QUESTIONS AND ANSWERS

**Ravens and Deaths (Analysis by Jason Lee)**

Hypothesis

Does the number of deaths (of major characters) or ravens sent per episode have any effect on episode sentiment?

Motivation

* Explore the ways in which the episode sentiment (Vader) had been affected by the occurrence of certain story developments.
  + **Ravens-** GoT uses ravens to communicate messages from character to character; for the writers, these plot devices serve as transition place-keepers to move the main story along. For the purposes of analysis, in the event a raven is sent, does this have an impact on overall episode sentiment (good/bad news)?
  + **Deaths-** GoT is infamous for its gratuitous violence, how will the deaths of the main characters affect the overall sentiment for each episode?
* Can we visualize the data to answer our questions above?

Summary

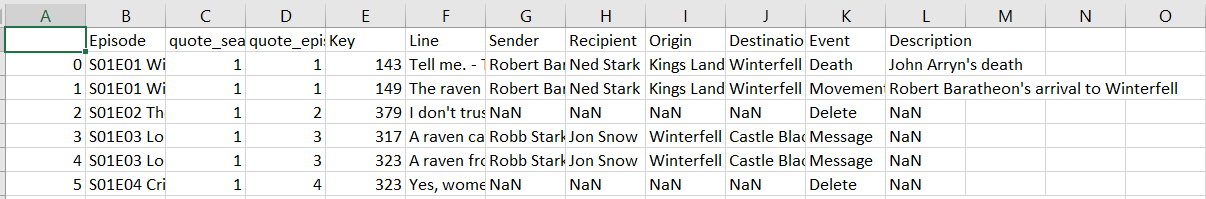
* Upon performing the analysis, I found that there were little to no relationships between the count of ravens/deaths in each episode to the overall sentiment of the episode script. The script contained a minimal amount of mentions for ‘ravens’ (compared to the books, one would assume) which made it difficult to establish a conclusion; but I did take note that in episodes with more than six high profile character deaths, the overall sentiment was always negative.
* Season six was particularly deadly for the main characters of GoT, especially the season finale.

Data Exploration & Cleaning

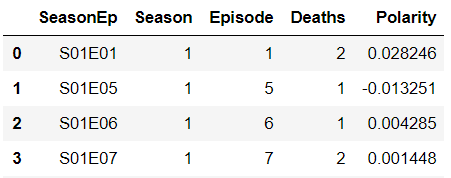
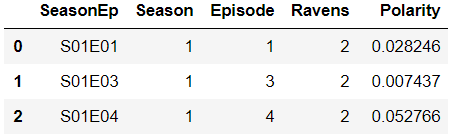
* The following datasets were used to perform my analysis:
  + <https://www.kaggle.com/gunnvant/game-of-thrones-srt> (JSON script for all 7 seasons)
  + <https://data.world/aendrew/game-of-thrones-deaths> (CSV file for major character deaths)
  + A Ravens CSV file I created with python that captured the keyword hits of (‘raven’,‘message’) along with the episode name.
  + Keyword searches through JSON files can only take you so far. The JSON script files don’t indicate which character speaks those lines.

Data Analysis

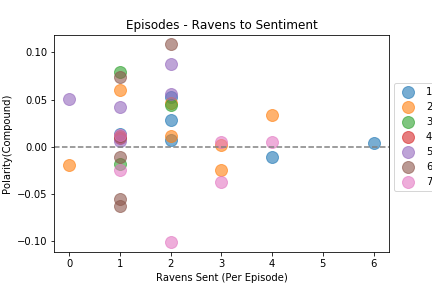
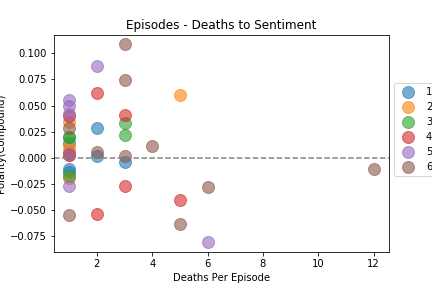
* In Pandas (Jupyter Notebook) I fed in three CSV datafiles:
  + Deaths - contains episode title/name, character name, manner of death, probability of return (if available)
  + Ravens- contains episode title/name, line, sender, recipient, origin, destination and message description.
  + Allseasons- contains episode title/name, sentiment means for each available episode (seven seasons)



* Creating Pandas DataFrame objects (for the Death/Ravens), grouping by episode, then merging the allseasons dataframe by episode title.



* Created two separate scatter plots for Deaths and Ravens, respectively, plotted against the episode sentiment. Plots are grouped by the season the episode was in and have their own color.



Discussion

* Although a trend between a count of ravens/deaths to episode sentiment was not found, the plots did visualize the data well.
* In episodes with more than six major character deaths, the overall sentiment was negative (to varying degrees)
* Season six was brutal and had an unprecedented amount of character deaths, the last episode contained deaths of several major characters thanks in part to Cersei Lannister, who created a ploy to bomb the citadel at Kings Landing (with several influential characters in attendance).

Post Mortem

* Cleaning the data for Ravens was tricky in part due to a lack of information about the characters who speak the line. (Although it would be interesting to create an algorithm to do this for you)
* If I could do a spider-web plot for all the messages (ravens) sent out, it would be easier to visualize the communication network that takes place in Westeros
* Analyzing the deaths of major characters by manner of death could be another possibility to explore.

**Jules’ information**

## Clean up

* Exploration was greatly aided by Kevin who found the scripts. The cleanup process was relatively clean and simple.
* Some insights I had were realizing how much information a key’s value in a dictionary can hold. I assigned entire sentences to them and they were easy to access.
* I found that paying money can get you access to older or more relevant/obscure data. Tweets retrieval was limited by how old the tweets were.
* Many helpful guides exist to help you get APIs and other data. Don’t reinvent the wheel!

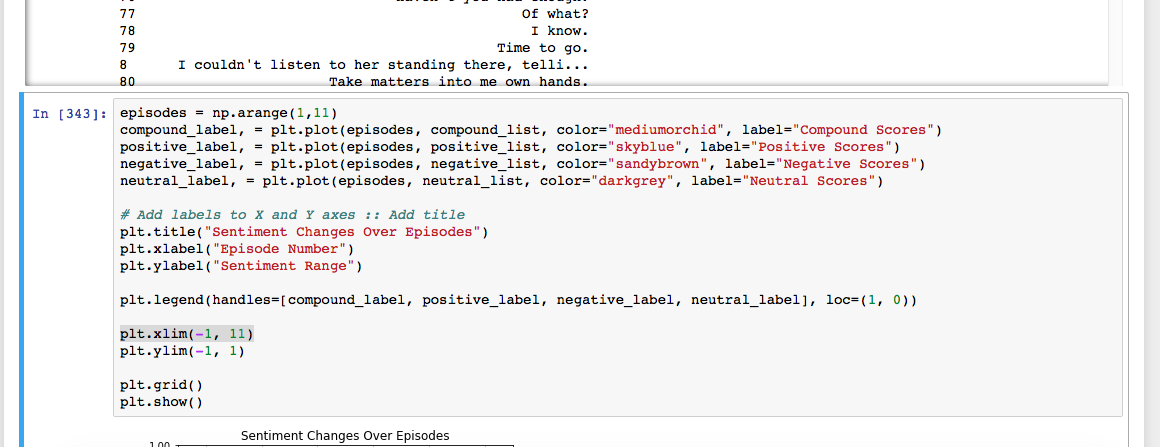
## Questions and Data

* What were the VADER sentiment analysis scores for each season? For me, particularly season four.
  + There seemed to be no better way to get an objective view of the seasons than to use VADER analysis. Also, as someone who does not watch GOT this would give me a window into the sentiment of the show without having to watch it. A downside to relying on VADER might be the use of irony in the scripts or words from languages specific to GOT
* Did these vary wildly?
  + For a show that is provocative and violent, the sentiment analysis did not vary widely from episode to episode, particuarly in season four. It seemed to maintain a uniform dark nature which is no doubt what fans expect. Certain compound scores dropped on particular episodes.
* Did they have any correlation with viewers’ sentiment, which was explored via tweets?
  + Getting tweets was more difficult as older tweets were behind a paywall.

## Data analysis

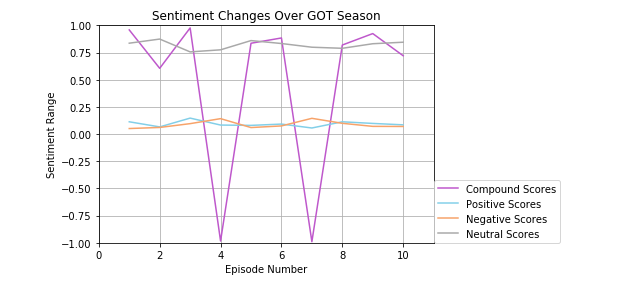
* For Season 4, I imported the season’s JSON file and put the episodes in a list. I then looped through these episodes and ran VADER analysis on each. I placed their sentiment scores into Compound, Positive, Negative, and Neutral lists.
* I began to create a line graph which would visualize the suspected ups and downs in sentiment over the course of season four. I created a plot, mapped the scores, added titles and labels, and adjusted the grid. All group members did so for various seasons.





## Discussion

* The information I found about Game of Thrones was all new to me. I don’t watch the show, so I had no idea what the sentiment was during certain episodes and seasons.
* However, I did know that it was a dark show that people reacted strongly to, so I suspected the sentiment would be MORE negative than it was. This assumption seemed to be incorrect.
* VADER of the episode scripts don’t tell me what viewers think, but I was able to get a feel for the sentiment of the episode. VADER analysis can be useful if you’re working with a subject you’re unfamiliar about. In our scenario where we’re selling ad space, you want to be able to tailor your ads regardless of whether or not you’re a hardcore fan or brand new to GOT.
* These findings showed that, for example in season four, the compound score varied widely.
* Compound scores pretty accurately reflected the average of the scores, ie. when episodes were highly positive or negative, the compound score reflects this
* Season four in its entirety didn’t seem to vary much, there were no large fluctuations in the positive, neutral, or negative scores



## Post mortem

* The difficulties that arose were finding tweets, finding information was difficult and you needed a premium membership to get some older tweets. What was good was that there were other people already interested in this pop culture topic, so APIs were easy to find.
* Other topics if we had more time: We started our project thinking we could focus on another topic and crime, and someone thought it’d be fun to track and analyze the crimes in GOT. That would be very entertaining to research.

Kevin’s Information

**What is the relationship between tweet sentiment and episode sentiment?**

As it turns out, twitter only keeps tweets on it’s user\_timeline api back to December 2016, or about 15-16 months, and the search api only returns results up to a week prior. Additionally, I was unable to find a way to filter through the api based on date outside of the since\_id and max\_id parameters to use another tweet’s id to set a time parameter. I had hoped to use the official Game of Thrones twitter to locate tweets at the start and finish of each episode, but even then the official twitter was inconsistent in tweeting out at the start and/or finish of each episode.

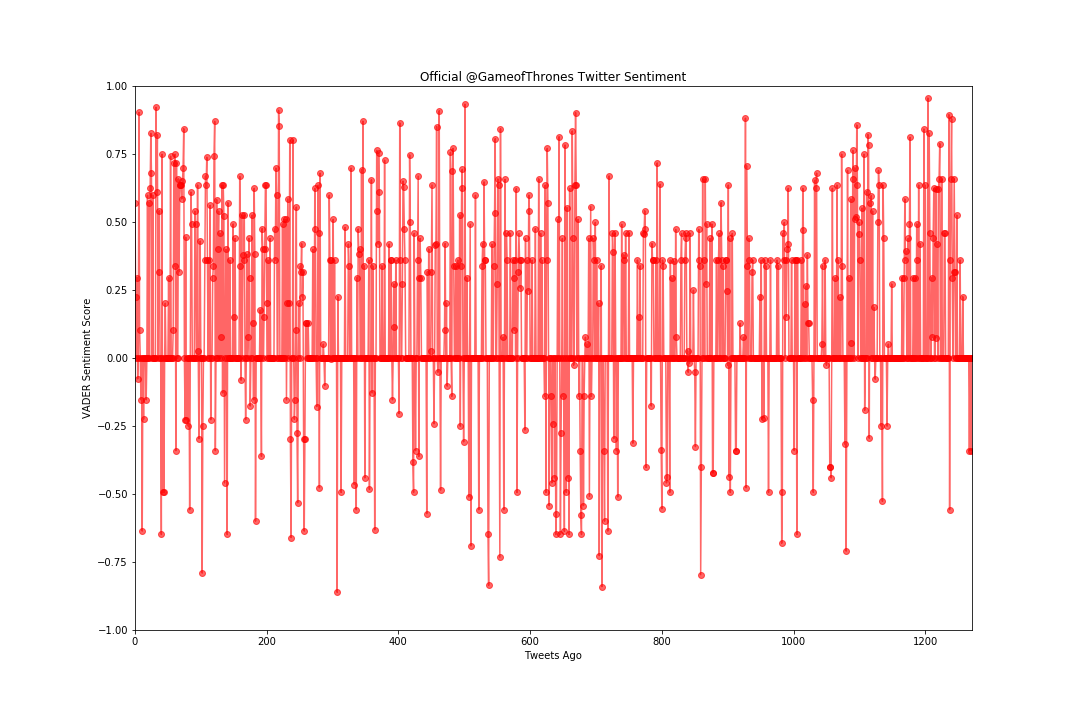


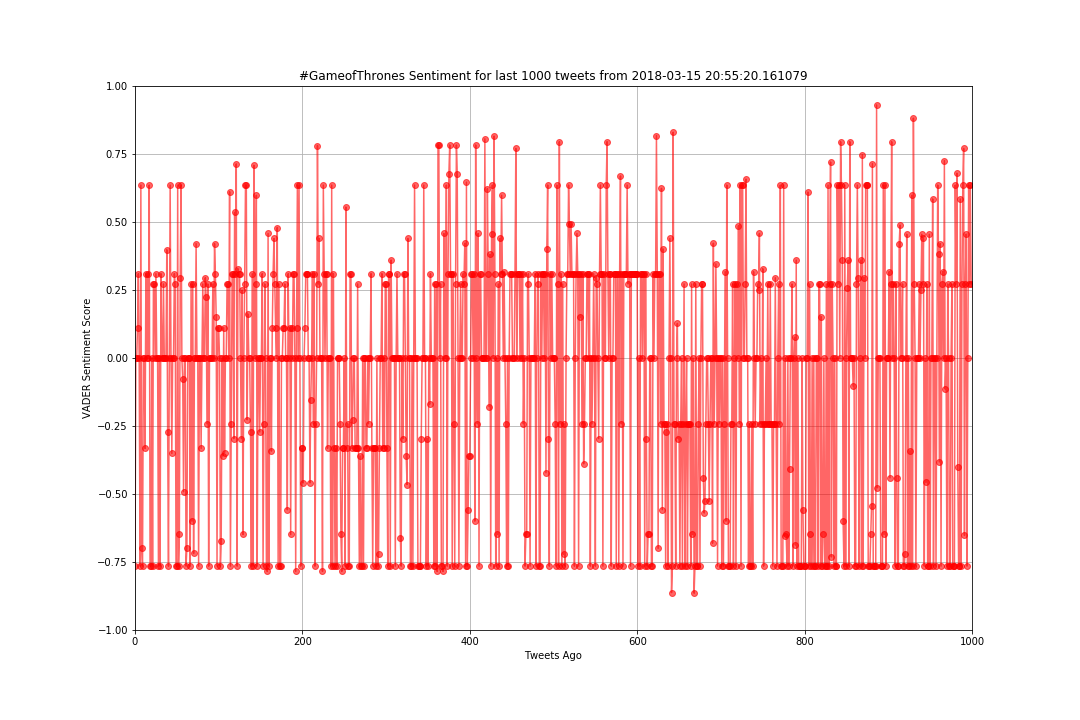
**What can I do?**

I have a massive file of all of the twitter data available on GameofThrones official account from the api. What kind of information can I find in here?

**How does the sentiment compare between GameofThrones official account and it’s fans?**

I first ran a sentiment analysis on all of the tweets I had pulled, and graphed it, revealing the top graph on the following page. Then I searched for #GameofThrones and plotted all of that sentiment, revealing the second graph. Comparing the two, it is clear that the official account is far more positive than the general tweeting public. In the future, it would be interesting to compare positive sentiments between varying official accounts, and see if different industries, genres, or other factors contribute to general twitter sentiment used for marketing.





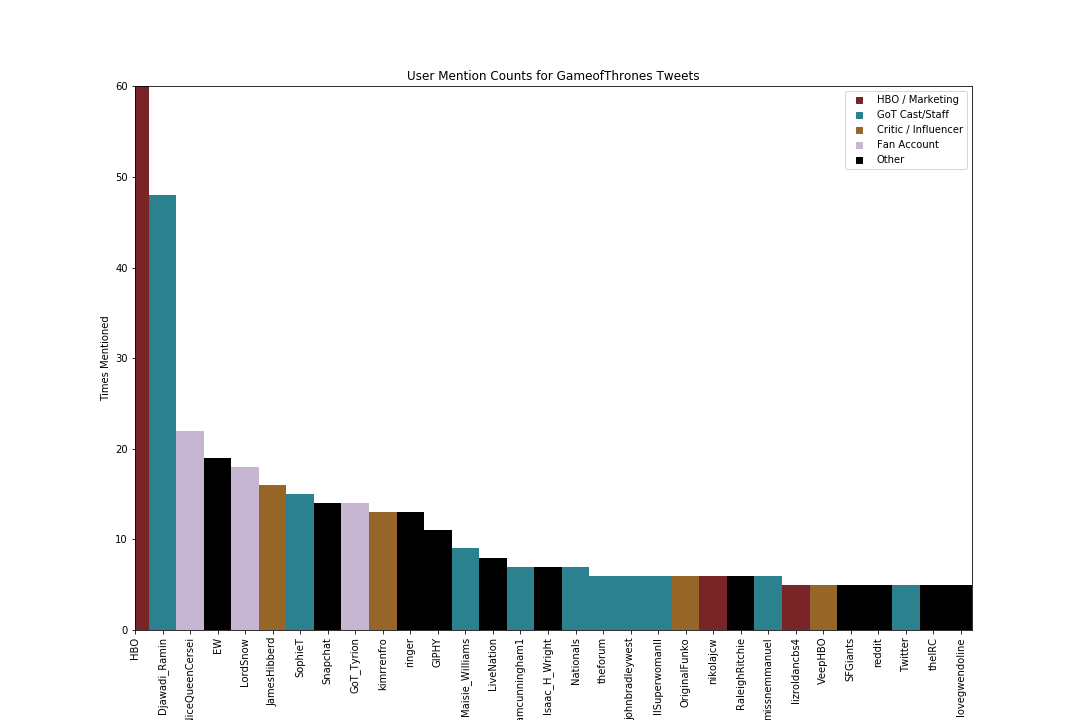
**What does GameofThrones’ twitter network look like?**

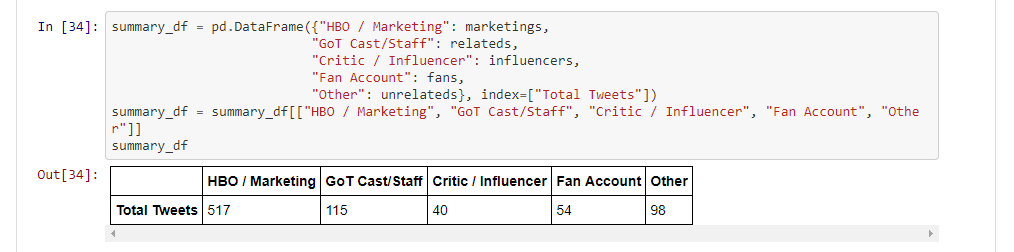
With the twitter information of course came any other user mentions, so I thought to count each users mentions and see if there are any patterns or curiosities. Interestingly enough, @GameofThrones mentioned itself 338 times in 1271 tweets. In presenting this, I decided to leave off itself and cut off HBO, the channel it is on, as those mentions are par for the course, but there is still some interest that HBO has a massive number of mentions so I left some of it in. I divided twitter accounts into five categories:

* HBO / Marketing - Twitter accounts for GameofThrones, HBO, other HBO shows, or companies that sell Game of Thrones products.
* GoT Cast/Staff - Game of Thrones staff twitter accounts, be they actors, actresses, composers, or anything else.
* Critic / Influencer - I put newspaper / internet critics and youtube influencer accounts together because they seem to fulfill a similar purpose to me: non Game of Thrones sources telling you how you should be watching this show.
* Fan Account - Fake accounts run by fans playing at being characters in the show. Interacting with the fan base through these, approving of their fan-fictions and fantasies is a great way to build a community.
* Other - Unrelated accounts, such as sports teams, music venues, GIPHY, and reddit. Could be that the service was used or being promoted by Game of Thrones.

We can see that, of course, marketing or HBO related tweets come in massively number one, followed by cast, fans, and influencers. But also of interest are the number of accounts associated with these mention numbers:

* HBO / Marketing: 4 Accounts, 517 Tweets, 100+TPA (tweets per account)
* GOT Cast/Staff: 10 Accounts, 115 Tweets, 11.5 TPA
* Critic / Influencer: 4 Accounts, 40 Tweets, 10 TPA
* Fan Account: 3 Accounts, 54 Tweets, 18 TPA
* Other: 10 Accounts, 96 Tweets, 9.6 TPA





Nikki’s Information

Questions and Data

The questions: Which episodes tended to do better? Building off of that, which season is the most popular season?

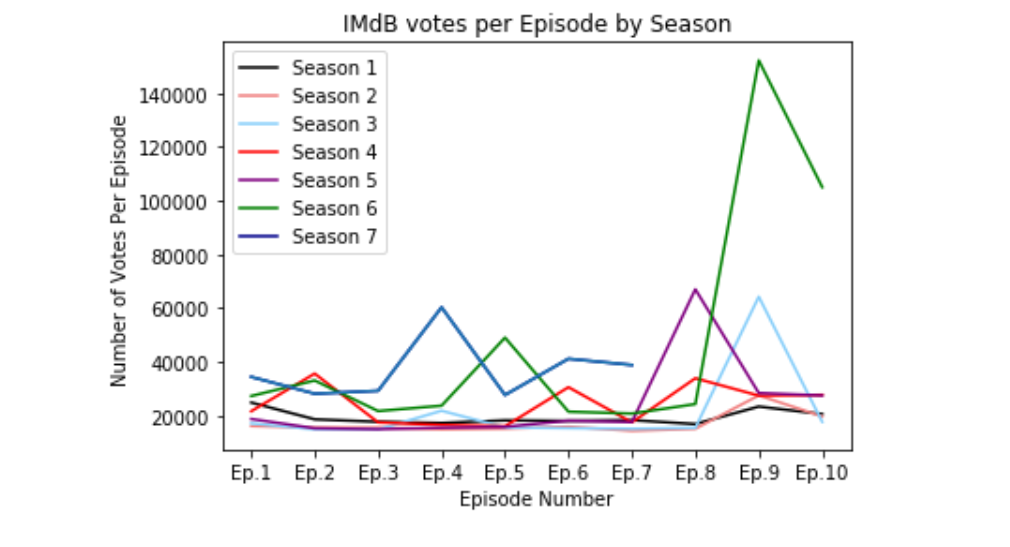
To retrieve this data, we pulled from the imdb api and extracted the rating score and the number of votes. Then with this data, we created a dataframe for each season and then plotted the graphs.

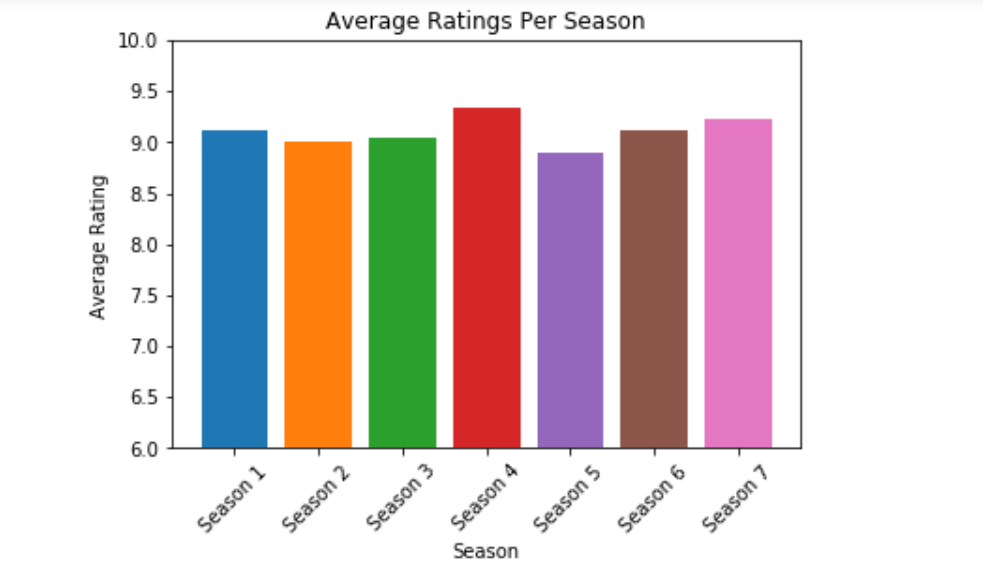
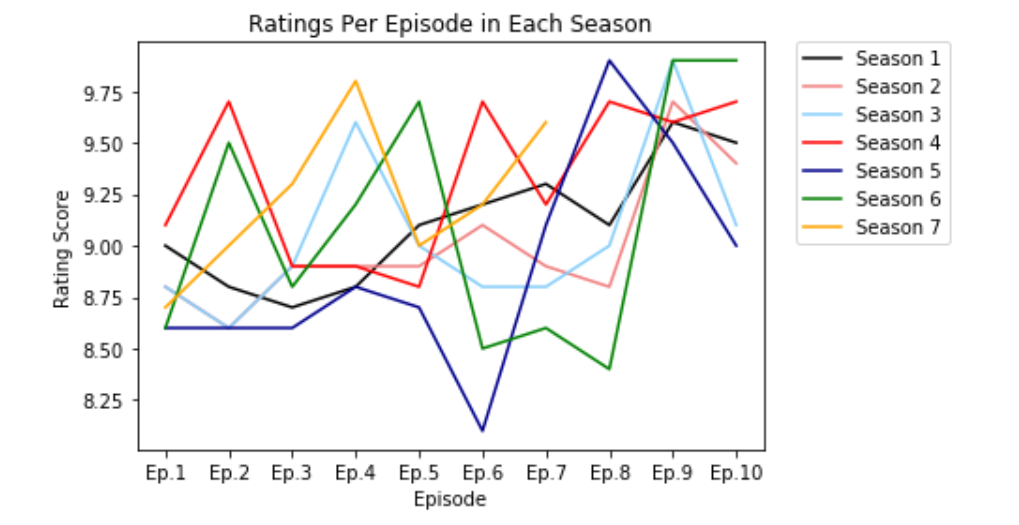
Clean up Exploration

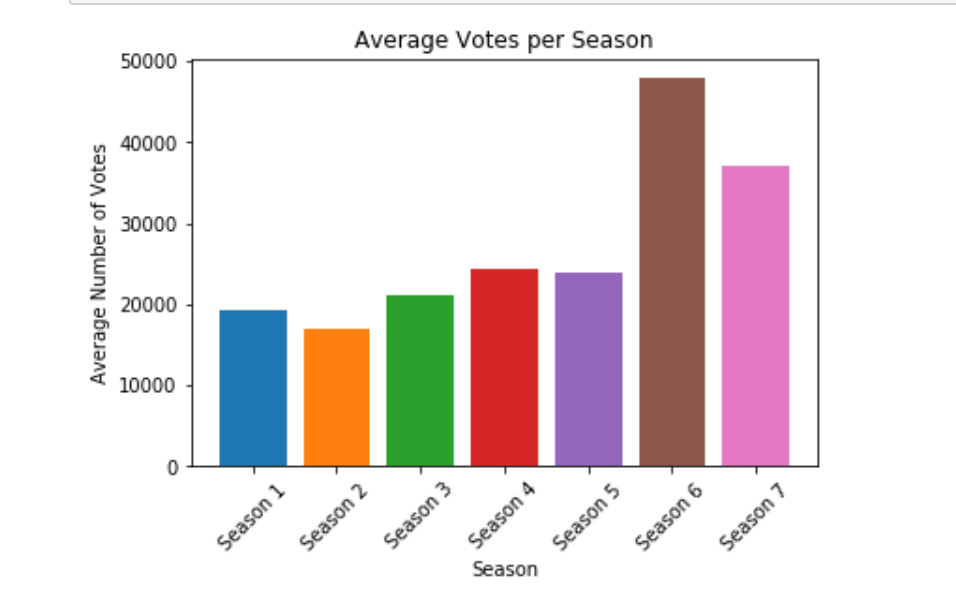
Some hiccups, that were encountered initially was converting the string values into integers or floating numbers so that the y axis of the plots are actually arranged in order.

Discussion

The primary purpose to analyze imdb reviews was to understand which seasons and episodes had the highest ratings. Another question that followed was which season was the most popular season. To answer these questions, we pulled data from the imdb api. In the imdb data we pulled the ratings per episode in each season as well as the number of votes that each episode received. Finally, to find out the most popular season, we analyzed the average ratings per season as well as the average amount of votes overall. Based on the data, most ratings and votes tend to increase towards the end of the season. Overall, season 4 was the most popular season by rating but season 6 had the highest votes overall. Which then leaves the question, which one is actually the most popular season? When looking into the the votes of both season 4 and season 6, they were pretty close. But the last two episodes of season 6 received higher votes.







Post Mortem:

If given more time, I would have loved to look into other review websites for comparison because there are more factors that go into determine which season is the best season. Also, Game of Thrones is still not over but it would also be interesting to see the ratings and votes throughout all of the seasons and see if there is a similar trend as the analysis of episodes.