

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

```
1  import nltk
2  from nltk.corpus import twitter_samples
3
4  nltk.download('twitter_samples', quiet=True)
5  nltk.download('stopwords', quiet=True)
6
7
8
9  all_positive_tweets = twitter_samples.strings('positive_tweets.json')
10 all_negative_tweets = twitter_samples.strings('negative_tweets.json')
11
12 Num_positive_tweets=len(all_positive_tweets)
13 Num_negative_tweets=len(all_negative_tweets)
14
15
16 print("Number of Positive Tweets:", positive_count)
17 print("Number of Negative Tweets:", negative_count)
18
19 print("The type of all_positive_tweets is:", type(all_positive_tweets))
20 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*):.

```
1  import nltk
2  from nltk.corpus import twitter_samples
3
4  nltk.download('twitter_samples', quiet=True)
5  nltk.download('stopwords', quiet=True)
6
7  #emptyset
8
9  data_set= []
10
11 all_positive_tweets = twitter_samples.strings('positive_tweets.json')
12 all_negative_tweets = twitter_samples.strings('negative_tweets.json')
13
14 #append
15
16 data_set.extend(all_positive_tweets)
17 data_set.extend(all_negative_tweets)
18
19 positive_count = len(all_positive_tweets)
20 negative_count = len(all_negative_tweets)
21
22 #create piechart
23 labels = 'Positive', 'Negative'
24 sizes = [positive_count, negative_count]
25 plt.pie(sizes, labels=labels, autopct='%1.1f%%')
26 shadow = True
27 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*):

```
1  import re
2  import random
3  import nltk
4  from nltk.corpus import twitter_samples
5  from termcolor import colored
6
7  nltk.download('twitter_samples',quiet=True)
8  nltk.download('stopwords',quiet=True)
9
10 all_positive_tweets = twitter_samples.strings('positive_tweets.json')
11 all_negative_tweets = twitter_samples.strings('negative_tweets.json')
12
13 rand1 = random.randint(0,len(all_positive_tweets))
14 len(all_positive_tweets)
15 ran2 = random.randint(0,len(all_negative_tweets))
16 len(all_negative_tweets)
17
18 positive_tweets = all_positive_tweets[rand1]
19 negative_tweets = all_negative_tweets[ran2]
20
21 print(colored(positive_tweets,'green'))
22 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*)

```

1  import re
2  import random
3  import nltk
4  from nltk.corpus import twitter_samples
5  from nltk.corpus import stopwords
6  from nltk.stem import PorterStemmer
7  from nltk.tokenize import TweetTokenizer
8  from nltk.stem import twitter_samples
9
10 nltk.download('twitter_samples',quiet=True)
11 nltk.download('stopwords',quiet=True)
12
13
14 #4.1 Remove old-style retweet text "RT"
15 tweet = re.sub(r'^RT[\s]+', '', tweet)
16
17 Tokenizer
18 TweetTokenizer(preserve_case=False, strip_handles=True, reduce_len=True)
19 tweet_tokens = tokenizer.tokenize(tweet)
20
21 tweet_clean = []
22 for word in tweet_tokens:
23     if (word not in stopwords_english and word not in string.punctuation):
24         stem_word = stemmer.stem(word):
25
26 tweets_clean.append(stem_word)
27 return tweets_clean
28
29 all_positive_tweets = twitter_samples.strings('positive_tweets.json')
30 all_negative_tweets = twitter_samples.strings('negative_tweets.json')
31
32 for word in all_positive_tweets:
33     stem_positive = process_tweet(word)
34     stemmer.stem(word) ''.join(stem_positive)
35     print(stem_tweet_positive)
36
37 for word in all_negative_tweets:
38     stem_negative = process_tweet(word)
39     stemmer.stem(word) ''.join(stem_positive)
40     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 twitter_samples
from nltk.corpus import stopwords
from nltk.stem import PorterStemmer
from nltk.tokenize import TweetTokenizer
from nltk.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_positive_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*)

```

1  import re
2  import random
3  import nltk
4  from nltk.corpus import twitter_samples
5  from nltk.corpus import stopwords
6  from nltk.stem import PorterStemmer
7  from nltk.tokenize import TweetTokenizer
8  from nltk.stem import twitter_samples
9
10 nltk.download('twitter_samples',quiet=True)
11 nltk.download('stopwords',quiet=True)
12
13
14 #4.1 Remove old-style retweet text "RT"
15 tweet = re.sub(r'^RT[\s]+', '', tweet)
16
17 # 4.2 Remove hyperlinks
18 tweet = re.sub(r'https?:\/\/\S+', '', tweet)
19 # 4.3 Remove hashtags (remove only the '#' character)
20 tweet = re.sub(r'#', '', tweet)
21 Tokenizer
22 TweetTokenizer(preserve_case=False, strip_handles=True, reduce_len=True)
23 tweet_tokens = tokenizer.tokenize(tweet)
24
25 tweet_clean = []
26 for word in tweet_tokens:
27     if (word not in stopwords_english and word not in string.punctuation):
28         stem_word = stemmer.stem(word):
29
30 tweets_clean.append(stem_word)
31 return tweets_clean
32
33 all_positive_tweets = twitter_samples.strings('positive_tweets.json')
34 all_negative_tweets = twitter_samples.strings('negative_tweets.json')
35

```

```

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)

```

5.1 utility.py (*1 point*)

```
1  import re
2  import random
3  import nltk
4  from nltk.corpus import twitter_samples
5  from nltk.corpus import stopwords
6  from nltk.stem import PorterStemmer
7  from nltk.tokenize import TweetTokenizer
8  from nltk.stem import twitter_samples
9
10 nltk.download('twitter_samples',quiet=True)
11 nltk.download('stopwords',quiet=True)
12
13 all_positive_tweets = twitter_samples.strings('positive_tweets.json')
14 all_negative_tweets = twitter_samples.strings('negative_tweets.json')
15
16 tweet = all_negative_tweets[2277]
17
18 print()
19 print('\033[92m' + tweet)
20 print('\033[94m')
21
22 #imported function
23 process_tweet =tweet(tweet); preprocess a given tweet
24
25 print('preprocessed tweet:')
26 print(tweet_stem)
27
```

cayabyab-charls-2020-8-0463

5.2 utility.py (*1 point*)

```
1  from utility import process_tweet
2
3
4  import nltk
5  from nltk.corpus import twitter_samples
6
7
8  nltk.download('twitter_samples',quiet=True)
9  nltk.download('stopwords',quiet=True)
10
11 all_positive_tweets = twitter_samples.strings('positive_tweets.json')
12 all_negative_tweets = twitter_samples.strings('negative_tweets.json')
13
14 √ for word in all_positive_tweets:
15 √     stem_positive = process_tweet(word)
16     stemmer.stem(word) ''.join(stem_positive)
17     print(stem_tweet_positive)
18
19 √ for word in all_negative_tweets:
20 √     stem_negative = process_tweet(word)
21     stemmer.stem(word) ''.join(stem_positive)
22     print(stem_tweet_negative)
```

cayabyab-charls-2020 -8-0463