

# Spis treści

|                       |           |
|-----------------------|-----------|
| <b>Przedmowa</b>      | <b>13</b> |
| <b>O autorze</b>      | <b>15</b> |
| <b>O recenzentach</b> | <b>16</b> |
| <b>Wstęp</b>          | <b>17</b> |

## **Część I. Najlepsze praktyki korzystania z języków R i Python w usłudze Power BI**

---

|                                                                                      |           |
|--------------------------------------------------------------------------------------|-----------|
| <b>Rozdział 1. Gdzie i jak używać w usłudze Power BI skryptów języka R i Python?</b> | <b>23</b> |
| Wymagania techniczne                                                                 | 24        |
| Wstrzykiwanie skryptów języka R lub Python do usługi Power BI                        | 24        |
| Ładowanie danych                                                                     | 24        |
| Przekształcanie danych                                                               | 27        |
| Wizualizacja danych                                                                  | 31        |
| Korzystanie z języków R i Python do interakcji z danymi                              | 32        |
| Ograniczenia stosowania języków R i Python dla różnych produktów usługi Power BI     | 34        |
| Podsumowanie                                                                         | 36        |
| <b>Rozdział 2. Konfigurowanie języka R na potrzeby usługi Power BI</b>               | <b>37</b> |
| Wymagania techniczne                                                                 | 38        |
| Dostępne silniki języka R                                                            | 38        |
| Dystrybucja CRAN R                                                                   | 38        |
| Dystrybucja Microsoft R Open i MRAN                                                  | 38        |

|                                                                             |           |
|-----------------------------------------------------------------------------|-----------|
| Klient Microsoft R                                                          | 42        |
| Stopniowe wycofywanie dystrybucji Microsoft R Open                          | 42        |
| <b>Wybór silnika języka R do zainstalowania</b>                             | <b>43</b> |
| Silniki języka R używane w usłudze Power BI                                 | 43        |
| Instalowanie sugerowanych silników języka R                                 | 45        |
| <b>Instalowanie środowiska IDE w celu programowania w języku R</b>          | <b>50</b> |
| Instalacja programu RStudio                                                 | 50        |
| <b>Konfigurowanie programu Power BI Desktop do pracy z językiem R</b>       | <b>54</b> |
| <b>Konfigurowanie usługi Power BI do pracy z językiem R</b>                 | <b>56</b> |
| Instalowanie lokalnej bramy danych w trybie osobistym                       | 56        |
| Udostępnianie raportów używających skryptów języka R w usłudze Power BI     | 60        |
| <b>Ograniczenia wizualizacji języka R</b>                                   | <b>61</b> |
| <b>Podsumowanie</b>                                                         | <b>63</b> |
| <b>Rozdział 3. Konfigurowanie języka Python na potrzeby usługi Power BI</b> | <b>64</b> |

|                                                                              |           |
|------------------------------------------------------------------------------|-----------|
| <b>Wymagania techniczne</b>                                                  | <b>64</b> |
| <b>Dostępne silniki Pythona</b>                                              | <b>65</b> |
| <b>Wybór silnika języka Python do zainstalowania</b>                         | <b>66</b> |
| Silniki języka Python wykorzystywane w usłudze Power BI                      | 66        |
| Instalowanie sugerowanych silników języka Python                             | 67        |
| <b>Instalowanie środowiska IDE na potrzeby programowania w języku Python</b> | <b>79</b> |
| Konfigurowanie obsługi Pythona w środowisku programu RStudio                 | 80        |
| Konfigurowanie języka Python w środowisku programu Visual Studio Code        | 82        |
| <b>Konfigurowanie programu Power BI Desktop do pracy z językiem Python</b>   | <b>86</b> |
| <b>Konfigurowanie usługi Power BI do pracy z językiem R</b>                  | <b>88</b> |
| Udostępnianie raportów używających skryptów języka Python w usłudze Power BI | 89        |
| <b>Ograniczenia wizualizacji w Pythonie</b>                                  | <b>89</b> |
| <b>Podsumowanie</b>                                                          | <b>91</b> |

## **Część II. Pozyskiwanie i przekształcanie danych za pomocą języków R i Python w usłudze Power BI**

|                                                                  |            |
|------------------------------------------------------------------|------------|
| <b>Rozdział 4. Importowanie nieobsługiwanych obiektów danych</b> | <b>95</b>  |
| <b>Wymagania techniczne</b>                                      | <b>96</b>  |
| <b>Importowanie plików RDS w języku R</b>                        | <b>96</b>  |
| Krótkie wprowadzenie do Tidyverse                                | 96         |
| Tworzenie zserializowanych obiektów w języku R                   | 97         |
| Korzystanie z plików RDS w usłudze Power BI                      | 103        |
| <b>Importowanie plików PKL w Pythonie</b>                        | <b>113</b> |
| Krótkie wprowadzenie do PyData                                   | 114        |
| Tworzenie zserializowanego obiektu Pythona                       | 115        |
| Korzystanie z pliku PKL w usłudze Power BI                       | 119        |
| <b>Podsumowanie</b>                                              | <b>126</b> |
| <b>Bibliografia</b>                                              | <b>127</b> |



|                                                                                      |            |
|--------------------------------------------------------------------------------------|------------|
| <b>Rozdział 5. Korzystanie z wyrażeń regularnych w usłudze Power BI</b>              | <b>128</b> |
| Wymagania techniczne                                                                 | 128        |
| Krótkie wprowadzenie do regeksów                                                     | 129        |
| Podstawowe informacje o regeksach                                                    | 130        |
| Sprawdzanie poprawności adresów e-mail                                               | 136        |
| Sprawdzanie poprawności dat                                                          | 139        |
| <b>Sprawdzanie poprawności danych przy użyciu regeksów w usłudze Power BI</b>        | <b>142</b> |
| Korzystanie z regeksów w usłudze Power BI                                            |            |
| do sprawdzania poprawności adresów e-mail z wykorzystaniem języka Python             | 143        |
| Korzystanie z regeksów w usłudze Power BI                                            |            |
| do sprawdzania poprawności adresów e-mail za pomocą języka R                         | 146        |
| Korzystanie z regeksów w usłudze Power BI                                            |            |
| do sprawdzania poprawności dat za pomocą języka Python                               | 149        |
| Korzystanie z regeksów w usłudze Power BI                                            |            |
| do sprawdzania poprawności dat za pomocą języka R                                    | 151        |
| <b>Ładowanie do usługi Power BI złożonych plików logów z wykorzystaniem regeksów</b> | <b>152</b> |
| Logi dostępu do serwera Apache                                                       | 153        |
| Importowanie logów dostępu serwera Apache w usłudze Power BI                         |            |
| za pomocą języka Python                                                              | 154        |
| Importowanie logów dostępu serwera Apache w usłudze Power BI                         |            |
| za pomocą języka R                                                                   | 156        |
| <b>Wyodrębnianie wartości z tekstu w usłudze Power BI przy użyciu regeksów</b>       | <b>157</b> |
| Jeden regex do zarządzania wszystkimi danymi                                         | 159        |
| Używanie regeksów w usłudze Power BI do wyodrębniania wartości                       |            |
| z wykorzystaniem języka Python                                                       | 160        |
| Korzystanie z regeksów w usłudze Power BI do wyodrębniania wartości                  |            |
| za pomocą języka R                                                                   | 162        |
| Podsumowanie                                                                         | 164        |
| Bibliografia                                                                         | 164        |
| <b>Rozdział 6. Anonimizacja i pseudonimizacja danych w usłudze Power BI</b>          | <b>165</b> |
| Wymagania techniczne                                                                 | 166        |
| Deidentyfikacja danych                                                               | 166        |
| Techniki deidentyfikacji                                                             | 166        |
| Istota pseudonimizacji                                                               | 170        |
| Czym jest anonimizacja?                                                              | 171        |
| <b>Anonimizacja danych w usłudze Power BI</b>                                        | <b>172</b> |
| Anonimizacja danych za pomocą języka Python                                          | 173        |
| Anonimizacja danych za pomocą języka R                                               | 176        |
| <b>Pseudonimizacja danych w usłudze Power BI</b>                                     | <b>180</b> |
| Pseudonimizacja danych za pomocą języka Python                                       | 181        |
| Pseudonimizacja danych w języku R                                                    | 184        |
| Podsumowanie                                                                         | 186        |
| Bibliografia                                                                         | 186        |



|                                                                                                     |            |
|-----------------------------------------------------------------------------------------------------|------------|
| <b>Rozdział 7. Zapisywanie danych z usługi Power BI do źródeł zewnętrznych</b>                      | <b>187</b> |
| Wymagania techniczne                                                                                | 187        |
| Zapis danych do plików CSV                                                                          | 188        |
| Zapis do plików CSV w języku Python                                                                 | 189        |
| Zapis informacji do plików CSV za pomocą języka R                                                   | 192        |
| Zapisywanie informacji do plików programu Excel                                                     | 195        |
| Zapisywanie informacji do plików Excela za pomocą języka Python                                     | 196        |
| Zapis do plików Excela za pomocą języka R                                                           | 200        |
| Zapis danych do serwera Azure SQL Server                                                            | 204        |
| Instalacja programu SQL Server Express                                                              | 205        |
| Tworzenie bazy danych w usłudze Azure SQL Database                                                  | 207        |
| Zapis danych do serwera Azure SQL Server za pomocą języka Python                                    | 212        |
| Zapis danych na serwerze Azure SQL za pomocą języka R                                               | 219        |
| Podsumowanie                                                                                        | 223        |
| Bibliografia                                                                                        | 224        |
| <b>Rozdział 8. Ładowanie do usługi Power BI zbiorów danych przekraczających dostępną pamięć RAM</b> | <b>225</b> |
| Wymagania techniczne                                                                                | 226        |
| Typowy scenariusz analityczny obejmujący wykorzystanie obszernego zestawu danych                    | 226        |
| Importowanie dużych zestawów danych w języku Python                                                 | 227        |
| Instalowanie Dask na laptopie                                                                       | 228        |
| Tworzenie obiektów DataFrame pakietu Dask                                                           | 229        |
| Wyodrębnianie informacji z obiektu DataFrame pakietu Dask                                           | 230        |
| Importowanie dużego zestawu danych w usłudze Power BI z wykorzystaniem języka Python                | 235        |
| Importowanie dużych zestawów danych za pomocą języka R                                              | 237        |
| Instalowanie pakietu disk.frame na laptopie                                                         | 238        |
| Tworzenie egzemplarza obiektu disk.frame                                                            | 238        |
| Wyodrębnianie informacji z obiektu disk.frame                                                       | 240        |
| Importowanie dużego zestawu danych w usłudze Power BI za pomocą języka R                            | 242        |
| Podsumowanie                                                                                        | 244        |
| Bibliografia                                                                                        | 244        |
| <b>Część III. Wzbogacanie danych w usłudze Power BI za pomocą języków R i Python</b>                | <b>245</b> |
| <b>Rozdział 9. Wywoływanie zewnętrznych interfejsów API w celu wzbogacania danych</b>               | <b>247</b> |
| Wymagania techniczne                                                                                | 248        |
| Czym jest usługa sieciowa?                                                                          | 248        |
| Rejestrowanie się w usłudze Bing Maps Web Services                                                  | 249        |
| Geokodowanie adresów z wykorzystaniem języka Python                                                 | 251        |
| Korzystanie z jawnego żądania GET                                                                   | 251        |
| Korzystanie ze współbieżnych żądań GET                                                              | 255        |
| Korzystanie z biblioteki Geocoder w trybie programowania współbieżnego                              | 257        |



|                                                                                              |            |
|----------------------------------------------------------------------------------------------|------------|
| <b>Geokodowanie adresów za pomocą języka R</b>                                               | <b>258</b> |
| Korzystanie z jawnych żądań GET                                                              | 259        |
| Korzystanie ze współbieżnych żądań GET                                                       | 261        |
| Korzystanie z pakietu tidygeocoder do współbieżnego wykonywania kodu                         | 262        |
| <b>Korzystanie z usług sieciowych z poziomu usługi Power BI</b>                              | <b>263</b> |
| Geokodowanie adresów w usłudze Power BI za pomocą języka Python                              | 265        |
| Geokodowanie adresów w usłudze Power BI za pomocą języka R                                   | 265        |
| <b>Podsumowanie</b>                                                                          | <b>266</b> |
| <b>Bibliografia</b>                                                                          | <b>267</b> |
| <b>Rozdział 10. Obliczanie kolumn przy użyciu złożonych algorytmów</b>                       | <b>268</b> |
| <b>Wymagania techniczne</b>                                                                  | <b>269</b> |
| <b>Obliczanie odległości między dwiema lokalizacjami geograficznymi</b>                      | <b>269</b> |
| Trygonometria sferyczna                                                                      | 269        |
| Twierdzenie cosinusów                                                                        | 271        |
| Wzór Haversinesa                                                                             | 272        |
| Wzór Vincenty'ego                                                                            | 273        |
| Jakiego wzoru odległości używać i kiedy?                                                     | 274        |
| <b>Implementacja algorytmów wyznaczania odległości w Pythonie</b>                            | <b>275</b> |
| Obliczanie odległości w języku Python                                                        | 275        |
| Obliczanie odległości w usłudze Power BI za pomocą języka Python                             | 278        |
| <b>Implementacja algorytmów wyznaczania odległości w języku R</b>                            | <b>279</b> |
| Obliczanie odległości za pomocą języka R                                                     | 279        |
| Obliczanie odległości w usłudze Power BI za pomocą języka R                                  | 282        |
| <b>Podstawy programowania liniowego</b>                                                      | <b>283</b> |
| Równania i nierówności liniowe                                                               | 283        |
| Formułowanie problemu optymalizacji liniowej                                                 | 285        |
| <b>Definicja problemu LP do rozwiązania</b>                                                  | <b>288</b> |
| Formułowanie problemu LP                                                                     | 289        |
| <b>Rozwiązywanie problemów optymalizacji w języku Python</b>                                 | <b>291</b> |
| Rozwiązywanie problemu LP za pomocą języka Python                                            | 291        |
| Rozwiązywanie problemu programowania liniowego<br>za pomocą języka Python w usłudze Power BI | 295        |
| <b>Rozwiązywanie problemów programowania liniowego za pomocą języka R</b>                    | <b>300</b> |
| Rozwiązywanie problemu LP w języku R                                                         | 300        |
| Rozwiązywanie problemu programowania liniowego<br>za pomocą języka R w usłudze Power BI      | 303        |
| <b>Podsumowanie</b>                                                                          | <b>305</b> |
| <b>Bibliografia</b>                                                                          | <b>306</b> |
| <b>Rozdział 11. Dodawanie statystyk: powiązania</b>                                          | <b>307</b> |
| <b>Wymagania techniczne</b>                                                                  | <b>307</b> |
| <b>Badanie powiązań między zmiennymi</b>                                                     | <b>308</b> |
| <b>Korelacja między zmiennymi liczbowymi</b>                                                 | <b>308</b> |
| Współczynnik korelacji Karla Pearsona                                                        | 310        |
| Współczynnik korelacji Charlesa Spearmana                                                    | 313        |
| Współczynnik korelacji Maurice'a Kendalla                                                    | 316        |
| Opis przypadku                                                                               | 316        |



|                                                                                                                |            |
|----------------------------------------------------------------------------------------------------------------|------------|
| Implementacja obliczeń współczynników korelacji w Pythonie                                                     | 317        |
| Implementacja obliczeń współczynników korelacji w języku R                                                     | 320        |
| Implementacja obliczania współczynników korelacji w usłudze Power BI<br>za pomocą języków Python i R           | 323        |
| <b>Korelacje między zmiennymi kategorycznymi a liczbowymi</b>                                                  | <b>325</b> |
| Związek między dwiema zmiennymi kategorycznymi                                                                 | 325        |
| Związki pomiędzy zmiennymi liczbowymi a kategorycznymi                                                         | 329        |
| Implementacja obliczania współczynników korelacji w Pythonie                                                   | 332        |
| Implementacja obliczania współczynników korelacji w języku R                                                   | 334        |
| Implementacja obliczania współczynników korelacji w usłudze Power BI<br>za pomocą języków R i Python           | 337        |
| <b>Podsumowanie</b>                                                                                            | <b>338</b> |
| <b>Bibliografia</b>                                                                                            | <b>339</b> |
| <b>Rozdział 12. Dodawanie statystyk: wartości odstające i wartości brakujące</b>                               | <b>340</b> |
| <b>Wymagania techniczne</b>                                                                                    | <b>340</b> |
| <b>Czym są wartości odstające i jak sobie z nimi radzić?</b>                                                   | <b>341</b> |
| Przyczyny istnienia wartości odstających                                                                       | 342        |
| Obsługa wartości odstających                                                                                   | 342        |
| <b>Identyfikacja wartości odstających</b>                                                                      | <b>343</b> |
| Jednowymiarowe wartości odstające                                                                              | 343        |
| Wielowymiarowe wartości odstające                                                                              | 344        |
| <b>Implementacja algorytmów wykrywania wartości odstających</b>                                                | <b>348</b> |
| Implementacja wykrywania wartości odstających w Pythonie                                                       | 349        |
| Implementacja wykrywania wartości odstających w języku R                                                       | 355        |
| Implementacja wykrywania wartości odstających w usłudze Power BI                                               | 358        |
| <b>Czym są wartości brakujące i jak sobie z nimi radzić?</b>                                                   | <b>359</b> |
| Przyczyny brakujących wartości                                                                                 | 360        |
| Obsługa wartości brakujących                                                                                   | 361        |
| <b>Diagnozowanie brakujących wartości z wykorzystaniem języków R i Python</b>                                  | <b>366</b> |
| <b>Implementacja algorytmów imputacji brakujących wartości</b>                                                 | <b>369</b> |
| Usuwanie brakujących wartości                                                                                  | 370        |
| Imputacja danych tabelarycznych                                                                                | 370        |
| Imputacja danych szeregów czasowych                                                                            | 373        |
| Imputacja brakujących wartości w usłudze Power BI                                                              | 374        |
| <b>Podsumowanie</b>                                                                                            | <b>376</b> |
| <b>Bibliografia</b>                                                                                            | <b>376</b> |
| <b>Rozdział 13. Korzystanie z uczenia maszynowego<br/>bez licencji Premium lub Embedded</b>                    | <b>378</b> |
| <b>Wymagania techniczne</b>                                                                                    | <b>379</b> |
| <b>Interakcje z mechanizmami uczenia maszynowego w usłudze Power BI<br/>z wykorzystaniem przepływów danych</b> | <b>379</b> |
| <b>Korzystanie z rozwiązań AutoML</b>                                                                          | <b>381</b> |
| PyCaret                                                                                                        | 382        |
| Azure AutoML                                                                                                   | 383        |
| RemixAutoML dla języka R                                                                                       | 383        |

|                                                                                             |            |
|---------------------------------------------------------------------------------------------|------------|
| <b>Osadzanie kodu szkoleniowego w dodatku Power Query</b>                                   | <b>384</b> |
| Szkolenie i korzystanie z modeli ML z wykorzystaniem biblioteki PyCaret                     | 384        |
| Korzystanie z biblioteki PyCaret w usłudze Power BI                                         | 387        |
| <b>Korzystanie z przeszkolonych modeli w dodatku Power Query</b>                            | <b>388</b> |
| Ocena obserwacji w dodatku Power Query przy użyciu przeszkolonego modelu PyCaret            | 389        |
| <b>Korzystanie z przeszkolonych modeli w wizualizacjach utworzonych za pomocą skryptów</b>  | <b>390</b> |
| Ocenianie obserwacji w skrypcie wizualizacji z wykorzystaniem przeszkolonego modelu PyCaret | 392        |
| <b>Wywoływanie usług sieciowych w dodatku Power Query</b>                                   | <b>401</b> |
| Korzystanie z usług Cognitive Services w dodatku Power Query                                | 408        |
| <b>Podsumowanie</b>                                                                         | <b>411</b> |
| <b>Bibliografia</b>                                                                         | <b>411</b> |

## **Część IV. Wizualizacja danych za pomocą języka R w usłudze Power BI**

---

### **Rozdział 14. Eksploracyjna analiza danych**

---

|                                                        |            |
|--------------------------------------------------------|------------|
| <b>Wymagania techniczne</b>                            | <b>415</b> |
| <b>Jaki jest cel EDA?</b>                              | <b>416</b> |
| Zrozumienie sensu danych                               | 416        |
| Oczyszczanie danych                                    | 417        |
| Odkrywanie powiązań między zmiennymi                   | 418        |
| <b>Analiza EDA z wykorzystaniem języków R i Python</b> | <b>418</b> |
| <b>Analiza EDA w usłudze Power BI</b>                  | <b>420</b> |
| Strona podsumowania zestawu danych                     | 421        |
| Eksploracja brakujących wartości                       | 424        |
| Eksploracja jednowymiarowa                             | 425        |
| Eksploracja wielowymiarowa                             | 433        |
| Powiązania zmiennych                                   | 438        |
| <b>Podsumowanie</b>                                    | <b>442</b> |
| <b>Bibliografia</b>                                    | <b>442</b> |

### **Rozdział 15. Zaawansowane wizualizacje**

---

|                                                                     |            |
|---------------------------------------------------------------------|------------|
| <b>Wymagania techniczne</b>                                         | <b>443</b> |
| <b>Tworzenie wykresu kołowego</b>                                   | <b>443</b> |
| <b>Implementacja kołowego wykresu słupkowego w języku R</b>         | <b>446</b> |
| <b>Implementacja kołowego wykresu słupkowego w usłudze Power BI</b> | <b>452</b> |
| <b>Podsumowanie</b>                                                 | <b>453</b> |
| <b>Bibliografia</b>                                                 | <b>454</b> |



|                                                                           |            |
|---------------------------------------------------------------------------|------------|
| <b>Rozdział 16. Interaktywne niestandardowe wizualizacje języka R</b>     | <b>455</b> |
| Wymagania techniczne                                                      | 455        |
| Zalety stosowania interaktywnych niestandardowych wizualizacji w języku R | 456        |
| Dodawanie elementów interaktywnych za pomocą biblioteki Plotly            | 458        |
| Wykorzystanie elementów interaktywnych za pośrednictwem widżetów HTML     | 460        |
| Tworzenie niestandardowych wizualizacji w usłudze Power BI                | 460        |
| Instalowanie pakietu pbiviz                                               | 461        |
| Tworzenie pierwszej niestandardowej wizualizacji R HTML                   | 462        |
| Importowanie pakietu niestandardowej wizualizacji do usługi Power BI      | 471        |
| Podsumowanie                                                              | 474        |
| Bibliografia                                                              | 475        |
| <b>Skorowidz</b>                                                          | <b>476</b> |