COGS 108 - Final Project Proposal

Names

- Jared Andrews
- Pedro Enciso
- Sieng Ly
- Erik Mumm

Group Members IDs

- A13437274
- A13839368
- A15692666
- A13865505

Research Question

Are there underlying attributes of Game of Thrones episodes that can be used to predict the audience

Background and Prior Work (Pedro to fill)

Game of Thrones is a television show that has aired on HBO for eight seasons and it's arguably the m of viewers around the world tuning in to view each of its eight seasons. According to IMDb, an online and tv shows, Game of Thrones is currently rated at 9.3 out of 10 overall (1). However, some seasons than the others. So, what attributes lead an episode to being more or less highly rated?

This question is of interest to our group because we are all fans of Game of Thrones and we are curio critics mind when rating an episode. Game of Thrones has aired its final episode in 2019 and the resp According to Rotten Tomatoes, a movies and tv review website, the last Episode of Season 8, "The Iro (48%) in the history of the show (2).

Having watched the whole Game of Thrones series and reviewed the viewer ratings of each show, we included a major battle. The episodes "Battle of the Bastards," "Hardhome," "The Winds of Westeros," contained a major battle or act of violence between two groups of significant characters. These episo the series, all receiving a 9.9 rating (2).

This research is important because it can shed light on the variables within a script that have an affec movie. Future script writers and directors can also reference this project to maximize the viewer rating

There have been many projects done with the Game of Thrones, most likely due to the overwhelming data stored in the script. For example, Yish Lim conducted an analysis to find the correlation of the sc series vs the amount of views of the episodes they appear in. She found that the three characters with (0.9038), Missandei (0.888) and Tormund Giantsbane (0.8864) (3). These findings are interesting bec in violence and play supporting roles within the series.

Another interesting study analyzes the popularity of each major character using votes from Winterisc of the show can vote for their favorite character in a bracket style competition. Although this project d gives us a good idea of who the viewers' favorite characters are, which may have an influence on the r This project found that the character who won the most votes overall was Jon Snow, and the characte (average of each round they appear in) was Davos Seaworth (4).

References (include links):

- 1) https://www.imdb.com/title/tt0944947/?pf_rd_m=A2FGELUUNOQJNL&pf_rd_p=12230b0e-0e 8d5353703cce&pf_rd_r=R4EKGB7W14R2850YVCJ3&pf_rd_s=center-1&pf_rd_t=15506&pf_rd_i=t
- 2) https://www.rottentomatoes.com/tv/game_of_thrones/s08
- 3) https://medium.com/@yishuen/data-science-in-westeros-a784a624ba80

Hypothesis

Our hypothesis is that the appearance of certain characters, the amount of violence in the episode, an the episode are the main factors that influence the ratings of the episode.

Data

To determine which attributes can be used to predict the audience rating of the episode, we will use a describe each episode. In this dataset, we have 73 observations, which each one of the 73 episodes o variable types, that this dataset contains are: Season (int), Episode (int), Plot (string), Writers (list of st (string), Number of words in the Episode (int), Common words in Script (list of strings), Number of Pro (list of strings), Number of Deaths (int), Average Death Importance (float), Main Character Death (boo and Rating (float).

To create this dataset, our group used a combination of preexisting datasets, python libraries and wek was the Game of Thrones Rating dataset downloaded RatinGraph. This dataset contained Season, Ep

Number of words in the Episode, Commmon words in Script, Number of Profanities and Characters w Script All Seasons dataset from Kaggle was utilized. This dataset contained every line, and who said t Additional feature engineering was needed to create the variables in our dataset. To extract episode p library Wikipedia was used. This library wraps the MediaWiki API, allowing for Wikipedia data to be ea Deaths, Average Death Importance , Main Character Death and Most Used Killing Method, a dataset fr This dataset contained information regarding each death in the tv show (e.g., who was killed, who did importance, etc.).

This data will be stored in a csv and made avaliable in the data folder of our group's GitHub repository Additionally, the raw dataset used to create our final dataset will also be stored in our GitHub repo.

Ethics & Privacy

Our team acknowledges the issues related to the unethical use of data. Personal data can endanger t source is Kaggle, and is presented to users in an open-source format. Our team also scraped data fro Both of these scraping efforts were conducted solely on articles about Game of Thrones, a fictional sh wikipedia or contributors to the Washington Post. Our data contains solely data about a fictional TV s

Because our data is about a fictional TV series, there are no issues in equitable analysis. Additionally, script of each episode, there is no way our data can be excluding any characters or populations.

Team Expectations

- *Every member of our group is expected to contribute equally throughout the project.*
- *We all are expected to meet our individual tasks deadlines.*
- *Share ideas and datas with all team members.*
- *Follow the project timeline as much as possible.*
- *Every member of our group are expected to treat one another respectfully.*

Project Timeline Proposal

Meeting Date	Meeting Time	Completed Before Meeting	
4/20	5 PM	Read & Think about COGS 108 expectations; brainstorm topics/questions	Determine best form of
4/21	4 PM	Do background research on topic	Discuss ideal dataset(s
4/23	7 PM	Edit, finalize, and submit proposal; Search for datasets	Discuss Wrangling and
5/1	5 PM	Import & Wrangle Data	Review/Edit wrangling/
5/21	5 PM	Finalize wrangling/EDA; Begin Analysis	Discuss/edit Analysis;

Meeting Date	Meeting Time	Completed Before Meeting	
6/5	5 PM	Complete analysis; Draft results/conclusion/discussion	Discuss/edit full projec
6/10	Before 11:59 PM	NA	Turn in Final Project &