Project Proposal for Capstone 3

Production of movies is a combination of multiple factors to create a successful, high revenue, movie. This project is to determine the effect of budget and summer releases on the overall revenue. Do either of these factors show a direct impact on the success of movies? More specifically, does a higher budget result in higher success? Do summer movies have an impact on revenue? These are concerns for both the director (budget) and the production companies.

The data set used is “Movie Dataset: Budgets, Genres, Insights” provided by Kaggle.com. Data provided includes budget, genre (multiple types per movie), date released, movie name, cast, director, and movie details such as original language, overview, tagline, runtime, and much more. Most data types are string, with float and integers.

There are a total of 4802 movies listed with release dates from 1916 to 2017. For the purposes of this project, only movies with a release date of 2011 to 2017 will be used and whose budget and revenue value not $0, 785 total movies. The first comparison will comprise of budget and revenue. The budget is in US dollars, mostly in thousands. Revenue is also US dollars to the dollar. The second comparison will be the month of release and revenue. The [meteorological seasons](https://www.calendarr.com/united-states/seasons-of-the-year-in-the-united-states/) state summer between June and August.

Hypothesis 1

Ho: There is no positive, direct variation between budget and revenue.

Ha: There is a positive, direct variation between budget and revenue.

A scatter plot will be used to visually determine if a correlation exists. Then the coefficient of determination, r squared, will be calculated to specify the strength and direction of correlation. This data is useful for a director to properly budget and not assume “blank check” when filming.

Hypothesis 2

Ho: Movies released in the summer months have no impact on revenue.

Ha: Movies released in the summer months have an impact on revenue.

A new column will be created called Summer Movie. This column will be a boolean, yes/no, to create two groups to conduct a Mann-Whitney test because the distribution of revenue is skewed and not normally distributed. The result will aid production companies to gauge release dates for maximum profits.

The result of these hypotheses will provide an understanding of how revenue is affected by budgetary constraints and if the movie is released in summer.

Sources:

(n.d.). Movie Dataset: Budgets, Genres, Insights. Kaggle.com. Retrieved July 18, 2023, from <https://www.kaggle.com/datasets/utkarshx27/movies-dataset>

(n.d.). Seasons of the Year in the United States. Calendarr.com. Retrieved July 18, 2023, from <https://www.calendarr.com/united-states/seasons-of-the-year-in-the-united-states/>