

## Analyzing Data Using Pandas :-

Python Pandas is used for relational or labeled data and provides various data structures for manipulating such data and time series. This library is built on top of the NumPy library. This module is generally imported as:

```
import pandas as pd
```

Here, pd is referred to as an alias to the Pandas. However, it is not necessary to import the library using the alias, it just helps in writing less amount code every time a method or property is called. Pandas generally provide two data structures for manipulating data,

They are:-

- Series
- Dataframe

### Series: -

Pandas Series is a one-dimensional labeled array capable of holding data of any type (integer, string, float, python objects, etc.).

The axis labels are collectively called indexes.

**Pandas Series is nothing but a column in an excel sheet.**

Labels need not be unique but must be a hashable type. The object supports both integer and label-based indexing and provides a host of methods for performing operations involving the index.



it can be created using the Series() function by loading the dataset from the existing storage like SQL, Database, CSV Files, Excel Files, etc., or from data structures like lists, dictionaries, etc.

## Python Pandas Creating Series

```
import pandas as pd
import numpy as np

# Creating empty series
ser = pd.Series()

print(ser)

# simple array
data = np.array(['o', 'm', 's', 'i', 'r'])

ser = pd.Series(data)
print(ser)
```

### Output:-

```
PS D:\archive> & "C:/Users/Big Data/AppData/Local/Programs/Python/Python311/python.exe" d:/archive/pandaseries.py
Series([], dtype: object)
0    o
1    m
2    s
3    i
4    r
```

### Dataframe:-

Pandas DataFrame is a two-dimensional size-mutable, potentially heterogeneous tabular data structure with labeled axes (rows and columns). A Data frame is a two-dimensional data structure, i.e., **data is aligned in a tabular fashion in rows and columns.** Pandas DataFrame consists of three principal components, the data, rows, and columns.

# Creating a dataframe using CSV files

First of install pandas: -

```
PS D:\archive> pip install pandas
Collecting pandas
  Downloading pandas-2.0.2-cp311-cp311-win_amd64.whl (10.6 MB)
----- 10.6/10.6 MB 8.7 MB/s eta 0:00:00
Collecting python-dateutil>=2.8.2
  Downloading python_dateutil-2.8.2-py2.py3-none-any.whl (247 kB)
----- 247.7/247.7 kB 7.7 MB/s eta 0:00:00
```

Csv data file `CardioGoodFitness.csv` :-

	A	B	C	D	E	F	G	H	I	J
1	Product	Age	Gender	Education	MaritalStatus	Usage	Fitness	Income	Miles	
2	TM195	18	Male		14 Single		3	4	29562	112
3	TM195	19	Male		15 Single		2	3	31836	75
4	TM195	19	Female		14 Partnered		4	3	30699	66
5	TM195	19	Male		12 Single		3	3	32973	85
6	TM195	20	Male		13 Partnered		4	2	35247	47
7	TM195	20	Female		14 Partnered		3	3	32973	66
8	TM195	21	Female		14 Partnered		3	3	35247	75
9	TM195	21	Male		13 Single		3	3	32973	85
10	TM195	21	Male		15 Single		5	4	35247	141
11	TM195	21	Female		15 Partnered		2	3	37521	85
12	TM195	22	Male		14 Single		3	3	36384	85
13	TM195	22	Female		14 Partnered		3	2	35247	66
14	TM195	22	Female		16 Single		4	3	36384	75
15	TM195	22	Female		14 Single		3	3	35247	75
16	TM195	23	Male		16 Partnered		3	1	38658	47
17	TM195	23	Male		16 Partnered		3	3	40932	75
18	TM195	23	Female		14 Single		2	3	34110	103
19	TM195	23	Male		16 Partnered		4	3	39795	94
20	TM195	23	Female		16 Single		4	3	38658	113
21	TM195	23	Female		15 Partnered		2	2	34110	38

Write code for readdata.py file:-

```
# Python program to illustrate
# creating a data frame using CSV files

# import pandas module
import pandas as pd

# creating a data frame
df = pd.read_csv("CardioGoodFitness.csv")
print(df.head())
```

Output:-

PROBLEMS OUTPUT DEBUG CONSOLE TERMINAL

Windows PowerShell

Copyright (C) 2015 Microsoft Corporation. All rights reserved.

PS D:\archive> & "C:/Users/Big Data/AppData/Local/Programs/Python/Python311/python.exe" d:/archive/readdata.py

	Product	Age	Gender	Education	MaritalStatus	Usage	Fitness	Income	Miles
0	TM195	18	Male	14	Single	3	4	29562	112
1	TM195	19	Male	15	Single	2	3	31836	75
2	TM195	19	Female	14	Partnered	4	3	30699	66
3	TM195	19	Male	12	Single	3	3	32973	85
4	TM195	20	Male	13	Partnered	4	2	35247	47

PS D:\archive>