In [1]:
```python
# 셀레니움
from selenium import webdriver
from selenium.webdriver.common.keys import Keys
from selenium.webdriver.common.by import By

import os
import time
import pandas as pd

# xlsx파일로 저장
import openpyxl
```

In [2]:
```python
# 파일 불러오기
xlsx = pd.read_excel('KOBIS_역대_박스오피스_내역.xlsx',engine="openpyxl")
```

In [4]:
```python
movie = []

# 박스오피스 영화 순위 100
for x in xlsx["영화명"][0:100]:
    movie.append(x)

print(movie)
```

```
['명량', '극한직업', '신과함께-죄와 벌', '국제시장', '어벤져스: 엔드게임', '겨울왕국 2
', '아바타', '베테랑', '괴물', '도둑들', '7번방의 선물', '암살', '알라딘', '광해,
왕이 된 남자', '왕의 남자', '신과함께-인과 연', '택시운전사', '태극기 휘날리며', '부산
행', '해운대', '변호인', '어벤져스: 인피니티 워', '실미도', '어벤져스: 에이지 오브 울트
론', '기생충', '겨울왕국', '인터스텔라', '보헤미안 랩소디', '검사외전', '엑시트', '설
국열차', '관상', '아이언맨 3', '캡틴 아메리카: 시빌 워', '해적: 바다로 간 산적', '수
상한 그녀', '국가대표', '디워', '백두산', '과속스캔들', '스파이더맨: 파 프롬 홈', '웰
컴 투 동막골', '공조', '트랜스포머 3', '히말라야', '미션임파서블:고스트프로토콜', '트랜
스포머: 패자의 역습', '밀정', '최종병기 활', '트랜스포머', '써니', '화려한 휴가', '스
파이더맨: 홈 커밍', 1987, '베를린', '마스터', '터널', '어벤져스', '내부자들', '인천
상륙작전', '럭키', '은밀하게 위대하게', '곡성', '범죄도시', '타짜', '좋은 놈, 나쁜 놈
, 이상한 놈', '늑대소년', '미녀는 괴로워', '군함도', '미션 임파서블: 폴아웃', '다크 나
이트 라이즈', '아저씨', '사도', '전우치', '킹스맨 : 시크릿 에이전트', '미션 임파서블:
로그네이션', '투사부일체', '연평해전', '반지의 제왕 : 왕의 귀환', '인셉션', '레미제라블
', '쉬리', '캡틴 마블', '미션 임파서블 3', '쥬라기 월드: 폴른 킹덤', '청년경찰', '가
문의 위기(가문의 영광2)', '숨바꼭질', '덕혜옹주', '더 테러 라이브', '스파이더맨: 노 웨
이 홈', '쥬라기 월드', '감시자들', '의형제', 2012, '앤트맨과 와스프', '닥터 스트레인
지', '검은 사제들', '안시성', '블랙 팬서']
```

In [8]:
```python
## 영화 코드 추출
## 영화 검색 -> 영화 목록 리스트 -> 리스트 중 하나를 검색하면 제일 먼저 앞에 나오는 걸
data2 = []

driver = webdriver.Chrome()

# 영화 목록

#movie_list = open("영화코드.txt","w")

for x in movie:
    driver.get("https://movie.naver.com/movie/search/result.naver?sect
    time.sleep(1)
    q = driver.find_element(By.XPATH, "/html/body/div/div[4]/div/div/
    w = q.find_element(By.TAG_NAME,"a")
    e = w.get_attribute("href")
    data2.append(e[53:]) # data2에 영화 코드 저장
    #movie_list.write(str(e[53:])+"\n") # movie_list.txt에 한줄씩 코드 작

movie_list.close()
```

In [11]:
```python
# 검색해서 다른 영화를 지목하는 경우가 있음 ->  #movie_list 주석처리 + 일일이 수정해

#코드 재확인
for x,i in enumerate(data2):
    print(x,i)
```

```
0 133253
1 167651
2 85579
3 102875
4 136900
5 136873
6 84024
7 115977
8 210812
9 210927
10 94775
11 121048
12 163788
13 83893
14 39894
15 167697
16 146469
17 36666
18 130966
19 45321
20 101901
21 136315
22 34501
23 98438
24 161967
25 136873
26 149545
27 156464
28 130903
29 203021
30 62328
31 93728
32 70254
33 122527
34 102817
35 107924
36 141824
37 39569
38 187940
39 51143
40 173123
41 39405
42 142384
43 70241
44 100647
45 53372
46 68052
47 137952
48 83084
49 123630
50 212745
51 58018
52 135874
53 158191
54 158100
55 145162
56 199913
57 136900
58 121788
59 142822
60 196284
61 92575
62 121051
63 192608
64 96951
65 65674
66 135725
67 39157
68 146506
69 154222
70 72054
71 122851
72 121922
73 48227
74 114249
75 95541
76 41450
77 102272
78 31796
79 191735
80 191469
81 19500
82 132623
83 43153
84 154285
85 153652
86 41438
87 195326
88 94767
89 211939
90 208077
91 191646
92 98146
93 52548
94 110457
95 144330
96 182016
97 197843
98 163533
99 137326
```

In [51]:
```python
## 검색후 코드 잘못 부른 영화 일일이 다른 항목들 다 수정함 ㅠ

#명량 0 : 93756
#아바타 6 : 62266
#괴물 8 : 39841
#아미
#겨울왕국 25 : 100931
#인터스텔라 26 : 45290
#엑시트 29 : 174903
#국가대표 37 : 47385
#트랜스포머49 : 61521
#써니 50 : 76016
#베를린 54: 89218
#터널 56: 141104
#어벤저스 57: 72363
#럭키 60: 140695
#범죄도시 63: 161242
#타자 64: 57723
#느대소년 66: 88253
#아저씨 71: 71509
#인셉션 79: 52515
#레미제라블 80: 89755
#숨바꼭질 87: 102824
#더테러라이브 89: 99794
#쥬라기월드 91: 67786
#2012 94: 49727
#닥터스트렌인지 96: 125459
#검은사제들 97: 120157

#data2[9] = "78726"
```

In [52]:
```python
# movie_real.txt 만들어서 data2의 요소들을 작성하자
movie_real = open("영화목록_리얼.txt","w")
for x in data2:
    movie_real.write(x+"\n")
```

In [53]:
```python
len(data2)
```

Out[53]: 100

In [55]:
```python
## 영화 관련 정보 추출
driver = webdriver.Chrome()
data = []

page = 20 # 내가 읽고 싶은 페이지 수

for z in data2:
    for a in range(1,page+1): # 페이지1장 당 20개의 리뷰
        #다음 장 넘기기
        driver.get("https://movie.naver.com/movie/point/af/list.naver?
        time.sleep(2)
        # 영화 이름 20개 이름, 별점, 리뷰내용 추출
        b = driver.find_elements(By.CLASS_NAME,"title")
        for x in range(10):
            # 영화 이름
            name = b[x].text.split("\n")[0]

            # 영화 평점 별점
            num = b[x].text.split("\n")[2]

            # 영화 평점 내용
            review = b[x].text.split("\n")[3][:-3]

            data.append(np.asarray((name,num,review)))

data_df = pd.DataFrame(data,columns=['name','num','review'])
```

In [57]:
```python
data_df
```

Out[57]:

	name	num	review
0	명량	10	역대 대명작이라 자부합니다^^
1	명량	10	진짜 이건 우리나라 영화계에 한 획을 그을 정도에 명작입니다 우리가 역사를 배워야 ...
2	명량	3	국뽕 범벅, 거품, 스크린독점 관객동원
3	명량	10	초등 고학년 아들과 진도에 다녀온후 같이 봤는데 정말 감사하고 감동이며 죄송스러운 ...
4	명량	10	깊게 숨을 들이쉬고 후하고 내뱉은 뒤, 이 영화를 보아라. 다보고 나서 이틀 뒤에 ...
...	...	...	...
9995	블랙 팬서	10	
9996	블랙 팬서	10	와칸다의 왕, 그의 죽음은 끝이 아니길..
9997	블랙 팬서	10	와칸다 포에버!! 당신이 떠났어도 왕국을 지키겠습니다
9998	블랙 팬서	10	처음 볼 때는 그저 재미있는 영화라고 생각을 하였습니다. 그런데 암 투병 중에 촬영...
9999	블랙 팬서	10	R.I.P 채드윅. 영원한 블랙팬서

10000 rows × 3 columns

In [58]:
```python
data_df.to_excel("영화목록3.xlsx",engine="openpyxl")

#나중에 누가 받을때 오류나면 아마 uft-8 cp949 관련 문제이므로 encoding을 지정하면 해
```

In [ ]:
```python
######### 노-력 흔적들 #########
```

In [91]:
```python
## 노-력의 흔적

#c = driver.find_elements(By.XPATH,"/html/body/div/div[4]/div/div/div
#c = driver.find_element(By.CLASS_NAME,"search_list_1")

#href 태그 내 이름 추출

'''q = driver.find_element(By.XPATH, "/html/body/div/div[4]/div/div/d
w = q.find_element(By.TAG_NAME,"a")
e = w.get_attribute("href")
#print(e)
print(e[53:]) # 53번째 부터 숫자가 나옴'''
```

```
https://movie.naver.com/movie/bi/mi/basic.naver?code=167651
167651
```

Out[91]: 'for x in q:\n   print(x.text)'

In [9]:
```python
# 영화 이름 10개 이름, 별점, 리뷰내용
b = driver.find_elements(By.CLASS_NAME,"title")
for x in b:
    print(x.text)
```

```
스파이더맨: 노 웨이 홈
별점 - 총 10점 중
10
눈은 즐거웠습니다 신고
스파이더맨: 노 웨이 홈
별점 - 총 10점 중
10
신고
스파이더맨: 노 웨이 홈
별점 - 총 10점 중
10
삼스파가 모였을 때 정말 좋았어요. 신고
스파이더맨: 노 웨이 홈
별점 - 총 10점 중
6
기억에 남는 액션이 없어요.. 신고
스파이더맨: 노 웨이 홈
별점 - 총 10점 중
8
전반적으로 너무 재밌게 봤는데 이제는 마블 스토리가 점점 복잡해져요 신고
스파이더맨: 노 웨이 홈
별점 - 총 10점 중
8
주인공 개 극암덩이샛기임 신고
스파이더맨: 노 웨이 홈
별점 - 총 10점 중
10
기존의 스파이더맨은 이번편을 위하여 만들어진 것이라 해도 과언이 아닐 정도로 좋았습니다.스파이
더맨 전 시리즈 보시고 노웨이홈 보는거 추천드려요 ㅎㅎㅎ 신고
스파이더맨: 노 웨이 홈
별점 - 총 10점 중
8
아주 길었지만 시간 가는줄 모르고 재밌게 봤어요 신고
스파이더맨: 노 웨이 홈
별점 - 총 10점 중
10
ㅠ 지렸다 역대급 스파이더맨.. 신고
스파이더맨: 노 웨이 홈
별점 - 총 10점 중
10
존잼 신고
```

In [21]:
```python
# 영화 이름
print(b[0].text.split("\n")[0])

# 영화 평점 별점
print(b[0].text.split("\n")[2])

# 영화 평점 내용
print(b[0].text.split("\n")[3][:-3])
```

```
스파이더맨: 노 웨이 홈
10
눈은 즐거웠습니다
```

In [ ]: