Unlocking Movie Metrics: A Look at Revenue, Ratings, and More

Data Quality and Cleaning:

- For this project, the MovieDataset data found on Kaggle was used. This dataset included the following variables:
    - Rank: the ranking of the movie based on the popularity
    - Description
    - Director
    - Actors
    - Title
    - Genre
    - Year
    - Runtime(Minutes)
    - Rating: the IMDb user rating of the movie on a scale from 1 to 10
    - Votes: the number of user votes for the movie on IMDb
    - Revenue(Millions): the box office revenue of the movie in millions of dollars
    - Metascore: this represents the aggregated critic scores on a scale of 1 to 100
- The data was pretty organized making the process of cleaning a lot easier. The cleaning process included:
    - Getting rid of rows that did not provide information for the insights that were to be made: rank, descriptions, director, actors
    - Deleting columns with extra spaces such as revenue and metascore which had blanks to make sure there were no null values
    - Removed duplicates so there would not be any mistakes when using the data to create visualizations

Visualization 1 Explanation:

- For the first visualization, a bar graph for revenue(millions) per movie genre was calculated in order to understand which genres were the most popular. To do this, the genres of movies(columns) were compared to the sum revenue(rows) that they made in millions. Furthermore, the average user votes and user ratings from IMDb were taken and added to the details section of the marks field in Tableau. To do this, a parameter allowing the user to select their measure of interest was created. This was in order to see their impact on revenue since if ratings and votes were relatively high, logically, the revenue should be high too**.** The genre with the highest sum revenue of $10,461.51 million was Action, Adventure, and Sci-Fi. The average user votes for this genre were 371,642 votes while the average user rating was a 6.724/10. The genre with the lowest sum revenue was a tie between Adventure and Horror, Mystery, and Sci-Fi. For Horror, the average rating was 5.8/10 as well as 9,247 average user votes. For Adventure, it was 2, 417 votes and a 6.1/10 rating making it a bit better in audience reviews than Horror

was. The revenue for this genre was about $0.15 million, but it should be taken into account that there was only movie in this genre, a product of the dataset as well as the unpopularity of the genre.

Visualization 2 Explanation:

- Based off what was learned in visualization one, a variation of it was created as visualization two to get a few more insights. A bubble chart was created with the same rows and columns as visualization one. In contrast, the movie title field was added to the rows shelf to create a hierarchy. This allowed Tableau to sort movies by revenue within each individual genre rather than sorting by the overall sum revenue. Now including title, the revenue was kept in decimal form in order to make it more detailed in order to understand what the most and least popular movies were, especially in the most and least popular genres. In Action, Adventure, Sci-Fi, *Jurassic World* made the most money with the sum revenue in millions being $652.2 with user votes being 455,169 and the rating being 7/10. The least popular movies in the least popular genres of Horror and Adventure in alphabetical order this time, with *Kicks*, the adventure movie being first, while *The Void*, the horror movie was second. The metrics for these movies are the same as in the previous visualization since there was only one movie in each of these genres. The lack of movies in this genre shows the unpopularity as well as stated previously. The most popular movie overall was *Star Wars: Episode VII - The Force Awakens* from the Action, Adventure, Fantasy genre with a revenue of 936.63 million. The average votes were 661,608 and the average rating was 8.1/10.

Visualization 3 Explanation:

- The third visualization was a scatter plot that was created in order to understand runtime(minutes) of a movie in relation to the box office revenue. To create the graph, the runtime(minutes) field was placed in columns while revenue(millions) was placed in the rows. Furthermore, reference lines were placed on the graph to understand the averages for both metrics. In the marks field, metascore and votes were placed. The size of the circle in the scatter plot depended on the amount of user votes meaning the more votes, the bigger the circle. In addition, metascore was added as the color where the darker the color, the higher the metascore. Both scales of reference for metascore and votes were also provided. The average runtime was 114.6 minutes while the average revenue was 84.6 million dollars. The values at the point at which the two average reference lines cross are metascore of 84, revenue of 82.6, runtime of 115 minutes, and votes of 211,348. That was around where majority of the clustering was occurring on the graph as well.
    - There were instances of higher revenue, higher votes. Around close to 533 million in revenue, the votes were 1,791,916 showing they can at times go hand in hand.

Visualization 4 Explanation:

- The final visualization was created in order to understand how metascore compares to user rating and votes. A heat map was used here ranging from metascores of 0-40, 41-60, 61-80, and 81-100 with 100 being the best. The average user rating ranged from 0-8 while the average votes ranged from 0K to 300K. The size of the squares in the heatmap were determined by the average rating meaning the higher the rating, the bigger the square. In a similar manner, the shade of color of the squares were determined by the average votes so the more votes, the darker the color. Overall, metascore, user rating, and votes were all consistent, with the squares being bigger as well as darker around the metascore range of 81-100. The average rating for this 81-100 range was 7.618 and had 297,192 average votes.

Insights and Conclusions:

1. **Genre in relation to Audience Demands:**
    a. Based on the data, the Action, Adventure, and Sci-Fi genre clearly dominates the rest in the revenue generation. This can suggest that audiences are more inclined to watch as well as invest in high-budget films with lots of spectacle in the plots. This could be incentive for filmmakers to create more action centered movies no matter what they are since there is a proven track record of success.
    b. However, other genres such as Adventure as well as Drama, Family, Music and an abundance of others have lackluster performances. This may be due to limited films in the dataset, but it also provides a niche opportunity for studios to hone in on this and explore fresh, innovate storytelling in these genres. Studios can capitalize on these underrepresented genres with unique narratives or even blending these genres as it could attract new audiences. Overall, exploring underrrepesented genresor maybe even experimenting with cross-genre films can unlock new fanbases and potential.
2. **Popular Franchises vs New Concepts:**
    a. Films such as *Star Wars: Episode VII* and *Jurassic World* both come from very popular established franchises therefore leading to generating significant revenues as seen in the data. This shows that audiences are willing to engage with films tied to something that is well known to them as long as it is not overdone.
    b. Investing in sequels in a safe bet for movie studios but it also suggests that there can be value in new, original concepts. These can break into the market and establish their own fanbases which can be a balance between diversifying the movies their studio puts out while also sticking to the classics.
3. **Impact of Runtime on Audience Engagement:**
    a. Analyzing the data from comparing runtime and revenue suggests that longer films, specifically those with an average runtime of 114.6 minutes can lead to high revenues. This can be due to more immersive experiences for the audience

since there is more time for deeper and more detailed storytelling. This correlation is not exact, so there will be outliers, so filmmakers should avoid unnecessary length for the sake of audiences engagement. The main takeaway is that focus should be on seeing if the runtime can be justified by the narrative and pacing to make sure it is engaging. Each movie is independent of another because for example, action-packed films may benefit from longer runtimes, while other genres such as Comedy may not.

4. **The Impact of Critical Reception:**
    a. As seen in the heatmap, there is a positive relationship between metascores, user ratings, and votes which shows the significant impact of critical reception on audience engagement. Although audiences have their own opinions on movies, this can show that critics due play a role on the general consensus. Films with higher metascores can indicate higher quality in the critics' eyes which can lead to receiving better user reviews and attracting more votes. For future productions, studios should consider prioritizing both critical acclaim and audience satisfaction. Both have their benefits as critical reception leads to increased marketing opportunities and awards while more audicenes means more viewership and revenue. Overall, quality filmmaking in terms of actors, directors, the storytelling can all pay off in the long run.