

SOCIAL WEB ANALYTICS

SOCIAL MEDIA SENTIMENT ANALYSIS

2023 March 27
Group 4

THE TEAM



**Gavriel Owens
Vincentio**



**Gargi
Asthana**



**Joshua
Benzon**



**Omkar
Shitole**



**Prakhar
Mishra**

OVERVIEW



HOUSE OF THE DRAGON

VS



RINGS OF POWER

PROJECT BACKGROUND

Data Scraping

Collecting data off of Twitter

Data Cleaning

Preparing the data for insights

Sentiment Analysis

Analysing sentiment behind each tweet

Visualisation

Visualising and summarising results

TOOLS USED

Python libraries:

- Snsccape
- Pandas + re
- VADER-Sentiment Analysis
- Matplotlib

DATA SCRAPING

SNSCRAPE

- JSON fetching using Twitter Public API
- Generates new random user agents to avoid blocking
- Includes cursor pagination and empty page fail safes
- Allows for almost limitless scraping without reaching API limits
- Can do other social media platforms as well

OUR METHODOLOGY

Tweet Attributes

- Datetime
- Tweet ID
- Username
- Text
- Reply Count
- Retweet Count
- Like Count
- Quote Count
- Mentions
- Hashtags
- Coords (if available)
- Link

Query Details

- Timeframe: Before, During, After Release
- Frequency: Monthly
- Language: English

DATA CLEANING

DATA CLEANING

- Libraries used for cleaning:
 - Pandas
 - re (Regular Expression)
- Necessary step before feeding data to VADER for sentiment analysis
- Notes on our Datasets
 - No missing or NaN values in the 'text' attribute
 - No tweet contained only @Mentions or #Hashtags

DATA CLEANING

- Steps followed for data cleaning:
 - Remove Duplicates and reset index
 - Remove any website links
 - Remove the '@' symbol for mentions and '#' symbol for hashtags
 - Feature Splitting 'datetime' feature (yyyy-mm-dd) into distinct features – Date, Month and Year

SENTIMENT ANALYSIS

SENTIMENT ANALYSIS WITH VADER

- VADER (Valence Aware Dictionary for Sentiment Reasoning)
- Used for text sentiment analysis that is sensitive to both polarity (positive/negative) and intensity (strength) of emotion.
- Uses human-centric approach using qualitative analysis and empirical validation.
- Originally all words have a score between -4 to 4. But the score returned is always between -1 and 1.
- Analyzes takes into consideration texts, lexical features and emoticons/emojis as well.

THE 5 HEURISTICS

Punctuations

If the score is positive, VADER adds a certain empirically-obtained quantity for every exclamation point (0.292) and question mark (0.18). If the score is negative, VADER subtracts.

Amazing

Amazing!!!

Capitalization

VADER takes this into account by incrementing or decrementing the sentiment score of the word by 0.733, depending on whether the word is positive or negative, respectively.

great work

GREAT work

Degree Modifiers

VADER maintains a booster dictionary which contains a set of boosters and dampeners.

kind of painful

very painful

Tri-gram

Negation is captured by multiplying the sentiment score of the sentiment-laden lexical feature by an empirically-determined value -0.74

Shift in Polarity using “but”

all sentiment-bearing words before the “but” have their valence reduced to 50% of their values, while those after the “but” increase to 150% of their values.

**we appreciate your efforts but,
the results are not acceptable.**

not good work

not bad work

ADJUSTING POLARITY SCORES

Polarity scores have been adjusted by the given formula below by attaching weights to the engagement.

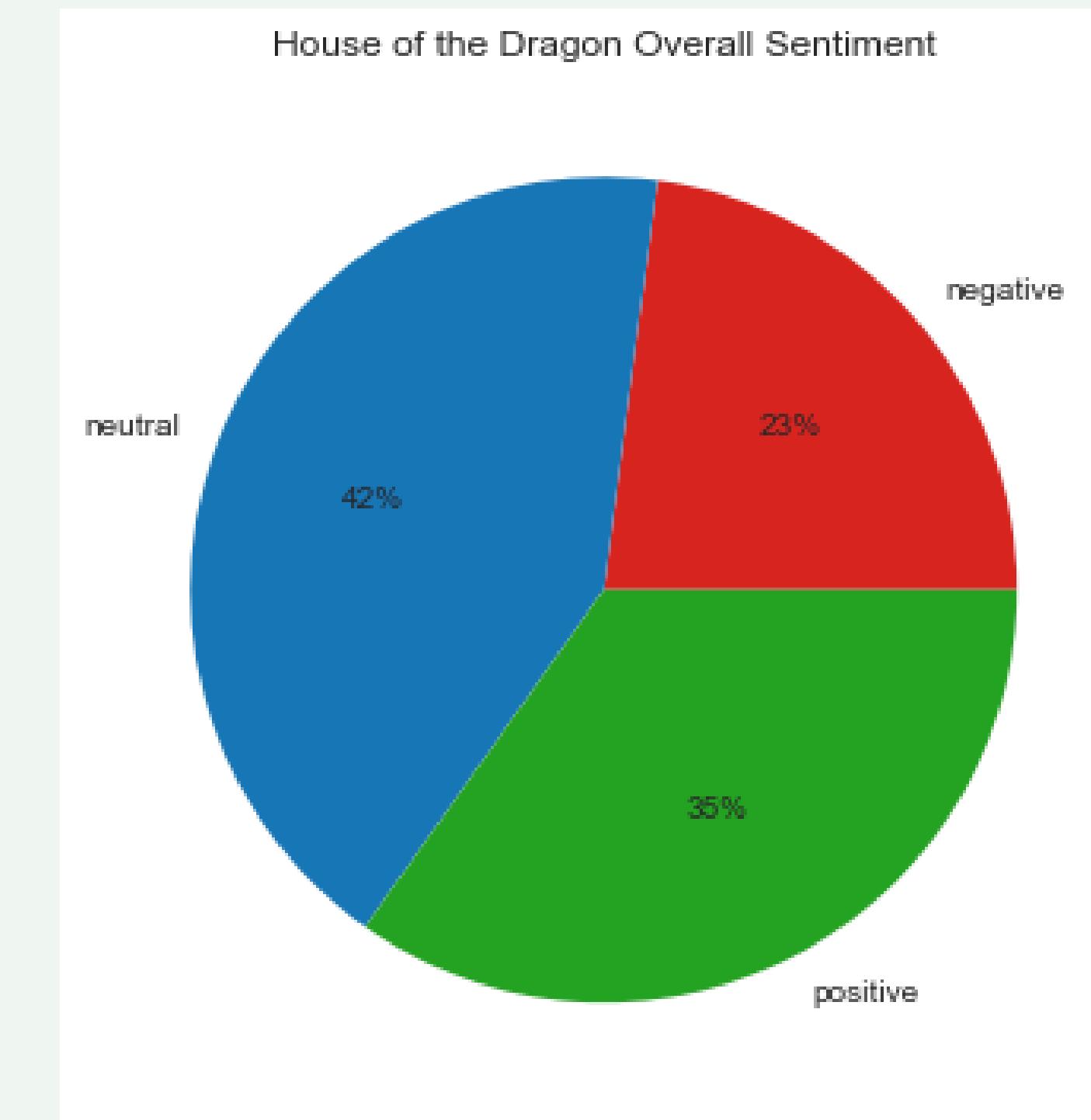
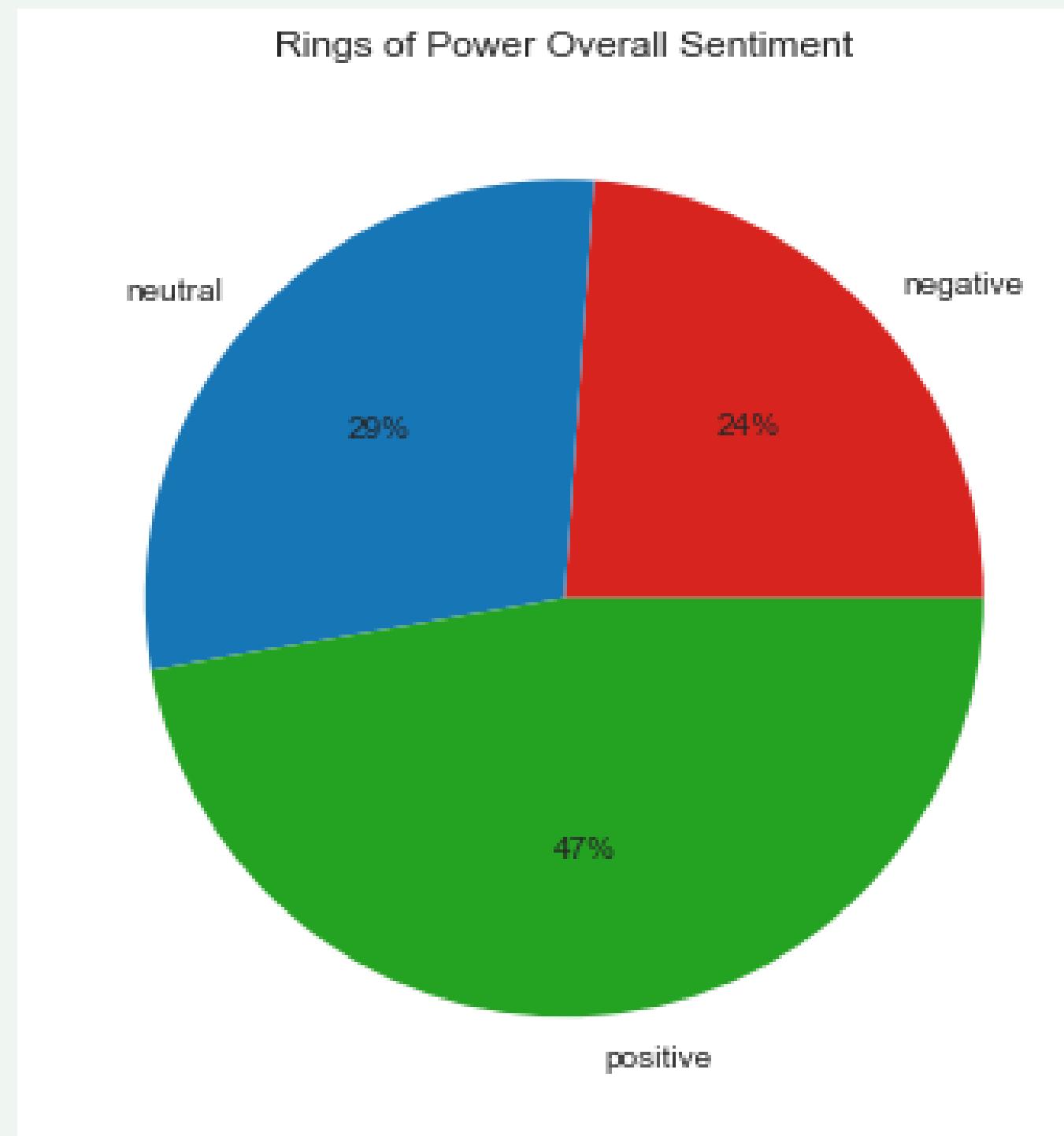
$$Adjusted\ Polarity = \frac{Polarity \times (Like\ Count + Retweet\ Count + 1)}{(Quote\ Count + Reply\ Count) * 0.1 + 1}$$

Polarity is directly proportional to like and retweet counts whereas inversely proportional to quote and reply counts (controversial reaction).

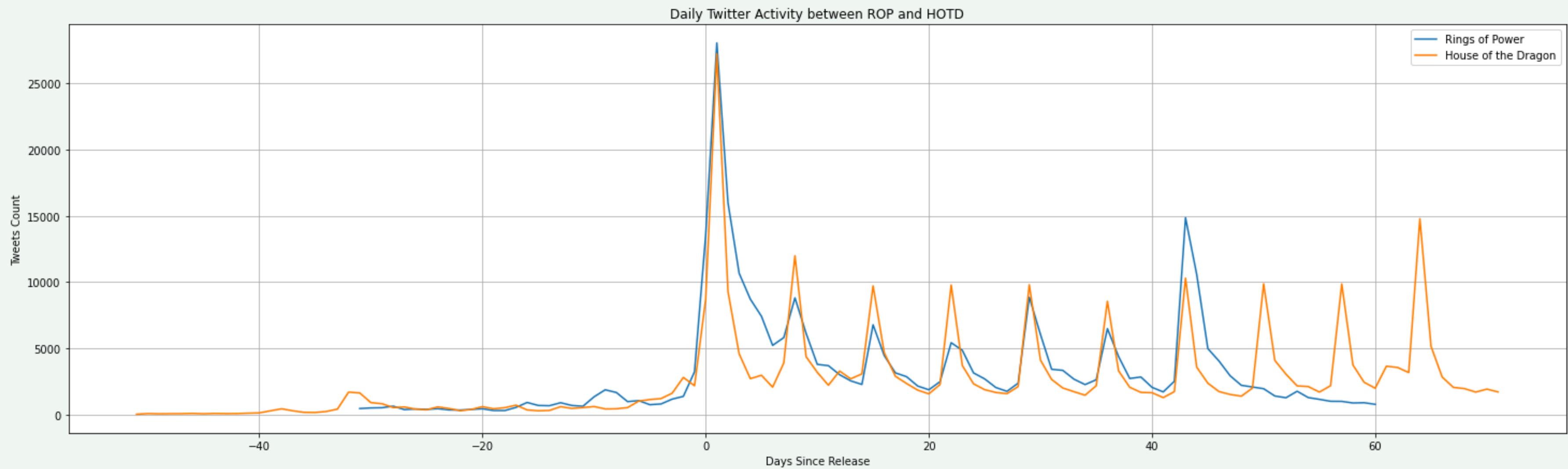
$$Adjusted\ Polarity\ Ratio = \frac{\sum |Positive\ Adjusted\ Compound\ Score|}{\sum |Adjusted\ Compound\ Score|}$$

RESULTS

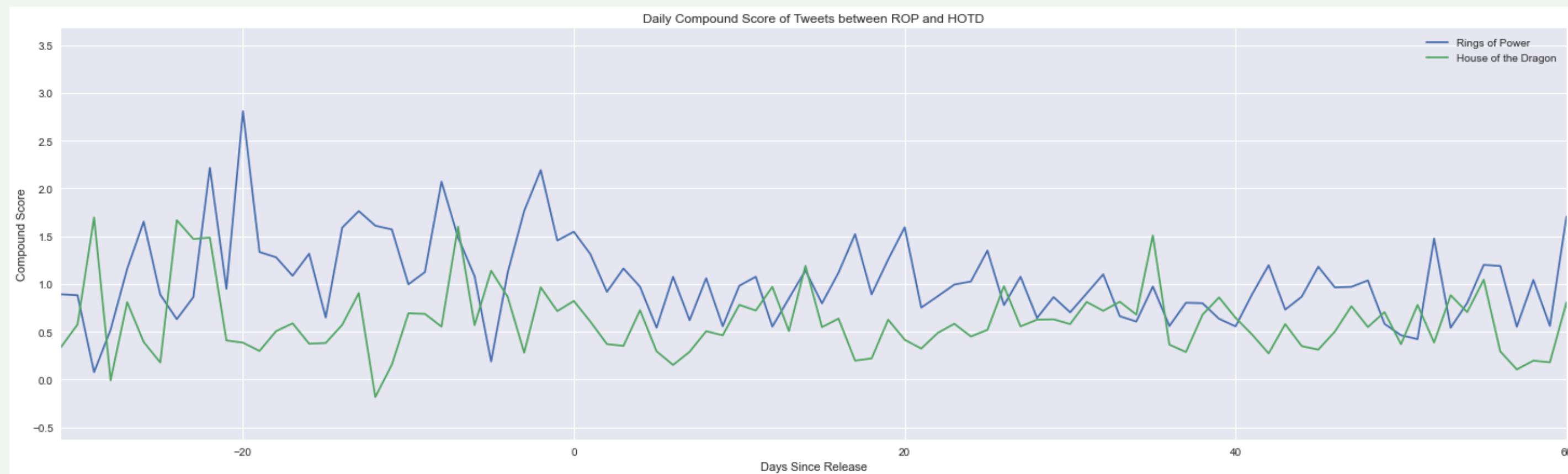
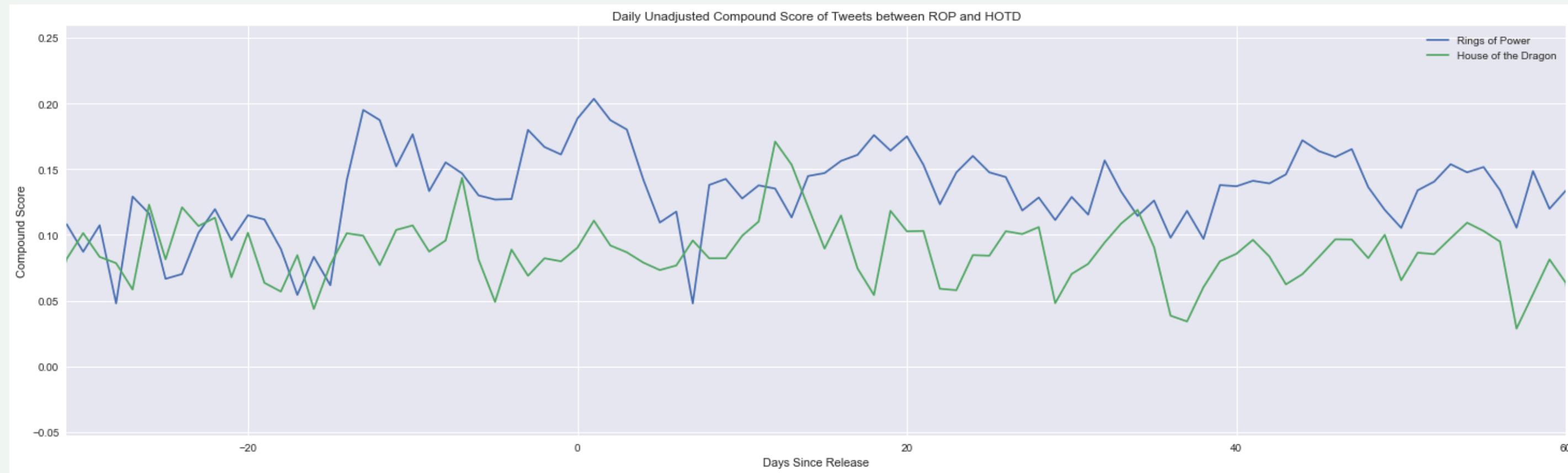
Overall Sentiment Ratio of Both Shows



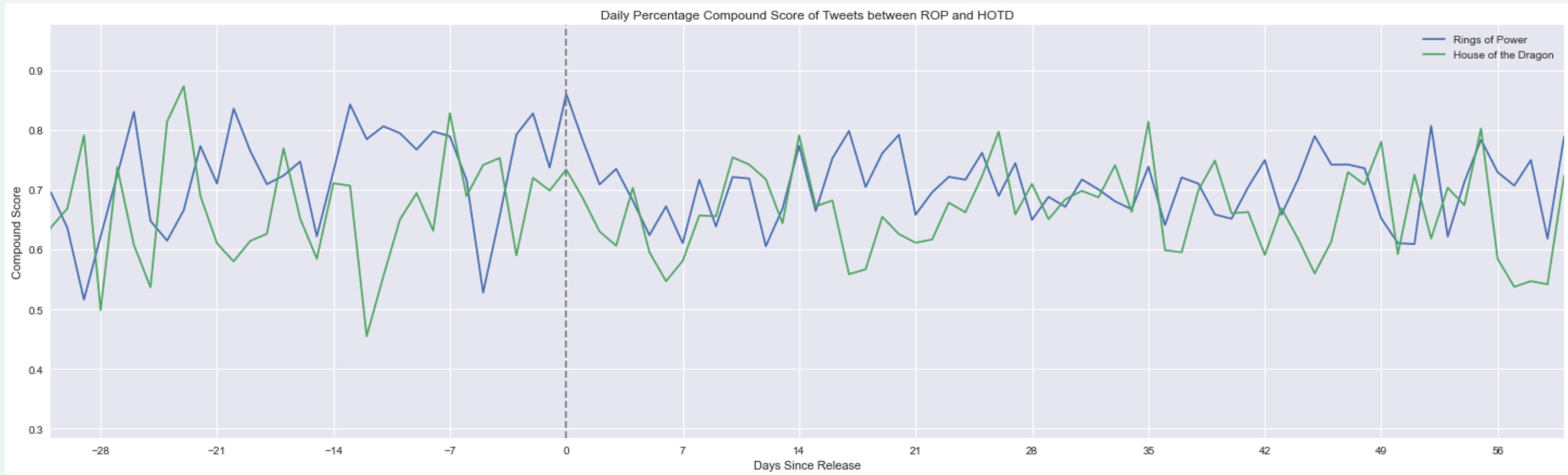
Twitter Activity (Counts) between Rings of Power and House of the Dragon



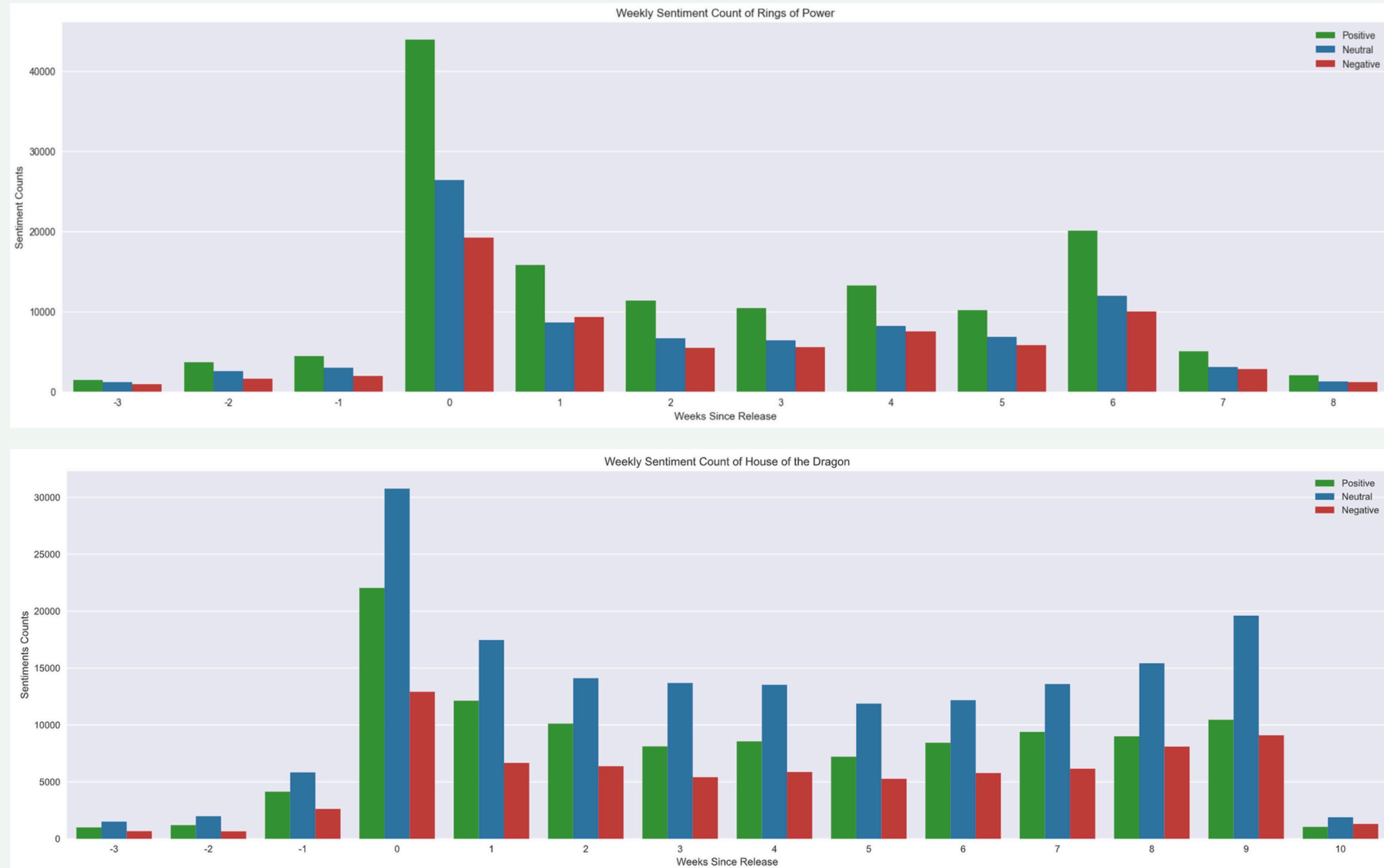
Before and After Adjustment of the Compound Score



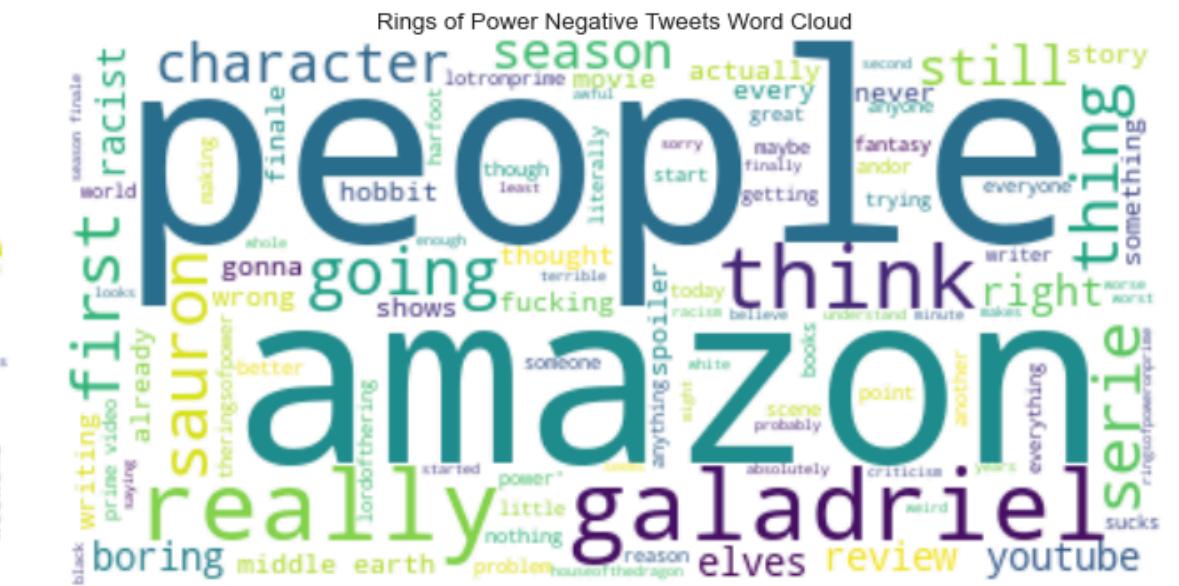
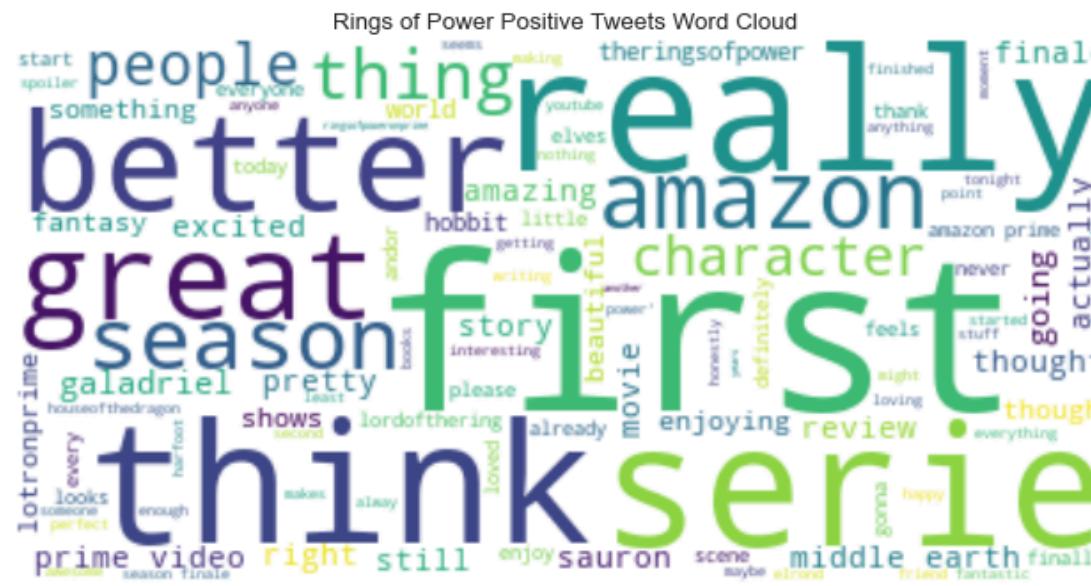
Positive Ratio of the Compound Score



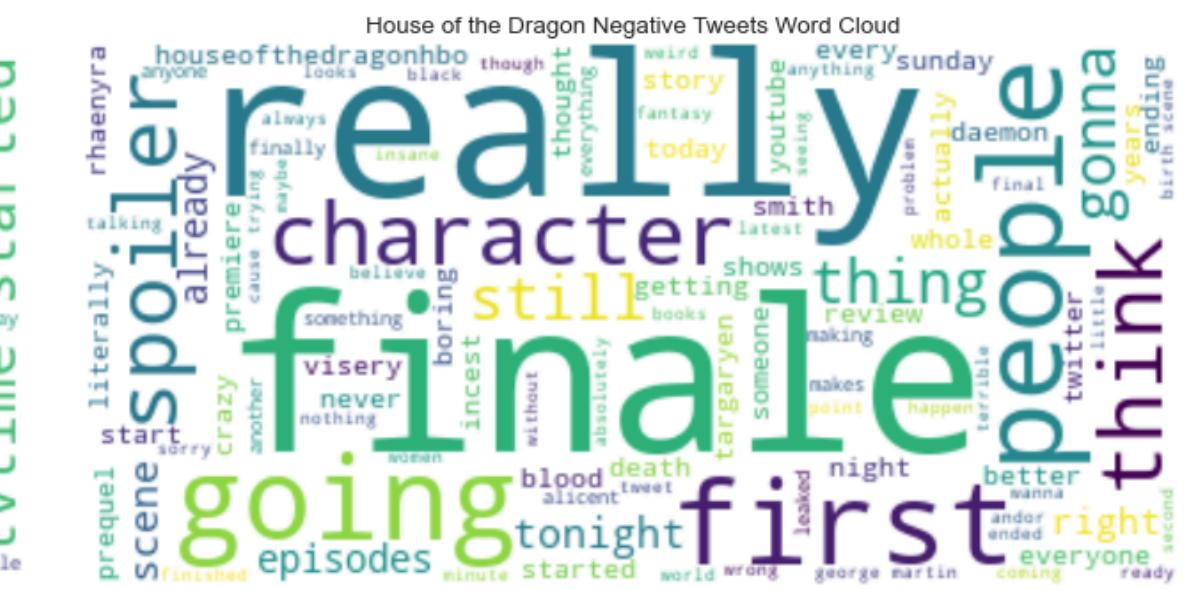
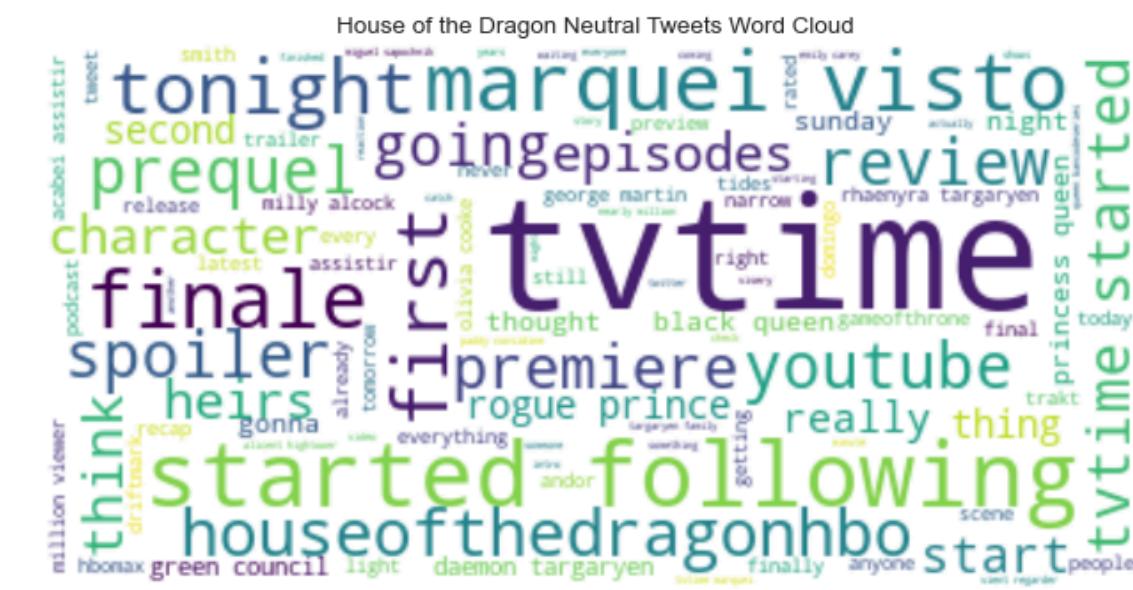
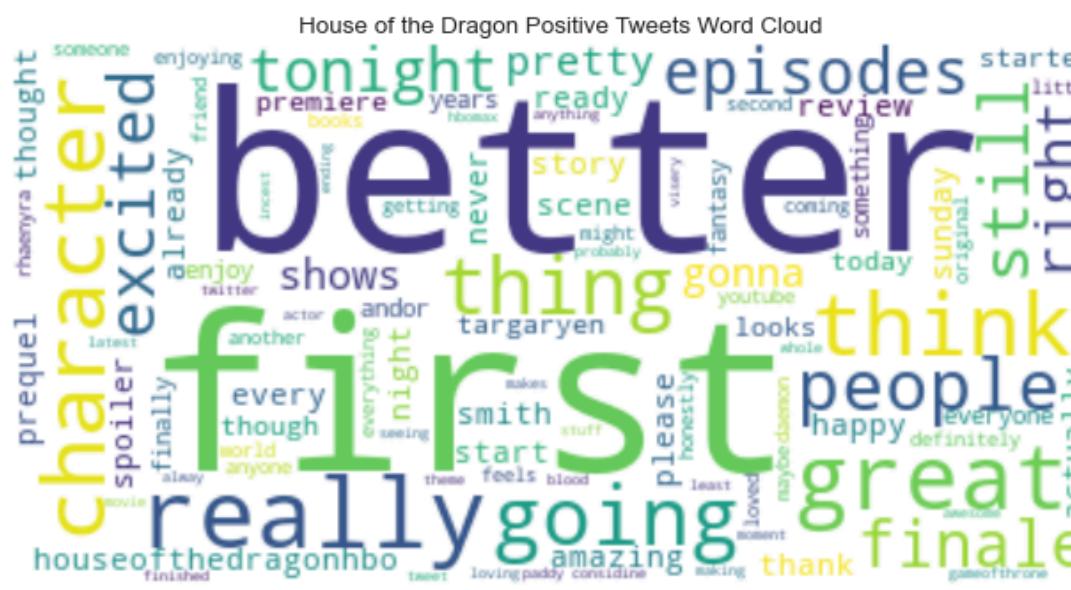
Weekly Sentiment Counts of Both Shows



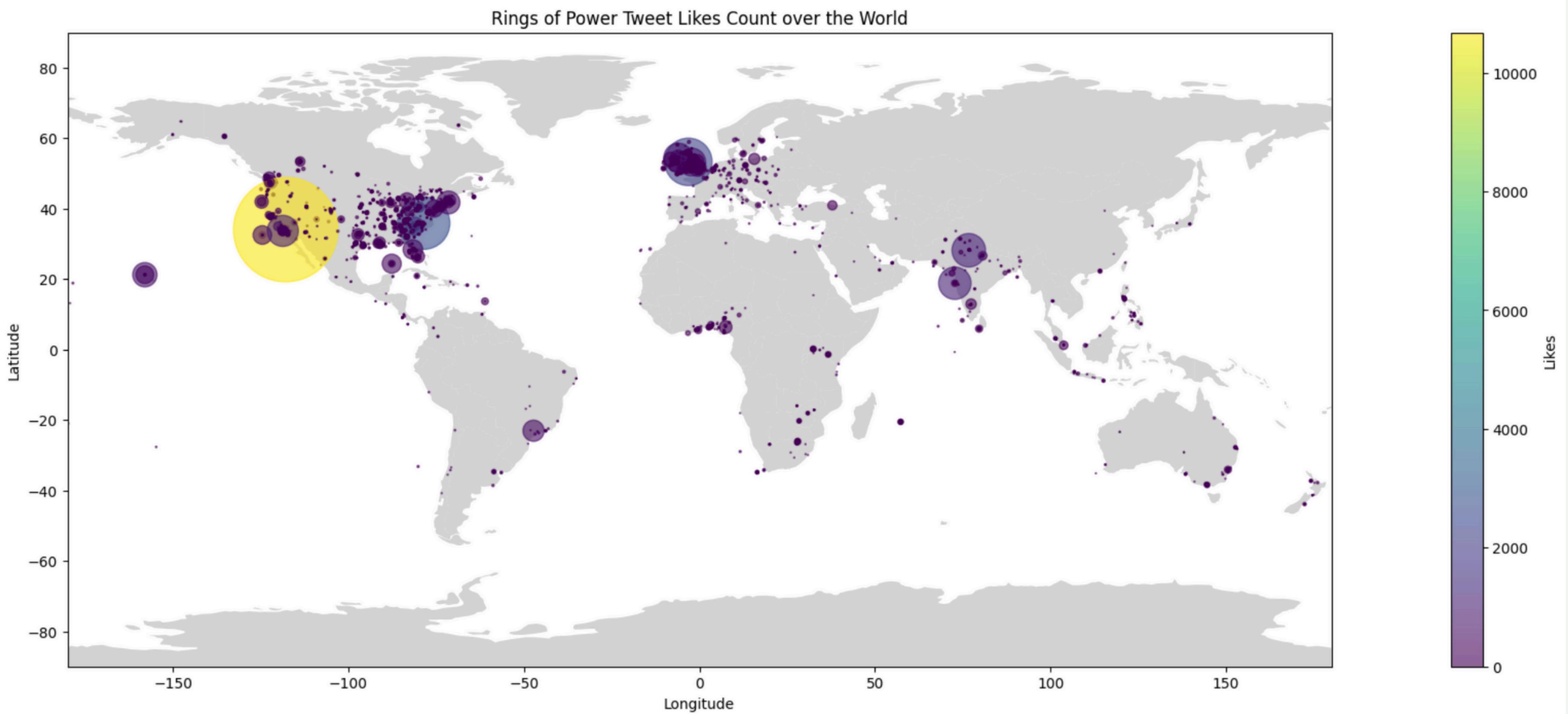
Word Cloud of Rings of Power



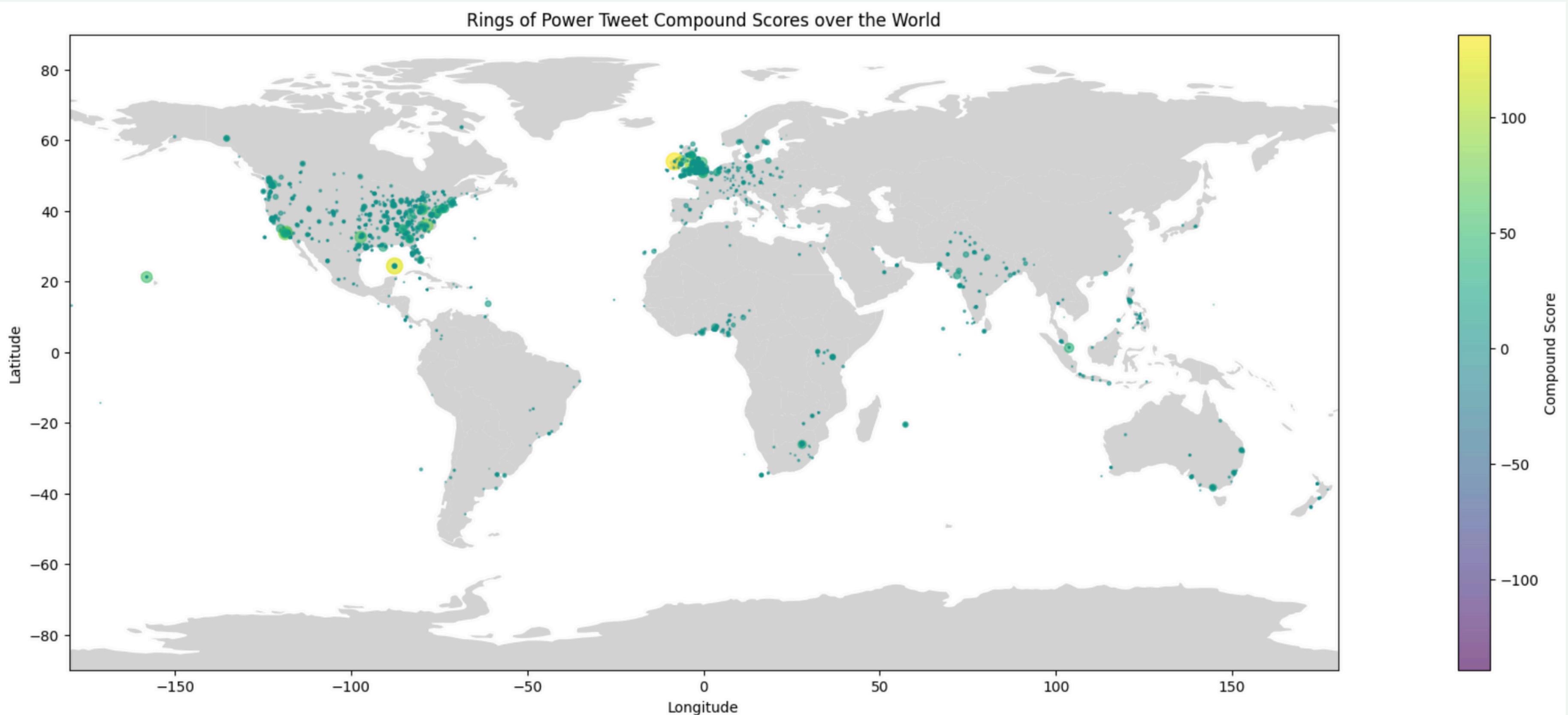
Word Cloud of House of the Dragon



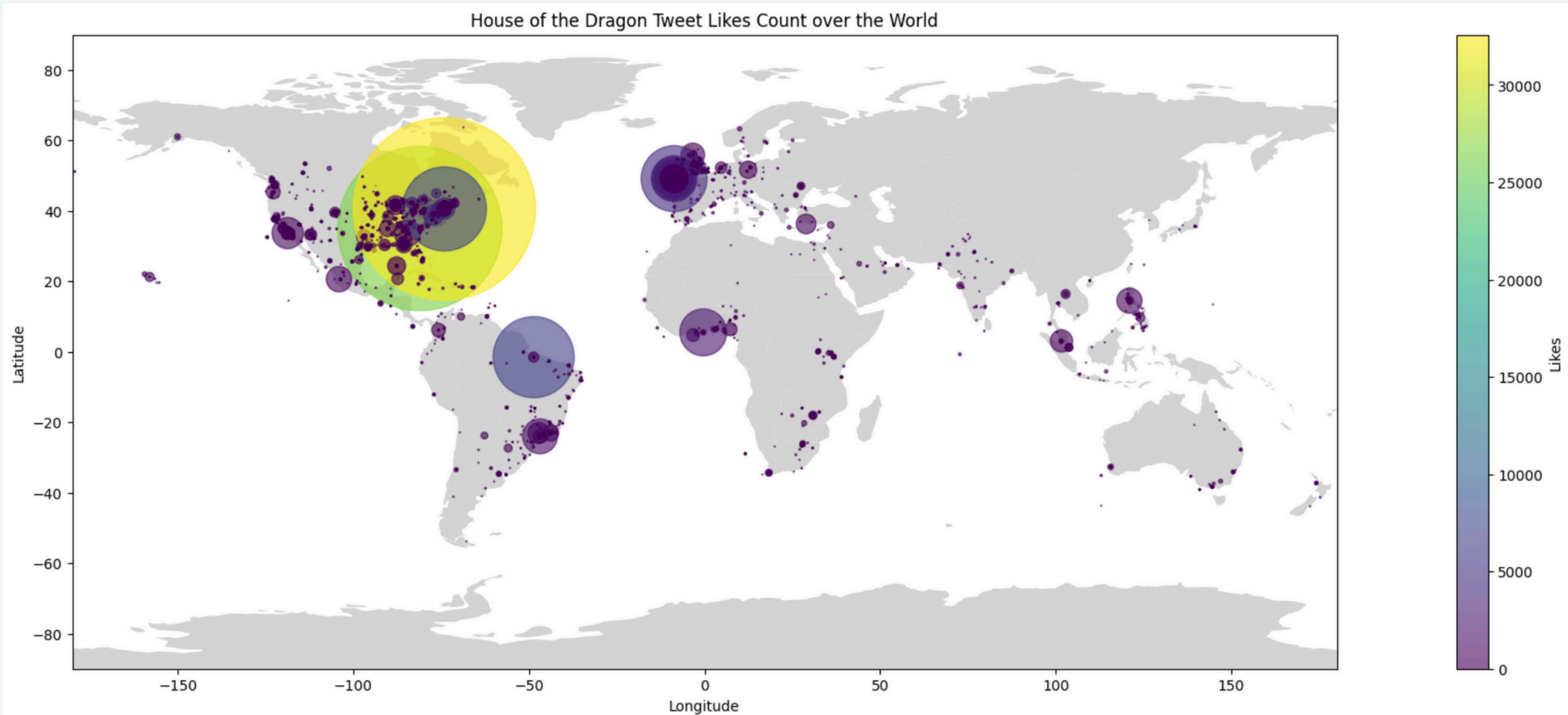
World Map of Tweet Likes Count of Rings of Power



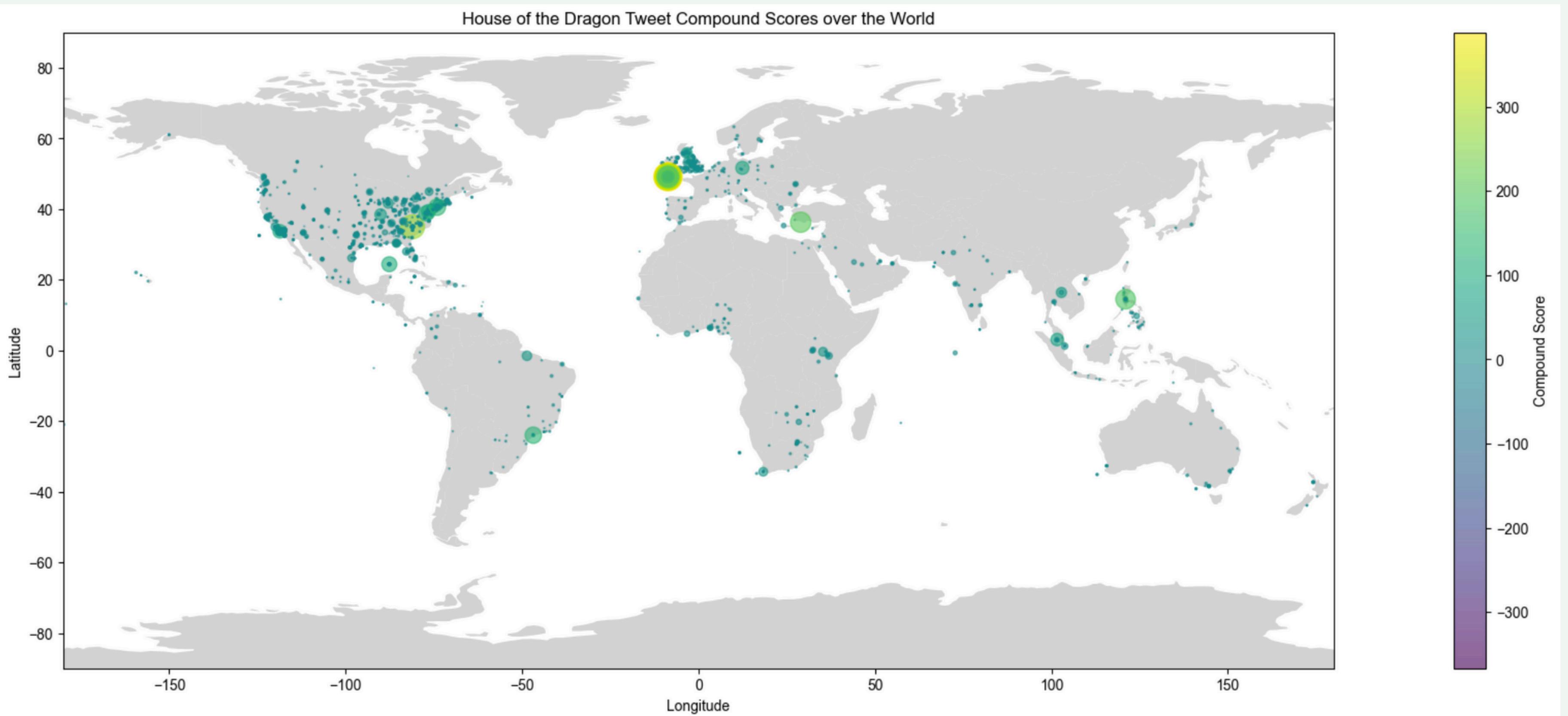
World Map of Tweet Compound Scores of Rings of Power



World Map of Tweet Likes Count of House of the Dragon



World Map of Tweet Compound Scores of House of the Dragon



CONCLUSION

Rotten Tomatoes Scores



SEASON 1

The Lord of the Rings: The Rings of Power



83%

TOMATOMETER

475 Reviews



38%

AUDIENCE SCORE

25,000+ Ratings



HOUSE OF THE DRAGON



93%

AVERAGE TOMATOMETER



82%

AVERAGE AUDIENCE SCORE

WHAT MIGHT THIS MEAN?

RINGS OF POWER'S OLDER CRITICAL AUDIENCE

Since Rings of Power comes from an **older franchise**, older fans might be more critical. These **older fans** may also not be on Twitter, as Twitter skews younger.

RINGS OF POWER'S SMALLER, MORE PASSIONATE FANBASE

Despite Rings of Power's **smaller audience** on Twitter, it seems to trend more **positively**. This might mean that RoP has a strong, niche fanbase.

HOUSE OF THE DRAGON'S WIDER AUDIENCE

As seen in the Twitter Activity graphs, **House of the Dragon** was consistently **more discussed** than **Rings of Power**, aside from RoP's finale.

AMERICAN EAST VS WEST COASTS

An odd discovery: **Rings of Power** tweets from the **west coast** (California, etc.) are significantly more liked than the east coast. **House of the Dragon** is more liked on the **east coast** than the west coast.

THANK YOU

2023 March 27
Group 4