Pandas

1) what is pandas?

↳ pandas is a python library.

↳ which is used for data Analysis & data science.

2) why pandas?

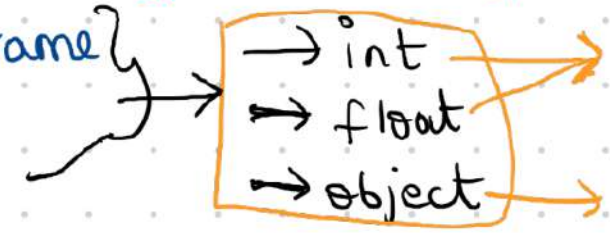
↳ pandas can handle large amount of data

↳ And also with pandas, we can do statistical Analysis of the data. (eg:- mean, median, mode)

↳ Also we can easily handle the null values in the Dataset, By removing (or) replacing the null values.

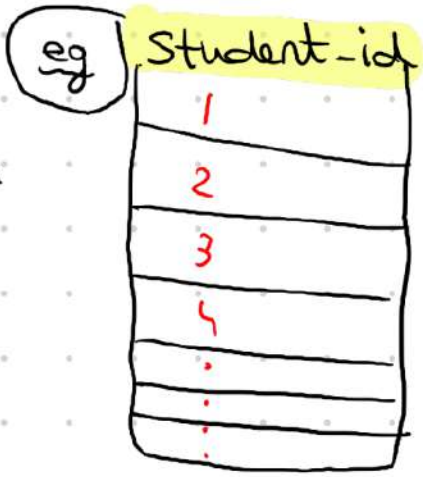
2) what are the two major data types used for Data Analysis?

↳(i) Data Frame
↳(ii) Series



4) what is Series?

↳ series, is one dimension
↳ because, it only contains one column And multiple Rows.



5) what is DataFrame?

↳ multiple series combines & forms a Data frame

↳ Data frame is 2 - Dimensional

↳ it has multiple columns & multiple rows.

eg:-

	roll_no	students_name	gender	class	section	subject	marks	grade	result
0	1	sanjay	male	10	A	Tamil	85	A	pass
1	2	bairoo	male	10	A	Tamil	95	A	pass
2	3	janani	female	10	A	Tamil	98	A	pass
3	4	kutty	male	10	A	Tamil	87	A	pass
4	5	jaanu	female	10	A	Tamil	92	A	pass
5	6	lavanya	female	10	A	Tamil	94	A	pass
6	7	kishore	male	10	A	Tamil	87	A	pass
7	8	karthik	male	10	A	Tamil	93	A	pass
8	9	sathish	male	10	A	Tamil	96	A	pass
9	10	john	male	10	A	Tamil	89	A	pass

6) Let's see the practical implementations of series ?

```
!pip install pandas
```

→ (PiP - Package installer Python) → pandas

```
Requirement already satisfied: pandas in c:\users\sanjay\anaconda3\lib\site-packages (1.5.3)
Requirement already satisfied: numpy>=1.21.0 in c:\users\sanjay\anaconda3\lib\site-packages (from pandas) (1.25.2)
Requirement already satisfied: pytz>=2020.1 in c:\users\sanjay\anaconda3\lib\site-packages (from pandas) (2022.7)
Requirement already satisfied: python-dateutil>=2.8.1 in c:\users\sanjay\anaconda3\lib\site-packages (from pandas) (2.8.2)
Requirement already satisfied: six>=1.5 in c:\users\sanjay\anaconda3\lib\site-packages (from python-dateutil>=2.8.1->pandas)
(1.16.0)
```

→ code to install pandas library

↳ In order to use pandas Library classes like Dataframe & series we need to install the pandas library first.

7) how to import the pandas library?

```
import pandas as pd
```

→ library name

→ (alias name)

↳ once you set the alias name for that Library name, instead of using that Library name again & again, we can use that alias name.

Eg:- pandas.Series() → without alias

Pd.Series() → with alias

8). How to create empty Series ?

```
In [2]: #creating empty series
        empty_Series = pd.Series()
        empty_Series
```

Series() → is used to create empty series

```
Out[2]: Series([], dtype: float64)
```

9). how to create a Series using list?

```
In [3]: #passing list as data to series without setting index
        #so it'll take default index from (0 to n) specific to number of datas
        names_list = ["sanjay","janani","bairoo","kutty"]
        Series1 = pd.Series(names_list)
        Series1

Out[3]: 0    sanjay
        1    janani
        2    bairoo
        3     kutty
        dtype: object
```

→ list

inorder to create a series using List, pass the List inside The Series() object

→ if index is not specified inside the Series(), the pandas itself assign the index numbers from 0 to (number of values in the list - 1)

10) How to set specific index to the series?

```
In [4]: #passing list as data to series with setting index
        index = ["a","b","c","d"]
        Series2 = pd.Series(data=names_list, index=index)
        Series2

Out[4]: a    sanjay
        b    janani
        c    bairoo
        d     kutty
        dtype: object
```

```
names_list = ["sanjay","janani","bairoo","kutty"]
```

we can pass the desired index while creating a series.

11) How to change the index of the series after the series is created?

```
In [4]: #passing List as data to series with setting index
        index = ["a","b","c","d"]
        Series2 = pd.Series(data=names_list, index=index)
        Series2

Out[4]: a    sanjay
        b    janani
        c    bairoo
        d     kutty
        dtype: object

In [5]: #changing the existing index with new index
        new_index = ["i","ii","iii","iv"]
        Series2.index = new_index
        Series2

Out[5]: i      sanjay
        ii     janani
        iii    bairoo
        iv      kutty
        dtype: object
```

→ first we created a series and setted the index during the time of series creation.

→ now you like to change the index of the existing series How?

```
Series2.index = new_index
```

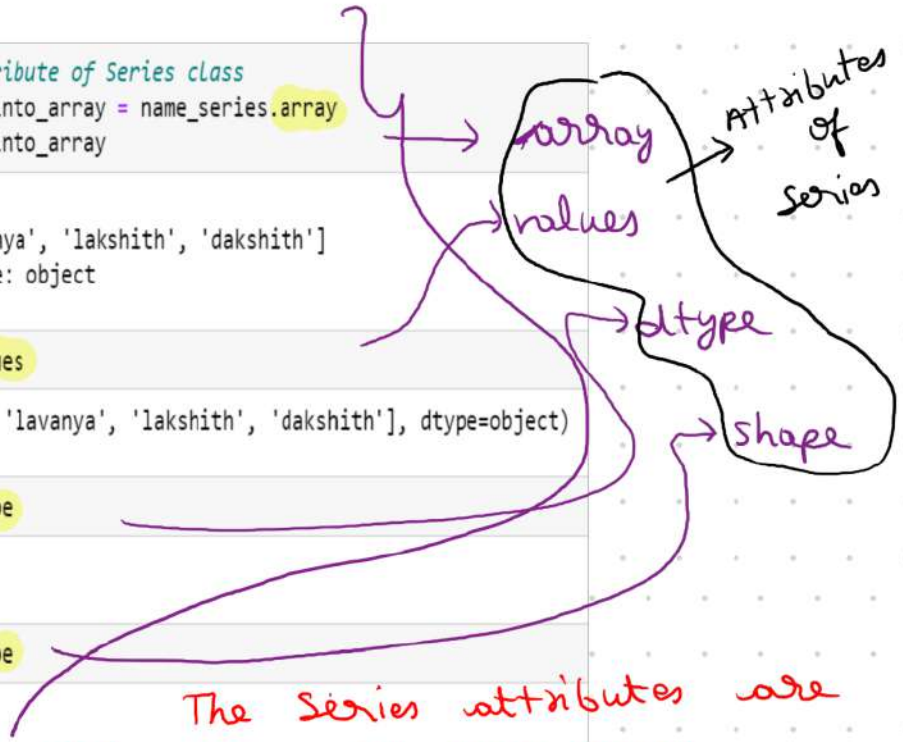
↳ with the help of index attribute called along with the series name, we can assign new index to that series

12) what are the attributes of Series ?

```
# passing list into a series object
name_list = ["jaanu","lavanya","lakshith","dakshith"]
index_list = ["i","ii","iii","iv"]
name_series = pd.Series(data=name_list, index=index_list)
print(name_series)
```

```
i         jaanu
ii      lavanya
iii    lakshith
iv     dakshith
dtype: object
```

```
In [19]: # array --> attribute of Series class
         convert_series_into_array = name_series.array
         convert_series_into_array

Out[19]: <PandasArray>
         ['jaanu', 'lavanya', 'lakshith', 'dakshith']
         Length: 4, dtype: object

In [16]: name_series.values

Out[16]: array(['jaanu', 'lavanya', 'lakshith', 'dakshith'], dtype=object)

In [17]: name_series.dtype

Out[17]: dtype('O')

In [18]: name_series.shape

Out[18]: (4,)
```



The Series attributes are used along with series name