

# amazon-prime-titles

August 15, 2023

```
[2]: import warnings
warnings.filterwarnings('ignore')
import pandas as pd
import numpy as np
import matplotlib.pyplot as plt
import seaborn as sns
%matplotlib inline
```

```
[3]: df = pd.read_csv("amazon_prime_titles.csv")
df.head()
```

```
[3]: show_id  type      title      director
0      s1  Movie  The Grand Seduction  Don McKellar \
1      s2  Movie  Take Care Good Night  Girish Joshi
2      s3  Movie  Secrets of Deception  Josh Webber
3      s4  Movie  Pink: Staying True  Sonia Anderson
4      s5  Movie  Monster Maker  Giles Foster

                                cast      country \
0  Brendan Gleeson, Taylor Kitsch, Gordon Pinsent  Canada \
1  Mahesh Manjrekar, Abhay Mahajan, Sachin Khedekar  India
2  Tom Sizemore, Lorenzo Lamas, Robert LaSardo, R...  United States
3  Interviews with: Pink, Adele, Beyoncé, Britney...  United States
4  Harry Dean Stanton, Kieran O'Brien, George Cos...  United Kingdom

    date_added  release_year  rating  duration      listed_in \
0  March 30, 2021      2014    NaN  113 min  Comedy, Drama \
1  March 30, 2021      2018    13+  110 min  Drama, International
2  March 30, 2021      2017    NaN   74 min  Action, Drama, Suspense
3  March 30, 2021      2014    NaN   69 min  Documentary
4  March 30, 2021      1989    NaN   45 min  Drama, Fantasy

                                description
0  A small fishing village must procure a local d...
1  A Metro Family decides to fight a Cyber Crimin...
2  After a man discovers his wife is cheating on ...
3  Pink breaks the mold once again, bringing her ...
```

4 Teenage Matt Banting wants to work with a famo...

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

```
[4]: (9668, 12)
```

```
df.info()
```

**Q1. Is there is a duplicate records in this dataset ? if yes, then remove them**

```
[5]: df.duplicated().sum()      # if we found any duplicate date then will use "df.  
    ↪drop_duplicates(inplace=True)"
```

```
[5]: 0
```

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

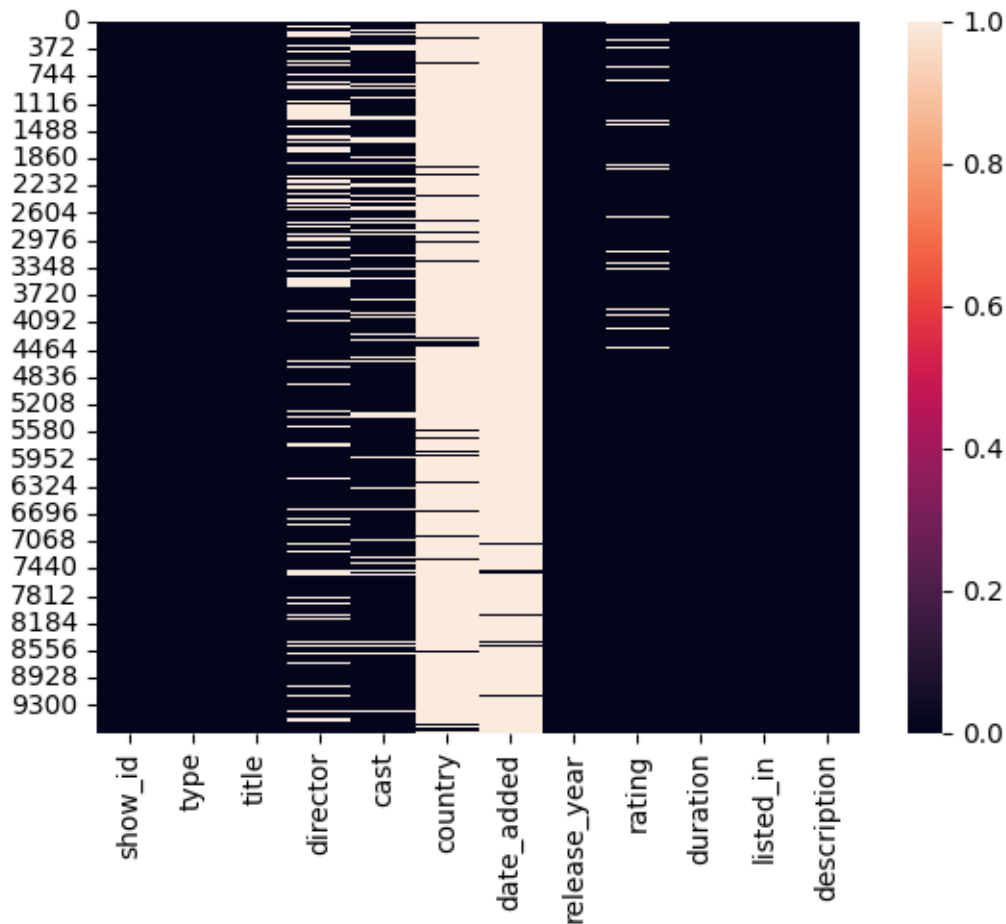
```
[6]: (9668, 12)
```

**Q2. Is there is a null values in this data ? If yes then show it in the heatmap**

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

```
[7]: show_id          0  
     type            0  
     title           0  
     director       2083  
     cast           1233  
     country        8996  
     date_added     9513  
     release_year    0  
     rating         337  
     duration        0  
     listed_in       0  
     description     0  
     dtype: int64
```

```
[8]: sns.heatmap(df.isnull())  
     plt.show()
```



Q3. For 'Monster Maker' movie what is the show\_id and who is the director of this movie ?

```
[9]: df.loc[df["title"]=="Monster Maker",["show_id","title","director"]]
```

```
[9]:  show_id      title      director
4      s5  Monster Maker  Giles Foster
```

Q4. In which year highest number of TV shows and movies released ? show it in the bar chart

```
[10]: df.dtypes # release_year column is in int64 format required to be change in date
```

```
[10]: 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
```

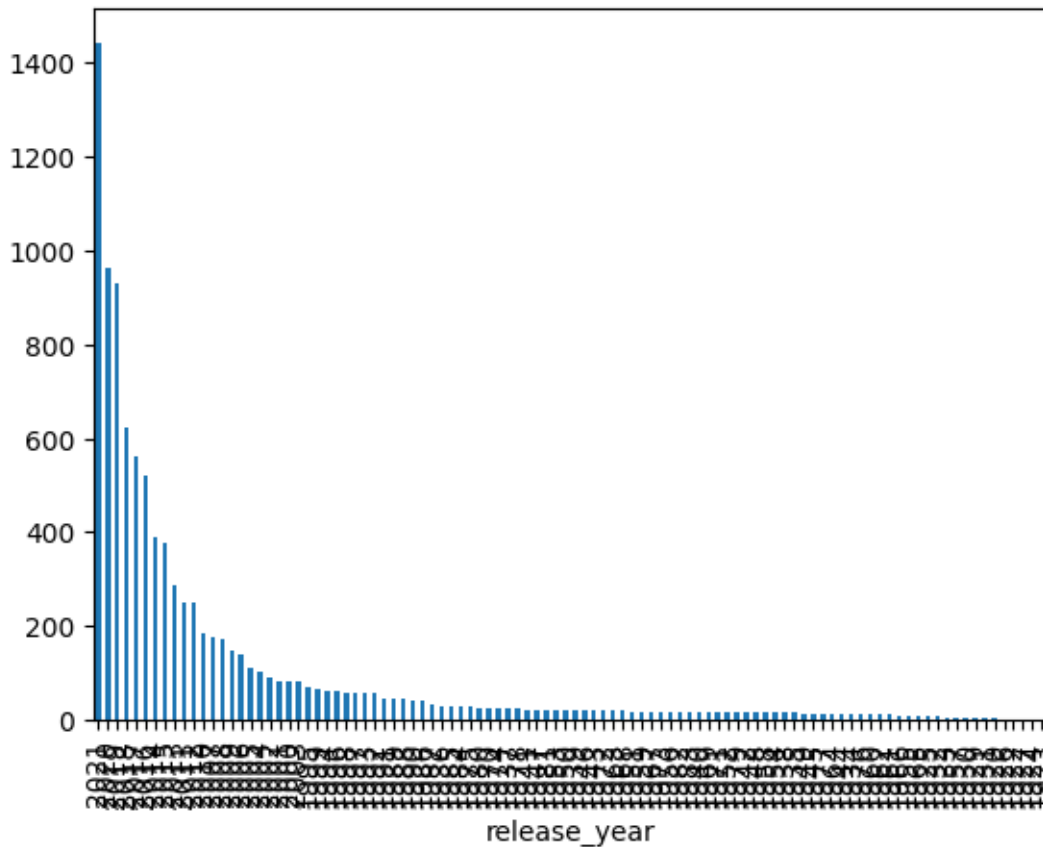
```
[11]: # df.release_year = pd.to_datetime(df.release_year)
      # df.release_year.dtype
```

```
[12]: results = df.groupby(df.release_year)["show_id"].count().
      ↪sort_values(ascending=False)
      results.head()
      # in 2021 highest number of movies and web series are released
```

```
[12]: release_year
      2021    1442
      2020     962
      2019     929
      2018     623
      2017     562
      Name: show_id, dtype: int64
```

```
[45]: df.groupby(df.release_year)["show_id"].count().sort_values(ascending=False).
      ↪plot.bar()
```

```
[45]: <Axes: xlabel='release_year'>
```



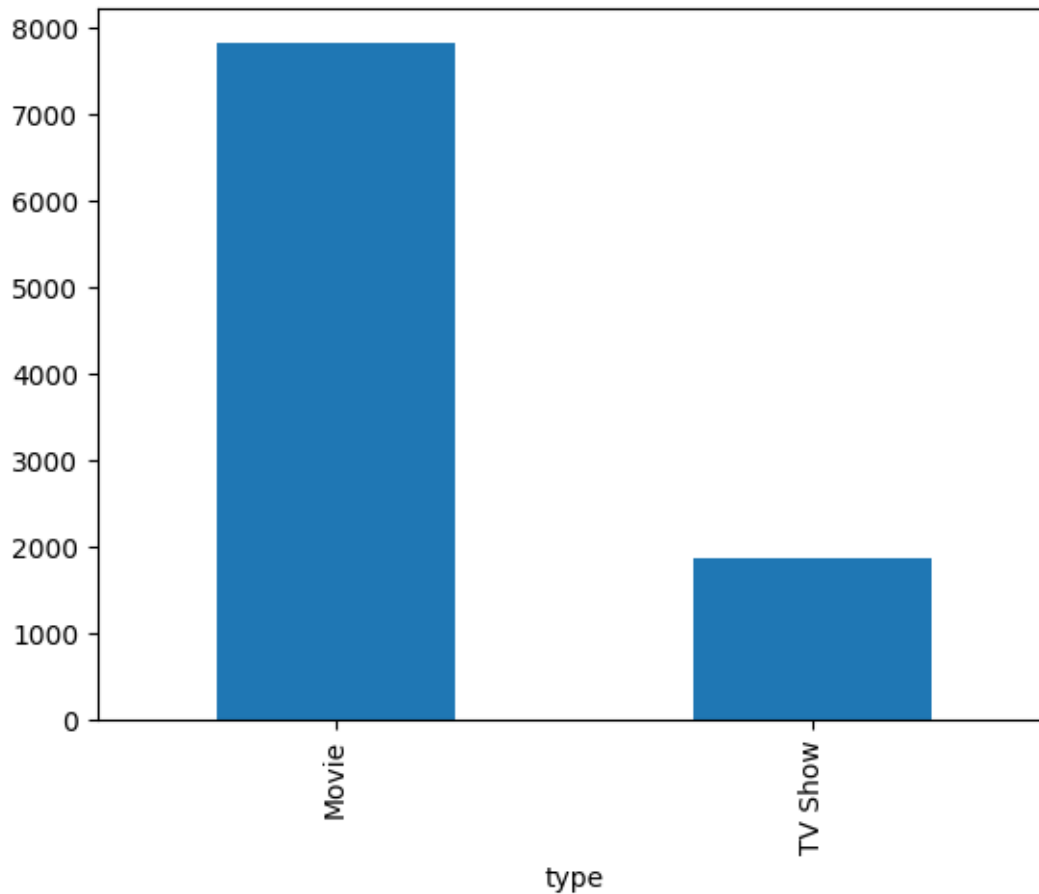
**Q5. How many Movies and TV series are in the dataset ? Show it in bar chart**

```
[14]: df["type"].value_counts()
```

```
[14]: type
      Movie      7814
      TV Show   1854
      Name: count, dtype: int64
```

```
[15]: df["type"].value_counts().plot.bar()
      # Movies are 3X as compare to webseries because web series trend has emerged
      # in recent 2-3 years
```

```
[15]: <Axes: xlabel='type'>
```



**Q6. How many 'movies' are made in year 2020 ?**

```
[16]: # to check both TV Shows and Movies
len(df[(df["type"].isin(["Movie", "TV Show"])) & (df["release_year"] == 2020)])
```

[16]: 962

```
[17]: # to check only movies made in 2020
print(len(df[(df["type"]=="Movie") & (df["release_year"]==2020)]), "'movies' are_
↳made in year 2020")
```

736 'movies' are made in year 2020

**Q7. How many Titles are released in India only ?**

```
[18]: print(len(df[df["country"]=="India"]), "Titles are released in India")
```

229 Titles are released in India

**Q8. Top 10 directors who gave highest number of movies or web series**

```
[19]: df["director"].value_counts().head(10)
```

```
[19]: director
Mark Knight          113
Cannis Holder        61
Moonbug Entertainment 37
Jay Chapman          34
Arthur van Merwijk   30
Manny Rodriguez      22
John English         20
1                    16
Brian Volk-Weiss     15
Baeble Music         14
Name: count, dtype: int64
```

**Q9.** Show the records where ‘category is movie’ and ‘type is comedy’ or ‘country is UK’

```
[20]: df[(df.type=="Movie") & (df["type"].str.contains("Comedy", case=False) |
↪(df["country"]=="United Kingdom"))].head(15)
```

```
[20]:      show_id  type      title
4         s5  Movie  Monster Maker \
5         s6  Movie  Living With Dinosaurs
14        s15  Movie  Elon Musk: The Real Life Iron Man
374       s375  Movie  The Zombie King
656       s657  Movie  The Flaw
892       s893  Movie  Stick Man
2095      s2096  Movie  Instant Death
2364      s2365  Movie  Get Duked!
2381      s2382  Movie  Gascoigne
2534      s2535  Movie  Everybody's Game
2809      s2810  Movie  Coldplay: A Head Full Of Dreams
3140      s3141  Movie  Bastille - ReOrchestrated
3258      s3259  Movie  Andy Murray: Resurfacing
3443      s3444  Movie  9/11 Phone Calls from the Towers
3955      s3956  Movie  Peterloo
```

```
      director
4         Giles Foster \
5         Paul Weiland
14        Sonia Anderson
374       Aidan Belizaire
656       David Sington
892      Jeroen Jaspaert, Daniel Snaddon
2095       Ara Paiaya
2364       Ninian Doff
2381       Jane Preston
```

2534	Beno Obano
2809	Mat Whitecross
3140	Tom Middleton, Toby L
3258	Olivia Cappuccini
3443	James Kent
3955	Mike Leigh

	cast	country
4	Harry Dean Stanton, Kieran O'Brien, George Cos...	United Kingdom \
5	Gregory Chisholm, Juliet Stevenson, Brian Hens...	United Kingdom
14	Elon Musk, Per Wimmer, Julie Anderson-Ankenbra...	United Kingdom
374	Edward Furlong, Corey Feldman, George McCluskey	United Kingdom
656	Andrew Luan, Robert Shiller, Louis Hyman	United Kingdom
892	Martin Freeman, Jennifer Saunders, Russell Tov...	United Kingdom
2095	Lou Ferrigno	United Kingdom
2364	Eddie Izzard, Kate Dickie, James Cosmo, Samuel...	United Kingdom
2381	Paul Gascoigne, Vinnie Jones, Gary Lineker, Jo...	United Kingdom
2534	Maro Itoje, Beno Obano, Anthony Watson, Ellis ...	United Kingdom
2809	Guy Berryman, Jonny Buckland, Will Champion, C...	United Kingdom
3140	Bastille	United Kingdom
3258	Andy Murray	United Kingdom
3443	Steven Mackintosh	United Kingdom
3955	Rory Kinnear, Maxine Peake, Neil Bell, Philip ...	United Kingdom

	date_added	release_year	rating	duration	
4	March 30, 2021	1989	NaN	45 min	\
5	March 30, 2021	1989	NaN	52 min	
14	May 2, 2021	2018	NaN	74 min	
374	NaN	2013	16+	96 min	
656	NaN	2011	13+	81 min	
892	NaN	2015	ALL	27 min	
2095	NaN	2017	18+	84 min	
2364	NaN	2020	16+	87 min	
2381	NaN	2015	16+	87 min	
2534	NaN	2020	16+	53 min	
2809	NaN	2018	PG-13	105 min	
3140	NaN	2021	16+	71 min	
3258	NaN	2019	16+	109 min	
3443	NaN	2009	16+	74 min	
3955	NaN	2019	PG-13	154 min	

	listed_in
4	Drama, Fantasy \
5	Fantasy, Kids
14	Documentary
374	Comedy, Horror
656	Special Interest



892	Adventure, Animation, Kids
2095	Action
2364	Adventure, Comedy
2381	Documentary
2534	Documentary
2809	Documentary, Music Videos and Concerts
3140	Documentary, Music Videos and Concerts
3258	Documentary
3443	Documentary, Special Interest
3955	Drama

	description
4	Teenage Matt Banting wants to work with a famo...
5	The story unfolds in a an English seaside town...
14	Discover the meteoric rise of Elon Musk, the m...
374	After heartbroken widower Samuel makes a deal ...
656	The definitive doc on the 2008 financial crash...
892	A half-hour animation based on the much-loved ...
2095	A gang war for drug dominance draws in a distu...
2364	Get Duked is an anarchic, hip-hop inspired com...
2381	Celebrating the life and achievements of natur...
2534	Coming from the most unlikely of rugby playing...
2809	Long-term Coldplay collaborator Mat Whitecross...
3140	A bracingly honest new documentary sourced fro...
3258	A documentary film that tracks the tennis star...
3443	September 11th 2001 is the most terrifying and...
3955	Included with your Prime membership. Internati...

**Q10. In how many movies / tv shows “salman khan” was casted ?**

```
[21]: len(df[df["cast"].str.contains("salman khan", case=False, na=False)])
```

```
[21]: 10
```

**Q11. What are the different Ratings provided by amazon prime ?**

```
[22]: print(df["rating"].unique())
```

```
[nan '13+' 'ALL' '18+' 'R' 'TV-Y' 'TV-Y7' 'NR' '16+' 'TV-PG' '7+' 'TV-14'
'TV-NR' 'TV-G' 'PG-13' 'TV-MA' 'G' 'PG' 'NC-17' 'UNRATED' '16' 'AGES_16_'
'AGES_18_' 'ALL_AGES' 'NOT_RATE']
```

**Q12. How many movies got the ‘AGES\_18\_’ rating in ‘india’ ?**

```
[23]: len(df[(df["rating"]=="18+") & (df["country"]=="India")])
```

```
[23]: 20
```

**Q13. How many movies get ‘16+’ rating in year ‘2019’ ?**

```
[24]: len(df[(df["rating"]=="16+") & (df["release_year"]==2019)])
```

```
[24]: 142
```

**Q14. What is the max duration of Movie/Show in this dataset ?**

```
[25]: df.duration.value_counts().head()
```

```
[25]: duration
1 Season      1342
2 Seasons      229
90 min         223
91 min         220
92 min         203
Name: count, dtype: int64
```

```
[26]: df[["Minutes","Unit"]] = df["duration"].str.split(" ", expand=True)
df[["Minutes","Unit"]].head()
```

```
[26]:   Minutes Unit
0      113 min
1      110 min
2       74 min
3       69 min
4       45 min
```

```
[29]: df["Minutes"].dtype
```

```
[29]: dtype('O')
```

```
[31]: df["Minutes"] = df["Minutes"].astype("int64")
df["Minutes"].dtype
```

```
[31]: dtype('int64')
```

```
[32]: print("The max duration in this dataset is ---->>>>",df["Minutes"].
      ↪max(),"Minutes")
```

The max duration in this dataset is ---->>>> 601 Minutes

**Q15. Which Individual Country has maximum number of TV Shows ?**

```
[34]: df_tv = df[df["type"]=="TV Show"]
df_tv["country"].value_counts().sort_values(ascending=False).head(3)
# United States has most number of TV Shows
```

```
[34]: country
United States    59
```

```
India          16
Spain          5
Name: count, dtype: int64
```

**Q16. Find out the instances where category is ‘movie’ and type is ‘drama’ ?**

```
[39]: df[(df["type"]=="Movie") & (df["listed_in"]=="Drama")].head(5)
```

```
[39]:
```

	show_id	type	title	director
9	s10	Movie	David's Mother	Robert Allan Ackerman \
24	s25	Movie	Zandalee	Sam Pillsbury
32	s33	Movie	Young Lovers	Ida Lupino
36	s37	Movie	You're Not You	George C. Wolfe
73	s74	Movie	Words On Bathroom Walls	Thor Freudenthal

	cast	country
9	Kirstie Alley, Sam Waterston, Stockard Channing	United States \
24	Nicolas Cage, Judge Reinhold, Erika Anderson, ...	NaN
32	Sally Forrest, Keefe Brasselle, Hugh O'Brian, ...	NaN
36	Hilary Swank, Emmy Rossum, Josh Duhamel	NaN
73	Charlie Plummer, Taylor Russell, Andy Garcia, ...	NaN

	date_added	release_year	rating	duration	listed_in
9	April 1, 2021	1994	NaN	92 min	Drama \
24	NaN	1991	R	94 min	Drama
32	NaN	1949	18+	86 min	Drama
36	NaN	2015	R	103 min	Drama
73	NaN	2020	PG-13	111 min	Drama

	description	Minutes	Unit
9	Sally Goodson is a devoted mother to her autis...	92	min
24	Zandalee has grown bored with her poet-turned-...	94	min
32	Talented and beautiful Carol Williams (Forrest...	86	min
36	A drama centered on a classical pianist who ha...	103	min
73	Diagnosed with a mental illness, a witty, intr...	111	min

```
[ ]:
```