

## SPRAWOZDANIE

Zajęcia: Nauka o danych I

Prowadzący: prof. dr hab. Vasyl Martsenyuk

Laboratorium Nr 4

Data 27.02.2025

Temat: Wizualizacja danych za  
pomocą pakietu MATHPLOTLIB

Wariant 14

Krzysztof Świerczek

Informatyka

II stopień, stacjonarne,  
1semestr, gr. A

1. Polecenie: Zadanie polega na wytestowaniu różnych wariantów wizualizacji danych.

2. Github:

<https://github.com/Krzycho165/STUDIA>

```
In [4]: # 1. Biblioteka Matplotlib Krzysztof Świerczek
# 1.1 Wykres liniowy z linią przerywaną

import pandas as pd
import matplotlib.pyplot as plt

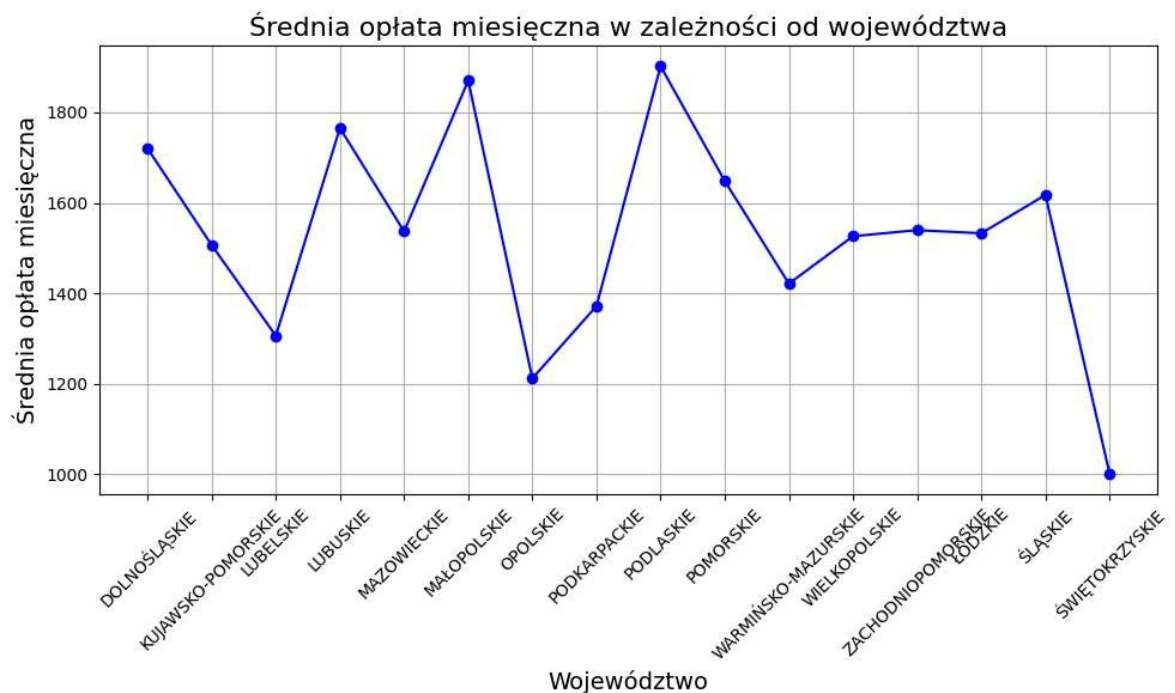
df = pd.read_csv("opiekuny_ms.csv")

df['Opłata miesięczna za pobyt'] = pd.to_numeric(df['Opłata miesięczna za pobyt'])
avg_oplata_miesieczna = df.groupby('Województwo')['Opłata miesięczna za pobyt'].
avg_oplata_miesieczna = avg_oplata_miesieczna[avg_oplata_miesieczna['Opłata mies

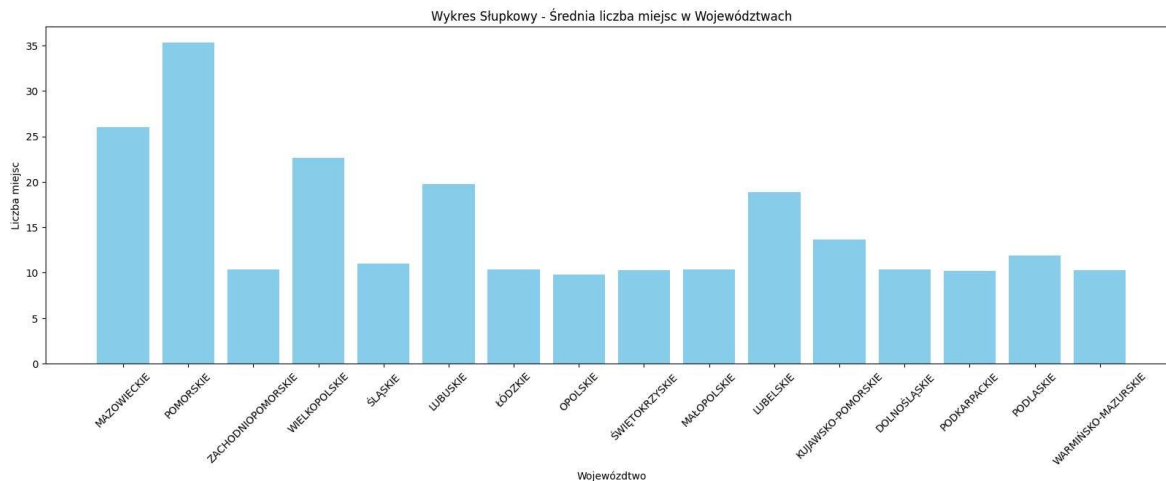
# Tworzenie wykresu liniowego
fig, ax = plt.subplots(figsize=(10, 6))
ax.plot(avg_oplata_miesieczna['Województwo'], avg_oplata_miesieczna['Opłata mies

# Dodanie tytułu i etykiety osi
ax.set_title('Średnia opłata miesięczna w zależności od województwa', fontsize=14)
ax.set_xlabel('Województwo', fontsize=14)
ax.set_ylabel('Średnia opłata miesięczna', fontsize=14)

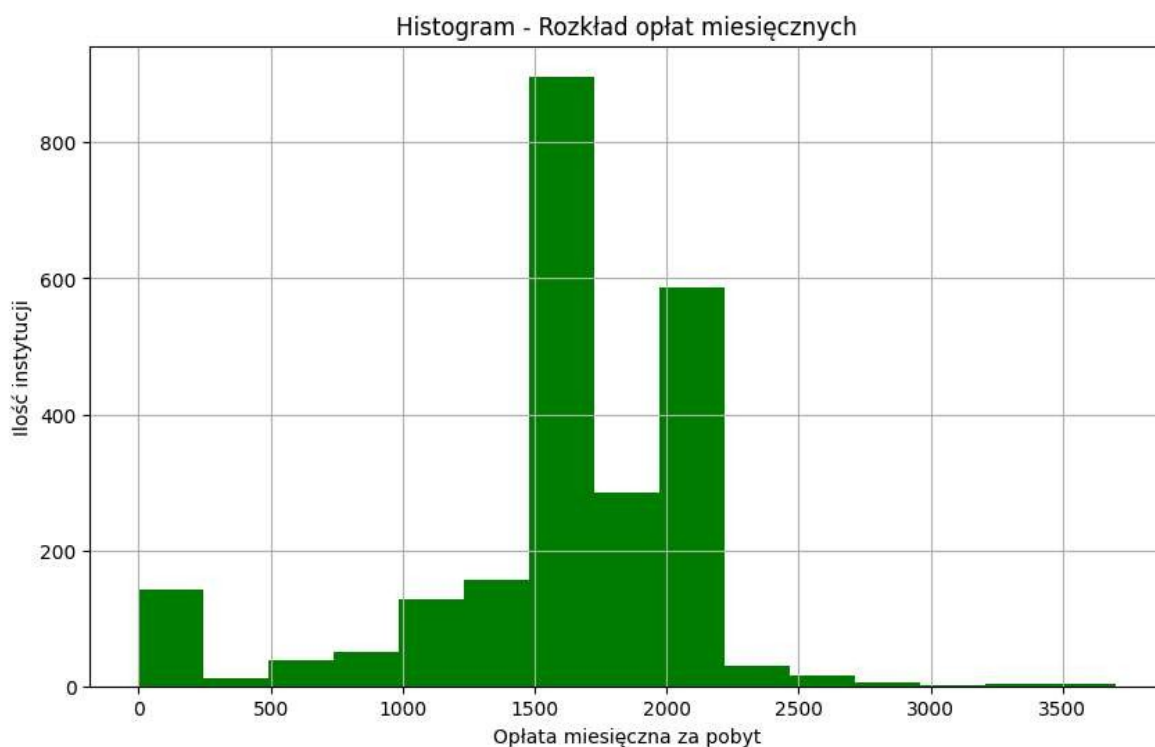
# Ustawienie wykresu
plt.xticks(rotation=45)
plt.grid(True)
plt.tight_layout()
plt.show()
```



```
In [5]: # 1.2 Wykres słupkowy ze zmianą koloru Krzysztof Świerczek
plt.figure(figsize=(20, 6))
plt.bar(df['Województwo'], df['Liczba miejsc'], color='skyblue')
plt.title('Wykres Słupkowy - Średnia liczba miejsc w Województwach')
plt.xlabel('Województwo')
plt.ylabel('Liczba miejsc')
plt.xticks(rotation=45)
plt.show()
```



```
In [6]: # 1.3 Histogram ze zmianą koloru Krzysztof Świerczek
plt.figure(figsize=(10, 6))
plt.hist(df['Opłata miesięczna za pobyt'], bins=15, color='green')
plt.title('Histogram - Rozkład opłat miesięcznych')
plt.xlabel('Opłata miesięczna za pobyt')
plt.ylabel('Ilość instytucji')
plt.grid(True)
plt.show()
```



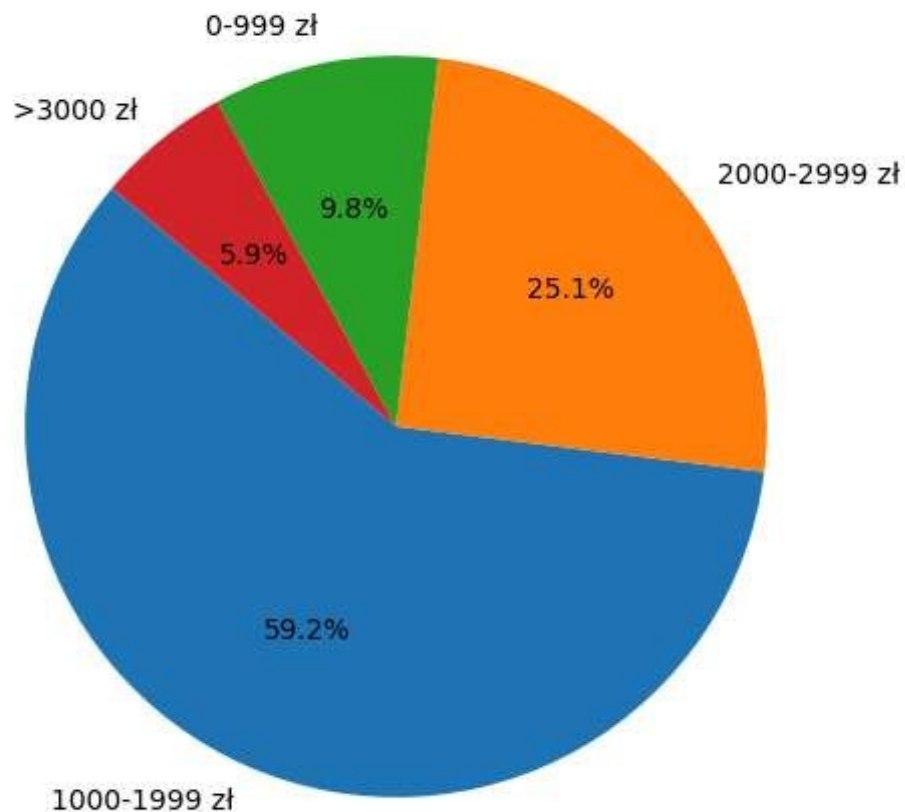
```
In [7]: # 1.4 Wykres kołowy Krzysztof Świerczek
def categorize_fee(fee):
    if fee <= 999:
        return '0-999 zł'
    elif fee <= 1999:
        return '1000-1999 zł'
    elif fee <= 2999:
        return '2000-2999 zł'
    else:
        return '>3000 zł'

df['Opłata_kategoria'] = df['Opłata miesięczna za pobyt'].apply(lambda x: float(
```

```
df['Opłata_kategoria'] = df['Opłata_kategoria'].apply(categorize_fee)
category_counts = df['Opłata_kategoria'].value_counts()

plt.figure(figsize=(10, 6))
plt.pie(category_counts, labels=category_counts.index, autopct='%1.1f%%', startangle=90)
plt.title('Procentowy udział opłat miesięcznych za pobyt w przedziałach')
plt.show()
```

### Procentowy udział opłat miesięcznych za pobyt w przedziałach



```
In [8]: # 1.6 Wykres 3D: Krzysztof Świerczek
fig = plt.figure(figsize=(15, 8))
ax = fig.add_subplot(111, projection='3d')

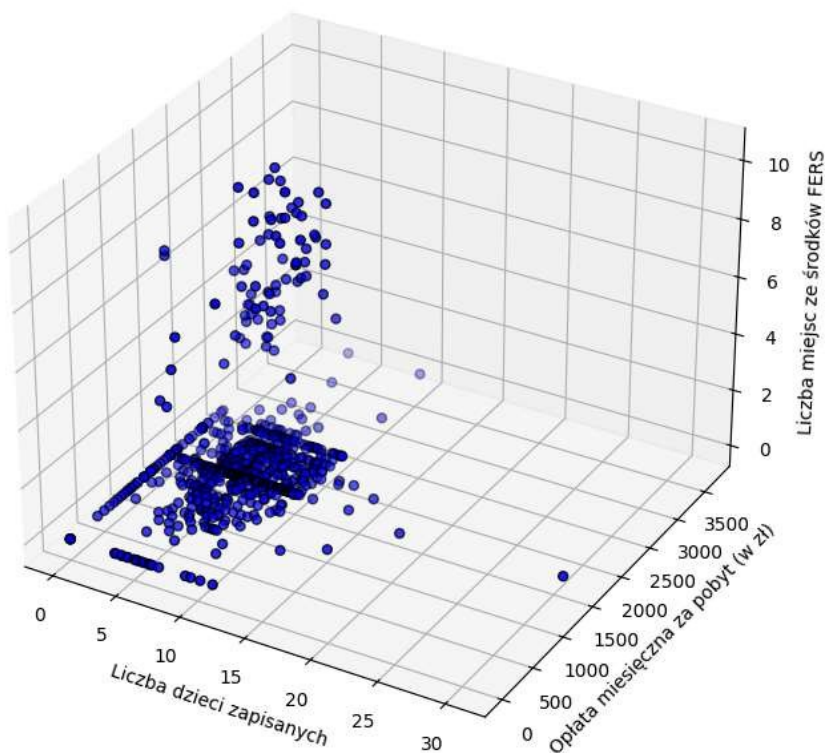
x = df['Liczba dzieci zapisanych']
y = df['Opłata miesięczna za pobyt'].apply(lambda x: float(x) if x != '' else 0)
z = df['Liczba miejsc ze środków FERS']

ax.scatter(x, y, z, c='blue', marker='.', edgecolors='black', s=100)

ax.set_title('Zależność liczby dzieci zapisanych od opłaty miesięcznej za pobyt')
ax.set_xlabel('Liczba dzieci zapisanych')
ax.set_ylabel('Opłata miesięczna za pobyt (w zł)')
ax.set_zlabel('Liczba miejsc ze środków FERS')

plt.show()
```

Zależność liczby dzieci zapisanych od opłaty miesięcznej za pobyt i miejsc ze środków FERS(Wykres 3D)



```
In [10]: # 2. Biblioteka Plotly: Krzysztof Świerczek
# 2.1 Wykres Liniowy (korzystamy z danych z punktu 1.1, dostosowanie motywu)
import plotly.express as px

fig = px.line(avg_oplata_miesieczna, x='Województwo', y='Opłata miesięczna za po
            title='Średnia opłata miesięczna w zależności od województwa',
            labels={'Opłata miesięczna za pobyt': 'Średnia opłata miesięczna',
            line_shape='linear', template='plotly_dark',
            color_discrete_sequence=['blue']) # Kolor linii

fig.update_layout(
    xaxis_title='Województwo',
    yaxis_title='Średnia opłata miesięczna',
)

fig.show()
```

```
In [11]: # 2.2 Wykres słupkowy (z dostosowanymi kolorami) Krzysztof Świerczek
fig = px.bar(df, x='Imię', y='Liczba miejsc', color='Województwo', title='Liczba
fig.update_layout(xaxis_title="Imię", yaxis_title="Liczba miejsc", barmode="grou

# Pokazanie wykresu
fig.show()
```

```
In [12]: # 2.3 Wykres kołowy Krzysztof Świerczek
# Podział kolumny "Opłata miesięczna za pobyt" na grupy używając funkcji categorize
df['Opłata_grupy'] = df['Opłata miesięczna za pobyt'].apply(lambda x: float(x) if x < 100 else 100)
df['Opłata_grupy'] = df['Opłata_grupy'].apply(categorize_fee)

group_counts = df['Opłata_grupy'].value_counts(normalize=True) * 100

fig = px.pie(values=group_counts.values, names=group_counts.index, title='Procenty')
fig.update_layout(showlegend=True)
fig.show()
```

```
In [13]: # 2.5 Wykres punktowy Krzysztof Świerczek

mdf = df[df['Liczba dzieci zapisanych na miejsca z FERS'] > 0] # Bierz pod uwagę
fig = px.scatter(mdf, x='Opłata miesięczna za pobyt', y='Liczba dzieci zapisanych')
fig.show()
```



```
In [14]: # 2.6 Wykres 3D Krzysztof Świerczek
fig = px.scatter_3d(df, x='Liczba dzieci zapisanych', y='Opłata miesięczna za po
fig.show()
```

### 3. Wnioski z wykonanego kodu:

Celem zadania była wizualizacja danych na różne możliwe sposoby. W tym celu wczytano plik `opiekuny_ms.csv`, a następnie dokonano konwersji danych, aby umożliwić dalsze operacje analityczne. Szczególną uwagę zwrócono na kolumnę „Opłata miesięczna za pobyt”, z której usunięto jednostkę „zł” i zamieniono wartości na typ numeryczny. Następnie obliczono średnią opłatę miesięczną dla każdego województwa, eliminując przy tym ewentualne brakujące dane.

Do wizualizacji wyników wykorzystano bibliotekę `matplotlib`, tworząc wykres liniowy przedstawiający średnie opłaty miesięczne w poszczególnych województwach. Na wykresie zastosowano znaczniki dla poszczególnych punktów danych, a także oznaczono osie, co pozwoliło na lepszą czytelność wyników. Nazwy województw zostały odpowiednio sformatowane poprzez rotację, a dodanie siatki ułatwiło analizę wartości. Przeprowadzona analiza pozwoliła na szybkie porównanie wysokości opłat w różnych regionach i wskazanie województw, w których koszty są najwyższe oraz najniższe. Warto zauważyć, że regionalne różnice w wysokości opłat mogą wynikać z wielu czynników, takich jak liczba placówek, koszty utrzymania czy poziom zamożności mieszkańców danego województwa.

W celu ulepszenia analizy możliwe jest zastosowanie wykresu słupkowego zamiast liniowego, co mogłoby lepiej oddać różnice między województwami. Dodatkowo uwzględnienie większej liczby zmiennych, np. liczby placówek w danym regionie, pozwoliłoby na uzyskanie bardziej kompleksowego obrazu sytuacji. W przyszłości warto również przeprowadzić analizę statystyczną w celu określenia istotności zaobserwowanych różnic. Biblioteka `Matplotlib` daje potężne możliwości wizualizacji danych i dopasowania metody wizualizacji do każdego konkretnego przypadku.