Movie Ratings and Box Office

[ECON 5753] — *University of Arkansas*

The movie dataset combines two datasets:

1. Scraped Letterbox Review

2. Information on the movie from The Movie Database

Table 1: Codebook

variable	description
movie_title	Title of movie
tmdb_id	The Movie Database Unique (TMDB) ID
imdb_id	The Internet Movie Database Unique ID
tmdb_link	The Movie Database Film Link
runtime	The film's runtime
year_released	The film's release year
release_date	The film's release date
letterboxd_vote_average	The average rating on letterboxd (1-10)
letterboxd_vote_count	The number of reviews on letterboxd (at time of scraping)
popularity	The film's popularity on tmdb
original_language	The film's original language
genre_1	The film's first listed genre
genre_2	The film's second listed genre
genre_3	The film's third listed genre
genre_4	The film's fourth listed genre
genre_5	The film's fifth listed genre

`production_country_1`	The film's first listed country of production
`production_country_2`	The film's second listed country of production
`production_country_3`	The film's third listed country of production
`production_country_4`	The film's fourth listed country of production
`production_country_5`	The film's fifth listed country of production
`spoken_language_1`	The film's first listed spoken language
`spoken_language_2`	The film's second listed spoken language
`tmdb_title`	The film's title on TMDB
`tmdb_runtime`	The film's runtime on TMDB
`tmdb_status`	The film's release status on TMDB
`tmdb_is_adult_movie`	If the film an adult movie according to TMDB
`tmdb_popularity`	The film's popularity on TMDB
`tmdb_release_date`	The film' release date on TMDB
`tmdb_budget`	The film's listed budget on TMDB
`tmdb_revenue`	The film's listed revenue on TMDB
`tmdb_vote_count`	The film's rating vote count on TMDB
`tmdb_vote_average`	The film's average rating on TMDB
`tmdb_production_company_name`	The film's first listed production company on TMDB
`tmdb_production_company_country`	The operating country of the film's first listed production company on TMDB

Table 1: Codebook (Continued)

Preview of data:

```
Rows: 25,813
Columns: 35
$ movie_title              <chr> "Desperately Seeking Santa", "A Deadly~
$ tmdb_id                  <dbl> 81330, 343921, 420644, 106717, 37821, ~
$ imdb_id                  <chr> "tt1967688", "tt4573800", "tt5827590",~
$ tmdb_link                <chr> "https://www.themoviedb.org/movie/8133~
$ year_released            <dbl> 2011, 2015, 2018, 2012, 2010, 2015, 20~
$ runtime                  <dbl> 86, 90, 101, 82, 100, 73, 112, 101, 97~
$ release_date             <date> 2011-11-27, 2015-06-20, 2018-08-17, 2~
$ letterboxd_vote_count    <dbl> 43, 68, 34, 85, 1819, 37, 2166, 15, 24~
$ letterboxd_vote_average  <dbl> 6.1, 5.0, 7.7, 5.3, 5.9, 6.0, 7.5, 5.6~
$ popularity               <dbl> 4.151, 5.060, 3.629, 6.645, 22.918, 2.~
$ original_language        <chr> "en", "en", "zh", "en", "en", "fr", "i~
$ genre_1                  <chr> "Romance", "Drama", "Drama", "Horror",~
$ genre_2                  <chr> "TV Movie", "Thriller", "Action", "Thr~
$ genre_3                  <chr> NA, NA, "Adventure", NA, "Thriller", N~
$ genre_4                  <chr> NA, NA, "Thriller", NA, "Romance", NA,~
$ genre_5                  <chr> NA, NA, NA, NA, NA, NA, NA, NA, NA, NA~
$ production_country_1     <chr> "Canada", "United States of America", ~
$ production_country_2     <chr> NA, NA, "Hong Kong", NA, NA, "France",~
$ production_country_3     <chr> NA, NA, NA, NA, NA, NA, NA, NA, NA, NA~
$ production_country_4     <chr> NA, NA, NA, NA, NA, NA, NA, NA, NA, NA~
$ production_country_5     <chr> NA, NA, NA, NA, NA, NA, NA, NA, NA, NA~
$ spoken_language_1        <chr> "English", "English", "普通话", "English"~
$ spoken_language_2        <chr> NA, NA, NA, NA, NA, NA, NA, NA, NA, NA~
$ tmdb_title               <chr> "Desperately Seeking Santa", "A Deadly~
$ tmdb_runtime             <dbl> 87, 84, 101, 82, 100, 73, 112, 101, 97~
$ tmdb_status              <chr> "Released", "Released", "Released", "R~
$ tmdb_is_adult_movie      <lgl> FALSE, FALSE, FALSE, FALSE, FALSE, FAL~
```

```
$ tmdb_popularity                  <dbl> 1.700, 2.864, 1.774, 3.785, 3.678, 1.8~
$ tmdb_release_date                <date> 2012-11-07, 2015-06-20, 2018-08-17, 2~
$ tmdb_budget                      <dbl> 0, 0, 0, 0, 75000000, 0, 1900000, 0, 0~
$ tmdb_revenue                     <dbl> 0, 0, 0, 0, 98159963, 54985, 0, 0, 0, ~
$ tmdb_vote_count                  <dbl> 48, 81, 50, 111, 2206, 51, 2501, 23, 2~
$ tmdb_vote_average                <dbl> 5.900, 4.900, 7.200, 5.050, 5.962, 5.6~
$ tmdb_production_company_name     <chr> "Muse Entertainment", "Gary Sanchez Pr~
$ tmdb_production_company_country  <chr> "CA", "US", "HK", NA, NA, "FR", "IT", ~
```