Ayan Sarkar

University of denver

Web scraping using python

**Data Sources**:

I have used three types of data sources for this assignment. I have used data from Reviews(Viewers & Critics) of the last season of the show: Game of Thrones; I have used the RSS feed of CNN to fetch the top news stories; I have used the tweet searches related to the recent 2019 Indian elections (#IndianElections2019) from Twitter. I have used Python to extract data from the three different sources.

I have extracted around 400 total reviews: viewers & critics combined; 24 top news from CNN RSS feed and 100 tweets related to the Indian 2019 elections from Twitter. I have created the word clouds from all these data sources for the top 25 words obtained from them. All the word clouds have been generated in Python.

**Word Clouds** :

**Game of Thrones: Critics Reviews**:



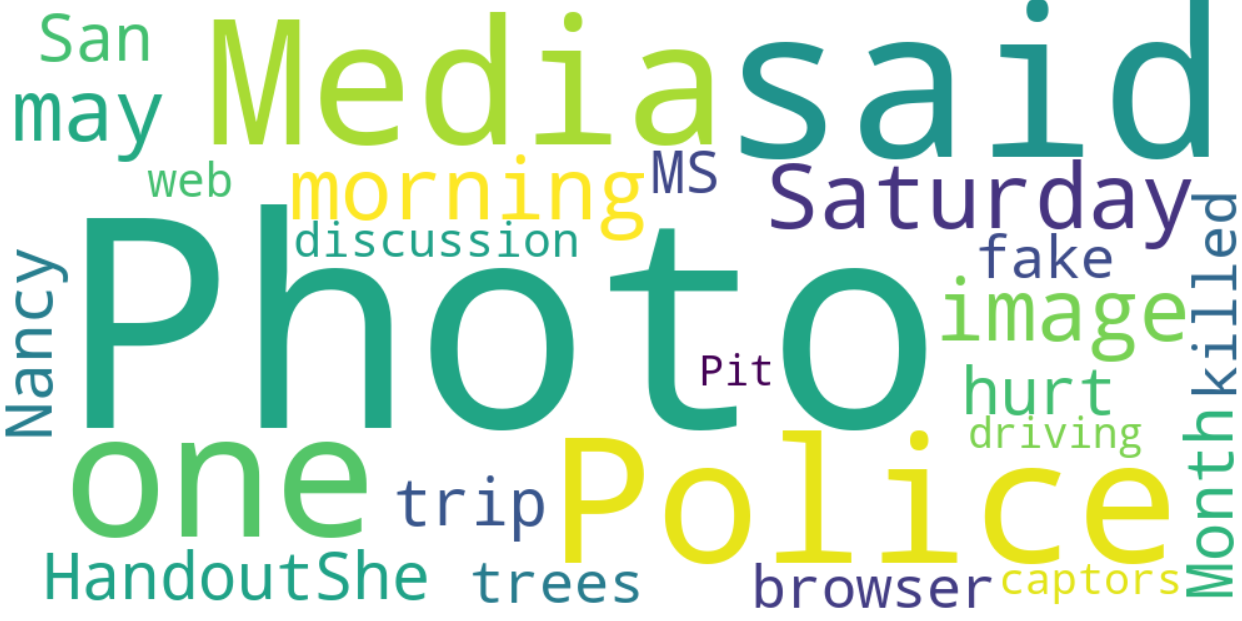
The critics review seems to resolve around story point, the characters, running time, the reunions in the story and the writer of the series, Song & Fire, Mr. George RR Martin. This is not a surprise that the critics reviews revolve around the story of the series as there was great displeasure regarding the story line about the series.

**Game of Thrones: Viewers Reviews**:



The reviews of the viewers revolve around the “one” savior that was promised to save Westeros, the Night King, the Wall and Winter fell; these lists of topics indicate that the viewers have allowed the plot of the story across all the episodes of the season and talked about the subjects related to the flow of the story.

**CNN RSS Top News Feed**:



The top news from CCN seem to revolve news around the Police, the speaker of the house, Nancy Peloci, killings, web, trees, handouts, the fake Media, captors of some crime and events that have taken place on Saturday, May.

**India 2019 elections**:



The word cloud generated for this topic is very reminiscent of the topics which were highlight of this mammoth election in my country. The elections were marred with controversies related to Muslim extremists, the election regulatory body, our Prime Minister, Mr. Modi and the two national parties of the country.

The next step that I would like to do in the future for all three data types is to perform a sentiment analysis on them and re produce the word cloud as per sentiment represented in the data.