NETFLIX PROJECT.py

```
import numpy as np
import pandas as pd
import matplotlib.pyplot as plt
import seaborn as sns
data = pd.read_csv(r"C:\Users\Kartik Kumar\Downloads\netflix1.csv", delimiter=',', quotechar='"', on_bad_lines='skip')
print(data.head())
print(data.info())
print(data.isnull().sum())

data=data.drop_duplicates()
print(data.dtypes)

freq=data['type'].value_counts()
fig, axes=plt.subplots (1,2, figsize=(8, 4))
sns.countplot(data, x=data['type'], ax=axes[0])
plt.pie(freq, labels=['Movie', 'TV Show'], autopct='%.0f%%')
plt.suptitle('Total Content on Netflix', fontsize=20)
plt.show()

data.info()


rating_types = ['TV-MA', 'TV-14', 'TV-PG', 'R', 'PG-13', 'TV-Y7', 'TV-Y', 'PG',
                'TV-G', 'NR', 'G', 'TV-Y7-FV', 'NC-17', 'UR']

rating_frequency = [3000, 2000, 1000, 800, 600, 400, 300, 250, 200, 150, 100, 50, 30, 10]
plt.figure(figsize=(10,6))
plt.bar(rating_types, rating_frequency,color='skyblue')
plt.xticks(rotation=45, ha='right')
plt.xlabel("Rating Types")
plt.ylabel("Rating Frequency")
plt.suptitle('Rating on Netflix', fontsize=20)
plt.show()

rating_types = ['TV-MA', 'TV-14', 'TV-PG', 'R', 'PG-13', 'TV-Y7', 'TV-Y', 'PG']
rating_percentage = [38, 26, 10, 9, 6, 4, 4, 3]
plt.figure(figsize=(8,8))
plt.pie(rating_percentage, labels=rating_types, autopct='%1.0f%%', colors=[''#1f77b4', '#ff7f0e', '#2ca02c', '#d62728',
                                                                           '#9467bd', '#8c564b', '#e377c2', '#7f7f7f'])
plt.title('Rating on Netflix')
plt.show()


print(data['country'].value_counts())

countries = ['United States', 'India', 'United Kingdom', 'Pakistan', 'Not Given',
             'Canada', 'Japan', 'South Korea', 'France', 'Spain']
frequency = [3000, 1200, 700, 500, 400, 300, 300, 250, 200, 150]

plt.figure(figsize=(10,6))
plt.bar(countries, frequency,color='skyblue')
plt.xlabel('Country')
plt.ylabel('Frequency')
plt.title('Content Distribution by Country on Netflix')
plt.xticks(rotation=45)
plt.tight_layout()
plt.show()



months = ['Jan', 'Feb', 'Mar', 'Apr', 'May', 'Jun', 'Jul', 'Aug', 'Sep', 'Oct', 'Nov', 'Dec']
movies = [550, 450, 530, 520, 500, 480, 510, 500, 470, 490, 520, 550]
series = [220, 200, 230, 240, 230, 250, 260, 240, 230, 220, 270, 300]

plt.figure(figsize=(8,6))
plt.plot(months, movies, label='Movies', color='skyblue', linewidth=2)
plt.plot(months, series, label='Series', color='orange', linewidth=2)
plt.title('Monthly releases of Movies and TV Shows on Netflix', fontsize=14, weight='bold')
plt.xlabel('Months', fontsize=12)
plt.ylabel('Frequency of releases', fontsize=12)
plt.grid(True)
plt.legend()
plt.show()
```

```
years = [2008, 2010, 2012, 2014, 2016, 2018, 2020]
movies = [0, 50, 100, 200, 600, 1400, 1100]
tv_shows = [0, 50, 100, 150, 300, 500, 400]

plt.figure(figsize=(8,6))
plt.plot(years, movies, label='Movies', color='skyblue', linewidth=2)
plt.plot(years, tv_shows, label='TV Shows', color='orange', linewidth=2)
plt.title('Yearly releases of Movies and TV Shows on Netflix', fontsize=14, weight='bold')
plt.xlabel('Years', fontsize=12)
plt.ylabel('Frequency of releases', fontsize=12)
plt.grid(True)
plt.legend()
plt.show()



genres = [
    'Dramas, International Movies', 'Documentaries', 'Stand-Up Comedy',
    'Comedies, Dramas, International Movies', 'Dramas, Independent Movies, International Movies',
    'Children & Family Movies', 'Children & Family Movies, Comedies',
    'Documentaries, International Movies', 'Children & Family Movies, Romantic Movies',
    'Comedies, International Movies'
]
frequency = [350, 350, 320, 300, 280, 270, 260, 250, 240, 230]


plt.figure(figsize=(10,6))
plt.bar(genres,Frequency, color='skyblue')
plt.title('Top 10 popular genres for movies on Netflix', fontsize=14, weight='bold')
plt.xlabel('Genres', fontsize=12)
plt.ylabel('Movies Frequency', fontsize=12)
plt.xticks(rotation=45, ha="right")
plt.tight_layout()
plt.show()



genres = ['Kids\' TV', 'International TV Shows, TV Dramas', 'International TV Shows',
          'Kids\' TV, TV Comedies', 'Reality TV', 'International TV Shows, TV Dramas',
          'Crime TV Shows, International TV Shows', 'Romantic TV Shows, International TV Shows',
          'Documentaries', 'TV Comedies']
frequency = [200, 120, 110, 105, 100, 95, 90, 85, 80, 75]

plt.figure(figsize=(10,6))
plt.bar(genres, frequency, color='skyblue')
plt.xlabel('Genres')
plt.ylabel('TV Shows Frequency')
plt.title('Top 10 popular genres for TV Shows on Netflix')
plt.xticks(rotation=45, ha='right')
plt.tight_layout()
plt.show()



directors=data['director'].value_counts().reset_index().sort_values(by='count', ascending=False)[1:15]
plt.bar(directors['director'], directors['count'])
plt.xticks(rotation=45, ha='right')


names = ['Rajiv Chilaka', 'Alastair Fothergill', 'Jan Suter', 'Raúl Campos', 'Suhas Kadav',
         'Marcus Raboy', 'Jay Karas', 'Cathy Garcia-Molina', 'Martin Scorsese',
         'Steven Spielberg', 'Youssef Chahine', 'Todd Kauffman', 'Mark Thornton',
         'Don Michael Paul', 'David Dhawan']
frequency = [20, 18, 17, 17, 16, 15, 15, 14, 14, 13, 12, 12, 11, 10, 10]


plt.figure(figsize=(10,6))
plt.bar(names, frequency, color='skyblue')
plt.xlabel('Names')
plt.ylabel('Frequency/Rating')
plt.title('Top Directors by Frequency or Rating')
plt.xticks(rotation=45, ha='right')
plt.tight_layout()
```

```
plt.show()
```