Efekt 3. Wstępna analiza danych

Damian Okoń, Mateusz Nowak, Robert Zamiar

# Opis zbioru danych, statystyki opisowe

W efekcie przekształceń danych otrzymano następujące zmienne:

Tabela 1. Zmienne z opisami

|  |  |  |
| --- | --- | --- |
| zmienna | wyjaśnienie | typ zmiennej |
| country | kraj | etykieta |
| GDP\_pc | Produkt krajowy brutto per capita w euro | zmienna numeryczna |
| high\_tech\_trade\_pc | Handel wysokimi technologiami per capita w euro | zmienna numeryczna |
| r&d\_gdp\_pct | Wydatki na badania i rozwój jako procent PKB | zmienna numeryczna |
| r&d\_bud\_pct | Wydatki na badania i rozwój jako procent budżetu | zmienna numeryczna |
| use\_cloud\_pct | Procent ludzi korzystających z chmury | zmienna numeryczna |
| weeknd\_work\_pct | Procent ludzi pracujących w weekendy (15-64) | zmienna numeryczna |
| emp\_deadline\_pct | Procent ludzi pracujących intensywnie i z deadlinem (15-64) | zmienna numeryczna |
| working\_pop\_pct | Procent populacji w wieku produkcyjnym (15-64) | zmienna numeryczna |
| phycisians\_per\_1000 | lekarze na 1000 mieszkańców | zmienna numeryczna |
| sea\_access | dostęp do morza | zmienna kategoryczna (binarna) |
| joined\_EU | dekada dołączenia do Unii Europejskiej | zmienna kategoryczna |
| is\_euro\_currency | Czy walutą jest euro? | zmienna kategoryczna (binarna) |
| nuclear\_electricity | Czy kraj posiada elektorwnie atomowe? | zmienna kategoryczna (binarna) |

Dane numeryczne jak i etykietę z nazwami krajów(nazwy w języku angielskim) jak i zmienne numeryczne zaczerpnięto z serwisu Eurostatu, natomiast zmienne kategoryczne zostały utworzone przez zespół na podstawie źródeł z serwisów geopolitycznych. Zmienne te mogą być wykazane w przyszłych, bardziej wnikliwych analizach – analizach głównych składowych czy analizach skupień.

## Obliczanie zmienności

Przygotowane przez nas dane są przekrojowe – pochodzą z roku 2015. więc nie są szeregiem czasowym. Z powodu przyszłego wykorzystania danych do analizy regresji obliczono współczynnik zmienności będący ilorazem odchylenia standardowego i średniej danych zmiennych – wskaźnik ten obliczono zarówno dla zmiennych numerycznych jak i binarnych. Tabela z wynikami na następnej stronie:

|  |  |  |  |
| --- | --- | --- | --- |
| Nazwa zmiennej | średnia | Odchylenie standardowe | Współczynnik zmienności |
| r&d\_pct of budget | 1,18 | 0,43 | 36,21 |
| pct of individuals who use cloud | 24,96 | 9,01 | 36,09 |
| pct of people who work on weekends | 24,43 | 7,98 | 32,68 |
| pct of people who work with deadlines | 43,22 | 8,68 | 20,07 |
| working population in thousands | 13111,58 | 15384,82 | 117,34 |
| phycisians\_per\_1000 | 10,41 | 10,61 | 101,94 |
| sea\_access | 0,84 | 0,37 | 44,54 |
| is\_euro\_currency | 0,64 | 0,49 | 76,55 |
| nuclear\_electricity | 0,56 | 0,51 | 90,47 |