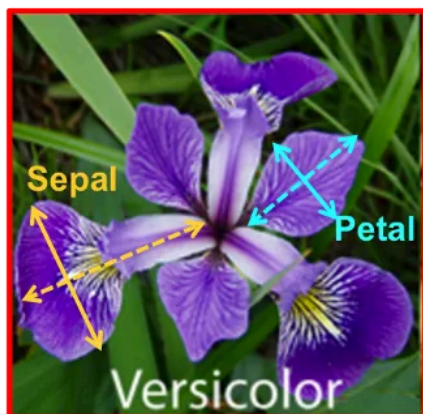


EDA Zestawu danych - Iris



1) Wstępna inspekcja danych:

Wyświetlenie losowych wierszy

	długość kielicha (sepal length)	szerokość kielicha (sepal width)	długość płatka (petal length)	szerokość płatka (petal width)	klasa (class)
95	5.7	3.0	4.2	1.2	Iris-versicolor
66	5.6	3.0	4.5	1.5	Iris-versicolor
82	5.8	2.7	3.9	1.2	Iris-versicolor
18	5.7	3.8	1.7	0.3	Iris-setosa
111	6.4	2.7	5.3	1.9	Iris-virginica
25	5.0	3.0	1.6	0.2	Iris-setosa
116	6.5	3.0	5.5	1.8	Iris-virginica
67	5.8	2.7	4.1	1.0	Iris-versicolor
32	5.2	4.1	1.5	0.1	Iris-setosa
2	4.7	3.2	1.3	0.2	Iris-setosa
57	4.9	2.4	3.3	1.0	Iris-versicolor
106	4.9	2.5	4.5	1.7	Iris-virginica
91	6.1	3.0	4.6	1.4	Iris-versicolor
87	6.3	2.3	4.4	1.3	Iris-versicolor
107	7.3	2.9	6.3	1.8	Iris-virginica
127	6.1	3.0	4.9	1.8	Iris-virginica
58	6.6	2.9	4.6	1.3	Iris-versicolor
47	4.6	3.2	1.4	0.2	Iris-setosa
33	5.5	4.2	1.4	0.2	Iris-setosa
51	6.4	3.2	4.5	1.5	Iris-versicolor
86	6.7	3.1	4.7	1.5	Iris-versicolor
133	6.3	2.8	5.1	1.5	Iris-virginica
48	5.3	3.7	1.5	0.2	Iris-setosa
56	6.3	3.3	4.7	1.6	Iris-versicolor
101	5.8	2.7	5.1	1.9	Iris-virginica
119	6.0	2.2	5.0	1.5	Iris-virginica
88	5.6	3.0	4.1	1.3	Iris-versicolor
54	6.5	2.8	4.6	1.5	Iris-versicolor
10	5.4	3.7	1.5	0.2	Iris-setosa
6	4.6	3.4	1.4	0.3	Iris-setosa

2) Czyszczenie danych:

Unikatowe wartości & typy danych & brakujące wartości

```
Długość_Kielicha      35
Szerokość_Kielicha     23
Długość_Płatka        43
Szerokość_Płatka      22
Odmiana                3
dtype: int64

Długość_Kielicha      float64
Szerokość_Kielicha     float64
Długość_Płatka        float64
Szerokość_Płatka      float64
Odmiana               object
dtype: object

Długość_Kielicha      0
Szerokość_Kielicha     0
Długość_Płatka        0
Szerokość_Płatka      0
Odmiana               0
dtype: int64
```

	Długość_Kielicha	Szerokość_Kielicha	Długość_Płatka	Szerokość_Płatka	Odmiana
34	4.9	3.1	1.5	0.1	Iris-setosa
37	4.9	3.1	1.5	0.1	Iris-setosa
142	5.8	2.7	5.1	1.9	Iris-virginica

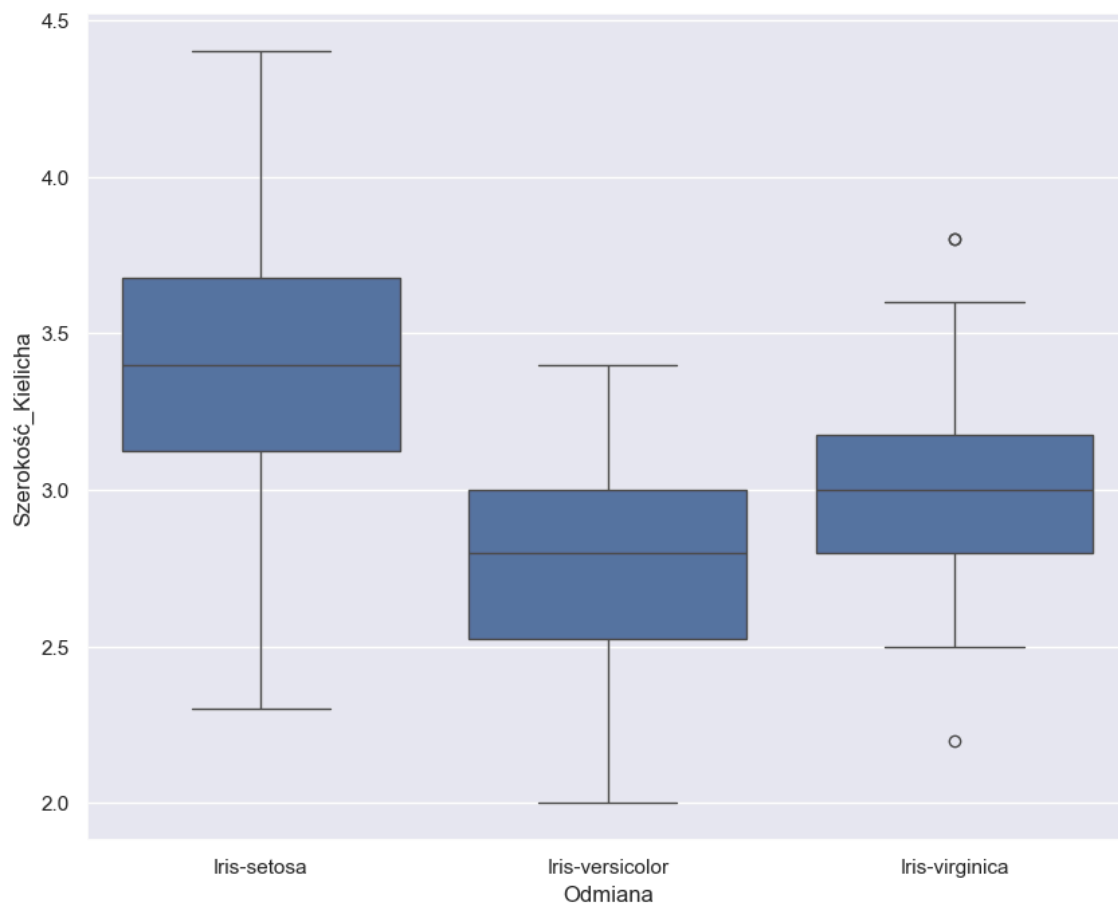
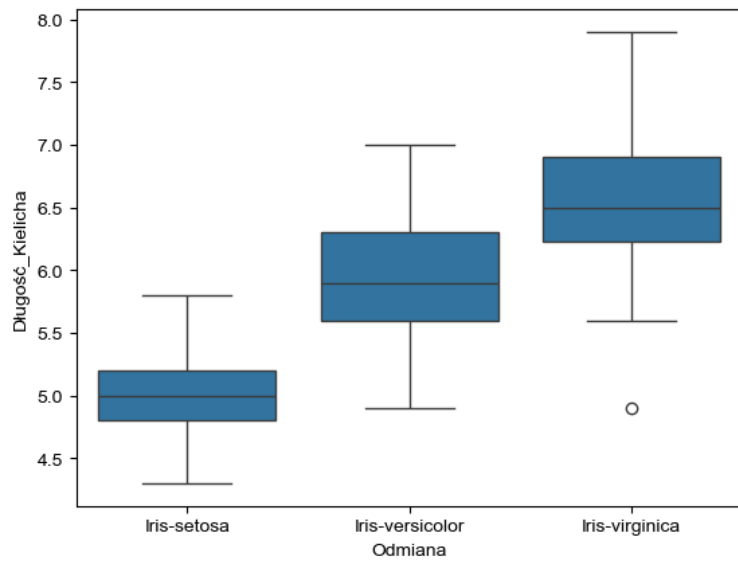
3) Analiza jednowymiarowa:

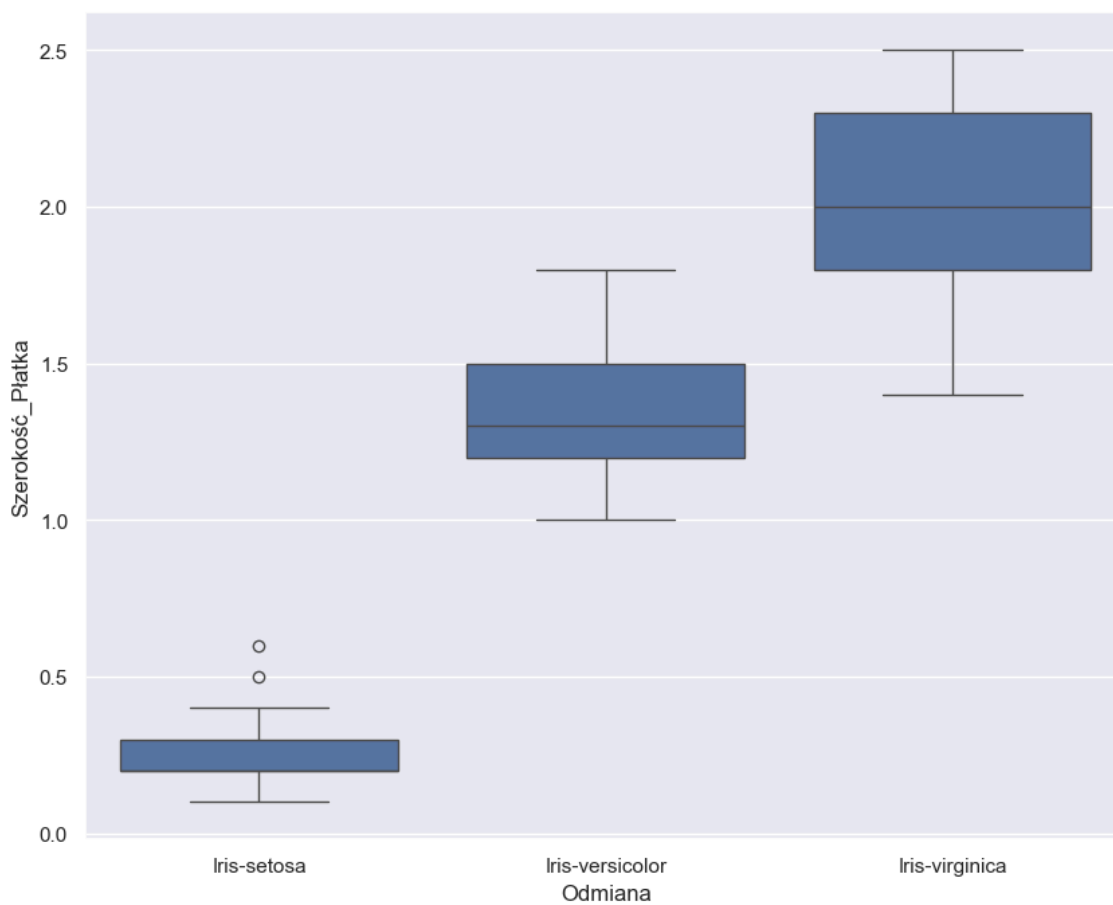
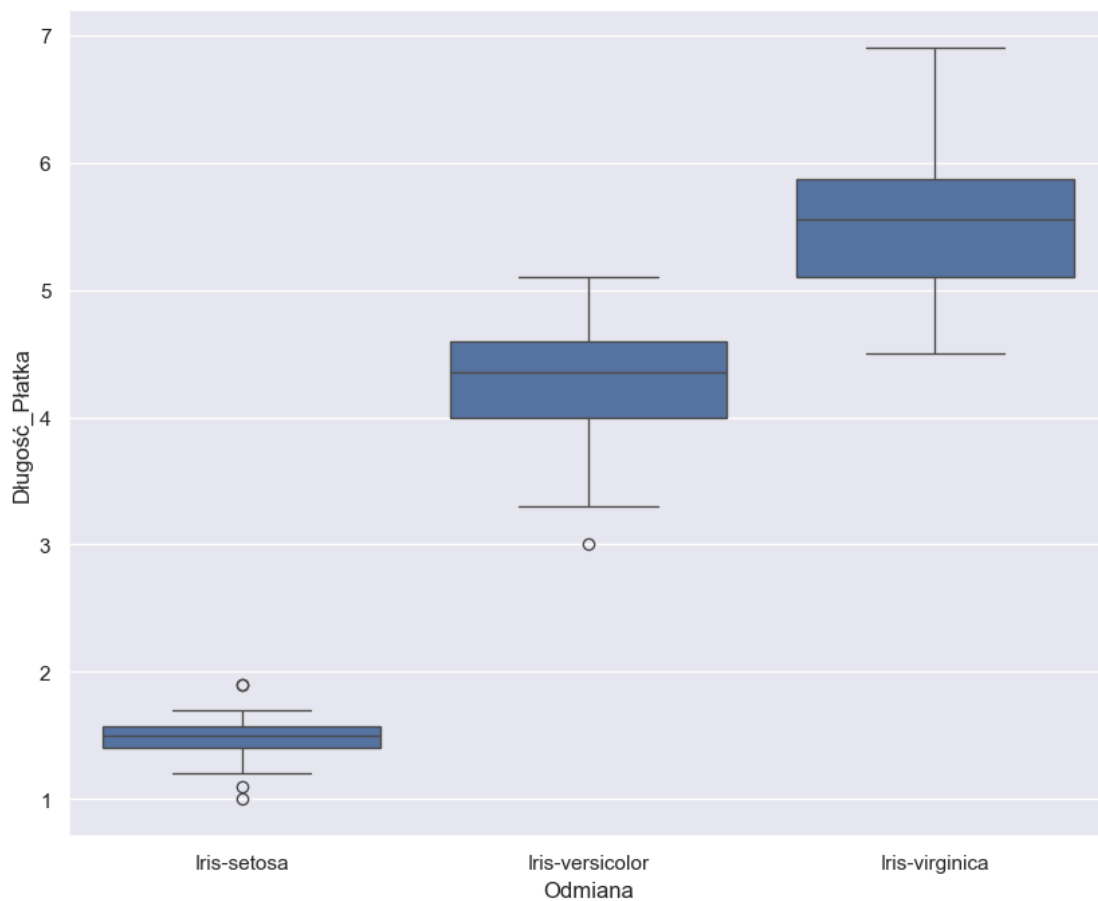
Przegląd kolumn numerycznych & Szczegółowa analiza odmian Irysa

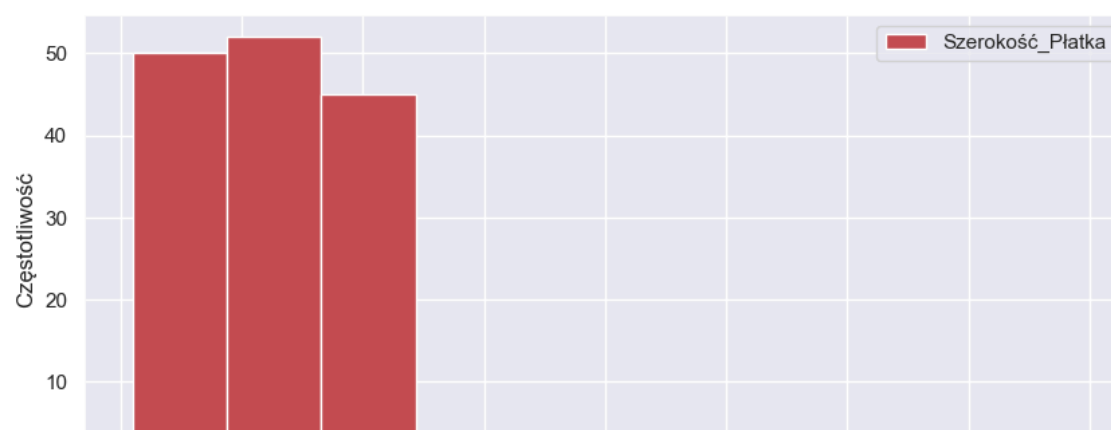
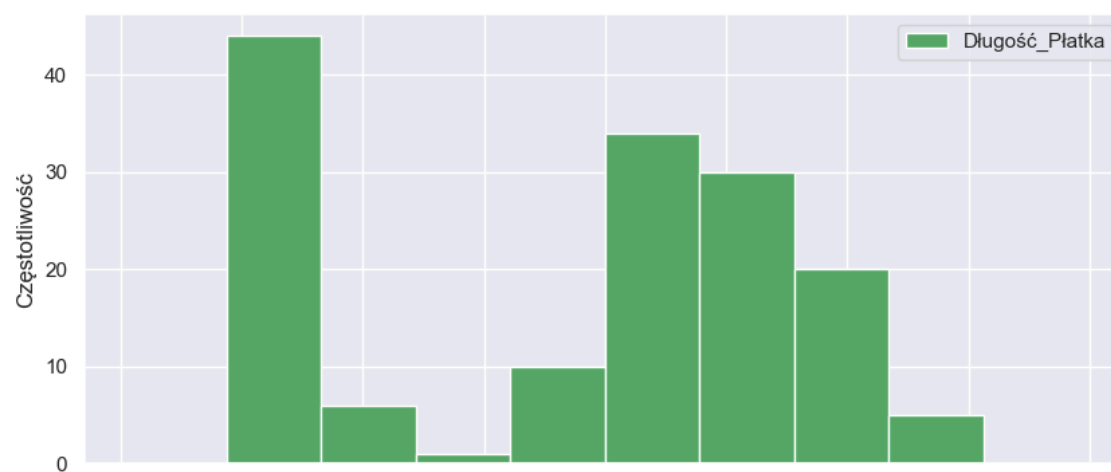
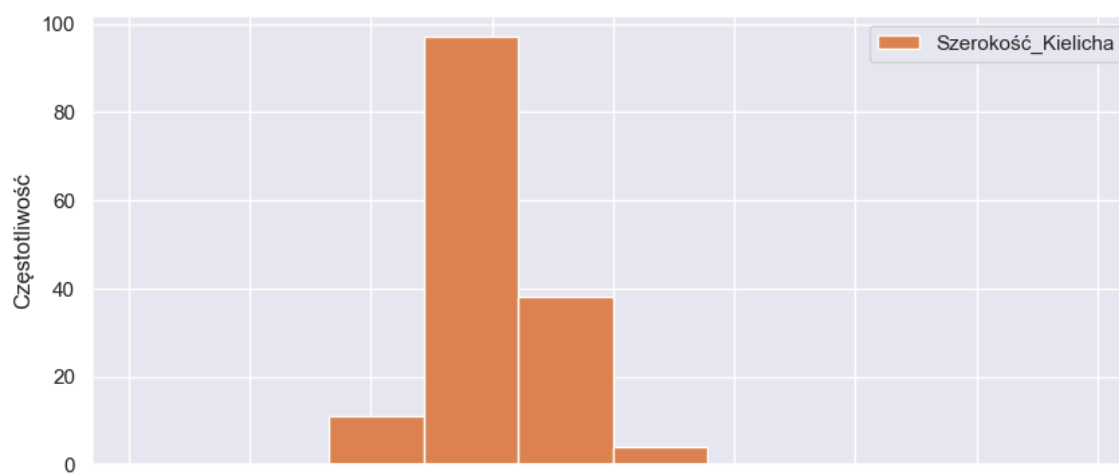
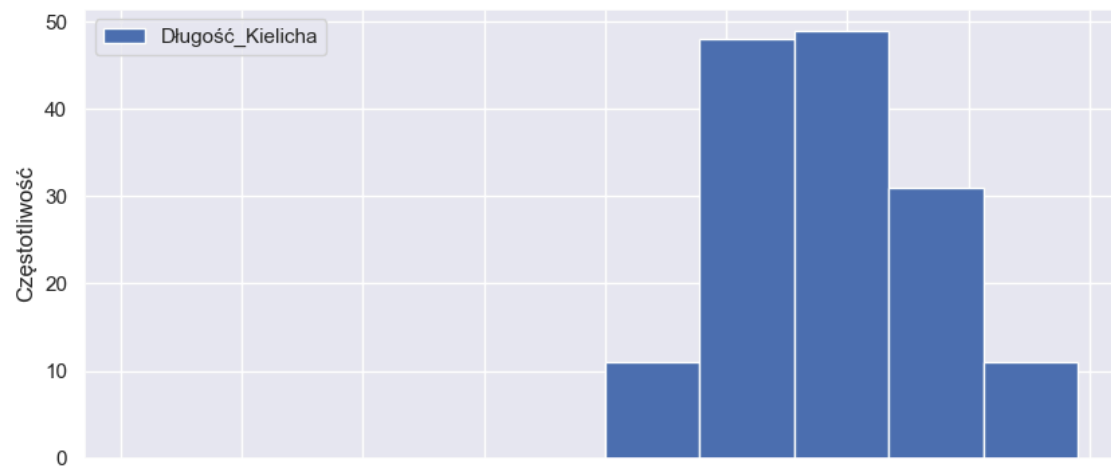
	Długość_Kielicha	Szerokość_Kielicha	Długość_Płatka	Szerokość_Płatka
count	150.000000	150.000000	150.000000	150.000000
mean	5.843333	3.054000	3.758667	1.198667
std	0.828066	0.433594	1.764420	0.763161
min	4.300000	2.000000	1.000000	0.100000
25%	5.100000	2.800000	1.600000	0.300000
50%	5.800000	3.000000	4.350000	1.300000
75%	6.400000	3.300000	5.100000	1.800000
max	7.900000	4.400000	6.900000	2.500000

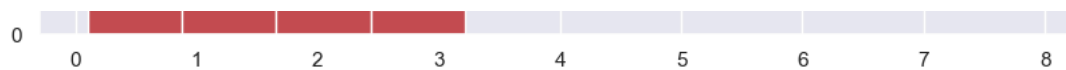
	Odmiana	Długość_Kielicha				Szerokość_Kielicha				Długość_Płatka				Szerokość_Płatka			
		min	max	mean	median	min	max	mean	median	min	max	mean	median	min	max	mean	median
0	Iris-setosa	4.3	5.8	5.006	5.0	2.3	4.4	3.418	3.4	1.0	1.9	1.464	1.50	0.1	0.6	0.244	0.2
1	Iris-versicolor	4.9	7.0	5.936	5.9	2.0	3.4	2.770	2.8	3.0	5.1	4.260	4.35	1.0	1.8	1.326	1.3
2	Iris-virginica	4.9	7.9	6.588	6.5	2.2	3.8	2.974	3.0	4.5	6.9	5.552	5.55	1.4	2.5	2.026	2.0

Wykresy

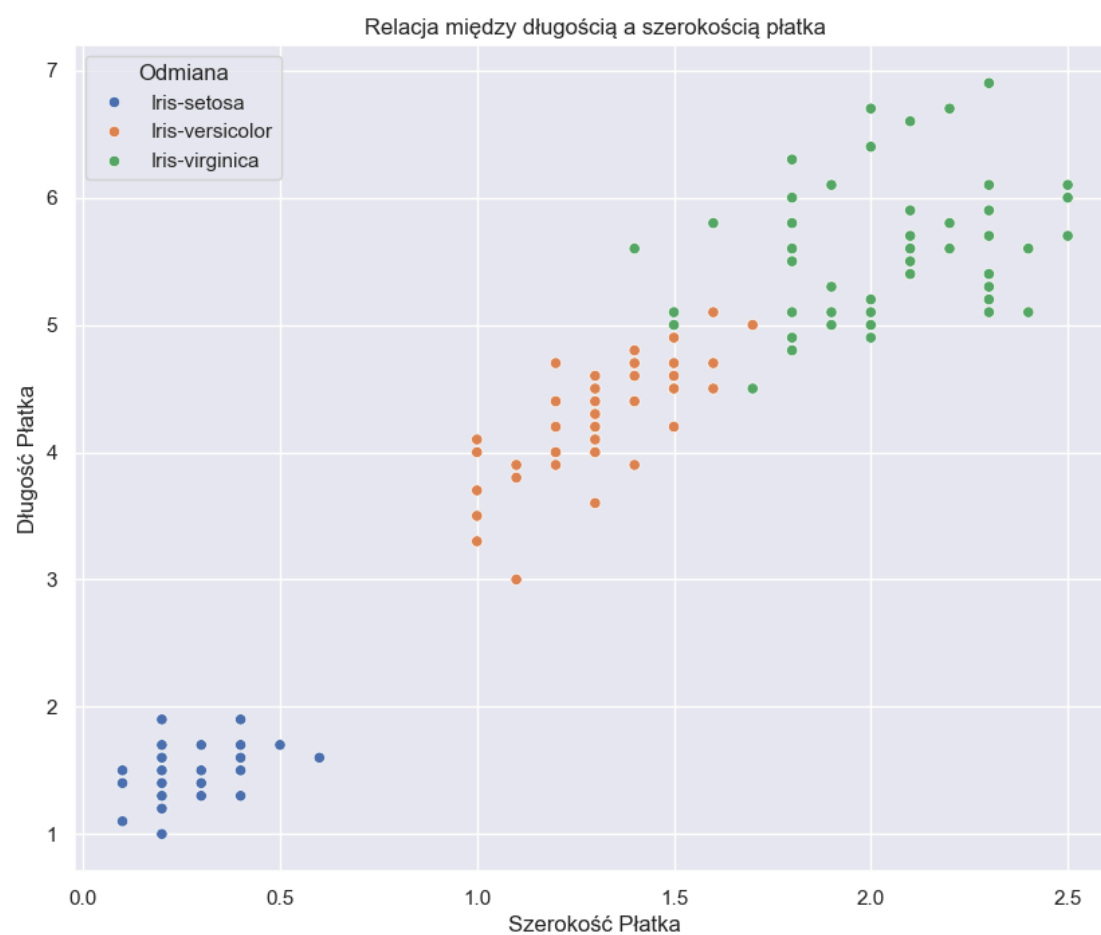




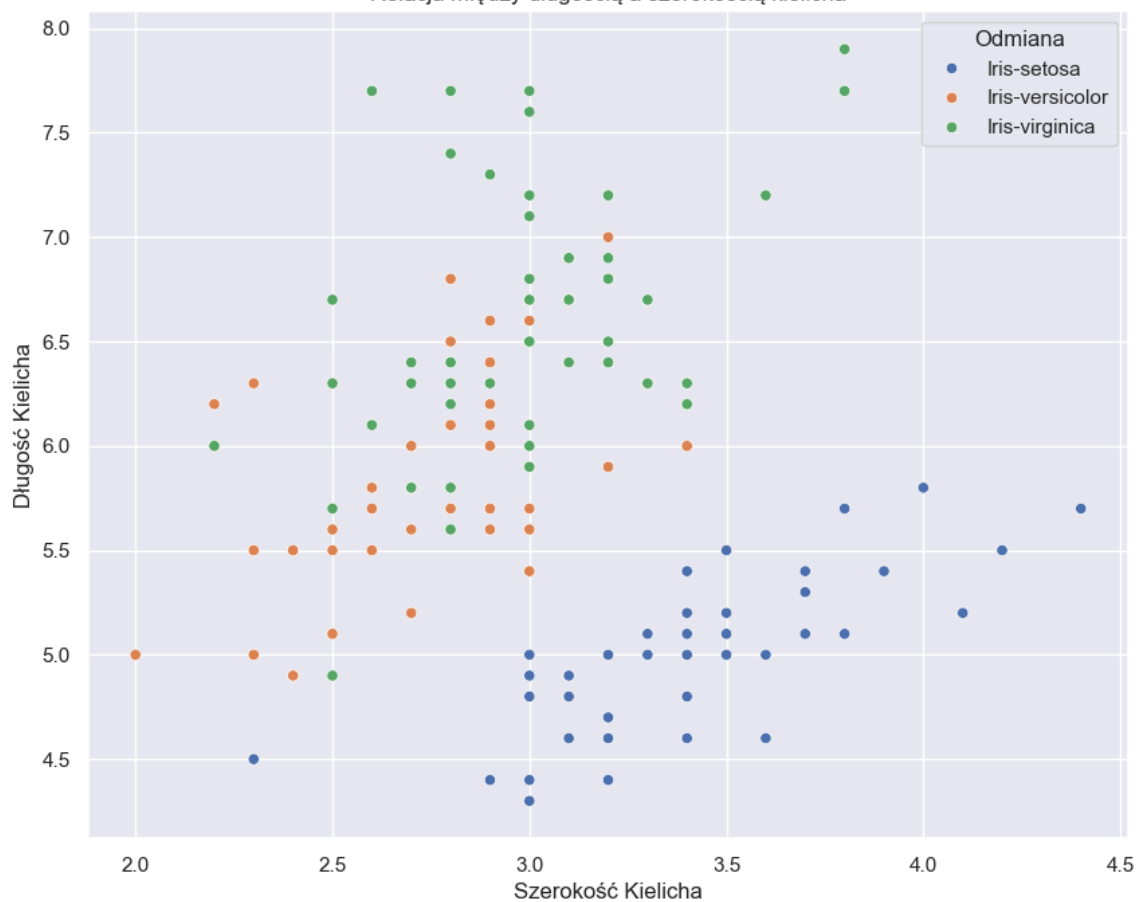


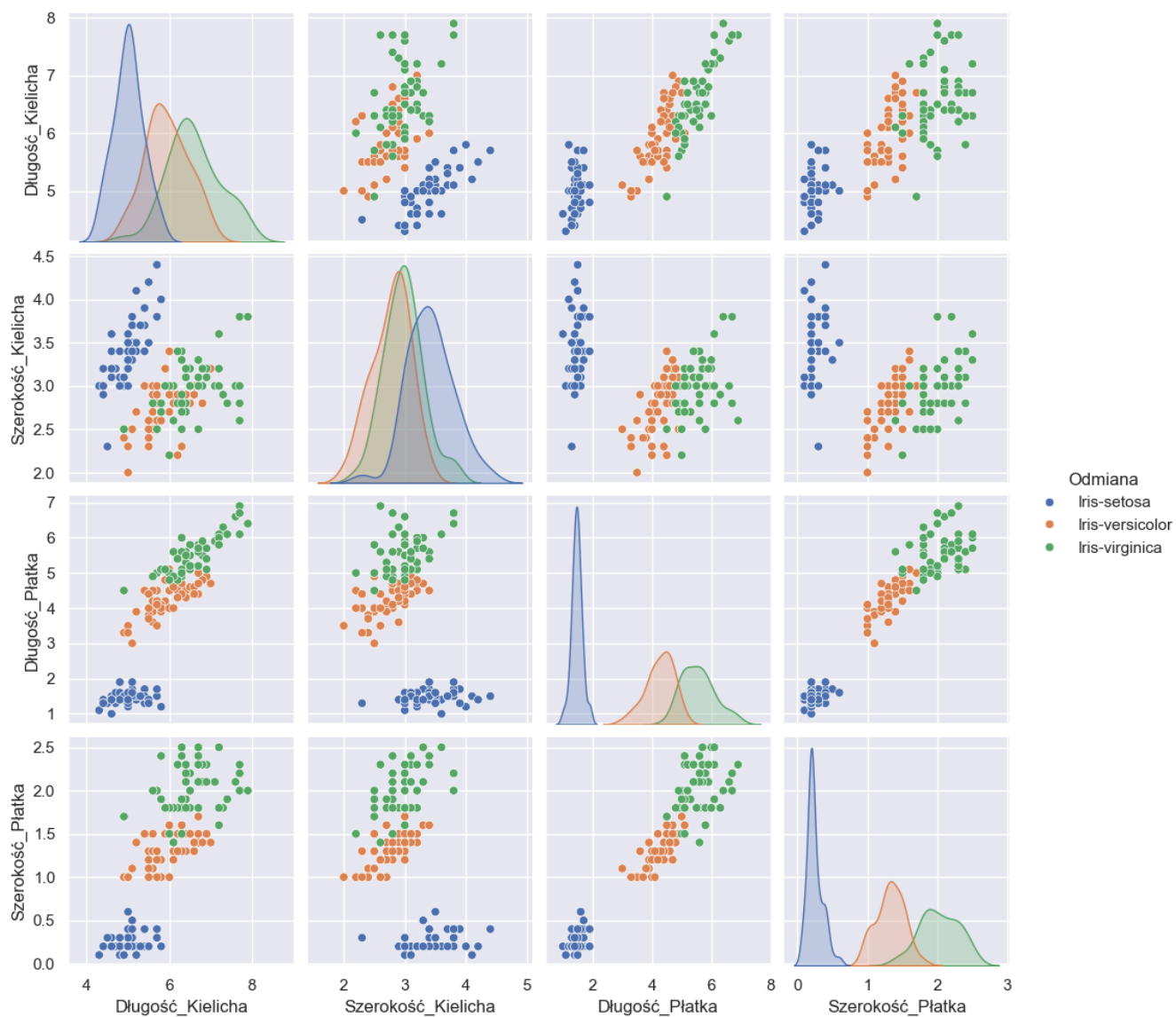


4) Analiza dwuwymiarowa:



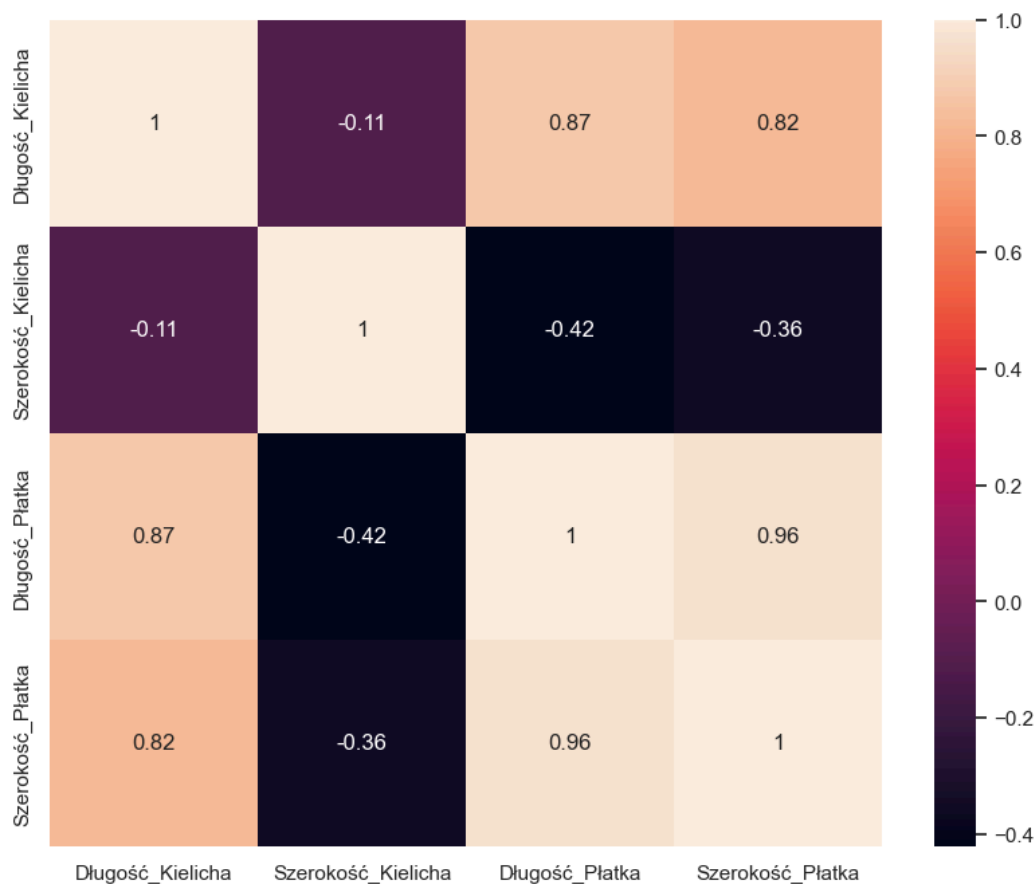
Relacja między długością a szerokością kielicha





Korelacja między danymi

	Długość_Kielicha	Szerokość_Kielicha	Długość_Płatka	Szerokość_Płatka
Długość_Kielicha	1.000000	-0.109369	0.871754	0.817954
Szerokość_Kielicha	-0.109369	1.000000	-0.420516	-0.356544
Długość_Płatka	0.871754	-0.420516	1.000000	0.962757
Szerokość_Płatka	0.817954	-0.356544	0.962757	1.000000



Wnioski końcowe:

Na podstawie przeprowadzonej analizy danych dotyczącej zbioru danych

Iris - można wyciągnąć następujące wnioski:

1) Struktura Danych:

- Zbiór danych jest kompletny i składa się z 150 wierszy oraz 5 kolumn. Każda odmiana Irysa zawiera 50 rekordów, co zapewnia równowagę w analizie danych.

2) Charakterystyka odmian:

- Setosa: Jest najmniejsza spośród trzech odmian. Średnia długość kielicha wynosi 5cm, a szerokość 3.4cm. Płatki tego gatunku charakteryzują się średnią długością 1.5cm i szerokością 0.2cm.
- * Versicolor: Stanowi średnią odmianę. Średnia długość kielicha to 5.9cm, a szerokość 2.8cm. Płatki mają średnią długość 4.35cm i szerokość 1.3cm.
- Virginica: Jest największą odmianą. Średnia długość jej kielicha to 6.5cm, z szerokością 3cm. Płatki tej odmiany mają średnią długość 5.55cm i szerokość 2cm.

3) Granice Między Gatunkami:

- Analiza wykazała, że już z użyciem zaledwie 2-3 wykresów można jednoznacznie określić różnice i granice między badanymi gatunkami Irysa. Każda odmiana ujawnia unikalne wzorce w wymiarach kielichów i płatków.

4) Zastosowanie Modelu Ucznia Maszynowego:

- Wyraźnie zdefiniowane granice między odmianami sugerują, że model uczenia maszynowego z wysokim prawdopodobieństwem będzie w stanie poprawnie nauczyć się identyfikować gatunki

na podstawie wymiarów, co może być praktyczne w klasyfikacji nowych próbek.

5) Zastosowania Praktyczne:

- Wyciągnięte wzorce i granice mogą mieć zastosowanie w klasyfikacji i identyfikacji gatunkowej na podstawie morfologii, co jest wartościowe w kontekście badań botanicznych oraz zastosowań przemysłowych.

Podsumowując, analiza EDA dostarcza istotnych wniosków na temat struktury i różnic między odmianami kwiatu Irys, co tworzy solidną podstawę do dalszych badań i rozwijania modeli predykcyjnych.