* 1. What you learned in the ICP

What I learned from this ICP was how to clean up my data. Few steps had to be considered before it could be analyzed. One for the first things I learned was how to remove unnecessary columns of data. Then learning how to remove Twitter handles, punctuations, and special characters . By following these steps, I was able to shrink the data size and only have the information I would need. Few other concepts were also applied like tokenization. Tokenization is breaking up sequence of strings or a sentence into just single words, symbols, phrases, or any other element that could be present. Also, lemmatization was applied to the data. The purpose of lemmatization is to reduce a word into its original dictionary root. Another concept I learned was TF-IDf which stands for “Term Frequency – Inverse Document Frequency”. The technique is used to quantify the words in a document and then compute the weight of each word, therefore signifying the importance of the word. The concept of POS tagging was also utilized to understand the concept. Also, learning how to check for missing values, and training, testing, and splitting of data. Few data visualization and analysis were also learned.

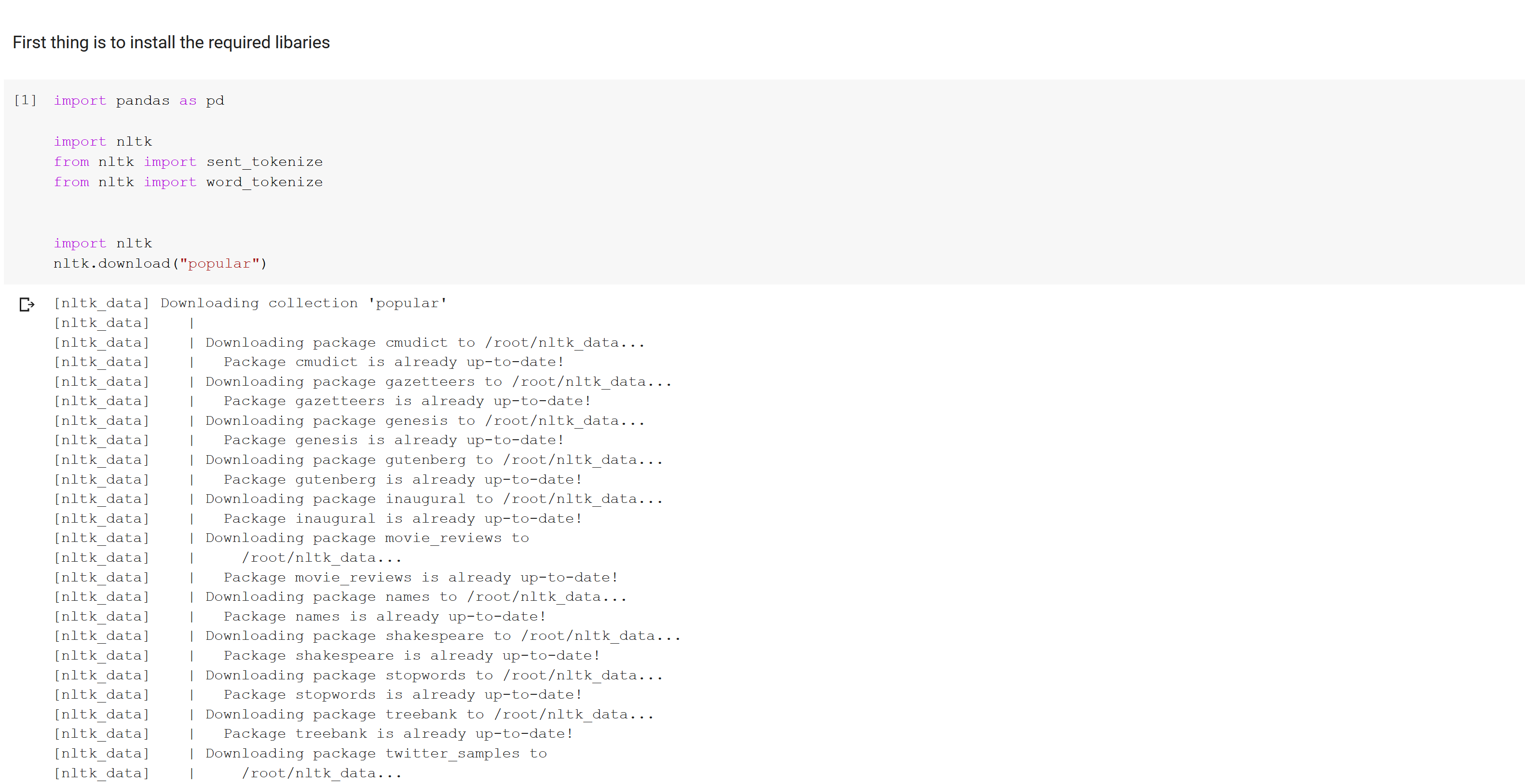
* 1. ICP description what was the task you were performing

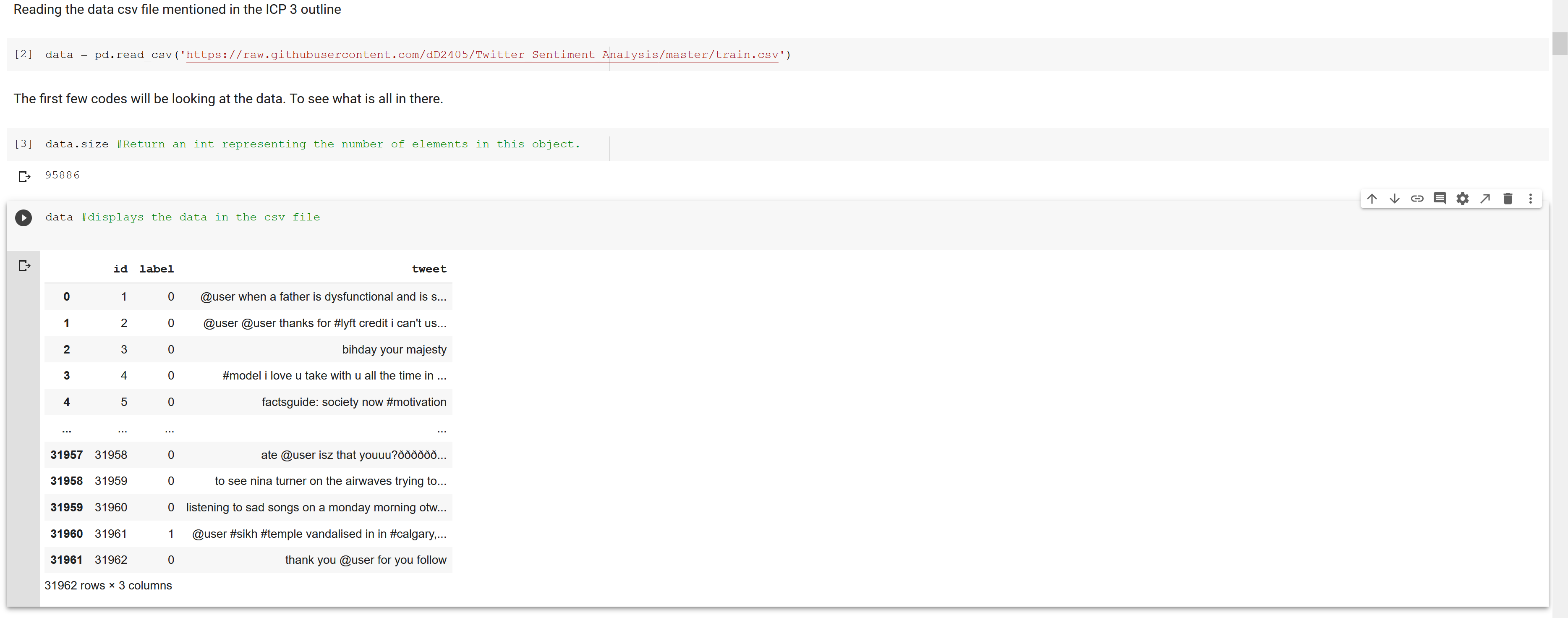
The following is general outline of the tasks performed.

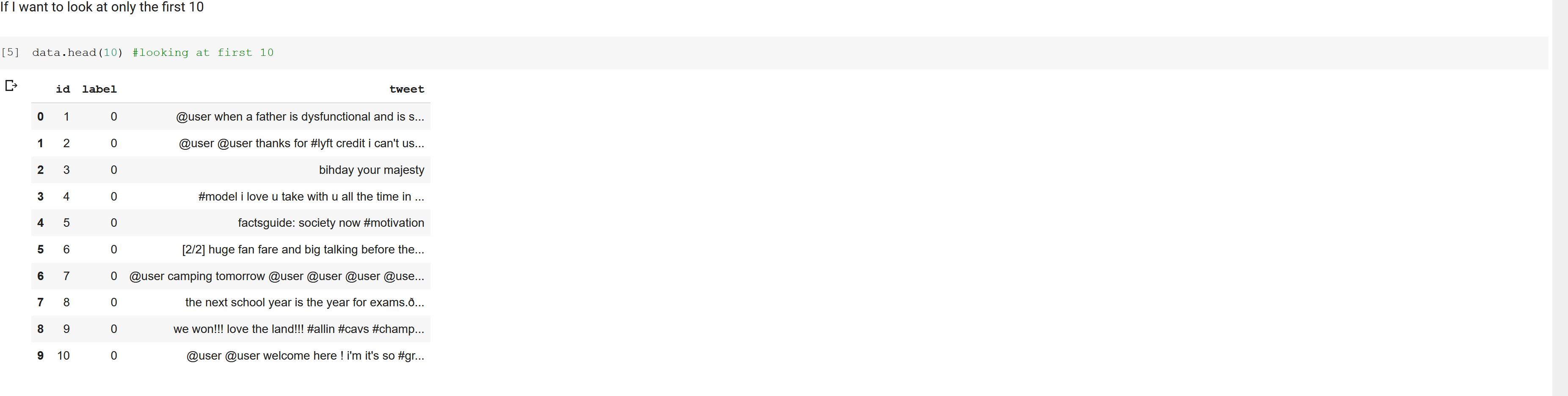
1. The required libraries were imported, and after doing so the data was read.
2. After reading the data few commands were used to see the basic information about what was in the data. EX: data.head(10) to see the columns, labels, and ID
3. Since the id column was no use it was removed using (data.drop() function ).
4. After successful completion of column removal, the first 20 tweets were converted into string. By doing so removes any ambiguity in the data. Therefore, that entire data is one data type, and making it easier to work with.
5. Then any special charters or elements that were present were removed from the tweets. The reason for doing so enables us to shrink the data size and remove unwanted data which as no meaning. A frequency method was applied to see the most common and then plotted.
6. Stop words were also removed which do not add much meaning.
7. After removal of stop words stemming and Lemmatization were implemented.
8. POS tagging and TDIDF method were also applied to the data
9. Performed a classification test to see the weighted avg of the data.
10. Checked for any missing data via the isnull() function.
11. Lastly, data visualization was applied using word cloud and bar plot and linear graph.
    1. Challenges that you faced

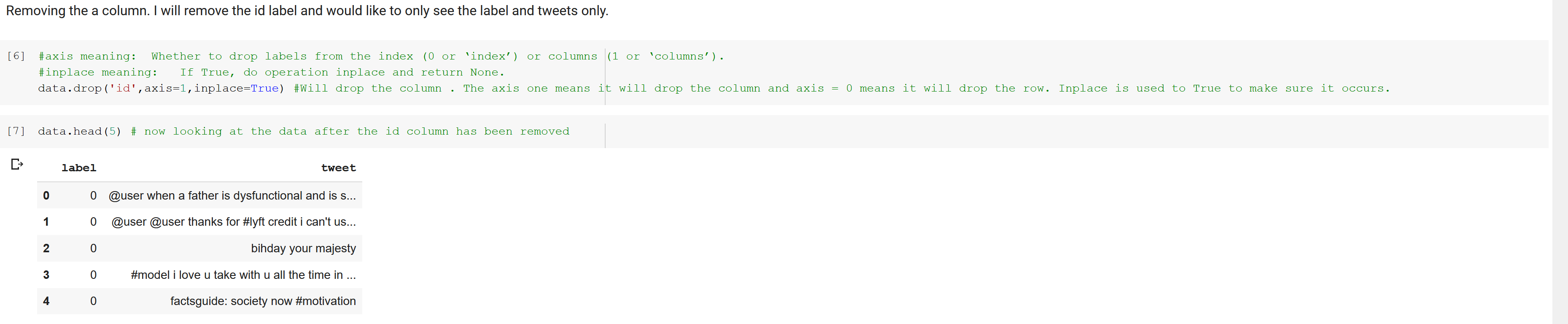
One of the main challenges that I faced was how to apply a predictive modeling to my data. Its something I am not very familiar with and hope to learn more about it as the semester goes on.

* 1. Screen shots that shows the successful execution of each required step of your code

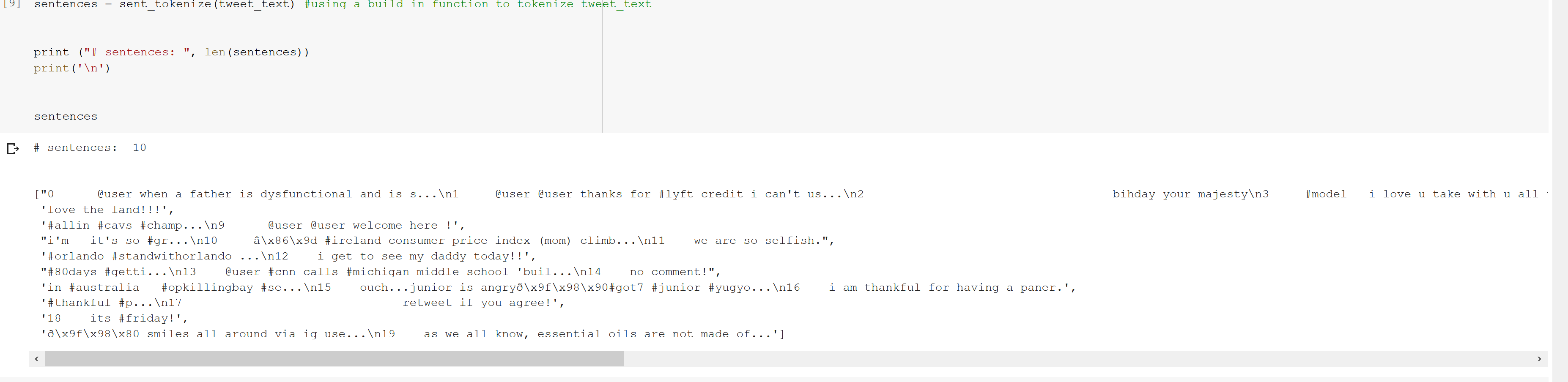


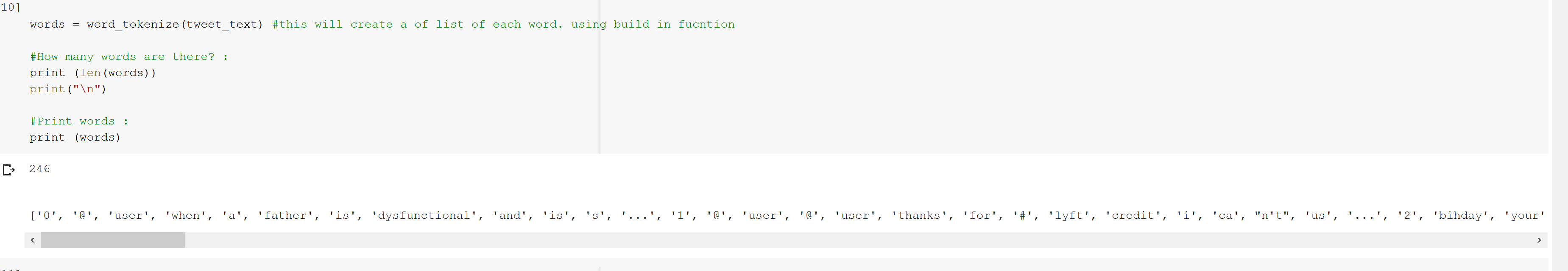


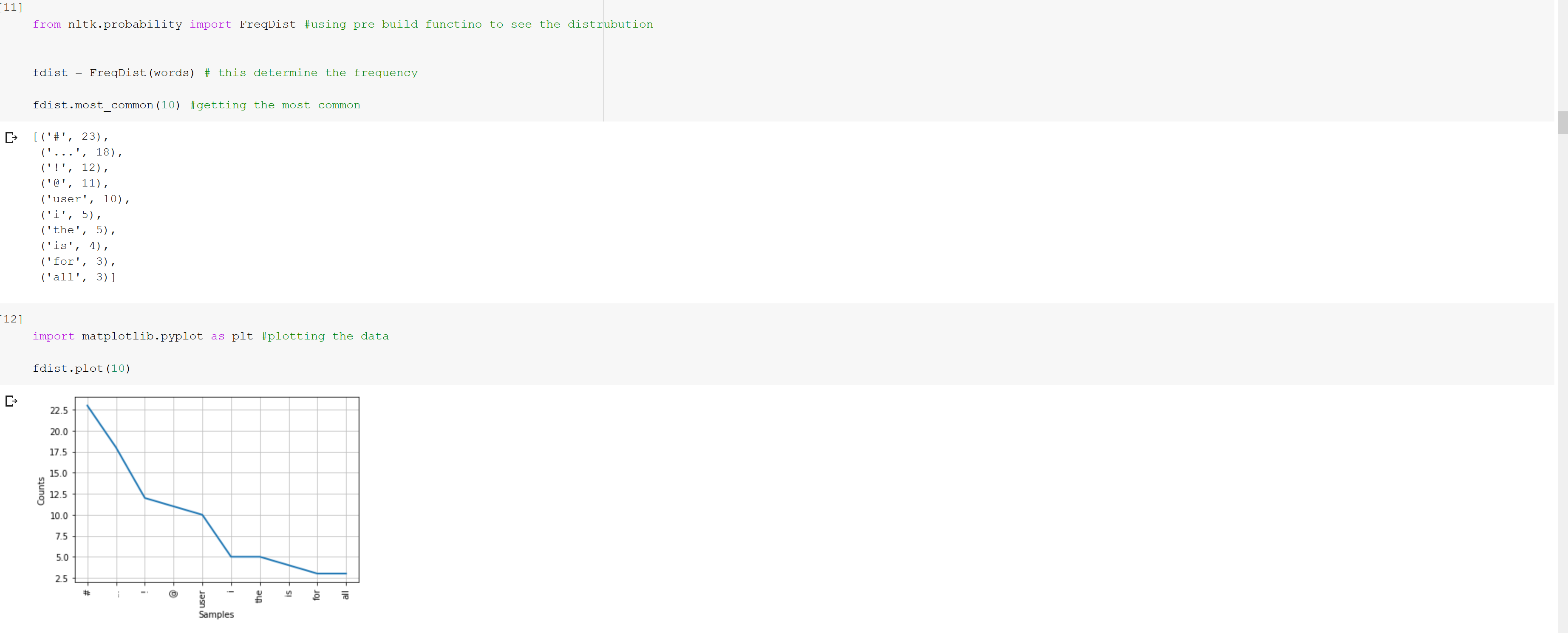




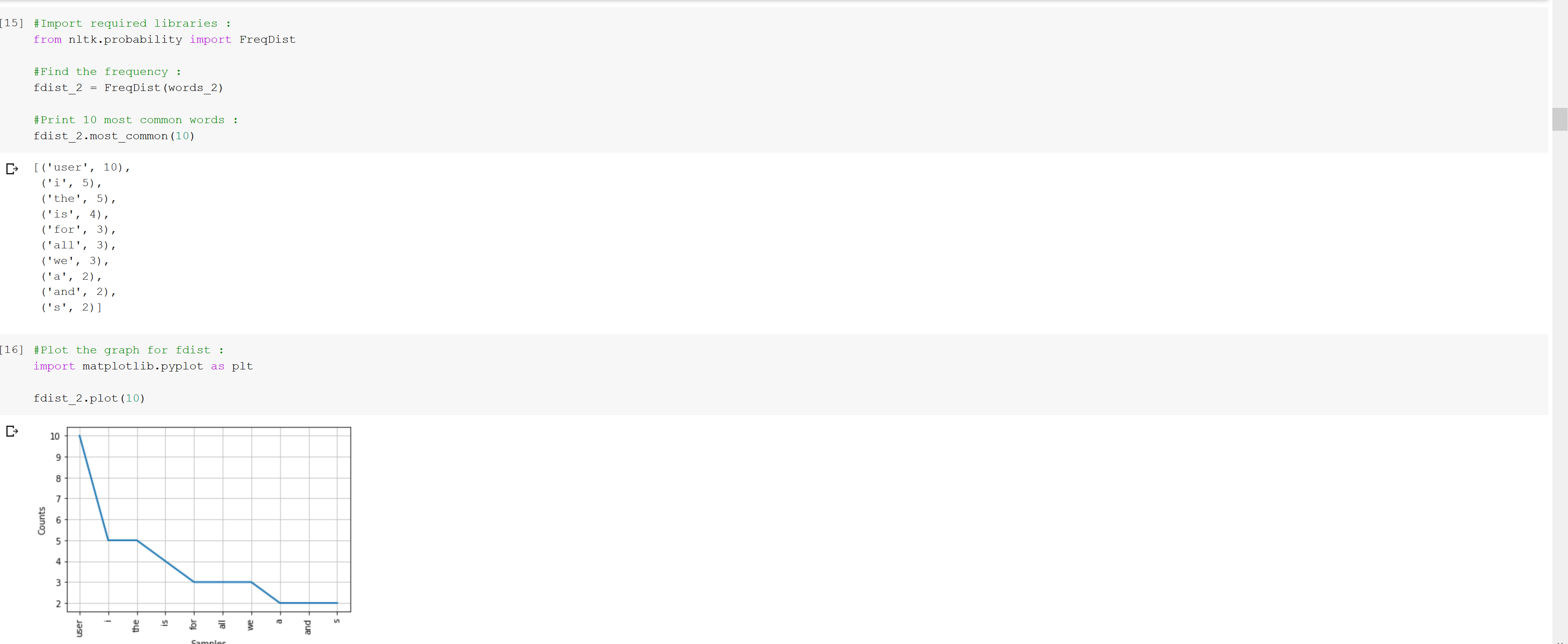


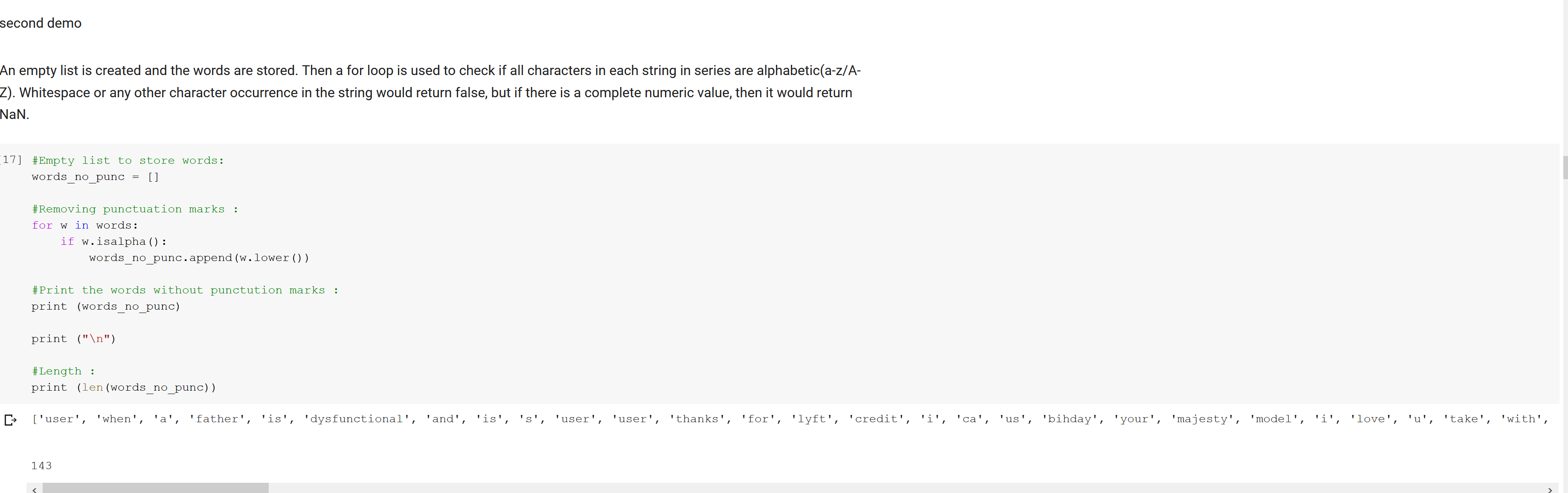


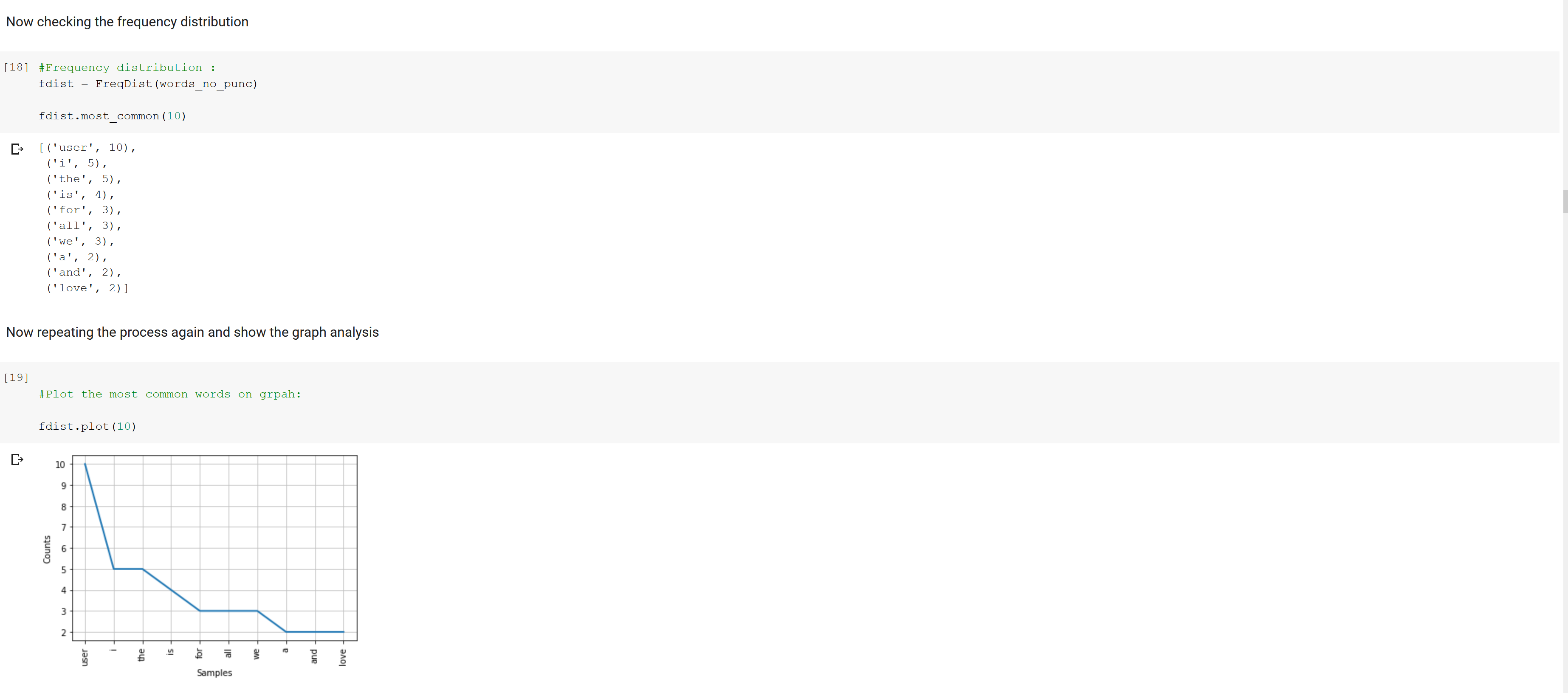


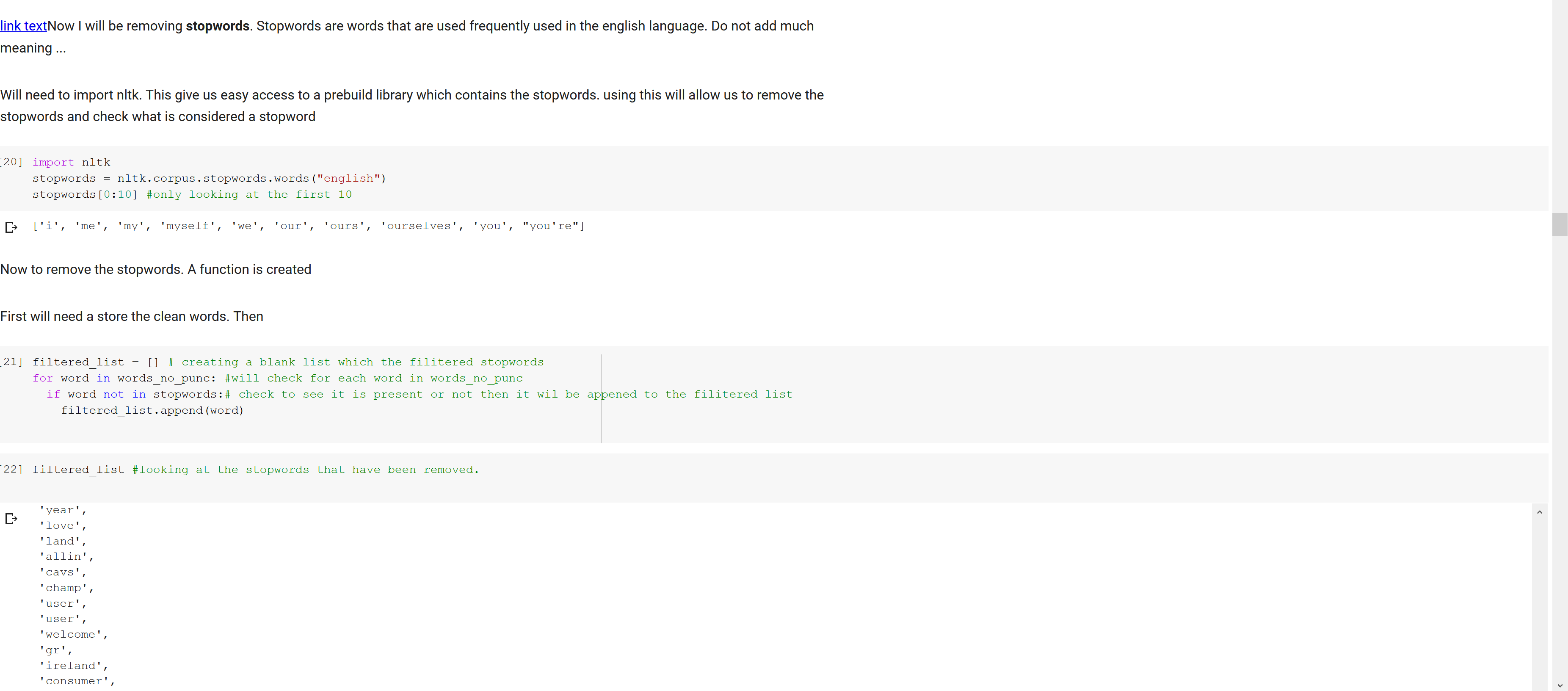


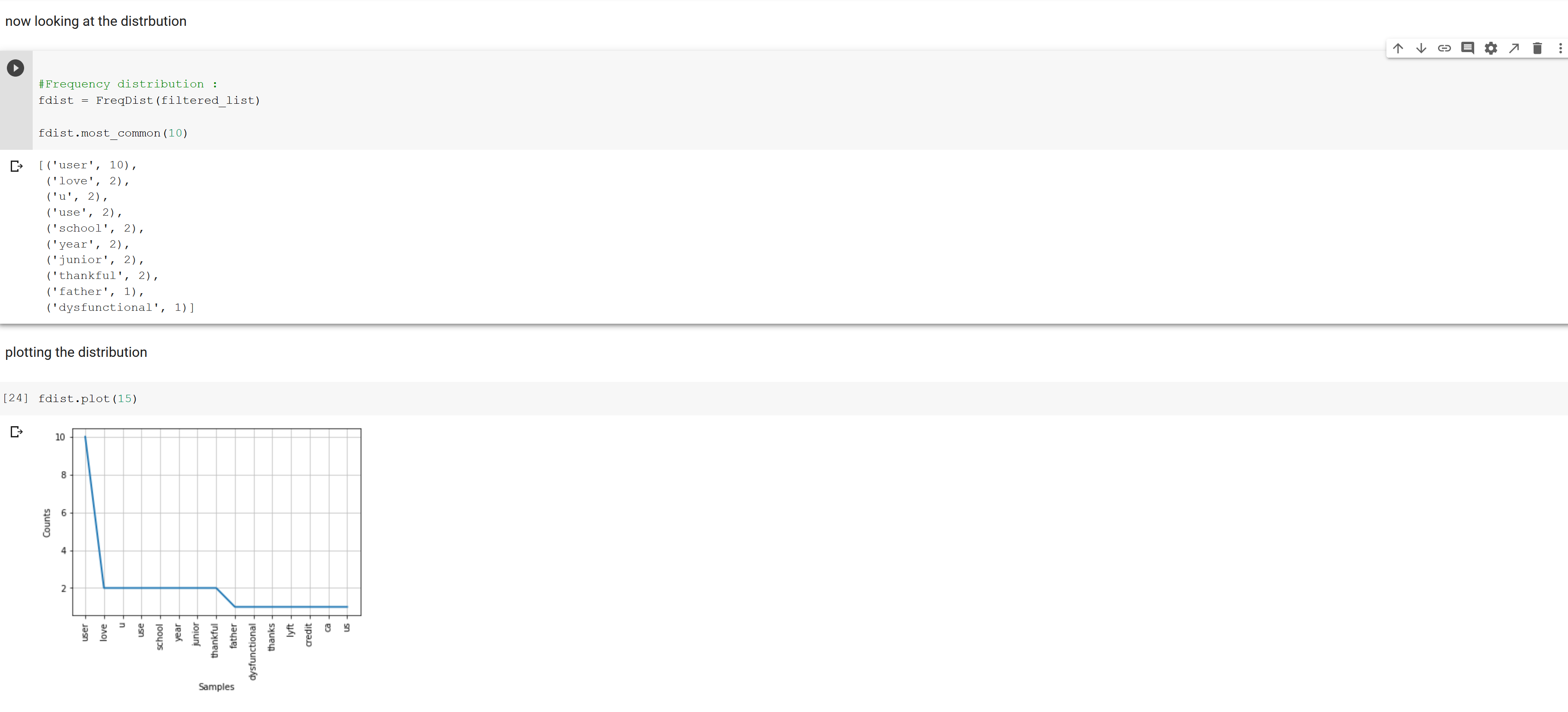






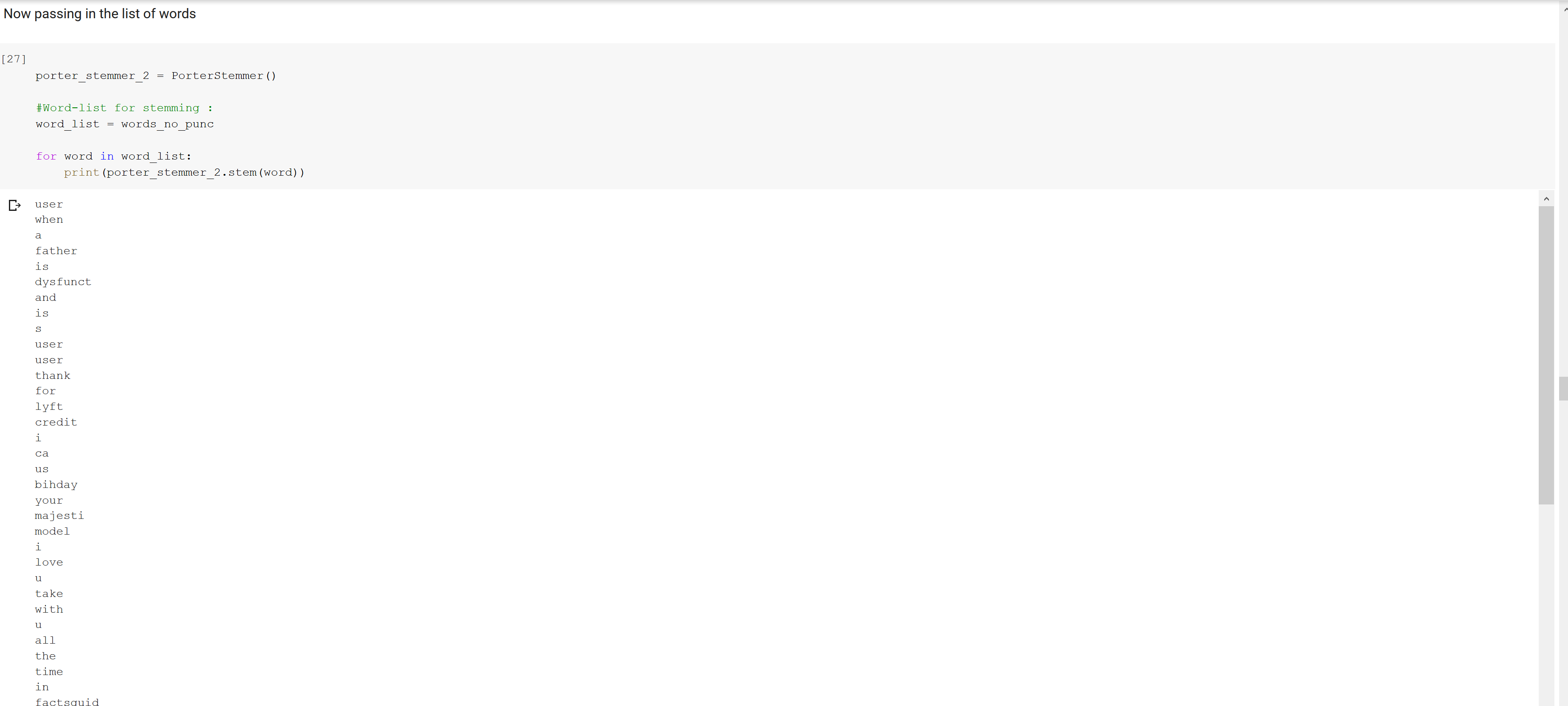


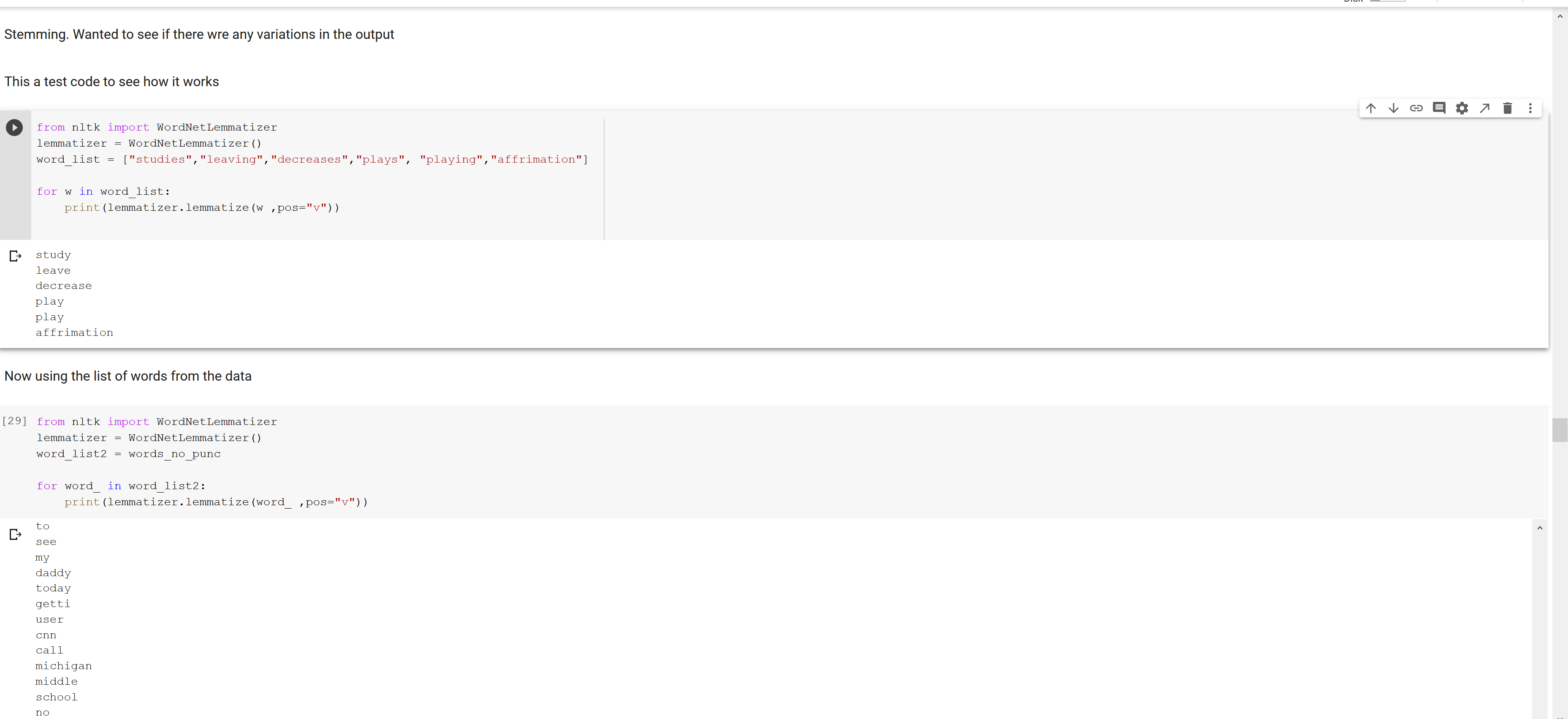


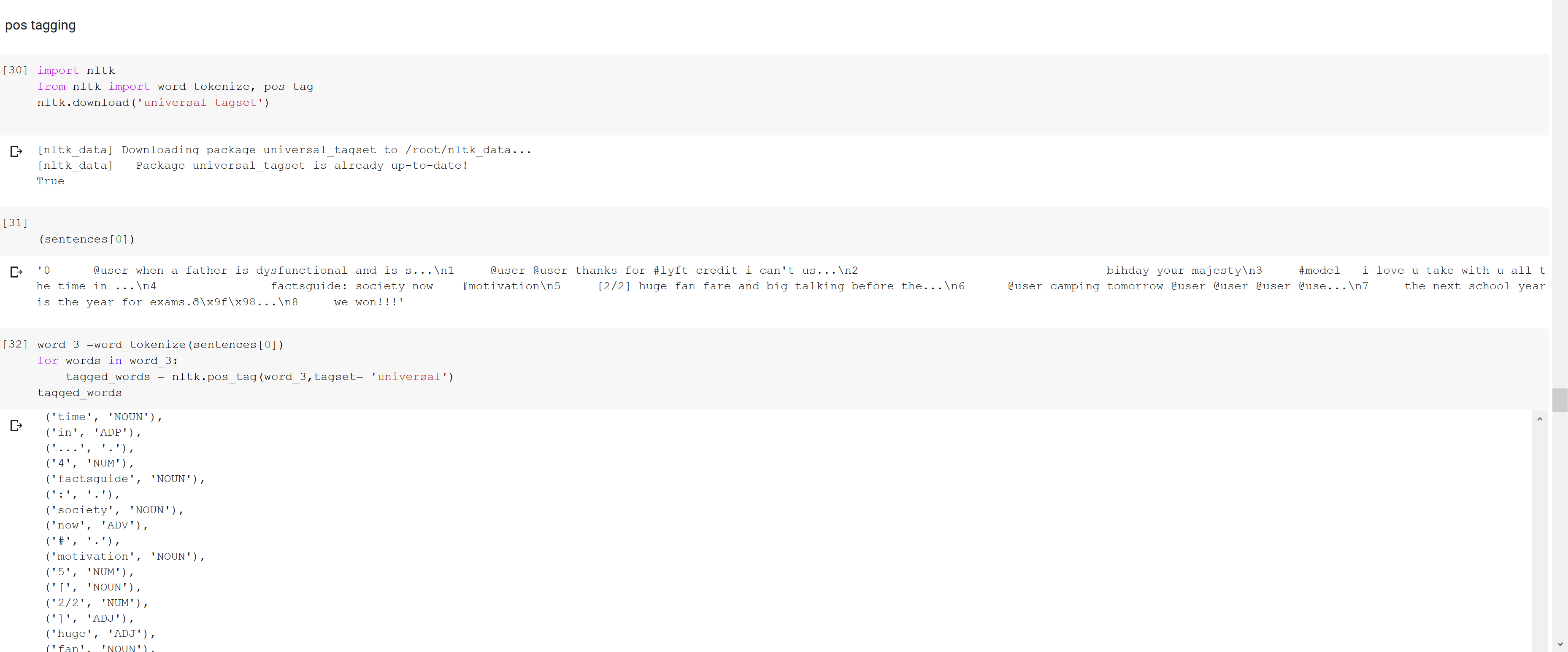






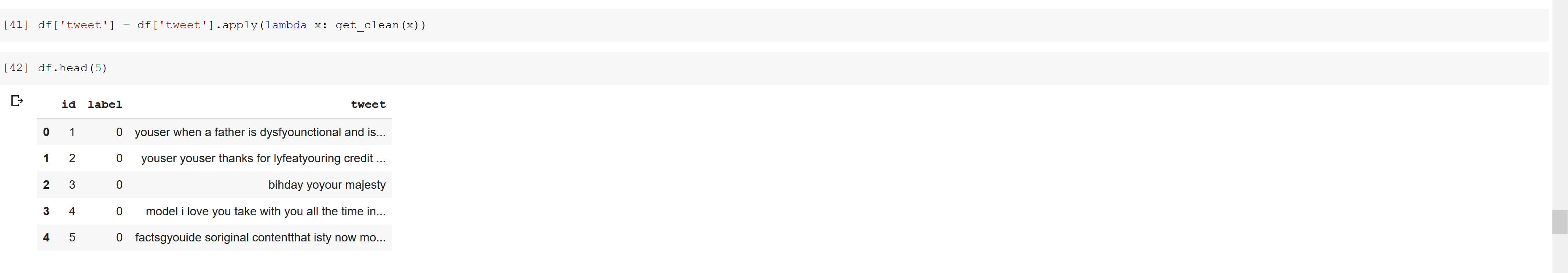


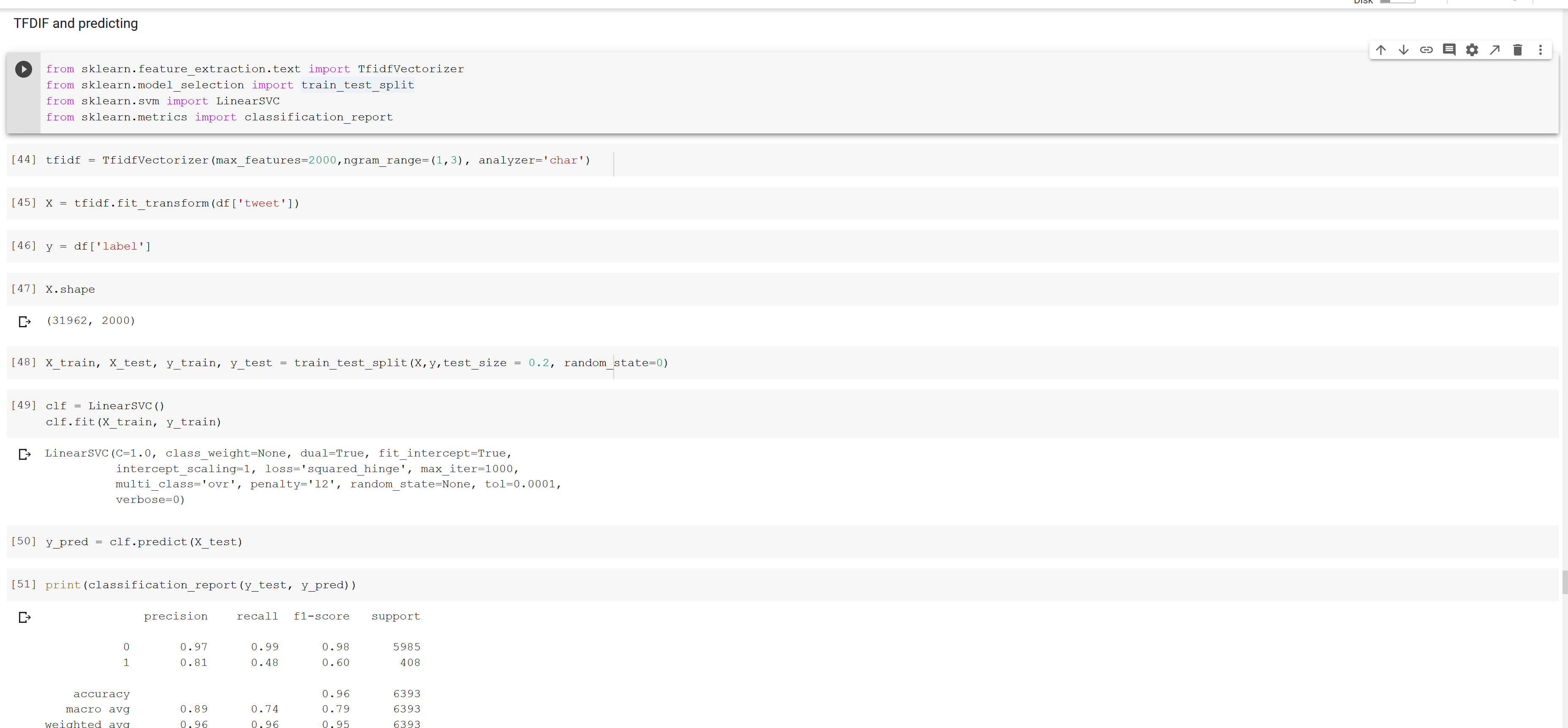






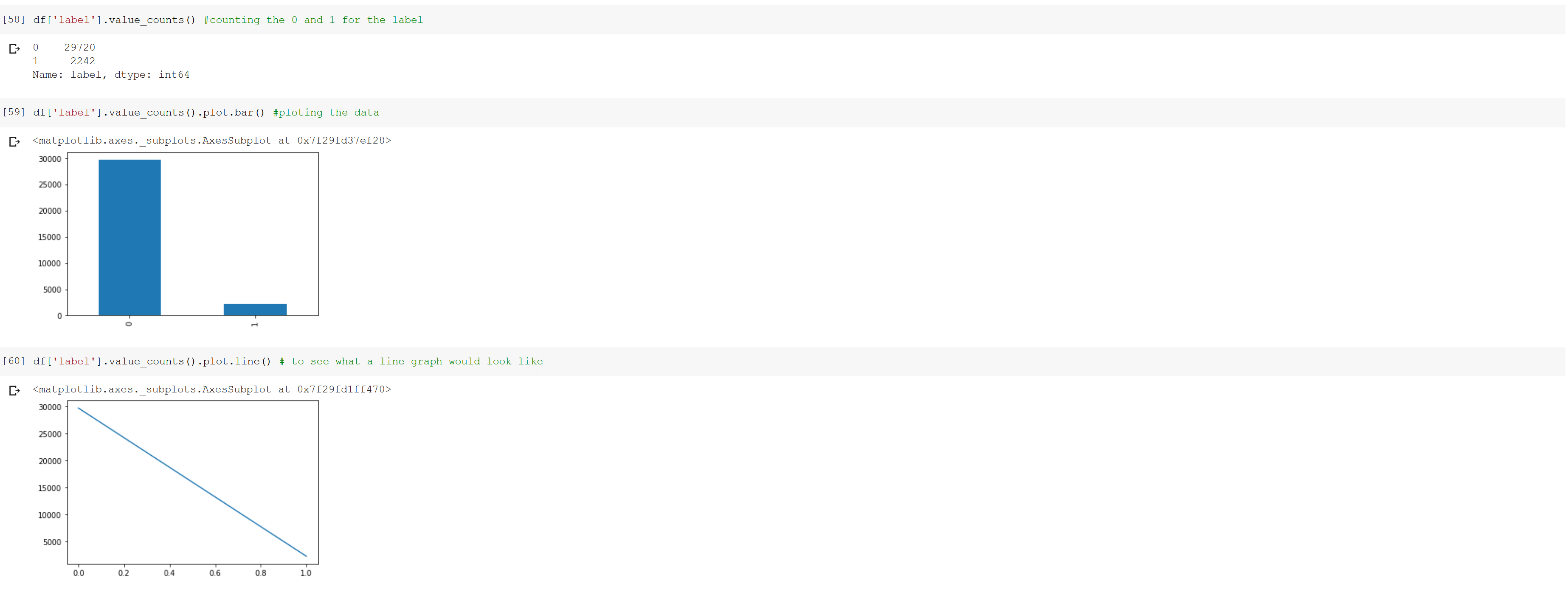












* 1. Output file link if applicable
  2. Video link (YouTube or any other publicly available video platform)

The link should. I opened the access where anyone with the link can view It. If counter any issues please let me know and I will change the setting to where you can view it. Thank you.

<https://umkc.hosted.panopto.com/Panopto/Pages/Viewer.aspx?id=fab7e9f4-a88c-4249-9f6c-ac37002bf092>

* 1. Any inside about the data or the ICP in general

I really enjoyed working on this ICP. The lecture was very helpful in understanding the core concepts, which allowed me better grasp the topics like stemming, tokenization, and few other discussed for this ICP. I think the due dates should be extended to this timeframe for future ICPs, this allows me to better understand the data and method I am working with. I talked to some of my classmates about this and they think the same. Just a consideration thought I add. Thank you