Exploring the factors behind horror movie revenue

The dataset chosen for this analysis is a collection of horror movie records as old as 1950. The dataset was extracted from The Movie Database via their TMDB's API and uploaded to this LINK. The data was inputted by movie enthusiasts into the TMDB website from all around the world and is collected to have a community-run movie database system. The data contains information like movie title, language, movie duration, user rating, budget and revenue.

The dataset contained a total of 32,540 records for horror movies split by original language. First, records which were missing data on runtime and user rating were excluded. In addition, movies with budget and revenue under 1000 dollars were also excluded. That resulted in a total of 969 remaining records.

Of them, 117 non-English language movies and 852 English language movies were left. Now looking at the distribution of revenue among English and non-English movies below it can be observed that English movies do better in terms of revenue.

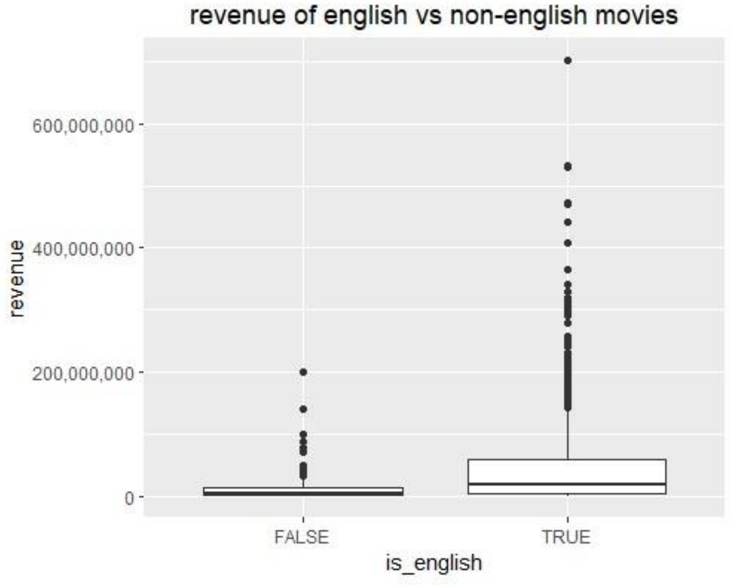

Figure 1

The median revenue for non-English movies is 3,604,598 USD. The median revenue for English movies is 18,501,252 USD.

Due to the data size imbalance, going forward analysis will only be on English language movies. The remainder of this report will systematically examine possible trends in horror movie revenue as influenced by several different variables and conditions.

The following is the distribution of revenue of English language horror movies.

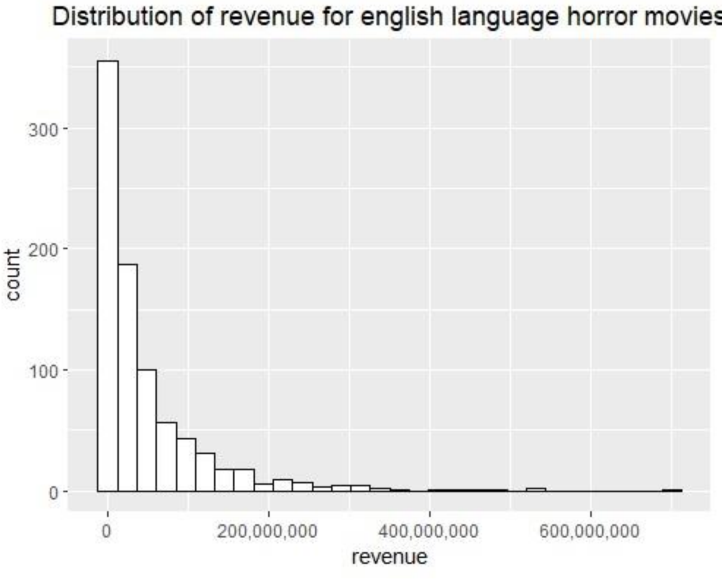

Figure 2

Since the distribution is heavily skewed, the median is calculated as average instead of mean. All other descriptive statistics can be found in the below table.

Number of observations	852
Median Revenue	18,501,253
Maximum Revenue	701,842,551
Minimum Revenue	1,677
Inter Quartile Range	55,221,148

The top 5 highest revenue horror movies in the dataset were:

1. It (2017) 701,842,551 USD
2. World War Z (2013) 531,865,000 USD
3. The Meg (2018) 530,243,742 USD
4. It Chapter Two (2019) 473,122,525 USD
5. Jaws (1975) 470,653,000 USD

Multi-Genre Fusion

Horror movies can contain aspects of several different film genres, for example besides what could be called "pure horror", there are also "horror comedies", "horror thriller mystery", etc. We can refer to this multi-genre production strategy as "genre fusion". In our dataset we can count the number of additional genres each horror movie is categorised in. It might be the case that the bigger the number of genres a movie belongs to, the larger appeal it has to a broader audience and results in more successful movies. Or conversely, viewers might dislike multi-genre horror films because they might be perceived as compromising on the horror aspects or trying to be too many things at once.

To test these hypotheses, plot the number of genres against the average user-submitted rating below.

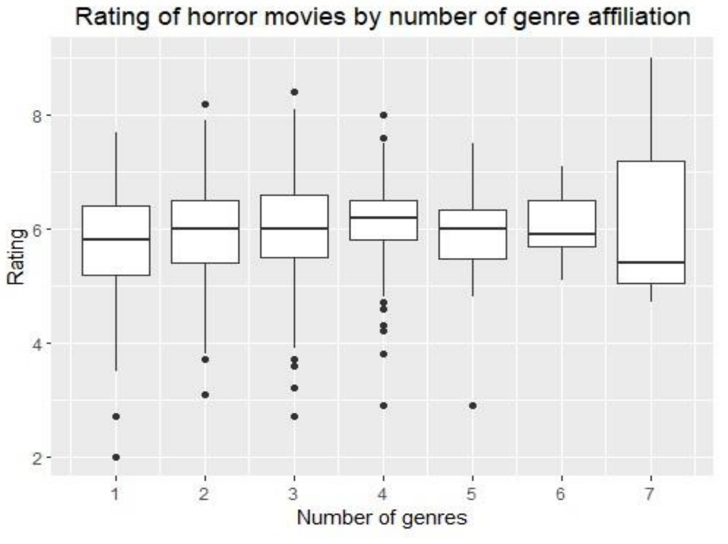

Figure 3

No particular pattern emerges from the plot, as median average ratings for horror movies belonging to multiple genres did not differ substantially from single-genre horror movies. It can be examined if the same trend translates to revenue.

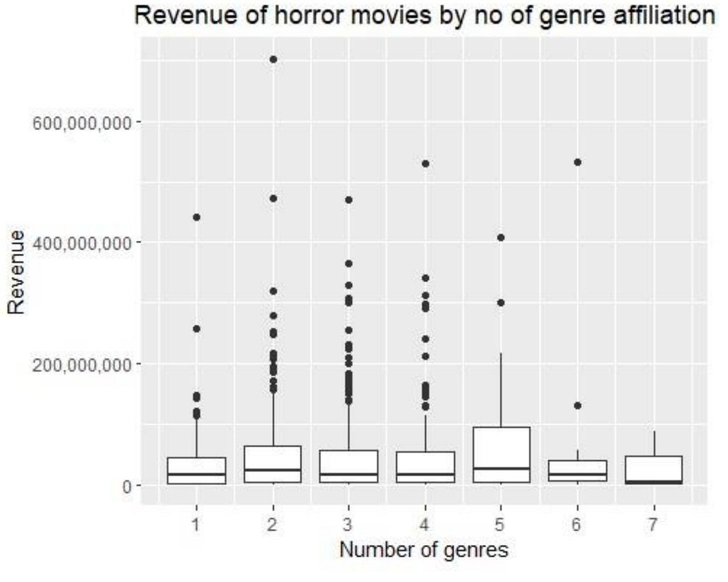

Figure 4

The plot shows the trend does not differ much in this scenario either, which suggests that multi-genre fusion does not affect the performance of horror movies.

Movie Runtime

Perhaps movie runtime could affect revenue. It is possible that runtime of a movie could be a deciding factor for moviegoers, who would want to watch more content for the price of admission. However, in popular culture movies get extended cuts released after their run in the cinemas, with an understanding that audiences prefer shorter movies compared to their longer counterparts.

A scatterplot of runtime against revenue can reveal any possible correlations. Since the distribution of movie revenue is skewed with a longer tail of higher grossing movies (as seen in Figure 2), a transformation of the revenue variable to a logarithmic scale is required before plotting.

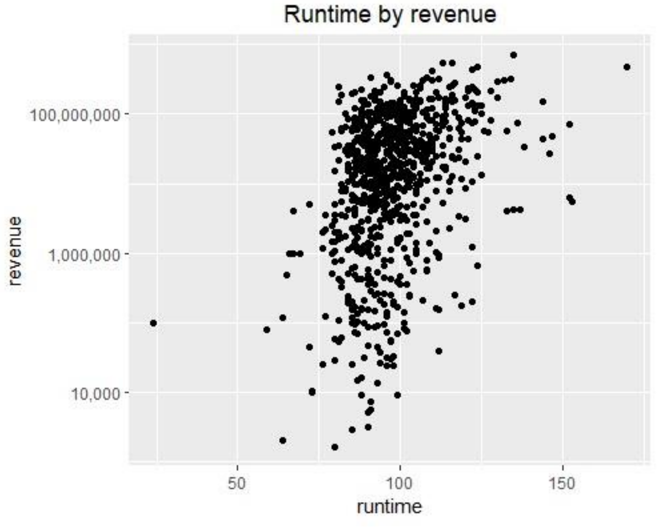

Figure 5

The data reveals a possible positive correlation between runtime and revenue. The longer the movie, the more revenue it can potentially generate.

Month of Release

Seasonal releases are common in the movie industry, especially Halloween during October is important for the horror movie genre. Similarly, Christmas-time or summer could be important as people tend to have vacation time during those periods. A bar plot can show the number of horror movie releases per month.

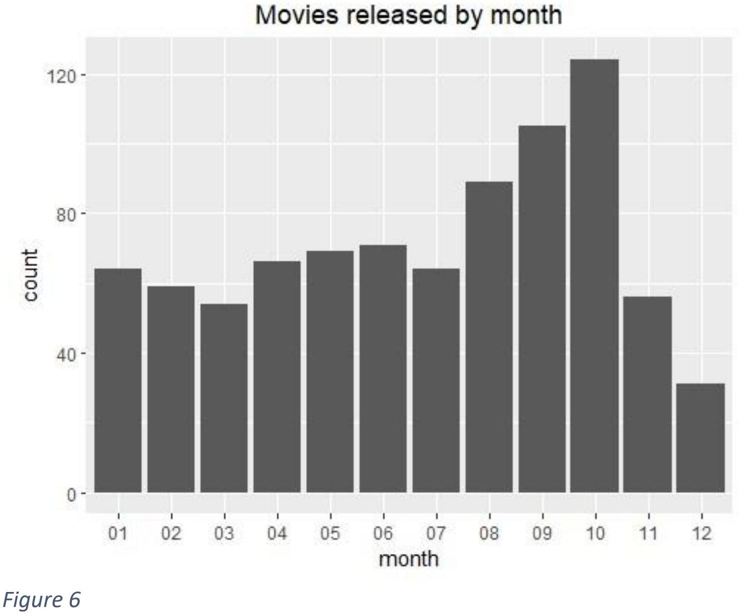

Figure 6

As expected, there are a larger number of horror releases during the month of October and the months leading up to it. On the other hand, December records the least number of horror movies released in a month, which could be because of a likely focus on family movies during that time.

From this information a reasonable hypothesis might propose that revenue for horror movies would be highest in October (Halloween).

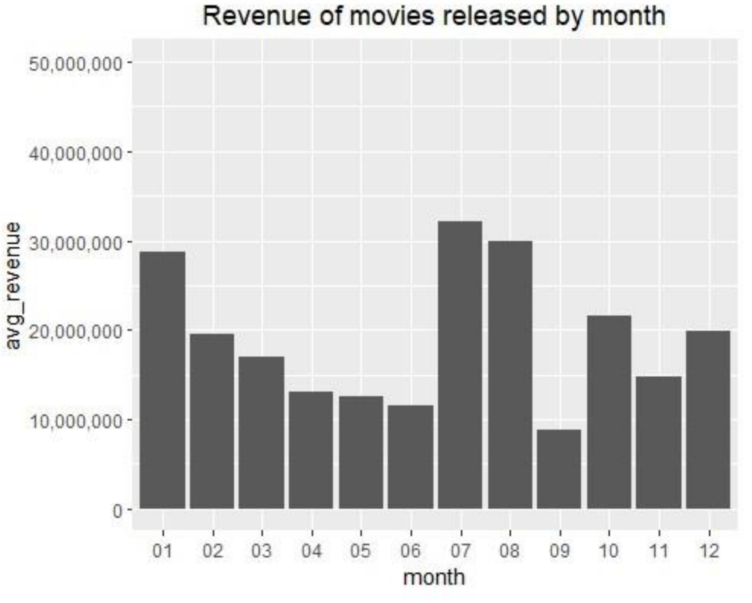

Figure 7

The plot for revenue by month shows that it is not the case, but that huge revenue is generated during summer i.e., months of July and August.

Movie Rating

Movies might perform better when the audiences agree that it is worth paying a ticket for admission. The audience's opinion is captured by the vote averages. A scatterplot reveals a possible positive correlation.

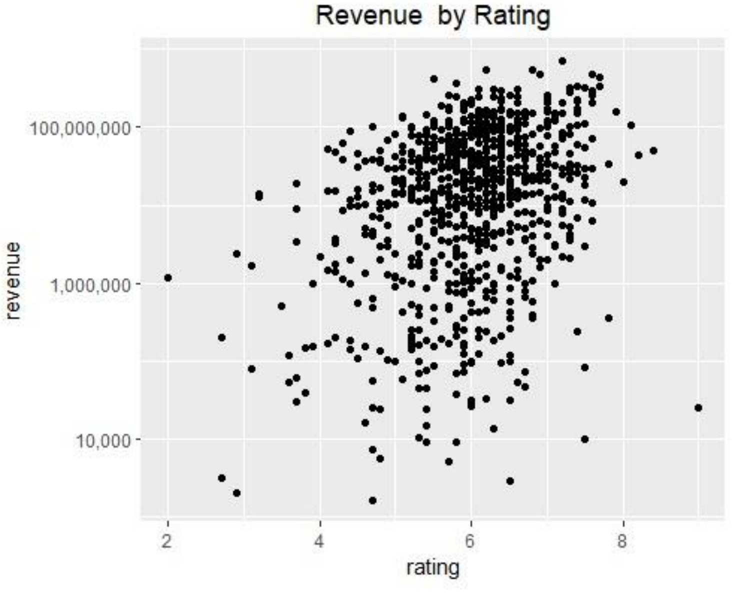

Figure 8

Conclusion

In conclusion, the findings of this exploratory analysis showed that:

- multi-genre horror movies did not perform better than single-genre (or “pure horror”);
- movie runtime was related to the generated revenue;
- month of release was also important for revenue with summertime containing higher grossing movies;
- audience rating was positively corelated with revenue.

However, it needs to be noted that the dataset had many missing values for revenue and other fields which could distort findings. Furthermore, revenue is a skewed distribution with many outliers – while using medians and occasionally a logarithmic transformation, as used in this report, can overcome this limitation, it still leaves a possibility for unexpected or misleading results (especially in a smaller, unrepresentative sample).

There might be other factors that influence revenue, which are not captured by this dataset, such as advertising or popularity of the cast, crew and studio, among others. In the future, merging this dataset with other sources of information could lead to more meaningful insights.