What kind of movies are more likely to get high score ?

Su Hang

30005019

November 30th,2018

Outline

- Idea: Questions and preliminary conclusion
- Data: Datasets I will use and their features
- Visualization Plans

Idea

Question: What types of movies are more likely to get high score?

Sub questions:

- What features do high rating films have (Country/ genres/ director/ investment)?
- Think about this from different age groups(different generations have different taste), different genders(men and women have distinctive preferences).

Preliminary conclusion: For old people, drama or history movies may get high score; for young people, action, adventure or science fiction movies may get high score.

Data

Dataset 1: Top rated English movie of 2010 – 2016 from IMDB

- Source: https://data.world/saipranav/top-rated-english-movies-of-this-decade-from-imdb
- Characteristic: More specific. Focus on English movies and split up the users’ votes according to age groups and genders.
- Variances: Name of the movies with the year of released/ Total average rating/ Total number of votes/ Genres/ Budget / Duration/ Number of votes of different rating/ Number of votes by different age groups and different genders
- Size: 41.43k, 118rows * 55columns
- Issues: Some missing entries in budgets and durations

Data

Dataset 2: IMDB 5000 movie dataset

- Source: https://www.kaggle.com/deepmatrix/imdb-5000-movie-dataset
- Characteristic: Large dataset, more complete features, additive information of dataset 1
- Variances: Director name/ Actors names/ Genres/ Movie title/ Number of voted users/ Language/IMDB score/ Budget/ Country/ Gross/
- Size: 1.43MB, 5043rows * 28columns
- Issues: Almost every column has some missing data

Visualization Plans

- Visualize ratings given by different age group for a genre: Bar charts
- Visualize ratings given by male and female for a genre: Bar charts
- Visualize the countries these movies released in: Map
- Visualize the Proportion of different genres of high rating movies: Pie chart
- Visualize the relationship between budget and scores: Scatter plot