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

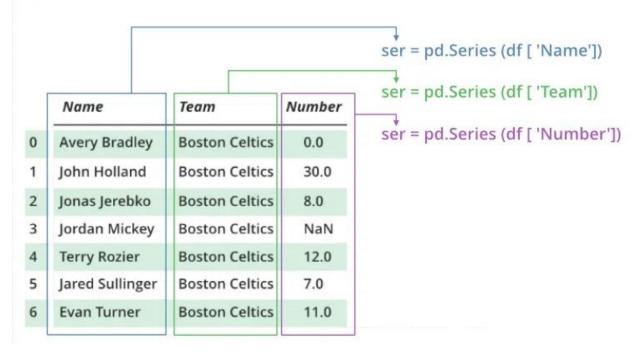
Series: -

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:-

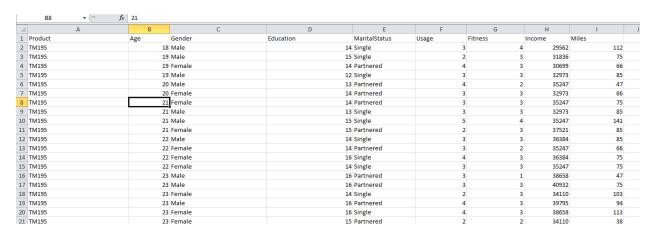
Dataframe:-

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: -

Csv data file CardioGoodFitness.csv :-



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 TM195 18 Male 14 Single 3 4 29562 112 0 TM195 TM195 19 Male 15 Single 31836 TM195 19 Female Partnered 30699 3 TM195 19 4 TM195 20 Single Male 12 32973 85 35247 Male Partnered PS D:\archive>