

## Ex no 6. Exploratory Data Analysis with python.

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-files.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-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')
```

(movie) dataframe of columns = columns - 10  
o/p -> columns = 10

< class: pandas.core.frame.data frame >

Range - index - 8807 entries, 0 to 8806

Data columns (total 10 columns)

#	column	Non-null count	dtype
0	showid	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	8297	non-null object
7	release-year	8807	non-null object
8	rating	8803	non-null object

dtypes = int 64 (1), object (4)

no. of unique countries = 748

no. of unique directors = 4528

type  
movie 6131  
TV show 2676

Name = count, dtype = int 64

```

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()

```

```

genres = df["listed-in"].str.split(',', expand=True)
top_genres = genres.value_counts().head(10)
top_genres = plot(kind='bar', label='label')
plt.title("Top 10 genres on netflix")
plt.xlabel("count")
plt.gca().invert_yaxis()
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 genres includes Documentaries. Missing data needs attention is done & successfully completed.