Ex. NO: 6
6/9/25

EXPLORATORY DATA ANALYSIS WITH PYTHON.

AIM:

To analyze the distribution of Netflix content based on type, release year, country, and genre, and to identify trends in content addition over time.

PROGRAM CODE:

```
import pandas as pd
import numpy as np
import matplotlib.pyplot as plt
import seaborn as sns
df = pd.read_csv("netflix_titles.csv")
Print(df.info())
print(df.head())
Print(df.describe(include='all'))
print("Number of unique countries:",
                df['country'].nunique())
print("Number of unique directors:",
                df['director'].nunique())
Print(df['type'].value_counts())
print(df['release_years'].value_counts(),
                head())
print(df.groupby(['country', 'type']),
                size().sort_values
                (ascending=False).head(10))
df['date_added'] = pd.to_datetime(df['date_added']
                format='mixed', errors='coerce')
```

Ex. NO: 6
6/9/25

AIM:

based
genr
add

PRO

OUTPUT:

<class 'pandas.core.frame.DataFrame'>
RangeIndex: 8807 entries, 0 to 8806
Data columns (total 12 columns):

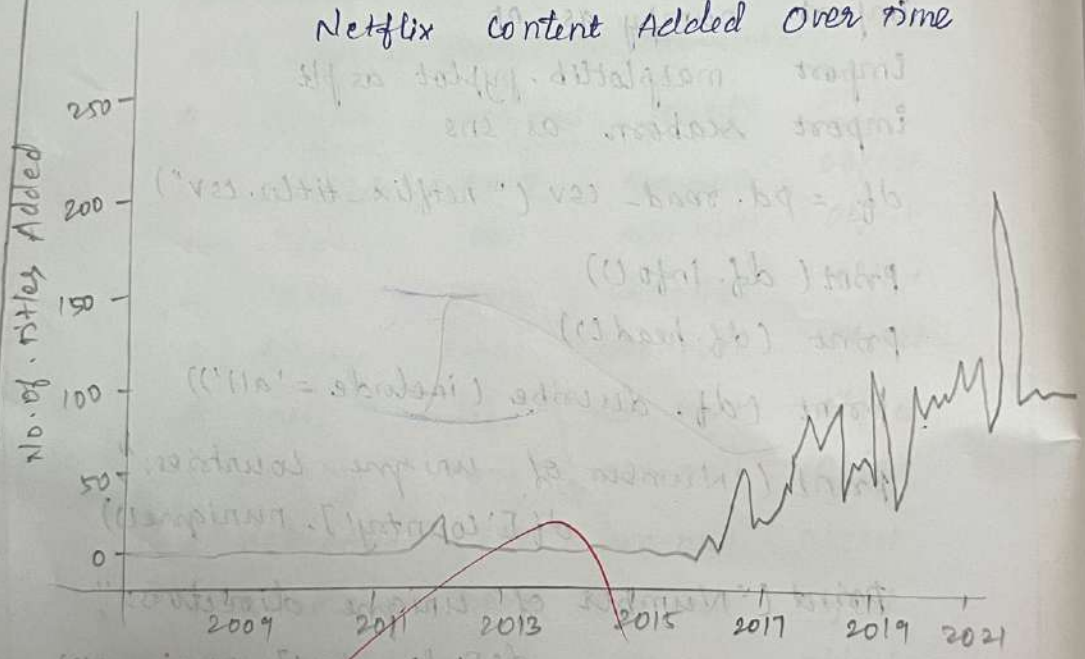
#	Column	Non-Null Count	Dtype
0	show_id	8807	non-null Object
1	type	8807	non-null Object
2	title	8807	non-null Object
3	director	6173	non-null Object
4	cast	7982	non-null Object
5	country	7976	non-null Object
6	date_added	8797	non-null Object
7	release_year	8807	non-null Object
8	rating	8803	non-null Object
9	duration	8804	non-null Object
10	listed_in	8807	non-null Object
11	description	8807	non-null Object

No. of. unique countries: 748
No. of. unique directors: 4528

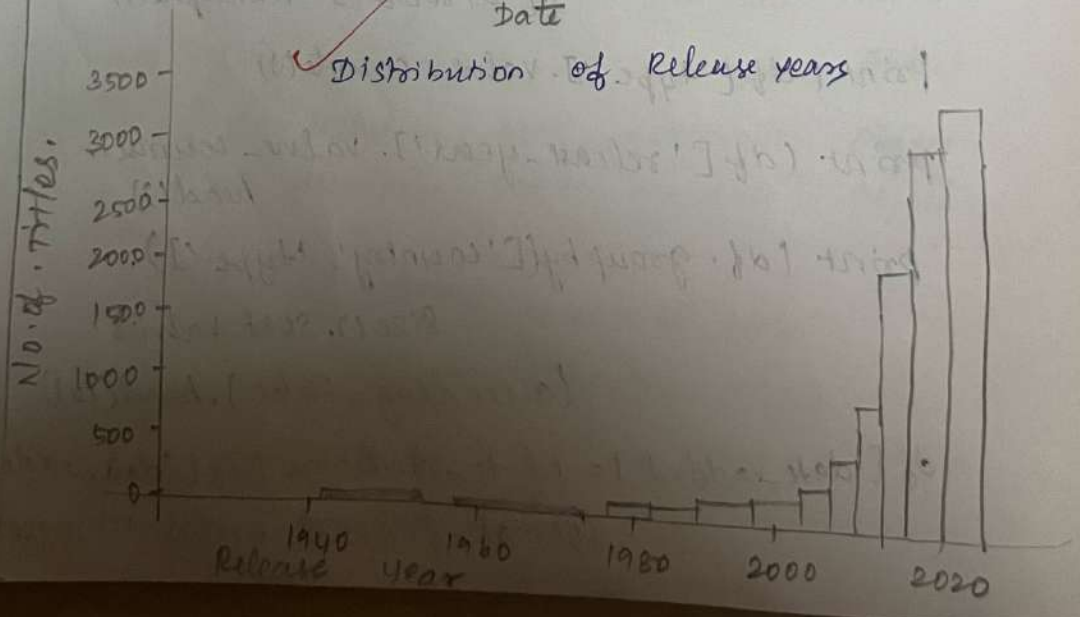
Name: count, dtype: int64

country	type	
United States	Movie	2058
India	Movie	893
United States	TV Show	760
United Kingdom	TV show	760
Spain	Movies	97
Egypt	Movie	92

Netflix Content Added Over Time

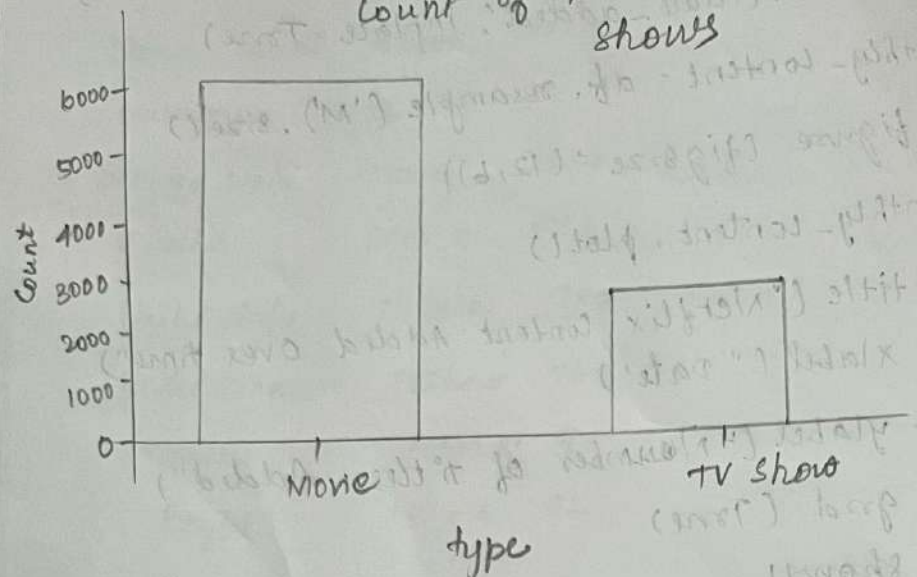


Distribution of Release years

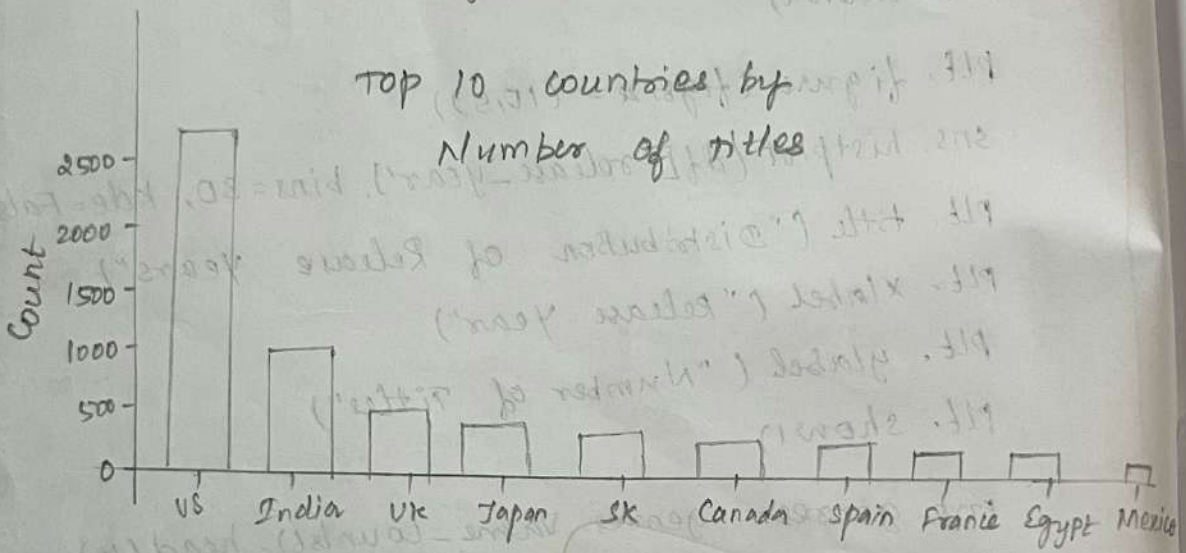


```
df.dropna(subset=['date_added'], inplace=True)
df.set_index('date_added', inplace=True)
monthly_content = df.resample('M').size()
plt.figure(figsize=(12,6))
monthly_content.plot()
plt.title("Netflix Content Added Over Time")
plt.xlabel("Date")
plt.ylabel("Number of Titles Added")
plt.grid(True)
plt.show()

plt.figure(figsize=(10,5))
sns.histplot(df['release_year'], bins=30, kde=False)
plt.title("Distribution of Release Years")
plt.xlabel("Release Year")
plt.ylabel("Number of Titles")
plt.show()

top_genres = genres.value_counts().head(10)
top_genres.plot(kind='barh', color='coral')
plt.title("Top 10 Genres on Netflix")
plt.xlabel("Count")
plt.gca().invert_yaxis()
plt.show()
```

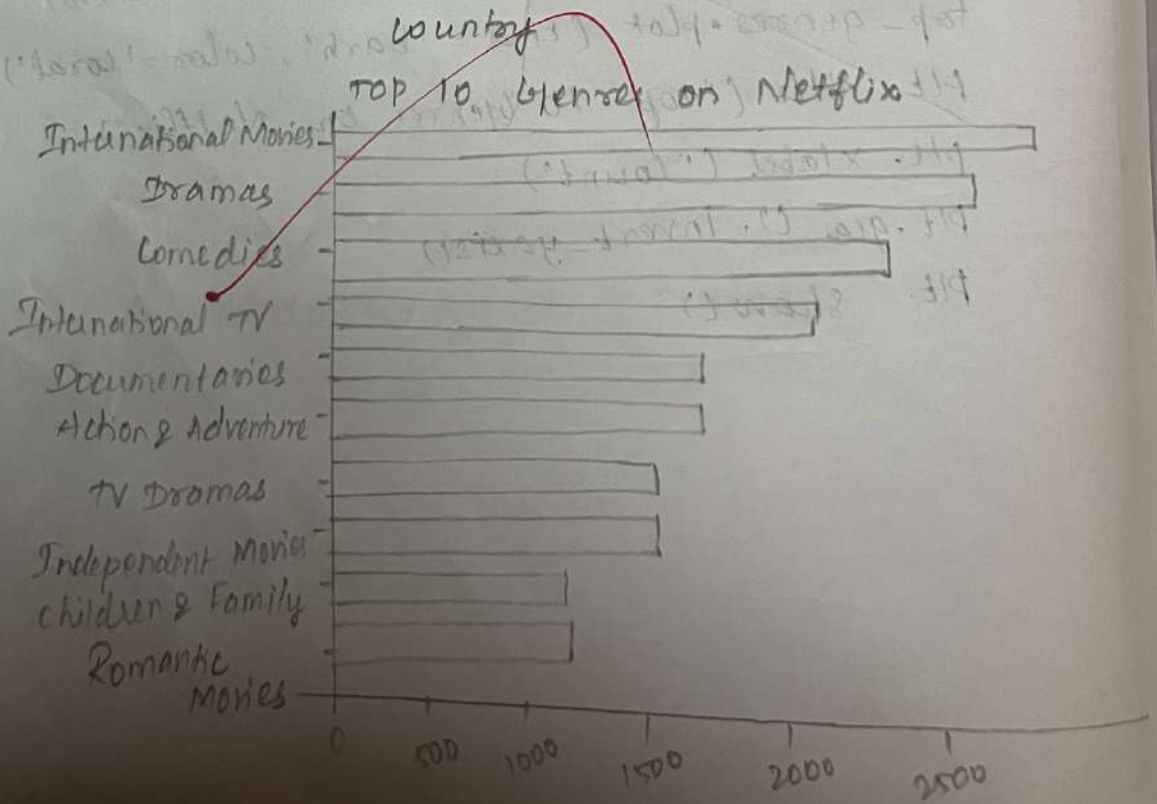
Count of Movies and TV shows



Top 10 countries by Number of titles



Top 10 Genres on Netflix



RESULT:

The analysis reveals trends in Netflix's content library, highlighting the growth of content over time, top countries, popular genres, and the balance between Movies & TV shows.