New_netflix_business_case

August 12, 2024

[1]:
```
import numpy as np
import pandas as pd
import matplotlib.pyplot as plt
import seaborn as sns
```

[2]:
```
path = r"C:\Users\Yuvra\OneDrive\Desktop\Python_Files\CSV_Files\netflix.csv"

df = pd.read_csv(path)

df.head()
```

[2]:

	show_id	type	title	director	cast	country	date_added	release_year	rating	duration	listed_in
0	s1	Movie	Dick Johnson Is Dead	Kirsten Johnson	NaN	United States	September 25, 2021	2020	PG-13	90 min	Documentaries
1	s2	TV Show	Blood & Water	NaN	Ama Qamata, Khosi Ngema, Gail Mabalane, Thaban…	South Africa	September 24, 2021	2021	TV-MA	2 Seasons	International TV Shows, TV Dramas, TV Mysteries
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…
3	s4	TV Show	Jailbirds New Orleans	NaN	NaN	NaN	September 24, 2021	2021	TV-MA	1 Season	Docuseries, Reality TV
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 …

```
                                          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…
```

[3]:
```
# Dropping if any row has all the values as null
df = df.dropna(axis=0, how="all")
df.head()
```

[3]:
```
  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…
```

[4]:
```
# Checking info about our data frame
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
```

[5]:
```
# No of unique values in our data
df.nunique()
```

[5]:
```
show_id         8807
type               2
title           8807
director        4528
cast            7692
country          748
date_added      1767
release_year      74
rating            17
duration         220
listed_in        514
description     8775
dtype: int64
```

[6]:
```
# Checking no of null values in our data
df.isna().sum()
```

[6]:
```
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
```

[7]:
```
# Checking basic metrics of our data
# We can see that only one column has numeric values that is release year
df.describe()
```

[7]:
```
       release_year
count   8807.000000
mean    2014.180198
std        8.819312
min     1925.000000
25%     2013.000000
50%     2017.000000
75%     2019.000000
max     2021.000000
```

[8]:
```
# I assigned original df to seperate new_df to keep the original data frame
 ↪intact
new_df = df
new_df.head()
```

[8]:
```
  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…
```

[9]:
```
# As we can see there are some ratings in min which does not make any sense and␣
 ↪got wrongly entered in the rating column instead of duration column.
new_df["rating"].unique()
```

[9]:
```
array(['PG-13', 'TV-MA', 'PG', 'TV-14', 'TV-PG', 'TV-Y', 'TV-Y7', 'R',
       'TV-G', 'G', 'NC-17', '74 min', '84 min', '66 min', 'NR', nan,
       'TV-Y7-FV', 'UR'], dtype=object)
```

[10]:
```
index = new_df.loc[new_df["duration"].isna()].index
```

[11]:
```
new_df.loc[index] = df.loc[index].ffill(axis=1)
```

```
C:\Users\Yuvra\AppData\Local\Temp\ipykernel_6764\1112725414.py:1: FutureWarning:
Setting an item of incompatible dtype is deprecated and will raise an error in a
future version of pandas. Value '[2017 2010 2015]' has dtype incompatible with
int64, please explicitly cast to a compatible dtype first.
  new_df.loc[index] = df.loc[index].ffill(axis=1)
```

[12]:
```
new_df.loc[index, "rating"] = "NR"
```

[13]:
```
new_df["rating"].unique()
```

[13]:
```
array(['PG-13', 'TV-MA', 'PG', 'TV-14', 'TV-PG', 'TV-Y', 'TV-Y7', 'R',
       'TV-G', 'G', 'NC-17', 'NR', nan, 'TV-Y7-FV', 'UR'], dtype=object)
```

[14]:
```
new_df.loc[new_df["rating"].isna()]
```

[14]:
```
     show_id     type                                              title  \
5989   s5990    Movie  13TH: A Conversation with Oprah Winfrey & Ava …
6827   s6828  TV Show                 Gargantia on the Verdurous Planet
7312   s7313  TV Show                                      Little Lunch
7537   s7538    Movie                           My Honor Was Loyalty
```

```
             director                                                cast  \
5989              NaN                          Oprah Winfrey, Ava DuVernay
6827              NaN  Kaito Ishikawa, Hisako Kanemoto, Ai Kayano, Ka…
7312              NaN  Flynn Curry, Olivia Deeble, Madison Lu, Oisín …
7537  Alessandro Pepe  Leone Frisa, Paolo Vaccarino, Francesco Miglio…

        country        date_added release_year rating  duration  \
5989        NaN  January 26, 2017         2017    NaN    37 min
6827      Japan  December 1, 2016         2013    NaN  1 Season
7312  Australia  February 1, 2018         2015    NaN  1 Season
7537      Italy     March 1, 2017         2015    NaN   115 min

                                listed_in  \
5989                               Movies
6827  Anime Series, International TV Shows
7312                Kids' TV, TV Comedies
7537                               Dramas

                                                 description
5989  Oprah Winfrey sits down with director Ava DuVe…
6827  After falling through a wormhole, a space-dwel…
7312  Adopting a child's perspective, this show take…
7537  Amid the chaos and horror of World War II, a c…
```

[15]:
```
new_df["rating"] = new_df["rating"].fillna("NR")
```

[16]:
```
new_df["rating"].unique()
```

[16]:
```
array(['PG-13', 'TV-MA', 'PG', 'TV-14', 'TV-PG', 'TV-Y', 'TV-Y7', 'R',
       'TV-G', 'G', 'NC-17', 'NR', 'TV-Y7-FV', 'UR'], dtype=object)
```

[17]:
```
new_df.loc[new_df["rating"] == "UR", "rating"] = "NR"
```

[18]:
```
new_df["rating"].unique()
```

[18]:
```
array(['PG-13', 'TV-MA', 'PG', 'TV-14', 'TV-PG', 'TV-Y', 'TV-Y7', 'R',
       'TV-G', 'G', 'NC-17', 'NR', 'TV-Y7-FV'], dtype=object)
```

[19]:
```
new_df["rating"].value_counts()
```

[19]:
```
rating
TV-MA       3207
TV-14       2160
TV-PG        863
R            799
PG-13        490
TV-Y7        334
```

```
TV-Y          307
PG            287
TV-G          220
NR             90
G              41
TV-Y7-FV        6
NC-17           3
Name: count, dtype: int64
```

[20]:
```
new_df.shape[0]
```

[20]:
```
8807
```

[21]:
```
# Directly changing the date_added column to date time won't work and will give␣
 ↪error due to some extra space in the existing dates
# So lets first strip some extra space from this column
new_df["date_added"] = new_df["date_added"].str.strip()

# Now changing the column to date time
new_df["date_added"] = pd.to_datetime(new_df["date_added"])
```

[22]:
```
new_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   datetime64[ns]
 7   release_year  8807 non-null   object
 8   rating        8807 non-null   object
 9   duration      8807 non-null   object
 10  listed_in     8807 non-null   object
 11  description   8807 non-null   object
dtypes: datetime64[ns](1), object(11)
memory usage: 825.8+ KB
```

[23]:
```
# Here we interpolated the NaT values
new_df["date_added"] = new_df["date_added"].interpolate()
```

[24]:
```
# We converted data type of release year column to int64
new_df["release_year"] = new_df["release_year"].astype("int64")
```

[25]:
```
new_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    8807 non-null   datetime64[ns]
 7   release_year  8807 non-null   int64
 8   rating        8807 non-null   object
 9   duration      8807 non-null   object
 10  listed_in     8807 non-null   object
 11  description   8807 non-null   object
dtypes: datetime64[ns](1), int64(1), object(10)
memory usage: 825.8+ KB
```

[26]:
```
new_df.head()
```

[26]:

	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	2021-09-25	2020	PG-13	90 min	Documentaries	As her father nears the end of his life, filmm…
1	s2	TV Show	Blood & Water	NaN	Ama Qamata, Khosi Ngema, Gail Mabalane, Thaban…	South Africa	2021-09-24	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	2021-09-24	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	2021-09-24	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	2021-09-24	2021	TV-MA	2 Seasons	International TV Shows, Romantic TV Shows, TV …	In a city of coaching centers known to train I…

```
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…
```

[27]:
```
# Creating month and year column separately
new_df["year_added"] = new_df["date_added"].dt.year
new_df["month_added"] = new_df["date_added"].dt.month_name()
```

[28]:
```
new_df.head()
```

[28]:
```
  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 2021-09-25          2020  PG-13     90 min
1 2021-09-24          2021  TV-MA  2 Seasons
2 2021-09-24          2021  TV-MA   1 Season
3 2021-09-24          2021  TV-MA   1 Season
4 2021-09-24          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  year_added month_added
0  As her father nears the end of his life, filmm…        2021   September
```

```
1  After crossing paths at a party, a Cape Town t…        2021   September
2  To protect his family from a powerful drug lor…        2021   September
3  Feuds, flirtations and toilet talk go down amo…        2021   September
4  In a city of coaching centers known to train I…        2021   September
```

[908]:
```
countries = new_df[["show_id", "country"]]
countries["country"] = countries["country"].str.split(",")
countries = countries.explode("country")
countries["country"] = countries["country"].str.strip()
countries["country"] = countries["country"].replace('', np.nan)
countries = countries.reset_index(drop=True)
countries.head()
```

```
C:\Users\Yuvra\AppData\Local\Temp\ipykernel_6764\941115102.py:2:
SettingWithCopyWarning:
A value is trying to be set on a copy of a slice from a DataFrame.
Try using .loc[row_indexer,col_indexer] = value instead

See the caveats in the documentation: https://pandas.pydata.org/pandas-
docs/stable/user_guide/indexing.html#returning-a-view-versus-a-copy
  countries["country"] = countries["country"].str.split(",")
```

[908]:

	show_id	country
0	s1	United States
1	s2	South Africa
2	s3	NaN
3	s4	NaN
4	s5	India

[30]:
```
genre = new_df[["show_id","listed_in"]]
genre["listed_in"] = genre["listed_in"].str.split(",")
genre = genre.explode("listed_in")
genre["listed_in"] = genre["listed_in"].str.strip()
genre = genre.reset_index(drop=True)
genre.head()
```

```
C:\Users\Yuvra\AppData\Local\Temp\ipykernel_6764\1486207755.py:2:
SettingWithCopyWarning:
A value is trying to be set on a copy of a slice from a DataFrame.
Try using .loc[row_indexer,col_indexer] = value instead

See the caveats in the documentation: https://pandas.pydata.org/pandas-
docs/stable/user_guide/indexing.html#returning-a-view-versus-a-copy
  genre["listed_in"] = genre["listed_in"].str.split(",")
```

[30]:

	show_id	listed_in
0	s1	Documentaries
1	s2	International TV Shows

```
2      s2              TV Dramas
3      s2            TV Mysteries
4      s3          Crime TV Shows
```

[31]:
```
cast = new_df[["show_id","cast"]]
cast["cast"] = cast["cast"].str.split(",")
cast = cast.explode("cast")
cast["cast"] = cast["cast"].str.strip()
cast = cast.reset_index(drop=True)
cast.head()
```

```
C:\Users\Yuvra\AppData\Local\Temp\ipykernel_6764\637374982.py:2:
SettingWithCopyWarning:
A value is trying to be set on a copy of a slice from a DataFrame.
Try using .loc[row_indexer,col_indexer] = value instead

See the caveats in the documentation: https://pandas.pydata.org/pandas-
docs/stable/user_guide/indexing.html#returning-a-view-versus-a-copy
  cast["cast"] = cast["cast"].str.split(",")
```

[31]:

	show_id	cast
0	s1	NaN
1	s2	Ama Qamata
2	s2	Khosi Ngema
3	s2	Gail Mabalane
4	s2	Thabang Molaba

[32]:
```
directors = new_df[["show_id","director"]]
```

[33]:
```
directors = new_df[["show_id","director"]]
directors["director"] = directors["director"].str.split(",")
directors = directors.explode("director")
directors["director"] = directors["director"].str.strip()
directors = directors.reset_index(drop=True)
directors.head()
```

```
C:\Users\Yuvra\AppData\Local\Temp\ipykernel_6764\3979949296.py:2:
SettingWithCopyWarning:
A value is trying to be set on a copy of a slice from a DataFrame.
Try using .loc[row_indexer,col_indexer] = value instead

See the caveats in the documentation: https://pandas.pydata.org/pandas-
docs/stable/user_guide/indexing.html#returning-a-view-versus-a-copy
  directors["director"] = directors["director"].str.split(",")
```

[33]:

	show_id	director
0	s1	Kirsten Johnson
1	s2	NaN

```
2      s3  Julien Leclercq
3      s4              NaN
4      s5              NaN
```

[34]:
```
movies = new_df.loc[new_df["type"] == "Movie"]
movies = movies[["show_id", "title", "duration"]]
movies = movies.reset_index(drop=True)
movies = movies.rename({"title":"movies"}, axis=1)
movies
```

[34]:

	show_id	movies	duration
0	s1	Dick Johnson Is Dead	90 min
1	s7	My Little Pony: A New Generation	91 min
2	s8	Sankofa	125 min
3	s10	The Starling	104 min
4	s13	Je Suis Karl	127 min
…	…	…	…
6126	s8802	Zinzana	96 min
6127	s8803	Zodiac	158 min
6128	s8805	Zombieland	88 min
6129	s8806	Zoom	88 min
6130	s8807	Zubaan	111 min

[6131 rows x 3 columns]

[35]:
```
tv_shows = new_df.loc[new_df["type"] == "TV Show"]
tv_shows = tv_shows[["show_id", "title", "duration"]]
tv_shows = tv_shows.reset_index(drop=True)
tv_shows = tv_shows.rename({"title":"tv_shows"}, axis=1)
tv_shows
```

[35]:

	show_id	tv_shows	duration
0	s2	Blood & Water	2 Seasons
1	s3	Ganglands	1 Season
2	s4	Jailbirds New Orleans	1 Season
3	s5	Kota Factory	2 Seasons
4	s6	Midnight Mass	1 Season
…	…	…	…
2671	s8796	Yu-Gi-Oh! Arc-V	2 Seasons
2672	s8797	Yunus Emre	2 Seasons
2673	s8798	Zak Storm	3 Seasons
2674	s8801	Zindagi Gulzar Hai	1 Season
2675	s8804	Zombie Dumb	2 Seasons

[2676 rows x 3 columns]

[36]:
```
tv_shows
```

[36]:
```
     show_id               tv_shows   duration
0         s2          Blood & Water  2 Seasons
1         s3              Ganglands   1 Season
2         s4  Jailbirds New Orleans   1 Season
3         s5           Kota Factory  2 Seasons
4         s6          Midnight Mass   1 Season
…      …                        …         …
2671   s8796        Yu-Gi-Oh! Arc-V  2 Seasons
2672   s8797             Yunus Emre  2 Seasons
2673   s8798              Zak Storm  3 Seasons
2674   s8801     Zindagi Gulzar Hai   1 Season
2675   s8804            Zombie Dumb  2 Seasons

[2676 rows x 3 columns]
```

[37]:
```
titles = new_df[["show_id","title", "rating", "date_added", "year_added",␣
 ↪"month_added" ]]
titles
```

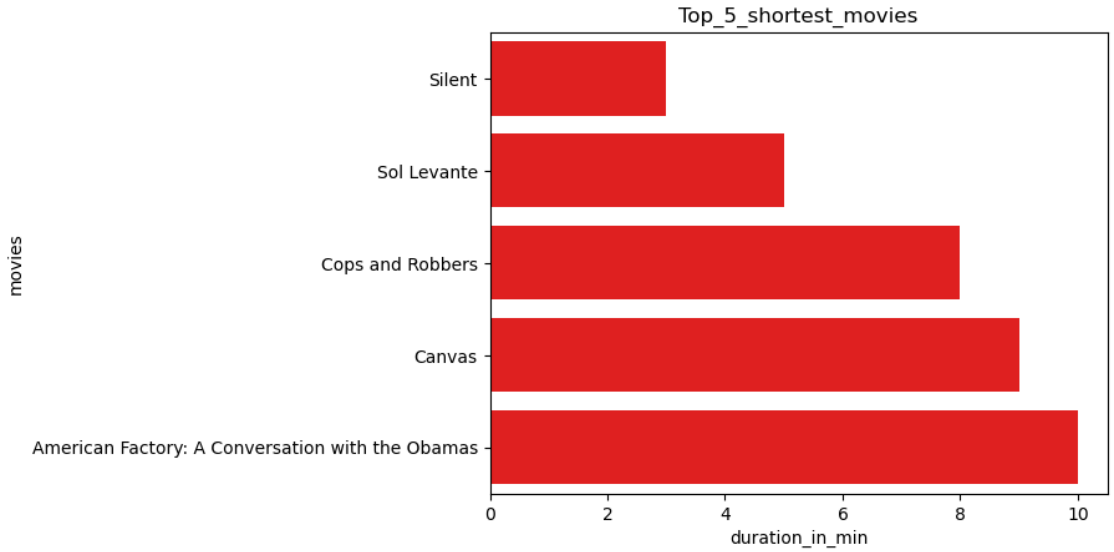
[37]:
```
     show_id                  title rating date_added  year_added month_added
0         s1   Dick Johnson Is Dead  PG-13 2021-09-25        2021   September
1         s2          Blood & Water  TV-MA 2021-09-24        2021   September
2         s3              Ganglands  TV-MA 2021-09-24        2021   September
3         s4  Jailbirds New Orleans  TV-MA 2021-09-24        2021   September
4         s5           Kota Factory  TV-MA 2021-09-24        2021   September
…      …                        …    …        …          …         …
8802   s8803                Zodiac      R 2019-11-20        2019    November
8803   s8804            Zombie Dumb  TV-Y7 2019-07-01        2019        July
8804   s8805             Zombieland      R 2019-11-01        2019    November
8805   s8806                   Zoom     PG 2020-01-11        2020     January
8806   s8807                 Zubaan  TV-14 2019-03-02        2019       March

[8807 rows x 6 columns]
```

[38]:
```
movies = movies.rename({"duration": "duration_in_min"}, axis=1)
```

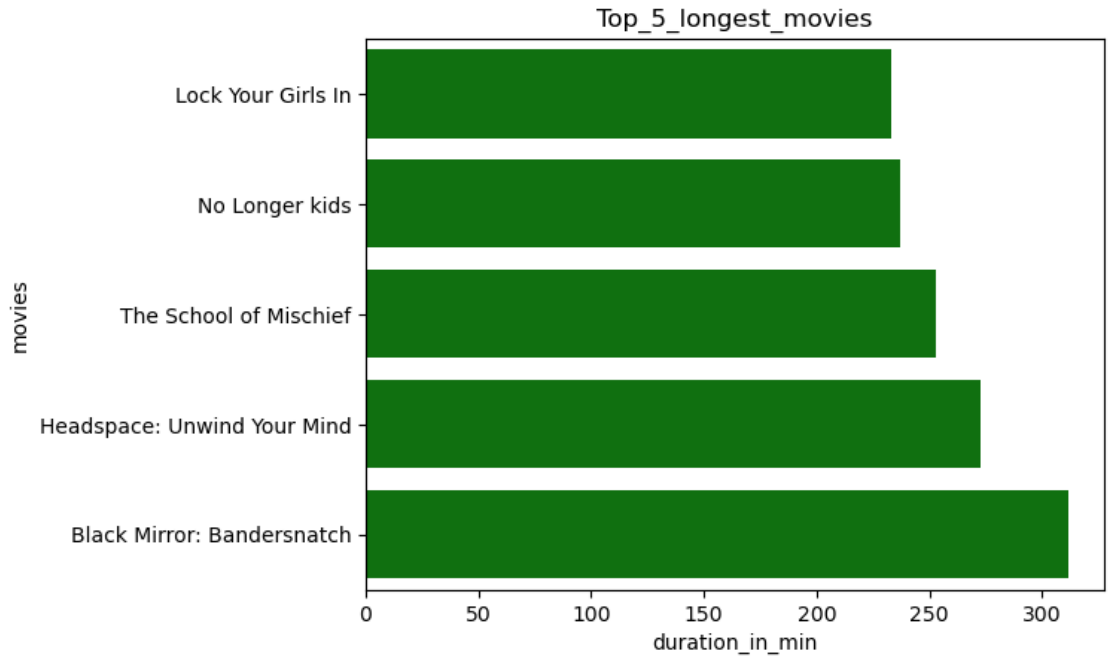
[39]:
```
movies["duration_in_min"] = movies["duration_in_min"].str.strip("min")
```

[40]:
```
movies["duration_in_min"] = movies["duration_in_min"].astype("int64")
```

[81]:
```
sns.barplot(data=movies.sort_values(by="duration_in_min").head(5),␣
 ↪x="duration_in_min", y="movies", color="red")
plt.title("Top_5_shortest_movies")
plt.show()
```



[83]:
```
sns.barplot(data=movies.sort_values(by="duration_in_min").tail(5),␣
 ↪x="duration_in_min", y="movies", color="green")
plt.title("Top_5_longest_movies")
plt.show()
```



[43]:
```
tv_shows = tv_shows.rename({"duration":"No_of_seasons"}, axis=1)
```

[45]:
```
tv_shows["No_of_seasons"] = tv_shows["No_of_seasons"].str.strip("Seasons")
tv_shows["No_of_seasons"] = tv_shows["No_of_seasons"].astype("int64")
```

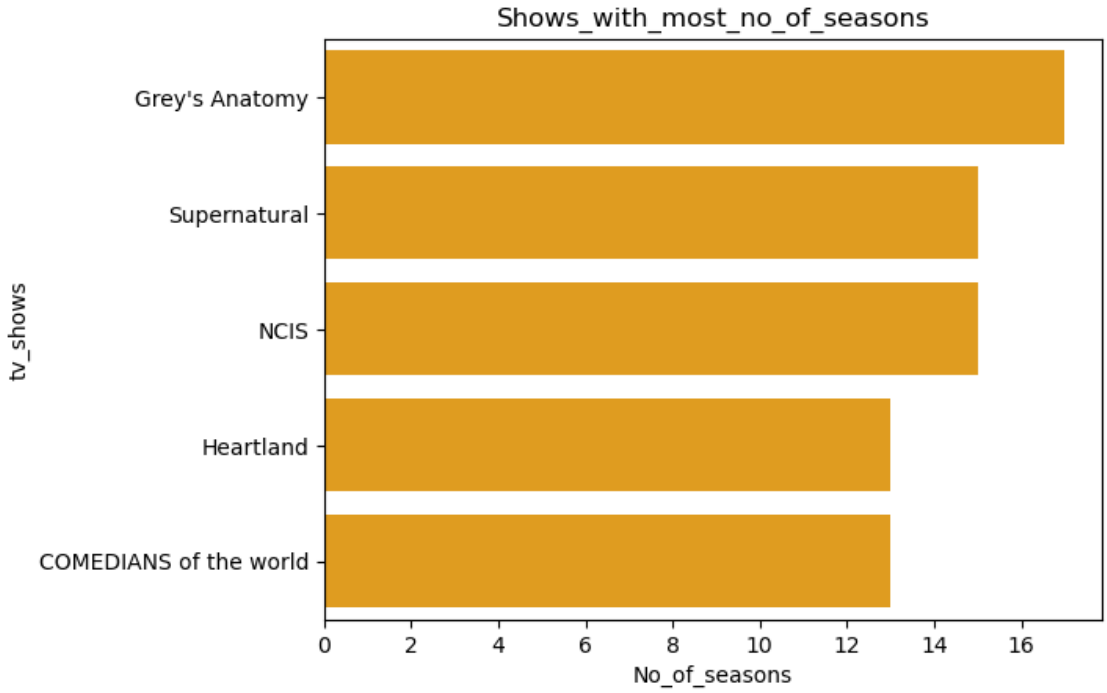
[49]:
```
tv_shows.sort_values(by="No_of_seasons", ascending=False).head(5)
```

[49]:

	show_id	tv_shows	No_of_seasons
204	s549	Grey's Anatomy	17
842	s2424	Supernatural	15
1726	s4799	NCIS	15
461	s1355	Heartland	13
1545	s4221	COMEDIANS of the world	13

[85]:
```
sns.barplot(data = tv_shows.sort_values(by="No_of_seasons", ascending=False).
 ↪head(5), y="tv_shows", x="No_of_seasons", color="orange")
plt.title("Shows_with_most_no_of_seasons")
```

[85]:
```
Text(0.5, 1.0, 'Shows_with_most_no_of_seasons')
```



[63]:
```
movies
```

[63]:

	show_id	movies	duration_in_min
0	s1	Dick Johnson Is Dead	90
1	s7	My Little Pony: A New Generation	91
2	s8	Sankofa	125

```
3        s10                    The Starling             104
4        s13                    Je Suis Karl             127
…        …                           …                   …
6126   s8802                       Zinzana                96
6127   s8803                        Zodiac               158
6128   s8805                     Zombieland               88
6129   s8806                           Zoom                88
6130   s8807                        Zubaan               111

[6131 rows x 3 columns]
```

[184]:
```
# Count of TV_shows in comparison with seasons
tv_shows_count = tv_shows.groupby("No_of_seasons")["tv_shows"].count().
 ↪reset_index()
```

[198]:
```
tv_shows_count["%"] = round((tv_shows_count["tv_shows"]/
 ↪tv_shows_count["tv_shows"].sum()*100),2)
```

[200]:
```
tv_shows_count
```

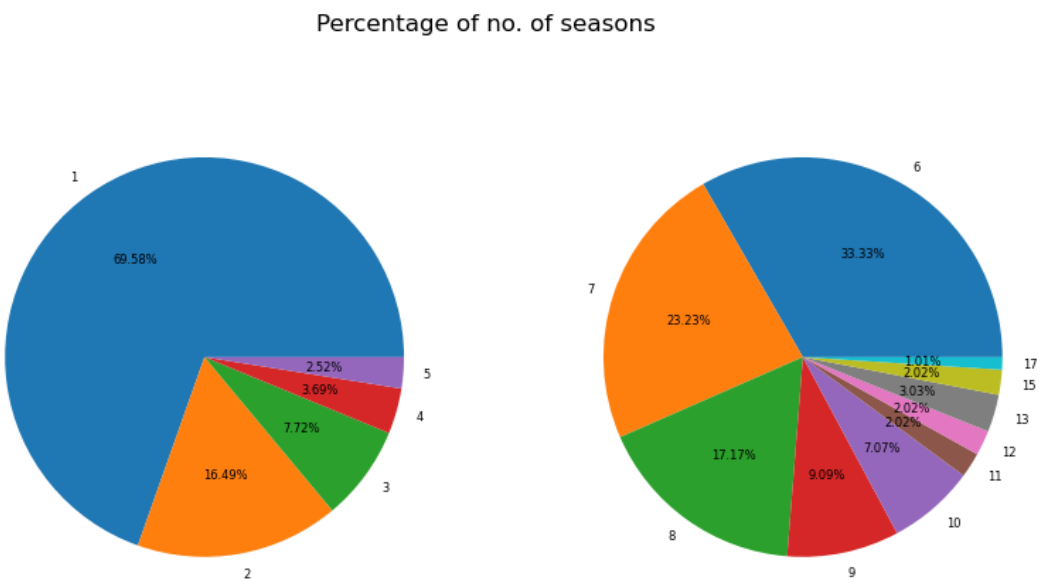
[200]:

	No_of_seasons	tv_shows	%
0	1	1793	67.00
1	2	425	15.88
2	3	199	7.44
3	4	95	3.55
4	5	65	2.43
5	6	33	1.23
6	7	23	0.86
7	8	17	0.64
8	9	9	0.34
9	10	7	0.26
10	11	2	0.07
11	12	2	0.07
12	13	3	0.11
13	15	2	0.07
14	17	1	0.04

[656]:
```
count = tv_shows_count["tv_shows"].to_list()
```

[654]:
```
plt.figure(figsize=(10,5))
plt.subplot(1,2,1)
plt.pie(count[:5], labels=tv_shows_count["No_of_seasons"].head(5), autopct= "%.
 ↪2f%%", textprops={"fontsize":6})

plt.subplot(1,2,2)
plt.pie(count[5::], labels=tv_shows_count["No_of_seasons"].tail(10), autopct=␣
 ↪"%.2f%%", textprops={"fontsize":6})
```

```
plt.suptitle("Percentage of no. of seasons")

plt.show()
```



As we can see most no of tv shows are in season 1 and then it keeps decreasing as no of season increases.

[658]:
```
movies
```

[658]:

	show_id	movies	duration_in_min
0	s1	Dick Johnson Is Dead	90
1	s7	My Little Pony: A New Generation	91
2	s8	Sankofa	125
3	s10	The Starling	104
4	s13	Je Suis Karl	127
…	…	…	…
6126	s8802	Zinzana	96
6127	s8803	Zodiac	158
6128	s8805	Zombieland	88
6129	s8806	Zoom	88
6130	s8807	Zubaan	111

[6131 rows x 3 columns]

[690]:
```
titles["type"] = new_df["type"]
```

[868]:
```
titles.groupby("type")[["title"]].count().reset_index()
```

[868]:

	type	title
0	Movie	6131
1	TV Show	2676

As we can see there are more movies than TV_shows on netflix.

[741]:
```
#countries.merge(genre, on="show_id").groupby(["listed_in"])[["country"]].
↪count().sort_values(by="country", ascending=False)
```

[765]:
```
#countries.merge(genre, on="show_id")[["listed_in"]]
```

[770]:
```
titles
```

[770]:

	show_id	title	rating	date_added	year_added	month_added	type
0	s1	Dick Johnson Is Dead	PG-13	2021-09-25	2021	September	Movie
1	s2	Blood & Water	TV-MA	2021-09-24	2021	September	TV Show
2	s3	Ganglands	TV-MA	2021-09-24	2021	September	TV Show
3	s4	Jailbirds New Orleans	TV-MA	2021-09-24	2021	September	TV Show
4	s5	Kota Factory	TV-MA	2021-09-24	2021	September	TV Show
…	…	…	…	…	…	…	…
8802	s8803	Zodiac	R	2019-11-20	2019	November	Movie
8803	s8804	Zombie Dumb	TV-Y7	2019-07-01	2019	July	TV Show
8804	s8805	Zombieland	R	2019-11-01	2019	November	Movie
8805	s8806	Zoom	PG	2020-01-11	2020	January	Movie
8806	s8807	Zubaan	TV-14	2019-03-02	2019	March	Movie

[8807 rows x 7 columns]

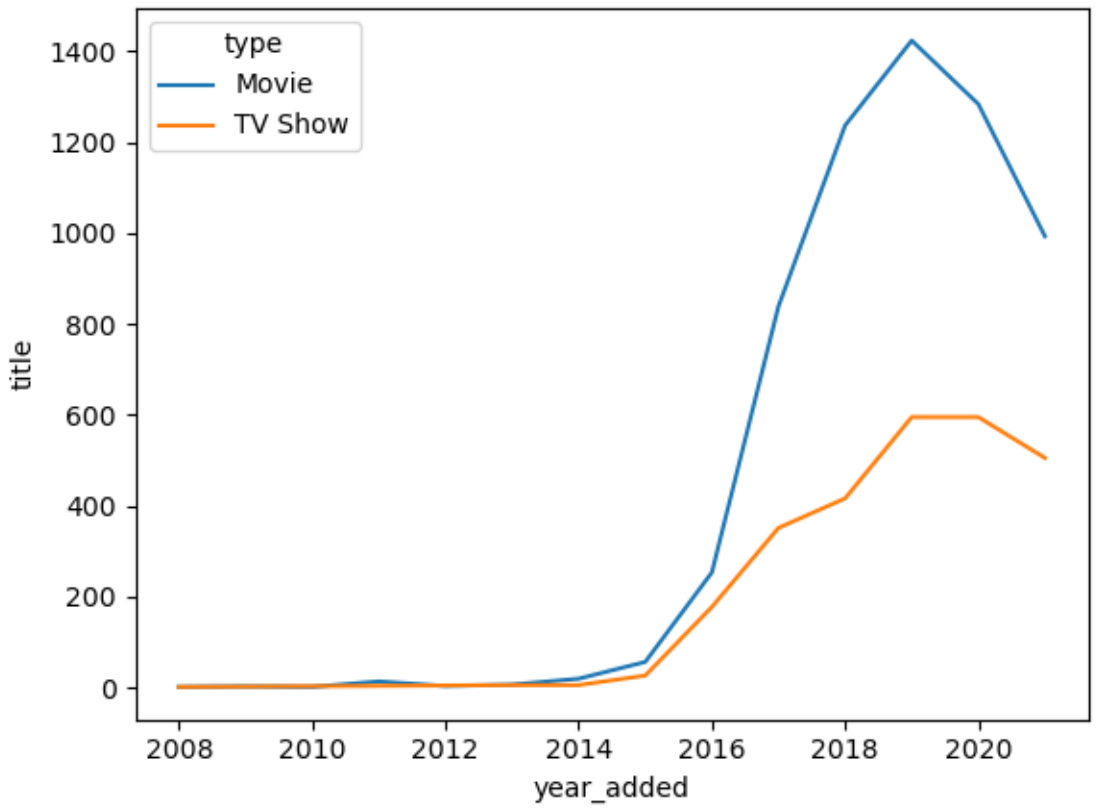
[874]:
```
count = titles.groupby(["year_added", "type"])[["title"]].count()
count
```

[874]:

		title
year_added	type	

```
2008       Movie        1
           TV Show      1
2009       Movie        2
2010       Movie        1
2011       Movie       13
2012       Movie        3
2013       Movie        6
           TV Show      5
2014       Movie       19
           TV Show      5
2015       Movie       56
           TV Show     26
2016       Movie      253
           TV Show    177
2017       Movie      839
           TV Show    351
2018       Movie     1237
           TV Show    416
2019       Movie     1424
           TV Show    595
2020       Movie     1284
           TV Show    595
2021       Movie      993
           TV Show    505
```

[872]:
```
sns.lineplot(data=count, x="year_added", y = "title", hue="type", )
plt.show()
```



As we can see there is sudden increase in the TV_shows and movies after 2015 and although still movies are added in higher numbers also observed that shows and movies started declining after 2020 may be due to worldwide lockdown.

[801]:
```
titles
```

[801]:
```
     show_id                  title rating date_added  year_added month_added  \
0         s1   Dick Johnson Is Dead  PG-13 2021-09-25        2021   September
1         s2           Blood & Water  TV-MA 2021-09-24        2021   September
2         s3               Ganglands  TV-MA 2021-09-24        2021   September
3         s4  Jailbirds New Orleans  TV-MA 2021-09-24        2021   September
4         s5           Kota Factory  TV-MA 2021-09-24        2021   September
…       …                       …    …          …           …
8802   s8803                 Zodiac      R 2019-11-20        2019    November
8803   s8804          Zombie Dumb  TV-Y7 2019-07-01        2019        July
8804   s8805            Zombieland      R 2019-11-01        2019    November
8805   s8806                  Zoom     PG 2020-01-11        2020     January
8806   s8807                Zubaan  TV-14 2019-03-02        2019       March

         type
0       Movie
```

```
1      TV Show
2      TV Show
3      TV Show
4      TV Show
…        …
8802     Movie
8803   TV Show
8804     Movie
8805     Movie
8806     Movie

[8807 rows x 7 columns]
```

[914]:
```
countries = countries.merge(genre, on="show_id")
```

[1071]:
```
countries_group = countries.groupby(["listed_in", "country"])[["listed_in"]].
↪count()
```

[1085]:
```
countries_group_new = pd.DataFrame(countries_group["listed_in"].
↪sort_values(ascending=False).head(20))
```

[1087]:
```
countries_group_new = countries_group_new.rename({"listed_in":"count"}, axis=1)
```

[1089]:
```
countries_group_new
```

[1089]:

		count
listed_in	country	
International Movies	India	864
Dramas	United States	835
Comedies	United States	680
Dramas	India	662
Documentaries	United States	512
Action & Adventure	United States	404
Children & Family Movies	United States	390
Independent Movies	United States	390
Comedies	India	323
Thrillers	United States	292
TV Comedies	United States	258
TV Dramas	United States	232
British TV Shows	United Kingdom	225
Romantic Movies	United States	225
Stand-Up Comedy	United States	216
Kids' TV	United States	214
International Movies	France	207
Horror Movies	United States	201
Dramas	United Kingdom	197
Docuseries	United States	192

As observed United states has most number of genre and then second is india and then other countries like United Kingdom and france

[ ]: