W203 Lab 02 Research Proposal - Team 02

Divya Menghani, Hamsini Sankaran, Israel Ayode, Sivakumar Thiyagarajan

The movie industry is a huge multi-billion dollar business, with movies created at various budgets ranging from low to high. However, one of the significant factors determining a movie's success is its revenue. In this proposal, we aim to investigate the relationship between the "Budget" and the "Revenue" of the film. Below are the research question, the X and the Y concept

- *Research Question: How does the budget of a movie influence its revenue?*
- *X Concept: Budget - budget (metric)*
- *Y Concept: Revenue - revenue (metric)*

The variables influencing the X concept "Budget" are the Movie length (runtime), No. Of Cast and Crew(derived from the 'credits' column) and Genre (genres). Likewise, the variables that influence the Y concept "Revenue" is the Movie length (runtime), Vote count (vote_count), Movie title length (no. of characters in movie title), No. Of Cast and Crew (derived from the 'credits' column), Genre (genres), and the Release Date. Movie investors are the key actors influencing the X and Y concepts.

We gathered the dataset from Kaggle. The fields like movie details, keywords, and credits are collected from the TMDB open API. The dataset comprises metadata for more than 700,000 movies listed in the TMDB Dataset. It is relevant and provides an opportunity to explore and analyze the relationship between the budget and the revenue of the movies. Since the dataset is unique on film, each movie is independent of one another and uniquely identified by an id column in the dataset. We operationalize the variables Budget(X) and Revenue(Y) in terms of $ amount.We also need to consider the impact of omitted variables such as the movie's storyline, competition from other movies released on the same date, and the film's rating given by the Motion Picture Association (MPA).

We operationalize the variables Budget(X) and Revenue(Y) in terms of $ amount.We also need to consider the impact of omitted variables such as the movie's storyline, competition from other movies released on the same date, and the film's rating given by the Motion Picture Association (MPA). Each row of data in the dataset is uniquely identified by a movie as indicated below.

- Budget is represented as $ amount
- Revenue is represented as $ amount
- Genre represents movie category
- Movie Length is represented as count
- Vote count is represented as count
- Movie title length is represented as count
- No. of Cast and Crew is represented as count

1

Budget vs Revenue Scatter Plot