

# zadanie

June 9, 2023

Zadanie analiza danych

Import bibliotek i wczytanie danych

```
[28]: import pandas as pd
import matplotlib.pyplot as plt
import seaborn as sns
```

Pobranie i wyświetlenie informacji o liczbie kolumn i wierszy:

```
[29]: print("Liczba kolumn",dt.shape[1])
print("Liczba wierszy",dt.shape[0])
dt = pd.read_csv("dane.csv")
print(dt)
```

Liczba kolumn 6

Liczba wierszy 10

	Duration	Average_Pulse	Max_Pulse	Calorie_Burnage	Hours_Work	\
0	30	80	120	240	10	
1	30	85	120	250	10	
2	45	90	130	260	8	
3	45	95	130	270	8	
4	45	100	140	280	0	
5	60	105	140	290	7	
6	60	110	145	300	7	
7	60	115	145	310	8	
8	75	120	150	320	0	
9	75	125	150	330	8	

	Hours_Sleep
0	7
1	7
2	7
3	7
4	7
5	8
6	8
7	8
8	8

9                    8

Wartość średnia

```
[30]: print(dt.mean())
```

```
Duration          52.5
Average_Pulse     102.5
Max_Pulse         137.0
Calorie_Burnage   285.0
Hours_Work         6.6
Hours_Sleep        7.5
dtype: float64
```

Odchylenie standardowe

```
[31]: print(dt.std())
```

```
Duration          16.201852
Average_Pulse     15.138252
Max_Pulse         11.352924
Calorie_Burnage   30.276504
Hours_Work         3.627059
Hours_Sleep        0.527046
dtype: float64
```

Mediana

```
[32]: print(dt.median())
```

```
Duration          52.5
Average_Pulse     102.5
Max_Pulse         140.0
Calorie_Burnage   285.0
Hours_Work         8.0
Hours_Sleep        7.5
dtype: float64
```

Wartość maksymalna

```
[33]: print(dt.max())
```

```
Duration          75
Average_Pulse     125
Max_Pulse         150
Calorie_Burnage   330
Hours_Work         10
Hours_Sleep        8
dtype: int64
```

Wartość minimalna

```
[34]: print(dt.min())
```

```
Duration          30
Average_Pulse     80
Max_Pulse        120
Calorie_Burnage   240
Hours_Work         0
Hours_Sleep        7
dtype: int64
```

Macierz korelacji dla zbioru danych.

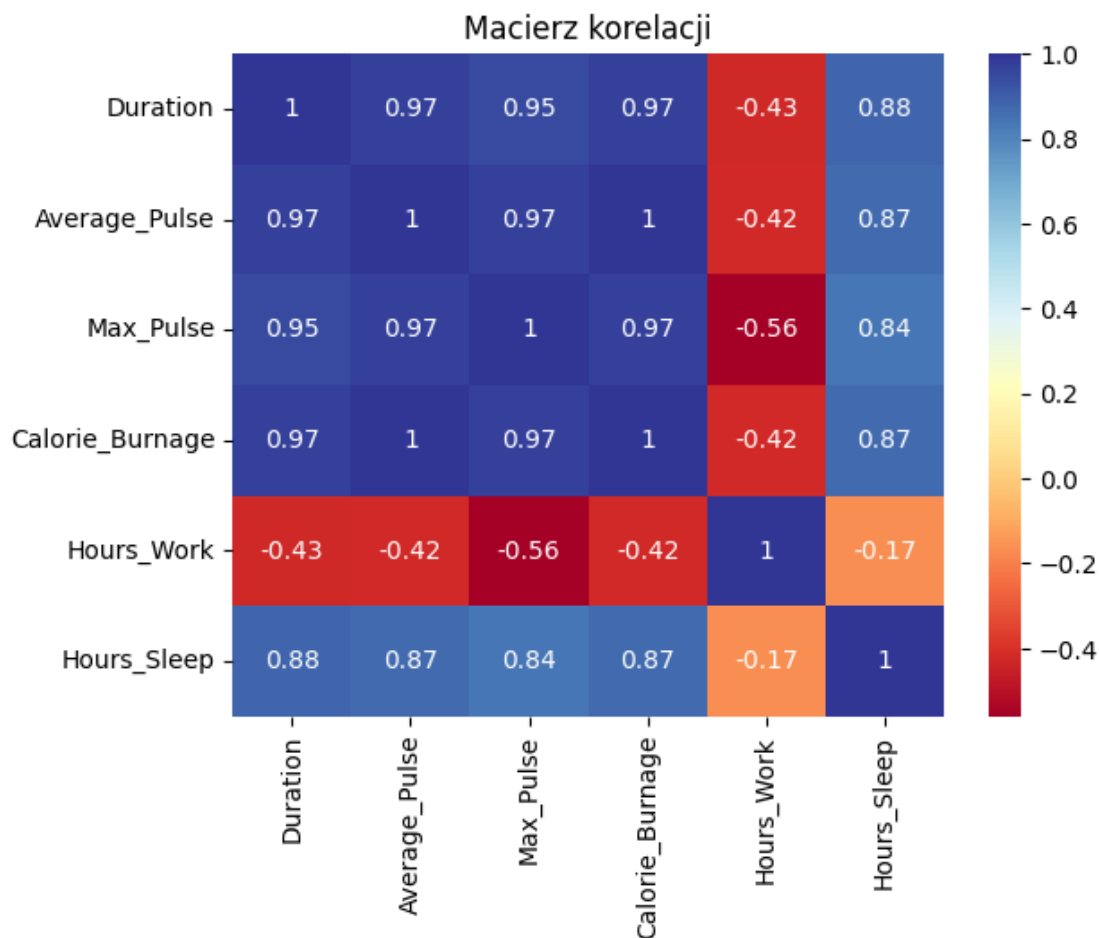
```
[35]: correlation_matrix = dt.corr().round(2)
      print(correlation_matrix)
```

	Duration	Average_Pulse	Max_Pulse	Calorie_Burnage	\
Duration	1.00	0.97	0.95	0.97	
Average_Pulse	0.97	1.00	0.97	1.00	
Max_Pulse	0.95	0.97	1.00	0.97	
Calorie_Burnage	0.97	1.00	0.97	1.00	
Hours_Work	-0.43	-0.42	-0.56	-0.42	
Hours_Sleep	0.88	0.87	0.84	0.87	

	Hours_Work	Hours_Sleep
Duration	-0.43	0.88
Average_Pulse	-0.42	0.87
Max_Pulse	-0.56	0.84
Calorie_Burnage	-0.42	0.87
Hours_Work	1.00	-0.17
Hours_Sleep	-0.17	1.00

Wykres korelacji wartości dla każdej z par kolumn.

```
[36]: sns.heatmap(correlation_matrix, annot=True, cmap="RdYlBu")
      plt.title("Macierz korelacji")
      plt.show()
```



Histogram rozkładu wartości dla wszystkich kolumn

```
[37]: for column in dt.columns:
    plt.hist(dt[column])
    plt.xlabel(column)
    plt.ylabel("Liczba wystąpień")
    plt.title(f"Histogram rozkładu wartości - {column}")
    plt.show()
```

