The Interplay of Genres and Tags in Movie Popularity and Ratings

Michael Wong

MicroMasters®

edX | UC San Diego

Dataset

I chose the MovieLens dataset we explored in the course for this analysis. The dataset contains movie information, user-provided associated tags, and ratings. This dataset is suitable for exploring patterns related to movie genres, tags, and ratings that can help content creators, distributors, and platforms to promote content that resonates with viewers.

Motivation

Understanding what drives movie popularity and high ratings in the entertainment industry is crucial for decision-making processes, from production to marketing.

Examining the correlation between genre-tag combinations and movie ratings can provide valuable insights and guidance for movie producers, streaming platforms, and movie recommendation systems, among others, by highlighting the types of movies that are both popular and receive high ratings.

Research Question

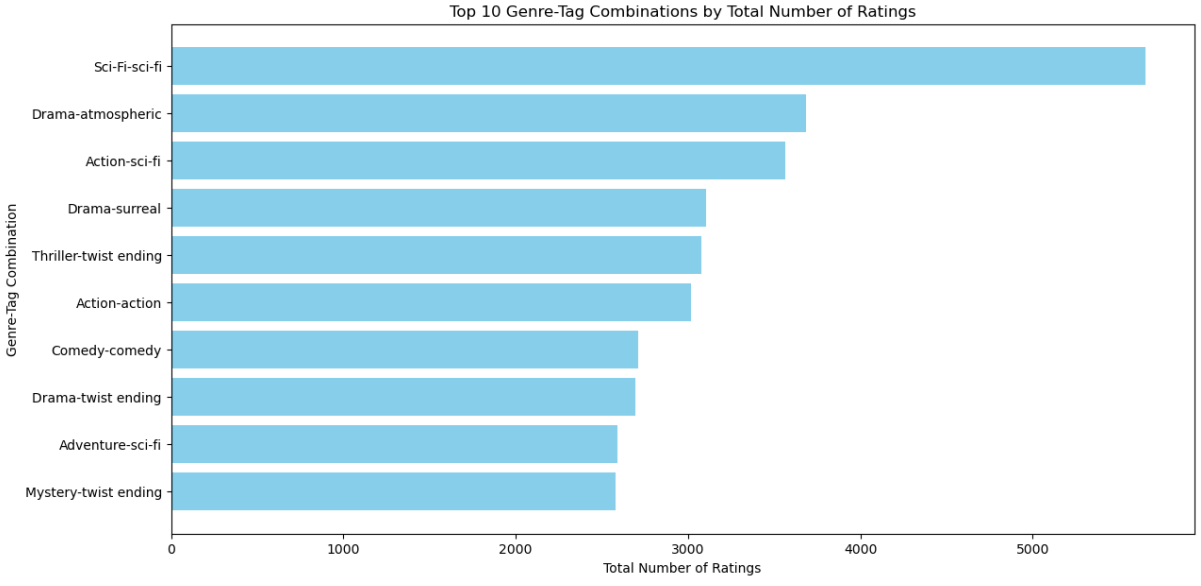
The research question I proposed is: "Do certain genre and tag combinations tend to receive more ratings in total (indicating popularity) while their average rating is above the average of all other combinations?"

This question aims to identify specific genre-tag combinations that are both popular (with a high number of ratings) and highly rated (with an average rating above the overall average), uncover the relationship between genre-tag combinations, and provide a nuanced understanding of viewer preferences.

MicroMasters®

edX® | UC San Diego

Findings



The first plot shows the total ratings for the top 10 genre-tag combinations.

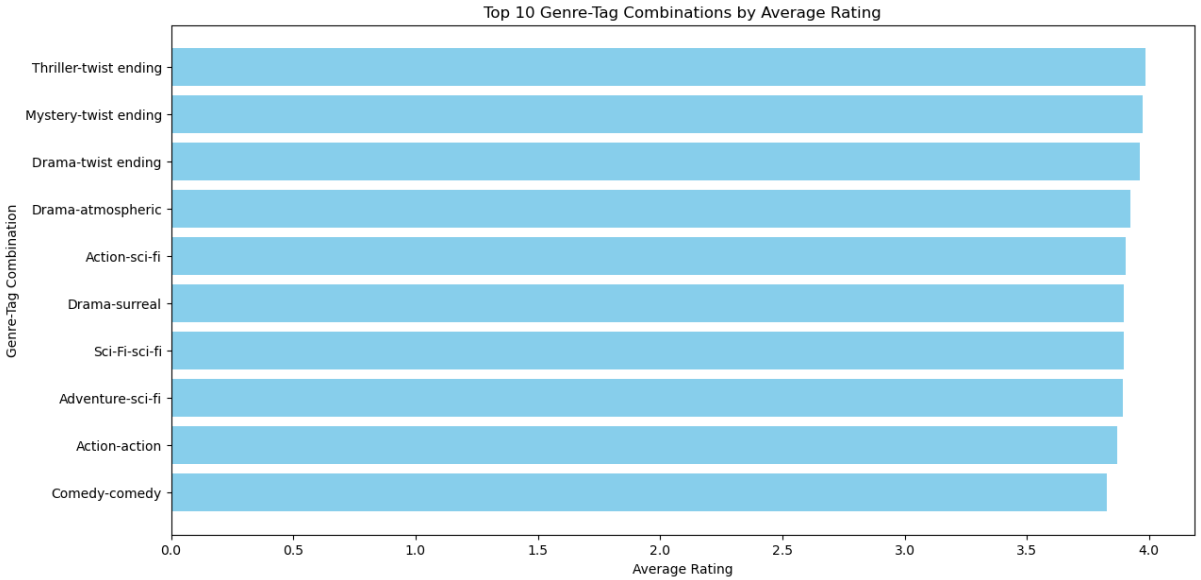
The Sci-Fi genre with the sci-fi tag combination received the highest number of ratings, indicating that this combination is the most popular among viewers.

Other genre-tag combinations, such as "Drama-atmospheric," "Action-sci-fi," and "Drama-surreal," also drew a high number of ratings, suggesting that these combinations significantly appeal to the audience and contribute to the movie's popularity.

MicroMasters®

edX® | UC San Diego

Findings



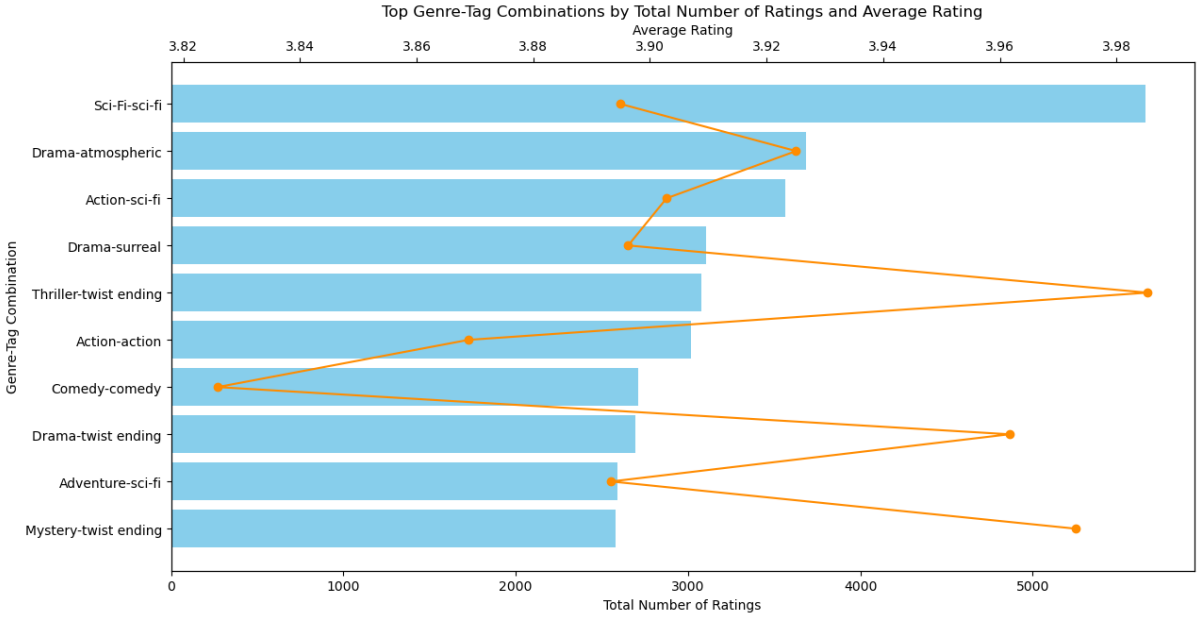
The second plot highlights that certain combinations of genres and tags not only captivate viewer interest but also receive a high appraisal, as all top 10 combinations have an average rating exceeding 3.8, significantly surpassing the overall average rating.

Specifically, "Thriller-twist ending, "Mystery-twist ending, "Drama-twist ending," and "Sci-Fi-sci-fi" combinations exhibit high average ratings, further indicating viewer preference and satisfaction for these specific genre-tag combinations.

MicroMasters®

edX® | UC San Diego

Findings



The third plot composites both the popularity (the bar part) and the average rating (the line part) for the overlapping genre-tag combinations.

Certain genre-tag combinations are both popular and highly rated. Specifically, "Sci-Fi-sci-fi" is the most popular combination with the highest total number of ratings and an above-average rating, indicating its broad appeal.

While not the most rated, the "Thriller-twist ending" combination has the highest average rating, suggesting that it is highly appreciated by those who watch it.

Other combinations, such as "Adventure-sci-fi" and "Drama-atmospheric," also demonstrate popularity and high ratings, implying a correlation between the number of ratings and the average rating.

MicroMasters®

edX® | UC San Diego

Final Thoughts

In conclusion, the analysis reveals that certain genre and tag combinations significantly impact a movie's popularity and ratings. Specifically, combinations associated with the genres of Sci-Fi, Drama, Action, Thriller, Adventure, and Mystery, paired with tags such as sci-fi, atmospheric, twist ending, surreal, action, and thought-provoking, tend to receive more ratings in total, indicating their popularity.

At the same time, these combinations also uphold an average rating higher than all other combinations, thereby indicating the potential for the creators to elevate the work to a commendable level.

Acknowledgements

I would like to thank Dr. Leo Porter, Dr. Ilkay Altintas, and other staff working behind the scenes at UC San Diego for delivering this fantastic content for DSE200x.

Also, I would like to thank my peers from DSE200x at UCSanDiegoX for providing valuable feedback when evaluating my project.

MicroMasters®

edX® | UC San Diego

References

The analysis conducted in this project was guided by understanding gained from the following article: F. Maxwell Harper and Joseph A. Konstan. 2015. The MovieLens Datasets: History and Context. ACM Transactions on Interactive Intelligent Systems (TiiS) 5, 4: 19:1–19:19. https://doi.org/10.1145/2827872

Beyond the insights from this article, the remainder of the work, including data preparation, analysis, and visualization, was based on standard practices in data science and carried out independently.

MicroMasters®

edX® | UC San Diego