PROJECT PART 1

Yes

Project title - 'Covid Trends' – Word frequency analysis

Description on the project – In this project I will be looking into the word frequency of 'Covid' in several different online UK newspaper articles over a period of time (most likely between 2020 – till the present time).

Outline of how you intend to approach the project: In this project I intend to source twenty-five articles (five articles for each newspaper company) from a variety of mainstream left- and right-wing UK newspapers, this will include The sun, Daily Mail, The Mirror, Evening standard and the Metro. In order to obtain legitimate text resources and data I will use the data base called Nexis UK which has over 20,000 full text sources. After sourcing my data, I will put each article in txt files so they are ready for the analysis, and I will ensure that they are all within the same folder so that no data be missing from the analysis. From this I will create a code within Juypter notebook and that will count the number of times the word 'covid' appears in each article. In order to test that my software works I will write the following: word_frequency('testMirror1.txt') changing the name to each txt filename and I will also manually count how many times the word appears within the extract to see if it matches with what the code is displaying. Once my data has been collected and I know that the code is running correctly I will then display my result in a bar graph using matplotlib within Jupyter notebook. Lastly, I will write a short analysis underneath on Juypter notebook as to why some articles have a higher word frequency than others.

or other charts. Be sure to label charts.

→ you could write a function to do that automatically
1) read all the filenames in current folder and put them in a list.
② iterate through that list calling word-freq (~) with the ~~correct~~ next filename.

Just a thought

Ag As always, be sure to get something working soon as you can, then iteratively improve it

Suggest you create some simple test data of your own. That will help you get started quicker.

Yes, all good