A.B. Ram Films Data Analysis Project Brief

Motivation

Movie companies want their audiences keep revisiting the theatres or pay online to watch the films they produce. This fictitious movie company named A.B. Ram Films wants to do the same and to maximize profit. A.B. Ram Films needs to determine what contributes to higher audience engagement.

Objective

Develop a launch strategy for this movie company, specifically where to produce films, what genres to film, and what actors to use.

Scope

A.B. Ram Films is a mobile company that tends to travel worldwide producing films. Overall, it earns a profit of $50 million on average from releasing films prior. To keep profit where it's at or even higher, the company is willing to spend at most $50 million for each film.

Questions

1. Find correlations between gross earnings, budget, votes, IMDb score, and runtime. Which movie factors are highly correlated with gross revenue?
2. Discover which region(s) earned the most profit releasing films on average.
3. Identify what types of films earn the most profit?
4. What types of film bring in the most gross? Actor?
5. How has the United States' profit of films varied over time? Has the US experienced years where they averaged a net loss?

Data Collection

The dataset "**Movie Industry: 4 Decades**" is an external source, which I found on Kaggle. The dataset is collected via web scraping from IMDb using Python.

movies.csv

Context

Is the movie industry dying? Is Netflix the new entertainment king? Those were the first questions that lead to the creation a dataset that focused on movie revenue and analyze it over the last 4 decades. There are factors that could've contribute to movie revenue, such as actors, genres, user ratings and more.

https://www.kaggle.com/danielgrijalvas/movies

Content

Movies and their quality values are present in movies.csv file (Original Data). This site contains 4 decades of films from 1980 to 2020. There are a total of 7668 films.

For the original dataset, each movie has the following attributes:

- budget: the budget of a movie. Some movies don't have this, so it appears as 0
- company: the production company
- country: country of origin
- director: the director
- genre: main genre of the movie.
- gross: revenue of the movie
- name: name of the movie
- rating: rating of the movie (R, PG, etc.)
- released: release date (YYYY-MM-DD) and release location of country
- runtime: duration of the movie
- score: IMDb user rating
- votes: number of user votes
- star: main actor/actress
- writer: writer of the movie
- year: year of release

Limitations and ethical considerations

The limitations in this dataset are the NaN values and the singular values for the genre, director, writer, and star columns. Movies can have more than 1 genre, director, write and actor. It can be assumed the creator of the dataset wants to simplify the columns by providing a single value that best describes each quality of the film.