**AC50002 Programming Languages for Data Engineering**

**Python Assignment**

**Name: Salar Junaid Qureshi**

**Student ID: 2432682**

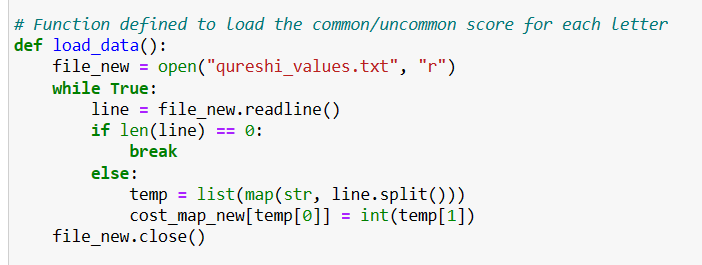
**GitHub Link:** <https://github.com/salarjunaid/Python_Assignment_2432682>

**Code Explanation:**

Three functions were defined:

**1: load\_data():**

This function reads the file qureshi\_values.txt file and splits each line through split() method on every whitespace character (default). It is stored in a list called temp which is then stored in a dictionary cost\_map\_new where the first value of the list is the key and the second value of the list is the value for the corresponding key. Qureshi\_values.txt file consists of letters with their common/uncommon score.

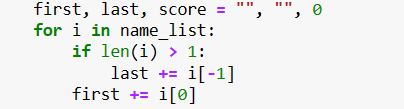


**2: calculate\_score():**

This function calculates score for each abbreviation generated for each name in the input file i.e. qureshi\_trees.txt.

Three variables were initialized: first, last and score. These variables keeps a track on the first letter of each word, last letter of each word and the total score respectively.

All last letters are appended to last variable and all first letters were appended to first variable.



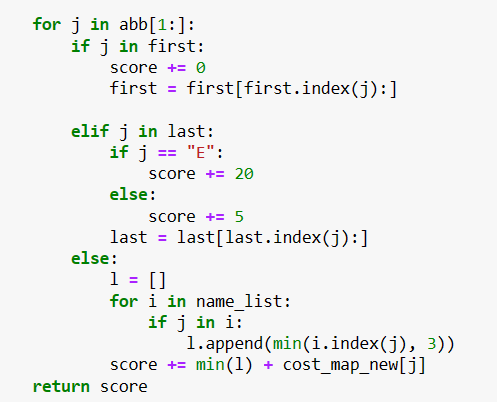
Through a for loop scores are calculated for each abbreviation.

If a character is in the first variable, the score is increased by 0.

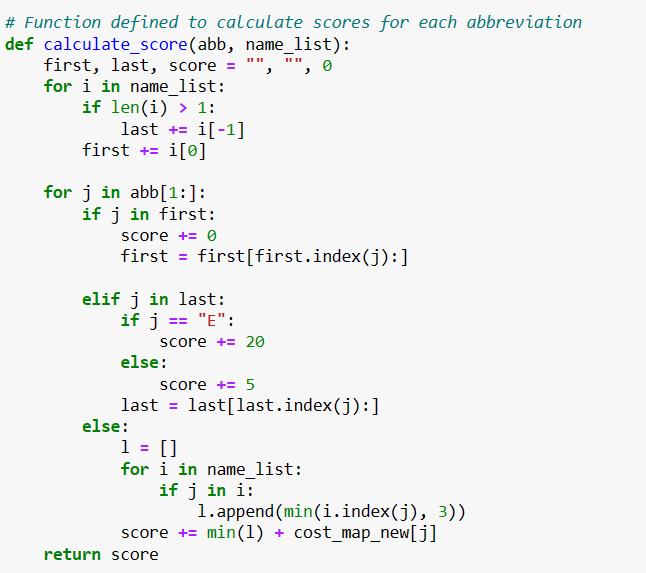
If the character E is in the last variable, the score is increased by 20. If there is any other character in last variable, the score is increased by 5.

If a letter is not in both first and last variable score is increased by the sum of positional score of a letter and the common/uncommon score of the letter.

The total score is returned by the function.



Whole function:

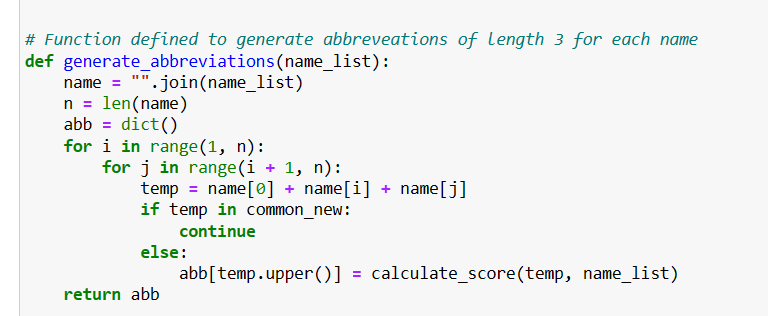


**3: generate\_abbreviations():**

This function generates abbreviations of length 3 for each name in the input file i.e. qureshi\_trees.txt.

Common abbreviations among names are ignored.

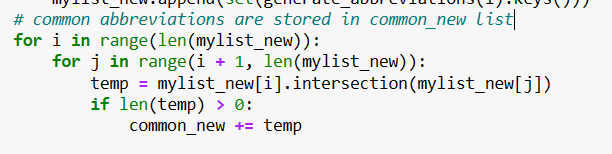
calculate\_score() function is called to calculate scores of each abbreviation.



**Main Function:**

In the main function input file i.e., qureshi\_trees.txt is read and defined functions were called to get the desired output.

All of the common abbreviations among all of the names in the input file are stored in a list called common\_new.



The abbreviations with the smallest score are written in an output file i.e., qureshi\_trees\_abbrevs.txt.

