Pandas

Why? (numpy)

- data handling
- no label in numpy
- missing values handling
- grouping of data
- Relational Data/Tabular Data

any label → Python

arr = [[2, 1, 2]
       [9, 2, 3]
       [6, 1, 1]]

Pandas → numpy

S.no.	name	roll no	class	year
1	SK	5	MCA	2
2	22	6	MS	2

Data Science

→ Collect

Warehouse → Storing

Pandas → Pre-processing

→ Summerise

→ → Visualize

| ML model

Pandas Objects

→ Series( )

* → Dataframes( )   → object

input

np·array ([_])

s1 = Student( )
Object

→ Dataframes( )

① Series

object

1, 2, 6, — — —] 1D

{

---

import (pandas) as (pd)

same power as numpy

$$a = [\ 2, 6, 6, 9, 2 \ ]$$

sr = pd. Series(a)

print (sr)

Why you take it as pd?

→ Law (Johny)

0

↓

label

	0 ⌐	2
	1	6
	2	6
	3	9
	4	2

Series
→ Pandas Object

→ 1D data

→ Table (Single Column)

* Kaggle *

Research Data

Dataframe

csv { Xcel excel

•XLs

} Prog.

{ tabular data

→ like a table

↓   ↓   ↓