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

Ciurea Claudia 1082C

Conținut

[Descriere proiect 2](#_Toc121955829)

[1.Descriere date 2](#_Toc121955830)

[2.Operații realizate 2](#_Toc121955831)

[3.Grafice 4](#_Toc121955832)

# Descriere proiect

## 1.Descriere date

In fisierul ”Time Americans Spend Slepping.csv ” se afla informatii privind 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). Informatiile sunt urmatoarele:

*Index*

*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

## 2.Operații realizate

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

In df s-au salvat datele din CSV-ul meu. 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.

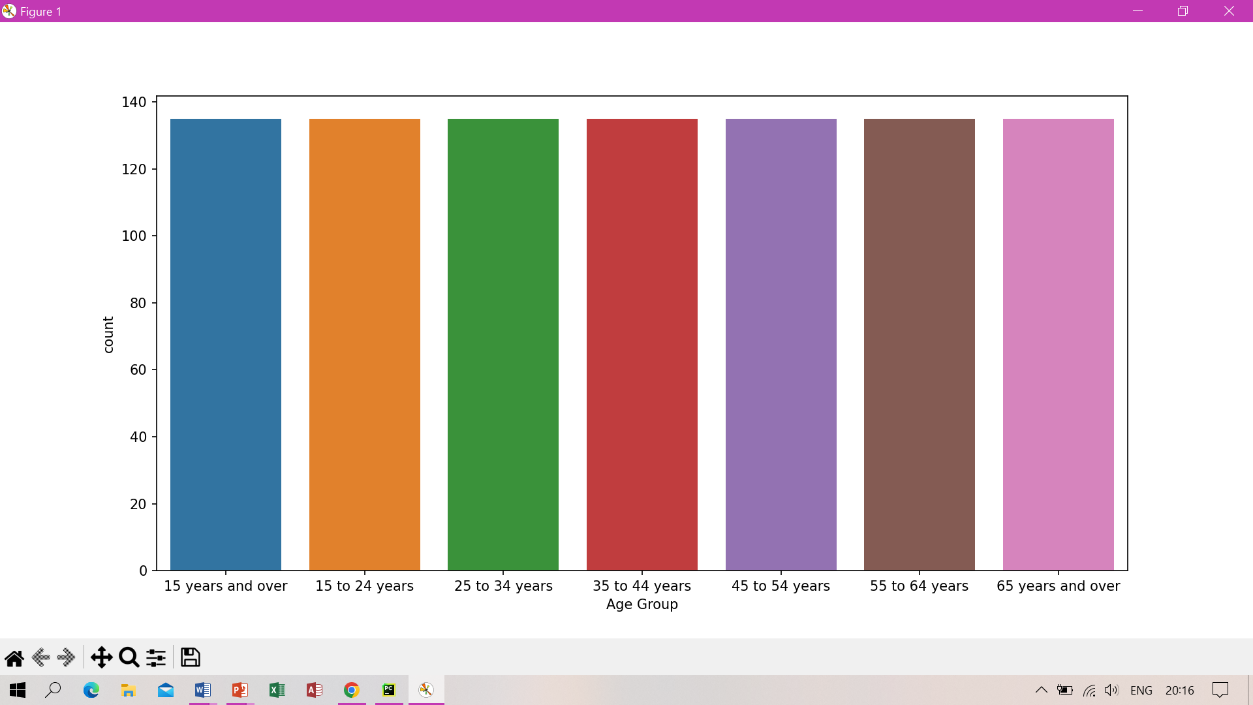
Rezultatele gruparilor si operatiilor efectuate pe setul de date s-au afisat in CSV-uri. De mentionat faptul ca s-a ales afisarea in csv a celor mai importante operatii pentru a nu ingreuna proiectul cu un numar foarte mare de CSV-uri.

|  |  |
| --- | --- |
| 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 |
| 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

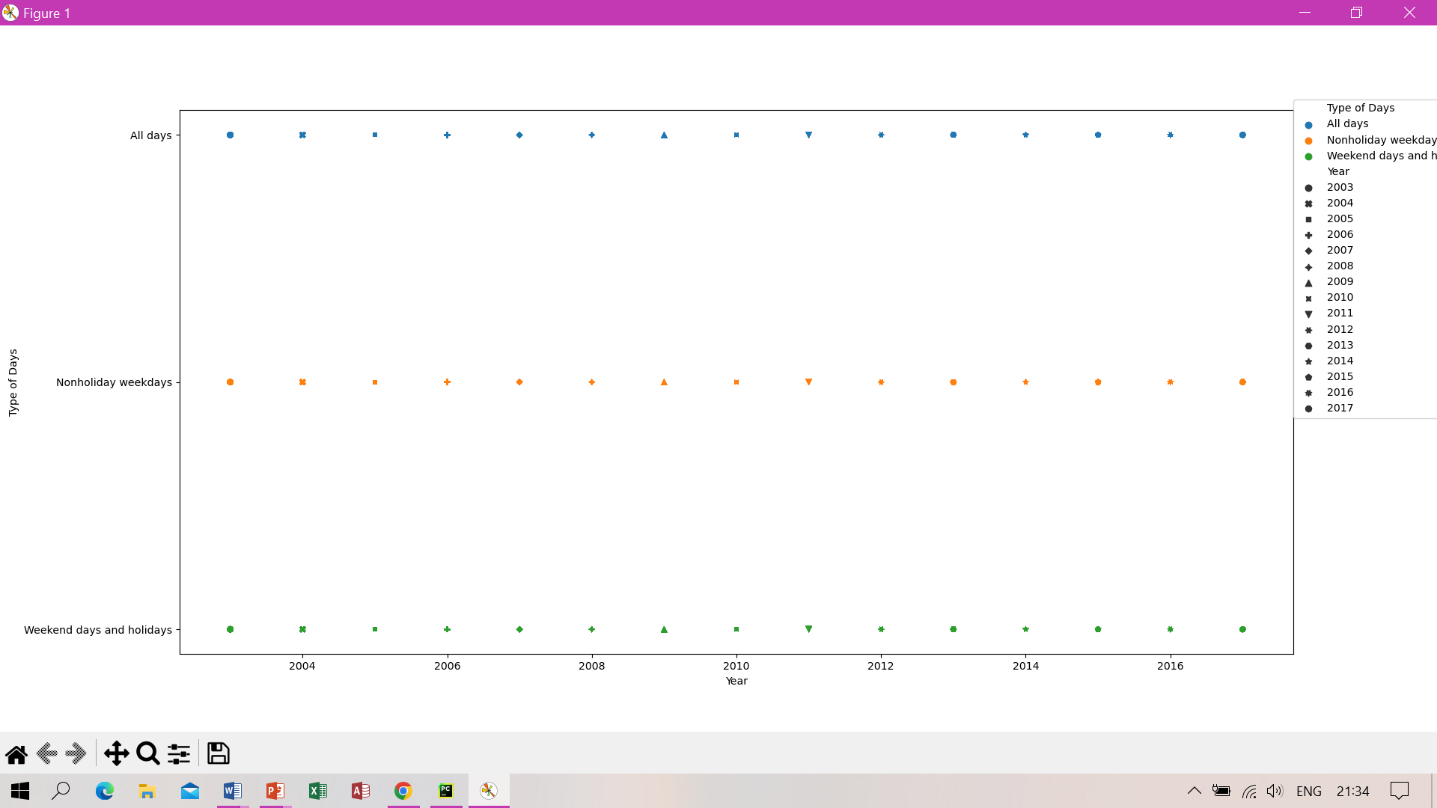
Grafic BarChart: pe axa Ox – grupele de varsta

Pe axa Oy-contor, cate observatii sunt din acea grupa de varsta



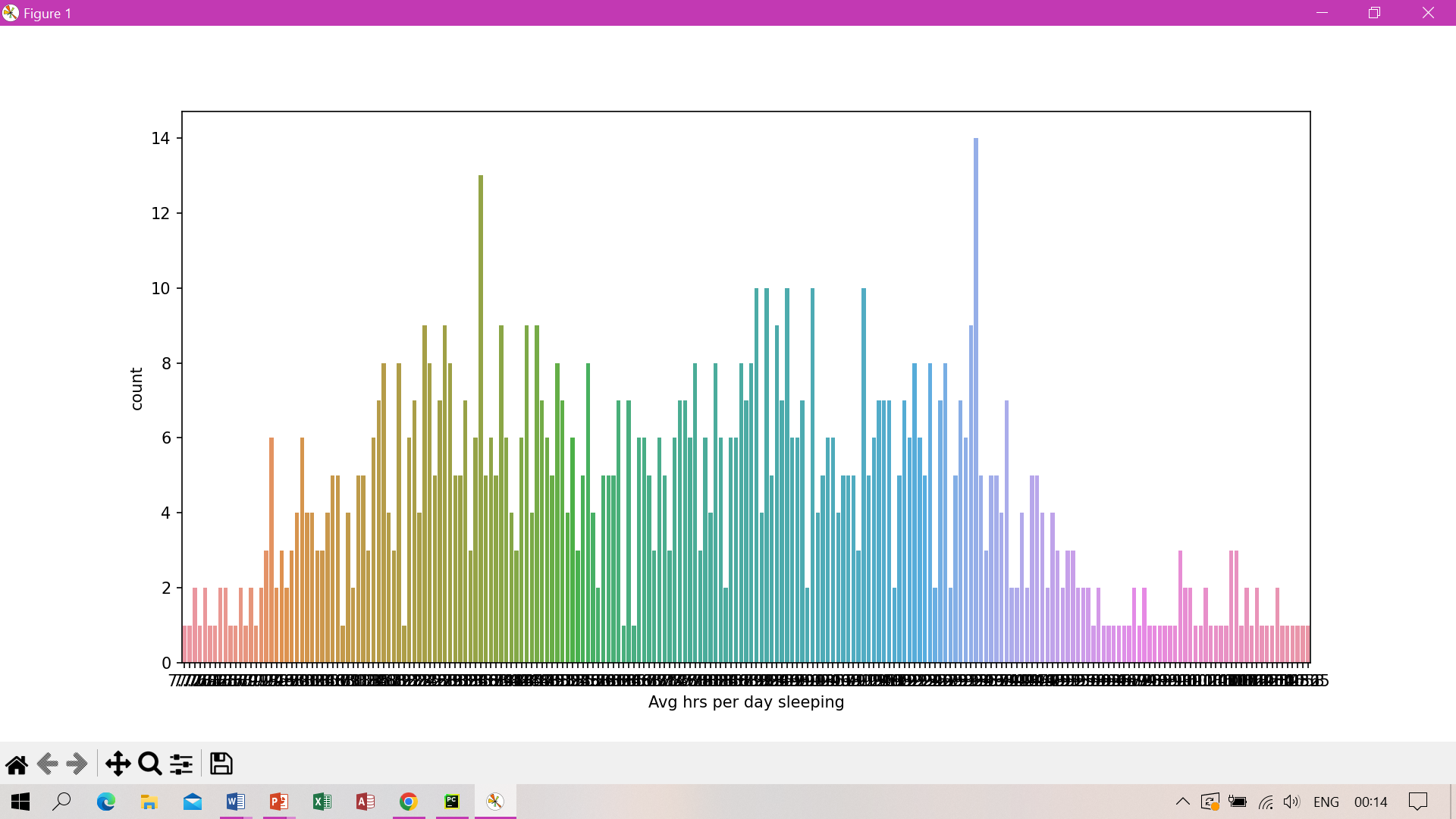
Grafic2 scatterPlot : Ox: Anii

Oy: Tipul de zi



Grafic3 BarChart Ox: numarul mediu de ore

Oy:numarul de observatii pentru un numar



Grafic4: barchart Ox: numarul mediu de ore

Oy: anii

