NLP - Basic - Home Exercise:

1. Perform tokenization on the following text:
   Text: "Tokyo, officially Tokyo Metropolis, is the capital city of Japan."
   For each token you should print its text, POS, tag, entity label and the explanation of each of them.
   Extract all punctuation marks as separate tokens.

2. Extract all noun chunks from the following paragraph.
   Text: "The twelve-year-old cat chased the mouse across the back yard."

3. Visualize named entities and their labels from the following paragraph, using SpaCy's displacy module.
   Text: "Apple Inc. is planning to buy a UK startup for $1 billion."

4. Perform lemmatization on the given paragraph and generate the new text with the lemmatization words.
   Text: "I am running. He runs. We will be running a marathon."

   Write a function that gets the original doc and the corresponding lemmatization doc and mark the words that changed according to lemmatization.

5. Identify and count the stop words in the given text. Next, filter out the stop words and generate a new text without any stop word.
   Text: "It is a truth universally acknowledged, that a single man in possession of a good fortune, must be in want of a wife."

6. Add a new stop word at your choice that is not currently segmented as a stop word. Provide a text doc that contains this new stop word and check that Spacy is recognizing your new customized stop word.

7. Create a Matcher object in SpaCy and define match patterns to find phrases in a text. Apply these patterns to another text and report what was found.

   Texts:

   Text1: "My grandmother has a pet cat."

   Text2: "Many people enjoy having pets. They calling them pat-cat or pat_cat"

   Text3: "Do you think a cat makes a good pet? Pat cats are the best!"

   Patterns to catch: all combinations of the words pat cat, including pat-cats