

Importing modules

```
In [41]: from bs4 import BeautifulSoup
import requests
import pandas as pd
import re
```

To specify url for fetching

```
In [42]: headers = {'User-Agent': 'Mozilla/5.0 (Windows NT 10.0; Win64; x64) AppleWebKit/537.36 (KHTML, like Gecko) Chrome/80.0.3987.163 Safari/537.36'}

page_url = "https://www.imdb.com/india/top-rated-indian-movies/"
response = requests.get(page_url, headers=headers)

response
```

Out[42]: <Response [200]>

```
In [43]: response.content
```

```
Out[43]: b'<!DOCTYPE html><html lang="en-US" xmlns:og="http://opengraphprotocol.org/schema/" xmlns:fb="http://www.facebook.com/2008/fbml"><head><meta charset="utf-8"/><meta name="viewport" content="width=device-width"/><title>Top Rated Indian Movies - IMDb</title><meta name="title" content="Top Rated Indian Movies - IMDb"/><meta name="description" content="Discover the best Indian movies based on IMDb user ratings"/><meta name="keywords" content="Best Indian movies,top indian movies,best Hindi movies,best tamil movies,best telugu movies"/><link rel="canonical" href="https://www.imdb.com/india/top-rated-indian-movies/"><meta property="imdb:pageType" content="imdb-feature d"/><meta property="imdb:subPageType" content="india"/><meta property="imdb:specialSectionPageId" content="india.top-rated-indian-movies"/><meta property="og:title" content="Top Rated Indian Movies - IMDb"/><meta property="og:description" content="Discover the best Indian movies based on IMDb user ratings"/><meta property="og:type" content="website"/><meta property="og:url" content="https://www.imdb.com/india/top-rated-indian-movies/"><meta property="twitter:title" content="Top Rated Indian Movies - IMDb"/><meta property="twitter:description" content="Discover the best Indian movies based on IMDb user ratings"/><meta property="twitter:card" content="summary_large_image"/><meta property="twitter:url" content="https://www.imdb.com/india/top-rated-indian-movies/"><meta name="next-head-count" content="16"/><script>
```

Use beautifulsoup to extract and parse info

```
In [44]: soup = BeautifulSoup(response.content, 'html.parser')
```

```
In [45]: span_tags = soup.find_all('span')
```

```
In [46]: len(span_tags)
```

```
Out[46]: 1386
```

```
In [47]: span_tags[:2]
```

```
Out[47]: [<span class="ipc-responsive-button__text">Menu</span>,
  <span class="navlinkcat__targetWrapper"><label aria-label="Expand Movies Nav
  Links" class="navlinkcat__item" data-testid="category-expando" for="nav-link-
  categories-mov" role="button" tabindex="0"><span class="navlinkcat__itemIco
  n"><svg class="ipc-icon ipc-icon--movie" fill="currentColor" height="24" role
  ="presentation" viewBox="0 0 24 24" width="24" xmlns="http://www.w3.org/2000/
  svg"><path d="M0 0h24v24H0V0z" fill="none"></path><path d="M18 4v1h-2V4c0-.55
  -.45-1-1-1H9c-.55 0-1 .45-1 1v1H6V4c0-.55-.45-1-1-1s-1 .45-1 1v16c0 .55.45 1
  1 1s1-.45 1-1v-1h2v1c0 .55.45 1 1 1h6c.55 0 1-.45 1-1v-1h2v1c0 .55.45 1 1 1s1
  -.45 1-1V4c0-.55-.45-1-1-1s-1 .45-1 1z" M8 17H6v-2h2v2zm0-4H6v-2h2v2zm0-4H6V7h2
  v2zm10 8h-2v-2h2v2zm0-4h-2v-2h2v2zm0-4h-2V7h2v2z"></path></svg></span><span c
  lass="navlinkcat__itemTitle">Movies</span><span class="navlinkcat__itemChevro
  n"><svg class="ipc-icon ipc-icon--chevron-right" fill="currentColor" height
  ="24" role="presentation" viewBox="0 0 24 24" width="24" xmlns="http://www.w
  3.org/2000/svg"><path d="M0 0h24v24H0V0z" fill="none"></path><path d="M9.29
  6.71a.996.996 0 0 0 0 1.41L13.17 12l-3.88 3.88a.996.996 0 1 0 1.41 1.41L14.59-
  4.59a.996.996 0 0 0 0-1.41L10.7 6.7c-.38-.38-1.02-.38-1.41.01z"></path></svg>
  </span></label><div aria-expanded="false" aria-hidden="true" class="navlinkcat
  _listContainer" data-testid="list-container"><div class="navlinkcat__listCo
  ntainerInner" role="presentation"><ul aria-orientation="vertical" class="ipc-
  list navlinkcat__list ipc-list--baseAlt" role="menu"><a aria-disabled="false"
  class="ipc-list__item nav-link sc-bBjSGg exGTOU ipc-list__item--indent-one" h
  ref="https://www.imdb.com/calendar/?ref=nv_mv_cal" role="menuitem" tabindex
  ="-1"><span class="ipc-list-item__text" role="presentation">Release Calendar
  </span></a><a aria-disabled="false" class="ipc-list__item nav-link sc-bBjSGg
  exGTOU ipc-list__item--indent-one" href="/chart/top/?ref=nv_mv_250" role="me
  nuitem" tabindex="-1"><span class="ipc-list-item__text" role="presentation">T
  op 250 Movies</span></a><a aria-disabled="false" class="ipc-list__item nav-li
  nk sc-bBjSGg exGTOU ipc-list__item--indent-one" href="/chart/moviemeter/?ref_
  =nv_mv_mpm" role="menuitem" tabindex="-1"><span class="ipc-list-item__text" r
  ole="presentation">Most Popular Movies</span></a><a aria-disabled="false" cla
  ss="ipc-list__item nav-link sc-bBjSGg exGTOU ipc-list__item--indent-one" href
  ="/feature/genre/?ref=nv_ch_gr" role="menuitem" tabindex="-1"><span class="i
  pc-list-item__text" role="presentation">Browse Movies by Genre</span></a><a a
  ria-disabled="false" class="ipc-list__item nav-link sc-bBjSGg exGTOU ipc-list
  __item--indent-one" href="/chart/boxoffice/?ref=nv_ch_cht" role="menuitem" t
  abindex="-1"><span class="ipc-list-item__text" role="presentation">Top Box Of
  fice</span></a><a aria-disabled="false" class="ipc-list__item nav-link sc-bBj
  SGg exGTOU ipc-list__item--indent-one" href="/showtimes/?ref=nv_mv_sh" role
  ="menuitem" tabindex="-1"><span class="ipc-list-item__text" role="presentatio
  n">Showtimes & Tickets</span></a><a aria-disabled="false" class="ipc-list
  __item nav-link sc-bBjSGg exGTOU ipc-list__item--indent-one" href="/news/movi
  e/?ref=nv_nw_mv" role="menuitem" tabindex="-1"><span class="ipc-list-item__t
  ext" role="presentation">Movie News</span></a><a aria-disabled="false" class
  ="ipc-list__item nav-link sc-bBjSGg exGTOU ipc-list__item--indent-one" href
  ="/india/toprated/?ref=nv_mv_in" role="menuitem" tabindex="-1"><span class
  ="ipc-list-item__text" role="presentation">India Movie Spotlight</span></a></
  ul></div></div></span>]
```

In [48]: `print(soup.prettify())`

```
<!DOCTYPE html>
<html lang="en-US" xmlns:fb="http://www.facebook.com/2008/fbml" xmlns:og="http://opengraphprotocol.org/schema/">
  <head>
    <meta charset="utf-8"/>
    <meta content="width=device-width" name="viewport"/>
    <title>
      Top Rated Indian Movies - IMDb
    </title>
    <meta content="Top Rated Indian Movies - IMDb" name="title"/>
    <meta content="Discover the best Indian movies based on IMDb user ratings" name="description"/>
    <meta content=" Best Indian movies,top indian movies,best Hindi movies,best tamil movies,best telugu movies" name="keywords"/>
    <link href="https://www.imdb.com/india/top-rated-indian-movies/" rel="canonical"/>
    <meta content="imdb-featured" property="imdb:pageType"/>
    <meta content="india" property="imdb:subPageType"/>
    <meta content="india.top-rated-indian-movies" property="imdb:specialSection"/>
  </head>
```

In [49]: `scraped_movies = soup.find_all('span',class_="sc-9910edf6-4 eyrnkD")`
`scraped_movies`

Out[49]: [`1Ramayana: The Legend of Prince Rama`,
`212th Fail`,
`3Nayakan`,
`4Gol Maal`,
`5Anbe Sivam`,
`6Rocketry: The Nambi Effect`,
`7777 Charlie`,
`8Pariyerum Perumal`,
`9Manichitrathazhu`,
`10...`]

```
In [55]: pattern = re.compile(r'\d')
movies = []
for movie in scraped_movies:
    new_name = pattern.sub(' ',movie.get_text())
    movies.append(new_name)
movies
```

```
Out[55]: [' Ramayana: The Legend of Prince Rama',
          '   th Fail',
          ' Nayakan',
          ' Gol Maal',
          ' Anbe Sivam',
          ' Rocketry: The Nambi Effect',
          '   Charlie',
          ' Pariyerum Perumal',
          ' Manichitrathazhu',
          ' #Home',
          '   Idiots',
          ' Apur Sansar',
          ' Kireedam',
          ' C/O Kancharapalem',
          ' Soorarai Pottru',
          ' Kumbalangi Nights',
          ' Black Friday',
          ' Sandesham',
          ' Jai Bhim',
          ' Taare Zameen Par',
          ' Jersey',
          ' Maya Bazaar',
          '   ',
          ' Natsamrat',
          ' Laapataa Ladies',
          ' Dangal',
          ' Kaithi',
          ' Sita Ramam',
          ' Drishyam ',
          ' Thevar Magan',
          ' Visaaranai',
          ' Asuran',
          ' Nadodikkattu',
          ' Thalapathi',
          ' Sarpatta Parambarai',
          ' Drishyam',
          ' Thani Oruvan',
          ' Nuvvu Naaku Nachchav',
          ' Anniyan',
          ' Vada Chennai',
          ' Jaane Bhi Do Yaaro',
          ' Khosla Ka Ghosla!',
          ' Peranbu',
          ' Aparajito',
          ' Ratsasan',
          ' Sardar Udham',
          ' Bangalore Days',
          ' Devasuram',
          ' Premam',
          ' Chupke Chupke',
          ' Pather Panchali',
          ' Gangs of Wasseypur',
          ' Satya',
          ' Agent Sai Srinivasa Athreya',
          ' Mahanati',
          ' Manjummel Boys',
          ' Chithram',
```

' Drishyam',
' Chhichhore',
' Bhaag Milkha Bhaag',
' Spadikam',
' Guide',
' Super Deluxe',
' Andhadhun',
' Vikram Vedha',
' Zindagi Na Milegi Dobara',
' Tumbbad',
' Vikram',
' Iruvar',
' Swades: We, the People',
' Kannathil Muthamittal',
' Sairat',
' Hera Pheri',
' Shahid',
' Mudhalvan',
' Aruvi',
' Munna Bhai M.B.B.S.',
' Papanasam',
' Pyaasa',
' Mandela',
' Paan Singh Tomar',
' Chak De! India',
' Uri: The Surgical Strike',
' Pudhu Pettai',
' Aaranya Kaandam',
' Jo Jeeta Wohi Sikandar',
' Black',
' Jana Gana Mana',
' Lagaan: Once Upon a Time in India',
' PK',
' Soodhu Kavvum',
' Drishyam ',
' Jigarthanda',
' Virumandi',
' Dhuruvangal Pathinaaru',
' Sholay',
' ',
' Kaakkaa Muttai',
' Sarfarosh',
' Barfi!',
' Udaan',
' Talvar',
' OMG: Oh My God!',
' Maheshinte Prathikaram',
' Angoor',
' Queen',
' Theeran Adhigaaram Ondru',
' Article ',
' Ustad Hotel',
' The Legend of Bhagat Singh',
' Baasha',
' Bāhubali : The Conclusion',
' Rang De Basanti',
' Pithamagan',

' Kahaani',
' Oru Vadakkan Veeragatha',
' Lion',
' Masaan',
' A Wednesday',
' Dil Chahta Hai',
' Shershaah',
' Nil Battey Sannata',
' Bajrangi Bhaijaan',
' Kaun Pravin Tambe?',
' Lage Raho Munna Bhai',
' Iqbal',
' Lucia',
' Alai Payuthey',
' Anand',
' Bommarillu',
' Roja',
' Pink',
' Charulata',
' Omkara',
' Indian',
' Bombay',
' Section 8',
' Viduthalai Part 1',
' Vaaranam Aayiram',
' Padayappa',
' Dilwale Dulhania Le Jayenge',
' K.G.F: Chapter 1',
' Rangasthalam',
' Athadu',
' Mughal-E-Azam',
' The Great Indian Kitchen',
' Thadam',
' Haider',
' Andaz Apna Apna',
' Kantara',
' Special Chabbis',
' Pelli Choopulu',
' Maqbool',
' Chithha',
' Maanaadu',
' Ulidavaru Kandantho',
' Gulaal',
' Hi Nanna',
' Only One Man Is Enough',
' Aadukalam',
' Naduvula Konjam Pakkatha Kaanom',
' Padosan',
' Maanagaram',
' Vaastav: The Reality',
' Deewaar',
' K.G.F: Chapter 2',
' Company',
' Saptha Sagaradaache Ello - Side A',
' Kshanam',
' Nayattu',
' Evaru',

```
'    Ugly',
'    Badhaai Ho',
'    Charlie',
'    Pranchiyettan and the Saint',
'    Garuda Gamana Vrishabha Vahana',
'    Ayyappanum Koshiyum',
'    Baishe Srabon',
'    My Name Is Khan',
'    M.S. Dhoni: The Untold Stor',
'    The Goat Life',
'    Deiva Thirumagal',
'    Gully Boy',
'    Thondi Muthalum Driksakshiyum',
'    Ship of Theseus',
'    Take Off',
'    Major',
'    Kal Ho Naa Ho',
'    Jab We Met',
'    Dev.D',
'    Vedam',
'    Manjhi: The Mountain Man',
'    Border',
'    Mukkabaaz',
'    Gargi',
'    Kirik Party',
'    Hridayam',
'    Por Thozhil',
'    Vinnaithaandi Varuvaayaa',
'    Bāhubali: The Beginning',
'    Salaam Bombay!',
'    Jigarthanda Double X',
'    Dil Bechara',
'    Karnan',
'    Super    ',
'    Traffic',
'    Pizza',
'    Hey Ram',
'    Padman',
'    Love Today',
'    Android Kunjappan Version .  ',
'    Baby',
'    Ankhon Dekhi',
'    Ghilli',
'    Malik',
'    Hindi Medium',
'    Memories',
'    Joseph',
'    The Tashkent Files',
'    Thuppakki',
'    Article    ',
'    Jalsaghar',
'    Anjaam Pathiraa',
'    English Vinglish',
'    Arjun Reddy',
'    Dor',
'    Johnny Gaddaar',
'    Lakshya',
```



```
' Pokiri',
' Airlift',
' Mera Naam Joker',
' Kaakha Kaakha',
' Okkadu',
' Vettaiyaadu Vilaiyaadu',
' Thulladha Manamum Thullum',
' Gangaaajal',
' Oopiri',
' The Lunchbox',
' Kaththi',
' Mumbai Police',
' Manam',
' Veer Zara',
' Leader',
' Unnaipol Oruvan',
' Chamkila',
' Ayirathil Oruvan',
' Ab Tak Chhappan',
' Poove Unakkaga',
' RangiTaranga',
' Aavesham']
```

```
In [51]: scraped_ratings = soup.find_all('div',class_="sc-e2dbc1a3-0 ajrIH")
scraped_ratings
```

```
Out[51]: [<div class="sc-e2dbc1a3-0 ajrIH" data-testid="ratingGroup--container"><span
aria-label="IMDb rating: 9.2" class="ipc-rating-star ipc-rating-star--base
ipc-rating-star--imdb ratingGroup--imdb-rating" data-testid="ratingGroup-
imdb-rating"><svg class="ipc-icon ipc-icon--star-inline" fill="currentColo
r" height="24" role="presentation" viewBox="0 0 24 24" width="24" xmlns="ht
tp://www.w3.org/2000/svg"><path d="M12 20.115.82 3.682c1.066.675 2.37-.322
2.09-1.584l-1.543-6.926 5.146-4.667c.94-.85.435-2.465-.799-2.567l-6.773-.60
2L13.29.89a1.38 1.38 0 0 0-2.581 0l-2.65 6.53-6.774.602C.052 8.126-.453 9.7
4.486 10.59l5.147 4.666-1.542 6.926c-.28 1.262 1.023 2.26 2.09 1.585L12 20.
099z"></path></svg>9.2</span><button aria-label="Rate Ramayana: The Legend
of Prince Rama" class="ipc-rate-button sc-e2dbc1a3-1 jbo0Qc ratingGroup--us
er-rating ipc-rate-button--unrated ipc-rate-button--base" data-testid="rate
-button"><span class="ipc-rating-star ipc-rating-star--base ipc-rating-star
--rate"><svg class="ipc-icon ipc-icon--star-border-inline" fill="currentCol
or" height="24" role="presentation" viewBox="0 0 24 24" width="24" xmlns="h
ttp://www.w3.org/2000/svg"><path d="M22.724 8.217l-6.786-.587-2.65-6.22c-.4
77-1.133-2.103-1.133-2.58 0l-2.65 6.234-6.772.573c-1.234.098-1.739 1.636-.8
2.446l5.146 4.446-1.542 6.598c-.28 1.202 1.023 2.153 2.09 1.511l5.818-3.495
5.819 3.509c1.065.643 2.37-.308 2.089-1.511-1.542-6.612 5.145-4.446c.94-.8
1.45 2.348 7.85 2.446-1.10 7.26 8.801 5.273 2.174 1.402 5.082 4.655 4.026 6
```

```
In [56]: ratings = []
for rating in scraped_ratings:
    ratings.append(rating.get_text())
ratings
```

```
Out[56]: ['9.2',
          '8.9',
          '8.7',
          '8.5',
          '8.6',
          '8.7',
          '8.7',
          '8.7',
          '8.7',
          '8.8',
          '8.4',
          '8.4',
          '8.9',
          '8.8',
          '8.7',
          '8.5',
          '8.4',
          '9.0',
          '8.7',
          '8.7']
```

Storing scraped data in csv file

```
In [61]: df = pd.DataFrame()
df['movies'] = movies
df['ratings'] = ratings
df.head()
```

```
Out[61]:
```

	movies	ratings
0	Ramayana: The Legend of Prince Rama	9.2
1	th Fail	8.9
2	Nayakan	8.7
3	Gol Maal	8.5
4	Anbe Sivam	8.6

```
In [63]: df.to_csv('IMDb Top 250 Movies.csv', index=False)
```

```
In [ ]:
```