

Irysy - analiza gatunków

O Danych

Zbiór danych zawiera informacje o trzech gatunkach irysów:

iris setosa



petal sepal

iris versicolor



petal sepal

iris virginica



petal sepal

O danych

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

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

1. Przegląd i analiza danych dotyczących Irysów

1.1 Przegląd losowych danych

1.1 Przegląd losowych danych

| | długość kielicha (sepal length) | szerokość kielicha (sepal width) | długość płatka (petal length) | szerokość płatka (petal width) | klasa (class) |
|-----|------------------------------------|-------------------------------------|----------------------------------|-----------------------------------|------------------|
| 128 | 6.4 | 2.8 | 5.6 | 2.1 | Iris-virginica |
| 72 | 6.3 | 2.5 | 4.9 | 1.5 | Iris-versicolor |
| 65 | 6.7 | 3.1 | 4.4 | 1.4 | Iris-versicolor |
| 23 | 5.1 | 3.3 | 1.7 | 0.5 | Iris-setosa |
| 119 | 6.0 | 2.2 | 5.0 | 1.5 | Iris-virginica |

1.1 Przegląd losowych danych

| | długość kielicha (sepal length) | szerokość kielicha (sepal width) | długość płatką (petal length) | szerokość płatką (petal width) | klasa (class) |
|-----|------------------------------------|-------------------------------------|----------------------------------|-----------------------------------|---------------------|
| 128 | 6.4 | 2.8 | 5.6 | 2.1 | Iris- virginica |
| 72 | 6.3 | 2.5 | 4.9 | 1.5 | Iris- versicolor |
| 65 | 6.7 | 3.1 | 4.4 | 1.4 | Iris- versicolor |
| 23 | 5.1 | 3.3 | 1.7 | 0.5 | Iris-setosa |
| 119 | 6.0 | 2.2 | 5.0 | 1.5 | Iris- virginica |

Na podstawie wyświetlonych wartości losowych widzimy, że mamy dane opisujące wielkości kielicha i płatków kwiatów z rodzaju lrys oraz rozróżnienie na ich gatunki.

1.2 Sprawdzenie wartości unikatowych.

1.2 Sprawdzenie wartości unikatowych.

| | |
|----------------------------------|----|
| 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 | |

1.2 Sprawdzenie wartości unikatowych.

```
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
```

```
klasa (class)
Iris-setosa          50
Iris-versicolor     50
Iris-virginica       50
Name: count, dtype: int64
```

1.2 Sprawdzenie wartości unikatowych.

```
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
```

```
klasa (class)
Iris-setosa      50
Iris-versicolor  50
Iris-virginica   50
Name: count, dtype: int64
```

Mamy dane na temat 3 gatunków lrysa - Setosa, Versicolor oraz Virginica. Dla każdego gatunku podane jest 50 wartości zawierających informacje o długości i szerokości kielicha oraz długości i szerokości płatka.

1.3 Przegląd danych statystycznych

1.3 Przegląd danych statystycznych

| | długość kielicha (sepal length) | | | | | szerokość kielicha (sepal width) | | | | |
|------------------------|---------------------------------|-------|--------|-----|----------|----------------------------------|-------|--------|-----|----------|
| | min | mean | median | max | std | min | mean | median | max | std |
| klasa (class) | | | | | | | | | | |
| Iris-setosa | 4.3 | 5.006 | 5.0 | 5.8 | 0.352490 | 2.3 | 3.418 | 3.4 | 4.4 | 0.381024 |
| Iris-versicolor | 4.9 | 5.936 | 5.9 | 7.0 | 0.516171 | 2.0 | 2.770 | 2.8 | 3.4 | 0.313798 |
| Iris-virginica | 4.9 | 6.588 | 6.5 | 7.9 | 0.635880 | 2.2 | 2.974 | 3.0 | 3.8 | 0.322497 |

1.3 Przegląd danych statystycznych

| | długość kielicha (sepal length) | | | | | szerokość kielicha (sepal width) | | | | |
|-----------------|---------------------------------|-------|--------|-----|----------|----------------------------------|-------|--------|-----|----------|
| | min | mean | median | max | std | min | mean | median | max | std |
| klasa (class) | | | | | | | | | | |
| Iris-setosa | 4.3 | 5.006 | 5.0 | 5.8 | 0.352490 | 2.3 | 3.418 | 3.4 | 4.4 | 0.381024 |
| Iris-versicolor | 4.9 | 5.936 | 5.9 | 7.0 | 0.516171 | 2.0 | 2.770 | 2.8 | 3.4 | 0.313798 |
| Iris-virginica | 4.9 | 6.588 | 6.5 | 7.9 | 0.635880 | 2.2 | 2.974 | 3.0 | 3.8 | 0.322497 |

| | długość płatka (petal length) | | | | | szerokość płatka (petal width) | | | | |
|-----------------|-------------------------------|-------|--------|-----|----------|--------------------------------|-------|--------|-----|----------|
| | min | mean | median | max | std | min | mean | median | max | std |
| klasa (class) | | | | | | | | | | |
| Iris-setosa | 1.0 | 1.464 | 1.50 | 1.9 | 0.173511 | 0.1 | 0.244 | 0.2 | 0.6 | 0.107210 |
| Iris-versicolor | 3.0 | 4.260 | 4.35 | 5.1 | 0.469911 | 1.0 | 1.326 | 1.3 | 1.8 | 0.197753 |
| Iris-virginica | 4.5 | 5.552 | 5.55 | 6.9 | 0.551895 | 1.4 | 2.026 | 2.0 | 2.5 | 0.274650 |

1.4 Zmiana nazw kolumn

1.4 Zmiana nazw kolumn

Zmieniam nazwy kolumn, na polskie aby lepiej się pracowało nad analizą oraz zwiększyć przejrzystość

2. Sprawdzenie kompletności danych

2. Sprawdzenie kompletności danych

```
długość kielicha      0
szerokość kielicha    0
długość płatka        0
szerokość płatka      0
gatunek               0
dtype: int64
```

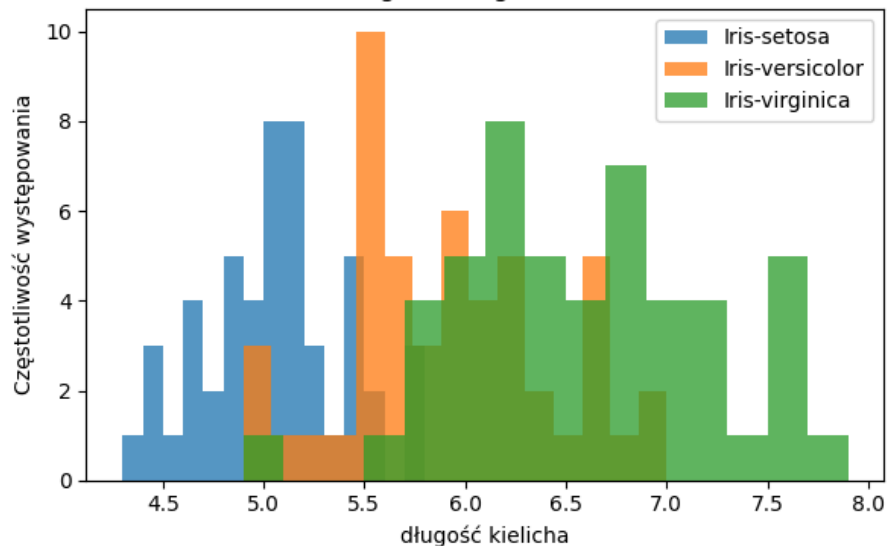
2. Sprawdzenie kompletności danych

```
długość kielicha      0
szerokość kielicha    0
długość płatka        0
szerokość płatka      0
gatunek               0
dtype: int64
```

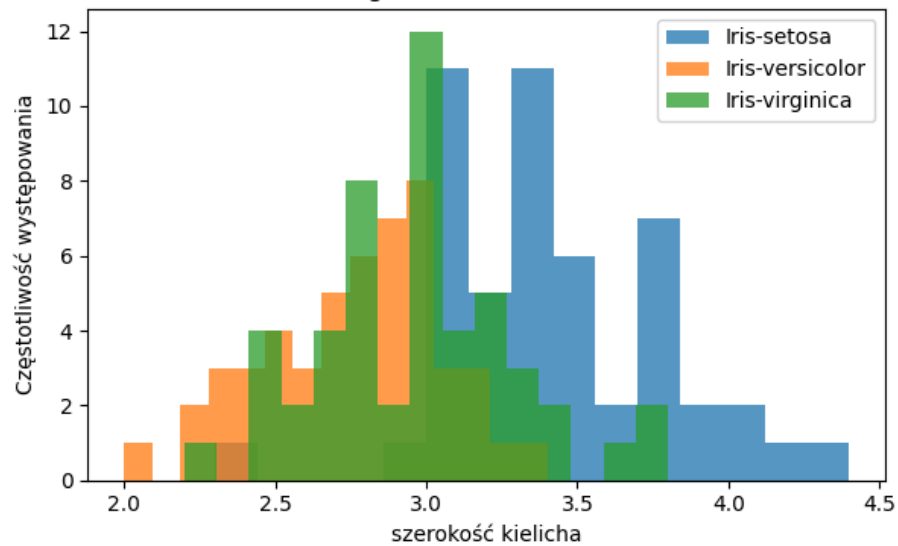
Posiadamy kompletne dane dla wszystkich przedstawionych gatunków kwiatów - brak pustych wartości.

3. Analiza danych poszczególnych danych

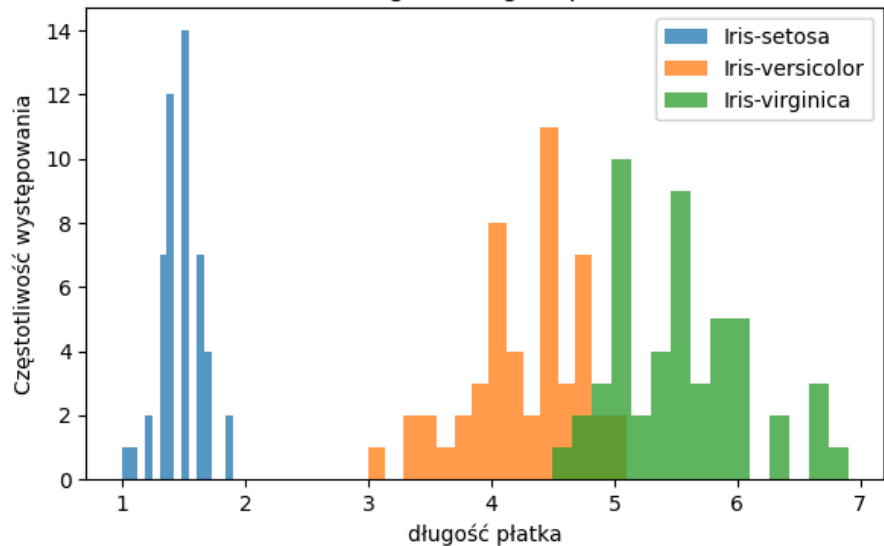
Histogram długość kielicha



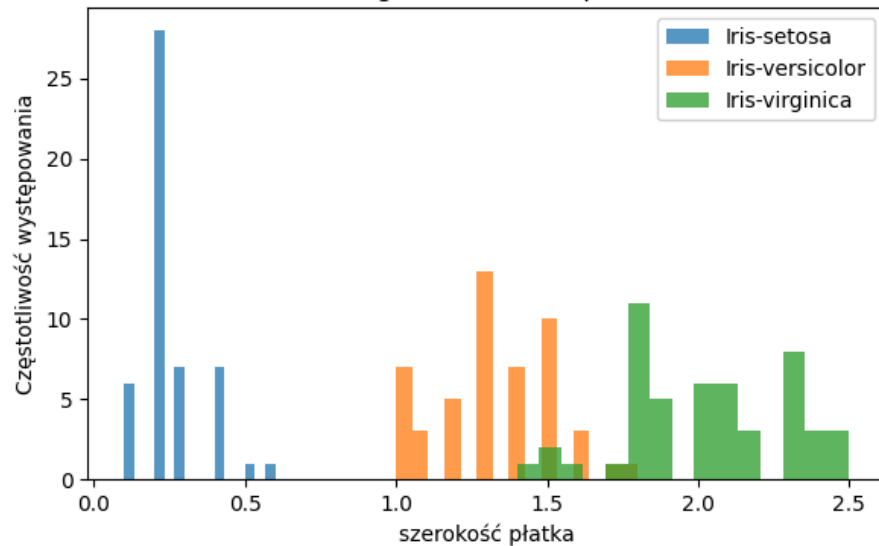
Histogram szerokość kielicha



Histogram długość płatka



Histogram szerokość płatka



Na podstawie histogramu możemy zauważyć, że gatunek *Iris-versicolor* oraz *Iris-virginica* mają zbliżoną budowę kielicha, jednak w przypadku *Iris-versicolor* więcej obserwacji wykazuje mniejszą długość kielicha, natomiast *Iris-virginica* cechuje większa rozpiętość w długości kielicha. Gatunek *Iris-setosa* posiada znacznie krótszy, ale szerszy kielich od pozostałych dwóch gatunków.

Obserwacje długości i szerokości płątka wykazują, że gatunek *Iris-virginica* posiada najkrótsze i najwęższe płatki, *Iris-setosa* najdłuższe i najszerze płatki, natomiast *Iris-versicolor* pod względem budowy płatków znajduje się pomiędzy dwoma powyższymi gatunkami.

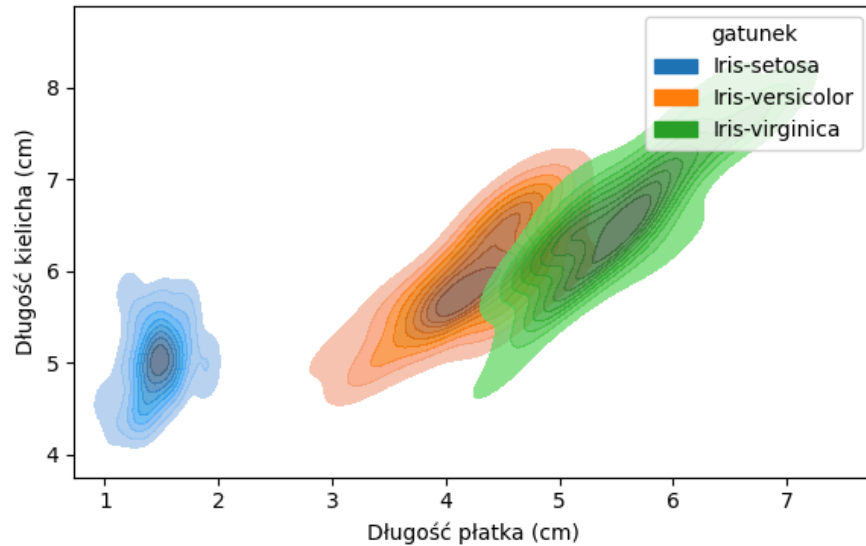
4. Naprawa brakujących wartości

4. Naprawa brakujących wartości

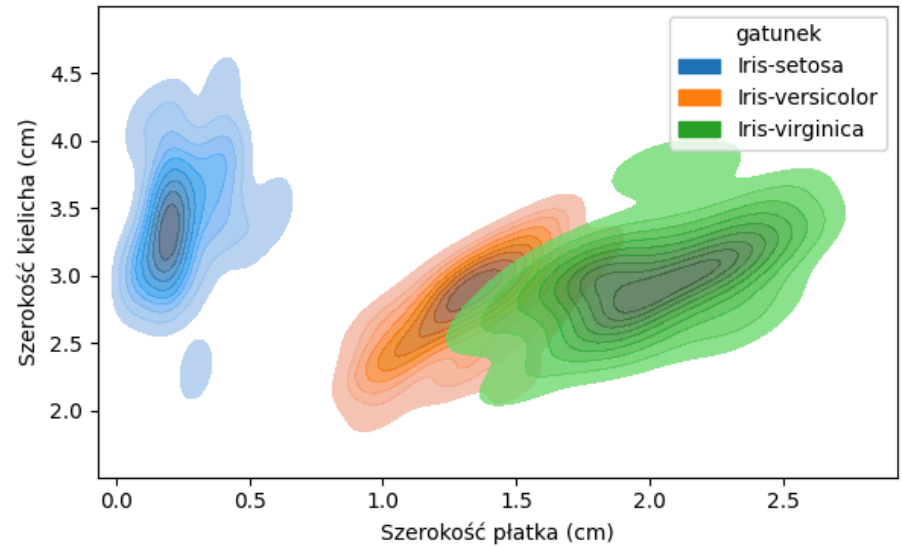
W tym przypadku mamy kompletne dane, więc pomijamy ten krok.

5. Analiza relacji między zmiennymi.

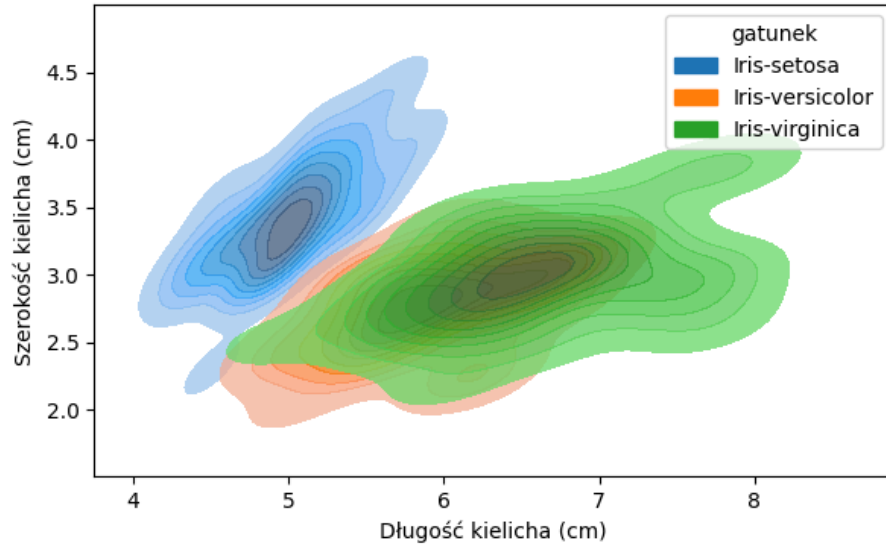
Długość płatka i kielicha według klasy



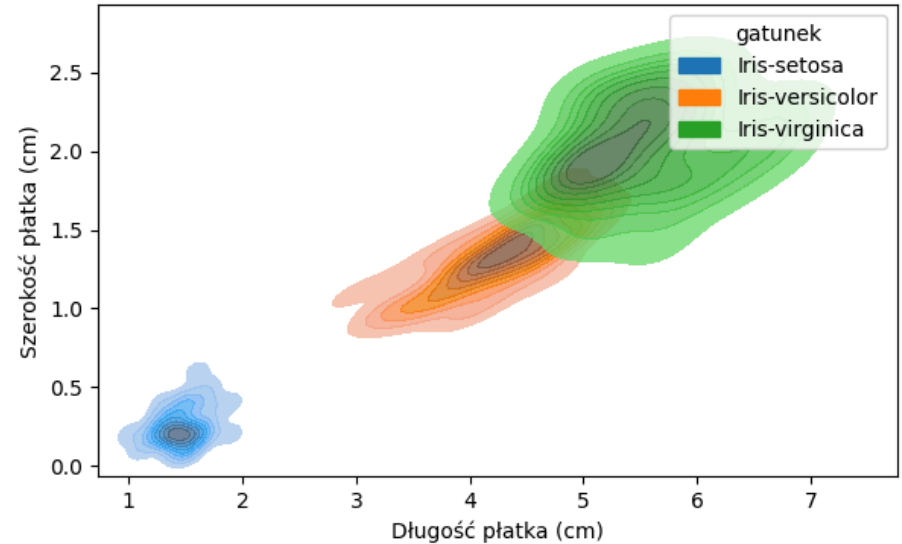
Szerokość płatka i kielicha według klasy



Długość i szerokość kielicha według klasy



Długość i szerokość płatka według klasy

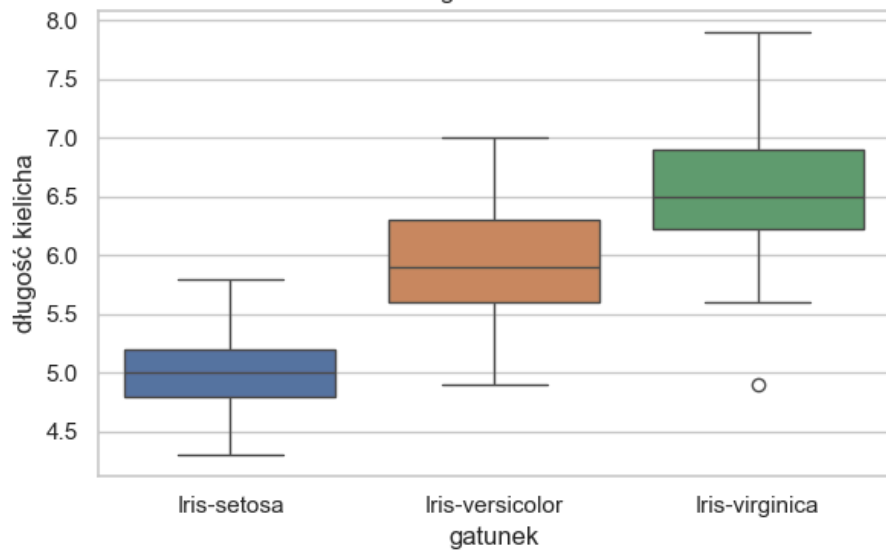


Na podstawie powyższych wykresów można stwierdzić, że dla gatunków *Iris-versicolor* i *Iris-virginica* zachodzą relacje pomiędzy długościami i szerokościami zarówno płatków jak i kielicha, są to zależności rosnące - im szerszy, tym dłuższy.

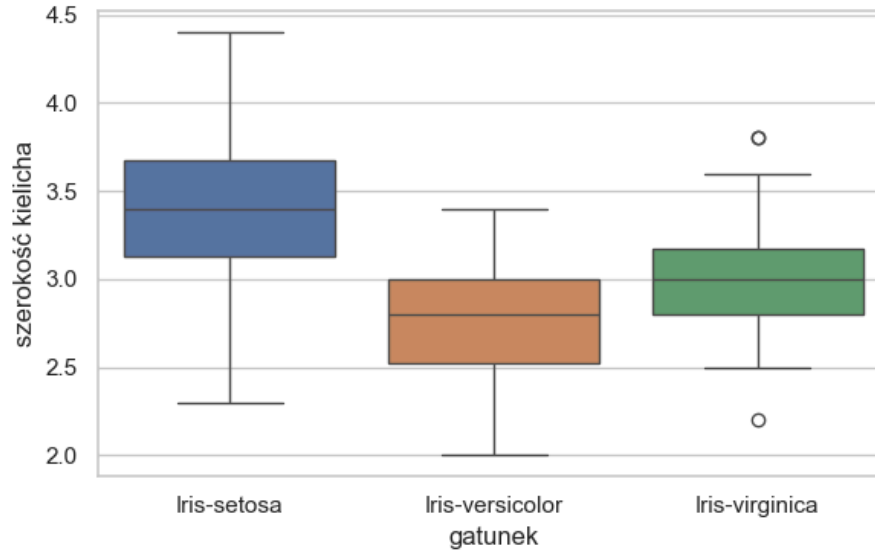
Iris-setosa wykazuje niewielką zależność długości od szerokości.

6. Wartości odstające.

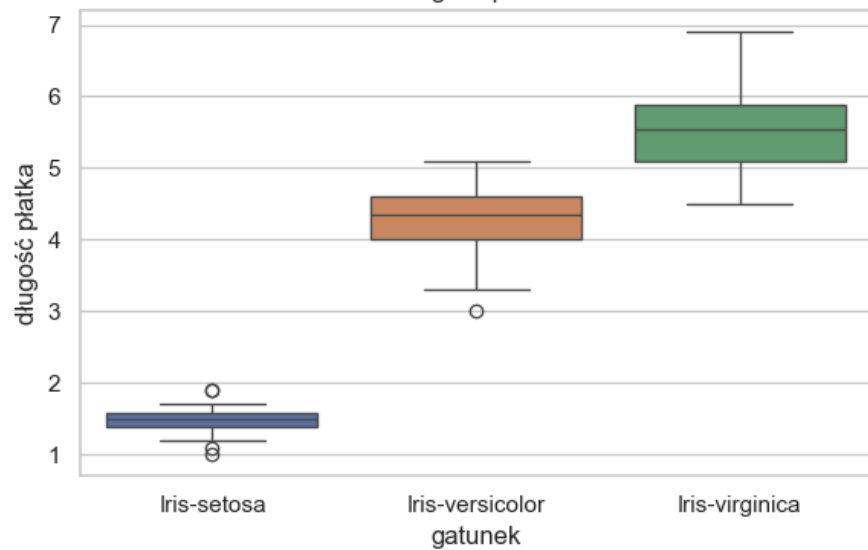
Długość kielicha



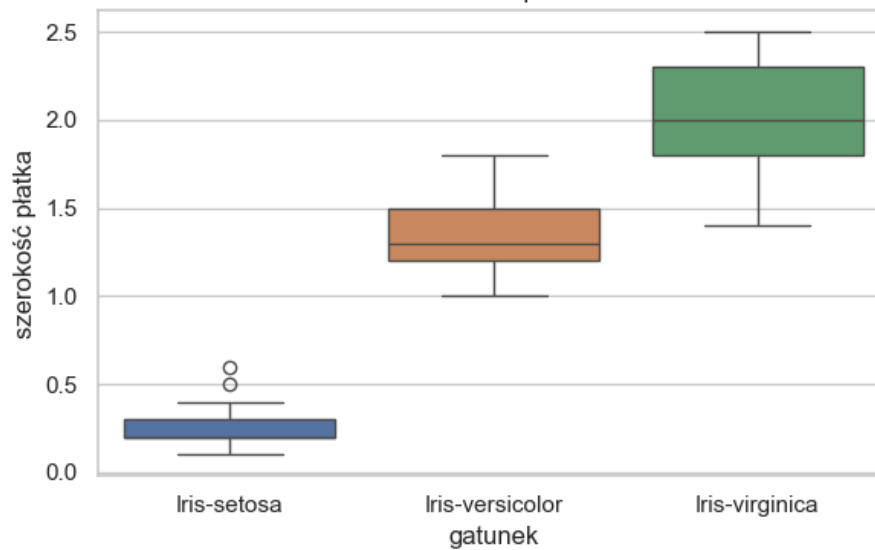
Szerokość kielicha



Długość płatka



Szerokość płatka



Występuję niewielka ilość danych odstających, która nie powinna wpłynąć znacząco na ogólne wykorzystanie danych.

Podsumowanie

- Dla każdego kwiatu zmierzono cztery cechy: długość i szerokość kielicha oraz długość i szerokość płatka (wszystko w centymetrach).
- Dane są kompletne – nie występują żadne brakujące wartości.
- Analiza statystyczna i wizualizacje (histogramy) pokazują, że poszczególne gatunki różnią się budową kielicha i płatków:
 - ■ Iris-setosa ma krótszy, ale szerszy kielich w porównaniu do pozostałych gatunków.
 - ■ Iris-virginica cechuje się większą rozpiętością długości kielicha.
 - ■ Iris-versicolor pod względem budowy płatków i kielicha plasuje się pomiędzy pozostałymi gatunkami.
- Występują niewielkie ilości wartości odstających, które nie mają istotnego wpływu na całość danych.
- Między niektórymi cechami (np. długość i szerokość płatka/kielicha) występują zależności – szczególnie dla gatunków versicolor i virginica.

Podsumowując, zebrane dane są kompletne i dobrze opisują różnice między gatunkami irysów, co czyni je użytecznymi do testowania algorytmów klasyfikacyjnych oraz nauki o danych.

