```
import numpy as np ## linear operations
import pandas as pd ##used for data preperation
import plotly.express as px ##used for data visualisation
from textblob import TextBlob #3used for sentiment analysis

df=pd.read_csv("netflix_titles.csv")
```

```
df
```

| | show_id | type | title | director | cast | country | date_added | release_year | rating | duration | listed_in | description |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| **0** | s1 | Movie | Dick Johnson Is Dead | Kirsten Johnson | NaN | United States | September 25, 2021 | 2020 | PG-13 | 90 min | Documentaries | As her father nears the end of his life, filmm... |
| **1** | s2 | TV Show | Blood & Water | NaN | Ama Qamata, Khosi Ngema, Gail Mabalane, Thaban... | South Africa | September 24, 2021 | 2021 | TV-MA | 2 Seasons | International TV Shows, TV Dramas, TV Mysteries | After crossing paths at a party, a Cape Town t... |
| **2** | s3 | TV Show | Ganglands | Julien Leclercq | Sami Bouajila, Tracy Gotoas, Samuel Jouy, Nabi... | NaN | September 24, 2021 | 2021 | TV-MA | 1 Season | Crime TV Shows, International TV Shows, TV Act... | To protect his family from a powerful drug lor... |
| **3** | s4 | TV Show | Jailbirds New Orleans | NaN | NaN | NaN | September 24, 2021 | 2021 | TV-MA | 1 Season | Docuseries, Reality TV | Feuds, flirtations and toilet talk go down amo... |
| **4** | s5 | TV Show | Kota Factory | NaN | Mayur More, Jitendra Kumar, Ranjan | India | September 24, 2021 | 2021 | TV-MA | 2 Seasons | International TV Shows, Romantic TV Shows, TV ... | In a city of coaching centers known to |

Next steps: View recommended plots New interactive sheet

```
df.shape
```

```
(8807, 12)
```

```
df.head()
```

1 to 5 of 5 entries Filter

| index | show_id | type | title | director | cast | country | date_added | release_year | rating | duration | listed_in | description |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| **0** | s1 | Movie | Dick Johnson Is Dead | Kirsten Johnson | NaN | United States | September 25, 2021 | 2020 | PG-13 | 90 min | Documentaries | As her father nears the end of his life, filmmaker Kirsten Johnson stages his death in inventive and comical ways to help them both face the inevitable. |
| **1** | s2 | TV Show | Blood & Water | NaN | Ama Qamata, Khosi Ngema, Gail Mabalane, Thabang Molaba, Dillon Windvogel, Natasha Thahane, Arno Greeff, Xolile Tshabalala, Getmore Sithole, Cindy Mahlangu, Ryle De Morny, Greteli Fincham, Sello Maake Ka-Ncube, Odwa Gwanya, Mekaila Mathys, Sandi Schultz, Duane | South Africa | September 24, 2021 | 2021 | TV-MA | 2 Seasons | International TV Shows, TV Dramas, TV Mysteries | After crossing paths at a party, a Cape Town teen sets out to prove whether a private-school swimming star is her sister who was abducted at birth. |

Next steps: View recommended plots New interactive sheet

```
df.columns ##listing the columns available in the dataset
```

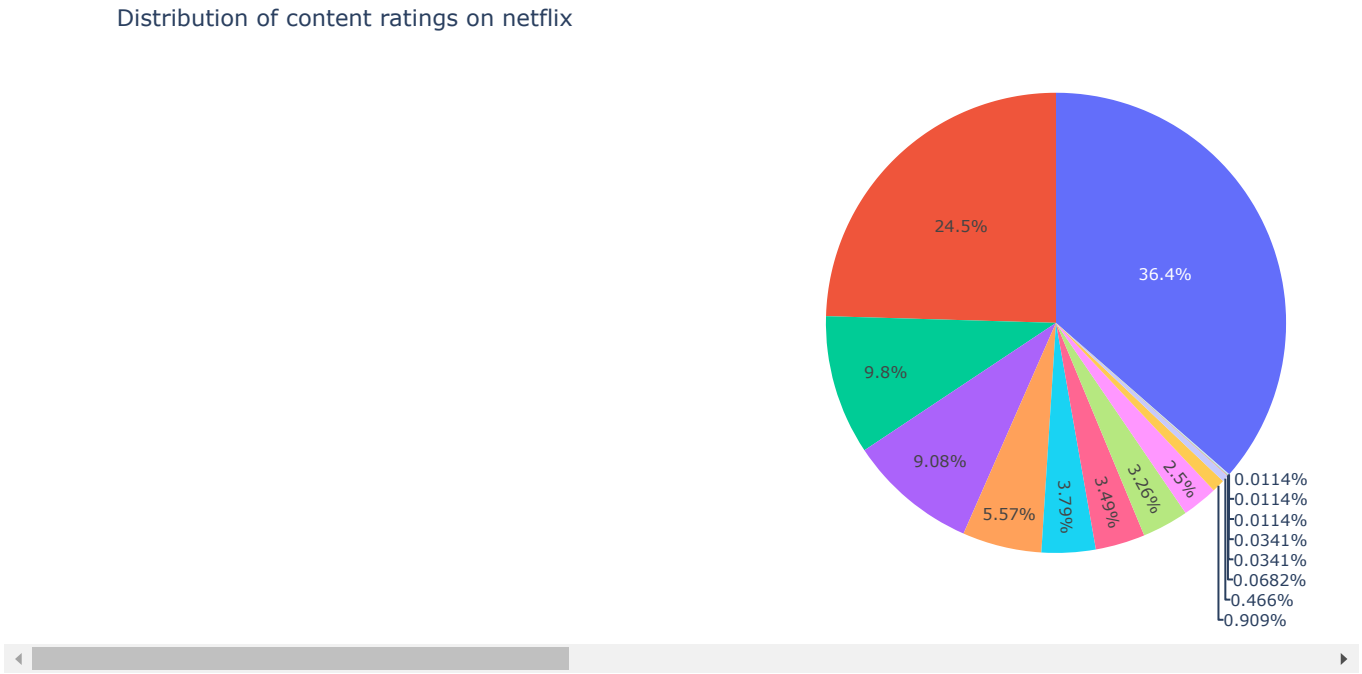
```
Index(['show_id', 'type', 'title', 'director', 'cast', 'country', 'date_added',
       'release_year', 'rating', 'duration', 'listed_in', 'description'],
      dtype='object')
```

```
a=df.groupby(["rating"]).size().reset_index(name="counts")
```

```
print(a)
```

```
       rating  counts
0     66 min       1
1     74 min       1
2     84 min       1
3          G      41
4      NC-17       3
5         NR      80
6         PG     287
7      PG-13     490
8          R     799
9      TV-14    2160
10      TV-G     220
11     TV-MA    3207
12     TV-PG     863
13      TV-Y     307
14     TV-Y7     334
15  TV-Y7-FV       6
16        UR       3
```

```
##plotting a pychart
piechart=px.pie(a,values="counts",names="rating",title="Distribution of content ratings on netflix")
piechart.show()
```




```
df["director"]=df["director"].fillna("Director not specified")##replacing nan values of director with the text d
df.head()
```

| | show_id | type | title | director | cast | country | date_added | release_year | rating | duration | listed_in | description |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| 0 | s1 | Movie | Dick Johnson Is Dead | Kirsten Johnson | NaN | United States | September 25, 2021 | 2020 | PG-13 | 90 min | Documentaries | As her father nears the end of his life, filmm... |
| 1 | s2 | TV Show | Blood & Water | Director not specified | Ama Qamata, Khosi Ngema, Gail Mabalane, Thaban... | South Africa | September 24, 2021 | 2021 | TV-MA | 2 Seasons | International TV Shows, TV Dramas, TV Mysteries | After crossing paths at a party, a Cape Town t... |

Next steps: View recommended plots New interactive sheet

```
directors_list=pd.DataFrame()
print(directors_list)
##creating data frame with no rows and coloumns next we fill data into it
```

```
Empty DataFrame
Columns: []
Index: []
```

```
directors_list=df["director"].str.split(",",expand=True).stack()##seperates the directors by commaa and then usi
print(directors_list)
```

```
0     0          Kirsten Johnson
1     0    Director not specified
2     0          Julien Leclercq
3     0    Director not specified
4     0    Director not specified
                   ...
8802  0             David Fincher
8803  0    Director not specified
8804  0           Ruben Fleischer
8805  0              Peter Hewitt
8806  0               Mozez Singh
Length: 9612, dtype: object
```

```
directors_list
```

| | | 0 |
|---|---|---|
| 0 | 0 | Kirsten Johnson |
| 1 | 0 | Director not specified |
| 2 | 0 | Julien Leclercq |
| 3 | 0 | Director not specified |
| 4 | 0 | Director not specified |
| ... | ... | ... |
| 8802 | 0 | David Fincher |
| 8803 | 0 | Director not specified |
| 8804 | 0 | Ruben Fleischer |
| 8805 | 0 | Peter Hewitt |
| 8806 | 0 | Mozez Singh |

9612 rows × 1 columns

Next steps: View recommended plots New interactive sheet

```
print(type(directors_list))
```

```
<class 'pandas.core.frame.DataFrame'>
```

```
directors_list.columns=["Director"]
print(directors_list)
```

```
                     Director
0    0         Kirsten Johnson
1    0   Director not specified
```

```
2    0         Julien Leclercq
3    0  Director not specified
4    0  Director not specified
...                       ...
8802 0           David Fincher
8803 0  Director not specified
8804 0         Ruben Fleischer
8805 0            Peter Hewitt
8806 0             Mozez Singh

[9612 rows x 1 columns]
```

```
directors = directors_list.groupby(["Director"]).size().reset_index(name="Total Count")
print(directors)
```

```
                     Director  Total Count
0                Aaron Moorhead            2
1                   Aaron Woolf            1
2      Abbas Alibhai Burmawalla            1
3              Abdullah Al Noor            1
4           Abhinav Shiv Tiwari            1
...                         ...          ...
5116               Çagan Irmak            1
5117           Ísold Uggadóttir            1
5118        Óskar Thór Axelsson            1
5119           Ömer Faruk Sorak            2
5120               Şenol Sönmez            2

[5121 rows x 2 columns]
```

```
directors=directors[directors.Director !="Director not specified"]
directors
```

1 to 25 of 5120 entries Filter

| index | Director | Total Count |
|---|---|---|
| 0 | Aaron Moorhead | 2 |
| 1 | Aaron Woolf | 1 |
| 2 | Abbas Alibhai Burmawalla | 1 |
| 3 | Abdullah Al Noor | 1 |
| 4 | Abhinav Shiv Tiwari | 1 |
| 5 | Abhishek Chaubey | 2 |
| 6 | Abosi Ogba | 1 |
| 7 | Aco Tenriyagelli | 1 |
| 8 | Adam Carolla | 3 |
| 9 | Adam Darke | 1 |
| 10 | Adam Egypt Mortimer | 1 |
| 11 | Adam Larson Broder | 1 |
| 12 | Adam Nee | 1 |
| 13 | Adam Schlesinger | 1 |
| 14 | Adirek Wattaleela | 1 |
| 15 | Ajay Bhuyan | 1 |
| 16 | Akihiko Yamashita | 1 |
| 17 | Akiva Schaffer | 2 |
| 18 | Akshay Shankar | 1 |
| 19 | Alain Gagnol | 1 |
| 20 | Alan Hicks | 1 |
| 21 | Alan Mak | 1 |
| 22 | Alan Poul | 1 |
| 23 | Alana Waksman | 1 |
| 24 | Alban Teurlai | 2 |

Show 25 per page

1 2 10 100 200 205

Next steps: View recommended plots New interactive sheet

```
directors=directors.sort_values(by=["Total Count"],ascending=False)##sorting in the Total count coloumn using th
print(directors)
```

```
                Director  Total Count
4021       Rajiv Chilaka           22
261            Jan Suter           18
4068         Raúl Campos           18
3236        Marcus Raboy           16
4652         Suhas Kadav           16
```

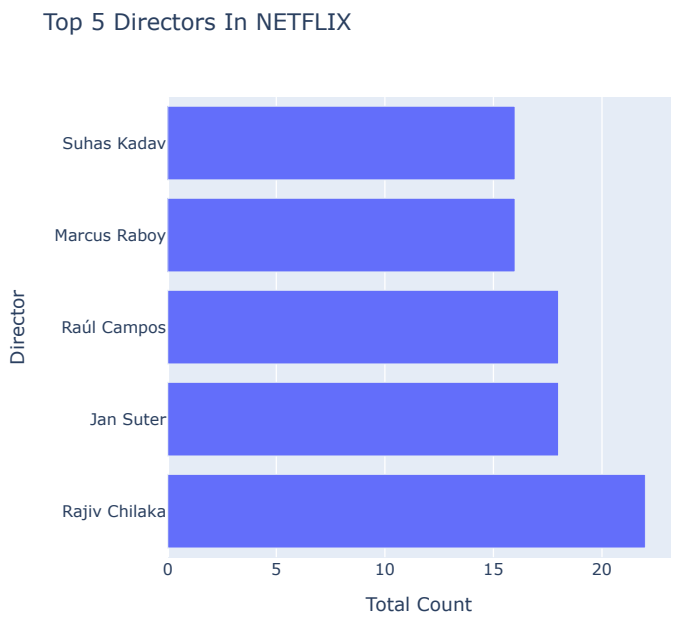
```
...                  ...          ...
4774      Thomas Sorriaux            1
1861       Elliot Hegarty            1
1863  Elsa Flores Almaraz            1
1307          Brian Smith            1
5014        Will Lovelace            1

[5120 rows x 2 columns]
```

```
topDirectors=directors.head()## taking the top 5 directors
topDirectors
```

| | Director | Total Count |
|---|---|---|
| **4021** | Rajiv Chilaka | 22 |
| **261** | Jan Suter | 18 |
| **4068** | Raúl Campos | 18 |
| **3236** | Marcus Raboy | 16 |
| **4652** | Suhas Kadav | 16 |

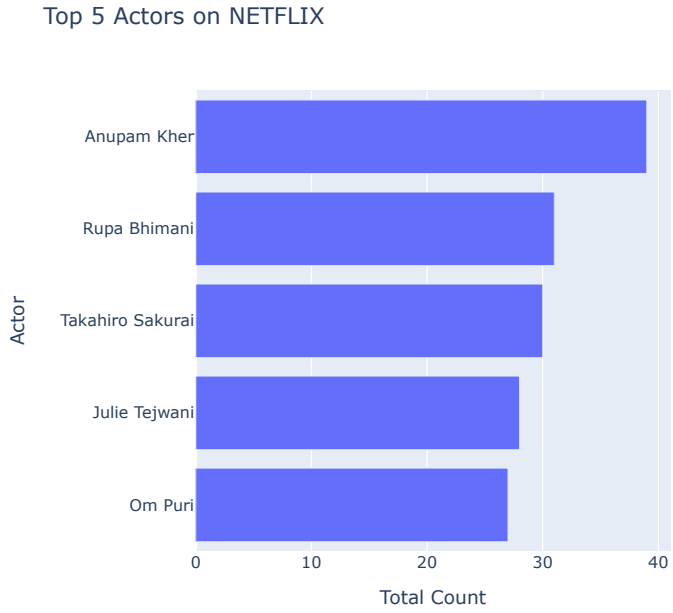
Next steps: View recommended plots New interactive sheet

```
barChart=px.bar(topDirectors,x="Total Count",y="Director",title="Top 5 Directors In NETFLIX")
barChart.show()
```




**Analysing the top 5 Actors/Actresses on NETFLIX**

```
df["cast"]=df["cast"].fillna("No cast specified")
cast_df=pd.DataFrame()
cast_df=df["cast"].str.split(",",expand=True).stack()
cast_df=cast_df.to_frame()
cast_df.columns=["Actor"]
actors=cast_df.groupby(["Actor"]).size().reset_index(name="Total Count")
actors=actors[actors.Actor!="No cast specified"]
actors=actors.sort_values(by=["Total Count"],ascending=False)
topActors=actors.head()
topActors=topActors.sort_values(by=['Total Count'])
barChart2=px.bar(topActors, x="Total Count",y="Actor",title ="Top 5 Actors on NETFLIX")
barChart2.show()
```




**ANALYZING THE CONTENT PRODUCED ON NETFLIX BASED ON YEARS**

```
df1=df[["type","release_year"]]
df1=df1.rename(columns={"release_year":"Release Year","type": "Type"})
df2 = df1.groupby(["Release Year", "Type"]).size().reset_index(name="Total Count")
print(df2)
```

```
     Release Year     Type  Total Count
0            1925  TV Show            1
1            1942    Movie            2
2            1943    Movie            3
3            1944    Movie            3
4            1945    Movie            3
..            ...      ...          ...
114          2019  TV Show          397
115          2020    Movie          517
116          2020  TV Show          436
117          2021    Movie          277
118          2021  TV Show          315

[119 rows x 3 columns]
```

```
df2=df2[df2["Release Year"]>=2000]
graph=px.line(df2,x="Release Year", y="Total Count",color="Type", title="Trend of Content Produced on NETFLIX Ev
graph.show()
```

Trend of Content Produced on NETFLIX Ever

**SENTIMENT ANALYSIS OF NETFLIX CONTENT**

```
df3=df[["release_year","description"]]
df3=df3.rename(columns={"release_year":"Release Year", "description":"Description"})
for index,row in df3.iterrows():
  d=row["Description"]
  testimonal=TextBlob(d)
  p=testimonal.sentiment.polarity
  if p==0:
    sent="Neutral"
  elif p>0:
    sent="Positive"
  else:
    sent="Negative"
  df3.at[index,"Sentiment"]=sent
df3=df3.groupby(["Release Year", "Sentiment"]).size().reset_index(name="Total Count")
df3=df3[df3["Release Year"]>2005]
barGraph=px.bar(df3,x="Release Year", y="Total Count", color="Sentiment", title="SENTIMENT ANALYSIS OF THE CONTE
barGraph.show()
```

SENTIMENT ANALYSIS OF THE CONTENT ON NETFLIX