ex1-ds-batch-2

December 16, 2023

[ ]:
```
# EX NO:1 Introduction to Data Science with Python Programming
```

[ ]:
```
DATE : 11/12/2023
REG NO:URK22AI1048
```

[ ]:
```
'''
AIM
    To Write a python program to demonstrate the basics of dictionary, list,␣
 ↪set, tuple, array and
csv files for data science application.

DESCRIPTION
            Python is a high-level general-purpose programming language. Python␣
 ↪is one of the most
popular and widely used programming languages in the world Python is used for␣
 ↪data
analytics, machine learning, and even design.

Lists
        Lists are mutable data types in Python. Lists is a 0-based index␣
 ↪datatype meaning the index
of the first element starts at 0. Lists are used to store multiple items in a␣
 ↪single variable.
Lists are one of the 4 data types present in Python.

Dictionary
            Dictionaries are mutable data types in nature meaning they can be␣
 ↪updated after they are
created. Syntactically they are written in a key, value pair format inside␣
 ↪curly braces.

Keys
        Keys are always unique and there cannot be any duplicates. There is no␣
 ↪index in the
dictionary meaning they are not ordered. Key is the default iterator and is␣
 ↪used to retrieve
the value.
```

```
Tuple
        Tuples are the same as lists are with the exception that the data once␣
 ↪entered into the tuple
cannot be changed no matter what. The only exception is when the data inside␣
 ↪the tuple is
mutable, only then the tuple data can be changed.

Sets
        Sets are a collection of unordered elements that are unique. Meaning␣
 ↪that even if the data is
repeated more than one time, it would be entered into the set only once. It␣
 ↪resembles the
sets that you have learnt in arithmetic. The operations also are the same as is␣
 ↪with the
arithmetic sets.

Numpy Array
        An array is a central data structure of the NumPy library. An array is␣
 ↪a grid of values and it
contains information about the raw data, how to locate an element, and how to␣
 ↪interpret an
element.

CSV Files
        CSV stands for "Comma Separated Values." It is the simplest form of␣
 ↪storing data in tabular
form as plain text. It is important to know to work with CSV because we mostly␣
 ↪rely on
CSV data in our day-to-day lives as data scientists.


'''
```

[5]:
```
#URK22AI1048 Q1

sub_dict = {}

num_entries = int(input("Enter the number of subjects: "))
for _ in range(num_entries):
    sub_code = input("Enter subject code: ")
    sub_name = input("Enter subject name: ")
    sub_dict[sub_code] = sub_name


search_sub_code = input("Enter a subject code to find its name: ")
```

```
sub_name_found = sub_dict.get(search_sub_code)


if sub_name_found is not None:
    print(f"The subject name for code {search_sub_code} is: {sub_name_found}")
else:
    print(f"Subject code {search_sub_code} not found in the dictionary.")
```

```
Enter the number of subjects:  2
Enter subject code:  13
Enter subject name:  maths
Enter subject code:  14
Enter subject name:  science
Enter a subject code to find its name:  14
```

```
The subject name for code 14 is: science
```

[6]:
```
#URK22AI1048 Q2

word_list = []

num_words = int(input("Enter the number of words: "))
for _ in range(num_words):
    word = input("Enter a word: ")
    word_list.append(word)


n = int(input("Enter the minimum length (n): "))


result_list = [word for word in word_list if len(word) > n]


print(f"List of words longer than {n}: {result_list}")
```

```
Enter the number of words:  2
Enter a word:  hi
Enter a word:  hari
Enter the minimum length (n):  3
```

```
List of words longer than 3: ['hari']
```

[7]:
```
#URK22AI1048 Q3

my_set = set()
```

```
num_values = int(input("Enter the number of values: "))
for _ in range(num_values):
    value = input("Enter a value: ")
    my_set.add(value)


search_value = input("Enter a value to check if it's present in the set: ")

if search_value in my_set:
    print(f"{search_value} is present in the set.")
else:
    print(f"{search_value} is not present in the set.")
```

```
Enter the number of values:  2
Enter a value:  1
Enter a value:  3
Enter a value to check if it's present in the set:  3
```

```
3 is present in the set.
```

[9]:
```
#URK22AI1048 Q4

my_tuple = ()


num_values = int(input("Enter the number of values: "))
my_tuple = tuple(int(input("Enter a value: ")) for _ in range(num_values))


search_number = int(input("Enter a number to count its occurrences in the tuple:
 ↪ "))


occurrences = my_tuple.count(search_number)


print(f"Number of occurrences of {search_number} in the tuple: {occurrences}")
```

```
Enter the number of values:  3
Enter a value:  1
Enter a value:  2
Enter a value:  2
Enter a number to count its occurrences in the tuple:  2
```

```
Number of occurrences of 2 in the tuple: 2
```

[11]:
```
#URK22AI1048 Q5
import numpy as np
```

```
matrix1 = np.array([[1, 2, 3], [4, 5, 6], [7, 8, 9]])
matrix2 = np.array([[9, 8, 7], [6, 5, 4], [3, 2, 1]])

result_matrix = matrix1 - matrix2

print("Matrix 1:")
print(matrix1)
print("\nMatrix 2:")
print(matrix2)
print("\nResult Matrix (Subtraction):")
print(result_matrix)
```

```
Matrix 1:
[[1 2 3]
 [4 5 6]
 [7 8 9]]

Matrix 2:
[[9 8 7]
 [6 5 4]
 [3 2 1]]

Result Matrix (Subtraction):
[[-8 -6 -4]
 [-2  0  2]
 [ 4  6  8]]
```

[3]:
```
#URK22AI1048 Q6;
import pandas as pd

df = pd.read_csv('titanic3.csv')

print(df.head())
```

```
   pclass  survived                                             name     sex  \
0     1.0       1.0                    Allen, Miss. Elisabeth Walton  female
1     1.0       1.0                   Allison, Master. Hudson Trevor    male
2     1.0       0.0                     Allison, Miss. Helen Loraine  female
3     1.0       0.0             Allison, Mr. Hudson Joshua Creighton    male
4     1.0       0.0  Allison, Mrs. Hudson J C (Bessie Waldo Daniels)  female

       age  sibsp  parch  ticket      fare    cabin embarked boat   body  \
0  29.0000    0.0    0.0   24160  211.3375       B5        S    2    NaN
1   0.9167    1.0    2.0  113781  151.5500  C22 C26        S   11    NaN
2   2.0000    1.0    2.0  113781  151.5500  C22 C26        S  NaN    NaN
3  30.0000    1.0    2.0  113781  151.5500  C22 C26        S  NaN  135.0
4  25.0000    1.0    2.0  113781  151.5500  C22 C26        S  NaN    NaN
```

```
                        home.dest
0                     St Louis, MO
1  Montreal, PQ / Chesterville, ON
2  Montreal, PQ / Chesterville, ON
3  Montreal, PQ / Chesterville, ON
4  Montreal, PQ / Chesterville, ON
```

[16]:
```
df.size
```

[16]:
```
18340
```

[15]:
```
df.shape
```

[15]:
```
(1310, 14)
```

[14]:
```
df.ndim
```

[14]:
```
2
```

[4]:
```
df.info
```

[4]:
```
<bound method DataFrame.info of       pclass  survived
name  \
0        1.0       1.0                    Allen, Miss. Elisabeth Walton
1        1.0       1.0                   Allison, Master. Hudson Trevor
2        1.0       0.0                     Allison, Miss. Helen Loraine
3        1.0       0.0             Allison, Mr. Hudson Joshua Creighton
4        1.0       0.0  Allison, Mrs. Hudson J C (Bessie Waldo Daniels)
…        …         …                                                 …
1305     3.0       0.0                          Zabour, Miss. Thamine
1306     3.0       0.0                       Zakarian, Mr. Mapriededer
1307     3.0       0.0                             Zakarian, Mr. Ortin
1308     3.0       0.0                           Zimmerman, Mr. Leo
1309     NaN       NaN                                           NaN

          sex      age  sibsp  parch  ticket      fare    cabin embarked boat  \
0      female  29.0000    0.0    0.0   24160  211.3375       B5        S    2
1        male   0.9167    1.0    2.0  113781  151.5500  C22 C26        S   11
2      female   2.0000    1.0    2.0  113781  151.5500  C22 C26        S  NaN
3        male  30.0000    1.0    2.0  113781  151.5500  C22 C26        S  NaN
4      female  25.0000    1.0    2.0  113781  151.5500  C22 C26        S  NaN
…       …       …    …    …    …      …         …       …       …  …
1305   female      NaN    1.0    0.0    2665   14.4542      NaN        C  NaN
1306     male  26.5000    0.0    0.0    2656    7.2250      NaN        C  NaN
1307     male  27.0000    0.0    0.0    2670    7.2250      NaN        C  NaN
1308     male  29.0000    0.0    0.0  315082    7.8750      NaN        S  NaN
```

```
1309     NaN      NaN    NaN    NaN     NaN       NaN      NaN       NaN  NaN

       body                        home.dest
0       NaN                      St Louis, MO
1       NaN  Montreal, PQ / Chesterville, ON
2       NaN  Montreal, PQ / Chesterville, ON
3     135.0  Montreal, PQ / Chesterville, ON
4       NaN  Montreal, PQ / Chesterville, ON
…     …                                 …
1305    NaN                             NaN
1306  304.0                             NaN
1307    NaN                             NaN
1308    NaN                             NaN
1309    NaN                             NaN

[1310 rows x 14 columns]>
```

[8]:
```
df.describe
```

[8]:
```
<bound method NDFrame.describe of       pclass  survived
name  \
0        1.0       1.0                    Allen, Miss. Elisabeth Walton
1        1.0       1.0                   Allison, Master. Hudson Trevor
2        1.0       0.0                     Allison, Miss. Helen Loraine
3        1.0       0.0             Allison, Mr. Hudson Joshua Creighton
4        1.0       0.0  Allison, Mrs. Hudson J C (Bessie Waldo Daniels)
…      …       …                                                 …
1305     3.0       0.0                          Zabour, Miss. Thamine
1306     3.0       0.0                       Zakarian, Mr. Mapriededer
1307     3.0       0.0                             Zakarian, Mr. Ortin
1308     3.0       0.0                           Zimmerman, Mr. Leo
1309     NaN       NaN                                          NaN

         sex      age  sibsp  parch  ticket      fare    cabin embarked boat  \
0     female  29.0000    0.0    0.0   24160  211.3375       B5        S    2
1       male   0.9167    1.0    2.0  113781  151.5500  C22 C26        S   11
2     female   2.0000    1.0    2.0  113781  151.5500  C22 C26        S  NaN
3       male  30.0000    1.0    2.0  113781  151.5500  C22 C26        S  NaN
4     female  25.0000    1.0    2.0  113781  151.5500  C22 C26        S  NaN
…      …       …    …    …    …     …       …       …       …  …
1305  female      NaN    1.0    0.0    2665   14.4542      NaN        C  NaN
1306    male  26.5000    0.0    0.0    2656    7.2250      NaN        C  NaN
1307    male  27.0000    0.0    0.0    2670    7.2250      NaN        C  NaN
1308    male  29.0000    0.0    0.0  315082    7.8750      NaN        S  NaN
1309     NaN      NaN    NaN    NaN     NaN       NaN      NaN      NaN  NaN

       body                        home.dest
```

```
0       NaN                      St Louis, MO
1       NaN  Montreal, PQ / Chesterville, ON
2       NaN  Montreal, PQ / Chesterville, ON
3     135.0  Montreal, PQ / Chesterville, ON
4       NaN  Montreal, PQ / Chesterville, ON
…       …                                …
1305    NaN                              NaN
1306  304.0                              NaN
1307    NaN                              NaN
1308    NaN                              NaN
1309    NaN                              NaN

[1310 rows x 14 columns]>
```

[7]:
```
'''
RESULT
The above programs were created and executed successfully.
'''
```

[7]:
```
'\nRESULT\nThe above programs were created and executed successfully.\n'
```