Introduction to Pandas

1. Pandas is an open source library in Python. It provides ready to use high-performance data structures and data analysis tools.

2. Pandas module runs on top of NumPy and it is popularly used for data science and data analytics.

3. NumPy is a low-level data structure that supports multi-dimensional arrays and a wide range of mathematical array operations. Pandas has a higher-level interface. It also provides streamlined alignment of tabular data and powerful time series functionality.

4. DataFrame is the key data structure in Pandas. It allows us to store and manipulate tabular data as a 2-D data structure.

5. Pandas provides a rich feature-set on the DataFrame. For example, data alignment, data statistics, slicing, grouping, merging, concatenating data, etc.

In this section of the course we will learn how to use pandas for data analysis. You can think of pandas as an extremely powerful version of Excel, with a lot more features.

- Series
- DataFrames
- Missing Data
- Merging, Joining and Concatenating
- Data Input and Output

NumPy is the fundamental package for scientific computing in Python.

It is a Python library that provides a multidimensional array object, various derived objects (such as masked arrays and matrices), and an assortment of routines for fast operations on arrays, including mathematical, logical, shape manipulation, sorting, selecting, I/O, discrete Fourier transforms, basic linear algebra, basic statistical operations, random simulation and much more.