**Pandas**

What is pandas , why to use pandas ?

pandas is used to manipulate the data.

# Ex. Read data from different sources, extract information,

filter rows, filter columns, group data, replace missing

values, replace values etc etc.

Import the library -

# import libraries

import numpy as nm

import pandas as pd

## # 1-D : Series()

## # 2-D : DataFrame()

## # 3-D : Panel

## # 4-D : Panel4 D

**Mean:** The mean is also known as the average, and it is calculated by adding up all the values in a data set and dividing by the total number of values.

**Median:** The median is the middle value of a data set, which separates the highest and lowest values equally. It is calculated by arranging the data set in order from lowest to highest and finding the value in the exact middle.

**Mode:** The mode is the value that appears most frequently in a data set.

**Sum:** Sum of all the values.

**loc** -- is label-based, which means that you have to specify rows and columns based on their row and column labels.

**iloc** -- is integer position-based, so you have to specify rows and columns by their integer position values (0-based integer position).

## What is a Series?

A Pandas Series is like a column in a table. It is a one-dimensional array holding data of any type.

a = [10, 20, 30, 40, 50] print(a, type(a))

arr = pd.Series(a) print(a, type(arr))

Indexing -

# In pandas default index of Series starts from 0.

# In Series we can change index values.

a = [10, 20, 30, 40, 50]

arr = pd.Series(a) arr1 = pd.Series(a,index = ['a', 'b', 'c', 'd', 'e'] ) print(arr1)

arr2 = pd.Series(a, index = [0,2,4,6,8]) print(arr2)

Accessing the elements -

a = [10, 20, 30, 40, 50] arr = pd.Series(a)

# To access the element arr[1] arr[2]

Slicing -

Slicing is same as we use in numpy.