**ACADEMIA DE STUDII ECONOMICE BUCUREŞTI**

**FACULTATEA DE CIBERNETICĂ STATISTICĂ ȘI INFORMATICĂ ECONOMICĂ**

A close up of a logo

Description automatically generated

**PROIECT**

**DEZVOLTARE SOFTWARE PENTRU ANALIZA DATELOR**

Analiza gradului de educație în cadrul unor țări europene

Realizat de:

|  |  |
| --- | --- |
| **Nume prenume** | **Gugiuman Irina** |
| **Grupa** | **1084** |

București, 2023

Cuprins

[1. Prezentare generală 3](#_Toc125821504)

[2. Analiza în componente principale 3](#_Toc125821505)

[3. Analiza cluster 6](#_Toc125821506)

[Bibliografie 9](#_Toc125821507)

# Prezentare generală

Educația la nivelul unei populații este influențată de o serie mare de factori, care pot fi atât economici, cât și politici, geografici, personali sau medicali.

Pentru a analiza gradul de educație în cadrul a 34 de țări europene pentru anul 2019, s-au luat în considerare următoarele elemente, exprimate în procente:

* Cheltuieli de cercetare și dezvoltare
* Resurse umane în știință și tehnologie
* Gospodării - nivelul de acces la internet
* Persoane care au abilități de informare de bază
* Persoane care au abilități de informare peste cele de bază
* Persoane care au abilități de comunicare de bază
* Persoane care au abilități de comunicare peste cele de bază
* Persoane care au abilități de bază de rezolvare a problemelor
* Persoane care au abilități superioare de rezolvare a problemelor
* Persoane care au abilități de bază de software
* Persoane care au abilități de software peste cele de bază
* Persoane care au competențe digitale generale scăzute
* Persoane care au abilități digitale generale de bază
* Persoane care au abilități digitale generale peste cele de bază
* Cheltuieli naționale pentru protecția mediului
* Copilărie timpurie până la învățământul terțiar
* Licență sau nivel echivalent
* Master sau nivel echivalent
* Rata totală a șomajului
* Moarte din cauza sinuciderii

# Analiza în componente principale

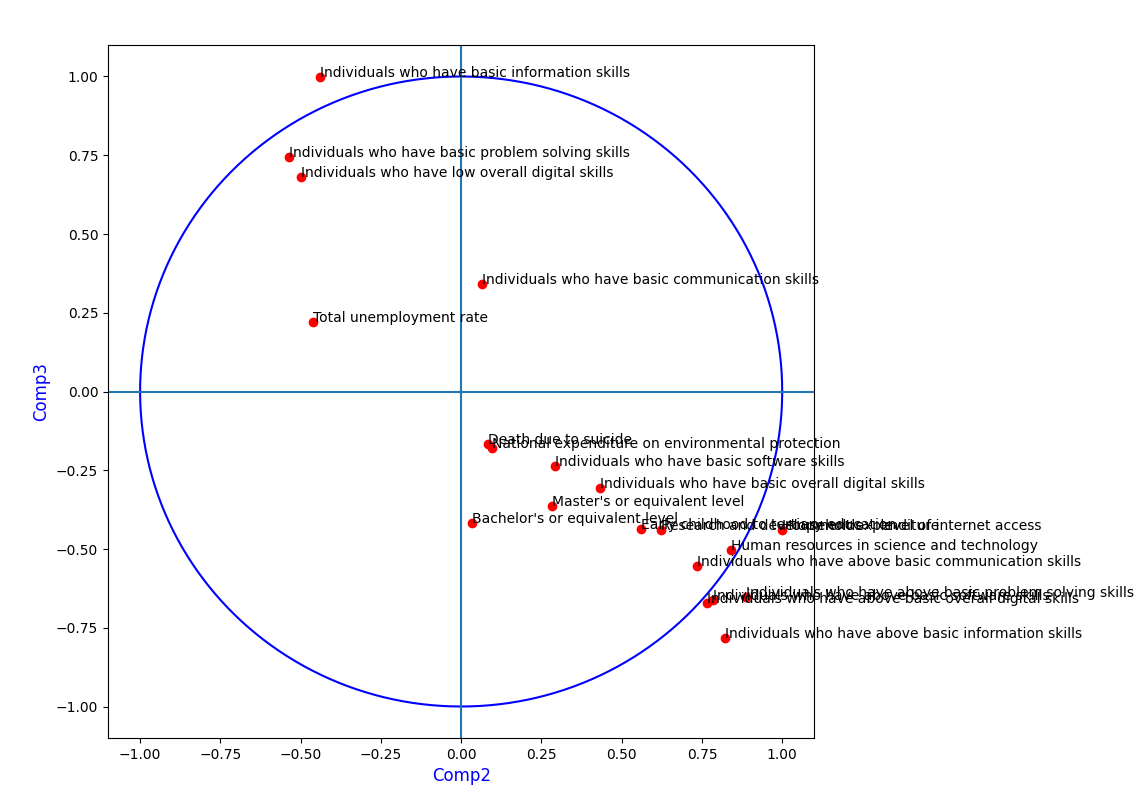
O primă imagine a indicatorilor analizați este dată de corelațiile dintre aceștia. Dacă între variabile este o legătură puternică, atunci și corelația dintre variabile va una puternică, astfel creându-se redundanță informațională în setul de date. Variabilele cu astfel de trăsături stau la baza componentelor principale.

Din corelograma se poate observa că există o legătură puternică pozitivă între resursele umane în știință și tehnologie și nivelul de acces la internet, indivizii care au competențe informaționale peste cele de bază, indivizii care au competențe de comunicare peste ce le de bază, indivizii care au abilități de rezolvarea problemelor peste cele de bază, indivizii care au abilități software peste cele de bază, indivizii care au abilități digitale generale peste cele de bază, dar și cu procentul copiilor care merg la școală până la învățământul terțiar. Există o legătură pozitivă dar mai slabă și între cheltuielile de cercetare și dezvoltare și variabilele menționate mai sus. Se pot observa și legături inverse între cheltuielile de cercetare și dezvoltare și indivizii care au competențe digitate generale scăzute, între indivizii care au abilități de bază de rezolvare a problemelor și indivizii care au competențe digitate generale Chart, treemap chart

Description automatically generatedscăzute, dar și cu rata șomajului. O legătură inversă se poate observa și între rata șomajului și nivelul de acces la internet, resurse umane în știință și tehnologie și indivizii care au abilități Line chart

Description automatically generated with low confidenceinformaționale peste cele de bază.

Conform criteriilor Cattel și Kaiser, sunt semnificative doar primele cinci componente, întrucât au o varianță mai mare decât 1. În ceea ce privește variabilitatea datelor, prima componentă acoperă 46, 61%, a doua componentă acoperă 11,69%, a treia componentă, 8,72%, a patra componentă, 7,12%, iar a cincea, 5,59%.

Cu ajutorul cercului de corelații dintre componentele 2 și 3, se poate observa că variabilele pozitiv corelate sunt grupate împreună, variabilele invers corelate sunt în partea opusă a diagramei, iar distanța dintre variabile și origine măsoară calitatea variabilelor.

Timeline

Description automatically generatedComunalitățile reprezintă proporția varianței fiecărei variabile care poate fi explicată prin componentele principale. Se pot observa variabilele care au legături mai puțin semnificative, precum licența, masterul sau moarte din cauza sinuciderii.

# Analiza cluster

Algoritmii de clusterizare sau clasificare nesupervizată sunt utilizați pentru a determina grupările naturale ale datelor sau pentru a furniza o împărțire convenabilă a datelor în grupuri. Datele supuse analizei sunt valori ale relațiilor dintre indivizii și variabilele aflate în studiu, luate câte două distanțe sau disimilarități. Înregistrarea acestor valori ia forma matricelor de distanță. Așadar, valorile supuse analizei reprezintă deja rezultatul unor calcule care au ca rezultat măsuri de disimilaritate dintre obiecte (instanțe sau variabile).

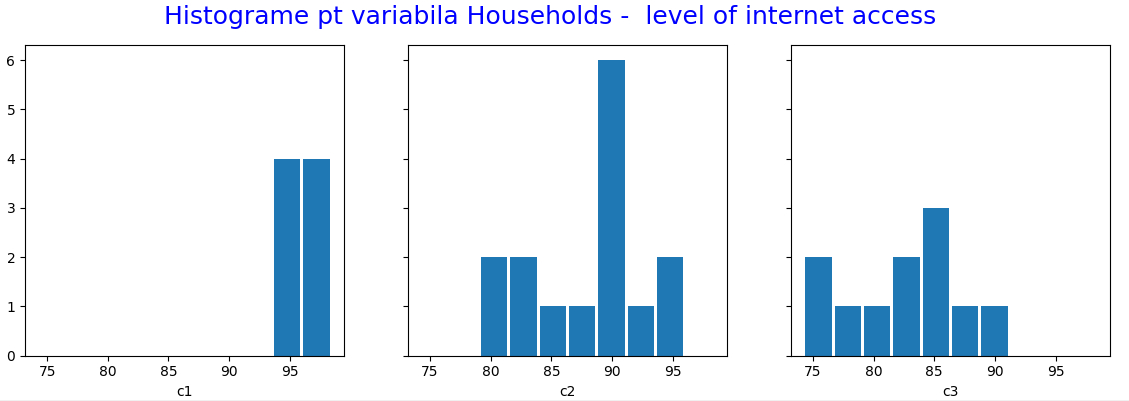
Chart, histogram

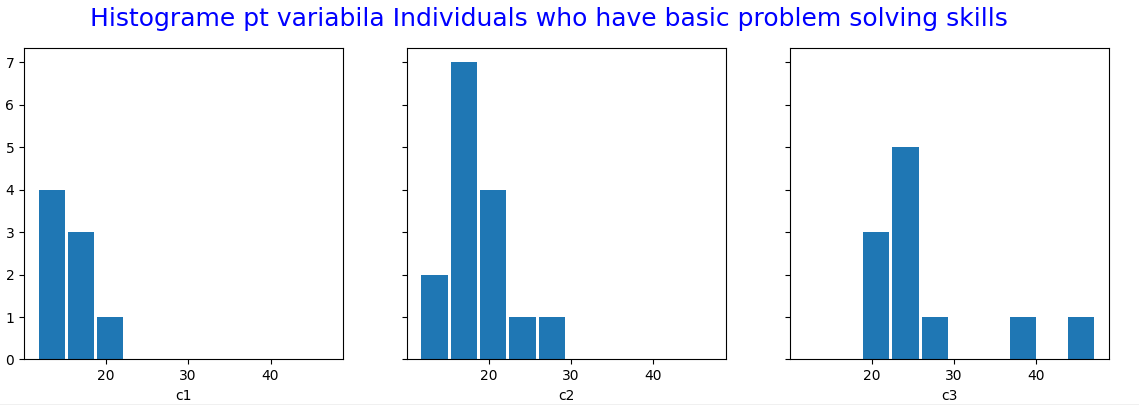
Description automatically generatedGrupele formate prin clasificare conțin indivizi asemănători intre ei, cu disimilarități mici. Privită din această perspectivă, a formării de grupe omogene, analiza de cluster este o metodă de sinteză informațională, așa cum este și analiza în componente principale, doar că se referă în principal la instanțe și nu la variabile.

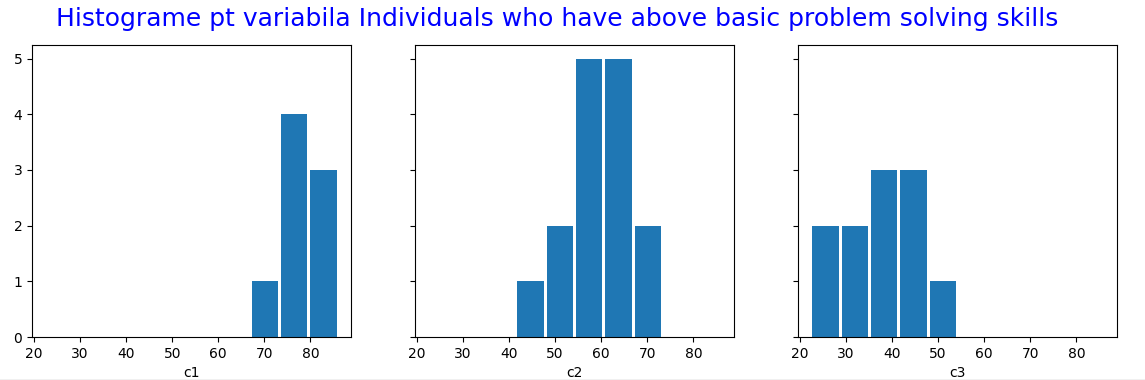
În figură este reprezentată partiția optimală a țărilor, formată din trei clustere. Pentru analiza acestora se urmărește distribuția fiecărui indicator pentru fiecare cluster, cu scopul de a se identifica particularitățile și diferențele dintre aceștia.

Danemarca, Olanda, Finlanda, Suedia, Islanda, Norvegia, Elveția și Regatul Unit fac partea din primul cluster, având procente ridicate pentru resurse umane în știință și tehnologie, nivelul de acces la internet al gospodăriilor, pentru indivizii care au competențe peste cele de bază și pentru copii care termină educația secundară.

În cel de-al doilea cluster sunt prezente în mare parte țări din Europa centrală și de vest, precum Germania, Austria, Belgia, Franța, Spania, Portugalia, Luxemburg, pentru care procentul de resurse umane este mai scăzut, apropiindu-se de media generală, accesul la internet rămâne ridicat, iar procentele pentru indivizii care au abilități peste cele de bază pentru informație, comunicare rezolvarea problemelor este mai ridicat decât pentru software și competențe digitate, dar mai scăzute față de primul cluster, cum este și cazul copiilor care termină educația secundată.

Cel de-al treilea cluster este format din țări majoritar din estul Europei, cum ar fi România, Bulgaria, Grecia, Serbia sau Muntenegru. În acest cluster, procentele de resurse umane și al caselor cu acces la internet sunt cele mai scăzute comparativ cu celelalte două clustere. Procentul indivizilor care au competențe peste medie este, de asemenea, mult mai scăzut față de celelalte clustere, iar în ceea ce privește competențele digitale generale, se poate observa un procentaj mai mare al indivizilor care au competențe scăzute. Este de remarcat că procentul sinuciderilor este în medie mai mic.



Prin analiza histogramelor, se poate observa că în cadrul primului cluster sunt prezente multe valori mari pentru variabile care ar putea fi reprezentative pentru un nivel mai ridicat de educație, ca de exemplu accesul la internet și abilități peste cele de bază. Pentru aceleași variabile, în al doilea cluster sunt prezente predominant valori medii, iar în al treilea valori sub medie.

Se poate astfel concluziona că, în clusterele 1 și 2 sunt incluse țări al căror nivel de educație este mai ridicat, față de cele care aparțin clusterului 3.

# Bibliografie

1. Cursuri Dezvoltare Software pentru analiza datelor, prof. univ. dr. Furtună Titus Felix
2. <https://ec.europa.eu/eurostat/databrowser/view/TSC00001/default/table>
3. <https://ec.europa.eu/eurostat/databrowser/view/TSC00025/default/table>
4. <https://ec.europa.eu/eurostat/databrowser/view/ISOC_CI_IN_H$DEFAULTVIEW/default/table>
5. <https://ec.europa.eu/eurostat/databrowser/view/ISOC_SK_DSKL_I__custom_4621414/default/table>
6. <https://ec.europa.eu/eurostat/databrowser/view/TEN00135/default/table>
7. <https://ec.europa.eu/eurostat/databrowser/view/EDUC_UOE_ENRA04__custom_4621523/default/table>
8. <https://ec.europa.eu/eurostat/databrowser/view/TPS00203/default/table>
9. <https://ec.europa.eu/eurostat/databrowser/view/TPS00122/default/table>