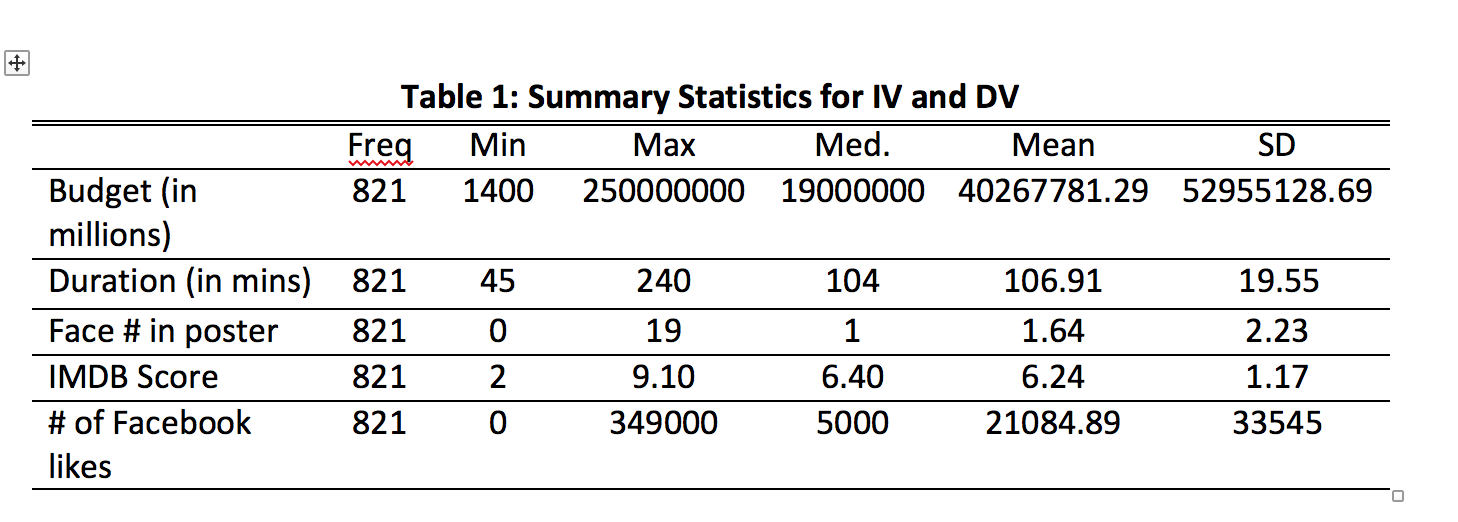
|  |  |
| --- | --- |
| To: | [Lebron’s sporting agency elite athletes, ESPN, Janzen Consulting Group] |
| From: | Janzen Consulting Group: Clarence Fernandes] |
| CC: | [ESPN’s data analytics hiring team] |
| Date: | [05/13/2019] |
| Re: | [Highest movie potential] |

LeBron’s recent moves to Los Angeles has not only caused a stir in the NBA, but also Hollywood. Movie directors are lining up for the chance to book LeBron in a movie. After LeBron’s historical movie success in train wreck the numbers show LeBron equals box office success. On the other hand, LeBron and his movie agency are looking to create and hire a director to make their own movie. Since this will be LeBron’s first full movie published by his agency, he wants the best possibility at success for the movie. LeBron and his team believe that there is no relationship between a movie's IMDB score and other variables. The more superstars he hires for his move, the more likely it will be rated higher.

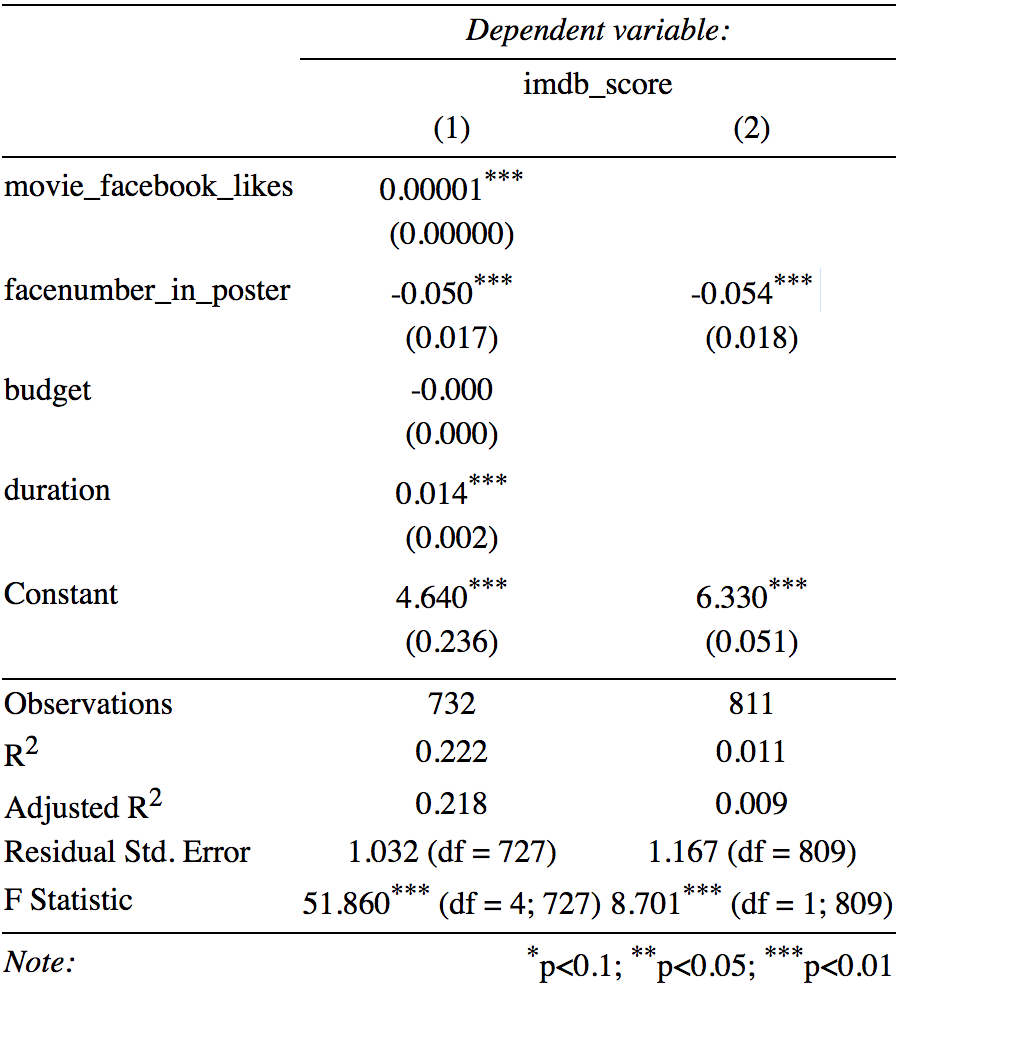
To test this hypothesis, we will be using the IMDB dataset to perform a multiple linear regression test. To predict the IMDB score of a movie based on their number of Facebook likes for a movie, number of faces on a movie poster, duration, and budget of a movie.Our target population for this study will be every movie from the year 2013 to 2016. Our sample is every movie listed in the IMDB data set within those years. Using an OLS model we predict having a lot of big name faces, large budget, a lot of Facebook likes and long duration associated with the movie will predict better ratings and success for the movie. We suspect that as the number of faces increases the IMDB score will go up. We think the more actors in a movie will draw more fans to the movie. Since the people watching the movie may be big fans of the actor it may correlate to a higher rating.

When investigating this model, we had to create a new subset data set from the original IMDB data set. With this data set we could focus only on the movies that were released from 2013 - 2016. Looking at this new subset we didn’t have any missing or outlier values. We also didn’t have to recode any of the variables used. In our dataset budget is recorded in millions and duration is recorded in minutes. 

The IMDB score variable contains values from 2.0 to 9.1. The average IMDB score from our data set was 6.8. So not very high to begin with. The faces in a movie poster variable contains values from 0 - 19. The average number of faces in a movie poster from 2013 - 2016 was 1.64. Lastly according to our summary of IV and DV on the side, we can see the four groups have the same frequencies.

After running our regression model, we had an overall p-value of 2.2. At an area of significance of 0.05, we can conclude that our findings are not statistically significant. This means we are safe to reject our null hypothesis. When checking our simple linear regression assumptions, we noticed almost all the assumptions are passed. The only assumption failed was influential outliers. Looking at our residuals vs leverage plot we can see there are some dotted lines. This lets us know there was a problem.

When it came to analyzing the data, we found that the overall adjusted R2 value of .22 tells us these values are not a good predictor of a movie's IMDB score. Our data also has an effect size of .28. This means there is a slightly large effect size of all our IV’s on our DV, but with a p > 0.05 this was not statistically significant.

Using the coefficients from the regression model and holding all variables constant we can tell that, a 1 unit change in the number of faces in a poster corresponds to an average decrease of 5.01 in the IMDB score, we can determine the results are statistically significant at P< 0.05. A one-unit change in budget amount for a movie corresponds to an average decrease of 1.04 in the IMDB score. We can determine that the results are not statistically significant at P > 0.05. We also found a one-unit change in duration of a movie corresponds to an average increase of 1.38 in the IMDB score. We can determine that the results are not statistically significant at P > 0.05. At a one-unit change in the amount of Facebook likes on a movie corresponds to an average increase of 1.08 in the IMDB score. At a P> 0.05 we can determine that the results are not statistically significant. 

We were able to determine for each 1 unit increase in budget, duration and Facebook like. These variables were not a significant predictor of a movie's IMDB score. However, it should be noted our study has some limitations. Our dataset has a small sample of movies with number of faces on a poster above 10. In the dataset, there are about 164 movies with 1 face, while there are about 14 movies with 10-15 faces. This small sample size of movies with a lot of faces may have contributed to the skewness of the data. Also, note our data only includes movies from the year 2013 – 2016. There could’ve been more movies to come out before 2013 that may have had a higher IMDB score with a bigger number if faces on a poster. This sample we drew from could’ve affected our overall results. Analyzing every movie listed on IMDB may have proved to be a more definitive result. Also, we aren’t able to generalize our findings to the population of movie’s rating. Our data is examined from the IMDB data set. Their rating system is taken into account by every member of IMDB. This means there could be people who may leave a biased review or leave a bad rating just because.

