Shawn Bryan Woodward.
Student Number: 190278830

Mini-Project Assignment – Part 1

Yes.

Good proposal.

My mini-project will be on the frequency analysis of 'COVID-19' in online publications from the Daily Mail and the Guardian news outlets.

I have chosen this word as it is very topical in our lives at the moment and there will be plenty of resources for me to explore as it is something that has consumed our society for almost two years. I want to investigate how the frequency of the usage of the word has varied in the past two years, but also on a monthly basis and potentially draw conclusions from this analysis. I will get an even number of articles from both news outlets and will be aiming for at least 20 from each so that I can gain a large data set to examine.

I have chosen the Daily Mail and the Guardian in order to give myself the best attempt of being comprehensive. This is because they are both believed to be on opposite sides of the political spectrum, allowing me to hopefully come close to nullifying the influence that potential political bias could play. The importance of this is paramount as I would like to try to limit contextual factors so that I can take the data for what it is and analysis it accordingly. I cannot simply pick one newspaper as I don't believe that it will provide me with enough data to draw conclusions from. I will however compare the results from each of these outlets as it would be interesting to see how they differ, but not conclude that this is from potential political bias. I am also able to gain access to these two sources very easily as they are open to view for free to the public online. In order to make them readable for my code, every article that I use for my data, I will put into separate text documents, meaning that I can use 'open file' functions.

good thoughts

And from Nexis?

My application will use two functions to form the base of my investigation. These are the Word_frequency and Count_words functions. These of course will be integral to counting how often COVID-19 appears in the articles. I will need to design these so that they can analyse the text documents. I will then need to work out how to create a function that then groups them by month and then year so that I can design different levels to my analysis. I will be running all code that I design in a python IDLE cell so that I make sure that it is able to run and make sure that it carries out its intended function. To check that it works, I will test the code on one article and once I can confirm that it does, I will then use that code on a larger data set.

Yes

or create some simple test data yourself.

I will also need to create an average function that allows me to input the data provided by the other functions and sort them by averages over a period of time, such as monthly and also by the News outlet so I can compare findings.

The python cells in Jupyter Notebooks will be useful as they will allow me to create bar charts and graphs displaying my findings. These will be showing the results of the word frequency by month, year and news outlet. This will give my report another medium to express the data that my code has produced – so it can be interpreted in a pictographic form, rather than just numbers and written analysis. After this I can draw conclusions from the data provided.

This is a nice proposal and well thought through.

P.T.O

Perhaps consider automating as much of the data input as possible with additional functions.

For example: get many extracts from Nexis, put them in files in same folder, build a function that successively takes each file through your anysis and then stores results in an output file (perhaps with links to charts as seperate files)

Just thoughts though.

Your original proposal is good.

PJ. Hill