Final Report

BAN 602

CSU East Bay

Prepared for Professor Bala

Oct 13, 2019

"The Most Profitable Hollywood Movies from 2007 to 2011"

Created by Group 4

Ankit Pushpam | Guojun Wang (Kevin) | Kanmani Natarajan

Sharmada Krishna | Uday Patel | Yinjia Liu

Table of Contents

Section I - Introduction

A.) Introduction (motivation)

Movies play a major role in the entertainment industry. People love to be entertained especially through watching movies. Indeed, they do not want to spend their valuable time and money to watch something boring. Hence it is useful to know the following questions before watching any movie.

- **What factors should be observed to select a movie?**
- **Whether an audience should rely on critic score?**
- **Does a box office hit movie mean higher ratings?**

The data set titled "Hollywood's most profitable stories" sheds some light on answering the above questions. The dataset contains information about Hollywood's 74 most profitable stories released from the year 2007 to 2011. Out of 74 movies, five movies are removed from analysis due to missing information. Overall, the entire dataset is quite clean and no major data tidiness issues.

B.) Data Cleaning & Description

Each movie contains information about the following:

1. **Title of the movie**
2. **Genre**

   a. Comedy
   b. Drama
   c. Romance
   d. Animation
   e. Action
   f. Fantasy

3. **Lead studio**

   a. Fox
   b. Universal
   c. 20th century
   d. Sony
   e. Disney
   f. Warner Bros.
   g. Lionsgate
   h. Summit
   i. The Weinstein Company
   j. New Line
   k. Paramount
   l. CBS
   m. Independent

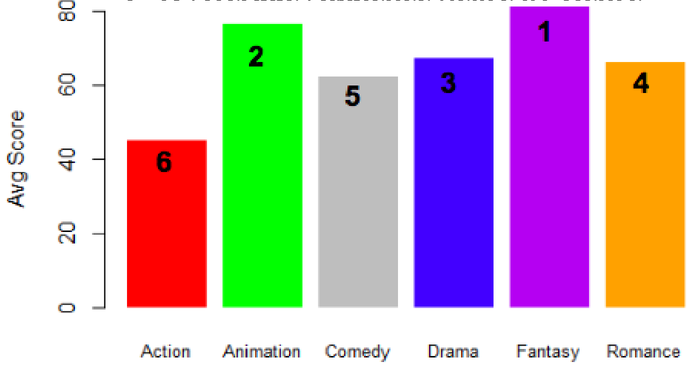
4. **Audience scores**: on a scale of 0-100 (0 being the worst, 100 being the best)
5. **Profitability**: a percentage of worldwide gross
6. **Rotten Tomatoes** on a scale of 0-100 (0 being the worst, 100 being the best)
7. **Worldwide Gross**: in millions of dollars
8. **Year**: between 2007-2011

Section II - Data Wrangling

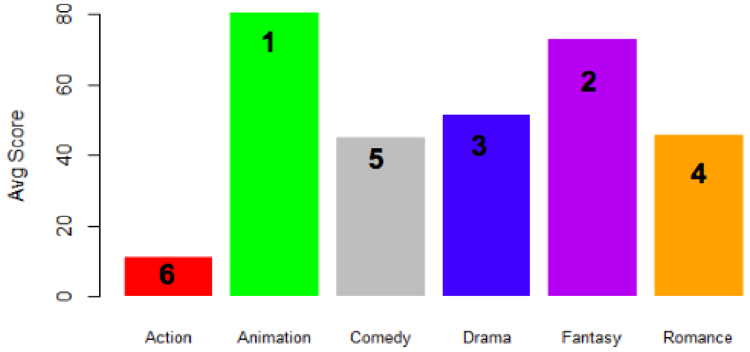
A.) Data Exploration

1.) Genre at glance

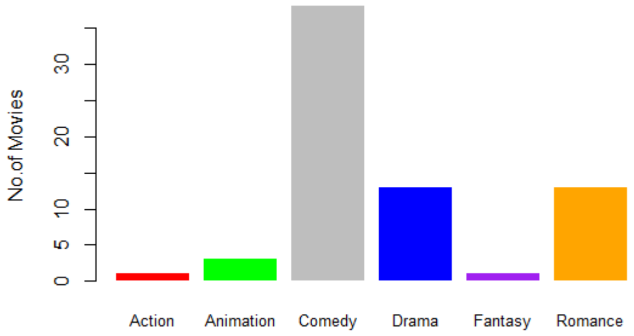
F-1: Average Audience Score by Genre



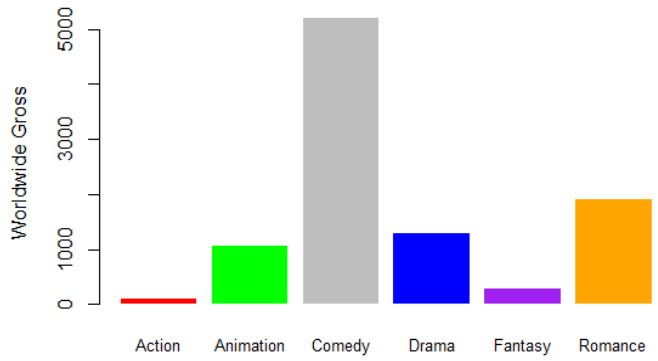
F-2: Average Rotten Tomato Score by Genre



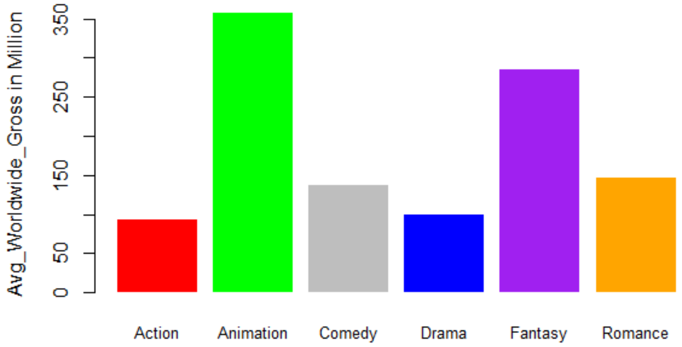
F-5: Number of Movies by Genre



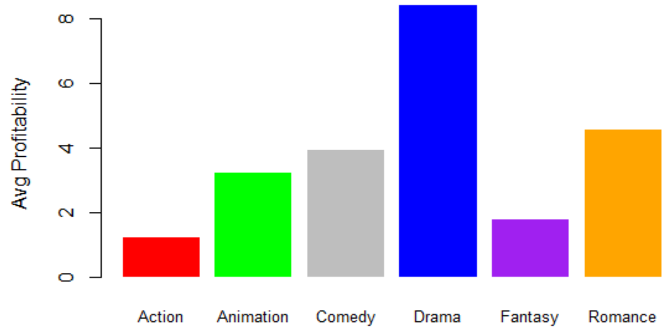
F-6: Total Worldwide Gross by Genre



F-3: Average Worldwide Gross by Genre



F-4: Average Profitability by Genre



Observation for Genre

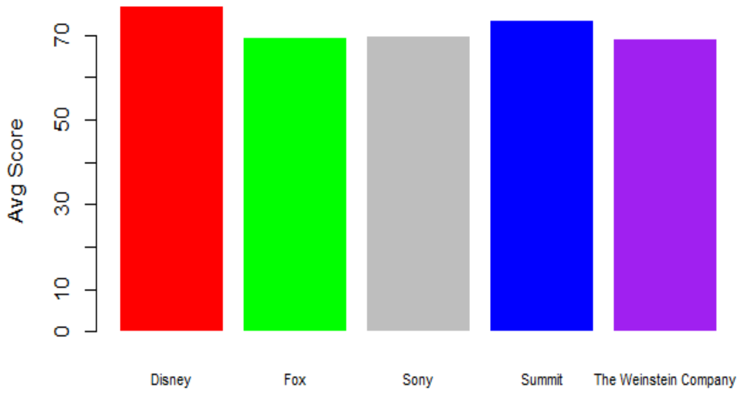
When comparing all the genres, Fantasy and Animation seem the best performing as the average audience and critics score is the highest for Fantasy: (81, 73) and Animation: (76, 80) respectively (F1-2). Animation is performing slightly better than Fantasy in the average worldwide gross. Drama is the most profitable of all genres at 8.4% average return. Comedy has the highest total worldwide gross (F-6): the worldwide gross is helpless as its biased to the number of films within a genre.

Analysis by Genre is lop-sided

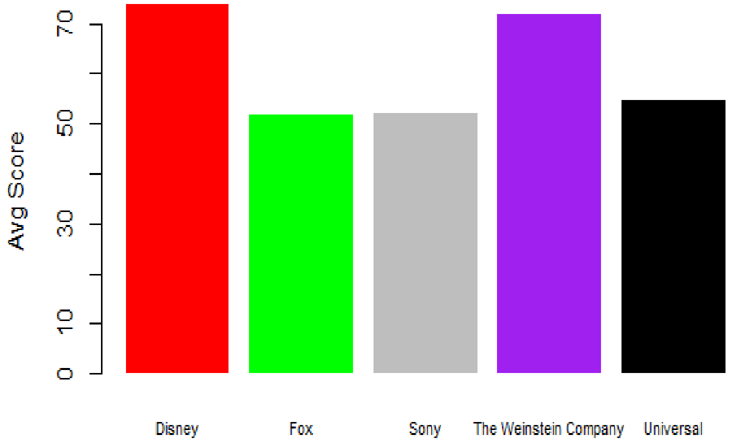
The reliability of the above observations may be challenged because the number of movie in a genre are unevenly distributed. For instance, Animation and Fantasy together account for only 4 out of 69 movies. Whereas Comedy represents 38 of 69 films. Although the ratings are averages, making an estimation of the film genres based on number of films as few as 3,1,1 (Animation, Action, Fantasy) is less likely to represent the population. Hence, less reliable.

2.) Top 5 Studios at glance

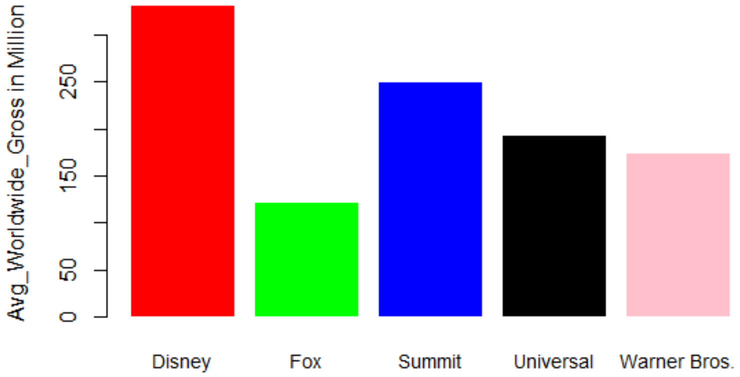
F-7: Average Audience Score by Studio



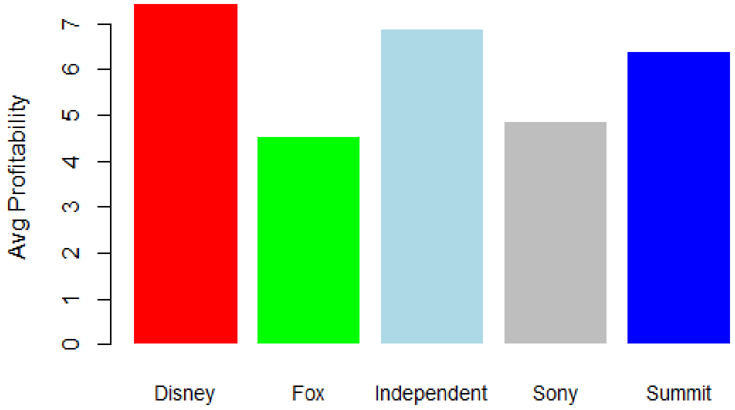
F-8: Average Rotten Tomato Score by Studio



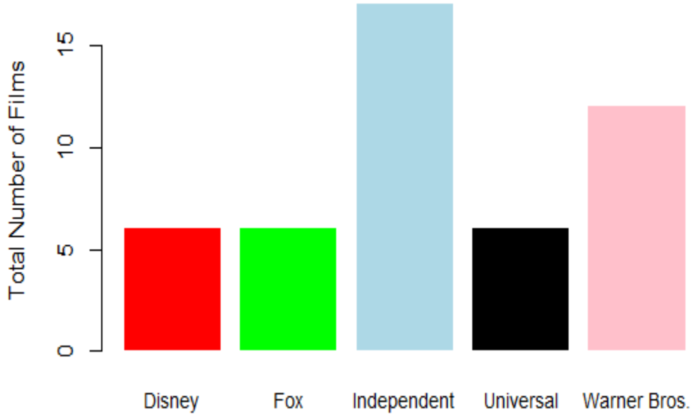
F-9: Average Worldwide Gross Score by Studio



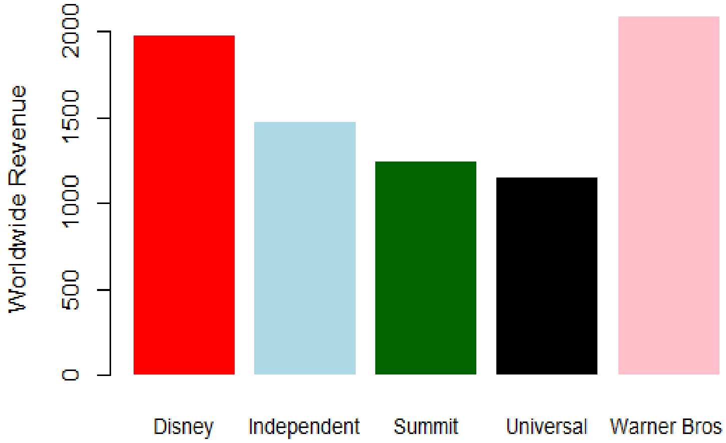
F-10: Average Profitability by Studio



F-11: Total Number of Films by Studio



F-12: Total Worldwide Gross by Studio



Observation for Studio

Disney has performed best in most categories:

a.) **Audience Score:** 76.5

b.) **Critic Score:** 73.67

c.) **Average Worldwide Gross**: 329.59

d.) **Average Profitability:** 7.40

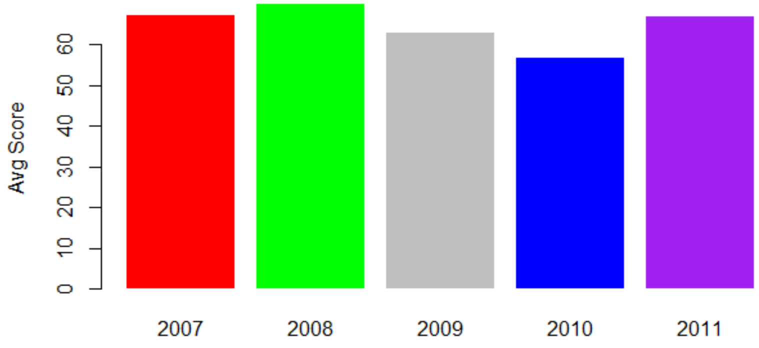
e.) **Number of films:** 6

f.) **Total Worldwide Gross:** $1977.56 Million

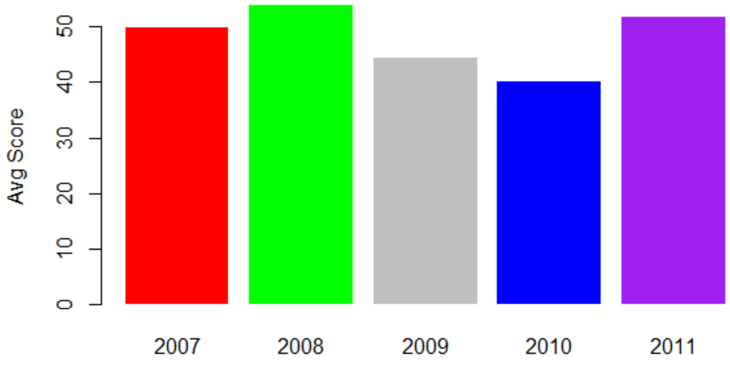
A combined look at genre and studio graphs, it can be said that Disney is responsible for the Animation genre performing well. The only three animation movies are produced by Disney. It produced the least number of movies and earned the second highest amount of total worldwide gross. The set of independent films represent a decent chunk of total films (17 out of 69); however, no independent film appears on the top five average Audience or Rotten tomato scores.

3.) Years at glance

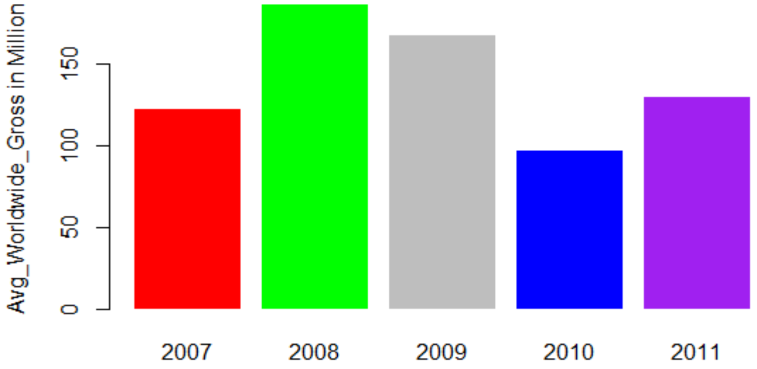
F-13: Average Audience Score by Year



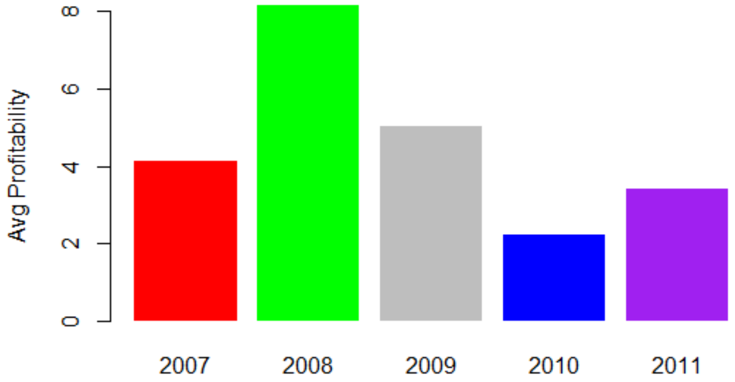
F-14: Average Rotten Tomato Score by Year



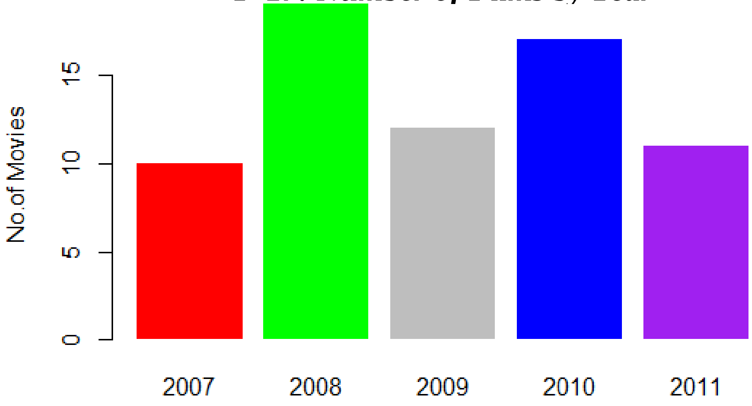
F-15: Average Worldwide Gross by Year



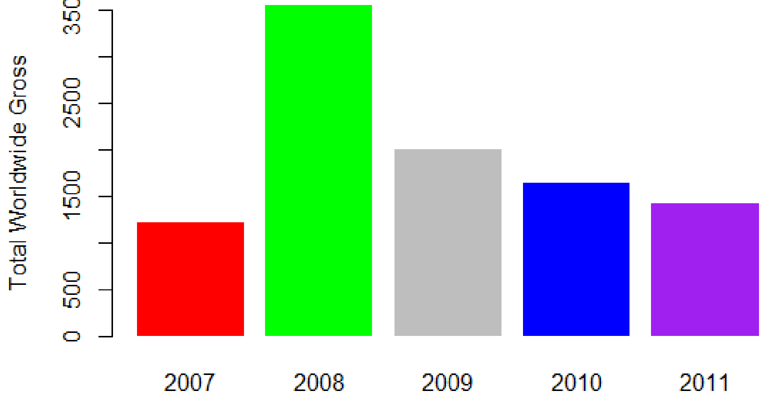
F-16: Average Profitability by Year



F-17: Number of Films by Year



F-18: Total Worldwide Gross by Year



Observation for Year

The above graph explains that the 2008 has the greatest number of movies and total worldwide gross income is the also highest. We can observe a pattern that if the number of films made in the year is more, then the worldwide gross income is also high for that year. Year 2008 is the best year as it wins in each category. We would conclude the best quality of movies were produced. 2010 is the worst year in terms of quality of movies produced because it does produce almost the same number of movies as in 2008 but the worldwide gross is very low compare to 2008. And, Average profitability and reviews is also the lowest for the year 2010 compare to 2008.

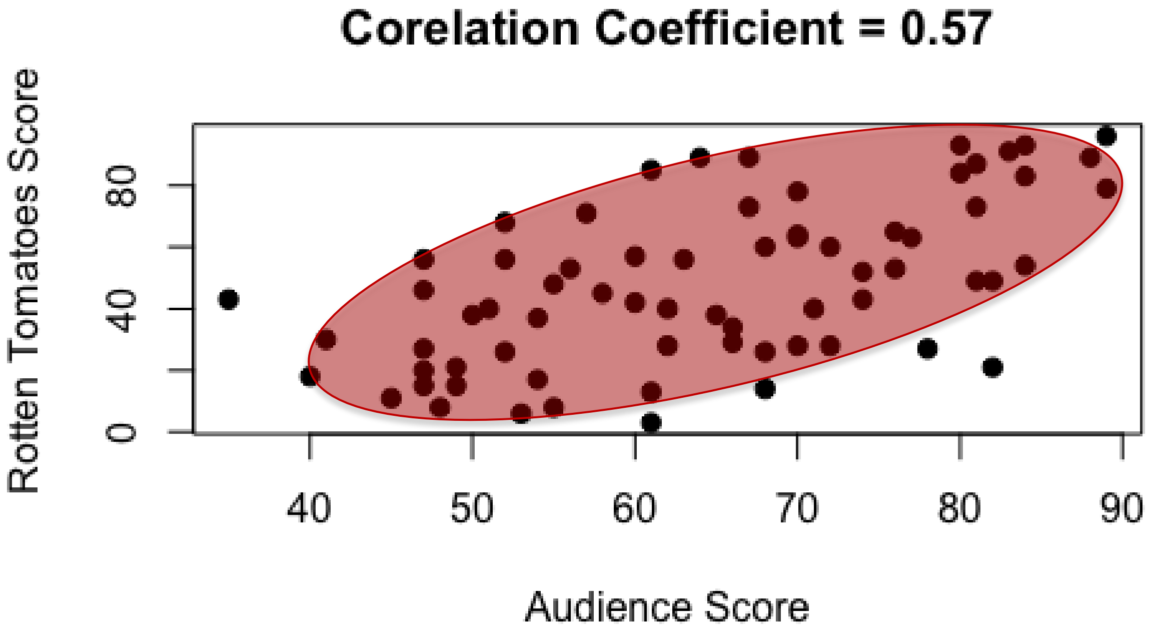
(2008 was a bad year for the economy, but definably not for the Film industry ☺.)

4.) Exploring the Ratings and Worldwide Gross

Statistic	Audience Score	Rotten Tomato Score	Worldwide Gross (in Millions of $)
Min	**35**	**3**	**0.025**
Quartile 1	**53**	**27**	**32.59**
Mean	**64.46**	**47.88**	**142.6**
Quartile 3	**76**	**65**	**205.30**
Median (Q2)	**65**	**46**	**79.18**
Max	**89**	**96**	**709.8**

a.) Five point Summary for Ratings

The mean score of the audience score is 64.46 and Rotten Tomato score is 47.88
Based on the 1st quartile and 2nd quartile values we observed that half of the score of Audience Score lies between 53 to 65 however the Rotten Tomato Score lies between 27 to 46, this shows that Audience give the score to movie more liberally then Rotten Tomato critics. The minimum score of the Audience Score 35 and Rotten Tomato score is 3 also explain the same thing.
Most of the movies worldwide gross income lies between $32.59 to $79.18 million.
There is a block buster movie $709.8 millions of worldwide gross.



b.) Correlation between Scores

The correlation coefficient is 0.57. This suggests a moderate positive relationship between the audience and rotten tomato scores. Furthermore, it can be evidenced that the scores follow a shared space on the plot above. There is some pattern in the two scores with will be analyzed in detail later.

Section III - Data Analysis

Statistic	Standard Deviation	F-Test P-value	Mean of Difference	Pair-Test P-value
Audience Score	**13.61**	**Less than 5%** (2.709742e-07)	**16.58** Range: 11-21	**Less than 5%** (1.246e-08)
Rotten Tomato Score	**25.98**			

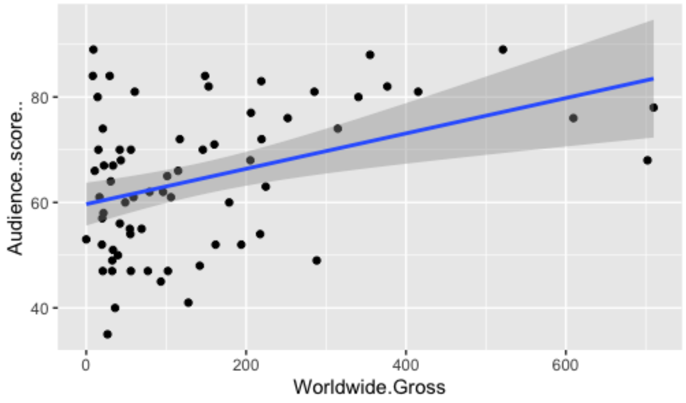
A.) Variance in Ratings (F-test)

The audience score varies less than rotten tomato score as evident by the standard deviation above. The F-test is conducted to confirm that claim. The p-value here is less than 5%; therefore, it confirms that the variation in the two scores is different. The higher variation in the Rotten Tomato score suggests that the critics are tough graders and compared to the audience.

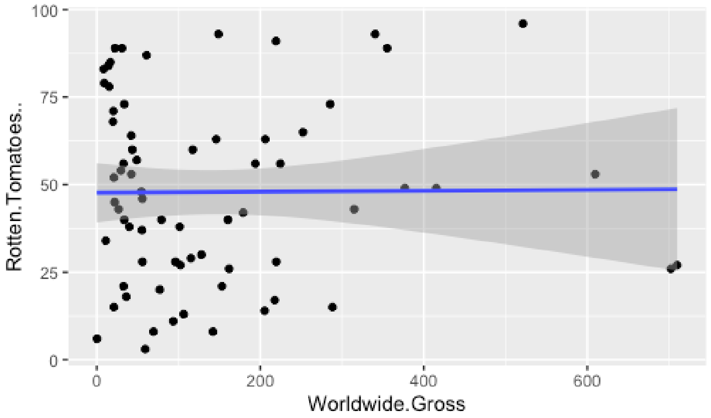
B.) Mean difference between Ratings (Paired Test)

Based on the p-value which is less than 5%, we can conclude that the population mean for audience scores are different from the population mean for tomato scores. This may be observed with the average difference between the movie ratings of 16.58. For example, an audience may rate a movie at 66.58 while the critic may rate it at 50. With the 95% confidence, we can estimate the difference interval for these two population means to be between 11.47 and 21.69.

F-19: Worldwide Gross & Audience Score



F-12: Worldwide Gross & Rotten Tomato Score



	b_1	**r square**	β_1
Audience score	**4.676**	**15.68%**	$\neq 0$
Rotten tomato score	**0.053**	Less than a 1% (7.379579e-05)	$= 0$

C.) Relationship between Ratings & Worldwide Gross (Simple Linear Regression)

Earlier, we propose the third question “Does a box office hit film mean higher ratings?”. We would like to refer to their relationship to give us the support for our final choice because a high box office film is often considered as a successful film.

In our dataset, the Worldwide Gross represents the box office and Audience Scorer Rotten tomato score represent ratings. In other words, we would like to know if Worldwide Gross are related positively to the Audience score or Rotten tomato score.

1.) Worldwide Gross & Audience Score

a. We use the least squares method to find the coefficients and obtain the estimated the regression equation:

$$\widehat{y} = -158.822219 + 4.676582 * x$$

The slope of the estimated regression equation (b1=4.676) is positive, implying that as Audience score increases, Worldwide Gross increase. In fact, we can conclude that an increase in the Audience score is associated with an increase of $4.676 million in Worldwide gross.

b. test for significance: p-value=0.000755 < 0.05, we can reject null hypothesis which give us evidence that $\beta_1 \neq 0$ that means the linear relationship between Worldwide gross and Audience score exists.

c. Assessment of the appropriateness using the regression equation:

We can compute the coefficient of determination r^2 =0.1568551, now we can conclude, there are about 15% of the variability in Worldwide gross can be explained by the linear relationship between Audience score and Worldwide gross. So, our regression equation is not so good.

2.) **Worldwide Gross & Rotten Tomato Score**

a.) We can obtain the regression equation:

$$\widehat{y} = 140.10324786 + 0.05314204 * x$$

The slope of the estimated regression equation (b1=0.053) is positive, implying that as rotten tomato score increases, Worldwide gross increase. Worldwide gross is expected to increase by $0.053 million per rotten tomato score.

b.) test for significance: p-value=0.9441 > 0.05, we cannot reject null hypothesis which give us evidence that $\beta_1 = 0$ that means the linear relationship between Worldwide gross and rotten tomato score does not exist.

c.) assess the appropriateness of using the regression equation:

We can compute the coefficient of determination r^2 =7.379579e-05. So, this time we obtain a worse model, then we would not use it to make any prediction.

Since there is no linear relationship between rotten tomato score and worldwide gross, we will not suggest anything this time. It’s up to you if you would like to refer to worldwide gross when a film has a high rotten tomato score.

Section IV - Conclusion

A.) Limitations and possible future research

The data set “Hollywood’s most profitable stories” had access to limited information. There are many other factors which may be influential in making a movie more successful in terms of worldwide gross and popularity. Some of the marketing strategies are identified and discussed below.

- Effect of actors in the movie

  Though it may be difficult to draw conclusions about the direction of causality, for some movies it is possible to say that the involvement of star actors is critical to the success of those movies. The stronger a cast already is, the greater the impact of a newly recruited star with a track record of box office successes or with a strong artistic reputation [1].

- Effect of title of the movie

  For instance, when the movie “Tangled” was first put into production, the film was promoted as having the title *Rapunzel Unbraided*, which was later changed to *Rapunzel*. To market the film to both sexes and additional age groups, Disney changed the film's name from *Rapunzel* to *Tangled* while also emphasizing Flynn Rider, the film's prominent male character, showing that his story is just as important as Rapunzel’s [2].

- Movie plot and connecting with large audience

  Movie plot with less controversial theme and its ability to connect with a large audience play a major role in successful movies. For instance, if the movie plot belongs to a small demographic/ ethnic group and unable to connect with a large audience, the movie may not become successful.

- Proportion of viewers who read or write reviews on the internet for the movies that they watch

  Reviewer score play a major role in attracting the audience to watch a movie. If a group of people control the rating of any movie, it may distort the reviewer score rating. Also, some reviewers may be biased against genre, actor or the film maker. Hence it is important to know the proportion of viewers who read or write reviews on the internet for the movies that they watch.

- List of unsuccessful movies with similar information as given in the data set “Hollywood’s most profitable stories”

  Data on unsuccessful movie may shed some light on the effect of audience & rotten tomatoes rating in successful & unsuccessful movies.

B.) Recommendation

1.) **What movie to watch from the dataset?**
We will recommend Watching Wall-E or Tangled: Audience Rating: 90/100 (Grossed over $350 mil).

2.) **Audience is a decent indicator to select a film**
Based on our analysis we can say that the Audience Score is relatively better indicator. However, Critic score & Worldwide gross are may not be the best estimate to choose a film. The audience watching a film may not have the same standards as the critics; hence; a lower critic score does not mean audience will dislike it.

A high worldwide grossing film does not necessarily mean high audience score. According to our regression model, worldwide gross is explain by only 15% of Audience score. Certain movies like “My Week with Marilyn” and “A Dangerous Method” received high audience score; yet, earned less than $10 million.

Section V - Reference

Dataset:
https://public.tableau.com/en-us/s/resources?qt-overview_resources=1#qt-overview_resources

Actor Effect
[1] http://www.people.hbs.edu/aelberse/papers/hbs_06-002.pdf

Title Effect
[2] https://www.latimes.com/archives/la-xpm-2010-mar-09-la-fi-ct-disney9-2010mar09-story.html