CS 5710 Machine Learning

Assignment 1

**Name:** Sriram Kethineni

**ID:** 700762460

**Question 1:**

The following is a list of 10 students ages:

𝑎𝑔𝑒𝑠 = [19, 22, 19, 24, 20, 25, 26, 24, 25, 24]

* Sort the list and find the min and max age
* Add the min age and the max age again to the list
* Find the median age (one middle item or two middle items divided by two)
* Find the average age (sum of all items divided by their number) ➢ Find the range of the ages (max minus min)

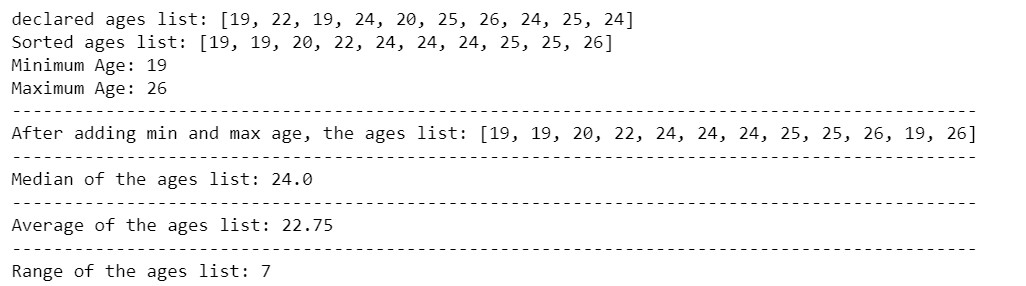
***So***

***urce code***

***:***



***Output:***



***Explanation:***

Here in the code, I have initially declared a list named “ages” with the ages of 10 students and sorted the list using the 𝒔𝒐𝒓𝒕() function.

𝑠𝑜𝑟𝑡(𝑎𝑔𝑒𝑠)

The functions 𝒎𝒊𝒏() and 𝒎𝒂𝒙() are used to find Minimum and the maximum age in the list.

𝑚𝑎𝑥(𝑎𝑔𝑒𝑠), 𝑚𝑖𝑛(𝑎𝑔𝑒𝑠)

𝒂𝒑𝒑𝒆𝒏𝒅() function is used to add the minimum and the maximum age again to the list.

𝑎𝑔𝑒𝑠. 𝑎𝑝𝑝𝑒𝑛𝑑(𝑚𝑎𝑥(𝑎𝑔𝑒𝑠))

𝑎𝑔𝑒𝑠. 𝑎𝑝𝑝𝑒𝑛𝑑(𝑚𝑖𝑛(𝑎𝑔𝑒𝑠))

The𝒎𝒆𝒅𝒊𝒂𝒏()function of 𝑺𝒕𝒂𝒕𝒊𝒔𝒕𝒊𝒄𝒔 module is used to find the median of the list. The 𝑺𝒕𝒂𝒕𝒊𝒔𝒕𝒊𝒄𝒔 module is imported.

𝑠𝑡𝑎𝑡𝑖𝑠𝑡𝑖𝑐𝑠. 𝑚𝑒𝑑𝑖𝑎𝑛(𝑎𝑔𝑒𝑠)

The average of list is shown by dividing sum of all ages by total number of ages in the list. To find the average of the list, functions 𝒔𝒖𝒎() and 𝒍𝒆𝒏() are used. 𝒔𝒖𝒎()function gives the sum of all values in the list and 𝒍𝒆𝒏() function gives the length of the list which determines the capacity (total number) of the list.

𝑎𝑣𝑒𝑟𝑎𝑔𝑒 = 𝑠𝑢𝑚(𝑎𝑔𝑒𝑠)/𝑙𝑒𝑛(𝑎𝑔𝑒𝑠)

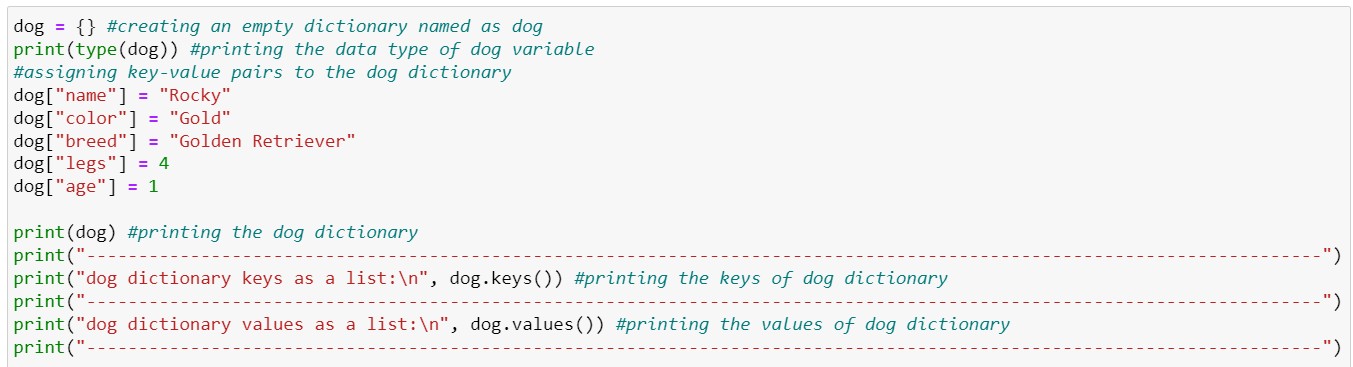
The range of the list is found by making the difference between maximum and minimum age in the list.

𝑅𝑎𝑛𝑔𝑒 = 𝑚𝑎𝑥(𝑎𝑔𝑒𝑠) − min (𝑎𝑔𝑒𝑠)

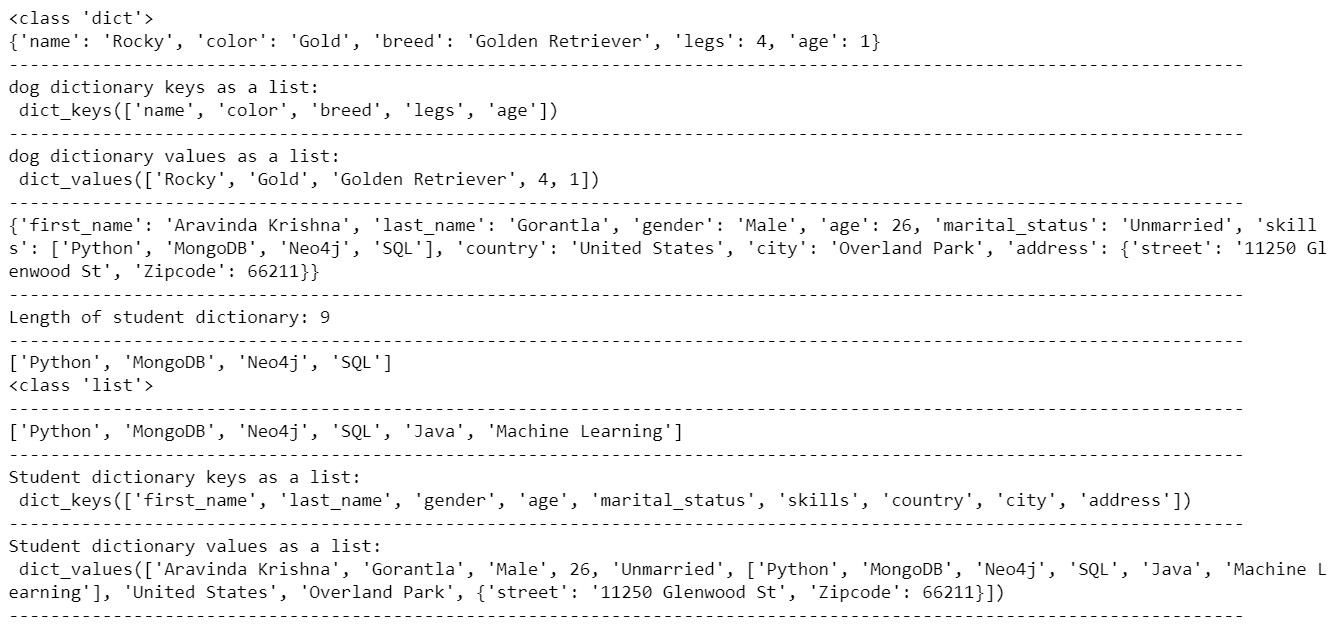
**Question 2:**

* Create an empty dictionary called dog
* Add name, color, breed, legs, age to the dog dictionary
* Create a student dictionary and add first\_name, last\_name, gender, age, marital status, skills, country, city, and address as keys for the dictionary
* Get the length of the student dictionary
* Get the value of skills and check the data type, it should be a list
* Modify the skills values by adding one or two skills
* Get the dictionary keys as a list
* Get the dictionary values as a list

***Source code:***



***Output:***



***Explanation:***

Here in the code, I have declared an empty dictionary named “dog” and the key value pairs are assigned as below.

𝑑𝑜𝑔 = {}

𝑑𝑜𝑔["𝑛𝑎𝑚𝑒"] = "𝑅𝑜𝑐𝑘𝑦"

𝑑𝑜𝑔["𝑐𝑜𝑙𝑜𝑟"] = "𝐺𝑜𝑙𝑑"

𝑑𝑜𝑔["𝑏𝑟𝑒𝑒𝑑"] = "𝐺𝑜𝑙𝑑𝑒𝑛 𝑅𝑒𝑡𝑟𝑖𝑒𝑣𝑒𝑟"

𝑑𝑜𝑔["𝑙𝑒𝑔𝑠"] = 4

𝑑𝑜𝑔["𝑎𝑔𝑒"] = 1

The dictionary dog is printed with its keys and values.

{′𝑛𝑎𝑚𝑒′: ′𝑅𝑜𝑐𝑘𝑦′, ′𝑐𝑜𝑙𝑜𝑟′: ′𝐺𝑜𝑙𝑑′, ′𝑏𝑟𝑒𝑒𝑑′: ′𝐺𝑜𝑙𝑑𝑒𝑛 𝑅𝑒𝑡𝑟𝑖𝑒𝑣𝑒𝑟′, ′𝑙𝑒𝑔𝑠′: 4, ′𝑎𝑔𝑒′: 1}

A dictionary named “student” is declared which includes the fields such as first\_name, last\_name, age, marital status, skills, country, city, and address as keys and its values are also assigned as shown below.

𝑠𝑡𝑢𝑑𝑒𝑛𝑡 = {′𝑓𝑖𝑟𝑠𝑡\_𝑛𝑎𝑚𝑒′: ′𝐴𝑟𝑎𝑣𝑖𝑛𝑑𝑎 𝐾𝑟𝑖𝑠ℎ𝑛𝑎′,

′𝑙𝑎𝑠𝑡\_𝑛𝑎𝑚𝑒′: ′𝐺𝑜𝑟𝑎𝑛𝑡𝑙𝑎′,

′𝑔𝑒𝑛𝑑𝑒𝑟′: ′𝑀𝑎𝑙𝑒′,

′𝑎𝑔𝑒′: 26,

′𝑚𝑎𝑟𝑖𝑡𝑎𝑙\_𝑠𝑡𝑎𝑡𝑢𝑠′: ′𝑈𝑛𝑚𝑎𝑟𝑟𝑖𝑒𝑑′,

′𝑠𝑘𝑖𝑙𝑙𝑠′: [′𝑃𝑦𝑡ℎ𝑜𝑛′, ′𝑀𝑜𝑛𝑔𝑜𝐷𝐵′, ′𝑁𝑒𝑜4𝑗′, ′𝑆𝑄𝐿′],

′𝑐𝑜𝑢𝑛𝑡𝑟𝑦′: ′𝑈𝑛𝑖𝑡𝑒𝑑 𝑆𝑡𝑎𝑡𝑒𝑠′,

′𝑐𝑖𝑡𝑦′: ′𝑂𝑣𝑒𝑟𝑙𝑎𝑛𝑑 𝑃𝑎𝑟𝑘′,

′𝑎𝑑𝑑𝑟𝑒𝑠𝑠′: {′𝑠𝑡𝑟𝑒𝑒𝑡′: ′11250 𝐺𝑙𝑒𝑛𝑤𝑜𝑜𝑑 𝑆𝑡′, ′𝑍𝑖𝑝𝑐𝑜𝑑𝑒′: 66211}}

The dictionary student is printed with its keys and values.

{′𝑓𝑖𝑟𝑠𝑡\_𝑛𝑎𝑚𝑒′: ′𝐴𝑟𝑎𝑣𝑖𝑛𝑑𝑎 𝐾𝑟𝑖𝑠ℎ𝑛𝑎′, ′𝑙𝑎𝑠𝑡\_𝑛𝑎𝑚𝑒′: ′𝐺𝑜𝑟𝑎𝑛𝑡𝑙𝑎′, ′𝑔𝑒𝑛𝑑𝑒𝑟′: ′𝑀𝑎𝑙𝑒′, ′𝑎𝑔𝑒′: 26,

′𝑚𝑎𝑟𝑖𝑡𝑎𝑙\_𝑠𝑡𝑎𝑡𝑢𝑠′: ′𝑈𝑛𝑚𝑎𝑟𝑟𝑖𝑒𝑑′, ′𝑠𝑘𝑖𝑙𝑙𝑠′: [′𝑃𝑦𝑡ℎ𝑜𝑛′, ′𝑀𝑜𝑛𝑔𝑜𝐷𝐵′, ′𝑁𝑒𝑜4𝑗′, ′𝑆𝑄𝐿′],

′𝑐𝑜𝑢𝑛𝑡𝑟𝑦′: ′𝑈𝑛𝑖𝑡𝑒𝑑 𝑆𝑡𝑎𝑡𝑒𝑠′, ′𝑐𝑖𝑡𝑦′: ′𝑂𝑣𝑒𝑟𝑙𝑎𝑛𝑑 𝑃𝑎𝑟𝑘′,

′𝑎𝑑𝑑𝑟𝑒𝑠𝑠′: {′𝑠𝑡𝑟𝑒𝑒𝑡′: ′11250 𝐺𝑙𝑒𝑛𝑤𝑜𝑜𝑑 𝑆𝑡′, ′𝑍𝑖𝑝𝑐𝑜𝑑𝑒′: 66211}}

The length of student dictionary is found using the function 𝑙𝑒𝑛().

𝑙𝑒𝑛(𝑠𝑡𝑢𝑑𝑒𝑛𝑡)

The values of skills are printed accessing the skills from student dictionary.

𝑠𝑡𝑢𝑑𝑒𝑛𝑡["𝑠𝑘𝑖𝑙𝑙𝑠"]

The data type of skills has been found using the 𝑡𝑦𝑝𝑒() function.

𝑡𝑦𝑝𝑒(𝑠𝑡𝑢𝑑𝑒𝑛𝑡["𝑠𝑘𝑖𝑙𝑙𝑠"])

The Skills list is modified by adding new skills “Java” and “Machine Learning” to the list using the 𝑎𝑝𝑝𝑒𝑛𝑑() function.

𝑠𝑡𝑢𝑑𝑒𝑛𝑡["𝑠𝑘𝑖𝑙𝑙𝑠"]. 𝑎𝑝𝑝𝑒𝑛𝑑("𝐽𝑎𝑣𝑎")

𝑠𝑡𝑢𝑑𝑒𝑛𝑡["𝑠𝑘𝑖𝑙𝑙𝑠"]. 𝑎𝑝𝑝𝑒𝑛𝑑("𝑀𝑎𝑐ℎ𝑖𝑛𝑒 𝐿𝑒𝑎𝑟𝑛𝑖𝑛𝑔")

The dictionary keys are printed using the 𝑘𝑒𝑦𝑠() function.

𝑑𝑜𝑔. 𝑘𝑒𝑦𝑠()

𝑠𝑡𝑢𝑑𝑒𝑛𝑡. 𝑘𝑒𝑦𝑠()

The dictionary values are printed using the 𝑣𝑎𝑙𝑢𝑒𝑠() function.

𝑑𝑜𝑔. 𝑣𝑎𝑙𝑢𝑒𝑠()

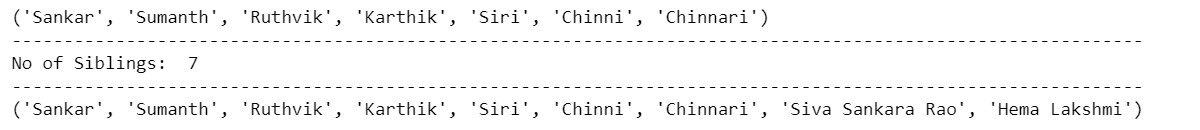
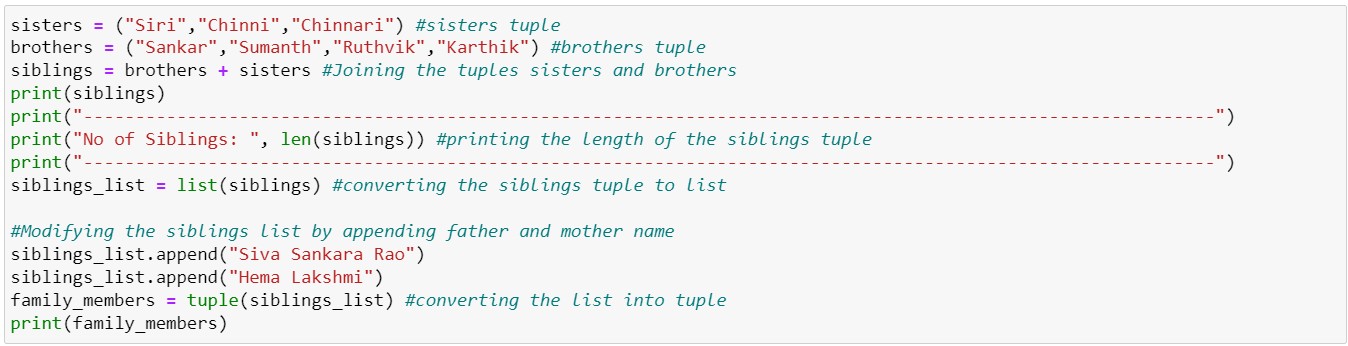
𝑠𝑡𝑢𝑑𝑒𝑛𝑡. 𝑣𝑎𝑙𝑢𝑒𝑠()

**Question 3:**

* Create a tuple containing names of your sisters and your brothers (imaginary siblings are fine)
* Join brothers and sisters’ tuples and assign it to siblings ➢ How many siblings do you have?
* Modify the sibling’s tuple and add the name of your father and mother and assign it to family\_members

***Source code:***

***Output:***



***Explanation:***

Here in this code, I have created two tuples named sisters and brothers with some imaginary values. Then joined both the tuples and created a new tuple named siblings.

Now converted the tuple into a list and then with the help of 𝑎𝑝𝑝𝑒𝑛𝑑() function in lists, I added the father and mother name. Then converted the list back to tuple and named it as family\_members.

**Question 4:**

𝑖𝑡\_𝑐𝑜𝑚𝑝𝑎𝑛𝑖𝑒𝑠 = {′𝐹𝑎𝑐𝑒𝑏𝑜𝑜𝑘′, ′𝐺𝑜𝑜𝑔𝑙𝑒′,′𝑀𝑖𝑐𝑟𝑜𝑠𝑜𝑓𝑡′,′𝐴𝑝𝑝𝑙𝑒′,′𝐼𝐵𝑀′,′𝑂𝑟𝑎𝑐𝑙𝑒′, ′𝐴𝑚𝑎𝑧𝑜𝑛′}

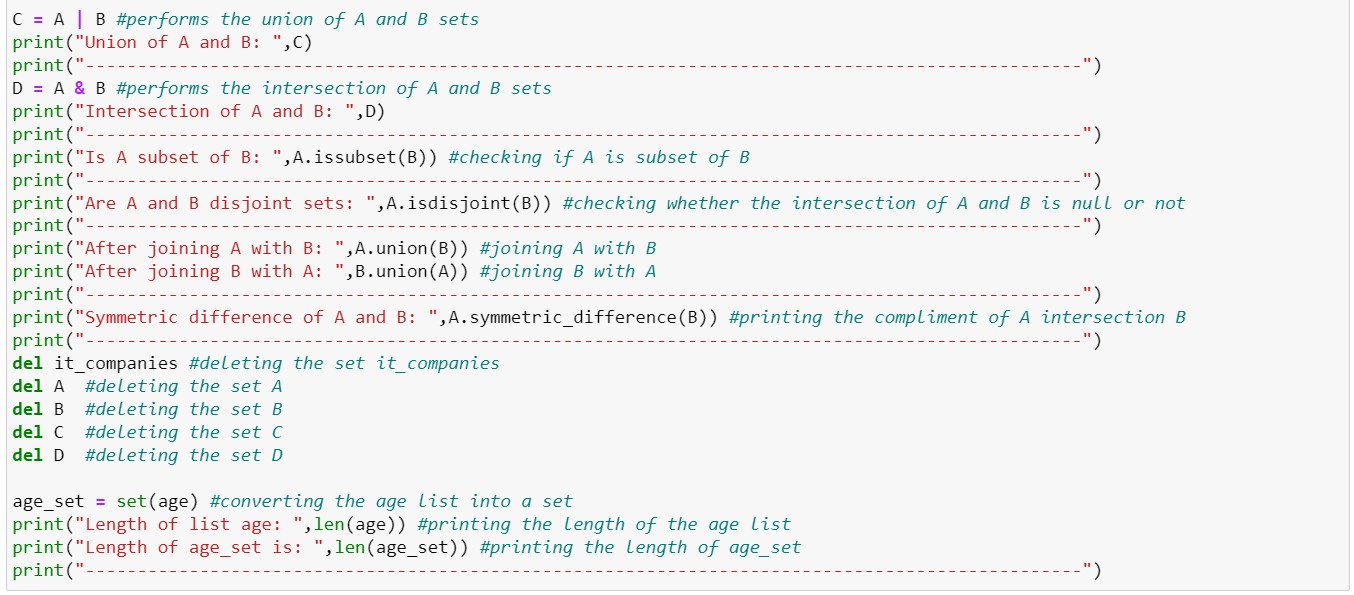
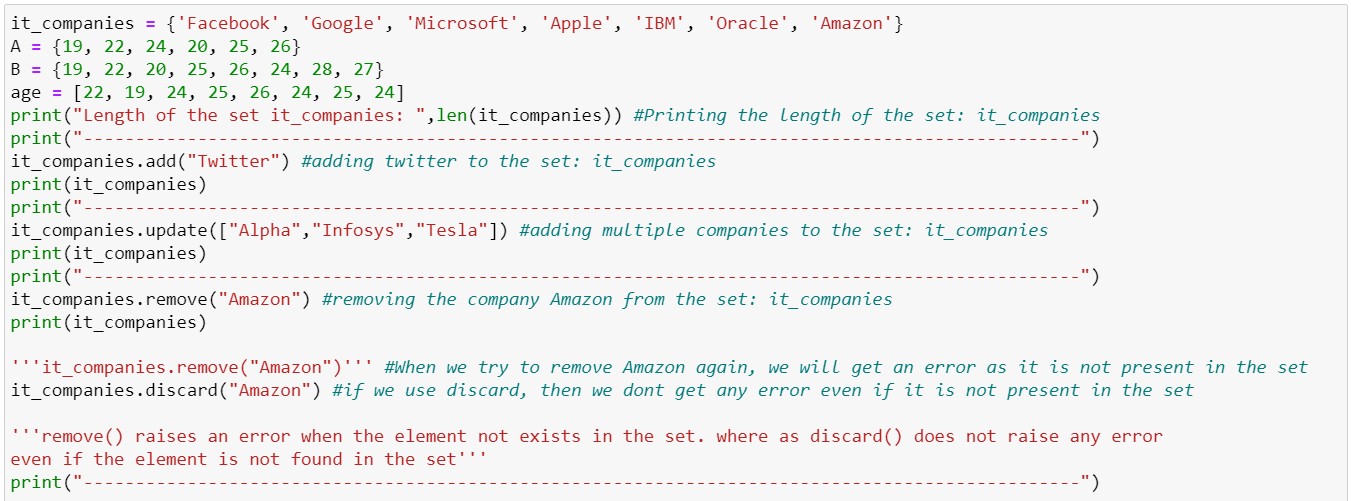
𝐴 = {19,22, 24, 20,25, 26}

𝐵 = {19, 22, 20,25, 26,24, 28,27}

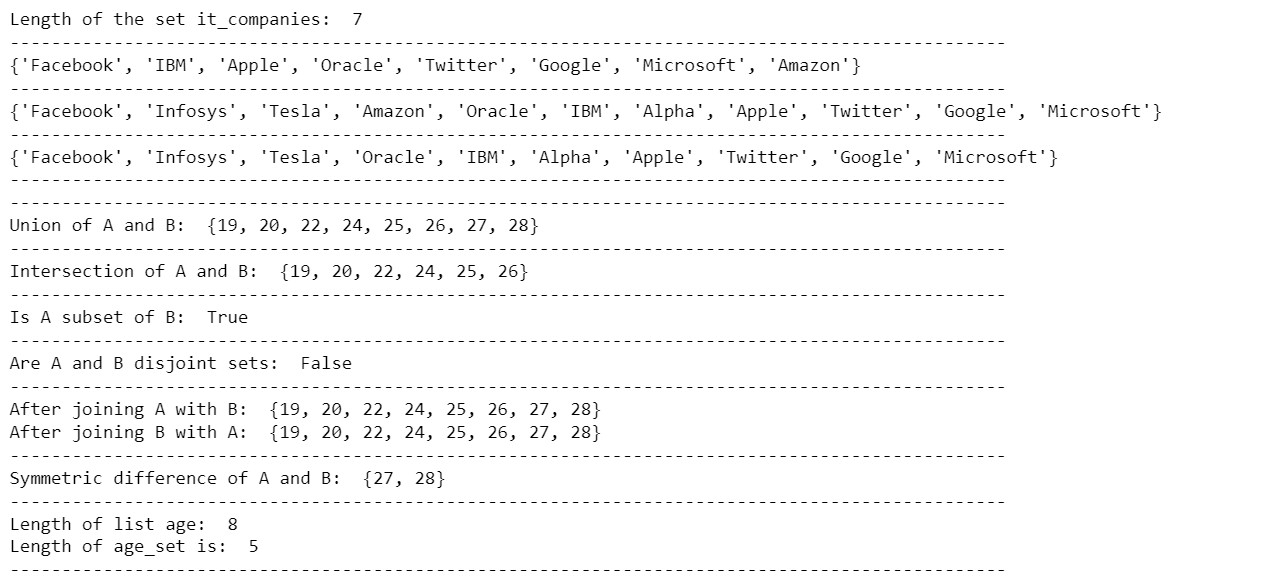
𝑎𝑔𝑒 = [22,19, 24,25,26, 24,25, 24]

* Find the length of the set it\_companies
* Add 'Twitter' to it\_companies
* Insert multiple IT companies at once to the set it\_companies
* Remove one of the companies from the set it\_companies
* What is the difference between remove and discard
* Join A and B
* Find A intersection B
* Is A subset of B
* Are A and B disjoint sets
* Join A with B and B with A
* What is the symmetric difference between A and B
* Delete the sets completely
* Convert the ages to a set and compare the length of the list and the set.

***Source code:***



***Output:***



***Explanation:***

Here in the code, I have declared a set named it\_companies and its Length is found using 𝒍𝒆𝒏() function. By using 𝒂𝒅𝒅() function, couple of elements are added one by one to the set it\_companies and printed the new set. Then we added multiple values into the set at once using 𝒖𝒑𝒅𝒂𝒕𝒆() function and printed the updated set.

Using the 𝒓𝒆𝒎𝒐𝒗𝒆() function removed a value and the resultant set is printed.

*Difference between* 𝑟𝑒𝑚𝑜𝑣𝑒() *and* 𝑑𝑖𝑠𝑐𝑎𝑟𝑑() *functions:*

𝑟𝑒𝑚𝑜𝑣𝑒() function raises an error when the element does not exist in the set. whereas 𝑑𝑖𝑠𝑐𝑎𝑟𝑑() function do not raise any error even if the element is not found in the set.

Now, I have declared two sets 𝐴 and 𝐵 with some values.

Set functions stated below are performed and printed the results.

𝒖𝒏𝒊𝒐𝒏() – Joins 2 sets. Can be performed using 𝐴 | 𝐵

𝒊𝒏𝒕𝒆𝒓𝒔𝒆𝒄𝒕𝒊𝒐𝒏() – the result set consists of common elements from both sets. Can be performed using 𝐴 & 𝐵

𝒊𝒔𝒔𝒖𝒃𝒔𝒆𝒕()– checks whether a set is subset of another

𝒊𝒔𝒅𝒊𝒔𝒋𝒐𝒊𝒏𝒕()– checks whether the intersection of 𝐴 and 𝐵 is a null or not 𝒔𝒚𝒎𝒎𝒆𝒕𝒓𝒊𝒄\_𝒅𝒊𝒇𝒇𝒆𝒓𝒆𝒏𝒄𝒆()– gives the complement of 𝐴 intersection 𝐵.

‘**del’** is used to delete the sets.

Now a list age is declared. Typecasting is done, i.e., converting the list into a set named age\_set. The length of the list and set are calculated. There is a difference in both the length values as the set does not allows duplicates.

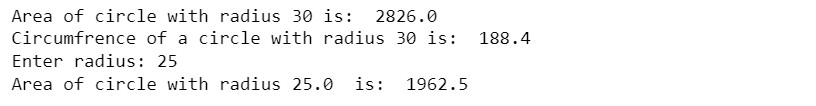
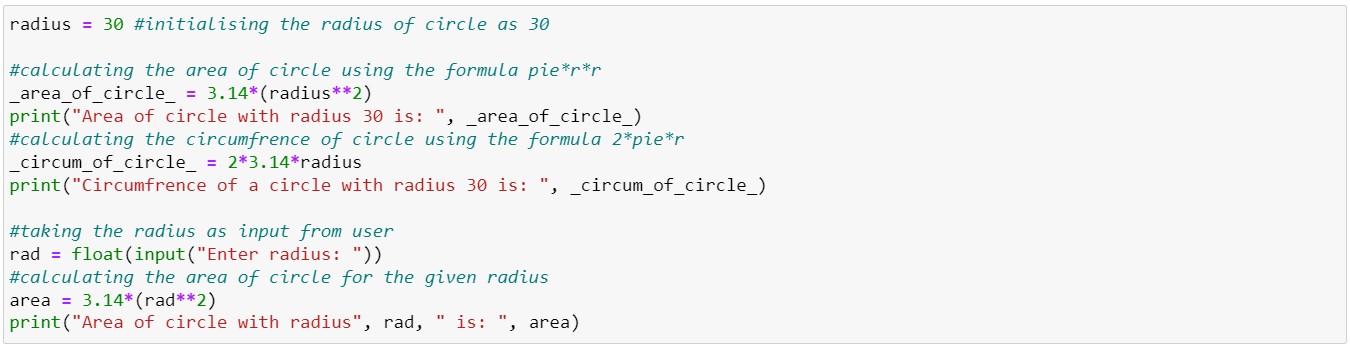
**5:**

The radius of a circle is 30 meters.

* Calculate the area of a circle and assign the value to a variable name of \_area\_of\_circle ➢ Calculate the circumference of a circle and assign the value to a variable name of \_circum\_of\_circle
* Take radius as user input and calculate the area.

***Source code:***

***Output:***



***Explanation:***

Here in the code, I have declared radius of the circle and initialized to 30.

With that radius, area and circumference of the circle are determined and printed.

Now radius is taken as the input from the user using 𝑖𝑛𝑝𝑢𝑡() function. With the user provided radius, area is calculated and printed.

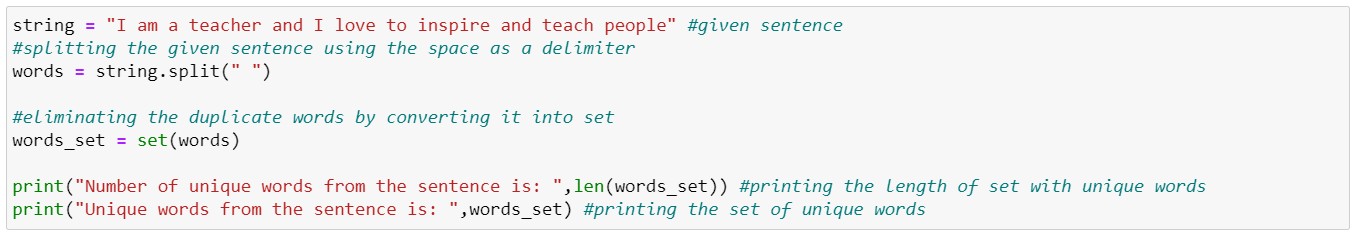
**6:**

“𝐼 𝑎𝑚 𝑎 𝑡𝑒𝑎𝑐ℎ𝑒𝑟 𝑎𝑛𝑑 𝐼 𝑙𝑜𝑣𝑒 𝑡𝑜 𝑖𝑛𝑠𝑝𝑖𝑟𝑒 𝑎𝑛𝑑 𝑡𝑒𝑎𝑐ℎ 𝑝𝑒𝑜𝑝𝑙𝑒”

How many unique words have been used in the sentence? Use the split methods and set to get the unique words.

***Source code:***

***Output:***



***Explanation:***

A variable named string is declared and initialized with the given text. The string is made into words as stored as a list using 𝑠𝑝𝑙𝑖𝑡() function and space as the delimiter. The list is typecasted as set because set will not allow duplicates.

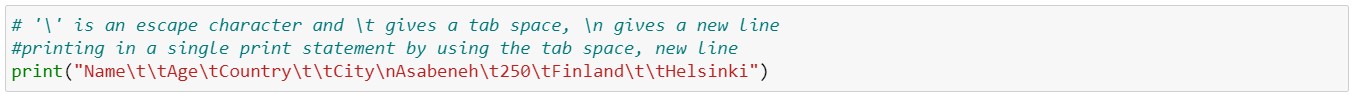
Now the length of the set gives number of unique words and set gives the unique words. Both the unique words and its count are printed. **Question 7:**

Use a tab escape sequence to get the following lines.

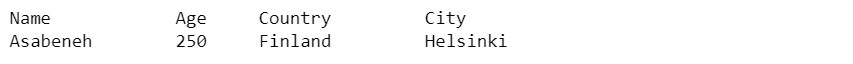
**Name Age Country City**

**Asabeneh 250 Finland Helsinki**

***Source code:***



***Output:***



***Explanation:***

Both the lines given are printed in the same pattern using the escape tab sequence “\t” which gives 3 space characters and “\n” used to get a new line.

**8:**

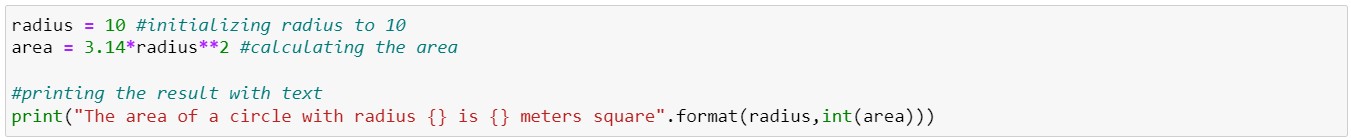
Use the string formatting method to display the following:

𝑟𝑎𝑑𝑖𝑢𝑠 

𝑎𝑟𝑒𝑎  𝑟𝑎𝑑𝑖𝑢𝑠 

“𝑇𝑒 𝑎𝑟𝑒𝑎 𝑜𝑓 𝑎 𝑐𝑖𝑟𝑐𝑙𝑒 𝑤𝑖𝑡 𝑟𝑎𝑑𝑖𝑢𝑠 10 𝑖𝑠 314 𝑚𝑒𝑡𝑒𝑟𝑠 𝑠𝑞𝑢𝑎𝑟𝑒. ”

***Source code:***



***Output:***



***Explanation:***

Here in the code, radius is initialized and area is calculated. The 𝑓𝑜𝑟𝑚𝑎𝑡() method formats the specified value(s) and inserts them inside the string's placeholder.

**9:**

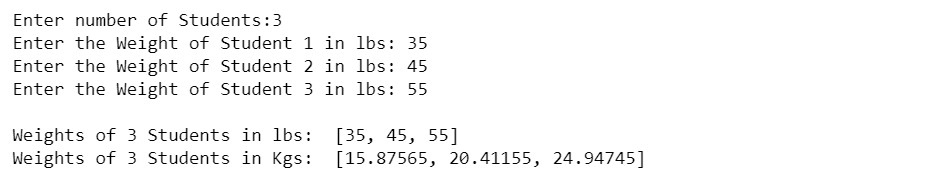
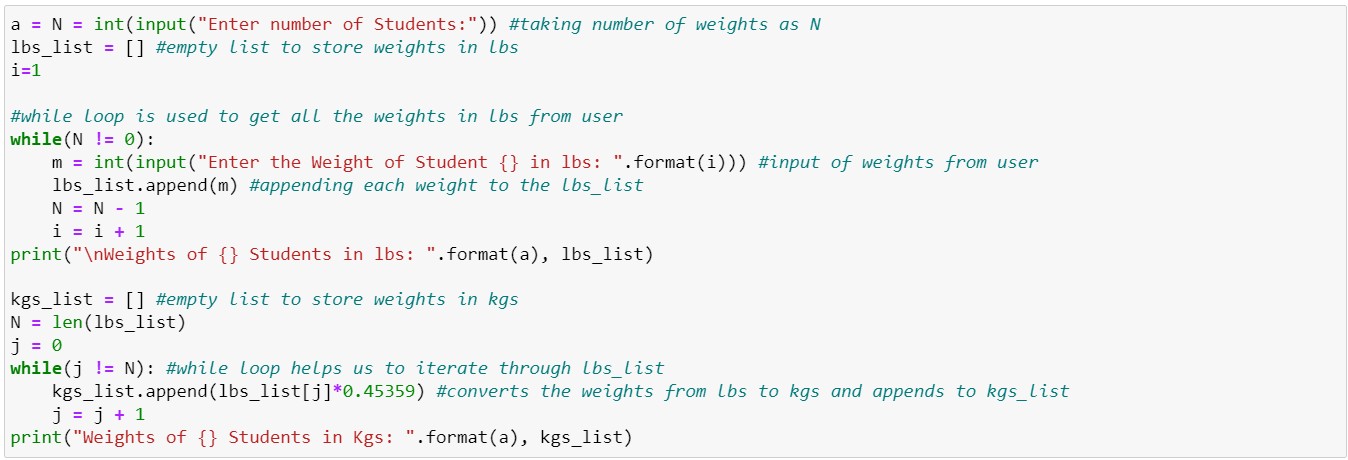
Write a program, which reads weights (lbs.) of N students into a list and convert these weights to kilograms in a separate list using Loop. N: No of students (Read input from user)

Ex: L1: [150, 155, 145, 148]

Output: [68.03, 70.3, 65.77, 67.13]

***Source code:***

***Output:***



***Explanation:***

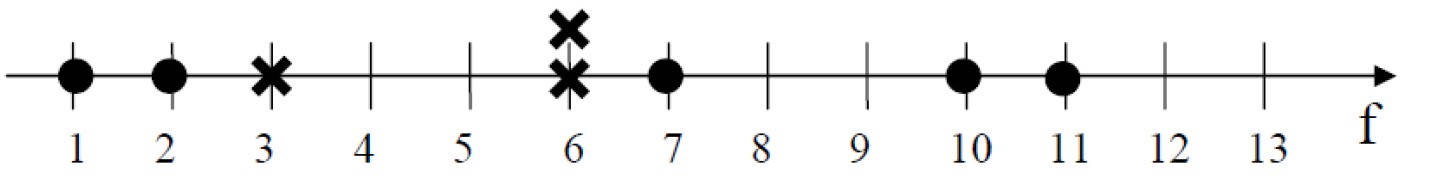
Here in the code, user input is taken for the number of students and an empty list is created. By using while loop, the individual weights of students in lbs. is taken as input and added to the empty list.

Another empty list is created. By using while loop, the individual weights of students in lbs. converted **to kgs.** is done and added to the new list, the list of converted values is printed.

Conversion formulae: 1 𝑘𝑔 = 0.45359 𝑙𝑏𝑠

**10:**

The diagram below shows a dataset with 2 classes and 8 data points, each with only one feature value, labeled 𝑓. Note that there are two data points with the same feature value of 6. These are shown as two x’s one above the other. Provide stepwise mathematical solution, do not write code for it.



1. Divide this data equally into two parts. Use first part as training and second part as testing. Using KNN classifier, for 𝐾 = 3, what would be the predicted outputs for the test samples? Show how you arrived at your answer.
2. Compute the confusion matrix for this and calculate accuracy, sensitivity, and specificity values.

***Solution:***

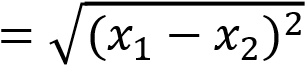
Given data elements from the diagram are taken in the tabular form as below.

|  |  |
| --- | --- |
| **Feature** | **Label** |
| 1 | O |
| 2 | O |
| 3 | X |
| 6 | X |
| 6 | X |
| 7 | O |
| 10 | O |
| 11 | O |

The given dataset is divided equally and first four rows of data in the table are considered to be the Training dataset and the next four rows are selected as the Testing dataset.

Using KNN Classifier considering 𝐾 = 3, the distance between the testing and training data is demonstrated below.

In KNN, we are going to find the Euclidean distance between the data points.

𝑑  (For a single feature)

The Euclidean distance between the data points is calculated and placed in the below table, where the columns are the training dataset and rows are the testing dataset.

The highlighted rows (Made bold) are the distanced values.

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
|  | 1(O) | 2(O) | 3(X) | 6(X) |
| 6 | 5 | **4** | **3** | **0** |
| 7 | 6 | **5** | **4** | **1** |
| 10 | 9 | **8** | **7** | **4** |
| 11 | 10 | **9** | **8** | **5** |

Let us assume ‘O’ as negative and ‘X’ as the positive values.

Confusion matrix for a prediction would be:

|  |  |
| --- | --- |
| TN | FP |
| FN | TP |

**True Positive (TP):** When the actual label is positive (X) and your machine learning model also predicts that label as positive (X).

**True Negative (TN):** When the actual label is negative (O) and your machine learning model also predicts that label as negative (O).

**False Positive (FP):** When the actual label is negative (O), but the machine learning model predicts that label as positive (X).

**False Negative (FN):** When the actual label is positive (X), but the machine learning model predicts that label as negative (O).

The prediction on testing data is below.

|  |  |  |  |
| --- | --- | --- | --- |
|  | True label | Predicted label | O/P |
| 6 | X | X | TP |
| 7 | O | X | FP |
| 10 | O | X | FP |
| 11 | O | X | FP |

The final Confusion matrix for the above prediction is:

|  |  |
| --- | --- |
| 0 | 3 |
| 0 | 1 |

Accuracy of the classifier = (𝑇𝑃 + 𝑇𝑁) = = 0.25

(𝑃 + 𝑁)

Sensitivity of the classifier = 𝑇𝑃 = 𝑇𝑃 = 1 = 1 𝑇𝑃+𝐹𝑁 𝑃 1

1

4

Specificity of the classifier = 𝑇𝑁 = 𝑇𝑁 = 0 = 0

𝐹𝑃+𝑇𝑁 𝑁 3

# - - - - End - - - -