

Game of Thrones Data Visualization Analysis

Inna Krmoyan, Lilit Galstyan

December, 2022

Table of Contents

Abstract	2
Introduction	2
Content.....	3
Research Methodology.....	4
Literature Review	4
Hypotheses.....	5
Exploatory Analysis	5
Conclusion.....	15
References.....	16

Abstract

This paper presents a set of visualizations and hypotheses about “Game of Thrones”, an American drama television series created for the HBO television network based on George R.R. Martin’s series of fantasy novels. In this report we will attempt to create useful and insightful data visualizations so that they might answer many of the questions of the fans and enthusiasts, create a better understanding of the series and overall make people interpret many factors of this television series easier. An interactive dashboard with trend charts and synchronized visual analytics components not only shows how often Game of Thrones events and characters are being mentioned by journalists and viewers, but also provides a real-time account of concepts that are being associated with the unfolding storyline and each new episode. Positive or negative sentiment is computed automatically, which sheds light on the perception of actors, new plot elements and many other factors.

Introduction

David Benioff and D. B. Weiss created the HBO fantasy drama television series Game of Thrones. It is based on George R. R. Martin’s A Song of Ice and Fire series of fantasy novels, the first of which is A Game of Thrones. The show was shot in several countries, including the United Kingdom, Canada, Croatia, Iceland, Malta, Morocco, and Spain. It debuted on HBO in the United States on April 17, 2011, and ended on May 19, 2019, with 73 episodes spread across eight seasons. Game of Thrones has a big ensemble cast and follows various story arcs throughout the show. It is set on the mythical continents of Westeros and Essos. The first major storyline revolves on the Iron Crown of Westeros, which is represented by a web of political struggles between noble families competing for the throne or battling for independence from whoever sits on it. A second centers on the last descendant of the realm’s overthrown ruling line, who has been banished to Essos and is preparing a return to the throne. The third follows the Night’s Watch, a military force tasked with defending Westeros against threats from beyond its northern border. Game of Thrones has a large, active, and multinational fan base and has earned record viewership on HBO. The series has received praises from critics for its performances, complex characters, story, scale, and production values. The final season garnered much criticism for its shorter length and creative ideas, with many viewing it as a letdown. The series won 59 Primetime Emmys, the most ever for a drama series, including Outstanding Drama Series in 2015, 2016, 2018, and 2019. Other honors and nominations include three Hugo Awards for Best Dramatic Presentation, a Peabody Award, and five Golden Globe nominations for Best Television Series - Drama. Many critics and publications have ranked the show one of the greatest television series of all time. As being curious young enthusiasts we are very interested in the future and its all possible developments. We strive to know whether it is possible to predict what can be awaiting for a certain series or company in the future, with the help of data visualization and its tools it surely is possible to predict patterns to a certain extent. The aim of our project is to understand the reasons, relationships between the variables, the main driving factors that affect the popularity of the show. This huge TV show was a hit for HBO television network and it gained a lot of profit by that. As the series was a huge success it gained many fans and there are some of them among us that want to understand the relationships between the characters, how they escalated or de-escalated, what were

the reasons for that and many other interesting hypotheses like this one. Seeing how the popularity grows of the TV show throughout each release of the episode is an important factor for the HBO company to see how many new subscribers the company potentially gains and so on. We chose this series for our project as it is very famous, has a very rich plot, plenty characters, fluctuating views and a very interesting database for exploration. All in all we want to understand our data make any necessary changes to make good visualizations that will show us the fluctuations of the variables and the relationship between them.

Content

Our chosen datasets contain various information about each Episode/Season of Game of Thrones. Some of the key features of this Datasets are - Season, No. of Episodes (Season), No. of Episodes (Overall), Title of the Episode, Running Time (Minutes), Directed by, Written by, Original Air Date, U.S. Viewers (Millions), Music by, Cinematography by, Editing by, IMDb Rating, Rotten Tomatoes Rating (Percentage), Metacritic Ratings, Season Ordered, Filming Duration, Novel(s) Adapted, Synopsis of each Episode.

All columns and their definitions:

Game_of_Thrones.csv

Season - No. of Seasons No. of Episode (Season) - No. of Episode in a particular Season
No. of Episode (Overall) - No. of Episode in the whole Series Title of the Episode - Name of the Episode
Running Time (Minutes) - Runtime of the Episode in Minutes Directed by - Name(s) of the Director
Written by - Name(s) of the Writer Original Air Date - Original Air Date of the Episode
U.S. Viewers (Millions) - No of U.S. Viewers of the Episode in Millions Music by - Name(s) of the Composer
Cinematography by - Name(s) of the Cinematographer Editing by - Name(s) of the Editor
IMDb Rating - IMDb Rating of the Episode (10 point scale) Rotten Tomatoes Rating (Percentage) - Rotten Tomatoes Rating of the Episode in Percentage
Metacritic Ratings - Metacritic Rating of the Episode Ordered - Date of Series/Season renewal
Filming Duration - Filming Duration of the Season Novel(s) Adapted - Adapted from which Novel(s)
Synopsis - Synopsis of the Episode

characters_v4.csv

Character - Character Actor/ess - The actor/ess who plays the particular character
Episodes_appeared - the episodes where the particular character appeared
First_appearance - the first appearance of the particular character Last_appearance - the last appearance of the particular character

GOT_episodes_v4.csv

Season - the season Episode - episode to the corresponding season Title - the title of the episode
Release_date - the date when the episode was released Rating - the rating of a particular episode
Votes - the votes of the particular episode Summary - the summary of the particular episode
Writer_1 - first writer of the particular episode Writer_2 - second writer of the particular episode
Star_1 - the first leading star of the particular episode Star_2 - the second leading star of the particular episode
Star_3 - the third leading star of

the particular episode Users_reviews - reviews of the users Critics_reviews - reviews of the critics US_viewers - the views of US Duration - the duration Director - the director Budget_estimate - the amount of money spent on the particular episode

battles.csv

name - name of the battle year - year of the battle battle_number - the number of the battle attacker_king - the attacker king defender_king - the defender king attacker_1 - the first attacker attacker_2 - the second attacker attacker_3 - the third attacker attacker_4 - the fourth attacker defender_1 - the first defender defender_2 - the second defender defender_3 - the third defender attacker_outcome - the outcome of the attacker whether they won or not battle_type - the type of the battle major_death - whether the battle had major death rate major_capture - whether the battle had major capture rate attacker_size - the size of people from the attacker's side defender_size - the size of people from the defender's side attacker_commander - the commander of the attacker side defender_commander - the commander of the defender side summer - whether it was summer or no location - the location of the battle region - the region of the location note - notes

Research Methodology

The data for our research was taken from Kaggle and the links will be provided in the reference page below. The data visualization tools that rendered information in a visual format throughout our project are: graphs, charts, heatmaps and many more. Everything used was covered during our course with some exceptions in the interactive plotting part. Everything applied was very helpful to get the required insights and understand the datasets more thoroughly.

All of the taken datasets are provided in '.csv' format, and for our project we decided to use the most useful and common tool for data visualizations that is the R programming language. R was used to explore, process, analyze, and make the desired charts. In order to make our project much more beautiful and useful, we created an interactive dashboard by using Shiny in order for viewers to investigate our plots and find answers to their questions about their favorite TV series Game of Thrones.

Our research was mainly concentrated on creating interactive charts so that it would be possible to see the exact parts of a certain plot. We have created barplots, heatmaps, connection graphs, donut plot, overlaid histograms and many more.

Literature Review

There are several literatures that we could find where TV show analysis is done.

[1] AN INTRODUCTION TO NETWORK VISUALIZATION FOR TELEVISION STUDIES: MODELS AND PRACTICAL APPLICATIONS

The purpose of the first research is to address the benefits, problems, and limitations of using visual models for large corpus analysis in television studies. Data visualization examples were presented, using a sample database of anthology TV shows retrieved via the

Wikidata Query Service. A visual model available on the RAWGraph platform was provided as a method of identifying production and distribution flows by examining the nation of origin and the industrial entities involved in the construction of such a network. By doing so, the author demonstrated how visualizing television programs as networks, nodes, or links using visual models can provide direction not just for contextualizing the text, but also for learning about our data. Cleaning data, resolving anomalies, and integrating information from peripheral television markets – or at the very least recognizing what is lacking – allows for new perspectives on media businesses.

[2] Analyzing the public discourse on works of fiction – Detection and visualization of emotion in online coverage about HBO's Game of Thrones

This paper describes a Web intelligence portal that collects and combines news and social media coverage of "Game of Thrones," an American drama television series made for the HBO television network based on George R.R. Martin's fantasy novel series. The system gathers content from Anglo-American news media websites as well as four social media platforms: Twitter, Facebook, Google+, and YouTube. An interactive dashboard with trend charts and synchronized visual analytics components not only displays how frequently Game of Thrones events and characters are mentioned by journalists and viewers, but it also gives a real record of concepts associated with the unfolding storyline and each new episode.

[3] Analysis and Visualization of Streaming Media Platforms Based on the R Language——Take Netflix as An Example

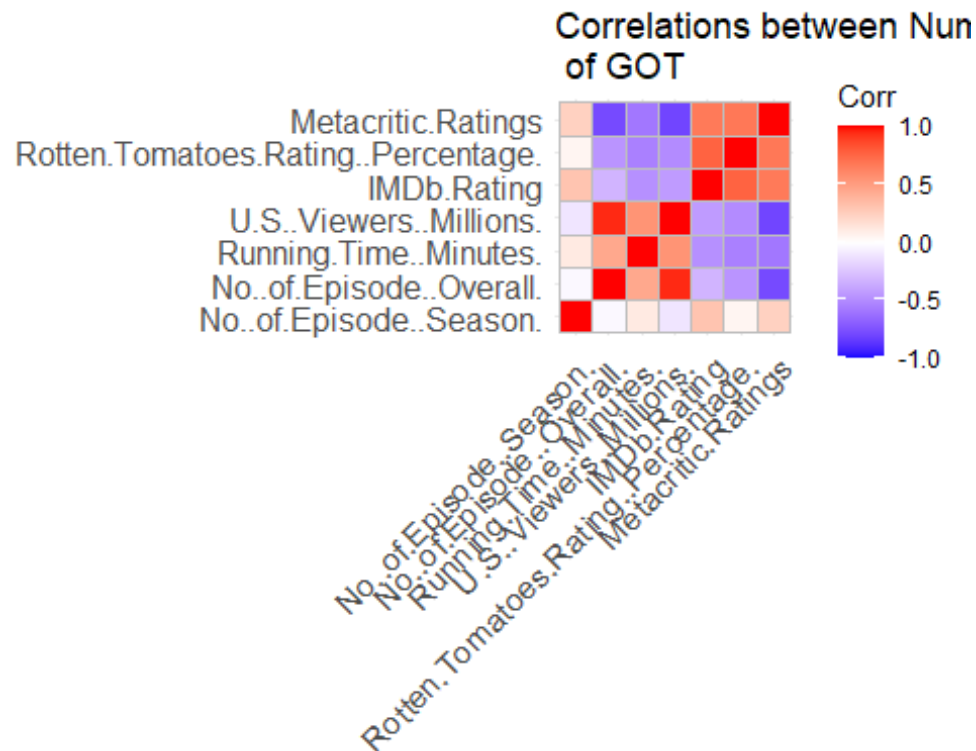
This article seeks to determine the business content given by Netflix through data analysis as well as the industry conditions. Netflix has been chosen to represent video platforms. The R programming language is the primary tool for analyzing data and displaying various characteristics in content providing. Today's streaming media technology is continuously evolving, and many streaming media providers place a premium on content richness. Mastering the content of resources aids in the creation of more accurate recommendations in order to retain consumers, and it can also serve as a model for the growth of streaming media in China.

Hypotheses

Our hypotheses are mainly concentrated at visualizing the relationships between the characters, the main factors that affect the views, the correlations between various elements in our datasets. We also want to understand how much the interest grows in GOT throughout the seasons, does that depend on the characters, the seasons, their episodes, the battles they were involved in and so on. Overall we want to analyze what factors contributed to the success of the Game of Thrones.

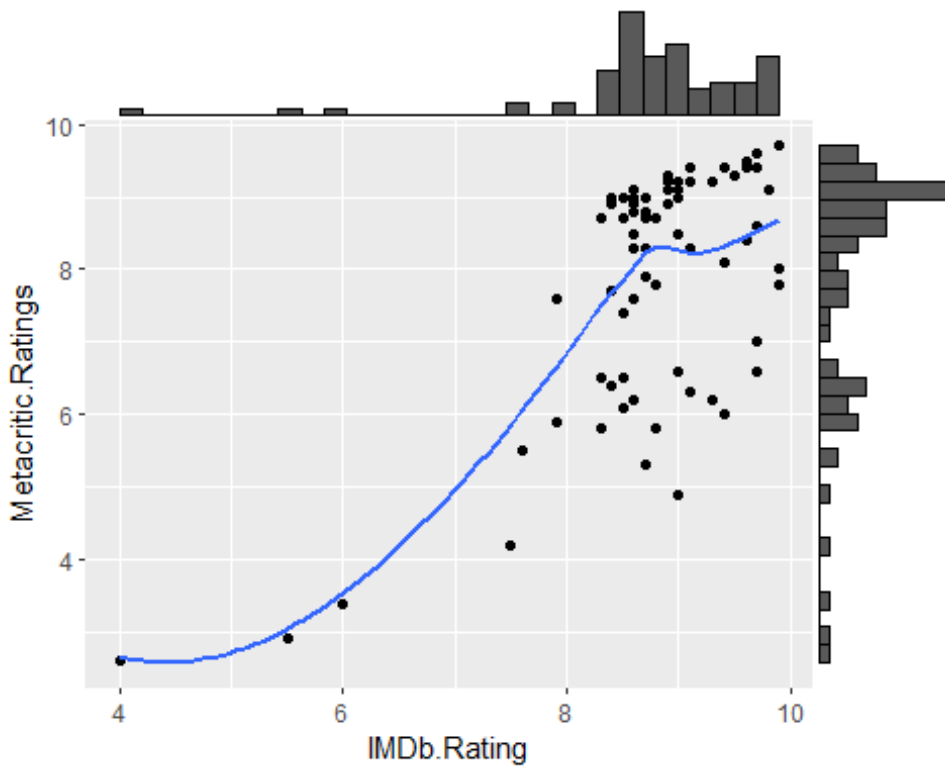
Exploratory Analysis

Correlation heatmap of the Game of Thrones dataset



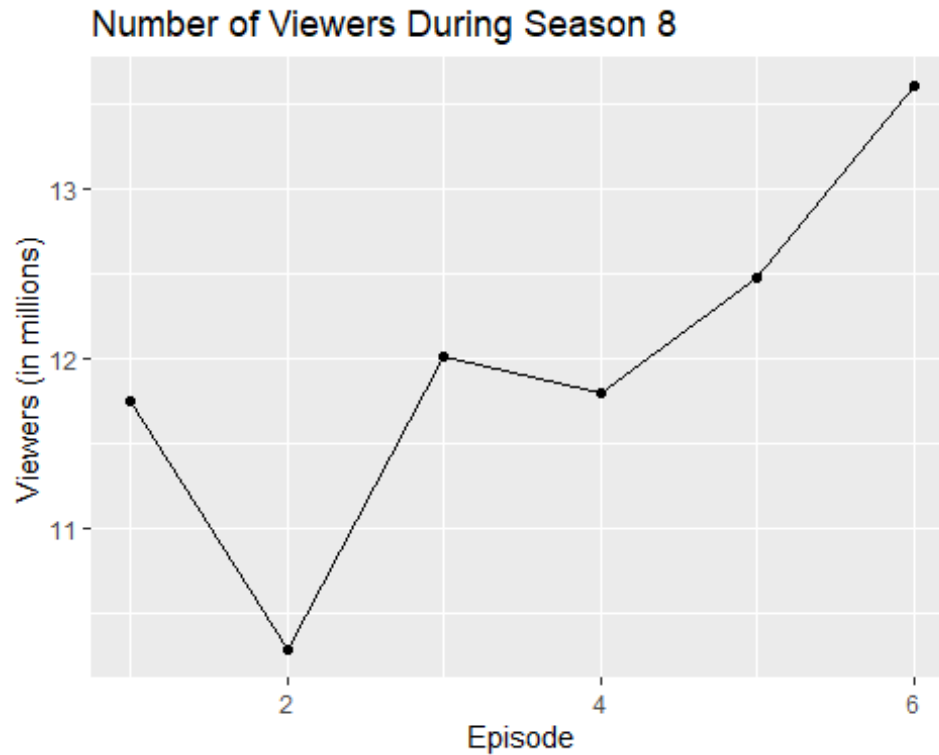
By looking at figure 1 it can be concluded, the highest correlated numeric features of GOT are IMDb Rating and Rotten Tomatoes Rating (Percentage), which makes a lot of sense as both of the websites have valid reviews so their correlation has to be positive.

User reviews vs critics how different are the distributions (reviews, episodes)



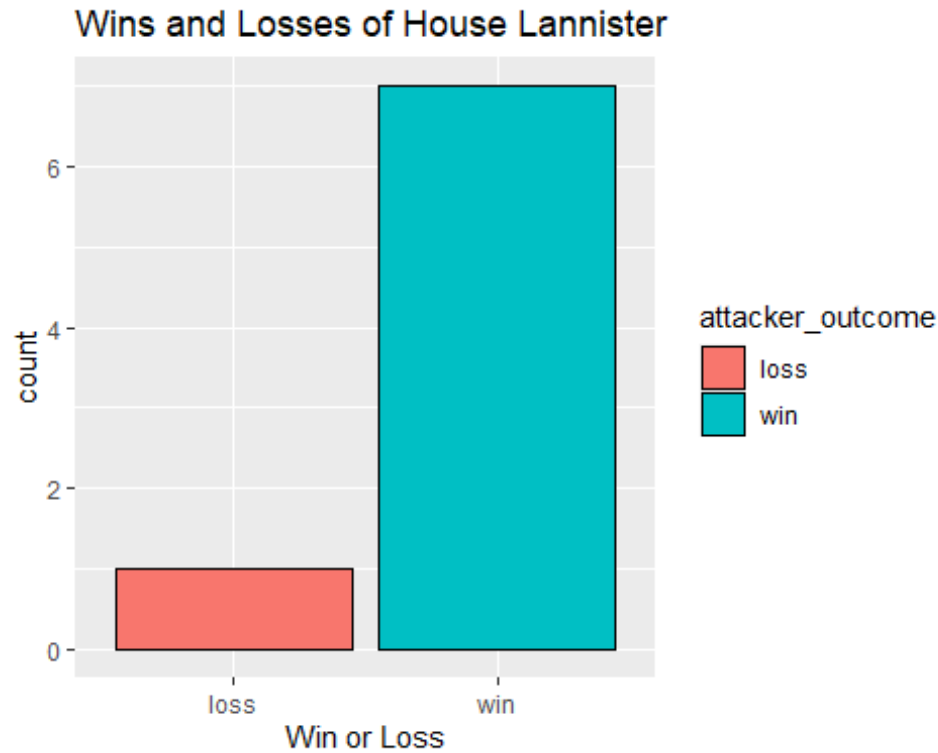
By looking at figure 2 it can be said that the IMDb and Metacritic ratings are generally in a positive relationship. Meaning that with the increase of the IMDb rating the Metacritic rating is also increased for each episode.

Did the last season's first episode cause a drastic change in views (episodes)



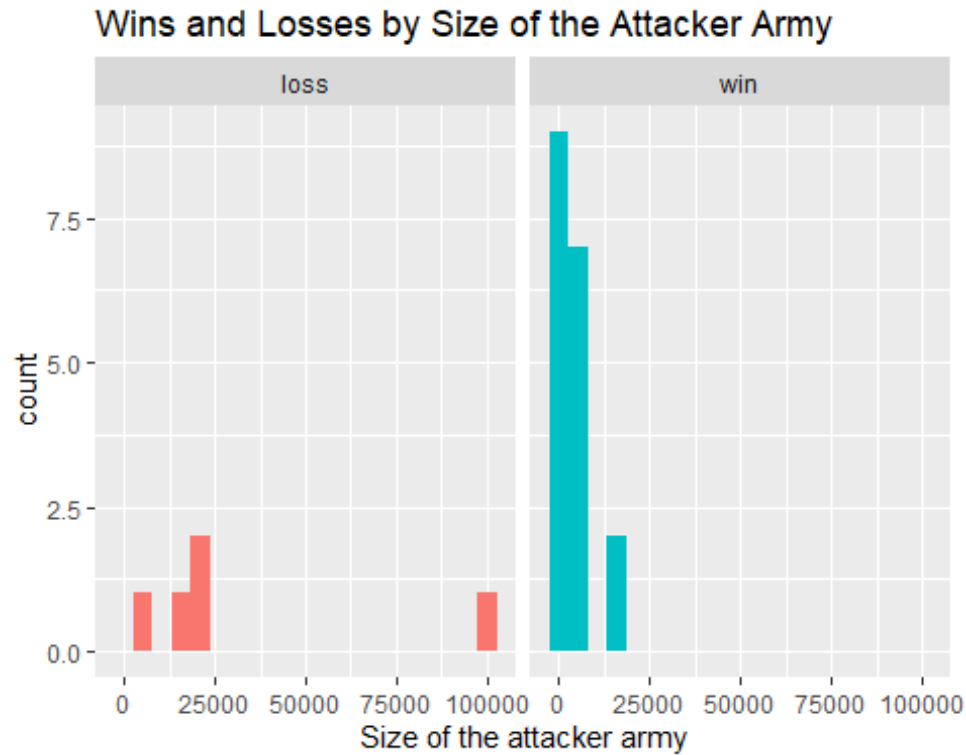
By looking at figure 3 it can be said that it is known that the 8th season is the most criticized because of its plot. Lets see if there was a change in number of views right after the first episode of the season. Just by looking at the plot, it is seen that there is decrease of views after the first episode of the 8th season. There is further increase in views after the second episode, but that is after people continued to watch the series after the first episode, but the fact is that many people stopped watch Game of Thrones after the first episode of the last season.

Lannisters were considered the most powerful house of Westeros, what percentage of battles did they win



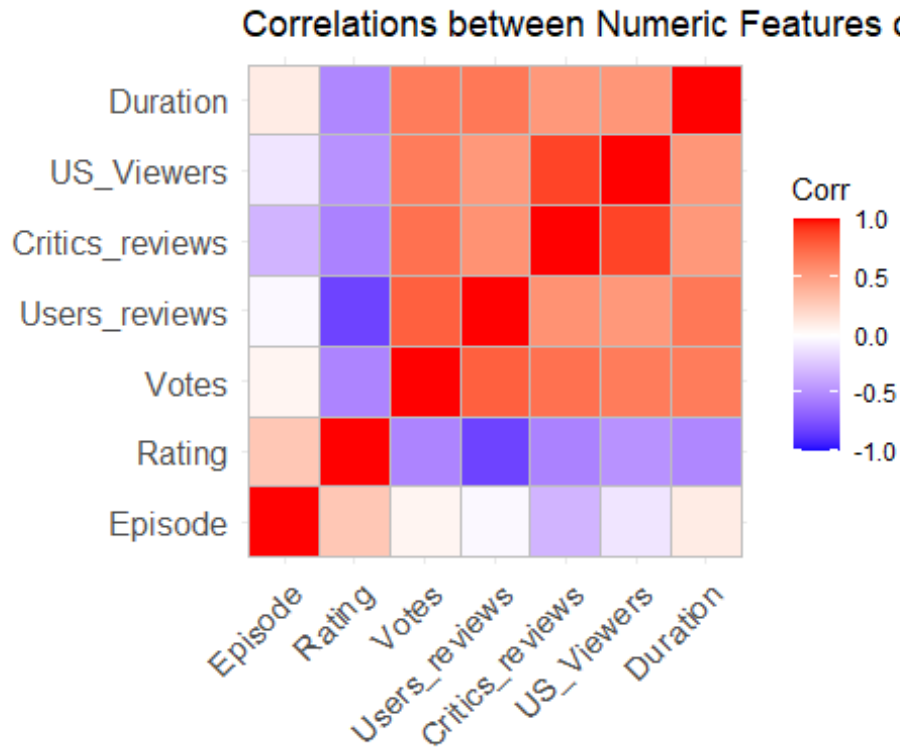
By looking at figure 4 it can be said that in total, there have been eight battles where the main attacker was house Lannister. Seven out of all eight battles have been won by Lannisters and only one lost. It can be concluded that Lannisters indeed have been very powerful.

Does always the side with the bigger size of army decisively win the battle?



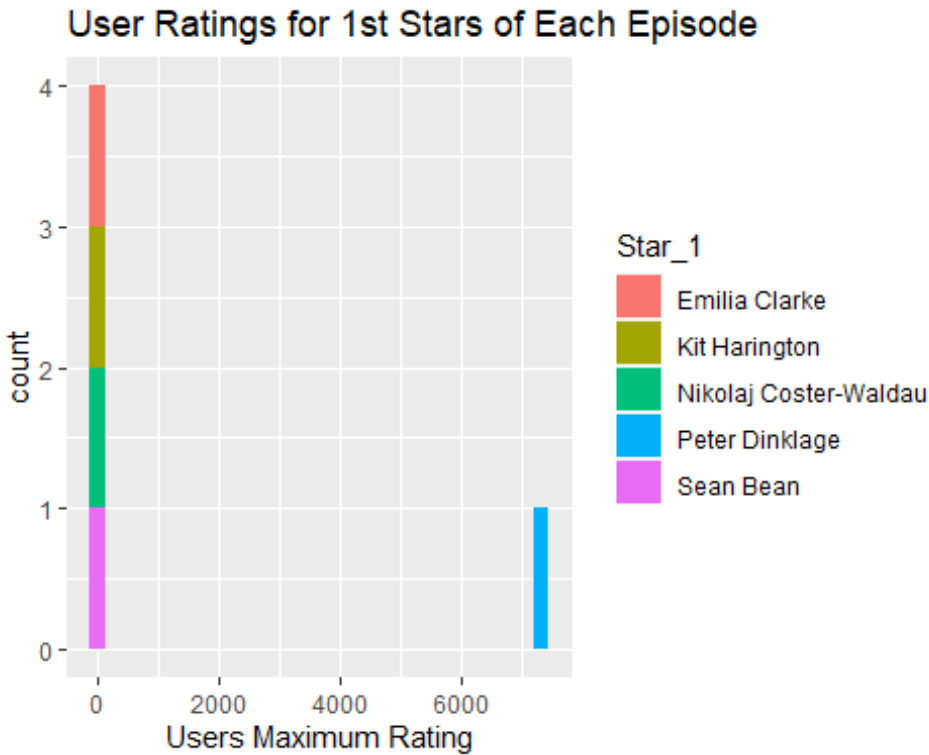
By looking at figure 5 it can be said that firstly for the losses, there are a few battles with smaller army size of attackers and there is one battle with a very big army size, close to 100k. However, there are more battles with small army size of attackers that resulted in winning. Therefore the size of the army does not really affect the outcome, whether the attacker would win or lose.

Correlation heatmap of Episodes

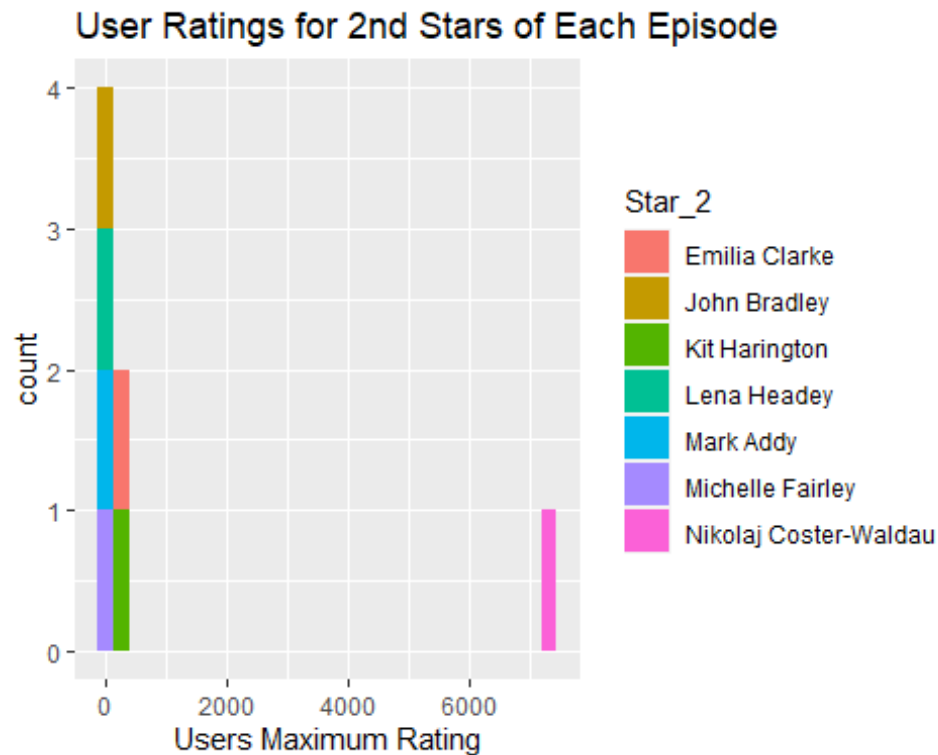


By looking at figure 6 it can be said that the highest correlated numeric features of episodes are US viewers and Critics reviews, which is quite logical because knowing the critics reviews is a good indicator of whether we would watch a TV series or not. The least correlated features are Users reviews and Rating.

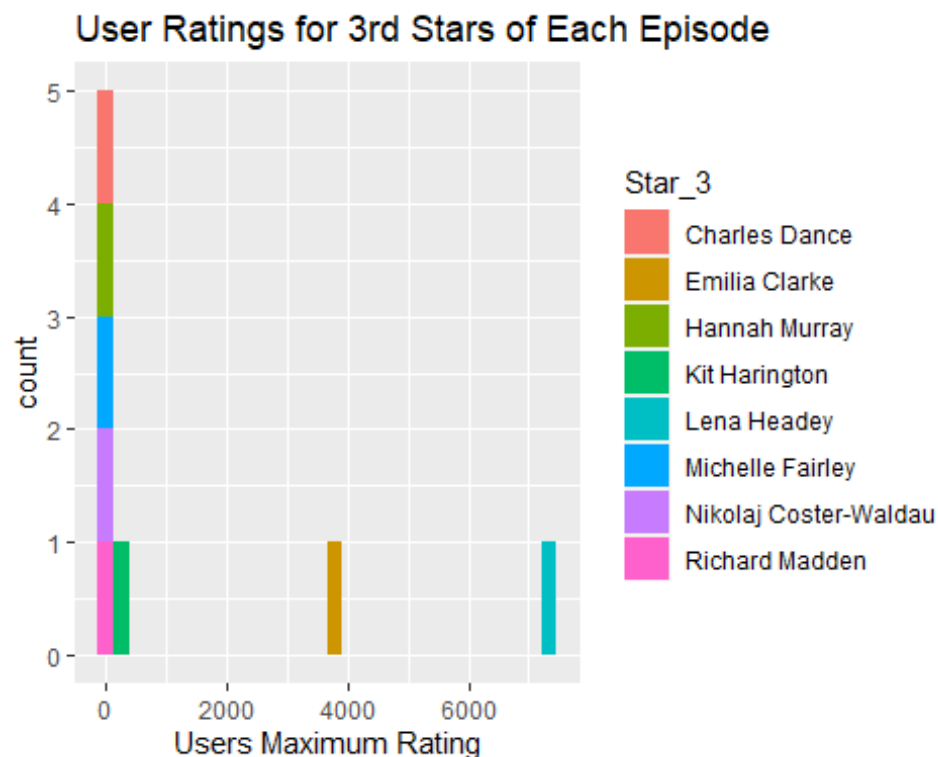
Does the starring actor of an episode have any relationship with the episode's rating?
Which stars get the highest ratings?



By looking at figure 7 it can be said that each episode has 3 main starring actors. Taken with the maximum ratings of each star, we analyze which stars get the highest ratings. There has been one episode that has been rated with a very big score of 7358 by the users. On the first plot it can be seen that Peter Dinklage, the actor of Tyrion Lannister, got to be Star 1 for the episode with the highest rating. Other actors for Star 1 got maximum ratings in the range of 36 to 106.

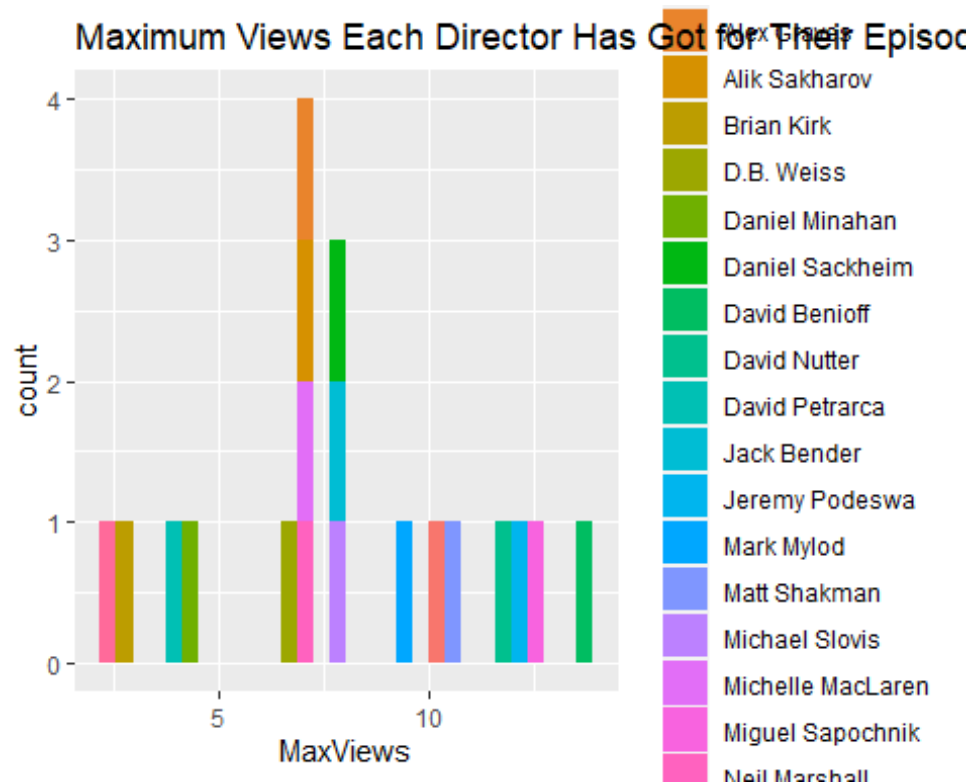


By looking at figure 8 it can be said that star 2 of the highest rated episode was Nikolaj Coster-Waldau, the actor of Jaime Lannister. Other actors who have been Star 2 got maximum ratings from 34 to 253.

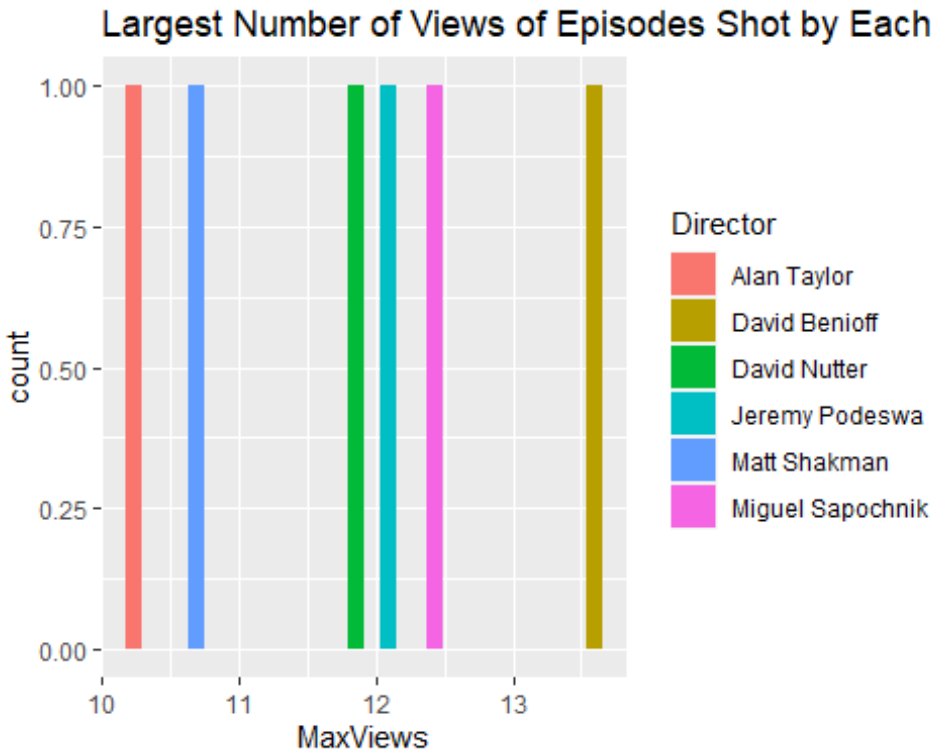


By looking at figure 9 it can be said that as Star 3 of the episode with the highest rating was Lena Headey, who is the actress of Cersei Lannister, and the other Star 3 actors got maximum ratings of 23 to 181, but there is also Star 3 who got a maximum rating of 3749 who is Emilia Clarke, the actress of Daenerys Targaryen.

Which directors get the biggest numbers of views for the episodes they have shot?



By looking at figure 10 it can be said that there are 19 directors, each has shot some amount of episodes. The goal is to find the maximum number of views each director has got for their works. This plot shows the directors for ranges of views. The smallest number of views of 2.22 million comes from the director Timothy Van Patten. But is this very informative? Lets have another hypothesis: which directors are responsible for the episodes that exceed the number of views of 10 million?



By looking at figure 11 it can be said that in this case we have only six directors left, the ones that have shot the episodes that have had views of more than 10 million. And to make the plot more informative and pleasant:

```
## Error in UseMethod("layout"): no applicable method for 'layout' applied to
an object of class "c('gg', 'ggplot')"
```

By looking at figure 12 it can be said that a bar plot of the six directors and the corresponding maximum ratings of their episodes. We see that the highest rating of 13.61 goes to the episode that was shot by David Benioff.

Conclusion

After analyzing our datasets we came to the conclusion that there are many driving factors that are contributing to the success of the Game of Thrones such as the authors, directors, characters, battles and many more. The Analysis showed that the variables that we mainly concentrated on were very useful in the sense of understanding the datasets and by using the tools of data visualization we could find the connections and answers to our hypotheses to come to a final conclusion. Game of Thrones is a very popular TV series that has many reasons to be this successful which are the spectacular characters, talented authors, impressive battles and of course the outstanding storyline.

References

- [1] Taurino, G. (2019, July 31). An introduction to network visualization for television studies: models and practical applications. series.unibo.it. Retrieved December 4, 2022, from <https://series.unibo.it/article/view/8975/9569>
- [2] Scharl, A., & Hubmann-Haidvogel, A. (2016, January). Analyzing the public discourse on works of fiction – Detection and visualization of emotion in online coverage about HBO's Game of Thrones. sciencedirect.com. Retrieved December 4, 2022, from <https://www.sciencedirect.com/science/article/pii/S0306457315000278>
- [3] Gong, X. (2022). Analysis and Visualization of Streaming Media Platforms Based on the R Language——Take Netflix as An Example. drpress.org. Retrieved December 4, 2022, from <https://drpress.org/ojs/index.php/EHSS/article/view/2766>