Movies Prediction Dataset

Out of the 28 variables, I am especially interested in know how does the IMDB rating score correlate with other variables. From the 3D gross-country-rating plot below, we can see that United States produced the largest amount of movies across the past 100 years (1905-2015). The sheer amount dwarfs other countries in the number of produced movies. The points at the top corner of the plot denote the movies having the highest gross in the movie history. Many countries produced great movies, but still there were quite a few bad movies. But this movie dataset re-defines about the bollywood data.

Movie rating prediction:

The prediction of movie ratings in this article is based on the following assumptions:

- The IMDB score reflects the greatness of movies. The higher, the better.
- Watching good movies is preferable to bad ones for many people.

With those 28 variables available for all scraped movies, can we predict movie rating? Before we begin, it is necessary to investigate the correlation of those variables.

Background

The movie dataset we are working with presents a rich and intriguing source of information. This dataset, sourced from Kaggle.com, encompasses a comprehensive compilation of movie data spanning the years 2000 to 2019. It

encompasses essential details such as budget, gross revenue, genre, directorship, and the cast of each film. This dataset is particularly captivating due to its capacity to address a wide spectrum of inquiries, offering valuable insights into the film industry's evolution over the years.

Motivation

Our project's motivation is driven by a collective fascination with the world of cinema and the desire to delve deeper into the intricacies of movie data. We harbor a genuine interest in gaining a more profound understanding of the dynamics within the film industry. Additionally, we are keen on expanding our knowledge of Tableau, a tool we are all familiar with but have yet to utilize for dashboard creation. We aspire to master the art of crafting interactive Tableau dashboards that provide in-depth insights into the movie domain. Our objective is to build a dashboard that not only engages users but also serves as a comprehensive repository of information related to the world of movies, encompassing facets like gross revenue, genre trends, actor influence, and directorial impact.

Project Objectives

Our project's primary objective is to create an interactive dashboard that provides a comprehensive overview of the movie industry. We aim to achieve this by addressing the following questions:

Q1: How has the gross revenue of the film industry evolved over time concerning genres, actors, and directors?

Q2: How does a film's profitability correlate with its viewer ratings?

Q3: What qualitative factors influence the profitability of a film?

Data

The dataset we are working with is sourced from Kaggle.com. It encompasses a comprehensive compilation of movie data spanning the years 2000 to 2017. It encompasses essential details such as budget, gross revenue, genre, directorship, and the cast of each film. The dataset is available as a series of CSV format files. The dataset is available at the following link: Credit Card Fraud Detection (kaggle.com)

Summary:

I conducted this data analysis project to uncover the key factors influencing movie popularity and to practice using the matplotlib and

Seaborn Python libraries. I chose this dataset because it spans movie genres, budgets, and release years. By applying data visualization techniques and statistical analysis, I tested my hypothesis that a movie's budget has a high correlation with it's gross revenue.