NLP Assignment

You are given a JSON file (tweets.json) that contains tweets (sentences) along with the name of the author.

Objective 1: Get the most frequent entities from the tweets. Objective 2: Find out the sentiment/polarity of each author towards each of the entities.

Sample Input: Assume we have only 4 tweets:

Tweet1 by Author1: Pink Pearl Apples are tasty but Empire Apples are not. Tweet2 by Author2: Empire Apples are very tasty. Tweet3 by Author3: Pink Pearl Apples are not tasty. Tweet4 by Author1: Pink Pearl Apples smells really good.

Sample output: Entities in the topics extracted: Share a CSV with extracted entities and the frequency of the extracted entity from all the tweets in the following format objective1.csv

entity	**frequency**
Pink Pearl Apples	2
Empire Apples	2

Sentiment/polarity of Authors: Share a CSV file with predicted sentiment values with extracted entities as columns and unique authors as rows. See the example CSV below.

objective2.csv

entity	**author_name**	**overall_polarity**
Pink Pearl Apples	Author1	Positive
Empire Apples	Author1	Negative
Empire Apples	Author2	Positive
Pink Pearl Apples	Author3	Negative

Python code for reading the JSON file:

```
import json

with open('tweets.json') as jfile: d = json.load(jfile)
```

d would be a dictionary with tweet_id as key and another dictionary as a value. The inner dictionary contains the information tweet_text and tweet_author. See the sample below.

{"1374140386071961602": ● {tweet_author:"Hematopoiesis News" ● tweet_text:"□ Scientists conducted a Phase II study of acalabrutinib in patients with relapsed/refractory #CLL who were ibrutinib-intolerant, and found an overall response rate of 73%. https://t.co/eJ6m4QpC5P https://t.co/kuZz6ZO47r"} ... }