# Horror Movies - the best Hollywood Investments?

Khalid Said

*November 19, 2023*

## Executive Summary

The dataset we chose for our project was, Horror Movies - the best Hollywood Investments? By Sujay Kapadnis, LINK. The analysis focuses on the business viability of horror movies in the context of the film industry. Within this dataset, qualitative and quantitative variables such as the movie title, production, domestic_gross, worldwide_gross, distributor, mpaa_rating, and genre can be observed. The objective is to understand the return-on-investment potential for horror movies and how it compares to big-budget summer blockbusters and action/adventure films. The findings aim to provide insights into the financial incentives for studios to continue producing horror movies despite their relatively smaller share of the total box office. The primary data analyzed includes the box office performance and production budgets of horror movies, particularly focusing on their return-on-investment metrics. Specifically, an example, such as "Paranormal Activity," is considered to illustrate extreme cases as it was made for $450,000 and pulled in $194 million which is 431 times the original budget. Comparative data for other genres, especially big-budget productions is also examined to provide context. The analysis reveals that while horror movies contribute only 3.7% to the box office, their return on investment potential is exceptional. Notable examples like "Paranormal Activity" demonstrate that horror films can achieve returns hundreds of times their production budgets. This contrasts with the performance of big-budget summer blockbuster films. Some key findings are that horror movies, on average, demonstrate a significantly higher return on investment compared to other genres, making them financially attractive for studios. Despite their lower box office share, the profitability of horror films suggests a strategic advantage for studios to diversify their film portfolios. The success of low-budget horror films, such as "Paranormal Activity," underscores the importance of cost-effective production strategies. Some recommendations are encouraging the production of horror movies to capitalize on their lucrative return-on-investment potential. It could also be considered that allocating resources judiciously and emphasizing cost-effective production strategies have proven successful in the horror genre. Some limitations are the market trends like external factors that may influence trends such as cultural shifts or global events that are not considered in this data analysis. For avenues of future pursuit, the analysis could be extended to examine trends in the horror genre's return on investment over different periods. Also, investigate external factors and market dynamics that may impact the performance of horror movies. This analysis will equip senior management with insights to make informed decisions regarding the studio's film production strategy.

**Five Measures of Location**

1) Midrange: $163,741,085.5

The midrange, calculated at $163,741,085.5, represents the midpoint between the minimum and maximum values of horror movie domestic gross. This measure provides a simple and intuitive insight into the central position of the data. The advantage of the midrange lies in its simplicity and ease of understanding. However, it is sensitive to extreme values, as they can disproportionately influence its value. Considering its straightforward interpretation, the midrange can serve as a quick reference point for senior management to understand the approximate central position of horror movie earnings.

2) Median:$25,321,609

The median, at $25,321,609, represents the middle value of horror movie domestic gross when the data is ordered. This measure is not influenced by extreme values, making it a robust indicator of central tendency. The advantage of the median is its resilience to outliers, providing a more accurate reflection of the typical horror movie earnings. For senior management seeking a measure less affected by extreme successes or failures, the median is a reliable choice.

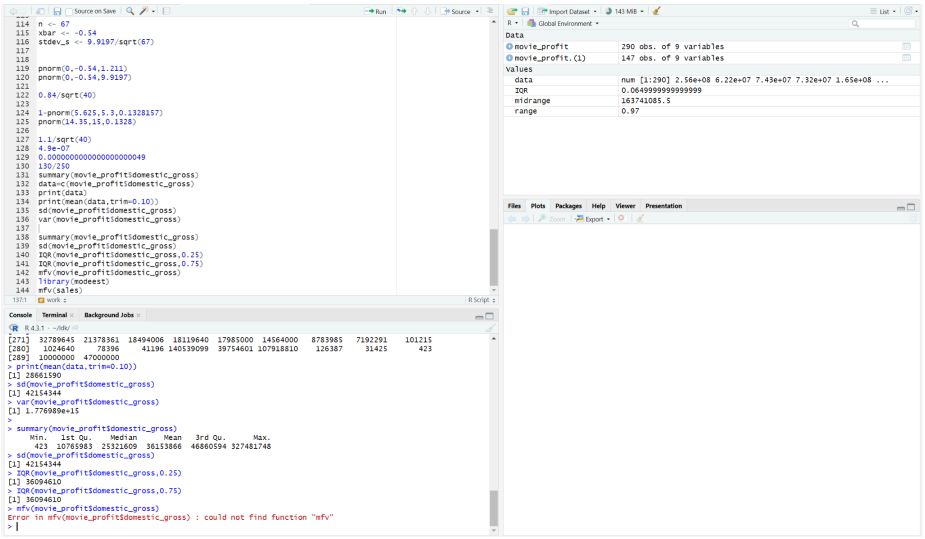
3) Mean: $36,153,866

The mean, calculated at $36,153,866, is the average horror movie domestic gross. This measure considers all data points, offering a comprehensive view of the central tendency. The mean is sensitive to extreme values, and in the context of horror movies, where blockbuster hits or significant flops can occur, it may be influenced by outliers. The advantage of the mean is its inclusivity, capturing the overall average performance. Senior management may find the mean useful when seeking a balanced overview of horror movie earnings, but caution is warranted due to its susceptibility to outliers.

4) Minimum: $423
5) Maximum: $327,481,748

The minimum and maximum values, at $423 and $327,481,748, respectively, represent the extremes of horror movie domestic gross. The advantage of these measures lies in their ability to showcase the breadth of financial success within the genre. However, they are highly sensitive to outliers, and extreme values can skew their interpretation. Senior management may find the minimum and maximum valuable for understanding the potential range of outcomes, but they should be used in conjunction with other measures to provide a more nuanced understanding of the data. There was no most frequent value.

In terms of business application, the choice of which measures to prioritize depends on the specific needs of senior management. If simplicity and a quick reference point are desired, the midrange could be valuable. For a robust central tendency indicator resistant to outliers, the median is a reliable choice. The mean offers a comprehensive average but should be approached with caution due to sensitivity to outliers. The minimum and maximum values provide insights into the extremes of horror movie earnings but should be considered alongside other measures to avoid misinterpretation. A balanced approach, considering the strengths and limitations of each measure, would be recommended for a well-informed decision-making process.




**Five Measures of Variability**

Include five measures when ‘weaving the story’. Follow the same instructions as the measures of location.

1) Range: $327,481,325

The range, spanning $327,481,325, illustrates the full extent of horror movie domestic gross values. The range provides a clear picture of the diversity in earnings within the horror genre. However, it is highly sensitive to outliers, and extreme values can disproportionately impact its magnitude. Senior management should be aware that the range may not fully capture the distribution's shape and may be skewed by extreme successes or failures.

2) Interquartile range: $36,094,610

The interquartile range, calculated at $36,094,610, represents the spread of the middle 50% of horror movie domestic gross values. This measure is less influenced by extreme values than the range, offering a more focused view of the central distribution. The IQR is advantageous

for senior management seeking a measure that is less sensitive to outliers, providing a more robust understanding of the typical range of horror movie earnings.

3) Standard Deviation: $42,154,344

The standard deviation, at $42,154,344, quantifies the average deviation of horror movie domestic gross values from the mean. This measure considers the entire distribution and provides insight into the dispersion of values. However, like the range, the standard deviation is sensitive to extreme values, potentially skewing its interpretation. Senior management should interpret the standard deviation cautiously, considering the genre's potential for blockbuster hits and significant flops.

4) Coefficient Variation: 116.60

The coefficient of variation, calculated at 116.60, expresses the standard deviation as a percentage of the mean, providing a relative measure of variability. The advantage of the CV is its ability to compare variability relative to the mean, making it useful for comparing the risk associated with different genres. However, caution is needed, as the CV can be sensitive to extreme values in the mean or standard deviation. Senior management may find the CV valuable for assessing the relative risk of horror movies compared to other genres.

5) Variance: 1.776989e+15

The variance, at 1.776989e+15, quantifies the average squared deviation of horror movie domestic gross values from the mean. While the variance provides a comprehensive measure of dispersion, it suffers from the same sensitivity to extreme values as the standard deviation. Senior management should approach the variance with caution, recognizing its potential for distortion due to outliers.

In terms of business application, these measures collectively offer a nuanced understanding of the variability in horror movie domestic gross. The range showcases the full spectrum of earnings, while the IQR provides a more focused view of the central distribution. The standard deviation and variance offer insights into the overall dispersion, with the coefficient of variation providing a relative comparison. Senior management should carefully consider these measures, considering the genre's propensity for extreme successes or failures, to make well-informed decisions about financial risk and opportunity within the horror movie market.

**Interpret Standard Deviation and Coefficient of Variation**

**Standard deviation:**

The standard deviation value of $42,154,344, serves as a key measure of variability in horror movie domestic gross. In the context of this dataset, the standard deviation quantifies the average deviation of individual movie earnings from the mean domestic gross. A higher standard deviation indicates greater variability in the data, which implies a wider spread of financial performance among horror movies. For senior management, the standard deviation provides insight into the dispersion of earnings within the horror genre. A higher standard deviation is interpreted as horror movies exhibiting a diverse range of financial outcomes, with some movies achieving exceptionally high box office success while others may experience significantly poor performance.

In terms of business applications, the standard deviation can be used to assess the level of risk associated with investing in horror movies. A higher standard deviation indicates a greater degree of uncertainty in predicting the financial success of each film within the genre. Studios and investors in the film industry should approach horror movie projects by being aware of this variability, adopting risk mitigation strategies, and diversifying their film portfolios to account for potential outliers.

**Coefficient of variation (CV):**

The coefficient of variation (CV) was 116.60. This value of CV provides a relative measure of variability, allowing for the comparison of risk across different movie genres or datasets. The higher the CV, the greater the level of dispersion around the mean in the variable. In other words, a higher CV value would indicate a proportionally larger standard deviation relative to the mean, signifying a higher level of relative variability in the data.

In terms of business applications, the CV is a valuable tool for senior management to assess how much volatility is assumed in comparison to the amount of return expected from investing in films. The lower the ratio of standard deviation to mean return, the better the risk-return tradeoff. Similarly, a CV of 116.60 means that, on average, the standard deviation is 116.60% of the mean domestic gross for horror movies. This emphasizes the significant variability within the genre, underlining the importance of cautious financial planning and risk management strategies. In conclusion, the high coefficient of variation highlights the need for careful consideration of the financial variability in the horror genre.

**Z-Score**

R code:

```r
# Function to calculate z-score
calculate_z_score <- function(data_point, mean, standard_deviation)
  {z_score <- (data_point - mean) / standard_deviation
  return(z_score)}

# Example usage
data_point <- 327481748
mean_value <- 36153866
standard_deviation_value <- 42154344

z_score_result <- calculate_z_score(data_point, mean_value, standard_deviation_value)

# Print the result
cat("The z-score for the given data point is:", z_score_result, "\n")
```

Zscore = 6.910981

The calculated Z-score of 6.910981 is a statistical measure that indicates how many standard deviations a particular data point (in this case, horror movie domestic gross) is from the mean of the dataset. A Z-score of 6.910981 is significantly high, suggesting that the data point is quite far from the mean. In the context of the business application and the project's focus on the financial performance of horror movies, the Z-score provides insights into the specific movie's domestic gross within the horror genre. A positive Z-score indicates that the movie's performance is exceptionally high compared to the average horror movie. For senior management and stakeholders in the film industry, a Z-score of 6.910981 implies that the particular horror movie associated with this Z-score has achieved a high level of financial success. This extreme Z-score suggests that the movie's box office earnings significantly deviate from the typical financial performance of horror movies.

The high Z-score highlights a potential outlier in the dataset. Senior management should investigate this specific horror movie to understand the factors contributing to its financial success. Identifying the characteristics of such outliers can inform future investment and production strategies. While a high Z-score signifies extraordinary success, it also implies a level of risk. Studios and investors should consider the factors that led to this movie's success and evaluate whether similar conditions can be replicated in future projects. Understanding the financial performance represented by the high Z-score can influence strategic decision-making. It may encourage studios to explore similar themes, production strategies, or marketing approaches that contributed to the success of this outlier. While high returns are possible, the horror genre's inherent variability means that not all movies will perform equally well. Diversification helps spread risk and ensures a more balanced financial outcome across a range of projects. In conclusion, the Z-score of 6.910981 provides valuable information for senior management to understand the exceptional financial success of a specific horror movie within the dataset. However, careful consideration and analysis are necessary to translate this statistical measure into actionable insights for future decision-making in the film industry.

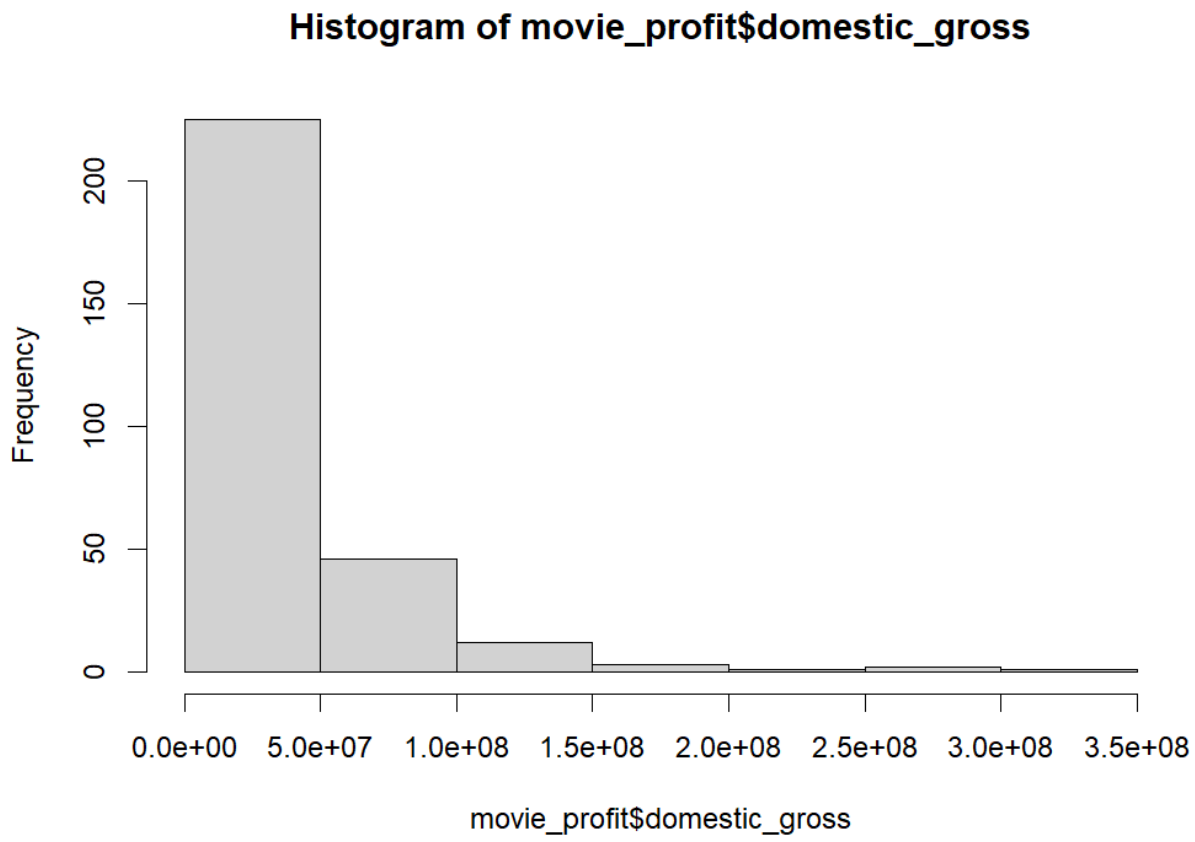
**Identify and Describe Outliers**

There are approximately 14 outliers however, the top three outliers are 327481748, 260000000, and 256393010. The presence of outliers in the horror movie dataset, as evidenced by the top three values of $327,481,748, $260,000,000, and $256,393,010, can be explained through various factors inherent to the film industry. Outliers often represent cases that deviate significantly from the typical financial performance of horror movies, and understanding their occurrence is crucial for informed decision-making.

One explanation for the outliers is the influence of substantial production budgets and extensive marketing campaigns. These outliers, such as the top three values mentioned, likely correspond to high-profile horror films that enjoyed massive success at the box office. Such films may have benefitted from large-scale promotional efforts, star casts, or groundbreaking special effects, which contributed to their earnings. The data point of $327,481,748, for instance, may be associated with a horror movie that became a cultural phenomenon, attracting a broad audience beyond the genre's typical fan base.

Moreover, outliers could be attributed to unique and rare circumstances, such as the release of highly anticipated sequels or adaptations of popular horror franchises. These events may generate significant buzz and fan interest, resulting in box-office success. The outlier at $260,000,000, for instance, might be linked to a horror movie that successfully tapped into a well-established and beloved horror franchise, driving enthusiastic attendance and financial returns.

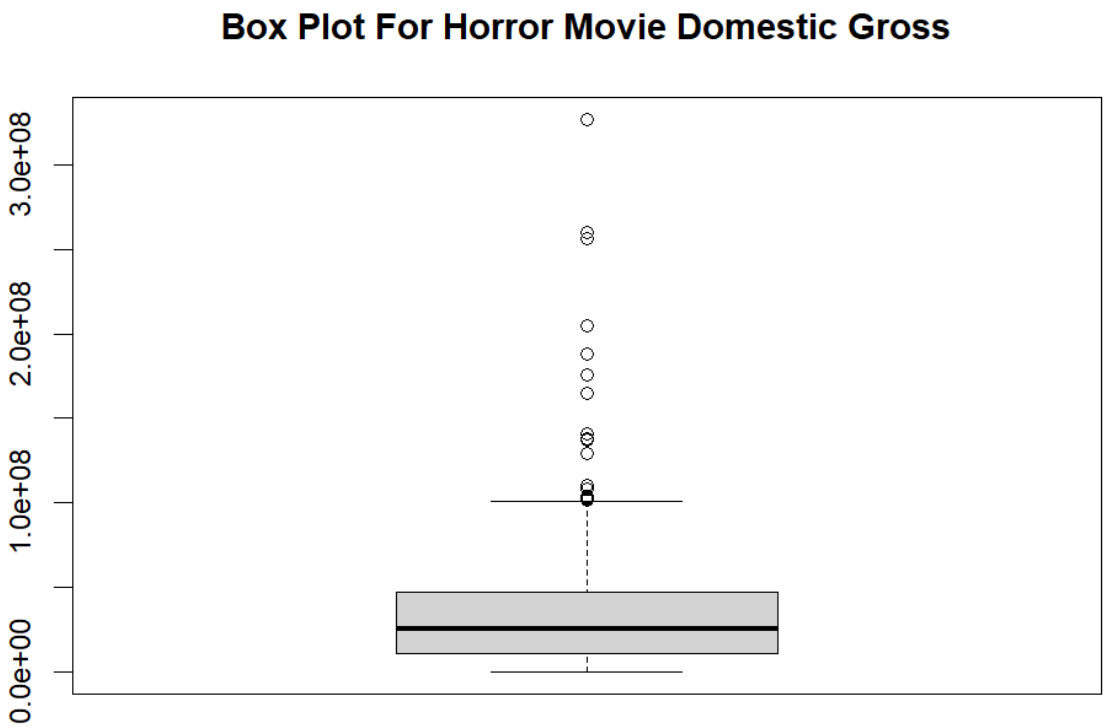
In conclusion, the outliers in the horror movie dataset, identified through their high z-scores, likely represent instances where certain films surpassed traditional expectations due to factors like extensive marketing, production values, or the leverage of established franchises. Understanding these outliers is crucial for studios and investors, as it provides insights into the potential financial success achievable within the horror genre under specific circumstances.




**Histogram/ Graphical Analysis/ Graph Enhancements: Text Boxes**

The histograms reveal the total domestic gross profit for horror movies for by 12/12/2023. The shape of the data set is skewed to the right. The first plots show the trends of the majority of horror movies produced in Hollywood, with over 200 movies grossing up to 50 million dollars. The second one has just under 50 movies between 50 million and 100 million. Looking at the data we can deduce that the majority of the movies analyzed fit the first range, and most often the movies analyzed will not pass 50 million dollars grossed. This data set is critical for senior management to view because it provides the success and failure rates for horror movies and it allows them to see how often this movie will fall into the higher ranges. They can compare this with other histograms for different genres to decide what is more reliable in producing the highest gross profit possible. The histogram represents the total number of domestic gross profits. It skewed to the right.

**Boxplot**




The boxplot can show the five summaries of the domestic gross profit for horror movies. There are over 14 outliers in the box plot. The vast majority of movies are expressed in the plot and the median is($25,321,609). The outliers are all able to surpass 100 million dollars and more. This information is important for the executives to know because the data expresses important data distributions allowing them to know what to expect for domestic gross profit. The boxplot being able to identify the outliers is important because it shows how many movies can rise above the median and become great successes. This data will be able to help them with important decision-making.

**Shape of the Data**

Both of the graphs for horror movies suggest that there is a distribution that is skewed to the right. One of the indicators that this is true is because the mean($36,153,866 ) is greater than the median($25,321,609). After viewing the graph for the histogram, the tail of the histogram curves down to the right. The median is also closer to the lower part of the box so it indicates that it is skewed to the right.