

Laborator 4

1. Exerciții cu Pandas

1. Incarcati setul de date in format tsv

<https://raw.githubusercontent.com/justmarkham/DAT8/master/data/chipotle.tsv>

(<https://raw.githubusercontent.com/justmarkham/DAT8/master/data/chipotle.tsv>) intr-un obiect Pandas DataFrame.

- A. (optional) Afisati primele 5 linii din tabel
- B. (optional) Care este numarul de inregistrari din DataFrame?
- C. (optional) Comparati iesirea metodelor `info()` si `describe()`.
- D. (optional) Care e numarul de coloane? afisati numele lor.
- E. (0.5 puncte) Cum e indexat setul de date? sunt valorile din index unice?
- F. (0.5 puncte) Care sunt marfurile vandute, in ordine alfabetica? Cate sunt?
- G. (0.5 puncte) Cate obiecte (items) au fost comandate, in total - suma pe coloana 'quantity'?
- H. (0.5 puncte) Transformati intr-un alt dataframe coloana 'item_price' din stringuri in valori floating point. Care sunt preturile (minim, maxim, mediu, median) cu care s-a vandut produsul 'Chicken Crispy Tacos'? Cate astfel de produse au fost comandate in total? Care sunt veniturile obtinute din vanzarea acestui produs? *Nota*: puteti folosi metoda `describe()` a unui dataframe/series.
- I. (0.5 puncte) Care sunt cele mai comandate 3 marfuri? Folositi `groupby()` pentru grupare, ne intereseaza cantitatile cele mai mari. *Nota*: sectiunea 'Aggregation and Grouping' din [Python Data Science Handbook](https://jakevdp.github.io/PythonDataScienceHandbook/) (<https://jakevdp.github.io/PythonDataScienceHandbook/>) e utila pentru intelegerea agregarilor.
- J. (0.5 puncte) De cate ori s-a comandat 'Veggie Salad Bowl'?
- K. (0.5 puncte) De cate ori intr-o comanda s-au cerut mai mult de o 'Canned Soda'?
- L. (0.5 puncte) Care sunt sumele incasate pentru fiecare comanda? atentie, pentru fiecare item comandat trebuie luat in considerare si cantitatea. Puteti crea coloane auxiliare (pret transformat din string in float, pret total pe acel item = pret_unitar * cantitate etc.)

2. Incarcati setul de date de la

https://raw.githubusercontent.com/jokecamp/FootballData/master/UEFA_European_Championship/Euro%202

(https://raw.githubusercontent.com/jokecamp/FootballData/master/UEFA_European_Championship/Euro%202) intr-un obiect Pandas DataFrame.

- A. (optional) Cate echipe au participat la Euro 2012?
- B. (optional) Sa se sorteze echipele dupa numarul de cartonase rosii, apoi dupa cele galbene, descrescator - sortare dupa criteriu dublu.
- C. (0.5 puncte) Care e numarul mediu de cartonase rosii date? Dar mediana?
- D. (0.5 puncte) Care sunt echipele care au luat mai multe cartonase rosii decat media?
- E. (obligatoriu, 0 puncte) Care sunt echipele al caror nume incepe cu litera G? *Indicatie*: documentati-va despre Vectorized String Operations, sectiunea omonima din Ref bib 1 din curs 1.
- F. (0.5 puncte) Care sunt Shooting accuracy pentru Anglia, Grecia, Rusia?

3. Incarcati setul de date de la <https://raw.githubusercontent.com/justmarkham/DAT8/master/data/drinks.csv>

(<https://raw.githubusercontent.com/justmarkham/DAT8/master/data/drinks.csv>). Folosind grupari (`groupby`)

dupa o documentare serioasa in sectiunea 'Aggregation and Grouping' din ref bibliografica [1] din primul curs, rezolvati:

- A. (0.5 puncte) Pe care continent se bea mai multa bere, in medie?
- B. (0.5 puncte) Pentru fiecare continent sa se afiseze statisticile pentru consumul de vin.
- C. (0.5 puncte) Sa se tipareasca min, max, media pentru consumul de bere pe fiecare continent. Folositi functia `agg`.

Precizari:

- Se acorda 3 puncte din oficiu.
- Predarea se va face in saptamana 25-29 martie 2019.
- Temele se pot lucra in echipa se se depun in acest caz in repository-urile github create pentru prima tema.
- Punctajul problemelor va fi precizat in decursul acestei saptamanini, urmariti documentul.

In []: