

16. час: Утврђивање градива

Сада ћемо мало застати и осврнути се на материјал са последња два часа.

Задатак 1. Погледај пажљиво наредбе у следећој ћелији, па одговори на питања:

```
In [1]: import pandas as pd

temp_anomalije = pd.read_csv("podaci/TemperaturneAnomalije.csv", header=None)
temp_anomalije1 = temp_anomalije.T
temp_anomalije1.columns = ["Година", "Аномалија"]

aditivi = pd.read_excel("podaci/Aditivi.xlsx", sheet_name="Адитиви", na_filter=False)

drzave = pd.read_csv("https://raw.githubusercontent.com/cs109/2014_data/master/countries.csv")
US = pd.read_html("https://simple.wikipedia.org/wiki/List_of_U.S._states", header=0)
```

1. Како бисмо учитали табелу `TemperaturneAnomalije.csv` када би она имала заглавље?
2. Зашто смо на табелу `temp_anomalije` применили функцију `T`?
3. Колико колона има табела `temp_anomalije1`?
4. Шта ће урадити функција `read_excel` ако се уклони аргумент `sheet_name="Адитиви"` (знамо да табела има само један радни лист)?
5. Шта значи аргумент `na_filter=False` функције `read_excel`?
6. Да ли је табела `drzave` формирана на основу података који су смештени локално, на твом рачунару?
7. У ком формату су представљени подаци на основу којих је формирана табела `US`?
8. Шта би се десило када би у интернет претраживач укуцао `https://simple.wikipedia.org/wiki/List_of_U.S._states`?

Задатак 2. У фолдеру `podaci` се налази датотека `StanovnistvoSrbije2017.csv` (која има заглавље). Табела има три колоне које се зову "Старост", "Мушко" и "Женско".

(а) Учитати датотеку у структуру података `DataFrame` и индексирати табелу колоном "Старост".

(б) На основу података из табеле израчунати колики је процењени број становника у следећим старосним групама:

- 0--17 година,
- 18--65 година, и
- 66 и више година,

и представити ова три податка секторским дијаграмом. (Упутство: следећи израз може бити од помоћи: `tabela.loc["0":"17", "Мушко":"Женско"]`)

Задатак 3. Ученици једног разреда су скакали у даљ. Сваки ученик је скакао три пута и резултати су дати у датотеци *SkokUDalj.csv* која се налази у фолдеру *podaci*. Табела има заглавље и састоји се од четири колоне: "Презиме и име", "Скок1", "Скок2" и "Скок3".

(a) Учитати датотеку у структуру података *DataFrame*.

(e) Издвојити из табеле оне редове које садрже ученике који су начинили бар један преступ. Преступ је у табели означен тако што је дужина одговарајућег скока постављена на 0.

Задатак 4. У табели *podaci/SO2.xlsx* налазе се резултати мерења концентрације сумпор-диоксида у 2017. години у неким градовима Србије. Табела има четири колоне:

- МернаСтаница = Мерна станица
- СГВ = Средња годишња вредност у микрограмима по кубном метру
- БД125 = Број дана са више од 125 микрограма по кубном метру
- МДВ = Максимална дневна вредност у микрограмима по кубном метру

(a) Учитати ову табелу у структуру података *DataFrame*.

(б) Сортирати подаке по колони СГВ и приказати вредности у овој колони хистограмом.

Задатак 5. У табели *podaci/Cricket.xlsx* се налазе подаци о најбољим играчима крикета. Ова табела има заглавље.

(a) Учитати ову табелу у структуру података *DataFrame* и исписати првих неколико редова табеле да видимо како изгледа. Индексирати табелу колоном "Player".

(б) Додати табели нову колону "YP" (*Years Played*) и у њу уписати колико година је сваки играч био активан. (За сваког играча од године у колони "To" одузети годину у колони "From").

(e) Додати табели нову колону "ARY" (*Average Runs per Year*) и у њу уписати количник бројева из колоне "Runs" и "YP". ($ARY = Runs / YP$).

(e) Сортирати табелу по колони "ARY" од највећих ка најмањим вредностима и приказати првих 25 редова табеле. У ком веку су били активни скоро сви од ових 25 играча? Шта мислиш зашто?

In []: