

Zaawansowana analiza danych : jak przejść z arkuszy Excela do Pythona i R / George Mount. – Gliwice, © 2022

Spis treści

Wprowadzenie	9
Część I. Podstawy analizy danych w Excelu	17
1. Podstawy eksploracyjnej analizy danych	19
Czym jest eksploracyjna analiza danych?	19
Obserwacje	21
Zmienne	21
Przykład: klasyfikacja zmiennych	24
Przypomnienie: typy zmiennych	26
Eksploracja zmiennych w Excelu	26
Eksploracja zmiennych kategoryjnych	27
Eksploracja zmiennych ilościowych	29
Wnioski	40
Ćwiczenia	40
2. Podstawy prawdopodobieństwa	41
Prawdopodobieństwo i losowość	41
Prawdopodobieństwo i przestrzeń zdarzeń elementarnych	41
Prawdopodobieństwo i eksperymenty	42
Prawdopodobieństwo bezwarunkowe i warunkowe	42
Rozkłady prawdopodobieństwa	42
Dyskretne rozkłady prawdopodobieństwa	43
Ciągłe rozkłady prawdopodobieństwa	46
Wnioski	53
Ćwiczenia	53
3. Podstawy wnioskowania statystycznego	54
Ramy wnioskowania statystycznego	54
Zbierz reprezentatywną próbkę	55
Sformułuj hipotezy	56
Stwórz plan analizy	57
Przeanalizuj dane	59
Podejmij decyzję	62
To Twój świat... Dane się tylko w nim znajdują	68
Wnioski	69
Ćwiczenia	70
4. Korelacja i regresja	71
Korelacja nie oznacza przyczynowości	71
Wprowadzenie do korelacji	72

Od korelacji do regresji	76
Regresja liniowa w Excelu	78
Zastanówmy się raz jeszcze — pozorne związki	84
Wnioski	85
Przejście do programowania	85
Ćwiczenia	85
5. Stos analizy danych	87
Statystyka, analiza danych, nauka o danych	87
Statystyka	87
Analiza danych	87
Analityka biznesowa	88
Nauka o danych	88
Uczenie maszynowe	88
Różne, ale nie rozłączne	89
Znaczenie stosu analizy danych	89
Arkusze kalkulacyjne	90
Bazy danych	92
Platformy analityki biznesowej	94
Języki programowania danych	94
Wnioski	95
Co dalej	96
Ćwiczenia	96
Część II. Z Excela do R	97
6. Pierwsze kroki w R dla użytkowników Excela	99
Pobieranie R	99
Pierwsze kroki w RStudio	99
Pakiety w R	108
Aktualizacja pakietów, RStudio i języka R	109
Wnioski	110
Ćwiczenia	110
7. Struktury danych w R	112
Wektory	112
Indeksowanie i wybór elementów z wektorów	114
Od tabel Excela do ramek danych R	115
Importowanie danych w R	117
Eksploracja ramki danych	120
Indeksowanie i wybór elementów z ramek danych	122
Zapisywanie ramek danych	123
Wnioski	124
Ćwiczenia	124
8. Przetwarzanie i wizualizacja danych w R	125
Przetwarzanie danych za pomocą dplyr	126
Operacje kolumnowe	126

Operacje wierszowe	128
Agregacja i łączenie danych	131
dplyr i potęga operatora potoku (%>%)	133
Przekształcanie danych za pomocą tidyr	135
Wizualizacja danych w ggplot2	137
Wnioski	142
Ćwiczenia	142
9. R w analizie danych	143
Eksploracyjna analiza danych	144
Testowanie hipotez	147
Test t-Studenta dla prób niezależnych	148
Regresja liniowa	150
Podział na zbiór uczący i testowy, walidacja	151
Wnioski	154
Ćwiczenia	154
Część III. Z Excela do Pythona	155
10. Pierwsze kroki w Pythonie dla użytkowników Excela	157
Pobieranie Pythona	157
Pierwsze kroki z Jupyterem	158
Moduły w Pythonie	166
Aktualizacja pakietów, Anacondy i Pythona	167
Wnioski	167
Ćwiczenia	168
11. Struktury danych w Pythonie	169
Tablice NumPy	170
Indeksowanie i wybieranie elementów z tablic NumPy	171
Ramki danych pandas	172
Importowanie danych w Pythonie	174
Eksploracja ramki danych	175
Indeksowanie i pobieranie wartości z ramek danych	177
Zapis ramek danych	178
Wnioski	178
Ćwiczenia	178
12. Przetwarzanie i wizualizacja danych w Pythonie	179
Operacje kolumnowe	180
Operacje wierszowe	182
Agregacja i łączenie danych	183
Przekształcanie danych	185
Wizualizacja danych	186
Wnioski	192
Ćwiczenia	192

13. Python w analizie danych	193
Eksploracyjna analiza danych	194
Testowanie hipotez	196
Test t-Studenta dla prób niezależnych	196
Regresja liniowa	197
Podział zbioru na zbiór treningowy i testowy oraz walidacja modelu	198
Wnioski	200
Ćwiczenia	200
14. Wnioski i kolejne kroki	201
Kolejne warstwy stosu	201
Projektowanie badań i eksperymenty biznesowe	201
Inne metody statystyczne	202
Nauka o danych i uczenie maszynowe	202
Kontrola wersji	202
Etyka	203
Idź naprzód i ciesz się danymi	203
Na pożegnanie	203
Skorowidz	204

oprac. BPK