

Statystyka COVID na świecie

Źródło danych

Dane zostały pobrane za strony <https://www.worldometers.info/coronavirus/> z tablic dla każdego kontynentu i połączone w jedną tabelę.

Data pobrania danych: 18 maja 2024 roku.

Przygotowanie danych

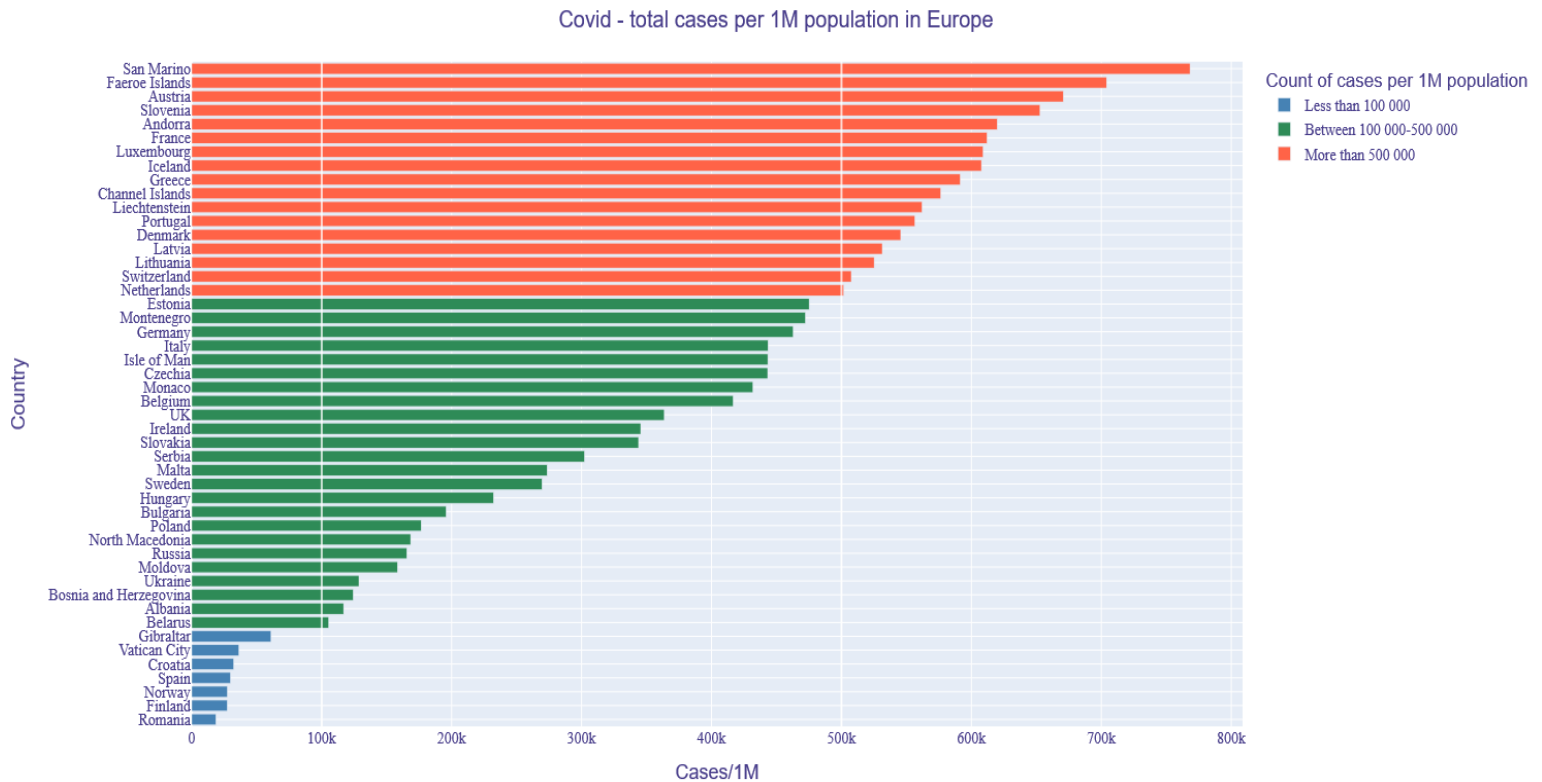
- Pobrane dane zostały przekształcone w następujący sposób:
- Usunięto kolumny: „New Cases”, „New Deaths”, „New Recovered”, „Serious, Critical”
- Dodano nową kolumnę: „Continent”
- Uzupełniono zawartość nową kolumny na podstawie nazw tabel ze strony powyżej
- Skrócono nazwy pozostałych kolumn
- Usunięto przecinki w liczbach
- Przystosowano zawartość tabeli do kodowania UTF-8
- Zapisano dane w formacie .csv

Przygotowanie wykresów

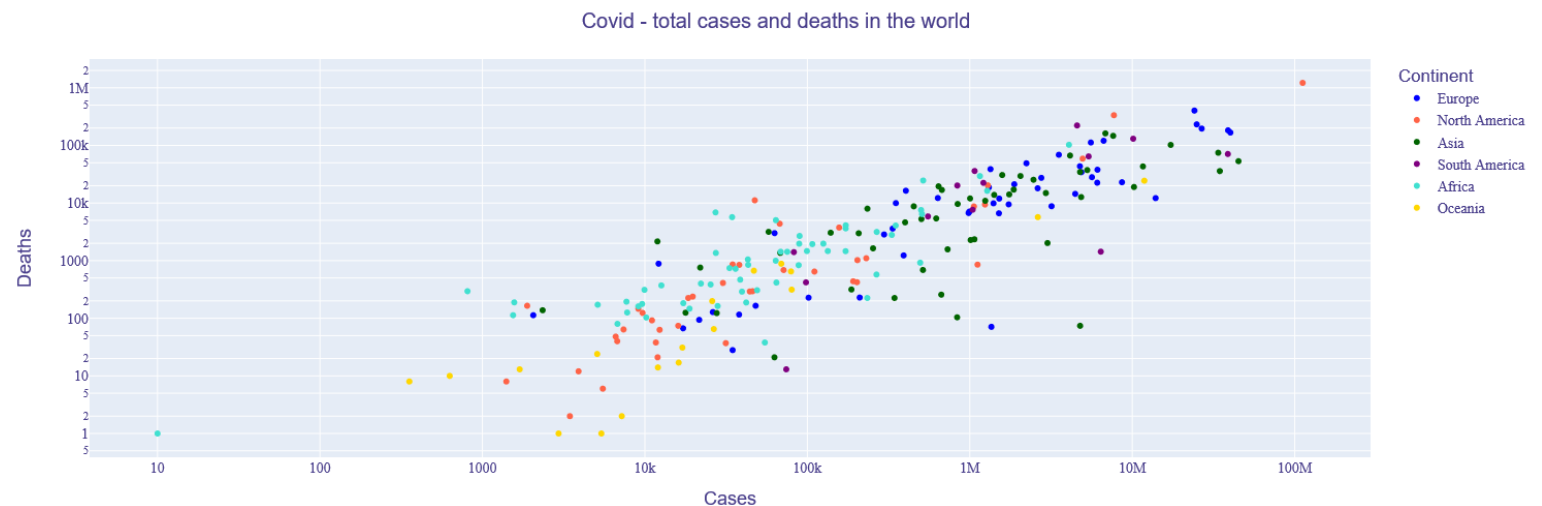
- Wybór środowiska - Python
- Importowanie modułów – Pandas, Plotly Express
- Przesłanie i rozpakowanie pliku .csv z danymi
- Utworzenie zmiennej z wybranymi kolumnami z danych
- Usunięcie wierszy z pustymi wartościami w zbiorze danych zmiennej
- Wizualizacja danych

Wygenerowane wykresy

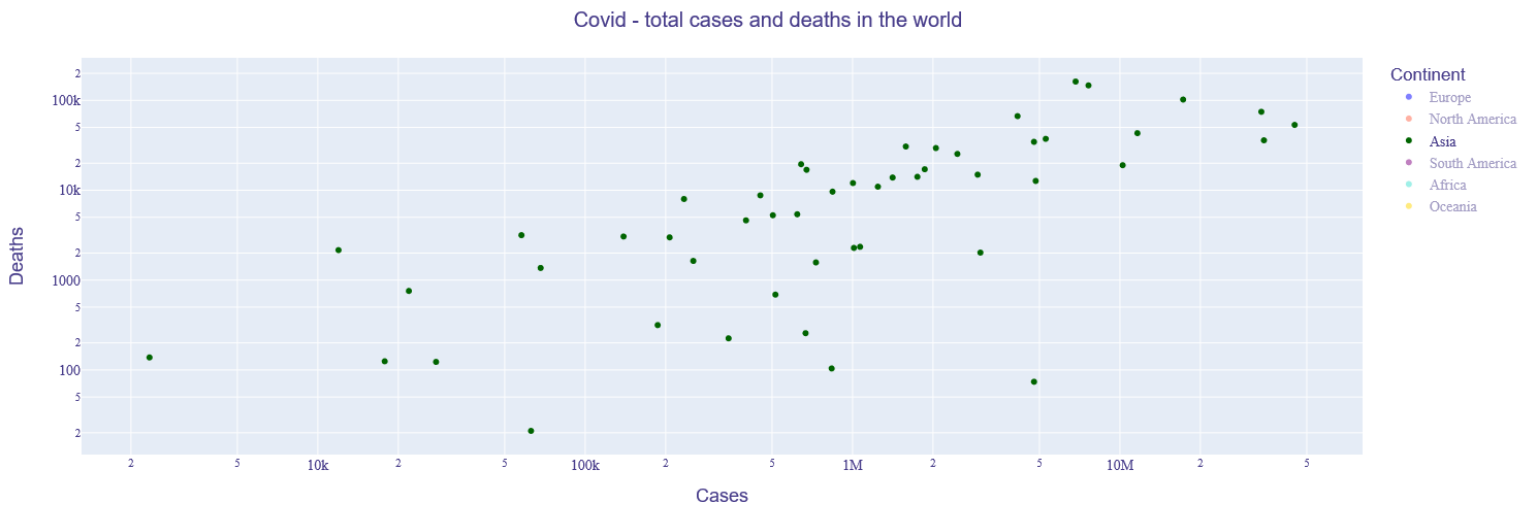
1) Wszystkie przypadki COVID na milion ludności w Europie



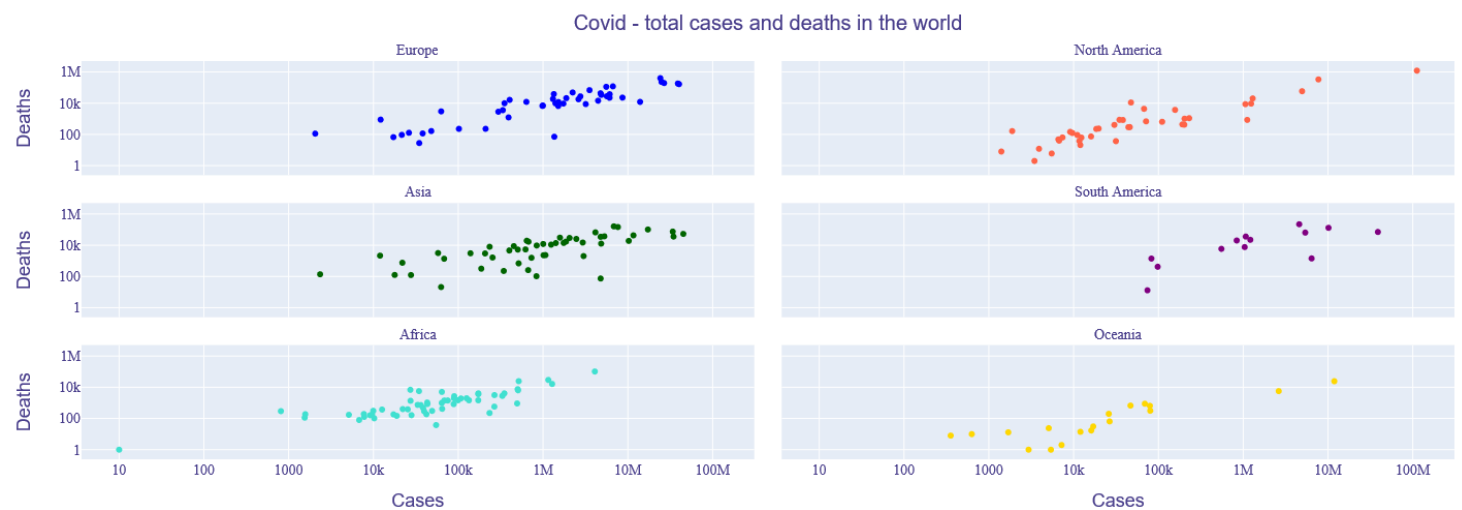
2) Wszystkie przypadki COVID i zgony na świecie



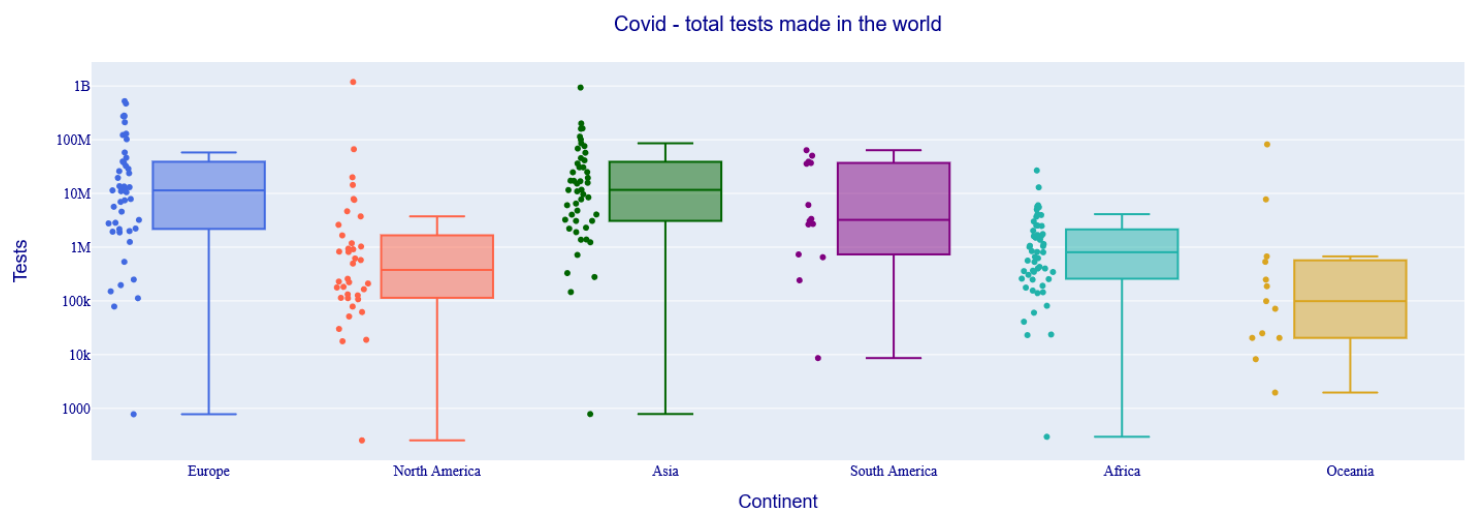
3) Wszystkie przypadki COVID i zgony w Azji



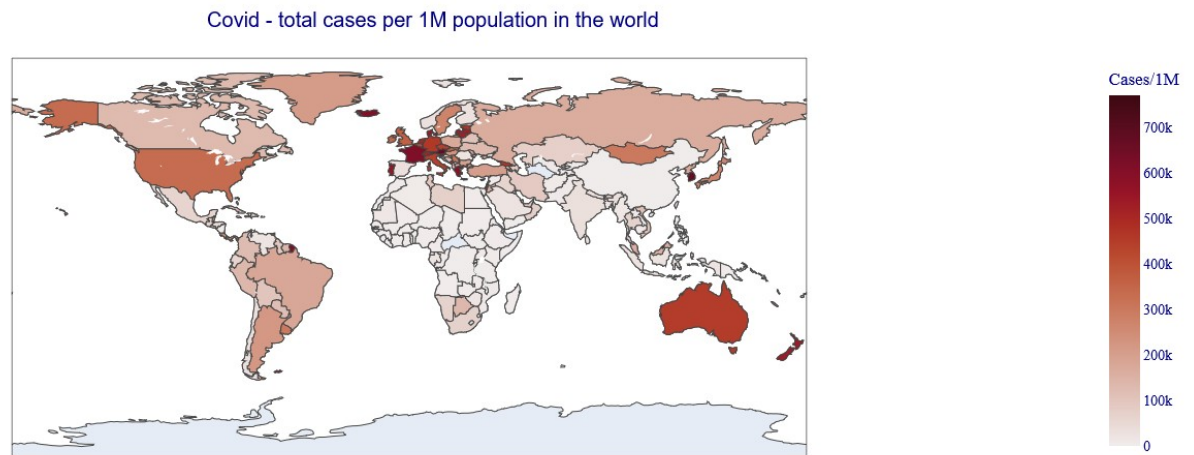
4) Wszystkie przypadki COVID i śmierci na świecie, kontynenty na osobnych wykresach



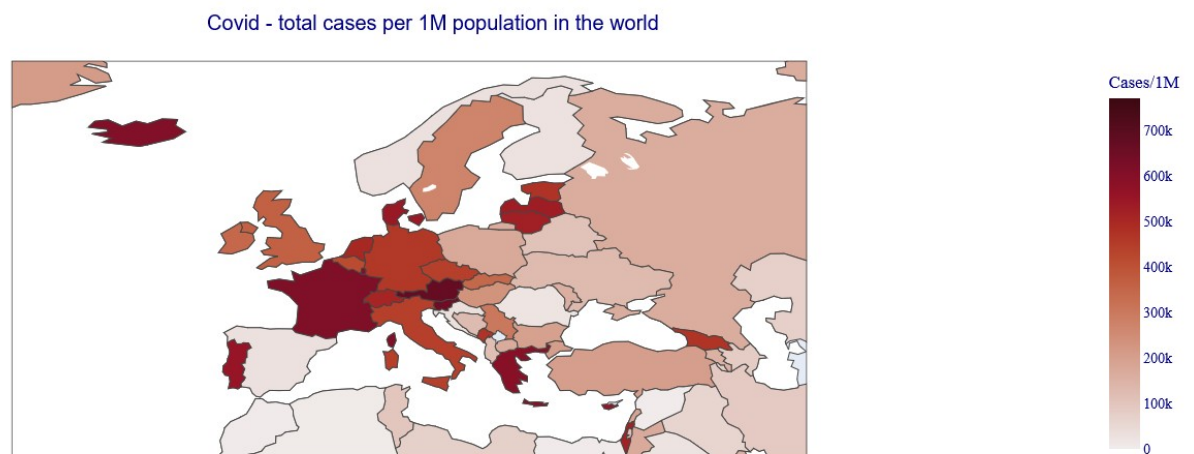
5) Wszystkie wykonane testy na COVID na świecie



6) Wszystkie przypadki COVID na milion ludności na mapie świata



7). Wszystkie przypadki COVID na milion ludności na mapie świata, z przybliżeniem na Europę



Odrzucone wykresy

a) Wszystkie przypadki COVID i zgony na świecie – wykres bąbelkowy, populacja jako wielkość punktów

Powód: Większość punktów staje się niewidoczna ze względu na niedużą populację, pozostają widoczne tylko wielkie państwa, takie jak Chiny, Indie, USA.

b) Wszystkie ozdrowienia, zgony i aktywne przypadki w Europie – wykres słupkowy warstwowy

Powód: liczba zgonów i aktywnych przypadków jest kilkadziesiąt razy mniejsza niż ozdowień, przez co nie są widoczne.

c) Wykres liniowy

Powód: brak zmiennej odnoszącej się do daty, brak możliwości przedstawienia zmian w czasie w zbiorze danych