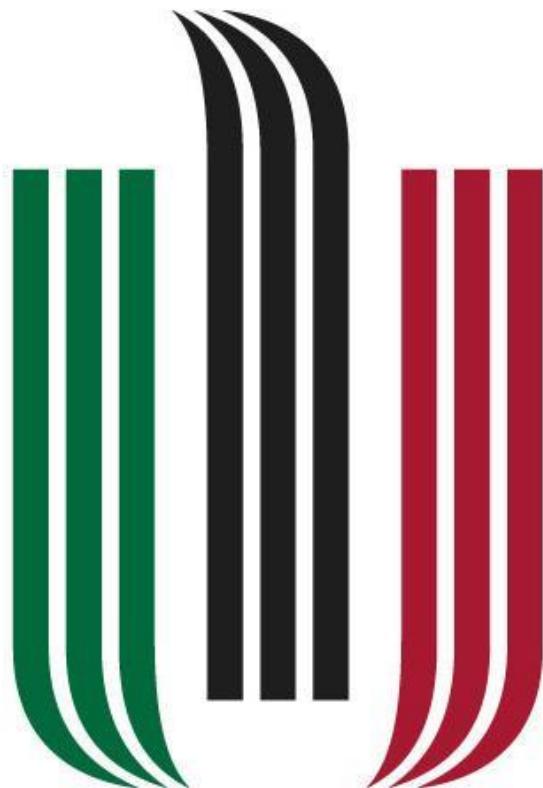


Dobrzański Tymoteusz
Makowski Tomasz

Eksploracja Danych



AGH

Temat projektu:

Analiza podobieństw między państwami na podstawie danych ekonomicznych i socjologicznych Banku Światowego.

1.	Zbiór danych	3
2.	Wstępne statystyki	4
3.	Wykorzystane parametry rozwoju	5
3.1.	Ekonomiczne:	5
3.2.	Geograficzne:	5
3.3.	Socjologiczne:	5
3.4.	Demograficzne:	5
4.	Preprzetwarzanie (wstępne przetwarzanie)	6
5.	Dynamic time warping (DTW)	7
6.	Skalowanie wielowymiarowe	8
6.1.	PCA	8
6.2.	TSNE	9
6.3.	TSNE na PCA	11
6.4.	MultiDimensional Scaling (MDS)	12
7.	Klasteryzacja hierarchiczna	14
7.1.	Dendrogramy	14
7.2.	Mapy ciepła:	16
7.3.	PCA	18
7.4.	TSNE	21
8.	Odnośniki	24
8.1.	Kod oraz uzyskane wyniki	24
8.2.	Spis obrazów	24

1. Zbiór danych

Dane wykorzystane w projekcie zawierają kolekcje indeksów rozwoju, skompletowane z oficjalnych międzynarodowych źródeł. Prezentują najbardziej bieżące i dokładne globalnie dostępne dane i zawierają narodowe, regionalne oraz światowe szacunki.

Typ danych: Szereg czasowy,

Okrzesowość: roczna,

Pokrycie czasowe: 1960-2020

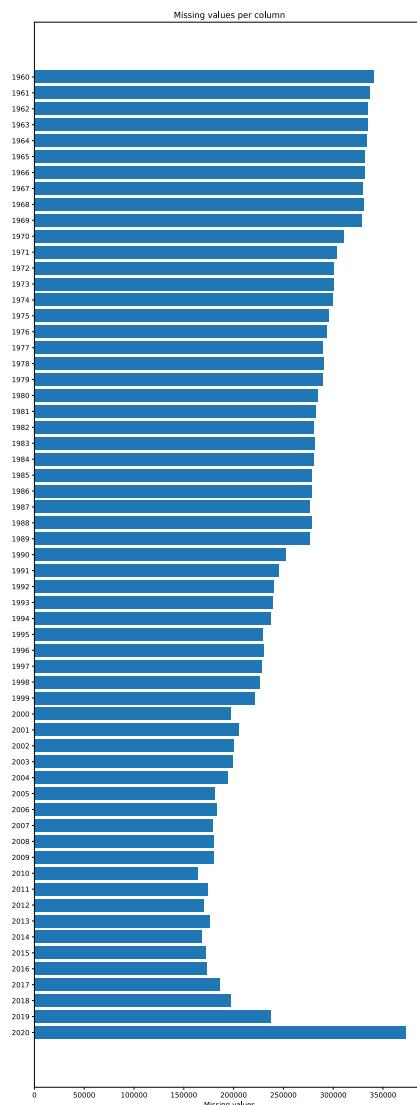
Źródło: <https://datacatalog.worldbank.org/dataset/world-development-indicators>

2. Wstępne statystyki

Liczba rekordów w zbiorze danych: 378576

Ilość dostępnych współczynników: 1434

Brakujące wartości:



Rysunek 1 Brakujące wartości

Ze względu na spore braki w danych w początkowych latach pokrycia czasowego, jak i w roku 2020, zbiór danych został okrojony do lat 1989 - 2019

Liczba krajów oraz regionów uwzględnionych w zbiorze: 264

3. Wykorzystane parametry rozwoju

Spośród 1434 dostępnych wskaźników, na potrzeby projektu, zostały wybrane następujące:

3.1. Ekonomiczne:

GDP (PKB - produkt krajowy brutto)

CPI (consumer price index - wskaźnik cen towarów i usług konsumpcyjnych)

Real interest rate (%) (realna stopa procentowa)

Adjusted net national income (skorygowany dochód narodowy netto)

Export of goods and services

3.2. Geograficzne:

Land area (sq. km)

Agricultural land (sq. km)

Access to electricity (% of population)

Urban population (% of total population)

Permanent cropland (% of land area)

Population density (people per sq. km of land area)

3.3. Socjologiczne:

Unemployment

school enrollment/ school attainment

suicide mortality rate

CPIA property rights and rule-based governance rating

Individuals using the Internet (% of population)

3.4. Demograficzne:

Population, total

Mortality rate, under-5 (per 1,000 live births)

Life expectancy at birth, total (years)

Age dependency ratio (% of working-age population)

Fertility rate, total (births per woman)

4. Preprzetwarzanie (wstępne przetwarzanie)

W pierwszym kroku preprzetwarzania zawężony został przedział lat, które zostaną wzięte pod uwagę, w dalszych procesach projektu z [1960-2020], do [1989-2019].

W drugim kroku wstępnego przetwarzania odfiltrowane lub uzupełnione zostały komórki, które zawierały w sobie wartości NaN (not a number ~ nie-liczba), czyli brakujących wartości. Proces ten został podzielony na kilka etapów:

- eliminacja krajów z największą liczbą brakujących wartości,
- eliminacja wskaźników ze sporą liczbą brakujących wartości,
- wyszukiwanie krajów, które w poszczególnych indeksach mają więcej niż $\frac{1}{3}$ brakujących wartości. Część z nich została uzupełniona za pomocą innych zbiorów danych, w których tych danych nie brakowało, a część z nich porzucona.
- uzupełnienie pozostałych brakujących wartości za pomocą metody "backward fill", czyli za pomocą wartości z roku następującego po komórce z wartością NaN.
- jedyne brakujące wartości zostały w ostatnich latach tj. 2016 - 2019 i zostały uzupełnione metodą 'forward fill', czyli za pomocą wartości z lat poprzedzających brakujące wartości

Kolejnym krokiem jest normalizacja poszczególnych wskaźników, normalizacja miała na celu uniezależnienie niektórych współczynników od wielkości badanego kraju (powierzchnia państwa, liczba ludności). Część dobranych indykatatorów była już wcześniej znormalizowana, jednak niektóre wymagały dalszego przetwarzania w celu poprawy jakości używanych danych.

Normalizacja cech wykorzystywanych do budowy modeli, przebiegała następująco:

Parametry takie jak:

Adjusted net national income (constant 2010 US\$)
Exports of goods and services (constant 2010 US\$)
GDP (constant 2010 US\$)

Zostały znormalizowane względem 1000 mieszkańców, aby te współczynniki rozwoju uniezależnić od liczby mieszkańców. Natomiast współczynnik:

Agricultural land (sq. km)

Został podzielony przez całą powierzchnię kraju, którego dotyczył, tworząc parametr określający jaką powierzchnię kraju stanowią ziemie uprawne.

Uzupełnione oraz znormalizowane dane zostały zapisane do nowego zbioru i wykorzystywane w kolejnych etapach projektu.

Zapisany zbiór danych zawierał wymienione w rozdziale 3 współczynniki zebrane dla 94 krajów.

Po przetworzeniu zbioru danych zostało w nim 1043 brakujących wartości w 1222 rekordach.

5. Dynamic time warping (DTW)

Korzystając z przygotowanych danych, został przeprowadzony eksperyment, polegający na zbadaniu miar podobieństwa pomiędzy szeregami czasowymi.

Badanie podobieństwa między szeregami czasowymi zostało wykonane przy wykorzystaniu metody dynamic time warping, metoda ta polega na pomiarze odległości pomiędzy dwoma szeregami czasowymi. Pomiar odległości opiera się na porównywaniu kolejno punktów będących składowymi szeregów czasowych. DTW zwraca sumę zebranych różnic w położeniu poszczególnych pasujących punktów. Korzystając z tego algorytmu zostały przygotowane macierze odległości pomiędzy szeregami, dla każdego współczynnika rozwoju powstała osobna macierz. Otrzymane macierze zostały wykorzystane w eksperymencie opisany w punkcie 6.4.

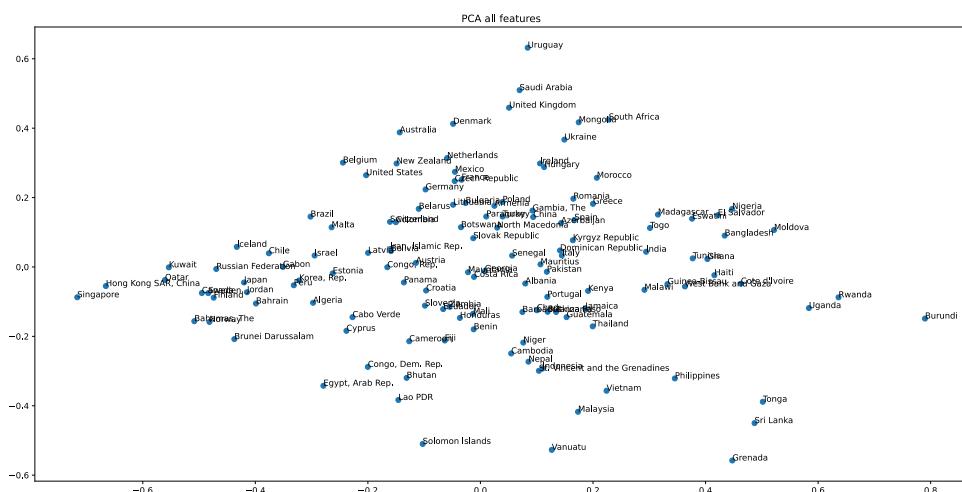
6. Skalowanie wielowymiarowe

Skalowanie wielowymiarowe dąży do rozmieszczenia obiektów jako punktów w przestrzeni n-wymiarowej tak, aby obiekty podobne do siebie znajdowały się bliżej.

6.1. PCA

Analiza głównych składowych (ang. principal component analysis, PCA) – jedna ze statystycznych metod analizy czynnikowej. Zbiór danych składający się z N obserwacji, z których każda obejmuje K zmiennych, można interpretować jako chmurę N punktów w przestrzeni K-wymiarowej. Celem PCA jest taki obrót układu współrzędnych, aby maksymalizować w pierwszej kolejności wariancję pierwszej współrzędnej, następnie wariancję drugiej współrzędnej itd.. Tak przekształcone wartości współrzędnych nazywane są ładunkami wygenerowanych czynników (składowych głównych). W ten sposób konstruowana jest nowa przestrzeń obserwacji, w której najwięcej zmienności wyjaśniają początkowe czynniki.

Rezultat przedstawia się następująco:



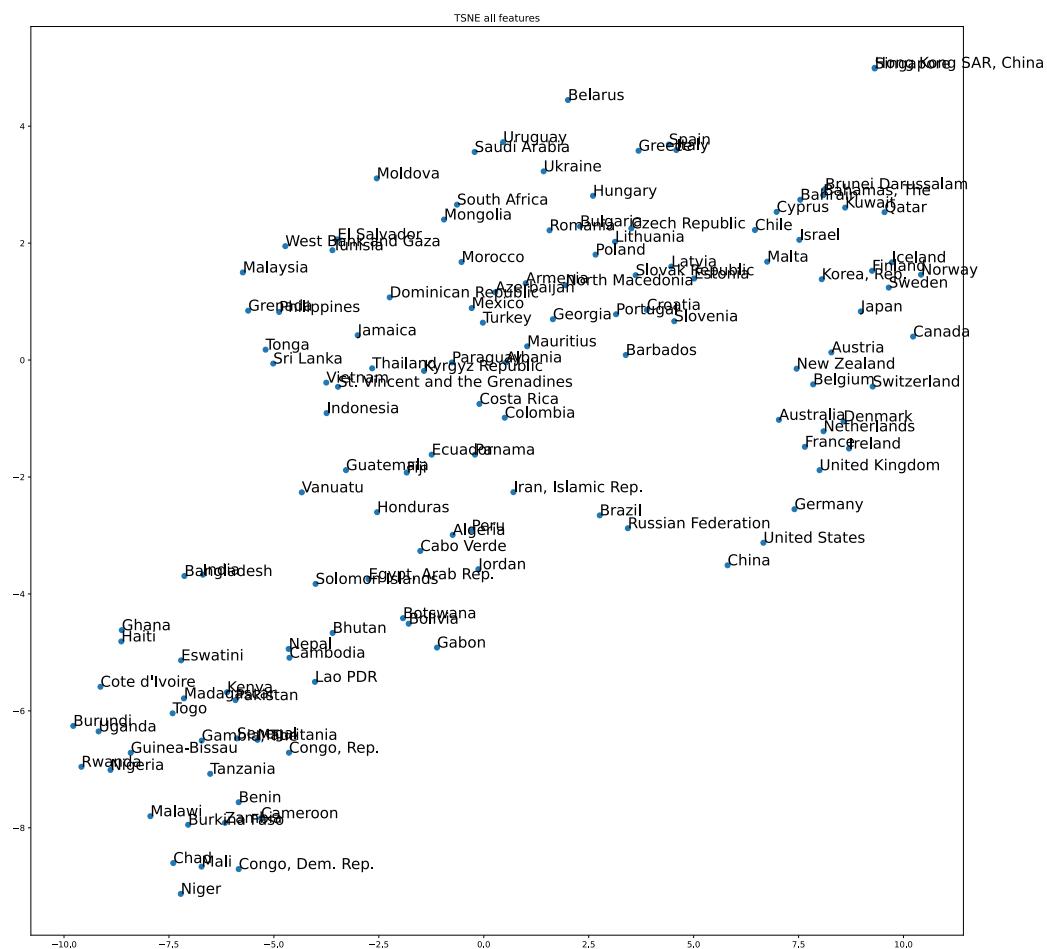
Rysunek 2 PCA

Wykorzystanie samego PCA zaczęło nam formułować kraje w pewne grupy. W lewym dolnym rogu widać tego dobry przykład, gdzie Singapur, Katar, Hong Kong, Islandia, Japonia, a dodatkowo kraje skandynawskie Norwegia, Szwecja, Finlandia, czyli kraje dobrze rozwijające się ale stosunkowo małe powierzchniowo. Następnie, trochę wyżej większość najbardziej rozwiniętych europejskich krajów i stany zjednoczone. Im bardziej w prawo tym pojawiają się nam kraje coraz mniej zamożne, jednak większość z nich skoncentrowana jest wokół państw, które ze sobą sąsiadują lub przynajmniej należą do tego samego regionu (klastry krajów Europy wschodniej, Ameryki Południowej, niektóre kraje Azji Centralnej). Całkowicie po prawo znajdują się państwa najmniej rozwinięte oraz najbiedniejsze - w większości są to kraje Afrykańskie.

6.2. TSNE

Skrót t-SNE oznacza stochastyczną metodę porządkowania sąsiadów w oparciu o rozkład t (t-Distributed Stochastic Neighbor Embedding). Jest to nieliniowa i nienadzorowana technika stosowana przede wszystkim do eksploracji i wizualizacji danych wielowymiarowych.

Algorytm t-SNE oblicza miarę podobieństwa między parami w przestrzeni o dużych wymiarach i przestrzeni o małych wymiarach. Następnie próbuje zoptymalizować te dwie miary podobieństwa.



Rysunek 3 TSNE

TSNE w porównaniu do PCA dał nam jeszcze lepsze pogrupowanie. Można doszukać się nawet 5 różnych wyraźnych grup, jak nie więcej.

Jedną z nich są bardzo małutkie, ale pomimo tego bardzo dobrze rozwijający się kraje i regiony, mianowicie, Singapur i Hong Kong. Po zeskalowaniu wszystkich cech do dwóch wymiarów właściwie znajdują się w jednym miejscu.

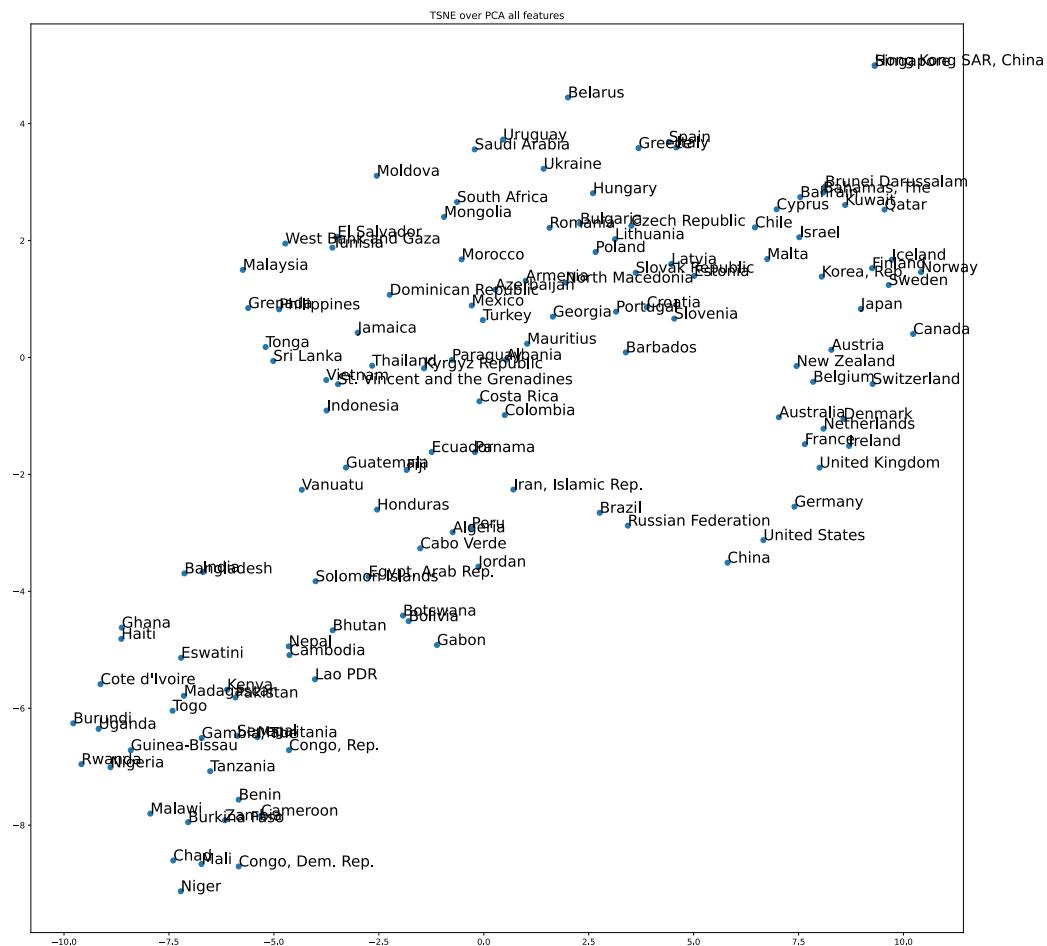
Następnie, ponad nimi bogate kraje środkowego wschodu (Katar, Kuwejt, Bahrajn), a po prawej od nich kraje skandynawskie, dodatkowo Islandia i Kanada. Powyżej, po lewej stronie, dość interesująca grupa krajów - Korea południowa, Izrael, Cypr, Malta i co ciekawe Chile. Dalej Japonia i Szwajcaria, które znajdują się pomiędzy poprzednimi dwoma klastrami, a klastrem mniejszych krajów Europy zachodniej (Austria, Dania, Belgia, Holandia) + Australii i Nowej Zelandii. Nad nimi większe kraje Europy Zachodniej (Wielka Brytania, Francja, Niemcy), aż w końcu dochodzimy do Stanów Zjednoczonych i Chin.

Te wszystkie kraje zdają się być krajami raczej bogatymi i dobrze rozwijającymi się.

Idąc w stronę centralną wykresu, znajdziemy wiele krajów Europy centralnej, południowej i wschodniej oraz większość krajów Ameryki Południowej i Azji. Gdziekolwiek można dostrzec, że kraje stosunkowo niedaleko znajdujące się na mapie, również pod względem czynników ekonomicznych i socjologicznych nie odbiegają od siebie za bardzo. W górnym lewym rogu większość państw, należą do grona słabo rozwiniętych albo biednych, w głównej mierze kraje afrykańskie.

6.3. TSNE na PCA

Najpierw redukujemy wymiarowość danych do 10 wymiarów, ponieważ tyle wymiarów wyjaśnia wariancję w ponad 99%, a następnie powstałe dane aplikujemy do metody TSNE i redukujemy wymiarowość do dwóch.



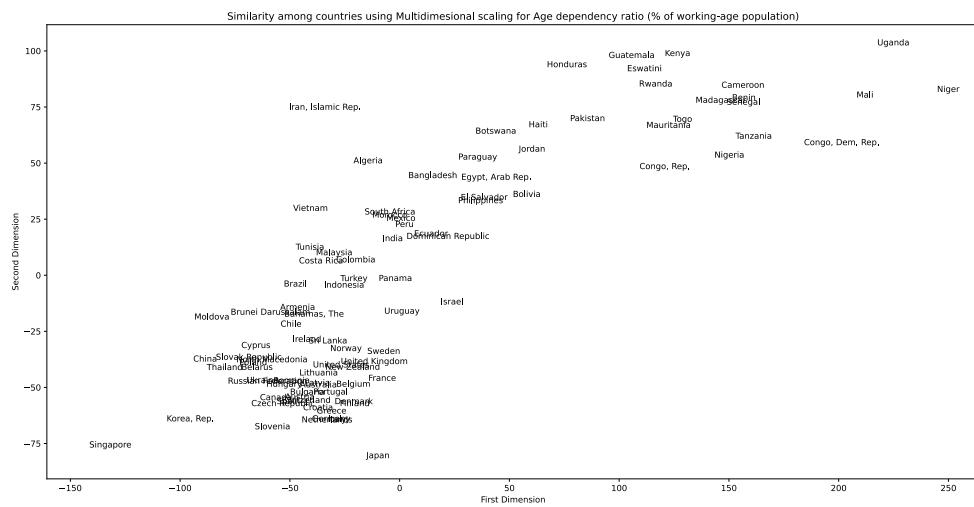
Rysunek 4 TSNE na PCA

W tym przypadku uzyskaliśmy lepszy podział między państwami średniozamożnym, a państwami biednymi. Prawa góra część wykresu, to państwa najlepiej rozwijające, środek, to państwa średnio rozwijające i lewy dół, to państwa słabo rozwijające.

6.4. MultiDimensional Scaling (MDS)

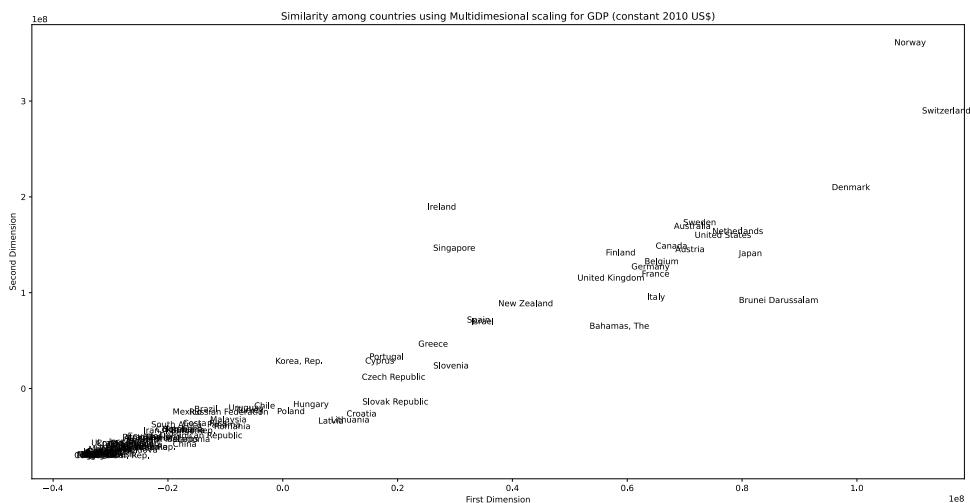
Korzystając z tej metody redukcji wielowymiarowości przygotowany model ma za zadanie wyznaczenie zmiennych ukrytych, pozornie niedostępnych. Zmienne ukryte są estymowane dla macierzy odległości pomiędzy szeregiem czasowymi, opis przygotowania macierzy znajduje się w punkcie 5.

Przykładowe wyniki:



Rysunek 5 MDS Age dependency ratio

Na wykresie uzyskanym dla parametru Age dependency ratio (procent populacji będącej w wieku pracującym), widać ciekawe wyniki eksperymentu. Kraje słabo rozwinięte, najczęściej kraje znajdujące się w Afryce) znajdują się w prawym górnym rogu wykresu, następnie wykres przechodzi przez kraje rozwijające się, aż do krajów wysoko rozwiniętych w lewej części wykresu, znajdują się tam między innymi Japonia, Singapur, spora część krajów europejskich. Można również zaobserwować wysokie podobieństwo pomiędzy krajami znajdującymi się w Europie, gdyż praktycznie wszystkie znajdują się bardzo blisko siebie, nie można zaobserwować tego podobieństwa pomiędzy krajami usytuowanymi na innych kontynentach.



Rysunek 6 MDS GDP

Kolejny wykres przygotowany dla parametru oznaczającego produkt krajowy brutto przeliczony na 1000 mieszkańców, współczynnik ten dobrze rozdziela kraje pod względem ekonomicznym, w prawym górnym rogu widać bardzo bogate kraje, które przechodzą w bardzo biedne kraje w lewym rogu. Również w tym przypadku można zaobserwować ciekawe wnioski. Na wykresie widać znaczące różnice pomiędzy krajami bogatymi, jest ich znacznie mniej dodatkowo są one rozmieszczone z większymi odległościami, natomiast kraje biedne są bardzo mocno skupione w jednym punkcie. Może to sugerować bardzo wysokie różnice pomiędzy państwami.

7. Klasteryzacja hierarchiczna

Klasteryzacja jest metodą uczenia nienadzorowanego, której zadaniem jest grupowanie podobnych wpisów nie posiadając etykiet.

Hierarchiczna klasteryzacja wykorzystuje macierze odległości aby określić podobieństwa pomiędzy poszczególnymi państwami. W algorytmie aglomeracyjnym, każde z państw najpierw znajduje się we własnym klastrze i za pomocą wybranej metryki obliczana jest różnica pomiędzy każdą grupą. Państwa, których różnica jest najmniejsza łączone są w jedną grupę i od tej pory brane pod uwagę jako jednostka (samodzielny klaster/grupa).

Najbardziej popularne metryki różnic to:

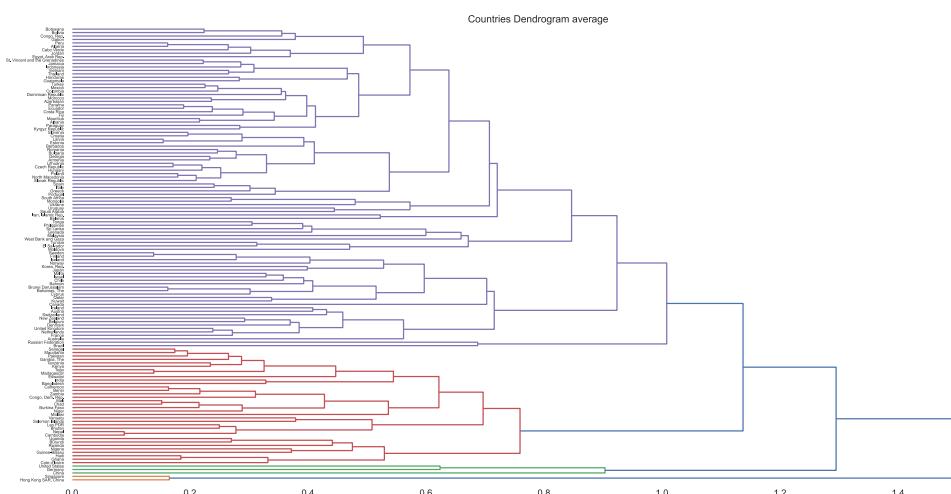
- maximum/ complete linkage, który minimalizuje maksymalny dystans pomiędzy obserwacjami,
- Average linkage, który minimalizuje średni dystans pomiędzy wszystkimi obserwacjami lub parą klastrów,
- single linkage, który minimalizuje odległość pomiędzy najbliższymi obserwacjami,
- ward, który minimalizuje sumy kwadratów różnic pomiędzy wszystkimi grupami.

Jednym z domyślnych sposobów przedstawiania klasteryzacji hierarchicznej są dendrogramy, czyli diagramem w kształcie drzewa ukazujący związki pomiędzy wybranymi elementami, w tym wypadku państwami.

7.1. Dendrogramy

Wygenerowane zostały dendrogramy dla każdej z podanej wyżej metod, aby rezultaty przedstawiają się następująco:

Average:

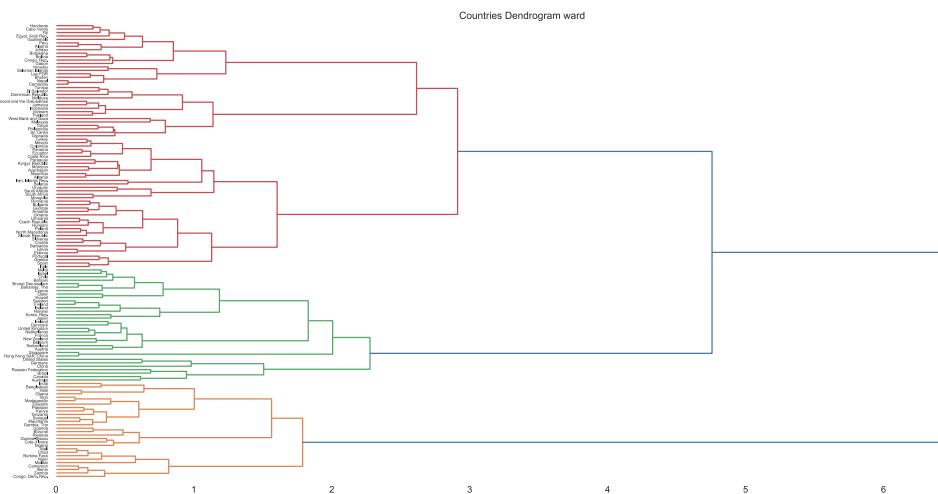


Rysunek 7 Dendrogram kryterium average

Dość ciekawy podział, który pogrupował kraje na 4 grupy. Jedna składa się z Singapuru i Hong Kongu, druga, do której należą Stany Zjednoczone, Chiny i Niemcy (co wskazuje na to, że

Niemcy odstają pod względem rozwoju i zamożności względem innych dobrze rozwijających się państw Europejskich. Trzecia grupa to kraje ubogie, do których co ciekawe należą Indie (można było to zauważyc już na wykresach PCA i TSNE, że lekko odstają od reszty, ale nie sądziłem, że aż tak). Czwarta grupa, to kraje średniozamożne lub bardziej zamożne.

Ward:



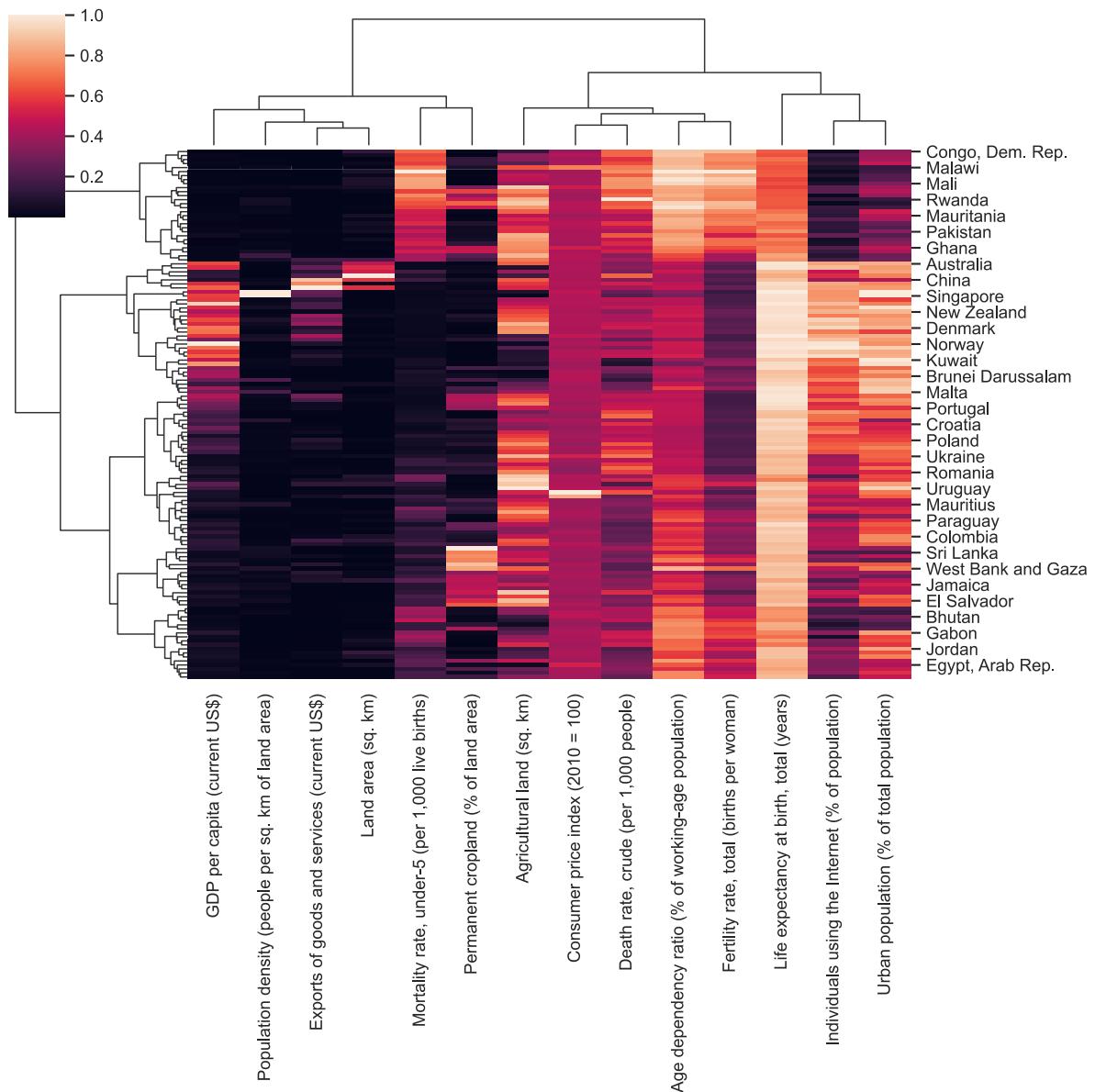
Rysunek 8 Dendrogram kryterium warda

Kraje zostały podzielone na 3 główne klastry - Bogate, średnio zamożne i biedne

7.2. Mapy ciepła:

Innym, bardzo ciekawym sposobem na wyświetlenie podobieństw pomiędzy państwami są mapy ciepła, gdyż nie tylko pokazują, które państwa zostały pogrupowane razem w klastry, ale również pokazują jak się do tego mają poszczególne wartości wykorzystanych wskaźników.

Przykładowa mapa ciepła dla metryki “Average Linkage”:



Rysunek 9 Mapa ciepła kryterium average

Powysze metody, cechują się tym, iż liczba utworzonych klastrów nie jest odgórnie zdefiniowana i wykorzystanie poszczególnych metryk prowadzi do powstania różnej liczby klastrów. Poniżej przedstawiony zostanie metoda biblioteki sklearn - AgglomerativeClustering, w która również dokonuje klasteryzacji hierarchicznej, jednakże w tym wypadku, liczba klastrów musi zostać zdefiniowana odgórnie. Podziały określone przez tę metodę zostaną nałożone na wykres państw poddany analizie głównych składowych

(PCA), w której liczba zmiennych określających poszczególne państwa została ograniczona do dwóch, aby możliwe było przedstawienie wszystkich krajów na wykresie dwuwymiarowym. Wstępnie zdefiniowana liczba klastrów to 7.

7.3. PCA



Rysunek 10 Klasteryzacja hierarchiczna na wykresie PCA

Zacznę od tych najmniej licznych grup.

Na różowo Chiny. Pomimo tego, że na wykresie znajdują się blisko krajów Europy środkowo-zachodniej i Europy południowej, to model klastrowania hierarchicznego wydzielił im osobną grupę.

W kolorze czerwonym, Singapur i Hong Kong po raz kolejny razem, malutkie, ale bardzo dobrze rozwijające się regiony.

Kolejny przykład, który dobrze ukazuje jak bardzo Niemcy odbiegają od reszty nawet tej zamożniejszej Europy, gdyż tworzą swoją własną grupę razem ze Stanami zjednoczonymi.

Na fioletowo dwa duże ale średnio rozwijające kraje - Rosja i Brazylia.

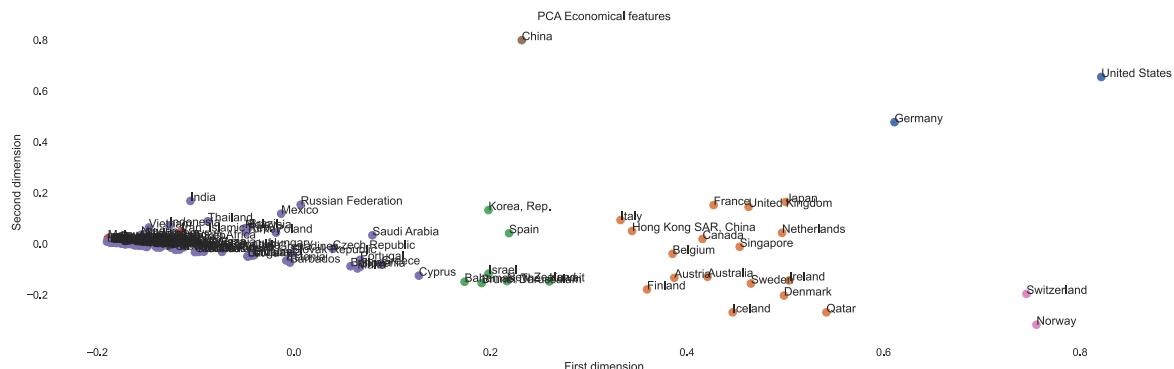
Na zielono kraje szybko rozwijające się i się i w większości przypadków bogatsze, takie jak kraje Europy zachodniej, Japonia, Korea południowa, Izrael, Chile, Nowa Zelandia i Australia.

Na niebiesko kraje rozwijające się, ale mniej zamożne. Większość pozostałych państw europejskich, kraje ameryki południowej oraz część państw Azjatyckich.

Na pomarańczowo kraje słabo rozwijające się oraz biedne. W głównej mierze kraje Afrykańskie, ale też kraje azjatyckie takie jak Indie, Pakistan, czy Nepal.

PCA z podziałem na kategorie

Ekonomiczne czynniki:

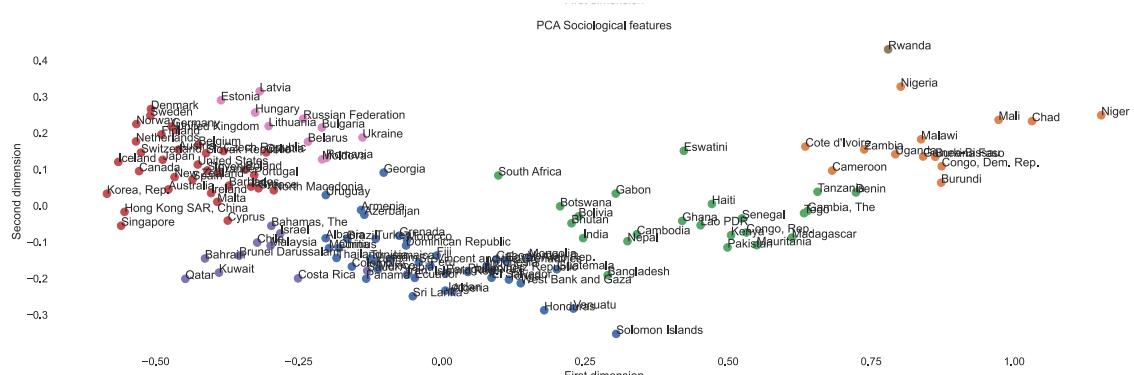


Rysunek 11 Klasteryzacja hierarchiczna na wykresie PCA czynniki ekonomiczne

Pod względem czynników ekonomicznych powstały następujące klastry:

- Chiny na brązowo, po raz kolejny pokazują, że na przestrzeni ubiegłych 30 lat nie próżnowały tylko się rozwijały i obecnie stanowią kluczową rolę jako światowa potęga.
- Niemcy i USA na niebiesko.
- Szwajcaria oraz Norwegia na różowo. Podobnie jak Niemcy na tyle odstają od reszty państw europejskich, że stanowią osobną grupę.
- Na pomarańczowo, większość dobrze rozwijających się krajów europejskich oraz najbogatsze państwa bliskiego wschodu, dodatkowo państwa takie jak Kanada, Singapur, region-Hong Kong,
- Na zielono kraje takie jak: Korea południowa, Hiszpania, Izrael, Kuwejt
- Na fioletowo kraje mniej zamożne, ale wciąż rozwijające się
- Na czerwono najuboższe.

Socjologiczne czynniki:



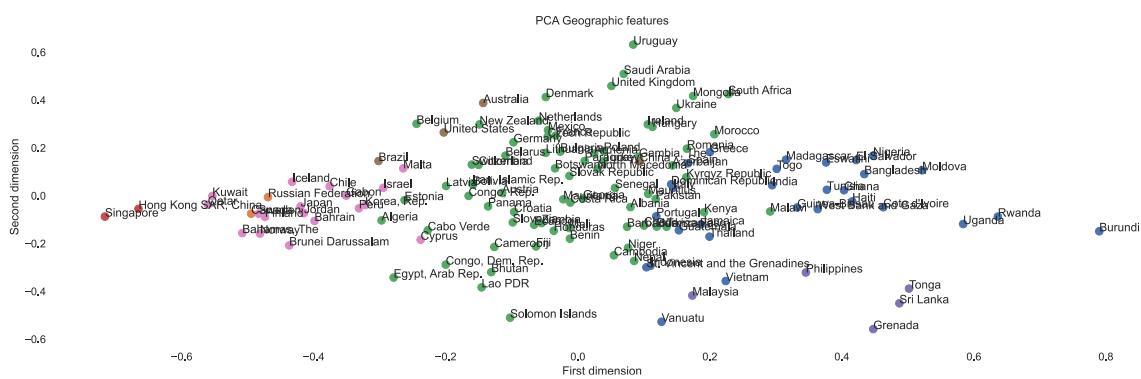
Rysunek 12 Klasteryzacja hierarchiczna na wykresie PCA czynniki socjologiczne

W tym wypadku uwarunkowało nam się kilka bardzo ciekawych klastrów.

- na różowo kraje europy wschodniej,
- na czerwono większość pozostałych części europy, do tego kraje takie jak Chiny, Singapur, Kanada, USA, Korea południowa.

- na fioletowo grupa zamożniejszych krajów bliskiego wschodu (Izrael, Kuwejt, Katar), pojedyncze, bardziej rozwinięte kraje ameryki południowej (Kostaryka, Chile), Bahamy, Malezja, Brunei Darussalam.
- Na niebiesko grupa złożona z krajów eurazji takich jak Azerbejdżan i Armenia. Kraje azjatyckie takie jak Sri Lanka, Honduras, kraje ameryki południowej - Kolumbia, Panama,
- W zielonym kolorze zaczynają pojawiać się kraje afrykańskie takie jak Republika południowej Afryki, Kambodża, madagaskar, ale też części państw azjatyckich takich jak Indie, Nepal, Haiti, Pakistan.
- Na pomarańczowo, same kraje afrykańskie - Nigeria, Kongo, Uganda, Kamerun
- i odstający nawet od krajów afrykańskich, na brązowo Rwanda.

Geograficzne czynniki:



Rysunek 13 Klasteryzacja hierarchiczna na wykresie PCA czynniki geograficzne

Pod względem czynników geograficznych powstały następujące klastry:

- czerwony, składający się z Singapuru i Hong Kongu (bardzo niewielkie, ale duże zaludnienie),
- pomarańczowy, w skład którego wchodzi Rosja i Kanada (dwa największe państwa),
- na brązowo - Australia, USA, Brazylia, Chiny (następne 4 największe kraje),
- na fioletowo - Sri Lanka, Filipiny, Malezja. Państw położone na wyspach między Azją a Oceanią + Grenada, czyli kraj na Karaibach. Za pewne cechujące się uprawami trwałymi takimi jak kakao, kawa, czy guma.
- na zielono i niebiesko dużo krajów o średniej wielkości, jednakże te w grupie zielonej podejrzewam, że cechują się większą gęstością populacji

7.4. TSNE

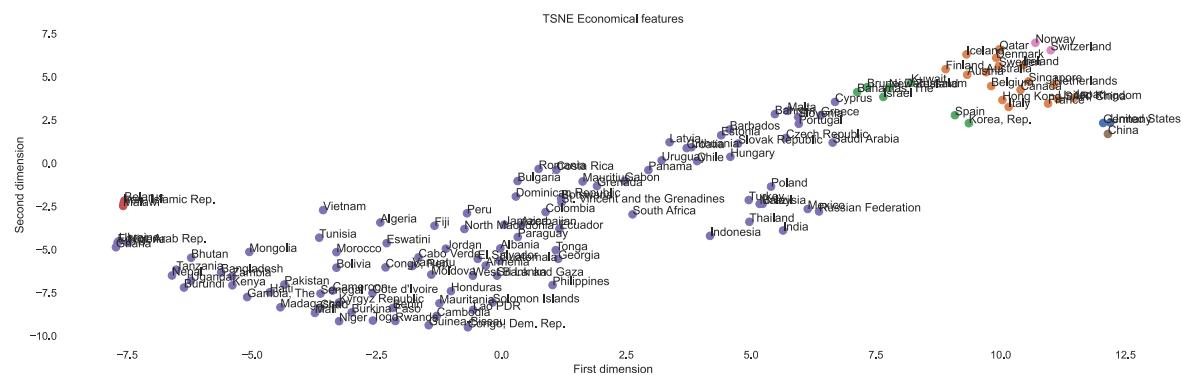


Rysunek 14 Klasteryzacja hierarchiczna na wykresie TSNE

TSNE po raz kolejny lepiej poradził sobie z separacją grup. W tym przypadku Chiny zostały oddzielone od całej reszty. Dodatkowo widać wyraźniejszą separację między grupami państw szybko, średnio i wolno rozwijającymi, oraz zamożnymi i biednymi.

TSNE z podziałem na kategorie:

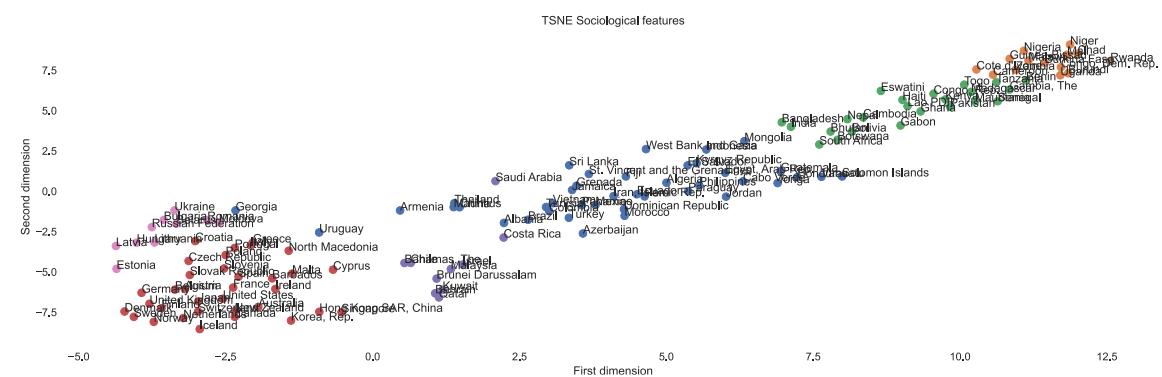
Ekonomiczne:



Rysunek 15 Klasteryzacja hierarchiczna na wykresie TSNE czynniki ekonomiczne

Podobnie jak w powyższym, znacznie lepsza separacja, w szczególności między państwami średnio zamożnymi i biednymi. Trochę gorsze, ale wciążauważalne różnice między państwami bogatymi i szybko rozwijającymi.

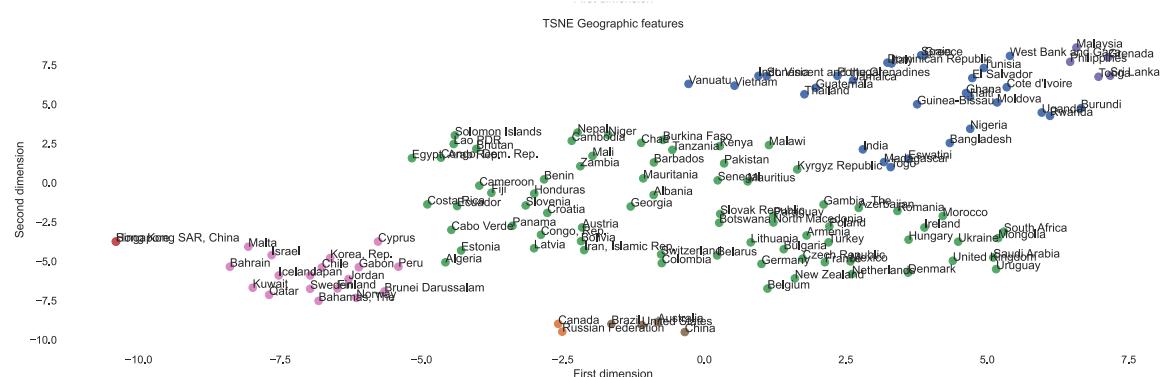
Socjologiczne:



Rysunek 16 Klasteryzacja hierarchiczna na wykresie TSNE czynnik socjologiczne

Najwidoczniejszy podział jest względem krajów o małej, średniej i dużej umieralności. Ale spośród tych o małej lub średniej umieralności dochodzi również podział na kraje z większym odsetkiem ludności posiadającym dostęp do internetu i mniejszym.

Geograficzne:



Rysunek 17 Klasteryzacja hierarchiczna na wykresie TSNE czynniki geograficzne

W tym przypadku znowu znacznie lepiej widoczne podziały. Największe kraje w lewym dolnym rogu. Powyżej małe kraje o małym odsetku ziemi uprawnej. Powyżej malutkie regiony bardzo dużym zaludnieniu. Całkowicie po prawej państwa nadmorskie o sporym odsetku upraw trwałych. A pośrodku, kraje o średniej wielkości. Te na zielono cechują się większą gęstością populacji, a na niebiesko mniejszą

8. Odnośniki

8.1. Kod oraz uzyskane wyniki

[ProjektEksploracjaDanych](#)

8.2. Spis obrazów

Rysunek 1 Brakujące wartości	4
Rysunek 2 PCA.....	8
Rysunek 3 TSNE	9
Rysunek 4 TSNE na PCA	11
Rysunek 5 MDS Age dependency ratio.....	12
Rysunek 6 MDS GDP	13
Rysunek 7 Dendrogram kryterium average	14
Rysunek 8 Dendrogram kryterium warda	15
Rysunek 9 Mapa ciepła kryterium average.....	16
Rysunek 10 Klasteryzacja hierarchiczna na wykresie PCA	18
Rysunek 11 Klasteryzacja hierarchiczna na wykresie PCA czynniki ekonomiczne.....	19
Rysunek 12 Klasteryzacja hierarchiczna na wykresie PCA czynniki socjologiczne	19
Rysunek 13 Klasteryzacja hierarchiczna na wykresie PCA czynniki geograficzne	20
Rysunek 14 Klasteryzacja hierarchiczna na wykresie TSNE	21
Rysunek 15 Klasteryzacja hierarchiczna na wykresie TSNE czynniki ekonomiczne.....	22
Rysunek 16 Klasteryzacja hierarchiczna na wykresie TSNE czynnik socjologiczne	22
Rysunek 17 Klasteryzacja hierarchiczna na wykresie TSNE czynniki geograficzne	22