

## The Movies Dataset – Column Information

### Labels (Target Variable of Prediction):

- revenue: the movie's box office revenue in Million USD (MUSD)

### Feature Variables:

- title: movie title (for information purposes only, not to be used in regression)
- main\_genre: the main genre of the movie (e.g. Drama)
- belongs\_to\_collection: Whether the movie belongs to a collection/franchise (True) or not (False). Example: Star Wars
- budget: the movie's budget in Million USD (MUSD)
- production\_companies: number of involved production companies
- vote\_average: Average vote by preview audience. From 1 (very bad) to 10 (excellent)
- popularity: the movie's popularity measured in google searches, social media posts, etc.
- runtime: runtime of the movie in minutes
- cast\_size: number of actors involved
- crew\_size: number of crew members involved
- cast\_tr: Track Record of the movie's three main actors (aggregated box-office revenues in MUSD prior to this movie)
- director\_tr: Track Record of the movie's director (aggregated box-office revenues in MUSD prior to this movie)
- production\_company\_tr: Track Record of the movie's production company (aggregated box-office revenues in MUSD prior to this movie)
- English: Whether the movie's original language is English (True) or not (False)
- title\_length: number of characters in the movie title
- rel\_year: the movie's year of release
- rel\_month: the movie's month of release