Assignment 3

In this module, you will create 3 Python notebooks. In all your programs, use markdown cells and be creative of summarizing/commenting your notebooks. You will also add detailed comments in your Python code (using "#" or triple quote signs)

Notebook 1: Prime Numbers

a. Write a Jupyter Notebook to find the 9991th to 10000th prime numbers.
b. Display the numbers to the notebook
c. Output the numbers to a data file called **prime.txt**.
d. Demonstrate that you used "continue" and "break" to improve efficiency and **explain why**

Notebook 2: ROT

ROT is a very simple cipher that is used to information hiding (https://en.wikipedia.org/wiki/ROT13) . Read the wiki page to understand how the encoder works.

a. Create an "encode_rot()" function to encode any given strings using ROT algorithm. The input should contain a key and a string of text. The key can be any integers both negative and positive (-12: turn left 12 positions, 36: turn right 36 positions). Only alphabet letters are encoded.
The following two lines of your code will generate an output of "*Ocejkpg ECP ngctp 2 !!!"*.
clear_text=" Machine CAN learn 2 !!!"
encode_rot(clear_text, 28)

b. Create a decode_rot() to decode a ciphertext. The input only contains the ciphertext. The output contains the cleartext and the key that was used to encode text. The key will be between 0 and 25. (hint: Compare your decoded clear text with a dictionary text file and decide which one has the most dictionary words.)
The following two lines will generate an output of

> *The clear text is : "Data is like people, interrogate it hard enough and it will tell you whatever you want to hear."*
> *The key is 16*

cipher_text= "Tqjq yi byau fuefbu, ydjuhhewqju yj xqht udekwx qdt yj mybb jubb oek mxqjuluh oek mqdj je xuqh."

decode_rot(cipher_text)

c. Use your function to test more ciphertexts and show the results in your notebook . (you can use www.rot13.com or your encode_rot() function generate cipher texts).

Notebook 3. Histogram of Top Words

Find the top frequently used words in the book of "Sense and Sensibility". The book is in the sense_andsensibility.txt file.

a. The words should not be case sensitive, meaning "Mother" and "mother" are considered the same word.
b. **Replace all the punctuation marks with a space**.
c. Use the "stopwords.txt" file to remove all the stop words in text. (**Do NOT modify the stopwords.txt file**)
d. Create a histogram similar to the "histogram.jpg" file. The diagram should contain the ranking, the top 30 words, the number of times they appeared in the book. The number of stars will be the number of appearance divided by 10. For example, "mother" appears 263 times; there are 26 stars displayed. (You may not have the exactly the same result as in the histogram.jpg)

```
1: heir(502times)        ****************************************************
2: sin(339times)         ********************************
3: sister(331times)      ********************************
4: ash(316times)         *******************************
5: time(308times)        ******************************
6: lie(304times)         ******************************
7: wood(298times)        *****************************
8: owe(296times)         *****************************
9: part(293times)        *****************************
10: dashwood(288times)   ****************************
11: pen(282times)        ****************************
12: led(268times)        **************************
13: mother(263times)     **************************
14: ring(255times)       *************************
15: miss(252times)       *************************
16: gain(242times)       ************************
17: ice(242times)        ************************
18: act(237times)        ***********************
19: day(237times)        ***********************
20: give(213times)       *********************
21: side(198times)       *******************
22: fee(197times)        *******************
23: long(191times)       *******************
24: rest(189times)       ******************
25: good(188times)       ******************
26: till(183times)       ******************
27: sat(179times)        *****************
28: house(179times)      *****************
29: low(178times)        *****************
30: pin(177times)        *****************
```

Submissions:

You will export your notebooks in both .html and .py formats. You will submit the following 6 files to Blackboard.

1. Firstname_Lastname_Notebook1.html
2. Firstname_Lastname_Notebook1.py
3. Firstname_Lastname_Notebook2.html
4. Firstname_Lastname_Notebook2.py
5. Firstname_Lastname_Notebook3.html
6. Firstname_Lastname_Notebook3.py

*The attached files include

Dictionary.txt :	the dictionary file for notebook 2
Sense_and_sensibility.txt :	the book for notebook 3
Stopwords.txt:	the stop words for notebook3
Histogram.jpg:	sample result of notebook3