GAME OF THRONES

City University of New York, Baruch College, Zicklin School of Business

Applied Natural Language Processing

CIS/STA 9665

Group 9

Project Report

The Beginner's Guide to Game of Thrones

Team Members

Sanket Patel	Sanket.Patel1@baruchmail.cuny.edu
Jingnan Qi	Jingnan.Qi@baruchmail.cuny.edu
Wei Chi Yeh	Weichi.Yeh@baruchmail.cuny.edu
Oje Williams	Oje.Williams@baruchmail.cuny.edu
Jack Yang	Jack.Yang@baruchmail.cuny.edu
Yichuan Hu	Yichuan.Hu@baruchmail.cuny.edu

GAME OF THRONES

Introduction and background

Game of Thrones, also known as GOT, is a fantasy drama television series that contains 73 episodes in 8 seasons. The series had won 62 awards and broken 10 Guinness World Records. According to Home Box Office (HBO), an American pay television network, Game of Thrones had notched a series record of 17.8 million viewers. This fantasy drama television series dominated the decade not because of wars and violence, instead, it was a triumph in storytelling of characters, plot twists, dialogues, and emotions.

The Beginner's Guide to Game of Thrones aims to help its users gain deeper insights of the stories and each main character using script analysis. By analyzing the GOT script of the main characters from all seasons, we aim to explore the following research topics: the change in main characters' sentiment, the aspect that reveals from the frequently used words in each family, and the comparison of the word usage of the two main characters. The goal of the research topics is to help the users understand the impact of the main characters and the main families. Utilizing sentiment analysis and aspect mining, we can provide better and deeper insights to the audience and professionals such as television producer or director, screenwriter, and actors. For those who have not watched the show, it will act as a beginner's guide to Game of Thrones.

Research Motivations

Characters are the heart of a novel, according to Roz Morris, a bestselling ghostwriter turned contemporary novelist and author mentor. Every story needs a plot, however, characters are critical to a plot because they bring a plot to life. A story becomes a great work of art when it is emotionally involved with the characters. This motivates us to use Sentiment Analysis to better understand the main characters in Game of Thrones.

According to ScienceDirect, a website that hosts over 18 million pieces of content from more than 4,000 academic journals, family influences are readily evident in a person's acquisition of specific skills and knowledge about the world. Therefore, family makes a great influence on a person's development of emotion and behavior. This inspires us to explore the main families in Game of Thrones.

Many ask, "who is the better queen, Cersei, or Daenerys?" The main premise of Game of Thrones is the fight for the coveted Iron Throne. Many kings and queens want to sit on it, but no one fights for it harder than Cersei Lannister and Daenerys Targaryen. In Matthew 5:7, it stated "Blessed are the merciful, for they shall obtain mercy." Also, according to "The Art and Science of Applied Leadership", by Dag Heward-Mills, a leader is someone who knows the importance of showing mercy. These led us to wonder who is the better queen: Cersei or Daenerys.

Text Dataset Description

The text dataset is a set of GOT for all seasons in a form of a table containing 6 columns with different data types used for various purposes.

The GOT script dataset is generated through scrapping the URLs provided on Genius.com for the Game of Thrones series and organized into csv files on Kaggle. Data columns contain release date, season number, episode number, title of each episode, name of a character in Game of Thrones and sentences spoken in the series with 22,300 unique values.

Text Preprocessing Description

The text was cleaned first by breaking up the strings into words and producing a list using tokenizers. Then we got rid of the punctuations and numbers. After that, the words are normalized to lowercase to ignore the case distinction. We also filtered out the stopwords so that more focus can be given to the remaining meaningful words. The last step was trying Porter stemmer to strip off any affixes.

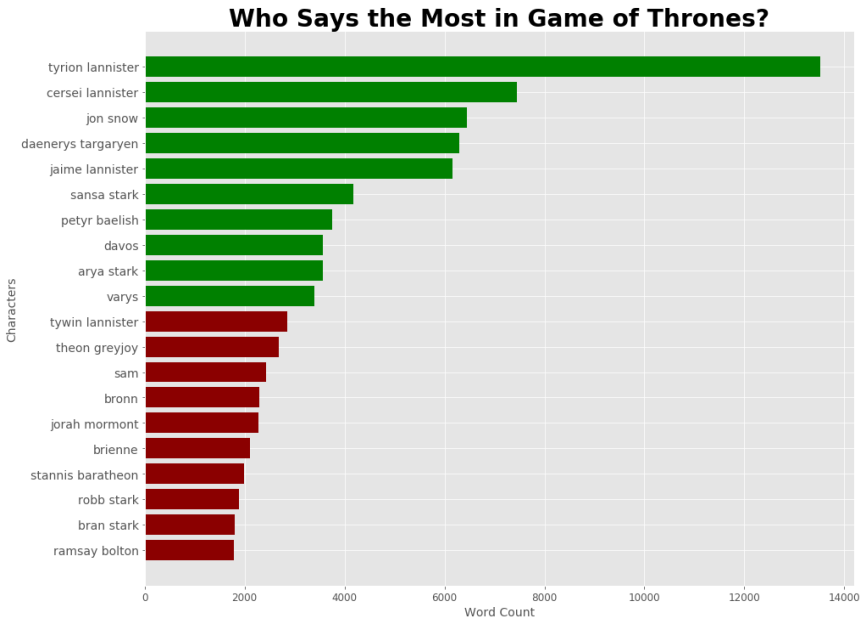
After importing necessary libraries and conducting preprocessing through cleaning, tokenization, normalization, stemming and part-of-speech tagging, we obtained the following post processing data (only shows the first 5 rows)

	Season	Name	Sentence	Tokenized Words	Tokenized Alphanumeric Words	Lowered_noPunc_noStop	Word Count	Pos_tag	Porter_stemmer
0	Season 1	waymar royce	What do you expect? They're savages. One lot s...	[What, do, you, expect, ?, They, 're, savages,...	[What, do, you, expect, They, savages, One, lo...	[what, expect, they, savages, one, lot, steals...	12	[(what, WP), (expect, VBP), (they, PRP), (sava...	[what, expect, they, savag, one, lot, steal, g...
1	Season 1	will	I've never seen wildlings do a thing like this...	[I, 've, never, seen, wildlings, do, a, thing,...	[I, never, seen, wildlings, do, a, thing, like...	[i, seen, wildlings, thing, like, i, seen, thi...	10	[(i, NN), (seen, VBN), (wildlings, NNS), (thin...	[i, seen, wildl, thing, like, i, seen, thing, ...
2	Season 1	waymar royce	How close did you get?	[How, close, did, you, get, ?]	[How, close, did, you, get]	[how, close]	2	[(how, WRB), (close, JJ)]	[how, close]
3	Season 1	will	Close as any man would.	[Close, as, any, man, would, .]	[Close, as, any, man, would]	[close, man]	2	[(close, JJ), (man, NN)]	[close, man]
4	Season 1	gared	We should head back to the wall.	[We, should, head, back, to, the, wall, .]	[We, should, head, back, to, the, wall]	[we, head, wall]	3	[(we, PRP), (head, VBP), (wall, VB)]	[we, head, wall]

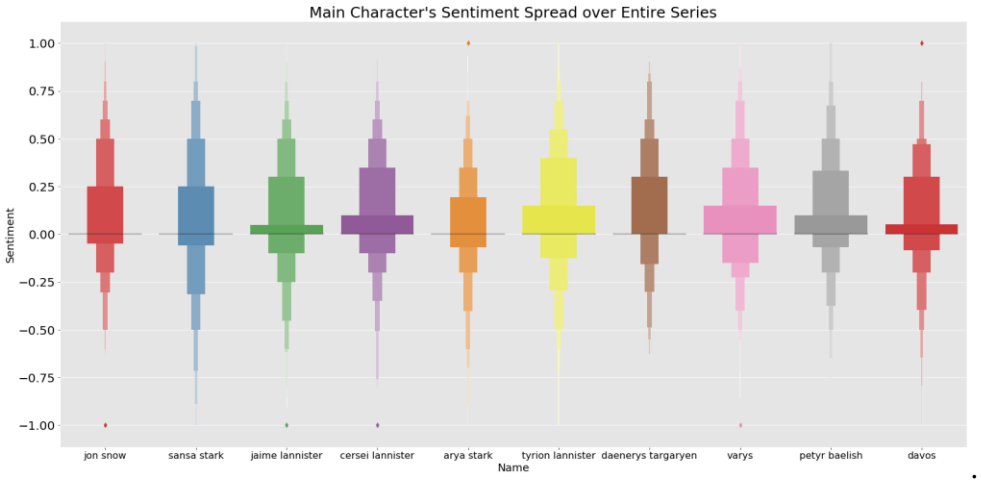
Text Analytic Method Description and Results

1. *How does the sentiment of characters change over time?*

For the first research question, in order to decide who are the main characters, we decide to first extract the top 20 according to who speaks the most (have the largest word count) over the entire series, and surprisingly we got not only does Tyrion dominate the whole series, overwhelmingly outpacing every other characters by at least 6000 words, but also do the Lannisters overall lead overwhelmingly in words. We decided to keep the top 10 (colored in green) as the main characters to conduct the following sentiment analysis over seasons.



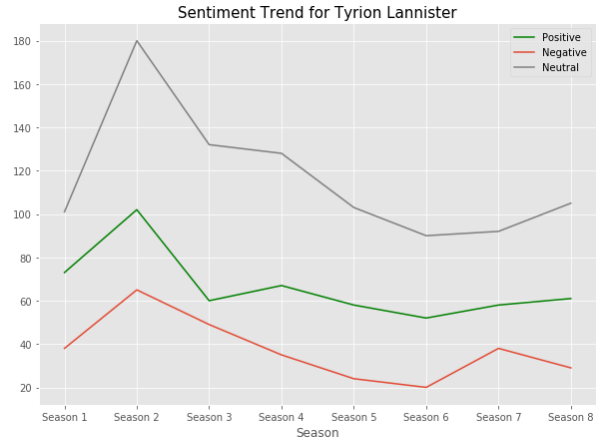
For this research question, the text analytic method we used is *TextBlob* library, which provides a simple API for diving into common natural language processing tasks such as sentiment analysis. The sentiment property returns a named tuple of the form *Sentiment (polarity, subjectivity)*. The polarity score is a float within the range [-1.0, 1.0]. The subjectivity is a float within the range [0.0, 1.0] where 0.0 is very objective and 1.0 is very subjective. By using the polarity score, we plotted a figure of sentiment spread of the 10 main characters' sentiment over the entire series using below seaborn boxplot:

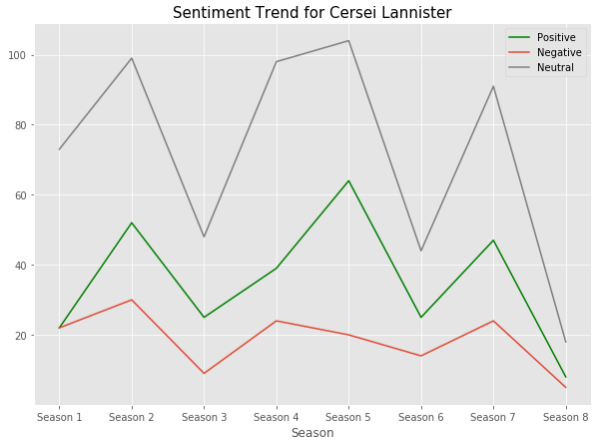


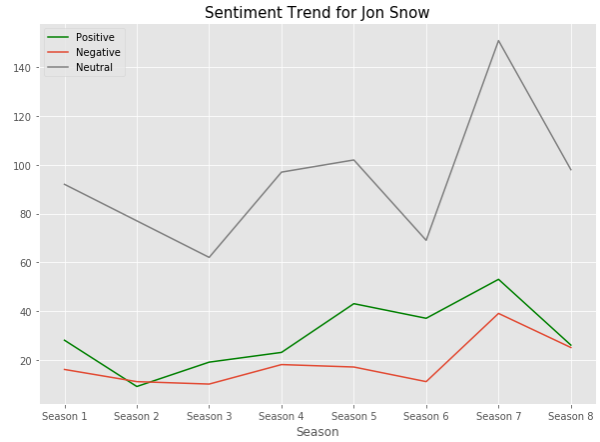
From the plot we can see that all characters had a very good sentiment spread with the largest spread being seen with Tyrion, Sansa and Cersei and the lowest spread being seen amongst Varys, Petyr and Davos. Another observation seen here is that even though there is a lot of spread with each character, each had a more neutral to positive sentiment over the entire season with the mean and median sentiment being nonnegative.

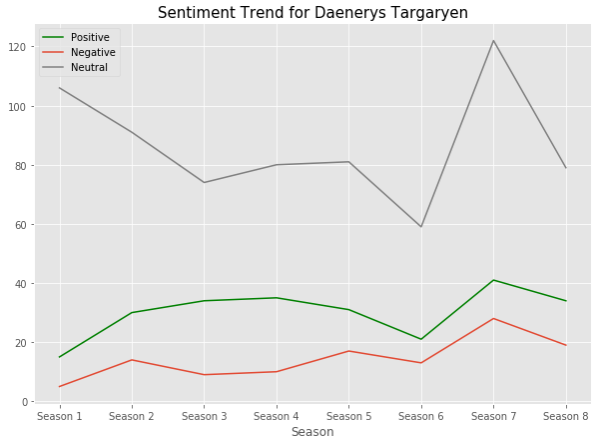
We then set the rules to determine the positive [0.05, 1.0], negative [-1.0, -0.05], and neutral [-0.05, 0.05] sentiment according to the polarity score.

Finally, we filtered out the 10 main characters and applied groupby to count each character's sentiment labels. Line graphs can show more clearly how each character's sentiment changes as the story develops. Here we only analyze the main 4 characters' sentiment changes over seasons (see the rest of the characters' charts in appendix 1).









Seeing from the count of the neutral sentiment we can tell if the character is relatively active in the season. For Tyrion Lannister, he is mostly active in season 2, while as the story develops, he speaks less in the following seasons. Tyrion had a very sharp rise in all sentiment over the first two seasons but following this all sentiment had a general decline over the entire series.

The second finding is that some of the characters talk more positive sentences than negative ones and maintain that throughout the whole series, such as Tyrion Lannister, Cersei Lannister, and Daenerys Targaryen. Cersei had a much more turbulent spread with sharp rises and declines in her sentiments, while Daenerys had a stable count for positive and negative sentiment the entire season with the sharpest changes happening in the neutral sentiment and over the last two seasons. Seeing from the sentiments of the two queens, we could also tell the difference

However, for the other characters, we can see the change even the reverse of their sentiment when they speak more sentences with a negative attitude, for example, Jon Snow talks more negatively in Season 2 than in other seasons. John had a generally a much calmer graph than Cersei and had a general slow rise in positive and negative sentiment over the series while his neutral sentiment had more sharp changes.

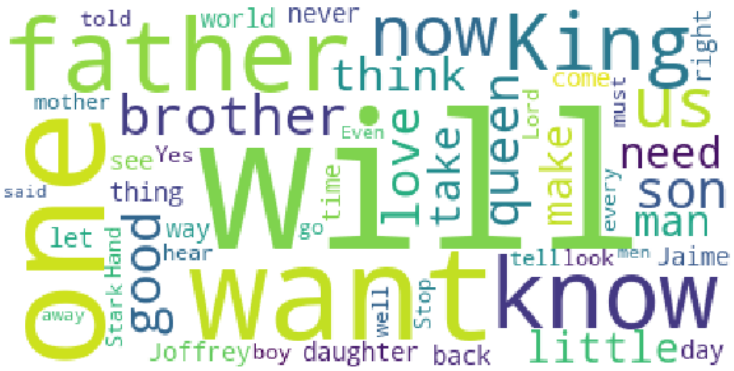
2. *Between Cersei and Daenerys, who said the word "kill" more often? Who said "love" more often?*

The second research question aims to compare these two potential queens by the words they said over the seasons. We assumed that the most important feature for a favorable queen is mercy. Therefore, by counting the words "kill" and "love," spoken by Cersei and Daenery, we can have a high-level understanding of who's the better queen in terms of mercy. Our result indicates that Daenerys said kill more frequently than Cersei did. Cersei said love more frequently than Daenerys did.

	Cersei	**Daenerys**
freq("kill")	**0.445%**	**0.628%**
freq("love")	**0.691%**	**0.240%**

However, the information from the frequency of kill and love is very limited, so deep exploration on Cersei and Daenerys' words cloud is conducted.

Cersei's Word Cloud:



Daenerys' Word Cloud:

Since Cercei is from the royal family so, from Cersei's word cloud, it's unsurprising that we have some royal terms such as King, queen. Other frequent words appeared like daughter, brother, father, son, Joffrey, who's her son, and Jaime, who's her brother. On the other hand, in Daenerys' word cloud, the dragon takes an important part over the whole season. Other frequent words like "people" that she used to say a lot of "my people" in the show, meaning that she cared about her people from a queen's perspective. Meereen is the city of Slaver's Bay that liberates Daenerys. The above-mentioned place and people are all significant to Daenerys.

3. *Based on the frequently used words, what aspect does each family reveal?*

The third research question aims to infer traits and aspects of the families (houses) from the series. To generate these insights, we counted the words most frequently used by each family to understand what topics they talk about the most and based on that what are the family's traits. We used the NLTK and WordCloud libraries to conduct the word count analysis and generate word cloud visualizations for the same.

The Targaryen family uses the words "know", "want", "men", "need", "king", "kill", "fight", "father" the most. The Greyjoy family uses the words "iron", "want", "father", "lord", "kill", "island" the most. The Lannister family uses the words "king", "want", "father", "kill", "Stark", "need" the most. The Stark family uses the words "know", "want", "lord", "king", "father", "kill", "men", "need" the most. The Baratheon family uses the words "king", "want", "father", "brother", "love", "Stark", "men", "need" the most. The Tully family uses the words "lord", "son", "king", "father", "Lannister" the most. The Martell family uses the words "sister", "kill", "children", "rape", "Lannister", "murder", "girl" The most. The Mormont family uses the words "khaleesi", "men", "queen", "dragon", "sword", "kill" the most.

Findings and Conclusions

As known, Game of Thrones dominated the decade due to a triumph in storytelling of characters, plot twists, dialogues, and emotions. The main characters are not just simply had positive sentiments or negative sentiments. They are more like figures in the real world, who stay positive but stand for different perspectives. There are no simple good people or bad people. As the story evolves, the sentiment of some characters may change with plot development.

If you believe that mercy is the most important quality of a leader, then Cersei is the better queen based on our research. However, after deep exploration of Cersei and Daenerys' words cloud. It can tell that Cersir used a lot of family-related terms and she's greatly devoted herself to her family. Combining the result of the word cloud and the comparison of kill and love frequency, we suppose that family, children are the two main things that Cercei cares about more. In contrast, Even though Daenery appeared more brutal from the words she spoke, she released the slaves from Meereen city and showed her caring for her people is also considered great mercy. From our analysis, it's actually hard to tell who's a better queen.

From the words each family is using, it can be concluded that Targaryen families are one of the key families fighting for the throne. They have armies of men to fight. They are ready to kill for the throne. The Greyjoy family are also one of the forerunners for the throne. They are a family that is near the seas and islands. They are also a patriarchal family because of the words “king” and “father” being used a lot. They are also ready to kill for the throne. The Lannister family is a family that is closely associated with the Stark family, though we cannot conclude whether it is a positive or a negative association. They are also a patriarchal family because of the words “king” and “father” being used a lot. They are also ready to kill for the throne. The Stark family are one of key families fighting for the throne. They have armies of men to fight. They are also a patriarchal family because of the words “king” and “father” being used a lot. They stay in a neutral position more and care for their families. The Tully family only have 1675 word count, so they are not a key family of the series and are a small party of it. They are a patriarchal family closely associated with the Lannisters. The Martell with only 870 word count. are not a key family of the series and are a small party of it. They are closely associated with the Lannisters. Based on the words, we can conclude that they are the most brutal families with words like rape and murder. The Mormont family with 2303 word count are not a key family of the series and are a small party of it. They are closely associated with the Khaleesi and the Targaryen. Based on the words, we can conclude that they work closely with dragons and swords and support the Targaryen in the fight for the throne. The Baratheon family is a patriarchal family because of the words “king” and “father” being used a lot. They also value brotherhood and love and are less likely to kill. They are also a family closely associated with the Stark family.

Practical implications

According to Dr. Paul Mroczka, an Associate Professor of Theatre and Director of Theatre at Plymouth State University, Script analysis is the major tool directors and others use in the theatre to turn a written script into the three-dimensional production, thus it is an important skill for directors, designers, and actors to master.

Not only can the audience gain deeper insights of Game of Thrones using script analysis, it also helps professionals such as television producer or director, screenwriter, and actors to better understand how the show is a triumph in storytelling of characters, plot twists, dialogues, and emotions.

References

- Dag Heward-Mills. ‘The Art and Science of Applied Leadership’.
- Loria, S. (n.d.). Simplified Text Processing, Textblob
- Mysiak, K. (2020, June 07). NLP Part 2: Preprocessing Text Data Using Python
- Morris, R. (n.d.). Roz Morris - Editor, https://reedsy.com/roz-morris
- Matthew 5:7. King James Bible. (n.d.). King James Bible Online
- Mroczka, P. (2016, January 12). Script Analysis Basics for Students: Understanding the Play and Plot, Part Two. Broadway Educators.
- Pallotta, F. (2019, April 30). 'Game of Thrones' epic battle sets viewership record, CNN
- Schermerhorn, A. C. (n.d.). Family Influence, Science Direct

Appendix 1 Sentiment Trend for other 6 Main Characters