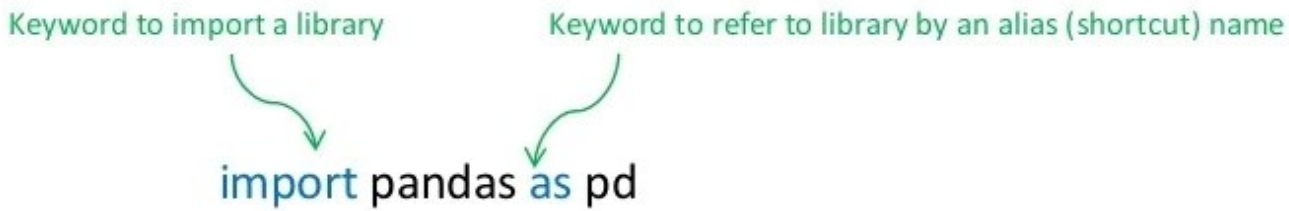
Introduction to Pandas and Matplotlib Library

Presented by: Dr. Muhammad Jawad Khan

Libraries - Pandas

- **A popular library for importing and managing datasets in Python for Machine Learning is 'pandas'.**



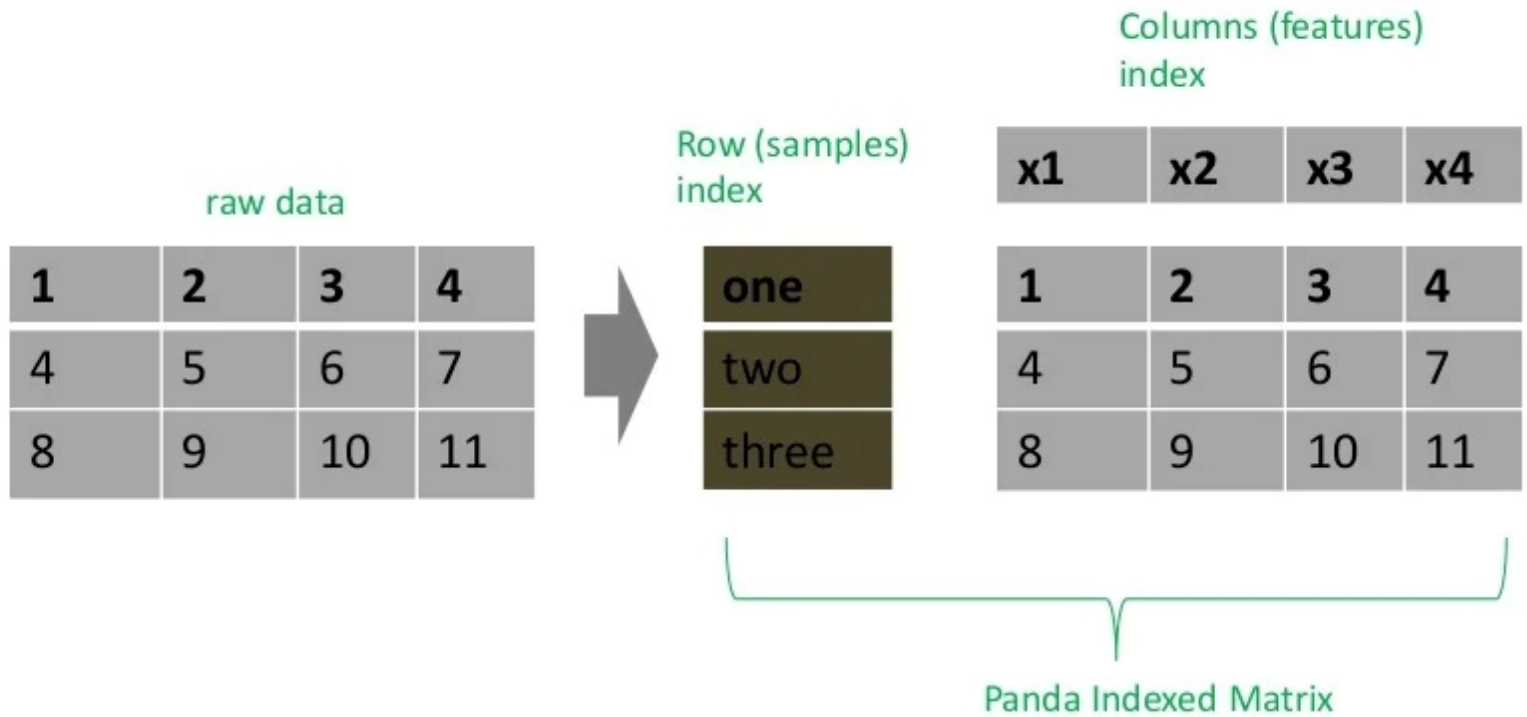
Used for:

- **Data Analysis**
- **Data Manipulation**
- **Data Visualization**

PyData.org : high-performance, easy-to-use data structures and data analysis tools for the Python programming language.

Pandas – Indexed Arrays

- **Pandas are used to build indexed arrays (1D) and matrices (2D), where columns and rows are labeled (named) and can be accessed via the labels (names).**

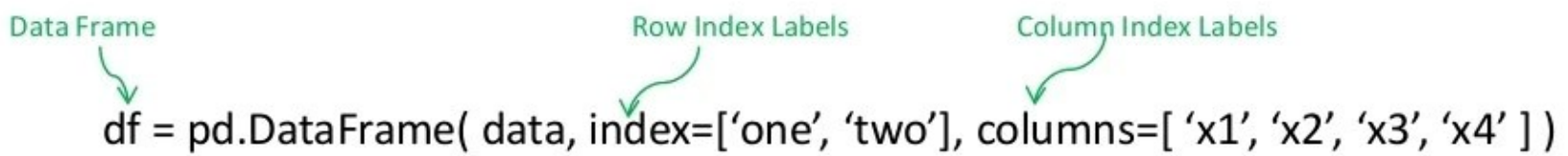


Pandas – Series and Data Frames

- **Pandas Indexed Arrays are referred to as Series (1D) and Data Frames (2D).**

- **Series is a 1D labeled (indexed) array and can hold any data type, and mix of data types.**

- **Data Frame is a 2D labeled (indexed) matrix and can hold any data type, and mix of data types.**



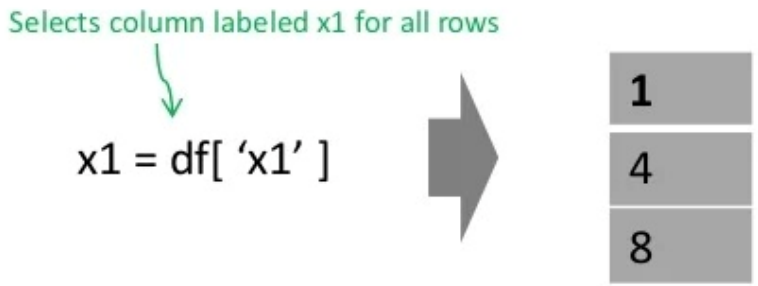
Example

- import pandas as pd
- data = [{'a': 1, 'b': 2},{'a': 5, 'b': 10, 'c': 20}]
- df = pd.DataFrame(data)
- print df

- import pandas as pd
- data = [{'a': 1, 'b': 2},{'a': 5, 'b': 10, 'c': 20}]
- df = pd.DataFrame(data, index=['first', 'second'])
- print df

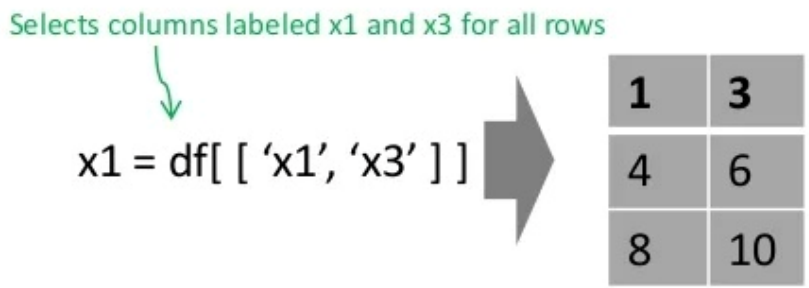
Pandas – Selecting

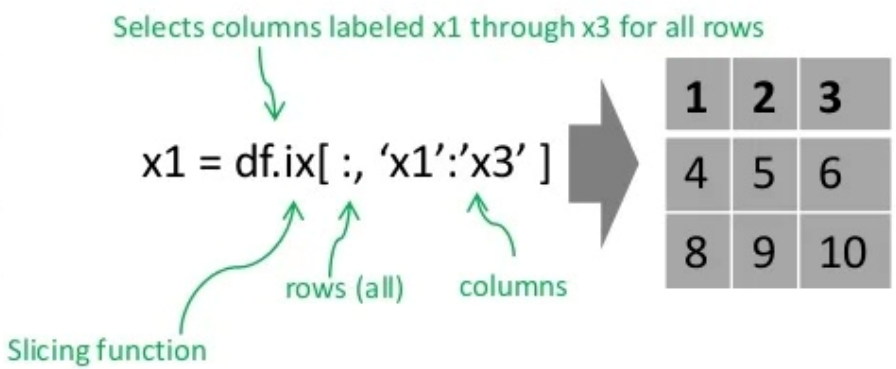
- **Selecting One Column**



- **Selecting Multiple Columns**

Note: df['x1':'x3'] this python syntax does not work!





And many more functions: merge, concat, stack, …

Slicing in python

- import pandas as pd
  d = {'one' : pd.Series([1, 2, 3], index=['a', 'b', 'c']),
  'two' : pd.Series([1, 2, 3, 4], index=['a', 'b', 'c',
  'd'])}
- df = pd.DataFrame(d)
- print df[2:4]

Addition of rows

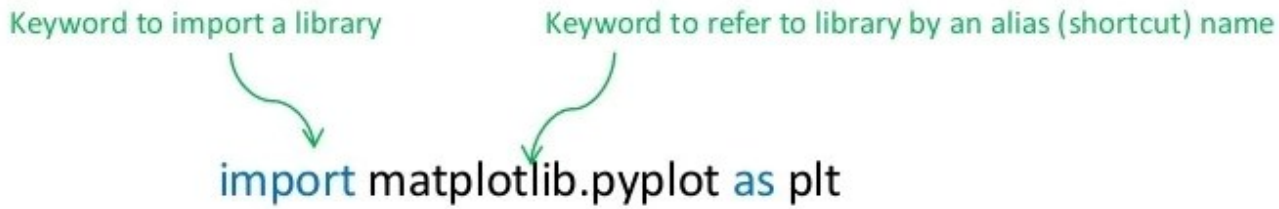
- Df2 = pd.DataFrame([[5,6], [7,8]], columns = [‘a’, ‘b’])
- Df = df.append(df2)
- Print df

Deletion of rows

- Df2 = pd.DataFrame([[5,6], [7,8]], columns = [‘a’, ‘b’])
- Df = df.drop(0)
- Print df

Libraries - Matplotlib

- **A popular library for plotting and visualizing data in Python**



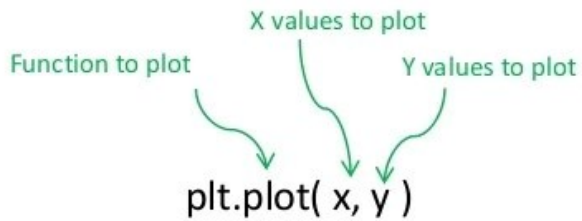
Used for:

- **Plots**
- **Histograms**
- **Bar Charts**
- **Scatter Plots**
- **etc**

matplotlib.org: Matplotlib is a Python 2D plotting library which produces publication quality figures in a variety of hardcopy formats and interactive environments across platforms.

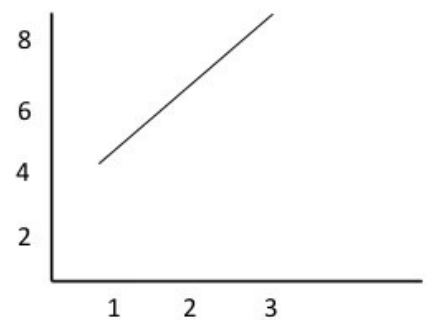
Matplotlib - Plot

- **The function plot plots a 2D graph.**



- **Example:**

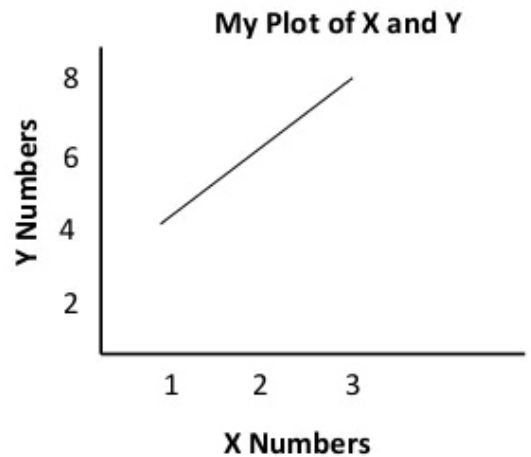
```
plt.plot( [ 1, 2, 3 ], [ 4, 6, 8 ] ) # Draws plot in the background
plt.show()                            # Displays the plot
```



Matplotlib – Plot Labels

- **Add Labels for X and Y Axis and Plot Title (caption)**

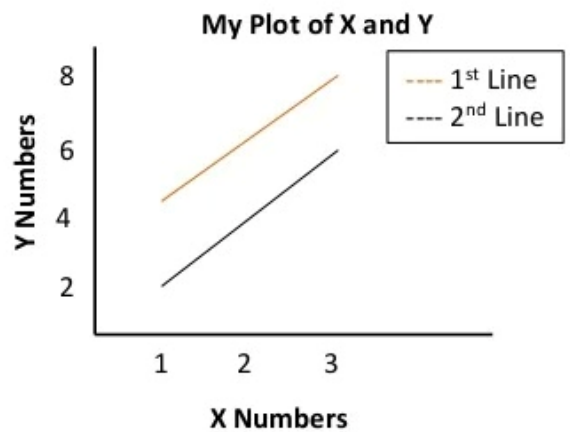
```
plt.plot( [ 1, 2, 3 ], [ 4, 6, 8 ] )
plt.xlabel( “X Numbers” )          # Label on the X-axis
plt.ylabel( “Y Numbers” )          # Label on the Y-axis
plt.title( “My Plot of X and Y”)   # Title for the Plot
plt.show()
```



Matplotlib – Multiple Plots and Legend

- **You can add multiple plots in a Graph**

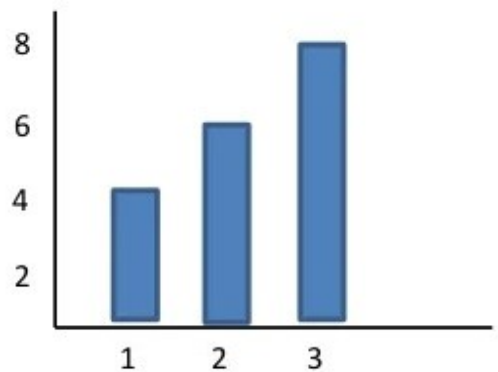
```
plt.plot( [ 1, 2, 3 ], [ 4, 6, 8 ], label=‘ 1st Line’ )      # Plot for 1st Line
plt.plot( [ 1, 2, 3 ], [ 2, 4, 6 ], label=‘2nd Line’ )      # Plot for 2nd Line
plt.xlabel( “X Numbers” )
plt.ylabel( “Y Numbers” )
plt.title( “My Plot of X and Y”)
plt.legend()                                                # Show Legend for the plots
plt.show()
```



Matplotlib – Bar Chart

- **The function bar plots a bar graph.**

```
plt.plot( [ 1, 2, 3 ], [ 4, 6, 8 ] ) # Plot for 1st Line
plt.bar()                             # Draw a bar chart
plt.show()
```



And many more functions: hist, scatter, …

Reading CSV file in Pandas as a DataFrame and Plotting in Matplotlib

```
#importing Libraries
import pandas as pd
import matplotlib.pyplot as plt

#Path to File
df = pd.read_csv("C:/Users/jkhan.smme/Desktop/Pandas Exercises/data.csv")

#Displaying CSV Data to Console
print(df.to_string())

#Plotting DataFrame
plt.plot(df)
plt.show()
```

```
In [6]: runfile('C:/Users/jkhan.smme/
wdir='C:/Users/jkhan.smme/Desktop/Pan
   Col1    Col2   Col3   Col4
0     3  215.63  25.96  35.00
1     5  152.69  38.95  68.02
2     9  142.96  28.45   2.99
3     6   95.86  12.69   3.99
4     1   89.76  18.96   5.94
5     2  112.85  24.96   4.95
6     8  199.95  11.95   7.72
7     5  103.56  20.75   6.22
8     4  139.67  10.87  35.00
9     9  218.96  23.75   8.33
```