Introduction to Data Science using Python(CSE 3054)

MINOR ASSIGNMENT-1

1. An anonymous dataset containing each user's salary (in dollars) and tenure as a data scientist (in years) is given.

```
salaries_and_tenures = [(83000, 8.7), (88000, 8.1), (48000, 0.7),
(76000, 6), (69000, 6.5), (76000, 7.5), (60000, 2.5), (83000, 10),
(48000, 1.9), (63000, 4.2)]
```

   Find out the average salary for each tenure and print a massage according to its value, i.e. "less than two", "between two and five" and "more than five" tenure and group together the salaries corresponding to each bucket. Compute the average salary for each group.

2. For the above data there seems to be a correspondence between years of experience and paid accounts Users with very few and very many years of experience tend to pay; users with average amounts of experience don't. Find out the condition for this correspondence and print it.

3. Write a Python Script to generate random passwords (alphanumeric). Ask users to enter the length of password and number of passwords they want to generate and then save all the generated passwords as a textfile named "MyPasswords.txt".

4. Given a file named "MyText.txt" containing several lines/paragraph, find all unique characters (ignore space, comma, full stop, brackets, and quotes etc.) present in the file. Capital and small letter are counted as same.
   Find the frequency (fi) of all characters in the file and print the output as follows.
   The character "a" is present _____ times in the document.
   The character "t" is present _____ times in the document.

5. Use the above program as a function and use it to write another function to compare contents of two files "MyText1.txt" and "MyText2.txt".

   a. The output must also give the following information.
      File MyText1 contain more (or less or equal) characters than MyText2.
   b. The output must be printed in the following format depending on content of the file.
      File MyText1 contain more (or less or equal)unique characters than MyText2.
   c. The frequency of each characters must be summarised.
      The frequency of character of character "x" in file MyText1 ismore (or less or equal)to characters than MyText2.
   d. The relative frequency of each characters also must be summarised.
      The relative frequency of character of character "x" in file MyText1 ismore (or less or equal)to characters than MyText2.

   The input files should be nonempty.

6. Read a lists named **StringList1** containing strings from the key board. Generate a string **MStringList1** that contains all items of StringList1 that are repeated twice or more number of times and print this list. By observing the outcome of MStringList1 perform the following tasks:

a. Check wather an item of MStringList1 occurs even number of times or odd number of times in StringList1.

b. Remove the i^{th} $(i \geq 2)$ occurrence of a given word in a StringList1.

7. From the file "MyText.txt" count frequencies of various alphabets (Convert upper case into lower case), plot the results for this as a bar chart with x-axis being the letter and y-axis as the corresponding frequency.

8. Use the following data to plot the number of applicant per year as a scatter plot.

```
year = [2020, 2019, 2018, 2017, 2016, 2015, 2014, 2013, 2012]
no_application_per_year = [921261, 929198, 1043739, 1186454,
                    1194938, 1304495, 1356805, 1282000, 479651]
```

9. Plot $xsinx$, x^2sinx , x^3sinx and x^4sinx in a single plot in the range $x \in [-10, 10]$.

10. Plot histogram for age of male and female in different plots for the following data of male and female age.

```
male_age = [53,51,71,31,33,39,52,27,54,30,64,26,21,54,52,20,59,32]
female_age = [53,65,68,21,75,46,24,63,61,24,49,41,39,40,25,54,42,
            32,48,23,23]
```

11. Plot the temperature extremes in certain region of India for each month, starting in January, which are given by (in degrees Celsius).

```
max:  17, 19, 21, 28, 33, 38, 37, 37, 31, 23, 19, 18
min:  -62, -59, -56, -46, -32, -18, -9, -13, -25, -46, -52, -58
```

12. Python Program to find all Numbers in a Range (given by user) which are Perfect Squares and Sum of all Digits in the Number is Less than 10.

13. Plot a bar chart with axis labels for given data:

```
mentions = [500, 505]
years = [2017, 2018]
```

Do not give any extra condition for x-axis as well as y-axis. Now again plot the bar chart for this data and start y-axis from 0.
State the difference in both the bar chart.

14. Plot the scatter plot for following data with unequal axis and then equal axis. Also state the difference in two.

```
test_1_grades = [ 99, 90, 85, 97, 80]
test_2_grades = [100, 85, 60, 90, 70]
```