

Wizualizacja danych w Python, semestr letni 2022

Anna Muranova

Ćwiczenie 9

Upewnij się, że w projekcie masz podpięta bibliotekę Pandas.

Ściągnij: https://github.com/pandas-dev/pandas/blob/master/doc/cheatsheet/Pandas_Cheat_Sheet.pdf

<https://www.amazeddeveloper.pl/blog/pandas-basics-1>

“the name is derived from the term “panel data”, an econometrics term for multidimensional structured data sets.”

program91.py

Stwórz program i wykonaj w nim poniższe czynności:

- stwórz jednowymiarową serię danych przechowującą liczby całkowite
- stwórz jednowymiarową serię danych przechowującą stringi
- stwórz listę a następnie przekształć ją na serię
- przekształć jedną z serii stworzonych wcześniej na listę
- wykonaj dodawanie, odejmowanie, mnożenie i dzielenie na dwóch dowolnych seriach danych z indeksami (nazwami). Jak to działa?
- stwórz serię danych przechowującą 10 liczb losowych z przedziału $[-10, 10]$ z krokiem 0.1 (jak to zrobić?), a następnie stwórz serię zawierającą liczby ujemne z wcześniej stworzonej serii w tym podpunkcie.

<https://pandas.pydata.org/docs/reference/api/pandas.Series.sample.html>

```
program91.py
```

Przekształć listę, słownik oraz serie danych na ramkę danych.

```
my_list = [1, 32, -37, 91, 12, 11, -5]
my_dict = {'a' : [1, 3, 2], 'b' : [2, 5, 7], 'c' : [3, 4, 8],
           'd' : [4, 10, 12]}
```

dodaj nazwy wierszy i kolumny w ramkach danych (jeżeli jeszcze nie ma).

Przekształć odwrotnie (w taki sam list, słownik, serie!).

https://pandas.pydata.org/docs/reference/api/pandas.DataFrame.to_dict.html

<https://www.geeksforgeeks.org/how-to-convert-pandas-dataframe-into-a-list/>

Stwórz ramkę danych i poćwicz na nią operacje związane z wyciągnięciem elementów, sortowaniem w kolumnie, zmianą kształtu.

	<i>a</i>	<i>b</i>	<i>c</i>	<i>d</i>	
<i>l1</i>	1	2	4	5	
<i>l2</i>	-3	8	0.5	10	
<i>l3</i>	2	-5	7	3	

Dla podanych tablic stwórz ramki danych (numery jako indeksy, nazwy Name, Age, ... jako nazwy kolumn)

ID	Name	Age
2312	Anna	21
2336	Zofia	40
2942	Sylwia	13
9840	Katarzyna	31
2794	Teresa	34
2933	Zenon	28

ID	Name	W	H	Glasses
2942	Sylwia	64	151	F
9840	Katarzyna	69	177	T
2794	Teresa	74	170	F
8891	Tomasz	61	157	T
8111	Cezary	66	151	F
2933	Zenon	61	153	T

- Połącz tabele najpierw po imieniu i po numerze (`my_df`) metodę `inner`.

```
df1 = pd.DataFrame
      ([[2942,9840,2794,8891,8111,2933,8301,9125], [ 'Sylwia',
        'Katarzyna', 'Teresa', 'Tomasz', 'Cezary', 'Zenon', '
        Filip', 'Adrian'],[13, 31, 34, 14, 13, 28, 20, 15]]).T
df1.columns = ['ID', 'Name','Age']
weight = [65, 80, 64, 69, 74, 61, 66, 61]
height = [179, 179, 151, 177, 170, 157, 151, 153]
glasses = [False, True, False, True, False, True, False,
           True]
df2 = pd.DataFrame
      ([[2312,2336,2942,9840,2794,8891,8111,2933], [ 'Anna', '
        Zofia', 'Sylwia', 'Katarzyna', 'Teresa', 'Tomasz', '
        Cezary', 'Zenon'],weight,height,glasses]).T
df2.columns = ['ID','Name','W', 'H','G1']
```

- posortuj imiona alfabetycznie
- stwórz tablicę przechowującą imiona osób noszących okulary (kolejność w tej tablicy musi odpowiednio zachować kolejność z wyjściowej tablicy)
- stwórz tablicę zawierającą imiona osób w wieku z przedziału lat [20,30]
- dodaj kolumnę z *bmi* dla wszystkich osób i wynik zapisz w tablicy ($bmi = \frac{waga}{wzrost^2}$).
- policz średni wiek i wyświetl na konsoli.

<https://stackoverflow.com/questions/51153010/slicing-with-a-logical-boolean-expression-a-pandas-dataframe>

Zapoznaj się z dokumentacją dotyczącą zapisu i odczytu plików <https://docs.python.org/3/tutorial/inputoutput.html#reading-and-writing-files>

program93.py

Za pomocą explorera plików i notatnika stwórz plik tekstowy tekst1.txt w tym samym folderze co projekt, wpisz do pliku dowolny tekst. Następnie za pomocą instrukcji open i różnych parametrów sprawdź różne tryby otwarcie plików.

Przykładowy kod:

```
f = open('tekst1.txt', 'r+')
s = f.read()
print(s)
print(type(s))
```

Następnie ustaw break pointa na trzeciej linii w powyższym przykładzie i spróbuj zmodyfikować i zapisać zmienioną zawartość pliku tekstowego przez notatnik, WprdPad, inny aplikacje.

Napraw sytuację wywołując metodę close.

Ew. zmień kod na taki z użyciem with.

Spróbuj dodać polskie znaki do pliku tekstowego i dodaj parametr encoding do funkcji open (poćwicz różne wersje kodowania).

Na koniec spróbuj zapisać coś nowego do pliku. Sprawdź w jakim trybie otwarcia plików jest to możliwe.

program94.py

Pobierz plik z cenami jaj w Polsce w wybranych sieciach (źródła danych: <http://wmii.uwm.edu.pl/~piojas/wp-content/uploads/2019/04/jajka1.csv>, <http://www.dlahandlu.pl/koszyk/towar/10-jaj-najtansze,38.html>) - <http://wmii.uwm.edu.pl/~muranova/WD/jajka.csv>.

Zapisz plik w folderze projektu. Załaduj go do środowiska np. poniższa komenda:

```
data = np.genfromtxt('jajka1.csv', delimiter=";", dtype=None)
https:
```

```
//numpy.org/doc/stable/reference/generated/numpy.genfromtxt.html
```

Dodaj najlepszy twoim zdaniem encoding.

Następnie przetwórz dane abyś mógł wykonać operacje:

- obliczyć średnią cenę wszystkich jaj.
- wyznaczyć w którym mieście i w jakiej sieci są najtańsze a w jakich najdroższe jajka. Wynik zapisz w postaci dwuwymiarowej tablicy przechowującej pary (Miasto, nazwa sieci).

Wszystkie operacje wykonaj używając funkcji wbudowanych w interpretator lub biblioteki NumPy.

Na kolejne ćwiczenia przygotuj 5 plików w formacie csv dostępnych na darmowych repozytoriach danych. Napisz do nich plik README i umieść informacje o źródle danych w postaci:

Nazwa pliku: sample1.csv

Opis: [krótki opis czego dotyczy dane zjawisko]

Zródło: [adres strony www, nazwa serwisu]

Licencja: [rodzaj licencji, o ile dotyczy]