SI 506 Section 8

Objectives

1. Print the text of the 5 tweets with the most likes from the Taylor Swift tweets data set.
    a. Optional challenge: Separately, print the 5 most common hashtags in the data
2. (If time) Write a CSV file with 2 columns, using the data from (1).

Materials (Section Resources > Section 008)

- **sample_tweet.txt** - A file with **data from one tweet**. You can use this file to understand the nested structure of each tweet in *tweets.txt*. We recommend either simply looking at this file or pasting it into http://jsoneditoronline.org/ to understand the structure of the single tweet, and if you are confused, do some programming investigation as well (see Step 1).
- **section8_sample.py** - A Python file you can use to practice extracting data from *sample_tweet.txt* (see Step 2).
- **tweets.txt** - A text file with 300 lines, each line containing data for a tweet. You will process the tweets in this file using *section8.py*.
- **section8.py** - A starter Python file for processing *tweets.txt*. This is the file you will use to accomplish the objectives above.

Step 1: Look at sample_tweet.txt and section8_sample.py

Download all the files from Canvas. Look at the provided files to make sure you understand what each file contains.

Step 2: Practice extracting data from sample_tweet.txt

1. Do some investigation with print statements.
    a. Check out the questions in Monday's lecture notes! (https://paper.dropbox.com/doc/SI-506-W18-Lecture-15-March-5-2018-More-parsing-nested-data-pip-introduction-to-JSON-file-format-and-data-from-the-internet-iPihUNCWTIZDnLyA8nFtj)
2. Add code in the file to parse the tweet and print each tweet's
    a. Tweet text
    `#idontwannaliveforever #fiftyshadesdarker\nhttps://t.co/PwER46Gt3F`
    b. The text of each hashtag
    `Idontwannaliveforever`

```
Fiftyshadesdarker
```

c. The ‘favorite_count’ associated with the tweet (i.e. number of ‘likes’)

```
43569
```

Step 3: Print the 5 most liked tweets in tweets.txt

In section8.py, write code to print the 5 most liked tweets in tweets.txt, i.e., the tweets with the highest ‘favorite_count’ values.

Hint: use the lambda function to sort the tweets by the ‘favorite_count’ field

Sample Output:

Favcount	**Tweet**
370408	That moment when Kanye West secretly records your phone call, then Kim posts it on the Internet. https://t.co/4GJqdyykQu
186814	I'm very happy to say the next single from 1989 will be 'New Romantics'.
180659	You used to call me on my elf phone. https://t.co/nMgUI0Igp6
175199	Today I begin my 26th year of freaking out over stuff. https://t.co/HjWAE7sluQ
173935	Bad Blood just won a Grammy for Best Music Video and we are not ok. @selenagomez https://t.co/EquWAq0UPg

Step 4: Create a .CSV file of the data from step 3

Remember, you need to open a file with a name that has a **.csv** extension for *writing*,
And then need to write strings of data line by line to that file, with commas separating the data you want to be in each cell, and \n (newline) characters separating each line. For more reminders, check out the textbook section on .CSV files that you read in your last reading assignment (for March 5).

Step 5: Print the 5 most common hashtags in tweets.txt

Print the 5 *most used hashtags* in tweets.txt. Note that each tweet can have multiple hashtags.

Hint: recall dictionary accumulation!

Sample Output:

Hashtag	**Count**
GRAMMYs	6
1989TourNashville	6
idontwannaliveforever	6
fiftyshadesdarker	6
1989TourStLouis	5