

În fișierul *Indicatori.csv* sunt prezentați indicatori ai activității firmelor la nivel de localitate pentru anul 2008. Indicatori: NR\_FIRME - numărul de firme, NSAL - numărul de salariați, CFA - cifra de afaceri, PROFITN - profitul net, PIERDEREN - pierderea netă. Variabila SIRUTA reprezintă codul Siruta al localității. În fișierul *PopulatieLocalitati.csv* se află codurile Siruta ale localităților, indicativele de județ și populația la nivel de localități.

### A. Cerințe

1. Să se determine localitățile cu cifra de afaceri mai mare decât valoarea medie a cifrei de afaceri pe țară. și să se salveze în fișierul *Cerinta1.csv*. Se vor salva codul Siruta și valorile indicatorilor, în ordinea descrescătoare a cifrei de afaceri. (2 puncte)

Exemplu:

```
SIRUTA,NR_FIRME,NSAL,CFA,PROFITN,PIERDEREN
179132,111826,1177345,330641496321.0,22997762362.0,16954653951.0
54975,17673,109526,23162554203.0,1341388982.0,818218913.0
...
```

**Criteriul de acordare a punctajului:** vizualizarea fișierului output și a codului sursă

2. Să se determine valorile indicatorilor raportate la populație (la 1000 locuitori), la nivel de județ și să se salveze în fișierul *Cerinta2.csv*. Valoarea pentru un indicator la 1000 locuitori se calculează astfel:  $v_{1000} = v * 1000 / p$ , unde  $v_{1000}$  este valoarea indicatorului la 1000 locuitori,  $v$  este valoarea indicatorului iar  $p$  este populația. (2 puncte)

Exemplu:

```
Judet,NR_FIRME,NSAL,CFA,PROFITN,PIERDEREN
ab,19.376,150.427,21641529.3,773726.106,1718918.399
ag,19.959,184.507,39377116.712,1585820.765,635601.084
...
```

**Criteriul de acordare a punctajului:** vizualizarea fișierului output și a codului sursă

**B.** În fișierul *LocationQ.csv* se găsesc indicatori de prezență a activității economice la nivel de județe între anii 2008 și 2021. Să se efectueze analiza de clusteri prin metoda Ward pe acest set de date (variabile sunt anii) și să se furnizeze următoarele rezultate:

1. Matricea ierarhie cu informații privind joncțiunile făcute. Pentru fiecare joncțiune se va specifica clusterii intrați în joncțiune, distanța dintre cei doi clusteri și numărul de instanțe în clusterul nou format. Matricea va fi afișată la consolă. (1 punct)

**Criteriul de acordare a punctajului:** vizualizarea outputului și a codului sursă

2. Graficul dendrogramă pentru partiția optimală. (2 puncte)

**Criteriul de acordare a punctajului:** vizualizarea graficului dendrogramă.

3. Componenta partiției optimale. Pentru fiecare instanță se determina clusterul de care aparține. Partiția se va salva în fișierul *popt.csv*. (2 puncte)

**Criteriul de acordare a punctajului:** vizualizarea fișierului output și a codului sursă

**1 punct** din oficiu

### Observație

*Punctajele se acordă pe baza rezultatelor obținute!*