Report: SussexBudgetProductions

Introduction

SussexBudgetProductions is facing financial challenges after the commercial failure of its last film,a comedy-action-thriller with a budget of £ £500K, which only grossed £100K.To uplift the company's finances,senior leadership,led by CEO Beff Jezos,has proposed the production of either a Horror or Romance film. The IMDb data can predict which genre could offer a better chance of financial success. The hypothesis shows that Romance and Horror movies with higher IMDb scores are more likely to gain higher gross earnings.

The IMDb dataset used for this analysis includes columns such as the movie budget, genre,gross revenue and IMDb score. Here I focused on data cleaning and exploratory analysis,followed by hypothesis testing to assess genre gain.Also according to genre, I'm suggesting a suitable lead actor and director.

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. Firstly,removed all the null values, as well as the duplicate records to avoid redundancy. Rearranged the columns based on local grouping and non-required columns were deleted.To simplify the analysis calculated the Budget and Gross column in millions and added new column(Profit) for the same.

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. Hypothesis test performed by 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 favourable indicator for production decisions.

Movie Title	IMDB Score	Budget (in million)
The Godfather: Part II	9.0	13.0
The Good, the Bad and the Ugly	8.9	1.2
Fight Club	8.8	63.0
Goodfellas	8.7	25.0
Seven Samurai	8.7	2.0
City of God	8.7	3.3
Modern Times	8.6	1.5

Movie Title	IMDB Score	Budget (in million)
American History X	8.6	7.5
Spirited Away	8.6	19.0
The Usual Suspects	8.6	6.0

Analysis of Top 5 Genres by Frequency

The most frequent genres in the dataset are **Crime,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 6.5,tend to have slightly higher average scores compared to Thriller (mean ~6.36) and Comedy (mean ~6.23).Drama,with a mean score of about 6.76,is the top genre in terms of average rating,indicating that audiences appreciates dramas the most.

- **Crime:**Mean IMDb score -6.54,with some outliers in the high-score range.
- **Drama:**Mean IMDb score -6.76,showing consistently strong ratings across movies.
- **Comedy:**Mean IMDb score -6.23,with moderate variability.
- **Thriller:**Mean IMDb score -6.36,with a slightly higher variance.
- **Romance:**Mean IMDb score -6.46,indicating steady audience appeal.

The histogram and box plot visualizations highlight that Crime, Drama and Romance scores are less variable,suggesting that they may provide a more stable audience response.Thriller 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 0.38 based on profit and 2.18 based on IMDb score.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 more desired output to the SussexBudgetProductions. The hypothesis test supports the decision to produce a Romance film,as this genre demonstrates both popularity and consistent positive audience feedback with low standard deviation,which could lead to higher revenue and recovery of production costs.

Our analysis identified Mike McCoy as a top director choice and J.K.Simmons as an ideal lead actor.While these recommendations are statistically justified based on IMDb data and profit,it's important to consider limitations such as market trends, seasonality,and production scale,which were not included in this analysis could also impact box office outcomes.

Desired Output: SussexBudgetProductions should proceed with producing a Romance movie,featuring Director Mike McCoy 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.