Memorandum

DATE: October 9, 2020
TO: Chris Acme
FROM: John Trotto, George Tavlarios, Rebecca Redmond, Wendy Merchant
SUBJECT: IMDB Project Proposal

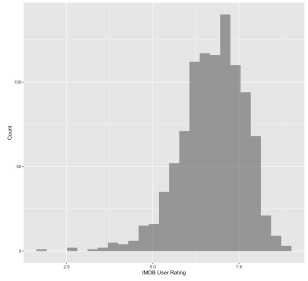
Summary:
This memo aims to educate you on our statistical analysis that we have recently begun that relates to movie performance at the box office. We are looking specifically at how much revenue a movie brings in and what factors help drive a movie's revenue. We are looking at a number of variables including both quantitative and categorical ones as well as some interaction variables in our analysis. This memo illustrates that our analysis could show an insight on the many different factors that go into a movie and how that affects its performance at the box office, which could be valuable to customers of Acme.

Introduction:
In this report, we are looking at data from the well-known movie site, IMDB. The dataset includes information on the 1,000 most popular films according to IMDB that were all made within the time period of 2006-2016. This dataset includes a lot of information on how well the movie performed through metrics like revenue and ratings. Along with this there is also information about the movie itself including the director, main actors, and genre. This dataset is neat and complete meaning there are no missing values for any of the 1,000 movies included. Our team has identified and analyzed a few variables that look to be important in understanding a movie's revenue performance including the difference between its rating and metascore, the genre of the film, and the runtime of the movie related to genre.

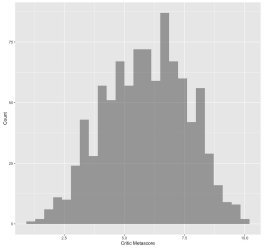
Data Overview:

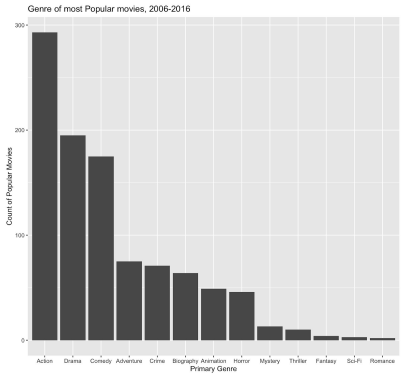
Comparing Rating and Metascore of movies.
The movies in the data set have two measures of likeness: a Rating and Metascore. The movie's rating is the average score (1 being the worst and 10 being the best) the movie received among IMDB users. The Metascore is the average score (out of 100) the movie received by critics. Metascores have a normal distribution, while user ratings are more left-skewed. It is possible an IMDB user may have more bias towards a movie they are rating, causing user ratings to be higher than metascores on average. This information is useful because it will be interesting to see which score better predicted a higher revenue for a movie. Are critic scores more relevant than user scores for a movie's financial success? Note: to have a better comparison, metascores were adjusted to be out of a scale of 10 and not 100 for better comparison to the ratings, that are out of 10.

Distribution of User Ratings



Distribution of Critic Metascores





Popularity of Genre.

The data set includes the genre each movie is categorized as, which will be helpful to see if the most popular genres had a high rating and revenue stream. As shown by the graph, the action genre counts for almost a third of all movies in the data set.

Genre & Revenue.

Since each movie can be categorized by multiple genres, we analyzed the effect of genre on revenue, in order to determine the optimal mix and/or focus of genres to maximize revenues. Additionally, the data's temporal attributes will allow us to build a model that attempts to demonstrate shifts in genre-trends. Parameters from our initial MLR–OLS model (*see table, right*) suggests a significant relationship for approximately half of the genres, with significantly different coefficients, which may inform our future customers on various genres on which they should focus or avoid.

	Coef.	Std.Err.	t	P>|t|	[0.025	0.975]
ACTION	41.6561	7.2622	5.7360	0.0000	27.4021	55.9101
ADVENTURE	74.4239	8.0558	9.2386	0.0000	58.6124	90.2354
ANIMATION	96.6134	14.7604	6.5455	0.0000	67.6424	125.5844
BIOGRAPHY	25.3401	11.9835	2.1146	0.0348	1.8195	48.8606
COMEDY	25.2451	6.6792	3.7796	0.0002	12.1353	38.3548
CRIME	19.3969	9.1443	2.1212	0.0342	1.4489	37.3448
DRAMA	2.2007	5.8737	0.3747	0.7080	-9.3279	13.7293
FAMILY	32.3454	13.7142	2.3585	0.0186	5.4279	59.2630
FANTASY	60.0575	10.4372	5.7542	0.0000	39.5720	80.5431
HISTORY	21.4762	18.9221	1.1350	0.2567	-15.6631	58.6156
HORROR	-11.2460	10.0181	-1.1226	0.2619	-30.9091	8.4170
MUSIC	23.4720	23.5823	0.9953	0.3199	-22.8143	69.7583
MUSICAL	50.9489	40.4209	1.2605	0.2078	-28.3873	130.2852
MYSTERY	12.3877	10.2124	1.2130	0.2255	-7.6567	32.4321
ROMANCE	13.0156	9.2603	1.4055	0.1602	-5.1602	31.1913
SCI-FI	60.7996	9.8935	6.1454	0.0000	41.3810	80.2181
SPORT	34.9890	23.2920	1.5022	0.1334	-10.7274	80.7053
THRILLER	31.9454	8.0914	3.9481	0.0001	16.0640	47.8269
WAR	9.9462	28.8110	0.3452	0.7300	-46.6026	66.4950
WESTERN	41.9748	40.3490	1.0403	0.2985	-37.2203	121.1700

Business Relevance:

We believe that the information this analysis will provide will be of use to the Acme Company and potential future customers. The movie industry is large and continues to grow and change, so companies in the industry will look for information on how to make critically and commercially successful films. We therefore believe that this analysis may provide future opportunities for Acme and is worth pursuing. Questions we would like to answer include:

- Which rating had a stronger relationship with the film's revenue?
- Do specific directors or actors tend to make high-earning movies?
- How does the genre of a movie relate to revenue?
- How does the runtime of a movie relate to its genre, and does this affect ratings?

Please let us know if you have any questions or concerns.