What makes a Movie a Great one?

Teammates: Mengyue Huang, Mallika Pallipuram

Dataset Materials: Movie Scripts, Movie Genres/Tags

Motivation and Audiences:

As movie lovers, we plan to explore what really makes a great movie, no matter in terms of reviews or box office. If the dataset contains enough data collected through the years, we plan to see how the audience’s taste and definition of an intriguing film changes over time.

Audiences, film industry professionals such as producers, directors, or screenwriters might find this project helpful for insights into the current market and gain inspiration on new film ideas after understanding the factors that contribute to a successful movie. The findings from the data can help them make decisions on scriptwriting, genre selection, and marketing strategies. Besides, in terms of Film Studies, the project contributes by providing insights into what makes a movie great and an exploration on how audience preferences and definitions of meaningful films change over time.

By leveraging movie scripts and other data sources, the project aims to apply computational methods to gain insights into the field of film studies, contributing to a data-driven approach to understanding art and culture in our society.

Description:

We will first explore the basic traits of films such as ratings, year, and genres of each movie in the dataset with their review scores and box office, in order to gain some basic expectation and understanding why it is a great film. After that, we would conduct some Language Processing from the scripts and conversations from the movie dataset and try to gain some insights from their shared similarities or differences on topics and plots, obtaining some comprehension on what makes a great movie in terms of the plot.

Most of the results will be represented with plots and graphs, with interpretation markdowns alongside explaining and discussing any discoveries from them.

Goals and Methods:

The project has two goals, one focused on characteristics of movies and one focused on the plots.

1. For our Exploratory Data Analysis (EDA) on the movie dataset, we plan to study the characteristics of each movie such as the box office, year of release, and ratings. We plan to use functions such as .describe and plots such as barplots, scatterplots, and graphs to explore the dataset and how factors such as ratings change over time. This can help us gain a better insight as to what factors contribute to higher ratings.

2. For exploration of the scripts and conversations, potential string slicing is needed. To be more specific, we plan to use Word Frequency Analysis and Text Similarity Analysis to dive deeper into the plots. The frequency analysis examines the most frequently used words or phrases to identify common elements in successful movie scripts, and the similarity analysis measures the similarity between different movie scripts to find patterns or trends.

3. In order to obtain a whole dataset containing both the basic traits and its scripts of selected movies, potential dataset merging based on indexes such as title and movie code may be needed.