Report: SussexBudgetProductions

**Introduction**

SussexBudgetProductions is facing financial challenges after the commercial failure of its last film,a comedy-action-thriller with a budget of £500,000,which only grossed £100,000.To revitalize the company’s finances,senior leadership,led by CEO Beff Jezos,has proposed the production of either a Romance or Horror film.This report explores these genres using IMDb data to provide a data-driven recommendation on which genre could offer a better chance of financial success.The hypothesis guiding this analysis is that Romance and Horror movies with higher IMDb scores are more likely to yield higher gross earnings.

The IMDb dataset used for this analysis includes crucial variables such as the movie budget,gross revenue,IMDb score,and genre.This analysis will focus on data cleaning and exploratory analysis,followed by hypothesis testing to assess genre profitability. Ultimately,We will provide a recommendation on which genre to produce and suggest a suitable director and lead actor.

**Methods**

**Data Wrangling and Cleaning (ref:week2 and week3 slides)**

Data cleaning was conducted to ensure the dataset was in an optimal state for analysis. First,duplicate entries were removed to avoid counting movies multiple times.Missing values were then addressed,as columns with critical financial and rating data,required complete information to ensure accurate genre analysis.The multi-genre entries in the 'genres' column were split into individual genres to assess each genre's performance independently.

**Exploratory Data Analysis (EDA)**

EDA focused on obtaining descriptive statistics and visualizing IMDb score distributions across genres.The IMDb score is potential for financial success.Histograms were created for IMDb scores,budget,and gross revenue,giving insight into the typical performance ranges within the dataset.Additionally,box plots of IMDb scores across the top five genres were generated to show score distributions,making it possible to identify genres that consistently perform well.

The analysis included:

1. Identifying the top 10 movies by IMDb score to understand top-performing films across genres.
2. Examining the top 5 genres by frequency to understand which genres are most represented in the dataset,their score distributions,and their summary statistics.
3. Conducting a hypothesis test comparing IMDb scores and gross profitability of Romance and Horror genres to determine if one genre is statistically more likely to yield higher scores,which could indicate higher audience satisfaction and, potentially,better box office performance.

**Hypothesis Testing (ref:week4 and week5 slides)**

The hypothesis test assessed whether IMDb scores between Romance and Horror movies differ significantly.The null hypothesis (H0H\_0H0​)states there is no difference in IMDb scores between Romance and Horror movies.The alternative hypothesis (H1H\_1H1​)posits that there is a significant difference between the two genres.A two-sample t-test was performed,to determine if there was a statistically significant difference in IMDb scores that could indicate a genre preference among audiences.

Finally,to provide additional guidance for production,I identified top directors and lead actors in the Romance genre based on the frequency of their appearances in high-grossing films,ensuring these recommendations align with successful Romance productions.

**Results**

**Top 10 Movies by IMDb Score**

The top 10 movies by IMDb score span a range of genres,demonstrating a broad appeal across various types of films.These movies,shown in Table 1,have IMDb scores of 8.8 and higher,indicating high audience ratings. While a high IMDb score does not guarantee financial success,these ratings reflect strong public reception,a favorable indicator for production decisions.

| **Rank** | **Movie Title** | **IMDb Score** | **Genres** |
| --- | --- | --- | --- |
| 1 | The Shawshank Redemption | 9.3 | Drama, Crime |
| 2 | The Godfather | 9.2 | Crime, Drama |
| 3 | The Dark Knight | 9.0 | Crime,Action,Thriller,Drama |
| 4 | The Godfather: part II | 9.0 | Crime,Drama |
| 5 | The Lord of the Rings | 8.9 | Action,Adventure,Drama,Fantasy |
| 6 | Schindler’s List | 8.9 | Biography,Drama,History |
| 7 | Pulp Fiction | 8.9 | Crime,Drama |
| 8 | The Good,the bad and the ugly | 8.9 | Western |
| 9 | Inception | 8.8 | Action,Adventure,Sci-Fi,Thriller |
| 10 | The Lord of the Rings:The fellowship of the R… | 8.8 | Action,Adventure,Drama,Fantasy |
|  |  |  |  |

**Analysis of Top 5 Genres by Frequency**

The most frequent genres in the dataset are **Action,Drama,Comedy,Thriller,**and **Romance**.Analyzing the IMDb score distributions for these genres reveals key insights into each genre's appeal.Romance movies,with a mean IMDb score of approximately 6.3,tend to have slightly higher average scores compared to Action (mean ~6.0) and Comedy (mean ~5.8).Drama,with a mean score of about 6.4,is the top genre in terms of average rating,indicating that audiences tend to appreciate well-made dramas.

* **Action:**Mean IMDb score -6.0,with some outliers in the high-score range.
* **Drama:**Mean IMDb score -6.4,showing consistently strong ratings across movies.
* **Comedy:**Mean IMDb score -5.8,with moderate variability.
* **Thriller:**Mean IMDb score -6.1,with a slightly higher variance.
* **Romance:**Mean IMDb score -6.3,indicating steady audience appeal.

The histogram and box plot visualizations highlight that Romance and Drama scores are less variable,suggesting that they may provide a more stable audience response.Action and Comedy,while popular,tend to have more variable IMDb scores,reflecting a broader range of audience reactions.

**Hypothesis Test Results**

The two-sample t-test comparing IMDb scores for Romance and Horror genres resulted in a p-value of 0.03,indicating a statistically significant difference in scores at the 5% significance level.This result allows us to reject the null hypothesis and accept that a significant difference in IMDb scores exists between Romance and Horror genres,with Romance showing slightly higher scores on average.This supports the hypothesis that producing a Romance film may yield a more favorable reception among audiences.

**Conclusion**

Based on the data,Romance films have consistent IMDb scores that align with positive audience reception.Although both genres offer unique benefits,the analysis suggests that Romance films may provide SussexBudgetProductions with a more reliable pathway to success.The hypothesis test supports the decision to produce a Romance film,as this genre demonstrates both popularity and consistent positive audience feedback,which could lead to higher revenue and recovery of production costs.

For casting,it is recommended to work with experienced professionals in the Romance genre.Our analysis identified Woody Allen as a top director choice and J.K.Simmons as an ideal lead actor.While these recommendations are statistically justified based on IMDb data,it's important to consider limitations.Factors such as market trends, seasonality,and production scale,which were not included in this analysis could also impact box office outcomes.

**Actionable Recommendation:** SussexBudgetProductions should proceed with producing a Romance movie,featuring Director Woody Allen and Actor J.K.Simmons,to optimize chances of box office success based on historical IMDb trends.The consistent audience appreciation for Romance films,supported by statistical evidence,suggests that this genre is a viable choice for the company’s financial recovery strategy.