

AIM:
To do exploratory data analysis with Python.

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(df.describe(include="all"))
print("Number of unique countries",
      df["countries"].nunique())
print("number of unique directors",
      df["directors"].nunique())
print(df["type"].value_counts())
print(df["release_year"].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")
```

output :

<class 'pandas.core.frame.DataFrame'
Range-index: 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 obj
2	title	8807	non-null obj
3	director	6173	non-null obj
4	cast	7982	non-null obj
5	country	7976	non-null obj
6	date added	8797	non-null obj
7	release-year	8807	non-null obj
8	rating	8803	non-null obj

dtypes = int64(1), object(4)

no. of unique countries = 748

no. of unique directors = 4528

dtype

movie 6131

TV Shows 2676

name = count, dtype = int64.


```

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("no. of titles added")
plt.grid(True)
plt.show()

```

```

plt.figure(figsize=(10, 5))
sns.histplot(df['release-year'], bins=30,
             rix=False)
plt.xlabel("Release years")
plt.ylabel("NO. of titles")
plt.show()

```

```

sns.countplot(data=df, x='type', palette='set3')
plt.show()

```

```

top_countries = df['country'].value_counts()
head(10)

```

```

plt.ylabel("count")
plt.xticks(rotation=45)
plt.show()

```

```
genres = df["listed_in"].str.split(',')  
expand = True.stack()  
top_genres = genres.value_counts().head(10)  
top_genres = genres  
top_genres.plot(kind='bar', title='total')  
plt.title("Top 10 genres on netflix")  
plt.xlabel("count")  
plt.gca().insert_yaxis(1)  
plt.show()
```

RESULT: Netflix's library is mainly movies with most content from the us, India and the UK. Recent year show peak release. Popular genre includes Dramas. missing data needs attention is done and Successfully completed.