

## SPRAWOZDANIE

Zajęcia: Nauka o danych I

Prowadzący: prof. dr hab. Vasyl Martsenyuk

Laboratorium Nr 2 Data 27.09.2025 Temat: "Wizualizacja Danych za Pomocą Biblioteki Matplotlib" Wariant 1	Artur Rolak Informatyka II stopień, stacjonarne, 1 semestr, gr. 1b
----------------------------------------------------------------------------------------------------------------------	-----------------------------------------------------------------------------

<https://github.com/rINkoo/Studia-Magisterskie/tree/main/Nauka%20o%20Danych%20L/Laboratorium%201>

### 1. Polecenie:

Celem laboratorium było zapoznanie się z podstawami wizualizacji danych w języku Python z wykorzystaniem bibliotek Matplotlib i Pandas.

Zadanie polegało na:

- pobraniu wybranego zbioru danych z serwisu Kaggle,
- obliczeniu podstawowych statystyk opisowych kolumn numerycznych,
- przygotowaniu czterech różnych wykresów najlepiej opisujących wybrany zbiór,
- dodaniu komentarzy interpretujących wyniki,
- oraz opracowaniu krótkiego podsumowania zawierającego pomysł na praktyczne wykorzystanie analizowanego zbioru danych.

Do realizacji zadania został wykorzystany zbiór danych „Netflix Movies and TV Shows” pochodzący z serwisu Kaggle (adres:

<https://www.kaggle.com/datasets/shivamb/netflix-shows>).

## 2. Opis programu opracowanego (kody źródłowe, rzuty ekranu)

### Laboratorium 3 - Wizualizacja danych przy użyciu Matplotlib i innych bibliotek

Bazowo zaczynamy od przygotowania danych

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

df = pd.read_csv("netflix_titles.csv") # <-- plik musi być w tym samym folderze co notatnik
print(df.columns)
df.head()
```

```
Index(['show_id', 'type', 'title', 'director', 'cast', 'country', 'date_added',
       'release_year', 'rating', 'duration', 'listed_in', 'description'],
      dtype='object')
```

	show_id	type	title	director	cast	country	date_added	release_year	rating	duration	listed_in	description
0	s1	Movie	Dick Johnson Is Dead	Kirsten Johnson	NaN	United States	September 25, 2021	2020	PG-13	90 min	Documentaries	As her father nears the end of his life, filmm...
1	s2	TV Show	Blood & Water	NaN	Ama Qamata, Khosi Ngema, Gail Mababalane, Thaban...	South Africa	September 24, 2021	2021	TV-MA	2 Seasons	International TV Shows, TV Dramas, TV Mysteries	After crossing paths at a party, a Cape Town L...
2	s3	TV Show	Ganglands	Julien Leclercq	Sami Bouajila, Tracy Gotoas, Samuel Jouy, Nabi...	NaN	September 24, 2021	2021	TV-MA	1 Season	Crime TV Shows, International TV Shows, TV Act...	To protect his family from a powerful drug kor...
3	s4	TV Show	Jailbirds New Orleans	NaN	NaN	NaN	September 24, 2021	2021	TV-MA	1 Season	Docuseries, Reality TV	Feuds, flirtations and toilet talk go down amo...
4	s5	TV Show	Kota Factory	NaN	Mayur More, Jitendra Kumar, Ranjan Raj, Alam K...	India	September 24, 2021	2021	TV-MA	2 Seasons	International TV Shows, Romantic TV Shows, TV ...	In a city of coaching centers known to train L...

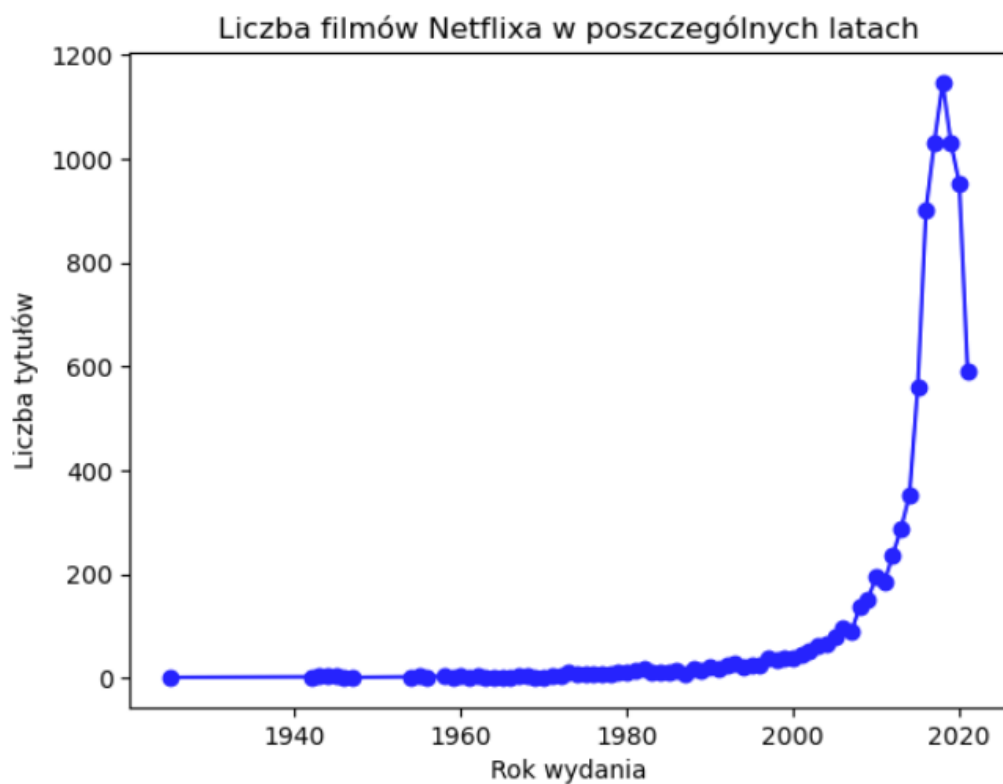
```
# Import matplotlib.pyplot as plt
# Import numpy as np
```

```
# x = [1, 2, 3, 4, 5]
# y = [2, 3, 5, 7, 11]
```

## 3.1 Rodzaje wykresów podstawowych

### ▼ 3.1.1 Wykres Liniowy

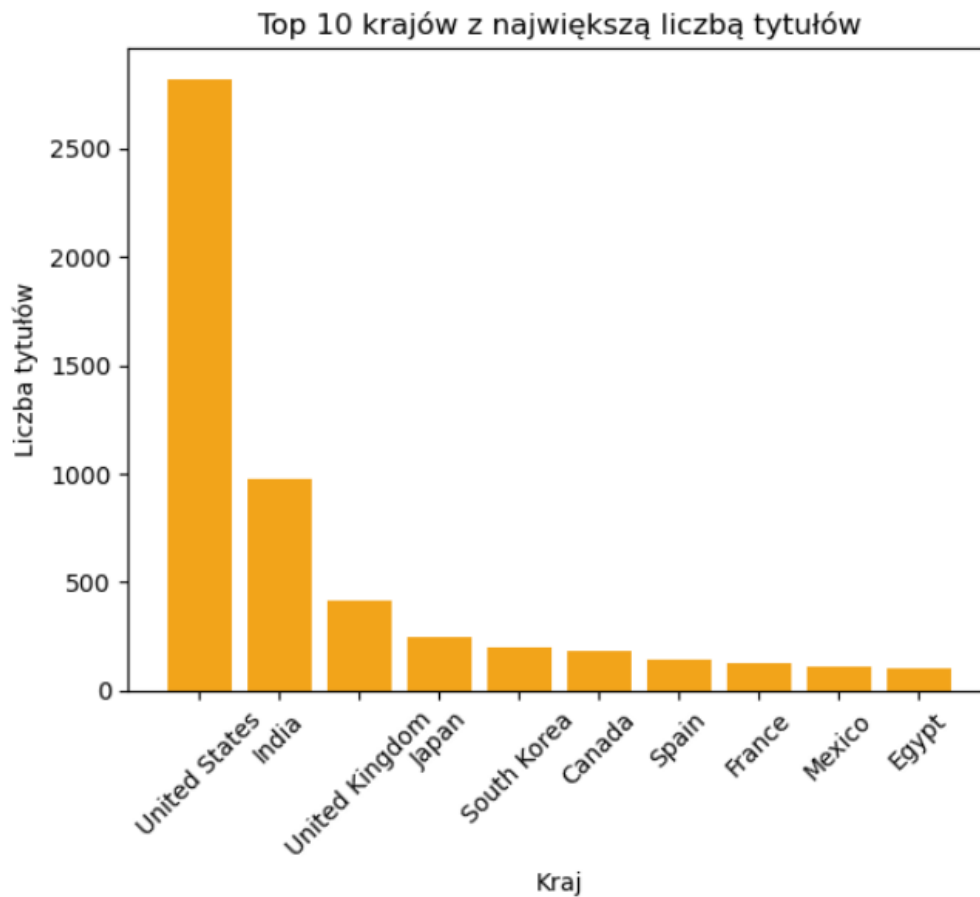
```
[15]: dane = df['release_year'].value_counts().sort_index()
plt.plot(dane.index, dane.values, marker='o', color='blue')
plt.title("Liczba filmów Netflixa w poszczególnych latach")
plt.xlabel("Rok wydania")
plt.ylabel("Liczba tytułów")
plt.show()
```



### 3.1.2 Wykres słupkowy

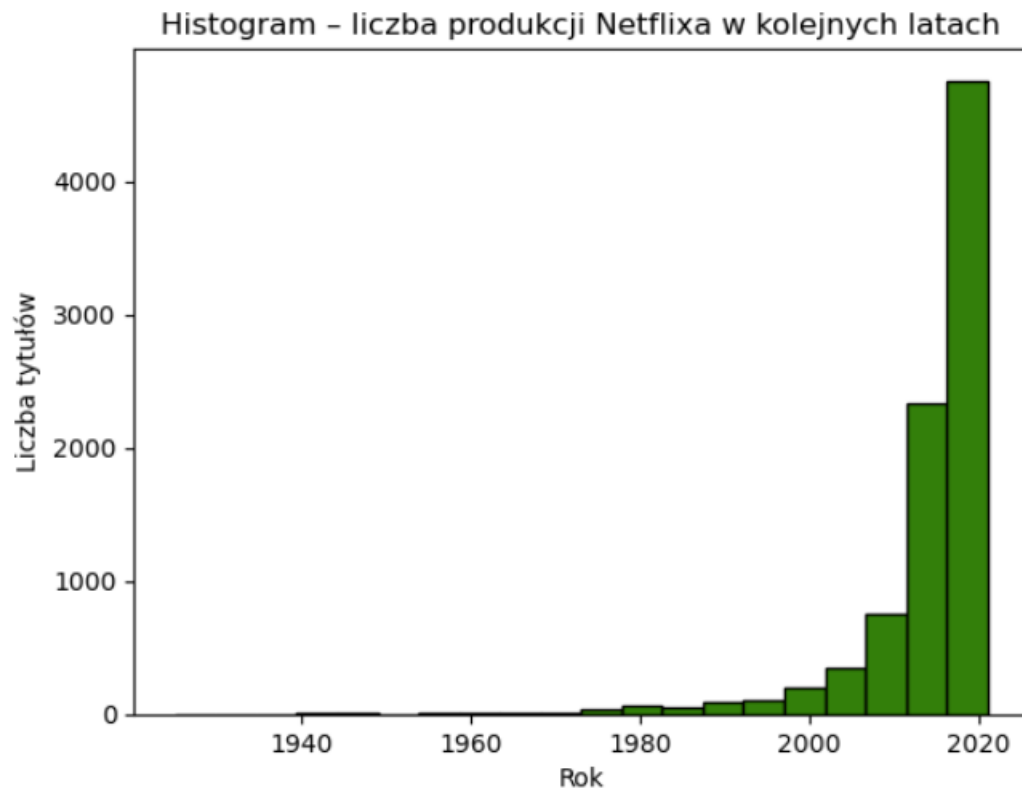
```
16]: kraje = df['country'].value_counts().head(10)

plt.bar(kraje.index, kraje.values, color='orange')
plt.title("Top 10 krajów z największą liczbą tytułów")
plt.xlabel("Kraj")
plt.ylabel("Liczba tytułów")
plt.xticks(rotation=45)
plt.show()
```



### 3.1.3 Histogram

```
[ ]: plt.hist(df['release_year'], bins=20, color='green', edgecolor='black')
plt.title("Histogram – liczba produkcji Netflix w kolejnych latach")
plt.xlabel("Rok")
plt.ylabel("Liczba tytułów")
plt.show()
```

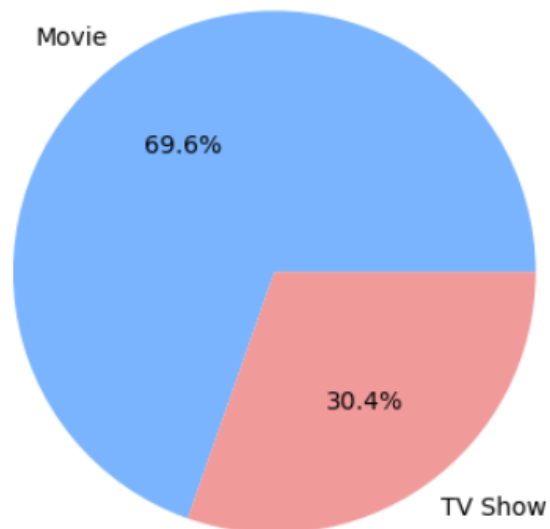


### ▼ 3.1.4 Wykres kołowy

```
] : typy = df['type'].value_counts()

plt.pie(typy.values, labels=typy.index, autopct='%1.1f%%', colors=['#66b3ff', '#ff9999'])
plt.title("Udział filmów i seriali w katalogu Netflix")
plt.show()
```

Udział filmów i seriali w katalogu Netflix



### 3. Wnioski

Wykonane laboratorium pozwoliło przećwiczyć praktyczne wykorzystanie bibliotek Pandas i Matplotlib w analizie rzeczywistych danych.

Na podstawie analizy datasetu „Netflix Movies and TV Shows” można zauważyć wyraźny wzrost liczby publikowanych tytułów w ostatniej dekadzie, co potwierdza globalny rozwój rynku streamingowego.

Zastosowanie różnych typów wykresów pozwoliło na wizualne przedstawienie struktury danych – od trendów czasowych po proporcje gatunkowe i geograficzne.

Pomysł na wykorzystanie datasetu

Zbiór danych o tytułach Netflixu może być wykorzystany do:

- stworzenia systemu rekomendacji filmów na podstawie gatunku, roku wydania i kraju,
- analizy trendów produkcyjnych w poszczególnych latach,
- badania różnorodności geograficznej treści,
- lub wizualizacji historii rozwoju Netflixu jako firmy medialnej.

Analiza tego typu danych może znaleźć praktyczne zastosowanie w marketingu, analizie trendów kulturowych oraz w badaniach nad rynkiem treści wideo.