

# Python Libraries for Data Science



*MD Arshad Ahmad*  
*15 Years+ Experience in Data Science*  
*Mentored 100+ people*

Many popular Python  
toolboxes/libraries:

- NumPy

- SciPy

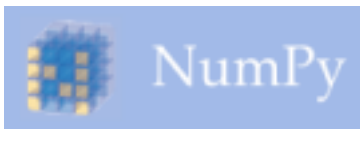
- Pandas

- SciKit-Learn

## Visualization libraries

- matplotlib • Seaborn

# Python Libraries for Data Science

*NumPy:* 

- introduces objects for multidimensional arrays and matrices, as well as functions that allow to easily perform advanced mathematical and statistical operations on those objects
- provides vectorization of mathematical operations on arrays and matrices which significantly improves the performance
- many other python libraries are built on

NumPy

Link: <http://www.numpy.org/>

4



# Python Libraries for Data Science

*SciPy:*

- collection of algorithms for linear algebra, differential equations, numerical integration, optimization, statistics and more
- part of SciPy Stack
- built on NumPy

Link: <https://www.scipy.org/scipylib/>

pandas  
 $y_{it} = \beta^t x_{it} + \mu_i + \epsilon_{it}$



# Python Libraries for Data Science

## *Pandas:*

- adds data structures and tools  
designed to work with table-like data  
(similar to Series and Data Frames in R)
- provides tools for data  
manipulation: reshaping, merging,  
sorting, slicing, aggregation etc.
- allows handling missing data

Link: <http://pandas.pydata.org/>



# Python Libraries for Data Science

## *SciKit-Learn:*

- provides machine learning algorithms: classification, regression, clustering, model validation etc.
- built on NumPy, SciPy and matplotlib

Link: <http://scikit-learn.org/>

# Python Libraries for Data Science

## *matplotlib:*

- python 2D plotting library which produces publication quality figures in a variety of hardcopy formats
- a set of functionalities similar to those of MATLAB
- line plots, scatter plots, barcharts, histograms, pie charts etc.
- relatively low-level; some effort



needed to create advanced visualization

Link: <https://matplotlib.org/>

8

# Python Libraries for Data Science

*Seaborn:*

- based on matplotlib
- provides high level interface for

drawing attractive statistical

graphics ▪ Similar (in style) to the

popular ggplot2 library in R

**Link:** <https://seaborn.pydata.org/>