Assignment #4

**Step 1**

Download the IMDB dataset from <https://www.kaggle.com/PromptCloudHQ/imdb-data> (or from FormazioneOnLine).

The dataset contains 5,000 popular movies on IMDB from 2006 to 2016. The features included are: Title, Genre, Description, Director, Actors, Year, Runtime, Rating, Votes, Revenue, Metascore.

**Step 2**

Perform a preprocessing analysis following the guidelines reported in this blog:

http://www.developintelligence.com/blog/2017/08/data-cleaning-pandas-python/ (also available on FormazioneOnLine).

In other words, reinsert the commands described in the blog and learn how to:

* Add a default value for the missing data;
* Get rid of (delete) the rows that have missing data;
* Get rid of (delete) the columns that have a high incidence of missing data;
* …