Problem Statement

Business Context

Streaming media services facilitate on-demand or real-time presentation and distribution of audio, video, and multimedia content across a communications route without downloading the files to their systems. This saves users time and storage, and at the same time provides the media owners with built-in copy protection. In today's digital space, streaming has become an influential medium for accessing information. Improved connectivity and advancement in technology have made streaming services accessible to almost everyone having an internet connection, and the surging demand for on-demand entertainment services such as entertainment programs and live matches is boosting the adoption of streaming media services globally.

Streamist is a streaming company that streams web series and movies to a worldwide audience. Every content on their portal is rated by the viewers, and the portal also provides other information for the content like the number of people who have watched it, the number of people who want to watch it, the number of episodes, duration of an episode, etc.

Objective

Streamist is currently focusing on the anime available in their portal and wants to identify the most important factors involved in rating an anime. As a data scientist, you are tasked with analyzing the portal's anime data and identifying the important factors by building a predictive model to predict the rating of an anime.

Data Dictionary

Each record in the database provides a description of an anime. A detailed data dictionary can be found below.

1. title: title of the anime
2. mediaType: format of publication
3. eps: number of episodes (movies are considered 1 episode)
4. duration: duration of an episode in minutes
5. startYr: the year that airing started
6. finishYr: the year that airing finished
7. description: the synopsis of the plot
8. contentWarn: content warning
9. watched: number of users that completed it
10. watching: number of users that are watching it
11. rating: average user rating
12. votes: number of votes that contribute to the rating
13. studio_primary: studios responsible for creation
14. studios_colab: whether there was a collaboration between studios for anime production
15. genre: genre to which the anime belongs