

netflix-bussiness-case1

May 27, 2024

```
[ ]: import numpy as np
import pandas as pd
```

```
[ ]: df = pd.read_csv('/content/netflix.csv')
```

```
[ ]: df.head(5)
```

```
[ ]: show_id    type    title    director \
0      s1      Movie    Dick Johnson Is Dead    Kirsten Johnson
1      s2  TV Show          Blood & Water          NaN
2      s3  TV Show          Ganglands    Julien Leclercq
3      s4  TV Show    Jailbirds New Orleans          NaN
4      s5  TV Show          Kota Factory          NaN

                                cast    country \
0                                NaN    United States
1    Ama Qamata, Khosi Ngema, Gail Mabalane, Thaban...    South Africa
2    Sami Bouajila, Tracy Gotoas, Samuel Jouy, Nabi...          NaN
3                                NaN          NaN
4    Mayur More, Jitendra Kumar, Ranjan Raj, Alam K...    India

    date_added    release_year    rating    duration \
0    September 25, 2021          2020    PG-13      90 min
1    September 24, 2021          2021    TV-MA    2 Seasons
2    September 24, 2021          2021    TV-MA      1 Season
3    September 24, 2021          2021    TV-MA      1 Season
4    September 24, 2021          2021    TV-MA    2 Seasons

                                listed_in \
0                                Documentaries
1    International TV Shows, TV Dramas, TV Mysteries
2    Crime TV Shows, International TV Shows, TV Act...
3                                Docuseries, Reality TV
4    International TV Shows, Romantic TV Shows, TV ...

                                description
0    As her father nears the end of his life, filmm...
```

- 1 After crossing paths at a party, a Cape Town t...
- 2 To protect his family from a powerful drug lor...
- 3 Feuds, flirtations and toilet talk go down amo...
- 4 In a city of coaching centers known to train I...

```
[ ]: df.shape
```

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

```
[ ]: df.isnull().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
```

```
[ ]: df_director_r = pd.DataFrame(df['director'].apply(lambda x:str(x).split(',')).
      ↳tolist(),index=df['title'])
```

```
[ ]: df_director = df_director_r.stack().reset_index()
      df_director.drop('level_1', axis=1, inplace=True)
      df_director.rename(columns={0:"Director"}, inplace=True)
      df_director.shape
      df_director.head()
```

```
[ ]:
      title      Director
0  Dick Johnson Is Dead  Kirsten Johnson
1      Blood & Water      nan
2      Ganglands      Julien Leclercq
3  Jailbirds New Orleans      nan
4      Kota Factory      nan
```

```
[ ]: df_cast_r = pd.DataFrame(df['cast'].apply(lambda x:str(x).split(',')).
      ↳tolist(),index=df['title'])
      df_cast = df_cast_r.stack().reset_index()
      df_cast.drop('level_1', axis=1, inplace=True)
      df_cast.rename(columns={0:"Cast"}, inplace=True)
      df_cast.shape
```

```
df_cast.sample(5)
```

```
[ ]:
      title      Cast
8911    Slasher  Mercedes Morris
41959  A Heavy Heart  Peter Schneider
36255    Rang De Basanti    Kiron Kher
32081  Lego DC Comics: Batman Be-Leaguered    Nolan North
11989    Bobby Jasoos    Aakash Dahiya
```

```
[ ]: df_country_r = pd.DataFrame(df['country'].apply(lambda x:str(x).split(',')).
    ↳tolist(),index=df['title'])
df_country = df_country_r.stack().reset_index()
df_country.drop('level_1', axis=1, inplace=True)
df_country.rename(columns={0:"Country"}, inplace=True)
df_country.shape
df_country.head()
```

```
[ ]:
      title      Country
0  Dick Johnson Is Dead  United States
1    Blood & Water    South Africa
2    Ganglands          nan
3  Jailbirds New Orleans    nan
4    Kota Factory      India
```

```
[ ]: df_listed_in_r = pd.DataFrame(df['listed_in'].apply(lambda x:str(x).split(',')).
    ↳tolist(),index=df['title'])
df_listed_in = df_listed_in_r.stack().reset_index()
df_listed_in.drop('level_1', axis=1, inplace=True)
df_listed_in.rename(columns={0:"listed_in"}, inplace=True)
df_listed_in.shape
df_listed_in.head()
```

```
[ ]:
      title      listed_in
0  Dick Johnson Is Dead    Documentaries
1    Blood & Water  International TV Shows
2    Blood & Water    TV Dramas
3    Blood & Water    TV Mysteries
4    Ganglands    Crime TV Shows
```

```
[ ]: df_new = df_director.merge(df_cast, how='inner', on='title')
```

```
[ ]: df_new1 = df_new.merge(df_country, how='inner', on='title')
```

```
[ ]: df_new2 = df_new1.merge(df_listed_in, how='inner', on='title')
```

```
[ ]: df_new2
```

```
[ ]:
      title      Director      Cast \
0    Dick Johnson Is Dead  Kirsten Johnson      nan
1      Blood & Water      nan      Ama Qamata
2      Blood & Water      nan      Ama Qamata
3      Blood & Water      nan      Ama Qamata
4      Blood & Water      nan      Khosi Ngema
...
202060      Zubaan      Mozez Singh      Anita Shabdish
202061      Zubaan      Mozez Singh      Anita Shabdish
202062      Zubaan      Mozez Singh      Chittaranjan Tripathy
202063      Zubaan      Mozez Singh      Chittaranjan Tripathy
202064      Zubaan      Mozez Singh      Chittaranjan Tripathy
```

```
      Country      listed_in
0    United States      Documentaries
1    South Africa      International TV Shows
2    South Africa      TV Dramas
3    South Africa      TV Mysteries
4    South Africa      International TV Shows
...
202060      India      International Movies
202061      India      Music & Musicals
202062      India      Dramas
202063      India      International Movies
202064      India      Music & Musicals
```

[202065 rows x 5 columns]

```
[ ]: df_final = df_new2.merge(df[['show_id', 'type', 'title',
    ↳ 'date_added', 'release_year', 'rating', 'duration', 'description']],
    ↳ how='inner', on='title')
```

```
[ ]: df_final
```

```
[ ]:
      title      Director      Cast \
0    Dick Johnson Is Dead  Kirsten Johnson      nan
1      Blood & Water      nan      Ama Qamata
2      Blood & Water      nan      Ama Qamata
3      Blood & Water      nan      Ama Qamata
4      Blood & Water      nan      Khosi Ngema
...
202060      Zubaan      Mozez Singh      Anita Shabdish
202061      Zubaan      Mozez Singh      Anita Shabdish
202062      Zubaan      Mozez Singh      Chittaranjan Tripathy
202063      Zubaan      Mozez Singh      Chittaranjan Tripathy
202064      Zubaan      Mozez Singh      Chittaranjan Tripathy
```

	Country	listed_in	show_id	type \
0	United States	Documentaries	s1	Movie
1	South Africa	International TV Shows	s2	TV Show
2	South Africa	TV Dramas	s2	TV Show
3	South Africa	TV Mysteries	s2	TV Show
4	South Africa	International TV Shows	s2	TV Show
...
202060	India	International Movies	s8807	Movie
202061	India	Music & Musicals	s8807	Movie
202062	India	Dramas	s8807	Movie
202063	India	International Movies	s8807	Movie
202064	India	Music & Musicals	s8807	Movie

	date_added	release_year	rating	duration \
0	September 25, 2021	2020	PG-13	90 min
1	September 24, 2021	2021	TV-MA	2 Seasons
2	September 24, 2021	2021	TV-MA	2 Seasons
3	September 24, 2021	2021	TV-MA	2 Seasons
4	September 24, 2021	2021	TV-MA	2 Seasons
...
202060	March 2, 2019	2015	TV-14	111 min
202061	March 2, 2019	2015	TV-14	111 min
202062	March 2, 2019	2015	TV-14	111 min
202063	March 2, 2019	2015	TV-14	111 min
202064	March 2, 2019	2015	TV-14	111 min

	description
0	As her father nears the end of his life, filmm...
1	After crossing paths at a party, a Cape Town t...
2	After crossing paths at a party, a Cape Town t...
3	After crossing paths at a party, a Cape Town t...
4	After crossing paths at a party, a Cape Town t...
...	...
202060	A scrappy but poor boy worms his way into a ty...
202061	A scrappy but poor boy worms his way into a ty...
202062	A scrappy but poor boy worms his way into a ty...
202063	A scrappy but poor boy worms his way into a ty...
202064	A scrappy but poor boy worms his way into a ty...

[202065 rows x 12 columns]

```
[ ]: df_final['Cast'].replace(['nan'], ['Unknown Actor'], inplace=True)
df_final['Director'].replace(['nan', 'Unknown Direcor'], inplace=True)
df_final.sample(1)
```

```
[ ]:
152221  Everybody Knows  Asghar Farhadi  Ricardo Darín  Spain \
```

	listed_in	show_id	type	date_added	release_year	\
152221	International Movies	s6714	Movie	June 6, 2019	2019	

	rating	duration	description
152221	R	133 min	Journeying back to her small Spanish hometown ...

```
[ ]: #df_final['unknown_directors'] = df_final[df_final['Director'] == 'Unknown_
↳Direcor']
unknown_directors
```

```
[ ]: Empty DataFrame
Columns: [title, Director, Cast, Country, listed_in, show_id, type, date_added,
release_year, rating, duration, description]
Index: []
```

```
[ ]: df_final[df_final['duration'].isnull()].head()
```

```
[ ]: Empty DataFrame
Columns: [title, Director, Cast, Country, listed_in, show_id, type, date_added,
release_year, rating, duration, description]
Index: []
```

```
[ ]: df_final["duration"] = df_final["duration"].fillna(df_final["rating"])
```

```
[ ]: df_final.sample(6)
```

	title	Director	Cast	\
132062	X: Past Is Present	Sudhish Kamath	Pooja Ruparel	
197122	Timeless Season	Sameer Vidwans	Hady	
124347	I Am Michael	Justin Kelly	Charlie Carver	
95954	A Second Chance	Cathy Garcia-Molina	Al Tantay	
7698	Two Fathers	Michael Schmitt	Lin Yo-Wei	
75816	Rock On!!	Abhishek Kapoor	Prachi Desai	

	Country	listed_in	show_id	type	\
132062	India	International Movies	s5838	Movie	
197122	nan	International TV Shows	s8593	TV Show	
124347	United States	Dramas	s5418	Movie	
95954	Philippines	International Movies	s4069	Movie	
7698	Taiwan	Romantic TV Shows	s322	TV Show	
75816	India	Music & Musicals	s3146	Movie	

	date_added	release_year	rating	duration	\
132062	July 1, 2016	2015	TV-MA	105 min	
197122	September 18, 2017	2009	TV-14	1 Season	
124347	June 26, 2017	2015	TV-MA	102 min	

95954	February 27, 2019	2015	TV-MA	130 min
7698	August 3, 2021	2013	TV-14	1 Season
75816	December 15, 2019	2008	TV-14	139 min

	description
132062	When a middle-aged filmmaker meets an alluring...
197122	A decade after college, a wake reunites five f...
124347	Gay activist Michael Glatze begins a painful j...
95954	Following their storybook wedding, Popoy and B...
7698	When two single guys learn a woman they both s...
75816	Four friends hoped that their rock band would ...

```
[ ]: #1. Defining Problem Statement and Analysing basic metrics
df_final
```

```
[ ]:
      title      Director      Cast \
0  Dick Johnson Is Dead  Kirsten Johnson  Unknown Actor
1      Blood & Water  Kirsten Johnson      Ama Qamata
2      Blood & Water  Kirsten Johnson      Ama Qamata
3      Blood & Water  Kirsten Johnson      Ama Qamata
4      Blood & Water  Kirsten Johnson      Khosi Ngema
...
202060      Zubaan      Mozez Singh      Anita Shabdish
202061      Zubaan      Mozez Singh      Anita Shabdish
202062      Zubaan      Mozez Singh  Chittaranjan Tripathy
202063      Zubaan      Mozez Singh  Chittaranjan Tripathy
202064      Zubaan      Mozez Singh  Chittaranjan Tripathy
```

	Country	listed_in	show_id	type	\
0	United States	Documentaries	s1	Movie	
1	South Africa	International TV Shows	s2	TV Show	
2	South Africa	TV Dramas	s2	TV Show	
3	South Africa	TV Mysteries	s2	TV Show	
4	South Africa	International TV Shows	s2	TV Show	
...	
202060	India	International Movies	s8807	Movie	
202061	India	Music & Musicals	s8807	Movie	
202062	India	Dramas	s8807	Movie	
202063	India	International Movies	s8807	Movie	
202064	India	Music & Musicals	s8807	Movie	

	date_added	release_year	rating	duration	\
0	September 25, 2021	2020	PG-13	90 min	
1	September 24, 2021	2021	TV-MA	2 Seasons	
2	September 24, 2021	2021	TV-MA	2 Seasons	
3	September 24, 2021	2021	TV-MA	2 Seasons	
4	September 24, 2021	2021	TV-MA	2 Seasons	

```

...
202060      March 2, 2019      2015  TV-14      111 min
202061      March 2, 2019      2015  TV-14      111 min
202062      March 2, 2019      2015  TV-14      111 min
202063      March 2, 2019      2015  TV-14      111 min
202064      March 2, 2019      2015  TV-14      111 min

```

```

                                description
0      As her father nears the end of his life, filmm...
1      After crossing paths at a party, a Cape Town t...
2      After crossing paths at a party, a Cape Town t...
3      After crossing paths at a party, a Cape Town t...
4      After crossing paths at a party, a Cape Town t...
...
202060  A scrappy but poor boy worms his way into a ty...
202061  A scrappy but poor boy worms his way into a ty...
202062  A scrappy but poor boy worms his way into a ty...
202063  A scrappy but poor boy worms his way into a ty...
202064  A scrappy but poor boy worms his way into a ty...

```

[202065 rows x 12 columns]

```
[ ]: df_final.info() #It will show how many columns and what are data types and null
      ↪values.
```

```

<class 'pandas.core.frame.DataFrame'>
RangeIndex: 202065 entries, 0 to 202064
Data columns (total 12 columns):
 #   Column          Non-Null Count  Dtype
---  -
 0   title           202065 non-null object
 1   Director        202065 non-null object
 2   Cast            202065 non-null object
 3   Country         202065 non-null object
 4   listed_in       202065 non-null object
 5   show_id         202065 non-null object
 6   type            202065 non-null object
 7   date_added      201907 non-null object
 8   release_year    202065 non-null int64
 9   rating          201998 non-null object
10   duration        202065 non-null object
11   description     202065 non-null object
dtypes: int64(1), object(11)
memory usage: 18.5+ MB

```

```
[ ]: df_final.keys() #It will show the all columns in the DataFrame
```



```
[ ]: Index(['title', 'Director', 'Cast', 'Country', 'listed_in', 'show_id', 'type',
          'date_added', 'release_year', 'rating', 'duration', 'description'],
          dtype='object')
```

```
[ ]: df_final.describe() #It will give the statistical values
```

```
[ ]:
      release_year
count  202065.000000
mean    2013.448950
std       9.013616
min     1925.000000
25%     2012.000000
50%     2016.000000
75%     2019.000000
max     2021.000000
```

```
[ ]: df_final.value_counts().sum() #it will give the It's used to count the
      ↪ occurrences of unique values in the each column.
```

```
[ ]: 201840
```

```
[ ]: df_final.nunique() # It will show the unique values in the each column.
```

```
[ ]: title           8807
      Director       5120
      Cast          39297
      Country        198
      listed_in       73
      show_id        8807
      type           2
      date_added     1767
      release_year    74
      rating         17
      duration       220
      description    8775
      dtype: int64
```

```
[ ]: df_final.head() #It will give first five rows in the data Frame.
```

```
[ ]:
      title           Director      Cast      Country \
0  Dick Johnson Is Dead  Kirsten Johnson  Unknown Actor  United States
1      Blood & Water  Kirsten Johnson    Ama Qamata  South Africa
2      Blood & Water  Kirsten Johnson    Ama Qamata  South Africa
3      Blood & Water  Kirsten Johnson    Ama Qamata  South Africa
4      Blood & Water  Kirsten Johnson    Khosi Ngema  South Africa

      listed_in  show_id      type      date_added  release_year \
```

0	Documentaries	s1	Movie	September 25, 2021	2020
1	International TV Shows	s2	TV Show	September 24, 2021	2021
2	TV Dramas	s2	TV Show	September 24, 2021	2021
3	TV Mysteries	s2	TV Show	September 24, 2021	2021
4	International TV Shows	s2	TV Show	September 24, 2021	2021

	rating	duration	description
0	PG-13	90 min	As her father nears the end of his life, filmm...
1	TV-MA	2 Seasons	After crossing paths at a party, a Cape Town t...
2	TV-MA	2 Seasons	After crossing paths at a party, a Cape Town t...
3	TV-MA	2 Seasons	After crossing paths at a party, a Cape Town t...
4	TV-MA	2 Seasons	After crossing paths at a party, a Cape Town t...

```
[ ]: df_final.tail() # it will give the last five rows in the dataframe.
```

```
[ ]:
      title      Director      Cast Country \
202060 Zubaan  Mozez Singh      Anita Shabdish  India
202061 Zubaan  Mozez Singh      Anita Shabdish  India
202062 Zubaan  Mozez Singh  Chittaranjan Tripathy  India
202063 Zubaan  Mozez Singh  Chittaranjan Tripathy  India
202064 Zubaan  Mozez Singh  Chittaranjan Tripathy  India
```

	listed_in	show_id	type	date_added	release_year	\
202060	International Movies	s8807	Movie	March 2, 2019	2015	
202061	Music & Musicals	s8807	Movie	March 2, 2019	2015	
202062	Dramas	s8807	Movie	March 2, 2019	2015	
202063	International Movies	s8807	Movie	March 2, 2019	2015	
202064	Music & Musicals	s8807	Movie	March 2, 2019	2015	

	rating	duration	description
202060	TV-14	111 min	A scrappy but poor boy worms his way into a ty...
202061	TV-14	111 min	A scrappy but poor boy worms his way into a ty...
202062	TV-14	111 min	A scrappy but poor boy worms his way into a ty...
202063	TV-14	111 min	A scrappy but poor boy worms his way into a ty...
202064	TV-14	111 min	A scrappy but poor boy worms his way into a ty...

```
[ ]: df_final.sample(4) #It will give the random 4 rows in the dataframe.
```

```
[ ]:
      title      Director      Cast \
14864 Quarantine Tales  Sidharta Tata  Arawinda Kirana
140917 Band of Robbers      Adam Nee  Creed Bratton
124839 Crashing      Salvador Calvo  Jonathan Bailey
100434 Kita Kita  Sigrid Andrea Bernardo  Alessandra de Rossi

      Country      listed_in show_id      type \
14864      nan  TV Comedies    s588  TV Show
140917 United States  Comedies    s6227  Movie
```

124839	United Kingdom	International TV Shows	s5445	TV Show
100434	Philippines	International Movies	s4298	Movie

	date_added	release_year	rating	duration	\
14864	July 1, 2021	2020	TV-MA	1 Season	
140917	April 15, 2016	2015	NR	95 min	
124839	June 2, 2017	2016	TV-MA	1 Season	
100434	December 15, 2018	2017	TV-14	84 min	

	description
14864	Traversing genres, five separate stories offer...
140917	Recast as modern-day men, Tom Sawyer and Huck ...
124839	In exchange for low rent, a hodgepodge group o...
100434	When a tour guide in Japan goes blind after se...

```
[ ]: df_final.shape #The shape attribute is commonly used in Python libraries such
↳as NumPy and pandas to determine the dimensions of a data structure like an
↳array or DataFrame.
```

```
[ ]: (202065, 12)
```

```
[ ]: df_final.isnull() #It will show the null/missing value in the DataFrame
```

	title	Director	Cast	Country	listed_in	show_id	type	\
0	False	False	False	False	False	False	False	
1	False	False	False	False	False	False	False	
2	False	False	False	False	False	False	False	
3	False	False	False	False	False	False	False	
4	False	False	False	False	False	False	False	
...	
202060	False	False	False	False	False	False	False	
202061	False	False	False	False	False	False	False	
202062	False	False	False	False	False	False	False	
202063	False	False	False	False	False	False	False	
202064	False	False	False	False	False	False	False	

	date_added	release_year	rating	duration	description
0	False	False	False	False	False
1	False	False	False	False	False
2	False	False	False	False	False
3	False	False	False	False	False
4	False	False	False	False	False
...
202060	False	False	False	False	False
202061	False	False	False	False	False
202062	False	False	False	False	False
202063	False	False	False	False	False

202064 False False False False False

[202065 rows x 12 columns]

```
[71]: # 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
df_final.shape #The shape attribute is commonly used in Python libraries such
      ↪ as NumPy and pandas to determine the dimensions of a data structure like an
      ↪ array or DataFrame.
```

[71]: (202065, 12)

```
[73]: df_final.dtypes # "dtype" stands for data type. It represents the type of data
      ↪ stored in an array or DataFrame
```

```
[73]: title           object
      Director        object
      Cast            object
      Country         object
      listed_in       object
      show_id         object
      type            object
      date_added      object
      release_year    int64
      rating          object
      duration        object
      description     object
      dtype: object
```

```
[81]: #Converting categorical columns are category datatype
categorical_columns = ['Director', 'Country', 'listed_in', 'type', 'rating',
      ↪ 'duration']

df_final[categorical_columns] = df_final[categorical_columns].astype('category')
df_final.dtypes
```

```
[81]: title           object
      Director        category
      Cast            object
      Country         category
      listed_in       category
      show_id         object
      type            category
      date_added      object
      release_year    int64
      rating          category
```

```

duration          category
description        object
dtype: object

```

```

[83]: #Missing/Null values finding
df_final.isnull()
df_final.isna()

```

```

[83]:      title  Director   Cast  Country  listed_in  show_id  type  \
0      False      False  False   False      False   False  False
1      False      False  False   False      False   False  False
2      False      False  False   False      False   False  False
3      False      False  False   False      False   False  False
4      False      False  False   False      False   False  False
...
202060  False      False  False   False      False   False  False
202061  False      False  False   False      False   False  False
202062  False      False  False   False      False   False  False
202063  False      False  False   False      False   False  False
202064  False      False  False   False      False   False  False

```

```

      date_added  release_year  rating  duration  description
0      False      False      False      False      False
1      False      False      False      False      False
2      False      False      False      False      False
3      False      False      False      False      False
4      False      False      False      False      False
...
202060      False      False      False      False      False
202061      False      False      False      False      False
202062      False      False      False      False      False
202063      False      False      False      False      False
202064      False      False      False      False      False

```

[202065 rows x 12 columns]

```

[86]: df_final.describe() #The describe() method in Python is primarily associated
    ↪with the pandas library and is used to generate descriptive statistics that
    ↪summarize the central tendency, dispersion, and shape of a dataset's
    ↪distribution, excluding NaN values.

```

```

[86]:      release_year
count  202065.000000
mean    2013.448950
std       9.013616
min     1925.000000
25%     2012.000000

```

```

50%      2016.000000
75%      2019.000000
max       2021.000000

```

[89]: #3. Non-Graphical Analysis: Value counts and unique attributes
`df_final.value_counts()` #The `value_counts()` method in Python is associated with
↳ the pandas library and is used to count the unique values in a Series or
↳ DataFrame column. It returns a Series containing counts of unique values,
↳ sorted in descending order.

```

[89]: title          Director          Cast          Country
listed_in          show_id  type    date_added    release_year
rating duration description
300 Miles to Heaven    Maciej Dejczer    Adrianna Biedrzyńska  France
International Movies    s6014    Movie    October 1, 2019    1989
TV-14    93 min    Hoping to help their dissident parents, two brothers sneak out
of Poland and land as refugees in Denmark, where they are prevented from
returning home.    2
Rust Creek            Jen McGowan            Micah Hauptman            United
States Thrillers            s1632    Movie    November 30, 2020    2018
R    108 min    A wrong turn in the woods becomes a fight for her life when a
career-seeking college student runs into two outlaw brothers looking to cook up
trouble.    2
300 Miles to Heaven    Maciej Dejczer    Adrianna Biedrzyńska  France
Dramas            s6014    Movie    October 1, 2019    1989
TV-14    93 min    Hoping to help their dissident parents, two brothers sneak out
of Poland and land as refugees in Denmark, where they are prevented from
returning home.    2
                                Poland
International Movies    s6014    Movie    October 1, 2019    1989
TV-14    93 min    Hoping to help their dissident parents, two brothers sneak out
of Poland and land as refugees in Denmark, where they are prevented from
returning home.    2
Dramas            s6014    Movie    October 1, 2019    1989
TV-14    93 min    Hoping to help their dissident parents, two brothers sneak out
of Poland and land as refugees in Denmark, where they are prevented from
returning home.    2
..
Haunted: Latin America  Adrián García Bogliano  Unknown Actor            Mexico
Reality TV            s1147    TV Show    March 31, 2021    2021
TV-14    1 Season  Real people's terrifying tales of the chilling, unexplained
and paranormal come to life with dramatic reenactments in this reality series.
1
Spanish-Language TV Shows  s1147    TV Show    March 31, 2021    2021
TV-14    1 Season  Real people's terrifying tales of the chilling, unexplained
and paranormal come to life with dramatic reenactments in this reality series.
1

```

```

International TV Shows      s1147      TV Show  March 31, 2021      2021
TV-14      1 Season  Real people's terrifying tales of the chilling, unexplained
and paranormal come to life with dramatic reenactments in this reality series.
1

```

```

United
States      Reality TV      s1147      TV Show  March 31, 2021      2021
TV-14      1 Season  Real people's terrifying tales of the chilling, unexplained
and paranormal come to life with dramatic reenactments in this reality series.
1

```

```

:      Young Jun Lee      Um Sang-hyun      nan
Children & Family Movies      s7109      Movie      September 1, 2018      2018
TV-Y7      68 min      Miniforce, a special task force of elite rangers, takes on the
Lizard Army to save Earth before it's too late, in this prequel to the TV
series.      1
Name: count, Length: 201833, dtype: int64

```

```

[91]: df_final.nunique() #The unique() method in Python is associated with the pandas
↳and NumPy libraries and is used to find the unique values in an array or a
↳Series.

```

```

[91]: title      8807
Director      5120
Cast      39297
Country      198
listed_in      73
show_id      8807
type      2
date_added      1767
release_year      74
rating      17
duration      220
description      8775
dtype: int64

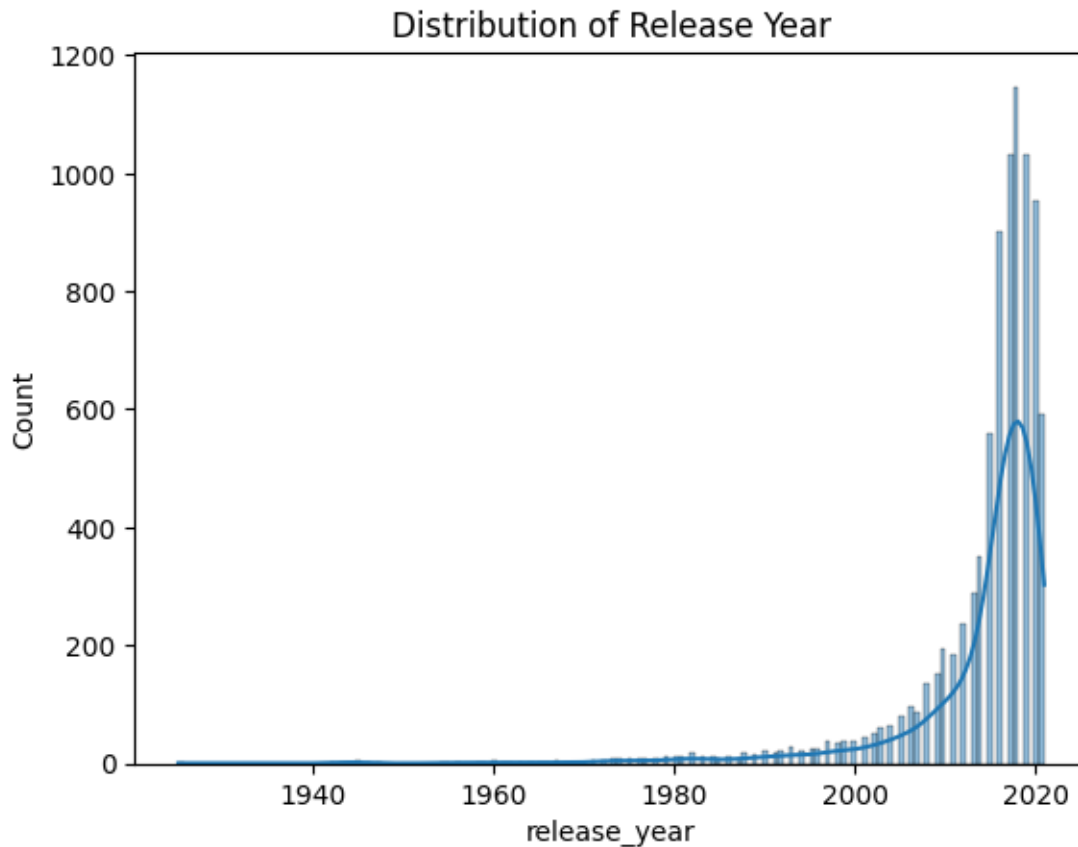
```

```

[94]: # 4.1 For continuous variable(s): Distplot, countplot, histogram for univariate
↳analysis
import seaborn as sns
import matplotlib.pyplot as plt

sns.histplot(df['release_year'], kde=True)
plt.title('Distribution of Release Year')
plt.show()

```



```
[101]: plt.figure(figsize=(10, 60))

# Plot the count plot
sns.countplot(y='Country', data=df_final, palette='viridis')

# Set title and labels
plt.title('Count of Shows by Country', fontsize=16)
plt.xlabel('Count', fontsize=14)
plt.ylabel('Country', fontsize=14)

# Annotate each bar with the count value
for p in plt.gca().patches:
    plt.gca().annotate(format(p.get_width(), '.0f'),
                        (p.get_width(), p.get_y() + p.get_height() / 2),
                        ha='center', va='center',
                        xytext=(20, 0),
                        textcoords='offset points',
                        fontsize=12)

# Show the plot
```

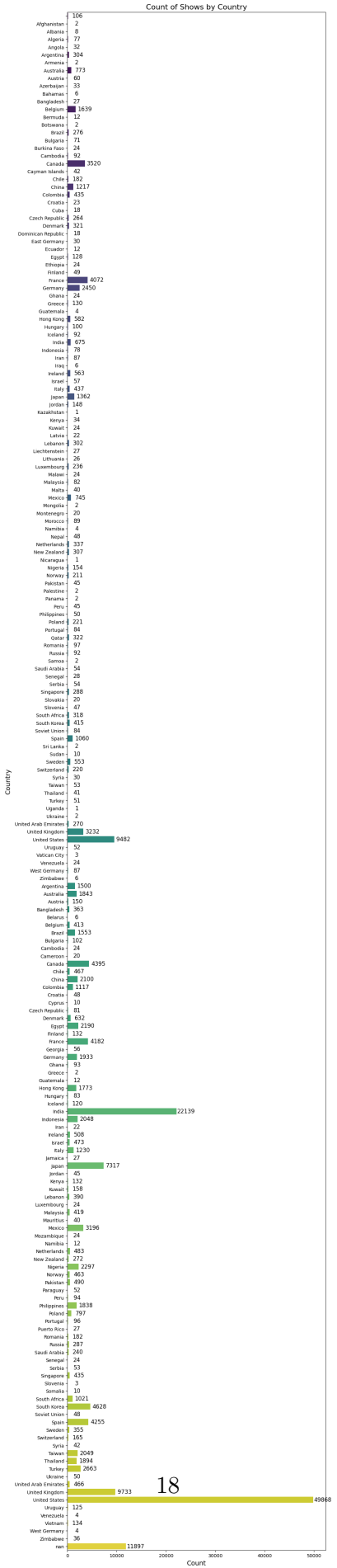


```
plt.show()
```

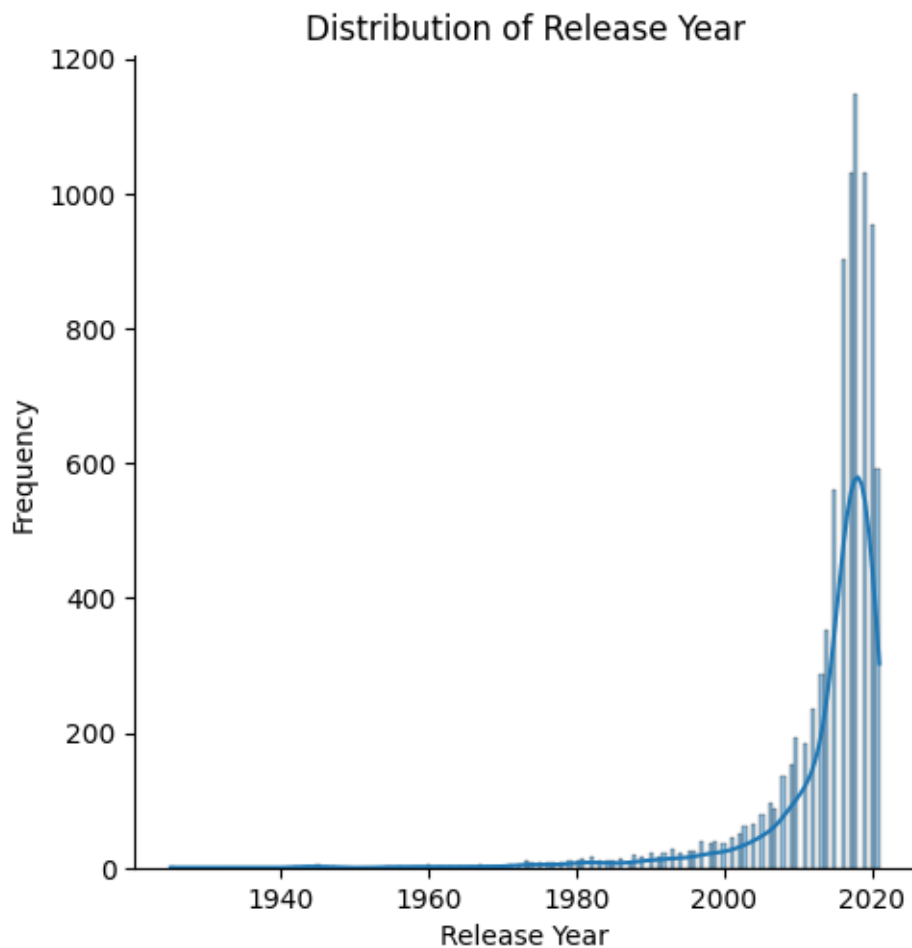
<ipython-input-101-dbe67473e551>:4: FutureWarning:

Passing `palette` without assigning `hue` is deprecated and will be removed in v0.14.0. Assign the `y` variable to `hue` and set `legend=False` for the same effect.

```
sns.countplot(y='Country', data=df_final, palette='viridis')
```



```
[97]: sns.displot(df['release_year'], kde=True)
plt.title('Distribution of Release Year')
plt.xlabel('Release Year')
plt.ylabel('Frequency')
plt.show()
```



```
[114]: # 4.2 For categorical variable(s): Boxplot
plt.figure(figsize=(10, 60))

# Create the box plot
sns.boxplot(x='release_year', y='Country', data=df_final, palette='viridis')

# Set title and labels
plt.title('Distribution of Release Year by Country', fontsize=16)
```

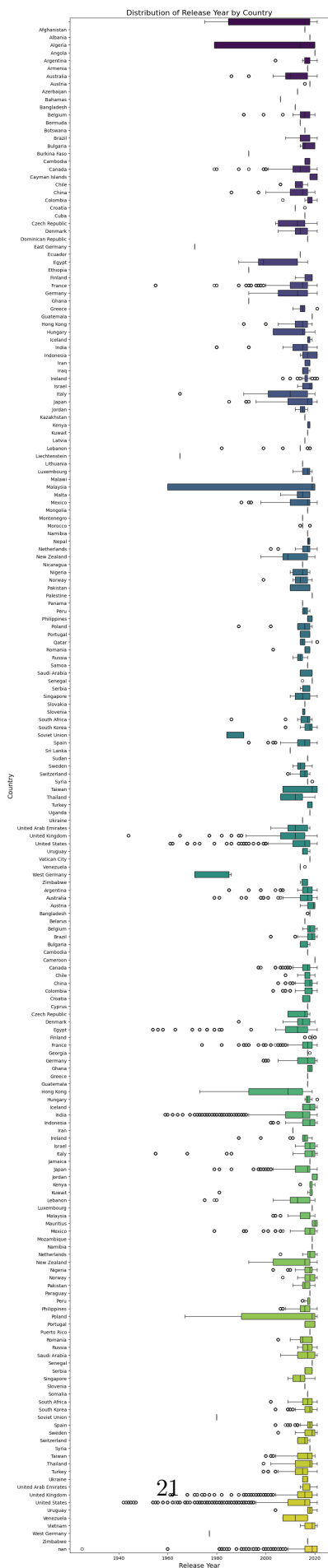
```
plt.xlabel('Release Year', fontsize=14)
plt.ylabel('Country', fontsize=14)

# Show the plot
plt.show()
```

<ipython-input-114-3991978e7826>:5: FutureWarning:

Passing `palette` without assigning `hue` is deprecated and will be removed in v0.14.0. Assign the `y` variable to `hue` and set `legend=False` for the same effect.

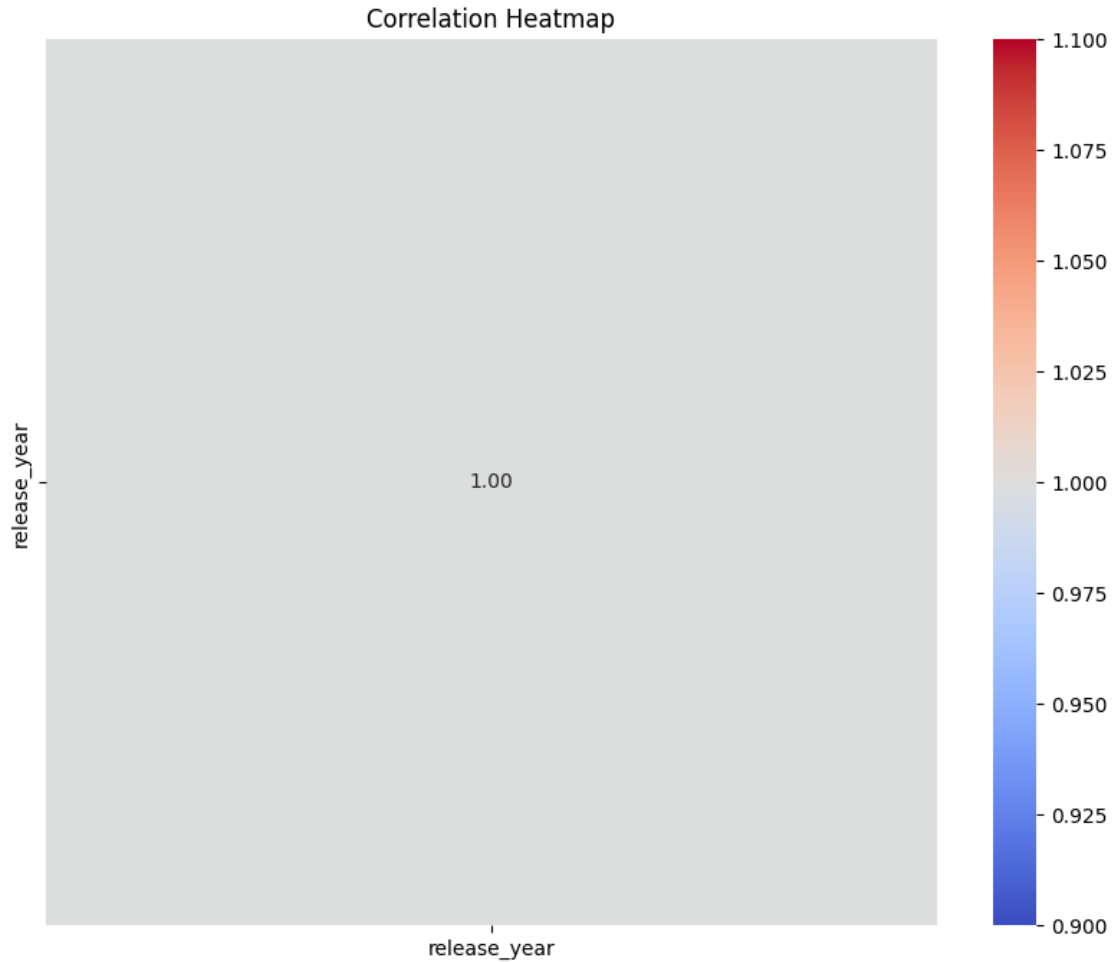
```
sns.boxplot(x='release_year', y='Country', data=df_final, palette='viridis')
```

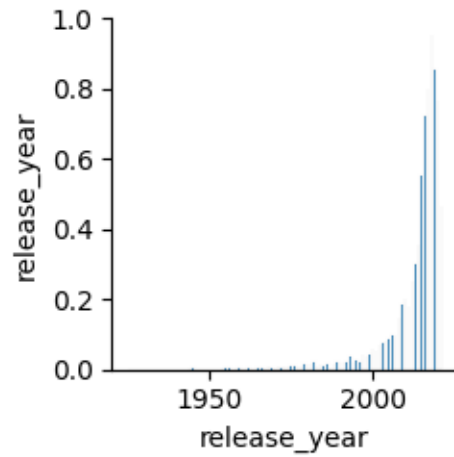


```
[122]: # 4.3 For correlation: Heatmaps, Pairplots
df_numeric = df_final.select_dtypes(include='number')

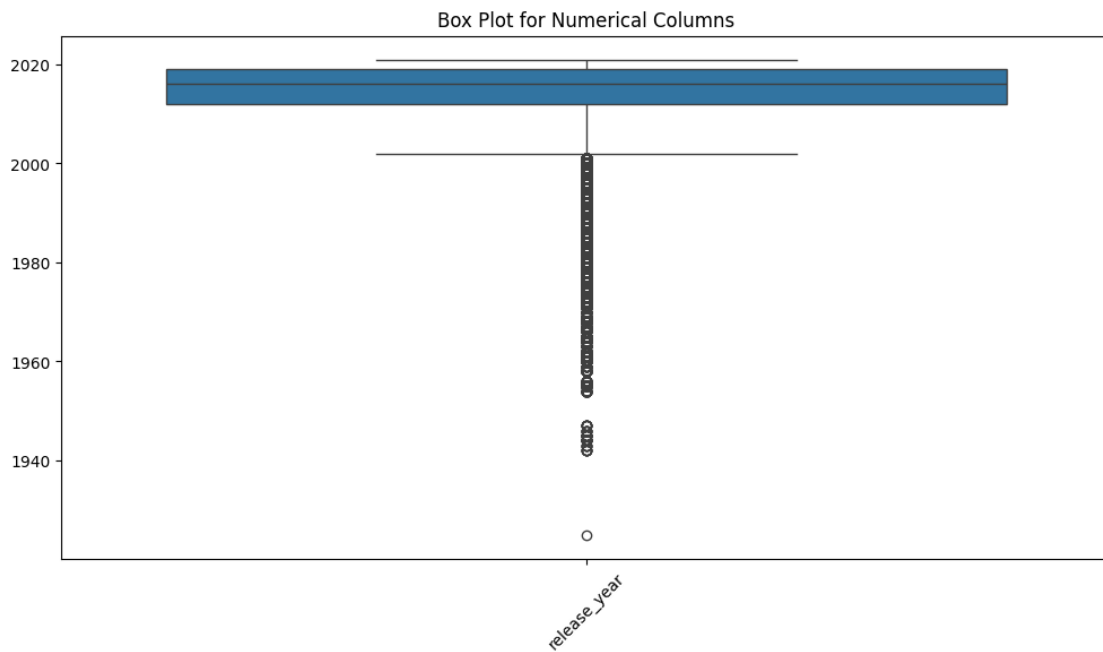
# Correlation heatmap
correlation_matrix = df_numeric.corr()
plt.figure(figsize=(10, 8))
sns.heatmap(correlation_matrix, annot=True, cmap='coolwarm', fmt=".2f")
plt.title('Correlation Heatmap')
plt.show()

# Pairplot
sns.pairplot(df_numeric)
plt.show()
```





```
[123]: # Missing Value & Outlier check (Treatment optional)
plt.figure(figsize=(12, 6))
sns.boxplot(data=df_final.select_dtypes(include='number'))
plt.title('Box Plot for Numerical Columns')
plt.xticks(rotation=45)
plt.show()
```



```
[124]: # Check for missing values
missing_values = df_final.isnull().sum()A
```

```
print(missing_values)
```

```
title          0
Director       0
Cast           0
Country        0
listed_in      0
show_id        0
type           0
date_added     158
release_year   0
rating         67
duration       0
description    0
dtype: int64
```

```
[ ]: ##Insights based on Non-Graphical and Visual Analysis
```

```
*There is a huge gaps in content of Netflix, approximately 70 percent content,
↳is movies and other 30 percent are tv shows.
```

```
*The movies and TV shows uploading on the Netflix started from the year 2008,
↳It had very lesser content till 2014.Contents are there from 1925 but they
↳are very less.
```


*Year 2015 move to a jump in the content getting uploaded on Netflix. It continues to 2019 marks the highest number of movies and TV shows added on the Netflix. Year 2020 and 2021 has seen the drop in content added on Netflix, possibly because of Pandemic. But still , TV shows content have not dropped as drastic as movies. Since 2018, A drop in the movies is seen , but rise in TV shows is observed clearly. Being in continuous uptrend , TV shows surpassed the movies count in mid 2020. It shows the rise in popularity of tv shows in recent years. Netflix has movies from variety of directors. Around 4993 directors have their movies or tv shows on Netflix. Netflix has movies from total 122 countries, United States being the highset contributor with almost 37% of all the content. The release year for shows is concentrated in the range 2005-2021. 50 mins - 150 mins is the range of movie durations, excluding potential outliers. 1-3 seasons is the range for TV shows seasons, excluding potential outliers. various ratings of content is available on netflix, for the various viewers categories like kids, adults , families. Highest number of movies and TV shows are rated TV-MA (for mature audiences). Content in most of the ratings is available in lesser quantity except in US. Ratings like TV-Y7 , TV-Y7 FV , PG ,TV-G , G , TV-Y , TV-PG are very less available in all countries except US. International Movies and TV Shows , Dramas , and Comedies are the top 3 genres on Netflix for both Movies and TV shows. Mostly country specific popular genres are observed in each country. Only United States have a good mix of almost all genres. Eg. Korean TV shows (Korea), British TV Shows (UK), Anime features and Anime series (Japan) and so on. Indian Actors have been acted in maximum movies on netflix. Top 5 actors are in India based on quantity of movies. Shorter duration movies have been popular in last 10 years.

*Since 2018, A drop in the movies is seen , but rise in TV shows is observed clearly. Being in continuous uptrend , TV shows surpassed the movies count in mid 2020. It shows the rise in popularity of tv shows in recent years.

*Netflix has movies from variety of directors. Around 4993 directors have their movies or tv shows on Netflix.

*Netflix has movies from total 122 countries, United States being the highset contributor with almost 37% of all the content.

The release year for shows is concentrated in the range 2005-2021. 50 mins - 150 mins is the range of movie durations, excluding potential outliers.

1-3 seasons is the range for TV shows seasons, excluding potential outliers.

[]: ##Business Insights

*Netflix have majority of content which is released after the year 2000. It is observed that the content older than year 2000 is very scarce on Netflix. Senior Citizen could be the target audience for such content, which is almost missing currently.

*Maximum content is for Adult. TV-MA - Content intended for mature audiences aged 17 and above. TV-14 - Content suitable for viewers aged 14 and above. TV-PG - Parental guidance suggested (similar ratings - PG-13 , PG) R - Restricted Content, that may not be suitable for viewers under age 17.

*These ratings' movies target Matured and Adult audience. Rest 20 % of the content is for kids aged below 13. It shows that Netflix is currently serving mostly Mature audiences or Children with parental guidance.

*Most popular genres on Netflix are International Movies and TV Shows , Dramas , Comedies, Action & Adventure, Children & Family Movies, Thrillers. Maximum content of Netflix which is around 75% , is coming from the top 10 countries. Rest of the world only contributes 25% of the content. More countries can be focussed in future to grow the business.

*Liking towards the shorter duration content is on the rise. (duration 75 to 150 minutes and seasons 1 to 3) This can be considered while production of new content on Netflix.

[]: #Recommendations

Movies :-

Preferd movies duration is between 90-100 minutes.

Netflix should add more movies for United States and India falling in category of Internation movies and comedies.

Netflix should add more movies for United States and India having rating of TV-MA & TV-14.

Top three countries where movies added are United States, India & United Kingdom.

Netflix shoud add TV Show on Friday than any other weekday.

TV Show:-

Preferd movies duration is 1-2 seeasons.

Netflix should focus on countries like Japan, South Korea and France in TV shows , as they prefer TV shows over movies.

Netflix should add TV Show on Friday than other weekday.

As per 2021 data, count of TV shows are more than movies , this means people
↳ wants more web-series as they have for leisure time may be due to work from
↳ home scenario.