

În fișierul *CastigulSalNet.csv* se află date privind câștigul salarial net la nivel de județe, pe ani, în perioada 2016-2021. În fișierul *PopulatieJudete.csv* se află populația pe județe și gruparea județelor pe regiuni.

A. Cerințe

1. Calculul mediei câștigurilor salariale nete pe perioada 2016-2021. Datele vor fi salvate în fișierul *Cerinta1.csv*, pe trei coloane: indicativul de județ, numele județului și media câștigurilor salariale nete, în ordinea descrescătoare a mediei. (1 punct)

Exemplu:

```
Indicativ,Judet,Media
b,Bucuresti,3779.6666666666665
cj,Cluj,3180.3333333333335
```

...

Criteriul de acordare a punctajului: vizualizarea fișierului output.

2. Să se determine câștigurile salariale medii la nivel de regiune. Se va calcula media ponderată, folosind ca pondere populația. Rezultatul va fi salvat în fișierul *Cerinta2.csv*. Se va salva codul regiunii și valorile medii pe fiecare an. (2 puncte)

Exemplu:

```
Regiune,2016,2017,2018,2019,2020,2021
b,2839.72,3194.16,3512.57,3821.96,4120.83,4436.6
```

...

Criteriul de acordare a punctajului: vizualizarea fișierului output

B. Să se efectueze analiza de clusteri prin metoda Ward, pentru câștigurile salariale și să se furnizeze următoarele rezultate:

1. Dendrograma partiției cu k clusteri. k se va inițializa prin cod. (1 punct)

Criteriul de acordare a punctajului: vizualizarea graficului

2. Componenta partiției formate din k clusteri. Pentru fiecare instanță se va specifica clusterul din care face parte și scorul Silhouette al instanței. Partiția va fi salvată în fișierul *partitie.csv*, pe 4 coloane: codul de țară, denumirea țării, clusterul din care face parte și scorul Silhouette. (2 puncte)

Criteriul de acordare a punctajului: vizualizarea fișierului output.

3. Trasarea plotului partiției din k clusteri în axe principale (primele 2 componente principale) (2 puncte)

Criteriu de acordare a punctajului: vizualizarea graficului

C. În urma efectuării unei analize canonice cu cinci variabile în prima grupă, (X_1, X_2, X_3, X_4, X_5) și patru variabile în grupa a doua, (Y_1, Y_2, Y_3, Y_4) s-au obținut următoarele corelații canonice (valorile proprii): 0.9, 0.3, 0.2, 0.1. În fișierul *rxz.csv* sunt salvate corelațiile dintre variabilele primei grupe (X) și variabilele canonice ale primei grupe.

Să se afișeze la consolă cantitatea de varianță comună pentru cele două seturi de date (redundanța) explicată de al treilea cuplu de variabile canonice.

Criteriul de acordare a punctajului: urmărirea consolei și a codului sursă

1 punct din oficiu

Observație

Punctajele se acordă pe baza rezultatelor obținute!