* Pandas Vs NumPy

Numpy	Pandas
⇒ Low-level data structure	⇒ High-level data Structure (data frame)
{ Supports large number of multi dimensional arrays and matrices }	{ More streamlined handling of tabular data and such time Series functionality }
⇒ A wide range of mathematical array Operations	

* Dependencies

[Scikit-learn] [Scikit-Images]

[SciPy]   [Pandas]   [Matplotlib]
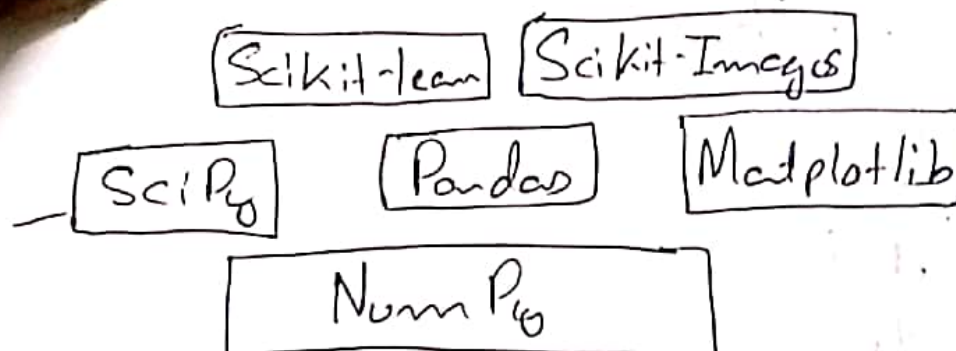
[NumPy]

* Loading hand-coded data into a dataframe

```
df = Pd. DataFrame (
    [['A', 1, 2],
     ['B', 2, 3],
     ['c', 3, 4]]   , Index = [0, 1, 2]
                    , Column = ['Letters', 'num1', 'num2'
Array                                              )
```

Pandas for Excel file (I/O)

```
import numpy as np
import pandas as pd

df = pd.read_excel ('Book1.xlsx')
```

Excel file

Workbook

|
|
|_____
          ↓   ↓
       Sheet 1   Sheet 2   • • •

data frame

* Writing a data frame to a file

```
writer = pd.ExcelWriter('new_book.xlsx')
df.to_excel(writer, 'new_sheet')
writer.save()
```

* Pandas

⇒ Pandas is an open source python library providing high-performance, easy to use data structures and data analysis tools.

⇒ Runs on top of NumPy

* **Statistical Summary of each Column**

  df.describe()

  → mean
    max
    min
    std
    25%
    50%
    75%

* **Sort records by any Column**

  df.Sort_Value ('Column Name', ascending
                                  = False)

* **Slicing by Column**

  df. ColumnName.

* **get data types, index, column, value**

  df.dtypes
  df.index
  df.columns
  df.Values.

* **Slicing by Row**

  df [2:4]
  ↘
    # grows 2&3