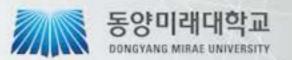


7주차 2차시

데이터 실전 분석 - 영화 평점 분석 [1]

32423432

77686787



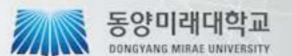
# 데이터 실전 분석 영화 평점 분석 [1]



# 학습개요

- 1/ 영화 평점 데이터 소개
- 2/ 영화 평점 분석 실습



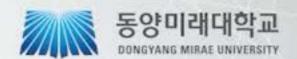




영화 평점 데이터 소개

# '영화 평점 데이터'







영화 평점 데이터 소개

# **→ 사용자 데이터 (약 6,000명)**

	사용자아이디	성별	연령	직업	지역
0	1	F	1	10	48067
1	2	M	56	16	70072
2	3	М	25	15	55117
3	4	M	45	7	02460
4	5	M	25	20	55455





(1) 영화 평점 데이터 소개

# **→ 영화 평점 데이러 (약 1,000,000건)**

	사용자아이디	영화아이디	평점	타임스탬프
0	1	1193	5	978300760
1	1	661	3	978302109
2	1	914	3	978301968
3	1	3408	4	978300275
4	1	2355	5	978824291

- ✓ 어떤 영화가 평점을 높게 받았는지 분석
- ✓ 특정 영화의 평점을 알기 어려움



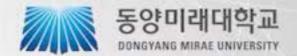


영화 평점 데이터 소개

# **→ 영화 데이터 (약 3,880건)**

	영화아이디	영화제목	장르
0	1	Toy Story (1995)	Animation Children's Comedy
1	2	Jumanji (1995)	Adventure Children's Fantasy
2	3	Grumpier Old Men (1995)	Comedy Romance
3	4	Waiting to Exhale (1995)	Comedy Drama
4	5	Father of the Bride Part II (1995)	Comedy



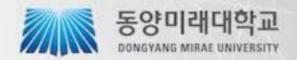




# 一 평점의 평균과 개수를 활용하여 보고 싶은 영화 찾기

	평균	개수	
영화제목			
American Beauty (1999)	4,317386	3428	
Godfather, The (1972)	4.524966	2223	
Matrix, The (1999)	4.315830	2590	
Princess Bride, The (1987)	4.303710	2318	
Raiders of the Lost Ark (1981)	4.477725	2514	
Saving Private Ryan (1998)	4.337354	2653	
Schindler's List (1993)	4.510417	2304	
Shawshank Redemption, The (1994)	4.554558	2227	
Silence of the Lambs, The (1991)	4.351823	2578	
Sixth Sense, The (1999)	4.406263	2459	
Star Wars: Episode IV - A New Hope (1977)	4.453694	2991	



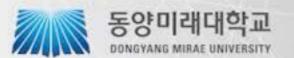




# **' 여자들이 좋아하는 영화 찾기**

	mean		count	
성별	F	М	F	М
영화제목				
American Beauty (1999)	4.238901	4.347301	946.0	2482.0
Being John Malkovich (1999)	4.159930	4.113636	569.0	1672.0
Braveheart (1995)	4.016484	4.297839	546.0	1897.0
Casablanca (1942)	4.300990	4.461340	505.0	1164.0
E.T. the Extra-Terrestrial (1982)	4.089850	3.920264	601.0	1668.0

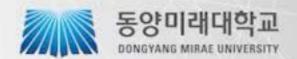






# **' 여자들이 좋아하는 영화들의** 장르 분석하기

Drama	12.0
Action	7.0
War	6.0
Comedy	6.0
Thriller	5.0
Adventure	5.0
Sci-Fi	4.0
Romance	4.0
Crime	3.0
Children's	3.0
Fantasy	2.0
Mystery	1.0
Musical	1.0
Film-Noir	1.0
Animation	1.0
	CACACA

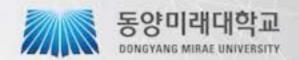




# 그 남녀의 호불호가 크게 갈리는 영화 찾기

	mean		count		평점차이
성별	F	м	F	М	
영화제목					
Dirty Dancing (1987)	3.790378	2.959596	291.0	396.0	0.830782
Good, The Bad and The Ugly, The (1966)	3.494949	4.221300	99.0	723.0	0.726351
Dumb & Dumber (1994)	2.697987	3.336595	149.0	511.0	0.638608
Evil Dead II (Dead By Dawn) (1987)	3.297297	3.909283	74.0	474.0	0.611985
Grease (1978)	3.975265	3.367041	283.0	534.0	0.608224
Caddyshack (1980)	3.396135	3.969737	207.0	760.0	0.573602
Animal House (1978)	3.628906	4.167192	256.0	951.0	0.538286
Exorcist, The (1973)	3.537634	4.067239	186.0	699.0	0.529605
Rocky Horror Picture Show, The (1975)	3.673016	3.160131	315.0	918.0	0.512885
Big Trouble in Little China (1986)	2.987952	3.485030	83.0	501.0	0.497078
1.00					

✓ 남자와 여자의 평점이 차이가 큰 영화





# 

연령대	10대 미만	10 <sup>CH</sup>	20대	30EH	40 <sup>C</sup> H	50대 이상
영화제목						
\$1,000,000 Duck (1971)	-	3	3.09091	3.13333	2	2.75
'Night Mother (1986)	2	4.66667	3.42308	2.90476	3.83333	3.75
'Til There Was You (1997)	3.5	2.5	2.66667	2.9	2.33333	2.6
'burbs, The (1989)	4.5	3.24444	2.65217	2.81818	2.54545	3.1
And Justice for All (1979)	3	3.42857	3.72414	3.65714	4.1	3.67442

- ✓ 연령 정보를 통해 연령대 계산
- ✓ 각 연령대의 평점 계산

### import

### 영화 평점 분석 실습

```
In [1]: import pandas as pd
from pandas import Series, DataFrame
import numpy as np
```

### 1. 영화 평점 데이터 적재 및 전처리

```
In []: # 사용자 데이터 읽어오기
users = pd.read_csv('data/movielens/users.dat', sep = '::', engine = 'python',
names = ['사용자아이디', '성별','연령','직업','지역'])
users.head()
```

```
In []: #평점 데이터 읽어오기
ratings = pd.read_csv('data/movielens/ratings.dat', sep = '::', engine = 'python',
names = ['사용자아이디', '영화아이디', '평점', '타임스탬프'])
ratings.head()
```

In []: #영화데이터 읽어오기

### ◎ 사용자 데이터(users), 평점 데이터(ratings), 영화 데이터(movies) 읽어오기

영화 데이터 인자값 살펴보기 "mes = ['사용자아이디', '영화아이디', '평점', '타임스탬프'])

Out[3]:

	사용자아이디	영화아이디	평점	타임스탬프
0	1	1193	5	978300760
1	1	661	3	978302109
2	1	914	3	978301968
3	1	3408	4	978300275
4	1	2355	5	978824291

```
In [4]: # 영화데이터 읽어오기
movies = pd.read_csv('data/movielens/movies.dat', sep = '::', engine = 'python',
names = ['영화아이디','영화제목','장르'], encoding = 'latin-1')
movies.head()
```

Out [4]:

장르	영화제목	├O├O □ □	영호
Animation Children's Comedy	Toy Story (1995)	1	0
Adventure Children's Fantasy	Jumanji (1995)	2	1

### ◎ 사용자 데이터(users), 평점 데이터(ratings), 영화 데이터(movies) 읽어오기

```
2355
                                     5 978824291
In [4]: # 영화데이터 읽어오기
        movies = pd.read_csv('data/movielens/movies.dat', sep = '::', engine = 'python',
                            names = ['영화아이디','영화제목','장르'], encoding = 'latin-1')
        movies.head()
Out[4]:
            영화아이디
                                                                     장르
                                         영화제목
                                   Toy Story (1995) Animation|Children's|Comedy
                                    Jumanji (1995)
                                                 Adventure|Children's|Fantasy
         2
                            Grumpier Old Men (1995)
                                                          Comedy|Romance
                                                            Comedy|Drama
                             Waiting to Exhale (1995)
                    5 Father of the Bride Part II (1995)
                                                                  Comedy
```

4 9/83002/5

In []: #3개의 데이터프레임을 하나로 합치기

### 2. 보고 싶은 영화 찾기

역하들이 편전 편규은 구하여 사라들에게 이전반는 (편전이 논은) 역하 찬기

### ◎ 3개의 데이터프레임을 하나로 합치기

- ratings 데이터를 기반으로 영화아이디, 영화제목을 같이 봐야 할 경우
- 사용자아이디를 users의성별,나이정보를 같이 평점을 분석할 경우

```
names = ['영화아이디', '영화제목', '장르'], encoding = 'latin-1')
        movies.head()
Out[4]:
            영화아이디
                                         영화제목
                                                                      장르
                                   Toy Story (1995)
                                                  Animation|Children's|Comedy
                    2
                                    Jumanji (1995)
                                                  Adventure|Children's|Fantasy
         2
                            Grumpier Old Men (1995)
                                                          Comedy|Romance
                    3
                                                             Comedy|Drama
         3
                             Waiting to Exhale (1995)
                    5 Father of the Bride Part II (1995)
                                                                  Comedy
         4
        #3개의 데이터프레임을 하나로 합치기
```

978300275

### 2. 보고 싶은 영화 찾기

### ◎ 3개의 데이터프레임을 하나로 합치기

```
✔ merge:사용자아이디를기준으로합침
```

✔ data에 users, ratings 2개의 dataframe을 합침

```
', sep = '::', engine = 'python',
배목','장르'], encoding = 'latin-1')
```

```
영화아이디
                                         영화제목
                                                                     장르
                                   Toy Story (1995)
                                                  Animation|Children's|Comedy
                    2
                                    Jumanji (1995)
                                                  Adventure|Children's|Fantasy
                            Grumpier Old Men (1995)
                                                          Comedy|Romance
         2
                             Waiting to Exhale (1995)
                                                            Comedy|Drama
         3
                    5 Father of the Bride Part II (1995)
                                                                  Comedy
In [5]:
        #3개의 데이터프레임을 하나로 합치기
        data = pd.merge(users, ratings)
        data
Out [5]:
                  사용자아이디
                                         직업
                                                    영화아이디
                                               지역
                                                                    타임스탬프
                                                                  5 978300760
                                              48067
                                                          1193
               0
```

48067

661

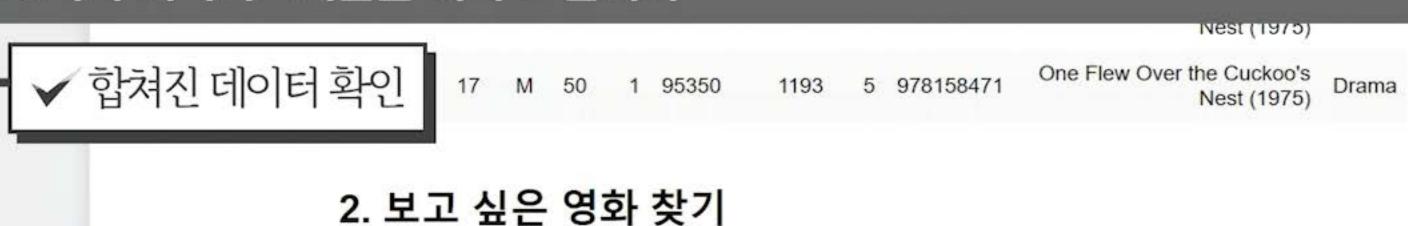
3 978302109

### ◎ 3개의 데이터프레임을 하나로 합치기

- ✔ data에 추가로 movies의 data를 합침
- ✔ on 인자는사용하지 않아도 영화아이디 공통 columns이 있으므로 상관 없음

```
Waiting to Exhale (1995)
                                                           Comedy|Drama
         3
                   5 Father of the Bride Part II (1995)
                                                                 Comedy
        #3개의 데이터프레임을 하나로 합치기
        data = pd.merge(users, ratings)
        data
                                          . . .
In [6]: data = pd.merge(data, movies)
In [7]:
       data
Out[7]:
                                    지역
                                                  타임스탬프
                                                                  영화제목
                                                                                          장르
                                                             One Flew Over
```

### ◎ 3개의 데이터프레임을 하나로 합치기



영화들의 평점 평균을 구하여, 사람들에게 인정받는 (평점이 높은) 영화 찾기

In [ ]:	# 영화들의 평점 평균을 구하여, 평점이 높은 영화 찾기
In [ ]:	
	평균 평점이 만점인 영화들이 최상위에 위치함. 일반적으로 평점이 만점인 경우는 대부분 평점
	개수가 매우 적은 경우이므로, 이를 확인하기 위해 평점의 개수도 함께 구해본다.
In [ ]:	

### [실습 #1] 여자들이 좋아하는 영화 찾기

- 여성 평점이 4.0 이상이고 여성 평점의 개수가 500개 이상인 영화

# <mark>보고 싶은 영화 찾기</mark> 평점의 평균과 개수를 활용하여 보고 싶은 영화 찾기

◎ 영화들의 평점 평균을 구하여, 평점이 높은 영화 찾기

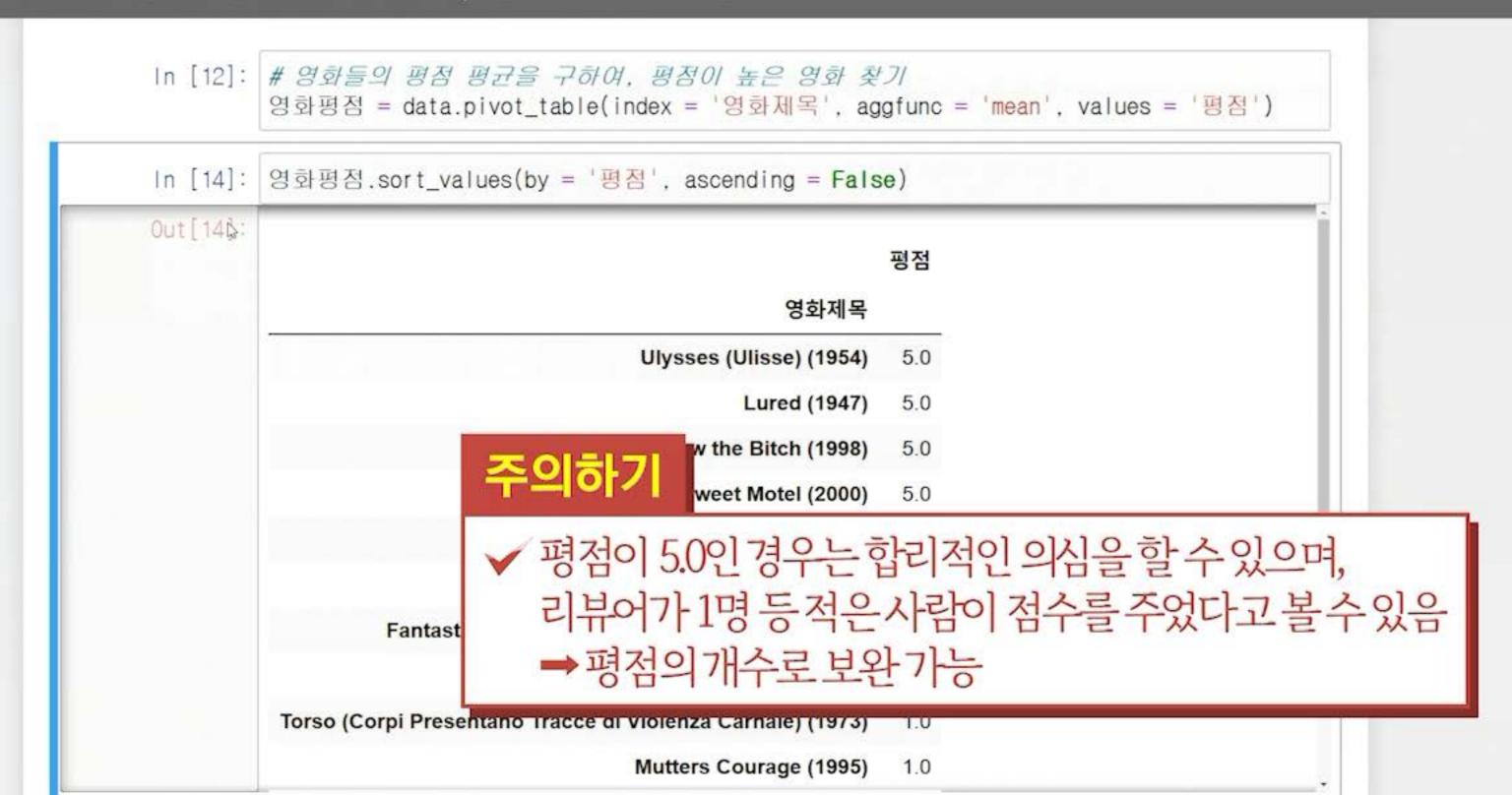


영화별로 평점이 매겨지는 것을 확인해 볼 수 있습니다.

Zeus and Roxanne (1997) 2.521739

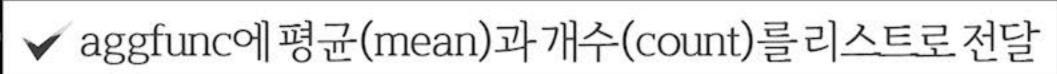
# 보고 싶은 영화 찾기 평점의 평균과 개수를 활용하여 보고 싶은 영화 찾기

◎ 영화들의 평점 평균을 구하여, 평점이 높은 영화 찾기



# <mark>보고 싶은 영화 찾기</mark> 평점의 평균과 개수를 활용하여 보고 싶은 영화 찾기

### ◎ 평점에 평점의 개수 확인하기



nean'. values = '평점')

```
In [14]: 영화평점.sort_values(by = '평점', ascending = False)
```

평균 평점이 만점인 영화들이 최상위에 위치함. 일반적으로 평점이 만점인 경우는 대부분 평점의 개수가 매우 적은 경우이므로, 이를 확인하기 위해 평점의 개수도 함께 구해본다.

```
In []: 영화평점 = data.pivot_table(index = '영화제목', aggfunc = ['mean', 'count'], values =
```

### [실습 #1] 여자들이 좋아하는 영화 찾기

- 여성 평점이 4.0 이상이고 여성 평점의 개수가 500개 이상인 영화

```
In [ ]:
In [ ]:
```

# 보고 싶은 영화 찾기 평점의 평균과 개수를 활용하여 보고 싶은 영화 찾기

### ◎ 평점에 평점의 개수 확인하기



✔ 정렬시 복사하여 사용하면 에러 발생

: 컬럼명이 2줄로 되어 있는 경우에는 계층 색인이 되므로 튜플의 형태를 사용해야 함

[실습 #1] 여자들이 좋아하는 영화 찾기

### 보고 싶은 영화 찾기

# 평점의 평균과 개수를 활용하여 보고 싶은 영화 찾기

### ◎ 기준 세우기

```
✔ 평점이 높은 영화 or 평점의 개수가 많은 영화 or 하이브리드
```

✔ 예: 평점 4.3이상이며, 평점의 개수가 2,000개 이상인 영화

3706 rows × 2 columns

In [19]: # 평점평균이 4.3 이상이고, 평점의 개수가 2000개 이상인 영화를 찾기 영화평점

Out[19]:

		mean	count
C <sub>2</sub>		평점	평점
٩	명화제목		
\$1,000,000 Duc	k (1971)	3.027027	37
'Night Mothe	r (1986)	3.371429	70
'Til There Was You	u (1997)	2.692308	52
'burbs, The	e (1989)	2.910891	303
And Justice for A	II (1979)	3.713568	199

# 보고 싶은 영화 찾기 평점의 평균과 개수를 활용하여 보고 싶은 영화 찾기

◎ 평점 4.3이상이며, 평점의 개수가 2,000개 이상인 영화 찾기

```
✔ 평점평균, 평점개수 2개의 columns으로 보기 좋게 변경
```

In [20]: # 평점평균이 4.3 이상이고, 평점의 개수가 2000개 이상인 영화를 찾기 영화평점.columns = ['평점평균', '평점개수'] 영화평점

Out [20]:

평점평균 평점개수

### 영화제목

	\$1,000,000 Duck (1971)	3.027027	37		
	'Night Mother (1986)	3.371429	70		
	'Til There Was You (1997)	2.692308	52		
	'burbs, The (1989)	2.910891	303		
	And Justice for All (1979)	3.713568	199		
	•••				
	Zed & Two Noughts, A (1985)	3.413793	29		
	Zero Effect (1998)	3.750831	301		
Zero Ke	lvin (Kiærlighetens kiøtere) (1995)	3.500000	2		

# 보고 싶은 영화 찾기 평점의 평균과 개수를 활용하여 보고 싶은 영화 찾기

(1980)

◎ 평점 4.3이상이며, 평점의 개수가 2,000개 이상인 영화 찾기

```
조건 색인: 평점평균>=4.3 & 평점개수>=2000
          In [21]: # 평점평균이 4.3 이상이고, 평점의 개수가 2000개 이상인 영화를 찾기
                  영화평점.columns = ['평점평균', '평점개수']
                  (영화평점.평점평균 >= 4.3) & (영화평점.평점개수 >= 2000)
          Out [21]: 영화제목
                  $1,000,000 Duck (1971)
                                                         False
                  'Night Mother (1986)
                                                         False
                  'Til There Was You (1997)
                                                         False
                  'burbs, The (1989)
                                                         False
                  ...And Justice for All (1979)
                                                         False
                                                          0.000
                  Zed & Two Noughts, A (1985)
                                                         False
                  Zero Effect (1998)
                                                         False
                  Zero Kelvin (Kjærlighetens kjøtere) (1995) False
                  Zeus and Roxanne (1997)
                                                         False
                  eXistenZ (1999)
                                                         False
                  Length: 3706, dtype: bool
```

[실습 #1] 여자들이 좋아하는 영화 찾기

### <mark>보고 싶은 영화 찾기</mark> 평점의 평균과 개수를 활용하여 보고 싶은 영화 찾기

◎ 평점 4.3이상이며, 평점의 개수가 2,000개 이상인 영화 찾기

조건 색인: 평점평균>=4.3 & 평점개수>=2000

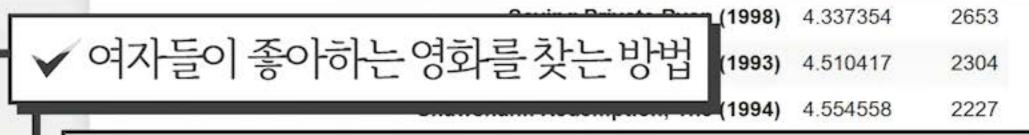
평점.평점개수 >= 2000)]

평점평균	평점개수
	O - 11 1

		영화제목
3428	4.317386	American Beauty (1999)
2223	4.524966	Godfather, The (1972)
2590	4.315830	Matrix, The (1999)
2318	4.303710	Princess Bride, The (1987)
2514	4.477725	Raiders of the Lost Ark (1981)
2653	4.337354	Saving Private Ryan (1998)
2304	4.510417	Schindler's List (1993)
2227	4.554558	Shawshank Redemption, The (1994)
2578	4.351823	Silence of the Lambs, The (1991)
2459	4.406263	Sixth Sense, The (1999)

이 조건을 충족하는 11개의 영화만 볼 수가 있습니다.

◎ 여성 평점이 4.0 이상이고 여성 평점의 개수가 500개 이상인 영화



①모든영화평점 데이터에서 여성 평점 데이터만 선택하여 영화별로

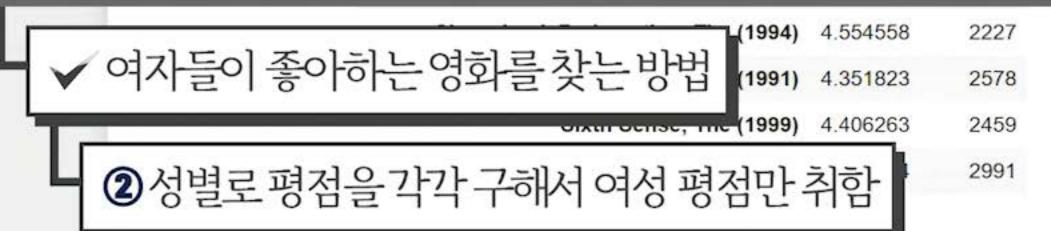
Star Wars: Episode IV - A New Hope (1977) 4.453694 2991

### [실습 #1] 여자들이 좋아하는 영화 찾기

- 여성 평점이 4.0 이상이고 여성 평점의 개수가 500개 이상인 영화

```
In [23]:
          data
Out [23]:
                     사용
자아
                                             아이
                                       지역
                                                      타임스탬프
                                                                       영화제목
                                                                                                장르
                     이디
                                                                  One Flew Over
                                                                   the Cuckoo's
                 0
                                             1193
                                                      978300760
                                                                                              Drama
                                                                    Nest (1975)
```

◎ 여성 평점이 4.0 이상이고 여성 평점의 개수가 500개 이상인 영화

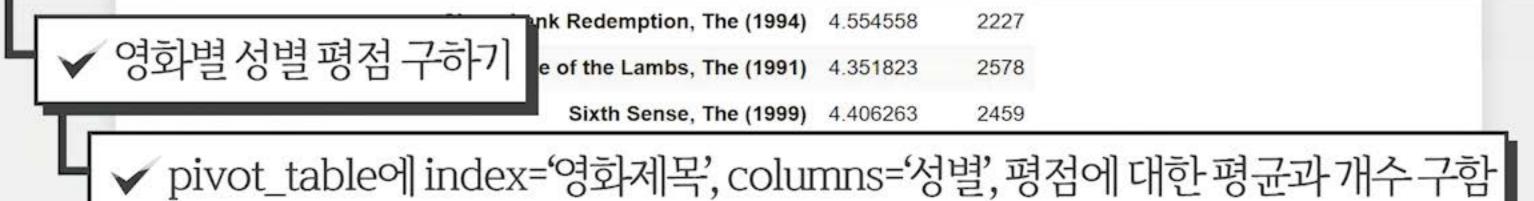


### [실습 #1] 여자들이 좋아하는 영화 찾기

- 여성 평점이 4.0 이상이고 여성 평점의 개수가 500개 이상인 영화 ¶

In [23]:	data										
Out [23]:		사용 자아 이디	성 별	연 령	직업	지역	영화 아이 디	평 점	타임스탬프	영화제목	장르
	0	1	F	1	10	48067	1193	5	978300760	One Flew Over the Cuckoo's Nest (1975)	Drama
	1	2	М	56	16	70072	1193	5	978298413	One Flew Over the Cuckoo's Nest (1975)	Drama

◎ 여성 평점이 4.0 이상이고 여성 평점의 개수가 500개 이상인 영화



### [실습 #1] 여자들이 좋아하는 영화 찾기

- 여성 평점이 4.0 이상이고 여성 평점의 개수가 500개 이상인 영화

```
data.pivot_table(index = '영화제목', columns = '성별',
                         aggfunc = ['mean', 'count'], values = '#0x')
Out [23]:
                                          영화
아이
                                    지역
                                                   타임스탬프
                                                                  영화제목
                                                                                          장르
                   이디
                                                              One Flew Over
                                                  978300760
                                                               the Cuckoo's
                               10 48067
                                          1193
                                                                                        Drama
                                                                Nest (1975)
                                                              One Flew Over
                                                5 978298413
                                                               the Cuckoo's
                      2 M 56 16 70072 1193
                                                                                        Drama
```

### ◎ 여성 평점이 4.0 이상이고 여성 평점의 개수가 500개 이상인 영화

- 여성 당염이 4.0 이성이고 여성 당염의 개구기 500개 이성인 당와

```
In [24]: = data.pivot_table(index = '영화제목', columns = '성별',
                          aggfunc = ['mean', 'count'], values = '평점')
Out [24]:
                                                                  count
                                               mean
          성별
                                                                       M
                                    영화제목
                          $1,000,000 Duck (1971) 3.375000 2.761905 16.0
                            'Night Mother (1986) 3.388889 3.352941 36.0
                        'Til There Was You (1997) 2.675676 2.733333 37.0
                              'burbs, The (1989) 2.793478 2.962085 92.0 211.0
                       ...And Justice for All (1979) 3.828571 3.689024 35.0 164.0
                     Zed & Two Noughts, A (1985) 3.500000 3.380952
```

평점의 평균과 개수가 각각 female과 male로 남성과 여성으로 각각 나뉘어서 데이터가 저장된 것을 확인해 볼 수 있습니다.

◎ 여성 평점이 4.0 이상이고 여성 평점의 개수가 500개 이상인 영화

I 심급 #1I 여사들이 좋아하는 영화 잦기

- ✔ 조건 색인
- ✔ 튜플의형태로, (mean, '여성'), 즉여성평점의평균을선택
- ✔ 4.0이상인경우만 True가 나옴

개 이상인 영화

; )

```
In [27]: (ex1[('mean', 'F')]>= 4.0
Out [27]: 영화제목
         $1,000,000 Duck (1971)
                                                       False
         'Night Mother (1986)
                                                       False
         'Til There Was You (1997)
                                                       False
         'burbs, The (1989)
                                                       False
         ... And Justice for All (1979)
                                                       False
         Zed & Two Noughts, A (1985)
                                                       False
         Zero Effect (1998)
                                                       False
         Zero Kelvin (Kjærlighetens kjøtere) (1995)
                                                        False
         Zeus and Roxanne (1997)
                                                       False
         eXistenZ (1999)
                                                       False
         Name: (mean, F), Length: 3706, dtype: bool
 In [ ]:
```

◎ 여성 평점이 4.0 이상이고 여성 평점의 개수가 500개 이상인 영화

I질급 #1I 여자들이 좋아하는 영화 잦기

✔ &로조건색인추가

In [ ]:

- ✔ 튜플의형태로, ('count', '여성'), 즉여성평점의개수를선택
- ✔ 2개를모두만족하는경우만 True가 나옴

개 이상인 영화

i

```
In [28]: ex1[(ex1[('mean', 'F')]>= 4.0) & (ex1[('count', 'F')]>= 500)
Out [28]: 영화제목
         $1,000,000 Duck (1971)
                                                       False
         'Night Mother (1986)
                                                       False
                                                       False
         'Til There Was You (1997)
         'burbs, The (1989)
                                                       False
         ... And Justice for All (1979)
                                                       False
         Zed & Two Noughts, A (1985)
                                                       False
         Zero Effect (1998)
                                                       False
         Zero Kelvin (Kjærlighetens kjøtere) (1995)
                                                        False
         Zeus and Roxanne (1997)
                                                       False
         eXistenZ (1999)
                                                       False
         Length: 3706, dtype: bool
```

### ◎ 여성 평점이 4.0 이상이고 여성 평점의 개수가 500개 이상인 영화



# 보고 싶은 영화 찾기 여성인기영화의 장르 분석

- ◎ 장르 정보는 movies에 존재
- ◎ 전체 영화에 대한 장르가 아닌 여성인기영화에 해당하는 장르만 필요함

### 식애 모사.

여성인기영화의 장르 통계 구하기

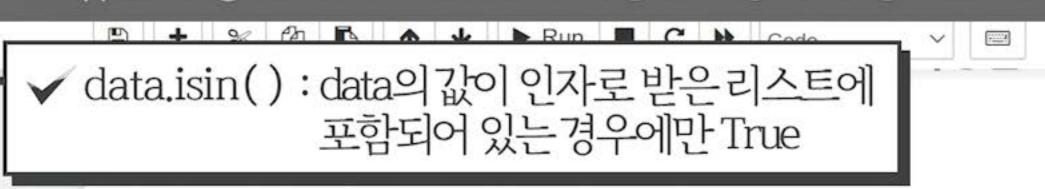
예를 들어, 여성인기영화 중 Drama 장르의 영화는 10개, Action 영화는 3개, ...

In [36]: movies I

Out [36]:

장르	영화제목	영화아이디		
Animation Children's Comedy	Toy Story (1995)	1	0	
Adventure Children's Fantasy	Jumanji (1995)	2	1	
Comedy Romance	Grumpier Old Men (1995)		2	
Comedy Drama	Waiting to Exhale (1995)	4	3	
Comedy	Father of the Bride Part II (1995)	5	4	
227	(202)		144	
Comedy	Meet the Parents (2000)	3948	3878	
Drama	Requiem for a Dream (2000)	3949	3879	
Drama	Tigerland (2000)	3950	3880	

### ◎ isin()을 사용하여 movies에서 여성인기영화의 장르만 선택



예를 들어, 여성인기영화 중 Drama 장르의 영화는 10개, Action 영화는 3개, ...

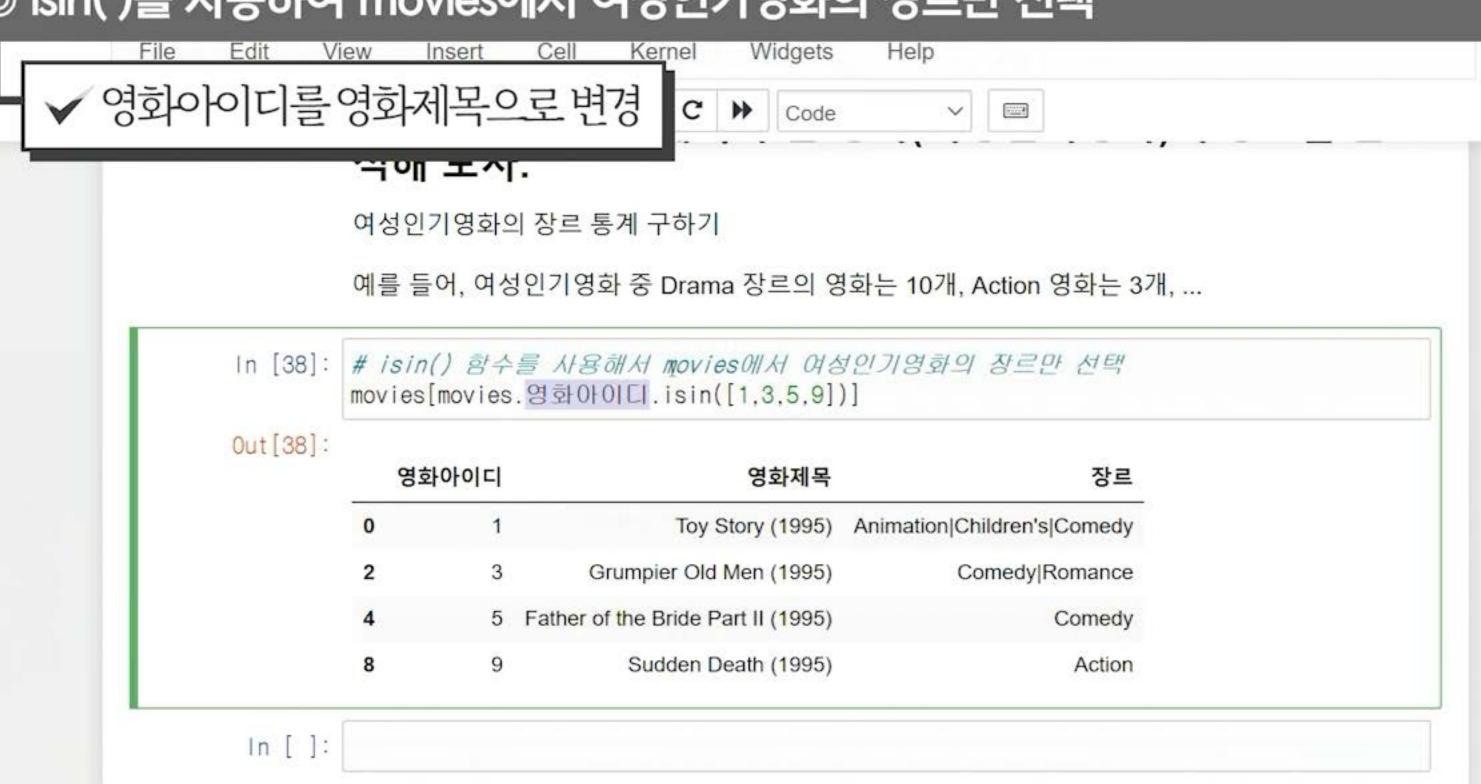
In [36]: # isin() 함수를 사용해서 movies에서 여성인기영화의 장르만 선택 movies.영화아이디.isin([1,3,5,9])

Out [36]:

장르	영화제목	영화아이디	
Animation Children's Comedy	Toy Story (1995)	1	0
Adventure Children's Fantasy	Jumanji (1995)	2	1
Comedy Romance	Grumpier Old Men (1995)	3	2
Comedy Drama	Waiting to Exhale (1995)	4	3
Comedy	Father of the Bride Part II (1995)	5	4
***	0.00	***	
Comedy	Meet the Parents (2000)	3948	3878
Drama	Requiem for a Dream (2000)	3949	3879

In [ ]:

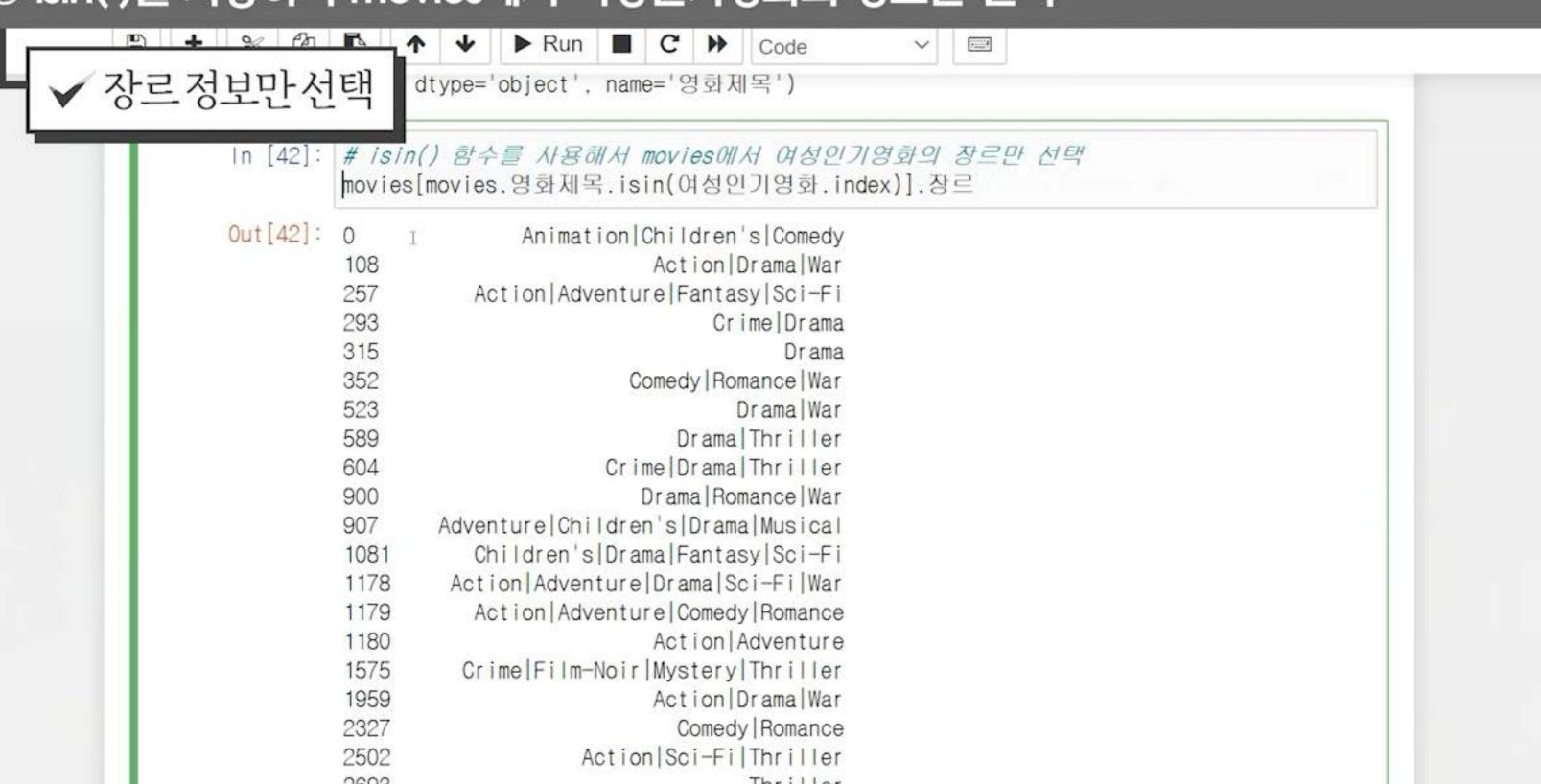
### ◎ isin()을 사용하여 movies에서 여성인기영화의 장르만 선택



◎ isin()을 사용하여 movies에서 여성인기영화의 장르만 선택

```
□ + 90 Pm FA A J Run ■ C NA Code
✔ 1,3,5,9대신 여성인기영화.index로 바꿈
                     예를 들어, 여성인기영화 중 Drama 장르의 영화는 10개, Action 영화는 3개, ...
             In [40]: 여성인기영화.index
             Out[40]: Index(['American Beauty (1999)', 'Being John Malkovich (1999)',
                            Braveheart (1995)', 'Casablanca (1942)',
                            'E.T. the Extra-Terrestrial (1982)', 'Fargo (1996)',
                            'Forrest Gump (1994)', 'L.A. Confidential (1997)', 'Matrix, The (1999)',
                            'Princess Bride, The (1987)', 'Pulp Fiction (1994)',
                            'Raiders of the Lost Ark (1981)', 'Saving Private Ryan (1998)',
                            'Schindler's List (1993)', 'Shakespeare in Love (1998)',
                            'Shawshank Redemption, The (1994)', 'Silence of the Lambs, The (1991)',
                            'Sixth Sense, The (1999)', 'Star Wars: Episode IV - A New Hope (1977)',
                            'Star Wars: Episode V - The Empire Strikes Back (1980)',
                            'Toy Story (1995)', 'Wizard of Oz, The (1939)'].
                           dtype='object', name='영화제목')
             In [38]: # isin() 함수를 사용해서 movies에서 여성인기영화의 장르만 선택
                     movies[movies.영화제목.isin([1,3,5,9])]
             Out [38]:
                         영화아이디
                                                  영화제목
                                                                           장르
```

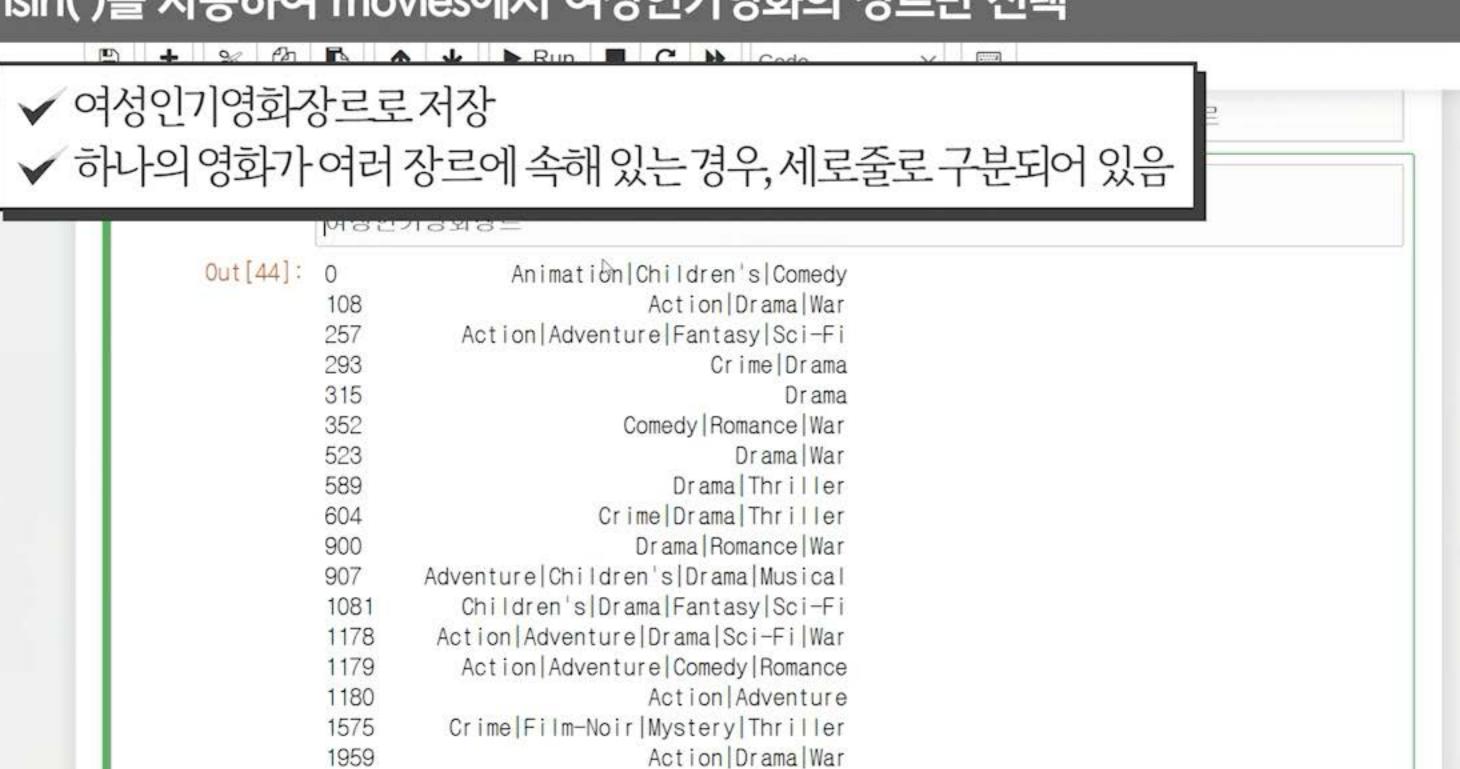
### ◎ isin()을 사용하여 movies에서 여성인기영화의 장르만 선택



2327

2502

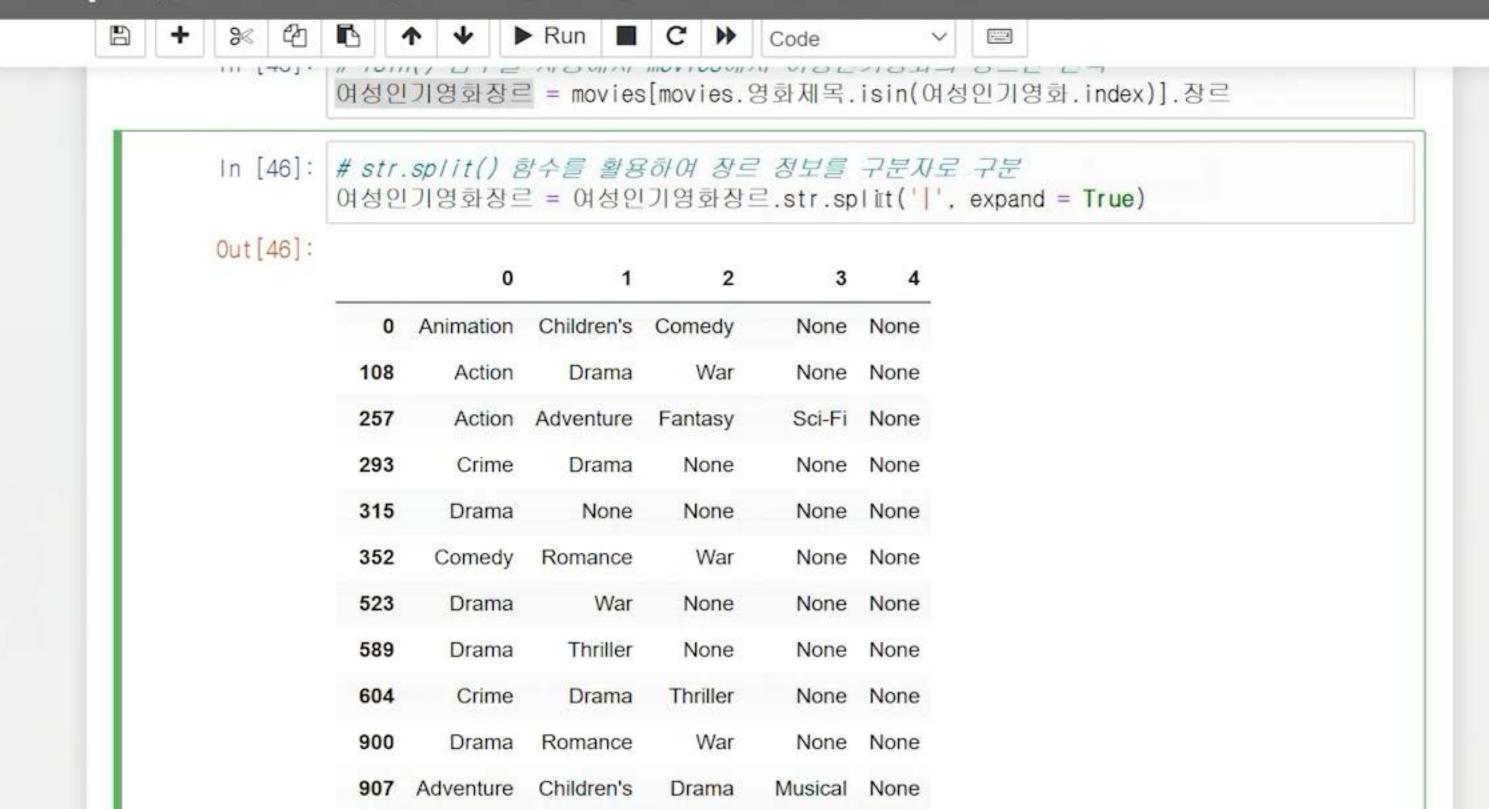
◎ isin()을 사용하여 movies에서 여성인기영화의 장르만 선택



Comedy Romance

Action|Sci-Fi|Thriller

### ◎ str.split() 함수를 활용하여 장르 정보를 구분자로 구분



### ◎ 영화장르 카운트하기

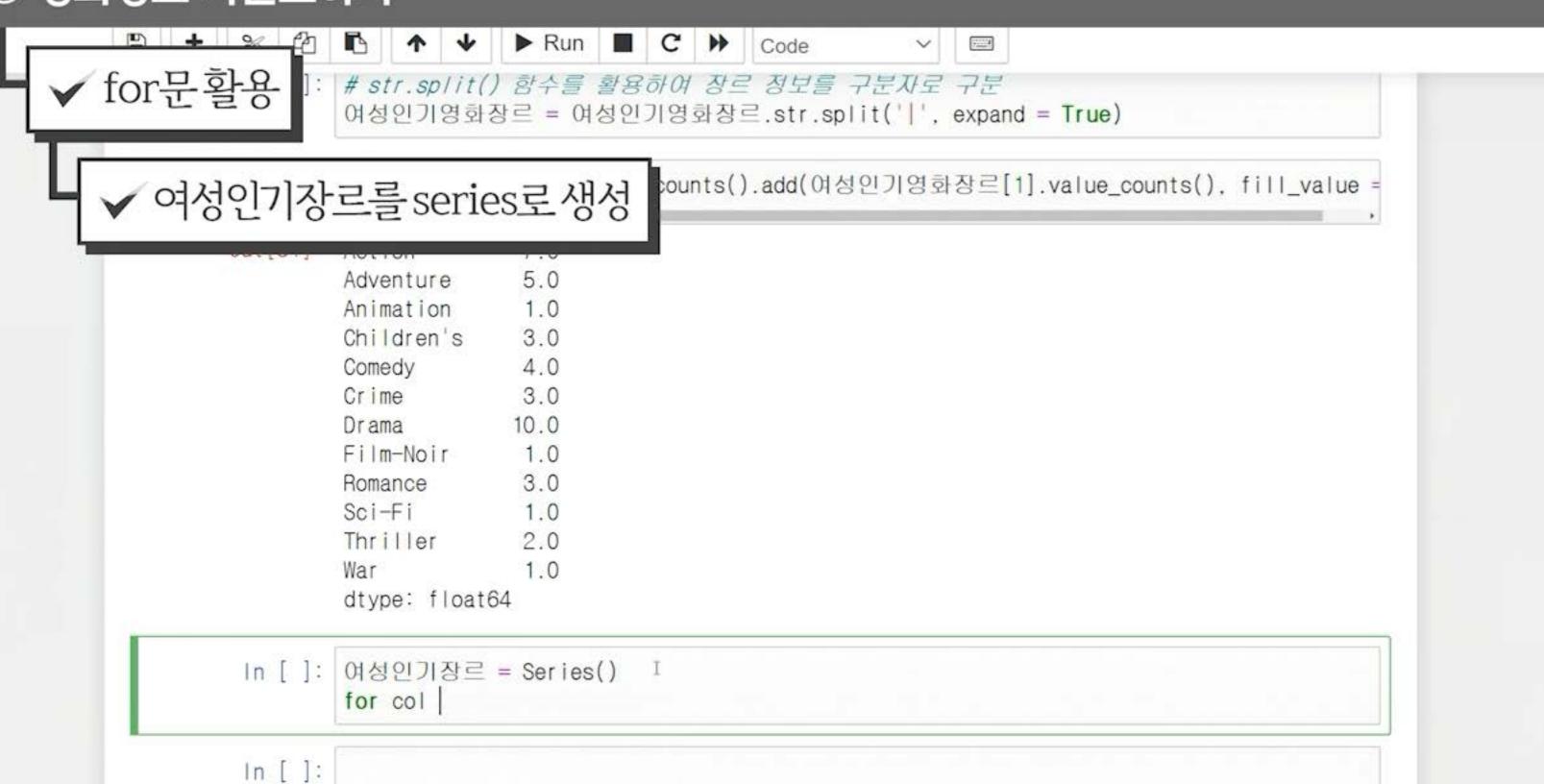
In [ ]:

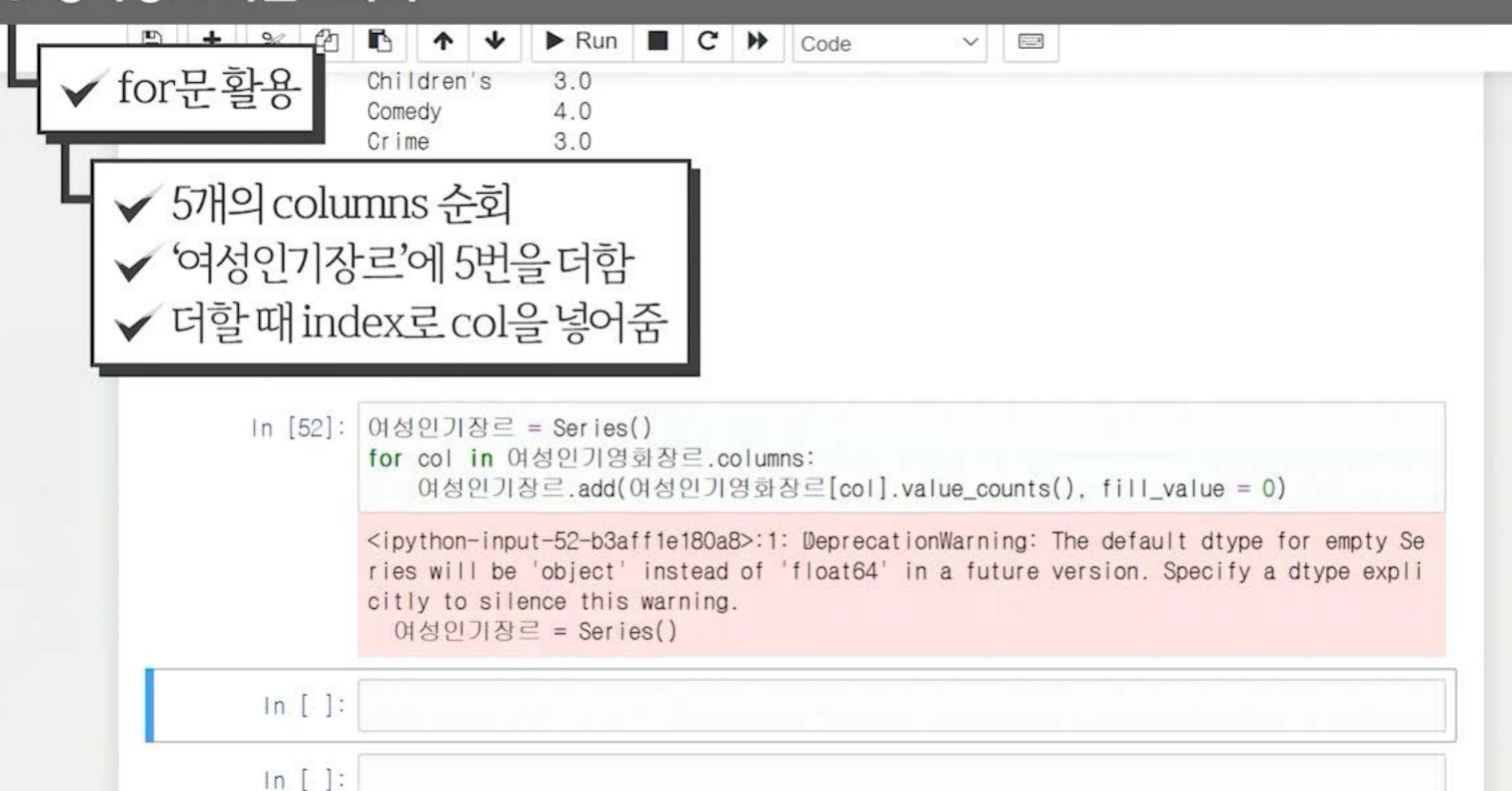


### 보고 싶은 영화 찾기 여성

### 여성인기영화의 장르 분석

```
15 + 90 Pm R. A
                             ▶ Run ■ C >>
                          4
                                                             ----
                                               Code
▶ +를 add로 변경
                      인기영화장르 = movies[movies.영화제목.isin(여성인기영화.index)].장르
fill value = 0
                     tr.split() 함수를 활용하여 장르 정보를 구분자로 구분
                   여행인기영화장르 = 여성인기영화장르.str.split('|', expand = True)
                  여성인기영화장르[0].value_counts().add(여성인기영화장르[1].value_counts(), fill_value =
           In [51]:
          Out[51]: Action
                               7.0
                               5.0
                  Adventure
                              1.0
                  Animation
                              3.0
                  Children's
                               4.0
                  Comedy
                  Crime
                               3.0
                              10.0
                  Drama
                  Film-Noir
                              1.0
                               3.0
                  Romance
                  Sci-Fi
                               1.0
                  Thriller
                               2.0
                  War
                               1.0
                  dtype: float64
           In [ ]:
```





E + 90 Pm R A L Run E C N Codo

#### ◎ 영화장르 카운트하기

In [ ]:

✓ DeprecationWarning : series 생성 시 object로 기본 데이터 타입이 설정되므로 float로 하기 위해서는 float로 변경하라는 뜻

```
FIIM-Noir
                      3.0
        Romance
        Sci-Fi
                      1.0
        Thriller
                      2.0
                      1.0
        War
        dtype: float64
In [52]: 여성인기장르 = Series()
        for col in 여성인기영화장르Icolumns:
            여성인기장르.add(여성인기영화장르[col].value_counts(), fill_value = 0)
        <ipython-input-52-b3aff1e180a8>:1: DeprecationWarning: The default dtype for empty Se
        ries will be 'object' instead of 'float64' in a future version. Specify a dtype expli
        citly to silence this warning.
          여성인기장르 = Series()
In [ ]:
```

```
P + 90 Pm R A 4 PRUD P Code
DeprecationWarning을 없애기 위해서는 dtype을 float64로 지정
                             10.0
                 Drama
                 Film-Noir
                             1.0
                             3.0
                 Romance
                 Sci-Fi
                             1.0
                 Thriller
                             2.0
                             1.0
                 War
                 dtype: float64
         In [52]: 여성인기장르 = Series(dtype = 'float64')
                 for col in 여성인기영화장르.columns:
                    여성인기장르.add(여성인기영화장르[col].value_counts(), fill_value = 0)
                 <ipython-input-52-b3aff1e180a8>:1: DeprecationWarning: The default dtype for empty Se
                 ries will be 'object' instead of 'float64' in a future version. Specify a dtype expli
                 citly to silence this warning.
                  여성인기장르 = Series()
          In [ ]:
          In [ ]:
```

