

Eksploracyjna analiza danych - laboratorium 3

November 24, 2023

```
[1]: import pandas as pd
import numpy as np
```

1 Zadania do samodzielnego wykonania

1.1 Zadanie 1

Podane są dwa obiekty typu Series:

```
[2]: np.random.seed(42)
s1 = pd.Series(np.random.rand(20))
s2 = pd.Series(np.random.randn(20))
```

Połącz te dwa obiekty w jeden obiekt DataFrame (jako dwie kolumny) i przypisz do zmiennej df. Nazwij kolumny odpowiednio col1 oraz col2. Wydrukuj wynik do konsoli.

1.2 Zadanie 2

Wyodrębnij wiersze z obiektu df (z zadania 1), dla którego wartość w kolumnie col2 zawiera się między 0.0 a 1.0 (włącznie). Wydrukuj wynik do konsoli.

1.3 Zadanie 3

Przypisz nową kolumnę o nazwie col3, która będzie miała wartość 1, gdy wartość w kolumnie col2 jest nieujemna i -1 przeciwnie. Wydrukuj wynik do konsoli.

1.4 Zadanie 4

Przypisz nową kolumnę col4, która obetnie wartości z kolumny col2 do zakresu $[-1.0, 1.0]$.

1.5 Zadanie 5

Wykonaj poniższe polecenia: - wydrukuj do konsoli 5 najwyższych wartości w kolumnie col2 - wydrukuj do konsoli pustą linię - wydrukuj do konsoli 5 najniższych wartości w kolumnie col2

1.6 Zadanie 6

Oblicz skumulowaną sumę dla każdej kolumny. Wydrukuj wynik do konsoli. Wskazówka: Użyj metody `pd.DataFrame.cumsum()`.

1.7 Zadanie 7

Oblicz medianę dla zmiennej col2. Wydrukuj wynik do konsoli.

1.8 Zadanie 8

Wyodrębnij wiersze, dla których zmienna col2 przyjmuje wartości większe niż 0.0. Wydrukuj wynik do konsoli.

1.9 Zadanie 9

Wyodrębnij pierwsze 5 wierszy obiektu df i przekonwertuj go na słownik. Wydrukuj wynik do konsoli.

1.10 Zadanie 10

Wyodrębnij pierwsze 5 wierszy obiektu df, przekonwertuj go na format HTML (jako obiekt typu str) i przypisz do zmiennej df_html. Wydrukuj zawartość zmiennej df_html do konsoli.

1.11 Zadanie 11

Podany jest następujący obiekt DataFrame:

```
[3]: np.random.seed(42)
df = pd.DataFrame(np.random.rand(10, 4), columns=list('ABCD'))
```

Wyodrębnij wiersze obiektu df dla których kolumna C jest większa niż 0.8. Wydrukuj wynik do konsoli.

1.12 Zadanie 12

Wyodrębnij wiersze obiektu df dla których kolumna C jest większa niż 0.3, a kolumna D jest mniejsza niż 0.7. Wydrukuj wynik do konsoli.

1.13 Zadanie 13

Dokonaj iteracji przez pierwsze pięć wierszy obiektu df. Wydrukuj każdy wiersz do konsoli.

1.14 Zadanie 14

Ustaw wartość w wierszu z indeksem 3 dla kolumny B jako np.nan (użyj df.iloc) oraz ustaw wartość w wierszu o indeksie 8 dla kolumny D jako np.nan (użyj df.loc). Przypisz wartości do zmiennej df.

Wydrukuj obiekt df do konsoli.

1.15 Zadanie 15

Usuń wszystkie wiersze z obiektu df zawierające brakujące wartości i pozostałą część obiektu przypisz do zmiennej df1. Wydrukuj zmienną df1 do konsoli.

1.16 Zadanie 16

Indeks po usunięciu dwóch wierszy w obiekcie df1 nie jest po kolei. Zresetuj indeks (nie dodawaj nowej kolumny). Wyświetl wartość zmiennej df1 do konsoli.

1.17 Zadanie 17

Określ liczbę brakujących elementów w obiekcie df dla każdej kolumny. Wydrukuj wynik do konsoli.

1.18 Zadanie 18

Uzupełnij brakujące wartości wartością 0 i wynik przypisz do zmiennej df. Wydrukuj zawartość zmiennej df do konsoli.

1.19 Zadanie 19

Zamień kolejność kolumn w obiekcie df na podaną poniżej: - D, A, B, C
i przypisz do zmiennej df. Wydrukuj zawartość zmiennej df do konsoli.

1.20 Zadanie 20

Usuń kolumnę D z obiektu df. Wydrukuj obiekt df do konsoli.