

レシピ34 複数のBoolean条件の構築(P91)

- pythonの論理演算子： and , or , not はpandasでは使えない
- pandasの論理演算子： & , | , ~ を使う
- さらに各論理式を括弧で括らないとエラーになる
 - 理由：
 - 比較演算子 < , > , == , != , <= , >= は、論理演算子 and , or , not よりも優先度が高い
 - pandasの新たな演算子 & , | , ~ はそれら比較演算子よりも優先度が高い
- データセットに対し適切なフィルタを適用するには、複数の論理式を組み合わせる部分集合を正確に抽出しなければならない
- このレシピでは、複数の論理式を作り、それらを組み合わせ、以下の条件の映画を求める
 - imdb_scoreが8より高く、
 - content_ratingがPG-13、
 - title_yearが2000年より前 または 2009年より後の映画

(1) 読み込む

```
In [4]: import pandas as pd
movie = pd.read_csv('movie.csv', index_col='movie_title')
movie.head(2)
```

Out[4]:

	color	director_name	num_critic_for_reviews	duration	director_facebook_likes	actor_3_facebook_likes
movie_title						
Avatar	Color	James Cameron	723.0	178.0	0.0	855.0
Pirates of the Caribbean: At World's End	Color	Gore Verbinski	302.0	169.0	563.0	1000.0

2 rows × 7 columns

(2) 基準を設け、変数に格納する

```
In [13]: criteria1 = movie.imdb_score > 8
criteria2 = movie.content_rating == 'PG-13'
# 括弧で括る(pandasの|は比較演算子より優先度が高いため先に計算されてしまうため)
criteria3 = ((movie.title_year < 2000) | (movie.title_year > 2009))

criteria2.head()
```

Out[13]:

movie_title	
Avatar	True
Pirates of the Caribbean: At World's End	True
Spectre	True
The Dark Knight Rises	True
Star Wars: Episode VII - The Force Awakens	False
Name: content_rating, dtype: bool	

(3) (2) を組み合わせ、1つのBooleanのSeriesにする

```
In [12]: criteria_final = criteria1 & criteria2 & criteria3
```

```
criteria_final.head()
```

Out[12]:

movie_title		
Avatar	False	
Pirates of the Caribbean: At World's End	False	False
Spectre	False	
The Dark Knight Rises	True	
Star Wars: Episode VII - The Force Awakens	False	

dtype: bool

(補足) 割愛