BATCH : B168 DATA SCIENCE

LESSON : PANDAS

DATE : 05.08.2023

SUBJECT : INTRODUCTION

techproeducation

techproeducation

techproeducation

techproeducation

techproedu

TECHPRO EDUCATION

techproeducation.com info@techproeducation.com +1 (917) 768-7466

TECHPRO EDUCATION

Pandas Introduction

Data Science - Pandas

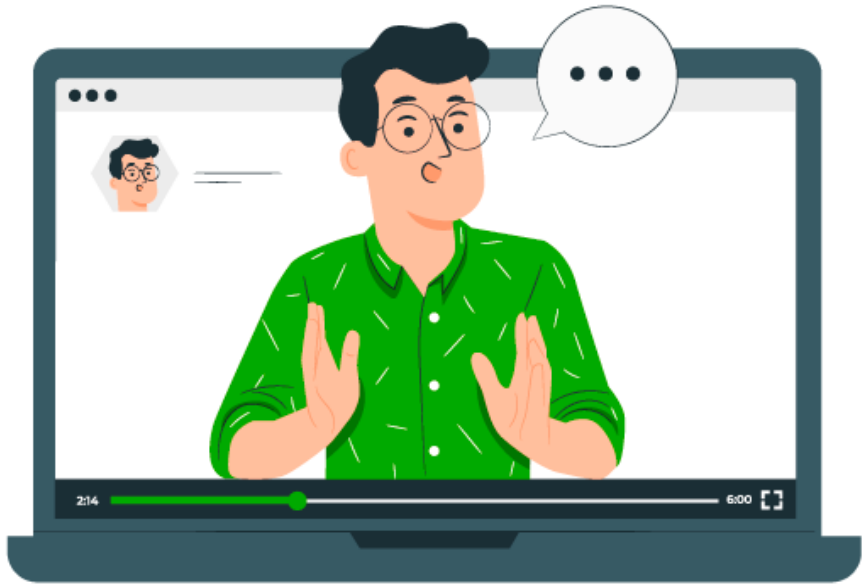
Session -1

NUMPY

TECHPRO EDUCATION

- Introduction
- Indexing, Slicing & Selection
- Groupby & Useful Operations
- Handling_with_Missing_Values
- Combining_Data_Frames
- Text_and_Time_Data

Session - 1 Content



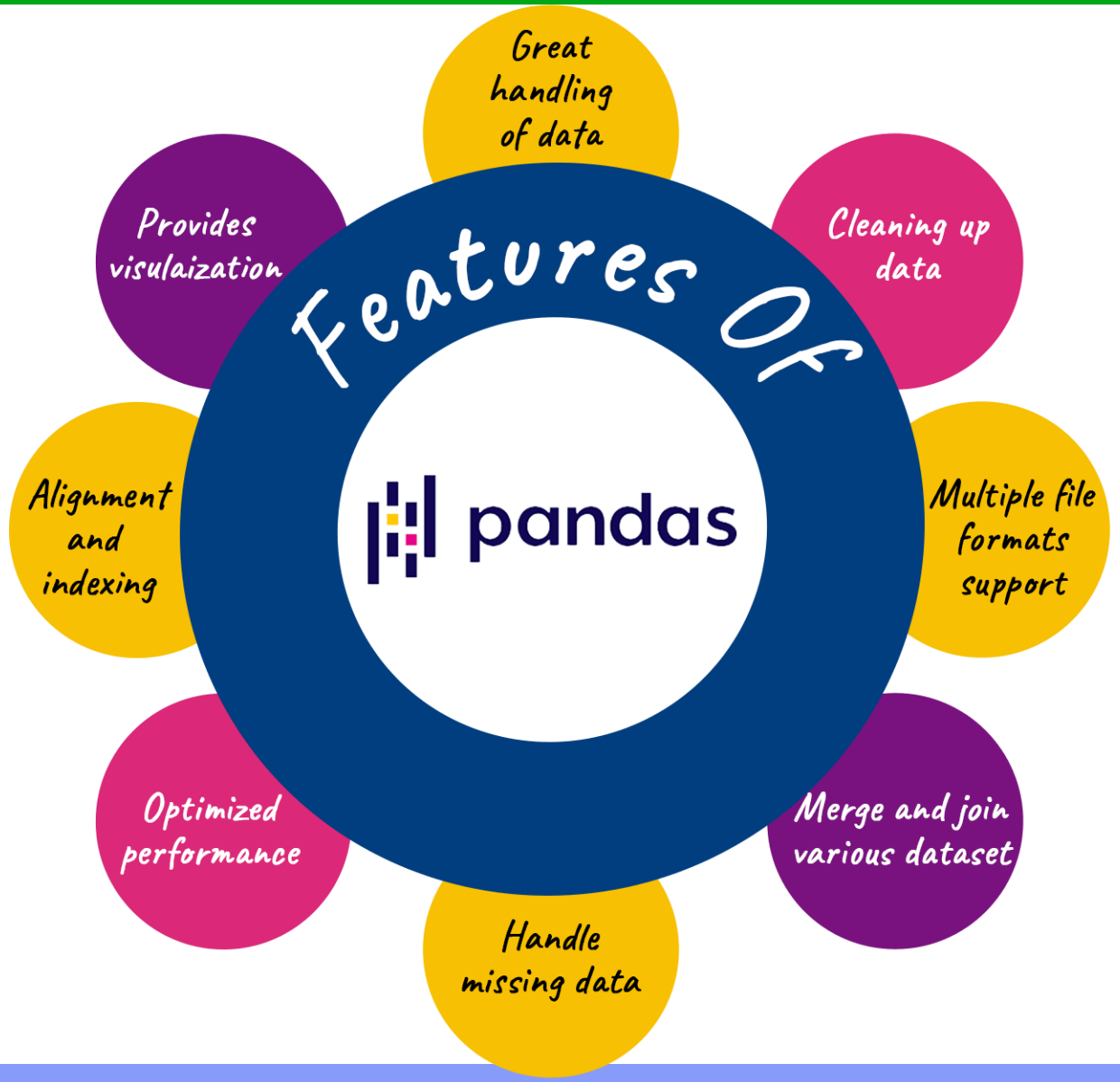
Bugün ne öğreneceğiz?

Introduction

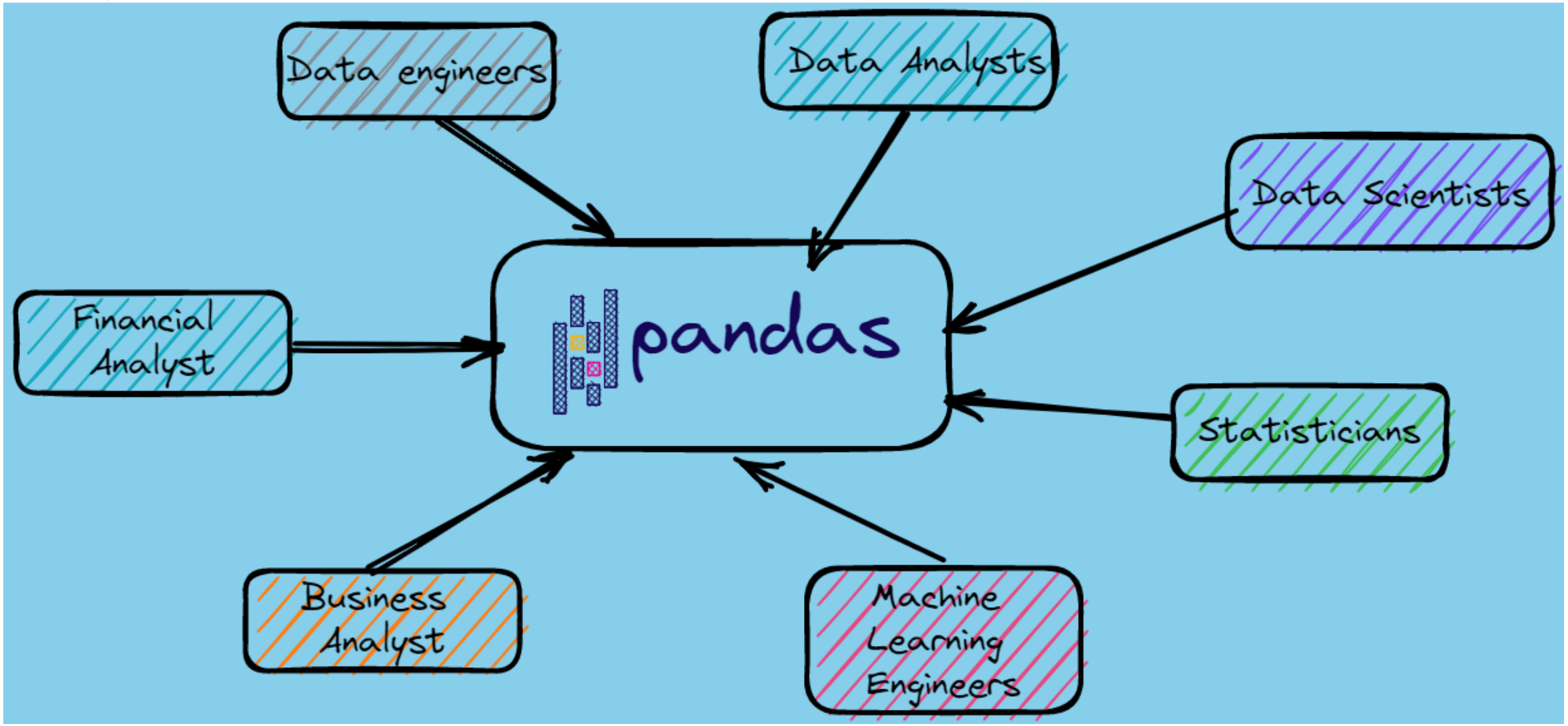
- Pandas Introduction
- Pandas Data Types
- Pandas Series
- Pandas DataFrame
- Creating Series- Notebook

Sizi bugünkü derse hazırlayacak pre-class materyalleri ile antrenman yaptınız mı?

Why Pandas


Features Of
pandas
Great handling of data
Cleaning up data
Multiple file formats support
Merge and join various dataset
Handle missing data
Optimized performance
Alignment and indexing
Provides visulaization

Why Pandas


Data engineers
Data Analysts
Data Scientists
Financial Analyst
pandas
Statisticians
Business Analyst
Machine Learning Engineers

Pandas

- Pandas, Python programlama dilinde veri manipülasyonu ve analizi için kullanılan açık kaynaklı bir kütüphanedir.

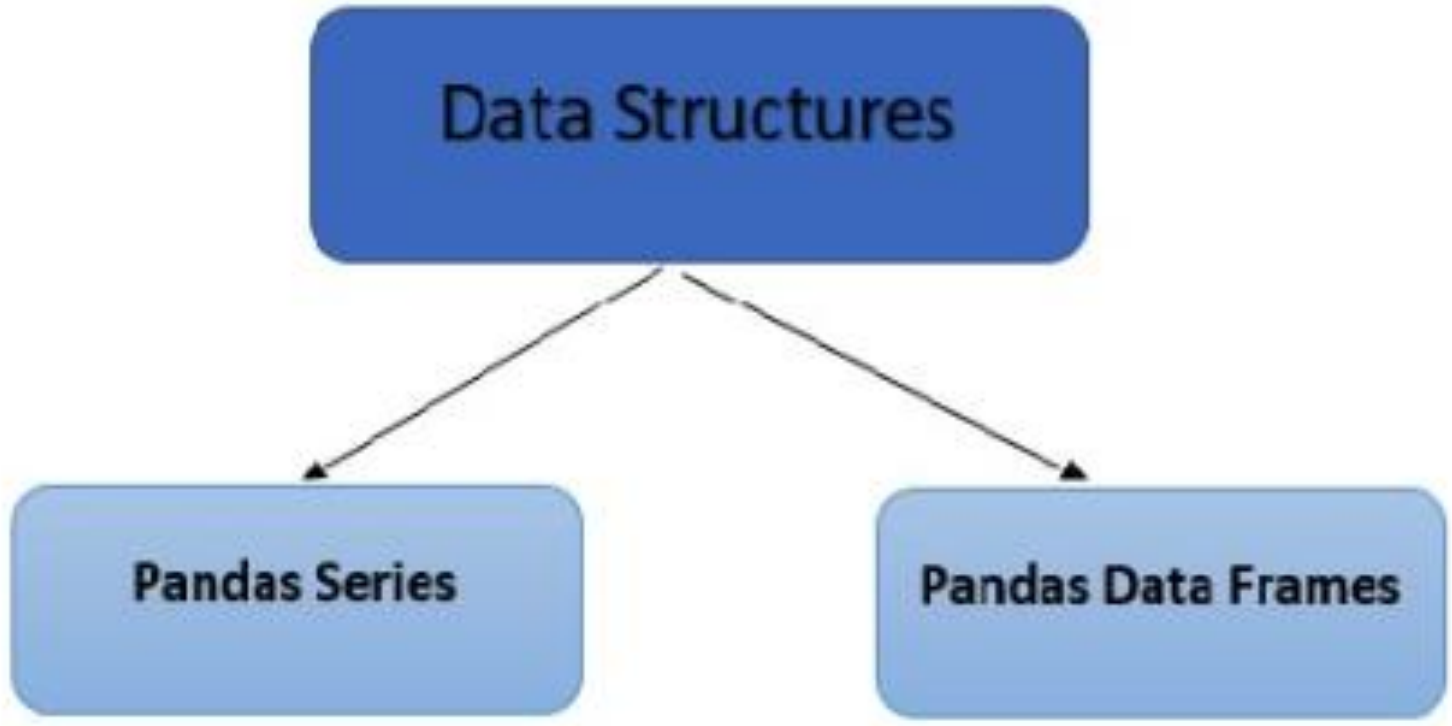
- Pandas'ın temel veri yapıları olan Series (Tek boyutlu veriler için) ve DataFrame (İki boyutlu veriler için) son derece kullanışlı ve esnek araçlardır.

Pandas

- **Veri temizleme ve ön işleme:** Eksik verilerin doldurulması, hatalı verilerin çıkarılması, verilerin formatının değiştirilmesi, veri dönüştürme işlemleri vb.
- **Veri analizi:** Verilerin analiz edilmesi ve özetlenmesi, istatistiksel analizlerin yapılması, özellik mühendisliği vb.
- **Veri görselleştirme:** Verilerin çeşitli grafiklerle görselleştirilmesi, verinin anlaşılmasını ve yorumlanmasını kolaylaştırmak için.
- **Makine öğrenmesi:** Veri setini makine öğrenmesi algoritmalarına beslemek üzere veriyi ön işleme ve hazırlama konusunda çok kullanışlıdır.

Pandas

Pandas ile veri analizi yaparken kullanacağımız temel veri yapıları Seriler ve DataFrame'lerdir.



Pandas

Pandas

Pandas = Panel Data System

- Pandas, Python programlama dili için yüksek performanslı, kullanımı kolay veri yapıları ve veri analiz araçları sağlayan açık kaynaklı bir kütüphanedir.

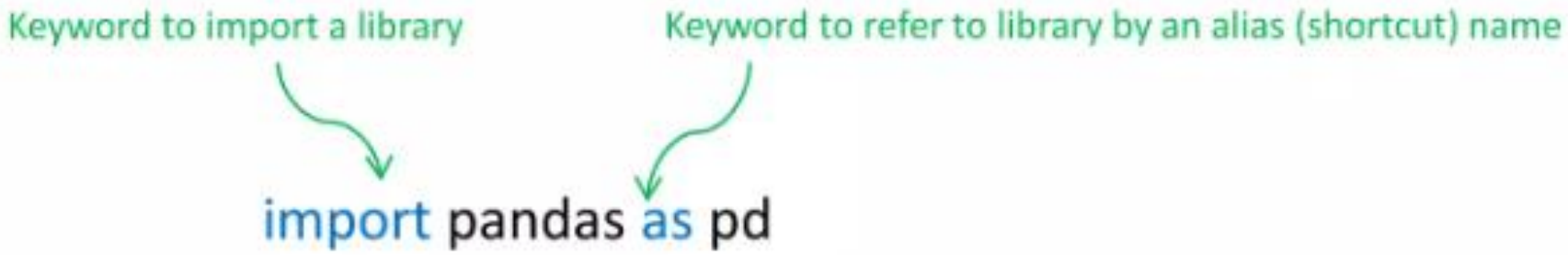
Pandas

Pandas = Panel Data System

- Pandas, Numpy'ın sütun adları ve homojen olmayan verilerle çalışamama gibi eksik kaldığı kısımlara çözümler üretir.

- Pandas ile veri analizi yaparken kullanacağımız temel veri yapıları Seriler ve DataFrame'lerdir.

Pandas



Used for:

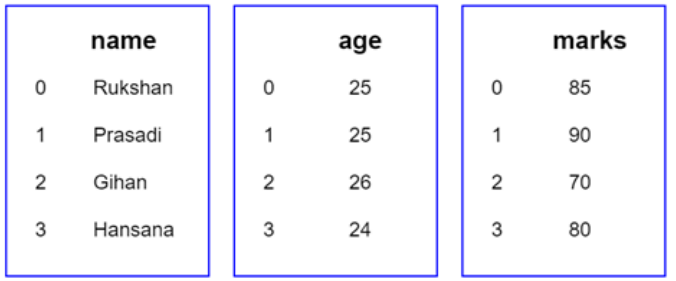
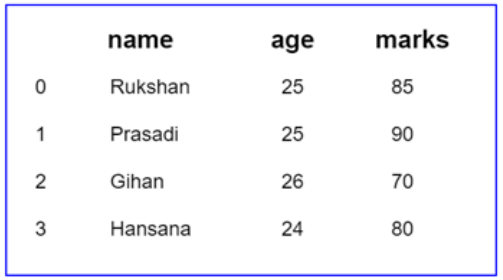
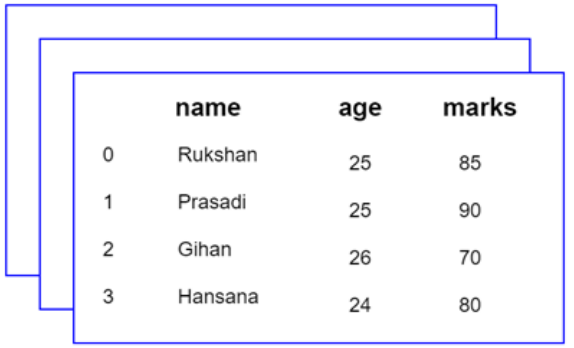
- **Data Analysis**
- **Data Manipulation**
- **Data Visualization**

Pandas

Data Structure	Dimensions	Description
Series	1	1D labeled homogeneous array, sizeimmutable.
Data Frames	2	General 2D labeled, size-mutable tabular structure with potentially heterogeneously typed columns.
Panel	3	General 3D labeled, size-mutable array.

Pandas

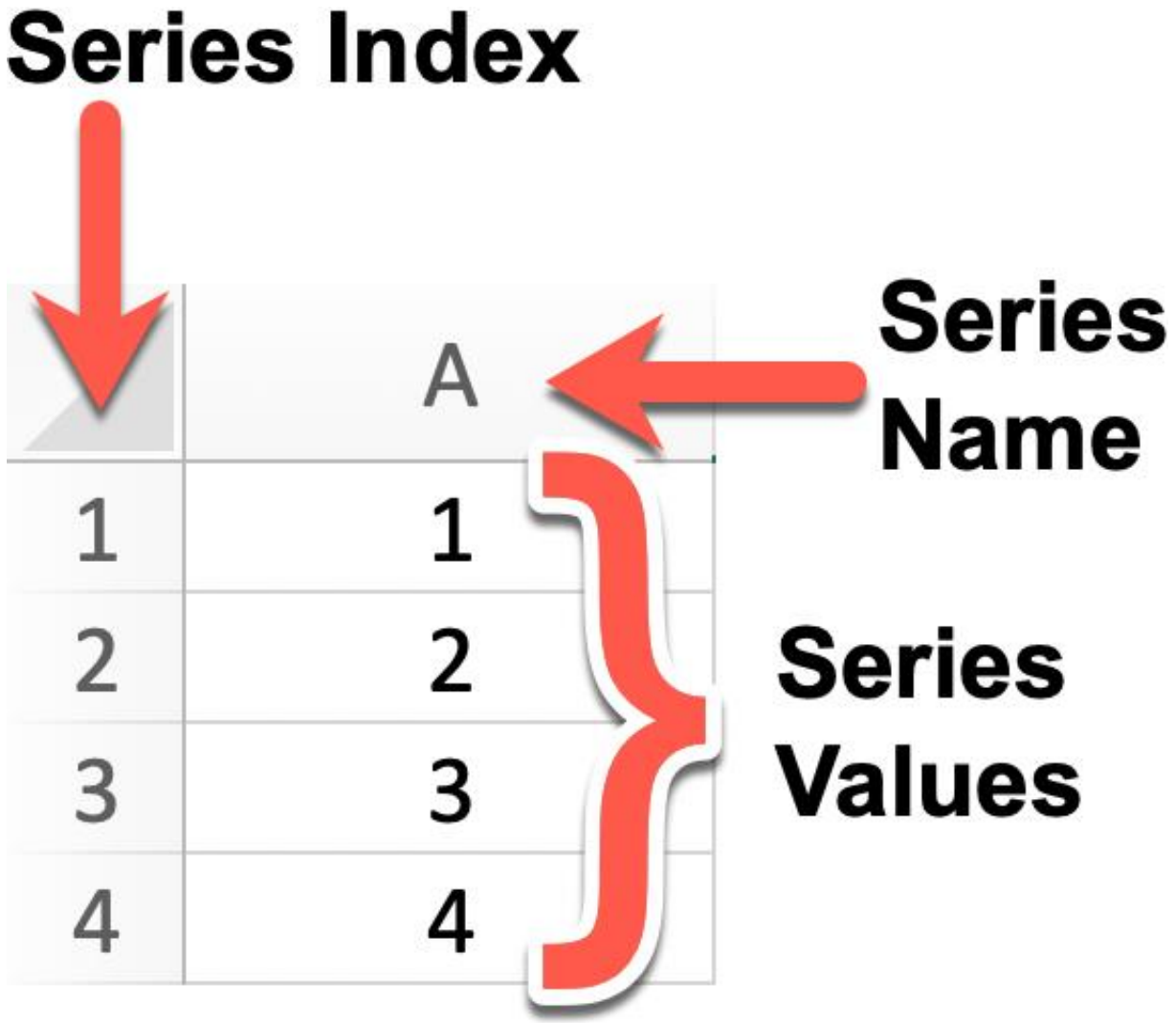
TECHPRO EDUCATION

Data Structure	Dimensionality	Format	View
Series	1D	Column	
DataFrame	2D	Single Sheet	
Panel	3D	Multiple Sheets	

Pandas Series

- Pandas Serisi, NumPy dizi nesnelerinin üzerine inşa edilmiştir ve çok benzerler.
- Herhangi bir veri tipinde veri tutabilen tek boyutlu etiketli bir dizidir.
- Etiket değerlerine ise indeks denir.
- Verinin kendisi sayılar, dizeler veya başka Python objelerinden oluşabilir.
- Serileri oluşturmak için ise listeler, sıralı diziler ya da sözlükler kullanılabilir.

Pandas



Pandas

```
import pandas as pd
series1 = pd.Series([10,20,30])
print(series1)
```

giving an alias name to pandas

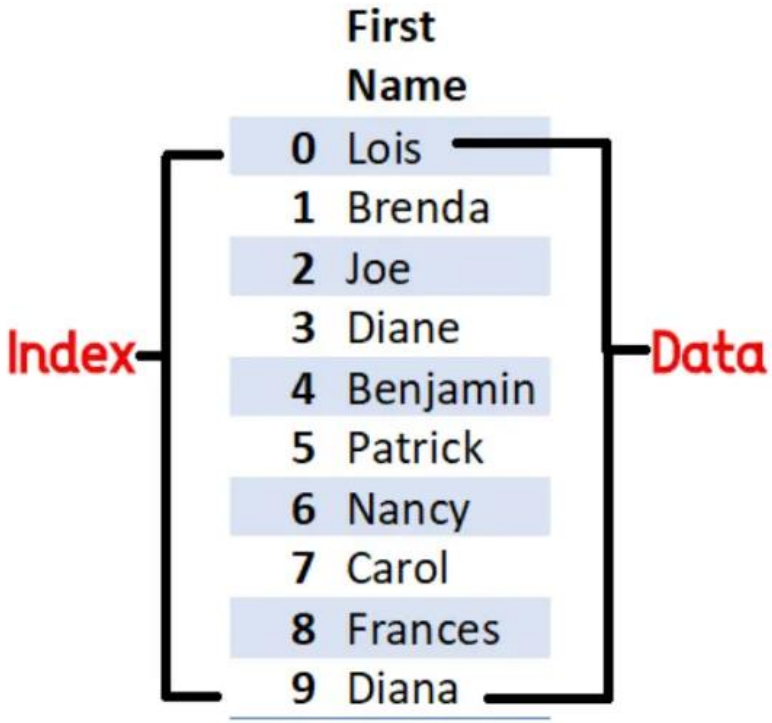
Series object

List

Output

```
0    10
1    20
2    30
dtype: int64
```

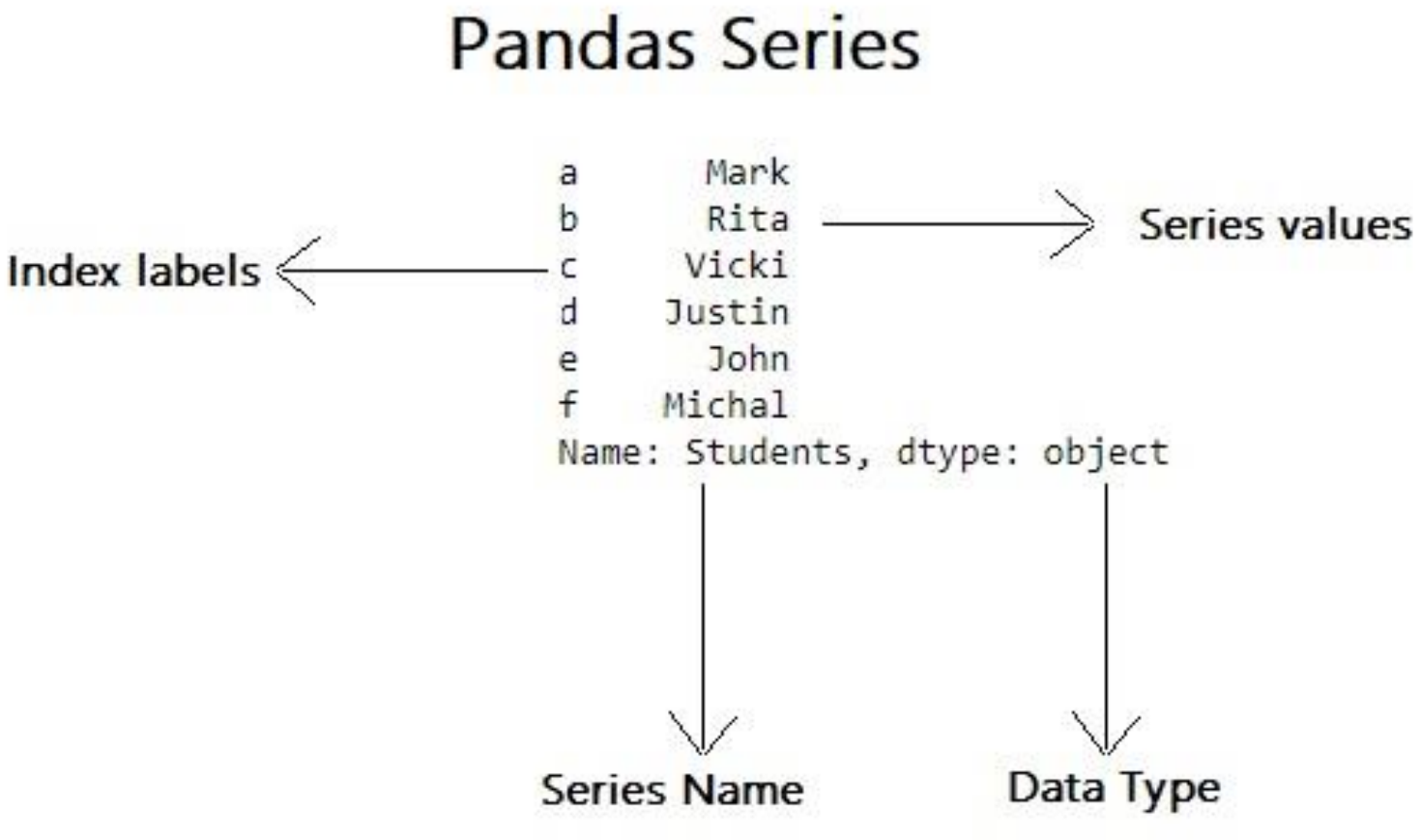
index



PANDAS

SERIES

Pandas


Pandas Series
a Mark
b Rita
c Vicki
d Justin
e John
f Michal
Name: Students, dtype: object
Index labels
Series values
Series Name
Data Type

Pandas

TECHPRO EDUCATION

Series 1

INDEX	DATA
0	A
1	B
2	C
3	D
4	E
5	F

Series 2

INDEX	DATA
A	1
B	2
C	3
D	4
E	5
F	6

Series 3

INDEX	DATA
0	[1, 2]
1	A
2	1
3	(4, 5)
4	{"a": 1}
5	6

Series 4

INDEX	DATA
Jan-18	11
Feb-18	23
Mar-18	43
Apr-18	21
May-18	17
Jun-18	6

Creating Series

```
import pandas as pd
s1 = pd.Series([1, 2, 3, 4])
```

```
s2 = pd.Series([1, 2, 3, 4], index=['A', 'B', 'C', 'D'])
```

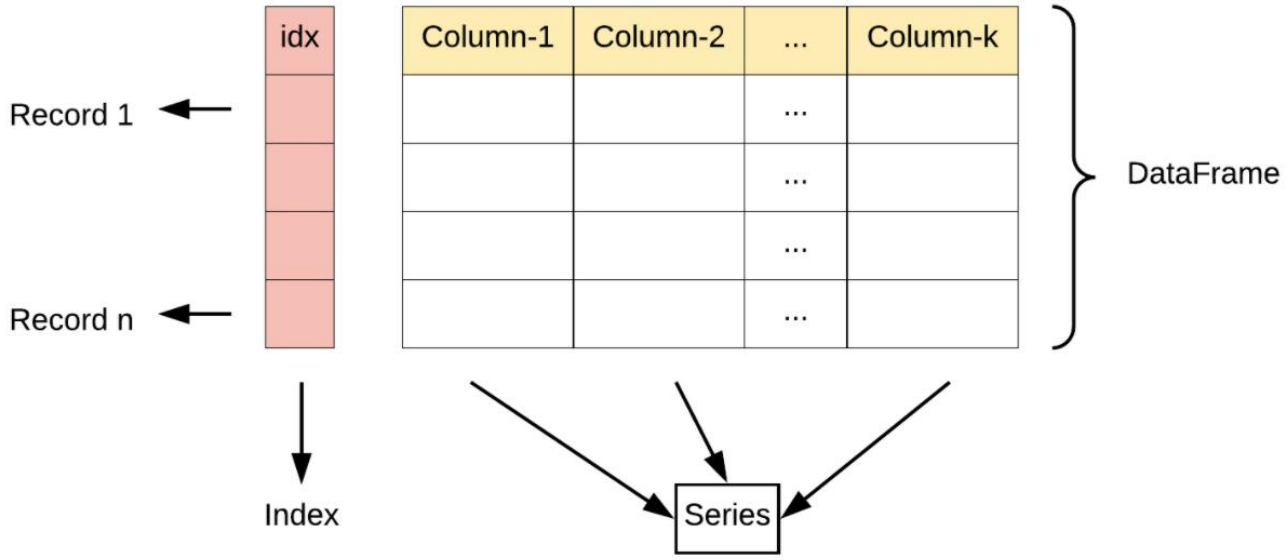
0	1
1	2
2	3
3	4

A	1
B	2
C	3
D	4

Pandas DataFrame

- Pandas Dataframe, satırları ve sütunları olan iki boyutlu etiketli veri yapısıdır.
- Pandas DataFrame'deki her sütun bir Pandas Serisidir.
- Verinin kendisi sayılar, dizeler veya başka Python objelerinden oluşabilir.
- Serileri oluşturmak için ise listeler, sıralı diziler ya da sözlükler kullanılabilir.

Pandas



Series 1

	Mango
0	4
1	5
2	6
3	3
4	1

+

Series 2

	Apple
0	5
1	4
2	3
3	0
4	2

+

Series 3

	Banana
0	2
1	3
2	5
3	2
4	7

=

DataFrame

	Mango	Apple	Banana
0	4	5	2
1	5	4	3
2	6	3	5
3	3	0	2
4	1	2	7

Pandas

TECHPRO EDUCATION

Series

	apples
0	3
1	2
2	0
3	1

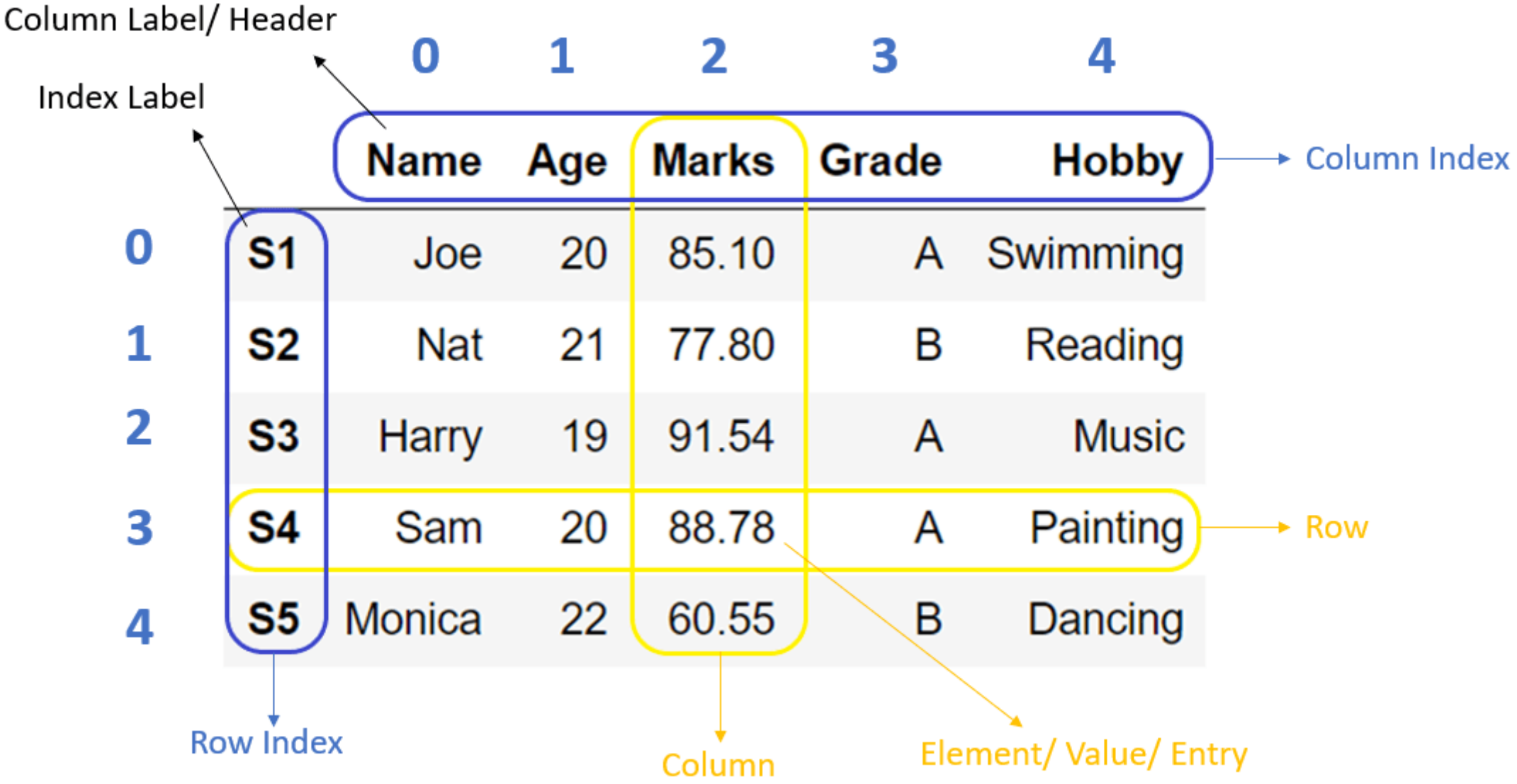
Series

	oranges
0	0
1	3
2	7
3	2

DataFrame

	apples	oranges
0	3	0
1	2	3
2	0	7
3	1	2

Pandas


Column Label/ Header
Index Label
0 1 2 3 4
Name Age Marks Grade Hobby
Column Index
0 S1 Joe 20 85.10 A Swimming
1 S2 Nat 21 77.80 B Reading
2 S3 Harry 19 91.54 A Music
3 S4 Sam 20 88.78 A Painting
4 S5 Monica 22 60.55 B Dancing
Row
Row Index
Column
Element/ Value/ Entry

DataFrame

- DataFrame is a two-dimensional array with heterogeneous data. For example,

Name	Age	Gender	Rating
Steve	32	Male	3.45
Lia	28	Female	4.6
Vin	45	Male	3.9
Katie	38	Female	2.78

Data Type of Columns

Column	Type
Name	String
Age	Integer
Gender	String
Rating	Float

Pandas Data Structures

Series

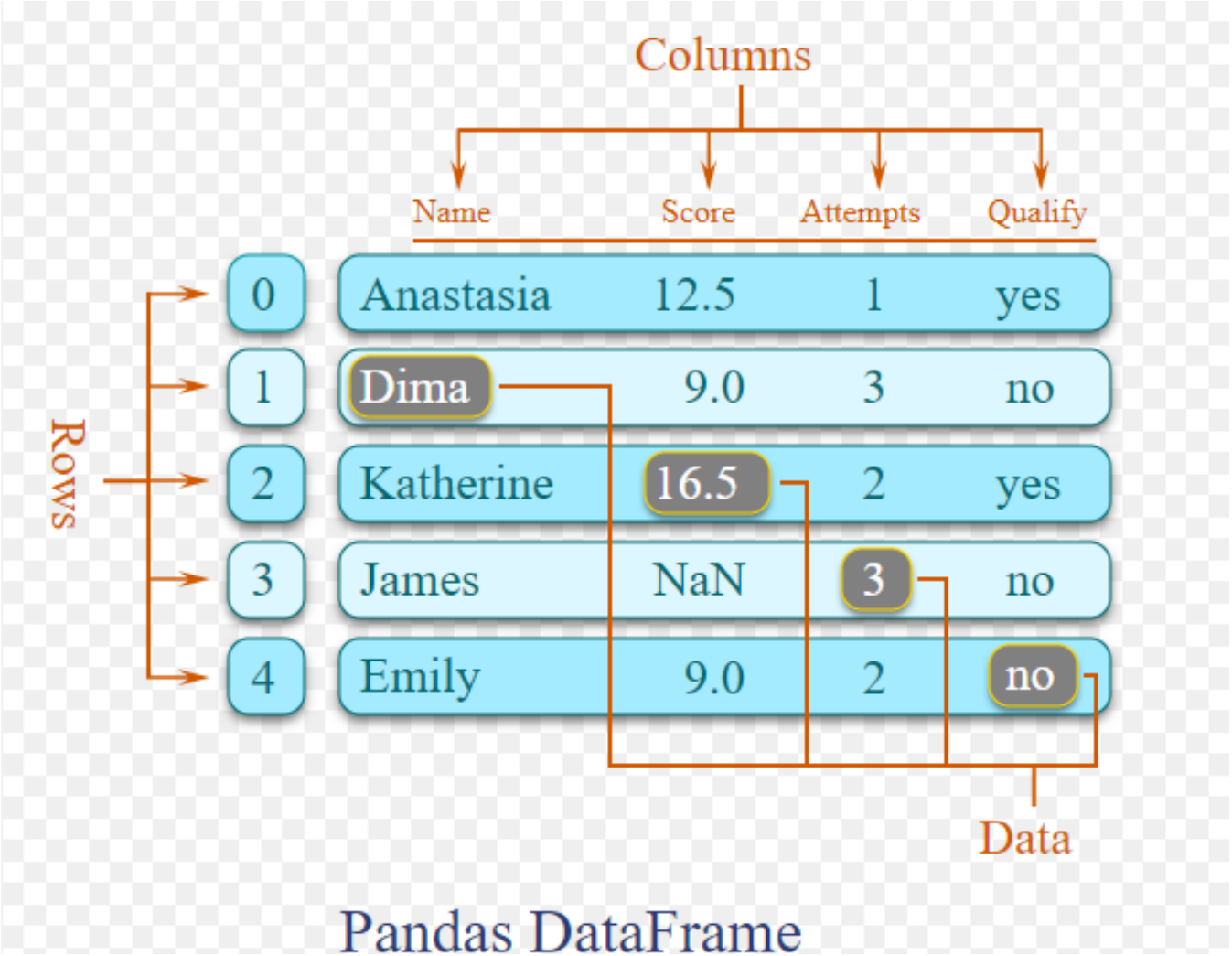
index	*values*
A	6
B	3.14
C	-4
D	0

DataFrame

index ⟵ *columns* ⟶

	foo	bar	baz
A	x	6	True
B	y	10	True
C	z	NaN	False

Pandas


Columns
Name
Score
Attempts
Qualify
Rows
0
Anastasia
12.5
1
yes
1
Dima
9.0
3
no
2
Katherine
16.5
2
yes
3
James
NaN
3
no
4
Emily
9.0
2
no
Data
Pandas DataFrame

Pandas

```
df = pd.DataFrame ( { 'month' : [ 2, 5, 8, 10 ],
 'year' : [ 2017, 2019, 2018, 2019 ],
 'sale' : [ 60, 45, 90, 36 ] } )
```

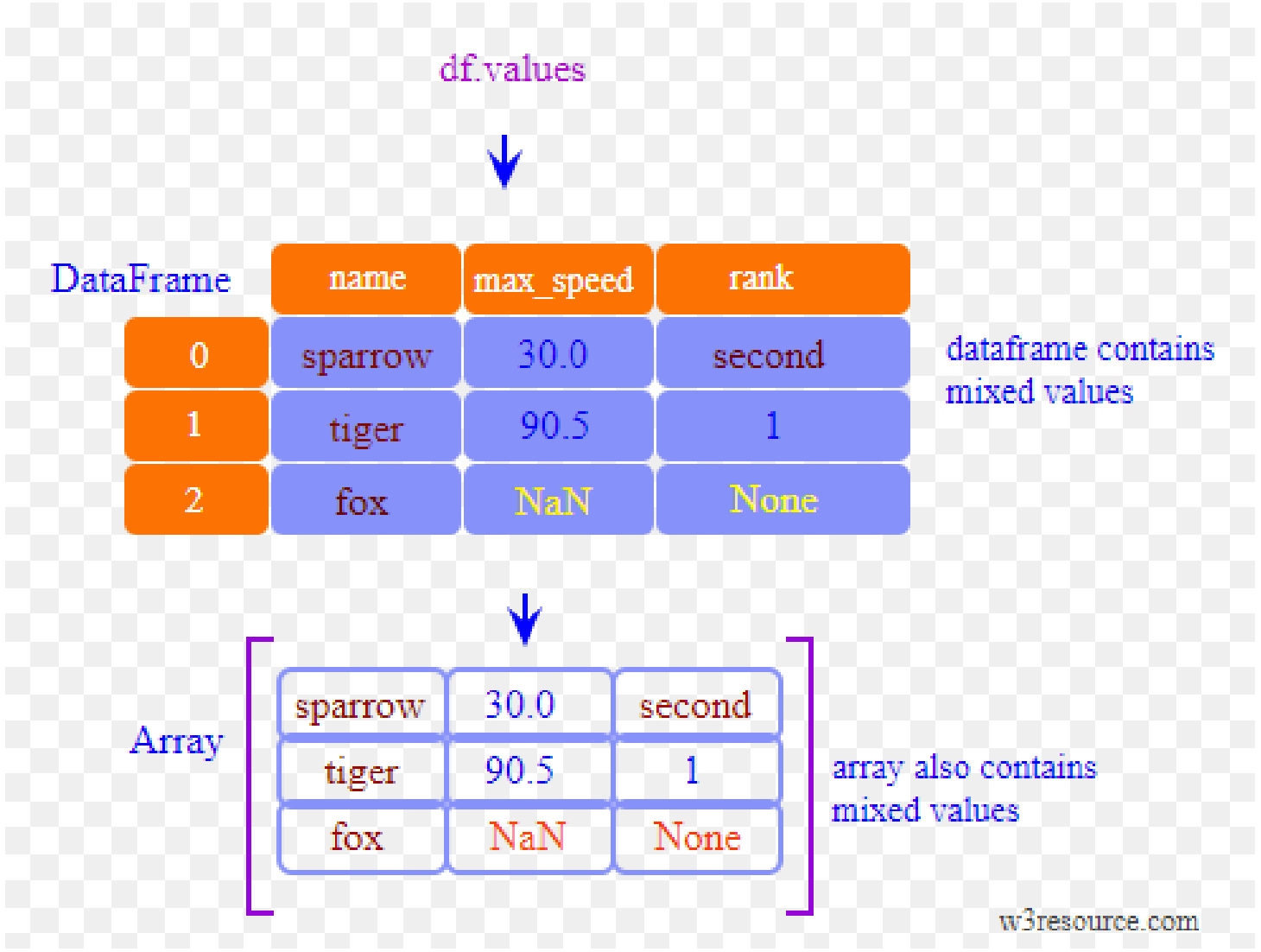
DataFrame
df

	month	year	sale
0	2	2017	60
1	5	2019	45
2	8	2018	90
3	10	2019	36

```
pd.DataFrame ( np.array ( ( [ 2, 3, 4 ], [ 5, 6, 7 ] ) ),
 index = [ 'tiger', 'lion' ],
 columns = [ 'one', 'two', 'three' ] )
```

DataFrame	one	two	three
tiger	2	3	4
lion	5	6	7

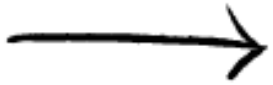
Pandas


df.values
DataFrame
name
max_speed
rank
0
sparrow
30.0
second
1
tiger
90.5
1
2
fox
NaN
None
dataframe contains mixed values
Array
sparrow
30.0
second
tiger
90.5
1
fox
NaN
None
array also contains mixed values
w3resource.com

Pandas

DATA SCIENCE
PARICHAY


```
{
    "Name": ["Jim",
"Dwight", "Angela",
"Tobi"],
    "Age": [26, 28, 27,
32],
    "Department": ["Sales",
"Sales", "Accounting",
"Human Resources"]
}
```

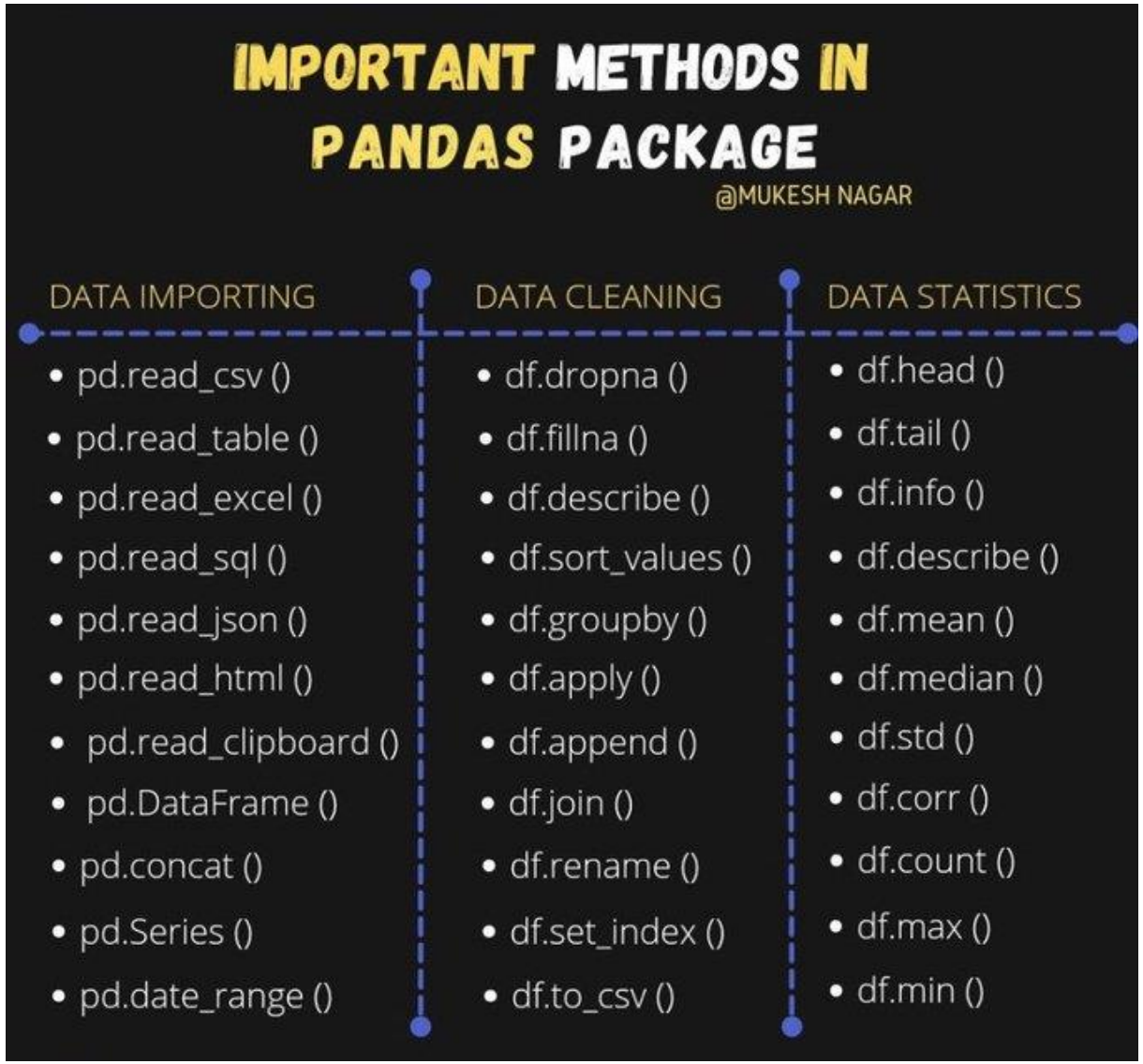


	Name	Age	Department
0	Jim	26	Sales
1	Dwight	28	Sales
2	Angela	27	Accounting
3	Tobi	32	Human Resources

DataFrame from Dictionary

Pandas Methods

TECHPRO EDUCATION




TIME TO PRACTICE

Tea break...

00:00