

# Wykrywanie podobieństw między państwami przy użyciu analizy danych ekonomicznych i socjologicznych Banku Światowego

Projekt Eksploracja Danych  
2022

Wiktoria Paluch  
Weronika Owsianka  
Kamil Szkoła

<b>Wprowadzenie</b>	<b>4</b>
<b>Przygotowanie danych</b>	<b>5</b>
2.1 Źródło danych	5
2.2 Zbiór danych	5
2.2.1 Lista państw	5
2.2.2 Wskaźniki rozwoju	5
2.2.2.1 Unemployment - procent bezrobocia	5
2.2.2.2 Immunization, measles (% of children ages 12-23 months) - procent szczepień przeciw odrze wśród dzieci	6
2.2.2.3 Death rate, crude (per 1,000 people) - wskaźnik śmiertelności na 1000 ludności	7
2.2.2.4 Fertility rate, total (births per woman) - całkowity współczynnik dzietności	8
2.2.2.5 GDP growth (annual %) - wzrost PKB	9
2.2.2.6 Inflation, GDP deflator (annual %) - Inflacja, deflator PKB (% roczny)	9
2.2.2.7 Population growth (annual %) - procentowy roczny wzrost populacji	10
2.2.2.8 Labor force, female (% of total labor force) - procent kobiet wśród grupy aktywnie zawodowo	10
2.2.2.9 Mortality rate, under-5 (per 1,000 live births) - wskaźnik śmiertelności wśród dzieci poniżej 5 lat	11
2.2.2.10 Urban population (% of total population) - % ludności miejskiej	12
2.2.2.11 Literacy rate, youth total (% of people ages 15-24) - % analfabetyzmu wśród dzieci	13
2.2.2.12 Exports of goods and services (% of GDP) - eksport towarów i usług	
2.2.2.13 Imports of goods and services (% of GDP) - import towarów i usług	
2.2.2.14 Population in urban agglomerations of more than 1 million (% of total population) - % populacji mieszkających w aglomeracjach dużych miejskich (ponad 1 mln mieszkańców)	14
2.2.2.15 Proportion of seats held by women in national parliaments (%) - procent miejsc zajmowanych przez kobiety w parlamencie krajowym	16
2.2.2.16 Labor force, total - Sia robocza	17
2.2.2.17 Military expenditure (% of GDP) - wydatki wojskowe (% PKB)	18
2.2.2.18 Taxes on income, profits and capital gains (% of revenue) - podatki od dochodu (procent przychodu)	18
2.2.2.19 Listed domestic companies, total - Krajowe spółki giełdowe	19
2.2.3 Redukcja danych	19
<b>3 Wstępna analiza danych</b>	<b>21</b>
3.1 Analiza podstawowych cech statystycznych za pomocą wykresów skrzynkowych	21
3.2 Korelacja między zmiennymi	23
3.2.1 Macierz korelacji	23
3.2.2 Biplots	25
3.2.3 Redukcja zmiennych	26

<b>3.2.4 Wizualizacja cech zrzutowanych na dwuwymiarową przestrzeń w celu znalezienia państw podobnych</b>	<b>26</b>
<b>4 Metoda klasteryzacji k-średnich</b>	<b>28</b>
4.1 Teoria	28
4.2 Klasteryzacja zbioru danych	28
4.2.1 Liczba skupień	28
4.2.2 Metoda elbow	29
4.2.3 Wizualizacja k-means	30
4.3 Klasteryzacja połączonego zbioru danych ze wszystkich lat	31
4.3.1 Liczba skupień	31
4.3.2 Metoda elbow	31
4.3.3 Wizualizacja k-means	31
4.3.4 Analiza utworzonych klastrów państw	32
<b>5. Hierarchiczna analiza skupień</b>	<b>36</b>
5.1 Teoria	36
5.2 Dendrogramy	36
5.3 Wizualizacja skupień AgglomerativeClustering	38
5.4 Analiza państw w poszczególnych klastrach	38
5.5 Analiza zawartości wśród klasyfikacji państw w poszczególnych latach.	42
5.5.1 Kostaryka	43
5.5.2 Azerbejdżan	47
5.5.3 Gruzja	50
5.5.4 Salvador	52
5.5.5 Ukraina	55
5.6 Ciekawe spostrzeżenia dotyczące rozwoju państw	57
<b>6. Wyciągnięcie wniosków poprzez zbadanie państw pod względem współczynnika szczęścia</b>	<b>58</b>
6.1 Teoria	58
6.2 Wykorzystywany zbiór danych	58
6.2.1 Pochodzenie zbioru	58
6.2.2 Dobór zbioru	58
6.2.3 Opisywane zmienne	59
6.2.4 Projekcja zbioru w 2D	59
6.3 Analiza współczynnika szczęścia w oparciu o grupy państw podobnych	61
6.3.1 Analiza indeksu na tle grup państw podobnych	61
6.3.2 Analiza współczynnika w grupie pierwszej	62
6.3.3 Analiza współczynnika w grupie drugiej	63
6.3.4 Analiza współczynnika w grupie trzeciej	64
<b>7. Zakończenie</b>	<b>66</b>

# 1. Wprowadzenie

W ciągu ostatnich dwóch dekad zastosowanie metod statystycznych i uczenia maszynowego do danych ekonomicznych i społecznych wzbudziły duże zainteresowanie. Grupowanie hierarchiczne używane jest zarówno jako wstępny krok analizy danych, jak i pełnoprawne narzędzie badawcze. Do najczęstszych zastosowań należy właśnie eksploracja danych (data mining), gdzie grupowanie używane jest np. do podziału klientów na pewne podgrupy. W pracy zastosowano hierarchiczne grupowanie wybranych krajów według różnych wskaźników np. populacji, pkb, eksport towarów i usług w celu określenia relacji między poziomami rozwoju krajów.

## 2. Przygotowanie danych

### 2.1 Źródło danych

Dane pochodzą ze strony World Bank Open Data <https://data.worldbank.org/>. Ze strony pobrano wartości wskaźników na przełomie lat 2000 - 2020 opisujące każde z 76 państw.

### 2.2 Zbiór danych

#### 2.2.1 Lista państw

Lista państw, które brane są pod uwagę:

```
['Angola' 'Armenia' 'Australia' 'Austria' 'Azerbaijan' 'Belgium'  
'Bulgaria' 'Belarus' 'Bolivia' 'Brazil' 'Canada' 'Switzerland' 'Chile'  
'China' 'Cameroon' 'Colombia' 'Costa Rica' 'Czech Republic' 'Germany'  
'Denmark' 'Algeria' 'Ecuador' 'Spain' 'Finland' 'France' 'United  
Kingdom' 'Georgia' 'Greece' 'Guatemala' 'Hungary' 'Indonesia' 'India'  
'Ireland' 'Iran, Islamic Rep.' 'Israel' 'Italy' 'Jordan' 'Japan'  
'Kazakhstan' 'Kenya' 'Cambodia' 'Korea, Rep.' 'Kuwait' 'Lebanon'  
'Morocco' 'Mexico' 'Mali' 'Mongolia' 'Mozambique' 'Malaysia' 'Nigeria'  
'Nicaragua' 'Netherlands' 'Norway' 'Nepal' 'New Zealand' 'Panama' 'Peru'  
'Philippines' 'Poland' 'Portugal' 'Paraguay' 'Romania' 'Russian  
Federation' 'Rwanda' 'Senegal' 'Singapore' 'El Salvador' 'Sweden'  
'Thailand' 'Turkey' 'Uganda' 'Ukraine' 'Uruguay' 'United States'  
'World']
```

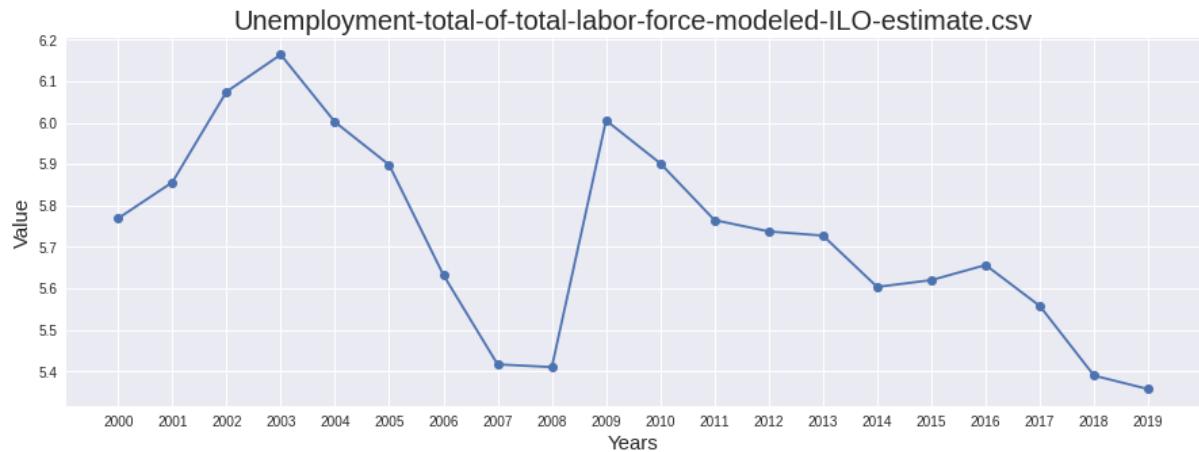
#### 2.2.2 Wskaźniki rozwoju

##### 2.2.2.1 Unemployment - procent bezrobocia

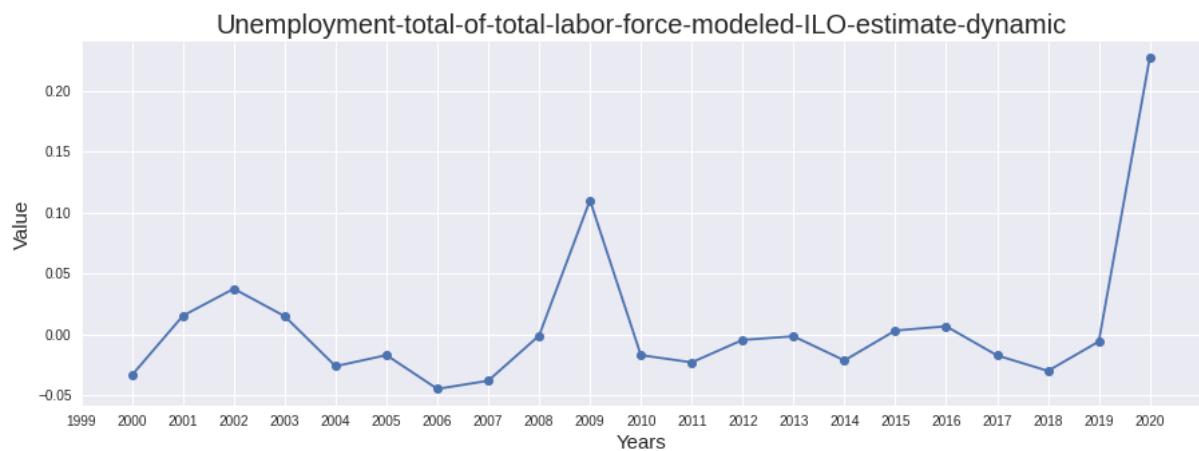
Definicja: Bezrobocie odnosi się do procentu całej siły roboczej, która nie ma pracy, ale jest dostępna i szuka zatrudnienia.

Opis: Standardowa definicja osób bezrobotnych to osoby bez pracy, poszukujące pracy w ostatnim okresie i obecnie dostępne do pracy, w tym osoby, które straciły pracę lub które dobrowolnie ją opuściły. Do bezrobotnych zalicza się również osoby, które nie szukały pracy, ale mają przygotowania do przyszłej pracy. Niektóre bezrobocie jest nieuniknione. W dowolnym momencie niektórzy pracownicy są czasowo bezrobotni między miejscami pracy, ponieważ pracodawcy szukają odpowiednich pracowników, a pracownicy szukają lepszej pracy. Podstawą tego wskaźnika jest siła robocza lub aktywna zawodowo część populacji, a nie ogół ludności. Szeregi stanowią część szacunków MOP (Międzynarodowa Organizacja Pracy) i są zharmonizowane w celu zapewnienia porównywalności między krajami i w czasie poprzez uwzględnienie różnic w źródłach danych, zakresie pokrycia, metodologii i innych czynnikach specyficznych dla danego kraju. Szacunki opierają się głównie na reprezentatywnych dla całego kraju badaniach siły roboczej, przy czym inne źródła (spisy

ludności i dane szacunkowe podawane na poziomie krajowym) są wykorzystywane tylko wtedy, gdy nie są dostępne żadne dane z badań. Na poniższym wykresie widać znaczny wzrost bezrobocia do roku 2019 bezrobocie spadało (nie na wykresie: w 2020 nastąpił wzrost związany z rozwojem pandemii covid-19).



Gdy zamienimy na dynamiczną widać skutki w bezrobociu kryzysu z 2008 roku:

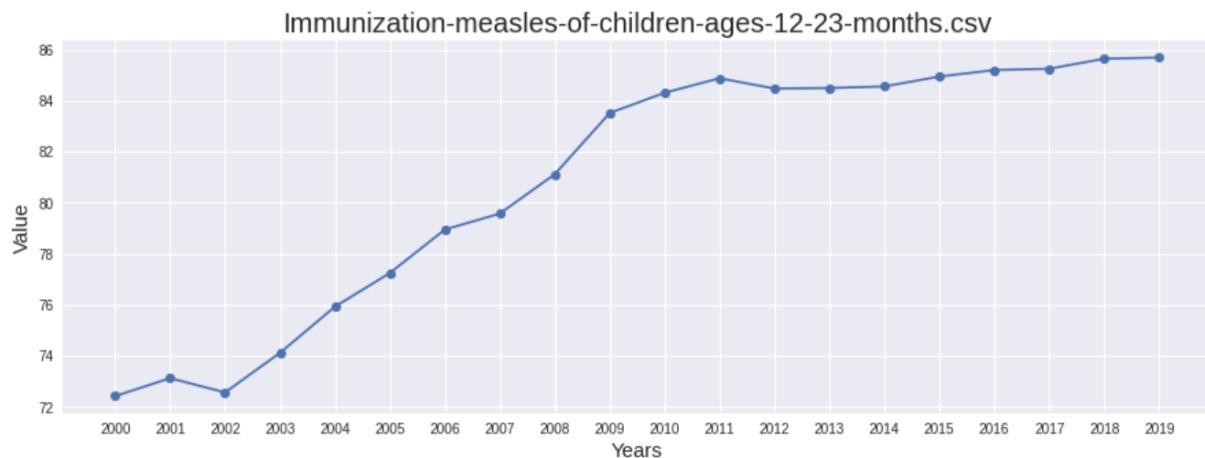


#### 2.2.2.2 Immunization, measles (% of children ages 12-23 months) - procent szczepień przeciw odrze wśród dzieci

Definicja: Szczepienia dzieci na odrę, mierzą odsetek dzieci w wieku 12-23 miesięcy. Dziecko uważa się za odpowiednio uodpornione przeciwko odrze po otrzymaniu jednej dawki szczepionki.

Opis: Kraje rozwijające się zazwyczaj finansują szczepienia przeciwko odrze i błonicy, krztuścowi (kokluszowi) i tężcowi (DTP) w ramach podstawowego pakietu zdrowia publicznego. Przedstawione dane opierają się na ocenie krajowych wskaźników szczepień dokonanych przez WHO i UNICEF. W ocenie uwzględniono zarówno dane administracyjne od dostawców usług, jak i dane ankietowe gospodarstw domowych dotyczące historii szczepień dzieci. Na podstawie dostępnych danych, rozważenia potencjalnych błędów systematycznych i wkładu lokalnych ekspertów, dla każdego roku określano najbardziej

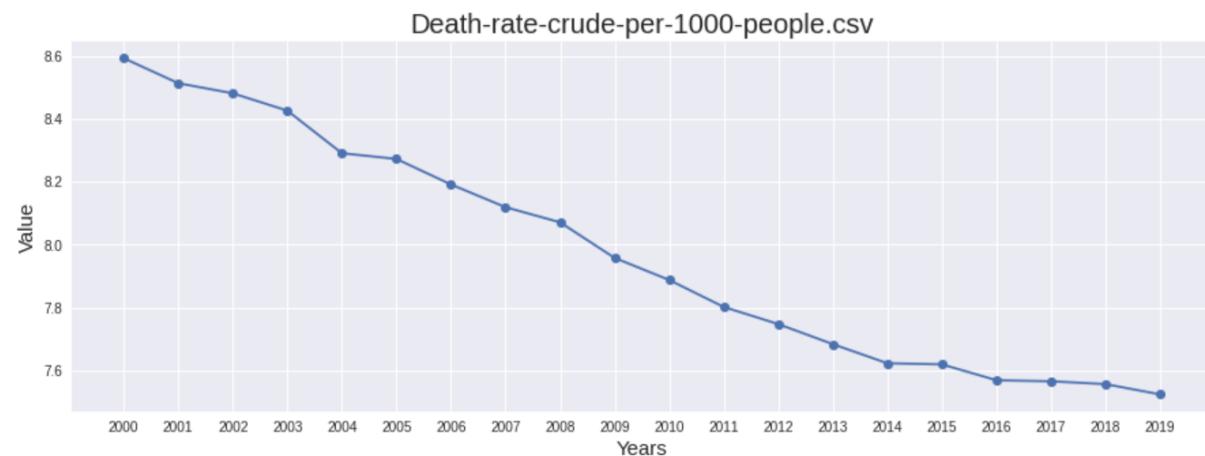
prawdopodobny prawdziwy poziom szczepień. Uwagi dotyczące agregatów regionalnych i globalnych: Gdy szczepionka nie jest wprowadzana do krajowego harmonogramu szczepień, zakłada się, że brakująca wartość wynosi zero (lub bliskie zeru) w średnich odpowiednich grupach.



#### 2.2.2.3 Death rate, crude (per 1,000 people) - wskaźnik śmiertelności na 1000 ludności

Definicja: Wskaźnik zgonów wskazuje liczbę zgonów występujących w ciągu roku na 1000 ludności oszacowaną w połowie roku.

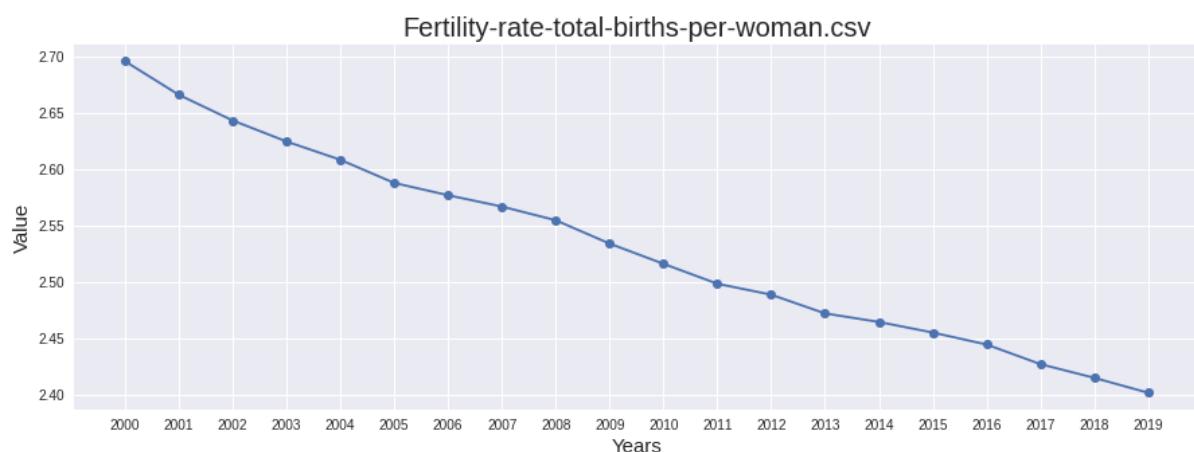
Opis: Surowy współczynnik zgonów oblicza się jako liczbę zgonów w danym okresie podzieloną przez populację narażoną w tym okresie na ryzyko zgonu. W przypadku populacji ludzkich okres ten wynosi zwykle jeden rok, a jeśli populacja zmienia się pod względem wielkości w ciągu roku, dzielnik przyjmuje się jako populację w połowie roku. Wskaźnik jest zwykle wyrażany w kategoriach 1000 osób: na przykład surowy wskaźnik zgonów wynoszący 9,5 (na 1000 osób) w populacji 1 miliona oznaczałby 9500 zgonów rocznie w całej populacji. Odejmując surowy wskaźnik śmiertelności od surowego wskaźnika urodzeń, otrzymujemy wskaźnik przyrostu naturalnego, który jest równy tempowi zmiany populacji w przypadku braku migracji. Wskaźniki żywotne oparte są na danych z systemów rejestracji urodzeń i zgonów, spisów powszechnych i badań próby prowadzonych przez krajowe urzędy statystyczne i inne organizacje lub na analizie demograficznej. Dane za ostatni rok dla niektórych krajów o wysokim dochodzie są wstępymi szacunkami opartymi na rejestrach stanu cywilnego. Szacunki dla wielu krajów są prognozami opartymi na ekstrapolacji poziomów i trendów z wcześniejszych lat lub interpolacji szacunków populacji i prognoz z Wydziału Ludności ONZ.



#### 2.2.2.4 Fertility rate, total (births per woman) - całkowity współczynnik dzietności

Definicja: Współczynnik dzietności – współczynnik określający przeciętną liczbę dzieci, które urodziłaby kobieta w ciągu całego okresu rozrodczego, przy założeniu, że w poszczególnych fazach tego okresu rodziłaby z intensywnością obserwowaną wśród kobiet w badanym roku.

Opis: Całkowite współczynniki dzietności opierają się na danych dotyczących zarejestrowanych żywych urodzeń z systemów rejestracji stanu cywilnego lub, w przypadku braku takich systemów, ze spisów lub badań reprezentacyjnych. Szacowane wskaźniki są ogólnie uważane za wiarygodne miary płodności w niedalekiej przeszłości. W przypadku braku danych empirycznych na temat współczynników dzietności w zależności od wieku, do oszacowania odsetka urodzeń nastolatków stosuje się model. W krajach, w których nie ma systemów rejestracji życiowych, współczynniki dzietności są generalnie oparte na ekstrapolacji trendów obserwowanych w spisach lub badaniach z lat wcześniejszych.



Ze względu na liniową tendencję zmieniono dodatkowo zmienną na zmienną dynamiczną:

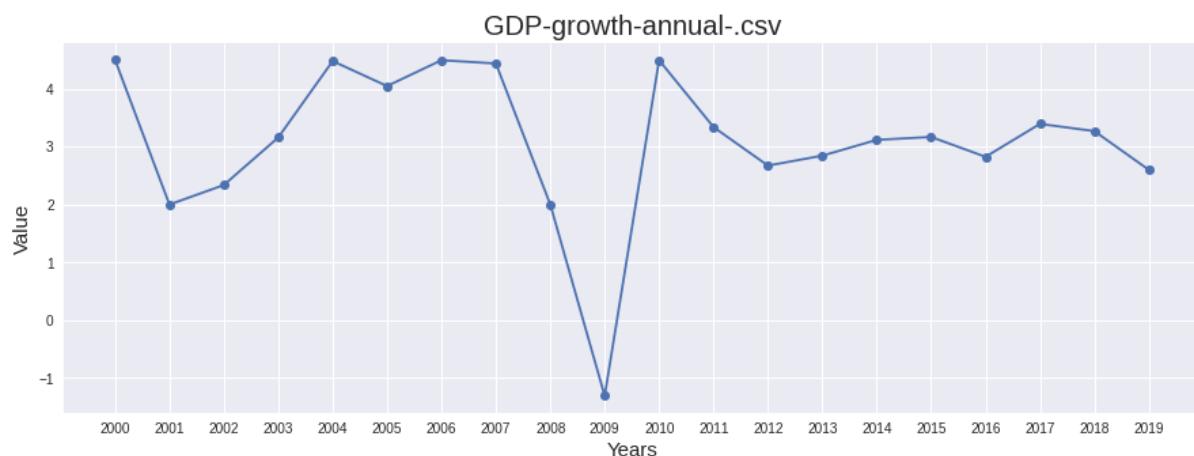


plik: "Fertility-rate-total-births-per-woman-dynamic.csv"

#### 2.2.2.5 GDP growth (annual %) - wzrost PKB

Definicja: Roczna procentowa stopa wzrostu PKB w cenach rynkowych w oparciu o stałą walutę lokalną. Agregaty są oparte na stałych cenach z 2015 r., wyrażonych w dolarach amerykańskich. PKB to suma wartości dodanej brutto wytworzonej przez wszystkich krajowych producentów w gospodarce powiększona o podatki od produktów i pomniejszona o wszelkie dotacje nieuwzględnione w wartości produktów. Oblicza się ją bez dokonywania odliczeń z tytułu amortyzacji sfabrykowanych aktywów lub wyczerpania i degradacji zasobów naturalnych.

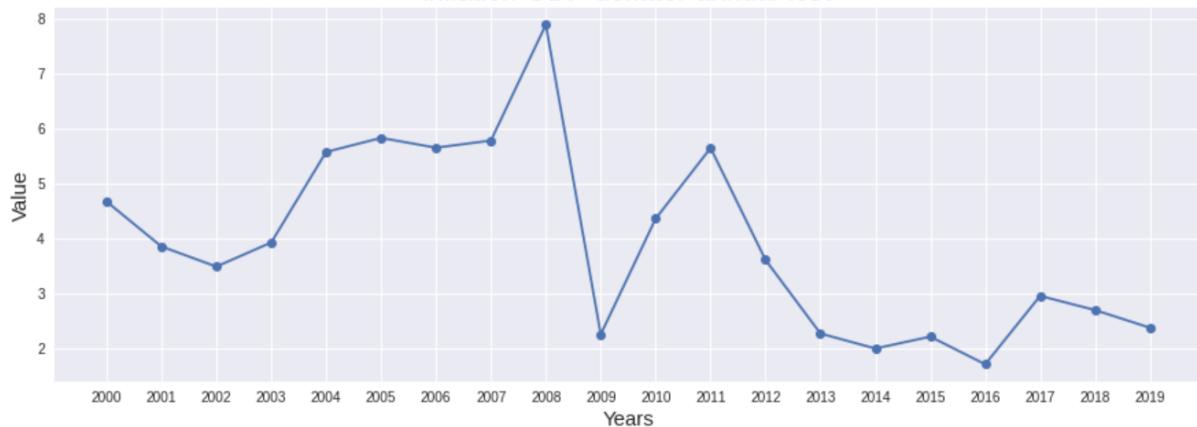
Opis: Produkt krajowy brutto (PKB) stanowi sumę wartości dodanej wszystkich jego producentów. Wartość dodana to wartość produkcji globalnej producentów pomniejszona o wartość dóbr i usług pośrednich zużytych w produkcji, przed rozliczeniem zużycia środków trwałych w produkcji. System Rachunków Narodowych Organizacji Narodów Zjednoczonych wzywa do wyceny wartości dodanej według cen podstawowych (z wyłączeniem podatków netto od produktów) lub cen producenta (w tym podatków netto od produktów płaconych przez producentów, ale z wyłączeniem podatków od sprzedaży lub podatku od wartości dodanej). Obie wyceny nie obejmują opłat transportowych, które są oddzielnie fakturowane przez producentów. Całkowity PKB jest mierzony w cenach zakupu. Wartość dodana przez przemysł jest zwykle mierzona w cenach podstawowych. Gdy wartość dodana jest mierzona w cenach producenta. Tempo wzrostu PKB i jego składników obliczane są metodą najmniejszych kwadratów i danymi o cenach stałych w walucie lokalnej. Cena stała w dolarach amerykańskich służy do obliczania wskaźników wzrostu dla regionu i grupy dochodowej. Serie walut lokalnych są przeliczane na stałe dolary amerykańskie przy użyciu kursu wymiany we wspólnym roku referencyjnym. Na poniższym wykresie zauważamy wyraźny spadek w 2008-2009 spowodowany światowym kryzysem finansowym.



#### 2.2.2.6 Inflation, GDP deflator (annual %) - Inflacja, deflator PKB (% roczny)

Definicja: Inflacja mierzona roczną stopą wzrostu deflatora ukrytego PKB pokazuje tempo zmian cen w gospodarce jako całości. Deflator ukryty PKB to stosunek PKB w bieżącej walucie lokalnej do PKB w stałej walucie lokalnej.

Inflation-GDP-deflator-annual-.csv

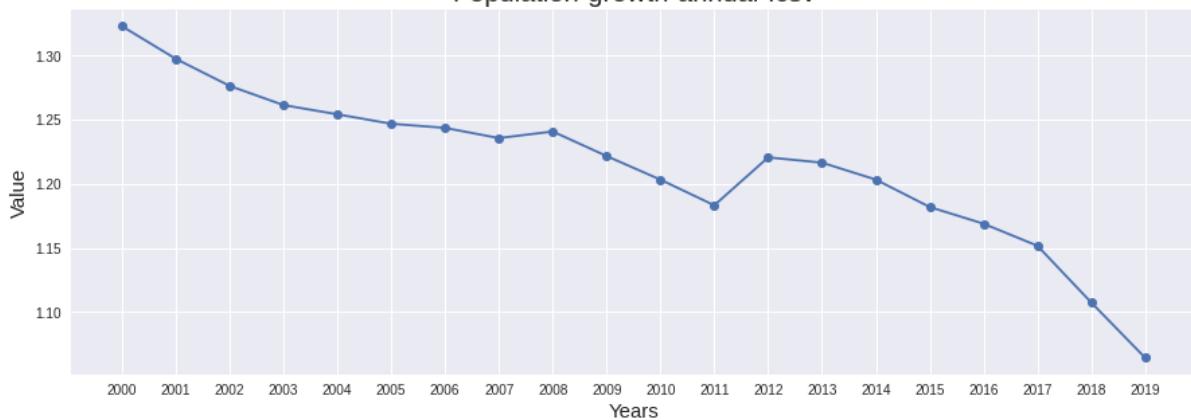


#### 2.2.2.7 Population growth (annual %) - procentowy roczny wzrost populacji

Definicja: Roczna stopa przyrostu ludności w roku  $t$  to wykładnicza stopa przyrostu ludności w połowie roku od roku  $t-1$  do  $t$ , wyrażona w procentach. Populacja opiera się na de facto definicji ludności, która obejmuje wszystkich mieszkańców bez względu na status prawnego czy obywatelstwa.

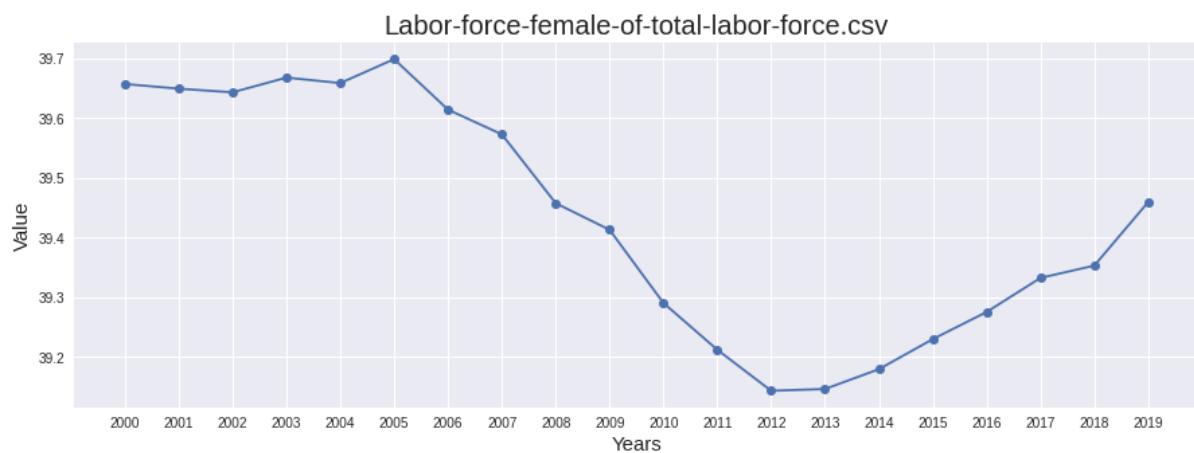
Opis: Całkowite tempo wzrostu populacji oblicza się przy założeniu, że tempo wzrostu jest stałe między dwoma punktami w czasie. Tempo wzrostu oblicza się za pomocą wzoru na wzrost wykładniczy:  $r = \ln(p_n/p_0)/n$ , gdzie  $r$  to wykładnicza stopa wzrostu,  $\ln()$  to logarytm naturalny,  $p_n$  to populacja z końca okresu,  $p_0$  to początek populacji okresu, a  $n$  to liczba lat pomiędzy nimi. Należy zauważyć, że nie jest to tempo wzrostu geometrycznego używane do obliczania wzrostu złożonego w okresach dyskretnych. Aby uzyskać informacje o całkowitej populacji, z której obliczane są wskaźniki wzrostu, zobacz całkowitą populację. Na wykresie widać spadek przyrostu populacji.

Population-growth-annual-.csv



#### 2.2.2.8 Labor force, female (% of total labor force) - procent kobiet wśród grupy aktywnie zawodowo

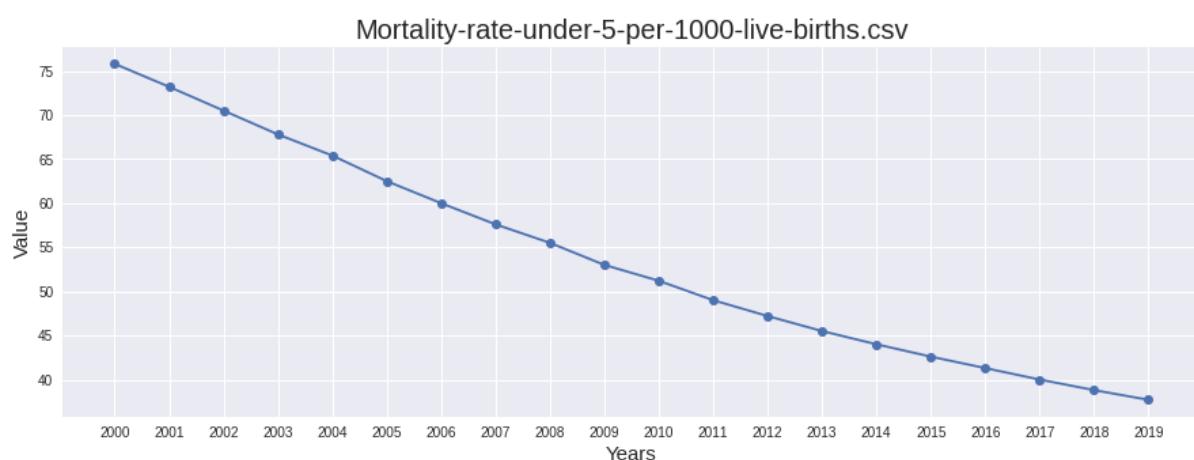
Definicja: Siła robocza kobiet jako odsetek całości pokazuje stopień, w jakim kobiety są aktywne zawodowo. Siła robocza obejmuje osoby w wieku 15 lat i starsze, które dostarczają siłę roboczą do produkcji towarów i usług w określonym okresie.



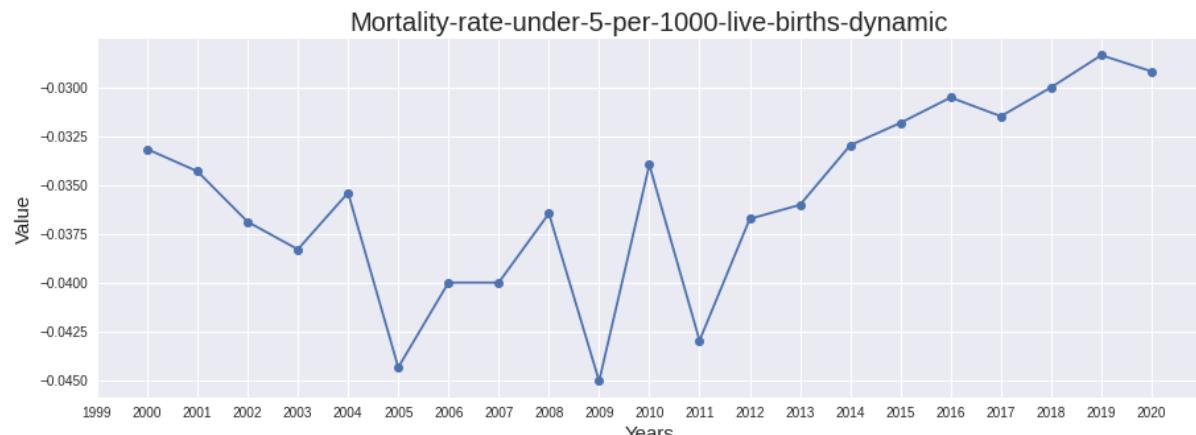
#### 2.2.2.9 Mortality rate, under-5 (per 1,000 live births) - wskaźnik śmiertelności wśród dzieci poniżej 5 lat

Definicja: Wskaźnik śmiertelności dzieci poniżej piątego roku życia na 1000 osób to prawdopodobieństwo, że noworodek umrze przed osiągnięciem piątego roku życia, jeśli podlega wskaźnikom śmiertelności określonym dla wieku w określonym roku.

Opis: Wskaźniki umieralności dla różnych grup wiekowych (niemowlęta, dzieci i dorośli) oraz wskaźniki śmiertelności ogólnej (oczekiwana długość życia w chwili urodzenia lub przeżycie do określonego wieku) są ważnymi wskaźnikami stanu zdrowia w kraju. Ponieważ dane dotyczące zachorowalności i rozpowszechnienia chorób są często niedostępne, wskaźniki śmiertelności są często wykorzystywane do identyfikacji zagrożonych populacji. I należą one do wskaźników najczęściej używanych do porównywania rozwoju społeczno-gospodarczego w różnych krajach.



Zamieniono na zmienną dynamiczną:

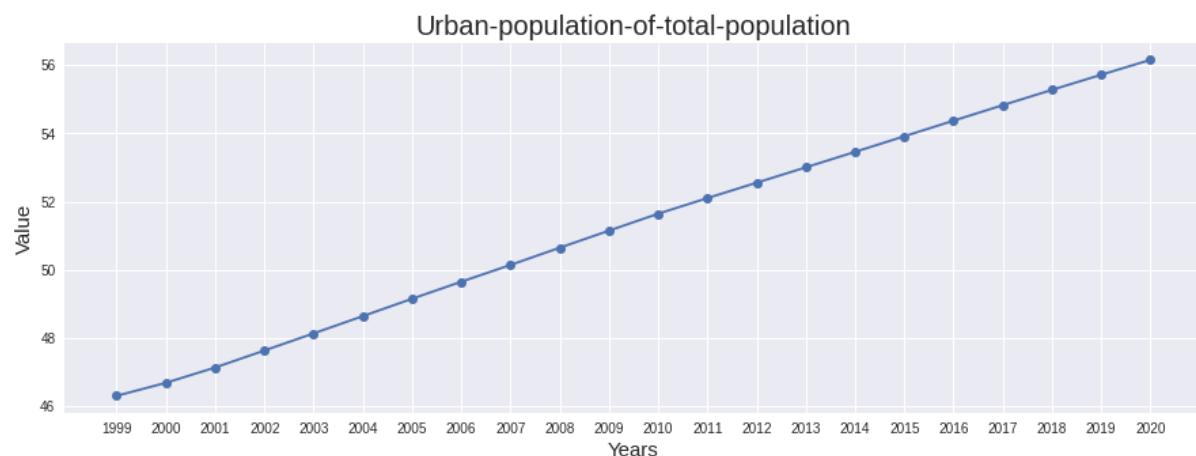


plik: Mortality-rate-under-5-per-1000-live-births-dynamic.csv

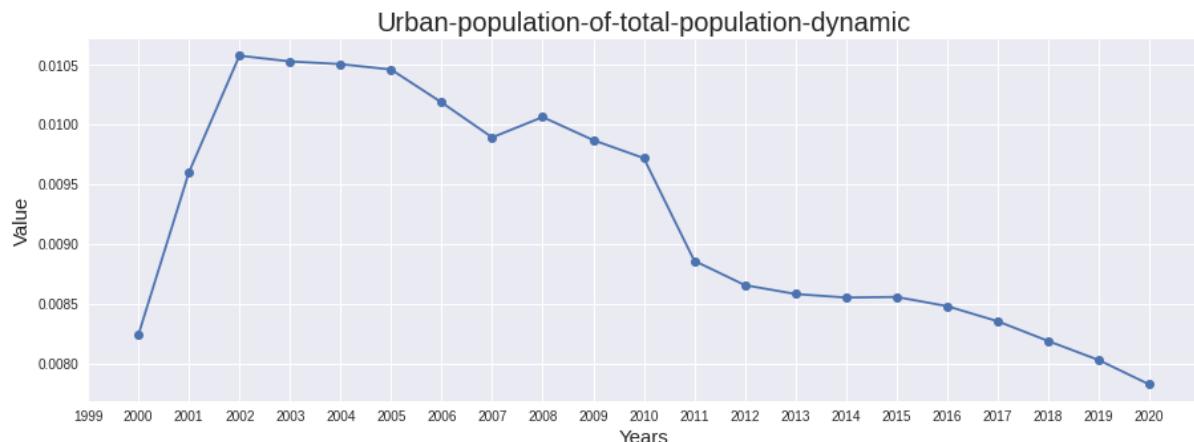
#### 2.2.2.10 Urban population (% of total population) - % ludności miejskiej

Definicja: Ludność miejska odnosi się do osób mieszkających na obszarach miejskich określonych przez krajowe urzędy statystyczne.

Opis: Gwałtowny wzrost miast na całym świecie oznacza przejście demograficzne ze wsi do miast i wiąże się z przejściem z gospodarki opartej na rolnictwie do przemysłu masowego, technologii i usług. W zasadzie miasta oferują korzystniejsze warunki do rozwiązywania problemów społecznych i środowiskowych niż obszary wiejskie. Miasta generują miejsca pracy i dochody, zapewniają edukację, opiekę zdrowotną i inne usługi. Miasta stwarzają również możliwości mobilizacji społecznej i wzmacnienia pozycji kobiet.



Zamieniono na zmienną dynamiczną:

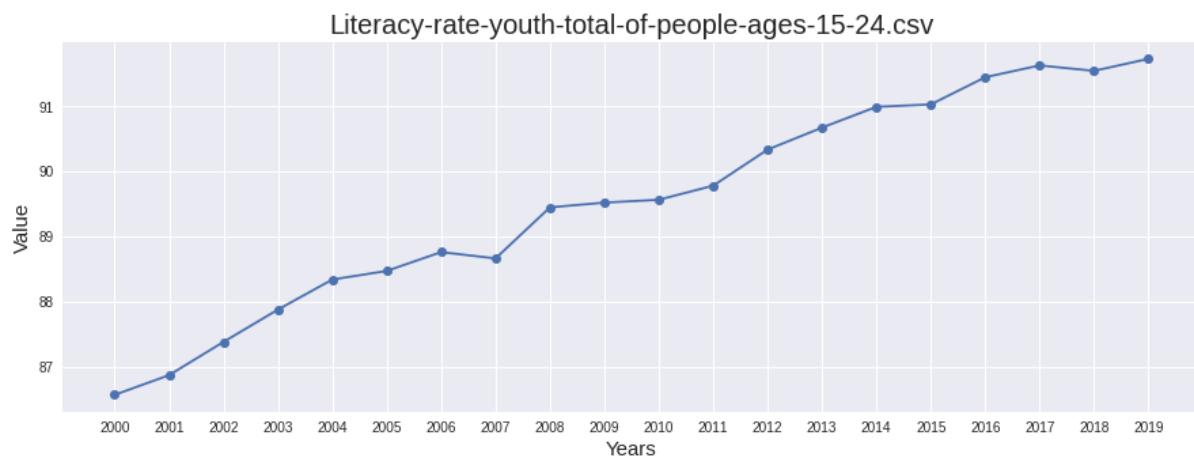


plik: "Urban-population-of-total-population-dynamic.csv"

#### 2.2.2.11 Literacy rate, youth total (% of people ages 15-24) - % analfabetyzmu wśród dzieci

Definicja: Wskaźnik umiejętności czytania i pisania wśród młodzieży to odsetek osób w wieku 15-24 lat, które potrafią czytać i pisać ze zrozumieniem krótkie, proste stwierdzenie dotyczące ich codziennego życia.

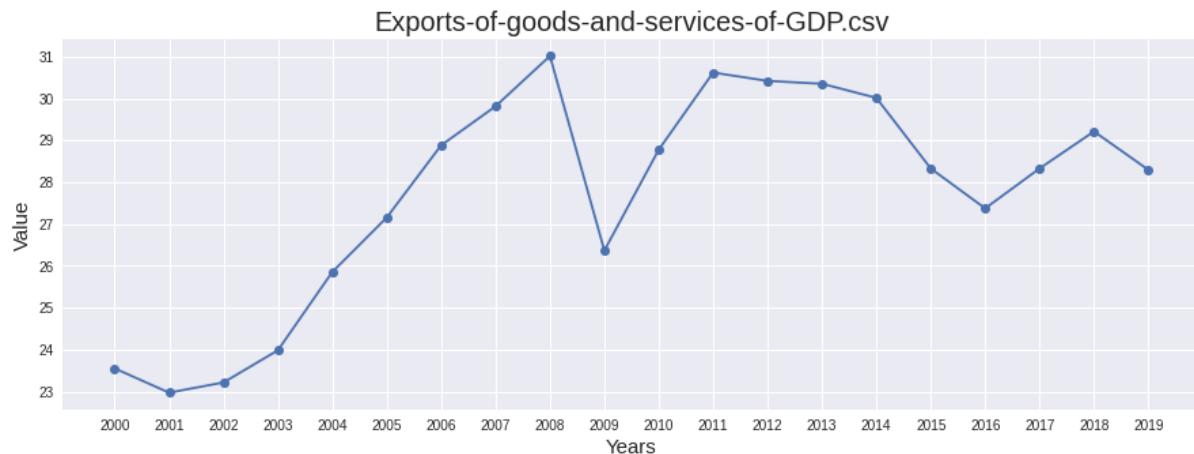
Opis: Wskaźnik umiejętności czytania i pisania jest wskaźnikiem wyniku służącym do oceny osiągnięć edukacyjnych. Dane te mogą przewidywać jakość przyszłej siły roboczej i mogą być wykorzystywane do tworzenia polityk dotyczących umiejętności życiowych mężczyzn i kobiet. Może być również używany jako narzędzie zastępcze do oceny skuteczności systemu edukacji; wysoki wskaźnik umiejętności czytania i pisania sugeruje zdolność systemu edukacji do zapewnienia dużej populacji możliwości nabycia umiejętności czytania i pisania. Skumulowane osiągnięcia edukacyjne mają fundamentalne znaczenie dla dalszego rozwoju intelektualnego oraz rozwoju społeczno-gospodarczego, choć niekoniecznie zapewniają jakość edukacji. Piśmienne kobiety oznaczają, że mogą wyszukiwać i wykorzystywać informacje w celu poprawy zdrowia, odżywiania i edukacji członków swoich gospodarstw domowych. Piśmienne kobiety są również upoważnione do odgrywania znaczącej roli.



#### 2.2.2.12 Exports of goods and services (% of GDP) - eksport towarów i usług

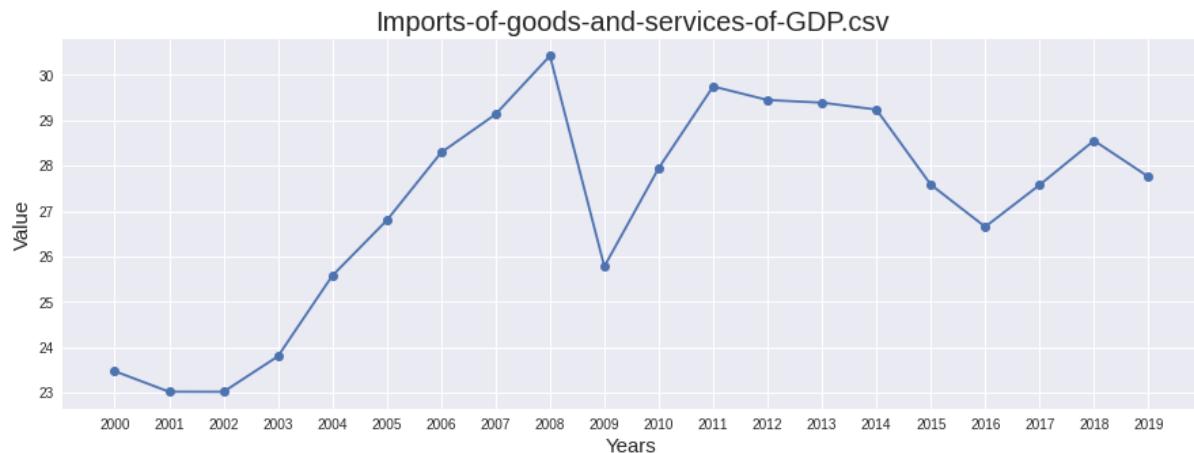
Definicja: Eksport towarów i usług reprezentuje wartość wszystkich towarów i innych usług rynkowych świadczonych na rzecz reszty świata. Obejmują one wartość towarów, frachtu, ubezpieczenia, transportu, podróży, tantiem, opłat licencyjnych i innych usług, takich jak usługi komunikacyjne,

budowlane, finansowe, informacyjne, biznesowe, osobiste i rządowe. Wykluczają wynagrodzenia pracowników i dochody z inwestycji (wcześniej nazywane usługami faktoringowymi) oraz płatności transferowe.



#### 2.2.2.13 Imports of goods and services (% of GDP) - import towarów i usług

Definicja: Import towarów i usług reprezentuje wartość wszystkich towarów i innych usług rynkowych otrzymanych z zagranicy. Obejmują one wartość towarów, frachtu, ubezpieczenia, transportu, podróży, tantiem, opłat licencyjnych i innych usług, takich jak usługi komunikacyjne, budowlane, finansowe, informacyjne, biznesowe, osobiste i rządowe. Wykluczają wynagrodzenia pracowników i dochody z inwestycji (wcześniej nazywane usługami faktoringowymi) oraz płatności transferowe.

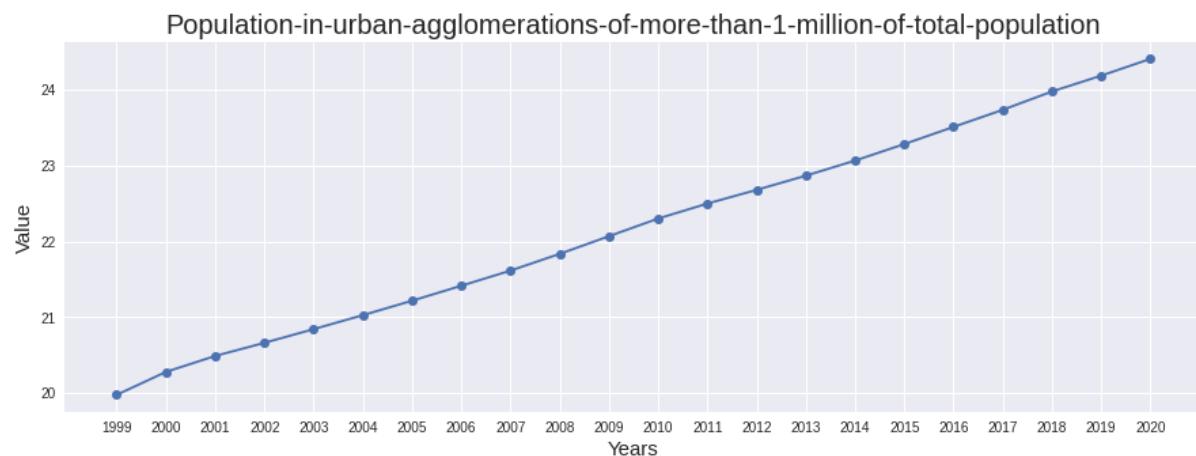


#### 2.2.2.14 Population in urban agglomerations of more than 1 million (% of total population) - % populacji mieszkających w aglomeracjach dużych miejskich (ponad 1 mln mieszkańców)

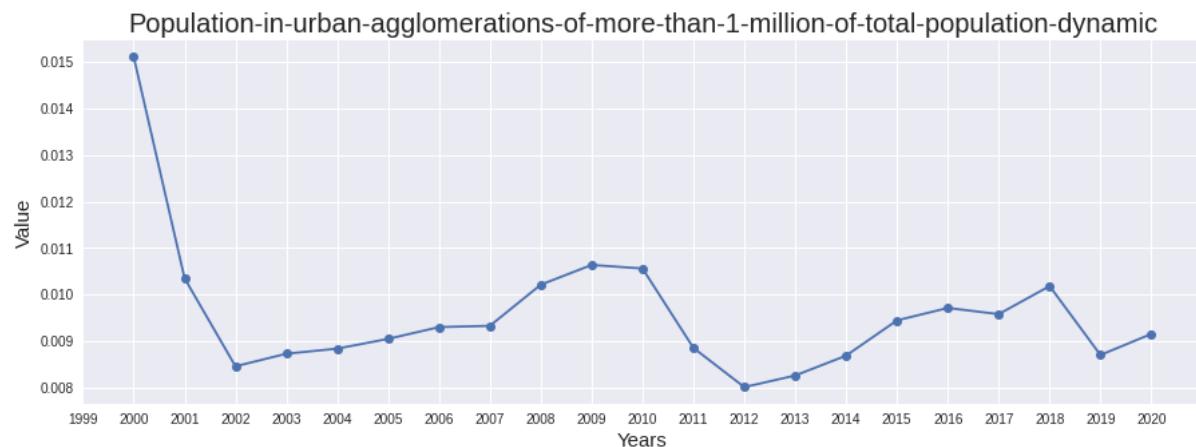
Definicja: Ludność w aglomeracjach miejskich powyżej miliona to odsetek ludności kraju zamieszkującego obszary metropolitalne, które w 2018 r. liczyły ponad milion osób.

Opis: Zgodnie z definicją Organizacji Narodów Zjednoczonych obszar metropolitalny obejmuje zarówno przylegające do siebie terytoria zamieszkałe na poziomach miejskich, jak i dodatkowe obszary otaczające o mniejszej gęstości zasiedlenia, które również znajdują się pod bezpośredniem wpływem miasta (np. dzięki częstym transportom, powiązaniom drogowym), dojazdy itp.). Gwałtowny wzrost miast na całym świecie oznacza przejście demograficzne ze wsi do miast i wiąże się z przejściem z gospodarki opartej na rolnictwie do przemysłu masowego, technologii i usług. Po

raz pierwszy większość światowej populacji mieszka w mieście i odsetek ten stale rośnie. Sto lat temu 2 na 10 osób mieszkało w obszarze miejskim. Do 1990 roku mniej niż 40 procent światowej populacji mieszkało w mieście, ale na początku 2010 roku ponad połowa wszystkich ludzi mieszkała na obszarach miejskich. Do 2030 r. w mieście będzie mieszkać 6 na 10 osób, a do 2050 r. odsetek ten wzrośnie do 7 na 10 osób. Około połowa wszystkich mieszkańców miast mieszka w miastach liczących od 100 000 do 500 000 osób, a mniej niż 10% mieszkańców miast mieszka w megamiasbach (miasto o populacji powyżej 10 milionów, zgodnie z definicją ONZ HABITAT). Obecnie liczba mieszkańców miast rośnie każdego roku o blisko 60 mln. W połowie XXI wieku populacja miejska prawie się podwoi, osiągając w 2050 roku 6,4 miliarda. Prawie cały wzrost liczby ludności miejskiej w ciągu najbliższych 30 lat nastąpi w miastach krajów rozwijających się. Szacuje się, że do połowy XXI wieku populacja miejska krajów rozwijających się podwoi się, osiągając prawie 5,2 miliarda w 2050 roku. Oczekuje się, że w krajach o wysokim dochodzie populacja miejska pozostanie w dużej mierze niezmieniona przez następne dwie dekady, osiągając do 2025 r. nieco ponad 1 miliard. W tych krajach imigracja (legalna i nielegalna) będzie odpowiadać za ponad dwie trzecie wzrostu miejskiego. Bez imigracji populacja miejska w tych krajach najprawdopodobniej zmniejszyłaby się lub pozostała statyczna. W zasadzie miasta oferują korzystniejsze warunki do rozwiązywania problemów społecznych i środowiskowych niż obszary wiejskie. Miasta generują miejsca pracy i dochody, zapewniają edukację, opiekę zdrowotną i inne usługi. Miasta stwarzają również możliwości mobilizacji społecznej i wzmocnienia pozycji kobiet. Ubóstwo rośnie szybciej w miastach niż na wsi. Według ONZ miliard ludzi żyje w miejskich slumsach, które są zazwyczaj przepienione, zanieczyszczone i niebezpieczne, i nie mają podstawowych usług, takich jak czysta woda i kanalizacja.



Zamiana na zmienną dynamiczną:

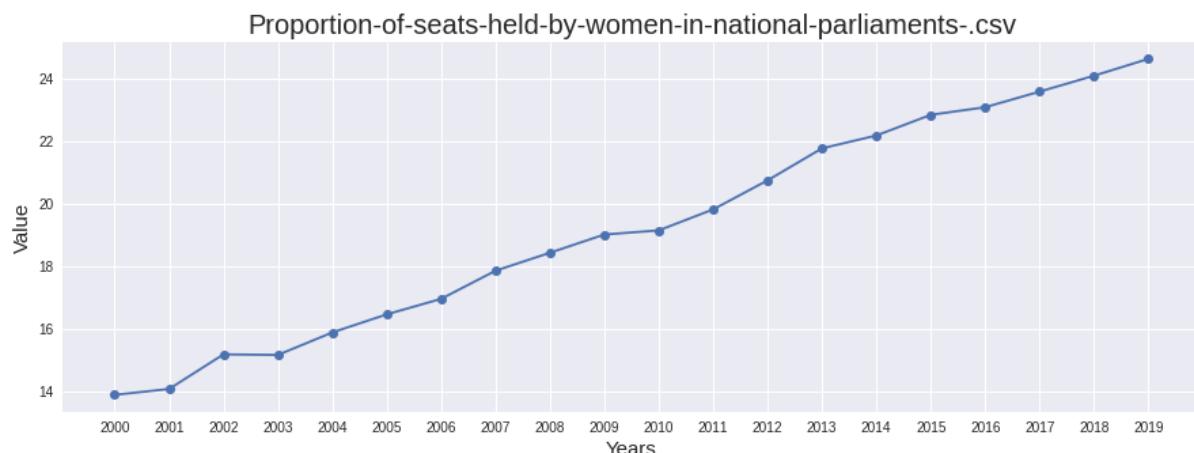


plik: "Population-in-urban-agglomerations-of-more-than-1-million-of-total-population-dynamic.csv"

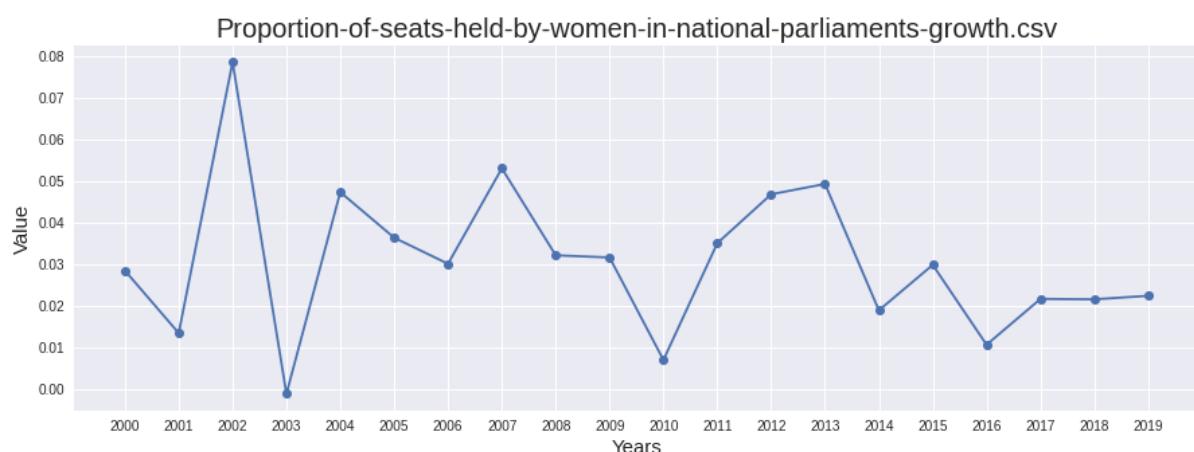
## 2.2.2.15 Proportion of seats held by women in national parliaments (%) - procent miejsc zajmowanych przez kobiety w parlamencie krajowym

Definicja: Kobiety w parlamentach to odsetek mandatów parlamentarnych przyznanych kobietom.

Opis: Pomimo znacznego postępu w ostatnich dziesięcioleciach, nierówności płci są wszechobecne w wielu wymiarach życia – na całym świecie. Ale chociaż rozbieżności istnieją na całym świecie, są one najbardziej rozpowszechnione w krajach rozwijających się. Nierówności płci w alokacji takich zasobów, jak edukacja, opieka zdrowotna, odżywianie i głos polityczny mają znaczenie ze względu na silny związek z dobrostanem, produktywnością i wzrostem gospodarczym. Te wzorce nierówności zaczynają się już w młodym wieku, a chłopcy rutynowo otrzymują większą część wydatków na edukację i zdrowie niż na przykład dziewczynki. Kobiety są zdecydowanie niedostatecznie reprezentowane na stanowiskach decyzyjnych w rządzie, chociaż istnieją pewne dowody na niedawną poprawę. Parytet płci w reprezentacji parlamentarnej jest wciąż daleki od realizacji. Bez reprezentacji na tym szczeblu kobietom trudno jest wpływać na politykę. Silna i żywa demokracja jest możliwa tylko wtedy, gdy parlament w pełni uwzględnia reprezentowaną przez siebie populację. Parlamenty nie mogą jednak uważać się za inkluzywne, dopóki nie będą mogły pochwalić się pełnym udziałem kobiet. Nie chodzi tu tylko o prawo kobiet do równości i ich wkład w prowadzenie spraw publicznych, ale także o wykorzystanie zasobów i potencjału kobiet do określania priorytetów politycznych i rozwojowych, które przynoszą korzyści społeczeństwu i społeczności globalnej.



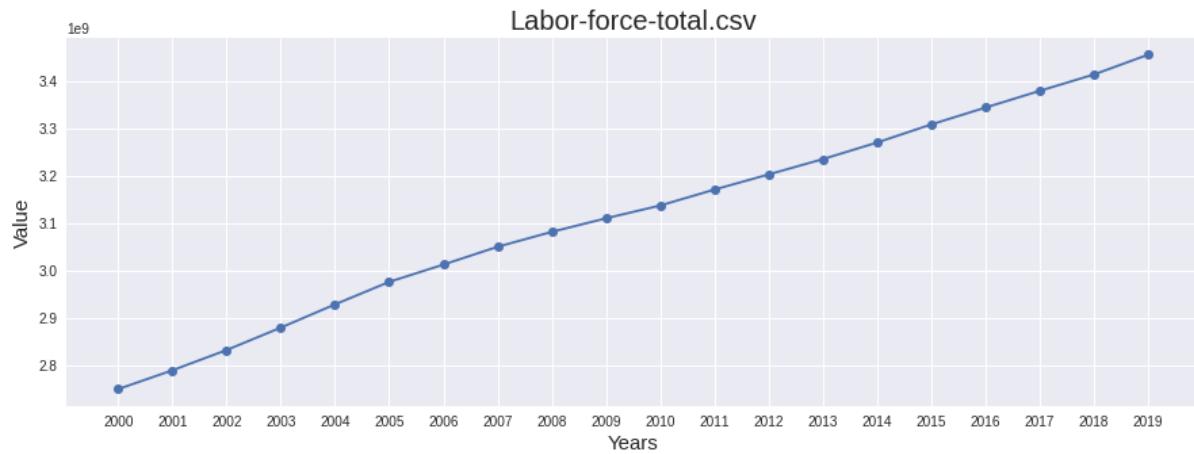
Zamiana na zmienną dynamiczną:



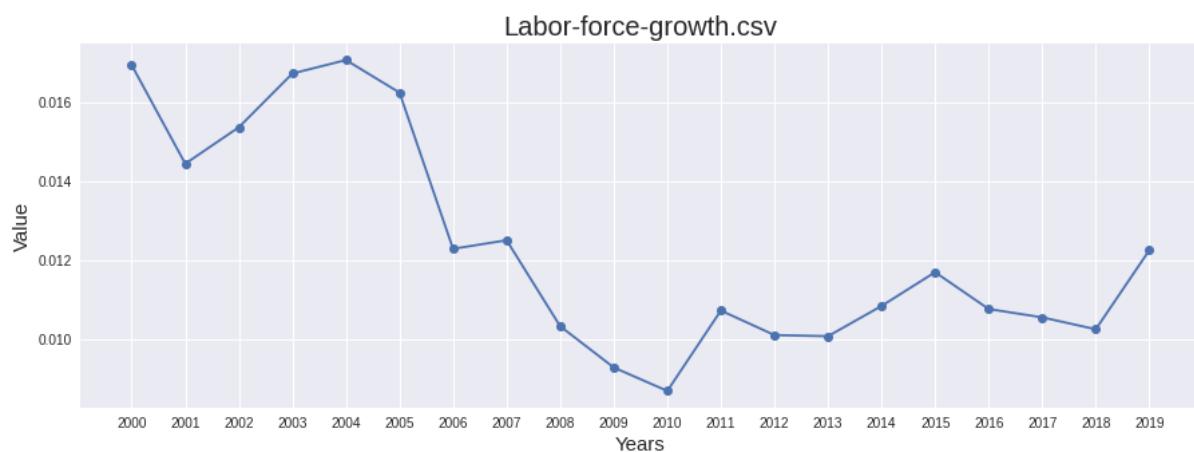
## 2.2.2.16 Labor force, total - Sia robocza

Definicja: Siła robocza obejmuje ludzi powyżej piętnastego roku życia, którzy zajmują się produkcją dóbr, dostarczaniem usług oraz osoby aktywnie poszukujące pracy w tych sektorach.

Opis: Siła robocza to zasób ludzi w wieku produkcyjnym, zdolnych do pracy i gotowych do jej podjęcia na typowych warunkach istniejących w gospodarce. Z definicji tej wynika, że zasób siły roboczej jest składową ogółu liczby osób w wieku produkcyjnym w danej gospodarce. W obecnej, dynamicznej gospodarce rzadko zdarza się by w danej chwili wszystkie osoby, które tworzą siłę roboczą były aktywne zawodowo. Ci, którzy pracują przyczyniają się do tworzenia tak zwanej grupy zatrudnionych. Jednakże ci, którym nie udało się znaleźć pracy, zasilają szeregi grupy osób nieaktywnych zawodowo- bezrobotnych. Ponadto nie wszystkie osoby pracujące zostały uwzględnione, jako że studenci czy pracownicy bez wynagrodzenia są często pomijani. Dodatkowo, wiele państw nie wlicza do siły roboczej członków sił zbrojnych. Liczność osób aktywnie zawodowo zmienia się w przeciągu danego roku na skutek pracowników sezonowych, którzy kończą pracę, gdy zapotrzebowanie znika.



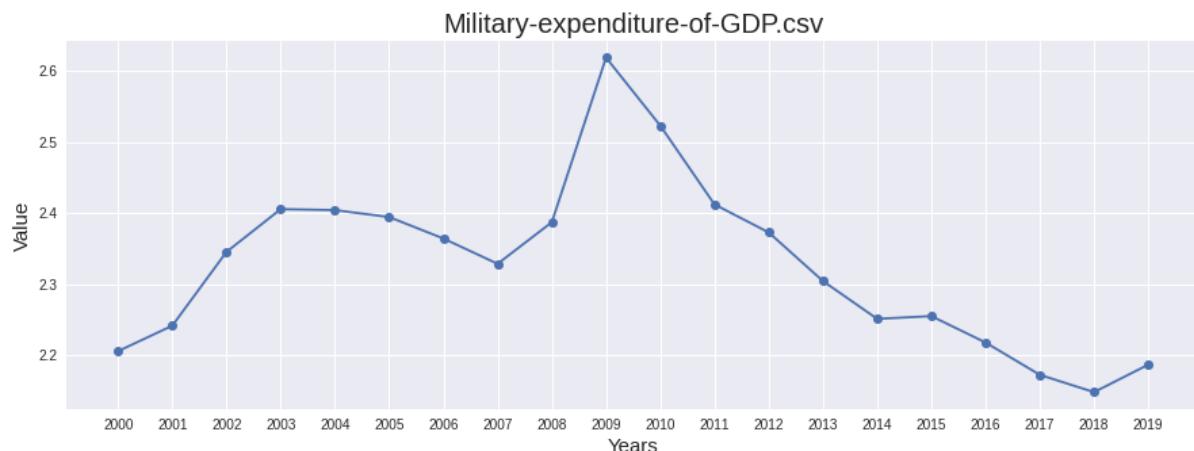
Zamiana na zmienną dynamiczną:



#### 2.2.2.17 Military expenditure (% of GDP) - wydatki wojskowe (% PKB)

Definicja: Wydatki wojskowe oznaczają wszystkie bieżące wydatki oraz nakłady inwestycyjne przekazane na wojsko w przeliczeniu na procent PKB.

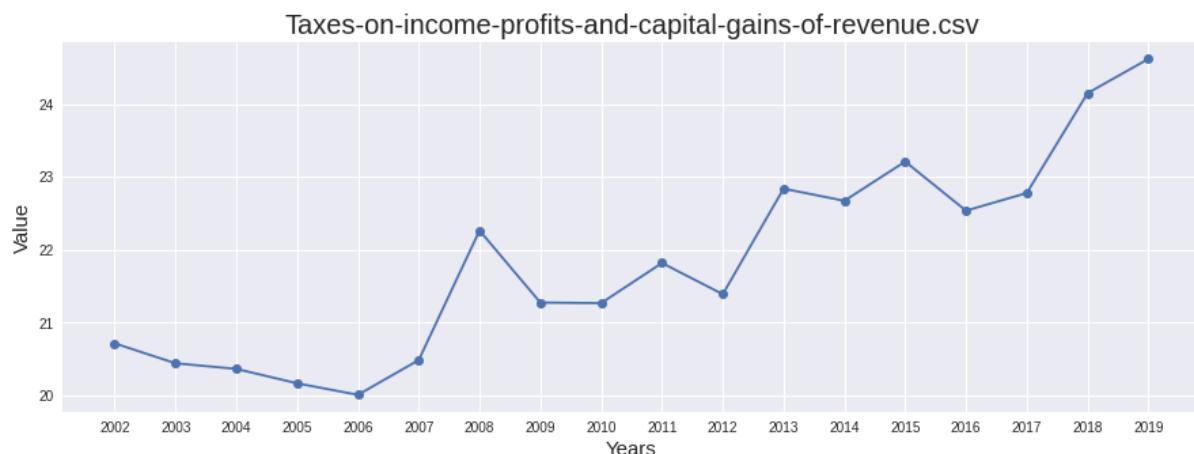
Opis: Wydatki wojskowe uwzględniają wszystkie bieżące wydatki oraz nakłady inwestycyjne przekazane na siły pokojowe, ministerstwa obrony oraz inne rządowe agencje zaangażowane w przedsięwzięcia obronne takie jak: siły paramilitarne i wojskowe działania kosmiczne. Wydatki te obejmują pieniądze przekazywane zarówno na personel wojskowy jak zwykłych obywateli, w skład których wchodzą emerytury, świadczenia socjalne, wyżywienie żołnierzy, pomoc medyczna, zaopatrzenia, badania wojskowe oraz rozwój wojska. Do wydatków wojskowych nie zostały włączone ochrona cywili oraz wydatki na zakończenie już konfliktów zbrojnych.



#### 2.2.2.18 Taxes on income, profits and capital gains (% of revenue) - podatki od dochodu (procent przychodu)

Definicja: Podatki od dochodu oznaczają pieniądze pobierane od rzeczywistego dochodu netto osoby fizycznej, zysku firm oraz zysku kapitałowego w przeliczeniu na procent PKB.

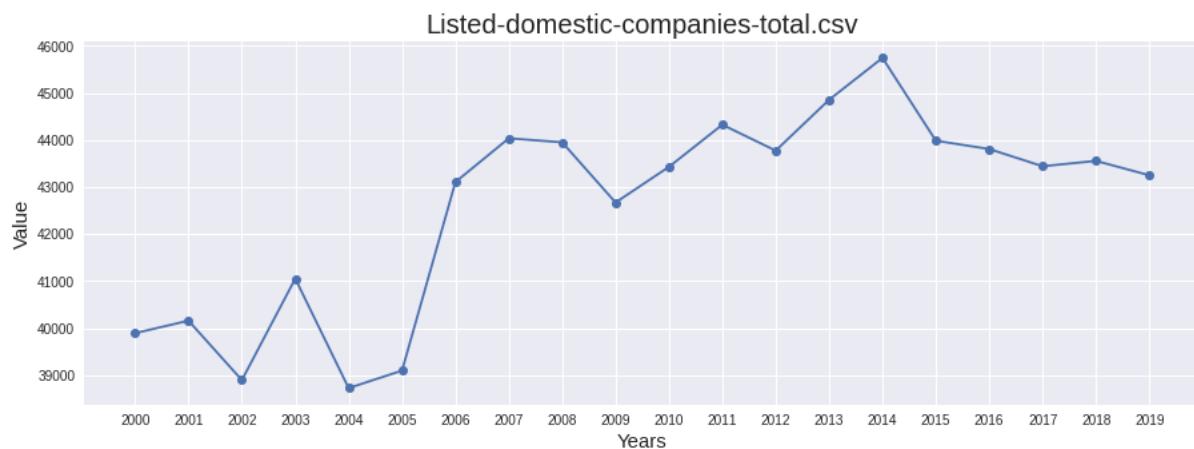
Opis: Metoda pobierania podatków od dochodu skupia się na wszystkich działaniach ekonomicznych mających wpływ na aktywa, pasywa, wynagrodzenia oraz wydatki, a nie tylko na samych przepływach gotówki. Metoda ta została zaproponowana przez IMF - International Monetary Fund we współpracy z SNA.



#### 2.2.2.19 Listed domestic companies, total - Krajowe spółki giełdowe

Definicja: Krajowe spółki giełdowe to spółki, które pod koniec danego roku posiadają akcje notowane na giełdzie.

Opis: Wielkość giełdy papierów wartościowych może być zmierzona w wiele sposobów, gdzie każdy z nich może wskazywać na różny ranking państw. Rozwój rynków finansowych jest znacząco powiązany z rozwojem gospodarczym. Dobrze funkcjonujące systemy finansowe dostarczają poprawnych i łatwo dostępnych informacji, które mogą znaczowo obniżyć opłaty za transakcje oraz zwiększyć wzrost gospodarczy. Dodatkowo, systemy bankowe oraz giełdy papierów wartościowych mają realny wpływ na zmniejszenie odsetka ubóstwa. W mniej rozwiniętych gospodarkach dominującą rolę odgrywają komercyjne banki. W bardziej rozwiniętych gospodarkach natomiast, giełdy są bardziej wydajne niż banki krajowe. Otwarte giełdy posiadające rozsądne zasady makroekonomii, odpowiednie systemy prawne oraz ochronę akcjonariuszy przyciągają kapitał, co za tym idzie posiadają większe rynki finansowe. Ostatnie badania na temat rozwoju giełdy pokazują, że współczesna środki komunikacji wraz z integracją rynków finansowych doprowadziły do większego transgranicznego przepływu kapitału czy silniejszej obecności firm na zagranicznych rynkach



#### 2.2.3 Redukcja danych

Opisane wyżej zbiory cechują się ogromną ilością danych, które nie są całkowicie spójne. W wielu przypadkach brakuje znaczającej liczby obserwacji, co stanowi problem w ich dalszej eksploatacji. Aby móc bezproblemowo badać podobieństwa między państwami i dokonać ich klasteryzacji postanowiono zredukować:

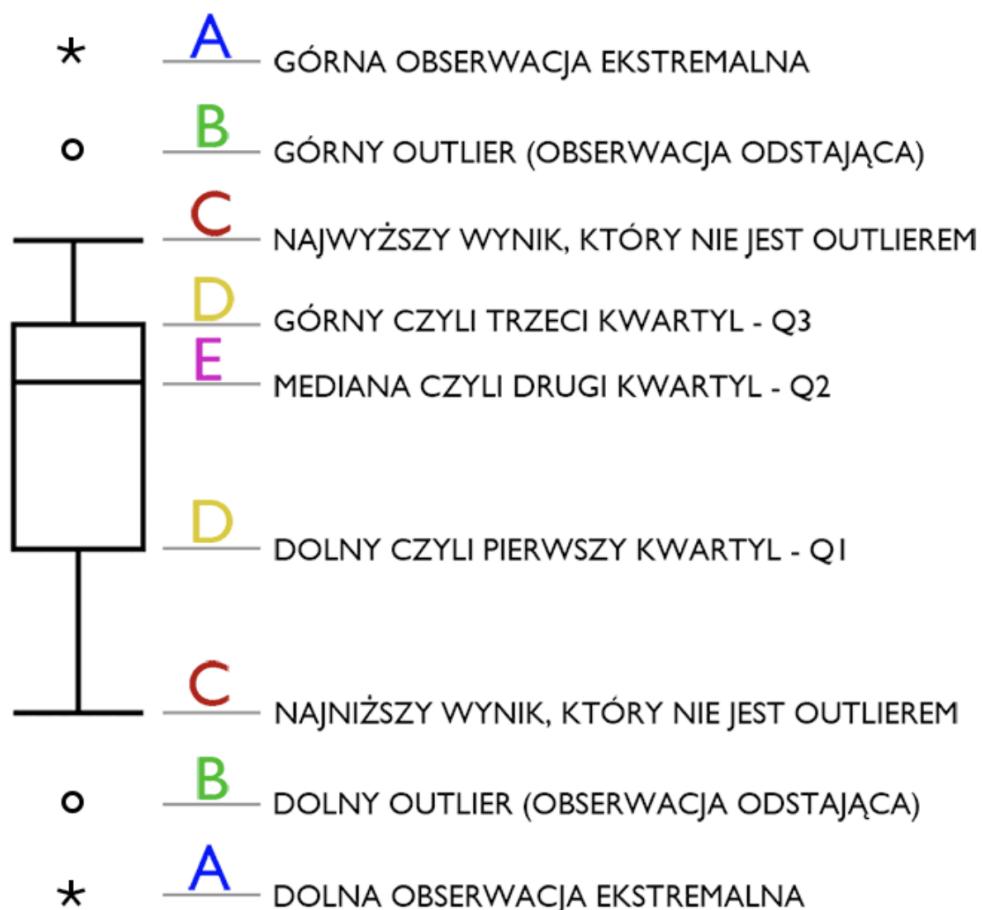
- **liczbę wskaźników** pozbywając się zbiorów "Literacy rate", "Taxes on Income" oraz "Listed Domestic Companies", które cechowały się dużą losowością, względem posiadanych informacji.
- **Ilość badanych państw** zmniejszając ich liczbę do 76 na skutek tego, iż dla większości państw brakowało istotnych wartości potrzebnych do przeprowadzenia analizy.
- **Przedział czasowy** skupiając się na latach 2000 - 2019, jako że ten okres skupiał największą gęstość danych.

- Pojedyncze **Wartości Nan**, które pozostały w zbiorze zastąpiono przy wykorzystaniu interpolacji liniowej. W interpolacji używana jest ostatnia ważna wartość przed brakiem danych oraz pierwsza ważna za brakiem. Jeśli pierwsza lub ostatnia obserwacja w szeregu zawiera brakujące wartości, wówczas używane są dwie najbliższe nie brakujące wartości na początku i na końcu serii.

### 3 Wstępna analiza danych

#### 3.1 Analiza podstawowych cech statystycznych za pomocą wykresów skrzynkowych

W celu dokonania pobiżnej analizy statystyk opisowych w zbiorach danych można posłużyć się wykresem skrzynkowym. Z uwagi na wartość informacyjną wykresów skrzynkowych są one często wykorzystywane w eksploracji. Z wykresów skrzynkowych można zauważać w takie cechy jak: mediana, minimalna wartość w zbiorze, maksymalna wartość w zbiorze, kwantyle rzędu 1/4, 3/4 - inaczej nazywane kwartylami, a także czy w danych występują obserwacje odstające. Bez potrzeby przeglądania, czasami gigantycznych tabel z podstawowymi statystykami opisowymi dzięki boxplot wystarczy szybki rzut oka na wykres. Dane można odczytać w taki sposób:

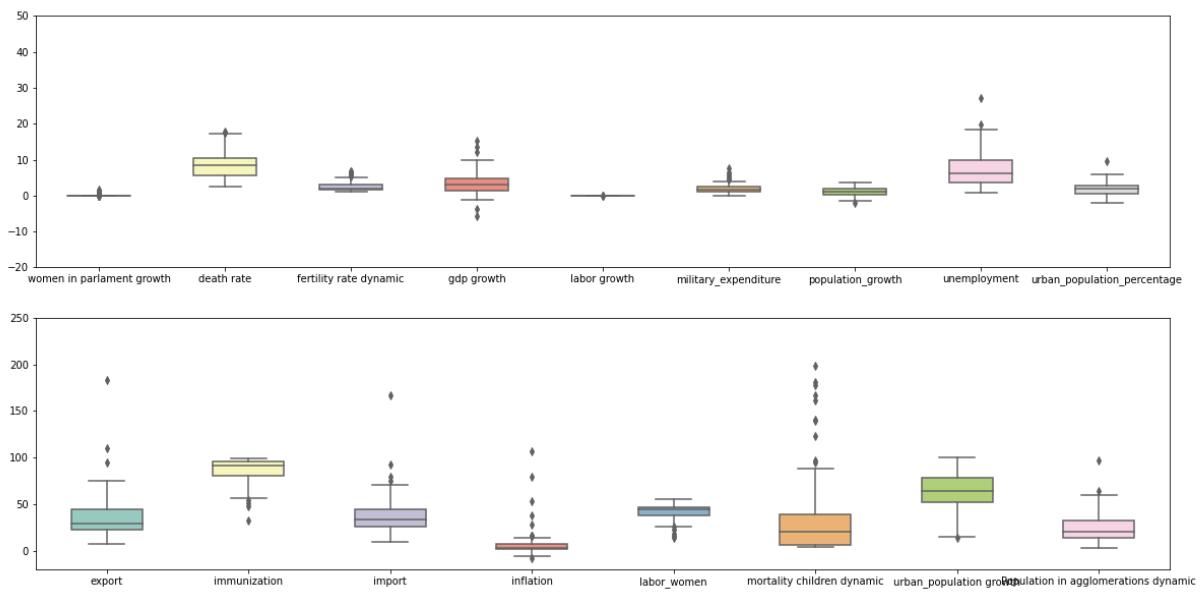


1

W celu lepszej wizualizacji, stworzono animacje wykresów skrzynkowych z każdego roku, aby móc dostrzec jak cechy zmieniały się w czasie.

<sup>1</sup> <https://pogotowiestatystyczne.pl/wykres-skrzynkowy-moc-informacji-na-jednym-rysunku/> dostęp 10.05.2022

Data from 2001



Z wykresu możemy dostrzec, że:

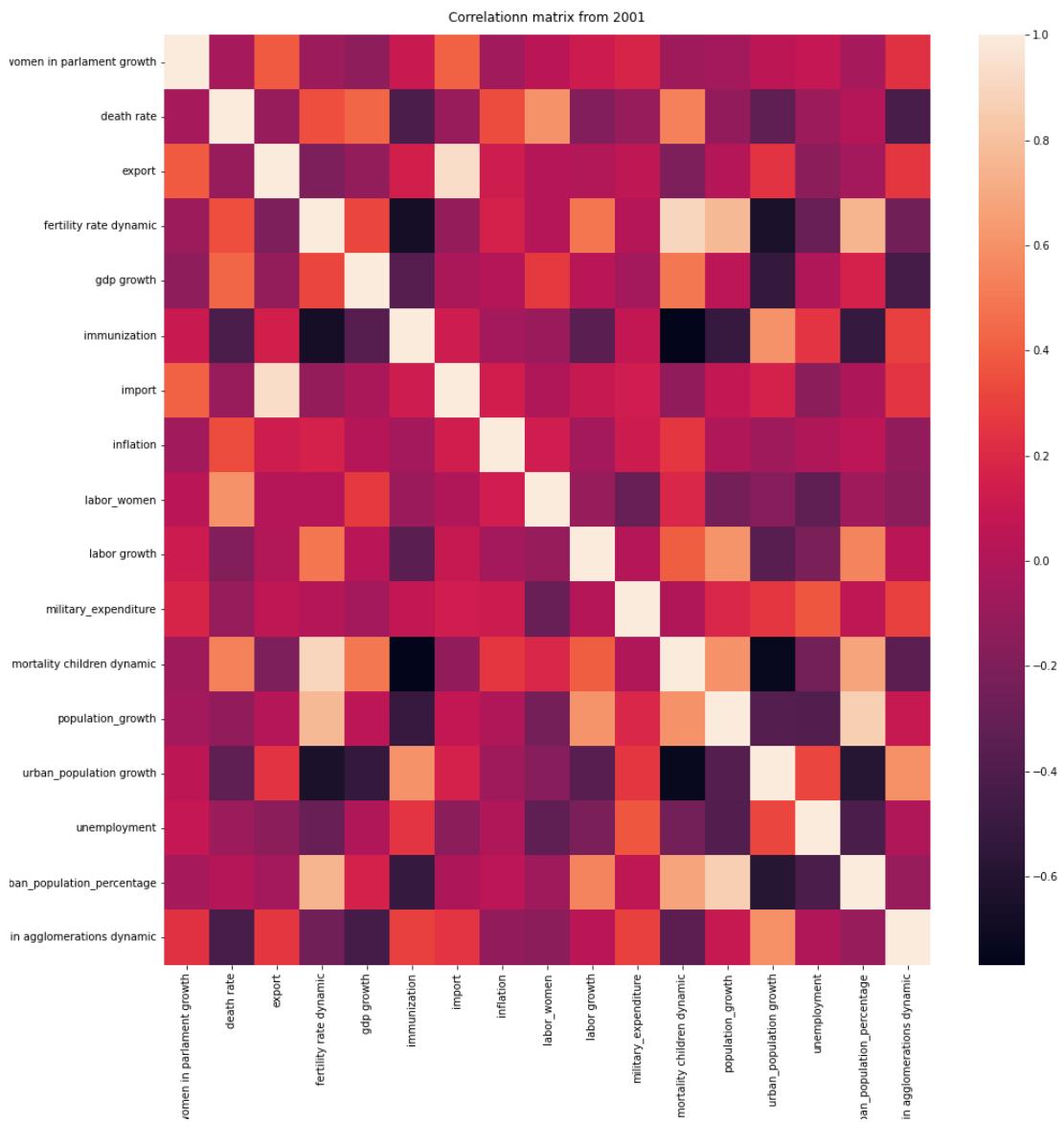
- najbardziej dynamicznie zmieniającą się w czasie zmienną była *gdp growth*;
- zmienne *Women in parliament growth* oraz *labor growth* mają wartości bardzo bliskie 0 i niewiele obserwacji odstających;
- wartości zmiennej *death rate* wraz z upływem czasu zmniejszają się i ich rozbieżność maleje, koncentrując się poniżej wartości 10;
- zmienność *fertility dynamic* również ma wartości bliskie 0, jednak można dostrzec też wartości odstające, które wraz z upływem czasu zbliżają się do 0;
- *military expenditure* nie ulega znacznym zmianom wraz z upływem lat, pojawia się kilka obserwacji odstających co może świadczyć o wzroście wydatków na wojsko w danych krajach;
- *population growth* - do 2012 roku można zauważać tendencję wzrostową, natomiast później następuje spadek w wartościach;
- *unemployment* - do 2008 roku można zauważać ogólny spadek wartości tej cechy w badanej grupie, natomiast po 2008 roku widać znaczną grupę obserwacji która ma ponadprzeciętnie wysoki współczynnik bezrobocia, następnie do 2019 roku te wartości zaczynają spadać;
- *urban\_population\_percentage* - wartości wraz z upływem lat koncentrują się wokół ok 2%;
- *export* - jest obserwacja znacznie odbiegająca od pozostałych, jej wartość praktycznie nie zbliża się do poziomu reszty obserwacji;
- *immunization* - odsetek zaszczepień wśród dzieci rośnie z roku na rok, w 2001 roku widać znaczną grupę obserwacji, która ma ten odsetek znacznie niższy niż reszta badanych państw, jednak można dostrzec tendencję zrównywania się tych wartości odstających z resztą grupy;
- *import* - podobnie jak przy zmiennej export, są tu obserwacje odstające znacznie przewyższające wartościami pozostałą grupę badanych państw;

- *inflation* - kolejna bardzo dynamicznie zmieniająca się w czasie zmienna, w różnych latach można dostrzec zarówno obserwacje ze skrajnie małą wartością inflacji jak i ze skrajnie dużą wartością inflacji;
- *labor\_women* - tendencja wzrostowa, jest grupa państw, która ma wartości tej cechy znacznie niższe niż pozostałe;
- *mortality\_children\_dynamic* - widać stopniowe zmniejszanie się wartości wśród badanych prób, również zmniejsza się ilość wartości odstających, które miały wartość tej zmiennej na wiele wyższym poziomie od pozostałych państw;
- *urban\_population\_growth* - stopniowe zwiększenie się wartości wraz z upływem lat, jednak nie są to duże zmiany
- *population\_in\_agglomerations\_dynamic* - stopniowe zwiększenie się wartości wraz z upływem lat, jednak nie są to duże zmiany, można dostrzec jedną obserwację ze znacznie wyższymi wartościami

## 3.2 Korelacja między zmiennymi

### 3.2.1 Macierz korelacji

Badania eksploracyjne danych w wielu przypadkach (czy to w data science, czy w nauce) wymagają sprawdzenia korelacji pomiędzy zmiennymi predykcyjnymi, jak również między zmiennymi predykcyjnymi a zmienną decyzyjną. W badaniu ze względu na cel klasteryzacji, do macierzy korelacji wzięto pod uwagę wszystkie cechu charakteryzujące państwa z każdego roku.



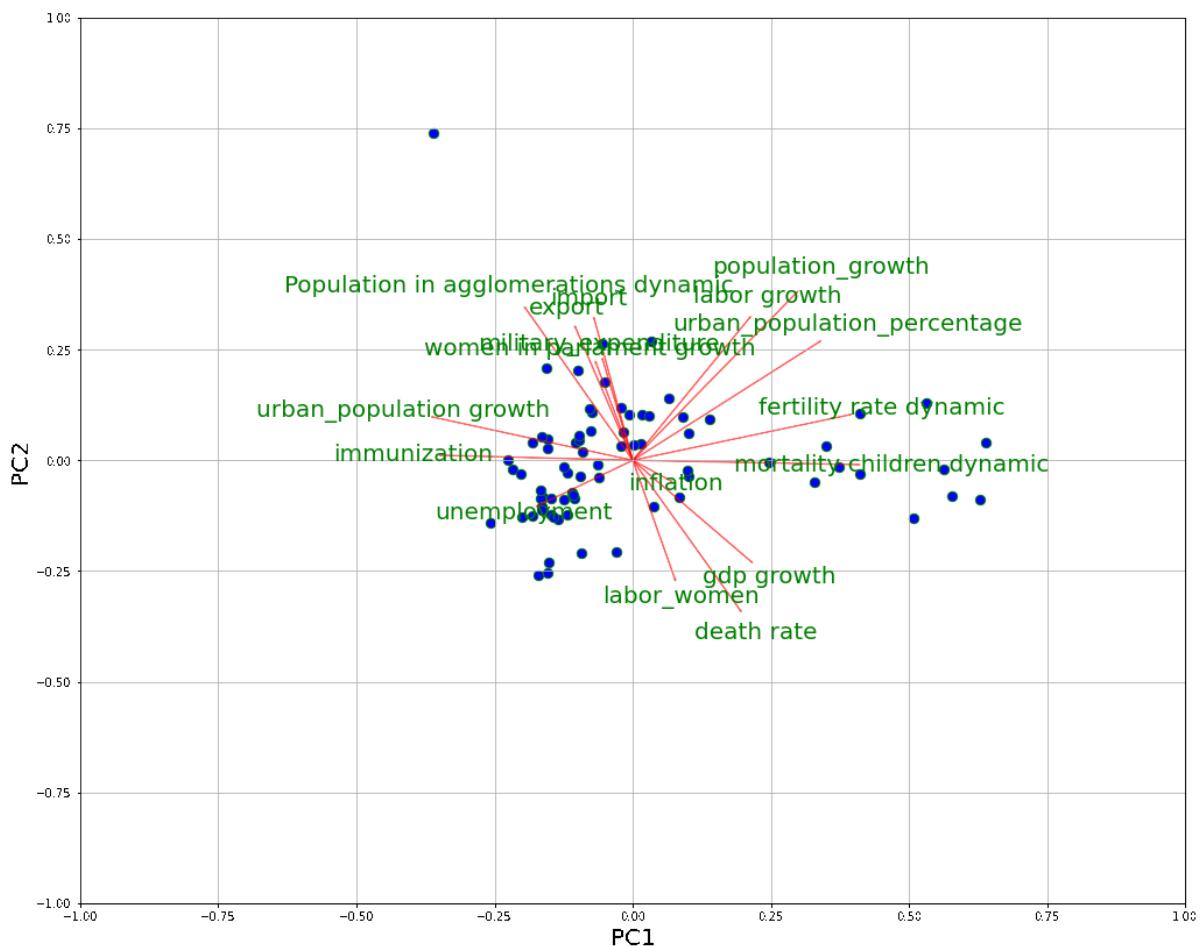
Z powyższej animacji da się zauważyc poniższe wnioski:

- jak się można było spodziewać zmienna `death_rate` jest ujemnie skorelowana z `population_growth`, co oznacza, że wraz ze wzrostem/spadkiem jednej zmiennej, druga zmienna zachowuje się odwrotnie i odpowiednio spada bądź rośnie. Zmienna `population_growth` jest za to dodatnio silnie skorelowana z `fertility_rate_dynamic`, `urban_population_percentage`, `labor_growth` i `mortality_children_dynamic`. Do dalszego badania można się pozbyć zmiennej `population_growth`.
- eksport jest silnie dodatnio skorelowany z importem, więc można wziąć pod uwagę jedynie jedną z nich
- Zmienne `Fertility_rate`, `mortality_children` i `urban_population` są nazwajem silnie skolerowane. Największe wartości współczynnika mają zmienne z cechą `urban_population`, dlatego to z niej należy zrezygnować. Również warto się zastanowić nad usunięciem zmiennej `mortality_children`.

- Korelacja wskaźnika PKB ze wskaźnikiem śmiertelności na przedziale 20 lat zmniejszyła się z 0,43 do 0,06.
- Korelacja wskaźnika dzietności ze wskaźnikiem śmiertelności na przedziale 20 lat zmniejszyła się z 0,35 do -0,24. Co oznacza, że 20 lat temu wartości obu zmiennych zmieniały się w tym samym kierunku, jednak w 2019 roku sytuacja stała się odwrotna. Jeżeli więc jedna wartość rośnie, to druga spada i na odwrót.

### 3.2.2 Biplots

Biplot to rodzaj wykresu na którym prezentowane są dwa rodzaje informacji. PCA biplot łączy score plot (pokazujący klastry punktów podobnych) oraz loading plot (pokazujący jak mocno każda z cech wpływa na główne składowe). Z takiego wykresu odczytać możemy również jak dane cechy są ze sobą skorelowane. Jeżeli dwa wektory są blisko siebie, oznacza to, że odpowiadające im cechy są pozytywnie skorelowane. Negatywną korelację możemy zauważać, jeżeli wektory tworzą kąt bliski 180 stopni. Natomiast, jeśli dwa wektory tworzą kąt prosty, to możemy stwierdzić, że odpowiadające im cechy nie są ze sobą skorelowane.



Z animacji możemy wysunąć wnioski pokrywające się z tymi wyciągniętymi z analizy macierzy korelacji. Mianowicie:

- Silnie dodatnio skorelowanymi zmiennymi są: import wraz z eksportem oraz labor\_women z death\_rate, co oznacza, że do dalszej analizy można zastąpić je jedną zmienną. Dodatkowo pozytywną korelację można również zauważać między GDP\_growth, a Urban\_population\_percentage
- Silnie ujemnie skorelowanymi zmiennymi są: immunization oraz mortality\_children, co można było podejrzewać, jako że im więcej dzieci jest szczepionych, tym mniejsza jest wśród nich śmiertelność. Co więcej kolejnymi ujemnie skorelowanymi zmiennymi są mortality\_children wraz z urban\_population\_growth
- Możemy również zauważać, że niektóre zmienne nie są ze sobą skorelowane np. fertility rate z GDP\_growth

### 3.2.3 Redukcja zmiennych

Po przeanalizowaniu powyższych metod należy zredukować zmienne, które wskazują na najwyższy stopień korelacji z innymi. Usunięte zostały trzy zmienne. Mianowicie:

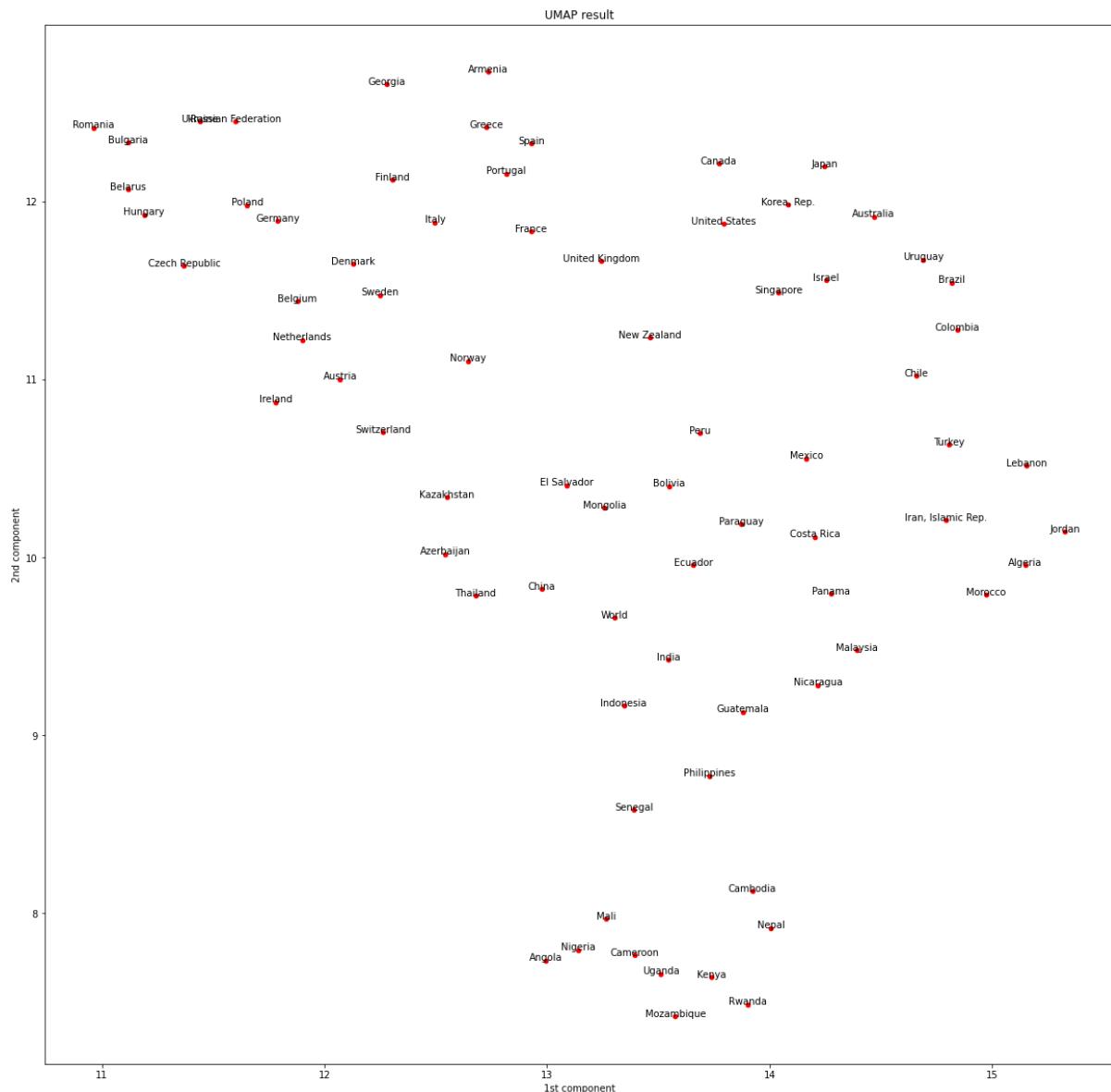
- Population\_growth, które wskazywało na silnie ujemną korelację ze zmienną Death\_rate
- Export, jako że jest bardzo zależny od importu.
- Mortality\_children, które wskazywało na silny związek ze zmiennymi immunization oraz urban\_population.

Do dalszych analiz wykorzystane zostaną dane pomniejszone o te trzy zmienne.

### 3.2.4 Wizualizacja cech zrzutowanych na dwuwymiarową przestrzeń w celu znalezienia państw podobnych

Wymiarowość danych oznacza liczbę cech reprezentujących dany zbiór. Żyjemy w trójwymiarowej rzeczywistości, co sprawia że nasz mózg bardzo dobrze się w niej odnajduje. Natomiast wszystko, co wykracza ponad trzy wymiary ciężko jest nam sobie wyobrazić. W takim razie, aby móc zwizualizować kilku wymiarowy zbiór danych należy zastosować redukcję wymiarów.

Redukcja wymiarów oznacza proces przekształcenia pierwotnego zbioru danych w zbiór o mniejszej liczbie wymiarów zachowując przy tym możliwie najwięcej informacji, które te dane ze sobą niosą. W celu łatwego i czytelnego zwizualizowania grup państw podobnych zastosujemy redukcję wymiarów do dwóch przy użyciu metody UMAP.



Na załączonym obrazku możemy zauważać przybliżoną “klasteryzację” państw według analizy wypisanych wyżej cech. Widocznym jest, że utworzyły się grupy państw podobnych.

W lewym górnym rogu w większości znajdują się państwa europejskie, gdzie kraje takie jak Ukraina, Rumunia, Bułgaria, Rosja i Białoruś są bardzo blisko siebie. Wynika to m.in. z bardzo podobnego ustroju politycznego danych państw. Kraje śródziemnomorskie (Hiszpania, Włochy, Portugalia, Grecja), których znaczny dochód pochodzi z turystyki również znajdują się blisko siebie. Kolejną ciekawą obserwacją wewnątrz krajów europejskich jest skupisko państw skandynawskich, które również są bardzo blisko siebie. Państwami do nich podobnymi okazały się Austria, Belgia, Holandia oraz Irlandia.

Dodatkowa informacja płynąca z niniejszego obrazka, to widoczne bardzo małe odległości między krajami takimi jak Kanada, USA, Korea pld, Japonia oraz Australia. Takie obserwacje potwierdzają wiarygodność analizy, jako że państwa te mają wiele cech wspólnych w rzeczywistości.

Na dole wykresu znajdują się ubogie kraje Afrykańskie takie jak: Kenya, Uganda, Nigeria czy Mozambique. Nad nimi plasują się Filipiny, Senegal, Indonesia czy Guatemała tworzące kolejne skupisko. W centrum znajduje się obserwacja world, która składa się z danych zebranych z całego świata. Możemy odczytać jej podobieństwo z Chinami oraz Indiami, co może być spowodowane wysokim zaludnieniem obu państw, przez co znaczna część statystyk dla obserwacji world jest od nich zależna.

## 4 Metoda klasteryzacji k-średnich

### 4.1 Teoria

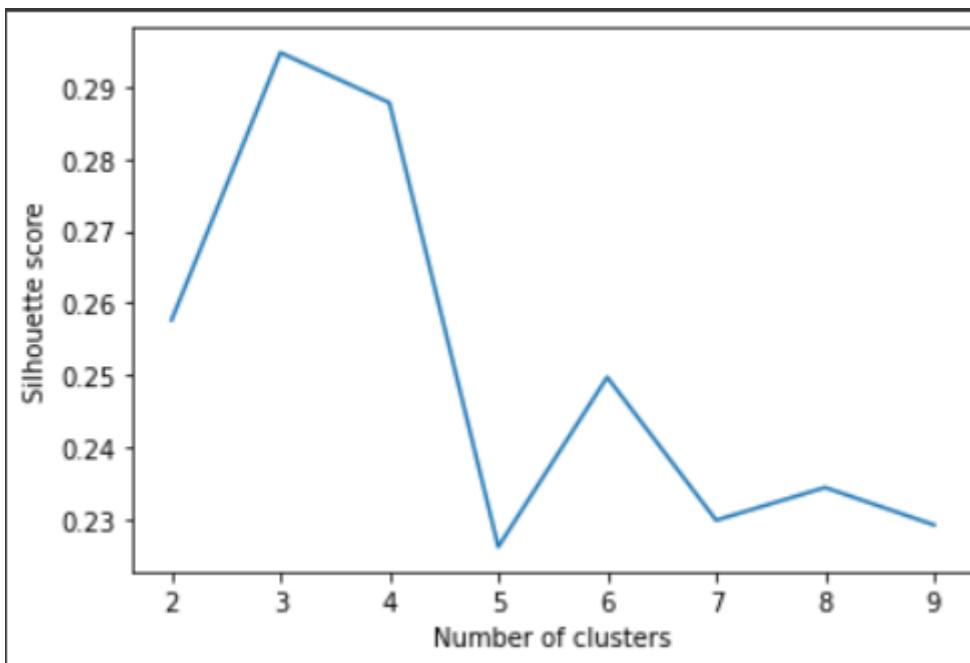
Algorytm k-średnich (z ang. k-means) inaczej zwany również algorytmem centroidów, służy do podziału danych wejściowych na z góry założoną liczbę klas. Jest to jeden z algorytmów stosowany w klasteryzacji (grupowaniu) i jest częścią uczenia nienadzorowanego w Machine Learning. Metoda k-średnich jest metodą należącą do grupy algorytmów analizy skupień tj. analizy polegającej na szukaniu i wyodrębnianiu grup obiektów podobnych (skupień). Reprezentuje ona grupę algorytmów niehierarchicznych. Główną różnicą pomiędzy niehierarchicznymi i hierarchicznymi algorytmami jest konieczność wcześniejszego podania ilości skupień.

Przy pomocy metody k-średnich zostanie utworzonych k różnych możliwie odmiennych skupień. Algorytm ten polega na przenoszeniu obiektów ze skupienia do skupienia tak długo aż zostaną zoptymalizowane zmienności wewnętrz skupień oraz pomiędzy skupieniami. Oczywistym jest, iż podobieństwo w skupieniu powinno być jak największe, zaś osobne skupienia powinny się maksymalnie od siebie różnić.

### 4.2 Klasteryzacja zbioru danych

#### 4.2.1 Liczba skupień

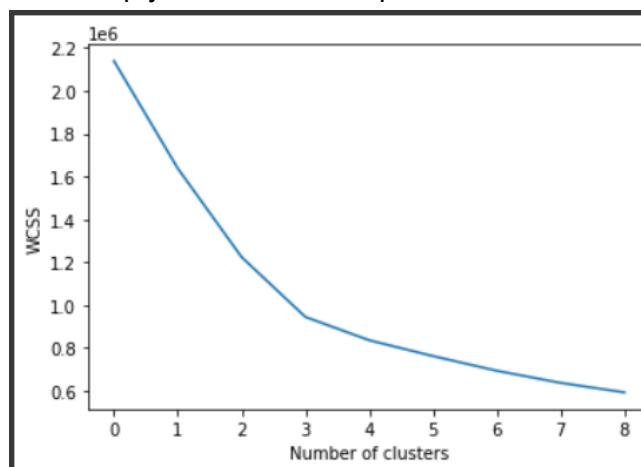
Jedną z metod ustalenia ilości skupień jest umowny jej wybór i ewentualna późniejsza zmiana tej liczby w celu uzyskania lepszych wyników. Wybór liczby skupień może być oparty również na wynikach innych analiz. W badaniu obliczono Silhouette Score dla różnych liczb klastrów.



Po analizie powyższego wykresu zdecydowano na liczbę klastrów równą 3, ze względu na największą wartość. Można powiedzieć, że klastry są od siebie bardzo oddalone, ponieważ wynik silhouette score jest najbliższy 1.

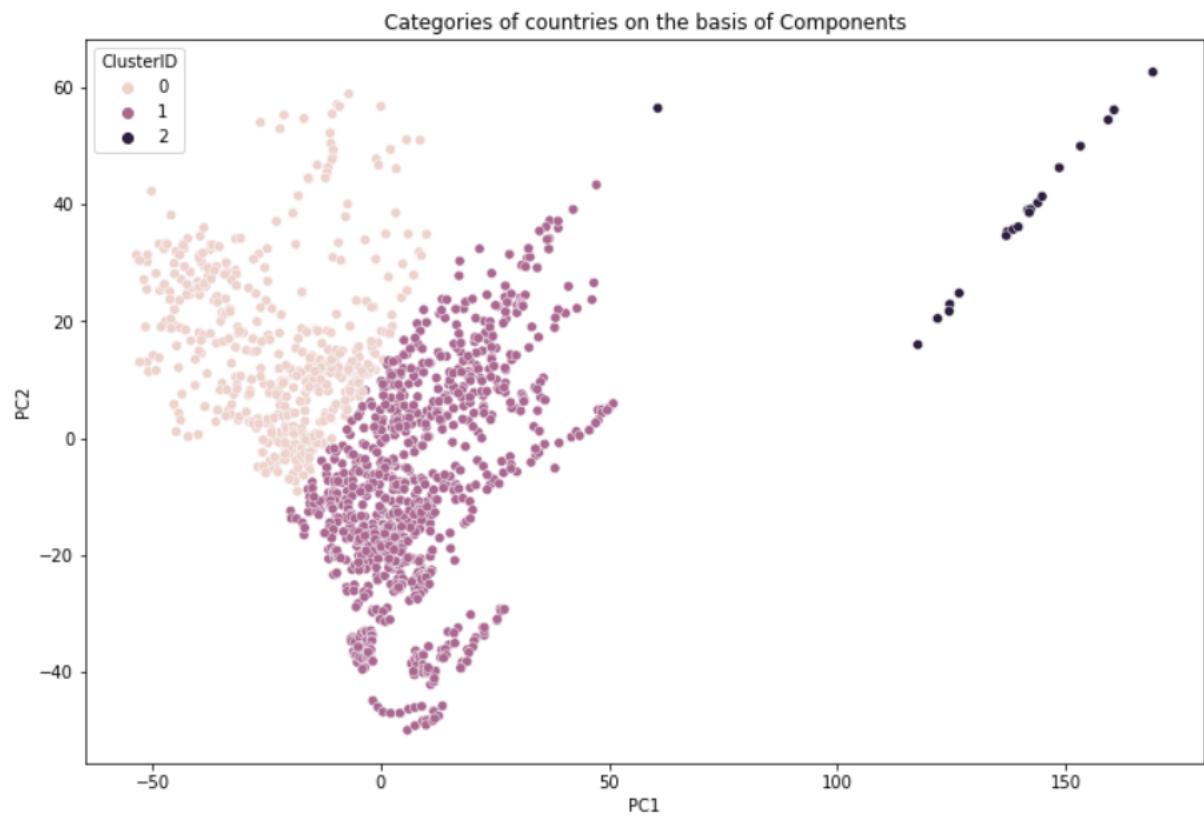
#### 4.2.2 Metoda elbow

W metodzie Elbow w rzeczywistości zmieniamy liczbę skupień i dla każdej wartości K obliczamy WCSS. WCSS to suma kwadratów odległości między każdym punktem a centroidem w klastrze. Kiedy wykreślmy WCSS z wartością K, wykres wygląda jak łokieć. Wraz ze wzrostem liczby klastrów wartość WCSS zacznie spadać. Analizując poniższy wykres, widzimy, że wykres gwałtownie zmieni się w punkcie dla  $k = 3$ , tworząc w ten sposób kształt łokcia. Od tego momentu wykres zaczyna się poruszać prawie równolegle do osi X. Wartość K równa 3 to optymalna liczba skupień.



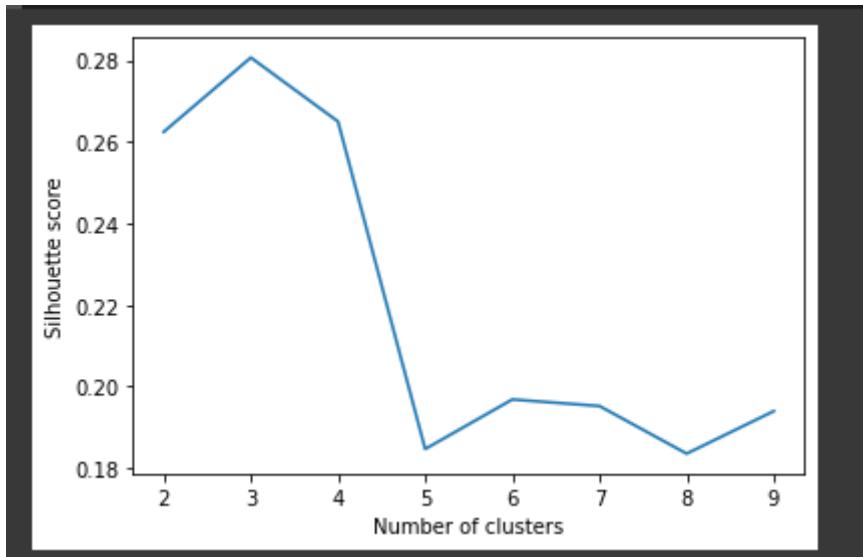
#### 4.2.3 Wizualizacja k-means

Do zwizualizowania metody k-means potrzebna jest redukcja wymiarów, dlatego skorzystano z PCA. Na poniższym wykresie wyraźnie wydzielili się 3 klastry, co ciekawe do trzeciego klastra przypisano mało obserwacji.



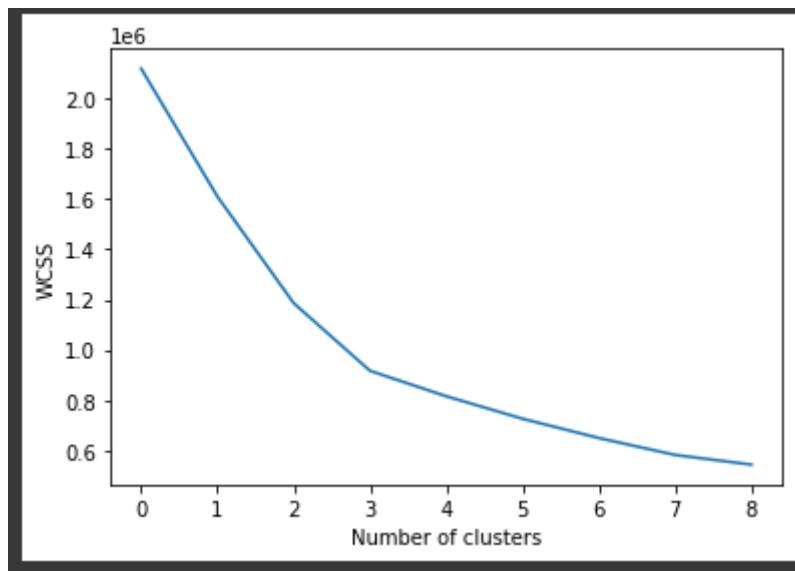
## 4.3 Klasteryzacja połączonego zbioru danych ze wszystkich lat

### 4.3.1 Liczba skupień



Po analizie powyższego wykresu zdecydowano na liczbę klastrów równą 3, ze względu na największą wartość. Można powiedzieć, że klastry są od siebie bardzo oddalone, ponieważ wynik silhouette score jest najbliższy 1.

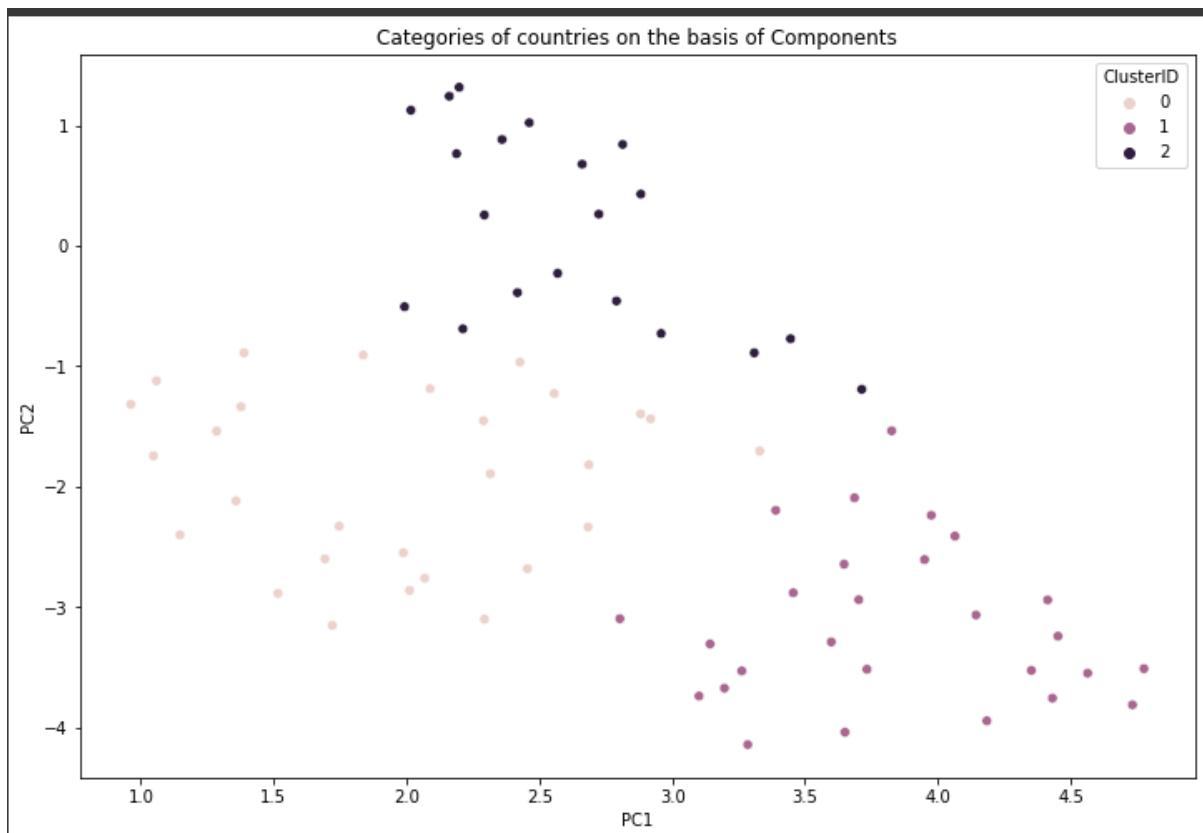
### 4.3.2 Metoda elbow



Metoda elbow sugeruje zastosowanie K=3.

### 4.3.3 Wizualizacja k-means

W tym przypadku do wizualizacji skorzystano z Umapa. Na poniższym wykresie również wyraźnie wydzielili się dość równomiernie 3 klastry.



#### 4.3.4 Analiza utworzonych klastrów państw

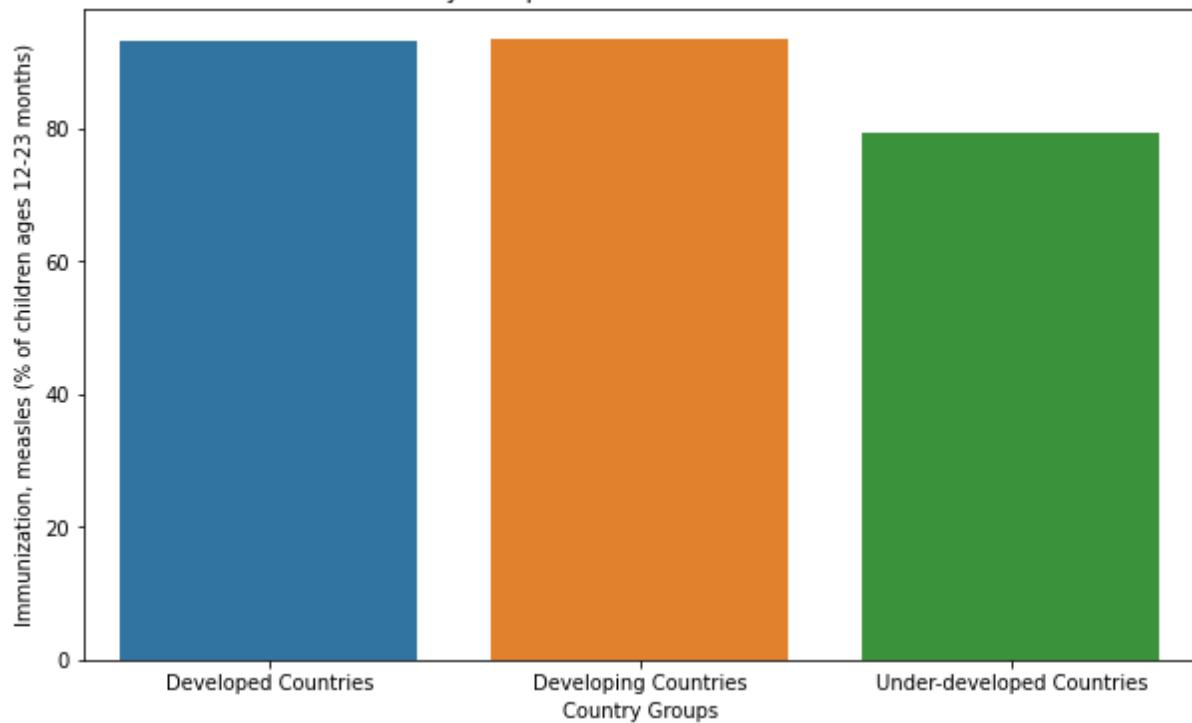
Obliczenie średniej wybranych zmiennych każdego klastra:

	immunization	military_expenditure	unemployment
ClusterID			
0	93.035714	2.230183	7.062947
1	93.415205	1.707525	8.377080
2	79.343912	1.633902	4.084247

Dokonano analizy poprzez porównanie, w jaki sposób powyższe wartości różnią się dla każdego klastra krajów, aby rozpoznać i odróżnić klastry krajów rozwiniętych od klastrów krajów słabo rozwiniętych.

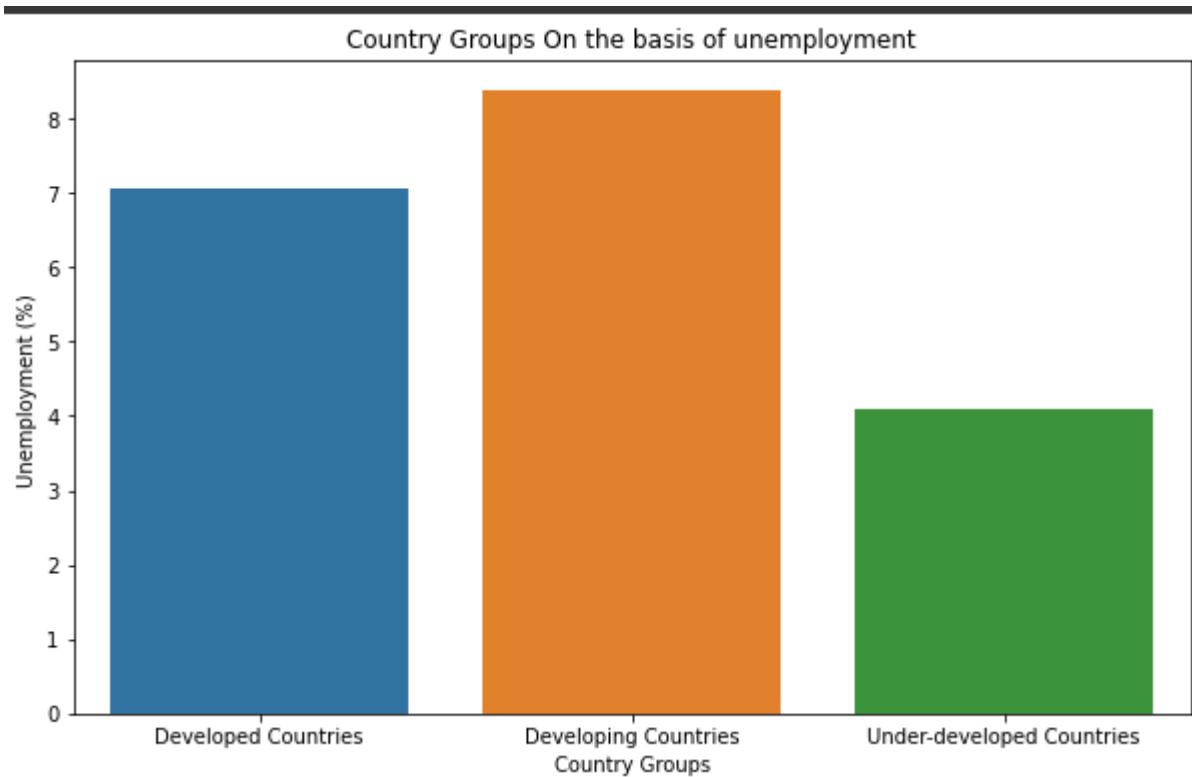
Z powyższego zestawu danych średnich wywnioskowano, że klaster 2 to grupa krajów średnie słabo rozwiniętych. Klaster 1 to kraje, wciąż rozwijające się a 0 kraje rozwinięte.

Country Groups On the basis of immunization

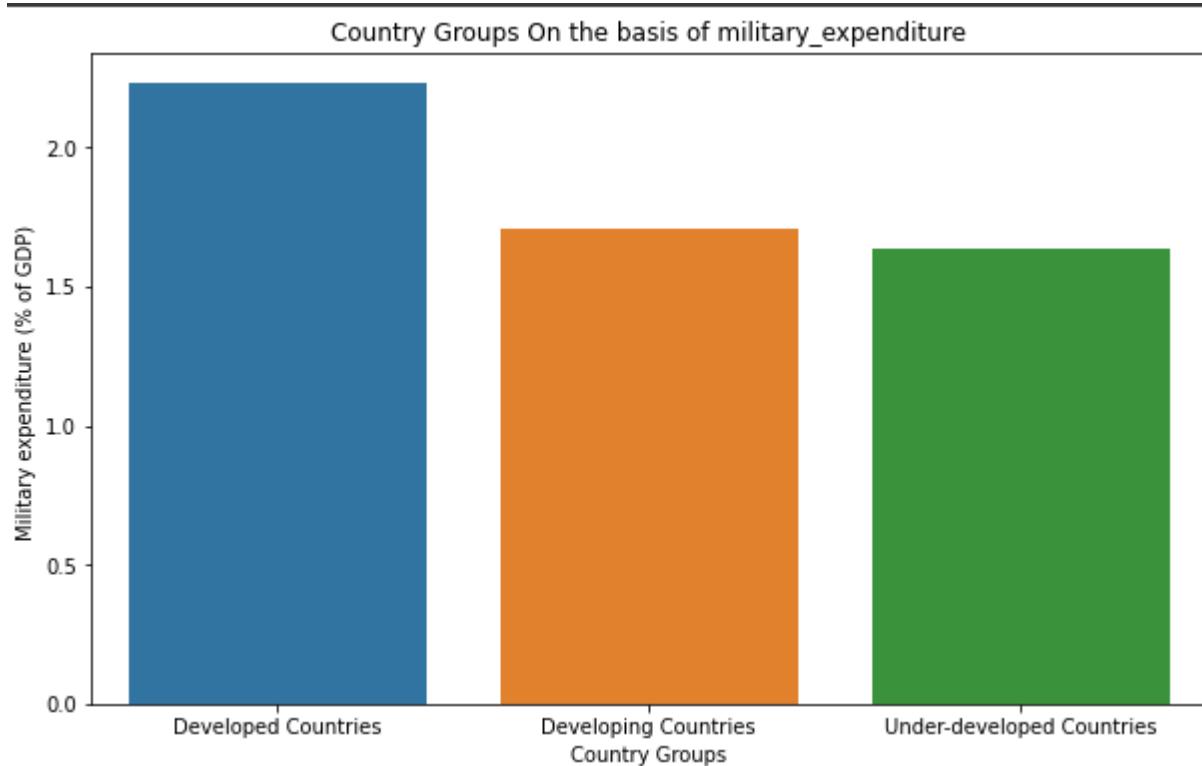


Powyższy wykres słupkowy pokazuje, że procent szczepień w krajach rozwiniętych jak i nadal rozwijających się jest wysoki około 93%. Zaś w krajach nierożwiniętych średnia wynosi już mniej 79%.

Country Groups On the basis of unemployment

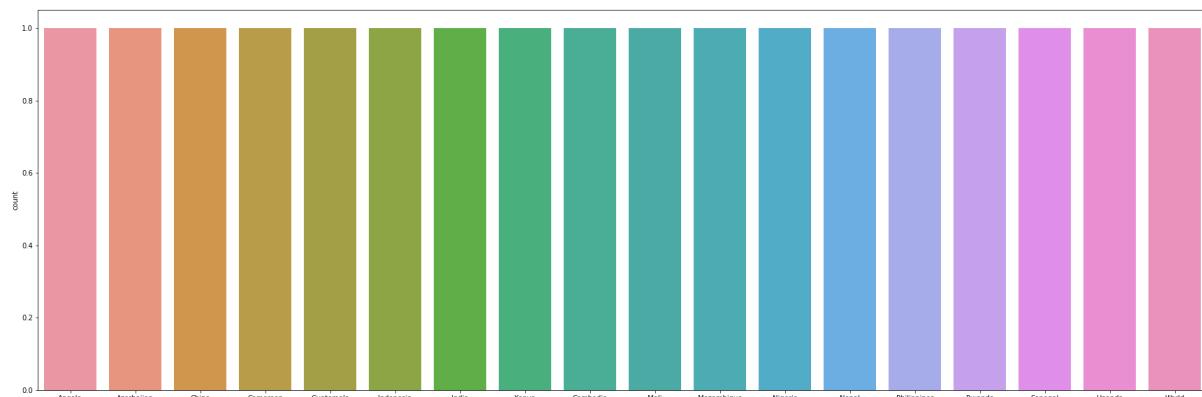


Odkryto, że średnie stopy bezrobocia są w rzeczywistości znacznie niższe w krajach biednych niż w krajach bogatych.



Jak można było się spodziewać w krajach rozwiniętych wydatki wojskowe są największe.

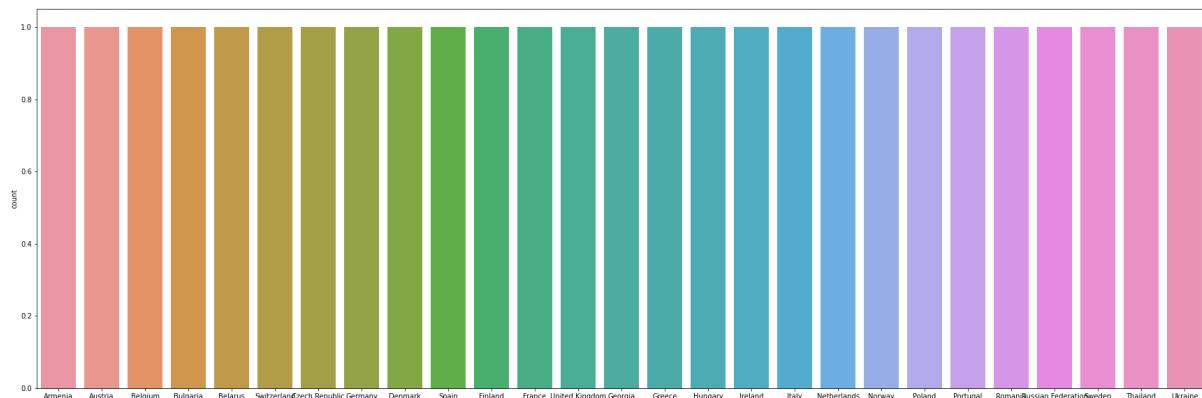
#### Klaster 0:



Państwa: Angola, Azerbejdżan, Chiny, Kamerun, Gwatemala, Indonezja, Indie, Kenia, Kambodża, Mali, Mozambik, Nigeria, Nepal, Filipiny, Rwanda, Senegal, Uganda i obserwacje dla świata.

Można zauważyc, że są to państwa głównie afrykańskie i azjatyckie.

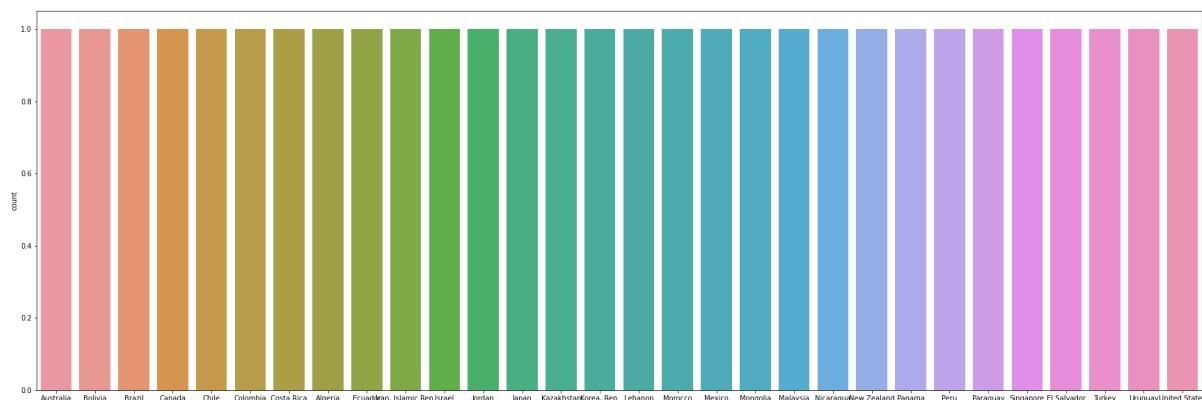
## Klaster 1:



Państwa: Armenia, Austria, Belgia, Bułgaria, Białoruś, Szwajcaria, Czechy, Niemcy, Dania, Hiszpania, Finlandia, Francja, Wielka Brytania, Gruzja, Grecja, Węgry, Irlandia, Włochy, Holandia, Norwegia, Polska, Portugalia, Rumunia, Rosja, Szwecja, Tajlandia i Ukraina.

W większości są to państwa Europejskie i kilka państw azjatyckich.

## Klaster 2:



Państwa: Australia, Boliwia, Brazylia, Kanada, Chile, Kolumbia, Kostaryka, Algieria, Ekwador, Iran, Izrael, Jordania, Japonia, Kazachstan, Korea Południowa, Lebanon, Maroko, Meksyk, Mongolia, Malezja, Nikaragua, Nowa Zelandia, Panama, Peru, Paragwaj, Singapur, Salwador, Turcja, Urugwaj, Stany Zjednoczone.

Wśród państw znajdują się światowe potęgi jak Stany Zjednoczone, Kanada, Australia, Japonia, Korea Południowa, także znajdują się tutaj państwa Ameryki Południowej.

# 5. Hierarchiczna analiza skupień

## 5.1 Teoria

Jedną z zalet algorytmów hierarchicznego grupowania jest możliwość generowania dendogramów (graficznego przedstawienia binarnej, hierarchicznej analizy skupień), pomagających w interpretowaniu wyników poprzez tworzenie istotnych taksonomii. Kolejną właściwością metod hierarchicznych jest brak konieczności odgórnego definiowania liczby skupień.

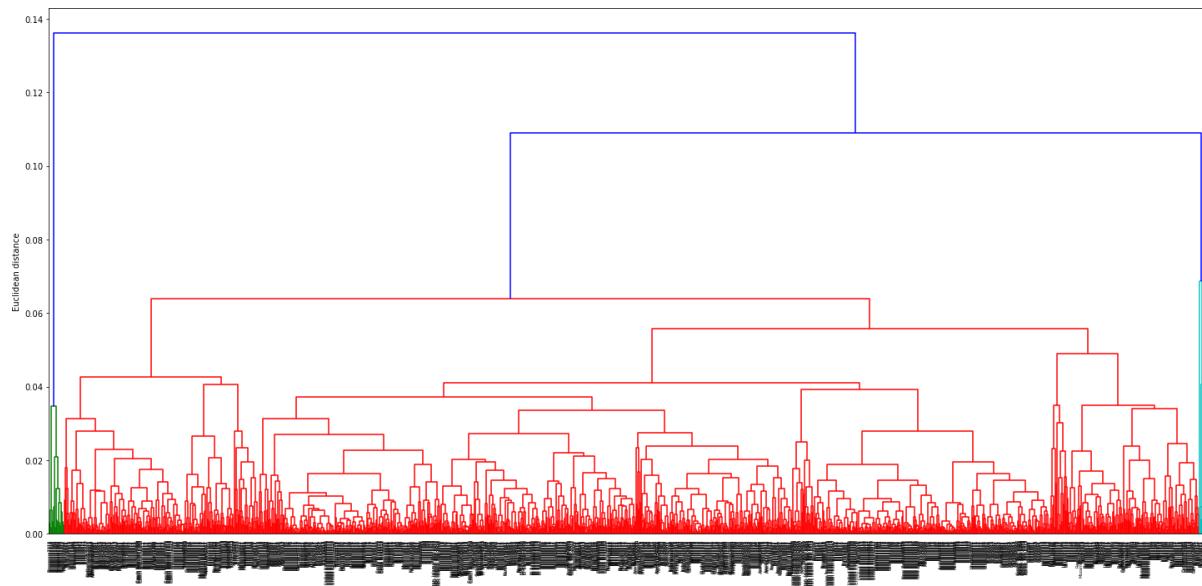
Dwema głównymi technikami wykorzystywanyimi w hierarchicznej analizie skupień są metody aglomeracyjne i deglomeracyjne. W klasteryzacji deglomeracyjnej zaczynamy od jednego klastra zawierającego w sobie wszystkie pozostałe skupienia i stopniowo dzielimy go na mniejsze grupy, aż do osiągnięcia stanu, w którym każdy klaster będzie zawierał tylko jedną próbkę. Klasteryzacja aglomeracyjna stanowi przeciwwieństwo grupowania deglomeracyjnego. W tym przypadku każda próbka stanowi osobne skupienie i dążymy dołączenia par najbliższych elementów, aż do uzyskania jednego wielkiego klastra.

Powszechnie stosowane algorytmy w aglomeracyjnej analizie skupień:

- algorytm pojedynczego wiązania (*ang. single linkage*) - wyszukujemy dwa punkty we wszystkich parach skupień(po jednym punkcie na każdy klaster), cechujących się najmniejszą odległością pomiędzy klastrami,
- algorytm pełnego wiązania (*ang. complete linkage*) - wyszukujemy dwa punkty we wszystkich parach skupień(po jednym punkcie na każdy klaster), cechujących się największą odległością pomiędzy klastrami,
- metoda średnich połączeń (*ang. average linkage*) - łączymy pary klastrów na podstawie minimalnych uśrednionych odległości pomiędzy wszystkimi elementami grupy w obydwu klastrach,
- metoda Warda (*ang. Ward's linkage*) - łączone są skupienia, dla których następuje najmniejszy wzrost wewnętrzgrupowej sumy kwadratów błędów.

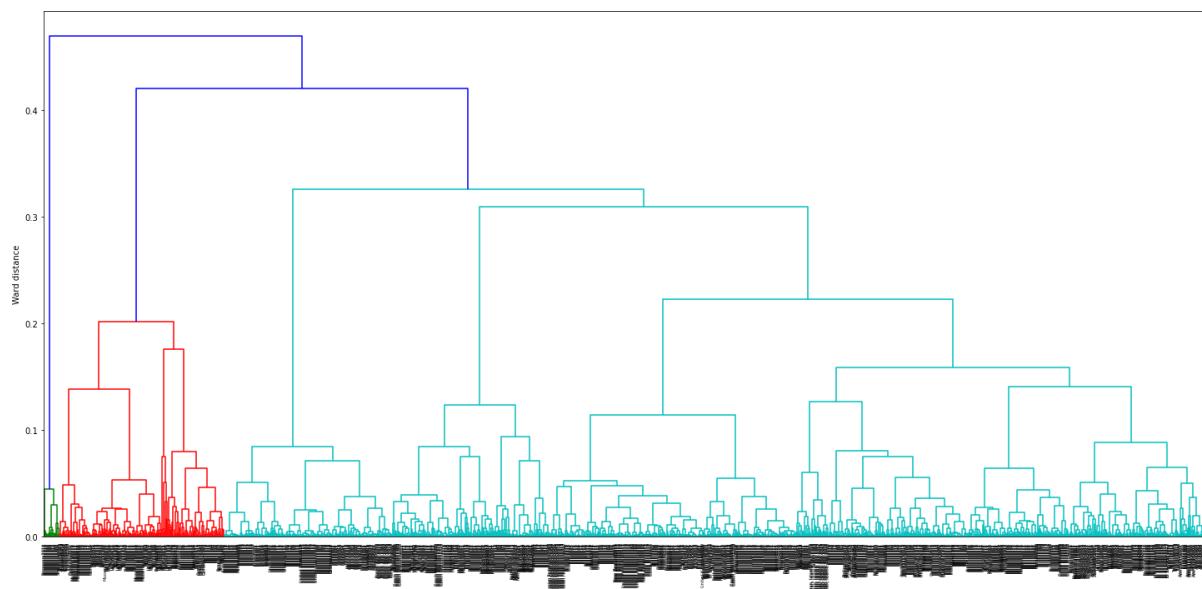
## 5.2 Dendrogramy

Przeprowadzono aglomeracje klastrów metodą pełnego wiązania za pomocą funkcji *linkage* (moduł *cluster.hierarchy*). Funkcja *linkage* zwraca macierz wiązania, którą można zwizualizować w postaci dendrogramu.



Na wykresie kolorami zaznaczono 3 różne klastry, zatem drzewo powinno być przycięte na wysokości 3 klastrów. Można zauważyć, że klastry nie są podobnej wielkości. Czerwony klaster jest największy i zajmuje większość zbioru, natomiast dwa pozostałe są bardzo małe. Zatem metoda pełnego wiązania może nie sprawdzić się na tych danych.

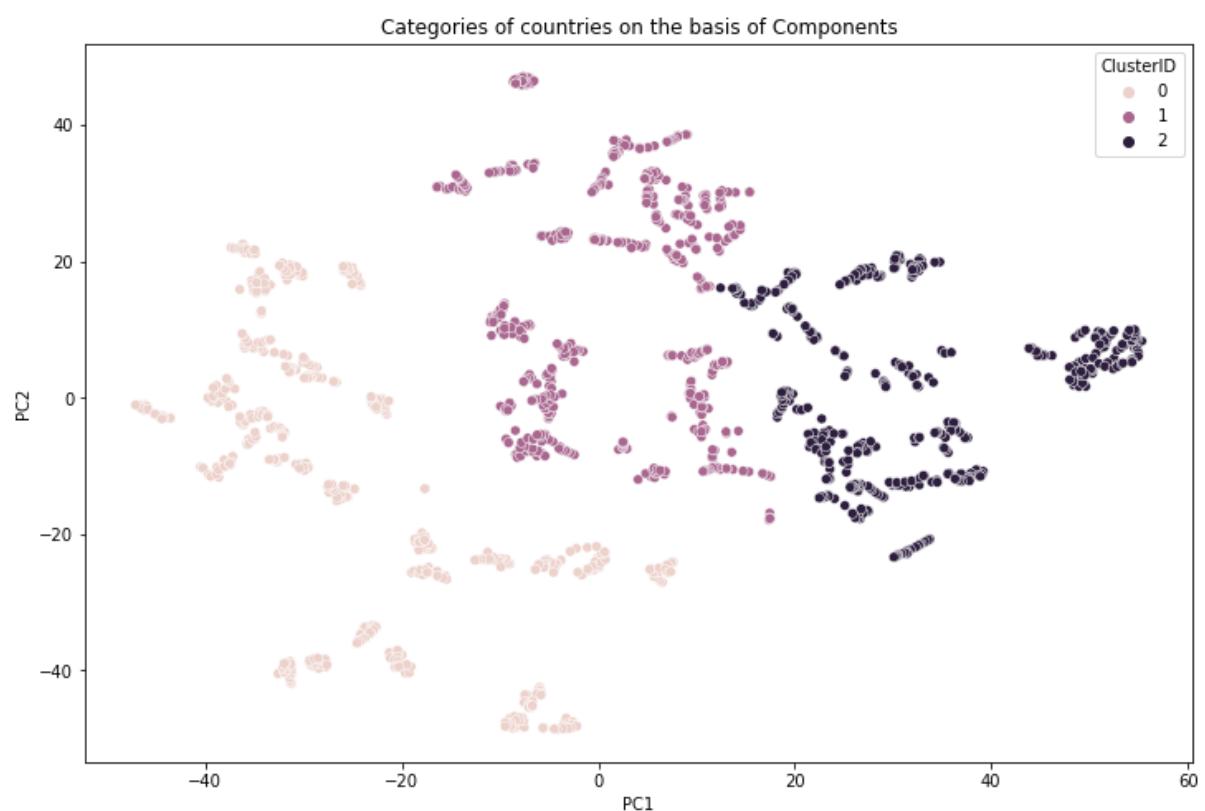
Przeprowadzono aglomeracje klastrów metodą Warda za pomocą funkcji *linkage* (moduł *cluster.hierarchy*)



Na wykresie kolorami zaznaczono 3 różne klastry, zatem drzewo powinno być przycięte na wysokości 3 klastrów. Dla metody Warda lepiej wygląda podział na klastry. Próbowano również innych metod i metryk, jednak wyniki nie były zadowalające.

### 5.3 Wizualizacja skupień AgglomerativeClustering

Wykorzystano teraz funkcje AgglomerativeClustering, w której można dobrać liczbę zwracanych skupień. Wybrano, aby funkcja zwracała 3 skupienia. Przed przystąpieniem do aglomeracyjnej hierarchicznej klasteryzacji dokonano standaryzacji danych, jak i zmniejszenie wymiarowości do 2 wymiarów za pomocą metody t-SNE.

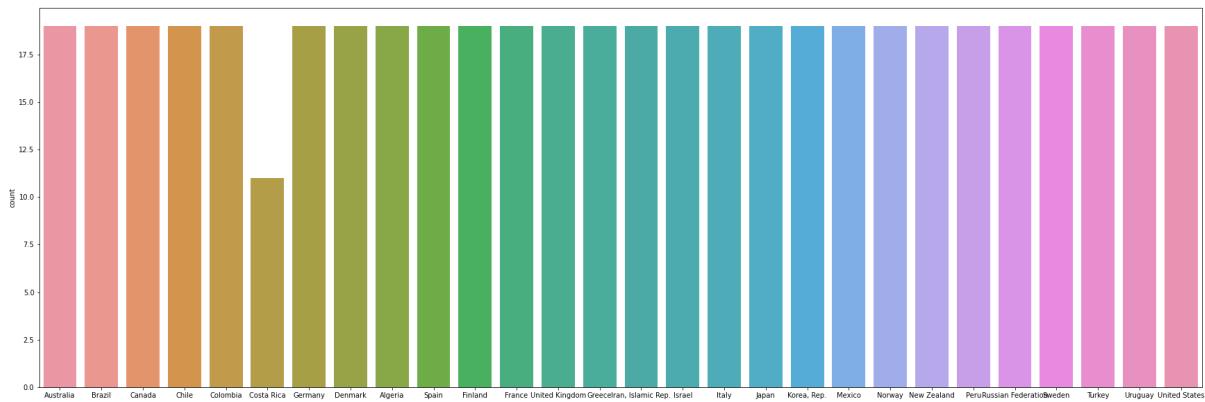


Można zauważać, że dwie grupy są do siebie zbliżone(1 i 2), natomiast 3 grupa ma obserwacje znacznie różne od pozostałych.

### 5.4 Analiza państw w poszczególnych klastrach

Do badań hierarchicznej analizy skupień wykorzystano tabelę stworzoną ze wszystkich obserwacji w ciągu 20 lat i lata były cechą, zatem każde państwo miało po 19 obserwacji dla lat 2001 - 2019. Zbiór został podzielony na trzy grupy.

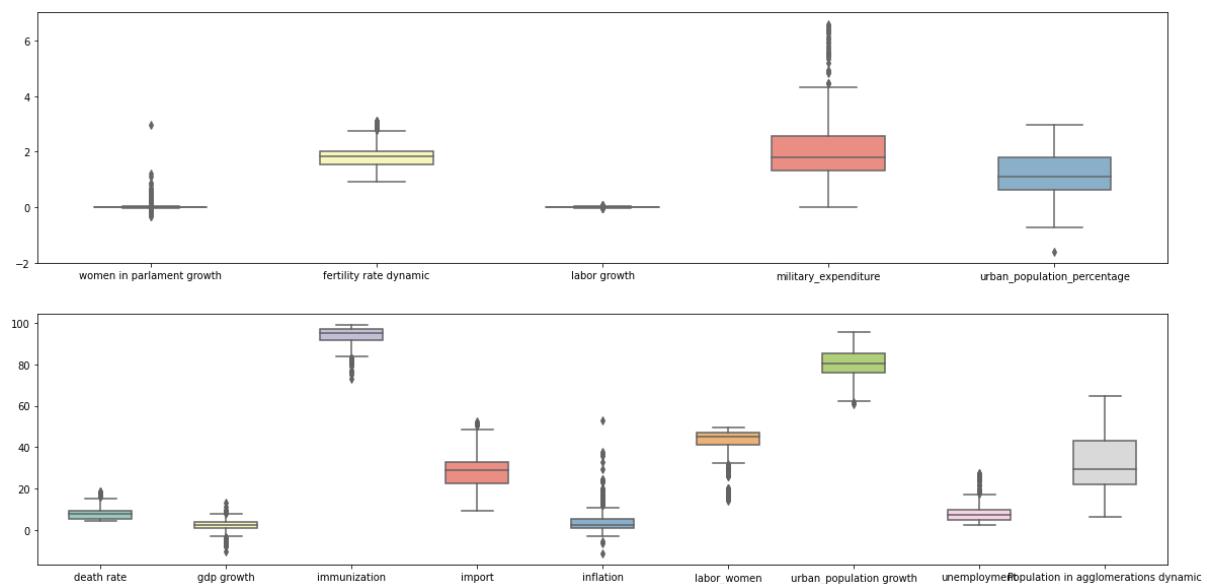
Pierwsza grupa państw, na wykresie “Categories of countries on the basis components” to grupa 0.



Widzimy, że w grupie tej wszystkie obserwacje mają państwa:

Australia, Brazylia, Kanada, Chile, Kolumbia, Niemcy, Dania, Algieria, Hiszpania, Finlandia, Francja, Wielka Brytania, Grecja, Iran, Izrael, Włochy, Japonia, Korea Południowa, Meksyk, Norwegia, Nowa Zelandia, Peru, Rosja, Szwecja, Turcja, Urugwaj, Stany Zjednoczone. Zatem w większości są to klasyfikowane jako dobrze rozwinięte państwa europejskie, jak i światowe.

W grupie pojawiło się państwo, które też nie uplasowało się w tej grupie przez cały badany okres, jest to Kostaryka. Obserwacje w grupie dla Kostaryki to lata 2009-2019.

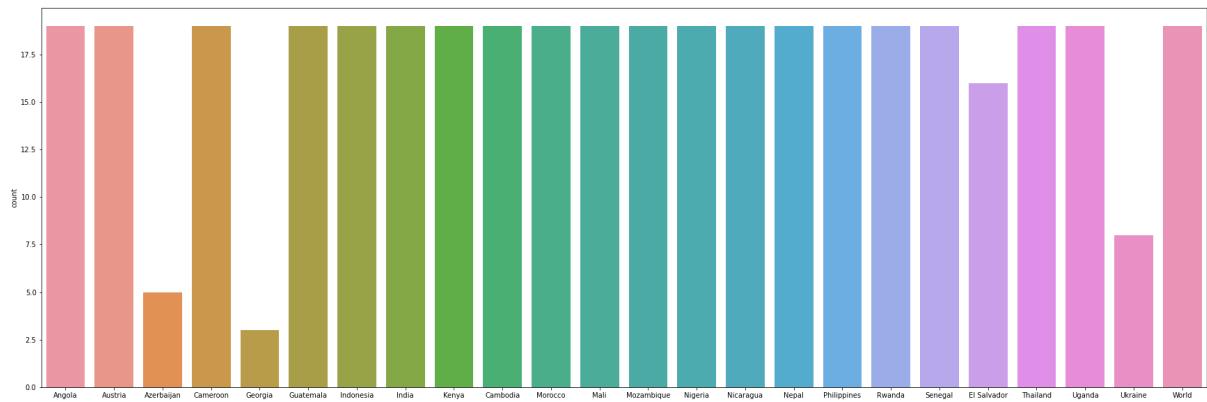


Grupa ta charakteryzuje się wysokimi wydatkami militarnymi, najmniejszą średnią death rate wśród grup, najmniejszym średnim gdp\_growth, niewielkim importem, największym współczynnikiem zaszczepień, najmniejszą inflacją oraz bardzo wysokim urban\_population growth.

Można spróbować zaklasyfikować te państwa jako **rozwinięte**, ponieważ wzrost PKB już nie jest tak duży jak w pozostałych grupach, państwa mogą sobie pozwolić na przeznaczenie większej ilości budżetu krajowego na wydatki militarne niż w pozostałych grupach, również służba zdrowia musi być tutaj dobrze rozwinięta o czym świadczy niski death\_rate,

dodatkowo w państwach rozwiniętych zmienia się struktura demograficzna i zazwyczaj rodziny decydują się na mniejszą ilość dzieci, co widzimy po wskaźniku `fertility_rate`. Niska wartość importu może wskazywać na gospodarczą niezależność tych państw od gospodarek innych państw.

Druga grupa państw, na wykresie “Categories of countries on the basis components” to grupa 1.

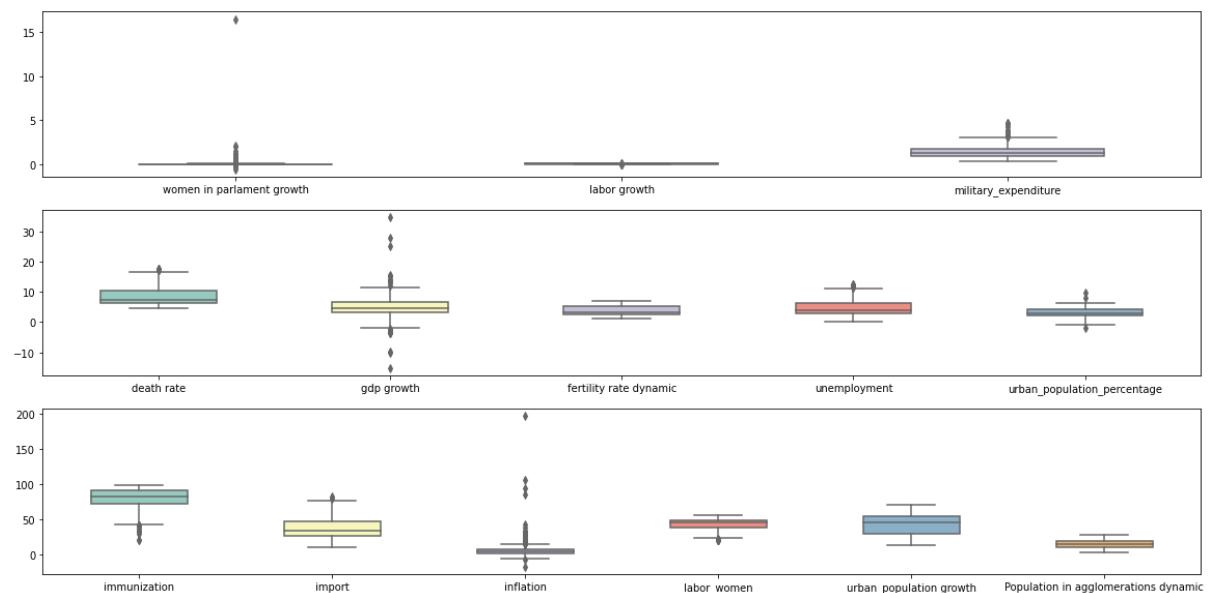


Widzimy, że w grupie tej wszystkie obserwacje mają państwa:

Angola, Austria, Kamerun, Gwatemala, Indonezja, Indie, Kenia, Kambodża, Maroko, Mali, Mozambik, Nigeria, Nikaragua, Nepal, Filipiny, Rwanda, Senegal, Tajlandia, Uganda oraz obserwacja ogólna dla świata.

W grupie pojawiły się państwa które też nie upłasowały się w tej grupie przez cały badany okres, są to : Ukraina, Salwador, Gruzja, Azerbejdżan.

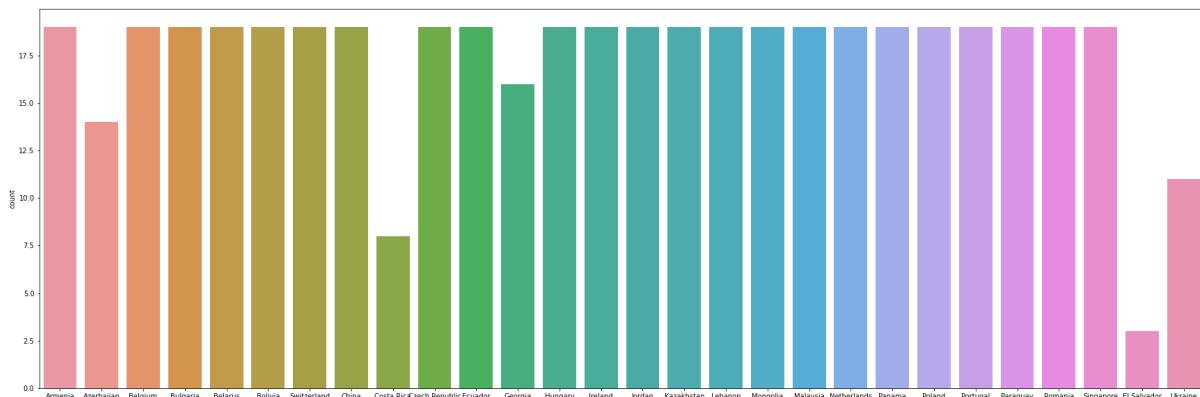
Obserwacje dla Azerbejdżanu pochodzą z lat 2001, 2002 oraz 2005-2007, dla Gruzji to lata 2001,2002,2003, dla Salwadoru w grupie znalazły się obserwacje z lat 2001-2016, dla Ukrainy lata 2009-2016



Grupa charakteryzuje się dużymi wartościami dla `woman_in_parlament_growth`, największym średnim współczynnikiem `death_rate`, największym `fertility_rate_dynamic` oraz najmniejszym współczynnikiem zaszczepień. W grupie tej występują niewielkie wartości dla zmiennej `population_in_agglomerations_dynamic` oraz najwyższym wśród badanych grup odsetkiem populacji mieszkającej w miastach.

Tą grupę państw można zaklasyfikować jako państwa **słabo rozwinięte**, wysoki wskaźnik `death_rate` oraz niski wskaźnik `immunization` mogą oznaczać problemy w segmencie ochrony zdrowia państwa. Dodatkowo widzimy, że dla tych państw następują największe wzrosty PKB, co nie jest często spotykanym zjawiskiem w krajach rozwiniętych.

Trzecia grupa państw, na wykresie “Categories of countries on the basis components” to grupa 2.



Widzimy, że w grupie tej wszystkie obserwacje mają państwa:

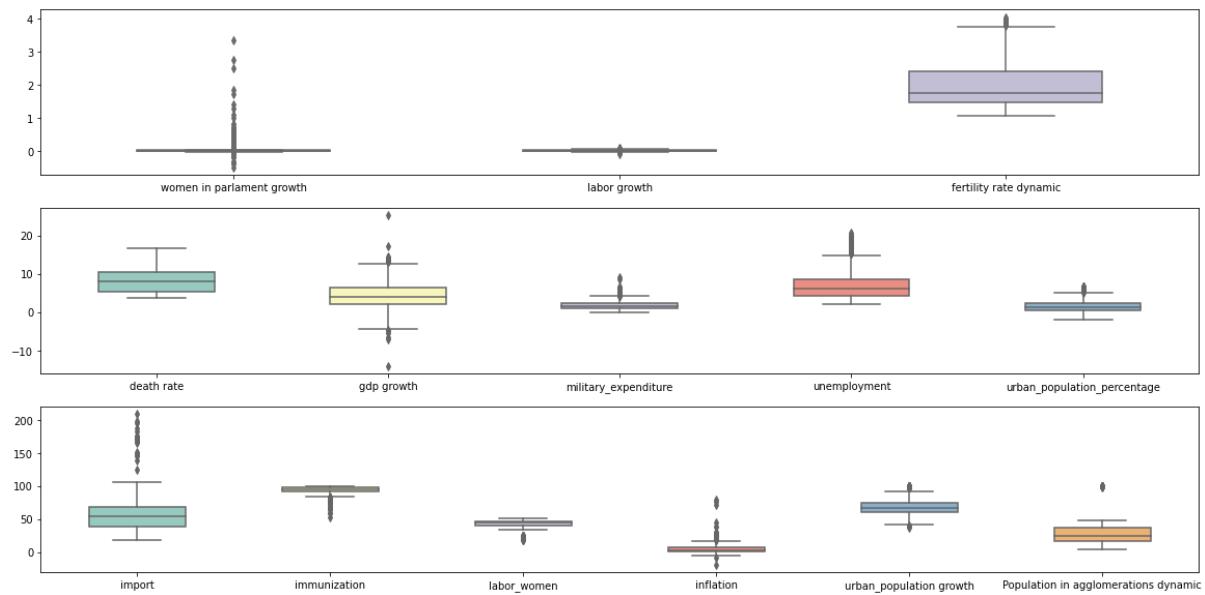
Armenia, Belgia, Bułgaria, Białoruś, Boliwia, Szwajcaria, Chiny, Czechy, Ekwador, Węgry, Irlandia, Jordania, Kazachstan, Lebanon, Mongolia, Malezja, Holandia, Panama, Polska, Portugalia, Paragwaj, Rumunia, Singapur.

W grupie pojawiły się państwa które też nie uplasowały się w tej grupie przez cały badany okres, są to : Azerbejdżan, Kostaryka, Gruzja, Salwador, Ukraina.

Szwajcaria, Gruzja, Meksyk,

W tej grupie pojawiły się pozostałe obserwacje dla Salwadoru z 2017, 2018 i 2019 roku. Dla Azerbejdżanu w tej grupie możemy zaobserwować próbki z lat 2003, 2004 oraz z lat 2008-2019. Dla kostaryki obserwacje pochodzą z lat 2001-2008 dla Gruzji to lata 2004-2019, natomiast obserwacje dla Ukrainy to lata 2001-2008 oraz 2017, 2018 i 2019.

Grupa charakteryzuje się największymi wartościami importu, wysokimi wartościami `GDP_growth`, wysokim poziomem zaszczepień. Wartości średnich pozostałych zmiennych znajdują się pomiędzy pozostałymi grupami, zatem tą grupę można nazwać tą ‘środkową’, a zatem są to **kraje rozwijające się**. Duże wartości importu świadczą o zależności gospodarczej od pozostałych państw, a co za tym idzie gospodarki tych państw nie są na tyle silne, aby być w pewnym sensie samowystarczalne.



Podsumowanie średnich wartości w formie tabeli przedstawiono poniżej, można na nim dostrzec cechy o których była mowa w opisach grup.

Średnie wartości zmiennych dla poszczególnych klastrów:

	women in parliament growth	death rate	fertility rate dynamic	gdp growth	immunization	import	inflation	labor_women	labor growth	military_expenditure	urban_population growth	unemployment	urban_population_percentage	Population in agglomerations dynamic	
clusters	0	0.044996	7.803256	1.860642	2.418765	93.597328	28.423097	4.209161	41.981059	0.012206	2.116732	80.253975	7.952437	1.182718	32.488956
	1	0.073096	8.436436	1.960042	4.249368	93.400818	57.385711	5.701050	42.474598	0.012294	1.935725	69.135051	7.219736	1.396161	27.193885
	2	0.104629	8.652230	3.695409	4.896903	79.732249	37.688299	6.730032	42.956719	0.020907	1.524463	42.065564	4.634504	3.062085	15.039079

## 5.5 Analiza zawirowań wśród klasyfikacji państw w poszczególnych latach.

Badając występowanie państw dla poszczególnych lat w grupach, można było zauważyc, że nie zawsze wszystkie państwa były dla całego badanego okresu umieszczone w jednej z grup. Występowały państwa, które dla pewnych lat były w jednej grupie, natomiast dla pozostałych występowały już w innej grupie. W tym podrozdziale przeanalizujemy i spróbujemy znaleźć przyczyny zmian w przydzielaniu do grupy dla poszczególnych państw w wybranych latach.

- Kostaryka z grupy państw rozwijających się do grupy państw rozwiniętych lata 2009-2019
- Azerbejdżan lata 2001, 2002 oraz 2005-2007 znalazły się w grupie państw słabo rozwiniętych, natomiast w pozostałych latach Azerbejdżan został sklasyfikowany jako państwo rozwijające się.
- Gruzja lata 2001,2002,2003 została sklasyfikowana jako państwo słabo rozwinięte, natomiast później dołączyła do grupy państw rozwijających się.
- Salwador w grupie państw słabo rozwiniętych znalazły się obserwacje z lat 2001-2016, późniejsze obserwacje to już grupa państw rozwijających się.
- Ukraina lata 2009-2016 to grupa państw słabo rozwiniętych, pozostałe obserwacje to grupa państw rozwijających się.

Można już po samych zmianach między grupami można zauważyc, że podział na państwa rozwinięte, rozwijające się i słabo rozwinięte tutaj bardzo pasuje, gdyż nie było ani jednego przypadku przejścia państwa z grupy państw słabo rozwiniętych do grupy państw rozwiniętych, a rotacje miały miejsce pomiędzy grupami rozwinięte i rozwijające się oraz słabo rozwinięte i rozwijające się, co ma sens.

### 5.5.1 Kostaryka

Pierwszym krokiem analizy, co miało wpływ na zmianę grupy dla Kostaryki w latach 2009-2019 będzie analiza PKB. Zazwyczaj po PKB można zauważyc rozwój ekonomiczny państwa.

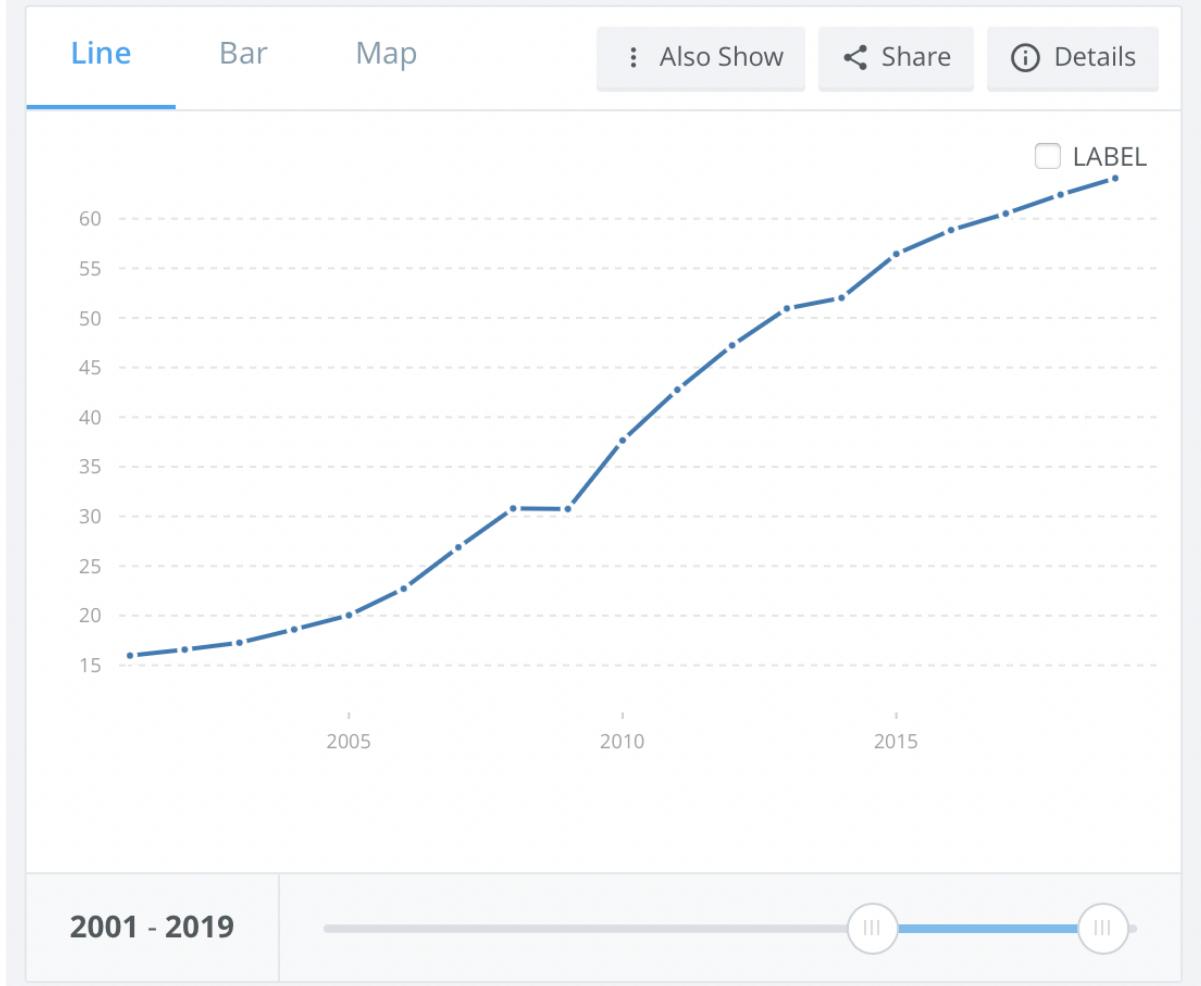
Na tle pozostałych państw środkowoamerykańskich, gospodarka Kostaryki od lat wyróżnia się pozytywnie. Jest stosunkowo stabilna, oparta na turystyce, rolnictwie i eksportie wyrobów przemysłu elektronicznego (zwłaszcza urządzeń medycznych), rozwija się w sposób zrównoważony, w niewielkim stopniu jest zależna od przekazów z zagranicy (natomiast znacznie bardziej od inwestycji zagranicznych) i zdecydowanie mniej skorumpowana niż inne kraje łatynoamerykańskie (z wyjątkiem Urugwaju i Chile).

Kostaryka jest jednym z największych celów przyjazdów cudzoziemców zainteresowanych ekoturystyką. Coraz szerzej rozwija się także turystyka medyczna. Z usług medycznych kostarykańskich klinik i szpitali skorzystało ca 70.000 cudzoziemców, a zyski z tej działalności przekraczają 500 mln dolarów. Ocenia się jednak, że Kostaryka nie wykorzystała jeszcze swojego potencjału w tej dziedzinie. Dwie najliczniejsze grupy turystów medycznych to Amerykanie i Kanadyjczycy. Pacjenci najczęściej korzystają z zabiegów medycyny sercowo-naczyniowej, nosowo- gardłowej, neurochirurgii, chirurgii plastycznej i ginekologii.

## GDP (current US\$) - Costa Rica

World Bank national accounts data, and OECD National Accounts data files.

License : CC BY-4.0 [i](#)

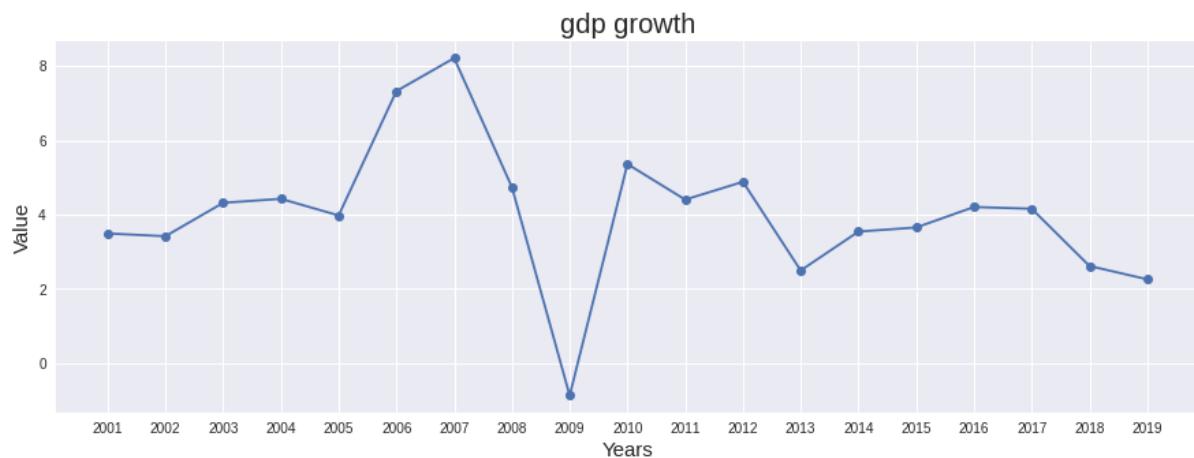


Jak można zauważyć w 2009 roku nastąpił silny wzrost PKB

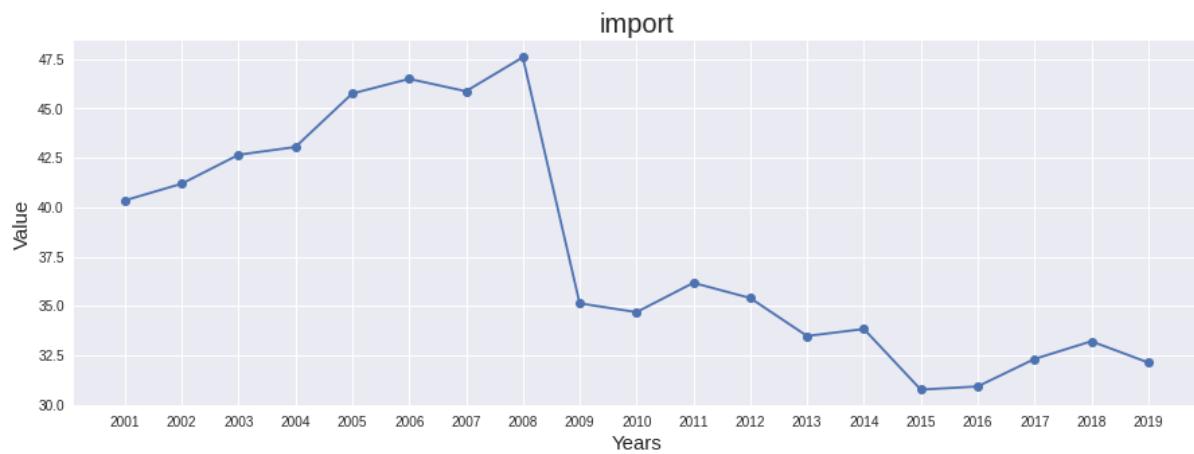
Z cech analizowanych dotychczas analizowanych można wymienić 3, które mogły mieć istotny wpływ w zaklasyfikowaniu Kostaryki w latach 2009-2019 do grupy państw rozwiniętych:

- gdp growth

W 2009 roku można zauważyć duży spadek dla zmiennej gdp growth, co oznacza, że PKB z 2008 na 2009 rok spadło. Po 2009 roku możemy zauważyć, że PKB stopniowo z roku na rok wzrastało, jednak wzrost ten nie był tak duży jak przed 2009 rokiem (ponad 8%), utrzymywał się w granicach 4% rocznie. Co jest cechą charakterystyczną dla grupy państw rozwiniętych.

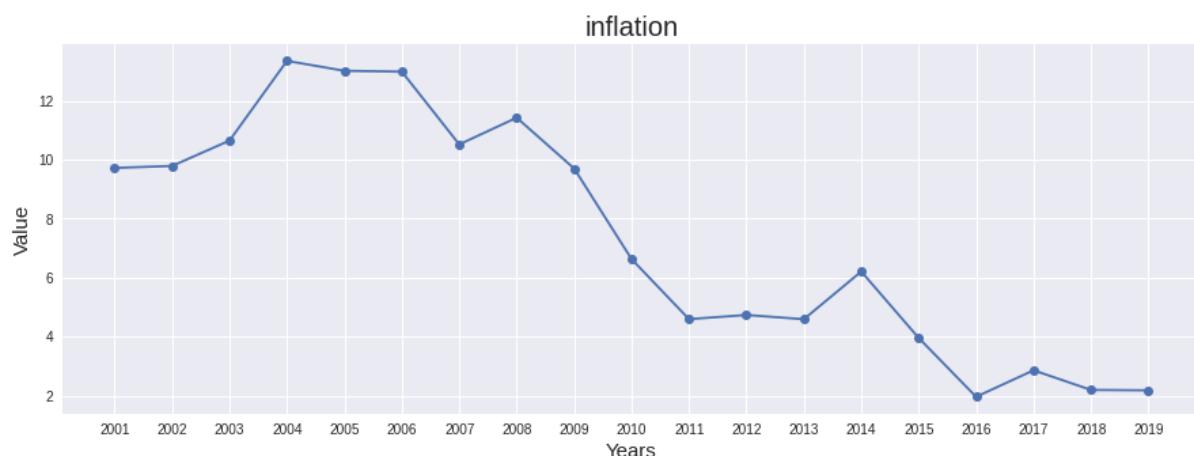


- import



W 2009 roku import znacznie spadł i potem jego wartość nie osiągnęła tak wysokich wartości jak przed 2009 rokiem. Kolejna cecha charakterystyczna dla grupy państw scharakteryzowanych jako rozwinięte.

- inflation



Inflacja od 2009 roku znacznie spadła i nie osiągnęła już w kolejnych latach wartości większej niż 6%.

Co mogło spowodować takie zmiany w gospodarce?

**Głównym powodem wszelkich przemian było zwiększenie się zainteresowania Kostaryką jako destynacją wakacyjną.**

#### All data for Costa Rica in detail

Year	Number of tourists	Receipts	% of GNP	Receipts per tourist
2020	1.15 m	1.48 bn \$	2.4 %	1,290 \$
2019	3.37 m	4.28 bn \$	6.7 %	1,272 \$
2018	3.31 m	4.07 bn \$	6.5 %	1,229 \$
2017	3.23 m	3.76 bn \$	6.2 %	1,163 \$
2016	3.20 m	3.78 bn \$	6.4 %	1,179 \$
2015	2.90 m	3.41 bn \$	6.0 %	1,176 \$
2014	2.76 m	3.14 bn \$	6.0 %	1,137 \$
2013	2.64 m	3.17 bn \$	6.2 %	1,201 \$
2012	2.54 m	2.77 bn \$	5.9 %	1,091 \$
2011	2.53 m	2.65 bn \$	6.2 %	1,048 \$
2010	2.50 m	2.43 bn \$	6.4 %	972 \$
2009	2.31 m	2.22 bn \$	7.2 %	960 \$
2008	2.41 m	2.81 bn \$	9.1 %	1,167 \$
2007	2.30 m	2.45 bn \$	9.1 %	1,066 \$
2006	2.07 m	2.07 bn \$	9.1 %	998 \$
2005	1.96 m	2.01 bn \$	10.02 %	1,025 \$
2004	1.77 m	1.76 bn \$	9.4 %	991 \$
2003	1.51 m	1.57 bn \$	9.1 %	1,039 \$
2002	1.34 m	1.43 bn \$	8.6 %	1,070 \$
2001	1.32 m	1.48 bn \$	9.3 %	1,121 \$

źródło: <https://www.worlddata.info/america/costa-rica/tourism.php>

Można zauważyc, że wartości z roku na rok rosną, przyczyniając się do wzrostu gospodarczego Kostaryki. Obecnie Kostaryka jest jednym z najbardziej popularnych kierunków wakacyjnych w Stanach Zjednoczonych, Kanadzie czy też Europie.

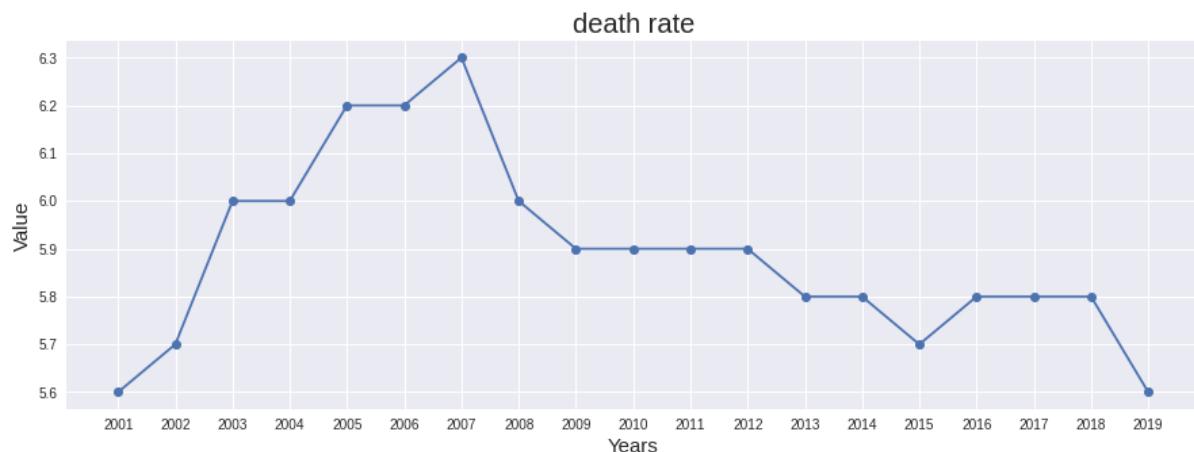
[OECD Tourism Trends and Policies 2020 - © OECD 2020](#)

## 5.5.2 Azerbejdżan

Dla lat 2001 i 2002 Azerbejdżan pod każdym względem miał bardzo niskie wartości, więc nie dziwi przydzielenie go do grupy państw zaklasyfikowanych jako słabo rozwinięte.

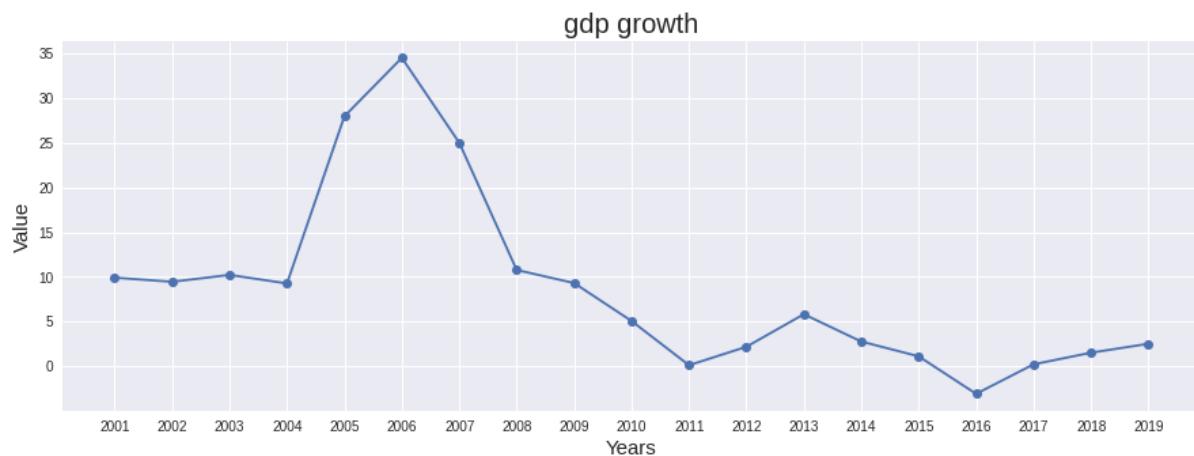
Z cech analizowanych dotychczas analizowanych można wymienić 3, które mogły mieć istotny wpływ w zaklasyfikowaniu Azerbejdżanu w latach 2001, 2002, 2005, 2006 i 2007 do grupy państw słabo rozwiniętych:

- death rate



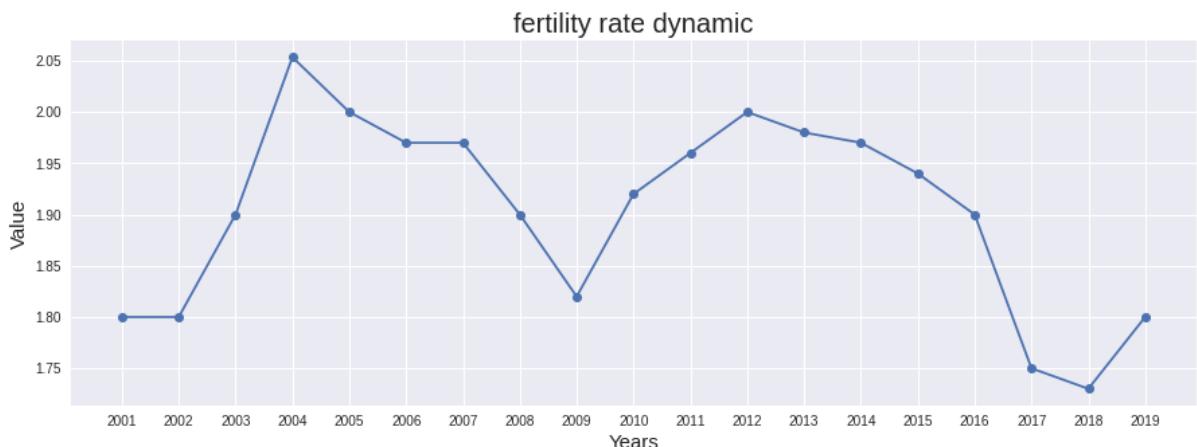
Bardzo wysoki wskaźnik death rate w badanym okresie można zauważyć dla lat 2005-2007 i dla tych lat też Azerbejdżan został zaklasyfikowany jako państwo słabo rozwinięte. Grupa państw słabo rozwiniętych charakteryzuje się wysokimi wartościami współczynnika death rate.

- gdp growth



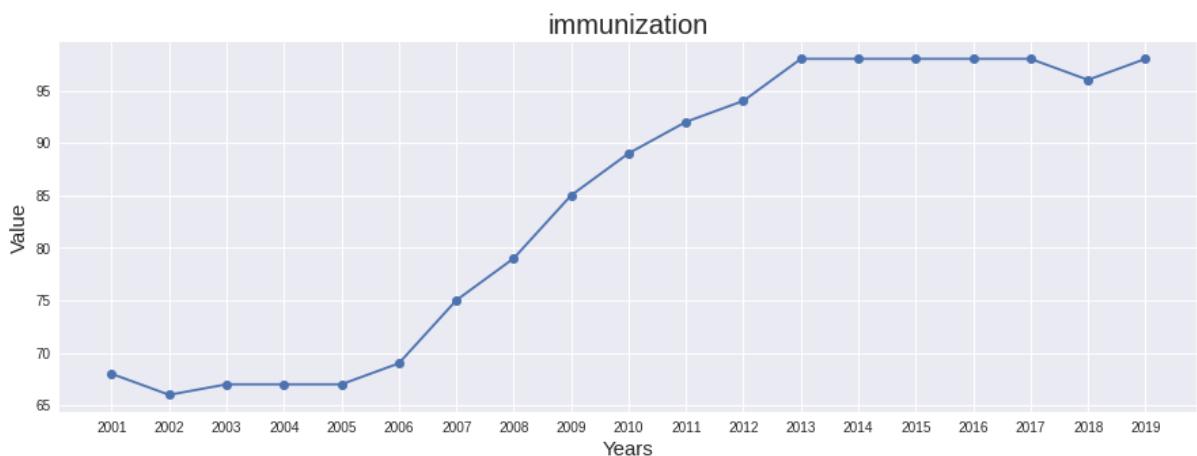
W latach 2005-2007 można zauważyć, że nastąpił największy wzrost PKB w badanym okresie, a wysokie wartości wzrostu PKB są kolejną charakterystyczną cechą grupy państw zaklasyfikowanych jako słabo rozwinięte.

- fertility rate dynamic

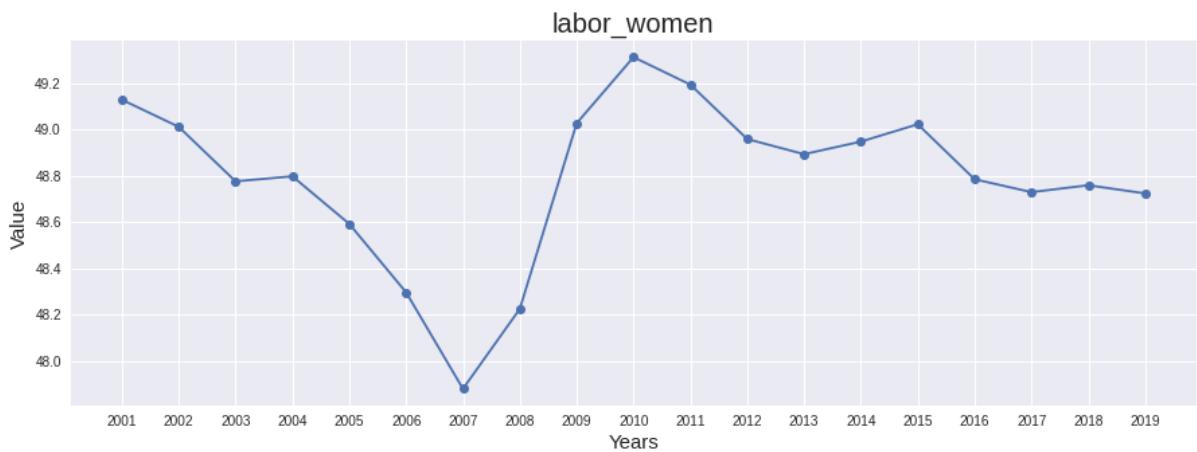


Wysokie wartości fertility rate dynamic dla lat 205-2007, cecha charakterystyczna dla grupy państw zaklasyfikowanych jako słabo rozwinięte.

Znacznej poprawie uległa sytuacja zaszczepienia społeczeństwa w Azerbejdżanie, co jest zapewne przejawem większego dostępu do służby zdrowia.



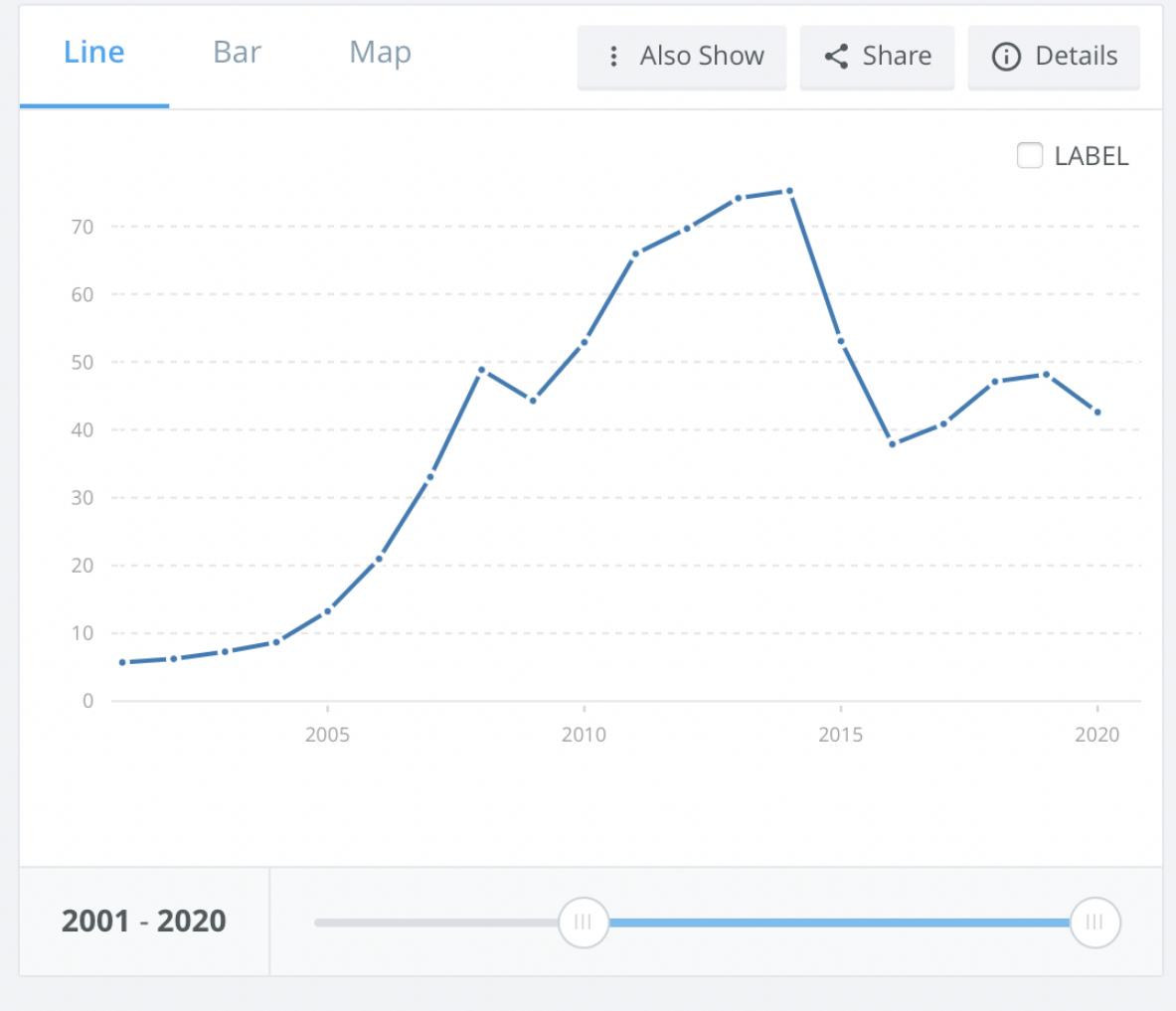
Poprawie też po 2007 roku uległ rynek pracy, jak można zauważyć do 2007 roku następował znaczny spadek odsetka kobiet pracujących. Po 2007 odsetek ten wzrósł, a następnie się utrzymywał na pewnym stałym poziomie



# GDP (current US\$) - Azerbaijan

World Bank national accounts data, and OECD National Accounts data files.

License : CC BY-4.0 [\(i\)](#)



Analizując PKB można zauważyc znaczny wzrost po 2007 roku.

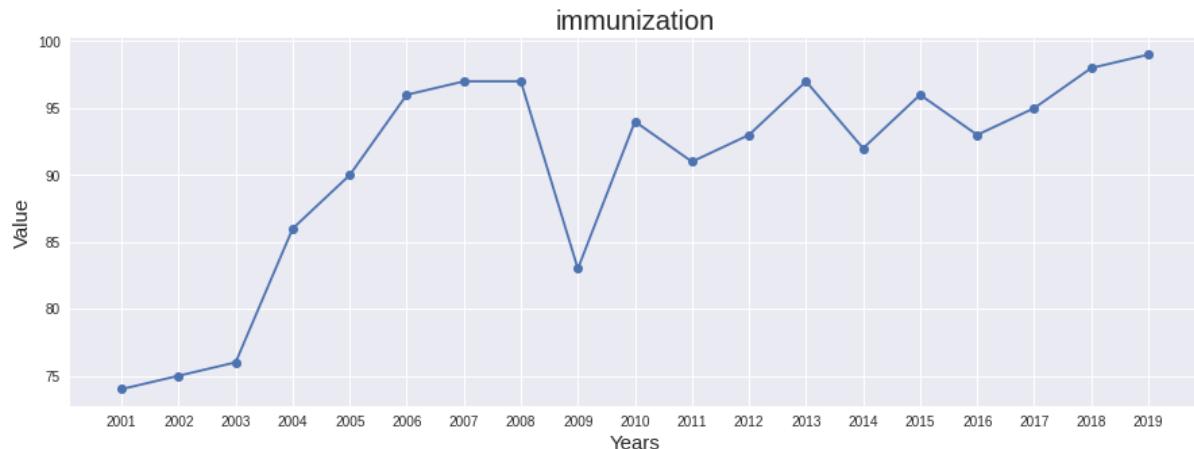
Dynamiczny rozwój kraju zawdzięcza sektorowi energetycznemu: wydobycie i przetwórstwo ropy naftowej oraz gazu ziemnego stanowi bowiem ok. 40 proc. PKB i aż 80 proc. azerbejdżańskiego eksportu. Wpływy z produkcji surowców naturalnych, wydobywanych przede wszystkim przez zachodnie kompanie naftowe (największym inwestorem w Azerbejdżanie jest BP), niepomiernie **wzrosły po ukończeniu budowy rurociągu Baku-Tbilisi-Ceyhan (2005 r.)**, umożliwiającego transport surowca na Zachód z ominięciem Rosji. Produkcja ropy wzrosła wówczas aż o 43 proc.

źródło: <https://www.cire.pl/pliki/2/azerbejdzanekst.pdf>

### 5.5.3 Gruzja

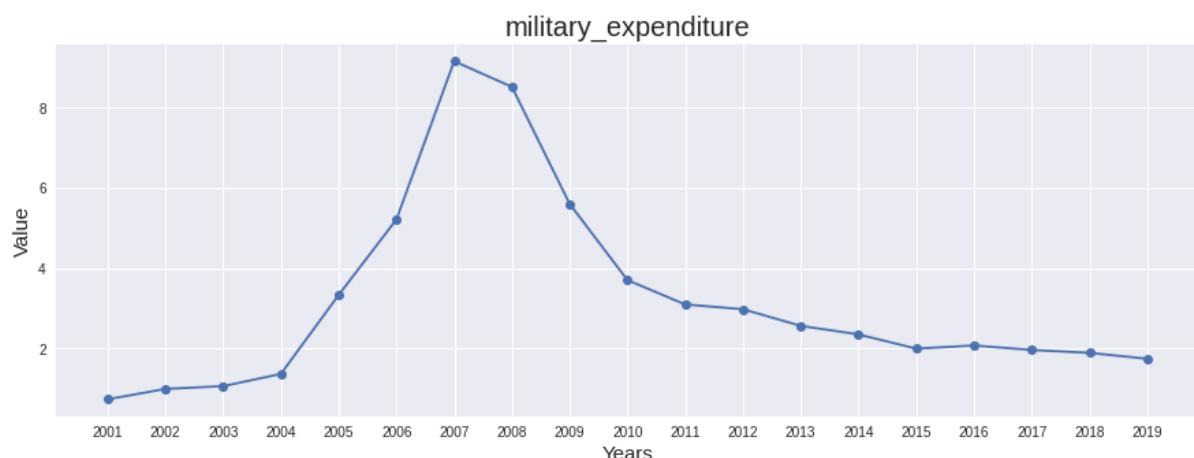
Z cech analizowanych dotychczas analizowanych można wymienić 3, które mogły mieć istotny wpływ w zaklasyfikowaniu Gruzji w latach 2001, 2002, 2003 do grupy państw słabo rozwiniętych:

- immunization



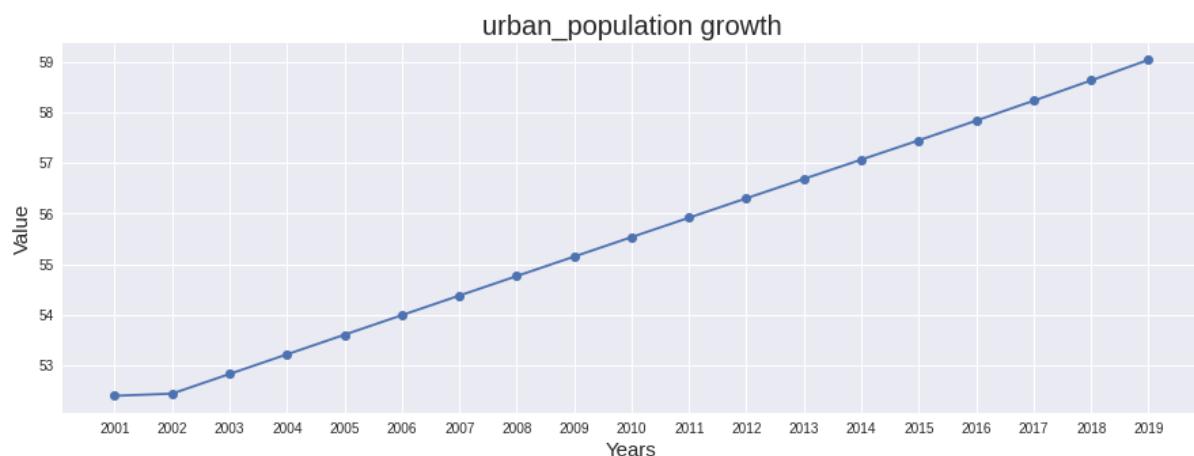
Bardzo niskie wartości dla lat 2001-2003, charakterystyczna cecha grupy państw zaklasyfikowanych jako słabo rozwinięte.

- military expenditure



Bardzo małe wartości dla lat 2001-2003, późniejszy wzrost wywołany jest konfliktem z Armenią. Jednak niskie wartości w wydatkach militarnych występują w grupie państw zaklasyfikowanych jako słabo rozwinięte.

- urban population growth

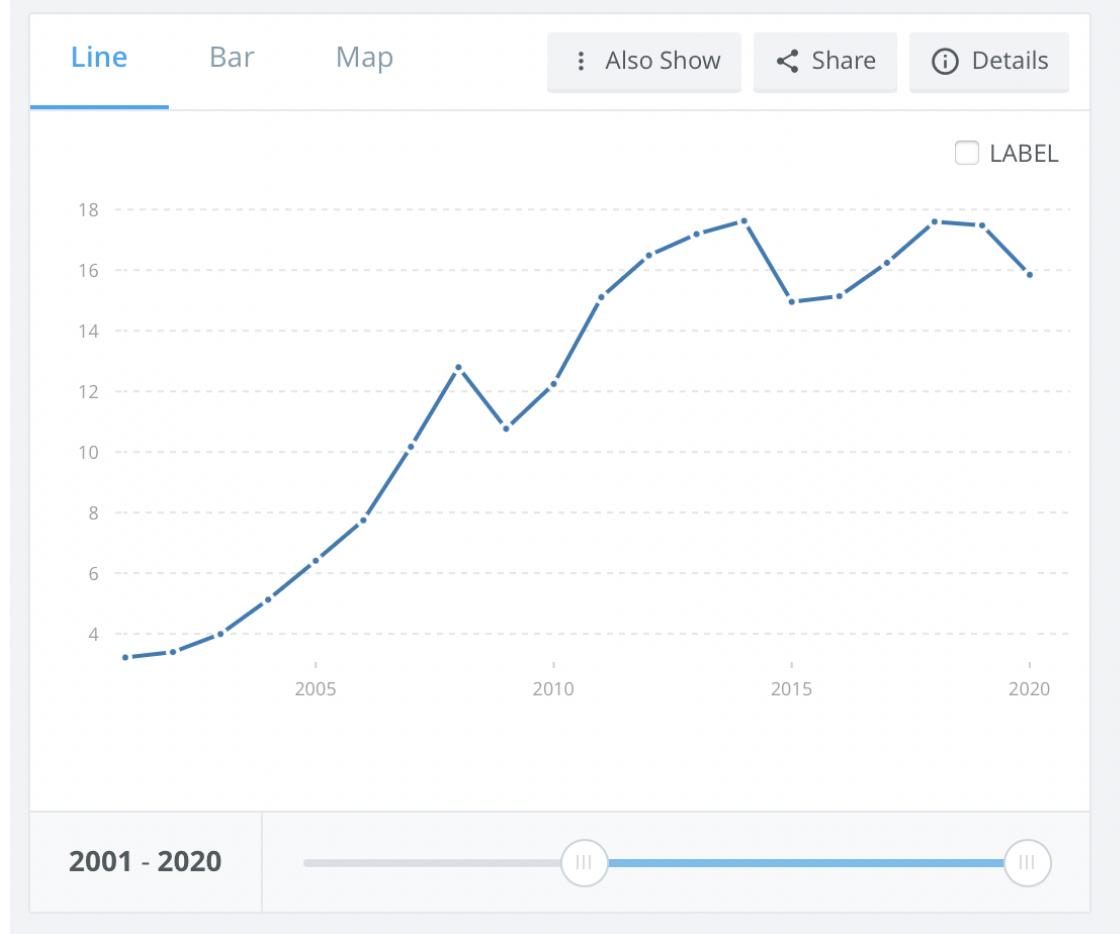


Bardzo mały wzrost w latach 2001-2003

## GDP (current US\$) - Georgia

World Bank national accounts data, and OECD National Accounts data files.

License : CC BY-4.0 [\(i\)](#)



Niskie wartości w latach 2001-2003, widać znaczny wzrost PKB po 2003 roku.

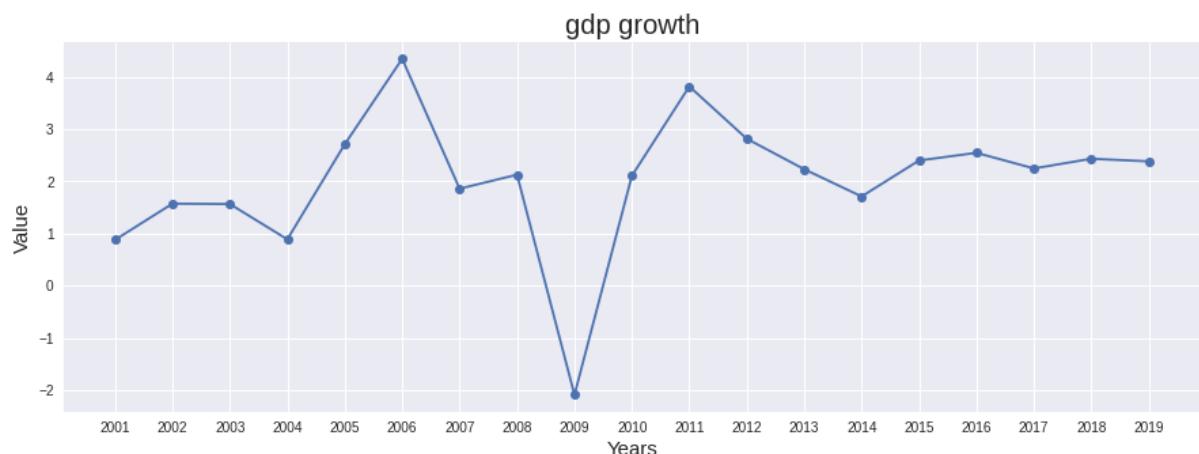
Na początku lat 90-tych XX w. w rezultacie rozpadu ZSRR, konfliktów wewnętrznych w Abchazji i Południowej Osetii i braku stabilizacji politycznej, Gruzja znalazła się w bardzo trudnej sytuacji gospodarczej. Jednakże w 1995 r., przy wsparciu Międzynarodowego Funduszu Walutowego i Banku Światowego, rozpoczęła proces stabilizacji gospodarki. Wprowadzono wymienialność waluty narodowej, spryatyzowano handel i drobną wytwórczość, stworzono bazę prawną dla rozwoju prywatnej przedsiębiorczości.

**Światowy kryzys ekonomiczny z 2009 r. nie zaszkodził Gruzji w takim stopniu jak niektórym krajom wschodnim.** Relatywnie niska zależność od eksportu i wysoki stopień rozwoju systemu bankowego oraz dopływ gotówki (przeznaczonej przede wszystkim na cele infrastrukturalne) związany z pomocą społeczności międzynarodowej spowodował, że kryzys nie wpłynął znaczco na dane makroekonomiczne kraju. Znacznie ważniejszym powodem ujemnej wartości wzrostu gospodarczego w 2009 r. był konflikt z Rosją z sierpnia 2008 r.

#### 5.5.4 Salwador

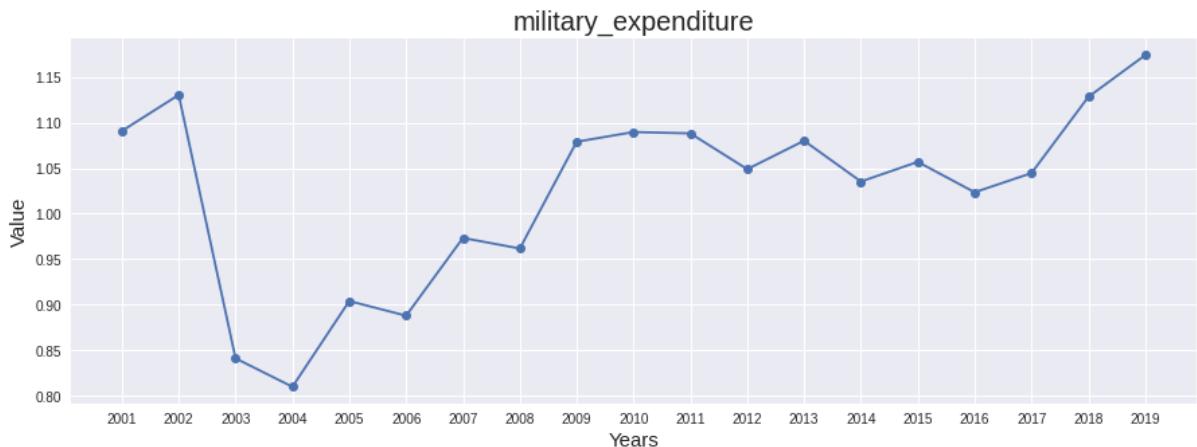
Z cech analizowanych dotychczas analizowanych można wymienić 4, które mogły mieć istotny wpływ w zaklasyfikowaniu Salwadoru w latach 2017, 2018, 2019 do grupy państw rozwijających się:

- gdp growth



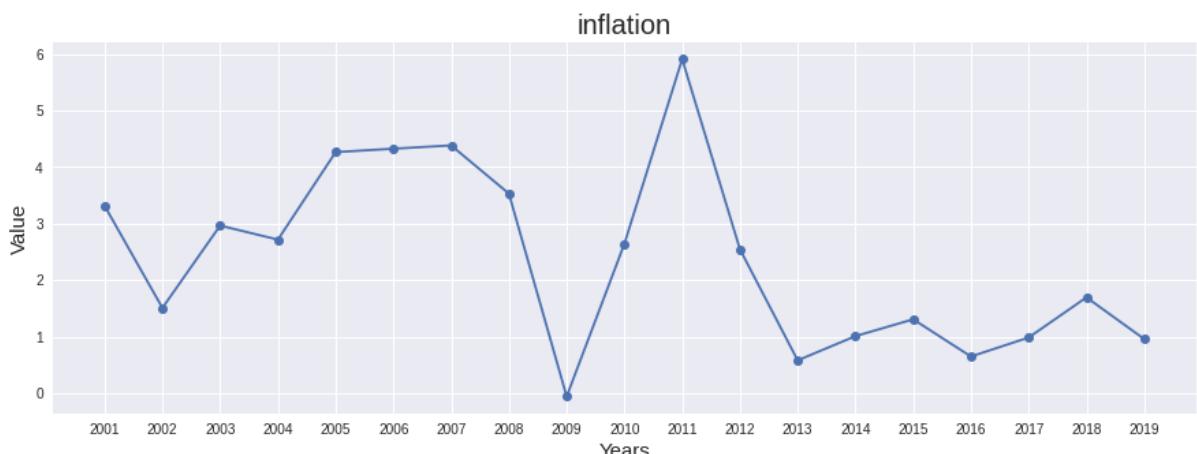
W latach 2015-2019 można zauważyć, że nastąpiło unormowanie się wzrostu PKB w badanym okresie

- military\_expenditure



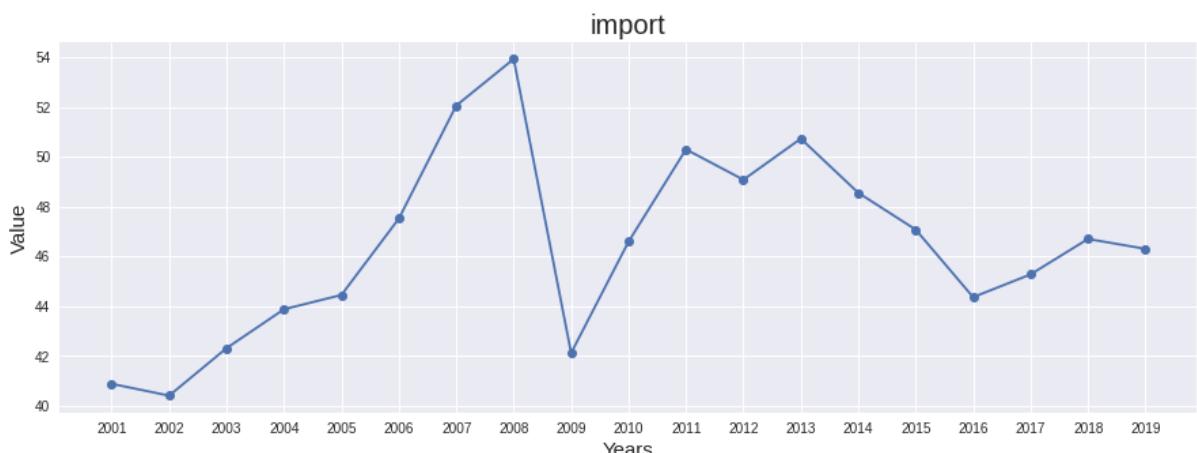
W latach 2017-2019 widać znaczny wzrost wydatków wojskowych, co mogło przyczynić się do tego, że Salwador został w tych latach zaklasyfikowany do grupy państw rozwijających się.

- inflation



Unormowanie się inflacji to kolejna cecha państw rozwijających się, można zauważyć, że przed 2013 rokiem jej wahania były znaczne.

- import



Unormowanie się importu to cecha państw rozwijających się, można zauważyc, że przed 2016 rokiem wahania były znaczne.

Od 2001 r. gospodarka Salwadoru jest całkowicie zdolaryzowana. **Salwador również aktywnie uczestniczy w rozszerzaniu regionalnej integracji ekonomicznej poprzez uczestnictwo w licznych regionalnych i ponadregionalnych inicjatywach gospodarczych.** Efektem tej polityki jest stały i umiarkowany rozwój kraju.

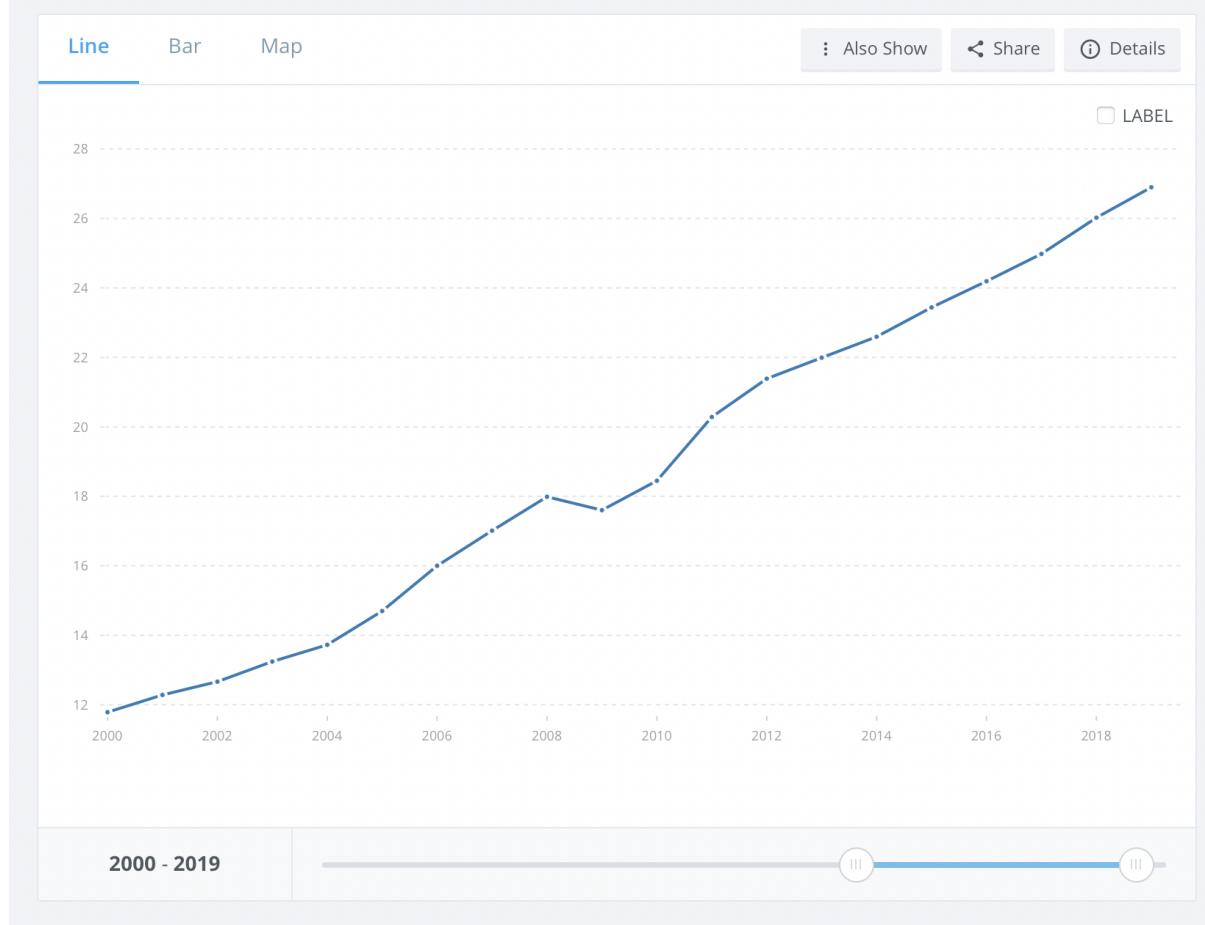
**Duże znaczenie (większe niż w innych państwach regionu) dla rozwoju kraju mają przekazy walutowe od emigrantów, zwłaszcza z USA. W 2019 r. wyniosły one 5,7 mld US\$ (przy poziomie eksportu 4,8 mld US\$).** Są one źródłem utrzymania dla blisko 25% ludności Salwadoru. Warto wspomnieć, że w USA mieszka blisko 2,8 mln Salvadorczyków, co stanowi blisko 0,8% populacji tego państwa.

Podobnie jak w innych państwach regionu, obserwuje się stopniowe odchodzenie od rolnictwa na rzecz rozwoju usług i przemysłu z uwagi na słabe możliwości eksportowe, niską kulturę gospodarki rolnej oraz zależność od katastrof naturalnych i wahań cen na rynkach światowych. Jest to kierunek popierany przez kolejne rządy.

### GDP (current US\$) - El Salvador

World Bank national accounts data, and OECD National Accounts data files.

License : CC BY-4.0 [i](#)



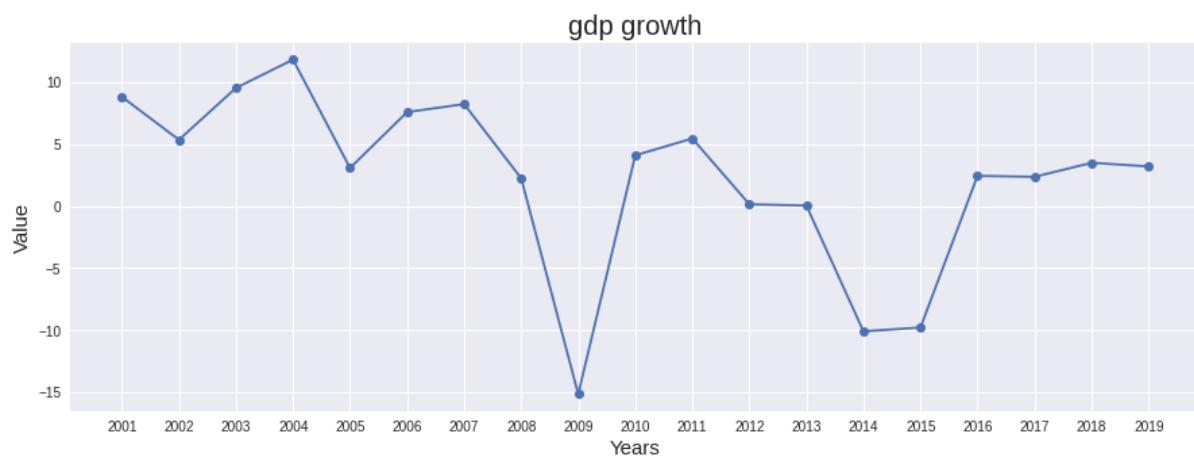
Najpoważniejszymi wyzwaniami dla Salwadoru (pomijając fenomeny naturalne) jest nadal wysoka przestępcość, silna korupcja oraz niski poziom życia. Jak dotychczas największe sukcesy w sprostaniu tym wyzwaniom Salwador osiągnął w zmniejszeniu poziomu ubóstwa z ponad 60% do 30%. Duże znaczenie w tym zakresie miała pomoc w wysokości 461 mln US\$ udzielona przez rządową organizację z USA Millennium Challenge Corporation przeznaczona na inwestycje w edukację, usługi publiczne, rozwój przedsiębiorczości i infrastruktury transportowej. Pomoc trafiła na północ kraju, która stanowiła główną strefę konfliktu w czasie wojny domowej.

Pomimo utrzymania dyscypliny fiskalnej i finansowej, problemem Salwadoru jest także wysoki i rosnący poziom dłużu publicznego.

### 5.5.5 Ukraina

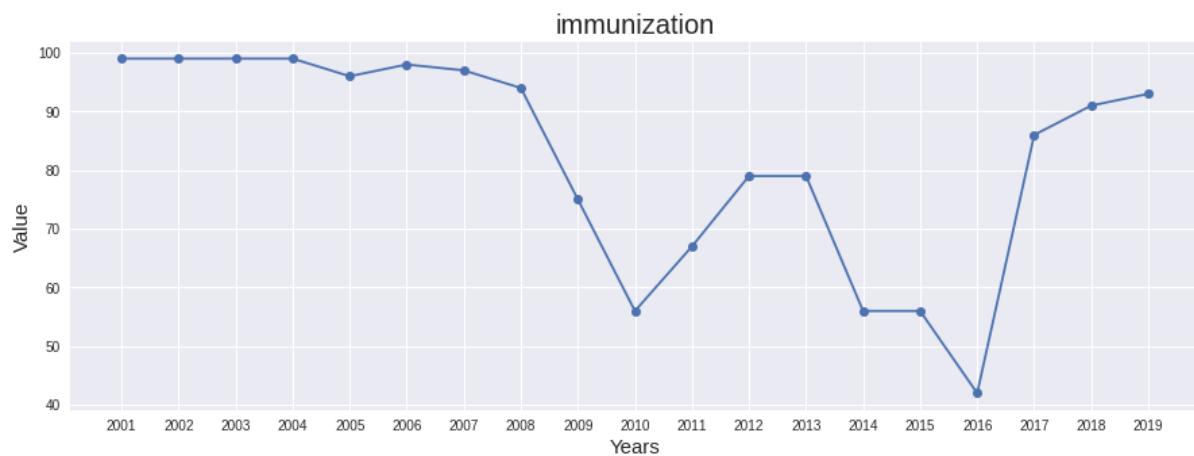
Z cech analizowanych dotychczas analizowanych można wymienić 2, które mogły mieć istotny wpływ w zaklasyfikowaniu Ukrainy w latach 2009-2016 to grupa państw słabo rozwiniętych:

- gdp growth

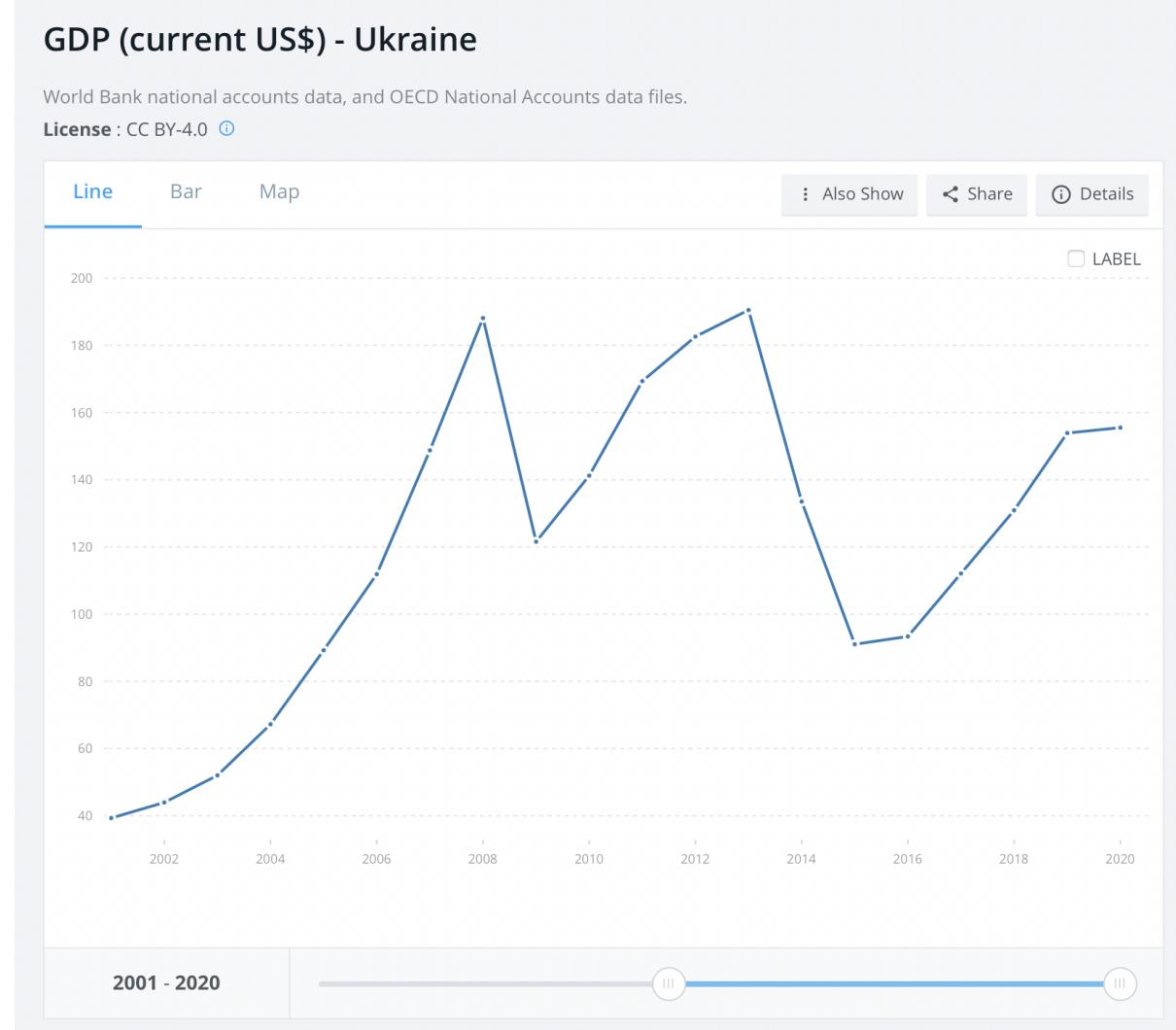


Między latami 2006-2016 wystąpiły duże wahania we wzroście PKB, co jest kolejną cechą państw słabowo rozwiniętych.

- immunization



Jak widać na wykresie można zauważać duży spadek współczynnika zaszczepień w latach 2009-2016, co może być głównym skutkiem w zaklasyfikowaniu Ukrainy w tych latach do grupy państw słabo rozwiniętych.



Z wykresu PKB można zauważyć bardzo duże wahania wartości PKB w latach 2008-2016, co może świadczyć o problemach gospodarczych w kraju.

W związku z polityką Rosji, wojną domową na wschodzie kraju, Ukraina borykała się z szeregiem problemów gospodarczych tj. postępującym spadkiem aktywności gospodarczej, wycofywaniem wkładów bankowych, olbrzymią inflacją czy spadem obrotów handlu zagranicznego. W 2014 i 2015 r. odnotowała ujemne PKB (odpowiednio -6,6% i -9,8%), z inflacją odpowiednio 24,9% i 43,3%. Od 2016 r. PKB zaczęło wzrastać, w 2018 r. inflacja nie osiągnęła już dwucyfrowego wyniku, a w 2019 r. wyniosła ok. 4,1%.

W ostatnich latach Ukraina rozpoczęła wdrażanie licznych reform gospodarczych, co wynikało również ze stowarzyszenia z UE i wymogów z tym związanych. W czerwcu 2016 r. przyjęto ustawę „O sądownictwie i statusie sędziów”, którą powołano nowy Sąd Najwyższy, jak również dwa wyższe sądy wyspecjalizowane: Najwyższy Sąd do Spraw Własności Intelektualnej oraz Najwyższy Sąd Antykorupcyjny. W 2018 r. kontynuowano reformy:

decentralizacyjną, opieki zdrowotnej i systemu oświaty. Przyjęto również ustawę o sądzie antykorupcyjnym i ustawę prywatyzacyjną. Pracę rozpoczęło Państwowe Biuro Śledcze, nowy organ dochodzeniowy, utworzony w celu zwalczania przestępstw popełnionych przez urzędników wysokiego szczebla, jak również organy antykorupcyjne: Krajowe Biuro Antykorupcyjne, Specjalna Prokuratura Antykorupcyjna, Narodowa Agencja ds. Zapobiegania Korupcji.

Sytuację Ukrainy pogorszyła Aneksja Krymu przez Rosję – nielegalne przyłączenie przez Federację Rosyjską terytorium Ukrainy (Autonomicznej Republiki Krymu oraz miasta wydzielonego Sewastopol, położonych na Półwyspie Krymskim) w drodze użycia siły zbrojnej w 2014 roku.

## 5.6 Ciekawe spostrzeżenia dotyczące rozwoju państw

Określenie podobieństw i różnic pomiędzy państwami wchodząymi w skład badania:

- Dwie najczęstsze metody przeliczania PKB na wspólną walutę to parytet nominalny i parytet siły nabywczej (PPP). Stany Zjednoczone i Chiny są dwiema największymi gospodarkami na świecie. Stany Zjednoczone są na szczycie pod względem nominalnym, podczas gdy Chiny znajdują się na szczycie PPP od 2017 roku po wyprzedzeniu USA. Od 2021 r. oba kraje łącznie posiadają 41,89% i 34,75% światowego PKB odpowiednio w ujęciu nominalnym i PPP. PKB obu krajów jest znacznie wyższy niż trzecie miejsce w rankingu Japonia (nominalnie) i Indie (PPP) z ogromnym marginesem. Dlatego tylko ci dwaj konkurują o pierwszeństwo.
- Poziom aktywności zawodowej kobiet w Polsce w ciągu 20 badanych lat był zdecydowanie mniejszy niż większość krajów europejskich. Istotnie mniejszy wskaźnik od Polski posiadają Włochy. Niestety wskaźnik zatrudnienia kobiet we Włoszech jest poniżej średniej europejskiej: przy czym nadal utrzymuje się ogromna różnica pomiędzy liczbą kobiet zatrudnionych na północy w stosunku do południowych regionów.
- Zdecydowana większość krajów o najwyższym współczynniku dzietności znajduje się w Afryce. Najwyższe wskaźnik dzietności mają takie państwa jak: Nigeria, Angola, Senegal, Mali. Dwa najbardziej zaludnione kraje świata, Chiny i Indie, mają współczynniki dzietności w dolnej środkowej części skali. Na obie te liczby wpływ ma polityka rządowa i oczekiwania kulturowe dotyczące reprodukcji w tych krajach. Na przykład Chiny utrzymywały „politykę jednego dziecka” od około 1980 do 2016 roku, ale w sierpniu 2021 uchwaliły ustawę, zgodnie z którą pary małżeńskie mogą mieć nawet troje dzieci.
- Bułgaria ma najwyższy wskaźnik śmiertelności na świecie/ Według Światowej Organizacji Zdrowia przyczyny zgonów w Bułgarii są podobne do tych w innych krajach europejskich, w tym choroby niezakaźne (choroby układu krążenia, pokarmowego lub oddechowego) oraz nowotwory.
- Również wysoki współczynnik zgonów ma Rosja. Ponad połowa zgonów w Rosji jest spowodowana chorobami układu krążenia. Drugą najczęszą przyczyną śmierci jest rak, a następnie samobójstwa, wypadki drogowe, zabójstwa i zatrucie alkoholem. Nadużywanie alkoholu to poważny problem w Rosji, zwłaszcza dla mężczyzn.
- Rwanda jest na czele listy jako kraj o najwyższym odsetku kobiet w parlamencie. W 2003 r. Rwanda uchwaliła nową konstytucję, która nakazywała, aby co najmniej 30 procent wszystkich miejsc ustawodawczych było zarezerwowanych dla kobiet. W

ciągu dziesięciu lat w parlamencie jest ponad dwa razy więcej kobiet niż wymuszona linia bazowa.

## 6. Wyciągnięcie wniosków poprzez zbadanie państw pod względem współczynnika szczęścia

### 6.1 Teoria

The World Happiness Report to publikacja United Nations Sustainable Development Solutions Network (UNSDSN), w której znajduje się między innymi ranking najszerzej ocenianych państw świata. Raport bazuje na udzielonych odpowiedziach obywateli, jak i niezależnej ocenie konkretnych czynników wpływających na jakość życia. Są to między innymi: dochód, oczekiwana długość życia w dobrym zdrowiu czy zaufanie mierzone brakiem korupcji w biznesie i rządzie. W każdej z kategorii państwa otrzymują punkty w 10-stopniowej skali. Pierwszy raport ukazał się w 2012 roku, a liczba badanych państw stopniowo zmniejsza się z roku na rok. W 2015 roku wynosiła 158, natomiast w 2022 roku jest równa 146.

### 6.2 Wykorzystywany zbiór danych

#### 6.2.1 Pochodzenie zbioru

Wykorzystywany zbiór pochodzi ze strony:

<https://www.kaggle.com/datasets/mathurinache/world-happiness-report>

#### 6.2.2 Dobór zbioru

Dane dotyczące poszczególnych lat znajdują się w odrębnych plikach, w których pojawiają się pewne nieścisłości co do liczby obserwacji czy charakteru zmiennych. Z tego powodu analiza ograniczy się do danych opisujących rok 2019. Dodatkowo, zbiór został zredukowany do państw, które byłybrane pod uwagę w poprzedniej analizie. Przykładowo, czwarte miejsce według współczynnika szczęścia w 2019 roku zajmowała Islandia, lecz kraj ten nie występował w poprzednich zbiorach pobranych z World Bank Data, więc tutaj również został usunięty. Badany zbiór danych prezentuje się następująco:

Country or region	Overall rank	Score	GDP per capita	Social support	Healthy life expectancy	Freedom to make life choices	Generosity	Perceptions of corruption
Finland	1	7.769	1.340	1.587	0.986	0.596	0.153	0.393
Denmark	2	7.600	1.383	1.573	0.996	0.592	0.252	0.410
Norway	3	7.554	1.488	1.582	1.028	0.603	0.271	0.341
Netherlands	5	7.488	1.396	1.522	0.999	0.557	0.322	0.298
Switzerland	6	7.480	1.452	1.526	1.052	0.572	0.263	0.343
...	...	...	...	...	...	...	...	...
Mali	128	4.390	0.385	1.105	0.308	0.327	0.153	0.052
Ukraine	133	4.332	0.820	1.390	0.739	0.178	0.187	0.010
Uganda	136	4.189	0.332	1.069	0.443	0.356	0.252	0.060
India	140	4.015	0.755	0.765	0.588	0.498	0.200	0.085
Rwanda	152	3.334	0.359	0.711	0.614	0.555	0.217	0.411

## 6.2.3 Opisywane zmienne

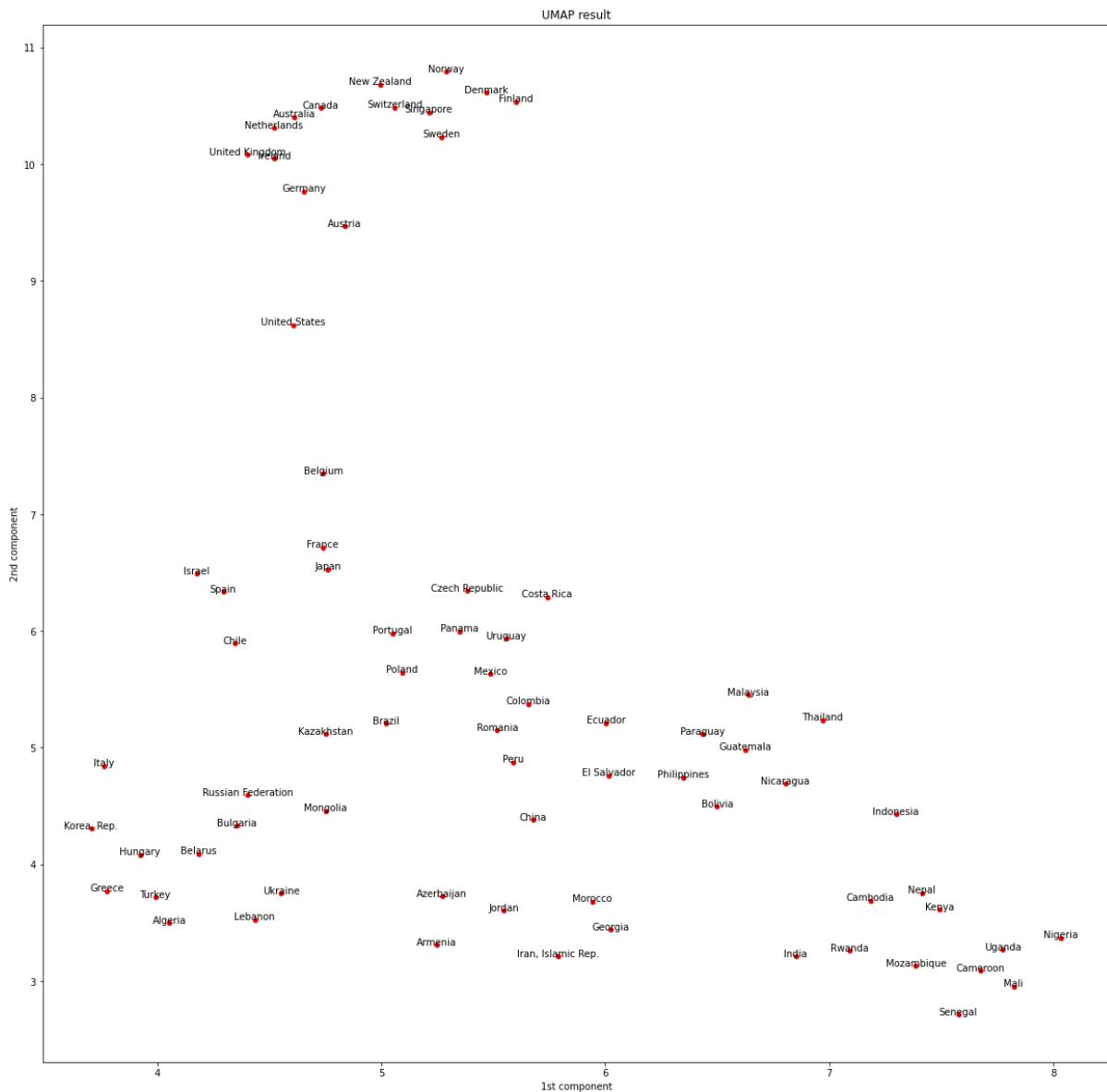
Zbiór składa się z ośmiu zmiennych, którymi są:

'Overall rank', 'Score', 'GDP per capita', 'Social support',  
 'Healthy life expectancy', 'Freedom to make life choices',  
 'Generosity', 'Perceptions of corruption'

Pierwsza z nich oznacza miejsce, które dany kraj zajął w rankingu, a druga liczbę przyznanych punktów. Kolejne sześć zmiennych opisuje wpływ danego czynnika na końcowy wynik.

## 6.2.4 Projekcja zbioru w 2D

Zbiór składa się z 73 obserwacji oraz 8 zmiennych, przez co jego wizualizacja jest niemożliwa do zrealizowania. Aby rozwiązać ten problem, tak jak poprzednio, została użyta redukcja do dwóch wymiarów przy użyciu metody UMAP.



Z prezentowanego wykresu możemy łatwo odczytać kraje podobne pod względem badanego współczynnika szczęścia w 2019 roku.

Na górze obrazka znajduje się najbardziej odseparowany klaster od reszty obserwacji, który wskazuje na państwa przodujące w danym rankingu. Znajdują się tam między innymi Finlandia, Dania, Norwegia, Szwecja, Nowa Zelandia czy Singapur.

Kolejny wyraźny klaster znajduje się w prawym dolnym rogu, który jest skupiskiem takich państw jak: Rwanda, Nigeria, Uganda, Nepal, Kenya czy Indie. Kraje te okupują końcowe miejsca w danym rankingu.

Krajami podobnymi do siebie są również Czechy, Kostaryka, Polska, Portugalia czy Urugwaj. Oraz z drugiej strony Grecja, Węgry, Turcja, Algieria i Białoruś.

Ciekawą obserwacją jest niskie położenie krajów takich jak Japonia czy Korea Płd. jako państw bardzo dobrze rozwiniętych o demokratycznym ustroju politycznym. Miejsca te

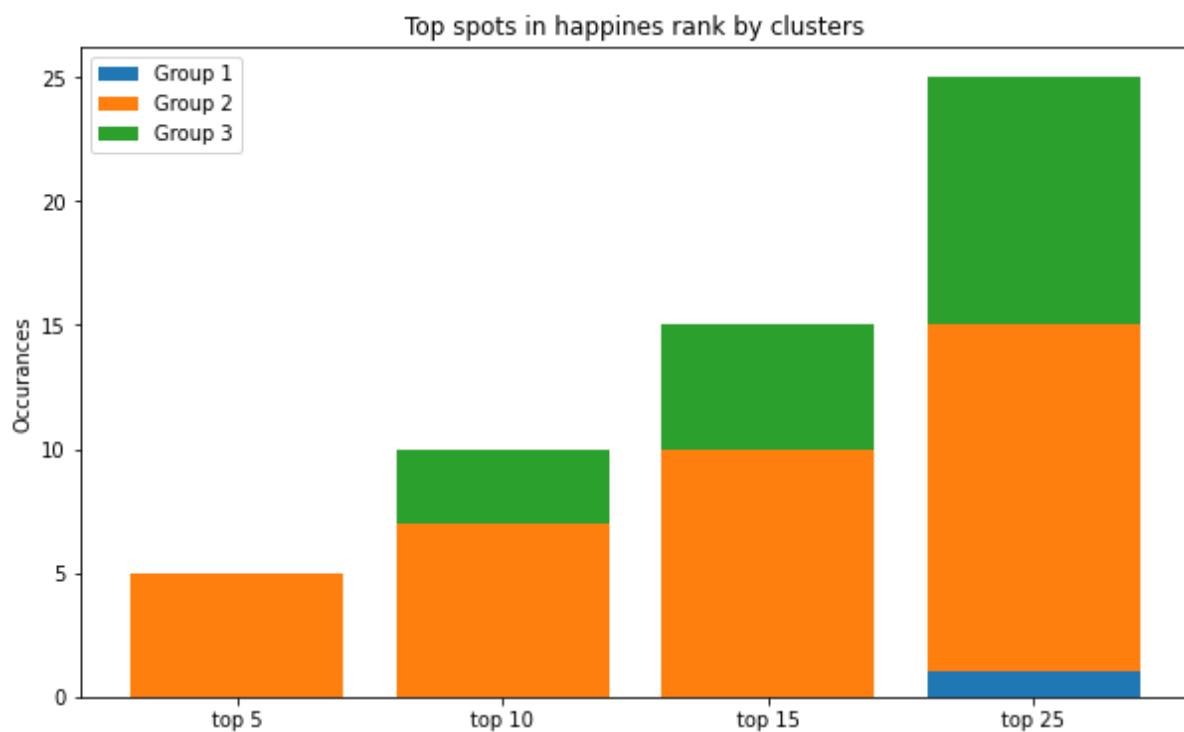
jednak cechują się wysokim współczynnikiem samobójstw, co może świadczyć o tym, że ludność zamieszkująca te tereny nie należy do najszczęśliwszych.

## 6.3 Analiza współczynnika szczęścia w oparciu o grupy państw podobnych

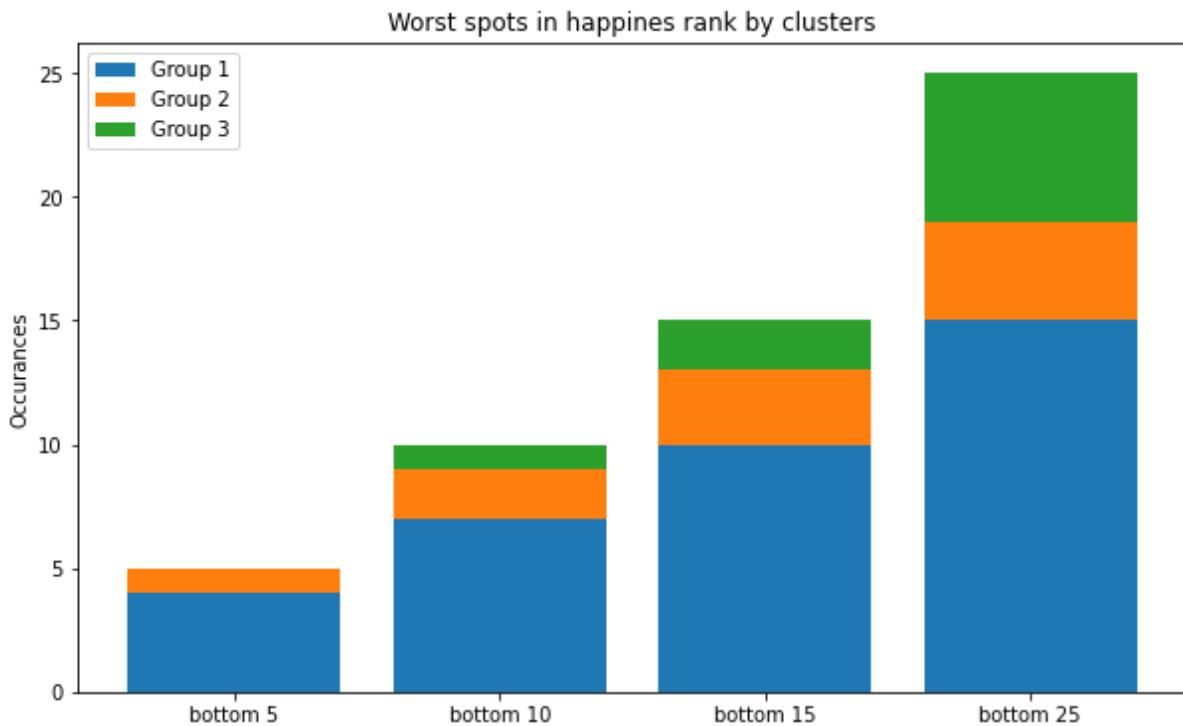
Z powyższych informacji możemy odczytać państwa przodujące i będące na końcowych miejscach rankingu objaśnianego przez współczynnik szczęścia. Może nasuwać się jednak pytanie, co jest głównym czynnikiem wpływającym na przyznawane punkty. Jak wygląda sytuacja w odniesieniu do państw podobnych? W celu porównania współczynnika szczęścia między państwami podobnymi wykorzystany zostanie podział na trzy grupy przy użyciu metody K Means.

### 6.3.1 Analiza indeksu na tle grup państw podobnych

Aby móc porównać zadowolenie z życia w danej grupie krajów i wyciągnąć pewne wnioski można zobrazować, która grupa państw zawiera najczęściej wystąpień w czołówce rankingu lub na jego dolnych pozycjach.



Na załączonym wykresie widoczne jest, że najszczęśliwi ludzie żyją w grupie drugiej, a następnie w grupie trzeciej. Przedstawiciel grupy pierwszej pojawia się dopiero w badaniu top 25.



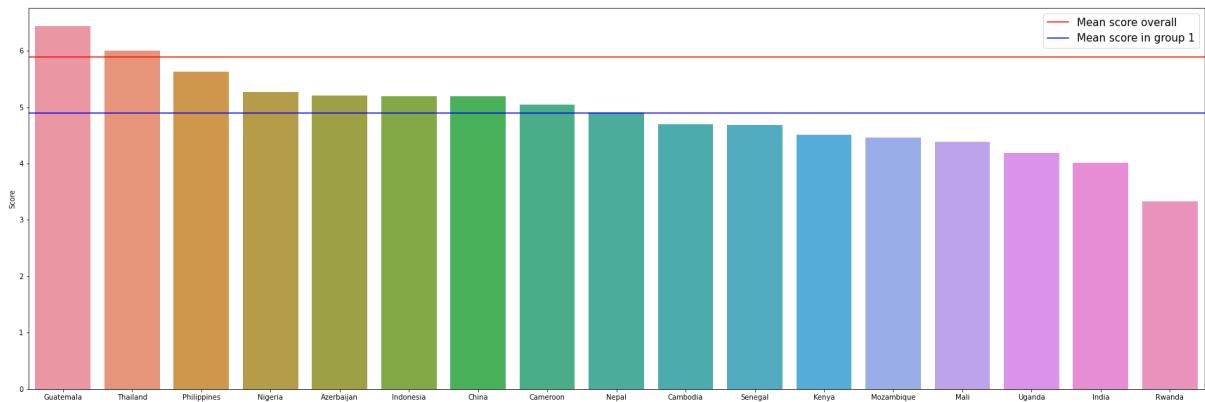
Z powyższego wykresu możemy odczytać, że najgorsze miejsca w większości zajmują kraje należące do grupy pierwszej. W analizie bottom 25 ponad połowa państw należy właśnie do tej grupy. Następnie aż do analizy bottom 15 pojawia się więcej obserwacji z grupy drugiej, ale zmienia się to podczas badania bottom 25, gdzie więcej wejść posiada grupa trzecia. Należy zauważyć jednak, że liczebność grupy pierwszej jest znacznie niższa od pozostałych, jako że znajduje się tam o około dziesięć obserwacji mniej.

Z powyższej analizy można wywnioskować, że kraje należące do danej grupy cechują się podobnymi wartościami współczynnika szczęścia. W krajach z grupy drugiej jakość życia jest na najwyższym poziomie, natomiast w państwach z grupy pierwszej na najniższym. Widocznym jest również, że grupa trzecia zajmuje głównie środkowe pozycje w rankingu.

### 6.3.2 Analiza współczynnika w grupie pierwszej

Skoro zauważono, że przynależność do grupy ma wpływ na zadowolenie z życia obywateli, to czym charakteryzują się poszczególne grupy?

Grupa pierwsza składa się z krajów, w których poziom szczęścia jest najniższy. Wynik poszczególnych państw można zobrazować za pomocą wykresu słupkowego.

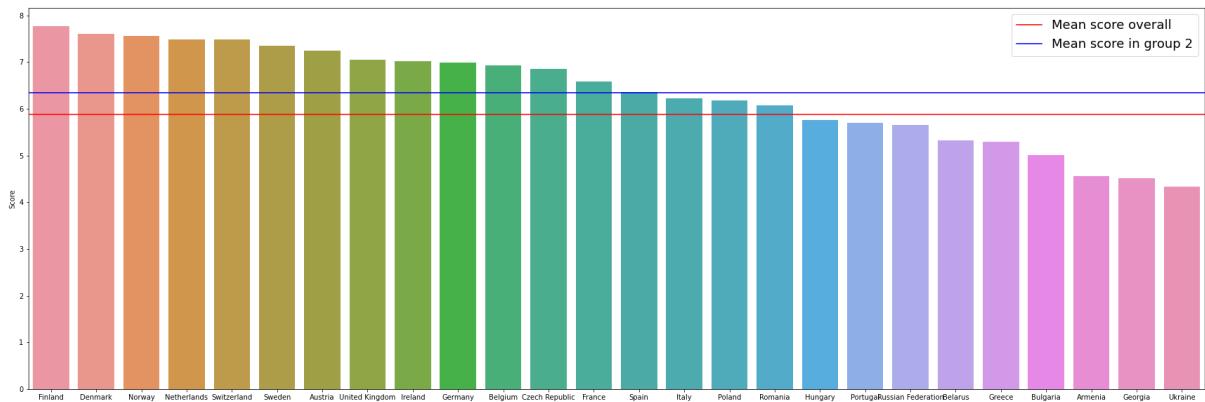


Na załączonym wykresie widocznym jest, że średni wynik w obrębie tej grupy wynosi 4.892 i jest znacznie mniejszy niż średnia całościowa z danego roku wynosząca 5.886. Dodatkowo zauważać można, że tylko dwie obserwacje posiadają wynik wyższy od średniej całego zbioru są to Guatemala i Tajlandia. Najniższy wynik w zbiorze, który jest jednocześnie najniższą wartością w danym roku, przypada Rwandzie.

	death rate	gdp growth	immunization	import	inflation	labor_women	labor growth	military_expenditure	urban_population growth	unemployment	Population in agglomerations dynamic
count	17.000000	17.000000	17.000000	17.000000	17.000000	17.000000	17.000000	17.000000	17.000000	17.000000	17.000000
mean	6.942176	4.794684	84.823529	34.720387	2.938729	44.389859	0.023620	1.566983	41.450412	3.566059	15.433635
std	1.804425	2.000644	13.634677	16.272727	2.447949	8.201331	0.013368	0.869228	14.318663	2.238492	6.818807
min	4.720000	2.208429	54.000000	17.480104	-0.241308	21.213170	-0.006438	0.459171	17.313000	0.147000	4.810101
25%	5.600000	3.475060	84.000000	20.957385	1.287452	40.197560	0.019597	1.086759	27.507000	2.190000	11.051462
50%	6.381000	4.756161	89.000000	36.122585	2.501055	45.710611	0.028854	1.376595	47.149000	3.466000	14.321416
75%	7.786000	6.118526	95.000000	40.458922	3.680606	49.464803	0.030447	1.769798	51.439000	4.850000	18.759187
max	11.630000	9.460682	99.000000	75.077972	10.384779	55.822213	0.042840	3.873412	60.308000	8.530000	28.437293

Z poniższej tabeli można odczytać, że kraje z grupy pierwszej charakteryzują się raczej dosyć niską śmiertelnością, inflacją i niskim współczynnikiem bezrobocia. Procent zaszczepionych niemowląt również jest dosyć niski, gdzie ponad połowa wartości znajduje się poniżej 90%, a średnia jest równa około 85%. Ludność tych krajów w większości nie kolonizuje miast, które w dodatku się nie rozrastają patrząc na wartości zmiennych urban\_population growth czy Population in agglomerations dynamic. Wartość importowanych dóbr w przeliczeniu na procent PKB również jest niska, chociaż w tym przypadku ma miejsce wysoki rozrzut danych przy wartości minimalnej równej 17,5% i maksymalnej wynoszącej aż 99%. Można przypuszczać, że niezadowolenie ludzi zamieszkujących kraje z grupy drugiej wynika z faktu, iż tempo wzrostu gospodarki tych państw jest bardzo powolne, co niesie za sobą stagnację i brak perspektyw na lepsze życie.

### 6.3.3 Analiza współczynnika w grupie drugiej

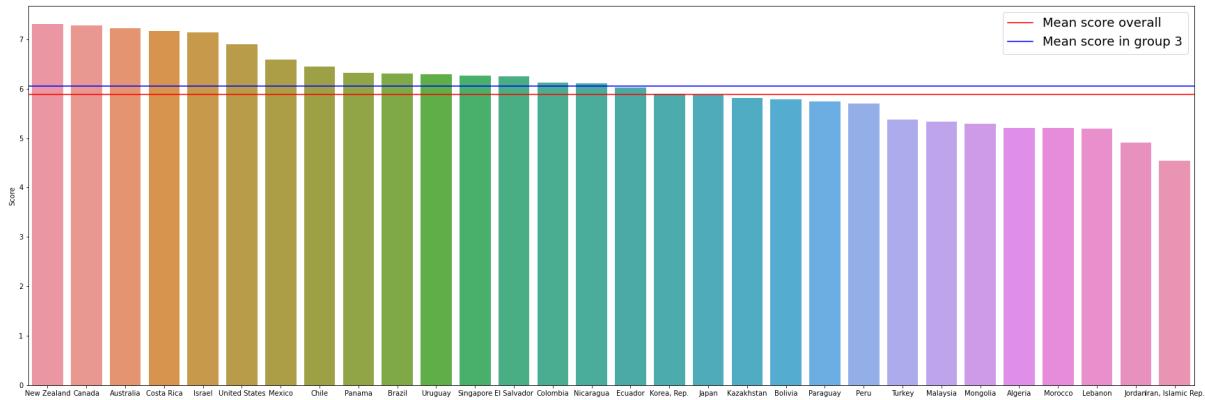


W grupie drugiej średni wynik wynosi 6.341, który jest znacznie wyższy od średniej wartości z danego roku wynoszącej 5.886. Krajami odstającymi od reszty są m.in. Ukraina, Gruzja, Armenia, Bułgaria, Grecja, Białoruś czy Rosja. Przy czym najniższa wartość należy do Ukrainy. Czołówka państw z tej grupy przoduje również w całym rankingu z danego roku. Mowa tutaj o takich krajach jak Finlandia, Dania, Norwegia czy Holandia.

	death rate	gdp growth	immunization	import	inflation	labor_women	labor growth	military_expenditure	urban_population growth	unemployment	Population in agglomerations dynamic
count	26.000000	26.000000	26.000000	26.000000	26.000000	26.000000	26.000000	26.000000	26.000000	26.000000	26.000000
mean	10.593308	2.658200	95.076923	52.405429	2.865517	46.487160	0.003551	1.776255	74.924423	6.595385	20.875100
std	2.283546	1.679527	2.827339	21.500686	2.416847	1.759890	0.009991	1.040980	11.138873	4.336596	8.375468
min	6.300000	0.410278	90.000000	20.907981	-0.472598	42.814771	-0.020224	0.279145	54.084000	2.010000	4.677758
25%	9.025000	1.536394	93.000000	40.038167	1.285108	45.452929	-0.002679	1.227986	66.691250	3.782500	15.765264
50%	10.167500	2.059087	95.500000	49.959510	2.035080	46.649628	0.004352	1.335321	74.967000	4.725000	21.489163
75%	12.518250	3.828070	97.000000	63.012178	4.141035	47.776013	0.010055	1.945080	82.139250	7.850000	25.408227
max	15.500000	7.600000	99.000000	124.407132	8.626733	49.276396	0.019854	4.768968	98.041000	18.299999	41.335731

Z powyższej tabeli możemy odczytać, że grupa ta w ogólności cechuje się wysokim współczynnikiem importu oraz dużą część społeczeństwa zamieszkuje metropole przez co można stwierdzić, że są to w większości kraje szybko rozwijające się. Można dodatkowo zauważyc, że w tych krajach około 46% pracowników stanowią kobiety, gdzie maksymalnie wartość dochodzi do prawie 50%. Ponadto w każdym z tych państw ponad 90% niemowląt jest szczepionych, gdzie maksymalna wartość wynosi nawet 99%, a tylko ćwierć badanych wartości nie przekracza 93%. Kraje te cechują się jednak najwyższym współczynnikiem śmiertelności oraz dosyć wysokim współczynnikiem bezrobocia dochodzącym nawet do około 18%, a jedna-czwarta danych osiąga wartość równą około 8%, co było najwyższą wartością w poprzedniej grupie. Warto dodać, że grupa ta osiąga najmniejsze wartości odchylenia standardowego.

#### 6.3.4 Analiza współczynnika w grupie trzeciej



Średni wynik w grupie trzeciej wynoszący 6.054 zbliżony jest bardzo do ogólnej średniej równej 5.886. Grupa ta jest najliczniejszą ze wszystkich. Posiada aż 30 obserwacji, z których co najmniej połowa ma wyniki bliskie średniej z danego roku. Najwyższe wartości należą odpowiednio do: Iranu, Jordanii, Libanu oraz Maroka. Najniższe wyniki przypadają natomiast Nowej Zelandii, Kanadzie, Australii czy Kostaryce.

	death rate	gdp growth	immunization	import	inflation	labor_women	labor growth	military_expenditure	urban_population growth	unemployment	Population in agglomerations dynamic
count	30.000000	30.000000	30.000000	30.000000	30.000000	30.000000	30.000000	30.000000	30.000000	30.000000	30.000000
mean	6.108167	1.208568	91.633333	37.168205	4.109900	39.491937	0.017101	2.079128	78.56610	6.741633	38.588031
std	1.550412	2.819709	7.227740	24.272450	7.052825	8.914447	0.010975	1.525342	11.19127	3.766495	18.349195
min	3.889000	-7.157247	73.000000	14.350575	-0.839325	17.659719	0.000449	0.000000	57.54000	2.400000	6.339446
25%	5.102000	0.485232	87.000000	24.047541	0.720725	38.480744	0.010530	1.040362	70.51625	3.802500	26.321362
50%	5.604500	1.917529	95.000000	31.168296	1.967374	42.202740	0.014209	1.650510	80.26000	5.185500	38.864729
75%	6.746500	2.351015	97.000000	43.009019	4.244135	45.597912	0.023710	2.826579	86.77175	9.790000	46.529662
max	11.100000	5.602244	99.000000	147.625337	36.910083	48.123254	0.047759	6.021046	100.00000	16.809999	100.000000

Kraje z grupy trzeciej podobnie jak te z grupy drugiej szybko się rozwijają o czym mogą świadczyć informacje na temat zamieszkiwania miast czy procent PKB odpowiadający importowi. Przy czym obie grupy charakteryzują się dużymi różnicami pomiędzy wartościami maksymalnymi i minimalnymi w danych kategoriach oraz można zauważyc dosyć duży rozrzut informacji względem wartości średniej. Możemy również dostrzec wysoki poziom zaszczepienia niemowląt. Skoro grupy druga i trzecia mają tak wiele wspólnego, to co może wskazywać na fakt, że ludność zamieszkująca kraje z grupy drugiej jest bardziej szczęśliwa? Warto zauważyć, że grupa trzecia cechuje się prawie dwukrotnie wyższym poziomem inflacji w jednym przypadku równym nawet 37%, wyższym współczynnikiem bezrobocia dochodzącym nawet do 17%, a ponad ćwierć danych przekracza wartość równą 9,7% oraz wyższymi wydatkami na wojsko.

## 7. Zakończenie

Utworzona klasteryzacja pozwoliła nam zauważyc rzeczywiste tendencje - państwa zostały pogrupowane w sposób oczekiwany. Koniecznym było jednak doprecyzowanie ze strony badaczy poprzez zastosowanie własnych doświadczeń i wiedzy w celu weryfikacji źle sklasyfikowanych państw. Nie udało nam się jednak pogrupować państw na kraje rozwinięte, rozwijające się oraz słabo rozwinięte, jako że brane pod uwagę były również czynniki socjologiczne i polityczne, które nie są w stanie w pełni odwzorować faktycznego rozwoju gospodarczego kraju. W pracy skupiliśmy się głównie na tempie rozwoju analizowanych państw, zamieniając większość zmiennych na zmienne dynamiczne.

### Bibliografia:

1. [Labor force, female \(% of total labor force\)](#) - procent kobiet wśród grupy aktywnie zawodowo
2. [Labor force, total](#) - siła robocza
3. [Population, total](#) - wielkość populacji
4. [Unemployment, total \(% of total labor force\) \(modeled ILO estimate\)](#) - procent bezrobocia
5. [Immunization, measles \(% of children ages 12-23 months\)](#) - procent zaszczepienia na odrę wśród dzieci
6. [Death rate, crude \(per 1,000 people\)](#) - wskaźnik śmiertelności
7. [Fertility rate, total \(births per woman\)](#) - współczynnik dzietności
8. [GDP growth \(annual %\)](#) - wzrost PKB
9. [Inflation, GDP deflator \(annual %\)](#) - inflacja
10. [Population growth \(annual %\)](#) - % wzrost populacji
11. [Mortality rate, under-5 \(per 1,000 live births\)](#) - wskaźnik śmiertelności wśród dzieci poniżej 5 lat
12. [Urban population \(% of total population\)](#) - % ludności miejskiej
13. [Literacy rate, youth total \(% of people ages 15-24\)](#) - % analfabetyzmu wśród dzieci
14. [Exports of goods and services \(% of GDP\)](#) - eksport towarów i usług
15. [Imports of goods and services \(% of GDP\)](#) - import towarów i usług
16. [Proportion of seats held by women in national parliaments \(%\)](#) - procent miejsc zajmowanych przez kobiety w parlamencie krajowym
17. [Military expenditure \(% of GDP\)](#) - wydatki wojskowe (% PKB)
18. [Taxes on income, profits and capital gains \(% of revenue\)](#) - podatki od dochodu %
19. [Population in urban agglomerations of more than 1 million \(% of total population\)](#) - % populacji mieszkających w aglomeracjach dużych miejskich (ponad 1 mln mieszkańców)
20. [World Happiness Report up to 2022](#) - wskaźnik szczęścia - nie ma
21. [Urban population growth \(annual %\)](#) - procentowy wzrost ludzi zamieszkujących miasta
22. [Listed domestic companies, total](#) - Krajowe spółki giełdowe ogółem
23. [Happiness report](#) - Współczynnik szczęścia