacji, nie wiedząc przy tym, jak dane są przechowywane i aktualizowane. Jeśli Twoja firma oferuje usługi w modelu abonamentowym, to gdzie znajdują się dane? Jeżeli ktoś zmieni swoją subskrypcję, to co się wtedy stanie? Czy ten rekord dotyczący abonenta jest aktualizowany, czy też do tabeli dodawany jest kolejny rekord? Czy musisz szukać sposobów na obchodzenie błędów lub niespójności danych? Jeśli nie znasz odpowiedzi na te pytania, nie będziesz w stanie udzielić dokładnej odpowiedzi na proste pytanie, takie jak „Ilu subskrybentów mieliśmy 2 marca 2019 roku?”.

Zrozumienie biznesu pomaga również osiągnąć wiedzę o tym, jakie pytania należy zadać. Zadane przez interesariusza pytanie „Co powinniśmy zrobić w następnej kolejności?” należałoby rozumieć raczej jak „Dlaczego nie mamy więcej pieniędzy?”. Ten rodzaj pytania rodzi dalsze. Zrozumienie podstaw działalności biznesowej (jak również zaangażowanych w nią osób) może pomóc w ocenie sytuacji. Możesz więc się dopytać: „Dla której linii produktów szukasz porady?” lub „Chciałbyś zwiększyć udział konkretnej grupy osób z naszej puli klientów?”.

Inną częścią zrozumienia biznesu jest rozwijanie ogólnych kompetencji w tym zakresie, takich jak umiejętność dostosowania prezentacji i raportów do różnych odbiorców. Raz będziesz dyskutować na temat tego, która metodyka jest lepsza, z pełną salą osób mających doktorat ze statystyki, a innym razem staniesz przed wiceprezesem, który ostatni raz był na lekcji matematyki 20 lat temu. Musisz informować swoich słuchaczy, nie traktując ich przy tym z góry ani nie komplikując zbytnio swojego przekazu.

Wreszcie, w miarę jak będziesz nabierać coraz więcej doświadczenia, częścią Twojej pracy stanie się określenie, w jakich obszarach firma może czerpać korzyści z badania danych. Jeśli chciałeś zbudować system prognozowania dla swojej firmy, ale nigdy nie otrzymałeś wsparcia ze strony kierownictwa, dołączenie do kadry zarządzającej może pomóc w rozwiązaniu tego problemu. Starszy badacz danych będzie szukał miejsc do wdrożenia uczenia maszynowego, ponieważ zna jego ograniczenia i możliwości, a także wie, jakie typy zadań skorzystałyby na automatyzacji.

Czy nauka o danych zaniknie?

U podstaw pytania o to, czy danologia wciąż będzie istnieć za dekadę lub dwie, leży obawa, że praca zostanie zautomatyzowana oraz że nauka o danych jest przereklamowana i bańka, która pojawiła się na rynku pracy, pęknie.

To prawda, że niektóre części danologicznych potoków mogą zostać zautomatyzowane. Zautomatyzowane uczenie maszynowe (ang. *Automated Machine Learning*, AutoML) może porównywać wydajność różnych modeli i wykonywać pewne części pracy związanej z przygotowaniem danych (np. skalowanie zmiennych). Jednak zadania te stanowią tylko niewielką część procesu danologicznego. Często trzeba na przykład samemu utworzyć dane; bardzo rzadko się zdarza, żeby czekały na Ciebie idealnie czyste dane. Ponadto tworzenie danych zazwyczaj wiąże się z rozmową z innymi osobami, takimi jak researcher badający doświadczenia użytkownika czy też inżynier przeprowadzający ankietę lub rejestrujący działania użytkownika, które mogą stanowić podstawę do przeprowadzenia analizy.

Jeśli chodzi o możliwość pęknięcia bańki na rynku pracy, to dobrym porównaniem będzie rozwój oprogramowania w latach 80. Ponieważ komputery stawały się coraz tańsze, szybsze i bardziej powszechne, pojawiły się obawy, że wkrótce komputer sam zrobi wszystko i nie będzie już zapotrzebowania na programistów. Sytuacja jednak rozwinęła się w zupełnie odwrotnym kierunku i teraz w samych Stanach Zjednoczonych mamy ponad 1,2 miliona programistów (<http://mng.bz/MOPo>). I chociaż takie tytuły zawodowe jak webmaster właściwie zaniknęły, obecnie więcej osób niż kiedykolwiek pracuje nad rozwijaniem, utrzymaniem i ulepszaniem stron internetowych.

Uważamy, że w ramach danologii nasili się specjalizacja, co może doprowadzić do zaniku ogólnego tytułu badacza danych, ale wiele firm jest wciąż na wczesnym etapie nauki o tym, jak wykorzystać badania danych, i ma przed sobą jeszcze wiele pracy na tym polu.

1.2. Różne rodzaje prac związanych z badaniem danych

Można łączyć i dopasowywać trzy podstawowe umiejętności z zakresu danologii (omówione w podrozdziale 1.1) do kilku stanowisk, w przypadku których powiązanie ich z tytułem badacza danych ma pewne uzasadnienie. Z naszej perspektywy umiejętności te mieszają się ze sobą w trzech głównych obszarach: analizie, uczeniu maszynowym i nauce o podejmowaniu decyzji. Każdy z tych obszarów służy innemu celowi firmy i co do zasady dostarcza czegoś innego.

Szukając pracy związanej z danymi naukowymi, powinieneś zwracać mniejszą uwagę na tytuły zawodowe, a większą na opisy stanowisk i to, o co jesteś pytany podczas rozmów kwalifikacyjnych. Rzuć okiem na doświadczenie zawodowe osób pełniących funkcje związane z badaniem danych, sprawdź np., gdzie i w jakim charakterze poprzednio pracowały, oraz to, jakie mają wykształcenie. Może się okazać, że ludzie, którzy pracują na stanowiskach o podobnie brzmiących nazwach, mają zupełnie inne tytuły zawodowe, lub że dwie osoby zatrudnione w charakterze danologa robią zupełnie inne rzeczy. Kiedy więc wspominamy na łamach niniejszej książki o różnych rodzajach pracy związanej z badaniem danych, należy pamiętać, że rzeczywiste tytuły zawodowe, używane w przedsiębiorstwach, mogą się różnić.

1.2.1. Analiza

Analitik bierze dane i przedstawia je właściwym osobom. Po tym, jak firma ustali roczne cele, można umieścić je na dashboardzie, aby kierownictwo mogło co tydzień śledzić postępy. Można również wbudować funkcje, które pozwalają menedżerom na łatwe podzielenie osiąganych rezultatów według kraju lub typu produktu. Praca ta wymaga dużo czyszczenia i przygotowywania danych, ale na ogół mniej pracy przy ich interpretacji. I chociaż należy umieć dostrzegać i rozwiązywać problemy związane z jakością danych, główną osobą podejmującą decyzje na podstawie tych danych jest partner biznesowy. Zadaniem analityka jest więc zbieranie danych z firmy, ich formatowanie, a także efektywne porządkowanie oraz dostarczanie innym.

Ponieważ rola analityka nie wiąże się z dużą ilością pracy z zakresu statystyki oraz uczenia maszynowego, niektóre osoby i przedsiębiorstwa uważają, że to stanowisko wykracza poza dziedzinę badania danych. Jednak znaczna część prac, takich jak tworzenie sensownych wizualizacji i podejmowanie decyzji o konkretnych przekształceniach danych, wymaga tych samych umiejętności, które są wykorzystywane w innych typach ról nauki o danych. Analityk może otrzymać zadanie polegające na „stworzeniu zautomatyzowanego dashboardu, który pokaże, jak liczba naszych abonentów zmienia się w czasie, i pozwoli nam filtrować dane tylko dla abonentów określonych produktów lub w określonych strefach geograficznych”. Analityk musiałby znaleźć odpowiednie dane wewnątrz firmy, dowiedzieć się, jak odpowiednio przekształcić dane (np. przez zamianę kategorii nowych subskrypcji z codziennych na cotygodniowe), a następnie stworzyć sensowny zestaw dashboardów, które byłyby atrakcyjne wizualnie, a także bezbłędnie automatycznie aktualizowane każdego dnia.

W skrócie: analityk tworzy *dashboardy i raporty dostarczające danych*.

1.2.2. Uczenie maszynowe

Inżynier zajmujący się *uczeniem maszynowym* opracowuje modele uczenia maszynowego i wprowadza je do środowiska produkcyjnego, gdzie pracują one w sposób ciągły. Może on zoptymalizować algorytm rankingu dla wyników wyszukiwania na stronie e-commerce, stworzyć system rekomendacji lub monitorować model w środowisku produkcyjnym, aby się upewnić, że jego wydajność nie spadła od czasu wdrożenia. Inżynier zajmujący się uczeniem maszynowym spędza mniej czasu na takich rzeczach jak tworzenie wizualizacji, które przekonają ludzi do czegoś, za to poświęca go więcej na pracę programistyczną w zakresie badania danych.

Dużą różnicą między tą rolą a innymi rodzajami stanowisk danologicznych jest to, że wyniki pracy są przede wszystkim przeznaczone do celów maszynowej konsumpcji. Możesz stworzyć modele uczenia maszynowego, które zostaną przekształcone na przykład w interfejsy programowania aplikacji (ang. *application programming interface*, API) dla innych maszyn. Pod wieloma względami na tym stanowisku będzie Ci bliżej do programisty niż innych ról z zakresu danologii. O ile każdy badacz powinien przestrzegać dobrych praktyk programistycznych, o tyle jako inżynier zajmujący się uczeniem maszynowym musisz to robić. Twój kod musi być wydajny, przetestowany i napisany w taki

sposób, żeby inni ludzie mogli z nim pracować. Z tego powodu wielu inżynierów zajmujących się uczeniem maszynowym wywodzi się ze środowiska programistycznego.

Pracując w roli inżyniera zajmującego się uczeniem maszynowym, możesz zostać poproszony o stworzenie modelu uczenia maszynowego, który będzie w czasie rzeczywistym przewidywał prawdopodobieństwo, że klient na stronie internetowej rzeczywiście dokończy swoje zamówienie. Inżynier zajmujący się uczeniem maszynowym musiałby w takiej sytuacji znaleźć w firmie historyczne dane, przeszkolić na nich model uczenia maszynowego, przekształcić model w API, a następnie wdrożyć to API, żeby strona mogła uruchomić model. Jeśli model z jakiegoś powodu przestanie działać, to właśnie inżynier zajmujący się uczeniem maszynowym zostanie wezwany do jego naprawy.

W skrócie: inżynier zajmujący się uczeniem maszynowym tworzy modele, które działają w sposób ciągły.

1.2.3. Nauka o podejmowaniu decyzji

Badacz pomagający w podejmowaniu decyzji przekształca surowe dane przedsiębiorstwa w informacje, które ułatwiają kierownictwu podejmowanie decyzji. Praca ta opiera się na dogłębnym zrozumieniu różnych metod matematycznych i statystycznych oraz znajomości procesu podejmowania decyzji biznesowych. Co więcej, badacze pomagający w podejmowaniu decyzji muszą być w stanie stworzyć przekonujące wizualizacje i tabele, aby osoby nietechniczne, z którymi rozmawiają, zrozumiały ich analizę. Mimo że badacze pomagający w podejmowaniu decyzji poświęcają mnóstwo czasu na programowanie, swoją pracę zazwyczaj wykonują tylko raz, aby przeprowadzić konkretną analizę, dzięki czemu może im ująć na sucho napisanie kodu, który jest nieefektywny lub trudny do utrzymania.

Badacz pomagający w podejmowaniu decyzji musi zrozumieć potrzeby innych ludzi w firmie i dowiedzieć się, jak wygenerować konstruktywne informacje. Na przykład dyrektor ds. marketingu może poprosić badacza pomagającego w podejmowaniu decyzji o doradzenie mu, które rodzaje produktów powinny być wyróżnione w firmowym przewodniku po prezentach świątecznych. Badacz pomagający w podejmowaniu decyzji może zbadać, jakie produkty dobrze się sprzedawały mimo nieumieszczenia ich w przewodniku, porozmawiać z zespołem zajmującym się badaniami użytkowników o przeprowadzeniu ankiety oraz zastosować zasady behawioryzmu, żeby wykonać analizę i zaproponować optymalne produkty. Wynikiem takiego działania będzie prawdopodobnie prezentacja PowerPoint lub raport, który zostanie udostępniony menedżerom produktów, wiceprezesom oraz innym pracownikom biznesowym.

Badacz pomagający w podejmowaniu decyzji często wykorzystuje swoją wiedzę z zakresu statystyki, żeby doradzić firmie w warunkach niepewności. Badacz pomagający w podejmowaniu decyzji mógłby na przykład być odpowiedzialny za wprowadzenie w firmie systemu analizy eksperymentów. Wiele firm prowadzi eksperymenty online lub testy A/B, żeby zmierzyć, czy zmiana jest skuteczna. Zmiana może być tak prosta jak dodanie nowego przycisku albo tak skomplikowana jak zmodyfikowanie systemu rankingowego wyników wyszukiwania, może też polegać na całkowitym przeprojektowaniu strony. Podczas testów A/B odwiedzający są losowo przypisywani do jednego z dwóch

(lub większej liczby) warunków, np. połowa do starej strony głównej, która jest *wersją A (bieżącą, zwaną również kontrolną)*, a połowa do nowej, która jest *wersją B (zmodyfikowaną, określaną także mianem kuracji)*. Następnie porównuje się działania osób odwiedzających poszczególne strony po rozpoczęciu eksperymentu, żeby sprawdzić, czy osoby przekierowane do wersji B (czyli poddane kuracji) mają wyższy wskaźnik wykonania pożądaných działań, takich jak zakup produktów.

Ze względu na przypadkowość rzadko się zdarza, aby metryka w wersji A i B była dokładnie taka sama. Załóżmy, że rzuciłeś dwiema monetami i że w przypadku jednej reszka wypadła 52 razy na 100, przy drugiej zaś 49 razy na 100. Czy stwierdziłbyś większe prawdopodobieństwo, że w przypadku pierwszej monety wypadnie reszka? Oczywiście, że nie! Jednak partner biznesowy może przyjrzeć się eksperymentowi, zobaczyć, że współczynnik konwersji wynosi 5,4% w wersji A i 5,6% w wersji B, po czym uznać kurację za skuteczną. Zadaniem badacza pomagającego w podejmowaniu decyzji jest wsparcie w interpretowaniu danych, egzekwowanie dobrych praktyk w zakresie projektowania eksperymentów i nie tylko.

W skrócie: badacz pomagający w podejmowaniu decyzji tworzy analizy, w których wyniku powstają *zalecenia*.

1.2.4. Pokrewne prace

Chociaż te trzy obszary, omówione w poprzednich punktach, stanowią główne rodzaje stanowisk pracy związanych z badaniem danych, można zauważyć kilka innych, odrębnych ról, które wykraczają poza te kategorie. Wymieniamy tutaj te stanowiska, ponieważ dobrze jest rozumieć dostępne oferty pracy, jak również dlatego, że być może trzeba będzie współpracować z ludźmi zatrudnionymi na tych stanowiskach. Jeżeli jednak jesteś zainteresowany jedną z tych ról, materiał zawarty w tej książce może być dla Ciebie mniej relevantny.

ANALITYK BIZNESOWY

Analitik biznesowy pracuje podobnie jak *analitik*, ale na ogół wykorzystuje mniej wiedzy statystycznej i programistycznej. Jego narzędziem (z wyboru) może być Excel zamiast Pythona, a poza tym istnieje spora szansa, że nigdy nie będzie tworzył modeli statystycznych. I chociaż jego stanowisko pracy jest podobne do funkcji *analitika*, to generuje on mniej wyrafinowane wyniki ze względu na ograniczenia będące rezultatem użytych narzędzi i technik.

Jeśli chcesz zajmować się uczeniem maszynowym lub programowaniem albo stosować metody statystyczne, stanowisko *analitika biznesowego* może być bardzo frustrujące, ponieważ nie pomoże Ci ono w zdobyciu nowych umiejętności. Poza tym te stanowiska są zazwyczaj gorzej opłacane niż prace związane z badaniem danych, a także uważa się je za mniej prestiżowe. Jednak praca *analitika biznesowego* może być dobrym punktem wyjścia do tego, by zostać badaczem danych w przyszłości, zwłaszcza jeśli nie pracowałeś wcześniej z danymi w środowisku biznesowym. Jeśli chcesz rozpocząć pracę jako *analitik biznesowy*, rozwijać się i przejść na stanowisko *danologa*, poszukaj takich ofert, które pozwolą Ci nauczyć się nowych umiejętności, np. programowania w R lub Pythonie.

INŻYNIER DANYCH

Inżynier danych koncentruje się na dbaniu o dane zawarte w bazach danych i na zapewnieniu tego, że ludzie mogą uzyskać potrzebne im dane. Nie sporządza raportów, nie analizuje i nie opracowuje modeli, lecz przechowuje — w uporządkowany sposób — dane w bazach danych po to, żeby inni mogli robić te wszystkie rzeczy. Inżynier danych może otrzymać zadanie utrzymywania wszystkich rekordów klientów w dużej bazie danych w chmurze i dodawania nowych tabel do tej bazy danych zgodnie z wymaganiami.

Inżynierowie danych wykonują inną pracę od danologów i są jeszcze rzadsi i bardziej poszukiwani. Inżynier danych może pomóc zbudować komponenty zaplecza (ang. *back-end*) danych dla wewnętrznego eksperymentalnego systemu firmy i aktualizować przepływ przetwarzania danych, gdy zadania zaczynają zabierać zbyt wiele czasu. Inżynierowie danych zajmują się również opracowywaniem i monitorowaniem środowisk wsadowych i strumieniowych, zarządzając danymi od momentu ich zebrania, poprzez ich przetwarzanie, aż po przechowywanie.

Jeśli jesteś zainteresowany inżynierią danych, będziesz potrzebować dużych umiejętności informatycznych — wielu inżynierów danych to byli programiści.

RESEARCHER

Researcher opracowuje i wdraża nowe narzędzia, algorytmy oraz metodyki, często w celu ich wykorzystania przez danologów, którzy zajmują się danymi w firmie. Tego typu stanowiska prawie zawsze wymagają zrobienia doktoratu, zwykle z dziedziny informatyki, statystyki, ilościowych nauk społecznych lub pokrewnej. *Researcher* może spędzić całe tygodnie na badaniu i wypróbowywaniu metod mających na celu zwiększenie skuteczności eksperymentów online, uzyskując o 1% większą dokładność w rozpoznawaniu obrazów w samochodach autonomicznych lub tworząc nowy algorytm głębokiego uczenia. Może nawet poświęcać czas na pisanie prac badawczych, które będą rzadko wykorzystywane w przedsiębiorstwie, ale które pomogą zwiększyć prestiż firmy i (w idealnym scenariuszu) poczynić postępy w tej dziedzinie. Ponieważ te stanowiska wymagają bardzo specyficznego wykształcenia i doświadczenia, nie skupiamy się na nich w niniejszej książce.

1.3. Wybór swojej drogi

W rozdziale 3. omawiamy niektóre możliwości pozyskania umiejętności w zakresie danologii, zalety i wady każdej z nich, a także kilka sugestii dotyczących dokonania wyboru. W tym momencie dobrze jest zacząć zastanawiać się nad obszarem badania danych, w którym chcesz się specjalizować. W jakich dziedzinach masz już doświadczenie? Widzieliśmy już danologów, którzy są byłymi inżynierami, profesorami psychologii, menedżerami ds. marketingu, studentami statystyki, a także pracownikami socjalnymi. W wielu przypadkach wiedza, którą zdobyłeś w innych zawodach i na studiach, może pomóc Ci stać się lepszym badaczem danych. Jeśli działasz już na polu danologii, warto teraz zastanowić się nad tym, w której części trójkąta się znajdujesz. Czy jesteś z tego zadowolony? A może chcesz zmienić pracę na inną związaną z badaniem danych? Często taka zmiana jest możliwa.

Vicki Boykis: Czy każdy może zostać danologiem?

Biorąc pod uwagę optymizm związany z badaniem danych (i potencjalnie duże zarobki, o których wspomina się w artykułach prasowych), łatwo jest zrozumieć, dlaczego danologia oferuje atrakcyjne możliwości rozwoju zawodowego — zwłaszcza że zakres obowiązków i różnorodność tytułów zawodowych związanych z badaniem danych stale się powiększają. Ale ważne jest, żebyś, jako nowicjusz w tej dziedzinie, miał realistyczny, zróżnicowany pogląd na to, w jakim kierunku będzie zmierzać rynek danologiczny w ciągu najbliższych kilku lat, i odpowiednio się do tego dostosować.

Istnieje kilka trendów wpływających dziś na dziedzinę badania danych. Po pierwsze, danologia jako dziedzina istnieje już od 10 lat i jako taka przeszła przez wczesne etapy cyklu rozgłosu: zyskanie rozgłosu w mass mediach, wczesne przyjęcie i konsolidację. Zyskała ona nadmierny rozgłos, mówiono o niej w środkach masowego przekazu, została zaadaptowana przez firmy z Doliny Krzemowej (i nie tylko przez nie), a teraz nastąpiło duże przyspieszenie w jej przyjmowaniu w większych firmach oraz standaryzacji przepływu badania danych z użyciem takich zestawów narzędzi jak Spark czy AutoML.

Po drugie, w wyniku tego mamy do czynienia z nadpodażą nowych badaczy danych, po intensywnych kursach szkoleniowych, nowo powstałych kierunkach danologii na uniwersytetach i wreszcie po kursach internetowych. Liczba kandydatów na każde stanowisko badacza danych, w szczególności na poziomie podstawowym, wzrosła z około 20 do 100 i więcej osób na miejsce. Liczba 500 nadesłanych życiorysów na nową ofertę pracy nie jest już rzadkością.

Po trzecie, standaryzacja zestawów narzędzi i gotowa podaż pracy, a także zapotrzebowanie na ludzi, którzy mają większe doświadczenie w tej dziedzinie, przyniosły zmianę w sposobie nazywania stanowisk pracy, jak również spowodowały stworzenie hierarchii zawodów i opisów z zakresu danologii. Na przykład w niektórych firmach praca w charakterze „badacza danych” może oznaczać tworzenie modeli, ale w innych wiąże się głównie z prowadzeniem analiz przy użyciu SQL, czyli stanowi odpowiednik tego, czym kiedyś zajmował się analityk danych.

Dla tych, którzy (jako nowicjusze) chcą wejść do świata danologii, oznacza to kilka rzeczy. Po pierwsze, i najważniejsze, rynek pracy może się okazać dla nich mocno konkurencyjny czy wręcz zatłoczony, zwłaszcza dla tych, którzy są zupełnie nowi w branży (np. absolwentów szkół wyższych), i dla osób przechodzących z innych branż i konkurujących z tysiącami takich samych kandydatów jak oni. Po drugie, mogą ubiegać się o pracę, która nie odzwierciedla w pełni obrazu danologii malowanego we wpisach na blogach i w popularnych artykułach prasowych, tj. nie będzie się wiązała tylko z pisananiem oraz implementowaniem algorytmów.

Biorąc pod uwagę te trendy, ważne jest to, by zrozumieć, że początkowo może być trudno wyróżnić się z grona innych kandydatów i dostać do finalnej rundy rozmowy kwalifikacyjnej. Choć strategie, o których przeczytasz w tej książce, mogą się wydawać bardzo pracochłonne, pomogą Ci się wyróżnić, co jest bardzo potrzebne w tym nowym, konkurencyjnym środowisku danologicznym.

1.4. Wywiad z Robertem Changiem, badaczem danych w Airbnb

Robert Chang jest danologiem w Airbnb, gdzie pracuje nad produktem Airbnb Plus. Wcześniej był zatrudniony w Twitterze, gdzie pracował w zespole Growth (ang. wzrost), zajmując się analizą produktów, tworzeniem potoków danych, prowadzeniem eksperymentów i tworzeniem modeli. Jego wpisy na blogu, dotyczące inżynierii danych, porad dla nowych i początkujących danologów oraz tego, jak wygląda praca w Airbnb i Twitterze, można znaleźć pod adresem: <https://medium.com/@rchang>.

Jak zaczęła się twoja podróż badacza danych?

Moją pierwszą pracą było stanowisko danologa w „The Washington Post”. W 2012 r. byłem gotów opuścić uczelnię i rozpocząć pracę w przemyśle, ale nie wiedziałem, co chcę robić. Miałem nadzieję, że zostanę badaczem zajmującym się wizualizacją danych, będąc pod wrażeniem pracy w „The New York Times”. Kiedy poszedłem na uczelniane targi kariery i zobaczyłem, że „The Washington Post” prowadzi rekrutację, w swojej naiwności po prostu założyłem, że musi robić podobne rzeczy jak „The New York Times”. Zaaplikowałem i dostałem tę robotę, a potem już nie sprawdziłem wystarczająco starannie swojego nowego pracodawcy.

Gdybyś zapytała o przykład, jak nie rozpoczynać kariery danologa, to na pewno sam bym się zgłosił na ochotnika! Zostałem przyjęty i miałem nadzieję na wizualizację lub modelowanie danych, ale bardzo szybko zdałem sobie sprawę, że moja praca będzie raczej zbliżona do tej, jaką wykonuje inżynier danych. Duża jej część polegała na budowaniu potoków ETL (ang. *extract transform load*), ponownym uruchamianiu skryptów SQL i próbach zapewnienia, że raporty działają, żebyśmy mogli dostarczać kadrze kierowniczej najważniejsze pomiary. Było to wtedy bardzo bolesne doświadczenie; zdałem sobie sprawę, że to, co chcę zrobić, nie jest zgodne z tym, czego firma naprawdę potrzebuje, i w końcu odszedłem z pracy.

Ale w następnych latach, które spędziłem w Twitterze i Airbnb, zdałem sobie sprawę z tego, że była to norma, a nie wyjątek. Kiedy tworzysz nowe możliwości w zakresie danych, musisz budować je warstwa po warstwie. Monica Rogati napisała słynny post na niezwykle popularnym blogu o hierarchii potrzeb danologicznych (<http://mng.bz/ad0o>). Ale w tym czasie byłem zbytnim nowicjuszem, żeby docenić to, jak prawdziwa, żywa praca z zakresu badania danych jest wykonywana.

Czego ludzie powinni szukać w pracy związanej z danologią?

Jeśli szukasz stanowiska z zakresu badania danych, powinieneś skupić się na stanie firmowej infrastruktury danych. Jeżeli bowiem w swojej firmie masz tylko garść surowych danych, które nie są przechowywane w hurtowni danych, to prawdopodobnie spędzisz całe miesiące, a być może nawet lata, żeby dojść do punktu, w którym będziesz mógł robić ciekawe analizy, eksperymentować lub uczyć maszynowo. Jeśli nie jest to coś, czego oczekujesz, to będzie istnieć zasadnicza rozbieżność między etapem, na którym znajduje się firma, a tym, jaki chcesz wnieść wkład w rozwój organizacji.

Aby to ocenić, można zadać pytania w rodzaju: „Czy macie zespół ds. infrastruktury danych? Od jak dawna istnieje? Jaki jest stos danych? Czy macie zespół ds. inżynierii danych? W jak sposób jego członkowie pracują z danologami? Czy kiedy tworzycie nowy produkt, macie proces instrumentalizacji logów, budowania tabel danych i umieszczania ich w swojej hurtowni danych?”. Jeśli tych rzeczy jeszcze nie ma, będziesz częścią zespołu, który jest odpowiedzialny za ich stworzenie, i możesz się spodziewać, że spędzisz na tym sporo czasu.

Druga rzecz, na którą należy zwrócić uwagę, to ludzie. Są trzy rodzaje ludzi, na których należy zwracać uwagę. Zakładając, że nie chcesz być pierwszym badaczem danych w firmie, powinieneś dołączyć do organizacji danologicznej, w której istnieje doświadczony

lider. Doświadczony lider wie, jak zbudować i utrzymać dobrą infrastrukturę i zapewnić przepływ pracy, żeby badacze danych byli produktywni. Po drugie, poszukaj menedżera, który będzie wspierał ciągłe uczenie się. Wreszcie jest bardzo ważne — zwłaszcza gdy jesteś nowy w pracy — żeby pracować z liderem technologicznym lub starszym badaczem danych, który ma praktyczne doświadczenie w tej robocie. W codziennej pracy to właśnie on będzie ci najbardziej pomagać.

Jakich umiejętności potrzeba, aby zostać badaczem danych?

Myślę, że to zależy od tego, jakiej pracy szukasz i jak wysoko pracodawca ustawią poprzeczkę. Najlepsze firmy na ogół mają wysoko ustawioną poprzeczkę, czasami wręcz absurdalnie wysoko, ponieważ wiele osób próbuje się w nich zatrudnić. Zazwyczaj szukają tzw. jednorożców — czyli osób mających umiejętności w zakresie obróbki danych przy użyciu R lub Pythona, a także osób z doświadczeniem w tworzeniu potoków ETL, w inżynierii danych, projektowaniu eksperymentów, tworzeniu modeli i wdrażaniu ich do środowiska produkcyjnego. Wywiera to dużą presję na kandydatów! Mimo że wszystkich tych umiejętności można się przecież nauczyć i mogą być one przydatne w rozwiązywaniu wszelkich problemów, z jakimi się borykasz, to jednak nie sądzę, żeby były one konieczne do rozpoczęcia przygody z danologią.

Jeśli znasz R lub Pythona i odrobinę SQL-a, jesteś już na całkiem niezłej pozycji wyjściowej do tego, żeby dostać się do domeny badania danych. Jeśli możesz zaplanować swoją karierę, ucząc się więcej „na zapas”, to jest to zawsze pomocne, ale nie sądzę, żeby stanowiło to wymóg. Ważniejsze jest zamiłowanie do uczenia się. Jeśli próbujesz zatrudnić się w najlepszych firmach technologicznych, potrzebujesz trochę większych umiejętności, ale to raczej po to, żeby przebić się przez sito selekcji, a nie dlatego, że naprawdę przydadzą ci się one w pracy. Pomocne będzie rozróżnienie między podstawowymi umiejętnościami, które są potrzebne do rozpoczęcia kariery w dziedzinie danologii, a innymi, które są mile widziane, jeśli chcesz się dostać do konkurencyjnej znanej firmy.

Podsumowanie

- Zestaw umiejętności w zakresie danologii różni się w zależności od osoby i stanowiska. Chociaż pewna wiedza jest fundamentalna, badacze danych nie muszą być ekspertami w każdej istotnej dziedzinie.
- Praca związana z danologią ma różne obszary, na których należy się skupić: przedstawianie interesariuszom właściwych, uporządkowanych danych (analiza), wdrażanie modeli uczenia maszynowego do środowiska produkcyjnego (uczenie maszynowe) oraz wykorzystywanie danych do podejmowania decyzji (nauka o podejmowaniu decyzji).

PROGRAM PARTNERSKI

— GRUPY HELION —

- 
1. ZAREJESTRUJ SIĘ
 2. PREZENTUJ KSIĄŻKI
 3. ZBIERAJ PROWIZJĘ

Zmień swoją stronę WWW w działający bankomat!

Dowiedz się więcej i dołącz już dzisiaj!

<http://program-partnerski.helion.pl>

GRUPA
Helion 

Nauka o danych, zwana danologią, zyskuje na znaczeniu. Dane dla gospodarki są tym, czym dotąd były węgiel, stal i ropa naftowa. Umiejętność korzystania z wiedzy zawartej w danych decyduje o efektywności prowadzenia działalności gospodarczej i determinuje rozwój nowych modeli, rozwiązań i relacji gospodarczych. Już teraz specjaliści danolodzy są rozchwytywani na rynku pracy. Aby jednak w pełni i do końca wykorzystać pojawiające się możliwości, trzeba wiedzieć, w jaki sposób podejść do trudnego zagadnienia, jakim jest budowanie ścieżki kariery i podążanie nią w odpowiednim dla siebie tempie.

To praktyczny przewodnik, dzięki któremu łatwiej zdobędziesz pierwszą pracę związaną z badaniem danych, szybciej staniesz się cenionym specjalistą i w miarę rozwoju zawodowego będziesz coraz trafniej wychwytywać pojawiające się możliwości awansu i zmiany pracy na atrakcyjniejszą. Dowiesz się, jak zdobyć podstawowe umiejętności i jak faktycznie wyglądają konkretne stanowiska pracy. Opisano tu również, jak pomyślnie przejść przez proces rekrutacji i zaaklimatyzować się w nowych warunkach. Nie zabrakło cennych wskazówek dotyczących awansowania na stanowiska kierownicze. Jako danolog prędko się przekonasz, że zawarta tutaj wiedza nietechniczna jest bardzo potrzebna do osiągnięcia sukcesu na polu badania danych.

Dzięki tej książce dowiesz się, jak:

- tworzyć świetne portfolio projektów z zakresu badania danych
- wyszukiwać, oceniać i negocjować oferty
- z klasą zmieniać miejsca pracy
- wybierać i skutecznie realizować scenariusze kariery
- poradzili sobie wybitni analitycy danych!

Danologia: nauka, pasja i sposób na życie!

Dr Jacqueline Nolis jest głównym analitykiem danych w Brightloom. Od lat zajmuje się doradzeniem największym firmom w zakresie korzystania z danych. Często występuje na prestiżowych konferencjach.

Emily Robinson jest starszym badaczem danych w Warby Parker, wcześniej pracowała w DataCamp oraz w Etsy. Regularnie prowadzi wykłady z zakresu nauki o danych.

Helion 



helion.pl



HELION SA
ul. Kościuszki 1c
44-100 Gliwice
tel.: 32 230 98 63
helion@helion.pl

Sprawdź nasze szkolenia!



AKADEMIA IT & BUSINESS

HELIONSZKOLENIA.PL

KOD KORZYŚCI
Sięgnij po więcej! ▶



ISBN 978-83-283-7293-1



9 788328 372931

INFORMATYKA W NAJLEPSZYM WYDANIU

Cena: 79,00 zł