Exercise sheet 5

Natural Language Processing

Hand-in (voluntarily):   11/11/2024 until 10:00 a.m. via Moodle
Please submit a `.py`, `.ipynb`, `.R` or `.rmd` file!

Task 1
In Moodle you will find the file `songs.csv`, containing the lyrics of songs found on Spotify ranging since 1980. Load the data set into your console.
We will now perform an elementary sentiment analysis. We will analyze, how positive or negative the mood in these song texts is and how it changes across time.

Task 2
Apply preprocessing to the given texts. Keep in mind, that we intend to use sentiment dictionaries to analyze the text later. How does this knowledge change your approach to preprocessing?

Task 3
In `dictionary.csv` you will find a sentiment dictionary. "Positive words" will have positive values while "Negative words" will have negative values.
Use this dictionary to calculate the sentiment score of each text, that is sum up all sentiment values to the corresponding words in said text. A negative score will thus indicate a negative text, while a positive value will indicate a positive text.

Task 4
Plot the average sentiment value for each year in the data set in one plot to compare how the sentiment changed over the years. Interpret the resulting graph.

Recommended packages & functions
R: `as.POSIXct()`, `tidyverse::group_by()`, `tidyverse::summarise()`
Python: `pandas.to_datetime()`, `pandas.groupby()`, `datetime.datetime.strptime()`, `matplotlib`