SettingWithCopyWarning

November 20, 2016

```
In [1]: import pandas as pd

In [2]: movies = pd.read_csv('http://bit.ly/imdbratings')

In [3]: movies.head()

Out[3]:    star_rating                    title content_rating    genre   duration
        0          9.3  The Shawshank Redemption              R    Crime        142
        1          9.2             The Godfather              R    Crime        175
        2          9.1     The Godfather: Part II             R    Crime        200
        3          9.0            The Dark Knight          PG-13   Action        152
        4          8.9               Pulp Fiction             R    Crime        154

                                              actors_list
        0  [u'Tim Robbins', u'Morgan Freeman', u'Bob Gunt...
        1    [u'Marlon Brando', u'Al Pacino', u'James Caan']
        2  [u'Al Pacino', u'Robert De Niro', u'Robert Duv...
        3  [u'Christian Bale', u'Heath Ledger', u'Aaron E...
        4  [u'John Travolta', u'Uma Thurman', u'Samuel L...

In [4]: movies.content_rating.isnull().sum() # count missing values in content_rati

Out[4]: 3

In [5]: movies[movies.content_rating.isnull()] # boolean passed through to give lis

Out[5]:      star_rating                          title content_rating  \
        187          8.2  Butch Cassidy and the Sundance Kid            NaN
        649          7.7                     Where Eagles Dare            NaN
        936          7.4                             True Grit            NaN

                  genre  duration                                  actors_list
        187  Biography        110  [u'Paul Newman', u'Robert Redford', u'Katharin...
        649     Action        158  [u'Richard Burton', u'Clint Eastwood', u'Mary ...
        936  Adventure        128    [u'John Wayne', u'Kim Darby', u'Glen Campbell']

In [6]: movies.content_rating.value_counts()
```

```
Out[6]: R            460
        PG-13        189
        PG           123
        NOT RATED     65
        APPROVED      47
        UNRATED       38
        G             32
        PASSED         7
        NC-17          7
        X              4
        GP             3
        TV-MA          1
        Name: content_rating, dtype: int64

In [9]: movies[movies.content_rating=='NOT RATED'].head() # find all movies with NO

Out[9]:    star_rating                          title content_rating    genre \
        5          8.9                  12 Angry Men      NOT RATED    Drama
        6          8.9  The Good, the Bad and the Ugly   NOT RATED    Western
        41         8.5                   Sunset Blvd.    NOT RATED    Drama
        63         8.4                             M    NOT RATED    Crime
        66         8.4            Munna Bhai M.B.B.S.    NOT RATED    Comedy

           duration                                    actors_list
        5        96  [u'Henry Fonda', u'Lee J. Cobb', u'Martin Bals...
        6       161  [u'Clint Eastwood', u'Eli Wallach', u'Lee Van ...
        41      110  [u'William Holden', u'Gloria Swanson', u'Erich...
        63       99  [u'Peter Lorre', u'Ellen Widmann', u'Inge Land...
        66      156   [u'Sunil Dutt', u'Sanjay Dutt', u'Arshad Warsi']

In [10]: movies[movies.content_rating=='NOT RATED'].content_rating.head()

Out[10]: 5     NOT RATED
         6     NOT RATED
         41    NOT RATED
         63    NOT RATED
         66    NOT RATED
         Name: content_rating, dtype: object

In [11]: import numpy as np

In [14]: movies[movies.content_rating=='NOT RATED'].content_rating = np.nan
         # movies[movies.content_rating=='NOT RATED'] => getter
         # .content_rating = np.nan => setter

/opt/conda/lib/python3.5/site-packages/pandas/core/generic.py:2701: SettingWithCopy
A value is trying to be set on a copy of a slice from a DataFrame.
Try using .loc[row_indexer,col_indexer] = value instead
```

```
See the caveats in the documentation: http://pandas.pydata.org/pandas-docs/stable/i
  self[name] = value


In [23]: movies.content_rating.isnull().sum() # didn't work => if it did would be 6
         # now 68 when re-run command

Out[23]: 68

In [22]: movies.loc[movies.content_rating=='NOT RATED', 'content_rating'] = np.nan

In [20]: movies.content_rating.isnull().sum()

Out[20]: 3

In [24]: top_movies = movies.loc[movies.star_rating >=9, :] # movies with star rati

In [25]: top_movies # returned list

Out[25]:    star_rating                         title content_rating   genre   duration
         0          9.3  The Shawshank Redemption              R   Crime        142
         1          9.2              The Godfather              R   Crime        175
         2          9.1     The Godfather: Part II              R   Crime        200
         3          9.0            The Dark Knight          PG-13  Action        152


                                 actors_list
         0  [u'Tim Robbins', u'Morgan Freeman', u'Bob Gunt...
         1    [u'Marlon Brando', u'Al Pacino', u'James Caan']
         2  [u'Al Pacino', u'Robert De Niro', u'Robert Duv...
         3  [u'Christian Bale', u'Heath Ledger', u'Aaron E...

In [26]: top_movies.loc[0, 'duration'] = 150 # set Shawshank Redemption to 150

/opt/conda/lib/python3.5/site-packages/pandas/core/indexing.py:465: SettingWithCopy
A value is trying to be set on a copy of a slice from a DataFrame.
Try using .loc[row_indexer,col_indexer] = value instead

See the caveats in the documentation: http://pandas.pydata.org/pandas-docs/stable/i
  self.obj[item] = s


In [27]: top_movies # hmm...it did change

Out[27]:    star_rating                         title content_rating   genre   duration
         0          9.3  The Shawshank Redemption              R   Crime        150
         1          9.2              The Godfather              R   Crime        175
         2          9.1     The Godfather: Part II              R   Crime        200
         3          9.0            The Dark Knight          PG-13  Action        152


                                 actors_list
```

```
0  [u'Tim Robbins', u'Morgan Freeman', u'Bob Gunt...
1    [u'Marlon Brando', u'Al Pacino', u'James Caan']
2  [u'Al Pacino', u'Robert De Niro', u'Robert Duv...
3  [u'Christian Bale', u'Heath Ledger', u'Aaron E...
```

In [28]: *# any time you are making a copy, you should use .copy method explicitly*
```
top_movies = movies.loc[movies.star_rating >=9, :].copy()
```

In [31]: `top_movies.loc[0, 'duration'] = 150`

In [32]: `top_movies` *# changed the copy*

Out[32]:
```
   star_rating                  title content_rating   genre  duration
0          9.3  The Shawshank Redemption             R   Crime       150
1          9.2             The Godfather             R   Crime       175
2          9.1     The Godfather: Part II            R   Crime       200
3          9.0             The Dark Knight        PG-13  Action       152

                                  actors_list  duation
0  [u'Tim Robbins', u'Morgan Freeman', u'Bob Gunt...    150.0
1    [u'Marlon Brando', u'Al Pacino', u'James Caan']      NaN
2  [u'Al Pacino', u'Robert De Niro', u'Robert Duv...      NaN
3  [u'Christian Bale', u'Heath Ledger', u'Aaron E...      NaN
```

In [33]: `movies.head()` *# not the original*

Out[33]:
```
   star_rating                  title content_rating   genre  duration
0          9.3  The Shawshank Redemption             R   Crime       142
1          9.2             The Godfather             R   Crime       175
2          9.1     The Godfather: Part II            R   Crime       200
3          9.0             The Dark Knight        PG-13  Action       152
4          8.9             Pulp Fiction             R   Crime       154

                                  actors_list
0  [u'Tim Robbins', u'Morgan Freeman', u'Bob Gunt...
1    [u'Marlon Brando', u'Al Pacino', u'James Caan']
2  [u'Al Pacino', u'Robert De Niro', u'Robert Duv...
3  [u'Christian Bale', u'Heath Ledger', u'Aaron E...
4  [u'John Travolta', u'Uma Thurman', u'Samuel L...
```

In [ ]: