

```
import pandas as pd
import numpy as np

df1 = pd.read_csv('C:/Users/as355/Downloads/netflix_titles.csv')
df1
```

	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	
...	...	...	...	...	
8802	s8803	Movie	Zodiac	David Fincher	
8803	s8804	TV Show	Zombie Dumb	NaN	
8804	s8805	Movie	Zombieland	Ruben Fleischer	
8805	s8806	Movie	Zoom	Peter Hewitt	
8806	s8807	Movie	Zubaan	Mozez Singh	

	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
...	...	...
8802	Mark Ruffalo, Jake Gyllenhaal, Robert Downey J...	United States
8803	NaN	NaN
8804	Jesse Eisenberg, Woody Harrelson, Emma Stone, ...	United States
8805	Tim Allen, Courteney Cox, Chevy Chase, Kate Ma...	United States
8806	Vicky Kaushal, Sarah-Jane Dias, Raaghav Chanan...	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	
...	...	...	...	...	

8802	November 20, 2019	2007	R	158 min
8803	July 1, 2019	2018	TV-Y7	2 Seasons
8804	November 1, 2019	2009	R	88 min
8805	January 11, 2020	2006	PG	88 min
8806	March 2, 2019	2015	TV-14	111 min

	listed_in \
0	Documentaries
1	International TV Shows, TV Dramas, TV Mysteries
2	Crime TV Shows, International TV Shows, TV Act...
3	Docuseries, Reality TV
4	International TV Shows, Romantic TV Shows, TV ...
...	...
8802	Cult Movies, Dramas, Thrillers
8803	Kids' TV, Korean TV Shows, TV Comedies
8804	Comedies, Horror Movies
8805	Children & Family Movies, Comedies
8806	Dramas, International Movies, Music & Musicals

	description
0	As her father nears the end of his life, filmm...
1	After crossing paths at a party, a Cape Town t...
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...
...	...
8802	A political cartoonist, a crime reporter and a...
8803	While living alone in a spooky town, a young g...
8804	Looking to survive in a world taken over by zo...
8805	Dragged from civilian life, a former superhero...
8806	A scrappy but poor boy worms his way into a ty...

[8807 rows x 12 columns]

```
df2 = pd.read_csv('C:/Users/as355/Downloads/Netflix TV Shows and
Movies.csv')
df2
```

	index	id	title	type \
0	0	tm84618	Taxi Driver	MOVIE
1	1	tm127384	Monty Python and the Holy Grail	MOVIE
2	2	tm70993	Life of Brian	MOVIE
3	3	tm190788	The Exorcist	MOVIE
4	4	ts22164	Monty Python's Flying Circus	SHOW
...	...	...	...	...
5278	5278	tm1040816	Momshies! Your Soul is Mine	MOVIE
5279	5279	tm1014599	Fine Wine	MOVIE
5280	5280	tm1045018	Clash	MOVIE
5281	5281	tm1098060	Shadow Parties	MOVIE
5282	5282	ts271048	Mighty Little Bheem: Kite Festival	SHOW

		description	release_year
\			
0		A mentally unstable Vietnam War veteran works ...	1976
1		King Arthur, accompanied by his squire, recrui...	1975
2		Brian Cohen is an average young Jewish man, bu...	1979
3		12-year-old Regan MacNeil begins to adapt an e...	1973
4		A British sketch comedy series with the shows ...	1969
...		...	...
5278		Three women with totally different lives accid...	2021
5279		A beautiful love story that can happen between...	2021
5280		A man from Nigeria returns to his family in Ca...	2021
5281		A family faces destruction in a long-running c...	2021
5282		With winter behind them, Bheem and his townspe...	2021

	age_certification	runtime	imdb_id	imdb_score	imdb_votes
0	R	113	tt0075314	8.3	795222.0
1	PG	91	tt0071853	8.2	530877.0
2	R	94	tt0079470	8.0	392419.0
3	R	133	tt0070047	8.1	391942.0
4	TV-14	30	tt0063929	8.8	72895.0
...	...	...	...	...	...
5278	NaN	108	tt14412240	5.8	26.0
5279	NaN	100	tt13857480	6.9	39.0
5280	NaN	88	tt14620732	6.5	32.0
5281	NaN	116	tt10168094	6.2	9.0
5282	NaN	0	tt13711094	8.8	16.0

[5283 rows x 11 columns]

## Merging Netflix files with imdb file

```
df = pd.merge(df1,df2[['title','imdb_score']],on="title",how='left')
df
```

	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
...	...	...	...	...
8845	s8803	Movie	Zodiac	David Fincher
8846	s8804	TV Show	Zombie Dumb	NaN
8847	s8805	Movie	Zombieland	Ruben Fleischer
8848	s8806	Movie	Zoom	Peter Hewitt
8849	s8807	Movie	Zubaan	Mozez Singh
				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
...				...
8845	Mark Ruffalo, Jake Gyllenhaal, Robert Downey J...			United States
8846				NaN NaN
8847	Jesse Eisenberg, Woody Harrelson, Emma Stone, ...			United States
8848	Tim Allen, Courteney Cox, Chevy Chase, Kate Ma...			United States
8849	Vicky Kaushal, Sarah-Jane Dias, Raaghav Chanan...			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
...	...	...	...	...
8845	November 20, 2019	2007	R	158 min
8846	July 1, 2019	2018	TV-Y7	2 Seasons
8847	November 1, 2019	2009	R	88 min
8848	January 11, 2020	2006	PG	88 min
8849	March 2, 2019	2015	TV-14	111 min
	listed_in			\
0	Documentaries			
1	International TV Shows, TV Dramas, TV Mysteries			
2	Crime TV Shows, International TV Shows, TV Act...			
3	Docuseries, Reality TV			

```

4      International TV Shows, Romantic TV Shows, TV ...
...
8845      Cult Movies, Dramas, Thrillers
8846      Kids' TV, Korean TV Shows, TV Comedies
8847      Comedies, Horror Movies
8848      Children & Family Movies, Comedies
8849      Dramas, International Movies, Music & Musicals

      description  imdb_score
0      As her father nears the end of his life, filmm...      7.4
1      After crossing paths at a party, a Cape Town t...      NaN
2      To protect his family from a powerful drug lor...      7.0
3      Feuds, flirtations and toilet talk go down amo...      6.6
4      In a city of coaching centers known to train I...      9.3
...
8845      A political cartoonist, a crime reporter and a...      NaN
8846      While living alone in a spooky town, a young g...      NaN
8847      Looking to survive in a world taken over by zo...      NaN
8848      Dragged from civilian life, a former superhero...      NaN
8849      A scrappy but poor boy worms his way into a ty...      NaN

[8850 rows x 13 columns]

```

## Keeping movies only

```

df = df[df['type']!= 'TV Show']
df.reset_index()

      index  show_id  type  title \
0         0      s1  Movie  Dick Johnson Is Dead
1         6      s7  Movie  My Little Pony: A New Generation
2         7      s8  Movie  Sankofa
3         9     s10  Movie  The Starling
4        12     s13  Movie  Je Suis Karl
...
6153    8844   s8802  Movie  Zinzana
6154    8845   s8803  Movie  Zodiac
6155    8847   s8805  Movie  Zombieland
6156    8848   s8806  Movie  Zoom
6157    8849   s8807  Movie  Zubaan

      director \
0      Kirsten Johnson
1  Robert Cullen, José Luis Ucha
2      Haile Gerima
3      Theodore Melfi
4  Christian Schwochow
...
6153    Majid Al Ansari
6154    David Fincher

```

6155	Ruben Fleischer
6156	Peter Hewitt
6157	Mozez Singh

	cast \
0	NaN
1	Vanessa Hudgens, Kimiko Glenn, James Marsden, ...
2	Kofi Ghanaba, Oyafunmike Ogunlano, Alexandra D...
3	Melissa McCarthy, Chris O'Dowd, Kevin Kline, T...
4	Luna Wedler, Jannis Niewöhner, Milan Peschel, ...
...	...
6153	Ali Suliman, Saleh Bakri, Yasa, Ali Al-Jabri, ...
6154	Mark Ruffalo, Jake Gyllenhaal, Robert Downey J...
6155	Jesse Eisenberg, Woody Harrelson, Emma Stone, ...
6156	Tim Allen, Courteney Cox, Chevy Chase, Kate Ma...
6157	Vicky Kaushal, Sarah-Jane Dias, Raaghav Chanan...

	country
date_added \	
0	United States September 25, 2021
1	NaN September 24, 2021
2	United States, Ghana, Burkina Faso, United Kin... September 24, 2021
3	United States September 24, 2021
4	Germany, Czech Republic September 23, 2021
...	...
...	...
6153	United Arab Emirates, Jordan March 9, 2016
6154	United States November 20, 2019
6155	United States November 1, 2019
6156	United States January 11, 2020
6157	India March 2, 2019

	release_year	rating	duration \
0	2020	PG-13	90 min
1	2021	PG	91 min
2	1993	TV-MA	125 min
3	2021	PG-13	104 min
4	2021	TV-MA	127 min
...	...	...	...
6153	2015	TV-MA	96 min

6154	2007	R	158 min
6155	2009	R	88 min
6156	2006	PG	88 min
6157	2015	TV-14	111 min

	listed_in \
0	Documentaries
1	Children & Family Movies
2	Dramas, Independent Movies, International Movies
3	Comedies, Dramas
4	Dramas, International Movies
...	...
6153	Dramas, International Movies, Thrillers
6154	Cult Movies, Dramas, Thrillers
6155	Comedies, Horror Movies
6156	Children & Family Movies, Comedies
6157	Dramas, International Movies, Music & Musicals

	description	imdb_score
0	As her father nears the end of his life, filmm...	7.4
1	Equestria's divided. But a bright-eyed hero be...	6.8
2	On a photo shoot in Ghana, an American model s...	7.0
3	A woman adjusting to life after a loss contend...	6.3
4	After most of her family is murdered in a terr...	NaN
...	...	...
6153	Recovering alcoholic Talal wakes up inside a s...	NaN
6154	A political cartoonist, a crime reporter and a...	NaN
6155	Looking to survive in a world taken over by zo...	NaN
6156	Dragged from civilian life, a former superhero...	NaN
6157	A scrappy but poor boy worms his way into a ty...	NaN

[6158 rows x 14 columns]

## Converting the type of date\_added column to DateTime Format

```
df['date_added'] = pd.to_datetime(df['date_added'])
df
```

C:\Users\as355\AppData\Local\Temp\ipykernel\_5836\478102478.py:1:  
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](https://pandas.pydata.org/pandas-docs/stable/user_guide/indexing.html#returning-a-view-versus-a-copy)

```
df['date_added'] = pd.to_datetime(df['date_added'])
```

	show_id	type	title \
0	s1	Movie	Dick Johnson Is Dead

6	s7	Movie	My Little Pony: A New Generation
7	s8	Movie	Sankofa
9	s10	Movie	The Starling
12	s13	Movie	Je Suis Karl
...	...	...	...
8844	s8802	Movie	Zinzana
8845	s8803	Movie	Zodiac
8847	s8805	Movie	Zombieland
8848	s8806	Movie	Zoom
8849	s8807	Movie	Zubaan

	director	\
0	Kirsten Johnson	
6	Robert Cullen, José Luis Ucha	
7	Haile Gerima	
9	Theodore Melfi	
12	Christian Schwochow	
...	...	
8844	Majid Al Ansari	
8845	David Fincher	
8847	Ruben Fleischer	
8848	Peter Hewitt	
8849	Mozez Singh	

	cast	\
0	NaN	
6	Vanessa Hudgens, Kimiko Glenn, James Marsden, ...	
7	Kofi Ghanaba, Oyafunmike Ogunlano, Alexandra D...	
9	Melissa McCarthy, Chris O'Dowd, Kevin Kline, T...	
12	Luna Wedler, Jannis Niewöhner, Milan Peschel, ...	
...	...	
8844	Ali Suliman, Saleh Bakri, Yasa, Ali Al-Jabri, ...	
8845	Mark Ruffalo, Jake Gyllenhaal, Robert Downey J...	
8847	Jesse Eisenberg, Woody Harrelson, Emma Stone, ...	
8848	Tim Allen, Courteney Cox, Chevy Chase, Kate Ma...	
8849	Vicky Kaushal, Sarah-Jane Dias, Raaghav Chanan...	

	country	date_added	\
0	United States	2021-09-25	
6	NaN	2021-09-24	
7	United States, Ghana, Burkina Faso, United Kin...	2021-09-24	
9	United States	2021-09-24	
12	Germany, Czech Republic	2021-09-23	
...	...	...	
8844	United Arab Emirates, Jordan	2016-03-09	
8845	United States	2019-11-20	
8847	United States	2019-11-01	
8848	United States	2020-01-11	
8849	India	2019-03-02	



```

      release_year rating duration \
0          2020   PG-13    90 min
6          2021     PG    91 min
7          1993   TV-MA   125 min
9          2021   PG-13   104 min
12         2021   TV-MA   127 min
...         ...     ...     ...
8844        2015   TV-MA    96 min
8845        2007     R   158 min
8847        2009     R    88 min
8848        2006     PG    88 min
8849        2015   TV-14   111 min

                                listed_in \
0                                Documentaries
6                                Children & Family Movies
7      Dramas, Independent Movies, International Movies
9                                Comedies, Dramas
12                               Dramas, International Movies
...         ...
8844      Dramas, International Movies, Thrillers
8845      Cult Movies, Dramas, Thrillers
8847      Comedies, Horror Movies
8848      Children & Family Movies, Comedies
8849      Dramas, International Movies, Music & Musicals

                                description  imdb_score
0      As her father nears the end of his life, filmm...      7.4
6      Equestria's divided. But a bright-eyed hero be...      6.8
7      On a photo shoot in Ghana, an American model s...      7.0
9      A woman adjusting to life after a loss contend...      6.3
12     After most of her family is murdered in a terr...      NaN
...         ...
8844     Recovering alcoholic Talal wakes up inside a s...      NaN
8845     A political cartoonist, a crime reporter and a...      NaN
8847     Looking to survive in a world taken over by zo...      NaN
8848     Dragged from civilian life, a former superhero...      NaN
8849     A scrappy but poor boy worms his way into a ty...      NaN

[6158 rows x 13 columns]

```

## Extracting year from date\_added column

```

df['year'] = df['date_added'].dt.year
df

C:\Users\as355\AppData\Local\Temp\ipykernel_5836\982568261.py:1:
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](https://pandas.pydata.org/pandas-docs/stable/user_guide/indexing.html#returning-a-view-versus-a-copy)

```
df['year'] = df['date_added'].dt.year
```

	show_id	type	title \
0	s1	Movie	Dick Johnson Is Dead
6	s7	Movie	My Little Pony: A New Generation
7	s8	Movie	Sankofa
9	s10	Movie	The Starling
12	s13	Movie	Je Suis Karl
...	...	...	...
8844	s8802	Movie	Zinzana
8845	s8803	Movie	Zodiac
8847	s8805	Movie	Zombieland
8848	s8806	Movie	Zoom
8849	s8807	Movie	Zubaan

	director \
0	Kirsten Johnson
6	Robert Cullen, José Luis Ucha
7	Haile Gerima
9	Theodore Melfi
12	Christian Schwochow
...	...
8844	Majid Al Ansari
8845	David Fincher
8847	Ruben Fleischer
8848	Peter Hewitt
8849	Mozez Singh

	cast \
0	NaN
6	Vanessa Hudgens, Kimiko Glenn, James Marsden, ...
7	Kofi Ghanaba, Oyafunmike Ogunlano, Alexandra D...
9	Melissa McCarthy, Chris O'Dowd, Kevin Kline, T...
12	Luna Wedler, Jannis Niewöhner, Milan Peschel, ...
...	...
8844	Ali Suliman, Saleh Bakri, Yasa, Ali Al-Jabri, ...
8845	Mark Ruffalo, Jake Gyllenhaal, Robert Downey J...
8847	Jesse Eisenberg, Woody Harrelson, Emma Stone, ...
8848	Tim Allen, Courteney Cox, Chevy Chase, Kate Ma...
8849	Vicky Kaushal, Sarah-Jane Dias, Raaghav Chanan...

	country	date_added \
0	United States	2021-09-25
6		NaN 2021-09-24
7	United States, Ghana, Burkina Faso, United Kin...	2021-09-24
9		United States 2021-09-24

12	Germany, Czech Republic	2021-09-23
...	...	...
8844	United Arab Emirates, Jordan	2016-03-09
8845	United States	2019-11-20
8847	United States	2019-11-01
8848	United States	2020-01-11
8849	India	2019-03-02

	release_year	rating	duration	\
0	2020	PG-13	90 min	
6	2021	PG	91 min	
7	1993	TV-MA	125 min	
9	2021	PG-13	104 min	
12	2021	TV-MA	127 min	
...	...	...	...	
8844	2015	TV-MA	96 min	
8845	2007	R	158 min	
8847	2009	R	88 min	
8848	2006	PG	88 min	
8849	2015	TV-14	111 min	

	listed_in	\
0	Documentaries	
6	Children & Family Movies	
7	Dramas, Independent Movies, International Movies	
9	Comedies, Dramas	
12	Dramas, International Movies	
...	...	
8844	Dramas, International Movies, Thrillers	
8845	Cult Movies, Dramas, Thrillers	
8847	Comedies, Horror Movies	
8848	Children & Family Movies, Comedies	
8849	Dramas, International Movies, Music & Musicals	

	description	imdb_score
year		
0	As her father nears the end of his life, filmm...	7.4
2021		
6	Equestria's divided. But a bright-eyed hero be...	6.8
2021		
7	On a photo shoot in Ghana, an American model s...	7.0
2021		
9	A woman adjusting to life after a loss contend...	6.3
2021		
12	After most of her family is murdered in a terr...	NaN
2021		
...	...	...
...		
8844	Recovering alcoholic Talal wakes up inside a s...	NaN
2016		

```

8845  A political cartoonist, a crime reporter and a...      NaN
2019
8847  Looking to survive in a world taken over by zo...      NaN
2019
8848  Dragged from civilian life, a former superhero...      NaN
2020
8849  A scrappy but poor boy worms his way into a ty...      NaN
2019

[6158 rows x 14 columns]

```

## Extracting numerical values from duration column

```

df['duration'] = df['duration'].str.extract('(\d+)')
df

<>:1: SyntaxWarning: invalid escape sequence '\d'
<>:1: SyntaxWarning: invalid escape sequence '\d'
C:\Users\as355\AppData\Local\Temp\ipykernel_5836\67282490.py:1:
SyntaxWarning: invalid escape sequence '\d'
    df['duration'] = df['duration'].str.extract('(\d+)')
C:\Users\as355\AppData\Local\Temp\ipykernel_5836\67282490.py:1:
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
    df['duration'] = df['duration'].str.extract('(\d+)')

```

	show_id	type	title \
0	s1	Movie	Dick Johnson Is Dead
6	s7	Movie	My Little Pony: A New Generation
7	s8	Movie	Sankofa
9	s10	Movie	The Starling
12	s13	Movie	Je Suis Karl
...	...	...	...
8844	s8802	Movie	Zinzana
8845	s8803	Movie	Zodiac
8847	s8805	Movie	Zombieland
8848	s8806	Movie	Zoom
8849	s8807	Movie	Zubaan

	director \
0	Kirsten Johnson
6	Robert Cullen, José Luis Ucha
7	Haile Gerima
9	Theodore Melfi
12	Christian Schwochow

...	...
8844	Majid Al Ansari
8845	David Fincher
8847	Ruben Fleischer
8848	Peter Hewitt
8849	Mozez Singh

	cast \
0	NaN
6	Vanessa Hudgens, Kimiko Glenn, James Marsden, ...
7	Kofi Ghanaba, Oyafunmike Ogunlano, Alexandra D...
9	Melissa McCarthy, Chris O'Dowd, Kevin Kline, T...
12	Luna Wedler, Jannis Niewöhner, Milan Peschel, ...
...	...
8844	Ali Suliman, Saleh Bakri, Yasa, Ali Al-Jabri, ...
8845	Mark Ruffalo, Jake Gyllenhaal, Robert Downey J...
8847	Jesse Eisenberg, Woody Harrelson, Emma Stone, ...
8848	Tim Allen, Courteney Cox, Chevy Chase, Kate Ma...
8849	Vicky Kaushal, Sarah-Jane Dias, Raaghav Chanan...

	country	date_added \
0	United States	2021-09-25
6	NaN	2021-09-24
7	United States, Ghana, Burkina Faso, United Kin...	2021-09-24
9	United States	2021-09-24
12	Germany, Czech Republic	2021-09-23
...	...	...
8844	United Arab Emirates, Jordan	2016-03-09
8845	United States	2019-11-20
8847	United States	2019-11-01
8848	United States	2020-01-11
8849	India	2019-03-02

	release_year	rating	duration \
0	2020	PG-13	90
6	2021	PG	91
7	1993	TV-MA	125
9	2021	PG-13	104
12	2021	TV-MA	127
...	...	...	...
8844	2015	TV-MA	96
8845	2007	R	158
8847	2009	R	88
8848	2006	PG	88
8849	2015	TV-14	111

	listed_in \
0	Documentaries
6	Children & Family Movies
7	Dramas, Independent Movies, International Movies

```

9          Comedies, Dramas
12         Dramas, International Movies
...
8844      Dramas, International Movies, Thrillers
8845      Cult Movies, Dramas, Thrillers
8847      Comedies, Horror Movies
8848      Children & Family Movies, Comedies
8849      Dramas, International Movies, Music & Musicals

```

	description	imdb_score
year		
0	As her father nears the end of his life, filmm...	7.4
2021		
6	Equestria's divided. But a bright-eyed hero be...	6.8
2021		
7	On a photo shoot in Ghana, an American model s...	7.0
2021		
9	A woman adjusting to life after a loss contend...	6.3
2021		
12	After most of her family is murdered in a terr...	NaN
2021		
...	...	...
...		
8844	Recovering alcoholic Talal wakes up inside a s...	NaN
2016		
8845	A political cartoonist, a crime reporter and a...	NaN
2019		
8847	Looking to survive in a world taken over by zo...	NaN
2019		
8848	Dragged from civilian life, a former superhero...	NaN
2020		
8849	A scrappy but poor boy worms his way into a ty...	NaN
2019		

```
[6158 rows x 14 columns]
```

```
df['duration'] = df['duration'].astype('float')
```

```
C:\Users\as355\AppData\Local\Temp\ipykernel_5836\3012677645.py:1:
```

```
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
```

```
df['duration'] = df['duration'].astype('float')
```

Removing data if missing values are found in columns  
"title", "listed\_in" or "ratings"

```
df.isnull().sum()
```

```
show_id      0
type         0
title        0
director    188
cast       475
country     442
date_added   0
release_year  0
rating       2
duration     3
listed_in    0
description  0
imdb_score  3800
year         0
dtype: int64
```

```
s_id = df['show_id'][df['rating'].isna()]
s_id
```

```
6032    s5990
7580    s7538
```

```
Name: show_id, dtype: object
```

```
df.set_index('show_id', inplace=True)
df
```

	type	title \
show_id		
s1	Movie	Dick Johnson Is Dead
s7	Movie	My Little Pony: A New Generation
s8	Movie	Sankofa
s10	Movie	The Starling
s13	Movie	Je Suis Karl
...	...	...
s8802	Movie	Zinzana
s8803	Movie	Zodiac
s8805	Movie	Zombieland
s8806	Movie	Zoom
s8807	Movie	Zubaan

	director \
show_id	
s1	Kirsten Johnson
s7	Robert Cullen, José Luis Ucha
s8	Haile Gerima
s10	Theodore Melfi

s13	Christian Schwochow
...	...
s8802	Majid Al Ansari
s8803	David Fincher
s8805	Ruben Fleischer
s8806	Peter Hewitt
s8807	Mozez Singh

	cast	\
show_id		
s1		NaN
s7	Vanessa Hudgens, Kimiko Glenn, James Marsden, ...	
s8	Kofi Ghanaba, Oyafunmike Ogunlano, Alexandra D...	
s10	Melissa McCarthy, Chris O'Dowd, Kevin Kline, T...	
s13	Luna Wedler, Jannis Niewöhner, Milan Peschel, ...	
...	...	
s8802	Ali Suliman, Saleh Bakri, Yasa, Ali Al-Jabri, ...	
s8803	Mark Ruffalo, Jake Gyllenhaal, Robert Downey J...	
s8805	Jesse Eisenberg, Woody Harrelson, Emma Stone, ...	
s8806	Tim Allen, Courteney Cox, Chevy Chase, Kate Ma...	
s8807	Vicky Kaushal, Sarah-Jane Dias, Raaghav Chanan...	

	country	date_added
\		
show_id		
s1	United States	2021-09-25
s7		NaN 2021-09-24
s8	United States, Ghana, Burkina Faso, United Kin...	2021-09-24
s10	United States	2021-09-24
s13	Germany, Czech Republic	2021-09-23
...	...	...
s8802	United Arab Emirates, Jordan	2016-03-09
s8803	United States	2019-11-20
s8805	United States	2019-11-01
s8806	United States	2020-01-11
s8807	India	2019-03-02

	release_year	rating	duration	\
show_id				



s1	2020	PG-13	90.0
s7	2021	PG	91.0
s8	1993	TV-MA	125.0
s10	2021	PG-13	104.0
s13	2021	TV-MA	127.0
...	...	...	...
s8802	2015	TV-MA	96.0
s8803	2007	R	158.0
s8805	2009	R	88.0
s8806	2006	PG	88.0
s8807	2015	TV-14	111.0

	listed_in \
show_id	
s1	Documentaries
s7	Children & Family Movies
s8	Dramas, Independent Movies, International Movies
s10	Comedies, Dramas
s13	Dramas, International Movies
...	...
s8802	Dramas, International Movies, Thrillers
s8803	Cult Movies, Dramas, Thrillers
s8805	Comedies, Horror Movies
s8806	Children & Family Movies, Comedies
s8807	Dramas, International Movies, Music & Musicals

	description	imdb_score
year		
show_id		
s1	As her father nears the end of his life, filmm...	7.4
2021		
s7	Equestria's divided. But a bright-eyed hero be...	6.8
2021		
s8	On a photo shoot in Ghana, an American model s...	7.0
2021		
s10	A woman adjusting to life after a loss contend...	6.3
2021		
s13	After most of her family is murdered in a terr...	NaN
2021		
...	...	...
...		
s8802	Recovering alcoholic Talal wakes up inside a s...	NaN
2016		
s8803	A political cartoonist, a crime reporter and a...	NaN
2019		
s8805	Looking to survive in a world taken over by zo...	NaN
2019		
s8806	Dragged from civilian life, a former superhero...	NaN
2020		

```
s8807    A scrappy but poor boy worms his way into a ty...    NaN
2019
```

```
[6158 rows x 13 columns]
```

```
df.drop(s_id,axis=0,inplace=True)
```

```
C:\Users\as355\AppData\Local\Temp\ipykernel_5836\1538357047.py:1:
SettingWithCopyWarning:
```

```
A value is trying to be set on a copy of a slice from a DataFrame
```

```
See the caveats in the documentation:
```

```
https://pandas.pydata.org/pandas-docs/stable/user\_guide/indexing.html#
returning-a-view-versus-a-copy
```

```
df.drop(s_id,axis=0,inplace=True)
```

```
df.reset_index(inplace=True)
```

## removing rows having nan values in ratings column

```
df.dropna(subset=['imdb_score'],inplace=True)
df
```

```
C:\Users\as355\AppData\Local\Temp\ipykernel_5836\1218822389.py:1:
SettingWithCopyWarning:
```

```
A value is trying to be set on a copy of a slice from a DataFrame
```

```
See the caveats in the documentation:
```

```
https://pandas.pydata.org/pandas-docs/stable/user\_guide/indexing.html#
returning-a-view-versus-a-copy
```

```
df.dropna(subset=['imdb_score'],inplace=True)
```

	show_id	type	title \
0	s1	Movie	Dick Johnson Is Dead
1	s7	Movie	My Little Pony: A New Generation
2	s8	Movie	Sankofa
3	s10	Movie	The Starling
5	s14	Movie	Confessions of an Invisible Girl
...	...	...	...
5914	s8512	Movie	The Social Network
5927	s8528	Movie	The Texas Chainsaw Massacre
6013	s8636	Movie	True Grit
6076	s8707	Movie	We the Animals
6116	s8758	Movie	World Trade Center

	director \
0	Kirsten Johnson
1	Robert Cullen, José Luis Ucha
2	Haile Gerima
3	Theodore Melfi
5	Bruno Garotti

...	...
5914	David Fincher
5927	Tobe Hooper
6013	Henry Hathaway
6076	Jeremiah Zagar
6116	Oliver Stone

	cast \
0	NaN
1	Vanessa Hudgens, Kimiko Glenn, James Marsden, ...
2	Kofi Ghanaba, Oyafunmike Ogunlano, Alexandra D...
3	Melissa McCarthy, Chris O'Dowd, Kevin Kline, T...
5	Klara Castanho, Lucca Picon, Júlia Gomes, Marc...
...	...
5914	Jesse Eisenberg, Andrew Garfield, Justin Timbe...
5927	Gunnar Hansen, Jim Siedow, Marilyn Burns, Edwi...
6013	John Wayne, Glen Campbell, Kim Darby, Jeremy S...
6076	Evan Rosado, Josiah Gabriel, Isaiah Kristian, ...
6116	Nicolas Cage, Michael Peña, Maggie Gyllenhaal,...

	country	date_added \
0	United States	2021-09-25
1	NaN	2021-09-24
2	United States, Ghana, Burkina Faso, United Kin...	2021-09-24
3	United States	2021-09-24
5	NaN	2021-09-22
...	...	...
5914	United States	2020-04-01
5927	United States	2019-10-22
6013	United States	2020-01-01
6076	United States	2019-02-11
6116	United States	2019-11-20

	release_year	rating	duration \
0	2020	PG-13	90.0
1	2021	PG	91.0
2	1993	TV-MA	125.0
3	2021	PG-13	104.0
5	2021	TV-PG	91.0
...	...	...	...
5914	2010	PG-13	121.0
5927	1974	R	83.0
6013	1969	G	128.0
6076	2018	R	93.0
6116	2006	PG-13	129.0

	listed_in \
0	Documentaries
1	Children & Family Movies
2	Dramas, Independent Movies, International Movies

```

3          Comedies, Dramas
5      Children & Family Movies, Comedies
...
5914          Dramas
5927      Cult Movies, Horror Movies, Independent Movies
6013          Classic Movies, Dramas
6076          Dramas, Independent Movies
6116      Action & Adventure, Dramas

```

	description	imdb_score
year		
0	As her father nears the end of his life, filmm...	7.4
2021		
1	Equestria's divided. But a bright-eyed hero be...	6.8
2021		
2	On a photo shoot in Ghana, an American model s...	7.0
2021		
3	A woman adjusting to life after a loss contend...	6.3
2021		
5	When the clever but socially-awkward Tetê join...	5.2
2021		
...	...	...
...		
5914	Director David Fincher's biographical drama ch...	7.8
2020		
5927	When some friends visit a vandalized graveyard...	6.2
2019		
6013	Teenage tomboy Mattie Ross enlists the help of...	7.6
2020		
6076	A trio of brothers cope with their parents' vo...	6.9
2019		
6116	Working under treacherous conditions, an army ...	6.0
2019		

[2358 rows x 14 columns]

removing rows having nan value in duration column

```

df.dropna(subset=['duration'],inplace=True)
df

```

C:\Users\as355\AppData\Local\Temp\ipykernel\_5836\3569969581.py:1:  
SettingWithCopyWarning:  
A value is trying to be set on a copy of a slice from a DataFrame

See the caveats in the documentation:  
[https://pandas.pydata.org/pandas-docs/stable/user\\_guide/indexing.html#returning-a-view-versus-a-copy](https://pandas.pydata.org/pandas-docs/stable/user_guide/indexing.html#returning-a-view-versus-a-copy)  
df.dropna(subset=['duration'],inplace=True)

	show_id	type	title \
0	s1	Movie	Dick Johnson Is Dead
1	s7	Movie	My Little Pony: A New Generation
2	s8	Movie	Sankofa
3	s10	Movie	The Starling
5	s14	Movie	Confessions of an Invisible Girl
...	...	...	...
5914	s8512	Movie	The Social Network
5927	s8528	Movie	The Texas Chainsaw Massacre
6013	s8636	Movie	True Grit
6076	s8707	Movie	We the Animals
6116	s8758	Movie	World Trade Center
		director \	
0		Kirsten Johnson	
1	Robert Cullen,	José Luis Ucha	
2		Haile Gerima	
3		Theodore Melfi	
5		Bruno Garotti	
...		...	
5914		David Fincher	
5927		Tobe Hooper	
6013		Henry Hathaway	
6076		Jeremiah Zagar	
6116		Oliver Stone	
		cast \	
0		NaN	
1	Vanessa Hudgens, Kimiko Glenn, James Marsden, ...		
2	Kofi Ghanaba, Oyafunmike Ogunlano, Alexandra D...		
3	Melissa McCarthy, Chris O'Dowd, Kevin Kline, T...		
5	Klara Castanho, Lucca Picon, Júlia Gomes, Marc...		
...		...	
5914	Jesse Eisenberg, Andrew Garfield, Justin Timbe...		
5927	Gunnar Hansen, Jim Siedow, Marilyn Burns, Edwi...		
6013	John Wayne, Glen Campbell, Kim Darby, Jeremy S...		
6076	Evan Rosado, Josiah Gabriel, Isaiah Kristian, ...		
6116	Nicolas Cage, Michael Peña, Maggie Gyllenhaal,...		
		country	date_added \
0		United States	2021-09-25
1		NaN	2021-09-24
2	United States, Ghana, Burkina Faso, United Kin...	United States	2021-09-24
3		United States	2021-09-24
5		NaN	2021-09-22
...		...	...
5914		United States	2020-04-01
5927		United States	2019-10-22
6013		United States	2020-01-01
6076		United States	2019-02-11

6116 United States 2019-11-20

	release_year	rating	duration	\
0	2020	PG-13	90.0	
1	2021	PG	91.0	
2	1993	TV-MA	125.0	
3	2021	PG-13	104.0	
5	2021	TV-PG	91.0	
...	...	...	...	
5914	2010	PG-13	121.0	
5927	1974	R	83.0	
6013	1969	G	128.0	
6076	2018	R	93.0	
6116	2006	PG-13	129.0	

	listed_in	\
0	Documentaries	
1	Children & Family Movies	
2	Dramas, Independent Movies, International Movies	
3	Comedies, Dramas	
5	Children & Family Movies, Comedies	
...	...	
5914	Dramas	
5927	Cult Movies, Horror Movies, Independent Movies	
6013	Classic Movies, Dramas	
6076	Dramas, Independent Movies	
6116	Action & Adventure, Dramas	

	description	imdb_score
year		
0	As her father nears the end of his life, filmm...	7.4
2021		
1	Equestria's divided. But a bright-eyed hero be...	6.8
2021		
2	On a photo shoot in Ghana, an American model s...	7.0
2021		
3	A woman adjusting to life after a loss contend...	6.3
2021		
5	When the clever but socially-awkward Tetê join...	5.2
2021		
...	...	...
...		
5914	Director David Fincher's biographical drama ch...	7.8
2020		
5927	When some friends visit a vandalized graveyard...	6.2
2019		
6013	Teenage tomboy Mattie Ross enlists the help of...	7.6
2020		
6076	A trio of brothers cope with their parents' vo...	6.9
2019		

```
6116 Working under treacherous conditions, an army ... 6.0
2019
```

```
[2357 rows x 14 columns]
```

```
df.isnull().sum()
```

```
show_id      0
type         0
title        0
director     57
cast        144
country     135
date_added   0
release_year  0
rating       0
duration     0
listed_in    0
description  0
imdb_score   0
year         0
dtype: int64
```

## deleting rows having nan values in country column

```
df.dropna(subset=['country'],inplace=True)
df
```

```
C:\Users\as355\AppData\Local\Temp\ipykernel_5836\731762678.py:1:
```

```
SettingWithCopyWarning:
```

```
A value is trying to be set on a copy of a slice from a DataFrame
```

```
See the caveats in the documentation:
```

```
https://pandas.pydata.org/pandas-docs/stable/user\_guide/indexing.html#
returning-a-view-versus-a-copy
```

```
df.dropna(subset=['country'],inplace=True)
```

	show_id	type	title	director
0	s1	Movie	Dick Johnson Is Dead	Kirsten Johnson
2	s8	Movie	Sankofa	Haile Gerima
3	s10	Movie	The Starling	Theodore Melfi
10	s25	Movie	Jeans	S. Shankar
12	s28	Movie	Grown Ups	Dennis Dugan
...	...	...	...	...
5914	s8512	Movie	The Social Network	David Fincher
5927	s8528	Movie	The Texas Chainsaw Massacre	Tobe Hooper
6013	s8636	Movie	True Grit	Henry Hathaway
6076	s8707	Movie	We the Animals	Jeremiah Zagar
6116	s8758	Movie	World Trade Center	Oliver Stone

	cast \
0	NaN
2	Kofi Ghanaba, Oyafunmike Ogunlano, Alexandra D...
3	Melissa McCarthy, Chris O'Dowd, Kevin Kline, T...
10	Prashanth, Aishwarya Rai Bachchan, Sri Lakshmi...
12	Adam Sandler, Kevin James, Chris Rock, David S...
...	...
5914	Jesse Eisenberg, Andrew Garfield, Justin Timbe...
5927	Gunnar Hansen, Jim Siedow, Marilyn Burns, Edwi...
6013	John Wayne, Glen Campbell, Kim Darby, Jeremy S...
6076	Evan Rosado, Josiah Gabriel, Isaiah Kristian, ...
6116	Nicolas Cage, Michael Peña, Maggie Gyllenhaal,...

	country	date_added \
0	United States	2021-09-25
2	United States, Ghana, Burkina Faso, United Kin...	2021-09-24
3	United States	2021-09-24
10	India	2021-09-21
12	United States	2021-09-20
...	...	...
5914	United States	2020-04-01
5927	United States	2019-10-22
6013	United States	2020-01-01
6076	United States	2019-02-11
6116	United States	2019-11-20

	release_year	rating	duration \
0	2020	PG-13	90.0
2	1993	TV-MA	125.0
3	2021	PG-13	104.0
10	1998	TV-14	166.0
12	2010	PG-13	103.0
...	...	...	...
5914	2010	PG-13	121.0
5927	1974	R	83.0
6013	1969	G	128.0
6076	2018	R	93.0
6116	2006	PG-13	129.0

	listed_in \
0	Documentaries
2	Dramas, Independent Movies, International Movies
3	Comedies, Dramas
10	Comedies, International Movies, Romantic Movies
12	Comedies
...	...
5914	Dramas
5927	Cult Movies, Horror Movies, Independent Movies
6013	Classic Movies, Dramas
6076	Dramas, Independent Movies



6116		Action & Adventure, Dramas	
		description	imdb_score
year			
0	As her father nears the end of his life, filmm...		7.4
2021			
2	On a photo shoot in Ghana, an American model s...		7.0
2021			
3	A woman adjusting to life after a loss contend...		6.3
2021			
10	When the father of the man she loves insists t...		6.4
2021			
12	Mourning the loss of their beloved junior high...		5.9
2021			
...		...	...
...			
5914	Director David Fincher's biographical drama ch...		7.8
2020			
5927	When some friends visit a vandalized graveyard...		6.2
2019			
6013	Teenage tomboy Mattie Ross enlists the help of...		7.6
2020			
6076	A trio of brothers cope with their parents' vo...		6.9
2019			
6116	Working under treacherous conditions, an army ...		6.0
2019			

[2222 rows x 14 columns]

```
df.isnull().sum()
```

```
show_id      0
type         0
title        0
director     49
cast        131
country      0
date_added   0
release_year  0
rating       0
duration     0
listed_in    0
description   0
imdb_score   0
year         0
dtype: int64
```

## adding 'genre' column

```
df['genre'] = df['listed_in'].apply(lambda x: x.split(',')[0])
df
```

C:\Users\as355\AppData\Local\Temp\ipykernel\_5836\4034668297.py:1:  
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](https://pandas.pydata.org/pandas-docs/stable/user_guide/indexing.html#returning-a-view-versus-a-copy)

```
df['genre'] = df['listed_in'].apply(lambda x: x.split(',')[0])
```

	show_id	type		title	director \
0	s1	Movie	Dick Johnson	Is Dead	Kirsten Johnson
2	s8	Movie		Sankofa	Haile Gerima
3	s10	Movie		The Starling	Theodore Melfi
10	s25	Movie		Jeans	S. Shankar
12	s28	Movie		Grown Ups	Dennis Dugan
...	...	...		...	...
5914	s8512	Movie		The Social Network	David Fincher
5927	s8528	Movie	The Texas Chainsaw	Massacre	Tobe Hooper
6013	s8636	Movie		True Grit	Henry Hathaway
6076	s8707	Movie		We the Animals	Jeremiah Zagar
6116	s8758	Movie		World Trade Center	Oliver Stone

		cast \
0		NaN
2	Kofi Ghanaba, Oyafunmike Ogunlano, Alexandra D...	
3	Melissa McCarthy, Chris O'Dowd, Kevin Kline, T...	
10	Prashanth, Aishwarya Rai Bachchan, Sri Lakshmi...	
12	Adam Sandler, Kevin James, Chris Rock, David S...	
...	...	...
5914	Jesse Eisenberg, Andrew Garfield, Justin Timbe...	
5927	Gunnar Hansen, Jim Siedow, Marilyn Burns, Edwi...	
6013	John Wayne, Glen Campbell, Kim Darby, Jeremy S...	
6076	Evan Rosado, Josiah Gabriel, Isaiah Kristian, ...	
6116	Nicolas Cage, Michael Peña, Maggie Gyllenhaal,...	

		country	date_added \
0		United States	2021-09-25
2	United States, Ghana, Burkina Faso,	United Kin...	2021-09-24
3		United States	2021-09-24
10		India	2021-09-21
12		United States	2021-09-20
...		...	...
5914		United States	2020-04-01
5927		United States	2019-10-22
6013		United States	2020-01-01

6076	United States	2019-02-11
6116	United States	2019-11-20

	release_year	rating	duration	\
0	2020	PG-13	90.0	
2	1993	TV-MA	125.0	
3	2021	PG-13	104.0	
10	1998	TV-14	166.0	
12	2010	PG-13	103.0	
...	...	...	...	
5914	2010	PG-13	121.0	
5927	1974	R	83.0	
6013	1969	G	128.0	
6076	2018	R	93.0	
6116	2006	PG-13	129.0	

	listed_in	\
0	Documentaries	
2	Dramas, Independent Movies, International Movies	
3	Comedies, Dramas	
10	Comedies, International Movies, Romantic Movies	
12	Comedies	
...	...	
5914	Dramas	
5927	Cult Movies, Horror Movies, Independent Movies	
6013	Classic Movies, Dramas	
6076	Dramas, Independent Movies	
6116	Action & Adventure, Dramas	

	description	imdb_score
year \		
0	As her father nears the end of his life, filmm...	7.4
2021		
2	On a photo shoot in Ghana, an American model s...	7.0
2021		
3	A woman adjusting to life after a loss contend...	6.3
2021		
10	When the father of the man she loves insists t...	6.4
2021		
12	Mourning the loss of their beloved junior high...	5.9
2021		
...	...	...
...		
5914	Director David Fincher's biographical drama ch...	7.8
2020		
5927	When some friends visit a vandalized graveyard...	6.2
2019		
6013	Teenage tomboy Mattie Ross enlists the help of...	7.6
2020		
6076	A trio of brothers cope with their parents' vo...	6.9

```
2019
6116 Working under treacherous conditions, an army ... 6.0
2019
```

```

      genre
0      Documentaries
2      Dramas
3      Comedies
10     Comedies
12     Comedies
...
5914    Dramas
5927    Cult Movies
6013    Classic Movies
6076    Dramas
6116    Action & Adventure

[2222 rows x 15 columns]
```

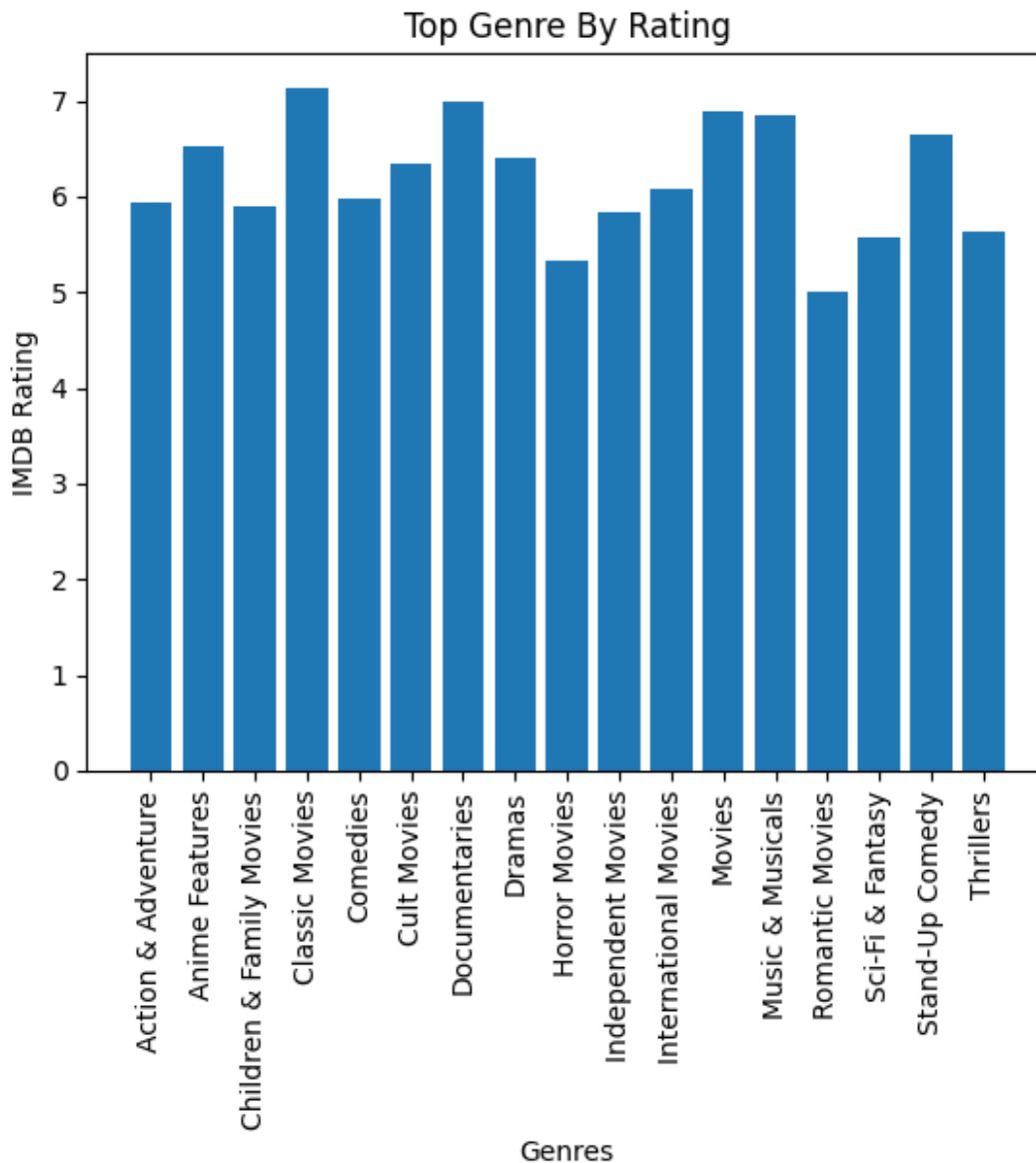
exporting the cleaned file

```
df.to_csv('C:/Users/as355/Downloads/cleaned_netflix_data.csv')
```

## Exploratory Data Analysis

genre by average rating

```
import matplotlib.pyplot as plt
plt.bar(genrel,ratings)
plt.xlabel("Genres")
plt.ylabel("IMDB Rating")
plt.title("Top Genre By Rating")
plt.xticks(genrel,rotation='vertical')
plt.show()
```



```
genre = df['genre'].unique()
genre1 = np.sort(genre)
genre1

array(['Action & Adventure', 'Anime Features', 'Children & Family
Movies',
      'Classic Movies', 'Comedies', 'Cult Movies', 'Documentaries',
      'Dramas', 'Horror Movies', 'Independent Movies',
      'International Movies', 'Movies', 'Music & Musicals',
      'Romantic Movies', 'Sci-Fi & Fantasy', 'Stand-Up Comedy',
      'Thrillers'], dtype=object)

ratings = df.groupby('genre')['imdb_score'].mean()
ratings
```

```

genre
Action & Adventure      5.940000
Anime Features          6.520000
Children & Family Movies 5.896933
Classic Movies          7.136364
Comedies                5.980483
Cult Movies             6.350000
Documentaries           6.981513
Dramas                  6.406815
Horror Movies           5.325743
Independent Movies      5.837500
International Movies    6.087879
Movies                  6.880000
Music & Musicals         6.857143
Romantic Movies         5.000000
Sci-Fi & Fantasy         5.566667
Stand-Up Comedy         6.649485
Thrillers               5.626667
Name: imdb_score, dtype: float64

```

## rating over years

```

ratings1 = df.groupby('year')['imdb_score'].mean()
ratings1

year
2009      7.200000
2011      6.550000
2013      7.133333
2014      7.360000
2015      6.488462
2016      6.593243
2017      6.217012
2018      6.233850
2019      6.240376
2020      6.223411
2021      6.148379
Name: imdb_score, dtype: float64

years = df['year'].unique()
years1 = np.sort(years)
years1

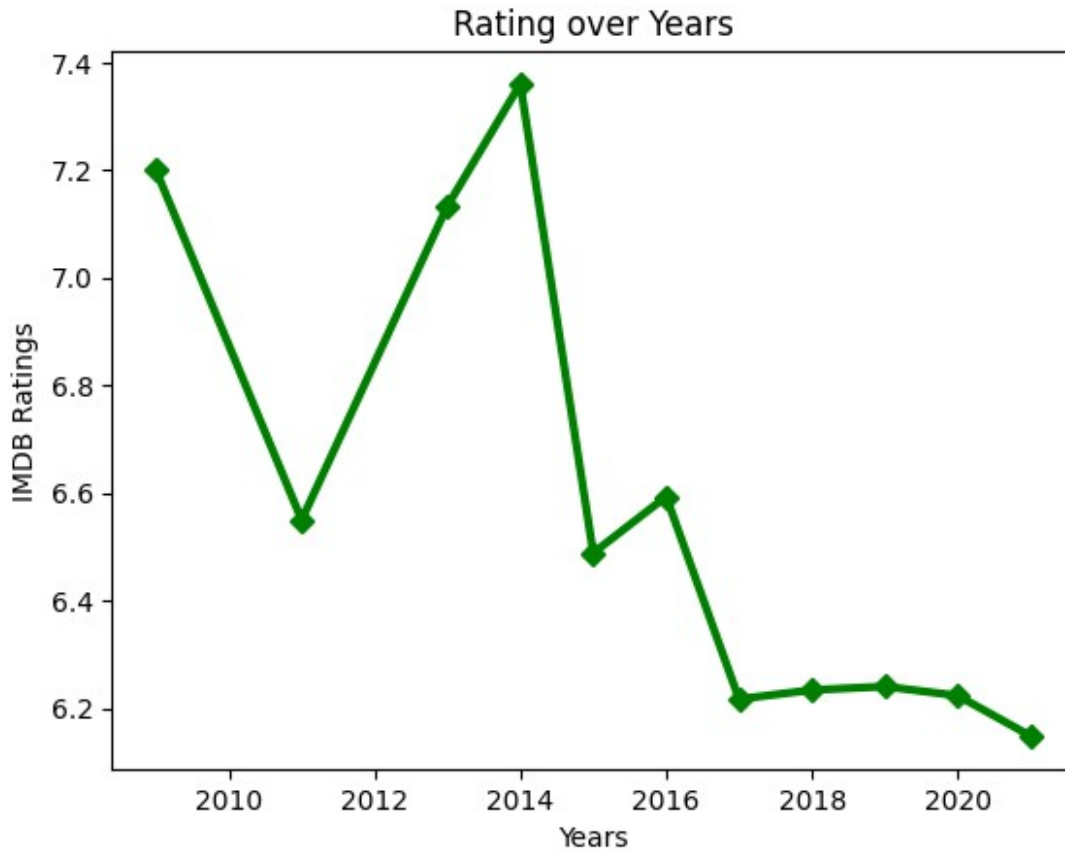
array([2009, 2011, 2013, 2014, 2015, 2016, 2017, 2018, 2019, 2020,
       2021],
      dtype=int32)

plt.xlabel("Years")
plt.ylabel("IMDB Ratings")

```

```
plt.title("Rating over Years")
plt.plot(years1,ratings1,color='green',linewidth=3,marker='D')

[<matplotlib.lines.Line2D at 0x1f8199125d0>]
```



## No. of movies by Duration

```
duration = df['duration'].unique()
duration
```

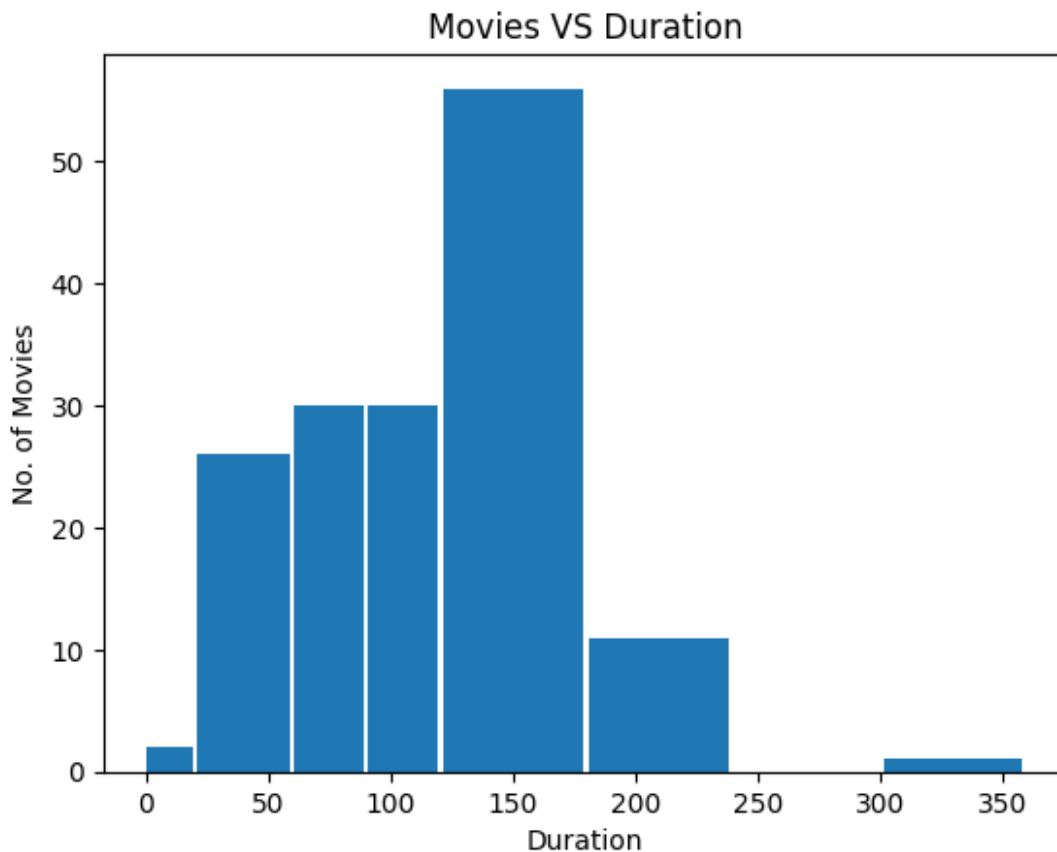
```
array([ 90., 125., 104., 166., 103.,  97.,  96., 102., 182., 106.,
        95.,  92.,  80., 128., 143., 114.,  94., 115., 121., 113.,  82.,
       109.,  88., 229.,  83., 105., 156., 129., 107., 127., 135., 124.,
       136., 165., 150., 133., 142.,  93., 116., 145., 108., 110.,  64.,
       70.,  98.,  86., 111., 100.,  69., 118.,  99., 148.,  91., 101.,
       81., 117., 120.,  89., 112.,  87., 119., 122.,  84.,  77.,  61.,
       72., 132.,  85., 131., 140., 149.,  76., 147., 141., 123.,  65.,
```

```

130., 78., 164., 173., 60., 139., 22., 58., 134., 36., 162.,
138., 52., 71., 161., 14., 126., 54., 8., 79., 48., 68.,
75., 151., 50., 44., 73., 28., 146., 137., 63., 157., 155.,
24., 33., 203., 46., 62., 55., 177., 67., 190., 180., 144.,
57., 174., 170., 192., 209., 187., 185., 160., 186., 59., 169.,
154., 29., 172., 56., 40., 152., 171., 66., 176., 74., 30.,
53., 49., 168., 42., 312., 158., 159., 214., 163., 51., 47.,
153., 26., 45.])

plt.hist(duration,bins=[0,20,60,90,120,180,240,300,360],rwidth=0.95)
plt.xlabel("Duration")
plt.ylabel("No. of Movies")
plt.title("Movies VS Duration")
plt.show()

```





## Category by rating

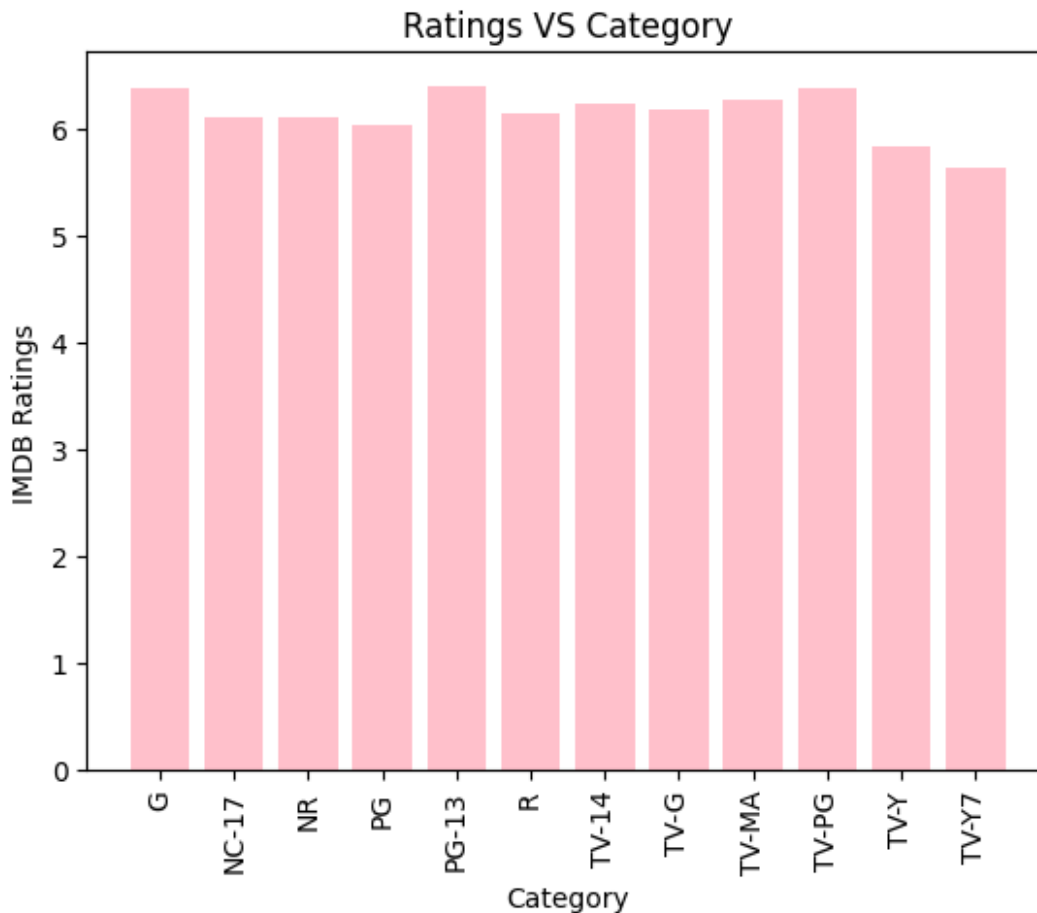
```
ratings2 = df.groupby('rating')['imdb_score'].mean()  
ratings2
```

```
rating  
G      6.380000  
NC-17   6.100000  
NR      6.100000  
PG      6.032857  
PG-13   6.391803  
R       6.147556  
TV-14   6.223883  
TV-G    6.169767  
TV-MA   6.263767  
TV-PG   6.380519  
TV-Y    5.827273  
TV-Y7   5.625000  
Name: imdb_score, dtype: float64
```

```
category = df['rating'].unique()  
category1 = np.sort(category)  
category1
```

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

```
plt.xlabel("Category")  
plt.ylabel("IMDB Ratings")  
plt.title("Ratings VS Category")  
plt.bar(category1, ratings2, color='pink')  
plt.xticks(category, rotation='vertical')  
plt.show()
```



creating a column 'duration\_type'

```
def type_dur(x):  
    if (x > 0 and x <= 90):  
        return "Short Length"  
    elif (x > 90 and x <= 150):  
        return "Medium Length"  
    else:  
        return "Long Length"  
df['duration_type'] = df['duration'].apply(type_dur)  
df
```

C:\Users\as355\AppData\Local\Temp\ipykernel\_5836\1912053336.py:8:

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](https://pandas.pydata.org/pandas-docs/stable/user_guide/indexing.html#returning-a-view-versus-a-copy)

```
df['duration_type'] = df['duration'].apply(type_dur)
```

	show_id	type	title		director	\
0	s1	Movie	Dick Johnson	Is Dead	Kirsten Johnson	
2	s8	Movie		Sankofa	Haile Gerima	
3	s10	Movie		The Starling	Theodore Melfi	
10	s25	Movie		Jeans	S. Shankar	
12	s28	Movie		Grown Ups	Dennis Dugan	
...	...	...		...	...	
5914	s8512	Movie		The Social Network	David Fincher	
5927	s8528	Movie	The Texas Chainsaw	Massacre	Tobe Hooper	
6013	s8636	Movie		True Grit	Henry Hathaway	
6076	s8707	Movie		We the Animals	Jeremiah Zagar	
6116	s8758	Movie		World Trade Center	Oliver Stone	

	cast					\
0	NaN					
2	Kofi Ghanaba, Oyafunmike Ogunlano, Alexandra D...					
3	Melissa McCarthy, Chris O'Dowd, Kevin Kline, T...					
10	Prashanth, Aishwarya Rai Bachchan, Sri Lakshmi...					
12	Adam Sandler, Kevin James, Chris Rock, David S...					
...	...					
5914	Jesse Eisenberg, Andrew Garfield, Justin Timbe...					
5927	Gunnar Hansen, Jim Siedow, Marilyn Burns, Edwi...					
6013	John Wayne, Glen Campbell, Kim Darby, Jeremy S...					
6076	Evan Rosado, Josiah Gabriel, Isaiah Kristian, ...					
6116	Nicolas Cage, Michael Peña, Maggie Gyllenhaal,...					

			country	date_added	\
0			United States	2021-09-25	
2	United States, Ghana, Burkina Faso,		United Kin...	2021-09-24	
3			United States	2021-09-24	
10			India	2021-09-21	
12			United States	2021-09-20	
...			...	...	
5914			United States	2020-04-01	
5927			United States	2019-10-22	
6013			United States	2020-01-01	
6076			United States	2019-02-11	
6116			United States	2019-11-20	

	release_year	rating	duration	\
0	2020	PG-13	90.0	
2	1993	TV-MA	125.0	
3	2021	PG-13	104.0	
10	1998	TV-14	166.0	
12	2010	PG-13	103.0	
...	...	...	...	
5914	2010	PG-13	121.0	
5927	1974	R	83.0	
6013	1969	G	128.0	
6076	2018	R	93.0	

6116            2006   PG-13       129.0

	listed_in \
0	Documentaries
2	Dramas, Independent Movies, International Movies
3	Comedies, Dramas
10	Comedies, International Movies, Romantic Movies
12	Comedies
...	...
5914	Dramas
5927	Cult Movies, Horror Movies, Independent Movies
6013	Classic Movies, Dramas
6076	Dramas, Independent Movies
6116	Action & Adventure, Dramas

	description	imdb_score
year \		
0	As her father nears the end of his life, filmm...	7.4
2021		
2	On a photo shoot in Ghana, an American model s...	7.0
2021		
3	A woman adjusting to life after a loss contend...	6.3
2021		
10	When the father of the man she loves insists t...	6.4
2021		
12	Mourning the loss of their beloved junior high...	5.9
2021		
...	...	...
...		
5914	Director David Fincher's biographical drama ch...	7.8
2020		
5927	When some friends visit a vandalized graveyard...	6.2
2019		
6013	Teenage tomboy Mattie Ross enlists the help of...	7.6
2020		
6076	A trio of brothers cope with their parents' vo...	6.9
2019		
6116	Working under treacherous conditions, an army ...	6.0
2019		

	genre	duration_type
0	Documentaries	Short Length
2	Dramas	Medium Length
3	Comedies	Medium Length
10	Comedies	Long Length
12	Comedies	Medium Length
...	...	...
5914	Dramas	Medium Length
5927	Cult Movies	Short Length
6013	Classic Movies	Medium Length

```
6076          Dramas  Medium Length
6116  Action & Adventure  Medium Length

[2222 rows x 16 columns]
```

## Duration VS Ratings

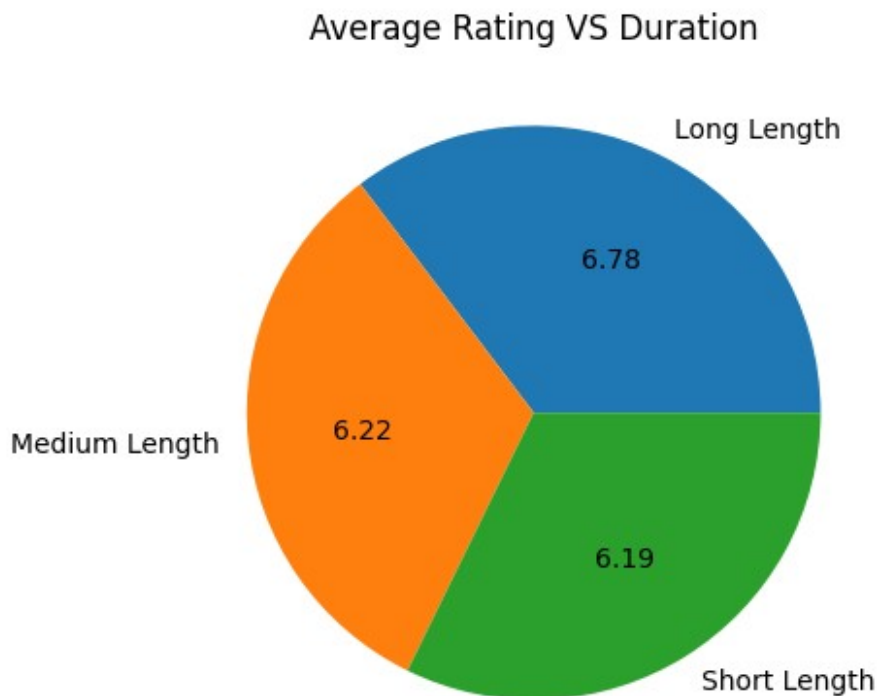
```
Ratings3 = df.groupby('duration_type')['imdb_score'].mean()
Ratings3

duration_type
Long Length      6.780000
Medium Length    6.221358
Short Length     6.188760
Name: imdb_score, dtype: float64

duration = df['duration_type'].unique()
duration1 = np.sort(duration)
duration1

array(['Long Length', 'Medium Length', 'Short Length'], dtype=object)

plt.title("Average Rating VS Duration")
plt.pie(Ratings3, labels=duration1, autopct=lambda x:
'{:.2f}'.format(x*sum(Ratings3)/100))
plt.show()
```



```
df.to_csv('C:/Users/as355/Downloads/cleaned_netflix_data.csv')

df.drop_duplicates(subset=['show_id','type','title','director','cast',
'country','date_added','release_year','rating','duration','listed_in',
'description'],keep='first',inplace=True)
df
```

C:\Users\as355\AppData\Local\Temp\ipykernel\_5836\4287186100.py:1:  
SettingWithCopyWarning:  
A value is trying to be set on a copy of a slice from a DataFrame

See the caveats in the documentation:  
[https://pandas.pydata.org/pandas-docs/stable/user\\_guide/indexing.html#returning-a-view-versus-a-copy](https://pandas.pydata.org/pandas-docs/stable/user_guide/indexing.html#returning-a-view-versus-a-copy)

```
df.drop_duplicates(subset=['show_id','type','title','director','cast',
'country','date_added','release_year','rating','duration','listed_in',
'description'],keep='first',inplace=True)
```

	show_id	type		title	director \
0	s1	Movie	Dick Johnson	Is Dead	Kirsten Johnson
2	s8	Movie		Sankofa	Haile Gerima
3	s10	Movie		The Starling	Theodore Melfi
10	s25	Movie		Jeans	S. Shankar
12	s28	Movie		Grown Ups	Dennis Dugan
...	...	...		...	...
5914	s8512	Movie		The Social Network	David Fincher
5927	s8528	Movie	The Texas Chainsaw	Massacre	Tobe Hooper
6013	s8636	Movie		True Grit	Henry Hathaway
6076	s8707	Movie		We the Animals	Jeremiah Zagar
6116	s8758	Movie		World Trade Center	Oliver Stone

		cast \
0		NaN
2	Kofi Ghanaba, Oyafunmike Ogunlano, Alexandra D...	
3	Melissa McCarthy, Chris O'Dowd, Kevin Kline, T...	
10	Prashanth, Aishwarya Rai Bachchan, Sri Lakshmi...	
12	Adam Sandler, Kevin James, Chris Rock, David S...	
...	...	...
5914	Jesse Eisenberg, Andrew Garfield, Justin Timbe...	
5927	Gunnar Hansen, Jim Siedow, Marilyn Burns, Edwi...	
6013	John Wayne, Glen Campbell, Kim Darby, Jeremy S...	
6076	Evan Rosado, Josiah Gabriel, Isaiah Kristian, ...	
6116	Nicolas Cage, Michael Peña, Maggie Gyllenhaal,...	

		country	date_added \
0		United States	2021-09-25
2	United States, Ghana, Burkina Faso,	United Kin...	2021-09-24
3		United States	2021-09-24
10		India	2021-09-21

12		United States	2021-09-20
...		...	...
5914		United States	2020-04-01
5927		United States	2019-10-22
6013		United States	2020-01-01
6076		United States	2019-02-11
6116		United States	2019-11-20

	release_year	rating	duration	\
0	2020	PG-13	90.0	
2	1993	TV-MA	125.0	
3	2021	PG-13	104.0	
10	1998	TV-14	166.0	
12	2010	PG-13	103.0	
...	...	...	...	
5914	2010	PG-13	121.0	
5927	1974	R	83.0	
6013	1969	G	128.0	
6076	2018	R	93.0	
6116	2006	PG-13	129.0	

	listed_in	\
0	Documentaries	
2	Dramas, Independent Movies, International Movies	
3	Comedies, Dramas	
10	Comedies, International Movies, Romantic Movies	
12	Comedies	
...	...	
5914	Dramas	
5927	Cult Movies, Horror Movies, Independent Movies	
6013	Classic Movies, Dramas	
6076	Dramas, Independent Movies	
6116	Action & Adventure, Dramas	

	description	imdb_score
year	\	
0	As her father nears the end of his life, filmm...	7.4
2021		
2	On a photo shoot in Ghana, an American model s...	7.0
2021		
3	A woman adjusting to life after a loss contend...	6.3
2021		
10	When the father of the man she loves insists t...	6.4
2021		
12	Mourning the loss of their beloved junior high...	5.9
2021		
...	...	...
...		
5914	Director David Fincher's biographical drama ch...	7.8
2020		

5927	When some friends visit a vandalized graveyard...	6.2
2019		
6013	Teenage tomboy Mattie Ross enlists the help of...	7.6
2020		
6076	A trio of brothers cope with their parents' vo...	6.9
2019		
6116	Working under treacherous conditions, an army ...	6.0
2019		

	genre	duration_type
0	Documentaries	Short Length
2	Dramas	Medium Length
3	Comedies	Medium Length
10	Comedies	Long Length
12	Comedies	Medium Length
...	...	...
5914	Dramas	Medium Length
5927	Cult Movies	Short Length
6013	Classic Movies	Medium Length
6076	Dramas	Medium Length
6116	Action & Adventure	Medium Length

[2197 rows x 16 columns]