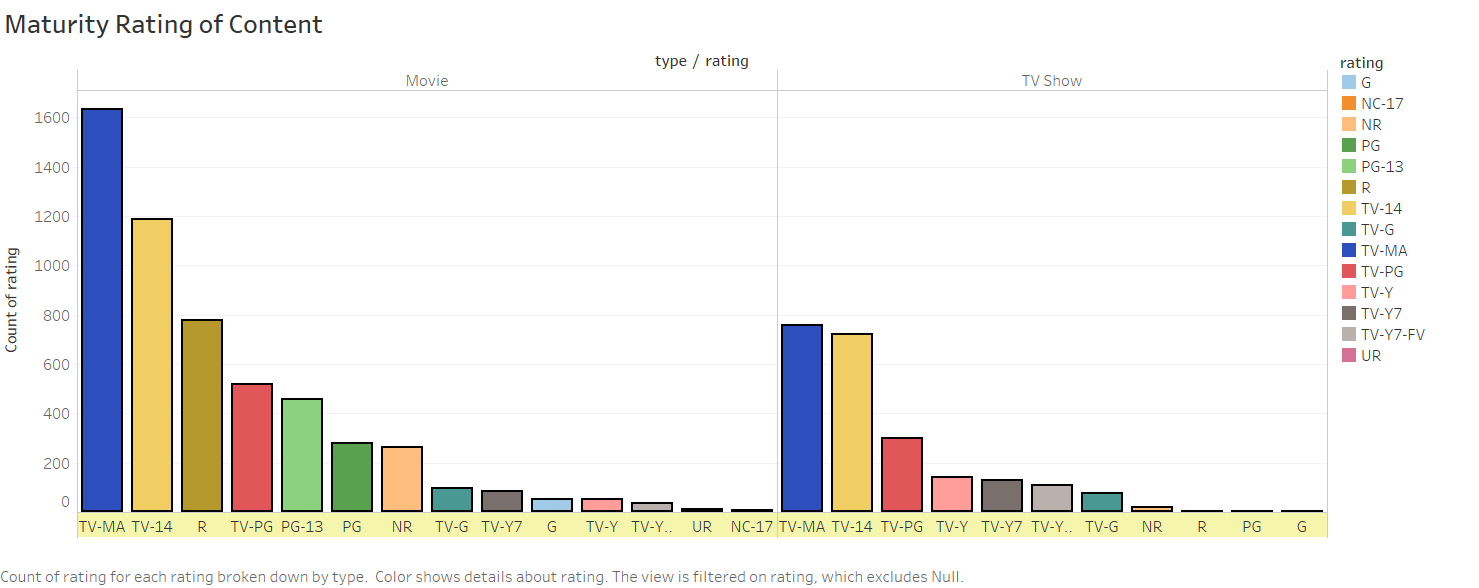
**Task 4**



Analysis

We wanted to find out the distribution of Netflix content in terms of maturity ratings and availability across the movie and Tv Show categories. We designed the visual by taking the count of ratings as Y-axis and type of rating as X-axis. The bar graph clearly shows the type of content that is available and it’s quantity. Separation on type further expresses the presence of different categories of content. The separation on type was important because it is evident that both categories have dissimilar availabilities of mature content.

Visualization

We had to determine the count of the maturity rating rows and divide them according the different ratings. Bar graph was the best suited visual for this task, because it clearly shows the frequency distribution, relative proportions between multiple categories and can be arranged in ascending order.

Implementation

For task 4, we had to summarise the type of content available on Netflix according to their maturity rating. The visualization displays the different maturity rating content in a bar graph. The graph was created on Tableau. The graph uses the following columns – Rating, Count(Rating) and Type (indicates whether Movie or Tv Show). We also filtered out null values – Rows where rating wasn’t present.

Evaluation:

For this set of tasks and visualization we used Think Aloud evaluation technique. This direct observation method was performed on two users which involved asking them to think out loud as they were performing a task. The user actions on this visualization pertaining to what they are looking at, what they are thinking, doing and feeling were noted down. This helped us in determining the user expectations and in identifying the aspects of the visualization that were unclear or perplexing.

The Bar graph was showcased to 2 participants addressing the following tasks:

* Task 1: Tell us what this visualization represents
* Task 2: Looking at the graph, identify the content rating which is the most available.
* Task 3 : Spot the difference, if any, between the 2 types of content available.

For all the tasks as listed above, the thoughts and actions of the participants were noted down. The single ease question score for these tasks was asked to every participant and noted down. This included just one question on - How difficult or easy did participants find the task? The range of the score was set from 1 – 5, with 1 being the toughest and 5 being the easiest. Below mentioned is the observation for the same.

|  |  |  |
| --- | --- | --- |
| **Single Ease Question** | **Participant 1** | **Participant 2** |
| Task 1 | 5 | 5 |
| Task 2 | 5 | 5 |
| Task 3 | 5 | 4 |

Summary of results

Maturity rating visualization indicates that most of the content – whether it’s movies or TV shows, belongs to Mature category. TV-14 rating is assigned to media which needs parental guidance for viewers below 14 years of age and this type of content is 2nd most common (across both Movies and Tv shows). It is also evident that there are far higher R rated movies than R rated Tv shows. Therefore, it can be said that majority of the content available on Netflix is not suitable for children.

Among the 2 categories of content – Movies and Tv Shows, it appears that there is more rating data for Movies. After analysing the dataset, we were able to conclude that the dataset indeed has more movies and fewer TV shows, thereby validating our conclusion.

Discussion

Netlfix offers to option to set profiles that block mature content for children. According to our analysis, vast majority of content on Netflix is not suitable for children. We believe that Netflix realised the fact that they have mature content and therefore decided to add this setting. The internet isn’t governed by the same censorship laws as broadcasted Television. This makes publishing R rated content easier and explains the presence of R rated movies on Netflix.

During the genre analysis, we learned that in general, “Drama” content is widely available on Netflix. However, after joining IMDB movie dataset we realised that “Action” movies tend to bring in more money for the production companies.

## Evaluation Technique for Appendix:

|  |  |
| --- | --- |
| Participant ID | 1 |
| Task ID | 1 |
| Overall Result | Success |
| Process | User looked at the bar graph  User has identified the intent of the graph. |
| SEQ Score | 5 |
| Errors |  |
| Other Notes |  |
| Participant ID | 1 |
| Task ID | 2 |
| Overall Result | Success |
| Process | User looked at the graph.  User was able to easily identify the most available maturity rated content. |
| SEQ Score | 5 |
| Errors |  |
| Other Notes |  |

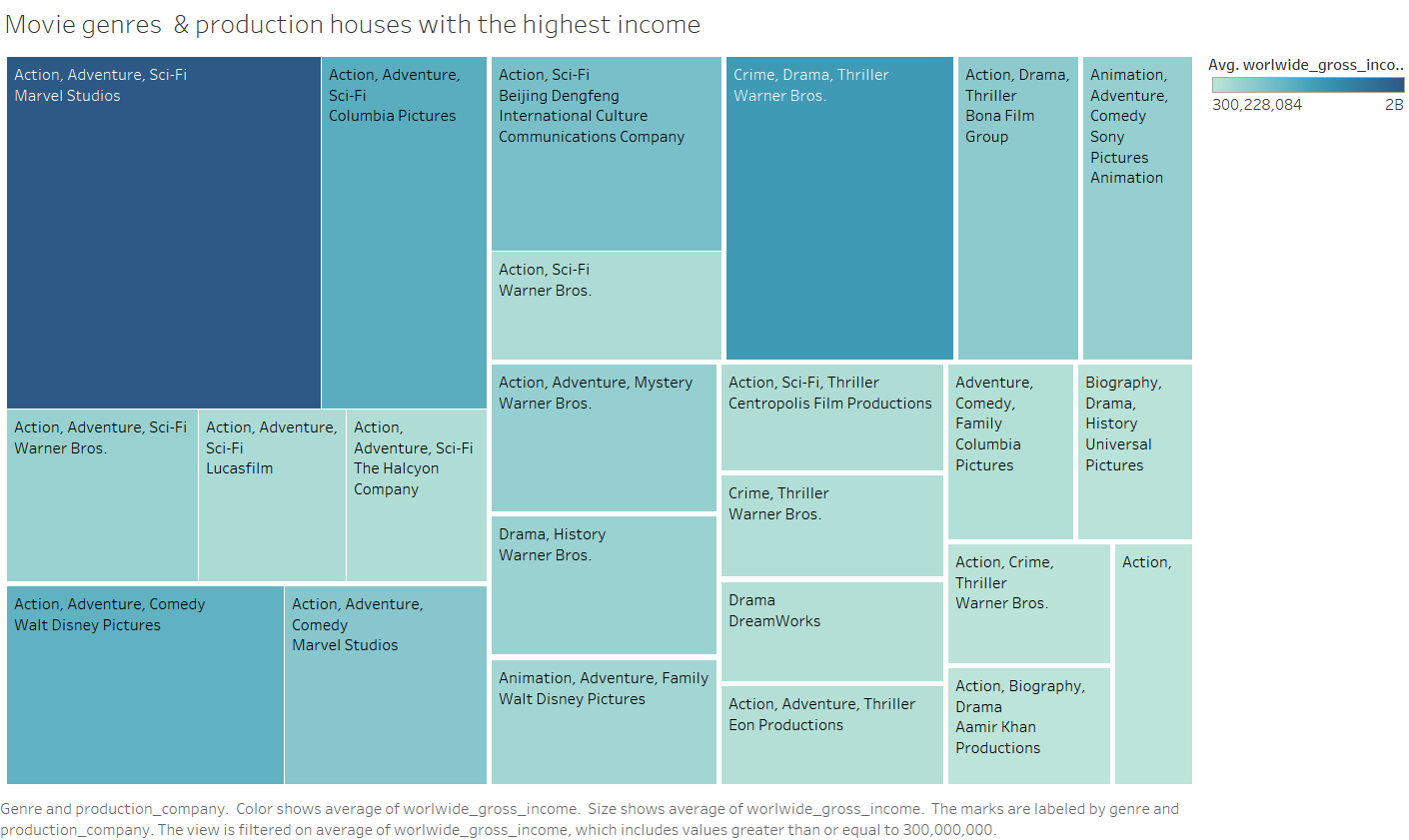
|  |  |
| --- | --- |
| Participant ID | 1 |
| Task ID | 3 |
| Overall Result | Success |
| Process | User looking at the stacked area chart.  User identifies that there is a higher amount of movie rated R content. |
| SEQ Score | 5 |
| Errors |  |
| Other Notes |  |

|  |  |
| --- | --- |
| Participant ID | 2 |
| Task ID | 1 |
| Overall Result | Success |
| Process | User looked at the bar graph  User has identified that the graph is about maturity ratings. |
| SEQ Score | 5 |
| Errors |  |
| Other Notes |  |

|  |  |
| --- | --- |
| Participant ID | 2 |
| Task ID | 2 |
| Overall Result | Success |
| Process | User looked at the graph.  User was able to easily identify the most available maturity rated content. |
| SEQ Score | 5 |
| Errors |  |
| Other Notes |  |

|  |  |
| --- | --- |
| Participant ID | 2 |
| Task ID | 3 |
| Overall Result | Success |
| Process | User looking at the stacked area chart.  User was able to spot the difference between the 2 categories- fewer R rated TV Shows. |
| SEQ Score | 4 |
| Errors |  |
| Other Notes |  |

**Task 5**



Analysis

For task 5, we wanted to find out the highest earning genres and production companies. In order to carry out this task, we had to refer to multiple columns of data. We also had to join another dataset, because Netflix dataset didn’t have the worldwide income and gross earning columns. This visualization only considers movies because the income data for tv shows was not available in the IMDB dataset. Tree maps was chosen to visualize multiple columns of data. The size of the tiles represent the popularity of Genres while the shade of the tiles represent the revenue they bring in. The combination of shape and shades simplify the message of the visualisation. Darker shades represent higher earnings by the production company and larger size represents higher earning from the genres.

Visualization

We had to visualize multiple columns to find out the highest earning genres and production companies. Tree maps was the visual of choice to represent multiple column magnitudes in a single frame. The size and shade express the magnitude of values from separate columns.

Implementation

To implement task 5, we performed the following pre-processing steps- Joined Netflix dataset with IMDB movies dataset. Inner join was done on Titles column using tableau prep. We chose Worldwide gross income, Genre and Production\_Companies columns for the visualisation on Tableau. The graph of choice was a treemap. Lastly, we filtered out the data where gross worldwide earnings were lower than 300,000,000 (300 million). This was done to improve the visual.

IMDB dataset source https://www.kaggle.com/stefanoleone992/imdb-extensive-dataset

Evaluation:

For task 5, we used cognitive walk through technique for evaluation. We asked 2 participants to evaluate the visualization against a set criterion and asked them questions relating to their experience with the visualisation. The questions asked were:

* Question 1: Is the user able to understand the visualization?
* Question 2: Is the visualization complex or easy to understand?

Results

|  |  |  |
| --- | --- | --- |
| Cognitive Walkthrough | Question 1 | Question 2 |
| Participant 1 |  |  |
| Readability of the visualization | Yes | Yes |
| Interpretability of visualization | Yes | Yes |
| Participant 2 |  |  |
| Readability of the visualization | Yes | Yes |
| Interpretability of visualization | Yes | Yes |

Summary of results

After visualizing 3 columns- Worldwide gross income, Genre and Production\_Companies, we were able to understand the type of movies that are popular and their relationship with profitability. We were also able to link the profitability with the production companies.

From the visual – “Movie genres and production houses with the highest profit”, we were able to identify the Genres and production companies that perform well at the box office. Action movies tend to perform well at the box office and Marvel studios has produced the highest earning movies in that genre. Marvel studios have so far earned $1,697,820,296 from movies that belong to “Action, Adventure and Sci-Fi” in gross revenue. Columbia Pictures is the 2nd highest earning production house with movies belonging to the same genre. 2nd highest earning genre is “Crime, Drama, Thriller” and has made Warner Bros $1,074,251,311.

Discussion

Marvel has been extremely successful at the box office in the “Action, Adventure and Sci-Fi” genre and this can be attributed to the success of their Avengers franchise. Columbia films (Owned by Sony corporation of America) have earned millions of dollars from the same genre and some of this success can be explained by the success of Spiderman movies. Warner bros are 3rd when it comes to earnings in the same genre indicating that their DC movies division isn’t as popular as marvel’s cinematic universe. However, they have the highest worldwide income in the “Crime, Drama, Thriller” category. Apart from Action, the other moneymaking genres are Drama and Comedy. The revenue from these movies on average is the same among the competing production companies.