QT2005_Assesment 2_00606851   This Document Contains 672 Words

Link to Codebook Below

[CODEBOOK](#)

Key

- *DM* = The Daily Mail
- *DE* = The Daily Express
- *BBC* = The British Broadcasting Cooperation
- *FT* = The Financial Times
- *TG* = The Guardian
- *TM* = The Mirror
- *NLP* = Natural Language Processing

Introduction to Project

This enquiry analyses YouTube titles published by UK news agencies. The publishers were chosen for their positioning on the political spectrum (Oxford Royale Academy, 2022)). Content on YouTube has a tremendous influence upon influencing public opinion, and its content has been used to understand political dynamics among various political groups (Munger & Phillips, 2020), therefore it was suitable for meaningful NLP analysis.

Gathering/Editing Data

Data was garnered through YouTube's API. My developer account limits me to 50 datapoints per account request. I narrowed my enquiry to the publishers' most recent videos; this to get a better understanding of each publisher through the way they cover similar topics/current affairs (e.g., Ukraine/Cost of Living). Once data was sourced from the API, errors present within the datasets (repeated, useless phrases such as #shorts) were removed.

Analysing Data

Textual analysis pertained to:

- Lemmatisation
- Tokenisation
- Case of text (lower/upper)
- Punctuation
- Stop words
- VAD characteristics
- Diversity of vocabulary
- PCA/Word Embeddings analysis
- PCA/Topic analysis.

(Lemmatisation, tokenisation, and stop words analysis was conducted for admin purposes, allowing me to do higher level VAD/PCA analysis)

Section A: VAD

The VAD model was successful in representing how a news' political leaning influenced the emotional makeup of their titles:
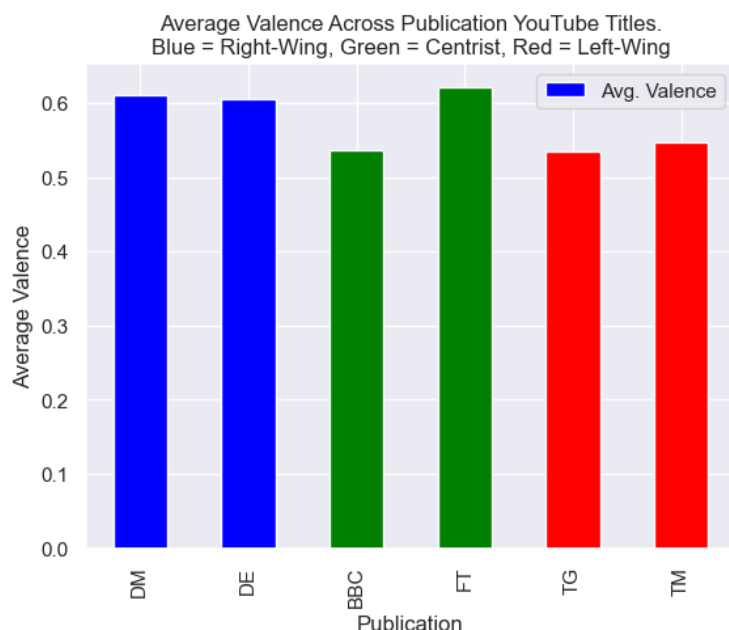

Figure 1

Right-wing publications, on average, use happier emotional language than centrist/left-wing publications. I have no plausible reasoning as to why.
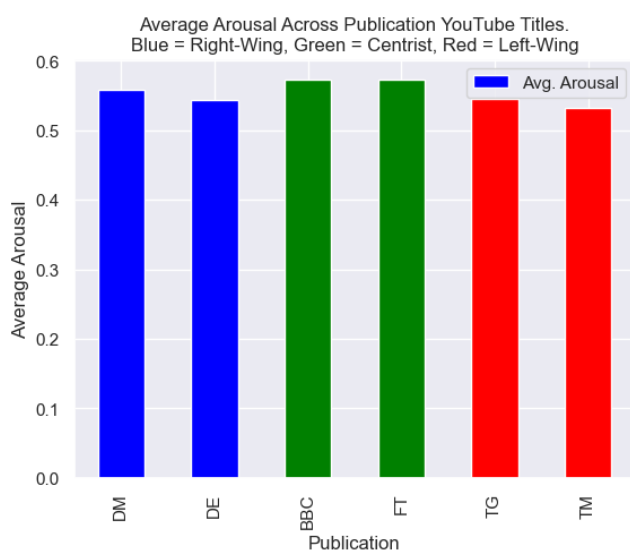

Figure 2

All publications feature a high level of arousal. This is to be expected as news titles readily use arousing text to maintain reader attention.

Average Dominance Across Publication YouTube Titles.
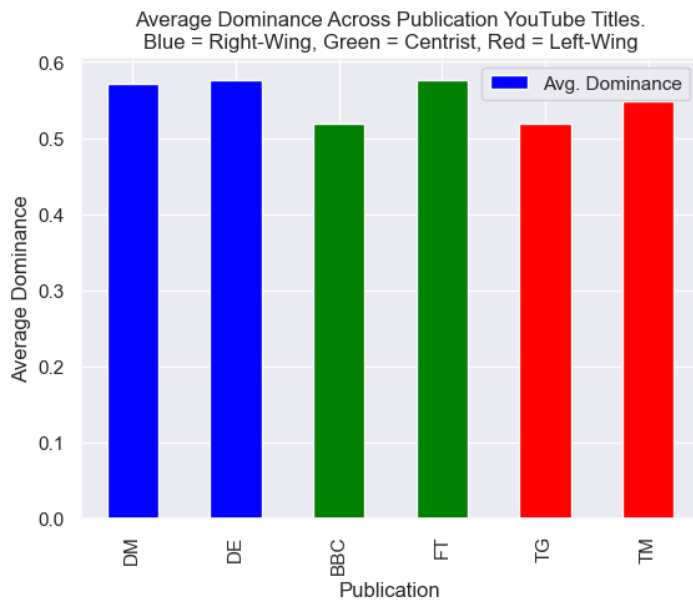Blue = Right-Wing, Green = Centrist, Red = Left-Wing

Figure 3

Right-wing publications, on average, utilise dominating language more than other publications. I propose that dominating language instils security in those that read them. This security maybe attractive to the right-wing audience who aims to also perpetuate secure, conservative values.

Comparing VAD Scores Across Right-Wing, Centrist, and Left-Wing UK Publications. Achieved Through Their A
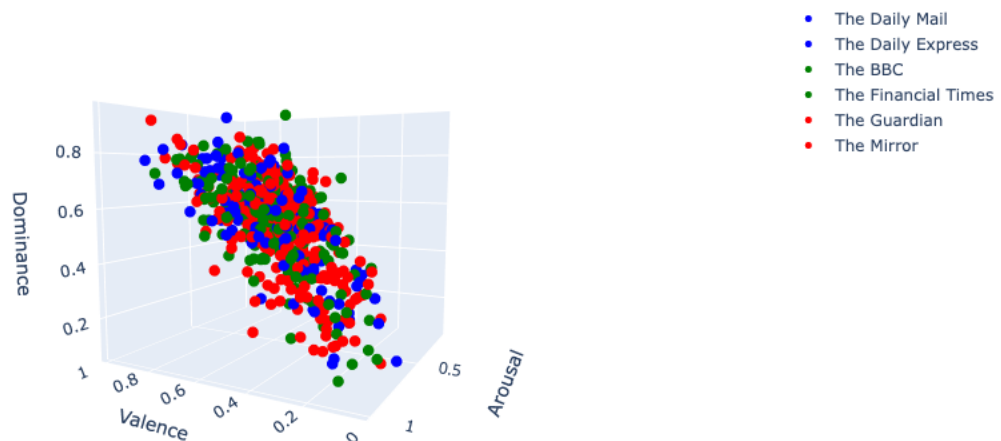
- The Daily Mail
- The Daily Express
- The BBC
- The Financial Times
- The Guardian
- The Mirror

Figure 4

VAD data was modelled in 3D, highlighting how similar publishers were to each other. They all follow an identical distribution to normal VAD models (negative linear correlation between Dominance and Valence).

Comparing VAD Scores Across Right-Wing, Centrist, and Left-Wing UK Publications. Achieved Through Their A
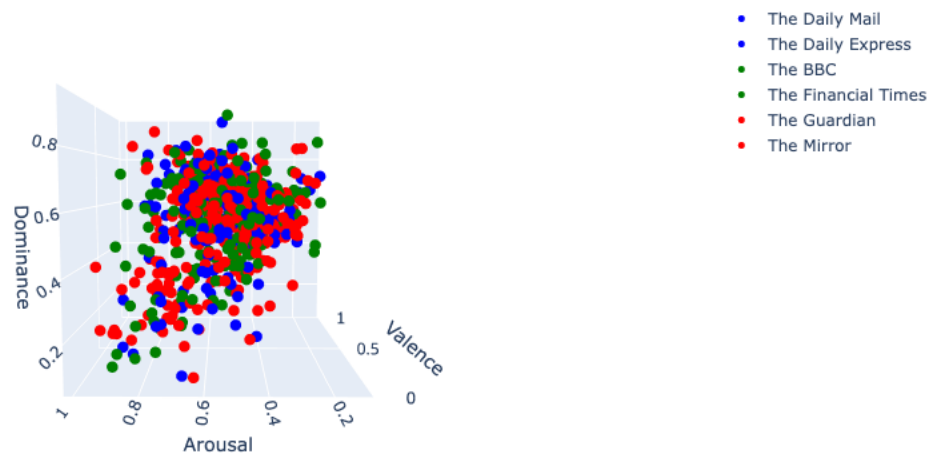

Figure 5

I did notice a missing patch in distribution between low arousal and dominance. This may also be a feature of common VAD models – I cannot confirm with the data available. Being in control is arousing, therefore, the inverse should be true.
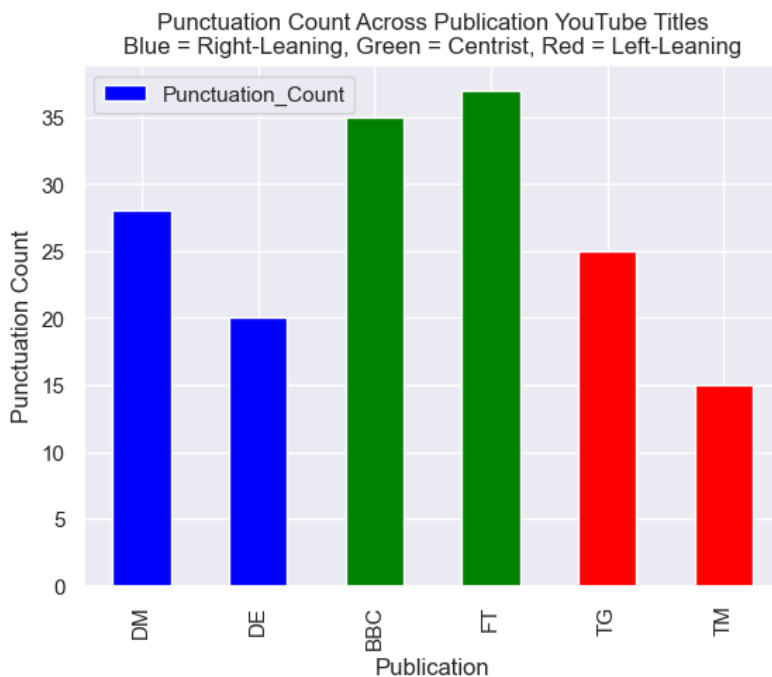
Section B: Punctuation

Figure 6

Centrist publishers use more punctuation, on average, than other publishers. I could not gain meaningful analysis from this chart as I was not specific enough about which types of punctuation were used, and whether lots of punctuation really means anything.
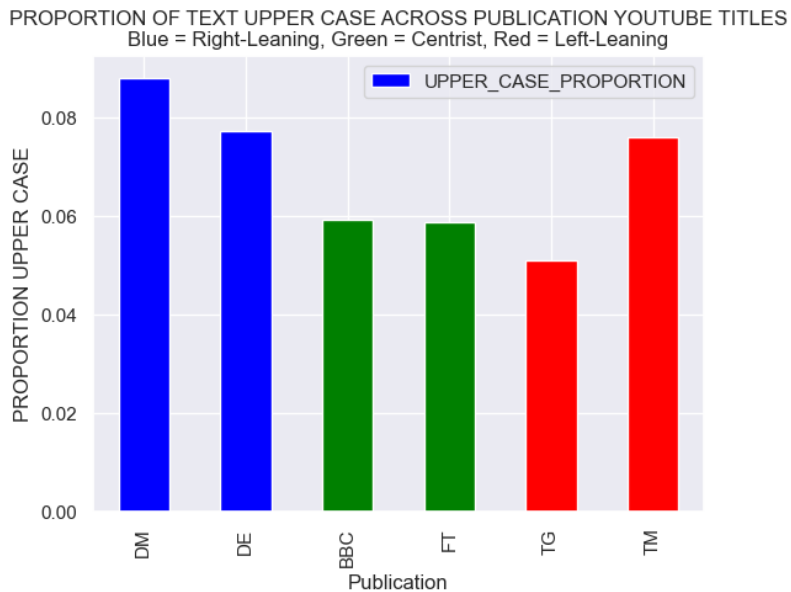
Section C: CASE OF TEXT

Figure 7

The right-wing uses more UPPER-CASE text, on average, than other publications. Upper case text increases the urgency of a headline, maintaining reader attention. This is not symbolic of the political leaning, more of the sensational nature of the title.

Section D: Vocabulary Diversity

The Daily Mail YouTube Titles Vocabulary Diversity

The Daily Express YouTube Titles Vocabulary Diversity

The BBC YouTube Titles Vocabulary Diversity

The Financial Times YouTube Titles Vocabulary Diversity

The Guardian YouTube Titles Vocabulary Diversity

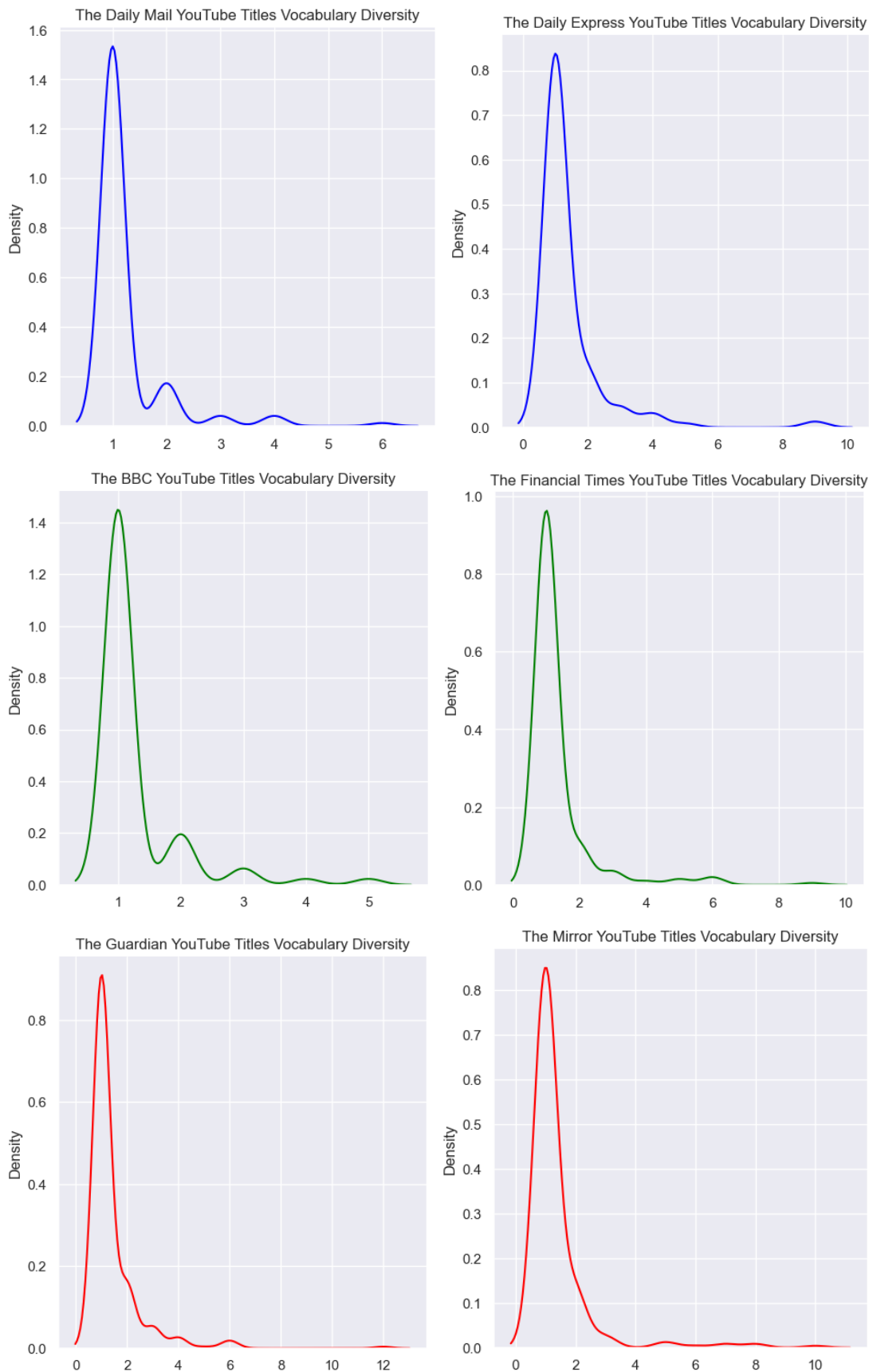The Mirror YouTube Titles Vocabulary Diversity

Figure 8

I could not find any meaning from analysis into vocabulary range as graphs were too inconsistent in spread. They did all, however, follow an expected zip's law distribution.

Section E: PCA/Word Embedding

YouTube Titles of Left/Centre/Right Wing UK news outlets – PCA analysis of Word Embedding
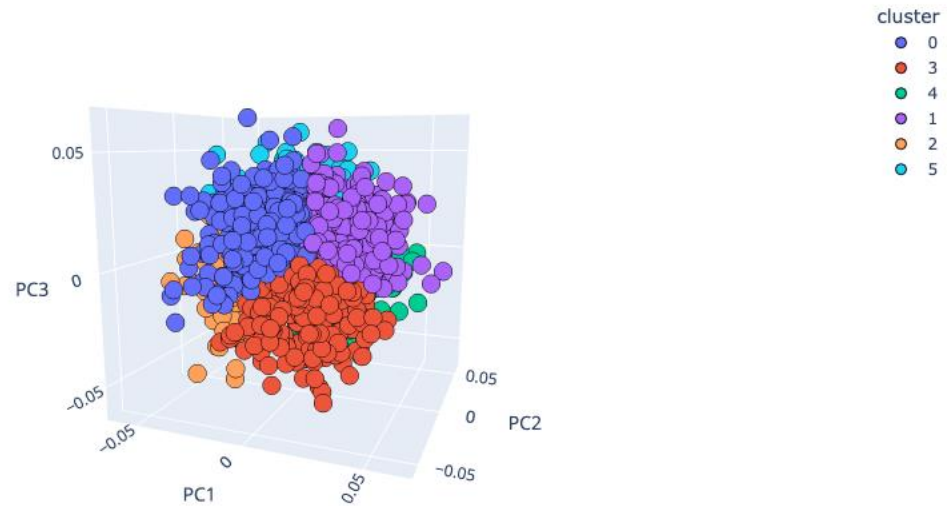

Figure 9

     This model separates the plots into a cube-like arrangement, whereby each side of the cube is another cluster from PCA analysis. The groupings are random in nature, with no common thread between any word groupings. This randomness within clusters was not due to how many clusters assigned to the model.

YouTube Titles of Left/Centre/Right Wing UK news outlets – PCA analysis of Word Embedding
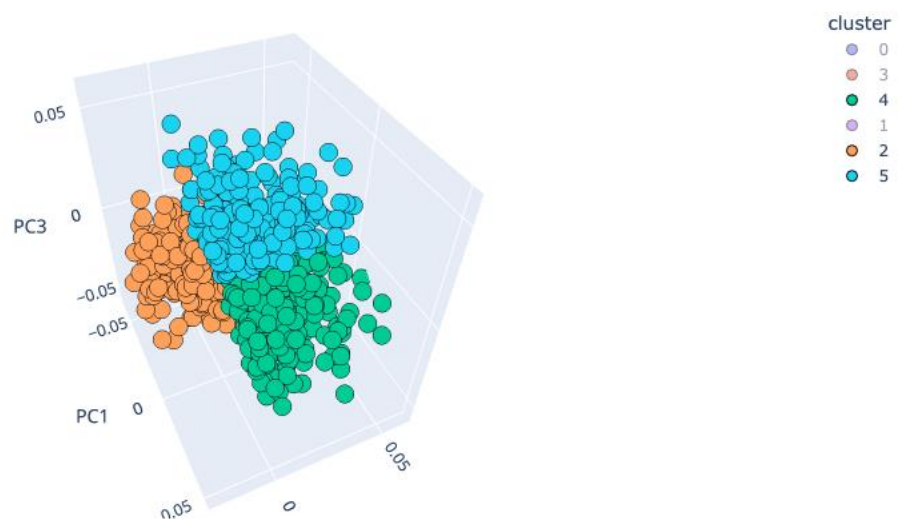

Figure 10

Figure 10 shows the back side of it, you can see how similar it is to a cube!

Section F: PCA/Topic Analysis

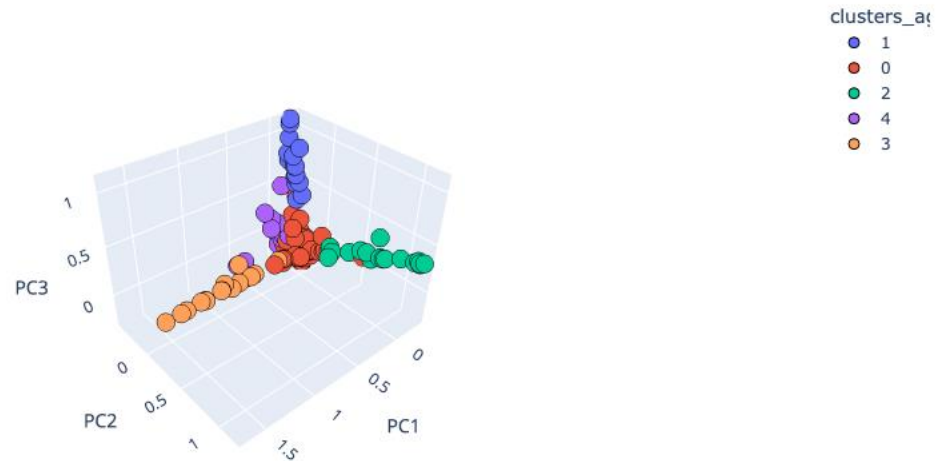YouTube Titles of Left/Centre/Right Wing UK news outlets – Topic analysis

Figure 11

PCA/Topic analysis was not a fruitful endeavour. The clusters are indecipherable, apart from the observation that it clusters the titles in accordance with their political ideology.

.

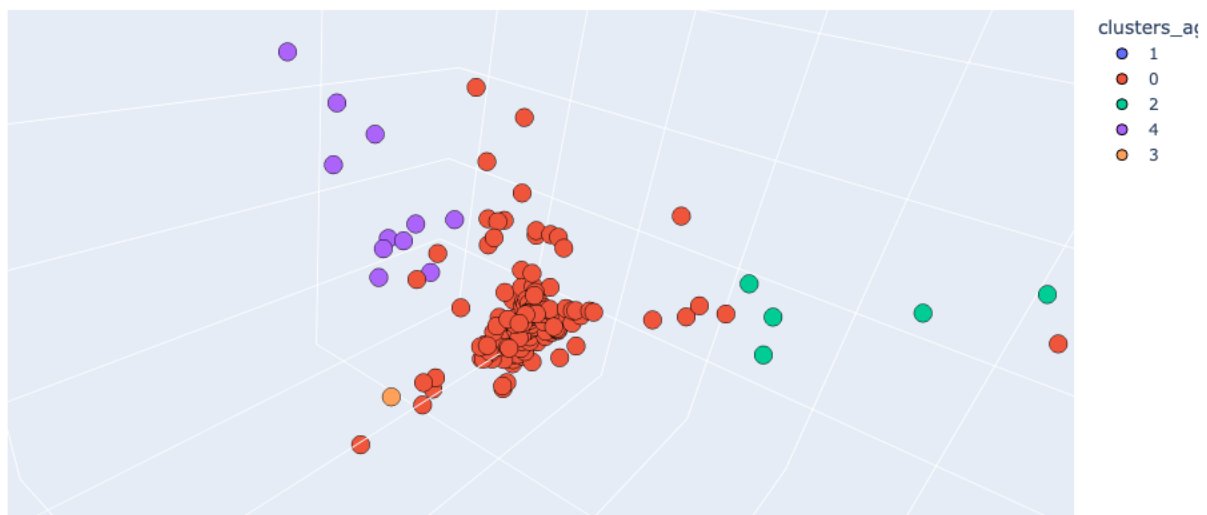YouTube Titles of Left/Centre/Right Wing UK news outlets – Topic analysis

Figure 12

Most Left/Right-Wing titles are dotted outside of the central blob. Centrist publications are practically all concentrated in the centre (Figure 12).

Conclusions

I struggled to build on top of the meaningful analysis conducted in my earlier inquiry. Whilst I am confident that my methodology is correct in the code, the results didn't tell me anything new, or that of which was new, utterly confusing. I guess that I had a small source dataset, and that this is not befitted to larger PCA analysis, however, I am still frustrated that nothing insightful has become of this inquiry.

I was also not logical in the methodology of my inquiry, several charts (punctuation/case of text) are more suited to different inquiries into the sensationalism of titles rather than their political ideology. However, I am very pleased in the novel coding techniques that this project entailed, PCA has an outstanding potential that I wish to utilize further.

Bibliography

- *Black and white and read all over: A guide to british newspapers* (2022) *Oxford Royale Academy*. Available at: https://www.oxford-royale.com/articles/a-guide-to-british-newspapers/ (Accessed: March 15, 2023).
- Munger, K. and Phillips, J. (2020) "Right-wing YouTube: A supply and demand perspective," The International Journal of Press/Politics, 27(1), pp. 186–219. Available at: https://doi.org/10.1177/1940161220964767.