



NETFLIX BUSINESS CASE STUDY

1. Defining Problem Statement and Analysing basic metrics

```
# Uploading the dataset
!wget https://d2beiqkhq929f0.cloudfront.net/public_assets/assets/000/000/940/original/netflix.csv -O 'netflix.csv'

--2024-02-13 15:44:06-- https://d2beiqkhq929f0.cloudfront.net/public_assets/assets/000/000/940/original/netflix.csv
Resolving d2beiqkhq929f0.cloudfront.net (d2beiqkhq929f0.cloudfront.net)... 18.172.139.46, 18.172.139.210, 18.172.139.61, ...
Connecting to d2beiqkhq929f0.cloudfront.net (d2beiqkhq929f0.cloudfront.net)|18.172.139.46|:443... connected.
HTTP request sent, awaiting response... 200 OK
Length: 3399671 (3.2M) [text/plain]
Saving to: 'netflix.csv'

netflix.csv      100%[=====>]  3.24M  --.-KB/s    in 0.08s

2024-02-13 15:44:07 (43.0 MB/s) - 'netflix.csv' saved [3399671/3399671]

# Importing the relevant modules and reading the file
import pandas as pd
import numpy as np
import matplotlib.pyplot as plt
import seaborn as sns
netflix=pd.read_csv('netflix.csv')
```

```
# First five rows to get an overview of the dataset
netflix.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 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	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...

.....

2. Observations on the shape of data, data types of all the attributes, conversion of categorical attributes to 'category' (If required), missing value detection, statistical summary.

```
# Table shape
netflix.shape
```

```
(8807, 12)
```

Insights: The table have 8807 rows and 12 columns.

```
# Data types of all the columns
netflix.dtypes
```

```
show_id      object
type         object
title        object
director     object
cast         object
country      object
date_added   object
release_year  int64
rating       object
duration     object
listed_in    object
description   object
dtype: object
```

```
# Conversion of categorical attributes like columns - type, country, rating, listed_in to 'category'
```

```
netflix['type']=netflix['type'].astype('category')
netflix['country']=netflix['country'].astype('category')
netflix['rating']=netflix['rating'].astype('category')
netflix['listed_in']=netflix['listed_in'].astype('category')
netflix.dtypes
```

```
show_id      object
type         category
title        object
director     object
cast         object
country      category
date_added   object
release_year  int64
rating       category
duration     object
listed_in    category
description   object
dtype: object
```

```
# Number of NaN values in different columns
```

```
netflix.isna().sum()
```

```
show_id      0
type         0
title        0
director    2634
cast        825
country     831
date_added   10
release_year  0
rating       4
duration     3
listed_in    0
description  0
dtype: int64
```

```
# Removing the missing values
```

```
netflix.dropna(inplace=True)
netflix.shape
```

```
(5332, 12)
```

```
# Statistical summary
```

```
netflix.describe()
```

	release_year
count	5332.000000
mean	2012.742123
std	9.625831
min	1942.000000
25%	2011.000000
50%	2016.000000
75%	2018.000000
max	2021.000000



```
# Statistical summary (Object Column)
```

```
netflix.describe(include = 'object')
```

	show_id	title	director	cast	date_added	duration	description
count	5332	5332	5332	5332	5332	5332	5332
unique	5332	5332	3945	5200	1453	198	5321
top	s8	Sankofa	Raúl Campos, Jan Suter	Samuel West	January 1, 2020	94 min	When pretty new neighbor Seema falls for their...
freq	1	1	18	10	92	135	2

.....

3. Non-Graphical Analysis: Value counts and unique attributes

```
# Value counts: Counting all the genre present in the above dataset
```

```
netflix['listed_in'].value_counts()
```

```
Dramas, International Movies          336
Stand-Up Comedy                      286
Comedies, Dramas, International Movies 257
Dramas, Independent Movies, International Movies 243
Children & Family Movies, Comedies    179
...
Crime TV Shows, TV Horror, TV Mysteries 0
International TV Shows, Romantic TV Shows, TV Action & Adventure 0
British TV Shows, Classic & Cult TV, International TV Shows 0
Anime Series, Teen TV Shows          0
Anime Series, International TV Shows, Romantic TV Shows 0
Name: listed_in, Length: 514, dtype: int64
```

```
# Insights : Top 5 Genre & top 5 country and
#           Top 5 countries in Netflix dataset
```

```
top_5_genre = netflix['listed_in'].value_counts().head()
top_5_genre
```

```
Dramas, International Movies          336
Stand-Up Comedy                      286
Comedies, Dramas, International Movies 257
Dramas, Independent Movies, International Movies 243
Children & Family Movies, Comedies    179
Name: listed_in, dtype: int64
```

```
top_5_countries = netflix['country'].value_counts().head()
top_5_countries
```

```
United States    1846
India            875
United Kingdom   183
Canada           107
Spain            91
Name: country, dtype: int64
```

.....

4. Visual Analysis - Univariate, Bivariate after pre-processing of the data

4.1. For continuous variable(s): Distplot, Countplot, histogram for univariate analysis

```
# Displot
```

```
sns.distplot(netflix['release_year'], bins=20, kde=True)
```

```
<ipython-input-18-8decddce83f>:1: UserWarning:
```

```
`distplot` is a deprecated function and will be removed in seaborn v0.14.0.
```

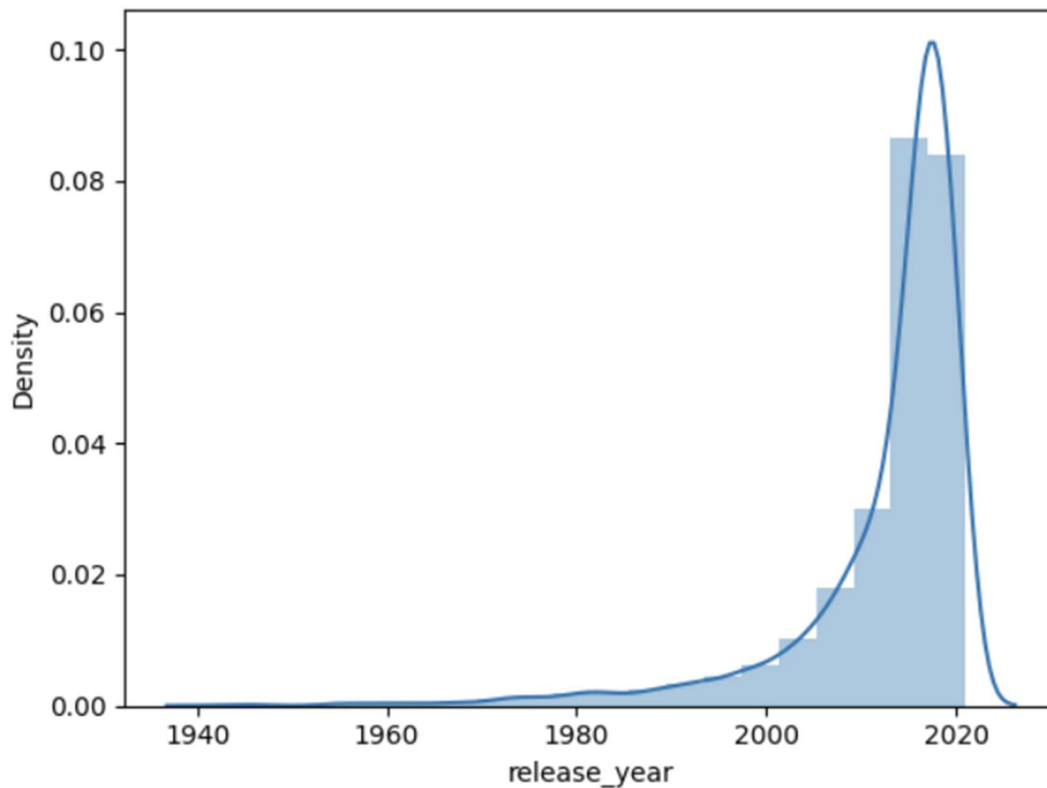
Please adapt your code to use either `displot` (a figure-level function with similar flexibility) or `histplot` (an axes-level function for histograms).

For a guide to updating your code to use the new functions, please see

<https://gist.github.com/mwaskom/de44147ed2974457ad6372750bbe5751>

```
sns.distplot(netflix['release_year'], bins=20, kde=True)
```

```
<Axes: xlabel='release_year', ylabel='Density'>
```



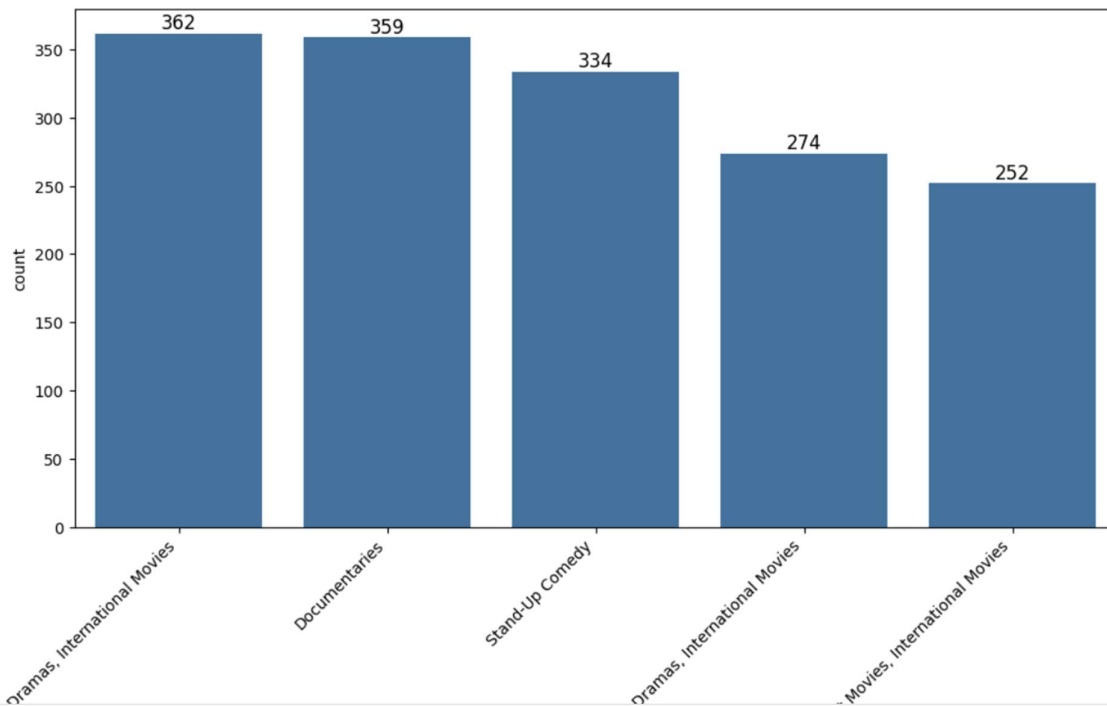
```

# Countplot |

df_IN1 = netflix[netflix['listed_in'] != 'International Movies']

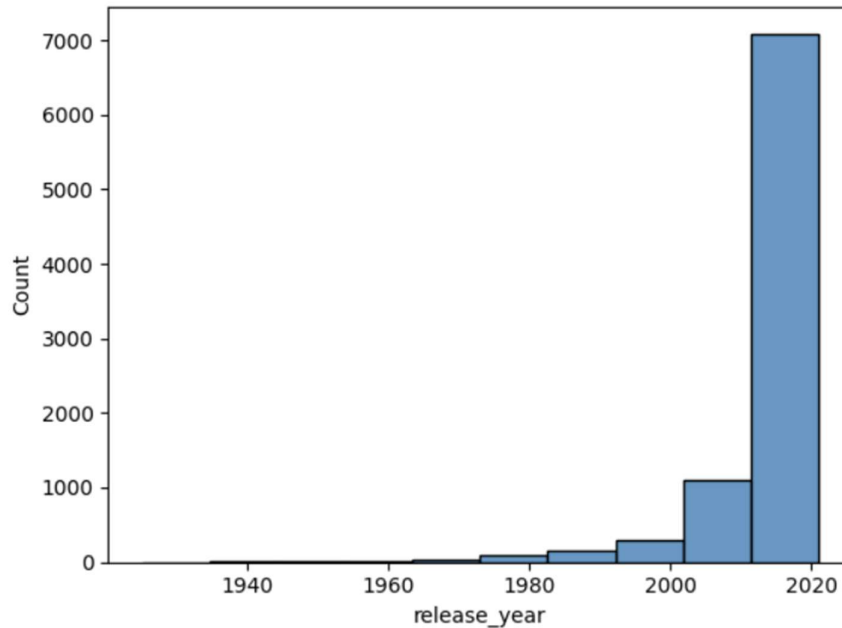
top_5_genre = netflix['listed_in'].value_counts().head()
top_5_genre_index = netflix['listed_in'].value_counts().head().index
x_bar = top_5_genre_index
y_bar = top_5_genre
plt.figure(figsize=(12, 6))
ax = sns.countplot(x='listed_in', data = netflix, order=top_5_genre_index)
# plt.xticks(rotation=90)
plt.xticks(rotation=45, ha='right')
for p in ax.patches:
    height = p.get_height()
    ax.annotate(f'{int(height)}', (p.get_x() + p.get_width() / 2., height), ha='center', va='bottom', fontsize=12)
plt.show()
print(top_5_genre)

```



```
# Histplot
sns.histplot(netflix["release_year"], bins=10)
```

<Axes: xlabel='release_year', ylabel='Count'>



4.2 For categorical variable(s): Boxplot

```
pd.isnull(netflix)
pd.isnull(netflix).sum()
netflix.dropna(inplace=True)
netflix['actors'] = netflix['cast'].apply(lambda x: x.split(', ') if not pd.isna(x) else [])
netflix.drop(columns='cast', inplace=True)
netflix['actors']
```

```
7      [Kofi Ghanaba, Oyafunmike Ogunlano, Alexandra ...
8      [Mel Giedroyc, Sue Perkins, Mary Berry, Paul H...
9      [Melissa McCarthy, Chris O'Dowd, Kevin Kline, ...
12     [Luna Wedler, Jannis Niewöhner, Milan Peschel,...
24     [Prashanth, Aishwarya Rai Bachchan, Sri Lakshm...
...
8801    [Ali Suliman, Saleh Bakri, Yasa, Ali Al-Jabri,...
8802    [Mark Ruffalo, Jake Gyllenhaal, Robert Downey ...
8804    [Jesse Eisenberg, Woody Harrelson, Emma Stone,...
8805    [Tim Allen, Courteney Cox, Chevy Chase, Kate M...
8806    [Vicky Kaushal, Sarah-Jane Dias, Raaghav Chana...
Name: actors, Length: 5332, dtype: object
```

```
netflix['country'] = netflix['country'].str.split(',')
df_country = netflix.explode('country')
df_country['country'] = df_country['country'].apply(lambda x: x.strip())
IND_data = df_country[df_country['country'] == "India"]
IND_data
```

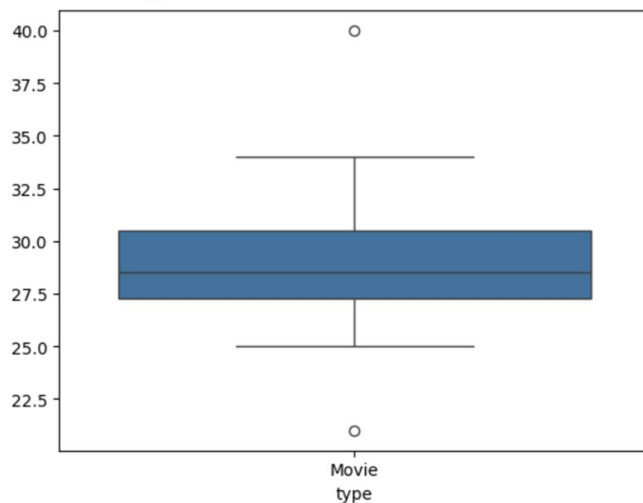
	show_id	type	title	director	country	date_added	release_year	rating	duration	listed_in	description	actors
24	s25	Movie	Jeans	S. Shankar	India	September 21, 2021	1998	TV-14	166 min	Comedies, International Movies, Romantic Movies	When the father of the man she loves insists t...	[Prashanth, Aishwarya Ra Bachchan, Sri Lakshm...
29	s30	Movie	Paranoia	Robert Luketic	India	September 19, 2021	2013	PG-13	106 min	Thrillers	Blackmailed by his company's CEO, a low-level ...	[Liam Hemsworth, Gar Oldman, Amber Heard Har..
105	s106	Movie	Angamaly Diaries	Lijo Jose Pellissery	India	September 5, 2021	2017	TV-14	128 min	Action & Adventure, Comedies, Dramas	After growing up amidst the gang wars of his h...	[Antony Varghese, Reshmi Rajan, Binny Rinky Be..
114	s115	Movie	Anjaam	Rahul Rawail	India	September 2, 2021	1994	TV-14	143 min	Dramas, International Movies, Thrillers	A wealthy industrialist's dangerous obsession ...	[Madhuri Dixit, Shah Rukh Khan, Tinnu Anand, J..
116	s117	Movie	Dhanak	Nagesh Kukunoor	India	September 2, 2021	2015	TV-PG	114 min	Comedies, Dramas, Independent Movies	A movie-loving 10-year-old and her blind littl...	[Krish Chhabria, Heta Gada, Vipin Sharma, Gu...

```
IND_data = IND_data.explode('actors')
top_10_IND_actors = IND_data["actors"].value_counts(normalize=True).head(10).index.to_list()
top_10_IND_data = IND_data[IND_data["actors"].isin(top_10_IND_actors)]
top_10_IND_actors
```

```
['Anupam Kher',
 'Shah Rukh Khan',
 'Naseeruddin Shah',
 'Om Puri',
 'Akshay Kumar',
 'Amitabh Bachchan',
 'Paresh Rawal',
 'Boman Irani',
 'Kareena Kapoor',
 'Ajay Devgn']
```

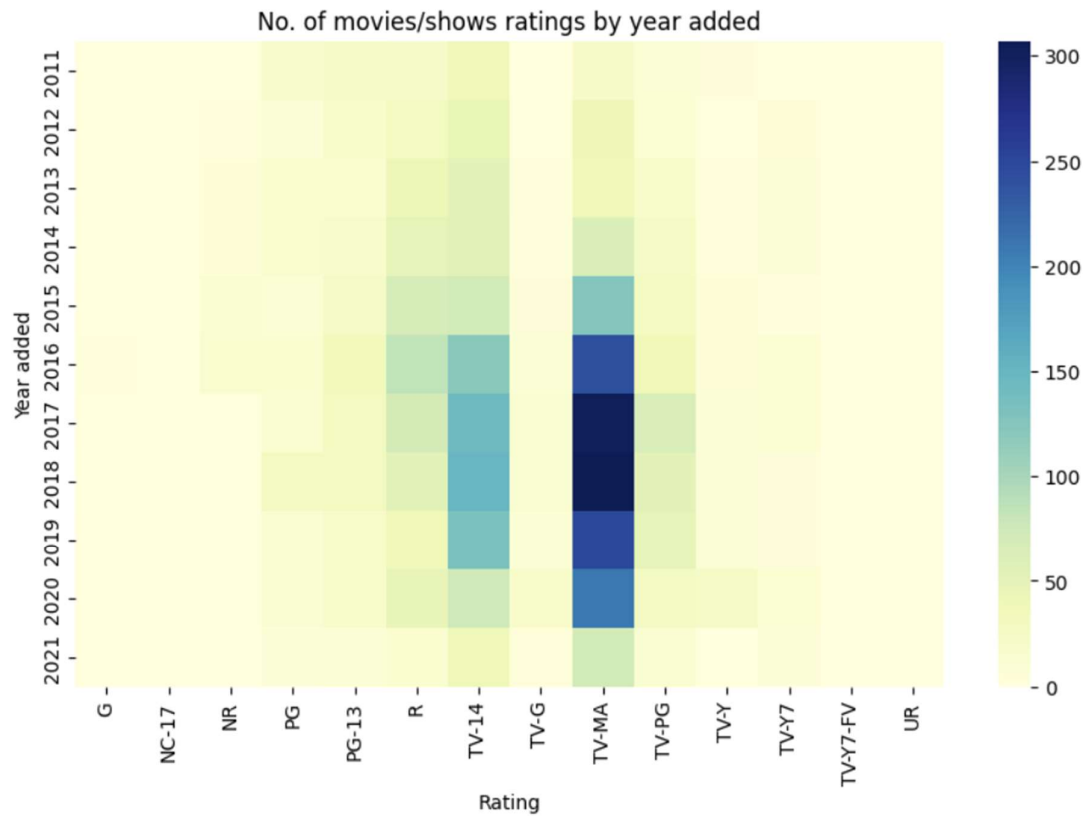
```
IND_data_pvt = top_10_IND_data.pivot_table(index="actors", columns="type", values="rating", aggfunc="count", fill_value=0)
sns.boxplot(IND_data_pvt)
```

<Axes: xlabel='type'>



4.3 For correlation: Heatmaps, Pairplots

```
heatmap_data = netflix[netflix['release_year']>=2011].pivot_table(index='release_year', columns='rating', values='type', aggfunc='count', fill_value=0)
plt.figure(figsize=(10, 6))
sns.heatmap(heatmap_data, cmap='YlGnBu', cbar=True)
plt.title('No. of movies/shows ratings by year added')
plt.xlabel('Rating')
plt.ylabel('Year added')
plt.show()
```



5. Missing Value & Outlier check (Treatment optional)

```
# Removing NaN values from dataframe
```

```
df_dropna = netflix.dropna()
df_dropna
```

	show_id	type	title	director	cast	country	date_added	release_year	rating	duration	listed_in	description
7	s8	Movie	Sankofa	Haile Gerima	Kofi Ghanaba, Oyafunmike Ogunlano, Alexandra D...	United States, Ghana, Burkina Faso, United Kin...	September 24, 2021	1993	TV-MA	125 min	Dramas, Independent Movies, International Movies	On a photo shoot in Ghana, an American model s...
8	s9	TV Show	The Great British Baking Show	Andy Devonshire	Mel Giedroyc, Sue Perkins, Mary Berry, Paul Ho...	United Kingdom	September 24, 2021	2021	TV-14	9 Seasons	British TV Shows, Reality TV	A talented batch of amateur bakers face off in...
9	s10	Movie	The Starling	Theodore Melfi	Melissa McCarthy, Chris O'Dowd, Kevin Kline, T...	United States	September 24, 2021	2021	PG-13	104 min	Comedies, Dramas	A woman adjusting to life after a loss contend...
12	s13	Movie	Je Suis Karl	Christian Schwochow	Luna Wedler, Jannis Niewöhner, Milan Peschel, ...	Germany, Czech Republic	September 23, 2021	2021	TV-MA	127 min	Dramas, International Movies	After most of her family is murdered in a terr...
24	s25	Movie	Jeans	S. Shankar	Prashanth, Aishwarya Rai Bachchan, Sri Lakshmi...	India	September 21, 2021	1998	TV-14	166 min	Comedies, International Movies, Romantic Movies	When the father of the man she loves insists t...
...
8801	s8802	Movie	Zinzana	Majid Al Ansari	Ali Suliman, Saleh Bakri, Yasa, Ali Al-Jabri, ...	United Arab Emirates, Jordan	March 9, 2016	2015	TV-MA	96 min	Dramas, International Movies, Thrillers	Recovering alcoholic Talal wakes up inside a s...

```
# Replacing NaN values with value 0
netflix.fillna(0)
```

	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	0	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	0	Ama Camata, Khosi Ngema, Gali Mababane, Thabani...	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 l...
2	s3	TV Show	Ganglands	Julien Leclercq	Sami Bouajila, Tracy Gotoas, Samuel Jouy, Nabil...	0	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	0	0	0	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	0	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...

.....

6. Insights based on Non-Graphical and Visual Analysis

Insights based on Non-Graphical -

- With help of info we can understand the datatypes of different columns
- Top 5 Genre {Dramas, Comedies, Action & Adventure, Independent Movies, Romantic Movies} this is most preferable genre for year 1940to2020
- Top 5 Country {United States, India, United Kingdom, Japan, South Korea} that makes and release the shows on Netflix
- Top 10 India Actors in Netflix dataset {Anupam Kher, Shah Rukh Khan, Naseeruddin Shah, Om Puri, Akshay Kumar, Amitabh Bachchan, Paresh Rawal, Boman Irani, Kareena Kapoor, Ajay Devgn}

Insights based on Visual Analysis

And till 2020 data from Netflix dataset

- Number of Movies is approximately twice the number of TV shows
- As we are moving forward with reference to years more movies and shows are released in 2020
- India is second in producing most movies & also producing more movies compared to TV shows

.....

7. Business Insights

- Most popular genres among users are Dramas, Comedies, Action & Adventure, Independent Movies, Romantic Movies, this can guide content acquisition and production decisions. This insight helps to understand this type of content attracts more viewers.
- In Indian market people prefer more movies compare to TV shows.
- But mostly international movies are preferred in all the other countries.

.....

8. Recommendations

1. With refer to Heatmap the target Audience age should be above 17+
2. So we should focus on college students and provide them subscription discounts
3. People prefer mature content so according to that most of the content should be released
4. India is the second highest market and still the open market for Netflix
5. People like to watch Dramas and Comedies contents