

Eksploracyjna analiza danych

Światowy program szczepień przeciwko COVID-19

Marek Grudkowski 156587
Kamil Kaczmarkiewicz 171701

24 kwietnia 2021

1 Ogólny opis danych

Zbiór danych dotyczy aktualnego postępu poszczególnych krajów w szczepieniach przeciwko COVID-19. W dobie pandemii jest to niezwykle gorący temat. Naszym zdaniem warto się na nim skupić, gdyż może zawierać wiele ukrytych informacji, które mogą przydać się w walce z pandemią i przyspieszyć sam proces szczepień.

2 Cel eksploracji i kryteria sukcesu

3 Charakterystyka zbioru danych

pochodzenie, format, liczba przykładów, ilość zbiorów nazwy typy - nominalne/numeryczne co oznaczają co oznaczają wartości - jednostka miary, wartości specjalne

Zbiór danych pochodzi z wielu różnych źródeł, którymi zazwyczaj są organy krajowe oraz lokalne i międzynarodowe organizacje. Dla każdego przykładu podane jest źródło i jego adres internetowy, co daje możliwość weryfikacji w przypadku jakichkolwiek wątpliwości co do poprawności danych. Dane zapisane są w formacie csv i podzielone są na następujące kolumny:

- **country** - kraj, dla którego podawane są informacje o szczepieniu, atrybut nominalny w postaci ciągu znaków
- **iso_code** - kod ISO dla danego kraju, atrybut nominalny w postaci ciągu znaków
- **date** - data wprowadzenia danych, atrybut nominalny opisujący datę
- **total_vaccinations** - bezwzględna liczba wszystkich szczepień ochronnych w danym kraju, atrybut numeryczny (liczba naturalna)
- **people_vaccinated** - liczba osób która otrzymała szczepionkę (przy dwóch dawkach liczona jest $\times 2$), atrybut numeryczny (liczba naturalna)

- **people_fully_vaccinated** - liczba osób, które otrzymały cały zestaw szczepień, atrybut numeryczny (liczba naturalna)
- **daily_vaccinations_raw** - dla danej pozycji liczba szczepień dla tej daty/kraju, atrybut numeryczny (liczba naturalna)
- **daily_vaccinations** - dla danej pozycji liczba szczepień dla tej daty/kraju, atrybut numeryczny (liczba naturalna)
- **total_vaccinations_per_hundred** - stosunek liczby szczepień do całkowitej liczby ludności danego dnia w kraju, atrybut numeryczny wyrażany w procentach
- **people_vaccinated_per_hundred** - stosunek liczby osób zaszczepionych do całkowitej liczby ludności danego dnia w kraju, atrybut numeryczny wyrażany w procentach
- **people_fully_vaccinated_per_hundred** - stosunek liczby osób uodpornionych do całkowitej liczby ludności danego dnia w kraju, atrybut numeryczny wyrażany w procentach
- **daily_vaccinations_per_million** - stosunek między liczbą szczepień a całkowitą liczbą ludności na bieżący dzień w kraju, dodania liczba rzeczywista
- **vaccines** - rodzaje szczepionek wykorzystanych w danym kraju, atrybut nominalny, ciągi znaków rozdzielone ukośnikiem
- **source_name** - źródło informacji, atrybut nominalny, ciąg znaków
- **source_website** - strona internetowa źródła informacji, atrybut nominalny, ciąg znaków

4 Wyniki eksploracyjnej analizy danych

rozkłady wartości atrybutów, korelacje pomiędzy wartościami atrybutów wstępne ustalenia dotyczące zawartości zbioru

5 Uwagi dotyczące jakości danych

dane brakujące, punkty oddalone, dane niespójne, dane niezrozumiałe,

6 Opis wyników eksploracji

w odniesieniu do celów eksploracji, czy dane są wystarczające, ewentualna rewizja celów