

Netflix Data Analysis

#I'll start this Netflix data analysis task with Python by importing the dataset and all the Python libraries needed #for this task:

```
In [1]: import pandas as pd
```

```
In [2]: import plotly.express as px
```

```
In [3]: data=pd.read_csv("C:\\Users\\MAMTA\\Documents\\netflix_titles.csv")
```

```
In [4]: data
```

Out[4]:

	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 Raj, Alam K...	India	September 24, 2021	2021	TV-MA	2 Seasons	International TV Shows, Romantic TV Shows, TV ...	In a city of coaching centers known to train l...
...
8802	s8803	Movie		Zodiac	David Fincher	Mark Ruffalo, Jake Gyllenhaal, Robert Downey J...	United States	November 20, 2019	2007	R	158 min	Cult Movies, Dramas, Thrillers	A political cartoonist, a crime reporter and a...
8803	s8804	TV Show		Zombie Dumb	NaN	NaN	NaN	July 1, 2019	2018	TV-Y7	2 Seasons	Kids' TV, Korean TV Shows, TV Comedies	While living alone in a spooky town, a young g...
8804	s8805	Movie		Zombieland	Ruben Fleischer	Jesse Eisenberg, Woody Harrelson, Emma Stone, ...	United States	November 1, 2019	2009	R	88 min	Comedies, Horror Movies	Looking to survive in a world taken over by zo...
8805	s8806	Movie		Zoom	Peter Hewitt	Tim Allen, Courteney Cox, Chevy Chase, Kate Ma...	United States	January 11, 2020	2006	PG	88 min	Children & Family Movies, Comedies	Dragged from civilian life, a former superhero...
8806	s8807	Movie		Zubaan	Mozez Singh	Vicky Kaushal, Sarah-Jane Dias, Raaghav Chanan...	India	March 2, 2019	2015	TV-14	111 min	Dramas, International Movies, Music & Musicals	A scrappy but poor boy worms his way into a ty...

8807 rows × 12 columns

```
In [5]: data.columns
```

```
Out[5]: Index(['show_id', 'type', 'title', 'director', 'cast', 'country', 'date_added', 'release_year', 'rating', 'duration', 'listed_in', 'description'], dtype='object')
```

```
In [6]: data.shape
```

```
Out[6]: (8807,12)
```

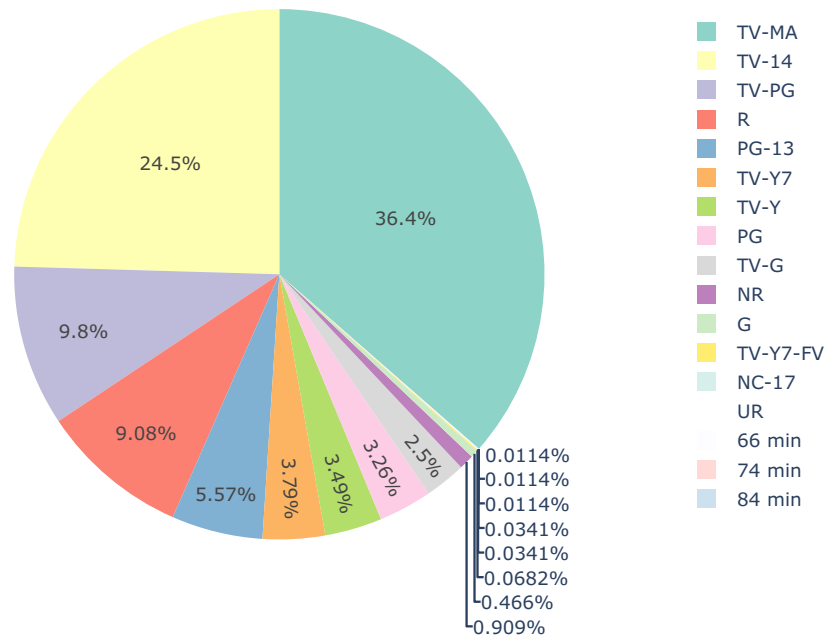
Distribution of Content:

To begin the task of analyzing Netflix data, I'll start by looking at the distribution of content ratings on Netflix:

```
In [7]: z = data.groupby(['rating']).size().reset_index(name='counts')
pieChart = px.pie(z, values='counts', names='rating',
                  title='Distribution of Content Ratings on Netflix',
                  color_discrete_sequence=px.colors.qualitative.Set3)
```

```
In [8]: pieChart
```

Distribution of Content Ratings on Netflix



The graph above shows that the majority of content on Netflix is categorized as TV-MA, which means that most of the content available on Netflix is intended for viewing by mature and adult audiences.

```
In [9]: data.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	NaN	Ama Qamata, Khosi Ngema, Gail Mababane, 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 Raj, Alam K...	India	September 24, 2021	2021	TV-MA	2 Seasons	International TV Shows, Romantic TV Shows, TV ...	In a city of coaching centers known to train l...

```
In [10]: data.groupby(data.rating).count()
```

Out[10]:

	show_id	type	title	director	cast	country	date_added	release_year	duration	listed_in	description
rating											
66 min	1	1	1	1	1	1	1	1	0	1	1
74 min	1	1	1	1	1	1	1	1	0	1	1
84 min	1	1	1	1	1	1	1	1	0	1	1
G	41	41	41	41	40	41	41	41	41	41	41
NC-17	3	3	3	2	3	3	3	3	3	3	3
NR	80	80	80	75	63	80	79	80	80	80	80
PG	287	287	287	286	279	281	287	287	287	287	287
PG-13	490	490	490	489	477	482	490	490	490	490	490
R	799	799	799	795	790	788	799	799	799	799	799
TV-14	2160	2160	2160	1457	1955	1930	2157	2160	2160	2160	2160
TV-G	220	220	220	118	183	190	220	220	220	220	220
TV-MA	3207	3207	3207	2117	2881	2931	3205	3207	3207	3207	3207
TV-PG	863	863	863	538	719	773	861	863	863	863	863
TV-Y	307	307	307	112	268	227	306	307	307	307	307
TV-Y7	334	334	334	132	310	236	333	334	334	334	334
TV-Y7-FV	6	6	6	4	4	5	6	6	6	6	6
UR	3	3	3	3	3	3	3	3	3	3	3

In [11]:

```
data.director=data.director.fillna('no director specified')
```

In [12]:

```
data.director.isnull().sum()
```

Out[12]:

0

In []:

In []:

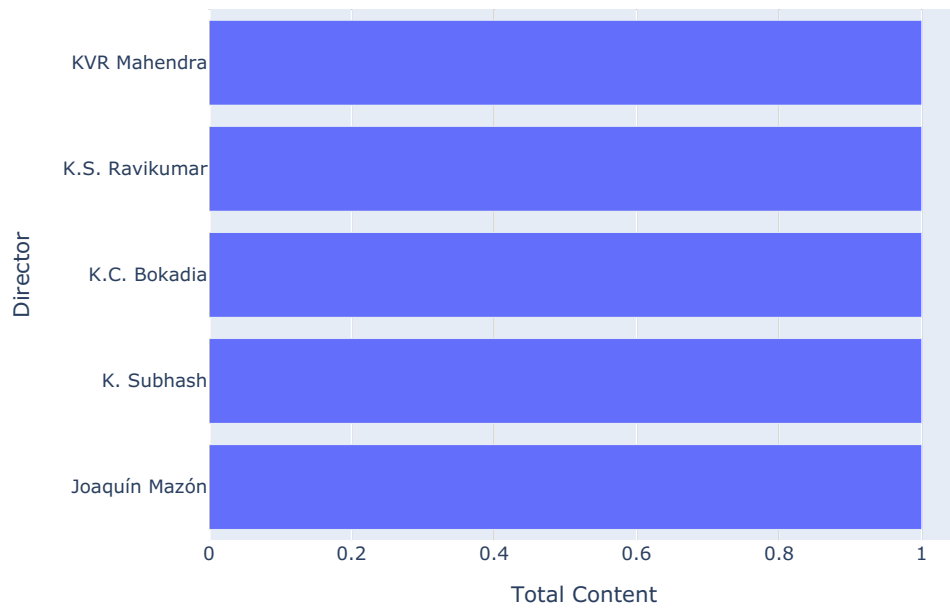
Top 5 Actors and Directors:

Now let's see the top 5 successful directors on this platform:

In [13]:

```
filtered_directors=pd.DataFrame()
filtered_directors=data['director'].str.split(',',expand=True).stack()
filtered_directors=filtered_directors.to_frame()
filtered_directors.columns=['Director']
directors=filtered_directors.groupby(['Director']).size().reset_index(name='Total Content')
directors=directors[directors.Director !='No Director Specified']
directors=directors.sort_values(by=['Total Content'],ascending=True)
directorsTop5=directors.head()
directorsTop5=directorsTop5.sort_values(by=['Total Content'])
fig1=px.bar(directorsTop5,x='Total Content',y='Director',title='Top 5 Directors on Netflix')
fig1.show()
```

Top 5 Directors on Netflix



From the above graph it is derived that the top 5 directors on this platform are:

- 1.KVR Mahendra
- 2.K.S. Ravikumar
- 3.K.C. Bokadia
- 4.K.Subhash
- 5.Joaquin Mazon

In []:

In []:

Now let's have a look at the top 5 successful actors on this platform:

```
In [14]: data['cast']=data['cast'].fillna('No Cast Specified')
filtered_cast=pd.DataFrame()
filtered_cast=data['cast'].str.split(',',expand=True).stack()
filtered_cast=filtered_cast.to_frame()
filtered_cast.columns=['Actor']
actors=filtered_cast.groupby(['Actor']).size().reset_index(name='Total Content')
actors=actors[actors.Actor !='No Cast Specified']
actors=actors.sort_values(by=['Total Content'],ascending=False)
actorsTop5=actors.head()
actorsTop5=actorsTop5.sort_values(by=['Total Content'])
fig2=px.bar(actorsTop5,x='Total Content',y='Actor', title='Top 5 Actors on Netflix')
fig2.show()
```

From the above plot, it is derived that the top 5 actors on Netflix are:

1. Anupam Kher
2. Rupa Bhimani
3. Takahiro Sakurai
4. Julie Tejjwani
5. Om Puri

```
In [15]: data.head()
```

Out[15]:	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	No Cast Specified	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	no director specified	Ama Qamata, Khosi Ngema, Gail Mablane, 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	no director specified	No Cast Specified	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	no director specified	Mayur More, Jitendra Kumar, Ranjan Raj, Alam K...	India	September 24, 2021	2021	TV-MA	2 Seasons	International TV Shows, Romantic TV Shows, TV ...	In a city of coaching centers known to train l...

```
In [ ]:
```

```
In [ ]:
```

The trend of production over the years on Netflix :

```
In [16]: data.rename(columns={'release_year':'Release Year'},inplace=True)
```

```
In [17]: data.head(2)
```

Out[17]:

	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	No Cast Specified	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	no director specified	Ama Qamata, Khosi Ngema, Gail Mabalan... 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...

```
In [18]: df=data.groupby(by=['type','Release Year']).size().reset_index(name='Total Content')
```

```
In [19]: df
```

Out[19]:

	type	Release Year	Total Content
0	Movie	1942	2
1	Movie	1943	3
2	Movie	1944	3
3	Movie	1945	3
4	Movie	1946	1
...
114	TV Show	2017	265
115	TV Show	2018	380
116	TV Show	2019	397
117	TV Show	2020	436
118	TV Show	2021	315

119 rows × 3 columns

```
In [20]: df=df[df['Release Year']>=2010]
```

```
In [21]: df
```

Out[21]:

	type	Release Year	Total Content
61	Movie	2010	154
62	Movie	2011	145
63	Movie	2012	173
64	Movie	2013	225
65	Movie	2014	264
66	Movie	2015	398
67	Movie	2016	658
68	Movie	2017	767
69	Movie	2018	767
70	Movie	2019	633
71	Movie	2020	517
72	Movie	2021	277
107	TV Show	2010	40
108	TV Show	2011	40
109	TV Show	2012	64
110	TV Show	2013	63
111	TV Show	2014	88
112	TV Show	2015	162
113	TV Show	2016	244
114	TV Show	2017	265
115	TV Show	2018	380
116	TV Show	2019	397
117	TV Show	2020	436
118	TV Show	2021	315

In [22]:

```
px.line(df, x="Release Year", y="Total Content", color='type',title='Trend of content produced over the years o
```

The above line graph shows that there has been a decline in the production of the content for movies since 2018 and for TV Shows since 2020

In []: