

R Notebook

Scraping top bollywood movies list from IMDb

```
library(rvest)
library(dplyr)
```

```
##
## Attaching package: 'dplyr'
```

```
## The following objects are masked from 'package:stats':
##
##   filter, lag
```

```
## The following objects are masked from 'package:base':
##
##   intersect, setdiff, setequal, union
```

```
link="https://www.imdb.com/india/top-rated-indian-movies/"
page = read_html(link)

TopBollywoodMovies = page %>% html_nodes(".titleColumn a") %>% html_text()

movie_links = page %>% html_nodes(".titleColumn a") %>% html_attr("href") %>% paste("https://www.imdb.com/title/",)

yr_of_release = page %>% html_nodes(".secondaryInfo") %>% html_text()

Star = page %>% html_nodes("strong") %>% html_text()

#t = page %>% html_table("table")

top_movies = data.frame(TopBollywoodMovies,yr_of_release,Star)

top_movies
```

```
##           TopBollywoodMovies yr_of_release Star
## 1                Jai Bhim      (2021)  8.7
## 2                 Nayakan      (1987)  8.5
## 3       Pariyerum Perumal      (2018)  8.5
## 4                 Anbe Sivam      (2003)  8.5
## 5       C/o Kancharapalem      (2018)  8.5
## 6                 Sardar Udham      (2021)  8.5
## 7                 Kireedam      (1989)  8.5
## 8       Manichitrathazhu      (1993)  8.5
```

## 9	Golmaal	(1979)	8.5
## 10	Apur Sansar	(1959)	8.5
## 11	Natsamrat	(2016)	8.4
## 12	Thevar Magan	(1992)	8.4
## 13	Kumbalangi Nights	(2019)	8.4
## 14	Black Friday	(2004)	8.4
## 15	96	(2018)	8.4
## 16	Pather Panchali	(1955)	8.4
## 17	#Home	(2021)	8.4
## 18	Soorarai Pottru	(2020)	8.4
## 19	Visaaranai	(2015)	8.4
## 20	Jersey	(2019)	8.4
## 21	3 Idiots	(2009)	8.4
## 22	Asuran	(2019)	8.3
## 23	Taare Zameen Par	(2007)	8.3
## 24	Kaithi	(2019)	8.3
## 25	Thalapathi	(1991)	8.3
## 26	Sarpatta Parambarai	(2021)	8.3
## 27	Drishyam 2	(2021)	8.3
## 28	Devasuram	(1993)	8.3
## 29	Dangal	(2016)	8.3
## 30	Peranbu	(2018)	8.3
## 31	Vada Chennai	(2018)	8.3
## 32	Chithram	(1988)	8.3
## 33	Aparajito	(1956)	8.3
## 34	Jaane Bhi Do Yaaro	(1983)	8.3
## 35	Pyaasa	(1957)	8.3
## 36	Thani Oruvan	(2015)	8.3
## 37	Guide	(1965)	8.3
## 38	Ratsasan	(2018)	8.3
## 39	Spadikam	(1995)	8.3
## 40	Iruvar	(1997)	8.2
## 41	Kannathil Muthamittal	(2002)	8.2
## 42	Agent Sai Srinivasa Athreya	(2019)	8.2
## 43	Chupke Chupke	(1975)	8.2
## 44	Super Deluxe	(2019)	8.2
## 45	Mahanati	(2018)	8.2
## 46	Aruvi	(2016)	8.2
## 47	Pudhu Pettai	(2006)	8.2
## 48	Khosla Ka Ghosla!	(2006)	8.2
## 49	Drishyam	(2013)	8.2
## 50	Anniyan	(2005)	8.2
## 51	Tumbbad	(2018)	8.2
## 52	Premam	(2015)	8.2
## 53	Kaakkaa Muttai	(2014)	8.2
## 54	Mudhalvan	(1999)	8.2
## 55	Vikram Vedha	(2017)	8.2
## 56	Bangalore Days	(2014)	8.2
## 57	Papanasam	(2015)	8.2
## 58	Satya	(1998)	8.2
## 59	Angoor	(1982)	8.2
## 60	Soodhu Kavvum	(2013)	8.2
## 61	Dhuruvangal Pathinaaru	(2016)	8.2
## 62	Mandela	(2021)	8.2

## 63	Pithamagan	(2003)	8.2
## 64	Andhadhun	(2018)	8.2
## 65	Shahid	(2012)	8.2
## 66	Jigarthanda	(2014)	8.2
## 67	Gangs of Wasseyypur	(2012)	8.1
## 68	Anand	(1971)	8.1
## 69	Paan Singh Tomar	(2012)	8.1
## 70	Sairat	(2016)	8.1
## 71	Bhaag Milkha Bhaag	(2013)	8.1
## 72	Maheshinte Prathikaaram	(2016)	8.1
## 73	Hera Pheri	(2000)	8.1
## 74	Sholay	(1975)	8.1
## 75	Theeran Adhigaaram Ondru	(2017)	8.1
## 76	Swades: We, the People	(2004)	8.1
## 77	Ustad Hotel	(2012)	8.1
## 78	Chak De! India	(2007)	8.1
## 79	Talvar	(2015)	8.1
## 80	Chhichhore	(2019)	8.1
## 81	Mughal-E-Azam	(1960)	8.1
## 82	Black	(2005)	8.1
## 83	Jo Jeeta Wohi Sikandar	(1992)	8.1
## 84	Zindagi Na Milegi Dobara	(2011)	8.1
## 85	Baasha	(1995)	8.1
## 86	Charulata	(1964)	8.1
## 87	Article 15	(2019)	8.1
## 88	Nil Battey Sannata	(2015)	8.1
## 89	Virumandi	(2004)	8.1
## 90	Drishyam	(2015)	8.1
## 91	Udaan	(2010)	8.1
## 92	A Wednesday	(2008)	8.1
## 93	Alai Payuthey	(2000)	8.1
## 94	Masaan	(2015)	8.1
## 95	Munna Bhai M.B.B.S.	(2003)	8.1
## 96	Sarfarosh	(1999)	8.1
## 97	Queen	(2013)	8.1
## 98	Roja	(1992)	8.1
## 99	Dil Chahta Hai	(2001)	8.1
## 100	OMG: Oh My God!	(2012)	8.1
## 101	Rang De Basanti	(2006)	8.1
## 102	Uri: The Surgical Strike	(2019)	8.1
## 103	Lagaan: Once Upon a Time in India	(2001)	8.1
## 104	Kahaani	(2012)	8.1
## 105	Andaz Apna Apna	(1994)	8.1
## 106	PK	(2014)	8.1
## 107	Lucia	(2013)	8.1
## 108	Bommarillu	(2006)	8.1
## 109	The Legend of Bhagat Singh	(2002)	8.1
## 110	Iqbal	(2005)	8.1
## 111	Barfi!	(2012)	8.1
## 112	Section 375	(2019)	8.0
## 113	Indian	(1996)	8.0
## 114	Bombay	(1995)	8.0
## 115	Maqbool	(2003)	8.0
## 116	Pink	(2016)	8.0

## 117	The Great Indian Kitchen	(2021)	8.0
## 118	Thadam	(2019)	8.0
## 119	Omkara	(2006)	8.0
## 120	Kshanam	(2016)	8.0
## 121	Pelli Choopulu	(2016)	8.0
## 122	Lage Raho Munna Bhai	(2006)	8.0
## 123	Shershaah	(2021)	8.0
## 124	Baahubali 2: The Conclusion	(2017)	8.0
## 125	Rangasthalam	(2018)	8.0
## 126	Naduvula Konjam Pakkatha Kaanom	(2012)	8.0
## 127	Deewaar	(1975)	8.0
## 128	Pranchiyettan and the Saint	(2010)	8.0
## 129	Athadu	(2005)	8.0
## 130	Evaru	(2019)	8.0
## 131	Padayappa	(1999)	8.0
## 132	Padosan	(1968)	8.0
## 133	Gulaal	(2009)	8.0
## 134	Thondimuthalum Dhripsakshiyum	(2017)	8.0
## 135	Dilwale Dulhania Le Jayenge	(1995)	8.0
## 136	Haider	(2014)	8.0
## 137	Take Off	(2017)	8.0
## 138	Maanagaram	(2017)	8.0
## 139	K.G.F: Chapter 1	(2018)	8.0
## 140	Aadukalam	(2011)	8.0
## 141	Vaaranam Aayiram	(2008)	8.0
## 142	Baishe Srabon	(2011)	8.0
## 143	Ugly	(2013)	8.0
## 144	Ulidavaru Kandanthe	(2014)	8.0
## 145	Mother India	(1957)	8.0
## 146	Special Chabbis	(2013)	8.0
## 147	Nayattu	(2021)	8.0
## 148	Ship of Theseus	(2012)	7.9
## 149	Badhaai ho	(2018)	7.9
## 150	Bajrangi Bhaijaan	(2015)	7.9
## 151	Dev.D	(2009)	7.9
## 152	Jalsaghar	(1958)	7.9
## 153	Deiva Thirumagal	(2011)	7.9
## 154	Karnan	(2021)	7.9
## 155	Vaastav: The Reality	(1999)	7.9
## 156	Kirik Party	(2016)	7.9
## 157	Vedam	(2010)	7.9
## 158	Mukkabaaz	(2017)	7.9
## 159	Gully Boy	(2019)	7.9
## 160	Ankhon Dekhi	(2013)	7.9
## 161	Company	(2002)	7.9
## 162	Manjhi: The Mountain Man	(2015)	7.9
## 163	Charlie	(2015)	7.9
## 164	Pizza	(2012)	7.9
## 165	Memories	(2013)	7.9
## 166	Arjun Reddy	(2017)	7.9
## 167	Malik	(2021)	7.9
## 168	Poove Unakkaga	(1996)	7.9
## 169	My Name Is Khan	(2010)	7.9
## 170	Ayyappanum Koshiyum	(2020)	7.9

## 171	Dil Bechara	(2020)	7.9
## 172	Jab We Met	(2007)	7.9
## 173	Kal Ho Naa Ho	(2003)	7.9
## 174	Border	(1997)	7.9
## 175	Vinnaithaandi Varuvaayaa	(2010)	7.9
## 176	Kaakha..Kaakha: The Police	(2003)	7.9
## 177	Mumbai Police	(2013)	7.9
## 178	Super 30	(2019)	7.9
## 179	Thulladha Manamum Thullum	(1999)	7.8
## 180	Angamaly Diaries	(2017)	7.8
## 181	Sonchiriyaa	(2019)	7.8
## 182	Unnaipol Oruvan	(2009)	7.8
## 183	Anjaam Pathiraa	(2020)	7.8
## 184	Hey Ram	(2000)	7.8
## 185	Bāhubali: The Beginning	(2015)	7.8
## 186	Lakshya	(2004)	7.8
## 187	Vettaiyaadu Vilaiyaadu	(2006)	7.8
## 188	Padman	(2018)	7.8
## 189	Dor	(2006)	7.8
## 190	Baby	(2015)	7.8
## 191	Airlift	(2016)	7.8
## 192	Hindi Medium	(2017)	7.8
## 193	Okkadu	(2003)	7.8
## 194	English Vinglish	(2012)	7.8
## 195	Gangaaajal	(2003)	7.8
## 196	Manam	(2014)	7.8
## 197	The Lunchbox	(2013)	7.8
## 198	Johnny Gaddaar	(2007)	7.8
## 199	M.S. Dhoni: The Untold Story	(2016)	7.8
## 200	Ab Tak Chhappan	(2004)	7.8
## 201	Aayirathil Oruvan	(2010)	7.8
## 202	Pokiri	(2006)	7.8
## 203	Joji	(2021)	7.8
## 204	Goodachari	(2018)	7.8
## 205	Badla	(2019)	7.8
## 206	Secret Superstar	(2017)	7.7
## 207	Nayak: The Real Hero	(2001)	7.7
## 208	Stanley Ka Dabba	(2011)	7.7
## 209	Vicky Donor	(2012)	7.7
## 210	RangiTaranga	(2015)	7.7
## 211	Mr. India	(1987)	7.7
## 212	Mimi	(2021)	7.7
## 213	Raazi	(2018)	7.7
## 214	Don	(1978)	7.7
## 215	Velaiilla Pattadhari	(2014)	7.7
## 216	Udta Punjab	(2016)	7.7
## 217	Ko	(2011)	7.7
## 218	Thuppakki	(2012)	7.7
## 219	Aligarh	(2015)	7.7
## 220	Veer-Zaara	(2004)	7.7
## 221	Rock On!!	(2008)	7.7
## 222	Dasvidaniya	(2008)	7.7
## 223	Kaththi	(2014)	7.7
## 224	Arya	(2004)	7.7

## 225	Guru	(2007)	7.7
## 226	Kai po che!	(2013)	7.7
## 227	Samsara	(2001)	7.7
## 228	Ghilli	(2004)	7.7
## 229	24	(2016)	7.7
## 230	Oye Lucky! Lucky Oye!	(2008)	7.7
## 231	Rockstar	(2011)	7.7
## 232	Eega	(2012)	7.7
## 233	Avane Srimannarayana	(2019)	7.6
## 234	Mumbai Meri Jaan	(2008)	7.6
## 235	Kushi	(2000)	7.6
## 236	Ugramm	(2014)	7.6
## 237	Sir	(2018)	7.6
## 238	Darr	(1993)	7.6
## 239	Kabhi Haan Kabhi Naa	(1994)	7.6
## 240	The Tashkent Files	(2019)	7.6
## 241	Kapoor & Sons	(2016)	7.6
## 242	Newton	(2017)	7.6
## 243	C U Soon	(2020)	7.6
## 244	Aamir	(2008)	7.6
## 245	Earth	(1998)	7.6
## 246	Detective Byomkesh Bakshy!	(2015)	7.6
## 247	Madras Cafe	(2013)	7.6
## 248	Agneepath	(1990)	7.6
## 249	Sarkar	(2005)	7.6
## 250	State of Siege: Temple Attack	(2021)	7.6