

Celem ćwiczenia jest zapoznanie się z podstawowymi etapami i narzędziami do przetwarzania danych wejściowych.

Ćwiczenie 1: Usuwanie Błędów w Danych

Cel ćwiczenia:

Nauczenie się technik usuwania błędów w danych.

Kroki:

Generowanie Zbioru Danych:

Stwórz zestaw danych z kilkoma błędami, takimi jak wartości odstające (*outliery*) lub brakujące dane.

Identyfikacja Błędów:

Użyj biblioteki *pandas* do zidentyfikowania błędów w danych, takich jak wartości odstające i brakujące dane.

Usuwanie Błędów:

Usuń błędy z danych, np. poprzez usunięcie próbek z wartościami odstającymi lub uzupełnienie brakujących danych.

```
import pandas as pd
import numpy as np
import matplotlib.pyplot as plt

# Generowanie danych losowych z rozkładu normalnego
np.random.seed(42)
data = {
    'A': np.random.normal(5, .5, 100),
    'B': np.random.normal(0, 200, 100)
}

# Dodanie wartości odstających (outliers) do danych
outliers = [40.5, 50.0, None, -4.0]
data['A'] = np.concatenate((data['A'], outliers))
data['B'] = np.concatenate((data['B'], outliers))

# Tworzenie ramki danych
df = pd.DataFrame(data)

# Wyczyszczenie danych użyj metody dropna() na DataFrame

# Tworzenie histogramu
```

```
# Oblicz wartość średnią i odchylenie standardowe zbioru A
mean = -1
std = -1
print(mean)
print(std)

# Analiza i usuwanie wartości odstających. Dobierz odpowiednie wartości
df = df[(df['A'] >= 0) & (df['A'] <= 0)]

# Wyświetlenie oczyszczonych danych (histogram)
```

Ćwiczenie 2: Standaryzacja Danych

Cel ćwiczenia:

Nauczenie się techniki standaryzacji danych.

Kroki:

Wizualizacja danych

Narysuj *scatterplot* serii **A** oraz **B** przed standaryzacją oraz po

Standaryzacja:

Użyj biblioteki `sklearn.preprocessing` do standaryzacji danych, tj. przekształć dane, aby miały średnią 0 i odchylenie standardowe 1.

```
from sklearn.preprocessing import StandardScaler
scaler = StandardScaler()
df[serie[[]]] = scaler.fit_transform(df[serie[[]]])
```