1. Construct a Dataframe of bag-of-words vectors
   1. Check for the similarities between sentences is to count the number of overlapping tokens using a dot product
   2. Tokenize informal text from social networks such as Twitter and Facebook (2.2.3)
2. Figure out Zipf’s law for other documents such as FOMC?

Section 3

1. Why the sklearn TF vectorizer has different values compared to the code we generated?
2. Typo: len(lexicon) -> len(tokens)