Ques 07: How do I `filter` rows of a pandas dataframe by column value ?

In [1]:
```
# IMDB - Internet Movie Database (Dataset)
import pandas as pd
movie_ratings_ds = pd.read_csv('http://bit.ly/imdbratings')
movie_ratings_ds.head(4)
```

Out[1]:

	star_rating	title	content_rating	genre	duration	actors_list
0	9.3	The Shawshank Redemption	R	Crime	142	[u'Tim Robbins', u'Morgan Freeman', u'Bob Gunt...
1	9.2	The Godfather	R	Crime	175	[u'Marlon Brando', u'Al Pacino', u'James Caan']
2	9.1	The Godfather: Part II	R	Crime	200	[u'Al Pacino', u'Robert De Niro', u'Robert Duv...
3	9.0	The Dark Knight	PG-13	Action	152	[u'Christian Bale', u'Heath Ledger', u'Aaron E...

In [2]:
```
movie_ratings_ds.shape
```

Out[2]:
```
(979, 6)
```

In [3]:
```
movie_ratings_ds.head()
```

Out[3]:

	star_rating	title	content_rating	genre	duration	actors_list
0	9.3	The Shawshank Redemption	R	Crime	142	[u'Tim Robbins', u'Morgan Freeman', u'Bob Gunt...
1	9.2	The Godfather	R	Crime	175	[u'Marlon Brando', u'Al Pacino', u'James Caan']
2	9.1	The Godfather: Part II	R	Crime	200	[u'Al Pacino', u'Robert De Niro', u'Robert Duv...
3	9.0	The Dark Knight	PG-13	Action	152	[u'Christian Bale', u'Heath Ledger', u'Aaron E...
4	8.9	Pulp Fiction	R	Crime	154	[u'John Travolta', u'Uma Thurman', u'Samuel L....

We want to find the records having duration >= 200.

Long Process

In [4]:
```
# Creating a python list containing boolean value for each row
booleans = []
for length in movie_ratings_ds.duration:
    if length >= 200:
        booleans.append(True)
    else:
        booleans.append(False)

booleans[0:6]   # check first 5 values
```

Out[4]:
```
[False, False, True, False, False, False]
```

In [5]:
```
len(booleans)
```

Out[5]:
```
979
```

In [6]:
```
# convert the boolean list into pandas series

is_long = pd.Series(booleans)
is_long.head()
```

Out[6]:
```
0    False
1    False
2     True
3    False
4    False
dtype: bool
```

In [7]:
```
# Pass the Series to dataframe. It will show only the rows having True value in the Series
movie_ratings_ds[is_long]
```

Out[7]:

	star_rating	title	content_rating	genre	duration	actors_list
2	9.1	The Godfather: Part II	R	Crime	200	[u'Al Pacino', u'Robert De Niro', u'Robert Duv...
7	8.9	The Lord of the Rings: The Return of the King	PG-13	Adventure	201	[u'Elijah Wood', u'Viggo Mortensen', u'Ian McK...
17	8.7	Seven Samurai	UNRATED	Drama	207	[u'Toshir\xf4 Mifune', u'Takashi Shimura', u'K...
78	8.4	Once Upon a Time in America	R	Crime	229	[u'Robert De Niro', u'James Woods', u'Elizabet...
85	8.4	Lawrence of Arabia	PG	Adventure	216	[u"Peter O'Toole", u'Alec Guinness', u'Anthony...
142	8.3	Lagaan: Once Upon a Time in India	PG	Adventure	224	[u'Aamir Khan', u'Gracy Singh', u'Rachel Shell...
157	8.2	Gone with the Wind	G	Drama	238	[u'Clark Gable', u'Vivien Leigh', u'Thomas Mit...
204	8.1	Ben-Hur	G	Adventure	212	[u'Charlton Heston', u'Jack Hawkins', u'Stephe...
445	7.9	The Ten Commandments	APPROVED	Adventure	220	[u'Charlton Heston', u'Yul Brynner', u'Anne Ba...
476	7.8	Hamlet	PG-13	Drama	242	[u'Kenneth Branagh', u'Julie Christie', u'Dere...
630	7.7	Malcolm X	PG-13	Biography	202	[u'Denzel Washington', u'Angela Bassett', u'De...
767	7.6	It's a Mad, Mad, Mad, Mad World	APPROVED	Action	205	[u'Spencer Tracy', u'Milton Berle', u'Ethel Me...

Short Process

In [8]:
```
is_long = movie_ratings_ds.duration >= 200
movie_ratings_ds[is_long].head()
```

Out[8]:

	star_rating	title	content_rating	genre	duration	actors_list
2	9.1	The Godfather: Part II	R	Crime	200	[u'Al Pacino', u'Robert De Niro', u'Robert Duv...
7	8.9	The Lord of the Rings: The Return of the King	PG-13	Adventure	201	[u'Elijah Wood', u'Viggo Mortensen', u'Ian McK...
17	8.7	Seven Samurai	UNRATED	Drama	207	[u'Toshir\xf4 Mifune', u'Takashi Shimura', u'K...
78	8.4	Once Upon a Time in America	R	Crime	229	[u'Robert De Niro', u'James Woods', u'Elizabet...
85	8.4	Lawrence of Arabia	PG	Adventure	216	[u"Peter O'Toole", u'Alec Guinness', u'Anthony...

In [9]:
```
# We can wrap it one line
movie_ratings_ds[ movie_ratings_ds.duration >= 200 ].head()
```

Out[9]:

	star_rating	title	content_rating	genre	duration	actors_list
2	9.1	The Godfather: Part II	R	Crime	200	[u'Al Pacino', u'Robert De Niro', u'Robert Duv...
7	8.9	The Lord of the Rings: The Return of the King	PG-13	Adventure	201	[u'Elijah Wood', u'Viggo Mortensen', u'Ian McK...
17	8.7	Seven Samurai	UNRATED	Drama	207	[u'Toshir\xf4 Mifune', u'Takashi Shimura', u'K...
78	8.4	Once Upon a Time in America	R	Crime	229	[u'Robert De Niro', u'James Woods', u'Elizabet...
85	8.4	Lawrence of Arabia	PG	Adventure	216	[u"Peter O'Toole", u'Alec Guinness', u'Anthony...

In [10]:
```
# Showing only one column based on filter

movie_ratings_ds[ movie_ratings_ds.duration >= 200 ].title.head()

# It has some limitation. So, there is a better approach for this same task
```

Out[10]:
```
2                          The Godfather: Part II
7     The Lord of the Rings: The Return of the King
17                                  Seven Samurai
78                      Once Upon a Time in America
85                                Lawrence of Arabia
Name: title, dtype: object
```

In [11]:
```
# use of "loc" method
movie_ratings_ds.loc[ movie_ratings_ds.duration >= 200, 'genre']
```

Out[11]:
```
2          Crime
7      Adventure
17         Drama
78         Crime
85     Adventure
142    Adventure
157        Drama
204    Adventure
445    Adventure
476        Drama
630    Biography
767       Action
Name: genre, dtype: object
```

In [ ]: