```
1 ! pip install mysql-connector-python
2 ! pip install PyMySQL
3 ! pip install ipython-sql
4 ! pip install mysqlclient
5 ! pip install pandas
6 ! pip install matplotlib
7 ! pip install seaborn
8 ! pip install scikit-learn
```

```
Requirement already satisfied: mysql-connector-python in c:\users\piyush\anaconda3\lib\site-packages (9.1.0)
Requirement already satisfied: PyMySQL in c:\users\piyush\anaconda3\lib\site-packages (1.1.1)
Requirement already satisfied: ipython-sql in c:\users\piyush\anaconda3\lib\site-packages (0.5.0)
Requirement already satisfied: prettytable in c:\users\piyush\anaconda3\lib\site-packages (from ipython-sql) (3.12.0)
Requirement already satisfied: ipython in c:\users\piyush\anaconda3\lib\site-packages (from ipython-sql) (8.27.0)
Requirement already satisfied: sqlalchemy>=2.0 in c:\users\piyush\anaconda3\lib\site-packages (from ipython-sql) (2.0.34)
Requirement already satisfied: sqlparse in c:\users\piyush\anaconda3\lib\site-packages (from ipython-sql) (0.5.3)
Requirement already satisfied: six in c:\users\piyush\anaconda3\lib\site-packages (from ipython-sql) (1.16.0)
Requirement already satisfied: ipython-genutils in c:\users\piyush\anaconda3\lib\site-packages (from ipython-sql) (0.2.0)
Requirement already satisfied: typing-extensions>=4.6.0 in c:\users\piyush\anaconda3\lib\site-packages (from sqlalchemy>=2.0->ipython
Requirement already satisfied: greenlet!=0.4.17 in c:\users\piyush\anaconda3\lib\site-packages (from sqlalchemy>=2.0->ipython-sql) (3
Requirement already satisfied: decorator in c:\users\piyush\anaconda3\lib\site-packages (from ipython->ipython-sql) (5.1.1)
Requirement already satisfied: jedi>=0.16 in c:\users\piyush\anaconda3\lib\site-packages (from ipython->ipython-sql) (0.19.1)
Requirement already satisfied: matplotlib-inline in c:\users\piyush\anaconda3\lib\site-packages (from ipython->ipython-sql) (0.1.6)
Requirement already satisfied: prompt-toolkit<3.1.0,>=3.0.41 in c:\users\piyush\anaconda3\lib\site-packages (from ipython->ipython-sq
Requirement already satisfied: pygments>=2.4.0 in c:\users\piyush\anaconda3\lib\site-packages (from ipython->ipython-sql) (2.15.1)
Requirement already satisfied: stack-data in c:\users\piyush\anaconda3\lib\site-packages (from ipython->ipython-sql) (0.2.0)
Requirement already satisfied: traitlets>=5.13.0 in c:\users\piyush\anaconda3\lib\site-packages (from ipython->ipython-sql) (5.14.3)
Requirement already satisfied: colorama in c:\users\piyush\anaconda3\lib\site-packages (from ipython->ipython-sql) (0.4.6)
Requirement already satisfied: wcwidth in c:\users\piyush\anaconda3\lib\site-packages (from prettytable->ipython-sql) (0.2.5)
Requirement already satisfied: parso<0.9.0,>=0.8.3 in c:\users\piyush\anaconda3\lib\site-packages (from jedi>=0.16->ipython->ipython-
Requirement already satisfied: executing in c:\users\piyush\anaconda3\lib\site-packages (from stack-data->ipython->ipython-sql) (0.8.
Requirement already satisfied: asttokens in c:\users\piyush\anaconda3\lib\site-packages (from stack-data->ipython->ipython-sql) (2.0.
Requirement already satisfied: pure-eval in c:\users\piyush\anaconda3\lib\site-packages (from stack-data->ipython->ipython-sql) (0.2.
Requirement already satisfied: mysqlclient in c:\users\piyush\anaconda3\lib\site-packages (2.2.7)
Requirement already satisfied: pandas in c:\users\piyush\anaconda3\lib\site-packages (2.2.2)
Requirement already satisfied: numpy>=1.26.0 in c:\users\piyush\anaconda3\lib\site-packages (from pandas) (1.26.4)
Requirement already satisfied: python-dateutil>=2.8.2 in c:\users\piyush\anaconda3\lib\site-packages (from pandas) (2.9.0.post0)
Requirement already satisfied: pytz>=2020.1 in c:\users\piyush\anaconda3\lib\site-packages (from pandas) (2024.1)
Requirement already satisfied: tzdata>=2022.7 in c:\users\piyush\anaconda3\lib\site-packages (from pandas) (2023.3)
Requirement already satisfied: six>=1.5 in c:\users\piyush\anaconda3\lib\site-packages (from python-dateutil>=2.8.2->pandas) (1.16.0)
Requirement already satisfied: matplotlib in c:\users\piyush\anaconda3\lib\site-packages (3.9.2)
Requirement already satisfied: contourpy>=1.0.1 in c:\users\piyush\anaconda3\lib\site-packages (from matplotlib) (1.2.0)
Requirement already satisfied: cycler>=0.10 in c:\users\piyush\anaconda3\lib\site-packages (from matplotlib) (0.11.0)
Requirement already satisfied: fonttools>=4.22.0 in c:\users\piyush\anaconda3\lib\site-packages (from matplotlib) (4.51.0)
Requirement already satisfied: kiwisolver>=1.3.1 in c:\users\piyush\anaconda3\lib\site-packages (from matplotlib) (1.4.4)
Requirement already satisfied: numpy>=1.23 in c:\users\piyush\anaconda3\lib\site-packages (from matplotlib) (1.26.4)
Requirement already satisfied: packaging>=20.0 in c:\users\piyush\anaconda3\lib\site-packages (from matplotlib) (24.1)
Requirement already satisfied: pillow>=8 in c:\users\piyush\anaconda3\lib\site-packages (from matplotlib) (10.4.0)
Requirement already satisfied: pyparsing>=2.3.1 in c:\users\piyush\anaconda3\lib\site-packages (from matplotlib) (3.1.2)
Requirement already satisfied: python-dateutil>=2.7 in c:\users\piyush\anaconda3\lib\site-packages (from matplotlib) (2.9.0.post0)
Requirement already satisfied: six>=1.5 in c:\users\piyush\anaconda3\lib\site-packages (from python-dateutil>=2.7->matplotlib) (1.16.
Requirement already satisfied: seaborn in c:\users\piyush\anaconda3\lib\site-packages (0.13.2)
Requirement already satisfied: numpy!=1.24.0,>=1.20 in c:\users\piyush\anaconda3\lib\site-packages (from seaborn) (1.26.4)
Requirement already satisfied: pandas>=1.2 in c:\users\piyush\anaconda3\lib\site-packages (from seaborn) (2.2.2)
Requirement already satisfied: matplotlib!=3.6.1,>=3.4 in c:\users\piyush\anaconda3\lib\site-packages (from seaborn) (3.9.2)
Requirement already satisfied: contourpy>=1.0.1 in c:\users\piyush\anaconda3\lib\site-packages (from matplotlib!=3.6.1,>=3.4->seaborn
Requirement already satisfied: cycler>=0.10 in c:\users\piyush\anaconda3\lib\site-packages (from matplotlib!=3.6.1,>=3.4->seaborn) (0
Requirement already satisfied: fonttools>=4.22.0 in c:\users\piyush\anaconda3\lib\site-packages (from matplotlib!=3.6.1,>=3.4->seabor
Requirement already satisfied: kiwisolver>=1.3.1 in c:\users\piyush\anaconda3\lib\site-packages (from matplotlib!=3.6.1,>=3.4->seabor
Requirement already satisfied: packaging>=20.0 in c:\users\piyush\anaconda3\lib\site-packages (from matplotlib!=3.6.1,>=3.4->seaborn)
Requirement already satisfied: pillow>=8 in c:\users\piyush\anaconda3\lib\site-packages (from matplotlib!=3.6.1,>=3.4->seaborn) (10.4
Requirement already satisfied: pyparsing>=2.3.1 in c:\users\piyush\anaconda3\lib\site-packages (from matplotlib!=3.6.1,>=3.4->seaborn
Requirement already satisfied: python-dateutil>=2.7 in c:\users\piyush\anaconda3\lib\site-packages (from matplotlib!=3.6.1,>=3.4->sea
Requirement already satisfied: pytz>=2020.1 in c:\users\piyush\anaconda3\lib\site-packages (from pandas>=1.2->seaborn) (2024.1)
Requirement already satisfied: tzdata>=2022.7 in c:\users\piyush\anaconda3\lib\site-packages (from pandas>=1.2->seaborn) (2023.3)
Requirement already satisfied: six>=1.5 in c:\users\piyush\anaconda3\lib\site-packages (from python-dateutil>=2.7->matplotlib!=3.6.1,
```

```
1 import  mysql.connector
2 import numpy as np
3 import pandas as pd
4 import matplotlib.pyplot as plt
5 import seaborn as sns
6 from sklearn.ensemble import RandomForestClassifier
7 from sklearn.model_selection import train_test_split
8 from sklearn.preprocessing import LabelEncoder
```

```
1 connection = mysql.connector.connect(
2 user='root',
3 password='1234',
4 host='localhost',
5 database='netflix'
6 )
7 # Create a cursor object to execute SQL queries
8 cursor = connection.cursor()
```

```
1 cursor.execute("SELECT * FROM netflix1")
```

```
1 netflix1 = pd.DataFrame(cursor.fetchall(), columns=[desc[0] for desc in cursor.description])
```

```
1 print(netflix1)
```

```
      show_id    type                                    title  \
0          s1   Movie                     Dick Johnson Is Dead
1          s3 TV Show                                Ganglands
2          s6 TV Show                             Midnight Mass
3         s14   Movie          Confessions of an Invisible Girl
4          s8   Movie                                   Sankofa
..        ...     ...                                       ...
523     s2041 TV Show             Transformers: Cyberverse
524     s2059 TV Show  The Witcher: A Look Inside the Episodes
525     s2124 TV Show                  DêrÃ¨: An African Tale
526     s2223 TV Show                Street Food: Latin America
527     s2237   Movie    Little Singham: Legend of Dugabakka

            director        country  date_added  release_year rating  \
0     Kirsten Johnson  United States    9/25/2021          2020  PG-13
1      Julien Leclercq         France    9/24/2021          2021  TV-MA
2       Mike Flanagan  United States    9/24/2021          2021  TV-MA
3       Bruno Garotti         Brazil    9/22/2021          2021  TV-PG
4        Haile Gerima  United States    9/24/2021          1993  TV-MA
..               ...            ...          ...           ...    ...
523        Not Given       Pakistan   2020-07-09          2020  TV-Y7
524        Not Given       Pakistan   2020-02-09          2020  TV-MA
525        Not Given       Pakistan    8-15-2020          2016  TV-PG
526        Not Given       Pakistan    7-21-2020          2020  TV-PG
527        Not Given       Pakistan    7-15-2020          2020  TV-Y7

        duration                                     listed_in
0         90 min                                 Documentaries
1       1 Season  Crime TV Shows, International TV Shows, TV Act...
2       1 Season             TV Dramas, TV Horror, TV Mysteries
3         91 min             Children & Family Movies, Comedies
4        125 min   Dramas, Independent Movies, International Movies
..           ...                                           ...
523    2 Seasons                                      Kids' TV
524     1 Season              Docuseries, TV Sci-Fi & Fantasy
525     1 Season  International TV Shows, Romantic TV Shows, TV ...
526     1 Season          Docuseries, Spanish-Language TV Shows
527       68 min             Children & Family Movies, Comedies

[528 rows x 10 columns]
```

```
1 netflix1.head(10)
```

	show_id	type	title	director	country	date_added	release_year	rating	duration	listed_in
0	s1	Movie	Dick Johnson Is Dead	Kirsten Johnson	United States	9/25/2021	2020	PG-13	90 min	Documentaries
1	s3	TV Show	Ganglands	Julien Leclercq	France	9/24/2021	2021	TV-MA	1 Season	Crime TV Shows, International TV Shows, TV Act...
2	s6	TV Show	Midnight Mass	Mike Flanagan	United States	9/24/2021	2021	TV-MA	1 Season	TV Dramas, TV Horror, TV Mysteries
3	s14	Movie	Confessions of an Invisible Girl	Bruno Garotti	Brazil	9/22/2021	2021	TV-PG	91 min	Children & Family Movies, Comedies
4	s8	Movie	Sankofa	Haile Gerima	United States	9/24/2021	1993	TV-MA	125 min	Dramas, Independent Movies, International Movies
5	s9	TV Show	The Great British Baking Show	Andy Devonshire	United Kingdom	9/24/2021	2021	TV-14	9 Seasons	British TV Shows, Reality TV
6	s10	Movie	The Starling	Theodore Melfi	United States	9/24/2021	2021	PG-13	104 min	Comedies, Dramas

```
1 print(netflix1.tail())
```

```
     show_id     type                                   title   director  \
523   s2041  TV Show                 Transformers: Cyberverse  Not Given
524   s2059  TV Show  The Witcher: A Look Inside the Episodes  Not Given
525   s2124  TV Show                      DÃ©rÃ¨: An African Tale  Not Given
526   s2223  TV Show                 Street Food: Latin America  Not Given
527   s2237    Movie      Little Singham: Legend of Dugabakka  Not Given

      country date_added  release_year rating   duration  \
523  Pakistan  2020-07-09          2020  TV-Y7  2 Seasons
524  Pakistan  2020-02-09          2020  TV-MA   1 Season
525  Pakistan   8-15-2020          2016  TV-PG   1 Season
526  Pakistan   7-21-2020          2020  TV-PG   1 Season
527  Pakistan   7-15-2020          2020  TV-Y7     68 min

                                             listed_in
523                                          Kids' TV
524                      Docuseries, TV Sci-Fi & Fantasy
525  International TV Shows, Romantic TV Shows, TV ...
526              Docuseries, Spanish-Language TV Shows
527             Children & Family Movies, Comedies
```

```
1 print(netflix1.info())
```

```
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 528 entries, 0 to 527
Data columns (total 10 columns):
 #   Column        Non-Null Count  Dtype
---  ------        --------------  -----
 0   show_id       528 non-null    object
 1   type          528 non-null    object
 2   title         528 non-null    object
 3   director      528 non-null    object
 4   country       528 non-null    object
 5   date_added    528 non-null    object
 6   release_year  528 non-null    int64
 7   rating        528 non-null    object
 8   duration      528 non-null    object
 9   listed_in     528 non-null    object
dtypes: int64(1), object(9)
memory usage: 41.4+ KB
None
```

```
1 print(netflix1.shape)
2
```

```
(528, 10)
```

```
1 netflix1['type'].value_counts()
```
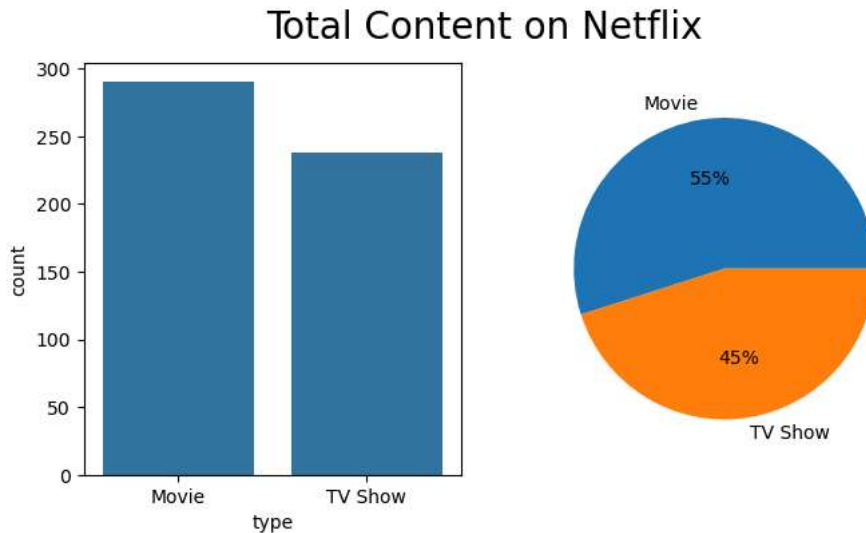
```
type
Movie      290
TV Show    238
Name: count, dtype: int64
```

```
1 # 1. WHAT IS THE TOTAL CONTENT ON NETFLIX?
2
3 freq=netflix1['type'].value_counts()
4 fig, axes=plt.subplots(1,2, figsize=(8, 4))
5 sns.countplot(netflix1, x=netflix1['type'], ax=axes[0])
6 plt.pie(freq, labels=['Movie', 'TV Show'], autopct='%.0f%%')
7 plt.suptitle('Total Content on Netflix', fontsize=20)
```

⇥ Text(0.5, 0.98, 'Total Content on Netflix')


```
1 Start coding or generate with AI.
```

```
1 netflix1['rating'].value_counts()
2
```

⇥
```
rating
TV-MA    154
TV-14     86
TV-PG     60
PG-13     56
TV-Y7     48
R         44
TV-Y      34
PG        28
TV-G      16
G          2
Name: count, dtype: int64
```

```
1 # 2. WHAT IS THE VISUAL REPRESENTAITON OF RATING FREQUENCY OF MOVIES AND TV SHOWS ON NETFLIX?
2
3 ratings=netflix1['rating'].value_counts().reset_index().sort_values(by='count',ascending=False)
4 plt.bar(ratings['rating'], ratings['count'])
5 plt.xticks(rotation=45, ha='right')
6 plt.xlabel("Rating Types")
7 plt.ylabel("Rating Frequency")
8 plt.suptitle('Rating on Netflix', fontsize=20)
```

```
Text(0.5, 0.98, 'Rating on Netflix')
```
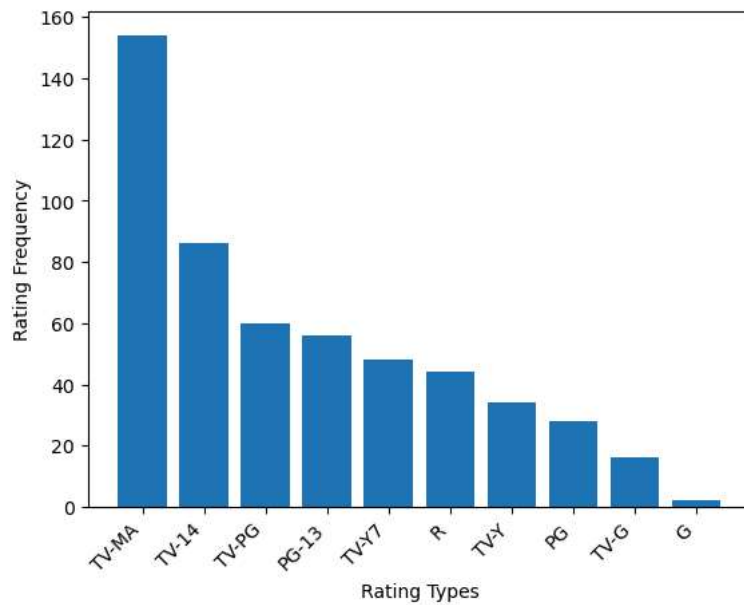


```
1  Start coding or generate with AI.
```

```
1  #Converting date_added column to datetime.
2  netflix1['date_added'] = pd.to_datetime(netflix1['date_added'], format='mixed', errors='coerce')
3
```

```
1  netflix1.describe()
```

	date_added	release_year
count	528	528.000000
mean	2021-05-24 00:19:05.454545408	2014.458333
min	2018-05-18 00:00:00	1961.000000
25%	2021-03-08 18:00:00	2010.750000
50%	2021-07-30 00:00:00	2020.000000
75%	2021-09-10 00:00:00	2021.000000
max	2021-12-06 00:00:00	2021.000000
std	NaN	10.519620

```
1  netflix1['country'].value_counts()
```

```
country
Pakistan          246
United States     136
Not Given          40
Japan              24
India              22
United Kingdom     14
France              8
Germany             8
South Africa        6
China               4
Nigeria             4
Brazil              2
Spain               2
Philippines         2
Australia           2
Argentina           2
Canada              2
Hong Kong           2
```

```
    Italy              2
    Name: count, dtype: int64
```

```
 1  # 3. WHAT ARE THE TOP 10 COUNTRIES WITH MOST CONTENT ON NETFLIX?
 2
 3  top_ten_countries=netflix1['country'].value_counts().reset_index().sort_values(by='count', ascending=False
 4
 5  plt.figure(figsize=(10, 6))
 6  plt.bar(top_ten_countries['country'], top_ten_countries['count'])
 7  plt.xticks(rotation=45, ha='right')
 8  plt.xlabel("Country")
 9  plt.ylabel("Frequency")
10  plt.suptitle("Top 10 countries with most content on Netflix")
11  plt.show()
```


Top 10 countries with most content on Netflix
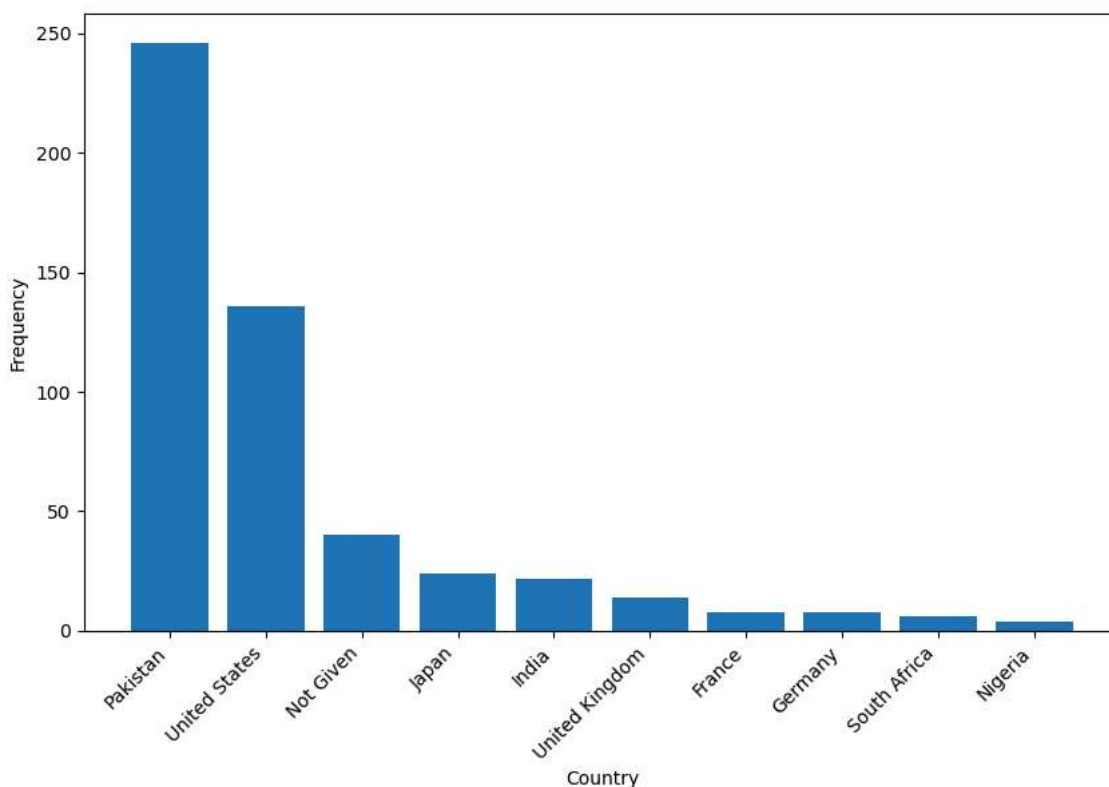
```
 1  Start coding or generate with AI.
```

```
 1  netflix1['year']=netflix1['date_added'].dt.year
 2  netflix1['month']=netflix1['date_added'].dt.month
 3  netflix1['day']=netflix1['date_added'].dt.day
```

```
 1  # 4. WHAT ARE THE MONTHLY RELEASES OF MOVIES AND TV SHOWS ON NETFLIX
 2
 3  monthly_movie_release=netflix1[netflix1['type']=='Movie']['month'].value_counts().sort_index()
 4  monthly_series_release=netflix1[netflix1['type']=='TV Show']['month'].value_counts().sort_index()
 5
 6  plt.plot(monthly_movie_release.index, monthly_movie_release.values, label='Movies')
 7  plt.plot(monthly_series_release.index, monthly_series_release.values,label='Series')
 8  plt.xlabel("Months")
 9  plt.ylabel("Frequency of releases")
10  plt.xticks(range(1, 13), ['Jan', 'Feb', 'Mar', 'Apr', 'May', 'Jun', 'Jul', 'Aug', 'Sep', 'Oct', 'Nov', 'De
11  plt.legend()
12  plt.grid(True)
```

```
13 plt.suptitle("Monthly releases of Movies and TV shows on Netflix")
14 plt.show()
```

Monthly releases of Movies and TV shows on Netflix
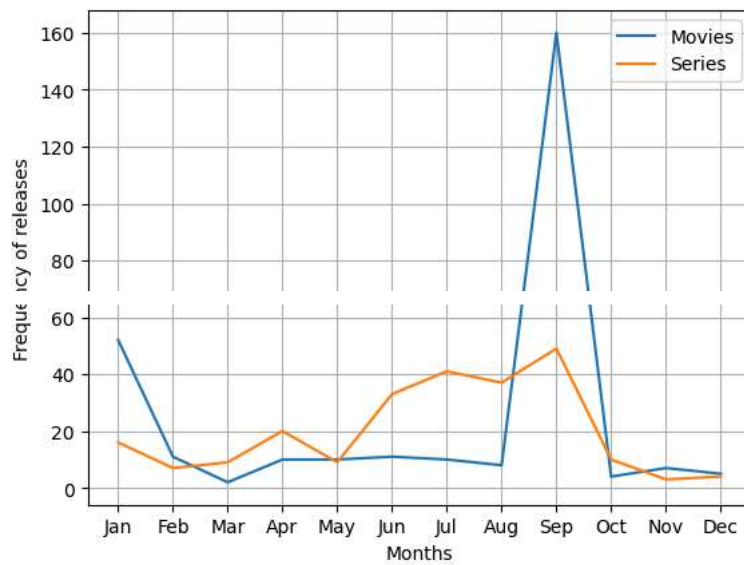

```
1 Start coding or generate with AI.
```

```
1 # 5. WHAT ARE THE YEARLY RELEASES OF MOVIES AND TV SHOWS ON NETFLIX
2
3 yearly_movie_releases=netflix1[netflix1['type']=='Movie']['year'].value_counts().sort_index()
4 yearly_series_releases=netflix1[netflix1['type']=='TV Show']['year'].value_counts().sort_index()
5
6 plt.plot(yearly_movie_releases.index, yearly_movie_releases.values, label='Movies')
7 plt.plot(yearly_series_releases.index, yearly_series_releases.values, label='TV Shows')
8 plt.xlabel("Years")
9 plt.ylabel("Frequency of releases")
10 plt.grid(True)
11 plt.suptitle("Yearly releases of Movies and TV Shows on Netflix")
12 plt.legend()
```

`<matplotlib.legend.Legend at 0x206137e6e70>`

Yearly releases of Movies and TV Shows on Netflix
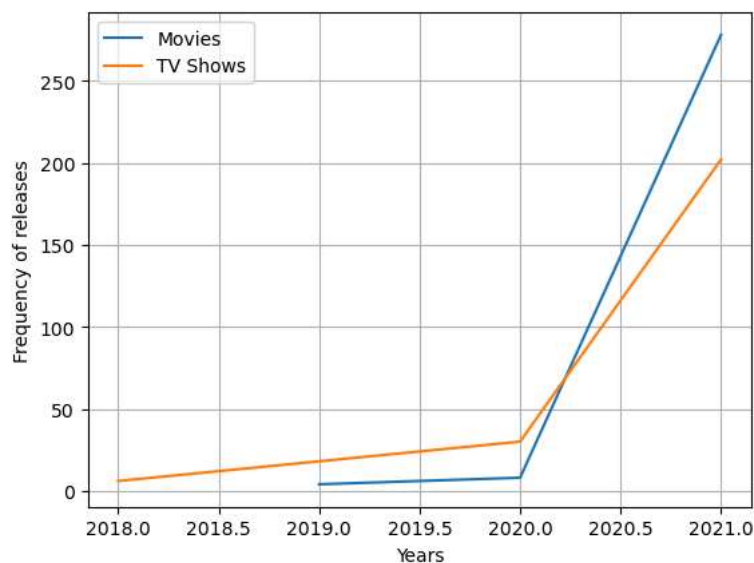
```
1 Start coding or generate with AI.
```

```
1 # 6. WHAT ARE THE TOP 10 POPULAR MOVIE GENRES
2
3 popular_movie_genre=netflix1[netflix1['type']=='Movie'].groupby("listed_in").size().sort_values(ascending=
4
5 plt.bar(popular_movie_genre.index, popular_movie_genre.values)
6 plt.xticks(rotation=45, ha='right')
7 plt.xlabel("Genres")
8 plt.ylabel("Movies Frequency")
9 plt.suptitle("Top 10 popular genres for movies on Netflix")
10 plt.show()
```


Top 10 popular genres for movies on Netflix
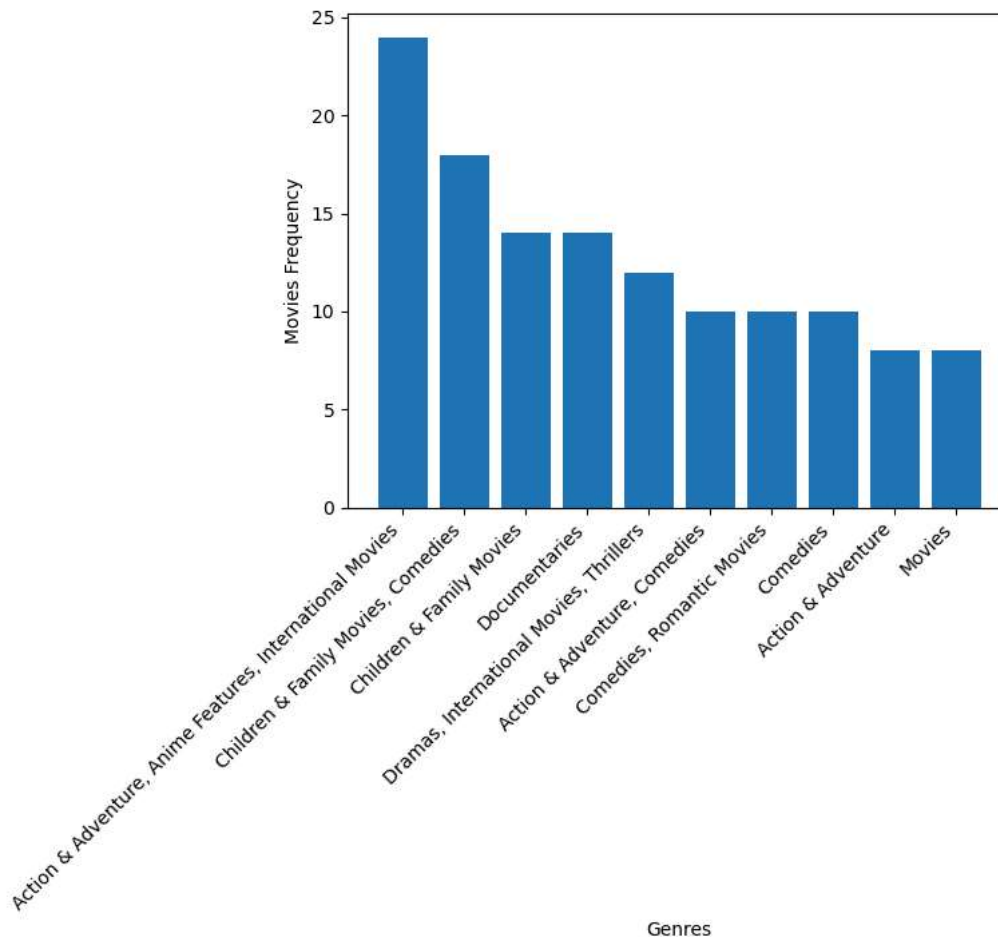
```
1 Start coding or generate with AI.
```

```
1 # 7. WHAT ARE THE TOP 10 DIRECTORS ACROSS NETFLIX WITH HIGH FREQUENCY OF MOVIES AND SHOWS
2
3 directors=netflix1['director'].value_counts().reset_index().sort_values(by='count',ascending=False)[1:15]
4 plt.bar(directors['director'], directors['count'])
5 plt.xticks(rotation=45, ha='right')
6
```

```
([0, 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13],
 [Text(0, 0, 'Suhas Kadav'),
  Text(1, 0, 'Krysia Plonka, Kristian Mercado'),
  Text(2, 0, 'Mark Thornton, Todd Kauffman'),
  Text(3, 0, 'Toshiya Shinohara'),
  Text(4, 0, 'Masahiko Murata'),
  Text(5, 0, 'Wolfgang Petersen'),
  Text(6, 0, 'Antoine Fuqua'),
  Text(7, 0, 'Hajime Kamegaki'),
  Text(8, 0, 'Lasse Hallström'),
  Text(9, 0, 'Delhiprasad Deenadayalan'),
  Text(10, 0, 'Reginald Hudlin'),
  Text(11, 0, 'Jane Campion'),
  Text(12, 0, 'Chapman Way, Maclain Way'),
  Text(13, 0, 'Christian Schwochow')])
```
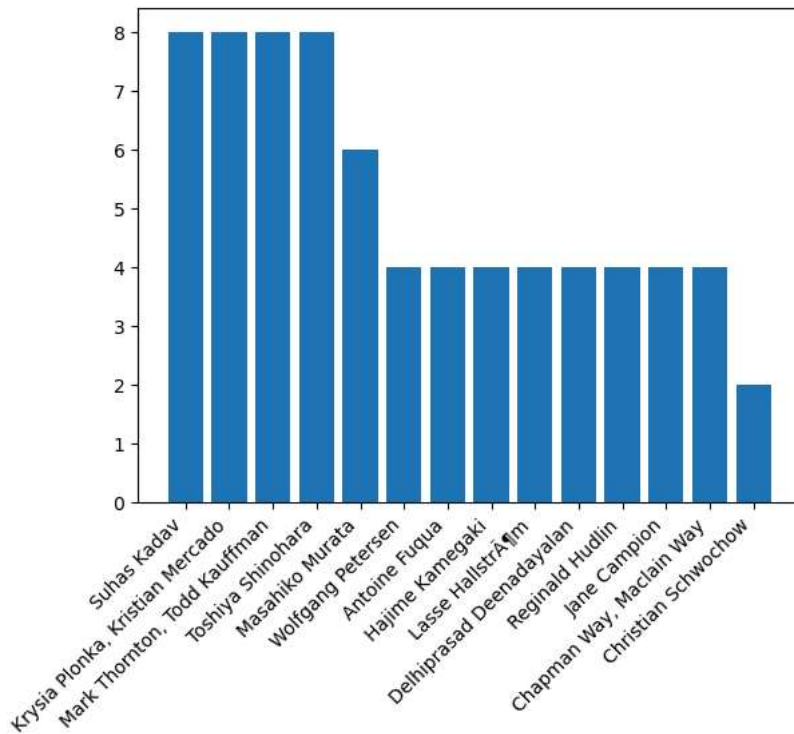


```
1 Start coding or generate with AI.
```

```
1 Start coding or generate with AI.
```

```
 1 # 8. HOW MANY MOVIES AND TV SHOWS WERE ADDED EACH YEAR?
 2
 3 netflix1['year_added'] = pd.to_datetime(netflix1['date_added']).dt.year
 4 yearly_additions = netflix1.groupby('year_added').size()
 5 yearly_additions.plot(kind='line', figsize=(10, 6), marker='o')
 6 plt.title('Movies and TV Shows Added Each Year')
 7 plt.xlabel('Year')
 8 plt.ylabel('Content Count')
 9 plt.show()
10
```

Movies and TV Shows Added Each Year
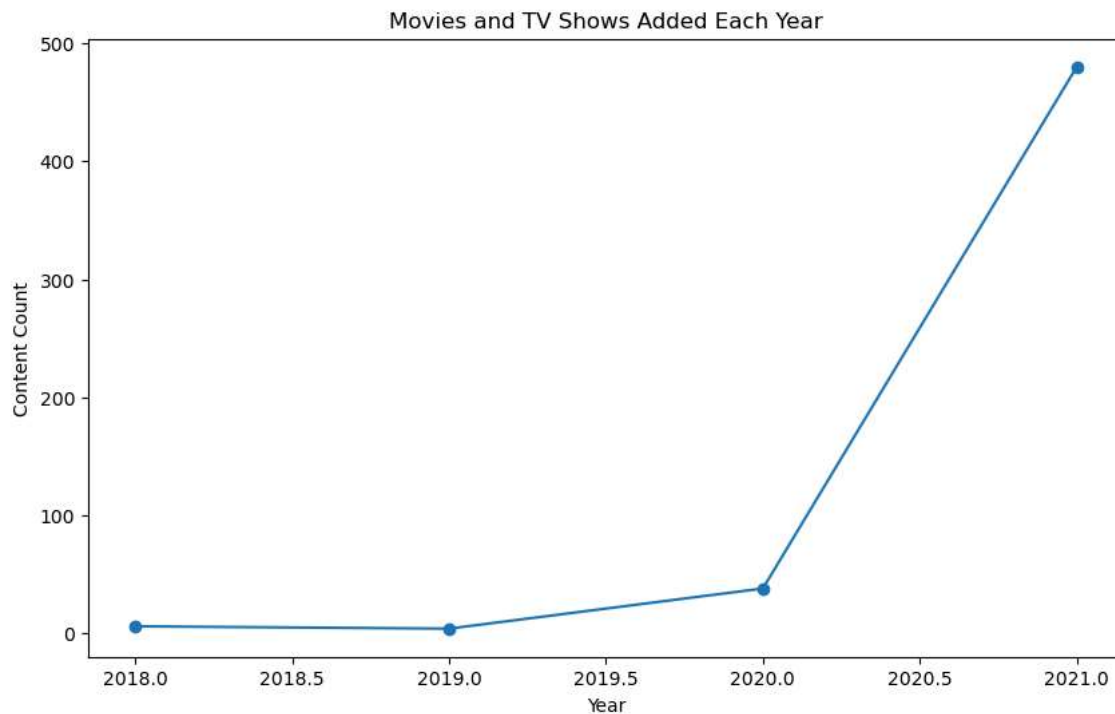

```
1 Start coding or generate with AI.
```

```
1 # 9. WHAT IS THE CORRELATON BETWEEN MOVIE DURATION AND RELEASE YEAR
2
3 plt.scatter(netflix1['release_year'], netflix1['duration_in_minutes'], alpha=0.5, color='purple')
4 plt.title('Movie Duration vs Release Year')
5 plt.xlabel('Release Year')
6 plt.ylabel('Duration (Minutes)')
7 plt.show()
```

Movie Duration vs Release Year
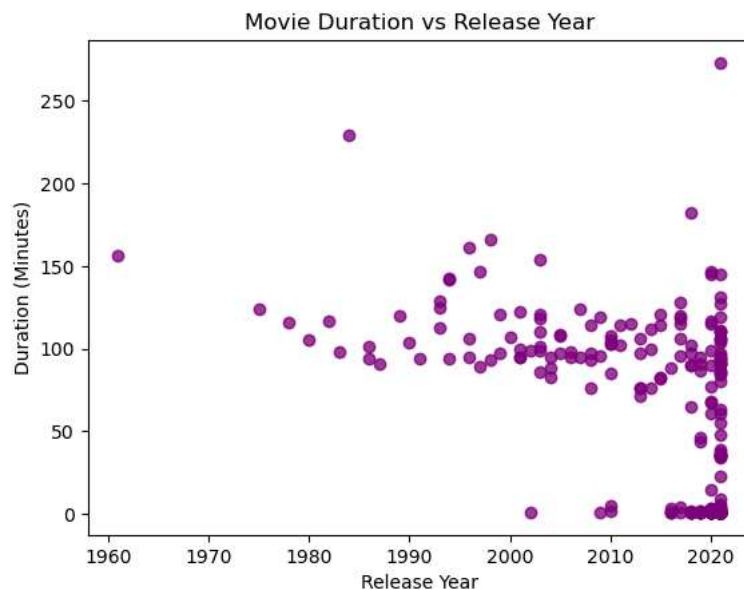

```
1 Start coding or generate with AI.
```
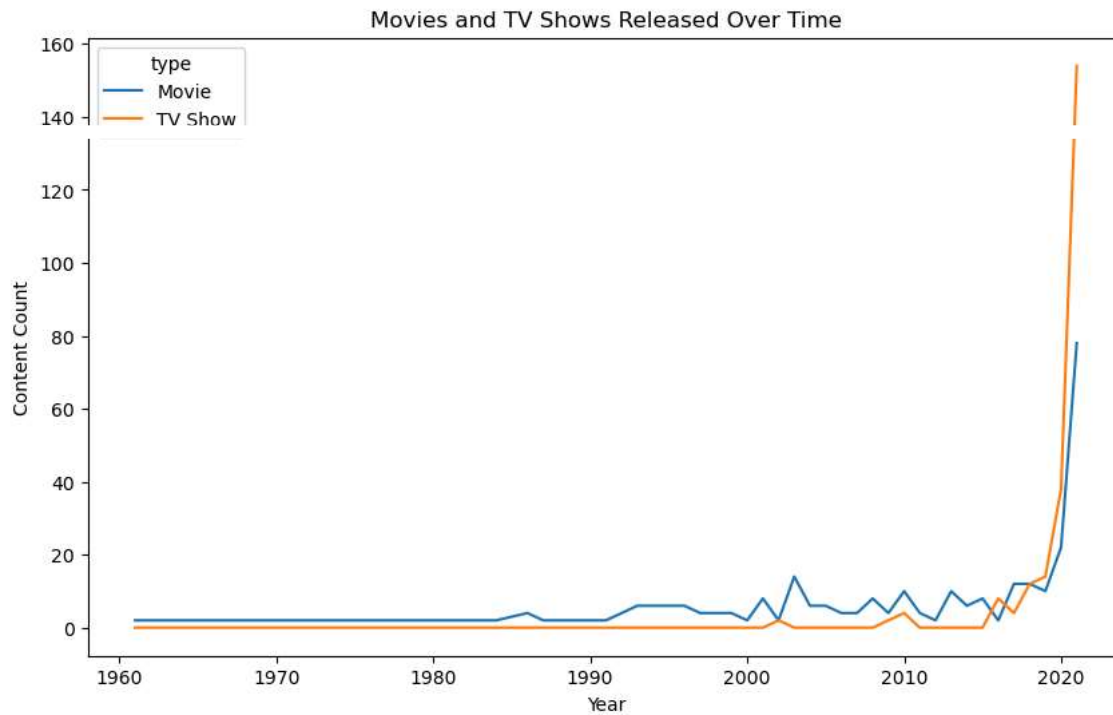
```
1 # 10. WHAT IS THE TREND OF TV SHOWS AND MOVIE RELEASES OVER TIME
2
3 tv_movies_over_time = netflix1.groupby(['release_year', 'type']).size().unstack().fillna(0
4 tv_movies_over_time.plot(kind='line', figsize=(10, 6))
```

```
5 plt.title('Movies and TV Shows Released Over Time')
6 plt.xlabel('Year')
7 plt.ylabel('Content Count')
8 plt.show()
```



```
1 Start coding or generate with AI.
```

```
1 # 11. WHAT IS THE DISTRIBUTION OF CONTENT RATINGS ACROSS DIFFERENT CONTENT TYPES?
2
3 sns.countplot(x='rating', hue='type', data=netflix1, palette='Set1')
4 plt.title('Content Rating Distribution for Movies vs TV Shows')
5 plt.xlabel('Rating')
```