



Amazon Prime Video

Data Analytics

Presented By
Ade Ariyo Yudanto

Data Understanding & Defining Problems

In this analysis, I will explore the Amazon Prime Movies and TV Shows dataset from Kaggle. This tabular dataset consists of listings of all the movies and TV shows available on Amazon Prime, along with details such as - cast, directors, ratings, release year, duration, etc. From all the details of the dataset, I want to know and get an insight about the:

1. How many shows by type of show?
2. How many shows by release year of the show?
3. Which rating has the most shows?
4. What movie has the longest duration?
5. What TV Show has the most seasons?
6. What is the average duration of a Movie by the top 10 directors with the most movies?

Amazon Prime Movies and TV Shows

Movies and TV Shows listings on Amazon Prime Video



[Data Card](#) [Code \(29\)](#) [Discussion \(2\)](#)

About Dataset

Amazon Prime Video - Movies and TV Shows

About this Dataset: Amazon Prime is another one of the most popular media and video streaming platforms. They have close to 10000 movies or tv shows available on their platform, as of mid-2021, they have over 200M Subscribers globally. This tabular dataset consists of listings of all the movies and tv shows available on Amazon Prime, along with details such as - cast, directors, ratings, release year, duration, etc.*

Usability ⓘ
10.00

License
[CC0: Public Domain](#)

Expected update frequency
Monthly

Tags

Tools



Google
Sheets



Looker

Google Sheets is used to clean, enhance, analyze, and visualize data. While Looker is also used for data visualization.

Data Cleaning

Remove Duplicate Data

	A	B	C	D	E	F	G	H	I	J	K	L	M	N
1	show_id	type	title	director	cast	country	date_added	release_year	rating	duration	listed_in	description		
2	s1	Movie	The Grand Sedu	Don McKellar	Brendan Gleeson	Canada	March 30, 2021	2014	113 min		Comedy, Drama	A small fishing village must procure a local doctor t		
3	s2	Movie	Take Care Good	Gresh Joshi	Malayesh Manjprek	India	March 30, 2021	2019	110 min		Drama, Internati	A Metro Family decides to fight a Cyber Criminal th		
4	s3	Movie	Secrets of Cereq	Josh Weidber	Tom Sciorone	United States	March 30, 2021	2017	74 min		Action, Drama	S After a man discovers his wife is cheating on him w		
5	s4	Movie	Pink: Staying Tn	Sonia Anderson	Interviews with	United States	March 30, 2021	2014	69 min		Documentary	Pink breaks the mold once again, bringing her care		
6	s5	Movie	Monster Maker	Giles Foster	Harry Dean Star	United Kingdom	March 30, 2021	1989	45 min		Drama, Fantasy	'Teenage Matt Bunting wants to work with a famous		
7	s6	Movie	Living With Dino	Paul Welland	Gregory Chishol	United Kingdom	March 30, 2021	1989	52 min		Fantasy, Kids	The story unfolds in a an English seaside town, wh		
8	s7	Movie	Hired Gun	Fran Shore	Alice Cooper, Li	United States	March 30, 2021	2017	98 min		Documentary, Si	They are the "First Call, A-las" musicians, just 20 h		
9	s8	Movie	Grease Live	Thomas Kail	Aki Julianne Hough, United States	March 30, 2021	2016	131 min			Comedy	This honest, uncompromising comedy chronicles t		
10	s9	Movie	Global Meltown	Daniel Gilboy	Michael Paré	Le Canada	March 30, 2021	2017	87 min		Action, Science I	A helicopter pilot and an environmental scientist lei		
11	s10	Movie	Drow's Mother	Robert Allan Ack	Kristie Arley, San United States	April 1, 2021	1994	92 min			Drama	Sally Goodson is a devoted mother to her autistic s		
12	s11	Movie	Forest Faires	Justin G. Dyck	Emily Wilder, Ad Canada	April 4, 2021	2016	88 min			Adventure, Kids	Amanda stumbles upon a hidden village of faeries i		
13	s12	Movie	Take Care	Lu Tuccillo	Leslie Bibb, Kev United States	April 10, 2021	2014	90 min			Comedy	When a car crash leaves 'Pammi immobilized, she		
14	s13	Movie	The Night Eats T	Dominique Roch	Andres Danielle	France	April 17, 2021	2016	94 min		Horror, Suspens	After waking up in an apartment the night after a		
15	s14	Movie	Resilencia	Jep Barcelona	Rafinha Alcantar	Spain	April 24, 2021	2020	49 min		Documentary, Si	The documentary follows the midlife's everyday		
16	s15	Movie	Elon Musk: The	Sonia Anderson	Elon Musk, Per United Kingdom	May 2, 2021	2018	74 min			Documentary	Discover the meteoric rise of Elon Musk, the man v		
17	s16	Movie	Summer '03	Becca Gleason	Joey King, Jack United States	June 3, 2021	2019	96 min			Comedy, Drama	'Summer '03 is the hilarious story of 15-year-old J		
18	s17	Movie	Zoombies	Gleem Miller	Marcus Anderson, Kaiwi Lyman, Andrew Asper	2016	13+	87 min			Horror, Science I	When a strange virus quickly spreads through a sa		
19	s18	TV Show	Zoo Babies		Narrator - Gillian Barlett	2008	ALL	1 Season			Kids, Special In	A heart warming and inspiring series that welcome		
20	s19	TV Show	Zoe Coombs Mar	Bossey Bottom	Zoe Coombs Mar	2020	19+	1 Season			Comedy, Talk Sh	Zoe Coombs Mar has been on hiatus. Sort of. For		
21	s20	Movie	Zoe	Drake Doremus	Ewan McGregor, Léa Seydoux, Thoe James, Mirar	2018	R	104 min			Science Fiction	ZOE tells a tale of forbidden love between an engi		
22	s21	TV Show	Zoboomfao			2001	TV-Y	1 Season			Kids	Chris and Martin Kutz bring their enthusiasm for		
23	s22	TV Show	Zo Zo Zombe	Mini-Series	Alex Cazares, Ben Diskin, Cristina Vee, Erika Ishi	2020	TV-Y7	1 Season			Action, Anime	C Z Zo Zombe is our undead hero who, with his be		
24	s23	Movie	Zis Boom Bah	William High	Grace Hayes, May	Healy	1941	18	62 min		Arts, Entertainm	'Hey kids, let's get together and put on a show!" T		
25	s24	TV Show	ZoroZeroZero		ANDREA RISEB	Italy	2020	19+	1 Season		TV Shows	ZoroZeroZero series has the journey of a cocar		
26	s25	Movie	Zandalee	Sam Pillsbury	Nicolas Cage, Judge Reinhold, Erika Anderson, Jo	1991	R	94 min			Drama	Zandalee has grown bored with her post-teeny-b		
27	s26	Movie	Zambozi: Force	Dir: Rudolf Lammers		2005	ALL	50 min			Documentary	The Zambozi river is the fourth-longest river in Afr		
28	s27	TV Show	Zahara: The Return		Natasha Heschelle, Curtis Morgan, Hannah Scott,	2020	19+	1 Season			TV Shows	The story explores the history of slavery and racis		
29	s28	TV Show	Z: The Beginning of Everything		Christina Ricci	United States	2017	19+	1 Season		Comedy, Drama	"Z: The Beginning of Everything" tells the story of 2		
30	s29	TV Show	Yu-Gi-Oh! Zexal		El Jey Marc Thompson, Sean Schemmel, Eileen C	2015	TV-Y7	3 Seasons			Animation, Aem	It's time to feel the fire as Yu-Gi-Oh! Zexal battles C		
31	s30	TV Show	Yu-Gi-Oh! ARC-V		Michael Liscio Jr., Emily Bauer	2016	TV-Y7	3 Seasons			Action, Adventur	Get set to swing into action as Yu-Gi-Oh! ARC-V ta		

amazon_prime_video_dataset

Count: 9738



	A	B	C	D	E	F	G	H	I	J	K	L	M	N
9640	s9639	Movie	Oven	Kishore Tummali	Lauren Ojeda		2019	7+	2 min		Drama	A young mom's lapse in attention leads to a calam		
9641	s9640	Movie	No Sleep 'Til Chi Phi Trail	Dave Annable, Odette Annable, Sheryl Lee Ralph,		2018	16+	85 min			Drama	Lizzie and Billy only fall asleep next to each other.		
9642	s9641	Movie	Meet the Spartans	Jason Friedberg	Sean Hannity, Carmen Electra, Ken Davitian, Kevi		2008	18+	86 min		Comedy	From the creators of Scary Movie and Date Movie i		
9643	s9642	Movie	Marley & Me	David Frankel	Owen Wilson, J	United States	2008	PG	116 min		Comedy, Drama	Jennifer Aniston and Owen Wilson unleash huge i		
9644	s9643	TV Show	Lunkplunkville		Michael de Avila		2021	13+	2 Seasons		Documentary	St Lunkplunkville's launches "Got Foks Fatsing", a socia		
9645	s9644	Movie	Love & Other Dr	Edward Zwick	K.K. Jake Gyllen	United States	2010	R	112 min		Comedy, Drama	A womanizing pifer drug rep with no sense of dre		
9646	s9645	Movie	Little Miss Sunst	Jonathan Dayton	Abigail Breslin, C	United States	2009	R	101 min		Comedy, Drama	A family determined to get their young daughter in		
9647	s9646	TV Show	Let's Go Cozy Cou	Cartoons for Kids		2021	ALL	1 Season			Kids	Cozy Coupe is an online series that helps kids to l		
9648	s9647	Movie	Knight and Day	20th century, fo	Tom Cruise, Cameron Diaz, Peter Sarsgaard, Jordi		2010	PG-13	117 min		Action, Adventur	SPECIAL EXTENDED EDITION! Get even MORE		
9649	s9648	TV Show	Kids In The Hall				1995	TV-14	5 Seasons		Arts, Entertainm	"The Kids in the Hall" is a unique and bizarre sketch		
9650	s9649	Movie	Judge Dredd	Danny Cannon	Sylvester Stallone, Armand Assante, Rob Schnede		1995	R	96 min		Suspense	Joseph Dredd is convicted for a crime he did not cc		
9651	s9650	Movie	Jack	Francis Ford Co	Robin Williams, Dane Lane, Brian Kernin, Jennifer		1996	PG-13	113 min		Drama, Sports	Robin Williams stars as a 15-year-old in a grown-up		
9652	s9651	Movie	I Can't Breathe	Kishore Tummali	Michele Anderson, Marc, Juan, Christopher David		2017	18+	32 min		Drama, Sports	An African American single mother's emotional jou		
9653	s9652	TV Show	Glass Room		Elizaveta Maskaya, Vsevolod Boldin, Alexander Ki		2019	16+	1 Season		Drama	Geography teacher Elena lives with her husband's		
9654	s9653	Movie	Fandoró	4K UH Pappy Romanoff	Stera Ekbasta, Dipo, Steve Aoki, Russian		2016	G	111 min		Documentary	Musclfin with unpublished images from the life of		
9655	s9654	Movie	Date Night	Exte Shaim Levy	Mark Wahlberg, Steve Carell, Tanaj P. Henson, Tin		2010	UNRATED	102 min		Comedy	A typical suburban couple's date night becomes m		
9656	s9655	Movie	Dard Divorce	Olaf Ittenbach	Marina Ittenbach, Daryl Jackson, Jaymes Butler, B		2008	18+	87 min		Drama, Horror	What started with a routine divorce between Natali		
9657	s9656	TV Show	Defundamentals	Series	Emile Crawford, Randall Crawford, Jon Ward		2021	TV-Y	1 Season		Animation, Kids	Learn "deFundamentals" through song and fun cha		
9658	s9657	Movie	Anaganaga Oka	Vamsi P	Vjey, Sveta chaudhary, NNR CHOWDARY, Hanyia		2021	18+	135 min		Action, Drama	S (Peter)hero) Who went to the forest to find a cure fo		
9659	s9658	Movie	A Place at the Ti	Kristen Carlin	Jeff Bridges, Tom Colicchio		2013	PG	84 min		Documentary, Si	The next chapter on America's food crisis from t		
9660	s9659	Movie	12 Dates of Christmas		Heath Ledger, Julia Stiles, Joseph Gordon-Levitt, L		2011	PG	87 min		Drama, Romance	Setup on a Christmas Eve date, a woman must reli		
9661	s9660	Movie	10 Things I Hate G	Jagger			1999	PG-13	97 min		Comedy, Drama	On the first day of his new school, Cameron instati		
9662	s9661	Movie	The Man in the I	John-Paul David	Caran Hinds, Stephen Dillane, Malvern		2021	13+	96 min		Comedy	The Man in the Hat pursues through France in a		
9663	s9662	TV Show	Temahashes		DENISE BRYER, WINDSOR DAVIES, JEREMY HI		1983	ALL	3 Seasons		TV Shows	This thrilling, action-packed series saw Gerry Andy		
9664	s9663	Movie	River	Emily Slaye	Maya Cameron Rogers, Alexandra Rose, Rob Mar		2021	16+	93 min		Drama, Science River	is a grounded Sci-Fi mystery. Thriller, that fol		
9665	s9664	Movie	Pride Of The Bo	Joseph H. Lewis	Lee Gorcey, Bobby Jordan		1940	7+	60 min		Comedy	New York City street principles get an East Side Ki		
9666	s9665	TV Show	Planet Patrol		DICK VOSSBURGH, RONNIE STEVENS, LIBBY MI		2018	13+	4 Seasons		TV Shows	This is Earth, 2100AD - and these are the adventur		
9667	s9666	Movie	Outpost	Steve Barker	Ray Stevenson, Julian Wadham, Richard Brake, M		2008	R	90 min		Action	In war-torn Eastern Europe, a world-weary group o		
9668	s9667	TV Show	Maradona: Blessed Dream		Esteban Recagno, Ezequiel Streinz, Luciano Vittor		2021	TV-14A	1 Season		Drama, Sports	The series tells the story of Diego Maradona, a boy		
9669	s9668	Movie	Harry Brown	Daniel Barber	Michael Caine, Emily Mortimer, Joseph Gilgun, Sei		2010	R	103 min		Action, Drama	A Harry Brown, starring two-time Academy Award wi		

amazon_prime_video_dataset

Count: 9662

9738 Data Before Cleaned

9662 Data After Cleaned

In this dataset, there are 9738 data which have several duplicates. After duplicate data is removed, there are 9662 unique data.

Data Cleaning

Fixing Title of Data

show_id	type	title
s3569	Movie	#Unfit: The Psychology of Donald Trump
s4353	Movie	#Lucky Number
s5548	Movie	"Mixed Up"
s7557	Movie	#Home
s3570	TV Show	#Lagira de #eldisco
s3571	TV Show	#IMomSoHard Live
s3573	TV Show	_DUPE_The Making of the Mob: Chicago
s5607	TV Show	#IGotThis
s5978	TV Show	"The Paramedic Angel"
s8853	TV Show	#WASHED



show_id	type	title
s3569	Movie	Unfit: The Psychology of Donald Trump
s4353	Movie	Lucky Number
s5548	Movie	Mixed Up
s7557	Movie	Home
s3570	TV Show	Lagira de eldisco
s3571	TV Show	IMomSoHard Live
s3573	TV Show	The Making of the Mob: Chicago
s5607	TV Show	IGotThis
s5978	TV Show	The Paramedic Angel
s8853	TV Show	WASHED

In this dataset, there are 10 data that contain uncommon characters or words as the title of a Movie or TV Show. So, They have to be fixed or cleaned.

Data Formatting

Handle Missing Data

There are several missing data in some columns:

1. 9507 out of 9662 data (98.40%) are empty in column **date_added**. Instead of deleting it, the data is labeled with “Missing Data”.
2. 2098 out of 9662 data (21.70%) are empty or have a number value in column **director**. Instead of deleting it, the data is labeled with “Unknown Director” if the type of show is Movie and labeled with “Missing Data” if the type of show is TV Show.
3. 1268 out of 9662 data (13.12%) are empty or have a number value in column **cast**. Instead of deleting it, the data is labeled with “Unknown Cast” if the type of show is Movie and labeled with “Missing Data” if the type of show is TV Show.
4. 8990 out of 9662 data (93.04%) are empty in column **country**. Instead of deleting it, the data is labeled with “Missing Data”.
5. 335 out of 9662 data (3.47%) are empty in column **rating**. Instead of deleting it, the data is labeled with “Unrated”.
6. 20 out of 9662 data (0.21%) are empty in column **description**. Instead of deleting it, the data is labeled with “Undescribed”.

show_id	type	title	director	cast	country	date_added
s21	TV Show	Zoboomafoo	Unknown Director	Unknown Cast	Missing Data	Missing Data
s34	Movie	Young Love (at the Sun Factory)	Unknown Director	Unknown Cast	Missing Data	Missing Data
s35	Movie	Young Love (at Bowery Ballroom)	Unknown Director	Unknown Cast	Missing Data	Missing Data
s73	TV Show	WordWorld	Unknown Director	Unknown Cast	Missing Data	Missing Data
s75	TV Show	WordGirl	Unknown Director	Unknown Cast	Missing Data	Missing Data
s79	TV Show	Wonder Balls	Unknown Director	Unknown Cast	Missing Data	Missing Data
s99	TV Show	Wildest Places	Unknown Director	Unknown Cast	Missing Data	Missing Data
s101	TV Show	Wild North	Unknown Director	Unknown Cast	Missing Data	Missing Data
s107	TV Show	Why I Ran	Unknown Director	Unknown Cast	Missing Data	Missing Data
s129	TV Show	What's Up In Space: The Solar System For Kids	Unknown Director	Unknown Cast	Missing Data	Missing Data
s135	TV Show	WGC-Dell Technologies Match Play Reveal Sho	Unknown Director	Unknown Cast	Missing Data	Missing Data
s139	TV Show	WellieWishers	Unknown Director	Unknown Cast	Missing Data	Missing Data
s150	TV Show	Watch the Game Like Never Before with X-Ray	Unknown Director	Unknown Cast	Missing Data	Missing Data
s171	TV Show	Waco: Madman or Messiah	Unknown Director	Unknown Cast	Missing Data	Missing Data
s172	TV Show	Voyage of the Continents	Unknown Director	Unknown Cast	Missing Data	Missing Data
s177	TV Show	Vocabulary and Numbers with Harry the Bunny	Unknown Director	Unknown Cast	Missing Data	Missing Data
s195	TV Show	Vella Raja (Tamil)	Unknown Director	Unknown Cast	Missing Data	Missing Data
s196	TV Show	Vella Raja (Hindi)	Unknown Director	Unknown Cast	Missing Data	Missing Data
s198	TV Show	Vatican Miracle Examiner	Unknown Director	Unknown Cast	Missing Data	Missing Data
s209	Movie	USAISAMONSTER	Unknown Director	Unknown Cast	Missing Data	Missing Data
s212	TV Show	UQ Holder!	Unknown Director	Unknown Cast	Missing Data	Missing Data
s248	TV Show	TuTu Town	Unknown Director	Unknown Cast	Missing Data	Missing Data
s261	TV Show	Troy The Train	Unknown Director	Unknown Cast	Missing Data	Missing Data
s268	TV Show	Trapped [English Subtitled]	Unknown Director	Unknown Cast	Missing Data	Missing Data
s299	TV Show	Tom's Paint Shop	Unknown Director	Unknown Cast	Missing Data	Missing Data

Data Transformation

Transform Type of Data

The key column in this analysis for solving the stated problems is **duration**. The duration column has a text value, so it will be converted to a number.

1. Add one column named **duration_extract**.
2. Extract data in column **duration** that only have the word "min" into only number value in **duration_extract** column.
3. Add one column again named **duration_num**.
4. Copy values from column **duration_extract** and paste it into column **duration_num** with paste special "values only".
5. Change the data type of column **duration_num** into a number.

duration	duration_extract	duration_num
96 min	96	96.00
87 min	87	87.00
1 Season	1 Season	1 Season
1 Season	1 Season	1 Season
104 min	104	104.00
1 Season	1 Season	1 Season
1 Season	1 Season	1 Season
62 min	62	62.00
1 Season	1 Season	1 Season
94 min	94	94.00
50 min	50	50.00
1 Season	1 Season	1 Season
1 Season	1 Season	1 Season
3 Seasons	3 Season	3 Season

Data Extraction

Grouping Data

Several Movie and TV Show ratings need to be corrected and adjusted in column **rating** based on the Maturity Rating in the United States which is divided into 2 local value ratings that are MPAA (Movie) and TVPG (TV Show).

1. Add one column named **rating_extract**.
2. Data with values “UNRATED”, “TV-NR”, “NR” and “NOT_RATE” in column **rating** filled with “Unrated” in **rating_extract** column.
3. Data with values “ALL” and “ALL_AGES” in column **rating** filled with “G” if Movie and “TV-G” if TV Show in **rating_extract** column.
4. Data with value “7+” in **rating** column filled with “PG” if Movie and “TV-Y7” if TV Show in column **rating_extract**.
5. Data with values “16”, “16+” and “AGES_16_” in column **rating** filled with “R” if Movie and “TV-14” if TV Show in **rating_extract** column.
6. Data with values “18” and “AGES_18_” in column **rating** filled with “NC-17” if Movie and “TV-MA” if TV Show in **rating_extract** column.
7. After grouping and adjusting the rating of Movies and TV Shows, there are 701 out of 9667 data (7.25%) labeled with “Unrated” in column **rating_extract**.

Data Cleaning

Exclude Outlier

The duration of the Movie needs to be adjusted by identifying whether the film is included in the outlier category or not.

1. Add one column named **outlier**.
2. Divide data using the quartile formula to divide the number of data points into four parts, or quarters, of more or less equal size and determine the lower bound and upper bound. Here is the result:

Q1	1	75.00
Q2	2	91.00
Q3	3	106.00
IQR		31.00
LB		28.50
UP		152.50

→ Lower Bound

→ Upper Bound

1. So, Movies with a duration of less than 30 minutes and more than 152 minutes are identified as outliers.
2. 623 out of 7808 data (7.98%) are identified as outliers. So, there are 7185 Movies that meet the criteria.

Data Extraction

Exclude Missing Data

In the previous handling missing data process, several empty data in some columns were found. That empty data will be excluded before visualized.

1. Movies that are identified as outliers or have missing data/unknown/unrated in column **director**, **cast**, and **rating_extract** will be excluded. 6116 Movies meet the criteria.
2. TV Shows that have missing data/unrated in column **rating_extract** will be excluded. 1714 TV Shows meet the criteria.
3. So, there are 7830 out of 9662 (81.04%) Movies and TV Shows which ready to serve and visualized after simple exploratory data analysis.

Data Visualization

Show by Type of Show

In this dataset, the proportion of Movies is 78.1%. So, Amazon Prime Video needs to take into consideration about to add TV Shows in its database.

Type of Show

Percentage of Show

TV Show

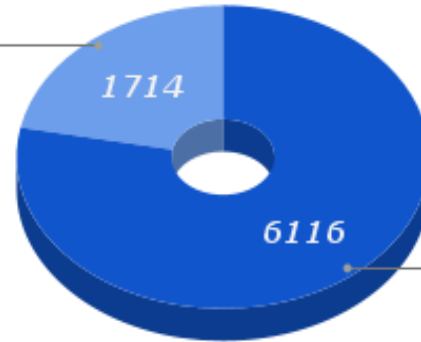
21.9%

1714

6116

Movie

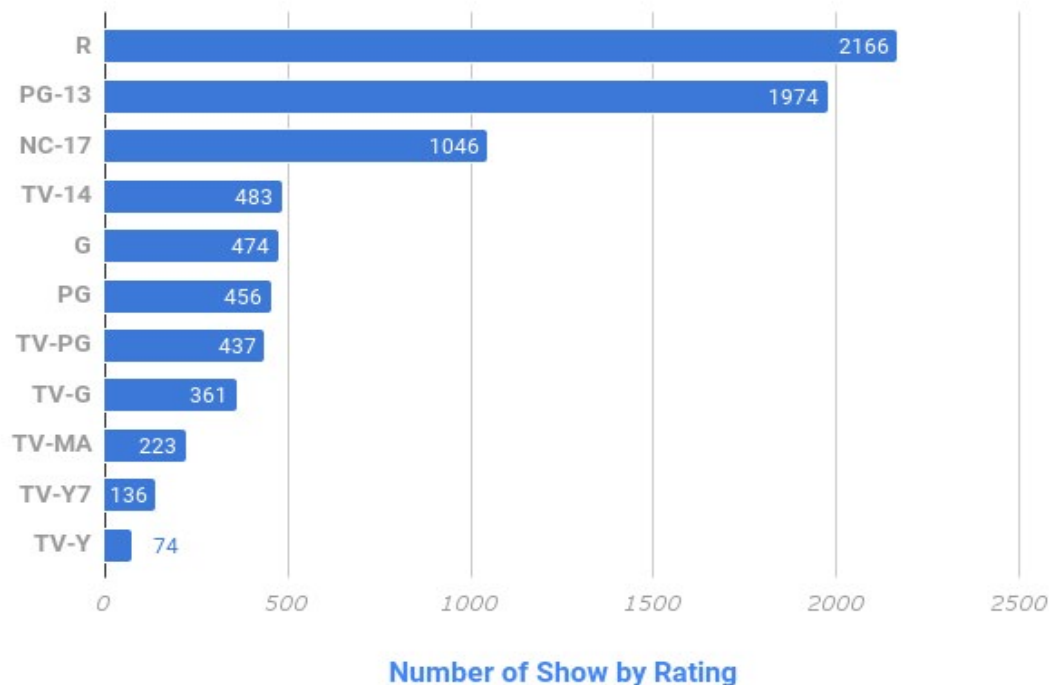
78.1%



Data Visualization

Number of Shows by Rating

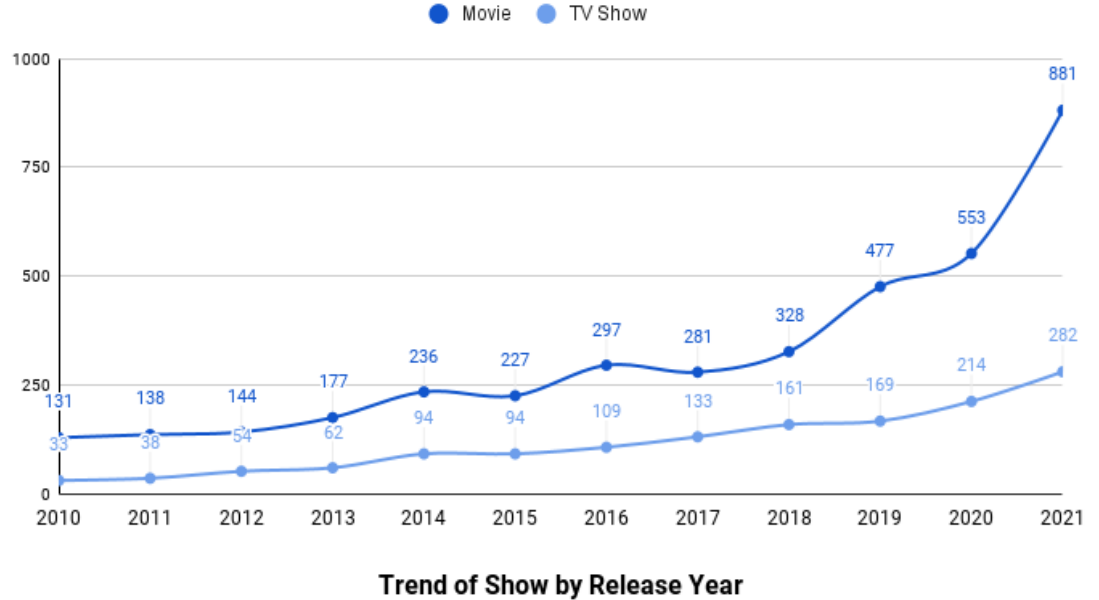
Most shows on Amazon Prime Video have a rating of "R" followed by PG-13. Amazon Prime Video needs to take a look at who its targeted users, are so it can add more Movies or TV Shows to its targeted users.



Data Visualization

Number of Shows by Release Year

The number of shows by release year at Movie significantly increased starting from 2019, although there was a slight decrease in 2015 & 2017. Meanwhile, TV Shows are increasing slowly year by year.



Data Visualization

Longest Duration Movie & Most Seasons TV Show

	Title	Duration (Minutes) ▾
1.	Dhoom:2	152
2.	Manadhai Thiruvittai	152
3.	Lovely	152
4.	Bachna Ae Haseeno	152
5.	Baba Keno Chakar	152
1 - 5 / 6116 < >		

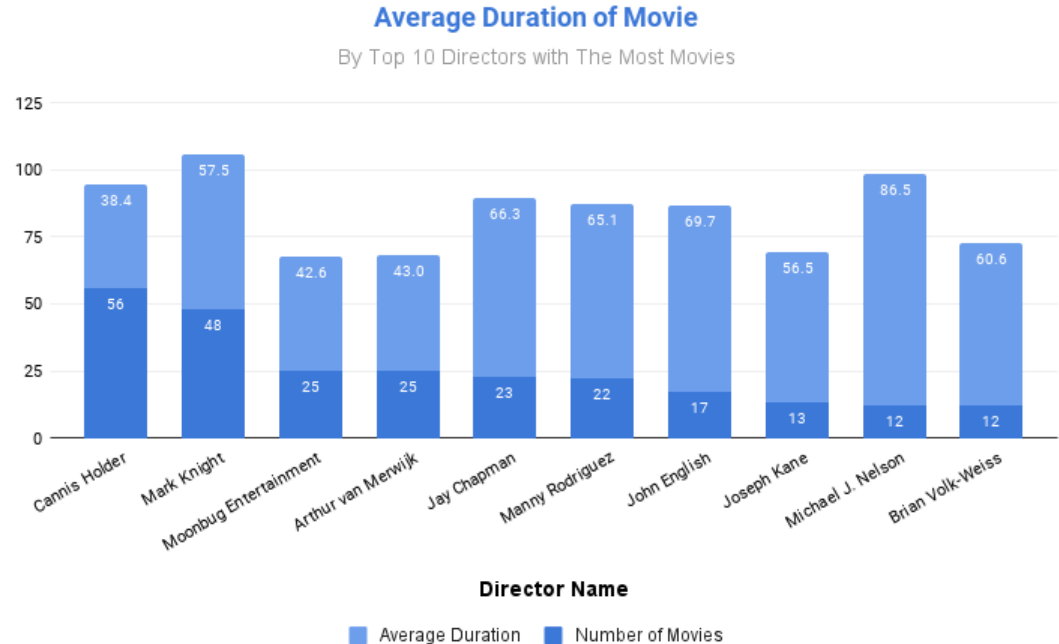
	Title	Number of Seasons ▾
1.	The Amazing Race	29
2.	Silent Witness	21
3.	Survivor	19
4.	Top Gear (UK)	15
5.	Paula's Best Dishes	14
1 - 5 / 1714 < >		

The movie has the longest duration "Dhoom: 2" with a duration of 152 minutes or approximately 2.5 hours. For TV Show, "The Amazing Race" has the most seasons with 29 seasons.

Data Visualization

Average Duration of Movie By Top 10 Directors

Cannis Holder is the most prolific director with 56 movies. However, the average duration of the movie is 38.4 minutes, which means he directs more of the short movies category. On the other hand, Michael J. Nelson just directed 12 Movies (bottom 2 out of 10), but the average duration of the movie is 86.5, highest than others.



MOVIE

6116

Total Count

TV SHOW

1714

Total Count

Longest Duration Movie

Dhoom: 2

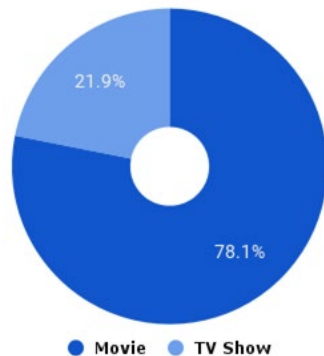
152 Minutes

Most Seasons TV Show

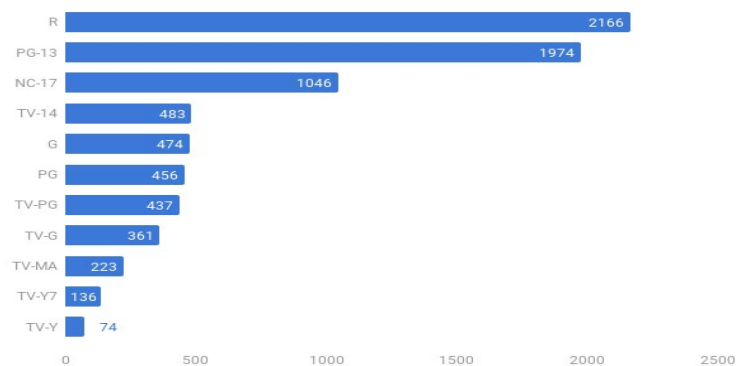
The Amazing Race

29 Seasons

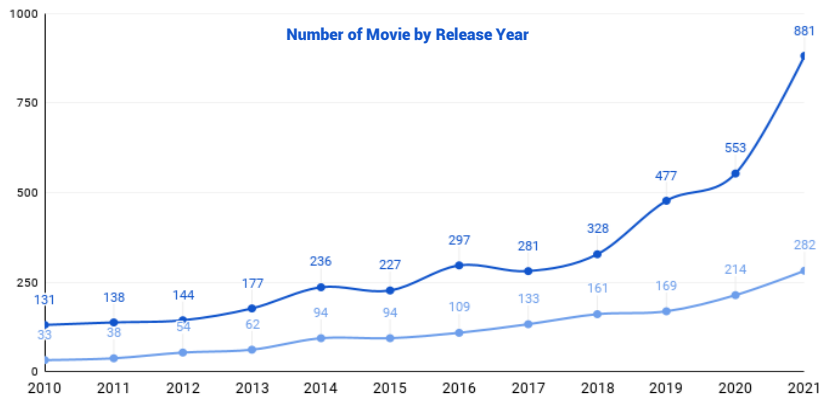
Percentage of Show



Number of Show by Rating



Number of Movie by Release Year



Average Duration of Movie by Top 10 Directors

