

COVID-19 – zagregowane dane światowe

Treść zadania

Dla każdego dnia marca 2021 roku należy wyznaczyć 1% (pierwszy percentyl) krajów o największym stosunku dziennego przyrostu osób wyzdrowiających (**Recovered**) do dziennego przyrostu liczby potwierdzonych zakażeń (**Confirmed**).

Źródło danych

Dane pochodzą z opublikowanego 6 kwietnia 2021 roku zbioru [time-series-19-covid-combined.csv](#), udostępnionego na platformie [github](#) w ramach katalogu [COVID-19 dataset](#).

Schemat danych

Nazwa pola	Typ pola	Komentarz
Date	znacznik czasu	dzień, miesiąc i rok obserwacji
Country/Region	ciąg tekstowy	kraj obserwacji
Province/State	opcjonalny ciąg tekstowy	wyróżniona dla Country/Region podlokalizacja obserwacji
Confirmed	liczba całkowita	<u>sumaryczna liczba</u> potwierdzonych zakażeń dla kraju (lub podlokalizacji) od początku prowadzenia obserwacji
Recovered	liczba całkowita	<u>sumaryczna liczba</u> osób wyzdrowiających dla kraju (lub podlokalizacji) od początku prowadzenia obserwacji
Deaths	liczba całkowita	<u>sumaryczna liczba</u> zgonów dla kraju (lub podlokalizacji) od początku prowadzenia obserwacji

Dostęp w środowisku laboratoryjnym

Plik płaski csv dostępny jest w rozproszonym systemie plików HDFS pod ogólnodostępną ścieżką `/input/assignments/time-series-19-covid/time-series-19-covid-combined.csv`

Wymagania

Stosunek dziennego przyrostu osób wyzdrowiających (**Recovered**) do dziennego przyrostu liczby potwierdzonych zakażeń (**Confirmed**) definiowany jest jako:

$$\frac{\text{dzienny przyrost } \mathbf{Recovered}}{\text{dzienny przyrost } \mathbf{Confirmed}}$$

dla danego kraju, a dzienny przyrost wartości X w dniu n jest wyznaczony jako różnica:

$$X_n - X_{n-1}$$

X_n – wartość X w dniu n ,

X_{n-1} – wartość X w dniu $n - 1$ (dzień poprzedzający n).

Ograniczenia na zakres przetwarzanych danych

1. Zakres przetwarzania danych (**Date**) to wyłącznie marzec 2021 roku.

Aspekty techniczne

1. Dienne wartości osób wyzdrowiających jak i potwierdzonych zakażeń należy rozpatrywać na poziomie całego kraju (**Country/Region**), bez podziału na podlokalizacje (**Province/State**). Innymi słowy, jeżeli dla danego kraju poziom szczegółowości raportowania obejmuje podlokalizacje, należy wszystkie dzienne wartości zsumować do jednej *krajowej*.
2. Wartości **Confirmed** oraz **Recovered** (a także **Deaths**) dostępne w danych źródłowych są wartościami sumarycznymi dla kraju (lub podlokalizacji) od początku prowadzenia obserwacji. Oznacza to, że w

ramach rozwiązania zadania należy wyznaczyć dzienne przyrosty rozumiane jako różnica wartości dla danego dnia i dnia poprzedniego.

3. Jeżeli w danym dniu w kraju obserwacji nie występują wartości **Confirmed** lub **Recovered** lub którakolwiek z nich jest ujemna, nie należy wyznaczać dziennego przyrostu z wykorzystaniem tego dnia.
4. Tak samo nie należy wyznaczać stosunku dziennych przyrostów dwóch liczb, jeżeli mianownik (dzienny przyrost **Confirmed**) wynosi 0.

Format odpowiedzi

W odpowiedzi wskazać należy 1% (pierwszy percentyl) krajów o największym stosunku dziennego przyrostu osób wyzdrowiałych (**Recovered**) do dziennego przyrostu liczby potwierdzonych zakażeń (**Confirmed**). Odpowiedź powinna zawierać wiersze w kolejności dni miesiąca rosnąco, a w ramach dni miesiąca w kolejności dziennego stosunku przyrostów malejąco. Jeden wiersz musi zawierać następujące dane:

1. numer dnia miesiąca;
2. nazwę kraju;
3. wartość dziennego stosunku przyrostów osób wyzdrowiałych (**Recovered**) do potwierdzonych zakażeń (**Confirmed**).