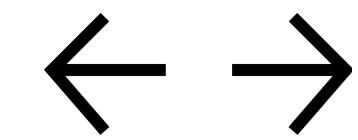


Wielowymiarowa Analiza i Profilowanie Państw Świata



Źródło i pochodzenie danych

- **Kurs:** Metody Profilowania Użytkownika w Środowiskach Inteligentnych
- **Rok akademicki:** 2025/2026, semestr zimowy
- **Grupa:** 3 (Piątek, 15:15)
- **Autor:** Dawid Boratyński
- **Indeks:** 268094

Cel projektu

Główym celem projektu jest stworzenie zaawansowanego systemu profilowania państw na podstawie danych statystycznych, wykraczającego poza tradycyjne, jednowymiarowe rankingi oparte wyłącznie na PKB.

Projekt wykorzystuje algorytmy uczenia maszynowego (nienadzorowane uczenie – algorytm K-Means) do automatycznego grupowania krajów w klastry o podobnej charakterystyce w pięciu kluczowych sektorach: Ekonomia, Edukacja, Zdrowie, Infrastruktura oraz Kwestie Społeczne. Efektem końcowym jest interaktywny dashboard analityczny, pozwalający na identyfikację globalnych wzorców (np. *Global Hubs, Developing Economies*).

Źródło i pochodzenie danych

Podstawowym źródłem danych wykorzystanych w projekcie jest baza World Bank Open Data (Bank Światowy), a konkretnie zbiór *World Development Indicators (WDI)*. Jest to najbardziej kompletna i autorytatywna globalna baza danych dotycząca rozwoju światowego, gromadząca statystyki z oficjalnie uznanych źródeł międzynarodowych.

Metodyka pobierania i przetwarzania

Proces pozyskiwania danych został zautomatyzowany przy użyciu języka Python oraz biblioteki pandas_datareader, która łączy się bezpośrednio z API Banku Światowego. Pozwoliło to na pobranie najświeższych dostępnych informacji bez konieczności ręcznego kopiowania plików.

Przegląd literatury i podstawy teoretyczne

Fundamentem teoretycznym przyjętego modelu jest koncepcja socjologiczna i ekonomiczna opracowana przez Immanuela Wallersteina, teoria Systemów-Światów (World-Systems Theory). Dzieli ona globalną gospodarkę na trzy strefy:

1. Centrum (Core): Kraje wysokorozwinięte, dominujące technologicznie i kapitałowo (w projekcie zidentyfikowane jako klastry *Developed* oraz *Independent Core*).
2. Peryferia (Periphery): Kraje dostarczające surowce i tanią siłę roboczą, uzależnione od inwestycji z zewnątrz (klastry *Struggling* oraz *Dependent Periphery*).
3. Półperiferia (Semi-periphery): Kraje aspirujące, łączące cechy obu grup, często będące hubami produkcyjnymi (klastry *Developing* oraz *Investment Hubs*).

W projekcie zastosowano tę teorię wprost, implementując Wskaźnik Zależności Kapitałowej (*Capital Dependency Ratio*) który bada relację napięcia inwestycji

Bibliografia

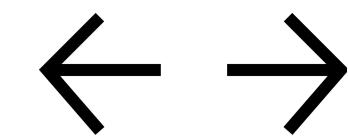
Wallerstein, I. (1974). *The Modern World-System I: Capitalist Agriculture and the Origins of the European World-Economy in the Sixteenth Century*. New York: Academic Press.
(Baza teoretyczna podziału na Centrum i Peryferia).

World Bank. (2023). *World Development Indicators (WDI) Database*. Washington, D.C.
(Źródło danych surowych).

Stiglitz, J. E., Sen, A., & Fitoussi, J. P. (2009). *Report by the Commission on the Measurement of Economic Performance and Social Progress*. (Krytyka PKB jako jedynej miary rozwoju).

Halkidi, M., Batistakis, Y., & Vazirgiannis, M. (2001). *On Clustering Validation Techniques*. Journal of Intelligent Information Systems.
(Podstawy metodologiczne walidacji klastrów użyte w doborze liczby grup).

United Nations Development Programme. (2022). *Human Development Report 2021-22: Uncertain Times, Unsettled Lives*. (Kontekst porównawczy dla wskaźników edukacyjnych i zdrowotnych).



Indicators explanation

Dobór wskaźników:

Zdefiniowano słownik kluczowych parametrów obejmujący 4 głównych sektorów:

- Ekonomia (np. PKB per capita, Inflacja, FDI)
- Edukacja (np. Partycypacja w szkolnictwie wyższym)
- Zdrowie (np. Oczekiwana długość życia, Wydatki na zdrowie)
- Infrastruktura (np. Dostęp do prądu i internetu)

Filtracja jednostek:

Surowe dane Banku Światowego zawierają również agregaty regionalne (np. "Arab World", "European Union", "High income"). W procesie czyszczenia (Data Cleaning) zdecydowano pozostawić te recordy w celu pokazania szerszej perspektywy.

Okres czasowy i obsługa braków:

Pobrano dane dla lat 2020–2022. Ze względu na różny czas raportowania przez poszczególne kraje, zastosowano algorytm wybierający najnowszą dostępną wartość dla każdego kraju. Jeśli kraj nie podał danych za 2022 rok, system automatycznie pobierał wartość z 2021 lub 2020 roku. Warości wzrostu ekonomicznego są uśrednionymi wartościami z tych lat

Interpretacja zmiennych i logika biznesowa

W celu nadania danym kontekstu analitycznego, do surowych liczb dołączono warstwę interpretacyjną (zawartą w pliku Indicators Explanation.csv).

Interpretacja ta nie pochodzi bezpośrednio z Banku Światowego, lecz została opracowana na potrzeby tego projektu w oparciu o:

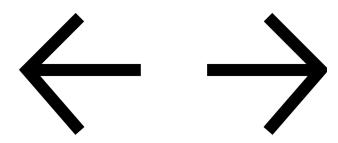
- Teorię ekonomii rozwoju (np. podział na kraje rozwinięte i rozwijające się wg struktury PKB).
- Logikę klastrowania: Zdefiniowano progi interpretacyjne (np. "Wysoki", "Niski", "Umiarkowany"), które pomagają zrozumieć charakterystykę wygenerowanych klastrów (np. *Global Hub vs Struggling*).

Dla każdego wskaźnika przygotowano:

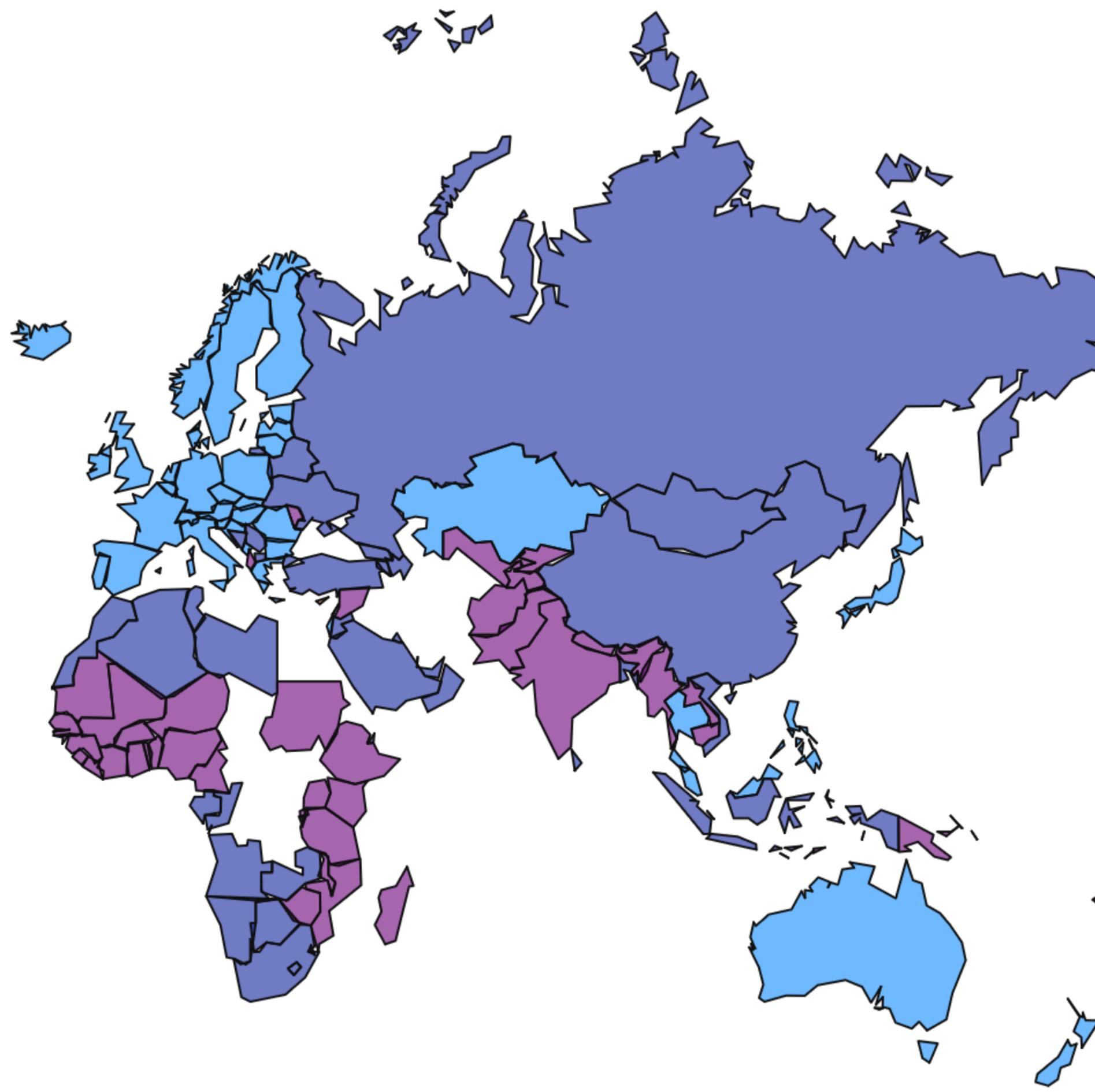
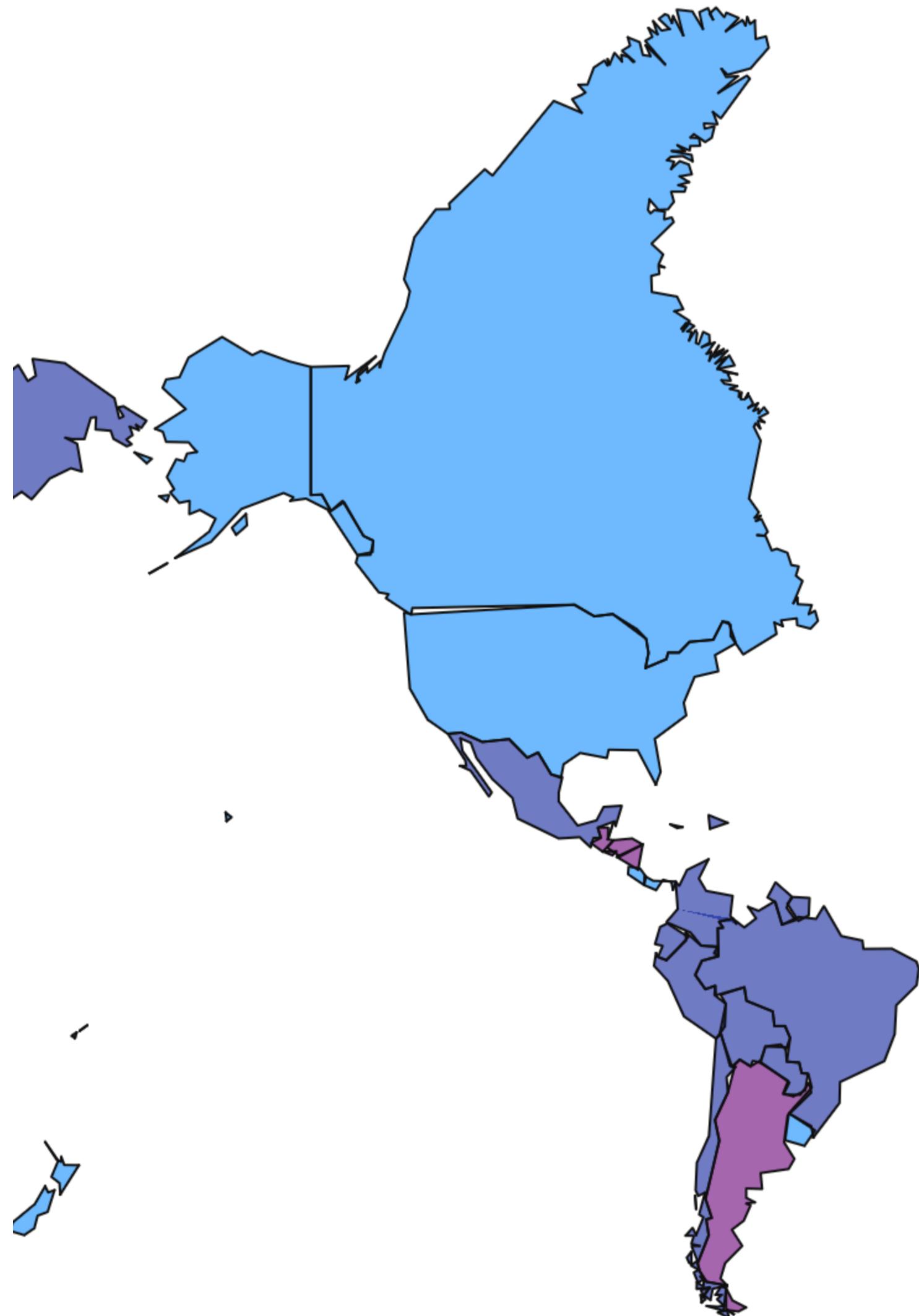
- Definicję: Wyjaśnienie, co dokładnie mierzy dany parametr.
- Jednostkę: (%), USD, lata).
- Wpływ na klasyfikację: Opis, jak wysoka lub niska wartość wpływa na ocenę kondycji państwa (np. wysoki udział rolnictwa jest sygnałem gospodarki rozwijającej się, podczas gdy wysoki udział usług charakteryzuje gospodarki rozwinięte).

Nazwa zmiennej (Kod)	Nazwa wskaźnika	Jednostka	Definicja / Wyjaśnienie	Interpretacja w analizie (Klastry)
Remittances_GDP	Przekazy od emigrantów	% PKB	Zależność od emigracji. Pieniądze wysyłane przez pracowników z zagranicy do domu.	Wysoki: Kraj "Eksportujący ludzi" (często Europa Wsch., Azja Płd.). Gospodarka zależna od zagranicy.
Inflation	Inflacja (CPI)	% (rocznie)	Wzrost cen towarów i usług w gospodarce.	Umiarkowana (2-5%): Zdrowa gospodarka. Wysoka (>10%): Kryzys, niestabilność waluty (częste w grupie Struggling).
Agriculture_Share_GDP	Udział rolnictwa w PKB	% PKB	Wskaźnik rozwoju strukturalnego (ile gospodarki opiera się na uprawie roli).	Wysoki: Gospodarka tradycyjna, często biedniejsza (Developing/Struggling). Niski: Gospodarka nowoczesna (Developed).
Population_Ages_65_plus	Odsetek seniorów (65+)	% populacji	Udział osób starszych w społeczeństwie. Mówią o etapie demoograficznym kraju.	Wysoki: Społeczeństwo starzejące się, obciążenie systemu emerytalnego i zdrowotnego (Developed). Niski: Młoda populacja.

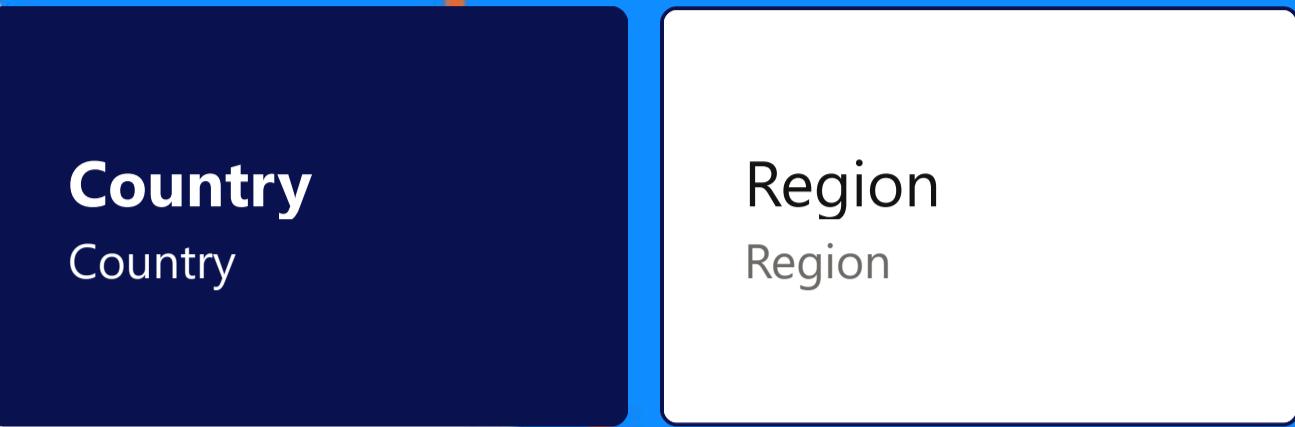
Economy - clustered map



Cluster_Name ● Developed ● Developing ● Economy Hub ● Struggling

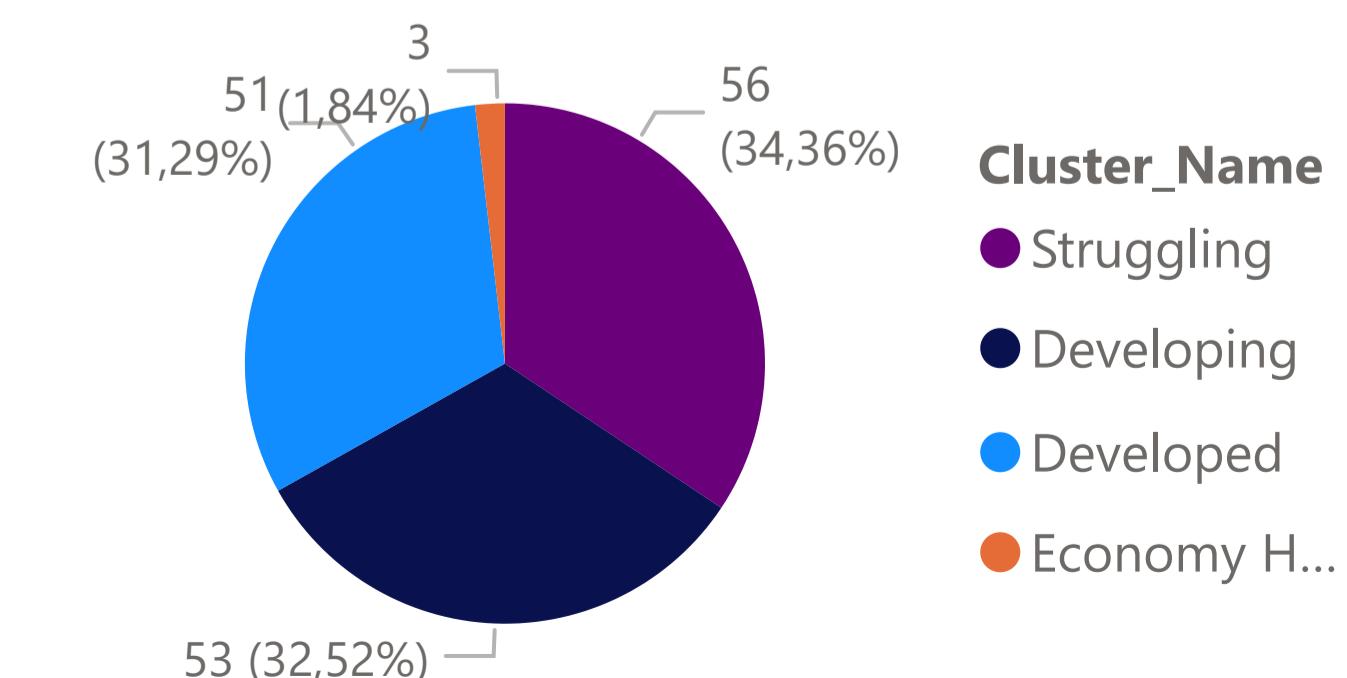


Liczba elementów według kategorii

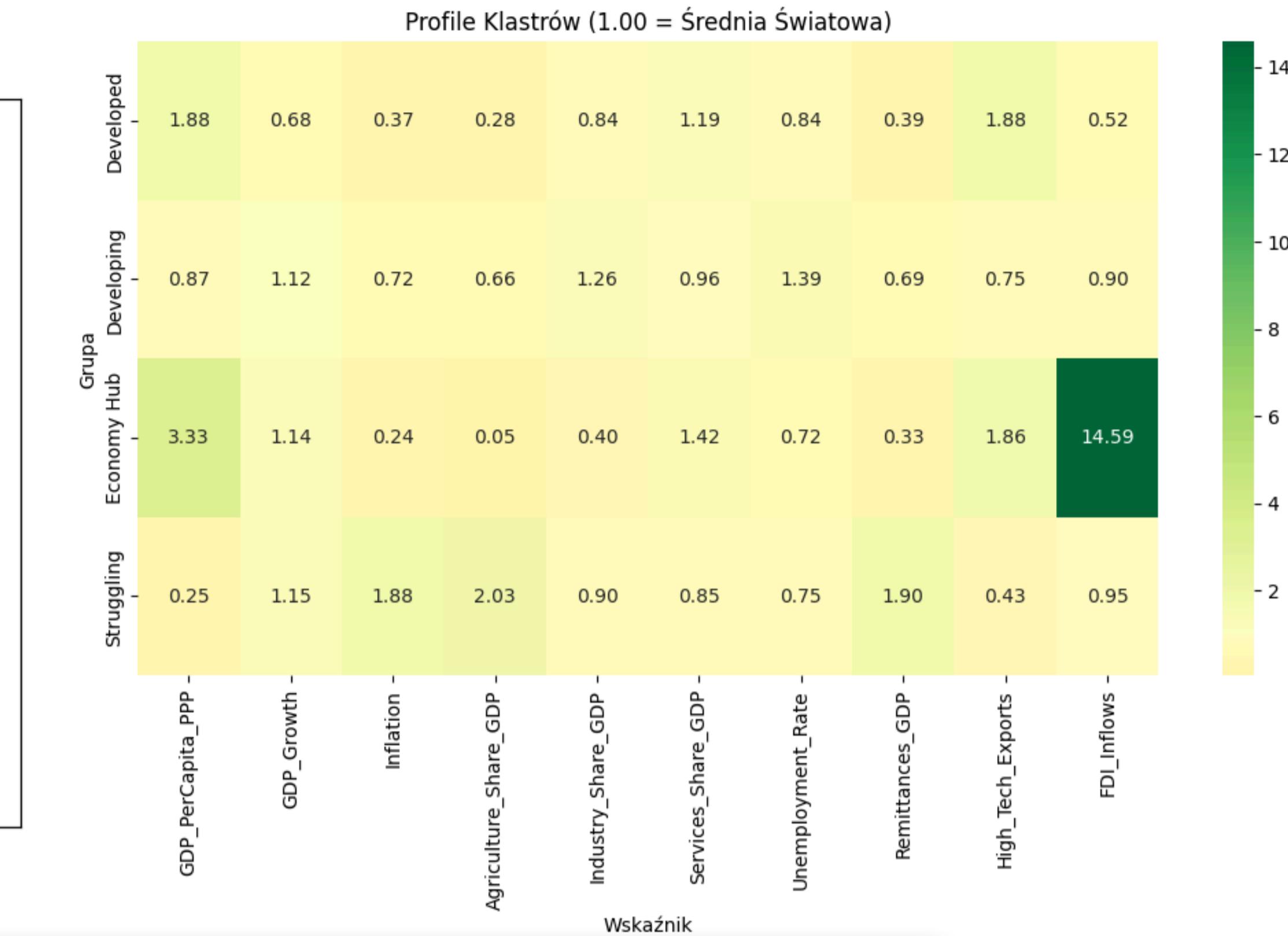
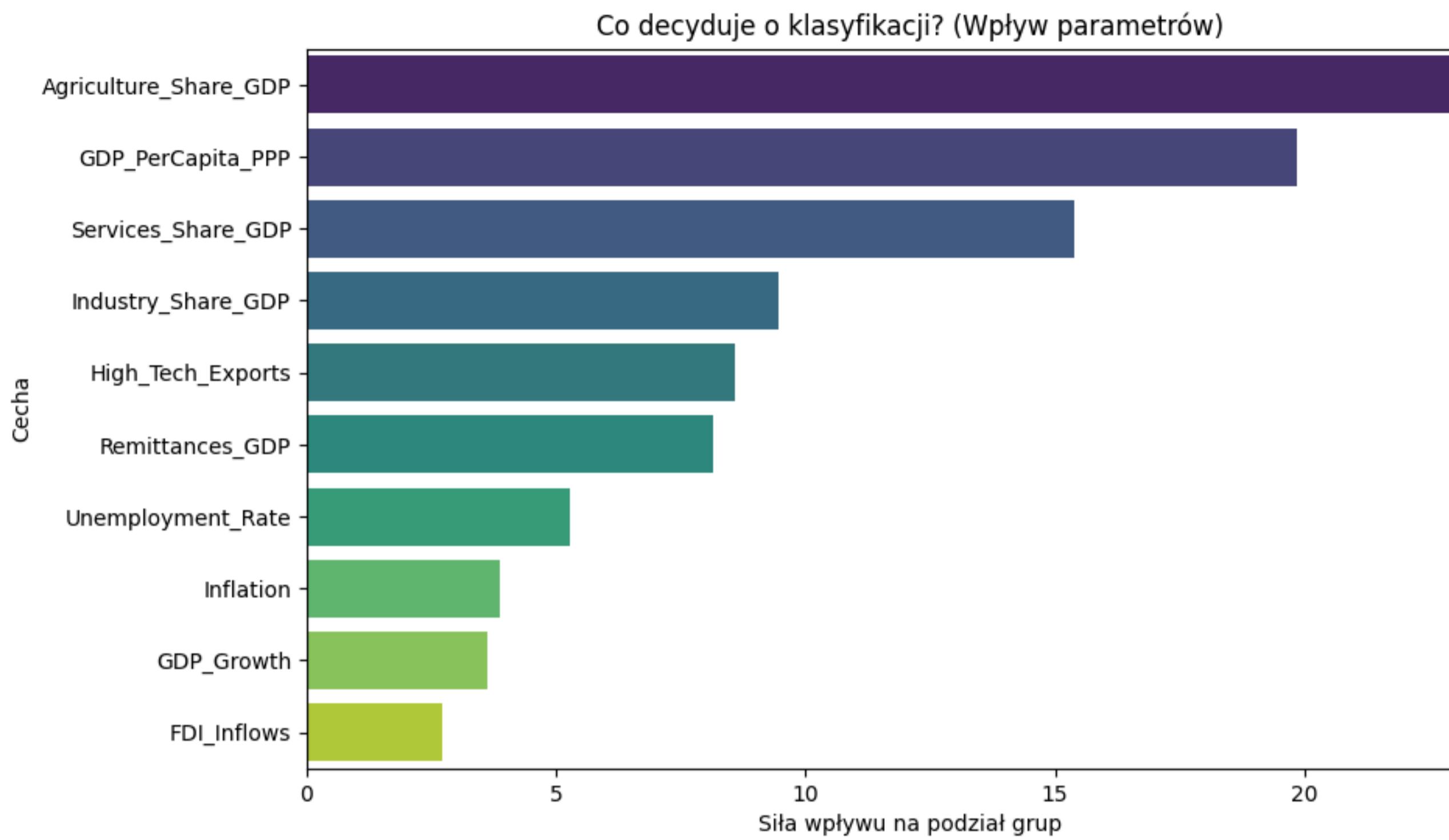
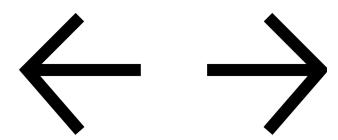


country	Cluster_Name
Afghanistan	Struggling
Albania	Struggling
Algeria	Developing
Angola	Developing
Argentina	Struggling
Armenia	Developing
Australia	Developed
Austria	Developed
Azerbaijan	Developing
Bahamas, The	Developed
Bangladesh	Developing

Liczba krajów według z kategorii



Economy - indicators



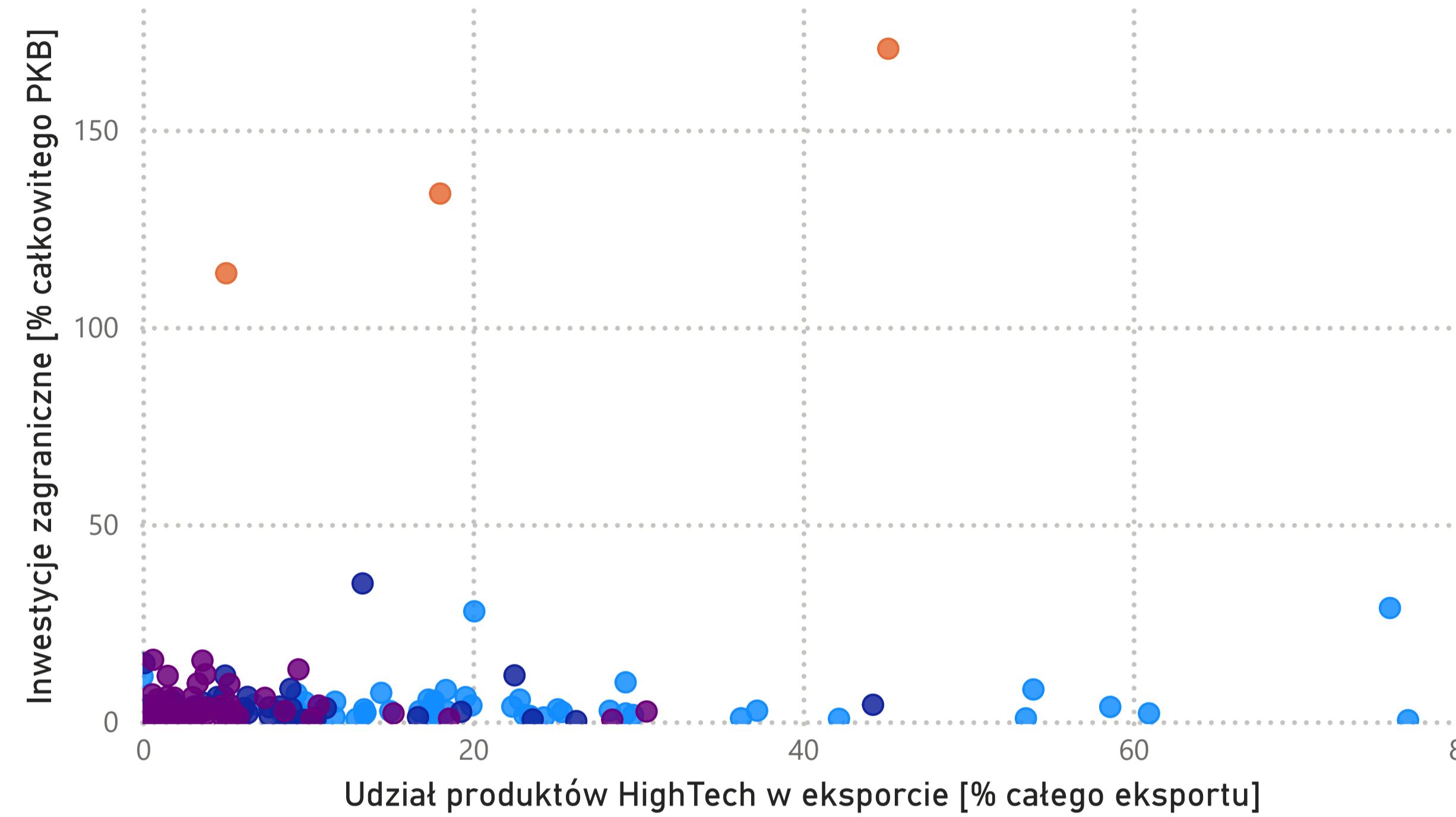
country	Agriculture Share GDP	GDP PerCapita	Services Share GDP	Industry Share GDP	High Tech Exports	Remittances GDP	Unemployment Rate	Inflation	GDP Growth	FDI Inflows
Uzbekistan	18,33	11 878,55	45,19	31,81	1,73	14,42	4,49	9,63	6,50	2,48
Georgia	5,49	28 285,07	61,98	19,58	3,74	11,87	11,48	1,11	9,68	4,69
Bosnia and Herzegovina	4,53	22 955,76	56,56	22,43	6,16	10,55	10,72	1,69	2,97	3,39
Moldova	7,11	18 614,99	62,27	16,78	2,93	10,54	1,43	4,68	0,10	2,52
Montenegro	3,49	34 225,87	63,77	12,18	14,46	10,34	14,10	3,34	3,16	7,25
Albania	15,45	26 693,19	48,69	22,40	1,53	8,41	10,25	2,22	4,05	6,32
Croatia	2,90	48 816,95	60,84	19,19	11,67	7,25	5,24	2,97	3,83	5,02
Serbia	3,17	32 233,76	58,78	23,07	4,55	6,40	7,39	4,67	3,95	6,20

- Category
 - Country
 - Region
- Cluster_Name
 - Developed
 - Developing
 - Economy Hub
 - Struggling

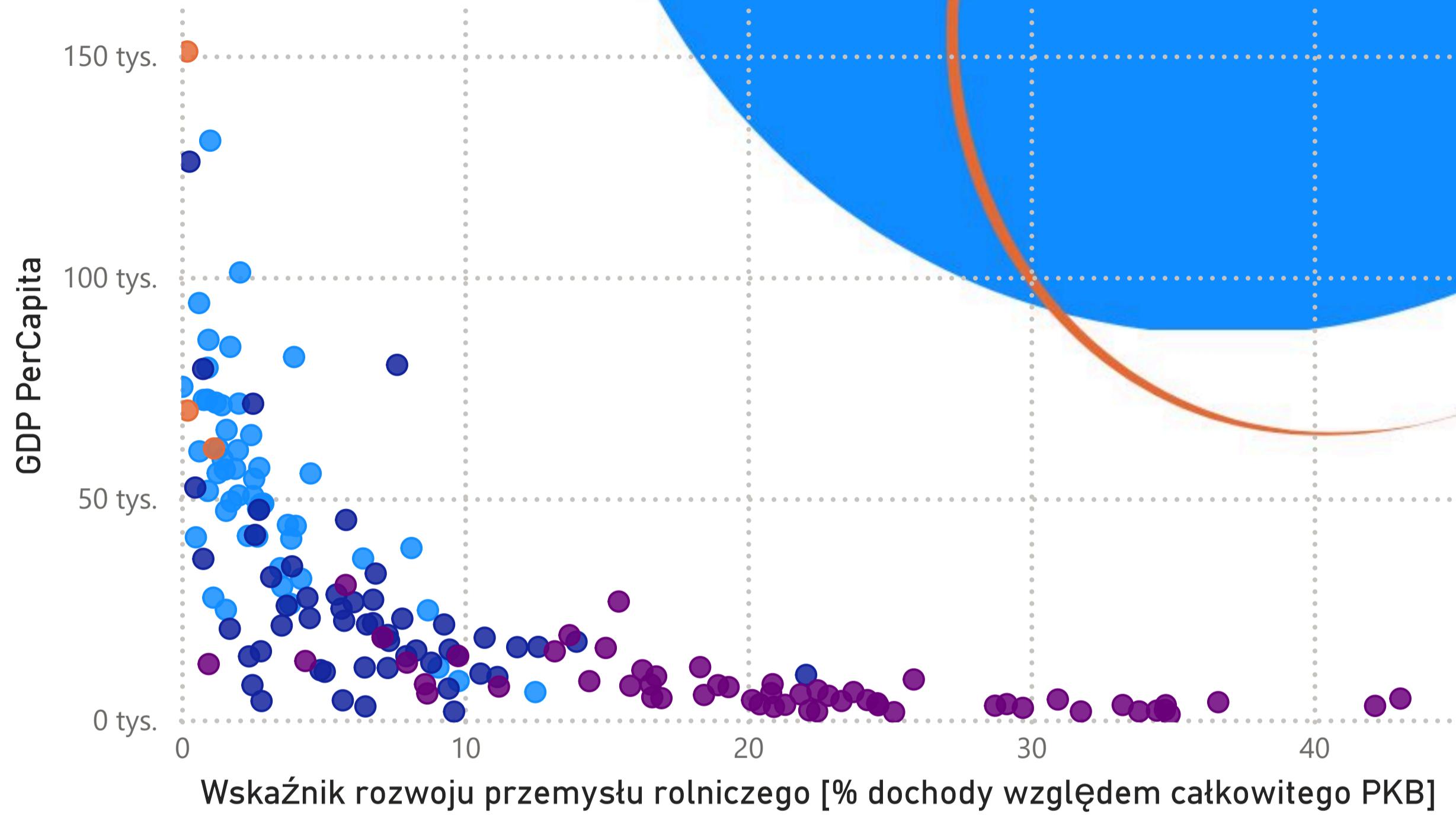
Economy - summary



Cluster_Name ● Developed ● Developing ● Economy Hub ● Struggling



Cluster_Name ● Developed ● Developing ● Economy Hub ● Struggling



	Developed	Developing	Economy Hub	Struggling
GDP PC MEDIAN	51 685,04	19 144,73	69 832,19	5 497,48
GDP Growth AVG	2,13	3,50	3,70	3,59
Inflation AVG	2,77	6,20	1,83	16,59
High Tech Exports AVG	22,94	7,05	22,76	4,74
Unemployment Rate AVG	5,62	10,35	4,75	4,97
FDI Inflows ABS AVG	4,46	3,56	139,33	3,57
Agriculture Share GDP AVG	2,92	6,42	0,53	21,37
Services Share GDP AVG	66,38	53,31	79,81	47,77
Industry Share GDP AVG	21,48	32,85	10,49	22,78
Promittances GDP AVG	0,01	1,01	1,51	5,12

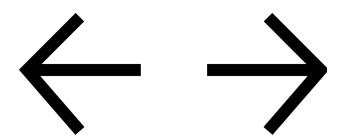
Strukturalny podział bogactwa: Analiza potwierdza silną korelację między zamożnością a strukturą PKB. Kraje klastra **Developed** opierają się na usługach (>70% PKB), podczas gdy klaster **Struggling** pozostaje uzależniony od rolnictwa.

Anomalia Global Hubs: Wykryto specyficzną grupę małych państw (np. Irlandia, Singapur), które przyciągają nieproporcjonalnie wysoki kapitał zagraniczny (FDI), pełniąc rolę "Rajów podatkowych" dla globalnego systemu finansowego.

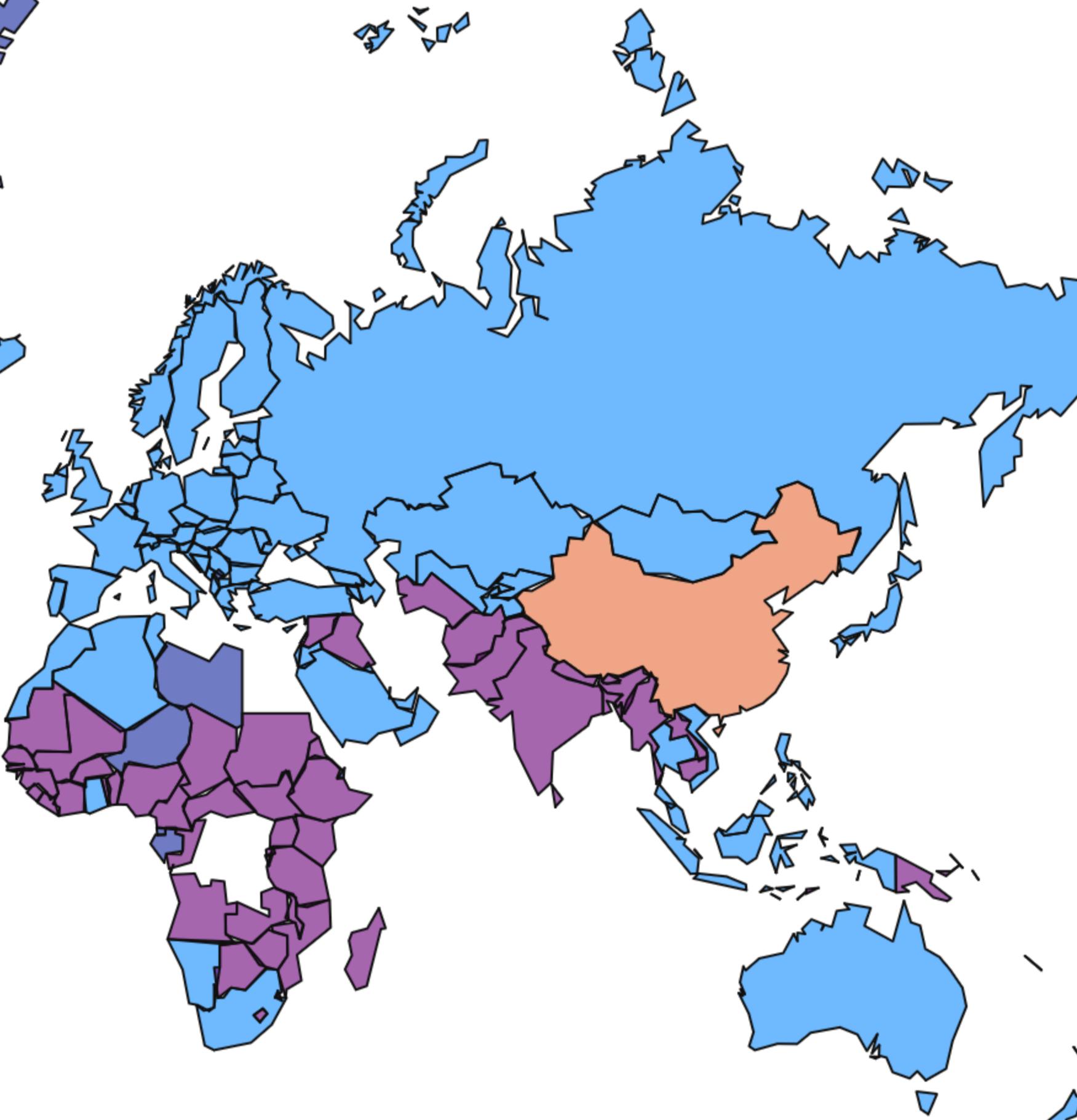
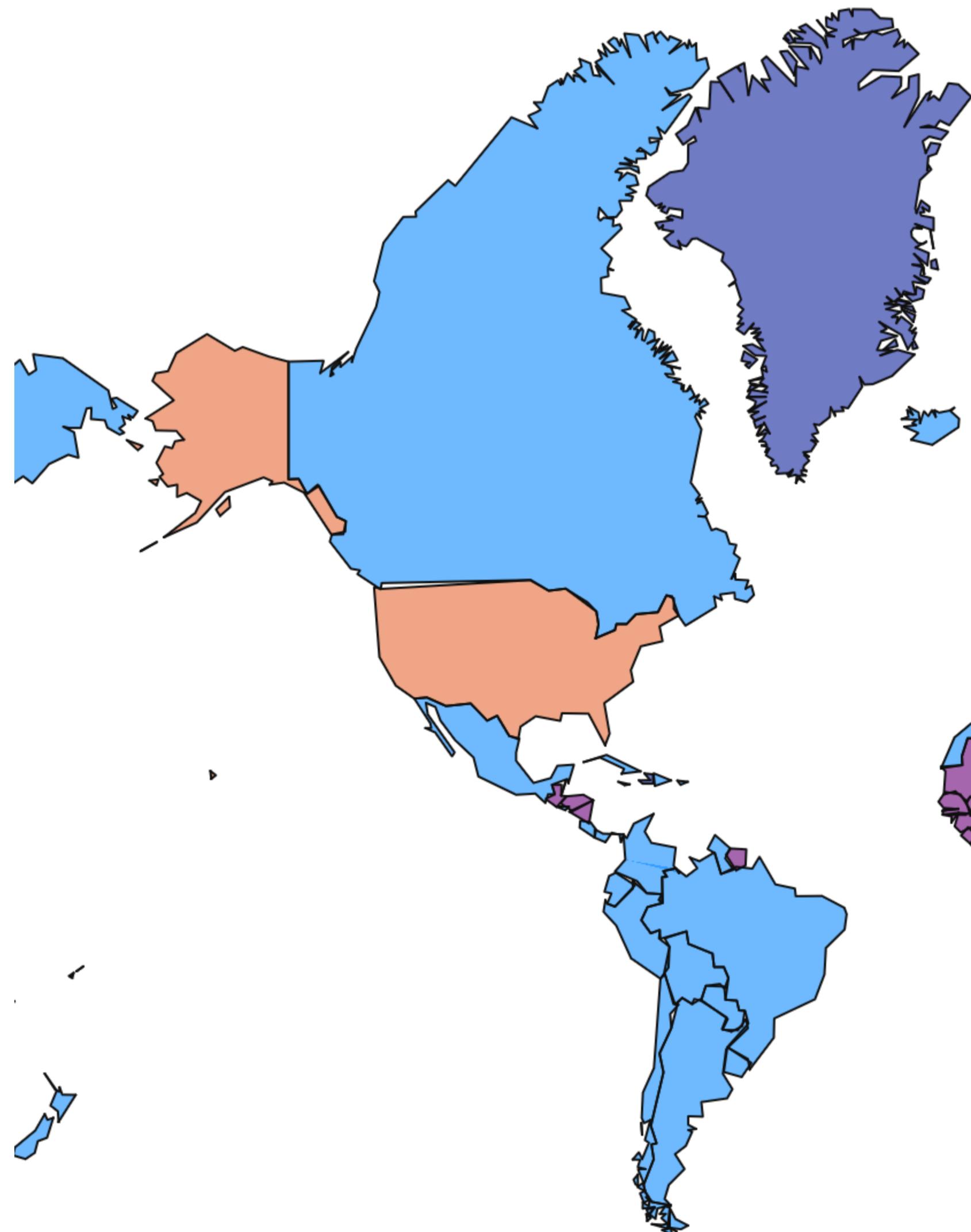
Niezależność vs Zależność: Analiza przepływów kapitałowych wyodrębniała **Niezależne Centrum** oraz **Zależne Peryferia**, gdzie rozwój jest warunkowany decyzjami inwestorów zewnętrznych.

Paradoks wzrostu: Najwyższe tempo wzrostu notują kraje z grupy **Developing**, co wynika z efektu niskiej bazy, jednak nie przekłada się to jeszcze na skokową poprawę jakości życia (PKB per capita).

Education - clustered map



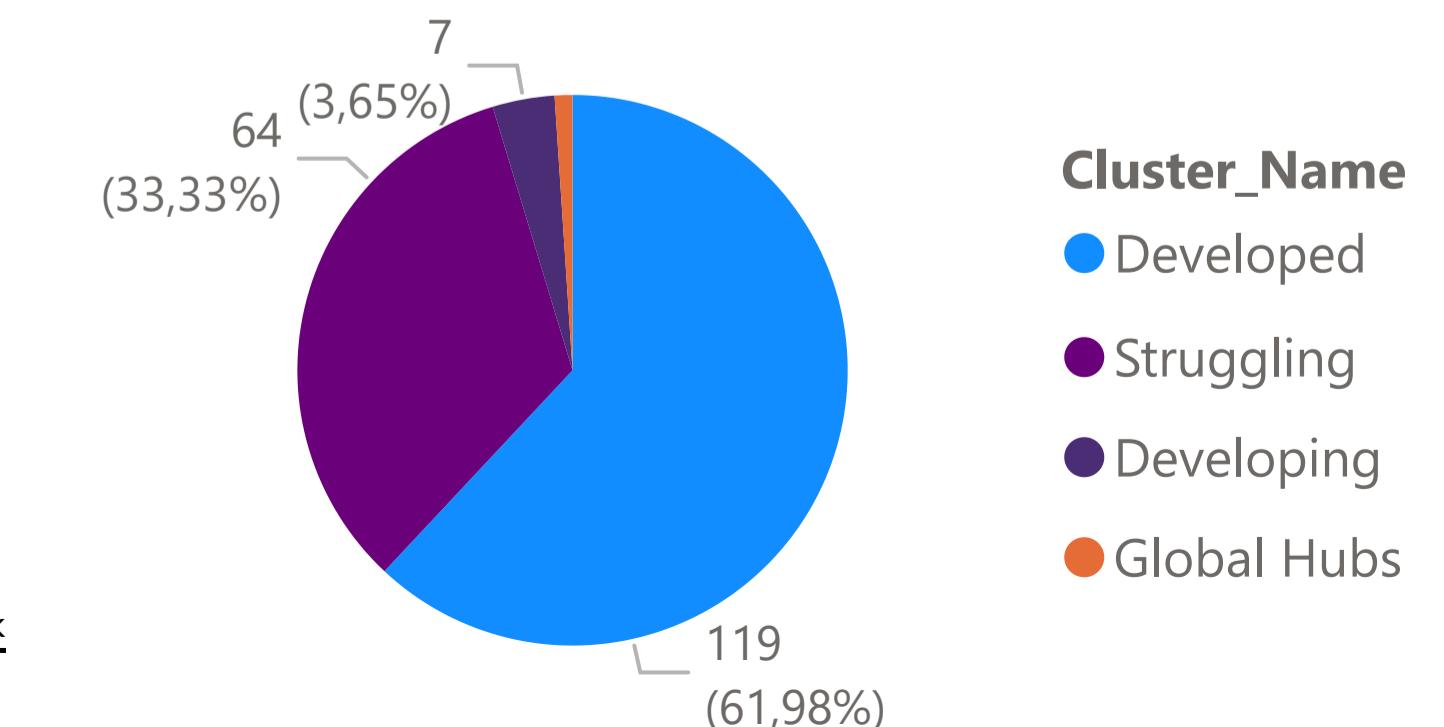
Cluster_Name ● Developed ● Developing ● Global Hubs ● Struggling



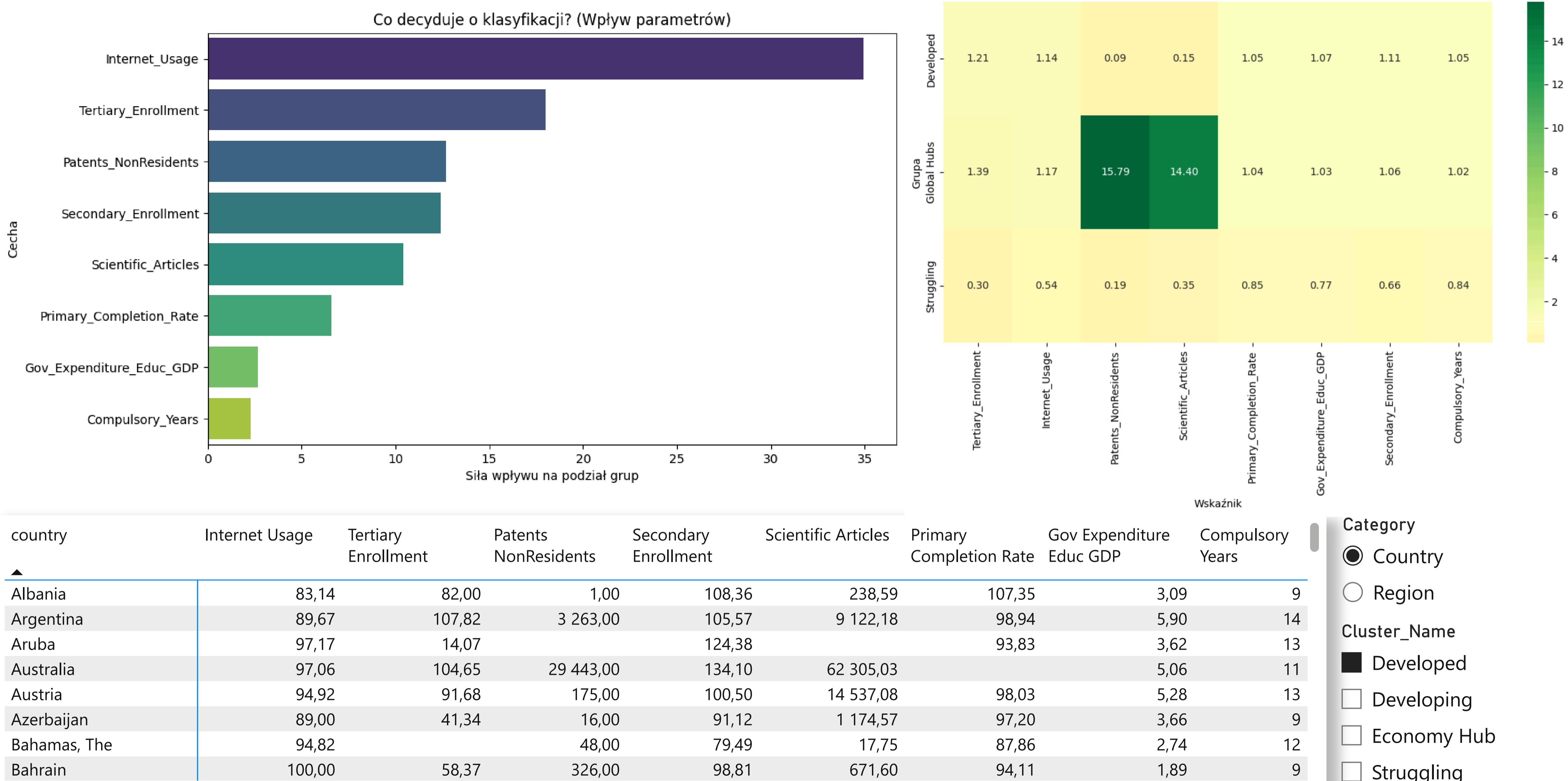
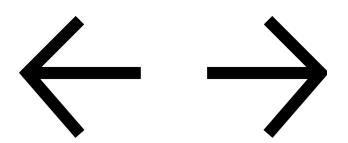
Liczba elementów według kategorii

Country	Region	Cluster_Name
Country	Region	
country		Cluster_Name
Afghanistan		Struggling
Albania		Struggling
Algeria		Developing
Angola		Developing
Argentina		Struggling
Armenia		Developing
Australia		Developed
Austria		Developed
Azerbaijan		Developing
Bahamas, The		Developed
Bangladesh		Developing

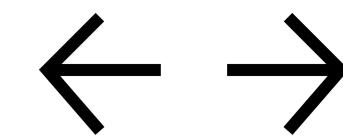
Liczba krajów według z kategorii



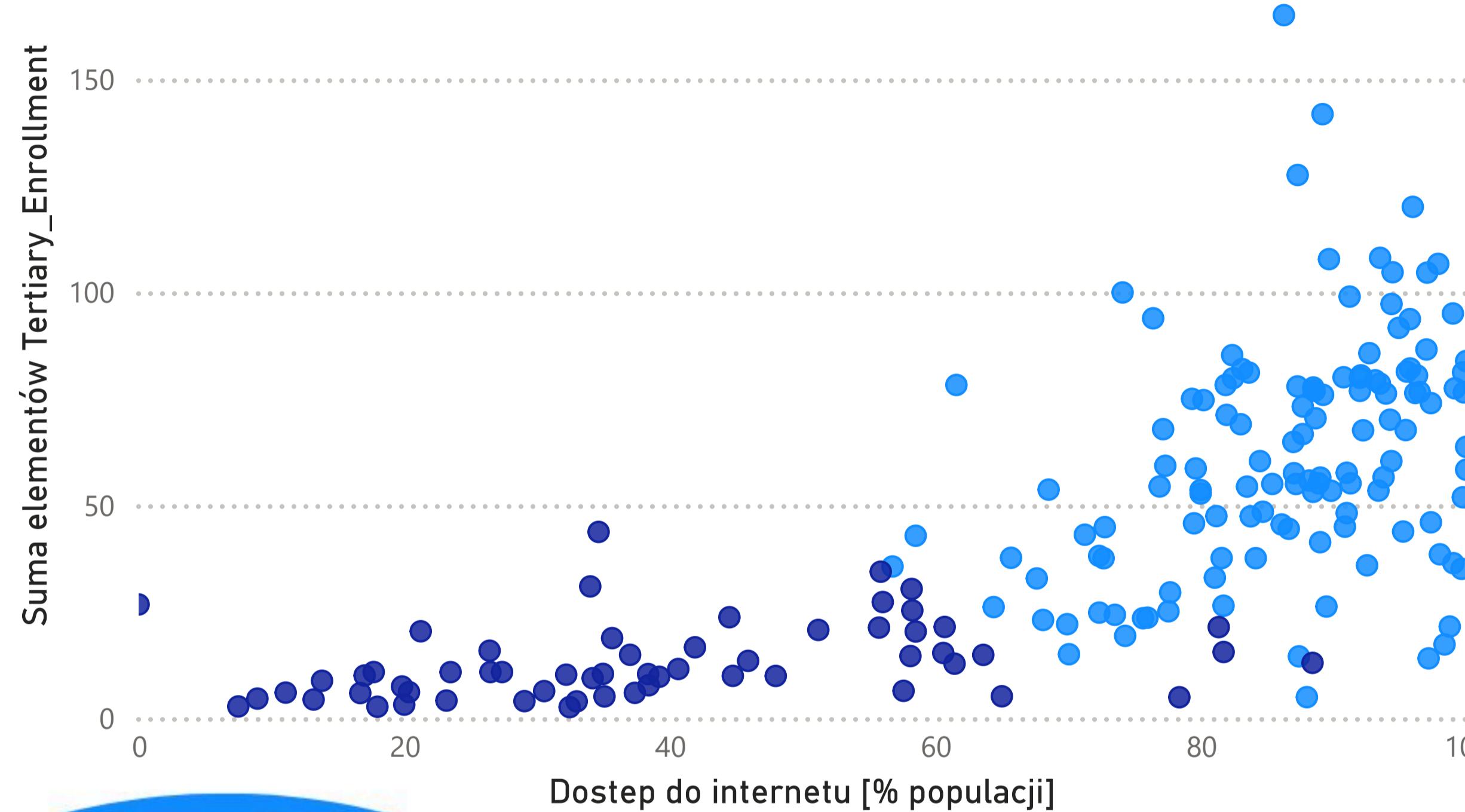
Education - indicators



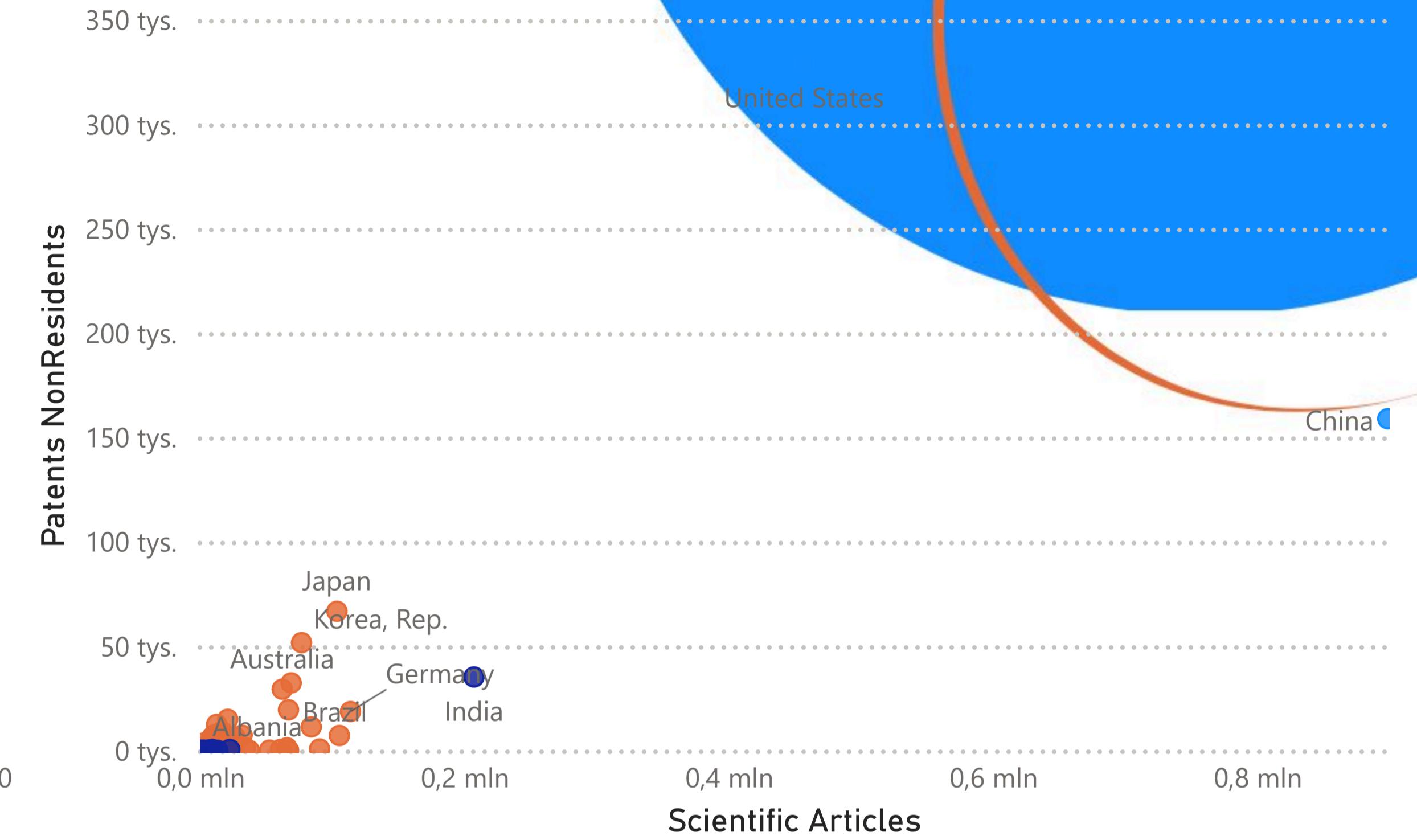
Education - summary



Cluster_Name ● Developed ● Developing ● Global Hubs ● Struggling



Cluster_Name ● Developed ● Developing ● Global Hubs ● Struggling



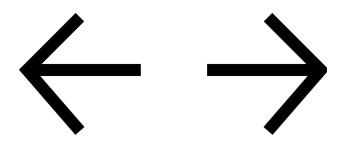
	Developed	Developing	Global Hubs	Struggling
Internet Usage AVG	86,36	66,57	92,57	38,81
Compulsory Years AVG	10,92	6,13	10,50	8,64
Tertiary Enrollment AVG	61,47	4,66	78,12	14,04
Secondary Enrollment AVG	99,44	41,77	94,87	62,12
Scientific Articles MEDIAN	2 187,40	73,56	678 142,17	156,66
Patents NonResidents MEDIAN	190,00	14,50	244 124,00	22,00
Primary Completion Rate	96,60	36,70	99,60	77,67

Wykluczenie Cyfrowe: Kluczową barierą rozwojową jest dostęp do sieci. W klastrze **Developed** wskaźnik *Internet Usage* przekracza 90%, podczas gdy w **Struggling** cyfryzacja jest luksusem, co trwale wyklucza te społeczeństwa z globalnej wymiany wiedzy.

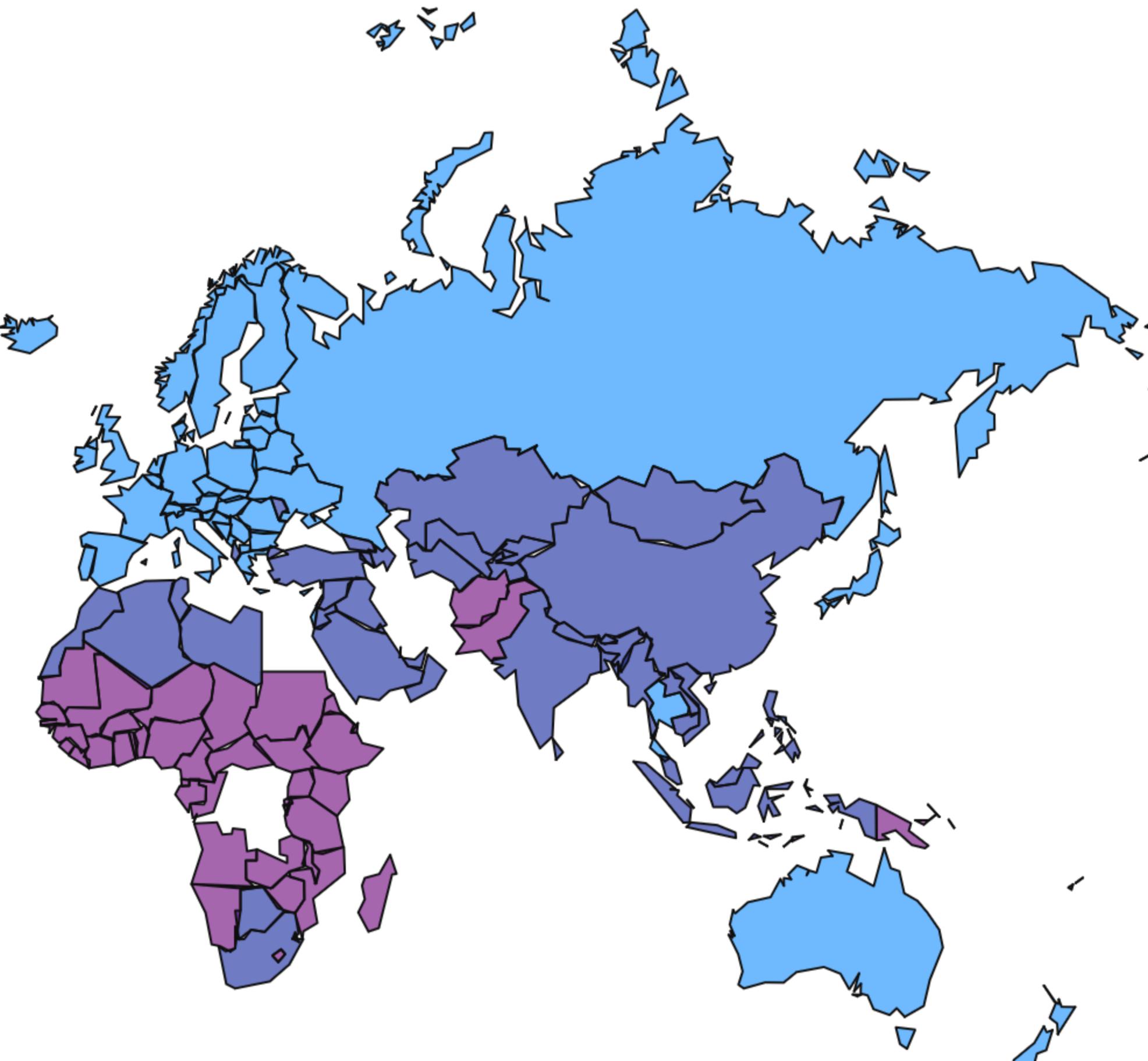
Szkłany sufit edukacji wyższej: Kraje **Developing** skutecznie nadrobiły zaległości w edukacji podstawowej (*Primary Completion* ~90%), jednak przepaść pojawia się na poziomie studiów. Wysoki wskaźnik *Tertiary Enrollment* (>70%) jest cechą ekskluzywną dla krajów bogatych, które przyciągają wielu studentów z całego świata, tworząc **Globalne Huby** wiedzy.

Monopol na innowacje: Analiza liczby patentów (*Patents NonResidents*) i artykułów naukowych ujawnia brutalną prawdę o systemie-świata: "Peryferia" konsumują technologie, a "Centrum" je tworzy.

Healthcare - clustered map



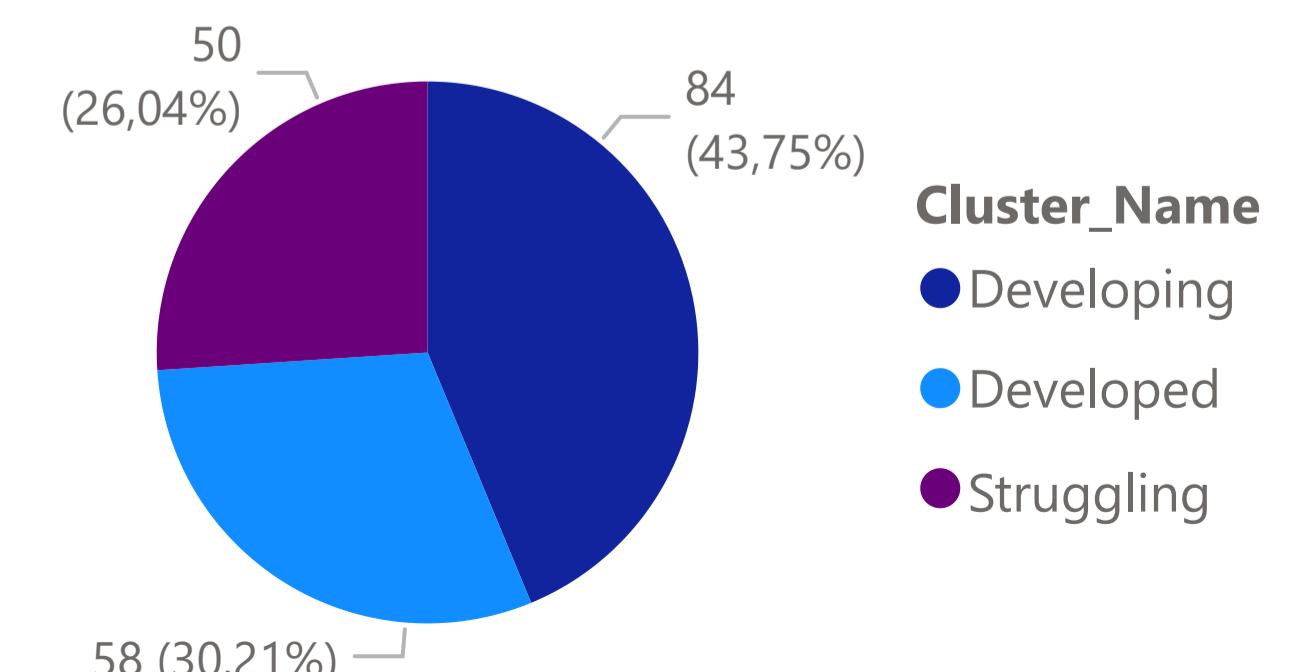
Cluster_Name ● Developed ● Developing ● Struggling



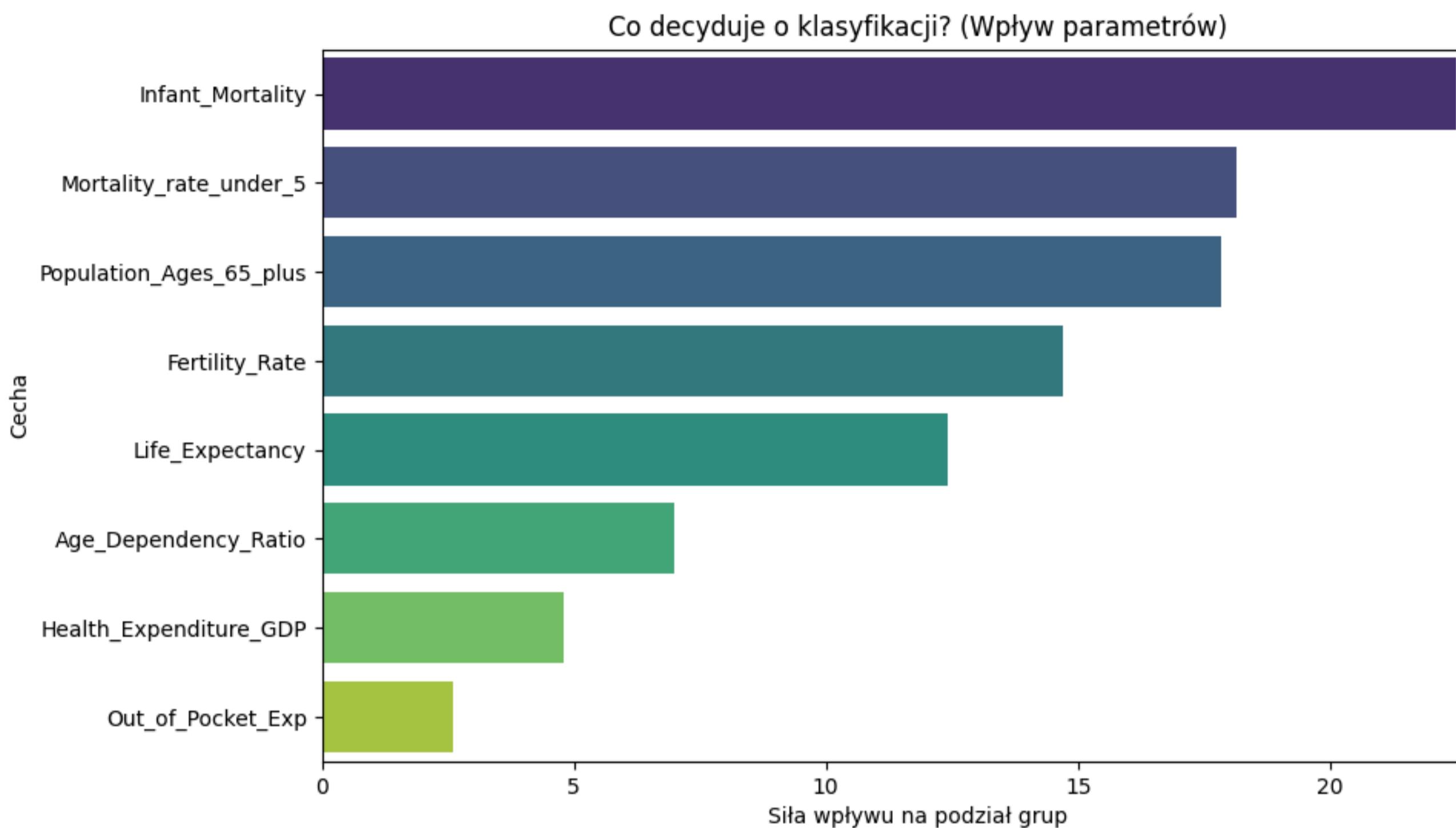
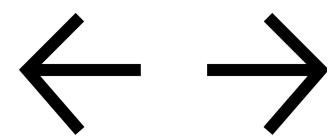
Liczba elementów według kategorii

Country	Region
192	48
country	Cluster_Name
Zimbabwe	Struggling
Zambia	Struggling
Yemen, Rep.	Struggling
West Bank and Gaza	Developing
Viet Nam	Developing
Venezuela, RB	Developing
Vanuatu	Developing
Uzbekistan	Developing
Uruguay	Developed
United States	Developed
United Kingdom	Developed

Liczba krajów w kazdej z kategorii



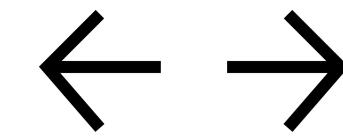
Healthcare - indicators



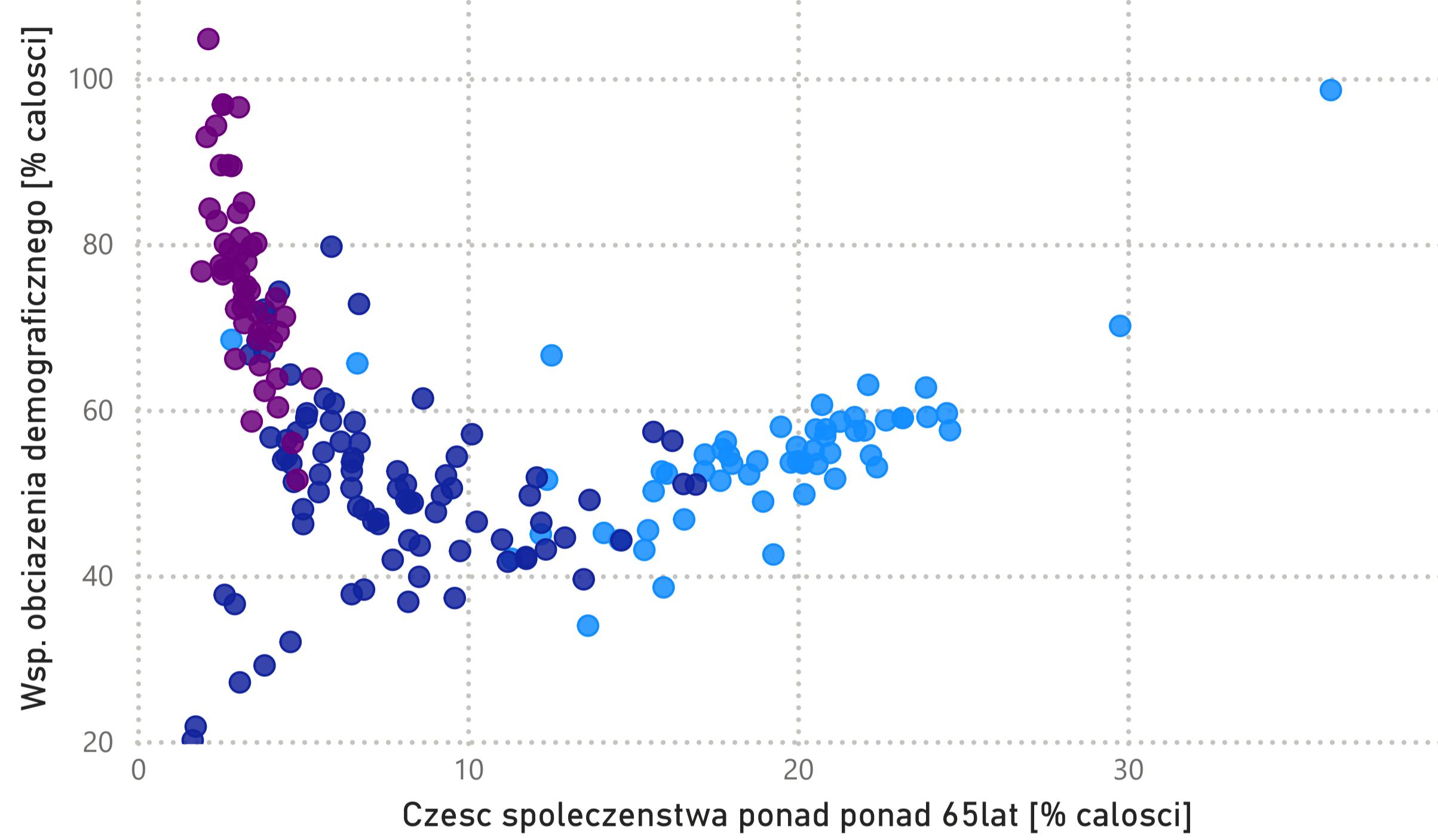
country	Infant Mortality	Mortality rate under 5	Population Ages 65 plus	Fertility Rate	Life Expectancy	Age Dependency Ratio	Health Expenditure GDP	Out of Pocket Exp
Albania	8,30	9,40	16,92	1,35	79,60	50,97	6,19	51,26
Argentina	8,20	9,60	12,42	1,50	77,40	51,56	9,86	26,40
Australia	3,10	3,70	17,73	1,50	83,05	55,21	9,93	15,28
Austria	2,60	3,10	20,61	1,32	81,54	53,46	10,88	16,14
Azerbaijan	13,30	18,60	8,56	1,55	74,43	43,61	3,98	64,47
Bahamas, The	11,40	12,70	11,78	1,37	74,55	42,22	7,22	24,25
Bahrain	7,20	8,60	3,85	1,82	81,28	29,15	3,84	23,47
Belarus	1,90	2,40	17,67	1,21	74,18	51,43	6,66	27,62

- Category
 - Country
 - Region
- Cluster_Name
 - Developed
 - Developing
 - Economy Hub
 - Struggling

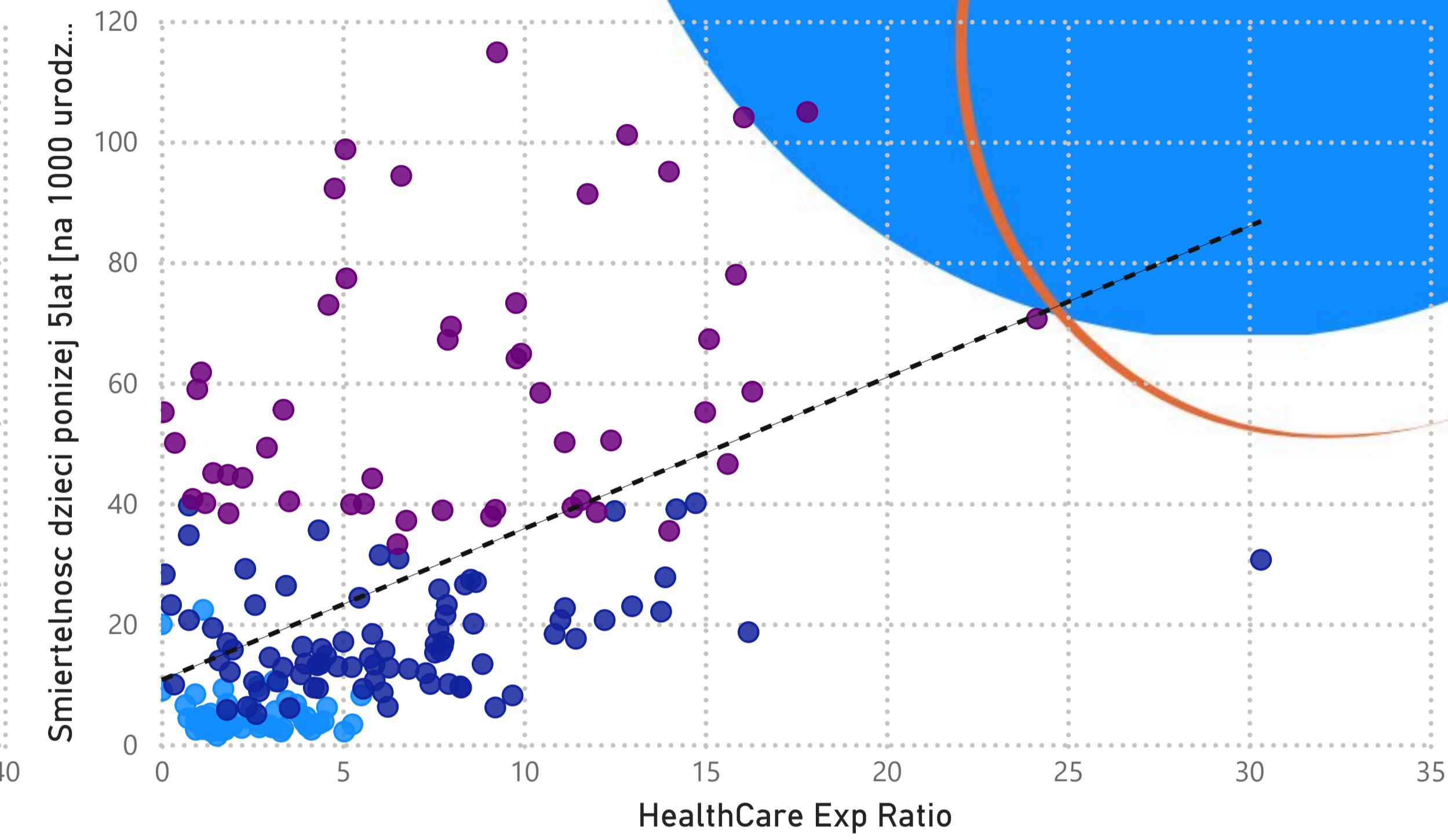
Healthcare - summary



Cluster Name ● Developed ● Developing ● Struggling



Cluster Name ● Developed ● Developing ● Struggling



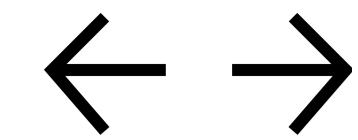
	Developed	Developing	Struggling
Health Expenditure GDP AVG	9,25	5,85	5,87
Infant Mortality AVG	4,18	15,10	42,55
Life Expectancy AVG	79,67	73,85	64,06
Mortality rate under 5 AVG	4,92	17,81	60,13
Out of Pocket Exp AVG	19,61	33,22	36,75
Population Ages 65 plus AVG	19,00	7,63	3,27
Fertility Rate AVG	1,48	2,16	4,15
Age Dependenc Ratio AVG	54,82	50,35	76,51
HealthCare Exp Ratio AVG	2,35	6,45	8,33

Cena życia (Nakłady vs Wyniki): Analiza wykazuje brutalną liniowość – kraje inwestujące poniżej 4% PKB w zdrowie (klaster **Struggling**) notują średnią długość życia krótszą o 15-20 lat w porównaniu do klastra **Developed**. Bariera finansowa jest tu głównym determinantem śmiertelności.

Paradoks Demograficzny: Najbogatsze gospodarki płacą wysoką cenę za sukces – zmagają się z gwałtownym starzeniem społeczeństwa (*High Age Dependency*). Z kolei kraje **Developing** posiadają ogromny potencjał "dywidendy demograficznej" (młode społeczeństwo)..

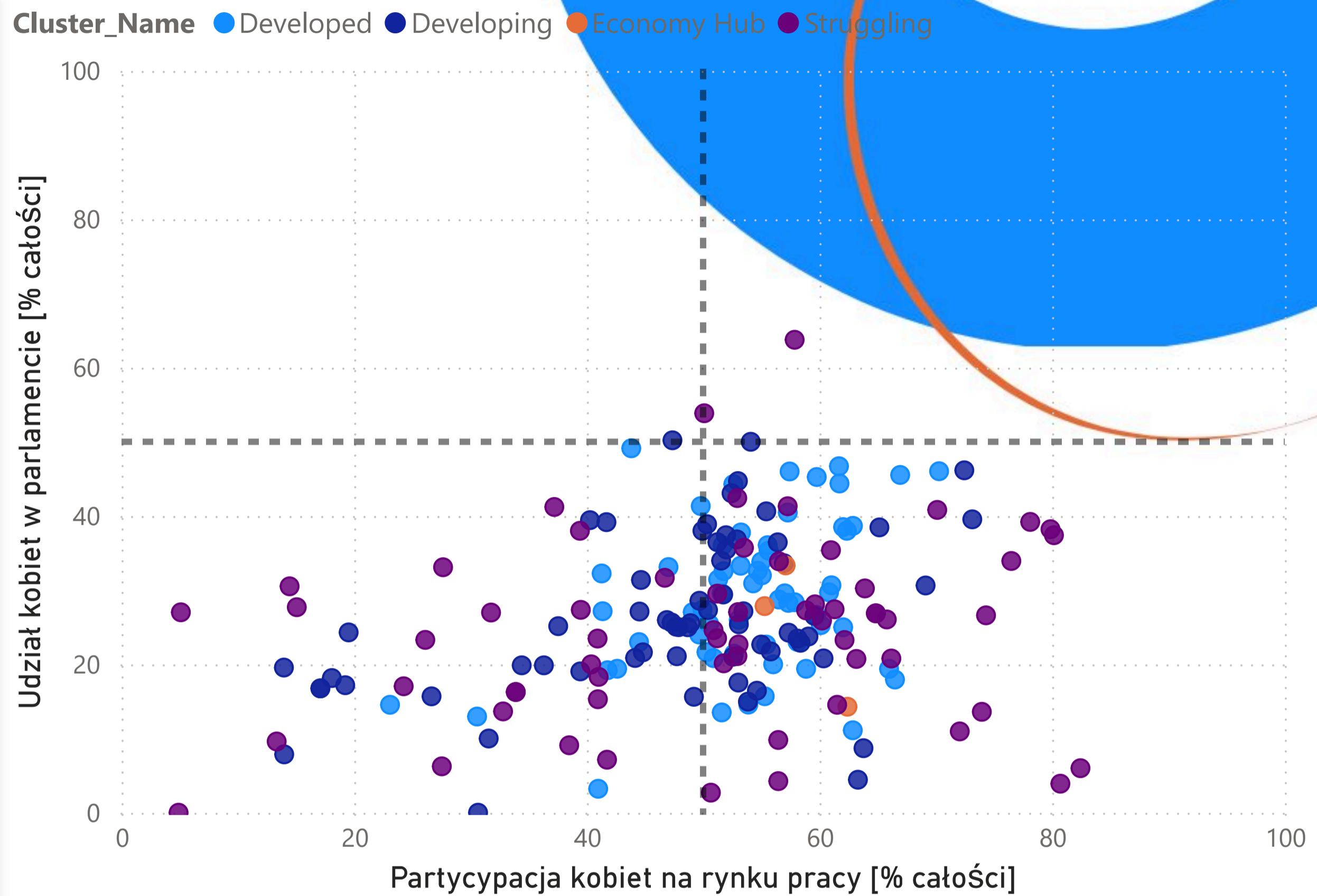
Pułapka "Z własnej kieszeni": W krajach **peryferyjnych** system ubezpieczeń nie działa – obywatele pokrywają nawet 60-70% kosztów leczenia prywatnie (*Out-of-Pocket*). W praktyce oznacza to, że centralne systemy ochrony zdrowia dają całkowicie lepsze rezultaty niż ochrona zdrowia finansowana "z własnej kieszeni".

Combined



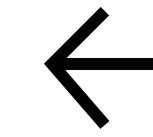
country	Economical	Educacion	Healthcare	Infrastructure	Social
Afghanistan	Struggling	Struggling	Struggling	Developing	Developing
Africa Eastern and Southern	Struggling	Struggling	Struggling	Struggling	Struggling
Africa Western and Central	Struggling	Struggling	Struggling	Struggling	Struggling
Albania	Struggling	Developed	Developing	Developed	Developed
Algeria	Developing	Developed	Developing	Developing	Developing
Angola	Developing	Struggling	Struggling	Struggling	Struggling
Arab World	Developing	Struggling	Developing		Developing
Argentina	Struggling	Developed	Developed	Developed	Developed
Armenia	Developing	Developed	Developing	Developing	Developed
Australia	Developed	Developed	Developed	Developed	Developed
Austria	Developed	Developed	Developed	Developed	Developed
Azerbaijan	Developing	Developed	Developing	Developed	Developed
Bahamas, The	Developed	Developed	Developing	Developed	Developed
Bangladesh	Developing	Struggling	Developing	Developing	Developing
Barbados	Developed	Developed	Developing		Developed
Belarus	Developing	Developed	Developed	Developed	Developed
Belgium	Developed	Developed	Developed	Developed	Developed

Analiza wykazuje, że wysokie wyniki w jednym sektorze nie gwarantują automatycznie sukcesu w innym. Bogactwo kraju nie zawsze przekłada się na długość życia czy jakość edukacji. Tradycyjny podział na "Bogatą Północ" i "Biedne Południe" jest przestarzały. Wiele państw rozwija się niesynchronicznie – mogą być globalnym hubem produkcyjnym, pozostając jednocześnie periferiami w kwestiach społecznych lub ekologicznych. W związku ze zmianami kulturowymi oraz globalizacją usług i produkcji nie można "szufladkować" państw jedną etykietą. Prawdziwy obraz kondycji państwa daje dopiero nałożenie na siebie niezależnych warstw: kapitału, demografii i innowacji.



Wykres pokazuje zależność między partycypacją kobiet na rynku pracy a ich udziałem w parlamencie w różnych klastrach krajów. Nie wykazano wpływu przynależności do żadnej z kategorii a stanem feminizacji społeczeństwa. Kraje rozwinięte charakteryzują się jedynie względną stabilnością obu wskaźników, natomiast kraje rozwijające się i zmagające się z problemami wykazują większe zróżnicowanie, co może wskazywać na istotną rolę czynników instytucjonalnych i kulturowych, które promują postawy pozwalające aspirować do przynależności do grup krajów rozwiniętych. Warto także zaznaczyć, że w znaczyć, że w jedynie 4 państwach odnotowano przynajmniej 50% udział kobiet w Parlamencie. Jest to wyraźna niedoreprezentacja kobiet w całej populacji.

Executive summary



Wnioski sektorowe:

Gospodarka (Economy): Analiza potwierdza wyraźną polaryzację świata. Kraje "Developed" opierają się na usługach (>70% PKB) i kapitale własnym, podczas gdy kraje "Struggling" pozostają uwięzione w rolnictwie. Wyodrębniono istotną grupę "Investment Hubs" – małych gospodarek, które mimo mniejszego potencjału wewnętrznego, pełnią kluczową rolę w globalnym przepływie kapitału (wysokie FDI).

Zdrowie (Healthcare): Wydajność systemu zdrowotnego jest silnie skorelowana z nakładami finansowymi. Kraje inwestujące <4% PKB w zdrowie notują drastycznie niższą długość życia. Zauważalny jest paradoks demograficzny: kraje najbogatsze zmagają się ze starzejącym się społeczeństwem (wysoki *Dependency Ratio* seniorów), podczas gdy kraje rozwijające się posiadają "dywidendę demograficzną" (młode społeczeństwo), której potencjał jest często hamowany przez śmiertelność dzieci.

Edukacja i Innowacje: To sektor, który najmocniej dzieli świat na "gospodarki wiedzy" i resztę. Kluczowym czynnikiem różnicującym nie jest już tylko analfabetyzm, ale wykluczenie cyfrowe (*Internet Usage*) oraz zdolność do tworzenia innowacji (*Patents*). Kraje peryferyjne często nadrabiają zaległości w edukacji podstawowej, ale tracą dystans w edukacji wyższej i technologicznej.

Wnioski ogólne:

Twarde Peryferia: Grupa państw (głównie Afryka Subsaharyjska) w każdym sektorze pozostaje w klastrze "Struggling". Brak kapitału blokuje tam start.

Niesynchroniczność: Wiele państw (*Developing*) rozwija się nierówno – np. są potęgą przemysłową, ale mają zaniedbaną służbę zdrowia lub środowisko.

Ocena przez samo PKB to błąd: Prawdziwy status "Centrum" (Core) wymaga spójnego sukcesu w Ekonomii, Zdrowiu i Edukacji jednocześnie.

Narzędzia :

- Środowisko: Python 3.10+ (Jupyter Notebook).
- Biblioteki: pandas (przetwarzanie danych), pandas_datareader (pobieranie API), scikit-learn (klastrowanie K-Means, RandomForest), matplotlib oraz seaborn (wizualizacja).
- Dane: World Bank Open Data API (okno czasowe 2020–2022, metoda *Latest Available Value*).
- Metoda: Nienadzorowane uczenie maszynowe (K-Means Clustering) na danych standaryzowanych.