

Anne Kroon,
Faculty of Social and Behavioural Sciences
Course: Big Data and Automated Content Analysis



UNIVERSITEIT VAN AMSTERDAM

Big Data and Automated Content Analysis

A comparison of news representation of BBC and CNN on
YouTube

Helge Moes
11348801
11348801@uva.nl

Date: 26.05.2023

Table of Content

Introduction	3
Research question:	4
Methodology	5
Analytical strategy	5
Dataset	6
Results	7
Exploratory Analysis	8
Sentiment Analysis	10
Conclusion	11
References	13
Appendix	15
Appendix 1: View Count Videos	15
Appendix 2: Like Count BBC News	16
Appendix 3: Like Count CNN	16
Appendix 4: Top 20 Liked Videos on CNN	17
Appendix 5: Top 20 Liked Videos on BBC News	17
Appendix 6: Top 20 Viewed Videos on CNN	17
Appendix 7: Top 20 Viewed Videos on BBC News	18
Appendix 8: Top 25 most mentioned words titles and descriptions of BBC News and CNN	18
Appendix 9: Sentiment Categories Description Comparison: BBC News vs CNN	19
Appendix 10: Sentiment Categories Title Comparison: BBC News vs CNN	20

Introduction

In recent years, the rise of social media and online platforms has greatly impacted the way people consume news and information. Among these platforms, YouTube has become a popular source of news. YouTube displays user-generated information alongside videos, including details like the number of views, likes, dislikes, and comments. The information within the views holds the secret to understanding what type of videos the audience prefer (Rui et al., 2019). It has become a common place for users to consume their daily news. Consequently, there have been multiple news outlets that have created a YouTube account in order to distribute their services through the internet.

Such news outlets are BBC News (at the time of recording 14,2 mln subscribers) and CNN (at the time of recording 15 mln subscribers). BBC News and CNN are two major news organizations that have a significant impact on the global news market. While BBC is widely recognized as the premiere news organization in the United Kingdom, CNN has a strong global presence and is a popular source of news in the United States. The Cable News Network, or CNN, is a cable news channel based in the United States, founded by Ted Turner in 1980 (*CNN Worldwide Fact Sheet*, n.d.). It was the first 24-hour television news channel and broadcasts primarily from CNN Center in Atlanta, Time Warner Center in New York City, and various studios in Washington, D.C., and Los Angeles. CNN is now owned by the Time Warner company and its U.S. news channel is owned by Turner Broadcasting System (*CNN Worldwide Fact Sheet*, n.d.). While CNN has numerous affiliates around the world, it primarily serves as a source of news in the United States. Nevertheless, the cable news channel was questioned on the content in their news stories.

In a study by Harvard University's Shorenstein Center on Media, Politics and Public Policy and the Project for Excellence in Journalism, it was found that CNN provided negative coverage of Republican presidential candidates at a ratio of three-to-one during the first five months of the 2007 presidential primaries. Negative coverage was given to all three main candidates with McCain receiving the worst (63% negative) and Romney receiving mostly neutral coverage. Although Obama fared the best, both Clinton and Edwards ended up with more negative than positive coverage overall ("The Invisible Primary — Invisible No Longer," 2007). This sentiment had a particular political origin, since CNN is identified as a left-leaning news channel (Mitchell, 2014).

On the other hand, the British Broadcasting Corporation, or BBC, is the largest broadcasting company in the world and is headquartered in London (*Review of BBC news and current affairs*, 2019). Under a Royal Charter, it primarily serves the United Kingdom, Channel Islands, and Isle of Man. It is received in all households, companies, and organizations in the UK that use equipment capable of receiving live television broadcasts. In recent years, the BBC News has faced criticism and challenges, including accusations of bias, funding cuts, and the changing media landscape (Gaddie, 2020; Lukis, 2021).

In July 2020, a BBC reporter used a racial slur in a report on the assault of a 21-year-old NHS worker and musician K-Dogg. This led to complaints to the BBC and raised questions about the racial sentiment the BBC has (Singh, 2020). Moreover, Indarjit Singh left his position at BBC Radio 4's Thought for the Day and criticized the BBC's political correctness in 2019. He claimed that an item about a Sikh guru who was executed in India for

opposing the forced conversion of Hindus to Islam in the 17th century was not broadcast because it might offend Muslims, despite the fact that it did not criticize Islam (Kennedy, 2023). Despite the sentiment held against these practices, the BBC News remains a highly respected and influential news organization, with a mission to inform and educate audiences around the world on the most important news and events of the day. It is therefore not a surprise that concerns have been raised about the potential for the sentiment in news on YouTube, as the platform has a largely unregulated ecosystem where anyone can create and share content (Lutz et al., 2021).

For this research project, a comparison between these two news organizations is logical, given their influence and reach, but mainly for their sentiment and controversies these news outlets hold. A viewer might be drawn or rather avoid certain news topics that are portrayed by the news channels. Based on the cases where sentiment played an integral role in news portrayal, the research shall examine whether sentiment plays a role for consumption of specific news topics these both channels portray in the views on YouTube. Moreover, the comparison between the two biggest news outlets of the United States and Kingdom, shall grant insights on how news is branded and presented on a platform such as YouTube (Vermeer et al., 2020).

Furthermore, comparing BBC News and CNN on YouTube with automated content analysis can be achieved by using various metrics and techniques. Since there are little to no automated content analyses on YouTube (Möller et al., 2019), the study shall consist of an explorative nature. Therefore, an analysis of the engagement metrics such as views and likes shall be conducted to observe the response of the viewers and how they interact on the channels' most viewed and liked content (Arcila, 2022). Another approach that shall be conducted is an analysis of the frequency of certain keywords or topics in both CNN articles and YouTube videos to compare the focus of their content (Arcila, 2022).

Finally, a sentiment analysis is conducted of video titles and descriptions on YouTube to determine the overall positive, negative, or neutral sentiment of the content (Boumans & Trilling, 2016). The aim of the sentiment analysis is to observe whether one can predict the number of likes and views based on the sentiment category of the title and descriptions of a dataset using a linear regression model. By analyzing the language, framing, and presentation of news stories on these channels, this research hopes to gain insights into how news are reflected and perpetuated on this popular platform by the viewer (Vermeer et al., 2020). Understanding these portrayals can inform discussions about the role of YouTube as a source of news and information in today's digital age. Based on these grounds, the following research question is formulated:

Research question:

To what extent can sentiment analysis be used to predict the view count of news on YouTube?

Sub Question:

How do CNN and BBC news channels perform based on their likes and view counts?

Methodology

In order to compare the YouTube channels of BBC News and CNN, this research implemented methods derived from *Computational Analysis of Communication* (Arcila, 2022) and *Toward open computational communication science: A practical road map for reusable data and code* (van Atteveldt et al., 2019). To answer the sub question, an exploratory analysis is conducted and both platforms are compared to each other based on their output. Besides the exploratory analysis, different sentiment analysis are made to examine the sentiment in the text of the videos presented on YouTube.

The first model of the sentiment analysis implements the sentiment form title as categorical variables to predict the view count as the dependent variable. A similar model shall be employed, but the sentiments of the descriptions shall be used in order to predict engagement. Lastly, given the potentially complex relationship between the variables as well as their types, this research shall implement a multiple regression model and test the accuracy, as this might be more robust than the previous tests. This section shall explain what steps were necessary to collect, process and analyze the data.

Analytical strategy

The preliminary steps consisted of the installation of various Python libraries using pip, which is a package manager for Python (van Atteveldt et al., 2019). The first library installed is google-api-python-client, which allows for interaction with various Google APIs. For this project, the YouTube Data API v3 was employed, since it allows access to the content presented on YouTube (*YouTube Data API*, n.d.). Secondly, the library is youtube-dl is implemented, which allows downloading videos from various websites, including YouTube (*YouTube Data API*, n.d.). The third library is TextBlob, which provides a simple API for natural language processing (NLP) tasks such as sentiment analysis, part-of-speech tagging, and noun phrase extraction (Arcila, 2022). Finally, the last library is word cloud, which is used for visual representations of text data where the size of each word is proportional to its frequency in the text (Holtz, n.d.).

Furthermore, a function `search_videos` was created in order to search for videos from a given YouTube channel using the YouTube Data API. The function takes a `channel_id` argument, which is the unique identifier for the YouTube channel to search (*YouTube Data API*, n.d.). Within the function, a `videos` list is created to store information about the videos found during the search. A `next_page_token` variable is initialized to an empty string and a `while` loop is used to iterate through the search results until at least 500 videos have been found.

Inside the loop, the `youtube.search().list()` method is used to retrieve up to 100 search results at a time, using the `channelId` parameter to filter the search by channel (*YouTube Data API*, n.d.). The video data for each search result is then extracted and added to the `videos` list as a dictionary. The `while` loop continues until at least 500 videos have been added to the `videos` list. This function is called twice at the end of the code to search for videos from the BBC News and CNN YouTube channels, and the search results are stored in the `bbc_news_videos` and `cnn_videos` variables, respectively. Finally, a data cleaning is performed on the 'title' and 'description' columns of the DataFrame 'df' by removing noise in

the form of certain characters and phrases. These cleaning operations are being performed using the `'replace()'` method of Pandas DataFrame.

To examine the data through an exploratory analysis of the two channels, the data is zoomed in on outliers of the most viewed and liked content by using `'sort_values()'`. This shall show the differences of content covered by the channels and what content can be regarded as preferred news that is consumed on YouTube. Various visualizations are used to portray these clear differences, such as word clouds and barcharts (Holtz, n.d.).

For the sentiment analysis, supervised machine learning has to be implemented (Arcila et al., 2022). Supervised machine learning explores statistical modeling and prediction, using regression analysis as an example (Bird et al., 2009). Moreover, it emphasizes the shift towards prediction in machine learning and addresses challenges in non-linear relationships and classification tasks (Arcila et al., 2022).

To create a sentiment classifier, several steps are involved. First, a `SentimentIntensityAnalyzer` named 'sia' is created to calculate scores for text data (Arcila et al., 2022). The sentiment analysis is then applied to the 'title' and 'description' columns. The lambda function utilizes the `'polarity_scores'` method of 'sia' on each value in the 'title_cleaned' and 'description_cleaned' columns. This function returns a dictionary of sentiment scores, and the 'compound' score (representing overall sentiment) is extracted and saved in new columns labeled 'title_sentiment' and 'description_sentiment.' In order to categorize the sentiment scores into positive, negative, or neutral categories, two additional columns are generated: 'sentiment_category_title' and 'sentiment_category_description.' Within the lambda function, the appropriate sentiment category ('Positive,' 'Negative,' or 'Neutral') is assigned based on the score conditions: scores greater than 0 for positive, less than 0 for negative, and otherwise classified as neutral.

The sentiment analysis is realized by implementing a Scikit-learn's `OneHotEncoder`. It is a class from the scikit-learn library in Python for converting categorical variables into binary vectors that can be used as input for machine learning algorithms and predictive modeling tasks (*Sklearn.Preprocessing.OneHotEncoder*, n.d.). The encoded categories were combined with the original dataset. Similarly, sentiments from descriptions were used to predict engagement. Training and test sets were created by using `'LinearRegression()'` and fitted by the `'fit()'` method, and a multiple regression model was employed for accurate predictions by using the `'predict()'` method (Arcila et al., 2022).

Dataset

For this study, 1000 videos were retrieved from the BBC News and CNN YouTube channels. This entailed that 500 videos of each channel were retrieved and were combined into a CSV file. Exploratory steps were implemented to examine the data presented in this file. In order to determine the first and last video registered on the BBC News and CNN channels, the code starts by filtering the pandas DataFrame (df) to select only videos from the BBC News channel. The resulting DataFrame is then sorted by the 'publishedAt' column. The first and last videos from the filtered DataFrame are then obtained using the 'iloc' method,

which is used to select rows by index. The same process is then repeated for the CNN channel. To assess the most viewed and liked videos on the two channels, the 'views' and 'likes' columns are converted into a numeric type, then filters the dataframe by the channel 'BBC News' and 'CNN'. In addition, the data is then sorted by views and likes, and selects the top 5 rows for each channel. Finally, the code prints the results, which are the top 5 most viewed and liked videos for each channel. Similar steps were taken for the other exploratory examinations.

The data consisted of: channel, video id (id), title of the video (title), description of the video (description), the number of views (views), the number of likes (likes) and the published date (publishedAt). Moreover, the posts by BBC News were published in a timeframe between the 26th of September 2014 to the 13th of May 2023. In the case of CNN, this was a timespan between the 30th of April 2013 until the 16th of May 2023. For BBC News the most viewed video is "Sex addiction: Five times a day 'wasn't enough' - BBC News" with 23,702,636 views and the most liked video is "Ben Shapiro: US commentator clashes with BBC Andrew Neil - BBC News" with 197,491 likes. Whereas for CNN, the most liked and viewed video is "Train conductor spots 3-year-old boy on tracks" with 268,471 likes and most viewed "Craziest moments at U.N. General Assembly" with 23098857 views (Also portrayed in Appendix 4, 5, 6 and 7).

Table 1. Overview of the data

Channel	BBC News	CNN
Begin date	2014-09-26	2013-04-30
End date	2023-15-13	2023-16-13
Number posts	500	500
Top viewed	23702636	23098857
Top liked	197491	268471

Besides cleaning the data, the dataset, new columns are added with 'sia' to give the data sentiment classifier attributes, which consist of: 'title_cleaned', 'description_cleaned', 'title_sentiment', 'description_sentiment', 'sentiment_category_title', 'sentiment_category_description.' These sentiment classifiers are used for the different tests that are run in order to determine whether one can deduce if the view count can be predicted based on the descriptions and titles. This results in the 'cleaned_df' and this dataset is used for exploring the news channels.

Results

In this section the results of the exploratory and sentiment analysis are presented. Firstly, the analysis of engagement metrics (views and likes) for 500 videos from BBC News and CNN revealed that BBC News had more videos generating higher likes. The word frequency analysis showed that BBC News frequently mentioned the word "BBC" in video

titles, while CNN focused more on news within the United States. Both channels covered similar topics related to global and European news, such as the Russian-Ukrainian war. Both portray that less political findings seem to catch the eye of the viewer on YouTube, yet both channels mention more political words in their content. To answer the subquestion, both channels perform similarly to each other and emphasize their own national news in their story coverage.

Secondly, three tests were conducted to predict view count based on sentiment analysis: sentiment on the title, sentiment on the description, and multiple regression. The sentiment analysis involved transforming categorical data into binary vectors and concatenating them with the original DataFrame. The models were evaluated using mean squared error (MSE) and R-squared scores. However, all three models showed high MSE values and negative R-squared scores, indicating they were not effective in predicting view count. Despite this, the sentiment analysis revealed that both BBC News and CNN had predominantly negative sentiment. This could be attributed to their coverage of severe news topics. However, these findings are speculative as they lack supporting outcomes.

Exploratory Analysis

Firstly, engagement descriptives such as views and likes were analyzed. All the 500 videos of each channel were plotted, in order to distinguish a difference in the views and likes. Appendix 1 and 2 prove that BBC News holds more videos that generate more views and likes. What Appendix 1, 2 and 3 ultimately portrays, is that the dataset is expansive. Consequently, further analyses shall draw on filters, such as 20 videos, that emphasize the most viewed or liked content on both channels.

By zooming in on the topics that generate these high likes and views count, the descriptives of Appendix 4, 5, 6, 7 point out that the news coverage of the BBC News generates a similar number of likes in comparison to the top 20 most liked videos. This also entails for the top 20 viewed videos. This might imply that both channels have a similar reach of distributing news.

The videos that generated a considerable number of views and likes for BBC News were “Ben Shapiro: US commentator clashes with BBCs Andrew Neil”, “FULL Interview: Prince Harry and Meghan Markle”, “Sex addiction: Five times a day 'wasn't enough'”, “The Queens advice on wearing a crown” and “The Queen vs The President.” These topics highlight the British sentiment the BBC News has and that the content that is validated can be considered as ‘entertainment’ rather than ‘political’, since these topics differ from world news that address severe topics such as “India overwhelmed by world’s worst Covid crisis” and “You killed a million people in Iraq George Galloway tells Jacqui Smith” (Möller et al., 2019).

For CNN, the most likes do not correspond to the most views. News stories such as: “Train Conductor spots 3-year-old boy on tracks”, “Jodi Thought the camera was off”, “Craziest moments at U.N. General Assembly”, “Migrant group attempts mass entry into US at Mexico Border” and “2 sets of identical twins switched at birth reunite” have many likes in correlation to the number of times they have been watched. In the lesser watched content,

Trump is mentioned a couple of times. Additionally, these topics seem to be focussed on local news, rather than world wide news.

In order to examine these news topics further, a word frequency analysis is conducted. the frequency of certain keywords and topics portrayed on the BBC News and CNN were compared. Table 2 portrays that the word ‘BBC’ is most frequently used (503), since it is mentioned in every title of a video. Although the word clouds portray a clear difference in news coverage, where CNN is mainly focussed on news that takes place in the United States and BBC News on news in the United Kingdom, the two news channels covered similar topics with regard to global and European news and portray more serious or political topics, such as the Russian-Ukrainian war (Appendix 8). However, the BBC News mentions the Russian-Ukrainian war more frequently than CNN and generates more likes for each video.

Table 2. Top 25 most mentioned words

Rank	CNN Words	Freq	BBC Words	Freq
1	Trump	145	BBC	503
2	former	90	ukraine	161
3	russian	84	russia	109
4	president	81	russian	95
5	donald	57	uk	91
6	trumps	56	president	85
7	new	55	war	71
8	us	55	says	62
9	ukraine	55	us	50
10	CNN	53	ukrainian	48
11	house	53	people	48
12	6	50	Putin	46
13	says	48	coronavirus	46
14	GOP	40	said	44
15	rep	36	city	40
16	republican	36	new	39
17	january	35	country	38
18	white	35	forces	35
19	cnns	33	minister	35
20	video	33	first	35
21	election	32	invasion	34
22	see	31	ukraines	32
23	john	30	troops	32
24	ukrainian	30	russias	32

Sentiment Analysis

To investigate whether one can predict the view count based on the sentiment analysis, three tests have been conducted: the sentiment on the title, the sentiment on the description and a multiple regression. As stated by Boumans & Trilling (2016), a sentiment analysis consists of a dictionary of a constructed list of words with attached tone scores and are dichotomous. This shall be realized in the first two tests.

In this case, the titles and descriptions are considered to be categorical and are transformed into binary vectors. Furthermore, the encoded categories are then concatenated with the original DataFrame. The features (X) were defined as the columns of the encoded categories, and the target variable (y) is set as 'views'. The data is split into training and test sets using the `train_test_split` function, with 80% of the data used for training and 20% for testing. The final test, multiple regression, was implemented to examine the robustness of the model.

Table 3. Overview Sentiment Analysis

Model	Mean Squared Error	R-squared
1	1439766912860.85	-0.01
2	1468847710290.61	-0.03
3	1462495343923.90	-0.03

Note: Model 1: sentiment title
Model 2: sentiment description
Model 3: multiple regression

The first sentiment analysis that was implemented to predict the number of views based on the sentiment title of the data performed on the 'sentiment_category_title' column.. The obtained results show an MSE of 1462495343923.90 and an R2 score of -0.01. The second test on 'sentiment_category' resulted in an MSE of 1439766912860.85 and an R2 score of -0.03. Lastly, The multiple regression test, containing 'Positive' 'Negative' 'Neutral,' showed that the score of MSE was 1468847710290.61 and it had an R2 score of -0.03, which implies that the model is not robust.

Based on these findings, the models are not fit to answer the research question. Despite this fact, the sentiment of BBC News is predominantly considered to be negative and CNN is a combination of neutral and negative (Appendix 9 and 10). This portrays that both channels are considered to be negative, which implies that both channels convey a negative sentiment, emotion or opinion. Moreover, this might be due to the fact that these channels portray severe news topics more often, such as: war, pandemic, and other negative news topics. This is an interesting finding, since the most viewed content was considered to be

light-hearted in nature that gained the most views. Then again, these findings are merely speculation, since the sentiment analysis did not contain supporting outcomes.

Conclusion

The present study attempted to investigate the extent to which sentiment analysis can be used to predict the view count of news videos on YouTube by comparing BBC News and CNN. The research utilized methods derived from previous studies in computational analysis of communication (Arcila, 2022) and open computational communication science (van Atteveldt et al., 2019) to compare the YouTube channels of BBC News and CNN. Moreover, a dataset consisting of 1000 videos (500 from each channel) was collected and analyzed and an exploratory analysis was conducted to examine engagement metrics such as views and likes. It was observed that BBC News and CNN generated similar output with regard to views and likes. However, when focusing on the top 20 most viewed and liked videos, both channels showed similar viewership, suggesting that both channels reach a broad audience on YouTube. Although both channels seem similar in output, BBC News received more validation through likes per video, indicating a stronger connection with viewers.

Further analysis involved examining the topics generating high view counts and likes. It was found that BBC News videos related to British sentiment and content that may be considered entertaining received the most views and likes. In contrast, CNN videos covering worldwide news received fewer likes compared to their view counts. These findings suggest potential differences in content preferences and viewer engagement between the two channels and that nationalities play an integral role for news consumption. Furthermore, a word frequency analysis was conducted to compare the topics covered by BBC News and CNN. The analysis revealed distinct news coverage, with CNN focusing primarily on events in the United States and BBC News emphasizing news within the United Kingdom. However, both channels covered similar topics related to global and European news.

To predict view counts based on sentiment analysis, three tests were performed: sentiment analysis of the title, sentiment analysis of the description, and a multiple regression. The sentiment analysis utilized the OneHotEncoder class from the scikit-learn library to convert categorical variables into binary vectors. However, the results of the sentiment analysis tests did not yield accurate predictions, as indicated by the high mean squared error (MSE) values and low R-squared (R²) scores. Despite the models' inability to predict view counts accurately, an interesting finding emerged regarding the sentiment of the channels. Both BBC News and CNN were predominantly associated with negative sentiment in the titles and descriptions, possibly due to the channels covering severe news topics such as wars and pandemics. However, these findings remain speculative, as the sentiment analysis did not provide supporting outcomes and therefore the research question cannot confirm whether sentiment analysis can be used for predicting view count on YouTube.

This raises questions for future research, since the results only considered titles and descriptions. However, it is interesting for future research to address the comment section and the content within the video's presented on these channels and to code whether these hold negative, neutral or positive sentiment. For this approach, a manual content analysis might grant valuable insights of the video content, whereas the comment section could also be

examined through automated measurements. Additionally, other models may be considered or implemented to research the sentiment of news channels. It would also be interesting to address sentiment of different news channels, such as local news and observe whether sentiment is more clearly portrayed.

In conclusion, to answer the research question: “To what extent can sentiment analysis be used to predict the view count of news on YouTube?”, the extent to which sentiment analysis can be used to predict view count of news on YouTube is limited based on the findings of the conducted tests. The models that employed sentiment analysis on the title, description, and multiple regression did not effectively predict view count. These results suggest that sentiment analysis alone may not be sufficient to accurately predict the number of views a news video will receive on YouTube. Other factors and variables likely play a significant role in determining view count, and additional research or alternative approaches may be needed to improve prediction accuracy. Nevertheless, the exploratory analysis to answer the sub question “How do CNN and BBC news channels perform based on their likes and view counts?”, revealed differences in viewer engagement and content preferences between BBC News and CNN. Further research is needed to explore alternative approaches and factors influencing the view counts of news videos based on sentiment on YouTube.

References

- Arcila, W. van A., Damian Trilling & Carlos. (2022, March 11). *Computational Analysis of Communication*. <https://cssbook.net/>
- Bird, S., Klein, E., & Loper, E. (2009). *Natural language processing with Python: analyzing text with the natural language toolkit*. O'Reilly Media, Inc.
- Boumans, J. W., & Trilling, D. (2016). Taking Stock of the Toolkit: An overview of relevant automated content analysis approaches and techniques for digital journalism scholars. *Digital Journalism*, 4(1), 8–23. <https://doi.org/10.1080/21670811.2015.1096598>
- CNN Worldwide Fact Sheet. (n.d.). Retrieved May 3, 2023, from <https://cnnpressroom.blogs.cnn.com/cnn-fact-sheet/>
- Gaddie. (2020). BBC Political Bias. *Medium*. <https://medium.com/@alexgadd83/bbc-political-bias-eec7ec29ef57>
- Holtz, Y. (n.d.). *Python Graph Gallery*. The Python Graph Gallery. Retrieved May 1, 2023, from <https://www.python-graph-gallery.com/>
- Kennedy, D. (2023). *Sikh peer leaves BBC Radio 4 show with swipe at 'thought police.'* <https://www.thetimes.co.uk/article/sikh-peer-leaves-bbc-radio-4-show-with-swipe-at-thought-police-m3vcjg77w>
- Lukis, L. (2021, September 9). BBC News vs. The U.S. News...Is There a Media Bias? *Medium*. <https://jasonmattlukis.medium.com/bbc-news-vs-the-u-s-news-is-there-a-media-bias-a9fd8cbf4f55>
- Lutz, M., Gadaginmath, S., Vairavan, N., & Mui, P. (2021). Examining Political Bias within YouTube Search and Recommendation Algorithms. *2021 IEEE Symposium Series on Computational Intelligence (SSCI)*, 1–7. <https://doi.org/10.1109/SSCI50451.2021.9660012>
- Mitchell, A. (2014, October 21). Section 1: Media Sources: Distinct Favorites Emerge on the Left and Right. *Pew Research Center's Journalism Project*. <https://www.pewresearch.org/journalism/2014/10/21/section-1-media-sources-distinct-favorites-emerge-on-the-left-and-right/>
- Möller, A. M., Kühne, R., Baumgartner, S. E., & Peter, J. (2019). Exploring User Responses to Entertainment and Political Videos: An Automated Content Analysis of YouTube. *Social Science Computer Review*, 37(4), 510–528. <https://doi.org/10.1177/0894439318779336>

Review of BBC news and current affairs. (2019). Ofcom.

https://www.ofcom.org.uk/data/assets/pdf_file/0025/173734/bbc-news-review.pdf

Rui, L., Afif, Z., Saedudin, R., Mustapha, A., & Razali, N. (2019). A regression approach for prediction of Youtube views. *Bulletin of Electrical Engineering and Informatics*, 8.

<https://doi.org/10.11591/eei.v8i4.1630>

Singh, A. (2020, July 31). BBC receives complaints after white reporter uses n-word on air. *The Telegraph*.

<https://www.telegraph.co.uk/news/2020/07/31/bbc-receives-complaints-white-reporter-uses-n-word-air/>

Sklearn.preprocessing.OneHotEncoder. (n.d.). Scikit-Learn. Retrieved May 13, 2023, from

<https://scikit-learn/stable/modules/generated/sklearn.preprocessing.OneHotEncoder.html>

The Invisible Primary—Invisible No Longer. (2007, October 26). *Shorenstein Center*.

<https://shorensteincenter.org/the-invisible-primary-invisible-no-longer/>

van Atteveldt, W., Strycharz, J., Trilling, D., & Welbers, K. (2019). Toward open computational

communication science: A practical road map for reusable data and code. *International Journal of Communication : IJoC*, 13.

<https://dare.uva.nl/search?identifier=e3b16032-b51d-48d6-b7ba-22050746bb2a>

Vermeer, S., Trilling, D., Kruikemeier, S., & de Vreese, C. (2020). Online News User Journeys: The Role of Social Media, News Websites, and Topics. *Digital Journalism*, 8(9), 1114–1141.

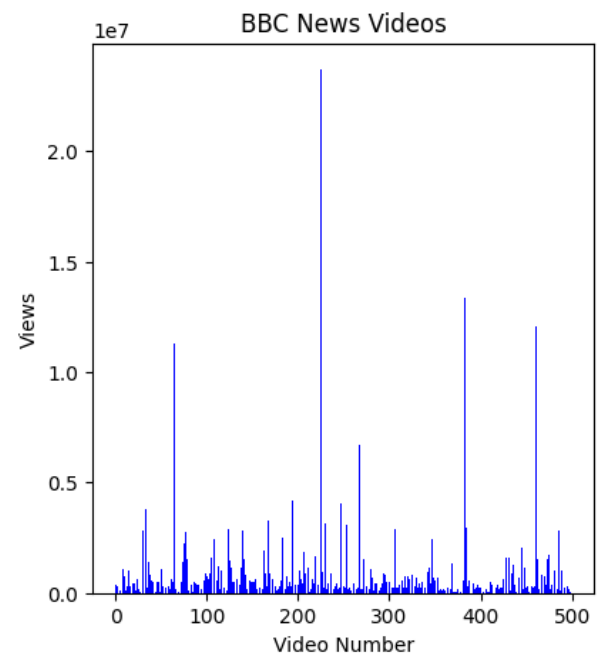
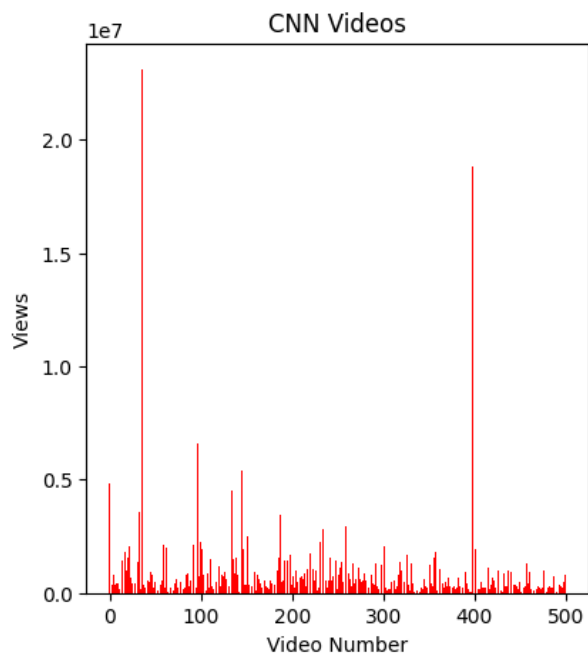
<https://doi.org/10.1080/21670811.2020.1767509>

YouTube Data API. (n.d.). Google Developers. Retrieved April 24, 2023, from

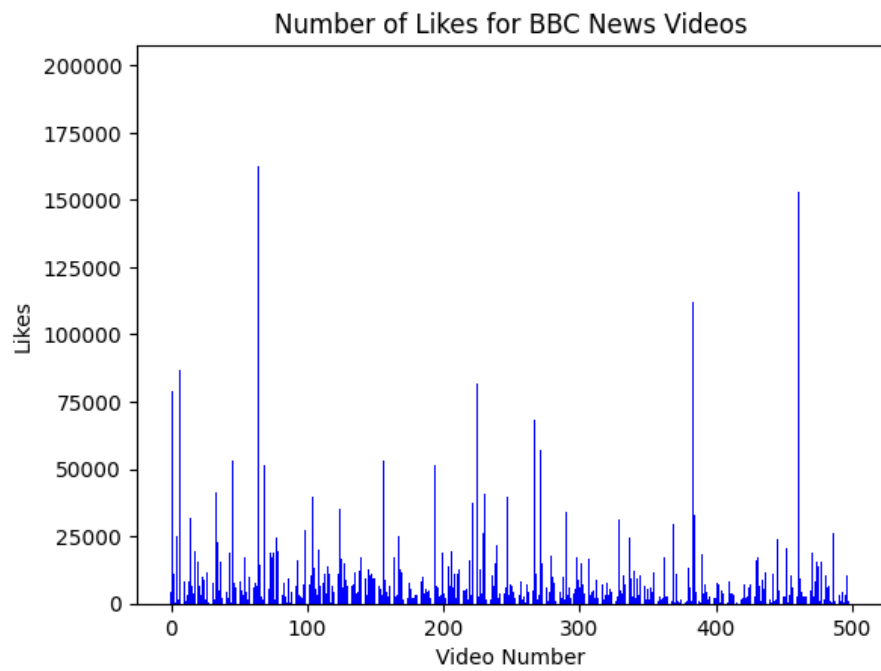
<https://developers.google.com/youtube/v3>

Appendix

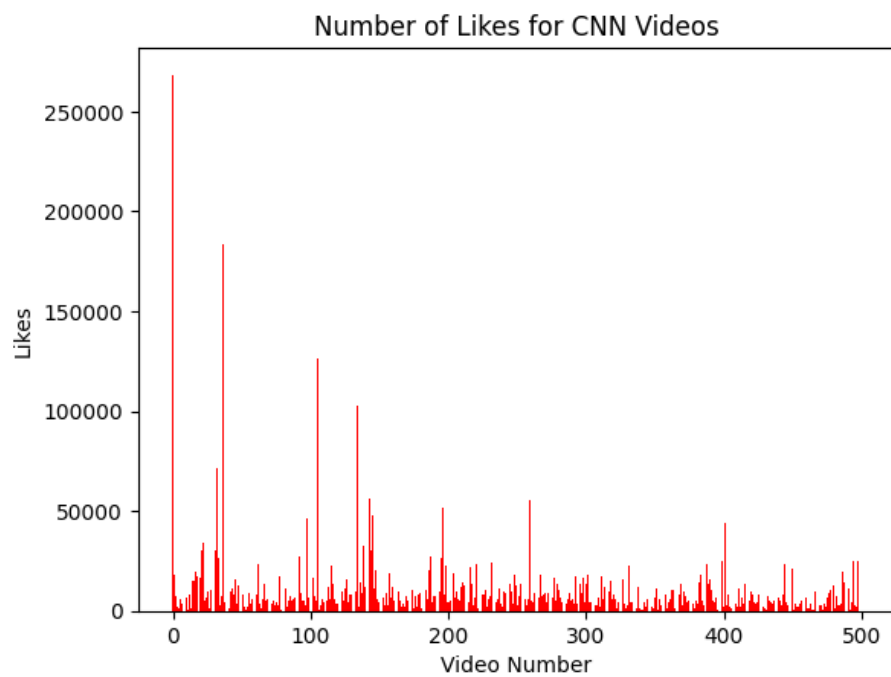
Appendix 1: View Count Videos



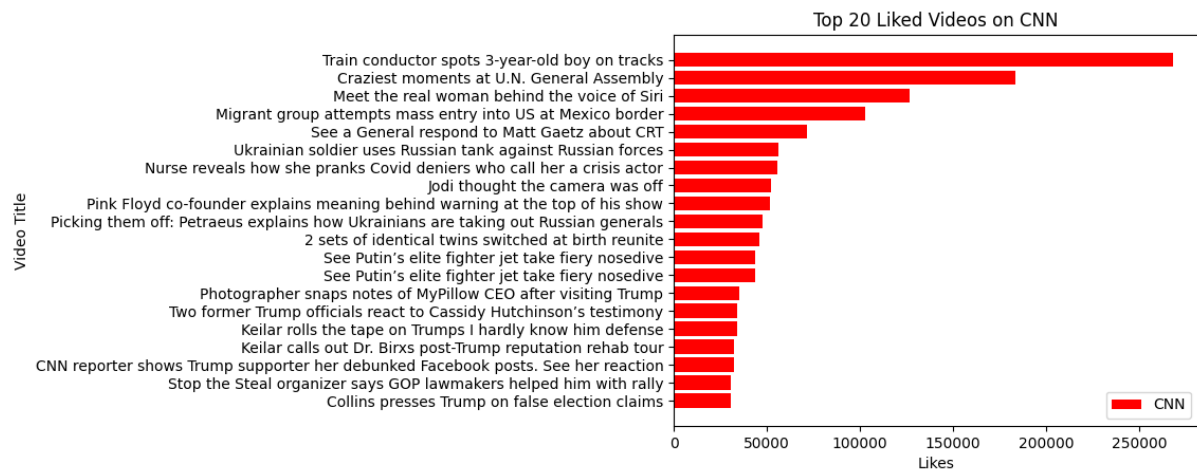
Appendix 2: Like Count BBC News



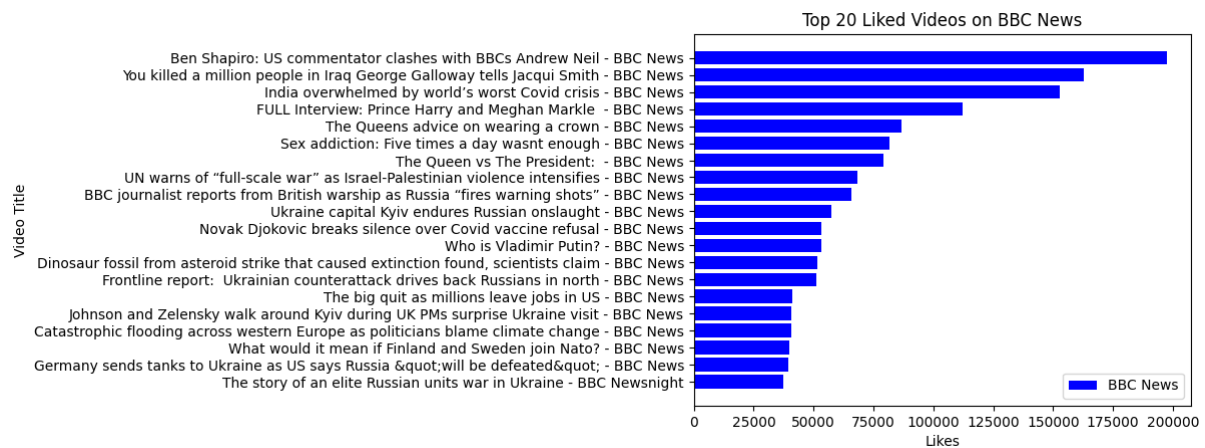
Appendix 3: Like Count CNN



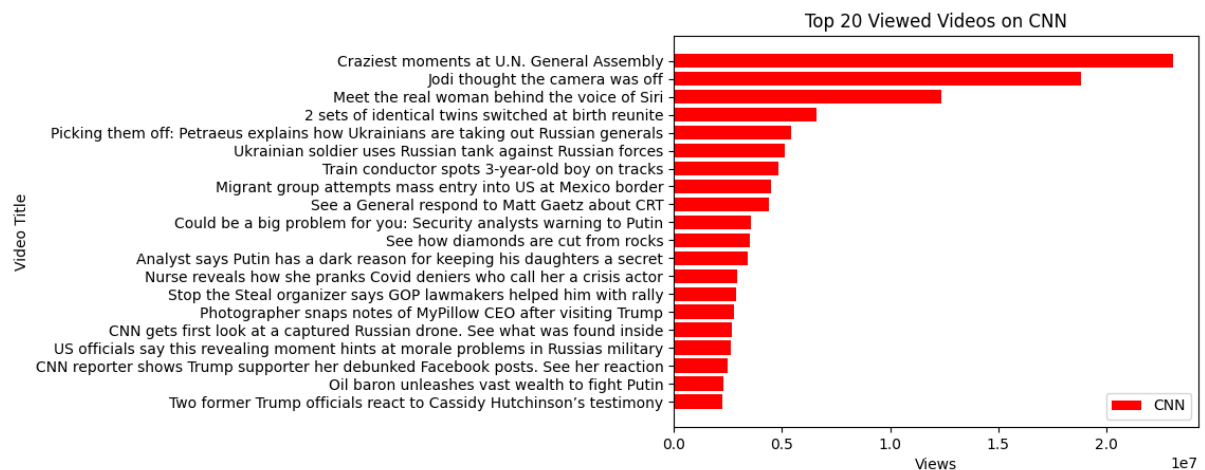
Appendix 4: Top 20 Liked Videos on CNN



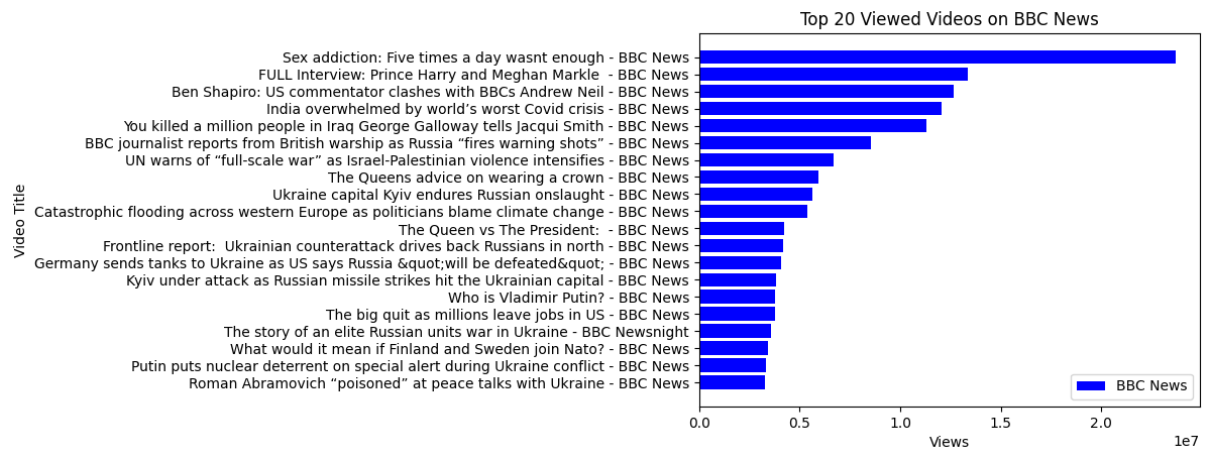
Appendix 5: Top 20 Liked Videos on BBC News



Appendix 6: Top 20 Viewed Videos on CNN



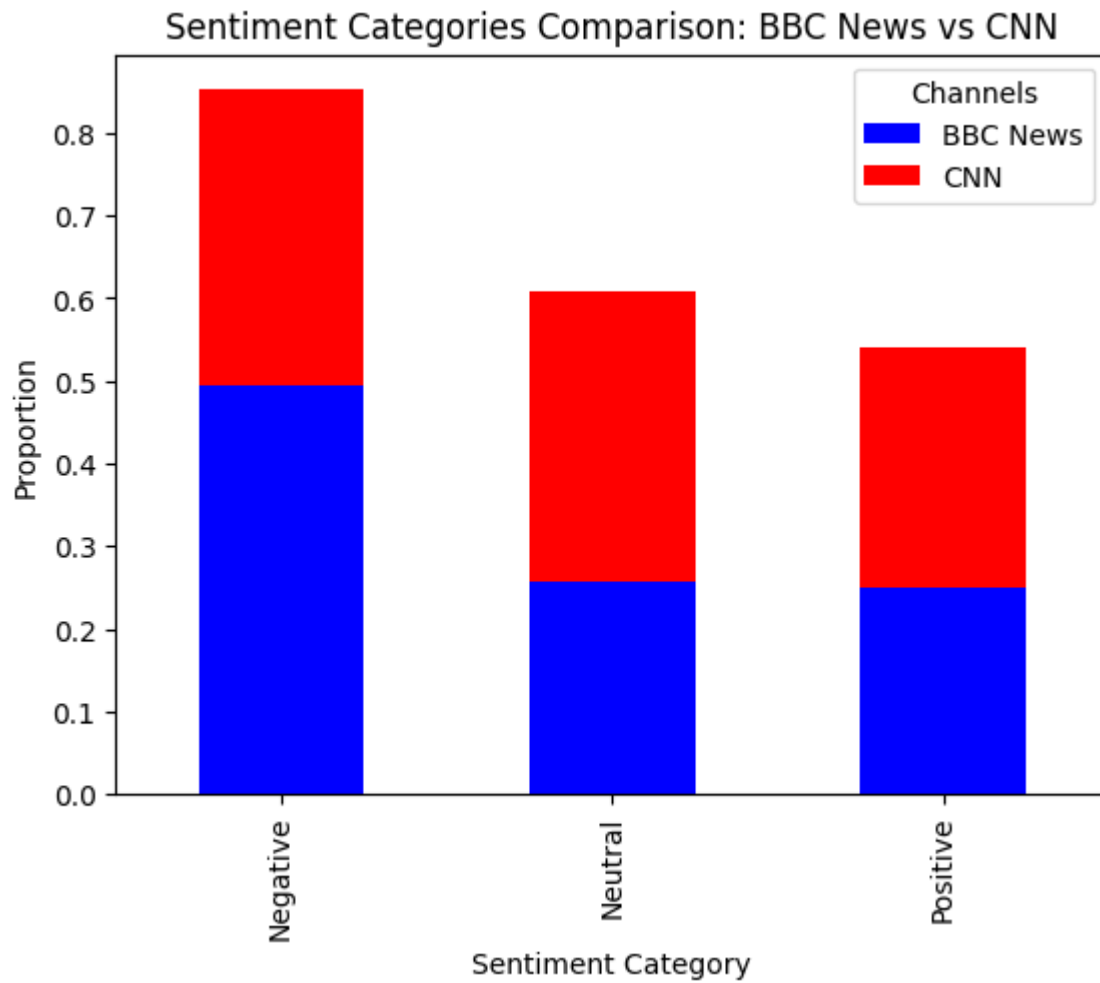
Appendix 7: Top 20 Viewed Videos on BBC News



Appendix 8: Top 25 most mentioned words titles and descriptions of BBC News and CNN



Appendix 9: Sentiment Categories Description Comparison: BBC News vs CNN



Appendix 10: Sentiment Categories Title Comparison: BBC News vs CNN

