**A blue and white logo with wings

Description automatically generated**

**Academia de Studii Economice București**

**Facultatea de Cibernetică, Statistică și Informatică Economică**

**Specializare: Informatică Economică**

**Proiect - Dezvoltare Software pentru analiza datelor**



**Analiza dezvoltării și stabilității în America Latină**

**Coordonator științific:**

**VINȚE Claudiu**

**Studenți:**

**Octavian GHEORGHIU**

**George-Daniel HEGHIȘ**

**Armand-Gabriel HANGU**

**Stefan-Daniel IGNAT**

# Cuprins

[**Cuprins 2**](#_Toc156147844)

[**Lista de figuri 3**](#_Toc156147845)

[**Introducere 4**](#_Toc156147846)

[**Capitolul 1 - Descrierea datelor 5**](#_Toc156147847)

[1.1. Anul de referință 5](#_Toc156147848)

[1.2. Descrierea variabilelor 5](#_Toc156147849)

[1.3. Descrierea observațiilor 8](#_Toc156147850)

[**Capitolul 2 - Abordarea privind analiza datelor 9**](#_Toc156147851)

[**Capitolul 3 - Analiza Componentelor Principale 10**](#_Toc156147852)

[3.1. Alegerea metodei de analiza a datelor 10](#_Toc156147853)

[3.2. Prezentarea rezultatelor 11](#_Toc156147854)

[3.2.1. Varianța explicată de componentele principale. 12](#_Toc156147855)

[3.2.2. Corelograma factorilor de corelație 13](#_Toc156147856)

[3.2.3. Determinarea scorurilor principale. 15](#_Toc156147857)

[3.2.4. Calitatea reprezentării punctelor pe axe. 16](#_Toc156147858)

[3.2.5. Contribuția instanțelor la varianta axelor 17](#_Toc156147859)

[3.2.6. Comunalitățile 17](#_Toc156147860)

[3.2.7. Cercurile corelațiilor 19](#_Toc156147861)

[3.2.8. Distribuția observațiilor în spațiul componentelor principale 20](#_Toc156147862)

[**Capitolul 4 - Analiza cluster 22**](#_Toc156147863)

[4.1. Alegerea metodei de analiza a datelor 22](#_Toc156147864)

[4.2. Prezentarea rezultatelor 23](#_Toc156147865)

[4.2.1. Metoda Elbow 25](#_Toc156147866)

[4.2.2. Metoda Silhouette 25](#_Toc156147867)

[4.3. Structurarea datelor pe 2 clustere. 26](#_Toc156147868)

[4.4. Structurarea datelor pe 3 clustere. 28](#_Toc156147869)

[**Concluzii 30**](#_Toc156147870)

[**Bibliografie 31**](#_Toc156147871)

# Lista de figuri

[Fig. 1: Setul de date complet. 11](#_Toc156150570)

[Fig. 2: Varianta componentelor principale. 12](#_Toc156150571)

[Fig. 3: Corelograma factorilor de corelație. 13](#_Toc156150572)

[Fig. 4: Rata inflației în componentele C1 și V2. 14](#_Toc156150573)

[Fig. 5: Scorurile componentelor principale standardizate. 15](#_Toc156150574)

[Fig. 6: Calitatea reprezentării rezultatelor. 16](#_Toc156150575)

[Fig. 7: Contribuția observatiilor la varianta axelor. 17](#_Toc156150576)

[Fig. 8: Harta comunalităților. 18](#_Toc156150577)

[Fig. 9: Corelația dintre variabile și C1, C2 19](#_Toc156150578)

[Fig. 10: Corelația dintre variabile și C3, C4 19](#_Toc156150579)

[Fig. 11: Distribuția observațiilor în spațiul C1, C2 20](#_Toc156150580)

[Fig. 12: Distribuția observațiilor în spațiul C3, C4 20](#_Toc156150581)

[Fig. 13: Matricea distanțelor , date standardizate 24](#_Toc156150582)

[Fig. 14: Valoarea funcției Elbow. 25](#_Toc156150583)

[Fig. 15: Partiția cu 2 clustere. 26](#_Toc156150584)

[Fig. 16: Dendrograma cu 2 clustere. 27](#_Toc156150585)

[Fig. 17: Partiția cu 3 clustere. 28](#_Toc156150586)

[Fig. 18: Dendrograma cu 3 clustere. 29](#_Toc156150587)

# Introducere

Continentul sud-american și America Centrala reprezintă o regiune diversă, cu o istorie bogată și o cultură vibrantă. Cu toate acestea, regiunea se confruntă, de asemenea, cu o serie de provocări, precum corupția, instabilitatea economică și criminalitatea.

Corupția, un fenomen major în regiune, a penetrat adânc în structurile de putere, afectând instituțiile cheie și subminând fundamentul democrației. Prin analiza riguroasă a datelor, devine evident că corupția nu este doar o problemă în sine, ci și o sursă de instabilitate și inechitate. O consecință directă a acestei practici este slăbirea serviciilor publice și a instituțiilor, de la sistemul de sănătate până la educație.

Instabilitatea economică este o altă problemă majoră cu care se confruntă regiunea. Multe țări din America Latină și Caraibe au avut o istorie de instabilitate economică, care a dus la sărăcie, inegalitate și conflicte. Criza economică globală din 2008 a avut un impact deosebit de puternic asupra regiunii, ducând la o creștere a sărăciei și inegalității. De asemenea, factori precum datoria externă, inflația și lipsa diversificării economice au adăugat presiune asupra capacității regiunii de a asigura bunăstarea populației sale. Desigur, există excepții, precum Argentina, Uruguay, si Chile care au înregistrat progrese economice semnificative și au avut perioade de creștere stabilă. Cu toate acestea, în alte țări sud-americane, disparitățile economice sunt evidente, iar o parte semnificativă a populației se confruntă cu sărăcia și lipsa accesului la servicii de bază.

În același timp, criminalitatea s-a manifestat sub diverse forme, de la infracțiuni obișnuite la rețele de trafic de droguri și conflicte armate. Analiza datelor relevă conexiuni complexe între sărăcie, inegalitate socială și ratele de criminalitate, conturând un peisaj în care autoritățile se confruntă cu provocări semnificative în asigurarea securității și aplicarea legii.

Corupția, instabilitatea economică și criminalitatea sunt probleme complexe care afectează negativ America Latină și Caraibe. Aceste probleme sunt interconectate și au un impact negativ asupra dezvoltării economice, sociale și politice a regiunii. Așadar, lucrarea de față își propune să ofere o explicație pentru problemele din această regiune a lumii.

# Capitolul 1 - Descrierea datelor

## 1.1. Anul de referință

Anul de referință ales pentru toate variabilele studiate este 2021.

## 1.2. Descrierea variabilelor

**1) Human Development Index**

Denumirea completă a variabilei**:** Indicele Dezvoltării Umane (Human Development Index)

Unitatea de măsură: HDI oferă o evaluare relativă a dezvoltării umane comparativ cu alte țări, dar nu exprimă o cantitate în termeni fizici.

Definiție: Indicele Dezvoltării Umane (HDI) este o măsură sumară a realizării medii în dimensiunile cheie ale dezvoltării umane: o viață lungă și sănătoasă, cunoștințe și un nivel de viață decent. HDI este media geometrică a indicilor normalizați pentru fiecare dintre cele trei dimensiuni.

Sursa: Human Development reports, <https://hdr.undp.org/data-center/human-development-index#/indicies/HDI>

**2) Life expectancy at birth**

Denumirea completă: Speranța de viață la naștere(Life expectancy at birth)

Unitate de măsură: Ani de viață

Definiție : Speranța de viață la naștere indică numărul de ani pe care i-ar trăi un nou-născut dacă tiparele predominante de mortalitate la momentul nașterii ar rămâne aceleași pe tot parcursul vieții.

Sursa: Human Development reports, [https://hdr.undp.org/data-center/](https://hdr.undp.org/data-center/human-development-index#/indicies/HDI)

**3) Mean years of schooling**

Denumirea completă a variabilei: Numărul mediu de ani de studii (Mean years of schooling)

Unitatea de măsură: Ani de școlarizare

Definiție: Numărul mediu de ani de studii finalizați ai populației unei țări cu vârsta de 25 de ani și mai mult, excluzând anii petrecuți repetarea claselor individuale.

Sursa: Human Development reports, [https://hdr.undp.org/data-center/](https://hdr.undp.org/data-center/human-development-index#/indicies/HDI)

**4) Gross national income (GNI) per capita**

Denumirea completă a variabilei: Venitul național brut per capita (VNB) (Gross national income (GNI) per capita)

Unitatea de măsură: Valuta țării analizate

Definiție: Venitul Național Brut pe locuitor este valoarea în dolari a venitului final al unei țări într-un an împărțită la populația acesteia. Valoarea reflectă venitul mediu înainte de impozitare al cetățenilor țării respective.

Sursa: Human Development reports, [https://hdr.undp.org/data-center/](https://hdr.undp.org/data-center/human-development-index#/indicies/HDI)

5) **Homicide Rate**

Denumirea completă a variabilei : Rata omuciderilor la 100,000 de locuitori (Homicide Rate per 100.000 people)

Unitate de măsură : număr de omucideri

Definiție: Rata omuciderilor sau rata crimelor cu omor reprezintă numărul de omucideri raportat la 100.000 de locuitori , pe o perioada de timp de un an , aceasta fiind o măsură importantă în analiza criminalității și a siguranței publice într-o anumită țară.

Sursa: Statista, <https://www.statista.com/statistics/947781/homicide-rates-latin-america-caribbean-country/>

6) **Political Stability**

Denumirea completă a variabilei : Stabilitate politică și absența violenței/terorismului (Political Stability and Absence of Violence/Terrorism)

Unitatea de măsură: Rangul percentilei (0-100) indică rangul țării dintre toate țările din lume. 0 corespunde rangului cel mai mic și 100 corespunde rangului cel mai mare.

Definiție: Stabilitatea politică și absența violenței/terorismului măsoară percepția asupra probabilității instabilității politice și/sau violenței motivate politic, inclusiv terorismul.

Sursa: World bank,  <https://www.worldbank.org/en/publication/worldwide-governance-indicators/interactive-data-access>

7) **Government Effectiveness**

Denumirea completă a variabilei : Eficacitatea Guvernului (Government Effectiveness)

Unitatea de măsură : Rangul percentilei (0-100) indică rangul țării dintre toate țările din lume. 0 corespunde rangului cel mai mic și 100 corespunde rangului cel mai mare.

Definiție : Eficacitatea guvernului surprinde percepțiile despre calitatea serviciilor publice, calitatea serviciului public și gradul de independență a acestuia față de presiunile politice, calitatea formulării și implementării politicilor și credibilitatea angajamentului guvernului față de astfel de politici.

Sursa: World bank,  <https://www.worldbank.org/en/publication/worldwide-governance-indicators/interactive-data-access>

8) **Control of Corruption**

Denumire completă a variabilei : Controlul Corupției (Control of Corruption)

Unitatea de măsură: Rangul percentilei (0-100) indică rangul țării dintre toate țările din lume. 0 corespunde rangului cel mai mic și 100 corespunde rangului cel mai mare

Definiție: Controlul corupției surprinde percepțiile cu privire la măsura în care puterea publică este exercitată în beneficiul privat, inclusiv atât mărunt, cât și măreț, forme de corupție, precum și „capturarea” statului de către elite și interese private.

Sursa: World bank,  <https://www.worldbank.org/en/publication/worldwide-governance-indicators/interactive-data-access>

**9) Drug Use**

Denumire completă a variabilei : Consumul de substanțe interzise Rata mortalitate (la 100.000 de oameni ) (Drug use death rate per 100 000 )

Unitatea de măsură: Oameni decedați

Definiție: Decese în care cauza de bază a decesului se datorează consumului de opioide, canabinoizi, sedative, cocaină și alte stimulente, inclusiv cofeina, halucinogene sau consumul de droguri multiple raportate la 100.000 .

Sursa:World population review, <https://worldpopulationreview.com/country-rankings/drug-use-by-country>

**10) Inflation rate**

Denumirea completă a variabilei : Rata inflației (Inflation rate )

Unitatea de măsură: Procente

Definiție: Inflația este rata de creștere a prețurilor într-o anumită perioadă de timp. Inflația este de obicei o măsură largă, cum ar fi creșterea globală a prețurilor sau creșterea costului vieții într-o țară.

Sursa: The global economy, <https://www.theglobaleconomy.com/rankings/Inflation/>

**11) Unemployment rate**

Denumirea completă a variabilei : Rata șomajului (Unemployment rate )

Unitatea de măsură: Procente

Definiție: Rata șomajului (conform criteriilor Biroului Internațional al Muncii) reprezintă raportul dintre numărul șomerilor și populația activă, exprimat procentual.

Sursa: World bank, <https://data.worldbank.org/indicator/SL.UEM.TOTL.ZS>

**12) Rule of Law**

Denumire completă a variabilei: Statul de drept (Rule of Law)

Unitatea de măsură: Rangul percentilei (0-100) indică rangul țării dintre toate țările din lume. 0 corespunde rangului cel mai mic și 100 corespunde rangului cel mai mare

Definiție: Statul de drept surprinde percepțiile cu privire la măsura în care agenții au încredere și respectă regulile societății și, în special, calitatea executării contractelor, drepturile de proprietate, poliția și instanțele, precum și probabilitatea crimei și violenței.

Sursa: World bank,  <https://www.worldbank.org/en/publication/worldwide-governance-indicators/interactive-data-access>

## 1.3. Descrierea observațiilor

Pentru analiza, sunt observate 24 tari aparținând Americii Latine și anume: Argentina, Belize, Bolivia, Brazilia, Chile, Columbia, Costa Rica, Cuba, Republica Dominicană, Ecuador, El Salvador, Guatemala, Guyana, Haiti, Honduras, Jamaica, Mexic, Nicaragua, Panama, Paraguay, Peru, Surinam, Uruguay, Venezuela.

# Capitolul 2 - Abordarea privind analiza datelor

În cadrul analizei detaliată prezentate, ne concentrăm pe observarea atentă a 12 variabile provenite din domenii diverse (Politică, Educație, Justiție, Economie), în contextul a 24 de state din regiunea Americii Latine. Obiectivul principal al acestui studiu este să identifice și să stabilească variabilele care au o influență semnificativă în determinarea stării actuale a acestui continent. Prin eliminarea redundanțelor, cercetarea își propune să identifice cauzele fundamentale care subliniază în mod preponderent problemele cu care se confruntă America Latină, precum creșterea criminalității și inflația, dar și să încerce sa răspundă la presupuneri precum: “Este criminalitatea un efect al sărăciei?”, “Este inflația un răspuns al unei guvernări deficitare?”, “Poate un sistem judiciar precar sa favorizeze o rata mai mare a omuciderilor?, dar și alte concluzii care nu au fost identificate inițial ca ipoteze.

De asemenea, odată cu găsirea acelor variabile care explică cel mai bine situația curentă a regiunii, analiza realizată ne va permite identificarea principalelor probleme, cât și luarea de decizii eficiente asupra acestora, care să conducă la o dezvoltare mult mai accelerată.

Concluzionând, prin analiza datelor colectate se urmărește:

* Redimensionarea tabloului de date prin păstrare într-un procent semnificativ a informației prezente pentru a facilita viitoare analize.
* Observarea legăturilor dintre variabile și cum se afectează reciproc în raportarea cu ritmul de dezvoltare a țărilor din regiune.
* Determinarea acelor repere fundamentale asupra cărora instituțiile abilitate ar trebui să acorde o deosebită atenție pentru a asigura statelor pe care le guvernează o dezvoltare sustenabilă.

Pentru a putea fi comparabile, datele vor fi standardizate, adică vor fi transformate astfel încât să aibă o medie de 0 și o deviație standard de 1. Această standardizare este crucială în analiza componentelor principale si analiza cluster, deoarece cele 2 sunt sensibile la variația absolută a datelor. Prin standardizarea datelor, toate variabilele vor avea aceeași scală, eliminând astfel posibilele distorsiuni introduse de diferențele de unități de măsură sau magnitudine între variabile.

# Capitolul 3 - Analiza Componentelor Principale

## 3.1. Alegerea metodei de analiza a datelor

Analiza Componentelor Principale (ACP) reprezintă o metodă esențială în investigarea relațiilor complexe dintre variabilele multiple și identificarea principalelor tendințe sau structuri subiacente în date. Alegerea de a utiliza ACP în analiza regiunii America Latină este susținută de câteva motive fundamentale care contribuie la obținerea unei înțelegeri mai profunde și mai sistematice a diversității socio-economice și politice din această regiune.

În primul rând, ACP se dovedește a fi deosebit de utilă în situațiile în care există un număr mare de variabile intercorelate, cum ar fi cele selectate pentru analiza noastră, cum ar fi HDI, speranța de viață, durata medie a școlarizării, venitul pe cap de locuitor și ratele de criminalitate. ACP permite reducerea dimensiunii setului de date, conservând în același timp informațiile semnificative. Acest aspect este esențial în cazul nostru, având în vedere complexitatea interacțiunilor dintre variabilele socio-economice și politice din America Latină.

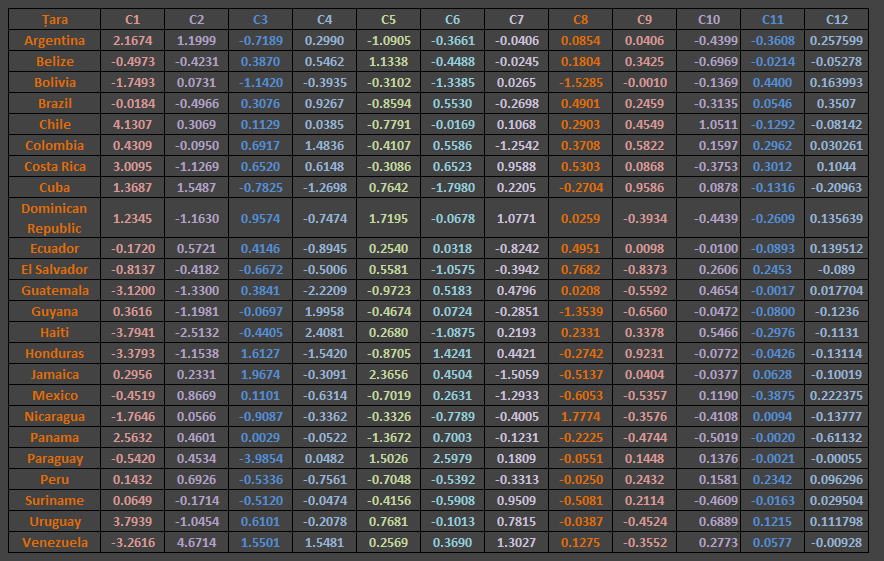
ACP are, de asemenea, capacitatea de a scoate la iveală modele ascunse în datele noastre, fără a fi nevoie de înțelegerea complicată a fiecărei variabile. Acest aspect este benefic într-un context complex precum America Latină, unde interacțiunile dintre factorii socio-economici și politici pot fi dificil de deslușit.

Astfel, ACP reprezintă o alegere logică pentru analiza noastră, oferindu-ne o cale de a face datele mai ușor de gestionat, de a le interpreta într-un mod uniform și de a descoperi conexiuni esențiale care ar putea rămâne ascunse într-o abordare mai tradițională. Alegerea ACP se aliniază obiectivelor noastre de a realiza o analiză comprehensivă și accesibilă, contribuind la o înțelegere mai profundă a complexității regiunii Americii Latine.

## 3.2. Prezentarea rezultatelor

În cadrul ACP, definim drept componente principale acele variabile vectoriale abstracte reprezentate sub forma unor combinații liniare ale variabilelor inițiale, combinații ce sunt independente două câte două și a căror sumă a pătratelor coeficienților care o definesc este egală cu unitatea.

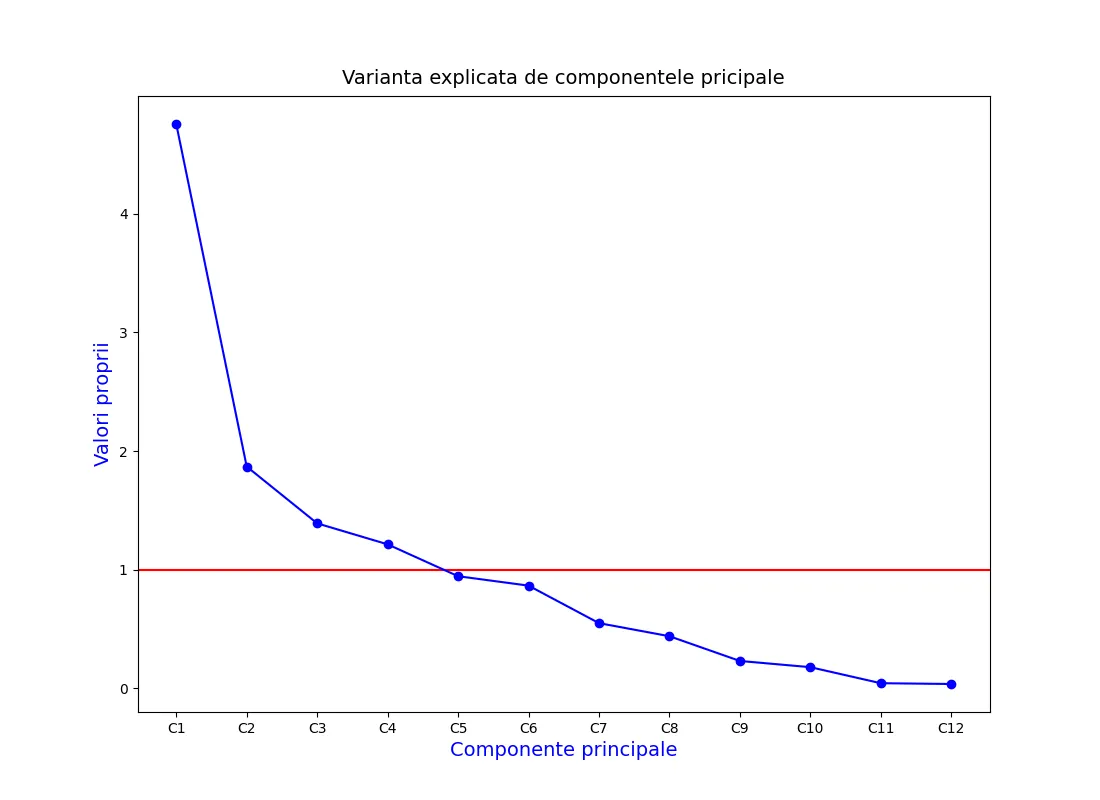
La nivelul analizei, au fost identificate 12 componente principale, și anume:



### Fig. 1: Setul de date complet.

Foarte important de reținut este faptul că deși numărul de componente principale este egal cu numărul variabilelor, nu toate componentele au o semnificație informațională considerabilă. Așadar, în cele ce urmează, va fi necesar să decidem care dintre componentele principale identificate conțin un grad ridicat de informație.

#### 3.2.1. Varianța explicată de componentele principale.



### Fig. 2: Varianta componentelor principale.

Odată ce am identificat cele 12 componente principale, focalizarea cercetării noastre se axează pe identificarea celor cu cea mai semnificativă varianță. Astfel, ne propunem să determinăm acele componente care includ nivelul maxim de informație din setul inițial de date și să efectuăm o reducere a dimensionalității acestuia. Acest demers urmărește eliminarea redundanței, menținând totuși o perspectivă globală asupra observațiilor. Din perspectiva graficului prezentat anterior, constatăm că primele 4 componente, selectate conform criteriului Kaiser (care implică alegerea numărului de axe pentru care valorile proprii sunt mai mari decât 1), rețin cea mai semnificativă cantitate de varianță din tabloul inițial de date.

alpha = [4.7567058 1.86951872 1.39083033 1.21399507 0.94530059 0.86592564 0.54992073 0.4395642 0.23104954 0.17858873 0.04348369 0.0368561 ]

Determinăm ponderea varianței celor 4 componente alese din totalul varianței de la nivelul setului de date:

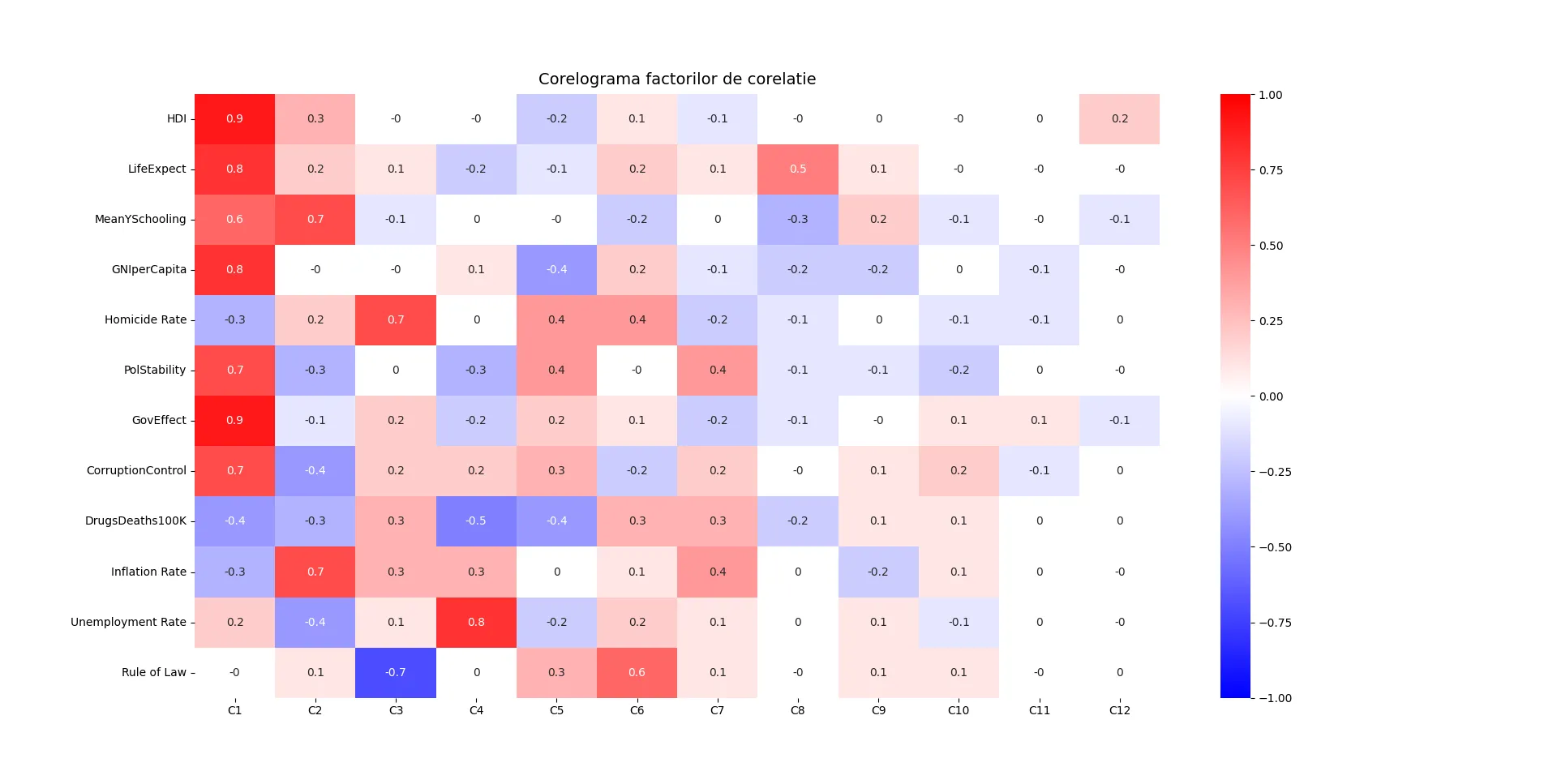
*S4=*4.7567058+1.86951872+1.39083033+1.21399507=**9.23104992**

S12=4.7567058+1.86951872+1.39083033+1.21399507+0.94530059+0.86592564+ 0.54992073+0.4395642+0.23104954+0.17858873+0.04348369+0.0368561=**12.52173914**

Pondere = 9.23104992 / 12.52173914 = **0.7372019028**

Astfel, prin redimensionarea tabloului inițial de date, păstrând doar primele 4 componente principale (întrucât ele conțin cea mai mare varianță) reușim să reținem aproximativ **73,7%** din informația prezentată. Diferența de aproximativ 26.3% în cazul nostru reprezintă o pierdere de informații.

#### 3.2.2. Corelograma factorilor de corelație



### Fig. 3: Corelograma factorilor de corelație.

Figura anterioară, surprinde corelograma dintre variabilele inițiale și componentele principale identificate pe parcursul analizei, reliefând astfel informații fundamentale asupra nivelului în care componentele principale sintetizează informația conținută în variabilele originale. Așadar, elementele acestei matrici – cunoscute drept intensități ale factorilor – reprezintă coeficienții de corelație dintre o anumită variabilă inițială și o componentă principală. Aceste intensități ale factorilor pot fi considerate indicator ai măsurii în care variabilele originale participă la formarea componentelor principale.

*Exemplu*:



### Fig. 4: Rata inflației în componentele C1 și V2.

Drept exemplificare pentru aspectele prezentate anterior, vom evalua intensitatea factorului dintre componenta principală C2 și variabila inițială Inflation\_Rate. În cadrul figurii 4, observăm că măsura în care Inflation\_Rate participă la formarea componentei principale C2 este foarte ridicată (valoarea indicatorului > 0.5). Așadar, în sens invers, putem spune despre componenta principală C2 că sintetizează într-o foarte mare măsură informația regăsită la nivelului variabilei originale Inflation\_Rate.

Totodată, matricea factor este un reper fundamental în cadrul cercetării efectuate întrucât, prin intermediul analizei elementelor sale putem identifica o serie de partiții sau clustere pe mulțimea variabilelor originale. Aceste partiții, prin asocierea cu anumite componente principale, permit stabilirea unor semnificații intuitive. Prin urmare, analiza intensităților conduce către determinarea acelor variabile inițiale care sunt reprezentate prin intermediul unei anumite componente principale, creând astfel posibilitatea de a identifica semnificații concrete pentru fiecare dintre acestea.

În cadrul cercetării efectuate, prin intermediul analizei matricei factor, deducem următoarele:

Componenta principală C1 sintetizează predominant informația din următoarele variabile: HDI (intensitate=0.92), LifeExpect (intensitate=0,76), GNIperCapita (intensitate=0,83), PolStability (intensitate=0,73), GovEffect (intensitate=0,90), CorruptionControl (intensitate=0,74). Putem astfel să caracterizăm această componentă ca fiind o *Guvernare bună și o calitate superioară a vieții.*

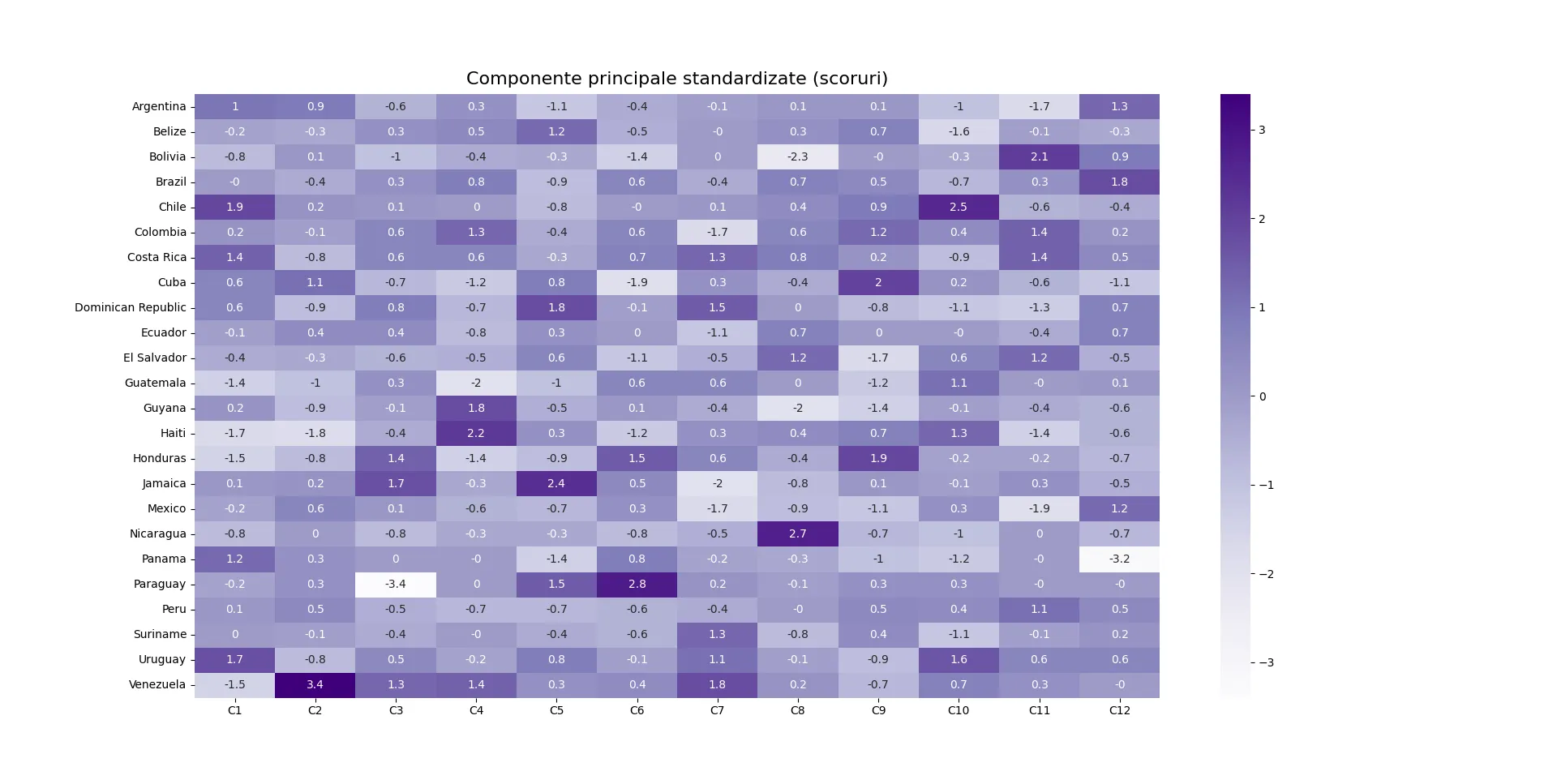
Componenta principală C2 sintetizează predominant informația din următoarele variabile: MeanYSchooling (intensitate=0,73), Inflation Rate(intensitate=0,74). În acest caz, datorită observației Venezuela care este un outlier, având hiperinflație, dar totodată prezentând o medie a anilor de școlarizare bună, vom vedea mai târziu în graficul scorurilor componentelor principale că această țară este predominant prezentă în calculul acestei componente. Prin urmare, vom denumi această componentă *Inflație și educație.*

Componenta principală C3 sintetizează predominant 2 variabile inițiale: Homicide Rate (intensitate=0,73) și Rule of Law (intensitate=-0,73). Aceste aspecte pot fi interconectate și, prin urmare, merită să ne întrebăm mai departe în ce fel se influențează cele 2 variabile una pe alta. Putem presupune că o rată înaltă a omuciderilor și un nivel redus al statului de drept pot sugera instabilitate socială, ineficiența sistemului de justiție penală, corupție, sărăcie și disparități sociale. Lipsa aplicării legii și a drepturilor individuale poate contribui la un mediu în care violența este mai frecventă și în care soluțiile pașnice sunt mai puțin probabile. Îmbunătățirea statului de drept și a sistemului de justiție poate juca un rol esențial în reducerea omuciderilor și a conflictelor. Concluzionând, componenta C3 surprinde *Criminalitatea și ineficiența sistemului judiciar.*

Componenta principală C4 sintetizează variabila Unemployment Rate (intensitate=0,80), dar și o intensitate semnificativă pentru DrugsDeaths100k (intensitate=-0,52). O interpretare preliminară ar putea fi că în țările cu rate mai mari de șomaj tind să aibă mai puține probleme cu drogurile, însă problema trebuie privită într-un context mai larg: În țările bogate, există unele tendințe care pot contribui la probleme legate de droguri, cum ar fi accesul mai ușor la substanțe, stresul și presiunile sociale. În aceste țări, sistemul de sănătate și programele de reducere a riscurilor pot ajuta la gestionarea efectelor negative ale consumului de droguri, iar resursele financiare pot fi direcționate către tratament și prevenție. În țările sărace, accesul limitat la resursele de sănătate, educarea precară și instabilitatea socio-economică pot contribui la problemele legate de droguri. În plus, lipsa unor programe eficiente de prevenire și tratament poate agrava situația. Așadar, pentru moment, doar vom denumi componenta ca *Rata șomajului și scăderea incidenței decesului de la droguri.*

#### 3.2.3. Determinarea scorurilor principale.

În cadrul ACP, coordonatele obiectelor analizate în spațiul redus (de dimensionalitate redusă) se mai numesc și scoruri principale. În urma aplicării algoritmului de analiză în componente principale obținem următoarea corelogramă:

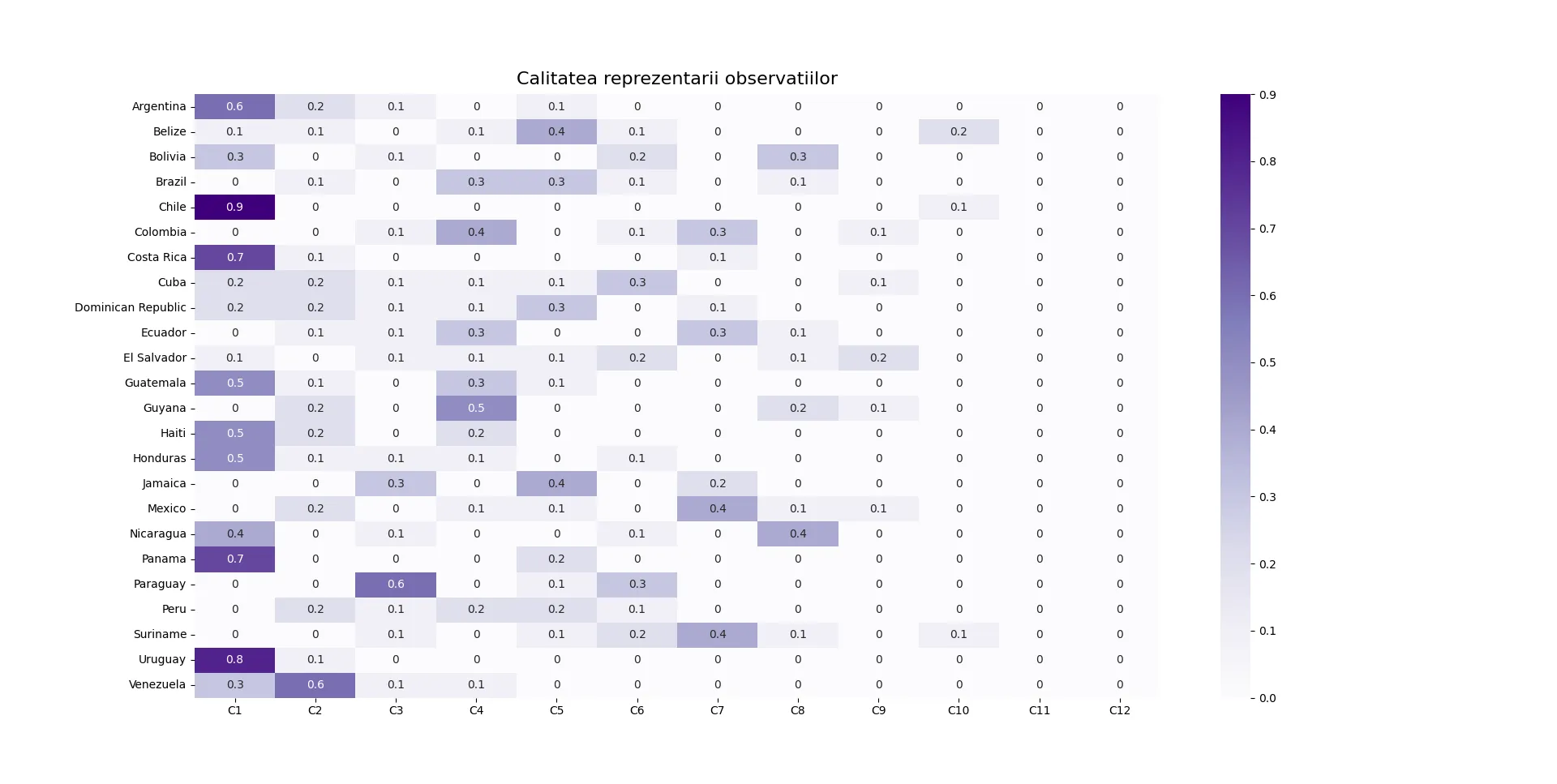


### Fig. 5: Scorurile componentelor principale standardizate.

Analizând matricea obținută (reprezentată în figura 5) putem afirma ca liniile sale reprezintă scorurile corespunzătoare noilor variabile sau, mai exact, observațiilor celor 12 componente principale. O valoare absolută a acestor factori de încărcare mai mare de 0.5 asociată variabilelor reflectă faptul că acestea contribuie după caz (într-o manieră pozitivă sau negativă) la componenta principală (Ex: Scorul componentei C1 pentru Argentina = 1. Este pozitiv, iar valoarea sa absolută > 0.5 => contribuție pozitivă. Scorul componentei C1 pentru Haiti = -1.5. Este negativ, iar valoarea sa absolută > 0.5 => contribuție negativă). Prin urmare scorurile au valori diferite, mai mari sau mai mici, pentru fiecare observație în parte. Un alt aspect esențial al scorurilor este faptul că acestea sunt mai puțin afectate de erori, deci putem spune că au o anumită invarianță în raport cu erorile, devenind astfel mult mai calitative din punct de vedere informațional în comparație observațiile inițiale. Date fiind aceste aspecte, scorurile principale pot să deservească drept substituenți ai observațiilor, simplificând în acest mod baza de date.

#### 3.2.4. Calitatea reprezentării punctelor pe axe.

O reprezentare bună pe axe indică cât de bine variabilele individuale sunt proiectate pe noile axe principale. În cadrul cercetării efectuate asupra celor 24 de observații, prin aplicarea ACP, calitatea reprezentării observațiilor este descrisă de figura următoare:

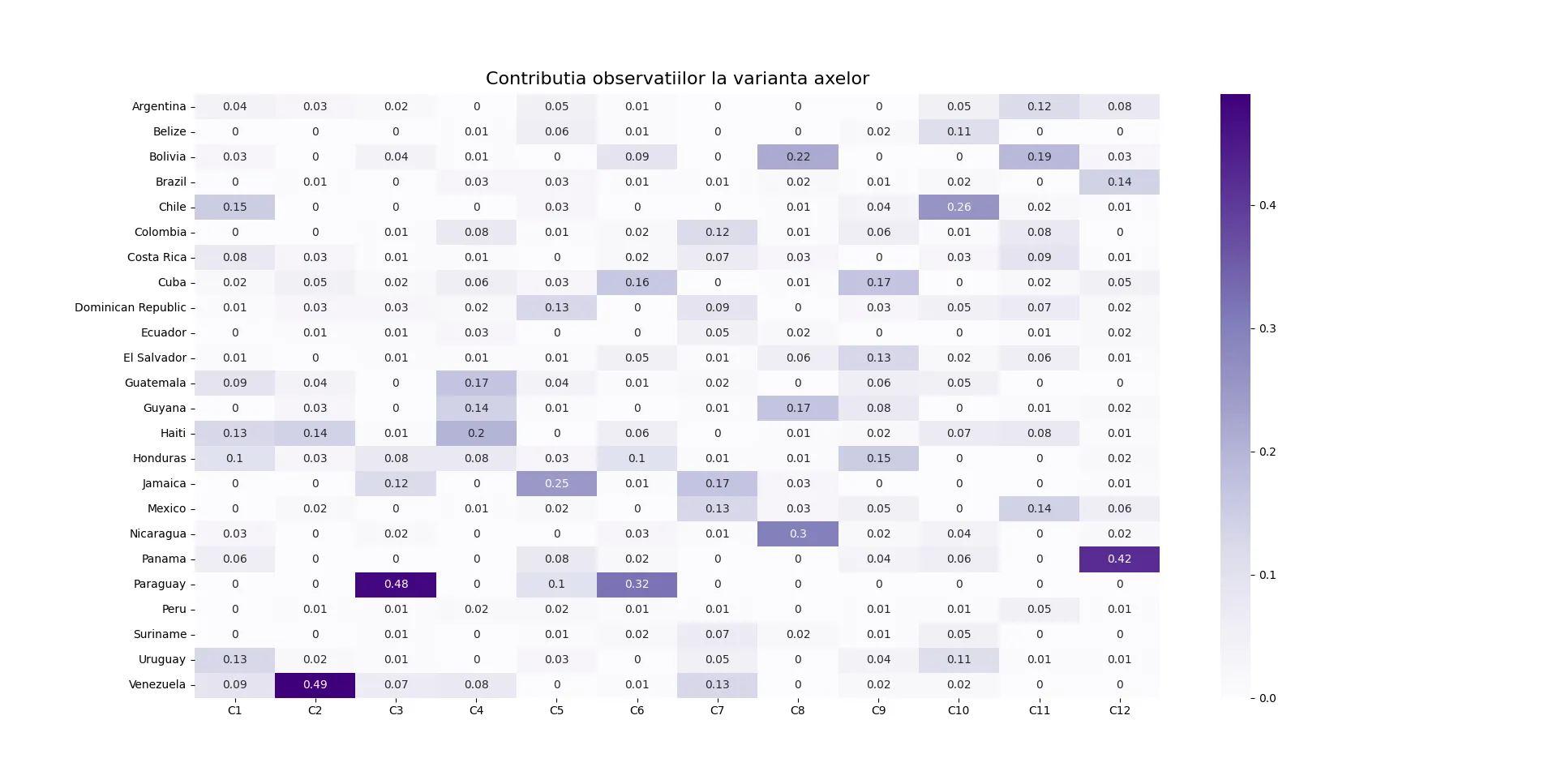


### Fig. 6: Calitatea reprezentării rezultatelor.

La nivelul Corelogramei calității reprezentării observațiilor sunt determinate pentru fiecare instanță în parte valoarea cosinusurilor dintre vectorii punct asociați acestora și axele asociate fiecărei componente principale. Astfel deducem că există variabile calitativ reprezentante în spațiul componentelor, precum Chile și Argentina în C1 datorită valorilor ridicate ale cosinusurilor (peste 0,5), dar și variabile mai puțin calitativ reprezentate de țări precum Belize în C1. Totodată observăm că există spații în care nicio observație nu este reprezentată calitativ, spații precum (C11, C12).

#### 3.2.5. Contribuția instanțelor la varianta axelor

În cadrul cercetării de față, este bine cunoscut faptul că fiecare observație în parte aduce componentelor un plus de varianță. Pentru a determina exact care este contribuția fiecărei observații, prin aplicarea PCA, s-a realizat următoarea figură:



### Fig. 7: Contribuția observatiilor la varianta axelor.

În cadrul Corelogramei contribuției instanțelor la varianța axelor componentelor, putem

observa aspecte fundamentale precum:

C1 --> contribuția cea mai mare este adusă de țări precum: Chile, Haiti și Uruguay .

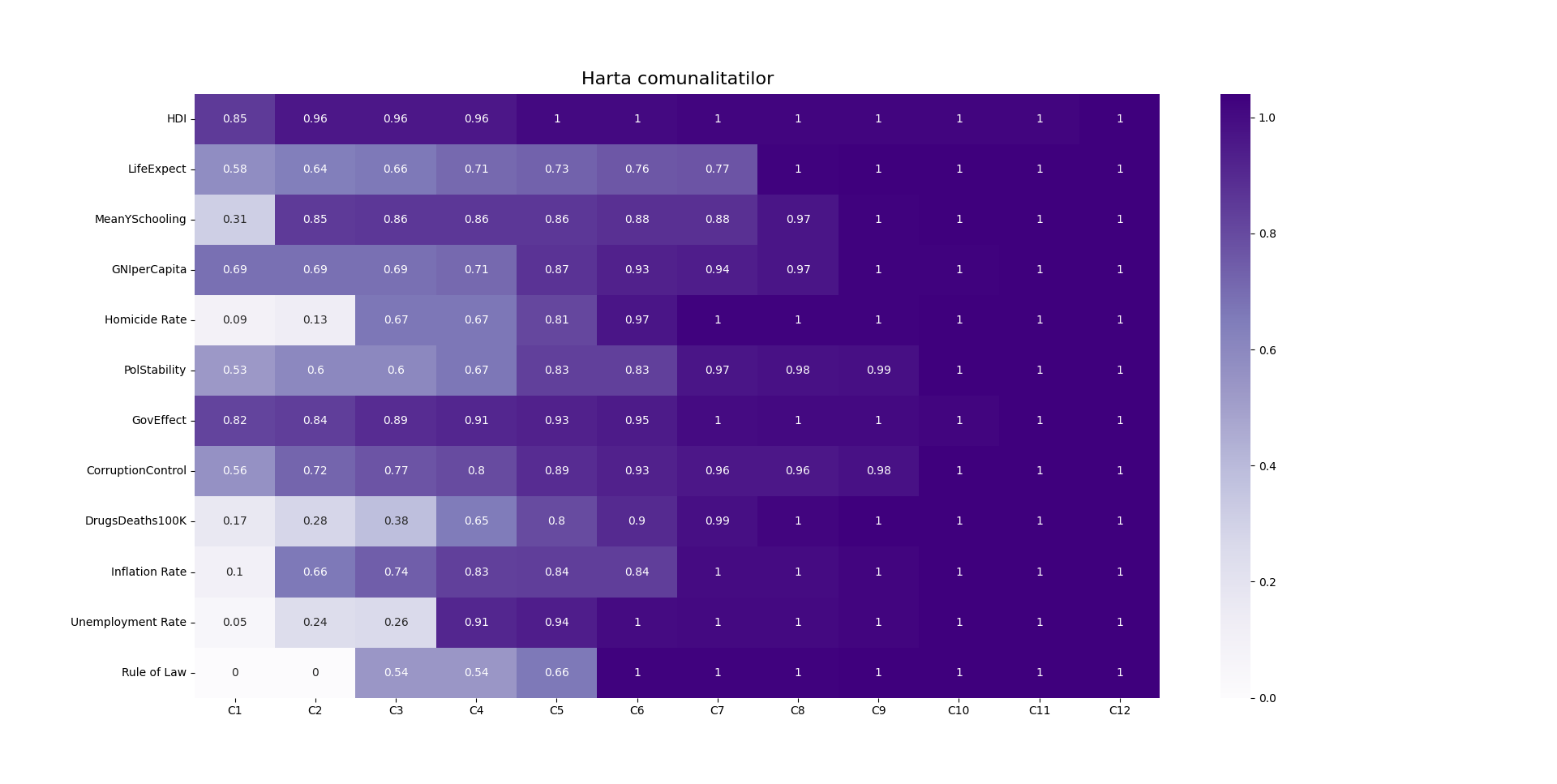
C2 --> contribuția cea mai mare este adusă de țări precum: Venezuela și Haiti.

C3 --> contribuția cea mai mare este adusă de țări precum: Paraguay și Jamaica.

C4 --> contribuția cea mai mare este adusă de țări precum: Haiti, Guatemala, Guyana.

#### 3.2.6. Comunalitățile

Cantitatea de varianță explicată în comun de către un grup de componente principale poartă numele de Comunalitate. În cadrul figurii de mai jos, prin intermediul aplicării ACP, au fost identificate comunitățile și cantitatea de informație din tabloul de date inițial pe care acestea le rețin:



### Fig. 8: Harta comunalităților.

Așadar, după cum am văzut anterior, componentele noastre principale înglobează aproximativ 73,7% din totalul varianței existente la nivelul variabilelor modelului. În acest sens, corelograma comunalităților ne permite să determinăm câtă informație este extrasă din fiecare variabilă de către colectivitatea primelor 4 componente principale, și anume:

* 96% din varianța variabilei HDI
* 71% din varianța variabilei LifeExpect
* 86% din varianța variabilei MeanYSchooling
* 71% din varianța variabilei GNIperCapita
* 67% din varianța variabilei Homicide Rate
* 67% din varianța variabilei PolStability
* 91% din varianța variabilei GovEffect
* 80% din varianța variabilei CorruptionControl
* 65% din varianța variabilei DrugsDeaths100K
* 83% din varianța variabilei Inflation Rate
* 91% din varianța variabilei Unemployment
* 54% din varianța variabilei Rule of Law

#### 3.2.7. Cercurile corelațiilor

### 

|  |  |
| --- | --- |
| Fig. 9: Corelația dintre variabile și C1, C2 | Fig. 10: Corelația dintre variabile și C3, C4 |

Cercul corelațiilor dintre două componente principale este o reprezentare grafică a relației dintre cele două componente obținute prin analiza componentelor principale. Acest lucru oferă o perspectivă vizuală asupra modului în care variabilele originale contribuie la formarea componentelor principale.

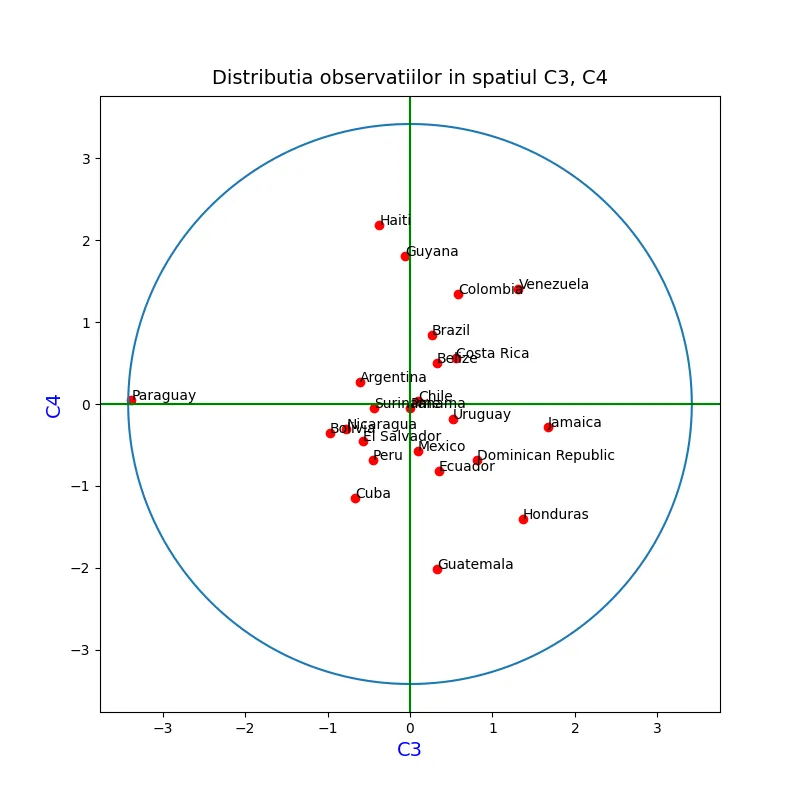
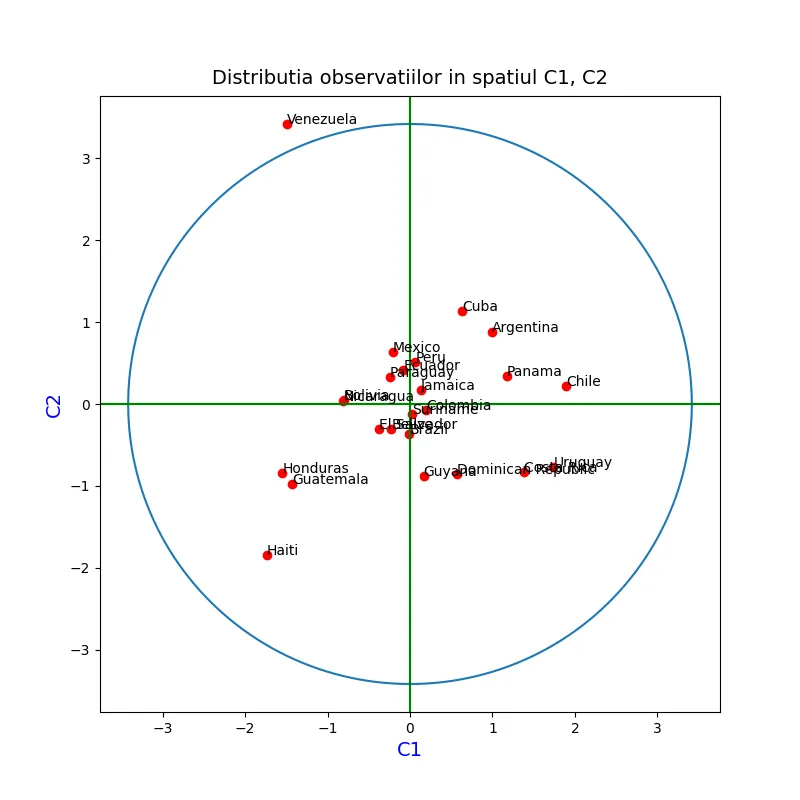
Analizând corelațiile dintre variabilele inițiale și componentele C1, C2, se evidențiază următoarele:

* Variabilele GNIperCapita, GovEffect, PolStability, CorruptionControl, Unemployment Rate sunt grupate preponderent în cadranul IV – rezultă legături puternice pe axa variabilei C1 (*Guvernare bună și o calitate superioară a vieții*);
* Variabila DrugsDeaths100K se afla în cadranul III – rezultă că nu există legături puternice ori medii pe axa variabilei C1 (*Guvernare bună și o calitate superioară a vieții*) , ori pe axa variabilei C2 (*Inflație și educație*);
* Variabilele Rule of Law, Inflation Rate și Homicide Rate sunt grupate în cadranul II– rezultă legături medii și respectiv puternice pe axa variabilei C2 (*Inflație și educație*);
* Variabilele Life Expect, HDI și MeanYSchooling sunt grupate preponderent în cadranul I – rezultă că există legături puternice pe axa variabilei C1 (*Guvernare bună și o calitate superioară a vieții*) , dar și pe axa variabilei C2 (*Inflație și educație*).

Analizând corelațiile dintre variabilele inițiale și componentele C3, C4, se evidențiază următoarele:

* Variabilele Life Expect, GovEffect, PolStability și DrugsDeaths100K sunt grupate preponderent în cadranul IV – fapt ce sugerează legături puternice pe axa variabilei C3 (*Criminalitatea și ineficiența sistemului judiciar*);
* Variabila HDI se afla în cadranul III – fapt ce sugerează că nu există legături puternice pe axa variabilei C3 (*Criminalitatea și ineficiența sistemului judiciar*) , ori pe axa variabilei C4 (*Rata șomajului și scăderea incidenței decesului de la droguri*);
* Variabilele Rule of Law, GNIperCapita si MeanYSchooling sunt grupate în cadranul II – fapt ce sugerează legături medii pe axa variabilei C4 (*Rata șomajului și scăderea incidenței decesului de la droguri*);
* Variabilele Unemployment Rate, Inflation Rate, Homicide Rate si CorruptionControl este grupată în cadranul I – fapt ce sugerează că există legături medii pe axa variabilei C3 (*Criminalitatea și ineficiența sistemului judiciar*), dar și legături puternice pe axa variabilei C4 (*Rata șomajului și scăderea incidenței decesului de la droguri*).

#### 3.2.8. Distribuția observațiilor în spațiul componentelor principale



|  |  |
| --- | --- |
| Fig. 11: Distribuția observațiilor în spațiul C1, C2 | Fig. 12: Distribuția observațiilor în spațiul C3, C4 |

La nivelul celor două distribuții, accentul este pus pe analizarea grupării observațiilor pentru componentele C1 și C2, în prima figură, și pentru componentele C3 și C4 în cadrul celei de-a doua.

La o primă observație, în cadrul primei figuri, putem afirma că între marea majoritate a țărilor analizate (spre exemplu Jamaica, Columbia, Suriname) și cel puțin una dintre componentele C1și C2 există o corelație medie, semnalând astfel că impedimentele care stau în calea dezvoltării mult mai rapide a continentului african sunt reprezentate într-o măsură medie de *Guvernare bună și o calitate superioară a vieții* și *Inflație și educație*. Totodată, analiza de față reușește să reliefeze și observații peste medie, la nivelul căreia guvernarea este atent gestionată de autoritățile competente, iar educația și o calitate superioară a vieții conferă un avantaj social al viețuirii într-o țară orientată către progres, făcând astfel din Argentina, Chile și Panama cele mai dezvoltate țări din America Latină prin raportarea la cele două componente. De asemenea, prima figură nuanțează și existența unor observații aflate sub medie, la nivelul cărora guvernarea precară și calitatea inferioară a vieții, a eficienței scăzute în ceea ce privește guvernarea și calitatea superioară a vieții mult prea scăzute reprezintă o piedică reală în cadrul procesului de dezvoltare. Astfel, țări precum Haiti, Guatemala și Honduras, prin prisma celor două deziderate (componente principale) vizate sunt cele mai subdezvoltate.

Analog cazului anterior, la nivelul celei de-a doua figuri, putem afirma că între marea majoritate a țărilor analizate (spre exemplu Chile, Mexico, Uruguay) și cel puțin una dintre componentele C3 și C4 există o corelație medie, semnalând astfel că impedimentele care stau în calea dezvoltării mult mai rapide a statelor latino-americane sunt reprezentate într-o măsură medie de *Rata șomajului și scăderea incidenței decesului de la droguri* și *Criminalitatea și ineficiența sistemului judiciar*. Analiza de față surprinde și țări precum Guatemala la nivelul cărora variabilele amintite anterior sunt adesea trecute cu vederea și ajung să cauzeze o încetinire a ritmului de dezvoltare, dar și țări precum Guyana unde atenția acordată oamenilor apți pentru câmpul muncii își spune pregnant cuvântul, ori Paraguay unde, dintre cele două componente, atenția cade asupra combaterii criminalității și a deținerii unui sistemul judiciar eficient.

# Capitolul 4 - Analiza cluster

## 4.1. Alegerea metodei de analiza a datelor

Utilizarea metodei de clusterizare în analiza datelor poate aduce multiple beneficii în înțelegerea structurii și a caracteristicilor țărilor din America Latină. Prin aplicarea ei, se pot evidenția grupuri omogene de țări, astfel analiza devenind mai accesibilă și interpretabilă, permițând investigarea asemănărilor și diferențelor semnificative dintre țări. Spre exemplu, poate scoate în evidență țări cu profiluri de dezvoltare similară sau poate evidenția diferențele dintre țările cu niveluri ridicate și scăzute de corupție, stabilitate politică sau rate de criminalitate.

Această metodă poate servi și ca o modalitate eficientă de segmentare a țărilor, contribuind la o înțelegere mai detaliată a specificităților fiecărui grup și la adaptarea strategiilor de dezvoltare sau analiză în funcție de aceste caracteristici comune.

Identificarea grupurilor omogene în date reprezintă un proces crucial în analiza datelor, evidențiind similitudini și diferențe semnificative între observații. Această metodă implică separarea datelor în categorii ușor de interpretat, facilitând înțelegerea contextului analizei. Tehnicile de clusterizare relevă modele și structuri ascunse în date, oferind o perspectivă asupra relațiilor complexe dintre variabile. Vizualizarea rezultatelor prin intermediul graficelor sau hărților îmbunătățește accesibilitatea interpretării datelor și facilitează identificarea distribuției grupurilor în setul de date.

Utilizarea clusterizării reprezintă o etapă crucială în pregătirea analizelor ulterioare, inclusiv modele predictive sau analize de regresie. Simplificarea analizei prin reducerea dimensiunii setului de date contribuie la eficiența procesului. În domeniul afacerilor, clusterizarea devine esențială pentru segmentarea piețelor sau a consumatorilor, permițând ajustarea strategiilor de marketing și personalizarea serviciilor.

Prin identificarea datelor care se disting de majoritatea observațiilor, clusterizarea devine un instrument valoros pentru detectarea potențialelor anomalii sau excepții. Eficiența în timp și resurse a algoritmilor de clusterizare facilitează analiza seturilor de date extinse într-un mod scalabil. Această metodă versatilă găsește aplicații într-o gamă variată de domenii, inclusiv știința datelor, marketing, biologie, medicină și altele.

## 4.2. Prezentarea rezultatelor

Prin analiza cluster se urmărește, în interiorul unor mulțimi de obiecte sau forme, identifi carea de clase, grupe sau clustere cu elementele cât mai asemănătoare în interiorul aceleiași clase (variabilitate minimă în interiorul claselor) și cât mai deosebite între ele dacă aceste elemente aparțin unor clase diferite (variabilitate maximă între clase). Rezultă că, analiza cluster permite examinarea similarităților și disimilarităților dintre obiectele aparținând unei anumite mulțimi, în scopul grupării acestora sub forma unor clase distincte între ele și omogene în interior. Fiecare obiect din mulțimea analizată este atribuit unei singure clase, iar mulțimea claselor este o mulțime discretă și neordonabilă. Clasele sau grupele sub forma cărora se structurează mulțimile de obiecte se numesc clustere.

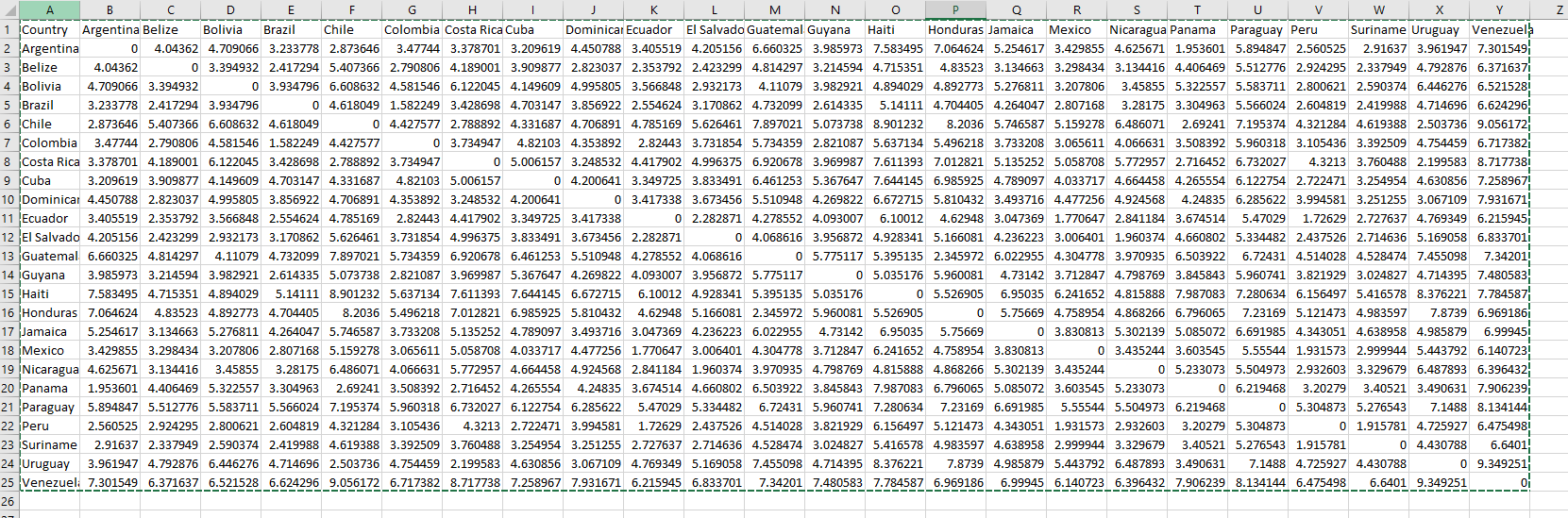
Clusterizarea ierarhică este considerată a fi un sistem de recunoaștere necontrolată, deoarece procesul clasificării pornește fără existența unor informații cu privire la numărul de clase și la apartenența formelor la aceste clase. În acest caz, clasele se construiesc pe măsura creșterii numărului de forme analizate, numărul de clase posibile determinându-se la finalul procesului de recunoaștere. Unii algoritmi de clasificare necontrolată, cum ar fi algoritmii de partiționare, presupun fixarea apriorică a numărului de clase în care vor fi împărțite obiectele analizate. Acest lucru nu înseamnă că în mod real este cunoscut numărul de clase, ci doar se face o presupunere cu privire la acest număr. Sistemele de recunoaștere necontrolată a formelor utilizează principii, metode, proceduri și tehnici, cunoscute în literatura de specialitate sub denumirea de tehnici de clasificare, clasificare nesupervizată sau analiză cluster.

Analiza cluster presupune fixarea formelor sau obiectelor în clustere sau grupe în mod progresiv, fără cunoașterea apriorică a numărului de clase și cu respectarea a două criterii fundamentale: a) Fiecare clasă să fie cât mai omogenă, adică să conțină obiecte sau forme cât mai similare în raport cu caracteristicile luate în considerare pentru clasificarea obiectelor; b) Fiecare clasă trebuie să conțină obiecte clasificate care să difere cât mai mult, din punct de vedere al caracteristicilor de clasificare, de obiectele clasificate în oricare din celelalte clase. În funcție de caracteristicile procedurilor utilizate, de ipotezele inițiale și de natura rezultatelor obținute, metodele de clusterizare ierarhică se împart în clusterizare prin:

• Agregare

• Divizare.

Proceduri specifice cunoscute în acest caz, sunt: metoda agregării simple, metoda agregării complexe, metoda agregării medii, metoda lui Ward. În cazul analizei unor cantități mari de date, caracterizate printr-un grad ridicat de eterogenitate, sistemele de recunoaștere necontrolată se utilizează mai mult în scopuri de sistematizare, grupare și sintetizare informațională. Întrucât aceste tehnici, care se bazează pe utilizarea conceptului de distanță, sunt utile și eficiente în activitatea de analiză preliminară a datelor, permit organizarea mai eficientă a datelor eterogene, precum și regăsirea și interpretarea mai ușoară și mai consistentă a informațiilor în cadrul unor date astfel structurate.In cazu nostru , distanta dintre componente / clustere este calculata utilizand tehnica euclidiană.



### Fig. 13: Matricea distanțelor , date standardizate

Matricea de distanțe reprezentată mai sus indica măsurile de similaritate sau distanța dintre țări în funcție de variabilele utilizate în analiză. Aceasta este o reprezentare numerică a cât de asemănătoare sau diferite sunt țările între ele, pe baza caracteristicilor măsurate. În general, în analiza clusterelor, se folosește o măsură de distanță, iar valorile mai mici indică o similaritate mai mare între obiecte.

În cazul nostru, valorile din matrice indică cât de aproape sau departe sunt țările unele de altele. Cu cât valoarea este mai mică, cu atât țările sunt considerate mai asemănătoare. De exemplu, dacă te uiți la Argentina și Belize, vei observa că distanța dintre ele este relativ mică (4.04), indicând că aceste țări sunt mai asemănătoare în ceea ce privește variabilele analizate comparativ cu alte țări din listă.

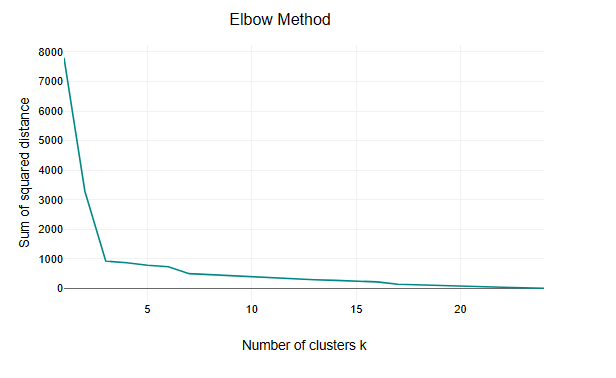
Pentru interpretare, se va folosi această matrice pentru a identifica grupuri de țări care sunt mai asemănătoare între ele. Clusterele pot fi formate în funcție de distanțele minime între țări.

La alegerea metodei de aplicare a algoritmului de clusterizare ierarhică am ales sa folosim metoda de calculare a distanței ward și metrica city-block, iar distanța dintre componente a fost calculata folosind metoda euclidiana .

Distanța Ward dintre două clustere măsoară variabilitatea intracluster cumulată, indusă de comasarea a două clustere, la nivelul configurației cluster rezultate. Prin comasarea a două clustere se urmărește obținerea unei omogenități maxime la nivelul tuturor clusterelor care aparțin unei configurații date a obiectelor pe clustere. Rezultă că distanța Ward este singura care ia în calcul minimizarea variabilității intracluster sau, cu alte cuvinte, maximizarea variabilității intercluster, adică a gradului de omogenitate a clusterelor. Trebuie precizat că, gradul de omogenitate a unui cluster se maximizează prin minimizarea sumei totale a pătratelor abaterilor intracluster.

#### 4.2.1. Metoda Elbow

Metoda Elbow este o tehnică utilizată pentru a determina numărul optim de clustere dintr-un set de date pentru algoritmi de clustering.

**

### Fig. 14: Valoarea funcției Elbow.

Se poate observa din figura ca punctul de cot este situat la o valoare aproximativ egala cu 2,5 de aceea am hotarat sa efectuam teste suplimentare.

#### 4.2.2. Metoda Silhouette

Pentru fiecare structură de clustere putem calcula silueta pentru a evidentia cat de bine este separat un cluster de vecinii săi, cât de apropiat este un element dintr-un cluster de elemente aceluiași cluster, mai exact distanța intracluster sau cat de departe este de celelalte clustere, adică distanța intercluster.

La nivel global silueta are valoarea de 0.35 ceea ce indica o structura moderata, nici slabă, dar nici puternică.

Dacă valoarea siluetei este apropiata de 1, structura de clustere este bine clasificată.

Dacă valoarea siluetei tinde catre 0 înseamnă ca obiectul este similar clasificat cu clusterele învecinate.

Dacă S(i) este apropiat de -1 înseamnă ca structura este clasificată eronat .

Am rulat funcția de testare pe intervalul [2,5] și avem următoarele rezultate:

For n\_clusters = 2, the average silhouette score is : 0.90.

For n\_clusters = 3, the average silhouette score is : 0.88.

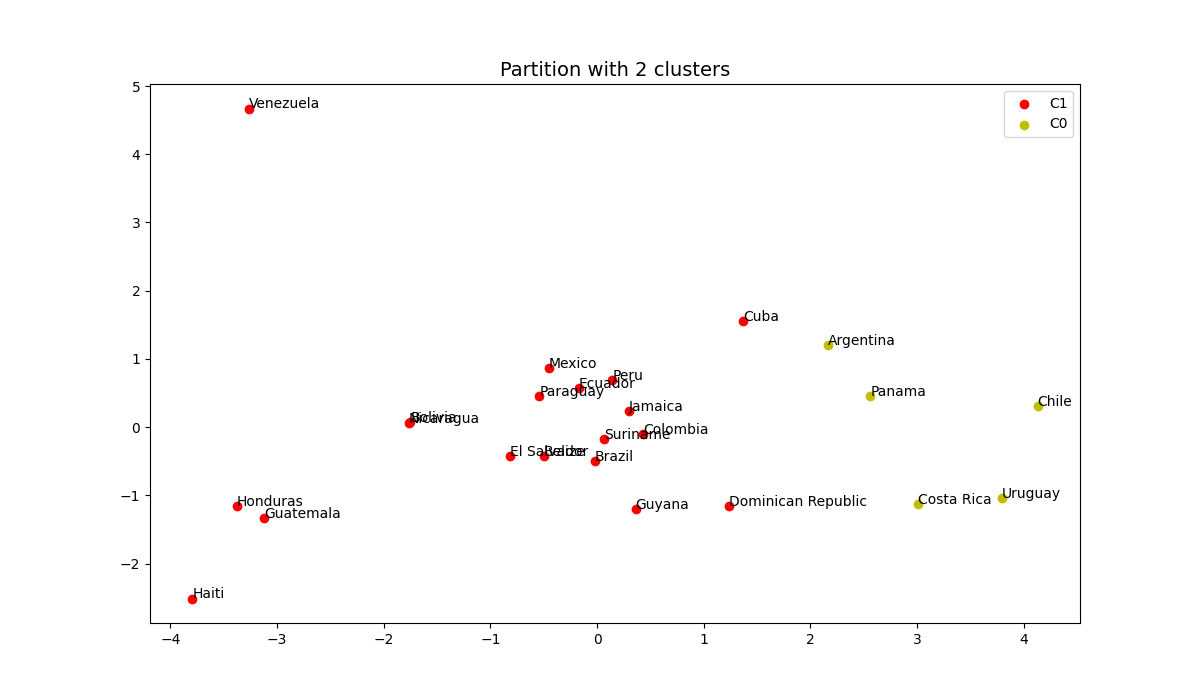
For n\_clusters = 4, the average silhouette score is : 0.38.

For n\_clusters = 5, the average silhouette score is : 0.25.

Concluzia finală este de a studia setul de date pe 2 structuri de tip cluster împărțite în 2 respectiv 3 clustere.

Am creat un grafic dispersiv care ilustrează distribuția datelor în cadrul clusterelor, evidențiind gruparea ierarhică pe datele standardizate.

## 4.3. Structurarea datelor pe 2 clustere.



### Fig. 15: Partiția cu 2 clustere.

Acest output reprezintă partajarea optimă a țărilor în clustere pe baza indicatorilor utilizați în analiză. Fiecare țară a fost atribuită unui anumit cluster, notat cu litere precum C0, C1. Interpretarea exactă a outputului depinde de natura și semnificația fiecărui indicator. Vom încerca să oferim o interpretare generală pentru fiecare cluster pe baza unor presupuneri generale despre acești indicatori.

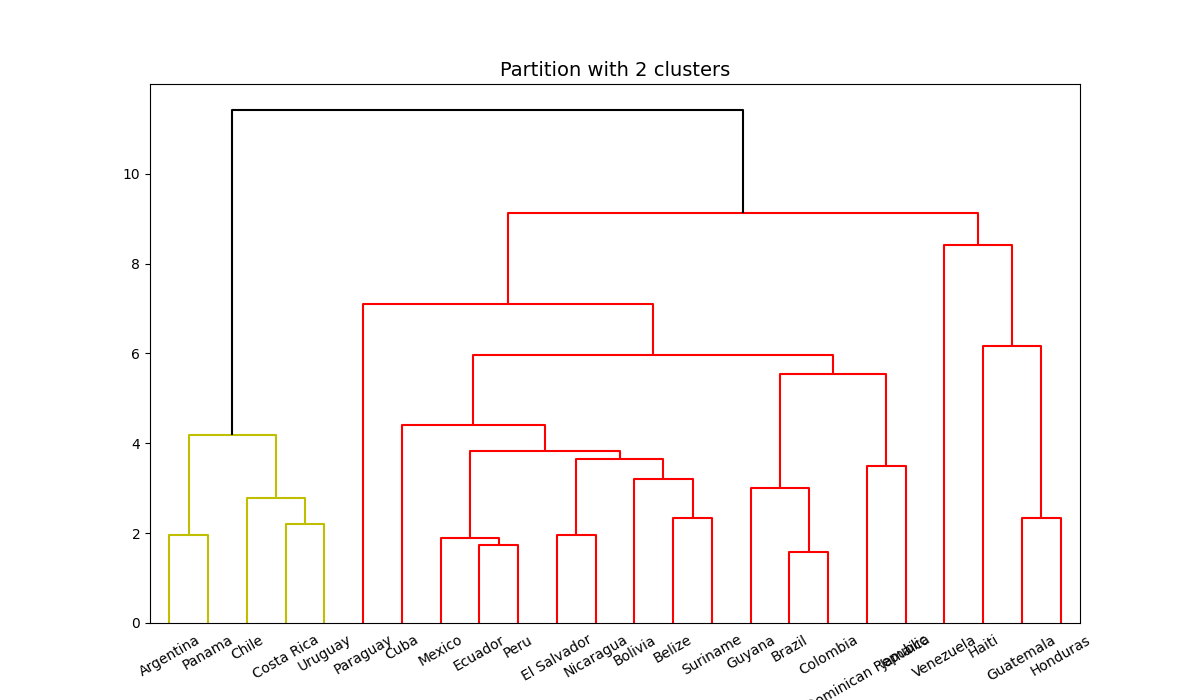
Țările ce au fost incluse de către algoritm în **Cluster 0** sunt: Argentina, Chile, Costa Rica, Panamá, Uruguay.

Aceste țări au fost grupate împreună, sugerând că acestea ar putea avea valori similare sau caracteristici comune în ceea ce privește indicatorii analizați. În funcție de natura indicatorilor folosită de către noi în acest studiu, aceasta ar putea sugera o stare de dezvoltare mai înaltă, stabilitate politică și alte caracteristici pozitive.

În cadrul acestui cluster distinct, observăm o aglomerare de țări care se remarcă printr-un nivel avansat de dezvoltare socio-economică. Indicii socio-economici relevanți reflectă o economie stabilă, o infrastructură bine dezvoltată, un nivel înalt al calității vieții și o forță de muncă instruită. Aceste țări prezintă inovație, acces la tehnologii moderne și servicii de înaltă calitate, marcându-se drept leaderi în cadrul economiei globale.

Țările ce au fost include de către algoritm în **Cluster 1**: Belize, Bolivia, Brazil, Colombia, Cuba, Dominican Republic, Ecuador, El Salvador, Guatemala, Guyana, Haití, Honduras, Jamaica, México, Nicaragua, Paraguay, Perú, Suriname, Venezuela.

Aceste țări au fost grupate împreună și sunt probabil caracterizate de alte seturi de valori ale indicatorilor. Aceasta ar putea indica o varietate mai mare în starea lor de dezvoltare sau în alte caracteristici socio-economice.

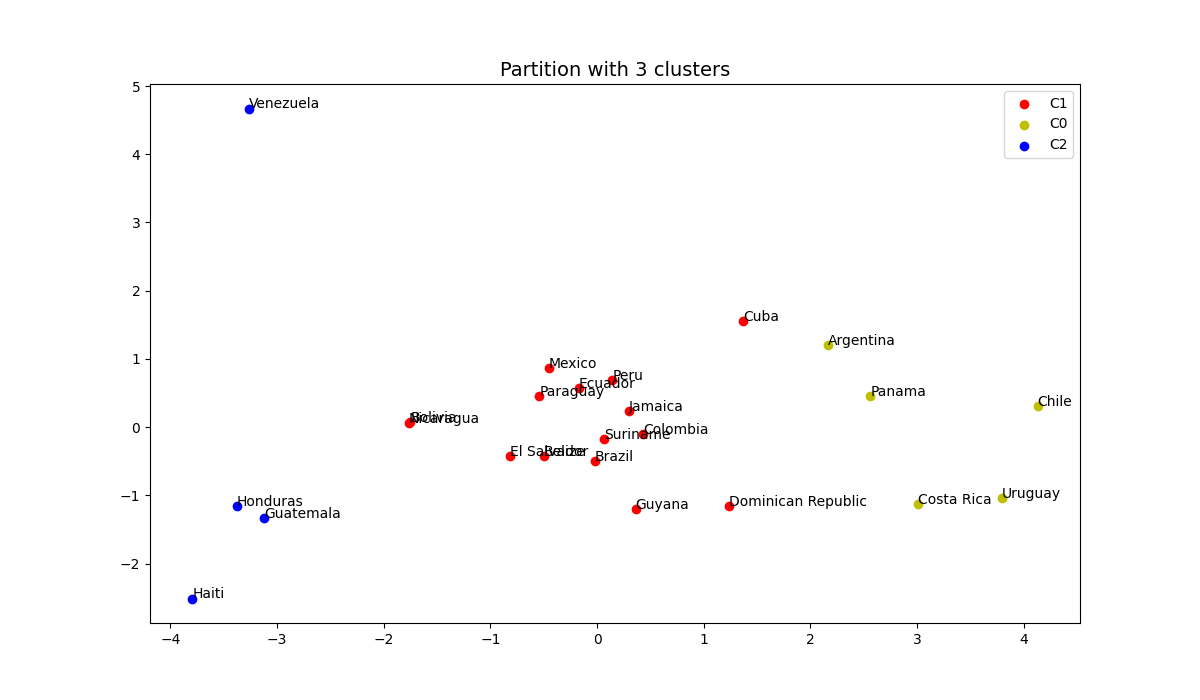


### Fig. 16: Dendrograma cu 2 clustere.

Prin evaluarea înălțimilor de conectare, observăm etapele în care obiectele se unesc, evidențiind două grupuri semnificative. Pragurile de tăiere conturează formarea acestor două clustere, oferind o perspectivă succintă asupra procesului de clusterizare în două grupuri distincte.

## 

## 4.4. Structurarea datelor pe 3 clustere.



### Fig. 17: Partiția cu 3 clustere.

Țările ce au fost incluse de către algoritm în **Cluster 0** sunt: Argentina, Panama, Costa Rica, Uruguay, Chile.

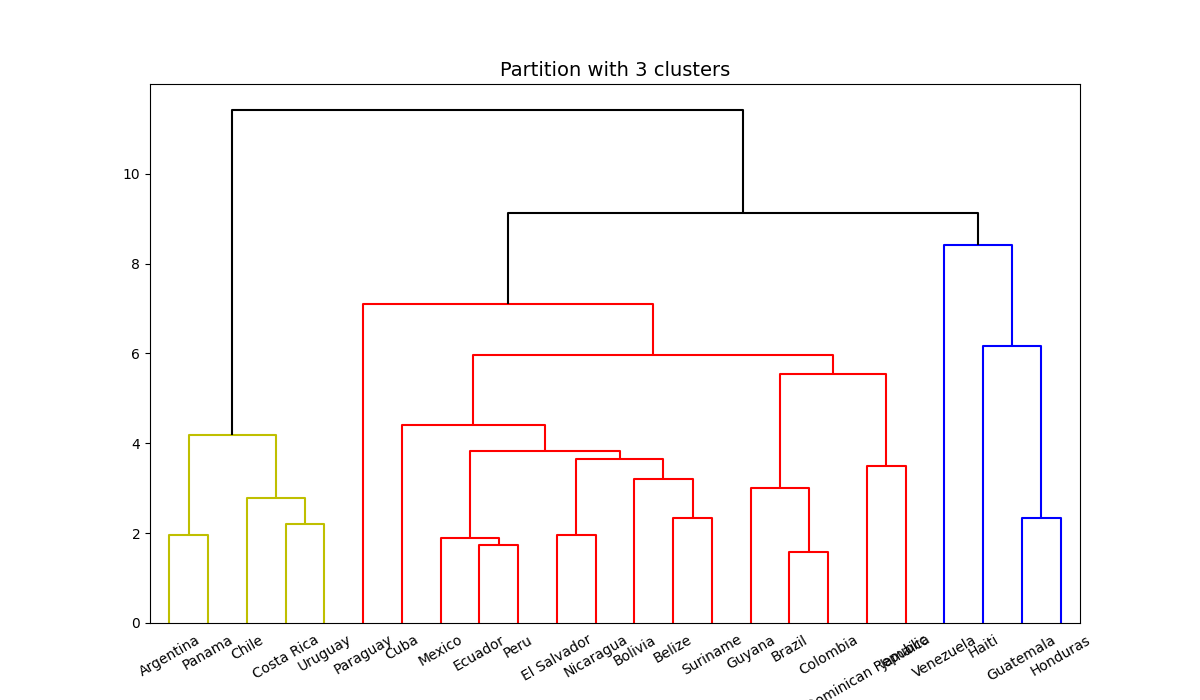
În cadrul acestui cluster distinct, observăm o aglomerare de țări care se remarcă printr-un nivel avansat de dezvoltare socio-economică. Indicii socio-economici relevanți reflectă o economie stabilă, o infrastructură bine dezvoltată, un nivel înalt al calității vieții și o forță de muncă instruită. Aceste țări prezintă inovație, acces la tehnologii moderne și servicii de înaltă calitate, marcându-se drept leaderi în cadrul economiei globale.

Țările ce au fost incluse de către algoritm în **Cluster 1** sunt: Belize, Bolivia, Brazil, Colombia, Cuba, Dominican Republic, Ecuador, El Salvador, Guyana, Haití, Jamaica, México, Nicaragua, Paraguay, Peru, Suriname.

În acest cluster, identificăm o grupare de țări care se află într-un stadiu intermediar al dezvoltării socio-economice. Indicii socio-economici relevanți indică progrese semnificative, cu accent pe industrializare, creșterea economică și extinderea infrastructurii. Aceste țări manifestă un potențial de creștere, dar întâmpină în continuare provocări în ceea ce privește echitatea socială, accesul la servicii și dezvoltarea regională.

Țările ce au fost incluse de către algoritm în **Cluster 2** sunt: Venezuela, Honduras, Guatemala, Haiti.

În acest cluster distinct, remarcăm țări caracterizate de niveluri reduse de dezvoltare socio-economică. Indicii socio-economici reflectă o vulnerabilitate crescută la șocuri economice, infrastructuri limitate și acces restricționat la servicii esențiale. Aceste țări se confruntă cu provocări semnificative în ceea ce privește sărăcia, accesul la educație și serviciile de sănătate, și necesită sprijin suplimentar pentru a depăși aceste deficiențe și a atinge un stadiu mai înalt de dezvoltare.



### Fig. 18: Dendrograma cu 3 clustere.

Această dendrogramă reprezintă o vizualizare grafică a procesului de grupare și formare a trei clustere distincte în setul nostru de date. Prin analiza înălțimilor de conectare și structurii ramurilor, putem observa cum obiectele individuale sau grupurile mici se unesc succesiv pentru a forma clustere mai mari. În această configurare, am identificat trei niveluri superioare de clusterizare, iar ramurile dendrogramei indică similaritățile și relațiile între aceste grupuri.

# Concluzii

In final, America Latina se dovedește a fi o regiune greu de înțeles pe deplin, cu dinamici sociale complicate si probleme diverse. Disparitățile sociale sunt adânc înrădăcinate, iar sărăcia și inegalitatea economică rămân problemele majore. Unele țări au avut succes în dezvoltarea economică și reducerea sărăciei, în timp ce altele continuă să lupte cu problemele legate de accesul la educație, servicii medicale și locuințe adecvate.

Rezultatele analizei componentelor principale au evidențiat asocieri semnificative între variabilele considerate. ACP a redus dimensionalitatea setului de date inițial, identificând noțiuni sau dimensiuni latente care explică cea mai mare parte a variației observate. De exemplu, factori precum indicatorii economici, calitatea vieții, și factorii politici au fost grupați în mod corespunzător, dezvăluind cea mai mare problema a acestei regiuni si anume lipsa unei guvernări eficiente.

De asemenea, ACP a permis identificarea țărilor care au avut comportamente sau profiluri similare pe aceste dimensiuni latente. Clasificarea țărilor în funcție de aceste factori a oferit o perspectivă mai clară asupra diversității și similarităților în cadrul regiunii, evidențiind grupuri de țări cu trăsături comune și, în același timp, subliniind diferențele semnificative între altele. Argentina, Chile și Panama reprezintă țările în care stabilitatea politică contribuie semnificativ la un standard de viată mai bun, chiar dacă în continuare Argentina are probleme mari cu rata inflației. La polul opus, țări precum Haiti, Guatemala și Honduras sunt cele mai subdezvoltate, in care lipsa accesului la educație, servicii medicale și oportunități economice echitabile creează un cerc vicios al sărăciei.

Analiza clusterelor a consolidat rezultatele ACP, contribuind la identificarea grupurilor omogene de țări în funcție de caracteristicile lor. Această abordare a dezvăluit modele complexe de asociere între țări, evidențiind aspecte precum dezvoltarea economică, stabilitatea politică, și indicatorii sociali.

Același grup de tari evidențiat mai sus, care se remarca pozitiv, la care se adaugă Uruguay, Costa Rica se afla in cluster 0, in timp ce restul in cluster 1, in care identificăm o grupare de țări care se află într-un stadiu intermediar al dezvoltării socio-economice.

În final, combinația de ACP și analiza clusterelor a furnizat o viziune comprehensivă asupra diversității și complexității țărilor din America Latină. Cu toate aceste provocări, America Latină are și oportunități considerabile. De exemplu, potențialul său economic, resursele naturale și forța de muncă calificată pot servi drept catalizatori pentru dezvoltare și inovație. Prin abordarea problemelor cu responsabilitate și prin investiții în educație, infrastructură și guvernare eficientă, regiunea poate aborda treptat aceste provocări și să-și realizeze întregul potențial.

# Bibliografie

1. [https://publications.iadb.org/en/2023-latin-american-and-caribbean-macroeconomic-report-preparing-macroeconomic-terrain-renewed](1.%09https:/publications.iadb.org/en/2023-latin-american-and-caribbean-macroeconomic-report-preparing-macroeconomic-terrain-renewed)
2. [https://hdr.undp.org/data-center/human-development-index#/indicies/HDI](2.%09https:/hdr.undp.org/data-center/human-development-index#/indicies/HDI)
3. [https://www.statista.com/statistics/947781/homicide-rates-latin-america-caribbean-country/](3.%09https:/www.statista.com/statistics/947781/homicide-rates-latin-america-caribbean-country)
4. [https://www.worldbank.org/en/publication/worldwide-governance-indicators/interactive-data-access](4.%09https:/www.worldbank.org/en/publication/worldwide-governance-indicators/interactive-data-access)
5. [https://worldpopulationreview.com/country-rankings/drug-use-by-country](5.%09https:/worldpopulationreview.com/country-rankings/drug-use-by-country)
6. [https://data.worldbank.org/indicator/SL.UEM.TOTL.ZS](6.%09https:/data.worldbank.org/indicator/SL.UEM.TOTL.ZS)
7. [https://www.theglobaleconomy.com/rankings/Inflation/](7.%09https:/www.theglobaleconomy.com/rankings/Inflation)