

netflix-business-case-study

September 28, 2024

Importing libraries

```
[201]: import numpy as np
import pandas as pd
import matplotlib.pyplot as plt
import seaborn as sns
```

Defining the data set

```
[202]: df = pd.read_csv('netflixdata.csv')
```

Knowing the Data

```
[203]: df.head(8)
```

```
[203]: show_id    type                title \
0      s1      Movie      Dick Johnson Is Dead
1      s2  TV Show      Blood & Water
2      s3  TV Show      Ganglands
3      s4  TV Show      Jailbirds New Orleans
4      s5  TV Show      Kota Factory
5      s6  TV Show      Midnight Mass
6      s7      Movie  My Little Pony: A New Generation
7      s8      Movie      Sankofa
```

```
                                director \
0                                Kirsten Johnson
1                                NaN
2                                Julien Leclercq
3                                NaN
4                                NaN
5                                Mike Flanagan
6  Robert Cullen, José Luis Ucha
7                                Haile Gerima
```

```
                                cast \
0                                NaN
1  Ama Qamata, Khosi Ngema, Gail Mabalane, Thaban...
2  Sami Bouajila, Tracy Gotoas, Samuel Jouy, Nabi...
```

```

3                                     NaN
4 Mayur More, Jitendra Kumar, Ranjan Raj, Alam K...
5 Kate Siegel, Zach Gilford, Hamish Linklater, H...
6 Vanessa Hudgens, Kimiko Glenn, James Marsden, ...
7 Kofi Ghanaba, Oyafunmike Ogunlano, Alexandra D...

```

```

                                     country      date_added \
0 United States September 25, 2021
1 South Africa September 24, 2021
2 NaN September 24, 2021
3 NaN September 24, 2021
4 India September 24, 2021
5 NaN September 24, 2021
6 NaN September 24, 2021
7 United States, Ghana, Burkina Faso, United Kin... September 24, 2021

```

```

release_year rating  duration \
0      2020 PG-13      90 min
1      2021 TV-MA  2 Seasons
2      2021 TV-MA   1 Season
3      2021 TV-MA   1 Season
4      2021 TV-MA  2 Seasons
5      2021 TV-MA   1 Season
6      2021 PG      91 min
7      1993 TV-MA  125 min

```

```

                                     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 ...
5 TV Dramas, TV Horror, TV Mysteries
6 Children & Family Movies
7 Dramas, Independent Movies, International Movies

```

```

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...
5 The arrival of a charismatic young priest brin...
6 Equestria's divided. But a bright-eyed hero be...
7 On a photo shoot in Ghana, an American model s...

```

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

[204]: (8807, 12)

[205]: df.info()

```
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 8807 entries, 0 to 8806
Data columns (total 12 columns):
#   Column          Non-Null Count  Dtype
---  -
0   show_id         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      8797 non-null   object
7   release_year    8807 non-null   int64
8   rating          8803 non-null   object
9   duration        8804 non-null   object
10  listed_in       8807 non-null   object
11  description      8807 non-null   object
dtypes: int64(1), object(11)
memory usage: 825.8+ KB
```

[206]: df.describe(include = 'object')

```
[206]:
```

| | show_id | type | title | director | \ |
|--------|---------|-------|----------------------|---------------|---|
| count | 8807 | 8807 | 8807 | 6173 | |
| unique | 8807 | 2 | 8807 | 4528 | |
| top | s1 | Movie | Dick Johnson Is Dead | Rajiv Chilaka | |
| freq | 1 | 6131 | 1 | 19 | |

| | cast | country | date_added | rating | duration | \ |
|--------|--------------------|---------------|-----------------|--------|----------|---|
| count | 7982 | 7976 | 8797 | 8803 | 8804 | |
| unique | 7692 | 748 | 1767 | 17 | 220 | |
| top | David Attenborough | United States | January 1, 2020 | TV-MA | 1 Season | |
| freq | 19 | 2818 | 109 | 3207 | 1793 | |

| | listed_in | \ |
|--------|------------------------------|---|
| count | 8807 | |
| unique | 514 | |
| top | Dramas, International Movies | |
| freq | 362 | |

| | description |
|--------|-------------|
| count | 8807 |
| unique | 8775 |

```
top      Paranormal activity at a lush, abandoned prope...
freq                                          4
```

While observing the data set we see coulmns having nested values and also the columns with null values. We have to clean the dataset and fill up the null values for more accurate analysis.

Un-nesting the column values

```
[207]: df['director'].isnull().any()
```

```
[207]: True
```

```
[208]: # Splitting & arranging the data
df['cast'] = df['cast'].apply(lambda x: str(x).split(', '))
df['director'] = df['director'].apply(lambda x: str(x).split(', '))
df['country'] = df['country'].apply(lambda x: str(x).split(', '))
df['listed_in'] = df['listed_in'].apply(lambda x: str(x).split(', '))
```

```
[209]: df = df.explode('cast')
df = df.explode('director')
df = df.explode('country')
df = df.explode('listed_in')

df
```

```
[209]:
```

| | show_id | type | title | director \ |
|------|---------|---------|----------------------|-----------------|
| 0 | s1 | Movie | Dick Johnson Is Dead | Kirsten Johnson |
| 1 | s2 | TV Show | Blood & Water | nan |
| 1 | s2 | TV Show | Blood & Water | nan |
| 1 | s2 | TV Show | Blood & Water | nan |
| 1 | s2 | TV Show | Blood & Water | nan |
| ... | ... | ... | ... | ... |
| 8806 | s8807 | Movie | Zubaan | Mozez Singh |
| 8806 | s8807 | Movie | Zubaan | Mozez Singh |
| 8806 | s8807 | Movie | Zubaan | Mozez Singh |
| 8806 | s8807 | Movie | Zubaan | Mozez Singh |
| 8806 | s8807 | Movie | Zubaan | Mozez Singh |

| | cast | country | date_added | release_year \ |
|------|----------------|---------------|--------------------|----------------|
| 0 | nan | United States | September 25, 2021 | 2020 |
| 1 | Ama Qamata | South Africa | September 24, 2021 | 2021 |
| 1 | Ama Qamata | South Africa | September 24, 2021 | 2021 |
| 1 | Ama Qamata | South Africa | September 24, 2021 | 2021 |
| 1 | Khosi Ngema | South Africa | September 24, 2021 | 2021 |
| ... | ... | ... | ... | ... |
| 8806 | Anita Shabdish | India | March 2, 2019 | 2015 |
| 8806 | Anita Shabdish | India | March 2, 2019 | 2015 |

| | | | | |
|------|-----------------------|-------|---------------|------|
| 8806 | Chittaranjan Tripathy | India | March 2, 2019 | 2015 |
| 8806 | Chittaranjan Tripathy | India | March 2, 2019 | 2015 |
| 8806 | Chittaranjan Tripathy | India | March 2, 2019 | 2015 |

| | rating | duration | listed_in \ |
|------|--------|-----------|------------------------|
| 0 | PG-13 | 90 min | Documentaries |
| 1 | TV-MA | 2 Seasons | International TV Shows |
| 1 | TV-MA | 2 Seasons | TV Dramas |
| 1 | TV-MA | 2 Seasons | TV Mysteries |
| 1 | TV-MA | 2 Seasons | International TV Shows |
| ... | ... | ... | ... |
| 8806 | TV-14 | 111 min | International Movies |
| 8806 | TV-14 | 111 min | Music & Musicals |
| 8806 | TV-14 | 111 min | Dramas |
| 8806 | TV-14 | 111 min | International Movies |
| 8806 | TV-14 | 111 min | Music & Musicals |

| | description |
|------|---|
| 0 | As her father nears the end of his life, filmm... |
| 1 | After crossing paths at a party, a Cape Town t... |
| 1 | After crossing paths at a party, a Cape Town t... |
| 1 | After crossing paths at a party, a Cape Town t... |
| 1 | After crossing paths at a party, a Cape Town t... |
| ... | ... |
| 8806 | A scrappy but poor boy worms his way into a ty... |
| 8806 | A scrappy but poor boy worms his way into a ty... |
| 8806 | A scrappy but poor boy worms his way into a ty... |
| 8806 | A scrappy but poor boy worms his way into a ty... |
| 8806 | A scrappy but poor boy worms his way into a ty... |

[201991 rows x 12 columns]

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

```
[210]: (201991, 12)
```

```
[211]: df.info()
```

```
<class 'pandas.core.frame.DataFrame'>
Index: 201991 entries, 0 to 8806
Data columns (total 12 columns):
#   Column          Non-Null Count  Dtype
---  -
0   show_id         201991 non-null object
1   type            201991 non-null object
2   title           201991 non-null object
3   director        201991 non-null object
```

```

4  cast          201991 non-null  object
5  country       201991 non-null  object
6  date_added    201833 non-null  object
7  release_year  201991 non-null  int64
8  rating        201924 non-null  object
9  duration      201988 non-null  object
10 listed_in     201991 non-null  object
11 description   201991 non-null  object
dtypes: int64(1), object(11)
memory usage: 20.0+ MB

```

```

[212]: #Replacing the nan values as string to actual Null values
df.replace('nan',np.nan,inplace = True)
df.head(8)

```

```

[212]:  show_id    type          title          director          cast \
0      s1      Movie  Dick Johnson Is Dead  Kirsten Johnson      NaN
1      s2  TV Show          Blood & Water          NaN      Ama Qamata
1      s2  TV Show          Blood & Water          NaN      Ama Qamata
1      s2  TV Show          Blood & Water          NaN      Ama Qamata
1      s2  TV Show          Blood & Water          NaN      Khosi Ngema
1      s2  TV Show          Blood & Water          NaN      Khosi Ngema
1      s2  TV Show          Blood & Water          NaN      Khosi Ngema
1      s2  TV Show          Blood & Water          NaN      Gail Mabalane

          country      date_added  release_year  rating  duration \
0  United States  September 25, 2021          2020  PG-13      90 min
1  South Africa  September 24, 2021          2021  TV-MA  2 Seasons
1  South Africa  September 24, 2021          2021  TV-MA  2 Seasons
1  South Africa  September 24, 2021          2021  TV-MA  2 Seasons
1  South Africa  September 24, 2021          2021  TV-MA  2 Seasons
1  South Africa  September 24, 2021          2021  TV-MA  2 Seasons
1  South Africa  September 24, 2021          2021  TV-MA  2 Seasons
1  South Africa  September 24, 2021          2021  TV-MA  2 Seasons

          listed_in          description
0      Documentaries  As her father nears the end of his life, filmm...
1  International TV Shows  After crossing paths at a party, a Cape Town t...
1      TV Dramas      After crossing paths at a party, a Cape Town t...
1      TV Mysteries  After crossing paths at a party, a Cape Town t...
1  International TV Shows  After crossing paths at a party, a Cape Town t...
1      TV Dramas      After crossing paths at a party, a Cape Town t...
1      TV Mysteries  After crossing paths at a party, a Cape Town t...
1  International TV Shows  After crossing paths at a party, a Cape Town t...

```

```

[213]: # Duration columns has different type of format as one is for the movies and
      ↳ other is for the dramas,

```

```
# so we will split the column and will consider only the numerical values to do
↳ the further analysis.
```

```
df['duration'] = df['duration'].apply(lambda x: str(x).split()[0])
df.head(5)
```

```
[213]:
```

| | show_id | type | title | director | cast | \ |
|---|---------|---------|----------------------|-----------------|-------------|---|
| 0 | s1 | Movie | Dick Johnson Is Dead | Kirsten Johnson | NaN | |
| 1 | s2 | TV Show | Blood & Water | NaN | Ama Qamata | |
| 1 | s2 | TV Show | Blood & Water | NaN | Ama Qamata | |
| 1 | s2 | TV Show | Blood & Water | NaN | Ama Qamata | |
| 1 | s2 | TV Show | Blood & Water | NaN | Khosi Ngema | |

| | country | date_added | release_year | rating | duration | \ |
|---|---------------|--------------------|--------------|--------|----------|---|
| 0 | United States | September 25, 2021 | 2020 | PG-13 | 90 | |
| 1 | South Africa | September 24, 2021 | 2021 | TV-MA | 2 | |
| 1 | South Africa | September 24, 2021 | 2021 | TV-MA | 2 | |
| 1 | South Africa | September 24, 2021 | 2021 | TV-MA | 2 | |
| 1 | South Africa | September 24, 2021 | 2021 | TV-MA | 2 | |

| | listed_in | description |
|---|------------------------|---|
| 0 | Documentaries | As her father nears the end of his life, filmm... |
| 1 | International TV Shows | After crossing paths at a party, a Cape Town t... |
| 1 | TV Dramas | After crossing paths at a party, a Cape Town t... |
| 1 | TV Mysteries | After crossing paths at a party, a Cape Town t... |
| 1 | International TV Shows | After crossing paths at a party, a Cape Town t... |

```
[214]: # date_added is of type object now we will convert it to type datetime
```

```
df['date_added'] = pd.to_datetime(df['date_added'], format='%B %d, %Y',
↳ errors='coerce')
df.head(5)
```

```
[214]:
```

| | show_id | type | title | director | cast | \ |
|---|---------|---------|----------------------|-----------------|-------------|---|
| 0 | s1 | Movie | Dick Johnson Is Dead | Kirsten Johnson | NaN | |
| 1 | s2 | TV Show | Blood & Water | NaN | Ama Qamata | |
| 1 | s2 | TV Show | Blood & Water | NaN | Ama Qamata | |
| 1 | s2 | TV Show | Blood & Water | NaN | Ama Qamata | |
| 1 | s2 | TV Show | Blood & Water | NaN | Khosi Ngema | |

| | country | date_added | release_year | rating | duration | \ |
|---|---------------|------------|--------------|--------|----------|---|
| 0 | United States | 2021-09-25 | 2020 | PG-13 | 90 | |
| 1 | South Africa | 2021-09-24 | 2021 | TV-MA | 2 | |
| 1 | South Africa | 2021-09-24 | 2021 | TV-MA | 2 | |
| 1 | South Africa | 2021-09-24 | 2021 | TV-MA | 2 | |
| 1 | South Africa | 2021-09-24 | 2021 | TV-MA | 2 | |

| | listed_in | description |
|---|------------------------|---|
| 0 | Documentaries | As her father nears the end of his life, filmm... |
| 1 | International TV Shows | After crossing paths at a party, a Cape Town t... |
| 1 | TV Dramas | After crossing paths at a party, a Cape Town t... |
| 1 | TV Mysteries | After crossing paths at a party, a Cape Town t... |
| 1 | International TV Shows | After crossing paths at a party, a Cape Town t... |

[215]: `df.info()`

```
<class 'pandas.core.frame.DataFrame'>
Index: 201991 entries, 0 to 8806
Data columns (total 12 columns):
#   Column          Non-Null Count  Dtype
---  -
0   show_id         201991 non-null object
1   type            201991 non-null object
2   title           201991 non-null object
3   director        151348 non-null object
4   cast            199845 non-null object
5   country         190094 non-null object
6   date_added      200245 non-null datetime64[ns]
7   release_year    201991 non-null int64
8   rating          201924 non-null object
9   duration        201991 non-null object
10  listed_in       201991 non-null object
11  description     201991 non-null object
dtypes: datetime64[ns](1), int64(1), object(10)
memory usage: 20.0+ MB
```

[216]: *# Now lets convert the data values in Year, Month & Day format to do the detailed analysis using dates.*

```
import datetime as dt
df['year'] = df['date_added'].dt.year
df['month'] = df['date_added'].dt.month
df['week_day_name'] = df['date_added'].dt.day_name()
```

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

| | show_id | type | title | director | cast | \ |
|---|---------|---------|----------------------|-----------------|-------------|---|
| 0 | s1 | Movie | Dick Johnson Is Dead | Kirsten Johnson | NaN | |
| 1 | s2 | TV Show | Blood & Water | NaN | Ama Qamata | |
| 1 | s2 | TV Show | Blood & Water | NaN | Ama Qamata | |
| 1 | s2 | TV Show | Blood & Water | NaN | Ama Qamata | |
| 1 | s2 | TV Show | Blood & Water | NaN | Khosi Ngema | |

| | country | date_added | release_year | rating | duration | \ |
|--|---------|------------|--------------|--------|----------|---|
|--|---------|------------|--------------|--------|----------|---|

| | | | | | |
|---|---------------|------------|------|-------|----|
| 0 | United States | 2021-09-25 | 2020 | PG-13 | 90 |
| 1 | South Africa | 2021-09-24 | 2021 | TV-MA | 2 |
| 1 | South Africa | 2021-09-24 | 2021 | TV-MA | 2 |
| 1 | South Africa | 2021-09-24 | 2021 | TV-MA | 2 |
| 1 | South Africa | 2021-09-24 | 2021 | TV-MA | 2 |

| | listed_in | description \ |
|---|------------------------|---|
| 0 | Documentaries | As her father nears the end of his life, filmm... |
| 1 | International TV Shows | After crossing paths at a party, a Cape Town t... |
| 1 | TV Dramas | After crossing paths at a party, a Cape Town t... |
| 1 | TV Mysteries | After crossing paths at a party, a Cape Town t... |
| 1 | International TV Shows | After crossing paths at a party, a Cape Town t... |

| | year | month | week_day_name |
|---|--------|-------|---------------|
| 0 | 2021.0 | 9.0 | Saturday |
| 1 | 2021.0 | 9.0 | Friday |
| 1 | 2021.0 | 9.0 | Friday |
| 1 | 2021.0 | 9.0 | Friday |
| 1 | 2021.0 | 9.0 | Friday |

```
[218]: # Since we already have the columns names Year, Month and week_day_name, we
      ↪ will drop the column date_added to avoid any confusion in further analysis
df.drop('date_added',axis = 1,inplace = True)
df.head(5)
```

```
[218]: show_id    type    title    director    cast \
0      s1    Movie  Dick Johnson Is Dead  Kirsten Johnson    NaN
1      s2  TV Show    Blood & Water    NaN    Ama Qamata
1      s2  TV Show    Blood & Water    NaN    Ama Qamata
1      s2  TV Show    Blood & Water    NaN    Ama Qamata
1      s2  TV Show    Blood & Water    NaN    Khosi Ngema
```

| | country | release_year | rating | duration | listed_in \ |
|---|---------------|--------------|--------|----------|------------------------|
| 0 | United States | 2020 | PG-13 | 90 | Documentaries |
| 1 | South Africa | 2021 | TV-MA | 2 | International TV Shows |
| 1 | South Africa | 2021 | TV-MA | 2 | TV Dramas |
| 1 | South Africa | 2021 | TV-MA | 2 | TV Mysteries |
| 1 | South Africa | 2021 | TV-MA | 2 | International TV Shows |

| | description | year | month \ |
|---|---|--------|---------|
| 0 | As her father nears the end of his life, filmm... | 2021.0 | 9.0 |
| 1 | After crossing paths at a party, a Cape Town t... | 2021.0 | 9.0 |
| 1 | After crossing paths at a party, a Cape Town t... | 2021.0 | 9.0 |
| 1 | After crossing paths at a party, a Cape Town t... | 2021.0 | 9.0 |
| 1 | After crossing paths at a party, a Cape Town t... | 2021.0 | 9.0 |

week_day_name

```
0      Saturday
1       Friday
1       Friday
1       Friday
1       Friday
```

Handling of missing values in the Dataset.

```
[219]: df['cast'].mode()[0]
```

```
[219]: 'Liam Neeson'
```

```
[220]: df['director'].mode()[0]
```

```
[220]: 'Martin Scorsese'
```

```
[221]: df['listed_in'].mode()[0]
```

```
[221]: 'Dramas'
```

```
[222]: df['country'].mode()[0]
```

```
[222]: 'United States'
```

```
[223]: df['month'].mode()[0]
```

```
[223]: 7.0
```

```
[224]: df['year'].mode()[0]
```

```
[224]: 2019.0
```

```
[225]: df['week_day_name'].mode()[0]
```

```
[225]: 'Friday'
```

```
[226]: # Filling rating column Null values by its type
movie_rating = df.loc[df['type'] == 'Movie', 'rating'].mode()[0]
tv_rating = df.loc[df['type'] == 'TV Show', 'rating'].mode()[0]
```

```
[227]: # Filling up all the columns with null values with mode values

df['cast'].fillna(df['cast'].mode()[0],inplace = True)
df['director'].fillna(df['director'].mode()[0],inplace = True)
df['country'].fillna(df['country'].mode()[0],inplace = True)
df['listed_in'].fillna(df['listed_in'].mode()[0],inplace = True)
df['month'].fillna(df['month'].mode()[0],inplace = True)
```

```

df['year'].fillna(df['year'].mode()[0],inplace = True)
df['week_day_name'].fillna(df['week_day_name'].mode()[0],inplace = True)
df['rating'] = df.apply(lambda x: movie_rating if x['type'] == 'Movie' and pd.
↳isna(x['rating'])
                                else tv_rating if x['type'] == 'TV Show' and pd.
↳isna(x['rating'])
                                else x['rating'], axis=1)

```

[228]: df

```

[228]:
   show_id  type  title  director \
0        s1  Movie  Dick Johnson Is Dead  Kirsten Johnson
1        s2  TV Show  Blood & Water  Martin Scorsese
1        s2  TV Show  Blood & Water  Martin Scorsese
1        s2  TV Show  Blood & Water  Martin Scorsese
1        s2  TV Show  Blood & Water  Martin Scorsese
...      ...  ...      ...      ...
8806    s8807  Movie  Zubaan  Mozez Singh
8806    s8807  Movie  Zubaan  Mozez Singh
8806    s8807  Movie  Zubaan  Mozez Singh
8806    s8807  Movie  Zubaan  Mozez Singh
8806    s8807  Movie  Zubaan  Mozez Singh

   cast  country  release_year  rating  duration \
0  Liam Neeson  United States  2020  PG-13  90
1  Ama Qamata  South Africa  2021  TV-MA  2
1  Ama Qamata  South Africa  2021  TV-MA  2
1  Ama Qamata  South Africa  2021  TV-MA  2
1  Khosi Ngema  South Africa  2021  TV-MA  2
...      ...      ...      ...      ...
8806  Anita Shabdish  India  2015  TV-14  111
8806  Anita Shabdish  India  2015  TV-14  111
8806  Chittaranjan Tripathy  India  2015  TV-14  111
8806  Chittaranjan Tripathy  India  2015  TV-14  111
8806  Chittaranjan Tripathy  India  2015  TV-14  111

   listed_in \
0  Documentaries
1  International TV Shows
1  TV Dramas
1  TV Mysteries
1  International TV Shows
...      ...
8806  International Movies
8806  Music & Musicals
8806  Dramas
8806  International Movies

```

8806 Music & Musicals

| | | description | year | month | \ |
|------|--|---|--------|-------|---|
| 0 | | As her father nears the end of his life, filmm... | 2021.0 | 9.0 | |
| 1 | | After crossing paths at a party, a Cape Town t... | 2021.0 | 9.0 | |
| 1 | | After crossing paths at a party, a Cape Town t... | 2021.0 | 9.0 | |
| 1 | | After crossing paths at a party, a Cape Town t... | 2021.0 | 9.0 | |
| 1 | | After crossing paths at a party, a Cape Town t... | 2021.0 | 9.0 | |
| ... | | ... | ... | ... | |
| 8806 | | A scrappy but poor boy worms his way into a ty... | 2019.0 | 3.0 | |
| 8806 | | A scrappy but poor boy worms his way into a ty... | 2019.0 | 3.0 | |
| 8806 | | A scrappy but poor boy worms his way into a ty... | 2019.0 | 3.0 | |
| 8806 | | A scrappy but poor boy worms his way into a ty... | 2019.0 | 3.0 | |
| 8806 | | A scrappy but poor boy worms his way into a ty... | 2019.0 | 3.0 | |

| | week_day_name |
|------|---------------|
| 0 | Saturday |
| 1 | Friday |
| 1 | Friday |
| 1 | Friday |
| 1 | Friday |
| ... | ... |
| 8806 | Saturday |
| 8806 | Saturday |
| 8806 | Saturday |
| 8806 | Saturday |
| 8806 | Saturday |

[201991 rows x 14 columns]

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

```
[229]: show_id      0
      type        0
      title       0
      director    0
      cast        0
      country     0
      release_year 0
      rating      0
      duration    0
      listed_in   0
      description 0
      year        0
      month       0
      week_day_name 0
      dtype: int64
```

```
[230]: # Observations - After cleaning the data by un-nsetting and filling up the null
      ↪ values for some specific columns,
      # we now have 201991 rows with 14 columns in the dataset

df.shape
```

```
[230]: (201991, 14)
```

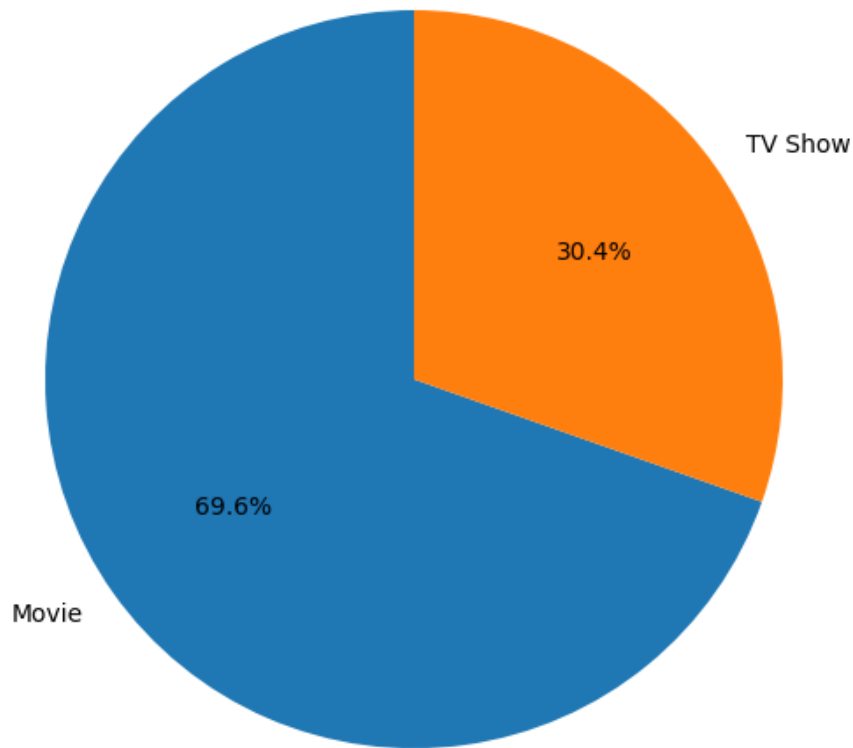
Analysis 1 - How many movies and TV Shows are listed on Netflix.

```
[231]: type_counts = df.groupby('type')['title'].nunique()
      type_counts
```

```
[231]: type
      Movie      6131
      TV Show    2676
      Name: title, dtype: int64
```

```
[232]: plt.figure(figsize=(8, 6))
      plt.pie(type_counts, labels=type_counts.index, autopct='%1.1f%%', startangle=90)
      plt.title('Distribution of Movies and TV Shows on Netflix')
      plt.axis('equal')
      plt.show()
```

Distribution of Movies and TV Shows on Netflix



Insights - After representing the data with a countplot, we observe that the netflix platform has more number movies(70%) as compared to TV Shows(30%).

Recommendations - Todays generation are more into the TV Shows like Web series, Dramas. so depending on the user preferences TV Shows must be increased in future.

Analysis 2 - What are the number of movies/TV shows released per year?

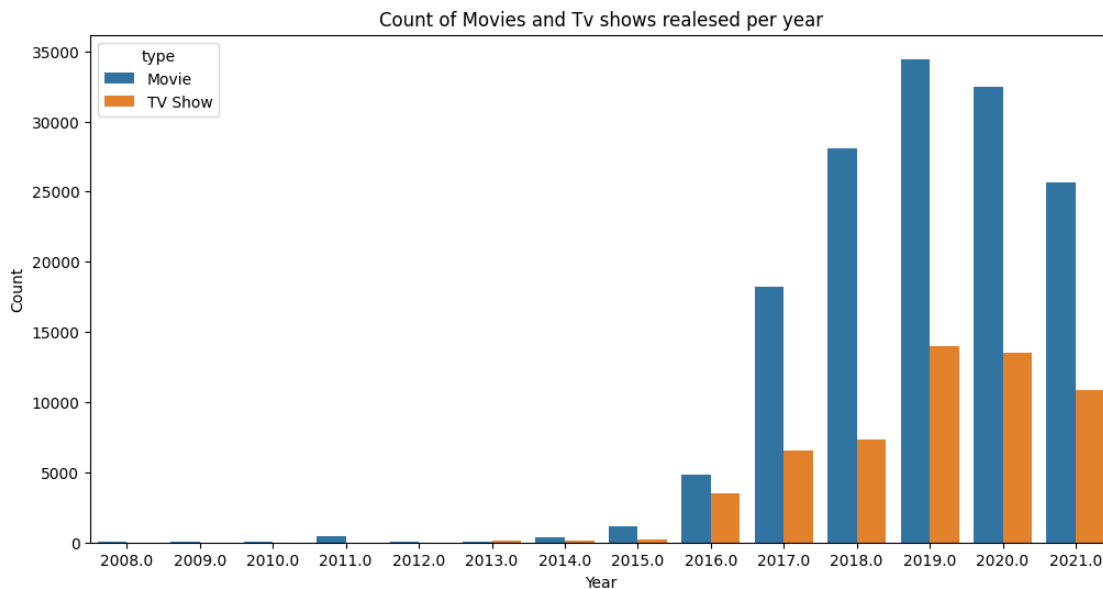
```
[233]: df.groupby(['year', 'type'])['type'].value_counts()
```

```
[233]: year    type
2008.0  Movie      18
        TV Show     1
2009.0  Movie     30
2010.0  Movie     20
2011.0  Movie    438
2012.0  Movie     36
2013.0  Movie     75
        TV Show    110
2014.0  Movie    343
        TV Show    104
```

| | | |
|--------|---------|-------|
| 2015.0 | Movie | 1125 |
| | TV Show | 232 |
| 2016.0 | Movie | 4858 |
| | TV Show | 3529 |
| 2017.0 | Movie | 18252 |
| | TV Show | 6531 |
| 2018.0 | Movie | 28049 |
| | TV Show | 7316 |
| 2019.0 | Movie | 34446 |
| | TV Show | 13955 |
| 2020.0 | Movie | 32462 |
| | TV Show | 13545 |
| 2021.0 | Movie | 25691 |
| | TV Show | 10825 |

Name: count, dtype: int64

```
[234]: plt.figure(figsize=(12,6))
sns.countplot(x='year', data = df, hue = 'type')
plt.title('Count of Movies and Tv shows realeased per year')
plt.xlabel('Year')
plt.ylabel('Count')
plt.show()
```



Insights - After adding a countplot for representing the number of movies over years, we observe that in last 4 to 5 Years more number of movies were added on netflix as compared to TV shows. Which states that Netflix has more focus on movies.

Analysis 3 - Best Time to launch a TVShow or Movies on Netflix

```
[235]: TV_Shows = df[df['type'] == 'TV Show']
Movies = df[df['type'] == 'Movie']
TV_Shows.head(5)
```

```
[235]: show_id      type      title      director      cast      country \
1      s2  TV Show  Blood & Water  Martin Scorsese  Ama Qamata  South Africa
1      s2  TV Show  Blood & Water  Martin Scorsese  Ama Qamata  South Africa
1      s2  TV Show  Blood & Water  Martin Scorsese  Ama Qamata  South Africa
1      s2  TV Show  Blood & Water  Martin Scorsese  Khosi Ngema  South Africa
1      s2  TV Show  Blood & Water  Martin Scorsese  Khosi Ngema  South Africa

      release_year rating duration      listed_in \
1      2021    TV-MA      2  International TV Shows
1      2021    TV-MA      2      TV Dramas
1      2021    TV-MA      2      TV Mysteries
1      2021    TV-MA      2  International TV Shows
1      2021    TV-MA      2      TV Dramas

      description      year      month \
1  After crossing paths at a party, a Cape Town t...  2021.0      9.0
1  After crossing paths at a party, a Cape Town t...  2021.0      9.0
1  After crossing paths at a party, a Cape Town t...  2021.0      9.0
1  After crossing paths at a party, a Cape Town t...  2021.0      9.0
1  After crossing paths at a party, a Cape Town t...  2021.0      9.0

      week_day_name
1      Friday
1      Friday
1      Friday
1      Friday
1      Friday
```

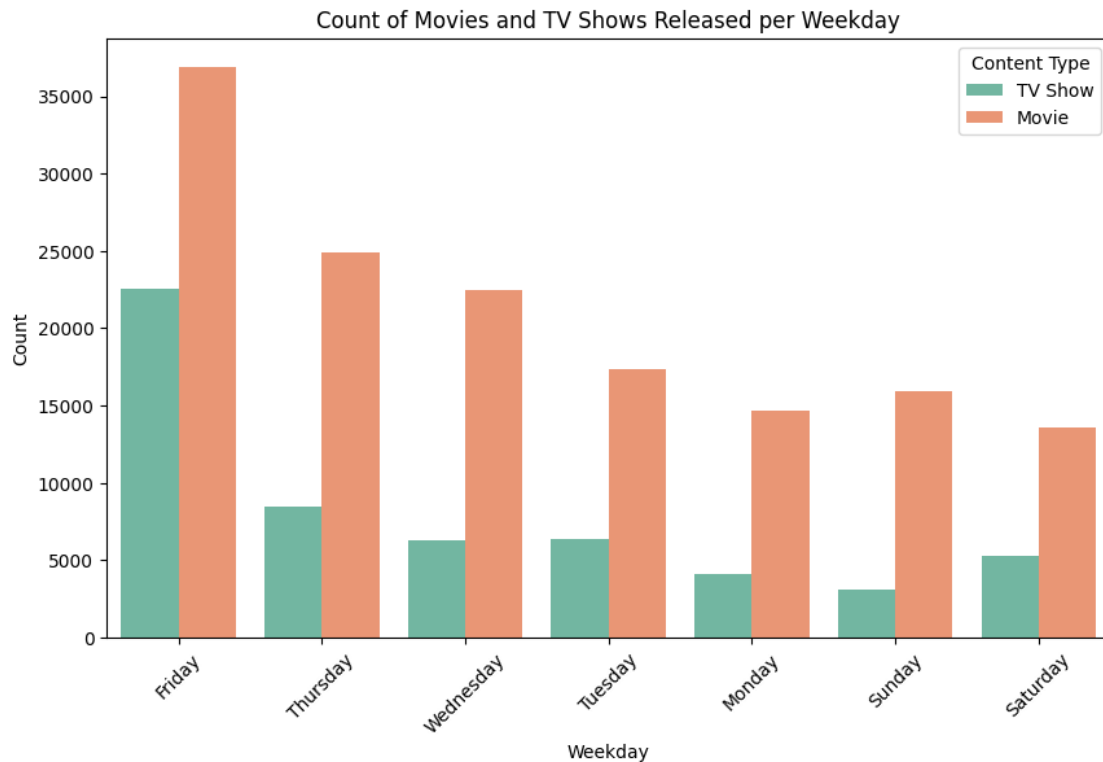
```
[236]: TV_Shows.groupby('week_day_name')['title'].nunique()
Movies.groupby('week_day_name')['title'].nunique()
```

```
[236]: week_day_name
Friday      1566
Monday       628
Saturday     557
Sunday       569
Thursday    1053
Tuesday      852
Wednesday   906
Name: title, dtype: int64
```

```
[237]: combined_df = pd.concat([TV_Shows[['week_day_name', 'type']],
↳ Movies[['week_day_name', 'type']]])
```



```
plt.figure(figsize=(10, 6))
sns.countplot(x='week_day_name', data=combined_df, hue='type', palette='Set2')
plt.title('Count of Movies and TV Shows Released per Weekday')
plt.xlabel('Weekday')
plt.ylabel('Count')
plt.xticks(rotation=45)
plt.legend(title='Content Type')
plt.show()
```



Insights - By observing the data we see that netflix has earlier launched most of the TV Shows & Movies on Fridays. And Yes that is the best time to do so as Its a Weekend, best time to relax and watch the content for viewers. This should be continued in the future.

[238]: *# Now lets check for best month to launch a new TV Show or Movie*

```
TV_Shows.groupby('month')['title'].nunique()
Movies.groupby('month')['title'].nunique()
```

```
[238]: month
1.0    546
2.0    382
3.0    529
4.0    550
```

```

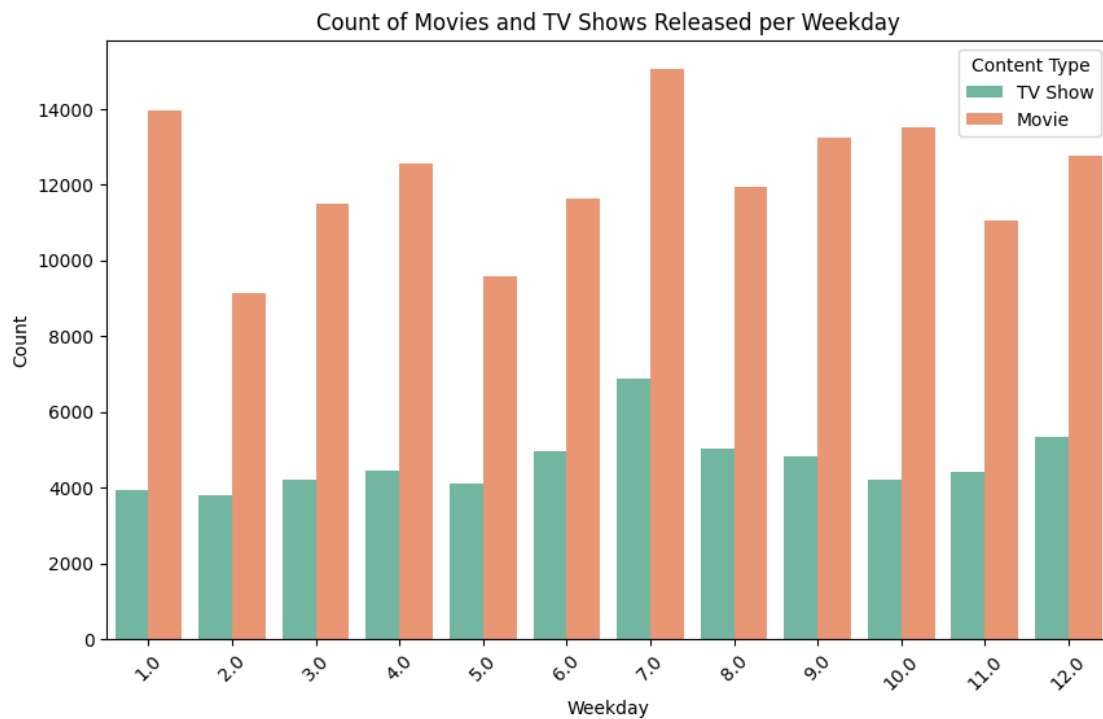
5.0    439
6.0    492
7.0    565
8.0    519
9.0    519
10.0   545
11.0   498
12.0   547
Name: title, dtype: int64

```

```

[239]: combined_df1 = pd.concat([TV_Shows[['month', 'type']], Movies[['month', 'type']])
plt.figure(figsize=(10, 6))
sns.countplot(x='month', data=combined_df1, hue='type', palette='Set2')
plt.title('Count of Movies and TV Shows Released per Weekday')
plt.xlabel('Weekday')
plt.ylabel('Count')
plt.xticks(rotation=45)
plt.legend(title='Content Type')
plt.show()

```



Insights - By observing the data, if you go month wise- we see that netflix has earlier launched highest number of the TV Shows & Movies in the month of July.

Recommendations - The best time to launch maximum amount of TV shows or movies can also be dependant on long holidays in specific countries, for eg. In India, the best time will be October & November(Diwali & Desehra). In USA, UK, NZ it can be December(Christmas, New Year).

Analysis 4 - No of movies released per year changed over the last 20-30 Years.

```
[240]: # lets check for the data for last 30 Years
```

```
recent_years_data = df.loc[(df['release_year'] > 1991) & (df['type'] == 'Movie')]
recent_years_data.head(5)
```

```
[240]:  show_id  type                title      director \
0      s1  Movie      Dick Johnson Is Dead  Kirsten Johnson
6      s7  Movie  My Little Pony: A New Generation    Robert Cullen
6      s7  Movie  My Little Pony: A New Generation    José Luis Ucha
6      s7  Movie  My Little Pony: A New Generation    Robert Cullen
6      s7  Movie  My Little Pony: A New Generation    José Luis Ucha
```

```
      cast      country  release_year  rating  duration \
0    Liam Neeson  United States      2020   PG-13      90
6  Vanessa Hudgens  United States      2021    PG      91
6  Vanessa Hudgens  United States      2021    PG      91
6    Kimiko Glenn  United States      2021    PG      91
6    Kimiko Glenn  United States      2021    PG      91
```

```
      listed_in \
0      Documentaries
6  Children & Family Movies
6  Children & Family Movies
6  Children & Family Movies
6  Children & Family Movies
```

```
      description  year  month \
0  As her father nears the end of his life, filmm...  2021.0    9.0
6  Equestria's divided. But a bright-eyed hero be...  2021.0    9.0
6  Equestria's divided. But a bright-eyed hero be...  2021.0    9.0
6  Equestria's divided. But a bright-eyed hero be...  2021.0    9.0
6  Equestria's divided. But a bright-eyed hero be...  2021.0    9.0
```

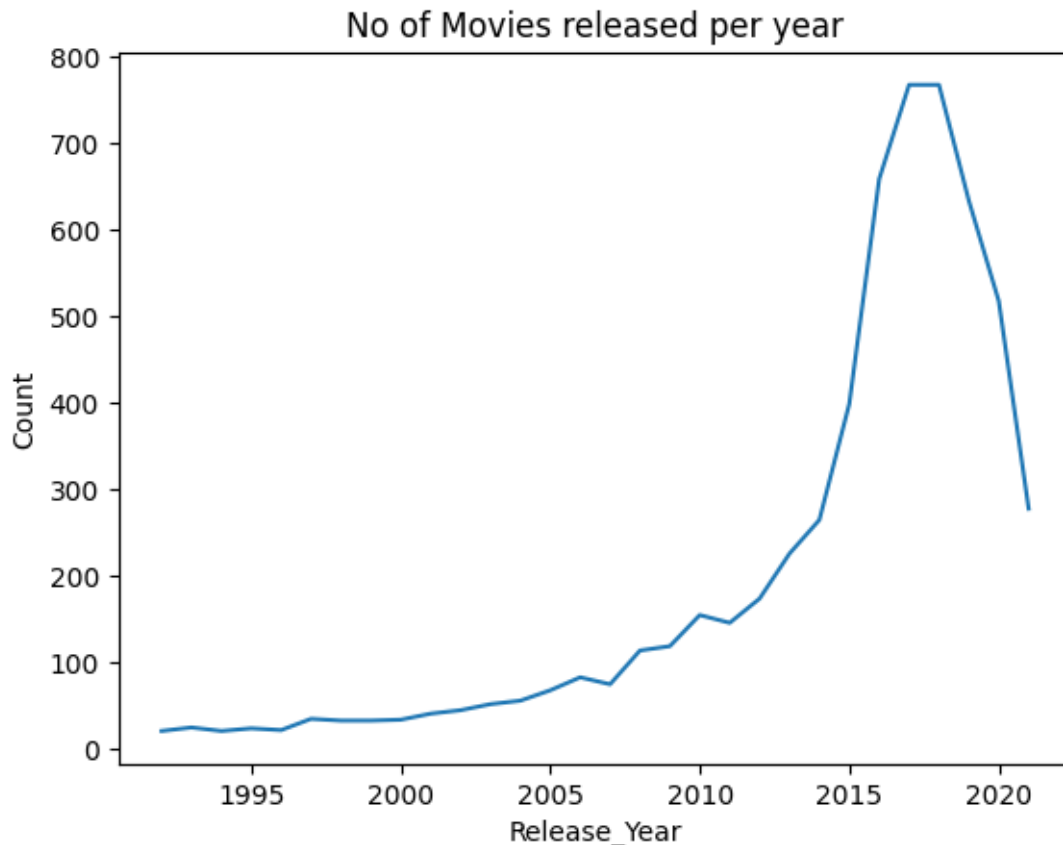
```
      week_day_name
0      Saturday
6      Friday
6      Friday
6      Friday
6      Friday
```

```
[241]: df2 = recent_years_data.groupby('release_year')['title'].nunique().reset_index()
df2
```

```
[241]:
```

| | release_year | title |
|----|--------------|-------|
| 0 | 1992 | 20 |
| 1 | 1993 | 24 |
| 2 | 1994 | 20 |
| 3 | 1995 | 23 |
| 4 | 1996 | 21 |
| 5 | 1997 | 34 |
| 6 | 1998 | 32 |
| 7 | 1999 | 32 |
| 8 | 2000 | 33 |
| 9 | 2001 | 40 |
| 10 | 2002 | 44 |
| 11 | 2003 | 51 |
| 12 | 2004 | 55 |
| 13 | 2005 | 67 |
| 14 | 2006 | 82 |
| 15 | 2007 | 74 |
| 16 | 2008 | 113 |
| 17 | 2009 | 118 |
| 18 | 2010 | 154 |
| 19 | 2011 | 145 |
| 20 | 2012 | 173 |
| 21 | 2013 | 225 |
| 22 | 2014 | 264 |
| 23 | 2015 | 398 |
| 24 | 2016 | 658 |
| 25 | 2017 | 767 |
| 26 | 2018 | 767 |
| 27 | 2019 | 633 |
| 28 | 2020 | 517 |
| 29 | 2021 | 277 |

```
[242]: sns.lineplot(x = df2['release_year'], y = df2['title'])  
plt.title('No of Movies released per year')  
plt.xlabel('Release_Year')  
plt.ylabel('Count')  
plt.show()
```



Insights - After observing the data, we see that the number of movies released kept increasing from 1992 till 2018, but suddenly after 2018 the number keeps decreasing every year in recent years that is from 2018 to 2021. This can be due various number of genres in entertainment industry like Dramas, Webseries, Documentories etc.

Analysis 5 - Who is the most famous Director on Netflix.

```
[243]: df.groupby('director')['title'].nunique().sort_values(ascending = False).
        head(10)
```

```
[243]: director
Martin Scorsese      2646
Rajiv Chilaka         22
Jan Suter             21
Raúl Campos          19
Suhas Kadav           16
Marcus Raboy          16
Jay Karas             15
Cathy Garcia-Molina   13
Youssef Chahine       12
Jay Chapman           12
```

```
Name: title, dtype: int64
```

Insights - The most famous director on Netflix is Martin Scorsese as we used Mode to fill out the null values, but if we consider the null values as missing data then the most famous director will be Rajiv Chilaka.

Analysis 6 - Who is the most famous Actor on Netflix.

```
[244]: df.groupby('cast')['title'].nunique().sort_values(ascending = False).head(10)
```

```
[244]: cast
Liam Neeson          839
Anupam Kher           43
Shah Rukh Khan       35
Julie Tejwani        33
Naseeruddin Shah     32
Takahiro Sakurai     32
Rupa Bhimani         31
Akshay Kumar         30
Om Puri              30
Yuki Kaji            29
Name: title, dtype: int64
```

Insights - The most famous director on Netflix is Liam Neeson, as we used Mode to fill out the null values, but if we consider the null values as missing data then the most famous actor will be Anupam Kher.

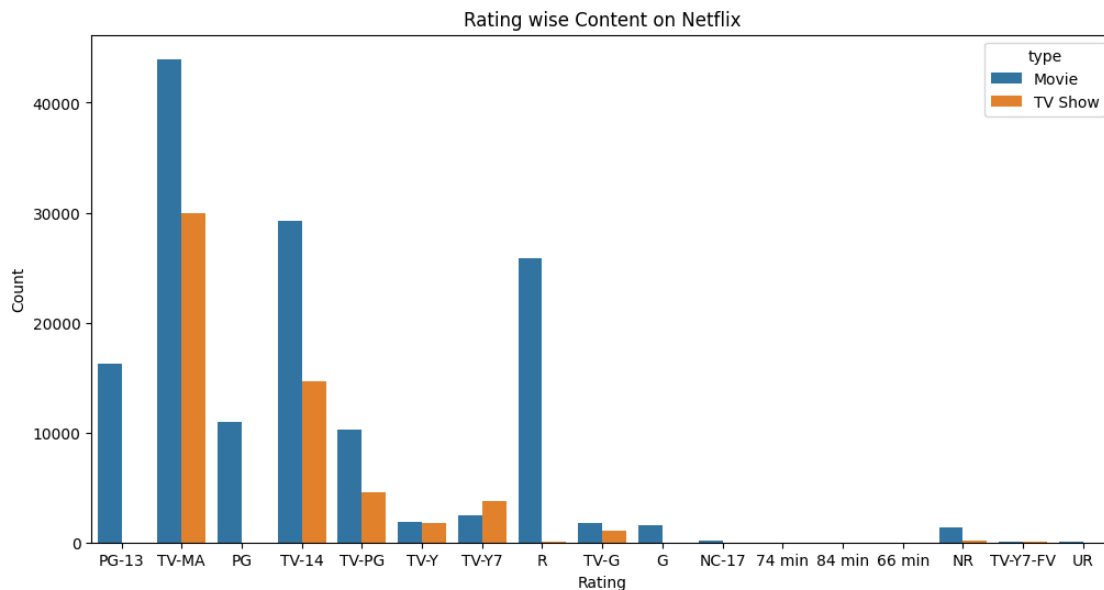
Analysis 7 - Rating wise Content on Netflix

```
[245]: df.groupby('rating')['title'].count().sort_values(ascending = False).head(10)
```

```
[245]: rating
TV-MA    73934
TV-14    43931
R         25860
PG-13     16246
TV-PG     14926
PG         10919
TV-Y7      6304
TV-Y       3665
TV-G       2779
NR         1573
Name: title, dtype: int64
```

```
[246]: plt.figure(figsize=(12,6))
sns.countplot(x = 'rating', data = df, hue = 'type')
plt.title('Rating wise Content on Netflix')
plt.xlabel('Rating')
plt.ylabel('Count')
```

```
plt.show()
```



Insights - After analysing the data on Netflix Top 3 Movie Content ratings are TV-MA, TV-14 & R. and for TV shows the top 3 are TV-MA, TV-14 & TV-PG.

Analysis 8 - What is the best Actor-Director pair is that is Popular on Netflix.

```
[247]: df3 = df.groupby(['cast', 'director'])['title'].nunique().sort_values(ascending=
      ↪ = False).reset_index()
df3.rename(columns = {'title': 'count_title'}, inplace = True)
df3.head(10)
```

```
[247]:
```

| | cast | director | count_title |
|---|------------------|-----------------|-------------|
| 0 | Liam Neeson | Martin Scorsese | 353 |
| 1 | Takahiro Sakurai | Martin Scorsese | 24 |
| 2 | Julie Tejwani | Rajiv Chilaka | 19 |
| 3 | Rajesh Kava | Rajiv Chilaka | 19 |
| 4 | Jigna Bhardwaj | Rajiv Chilaka | 18 |
| 5 | Rupa Bhimani | Rajiv Chilaka | 18 |
| 6 | Yuki Kaji | Martin Scorsese | 18 |
| 7 | Fortune Feimster | Martin Scorsese | 16 |
| 8 | Vatsal Dubey | Rajiv Chilaka | 16 |
| 9 | Yuichi Nakamura | Martin Scorsese | 16 |

Insights - The Most famous Actor-Director Combos on Netflix is Actor Liam Neeson with director Martin Scorsese.

Analysis 9 - Distribution of TV Shows and Movies accross the Countries on Netflix.

[248]: df

```
[248]:
```

| | show_id | type | title | director | \ |
|------|---------|---------|----------------------|-----------------|---|
| 0 | s1 | Movie | Dick Johnson Is Dead | Kirsten Johnson | |
| 1 | s2 | TV Show | Blood & Water | Martin Scorsese | |
| 1 | s2 | TV Show | Blood & Water | Martin Scorsese | |
| 1 | s2 | TV Show | Blood & Water | Martin Scorsese | |
| 1 | s2 | TV Show | Blood & Water | Martin Scorsese | |
| ... | ... | ... | ... | ... | |
| 8806 | s8807 | Movie | Zubaan | Mozez Singh | |
| 8806 | s8807 | Movie | Zubaan | Mozez Singh | |
| 8806 | s8807 | Movie | Zubaan | Mozez Singh | |
| 8806 | s8807 | Movie | Zubaan | Mozez Singh | |
| 8806 | s8807 | Movie | Zubaan | Mozez Singh | |

| | cast | country | release_year | rating | duration | \ |
|------|-----------------------|---------------|--------------|--------|----------|---|
| 0 | Liam Neeson | United States | 2020 | PG-13 | 90 | |
| 1 | Ama Qamata | South Africa | 2021 | TV-MA | 2 | |
| 1 | Ama Qamata | South Africa | 2021 | TV-MA | 2 | |
| 1 | Ama Qamata | South Africa | 2021 | TV-MA | 2 | |
| 1 | Khosi Ngema | South Africa | 2021 | TV-MA | 2 | |
| ... | ... | ... | ... | ... | ... | |
| 8806 | Anita Shabdish | India | 2015 | TV-14 | 111 | |
| 8806 | Anita Shabdish | India | 2015 | TV-14 | 111 | |
| 8806 | Chittaranjan Tripathy | India | 2015 | TV-14 | 111 | |
| 8806 | Chittaranjan Tripathy | India | 2015 | TV-14 | 111 | |
| 8806 | Chittaranjan Tripathy | India | 2015 | TV-14 | 111 | |

| | listed_in | \ |
|------|------------------------|---|
| 0 | Documentaries | |
| 1 | International TV Shows | |
| 1 | TV Dramas | |
| 1 | TV Mysteries | |
| 1 | International TV Shows | |
| ... | ... | |
| 8806 | International Movies | |
| 8806 | Music & Musicals | |
| 8806 | Dramas | |
| 8806 | International Movies | |
| 8806 | Music & Musicals | |

| | description | year | month | \ |
|---|---|--------|-------|---|
| 0 | As her father nears the end of his life, filmm... | 2021.0 | 9.0 | |
| 1 | After crossing paths at a party, a Cape Town t... | 2021.0 | 9.0 | |
| 1 | After crossing paths at a party, a Cape Town t... | 2021.0 | 9.0 | |
| 1 | After crossing paths at a party, a Cape Town t... | 2021.0 | 9.0 | |
| 1 | After crossing paths at a party, a Cape Town t... | 2021.0 | 9.0 | |


```

...
8806 A scrappy but poor boy worms his way into a ty... 2019.0 3.0
8806 A scrappy but poor boy worms his way into a ty... 2019.0 3.0
8806 A scrappy but poor boy worms his way into a ty... 2019.0 3.0
8806 A scrappy but poor boy worms his way into a ty... 2019.0 3.0
8806 A scrappy but poor boy worms his way into a ty... 2019.0 3.0

```

```

week_day_name
0 Saturday
1 Friday
1 Friday
1 Friday
1 Friday
...
8806 Saturday
8806 Saturday
8806 Saturday
8806 Saturday
8806 Saturday

```

[201991 rows x 14 columns]

```

[249]: df4 = df.groupby(['country', 'type'])['title'].nunique().sort_values(ascending_
      ↪= False).reset_index()
df4.rename(columns = {'title':'count_title'},inplace = True)
df4.head(10)

```

```

[249]:
country    type  count_title
0  United States  Movie      3191
1  United States  TV Show     1329
2      India      Movie      962
3  United Kingdom  Movie      532
4      Canada      Movie      319
5      France      Movie      303
6  United Kingdom  TV Show      272
7      Japan      TV Show      199
8      Germany      Movie      182
9      Spain      Movie      171

```

Insights - Netflix has maximum number of content for both Movies and TV Shows in United States.

Analysis 10 - Popular listed in based on the country on Netflix

```

[250]: df5 = df.groupby(['country', 'listed_in'])['title'].nunique().
      ↪sort_values(ascending = False).reset_index()
df5.rename(columns = {'title':'count_title'},inplace = True)

```

```
max_counts = df5.loc[df5.groupby('country')['count_title'].idxmax()].
↳sort_values(by='count_title', ascending=False)
max_counts.head(10)
```

```
[250]:
```

| | country | listed_in | count_title |
|----|----------------|------------------------|-------------|
| 0 | United States | Dramas | 945 |
| 1 | India | International Movies | 864 |
| 18 | United Kingdom | British TV Shows | 225 |
| 20 | France | International Movies | 207 |
| 29 | South Korea | International TV Shows | 152 |
| 30 | Japan | International TV Shows | 151 |
| 32 | Spain | International Movies | 140 |
| 41 | Egypt | International Movies | 99 |
| 43 | Canada | Comedies | 94 |
| 44 | Germany | International Movies | 94 |

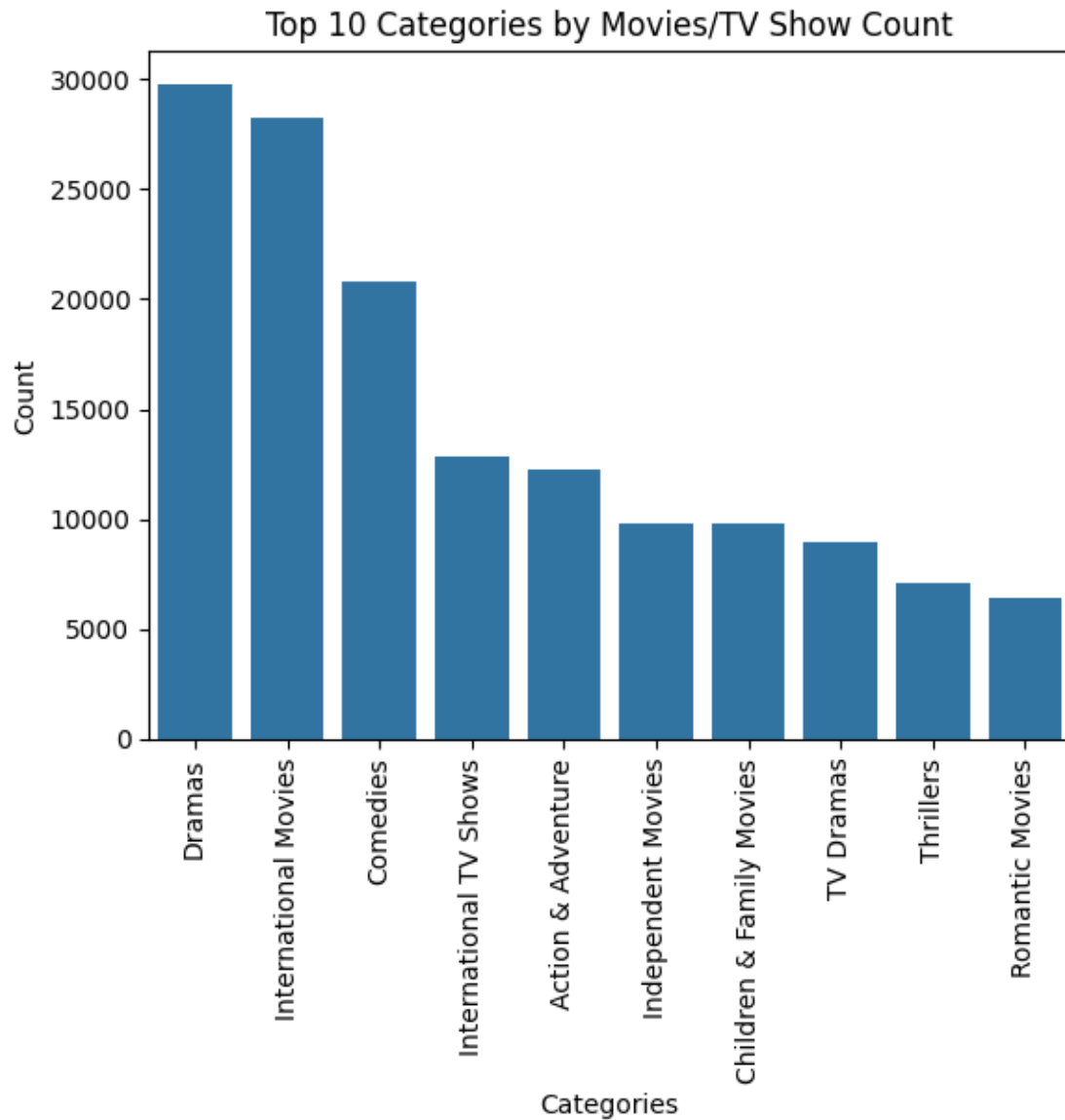
Insights - the above analysis shows which listed_in is popular in all the countries, eg United States has highest number of Dramas, meanwhile India has highest number of international movies on Netflix.

Analysis 11 - Top 10 Categories by Movies/ TV Show Count

```
[183]: listed_in_counts = df['listed_in'].value_counts()
listed_in_counts.head(10)
```

```
[183]: listed_in
Dramas                29775
International Movies   28211
Comedies              20829
International TV Shows 12845
Action & Adventure    12216
Independent Movies     9834
Children & Family Movies 9771
TV Dramas             8942
Thrillers             7107
Romantic Movies       6412
Name: count, dtype: int64
```

```
[187]: sns.barplot(x = listed_in_counts.head(10).index, y = listed_in_counts.head(10).
↳values)
plt.title('Top 10 Categories by Movies/TV Show Count')
plt.xlabel('Categories')
plt.ylabel('Count')
plt.xticks(rotation=90)
plt.show()
```



Insights - The most famous category is Dramas following with international movies and comedies

Analysis 12 - Average Duration of TV Shows and Movies on Netflix.

[251]:

```
df
```

```
[251]:
```

| | show_id | type | title | director \ |
|---|---------|---------|----------------------|-----------------|
| 0 | s1 | Movie | Dick Johnson Is Dead | Kirsten Johnson |
| 1 | s2 | TV Show | Blood & Water | Martin Scorsese |
| 1 | s2 | TV Show | Blood & Water | Martin Scorsese |
| 1 | s2 | TV Show | Blood & Water | Martin Scorsese |
| 1 | s2 | TV Show | Blood & Water | Martin Scorsese |

| | | | | | | |
|------|-------|-------|--------|-------------|-----|-----|
| ... | ... | ... | ... | ... | ... | ... |
| 8806 | s8807 | Movie | Zubaan | Mozez Singh | | |
| 8806 | s8807 | Movie | Zubaan | Mozez Singh | | |
| 8806 | s8807 | Movie | Zubaan | Mozez Singh | | |
| 8806 | s8807 | Movie | Zubaan | Mozez Singh | | |
| 8806 | s8807 | Movie | Zubaan | Mozez Singh | | |

| | cast | country | release_year | rating | duration | \ |
|---|-------------|---------------|--------------|--------|----------|---|
| 0 | Liam Neeson | United States | 2020 | PG-13 | 90 | |
| 1 | Ama Qamata | South Africa | 2021 | TV-MA | 2 | |
| 1 | Ama Qamata | South Africa | 2021 | TV-MA | 2 | |
| 1 | Ama Qamata | South Africa | 2021 | TV-MA | 2 | |
| 1 | Khosi Ngema | South Africa | 2021 | TV-MA | 2 | |

| | | | | | | |
|------|-----------------------|-------|------|-------|-----|-----|
| ... | ... | ... | ... | ... | ... | ... |
| 8806 | Anita Shabdish | India | 2015 | TV-14 | 111 | |
| 8806 | Anita Shabdish | India | 2015 | TV-14 | 111 | |
| 8806 | Chittaranjan Tripathy | India | 2015 | TV-14 | 111 | |
| 8806 | Chittaranjan Tripathy | India | 2015 | TV-14 | 111 | |
| 8806 | Chittaranjan Tripathy | India | 2015 | TV-14 | 111 | |

| | listed_in | \ |
|---|------------------------|---|
| 0 | Documentaries | |
| 1 | International TV Shows | |
| 1 | TV Dramas | |
| 1 | TV Mysteries | |
| 1 | International TV Shows | |

| | |
|------|----------------------|
| ... | ... |
| 8806 | International Movies |
| 8806 | Music & Musicals |
| 8806 | Dramas |
| 8806 | International Movies |
| 8806 | Music & Musicals |

| | description | year | month | \ |
|---|---|--------|-------|---|
| 0 | As her father nears the end of his life, filmm... | 2021.0 | 9.0 | |
| 1 | After crossing paths at a party, a Cape Town t... | 2021.0 | 9.0 | |
| 1 | After crossing paths at a party, a Cape Town t... | 2021.0 | 9.0 | |
| 1 | After crossing paths at a party, a Cape Town t... | 2021.0 | 9.0 | |
| 1 | After crossing paths at a party, a Cape Town t... | 2021.0 | 9.0 | |

| | | |
|------|---|------------|
| ... | ... | ... |
| 8806 | A scrappy but poor boy worms his way into a ty... | 2019.0 3.0 |
| 8806 | A scrappy but poor boy worms his way into a ty... | 2019.0 3.0 |
| 8806 | A scrappy but poor boy worms his way into a ty... | 2019.0 3.0 |
| 8806 | A scrappy but poor boy worms his way into a ty... | 2019.0 3.0 |
| 8806 | A scrappy but poor boy worms his way into a ty... | 2019.0 3.0 |

week_day_name

```

0          Saturday
1           Friday
1           Friday
1           Friday
1           Friday
...
8806        Saturday
8806        Saturday
8806        Saturday
8806        Saturday
8806        Saturday

```

[201991 rows x 14 columns]

```
[149]: df['duration'].fillna(df['duration'].mode()[0], inplace = True)
df['duration'].replace('nan', np.nan, inplace = True)
```

```
[148]: df['duration'] = pd.to_numeric(df['duration'], errors='coerce')
df.info()
```

```

<class 'pandas.core.frame.DataFrame'>
Index: 201991 entries, 0 to 8806
Data columns (total 14 columns):
 #   Column          Non-Null Count  Dtype
---  -
 0   show_id         201991 non-null object
 1   type            201991 non-null object
 2   title           201991 non-null object
 3   director        201991 non-null object
 4   cast            201991 non-null object
 5   country         201991 non-null object
 6   release_year    201991 non-null int64
 7   rating          201991 non-null object
 8   duration        201991 non-null int64
 9   listed_in       201991 non-null object
10   description     201991 non-null object
11   year            201991 non-null float64
12   month           201991 non-null float64
13   week_day_name   201991 non-null object
dtypes: float64(2), int64(2), object(10)
memory usage: 23.1+ MB

```

```
[155]: df.groupby('type')['duration'].mean()
```

```

[155]: type
Movie      106.854275
TV Show     1.928101

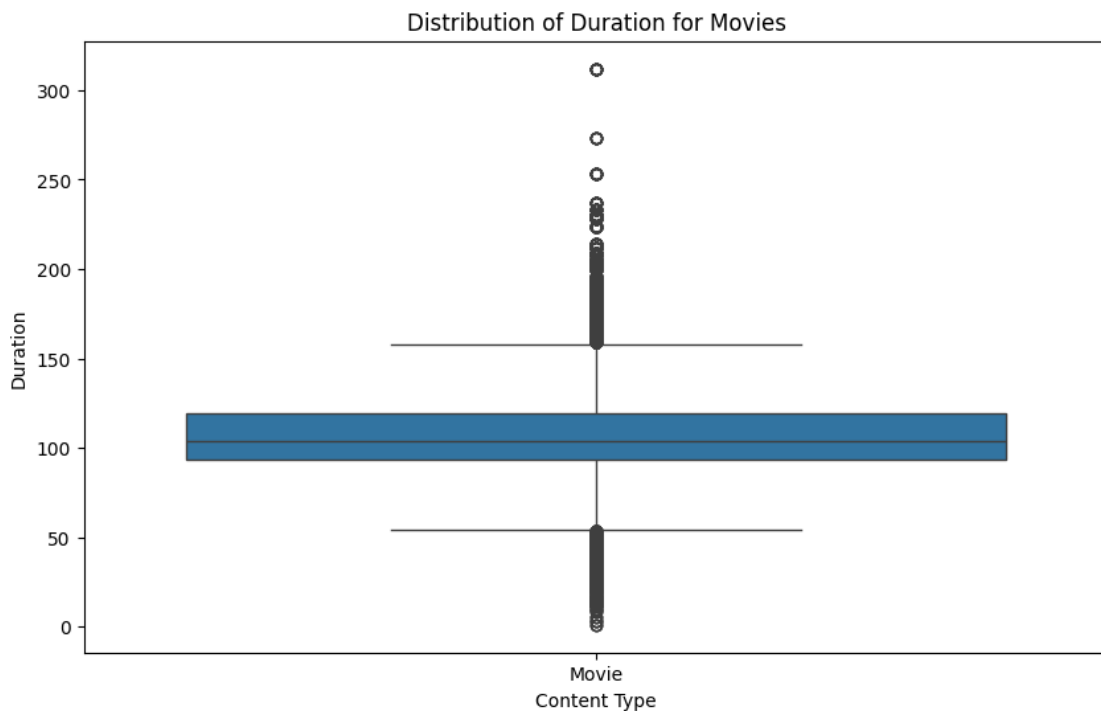
```

Name: duration, dtype: float64

Insights- Average duration for movies is approx 107 mins and for TV Shows is approx 2 Seasons.

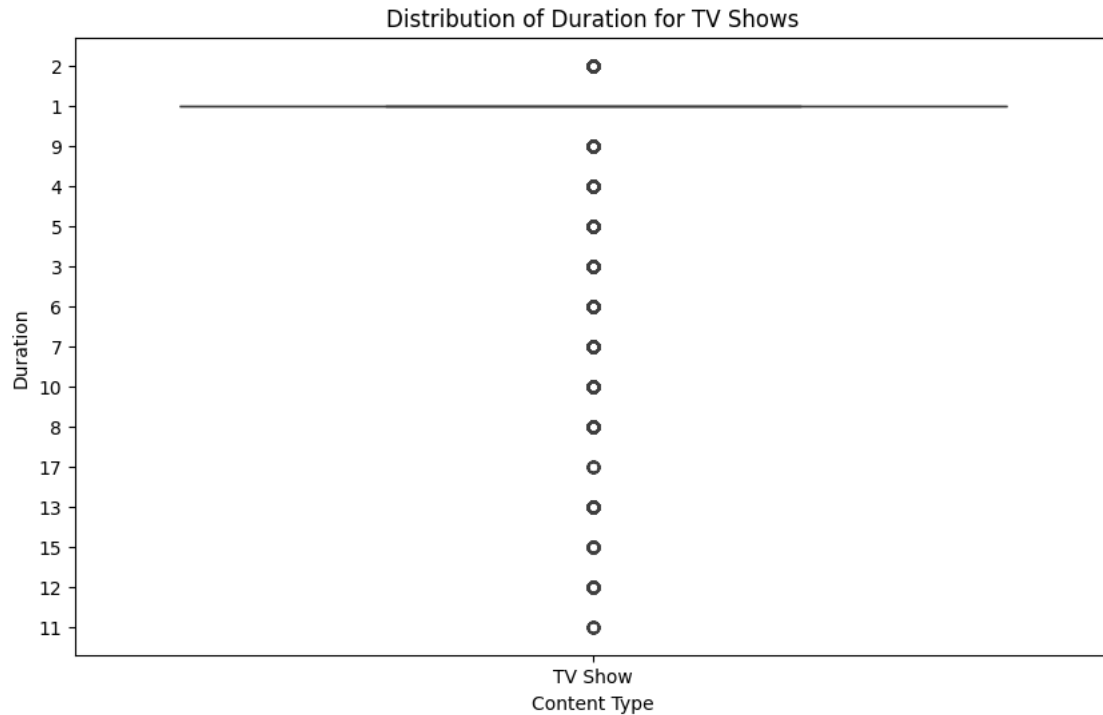
```
[252]: df_tv_shows = df[df['type'] == 'TV Show']  
df_movies = df[df['type'] == 'Movie']
```

```
[158]: plt.figure(figsize=(10, 6))  
sns.boxplot(data=df_movies, x='type', y='duration')  
plt.xlabel('Content Type')  
plt.ylabel('Duration')  
plt.title('Distribution of Duration for Movies')  
plt.show()
```



Insights- If we check for the movies boxplot, we can see that most movies fall within a reasonable duration range, with few outliers exceeding approximately 2 hours. This suggests that most movies on Netflix are designed to fit within a standard viewing time.

```
[263]: plt.figure(figsize=(10, 6))  
sns.boxplot(data=df_tv_shows, x='type', y='duration')  
plt.xlabel('Content Type')  
plt.ylabel('Duration')  
plt.title('Distribution of Duration for TV Shows')  
plt.show()
```



Insights - For TV shows, the box plot reveals that most shows have one to four seasons, with very few outliers having longer durations. This totally aligns with the earlier trends, indicating that Netflix focuses on shorter series formats.

Recommendations -

1. Netflix should also focus more on TV Shows, as our analysis states that Netflix has added more movies than TV Shows.
2. Netflix should also focus on the other ratings TV-Y7, TV-Y, TV-G.
3. The number of releases per year are high as comparison to the content added on Netflix. So Netflix should increase the number of addition of content.
4. If we check the duration of Movies, most of the movies are 2 or 2.5 hrs long which is a perfect duration for the viewers and it should be continued. However, for TV Shows i see that duration is 1 to 2 Seasons which can be extended for the viewers to be more active on the netflix format.
5. The best time to launch maximum amount of TV shows or movies can also be dependant on long holidays in specific countries, for eg. In India, the best time will be October & November(Diwali & Desehra). In USA, UK, NZ it can be December(Christmas, New Year).