

Coding Mini Project Part 1:

Yes

A project development proposal (500 words) 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:
  1. The data and text resources you will need
  2. Where will you get free access to sufficient quantities of data?
  3. The functionality of the application you will create, perhaps with a tentative list of the functions you will build
  4. How will you test your software to confirm it works correctly?
  5. The software tools you will use e.g. IDLE & MS Word or Jupyter & Markdown or something else
4. An outline project plan

Great project proposal.

The project proposal for this assignment will investigate the language used by individuals when reviewing products on Amazon, or more specifically how the language used in Amazon product reviews reflect on the product itself.

Since its founding in 1994, Amazon has become one of the largest (if not the largest) online retailing company in the world. Although it specializes in more than just e-commerce, in this project I will be focusing more on the e-commerce side of the business. Amazon's reputation has been unstable throughout the years, with many claims on its handling of bad products and inhumane working conditions but is still frequently used by many customers around the world. The project that I will be researching will investigate specific products on their website ([www.amazon.co.uk](http://www.amazon.co.uk)) and look at the language used to describe the reviews of the product and whether the reviews are more positive or negative.

To collect the data, I will be using the official Amazon website ([www.amazon.co.uk](http://www.amazon.co.uk)) to search up specific products (clothing, technology, homeware, children's toys etc.). Then when selecting a specific product, I will look at their ratings and reviews and select a range of comments that provide different views on the product to then input into my code to see if the reviews correctly relate to the number of stars that the reviewer has given for the product. When selecting the data that I will need to include a significant number of reviews, ranging from 1-5 stars per product (depending on the choice of product I will check the reviews for).

Look for free s/w that does this

If the code is successful, this could be used to categorise good and bad reviews for products and may even be used on a broader scope for generalized good and bad comments on businesses and could even be used to hide particularly bad/critical reviews or comments made by the public. To create this code, I will need to categorise terms that are consistently positive and negative and will need an input() function to copy and paste the reviews that I will be investigating into the code. To represent the positive terms, I could need to create a list of positive terms that include 'good' and 'great' as well as phrases such as 'would recommend' and 'I am happy with this product'. The reverse list will be made for negative

terms 'bad', 'unhappy', 'not good' and phrases 'I would not recommend this product' and 'not worth it'.

To test the software, I will have to check without using code whether the reviews will be positive or negative. This could also be done by checking the number of stars given for a review to deem whether it's more positive or negative. There could be instances where the review is neutral and could be difficult to determine whether the review is more positive or negative overall so a 'neutral' or 'unknown' result may also have to be included into the code. The software tools that I would use to help create this code will be Jupyter Notebook (to test the code and make sure that it's functioning correctly) and Microsoft Word (to place all the Amazon reviews into one document that can be easily accessible and so I wouldn't have to constantly go back onto the Amazon website to look for the product and find the review again if I lose them).

- ① There is, I think free s/w on web for <sup>prefer text and csv</sup> ~~scapping~~ files. info from Amazon pages. - would save you a lot of time. Have a look at Pypi and GitHub as well ~~as~~ as Google search.
- ② Get the basics working with cut & paste techniques <sup>first</sup> ~~but~~. Then try and automate that with s/w tools and your own Python code
- ③ A graphical plot of info is always the icing on the cake. Not sure what would be best but try a few charts and compare.

Yes this is a good project proposal