

Executive Summary

The shift in consumer behavior relating to the way TV is consumed has resulted from the rise of the internet and the tools it provides to watchers. Everybody can look up for a title before start watching a show and find thousands of reviews and relevant information about that specific title. They can decide if it is worth consuming even without watching a single episode.

As a result of this behavior massive data can be available to help producers to decide on what they want to invest next. It is the two-way era where what consumers think of a product is massively spread throughout the internet and serve as recommendation feature for future customers. Companies now can have access to data on what success look like based on the customer's perspective and produce their shows based on what they value most and are interested in / expecting to see.

Based on this premise, the Data Science team from MRO was hired to analyze IMDB TV Shows data to identify what can predict the success of a show and what attributes predict the success of a show (as measured by the average IMDB rating).

To conduct this analysis, we were first asked to use the list of top 250 TV shows in IMDB but our team decided to expand this title's list by utilizing web scraping techniques to gather information about TV Shows in the US in the last 27 years that had a minimum user-rate number of 5,000, with ratings between 1 and 10. We came up with 615 titles to be used in the analysis. The main purpose of expanding the list was the need to have low rate shows in order to build a model that could make predictions based not only on successful shows but low rated ones as well.

After selecting the titles, information such release data, runtime, creator, cast, production, rating and reviews was collected on each of them and used to create a model to predict whether a title would be rated above or below average based on the available rating for each of the shows in the list. Due to limitation of data availability, after running the analysis, our model showed an accuracy of 74% when predicting the rating of a show. More data would be crucial to create more accurate models to make these predictions.

By analyzing the available data, we could come up with some attributes that are relevant to a show success. Not considering the attributes related with the rating (rate counts, reviews, popularity) because those are intended to exist only after a show is launched, the model showed that runtime, number of episodes, genre and creator tend to be the most powerful features to predict a show success.

As final recommendation, the acquisition of more data would be extremely significant to improve the model and make more accurate predictions. Detailed information about the analysis can be found in the attached report.

Marina Oliveira
CEO – MRO Analysis