**AI-Statistical Machine Learning Approaches to Liver Disease Prediction**

**Team ID: PNT2022TIMD48272**

**Faculty Mentor:** **Team Leader :** G.Lydia

D.Pradhiba **Team Member :** R.Priya

**Team Member :** U. Lavanya sri

**Team Member :** G. Nagalakshmi

**Packages Installation**

To build Machine learning models you must require the following packages.

**Numpy:**

It is an open-source numerical Python library. It contains a multidimensional array

and matrix data structures and can be used to perform mathematical operations.

**Python NumPy** is a general-purpose array processing package which provides

tools for handling the n-dimensional arrays. It provides various computing tools

provides both the flexibility of Python and the speed of well-optimized compile C

code. It’s easy to use syntax makes it highly accessible and productive for

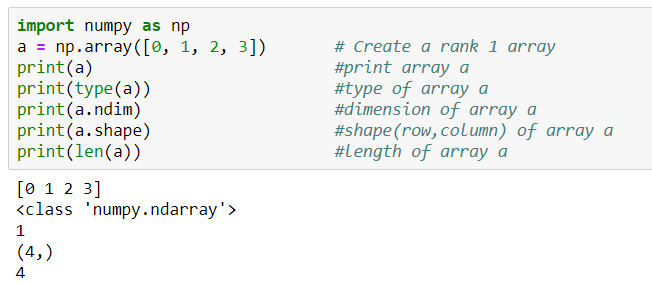
programmers from any background. This NumPy tutorial helps you learn the

fundamentals of NumPy from Basics to Advance, like operations on NumPy

array, matrices using a huge dataset of NumPy – programs and projects. Now to

use numpy in the program we need to import the module.

Generally, numpy package is defined as np of abbreviation for convenience.



In the example figure above we can observe numpy is imported first and then a l-

d numpy array a is defined. Then we can examine the type, dimension, shape, and

length of the array using mentioned commands.

**Pandas**

Panda is an open-source library built on top of numpy providing high

performance,easy-to-use data structures and data analysis tools for the Python

programming language. It allows for fast analysis and data cleaning and

preparation. It excels in performance and productivity. It can work with data from a

wide variety of sources. pandasis suited for many different kinds of data: tabular

data, time-series data, arbitrary matrix data with row and column labels, and Any

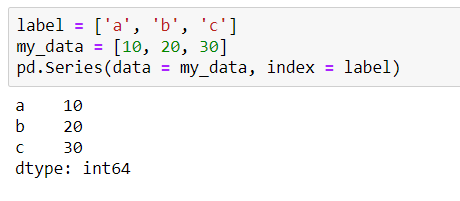
other form of observational/statistical data sets. To install pandas in your system

you can use this command pip install pandas or conda install pandas . To make

series in pandaswe need to use pd.Series(data, index)format where data are input

data and index are selected index for data. To understand it fully we can follow the

below example.



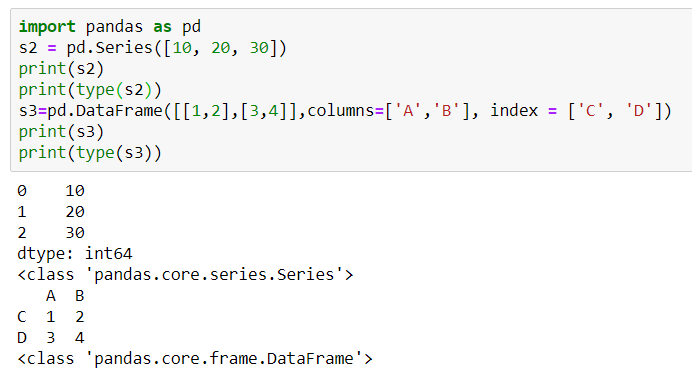
**Pandas DataFrames**create a tabular data structure with labeled axes(rows and

columns). The default format of a DataFrame would be pd.Dataframe(data, index,

column) . You need to mention the data, index and columns value to generate a

DataFrame. Data should be at least two-dimensional*,*index will be the row name

and columns values for the columns.



**Scikit-learn:**

It is a free machine learning library for Python. It features various algorithms like

support vector machine, random forests, and k-neighbours, and it also supports

Python numerical and scientific libraries like NumPy and SciPy. Scikit-learn

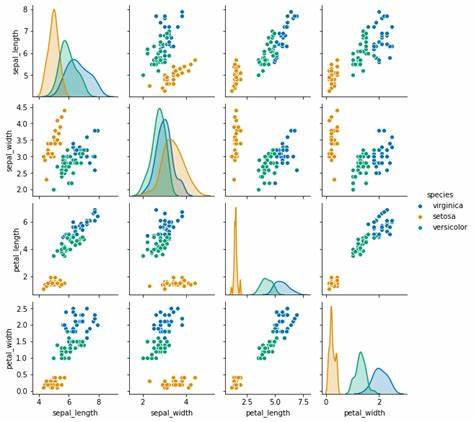
(Sklearn) is the most useful and robust library for machine learning in Python. It

provides a selection of efficient tools for machine learning and statistical modeling

including classification, regression, clustering and dimensionality reduction via a

consistence interface in Python. This library, which is largely written in Python, is

built upon NumPy, SciPy and Matplotlib.



**Learning problems fall into a few categories:**

**C**[lassification](https://en.wikipedia.org/wiki/Classification_in_machine_learning)**:**

samples belong to two or more classes and we want to learn from already labeled

data how to predict the class of unlabeled data. An example of a classification

problem would be handwritten digit recognition, in which the aim is to assign each

input vector to one of a finite number of discrete categories. Another way to think

of classification is as a discrete (as opposed to continuous) form of supervised

learning where one has a limited number of categories and for each of the n

samples provided, one is to try to label them with the correct category or class.

**R**[egression](https://en.wikipedia.org/wiki/Regression_analysis)**:**

If the desired output consists of one or more continuous variables, then the task is

called regression. An example of a regression problem would be the prediction of

the length of a salmon as a function of its age and weight. [**unsupervised learning**](https://en.wikipedia.org/wiki/Unsupervised_learning),

in which the training data consists of a set of input vectors x without any

corresponding target values. The goal in such problems may be to

discover groups of similar examples within the data, where it is called [**clustering**](https://en.wikipedia.org/wiki/Cluster_analysis),

or to determine the distribution of data within the input space, known as [**density estimation**](https://en.wikipedia.org/wiki/Density_estimation) , or to project the data from a high-dimensional space down to two or

three dimensions for the purpose of visualization.

**Matplotlib and Seaborn:**

Matplotlib is mainly deployed for basic plotting. Visualization

using Matplotlib generally consists of bars, pies, lines, scatter plots and so

on. Seaborn: Seaborn, on the other hand, provides a variety of visualization

patterns. It uses fewer syntax and has easily interesting default themes.

**Matplotlib:**

Matplotlib is an amazing visualization library in Python for 2D plots of arrays.

Matplotlib is a multi-platform data visualization library built on NumPy arrays and

designed to work with the broader SciPy stack. It was introduced by John Hunter

in the year 2002. One of the greatest benefits of visualization is that it allows us

visual access to huge amounts of data in easily digestible visuals. Matplotlib

consists of several plots like line, bar, scatter, histogram etc. As per above

definition, Matplotlib is used for visualizing the data. (Huge or small)

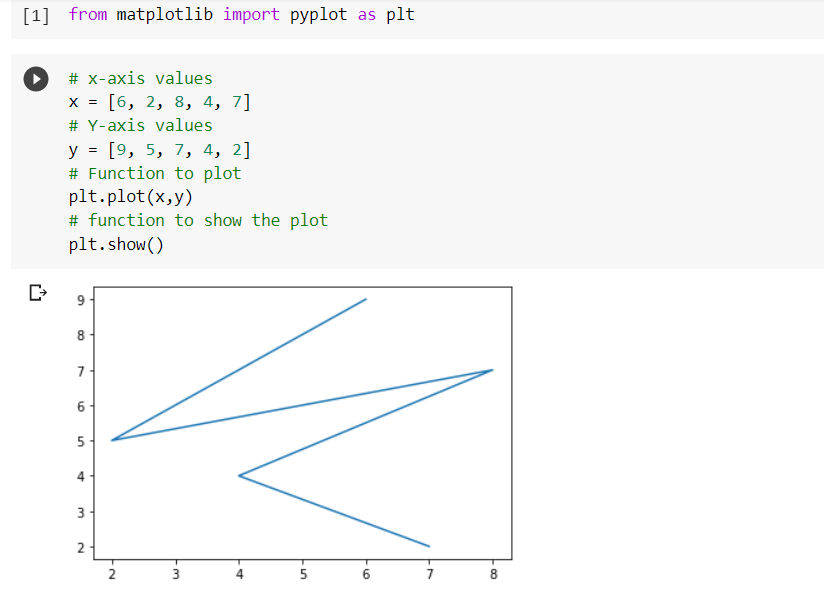
## Basic plots in Matplotlib :

Matplotlib comes with a wide variety of plots. Plots helps to understand trends,

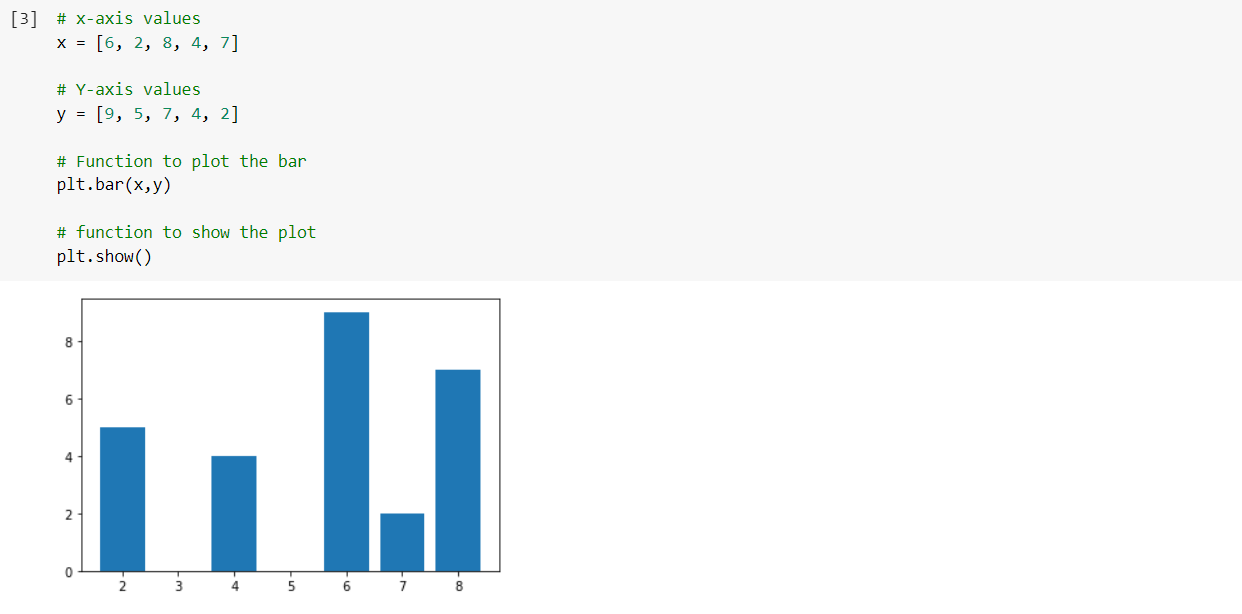
patterns, and to make correlations. They’re typically instruments for reasoning

about quantitative information. Some of the sample plots are covered here.

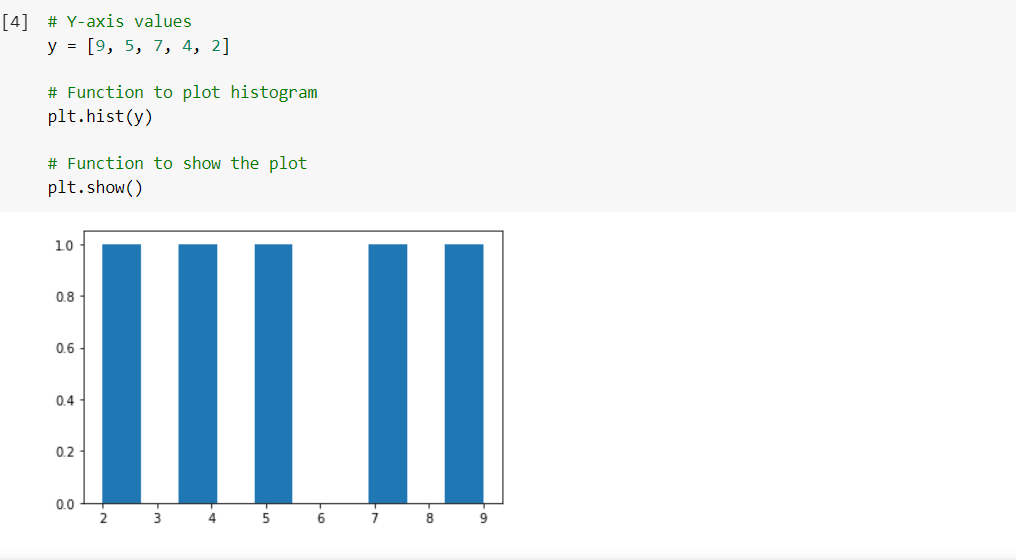
**Line plot :**



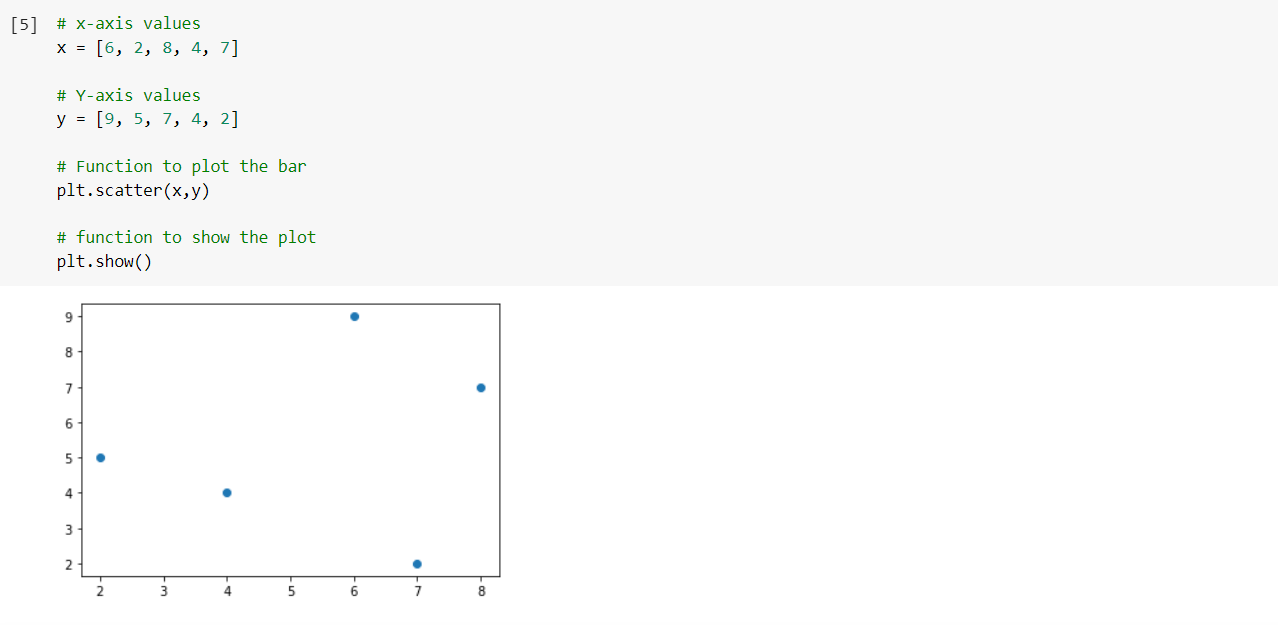
**Bar plot:**



**Histogram :**



**Scatter Plot:**



# Seaborn:

Seaborn is an amazing visualization library for statistical graphics plotting in

Python. It provides beautiful default styles and color palettes to make

statistical plots more attractive. It is built on the top of [**matplotlib**](https://www.geeksforgeeks.org/python-introduction-matplotlib/) library

and also closely integrated to the data structures from pandas. Seaborn

aims to make visualization the central part of exploring and understanding

data. It provides dataset-oriented APIs, so that we can switch between

different visual representations for same variables for better understanding

of dataset. As per definition Seaborn also aims for data visualization, the

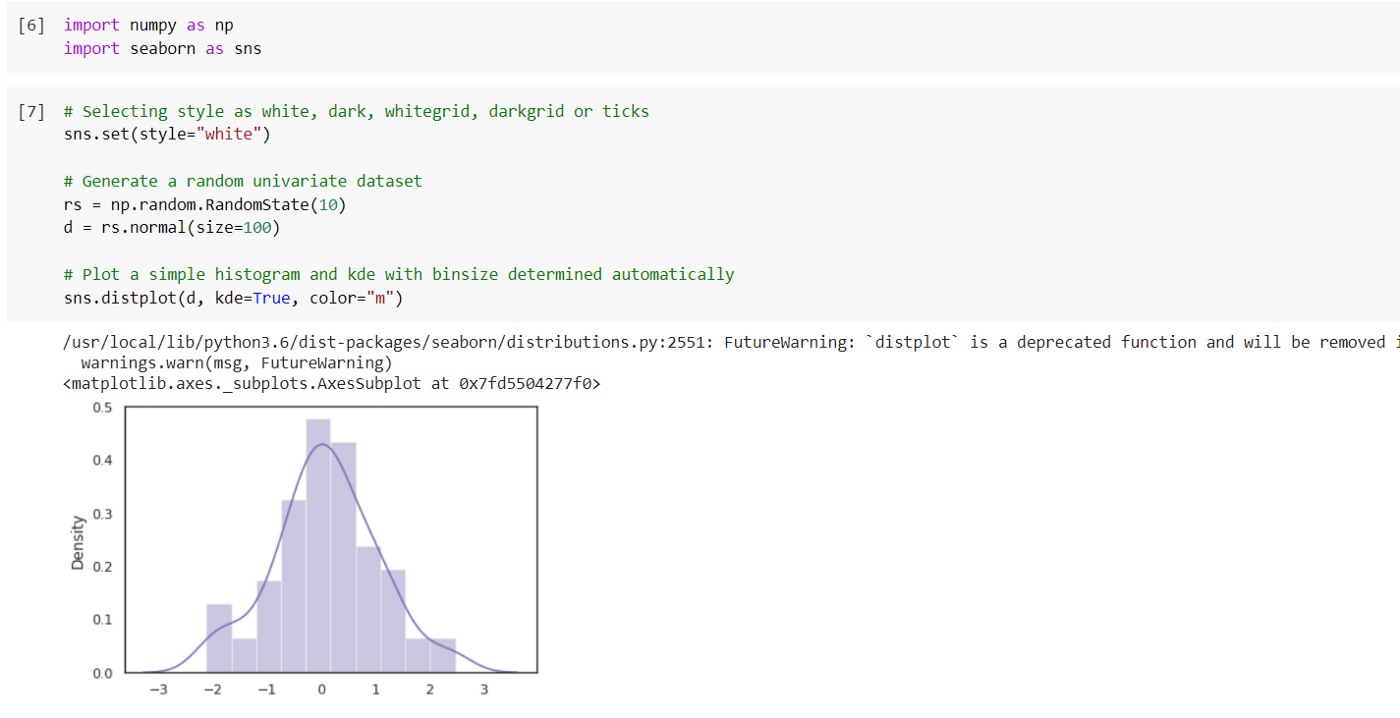
major difference is it aims for central part of exploring and understanding data.

## Some basic plots using seaborn:

## Dist plot:

## Seaborn dist plot is used to plot a histogram, with some other variations like

## kdeplot and rugplot.

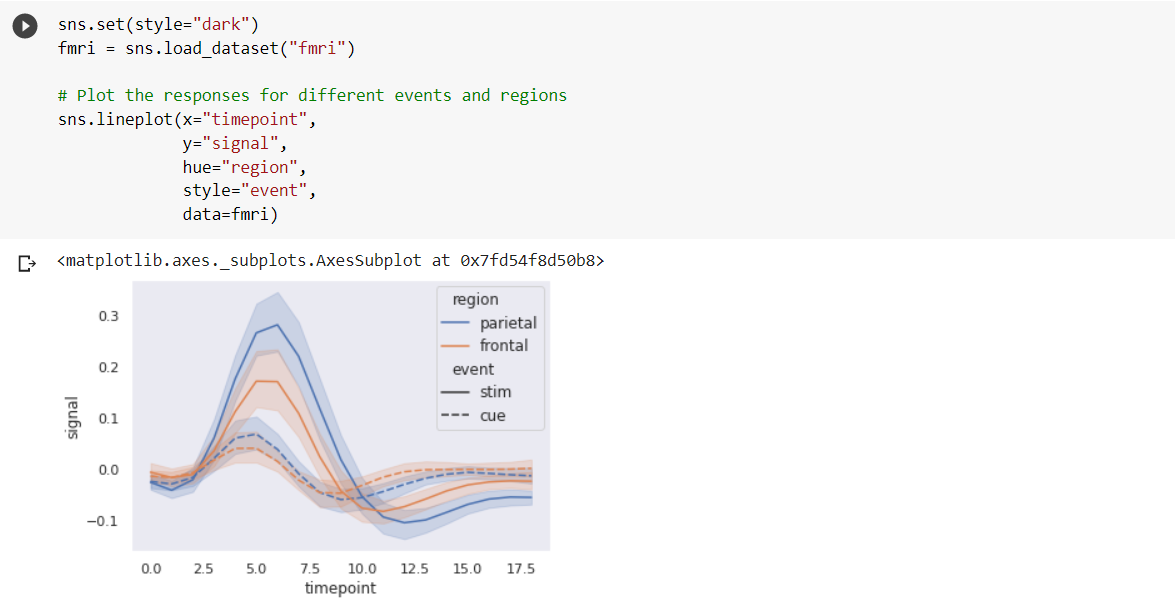


**Line plot:**

The line plot is one of the most basic plots in seaborn library. This plot is

mainly used to visualize the data in form of some time series, i.e., in

continuous manner.

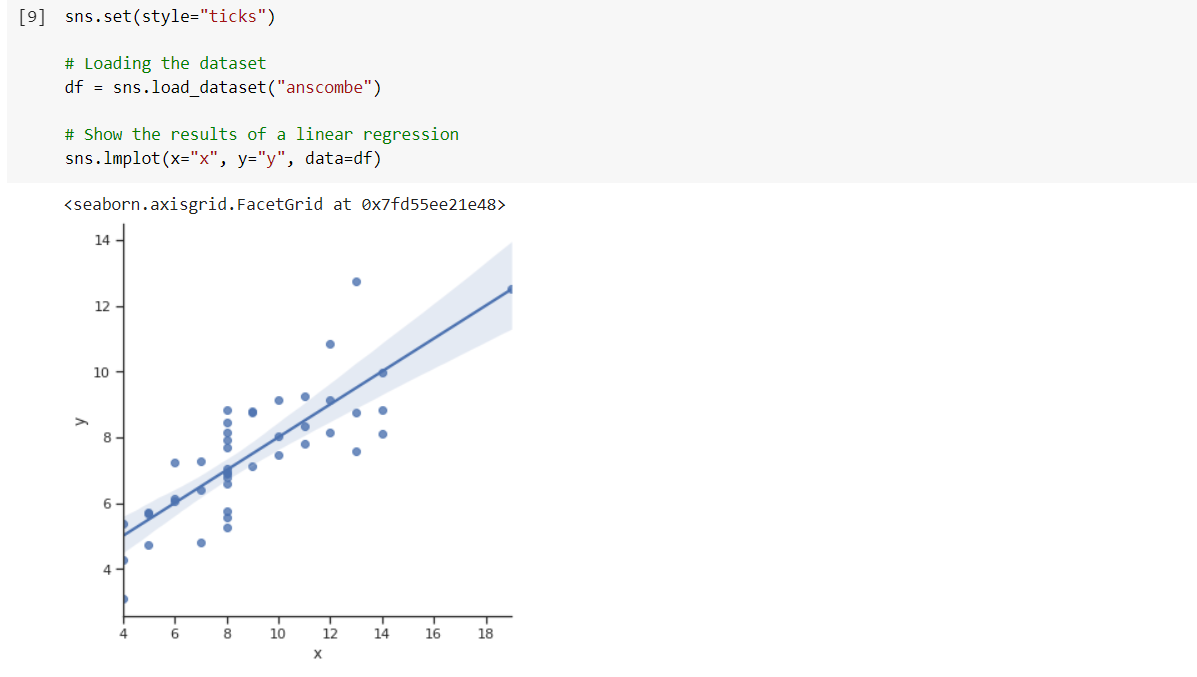


**Lmplot :**

The lmplot is another most basic plot. It shows a line representing a linear

regression model along with data points on the 2D-space and x and y can be

set as the horizontal and vertical labels respectively.



**Pickle:**

The pickle module implements serialization protocol, which provides an ability to

save and later load Python objects using special binary format.

If you are using **anaconda navigator**, follow below steps to download required

packages:

* Open the anaconda prompt.
* Type “pip install jupyter notebook” and click enter.
* Type “pip install spyder” and click enter.
* Type “pip install numpy” and click enter.
* Type “pip install pandas” and click enter.
* Type “pip install matplotlib” and click enter.
* Type “pip install seaborn” and click enter.
* Type “pip install sklearn” and click enter.
* Type “pip install Flask” and click enter.

If you are using Pycharm IDE, you can install the packages through the command

prompt and follow the same syntax as above.