In [1]:
```
import requests as req
from bs4 import BeautifulSoup
from time import sleep
import random
from numpy import random
import pandas as pd
```

In [2]:
```
from selenium.webdriver import Chrome
driver = Chrome(executable_path='chromedriver')
```

In [3]:
```
url = "https://www.imdb.com/chart/top"
driver.get(url)
html = driver.page_source
s = BeautifulSoup(html, "lxml")
```

In [4]:
```
genres_links = []
genres = []
films = s.find('ul', class_='quicklinks').findAll('li', class_='subnav_item_main')
for film in films:
    link = "https://www.imdb.com"+film.find('a').get('href')
    genre = film.find('a').text
    genres_links.append([link])
    genres.append(str(genre)[1:-1])
```

In [9]:
```
data =[]
i = 0
for url in genres_links:
    sleep(random.uniform(3, 7))
    driver.get(str(url)[2:-2])
    html = driver.page_source
    s = BeautifulSoup(html, "lxml")
    print(genres[i])
    blocks = s.findAll('div', class_='lister-item mode-advanced')

    for block in blocks:
        title = block.find('div', class_='lister-item-content').find('a').text
        year = block.find('div', class_='lister-item-content').find('span', class_='
        score = block.find('div', class_='lister-item-content').find('strong').text
        data.append([title, year, genres[i], score])

    i = i + 1
```

```
Action
Adventure
Animation
Biography
Comedy
Crime
Drama
Family
Fantasy
Film-Noir
History
Horror
Music
Musical
Mystery
```

```
Romance
Sci-Fi
Sport
Thriller
War
Western
```

In [10]:
```
len(data)
```

Out[10]:
```
1030
```

In [11]:
```
head = ['title', 'year', 'genre', 'score']
dt = pd.DataFrame(data, columns=head)
dt.to_csv('imdb_data.csv', sep=';', encoding='utf-16')
```

In [ ]:
```

```