

Aim

use python libraries (numpy, pandas) to analyze a dataset. To perform descriptive statistics and visualize data

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(df['type'].value_counts())
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("No. of Titles")
plt.show()
```

### Output

Data columns (total 12 columns):

#	columns	non-null count	dtype
0	show_id	8807 non null	object
1	type	8807 non null	object
2	title		
3	director		
4	cast		
5	country		
6	data_added		
7	release_year		
8	rating		

	showid	type	title	director
count	8807	8807	8807	6173
unique	8807	2	8807	4528
max			Nan	

Number of unique countries : 748

Number of unique directors : 4528

type

Movie

TV show

6131

2676

Name : count, dtype: int64

sns. pairplot (data = df, x = 'Type', palette = 'tab10')  
plt. title (" count of movie & TV shows")  
plt. show ()

top\_countries = df [ 'country' ]. value - counts ()

top\_countries . plot ( kind = 'bar' , color = 'grayblue' )  
head (10)

plt. title (" Top 10 countries by number of titles")

plt. ylabel (" count ")

plt. xlabel ("titles")

plt. show ()

genres = df [ 'listed\_in' ]. str . split ( ' , ' ,

expand = True ). stack ()

top\_genres = genres . value - counts () . head (10)

top\_genres . plot ( kind = 'barh' , color = " teal " )

plt. title ("Top 10 genres on Netflix")

plt. xlabel ("count")

plt. gca () . invert\_yaxis ()

plt. show ()

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

df [ 'release\_year' ]. value - counts () . sort\_index ()  
plot ( kind = 'line' )

plt. title (" Number of Netflix titles Released  
per year ")

plt. xlabel (" Year ")

plt. grid ( True )

plt. show ()

release year

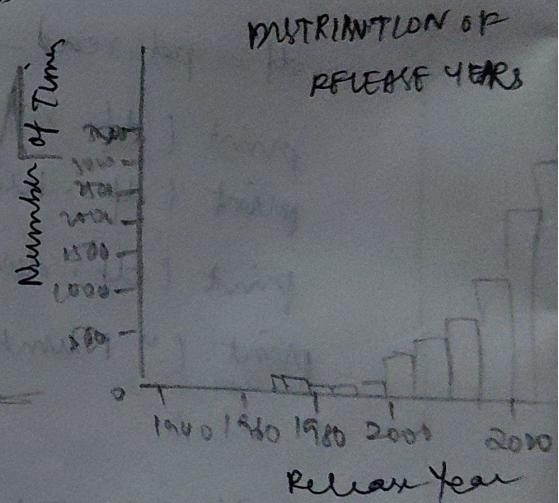
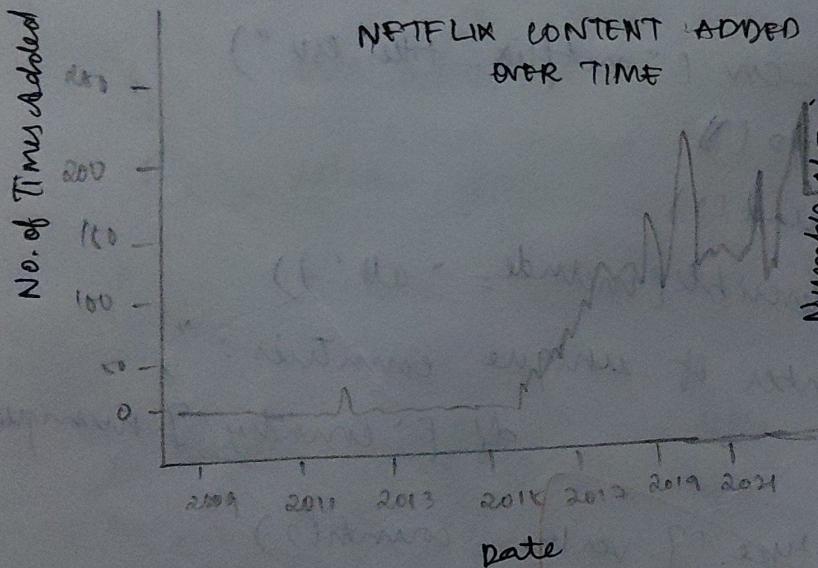
2018	1147
2017	1032
:	:
:	:

Name: count, dtype: int64

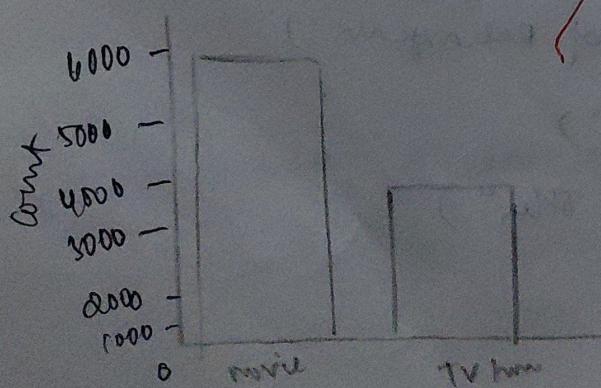
country	type	
US	Movie	2088

India	Movie	893
:	:	:

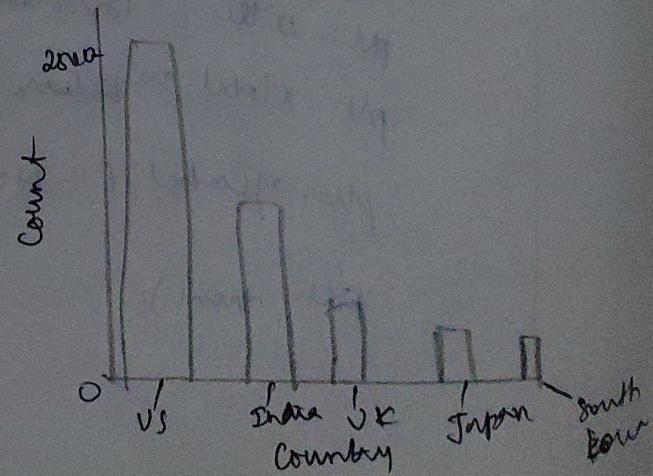
dtype: int64



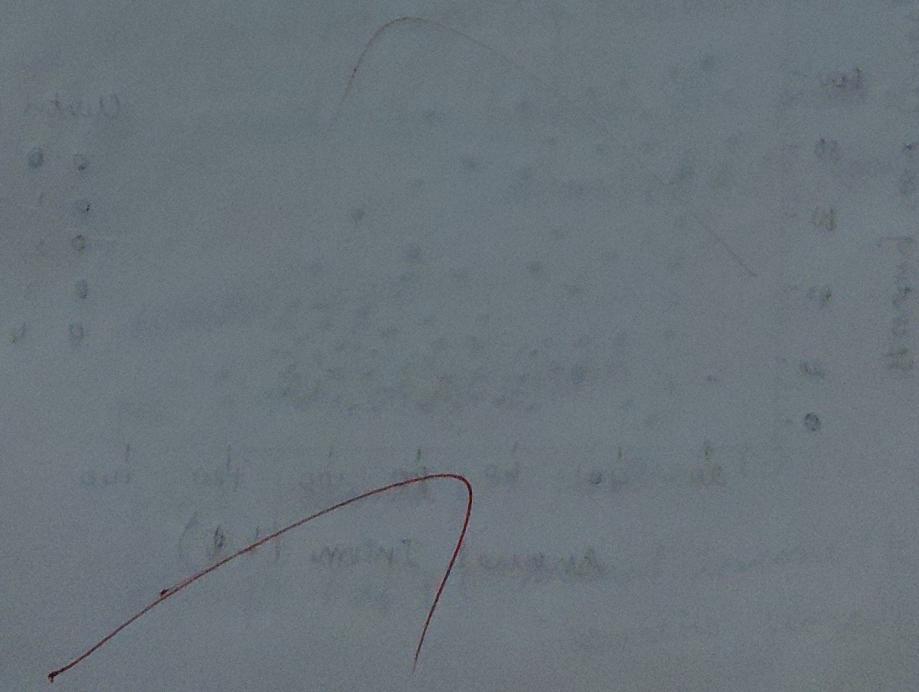
Count of Movies & TV Shows



Top 10 countries by Number of Titles



as well as no further analysis



RESULT  
\*  
THUS the BDA has been done successfully