
Top four Action, Animation, Comedy, Drama, & Romantic movies

Movie ratings are often misleading because of a disproportionate number of people rating a movie. Hence, we should calculate the weighted rating for a movie. As we estimate weighted rating, we run into another problem of a recent movies have a smaller number of people rating that movie than the number of people rating a movie which has been released sometime back or earlier. eg, less number of people have seen and rated "Blockers (2018)" than "The dark knight rises (2012)". We solve this problem by considering votes a movie got in a year(v). We use this votes rate in our below formulae to calculate the weighted rating of a movie.

$$\text{Weighted Rating}(R) = \frac{\sum_{k=1}^{10} (R_k * k)}{\sum_{k=1}^{10} R_k}$$

$$\text{Corrected Rating} = \frac{v}{(v + m)} * R + \frac{m}{(v + m)} * C$$

Where:

R = Weighted rating for the movie (Out of 10)

R_k = Number of votes for rating k ($k = 1, 2, 3, 4 \dots 9, 10$)

m = Minimum votes per year required to be considered to be listed in top four (200 our case)

$$v = \frac{\text{Number of votes}}{\text{Year of movie} - 2018} = \frac{\text{Number of votes}}{\text{Age of the movie}}$$

C = The mean weighted rating across the whole set of movies

The above equations are used in Tableau as calculated fields.

The other genres are filtered out since I only wanted to show movies from Action, Animation, Comedy, Drama, & Romance genre.

How the data understanding, cleaning, and design was approached:

- Movies may have same title; hence a new field was created by concatenating movie title with the year it was related [*Movie Title (Year of release)*]. The title also had a problem with trailing “, The” string. Below is how the dimension “Title_Year” was creation.

```
IF ENDSWITH(TRIM([Title]), ", The")
  THEN "The " + REPLACE([Title], ", The", "") + " (" + str([Year]) + ")"
ELSE
  [Title] + " (" + str([Year]) + ")"
END
```

- No legend is added to the graph since the image on Y axis serves the purpose.
- The background is made white to make the title of movies look clean.

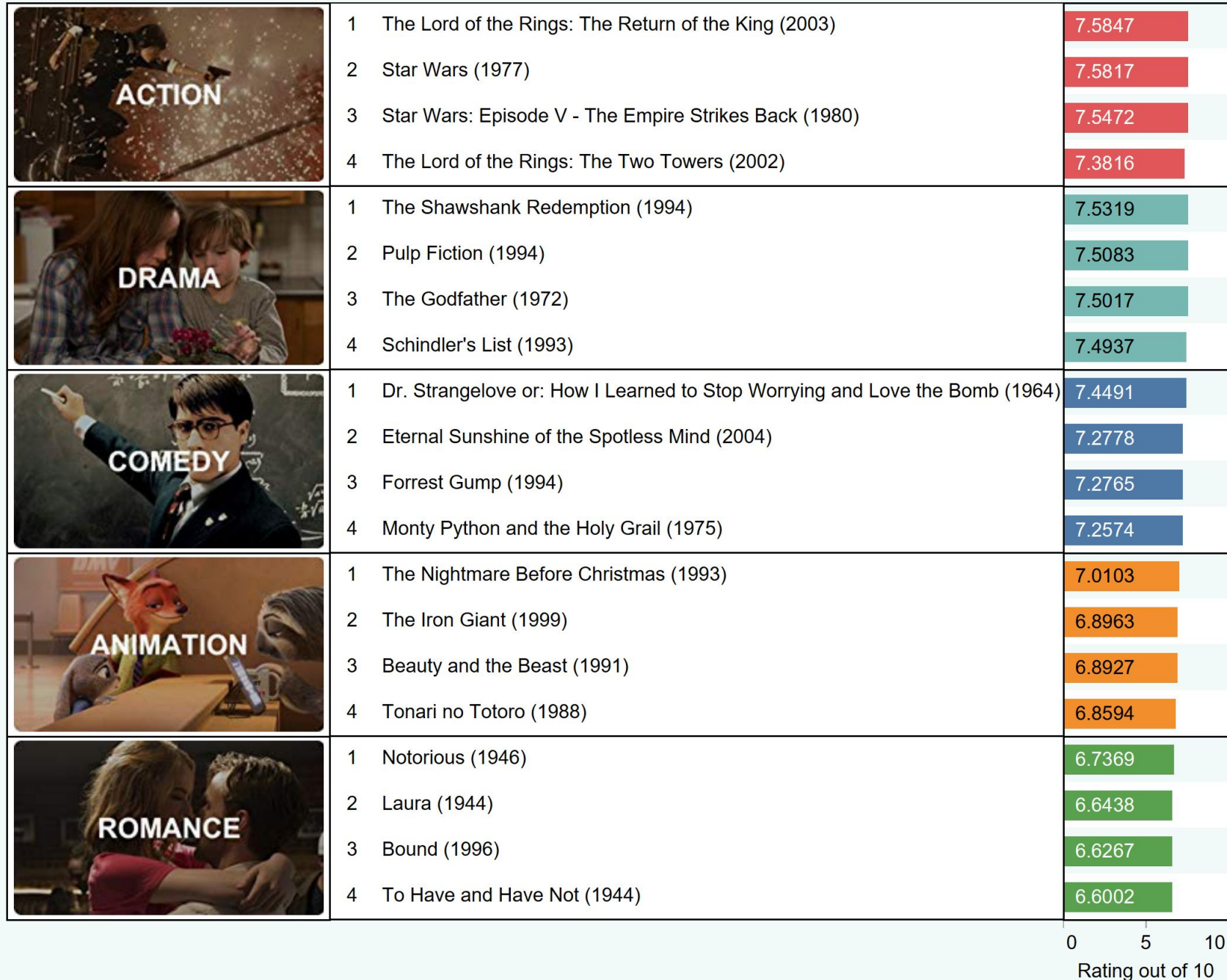
- The horizontal bar chart is used to show the relative ratings of the movies.
- The color of the bar chart is chosen in a way that they are comfortable to the user eyes.
- The Genre column was created from Action, Romance, Animation, Drama, & Comedy columns by creating a calculated field with IF and ELIF in tableau.
- An index was created as calculated field to create ranking for each genre.
- The scale of ratings was selected such that the comparison between each rating is possible.
- The movies were sorted based on their weighted rating. Top 4 movies filter was added to the rank column.
- The theme selected for the workbook is “Modern”
- The marks in tableau were created in such a way that the alignment of the title, bar chart, and weighted rating is consistent.
- Custom shapes were created to show the images that are seen on the left side column of the visualization.

The screenshot displays the Tableau interface with the following configuration:

- Filters:** Votes, Rank, Genre
- Columns:** SUM(Weighted Rating)
- Rows:** Genre, Rank, Title_Year
- Marks:**
 - Card type: Automatic
 - Fields: Genre, Title_Year, SUM(Weighted..)

Next page shows the chart created.

Top Rated Movies In Each Genre



Data source: IMDB Ratings.
Created by Aman Jiddewar.