

Seasons of Success: Effects of Seasonality on Film Profitability

Team name: LAME

Members: malexan5, eshelley, areed6, lgamalel

Hypothesis

Over the years, the terms “summer blockbuster” and “Oscar season” have become part of the vocabulary of even non-movie-buffs. So, our group was interested in determining whether the implications of these terms may be found manifested in the data. For that reason, we were interested in testing the hypothesis that certain genres of film are generating more or less profit over time in certain seasons. More specifically, we were interested to test the classic tropes and whether they have become more or less true over time, namely that, all else equal, summer-released action, comedies, and kids films, dramas in the winter, and horror films in the fall show increased profit over time.

Data

The data is gathered from two Kaggle datasets (Rotten Tomatoes and IMDB). These datasets were joined on title and release date using the Python pandas library. From the IMDB dataset, we used release_date, revenue, and budget. From the Rotten Tomatoes dataset, we used tomatometer rating. We cleaned our dataset of several rows with incomplete column values and those with nonsensical values (e.g. zero budget and/or revenue). The data is mildly skewed towards smaller budget and more recent films. The release date was used to split the data into four seasons using dummy variables. Our dataset contained the following columns of interest: original_title, budget, revenue, year_released, tomatometer_rating, inflationAdjustedProfit (created using historical CPI data) along with the 4 season dummy variables, and five genre dummy variables (comedy, action, horror, drama, kids).

Findings

Claim #1: The distribution of profit by season is not uniform and has become less uniform over time.

Support for Claim #1: We separated the population of films into three categories: films released before 1980, films released from 1980 to 2000, and films released from 2000 onward. We performed chi squared tests on each of these three groups. The first two chi squared tests resulted in decreasing, though statistically insignificant p-values. The final test resulted in a statistically significant p-value of 0.001. Furthermore, when performed on the entire population of films, the test resulted in a statistically significant p-value of 0.0002. Therefore, we can conclude that the distribution of profit by season is non-uniform. Given that we accepted the null hypothesis that the variables were independent for the chi-squared tests on films before 1980 and from 1980 to 2000, and we rejected the null hypothesis for films from 2000 onward, this indicates that the distribution became less uniform from the first two periods to the third.

Claim #2: There is statistically significant evidence to support that, all else equal, over time, summer-released action and kid movies showed increased average profitability year to year.

Support for Claim #2: We ran a multiple regression controlling for year, budget, and critic review looking at the effect of interaction terms between each season and year (effects of season over time). When run on movies from 1960 onwards, we find statistically significant p-values for interaction terms for summer-released action and kids movies with positive coefficient. This suggests that there is a strong probability of correlation where, all else equal, summer-released action and kids movies have increased profits over time on average. See table below.

Action Films post-1960 (n=581)			Horror Films post-1960 (n=225)		
Variable	Coefficient	p-Value	Variable	Coefficient	p-Value
Winter-Year Interaction	-1.857e+06	0.714	Winter-Year Interaction	1.394e+06	0.178
Spring-Year Interaction	-2.813e+06	0.536	Spring-Year Interaction	-6.903e+05	0.490
Summer-Year Interaction	7.872e+06	0.044*	Summer-Year Interaction	-1.675e+06	0.091†
Fall-Year Interaction	-1.677e+06	0.691	Fall-Year Interaction	9.109e+05	0.348

Drama Films post-1960 (n=906)			Kid Films post-1960 (n=171)		
Variable	Coefficient	p-Value	Variable	Coefficient	p-Value
Winter-Year Interaction	-2.145e+06	0.002**	Winter-Year Interaction	-1.823e+06	0.517
Spring-Year Interaction	1.976e+05	0.804	Spring-Year Interaction	5.071e+06	0.219
Summer-Year Interaction	4.709e+04	0.945	Summer-Year Interaction	4.716e+06	0.045*
Fall-Year Interaction	3.789e+05	0.088†	Fall-Year Interaction	-4.596e+06	0.066†

Comedy Films post-1960 (n=723)		
Variable	Coefficient	p-Value
Winter-Year Interaction	-1.222e+06	0.161
Spring-Year Interaction	8.439e+05	0.389
Summer-Year Interaction	5.765e+05	0.474
Fall-Year Interaction	-1.005e+06	0.265

† = Nearly Significant
 * = Significant
 ** = Highly Significant

Claim #3: There was statistically significant evidence to support that, all else equal, winter-released dramas had decreased profits over time.

Support for Claim #3: Using the same multiple regression and data stated in claim #2, we find a statistically significant p-value for winter-year interaction term for drama movies (negative coefficient). This suggests that there is a strong probability of correlation where, all else equal, winter-released dramas have decreased profits over time on average. All other interaction terms showed no statistical significance, although some did show near statistical significance. See table in claim #2.