

```
In [2]: #importing Libraries
import pandas as pd
import numpy as np
import requests
from bs4 import BeautifulSoup
from random import randint
```

```
In [3]: #creating a empty list
movie_name = []
year = []
time = []
genre = []
rating = []
description = []
```

```
In [4]:
pages = np.arange(1,1000,50)
```

```
In [5]: #fetching the data from the pages
for page in pages:
    page = requests.get("https://www.imdb.com/search/title/?groups=top_1000&sort=
    soup = BeautifulSoup(page.text,"html.parser")
    movie_data = soup.findAll("div",attrs= {'class' : 'list-item mode-advanced'})
    for store in movie_data:
        name = store.h3.a.text
        movie_name.append(name)

        year_of_movie = store.h3.find('span', class_ = 'list-item-year text-muted')
        year.append(year_of_movie)

        timing_of_movie = store.p.find('span', class_ = 'runtime').text
        time.append(timing_of_movie)

        genre_of_movie = store.p.find('span', class_ = 'genre').text.replace('/n', ' ')
        genre.append(genre_of_movie)

        rating_of_movie = store.find('div', class_ = 'inline-block ratings-imdb-rating')
        rating.append(rating_of_movie)

        descrip = store.find_all('p', class_ = 'text-muted')
        description_ = descrip[1].text.replace('/n', ' ')
        description.append(description_)
```

```
In [6]: movie_list = pd.DataFrame({'Movie Name': movie_name, 'Year': year, 'Watch Time':
```

In [7]: movie_list

Out[7]:

	Movie Name	Year	Watch Time	Genre	Rating	Description
0	The Shawshank Redemption	(1994)	142 min	\nDrama	\n\n9.3\n	\nTwo imprisoned men bond over a number of yea...
1	The Godfather	(1972)	175 min	\nCrime, Drama	\n\n9.2\n	\nThe aging patriarch of an organized crime dy...
2	The Dark Knight	(2008)	152 min	\nAction, Crime, Drama	\n\n9.0\n	\nWhen the menace known as the Joker wreaks ha...
3	The Lord of the Rings: The Return of the King	(2003)	201 min	\nAction, Adventure, Drama	\n\n9.0\n	\nGandalf and Aragorn lead the World of Men ag...
4	Schindler's List	(1993)	195 min	\nBiography, Drama, History	\n\n9.0\n	\nIn German-occupied Poland during World War I...
...
995	American History X	(1998)	119 min	\nCrime, Drama	\n\n8.5\n	\nA former neo-nazi skinhead tries to prevent ...
996	The Usual Suspects	(1995)	106 min	\nCrime, Drama, Mystery	\n\n8.5\n	\nA sole survivor tells of the twisty events I...
997	Léon	(1994)	110 min	\nAction, Crime, Drama	\n\n8.5\n	\n12-year-old Mathilda is reluctantly taken in...
998	The Lion King	(1994)	88 min	\nAnimation, Adventure, Drama	\n\n8.5\n	\nLion prince Simba and his father are targete...
999	Nuovo Cinema Paradiso	(1988)	155 min	\nDrama, Romance	\n\n8.5\n	\nA filmmaker recalls his childhood when falli...

1000 rows × 6 columns

In [8]: movie_list.to_excel("Top rated 1000 Movies by IMDb.xlsx")

```
In [9]: movie_list["Movie Name"]
```

```
Out[9]: 0          The Shawshank Redemption
        1          The Godfather
        2          The Dark Knight
        3  The Lord of the Rings: The Return of the King
        4          Schindler's List
        ...
        995        American History X
        996        The Usual Suspects
        997          Léon
        998        The Lion King
        999        Nuovo Cinema Paradiso
        Name: Movie Name, Length: 1000, dtype: object
```

```
In [15]: movie_list['Genre'].isnull().sum()
```

```
Out[15]: 0
```

```
In [17]: movie_list['Rating']
```

```
Out[17]: 0      \n\n9.3\n
        1      \n\n9.2\n
        2      \n\n9.0\n
        3      \n\n9.0\n
        4      \n\n9.0\n
        ...
        995    \n\n8.5\n
        996    \n\n8.5\n
        997    \n\n8.5\n
        998    \n\n8.5\n
        999    \n\n8.5\n
        Name: Rating, Length: 1000, dtype: object
```

```
In [ ]:
```