1. Write Python code that outputs no. of positve and negative tweets (*2 point*).

```
import nltk
from nltk.corpus import twitter_samples

nltk.download('twitter_samples', quiet=True)
nltk.download('stopwords', quiet=True)



all_positive_tweets = twitter_samples.strings('positive_tweets.json')
all_negative_tweets = twitter_samples.strings('negative_tweets.json')

Num_positive_tweets:len(all_positive_tweets)
Num_negative_tweets:len(all_negative_tweets)


print("Number of Positive Tweets:", positive_count)
print("Number of Negative Tweets:", negative_count)

print("The type of all_positive_tweets is:", type(all_positive_tweets))
print("The type of a tweet entry is:", type(all_negative_tweets[0]))
```

cayabyab-charls-2020-8-0463

2. Write a Python code that outputs the following (*3 points*):.

```
import nltk
from nltk.corpus import twitter_samples

nltk.download('twitter_samples', quiet=True)
nltk.download('stopwords', quiet=True)

#emtpyset

data_set= []

all_positve_tweets = twitter_samples.strings('positive_tweets.json')
all_negative_tweets = twitter_samples.strings('negative_tweets.json')

#append

data_set.extend(all_positve_tweets)
data_set.extend(all_negative_tweets)

positive_count = len(all_positve_tweets)
negative_count = len(all_negative_tweets)

#create piechart
labels = 'Positive', 'Negative'
sizes = [positive_count, negative_count]
plt.pie(sizes, labels=labels, autopct='%1.1f%%')
shadow = True
plt.show()
```

cayabyab-charls-2020-8-0463

3. Write a python code that will write random tweets, colormark positive tweets in GREEN,

and negative tweets in RED. I.e. below (*2 point*):

```
import re
import random
import nltk
from nltk.corpus import twittter_samples
from termcolor import colored

nltk.download('twitter_samples',quiet=True)
nltk.download('stopwords',quiet=True)

all_positve_tweets = twitter_samples.strings('positive_tweets.json')
all_negative_tweets = twitter_samples.strings('negative_tweets.json')

rand1 = random.randint(0,len(all_positve_tweets))
len(all_positve_tweets)
ran2 = random.randint(0,len(all_negative_tweets))
len(all_negative_tweets)

positive_tweets = all_positve_tweets[rand1]
negative_tweets = all_negative_tweets[ran2]

print(colored(positive_tweets,'green'))
print(colored(negative_tweets,'red'))
```

cayabyab-charls-2020-8-0463

remove old style retweet text "RT"

4.1 Write codes here to show the same output below (*2 points*)

```
import re
import random
import nltk
from nltk.corpus import twittter_samples
from nltk.corpus import stopwords
from nltk.stem import PorterStemmer
from nltk.tokenize import TweetTokenizer
from nltjk.stem import twitter_samples

nltk.download('twitter_samples',quiet=True)
nltk.download('stopwords',quiet=True)


#4.1 Remove old-style retweet text "RT"
tweet = re.sub(r'^RT[\s]+', '', tweet)

Tokenizer
TweetTokenizer(preserve_case=False, strip_handles=True, reduce_len=True)
tweet_tokens = tokenizer.tokenize(tweet)

tweet_clean = []
 for word in tweet_tokens:
     if (word not in stopwords_english and word not in string.punctuation):
         stem_word = stemmer.stem(word):

tweets_clean.append(stem_word)
return tweets_clean

all_positve_tweets = twitter_samples.strings('positive_tweets.json')
all_negative_tweets = twitter_samples.strings('negative_tweets.json')

for word in all_positive_tweets:
    stem_positive = process_tweet(word)
       stemmer.stem(word) ''.join(stem_positive)
       print(stem_tweet_positive)
```

```
 for word in all_negative_tweets:
     stem_negative = process_tweet(word)
        stemmer.stem(word) ''.join(stem_positive)
        print(stem_tweet_negative)sss
```

cayabyab-charls-2020-8-0463

remove hyperlinks

4.2 Write codes here to show the same output below *(2 points*)

```
import re
import random
import nltk
from nltk.corpus import twittter_samples
from nltk.corpus import stopwords
from nltk.stem import PorterStemmer
from nltk.tokenize import TweetTokenizer
from nltjk.stem import twitter_samples

nltk.download('twitter_samples',quiet=True)
nltk.download('stopwords',quiet=True)


#4.1 Remove old-style retweet text "RT"
tweet = re.sub(r'^RT[\s]+', '', tweet)

# 4.2 Remove hyperlinks
tweet = re.sub(r'https?:\/\/\S+', '', tweet)
Tokenizer
TweetTokenizer(preserve_case=False, strip_handles=True, reduce_len=True)
tweet_tokens = tokenizer.tokenize(tweet)

tweet_clean = []
 for word in tweet_tokens:
^    if (word not in stopwords_english and word not in string.punctuation):
        stem_word = stemmer.stem(word):

tweets_clean.append(stem_word)
return tweets_clean

all_positve_tweets = twitter_samples.strings('positive_tweets.json')
all_negative_tweets = twitter_samples.strings('negative_tweets.json')

for word in all_positive_tweets:
    stem_positive = process_tweet(word)
      stemmer.stem(word) ''.join(stem_positive)
      print(stem_tweet_positive)
```

```
for word in all_negative_tweets:
    stem_negative = process_tweet(word)
      stemmer.stem(word) ''.join(stem_positive)
      print(stem_tweet_negative)
```

cayabyab-charls-2020-8-0463

remove hashtags

4.3 Write codes here to show the same output below (*2 points*)

```
import re
import random
import nltk
from nltk.corpus import twittter_samples
from nltk.corpus import stopwords
from nltk.stem import PorterStemmer
from nltk.tokenize import TweetTokenizer
from nltjk.stem import twitter_samples

nltk.download('twitter_samples',quiet=True)
nltk.download('stopwords',quiet=True)


#4.1 Remove old-style retweet text "RT"
tweet = re.sub(r'^RT[\s]+', '', tweet)

# 4.2 Remove hyperlinks
tweet = re.sub(r'https?:\/\/\S+', '', tweet)
# 4.3 Remove hashtags (remove only the '#' character)
tweet = re.sub(r'#', '', tweet)
Tokenizer
TweetTokenizer(preserve_case=False, strip_handles=True, reduce_len=True)
tweet_tokens = tokenizer.tokenize(tweet)

tweet_clean = []
 for word in tweet_tokens:
      if (word not in stopwords_english and word not in string.punctuation):
          stem_word = stemmer.stem(word):

tweets_clean.append(stem_word)
return tweets_clean

all_positve_tweets = twitter_samples.strings('positive_tweets.json')
all_negative_tweets = twitter_samples.strings('negative_tweets.json')
```

```
for word in all_positive_tweets:
    stem_positive = process_tweet(word)
       stemmer.stem(word) ''.join(stem_positive)
       print(stem_tweet_positive)

for word in all_negative_tweets:
    stem_negative = process_tweet(word)
       stemmer.stem(word) ''.join(stem_positive)
       print(stem_tweet_negative)
```

cayabyab-charls-2020-8-0463

5.1 utility.py (*1 point*)

```
import re
import random
import nltk
from nltk.corpus import twittter_samples
from nltk.corpus import stopwords
from nltk.stem import PorterStemmer
from nltk.tokenize import TweetTokenizer
from nltjk.stem import twitter_samples

nltk.download('twitter_samples',quiet=True)
nltk.download('stopwords',quiet=True)

all_positve_tweets = twitter_samples.strings('positive_tweets.json')
all_negative_tweets = twitter_samples.strings('negative_tweets.json')

tweet = all_negative_tweets[2277]

print()
print('\033[92m' + tweet)
print('\033[94m')

#imported function
process_tweet =tweet(tweet); preprocess a given tweet

print('preprocessed tweet:')
print(tweet_stem)

```

cayabyab-charls-2020-8-0463

5.2 utility.py (*1 point*)

```
from utility import process_tweet


import nltk
from nltk.corpus import twitter_samples


nltk.download('twitter_samples',quiet=True)
nltk.download('stopwords',quiet=True)

all_positve_tweets = twitter_samples.strings('positive_tweets.json')
all_negative_tweets = twitter_samples.strings('negative_tweets.json')

for word in all_positive_tweets:
    stem_positive = process_tweet(word)
        stemmer.stem(word) ''.join(stem_positive)
        print(stem_tweet_positive)

for word in all_negative_tweets:
    stem_negative = process_tweet(word)
        stemmer.stem(word) ''.join(stem_positive)
        print(stem_tweet_negative)

```

cayabyab-charls-2020 -8-0463