Movie Analysis

Chin-hung Yeh, 48341011 Michael Zhao, 48311436 Jess Pei, 48422157
Sunny Zhang, 48377801 Ryan Li, 47038567

March 01, 2021

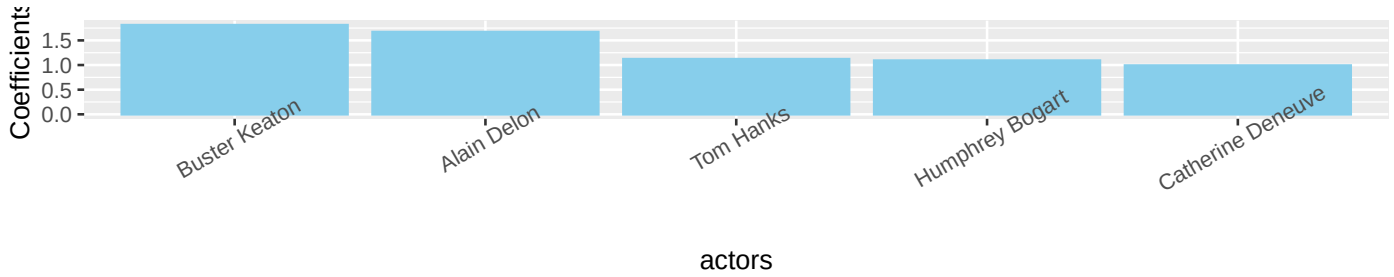
I. Introduction

Movies have always been part of people's life. When we decide a movie that is good or bad, there is always a certain rationale behind it. Sometimes, *major actors* can *boost the ratings* tremendously because of their skills and popularity in the community.However, sometimes, *some genres* yield a *better attraction* than some others. In this assignment, we are going to ***explore multiple different factors that influence the ratings***, and ***draw conclusions on how each variable affects the rating of the movies***.

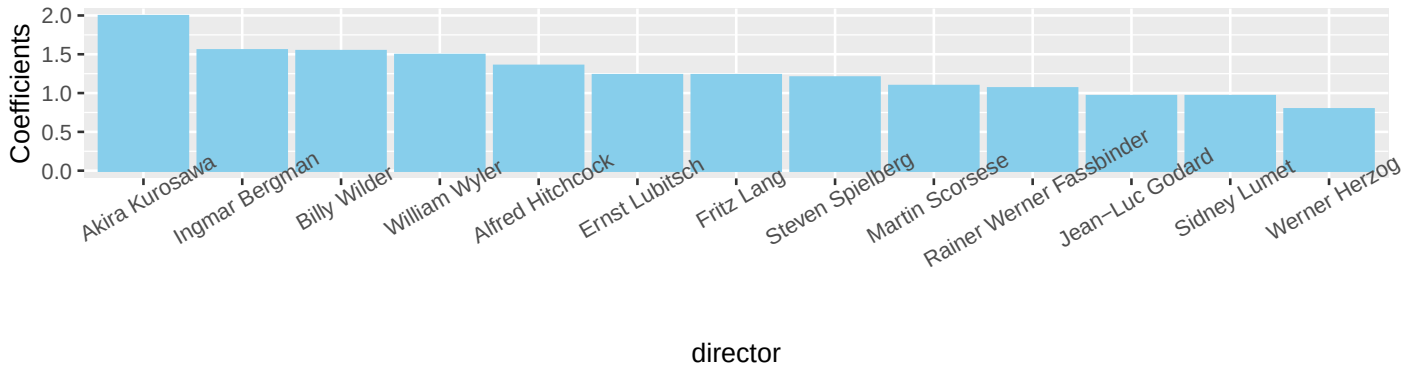
II. Model Insights and Recommendations

A. Actors & Directors

Taking a quick glance of the list of casts before making the movie decisions has become a force of habit for almost all audiences. Because we know by heart that certain actors are the indicators of good movies by gracing the movies with their presences. Hopefully, our model will help you stay away from mediocre or bad movies by giving you recommendations of actors to look for based on the patterns our model has found. Here are the six actors from multiple countries, different age range that will more likely guarantee you a good movie:

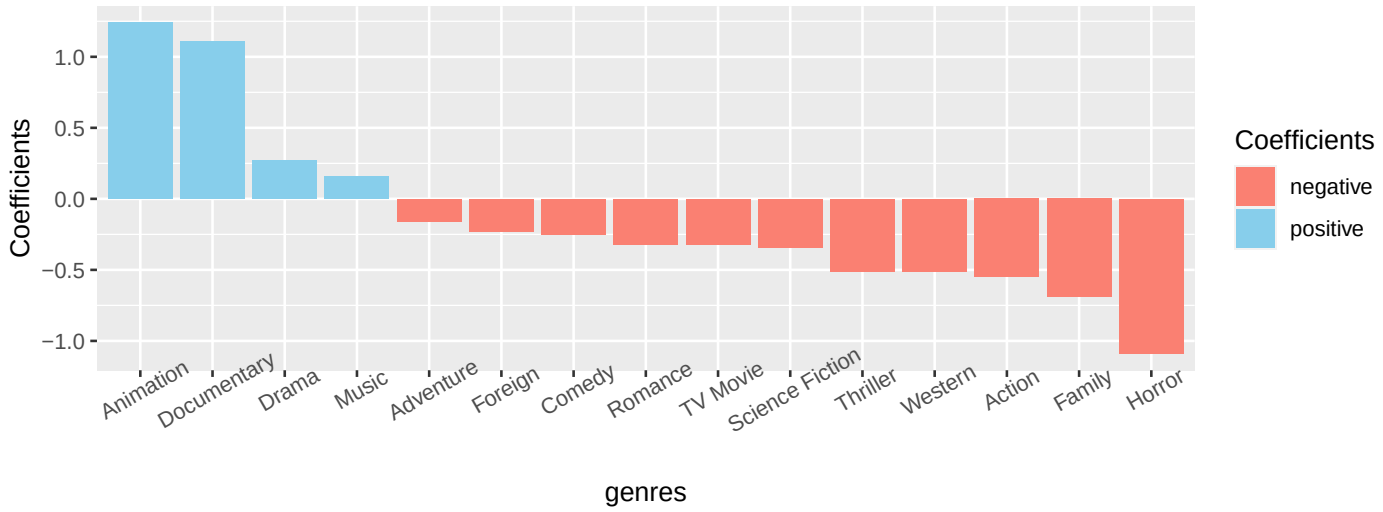


Same goes to the movie directors. We found that the following 12 directors have more possibilities of make good movies than others:



B. Genres

Besides actors and directors, a generic label also plays an important role in movie selection. It provides audiences a rough indication of the movie's content and narrative form. Not surprisingly the factors that most determine the relation between genres and viewer preferences are based on biology, namely age and gender. Based on our model, we discovered the following genres can greatly impact audience's in selecting movies.



Higher Coefficients means higher probability of getting good movie ratings. According to the genres graph, Animation and Documentary are in the top list among all these genres. On the other side, Horror and family type of movies are still struggling with getting higher ratings.

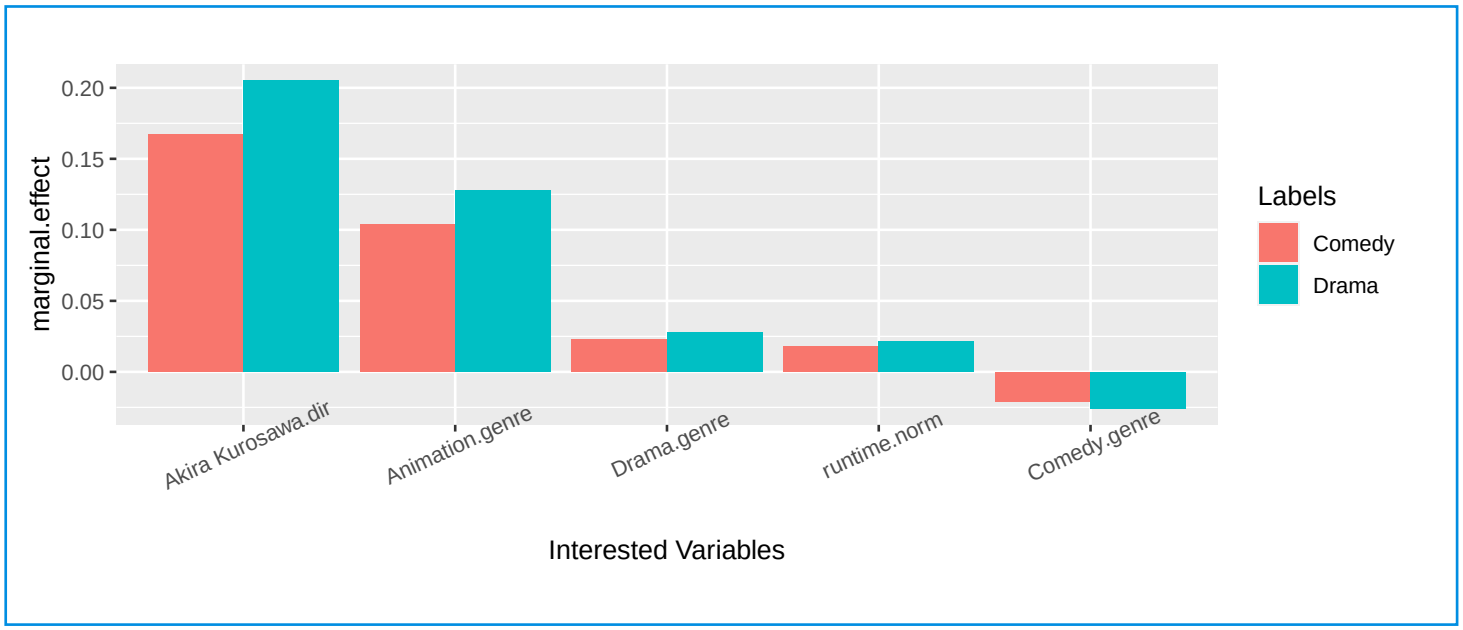
An interesting question raised from this genres analyis is ***Does the audience usually pay close attention to which production company is when selecting movies? Will you?*** From our model, the answer is no. Movie production companies are not as heavy as other factors do in affecting movie selection and movie rating. Quite accidentally, we discover that the movies from production company Columbia Pictures tend to get low movie ratings(reference by 0.904(90.4%) of their movies get low movie rates in table below).

```
   Cell Contents
|-------------------------|
|                       N |
|           N / Row Total |
|-------------------------|


Total Observations in Table:  40675


                          | data6$vote_average.norm
data6$"Columbia Pictures" |         0 |         1 | Row Total |
--------------------------|-----------|-----------|-----------|
                        0 |     33830 |      6427 |     40257 |
                          |     0.840 |     0.160 |     0.990 |
--------------------------|-----------|-----------|-----------|
                        1 |       378 |        40 |       418 |
                          |     0.904 |     0.096 |     0.010 |
--------------------------|-----------|-----------|-----------|
             Column Total |     34208 |      6467 |     40675 |
--------------------------|-----------|-----------|-----------|
```

C. Recommendations and its Marginal Effect

Taking a good movie to spend your downtime is always interesting, and our analysis and modeling technique helps you make the decision. From the actor level, there are 6 actors that would give you a significant boost in choosing a good movie. Buster Keaton and Alain Delon are the top two choices. In terms of the directors, 12 of them would likely yield a great result, and Akira Kurosawa and Ingmar Bergman are the two top directors that people should keep an eye on. Genre has been a big factor in this decision as animation, documentary, and drama are more popular ***(The reason some less common genres such as documentary turn out to contribute substantially to probability of being rendered good movies will be discussed later in Model Interpretability part)***. When you see the Columbia pictures, the audience should think twice about whether they should invest their time in it or not.

To further analyze marginal affect when people switch from one genre (or multiple genres) to another, we assume that Buster Keaton - main character - and Akira Kurosawa - director - are both invited because of their popularity, and we try to see the marginal effect on comedy and drama because these two genres appear to have most records in our original data sets. In the first tab, we can see that without Akira Kurosawa, there is a larger marginal effect on drama than comedy. If we increase an animation genre into either comedy or drama, there is a bigger marginal effect on drama. If there is a change in runtime, we can see a bigger effect on drama than comedy. If there is another increase in drama, we can determine that drama still yields a larger effect. If there is another change in comedy, we can see a negative effect on both, with a larger marginal effect on drama. To conclude, suppose you are originally watching films from director Akira Kurosawa and it's of main character of Buster Keaton and of both Drama and Comedy genres, if you choose to seek for another movies from the same director and main actor, you would probably want to get rid of the comedy genre and look for pure drama genre to expect a better rating movies. In comparison, if you choose to drop either comedy or drama to find a movie of single genre, marginal effect of dropping drama on all others inetrested variables are expected to be larger. That is to say, if you choose to drop drama and look for movies from the same setting but just comedy genre, you should expect a higher probability of choosing either a good movie or bad movie, as dropping drama genre will influence probability of being rendored a good movie more substatially than dropping comedy genre.



III. Model Interpretability

A. Model Deficiencies

Two things that would make this model less reliable would be ***Collinearity*** and ***self selection***. The reason behind this is that those dependent variables in our regression model are potentially highly related. Therefore, it becomes difficult to distinguish the individual effects of this model. For example, In our model we had Columbia Pictures, Comedy and Animation, those three variables could face the multicollinearity problem since the Columbia pictures mainly distribute the Disney's cartoons movies. Those three factors may cause some relationships within each other that would be ignored

in this model. In addition, we noticed as well some minor genres in our model were highly rated by the minority, indicating a pattern of self selection data in terms of those minor genres. For example, people who would go to a movie theater for Documentary kind of genres and rate it are themselves interested in such specific genre and thus are highly possible to rate those movies positively. On the contrary, people who are not even interested in such kind of films wouldn't even spend time watching movies in such genres, and thus there is not negatively rated record for some specific kind of genres. Under such circumstance, some minor genres become a substantial contributor to log-odds ratio in our model, ratio we later convert into probability whether the movie is good or not.

B. Model Confounds

Besides two major model deficiencies described above, there are some additional features - features that are not in our model - that could help explain whether a movie is good or not. Some specific features are ignored in our model due to a lack of precise information in its record. For instance, movie budget, revenue, popularity, etc. are dropped because of their nullity - 30000+ records with zero-value for these features. Additionally, some features such as comfortability in specific movie theater, holiday seasons, competitive assets, economy, etc. are just not listed in the original data sets. For example, whether feeling comfortable or not during movies might affect degree on how much people enjoy movies. In other words, to general public, the rating does not lie purely in whether movies themselves are good but also in whether they enjoy them - both plots and watching environment.

C. Conclusion

To conclude, our model does provide ways to assess whether movies are good or not. Nevertheless, when our model is applied on other movies, predicted results should be evaluated more carefully on its potential problems. As the above confounds and deficiencies are two major drawbacks, we suggest to have these two drawbacks as a starting point to think over potential issues - if any - on results attained from our model.