Overview:

There are many standards to judge whether one movie was good or not.Based on the dataset provided, we choose aggregate rating and revenue as our standards.Thus, we developed two models to investigate which factors will influence rating and revenue of a movie separately. Additionally, our models can be used to predict the probability that a movie can be rated as "good" and its revenue.The comprehensive analysis are stated as follows.

Dependent Variable:

Rating model:Since we have rating data from two websites, we combine the rating score together to get more accurate results.We firstly rescaled the rating data from TMDB to match another website and then calculated the weighted average of rating score for each movie based on the number of reviews.Finally, we set 3.48 as the threshold to divide all the movies into "good" and "bad" movie, which is the third quartile for all the aggregate rating scores.

Independent Variable Pre-procession:

1. **Production Company**
   Since one movie may have many production companies, we only chose main production company as our independent variable.We also divided these production companies into 4 tiers based on the average rating scores of the movies they produced. Each tier has the same number of companies. The following are ranges of rating score for all 4 tiers:
   Tier 1:3.47 - 5, Tier 2: 3.18 - 3.47, Tier 3:2.86 - 3.18, Tier 4: 0 - 2.86

2. **Production Country**
   Since one movie may have many production companies, we only chose main production company as our independent variable. Production countries were grouped the same way as production company. The following are ranges of rating score for all 4 tiers:
   Tier 1: 3.51 - 5, Tier 2: 3.29 - 3.51, Tier 3: 3.11 - 3.29,Tier 4: 0 - 3.11

3. **Genre of Movie**
   One movie can be classified as various genres.We set genres as dummy variables so we could analyze the impact of each genre. "1" represents one movie was classified as a specific genre. "0" represents one movie was not classified as a specific genre.

4. **Budget**
   There are many observations that have zero budget in our dataset. We threw out all the rows with zero budget to build more accurate models.

Model interpretation:

Rating model

Our relation equation for the rating model is:
Score ~ budget + popularity + runtime + tier_com + release_month + genre_Action + genre_Documentary + genre_Drama + genre_Comedy + genre_Horror + genre_Family + genre_War + genre_Romance + genre_Thriller + genre_Crime + genre_Animation + genre_Western

Model result Interpretation:

The coefficient of budget is negative, which indicates that increasing budget will decrease the probability that a movie will be rated as a good movie. Both popularity and runtime have positive coefficients, which implies that increasing popularity and runtime will make a movie more likely to receive positive feedback. Regarding the production company, movies from a tier 4 production company will have a much lower probability to acquire a good rating than movies from a tier 1 production company.

Apart from that, release months also have significant impact on rating score. Specifically, movies released in May, June, October, November, and December have higher probability of becoming superior ones among peers. Among the total 20 genres, movies that belong to documentary, drama, war, crime, animation, or western genre are more likely to be judged as "good" ones. On the contrary, genres including action, comedy, horror, family, romance, and thriller have negative impact on the probability. Due to the difference of coefficient for each genre, impact varies accordingly, and we will discuss that later.

Revenue model

Our relation equation for the revenue model is:
log(revenue) ~ rating_all + log(budget) + runtime + release_month + genre_Action + genre_Drama + genre_Horror + genre_Family + genre_Adventure + genre_Science_Fiction + genre_Western + genre_Foreign

Model result Interpretation:

From our results, we found that the coefficient for rating is positive, suggesting that a higher rating leads to a higher revenue for a movie when holding everything else constant. It is also true that a relatively higher budget or a relatively higher runtime would lead to a higher revenue for a movie. And different release months have different impact on the revenue of the movie. Basically, movies released in June, July and December have a higher revenue in the end. As for the genre of the movie, action, horror, family and adventure movies would have a higher revenue while drama and science fiction movies would have a lower revenue compared to other types of movies.

Recommendation and Insights:

From our rating model, we have the following insights and recommendations for movie investors and managers of movie production companies:

1. Movies with longer runtime are more likely to be rated as good movie.For instance,typically, the length of movies is 90 minutes or 120 minutes. Assuming a movie whose budget is $26,957,381, popularity index is 9, production company belongs to tier 3, release month is February and belongs to genre music(hypothesis 1). For such a movie, if the runtime is 90 minutes, the probability of being rated as a good movie is 14.53%.At this point, if the runtime increase by 1 minute, the probability will increase by 0.2%. If the runtime is 120 minutes, the probability of being rated as a good movie is 21.57%.At this point, if the runtime increases by 1 minute, the probability will increase by 0.27%.
   Therefore, we highly recommend investors to choose one with longer runtime when they compare two movies if other traits of these two movies are similar.

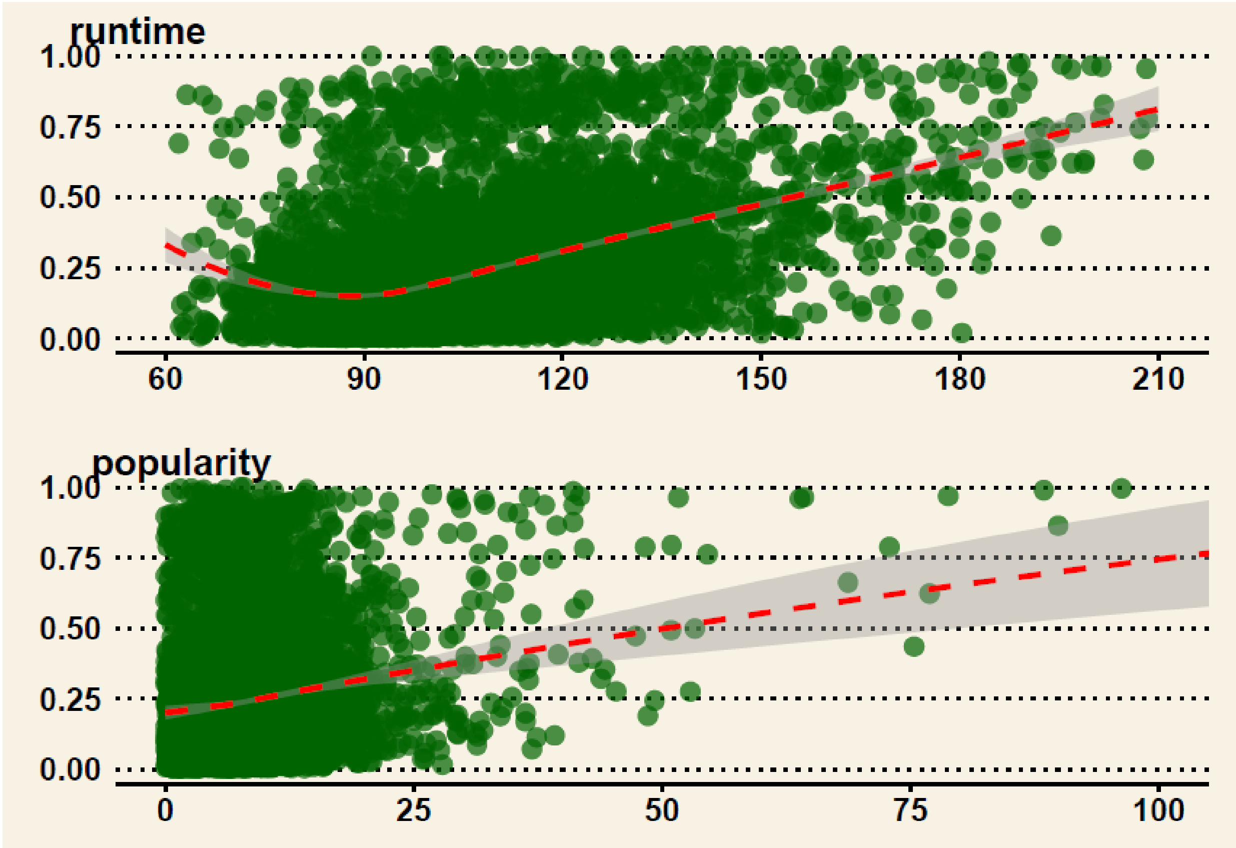
2. Movies with higher popularity are more likely to be rated as good movie.We choose first and third quartiles of popularity index as examples.If the popularity index is 4.6935, the probability of being rated as a good movie is 14.49%.At this point, if the index increases by 1 unit, the probability will increase by 0.87%. If the index is 11.1533, the probability of being rated as a good movie is 20.84%.At this point, if the runtime increases by 1 unit, the probability of being rated as a good movie will increase by 1.15%.
   The positive relationship between popularity and score indicates that investors should attach more importance on most popular movies and give up unpopular ones.

3. Budget: As mentioned in the model interpretation, the budget has a negative impact on rating according to the result of our model. The potential explanation for this unusual result is that many high budget movies are commercial movies produced by big film company who focus more on the revenue rather than the ratings on the movie. We suggest managers of production companies should consider more about other variables rather than budget if they want to produce good movies since high budget may not guarantee a good movie.

4. Company Tier: The higher the tier, the higher the probability of being rated as good movie. For example, if we use the hypothesis 1, the probability will be 68.17% for movie from tier 1 company, 19.16% for tier 2 company, 13.78% for tier 3 company, and 2.36% for tier 4 company. For other scenarios, the probability will vary but the relative position of each tier will remain the same. Thus, we highly recommend investors to choose movies from a tier1 production company.

5. Release month: We know movies released in May, June, October, November, and December have higher probability of being a good movie, but the impact of each month is different. Specifically, if we hold other variables the same as hypothesis 1, movies released in June will increase the probability of being a good movie by 5.2% compared with the seven “normal” months. The ratio will be 9.2% if we release in December.
   According to the above comparison, we see that the gap of probability between “good” month and “normal” month is big. Therefore, we highly recommend managers to release movies in these “good” months.

6. Genre: Different genres have different impact on the ratings of the movie.Specifically, animation movies will increase the probability of being a good movie by 14.8% compared to “normal” genres; horror movies will decrease the probability of being a good movie by 7.92% compared to “normal” genres.So it is important for investors to pay close attention to genres when making investment decisions. This model can also be used by production companies to decide the runtime and ad spend.Longer runtime means higher cost but also higher rating. More ad spend means higher popularity but also higher cost. However, the incremental value of probability will decrease at certain points for both variables. Production companies need to pay attention to these points to make better trade-off decisions.



What's more, this model can also be used to calculate the probability of being a good movie by setting parameters. Quantified probability can be utilized in many ways to support business decisions.

Note: All the numbers in hypothesis 1 are the mean value.Regarding release month and movie genre, we just randomly choose one.The probability for movies of different genre, release month, budget, popularity index, runtime may vary. Model users can set parameters to meet different needs.

As for the revenue model, we have the following recommendations. First, because a higher rating leads to a higher revenue for a movie, we suggest the managers of production companies to pay efforts in improving the ratings for the movie. Then, we recommend managers to increase the runtime in a rational range because a higher runtime means a higher revenue for a movie when holding everything else constant. As for the release months, we think managers should choose June, July and December to release the movies. At last, investors should avoid investing movies of drama and science fiction movies because movies of these genres would lower the revenue. And investors should invest in action, horror, family and adventure genres because they make a relatively higher revenue at the end.