Paige Anderson.

Yes

A project development proposal describing the ideas and objectives you intend to explore and an outline of the software you will build. As well as descriptive text, the document you submit for this assignment can include diagrams, lists, sketches, and so on. It should include the following sections:

1. The project title
2. A description of the project and the question it will explore
3. An outline of how you intend to approach the project:
    i. The data and text resources you will need
    ii. Where you will get free access to sufficient quantities of data?
    iii. The functionality of the application you will create, perhaps with a tentative list of the functions you will build
    iv. How you will you test your software to confirm it works correctly?
    v. The software tools you will use e.g. IDLE & MS Word or Jupyter & MarkDown or something else

250-500 words.

Yes this is a good project.

'Brexit' trends

- Frequency analysis of the word 'Brexit' over time in various publications.
    - Can compare three newspapers articles about Brexit from same day across 3 times, (pick significant days of Brexit and two Broadsheets with one Tabloid and get whole newspaper as text file if possible). Label the text files as title of newspaper and year. Then can present results as a comparison.

The frequency analysis of the word 'Brexit' over time and various publications.

The project will focus on three newspapers coverage across three specific dates relating to 'Brexit' and the frequency of the word exploring the question 'Is there a difference in the use of the word 'Brexit' between the vote for Brexit and the leaving of the European Union?'.

The data I need are three articles written by three UK News publications. The three articles that are collected will be published on the three following dates: 23/07/2016, 14/03/2019 and 31/01/2020. This is because on the date 23/07/2016 the referendum was held on and the majority of those who voted chose to leave the European Union; on the 14/03/2019 the UK Government sought permission from the EU to extend Article 50 and agree a later Brexit date (rather than 29/03/2019); and finally, on 31/01/2020 the UK finally left the European Union.

I will get access of the data using the database and research website Nexis. The website Nexis provides news and business information from a range of sources, including UK national and regional newspapers. Due to it being a database it is possible to search groups of sources (for example, UK Broadsheets, Major World Newspapers) or an individual publication. I will filter by

country of publish (UK) and dates (previously stated) and pick the first three results made by The Guardian, The Independent and The Sun. I will separate the articles by saving them as separate text files titled by publication and date. Nexis' coverage dates from the 1980s and is updated daily. I will collect the data by searching the articles online and then creating them as separate text files by both year and publication.

The functionality of the application I plan to create is to be widely used as a tool to help with word frequency comparison. By creating this application, I will have to create functions such as a word count that can recognise word boundaries, including those followed with a space and/or punctuation. I will also have to make sure that I can make the article into a list of words and their occurrence, so I am able to compare the frequency of 'Brexit' across all 9 text files.

To test my function, I will have to execute the code multiple times using Idle and a test text file that I have created myself. In this file the word 'Brexit' will occur a certain number of times to which I will be able to compare the results returned by the code and their actual occurrence with. To do this I will have to count the occurrences of the word myself in the text file beforehand. To overcome any problems with my own counting, I plan to use the list function to return all the words used in the text file with a numerical value. Something which I have previously coded and has passed multiple tests.

For this project I plan to use the software Idle as the main software. This is because I find it easier to read my syntax errors on this software. I plan to use the software Jupyter to create a bar graph to present a comparison of the word frequency across the time and newspapers analysed.

or make your own simple + short test files.

Once you have an early version working you can then perhaps expand the amount of data it can process. If nexis exports the articles in separate files then ~~python~~ you should be able to write a function to read all of those files into your 'Brexit' functions automatically. And explore graphs. Barcharts are good, try a few others as well just to compare which present data best.