

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              Nayakan          (1987)  8.5
## 2              Anbe Sivam          (2003)  8.5
## 3      Pariyerum Perumal          (2018)  8.5
## 4      C/o Kancharapalem          (2018)  8.5
## 5              Golmaal          (1979)  8.5
## 6      Manichitrathazhu          (1993)  8.5
## 7              Kireedam          (1989)  8.5
## 8              Apur Sansar          (1959)  8.5
```

## 9	Natsamrat	(2016)	8.4
## 10	96	(2018)	8.4
## 11	Thevar Magan	(1992)	8.4
## 12	Kumbalangi Nights	(2019)	8.4
## 13	Pather Panchali	(1955)	8.4
## 14	Black Friday	(2004)	8.4
## 15	Soorarai Pottru	(2020)	8.4
## 16	Visaaranai	(2015)	8.4
## 17	3 Idiots	(2009)	8.4
## 18	#Home	(2021)	8.3
## 19	Jersey	(2019)	8.3
## 20	Taare Zameen Par	(2007)	8.3
## 21	Drishyam 2	(2021)	8.3
## 22	Asuran	(2019)	8.3
## 23	Thalapathi	(1991)	8.3
## 24	Kaithi	(2019)	8.3
## 25	Sarpatta Parambarai	(2021)	8.3
## 26	Dangal	(2016)	8.3
## 27	Devasuram	(1993)	8.3
## 28	Aparajito	(1956)	8.3
## 29	Jaane Bhi Do Yaaro	(1983)	8.3
## 30	Ratsasan	(2018)	8.3
## 31	Pyaasa	(1957)	8.3
## 32	Vada Chennai	(2018)	8.3
## 33	Peranbu	(2018)	8.3
## 34	Guide	(1965)	8.3
## 35	Thani Oruvan	(2015)	8.3
## 36	Kannathil Muthamittal	(2002)	8.2
## 37	Spadikam	(1995)	8.2
## 38	Iruvar	(1997)	8.2
## 39	Agent Sai Srinivasa Athreya	(2019)	8.2
## 40	Chupke Chupke	(1975)	8.2
## 41	Super Deluxe	(2019)	8.2
## 42	Mahanati	(2018)	8.2
## 43	Aruvi	(2016)	8.2
## 44	Khosla Ka Ghosla!	(2006)	8.2
## 45	Pudhu Pettai	(2006)	8.2
## 46	Drishyam	(2013)	8.2
## 47	Tumbbad	(2018)	8.2
## 48	Vikram Vedha	(2017)	8.2
## 49	Anniyan	(2005)	8.2
## 50	Premam	(2015)	8.2
## 51	Mudhalvan	(1999)	8.2
## 52	Kaakkaa Muttai	(2014)	8.2
## 53	Bangalore Days	(2014)	8.2
## 54	Satya	(1998)	8.2
## 55	Angoor	(1982)	8.2
## 56	Papanasam	(2015)	8.2
## 57	Andhadhun	(2018)	8.2
## 58	Soodhu Kavvum	(2013)	8.2
## 59	Dhuruvangal Pathinaaru	(2016)	8.2
## 60	Anand	(1971)	8.2
## 61	Shahid	(2012)	8.2
## 62	Mandela	(2021)	8.2

## 63	Pithamagan	(2003)	8.2
## 64	Gangs of Wasseyypur	(2012)	8.1
## 65	Jigarthanda	(2014)	8.1
## 66	Paan Singh Tomar	(2012)	8.1
## 67	Bhaag Milkha Bhaag	(2013)	8.1
## 68	Sairat	(2016)	8.1
## 69	Hera Pheri	(2000)	8.1
## 70	Sholay	(1975)	8.1
## 71	Swades: We, the People	(2004)	8.1
## 72	Maheshinte Prathikaaram	(2016)	8.1
## 73	Talvar	(2015)	8.1
## 74	Chak De! India	(2007)	8.1
## 75	Ustad Hotel	(2012)	8.1
## 76	Mughal-E-Azam	(1960)	8.1
## 77	Black	(2005)	8.1
## 78	Zindagi Na Milegi Dobara	(2011)	8.1
## 79	Jo Jeeta Wohi Sikandar	(1992)	8.1
## 80	Theeran adhigaaram ondru	(2017)	8.1
## 81	Drishyam	(2015)	8.1
## 82	Chhichhore	(2019)	8.1
## 83	Nil Battey Sannata	(2015)	8.1
## 84	Article 15	(2019)	8.1
## 85	Charulata	(1964)	8.1
## 86	Udaan	(2010)	8.1
## 87	A Wednesday	(2008)	8.1
## 88	Baasha	(1995)	8.1
## 89	Masaan	(2015)	8.1
## 90	Munna Bhai M.B.B.S.	(2003)	8.1
## 91	Alai Payuthey	(2000)	8.1
## 92	Sarfarosh	(1999)	8.1
## 93	Queen	(2013)	8.1
## 94	Virumandi	(2004)	8.1
## 95	Dil Chahta Hai	(2001)	8.1
## 96	Rang De Basanti	(2006)	8.1
## 97	OMG: Oh My God!	(2012)	8.1
## 98	Roja	(1992)	8.1
## 99	Uri: The Surgical Strike	(2019)	8.1
## 100	Lagaan: Once Upon a Time in India	(2001)	8.1
## 101	Kahaani	(2012)	8.1
## 102	Andaz Apna Apna	(1994)	8.1
## 103	PK	(2014)	8.1
## 104	Lucia	(2013)	8.1
## 105	Bommarillu	(2006)	8.1
## 106	The Legend of Bhagat Singh	(2002)	8.1
## 107	Iqbal	(2005)	8.1
## 108	Barfi!	(2012)	8.1
## 109	Section 375	(2019)	8.0
## 110	Pink	(2016)	8.0
## 111	Maqbool	(2003)	8.0
## 112	Bombay	(1995)	8.0
## 113	Shershaah	(2021)	8.0
## 114	Indian	(1996)	8.0
## 115	Omkara	(2006)	8.0
## 116	The Great Indian Kitchen	(2021)	8.0

## 117	Thadam	(2019)	8.0
## 118	Kshanam	(2016)	8.0
## 119	Lage Raho Munna Bhai	(2006)	8.0
## 120	Pelli Choopulu	(2016)	8.0
## 121	Baahubali 2: The Conclusion	(2017)	8.0
## 122	Deewaar	(1975)	8.0
## 123	Pranchiyettan and the Saint	(2010)	8.0
## 124	Naduvula Konjam Pakkatha Kaanom	(2012)	8.0
## 125	Rangasthalam	(2018)	8.0
## 126	Padosan	(1968)	8.0
## 127	Gulaal	(2009)	8.0
## 128	Athadu	(2005)	8.0
## 129	Dilwale Dulhania Le Jayenge	(1995)	8.0
## 130	Haider	(2014)	8.0
## 131	Thondimuthalum Dhripsakshiyum	(2017)	8.0
## 132	Padayappa	(1999)	8.0
## 133	Take Off	(2017)	8.0
## 134	Ugly	(2013)	8.0
## 135	K.G.F: Chapter 1	(2018)	8.0
## 136	Aadukalam	(2011)	8.0
## 137	Mother India	(1957)	8.0
## 138	Vaaranam Aayiram	(2008)	8.0
## 139	Special Chabbis	(2013)	8.0
## 140	Ulidavaru Kandanthe	(2014)	7.9
## 141	Nayattu	(2021)	7.9
## 142	Ship of Theseus	(2012)	7.9
## 143	Badhaai ho	(2018)	7.9
## 144	Dev.D	(2009)	7.9
## 145	Bajrangi Bhaijaan	(2015)	7.9
## 146	Jalsaghar	(1958)	7.9
## 147	Vaastav: The Reality	(1999)	7.9
## 148	Gully Boy	(2019)	7.9
## 149	Deiva Thirumagal	(2011)	7.9
## 150	Mukkabaaz	(2017)	7.9
## 151	Ankhon Dekhi	(2013)	7.9
## 152	Vedam	(2010)	7.9
## 153	Company	(2002)	7.9
## 154	Manjhi: The Mountain Man	(2015)	7.9
## 155	Karnan	(2021)	7.9
## 156	Charlie	(2015)	7.9
## 157	Pizza	(2012)	7.9
## 158	Arjun Reddy	(2017)	7.9
## 159	Memories	(2013)	7.9
## 160	Malik	(2021)	7.9
## 161	My Name Is Khan	(2010)	7.9
## 162	Jab We Met	(2007)	7.9
## 163	Kal Ho Naa Ho	(2003)	7.9
## 164	Dil Bechara	(2020)	7.9
## 165	Border	(1997)	7.9
## 166	Ayyappanum Koshiyum	(2020)	7.9
## 167	Super 30	(2019)	7.9
## 168	Vinnaithaandi Varuvaayaa	(2010)	7.9
## 169	Mumbai Police	(2013)	7.9
## 170	Poove Unakkaga	(1996)	7.8

## 171	Angamaly Diaries	(2017)	7.8
## 172	Unnaipol Oruvan	(2009)	7.8
## 173	Kaakha..Kaakha: The Police	(2003)	7.8
## 174	Sonchiriya	(2019)	7.8
## 175	Lakshya	(2004)	7.8
## 176	Bāhubali: The Beginning	(2015)	7.8
## 177	Anjaam Pathiraa	(2020)	7.8
## 178	Hey Ram	(2000)	7.8
## 179	Padman	(2018)	7.8
## 180	Vettaiyaadu Vilaiyaadu	(2006)	7.8
## 181	Baby	(2015)	7.8
## 182	Dor	(2006)	7.8
## 183	Thulladha Manamum Thullum	(1999)	7.8
## 184	Airlift	(2016)	7.8
## 185	Hindi Medium	(2017)	7.8
## 186	English Vinglish	(2012)	7.8
## 187	Okkadu	(2003)	7.8
## 188	The Lunchbox	(2013)	7.8
## 189	Gangaaajal	(2003)	7.8
## 190	Manam	(2014)	7.8
## 191	Johnny Gaddaar	(2007)	7.8
## 192	Ab Tak Chhappan	(2004)	7.8
## 193	M.S. Dhoni: The Untold Story	(2016)	7.8
## 194	Joji	(2021)	7.8
## 195	Pokiri	(2006)	7.8
## 196	Badla	(2019)	7.8
## 197	Goodachari	(2018)	7.7
## 198	Secret Superstar	(2017)	7.7
## 199	Mimi	(2021)	7.7
## 200	Nayak: The Real Hero	(2001)	7.7
## 201	Vicky Donor	(2012)	7.7
## 202	Aayirathil Oruvan	(2010)	7.7
## 203	Stanley Ka Dabba	(2011)	7.7
## 204	Mr. India	(1987)	7.7
## 205	Raazi	(2018)	7.7
## 206	Don	(1978)	7.7
## 207	Udta Punjab	(2016)	7.7
## 208	RangiTaranga	(2015)	7.7
## 209	Aligarh	(2015)	7.7
## 210	Ko	(2011)	7.7
## 211	Velaiilla Pattadhari	(2014)	7.7
## 212	Veer-Zaara	(2004)	7.7
## 213	Rock On!!	(2008)	7.7
## 214	Thuppakki	(2012)	7.7
## 215	Dasvidaniya	(2008)	7.7
## 216	Arya	(2004)	7.7
## 217	Guru	(2007)	7.7
## 218	Kaththi	(2014)	7.7
## 219	Kai po che!	(2013)	7.7
## 220	Samsara	(2001)	7.7
## 221	Oye Lucky! Lucky Oye!	(2008)	7.7
## 222	24	(2016)	7.7
## 223	Ghilli	(2004)	7.7
## 224	Rockstar	(2011)	7.7

## 225	Eega	(2012)	7.6
## 226	Mumbai Meri Jaan	(2008)	7.6
## 227	Avane Srimannarayana	(2019)	7.6
## 228	Darr	(1993)	7.6
## 229	Kabhi Haan Kabhi Naa	(1994)	7.6
## 230	Sir	(2018)	7.6
## 231	C U Soon	(2020)	7.6
## 232	Kapoor & Sons	(2016)	7.6
## 233	Newton	(2017)	7.6
## 234	The Tashkent Files	(2019)	7.6
## 235	Aamir	(2008)	7.6
## 236	Ugramm	(2014)	7.6
## 237	Earth	(1998)	7.6
## 238	Detective Byomkesh Bakshy!	(2015)	7.6
## 239	Agneepath	(1990)	7.6
## 240	State of Siege: Temple Attack	(2021)	7.6
## 241	Madras Cafe	(2013)	7.6
## 242	Sarkar	(2005)	7.6
## 243	Bheja Fry	(2007)	7.6
## 244	Neerja	(2016)	7.6
## 245	Pyaar Ka Punchnama	(2011)	7.6
## 246	Kannum Kannum Kollaiyadithaal	(2020)	7.6
## 247	Magadheera	(2009)	7.6
## 248	Highway	(2014)	7.6
## 249	Baazigar	(1993)	7.6
## 250	1 - Nenokkadine	(2014)	7.6