

Codebook for Movies dataset

Data

The data set is comprised of 651 randomly sampled movies produced and released before 2016.

Some of these variables are only there for informational purposes and do not make any sense to include in a statistical analysis. It is up to you to decide which variables are meaningful and which should be omitted. For example information in the the actor1 through actor5 variables was used to determine whether the movie casts an actor or actress who won a best actor or actress Oscar.

You might also choose to omit certain observations or restructure some of the variables to make them suitable for answering your research questions.

When you are fitting a model you should also be careful about collinearity, as some of these variables may be dependent on each other.

Codebook

1. title : Title of movie
2. title_type : Type of movie (Documentary, Feature Film, TV Movie)
3. genre : Genre of movie (Action & Adventure, Comedy, Documentary, Drama, Horror, Mystery & Suspense, Other)
4. runtime : Runtime of movie (in minutes)
5. mpaa_rating : MPAA rating of the movie (G, PG, PG-13, R, Unrated)
6. studio : Studio that produced the movie
7. thtr_rel_year : Year the movie is released in theaters
8. thtr_rel_month : Month the movie is released in theaters
9. thtr_rel_day : Day of the month the movie is released in theaters
10. dvd_rel_year : Year the movie is released on DVD
11. dvd_rel_month : Month the movie is released on DVD
12. dvd_rel_day : Day of the month the movie is released on DVD
13. imdb_rating : Rating on IMDB
14. imdb_num_votes : Number of votes on IMDB
15. critics_rating : Categorical variable for critics rating on Rotten Tomatoes (Certified Fresh, Fresh, Rotten)
16. critics_score : Critics score on Rotten Tomatoes
17. audience_rating : Categorical variable for audience rating on Rotten Tomatoes (Spilled, Upright)
18. audience_score : Audience score on Rotten Tomatoes
19. best_pic_nom : Whether or not the movie was nominated for a best picture Oscar (no, yes)
20. best_pic_win : Whether or not the movie won a best picture Oscar (no, yes)
21. best_actor_win : Whether or not one of the main actors in the movie ever won an Oscar (no, yes) – note that this is not necessarily whether the actor won an Oscar for their role in the given movie
22. best_actress_win : Whether or not one of the main actresses in the movie ever won an Oscar (no, yes) – not that this is not necessarily whether the actresses won an Oscar for their role in the given movie
23. best_dir_win : Whether or not the director of the movie ever won an Oscar (no, yes) – not that this is not necessarily whether the director won an Oscar for the given movie
24. top200_box : Whether or not the movie is in the Top 200 Box Office list on BoxOfficeMojo (no, yes)
25. director : Director of the movie
26. actor1 : First main actor/actress in the abridged cast of the movie

27. actor2 : Second main actor/actress in the abridged cast of the movie
28. actor3 : Third main actor/actress in the abridged cast of the movie
29. actor4 : Fourth main actor/actress in the abridged cast of the movie
30. actor5 : Fifth main actor/actress in the abridged cast of the movie
31. imdb_ur1 : Link to IMDB page for the movie
32. rt_ur1 : Link to Rotten Tomatoes page for the movie