Module 4: Data Info in Python

contributed by

Mr. Anik Paul
Intern, R Team, FOSSEE, IIT Bombay
Mathematics Department, IIT Bombay

Ms. Usha Viswanathan
Sr. Project Manager
FOSSEE, IIT Bombay

17 December 2022

data.info() command in Python

This module explains the command data.info() and the different data types. For this module, the data from the previous modules will be used, i.e., the Indian Agricultural data.

- A short information of the types of the variables in the collected dataset can be fetched using the info() function.

```
Agriculture_data.info()
```

```
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 525 entries, 0 to 524
Data columns (total 15 columns):
 #   Column      Non-Null Count  Dtype
---  ------      --------------  -----
 0   State       525 non-null    object
 1   Zone        525 non-null    object
 2   GSDP        525 non-null    int64
 3   Area        525 non-null    float64
 4   Prod.       525 non-null    float64
 5   Prod./Area  525 non-null    float64
 6   NFS         525 non-null    object
 7   PFS         525 non-null    object
 8   KFS         525 non-null    object
 9   Dist.       525 non-null    object
 10  SML         525 non-null    float64
 11  SMV         525 non-null    float64
 12  SDN         525 non-null    float64
 13  SDP         525 non-null    float64
 14  SDK         525 non-null    float64
dtypes: float64(8), int64(1), object(6)
memory usage: 61.6+ KB
```

Figure 1: info() function to get the datatypes of the columns in the data

Short note of the above data types:

- **Object:** an array in python with strings as elements.
- **Int64:** integer, numerical type variable with 64-bit memory size. This type of variable can take only integer type data, and cannot take decimal data.
- **Float64:** numerical type variable with 64-bit memory size. This type of variable can take values of decimal type unlike int64.

Note that,

- Like 'int64', there are also int8, int16 and int32, defined as integer type numerical type variables with 8-bit, 16-bit, 32-bit memory size respectively.
- Like 'float64', there are float8, float16, float32 defined in the similar manner.
- int8, int16, int32, int64 etc are not in-built data types in python. They are created under the pandas library for data analysis and storage management.

Data types in Python (In-built):

- **Numeric:** numeric data type represents data which has numeric value.

1. **Integers:**
   - Represented by '*int*' class.
   - Positive or negative whole numbers(without fraction or decimal)
2. **Float:**
   - Represented by '*float*' class.
   - Real number with a floating-point representation, specified by a decimal point.
3. **Complex numbers:**
   - represented by *'complex'* class.
   - specified by (real part) + (imaginary part)j. For example: 2+3j.

- **Sequence type:** Ordered collection of similar or different data types.

1. **String:** It is represented by '*str*' class.
   - arrays of bytes representing Unicode characters.
   - Collection of one or more characters. A character is a string of length 1.
2. **List:** lists are like arrays, where the items need not to be of the same type.

3. **Tuple:** Tuples are also like arrays but the difference between list and tuple is that tuples are immutable i.e., tuples cannot be modified after it is created.

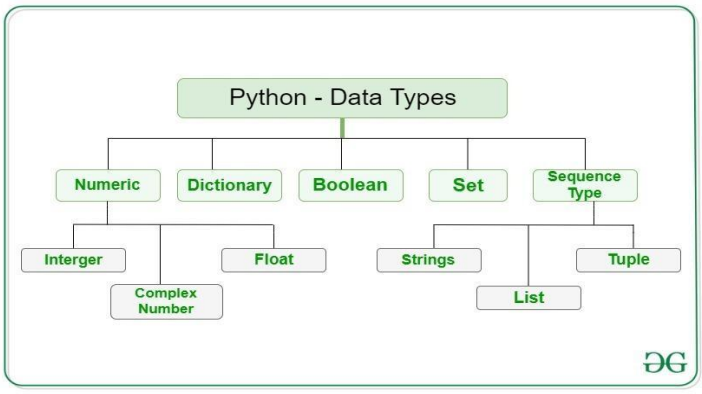

Figure 2: Data types in Python

- **Boolean:** Data types with one of the two built-in values, True or False.
- **Set:** Set is an unordered collection of data type that is iterable, mutable and has no duplicate elements.
- **Dictionary:**

  - Represented by '*dict*' class.
  - Dictionary is an unordered collection of data values,
  - used to store data like maps. It holds a '*key: value*' pair.
  - Each key-value pair is separated by a colon (:) , whereas each key is separated by a comma (,) .

Further details: For further detailed study, you can follow the link: Python Data Types - GeeksforGeeks and Python Data Types (w3schools.com) .