Streaming Service Movie Reviews

This business model is a comparison between streaming apps that a large amount of the population is using today. The dataset labeled “Movies on Netflix, Prime Video, Hulu and Disney+” which includes the id for the movie, the movie title, year released, age range allowed for movie, which application the movie was on, and a rating out of 100. Exploratory analysis was done to see the number of movies on each platform, scatter plot of movies release date, and histograms showing rotten tomato ratings overall and on the two most popular streaming services (Netflix and Prime Video).

After EDA, I dropped unnecessary columns like the ‘title’ and ‘type’ column, removed duplicates, and looked for missing values. The age was then converted into a useable form and missing values were replaced with the average age of the dataset. Following this, I double checked for missing values and moved towards models.

A random forest and linear regression model were done to understand the relationships in the data more between the ratings and streaming services. Both were quantified by their Mean Standard Error and R Squared Value, in addition to the top features and their coefficient score.

Post model analysis, suggestions are made for streaming service businesses to have a focus on their future data to better understand ways to stay relevant against competition.