✅ Google Colab Code Block:

# 📦 Step 1: Upload the CSV file

from google.colab import files

uploaded = files.upload()

# 📥 Step 2: Load dataset

import pandas as pd

df = pd.read\_csv('netflix\_movies\_detailed\_up\_to\_2025.csv')

print("Initial data shape:", df.shape)

df.head()

# 🧹 Step 3: Data Cleaning

# -- Convert 'date\_added' to datetime

df['date\_added'] = pd.to\_datetime(df['date\_added'], errors='coerce')

# -- Drop duplicates

df.drop\_duplicates(inplace=True)

# -- Clean text columns

text\_cols = ['title', 'director', 'cast', 'country', 'rating', 'genre']

for col in text\_cols:

if col in df.columns:

df[col] = df[col].astype(str).str.strip().str.lower()

# -- Split genres for further analysis

df['genre\_list'] = df['genre'].str.split(',')

# -- Extract year from date\_added

df['year\_added'] = df['date\_added'].dt.year

# 🔍 Check cleaned data

print("Cleaned data shape:", df.shape)

df.info()

📊 Charts & Visualizations:

# 📈 Plotting setup

import matplotlib.pyplot as plt

import seaborn as sns

sns.set(style='whitegrid')

plt.rcParams['figure.figsize'] = (12, 6)

# 📅 Number of Movies Added Each Year

df['year\_added'].value\_counts().sort\_index().plot(kind='bar')

plt.title('Number of Movies Added to Netflix Each Year')

plt.xlabel('Year')

plt.ylabel('Number of Movies')

plt.xticks(rotation=45)

plt.show()

# 🎭 Top 10 Genres

from collections import Counter

genre\_counts = Counter([g.strip() for sublist in df['genre\_list'].dropna() for g in sublist])

top\_genres = pd.DataFrame(genre\_counts.most\_common(10), columns=['Genre', 'Count'])

sns.barplot(x='Count', y='Genre', data=top\_genres)

plt.title('Top 10 Genres on Netflix')

plt.xlabel('Count')

plt.ylabel('Genre')

plt.show()

# 📊 Ratings Distribution

df['rating'].value\_counts().plot(kind='bar', color='skyblue')

plt.title('Content Ratings Distribution')

plt.xlabel('Rating')

plt.ylabel('Number of Titles')

plt.xticks(rotation=45)

plt.show()

# 🌍 Top 10 Countries

df['country'].value\_counts().head(10).plot(kind='bar', color='orange')

plt.title('Top 10 Countries Producing Netflix Content')

plt.xlabel('Country')

plt.ylabel('Number of Titles')

plt.xticks(rotation=45)

plt.show()