

Zadanie domowe - EDA Danych dotyczących Irysów

- Twoim zadaniem jest wykonanie analizy danych zawartych w załączonym pliku CSV (`25__iris.csv`).
- Plik znajduje się pod video. Pamiętaj, żeby wrzucić plik do tego samego folderu, w którym znajduje się ten notebook.
- Wykorzystaj nowo nabytą wiedzę z biblioteki `Pandas`
- Jeżeli jest taka konieczność posiłkuj się [code explainerem](#) lub / i [data chatterem](#) (który również te dane ma dostępne).
- Prześlij na discordzie notebook z rozwiązaniem (`#modul-4-zad1`)
- Pamiętaj:
 1. Nie spiesz się
 2. Potraktuj zadanie domowe jak prawdziwy projekt
 3. Dodawaj opisy, obserwacje, wnioski
 4. Dodaj wizualizacje
 5. Dodaj konkluzję i wnioski końcowe

Powodzenia!!! I mega zabawy!

O Danych

Zbiór danych zawiera informacje o trzech gatunkach irysów: Iris setosa, Iris versicolor, i Iris virginica.

Dane obejmują pomiary czterech cech: długość i szerokość działki kielicha oraz długość i szerokość płatków.

Każdy wiersz w zbiorze danych reprezentuje pojedynczy kwiat, a wartości pomiarów są podane w centymetrach.

Zbiór składa się z 150 próbek, po 50 dla każdego gatunku, i jest szeroko stosowany jako podstawowy zbiór danych do testowania algorytmów klasyfikacji oraz w nauce o danych i uczeniu maszynowym.

Kolumny:

- **długość kielicha (sepal length)** - Długość kielicha w cm
- **szerokość kielicha (sepal width)** - Szerokość kielicha w cm
- **długość płatków (petal length)** - Długość płatków w cm
- **szerokość płatków (petal width)** - Szerokość płatków w cm
- **klasa (class)** - Klasa irysa (setosa, versicolor, virginica)

Prezentacja danych

Początkowe wiersze w dataframe

	długość kielicha (sepal length)	szerokość kielicha (sepal width)	długość płatka (petal length)	szerokość płatka (petal width)	klasa (class)
0	5.100000	3.500000	1.400000	0.200000	Iris-setosa
1	4.900000	3.000000	1.400000	0.200000	Iris-setosa
2	4.700000	3.200000	1.300000	0.200000	Iris-setosa
3	4.600000	3.100000	1.500000	0.200000	Iris-setosa
4	5.000000	3.600000	1.400000	0.200000	Iris-setosa

Losowe wiersze

	długość kielicha (sepal length)	szerokość kielicha (sepal width)	długość płatka (petal length)	szerokość płatka (petal width)	klasa (class)
88	5.600000	3.000000	4.100000	1.300000	Iris-versicolor
115	6.400000	3.200000	5.300000	2.300000	Iris-virginica
58	6.600000	2.900000	4.600000	1.300000	Iris-versicolor
123	6.300000	2.700000	4.900000	1.800000	Iris-virginica
138	6.000000	3.000000	4.800000	1.800000	Iris-virginica
7	5.000000	3.400000	1.500000	0.200000	Iris-setosa
84	5.400000	3.000000	4.500000	1.500000	Iris-versicolor
3	4.600000	3.100000	1.500000	0.200000	Iris-setosa
24	4.800000	3.400000	1.900000	0.200000	Iris-setosa
37	4.900000	3.100000	1.500000	0.100000	Iris-setosa

Ostatnie wartości

	długość kielicha (sepal length)	szerokość kielicha (sepal width)	długość płatka (petal length)	szerokość płatka (petal width)	klasa (class)
145	6.700000	3.000000	5.200000	2.300000	Iris- virginica
146	6.300000	2.500000	5.000000	1.900000	Iris- virginica
147	6.500000	3.000000	5.200000	2.000000	Iris- virginica
148	6.200000	3.400000	5.400000	2.300000	Iris- virginica
149	5.900000	3.000000	5.100000	1.800000	Iris- virginica

Informacje o kolumnach

```
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 150 entries, 0 to 149
Data columns (total 5 columns):
#   Column                                Non-Null Count  Dtype
---  -
0   długość kielicha (sepal length)      150 non-null    float64
1   szerokość kielicha (sepal width)      150 non-null    float64
2   długość płatka (petal length)        150 non-null    float64
3   szerokość płatka (petal width)       150 non-null    float64
4   klasa (class)                        150 non-null    object
dtypes: float64(4), object(1)
memory usage: 6.0+ KB
```

Zbiorcze informacje o danych

Dane pogrupowane są w pięciu kolumnach, czterech liczbowych i jednej opisowej.

	długość kielicha (sepal length)	szerokość kielicha (sepal width)	długość płatka (petal length)	szerokość płatka (petal width)
count	150.000000	150.000000	150.000000	150.000000
mean	5.843333	3.054000	3.758667	1.198667
std	0.828066	0.433594	1.764420	0.763161
min	4.300000	2.000000	1.000000	0.100000
25%	5.100000	2.800000	1.600000	0.300000
50%	5.800000	3.000000	4.350000	1.300000
75%	6.400000	3.300000	5.100000	1.800000
max	7.900000	4.400000	6.900000	2.500000

```
długość kielicha (sepal length)    35
szerokość kielicha (sepal width)    23
długość płatka (petal length)      43
szerokość płatka (petal width)     22
klasa (class)                       3
dtype: int64
```

Analiza brakujących wartości

	długość kielicha (sepal length)	szerokość kielicha (sepal width)	długość płatka (petal length)	szerokość płatka (petal width)	klasa (class)
0	False	False	False	False	False
1	False	False	False	False	False
2	False	False	False	False	False
3	False	False	False	False	False
4	False	False	False	False	False
...
145	False	False	False	False	False
146	False	False	False	False	False
147	False	False	False	False	False
148	False	False	False	False	False
149	False	False	False	False	False

150 rows × 5 columns

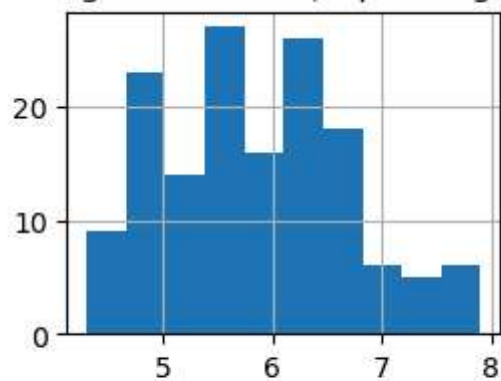
```
długość kielicha (sepal length)    0
szerokość kielicha (sepal width)    0
długość płatka (petal length)      0
szerokość płatka (petal width)     0
klasa (class)                       0
dtype: int64
```

Podsumowanie brakujących wartości

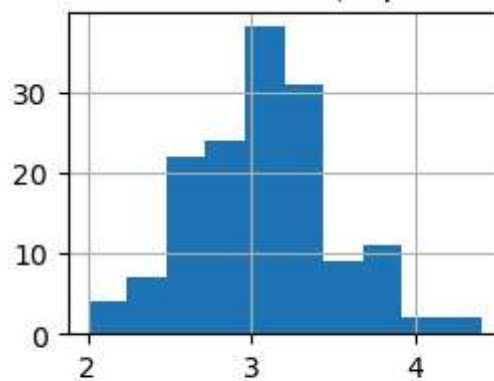
Dane są kompletne, nie zawierają braków.

Analiza pojedynczych zmiennych

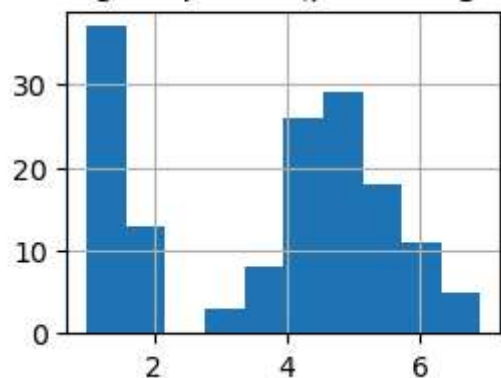
długość kielicha (sepal length)



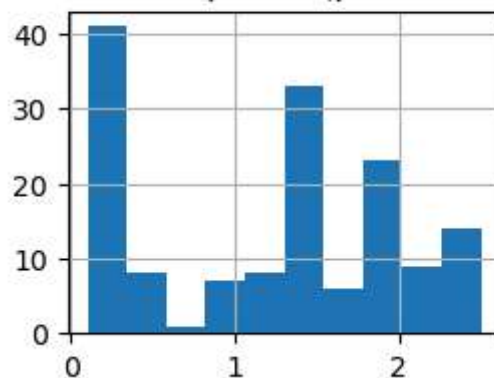
szerokość kielicha (sepal width)



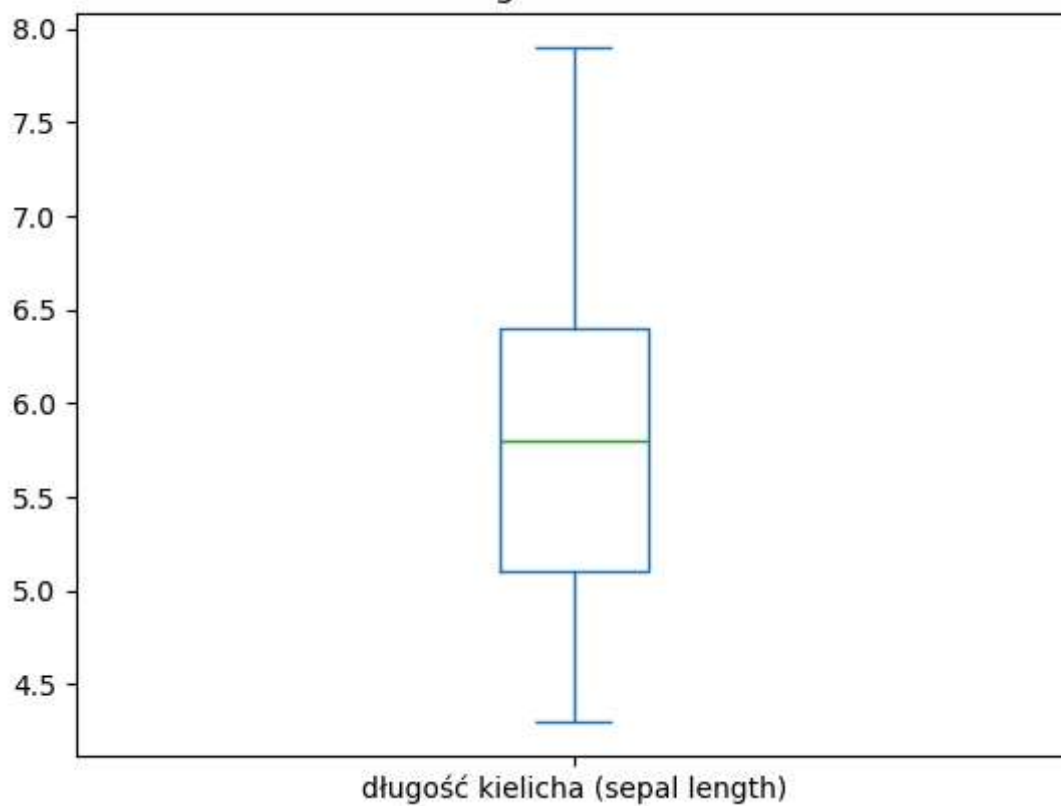
długość płatka (petal length)

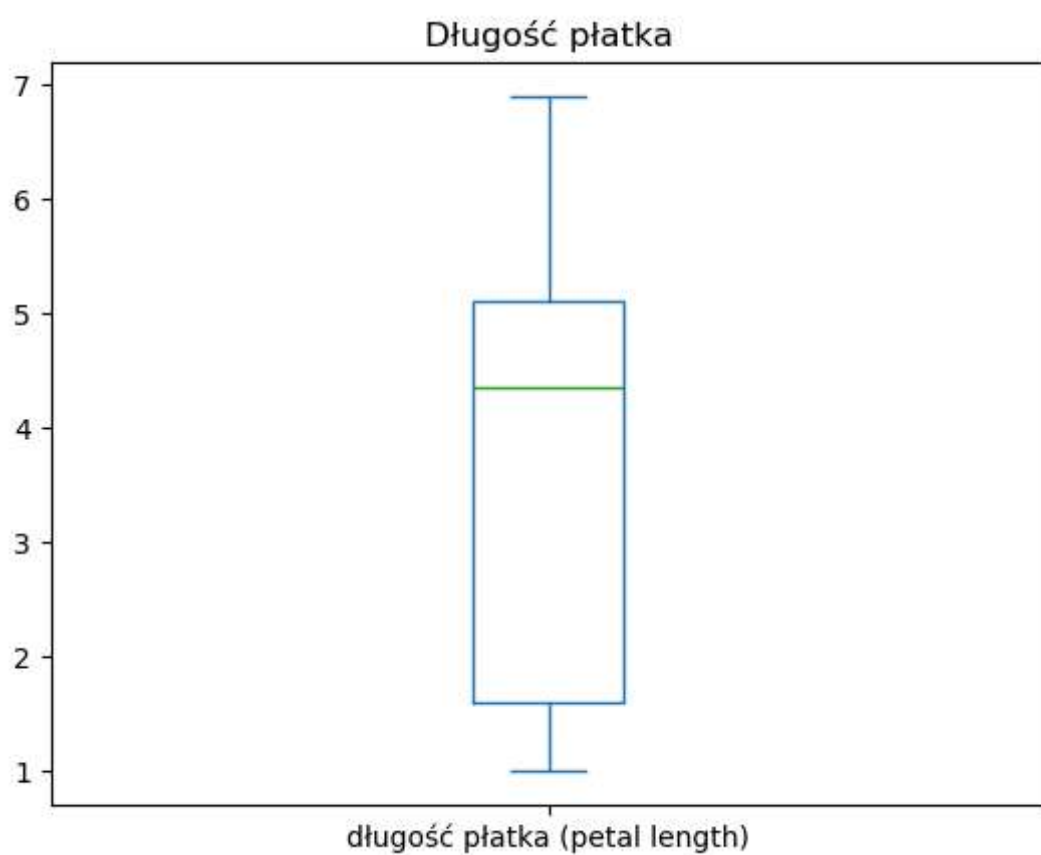
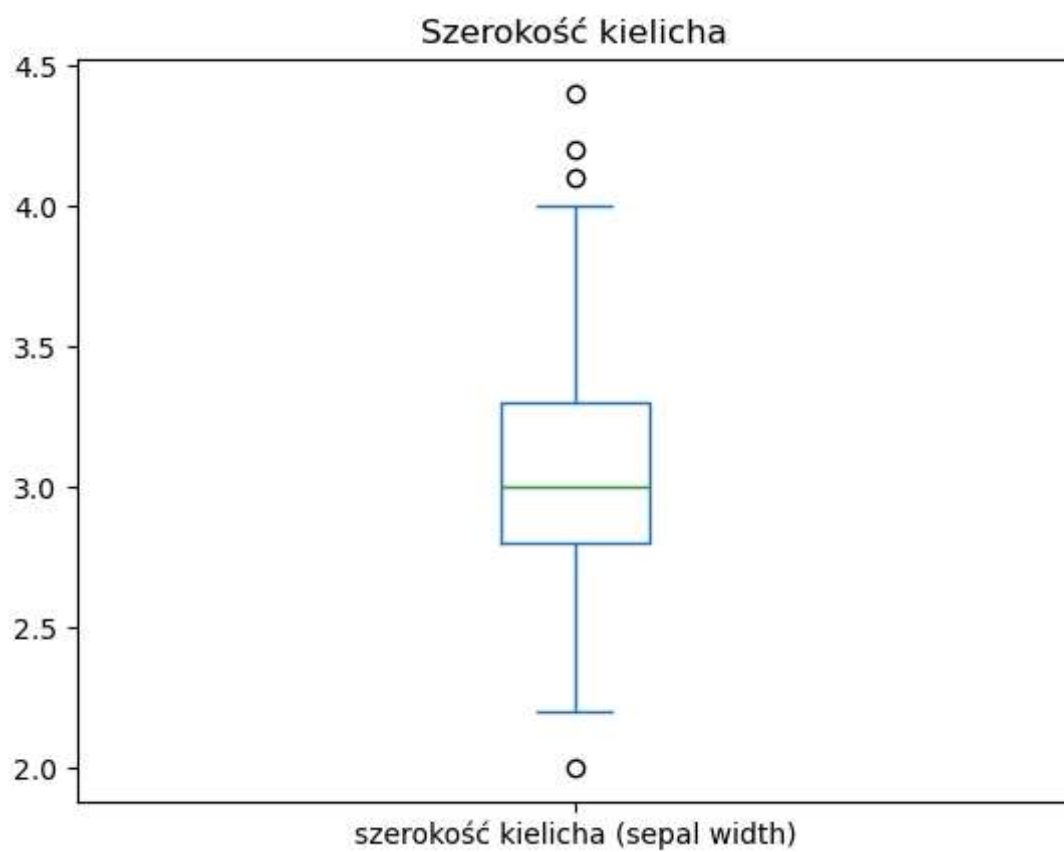


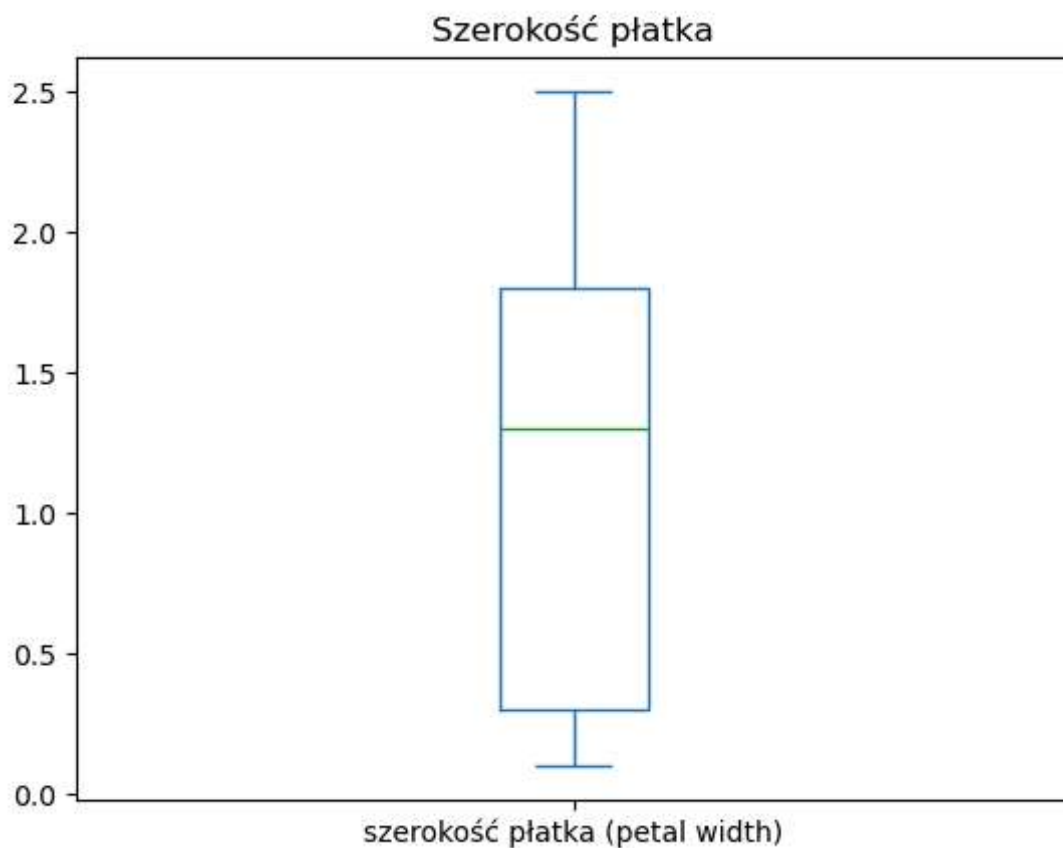
szerokość płatka (petal width)



Długość kielicha

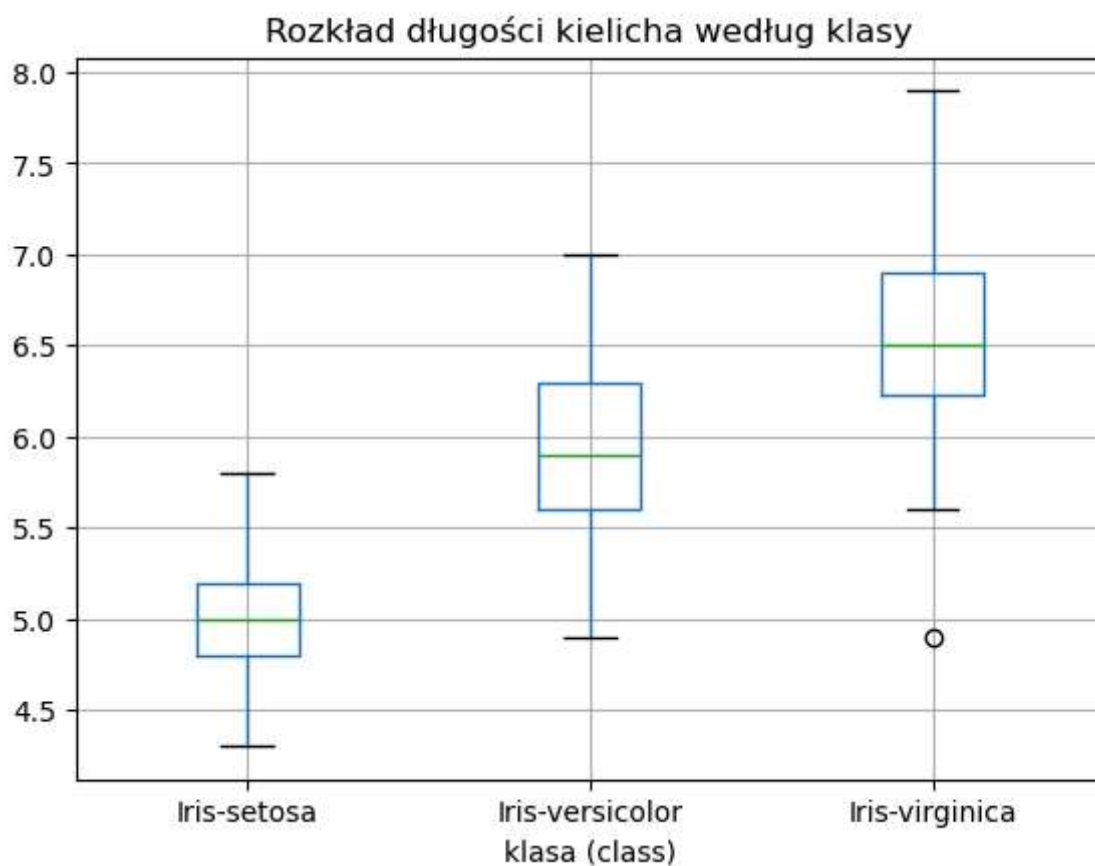






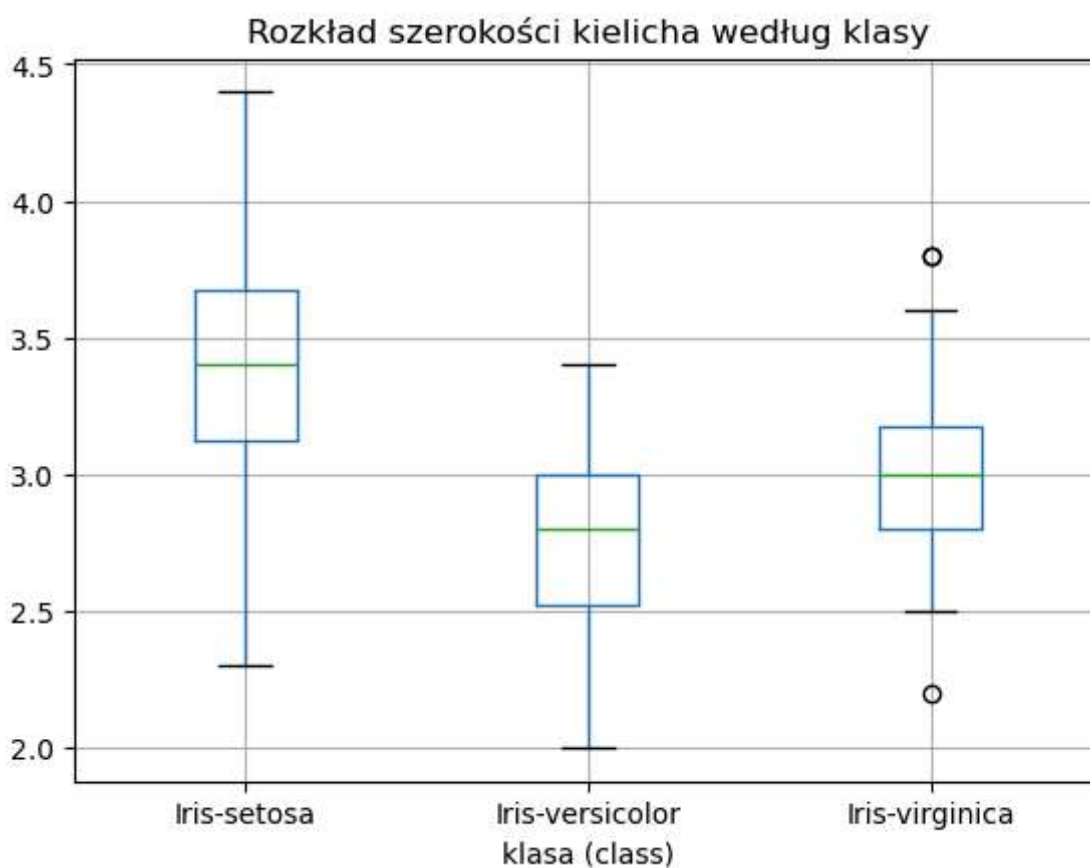
Analiza zależności pomiędzy zmiennymi

Długość kielicha, a klasa irysa



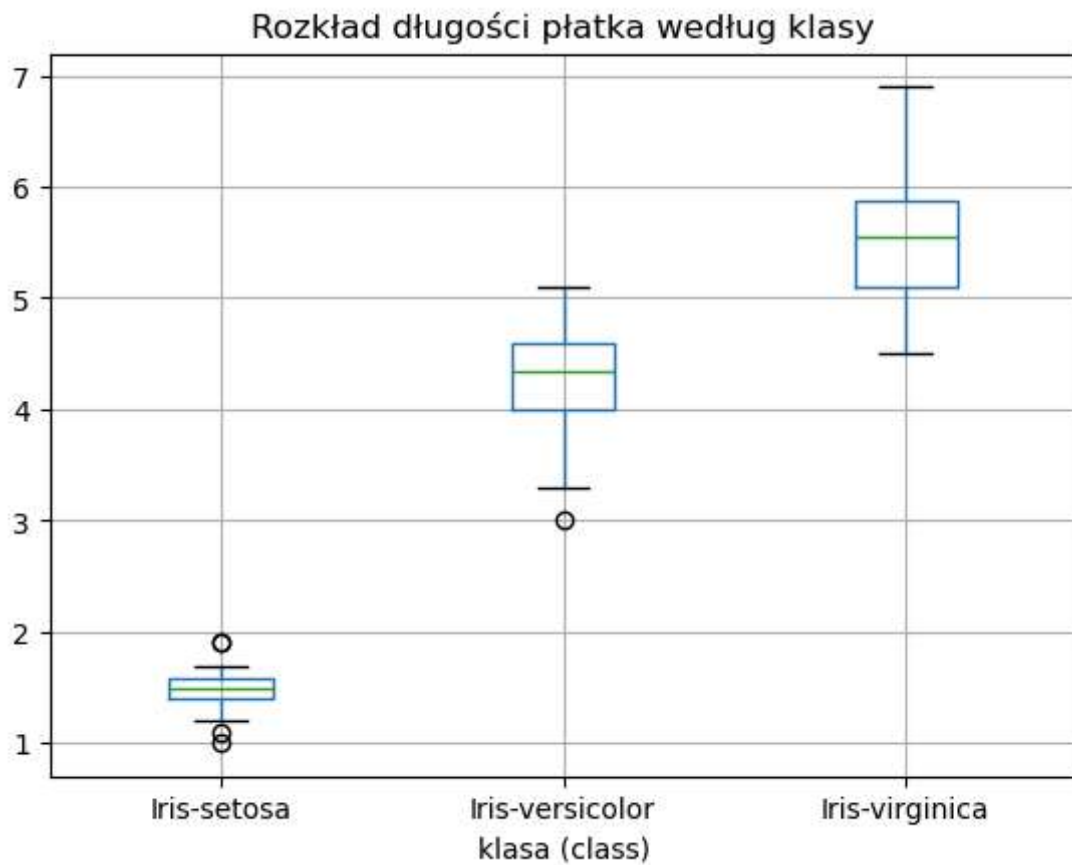
klasa (class)		długość kielicha (sepal length)					
		min	mean	median	max	std	count
0	Iris-setosa	4.300000	5.006000	5.000000	5.800000	0.352490	50
1	Iris-versicolor	4.900000	5.936000	5.900000	7.000000	0.516171	50
2	Iris-virginica	4.900000	6.588000	6.500000	7.900000	0.635880	50

Szerokość kielicha, a klasa irysa



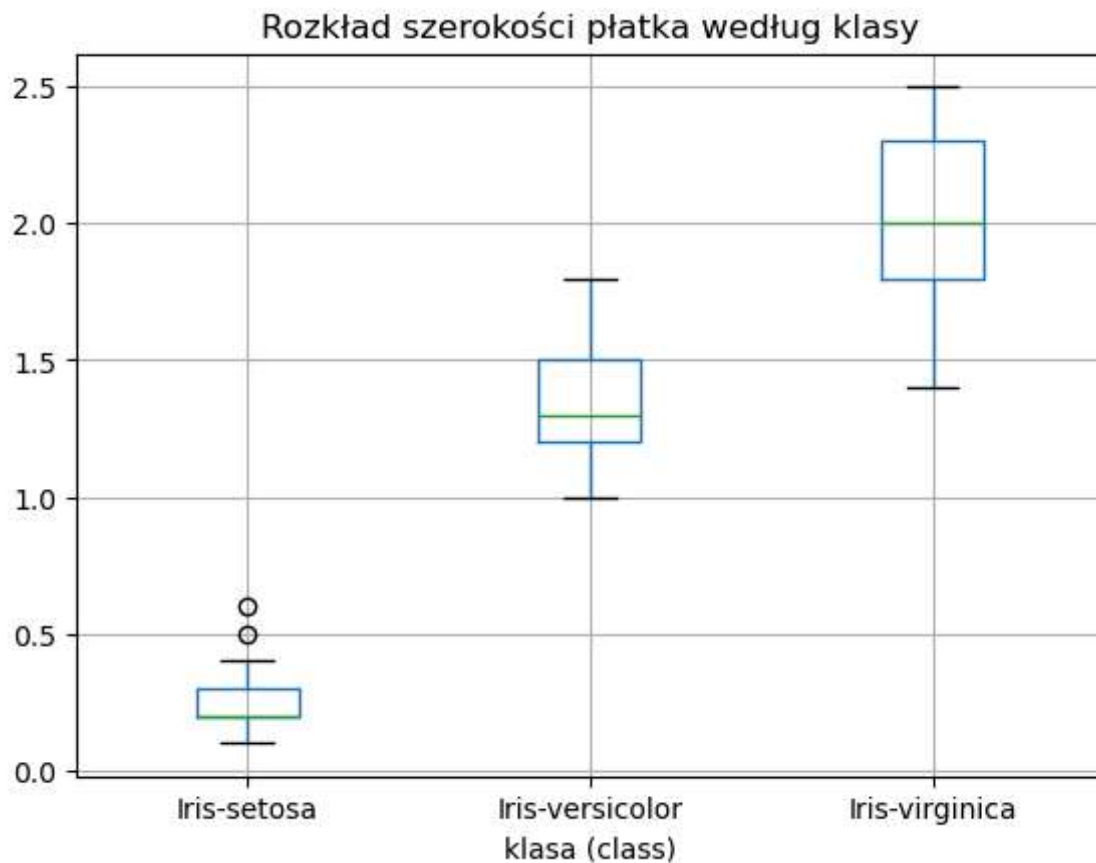
klasa (class)		szerokość kielicha (sepal width)					
		min	mean	median	max	std	count
0	Iris-setosa	2.300000	3.418000	3.400000	4.400000	0.381024	50
1	Iris-versicolor	2.000000	2.770000	2.800000	3.400000	0.313798	50
2	Iris-virginica	2.200000	2.974000	3.000000	3.800000	0.322497	50

Długość płatk, a klasa irysa



klasa (class)		długość płatka (petal length)					
		min	mean	median	max	std	count
0	Iris-setosa	1.000000	1.464000	1.500000	1.900000	0.173511	50
1	Iris-versicolor	3.000000	4.260000	4.350000	5.100000	0.469911	50
2	Iris-virginica	4.500000	5.552000	5.550000	6.900000	0.551895	50

Szerokość płatka, a klasa irysa



klasa (class)		szerokość płatka (petal width)					count
		min	mean	median	max	std	
0	Iris-setosa	0.100000	0.244000	0.200000	0.600000	0.107210	50
1	Iris-versicolor	1.000000	1.326000	1.300000	1.800000	0.197753	50
2	Iris-virginica	1.400000	2.026000	2.000000	2.500000	0.274650	50

Analiza zależności

	długość kielicha (sepal length)	szerokość kielicha (sepal width)	długość płatka (petal length)	szerokość płatka (petal width)
długość kielicha (sepal length)	1.000000	-0.109369	0.871754	0.817954
szerokość kielicha (sepal width)	-0.109369	1.000000	-0.420516	-0.356544
długość płatka (petal length)	0.871754	-0.420516	1.000000	0.962757
szerokość płatka (petal width)	0.817954	-0.356544	0.962757	1.000000

Wnioski

W wyniku analizy danych otrzymaliśmy następujące dane:

1. Iris-setosa mają:

- najkrótsze łodygi z badanych gatunków irysów
- najmniejszą rozpiętość długości łodyg
- łodygi o średniej długości 5cm
- najszersze kielichy z badanych gatunków irysów
- kielichy o średniej szerokości 3,4cm
- najkrótsze płatki z badanych gatunków irysów
- płatki o średniej długości cm
- najwęższe płatki z badanych gatunków irysów
- płatki o średniej szerokości 0,2cm
- najkrótsze kielichy z badanych gatunków irysów
- kielichy o średniej długości 5cm

2. Iris-versicolor mają:

- łodygi o średniej długości 5,9cm
- najwęższe kielichy z badanych gatunków irysów
- kielichy o średniej szerokości 2,8cm
- płatki o średniej długości 4,3cm
- płatki o średniej szerokości 1,3cm
- kielichy o średniej długości 5,9cm

3. Iris-virginica mają:

- najdłuższe łodygi z badanych gatunków irysów
- największą rozpiętość długości łodyg
- łodygi o średniej długości 6,5cm
- kielichy o średniej szerokości 3cm
- najdłuższe płatki z badanych gatunków irysów
- płatki o średniej długości 5,5cm
- najszersze płatki z badanych gatunków irysów
- płatki o średniej szerokości 2cm
- najdłuższe kielichy z badanych gatunków irysów
- kielichy o średniej długości 6,5cm o średniej długości 6,5cm

4. Szerokość płatka jest silnie skorelowana z jego długością. Wraz ze wzrostem szerokości płatka, wzrasta jego długość

5. Długość kielicha jest silnie skorelowana z długością i szerokością płatka. Im kielich dłuższy, tym płatek dłuższy i szerszy kielichy o średniej długości 6,5cm