

PanDas:-

Pan-Da coined from Panel-Data: 2D data usually described in form a table or spread-sheet
excel, google-sheet.

	1	2	3	4	5
-					
-					
-					
-					
-					

Pandas is a open-source python library that provides fast, reliable & robust operation on

data those are usually described in row-column fashion.

Benefits:- 1) PanDas provide fast & reliable APIs (function call) to retrieve, store, create data from various sources.

Various Sources:- (1) csv file (comma separated values)

name, rollno., marks
Joy, 211, 765
Vincent, 212, 720
...

(.csv format)

Benefits:- various sources:- (2) excel file (.xls, .xlsx format)

(3) SQL table.

(4) HTML

(5) JSON

2) Pandas is built on top of Numpy. So it is fast while operating on massive datasets.

3) Pandas is a very well known open-source project & we can find a wide community that uses Pandas regularly.

The Pandas library has two types of- Data structure

(1) Series.

(2) DataFrame.

Series: (pandas.Series).

Fruits →	
0	Apple
1	Orange
2	Banana
3	Mango.

indexes.

Series is a One-dimensional labelled array capable of holding any type of data.

DataFrame (pandas.DataFrame)
Combination of few series.

columns.		
	Fruits	Price
0	Apple	130
1	Orange	60
2	Banana	40
3	Mango	100

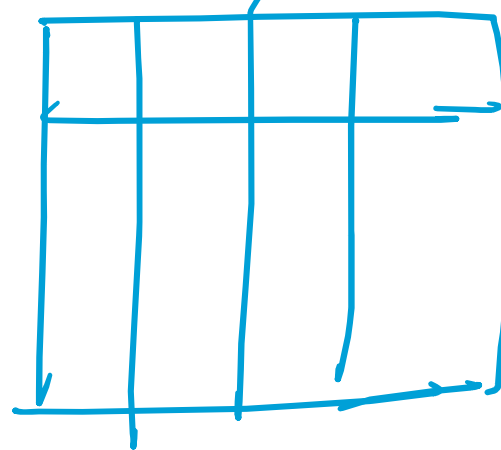
Index.

DataFrame is a two-dimensional data structure comprises of rows (index) & columns & can store values like tables or spread-sheet.

$$3245 = \frac{3 \cdot 245 \times 10^3}{10^3}$$

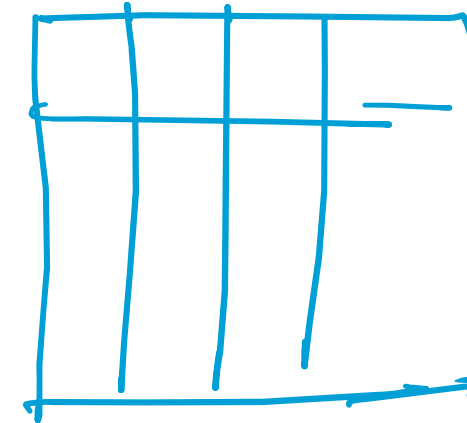
$$3 \cdot 245 e 3$$

memory-1



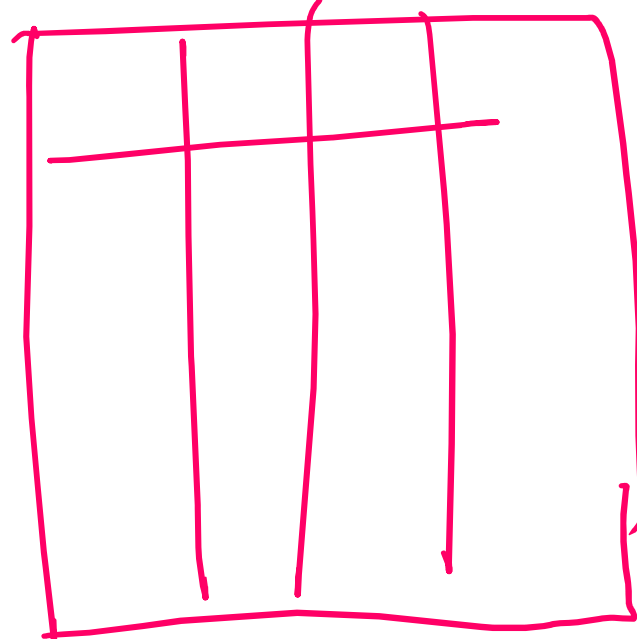
memory-2

operation



inplace = True

mem-1



mem-2

