SOURCES code

```
# THEN FEEL FREE TO DELETE THIS CELL.

# NOTE: THIS NOTEBOOK ENVIRONMENT DIFFERS FROM KAGGLE'S PYTHON

# ENVIRONMENT SO THERE MAY BE MISSING LIBRARIES USED BY YOUR

# NOTEBOOK.

import kagglehub

parasharmanas_movie_recommendation_system_path =
kagglehub.dataset_download('parasharmanas/movie-recommendation-system')


print('Data source import complete.')


print('Jumlah Data :', len(df1.iloc[:,1]))

print('Jumlah Fitur :', len(df1.iloc[1,:]))

print(f'Terdapat {len(df1.iloc[1,:])} Kolom Fitur pada Dataset yaitu:')

print('Fitur Data :', df1.columns.tolist()[:])

pd.options.display.max_columns = None

df1.head()

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

RangeIndex: 62423 entries, 0 to 62422

Data columns (total 3 columns):

 #   Column   Non-Null Count  Dtype

---  ------   --------------  -----

 0   movieId  62423 non-null  int64

 1   title    62423 non-null  object

 2   genres   62423 non-null  object
```

dtypes: int64(1), object(2)

memory usage: 1.4+ MB

Pisahkan genre menggunakan pemisah '|'

df1['genres'] = df1['genres'].str.split('|')

Bersihkan judul film

df1['title'] = df1['title'].apply(clean_title)

Perbarui movies_data

movies_data = df1[['movieId', 'title', 'genres']]

Mendapatkan genre unik dari semua film

unique_genres = pd.Series([genre for genres_list in movies_data['genres'] for genre in genres_list]).unique()

Output hasil

print(movies_data.head())

Terdapat 62298 Judul Film

Terdapat 20 Genre Film.

Genre Film: ['Adventure' 'Animation' 'Children' 'Comedy' 'Fantasy' 'Romance' 'Drama'

'Action' 'Crime' 'Thriller' 'Horror' 'Mystery' 'Sci-Fi' 'IMAX'

'Documentary' 'War' 'Musical' 'Western' 'Film-Noir' '(no genres listed)']

Periksa jumlah baris dengan '(no genres listed)'

no_genres_count = movies_data[movies_data['genres'].apply(lambda x: '(no genres listed)' in x)].shape[0]

print(f"Terdapat {no_genres_count} film tanpa genre.")

```
# Hapus baris dengan '(no genres listed)'

movies_data = movies_data[~movies_data['genres'].apply(lambda x: '(no genres listed)' in x)]


# Perbarui daftar genre unik

unique_genres = pd.Series([genre for genres_list in movies_data['genres'] for genre in genres_list]).unique()


# Tampilkan hasil setelah penghapusan

print(f"Setelah penghapusan, terdapat {movies_data['title'].nunique()} Judul Film.")

print(f"Terdapat {len(unique_genres)} Genre Film setelah pembaruan.")

print("Genre Film:", unique_genres)


import pandas as pd

import matplotlib.pyplot as plt


# Menghitung jumlah film per genre

genre_counts = pd.Series([genre for genres_list in movies_data['genres'] for genre in genres_list]).value_counts()


plt.figure(figsize=(12, 6))

genre_counts.plot(kind='bar', color='skyblue')

plt.title('Jumlah Film per Genre', fontsize=16)

plt.xlabel('Genre', fontsize=14)

plt.ylabel('Jumlah Film', fontsize=14)

plt.xticks(rotation=45, ha='right')
```

```
plt.show()

print('Jumlah Data :', len(df2.iloc[:,1]))

print('Jumlah Fitur :', len(df2.iloc[1,:]))

print(f'Terdapat {len(df2.iloc[1,:])} Kolom Fitur pada Dataset yaitu:')

print('Fitur Data :', df2.columns.tolist()[:])

pd.options.display.max_columns = None

df2.head()

print("Distribusi Rating:")

print(df2['rating'].value_counts())

print("\nRating Rata-Rata per Film:")

print(df2.groupby('movieId')['rating'].mean().head())

print("\nRating Rata-Rata per Pengguna:")

print(df2.groupby('userId')['rating'].mean().head())
```