A picture containing sketch, clipart, logo, line art

Description automatically generated

Academia de Studii Economice din București

Facultatea de Cibernetică, Statistică și Informatică Economică

Specializarea Informatică Economică

RAPOARTE AGREGATE

Analiza numărului mediu de ore pe care un american îl doarme în medie pe zi într-un an

Coordonator

Prof. univ. Andreescu Ioana

Student

Cazacu Irina Claudia

Grupa 1081, Seria C

Ciurea Claudia Maria

Grupa 1082, Seria C

Cuprins

[1.Descrierea setului de date 2](#_Toc135762832)

[2.Operații realizate 4](#_Toc135762833)

[3.Grafice 5](#_Toc135762834)

## 1.Descrierea setului de date

Somnul are o importanță crucială pentru sănătatea și bunăstarea generală a unei persoane pentru ca somnul adecvat susține sănătatea fizică. In cele ce urmeaza se prezinta o analiza a unui set de date ce cu privire la numarul mediu de ore pe care un American il doarme intr-o zi in functie de sex, grupa de varsta si tipul de zi (zi normala, weekend normal si weekend in perioada vacantei), obervatiile fiind facute pe un set de date de tip panel (2003-2017).

Variabilele sunt urmatoarele:

*Index* Numarul observatiei

*Year* Anul in care a fost preluata observatia

*Period* Perioada de timp careia ii corespunde media

Avg hrs per day sleeping Valoarea medie a numarului de ore dormite

*Standard Error* Eroarea standard

*Type of Days* Tipul zilei

*Age Group* grupele de varsta pe care s-au realizat observatiile

*Activity* activitatea persoanelor pe care s-au realizat observatiile

*Sex* sexul persoanelor

Setul de date ”Time Americans Spend Slepping.csv ” a fost preluat de pe

<https://www.kaggle.com/datasets/thedevastator/how-much-sleep-do-americans-really-get>

## 2.Operații realizate

Python este un limbaj de programare puternic și versatil, potrivit pentru prelucrarea analitică a datelor și crearea de rapoarte agregate. Există mai multe biblioteci și instrumente disponibile în Python care facilitează acest proces: există biblioteci specializate pentru analiza datelor, cum ar fi pandas pentru manipularea datelor tabulare și matplotlib/seaborn pentru vizualizarea grafică, lucrul cu tipul de date DataFrame, operatii de filtrare, agregare si alte transformari..

Pe baza setului de date, s-au realizat mai multe grupari, agregari dupa diferite tipuri de variabile, folosind bibliotecile numPy si Pandas.

In df s-au salvat datele din CSV. Pe baza lui df s-au creat noi dataframes in functie de diferite variabile folosind groupBy si s-au aplicat metode agregate(count, sum , min, max,mean) si filtre(sorare crescatoare), folosind bibliotecile numpy si pandas.

De asemenea, datele au fost grupate dupa diferite variabile si salvate in numpy array-uri sau dataframeuri. Exista si o matrice, care cuprinde tot setul de observatii.

Rezultatele gruparilor si operatiilor efectuate pe setul de date s-au afisat in CSV-uri. De mentionat faptul ca s-a ales afisarea in csv a celor mai importante operatii pentru a nu ingreuna proiectul cu un numar foarte mare de CSV-uri.

|  |  |
| --- | --- |
| CSV-uri in care s-au afisat: | Preluate din: |
| Avg\_hrs\_Sex\_Year.csv | df2 |
| AgeGroup\_Sex\_Year\_functii.csv | df3 |
| Count\_TypeOfDays.csv | df4 |
| Count\_Year.csv | df7 |
| TypeOfDay\_Sex\_functii.csv | df8 |
| Sex\_AgeGroup\_first.csv | df5 |
| Ranked\_age\_sex.csv | ranked\_df |
| TypeofDays\_Sex\_hours.csv | df6 |
| Women.csv | dff- f\_filter |
| Men.csv | df- m\_filter |
| Both.csv | df- b\_filter |
| MinMax.csv | f\_avg |
| Hours.csv | avg\_hours |
| Grupare\_dupa\_sex.csv | sex |
| Grupare\_dupa\_Zi.csv | day |
| Grupare\_dupa\_2003.csv | xx |
| Grupare\_dupa\_2003\_BOTH.csv | xxx |
| Sortare\_cresc\_Both.csv | zz |

## 3.Grafice

Grafic:

1. Barchart Ox: numarul mediu de ore Oy: anii



1. Piechart – Ponderi ore somn în funcție de ani si tipul zilei

A picture containing screenshot, colorfulness, text, design

Description automatically generated