

Counting and plotting word frequency from a file

```
In [1]: from collections import Counter # importing Counter from collections
```

```
In [2]: # Remove Punctuations funcation
def remove_punctuations(char):
    punctuations = '!'() - [{}];: '"\,<>./?@#$$%^&*~''
    l = []
    for i in char:
        if i not in punctuations:
            l.append(i)
    return ''.join(l)
```

```
In [3]: file = open('Sense_and_Sensibility.txt') #open sense and sensibility file
stopwords = [] #stopword
for lines in open('stopwords.txt'): #opening stopword file
    stopwords.append(lines.strip()) # removing spaces from left and right
data_set = file.read() #reading the Sense and Sensibility file
split_it = data_set.split() #splitting the Sense and sensibility file based on spaces
```

```
In [4]: for i in range(0,len(split_it)): #Looping thought each word
        split_it[i] = remove_punctuations(split_it[i]) # removing punctuations
new_list = []
for i in split_it:
    if i.lower() not in stopwords: #checking in stopwords
        new_list.append(i)
split_it = new_list
Counter = Counter(split_it)
most_occur = Counter.most_common(30) #getting the most common 30 words
histogram = {}
for i in most_occur:
    word, count = i
    histogram[word] = count
```

```
In [5]: i = 0
for keys, value in histogram.items(): #displaying the histogram
    i += 1
    text1 = f'{i}: {keys}({value}times)'.ljust(35)
    star = '*'*int(value/10)
    text2 = f'{star}'.ljust(50)
    print(f'{text1}{text2}')
```

```
1: time(235times) *****
2: Dashwood(224times) *****
3: sister(213times) *****
4: Miss(207times) *****
5: mother(200times) *****
6: Colonel(162times) *****
7: house(146times) *****
```

8: John(146times)	*****
9: make(142times)	*****
10: day(138times)	*****
11: made(137times)	*****
12: good(136times)	*****
13: give(129times)	*****
14: heart(123times)	*****
15: thought(116times)	*****
16: Brandon(116times)	*****
17: Ferrars(108times)	*****
18: dear(106times)	*****
19: felt(104times)	*****
20: young(103times)	*****
21: long(99times)	*****
22: sisters(99times)	*****
23: replied(99times)	*****
24: Lady(99times)	*****
25: left(98times)	*****
26: room(97times)	*****
27: mind(95times)	*****
28: kind(93times)	*****
29: happy(93times)	*****
30: world(89times)	*****

In []: