**NAME: CLAIRE NDOFOR**

**COURSE: KNOWLEDGE AND DISCOVERY MANAGEMENT**

**ICP 1**

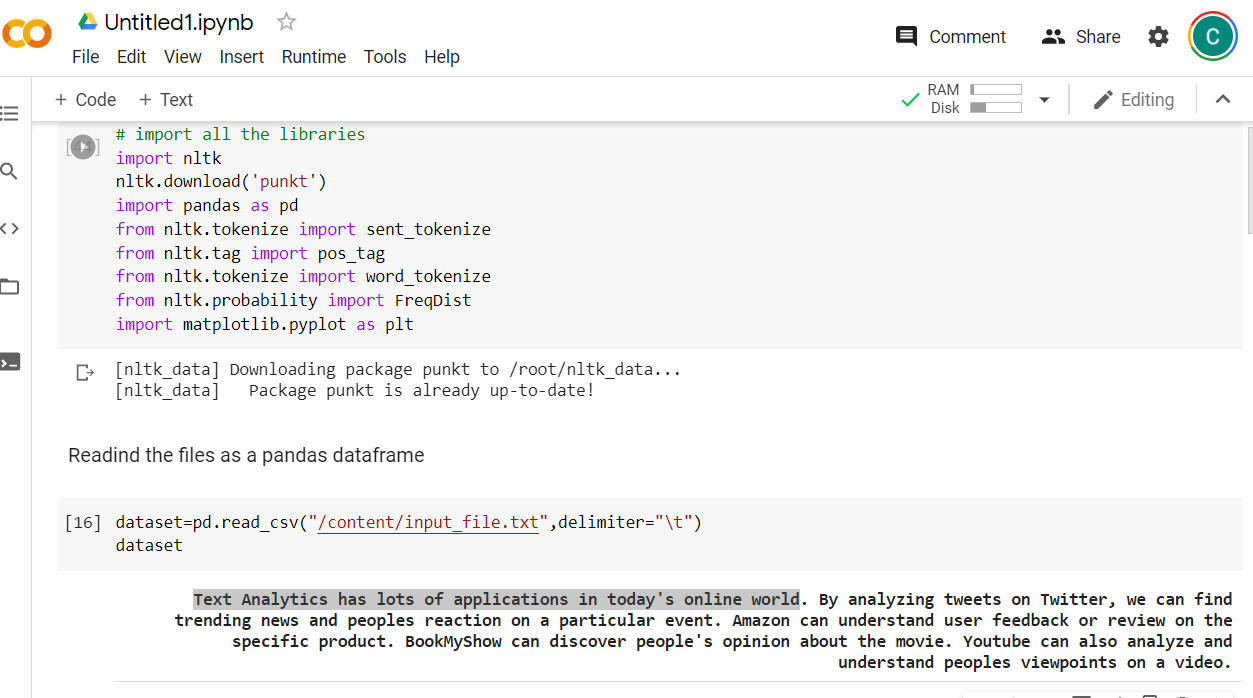
**WHAT I LEARNED IN THIS ICP**

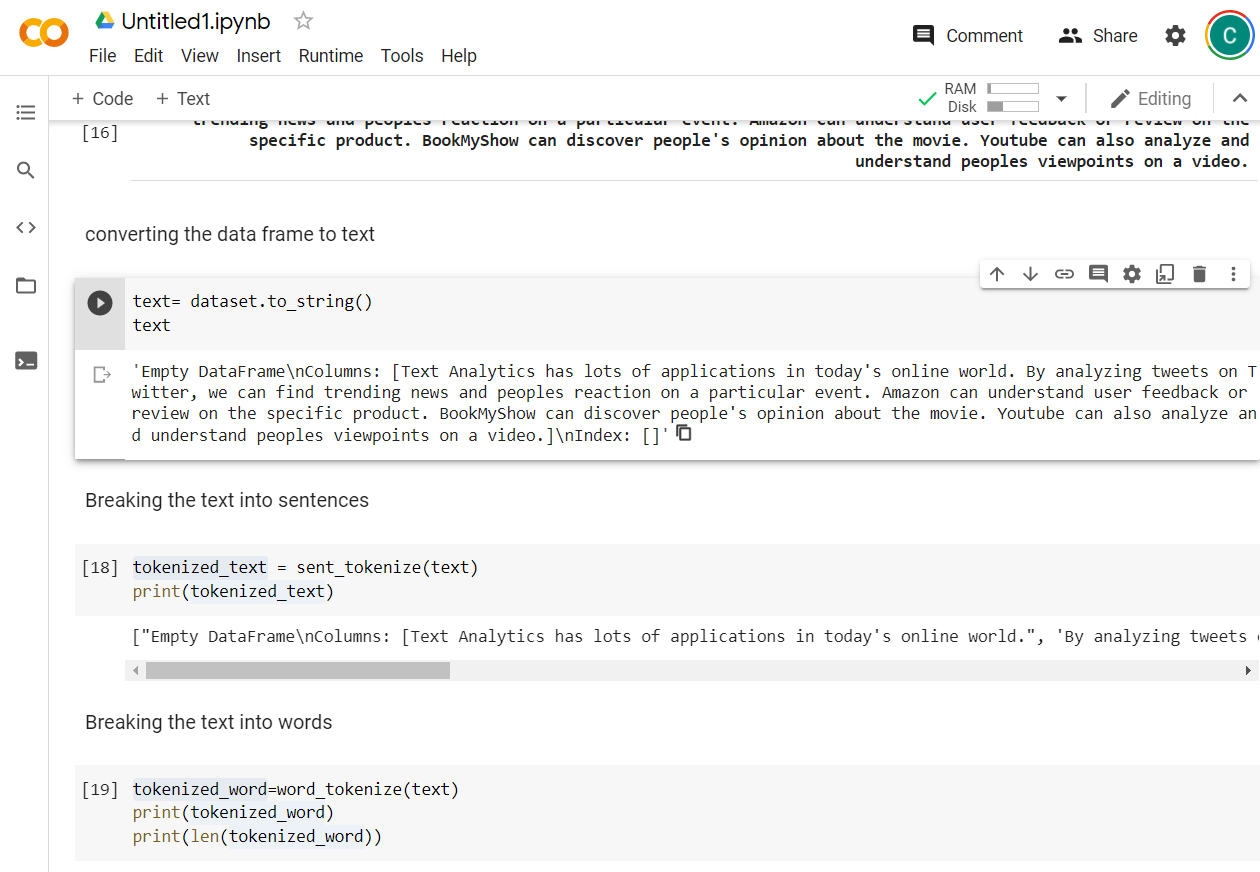
In this ICP, I learned about the fundamentals of big data and how it is made up of structured and unstructured data

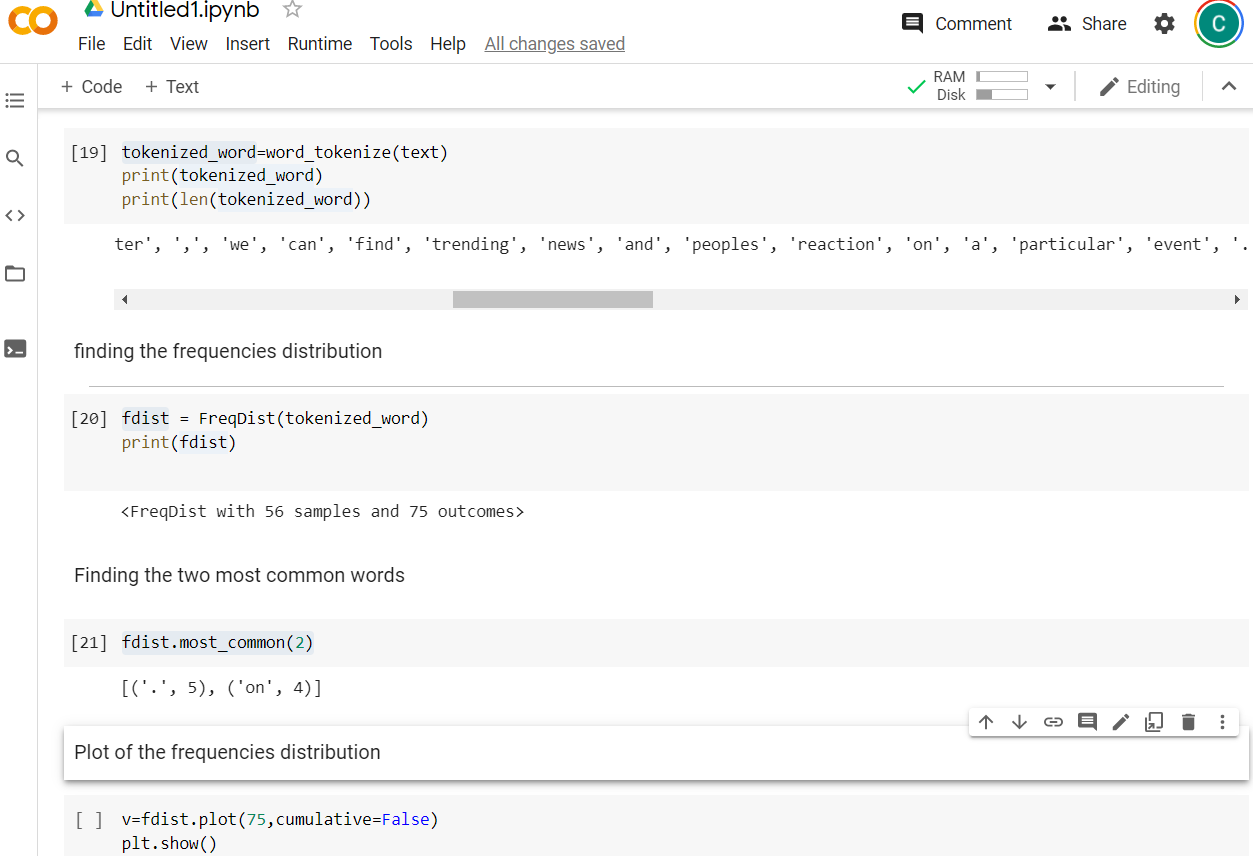
i also learned the differences between them. Moreso, this icp was focusing on some few data analytics tasks applied to a text

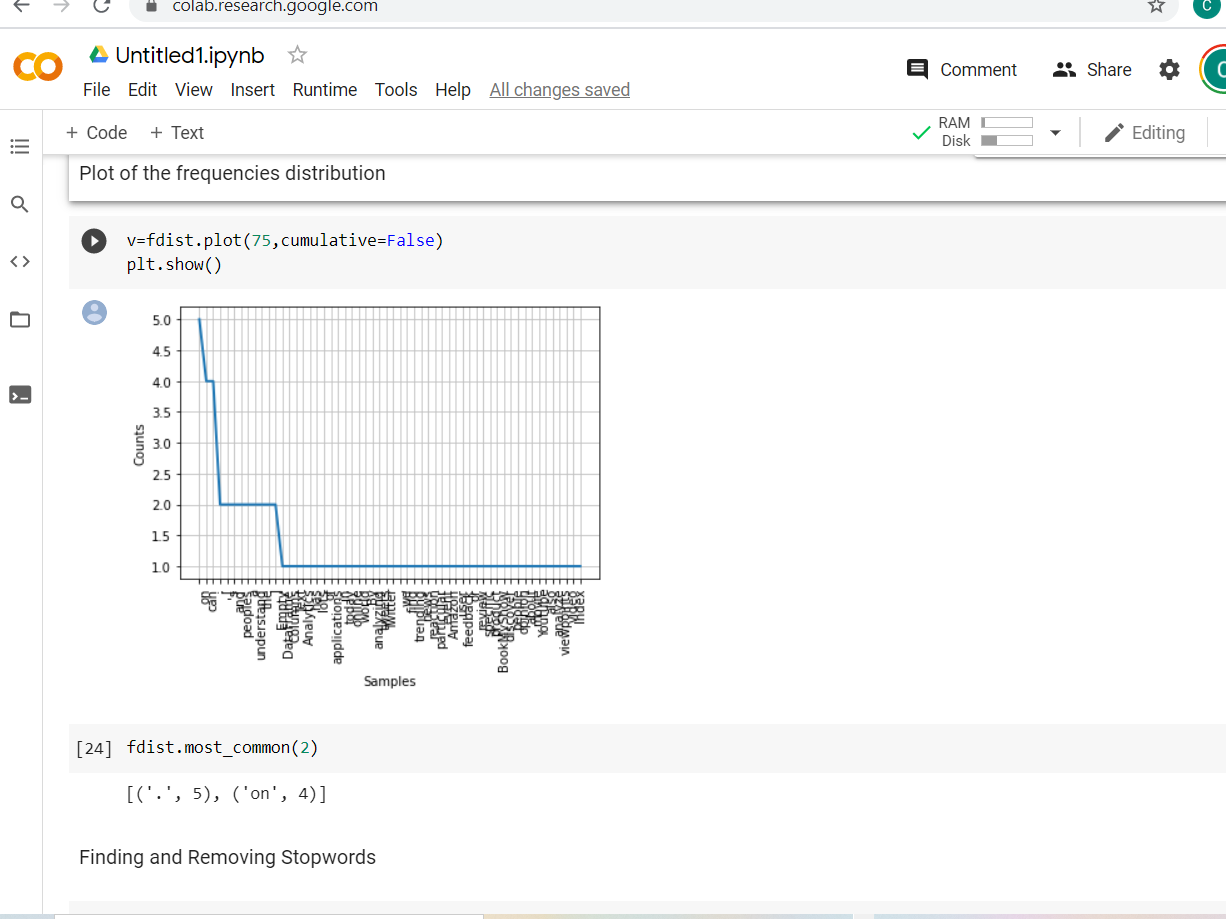
like stemming, lemmatization, parts of speech tagging and Stopwords.

**SCREENSHOTS OF CODE IN GOOGLE COLAB**









**DESCRIBTION ON WHAT I DID FOR THIS ICP**

This Icp required 3 data analytics tasks to be performed on a text provided.

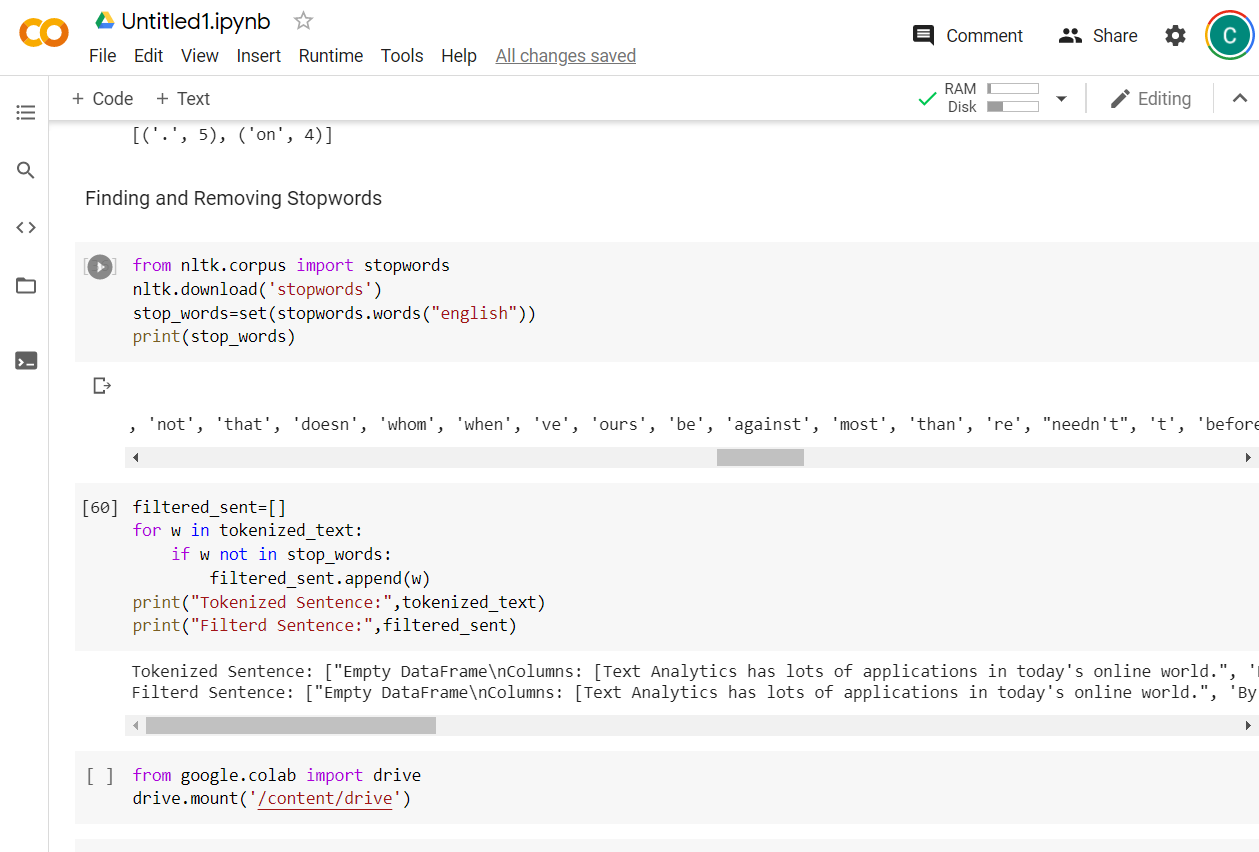
i performed the following tasks

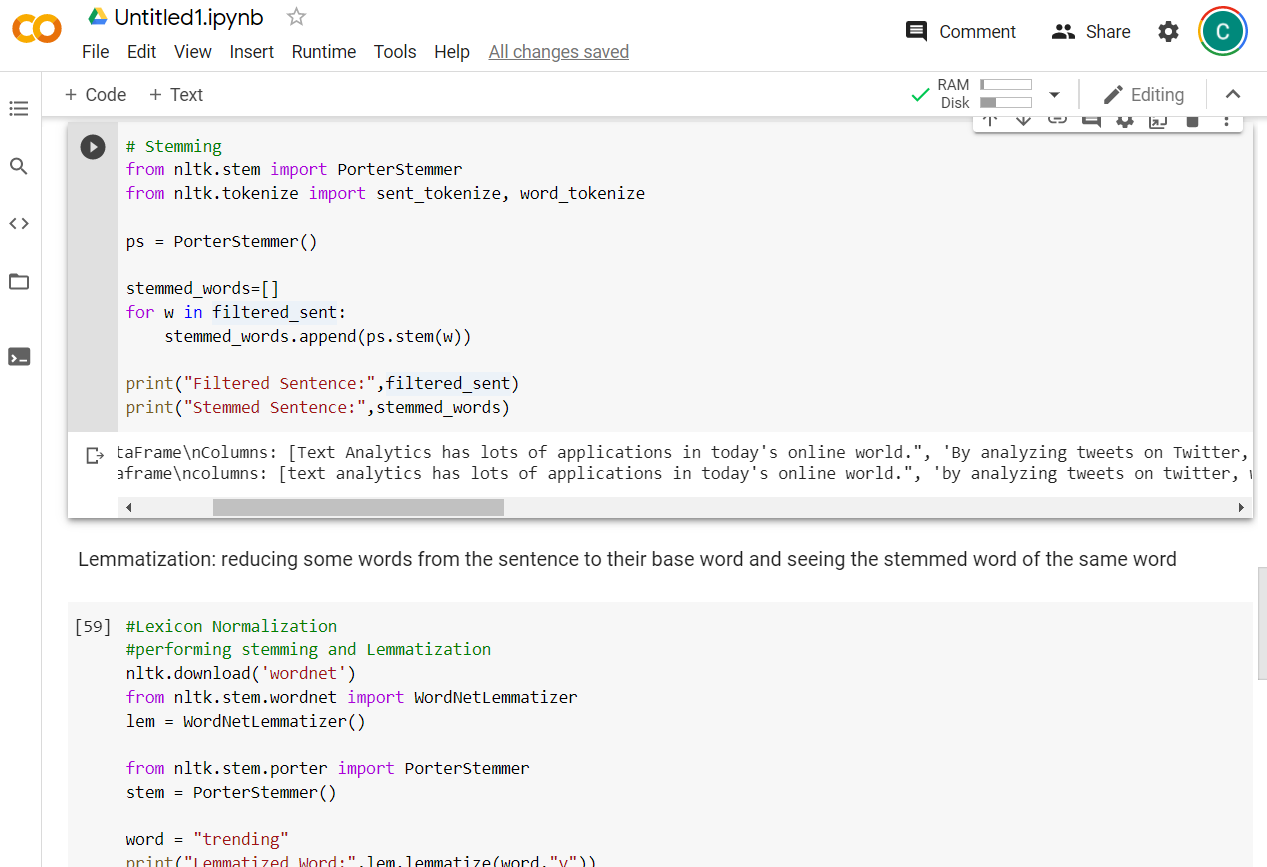
**Lexicon Normalization**(stemming and lemmatization): this is another way of reducing noise in a text. it helps to reduce words. so i applied stemming and lemmatization to some of the words from the text

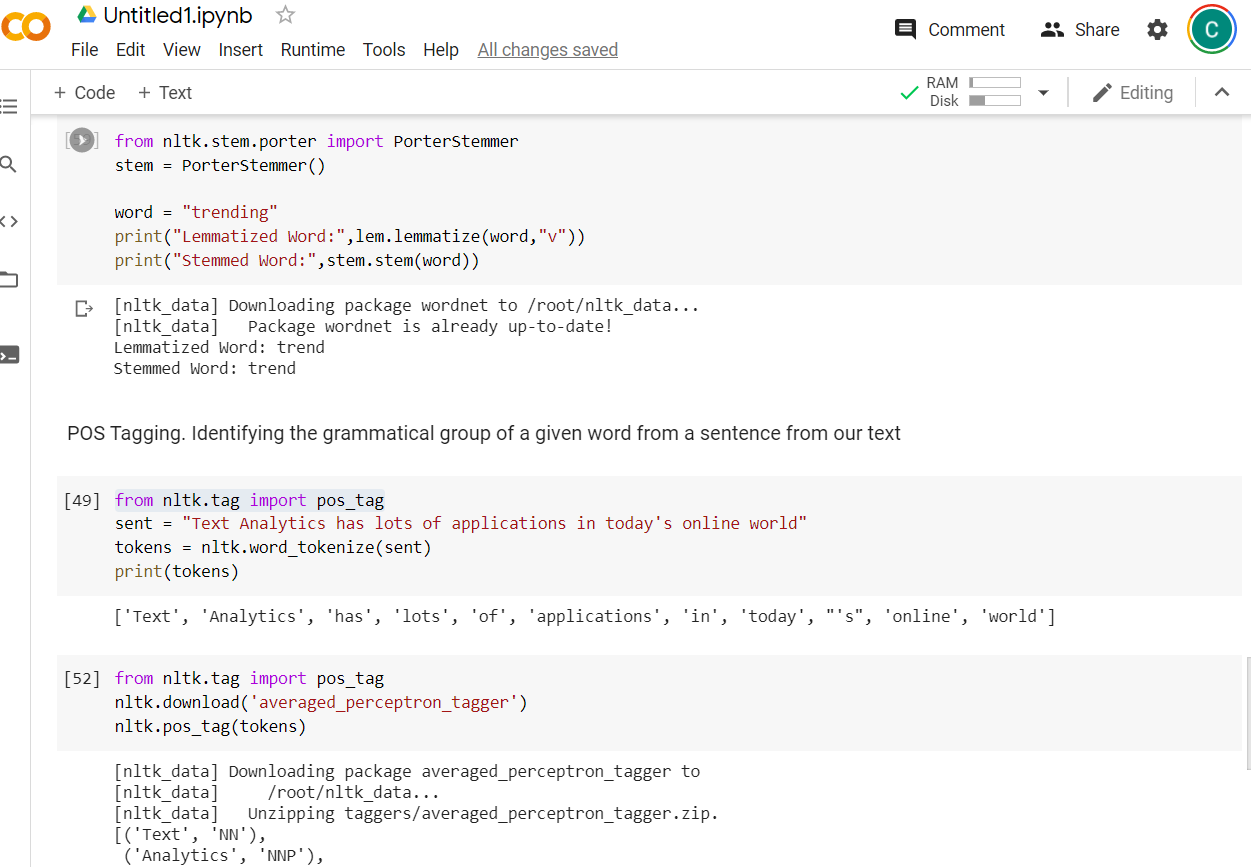
**stop words**: stop words are words like am, are, is a , an , the etc and this words are referred to as noise in a text so

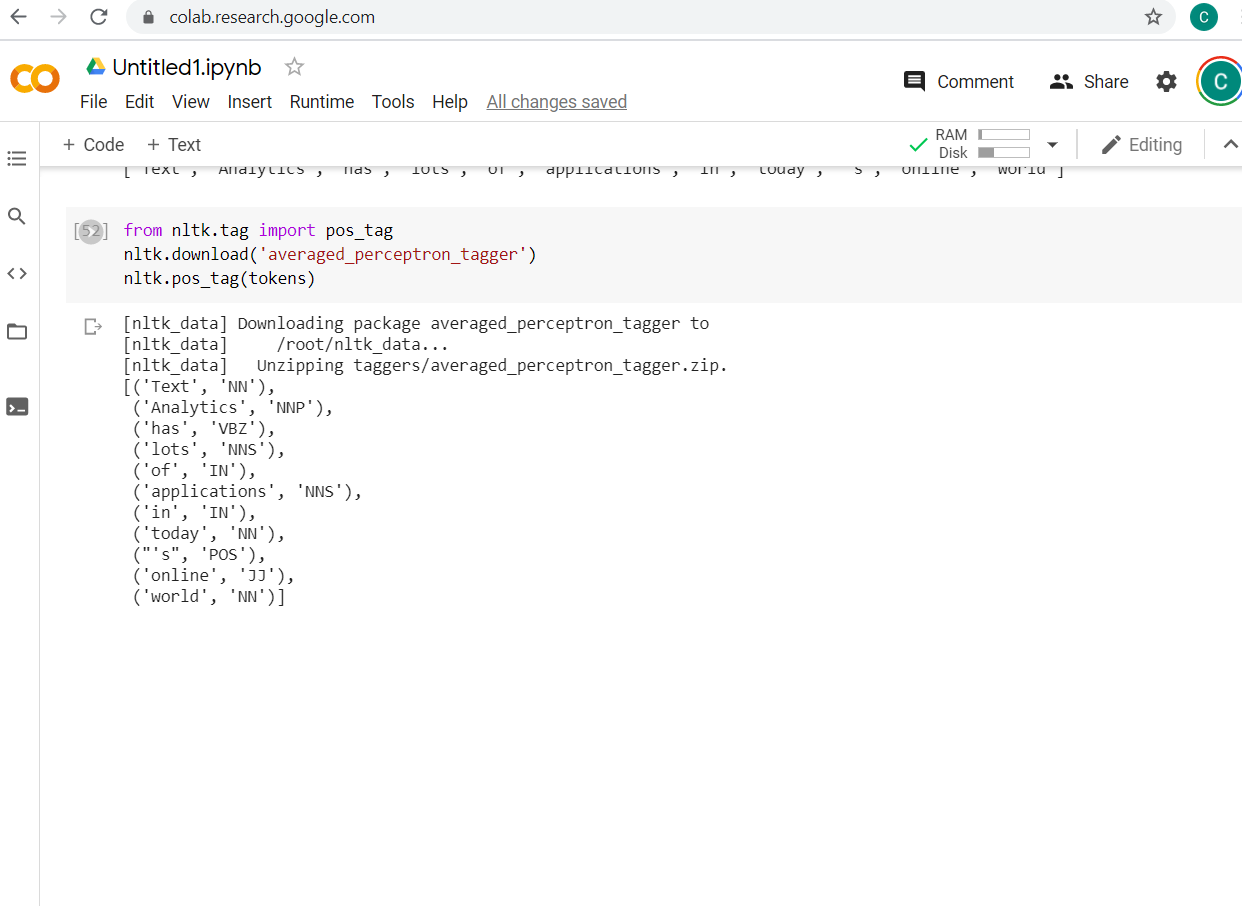
removing them may improve on your analysis. For this icp i identified stop words in my text and filtered out my list of tokens from the stop words identified

**POS Tagging**: i used this to identify the grammatical group of a given word like noun,pronoun, verb etc based on the context









**SCREENSHOT OF GITHUB REPOSITORY CREATED**

