Popularity of TV Shows

Anthony Gardner

**Abstract— Television has played a transformative role in shaping global culture, serving as a platform for storytelling, information dissemination, and social connection. With the evolution of streaming services and the decline of traditional cable TV, the dynamics of television popularity and ratings have shifted. This study explores the relationship between TV show popularity and critical reception. Through data analysis and visualization techniques, the research investigates correlations between popularity, ratings, and viewership trends over time. Results reveal minimal correlation between popularity and ratings, highlighting the distinction between ratings and appeal. These findings offer insights into the evolving landscape of television consumption.**

**Keywords—Television, Popularity, Ratings, Streaming, Data Analysis**

# Introduction

Tv has had a major impact on culture across the world. It being a medium to share stories, learn about current events, and connect people globally. With the rise of streaming and the death of cable television what effect does this have on landscape of television? Does popularity and ratings correlate? Are the best shows the ones getting the time to shine? Looking at different reviews and viewership numbers we can get this information.

# Datasets

## Source of dataset

The data used was found on Kaggle, a credible data source. The data is from 2022 and collected the ratings and popularity of a ton of different shows. The data was gathered by The Movie Database API

## Character of the datasets

the dataset is from 2023 and contains 2616 entries each with 8 different data points

1) first\_air\_date - The date when the show was first aired on television

2) origin\_country - The country where the show was created / originates from

3) original\_language - The original language of the show

4) name - Name of the show in English. Note that names in original language are not included in this dataset.

5) popularity - A metric that measures how popular a TV show is based on consumer views

6) vote\_average - Average of the total number of votes the show received

7) vote\_count - The number of votes the show received

8) overview - A brief description of the show

These data points are all very clear with only a couple null values

# Methodology

## Data Overview

The first thing I did was create look at general information about that data. A table was created with the necessary information containing mean, standard deviation, min, max, and 25th and 75th percentile using the data describe function in pandas

I also wanted to take a closer look at some of the more popular shows, so I mad another table containing the top 5 most popular shows and then the highest rated shows

## Heat map

The next thing I wanted to look at was the correlation between quality and popularity, I also wanted to see if the amount of votes a show received impacted the rating as well. Using the seaborn correlation function I was able to produce a heatmap

Although the heatmap is a very simplistic way of looking at the data and needed to be preprocessed for null values. It is a visually clear and easy to understand way of visualizing the data

## Year vs Popularity Graph

Example: next I wanted to look at papulation vs the year using graph for this I used MATLAB. The first attempt did not go as planned the graph look awkward having a big spike in popularity in 2023. I decided that the data needed to be filtered to give a more accurate graph. I removed any extreme outliers and checked how many data points per year there were.

As the years increased so did the amount of data points beside for 2023 this being 2-year-old data 2023 having few data points made sense as the study was conducted in 2023. With the filtered data I created a new graph this time feeling cleaner.

The graph is a very effective was of showing the trends in show popularity although its hard to explain some of the trend like the spike at around 1963

# Results

## Data Overview

|  | Data | | |
| --- | --- | --- | --- |
| Popularity | Vote Average | Vote Count |
| count | 2617 | 2617 | 2617 |
| mean | 59.805 | 7.692 | 604.823 |
| std | 22.409 | 0.617 | 1223.234 |
| min | 0.866 | 0.6 | 99 |
| 25% | 16.567 | 7.3 | 150 |
| 50% | 27.489 | 7.7 | 257 |
| 75% | 49.765 | 8.1 | 569 |
| max | 6684.611 | 9.0 | 19459 |

Fig. 1. General description table

This is a general overview of the data we received. I would like to draw attention to a couple key points. First the lowest voting average was a 0.6 while the 25th percentile has a 7.3 being much higher than then the lows showing that most show have an above average (5/10) reception. If we look at the data provided for vote count, we can see that most shows have a very low number of vote for there rating the 75th percentile being 569 but the max being 19459 severely driving our mean up giving us 604.84

When looking at the popularity we can see the max is 6684.611 which is a huge difference from most other popularity scores given in the data this is most likely the data point that messed up our results in part C

## Heat Map

A red and blue squares

Description automatically generated

Fig. 2. Popularity and vote correlation heat map

The heat map shows that there is very little correlation between the popularity of a show and the rating of a show. This to me is surprising I thought that the more popular a show was the higher reviews it would have but this does not seem to be the case

Popular things are not necessarily the best but this I think that the low number of votes collected play a role in this. The votes that were collected were from critics and thus their job is to look at shows more objectively popularity not being a factor

## Year vs Popularity Graph

#### Fig. 3. Unfilter year vs popularity graph

#### A graph showing a number of release Description automatically generated

#### Fig. 4. Filtered year vs popularity graph

I already talked a little bit about why the first unfiltered one went wrong but to expand on that this data was published in early 2023 and as such the only data that they had were super popular shows that are at the start of the year this explains the second spike as the data for 2023 is incomplete

The spike in 1963 can also be explained the range of shows during that period is very small each year only havening 5 or 6 shows to pick from 1963 in particular saw the start of 2 particularly popular shows Doctor who and General Hospital.

These results leave me a bit unsatisfied as I expected the popularity of shows to go down over time with much more content being made and the rise of streaming but the graph aside from a couple points seems to fluctuate in the same range

# Discussion

I feel like the dataset I had was very good, but I wish I could have seen more about the effects of streaming, I am happy with the visualization methods I used but I wish I could have looked at a scatter plot of vote and popularity to see all the different data points. I also wish that the 2023 data was complete as it frustrating to have such a large spike

# Conclusion

The Creation of new shows has not stop and the popularity on shows haven’t gone anywhere. Storytelling has always been a part of human culture since verbal stories that were passed down by mouth. This has been our culture and will continue to evolve as people still tune in week after week

##### References

1. **T. Saha**, "Top Rated TV Shows," Kaggle, Accessed: Dec. 09, 2024. [Online]. Available: <https://www.kaggle.com/datasets/titassaha/top-rated-tv-shows>