PandaCommands_Movie_database

November 16, 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]: # as long as >1 shows descriptive statistics of each column -> star rating
        movies.describe()
```

```
Out[4]:        star_rating    duration
        count   979.000000  979.000000
        mean      7.889785  120.979571
        std       0.336069   26.218010
        min       7.400000   64.000000
        25%       7.600000  102.000000
        50%       7.800000  117.000000
        75%       8.100000  134.000000
        max       9.300000  242.000000
```

```
In [5]: movies.shape #shows tuple of rows and columns
```

```
Out[5]: (979, 6)
```

```
In [6]: movies.dtypes #tells us data types of six columns
```

```
Out[6]: star_rating       float64
        title              object
        content_rating     object
        genre              object
        duration            int64
        actors_list        object
        dtype: object
```

```
In [8]: type(movies) # -> is a data frame
```

```
Out[8]: pandas.core.frame.DataFrame
```

```
In [9]: movies.describe(include=['object'])
```

```
Out[9]:                  title content_rating  genre  \
        count              979            976    979
        unique             975             12     16
        top     Les Miserables              R  Drama
        freq                 2            460    278

                                                     actors_list
        count                                                979
        unique                                               969
        top     [u'Daniel Radcliffe', u'Emma Watson', u'Rupert...
        freq                                                   6
```

```
In [ ]: # shift + tab for description of parentheses-object -> several times for sp
```