

## IRIS dataset

<http://archive.ics.uci.edu/dataset/53/iris>

Variable Name	Role	Type	Description	Units	Missing Values
sepal length	Feature	Continuous		cm	no
sepal width	Feature	Continuous		cm	no
petal length	Feature	Continuous		cm	no
petal width	Feature	Continuous		cm	no
class	Target	Categorical	class of iris plant: Iris Setosa, Iris Versicolour, or Iris Virginica		no

## Ćwiczenie 1: Import i Przegląd Danych

Zadanie:

1. Importuj bibliotekę `pandas`.
2. Wczytaj zbiór danych Iris, dostępny jako wbudowany zbiór danych w bibliotece `sklearn`.
3. Wyświetl pierwsze kilka wierszy zbioru danych, aby zapoznać się z jego strukturą.

Kod:

```
import pandas as pd
from sklearn.datasets import load_iris

# Wczytanie danych
iris = load_iris()
df = pd.DataFrame(iris.data, columns=iris.feature_names)
df['target'] = iris.target

# Wyświetlenie pierwszych wierszy
print(df.head())
```

## Ćwiczenie 2: Oczyszczanie Danych

Zadanie:

1. Sprawdź, czy istnieją brakujące dane w zbiorze.

2. Jeśli istnieją brakujące dane, zdecyduj, jak je obsłużyć: usuń wiersze lub uzupełnij brakujące dane.
3. Jakie wnioski można wyciągnąć z ćwiczenia?

**Kod:**

```
# Sprawdzenie brakujących danych
missing_data = df.isnull().sum()
print("Brakujące dane:")
print(missing_data)

# Usunięcie wierszy z brakującymi danymi
df_cleaned = df.dropna()
```

## Ćwiczenie 3: Wizualizacja Danych

**Zadanie:**

1. Wykorzystaj bibliotekę `matplotlib` do wygenerowania wykresu punktowego (scatter plot) przedstawiającego dane z dwóch wybranych cech (na przykład "sepal length (cm)" i "sepal width (cm)").
2. Oznacz punkty na wykresie różnymi kolorami w zależności od gatunku irysa.
3. Zmodyfikuj kod tak, aby pokazać "petal length (cm)" i "petal width (cm)"
4. Jakie wnioski można wyciągnąć z ćwiczenia?

**Kod:**

```
import matplotlib.pyplot as plt

# Wykres punktowy
plt.figure(figsize=(8, 6))
for species in df_cleaned['target'].unique():
    subset = df_cleaned[df_cleaned['target'] == species]
    plt.scatter(subset['sepal length (cm)'], subset['sepal width (cm)'],
                label=iris.target_names[species])
plt.xlabel('sepal length (cm)')
plt.ylabel('sepal width (cm)')
plt.legend()
plt.show()
```

## Ćwiczenie 4: Analiza Liczebności Klas

**Zadanie:**

1. Użyć biblioteki `pandas`, metody `value_counts()` do wyświetlenia liczby próbek dla każdej z klas w zbiorze danych.
2. Metodę użyj na zbiorze `df['target']` wyliczonym w poprzednich ćwiczeniach.

3. Wyświetl wynik za pomocą funkcji `print`

Kod:

Brak 😊

## Ćwiczenie 5: Histogramy

Zadanie:

1. Wybrać jedną z cech, np. "sepal length (cm)".
2. Użyć biblioteki `matplotlib` do stworzenia histogramu wybranej cechy.
3. Wyświetlić histogram
4. Zrobić to samo ćwiczenie dla klasy (*target*)

Kod

```
# Wybór cechy do analizy
chosen_feature = "[...]"

# Tworzenie histogramu
plt.figure(figsize=(8, 6))
plt.hist(df[chosen_feature], bins=20, edgecolor='k')
plt.xlabel(chosen_feature)
plt.ylabel('Liczba Próbek')
plt.title(f'Histogram {chosen_feature}')
plt.show()
```