

## **Pandas Overview:**

Pandas is an open source tool, high performance tool built on Numpy package. It's an high level data manipulation tool used for data cleansing, data wrangling and data analysis purposes. The name Pandas is derived from the word Panel Data – an Econometrics term also called Multidimensional data.

Pandas is used in solving different kind of data analysis including:

- ✓ data extraction and loading,
- ✓ data manipulation [cleaning and transforming],
- ✓ handling missing values,
- ✓ detecting outliers,
- ✓ data modelling and analyzing,
- ✓ indexing
- ✓ Group by and aggregation etc.

It is built on the Numpy package which means it is fast and its key data structure is called the DataFrame. DataFrames allow you to store and manipulate tabular data in rows of observations and columns of variables.

## **Key features of Pandas:**

- Fast [as its built on Numpy package]
- Loading data from different file formats
- Handling missing data
- Detecting outliers
- Reshaping and Pivoting
- Label based indexing and slicing,
- Subsetting large data sets
- Binnig data
- Creating new variables
- Group by for aggregation and transformation
- Merging and joining data sets
- Times series etc.