Project: Visualizing Movie Data

Complete each section. When you are ready, save your file as a PDF document and submit it here.

Step 1: Data Cleanup and Attribute Selection

- Clean up any missing information and choose the most important attributes you will explore further in your visualizations.
- List out the attributes (or variables) you plan to dive further with your visualizations. You should explore no more than 8 attributes.
- Please refer back to the Data Cleanup course to help you clean up your data.

Step 2: Tableau Visualizations

- Please make sure you follow the rubric and include Tableau Dashboards, Stories, and the appropriate visualizations (small multiples, scatter plot, bar chart, etc..) your reviewer expects your visualizations to contain. Remember: You need one Dashboard for every question (Q1-Q4) and in addition, you also need one Story, pertaining to a question of your choosing.
- Attach your visualizations as Tableau Workbooks in a zip file along with this report.

IMPORTANT: Please upload the workbooks to **Tableau Public** to allow reviewers to access your workbooks. Note that simply saving your file as a ".twbx" is not enough to allow all reviewers to access. Instructions on how to do this.

Step 3: Questions

- Answer the following questions. Refer to your online visualizations to back up your answers:
    - **Question 1:** How have movie genres changed over time?
      We can see the number of films increases exponentially starting in 1980. Drama is the most popular genre in 1980 (32 films released), and maintains its position in 2015 (260 films released). However, having the most films released does not necessarily equate to the highest profit. Action, Adventure, Comedy, Drama, and Thriller are the top 5 most profitable genres of the last 10 years. We can see that despite having less films released, Action and Adventure movies have become far more profitable than all other genres, including Drama. Adventure has been the top grossing genre of the past 10 years. In fact, 8 of the top 10 highest grossing movies of all time are Adventure. The only 2 non-Adventure movies in the top 10 were Titanic and Furious 7.

Tableau Public Link:
https://public.tableau.com/profile/jason.grenig#!/vizhome/FilmGenreGrowthandProfitabilitySincethe1960s/GenreStory

- **Question 2:** How do the attributes differ between Universal Pictures and Paramount Pictures?
  I compared revenue (adj.) over and vote avg (rating), between Universal Pictures and Paramount Pictures. These 2 companies are quite similar in both of these metrics, with some minor differences: across all years, Paramount has produced 426 films vs 460 for Universal, even though data for Paramount goes back to 1960, whereas Universal's first movie was produced in 1962. These are both mass market production companies, and their avg votes reflect mainstream moviegoer tastes. The avg vote is 5.97 for all movies, and they stack up at 6.14 (Paramount) vs 6.07 (Universal). Both companies also have a running total of around $54B in revenue across all years.
  Tableau Public Link:
  https://public.tableau.com/profile/jason.grenig#!/vizhome/RevenueAdjustedandFilmRatingComparisonforParamountPicturesandUniversalPictures/PvsU

- **Question 3:** How have movies based on novels performed relative to movies not based on novels?
  Movies based on novels have an avg revenue (adj.) of around $103M, with a total revenue of $13B vs $51M avg revenue (adj.) and $545B total revenue for all other film types.
  Movies based on novels comprise 1.2% of total movie types, however they are almost twice as popular (based on the Popularity field in the dataset), having an avg popularity of 1.29 vs 0.64 for all other films.
  Tableau Public Link:
  https://public.tableau.com/profile/jason.grenig#!/vizhome/RevenueComparisonFilmsBasedonNovelsvsAllOtherFilms_0/bonovel

- What is your additional question that you proposed? What is the answer? How did you come up with this question?
  - **Question 4**: How do the number of votes a film receives correspond with its vote average (rating)? From the scatterplot, we can see the distribution is left-skewed, with an avg vote count of 217.4 and avg vote avg (mean) of 5.97. This makes sense as people generally wouldn't spend time watching movies they really dislike or have no interest in. Further, if we look at the vote counts in 1,000 vote buckets (following the scatterplot lines), we notice that the worst rated movie in each successive bucket has a higher and higher vote avg. For ex: the worst rated movie w/ over 1,000 votes is Fantastic Four w/ a 4.4 rating, the worst rated movie w/ over 2,000 votes is A Good Day to Die Hard w/ a 5.2 rating, and the worst rated movie w/ over 8,000 votes is Avatar at a

7.1 avg vote rating. We can conclude that the more votes a movie has, the less chance it has to have a low vote avg.
I also added a third continuous dimension to show the popularity of each film (relative number of page views on The Movie Database). The average popularity is 0.65, with a max of 32.99. Only 3 films have a popularity of over 15.00: Jurassic World, Interstellar, and Mad Max: Fury Road.
I came up with this question because I wanted to examine the vote relationship with the movies.
Tableau Public Link:
https://public.tableau.com/profile/jason.grenig#!/vizhome/VoteCountbyVoteAveragebyFilm/VoteCountvsVoteAvg2