# BDA Analytics Challenge 2020 – Data Set

## ratings.csv

Ratings given by users to movies.

Columns:

- user_id (int)
- movieID (int)
- rating (float): Ratings in steps of 0.5 (worst: 0.5, best: 5.0)

## movies.csv

Metadata about the movies.

Columns:

- id (int)
- title (str)
- imdbID (int): IMDB ID of the movie
  - IMDB URL: https://www.imdb.com/title/tt<imdbID> with imdbID zero-padded to 7 digits
  - E.g. ID 25464 → URL https://www.imdb.com/title/tt0025464
- spanishTitle (str)
- imdbPictureURL (str)
- year (int)
- rtID (str): Rotten Tomatoes (RT) ID of the movie
  - RT URL: https://www.rottentomatoes.com/m/<rtID>
- rtPictureURL (str)

## actors.csv

Mapping of movies to the actors in the movies.

Columns:

- movieID (int)
- actorID (str)
- actorName (str)
- ranking (int): order in which actors appear on the movie IMDb cast web page

## countries.csv

Mapping of movies to their countries of origin.

Columns:

- movieID (int)
- country (str)

## directors.csv

Mapping of movies to their directors.

Columns:

- movieID (str)
- directorID (str)
- directorName (str)

## genres.csv

Mapping of movies to one or more genres.

Columns:

- movieID (int)
- genre (str)

## locations.csv

Mapping of movies to their filming locations.

Columns:

- movieID (int)
- location1 – location4 (str): up to four filming locations per movie

## movie_tags.csv

Mapping of movies to tags which were assigned by users.

Columns:

- movieID (str)
- tagID (str)
- tagWeight (int): Weight / strength of tag assignment

## tags.csv

User-generated tags.

Columns:

- id (int)
- value (str)