

**Ex. No. 1****STUDY OF PYTHON DATA SCIENCE ENVIRONMENT****AIM:**

To study the Python Data Science Environment (NumPy, SciPy, Pandas, Matplotlib).

**PROBLEM DEFINITION:**

Study the features of Python, packages required for data science operations and their installation procedure required for Data Science programming.

**a) PYTHON DATA SCIENCE ENVIRONMENT**

Data Science is a branch of computer science that deals with how to store, use and analyze data for deriving information from it. Analyzing the data involves examining it in ways that reveal the relationships, patterns, trends, etc. that can be found within it. The applications of data science range from Internet search to recommendation systems to customer services and Stock market analysis. The data science application development pipeline has the following elements: Obtain the data, wrangle the data, explore the data, model the data and generate the report. Each element requires skills and expertise in several domains such as statistics, machine learning, and programming. Data Science projects require a knowledge of the following software:

**PYTHON:** Python is a high-level, interpreted, interactive and object-oriented scripting language that provides very high-level dynamic data types and supports dynamic type checking. It is most suited for developing data science projects.

**NUMPY:** NumPy provides n-dimensional array object and several mathematical functions which can be used in numeric computations.

**SCIPY:** SciPy is a collection of scientific computing functions and provides advanced linear algebra routines, mathematical function optimization, signal processing, special mathematical functions, and statistical distributions.

**PANDAS:** Pandas is used for data analysis and can take multi-dimensional arrays as input and produce charts/graphs. Pandas can also take a table with columns of different datatypes and may input data from various data files and database like SQL, Excel, CSV.

**MATPLOTLIB:** Matplotlib is scientific plotting library used for data visualization by plotting line charts, bar graphs, scatter plots.

**b) INSTALLATION OF PYTHON AND DATA SCIENCE PACKAGES**

The following documentation includes setting up the environment and executing programming exercises targeted for users using Windows 10 with Python 3.7 or later version. Steps should work on most machines running Windows 7 or 8 as well.

Sections that are indicated as optional are marked with **[Optional]**. Though optional, students are strongly encouraged to try out these sections.

We use the default python package management system - pip to install packages through one may prefer to install using conda.

## Setting up Environment:

### Python:

1. To install Python 3 on Windows, navigate to <https://www.python.org/downloads/> on your web browser, download and install the desired version.
2. For example to install Python 3.7.9:
  1. Navigate to <https://www.python.org/downloads/>
  2. Scroll down to “Looking for a specific release?” section and click on Python 3.7.9 as shown below:

Looking for a specific release?

Python releases by version number:

Release version	Release date	Click for more
<a href="#">Python 3.8.7</a>	Dec 21, 2020	<a href="#">Download</a> <a href="#">Release Notes</a>
<a href="#">Python 3.9.1</a>	Dec 7, 2020	<a href="#">Download</a> <a href="#">Release Notes</a>
<a href="#">Python 3.8.4</a>	Oct 3, 2020	<a href="#">Download</a> <a href="#">Release Notes</a>
<a href="#">Python 3.8.3</a>	Sept. 24, 2020	<a href="#">Download</a> <a href="#">Release Notes</a>
<a href="#">Python 3.8.10</a>	Sept. 5, 2020	<a href="#">Download</a> <a href="#">Release Notes</a>
<a href="#">Python 3.7.9</a>	Aug. 17, 2020	<a href="#">Download</a> <a href="#">Release Notes</a>
<a href="#">Python 3.8.12</a>	Aug. 17, 2020	<a href="#">Download</a> <a href="#">Release Notes</a>
<a href="#">Python 3.8.6</a>	Aug. 16, 2020	<a href="#">Download</a> <a href="#">Release Notes</a>

- c. Scroll down to “Files” section and click on “Windows x86-64 executable installer” (Indicated [A]) if running a 32 bit machine or “Windows x86 executable installer” (indicated [B]) if running a 64 bit machine. If not sure if your machine is 32 or 64 bit, we recommend installing the 32 bit version.

Files						
Version	Operating System	Description	MD5 Sum	File Size	GPG	
Gzipped source tarball	Source release		bcd9f22cf531efcf06ca62b35219bdc...	2327710	<a href="#">SIG</a>	
tar compressed source tarball	Source release		399d3e021b4d97c141d95421da1743b...	13388636	<a href="#">SIG</a>	
macOS 64-bit installer	Mac OS X	for OS X 10.9 and later	4b544fbfa0c3c5cf8fb67dede23edcb79e...	25005253	<a href="#">SIG</a>	
Windows help file	Windows		1094c809c08ad1ad263ca57ccb3b927...	8186795	<a href="#">SIG</a>	
Windows x86-64 embeddable zip file	Windows	for AMD64/EM64T/x64	60f7740b3030b22999d5d14883a43...	7502379	<a href="#">SIG</a>	
<a href="#">Windows x86-64 executable installer</a>	Windows	for AMD64/EM64T/x64	7083fe0513c3c3+ea655211df9ade27...	16940592	<a href="#">SIG</a>	[A]
Windows x86-64 web-based installer	Windows	for AMD64/EM64T/x64	da0b17ae84d857fb3cf3eb24927f3125...	134894	<a href="#">SIG</a>	
Windows x86 embeddable zip file	Windows		97c65580479d51bf4483066eadfc1e...	6659999	<a href="#">SIG</a>	[B]
Windows x86 executable installer	Windows		1e6d31c98c68722541f0821b9c15d52...	25815560	<a href="#">SIG</a>	
Windows x86 web-based installer	Windows		22b687e533c49d036e035f8aa2...	131964	<a href="#">SIG</a>	

- d. Double click the downloaded exe to run the installer. Follow the prompts on the screen and install with default options.

3. To verify installation, go to Start->Command Prompt. Type in “python --version” and hit Enter key. This will display “Python 3.7.9” or similar in the next line. If you do not see this or see any other error, please revisit the above steps.
4. Advanced Windows users or users facing issues can refer to  
<https://docs.python.org/3/using/windows.html>
5. To install Python on other distributions refer to:
  - a. Macintosh OS: <https://docs.python.org/3/using/mac.html>
  - b. Unix distros: <https://docs.python.org/3/using/unix.html>

#### Additional Resource:

<https://docs.python.org/3/installing/index.html#basic-usage>

### **pip**

Python installation comes with a default package management/install system (pip - “pip installs Package”). Make sure to verify this by:

1. Start->Command Prompt.
2. Type in “pip --version” and hit Enter key.
3. This will display “pip 20.0.2 from “c:\users\DELL\appdata\local\programs\python\python37\lib\site-packages\pip (python 3.7)” or similar in the next line.

### **Virtual Environment (venv) [Optional]**

Follows steps from here to install/use virtual environment:

<https://docs.python.org/3/tutorial/venv.html#creating-virtual-environments>

### **Jupyter Notebook [Optional]**

Jupyter Notebook is a web based interactive development environment, usually preferred for quick prototyping.

To install:

1. Start->Command Prompt.
2. Type in “pip install jupyter” and hit Enter key.

To use:

1. In Command Prompt, type “jupyter notebook” and hit Enter key.
2. By default a web browser tab with jupyter notebook will open. If not, type in the following URL to open - <http://localhost:8888/tree>
3. Do not close this Command Prompt opened in Step 1.
4. Click on New -> Python 3 (right top) to open a new Notebook.
5. To close (also called as “Shut down Jupyter”), close all newly created notebook tabs and click on “Quit”.

More on Jupyter Notebooks at <https://jupyter.org/>

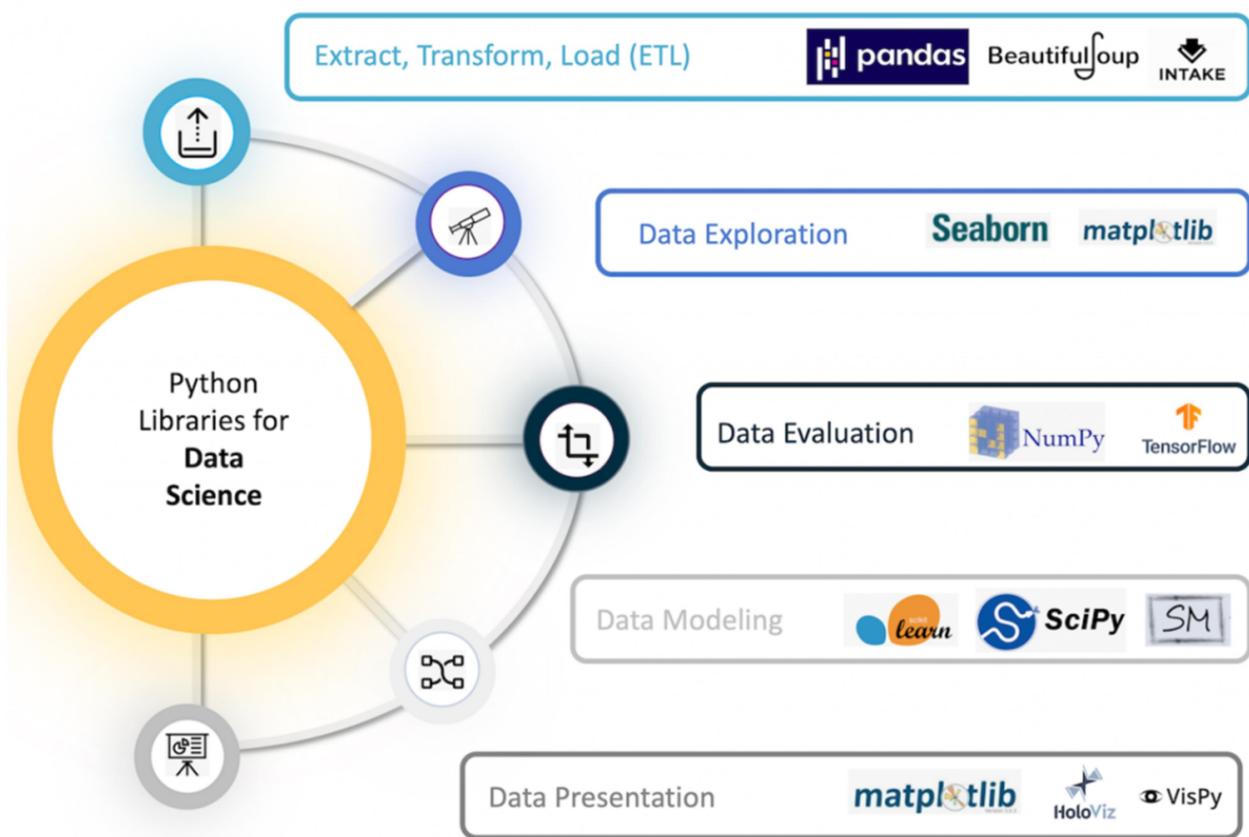
## Packages

We will install the following packages: numpy, scipy, matplotlib, pandas, scikit-learn (sklearn), bokeh.

1. Start->Command Prompt.
2. Type in “pip install numpy” and hit Enter key\*\*.  
\*\*If one encounters issue with installing/using numpy, try “pip install numpy==1.19.3”
3. Type in “pip install scipy matplotlib pandas sklearn bokeh” and hit Enter key.
4. To verify installation:
  - a. Type in “python”, hit enter.
  - b. Type in

```
import <package_name>
<package_name>.__version__
```
  - c. This will display the desired package with it's version number if properly installed as indicated below:

```
Python 3.7.5 (tags/v3.7.5:5c02a39a0b, Oct 15 2019, 00:11:34) [MSC v.1916 64 bit (AMD64)] on win32
Type "help", "copyright", "credits" or "license" for more information.
>>> import numpy
>>> numpy.__version__
'1.19.3'
>>> import scipy
>>> scipy.__version__
'1.5.4'
>>> import matplotlib
>>> matplotlib.__version__
'3.3.3'
>>> import pandas
>>> pandas.__version__
'1.2.0'
>>> import sklearn
>>> sklearn.__version__
'0.24.0'
>>> import bokeh
>>> bokeh.__version__
'2.2.3'
>>>
```



## RESULT:

A study on the Python Data Science environment was carried out to understand and install the software packages required for Data Science experiments.