Exp 6: Exploratory Data Analysis with python

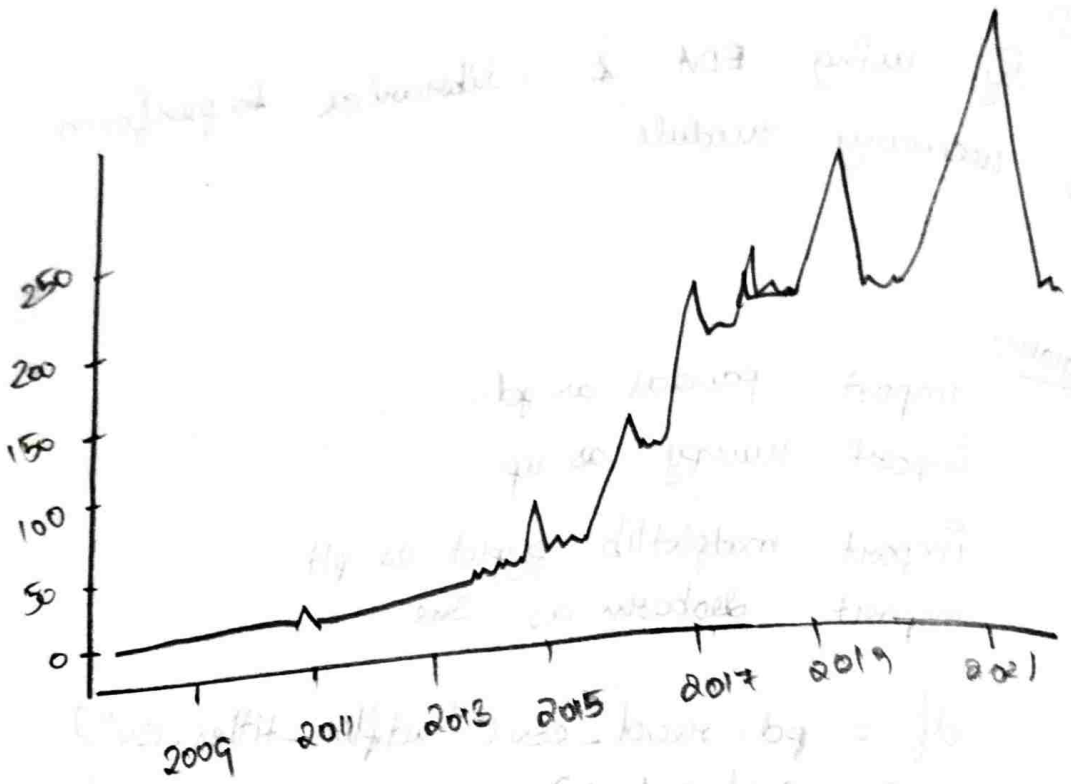
Aim:

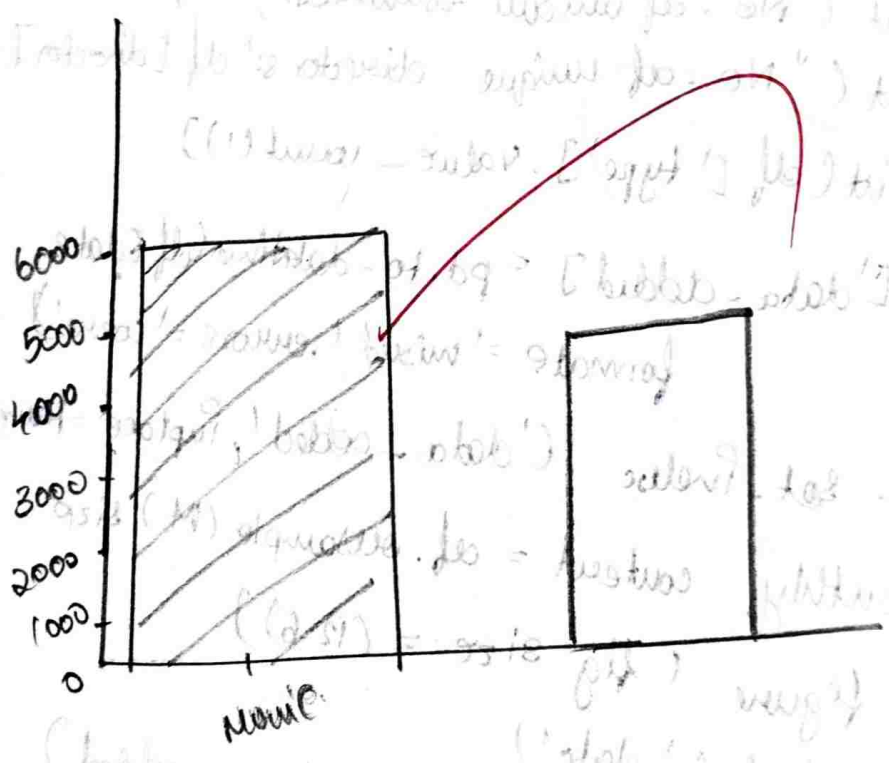
By using EDA & libraries to perform the necessary module.

Program:

```
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("No. of unique countries:", df['country'])
print("No. of unique director s:" df['director']
print(df['type'].value_counts())
df['data_added'] = pd.to_datetime(df['date
                    format='mixed', errors='coerce')
df.set_index('data_added', inplace=True)
monthly_content = df.resample('M').size
plt.figure(fig size = (12,6))
plt.xlabel('date')
plt.ylabel("Number of titles added")
plt.grid(True)
```


250
200
150
100
50
0
2009
2011
2013
2015
2017
2019
2021


6000
5000
4000
3000
2000
1000
0
Music

Content type count

```
sns.countplot(data=df, x='type', palette='Set2')
plt.title('Count of movies tv shows')
plt.show()
```

Top 10 countries with most count.

```
top_countries = df['country'].value_counts().
top_countries.plot(kind='bar', color='sky blue')
plt.title("Top 10 countries by titles")
```

Genre Frequency.

```
genres = df['listed in'].str.split(',', expand=True)
         .stack()

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

Dec 2) 19/25

Result: Result the required programming for EDA has been executed succesfully.