Analysis of Key Factors Influencing TV Shows' Success

Aakarsha LNU
alnu2@binghamton.edu
SUNY Binghamton
Binghamton, New York, USA

Riya Yeshwant Thakur
rthakur1@binghamton.edu
SUNY Binghamton
Binghamton, New York, USA

Aayushi Ahlawat
aahlawa1@binghamton.edu
SUNY Binghamton
Binghamton, New York, USA

Ishan Bagchi
ibagchi1@binghamton.edu
SUNY Binghamton
Binghamton, New York, USA

Tarun Tiwari
ttiwari1@binghamton.edu
SUNY Binghamton
Binghamton, New York, USA

ABSTRACT

These days, television and movies are the two main forms of entertainment. Everybody enjoys watching and reading fascinating stories. When we watch television, we observe that many people discuss their experiences. On social media, anyone can publicly voice their opinions on any subject. First off, the big social networking sites like Twitter, Instagram, TMDB, and Reddit create a ton of information about user behaviors, news, rising trends, etc. There is a huge amount of user-generated content available, and this al- ters the environment significantly. The genre, cast, number of seasons, number of episodes, production companies, networks, episode run time all play a role in determining whether a television program is successful or not. Thus, by examining these elements for a range of programs, we aim to predict which TV show has a higher chance of becoming successful.

KEYWORDS

Tv shows, Genre, Cast, Budget, Reviews, Region, Twitter, Reddit, TMDB

1 INTRODUCTION

We all watch a lot of television. Who does not enjoy having some spare time on the weekend? With the increasing popularity of OTT platforms and the media business, individuals have more options for different TV series and movies, which has increased competition in the media sector. People in the entertainment industry are always seeking for new strategies to make their series or movie a success. Thus, data analysis is becoming increasingly important in managerial decision-making in media enterprises throughout the world. The dependence of subscription video on demand providers, such as Netflix, on data analytics to support choices regarding new content investment is widely documented. Tmdb is a website that provides millions of movies, TV shows for people to explore. It provides us with the ability to search for the various genre movies, TV shows, the most popular movies or TV shows currently streaming on TV, theaters or available for rent. The platform also provides users to post reviews about the movies/TV shows and provide ratings. The website provides recommendations based on the movies searched or watched by the user. We will make use of the TMDB API which is available on the TMDB API Developers page. The aim of this project is to determine the success of the TV show based on major factors such as budget,cast etc.

Twitter is an open social network that people use to converse with each other in short messages, known as tweets. People publish their comments regarding actors, production companies, and TV shows, which is useful for our study of the most popular and trending shows and their dependence on various factors. Reddit is a social news aggregation, content review, and debate platform. The website's material, which includes links, text entries, photographs, and videos, is contributed by registered users and is then rated by other users. Posts are arranged into user-made boards called "communities" or "subreddits" according to their subjects. When there are enough upvotes, posts that have received the most upvotes will eventually appear on the first page of the website.

2 DATA COLLECTION

In this project we are collecting data from 3 data sources, namely, TMDB, Twitter, and Reddit which would be more than enough to find the essential factors which are contributing towards the success rate or the popularity of a TV Show.

3 RESEARCH QUESTIONS

1. How are elements like the genre, cast, number of seasons, number of episodes, production firms, networks, and length of the episodes affecting the success of TV shows?
2. To what extent are popular TV series on TMDB hyped on Twitter, and if this hype is positive or negative?
3. What phrases/terms from major TV series are becoming popular and most used or admired by the public on their respective subreddits?

4 METHODOLOGY

We are going to perform the following analysis on the available dataset:

4.1 DATA PREPROCESSING

We are going to perform basic text pre-processing steps like tokenization, lower casing, stop words removal, stemming, and lemmatization.

4.2 Descriptive and Predictive Analysis

Using both recent and old data, descriptive analytics seeks out patterns and connections. It is one of the most crucial processes in the examination of statistical data. It provides you with an analysis of the distribution of your data, aids in the detection of errors

and outliers, and enables you to spot patterns between variables, preparing you for future statistical analysis.

For predictive analysis, we will be considering the independent variables such as genre, cast, number of seasons, number of episodes, production companies, networks, episode run time to determine the popularity of TV shows which are already aired. We will create a machine learning model to predict the TV shows' popularity which will be aired in the future based on the factors considered above.

We will analyze multiple factors like Genre, cast, number of seasons, number of episodes, production companies, networks, episode run time etc which will impact the popularity of a tv show and then plot a graph.

4.3 Qualitative Analysis

The process of gathering, analyzing, and interpreting qualitative data—non-numeric, conceptual information, and user feedback—in order to identify themes and patterns, respond to research questions, and decide what steps need to be taken to improve the data collected for a particular product or website is basically what we do in qualitative analysis.

For qualitative analysis, we will get the most popular TV shows from TMDB database and can predict its popularity/hype on twitter based on tweets and frequency of keywords extracted from the tweets. We will also establish the sentiment of the hype if it's positive or negative.

4.4 Text Analysis

We will analyze the popular phrases of a few specific television series from their respective subreddits and create a word cloud for each of them. It will provide a close analysis of most frequently used words which are directly linked to the popular dialogues.

5 CONCLUSION

Our goal is to do data analysis on the information obtained from TMDB, Twitter, and Reddit following the data gathering process. In order to respond to the research questions, we will do predictive, qualitative, and text analysis. Using data visualization methods, we will also visually depict the outcomes and trends.