```
In [1]: import pandas as pd
        import numpy as np
        import seaborn as sns
        import matplotlib.pyplot as plt
        %matplotlib inline
        import json
        from nltk.corpus import inaugural
        from sklearn.feature_extraction.text import CountVectorizer
        from sklearn.model_selection import train_test_split
        from sklearn.tree import DecisionTreeClassifier
        import re
        import nltk
```

```
In [2]: import os
        os.getcwd()
```

```
Out[2]: 'C:\\Users\\rahul'
```

```
In [3]: os.chdir('C:\\Users\\rahul')
```

```
In [4]: data=pd.read_json("tweets.json")
```

```
In [5]: data1=data.T.copy()
```

Objective 1: Get the most frequent entities from the tweets.

```
In [6]: data1.head()
```

Out[6]:

	tweet_author	tweet_text
2013-07-18 09:39:46.071961602	Hematopoiesis News	💲 Scientists conducted a Phase II study of ac...
2013-07-17 03:40:32.173842437	Michael Wang, MD	This phase 2 Acalabrutinib-Venetoclax (AV) tri...
2013-07-15 15:41:16.553048065	1stOncology	#NICE backs #AstraZenecas #Calquence for #CLL ...
2013-07-12 19:19:42.367813635	Toby Eyre	#acalabrutinib is a valuable option in pts int...
2013-07-04 12:40:34.334232586	Lymphoma Hub	NICE has recommended the use of acalabrutinib ...

In [7]: `data1.info()`

```
<class 'pandas.core.frame.DataFrame'>
DatetimeIndex: 43347 entries, 2013-07-18 09:39:46.071961602 to 1987-06-13 10:
44:06.537678849
Data columns (total 2 columns):
 #   Column        Non-Null Count  Dtype
---  ------        --------------  -----
 0   tweet_author  43347 non-null  object
 1   tweet_text    43347 non-null  object
dtypes: object(2)
memory usage: 2.0+ MB
```

In [8]: `data1.tweet_author.value_counts().head()`

Out[8]:
```
Patient Power          1603
Paperbirds_Hematology  1510
VJHemOnc               1079
Oncology Tube           714
Medivizor               663
Name: tweet_author, dtype: int64
```

In [9]: `data1.groupby("tweet_author").sum().head(10)`

Out[9]:

	tweet_text
tweet_author	
Camilla White	The pleotropic activity of... http://t.co/6u15...
Emilie Thompson	The pleotropic activity of... http://t.co/mr9S...
Hannah Wright	The pleotropic activity of... http://t.co/lLyP...
Yvianna 🔡	Saudade #Cll 😞
#DestroyTheAadhaar TwiLightOFTheGODS	UK hospitals to trial five new drugs in search...
#Endsars protest	Medical News Today: What is the outlook for ch...
#Enritchansrajpandey	☀️ Good Morning All ☀️ \n 🟠 Exclusive Morning Mark...
#KING OF FLORIDA 👑	#360WiseNews : Venetoclax Is A Powerful New Ki...
#LeoWeichafe	Esperando a que comience la lucha, acá en @chi...
#Memory	Antitumor Potency of an Anti-CD19 Chimeric Ant...

In [10]: `df=data1.groupby("tweet_text").sum().value_counts()`

In [11]: `df.head(22)`

Out[11]:
```
tweet_author
Patient Power                  1573
Paperbirds_Hematology          1510
VJHemOnc                       1075
Oncology Tube                   714
Medivizor                       663
Club Libertad Limpeño           505
Lymphoma Hub                    489
Raul Cordoba, MD, PhD           488
CLL Ireland                     482
OncLive.com                     475
Targeted Oncology               450
Kathleen D. Hoffman             373
Cancer Papers                   372
Patient Empowerment Network     367
chadi nabhan MD, MBA, FACP      319
CLLSA                           282
Andrew Schorr                   271
CLL Society Inc.                267
Lymphoma Papers                 236
Anna Deveau                     210
CancerWallonia                  206
Toby Eyre                       194
dtype: int64
```

objective1.csv

In [12]: `df.to_csv (r'C:\Users\rahul\OneDrive\Desktop\Akaike Technologies Assignment_Ra`

In [13]: `data1.groupby("tweet_author").sum().value_counts()`

Out[13]:
```
tweet_text
#CLL
4
chronic lymphocytic leukemia
2
ASTRAZENECA: LYMPHOMA DRUG ACALABRUTINIB A POTENTIAL THERAPY FOR SEVERE COVID
-19ASTRAZENECA'S CANCER DRUG CALQUENCE SAID TO SHOW INITIAL SIGNS OF HELPING
HOSPITALISED COVID-19 PATIENTS - RTRS
2
Among older patients with untreated chronic lymphocytic leukemia, treatment w
ith ibrutinib, either alone or in combination with rituximab, was superior to
treatment with bendamustine + rituximab with regard to progression-free survi
val. #ASH18      2
#cll
2

..
Congrats to the @WEHI_research @TheRMH @PeterMacCC team for the recognition a
t #VPHAwards. New paradigm for treating #cancer #CLL https://t.co/PEK0rxauQg
(https://t.co/PEK0rxauQg)
1
Congratulation to my colleagues from the #GCLLSG on this very important trial
in a #CLL patient population in need for innovative therapies #ASCO19 http
s://t.co/YKKSbS1NsM (https://t.co/YKKSbS1NsM)
1
Congratulations @FerranNadeu!!! @idibaps @hospitalclinic #CLL https://t.co/E1
Z9QHh9jy (https://t.co/E1Z9QHh9jy)
1
Congratulations @genentech for being among companies transforming chronic lym
phocytic #leukemia treatment. #ICYMI : http://t.co/gzmdDuiYNb (http://t.co/gz
mdDuiYNb)
1
🔲 urbano encuéntranos en la #cll-37-8-51 barrio gaitan
1
Length: 9282, dtype: int64
```

Objective 2: Find out the sentiment/polarity of each author towards each of the entities.

In [14]: `from vaderSentiment.vaderSentiment import SentimentIntensityAnalyzer`

```
In [15]: sentiment = SentimentIntensityAnalyzer()

         text_1 = "Pink Pearl Apples are tasty but Empire Apples are not."
         text_2 = "Empire Apples are very tasty."
         text_3 = "Pink Pearl Apples are not tasty."
         text_4 = "Pink Pearl Apples smells really good."
         sent_1 = sentiment.polarity_scores(text_1)
         sent_2 = sentiment.polarity_scores(text_2)
         sent_3 = sentiment.polarity_scores(text_3)
         sent_4 = sentiment.polarity_scores(text_4)
         print("Sentiment of text 1:", sent_1)
         print("Sentiment of text 2:", sent_2)
         print("Sentiment of text 3:", sent_3)
         print("Sentiment of text 4:", sent_4)
```

```
Sentiment of text 1: {'neg': 0.0, 'neu': 1.0, 'pos': 0.0, 'compound': 0.0}
Sentiment of text 2: {'neg': 0.0, 'neu': 1.0, 'pos': 0.0, 'compound': 0.0}
Sentiment of text 3: {'neg': 0.0, 'neu': 1.0, 'pos': 0.0, 'compound': 0.0}
Sentiment of text 4: {'neg': 0.0, 'neu': 0.61, 'pos': 0.39, 'compound': 0.492
7}
```

```
In [16]: nltk.download('stopwords')
         stemmer = nltk.SnowballStemmer("english")
         from nltk.corpus import stopwords
         import string
         stopword=set(stopwords.words('english'))

         def clean(text):
             text = str(text).lower()
             text = re.sub('\[.*?\]', '', text)
             text = re.sub('https?://\S+|www\.\S+', '', text)
             text = re.sub('<.*?>+', '', text)
             text = re.sub('[%s]' % re.escape(string.punctuation), '', text)
             text = re.sub('\n', '', text)
             text = re.sub('\w*\d\w*', '', text)
             text = [word for word in text.split(' ') if word not in stopword]
             text=" ".join(text)
             text = [stemmer.stem(word) for word in text.split(' ')]
             text=" ".join(text)
             return text
         data1['tweet_text'] = data1['tweet_text'].apply(clean)
```

```
[nltk_data] Downloading package stopwords to
[nltk_data]     C:\Users\rahul\AppData\Roaming\nltk_data...
[nltk_data]   Package stopwords is already up-to-date!
```

```
In [17]: from nltk.sentiment.vader import SentimentIntensityAnalyzer
         nltk.download('vader_lexicon')
         sentiments = SentimentIntensityAnalyzer()
         data1["Positive"] = [sentiments.polarity_scores(i)["pos"] for i in data1['twee
         data1["Negative"] = [sentiments.polarity_scores(i)["neg"] for i in data1['twee
```

```
         [nltk_data] Downloading package vader_lexicon to
         [nltk_data]     C:\Users\rahul\AppData\Roaming\nltk_data...
         [nltk_data]   Package vader_lexicon is already up-to-date!
```

```
In [18]: data1 = data1[['tweet_text', "Positive",
                        "Negative"]]
         print(data1.head())
```

```
                                                                  tweet_t
         ext  \
         2013-07-18 09:39:46.071961602   ⚕ scientist conduct phase ii studi acalabrut
         i...
         2013-07-17 03:40:32.173842437  phase  acalabrutinibvenetoclax av trial still
         ...
         2013-07-15 15:41:16.553048065                   nice back astrazeneca calquenc c
         ll
         2013-07-12 19:19:42.367813635  acalabrutinib valuabl option pts intoler ibru
         t...
         2013-07-04 12:40:34.334232586  nice recommend use acalabrutinib patient trea
         t...

                                        Positive  Negative
         2013-07-18 09:39:46.071961602     0.000       0.0
         2013-07-17 03:40:32.173842437     0.104       0.0
         2013-07-15 15:41:16.553048065     0.412       0.0
         2013-07-12 19:19:42.367813635     0.162       0.0
         2013-07-04 12:40:34.334232586     0.346       0.0
```

objective2.csv

```
In [19]: data3=data1.copy()
```

```
In [20]: data3.head()
```

Out[20]:

	tweet_text	Positive	Negative
2013-07-18 09:39:46.071961602	⚕ scientist conduct phase ii studi acalabruti...	0.000	0.0
2013-07-17 03:40:32.173842437	phase acalabrutinibvenetoclax av trial still ...	0.104	0.0
2013-07-15 15:41:16.553048065	nice back astrazeneca calquenc cll	0.412	0.0
2013-07-12 19:19:42.367813635	acalabrutinib valuabl option pts intoler ibrut...	0.162	0.0
2013-07-04 12:40:34.334232586	nice recommend use acalabrutinib patient treat...	0.346	0.0

```
In [21]: data3.to_csv (r'C:\Users\rahul\OneDrive\Desktop\Akaike Technologies Assignment
```

```
In [22]: x = sum(data1["Positive"])
         y = sum(data1["Negative"])


         def sentiment_score(a, b):
             if (a>b) :
                 print("Positive 😊 ")
             elif (b>a) :
                 print("Negative 😠 ")

         sentiment_score(x, y)
```

Positive 😊

```
In [23]: print("Positive: ", x)
         print("Negative: ", y)
```

```
Positive:  3036.0440000000335
Negative:  1902.5980000000004
```

Thanks

```
In [ ]:
```