**ECONOMETRICS CIA 2**

**Course Code: Eco.5.08**

THE HUNDRED CRORE CLUB

**Name Roll No: UID No**

Divya Somakumar 036 131141

Roveena D’Souza 037 131406

Sayali Jadhav 044 131004

Jerrin Anna Joe 046 131288

**INTRODUCTION**

According to India times in 2013-14, Bollywood as an industry had contributed around 1% to the total GDP of our economy. Bollywood has been one of the fastest growing industries in India and according to the latest reports in 2012, 2013 and 2014 its contribution has been about 1 % to 1.5%. of the GDP. Apart from its contribution to GDP it is a growing part of the economy that generates employment to people as well. The film industry has been an influential player in the global economy. Bollywood movies have been doing extremely well in India as well as other parts of the world.. With respect to movies producers and directors may use all strategies like budget, cast screens locations to make their movie more attractive and appealing to the audience but how well a movie does at the box office and its fate over there is completely decided by the audience. The success of the movie is highly unpredictable. Thus through this analysis we have tried to find factors that influence the box office revenue.

The main objective of this analysis is to find the factors that influence Box Office revenue.The body of knowledge regarding the significance and importance of variables that tend to affect the box office collection has been constantly maturing. The first multiple regression model to predict to predict the financial success of a film that is precisely box office collection was developed by Litman in 1983.According to him films are very unpredictable. According to him no one knows how a movie will fare in the market place. With science and technology on the rise the film industry in our case bollywood has turned extremely competitive. There is rat race among producers and directors to enter the 100 crore category. The question rises as to why 100 crores? It is because according to our secondary sources of data that was the minimum benchmark that was set for the top grossing movies in the period 2013-2015. Each and every producer wants to spend exorbitantly and wants to revive that amount too. So basically how does this happen? It is primarily through box office collection. In our analysis we have tried to find the factors that influence the box office collection. The dependent variable that we have chosen in our study is **box office** collections in India as well as overseas. We also identified a few independent variables which contribute to the success of the movie These variables are **Budget** of the movie, **Critics review**, **YouTube likes** , **YouTube views** , Fan **following** on social media, **Rank of the actors**.

**DATA COLLECTION**

For this analysis we obtained secondary data. Our dependent variable is Box office collection. The Box office collection is quantified in terms of Indian Rupees.We have only considered the domestic box office collection. We have not included the box office revenue earned from other parts of the world. We collected the data from the website “Box office India”. The link for the same is provided below. We considered those movies which grossed above 100 crores in the Box Office during the years 2012 to 2015. We have not adjusted Box office revenue with inflation rate because we have only considered three years i.e. 2012 to 2015. In this analysis we have six independent variables.

**INDEPENDENT VARIABLES**

The first independent variable is the budget of the movie. The budget of the movie includes three components these are total production cost, Total promotion cost , and total distribution cost. These three totals were aggregated for each film and this generated a total value of budget of each film. We collected the data regarding budget of the movie from the website “Box Office India”.

The second independent variable is the Critics review. We first collected the critics review from several sites like The Times of India , Hindustan Times, NDTV , Rajeev Masand. We quantified the critics review by taking an average of the critics ratings of the above sites.The links for the same is provided in the Reference section.

Our third independent variable is the Rank of the actors .We chose Rank of the actors as an Independent variable because the success of the movie depends on the lead actor of the movie.The rank of actors was extracted from the TIMESCELEBZ database where ranks of actors are published in every 6 months according to the performance of the actors and the hits delivered from those actors in six months. The recent ranking of June 2015 have been used for our analysis.

Our fourth independent variable is the Fan following on social media because We used the number of likes on Face book and number of followers on twitter in order to quantify our independent variable. The fifth independent variable was Youtube likes for the movie trailer. We chose this variable because Trailers play a significant role in the success of the movie. We used the number of likes on YouTube for the movie trailer.

**THE R RESULT**

The main objective of our project is to determine the factors that contribute to the success of a movie. In our analysis we have considered six independent variables and one dependent variable. Through the study we have tried to find out the functional relationship between the dependent variable i.e. Box Office collection and independent variables like budget of the movie, critics review, fan following on social media , Ranks of the actors, YouTube views for the trailer and YouTube likes for the trailer.

For this analysis we took a sample of 20 Bollywood movies which earned more than 100 crores at the Box office during the years 2012 and 2015.We took a small time interval because the value of money doesn’t remain constant in the long run.

The Regression equation is given by

Y = ß1X1 + ß2X2 + ß3X3 + ß4X4 + ß5X5 + ß6X6 + U

Where Y = Box office collection

|  |  |  |  |
| --- | --- | --- | --- |
| Independent Variable(X) | Estimate ( ß vale) | Std. Error | P value |
| Budget | 1.249e+00 | 4.902e-01 | 0.0243 |
| YouTube likes | -1.388e+04 | 1.828e+04 | 0.4614 |
| Rank of actors | -7.702e+07 | 4.246e+07 | 0.0929 |
| Critics | 3.150e+08 | 3.150e+08 | 0.0402 |
| Social media | 3.088e+01 | 9.751e+01 | 0.7565 |
| Youtubeviews | 8.676e+01 | 8.142e+01 | 0.3060 |

From the above table we can predict whether our hypothesis is accepted or rejected .In all there are six independent variables and there are six hypotheses which were tested for possible rejection.

**Hypothesis 1**: There is a positive correlation between Budget of the movie and box office collection.

In the First case, we considered the effect of Budget of the movie on the Box office collection. The positive ß value of 1.249e+00 indicates positive relation between the two variables. The p value 0.0243 validates our hypothesis at 5 % level of significance. This implies that there is a positive correlation between budget of the movie and the box office collection i.e. a 1% increase in the budget leads to a 1.249% increase in the revenue collected at the box office.

**Hypothesis 2**: There is a positive correlation between YouTube likes of the trailer of the movie and box office collection.

In the second case, we tried to find a relation between YouTube likes for the trailer and Box office collection. We assumed that there would be a positive relationship between the two as we live in the 21 st century which is also called as information age. But unfortunately the ß value for YouTube likes for the trailer is -1.388e+04 which imply a negative relationship. Thus YouTube likes for the trailer doesn’t play a significant role in the box office collection. Thus our hypothesis is rejected at 5% level of significance due to the high p value of 0.4614.This may have happened due to sampling error.

**Hypothesis 3**: There is a positive correlation between the Rank of the actors in the movie and box office collection.

We also tried to find whether the lead actors affect the success of the movie. The p value for the rank of the actors was 0.0929 which indicates that the lead actors of the movie also play a significant role in the success of the movie. But the negative ß value of -7.702e+08 imply that the box office collection is not always determined by the rank of the actors. Higher the rank of the actor doesn’t really imply that the box office collection would be high.

**Hypothesis 4**: There is a positive correlation between critic’s review of the movie and box office collection.

In case of the critic’s review the p value of 0. 0402 that we obtained indicates that our hypothesis is not rejected. The ß value of 3.150e+08 indicates that there is a strong positive relation between the number of stars awarded by the critics for the film and the box office collection.

**Hypothesis 5**: There is a positive correlation between fan following of the movie on social media and box office collection.

Our R results indicated that there is a positive relation between the fan following on the social media sites like twitter and face book and the box office collection ie the ß value was 3.08e+01. Thus higher the fan following on social media higher the revenue collected at the box office. But the p value which is just 0.7565 indicates social media is not a very important factor that determine the collection at the box office. As a result our hypothesis is rejected at 5 % level of significance.

**Hypothesis 6**: There is a positive correlation between number of views of the movie trailer on YouTube and box office collection.

The ß value for the number of views of the trailer on You Tube is 8.676e+01which shows a positive relationship between number of views of the trailer of the movie and the box office collection. But the p value of 0.3060 indicates that this variable doesn’t play a significant role in the collection at the box office.

The value of the multiple is the coefficient of multiple determinations in case of multiple regression model. It helps in identifying how well the regression model fits the data. Our value of the multiple is 0.579 which indicates that only 50% of the variation in the box office collection is determined by the independent variables. We have got a lower value because the box office collection is not only determined by the above variables but also the human behaviour and human behaviour is not predicatble.

The adjusted is a modified version of the that has been adjusted to the number of predictors in the model. It compares the explanatory power of regression models that contain different numbers of predictors. The adjusted R-squared increases only if the new term improves the model more than would be expected by chance. It decreases when a predictor improves the model by less than expected by chance. The adjusted R-squared can be negative, but it’s usually not.  It is always lower than the. Our adjusted is 0.3847.

Our overall p value is 0.04672 which is lower than 0.05. Thus it can be inferred that our hypothesis is not rejected at 5% level of significance. Thus through this analysis we can conclude that budget of the movie, Rank of the actors, critics review, fan following on social media and YouTube likes and views of the trailer have only 50% impact on the Box office collection.

**Weakness in our Regression model**

The ß value of the intercept is -9.331e+08 which indicates that there is an overall inverse relationship between the dependent variables and independent variables. The large standard deviation which has been observed in the table has been caused due to the small sample size. There is sharp multicollinearity in the independent variables of YouTube likes and YouTube views. The number of likes is the subset of number of views. Several independent variables are considered insignificant due to faulty selection of the sample. Our model only explains a 50% effect on the box office. The remaining 50% of the box office collection is determined by other proxy variables. Movies are associated with people. Moods of people keep on varying. As a result we cannot capture the main factors that attribute to the success of a movie

Several weaknesses in the data analysis might have been responsible for the weak support. The success of the film to a large extent depends upon the cast of the movie, the story, production house, performance of the actors etc. We were unable to capture the effectiveness of the story, genre, performance etc. It was difficult to determine whether a release date affects the box office collection. Had there been any proxy variables available to quantify the above factors a strong regression model would have been constructed.

**CONCLUSION**

The main purpose of this research was to establish which factors, if any, contribute to the success of the films at the box office. Using a multiple regression model, a number of variables believed to influence box office success were evaluated. The results of this regression analysis provide evidence that specific variables are factors in determining box office success. Out of the six variables considered in this study, the most significant contributors to the box office were Budget, Critics acclaim and the Rank of Actors. Budget of the movie has a positive association with the success at the box office while YouTube likes and rank of actors showed a negative relation with the box office revenue. There is direct relation between fan following on Social networking sites and box office collection. The analysis has not showed a strong evidence to support our hypothesis i.e. it showed partial (50%) support. The other three variables came forward as contrary to our expectations.

Funding a movie is a high risk/ high return proposition as in any production decision in the field of entertainment. Hence budget plays an instrumental role which calls for ‘smart budgeting’. Film makers must spend more on promotional activities which determines maximum revenue for the film. Importance must be given to critic review since they are the perfect judges of the quality of films. The regression analysis reveals that Rank of the actors even though less significant does have influence on the box office success which might be because of the consumer bias towards certain actors.

**REFERENCES**

1. Budget and box office collection <http://www.boxofficeindia.com/>
2. Budget of the movie <http://www.koimoi.com/>
3. Critics Review <http://www.hindustantimes.com/>

<http://timesofindia.indiatimes.com/>

<http://www.rajeevmasand.com/>

1. Youtube Views and Youtube likes <https://www.youtube.com/>
2. Fan Following on Social media <https://www.facebook.com/>
3. Rank of the actors <http://timescelebex.com/>
4. Basic Econometrics, Gujarati Damodar, Mc Graw Hill.Fourth Edition.
5. For calculating the Multiple regression we have used the Statistical Software Called R.