Academia de Studii Economice din București

CSIE, Informatică Economică

Educația în România

Student: Daniela Monica MITROI

Profesor coordonator: Claudiu VINȚE

1. Descrierea datelor:

Datele utilizate sunt preluate din bazele de date ale Institutului Național de Statistică[[1]](#footnote-1) și fac parte din categoria statisticii sociale, grupa educație. INS pune la dispoziție date care se leagă de anumite elemente esențiale pentru sistemul de învățământ, ele referindu-se la numărul de unități școlare, populația școlară, personalul didactic, infrastructura învățământului sau numărul de absolvenți, creând astfel o privire de ansamblu asupra educației din România.

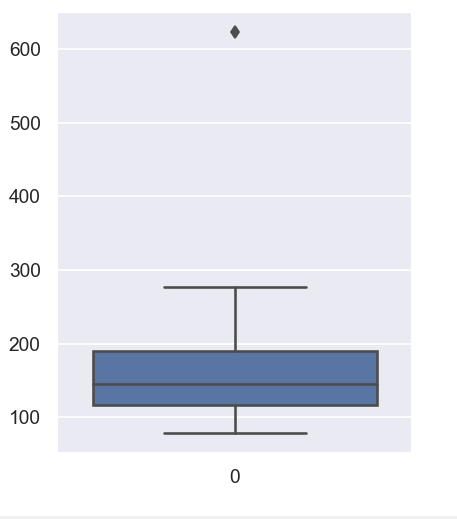
Fiecare categorie de dată a fost salvată individual în directorul denumit “DataIN”. În plus față de cele preluate de pe INS sunt datele cu privire la împărțirea geografică a României în județe, regiuni, respectiv macroregiuni care pe parcursul analizei au ajutat la agregarea datelor și la o vizualizare mai simplă a anumitor caracteristici ce țin de educația românească.

Toate datele sunt din anul 2021, cu excepția celor referitoare la numărul de absolvenți care sunt din 2020, acesta fiind unul dintre motivele pentru care analiza absolvenților a fost facută separat de-a lungul proiectului. De asemenea, toate datele sunt cantitative și absolute.

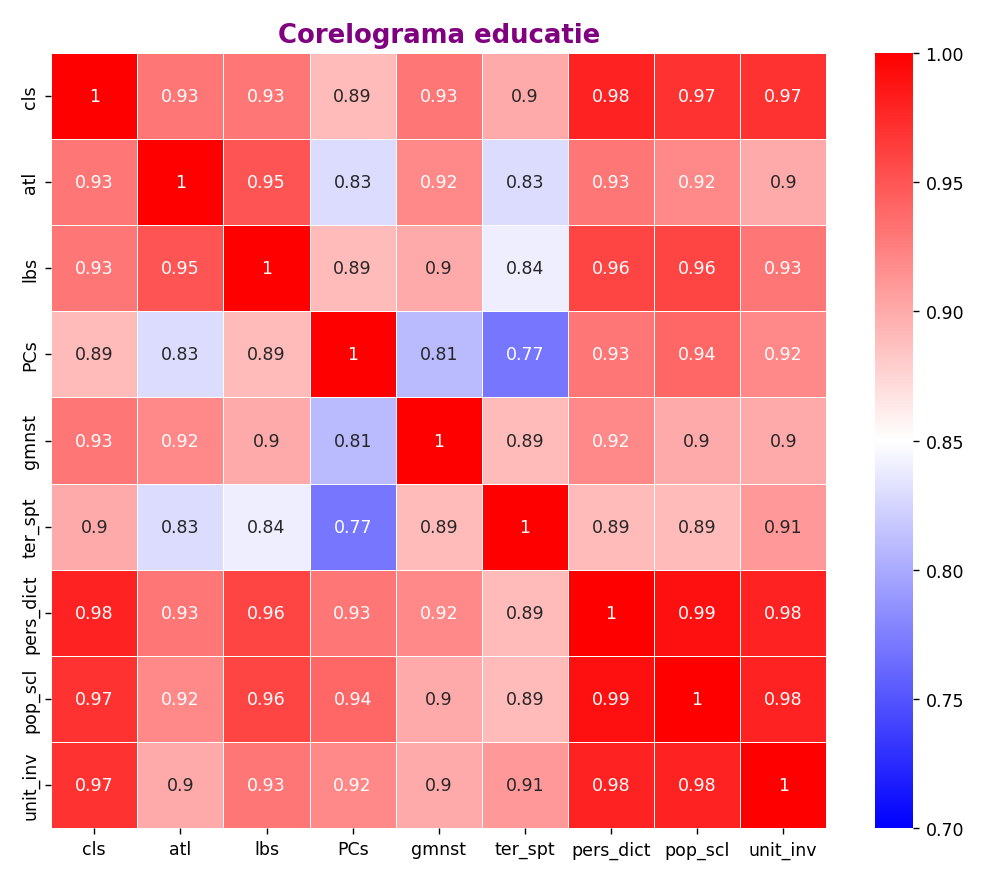
Prelucrările au fost realizate prin intermediul structurilor de date implicite limbajului python (liste, tupluri, dicționare și șiruri de caractere), dar și a unora externe preluate din bibliotecile pandas și numpy (DataFrame, Series și ndarray).

1. Prelucrările realizate asupra datelor

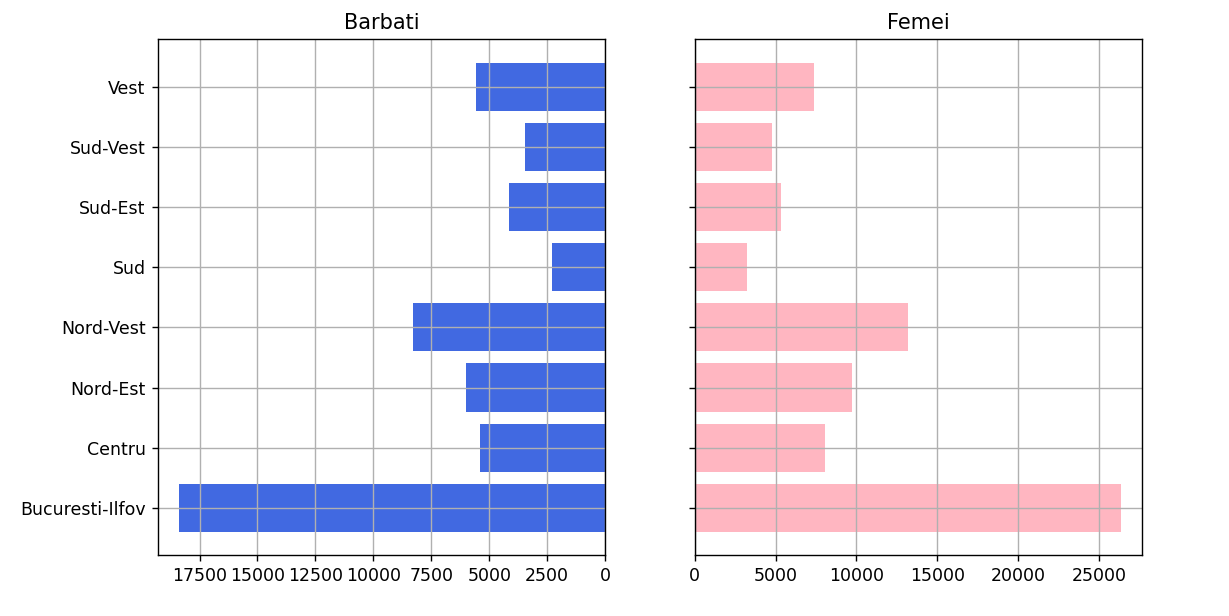
Inițial, datele necesare analizei sunt preluate prin intermediul metodelor specializate din biblioteca “pandas”.

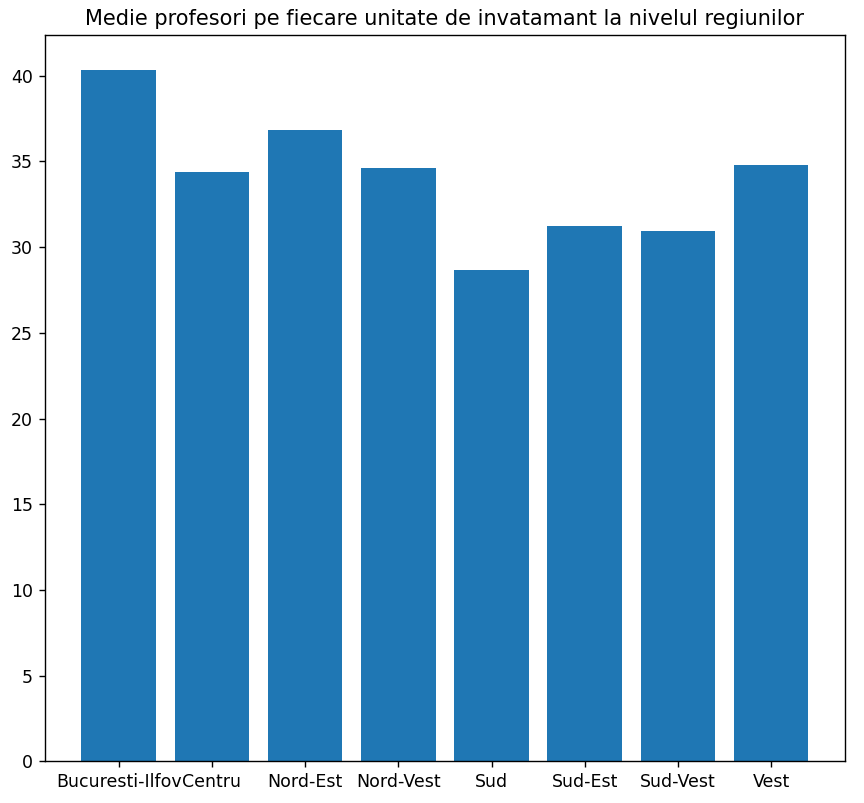
Prima prelucrare este realizată asupra unităților de învățământ. Se măsoară tendința centrală a acestei variabile prin calcularea mediei simple. Cu toate observațiile luate în considerare media era de aproximativ 167 de unități pe județ. Totuși, calculând coeficientul de variație, care depășea pragul de 35%, am putut să-mi dau seama de faptul că media ințială nu era reprezentativă. Din acest motiv, am determinat valorile aberante, atât vizual prin crearea graficului box-plot, cât și matematic prin utilizarea regulii lui Cebâșev și ulterior le-am scos din calcul. După această etapă, media a devenit reprezentativă și astfel pot afirma că în fiecare județ se află în medie aproximativ 156 de unități de învățământ.

Ulterior, am analizat existența, sensul și intensitatea legăturile tuturor variabilelor studiate (cu excepția numărului de absolvenți) prin crearea matricelor de covarință și corelație care au fost realizate cu ajutorul funcțiilor specifice din biblioteca “numpy” . De asemenea, pentru a vizualiza mai ușor rezultatele obținute, am creat o corelogramă sub forma unui grafic de tip heatmap.



Se poate observa că toate variabile sunt corelate puternic, ceea ce este foarte logic.

În următoarea etapă a proiectului am decis să studiez diferențele de sex existente în România din punctul de vedere al numărului de absolvenți ai studiilor universitare (această categorie include diploma de licență, master, cursuri postuniversitare, doctorat și programe postdoctorale de cercetare avansată). Pentru a interpreta mai ușor existența unor posibile dezechilibre am creat grafice de tip piramidal. Se observă că numărul femeilor absolvente este mai mare decât cel al bărbaților, iar acest lucru poate fi explicat de faptul că majoritatea persoanelor de sex masculin încep să lucreze de timpuriu și nu mai pun accentul pe educație atât de mult.

Pentru a observa mai ușor anumite tendințe și din cauza faptului că unele județe aveau pentru majoritatea variabilelor valori aberante, am decis să analizez datele și la nivelul regiunilor, respectiv a macroregiunilor. Inițial, am căutat pentru fiecare variabilă regiunea cu valoarea maximă, iar analog am făcut și pentru macroregiuni. Astfel, de exemplu, am putut observa că regiunea nord-est este bine pregătită din perspectiva infrastructurii educației. De asemenea, am calculat numărul mediu de profesori pe unitatea de învățământ și numărul de absolvenți pe sexe pentru fiecare regiune. Pentru analiza personalului didactic am creat și un grafic cu bare.

În final, am transformat datele inițiale, standardizându-le, această etapă fiind importantă pentru analizele următoare.

Pe parcursul întregului proiect, rezultatele matriciale mai importante au fost salvate în fișiere cu extensia csv sau xlsx.

1. [*http://statistici.insse.ro:8077/tempo-online/#/pages/tables/insse-table*](http://statistici.insse.ro:8077/tempo-online/#/pages/tables/insse-table) [↑](#footnote-ref-1)