netflix

March 1, 2024

0.0.1 About Netflix

Netflix is one of the most popular media and video streaming platforms. They have over 10000 movies or tv shows available on their platform, as of mid-2021, they have over 222M Subscribers globally. This tabular dataset consists of listings of all the movies and tv shows available on Netflix, along with details such as - cast, directors, ratings, release year, duration, etc.

0.0.2 Business Problem

Analyze the data and generate insights that could help Netflix in deciding which type of shows/movies to produce and how they can grow the business in different countries.

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

[2]:
```
df = pd.read_csv("netflix.csv")
```

0.0.3 About Dataset

1. `Show_id`: Unique ID for every Movie / Tv Show
2. `Type`: Identifier - A Movie or TV Show
3. `Title`: Title of the Movie / Tv Show
4. `Director`: Director of the Movie
5. `Cast`: Actors involved in the movie/show
6. `Country`: Country where the movie/show was produced
7. `Date_added`: Date it was added on Netflix
8. `Release_year`: Actual Release year of the movie/show
9. `Rating`: TV Rating of the movie/show
10. `Duration`: Total Duration - in minutes or number of seasons
11. `Listed_in`: Genre
12. `Description`: The summary description

[3]:
```
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…
```

0.0.4 1. Defining Problem Statement and Analysing basic metrics (10 Points)

Probem statement: How has the number of movies released per year changed over the last 20-30 years?

0.0.5 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 (10 Points)

[4]:
```
print("Shape of data:", df.shape)
```

```
Shape of data: (8807, 12)
```

[5]:
```
# data type of all the attributes
df.dtypes
```

[5]:
```
show_id         object
type            object
```

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

[6]:
```
# find missing value before handling missing value
print("Missing Value in all Columns:")
df.isnull().sum()
```

```
Missing Value in all Columns:
```

[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
```

0.0.6 Handle Missing Value

[7]:
```
df = df.dropna().reset_index(drop=True)
```

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

[8]:

	show_id	type	title	director	\
0	s8	Movie	Sankofa	Haile Gerima	
1	s9	TV Show	The Great British Baking Show	Andy Devonshire	
2	s10	Movie	The Starling	Theodore Melfi	
3	s13	Movie	Je Suis Karl	Christian Schwochow	
4	s25	Movie	Jeans	S. Shankar	

	cast	\
0	Kofi Ghanaba, Oyafunmike Ogunlano, Alexandra D…	

```
1  Mel Giedroyc, Sue Perkins, Mary Berry, Paul Ho…
2  Melissa McCarthy, Chris O'Dowd, Kevin Kline, T…
3  Luna Wedler, Jannis Niewöhner, Milan Peschel, …
4  Prashanth, Aishwarya Rai Bachchan, Sri Lakshmi…

                                          country          date_added  \
0  United States, Ghana, Burkina Faso, United Kin…  September 24, 2021
1                                   United Kingdom  September 24, 2021
2                                    United States  September 24, 2021
3                         Germany, Czech Republic  September 23, 2021
4                                            India  September 21, 2021

   release_year rating   duration  \
0          1993  TV-MA    125 min
1          2021  TV-14  9 Seasons
2          2021  PG-13    104 min
3          2021  TV-MA    127 min
4          1998  TV-14    166 min

                                       listed_in  \
0  Dramas, Independent Movies, International Movies
1                    British TV Shows, Reality TV
2                              Comedies, Dramas
3                    Dramas, International Movies
4   Comedies, International Movies, Romantic Movies

                                         description
0  On a photo shoot in Ghana, an American model s…
1  A talented batch of amateur bakers face off in…
2  A woman adjusting to life after a loss contend…
3  After most of her family is murdered in a terr…
4  When the father of the man she loves insists t…
```

[9]:
```
# find missing value after handling missing value
print("Missing Value in all Columns:")
df.isnull().sum()
```

```
Missing Value in all Columns:
```

[9]:
```
show_id         0
type            0
title           0
director        0
cast            0
country         0
date_added      0
release_year    0
```

```
rating          0
duration        0
listed_in       0
description     0
dtype: int64
```

[10]:
```
# shape after handling missing value
df.shape
```

[10]:
```
(5332, 12)
```

[11]:
```
#Statistical Summary of object features
df.describe(include='object')
```

[11]:
```
       show_id   type    title                director         cast  \
count     5332   5332     5332                    5332         5332
unique    5332      2     5332                    3945         5200
top         s8  Movie  Sankofa  Raúl Campos, Jan Suter  Samuel West
freq         1   5185        1                      18           10

               country       date_added rating duration  \
count             5332             5332   5332     5332
unique             604             1453     14      198
top     United States  January 1, 2020  TV-MA   94 min
freq              1846               92   1822      135

                          listed_in  \
count                          5332
unique                          335
top     Dramas, International Movies
freq                            336

                                                 description
count                                                   5332
unique                                                  5321
top     When pretty new neighbor Seema falls for their…
freq                                                       2
```

[12]:
```
#Statistical Summary of integer features
df.describe(include='int')
```

[12]:
```
       release_year
count   5332.000000
mean    2012.742123
std        9.625831
min     1942.000000
25%     2011.000000
```

```
50%     2016.000000
75%     2018.000000
max     2021.000000
```

0.0.7 3. Non-Graphical Analysis: Value counts and unique attributes (10 Points)

[13]:
```
def non_graphica_analysis(column):
    count = df[column].value_counts()
    unique = df[column].unique()
    print("Unique Attrubutes:", unique.tolist())
    print("Value Counts:")
    print(count)
```

[14]:
```
df.columns
```

[14]:
```
Index(['show_id', 'type', 'title', 'director', 'cast', 'country', 'date_added',
       'release_year', 'rating', 'duration', 'listed_in', 'description'],
      dtype='object')
```

[15]:
```
print("-"*25+"show_id"+"-"*25)
non_graphica_analysis('show_id')
```

```
-------------------------show_id-------------------------
Unique Attrubutes: ['s8', 's9', 's10', 's13', 's25', 's28', 's29', 's30', 's39',
's42', 's43', 's44', 's45', 's47', 's49', 's52', 's53', 's54', 's55', 's57',
's58', 's59', 's60', 's61', 's62', 's63', 's64', 's74', 's82', 's85', 's91',
's95', 's97', 's106', 's108', 's115', 's116', 's117', 's119', 's123', 's127',
's128', 's130', 's132', 's134', 's135', 's136', 's137', 's138', 's139', 's140',
's141', 's142', 's143', 's144', 's145', 's146', 's147', 's150', 's151', 's152',
's153', 's156', 's157', 's158', 's159', 's160', 's162', 's163', 's164', 's165',
's167', 's168', 's169', 's170', 's171', 's172', 's173', 's174', 's175', 's176',
's177', 's178', 's179', 's180', 's183', 's184', 's189', 's191', 's192', 's193',
's196', 's199', 's200', 's201', 's202', 's203', 's204', 's205', 's206', 's207',
's208', 's209', 's210', 's211', 's212', 's216', 's217', 's218', 's228', 's229',
's230', 's232', 's248', 's252', 's254', 's260', 's265', 's268', 's271', 's274',
's277', 's279', 's280', 's282', 's283', 's285', 's286', 's293', 's294', 's296',
's298', 's299', 's300', 's301', 's302', 's304', 's305', 's306', 's307', 's308',
's309', 's315', 's316', 's317', 's326', 's327', 's328', 's329', 's330', 's331',
's333', 's334', 's335', 's336', 's338', 's341', 's342', 's343', 's344', 's345',
's346', 's347', 's348', 's349', 's350', 's351', 's352', 's353', 's354', 's355',
's356', 's357', 's358', 's359', 's360', 's361', 's366', 's375', 's378', 's381',
's384', 's390', 's393', 's394', 's403', 's411', 's416', 's427', 's430', 's434',
's437', 's438', 's440', 's446', 's447', 's448', 's449', 's450', 's451', 's452',
's453', 's461', 's464', 's470', 's475', 's487', 's490', 's492', 's493', 's495',
's497', 's503', 's504', 's507', 's509', 's519', 's529', 's530', 's532', 's535',
's542', 's543', 's552', 's555', 's556', 's557', 's558', 's559', 's560', 's562',
's563', 's564', 's565', 's567', 's568', 's569', 's570', 's573', 's574', 's575',
```

```
's8360', 's8361', 's8362', 's8363', 's8364', 's8365', 's8366', 's8368', 's8370',
's8373', 's8375', 's8376', 's8378', 's8381', 's8382', 's8383', 's8384', 's8385',
's8386', 's8387', 's8388', 's8390', 's8392', 's8394', 's8395', 's8396', 's8397',
's8399', 's8400', 's8402', 's8403', 's8405', 's8406', 's8407', 's8409', 's8410',
's8411', 's8413', 's8414', 's8415', 's8416', 's8417', 's8418', 's8419', 's8421',
's8426', 's8427', 's8428', 's8431', 's8433', 's8435', 's8436', 's8439', 's8442',
's8444', 's8445', 's8446', 's8447', 's8448', 's8449', 's8450', 's8452', 's8453',
's8454', 's8455', 's8456', 's8457', 's8458', 's8459', 's8460', 's8461', 's8462',
's8464', 's8465', 's8466', 's8467', 's8468', 's8469', 's8470', 's8471', 's8473',
's8474', 's8478', 's8479', 's8481', 's8484', 's8485', 's8486', 's8487', 's8488',
's8489', 's8490', 's8492', 's8493', 's8494', 's8495', 's8496', 's8497', 's8498',
's8499', 's8500', 's8502', 's8505', 's8506', 's8507', 's8508', 's8509', 's8510',
's8512', 's8513', 's8514', 's8515', 's8516', 's8518', 's8521', 's8523', 's8526',
's8527', 's8528', 's8529', 's8530', 's8531', 's8532', 's8533', 's8535', 's8536',
's8537', 's8538', 's8539', 's8541', 's8543', 's8544', 's8547', 's8548', 's8549',
's8551', 's8553', 's8554', 's8555', 's8556', 's8557', 's8559', 's8560', 's8563',
's8564', 's8565', 's8566', 's8567', 's8569', 's8570', 's8571', 's8572', 's8573',
's8577', 's8579', 's8580', 's8581', 's8582', 's8583', 's8584', 's8585', 's8587',
's8588', 's8590', 's8591', 's8592', 's8595', 's8596', 's8599', 's8600', 's8604',
's8607', 's8608', 's8609', 's8611', 's8612', 's8613', 's8614', 's8615', 's8616',
's8617', 's8618', 's8619', 's8622', 's8624', 's8625', 's8626', 's8627', 's8628',
's8629', 's8630', 's8631', 's8632', 's8635', 's8636', 's8637', 's8638', 's8639',
's8640', 's8641', 's8642', 's8643', 's8644', 's8646', 's8649', 's8650', 's8652',
's8653', 's8655', 's8656', 's8658', 's8659', 's8660', 's8662', 's8663', 's8664',
's8665', 's8666', 's8667', 's8668', 's8669', 's8671', 's8672', 's8673', 's8675',
's8676', 's8677', 's8678', 's8679', 's8681', 's8682', 's8683', 's8684', 's8686',
's8687', 's8688', 's8689', 's8690', 's8692', 's8693', 's8694', 's8695', 's8696',
's8697', 's8698', 's8699', 's8700', 's8704', 's8705', 's8706', 's8707', 's8709',
's8710', 's8712', 's8714', 's8715', 's8716', 's8717', 's8718', 's8720', 's8723',
's8725', 's8726', 's8730', 's8731', 's8732', 's8733', 's8734', 's8735', 's8736',
's8738', 's8743', 's8744', 's8746', 's8749', 's8751', 's8752', 's8753', 's8754',
's8755', 's8758', 's8761', 's8762', 's8763', 's8765', 's8766', 's8767', 's8768',
's8769', 's8770', 's8771', 's8772', 's8773', 's8774', 's8775', 's8777', 's8778',
's8779', 's8780', 's8782', 's8783', 's8787', 's8788', 's8789', 's8790', 's8791',
's8792', 's8793', 's8794', 's8795', 's8799', 's8800', 's8802', 's8803', 's8805',
's8806', 's8807']
Value Counts:
show_id
s8       1
s6285    1
s6283    1
s6281    1
s6279    1
        ..
s3225    1
s3224    1
s3219    1
s3218    1
```

```
s8807    1
Name: count, Length: 5332, dtype: int64
```

[16]:
```
print("-"*25+"type"+"-"*25)
non_graphica_analysis('type')
```

```
-------------------------type-------------------------
Unique Attrubutes: ['Movie', 'TV Show']
Value Counts:
type
Movie      5185
TV Show     147
Name: count, dtype: int64
```

[17]:
```
print("-"*25+"title"+"-"*25)
non_graphica_analysis('title')
```

```
-------------------------title-------------------------
Unique Attrubutes: ['Sankofa', 'The Great British Baking Show', 'The Starling',
'Je Suis Karl', 'Jeans', 'Grown Ups', 'Dark Skies', 'Paranoia', 'Birth of the
Dragon', 'Jaws', 'Jaws 2', 'Jaws 3', 'Jaws: The Revenge', 'Safe House',
'Training Day', 'InuYasha the Movie 2: The Castle Beyond the Looking Glass',
'InuYasha the Movie 3: Swords of an Honorable Ruler', 'InuYasha the Movie 4:
Fire on the Mystic Island', 'InuYasha the Movie: Affections Touching Across
Time', 'Naruto Shippuden the Movie: Blood Prison', 'Naruto Shippûden the Movie:
Bonds', 'Naruto Shippûden the Movie: The Will of Fire', 'Naruto Shippuden: The
Movie', 'Naruto Shippuden: The Movie: The Lost Tower', 'Naruto the Movie 2:
Legend of the Stone of Gelel', 'Naruto the Movie 3: Guardians of the Crescent
Moon Kingdom', 'Naruto the Movie: Ninja Clash in the Land of Snow', 'King of
Boys', 'Kate', 'Omo Ghetto: the Saga', 'Paradise Hills', 'Show Dogs', 'If I
Leave Here Tomorrow: A Film About Lynyrd Skynyrd', 'Angamaly Diaries', 'A
Champion Heart', 'Anjaam', 'Bright Star', 'Dhanak', 'Gurgaon', 'In the Cut',
'Shikara', 'A Cinderella Story', 'An Unfinished Life', 'Blade Runner: The Final
Cut', 'Chappie', 'Clear and Present Danger', 'Cliffhanger', 'Cold Mountain',
'Crocodile Dundee in Los Angeles', 'Dear John', 'Do the Right Thing', 'El
patrón, radiografía de un crimen', 'Extraction', 'Freedom Writers', 'Green
Lantern', 'House Party', 'House Party 2', 'House Party 3', 'I Got the Hook Up',
'In Too Deep', 'Initial D', 'Janoskians: Untold and Untrue', 'Labyrinth',
'Letters to Juliet', 'Level 16', "Love Don't Cost a Thing", 'Love in a Puff',
'Mars Attacks!', 'Marshall', "My Boss's Daughter", 'Mystery Men', 'Once Upon a
Time in America', 'Open Season 2', 'Osmosis Jones', 'Poseidon', 'Rhyme &
Reason', 'Same Kind of Different as Me', 'School of Rock', 'Snervous Tyler
Oakley', 'Tears of the Sun', 'The Blue Lagoon', 'The Golden Child', 'The Guns of
Navarone', 'The Interview', 'The Nutty Professor', 'Welcome Home Roscoe
Jenkins', 'In the Line of Fire', '2 Alone in Paris', 'Thimmarusu', 'Wind River',
'C Kkompany', 'EMI: Liya Hai To Chukana Padega', 'King of Boys: The Return of
the King', 'Koi Aap Sa', 'Krishna Cottage', 'Kucch To Hai', 'Kyaa Kool Hai Hum',
'Kyaa Kool Hain Hum 3', 'Kyaa Super Kool Hain Hum', 'Kyo Kii… Main Jhuth Nahin
```

```
in the Air', 'Urvi', 'Used Goods', 'USS Indianapolis: Men of Courage', 'V for
Vendetta', 'Vantage Point', 'Veerappan', 'Veerey Ki Wedding', "Viceroy's House",
'Victor', 'Viking Destiny', 'Vikings Unearthed', 'Vincent N Roxxy',
'Viswasapoorvam Mansoor', 'Vitti Dandu', 'Vizontele Tuuba', 'Vodka Diaries',
'VS.', 'Waar', 'Waarrior Savitri', 'Waiting', 'Waiting…', 'Wakefield', 'Walk
with Me', 'Walking Out', 'War', 'War Chhod Na Yaar', 'War Horse', 'War on
Everyone', 'Warda', 'Warehoused', 'We Are Your Friends', 'We Belong Together',
'We Need to Talk', 'We the Animals', "We're No Animals", 'Weather forecast',
'Weeds on Fire', 'Welcome 2 Karachi', 'Welcome to Monster High: The Origin
Story', 'Welcome to New York', 'Welcome to Willits', 'Westerplatte Resists',
'What a Wonderful Family!', 'What Lies Beneath', 'What Still Remains', "What's
Up With Love?", 'Where the Money Is', "Where's the Money", "While We're Young",
'White Chamber', 'White Island', 'Who the F**k Is That Guy?', "Who's That
Knocking at My Door?", 'Why Are We Getting So Fat?', 'Wild Wild West',
'Wildlife', 'Willy Wonka & the Chocolate Factory', 'Winter of Our Dreams',
"Winter's Bone", 'Wish I Was Here', 'Wish Man', 'Withdrawn', 'Wolves', 'World
Trade Center', 'Would You Rather', 'Wrong No.', 'Wrong Side Raju', 'Wyatt Earp',
'XX', 'XXx', 'XXX: State of the Union', 'Y Tu Mamá También', 'Y.M.I.: Yeh Mera
India', 'Yaadein', 'Yaara O Dildaara', 'Yamla Pagla Deewana 2', 'Yanda Kartavya
Aahe', '     ', 'Yellowbird', 'Yes or No', 'Yes or No 2', 'Yes or No 2.5',
'Yo-Kai Watch: The Movie', 'Yoga Hosers', 'You Can Tutu', 'You Can't Fight
Christmas', 'You Carry Me', 'You Changed My Life', "You Don't Mess with the
Zohan", 'Young Adult', 'Young Tiger', 'Yours, Mine and Ours', '    ', 'Zed
Plus', 'Zenda', 'Zinzana', 'Zodiac', 'Zombieland', 'Zoom', 'Zubaan']
Value Counts:
title
Sankofa                             1
Benji's Very Own Christmas Story    1
Beneath the Leaves                  1
Below Her Mouth                     1
Being AP                            1
                                   ..
Evvarikee Cheppoddu                 1
Defiance                            1
Holiday Rush                        1
The Island                          1
Zubaan                              1
Name: count, Length: 5332, dtype: int64
```

[18]:
```
print("-"*25+"director"+"-"*25)
non_graphica_analysis('director')
```

```
-------------------------director-------------------------
Unique Attrubutes: ['Haile Gerima', 'Andy Devonshire', 'Theodore Melfi',
'Christian Schwochow', 'S. Shankar', 'Dennis Dugan', 'Scott Stewart', 'Robert
Luketic', 'George Nolfi', 'Steven Spielberg', 'Jeannot Szwarc', 'Joe Alves',
'Joseph Sargent', 'Daniel Espinosa', 'Antoine Fuqua', 'Toshiya Shinohara',
'Masahiko Murata', 'Hajime Kamegaki', 'Hirotsugu Kawasaki', 'Toshiyuki Tsuru',
```

```
'Na Hyeon', 'John Papola', 'Grant Korgan, Brian Niles', 'Francis Ford Coppola',
'Kevin Derek', 'Shaina Allen', 'Alison MacLean', 'Steve Barker', 'Dejan
Zečević', 'Hadi El Bagoury', 'Lone Scherfig', 'N.D. Wilson', 'Curtis Hanson',
'Christian E. Christiansen', 'Bruce Robinson', 'Austin Stark', 'Kristian
Levring', 'Chris Renaud', 'Giancarlo Esposito', 'William Eubank, Will Eubank',
'Sunil Soraya', 'Krzysztof Zanussi', 'Kyle Patrick Alvarez', 'Steven R. Monroe',
'Stephan Rick', 'Tobe Hooper', 'Mohammed El-Tahawy, Mohamed Mostafa', 'Lena
Khan', 'Leanne Gater', 'Robert Schwentke', 'Brendan Malloy, Emmett Malloy',
'David Koepp', 'Ben Brewer, Alex Brewer', 'David Briggs', 'Kevin Donovan',
'Jean-Pierre Dardenne, Luc Dardenne', 'Mark Neveldine', 'Ossama Abu El Atta,
Tarek Al Eryan', 'Russell Crowe', 'Jay Russell', 'Robert Eggers', 'Kyle Rankin',
'Emil Ben-Shimon', 'Laurent Cantet', 'Michael Apted', 'Victoria Bromley',
'George Mendeluk', 'Peter Foott', 'Peter Svatek', 'Saheed Arafath', 'Charlotte
Zwerin', 'Derek Hui', 'Raam Reddy', 'Bin Bunluerit', 'Benjamin Ross', 'Stephen
Hopkins', 'Jordan Ross', 'William Wyler, John Sturges', 'Li Jun', 'Steve Loter',
'Sonny Marler', 'Beeban Kidron', 'Michael Cumming', 'Roger Spottiswoode',
'Buddhadev Dasgupta', 'Sukum Maetawanitch', 'Jeffrey Nachmanoff', 'Wally
Pfister', 'Richard Bates Jr.', 'Ron Underwood', 'Brent Maddock', 'S.S. Wilson',
'Adam Smith', 'Jitender Pawar', 'Nicholas Zeig-Owens', 'Tekin Girgin', 'Henry
Hathaway', 'Preston A. Whitmore II', 'Frank Capra, John Huston, Hugh Stewart,
Roy Boulting, Anthony Veiller', 'David Soren', 'Hidetaka Inazuka', 'Buz
Wallick', 'Gary Young', 'Benedict Andrews', 'Paul Feig', 'Chris Burkard',
'Jonathan Glazer', 'Sanjay Patel', 'Sherif Mandour', 'Jason Reitman', 'Pradeep
Verma', 'Mario Van Peebles', 'Ashu Trikha', 'Gurinder Chadha', 'Brandon
Dickerson', 'David L.G. Hughes', 'Harvey Lilley', 'Gary Michael Schultz', 'P.T.
Kunju Muhammad', 'Ganesh Kadam', 'Kushal Srivastava', 'Ed Lilly', 'Bilal
Lashari', 'Param Gill', 'Anu Menon', 'Rob McKittrick', 'Marc Francis, Max Pugh',
'Alex Smith, Andrew J. Smith', 'Philip G. Atwell', 'Faraz Haider', 'John Michael
McDonagh', 'Jack Zagha Kababie', 'Max Joseph', 'David Serrano', 'Jeremiah
Zagar', 'Antoni Krauze', 'Chi Fat Chan', 'Ashish R. Mohan', 'Stephen Donnelly,
Olly Reid, Jun Falkenstein', 'Abel Ferrara', 'Trevor Ryan', 'Huang Lei', 'Josh
Mendoza', 'Rudy Soedjarwo, Riri Riza', 'Marek Kanievska', 'Paul Raschid',
'Benjamin Turner', 'Drew Stone', 'Milla Harrison-Hansley, Alicky Sussman', 'Paul
Dano', 'Mel Stuart', 'Debra Granik', 'Zach Braff', 'Theo Davies', 'Adrian
Murray', 'Bart Freundlich', 'David Guy Levy', 'Yasir Nawaz', 'Lawrence Kasdan',
'Jovanka Vuckovic, Annie Clark, Roxanne Benjamin, Karyn Kusama', 'N. Chandra',
'Abu Bakr Shawky', 'Christian De Vita', 'Saratswadee Wongsomphet', 'Kirati
Nakintanon', 'Mark Risley', 'James Brown', 'Ivona Juka', 'Mu Chu', 'Mohamed
Diab', 'Chandra Prakash Dwivedi', 'Majid Al Ansari', 'Peter Hewitt', 'Mozez
Singh']
Value Counts:
director
Raúl Campos, Jan Suter          18
Marcus Raboy                    15
Jay Karas                       14
Cathy Garcia-Molina             13
Martin Scorsese                 12
                                ..
```

```
Igor Kovalyov, Norton Virgien     1
Danny Cannon                      1
Mana Yasuda                       1
Gupse Özay                        1
Mozez Singh                       1
Name: count, Length: 3945, dtype: int64
```

[19]:
```
print("-"*25+"cast"+"-"*25)
non_graphica_analysis('cast')
```

```
-------------------------cast-------------------------
Unique Attrubutes: ['Kofi Ghanaba, Oyafunmike Ogunlano, Alexandra Duah, Nick
Medley, Mutabaruka, Afemo Omilami, Reggie Carter, Mzuri', 'Mel Giedroyc, Sue
Perkins, Mary Berry, Paul Hollywood', "Melissa McCarthy, Chris O'Dowd, Kevin
Kline, Timothy Olyphant, Daveed Diggs, Skyler Gisondo, Laura Harrier, Rosalind
Chao, Kimberly Quinn, Loretta Devine, Ravi Kapoor", 'Luna Wedler, Jannis
Niewöhner, Milan Peschel, Edin Hasanović, Anna Fialová, Marlon Boess, Victor
Boccard, Fleur Geffrier, Aziz Dyab, Mélanie Fouché, Elizaveta Maximová',
'Prashanth, Aishwarya Rai Bachchan, Sri Lakshmi, Nassar', 'Adam Sandler, Kevin
James, Chris Rock, David Spade, Rob Schneider, Salma Hayek, Maria Bello, Maya
Rudolph, Colin Quinn, Tim Meadows, Joyce Van Patten', 'Keri Russell, Josh
Hamilton, J.K. Simmons, Dakota Goyo, Kadan Rockett, L.J. Benet, Rich Hutchman,
Myndy Crist, Annie Thurman, Jake Brennan', 'Liam Hemsworth, Gary Oldman, Amber
Heard, Harrison Ford, Lucas Till, Embeth Davidtz, Julian McMahon, Josh Holloway,
Richard Dreyfuss, Angela Sarafyan', 'Billy Magnussen, Ron Yuan, Qu Jingjing,
Terry Chen, Vanness Wu, Jin Xing, Philip Ng, Xia Yu, Yu Xia', 'Roy Scheider,
Robert Shaw, Richard Dreyfuss, Lorraine Gary, Murray Hamilton, Carl Gottlieb,
Jeffrey Kramer, Susan Backlinie, Jonathan Filley, Ted Grossman', 'Roy Scheider,
Lorraine Gary, Murray Hamilton, Joseph Mascolo, Jeffrey Kramer, Collin Wilcox
Paxton, Ann Dusenberry, Mark Gruner, Barry Coe, Susan French', 'Dennis Quaid,
Bess Armstrong, Simon MacCorkindale, Louis Gossett Jr., John Putch, Lea
Thompson, P.H. Moriarty, Dan Blasko, Liz Morris, Lisa Maurer', 'Lorraine Gary,
Lance Guest, Mario Van Peebles, Karen Young, Michael Caine, Judith Barsi,
Mitchell Anderson, Lynn Whitfield', 'Denzel Washington, Ryan Reynolds, Vera
Farmiga, Brendan Gleeson, Sam Shepard, Rubén Blades, Nora Arnezeder, Robert
Patrick, Liam Cunningham, Joel Kinnaman', 'Denzel Washington, Ethan Hawke, Scott
Glenn, Tom Berenger, Harris Yulin, Raymond J. Barry, Cliff Curtis, Dr. Dre,
Snoop Dogg, Macy Gray, Eva Mendes', 'Kappei Yamaguchi, Satsuki Yukino, Mieko
Harada, Koji Tsujitani, Houko Kuwashima, Kumiko Watanabe, Noriko Hidaka, Kenichi
Ogata, Toshiyuki Morikawa, Izumi Ogami', 'Kappei Yamaguchi, Satsuki Yukino, Koji
Tsujitani, Houko Kuwashima, Kumiko Watanabe, Ken Narita, Akio Otsuka, Kikuko
Inoue', 'Kappei Yamaguchi, Satsuki Yukino, Koji Tsujitani, Houko Kuwashima,
Kumiko Watanabe, Noriko Hidaka, Ken Narita, Cho, Mamiko Noto, Nobutoshi Canna',
'Kappei Yamaguchi, Satsuki Yukino, Koji Tsujitani, Houko Kuwashima, Kumiko
Watanabe, Kenichi Ogata, Noriko Hidaka, Hisako Kyoda, Ken Narita, Tomokazu
Seki', 'Junko Takeuchi, Chie Nakamura, Rikiya Koyama, Kazuhiko Inoue, Masaki
Terasoma, Mie Sonozaki, Yuichi Nakamura, Kengo Kawanishi, Kosei Hirota, Masako
Katsuki', 'Junko Takeuchi, Chie Nakamura, Noriaki Sugiyama, Unsho Ishizuka,
```

```
7
Kevin Hart
6
Bill Burr
5
Jim Gaffigan
5
                                                        ..
Ivie Okujaye, Sambasa Nzeribe, Tope Tedela, Gina Castel, Majid Michel, Victor
Erabie, Brutus Richard, Adebayo Thomas
1
Yoshitsugu Matsuoka, Ai Kayano, Yoko Hikasa, Yukari Tamura, Yuka Iguchi, Mamiko
Noto, Miyuki Sawashiro, Rie Kugimiya, Toru Okawa, Kazuyuki Okitsu, Junichi
Suwabe, Lynn, Yui Horie, Nanako Mori, Ryota Takeuchi     1
Scarlett Johansson, Adam Driver, Laura Dern, Alan Alda, Ray Liotta, Julie
Hagerty, Merritt Wever, Azhy Robertson
1
Chigul, Patrick Diabuah, Adetomiwa Edun, Uche Jumbo, Saidi Balogun, Ali Nuhu,
Bimbo Manuel, Akah Nnani, Damilola Adegbite, Dorcas Shola Fapson
1
Vicky Kaushal, Sarah-Jane Dias, Raaghav Chanana, Manish Chaudhary, Meghna Malik,
Malkeet Rauni, Anita Shabdish, Chittaranjan Tripathy
1
Name: count, Length: 5200, dtype: int64
```

[20]:
```
print("-"*25+"country"+"-"*25)
non_graphica_analysis('country')
```

```
-------------------------country-------------------------
Unique Attrubutes: ['United States, Ghana, Burkina Faso, United Kingdom,
Germany, Ethiopia', 'United Kingdom', 'United States', 'Germany, Czech
Republic', 'India', 'United States, India, France', 'China, Canada, United
States', 'South Africa, United States, Japan', 'Japan', 'Nigeria', 'Spain,
United States', 'United Kingdom, United States', 'United Kingdom, Australia,
France', 'United Kingdom, Australia, France, United States', 'United States,
Canada', 'Germany, United States', 'South Africa, United States', 'United
States, Mexico', 'United States, Italy, France, Japan', 'United States, Italy,
Romania, United Kingdom', 'Australia, United States', 'Argentina, Venezuela',
'United States, United Kingdom, Canada', 'China, Hong Kong', 'Canada', 'Hong
Kong', 'United States, China, Hong Kong', 'Italy, United States', 'United
States, Germany', 'France', 'United Kingdom, Canada, United States', 'United
States, United Kingdom', 'India, Nepal', 'New Zealand, Australia, France, United
States', 'Italy, Brazil, Greece', 'Spain', 'Colombia', 'United States, Japan',
'Mexico', 'Switzerland, United Kingdom, Australia', 'South Africa', 'Canada,
United States', 'Argentina', 'Argentina, Spain', 'United States, Nigeria',
'Taiwan', 'Bulgaria, United States', 'Spain, United Kingdom, United States',
'United States, China', 'United States, France', 'Spain, France, United Kingdom,
United States', ', France, Algeria', 'Poland', 'France, Israel, Germany, United
```

```
France, Poland, United States', 'India, Switzerland', 'Canada, India', 'United
States, Morocco', 'Singapore, Japan, France', 'Canada, Mexico, Germany, South
Africa', 'United Kingdom, United States, Canada', 'Germany, France, United
States, Canada, United Kingdom', 'India, Canada', 'Ireland, Canada, United
Kingdom, United States', 'United States, Germany, Australia', 'Australia,
France, Ireland', 'Australia, India', 'United States, United Kingdom, Canada,
Japan', 'Sweden, United Kingdom, Finland', 'Hong Kong, Taiwan', 'United States,
United Kingdom, Spain, South Korea', 'Guatemala', 'United States, Germany,
United Kingdom, Australia', 'Italy, France, Switzerland', 'Canada, France,
United States', 'Switzerland, United States', 'Thailand, Canada, United States',
'China, Hong Kong, United States', 'United Kingdom, New Zealand', 'Czech
Republic, United Kingdom, France', 'Australia, United Kingdom, Canada',
'Jamaica, United States', 'Australia, United Kingdom', 'United Kingdom, France,
Canada, Belgium, United States', 'Denmark, United Kingdom, Sweden', 'United
States, Hong Kong', 'Argentina, France, United States, Germany, Qatar', 'United
States, Germany, United Kingdom', 'United States, Germany, United Kingdom,
Italy', 'United States, New Zealand, United Kingdom', 'Finland, United States',
'Spain, France, Uruguay', 'France, Canada, United States', 'United States,
Canada, China', 'Ireland, Canada, Luxembourg, United States, United Kingdom,
Philippines, India', 'United States, Czech Republic, United Kingdom', 'Israel,
Germany', 'Mexico, France', 'Israel, Germany, Poland, Luxembourg, Belgium,
France, United States', 'Austria, United States', 'United Kingdom, Lithuania',
'United States, Greece, United Kingdom', 'United Kingdom, China, United States,
India', 'United States, Sweden, Norway', 'United Kingdom, United States,
Morocco', 'United States, United Kingdom, Morocco', 'Spain, Canada, United
States', 'United States, India, United Arab Emirates', 'United Kingdom, Canada,
France, United States', 'Belgium, Ireland, Netherlands, Germany, Afghanistan',
'France, Canada, Italy, United States, China', 'Ireland, United Kingdom, Greece,
France, Netherlands', 'New Zealand, United States', 'United States, Australia,
South Africa, United Kingdom', 'United States, Germany, Mexico', 'United States,
Czech Republic', 'Somalia, Kenya, Sudan, South Africa, United States', 'United
States, Canada, Japan, Panama', 'United Kingdom, Spain, Belgium', 'Serbia, South
Korea, Slovenia', 'Denmark, United Kingdom, South Africa, Sweden, Belgium',
'Germany, Canada, United States', 'New Zealand, United Kingdom, Australia',
'United Kingdom, Australia, Canada, United States', 'United Kingdom, China,
United States', 'United Kingdom, Brazil, Germany', 'Canada, United States,
India, United Kingdom', 'Switzerland, United Kingdom, United States', 'United
States, United Kingdom, Germany', 'United Kingdom, India, Sweden', 'United
States, Argentina', 'Germany, United States, United Kingdom, Canada', 'United
States, East Germany, West Germany', 'Egypt, Austria, United States', 'Croatia,
Slovenia, Serbia, Montenegro', 'United Arab Emirates, Jordan']
Value Counts:
country
United States                                  1846
India                                           875
United Kingdom                                  183
Canada                                          107
Spain                                            91
```

```
                                                    ...
Uruguay, Guatemala                                    1
Romania, Bulgaria, Hungary                            1
Philippines, United States                            1
India, United Kingdom, Canada, United States          1
United Arab Emirates, Jordan                          1
Name: count, Length: 604, dtype: int64
```

[21]:
```
print("-"*25+"date_added"+"-"*25)
non_graphica_analysis('date_added')
```

```
-------------------------date_added-------------------------
Unique Attrubutes: ['September 24, 2021', 'September 23, 2021', 'September 21,
2021', 'September 20, 2021', 'September 19, 2021', 'September 16, 2021',
'September 15, 2021', 'September 14, 2021', 'September 10, 2021', 'September 9,
2021', 'September 8, 2021', 'September 7, 2021', 'September 5, 2021', 'September
4, 2021', 'September 2, 2021', 'September 1, 2021', 'August 31, 2021', 'August
28, 2021', 'August 27, 2021', 'August 25, 2021', 'August 20, 2021', 'August 19,
2021', 'August 18, 2021', 'August 16, 2021', 'August 15, 2021', 'August 13,
2021', 'August 12, 2021', 'August 11, 2021', 'August 8, 2021', 'August 7, 2021',
'August 6, 2021', 'August 5, 2021', 'August 4, 2021', 'August 3, 2021', 'August
1, 2021', 'July 30, 2021', 'July 29, 2021', 'July 28, 2021', 'July 27, 2021',
'July 24, 2021', 'July 23, 2021', 'July 22, 2021', 'July 21, 2021', 'July 20,
2021', 'July 19, 2021', 'July 17, 2021', 'July 16, 2021', 'July 15, 2021', 'July
14, 2021', 'July 13, 2021', 'July 10, 2021', 'July 9, 2021', 'July 8, 2021',
'July 7, 2021', 'July 6, 2021', 'July 2, 2021', 'July 1, 2021', 'June 30, 2021',
'June 28, 2021', 'June 26, 2021', 'June 25, 2021', 'June 24, 2021', 'June 22,
2021', 'June 21, 2021', 'June 19, 2021', 'June 18, 2021', 'June 17, 2021', 'June
16, 2021', 'June 15, 2021', 'June 13, 2021', 'June 11, 2021', 'June 10, 2021',
'June 9, 2021', 'June 7, 2021', 'June 5, 2021', 'June 4, 2021', 'June 3, 2021',
'June 2, 2021', 'May 30, 2021', 'May 28, 2021', 'May 27, 2021', 'May 26, 2021',
'May 25, 2021', 'May 21, 2021', 'May 20, 2021', 'May 19, 2021', 'May 18, 2021',
'May 16, 2021', 'May 15, 2021', 'May 14, 2021', 'May 13, 2021', 'May 12, 2021',
'May 11, 2021', 'May 10, 2021', 'May 9, 2021', 'May 8, 2021', 'May 7, 2021',
'May 6, 2021', 'May 5, 2021', 'May 4, 2021', 'May 1, 2021', 'April 30, 2021',
'April 28, 2021', 'April 27, 2021', 'April 24, 2021', 'April 23, 2021', 'April
22, 2021', 'April 21, 2021', 'April 19, 2021', 'April 17, 2021', 'April 16,
2021', 'April 15, 2021', 'April 14, 2021', 'April 12, 2021', 'April 10, 2021',
'April 9, 2021', 'April 8, 2021', 'April 5, 2021', 'April 4, 2021', 'April 3,
2021', 'April 2, 2021', 'April 1, 2021', 'March 31, 2021', 'March 30, 2021',
'March 29, 2021', 'March 28, 2021', 'March 26, 2021', 'March 25, 2021', 'March
23, 2021', 'March 22, 2021', 'March 20, 2021', 'March 19, 2021', 'March 18,
2021', 'March 17, 2021', 'March 16, 2021', 'March 15, 2021', 'March 12, 2021',
'March 11, 2021', 'March 10, 2021', 'March 8, 2021', 'March 5, 2021', 'March 4,
2021', 'March 3, 2021', 'March 2, 2021', 'March 1, 2021', 'February 26, 2021',
'February 25, 2021', 'February 24, 2021', 'February 23, 2021', 'February 21,
2021', 'February 20, 2021', 'February 19, 2021', 'February 18, 2021', 'February
16, 2021', 'February 15, 2021', 'February 14, 2021', 'February 13, 2021',
```

```
19, 2016', 'July 29, 2019', 'March 19, 2016', 'September 29, 2015', 'July 7,
2015', 'July 6, 2017', 'April 9, 2019', 'October 17, 2018', 'October 30, 2018',
'June 12, 2018', 'December 1, 2015', 'December 17, 2014', 'March 10, 2018',
'September 9, 2019', 'April 14, 2018', 'January 3, 2019', 'August 25, 2019',
'November 27, 2018', 'May 1, 2016', 'December 18, 2014', 'March 9, 2017', 'July
13, 2019', 'April 22, 2015', 'February 19, 2019', 'January 1, 2016', 'April 6,
2019', 'June 11, 2019', 'May 26, 2019', 'December 11, 2017', 'March 24, 2018',
'June 13, 2016', ' May 18, 2017', 'May 10, 2015', 'October 14, 2013', 'April 22,
2017', 'July 10, 2017', 'May 19, 2019', 'July 6, 2019', 'December 3, 2018',
'February 14, 2016', 'July 4, 2019', 'January 16, 2019', 'October 2, 2017',
'September 22, 2015', 'December 18, 2017', 'July 22, 2020', 'January 10, 2021',
'June 1, 2016', 'February 11, 2017', 'October 14, 2015', 'July 10, 2016', 'May
24, 2018', 'February 12, 2019', 'May 10, 2017', 'May 18, 2019', 'October 10,
2015', 'February 20, 2018', 'March 23, 2020', 'February 4, 2017', 'May 8, 2016',
'July 23, 2019', 'April 26, 2017', 'August 31, 2017', 'September 19, 2018',
'April 6, 2017', 'September 20, 2016', 'September 11, 2017', 'January 5, 2021',
'December 5, 2016', 'April 18, 2020', 'February 26, 2018', 'September 26, 2018',
'June 19, 2017', 'July 8, 2016', 'July 7, 2019', 'June 2, 2018', 'June 4, 2018',
'January 13, 2020', 'May 26, 2016', 'October 26, 2019', 'September 30, 2017',
'December 15, 2015', 'July 15, 2015', 'June 15, 2014', 'November 18, 2017',
'October 21, 2018', 'June 17, 2017', 'April 9, 2018', 'October 7, 2019', 'March
8, 2017', 'September 17, 2018', 'November 10, 2018', 'October 14, 2017', 'June
5, 2018', 'November 2, 2016', 'May 17, 2015', 'February 2, 2017', 'September 12,
2019', 'July 24, 2017', 'February 11, 2019', 'September 27, 2018', 'March 23,
2017', 'February 6, 2018', 'January 22, 2018', 'October 23, 2017', 'September
25, 2017', 'August 5, 2015', 'December 6, 2018', 'March 9, 2016', 'January 11,
2020']
Value Counts:
date_added
January 1, 2020      92
March 1, 2018        67
November 1, 2019     67
December 31, 2019    64
October 1, 2018      60
                     ..
June 27, 2017         1
June 21, 2017         1
April 5, 2021         1
April 8, 2021         1
January 11, 2020      1
Name: count, Length: 1453, dtype: int64
```

[22]:
```
print("-"*25+"release_year"+"-"*25)
non_graphica_analysis('release_year')
```

```
-------------------------release_year-------------------------
Unique Attrubutes: [1993, 2021, 1998, 2010, 2013, 2017, 1975, 1978, 1983, 1987,
2012, 2001, 2002, 2003, 2004, 2011, 2008, 2009, 2007, 2005, 2006, 2018, 2020,
```

```
2019, 1994, 2015, 1982, 1989, 2014, 1990, 1991, 1999, 2016, 1986, 1996, 1984,
1997, 1980, 1961, 1995, 1985, 1992, 2000, 1976, 1959, 1988, 1972, 1981, 1964,
1954, 1979, 1958, 1956, 1963, 1970, 1973, 1960, 1974, 1966, 1971, 1962, 1969,
1977, 1967, 1968, 1965, 1945, 1946, 1955, 1942, 1947, 1944]
Value Counts:
release_year
2017    657
2018    648
2016    577
2019    519
2020    442
       …
1946      1
1961      1
1942      1
1947      1
1944      1
Name: count, Length: 72, dtype: int64
```

[23]:
```
print("-"*25+"rating"+"-"*25)
non_graphica_analysis('rating')
```

```
-------------------------rating-------------------------
Unique Attrubutes: ['TV-MA', 'TV-14', 'PG-13', 'PG', 'R', 'TV-PG', 'G', 'TV-Y7',
'TV-G', 'TV-Y', 'NC-17', 'NR', 'TV-Y7-FV', 'UR']
Value Counts:
rating
TV-MA       1822
TV-14       1214
R            778
PG-13        470
TV-PG        431
PG           275
TV-G          84
TV-Y7         76
TV-Y          76
NR            58
G             40
TV-Y7-FV       3
UR             3
NC-17          2
Name: count, dtype: int64
```

[24]:
```
print("-"*25+"duration"+"-"*25)
non_graphica_analysis('duration')
```

```
-------------------------duration-------------------------
Unique Attrubutes: ['125 min', '9 Seasons', '104 min', '127 min', '166 min',
```

```
'103 min', '97 min', '106 min', '96 min', '124 min', '116 min', '98 min', '91
min', '115 min', '122 min', '99 min', '88 min', '100 min', '102 min', '93 min',
'95 min', '85 min', '83 min', '182 min', '147 min', '90 min', '128 min', '143
min', '119 min', '114 min', '118 min', '108 min', '117 min', '121 min', '142
min', '113 min', '154 min', '120 min', '82 min', '94 min', '109 min', '101 min',
'105 min', '86 min', '229 min', '76 min', '89 min', '110 min', '156 min', '112
min', '129 min', '107 min', '1 Season', '135 min', '136 min', '165 min', '150
min', '133 min', '145 min', '92 min', '2 Seasons', '64 min', '59 min', '111
min', '87 min', '148 min', '189 min', '141 min', '130 min', '7 Seasons', '68
min', '131 min', '126 min', '155 min', '123 min', '84 min', '4 Seasons', '13
min', '77 min', '74 min', '49 min', '72 min', '78 min', '70 min', '132 min',
'140 min', '81 min', '138 min', '149 min', '15 min', '224 min', '162 min', '60
min', '65 min', '137 min', '75 min', '32 min', '158 min', '164 min', '173 min',
'181 min', '73 min', '6 Seasons', '24 min', '80 min', '139 min', '151 min', '22
min', '134 min', '58 min', '52 min', '71 min', '161 min', '53 min', '8 min', '46
min', '57 min', '28 min', '66 min', '50 min', '79 min', '26 min', '54 min', '48
min', '45 min', '171 min', '42 min', '27 min', '51 min', '69 min', '47 min', '33
min', '44 min', '29 min', '146 min', '61 min', '63 min', '157 min', '203 min',
'25 min', '30 min', '62 min', '194 min', '55 min', '5 Seasons', '15 Seasons',
'177 min', '237 min', '195 min', '253 min', '152 min', '67 min', '190 min', '160
min', '208 min', '180 min', '144 min', '174 min', '170 min', '192 min', '209
min', '187 min', '185 min', '172 min', '186 min', '8 Seasons', '193 min', '176
min', '17 min', '56 min', '169 min', '40 min', '20 min', '12 min', '168 min',
'153 min', '159 min', '214 min', '31 min', '163 min', '14 min', '179 min', '38
min', '23 min', '3 Seasons', '43 min', '200 min', '196 min', '167 min', '41
min', '37 min', '35 min', '178 min', '228 min', '18 min', '205 min', '201 min',
'191 min']
Value Counts:
duration
94 min     135
93 min     130
97 min     129
95 min     127
90 min     123
          …
189 min      1
229 min      1
237 min      1
214 min      1
191 min      1
Name: count, Length: 198, dtype: int64
```

[25]:
```
print("-"*25+"listed_in"+"-"*25)
non_graphica_analysis('listed_in')
```

```
-------------------------listed_in-------------------------
Unique Attrubutes: ['Dramas, Independent Movies, International Movies', 'British
TV Shows, Reality TV', 'Comedies, Dramas', 'Dramas, International Movies',
```

```
Movies, Romantic Movies', 'Cult Movies, Dramas, Thrillers']
Value Counts:
listed_in
Dramas, International Movies                                336
Stand-Up Comedy                                             286
Comedies, Dramas, International Movies                      257
Dramas, Independent Movies, International Movies            243
Children & Family Movies, Comedies                          179
                                                           ...
Comedies, Documentaries                                       1
International TV Shows, Romantic TV Shows, TV Mysteries       1
Horror Movies, International Movies, Sci-Fi & Fantasy         1
Reality TV                                                    1
Cult Movies, Dramas, Thrillers                                1
Name: count, Length: 335, dtype: int64
```

[26]:
```
print("-"*25+"description"+"-"*25)
non_graphica_analysis('description')
```

```
-------------------------description-------------------------
Unique Attrubutes: ['On a photo shoot in Ghana, an American model slips back in
time, becomes enslaved on a plantation and bears witness to the agony of her
ancestral past.', "A talented batch of amateur bakers face off in a 10-week
competition, whipping up their best dishes in the hopes of being named the
U.K.'s best.", "A woman adjusting to life after a loss contends with a feisty
bird that's taken over her garden - and a husband who's struggling to find a way
forward.", 'After most of her family is murdered in a terrorist bombing, a young
woman is unknowingly lured into joining the very group that killed them.', 'When
the father of the man she loves insists that his twin sons marry twin sisters, a
woman creates an alter ego that might be a bit too convincing.', 'Mourning the
loss of their beloved junior high basketball coach, five middle-aged pals
reunite at a lake house and rediscover the joys of being a kid.', 'A family's
idyllic suburban life shatters when an alien force invades their home, and as
they struggle to convince others of the deadly threat.', "Blackmailed by his
company's CEO, a low-level employee finds himself forced to spy on the boss's
rival and former mentor.", 'A young Bruce Lee angers kung fu traditionalists by
teaching outsiders, leading to a showdown with a Shaolin master in this film
based on real events.', 'When an insatiable great white shark terrorizes Amity
Island, a police chief, an oceanographer and a grizzled shark hunter seek to
destroy the beast.', 'Four years after the last deadly shark attacks, police
chief Martin Brody fights to protect Amity Island from another killer great
white.', 'After the staff of a marine theme park try to capture a young great
white shark, they discover its mother has invaded the enclosure and is out for
blood.', 'After another deadly shark attack, Ellen Brody has had enough of Amity
Island and moves to the Caribbean - but a great white shark follows her there.',
"Young CIA operative Matt Weston must get a dangerous criminal out of an agency
safe house that's come under attack and get him to a securer location.", "A
rookie cop with one day to prove himself to a veteran LAPD narcotics officer
```

each face wrenching family turmoil, from teen rebellion and marital infidelity to Alzheimer's.", "Laida and Miggy think they've found perfect happiness with each other, until their burgeoning careers force them into a long-distance relationship.", 'An Israeli counterterrorism soldier with a secretly fabulous ambition to become a Manhattan hairstylist will do anything to make his dreams come true.', "When a divorced writer gets a letter from an old boyfriend announcing his wife's had a baby, she returns to her hometown to reconnect with her ex.", 'Aided only by a tough female police officer, a Hong Kong taxi driver tries to escape from gangsters pursuing incriminating evidence left in his cab.', "When a father of eight and a mother of 10 prepare to wed, the couple's children attempt to sabotage the union in this remake of a 1968 comedy.", "Amid the tumult following Egyptian President Morsi's ouster, demonstrators from divergent backgrounds are brought together inside a police truck.", "A philandering small-town mechanic's political ambitions are sparked when the visiting prime minister mistakenly grants him special security clearance.", "A change in the leadership of a political party sparks bitter conflict and the party's division into two rival factions.", "Recovering alcoholic Talal wakes up inside a small-town police station cell, where he's subject to the mind games of a psychotic sadist.", "A political cartoonist, a crime reporter and a pair of cops investigate San Francisco's infamous Zodiac Killer in this thriller based on a true story.", 'Looking to survive in a world taken over by zombies, a dorky college student teams with an urban roughneck and a pair of grifter sisters.', 'Dragged from civilian life, a former superhero must train a new crop of youthful saviors when the military preps for an attack by a familiar villain.', "A scrappy but poor boy worms his way into a tycoon's dysfunctional family, while facing his fear of music and the truth about his past."]
Value Counts:
description
When pretty new neighbor Seema falls for their shy roommate Sid, jealous womanizers Omi and Jai plot to break up the new lovebirds.
2
Mistakenly accused of an attack on the Fourth Raikage, ninja Naruto is imprisoned in the impenetrable Hozuki Castle and his powers are sealed.
2
With their biggest foe seemingly defeated, InuYasha and his friends return to everyday life. But the peace is soon shattered by an emerging new enemy. 2
When Elastigirl gets recruited for a high-profile crime-fighting mission, Mr. Incredible takes on his toughest assignment ever: full-time parenting. 2
After devastating terror attacks in Norway, a young survivor, grieving families and the country rally for justice and healing. Based on a true story. 2
..
When caste differences throw a wrench into their otherwise blossoming relationship, a couple must somehow convince the girl’s father to let them marry. 1
In this action-packed drama based on an extraordinary true story, four brothers protect more than 1,000 Jewish refugees during World War II. 1
A widowed radio DJ and his four spoiled kids need to tighten their belts and pull together when their life of luxury turns lean just before Christmas. 1

```
When a colonel uncovers controversial intel about the government, he makes a
shocking discovery and must decide whether to reveal it or risk his life.     1
A scrappy but poor boy worms his way into a tycoon's dysfunctional family, while
facing his fear of music and the truth about his past.                        1
Name: count, Length: 5321, dtype: int64
```

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

Note: Pre-processing involves unnesting of the data in columns like Actor, Director, Country

4.1 For continuous variable(s): Distplot, countplot, histogram for univariate analysis (10 Points)

4.2 For categorical variable(s): Boxplot (10 Points)

4.3 For correlation: Heatmaps, Pairplots (10 Points)

[27]:
```
# unnesting director column
director_df = df['director'].str.split(', ', expand=True).
 ↪set_index(df['title']).stack().reset_index().rename(columns={0:"Directors"}).
 ↪drop(['level_1'], axis=1)
director_df
```

[27]:

	title	Directors
0	Sankofa	Haile Gerima
1	The Great British Baking Show	Andy Devonshire
2	The Starling	Theodore Melfi
3	Je Suis Karl	Christian Schwochow
4	Jeans	S. Shankar
…	…	…
5955	Zinzana	Majid Al Ansari
5956	Zodiac	David Fincher
5957	Zombieland	Ruben Fleischer
5958	Zoom	Peter Hewitt
5959	Zubaan	Mozez Singh

[5960 rows x 2 columns]

[28]:
```
# unnesting cast_df column
cast_df = df['cast'].str.split(", ", expand=True).set_index(df['title']).
 ↪stack().reset_index().rename(columns={0:'casts'}).drop(['level_1'], axis=1)
cast_df
```

[28]:

	title	casts
0	Sankofa	Kofi Ghanaba
1	Sankofa	Oyafunmike Ogunlano
2	Sankofa	Alexandra Duah
3	Sankofa	Nick Medley
4	Sankofa	Mutabaruka
…	…	…

```
42706   Zubaan       Manish Chaudhary
42707   Zubaan           Meghna Malik
42708   Zubaan          Malkeet Rauni
42709   Zubaan         Anita Shabdish
42710   Zubaan  Chittaranjan Tripathy

[42711 rows x 2 columns]
```

[29]:
```
# unnesting country column
df_country = df['country'].str.split(', ', expand=True).set_index(df['title']).
 ↪stack().reset_index().drop(['level_1'], axis=1).rename(columns={0:"country"})
df_country
```

[29]:
```
           title         country
0        Sankofa   United States
1        Sankofa           Ghana
2        Sankofa    Burkina Faso
3        Sankofa  United Kingdom
4        Sankofa         Germany
…          …              …
6874     Zinzana          Jordan
6875      Zodiac   United States
6876  Zombieland   United States
6877        Zoom   United States
6878      Zubaan           India

[6879 rows x 2 columns]
```

[30]:
```
listed_df = df['listed_in'].str.split(", ", expand=True).set_index(df['title']).
 ↪stack().reset_index().drop(['level_1'], axis=1).rename(columns={0:
 ↪'listed_in'})
listed_df
```

[30]:
```
                              title                 listed_in
0                           Sankofa                    Dramas
1                           Sankofa       Independent Movies
2                           Sankofa     International Movies
3      The Great British Baking Show         British TV Shows
4      The Great British Baking Show                Reality TV
…                                …                       …
11853                          Zoom  Children & Family Movies
11854                          Zoom                  Comedies
11855                        Zubaan                    Dramas
11856                        Zubaan     International Movies
11857                        Zubaan         Music & Musicals
```

```
[11858 rows x 2 columns]
```

[31]:
```
new_df = cast_df.merge(director_df, on=['title'], how='inner').
↪merge(df_country, on=['title'], how='inner').merge(listed_df, on=['title'],␣
↪how='inner')
new_df
```

[31]:
```
          title                  casts     Directors        country  \
0       Sankofa           Kofi Ghanaba  Haile Gerima  United States
1       Sankofa           Kofi Ghanaba  Haile Gerima  United States
2       Sankofa           Kofi Ghanaba  Haile Gerima  United States
3       Sankofa           Kofi Ghanaba  Haile Gerima          Ghana
4       Sankofa           Kofi Ghanaba  Haile Gerima          Ghana
…         …                      …           …             …
143087   Zubaan         Anita Shabdish   Mozez Singh          India
143088   Zubaan         Anita Shabdish   Mozez Singh          India
143089   Zubaan  Chittaranjan Tripathy   Mozez Singh          India
143090   Zubaan  Chittaranjan Tripathy   Mozez Singh          India
143091   Zubaan  Chittaranjan Tripathy   Mozez Singh          India

                   listed_in
0                     Dramas
1          Independent Movies
2       International Movies
3                     Dramas
4          Independent Movies
…                        …
143087  International Movies
143088      Music & Musicals
143089                Dramas
143090  International Movies
143091      Music & Musicals

[143092 rows x 5 columns]
```

[32]:
```
new_df.isnull().sum()
```

[32]:
```
title        0
casts        0
Directors    0
country      0
listed_in    0
dtype: int64
```

[33]:
```
df.drop(['director','cast','country', 'listed_in' ], axis=True, inplace=True)
df.head()
```

[33]:
```
  show_id     type                          title          date_added  \
0      s8    Movie                       Sankofa  September 24, 2021
1      s9  TV Show  The Great British Baking Show  September 24, 2021
2     s10    Movie                    The Starling  September 24, 2021
3     s13    Movie                    Je Suis Karl  September 23, 2021
4     s25    Movie                           Jeans  September 21, 2021

   release_year rating   duration  \
0          1993  TV-MA    125 min
1          2021  TV-14  9 Seasons
2          2021  PG-13    104 min
3          2021  TV-MA    127 min
4          1998  TV-14    166 min

                                         description
0  On a photo shoot in Ghana, an American model s…
1  A talented batch of amateur bakers face off in…
2  A woman adjusting to life after a loss contend…
3  After most of her family is murdered in a terr…
4  When the father of the man she loves insists t…
```

[34]:
```
final_df = new_df.merge(df, on=['title'], how='left')
final_df.head()
```

[34]:
```
     title         casts     Directors        country             listed_in  \
0  Sankofa  Kofi Ghanaba  Haile Gerima  United States                Dramas
1  Sankofa  Kofi Ghanaba  Haile Gerima  United States    Independent Movies
2  Sankofa  Kofi Ghanaba  Haile Gerima  United States  International Movies
3  Sankofa  Kofi Ghanaba  Haile Gerima          Ghana                Dramas
4  Sankofa  Kofi Ghanaba  Haile Gerima          Ghana    Independent Movies

  show_id   type          date_added  release_year rating duration  \
0      s8  Movie  September 24, 2021          1993  TV-MA  125 min
1      s8  Movie  September 24, 2021          1993  TV-MA  125 min
2      s8  Movie  September 24, 2021          1993  TV-MA  125 min
3      s8  Movie  September 24, 2021          1993  TV-MA  125 min
4      s8  Movie  September 24, 2021          1993  TV-MA  125 min

                                         description
0  On a photo shoot in Ghana, an American model s…
1  On a photo shoot in Ghana, an American model s…
2  On a photo shoot in Ghana, an American model s…
3  On a photo shoot in Ghana, an American model s…
4  On a photo shoot in Ghana, an American model s…
```

[35]:
```
final_df.isnull().sum()
```

[35]:
```
title           0
casts           0
Directors       0
country         0
listed_in       0
show_id         0
type            0
date_added      0
release_year    0
rating          0
duration        0
description     0
dtype: int64
```

4.1 For continuous variable(s): Distplot, countplot, histogram for univariate analysis (10 Points)

observation 1: How many movies release each year?

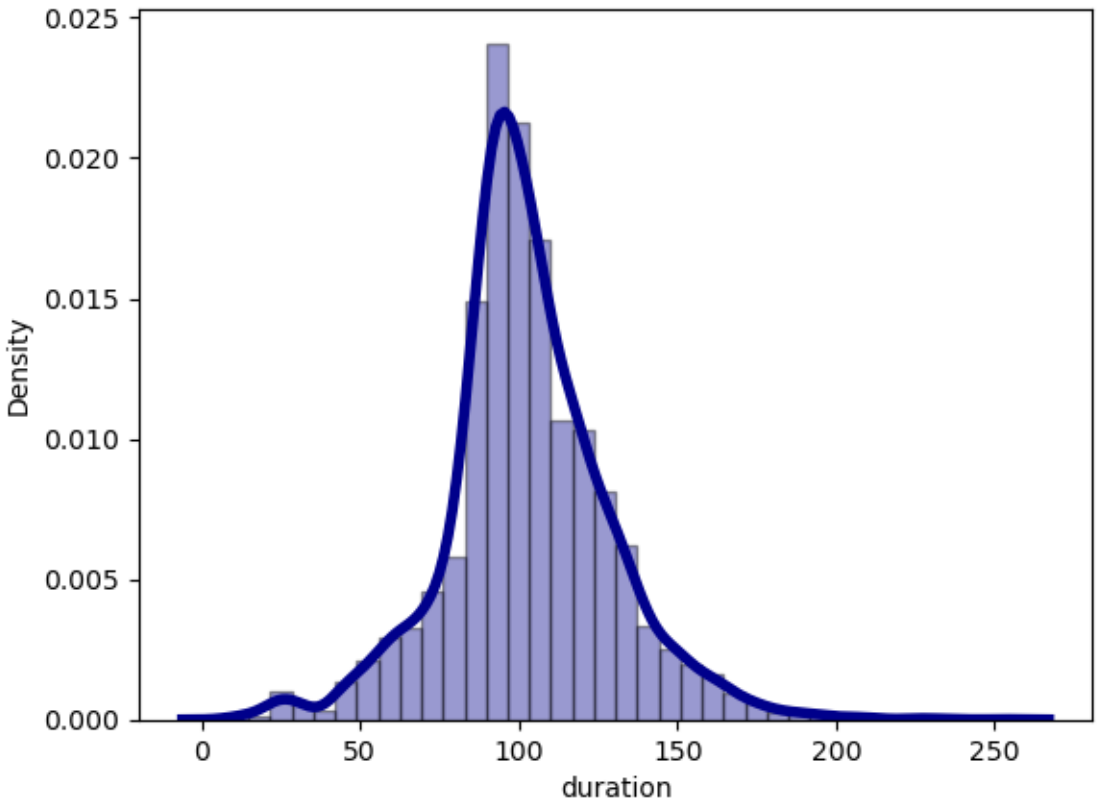
[36]:
```
movie_duration = final_df[final_df['duration'].str.contains(" min")].
 ↪drop_duplicates(subset='title').reset_index(drop=True)
movie_duration['duration'] = movie_duration['duration'].str.replace(" min", "").
 ↪astype(int)
movie_duration.head()
```

[36]:

	title	casts	Directors	country	listed_in	show_id	type	date_added	release_year	rating	duration	description
0	Sankofa	Kofi Ghanaba	Haile Gerima	United States	Dramas	s8	Movie	September 24, 2021	1993	TV-MA	125	On a photo shoot in Ghana, an American model s…
1	The Starling	Melissa McCarthy	Theodore Melfi	United States	Comedies	s10	Movie	September 24, 2021	2021	PG-13	104	A woman adjusting to life after a loss contend…
2	Je Suis Karl	Luna Wedler	Christian Schwochow	Germany	Dramas	s13	Movie	September 23, 2021	2021	TV-MA	127	After most of her family is murdered in a terr…
3	Jeans	Prashanth	S. Shankar	India	Comedies	s25	Movie	September 21, 2021	1998	TV-14	166	When the father of the man she loves insists t…
4	Grown Ups	Adam Sandler	Dennis Dugan	United States	Comedies	s28	Movie	September 20, 2021	2010	PG-13	103	Mourning the loss of their beloved junior high…

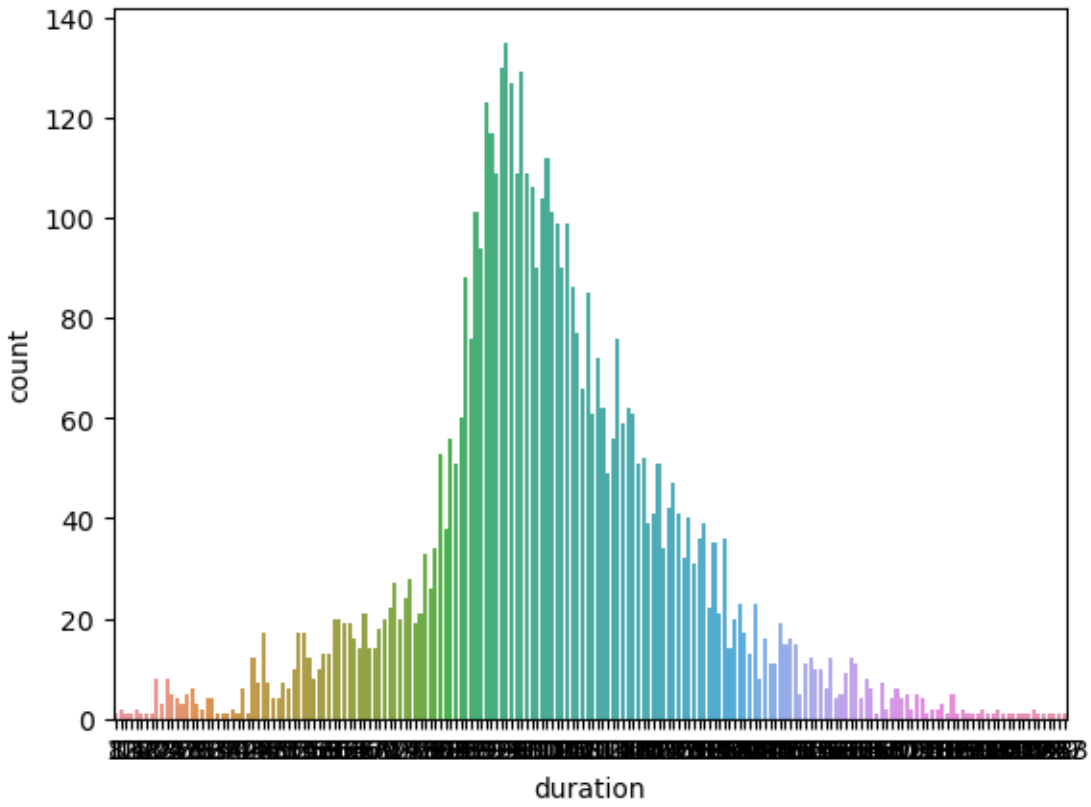
[37]:
```
#displot
import seaborn as sns
import matplotlib.pyplot as plt
sns.distplot(movie_duration['duration'], hist=True, kde=True,
bins=int(36), color = 'darkblue',
hist_kws={'edgecolor':'black'},
kde_kws={'linewidth': 4})
plt.show();
```

```
C:\Users\visha\AppData\Local\Temp\ipykernel_27788\1686422039.py:4: UserWarning:

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

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

For a guide to updating your code to use the new functions, please see
https://gist.github.com/mwaskom/de44147ed2974457ad6372750bbe5751

  sns.distplot(movie_duration['duration'], hist=True, kde=True,
```

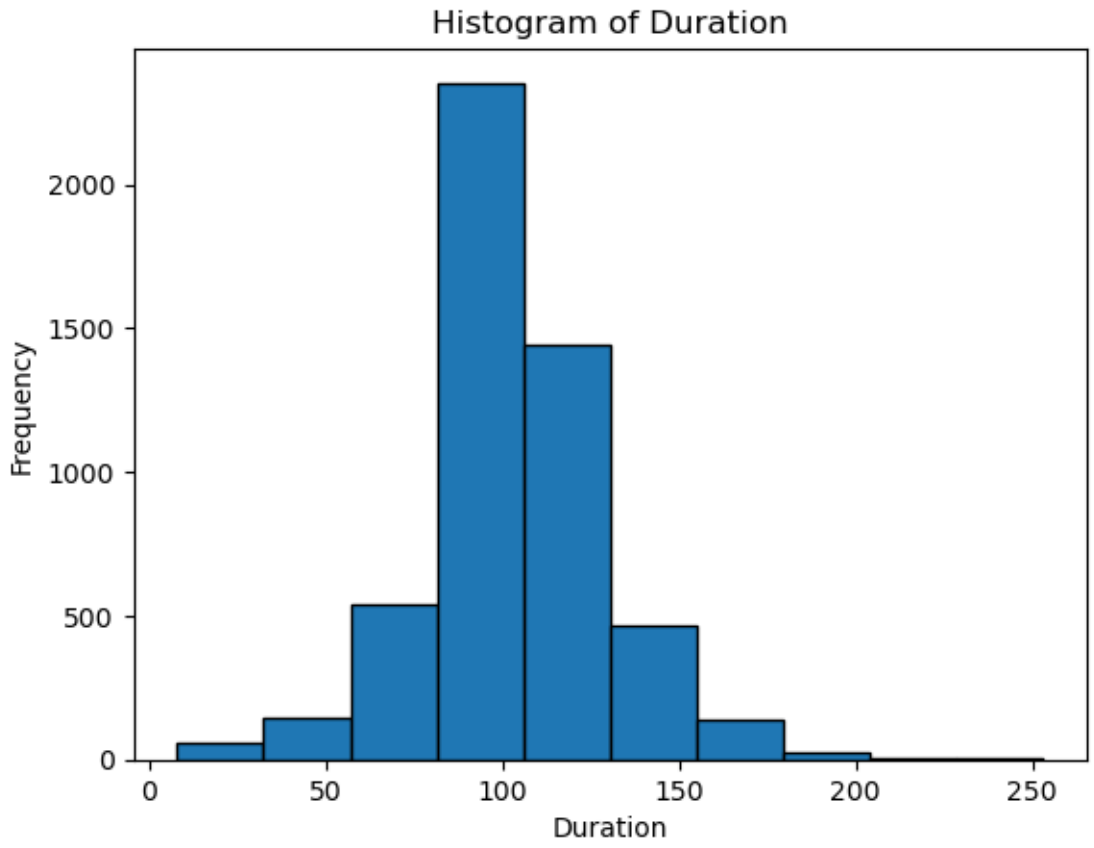


[38]:
```
sns.countplot(data=movie_duration, x='duration')
```

[38]:
```
<Axes: xlabel='duration', ylabel='count'>
```



[39]:
```
# Create histogram of the 'duration' column
plt.hist(movie_duration['duration'], bins=10, edgecolor='black')

# Add title and labels
plt.title('Histogram of Duration')
plt.xlabel('Duration')
plt.ylabel('Frequency')

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



[40]:
```
show_duration = final_df[final_df['duration'].str.lower().str.
 ↪contains("season")].drop_duplicates(subset='title').reset_index(drop=True)
show_duration['duration'] = show_duration['duration'].str.extract(r'(\d+)').
 ↪astype(int)
show_duration.head()
```

[40]:

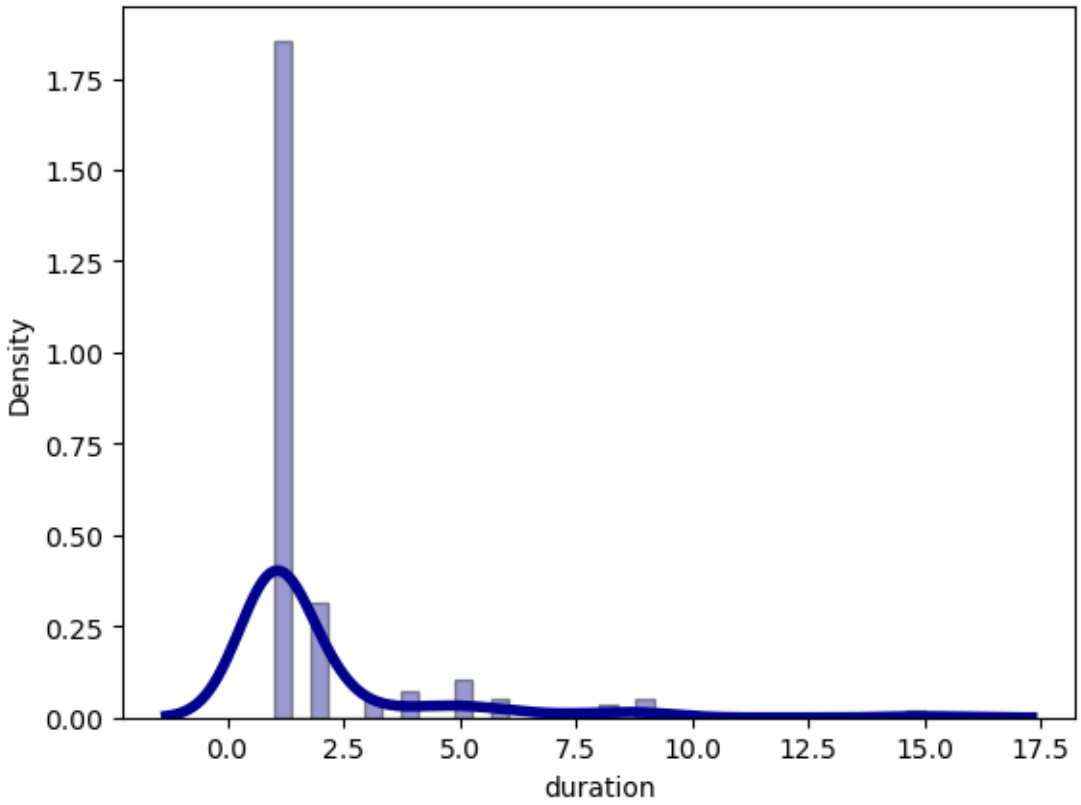
	title	casts	\
0	The Great British Baking Show	Mel Giedroyc	
1	King of Boys: The Return of the King	Sola Sobowale	
2	Gone for Good	Finnegan Oldfield	
3	Valeria	Diana Gómez	
4	Navarasa	Suriya	

	Directors	country	listed_in	show_id	\
0	Andy Devonshire	United Kingdom	British TV Shows	s9	
1	Kemi Adetiba	Nigeria	Crime TV Shows	s199	
2	Juan Carlos Medina	France	Crime TV Shows	s274	
3	Inma Torrente	Spain	International TV Shows	s277	
4	Bejoy Nambiar	India	TV Shows	s298	

```
      type          date_added  release_year rating  duration  \
0  TV Show  September 24, 2021          2021  TV-14         9
1  TV Show     August 27, 2021          2021  TV-MA         1
2  TV Show     August 13, 2021          2021  TV-MA         1
3  TV Show     August 13, 2021          2021  TV-MA         2
4  TV Show      August 6, 2021          2021  TV-MA         1

                                         description
0  A talented batch of amateur bakers face off in…
1  Alhaja Eniola Salami starts anew and sets her …
2  Ten years after losing two loved ones, a man i…
3  A writer in creative and marital crises finds …
4  From amusement to awe, the nine human emotions…
```

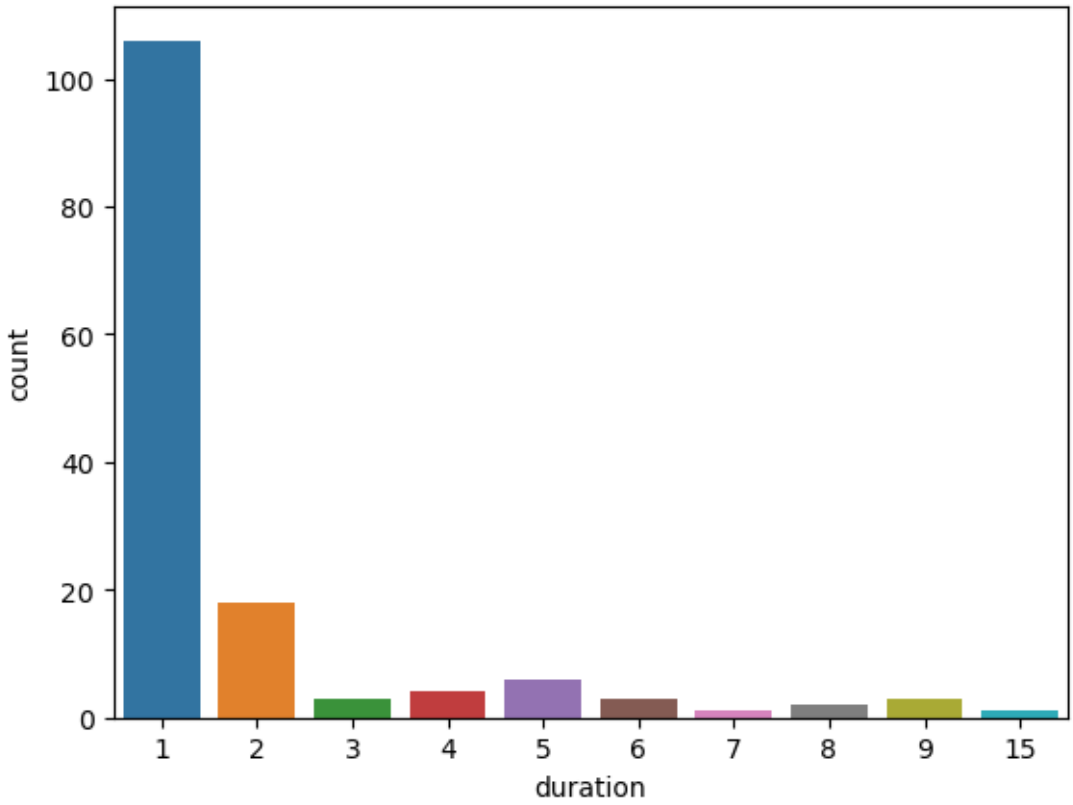
[41]:
```
sns.distplot(show_duration['duration'], hist=True, kde=True,
bins=int(36), color = 'darkblue',
hist_kws={'edgecolor':'black'},
kde_kws={'linewidth': 4})
plt.show();
```

```
C:\Users\visha\AppData\Local\Temp\ipykernel_27788\2253031695.py:1: UserWarning:

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

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

For a guide to updating your code to use the new functions, please see
https://gist.github.com/mwaskom/de44147ed2974457ad6372750bbe5751

  sns.distplot(show_duration['duration'], hist=True, kde=True,
```

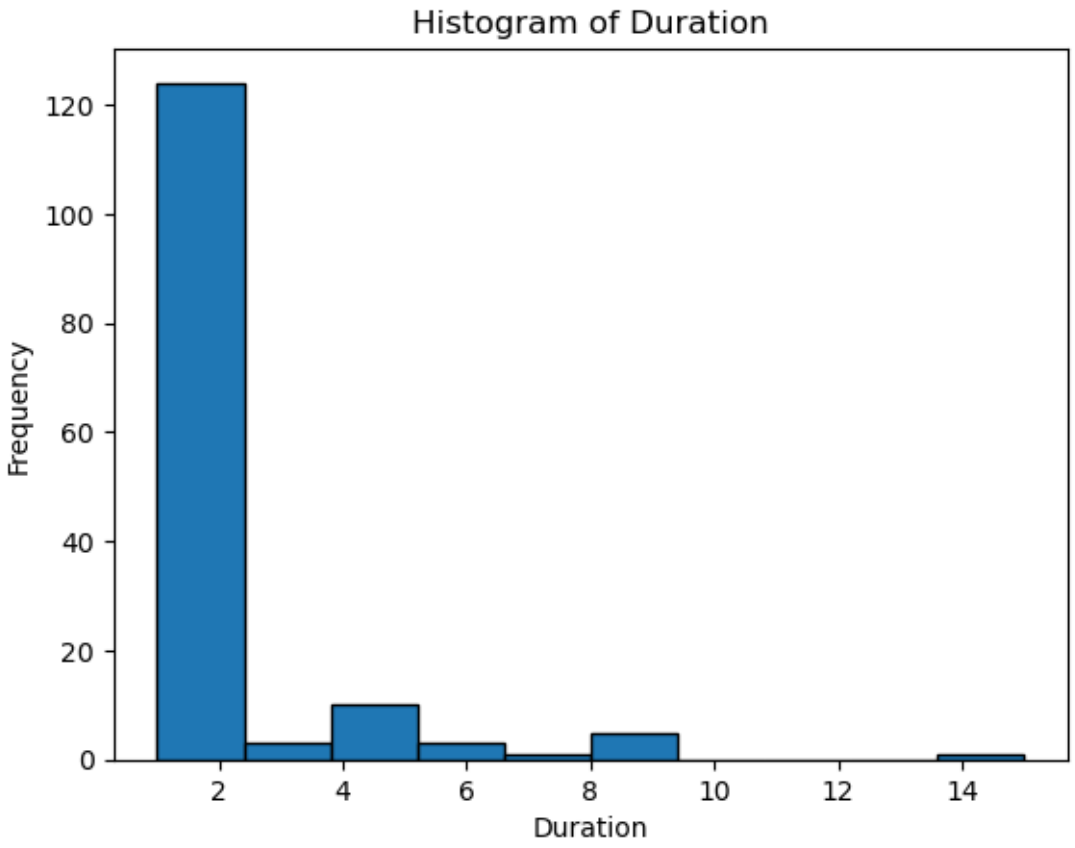


[42]:
```
sns.countplot(data=show_duration, x='duration')
```

[42]:
```
<Axes: xlabel='duration', ylabel='count'>
```

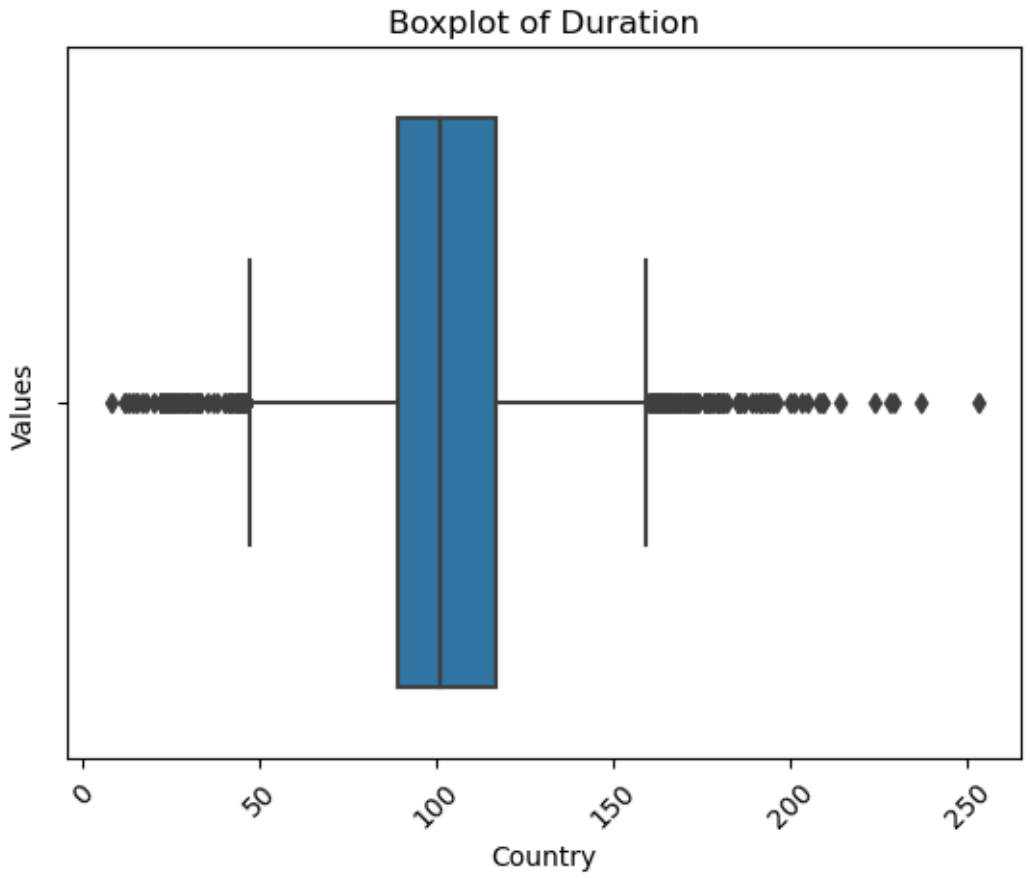


[43]:
```
# Create histogram of the 'duration' column
plt.hist(show_duration['duration'], bins=10, edgecolor='black')

# Add title and labels
plt.title('Histogram of Duration')
plt.xlabel('Duration')
plt.ylabel('Frequency')

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

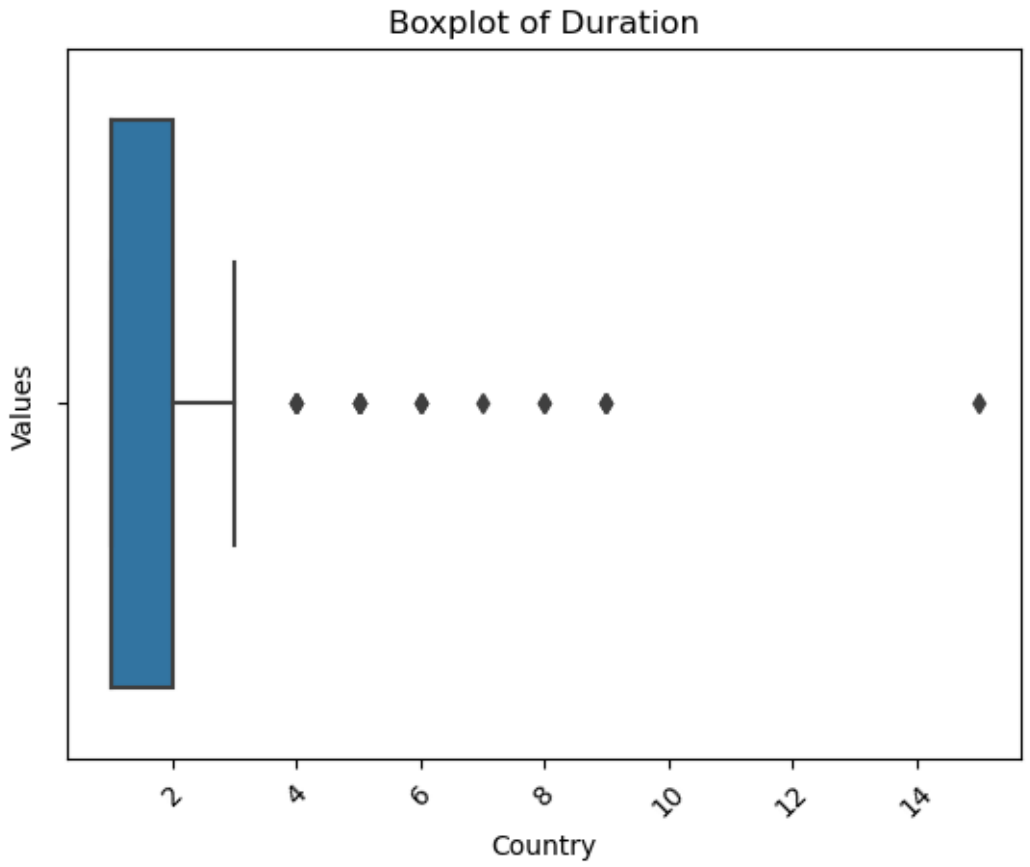


4.2 For categorical variable(s): Boxplot (10 Points)

[44]:
```
# Create boxplot for the 'country' column
sns.boxplot(data=movie_duration, x='duration')

# Add title and labels
plt.title('Boxplot of Duration')
plt.xlabel('Country')
plt.ylabel('Values')

# Rotate x-axis labels for better visibility
plt.xticks(rotation=45)

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



[45]:
```
# Create boxplot for the 'country' column
sns.boxplot(data=show_duration, x='duration')

# Add title and labels
plt.title('Boxplot of Duration')
plt.xlabel('Country')
plt.ylabel('Values')


# Rotate x-axis labels for better visibility
plt.xticks(rotation=45)

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



[46]:
```
final_df.sample(10)
```

[46]:

	title	casts	Directors	country	listed_in	show_id
62431	Bad Seeds	Hakou Benosmane	Kheiron	France	Comedies	s4268
141604	While We're Young	Charles Grodin	Noah Baumbach	United States	Dramas	s8732
76449	The Lion Woman	Henrik Mestad	Vibeke Idsøe	Germany	Dramas	s5249
74456	Rip Tide	Valerie Bader	Rhiannon Bannenberg	Australia	Sports Movies	s5082
30540	Afonso Padilha: Classless	Afonso Padilha				
8294	Mary Magdalene	Ryan Corr				
43930	Hugo	Edmund Kingsley				
92095	Bean: The Ultimate Disaster Movie	Peter James				
18822	Our Idiot Brother	Shirley Knight				
81981	Coin Heist	Sasha Pieterse				

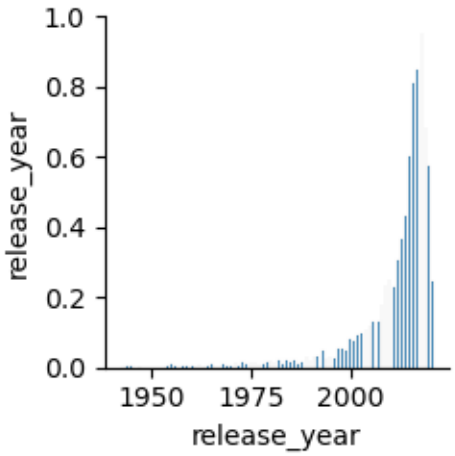
```
30540        Junior Carelli          Brazil           Stand-Up Comedy   s2051
8294            Garth Davis       Australia                   Dramas    s577
43930       Martin Scorsese  United Kingdom  Children & Family Movies   s2861
92095             Mel Smith   United States                 Comedies   s6261
18822          Jesse Peretz   United States                 Comedies   s1265
81981          Emily Hagins   United States                   Dramas   s5643

         type         date_added  release_year rating duration  \
62431   Movie  December 21, 2018          2018  TV-MA  105 min
141604  Movie   October 23, 2017          2015      R   97 min
76449   Movie    October 1, 2017          2016  TV-14  118 min
74456   Movie   January 15, 2018          2017  TV-PG   87 min
30540   Movie  September 3, 2020          2020  TV-MA   63 min
8294    Movie       July 1, 2021          2019      R  120 min
43930   Movie      March 1, 2020          2011     PG  127 min
92095   Movie     August 1, 2018          1997  PG-13   89 min
18822   Movie  February 26, 2021          2011      R   90 min
81981   Movie    January 6, 2017          2017  TV-14   98 min

                                                 description
62431   Troubled by his past, a scam artist who runs a…
141604  A documentary filmmaker struggling with a crea…
76449   A girl born with a condition that causes abnor…
74456   Following an embarrassing viral video, a New Y…
30540   Brazilian comedian Afonso Padilha dives into h…
8294    A woman with healing powers abandons her villa…
43930   Living inside a Paris train station, an orphan…
92095   Disaster-prone Mr. Bean is sent to Los Angeles…
18822   A seemingly clueless idealist relies on his ex…
81981   When a crisis threatens to destroy their high …
```

4.3 For correlation: Heatmaps, Pairplots (10 Points)

[47]:
```
sns.pairplot(final_df)
```

```
C:\Users\visha\anaconda3\Lib\site-packages\seaborn\axisgrid.py:118: UserWarning:
The figure layout has changed to tight
  self._figure.tight_layout(*args, **kwargs)
```

[47]:
```
<seaborn.axisgrid.PairGrid at 0x173483df210>
```



[48]:
```
# Calculate correlation matrix
temp = show_duration[['release_year', 'duration']]
temp.corr()
```

[48]:

	release_year	duration
release_year	1.000000	0.033172
duration	0.033172	1.000000

[49]:
```
final_df
```

[49]:

	title	casts	Directors	country	\
0	Sankofa	Kofi Ghanaba	Haile Gerima	United States	
1	Sankofa	Kofi Ghanaba	Haile Gerima	United States	
2	Sankofa	Kofi Ghanaba	Haile Gerima	United States	
3	Sankofa	Kofi Ghanaba	Haile Gerima	Ghana	
4	Sankofa	Kofi Ghanaba	Haile Gerima	Ghana	
…	…	…	…	…	
143087	Zubaan	Anita Shabdish	Mozez Singh	India	
143088	Zubaan	Anita Shabdish	Mozez Singh	India	
143089	Zubaan	Chittaranjan Tripathy	Mozez Singh	India	
143090	Zubaan	Chittaranjan Tripathy	Mozez Singh	India	
143091	Zubaan	Chittaranjan Tripathy	Mozez Singh	India	

	listed_in	show_id	type	date_added	release_year	\
0	Dramas	s8	Movie	September 24, 2021	1993	
1	Independent Movies	s8	Movie	September 24, 2021	1993	
2	International Movies	s8	Movie	September 24, 2021	1993	
3	Dramas	s8	Movie	September 24, 2021	1993	
4	Independent Movies	s8	Movie	September 24, 2021	1993	
…	…	…	…	…	…	
143087	International Movies	s8807	Movie	March 2, 2019	2015	

```
143088      Music & Musicals   s8807  Movie       March 2, 2019          2015
143089                Dramas   s8807  Movie       March 2, 2019          2015
143090  International Movies   s8807  Movie       March 2, 2019          2015
143091      Music & Musicals   s8807  Movie       March 2, 2019          2015

       rating duration                                          description
0       TV-MA  125 min  On a photo shoot in Ghana, an American model s…
1       TV-MA  125 min  On a photo shoot in Ghana, an American model s…
2       TV-MA  125 min  On a photo shoot in Ghana, an American model s…
3       TV-MA  125 min  On a photo shoot in Ghana, an American model s…
4       TV-MA  125 min  On a photo shoot in Ghana, an American model s…
…         …       …                                                 …
143087  TV-14  111 min  A scrappy but poor boy worms his way into a ty…
143088  TV-14  111 min  A scrappy but poor boy worms his way into a ty…
143089  TV-14  111 min  A scrappy but poor boy worms his way into a ty…
143090  TV-14  111 min  A scrappy but poor boy worms his way into a ty…
143091  TV-14  111 min  A scrappy but poor boy worms his way into a ty…

[143092 rows x 12 columns]
```

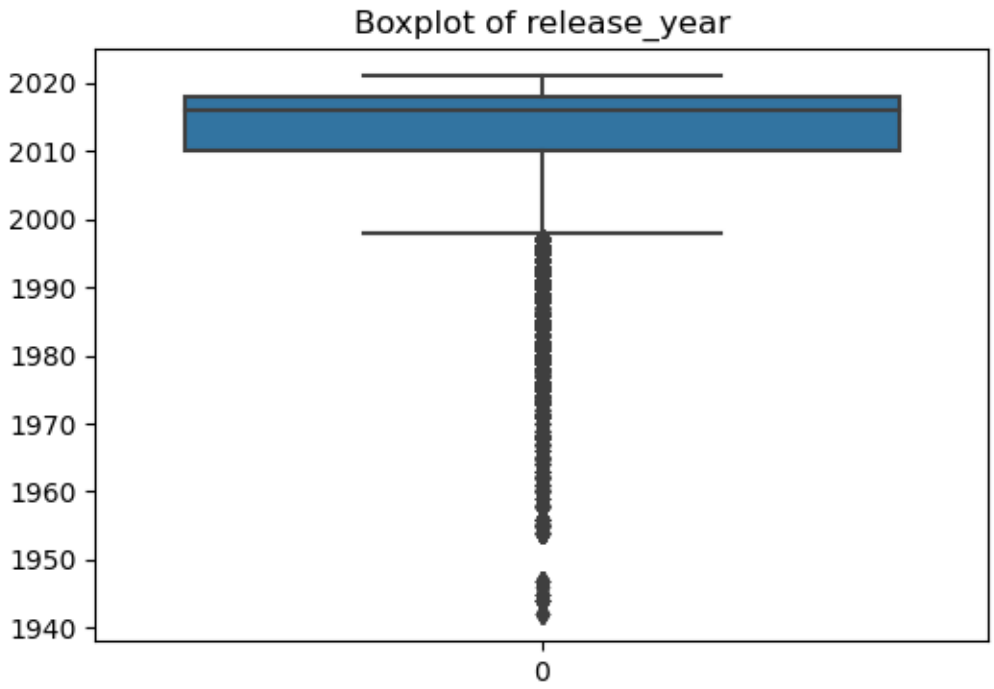
0.0.9 5. Missing Value & Outlier check (Treatment optional)

[50]:
```
# Missing Value Check
missing_values = final_df.isnull().sum()
print("Missing Values:\n", missing_values)
```

```
Missing Values:
 title           0
casts           0
Directors       0
country         0
listed_in       0
show_id         0
type            0
date_added      0
release_year    0
rating          0
duration        0
description     0
dtype: int64
```

[51]:
```
# Outlier Check
numerical_columns = final_df.select_dtypes(include=['int64', 'float64']).columns
for col in numerical_columns:
    plt.figure(figsize=(6, 4))
    sns.boxplot(data=final_df[col])
    plt.title(f'Boxplot of {col}')
```

```
    plt.show()
```



0.0.10 6. Insights based on Non-Graphical and Visual Analysis (10 Points)

6.1 Comments on the range of attributes

6.2 Comments on the distribution of the variables and relationship between them

6.3 Comments for each univariate and bivariate plot

0.0.11 7. Business Insights (10 Points)

- Should include patterns observed in the data along with what you can infer from it

0.0.12 8. Recommendations (10 Points)

- Actionable items for business. No technical jargon. No complications. Simple action items that everyone can understand

[52]:
```
final_df.rename(columns={'listed_in':'Genre'}, inplace=True)
final_df
```

[52]:

	title	casts	Directors	country	\
0	Sankofa	Kofi Ghanaba	Haile Gerima	United States	
1	Sankofa	Kofi Ghanaba	Haile Gerima	United States	
2	Sankofa	Kofi Ghanaba	Haile Gerima	United States	

```
3       Sankofa           Kofi Ghanaba  Haile Gerima          Ghana
4       Sankofa           Kofi Ghanaba  Haile Gerima          Ghana
…         …                     …           …            …
143087   Zubaan         Anita Shabdish   Mozez Singh          India
143088   Zubaan         Anita Shabdish   Mozez Singh          India
143089   Zubaan  Chittaranjan Tripathy   Mozez Singh          India
143090   Zubaan  Chittaranjan Tripathy   Mozez Singh          India
143091   Zubaan  Chittaranjan Tripathy   Mozez Singh          India

                     Genre show_id   type          date_added  release_year  \
0                   Dramas      s8  Movie  September 24, 2021          1993
1         Independent Movies      s8  Movie  September 24, 2021          1993
2       International Movies      s8  Movie  September 24, 2021          1993
3                   Dramas      s8  Movie  September 24, 2021          1993
4         Independent Movies      s8  Movie  September 24, 2021          1993
…                       …     …    …                   …           …
143087  International Movies   s8807  Movie       March 2, 2019          2015
143088     Music & Musicals   s8807  Movie       March 2, 2019          2015
143089              Dramas   s8807  Movie       March 2, 2019          2015
143090  International Movies   s8807  Movie       March 2, 2019          2015
143091     Music & Musicals   s8807  Movie       March 2, 2019          2015

       rating duration                                          description
0       TV-MA  125 min  On a photo shoot in Ghana, an American model s…
1       TV-MA  125 min  On a photo shoot in Ghana, an American model s…
2       TV-MA  125 min  On a photo shoot in Ghana, an American model s…
3       TV-MA  125 min  On a photo shoot in Ghana, an American model s…
4       TV-MA  125 min  On a photo shoot in Ghana, an American model s…
…         …       …                                                 …
143087  TV-14  111 min  A scrappy but poor boy worms his way into a ty…
143088  TV-14  111 min  A scrappy but poor boy worms his way into a ty…
143089  TV-14  111 min  A scrappy but poor boy worms his way into a ty…
143090  TV-14  111 min  A scrappy but poor boy worms his way into a ty…
143091  TV-14  111 min  A scrappy but poor boy worms his way into a ty…

[143092 rows x 12 columns]
```

[53]:
```
final_df.groupby(by='Genre').agg({'title':'nunique'}).sort_values(by='title',␣
↪ascending=False)
```

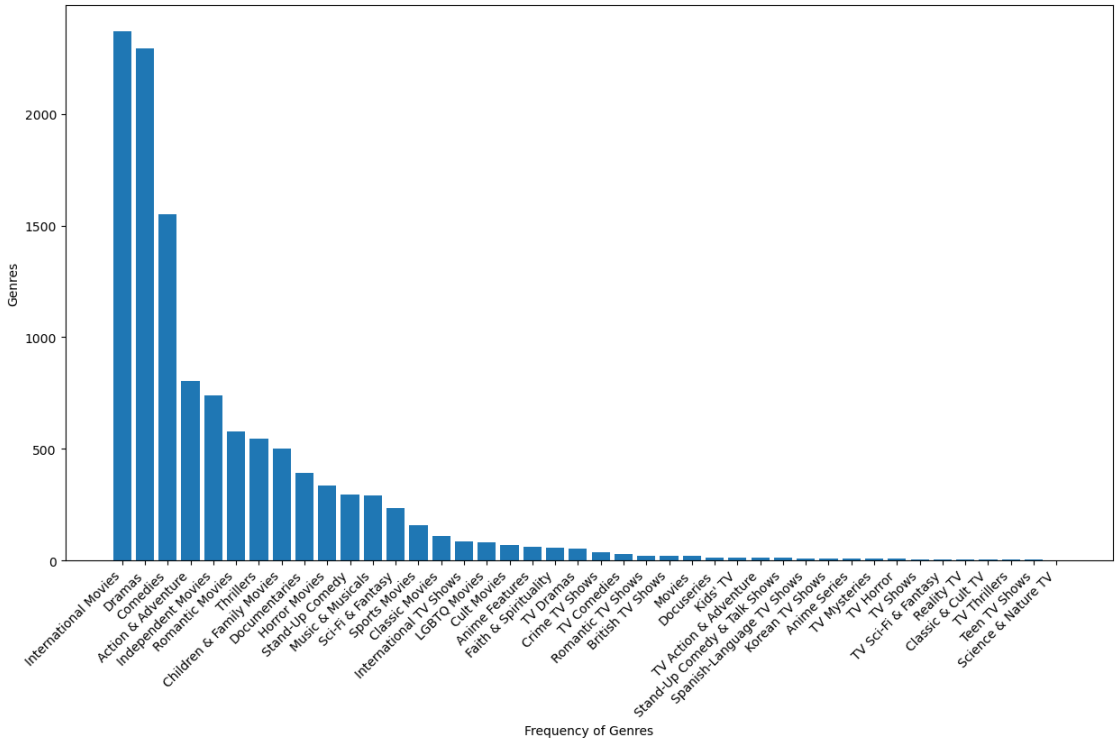
[53]:
```
                           title
Genre
International Movies        2369
Dramas                      2293
Comedies                    1553
Action & Adventure           806
Independent Movies           740
```

```
Romantic Movies                 579
Thrillers                       547
Children & Family Movies        503
Documentaries                   391
Horror Movies                   336
Stand-Up Comedy                 294
Music & Musicals                292
Sci-Fi & Fantasy                236
Sports Movies                   156
Classic Movies                  108
International TV Shows           87
LGBTQ Movies                     80
Cult Movies                      69
Anime Features                   61
Faith & Spirituality             58
TV Dramas                        52
Crime TV Shows                   38
TV Comedies                      30
Romantic TV Shows                21
British TV Shows                 21
Movies                           20
Docuseries                       14
Kids' TV                         13
TV Action & Adventure            13
Stand-Up Comedy & Talk Shows     11
Spanish-Language TV Shows        10
Korean TV Shows                  10
Anime Series                     10
TV Mysteries                      8
TV Horror                         7
TV Shows                          5
TV Sci-Fi & Fantasy               4
Reality TV                        3
Classic & Cult TV                 3
TV Thrillers                      3
Teen TV Shows                     3
Science & Nature TV               1
```

Observation: Director Mostly make `interntional`, `Dramas`, and `Comedies` Genre movie.‘

[54]:
```
total_genre = final_df.groupby(by='Genre').agg({'title':'nunique'}).
 ↪sort_values(by='title', ascending=False)
plt.figure(figsize=(15,8))
plt.bar(total_genre.index, total_genre['title'])
plt.xticks(rotation=45, ha='right')
plt.xlabel('Frequency of Genres')
plt.ylabel('Genres')
```

```
plt.show()
```



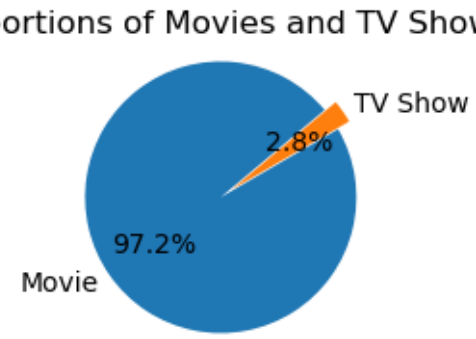
[55]:
```
#Total Number of Movie and Tv Show
temp = final_df.groupby(by='type').agg({'title':'nunique'})
temp
```

[55]:

type	title
Movie	5185
TV Show	147

Observation: Movie and TV show and 97:3 ration

[56]:
```
# Plotting
plt.figure(figsize=(2,2))
plt.pie(temp['title'], explode=(0.05,0.05), labels=temp.index, autopct='%1.
↪1f%%', startangle=40)
plt.title('Proportions of Movies and TV Shows')
plt.axis('equal')
plt.show()
```



[57]:
```
#number of movie release on a country
temp = final_df.groupby(by='country').agg({'title':'nunique'}).
 ↪sort_values(by='title', ascending=False)
temp
```

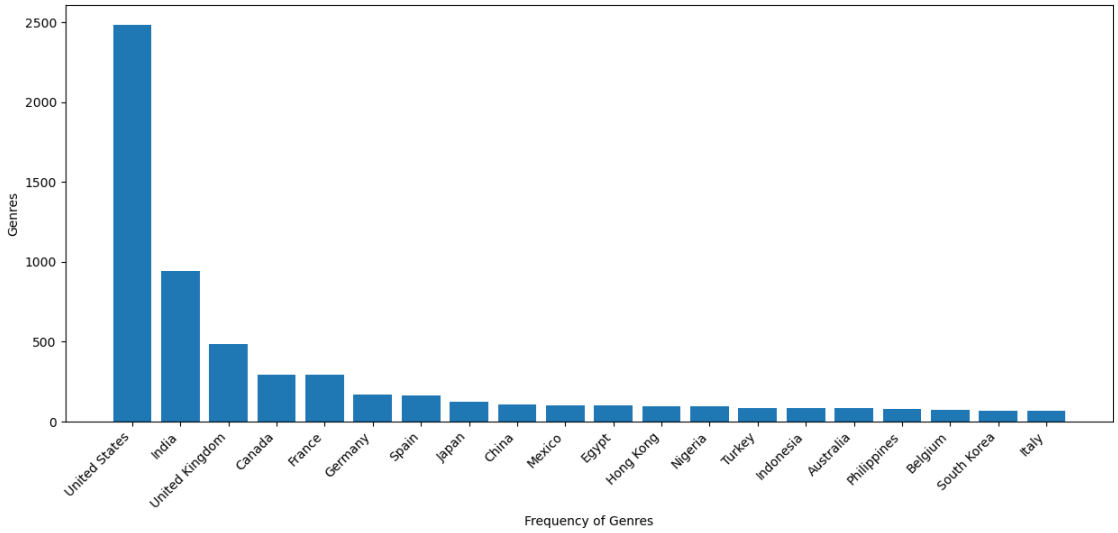
[57]:

country	title
United States	2485
India	940
United Kingdom	484
Canada	295
France	293
…	…
Mozambique	1
Namibia	1
Nicaragua	1
Panama	1
Zimbabwe	1

[113 rows x 1 columns]

Observation: United States, India and United Kingdom are the top 3 country where most of the movie is release US,India,UK,Canada and France are leading countries in Content Creation on Netflix

[58]:
```
plt.figure(figsize=(15,6))
temp = temp.head(20)
plt.bar(temp.index, temp['title'])
plt.xticks(rotation=45, ha='right')
plt.xlabel('Frequency of Genres')
plt.ylabel('Genres')
plt.show()
```

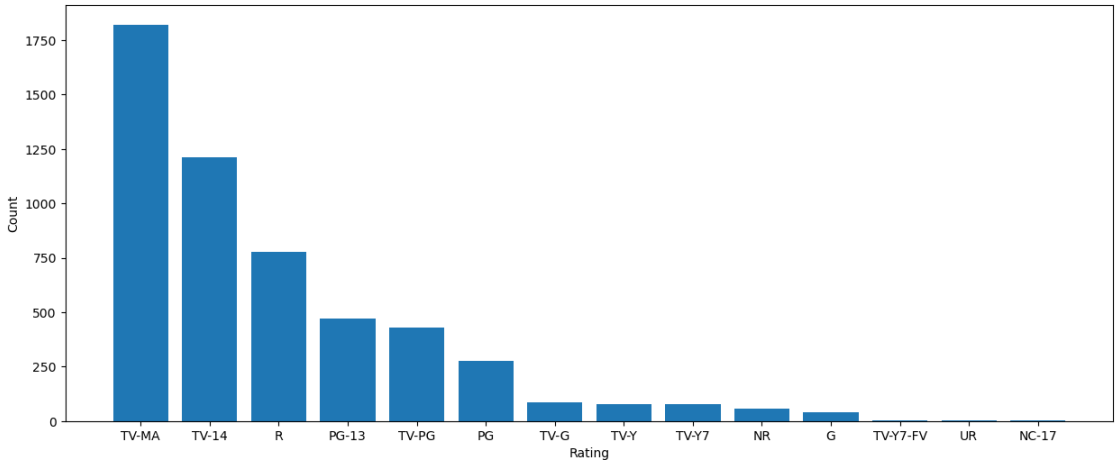


[59]:
```
# Different Rating Received by different show
temp = final_df.groupby(by='rating').agg({'title':'nunique'}).
 ↪sort_values(by='title', ascending=False)
temp
```

[59]:

rating	title
TV-MA	1822
TV-14	1214
R	778
PG-13	470
TV-PG	431
PG	275
TV-G	84
TV-Y	76
TV-Y7	76
NR	58
G	40
TV-Y7-FV	3
UR	3
NC-17	2

[60]:
```
plt.figure(figsize=(15,6))
plt.bar(temp.index, temp['title'])
plt.xlabel("Rating")
plt.ylabel("Count")
plt.show()
```



[61]:
```
final_df['casts']=final_df.casts.str.strip()
```

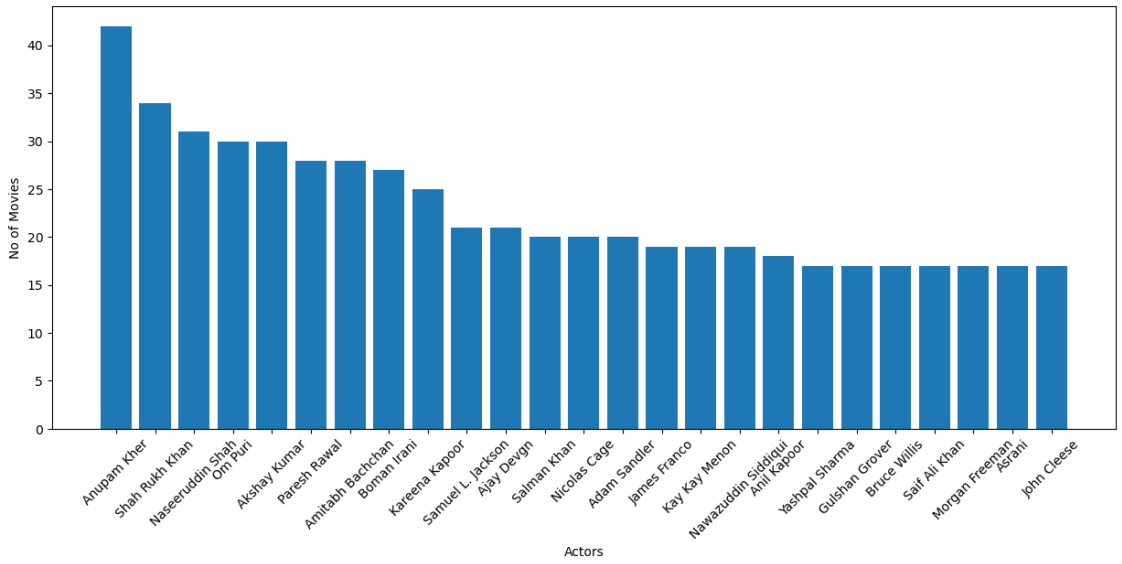
[62]:
```
#number of movie done by each cast
temp=final_df.groupby(by='casts').agg({'title':'nunique'}).
 ↪sort_values(by='title', ascending=False)
temp
```

[62]:
```
                  title
casts
Anupam Kher          42
Shah Rukh Khan       34
Naseeruddin Shah     31
Om Puri              30
Akshay Kumar         30
…                    …
Jacob Gregory         1
Jacob Ewaniuk         1
Jacob Davich          1
Jacob Craner          1
Ṣọpẹ́ Dìrísù          1

[25458 rows x 1 columns]
```

Observation: Anupam Kher, Shah Rukh Khan, Naseeruddin Shah, Om Puri, Akshay Kumar are the top 5 actor who did most of the movies in netflix platform

[63]:
```
temp1 = temp.head(25)
plt.figure(figsize=(15,6))
plt.bar(x=temp1.index, height=temp1.title)
plt.xlabel("Actors")
plt.ylabel('No of Movies')
```

```
plt.xticks(rotation=45)
plt.show()
```



[64]:
```
#number of different movie directed by differnt director
temp = final_df.groupby(by='Directors').agg({'title':'nunique'}).
 ↪sort_values(by='title', ascending=False)
temp
```

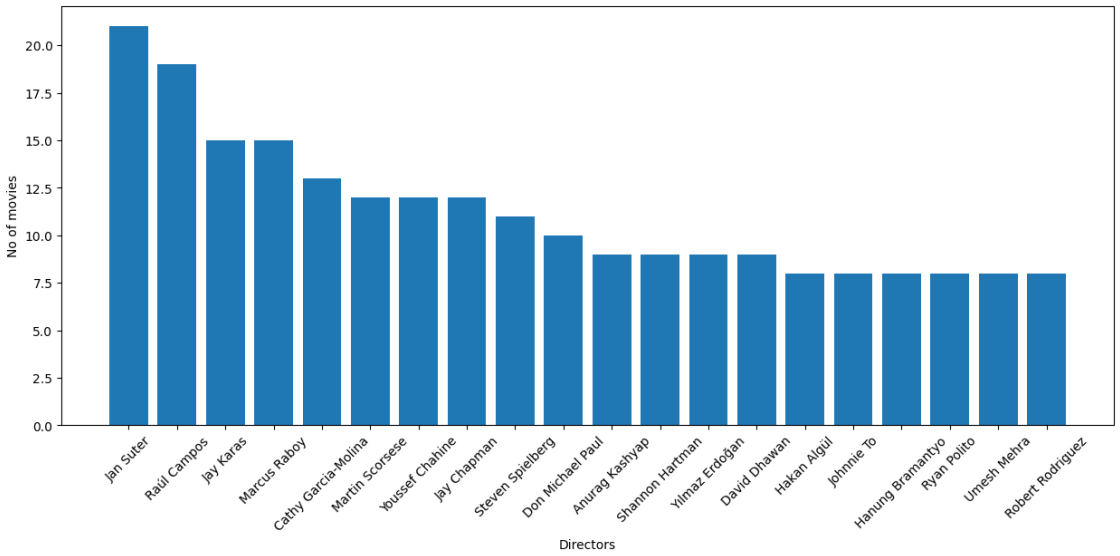
[64]:

Directors	title
Jan Suter	21
Raúl Campos	19
Jay Karas	15
Marcus Raboy	15
Cathy Garcia-Molina	13
…	…
Joshua Z. Weinstein	1
Josiah Ng	1
José Eduardo Belmonte	1
José Luis Gutiérrez	1
Kevin Wotton	1

[4311 rows x 1 columns]

Observation: Top 3 Director to produce movie on netflix. 1. Jan Suter 21 1. Raúl Campos 19 1. Jay Karas 15

[65]:
```
temp_df = temp.head(20)
plt.figure(figsize=(15,6))
plt.bar(x=temp_df.index, height=temp_df.title)
plt.xlabel("Directors")
plt.ylabel("No of movies")
plt.xticks(rotation = 45)
plt.show()
```



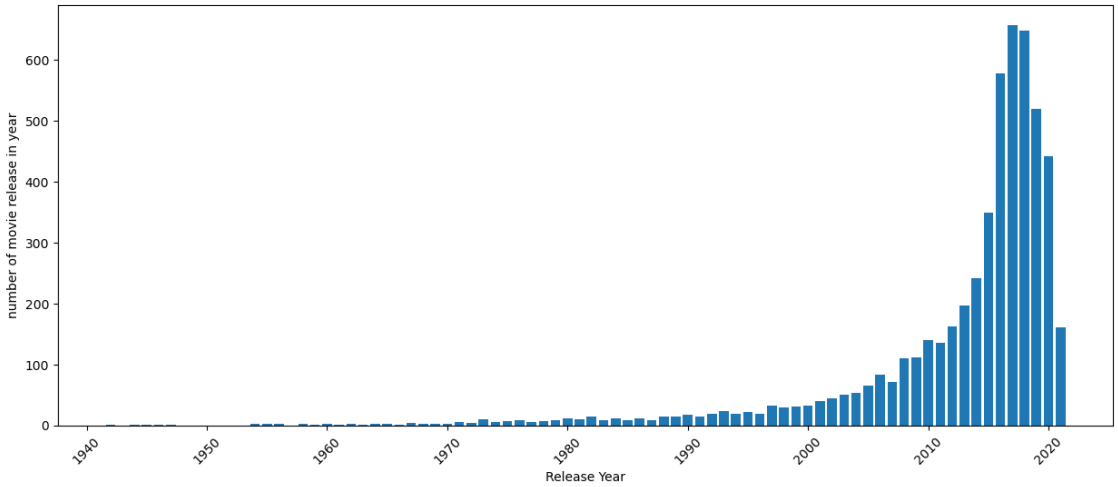
[66]:
```
#which year was best for netflix in term of movie or tv show added to there platform
temp = final_df.groupby(by=['release_year']).agg({'title':'nunique'}).sort_values(by='title', ascending=False).reset_index()
temp
```

[66]:

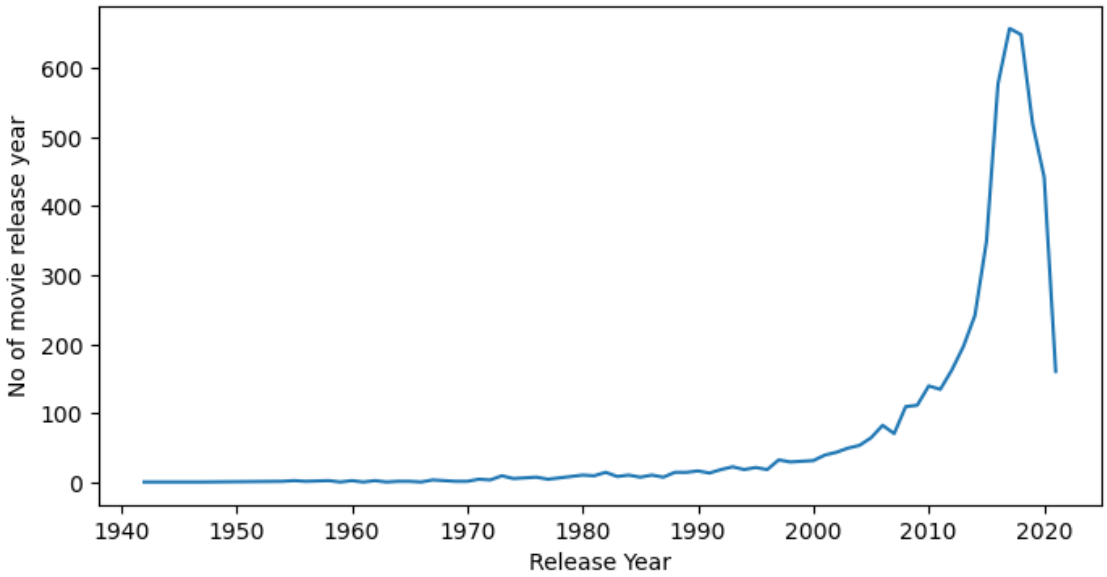
	release_year	title
0	2017	657
1	2018	648
2	2016	577
3	2019	519
4	2020	442
..	…	…
67	1959	1
68	1961	1
69	1963	1
70	1966	1
71	1942	1

[72 rows x 2 columns]

[67]:
```
plt.figure(figsize=(15,6))
plt.bar(x=temp.release_year, height=temp.title)
plt.xlabel("Release Year")
plt.ylabel("number of movie release in year")
plt.xticks(rotation=45)
plt.show()
```



[68]:
```
plt.figure(figsize=(8, 4))
sns.lineplot(data=temp, x='release_year', y='title')
plt.xlabel("Release Year")
plt.ylabel("No of movie release year")
plt.show()
```



[69]:
```
# Remove leading whitespace from 'date_added' column
final_df['date_added'] = final_df['date_added'].str.strip()

# Convert 'date_added' column to datetime format
final_df['date_added'] = pd.to_datetime(final_df['date_added'], errors='coerce')
```

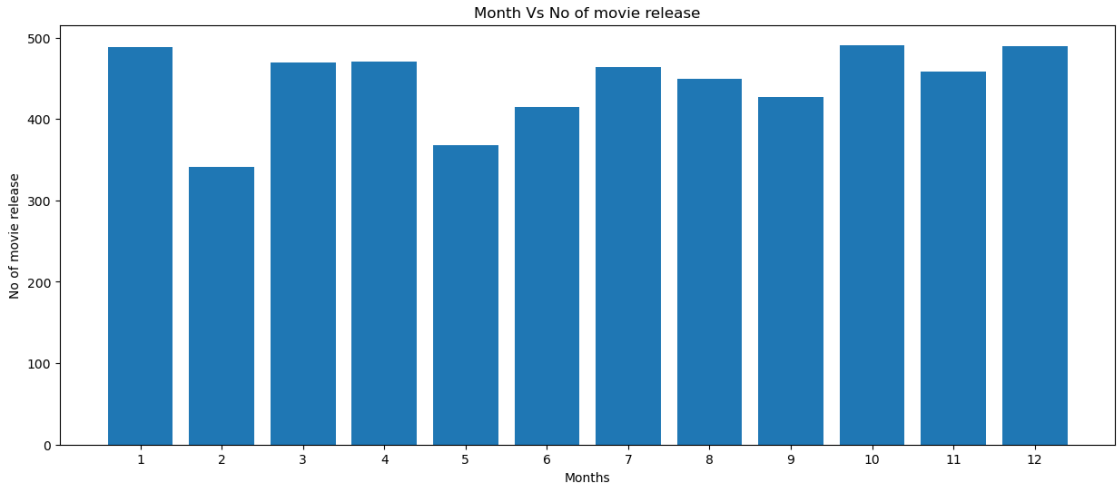
[70]:
```
final_df['month'] = final_df['date_added'].dt.month
final_df['day'] = final_df['date_added'].dt.day_name()
```

[71]:
```
temp = final_df.groupby(by='month').agg({'title':'nunique'}).reset_index()
temp
```

[71]:

	month	title
0	1	489
1	2	341
2	3	469
3	4	471
4	5	368
5	6	415
6	7	464
7	8	449
8	9	427
9	10	491
10	11	458
11	12	490

[72]:
```
plt.figure(figsize=(15,6))
plt.bar(x=temp.month, height=temp.title)
plt.title("Month Vs No of movie release")
plt.xlabel("Months")
plt.ylabel("No of movie release")
plt.xticks(temp.month)
plt.show()
```



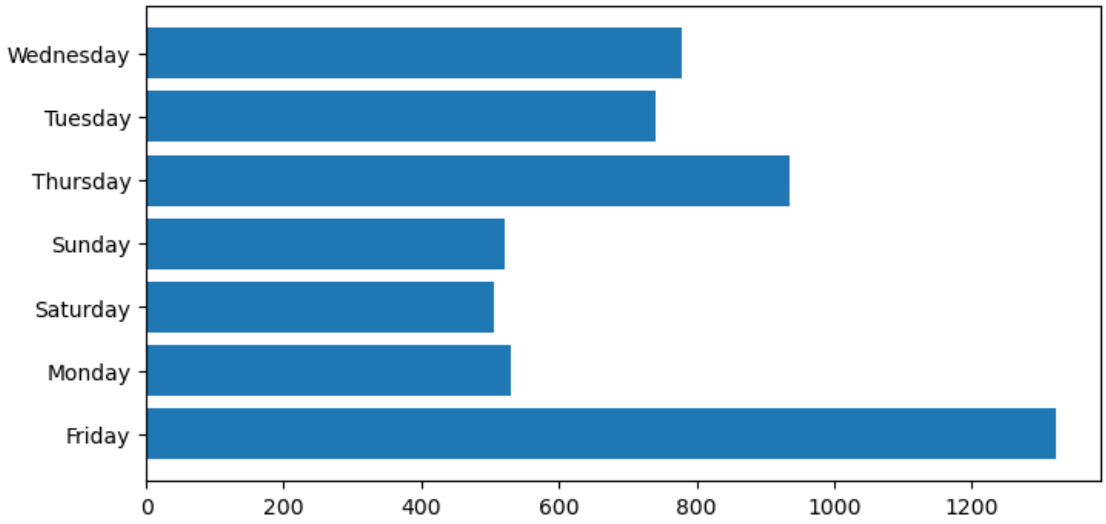
[73]:
```
temp = final_df.groupby(by='day').agg({'title':'nunique'})
temp
```

[73]:

day	title
Friday	1321
Monday	529
Saturday	506
Sunday	521
Thursday	935
Tuesday	741
Wednesday	779

Observation: Most of the movie release in `friday` and then `thursday`

[74]:
```
plt.figure(figsize=(8, 4))
plt.barh(temp.index, temp.title)
plt.show()
```



[75]:
```
final_df.groupby(by=[''])
```

[75]:
```
          title                  casts     Directors        country  \
0       Sankofa           Kofi Ghanaba  Haile Gerima  United States
1       Sankofa           Kofi Ghanaba  Haile Gerima  United States
2       Sankofa           Kofi Ghanaba  Haile Gerima  United States
3       Sankofa           Kofi Ghanaba  Haile Gerima          Ghana
4       Sankofa           Kofi Ghanaba  Haile Gerima          Ghana
…         …                      …           …             …
143087   Zubaan         Anita Shabdish   Mozez Singh          India
143088   Zubaan         Anita Shabdish   Mozez Singh          India
143089   Zubaan  Chittaranjan Tripathy   Mozez Singh          India
143090   Zubaan  Chittaranjan Tripathy   Mozez Singh          India
143091   Zubaan  Chittaranjan Tripathy   Mozez Singh          India

                       Genre show_id   type date_added  release_year rating  \
0                     Dramas      s8  Movie 2021-09-24          1993  TV-MA
1          Independent Movies      s8  Movie 2021-09-24          1993  TV-MA
2       International Movies      s8  Movie 2021-09-24          1993  TV-MA
3                     Dramas      s8  Movie 2021-09-24          1993  TV-MA
4          Independent Movies      s8  Movie 2021-09-24          1993  TV-MA
…                        …     …    …        …            …     …
143087  International Movies   s8807  Movie 2019-03-02          2015  TV-14
143088      Music & Musicals   s8807  Movie 2019-03-02          2015  TV-14
143089                Dramas   s8807  Movie 2019-03-02          2015  TV-14
143090  International Movies   s8807  Movie 2019-03-02          2015  TV-14
143091      Music & Musicals   s8807  Movie 2019-03-02          2015  TV-14

       duration                                        description  month  \
```

```
0       125 min  On a photo shoot in Ghana, an American model s…      9
1       125 min  On a photo shoot in Ghana, an American model s…      9
2       125 min  On a photo shoot in Ghana, an American model s…      9
3       125 min  On a photo shoot in Ghana, an American model s…      9
4       125 min  On a photo shoot in Ghana, an American model s…      9
…         …                                                  …    …
143087  111 min  A scrappy but poor boy worms his way into a ty…      3
143088  111 min  A scrappy but poor boy worms his way into a ty…      3
143089  111 min  A scrappy but poor boy worms his way into a ty…      3
143090  111 min  A scrappy but poor boy worms his way into a ty…      3
143091  111 min  A scrappy but poor boy worms his way into a ty…      3

              day
0          Friday
1          Friday
2          Friday
3          Friday
4          Friday
…           …
143087  Saturday
143088  Saturday
143089  Saturday
143090  Saturday
143091  Saturday

[143092 rows x 14 columns]
```

1 Business Insights

1.1 1. Genre Preferences

The data indicates a consistent preference for drama, comedy, and international TV shows/movies across various countries. This pattern suggests that investing in content aligned with these genres could yield higher viewership and engagement.

1.2 2. Seasonal Release Strategy

Timing plays a crucial role in content consumption, as evident from the recommendation to release TV shows during July/August and movies towards the end of the year or the beginning of the next. This observation suggests that aligning release dates with seasonal trends can optimize audience reach and impact.

1.3 3. Duration Optimization

There's a clear indication that the optimal duration for movies in the USA and UK falls within the 80-120 minute range. This insight implies that maintaining an appropriate duration for content can enhance audience satisfaction and retention.

1.4 4. Rating Consideration

Tailoring content ratings to match audience preferences is crucial for maximizing engagement. The recommendation to target 14+ and above ratings in the USA and India, and mature/R ratings in the UK, underscores the importance of aligning content with audience demographics and preferences.

1.5 5. Audience Trends

The declining trend in movie consumption among Indian audiences since 2018 highlights the need for strategic interventions to re-engage this audience segment. This insight suggests exploring new content strategies or marketing initiatives to revitalize interest and regain market share.

1.6 6. Regional Content Preferences

Recognizing and catering to regional content preferences is essential for capturing diverse audience segments. The popularity of anime genre content in Japan and romantic genre TV shows in South Korea underscores the importance of localized content strategies to resonate with specific cultural preferences.

1.7 7. Talent Utilization

Leveraging the popularity of actors and directors, as well as exploring successful director-actor combinations, can significantly impact content performance. This insight suggests that collaborations with renowned talent could enhance content visibility and appeal to a broader audience base.

1.8 8. Data-Driven Decision Making

The recommendations provided are based on data analysis and insights derived from observed patterns. Embracing a data-driven approach to content development and distribution can lead to more informed decision-making and higher chances of success in the competitive entertainment industry.

1.9 9. Continuous Monitoring

Given the dynamic nature of audience preferences and market trends, continuous monitoring of viewership data and industry developments is essential. Regular analysis and adaptation of content strategies based on emerging patterns can help maintain relevance and competitive advantage.

1.10 10. Iterative Improvement

Incorporating feedback from audience engagement metrics and market response allows for iterative improvement of content offerings. By iteratively refining content based on performance insights, content creators can optimize audience satisfaction and drive long-term success.

2 Recommendations

1. **Popular Genres**: Drama, comedy, and international TV shows/movies emerge as the most favored genres across various countries and platforms like TV shows and movies. Hence, creating content that aligns with these genres is advisable.

2. **Release Timing**: Consider introducing TV shows during the months of July and August, and movies during the last week of the year or the first month of the following year.
3. **Duration Recommendation (USA)**: For the audience in the USA, movies with a duration between 80-120 minutes are recommended. There is also a preference for kids' TV shows alongside the aforementioned genres.
4. **Duration Recommendation (UK)**: Similarly, for the UK audience, it is advisable to produce movies within the same duration range as the USA (80-120 minutes).
5. **Rating Recommendations**: When targeting audiences in the USA and India, content with a rating of 14+ and above is recommended, while for the UK audience, it is advised to focus on content with a mature or R rating.
6. **Audience Trends (India)**: Considering the declining trend in movie consumption among Indian audiences since 2018, incorporating more movies aimed at the Indian audience is suggested.
7. **Regional Preferences**: In Japan, anime genre content is highly recommended, while for South Korean audiences, romantic genre TV shows are popular.
8. **Consideration of Talent**: While developing content, it is essential to factor in the popularity of actors and directors within each country. Additionally, exploring the synergy between directors and actors, which is well-received, is strongly recommended.