

# STA 610 Case Study 1

Richard Fremgen

14 October, 2022

---

## 1. Introduction

Anytime a new movie begins production or simply starts off as an idea on a piece of paper, directors and producers strive to obtain one main goal - create the next box office hit that will bring in maximum returns. Engineering a successful movie is a challenging task as there are a variety of features that go into whether or not a film is a boom or bust. Producers consider a multitude of features ranging from selecting which production company(s) should sponsor the film, to how much money should be allocated for production (budget), and to even consider when is the optimal time to release a film. All of which play some role in how successful, measured in net profit, a film is while in theaters. With that in mind, the objective of this case study is to investigate what makes for a successful movie; in particular, we are asked to evaluate whether a film's budget and IMDb's critic score are predictive of a film's net profits. The data set used in this project is the `2019_United_States_Film_Releases.csv`, which contains data about films released in the United States in 2019. Included in this data set are variables containing each film's release date, production company, lead cast members, director, box office revenue, budget, run time, IMDb critic score and genre of the film. While these variables are conducive when trying to build a model to predict a film's success, caution had to be taken to transform some variables and completely avoid including certain variables in the models fit to prevent the possibility of overfitting.

## 2. Data Handling

The data used for this report, like most real-world data sets was not short of imperfections that required preprocessing before any exploratory data analysis could occur. Since the response variable (net profit) was not directly included in the film provided, a new variable `Net_Profit` was appended to the data set which is the difference between the `Box.Office` and `Budget` variables. Therefore, `Net_Profit` depicts how much money the movie generated while in theaters, after accounting for expenses, and is a common threshold to gauge whether or not a film is a boom or a bust. Initial analysis of the data set revealed an extensive amount of missing data, as there were 48 film with missing `Box.Office` values and 75 films with missing `Budget` values (missing values were labeled as N/A). Since both `Box.Office` and `Budget` are used to directly calculate the response variable, `Net_Profit`, any rows that contained N/A for either the `Box.Office` or `Budget` columns were removed from the data set. This decision ensured that the every data point used within the model(s) was an actual net profit value, and not an imputed value, which could give misleading results if the data was very noisy for certain groups with small sample sizes.

In addition to choosing to remove these missing values, another area of focus during this initial analysis was to evaluate whether or not there were any variables in the data set that had group-level effects. The first variable that came to mind was `Release.Date`, which corresponds to the the date in which the film was released; the main component of `Release.Date` that drew initial interest was the first two numerical digits of the variable that corresponded to the month in which the movie is released. From my own experience, *a priori* it was believed that the release date month had some type of group effect on the success of a film, as it is common for the most anticipated movies of the year to be released in the month of December. December is typically an attractive month to release a movie, since this is a time when most children are off from school for winter

break, and parents usually take time off for the holidays, therefore making it an opportune time to take the family to the movies. This is also supported by the fact that four out of the six highest-grossing films of all time were all released in December. Aside from December, *a priori* it was believed that some of the summer months like July would have some of the most profitable films, due to similar reason as the December logic, with kids being home for summer break. On the contrast, it is widely known in the film industry that some of the winter months at the beginning of the calendar year (January, February) are referred to as *dump months* for films, where typically the average attendance for movies are down, and new movie releases tend to be lower quality and usually on a lower budget. With this *a priori* knowledge, a new variable called **Month** was added to the data set that represented a two digit encoding of the release date month for each film.

Another hypothesis that was initial considered was if a movie's success could be based on the production company involved with sponsoring the film. Like any industry, production companies come in all shapes and sizes, and with that comes access to different levels of resources in the film industry. *A priori* it was believed that larger production companies that control a larger market share will have the ability to scale faster, and support large film budgets, which was originally believed to have some effect on how successful a film is. These large, mainstream production companies have the breath and reach to not only supply the resources to produce films at the highest quality, but also extensively market the release of a new movie, which can have immense effects on how a film does at the box office.

Movie critic experts typically categorize production companies based on market share into the following three categories:

- **Major Studios:** these are studios such as Universal, Paramount, Warner Brothers, Disney, and Columbia that control majority of the market capitalization in the film industry and release the greatest number of films each year; usually at budgets much higher than films sponsored by the other two categories. Each one of these studios contains at least 5% market share, and contain smaller studios such as Lucasfilm that are own by larger studios, Disney.
- **Mini-Major Studios:** these are studios such as STX Entertainment, Lionsgate, or Metro-Goldwyn-Mayer (MGM). These studios compete directly with the major studios listed above, but having a smaller market share, and typically generate a fewer number of films each year.
- **Other Studios:** these are usually small, independently-owned studios that produce a couple of movies each year, usually on a small budget have only own a small fractional percentage of the market share.

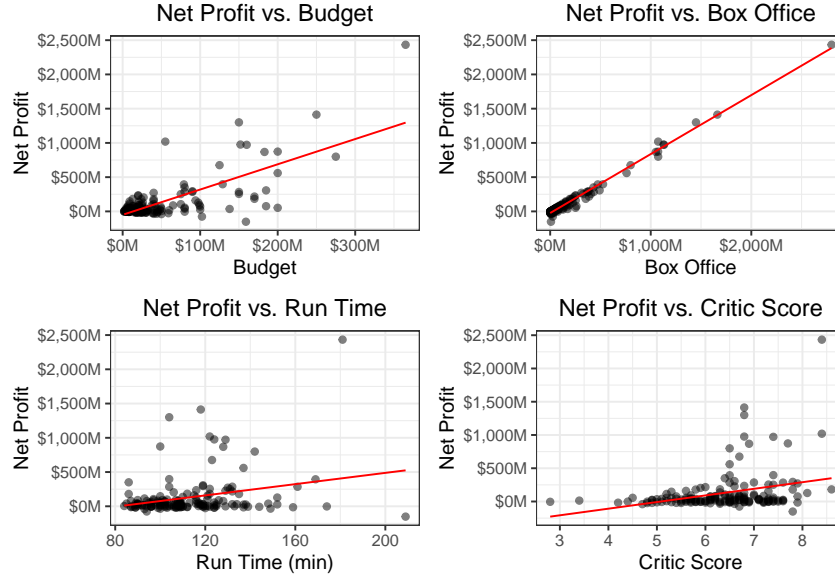
In addition to these three well-known movie studio categories, another studio categorization considered was *Streaming Studios*, such as Netflix and Amazon that have a completely different business model when it comes to releasing movies. Streaming services have begun in the past decade to purchase the rights to movies produced by other studios or produce films on their own. These films are temporarily shown in theaters for a limited amount of time (anywhere from a couple of days to a couple of weeks), before becoming available on their streaming platform for their subscribers to access. This of course means that since our data set only considers the amount of USD generated at the box office to calculate net profit, our values for films sponsored by *Streaming Studios* should differ greatly from traditionally produced movies, because logically it makes sense that a movie released in theaters for a limited period of time will generate less revenue than if the same film was released on theaters for the usual 2-2.5 month period. With that in mind, an additional categorical variable called **Studio** that took on a value of **Major**, **Mini\_Major**, **Other**, or **Streaming** was created to represent the *type* of studio that sponsored the film. In instances where multiple *types* of studios sponsored the same film, the studio with the larger market share was selected for **Studio**. For example if Paramount (Major Studio) and MGM (Mini-Major Studio) both sponsored the same film, then the film's **Studio** variable became **Major** based on the assumption that even though the film was co-sponsored, it would still have access to the same resources as other **Major** Studio sponsored films.

Additionally, the **Genre** variable was investigated for group-level effects and was transformed since the original data set **Genre** variable contained multiple genre types per movie (in some instances). To cope with this, each movie was placed into one of the six most common genres: **Action**, **Comedy**, **Drama**, **Horror**, **Thriller**, or **Other** if the genre wasn't one of the previous five. In instances where a movie had multiple genre's, this

movie was repeated in order to capture individual entries for each genre. For example, if a film was classified as Comedy/Drama in the initial data set, then the film would appear twice in this data set in both **Genre** groups.

### 3. Exploratory Data Analysis (EDA)

Based on the *a priori* beliefs described, EDA was performed to evaluate the relationship between the response variable, **Net\_Profit** and variables of interest that could be included in the fitted models.

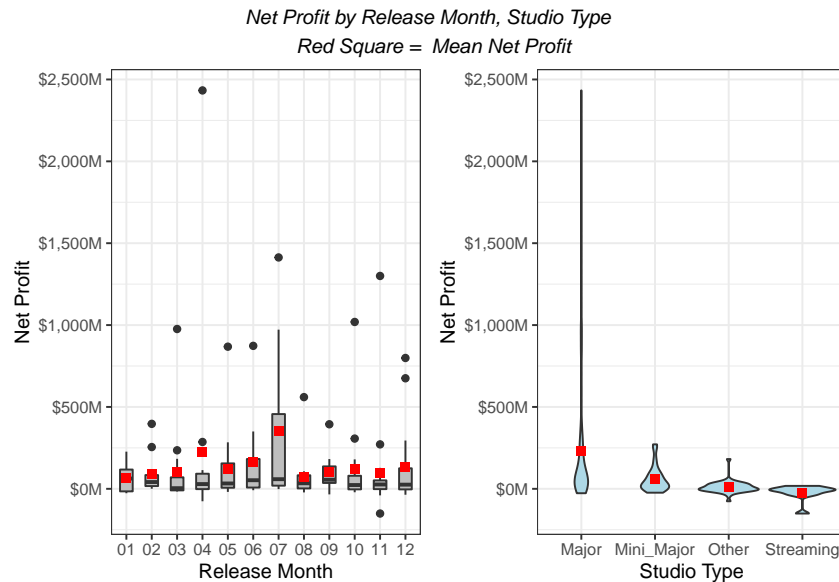


Looking at the plots above, it appears that between *Net Profit* and *Budget* there is relative, strong positive relationship between the two variables, which is supported by the 0.73 correlation coefficient between the two variables. This supports the idea that generally speaking, higher budget films were able to have higher net profits when compared to films produced on a minimal budget. On the contrast, **Critic.Score** only had a relatively weak positive relationship with **Net\_Profit** and a correlation coefficient of 0.31. Though it should be mentioned that the most successful films did generally have critic scores that were above the average score of 6.4 and median score of 6.5. The **Run.Time** variable had a similar relationship with **Net\_Profit**, posting a correlation of 0.27 and a relatively weak positive relationship based on the scatterplot above. **Net\_Profit** and **Box.Office** were nearly perfectly correlated, with a correlation coefficient of 0.99, which should come as no surprise, since **Box.Office** is a measure of revenue, and increasing a film's revenue, directly increases profit. Lastly, it should be mentioned that **Box.Office** and **Budget** were found to be highly correlated, which supports the remark above that increasing the budget of a film, tends to increase the amount of money brought in at the box office. This is important to note because care must be taken when fitting a model to not include both of these variables simultaneously, as one could run the risk of multicollinearity and have a model that is non-identifiable.

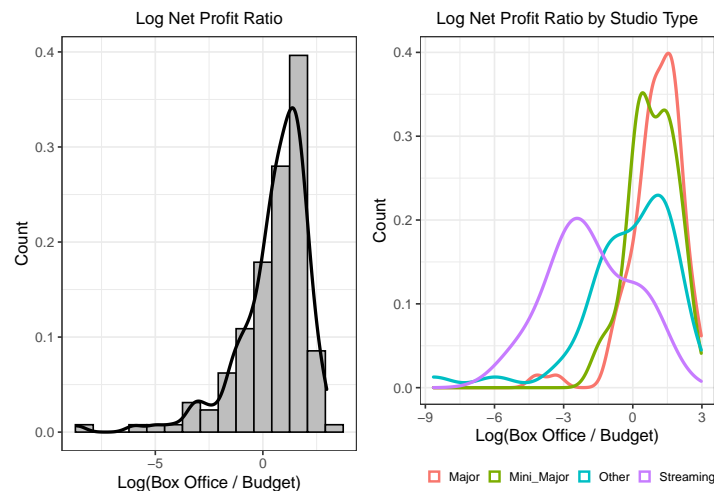
The EDA plots on Page 4 present look at the potential for **Studio** and **Month** to have group-level effects. When considering how **Net\_Profit** varies by release month, it appears that 07 or July was the month with the highest average net profit with \$352 M, whereas 01 or January narrowly out edged July for the highest median net profit with \$63 M. This was both surprising and expected, since we initially suspected that the summer months of June/July would be some of the higher profit generating months, however, January performing so well was a surprise, especially by how much its median value outperformed 12 or December profit values. While grouping by month does provide some interesting insight, one shortcoming of this approach is that the monthly group sizes are very small, as each month has between 7 – 20 data points; both the January and July month groups had the smallest group sizes with only 7, 8 data points respectively.

Additionally, the violin plots depict the relationship between the four different types of studio described in the **Data Handling** section. As expected the **Major** type of studio the highest average net profit with \$232 M, highest median net profit with \$77M, and had a budget that was on average 2 – 4 times the size of the other

studio types. This aligned with our *a priori* belief that the type of studio plays a role in determining the magnitude of a movie's net profit. The violin plots also provide some insight into the distribution of net profit for each studio type. We can see each of the four studio type appear to be unimodal, but have varying degrees of skewness, due to outliers. Additionally, it can be seen that the **Streaming** category exhibits some left skewness, primarily due to the fact that production companies in this studio type (Netflix, Amazon) are more prone to having negative net profits based on the computation method used.



In the later model fitting sections of this report, we will discuss some of the assumptions we make in order to fit certain models; one of which has to do with the normality of the data, which is clearly violated looking at the violin plots on Page 4. To cope with this, a transformation is required in order to make the `Net_Profit` data become more normal, however, such an approach does present challenges, since `Net_Profit` contains negative values, as taking the normal  $\log(\text{Net\_Profit})$  will not work since the  $\log$  is undefined. Another approach is rather than considering net profit as being the difference between `Box.Office` and `Budget`, one can use  $\log\left(\frac{\text{Box\_Office}}{\text{Budget}}\right)$  as the response variable. This so-called *Log Net Profit Ratio* captures the same information that the original `Net_Profit` variable captured, because films that have a higher *Net Profit Ratio* are indicative of greater success, where as films with *Net Profit Ratios* less than 1 indicate that the film lost money, which is also captured by this variable since  $\log(x) < 0 \quad \forall 0 < x < 1$ . The result of implementing this data transformation are depicted below, as it is evident that using this new response variable appears to be much closer to a normal distribution when compared to `Net_Profit`. Notwithstanding, skewness still exists in the density plots below, but nowhere near the level seen on the studio specific violin plots.



The last area where EDA was performed was on the `Genre` variable; due to the page limit requirements of this

report, the box plot comparing the six **Genre** types by **Net Profit** was placed in the appendix of the report. Overall, it is difficult to discern differences in **Net Profit** between the different **Genre** type. Even though the **Other** category has the highest mean and median **Net\_Profit**, such could just be due to how the data was aggregated and how the genre’s were assigned in the first place.

## 4. Variable Selection

Whenever trying to determine which variables in a data set are predictive of a response, it is important to annotate which variables were not included in your analysis and the reasoning behind such a decision. There were numerous columns that were initially dropped from the dataset after initial analysis was performed. For example **Title** and **Director** were the first two variables excluded from the analysis, primarily because these are categorical variables, where each entry is essentially unique, therefore, making it useless to try to use these variables to study how group effects could impact net profit. Additionally, **Lead.Cast.1**, **Lead.Cast.2**, and **Lead.Cast.3** were dropped from the analysis because out of the 726 possible values for these three columns, there were 621 unique actor/actresses recorded. This essentially means that majority of the actors/actresses listed in this data set are used only once and two times at maximum, therefore, running into a similar issue described above where you essentially have a categorical variable with only unique group, which can lead to severe overfitting if used within a model. Additionally, **Box.Office** variable was not considered as a covariate in the succeeding section to prevent there being collinearity between **Box.Office** and **Budget**.

## 5. Model Fitting

### 5.1 One Way ANOVA Analysis

The first step when starting the model selection process was to first evaluate the three categorical values that were transformed: **Month**, **Genre**, and **Studio**. In order to test there is statistical evidence that the means of the groups in these variables are statistically significant, we can use a One-Way ANOVA, which we can express as  $y_{ij} = \mu + \alpha_j$ , where  $\mu_j = \mu + \alpha_j$  and  $\varepsilon_{ij} \sim N(0, \sigma^2)$ . For each of the three models run,  $\alpha_j$  represents the group-level effect for the variable of interest, as the null hypothesis test for each model would be  $H_0 : \alpha_1 = \dots = \alpha_J$ , if we assume that each of the fixed effects variables have levels:  $1, \dots, J$

The actual ANOVA tables and output are depicted in Tables 9-11 of the Appendix when using  $\log(\frac{\text{Box Office}}{\text{Budget}})$  as the response variable. To summarize these findings, there is statistical evidence to reject the null hypothesis for the **Genre** and **Studio** variables, but not for the **Month** variables, indicating that statistically significant differences exist for the different studio and genre types.

### 5.2 Multiple Linear Regression

Another approach to model the success of a movie is to use the covariates in a multiple linear regression model, such as  $y_i = \beta_0 + \beta_1 x_{1i} + \beta_2 x_{2i} + \beta_3 x_{3i} + \varepsilon_i$ , where  $\beta_1$  is the coefficient for **Run.Time**,  $\beta_2$  is the coefficient for **Budget** and  $\beta_3$  is the coefficient for **Critic.Score**. The results of running this model are depicted below, and illustrated that run time, budget, and critic score are all statistically significant predictor based on a significance level of 0.05. While this knowledge is useful for our analysis, this model had very poor fit with an adjusted R-squared value of only 0.09.

Table 1: MLR Model Results

	Estimate	Std. Error	t-value	P-Value
(Intercept)	-1.1609	0.9662	-1.2014	0.2315
Run_Time	-0.0199	0.0079	-2.5207	0.0128
Budget	0.0000	0.0000	2.1824	0.0306
Critic_Score	0.5719	0.1588	3.6008	0.0004

### 5.3 Random Effects Model:

While one can obtain decent results using a fixed effects or classical multiple linear regression model, incorporating random effects into the model could be very beneficial when trying to model the effect that **Studio** or **Genre** has on the net profit of a movie. In such a case we can model a random intercept model for **Studio** or **Genre** by using the model:  $y_{ij} = \mu + \alpha_j$ , where  $\mu_j = \mu + \alpha_j$ ,  $\varepsilon_{ij} \sim N(0, \sigma^2)$ , and  $\mu_j \sim N(\mu, \tau^2)$ , which is equivalent to  $\alpha_j \sim N(0, \tau^2)$ . Similar to the One-Way ANOVA model,  $\alpha_j$  represents the group-level effects that **Studio** (or **Genre**) have on net income for groups,  $j = 1, \dots, J$ . The major advantage of placing a random effect on these two categorical variables, is that it uses *shrinkage* estimators to share information between groups. Such a feature is crucial when a data set has groups that contain a couple of data points, as is the case with the Streaming group and with a couple of the genre groups. Additionally, placing a random effect on a variable places a distributional assumption on such a variable, which in our case we assume that  $\alpha_j$  or  $\mu_j$  (depending on the parameterization) is normally distributed; which is why transforming the net profit variable to reduce some of the variables skewness was such a crucial data transformation step. With that in mind, two random intercept models were fit (one for **Studio** and one for **Genre**) and are compared below. Clearly, based on a lower BIC, AIC and Deviance using a random intercept model with **Studio** provided a much better model fit than a random intercept model with **Genre**.

Table 2: Random Intercept Model Comparison

Random Effect	AIC	BIC	logLik	Deviance	df.resid
Studio	585.8904	595.0207	-289.9452	579.8904	152
Genre	842.1266	852.2246	-418.0633	836.1266	211

### 5.3 Mixed Effects Model:

While fitting a random slope model produced promising results, the purpose of conducting this case study was to evaluate whether a film's budget and critical score are predictive of its net profits. To incorporate these covariates into the model, I treated **Budget**, **Critic.Score** and **Run.Time** as fixed effects and incorporated them into the random slope model for **Studio**, therefore creating a linear mixed effects model. Such a model could be written as  $y_{ij} = \mu_j + \beta_1 x_{ij} + \beta_2 x_{ij} + \beta_3 x_{ij} + \varepsilon_{ij}$  where  $\mu_j \sim N(\mu, \tau^2)$ ,  $\varepsilon_{ij} \sim N(0, \sigma^2)$ , and  $\beta_1, \beta_2, \beta_3$  represent the fixed effects for the covariates of interest. One area of evaluation that was considered when fitting these mixed effects models was whether or not to include the **Run.Time** covariate, as EDA revealed mixed results for this variable's relationship with net profit. The results of fitting a mixed effects model with and without **Run.Time** are depicted below.

Table 3: Mixed Effects Model Comparison

Random Effect	AIC	BIC	logLik	Deviance	df.resid
With Run Time	574.5868	592.8473	-281.2934	562.5868	149
Without Run Time	576.5958	591.8130	-283.2979	566.5958	150

Model comparison reveals mixed results when comparing AIC to BIC, as the values for each model are very similar, but certainly both represent an improvement from only fitting a random intercept model. With that being said I decided to select the parsimonious model without the **Run.Time** variable since it had a minimal impact on the model, which supports what was discovered during EDA. In addition to the two mixed effects models fit above, several additional mixed effects models were fit with a random slope effect on **Budget**, however, each of these models resulted in models with higher AIC/BIC values and did not improve model performance. (See Appendix for further details)

## 6. Final Model

As mentioned above, the final model selected was the mixed effects model with **Budget** and **Critic.Score** as fixed effects along with a random effect on **Studio**; mathematically this model can be expressed as  $y_{ij} = \mu_j + \beta_1 x_{1ij} + \beta_2 x_{2ij} + \varepsilon_{ij}$  where  $\mu_j \sim N(\mu, \tau^2)$ ,  $\varepsilon_{ij} \sim N(0, \sigma^2)$ , and  $\beta's$  = Fixed Effects for Budge and Critic Score. One thing to note about the model fit below, was that in order to get the parameters on the same scale, **log(Budget)** had to be used as a covariate, rather than **Budget** in the **lmer()** function. This was due to the fact that the values in **Budget** were so immensely large (on the order of tens or hundreds of millions) when compared to the **Critic\_Score** or the Log Net Profit Ratio that ended up being used as the response. From an analysis perspective, when considering the uncertainty quantification of it is important to evaluate which intervals of our estimates of interest contain zero and do not contain, in order to objectively say whether or not a variable has a significant effect on the response. Looking at the results below we can see that the confidence interval for **Critic\_Score** is **(0.22, 0.72)**, once again supporting the idea that **Critic\_Score** has a statistically significant effect on a film's net profit. Additionally it can be seen that the estimator for **log(Budget)** had a confidence interval of **(-0.34, 0.12)**, which casts doubt on the statistical significance of budget on how successful a film is, a finding that counteracts our initial *a priori* belief and results found during EDA. When considering the random effect of **Studio**, we can visualize the model's coefficients (Appendix), which depict that **Major** studio had the greatest studio-group effect on the Net Profit Ratio, which supports our *a priori* belief and EDA work that large studios (**Major**) would be in a more advantageous situation to produce successful films.

Table 4: Final Model: Random Effects

Groups	Name	Variance	Std. Dev
Studio	(Intercept)	1.359144	1.165823
Residual	NA	2.094618	1.447280

Table 5: Final Model: Fixed Effects Coefficients

	Estimate	Std. Error	t-value
(Intercept)	-1.2275	2.0784	-0.5906
log(Budget)	-0.1073	0.1168	-0.9192
Critic_Score	0.4716	0.1261	3.7402

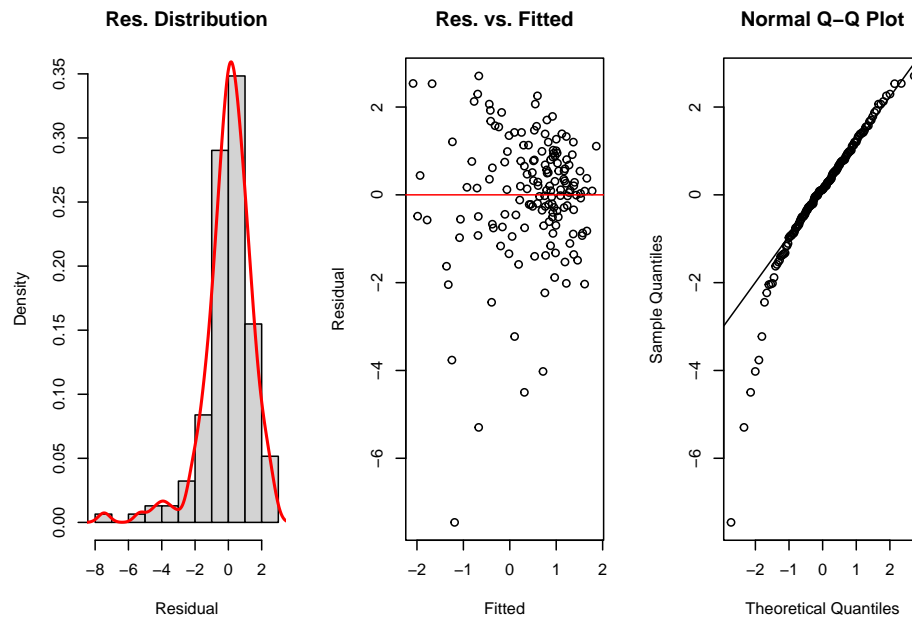
Table 6: Final Model: Confidence Intervals

	2.5 %	97.5 %
.sig01	0.5765	2.9968
.sigma	1.2980	1.6275
(Intercept)	-5.3267	2.8904
log(Budget)	-0.3383	0.1244
Critic_Score	0.2220	0.7205

## 7. Model Diagnostics

From a model assessment perspective, one of the main things we want to check is that the assumptions we made about the data hold after we fit the model. For example, we can use the residual histogram below to verify that our assumption of  $\varepsilon_{ij} \sim N(0, \sigma^2)$  holds. Looking at the distribution plot below, we can see that based on the distribution plot,  $\varepsilon_{ij}$  appears to be approximately normally distributed, although there does appear to be some left skewness in the residual histogram. Because residuals represent the variation that is left unexplained by

the model, it is important to compare Residual values to Fitted values to evaluate whether or not there are distinct patterns present. Ideally, we would like to see the residuals be plotted randomly around the dotted zero reference line, indicating that the variances of the error terms are equal, and a linearity assumption about the relationship holds. Looking at the plot below, we can see that for the most part such is found to be true, as the residuals are generally equally spread on both sides of the dotted reference line without an explicit pattern present. One could, however, say that there are slight degrees of the residuals starting to “fan out” as the fitted value decreases, potentially due to outliers, but overall this plot generally supports that a linear relationship exists. Finally, the Normal Q-Q plot is used to visualize whether the residuals are normally distributed. Looking at the plot below, we can see that the residuals generally follow the dotted reference line, indicating normality, but do however veer off at the beginning of the line, thus indicating some left skewness in the data as previously mentioned. Overall analysis of the residuals revealed that our normality and linearity assumptions generally held, although it is evident that skewness still remains in the data, despite attempts at data transformations.



## 8. Conclusion and Model Limitations:

In this project, factors that impact a movie’s success were analyzed and modeled to determine what quantitative or qualitative variables could be predictive of how much net income a film will generate. In particular, a movie’s budget and IMDb critic rating were considered to evaluate whether or not these variables are associated with net profits. EDA revealed a strong positive correlation between budget and net profit, while critic score was weakly positively correlated with net profit. The type of studio that sponsored a film was found to have significant group-level effects on a movie’s success and was included in our evaluation of the data. Modeling of the data set revealed that implementing a parsimonious mixed effect model with studio as random effect and critic’s score and budget as fixed effects produced the best fit model that maximized the amount of variation that could be explained by the model. Output of the final model revealed that the critic’s score was a significant predictor of a movie’s success, while budget was found to not be a statistically significant predictor of net profit. There are several limitations of this analysis, one of which is evidence in the **Model Diagnostics** section, where it is still evident that the data fit to the model still has varying degrees of skewness, even though the log ratio transformation did help the analysis to an immense degrees. Further analysis regarding this data set, should focus on optimizing the transformation on the response variable, in order to fully meet the normality assumptions expression. Additionally, analyzing how to properly transform the budget variable for the final mixed effects model is another area that could be improved upon, to ensure that all covariates are somewhat scaled properly. Lastly, one of the biggest hurdles to this analysis was the amount of missing data present in the initial data set. While removing these instances where missing budget or box office values was necessary for this analysis performed, a potential way to improve the quality of the model developed is to place greater focus on the data collection process, or to study a data imputation method to fill-in the missing values.



## 9. Appendix

**Note:** this section of the case study is not to be included in the 8 page requirement and is used to portray supplemental material. Attached to this Sakai submittal is also the .Rmd that was used to generate this document along with all supporting code

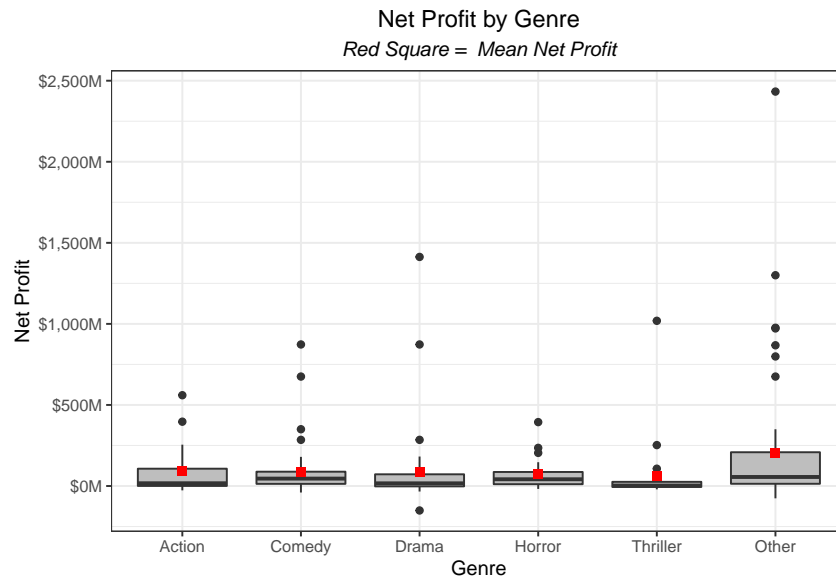
### 9.1 EDA

Table 7: Correlation Between Quantitative Variables

	Box Office	Budget	Run Time	Critic Score	Net Profit
Box Office	1.00	0.81	0.30	0.30	0.99
Budget	0.81	1.00	0.38	0.21	0.73
Run Time	0.30	0.38	1.00	0.51	0.27
Critic Score	0.30	0.21	0.51	1.00	0.31
Net Profit	0.99	0.73	0.27	0.31	1.00

Table 8: Studio Type Summary Statistics

Studio	Count	Avg Net Profit (\$ M)	Med Net Profit (\$ M)	Avg Budget (\$ M)	Avg Critic Score
Major	80	231.80	77.40	72.61	6.48
Mini_Major	25	60.38	30.50	31.19	6.26
Other	40	10.04	3.20	16.46	6.19
Streaming	10	-22.74	-7.52	32.70	6.85



### 9.2 One Way ANOVA Results

Table 9: Month One-Way ANOVA Model Results

	Df	Sum Sq	Mean Sq	F value	Pr(>F)
Month	11	28.9134	2.6285	0.9025	0.5398
Residuals	143	416.4939	2.9125	NA	NA

Table 10: Studio One-Way ANOVA Model Results

	Df	Sum Sq	Mean Sq	F value	Pr(>F)
Studio	3	98.5028	32.8343	14.292	0
Residuals	151	346.9045	2.2974	NA	NA

Table 11: Genre One-Way ANOVA Model Results

	Df	Sum Sq	Mean Sq	F value	Pr(>F)
Genre_Type	5	45.8444	9.1689	3.2459	0.0076
Residuals	208	587.5560	2.8248	NA	NA

### 9.3 Two Way ANOVA Results

Table 12: Two-Way ANOVA Model with Interaction Results

	Df	Sum Sq	Mean Sq	F value	Pr(>F)
Studio	3	116.91	38.97	17.00	0.00
Genre_Type	5	18.81	3.76	1.64	0.15
Studio:Genre_Type	13	57.42	4.42	1.93	0.03
Residuals	192	440.26	2.29	NA	NA

### 9.4 Mixed Effects Model Results - Random Slope

```
## boundary (singular) fit: see ?isSingular
```

```
## boundary (singular) fit: see ?isSingular
```

Table 13: Mixed Effects Model Comparison - Random Slope

Random Effect	AIC	BIC	logLik	Deviance	df.resid
With Run Time	578.5647	602.9121	-281.2823	562.5647	147
Without Run Time	580.5803	601.8843	-283.2902	566.5803	148

### 9.5 Final Model Supplemental Information

Table 14: Final Model Studio Coefficients

Studio	Coefficient
Major	-0.5476324
Mini_Major	-0.7924375
Other	-2.1125485
Streaming	-4.2433215

## 10 References

- *A First Course in Bayesian Statistical Methods* (Hoff)
  - *CSS/STAT 560 - Lecture Notes on Hierarchical Modeling* (Hoff)
  - *STA 610 Lecture Notes*
  - *STA-602 Lecture Notes*
  - *Statistics: Market Share of Leading Film Studios in the US from 2010-2021*
    - <https://www.statista.com/statistics/187171/market-share-of-film-studios-in-north-america-2010/>
-