# Report

Here we have a data about `philosophy` , which contains the philosophy sentences said by philosophers. Each row represents one philosophy sentence, the name of philosopher and the school that the philosopher belongs to. Actually, I am not familiar with philosophy, so I am going to using `Python` to do some `EDA` and find something interesting. What's more, I really want to be a philosopher so that I can speak like them. How can we speak like them?

I found there are two ways to speak like them. First is find the most frequency vocabulary they use and the second way is use Markov Chain Model to generate their sentences.

## I . Vocabulary

First of all, I really want to know what kind of words that the philosopher uses most, because I am not a philosopher, but I really want to be a philosopher. So, I need to learn from their words. After doing the text tokenized, we can get all the vocabulary, then we can plot the histogram to see what kinds of words are used most.
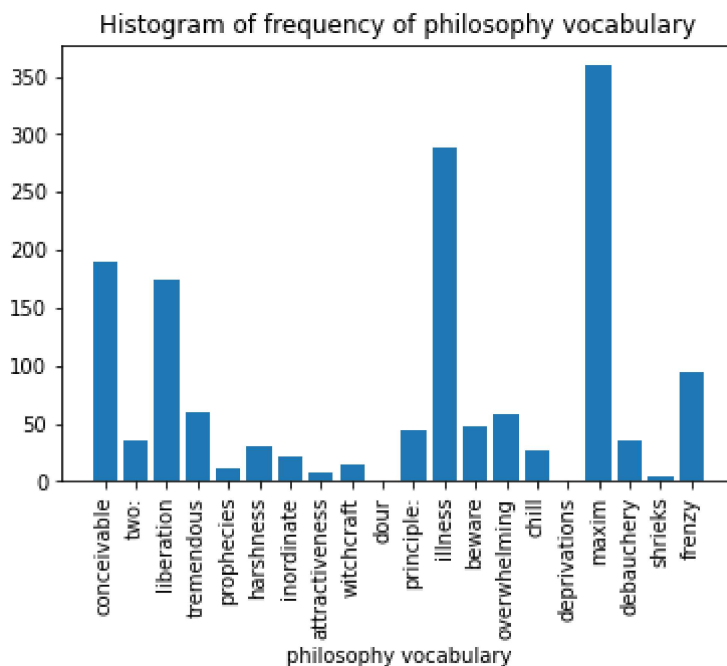


Except the preposition and after plotting many histograms, I found that most philosophers like to say some abstract words, like `principle` , `prosecute` etc. Some words may not be abstract, but can represent the style of that philosopher. Every philosopher may have different styles, such `Plato` . Plato usually uses `king` , `spend` , `leave` etc. most. Also, we can use the same way to see how other philosophers speak.

## Frequency of Plato's vocabulary
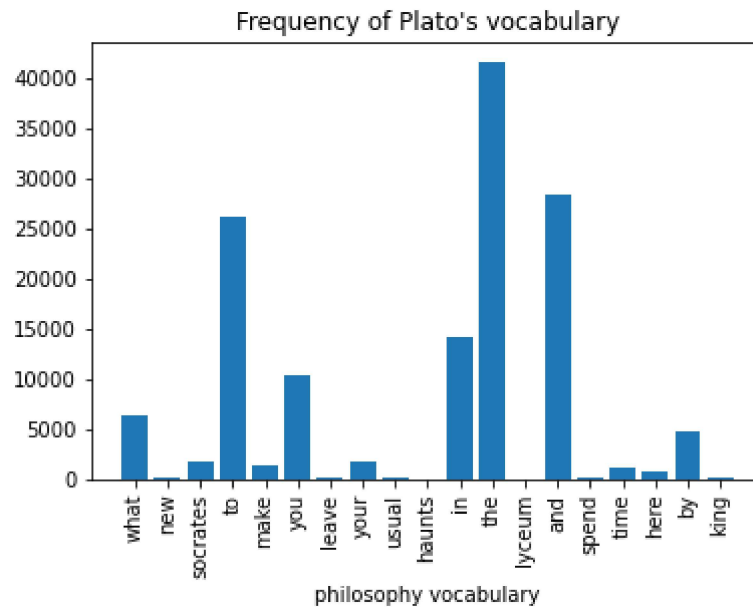


## II. Markov Chain

Let's suppose we have a string, `the` . We need to find the character that is best suited after the character `e` in the word `the` based on our training corpus.

Our text generator would determine that `y` is sometimes after `e` and would form a completed word. In other words, we are going to generate the next character for that given string. So, that we get the new word `they` .

But why don't we choose `n` after `e` in the word `the` , so that we can still get a new vocabulary? The reason is we need to see the frequency of all vocabularies. For example, the frequency of `they` is larger than `then` , then we would define to use `y` after `e` . That's the principle of how **Markov Chain Model** works. It's based on the word that philosopher said.

If I want to simulate how **Plato** speaking, I just need to use the Markov Chain Model we built before. Here we can see the result.

### Original Sentences of Plato

> what's new, socrates, to make you leave your usual haunts in the lyceum and spend your time here by the king archon's court?surely you are not prosecuting anyone before the king archon as i am?the athenians do not call this a prosecution but an indictment, euthyphro.what is this you say?someone must have indicted you, for you are not going to tell me that you have indicted someone else.

### Sentences of Plato after Simulation

> Plato said: what's new, socrates, but, as i was saying just now, for, my friend, you did not teach me adequately when i asked you what the impious, do you say?i say that the law is so.i have already said to others that such things.it is indeed most important, my admirable euthyphro the city by attempting to wrong you.tell me, what does he say you do to corrupt the young plants first, and of the others later.

## III. Conclusion

Finally, I become a 'fake' philosopher. From this data, after doing some `Natural Language Process`, we can get a lot of meaningful things. In this case, I showed how to speak like other philosophers. You can also simulate any philosophers that you liked. Not just philosophy, we can also use these ways to simulate other's speaking, like Trump, Obama etc.

> ***Thank you!***