Pandas: DataFrame and Series

Pandas is a popular library for data analysis built on top of the Python programming language.
Pandas generally provide two data structures for manipulating data, They are:

DataFrame
Series

A **DataFrame** is a **two-dimensional data structure**, i.e., data is aligned in a tabular fashion in rows and columns.

A Pandas DataFrame will be created by loading the datasets from existing storage.
Storage can be SQL Database, CSV file, an Excel file, etc.
It can also be created from the lists, dictionary, and from a list of dictionaries.

A **Series** represents a **one-dimensional array** of indexed data. It has two main components :

1. An array of actual data.
2. An associated array of indexes or data labels.

The index is used to access individual data values. You can also get a column of a dataframe as a Series.
You can think of a Pandas series as a 1-D dataframe.

Concatination operation of two datasets:

1. Vetically
2. Horizontally

Merge	Based on columns/features	Default = 'inner'
Join	Based on indexes	Default = 'left'

Series	One dimensional array
Dataframe	Two dimensional array

For selecting row labels	Axis = 0
For selecting column names	Axis = 1

* 0 or 'index': apply function to each column.
* 1 or 'columns': apply function to each row.

Pandas and Numpy are used for data manipulation