

Project: Visualizing Movie Data

Step 1: Data Cleanup and Attribute Selection

I used Alteryx to clean up the data. I first selected the attributes that will be needed to answer the questions.

I decided to keep working with the followings :

- id
- popularity
- original_title
- keywords
- genres
- productio_companies
- release_year
- budget_adj
- revenue_adj

I then removed the movies with 0 budget and 0 revenue because I considered that those movies will not have much impact on the study.

After saving the movie_cleaned dataset, I worked on 4 different datasets for the 4 questions and this by selecting the needed attributes for each question.

Questions

[Question 1:](#) How have movie genres changed over time? [Tableau Link](#)

First, I compared the movie release through years by genre. We can clearly see a jump in production after the 90s, before that the movie releases were almost stagnating.

We can also see that Drama is the most released genre through time.

Then, I compared the movie budget through the years. We can notice an opposite relation between the budget and the number of movie released. Drama despite having the highest release has a very low average budget, on the contrary, adventure movies have the highest budget and the lowest release number because they are expensive to produce.

Finally, I compared the popularity of movies genre through the years. We can clearly see that Action and Adventure movies are way more popular than other genres. The jump in popularity in last years may be related to new special effects and technology in blockbusters.

[Question 2:](#) How do the attributes differ between Universal Pictures and Paramount Pictures?
[Tableau Link](#)

We can see that Universal Pictures has a higher movie release than Paramount Pictures during most of the years. When it comes to the average budget, we can see that Paramount Pictures has usually a higher budget spent on movies, although Universal Pictures overtook Paramount Pictures in some years.

For the Revenue by Movie, we can see that except Titanic (Paramount Pictures) which had the highest revenue of all movies by the two production companies, the other top 10 movies have an almost equivalent revenue and this although Paramount Pictures release fewer movies.

[Question 3:](#) How have movies based on novels performed relative to movies not based on novels? [Tableau Link](#)

We can clearly see for the dashboard that the number of novel based movie is far less than their opposites, their budget and popularity, on the other hand, is usually higher. We can notice peaks that can be related to novel based blockbusters like Harry Potter and Lord of the Ring. This type of movies is in a constant rise in popularity.

[Question 4:](#) What is the impact of a movie budget on its popularity and revenue?
[Tableau Link](#)

I wanted first to explore the average revenue of each genre compared to the average budget. We can notice that Animation, Adventure, and Fantasy genre although having a high budget usually have a high revenue as well. That is not the case for Western, War and History movies which are expensive to produce but don't necessarily generate high revenue.

When comparing the popularity of genres to the average budget, we can see that Action movies are spread through the whole sheet. Some Action movies even with the highest budget can score really low popularity.

I came up with this question, so we would have an approximative prediction when deciding on a budget for an upcoming movie. We can see if the the budget has an impact on its popularity and the revenue of the future movie and this by comparing it to similar movies in the same genre and similar budget.