

Nama : Try Yani Rizki Nur Rohmah

NIM : 122450020

✓ Tugas VDI 3 (Data Visualitation Methodologi & Define Purpose and parameters)

1. Tentukan data yang akan anda gunakan dalam visualisasi
2. Berdasarkan data yang anda pilih tentukan tujuan, user, fungsi dan tone (jika ada) yang sesuai dengan visualisasi yang akan anda buat
3. Buatlah visualisasi dengan menggunakan library python seperti matplotlib, seaborn atau tools lain

```
# Import library yang diperlukan
import pandas as pd
import numpy as np
import os
import matplotlib.pyplot as plt
import seaborn as sns
```

1. Dataset yang saya gunakan pada tugas ini ialah Kumpulan data ini berisi hampir 1 Juta ulasan film unik dari 1150 film IMDb dengan genre animasi yang saya dapatkan di kaggle.

```
# melakukan import dataset
df = pd.read_csv('/content/Animation (1).csv')
# menampilkan informasi umum dataset
print(df.head())
print(df.info())
print(df.describe())
```

```
↩
   name  year  movie_rating  run_length  genres \
0  WALL·E  2008           G    1h 38min  Animation; Adventure;
1  Finding Nemo  2003           G    1h 40min  Animation; Adventure; Comedy;
2  The Lion King  1994           G    1h 28min  Animation; Adventure; Drama;
3    Up  2009          PG    1h 36min  Animation; Adventure; Comedy;
4  Toy Story  1995           G    1h 21min  Animation; Adventure; Comedy;
```

```
   release_date  rating  num_raters  num_reviews \
0  27 June 2008 (USA)    8.4    967917    1268
1  30 May 2003 (USA)    8.1    922080     968
2  24 June 1994 (USA)    8.5    910179    1009
3  29 May 2009 (USA)    8.2    905044     891
4  22 November 1995 (USA)  8.3    858215     606
```

```
   review_url
0  https://www.imdb.com/title/tt0910970/reviews/...
1  https://www.imdb.com/title/tt0266543/reviews/...
2  https://www.imdb.com/title/tt0110357/reviews/...
3  https://www.imdb.com/title/tt1049413/reviews/...
4  https://www.imdb.com/title/tt0114709/reviews/...
```

```
<class 'pandas.core.frame.DataFrame'>
```

```
RangeIndex: 100 entries, 0 to 99
```

```
Data columns (total 10 columns):
```

```
#   Column      Non-Null Count  Dtype
---  ---
0  name         100 non-null    object
1  year         100 non-null    int64
2  movie_rating 100 non-null    object
3  run_length   100 non-null    object
4  genres       100 non-null    object
5  release_date 100 non-null    object
6  rating       100 non-null    float64
7  num_raters   100 non-null    int64
8  num_reviews  100 non-null    int64
9  review_url   100 non-null    object
```

```
dtypes: float64(1), int64(3), object(6)
```

```
memory usage: 7.9+ KB
```

```
None
```

2. Berdasarkan data yang anda pilih tentukan tujuan, user, fungsi dan tone (jika ada) yang sesuai dengan visualisasi yang akan anda buat

jawab:

1. Tujuan saya ialah:

- Menganalisis tren dengan melihat bagaimana industri animasi berkembang seiring waktu, misalnya dengan melihat jumlah film animasi yang dirilis setiap tahun.
- Melihat pola dalam genre yang paling populer, studio produksi, atau durasi rata-rata film animasi.
- Mengevaluasi popularitas jika terdapat data tentang rating atau pendapatan dengan melihat faktor-faktor yang mempengaruhi popularitas suatu animasi.

2. User (Pengguna) yang mungkin memanfaatkan visualisasi ini ialah :

- Analis Industri Animasi: Mereka yang bekerja di industri media atau animasi yang memerlukan data untuk memantau perkembangan atau potensi pasar.
- Mahasiswa dan Akademisi: Peneliti atau mahasiswa yang mempelajari tren dalam industri animasi untuk tugas atau penelitian.
- Penonton Umum: Orang yang memiliki ketertarikan pada tren animasi, misalnya untuk melihat film atau genre apa yang sedang populer.

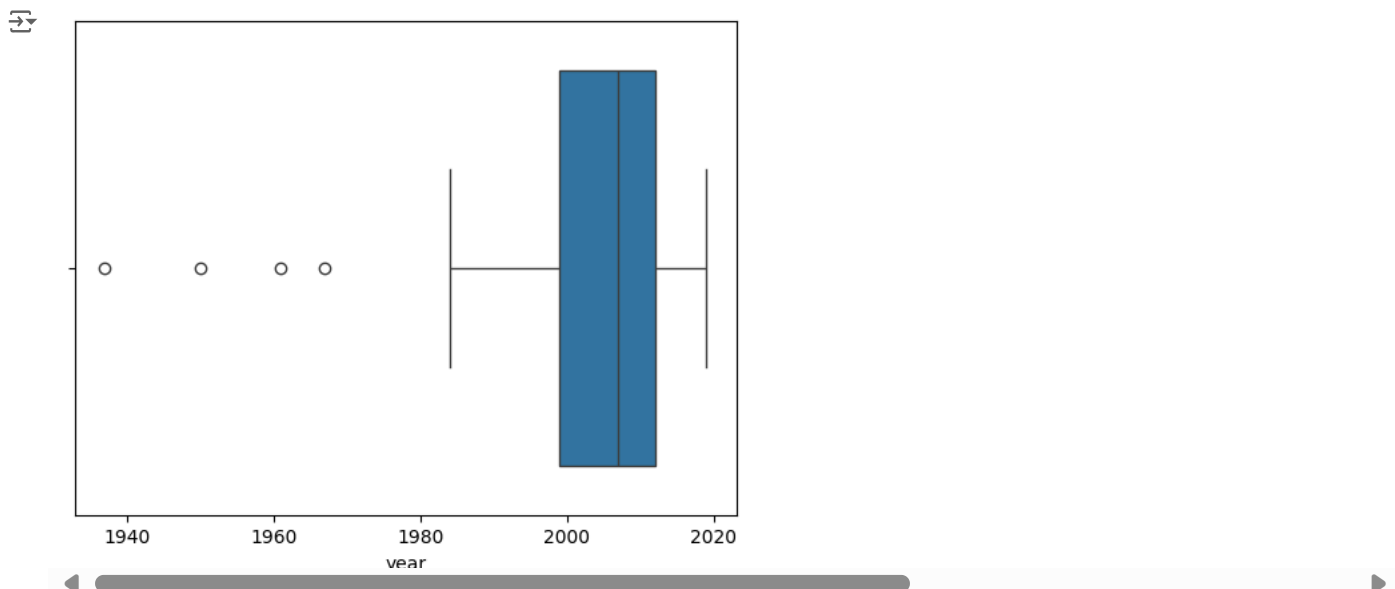
3. Fungsi visualisasi ialah untuk menyajikan tren, perbandingan, dan pemahaman data secara efektif. Visualisasi dapat digunakan untuk menampilkan perubahan jumlah produksi animasi dari tahun ke tahun, memungkinkan perbandingan antara berbagai genre, studio, atau film guna mengidentifikasi yang paling populer, serta membantu pengguna memahami distribusi durasi film atau jumlah film per kategori. Dengan demikian, pola atau tren tertentu dalam data menjadi lebih mudah dikenali dan diinterpretasikan.

4. Tone visualisasi mengacu pada cara penyampaian informasi melalui visual dan disesuaikan dengan audiens. Untuk analisis formal, visualisasi harus bersifat profesional dan data-driven, dengan grafik yang sederhana, jelas, serta menggunakan skala atau label yang akurat. Jika ditujukan untuk penonton umum atau non-teknis, tone dapat dibuat lebih ringan dan informatif, dengan visual yang menarik menggunakan warna-warna cerah namun tetap mempertahankan kejelasan dan akurasi informasi yang disampaikan.

```
for col in df.select_dtypes(include='object').columns:
    df[col] = df[col].astype('category')
```

3. visualisasi dengan menggunakan library python seperti matplotlib, seaborn atau tools lain

```
#Mengecek Outliner dengan boxplot
sns.boxplot(x=df['year'])
plt.show()
```



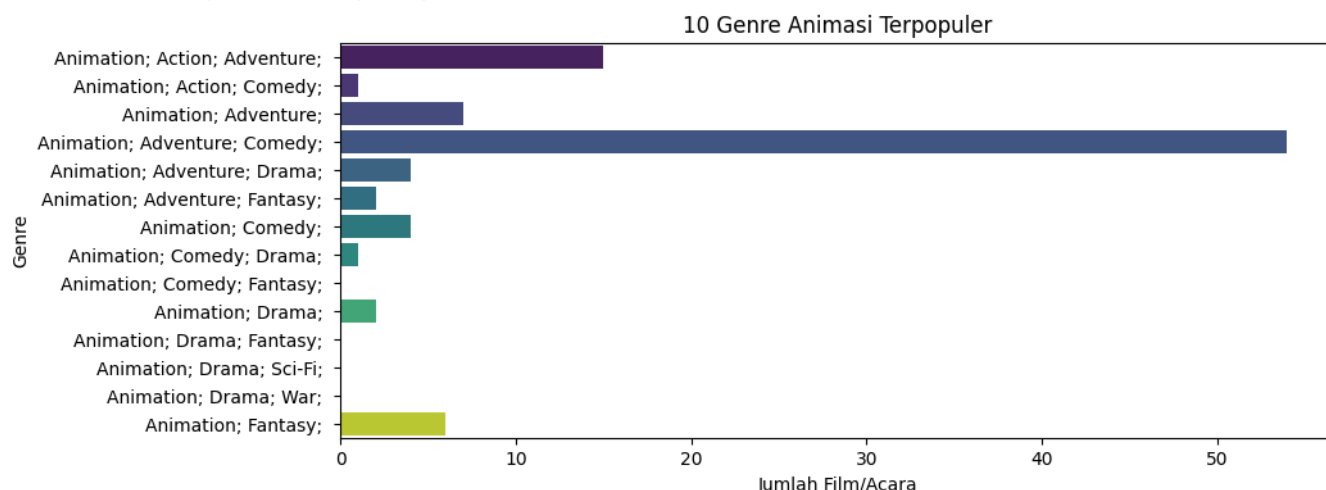
```
# grafik untuk melihat 10 genre teratas
top_genres = df['genres'].value_counts().head(10)

plt.figure(figsize=(10,4))
sns.barplot(x=top_genres.values, y=top_genres.index, palette='viridis')
plt.title('10 Genre Animasi Terpopuler')
plt.xlabel('Jumlah Film/Acara')
plt.ylabel('Genre')
plt.show()
```

```
<ipython-input-5-dfaedf4da681>:5: FutureWarning:
```

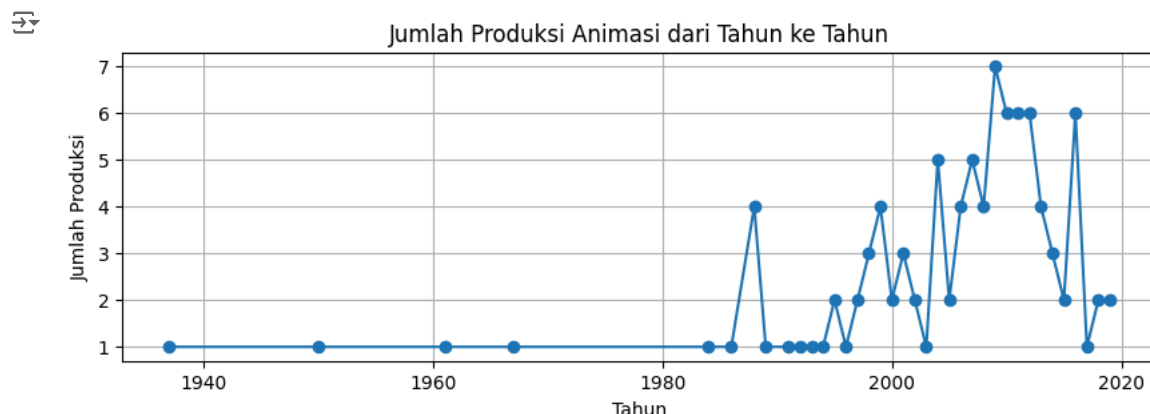
Passing `palette` without assigning `hue` is deprecated and will be removed in v0.14.0. Assign the `y` variable to `hue` and set `le`

```
sns.barplot(x=top_genres.values, y=top_genres.index, palette='viridis')
```



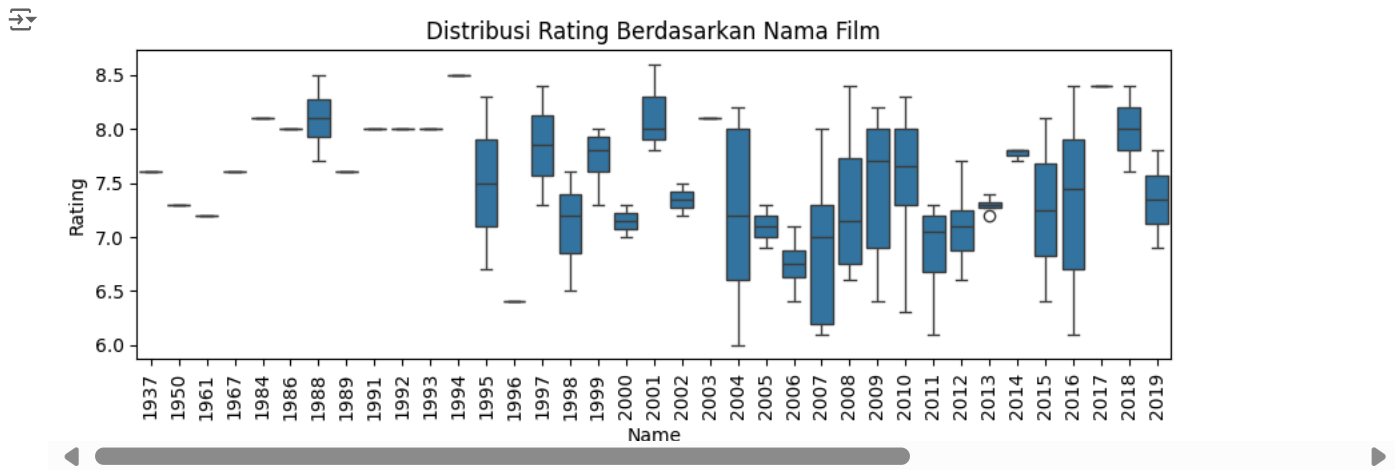
Grafik di atas menunjukkan 10 genre animasi yang paling populer berdasarkan jumlah film atau acara yang ada. Genre "Animasi, Adventure, Comedy" menempati posisi teratas dengan jumlah terbanyak, diikuti oleh "Animasi, Action, Adventure". Genre-genre lainnya seperti "Animasi, Fantasy", "Animasi, Drama", dan kombinasi genre lainnya juga cukup populer. Grafik ini memberikan gambaran umum tentang preferensi penonton terhadap genre-genre tertentu dalam animasi, di mana kombinasi antara petualangan, komedi, dan aksi menjadi yang paling diminati.

```
# Plot tren produksi animasi berdasarkan tahun (misal ada kolom 'year')
plt.figure(figsize=(10,3))
df.groupby('year').size().plot(kind='line', marker='o')
plt.title('Jumlah Produksi Animasi dari Tahun ke Tahun')
plt.xlabel('Tahun')
plt.ylabel('Jumlah Produksi')
plt.grid(True)
plt.show()
```



Grafik ini menunjukkan adanya peningkatan produksi animasi terutama setelah tahun 1980, dengan beberapa fluktuasi hingga 2020.

```
# Boxplot rating berdasarkan nama film
plt.figure(figsize=(10,3))
sns.boxplot(x='year', y='rating', data=df)
plt.title('Distribusi Rating Berdasarkan Nama Film')
plt.xlabel('Name')
plt.ylabel('Rating')
plt.xticks(rotation=90)
plt.show()
```



Pada tahun 1995, terdapat variasi besar pada rating film dengan beberapa outliers di bagian bawah, sementara tahun 2008 menunjukkan rating yang lebih rendah dibanding tahun lainnya. Pada 2019, median rating cukup tinggi dengan distribusi yang lebih sempit, menandakan film-film pada tahun tersebut mendapat rating lebih baik dan lebih seragam. Secara keseluruhan, boxplot ini memudahkan perbandingan performa rating film animasi berdasarkan tahun rilis dan memperlihatkan tren atau variasi kualitas animasi dari waktu ke waktu.